

Statistik in der Archäologie

Frank Siegmund



Statistik in der Archäologie:  
eine anwendungsorientierte Einführung  
auf Basis freier Software

Frank Siegmund

(Vers. 0.9)

Pre-Print der Febr. 2020 im Druck erschienenen Fassung

[ Impressum ]

Lizensierung: CC BY

Skripte, Beispiel- und Übungsdateien auf der Website des Autors:

[www.frank-siegmund.de](http://www.frank-siegmund.de)

Direkter Link:

[https://frank-siegmund.de/images/opendata/2020-archaeostatistik/Siegmund\\_2020\\_Daten.zip](https://frank-siegmund.de/images/opendata/2020-archaeostatistik/Siegmund_2020_Daten.zip)

## Inhaltsverzeichnis

1	Vorwort.....	15
2	Das didaktische Konzept.....	16
3	Konventionen.....	18
4	Teil I: Einleitung.....	19
4.1	Warum „Statistik in der Archäologie“?.....	19
4.2	Was heißt hier „Archäologie“?.....	21
4.3	Warum freie Software?.....	21
4.4	Beschreibende und schließende Statistik.....	23
5	Über den Aufbau des Buches.....	24
5.1	Über die Sprache: deutsch oder englisch.....	27
5.2	Was kaum oder nicht behandelt wird.....	27
6	Skepsis gegenüber der Anwendung statistischer Methoden.....	28
6.1.1	Einwand: zu wenig Zahlen.....	30
6.1.2	Einwand: Statistik ist (zu) kompliziert.....	30
6.1.3	Einwand: mir fehlen die Programme.....	31
7	Ein kurzer Rückblick in die Forschungsgeschichte: die Ko-Evolution von angewandter Statistik und Archäologie.....	31
7.1.1	Die Beginne in der Archäologie der 1970er und -80er Jahre.....	32
7.1.2	Die 1980er und 1990er Jahre.....	38
7.1.3	Die 1990er und 2000er Jahre.....	40
7.1.4	Neue Vielfalt und Verschiebungen in den 2000ern.....	43
7.1.5	Stagnation in den 2010er Jahren?.....	44
8	Erster Schritt: persönliche Bedürfnisse und Möglichkeiten klären.....	47
8.1	„mal schnell ein paar Statistiken bekommen“.....	47
8.2	Öfter einfache Statistiken: Tabellenkalkulation.....	47
8.3	Ernsthaft auch komplexere Aufgaben lösen, ohne Profi werden zu wollen.....	48
8.4	Ernsthaft und nachhaltig selbstständig arbeiten.....	48
8.5	Empfehlungen für das allmähliche Sich-Vortasten.....	49

9	Teil II: Werkzeuge und Grundlagen kennenlernen und auswählen.....	51
9.1	Tabellenkalkulation LO-Calc .....	51
9.1.1	LibreOffice Calc: Installation und erste Schritte.....	52
10	Datenerfassung und -verwaltung mit einer Tabellenkalkulation .....	53
10.1	Übliche Anfängerfehler.....	53
10.1.1	Praxistipp: Tabelle einrichten & Kopfzeile fixieren .....	56
10.1.2	Praxistipp: Datenformular statt Tabelle zu Dateneingabe nutzen .....	56
10.1.3	Frage: ...oder stattdessen mehrere Erfassungstabellen anlegen? .....	57
10.1.4	Praxistipp: zwei Tabellen verbinden.....	58
10.1.5	Hinweis: Epi Info .....	59
11	„Skalenniveau“: über unterschiedliche Arten von Informationen .....	60
11.1	Nominalskala .....	61
11.2	Ordinalskala (Rangskala).....	61
11.3	Quantitative Skala .....	61
11.4	Exkurs Farbbestimmungen.....	62
12	Einen ersten Datensatz erfassen .....	63
12.1.1	Rechnen mit einer Tabellenkalkulation.....	65
12.2	Einfache Datenverwaltung und erste Datensichtung mit LO-Calc .....	67
12.2.1	Sortieren .....	67
12.2.2	Filtern.....	67
12.2.3	Pivot-Tabellen und erste Grafiken .....	71
12.2.4	Zweidimensionale Pivot-Tabelle.....	76
12.2.5	Pivot-Tabelle für eine Variable mit quantitativer Skala.....	78
12.2.6	PS: Pivot-Tabellen bei MS-Excel.....	78
12.3	Datenfehler und erste Fehlerprüfung.....	79
12.3.1	Fehler bei Daten in einer Nominalskala .....	80
12.3.2	Fehler bei Daten in einer Verhältnisskala.....	81
12.4	Vertiefte Fehlerprüfung .....	81
12.5	Beobachterfehler – „Observer Error“ .....	82
12.6	Pilotstudie.....	85
12.7	Datensicherung .....	85
13	Über Namen und Messgenauigkeiten .....	87

13.1.1	Namensvergabe: Namen, Label, Etiketten .....	87
13.1.2	Messgenauigkeit.....	88
13.2	Dokumentation und Datenbeschreibung.....	89
13.3	Beispieldatensätze.....	90
13.3.1	[1] Geldbörsen 1989 ff.: „Muenzen“ .....	90
13.3.2	[2] Neugeborene aus Braunschweig.....	91
13.3.3	[3] Anthropologie Gräberfeld Zwentendorf (Niederösterreich) .....	91
13.3.4	[4] Howells' world data .....	92
13.3.5	[5] Anatomieleichen aus Lyon: Rollet 1888 .....	94
13.3.6	Dateiformate .....	94
14	Freie Statistik-Programme .....	95
14.1	PAST.....	96
14.1.1	PAST: Installation und erste Schritte .....	96
14.1.2	Datenimport nach PAST: Frust vermeiden.....	98
14.1.3	Schnell ein paar Statistiken und Grafiken mit PAST .....	99
14.1.4	Feintuning der Datentabelle unter PAST .....	100
14.1.5	Datenmanagement mit PAST .....	101
14.1.6	Bilanz zu PAST.....	102
14.2	JASP und JAMOVI.....	102
14.2.1	JASP: Installation und erste Schritte.....	103
14.2.2	JAMOVI: Installation und erste Schritte .....	104
14.2.3	Bilanz JASP und JAMOVI .....	108
14.3	Kurz vorgestellt: SOFA, Easy R, Epi Info und SPSS.....	108
14.3.1	SOFA – Statistics open for all .....	108
14.3.2	EZR – Easy R.....	109
14.3.3	Epi Info.....	110
14.3.4	Das Bezahlprogramm SPSS .....	110
14.4	Datenbanken .....	111
15	R als Arbeitsumgebung.....	113
15.1	Graphische Benutzeroberflächen (GUIs) zu R.....	115
15.2	Dokumentation und allgemeine Hilfe zu R.....	116
15.3	R installieren.....	117
15.4	Erste Schritte in R .....	117
15.4.1	Rechnen mit R, mathematische und logische Funktionen .....	118

15.4.2	Beispiel: Logarithmus rechnen & „Umkehrfunktion“ .....	120
15.4.3	Zuweisungen in R: Objekte und Objektnamen.....	121
15.4.4	R-Kommandozeilen .....	123
15.4.5	Der Workspace .....	124
15.4.6	Vektoren.....	125
15.4.7	Matrizen.....	126
15.4.8	Datentypen.....	127
15.4.9	Funktionen.....	128
15.5	RStudio .....	129
15.6	Das Installieren von „Paketen“ in R .....	134
15.6.1	Ein Paket erstmals beziehen .....	134
15.6.2	Installierte Pakete bei Bedarf laden .....	135
15.7	Den R-Commander kennenlernen.....	136
15.7.1	... und wenn's nicht klappt mit dem Datenimport? .....	137
15.7.2	Hinweis: Editor Notepad++ .....	138
15.7.3	Erste Schritte jenseits des Datenimports im R-Commander.....	139
15.7.4	Hinweis: Stabilität des R-Commanders .....	140
15.7.5	Hinweis: R-Commander um weitere Funktionen erweitern .....	140
15.7.6	Nachtrag: was bedeutet „in Faktoren wandeln“?.....	141
15.8	BlueSky Statistics – eine gute Alternative zum R-Commander .....	141
15.8.1	BlueSky installieren.....	142
15.8.2	BlueSky: erste Schritte.....	143
15.9	Die R-interne Hilfe .....	144
15.10	Zwischenbilanz R, RStudio, R-Commander, BlueSky .....	145
15.11	Swirl, YouTube u. a. ....	146
16	Teil III: Einführung in die angewandte Statistik .....	147
17	Die grafische Darstellung von Daten.....	147
17.1	Empfehlungen zu statistischen Grafiken.....	147
17.1.1	Schätzung der empirischen Dichtefunktion.....	149
17.2	Beispiel-Grafiken .....	150
17.3	Balkendiagramm und Kreisdiagramm .....	151
17.3.1	Einschub: die gleichen Prozeduren mit BlueSky.....	157
17.4	Histogramm .....	158

17.4.1	Praxis PAST: Einstellen von Klassenweite und Startwert bei Histogrammen 161	
17.4.2	Praxis R-Commander: Einstellen von Klassenweite und Startwert bei Histogrammen .....	163
17.4.3	Weitere Histogramm-Varianten.....	167
17.4.4	Datenreihen im Balkendiagramm vergleichen .....	169
17.4.5	Datenreihen vergleichen durch verbundene Histogramme ( <i>facets</i> ).....	171
17.5	Streuungsdiagramm: Zwei Variablen im Zusammenhang .....	174
17.5.1	Praxis PAST: Anlegen eines Streuungsdiagramms.....	178
17.5.2	Praxis R-Commander: Anlegen eines Streuungsdiagramms.....	178
17.6	Grafiken zum Zusammenhang von mehr als zwei Variablen: Dreiecksdiagramm, Streuungsdiagramm-Matrix .....	180
17.7	Allgemeines zu den Grafik-Möglichkeiten in R.....	183
17.7.1	ggplot2-Graphiken mit dem R-Commander.....	184
18	Häufigkeiten, Prozentzahlen, Häufigkeiten vergleichen ... ..	185
18.1.1	Rundungsregeln .....	187
18.2	Prozentzahlen und ihre Zuverlässigkeit .....	187
18.3	Konfidenzintervalle.....	188
18.3.1	Berechnen von Konfidenzintervallen mit R.....	191
18.3.2	Asymmetrie der Konfidenzintervalle.....	192
18.3.3	Berechnen von Konfidenzintervallen mit PAST.....	193
18.3.4	Caveat: endliche Grundgesamtheiten .....	193
18.3.5	Konfidenzintervall: Anwendungsbeispiel .....	194
18.4	Kontingenztafeln (alias Kreuztabellen, Mehrfeldertafeln) und das Chi- Quadratverfahren .....	195
18.4.1	Kritischer Wert von Chi-Quadrat und Chi-Quadrat-Test .....	200
18.4.2	Zwei wichtige Begriffe: Freiheitsgrad und Irrtumswahrscheinlichkeit .....	202
18.4.3	Nullhypothese .....	203
18.4.4	Zwischenbilanz Chi-Quadrat-Test.....	203
18.4.5	Chi-Quadrat-Test praktisch rechnen .....	204
18.4.6	Der p-Wert, oder: die Sache mit der Signifikanz .....	206
18.4.7	Chi-Quadrat-Verfahren: Schwach besetzte Tafeln (I) .....	207
18.4.8	Exakter Test nach Fisher .....	208
18.4.9	Chi-Quadrat-Verfahren: Schwach besetzte Tafeln (II) .....	211
18.4.10	Testvoraussetzung Binomial- oder Poisson-Verteilung.....	212
18.4.11	Informationsstatistik.....	213

18.4.12	Chi-Quadrat signifikant: und nun? – Konfigurationsfrequenzanalyse (KFA) 213	
18.4.13	Chi-Quadrat-Test und große Grundgesamtheiten.....	214
18.4.14	Praxistipp: Tests.....	215
18.5	Chi-Quadrat-Krise, Mindestindividuenzahl, und warum die Prüfung der Verteilung so wichtig ist .....	215
19	Analyse von Messwerten: Mittelwert, Standardabweichung ... ..	218
19.1.1	Begriffsklärung unimodal, bimodal, multimodal.....	218
19.2	Zentralmaße: Modalwert, Median, Mittelwert .....	221
19.2.1	Zentralmaße selbst gerechnet.....	221
19.3	Streuungsmaße: Interquartilabstand, Standardabweichung.....	223
19.3.1	Minimum, Maximum, Spannweite.....	223
19.3.2	Interquartilabstand (IQR).....	224
19.3.3	Standardabweichung .....	225
19.3.4	Quantile als robustere Alternativen zu Minimum und Maximum .....	226
19.3.5	Kurze Zusammenschau .....	227
19.3.6	grafische Darstellungen, u. a. als Selbstkontrolle .....	227
19.4	Praxis PAST: Zentral- und Streuungsmaße mit PAST .....	229
19.5	Praxis R: Zentral- und Streuungsmaße – am Beispiel Zwentendorf .....	231
19.6	„Jackknifing“.....	233
20	Stetige Wahrscheinlichkeitsverteilungen, Normalverteilung, Vergleich von Mittelwerten 235	
20.1	Normalverteilung .....	237
20.2	Schiefe und Wölbung .....	239
20.2.1	Schiefe.....	239
20.2.2	Wölbung und Exzess .....	239
20.2.3	Praxis: Schiefe und Wölbung.....	240
20.3	Ist eine Variable normalverteilt?.....	242
20.3.1	Simulierte Daten als Vergleich .....	243
20.3.2	Test auf Vorliegen einer Normalverteilung.....	243
20.3.3	Praxis: KS-Test auf Vorliegen einer Normalverteilung mit R .....	244
20.3.4	Praxis PAST: KS-Test auf Vorliegen einer Normalverteilung .....	247
20.3.5	QQ-Plot – wie sehen die Abweichungen von der Normalverteilung aus? .....	248
20.3.6	Option Variablen-Transformationen .....	249

20.4	Zurückgefragt: und wozu der Stress mit der Normalverteilung?	250
20.4.1	Vorab: Prozentwerte und Verhältniszahlen sind nicht normalverteilt	251
21	Der Vergleich von Mittelwerten: Vorüberlegungen und Auswahl des geeigneten Tests	252
21.1.1	parametrische / nicht-parametrische Verfahren	252
21.1.2	unabhängige Stichproben / abhängige Stichproben	253
21.1.3	einseitige oder zweiseitige Fragestellung	253
21.1.4	$n = 2$ oder $n > 2$	253
21.2	Testverfahren auf Mittelwertunterschiede	254
21.2.1	t-Test	254
21.2.2	einfache / einfaktorielle Varianzanalyse	255
21.2.3	U-Test (nach Wilcoxon, Mann & Whitney)	255
21.2.4	Kolmogorov-Smirnov-Test für zwei Stichproben	255
21.2.5	Rangvarianzanalyse nach Kruskal & Wallis (auch: H-Test)	256
21.2.6	Praxis R: Mittelwertvergleich zweier Stichproben	256
21.2.7	Rekapitulation Mittelwertvergleiche	262
22	Einschub: Sampling und Bayes-Statistik	262
22.1	Sampling (Bootstrapping)	263
22.2	Bayes-Statistik	265
23	Über den Zusammenhang von zwei Messwerten: Korrelation und Regression	265
23.1	Begriffsklärung Korrelation und Regression	265
23.2	Korrelation	266
23.2.1	Begriffsklärungen	267
23.2.2	Der Wert des Korrelationskoeffizienten	267
23.2.3	Korrelationskoeffizient $r$	267
23.2.4	Rangkorrelationskoeffizient $(\rho, \tau)$	268
23.2.5	Kontingenzkoeffizienten C und K	268
23.2.6	$\Phi$ -Koeffizient	268
23.2.7	Yule-Koeffizient	268
23.2.8	Korrelationskoeffizient und Signifikanz	269
23.2.9	Praxis R: Korrelation	269
23.3	Regression	273
23.3.1	Lineare oder nicht-lineare Regression?	274
23.3.2	Lineare Regression	274
23.3.3	Wie jetzt: zwei Regressionsgeraden?	274

23.3.4	Praktisches Vorgehen und Ausblick.....	275
23.3.5	„Residuen“: der Vergleich beobachtet vs. erwartet .....	276
23.3.6	Praxis R: lineare Regression .....	277
23.3.7	Erweiterung der linearen Regression.....	279
23.3.8	RMA – Reduced major axis regression.....	280
23.3.9	Logistische Regression.....	284
24	(Diskrete) Wahrscheinlichkeitsverteilungen .....	287
24.1	Gleichverteilung .....	288
24.2	Poisson-Verteilung.....	289
24.3	Binomialverteilung.....	290
24.4	Negative Binomialverteilung.....	291
25	Teil IV: Anwendungsbeispiele .....	292
25.1	Mehr Genauigkeit mit einfachem GPS .....	293
25.1.1	Anlass, Fragestellung.....	293
25.1.2	Erste Eindrücke .....	295
25.1.3	Begriffsklärung Genauigkeit: Präzision und Richtigkeit .....	296
25.1.4	praktische Anwendung der Präzision (DOP) .....	297
25.1.5	Fehlerangaben: Standardabweichung und RMS .....	297
25.1.6	Signale, Geräte und Prozessoren .....	299
25.1.7	WAAS und EGNOS.....	300
25.1.8	Empirie & Praxis: Wie genau misst mein GPS?.....	300
25.1.9	Statische Messungen.....	302
25.1.10	Lokale Messnetze mit Netzausgleichsrechnungen.....	303
25.2	Kulturmodelle und Identitäten quantifiziert .....	307
25.3	Saxe: Beispiel für eine Metrisch gestützte Typologie.....	311
25.4	Demographie Zwentendorf .....	322
25.4.1	Datensichtung und -aufbereitung .....	322
25.4.2	Analyse der Geschlechtszusammensetzung.....	325
25.4.3	Alterszusammensetzung I: der Kinderanteil.....	327
25.4.4	Alterszusammensetzung II: Demographie der Kinder .....	328
25.4.5	Alterszusammensetzung III: Sterbetafeln (life tables).....	330
25.4.6	Alterszusammensetzung IV: Demographie der Erwachsenen .....	334
25.4.7	Schätzung der Lebendbevölkerung .....	338
25.5	Soziale Fragestellung am Beispiel Sandhofen .....	339
25.5.1	Theorie, Modelle, Ansatzpunkte, Gegenwartsbezug.....	339
25.5.2	Datensichtung und -aufbereitung .....	342

25.5.3	Vergleich der Befunde mit urgeschichtlichen und modernen Gesellschaften 350	
25.5.4	Literatur zum Kapitel .....	352
25.6	Regression: Körperhöschätzung aus Langknochen.....	353
25.6.1	Planen vor dem Handeln .....	353
25.6.2	Verfahren-Auswahl, Prüfen der Voraussetzungen, Vorgehen.....	354
25.6.3	Zwei Zufallsstichproben bilden: Referenzserie und Testserie .....	356
25.6.4	Lineare Regression anhand der Referenzserie .....	358
25.6.5	Anwendung des Regressions-Ergebnisses auf die Testserie.....	360
25.6.6	Regression: Ausreißer, Gütekriterien, Optimierungen .....	360
25.6.7	Lineare Regression mit Resampling .....	366
25.6.8	Prüfung anhand externer Testserie .....	369
25.6.9	Fallstudie Multiple Lineare Regression.....	371
25.7	Klimadaten: Eisbohrkerne u.a.....	376
25.7.1	Woher die Daten kommen: Archive für Klimadaten .....	376
25.7.2	Das $\delta^{18}O$ -Verhältnis seit 12.000 v.H. als Schätzer der Paläotemperatur ...	378
25.7.3	Quellen / Archive zum Kapitel .....	387
25.7.4	Literatur zum Kapitel .....	388
26	Schlusswort.....	389
27	Kommentierte Literaturlisten.....	389
27.1	Vertiefende Literatur zu R.....	389
27.2	Hinweise für Freunde des Online-Lesens.....	392
27.2.1	Web 1.0 .....	392
27.2.2	Twitter.....	394
27.3	Foren.....	394
27.4	Persönlich, von Mensch zu Mensch .....	394
27.5	Nachschlagewerke Statistik sowie Archäologie & Statistik.....	395
27.6	Literatur zum Thema grafische Darstellungen.....	397
28	Nachweis der zitierten Literatur .....	398
29	finis.....	400
30	Stichwortverzeichnis .....	401



# 1 VORWORT

---

Dieses Buch ist für den Autor eine Etappe in einem bereits lange währenden „Work in Progress“. Es entstand aus Unterlagen und diversen Skripten zu meinen Lehrveranstaltungen über Statistik in der Archäologie, die ich seit dem Wintersemester 1999/2000 im Abstand von jeweils ein bis zwei Jahren an verschiedenen Orten und für unterschiedliche Gruppen und Niveaus durchgeführt habe, zunächst auf der Basis der Software SPSS für die Statistik und Datenverwaltung sowie auf der Basis von WinBASP für die Seriation und Korrespondenzanalyse. Die nunmehrige Fassung ist systematischer und vollständiger als die bislang eher *ad hoc* für die jeweiligen Lehrveranstaltungen erzeugten Materialien und ersetzt hinsichtlich der Software das (gute) Bezahlprogramm SPSS durch die Arbeit mit dem kostenlosen Programm PAST und der Open-Source-Software **R**. Trotz der Reifezeit wird auch diese Fassung nicht perfekt sein, Fehler beinhalten und vor allem schnell wieder altern: unvermeidbar in einer Zeit schneller Evolution. Wobei es seit den 1980er Jahren mehr die technischen Geräte und die Software sind, die sich schnell entwickeln, weniger die Methoden, d. h. die Statistik und das fruchtbare Zusammenwirken von Archäologie und Statistik. Die von mir für dieses Skript gewählte Buchform entspricht meinen persönlichen Arbeits- und Lesepräferenzen; so anregend und hilfreich auf Youtube eingestellte Videos punktuell auch sein mögen, für ein systematisches Arbeiten und auch das immer wieder notwendige Nachschlagen, Notieren und Blättern erscheint mir ein Buch zielführender.

Diese Einführung ist als erster Schritt gedacht. Für tiefer Interessierte könnte man danach die hier bewusst verfolgte Parallelspurigkeit mehrerer Programme beenden und sich auf **R** konzentrieren. Wesentliche nächste Themen wären mir dann eine Vertiefung des kommandozeilen-orientierten Umgangs mit **R** (Leseempfehlung: Kabacoff 2015), das nicht nur in der Archäologie sehr wichtige Thema Datenmanagement mit **R** (Leseempfehlung: Wickham & Golemund 2018) und das Verfertigen moderner Grafiken mit dem **R** Paket *ggplot2* (Handbuch: Chang 2013). Das vierte Vertiefungsthema wären multivariate Verfahren wie z. B. Hauptkomponenten- und Faktorenanalyse, Clusteranalyse etc. sowie das für die Archäologie so wichtige Feld Seriation und Korrespondenzanalyse. Multivariate Verfahren kommen zwar im Forschungsalltag deutlich seltener zum Einsatz als das in diesem Buch vermittelte Basiswissen, aber zumindest das sichere Lesen-, Verstehen- und Interpretieren-Können solcher Analysen ist auch in der Archäologie wichtig. Diese Skizze macht zugleich deutlich, was die vorliegende Einführung ausklammert. Der Umfang dieser Einführung entspricht dem, was nach meinen langjährigen Erfahrungen in der Lehre mit Anstrengung, aber ohne

Stress im Laufe eines Semesters in einer Einführungsübung erfolgreich von Studierenden der Archäologie erlernt werden kann, um dann dauerhafter Bestandteil des persönlichen Arbeitsinstrumentariums zu sein.

Ich danke allen Studierenden, die in meinen Kursen durch Mitwirken, Fragen, Verständnis und Unverständnis zum Reifen dieses Textes beigetragen haben. Als langjährige Mitwirkende an meinen Lehrveranstaltungen hat Sandra Viehmeier (Basel) wesentlich zum Reifen der seinerzeitigen Skripte und HTML-Fassungen beigetragen, auf denen diese Fassung beruht. Für ihr aufmerksames Lektorat der vorliegenden Fassung danke ich Rashida Hussein-Oglü (Münster) und Michaela Schauer (München).

## 2 DAS DIDAKTISCHE KONZEPT

---

Diese Einführung ist zum Selbststudium gedacht und vor allem an Diejenigen gerichtet, denen ein Zugang zu guten Kursen mit persönlichem Unterricht verwehrt ist. Das Buch tut so, als würden sich seine Leser ohne Vorkenntnisse in das Thema hineinbegeben – was tatsächlich nur selten der Fall sein dürfte. Ebenso geht es davon aus, dass die Nutzer das ganze Buch systematisch von Anfang bis Ende durcharbeiten wollen und die nötige Zeit dafür haben, was ebenfalls wohl eine Ausnahme bleiben dürfte. Daher ist der Text stark gegliedert. Dies ermöglicht es, sich individuell für das Selbststudium überschaubare Portionen zu schnüren, die man nach eigenem Bedarf auswählt und schrittweise durcharbeitet. Es ermöglicht zugleich allen bereits etwas Eingearbeiteten, das ihnen bereits Bekannte schnell zu identifizieren, es zu überblättern und nur die ihnen relevant erscheinenden Kapitel genauer durcharbeiten.

Die Idee dieser Einführung ist es, dass Leser, die dies wollen, möglichst schnell arbeitsfähig werden. Die überwiegende Mehrheit archäologischer Fragestellungen und Probleme lässt sich mit einem durchaus übersichtlichen Set von Instrumenten lösen. Auch wer im Alltag schlicht von A nach B gelangen will, wird dies meist mit einem einfachen Instrument wie Zu-Fuß-Gehen, Fahrrad-Fahren oder dem Einsatz eines einfachen PKWs tun, wiewohl man mit High-Tech-Einsatz möglicherweise schneller zum Ziel käme. Doch nur wer wirklich komplexe Fragestellungen und Daten hat, braucht auch mächtigere Werkzeuge. Das komplexeste der hier verwendeten Software-Werkzeuge ist gewiss **R**. Es ist sehr vielfältig, flexibel und kann ganz unterschiedlich bedient werden. Um das Buch, aber auch Ihr Lernen zu vereinfachen, verfolge ich hierzu einen klaren Pfad, der denkbare Alternativen außen vorlässt. Bitte ignorieren

Sie in der folgenden Liste alle ggf. noch unbekanntem Begriffe, sie werden später im Text erläutert.

- Die Datenerfassung erfolgt in LibreOffice Calc (künftig: LO-Calc). Wer bereits eingehende Erfahrungen mit und Zugang zu MS-Excel hat, kann ohne große Änderungen alternativ auch damit arbeiten.
- Eine einfache Datenverwaltung erfolgt zunächst in LO-Calc.
- Die Statistik wird – nach ersten Anfängen mit LO-Calc und Hinweisen u. a. auf JAMOVI und BlueSky Statistics – parallel mit PAST und mit **R** gerechnet.
- Die Bedienung von **R** erfolgt mit Hilfe von RStudio (statt des schwierigeren Arbeitens mit dem „rohen“ **R** im Kommandozeilenmodus oder mit Alternativen zu RStudio wie z. B. JGR).
- Nach einem ersten Schnuppern an **R** mithilfe von RStudio erfolgt das weitere statistische Arbeiten jedoch über eine relativ einfach zu erlernende graphische Benutzeroberfläche (*graphical user interface*, GUI), und zwar mit dem R-Commander und gelegentlich – als mögliche Alternative zum R-Commander – mit BlueSky (anstelle von Alternativen wie z. B. Deducer, Rattle). Das Bedienen von **R** ohne graphische Benutzeroberfläche wird zwar kurz angerissen, ist aber m.E. für ein einführendes Niveau nicht notwendig und soll einem vertiefenden Werk vorbehalten bleiben.
- Zuerst geht es bei diesen Instrumenten um eine Einführung in die Grundlagen, in die einfache Standard-Statistik für Häufigkeiten inkl. Chi-Quadrat-Test, für Messwerte inkl. T-Test und Rangtests sowie um Korrelation und Regression. Mit den am Ende dieser Einführung erreichten Fertigkeiten sind etwa 90 % der in der Archäologie normalerweise auftretenden Aufgaben und Fragestellungen bearbeitbar. Zudem ist ein eingehendes, nachvollziehendes Verständnis der meisten in der Archäologie publizierten Auswertungen möglich, die mit statistischen Analysen argumentieren. Dieser Aufbau soll es auch erlauben, das Durcharbeiten des Buches an jeder Stelle zu unterbrechen oder auch abzubrechen, wenn man sein persönliches Ziel erreicht hat.

Das Buch geht von Lesern aus, die die angebotenen Übungen auch praktisch nachvollziehen. Wer indes für sein persönliches Projekt „nur“ eine Lesefähigkeit erreichen will, d. h. ein besseres Verständnis von Texten, die statistisch argumentieren und die man verstehen will, ohne selbst aktiv eigene Probleme zu verfolgen, kann es auch beim Lesen ohne praktische Übungen belassen.

Das Buch ist nicht als Nachschlagewerk, Lexikon oder Handbuch zur Statistik gedacht, im Gegenteil: Eine gute Einführung in die Statistik sollten Sie ggf. parallel zurate ziehen. Dazu finden Sie am Ende des Buchs außer der Liste der zitierten Literatur auch eine kommentierte Literaturliste, die dazu einige Vorschläge macht. Das vorliegende Buch ist vielmehr als praxisorientierte Arbeitsunterlage gedacht mit konkreten Fallbeispielen und Übungen. Nicht jede hier eingeführte Software müssen Sie auf Ihrem PC/Laptop installieren und sich darin einüben, aber das Buch geht davon aus, dass Sie mit der von Ihnen ausgewählten freien Software möglichst viele der Beispiele und Übungen ganz praktisch nacharbeiten. Nur durch Nacharbeiten wird es Ihnen gelingen, Fertigkeiten und Routine auch für die spätere Umsetzung Ihrer eigenen Projekte zu gewinnen. Später, wenn Sie an eigenen Daten und Problemen arbeiten und selbstständig gute Lösungen finden müssen, sollten Sie sich auf Ihre inhaltlichen Fragestellungen und die Wahl adäquater statistischer Methoden konzentrieren können, weil der anfängliche Kampf mit der Technik und Software für Sie nunmehr Routine geworden ist.

### 3 KONVENTIONEN

---

- **Notation von Zahlen.** Das Buch ist in deutscher Sprache geschrieben und behält die im Deutschen übliche Schreibweise von Zahlen bei: Das Komma dient als Trennzeichen bei nicht-ganzen Zahlen (also 10,5 für zehneinhalb), für hohe Zahlenwerte wird der Punkt als Tausenderzeichen eingesetzt (also 1.000 € für eintausend Euro). Die eingesetzten Programme operieren jedoch mehrheitlich in englischer Sprache, wo der Punkt statt des Kommas als Trenner für nicht-ganze Zahlen dient. Da, wo Ausgaben dieser Programme zitiert werden, behalten wir deren originale Zahlennotation bei.

Bei den üblichen Tabellenkalkulationsprogrammen wie z. B. LO-Calc ist es möglich, diese Konventionen zu wählen und einzustellen. Sie sollten dies (a) bewusst tun, und (b) sich daran erinnern, dass darin beim konkreten Arbeiten und Verwenden von Fremddaten auch ein (leicht behebbarer) Fehlerquell liegt.

- **x-Achse, y-Achse.** International werden Koordinatensysteme unterschiedlich gesehen. In Deutschland ist es üblich, bei der Waagerechten von x-Achse zu sprechen und die Senkrechte als y-Achse zu bezeichnen. In den USA ist der Gebrauch oft – aber

nicht immer – genau andersherum. Klarstellung: In diesem Buch bildet die Waagerechte die von links nach rechts verlaufende x-Achse und die Senkrechte die von unten nach oben verlaufende y-Achse.

- **Grundschrift.** Der Text des Buches ist in der Garamond gesetzt, für Fußnoten und Abbildungsunterschrift wird, klein gesetzt, die gut zu unterscheidende Schrift Calibri verwendet. Die serifenlose Monospace-Schrift Anonymous Pro wird eingesetzt, wenn Ausgaben der verwendeten Software als Zitat verwendet werden.

- **Spezielle Symbole.** Für die Bedienung der Programme werden bisweilen Anleitungen Schritt für Schritt gegeben. In diesen Anleitungen verwende ich folgende Symbol resp. Kürzel:

=> steht für „gehe zu“, in der Regel mit der Maus zu bestimmten Schaltflächen oder Eingabefeldern.

OK steht für „o.k.“, d. h. entweder das Klicken auf eine Schaltfläche OK oder die Eingabebestätigung per Enter-Taste.

## 4 TEIL I: EINLEITUNG

---

### 4.1 WARUM „STATISTIK IN DER ARCHÄOLOGIE“?

Lehrbücher und Nachschlagewerke zur Statistik gibt es zahlreiche und gute, ebenso Anleitungen zur Bedienung einschlägiger Programme (Software). Warum also ein weiteres Buch, speziell für die Archäologie? Weil ich nach vielen Jahren der Lehrerfahrung und des Coachings von Projekten und Kollegen einen Bedarf dafür sehe. Das praktische Arbeiten mit quantitativen Methoden ist für Interessierte ein komplexer Mehrkampf: mit der statistischen Theorie, mit der Software, mit der Datenerfassung und ihrer Verwaltung und nicht zuletzt mit der Archäologie. Bei einem konkreten Forschungsvorhaben gilt es gleich von Anbeginn an, diese unterschiedlichen Felder gleichzeitig „im Griff zu haben“, die isolierte Beschäftigung mit einem der Aspekte führt nicht weiter. Man muss lernen, die vorhandenen archäologischen Informationen so zu formalisieren, zu erfassen und zu verarbeiten, dass daraus auswertbare Daten entstehen. Man muss lernen, die am Ausgangspunkt stehenden archäologischen Sachverhalte und die historischen Fragestellungen so aufzufassen und zu formalisieren,

dass daraus exakte Fragestellungen werden und aus der Kombination von passenden Daten, geschickt gestellten Fragen und den geeigneten statistischen Werkzeugen und Methoden konkrete Antworten und historisch verwendbare Erkenntnisse resultieren. Meine Erfahrung zeigt, dass es für die beiden Themen Statistik und Softwarebedienung reichlich Einführungs- und Nachschlagewerke gibt und beide Bereiche anfangs, z. B. in einer einschlägigen Lehrveranstaltung, auch zügig bewältigt werden können. Das selbstständige Fragen, Denken, Auswerten und Argumentieren im Dreieck Statistik, Software und Archäologie bleibt jedoch für längere Zeit ein schwieriges Feld. Daher möchte ich mit dieser Einführung versuchen, Interessierten anhand konkreter Beispiele aus der Archäologie (und manchmal auch aus der archäologie-relevanten biologischen Anthropologie) den Einstieg in diesen Mehrkampf und in das selbstständige Arbeiten zu erleichtern. Denn die entscheidenden Fortschritte im Aufbau der nötigen Erfahrungen und auch von Selbstsicherheit entstehen am ehesten aus konkreten Anwendungen heraus. Folglich versteht sich dieses Buch ausdrücklich als anwendungsorientierte Einführung; es orientiert sich vor allem an der archäologischen Praxis und an geläufigen Problemstellungen und wird für den theoretischen Hintergrund auf gute Handbücher verweisen.

Ein weiterer Grund für ein eigenes Werk zur Archäologie ist die Weite des Feldes Statistik und auch der zugehörigen Software. Heute ist „die Statistik“ (und deren Schwester Informatik) eine selbstständige Wissenschaft, die man z. B. auch als solche studieren kann; sie steht für ein gereiftes und sehr vielfältiges Bündel an Verfahren, Methoden und auch an unterschiedlichen Schulen und Stilen. Doch nicht alles aus diesem breiten Spektrum wird in der Praxis von Archäologie und Anthropologie auch benötigt. Vielmehr stehen in der Archäologie, wie bei anderen Wissensbereichen auch, bestimmte Arten von Fragestellungen und bestimmte Arten von Informationen im Vordergrund, während anderes nicht oder nur seltener relevant ist. Das ermöglicht es für diese Einführung, auf große Bereiche des Feldes Statistik zu verzichten und andererseits einige Themen gründlicher zu entwickeln, die außerhalb der Archäologie vielleicht weniger bedeutsam sind. So spielen z. B. in der Archäologie – wie auch in den Sozialwissenschaften – Häufigkeiten und deren Vergleich eine besonders große Rolle, während z. B. der in der Medizin wichtige Vergleich von Medikamentendosen mit multiplen Messwerten (Patienten-Reaktionen) in der Archäologie eine eher seltene Fragestellung ist. So werden andere Wissensbereiche aufgrund ihrer spezifischen Daten und Fragestellungen wiederum oft andere Bereiche der Statistik mehr betonen müssen. Gute Lehrbücher und wertvollen Rat außerhalb der Archäologie, der oft auf

archäologische Themen gut anwendbar ist, findet man vor allem bei den Sozialwissenschaften und im Bereich Biologie/Ökologie.

## **4.2 WAS HEIßT HIER „ARCHÄOLOGIE“?**

Der Begriff Archäologie umfasst ebenfalls ein weites Feld. Statistik kommt in der Archäologie vorwiegend dort zum Tragen, wo relativ zahlreiche Informationen verfügbar sind, Häufigkeiten und das Erkennen von Regel und Ausnahme eine große Rolle spielen, die einzelnen Informationsträger wie z. B. Artefakte jedoch relativ informationsarm sind. Jene Bereiche der Archäologie, die sich intensiv mit einzelnen, oft exzeptionellen (Kunst-) Werken der Antike beschäftigen wie z. B. griechischen Statuen oder monumentalen Bauten wie Tempeln oder Theatern, argumentieren selten mit Statistik. Daher steht der Begriff „Archäologie“ hier vor allem für die Themenfelder prähistorische Archäologie, Ur- und Frühgeschichte, Frühmittelalterarchäologie und Archäologie des Mittelalters, während Beispiele und Problemstellungen z. B. aus der Klassischen Archäologie oder der Ägyptologie keine Rolle spielen.

Gelegentlich wird hier auch auf die „Anthropologie“ und Beispiele aus diesem Feld ausgegriffen, und zwar im Sinne von Biologischer Anthropologie alias Physischer Anthropologie, also jenem Bereich der Anthropologie, der sich vorwiegend anhand von archäologisch überlieferten Menschenknochen und der an ihnen ablesbaren Informationen (Maße und Merkmale) mit den Menschen, ihrem Leben und ihrer Geschichte beschäftigt. Das zunehmend wichtige Feld der Untersuchung menschlicher DNA bewegt sich auch hinsichtlich der Statistik in einem ganz anderen Methodenspektrum. Archäologie und Anthropologie im skizzierten Sinne sind enge Nachbarn; bei der Analyse von ur- und frühgeschichtlichen Gräberfeldern wirken sie eng zusammen. Daher sollten Archäologen über Grundkenntnisse der Auswertung physisch-anthropologischer Daten verfügen resp. vorliegende Auswertungen lesen und einordnen können, wie auch das Umgekehrte gilt: die physisch-anthropologische Analyse eines urgeschichtlichen Gräberfeldes ist ohne den Blick auf die Archäologie ärmer.

## **4.3 WARUM FREIE SOFTWARE?**

Für die Auswertung von Daten stehen heute elaborierte und wirklich gute „Statistikpakete“ bereit wie z. B. SPSS oder SAS, die an den Universitäten weit verbreitet sind und über eine Campuslizenz den Studierenden, Examenskandidaten und Angestellten kostenlos oder sehr preiswert zur Verfügung gestellt werden. Folglich gibt es an vielen Universitäten auch regelmäßig gute Kurse dazu. Eigentlich läge es nahe, darauf auch für dieses Buch zurückzugreifen, umso mehr, als der Autor seit den späten 1970er

Jahren umfangreiche eigene Erfahrung im Arbeiten mit SPSS hat, erst als Anwender, dann als Berater und seit dem Wintersemester 1999/2000 auch formal als Lehrender. Doch außerhalb der Welt von Forschung und Lehre sind diese Bezahlprogramme (alias proprietäre Software) relativ teuer und in der staatlichen Archäologie und Denkmalpflege nicht weit verbreitet.<sup>1</sup> Zwar begann spätestens mit der Einführung des „Personal Computers“ (IBM-PC) im August 1981 der sukzessive Einzug von Computern am persönlichen Arbeitsplatz der Wissenschaftler auch in den Museen und den Fachämtern für die Archäologie. Doch bis heute geht es dort vorwiegend um Büro- und Kommunikationsaufgaben, die PC-gestützt ausgeführt werden (erste E-Mail in Deutschland im August 1984, zunehmend weite Verbreitung seit 1988; Internet in Deutschland seit 1991/93). Die ehemals händische Verwaltung von Foto- und Planarchiven oder (Boden-) Denkmälerverzeichnissen erfolgt heute nahezu flächendeckend mit Datenbanken am Computer. Für die Handhabung von Karten und die Verschneidung geografischer und archäologischer Informationen werden „Bodendenkmälerdatenbanken“ und GIS-Systeme genutzt. Auch für die Foto- und Planbearbeitung stehen zumindest den in den Ämtern damit betrauten Spezialisten geeignete Werkzeuge zur Verfügung wie z. B. Photoshop und Auto-CAD. Die statistische Analyse der vorhandenen Informationen ist jedoch nur zögernd und deutlich seltener in den Museen, Denkmalämtern oder der privatwirtschaftlichen Archäologie zur alltäglichen Praxis geworden. So werden z. B. Besucherbefragungen in Museen und deren Auswertung zumeist ausgelagert und von externen Spezialisten ausgeführt. In der Regel wird in den Fachämtern jenseits der in den weit verbreiteten Office-Paketen enthaltenen Tabellenkalkulation wie z. B. MS-Excel oder LibreOffice Calc keine Software für statistische Analysen vorgehalten. Es gibt auch heute – gut 30 Jahre nach dem geschilderten grundsätzlichen Wandel der Büroarbeitsplätze von Archäologen – keine Anzeichen dafür, dass sich dieser Zustand ändert. Vielmehr werden vor allem aus Sicherheitsgründen die Computerarbeitsplätze in vielen archäologischen Dienststellen abgeschottet, d. h. das denkbare Mitbringen und Andocken von USB-Sticks und ggf. eigener Software ist zumindest unerwünscht und vielfach sogar technisch verunmöglicht. In der privatwirtschaftlichen Archäologie, die freier agieren kann, dürfte die Frage der Kosten entscheidend sein, und gerade kleine Firmen oder Dienstleister werden vor

---

<sup>1</sup> Eine aktuelle Übersicht über die Gepflogenheiten, Nutzungsintensitäten und auch den Job-Markt zu Programmiersprachen, Datenbankprogrammen und Statistik-Software bietet Bob Muenchen in seinem Blog: Muenchen, Robert A. (2017, 2019). The Popularity of Data Science Software. *Blog R4stats*, <http://r4stats.com/articles/popularity/> [24.1.2020].

allem die ihres Erachtens dringlichsten Lizenzen erwerben wollen, d. h. sich eher auf unmittelbar Dokumentationsrelevantes fokussieren wollen und auf das, was sich aus den Übergabepflichten an Auftraggeber und Denkmalbehörden ergibt. Dies insgesamt führt aus der Sicht von Forschung und universitärer Lehre zu der unbefriedigenden Situation, dass die Ausbildung und Anfangserfahrung von Studierenden an hochwertigen proprietären Statistik-Programmen, die im Universitätskontext kostenlos oder sehr preiswert zur Verfügung stehen, die darin nunmehr Erfahrenen in eine Sackgasse führt: Kaum beherrscht man die Instrumente, kann oder darf man sie im Berufsleben nicht mehr anwenden. Es führt auch dazu, dass in der deutschen Archäologie weitaus weniger statistische Analysen durchgeführt werden, als es sachlich geboten wäre.

Gewiss: Man darf hoffen, dass sich dies langfristig ändert. Aber längeres Warten oder die Hoffnung auf einen erfolgreichen Marsch durch die Institutionen sind keine Lösung für das Hier & Heute. Daher setzt dieses Buch auf freie Software, und zwar vollständig. Hier stehen inzwischen sehr leistungsfähige Alternativen zu den Bezahlprogrammen zur Verfügung und das Lernen und Einarbeiten darin lohnt sich, weil es jedem Willigen auch jenseits des universitären Kontextes die Option auf eine nachhaltig verfügbare gute Arbeitsumgebung bietet. Arbeitgeber und Dienstherren profitieren davon, denn sie gewinnen ohne Umlern-Aufwand und zusätzliche Softwarekosten unmittelbar einsatzfähiges Personal. Für die auch im Beruf forschen wollenden Kollegen gilt: Selbst wenn's im Dienst nicht möglich ist oder nicht gewollt wird, kann man zumindest in der sog. Freizeit weiterhin mit guten und nunmehr gewohnten Werkzeugen weiterarbeiten. Daher: freie Software.

Dass diese noch andere Vorzüge hat, sei nur angedeutet: Bei freier Software liegen die verwendeten Algorithmen und der Programmcode (meist) offen, d. h. sie sind wissenschaftlich nachprüfbar, und die Produkte, die man mit ihnen erarbeitet hat, können weitergegeben werden (Open Science). Später in diesem Buch werden wir dies unter dem Thema „Skripte“ wieder aufgreifen, worunter man Folgen von statistischen Prozeduren versteht, die ggf. auch Teil einer wissenschaftlichen Publikation sein können. Mit Bezahlprogrammen ist dies nicht resp. nur mit erheblichen Einschränkungen möglich bzw. sinnvoll.

#### **4.4 BESCHREIBENDE UND SCHLIEßENDE STATISTIK**

Der Begriff „Statistik“ meinte ursprünglich das Sammeln und übersichtliche Zusammenstellen von Beobachtungen (Daten, Fakten), und wird in diesem Wortsinn auch weiterhin verwendet. Diese „beschreibende Statistik“ unterscheidet sich in mancherlei Hinsicht von der sog. „schließenden Statistik“. Am leichtesten lässt sich dies anhand

von Wahlen und Wahlergebnissen erläutern. Nach einer Parlamentswahl werden die Stimmen gezählt, die Zählungen geprüft und die Ergebnisse dokumentiert und öffentlich bekannt gegeben. In der Regel werden Prozentzahlen kommuniziert und als Hintergrundinformationen die absoluten Stimmzahlen genannt. Das ist „beschreibende Statistik“, und die Prozentzahl ist eindeutig, exakt und zuverlässig, sie ist als Beschreibung des Wahlergebnisses so zu nehmen, wie sie ist.

Bei der Entwicklung und Einführung neuer Produkte oder – öffentlich bekannter – im Vorfeld wichtiger Wahlen finden regelmäßig Umfragen statt, teils werden die Ergebnisse öffentlich genannt, teils verbleiben sie beim Auftraggeber. Da hier aus praktischen wie finanziellen Gründen nie alle Mitglieder der Zielgruppe wie z.B. alle Wähler befragt werden, werden wohlgeplant Stichproben befragt, meines Wissens oft circa 1.000 oder 2.000 Menschen, bisweilen bis 4.000 Wähler. Die hieraus resultierenden Prozentzahlen kann man im Sinne der beschreibenden Statistik als zuverlässige, sichere Aussagen über diese Befragten nehmen. Aber das ist nicht das Ziel. Vielmehr streben Auftraggeber wie Statistik-Dienstleister an, aufgrund der (kleinen) Stichprobe eine Aussage über die (große) Gesamtkohorte aller Kunden oder Wähler zu machen, eine Wahlprognose eben. Wenn dies angestrebt ist – auf Basis einer Stichprobe eine Aussage über die Allgemeinheit resp. eine weitaus größere Grundgesamtheit zu machen – spricht man von „schließender Statistik“. Deren Zahlen sind, auch das wissen wir z. B. von Wahlprognosen, mit (kalkulierbaren) Unsicherheiten behaftet. Eine Prozentzahl aus dem Bereich schließender Statistik hat demnach eine andere Bedeutung als eine Prozentzahl aus dem Bereich beschreibender Statistik – was selbstverständlich für alle anderen statistischen Kennzahlen wie z. B. Mittelwerte oder Streuungsmaße ebenfalls gilt. Insofern unterscheiden sich die Aussagen der beschreibenden und der schließenden Statistik deutlich.

## 5 ÜBER DEN AUFBAU DES BUCHES

---

Das Buch möchte den in der Praxis üblichen Arbeitsprozess und die übliche Reihenfolge der Entscheidungen und Analyseschritte abbilden. Idealerweise haben Sie ein ganz konkretes eigenes Projekt im Sinn und können nun parallel zum Durcharbeiten des Buchs Ihr Projekt Schritt für Schritt bearbeiten, weiterverfolgen und auswerten. Der Aufbau des Werks soll ihnen dabei helfen und entgegenkommen. Dazu wird im nächsten Kapitel zunächst dazu angeregt, die eigenen Bedürfnisse und die Möglichkeiten aus der Benutzer- und Anwendersicht zu klären. Denn danach richtet sich die

Frage, wie tief man in die Materie einsteigt, und die Wahl der Instrumente und statistischen Verfahren. Anschließend wird kurz die bei Geisteswissenschaftlern weit verbreitete Skepsis gegen die Anwendung von statistischen Methoden und Argumentationen thematisiert, denn ohne eine gewisse Motivation und Überzeugtheit macht das weitere Arbeiten entlang dieses Buchs wenig Sinn.

Im anschließenden Kapitel stelle ich exemplarisch einige m. E. interessante freie, kostenlose Werkzeuge (Software) vor, die im Folgenden eingesetzt werden, wobei sich die persönliche Auswahl des Lesers an den zuvor getroffenen Entscheidungen hinsichtlich seiner persönlichen Bedürfnisse und Möglichkeiten orientiert. Es werden jeweils die Bezugsquelle, die Installation sowie erste Schritte im Umgang damit eingeführt. Danach wird der Weg einer parallelen Beschreibung mehrerer Softwareoptionen jedoch verlassen, um das Buch lesbar zu halten: Diejenigen, die sich tiefer für das Arbeiten z. B. mit PAST, JAMOVI, BlueSky Statistics oder dem Bezahlprogramm SPSS entscheiden, werden dann selbstständig weiterfinden müssen, denn der anschließende Teil dieses Buchs beruht softwareseitig auf LO-Calc und **R** mit R-Commander. Keine Sorge: Diejenigen, die sich aufgrund ihrer anders gelagerten Bedürfnisse nicht auf **R** einlassen können/wollen/müssen, finden insbesondere bei Lösungen wie z. B. PAST oder JAMOVI selbstständig gute Hilfen und Dokumentationen; zudem stößt man nach kurzer Recherche z. B. bei YouTube auf zahlreiche Anleitungen, die beim praktischen Weiterarbeiten helfen. Das weitere Durcharbeiten dieses Buchs bleibt aber auch für sie sinnvoll, denn die statistische Theorie und die Verknüpfung von archäologischen Problemstellungen mit den dafür geeigneten statistischen Verfahren, um die es hier im Kern geht, kann anhand dieses Buches auch ohne das Erlernen von **R** erfahren werden. Der Unterschied ist nur, dass **R** weiter führt, ohne dass Sie je an ein „gläsernes Dach“ eingeschränkter Möglichkeiten durch die Wahl ihres Statistikprogramms geraten.

Nach dem Aufbau einer Arbeitsumgebung mit **R** beschäftigen wir uns dann als ersten weiterführenden Schritt mit dem Thema Datenerfassung (und -verwaltung) und dem wichtigen Thema der unterschiedlichen Art von Informationen (Stichwort Skalenniveau), weil dessen Verständnis bereits bei der Datenerfassung wichtig ist. Hat man die für ein Projekt benötigten Informationen erfasst, liegt es nahe, sich zuerst ohne aufwendige Statistiken anhand von grafischen Darstellungen einen ersten Einblick in die Daten und deren Zusammenhänge zu verschaffen. Daher werden zunächst die grundlegenden statistischen Grafiken besprochen. Schon hier kann man viel falsch oder eben auch richtig und professionell machen. Denn schon bei Befolgen relativ weniger Grundregeln gewinnt man gute und publikationsreife Grafiken.

Einerseits hoffe ich, dass Sie das Buch bereits früh anhand eigener Daten und Problemstellungen durcharbeiten, andererseits ist zu vermuten, dass ihnen nicht zu jedem Thema schon passende Daten zur Verfügung stehen. Daher werden hier einige Beispieldatensätze eingeführt, die wir im Folgenden exemplarisch immer wieder für die verschiedenen Fragestellungen und Analysen benutzen. Wiewohl der Autor über zahlreiche Datensätze verfügt, wird versucht, die Anzahl dieser Beispieldatensätze klein zu halten, damit die Leser fokussiert an wenigen Modellen intensiver üben können.

Danach geht es darum, von einem fundierten Grundwissen ausgehend möglichst schnell arbeitsfähig zu werden. Ein umfangreiches Kapitel beschäftigt sich mit dem Themenkomplex Häufigkeiten und Prozentzahlen, weil deren Analyse und Vergleich in der Archäologie eine große Rolle spielen. An praktischen Beispielen werden alle grundlegenden Themen behandelt. Teil der Beschäftigung mit Häufigkeiten ist die – hier recht ausführliche – Einführung des Chi-Quadrat-Tests. Die Ausführlichkeit begründet sich darin, dass dieses einfache Verfahren in der Archäologie tatsächlich relativ wichtig ist, vor allem aber auch darin, dass man an diesem einfachen Modell lernen kann, wie statistische Tests überhaupt funktionieren. Für später einzuführende andere Testverfahren kann dann auf diese Ausführlichkeit verzichtet werden.

Ein weiteres umfangreiches Kapitel führt dann in die Analyse von Messwerten ein, wie sie uns beispielsweise bei den Maßen von Waffen, Keramikgefäßen oder Langknochen begegnen. Seitens der Statistik werden hier Begriffe wie Mittelwert und Standardabweichung, Mittelwertvergleiche, Korrelation und Regression etc. behandelt werden. Leser, die sich bis hierhin vorgearbeitet haben, sollten für geschätzt 90 % ihrer künftigen Problemstellungen und deren angemessene Lösung gerüstet sein.

Nach dieser grundlegenden Einführung, die Sie mit der nötigen Theorie und den Werkzeugen vertraut macht, werden im letzten Teil des Buchs einige m. E. in der Archäologie typische und immer wieder vorkommende Anwendungsfälle behandelt. Hier soll das zuvor Erlernete praxisnah angewendet werden. Eine gewisse Künstlichkeit wird sich dabei nicht vermeiden lassen, denn die gewählten Themen sind ja – notwendigerweise – schon beforscht, während ihr eigenes Thema noch unerforscht ist und mit selbstständigen Schritten in das noch Unbekannte einhergeht. Gleichwie sollen Ihnen diese Beispiele Orientierung und Anregung geben, damit Sie selbstständig in ihre eigenen Analysen einsteigen können.

## 5.1 ÜBER DIE SPRACHE: DEUTSCH ODER ENGLISCH

Dieses Buch ist in deutscher Sprache geschrieben, weil nach dem guten Buch von Peter Ihm (et al., 1978) eine zeitgemäße Einführung in deutscher Sprache m. E. ein Desiderat ist. Für den englischen Sprachraum stehen archäologie-orientierte Einführungen in die Statistik zur Verfügung, als „Klassiker“ z. B. Clive Orton (1980), Stephen Shennan (1988) oder Robert D. Drennan (2010), und zuletzt das empfehlenswerte Buch von Grant S. McCall (2018). Einführungs- oder Nachschlagewerke aus Nachbarfeldern der Archäologie wie z. B. Sozialwissenschaften und Biologie/Ökologie sind z. T. weiterführender, z. B. Sokal & Rohlf (1995), Hair, Black, Babin & Anderson (2010) oder Legendre & Legendre (2012). Die Erfahrung zeigt, dass der Dreikampf Archäologie – Statistik – Software hinreichend komplex ist und daher Einführungs- werke in der Muttersprache hilfreich sind, um wenigstens eine weitere (wenn auch geringe) Hürde zu vermeiden. Andererseits: Sehr viel Literatur muss in Englisch gelesen werden und die hier benutzten freien Programme PAST und **R** „sprechen Englisch“. Ich werde daher bei der Einführung wichtiger Fachbegriffe stets auch das englische Wort in Klammern (*Kursivschrift*) nachstellen und die englischen Begriffe mit ins Register aufnehmen, weil dies für ein selbstständiges Arbeiten hilfreich ist.

**Praxistipp.** Was tun, wenn man in englischen Texten auf ein Fachwort stößt, an dem auch das persönlich bevorzugte Online-Wörterbuch scheitert? Mir hilft dabei zumeist die Wikipedia: Auf die entsprechende Wikipedia gehen, also z. B. die englische Fassung, das betreffende Stichwort suchen (und hoffen, dass es zu einem geeigneten Artikel führt). Findet man etwas, kann man dann in der Navigationsleiste links die passende anderssprachige Wikipedia anfahren, z. B. „Deutsch“, und landet meist beim treffenden deutschen Fachwort. Funktioniert bei Bedarf auch andersrum.

## 5.2 WAS KAUM ODER NICHT BEHANDELT WIRD

Einige Themen, die Leser eventuell erwarten, werden in diesem Buch nicht behandelt werden. An erster Stelle nenne ich das für die Archäologie sehr relevante Thema Seriation und Korrespondenzanalyse. Ich klammere das Thema hier aus dem einfachen Grund aus, dass ich dazu bereits ein kleines Lehrbuch veröffentlicht habe (Siegmund 2015) und mit Müller & Zimmermann (1997) ein gutes weiterführendes Werk vorliegt.

Nur sehr wenig werde ich auf das seit einigen Jahren viel erwähnte Thema Bayes-Statistik eingehen. Aus meiner Sicht ist dieser Bereich seitens der Statistik unverhältnismäßig kompliziert und bis auf wenige Ausnahmen sind die daraus abgeleiteten zusätzlichen Erkenntnisgewinne in der Archäologie gegenüber dem, was mit traditioneller Statistik erreicht werden kann, bisher recht übersichtlich. Das Sprichwort

von Kanonen und Spatzen kommt mir hierbei in den Sinn. Im Grunde würde eine breite Integration der Bayes-Statistik letztlich auf ein Outsourcing der statistischen Analysen an Spezialisten hinauslaufen (so, wie es derzeit in der aDNA-Forschung üblich ist), und gerade das halte ich für einen falschen Weg. Das archäologie-adäquate Formalisieren der Fragestellungen, das Entwerfen und Planen einer angemessenen Datenerhebung und das Entwickeln und Verfolgen eines „roten Fadens“ der Argumentation und Interpretation muss aus der Archäologie selbst heraus geleistet werden, und folglich müssen quantitativ argumentierende Archäologen ein hinreichendes Verständnis der Statistik haben, die sie dafür einsetzen.

## 6 SKEPSIS GEGENÜBER DER ANWENDUNG STATISTISCHER METHODEN

---

Gerade bei Archäologen – wie Geisteswissenschaftlern allgemein – stößt man oft auf eine große Skepsis Zahlen und Statistiken gegenüber. Der Winston Churchill zugeschriebene Spruch „Traue keiner Statistik, die du nicht selbst gefälscht hast“<sup>2</sup> gehört quasi zur Allgemeinbildung. Das ist seltsam. Denn wer im Alltag politische Diskussionen verfolgt, etwa um die Rentenversicherung, die Kosten im Gesundheitswesen, die Strompreise oder die nächste Lohnrunde, beobachtet Konflikte, die sich sehr oft in Zahlen ausdrücken, wobei es vielfach „nur“ um Zehntelprozente geht. Und dennoch können diese Auseinandersetzungen heftig sein, und fast jeder Betroffene weiß zu bewerten, was so abstrakte Zahlen wie „+1,9 %“ oder „+1,2 %“ für ihn persönlich bedeuten, wenn es um eine Lohnerhöhung geht. Welcher Autofahrer in drängendster Eile hat nicht schon einmal überschlägig kalkuliert, wie schnell man hier fahren darf und wie die Strafe aussähe, wenn man sich ein wenig jenseits einer Grenze bewegt? Zahlen sind ein selbstverständlicher Teil unseres Alltags. Sind sie indes etwas spezifisch Technisches oder Naturwissenschaftliches und den Geisteswissenschaften und historischen Prozessen Fremdes? Nein. In allen Wissenschaften sind Begriffe wichtig, d. h. gedankliche Abstraktionen, die eine bunte Vielfalt von Einzelerscheinungen unter bestimmten Aspekten subsumieren und damit als Gedanken handhabbar machen. Es gibt viele gotische Kirchen, jede ist anders, und doch scheint es nützlich, einen

---

<sup>2</sup> Zur Authentizität siehe z. B. Wikiquote: [https://de.wikipedia.org/wiki/Liste\\_gefl%C3%BCgelter\\_Worte/T#Traue\\_keiner\\_Statistik,\\_die\\_du\\_nicht\\_selbst\\_gef%C3%A4lscht\\_hast](https://de.wikipedia.org/wiki/Liste_gefl%C3%BCgelter_Worte/T#Traue_keiner_Statistik,_die_du_nicht_selbst_gef%C3%A4lscht_hast) [24.1.2020].

Epochenbegriff „die Gotik“ zu prägen, um eine Epoche und einen Zeitstil als Ganzes zu fassen. Gerade dies, die Fähigkeit zur Abstraktion, gilt sogar als wesentliche Leistung des menschlichen Geistes. Man kann auch die Bibel so verstehen: „Am Anfang war das Wort“ (Joh. 1,1), also ein Begriff und die begriffliche Ordnung der Welt. Wenn wir Beobachtungen in Zahlen fassen, ist dies ein erster Schritt zur Abstraktion, auch ein Schritt zur gegenseitigen Verständigung. Verfolgen wir den Dialog über ein Neugeborenes: „Wie groß ist das Kind denn?“ „Ziemlich, na so ... [es folgt eine Handbewegung]“ – Gesten und Zeichen sind sicherlich eine gängige Art, Information zwischen Personen auszutauschen. Indes, eine klare Zahl wie „5.230 g, 51 cm“ kann hilfreich sein; sehr viele Geburtsanzeigen, aufgegeben von stolzen Eltern, nennen solche Daten. „Oh, ein recht schweres Kind!“ – könnte dem Leser dieser Nachricht durch den Kopf gehen. Offenbar steht im Hintergrund eine Erfahrung, wie groß und schwer Neugeborene üblicherweise sind. Auch diese Erfahrung ließe sich objektivierend und somit besser vermittelbar in abstrakte Zahlen fassen – und schon geraten wir in das Feld der beschreibenden Statistik. Wir sollten Zahlen und Statistik als eine selbstverständliche Form der Abstraktion verstehen, die der Verständigung dient und sich nicht grundsätzlich von sprachlichen Begriffen unterscheidet.

Man könnte sogar Statistik als etwas der Archäologie besonders Naheliegender betrachten: Archäologische Quellen sind als einzelnes Objekt oder Befund oft nur wenig informativ, aber häufig; Beispiele hierfür sind unscheinbare und als Einzelstücke auch kaum besonders spektakuläre Steinartefakte, Gefäßscherben oder Hauspfosten. Vielfach stehen nicht Individuen und Einzelercheinungen im Fokus archäologischer Forschung, sondern Gruppen und Prozesse. In diesem Kontext ist das Erkennen von Häufigkeiten, Regelmäßigkeiten und eben Ausnahmen ein geläufiges Argument. Archäologie fußt heute auf einer etwa 150-jährigen Tradition des wissenschaftlichen Sammelns, Ausgrabens und Bewahrens von Funden bei wachsender Professionalisierung der Archäologie. Heutige Archäologen haben dadurch anders als frühere Forschergenerationen tatsächlich viele Funde.<sup>3</sup> Das Zusammenfassen von Tatsachen in Zahlen und Tabellen schafft Übersicht; Statistik hilft, Klarheit und Eindeutigkeit zu gewinnen und die Sicherheit oder auch Unsicherheit von Aussagen erfassen und beschreiben zu können.

---

<sup>3</sup> „Um eine sichere Chronologie für die vorgeschichtlichen Zeiten zu erhalten, muss man ein grosses Material und eine gute Methode haben.“ (Montelius, Oscar [1903]. *Die typologische Methode*. [Separatdruck aus „Die älteren Kulturperioden im Orient und in Europa“]. [S. 2]. Stockholm: Selbstverlag des Verfassers). Hervorhebungen von Montelius.

### 6.1.1 Einwand: zu wenig Zahlen

Gelegentlich hört oder liest man als Argument gegen die Anwendung statistischer Methoden die Aussage: „Ich habe gar nicht so viele Funde, da kann man statistisch nichts machen“. Diese These ist, zurückhaltend formuliert, nicht ganz richtig. Denn eine geläufige Aufgabe für die schließende Statistik beruht gerade auf geringen Quantitäten. Ein Beispiel: Mediziner testen ein neues Medikament erstmals an Menschen. Der Versuch ist vernünftigerweise auf wenige Patienten begrenzt. Nun sind wider Erwarten in Folge bei den ersten drei Probanden gravierende Verschlechterungen ihres Befindens aufgetreten – was soll man tun? Die Testreihe abbrechen, um weiteren Patienten Schaden zu ersparen? Oder liegen die drei Fehlversuche noch im Rahmen der zu erwartenden Zufallsstreuung, kann man es also verantworten, die Versuchsreihe fortzusetzen? Dies sind konkrete und wichtige Fragen, zu denen man Rat in der Statistik suchen und finden kann. Peter Ihm etwa, der das Lehrbuch „Statistik in der Archäologie“ (1978) verfasst hat, hat genau dies im Hauptberuf getan: an der Universität Marburg in der medizinischen Statistik gearbeitet und beispielsweise zu derartigen Problemen der kleinen Zahl mithilfe der Statistik beraten. Kurz: Gerade weil die Datengrundlage in der Archäologie oft nicht, wiewohl wünschenswert, gezielt verbessert werden kann, und weil die Ressourcen endlich sind, bedarf es der Statistik: Sie hilft, objektivierend aufzuzeigen, ob aufgrund des vorliegenden Datenmaterials eine bestimmte These verifiziert oder falsifiziert werden kann.

### 6.1.2 Einwand: Statistik ist (zu) kompliziert

Manchem erscheint der Weg in die Statistik kompliziert, d. h. mit allzu viel Mathematik und Computerfertigkeiten behaftet. Im Verlauf dieses Buchs wird sich hoffentlich herausstellen, dass auch dieses Vorurteil der Korrektur bedarf. In der Tat benötigt man gelegentlich komplexere Berechnungen, doch die Masse dessen, was in der Archäologie wichtig und weiterführend ist, kann mit Hilfe einfacher Grafiken dargestellt werden und sogar rechnerisch mit Hilfe der vier Grundrechenarten und des Wurzelziehens selbst erarbeitet werden. Die meisten Dinge lassen sich auch mit Papier und Stift oder einem Taschenrechner als Hilfsmittel gut nachvollziehen. Allerdings sind viele der nötigen Rechnereien oft und wiederholt durchzuführen, weshalb auf Dauer und bei größeren Serien das Arbeiten mit dem Taschenrechner etwas langweilig und in der Praxis auch etwas fehleranfällig wäre. An dieser Stelle tritt der „Computer“ im Wortsinn hilfreich zur Seite, der die immer gleichen, langweiligen Rechnereien zuverlässig übernimmt. Man lasse sich nicht von einzelnen komplizierten Verfahren abschrecken: circa 90 % der Leistungen der Statistik für die Archäologie können mit sehr wenigen und zugleich sehr einfachen Verfahren erbracht werden. Oft benötigt das Verwalten der

Informationen, d. h. das Erfassen, Sortieren, Filtern und Anpassen der Daten, weitaus mehr Zeit und Aufmerksamkeit als die reine statistische Analyse, wobei diese Datenverwaltung meist auch dann nötig ist, wenn man sein Material ohne Statistik auswerten möchte.

### **6.1.3 Einwand: mir fehlen die Programme**

Stimmt nicht, und genau daher gibt es dieses Buch, das auf den Einsatz freier Software abzielt: Sofern Sie überhaupt mit einem Computer arbeiten, z. B. um Texte zu schreiben, im Internet zu surfen oder Emails abzurufen, können Sie sich die nötigen Programme ohne große Mühen kostenlos besorgen. Alles Weitere liegt nur an Ihnen, nicht am Geld. Mit etwas Offenheit, Lernbereitschaft und Geduld können Sie sich einarbeiten – dieses Buch wird ihnen dabei helfen.

## **7 EIN KURZER RÜCKBLICK IN DIE FORSCHUNGSGESCHICHTE: DIE KO-EVOLUTION VON ANGEWANDTER STATISTIK UND ARCHÄOLOGIE**

---

Nicht nur die Archäologie ist eine Wissenschaft, die sich weiterentwickelt und die auch Moden und unterschiedliche Schulen wie z. B. Prozessuale und Post-Prozessuale Archäologie kennt. Die Idee, Statistik sei demgegenüber ein starres Feld mit ausgereiften, erratisch dastehenden Regeln wäre irrig. Richtig ist, dass auch hier neben etabliertem Allgemeinwissen und allgemein als gültig anerkannten Regeln Dinge in Bewegung sind, ähnliche Ziele auf verschiedenen Wegen erreicht werden können und es unterschiedliche Stile und auch Moden gibt. Viele Lehrbücher neigen dazu, die Wissensfortschritte nach einer gewissen Reifezeit akkumulierend zusammenzustellen, was für uns nützlich ist, aber manchmal den Blick auf die durchaus unterschiedlichen Ansätze und Grundhaltungen innerhalb der Statistik verstellt. Zudem sind für unser Thema die Technikgeschichte sowie die Forschungsgeschichten von Statistik und Archäologie miteinander verknüpft; dafür soll dieses Kapitel die Augen schärfen.

Ursprünglich verstand man unter „Statistik“ die beschreibende Darstellung von Staat und Bevölkerung. In diesem Sinne wird der Begriff im 17. Jahrhundert üblich und dann lange Zeit benutzt, d. h. im Sinne von zielgerichtetem Sammeln und Zusammenstellen von Informationen, vor allem zum Zwecke der staatlichen Verwaltung und Wirtschaftsplanung. Noch Meyers Konversationslexikon von kurz vor 1900

versteht „Statistik“ vorwiegend in diesem Sinne, und auch heute ist dieser Gebrauch als eines der Bedeutungsfelder des Begriffs „Statistik“ noch üblich. Eine andere wichtige Nährquelle unseres heutigen Verständnisses von Statistik ist das Glückspiel. Schon früh haben sich Menschen für das „Glück im Spiel“ interessiert und wissen wollen, wie es um die Chancen etwa im Würfelspiel oder beim Roulette steht. Der Basler Professor Jacob Bernoulli etwa machte Ende des 17. Jahrhunderts grundlegende Überlegungen zur Wahrscheinlichkeits-Theorie, angewandt auf das Würfelspiel als „ars conjectandi“ 1713 in Basel publiziert. Weitere wichtige Entdeckungen fallen in das 19. Jahrhundert – ich nenne Namen wie Gauss, Cantor oder Laplace – und bald geht es dabei auch um Fragen des Versicherungswesens etc., d. h. auch diese Facette der Statistik erhält konkrete ökonomische Bedeutung. Indes: Letztlich ist die Statistik in unserem heutigen Sinne recht jung. Viele der heute üblichen Begriffe und Standard-Verfahren sind erst im Laufe des 20. Jahrhunderts entwickelt worden – was sicherlich auch mit der technischen Entwicklung zusammenhängt. Denn die mühsamen Berechnungen, vielfach mit Hilfe von Logarithmentafeln und Tabellenwerken, waren lange Zeit Sache weniger Spezialisten. Eine weite Verbreitung – d. h. eine Verwendung über die Staatsverwaltung, die Ökonomie und das Versicherungswesen hinaus – und dann auch einen inhaltlichen Aufschwung nimmt die Statistik mit dem Aufkommen der Computer in den 1950er und -60er Jahren.

### **7.1.1 Die Beginne in der Archäologie der 1970er und -80er Jahre**

Während des 2. Weltkriegs entstehen als neue technische Geräte „Computer“ und als neuer Gedanke „Programmiersprachen“, welche die einfache und flexible Steuerung solcher Anlagen erlauben. Zu einem auch in der Archäologie spürbaren Schub führt indes erst das Aufkommen und die weite Verbreitung von sogenannten Großrechnern (*mainframe*) in der 2. Hälfte der 1960er und der 1. Hälfte der 1970er Jahre an den Universitäten (**Abb. 1**). Ergänzend standen ab 1974 – für heutige Verhältnisse sehr teure – programmierbare Taschenrechner wie z.B. der HP-65 als persönliche Arbeitsinstrumente zur Verfügung, die von vielen Statistikern genutzt wurden, um das zuvor übliche Arbeiten mit Papier und Bleistift, Rechenschieber und Logarithmentafeln abzulösen. Parallel zur Entwicklung der Hardware entstehen „Programmpakete“ für Statistik (z. B. BMDP 1965 ff.; SAS 1966 ff.; SPSS 1968 ff.); mit ihnen ersetzen fertige, ausgetestete professionelle Programmsammlungen mit standardisierten Verfahren, die mit geringerem Aufwand an die individuellen Bedürfnisse angepasst werden konnten, die bislang notwendige individuelle Programmierung statistischer Prozeduren. Sie erkennen die Werke dieser Epoche leicht an den typischen, kaum lesbaren Ausdrucken der

Schnelldrucker dieser Großrechner. Das Lehrbuch von Peter Ihm „Statistik in der Archäologie“ (1978) liegt stilistisch noch ganz in dieser Zeit und ihren technischen Möglichkeiten.



**Abb. 1** Blick auf eine CDC Cyber 76. An einem solchen „Großrechner“ und in Räumen ähnlicher Atmosphäre machte der Autor seine ersten IT- und Statistik-Erfahrungen. (Quelle: Prospekt der Firma CDC, Nov. 1975; vgl. <https://www.youtube.com/watch?v=HvLlFXZMBXg> [24.1.2020]).

Die seinerzeit schnell wachsenden Rechenkapazitäten machten es einfacher als bisher, nicht nur einen einzelnen Aspekt im jeweiligen Datensatz zu verfolgen, sondern mehrere Aspekte gleichzeitig zu untersuchen – wir nennen dies „multivariate Verfahren“. Also etwa bei der Frage nach der Form von Töpfen nicht eine Variable allein wie z. B. die Gefäßhöhe zu betrachten und sich dann in einem weiteren Analyseschritt z. B. mit dem Mündungsdurchmesser zu beschäftigen usw., sondern gleichzeitig mehrere Vari-

ablen zu untersuchen, also etwa Gefäßhöhe, Mündungs- und Bauchdurchmesser gemeinsam. Hierzu dienen statistische Verfahren wie Faktorenanalyse, Clusteranalyse etc., die in jenen Jahren außerordentlich beliebt waren. Klassische Lehrbücher dazu stammen aus den frühen 1970er Jahren (z. B. Überla 1968, 1977; Bock 1974). Vergleichbares ohne Computer zu rechnen ist zwar möglich, aber schlicht quälend. Letztlich gehört auch das Thema „Seriation“ in diesen Kontext, dessen moderne Grundlagen aus archäologischer Sicht Ende der 1960er Jahre gelegt wurden (Goldmann 1972; Ihm 1983).

In der Archäologie steht diese Zeit für das Aufkommen der *New Archaeology* (heute: Prozessuale Archäologie) – also eines Archäologie-Verständnisses, das weniger nach den individuellen historischen Ereignissen, sondern nach allgemeingültigen Regeln menschlichen Verhaltens fragt, und folglich naheliegenderweise die neuen Möglichkeiten der Statistik begierig aufgriff. Einflussreiche Werke in diesem Sinne waren insbesondere die beiden Bücher von David Leonard Clarke „Analytical archaeology“ (London 1968) und „Beaker pottery of Great Britain and Ireland“ (Cambridge 1970). Rasch erschienen erste Lehrbücher zur Statistik in der Archäologie: Im englischsprachigen Raum das grundlegende Werk von James E. Doran und Frank Roy Hodson „Mathematics and computers in archaeology“ (Edinburgh 1975), in Deutschland Peter Ihm „Statistik in der Archäologie“ (Bonn 1978).

Dabei spielt die quantitative Entwicklung der Archäologie sicherlich ebenfalls eine wichtige Rolle: Die neuen Denkmalschutzgesetze und der relative Wohlstand in den 1970er Jahren führten in Deutschland einerseits zu einer festeren Etablierung der staatlichen Ämter für Bodendenkmalpflege und damit auch zu einer gesteigerten archäologischen Tätigkeit der Amtsarchäologie,<sup>4</sup> andererseits entstanden parallel dazu an einigen Universitäten ambitionierte Forschungsprojekte von bislang kaum gekannten Größenordnungen – mit entsprechender Fund- und Informationsmenge. Für die deutsche Archäologie spielt das 1971 in Köln gestartete und bis 1983 laufende Projekt zur Siedlungsarchäologie des Neolithikums im rheinischen Braunkohlerevier (SAP) eine wichtige Rolle,<sup>5</sup> da es sich frühzeitig intensiv der Möglichkeiten des Computers und der Statistik bediente. Wenig später entstand auch in der Bodendenkmalpflege das Bedürfnis, die nun dramatisch anwachsende Menge an Quellen besser zu verwalten:

---

<sup>4</sup> Für das Rheinland z.B. Siegmund (1988), 8 Abb. 3.

<sup>5</sup> Rückblickende Zusammenfassung: Kuper R. & Lüning, J. (2018). Mit dem Bagger gegen den Bagger: das Projekt "Siedlungsarchäologie des Neolithikums der Aldenhovener Platte" (SAP). In J. Richter (Hrsg.), *111 Jahre prähistorische Archäologie in Köln*. (Kölner Studien zur Prähistorischen Archäologie, 9). (p. 198-207). Rahden /Westf.: Leidorf.

Seit ca. 1980 wurde für die rheinische Landesarchäologie die sog. „Bonner Bodendenkmälerdatenbank“ entwickelt, die zwar auf das Rheinland beschränkt blieb, aber im deutschsprachigen Raum für fast alle danach entwickelten Datenbanken der staatlichen Ämter ein wichtiges Modell abgab.



**Abb. 2** Eine DEC PDP-11. An einem solchen sog. "Minicomputer" lief seit Mitte der 1970er Jahre unter dem unix-ähnlichen Betriebssystem RSX 11 das von Irwin Scollar und Irmela Herzog entwickelte Seriationsprogramm BASP, das dann Mitte der 1980er Jahre für PCs unter MS-DOS portiert wurde (vgl. <https://www.youtube.com/watch?v=tWofN9p5E8k> [24.1.2020]).

In der Schweiz scheint mir das große Grabungsprojekt Twann 1974-1976 eine vergleichbare Bedeutung zu haben:<sup>6</sup> Ausgrabungen in einer bis dahin kaum gekannten Dimension mit entsprechenden Fund- und Informationsmengen. Viele Konzepte und Modelle der Schweizer Archäologie, die später auch in den Kantonsarchäologien verfolgt wurden, haben im Projekt Twann ihren Ursprung. Im Hinblick auf die großen

---

<sup>6</sup> Rückblickend: Stöckli, W. E. (2018). *Twann: Ausgrabungen 1974-1976, Auswertungen 1976-1982, Schlussbericht von 1981/82, Kommentar von 2017*. (Die neolithischen Ufersiedlungen von Twann, 21; Hefte zur Archäologie im Kanton Bern, 3). Bern: Archäologischer Dienst des Kantons Bern.

Grabungsprojekte am Zürichsee publizierte Ulrich Ruoff 1984 ein in BASIC geschriebenes „Microcomputerprogramm für den Archäologen“ (Zürcher Studien zur Archäologie 2 [Zürich 1984]), das ebenfalls der besseren Verwaltung großer Grabungen dienen sollte.

Dass die Beschäftigung mit Computern und Statistik meist von jungen Archäologen getragen wurde, die ihre neuen Ansätze in die traditionelle Archäologie der Etablierten einbringen wollten, liegt nahe. Ebenso naheliegend ist das Bedürfnis zum „Verein“, d. h. zum Zusammenfinden und regelmäßigen Treffen Gleichgesinnter. Anfang der 1970er Jahre, wohl 1973, formierte sich in England eine Gruppe, die seitdem jährlich eine Tagung „Computer applications in archaeology“ (kurz CAA) organisiert, eine Tagung, die seit dem Jahr 2000 auch häufiger außerhalb Englands stattfindet. Zu den Treffen gehören auch entsprechende Tagungsberichte. Eine wichtige Bedeutung für den deutschen Sprachraum gewann zeitweise der Workshop „Archäologie und Computer“, der seit 1996 Jahren regelmäßig von der Stadtarchäologie in Wien ausgerichtet wurde und über viele Jahre hinweg auch international Teilnehmer anzog. In Deutschland ist es – mit vergleichbaren Interessen – die ebenfalls Anfang der 1970er Jahre gegründete „Arbeitsgemeinschaft Quantitative Methoden in der Archäologie“, die sich jährlich trifft, und zwar über lange Jahre hinweg in festem Turnus einmal bei den Tagungen der Altertumsverbände und einmal bei der „Gesellschaft für Klassifikation“ (GfKl), einem wichtigen Statistiker-Verband. Diese Gruppe publizierte ihre Tagungsberichte 1983 bis 1999 in der Zeitschrift „Archäologische Informationen“, eine Liste findet sich im entsprechenden Wikipedia-Artikel.<sup>7</sup> Nach einem Zwischenschritt im Jahr 2005 konstituierte sich diese AG im Jahr 2011 förmlich zum Verein und führt seitdem den Namen „AG CAA“, weil er sich zugleich als nationaler Teilverband der internationalen CAA versteht. In der Schweiz formierte sich 1992 die AGAI, die „Arbeitsgruppe Archäologie und Informatik“ ([www.agai.ch](http://www.agai.ch)), die jedoch in den 2000er Jahren kaum mehr aktiv war.

In dieser Pionierzeit formieren sich auch unterschiedliche „Schulen“, deren Konturen zwar in den 2010er Jahre verwischen, deren Kenntnis aber hilft, die Literatur der 1970er bis 2000er Jahre besser zu verstehen. Die „Marburger/Kölner Schule“ um das Forschungsprojekt SAP in Köln wurde bereits genannt; eine zentrale Figur darin war damals Andreas Zimmermann, der mit der Bearbeitung der bandkeramischen

---

<sup>7</sup> Wikipedia „Arbeitsgemeinschaft Computer-Anwendungen und Quantitative Methoden in der Archäologie“ (28.6.2017); ggf. auch ältere Artikelversionen konsultieren.

Mahlsteine und Feuersteinartefakte betraut war. Die rasche Entwicklung der Methoden wird deutlich, wenn man die aufeinander folgenden Publikationen der Siedlungen Langweiler 2 (1973), Langweiler 9 (1977) und Langweiler 8 (1988) nachliest. Ein steter Berater dieser Arbeitsgruppe war der Marburger Statistiker Peter Ihm, Autor des genannten Lehrbuches. Mit dem um 1980 erfolgten Ruf von Jens Lüning, dem Leiter dieser Kölner Forschungsgruppe, an die Universität Frankfurt wurde dieses Know-how auch nach Frankfurt getragen, sodass man seitdem von einer Marburg-Köln-Frankfurter Schule sprechen könnte.

Die „Schleswiger Schule“: ihr Kern ist Michael Gebühr, dessen Interesse vor allem der Römischen Kaiserzeit im Barbaricum galt, und der lange am Archäologischen Landesmuseum in Schleswig tätig war.<sup>8</sup> Typisch für ihn und die von ihm Beflügelten sind eine Freude am eigenen Programmieren und an der Analyse vor allem von Grabfunden. Seinerzeit ausnehmend innovativ waren seine Simulationen, d. h. die Erarbeitung sehr konkreter Modelle und daraus abgeleiteter Modellrechnungen, die als Simulationen den Vergleich mit archäologischen Realdaten ermöglichen, um im Rückschluss wiederum die Parameter der Simulationsmodelle schärfen zu können.<sup>9</sup>

In der Schweiz ist ein früher Pionier Ulrich Ruoff, auf den hier wesentliche Anstöße zurückgehen, und der nach meiner Erinnerung um 1984 in der „AG Quantitative Methoden“ in Deutschland als Erster mit einem tragbaren Computer erschien, einem ca. 18 kg leichten „Osborne“.<sup>10</sup> Wer einmal in Ruhe in den Band 7 der „Basler Beiträge zur Ur- und Frühgeschichte“ blättert, wird sehen, dass Ende der 1970er Jahre unter den Schülern von Ludwig Berger in Basel ein Kern der Anwendung quantitativer Methoden zu entstehen begann, was jedoch mit dem Fortgang von Andres Furger an das Schweizerische Landesmuseum Zürich und der Wiederaufnahme großer Grabungen im Bereich der Siedlung Basel-Gasfabrik seitens der kantonalen Denkmalpflege keine Nachfolge fand.

In Österreich bildet(e) Peter Stadler in Wien einen Kern; seine Interessen galten vor allem der Seriation awarischer Grabfunde, einige Awarenforscher haben sein selbstgeschriebenes und beständig erweitertes Programm „Serion“ – heute WinSerion – intensiv genutzt.<sup>11</sup>

---

<sup>8</sup> Z.B. *Archäologische Informationen* 5, 1983, 11-19; ebd. 13(2), 1990, 106-120.

<sup>9</sup> Gebühr, M. (1997). Überlegungen zum archäologischen Nachweis von Wanderungen am Beispiel der angelsächsischen Landnahme in Britannien. *Archäologische Informationen* 20(1), 11-24.

<sup>10</sup> Empfehlung: Google Bildersuche nach „osborne portable computer“.

<sup>11</sup> Aus den Anfängen: Stadler, P. (1984). Überlegungen zur computerunterstützten Seriation. In F. Daim & A. Lipert, *Das awarische Gräberfeld von Sommerein am Leithagebirge, NÖ*. (Studien zur Archäologie der Awaren, 1).



**Abb. 3** Der im August 1981 eingeführte IBM-PC 5150 in einer typischen Ausstattung mit zwei 5 1/4“ Diskettenlaufwerken. (Quelle: Wikimedia, Ruben de Rijcke, CC BY SA 3.0).

### 7.1.2 Die 1980er und 1990er Jahre

Zurück zur Statistik. Die klassische Statistik macht vielfach Annahmen über die Verteilung der Variablen in den Datensätzen; so ist beispielsweise eine wichtige Voraussetzung vieler multivariater Verfahren, dass die Daten in etwa normalverteilt sind. Die sich daraus ergebenden Verfahren nennt man „parametrische Verfahren“ – Begriffe und Themen, die wir an geeigneter Stelle noch vertiefen werden. Theoretische Überlegungen und statistische Prüfungen, ob die vorliegenden Daten die Voraussetzungen des angestrebten Analyse-Verfahrens erfüllen, spielen für statistische Tests und multivariate Verfahren eine zentrale Rolle. In einem späteren Kapitel wird zumindest an einem Beispiel demonstriert, was bei Nicht-Erfüllung der zu Grunde liegenden Verteilungs-Annahmen geschehen kann. Einen Ausweg aus der Not bieten zwei Wege:

---

(p. 171-179). Ort: Verlag. – Stadler, P. (1985). *Die Seriation awarischer Gürtelgarnituren*. Diss. Univ. Wien. – Aktuell: Stadler, P. (2019). Homepage of "Serion Suite": <https://homepage.univie.ac.at/peter.stadler/index.html> [24.1.2020].

die Transformation von Variablen (so, dass sie annähernd normalverteilt sind – auch das wird noch vertieft werden) oder die Anwendung von „nicht-parametrischen Verfahren“. Aus solchen Verfahren, die weniger strenge Voraussetzungen an die Daten erheben, dabei allerdings bisweilen auch etwas weniger leistungsfähig sind, bildete sich in der 2. Hälfte der 1970er Jahre in der Statistik eine neue Philosophie, eine neue Schule, die sog. „Explorative Datenanalyse“, abgekürzt EDA. Der „Papst“ dieser Richtung und immer wieder zitierte Klassiker ist John Wilder Tukey „Exploratory data analysis“ (1977). Die EDA setzt konsequent solche Verfahren ein, die weniger strenge Annahmen an die Qualität der Daten machen. Also beispielsweise statt Mittelwert und Standardabweichung wie in der klassischen Statistik nun Median und Interquartilabstand verwenden; so genannte Boxplots sind eine typische Grafik dieser anderen Statistik-Philosophie.

An dieser Stelle muss an einen für die Forschungspraxis wichtigen Schritt der technischen Entwicklung erinnert werden: Im August 1981 brachte die Firma IBM den „Personal Computer“, den (IBM-) PC auf den Markt (**Abb. 3; Abb. 4**). Dieser sofort viel kopierte Standard und sein Betriebssystem, MS-DOS (ebenfalls 1981 ff.) leiteten in vielen Bereichen der IT den Abschied von den sog. Großrechnern ein und setzten sich bald gegen eine seinerzeit sehr lebendige Szene unterschiedlichster Rechner-Standards und Betriebssysteme als Marktführer durch. Die ehemals auf Großrechnern betriebenen Statistikprogramme wie z. B. SPSS und SAS waren seit etwa Mitte der 1980er Jahre auch für diese PCs verfügbar. Damit verlagerte sich der übliche Arbeitsplatz des Wissenschaftlers vom universitären Großrechenzentrum hin zum eigenen Schreibtisch. Ein Hinweis für die Stilkundler und Typologen: Auch diese Zeit hat ihr typisches Layout, denn nun ersetzen Matrixdrucker mit acht Nadeln die sog. Schnelldrucker der Rechenzentren; die geringe Auflösung und die vor allem unter Studenten gerne etwas zu lange gebrauchten Farbbänder führten oft zu nur bei sehr gutem Willen lesbaren Ausdrucken.

Mit diesem Übergang des Arbeitsplatzes an den PC einher geht ein beträchtlicher Zuwachs an allgemein verfügbarer Rechenkapazität, was heute fast vergessen ist. Der Rechner, auf dem die erste Fassung jenes Bonner Seriationsprogrammes BASP (später: WinBASP) lief, das seit 1975 bis in die Zeit um 2010 viel verwendet wurde,<sup>12</sup> war eine „PDP 11“ unter einem UNIX-ähnlichen Betriebssystem; sie hatte einen Kernspeicher von 64 kB (ähnlich **Abb. 2**). Zum Vergleich: im Jahr 2019 sind beim PC

---

<sup>12</sup> Die jüngste Fassung von WinBASP wurde 1997 herausgegeben und ist nur auf 32-Bit-Systemen lauffähig. WinBASP wird nicht mehr weiterentwickelt. Website: <http://www.uni-koeln.de/~al001/basp.html> [24.1.2020].

ca. 4 GB Rechenpeicher ein unteres Niveau, das man gerne auf 8 oder 16 GB hochrüstet, d. h. das 40.000-fache. Auch Rechenzeit war kostbar: Wenn ich um 1980 an dem ersten Großrechner, den ich selbst benutzt habe, der „CDC Cyber 76“ an der Universität Köln (vgl. **Abb. 1**), den Kernspeicher für eine relativ umfangreiche Seriations-Aufgabe länger als 60 Sekunden exklusiv nutzen wollte, war dies oft nur nachts – in Zeiten geringerer Auslastung durch andere Nutzer – möglich.



**Abb. 4** Eine kurze Geschichte von Speichermedien anhand der Sammlung des Autors. In stratigraphischer Folge im Untergrund eine Lochkarte aus den 1970er Jahren, eine 8“-Floppy-Disk (180 kb Speicher Vermögen; Mitte 1970er Jahre ff.), eine 5,25“-Diskette (720 kb oder 1200 kb; 1982/84 ff.), eine 3,5“-Diskette (1,4 MB; 1987 ff.) – heute noch tradiert als das Icon für das Speichern von Daten – und eine CD-ROM (650 MB), die seit Mitte der 1990er Jahre die „Flop-pys“ ablöste. Die heute vielverwendeten USB-Speichersticks wurden erst Ende 2000 eingeführt.

### 7.1.3 Die 1990er und 2000er Jahre

Der Rückblick auf die technische Geschichte ist weit mehr als nur Anekdote. Denn mit den mächtiger werdenden PCs entstand sukzessive eine neue Schule der Statistik, eine besonders rechenintensive Art, die praktisch erst seit den 1990er Jahren möglich wurde. Am Anfang stehen die Schlagworte „Bootstrapping“ (i. e. sich am Schnürsen-

kel aus dem Sumpf ziehen) und „Jackknifing“ (i. e. mit dem Taschenmesser zustecken), ein heute klassisches Lehrbuch dazu wurde 1993 publiziert.<sup>13</sup> Inzwischen spricht man erweitert von „Bayes-Statistik“. Im Kern geht es in der Theorie darum, nicht vorab Annahmen über die Natur der Daten machen zu müssen, sondern im Verlauf der Analysen aus den Daten selbst zu lernen; „a-posteriori-Wahrscheinlichkeit“ ist ein zentrales Schlagwort. Die praktische Umsetzung macht das statistische Arbeiten wieder kompliziert, und trotz des 1996 erschienenen Buches von Caitlin Buck über Bayes-Statistik in der Archäologie werden solche Verfahren in der Archäologie bislang nicht regelhaft eingesetzt. Einige wichtige Ausnahmen seien genannt, damit Interessierte dort weiterlesen können: (a) die Kalibration von <sup>14</sup>C-Messungen – das beliebte Programm OxCAL benutzt Techniken der Bayes-Statistik; (b) die von Werner Vach statistisch betreute Auswertung der epigenetischen Merkmale für Verwandtschaftsdiagnosen, die Kurt W. Alt (2004) vorgestellt hat, und (c) das von John Hines und Alex Bayliss publizierte Chronologieprojekt über die Merowingerzeit in England (2013). Praktisch machbar ist Bayes-Statistik nur, in dem bestimmte umfangreiche Rechenprozeduren sehr oft wiederholt werden, sprich: erst mit Hilfe der seit der Jahrtausendwende verfügbaren Hardware.

Ein weiteres Schlagwort der 1990er Jahre ist „fuzzy-coding“ oder „Fuzzy-Logik“. Der dahinterstehende Gedanke lässt sich im Rückgriff auf Schulstoff leicht erklären. Es geht um die Mengenlehre. Ein erster zentraler Satz ist: „A ist Element einer Menge oder ist nicht Element einer Menge“. Sprich: eindeutig Ja oder eindeutig Nein sind gefragt und die zentrale Voraussetzung jeder weiteren logischen Arbeit. Demgegenüber versucht die „Fuzzy-Logik“, unscharfe Informationen besser zu verwenden und ohne diese strenge Forderung auszukommen. Salopp formuliert: Sie versucht, ein Ja zuzulassen, im Sinne „A ist ein bisschen Element der Menge und ist es auch ein bisschen nicht“. Umfangreiche Anwendungen solcher Verfahren in der Archäologie sind selten geblieben, aber in den 1990er/2000er Jahren wurden mehrere Artikel publiziert, in denen Daten einem „fuzzy-coding“ unterworfen wurden – also einer Codierung, die auf eine Eindeutigkeit der Zuordnung verzichtet.<sup>14</sup>

---

<sup>13</sup> Efron, B. & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. (Monographs on Statistics and Applied Probability, 57). New York: Chapman & Hall. Weiterhin: Michael R. Chernick, M. R. (1999). *Bootstrap Methods. A practitioner's guide*. (Wiley Series in probability and statistics). New York: John Wiley & Sons; Good, Ph. I. (1999). *Resampling methods. A practical guide to good data analysis*. Boston: Birkhäuser.

<sup>14</sup> Ein m. W. frühes Beispiel: Höhn, B. & Mattheußer, E. (1991). «Fuzzy coding» oder die Möglichkeit zur Verarbeitung rangskalierteter Variablen in der Korrespondenzanalyse am Beispiel Michelsberger Keramikdaten. *Archäologische Informationen* 14(2), 267-273.

Technisch erbrachten die 1990er Jahre eine schleichende Veränderung des Verhaltens der typischen PC-Benutzer. Die grafische Benutzeroberfläche Microsoft-Windows wurde zwar seit 1985 angeboten (bei Apple mit dem wenig erfolgreichen „Apple Lisa“ seit 1983 und dem erfolgreichen Macintosh seit 1984), doch erst seit den Versionen Windows 3.1 (1992) und vor allem Windows NT (1993) als ernsthaftes, im Beruf taugliches Betriebssystem wahrgenommen und wirklich weit verbreitet. Zuvor waren PCs mit direkten Texteingaben in sog. Kommandozeilen gesteuert worden. Parallel zum Betriebssystem veränderten sich nun auch die Statistikprogramme und mit ihnen die Gewohnheiten und Bedürfnisse der Anwender hin zum „Point-and-Click“, also der Auswahl von Prozeduren per Mausklick aus Menüs. Anwender, die das kommandozeilen-orientierte Arbeiten am PC vor Mitte der 1990er Jahre nicht mehr kennengelernt haben, tun sich mit der Steuerung von Programmen per Kommandozeile erfahrungsgemäß schwer, etwas, was hinsichtlich der hier verwendeten Software **R** sehr relevant ist.

Nach diesem Ausblick auf die jüngeren Entwicklungen in der Statistik wieder zurück zur Archäologie: In den späten 1970er und vor allem in den 1980er Jahren wurde die Beschäftigung mit räumlicher Information intensiviert. Frühe Pionierarbeiten sind die von C. Vita-Finzi & E. S. Higgs zu Palästina und von H. Hietala in den USA.<sup>15</sup> Im Schwerpunkt geht es um die Verbreitung archäologischer Funde in der Fläche, einmal auf Fundplatz-Niveau – „intrasite spatial analysis“ genannt – und einmal für Fundplätze in der Fläche („intersite spatial analysis“). Das danach allmählich aufkommende Schlagwort GIS „Geografisches Informationssystem“ hat seinen Ursprung in der Geografie, die bemüht ist, flächige Informationen – also weniger Punktdaten – auszuwerten, insbesondere durch Überlagerung und Verschneidung von Karten.<sup>16</sup> In der Praxis wird heute jedoch weniger die statistische Analyse, sondern bereits die pure Visualisierung von flächigen Informationen in Karten als „GIS“ bezeichnet.

---

<sup>15</sup> Vita-Finzi, C. & Higgs, E. S. (1970). Prehistoric Economy in the Mount Carmel Area of Palestine: Site Catchment Analysis. *Proceedings of the Prehistoric Society* 36, 1-37. – Hietala, H. J. (1984). *Intrasite spatial analysis in archaeology*. Cambridge: Cambridge Univ. Press.

<sup>16</sup> Für die deutsche Archäologie wichtige Pionierarbeiten: Sielmann, B. (1971). Der Einfluß der Umwelt auf die neolithische Besiedlung Südwestdeutschlands unter besonderer Berücksichtigung der Verhältnisse am nördlichen Oberrhein. *Acta Praehistorica et Archaeologia* 2, 65-197. – Linke, W. (1976). *Frühestes Bauerntum und geographische Umwelt. Eine historisch - geographische Untersuchung des Früh- und Mittelneolithikums westfälischer und nordhessischer Bördenlandschaften*. (Bochumer Geographische Arbeiten, 28). Paderborn: Schöningh. – Dohrn-Ihmig, M. (1979). Bandkeramik an Mittel- und Niederrhein. In: *Beiträge zur Urgeschichte des Rheinlandes*, 3. (Rheinische Ausgrabungen, 19). (S. 191-362). Köln: Rheinland-Verlag.

Es ist naheliegend, dass das Kartieren in diesem Sinne eine Routineaufgabe in der Bodendenkmalpflege ist, und das Thema GIS heute gerade dort ein oft verwendetes Schlagwort ist.

#### **7.1.4 Neue Vielfalt und Verschiebungen in den 2000ern**

Die beschriebenen „Schulen“ erhalten in den 2000er Jahren Zuwachs durch neue Zentren, die nicht aus diesen alten Schulen erwachsen, sondern eher autonom entstanden sind. Ohne Anspruch auf Vollständigkeit sei hier W. Schier genannt, der – seit 2006 an der FU Berlin etabliert – vor allem landschaftsarchäologische Projekte verfolgt und dafür intensiv GIS einsetzt, sowie J. Müller (seit 2004), der in Kiel dank starker Drittmittelförderung viele Doktoranden und Postdoktoranden an sich zog, die vor allem auf freie Software und **R** setzen. Nicht zuletzt wird die Szene der Aktiven bereichert durch Kai-Christian Bruhn, seit 2008 an der Hochschule Mainz und seit 2015 Direktor von „mainzed“, dem „Mainzer Zentrum für Digitalität in den Geistes- und Kulturwissenschaften“, einem multidisziplinären Netzwerk zur Bündelung digitaler Kompetenzen von sechs wissenschaftlichen Einrichtungen am Wissenschaftsstandort Mainz, zu denen auch das RGZM gehört. Damit ist die Landschaft der Akteure etwas unübersichtlicher, vor allem aber vielfältiger geworden.

Der bisherige Rückblick auf die Entwicklung betraf vor allem die Ur- und Frühgeschichte, und zwar nicht nur, weil der Autor dieser archäologischen Richtung entstammt: Statistik und IT hielten in vielen anderen Archäologien erst deutlich später Einzug. Dies änderte sich in Deutschland seit Mitte der 1990er Jahre, als in Göttingen (J. Bergemann) und Köln (R. Förtsch) Bildarchive aufgebaut wurden, sprich: bedeutende analoge Fototheken und Diatheken digitalisiert wurden (Köln: ARACHNE). Mit diesem weniger auf Kontexte und Datenanalyse zielenden Forschungsbedarf, denen es vor allem um eine effiziente Objektverwaltung und -erschließung ging, veränderten sich die inhaltlichen Schwerpunktsetzungen. Denn nach längerer intensiver Kooperation zwischen der Kölner Klassischen Archäologie und dem DAI in Berlin wechselte R. Förtsch 2012 ans DAI und wurde dort als Direktor der IT-Abteilung tätig. Ganz im Sinne dieser Verschiebung des Interessenschwerpunktes entstand am DAI das große, drittmittelgeförderte Projekt IANUS (2011-2017),<sup>17</sup> das – weit über ein Fotoarchiv hinauszielend – ein universelles Archiv für Forschungsdaten aufbauen wollte, das insbes. auch Grabungsdokumentationen aufnehmen soll. In dem vom

---

<sup>17</sup> IANUS Forschungsdatenzentrum Archäologie & Altertumswissenschaften: <https://www.ianus-fdz.de/> [24.1.2020].

Bund 2019 ff. lancierten Vorhaben, eine „Nationale Forschungsdateninfrastruktur“ (NFDI) zu schaffen, findet diese Ausrichtung innerhalb der Archäologie vermutlich eine erneute Stärkung und Fortführung („NFDI4objects“). Mit dem Bedürfnis, nachhaltig zu Archivieren, kommen ganz andere technische Aspekte und auch Themenschwerpunkte auf, wie z.B. Thesauri, Metadaten, die Fürsorge für und Anschlussfähigkeit an komplexere und stärker unterschiedliche Datenformate und auch das ebenso wichtige wie komplexe Thema von Autorenschaften und Zugriffsrechten. Ein gegen Ende des Projekts IANUS veröffentlichtes Positionspapier von M. Heinrich verdeutlicht die Akzentverschiebung exemplarisch:<sup>18</sup> Von 13 dort aufgeführten Themenfeldern, in denen junge Archäologen s. E. Mindestfähigkeiten sollten vorweisen können, sind Statistik und GIS nur mehr zwei Felder.

Den Autor befremdet diese Schwerpunktsetzung, denn die Betonung von Infrastrukturen, Datenarchiven, Metadaten, rechtlichen Aspekten und der Möglichkeit der Nachnutzung von Daten verschiebt die Aktivitäten (wie auch Stellen und Mittel) weg von der Archäologie hin auf eine IT, und zunächst einmal weg von der Erkenntnisgewinnung hin zur Wissensverwaltung. Nicht zuletzt erfordert sie zunächst einmal das Vorhandensein von Daten und die Bereitschaft der Datenautoren, diese auch zu publizieren. Angesichts der seit Anfang der 2010er Jahre auch in der deutschen Archäologie wachsenden Schere zwischen der gegebenen Möglichkeit, Daten sorgsam in öffentlichen Archiven zu publizieren, und der tatsächlichen, nämlich ausnehmend geringen Nutzung dieser Möglichkeiten, ist mir der Bedarf an komplexen Archiven in dem von den genannten Projekten angestrebten Umfang nicht ersichtlich.

### **7.1.5 Stagnation in den 2010er Jahren?**

Die Gegenwart ist stets schwieriger einzuordnen als die Vergangenheit, Prognosen erweisen sich im Nachhinein oft als falsch. Dennoch sei der Versuch einer kurzen subjektiven Situationsbeschreibung unternommen. Ansehnliche farbige Kartierungen, farbige Befundbilder, 3D-Darstellungen und LIDAR-Bilder ziehen derzeit verstärkt auch in die gedruckten Publikationen ein, während zugleich die theoretische Beschäftigung mit Statistik in der Archäologie und die Methodenentwicklung zu stagnieren scheinen. Am augenfälligsten ist dies bei der Verwendung des Begriffs GIS, wo das

---

<sup>18</sup> Heinrich, M. (23.10.2017). Statement zu minimalen IT-Kenntnissen für Studierende der Altertumswissenschaften. DAI IANUS. [https://www.ianus-fdz.de/attachments/download/951/Statement-zu-IT-Kenntnissen\\_2017-10-23.pdf](https://www.ianus-fdz.de/attachments/download/951/Statement-zu-IT-Kenntnissen_2017-10-23.pdf) [24.1.2020].

Anfertigen schöner, gut gestalteter Karten bereits als GIS gilt, während man ursprünglich die statistische Analyse von raumbezogenen Informationen mit dem Begriff GIS verband. Die Stagnation der Anwendung statistischer Methoden in der Archäologie steht m. E. in einem Zusammenhang mit geänderten Forschungsinteressen. Während die New Archaeology / Prozessuale Archäologie am Aufdecken von Regeln und Allgemeingültigkeiten interessiert war, zielt die Postprozessuale Archäologie mehr auf Ereignisse und Individuen, blickt darauf, inwieweit wir Heutigen eine Vergangenheit konstruieren. An die Stelle von Modellbildung und deduktiven Schlüssen tritt der Analogieschluss, viele Archäologen suchen Anregungen von Nachbarwissenschaften weniger als zuvor bei den Naturwissenschaften, sondern z. B. bei der Soziologie und der Philosophie. Wo man über „Bilder“, „Wahrnehmungen“ und „Diskurse“ forscht und mit Dekonstruktionen im Umfang von Monografien wissenschaftliche Karrieren machen kann, haben es Tatsachen, Fakten und quantitative Methoden schwer. Nicht, dass sich der Forschungsstand wirklich geändert hat – aber die Stimmung und die Forschungsinteressen scheinen sich verschoben zu haben.

Notwendig ist dieser Weg nicht. Denn im wissenschaftlichen Publikationswesen setzt sich in den 2010er Jahre sukzessive das Prinzip Open Access durch, es ist ein zentrales Anliegen der EU-Forschungsförderung der Jahre 2014-2020 und seit Mitte der 2010er Jahre auch der deutschen Forschungsförderung. Drittmittelgeförderte Projekte sind mehr und mehr gehalten, bei Projektabschluss nicht nur ihre Auswertungen, sondern auch ihre Daten frei zugänglich zu publizieren. Auch die um 2010 stark in die Öffentlichkeit gerückten Plagiats- und Fälschungsskandale haben dazu beigetragen, die Forderung nach Open Data zu nähren: Die Veröffentlichung wissenschaftlicher Daten soll so erfolgen, dass sie nach ihrer Primärauswertung frei zugänglich und weiterverwertbar sind. Damit entsteht derzeit, eher bald als mittelfristig, ein großer Fundus an frei zugänglichen und elektronisch lesbaren wissenschaftlichen Daten und Information, in denen hohes Potenzial für eine weitere Intensivierung der statistischen Analysen liegt. Denn zu jedem eigenen Thema und Datensatz steht dann eine wachsende Anzahl an relativ unaufwendig erreichbaren Vergleichsinformationen zu Verfügung, was zu gemeinsamen Analysen und Re-Analysen „alter“ Daten einlädt.

Ich sehe in diesem Ansatz, dem gezielten Hinzuziehen bereits publizierter Daten in die eigenen Daten und Analysen, mehr Potenzial als in den derzeit auch in den Geisteswissenschaften beliebten Begriffen „Data Mining“ und „Big Data“. Denn einerseits wird in den 2010er Jahren viel von Big Data gesprochen und versucht, Infrastrukturen für das Speichern und Wiederfinden von Open Data möglichst in großen, zentralen Repositorien aufzubauen, doch konkrete Anwendungen dieser Ansätze (*sensu* wissenschaftlich ertragreiche Auswertungsprojekte) sind in den Archäologien

m.W. so gut wie inexistent.<sup>19</sup> Die Techniken, mit denen Unternehmen wie z. B. Google, Facebook oder Amazon die Daten ihrer Nutzer auswerten, mögen dort (im Sinne dieser Firmen) zielführend sein, für die Archäologien bezweifle ich dies.

Ein gewisses Befremden gegenüber dem vorherrschenden Zeitgeist in der Archäologie wird dem Autor erlaubt sein. Das Ungefähre in der derzeit als modern geltenden „postmodernen“ (postprozessualen) Archäologie, die lieber ihr Nicht-Wissen-Können betont als ihr Wissen, die lieber spektakuläre Einzelfälle des Typs „Luthers Kloake“ oder „das verliebte Paar im Grab“ denn ganze Gesellschaften und das Allgemeine untersucht, die sich lieber mit der Forschungsgeschichte als mit der Forschung beschäftigt, während man zugleich ebenso begeistert wie bemerkenswert unkritisch die Angebote des (scheinbar) Exakten und Sicherem aus dem Labor, d. h. aus den Naturwissenschaften aufgreift, führt die Archäologie ins Abseits. Ja, die Zahlen aus dem Labor wie etwa zum Verhältnis stabiler Isotopen sind exakt. Doch solche Zahlen ohne eine geisteswissenschaftliche Quellenkritik, ohne das Herstellen des archäologischen Kontextes und ohne eine historische Deutung sagen wenig, d. h. hier ist statt des Laboranten der fähige und auch statistik-kundige Archäologe gefragt. Aber vor allem: Auch die Beobachtung von Stratigrafien, das Auszählen und Vermessen von Silices, die Klassifikation von Keramikgefäßen etc. wie auch die Analyse und der Vergleich von Ensembles („geschlossener Fund“) sind exakt und können als nachprüfbare Prozesse aufgesetzt werden. Es besteht kein Grund für die Archäologie, ihre ureigensten Quellen und Methoden für geringerwertig zu halten als Laborergebnisse. Bei diesem Arbeiten mit den archäologischen Quellen im engeren Sinne ist eine sorgfältige Methodik inklusive der Statistik notwendig – so die Überzeugung und Haltung des Autors. Insofern mag diese Einführung manchem als ein unzeitgemäßes Werk erscheinen. Es möchte wieder dazu anregen und motivieren, dass die Archäologie selbst auf den Weg der Tatsachen, des Wissens und der exakten Darlegung und Analyse ihrer Informationen zurückfindet.

Pessimistisch? Nein. Denn mit dem Aufblühen der Themen „Open Data“, „Open Science“ und der wieder neu aufkommenden wissenschaftlichen Ethik einer Nachvollziehbarkeit von Ergebnissen gegen Ende der 2010er Jahre scheint sich ein

---

<sup>19</sup> Mitte der 1990er Jahre wurde in Großbritannien mit „ADS Archaeological Data Service“ das m.W. erste und heute noch aktive, moderne und leistungsfähige Datenarchiv gegründet, das zu Metaanalysen geradezu einlädt (<https://archaeologydataservice.ac.uk/> [24.1.2020]). Vgl. dazu die Beiträge im Fokus „Open Access und Open Data“ in den *Archäologischen Informationen* 38 (2015). <https://journals.ub.uni-heidelberg.de/index.php/arch-inf/issue/view/2578> [24.1.2020].

Generationenwechsel auch in der Archäologie anzudeuten, mit dem einher das empirische Arbeiten – nun oft eben mit Hilfe von **R** – wieder zurückkehrt und an Reputation gewinnt.

## **8 ERSTER SCHRITT: PERSÖNLICHE BEDÜRFNISSE UND MÖGLICHKEITEN KLÄREN**

---

Diese Einführung zielt im Grunde auf Interessierte, die sich wirklich tief einarbeiten wollen. Aber genau dieser Aspekt sollte vorab geklärt werden: Was sind Ihre persönlichen Bedürfnisse und Möglichkeiten? Denn die Wahl der Wege und Werkzeuge sollte sich an den individuellen Bedürfnissen orientieren. Daher sei zunächst versucht, typische Interessenlagen zu skizzieren und die dazu passenden Wege, diesen Bedürfnissen in geeigneter Weise nachzugehen.

### **8.1 „MAL SCHNELL EIN PAAR STATISTIKEN BEKOMMEN“**

Manchmal liegt nur ein flaches Interesse am Arbeiten mit Statistik vor, oder einfach nur zu wenig Zeit samt Termindruck: Man braucht einmalig und schnell ein paar Zahlen und Grafiken. Man hat keine Zeit, sich einzuarbeiten und möglicherweise auch das Gefühl, dass man mit Zahlen noch nie gut konnte.

Lösung: Bitten Sie einen Experten um Beratung und Hilfe. Denn ein schnelles „sich eben Mal Einarbeiten“ gibt es auf diesem Felde nicht. Statt husch-husch etwas ggf. Unseriöses oder gar Falsches abzuliefern, sollten Sie sich Rat von Erfahrenen holen, ggf. auch professionelle Beratung einkaufen. Denn Ihre Lern- und Einarbeitungszeit, um selbstständig verlässlich gute Ergebnisse erzielen zu können, wird nicht klein sein. Ersparen Sie sich diesen Stress.

### **8.2 ÖFTER EINFACHE STATISTIKEN: TABELLENKALKULATION**

Wer öfter einmal relativ einfache Statistiken und statistische Grafiken benötigt wie z. B. eine Mittelwertberechnung, den seriösen Vergleich von Häufigkeiten („Prozentzahlen“) oder ein Säulendiagramm, dem ist mit einer Tabellenkalkulation wie z. B. LO-Calc oder dem weitverbreiteten MS-Excel gut gedient. Damit kann man Daten erfassen, sortieren und filtern (also Informationen fragestellungsorientiert auswählen), und anschließend überraschend viele Statistiken rechnen und Grafiken anfertigen. Wir

werden dies in späteren Kapiteln vertiefen, Ihre Bibliothek oder eine gute Buchhandlung liefern ggf. unterstützende Literatur. Wer aller Voraussicht nach in absehbarer Zukunft nicht mehr braucht, kann den Aufwand für das Erlernen weiterer Programme sparen. Denn es ist besser, ein Instrument gut zu beherrschen, als viele Programme nur oberflächlich.

Wir werden in diesem Buch das beispielhafte Arbeiten mit LO-Calc relativ früh wieder zugunsten anderer Programme aufgeben. Das heißt nicht, dass diese Themen nicht mit LO-Calc bewältigbar wären. Mit etwas Geduld und Suchen geht hier mehr, als die übersichtliche Oberfläche auf den ersten Blick zeigt.

### **8.3 ERNSTHAFT AUCH KOMPLEXERE AUFGABEN LÖSEN, OHNE PROFI WERDEN ZU WOLLEN**

Wer ernsthaft als Archäologe selbstständig mit Statistik arbeiten möchte und tiefgehende Fragen resp. Anforderungen an das Methodenspektrum hat, als es eine Tabellenkalkulation beantwortet, sollte sich in eine professionelle Software einarbeiten. In dieser Einführung werden wir dazu das Statistikpaket **R** benutzen, das ausnehmend mächtig und flexibel ist, doch leider eine nennenswerte Lern- und Einarbeitungszeit benötigt. Wer diese Zeit nicht hat, aber am Ende der Möglichkeiten seiner Tabellenkalkulation angelangt ist, kann eine Zwischenlösung verfolgen: Datenerfassung und -verwaltung mit der gewohnten Tabellenkalkulation, anschließend eine statistische Auswertung mit Programmen wie z. B. PAST (das hier nachdrücklich empfohlen sei), JAMOVI oder dem R-nahen BlueSky Statistics. Die genannten Programme sind durchweg kostenlos, seriös und vergleichsweise schnell erlernbar. Sie werden im nächsten Kapitel näher vorgestellt werden.

### **8.4 ERNSTHAFT UND NACHHALTIG SELBSTSTÄNDIG ARBEITEN**

Wer sich gründlich in das Thema statistischer Auswertungen für archäologische Fragestellungen einarbeiten will (oder muss) und anfangs etwas mehr Lernaufwand investieren kann, sollte zu einem der bewährten Statistikpakete greifen. Unter den Bezahlprogrammen, die hier nicht weiterverfolgt werden, nenne ich exemplarisch SPSS oder SAS, die an Forschungsinstituten und Universitäten weit verbreitet sind. Für sie gibt es zahlreiche Kurse sowie gute Handbücher und Einführungsliteratur. Beide Programme sind einander ähnlich, SPSS ist in den Geistes- und vor allem den Sozialwissenschaften bekannter und Quasi-Standard, SAS eher in der Medizin und vielen Naturwissenschaften – wohinter aber eher fachspezifische Traditionen denn Unterschiede im Leistungsvermögen stecken.

In dieser Einführung arbeiten wir stattdessen mit **R**, einem seit ca. 1995 aus „S“ (einem kommerziellen Statistikpaket) heraus entwickelten Open-Source-Projekt. Im Jahr 2000 erschien die Version **R** 1.0 – also die erste vollständige und reife Programmversion – seitdem wächst das Programm weiterhin stark, sowohl in seinem Leistungsumfang wie auch in seinem Benutzerkreis. Mitte der 2010er Jahre darf man **R** durchaus als das Werkzeug professioneller Statistiker bezeichnen. **R** ist kostenlos und läuft unter Windows, unter Linux und auf dem Mac. Der Vorteil einer Entscheidung für **R** liegt nicht nur in der freien, kostenlosen Verfügbarkeit. In **R** kann man typische und häufiger gebrauchte Arbeitsabläufe als „Skripte“ speichern und wiederbenutzen, also sie auch wie ein Computerprogramm an Dritte weitergeben, was dank der offenen Lizenz rundum erlaubt ist und beim Gegenüber auch keine Kosten verursacht. In **R** kann man ausnehmend komplexe Aufgaben verfolgen und erfolgreich realisieren, d. h. **R** wächst gemeinsam mit Ihren Problemstellungen und Ihrer Expertise, während das Arbeiten beispielsweise mit Tabellenkalkulationen recht bald an eine gläserne Decke stoßen lässt: Dinge, die Sie gerne realisieren möchten, sind nicht möglich resp. nicht mit vertretbarem Aufwand umsetzbar.

## 8.5 EMPFEHLUNGEN FÜR DAS ALLMÄHLICHE SICH-VORTASTEN

Es gäbe dieses Buch nicht, wenn ich die eingangs skizzierte Option „Finger davonlassen“ & „Spezialisten hinzuziehen“ als die universell angemessene Lösung empfehlen wollte. Sich mal über ca. eine Woche hinweg täglich ein paar Stunden mit einer Tabellenkalkulation wie z. B. LO-Calc näher zu befassen, um damit Datenerfassung, -verwaltung, Rechnen und einfache statistische Auswertungen zu erlernen, ist ein sinnvolles Unterfangen. Was man sich hierbei erarbeitet, bleibt dauerhaft nützlich, weil sich mit diesem Werkzeug nach kurzem Einarbeiten viele kleine Alltags-Aufgaben gut und effizient erledigen lassen.

Willkommen, wer danach eine weitere Zeitdosis dieser Menge in Richtung auf die weitergehende Archäo-Statistik investieren kann! Die Vielfalt der nachfolgend dargestellten Instrumente ordne ich persönlich nach Stand 2020 wie folgt:

- PAST kann (sehr) viel, und wer als Archäologe in das Thema Korrespondenzanalyse einsteigen will (Siegmond 2015), findet hier ein gutes Werkzeug für alle nicht allzu umfangreichen Projekte. Nach meinen Erfahrungen ist jedoch bei vielen archäologischen Auswertungen die Datenverwaltung ein sehr wichtiger Teil, der zudem nicht einmal durchlaufen wird und dann abgeschlossen ist, sondern der die Projekte anhaltend begleitet. Das ist in PAST nur holprig möglich, und der in diesem Buch skizzierte,

mögliche Weg eines steten Hin und Her zwischen LO-Calc & PAST ist mir persönlich auf Dauer zu mühsam.

- Viele mit **R** vertraute Anwender empfehlen, **R** nicht über eine grafische Benutzeroberfläche anzugehen, sondern gleich „richtig **R** zu lernen“, d. h. **R** im Kommandozeilen-Modus zu bedienen. Davon rate ich ab. Denn ich persönlich brauche Erfolgserlebnisse, die nicht allzu lange auf sich warten lassen, und wichtiger: meist steht vor mir eine konkrete Aufgabe, die nicht zuletzt auch mit eng gesetzten Zeitzielen verbunden ist. Wer als Anfänger zuverlässig zeitnah gute Lösungen erzielen muss (Beispiel: BA- oder MA-Arbeit), sollte **R** zunächst mit einer grafischen Benutzeroberfläche bedienen. Und ja: er sollte diese Option wählen, statt sich in PAST zu verlieren. Eine Ausnahme macht allein die Korrespondenzanalyse, die in **R** zwar sehr gut vorhanden ist, aber z. Zt. nur via Kommandozeilen-Modus durchführbar ist.

- Von den nachfolgend vorgestellten grafischen Benutzeroberflächen zu **R** empfehle ich, sich JAMOVI, BlueSky und vor allem das Arbeiten mit dem Paar **R** & R-Commander näher anzuschauen. „Näher anschauen“ heißt, die kostenlosen Programme zu installieren, die entsprechenden Kapitel in diesem Buch nachzuarbeiten und jeweils ein kleines eigenes Projekt darin zu lösen, um wirklich vergleichen zu können. Wer bereits weiß, dass er danach den Weg weitergehen will, **R** auch im Kommandozeilen-Modus zu erlernen, ist m. E. bei JAMOVI oder vor allem dem Paar **R** & R-Commander besser aufgehoben. Wer daran zweifelt, ob er sich je dem Kommandozeilen-Modus nähern will, kann mit BlueSky dank dessen Datenverwaltungswerkzeugen und reichem Statistik-Angebot (weit über die Möglichkeiten von z. B. LO-Calc hinaus) sehr weit kommen und viele Aufgaben lösen, bevor er eine gläserne Decke erreicht und das Umsteigen zur Kommandozeile erwägen muss.

- Caveat. Es ist offensichtlich, dass **R** im Bereich der professionellen Anwendungen im Laufe der 2010/20er Jahre die älteren Software-Klassiker wie z. B. SPSS oder SAS ablöst und zum neuen Standard wird – und es für längere Zeit auch bleiben wird.<sup>20</sup> Doch alles Weitere entwickelt sich z. Zt. sehr schnell und mit unklarer Prognose. So wächst z. B. derzeit um das moderne und elegante JAMOVI herum eine lebendige Entwicklergemeinde, was dazu führen könnte, dass JAMOVI demnächst *die* grafische Benutzeroberfläche von **R** werden wird. Doch ebenso gut ist denkbar, dass andere

---

<sup>20</sup> Dazu z.B.: Muenchen, Robert A. (2017, 2019). The Popularity of Data Science Software. *Blog R4stats*, <http://r4stats.com/articles/popularity/> [24.1.2020].

Optionen diesen Platz einnehmen werden.<sup>21</sup> Insofern ist die grobe Richtung der Entwicklung offensichtlich, wichtige Details jedoch sehr offen.

Nicht zu vergessen: Es geht nicht allein um die Wahl und das Erlernen einer Software. Wer auf dem Feld Archäologie & Statistik vorankommen will, lebt im Dreikampf Archäologie – Statistik-Theorie – Software. Alle Überlegungen und Zeitkalkulationen zum Einarbeitungsaufwand und zur Lernkurve, die sich allein an der Software orientieren, unterschätzen die beiden anderen Komponenten und das notwendige Ringen um das fruchtbare Zusammenwirken dieser Trias. Daher wiederum sollte man nicht seine kompletten Ressourcen allein für das Erlernen einer Software verplanen.

## 9 TEIL II: WERKZEUGE UND GRUNDLAGEN

### KENNENLERNEN UND AUSWÄHLEN

---

Im folgenden Teil geht es darum, einige wichtige Werkzeuge kennenzulernen. Sie benötigen später nicht alle diese Werkzeuge, sondern die Auswahl hängt auch von Ihren persönlichen Bedürfnissen ab. Aber um das für Sie Passende zu finden und ein gutes Zusammenspiel von Bedürfnissen und Werkzeugen zu erreichen, werden hier verschiedene Optionen vorgestellt. Dabei ist das erste Werkzeug, die Tabellenkalkulation, grundlegend. Vermutlich kennen Sie es bereits und können diesen Abschnitt überfliegen. Wenn nicht, sollten Sie sich jetzt darin mindestens so weit einarbeiten wie nachfolgend skizziert und sich dann mit dem Thema Datenerfassung und -verwaltung beschäftigen. Bei den anschließend vorgestellten Instrumenten PAST, JASP, JAMOVI und BlueSky Statistics geht es darum, für Sie einfach zu erlernende Programme vorzustellen, damit Sie Ihre Aufgaben ggf. niederschwelliger, d. h. auch ohne das Arbeiten mit **R**, weiterverfolgen können. Abschließend stelle ich **R** vor. Im gesamten Abschnitt geht es also nicht zuvorderst um Statistik und Archäologie, sondern darum, einen Einblick in die Software zu gewinnen, sodass man sich danach auf die Inhalte konzentrieren kann.

#### 9.1 TABELLENKALKULATION LO-CALC

Für die Datenerfassung und -verwaltung sowie einfache statistische Berechnungen und Grafiken kann eine Tabellenkalkulation nützlich sein. Statt des weit verbreiteten

---

<sup>21</sup> So warte ich z. B. mit Neugier auf den zunächst für Ende 2018, dann für März 2020 angekündigten Start des spannenden französischen Projekts „r++“ (alias „Rplusplus“): <https://rplusplus.com/en/> [24.1.2020].

kommerziellen Programms MS-Excel verwenden wir hier die Open-Source-Alternative Calc aus dem Office-Paket LibreOffice (LO). LibreOffice wird für Windows-, MAC- und Linux-Betriebssysteme angeboten. Neben ihrem eigenen Datenformat (\*.ods) lesen und schreiben die Programme von LO unter anderem auch die weit verbreiteten Formate des Microsoft-Office-Paketes (\*.xls, \*.xlsx), sodass ein Dateiaustausch mit Dritten gemeinhin problemlos ist, auch wenn dort ein anderes Programm zur Anwendung kommt. Die Installation von LibreOffice rundet die persönliche Arbeitsumgebung ab, das Programm findet sich hier: <https://de.libreoffice.org/> Anfang 2020 waren die Versionen 6.2.8 (sicher, „stable“) resp. 6.3.4 („innovativ“) aktuell; die Anleitungen hier und die Screen-Shots beziehen sich auf die Version 6.3.4.

Wer unter Linux arbeitet, kann alternativ zu LO-Calc das Arbeiten mit der sehr leistungsfähigen Tabellenkalkulation „Gnumeric“ erwägen, die hier jedoch nicht weiter behandelt wird.<sup>22</sup>

### 9.1.1 LibreOffice Calc: Installation und erste Schritte

LibreOffice weist auf seiner Website seit einiger Zeit zwei Programmversionen aus: eine stabile Fassung („stable“) für den sicherheitsbedürftigen, konservativen Benutzer und eine zwar getestete, aber noch in der Entwicklung befindliche aktuellere Fassung („innovativ“, Early Adaptor) für erfahrene Nutzer, welche die angebotenen Neuerungen der demnächst kommenden stabilen Fassung schon früh benötigen. Wählen Sie das für Sie passende aus – auch hinsichtlich des Betriebssystems –, laden dieses herunter und installieren Sie das Programm: Doppelklick auf das Bildsymbol (Icon) des heruntergeladenen Pakets und dann den Anweisungen folgen. Wegen der Größe des Programms dauert das Herunterladen und vor allem das Installieren meist ein paar Minuten. Wenn noch genug Platz auf Ihrer Festplatte vorhanden ist, dürfte es sinnvoll sein, auch die Offline-Hilfe zu LibreOffice in Ihrer Wunschsprache herunterzuladen und zu installieren. Vergessen Sie nicht, nach erfolgreicher Installation die Installationsdateien wieder aus ihrem Download-Ordner zu löschen, denn LibreOffice macht das (noch?) nicht automatisch. Nach der Installation befindet sich (sofern Sie es nicht anders bestimmt haben) das Programm-Icon von LibreOffice auf Ihrem Desktop. Nach einem Doppelklick auf dieses Icon startet LibreOffice und bietet in der Navigationsleiste links verschiedene Programme zur Auswahl an: Writer, Calc, Impress usw.

---

<sup>22</sup> Website: <http://www.gnumeric.org/> [24.1.2020]. Die bis August 2014 vorhandene Windows-Version von Gnumeric wird nicht weiterentwickelt.

„Calc“ ist die MS-Excel-ähnliche Tabellenkalkulation von LibreOffice, mit der wir arbeiten wollen.

Da das Arbeiten mit Tabellenkalkulationen weit verbreitet und vielen bekannt ist und es ggf. gute Anleitungen gibt, folgt jetzt keine ausführliche Einführung in LO-Calc. Wir werden aber bei den Themen Datenerfassung und -verwaltung sowie einfache Statistiken und Grafiken auf LO-Calc zurückkommen.

## 10 DATENERFASSUNG UND -VERWALTUNG MIT EINER TABELLENKALKULATION

---

### 10.1 ÜBLICHE ANFÄNGERFEHLER

Datentabellen sind keine Texttabellen, Zahlen sind keine Buchstaben. Datentabellen sollten von Anbeginn an auf eine leichte Erfassung der Daten und ihre spätere Auswertbarkeit hin optimiert werden. Anhand typischer Anfängerfehler sollen ein paar grundlegende Regeln für Datentabellen erklärt werden.

- Bis auf seltene Ausnahmen gilt für eine gut verwendbare Datentabelle: Jeder Fall bildet eine eigene Zeile, jede Information erhält eine eigene Spalte.
- *Problem:* Am Beispiel der Erfassung von Messstrecken wie z. B. der Höhe von Gefäßen oder der Länge von Oberschenkelknochen lässt sich ein weiteres Thema gut beschreiben. Neben dem Messwert selbst hat der Erfasser oft das Bedürfnis, wichtige zusätzliche Informationen zu speichern, wie z. B. statt „428 mm“ korrekter „ca. 428 mm“, „mind. 428 mm“ oder „428 ?“. Sobald man dies wie gedacht in die Zelle einer Tabelle einträgt, beinhaltet diese Zelle Zahlen und Buchstaben, wodurch auch die Zahlen nicht mehr als Zahlen, sondern technisch als Buchstaben (Text) gelesen werden, was einen großen Unterschied ausmacht. Aus der Zahl 428 kann man zusammen mit weiteren Zahlen einen Mittelwert bilden, mit der Zahl 428 kann man rechnen, z. B. Verhältniszahlen (Indizes) bilden. Mit dem Text „ca. 428 mm“ hingegen kann man nicht rechnen und ihn auch in keine Mittelwertberechnung einfließen lassen, und auch das automatische Sortieren führt nicht zu einem geeigneten Ergebnis.

*Lösung:* Die nähere Bezeichnung des Wertes wie z. B. m, cm, mm gehört nicht in die Zelle, sondern sie ist Teil der Definition einer Spalte. Sie gilt für die gesamte Spalte, also alle Fälle, d. h. darf nicht innerhalb einer Spalte wechseln (z. B. mal mm, mal cm). Zahlen in einer Spalte sind die als zuverlässig erachteten Informationen. Will

man unbedingt auch circa-Werte erfassen oder Unsicherheiten dokumentieren, gehört dies unbedingt in eine separate Spalte, die man der eigentlichen Zahlenspalte voran- oder nachstellt, d. h. eine Spalte für die Zahlen und eine eigene, voran- oder nachgestellte Spalte, in der man „ca.“, „?“ , „mind.“ o.ä. dokumentiert. So ist die Sicherheit und Unsicherheit festgehalten und bei Bedarf auch technisch verfügbar, aber die Werte selbst sind als verarbeitbare Zahl abgelegt.

- **Fehlende Informationen:** In der Archäologie lassen sich nicht immer an allen Fällen alle Beobachtungen machen, insbesondere, weil Objekte bisweilen nicht vollständig erhalten sind. Daraus resultieren fehlende Werte, d. h. leere Zellen. Genau! Leere Zellen, und nicht etwa Einträge wie „?“ oder „-“ in diesen Zellen. Denn eine leere Zelle in einer Spalte mit Zahlen gilt als eine nicht vorhandene Zahl (alias fehlender Wert, *missing value*, NA), während ein dort eingetragenes Fragezeichen die gesamte Spalte wiederum in eine Spalte mit Text verwandelt.

- **Problem:** Oft liegen uns Informationen nicht als einzelne Zahlen, sondern als Spannen vor: Eine Datierung kann nicht als Mittelwert angegeben werden, sondern in zwei Jahreszahlen, von denen die eine die Untergrenze, die andere die Obergrenze der Spanne formuliert. Altersschätzungen an Skeletten resultieren oft in Angaben wie „20-40“ Jahre oder „40-60“ Jahre. Trägt man solche Spannen wie hier geschrieben in eine Spalte ein, sind es wegen des „bis“-Zeichens technisch wieder Texte, die nur mit zusätzlichem Aufwand auswertbar sind.

*Lösung:* Spannen als Zahlen eingeben und stets auf zwei Spalten aufteilen im Sinne eines „von“ und „bis“. Aus zwei solcher Spalten lässt sich die Differenz rechnen, also die Spannweite, die mal größer und mal kleiner ausfallen kann, und vor allem auch der mittlere Wert wie z. B. 30 aus „20 bis 40“, sprich: mit Zahlen lässt sich rechnen und gut auswerten.

- Wo die Wahl besteht, sollten Sie die Informationen in Zahlen statt in Text erfassen, weil das Arbeit spart, weil sich mit Zahlen leichter rechnen lässt und das Setzen von Filtern der Art größer als / kleiner als möglich ist. In der Anthropologie beispielsweise sind Altersbestimmungen von Skeletten auch als „ad.“ für adult (20-40 Jahre) oder „mat.“ resp. „matur“ für 40-60 Jahre üblich, samt Zwischenlösungen wie „ad./mat.“ für: nicht näher bestimmbar, ob adult oder matur. Man kann das so als Text in die Datentabelle eintippen und die Daten später in eine brauchbare Form überführen, man kann aber auch gleich die passenden Zahlen einsetzen.

- Denken Sie daran, dass Datenerfassung mühsam sein kann, sparen Sie im eigenen Interesse Tastatur-Anschläge! Tragen wir eine Messstrecke als 42,8 cm in die Spalte ein, bedeutet das für Sie vier Anschläge für „42 Komma 8“. Notiert man diese Strecke durchweg in Millimetern, sind es mit 428 drei Anschläge, eine Ersparnis von 25 % Tipparbeit. Das mag kleinlich klingen, aber die Masse macht's. Wenn man beispielsweise die Maße eines dreigliedrigen Gefäßes erfasst mit Randhöhe und -durchmesser, Halshöhe und -durchmesser, Bauchhöhe und -durchmesser sowie Bodendurchmesser, sind dies pro Gefäß sieben Werte, also entweder 28 Anschläge (4 pro Wert) oder 21 Anschläge (3 pro Wert), was bei vielleicht 500 Gefäßen dann doch einen Unterschied macht. Nicht vergessen: mit Zahlen kann man rechnen; im Nachhinein aus Millimetern neu Zentimeter zu machen, ist eine simple Division durch zehn, und fertig.
- In der Archäologie arbeitet man häufig mit Kategorien, wie z. B. Abfallgrube, Pfostengrube, Längsgrube, usw. Dies im Klartext bei der Datenerfassung einzutragen, bedeutet viele Anschläge. Oft spart man erheblich Arbeit, wenn man stattdessen mit Codelisten arbeitet, also z. B. 1 für Pfostengrube, 2 für Längsgrube, 3 für Abfallgrube, usw., und nur die jeweilige Zahl eintippt. Das mag am Beginn einer Datenerfassung etwas fehleranfällig sein und häufiger mal einen Blick in die Codeliste erfordern, spart auf Dauer aber erheblich an Erfassungszeit.
- Die praktische Erfassung von Daten folgt in der Regel auch einem mechanischen Muster am Objekt oder am Buch bzw. Bild. Bei der Vermessung von Gefäßen kann es z. B. effizienter sein, erst alle Höhen auszumessen und dann alle Durchmesser. Ordnen Sie die Spalten ihrer Datentabelle in der Reihenfolge an, in der Sie die Messungen vornehmen. Später die Spalten so umzuordnen, wie es für das Kontrollieren und Auswerten besser geeignet ist, ist nur ein geringer Aufwand.
- Der Mensch plant und denkt, die Maschinen rechnet! Gerade mit Tabellenkalkulationen lässt sich hervorragend rechnen, das sollte man nutzen. Oft liegen z. B. die Abbildungen von Gefäßen in unterschiedlichen Maßstäben vor, wie z. B. 1:2, 1:3 oder 1:4. Fangen Sie nicht an, an den Zeichnungen die mm abzulesen und dann selbst zu rechnen. Tragen sie Ihre originale mm-Ablesung in die Tabelle ein, vermerken in einer zweiten Spalte den Maßstab, und lassen die Maschine später aus Maßstab und originalen Ablesungen die tatsächlichen Strecken errechnen. Das geht schneller und ist weniger fehleranfällig gegenüber dem individuellen Rechnen (Kopfrechnen, oder gar Taschenrechner) und dann Eintragen in die Datentabellen.

### 10.1.1 Praxistipp: Tabelle einrichten & Kopfzeile fixieren

Üblicherweise setzt man einen kurzen (!), treffenden Namen in die oberste Zeile jeder Datenspalte, die deren Inhalt angibt, also beispielsweise 9 Spalten mit dem Text „Name“, „Vorname“, „Straße“, „Nr.“, „PLZ“, „Ort“, „Alter“, „Körperhöhe“, „Gewicht“, usw. Mir gefällt es, wenn diese Kopfzeilen (als einzige) in Fettschrift gesetzt werden (markieren & „F“ für fett). Ich schätze es, wenn ich die Schrift und Daten in der Tabelle gut lesen kann. Dazu Klick in die linke obere Ecke der Tabelle (das kleine graue Feld links von A und oberhalb von 1), womit die ganze Tabelle markiert wird; dann eine geeignete Schrift herausuchen, z. B. Calibri 12 oder 14 (o.ä.). Die Namen in den Spaltenköpfen sind unterschiedlich lang, möglicherweise möchten sie einzelne Spalten breiter schieben oder schmaler schieben. Schneller geht's automatisch: => Format; => Spalten; => „optimale Breite“ passt jede Spalte an den Inhalt an, das ist für's Erste schneller als das manuelle Zurechtschieben.

Nun kann die Dateneingabe beginnen. Nachdem alle Zeilen einer Bildschirmseite – je nach Schriftgröße vielleicht etwa 40 Fälle – vollgeschrieben sind, wandert die Tabelle mit jeder neuen Zeile, die man unten anfügt, nach oben. Irgendwann sind die Kopfzeilen verschwunden, genauer: oben nicht mehr sichtbar. Unrund. Lösung: oberste Zeile fixieren (funktioniert ggf. auch mit der ersten Spalte): => Ansicht; => Zellen fixieren; => Erste Zeile fixieren. Nun bleibt die oberste Zeile stehen, und die Folgezeilen schieben sich (scheinbar) darunter durch. Ergebnis: Auch wenn die Tabelle sehr reich an Datenzeilen ist, bleibt stets die Kopfzeile sichtbar. Man kann den Effekt auf dem gleichen Weg auch wieder ausschalten. Manchmal ist es sinnvoll, das gleiche auch mit der ersten Spalte zu tun, z. B. wenn man dort die Fallnummer, Objekt-ID oder ähnliches eingetragen hat, die man stets sehen möchte, auch wenn die Tabelle sehr spaltenreich ist.

*PS für Freunde von MS-Excel:* Hier versteckt sich die Funktion des Fixierens hinter: => Ansicht; => Fenster fixieren; => Oberste Zeile fixieren.

### 10.1.2 Praxistipp: Datenformular statt Tabelle zu Dateneingabe nutzen

Die Dateneingabe in eine Tabelle, die sehr viele Spalten hat – vor allem, wenn es mehr sind, als man auf einer Bildschirmseite sehen kann – wird schnell unübersichtlich. Ebenso misslich sind Tabellen mit vielen leeren Zellen; hier fällt die optische Orientierung schwer und schnell landet eine Eingabe in der falschen Spalte. Lösung: Für die Eingaben das Datenformular nutzen. Das geht so: Alle auszufüllenden Spalten der Tabelle und circa 10 bis 20 Zeilen markieren, also ein Rechteck. Dann: => Daten; => Formular. Es öffnet sich ein neues Fenster „Datenformular“, in dem die Spalten untereinander angeordnet sind. Klicken Sie auf die Schaltfläche „Neu“ (rechts) und füllen

die Felder aus (mit einem TAB springt man jeweils in das Folgefeld). Am Ende klicken Sie auf „NEU“, und der soeben eingegebene Datensatz wird in die nächste freie Zeile der Tabelle eingespielt, währenddessen das „Datenformular“ geleert wurde und nun die Eingabe des nächsten Falles erwartet. Mit „nächster Datensatz“ resp. „vorheriger Datensatz“ kann man in dem Datenformular auch blättern, z. B. um Tippfehler zu korrigieren. Bei spaltenreichen Tabellen ist diese Art der Eingabe weitaus sicherer, d. h. weniger anfällig für Fehler als direkt über die Tabelle.

*PS für Freunde von MS-Excel:* Datenformulare sind auch mit Excel möglich, aber in der aktuellen Version (Anfang 2020) von Office 365 standardmäßig nicht mehr in den Menüleisten enthalten und daher ziemlich versteckt. Was tun? In der allerersten (farbigen, meist grünen) Zeile, wo u. a. das Symbol für Speichern angelegt ist, zu dem kleinen ‚Pfeil nach unten‘ ganz rechts gehen und einen Klick mit der rechten Maustaste draufsetzen. Es erscheint ein Pop-up-Fenster, wo man ziemlich unten „Weitere Befehle“ findet. Draufklicken, es öffnet sich ein neues Fenster. Wählen Sie in der Auswahlzeile in der Mitte oben statt voreingestellt „Häufig verwendete Befehle“ neu „alle Befehle“. Nach wenigen Sekunden (ja: warten) ändert sich das längliche Auswahlfeld darunter und zeigt genau das an: alle Befehle. Zu „Maske“ scrollen, und mit der Schaltfläche „hinzufügen“ in das rechte Feld bewegen, dann mit „ok“ (unten rechts) schließen. Nun findet sich in der Ausgangstabelle, und zwar wieder in der (grünen) Zeile ganz oben ein neues Icon, mit der man das Datenformular ein- und ausschalten kann. Ansonsten funktioniert es ähnlich wie bei LO-Calc.

Erwähnte ich schon, dass das Gesuchte in unterschiedlichen Excel-Versionen an unterschiedlichen Stellen versteckt ist und auch unterschiedlich heißt? Suchen Sie ggf. nach „Formular“, oder „Datenformular“. Nachdem Sie also beim Einrichten ggf. schier verzweifelt sind (oder auch alles sofort „wie von alleine“ gefunden haben), lässt sich mit dem Formular gut arbeiten – mit einer nicht unwichtigen Einschränkung: sehr viele Spalten akzeptiert das Formular nicht. M.W. sind bei MS-Excel maximal 32 Spalten resp. Datenfelder möglich, d. h. den eigentlichen hier erwünschten Zweck – Hilfe bei sehr spaltenreichen Tabellen – erfüllt es nicht. LO-Calc hat diese Begrenzung nicht. Wer also diese Art der Dateneingabe als sehr nützlich empfindet und benötigt, sollte ggf. erwägen, die Datenerfassung mit LO-Calc durchzuführen, das ja \*.xlsx-Dateien lesen und schreiben kann.

### **10.1.3 Frage: ...oder stattdessen mehrere Erfassungstabellen anlegen?**

Im Falle von absehbar sehr spaltenreichen Tabellen sollte auch eine pragmatische Alternative zum beschriebenen Weg der Datenformulare bedacht werden, sprich: man

sollte sich diese Frage gleich zu Anfang einer ggf. komplexen Datenerfassung stellen: Müssen die Daten wirklich in einer einzigen Tabelle erfasst werden? Alternative: Man legt pro Thema eine Tabelle mit einer übersichtlichen Menge an Spalten an, beginnt jeden Fall – wie eigentlich sowieso – mit einer eindeutigen Kennung (ID) des Falles, und fügt weitere Teiltabellen später mit Softwareunterstützung anhand dieser ID zusammen. So könnte man Gräber beispielsweise erfassen mit einer Tabelle betreffs Anthropologie, einer weiteren zu Grabbau und Bestattungsweise, einer weiteren zum Beigabenspektrum usw., und diese Informationen später anhand der Grabnummer miteinander verknüpfen.

Nach der Datenerfassung in solchen Teiltabellen kann man dann für die weitere Auswertung per Copy & Paste oder z. B. mit den Werkzeugen von **R** & R-Commander aus diesen Teiltabellen eine gemeinsame große Datentabelle machen, an der die weitere Auswertung stattfindet. Hier für den R-Commander nur in Stichworten: beide (oder drei, vier, ...) Datentabellen in den R-Commander einlesen; dann: => Datenmanagement; => Datenmatrizen zusammenfügen; dort die beiden zusammenzufügenden Datensätze auswählen, einen passenden Namen für den neuen, gemeinsamen Datensatz vergeben; je nach vorliegendem Fall die passende Art der Zusammenführung wählen, in unserem Beispiel „füge Spalten zusammen“. Am Ende wie immer: Ergebnis unbedingt stichprobenartig kontrollieren, d. h. vor jeder weiteren Arbeit die neue Datenmatrix anschauen und prüfen.

#### **10.1.4 Praxistipp: zwei Tabellen verbinden**

Eine solche Datenerfassung in Teiltabellen kann weitaus sicherer unternommen werden, wenn man die Teiltabellen anhand einer Identifikationsspalte wie z.B. einer Individualnummer oder Grabnummer fest miteinander verbindet.

##### **10.1.4.1 Der Weg mit LO-Calc**

- Tabelle mit erstem Datenblatt anlegen, z.B. „Anthro“ mit der Grabkennung / ID in der ersten Spalte und dann den anthropologischen Daten.
- Weiteres Datenblatt in dieser Tabelle anlegen, z.B. „Frau“ für die Liste der weiblichen Beigaben (später Datenblatt Mann; neutral, usw.). Dort, d. h. in der Tabelle „Frau“ die erste Spalte mit der ersten Spalte des vorangehenden Datenblattes verbinden. Was am schnellsten und sichersten „händisch“ erfolgt, indem man z. B. in die Zelle A2 folgendes eingibt: =Anthro.A2. Danach erscheint in der zweiten Zeile der ersten Spalte von „Frau“ der Wert, der dort auch im Datenblatt „Anthro“ steht.
- Anschließend diese Verknüpfung per Copy & Paste auf die weiteren Zeilen des zweiten Datenblatts ausdehnen. Mit der Maus in Zelle „Frau“ A2 gehen, copy, weitere

Zellen darunter in der Spalte A markieren und „paste“ eingeben: Alle (markierten) Zeilen sind jetzt mit dem ersten Datenblatt „Anthro“ fest verbunden. Ändert man dort einen Wert oder fügt neue hinzu, geschieht dies auch im zweiten Datenblatt „Frau“. Auf diese Weise verhindert man, dass Zeilenfehler entstehen, d. h. in einer der Datenblätter eine Zeile fehlt, womit das spätere Zusammenfügen etwas schwieriger werden könnte.

#### **10.1.4.2 Der Weg in MS-Excel**

- Tabelle mit erstem Datenblatt anlegen, z.B. „Anthro“ mit der Grabkennung / ID in der ersten Spalte und dann den anthropologischen Daten.
- Weiteres Datenblatt in dieser Tabelle anlegen, z.B. „Frau“ für die Liste der weiblichen Beigaben (später Datenblatt Mann; neutral, usw.). Dort die erste Spalte mit der ersten Spalte des vorangehenden Datenblattes verbinden. Weil das im Point-&-Click-Verfahren mit MS-Excel sehr viele Klicks erfordert, lieber „händisch“:
  - Klick in die erste Datenzelle der ersten Spalte, vermutlich die Zelle A2; Gleichheitszeichen eingeben und fortfahren: =’Anthro’!A2 & Enter. Die Zelle (z.B.) Frau A2 ist jetzt automatisch verbunden mit der Zelle „Anthro“A2.
  - Die gerade angelegte Zelle des „abhängigen“ Datenblatts markieren, kopieren, und in den folgenden Zellen (z.B. A3 - A25) einfügen. Nun sind alle Zellen A2 – A25 beider Datenblätter miteinander verbunden. Alle Eingaben und Änderungen im ersten Datenblatt („Anthro“) werden automatisch, spätestens beim Speichern auch im Datenblatt (z.B.) „Frau“ übernommen.
- Ergebnis: Man hat am Ende zwei oder mehrere Datenblätter, die sich leichter ausfüllen lassen als eine sehr breite Tabelle, aber fest (und damit frei von Zeilenfehlern) über eine gemeinsame ID-Spalte miteinander verbunden sind.

#### **10.1.5 Hinweis: Epi Info**

Wenn eine umfangreiche Datenerfassung mit vielen Variablen ansteht, ggf. auch mit mehreren Erfassern, sodass die zuvor geschilderten Formulare (statt der Erfassung in einer Tabelle) etwas unkomfortabel erscheinen: Epi Info anschauen! Das Programm wird vom US-amerikanischen „Centers for Disease Control and Prevention (CDC)“ als Freeware herausgegeben und weiterentwickelt, im Jahr 2020 war die Version 7.2.3 aktuell: <https://www.cdc.gov/epiinfo/pc.html> [24.1.2020]. Es zielt auf medizinische, genauer: epidemiologische Studien ab, außerhalb dieser Sphäre ist es wenig bekannt. Epi Info „spricht“ außer Englisch auch Französisch und Spanisch und ist in einem ausführlichen Handbuch gut dokumentiert. Epi Info bietet die Möglichkeit, eigene Datenerfassungsformulare aufzubauen, und zwar relativ komfortabel und einfach. Ich

werde diesen Aspekt hier nicht detaillierter beschreiben, möchte aber Mut machen: Nach relativ kurzer Einarbeitungszeit unter Zuhilfenahme von Anleitungen auf YouTube war ein taugliches Erfassungsformular entworfen. Für Besucherbefragungen beispielsweise oder die systematische Erfassung von Silices und das Management von Surveydaten m. E. ein Werkzeug, dessen Einsatz man erwägen sollte. Die Daten werden im MS-Access-Format abgelegt – ein weit verbreiteter Standard, sodass einer Weiterverarbeitung in anderen Programmen nichts entgegensteht. Der Vorteil solcher Lösungen gegenüber einer Tabellenkalkulation besteht darin, dass sich in solchen Formularen Fehlerprüfungen einbauen lassen, z. B. erlaubte und unerlaubte Wertebereiche, was Datenfehler bereits bei der Erfassung minimiert.

Sofern eine netzbasierte arbeitsteilige Datenerfassung im Zentrum des Interesses resp. Projektes steht, empfehle ich zusätzlich zu Epi Info auch einen Blick auf „shiny“ und „lab.js“ (googeln). Möglicherweise bieten diese sehr modernen Ansätze das gesuchte Instrument. Wenn es um eine mobile Datenerfassung per Smartphone oder Tablet geht, führen die Suchworte Open Data Kit (ODK) sowie GeoODK oder auch KoboToolBox zu interessanten Lösungen – mit denen der Autor allerdings keine eigenen praktischen Erfahrungen hat.

## 11 „SKALENNIVEAU“: ÜBER UNTERSCHIEDLICHE ARTEN VON INFORMATIONEN

---

Das zuvor Beschriebene hat nicht nur etwas mit Pragmatik und Effizienz zu tun, sondern auch einen tieferreichenden statistischen Hintergrund, der sich mit dem Stichwort „Skalenniveau“ verbindet. Es ist nützlich, sich schon früh mit diesem Aspekt zu beschäftigen, bevor man in die erste richtige Datenerfassung einsteigt. Es geht um die Tatsache, dass es unterschiedliche Arten von Informationen gibt. Man unterscheidet Nominalskala, Ordinalskala und quantitative Skala. Je höher das Skalenniveau, desto mehr statistische Analyseverfahren stehen bereit und desto leichter sind automatisierte Fehlerprüfungen. Jedes Skalenniveau beinhaltet spezifische Möglichkeiten der Abstraktion, der grafischen Darstellung und der statistischen Aussagen.

## 11.1 NOMINALSKALA

Beispiele: Namen, E-Mail-Adressen, Studienfächer, Farben, Währungen, Prägeorte von Münzen.

Was geht? Häufigkeiten auszählen. Statistik: u. a. Prozentwerte, Vergleich von Prozentwerten, Mehrfeldertafeln, Chi-Quadratstest.

Mögliche Rechenoperationen: zählen, Summen bilden. Unterschiedliche Kategorien können eventuell zu abstrakteren Gruppen zusammengefasst werden, z. B. Schwerter, Lanzen, Schilde etc. zu Waffen.

## 11.2 ORDINALSKALA (RANGSKALA)

Geordnete Reihe. Die Reihenfolge ist definiert, die Abstände sind unklar.

Beispiele: Ränge beim Sport wie Gold-, Silber-, Bronze-Medaille; Noten oder Beliebtheitskalen wie z. B. bei Lehr-Evaluationen; Farben, wenn man die einzelnen Kategorien ordnet, z. B. im Sinne eines Farbenkreises; Stratigrafien und relative Chronologien in der Archäologie (Schicht 1, Schicht 2, ..., oder Phase 1, Phase 2, Phase 3, ...) bilden eine Ordinalskala.

Was geht? Alles, was bei Nominalskalen möglich ist. Zusätzlich aber auch: einen mittleren Rang berechnen (während in der Nominalskala z. B. ein „mittleres Studienfach“ keinen Sinn macht), Gruppen auf Unterschiede des mittleren Rangs testen.

## 11.3 QUANTITATIVE SKALA

Messwerte. Hier unterscheidet man zwischen Intervallskala und Verhältnisskala:

- Intervallskala: Die Abstände und die Differenz zwischen den Werten sind definiert, die Skala hat keinen echten Nullpunkt.

Beispiele: Temperatur Celsius; schwimmende Dendro-Kurven; Häufigkeiten aus zyklischen Beobachtungen, z. B. sogenannte Jahresarbeitskalender (nach dem 12. Monat folgt nicht der 13. Monat, sondern der 1. Monat).

Was geht? Alles, was bei Nominal- und Ordinalskalen möglich ist. Zusätzlich jedoch können Summen und Differenzen gebildet werden. Statistik: Man kann Mittelwert und Standardabweichung berechnen, man kann Mittelwerte vergleichen und auf Mittelwertunterschiede testen.

- Verhältnisskala: Intervallskala mit „echtem“ Nullpunkt.

Beispiele: Semesterzahl, Körpergröße, Gewicht, Bargeldsumme, Prägejahr, Münzwert.

Was geht? Alles, was bei Nominalskalen, Ordinalskalen und Intervallskalen möglich ist. Zusätzlich jedoch auch Multiplikation und Division, sowie die Bildung von Verhältnissen (Indexzahlen).

Weiterhin unterscheidet man bei Daten in einer quantitativen Skala zwischen stetiger Skala und diskreter Skala:

- Diskrete Skala: Der Zahlenstrahl weist Lücken auf, zwischen denen keine Werte existieren.

Beispiel: Semesterzahl, Nominale von Geldmünzen, Prägejahr von Münzen.

- Stetige Skala: Der Zahlenstrahl ist mit Werten beliebig dicht besetzt.

Beispiele: Zentimeter, Zeit; Körpergewicht; Körpergröße.

Je höher das Skalenniveau ist, desto mehr Auswertungsmöglichkeiten gibt es. Folglich wird man bemüht sein, bei Beobachtungen jeweils ein möglichst hohes Skalenniveau anzustreben. Das ist nicht immer, aber manchmal eben doch möglich. Wie? Das lässt sich am Beispiel von Farben gut darlegen. Farben werden im Alltag nach einfachen Kategorien benannt, wie z. B. rot, blau, grün, oliv, aubergine usw. Dies ergibt Informationen auf dem Niveau einer Nominalskala. Nutzen wir einen der üblichen Farbkreise und ordnen die Farben entsprechend an, gewinnen wir eine Ordinalskala. Bestimmt man die Farben z. B. nach dem „Munsell Book of Colors“ (resp. den „Munsell Soil Color Charts“), gewinnt man sogar eine diskrete Intervallskala.

#### 11.4 EXKURS FARBBESTIMMUNGEN

Die Überlegungen zum Skalenniveau und zur Fehlerprüfung sind bei Farbbestimmungen ein wichtiger Grund für die Anwendung der scheinbar komplizierten „Munsell Soil Color Charts“. Beim leider üblichen Vorgehen von Farbangaben nach der Nominalskala (rot, blau, grün, usw.) führt eine Fehlbestimmung (seltener: eine Meinungsverschiedenheit zwischen unterschiedlichen Bearbeitern) schlicht zu einer kaum systematisch auffindbaren Fehlbestimmung, die diese spezielle Beobachtung in eine gänzlich andere, falsche Klasse führt. Codiert man die Farben nach den „Munsell Color Charts“ oder ähnlichen Systemen, erhält man eine diskrete Intervallskala. Eine Fehlbestimmung – z. B. durch unglückliches Licht – oder eine Meinungsverschiedenheit zwischen unterschiedlichen Bearbeitern führt meist in ein eng benachbartes Feld, d. h. sie wird bei späteren zusammenfassenden Statistiken die Gesamtbilanzen nicht wesentlich verfälschen.

Literatur zur Farbbestimmung: Munsell, Albert Henry (1951). *Atlas of the Munsell Color System*. Malden; Munsell, Albert Henry (1929). *Book of color. Defining, explaining, and illustrating the fundamental characteristics of color* (rev. and ext. of „The atlas of the Munsell color system“, by A. H. Munsell; Baltimore 1929). Weit verbreitete Auszüge sind: Munsell Soil Color Charts (Baltimore 1975, rev.ed. 1992). Munsell Rock Color Charts (1991, 1995). – Man findet für diese Werke sehr unterschiedliche Preisangaben, auch sehr unterschiedliche Auszüge, sie kosten oft um 100-200 €. – Das Munsell-System ist u. a. erklärt bei: Shepard, Anne O. (1956). *Ceramics for the Archaeologists*. Washington: Carnegie Institution of Washington; Spennemann, Dirk R. (1986). Munsell-Farbtafel: Eine Notwendigkeit für Archäologen? *Acta Praehistorica et Archaeologia*, 18, 177-187.

Alternativ: Michel-Farbenführer. Farbentafeln für Briefmarkensammler (31. Aufl. München). Zum Thema Farbbestimmung allgemein: Völz, Hans G. (1990). *Industrielle Farbprüfung. Grundlagen und Methoden. Farbmetrische Testverfahren für Farbmittel in Medien* (Weinheim 1990). – Itten, Johannes (1961). *Kunst der Farbe. Subjektives Erleben und objektives Erkennen als Wege zur Kunst* (1. Aufl. 1961, hier: Ravensburg 1987).

## 12 EINEN ERSTEN DATENSATZ ERFASSEN

---

Lassen Sie uns in diesem Sinne einen ersten Datensatz erfassen, der uns auch im Folgenden zum Weiterarbeiten dienen kann: die Münzen in Ihrer Geldbörse. Später wollen wir auswerten können, welche Nominalen (Münzwerte) vorhanden waren, welche Währungen, Prägestätten und Prägejahre. Wollen Sie dazu mit Hilfe von LO-Calc eine Tabelle anlegen? Es wäre eine gute Übung.

Wir legen also eine Tabelle in LO-Calc und geben ein: den Münzwert (Nominal), das Prägejahr, die Währung, das Land und den Prägeort (sofern angegeben). Dazu ein paar Tipps und Hinweise:

- Die oberste Zeile der Tabelle benennt die Spalten (kurz); man setzt sie in Fettschrift und fixiert sie wie oben beschrieben.
- Die Angaben auf den Münzen, gerade bei den 1-, 2- und 5-Cent-Münzen, sind oft sehr klein. Ziehen Sie eine Lupe hinzu oder nutzen Sie die Kamera eines Smartphones und die Möglichkeit, dort ein Bild zu vergrößern. Für das Erkennen und Zuweisen der Münzbilder bei den Euromünzen, die auf der Vorderseite (Avers) gleich sind, sich auf der Rückseite (Revers) nach Ländern unterscheiden, hilft z. B. der Wikipedia-Artikel

„Euromünzen“, wo sich Bilder finden und auch eine Liste der Staaten und deren Prägeorte. Für die Codierung der Staaten benutze man systematisch die relativ bekannten Kürzel der KFZ-Kennzeichen (Wikipedia: „Liste der Kfz-Nationalitätszeichen“).

- Wir erinnern uns an das oben unter „übliche Anfängerfehler“ Geschriebene und geben den Wert der Münze („Nominal“) stets in Cent an, tragen also für eine 2-Euro-Münze die Zahl 200 ein. Das hält die Datenspalte in sich konsistent und spart – im Vergleich zu anderen Optionen wie z. B. 1 Cent = 0,01 Euro – Tipparbeit.

- Um spätere Mehrarbeit zu vermeiden und Fehler von Anbeginn an zu reduzieren: Geben Sie die Daten zu einer bis fünf Münzen ein, und formatieren Sie nun die Spalten dann so, wie es der Art der Informationen und dem Skalenniveau entspricht. In unserem Fall also:

- Bei den Spalten Nominal und Prägejahr: Spalte markieren, => Format, => Zellen. Dort beim Reiter Zahlen: „Dezimalzahl“, „0“ Nachkommastellen, „0“ führende Nullen & OK; Reiter „Ausrichtung“, Horizontal „rechts“ & OK.

- Bei den Spalten Währung, Land und Prägeort: Spalte(n) markieren, => Format, => Zellen. Dort Reiter „Zahlen“, „Text“ & OK; Reiter „Ausrichtung“ „zentriert“ & OK.

- Wenn Sie sich angewöhnen, in solchen Tabellen Zahlen stets rechtsbündig zu schreiben, Text auf Nominalskalenniveau zentriert und anderen Text wie üblich linksbündig, und dies in der Tabellenkalkulation frühzeitig zu vereinbaren (d. h. gezielt so einzustellen), vermeiden Sie spätere Fehler und gewinnen an Übersicht. Übersicht in Tabellen unterstützt die optische Fehlererkennung.

- Zusätzlich legen wir noch eine Spalte „Name“ an und tragen dort den Namen des Geldbörseninhabers ein und eine Spalte Datum, in der das Erhebungsdatum steht (im Format TT.MM.JJJJ). Man braucht dies nur einmal zu schreiben und kann die folgenden Zeilen später per Copy-and-Paste füllen.

Ihre Tabelle sollte nun ähnlich aussehen wie in **Abb. 5**. Obwohl wir mit LO-Calc arbeiten, empfehle ich ein Abspeichern der Daten als \*.xlsx-file (MS-Excel-Format), weil dies ein weit verbreitetes Austauschformat ist. Eine wichtige Alternative ist das \*.csv-Dateiformat („comma separated values“), doch dabei gehen die meisten der soeben vereinbarten Formatierungen verloren.

	A	B	C	D	E	F	G	H	
	Nominal	Prägejahr	Währung	Land	Prägeort	Name	Erhebungsjahr	Erhebungsort	
1		1	2011	EU	D	A	FS	2019	Münster
2		1	2012	EU	D	F	FS	2019	Münster
3		5	2002	EU	D	F	FS	2019	Münster
4		200	2008	EU	B		FS	2019	Münster
5		10	2012	CHF	CH		FS	2019	Bern
6		10	2012	CHF	CH		FS	2019	Bern
7		50	1986	CHF	CH		FS	2019	Bern
8		100	2010	CHF	CH		FS	2019	Bern
9		200	2012	CHF	CH		FS	2019	Bern
10		200	1995	CHF	CH		FS	2019	Bern
11		200	1987	CHF	CH		FS	2019	Bern
12		500	1974	CHF	CH		FS	2019	Bern

**Abb. 5** Beispiel für die Erfassung der Münzen aus zwei Geldbörsen des Autors im Jahr 2019.

Ich persönlich empfinde die in **Abb. 5** ersichtliche Tabellenform als übersichtlich und gut bearbeitbar. Alternativ kann man auch wie oben beschrieben mit einem Eingabeformular arbeiten.

### 12.1.1 Rechnen mit einer Tabellenkalkulation

Tabellenkalkulationen eignen sich sehr gut dazu, mit ihnen zu rechnen. Lassen Sie uns – ein etwas künstliches Beispiel – ein Interesse am Wert der Münzen in einer fremden Währung wie z. B. dem Schweizer Franken (CHF) haben. Im Sommer Herbst 2019 war 1 CHF ca. 0,91 Euro wert. Setzen wir dies um:

=> Tabelle „Geldbörsen“ (am Beispiel **Abb. 5**); Klick in den Kopf einer noch leeren Spalte, z. B. „I“; wir schreiben dort „Wert“ hinein, setzen den Titel fett und zentrieren ihn; klicken in die sechste Zeile, d. h. die Zelle I6, und schreiben dort: =I6\*0,91 [ENTER]. Wir lesen: 9,1. Dabei steht „I6“ für den Inhalt der Zelle I6, also „10 Cent“ (resp. Rappen). Es fällt auf: Unsere Nominale sind in Cent gedacht, also ist auch der CHF um das Hundertfache geringer anzusetzen, also Cent. Daher korrigieren wir die Formel in: =(I2\*0,91)/100 [Enter], woraufhin mit 0,091 Cent der richtige Betrag erscheint.

Jetzt klicken wir in das (gerade errechnete) Feld I6, rufen mit der rechten Maustaste ein „Kopieren“ (*copy*) auf, markieren die darunterliegenden übrigen, noch leeren Felder der Spalte, und rufen mit der rechten Maustaste ein „Einfügen“ (*paste*) auf. Die Felder werden mit Zahlen gefüllt.

Klicken Sie in ein beliebiges Feld der Spalte, z. B. I8, und sehen Sie sich in der Kopfleiste an, was in der Zelle steht:  $= (A8 * 0,91) / 100$ . Es wurde also automatisch beim Copy-&-Paste aus der Eingabe A6 für die sechste Zeile richtigerweise ein „A8“ für den Bezug in der 8. Zeile gemacht – ein nützlicher, viel Arbeit sparender Automatismus.

Ein Automatismus, den man bei Bedarf allerdings unterbinden kann. Ein vorangestelltes Dollarzeichen „\$“ fixiert die Adresse der Zelle, was in unserem Beispiel wie folgt ausgesehen hätte:  $= (\$A\$6 * 0,91) / 100$

Indes, in unserem Fall ist das Richtige geschehen. Nun schnell noch für Schönheit und Übersicht sorgen, d. h. die Zellen angemessen formatieren:

=> Spalte I markieren; => Format; => Zellen; im aufpoppenden Menü einstellen beim Reiter „Zahlen“ auf Dezimalzahlen, 2 Kommastellen, 0 führende Nullen, und beim Reiter „Ausrichtung“ ggf. auf „rechts“ setzen, & OK. Nun zeigt Spalte I den Wert der einzelnen Münzen aus der Schweiz in Cent des Euros an.

Lassen Sie uns noch schnell die Summe bilden und dabei eine weitere Fähigkeit solcher Tabellenkalkulationen kennenlernen. Wir klicken in die leere Zelle unterhalb unserer Zahlenkolonne und danach auf das Summenzeichen „Σ“ in der Kopfzeile, dann [ENTER]. In der Zelle steht nun die Summe der Zahlenkolonne.

Das System ist weitaus flexibler! Setzen Sie bitte in der Kopfzeile einen Doppelklick nicht auf das Summenzeichen, sondern auf das Zeichen „fx“ links daneben. Der nun aufgeklappte Funktionsassistent zeigt diverse andere Funktionen auf, wie z. B. Abrunden, Abs, ... Klicken wir auf „Anzahl“, erscheint oben rechts die Erklärung dieser Funktion: „zählt, wie viele Zahlen eine Liste von Argumenten enthält“. Klickt man anschließend auf „Weiter“, verändert sich das Menü. Gehen Sie mit dem Mauszeiger in die Datentabelle und markieren alle Zellen in der Spalte I: Im Menü wird die erste Zeile nun automatisch entsprechend Ihrer Markierung ausgefüllt, z. B. mit I2:I12. Heißt: Alle Zellen von I2 bis I12. Rechts unten im Menü sehen wir zugleich, welche Formel wir gerade bestellen: „=ANZAHL(I2:I12)“. Mit einem „OK“ wird das Resultat in die Zelle der Datentabelle geschrieben, die man mit dem Mauszeiger ausgesucht hat.

Auf diese Weise stehen alle hinter „fx“ steckenden Funktionen zu Verfügung, u. a. auch diverse statistische Maßzahlen wie z. B. der Mittelwert usw. – doch das soll erst später Thema werden. Hier ging es nur darum, das Rechnen und den Umgang mit Funktionen einmal angesehen zu haben. Spielen und experimentieren Sie ein bisschen mit den Möglichkeiten.

## 12.2 EINFACHE DATENVERWALTUNG UND ERSTE DATENSICHTUNG MIT LO-CALC

Mit einer Tabellenkalkulation können die erfassten Daten verwaltet und auch statistisch analysiert werden, letzteres viel weitgehender als vielen Anwendern klar ist. Wir wollen diesen Weg jedoch nicht vertiefen, sondern die Tabellenkalkulation „nur“ zur Erfassung, Verwaltung und ggf. ersten Aufbereitung von Daten benutzen, um die Daten danach an spezielle Auswertungsprogramme wie z. B. PAST oder R übergeben zu können. Für die Umsetzung dieser reduzierten Ansprüche benötigen wir nur drei Instrumente, die hier kurz besprochen werden sollen: Sortieren, Filtern und gelegentlich auch „Pivot-Tabellen“.

### 12.2.1 Sortieren

Das Sortieren einer Tabelle nach der ersten Erfassung der Daten kann rein ästhetische Gründe haben. Es kann aber auch helfen, Eingabefehler zu erkennen und zu bereinigen. Technisch ist es leicht durchzuführen: Tabelle markieren; => Daten; => Sortieren, und in dem aufpoppenden Menü beim Reiter „Sortierkriterien“ die erwünschten Sortierungen aussuchen, bei unserem Münzbeispiel z. B. nach 1. Prägejahr und 2. Nominal. Nach „OK“ wird die bestellte Sortierung ausgeführt. Der Reiter „Optionen“ bietet Möglichkeiten zu manchmal nötigen Feineinstellungen.

Wieso hilft das betreffs Eingabefehlern? Weil Eingabefehler oft Kommafehler sind oder mit zu viel resp. zu wenig Ziffern einhergehen. In unserem Beispiel mit den Münzen erwarten wir Prägejahre irgendwo zwischen 1900 und 2100, aber nicht z. B. 1015 (Typo vorne, 1 statt 2), 215 (statt 2015, eine Ziffer fehlt) oder 20017 (eine Ziffer zu viel). Beim Sortieren geraten diese Fälle ganz an den Anfang resp. an das Ende der Tabelle und können so schnell entdeckt (und bereinigt) werden. Ähnliches bei der Eingabe von Text, z. B. dem Land der Münzen in Buchstaben. Bei Sortieren fällt auch auf, was schwer sichtbar ist: versehentlich ein Leerzeichen vor oder nach dem (eigentlich richtigen) Buchstaben. Kurz: Den Datensatz sukzessive einmal nach jeder Spalte Sortieren und Sichten ist ein erster Schritt, Datenfehler zu identifizieren. Wenig Aufwand, viel Effekt.

### 12.2.2 Filtern

Daten filtern, d. h. gezielt eine Untermenge aus einem Datensatz heraussuchen, ist eine Standardaufgabe, die für fast alle weiteren Analysen ständig und mit wechselnden Zielen benötigt wird. LO-Calc bietet dazu zwei Modi an, „Autofilter“ und „Weitere Filter“. Schauen wir uns zunächst die „Autofilter“ an: => Daten; => Autofilter. Danach sollten an der Kopfzeile jeder Spalte neu Kästchen mit Pfeilen eingespiegelt sein

(Abb. 6). Klickt man auf einen dieser Pfeilkästchen, erscheint eine Schaltfläche mit den Daten dieser Spalte, genauer: mit allen unterschiedlichen Ausprägungen einer Variablen, die in dieser Spalte vorkommen. In meinem Beispieldatensatz treten unter „Stichort“ die Kästchen Basel, Berlin, Duesseldorf, usw. auf, jeweils mit einem Häkchen. Nehme ich das Häkchen von „Basel“ weg, werden die übrigen Fälle weggefiltert: Nach „OK“ zeigt die Tabelle nur noch die anderen Münzen, ohne die in Basel beobachteten Geldbörsen. Die Tatsache, dass ein Filter gesetzt ist (die anderen Fälle sind nicht weg, sondern nur unsichtbar!) wird durch die Blaufärbung des Pfeilkästchens angezeigt (in Abb. 6 bei „Ort“).

Nach kurzem Nachdenken und Spielen mit diesem Werkzeug wird deutlich: Damit lässt sich (sehr) viel erreichen. Denn die Filter sind kombinierbar, es ist leicht, z. B. alle 1-Euro-Münzen aus Deutschland mit dem Prägejahr 2003 herauszufischen – und zu zählen, wie viele es davon in dem Datensatz gibt. Mit einem Klick auf das blaue Pfeilkästchen und dann die Schaltfläche „Alle“ (unten links) wird ein eingestellter Filter wieder aufgehoben.

	A	B	C	D	E	F	G	H	I
	Nominal	Jahr	Waehrung	Land	Ort	ID	Stichjahr	Stichort	Alter
572	1	1950	DM	D	5	1036	1993	Aufsteigend sortieren	43
573	1	1991	DM	D	3	1036	1993	Absteigend sortieren	2
574	2	1975	DM	D	5	1036	1993	Top 10	18
575	2	1991	DM	D	1	1036	1993	Nicht leer	2
576	10	1970	DM	D	3	1036	1993	Standardfilter...	23
577	10	1981	DM	D	2	1036	1993	Elemente suchen...	12
578	1	1970	DM	D	5	1037	1993	Alle	23
579	1	1983	DM	D	4	1037	1993	Berlin	10
580	1	1985	DM	D	2	1037	1993	Duesseldorf	8
581	2	1966	DM	D	2	1037	1993	Frankfurt	27
582	2	1976	DM	D	4	1037	1993	Goettingen	17
583	2	1983	DM	D	2	1037	1993	Hannover	10
584	5	1950	DM	D	4	1037	1993	Heidelberg	43
585	5	1981	DM	D	4	1037	1993	Jena	12
586	5	1985	DM	D	3	1037	1993	Koeln	8
587	5	1987	DM	D	5	1037	1993	Goettingen	6
588	5	1989	DM	D	3	1037	1993	Goettingen	4
589	5	1989	DM	D	3	1037	1993	Goettingen	4

Abb. 6 Münzen in der LO-Calc-Tabelle mit aktiviertem Autofilter bei „Ort“ (Prägeort) und aufgeklappt bei „Stichort“ (Ort der Erhebung).

Nun probieren wir die Option „Weitere Filter“ aus: => Daten; => Weitere Filter; => Standardfilter. Das neue Menü bietet an, unter „Feldname“ eine Spalte auszuwählen, unter „Bedingung“ verschiedene logische Operatoren und unter „Wert“ kann man einen Wert eingeben. Ein Ausfüllen mit dem Feldnamen „Waehrung“, Bedingungen „=“ und Wert „EUR“ und ein abschließendes „OK“ erreicht das gleiche wie der Autofilter zuvor: Wir fischen alle Euro-Münzen heraus. Wir sehen aber an dem Menü, dass unter „Bedingungen“ erheblich mehr Filtermöglichkeiten angeboten, wie

z. B. kleiner als, größer als, oder auch „enthält nicht“. Es ist also leicht, z. B. auch alle Nicht-Euros herauszufischen.

Auch dieser Filter erlaubt Kombinationen, wie das Beispiel in **Abb. 7** zeigt. Die zweite Zeile des Menüs beginnt mit „Operator“, wo man ein logisches „und“ resp. ein „oder“ aussuchen kann, und dann die Eingangsbedingung mit einer zweiten Bedingung verknüpfen kann. So werden beispielsweise Aufgaben der Art lösbar: Alle Euro-Münzen (1. Zeile), deren Nominal größer als 100 ist („und“ plus Bedingung in der 2. Zeile) und die vor dem Jahr 2015 erhoben wurden (3. Zeile), herausfiltern.

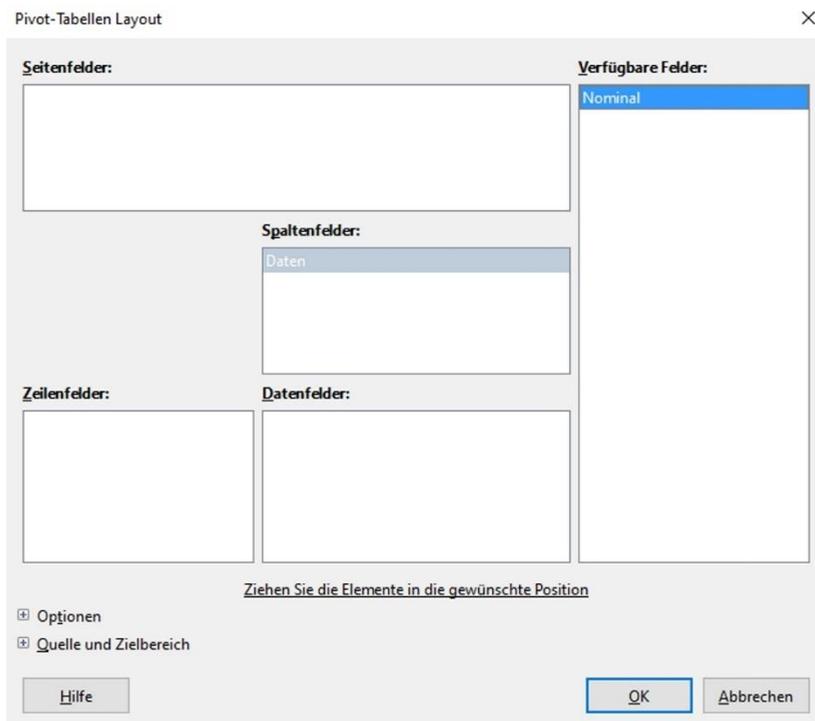
	A	B	C	D	E	F	G	H	I	J
1	Nominal	Jahr	Währung	Land	Ort	ID	Stichjahr	Stichort	Alter	
2642	200	2002	EUR	D		1	1169	2002	Berlin	0
2643	200	2002	EUR	D		1	1169	2002	Berlin	0
2644	200	2002	EUR	D						
2645	200	2002	EUR	D						
2665	200	2002	EUR	D	#					
2666	200	2002	EUR	D	#					
2667	200	2002	EUR	D	#					
2668	200	2002	EUR	F	#					
2704	200	2002	EUR	D	#					
2705	200	2002	EUR	D	#					
2706	200	2002	EUR	D	#					
2707	200	2002	EUR	D	#					
2717	200	2002	EUR	D	#					
2756	200	2002	EUR	D	#					
2757	200	2002	EUR	D	#					
2793	200	2002	EUR	F	#					
2798	200	2002	EUR	D	#					
2815	200	2002	EUR	D	#NULL!		1179	2002	Berlin	0
2816	200	2002	EUR	D	#NULL!		1179	2002	Berlin	0
2826	200	2002	EUR	D	#NULL!		1180	2002	Berlin	0
2833	200	2002	EUR	D	#NULL!		1181	2002	Berlin	0

**Abb. 7** Tabelle der Münzen mit gesetztem Filter („Weitere Filter“, „Standardfilter“): Währung gleich Euro & Nominal größer 100 (d. h. alle 200 Euro-Münzen) & Erhebungsjahr kleiner (d. h. vor) 2015.

Sortieren und Filtern sind zusammengenommen mächtige Werkzeuge, der Umgang damit sollte geübt und Routine werden. Ein Beispiel? Spätere Themen vorausgreifend, rufen wir schnell ein paar Statistiken ab: Man markiere die Spalte Prägejahr, dann: => Daten; => Statistiken (ganz unten); => Beschreibende Statistik; dann ein Klick in ein freies Feld unterhalb oder rechts der Datentabelle mit dem Mauszeiger klicken, und in der Menüfläche wieder „OK“ setzen. Rechts in die Datentabelle wurden nun eingefügt diverse Statistiken für das Prägejahr, z. B. der Mittelwert (hier: 1996,8). Diese Möglichkeit kann man nun mit dem Sortieren und Filtern kombinieren: Zum Beispiel alle Münzen aus dem Land D herausfiltern, die entsprechenden Daten in eine neue Tabelle

copy-pasten und für diese Untermenge die beschreibenden Statistiken errechnen lassen. [Innerhalb der Originaltabelle ist das nicht möglich. Daher: markieren inkl. der Spaltenköpfe, kopieren, über „Datei; neu; Tabellendokument“ eine neue Tabelle öffnen und per Paste die Teil-Tabelle einfügen.]

Die Möglichkeiten dieser „Statistiken“ wollen wir hier nicht weiterverfolgen, weil dazu im Folgenden andere, mächtigere Programme genutzt werden sollen. Da PAST u. a. jedoch sehr eingeschränkte Möglichkeiten der Datenverwaltung aufweist, soll das Beispiel zeigen, wie eine gute Zusammenarbeit zwischen Tabellenkalkulation und z. B. PAST aussehen könnte: Datenerfassung und -verwaltung insbesondere durch Sortieren und Filtern mit der Tabellenkalkulation, dann Übergabe der Daten resp. einer gezielt ausgesuchten Untermenge an das Statistikprogramm.



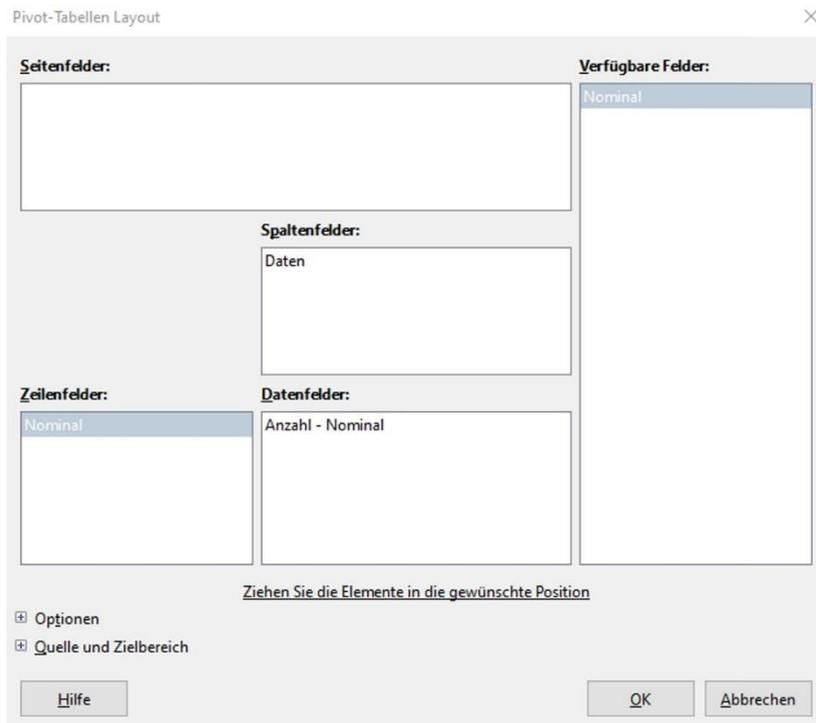
**Abb. 8** Fenster zum Anlegen einer Pivot-Tabelle.

### 12.2.3 Pivot-Tabellen und erste Grafiken

Statt langer Erklärungen, was Pivot-Tabellen sind, machen wir schnell eine: Eine Datenspalte markieren, z. B. „Nominal“, dann => Daten; => Pivot-Tabelle; => Einfügen oder Bearbeiten (in älteren Versionen: „Erstellen“), in dem aufpoppenden Menü „aktuelle Auswahl“ mit OK bestätigen. Neu erscheint ein zunächst kryptisch wirkendes neues Fenster (**Abb. 8**).

Folgende Schritte sind zu tun:

- aus dem Fenster „Verfügbare Felder“ die Option „Nominal“ in das Fenster „Zeilenfelder“ (links) ziehen;



**Abb. 9** Das Fenster zum Anlegen einer Pivot-Tabelle nach den nötigen Eingaben, unmittelbar vor „OK“.

- im Fenster „Zeilenfelder“ nun auf „Nominal“ doppelklicken; die im neu aufklappenden Fenster angebotene Auswahl (oben) von „Keine“ auf „Benutzerdefiniert“ schalten; aus der nun statt grau schwarz gewordenen Auswahl darunter „Anzahl“ wählen und „OK“ klicken; nun (etwas verwirrend) mit der Maus (linke Maustaste) auf das „Nominal“ in „Zeilenfelder“ gehen und dieses nach rechts in die „Datenfelder“ ziehen, wo nun ein blau hervorgehobenes „Anzahl – Nominal“ auftaucht (**Abb. 9**).

Mit „OK“ bestätigen. Es poppt eine neue Tabelle auf (**Abb. 10**), die wir studieren: Die Tabelle enthält zwei Spalten: „Nominal“ und „Anzahl – Nominal“. Die erste Spalte weist alle vorkommenden Werte (Ausprägungen der Variable) der Variable „Nominal“ auf, die zweite Spalte weist nach, wie oft dieser Wert in der Ausgangstabelle vorkommt.

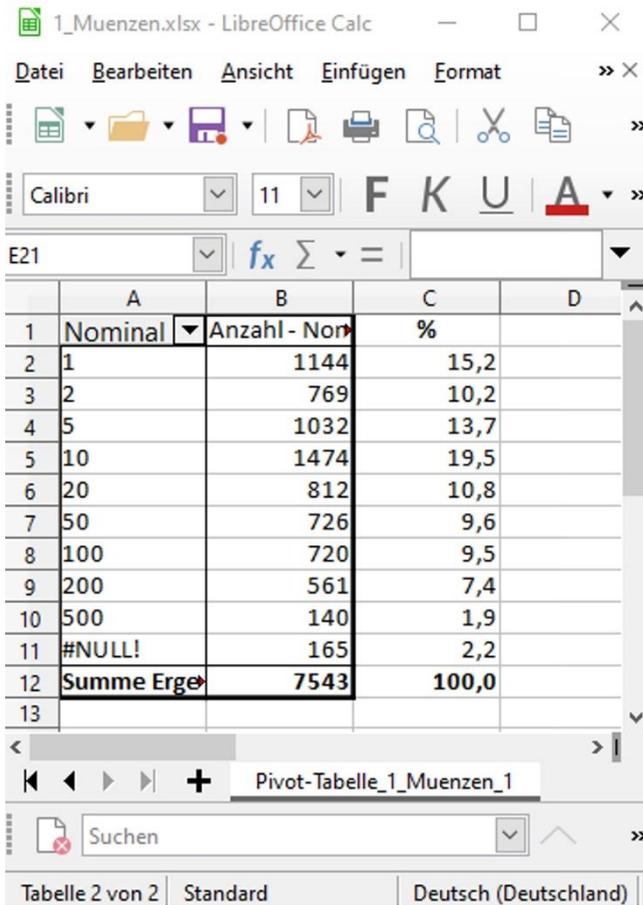
	A	B	C	D
1	Nominal	Anzahl - Nominal		
2	1	1144		
3	2	769		
4	5	1032		
5	10	1474		
6	20	812		
7	50	726		
8	100	720		
9	200	561		
10	500	140		
11	#NULL!	165		
12	Summe Erge	7543		

**Abb. 10** Die aus den ursprünglichen Daten erzeugte Pivot-Tabelle.

Aha: Pivot-Tabellen sind Zusammenfassungen (Aggregationen) von Daten aus einer Ausgangstabelle heraus. Im vorliegenden Fall haben wir die Häufigkeit der Münznominale auszählen lassen. Mit den so gewonnenen Zahlen kann man weiterarbeiten.

Beispiel: Wir berechnen daraus die Prozentzahlen der beobachteten Häufigkeiten.

- Klick ins Feld C2, dort schreiben:  $= (B2/B\$12)*100$  [„B12“ ist das Feld mit der Gesamtsumme];
- Klick;
- ins Feld C2 gehen; Copy; Feld C3 bis C 12 markieren; Paste;
- nun sollten alle Prozentzahlen sichtbar sein. Beim Operieren in der beschriebenen Weise mit Copy-&-Paste sorgt das Programm dafür, dass die Adressen der Felder automatisch angepasst werden, d. h. aus „B2“ ist in den Folgezeilen „B3“, „B4“ usw. geworden. Das Feld „B12“ als Divisor ist hingegen stabil geblieben, weil wir die „12“ mit dem führenden Dollarzeichen als beizubehalten markiert haben.
- Etwas mehr Schönheit erreichen Sie, wenn Sie die neue Spalte noch angemessen formatieren: In die Kopfzeile „%“ hineinschreiben, zentrieren, dann die übrigen Zellen markieren, dann => Format; => Zellen, und einstellen auf „Dezimalzahl“, 1 Nachkommastelle, null „Führende Nullen“ und „OK“ (Abb. 11).



**Abb. 11** Die um Prozentzahlen erweiterte Pivot-Tabelle.

Ja, dieses spezielle Ziel Prozentzahlen zu berechnen, kann man anderweitig auch schneller erreichen, aber mit dem hier dargelegten Weg verfolgen Sie das Ziel auf eine flexible Weise, mit der ggf. auch andere Ziele erreichbar sind.

Nun noch schnell eine Grafik aus der Pivot-Tabelle ableiten. Zunächst noch einmal mit der Maus irgendwo in die Pivot-Tabelle klicken, um diese aktiv zu schalten. Dann öffnet ein Klick auf das kleine Säulendiagramm (oder Kuchendiagramm) in der Schaltleiste oben (ganz oben, mehr rechts) das Fenster „Diagramm-Assistent“ – und eigentlich taucht jetzt schon ein Säulendiagramm mit der Häufigkeit unserer Nominalen

auf. Das Diagramm kann man nach einem Anklicken verschieben, z. B. neben die Tabelle setzen. Innerhalb des Diagramms lassen sich die einzelnen Elemente nun durch Anklicken (Doppelklick) verändern. Ich würde z. B. die Zahlen links (y-Achse) in Fettschrift setzen, die Zahlen unten (x-Achse) in Fettschrift setzen und z. B. auf Schriftgröße 12 bringen, und die blaue Farbe in den Säulen (reinklicken!) durch ein Grau ersetzen (Reiter „Fläche“) sowie den Säulen einen Linienrahmen geben (Reiter „Linie“) (Abb. 12).

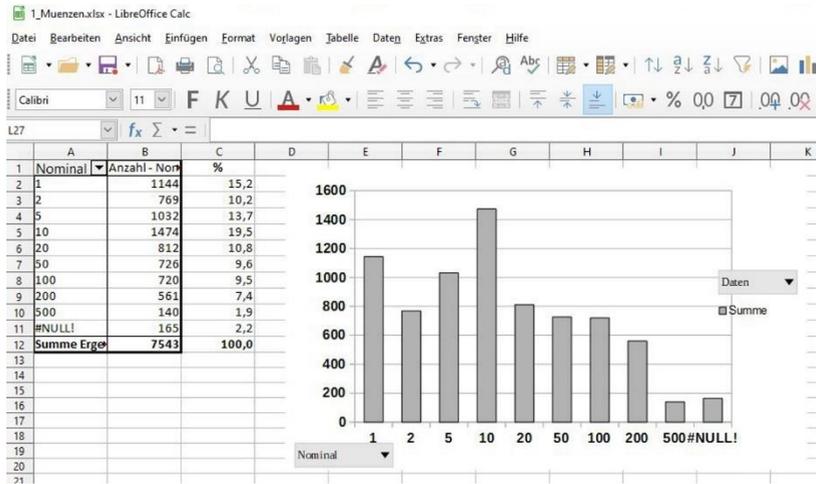


Abb. 12 Ein aus der Pivot-Tabelle abgeleitetes und grafisch angepasstes Balkendiagramm.

Wiederholen Sie das Ganze doch ab Mausclick in die Pivot-Tabelle und bestellen Sie statt des Säulendiagramms ein „Kreisdiagramm“ und verändern es, bis es Ihren ästhetischen Anforderungen entspricht. Auch diesen Aspekt und das Thema grafische Darstellungen werden wir noch vertiefen – an dieser Stelle geht es nur um erste Eindrücke und erste Erfolgserlebnisse bei der Auswertung der eingegebenen Daten.

Richtig, die unterste Datenzeile #NULL! in der Tabelle (Abb. 11) bedarf eines kurzen Nachdenkens. Es handelt sich um fehlende Werte, d. h. in insgesamt 165 Fällen wurden zwar diverse andere Informationen erfasst, nicht aber der Wert der Münze. Nicht ungewöhnlich, in der Archäologie haben wir es sehr oft mit Fällen zu tun, wo – meist wegen mangelnder Erhaltung – nicht alle Merkmale erfasst werden können. Für die Prozentzahlen und das Balkendiagramm haben wir diese Fälle hier mitberücksichtigt, d. h. eingeschlossen. Alternativ wäre jedoch zu erwägen, ob man die Hundertsumme

nicht ohne diese Fälle bildet und sie auch aus dem Balkendiagramm entfernt. Je nach Fragestellung kann das eine richtiger sein als das andere, weshalb es hier nur darum geht, auf dieses oft wiederkehrende Thema „fehlende Werte“ aufmerksam zu machen.

Pivot-Tabellen sind ein wichtiges und mächtiges Instrument. Wie gesehen, ist ihr Erzeugen jedoch mit ein paar mehr Klicks und Schritten verbunden. Daher empfehle ich: mehrmals mit unterschiedlichen Tabellen und Aufgaben üben, um Routine zu gewinnen.

#### 12.2.4 Zweidimensionale Pivot-Tabelle

Lassen Sie uns die erste Pivot-Tabelle noch etwas ausbauen: Es sollen zwei Variablen gemeinsam ausgezählt werden, z. B. Nominal und Währung. Start (fast) wie zuvor, denn erst einmal markieren wir alle Spalten in der Ausgangs- /Daten-Tabelle, z. B. durch einen Klick in das kleine Fenster links oben (über Zeile 1, links von Spalte A); dann wie zuvor: => Daten; => Pivot-Tabelle; => Einfügen oder Bearbeiten, in dem aufpoppenden Menü „aktuelle Auswahl“ mit OK bestätigen. In dem neu aufklappenden Menü im Fenster rechts unter „verfügbare Felder“ aus der Liste die Variable „Nominal“ ins Fenster „Zeilenfelder“ ziehen, dort auf das blau hinterlegte Nominal doppelklicken, auf „Benutzerdefiniert“ umschalten und „Anzahl“ auswählen, mit OK bestätigen. Anschließend das blau hinterlegte Zeilenfeld „Nominal“ in das Fenster „Datenfelder“ ziehen.

Nun aus dem Fenster „Verfügbare Felder“ die Variable „Währung“ in das Fenster „Spaltenfelder“ ziehen (**Abb. 13**), dann mit OK bestätigen. Es öffnet sich eine neue Tabelle (Pivot-Tabelle), die die Häufigkeiten der Nominale für jede der drei Währungen und an den Rändern die Zeilensummen und die Spaltensummen ausweist. Die erwünschte neue Tabelle poppt auf (**Abb. 14**), eine Kreuztabelle (auch: „Mehrfeldertafel“), die in den Zellen die Häufigkeiten der Nominale zur jeweiligen Währung ausweist.

**Seitenfelder:**

**Spaltenfelder:**

**Zeilenfelder:**

**Datenfelder:**

**Verfügbare Felder:**

Nominal  
Jahr  
Währung  
Land  
Ort  
ID  
Stichjahr  
Stichort  
Alter

Ziehen Sie die Elemente in die gewünschte Position

☑ Optionen  
☑ Quelle und Zielbereich

Hilfe OK Abbrechen

**Abb. 13** Menü zum Zusammenstellen der Pivot-Tabelle (Mehrfeldertafel).

	A	B	C	D	E	
1	Anzahl - Nominal	Daten				
2	Nominal	CHF	DM	EUR		Summe Ergebnis
3	1	9	470	665		1144
4	2	2	216	551		769
5	5	315	239	478		1032
6	10	355	697	422		1474
7	20	406		406		812
8	50	299	154	273		726
9	100	191	182	347		720
10	200	147	123	291		561
11	500	68	72			140
12	#NULL!	69	65	31		165
13	Summe Ergebnis	1861	2218	3464		7543
14						

**Abb. 14** Die resultierende Mehrfeldertafel mit den Nominalen als Zeilen und den Währungen als Spalten.

### 12.2.5 Pivot-Tabelle für eine Variable mit quantitativer Skala

Bei unserem Münzbeispiel bieten die Nominale und die Prägejahre eine quantitative Skala, jeweils mit echtem Nullpunkt. Allerdings sind die Nominale recht speziell. Zwar darf man mit ihnen rechnen, Mittelwerte bilden etc., aber die Skala von 1 bis 200 (resp. 500) Cent ist nicht dicht mit (möglichen) Werten besetzt, sondern es gibt nur die neun Ausprägungen 1, 2, 5, 10, 20, 50, 100, 200 Cent (Euro) und 500 (CHF: Rappen). Hier kann man, um eine Übersicht zu gewinnen, mit Hilfe einer Pivot-Tabelle deren Häufigkeit auszählen lassen und z. B. daraus auch ein Balkendiagramm ableiten. Die Prägejahre der Münzen hingegen bilden eine stetige Skala, jede Jahreszahl kann vorkommen. Das ergibt für eine sinnvolle Häufigkeitstabelle und auch ein Balkendiagramm zu viele mögliche Werte. In solchen Fällen ist es üblich, die Zahlen zu Klassen zusammenzufassen, die jeweils mehrere der Ausprägungen umfassen. Als Übungsbeispiel wählen wir eine recht starke Zusammenfassung, zu Jahrzehnten. Wie geht das? Mit Rechnen: Das Prägejahr durch 10 teilen, vom Ergebnis den Wert hinter der Null abschneiden, und das Ganze wieder mit 10 multiplizieren. Resultat: Jahrzehnte statt Jahre. Vielleicht vollziehen Sie diese Anleitung an einem beliebigen Beispiel zunächst einmal mit Papier und Bleistift nach. Mit der Tabellenkalkulation lässt sich das relativ schnell umsetzen. In unserem Beispiel stehen die Prägejahre in Spalte B der Ausgangstabelle. In einer freien Spalte schreiben wir:  $=(\text{Ganzzahl}(B2/10))*10$ . Mit Copy-&-Paste schreiben wir die ganze Spalte mit dieser Formel voll, und prüfen hie und da, ob auch richtig gerechnet wurde. Danach via Pivot-Tabelle die Häufigkeiten der Jahrzehnte auszählen lassen und ein Säulendiagramm bilden, so, wie wir es bereits kennengelernt haben.

Mehr müssen wir derzeit nicht erreichen; ein paar grundlegende, mächtige Funktionen der Tabellenkalkulation, die man für das weitere Arbeiten braucht, sind erlernt.

### 12.2.6 PS: Pivot-Tabellen bei MS-Excel

Die Erstellung von Pivot-Tabellen mit MS-Excel verläuft ähnlich und sei hier nur kurz skizziert. Man gehe in die Ausgangstabelle und markiere den Bereich, der in die Pivottable aufgenommen werden soll, meist die ganze Tabelle. Über die oberste Menüleiste kommt man mit „Einfügen“; => „Tabelle“; => „Pivottable“ zur relevanten Schaltfläche. Ich empfehle als Standard, das Anlegen einer neuen Tabelle zu wählen. Dann schiebt man die gewünschte Zeilenvariable (Feld rechts oben) nach unten in das Feld Zeilen (im Falle des Geldbörsen-Beispiels z.B. „Währung“) und die zweite Variable (z. B. Nominal) in das Feld Spalten. Sodann die relevante Variable – hier z. B.

Nominal – in das Feld „ $\Sigma$  Werte“, wonach dort automatisch „Summe von Variable“ (hier: Nominal) auftaucht und die Pivottabelle angelegt wird. Mit Klicken auf „Summe von Nominal“ kann man in einem neu aufklappenden Fenster „Wertfeldeinstellungen“ auswählen und so die Funktion des Zusammenziehens von Werten beeinflussen, in unserem Beispiel z.B. statt „Summe“ (Voreinstellung) neu „Anzahl“ wählen. Dann wird jede Zelle mit der Anzahl der Nominalen in der jeweiligen Währung belegt.

Es lohnt auch hier, durch Üben etwas Routine aufzubauen, denn in der Kombination mit den angebotenen Filterfunktionen ist das Erstellen von Pivottabellen ein mächtiges Werkzeug, mit dem viele Auswertungsfragen schnell beantwortet werden können.

### 12.3 DATENFEHLER UND ERSTE FEHLERPRÜFUNG

Bei jeder Erhebung von Informationen kommt es zu Fehlern. Der erste Schritt jeder Datenanalyse dient im Grunde nicht der Auswertung, sondern der Prüfung auf Fehler und möglichst der Identifikation falscher Datensätze. Gut identifizierte Fehler können durch eine Nacherhebung korrigiert werden. Ist das nicht möglich, wird man solche Fehler als sehr unwahrscheinliche Beobachtungen aus der weiteren Untersuchung ausschließen, entweder, indem man den ganzen Fall löscht, oder indem man diesen speziellen Wert löscht, d. h. ihn künftig als fehlende Beobachtung behandelt. Höhere Skalenniveaus führen zu Fehlern, die leichter automatisch identifizierbar sind, resp. die nur kleine Auswirkungen auf das Ergebnis haben. Ein kleiner Fehler bei quantitativen Skalen, z. B. eine in einer Datenserie einmal aus einem Ablesefehler oder Tippfehler entstandene Zahl „9“ statt „8“, fällt in eine nah benachbarte Klasse. Solch ein Fehler wird bei Statistiken das Gesamtergebnis wie z. B. den Mittelwert nur wenig verändern. Anders ist dies bei Variablen auf Nominalskalenniveau, z. B. „Prägeort“, wo der Fehler im Ergebnis den Fall in eine gültige andere Klasse führt und nicht auffällig ist. Bei quantitativen Skalen fallen große Fehler, vor allem solche des Typs Kommafehler oder Spaltenfehler, meist rasch auf; ein Nominal z. B. von „2000 Cent“ (statt 200 Cent) würde unmittelbar zu einer Überprüfung führen.

Oft stehen Informationen nicht isoliert, sondern in einem Zusammenhang miteinander. Solche Zusammenhänge können überprüft werden. So machen z. B. einige Währungen die Prägeorte auf ihren Münzen erkennbar, andere nicht. Entsprechende Prüfungen können Aufnahmefehler aufdecken. Für die Fehlerprüfung ist es sinnvoll, sich solche Abhängigkeiten innerhalb des Datensatzes systematisch bewusst zu machen und dieses Wissen zu nutzen, um nach einer ersten einfachen Fehlerbereinigung eine zweite Runde einzulegen, die auf solchen Abhängigkeiten beruht.

Die zuvor erlernten Techniken Sortieren, Filtern und die Pivot-Tabellen sowie die daraus abgeleiteten einfachen Grafiken können helfen, nach der Datenerfassung die nötigen Fehlerprüfungen vorzunehmen. Für richtige Projekte sollte zuvor geplant werden, wann und wie oft Datenprüfungen und Fehlerbereinigungen vernünftigerweise erfolgen. Erfasst man beispielsweise Daten aus der Literatur, die ja länger verfügbar ist, kann man die Datenprüfung einmal am Ende der primären Datenerfassung vornehmen – was in der Regel der effizientere Weg ist. Erfasst man hingegen Daten an Funden, die z. B. auf einer Museumsreise nur wenige Tage zur Verfügung stehen, sollten Datenprüfungen bereits erfolgen, so lange das Material noch relativ bequem verfügbar ist, also ggf. sogar täglich am Ende jedes Arbeitstages. Für tägliche Routineprüfungen kann der hier beschriebene Weg auf die Dauer etwas mühsam werden, weshalb wir auf das Thema – allerdings weitaus später – nochmals zurückkommen werden.

Größere Datensätze mit mehreren oder gar vielen Variablen werden auch nach diesen ersten Bereinigungen einzelne Fehler enthalten, auf die man erst später stößt. Bisweilen geht es ja auch nicht um falsche Messungen oder Tippfehler; manchmal sind Beobachtungen auch mit Entscheidungen verbunden, die man nach tieferer Kenntnis des Materials und der Daten in Einzelfällen mit erfahreneren Augen anders treffen möchte. Im Grunde sind große Datensätze erst am Ende einer Auswertung wirklich fehlerfrei.

### **12.3.1 Fehler bei Daten in einer Nominalskala**

Die Frage nach einem Namen oder einem Studienfach führt im Falle von falschen Angaben oder irrigen Notizen zu anderen Häufigkeiten. Der Fehler fällt nicht auf, er ist nur mit Aufwand systematisch suchbar. Daher ist eine EDV-gestützte Fehlersuche bei Nominaldaten schwierig. Umfasst der Datensatz relativ viele Fälle, ist eine Auszählung nach Häufigkeit immer sinnvoll. Beim Beispiel der Geldbörsen sollten die Prägeorte der deutschen Münzen jeweils mehrfach vorkommen; ein in einer großen Stichprobe nur einmal vorkommender Prägeort dürfte auf einen Erfassungsfehler zurückgehen. Daher zielt die Fehlersuche bei Nominaldaten zunächst auf singuläre oder seltene Beobachtungen, da sie oft auf Tippfehler zurückgehen.

Liegen für jeden Fall mehr als eine Variable vor, kann man anschließend versuchen, nach unwahrscheinlichen Kombinationen von Ausprägungen der Variablen zu suchen. Im Falle der Münzen beispielsweise gibt es in einigen Staaten der Euro-Welt die Angabe von Prägestätten, in anderen nicht, und manche Prägestätten eben

nur in einem Staat; fragt man nach diesen Zusammenhängen, können einzelne zunächst nicht auffällige Fehler über die nicht mögliche Kombination von zwei Merkmalen entdeckt werden.

### 12.3.2 Fehler bei Daten in einer Verhältnisskala

Bei der Frage z. B. nach Semesterzahl, Bargeldsumme o. ä. fallen große Fehler, die z. B. typischerweise auf eine falsche Kommasetzung zurückgehen, als erheblich falsche Angaben auf bzw. erscheinen als „verdächtig“. Ein Teilnehmer z. B. mit der Semesterzahl „80“ statt „8“ ist möglich, aber doch auffällig und folglich ein Fall, den man nach Identifikation auf Richtigkeit prüfen würde. Berechnet man aus den Münzen die Bargeldsumme, sind Werte zwischen Null und vielleicht 20 Euro zu erwarten; einige Hundert Euro Bargeld in Münzen wären ein unerwartetes, auffälliges Ergebnis, und daher bei einer Fehlerprüfung gut zu identifizieren. Im Grunde lohnt es, bei allen quantitativen Daten zunächst einmal nach den kleinsten und den höchsten Werten zu schauen und zu prüfen, ob diese plausibel sind. Das zielgerichtete Sortieren der Datentabellen führt schnell zum Ziel, auch das Anfertigen einfacher Histogramme ist hilfreich.

## 12.4 VERTIEFTE FEHLERPRÜFUNG

Die zuvor skizzierten Prüfroutinen sind sinnvoll, sie helfen, offensichtliche Mängel zu bereinigen. Tieferliegende Fehler werden in der Praxis oft erst bei der Auswertung bemerkt, wenn irgendwelche Zwischenergebnisse seltsam erscheinen und sich bei näherer Prüfung nicht als besonderer Einzelfall herausstellen, sondern als schlichter Fehler. Erfahrene Auswerter wissen, dass oft erst ein bis zu Ende ausgewerteter Datensatz fehlerfrei ist. Doch über das zuvor Geschilderte hinaus ist eine vertiefte Fehlerprüfung möglich und sinnvoll, bevor man in die Auswertung eintritt. Denn man kann sich Abhängigkeiten zwischen den Variablen zunutze machen. Bei der Erfassung von Silices beispielsweise erwartet man für die Grundformen Abschlag und Klinge jeweils ein spezifisches Verhältnis zwischen Länge und Breite resp. das Einhalten gewisser Grenzen. Messer, Lanzenspitzen oder Oberschenkelknochen sollten stets länger als breit sein, usw. Es bestehen Grenzen wahrscheinlicher Werte und es bestehen Abhängigkeiten zwischen den einzelnen Variablen eines Falles, die man systematisch überprüfen kann. Bei umfangreichen Datenerhebungen sollte am Ende der ersten Fehlerprüfung eine weitere Runde folgen, in der man den Datensatz systematisch auf mögliche Verletzungen der inneren Logik zwischen den Variablen prüft.

Klingt komplizierter, als es ist. Wer mit **R** arbeitet, würde sein Wissen über die Logik und die Zusammenhänge der Variablen formalisieren, in ein Skript umsetzen

und dieses über den Datensatz „laufen“ lassen, sodass Fehler relativ automatisiert ausgewiesen würden. Das Erarbeiten eines Skripts zur Fehlerprüfung empfehle ich vor allem dann, wenn diese Prüfung z. B. wegen einer mobilen Datenerfassung häufiger stattfinden muss, also etwa allabendlich oder wöchentlich. Der Aufwand lohnt, weil ein zusätzliches Zeitinvestment – noch vor der eigentlichen Auswertung – viel Zeit sparen kann, die man sonst ggf. in das „alles wieder neu rechnen“ steckt, wenn Datenfehler erst spät während der Auswertung erkannt werden.

## 12.5 BEOBACHTERFEHLER – „OBSERVER ERROR“

Das zuvor Beschriebene zielt auf die Bereinigung und Fehlerprüfung der eingegebenen Daten, im Grunde auf das Ausmerzen von Tippfehlern und Ähnlichem. Unter dem Thema „Beobachterfehler“ geht es nun um den systematischen Blick auf das Problem nach Bereinigung solcher primären Mängel. Wie steht es um die Messungen selbst und deren Zuverlässigkeit? Dieses Thema – der Beobachterfehler (*observer error*) – ist in der Archäologie kaum bekannt, spielt hingegen in den Naturwissenschaften eine wichtige Rolle, nicht zuletzt in der Anthropologie. Wird mit Maschinen gemessen wie z. B. mit Waagen das Gewicht ermittelt, sind deren Messfehler in der Regel vom Hersteller ermittelt worden und dokumentiert; die Angabe des verwendeten Geräts ist also hinreichend, um gerätebedingte Fehlerbandbreiten wissenschaftlich zu dokumentieren und Dritten mitzuteilen. Was aber, wenn menschliche Beobachter die Daten erheben, z. B. die Metrik von Töpfen erfassen oder die Länge von Knochen ermitteln? Hier ist die Frage zu beantworten, wie verlässlich – sprich: reproduzierbar – die Daten sind. Dabei unterscheidet man – je nach vorliegendem Fall – zwischen „*intra-observer error*“ und „*inter-observer error*“. Der Intra-Observer Error prüft, ob die Messungen eines Beobachters reproduzierbar sind; er stellt den Regelfall dar. Welche Messwerte kommen zustande, wenn der Bearbeiter seine eigenen Messungen ein paar Tage später wiederholt? Bei umfangreichen oder auch langwierigen Projekten sind bisweilen mehrere Beobachter an den Datenerhebungen beteiligt, die Datenerfassung erfolgt arbeitsteilig. Hier greift die wichtige Frage nach dem „Inter-Observer Error“, also ob Messungen am gleichen Objekt durch verschiedene Beobachter zum gleichen Ergebnis führen.

Das Vorgehen zur Beantwortung dieser Frage ist in beiden Fällen grundsätzlich ähnlich: Es wird eine zufällige Stichprobe von Fällen ausgewählt, die mehrfach vermessen werden, um dann die Ergebnisse beider Messungen zu vergleichen. Klar, am besten wäre es, relativ viele Fälle doppelt zu vermessen, aber der betroffene Bearbeiter wird stets bemüht sein, die Zahl der Doppelerhebungen klein zu halten. Oft

werden nur 10 oder 20 Fälle doppelt vermessen. Wichtig ist es, die beiden Serien zufällig auszuwählen. Bei einer langwierigen Keramikaufnahme, beispielsweise, könnte man am Ende der Materialerhebung ein paar Fälle hervorholen, die man nach der ersten Einarbeitung – sagen wir nach 10 % der Materialerfassung – schon einmal aufgenommen hat. Wichtig ist, dass man die damaligen Messungen „vergessen“ hat, also nicht die Zahlen der ersten Erhebung quasi noch aus dem Gedächtnis heraus reproduziert, sondern wirklich neu misst. Für den kurzen Abschnitt oder auch die Fußnote in einer wissenschaftlichen Publikation, in dem man den Observer-Error einer Erhebung darlegt, sollte die Anzahl der Wiederholungsmessungen und die Art der Stichprobenziehung kurz beschrieben werden.

Für die konkreten Berechnungen folgen wir dem Weg, der sich seit Mitte der 2000er Jahre in der physischen Anthropologie beim Vermessen von Knochen etabliert hat und ermittelt den „(absolute) technical error of measurement“ TEM, den „relative technical error of measurement“ rTEM und den Reliabilitäts-Koeffizienten R.<sup>23</sup> Wir beschränken uns auf den üblichen Fall, dass zwei Messreihen verglichen werden sollen, d. h. der Intra-Observer Error bestimmt wird, oder der Inter-Observer Error zwischen zwei Beobachtern zu ermitteln ist. Die Formeln für den Inter-Observer Error bei mehr als zwei Beobachtern sind etwas komplexer.<sup>24</sup> Ich verzichte hier auf die Wiedergabe der Formeln und schildere den Rechenweg als Algorithmus. Das Ganze ist recht schlicht und kann bereits mit einer Tabellenkalkulation zuverlässig umgesetzt werden. Wir beginnen mit TEM:

- (a) Vergleich 1. Messung und 2. Messung; man berechne die Differenz beider Messungen;
- (b) Ergebnis quadrieren;
- (c) die Summe all dieser quadrierten Differenzen berechnen;
- (d) diese Summe teilen durch „Anzahl der Messpaare mal 2“;
- (e) aus dieser Zahl die Quadratwurzel ziehen;
- (f) das Ergebnis ist TEM.

Der Wert von TEM hat die gleiche Skala wie die Messdaten; wenn also mm oder kg erfasst wurden, ist auch TEM auf mm oder kg bezogen. Er kann als Standardabweichung der Wiederholungsmessungen aufgefasst werden.

---

<sup>23</sup> Nach Perini et al. 2005; Weinberg et al. 2005; Goto & Mascie-Taylor 2007. Eine später oft als zusammenfassende Referenz zitierte Anwendung ist Gapert et al. 2009.

<sup>24</sup> Goto & Mascie-Taylor 2007, 254.

Die Kennzahl  $r_{TEM}$  ist im Grunde eine Art Prozentwert zu TEM, indem der ermittelte Wert von TEM auf den Mittelwert der gemeinsamen Ergebnisse beider Messreihen bezogen wird:

- (a) Berechne das arithmetische Mittel (Mittelwert; *mean*) für beide Messreihen;
- (b) teile TEM durch diesen Mittelwert;
- (c) und multipliziere das Ergebnis mit 100;
- (d) das Ergebnis ist  $r_{TEM}$ .

Der  $r_{TEM}$  besagt im Grunde, wie groß im Sinne eines Prozentwertes der Messfehler ist.  $r_{TEM}$ -Werte bis etwa 5 (%) gelten in der anthropologischen Literatur als akzeptabel.

Als dritte Kenngröße wird der Reliabilitätskoeffizient  $R$  berechnet. Dazu ist die gemeinsame Standardabweichung beider Messreihen zu ermitteln, was ggf. eine Tabellenkalkulation übernimmt, resp. hier weiter unten beim Thema „Streuungsmaße: Standardabweichung“ erklärt und vorgerechnet wird.

- (a) TEM quadrieren ( $TEM^2$ );
- (b) die gemeinsame Standardabweichung beider Reihen quadrieren ( $SD^2$ );
- (c)  $TEM^2$  durch  $SD^2$  dividieren;
- (d) den berechneten Wert von der Zahl 1 abziehen;
- (e) das Ergebnis ist „ $R$ “.

$R$  kann Werte zwischen 0 (völlig unzuverlässig) und 1 (Messungen gänzlich identisch) annehmen. Je näher  $R$  dem Wert 1 kommt, desto zuverlässiger sind die Messungen. In der anthropologischen Literatur gelten  $R$ -Werte von größer 0,95 als akzeptabel, d. h. ein menschlicher Beobachterfehler von unter 5 %.

Wer seinen Beobachterfehler mit ca. 20 Wiederholungsmessungen in dieser Weise ermittelt und darlegt, weist nach, wie sorgfältig und verlässlich er arbeitet und wie tragfähig die seiner Untersuchung zugrundeliegenden Daten sind.

### Literatur zu TEM und rTEM

- Gapert, R., Black, S. & Last, J. (2009). Sex Determination from the Occipital Condyle: Discriminant Function Analysis in an Eighteenth and Nineteenth Century British Sample. *American Journal of Physical Anthropology*, 138, 384-394.
- Goto, R. & Mascie-Taylor, C. G. (2007). Precision of measurement as a component of human variation. *Journal of Physiological Anthropology*, 26(2), 253–256.
- Perini, T. A., de Oliveira, G. L., dos Santos Omellas, J. & de Oliveira, F. P. (2005). Technical error of measurement in anthropometry. *Revista Brasileira de Medicina do Esporte*, 11(1), 86–90.
- Weinberg, S. M., Scott, N. M., Neiswanger, K. & Marazita, M. L. (2005). Intraobserver error associated with measurements of the hand. *American Journal of Human Biology*, 17(3), 368–371.

## 12.6 PILOTSTUDIE

Erst mal alle Daten erfassen, dann auswerten: Das ist keine gute Idee! Richtig ist, die Datenerfassung sauber zu planen und dann erste Daten zu erfassen. Geht es z. B. um 500 Töpfe, erfasst man erst einmal circa 20 oder 50 Gefäße. Daran lernt man, was man nicht bedacht hat, und justiert die zuvor erdachte Systematik nach. Diese Stichprobe wertet man dann nach den Fragen aus, die man an das Material hat, um zu testen, ob die Daten auch im erwünschten Sinne auswertbar sind. Dann merkt man z. B. schnell, dass sich – wie oben als üblicher Anfängerfehler beschrieben – aus „428“, „ca. 422“ und „mind. 432“ kein Mittelwert errechnen lässt. Sprich: Man kann an einer kleinen Stichprobe, bevor man möglicherweise sehr viele Zahlen ungeschickt erfasst hat, das eigene Vorgehen prüfen und optimieren. So etwas nennt man eine Pilotstudie. Man sollte sie unbedingt durchführen, bevor man eine sehr umfangreiche Datenerfassung vornimmt.

Ich empfehle sehr, bei einer Pilotstudie auch die eigene Arbeitszeit zu protokollieren. Das hilft bei der Arbeitsplanung für die eigentliche Studie. Man kann davon ausgehen, dass die Erfassung mit aufkommender Routine schneller erfolgt, aber die Pilotstudie hilft, den Ressourcenbedarf für das Erfassungsprojekt besser einzuschätzen.

## 12.7 DATENSICHERUNG

Erfasste Daten sind kostbar. Technische Unfälle ereignen sich bekanntermaßen vor allem dann, wenn man sie nicht erwartet und man sie am wenigsten brauchen kann. Die Examensarbeit muss auf einen festen Termin hin angegeben werden: Ausgerechnet jetzt fällt erst die Festplatte aus, dann ist plötzlich der Toner oder die Tinte des

Druckers erschöpft, das Papier verbraucht – und natürlich ist an diesem Wochenende der Copy-Shop des Vertrauens wegen einer Familienfeier ausnahmsweise geschlossen. Bezogen auf Ihre kostbaren Texte und Daten bedeutet das: Datensicherung ist eine permanente Notwendigkeit. Datensicherungen erfolgen auf externen Medien, die auch an einem anderen PC lesbar sind, also z. B. auf externen Festplatten, USB-Sticks, usw. Alternativ kann man Sicherungen auch in der Cloud vornehmen. Dies erscheint bequemer, doch bei größeren Datenmengen kann die Übertragung der Daten recht lange dauern, sodass das Sichern auf externen Medien letztlich effizienter sein kann.

Eine Sicherung ist besser als keine Sicherung, aber zwei Sicherungen erhöhen das Sicherheitsniveau erheblich. Wie man das besonders geschickt macht? Der heutige Arbeitsstand landet auf der externen Festplatte Nr. 1. Der morgige Arbeitstag landet auf Festplatte Nr. 2. Übermorgen wird die Sicherung auf Festplatte Nr. 1 überschrieben, usw. Mit einem System gerade Tage/ungerade Tage lässt sich solch ein Wechsel leicht merken. Ergebnis: Es liegt immer extern der jüngste und auch der vorangehende Arbeitsstand vor. Sollte – warum auch immer – die aktuelle Datei auf Ihrem PC defekt sein, ist in Folge auch die aktuell gesicherte Datei defekt! In diesem Fall hätten Sie aber immerhin noch den vorletzten Arbeitsstand in einer intakten Sicherung vorliegen. Sicherheitsfanatiker spielen dieses rollierende Schema sogar mit drei statt zwei externen Medien und erreichen damit ein schon fast professionelles Sicherheitsniveau.

Übrigens: Man ist kein Sicherheitsfanatiker, sondern einfach nur vernünftig, wenn man die externe(n) Festplatte(n) oder die USB-Sticks mit den Sicherheitskopien nicht genau da aufbewahrt, wo sich das Daten-Original, also z. B. der Laptop, befindet. Man hat schon von Einbrüchen gehört, von Zimmer- und Wohnungsbränden...

Daten sichern. Wirklich.

Auf einem externen Medium: Stick, externe Festplatte... & diese Sicherung getrennt vom PC/Laptop in anderem Raum aufbewahren.

Einmal ist besser als nie. Zwei Mal ist gut.

## 13 ÜBER NAMEN UND MESSGENAUIGKEITEN

---

Nun, da die ersten Datensätze erfasst, auf Fehler geprüft und gut gesichert sind, ist es vor dem Einstieg in die richtige Auswertung an der Zeit, über zwei Elemente nochmals nachzudenken: Über die Namensvergabe und über die Messgenauigkeit Ihrer quantitativen Daten.

### 13.1.1 Namensvergabe: Namen, Label, Etiketten

„Damals“, in den 1970er und -80er Jahren, waren die Regeln einfach und übersichtlich: Ein Computerwort konnte das Äquivalent von exakt acht Zeichen gut verarbeiten. Aus dieser Zeit stammt die Sitte, zwischen „Namen“ und „Label“ (Etiketten) zu unterscheiden. Eine Variable erhielt einen Namen aus Buchstaben und Zahlen, acht Zeichen lang, vornan stets ein Buchstabe, und im Namen keine Sonderzeichen. Umlaute wie ä, ö, ü wurden vermieden, da viele Programme mit ihnen nicht zurechtkamen. Wenn das zu kurz war, gab es zusätzlich die Option, ein Label zu vereinbaren, d. h. ein Etikett, das – so möglich – z. B. bei Grafiken zusätzlich zum Namen ausgespielt wurde. Die Arbeit im Inneren der Prozesse geschah jedoch auf Basis der 8-Zeichen-Namen. Heutzutage sind diese Restriktionen technischer Natur (meist) nicht mehr gegeben. Dementsprechend wurden auch die meisten Programme angepasst und haben diese Schranken aufgegeben. Doch bevor wir diese Freiheit nun bedenkenlos genießen: Macht es auch Sinn, auszunutzen, was geht? Nein. Weiterhin sind – auch ohne starre Schranken – wohlüberlegte und kurze Namen sinnvoll. Bitte bedenken Sie:

- Sehr lange Namen z. B. in Spaltenköpfen behindern die Übersichtlichkeit der Tabellen, und später auch der Ausgaben.
- Kurze Namen sind praktisch. Muss man z. B. Berechnungen durchführen, in denen Variablennamen eingesetzt werden, schreibt sich „RDm + BDm / H“ statt „Randdurchmesser + Bauchdurchmesser / Gesamthöhe“ schneller und auch fehlerärmer als die Langfassung. Indes:
- Kryptische Abkürzungen, die später auch in Grafiken und Tabellen auftauchen und folglich in Abbildungsunterschriften wieder mühsam aufgelöst werden müssen, sind ebenfalls nicht hilfreich. Bedenken Sie, dass sogar Sie gelegentlich ein Projekt auf die Seite legen (müssen). Wenn Sie dann nach Monaten wieder hineinfinden und ihre Kürzel schnell und zuverlässig verstehen wollen: unrund, wenn Sie selbst Ihre Abkürzungen vergessen haben.
- Weiterhin mögen manche Programme keine deutschen Umlaute. Es ist klug, in Variablennamen auf ä, ö, ü, ß u.ä. zu verzichten.

Bilanz: Eine wohlüberlegte Namengebung ist wichtig. Spätestens jetzt, nach der Datenerfassung und -prüfung, sollten Sie diese einrichten.

Die Überlegungen gelten auch für Nominalvariablen und codierte Variablen, also solche, wo man 1 = rot, 2 = gelb und 3 = grün vereinbart hat. Denn meist können für die tatsächlich gespeicherten Zahlen 1, 2, 3, zusätzlich Label (Etiketten) vereinbart werden, die dann vor allem in den grafischen Ausgaben auftauchen sollen. Folglich gilt für die Label das gleiche wie für die Variablennamen: Kryptische Abkürzungen sind wenig sinnvoll, epische Länge noch weniger.

### 13.1.2 Messgenauigkeit

Man kann das gleiche Phänomen unterschiedlich genau messen, und dies macht Sinn. Nehmen wir als Beispiel das Messen von Strecken. Die Körperhöhe von Menschen wird, z. B. beim Arztbesuch oder bei der Rekruten-Musterung, in cm gemessen und taucht so z. B. auch im Personalausweis auf. Dahinter steht das Wissen, dass zwar technisch auch eine Messung auf Millimeter möglich ist, angesichts der Tagesschwankungen der Körperhöhe, die beim Menschen gerne mal um 2 cm beträgt, ein Messen in Millimetern jedoch eine Exaktheit vortäuschen würde, die das vermessene Phänomen gar nicht hat. Also notiert man in solchen Fällen z. B. 160 cm, 180 cm usw., und nicht 160,0 cm oder 1,6 m.

Bisweilen steht aber diese Messgenauigkeit gar nicht zu Verfügung. Jemand beobachtet Menschen, z. B. Besucher in einem Museum und schätzt unter anderem – warum auch immer – zusätzlich deren Körperhöhe. Das mag mit etwas Erfahrung auf den Dezimeter genau treffend möglich sein, aber nur wenige Menschen können dies zuverlässig im Zentimeterbereich ermitteln. In solchen Fällen würde man adäquat notieren: 1,6 m, 1,8 m usw. Eine Notation signalisiert immer auch, wie genau die Messungen erfolgten, im Eingangsbeispiel auf cm, im zweiten Beispiel auf Dezimeter.

Diese Orientierung am eingesetzten Messinstrument ist zu ergänzen um eine Orientierung an der Sache. Beim Vermessen von Keramikgefäßen macht es einen Unterschied, ob wir es mit Horgener Keramik zu tun haben (grob, groß, handgemacht, meist nur in kleinen Fragmenten überliefert) oder mit Terra Sigillata der ersten Jahrzehnte nach Christi Geburt (klein, dünnwandig, Feinware, Drehscheibe, genormte Manufakturware, wenn nicht vollständig, so doch oft sehr exakt rekonstruierbar). Jeder Bearbeiter muss sich selbst Rechenschaft darüber ablegen, ob er bei einem rekonstruierten Gefäßdurchmesser die Angabe von Millimetern vertreten möchte, oder eher

nicht. Ein Selbstversuch kann helfen: Die gleichen Gefäße an drei aufeinander folgenden Tagen vermessen, und die drei Protokolle miteinander vergleichen.<sup>25</sup> Obwohl es auf Unerfahrene unsystematisch wirken kann: Es kann sinnvoll sein, an Gefäßen die Höhen und Durchmesser in cm anzugeben, die mit der Schieblehre vermessenen Wandungsstärken hingegen in mm.

Die Genauigkeit, die Ihren Daten innewohnt, sollte sich in der Skala in Ihrer Datentabelle niederschlagen. Sollten gegenüber der Skala, in der die Informationen während der Erfassung zunächst abgelegt wurden, Änderungen sinnvoll sein, lassen sich diese in der Regel durch eine Multiplikation resp. Division durch 10, ggf. 100, angemessen umsetzen. Sprich: Der Aufwand für die Umsetzung ist gering, das diesbezügliche Überlegen und ggf. Korrigieren spätestens an dieser Stelle jedoch nötig.

### **13.2 DOKUMENTATION UND DATENBESCHREIBUNG**

Inzwischen steckt viel Wissen und manche Konvention in Ihrem Datensatz. Möglicherweise müssen Sie Ihre Arbeit irgendwann einmal für längere Zeit unterbrechen und werden erst später wieder fortfahren können. Möglicherweise gehören Sie zu den glücklichen Menschen mit einem guten Gedächtnis und haben alles noch im Kopf, was Sie beim ersten Teil ihrer Datenerfassung mit sich selbst vereinbart haben. Aber auch dann werden Sie Ihre Daten später einmal mit Kollegen teilen oder sie als Open Data veröffentlichen wollen, wie es heute in der Wissenschaft mehr und mehr verlangt und zur Regel wird. Kurz: Es ist sinnvoll, die Datenaufnahme und die dabei verfolgten Konventionen auch zu dokumentieren. Also über die Datentabelle hinaus ein kurzes, nüchternes Textdokument anzulegen, in dem festgehalten ist, wer den Datensatz erfasst hat, was erfasst wurde, und was die einzelnen Variablen (Spalten) bedeuten, welche Konventionen verfolgt wurden und welche Entscheidungen bei der Erfassung getroffen wurden. Eine solche Datenbeschreibung ist notwendig und sie anzulegen (und ggf. immer wieder zu aktualisieren) mag circa so attraktiv sein wie das Ausfüllen einer Steuererklärung, aber (wie bei der Steuererklärung): mit dem Aufschieben wird's nicht besser. Also beginnen Sie mit diesem Dokument – oder ggf. Dokumenten: für jeden Datensatz eines – frühzeitig; die Arbeit muss eh getan werden.

---

<sup>25</sup> Das ist die einfache Fassung des wichtigen Themas „Observer Error“, das wir schon behandelt haben.

### 13.3 BEISPIELDATENSÄTZE

Für die praktischen Übungen in diesem Buch benötigen wir Daten. Gerne bereits Ihre eigenen. Sollten solche jedoch (noch) nicht zur Verfügung stehen, können die hier mitgegebenen Beispieldatensätze verwendet werden. Sie sollen nachfolgend kurz vorgestellt werden.

#### 13.3.1 [1] Geldbörsen 1989 ff.: „Muenzen“

Seit 1989 habe ich – meist in Proseminaren zur Einführung in die Ur- und Frühgeschichte oder Übungen zur „Einführung in die Archäo-Statistik“ – Studierende gebeten, ihre Geldbörse zu erfassen: den Wert der Münze („Nominal“), das Prägejahr, die Währung, das die Münze emittierende Land und – sofern gegeben – den Prägeort. Hinzugefügt wurde das Jahr dieser Erhebung („Stichjahr“), der Ort der Erhebung sowie eine Fallnummer/ID, um die individuellen Geldbörsen (anonym) zu kennzeichnen. An diesem Modell lässt sich das für die frühgeschichtliche Archäologie so wichtige Thema „*terminus post quem*“ sowie Münz- und Emissionsgeschichte handfest erfahrbar darlegen, sodann bieten die Münzen – statistisch gesehen – mehrere unterschiedliche, typische Skalenniveaus und daher einen guten Übungsfall. Vor allem im Umfeld der Einführung des Euro an Stelle der Deutschen Mark (DM) am 1. Januar 2002 haben auch Kollegen an anderen Orten zu dieser Sammlung beigetragen. Aktuell umfasst der Datensatz mehr als 7.700 Münzen aus über 450 Geldbörsen, das jüngste Erfassungsjahr ist 2019. Allen Beitragenden sei nochmals herzlich gedankt! Der originale Datensatz wird hier in einer reduzierten Fassung verwendet, nämlich bereinigt um alle (seltenen) Münzen jenseits der drei Währungen DM, CHF und Euro, die im Originaldatensatz etwa 0,35 % ausmachen. Die Emissionsländer wurden mit den international üblichen Auto-Länderkennzeichen abgekürzt. Die Prägeorte wurden hier nur für die in Deutschland geprägten Münzen tabelliert und sind als Zahl codiert: 1: „A“ Berlin; 2: „D“ München; 3: „F“ Stuttgart; 4: „G“ Karlsruhe und 5 „J“ Hamburg.

Wenn Sie zum weiteren Wachstum des Datensatzes beitragen wollen: gerne! Senden Sie mir Ihre digital erfasste(n) Geldbörse(n), im gleichen Format wie dieser Beispieldatensatz, als \*.csv oder \*.xls/\*.xlsx-Tabelle an: [mail@frank-siegmund.de](mailto:mail@frank-siegmund.de)

Technisch wird dieser Datensatz hier als \*.csv-Tabelle eingeführt, wobei – entgegen der Standardbezeichnung „csv“ alias „comma separated values“ – tatsächlich das Semikolon als Trennzeichen verwendet wird. Solche csv-Tabellen sind u. a. auch mit LO-Calc lesbar. Ich empfehle, zumindest einmal dazu den „editor“ zu nutzen, um den Datensatz zu öffnen und anzuschauen: Die erste Zeile weist, jeweils mit Semikolon getrennt, die Variablennamen auf, die Folgezeilen enthalten die Daten. Zwei aufeinanderfolgende Semikolons signalisieren einen fehlenden Wert.

### 13.3.2 [2] Neugeborene aus Braunschweig

Der Datensatz „Babies“ geht auf die Website des Klinikums Braunschweig zurück, das in der Region u. a. für seine Abteilung Frauenheilkunde und Geburtshilfe einen guten Ruf genießt. Als Bestandteil des Leistungsangebotes seiner Geburtsklinik, die nach eigenen Angaben (Stand 2019) jährlich ca. 2.000 Geburten betreut, führt die Klinik eine jeweils aktuelle öffentliche „Babygalerie“ auf seiner Website, in der die Neugeborenen mit Namen, Bild und einigen Daten vorgestellt werden: <https://klinikum-braunschweig.de/aktuelles-veranstaltungen/babygalerie.php> [24.1.2020]. Aus zwei Zeitabschnitten zurückliegender Jahre (März bis Oktober 2007; April bis Juni 2008) wurden die Angaben für insgesamt 519 Neugeborene erfasst. Der Datensatz ist gut geeignet, um etwa den Zusammenhang von Körpergröße und -gewicht zu untersuchen oder um Beobachtungen zu Körpergröße und -gewicht bei Jungen und Mädchen, bei Zwillingen und Nicht-Zwillingen zu vergleichen. Es wäre durchaus reizvoll, den Datensatz um eine aktuelle Stichprobe aus der gleichen Quelle zu erweitern, weil sich dann auch nach zeitlichen Effekten fragen ließe.

Der Datensatz wird wieder als \*.csv-Datei mitgegeben, sein Inhalt versteht sich weitgehend von selbst. Bei der Variable „Zwilling“ bedeutet die Ausprägung „0“ naheliegenderweise „kein Zwilling“, „1“ ja, Zwilling. Bei der Variable „Geschlecht“ steht 1 für Knabe, 2 für Mädchen.

### 13.3.3 [3] Anthropologie Gräberfeld Zwentendorf (Niederösterreich)

Als Beispiel für einen Datensatz von einer physisch-anthropologischen Aufarbeitung eines Gräberfeldes sei hier – aus einer Vielzahl anderer Optionen – beispielhaft das mittelalterliche Gräberfeld von Zwentendorf an der Donau (Bezirk Tulln, Niederösterreich) herausgegriffen, das im Ruinenareal eines römischen Kastells entdeckt wurde und seit dem späten 10. Jahrhundert und vor allem im 11. Jahrhundert n. Chr. belegt wurde (Giesler, 1997, passim, insbes. 192). Die 254 Körpergräber sind nahezu beigabenlos, die wenigen Beigaben ordnen die Bestattungsgemeinschaft der sog. Köttlacher Kultur zu. Die anthropologische Bearbeitung hat Wolfgang Heinrich (2001) publiziert, die verwendeten Daten stammen, leicht verändert, von der der Publikation beiliegenden CD.

Für unsere Zwecke wurde der originale Datensatz stark reduziert. Der Übungsdatensatz mit den Knochenmaßen klammert alle Nicht-Erwachsenen aus und er verzichtet auf viele Schädelmaße und Langknochenmaße. Bei den Langknochen wurden zudem die Werte der rechten und linken Körperseite zusammengezogen; sprich: Im Falle des Vorliegens beider Werte der Mittelwert gebildet, sonst der jeweils verfügbare Wert eingesetzt.

- Heinrich, W. (2001). *Zwentendorf – Ein Gräberfeld aus dem 10. – 11. Jahrhundert. Anthropologische Auswertung* (Mitteilungen der Prähistorischen Kommission 42). Wien: Österreichische Akademie der Wissenschaften.
- Giesler, J. (1997). *Der Ostalpenraum vom 8. bis 11. Jahrhundert. Studien zu den archäologischen und schriftlichen Zeugnissen. Tl. 2: Historische Interpretation* (Frühgeschichtliche und Provinzialrömische Archäologie: Materialien und Forschungen 1). Rahden/Westf.: Leidorf.

### 13.3.4 [4] Howells‘ world data

Der US-amerikanische Anthropologe William W. Howells (1908-2005) erfasste in den Jahren 1965 bis 1980 verschiedene Maße an rezenten und subrezentem menschlichen Schädeln, und zwar an 2.524 Schädeln von 28 unterschiedlichen Populationen aus allen Kontinenten. Ein weiterer, sog. Testdatensatz enthielt 524 Schädel. Die Maße wurden nach dem US-amerikanischen Anthropologie-Standard erfasst, der ursprünglich auf das Lehrbuch von Rudolf Martin (1914) zurückgeht, sich davon jedoch etwas weiter- und wegentwickelt hat. Vor allem sind die seit Martin (1914) in Europa nahezu überall verwandten Benennungen der Messstrecken heute in US-Amerika anders, und manchmal differieren auch die Definitionen der Messstrecken leicht. In dem hier benutzten Datensatz sind die Strecken – soweit möglich – wieder nach dem Martin’schen System benannt. Howell hat seine Studie in drei Monografien publiziert (1973; 1989; 1995) sowie einen kurzen Aufsatz veröffentlicht, der nur die Genese des Datensatzes und das Messsystem schildert (Howells 1996). Beide Datensätze von Howell wurden später von dem US-amerikanischen Anthropologen Benjamin M. Auerbach in moderne Excel-Tabellen übertragen und frei verfügbar gemacht.<sup>26</sup> Die von mir wie beschrieben aufbereitete Fassung der World Data von Howells (i. e. 2.524 Schädel) dient hier als Übungsbeispiel. Bei den Maßen handelt es sich um biologische Variablen, der Verdacht auf Normalverteilung liegt nahe. An ihm können verschiedene Test-Verfahren erlernt werden, das Berechnen und Auswerten von Indizes geübt werden, und der Datensatz eignet sich sehr gut für zwei- und mehrdimensionale Betrachtungen („multivariate Verfahren“).

- Howells, W. W. (1973). *Cranial Variation in Man. A Study by Multivariate Analysis of Patterns of Differences Among Recent Human Populations* (Papers of the Peabody Museum of Archaeology and Ethnology 67). Cambridge, Mass.: Peabody Museum.

---

<sup>26</sup> Quelle: <http://web.utk.edu/~auerbach/HOWL.htm> [24.1.2020].

- Howells, W. W. (1989). *Skull Shapes and the Map. Craniometric Analyses in the Dispersion of Modern Homo* (Papers of the Peabody Museum of Archaeology and Ethnology 79). Cambridge, Mass.: Peabody Museum.
- Howells, W. W. (1995). *Who's Who in Skulls. Ethnic Identification of Crania from Measurements* (Papers of the Peabody Museum of Archaeology and Ethnology 82). Cambridge, Mass.: Peabody Museum.
- Howells, W. W. (1996). Howells' Craniometric Data on the Internet. *American Journal of Physical Anthropology* 101, 1996, p. 441-442. <http://web.utk.edu/~auerbach/Howells.pdf> [24.1.2020].
- Martin, Rudolf (1914). *Lehrbuch der Anthropologie in systematischer Darstellung; mit besonderer Berücksichtigung der anthropologischen Methoden; für Studierende, Ärzte und Forschungsreisende*. Jena: Gustav Fischer. – Online-Ausgabe (CC BY SA): <http://digi.ub.uni-heidelberg.de/diglit/martin1914> [24.1.2020].

Bei anthropologischen Aufnahmen z. B. an Gräberfeld-Toten in Mitteleuropa werden in der Praxis insbes. am Schädel erheblich weniger Maße erfasst, als Martin sie definiert und auch als Howells gemessen hat. Für Übungszwecke ist es vernünftig, sich auf die wichtigsten Maße zu beschränken; jene, die häufig beobachtet und ausgewertet werden. In der europäischen Tradition werden oft erfasst und ausgewertet:

#### *Hirnschädel*

M1: Größte Hirnschädellänge

M5: Schädelbasislänge

M8: Größte Hirnschädelbreite

M9: Kleinste Stirnbreite<sup>27</sup>

M17: Basion-Bregma-Höhe

M20: Ohr-Bregma-Höhe

#### *Schädelbasis, Hinterhauptsloch*

M7: Länge Foramen magnum

M16: Breite Foramen magnum

#### *Gesichtsschädel*

M40: Gesichtslänge

M45: Jochbogenbreite

M47: Gesichtshöhe

M48: Obergesichtshöhe

M51: Orbitalbreite (ca. Howell m51a)

M52: Orbitalhöhe

M54: Nasenbreite; M55: Nasenhöhe

---

<sup>27</sup> Einrückung: Dieses Maß wird bei Howells leider nicht erfasst.

Daher wird hier zur leichteren Handhabbarkeit neben dem vollständigen Datensatz ein auf diese Messstrecken reduzierter Datensatz verwendet. Bilder des Schädels mit Eintragung der Messstrecken finden sich in der online verfügbaren Ausgabe von Martin (1914), ab S. 520 Fig. 217:

<http://digi.ub.uni-heidelberg.de/diglit/martin1914/0542?sid=4a68925af2c50640691c442911c26e1d> [24.1.2020].

### 13.3.5 [5] Anatomieleichen aus Lyon: Rollet 1888

Der Datensatz enthält Geschlecht, Alter und die beobachtete Leichenlänge von je 50 Männern und Frauen, an denen der Anatom Etienne Rollet (1888) auf beiden Körperseiten die Länge der Langknochen dokumentierte. Dieser Datensatz wiederum war für Karl Pearson Anlass für einen Versuch, mathematische Formeln zur Körperhöhen-schätzung aus Langknochen zu suchen. In seinem Aufsatz entwickelte er dazu das (mit wenigen Abwandlungen) bis heute übliche Verfahren der linearen Regression und zugleich Schätzformeln für die Körperhöhe, die bis heute angewendet werden. Der Datensatz ist gut geeignet, eigene Erfahrungen mit dem Thema Korrelation und Regression zu sammeln.

Unter „Sex“ steht 1 für Männer, 2 für Frauen; „Alter“ nennt das erreichte Lebensalter; „Knoche“ nennt die beobachtete Leichenlänge in cm, die als etwa 2,0 – 2,5 cm länger gilt als die Lebendhöhe. Alle Knochenmaße sind in mm angegeben und nach den Definitionen im Lehrbuch von Martin (1914) benannt; es wurden jeweils die rechte (-re) und linke Körperseite (-li) gemessen, wobei es einige wenige fehlende Werte gibt. Am Ende des Datensatzes werden die Langknochenmaße wiederholt, gemittelt aus beiden Körperseiten.

• Rollet, E. (1888). *De la mensuration des os longs des membres dans ses rapports avec l'anthropologie, la clinique et la médecine judiciaire*. Lyon: A. Storck.

<https://gallica.bnf.fr/ark:/12148/bpt6k97653475.texteImage> [24.1.2020].

• Pearson, K. (1899). *On the reconstruction of the stature of prehistoric races. Mathematical contributions to the theory of evolution 5*. (Philosophical transactions of the Royal Society of London A 192). (p. 169-244). London: Royal Society of London.

<https://archive.org/details/philtrans06400427> [24.1.2020].

### 13.3.6 Dateiformate

Leider sind manchmal Datenformate trotz identischen Namens nicht wirklich gleich, insbesondere hinter den \*.xls- und \*.xlsx-Dateien stecken oft unerwartet technisch unterschiedliche Inhalte. Um Frust beim Import und Verwenden der Beispieldaten zu

vermeiden, geben wir die Beispieldatensätze in diversen Formaten mit, in der Hoffnung, dass schon irgendeines für die speziellen Einstellungen Ihres Computers und Ihrer Software passt. Sowohl beim R-Commander als auch bei BlueSky hat sich das Einlesen der SPSS-Dateien (\*.sav) als meist problemlos erwiesen, die Formate \*.csv und \*.dat („Tab separated“) als unkomplizierter denn \*.xlsx.

## 14 FREIE STATISTIK-PROGRAMME

---

Nun schauen wir uns vier freie Statistikprogramme näher an, die für Anwender gut geeignet sind, die nur einmal kurz ein paar Kleinigkeiten berechnen und analysieren möchten und sich dafür nicht lange in eine Software einarbeiten wollen. Die genannten Programme sind bewährt und leistungsfähig, in ihrem Umfang jedoch begrenzt. Insbesondere ist eine komfortable Verwaltung der Daten – ein für umfassendere Analysen wichtiger Aspekt – mit ihnen (derzeit) nur eingeschränkt möglich. Gleichwie können sie wertvolle Instrumente da sein, wo die Statistik-Möglichkeiten von LO-Calc ausgeschöpft sind und man dort nicht abgedeckte Verfahren einsetzen will.

Die vier Programme unterscheiden sich deutlich. Eine persönliche Einordnung ihrer jeweiligen Stärken nach Stand 2020 in Kürze: PAST hat einen großen Leistungsumfang, es eignet sich (wenn man nicht mit **R** oder z. B. SPSS arbeiten will) derzeit wohl am meisten für professionelles Arbeiten. JASP hilft nach Ansicht seiner Entwickler, nebenbei auch niederschwellig erste Erfahrungen mit der Bayes-Statistik zu machen. JAMOVI ist eine vielversprechende neue Entwicklung mit schnell erlernbarer Bedienung; es ermöglicht zugleich, nebenbei auch erste Erfahrungen mit **R** zu sammeln. JAMOVI ist elegant und hat sehr klar und modern wirkende Grafiken, seine weitere Entwicklung scheint zügiger zu verlaufen als diejenige von JASP. BlueSky Statistics beruht wie JASP und JAMOVI im Hintergrund auf **R**, kommt aber als voll- und selbstständiges Programm zum Nutzer, was den Start unkompliziert macht. Es wirkt weniger modern als JAMOVI, ist in vielem an SPSS orientiert, in seiner Bedienung manchmal etwas störrisch, beinhaltet aber neben der üblichen Statistik derzeit in dieser Gruppe die weitestgehenden Werkzeuge zur Datenverwaltung. Insofern ist BlueSky

Stand Anfang 2020 statt des Pärchens LO-Calc & PAST eine ernstzunehmende All-in-One Alternative für alle Diejenigen, die den direkten Weg zu **R** (noch) scheuen.<sup>28</sup>

## 14.1 PAST

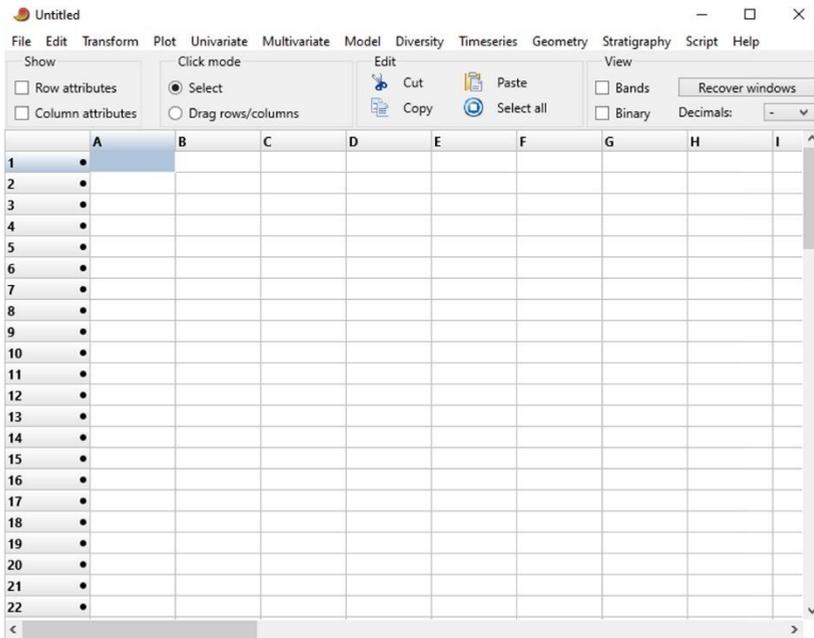
PAST – abgekürzt für „**PA**leontological **ST**atistics“ – wird seit 2001 als freie Software von Øyvind Hammer (Naturhistor. Museum, Univ. Oslo) angeboten und stetig weiterentwickelt. PAST wird primär für Windows entwickelt, aber mit etwas Zeitverzögerung gibt es jeweils auch eine Version für den MAC, während eine LINUX-Version nicht vorgesehen ist. Im Januar 2020 lag PAST in der Programmversion 4.0 vor, die hier zu Grunde gelegt ist. Der Autor beschreibt das Programm kurz als „*free software for scientific data analysis, with functions for data manipulation, plotting, univariate and multivariate statistics, ecological analysis, time series and spatial analysis, morphometrics and stratigraphy.*“ In der Verfahrensauswahl und der Weiterentwicklung liegt der Fokus neben den üblichen Standards deutlich auf paläontologischen Aufgabenstellungen, was in vielen Themen von der Archäologie nicht weit entfernt ist. Mit PAST lassen sich wirklich viele auch anspruchsvolle Auswertungen machen, inklusive einer soliden Korrespondenzanalyse. Die Bedienung zu erlernen dauert kaum mehr als eine halbe Stunde, das mitgelieferte Handbuch vermittelt alles Nötige. Alles Weitere findet sich auf der Website des Programms: <https://folk.uio.no/ohammer/past/> [24.1.2020].

### 14.1.1 PAST: Installation und erste Schritte

Das Programm kann von der genannten Website heruntergeladen werden, es landet als ZIP-gepackte Datei auf dem eigenen Rechner. Laden Sie am besten gleich auch das Handbuch herunter. Danach muss es entpackt (entzippt) werden. Nun sollte die Datei mit ihrem markanten orange-roten Ikon verfügbar sein. Das war's: PAST bedarf keiner weiteren Installation, die \*.exe-Datei ist alles Nötige. Das Arbeiten mit PAST wird vereinfacht, wenn Daten und PAST im gleichen Ordner liegen. Daher empfehle ich, für ein Projekt einen eigenen Ordner anzulegen, in den Sie die Daten sowie die PAST.exe hineinkopieren. Ein Doppelklick auf das PAST-Ikon startet das Programm, es öffnet sich eine Oberfläche, die auf den ersten Blick einer Tabellenkalkulation ähnelt (**Abb. 15**).

---

<sup>28</sup> Siehe zu diesem Thema die sorgfältigen Tests und Analysen von Robert Muenchen mit einem abschließenden Vergleich der Tools; seine Bewertungen entsprechen meinen Erfahrungen vollauf: Muenchen, Bob (2019). Comparing Point-and-Click Front Ends for R. *R-bloggers*, April 22, 2019. <https://www.r-bloggers.com/comparing-point-and-click-front-ends-for-r/> [24.1.2020].



**Abb. 15** Der leere Programmbildschirm von PAST (Vers. 4.0).

PAST liest und schreibt u. a. \*.xls-Dateien. Lesen wir doch die Geldbörse(n), die wir erfasst haben, in PAST ein: => File; => Open; es öffnet sich das übliche Dateifenster, wo Sie an gegebener Stelle auf Ihre Datei „Geldboersen“ klicken. PAST öffnet – in englischer Sprache – ein Menü und fragt nach, wie der Inhalt der Zeilen („rows contain“) und Spalten („columns contain“) zu lesen ist. Sollte Ihre Datei so vorliegen wie eingangs vereinbart, ist für die Zeilen (rows) „only data cells“ richtig und für die Spalten (columns) die Angabe „Names, data“. Bitte so markieren & OK absenden. Ihre Münzen müssten nun sichtbar sein, PAST ist bereit zur Datenanalyse (**Abb. 16**). Erleichtern Sie sich das Lesen der Daten, indem Sie rechts oben bei „View“ den Schaltknopf „Bands“ anklicken.

Warum wir im vorliegenden Fall für Zeilen und Spalten eine unterschiedliche Auswahl bei der Spezifizierung für den Import treffen? Unsere \*.xls-Tabelle enthält – wie es üblich ist – in den Zeilen nur Daten, während in den Spalten die oberste Zelle statt Daten jeweils den Spaltenkopf, d.h. den Namen und die Bedeutung der Spalte beinhaltet, und erst die folgenden Zellen dann echte Datenwerte enthalten. Indem wir die Wahl wie beschrieben passend wählen, übernimmt PAST auch die Spaltenköpfe.

Wer's näher studieren möchte: Die Ausgangsdaten werden ja nicht zerstört, wenn man etwas falsch macht. Also kann man durchaus mit den Einstellungen experimentieren. Am Vergleich zwischen Einstellungen und Ergebnis wird schnell deutlich, wann man was vernünftigerweise anklickt.

	Nominal	Jahr	Waehrung	Land	Ort	ID	Stichjahr	Stichoort	Alter
1	• 1	1968	DM	D	#NULL!	1001	1989	Goettingen	21
2	• 1	1972	DM	D	#NULL!	1001	1989	Goettingen	17
3	• 1	1982	DM	D	#NULL!	1001	1989	Goettingen	7
4	• 1	1983	DM	D	#NULL!	1001	1989	Goettingen	6
5	• 2	1963	DM	D	#NULL!	1001	1989	Goettingen	26
6	• 2	1976	DM	D	#NULL!	1001	1989	Goettingen	13
7	• 2	1979	DM	D	#NULL!	1001	1989	Goettingen	10
8	• 2	1982	DM	D	#NULL!	1001	1989	Goettingen	7
9	• 2	1950	DM	D	#NULL!	1002	1989	Goettingen	39
10	• 2	1974	DM	D	#NULL!	1002	1989	Goettingen	15
11	• 10	1950	DM	D	#NULL!	1002	1989	Goettingen	39
12	• 10	1950	DM	D	#NULL!	1002	1989	Goettingen	39
13	• 10	1966	DM	D	#NULL!	1002	1989	Goettingen	23
14	• 10	1975	DM	D	#NULL!	1002	1989	Goettingen	14
15	• 10	1979	DM	D	#NULL!	1002	1989	Goettingen	10
16	• 10	1979	DM	D	#NULL!	1002	1989	Goettingen	10

Abb. 16 Die erfassten Geldbörsen mit ihren Münzen im Startbildschirm von PAST.

### 14.1.2 Datenimport nach PAST: Frust vermeiden

Ganz so einfach wie hier ist es vielleicht nicht immer, daher ein paar Hinweise zu diesem wichtigen und sich häufig wiederholenden Arbeitsschritt.

- PAST liest \*.xls-Dateien, aber nicht die moderneren \*.xlsx-Dateien. Dies ist beim Abspeichern aus dem ursprünglichen Programm heraus zu beachten.
- Bei den \*.csv-Dateien ist ein steter Quell von Ärgernissen, ob ein Komma oder ein Semikolon als Trennzeichen eingesetzt wird – was meist davon abhängt, welche Ländereinstellungen im Quellprogramm gemacht wurden. Scheitert der Import von \*.csv-Dateien, ist meist die Frage Semikolon oder Komma der Grund. Oft ist es weitaus weniger fehleranfällig, auf „TAB delimited“ als Dateiformat für den Austausch auszuweichen, wo ein Tabulator als Trennzeichen eingesetzt wird; erfahrungsgemäß sind diese Dateien (erheblich) unproblematischer beim Import nach PAST.
- Ein weiterer Störfaktor können die Tausenderzeichen bei Zahlen sein, die beim Import von PAST gerne als Punkte wahrgenommen werden, weshalb die Zahlen dann als Buchstaben gelesen werden. Das kann man nur verhindern, indem man vor dem

Export im Quellprogramm alle Zahlen entsprechend formatiert, also ohne Tausenderzeichen.

- Ein weiterer Störfaktor können die fehlenden Werte sein, resp. deren Kodierung. Im Idealfall ist ein fehlender Wert in einer Datentabelle schlicht eine leere Zelle – so sollte man es bei der Dateneingabe halten. Manchmal aber füllt die verwendete Software diese leeren Zellen automatisch auf, z. B. mit „#NULL!“ . Programm-intern kein Problem. Beim Import nach PAST aber werden Spalten, in denen #NULL! auftaucht, also Zeichenketten, insgesamt als Spalten mit alphanumerischen Inhalten (i. e. Text) verstanden, d. h. auch die Zahlen in den übrigen Feldern werden nicht als Zahlen, sondern als Text aufgefasst. Daher ist es vorteilhaft, vor dem Datenexport im Quellprogramm ggf. die nötigen Einstellungen vorzunehmen.

Man kann dies auch im Nachhinein in PAST via Suchen & Ersetzen bereinigen: => Edit; => Replace; es öffnet sich ein Suchen-&-Ersetzen-Menü, wo man mit Hilfe der passenden Eingaben die störenden Zeichen löschen und durch nichts, als ein leeres Feld ersetzen lassen kann.

- Nicht elegant, aber schnell: PAST übernimmt Daten meist problemlos und schnell per Copy-&-Paste. Wobei das Einfügen nicht wie üblich mit der rechten Maustaste gelingt, sondern über das Menü von PAST erfolgen muss: => Edit; => Paste.

Ist der Datenimport wie eigentlich vorgesehen gelungen, haben die Spalten in PAST auch den von Ihnen zuvor vergebenen Namen (Variablenamen) übernommen. Wenn nicht, ist hier noch eine Bereinigung nötig. Oben links unter „Show“ ein Häkchen an „Column attributes“ setzen. Nun klappen unter der ursprünglichen Kopfzeile des Datensatzes zusätzlich zwei farbunterlegte Zeilen auf: „Type“ und „Name“. Tragen Sie in die Zeile „Name“ die Variablenamen ein. Wenn Sie das Häkchen von den „Column attributes“ wieder wegnehmen, klappen diese Zeilen weg und der Datensatz zeigt oben die richtigen Variablenamen.

Es lohnt, diesen Datenimport geduldig zu üben und ggf. etwas zu experimentieren, welche Einstellungen bei dem von Ihnen verwendeten Quell- und beim Zielprogramm am besten funktionieren, denn es ist ein wichtiger und häufig wiederkehrender Arbeitsschritt.

### 14.1.3 Schnell ein paar Statistiken und Grafiken mit PAST

Die oberste Zeile von PAST „File; Edit; Transform; ...“ enthält wichtige Untermenüs, die sich jeweils aufklappen, wenn man auf den obersten Begriff klickt. Wir wollen im

Folgenden schnell ein paar Statistiken und Grafiken abrufen, nicht, um nun tiefreichende Erkenntnisse über die Geldbörse zu gewinnen, sondern um PAST kennenzulernen.

Mit einem Klick auf den Rand, d. h. den Zeilenanfang oder den Spaltenkopf, aktiviert man eine Zeile oder Spalte in PAST. Hält man die Shift-Taste gedrückt (die Taste, die sonst von Klein- zu Großschreibung wechselt), kann man statt einer Spalte auch eine oder mehrere weitere Spalten als gleichzeitig aktiv markieren. Aktive (ausgewählte) Spalten (oder Bereiche) erscheinen in PAST hellblau hinterlegt. Wir markieren „Nominal“ und „Präcejahr“.

Die (gedrückt gehaltene) Shift-Taste ermöglicht es, Bereiche von Spalten zu markieren. Wählt man statt der Shift-Taste die STRG-Taste und hält sie gedrückt, kann man mehrere Spalten einzeln markieren, die nicht nebeneinanderstehen.

Wir rufen einfache Statistiken ab: (oben) => Univariate; => Summary Statistics; ...und das aufpoppende neue Menü zeigt das Ergebnis, die univariaten Statistiken für die Variablen Nominal und Präcejahr, u. a. die Anzahl der Fälle („N“), den kleinsten Wert („Min“), den höchsten Wert („Max“), den Mittelwert („Mean“), usw.

Wir rufen ein Histogramm ab, aber nur für „Nominal“, d. h. nur die Spalte Nominal markieren, dann (oben) => Plot; => Histogram; ...und in einem neuen Fenster taucht ein Histogramm auf. Rechts neben dem Histogramm sind ausgewiesen „Start“, der kleinste Wert, „End“, der höchste Wert und „Bins“, die Anzahl der automatisch gesetzten Klassen (*bin* = GefäÙe, Behälter). Mit „Flip axis“ kann man die Grafik um 90 Grad drehen, und die Anzahl der „bins“ kann mit den kleinen Pfeilen daneben schrittweise erhöht oder reduziert werden, um die Detailliertheit der Klassen zu verändern.

Nun, die von PAST automatisch gesetzten Details der Grafik sind noch nicht perfekt, Grafiken wie Statistiken kann man verbessern resp. verschönern – doch das soll erst später an passender Stelle vertieft werden. Hier geht es nur darum, erste Eindrücke von PAST und dessen Bedienung zu gewinnen.

#### 14.1.4 Feintuning der Datentabelle unter PAST

Die Eigenschaften der Zeilen und der Spalten lassen sich spezifizieren, oft ist es sinnvoll, dies angemessen zu tun. Im vorliegenden Falle geht es vor allem um die Spalten (Variablen). Dazu aktivieren wir die Schaltfläche (oben, links) „Column attributes“. Die Kopfzeile klappt nun aus und zeigt, leicht orangefarben hinterlegt, den „Type“ (Typ) der Variable und unter „Name“ ihren Namen. An letzterem müssen wir nichts tun. Den Typ der Variablen hingegen passen wir an: Variable „Nominal“ auf „nomi-

nal“, wie auch Wahrung, Land und Prageort. Dem Namen wurde ich „String“ (Zeichenkette) zuweisen. Das Pragejahr belassen wir, wie automatisch richtig von PAST angesetzt. Rufen Sie jetzt einen Plot & Histogramm ber Nominal auf, ist die Klasseinteilung von Anbeginn an richtiger – eben, weil wir jetzt PAST die Skala mitgeteilt haben, in der die Variable vorliegt. Nach Abschluss dieser Spezifizierungen das Hakchen bei „Columns attributes“ wieder wegnehmen.

#### 14.1.5 Datenmanagement mit PAST

Die Moglichkeiten der Datenverwaltung mit PAST sind begrenzt, aber gegeben. Was geht? Man kann die Daten auf- oder absteigend sortieren, was als Vorbereitung zur Auswahl von Fallen sehr hilfreich sein kann: => Edit; => Rearrange; => Sort ascending/descending.

Das Auswahlen von Fallen geschieht ber das Markieren der aktiven Bereiche. Man kann Teile von Spalten auswahlen/markieren: In die oberste auszuwahlende Zelle maus-klicken, Shift-Taste drucken und halten; in die unterste der auszuwahlenden Zellen klicken: Der markierte, d. h. ausgewahlte Bereich erscheint blau hinterlegt. Ruft man jetzt Grafiken oder Statistiken ab, werden sie nur fur den markierten Bereich berechnet.

Bei nicht zu groen Datensatzen – sagen wir: bis zu zwei Bildschirmseiten rechts/links wie oben/unten – erlauben diese rudimentaren Moglichkeiten ein recht schnelles Datenmanagement. Ab einer gewissen Dimension der Datentabelle wird dies unbersichtlich, d. h. man wird die primare Datenhaltung und -verwaltung auerhalb von PAST z. B. in der Tabellenkalkulation vornehmen.

Hinter einer weiteren Fahigkeit von PAST verbirgt sich ein wichtiges Werkzeug der PAST-internen Datenverwaltung: Man kann relativ leicht Zeilen und Spalten verschieben. Oben in der zweiten Bedienleiste von PAST finden sich die vier Bereiche „Show“, „Click mode“, „Edit“ und „View“. Im Bereich „Click mode“ steht der Schalter blicherweise (voreingestellt) auf „Select“ = Auswahlen. Bitte setzen Sie den Schalter auf „Drag rows/columns“. Klickt man jetzt in die Spaltenkopfe, kann man die Spalten gegeneinander verschieben. Dies ermoglicht es, sich fallweise im aktiven Datensatz die Variablen zusammenzustellen, die man gemeinsam analysieren mochte. Also z. B. zuerst zwei Variablen so zu verschieben, dass sie nebeneinanderstehen, dann („Click mode“ wieder auf „Select“ setzen) beide Spalten zu markieren, um deren Zusammenhang zu untersuchen. Insbesondere fur die multivariaten Verfahren ist diese Moglichkeit ungemein hilfreich.

PAST selbst verwendet ein einfaches ASCII-Zeichen-basiertes Format zum Speichern der Daten (\*.DAT). Wollen Sie die vorgenommenen anderungen fixieren

und den letzten Arbeitsstand schnell wieder aufrufen können, ist das ein guter Weg:  
=> File; => Save as; usw.

#### 14.1.6 Bilanz zu PAST

Nach kurzem Einüben lassen sich mit PAST viele Statistik- und Grafik-Aufgaben un-  
gemein schnell erledigen, die angebotenen Möglichkeiten gehen weit über das Einstei-  
gerniveau und auch den Inhalt dieser Einführung hinaus. Das Programm wird seit  
längeren Jahren vom Programmator gut gepflegt, weiterentwickelt und ggf. vorhan-  
dene Fehler mit der nächsten Version bereinigt. Ob dieser Stabilität ist das Einarbeiten  
in PAST auch keine Sackgasse.

Eine komplexere Datenverwaltung ist mit PAST jedoch nur schwer umsetzbar.  
Der – wie hier erlebt – leichte Import von Daten aus einer Tabellenkalkulation zeigt  
die Lösung an für alle Fälle, wo eine aufwendigere Datenaufbereitung nötig ist: Man  
kombiniert Tabellenkalkulation und PAST. Das Sortieren, Filtern und Aufbereiten der  
Daten, ggf. inkl. dem Bilden von Pivot-Tabellen, geschieht mit Hilfe einer Tabellen-  
kalkulation, für die Statistik und Grafik wechselt man zu PAST.

### 14.2 JASP UND JAMOVI

Wesentlich jünger am Markt als PAST sind die beiden kostenlosen Open-Source-Pro-  
gramme JASP und JAMOVI, deren Entwicklung ca. 2013 begann und die im Jahr  
2020 beide noch in einem in einem frühen Entwicklungsstadium vorliegen, aber un-  
bedingt einen Blick wert. Beide starten wie eine Tabellenkalkulation, arbeiten mit einer  
angenehm einfachen Point-&-Click-Bedienung und setzen das Bestellte unmittelbar,  
„live“ um. Beide Programme beruhen im Hintergrund (zum Teil) auf R – was dem  
Benutzer aber nicht aufgedrängt wird. Im Look-&-Feel sind JASP und JAMOVI ei-  
nander sehr ähnlich, weil JAMOVI eine Abzweigung des noch jungen JASP ist. Beide  
Programme wollen dem Benutzer ohne lange Anlernphase einfache wie komplexe stati-  
stische Auswertungen ermöglichen.

Dabei möchte JASP die Nutzer unerschwerlich an die Bayes-Statistik heran-  
führen: Erst werden die klassischen statistischen Verfahren angeboten, zusätzlich aber  
leicht erreich- und umsetzbar ihre Bayes-Varianten. So soll der Nutzer motiviert wer-  
den, letztere mal auszuprobieren und sich sukzessive anzueignen. Wer Bayes-Statistik  
anschaulich und in der Praxis lernen will, findet hier einen guten Start.<sup>29</sup>

---

<sup>29</sup> Rezensionen: Love, J. et al. (2019). JASP: Graphical Statistical Software for Common Statistical Designs. *Journal of Statistical Software*, 88(2). DOI 10.18637/jss.v088.i02 Open Access: <https://www.jstatsoft.org/article/view/v088i02> [24.1.2020]. // Muenchen, Bob (2019). A Comparative Review of the JASP Statistical Software.

JAMOVI hingegen möchte vor allem die Kooperation mit **R** erleichtern.<sup>30</sup> Wer will, kann sich seine in JAMOVI zusammengeklickte Analyse auch als R-Code anzeigen lassen und auf diesem Wege sukzessive das Schreiben von R-Code erlernen. Im Gegensatz zu JASP möchte JAMOVI auch von den Nutzern selbst weiterentwickelt werden, d. h. man kann eigene Routinen schreiben und in JAMOVI integrieren. Dieser Plan der Entwickler scheint aufzugehen, d. h. es finden sich tatsächlich Externe, die zusätzliche PlugIns für JAMOVI entwickeln. Angesichts der hübschen Oberfläche und der leichten Bedienbarkeit könnte sich JAMOVI tatsächlich zu einer neuen, beliebten grafischen Oberfläche für **R** entwickeln.

Näheres findet man auf den jeweiligen Websites (Stand Anfang 2020):

- JASP (0.11.1): <https://jasp-stats.org/> [24.1.2020]; verfügbar auch für MAC und Linux.
- JAMOVI (1.1.9 “current”): <https://www.jamovi.org/> [24.1.2020]; verfügbar auch für MAC und Linux. JAMOVI kann auch als R-Paket installiert werden (Erklärung unten, bei „R“).

Da mir der Gedanke einer Hinführung der Benutzer an **R** und des Aufbaus einer aktiven Nutzer- und Entwicklergemeinschaft, wie JAMOVI ihn verfolgt, wichtiger und zukunftsweisender erscheint als der einer subkutanen Annäherung an die Bayes-Statistik wie bei JASP, gehe ich im Folgenden vor allem auf JAMOVI ein. Zudem ist die Website von JAMOVI und die dort zu findende Hilfe und Einführung m. E. etwas geschmeidiger als bei JASP. Es spricht aber nichts dagegen, beide Programme parallel zu nutzen, denn das Hin- und Herwechseln ist ob ihrer 2020 noch gegebenen hohen Ähnlichkeit problemlos.

### 14.2.1 JASP: Installation und erste Schritte

Das Programm JASP findet sich auf der genannten Website, es sind Versionen für Windows, MAC und Linux (Ubuntu) verfügbar. Die Installation erfolgt nach den heute üblichen Standards, weshalb sie hier nicht detaillierter beschrieben werden muss. Mit Klicken auf das Programmsymbol öffnet sich der übersichtliche Arbeitsbildschirm. Über den Reiter „File“ kann man einen der beigefügten Beispieldatensätze

---

*R-Bloggers*, 17.4.2019: <https://www.r-bloggers.com/a-comparative-review-of-the-jasp-statistical-software/> [24.1.2020].

<sup>30</sup> Rezension: Muenchen, Robert A. (2018). A Comparative Review of the jamovi GUI for R. *r4stats.com* <https://r4stats.com/articles/software-reviews/jamovi/> [24.1.2020].

aktivieren oder mit Hilfe der Schalter „Open“ und „Browse“ eigene Daten einlesen. Neben seinem eigenen Datenformat (eine direkte Dateneingabe nach JASP ist möglich) liest JASP auch die Dateiformate \*.csv, \*.txt, \*.ods und \*.sav (SPSS) ein. Sofern die Automatik nicht perfekt funktioniert, kann man nachträglich die Notation für fehlende Werte verändern (leere Zelle, Leerzeichen, Punkt oder NA) und auch Daten in der Tabelle editieren. JASP unterscheidet beim Einlesen die Variablen automatisch in Zeichenketten, Nominaldaten, Ordinaldaten und Daten einer quantitativen Skala; mit Doppelklick auf den Spaltenkopf lässt sich dies – so inhaltlich möglich – anpassen, z. B. ein Schalten zwischen Nominal- und Ordinalskala, was später die Auswahl der möglichen Statistiken beeinflusst. Achtung: JASP spricht English und kann mit deutschen Umlauten in den Variable-Namen nicht umgehen, sie werden durch einen Punkt ersetzt. Man sollte dies also vorher in seiner Datentabelle angepasst haben, z. B. das Umschreiben der Umlaute in „ae“, „ue“, „oe“ usw. Für die weitere Bedienung verweise ich auf gute Einführungen auf Youtube, die man mit dem Suchwort „JASP“ schnell findet. Bitte praktizieren Sie das so Erlernte, indem Sie einmal mit unserem Datensatz „Geldboersen“ üben.

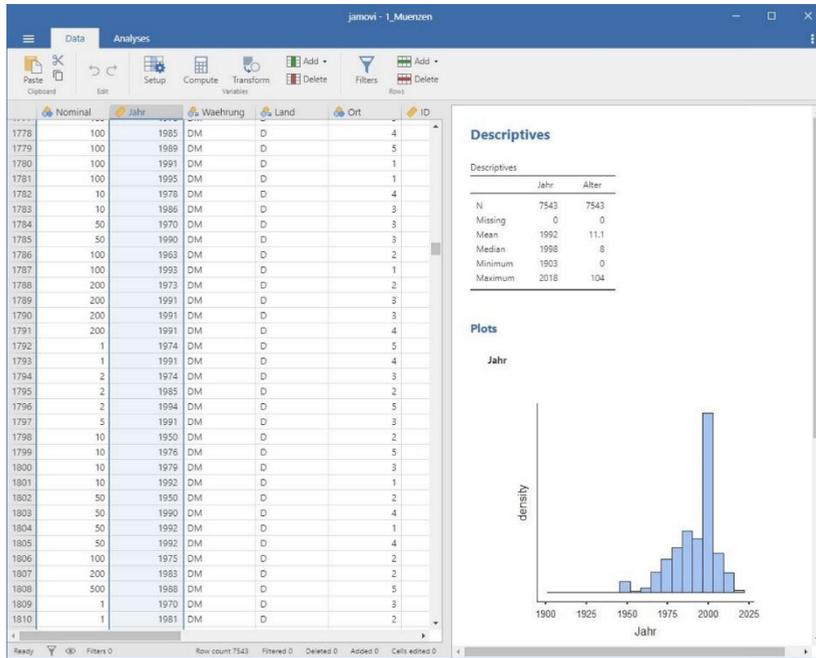
### 14.2.2 JAMOVI: Installation und erste Schritte

Das Programm JAMOVI findet sich auf der genannten Website, es sind Versionen für Windows, MAC und Linux (Ubuntu) verfügbar.<sup>31</sup> Die Installation (wie bei JASP) erfolgt nach den heute üblichen Standards, weshalb sie hier nicht detaillierter beschrieben werden muss.

Mit Klicken auf das Programmsymbol öffnet sich der Arbeitsbildschirm. JAMOVI liest keine Excel-Tabellen (\*.xls, \*.xlsx) ein, daher sollten wir den Beispiel-Datensatz „Geldboersen“ zunächst in der Tabellenkalkulation in das \*.csv-Format („comma separated values“) exportieren. Nach dem Start meldet sich JAMOVI mit seiner leeren Arbeitsoberfläche. Das Zeichen in der linken oberen Ecke öffnet den JAMOVI-Dateimanager, über den Datensätze neu angelegt, geöffnet und gesichert werden können. Laden Sie bitte den Datensatz „Geldboersen“. JAMOVI kann im Gegensatz zu JASP mit deutschen Umlauten umgehen und löst das oben beschriebene Problem „Komma oder Semikolon als Trenner“ selbsttätig.

---

<sup>31</sup> Lesenswerte Rezension: Muenchen, Bob (2018). Jamovi for R: Easy but controversial. *R-Bloggers*, 13.2.2018. Auch: <https://r4stats.com/2018/02/13/jamovi-for-r-easy-but-controversial/> [24.1.2020]. Gute Einführung: Navarro, D. J. & Foxcroft, D. R. (2019). *Learning statistics with JAMOVI: a tutorial for psychology students and other beginners*. DOI 10.24384/hgc3-7p15 Open Access: <https://sites.google.com/brookes.ac.uk/learning-stats-with-jamovi> [24.1.2020].

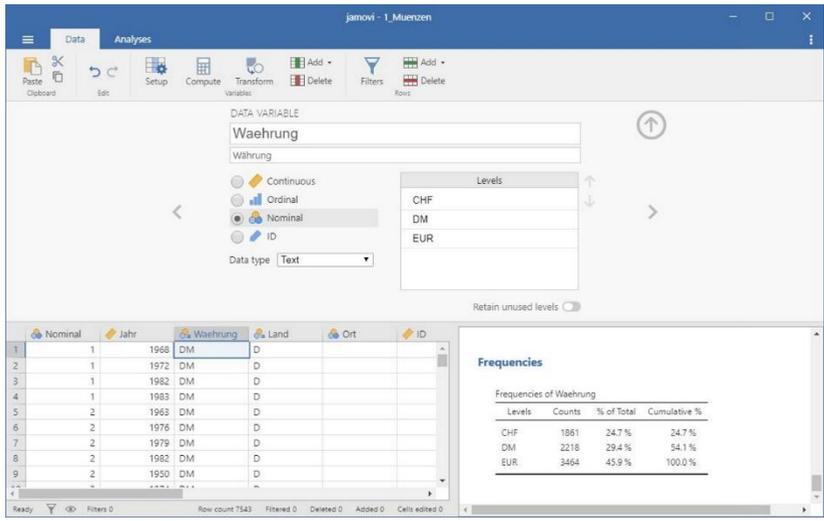


**Abb. 17** JAMOVI-Arbeitsfläche mit dem geladenen Datensatz „Geldbörsen“ im Daten-Fenster links und einfachen Statistiken und einem Histogramm im Ergebnis-Fenster rechts.

Die erfassten Münzen sollten nun im linken Bereich der JAMOVI-Oberfläche aufgeführt sein (Abb. 17, linke Seite), womit schon alles Nötige erreicht ist, um mit den Auswertungen zu beginnen: => Reiter „Analyses“; => Exploration; => Descriptives; Variablen „Prägejahr“ und „Alter“ vom Fenster links in das Fenster „Variables“ (rechts davon) schieben (auf die Variable(n) klicken, sie dadurch markieren, dann Pfeiltaste rechts drücken, ...). Fertig. Noch während der Befehlseingabe tauchen in der rechten Hälfte der JAMOVI-Oberfläche die Ergebnisse auf: Einfache deskriptive Statistiken der Daten, nämlich Anzahl der Fälle (n), Anzahl der fehlenden Beobachtungen (Missing), Mittelwert (Mean), Median, Minimum und Maximum.

Auf der linken Seite der Oberfläche kann man unterhalb der Variablenauswahl die Leiste „Statistics“ ausklappen (ggf. wiederholen: => Analyses; => Exploration; => Descriptives) und weitere Statistiken bestellen (was wir jetzt nicht tun wollen) und die Leiste „Plots“ ausklappen und dort Grafiken bestellen. Untersuchen wir „Plots“: => Häkchen an „Histogram“, und JAMOVI legt im Ergebnisfenster ein Bal-

kendiagramm für Prägejahr und Nominal an (Abb. 17, rechte Seite). Das war's, so einfach und intuitiv ist die Bedienung von JAMOVI. Allerdings auch eingeschränkt, denn die Ergebnisse lassen sich kaum anpassen. Der naheliegende Wunsch, beispielsweise, die Beschriftung der y-Achse des Balkendiagramms anzupassen und das englische „Counts“ durch „Häufigkeit“ oder „Anzahl“ zu ersetzen, lässt sich innerhalb von JAMOVI nicht umsetzen.



**Abb. 18** JAMOVI: Oberhalb des Datenfensters (links) die Schaltfläche zur Vereinbarung der Variableneigenschaften. Rechts unten „Frequencies“ – die Häufigkeitstabelle der Währungen.

Ein paar wenige Dinge sollten wir uns noch anschauen. Die Datenansicht links zeigt oben in den Spaltenköpfen vor dem Namen jeweils ein Symbol. Es steht für das Skalenniveau dieser Variablen: „Continuous“ für quantitative Skala, „Ordinal“ für Rangskala, „Nominal“ für Nominalskala und „ID“ für die Fallkennung. Mit einem Klick auf einen Variablennamen öffnet sich nach oben hin ein Menü (Abb. 18), das einerseits näher über die Variable informiert, vor allem aber – so grundsätzlich möglich – eine Änderung des Skalenniveaus erlaubt. Bei unseren Münzwerten z. B. wurde beim Einlesen seitens JAMOVI das Skalenniveau auf „Nominalskala“ gesetzt. Das ist inhaltlich nicht falsch, aber für manche Fragestellungen kann es Sinn machen, auf Rang- oder metrische Skala umzustellen. Denn: JAMOVI lässt nur jene Statistiken und Grafiken zu, die zum jeweiligen Skalenniveau sinnvoll sind. Daher kann ein gelegentlicher Wechsel des Skalenniveaus nötig sein. Im beschriebenen Menü findet man unter den

Skalenniveaus die Auswahlzeile „Data type“ mit den Optionen „Integer“, „Decimal“ und „Text“. Die beiden letztgenannten verstehen sich von selbst, „integer“ meint Ganzzahlen, d. h. Zahlen ohne Werte hinter dem Komma. Auch hier kann ggf. ein nachbessern oder auch umschalten sinnvoll sein.

Nach Klicken auf den Reiter „Data“ wird oben eine Leiste aktiviert, die verschiedene Optionen enthält, die Daten zu bearbeiten, insbes. das Löschen oder Hinzufügen von einzelnen Spalten oder Zeilen. Nicht zuletzt gibt es hier die Möglichkeit, einzelne Daten direkt in der Tabelle zu bearbeiten, um z. B. Korrekturen auszuführen, oder auch Daten zu erfassen. Da sich die Daten wiederum bequem als \*.csv-File aus JAMONI heraus exportieren lassen, ist eine Weiterverarbeitung der via JAMONI eingegebenen oder veränderten Daten mit Hilfe anderer Werkzeuge problemlos möglich. Näher studieren sollte man die Option „Filters“, die nach dem Anklicken ein Fenster mit den Filtermöglichkeiten öffnet. Sehr mächtig und flexibel ist dieser Bereich (noch?) nicht, aber immerhin lassen sich dort Filter setzen und auch mehrere Filter kombinieren. Die Filter selbst werden in der Art von R-Code geschrieben. So bewirkt z. B. die Eingabe von „Jahr > 1948“ ein Herausfiltern aller Münzen ab 1949 (Jahr der Währungsreform in Deutschland), was in unserem Datensatz wegen des Wegfalls der sehr wenigen älteren Münzen z. B. die Histogramme verbessert. Entsprechend der R-Gepflogenheiten wird die Relation „ist gleich“ als „==“ geschrieben, also mit zwei hintereinander gesetzten Gleichheitszeichen. Das Ergebnis des Filterns wird im linken Fenster mit der Datentabelle angezeigt – die Symbolik mit grünen Häkchen und roten Kreuzen erklärt sich selbst.

Werfen wir noch einen kurzen Blick auf die kleine Reihe von drei senkrechten Punkten ganz rechts in der obersten, blauen Leiste. Ein Maus-Klick dort öffnet eine Schaltfläche, wo man (ca. zwischen erstem und zweitem Drittel des Fensters) ein Häkchen an „Syntax Mode“ setzen kann. Nun erscheinen die zuvor bestellten Statistiken zusätzlich auch als R-Skript, d. h. es wird eingespielt, welchen R-Code die grafische Benutzeroberfläche von JAMONI tatsächlich erzeugt und als R-Skript an den Programmhintergrund weitergeleitet hat. Sprich: Sie sehen hier die bestellte Aufgabe in der Skriptsprache von **R**. Hingewiesen werden soll noch auf den Bereich „Plots“ dort, der es ermöglicht, ein anderes „theme“ und eine andere Farbpalette auszuwählen – weiteres erfährt man am leichtesten durch Ausprobieren. Genug JAMONI für's Erste, oder besser: viel Spaß beim selbstständigen Ausprobieren und Weiterarbeiten!

### 14.2.3 Bilanz JASP und JAMOVI

Im Vergleich ist im Zustand Anfang 2020 PAST noch das weitaus mächtigere Werkzeug. Aber vor allem JAMOVI hat viel Potenzial, denn der Einstiegsaufwand in das Programm ist gering, das Look-&-Feel modern und attraktiv. Es könnte sich, ganz wie Bob Muenchen es einschätzt, künftig zur viel verwendeten R-Benutzeroberfläche entwickeln. In den Funktionen, die JAMOVI bereits abdeckt, ist es einfacher und eleganter als das Arbeiten mit dem Paar **R** und R-Commander.

### 14.3 KURZ VORGESTELLT: SOFA, EASY R, EPI INFO UND SPSS

Der Vollständigkeit halber führen wir einige freie Programme an, auf die Interessierte stoßen könnten, deren produktive Nutzung m. E. jedoch angesichts der vorhandenen Alternativen nicht empfehlenswert ist. Gäbe es nicht BlueSky Statistics (dazu unten Kap. 15.8), könnte Easy R für Anwender, die **R** nutzen wollen, ohne sich tiefer einzuarbeiten, eine ernsthaft zu erwägende Option sein, da es in seinem Funktionsangebot etwas umfassender ist als die Kombination von **R** und R-Commander. Zudem sei kurz das mächtige Bezahlprogramm SPSS vorgestellt, das ein sehr ausgereiftes Produkt ist und für viele Anwender die nahezu perfekte All-in-One-Lösung ihres alltäglichen statistischen Arbeitens darstellt.

#### 14.3.1 SOFA – Statistics open for all

Sehr einfach in der Bedienung ist „SOFA“, die Abkürzung für „Statistics Open For All“. Das kostenlose, freie Programm ist für Windows wie für MAC und Linux (Ubuntu) verfügbar und einfach zu installieren. Die Programmoberfläche spricht auf Wunsch auch deutsch. SOFA liest Tabellen der üblichen Formate ein (u. a. \*.xlsx) und erlaubt eine Fülle an einfachen Statistiken und Berichten. Es besticht durch eine tatsächlich weitgehend selbsterklärende Bedienung und könnte vielen Anwendern für erste einfache Analysen ausreichen. Indes: Es ist nicht weiterführend, d. h. es hat weder den Funktionsumfang von PAST noch weitergehende Lerneffekte wie z. B. JAMOVI, zudem ist die Steuerbarkeit des Layouts eng begrenzt. Die Webseite von SOFA (1.5.3.), findet sich hier: <http://www.sofastatistics.com/home.php> [24.1.2020].

Die Website von SOFA zeigt hinter dem Reiter „Screenshots“ ein paar typische Bilder vom Programm, sodass Sie sehen können, was Sie erwartet. Hinter dem Reiter „Video“ findet man eine Folge von guten einführenden Videos. Probleme mit dem Player? – dann sucht man alternativ die Videos bei Youtube, indem man nach

„SOFA Statistics“ oder dem Autor „J David Eisenberg“ sucht und dann weiter stöbert. Weil wir SOFA hier nicht weiterverwenden wollen, belassen wir es bei diesem Hinweis.

### 14.3.2 EZR – Easy R

„Easy R“ (EZR) ist **R** und doch nicht **R**. Easy R ist eine kostenlose, frei verfügbare grafische Benutzeroberfläche zu **R**, die man als Plugin zum R-Commander installieren kann und die diesen erweitert und – so die Programmierer von EZR – in der Bedienung nochmals vereinfacht (Quelle: CRAN: RcmdrPlugin.EZR). Andererseits liegt EZR auch in einer Stand-alone-Fassung vor, d. h. es kann autonom ohne **R** benutzt werden. Scheinbar, denn auf Ihrem PC/Laptop müssen **R** und der R-Commander installiert sein, damit EZR im Hintergrund darauf zurückgreifen kann – aber für den Benutzer funktioniert das Programm scheinbar selbstständig. Das in Japan entwickelte Programm „spricht“ japanisch und englisch, die eigentliche Dokumentation ist nur in japanischer Sprache verfügbar. Aber ein seriös publizierter Text in englischer Sprache stellt es vor (Kanda 2013) und die Menüs funktionieren vollständig in englischer Sprache. Soweit Anfang 2020 erkennbar (aktuelle Version: 1.40), wird EZR gut weiterentwickelt, es scheint keine kurzlebige Sackgasse zu sein. Alle nötigen Angaben zum Programm und zur Installation stehen auf einer englischsprachigen Seite des „Jichi Medical University Saitama Medical Centers“ zu Verfügung:  
<http://www.jichi.ac.jp/saitama-sct/SaitamaHP.files/statmedEN.html> [24.1.2020].

Da es m. E. richtiger ist, den wichtigen R-Commander in seiner originalen Standardversion nutzen zu können, empfehle ich bei Interesse das Installieren der Stand-Alone-Version von EZR: Erst, wie später beschrieben, **R** installieren und den R-Commander, dann EZR. Das Herunterladen der Stand-Alone-Version von EZR dauert einige Zeit (ca. 140 MB), danach erfolgt die – ebenfalls nicht schnelle – Installation in der für Windows-Programme üblichen Art. Am Ende wird ein Programm-Icon auf dem Desktop abgelegt.

Das Programm öffnet sich nach einem Doppelklick auf das Icon (Geduld!, ebenfalls nicht schnell),<sup>32</sup> und zwar mit zwei Fenstern: ein mit „R Console“ beschriebenes Fenster, das man im Weiteren (a) gänzlich ignorieren kann, es (b) aber keinesfalls

---

<sup>32</sup> Die Programmautoren von Easy R mögen es mir verzeihen, aber der Hinweis auf die unerwartete Langsamkeit des Downloads, der Installation und auch des Programmstarts von Easy R sind notwendig, weil nicht vorgewarnte Benutzer denken könnten, ihr Computer sei abgestürzt. Allen Geduldigen, die diese Schritte hinter sich gebracht haben, sei versichert: Das normale Arbeiten mit Easy R geht zügig von statten, es ist ein wirklich nützliches und flinkes Werkzeug.

schließen/wegklicken darf; es muss während des Arbeitens mit den eigentlichen EZR geöffnet bleiben. Das eigentliche Arbeitsfenster von EZR ist irritierenderweise mit „R Commander“ beschriftet; es ähnelt diesem im Layout, weist aber andere Elemente und Schaltflächen auf und ist zudem in seinen Funktionen gegenüber dem R-Commander deutlich erweitert. Auf eine detaillierte Vorstellung sei verzichtet, weil man sich schnell selbst einfindet, wenn man ein paar erste Schritte im R-Commander hinter sich hat. Persönlich habe ich diese grafische Benutzeroberfläche als gut bedienbar wahrgenommen.

### 14.3.3 Epi Info

Das oben als Werkzeug für eine mögliche formularbasierte Datenerfassung empfohlene „Epi Info“ beinhaltet daneben viele wichtige Werkzeuge zur Datenverwaltung und zur statistischen Analyse, mit einer vor allem an den Belangen der Medizin und Epidemiologie orientierten Schwerpunktsetzung in der Methodenauswahl. Eine weitere Besonderheit sind integrierte Werkzeuge zur Erstellung von Kartierungen. Neben dem Schreiben und Lesen des von Epi Info bevorzugten MS-Access-Formats können z. B. auch \*.xls, \*.xlsx und \*.csv-Daten eingelesen werden, d. h. die nötige Datenkommunikationsfähigkeit ist gegeben. In die Bedienung der auf den ersten Blick etwas ungewöhnliche erscheinenden Benutzeroberfläche kann man sich schnell einarbeiten: Ein insgesamt interessantes Werkzeug, das m. W. bislang noch nicht in der Archäologie eingesetzt wurde und je nach Anwendungsfall eine erwägenswerte Alternative z. B. zu PAST sein kann.

### 14.3.4 Das Bezahlprogramm SPSS

SPSS, aufgelöst „Statistical Package for the Social Sciences“ (2020: Version 26), ist eines der großen, seit den 1960er Jahren entwickelten und sehr weit verbreiteten professionellen „Statistikpakete“ und soll zumindest kurz erwähnt werden. SPSS ist sehr umfangreich und enthält an vielen Stellen weit mehr Funktionen, als man üblicherweise braucht. Dennoch lohnt es, damit umzugehen zu lernen, denn „der Appetit kommt beim Essen“. Wer in dann gewohnter Umgebung weitere Möglichkeiten rasch zur Hand hat, wird gelegentlich mehr tun, als er zunächst vorhatte, rasch auch einmal etwas Neues ausprobieren – und auf diesem Wege vielleicht kreativ weiterkommen als geplant. Auf dem Markt dieser kommerziellen Statistikpakete gibt es andere Produkte, wie z. B. das sehr leistungsfähige Paket „SAS“. Der Charme von SPSS liegt in seiner modernen Benutzeroberfläche, einer vergleichsweise leichten Erlernbarkeit und der Tatsache, dass es – anders als z. B. PAST oder JAMOVI – in dieser Oberfläche auch viele mächtige Werkzeuge zur Datenverwaltung integriert hat. Mit SPSS kann man

sehr vieles Notwendige tun, ohne diese dann gewohnte Arbeitsumgebung je verlassen zu müssen. An praktisch allen Universitäten wird eine Campus-Lizenz vorgehalten, zumeist können Studierende und Doktoranden auch eine Jahreslizenz für ihren PC/Laptop zu vertretbaren Kosten erwerben. Aber wie eingangs beschrieben: Sobald man die Sphäre von Universitäten und Forschungseinrichtungen verlässt, wird der Zugang zu dieser Software schwierig und für den Privatgebrauch arg teuer.

In der Open-Source-Gemeinde ist ein Versuch entstanden, eine freie Alternative zu SPSS zu schaffen; sie wird PSPP genannt. Wer z. B. nach einer universitären Arbeitsphase mit SPSS keinen Zugang mehr zu SPSS hat, könnte prüfen, ob er mit PSPP weiterarbeiten möchte. Aus meiner persönlichen Sicht ist dies jedoch kein tauglicher Weg, weil PSPP nicht hinreichend umfassend ist und nicht schnell weiterentwickelt wird. Wer unter SPSS-Entzug leidet, tut besser daran, auf BlueSky umzusteigen.

#### 14.4 DATENBANKEN

Schnell ein paar Daten erfassen und dann die Statistiken abrufen: Es wäre schön, wenn es so einfach wäre. Tatsächlich nimmt nach der Ersterfassung der Daten die Prüfung, Pflege und Verwaltung recht viel Zeit in Anspruch, die jeder statistischen Analyse vorangeht und meist auch permanent begleitet. Hierzu werden oft „Datenbanken“ eingesetzt, spezielle Programme zur Erfassung, Speicherung und Verwaltung von Daten. Wir kennen sie alle aus dem Alltag: Unsere Bank und deren Online-Banking benötigen eine Datenbank, die Fahrplanauskunft und das Online-Buchungssystem der Bundesbahn und der Online-Händler unseres Vertrauens arbeiten im Hintergrund mit einer Datenbank, und auch eine Bodendenkmälerdatenbank ist eben dieses: eine Datenbank. Brauchen wir das also auch, eine Datenbank? Meistens nein. Im Grunde basieren die oben genannten Programme von der Tabellenkalkulation über PAST zu **R** auf der Idee einer sog. rechteckigen Datenmatrix: Jeder Fall ist eine Zeile und jedes Merkmal wie z. B. Name, Straße, Wohnort, Telefonnummer und E-Mailadresse bildet eine Spalte in einer Tabelle. Bis auf Fehlstellen, wo einzelne dieser Informationen fehlen, hat man zu allen Fällen ein gleiches Bündel an gleichartigen Informationen. Die zuvor genannten Beispiele sind jedoch komplexer: Eine Bank hält Informationen über ihre Kunden mit Namen und Adressen vor, ähnlich einer Adresskartei, aber auch solche über ihre Partnerbanken, zu denen Geld überwiesen wird oder von denen es hereinkommt; sie braucht ein Protokoll aller Buchungsvorgänge und die Bank unterhält Depots für Ihre Kunden, die wiederum aus unterschiedlichen Wertpapieren, Anleihen, Aktien u. ä. bestehen, zu denen es eigene Begleitinformationen gibt wie z. B. die oft

wechselnde Börsenkurse. Zudem werden die Mitarbeiter der Bank verwaltet, schließlich muss auch geregelt sein, welcher Mitarbeiter welche Daten sehen und ggf. auch bearbeiten darf. Diese Informationen verwaltet man nicht mehr gemeinsam in einer riesigen gemeinsamen Tabelle (*sensu* rechteckige Datenmatrix), sondern in vielen verschiedenen, jeweils spezifischen Tabellen, die bei Bedarf verknüpft werden können: Die Informationen über eine Aktie mit einem Depot mit der Adresse eines Kunden und wiederum mit den Daten seines Kundenberaters usw. Eine Bodendenkmälerdatenbank ist ähnlich komplex: Eine archäologische Fundstelle liegt auf mehreren Grundstücken, zu denen es wiederum Informationen über deren jeweilige Besitzer gibt, sodann Informationen zu dort geborgenen Funden und deren Verbleib, ggf. deren Ausleihe an Werkstätten oder in eine Museumsausstellung, sodann Informationen zum ehrenamtlichen Finder, und die Fundstelle selbst umfasst oft verschiedene Aktivitäten (Begehungen, Sondagen, Grabungen etc.), die über Jahre hinweg stattfanden, usw. usf.

Forschungsdaten sind jedoch in der Regel weitaus schlichter, weil exakter auf eine spezielle Fragestellung hin orientiert und zusammengestellt: Man analysiert Töpfe und deren Maße, oder Gräber und deren Beigaben. Die meisten Forschungsdaten lassen sich gut in einfachen Tabellen abspeichern. Ja, gelegentlich benötigt man Verknüpfungen, z. B. aus der Datentabelle des Anthropologen die Informationen zu Alter und Geschlecht der Toten und aus der Tabelle zu den Gräbern Informationen zu den Beigaben oder zur Grabgrubengröße. Aber solche Informationen lassen sich mit Hilfe der genannten Statistikpakete wie z. B. **R** auch aus zwei oder drei unterschiedlichen Tabellen miteinander verknüpfen. Kurz: Ist das Verknüpfen unterschiedlicher Datentabellen eine häufig benötigte Daueraufgabe, wie es das z. B. bei Banken oder auch Bodendenkmälerdatenbanken der Fall ist, kommt man um eine echte Datenbanksoftware nicht herum. Benötigt man solche Verknüpfungen eher selten oder gelegentlich, kann man (und sollte man) auf eine Datenbanksoftware verzichten. Im Forschungsalltag stehen meist die Informationen aus einer einzelnen Tabelle im Vordergrund und die tendenziell selten benötigten Verknüpfungen lassen sich z. B. mit Hilfe von **R** erzeugen und verwalten.

Ein Spezialfall liegt m. E. dann vor, wenn man selbst mehr Unterstützung für die Erfassung benötigt und man Datenprüfungen gleich während der Datenerfassung vornehmen möchte – also eine Maske zur Datenerfassung benötigt. Richtig programmiert, kann man dabei z. B. unsinnige Eingaben gleich ausschließen (z. B. Buchstaben verhindern, wo nur Zahlen eingegeben werden dürfen) und die Logik zwischen verschiedenen Eingaben noch während der Eingabe abprüfen. So werden bei einem generell großen Skelett tendenziell alle Langknochen eher lang sein, weshalb ein auffällig

kleiner Wert eher auf einen Mess- oder Tippfehler zurückgehen dürfte als auf einen einzelnen, tatsächlich vergleichsweise kurz geratenen Knochen. Datenbanken ermöglichen es, Eingabemasken zu programmieren, die solche Datenprüfungen gleich bei der Erfassung vornehmen. Doch lohnt sich das? In der Regel ist der zusätzliche Aufwand für das Erstellen solcher Eingabemasken recht hoch (sprich: zu hoch) und eine Prüfung der Korrektheit der Daten auch im Nachhinein noch auf andere Weise möglich. Bis auf seltene Ausnahmen sind bei den üblichen Projekten solche Eingabemasken und der daraus resultierende Einsatz einer Datenbanksoftware nicht sinnvoll.

Wer sich trotz meiner persönlichen Skepsis solche Datenbanken und Eingabemasken in der Praxis ansehen möchte, sei auf das gute und anregende System für die Erfassung von Silices,<sup>33</sup> das aus dem Kölner SAP-Projekt hervorgegangen ist, und auf eine Datenbank für die Erfassung von Megalithgräbern hingewiesen.<sup>34</sup>

Kurz: Obwohl das Prüfen und Verwalten der Daten eine regelmäßige Aufgabe während des Forschungsprozesses ist, die mehr Zeit in Anspruch nimmt, als man anfangs ahnt, empfiehlt sich der Einsatz spezieller Datenbankprogramme nur in Ausnahmefällen. Fast alles Nötige lässt sich mit einer Tabellenkalkulation oder mit **R** (oder kommerziellen Paketen wie SPSS) gut bewältigen. Daher sind Datenbanken im Folgenden kein Thema mehr.

## 15 R ALS ARBEITSUMGEBUNG

---

Das seit ca. 1995 entwickelte Statistikprogramm **R** ist kostenlos und „Open Source“, d. h. der Quellcode ist für alle Interessierten frei einsehbar. Es liegt im Jahr 2020 in der Version 3.6.1 vor und hat sich spätestens im Laufe der 2010er Jahre zu einem (genauer: dem) zentralen Werkzeug professioneller Statistiker entwickelt. Fertig kom-

---

<sup>33</sup> Drafehn, A., Bradtmöller, M. & Mischka, D. (2008). SDS – Systematische und digitale Erfassung von Steinartefakten (Arbeitsstand SDS 8.05). *Journal of Neolithic Archaeology*, 10 (23.9.2008): <https://doi.org/10.12766/jna.2008.25> sowie ergänzend: Drafehn, A. (2014). SDS – Systematic, digital collection of data sets of stone artefacts - Supplement K-list, vers. 1.0.12.11 (Classification and denomination of projectiles and inset-blades), list of accordance of the west and central European forms. *Journal of Neolithic Archaeology*, 16 (14.4.2014): <https://doi.org/10.12766/2014.2>

<sup>34</sup> Furholt, M. Mischka, D., Rassmann, K. & Schafferer, G. (2010). MegaForm – Ein Formalisierungssystem für die Analyse monumentaler Baustrukturen des Neolithikums im nördlichen Mitteleuropa. *Journal of Neolithic Archaeology*, 12 (25.11.2010): <https://doi.org/10.12766/jna.2010.60>

pilierte Programmversionen werden jeweils für Windows, Linux und Mac-OS bereitgestellt. In **R** kann man typische und häufiger gebrauchte Arbeitsabläufe als „Skripte“ speichern und später wiederbenutzen; folglich können R-Skripte auch wie ein Computerprogramm publiziert, an Dritte weitergeben und von diesen ohne Lizenzprobleme genutzt werden. Heute werden neue statistische Prozeduren meist in Fachzeitschriften theoretisch begründet und erläutert und zugleich als R-Skript bereitgestellt, damit Interessierte die neuen Verfahren sogleich prüfen und anwenden zu können. Fertige, veröffentlichte Skripte – bei **R** „Pakete“ genannt – werden in einigen zentralen Depots bereitgestellt, wo sie offen erreichbar sind. Das wichtigste Depot ist CRAN „The Comprehensive **R** Archive Network“: <https://cran.r-project.org/> [24.1.2020]. Ein weiteres viel verwendetes Repositorium für R-Pakete ist Bioconductor: Open Source Software for Bioinformatics (<https://www.bioconductor.org/> [24.1.2020]), wo u. a. Werkzeuge für aDNA-Auswertungen deponiert werden.

In seiner Bedienung unterscheidet sich die Software **R** in vielfältiger Weise von den heute gewohnten Programmen, was gleich eingangs eine Rolle spielt. Im Kern arbeitet **R** ohne eine grafische Benutzeroberfläche, sondern wird durch Texteingaben in eine sog. Kommandozeile gesteuert, mit denen man Daten und Befehle eingibt. Wir werden uns dieses „rohe“ **R** noch anschauen. Die meisten Nutzer installieren zusätzlich RStudio, was die Bedienbarkeit des rohen **R** wesentlich erleichtert. Zwar gibt es zahlreiche Alternativen zu RStudio wie z. B. JGR, RKWard, StatET/Eclipse, Rattle GUI, aber für sehr viele R-Nutzer bildet die Kombination aus **R** und RStudio ihre Standardarbeitsumgebung. Wir wollen es hier ebenso halten, also dem Mainstream folgen.

Mit **R** & RStudio hat man eine leistungsfähige Grundversorgung erreicht, die vielen Nutzern ausreicht. Der besondere Reichtum von **R** besteht jedoch darin, dass man zusätzlich viele – sehr viele – „Pakete“ installieren kann für spezielle Aufgaben, die nicht mit der Basisausstattung abgedeckt sind. Dies geschieht in zwei Schritten: (a) ein Paket wird aus einem der Programmdepots („Repositorium“) heruntergeladen und auf dem PC/Laptop installiert, (b) das Paket wird zum laufenden Betrieb von **R** hinzugeladen. Beides geschieht am einfachsten innerhalb von **R**, d. h. aus der bereits laufenden Basisinstallation heraus. Während der erste Schritt (a) nur beim ersten Mal nötig ist, ist (b) das Aktivieren der Zusatzpakete in den laufenden Betrieb von **R** hinein nach jedem Programmstart von **R** wieder neu nötig. Keine Sorge: das geht einfach und schnell.

## 15.1 GRAPHISCHE BENUTZEROBERFLÄCHEN (GUIs) ZU R

Für geübte Benutzer, die die Komponenten und Befehle von **R** gut kennen, ist das Arbeiten mit der direkten Befehlseingabe sehr effizient, für Anfänger bedeutet diese Art der Bedienung jedoch ein längeres Lernen, bevor man ins eigentliche inhaltliche Arbeiten eintreten kann. Daher wurden verschiedene grafische Benutzeroberflächen (*graphic user interface*, abgekürzt GUI) geschaffen, welche für weniger geübte Anwender das Arbeiten mit **R** vereinfachen. Entweder man bleibt auch längerfristig bei dieser Art der etwas bequemereren Bedienung, oder man benutzt sie für den Einstieg, um schneller ins produktive Arbeiten eintreten zu können, und geht später schleichend in das Arbeiten via Kommandozeile über. Letzteres ist der Weg, den wir hier verfolgen. Der Nachteil dieser GUIs ist, dass sie nicht den vollen Leistungsumfang von **R** anbieten, sondern lediglich (unterschiedliche) Basisfunktionalitäten und -pakete umfassen. Für Spezialisten ist es demnach notwendig, beizeiten auch ohne eine GUI mit **R** arbeiten zu können.

Es gibt mehrere Projekte für solche grafischen Benutzeroberflächen zu **R**. In dieser Einführung wird vor allem der „R-Commander“ (auch: „RCommander“, abgekürzt Rcmdr) benutzt, der seit ca. 2003/2005 verfügbar ist und weiterhin rege von einer relativ großen Gemeinde entwickelt wird. Im folgenden Abschnitt werden erst das unveränderte, „rohe“ **R** und dann RStudio eingeführt, danach als GUI der R-Commander. Das konkrete inhaltliche Arbeiten mit **R** wird anschließend vor allem mit Hilfe des R-Commanders geschehen.

Als Alternative zum R-Commander ist „Deducer“ einer Erwähnung wert, eine jüngere, seit ca. 2010 verfügbare grafische Oberfläche. Deducer versucht noch stärker, eine per Menüauswahl und Klick-Bedienung steuerbare Oberfläche zu sein, es will erklärtermaßen in der Einfachheit der Bedienung dem Programm SPSS ähneln. Deducer legt seinen Schwerpunkt auf statistische Grafiken und ummantelt das moderne **R** Grafikpaket „ggplot2“. Auf den ersten Blick (sprich: in den Demo-Videos auf YouTube) wirkt dieser Ansatz sehr interessant. Neugierige seien ermuntert, Deducer selbst auszuloten resp. seine weitere Entwicklung zumindest im Blick zu behalten. Der Verfasser hat sich damit nicht anfreunden können: Die Installation ist tricky und erfordert ein problemloses Interagieren mit einer JAVA-Laufzeitumgebung – hier liest man in vielen Foren von Problemen. Deducer erfordert zunächst das Einrichten der Benutzeroberfläche JGR (als „Jaguar“ aussprechen), die jedoch ansonsten dem RStudio deutlich unterlegen ist. Die Grafiken sind dann mit Deducer schnell erstellt, bieten aber nach dem ersten Einrichten keine komfortable Möglichkeit zum (nötigen) Feinschliff per Point-and-Click – und die händische Nacharbeit im Paar JGR & Deducer ist dann nicht einfacher als via RStudio & R-Commander.

Die erst Mitte der 2010er Jahre aufgekommene Benutzeroberfläche BlueSky Statistics, die unten näher vorgestellt wird, verfolgt ähnliche Ziele wie Deducer und erreicht diese einfacher und problemloser. Der Einsatz von BlueSky kann m. E. ernsthaft als Alternative zum R-Commander erwogen werden (s. unten Kap. 15.8).

## 15.2 DOKUMENTATION UND ALLGEMEINE HILFE ZU R

**R** selbst und viele seiner Ergänzungen stehen unter einer offenen Lizenz auf CRAN bereit. Um den internationalen Datenverkehr nicht unnötig zu belasten, wird CRAN an vielen Orten der Welt gespiegelt, d. h. es stehen dort aktuelle Kopien von CRAN bereit. Daher fragt **R** den Nutzer gelegentlich, welches Repositorium er ansteuern möchte. Die Idee dahinter ist, dass die Nutzer auf eine ihnen räumlich möglichst nahe gelegene CRAN-Kopie zurückgreifen.

**R** wird inzwischen weltweit und von einer sehr großen Nutzergemeinde eingesetzt, daher bieten Google und YouTube schnell Hilfe, sofern man das geeignete Suchwort kennt. Sodann bietet **R** selbst eine integrierte Hilfe, die wir zusammen mit der Basisinstallation anschauen werden. Spezielle Bücher werde ich fallweise nennen. Die eigentliche Dokumentation, die zu jedem Paket von **R** dazugehört und Open Access bereitsteht, ist eher spröde und knapp; sie kann Anfängern nicht empfohlen werden. Diese Dokumentationen sind wichtig, wenn es um die inhaltlichen Spezifika der Pakete geht, also die statistischen Aspekte, helfen aber kaum bei allgemeineren Fragen zur Bedienung. Zu erwähnen sind – eher für Fortgeschrittene – die im Open Access publizierte R-eigene Fachzeitschrift „The R Journal“ (<https://journal.r-project.org/>) sowie – sehr nützlich auch für Anfänger – die Website „R-bloggers“ (<https://www.r-bloggers.com/>), zu der mehr als 750 R-Nutzer beitragen; R-bloggers liefert neben vielen Tipps und Tricks auch einen gut lesbaren Überblick über wichtige Neuentwicklungen, es ist eine gute Quelle, um „auf dem Laufenden zu bleiben“.

Ein feines, um 2013 erstmals vorgestelltes Werkzeug, das in diesen Kontext gehört, ist „Swirl“, ein in **R** eingebetteter R-Kurs (<http://swirlstats.com/> [24.1.2020]). Swirl soll nach dem Prinzip „learning by doing“ **R** vermitteln; schauen Sie ggf. selbst unter der genannten Adresse nach.

Da **R** inzwischen sehr breit weltweit in Statistikkursen eingesetzt wird, gibt es zahlreiche Videos auf YouTube, die **R** resp. einzelne Aspekte von **R** erklären; das Material ist leicht zu finden, schwer überschaubar, unterschiedlich lang und intensiv und von sehr heterogener Qualität. Auf jeden Fall nützlich, um erste Eindrücke und Anregungen zu erhalten; für mich persönlich für ein systematisches Arbeiten weniger geeignet

– weshalb sich am Ende dieser Einführung eine kommentierte Literaturliste zu **R** findet, denn gute Bücher sind m. E. weiterhin der beste Weg, sich in dieses etwas schwierigere Feld einzuarbeiten.

Zu den mit **R** mitgelieferten Hilfe-Angeboten siehe unten den Abschnitt „Die R-interne Hilfe“.

### 15.3 R INSTALLIEREN

Besuchen Sie das offizielle R-Repositorium <https://cran.r-project.org/> und wählen dort das für Ihr Betriebssystem passende **R**, z. B. Windows 64-bit. Installieren Sie das Programm nach den Angaben auf dieser Website. Ein typischer Weg über die genannte Website ist: => Download R for Windows; => base; => install R for the first time; => Download R 3.6.1 for Windows (62 megabytes, 32/64 bit); dann lädt das Programm herunter. Dann in den Download-Ordner des eigenen PC/Laptop gehen und dort auf das neu sichtbare Icon des Installers klicken („R-3.6.1.win“) und die weitere Installation ablaufen lassen. Die Installation erfolgt weitgehend automatisch, wie man es heute auch von anderen Programmen her gewohnt ist; die Standardinstallation von **R** benötigt ca. 200 MB Platz. Es empfiehlt sich, die Voreinstellungen, die bei der Installation vorgeschlagen werden, anzunehmen.

Allerdings: Der R-Commander sagt von sich selbst, dass er besser funktioniert, wenn **R** bei der Installation mit der Option „SDI“ eingerichtet werde. Wer also den R-Commander einzusetzen gedenkt, sollte die entsprechende Auswahlfrage bei der Installation von **R** mit „SDI“ beantworten. (Kein Stress, man kann das Passende auch nachträglich in **R** einstellen). Ansonsten verzichten wir hier und im Folgenden darauf, die Installation von **R** und weiteren Komponenten ausführlicher vorzustellen. Einerseits, weil dies erfahrungsgemäß nur in Ausnahmefällen nötig ist, und andererseits, weil ggf. eine kurze Websuche zu unzähligen Skripten und insbesondere YouTube-Clips führt, die dies tun.

### 15.4 ERSTE SCHRITTE IN R

Gerade Anfängern empfehle ich, **R** mit Hilfe einer grafischen Benutzeroberfläche zu bedienen, z. B. mit dem R-Commander. Die weiteren Schritte im Dreikampf Statistik, **R** und Archäologie werden wir in diesem Buch mit dem R-Commander (oder mit BlueSky) gehen und die Daten aus einer Tabellenkalkulation einlesen, d. h. **R** nicht wie im jetzt folgenden Abschnitt bedienen. Dennoch ist es für den Anfang sinnvoll, **R** einmal in seinem „rohen“ Modus kennen zu lernen, weil man so das Besondere der

Arbeitsweise und ein paar grundlegende Begriffe und Konzepte versteht und die Hilfen und die Literatur besser nachvollziehen kann.

#### 15.4.1 Rechnen mit R, mathematische und logische Funktionen

Mit einem Doppelklick auf das Programmsymbol startet **R** und es öffnet sich die R-Console. In der Standardeinstellung ist das ein weiß hinterlegtes Feld mit blauer Schrift, in dem unten ein „Prompt“, die Eingabeaufforderung in roter Schrift steht. Wenn Sie dort einen Befehl eingeben und mit ENTER dessen Ausführung bestellen, geschieht dies unmittelbar und **R** gibt das Ergebnis Ihrer Anforderung aus. Tippen Sie z. B. „2 + 4“ & ENTER, erscheint in der Folgezeile in blauer Schrift „[1] 6“: 2 plus 4 gleich 6. Auf diese Weise lässt sich mit **R** rechnen in der Notation, wie man sie auf der Schule gelernt hat, d. h. mit den üblichen Zeichen für Plus (+), Minus (-), Multiplikation (\*) und Division (/) sowie Potenzieren (^). Dabei gilt die übliche Regel „Punktrechnung geht vor Strichrechnung“, weshalb „1 + 3 / 2“ 2,5 ergibt, nicht 2,0. Meint man die Aufgabe anders, muss man dies durch Klammersetzung klären, z. B. als „(1 + 3) / 2“, was 2,0 ergibt.

Außer den vier Grundrechenarten, dem Potenzieren und dem Setzen von Klammern stehen ein paar häufig benötigte mathematische Funktionen bereit. So ergibt z. B. die Eingabe von „sqrt(2)“ – d. h. Quadratwurzel aus 2 – die Ausgabe „1.414214“. **Abb. 19** zeigt die Rechen-Operatoren von **R**, **Abb. 20** und **Abb. 21** zeigen die mathematischen und logischen Funktionen von **R**.

Funktion	Aufgabe
+	plus; addieren
-	minus; subtrahieren
*	mal; multiplizieren
/	durch; dividieren
%/%	ganzzahliges dividieren, d.h. ohne Rest
%%	Modulo Division, gibt den Rest einer Ganzzahldivision aus
^ (oder) **	potenzieren; x hoch y

**Abb. 19** Rechnen mit R.

Funktion	Aufgabe
<	kleiner als
<=	kleiner oder gleich
>	größer als
>=	größer oder gleich
==	ist gleich (identisch)
!=	nicht gleich; ungleich
!x	nicht x
x   y	x oder y
x & y	x und y
isTRUE(x)	testet, ob x wahr ist

**Abb. 20** Logische Funktionen in R. Hinweis: Das Zeichen „|“ für „oder“ findet sich auf deutschen Tastaturen meist mit „AltGr“ auf der Taste <>.

Funktion	Aufgabe
sqrt(x)	Quadratwurzel; sqrt(25) gibt 5 aus; identisch: 25^(1/2)
abs(x)	Absoluter Betrag; abs(-12) gibt 12 aus
round(x)	Runden auf nächste Ganzzahl nach den math. Rundungsregeln round(3.14159) gibt 3 aus round(3.51514) gibt 4 aus
round(x, digits=n)	Runden nach den math. Rundungsregeln auf spezifizierte Zahl v. Nachkommastellen round(3.14159, digits=2) gibt 3.14 aus
floor(x)	Abrunden auf nächste Ganzzahl; floor(3.14159) gibt 3 aus
ceiling(x)	Aufrunden auf nächste Ganzzahl; ceiling(3.14159) gibt 4 aus
trunc(x)	Tranchieren, Nachkommastellen abschneiden
log(x)	Natürlicher Logarithmus (log <sub>e</sub> ; ln)
log10(x)	Logarithmus zur Basis 10 (log <sub>10</sub> )
Log(x, base=n)	Logarithmus x zur Basis n
exp(x)	Exponentialfunktion
factorial(x)	Fakultät
choose(n,k)	Binomialkoeffizient (n über k)
sin(), cos(), tan()	Winkelfunktionen Sinus, Cosinus, Tangens
asin(), acos(), atan()	Arcus-Winkelfunktionen

**Abb. 21** Mathematische Funktionen in R.

## 15.4.2 Beispiel: Logarithmus rechnen & „Umkehrfunktion“

Um diese Möglichkeiten kennenzulernen, möchten wir den Logarithmus zur Basis 10 für die Zahlen 1, 10, 100 und 1.000 berechnen lassen. Folglich geben wir in der Kommandozeile ein:

```
> log10(1); log10(10); log10(100); log10(1000)
```

Dabei ist das rote „>“ von alleine da und muss nicht eingegeben werden. Es ist das „Prompt“ von **R** und signalisiert „hallo, ich bin arbeitsbereit und warte auf eine Eingabe“. Das Semikolon „;“ wirkt als Trennzeichen; so kann man in einer Zeile mehr als eine Berechnung bestellen. Nach ENTER sehen wir das Ergebnis:

```
[1] 0  
[1] 1  
[1] 2  
[1] 3
```

In vier Zeilen erhalten wir die Werte null, eins, zwei und drei – und so haben wir nebenbei erfahren, was eine Logarithmuskala bedeutet: (sehr) hohe Werte werden relativ stärker verkleinert als niedrige Werte.

Wir müssen noch lernen, was bei „verbotenen“ Operationen geschieht. Dazu geben Sie bitte ein:

```
> log10(-1)
```

Nach ENTER erfolgt die Meldung:

```
[1] NaN  
Warnmeldung:  
NaNs wurden erzeugt
```

Logarithmen von negativen Zahlen sind mathematisch nicht definiert, daher wird das Ergebnis „NaN“ für „Not a Number“ ausgegeben, die R-Notation für einen fehlenden Wert.

### 15.4.2.1 Stichwort „Rücktransformation“ oder „Umkehrfunktion“

Auch wenn wir das Logarithmieren hier nur als kleines Anschauungsbeispiel eingeführt haben: solche Variablentransformationen sind in der Statistik wichtig. Nicht selten muss man transformierte Variablen später wieder „rücktransformieren“. Zu jeder Transformation gibt es eine Umkehrfunktion, die wieder in den ursprünglichen Stand zurückführt. Die folgende Tabelle (**Abb. 22**) gibt ein paar Beispiele.

Transformation	Rücktransformation
Wurzel ziehen $\text{sqrt}(x)$ ; oder: $x^{1/2}$	Potenzieren ( $x$ hoch 2) $x^{**2}$
Kubikwurzel ziehen $x^{1/3}$	dritte Potenz berechnen $x^{**3}$
Logarithmus zur Basis 10 ( $\log_{10}$ ) $\log_{10}(x)$	10 hoch $x$ $10^{**x}$ (oder: $10^x$ )
Logarithmus zur Basis $e$ , auch „Logarithmus naturalis“ genannt ( $\ln_x$ )	$e$ hoch $x$ , oder: Exponentialfunktion( $x$ ) $\exp(x)$

**Abb. 22** Einige Transformationen und ihre Rücktransformation ("Umkehrfunktion"). „ $x$ “ steht für die jeweils frei einsetzbare Zahl.

Probieren Sie diese Paare doch einmal aus. So ergibt z. B.

```
> log(300)
[1] 5.703782
```

d. h. der natürliche Logarithmus von 300 hat den Wert (circa) 5,7. Nun die Rücktransformation:

```
> exp(5.7)
[1] 298.8674
```

d. h. „fast“ 300, weil wir schreibfaul nicht den exakten Wert eingetragen haben.

Für ein Keramikgefäß mit einem Fassungsvermögen von 2,5 Litern = 2.500 cm<sup>3</sup> berechnen wir über die Kubikwurzel:

```
> 2500^(1/3)
[1] 13.57209
```

dass sein Fassungsvermögen einem Würfel von ca. 13,6 cm Kantenlänge entspricht, usw.

### 15.4.3 Zuweisungen in R: Objekte und Objektnamen

Wir haben bisher allein mit der Konsole gearbeitet: Die Ergebnisse der Berechnungen wurden ausgegeben, sind aber (außer auf dem Bildschirm) nicht weiter verfügbar. Wenn man die Ergebnisse erhalten will, um mit ihnen weiterzuarbeiten, muss man neue Variablen bilden und diesen das Ergebnis der Berechnungen zuweisen. Warum das sehr nützlich ist, zeigt das Beispiel BMI wenige Zeilen später, doch erst einmal schauen wir, wie solch eine Wertzuweisung vorgenommen wird. Sie geschieht in **R** traditionell mit den beiden Zeichen „<-“, neu auch mit dem weitaus vertrauenswürdigeren Gleichheitszeichen „=“. Ich folge hier jedoch dem R-Mainstream und verwende das Zuweisungszeichen „<-“.

```
> A <- log10(100)
```

Nach einem ENTER hat **R** den Logarithmus zur Basis 10 für die Zahl 100 berechnet, und wartet auf die nächste Eingabe. Schreiben Sie jetzt:

```
> A
```

antwortet **R** mit dem Wert, den die (neue) Variable A aktuell hat, nämlich:

```
[1] 2
```

Sehr wichtig ist, dass **R** „case sensitive“ ist, d. h. zwischen Klein- und Großschreibung unterscheidet: „a“ ist nicht gleich „A“.

Die Namen von Variablen (Objekten) können:

- beliebig lang sein;
- aus Buchstaben, Ziffern und Punkten bestehen (z. B. „Logarithmus.Basis.10“ ist erlaubt);
- und müssen, wenn sie mit einem Punkt anfangen, zumindest auch Buchstaben beinhalten.

Die Freiheit, die **R** bei der Namengebung von Variablen einräumt, ist Fluch und Segen zugleich. Praktiker empfehlen, sich zwecks besserer Lesbarkeit seines Codes sehr zu disziplinieren und Variablennamen eindeutig und eher kurz zu halten und vor allem eine stets gleiche Schreibweise zu verfolgen. Bei Wortzusammensetzungen sind in der R-Szene drei verschiedene Modi üblich: „*camel case*“ (Binnenmajuskel) mit Kleinbuchstaben am Anfang und Großbuchstaben als Zeichen für den Beginn des zweiten Elements (z.B. „logBasis10“), die Binnengliederung mit Punkten, z. B. „log.basis.10“, und „*snake case*“, die Binnengliederung mit Unterstrichen, also z. B. „log\_basis\_10“. Da man als Anwender außer für Variablen auch Namen für Funktionen vergibt (was das ist, lernen wir sogleich), könnte es sinnvoll sein, für Variablennamen ggf. stets den Punkt zu verwenden und für Funktionsnamen stets den Unterstrich.

Mit den bislang erlernten Instrumenten können wir schnell mal einen BMI (Body Mass Index) berechnen für eine Person, die 1,8 m groß ist und 90 kg wiegt:

```
> BMI <- 90 / (1.80^2)
```

und rufen das Ergebnis ab:

```
> BMI  
[1] 27.77778
```

Etwas eleganter und vor allem wiederverwendbarer ist folgendes Vorgehen:

```
> Körperhöhe <- 1.80  
> Gewicht <- 90  
> BMI <- Gewicht/Körperhöhe^2  
> BMI
```

[1] 27.77778

In diesem Fall kann die Wertezuweisung „BMI“ auch auf eine ganze Zahlenreihe (in der Sprache von **R**: einen Vektor) oder auf eine Datentabelle angewendet werden, wenn dort die Variablen „Körperhöhe“ und „Gewicht“ enthalten sind.

Mit dem Befehl

```
> ls()
[1] "BMI"      "Gewicht"   "Körperhöhe"
```

erfahren wir, dass im Gedächtnis von **R** nun die Variablen Körperhöhe, Gewicht und BMI vorhanden sind. Dadurch, dass wir nicht nur gerechnet haben, sondern Wertzuweisungen vorgenommen haben, ist der errechnete Wert nun als „BMI“ zur Weiterverwendung verfügbar, z. B. um eine Mittelwertberechnung durchzuführen, ein Histogramm zu erzeugen oder ein Streudiagramm. Klar, das macht hier bei nur einem Wert keinen Sinn, aber wir werden in wenigen Schritten sehen, wie man ganze Zahlenreihen so bereithalten kann.

#### 15.4.4 R-Kommandozeilen

Das, was wir hier kurz angeschnitten und eingeübt haben, ist die Bedienung des rohen **R** per Kommandozeile. Zwei Dinge müssen noch vermittelt werden, um die Eingaben und Ausgaben besser lesen zu können:

- Eine Kommandozeileingabe ist im Grunde als eine Zeile gedacht. Mit jedem ENTER endet eine solche Zeile und die Ausführung des Kommandos wird aktiviert. Aber nicht immer passt so viel Text in eine Zeile, zudem neigen allzu lange Zeilen dazu, unübersichtlich zu werden. Daher können Zeilen in **R** gebrochen und in einer Folgezeile fortgeführt werden. Das Fortführungszeichen in ein „+“ am Beginn der Folgezeile. Beginnt eine Zeile mit einem +-Zeichen, wird sie durch **R** als Fortsetzung der vorangehenden Zeile verstanden.
- Wenn man fremden R-Code liest, begegnet nicht selten ein Erläuterungstext innerhalb des Codes. Dahinter steckt die Konvention, dass man (1) längeren Code kommentiert – auch für sich selbst – und (2) alles, was dem #-Zeichen folgt, Kommentar und kein auszuführender Code mehr ist. Etwas eleganter und R-mäßiger hätte das letztgenannte Beispiel also auch so aussehen können:

```
> Körperhöhe <- 1.80 # Setzen der Variable Körperhöhe auf 1.80 m
> Gewicht <- 90      # Setzen der Variable Gewicht auf 90 kg
> BMI <- Gewicht / Körperhöhe^2 # Definieren der Zuweisung BMI
> BMI
[1] 27.77778
```

Dabei ist nur ein Abstand von einem Leerzeichen zwischen dem Code im engeren Sinne und dem mit dem Leiterzeichen # eingeleiteten Kommentar nötig – aber wer R-Code schreibt, achtet in der Regel auf eine gute Lesbarkeit und nutzt besser sichtbare und übersichtliche Einrückungen.

#### 15.4.5 Der Workspace

Anders als bei einer Tabellenkalkulation sieht man normalerweise seine Variablen, Funktionen und Daten in **R** nicht fortwährend. Nicht selten fragt man daher: Wo sind meine Sachen jetzt? Tippen Sie bitte (einschließlich der leeren Klammern) ein:

```
> ls()
```

**R** unterrichtet Sie darüber, dass „A“, „BMI“, „Gewicht“ und „Körperhöhe“ in seinem Gedächtnis existieren, denn anders als beim einfachen Rechnen wie eingangs haben wir zuletzt die Ergebnisse auf die Variable A, BMI sowie Gewicht und Körperhöhe „rausgeschrieben“. Die Bedeutung davon erfährt man, wenn man **R** verlässt, wie gewohnt über das Menü (oben) Daten => Beenden: Es öffnet sich ein Fenster, das fragt „Workspace sichern?“. Antworten Sie bitte mit „ja“, danach wird **R** beendet. Rufen Sie nun bitte **R** erneut auf, und geben „ls()“ & ENTER ein und anschließend „BMI“ & ENTER. **R** meldet, dass es den Workspace wiederhergestellt hat, dass sich darin die Variablen A, BMI, Gewicht und Körperhöhe befinden, und dass die Variable BMI den Wert „27.77778“ innehat. Der Workspace ist das Arbeitsgedächtnis von **R**, es bleibt erhalten; wenn man es sichert, steht es nachfolgend wieder zu Verfügung.

Nein, nicht auf ewig, man kann Inhalte auch wieder löschen. Über das Menü ganz oben: unter „Verschiedenes“ findet man „Liste Objekte auf“, was der Funktion „ls()“ entspricht, und „Entferne alle Objekte“, was genau dieses tut. Alternativ und zielgenauer kann man, beispielsweise, mit der Eingabe von „rm(A)“ (für remove) in der Kommandozeile die Variable A aus dem Workspace entfernen. Tun Sie dies bitte und prüfen das Ergebnis mit „ls()“ nach.

Indes, dass Informationen „irgendwo“ im Computer vorhanden sind, ist für Freunde der Datensicherheit kein wirklich beruhigendes Wissen. Man will schon genauer wissen, wo, denn nur so ist z. B. die Sicherung auf ein externes Medium (USB-Stick, externe Festplatte) möglich. **Abb. 23** zeigt die nötigen Befehle:

Befehl	Wirkung
> ls()	Listet den Inhalt des aktuellen Workspaces auf
> rm(yxz)	löscht das Element xyz aus dem Workspace
> getwd()	sagt, wo der Workspace sich aktuell befindet
> setwd(„D:/RDaten“)	ordnet an, wo der Workspace künftig abgelegt werden soll, hier z. B. im Verzeichnis /RDaten auf D:
> save.image(file = „D:/RDaten/heute.Rdata“)	setzt den Namen (hier: „heute“) und den Ablageort des Workspaces
> load(file = „D:/RDaten/heute.Rdata“)	lädt den Workspace gezielt in das Arbeitsgedächtnis

**Abb. 23** Der Workspace von R: Benennen und Steuern des Ablageortes. „Rdata“ ist die von R erwartete Dateinamenergänzung; sie ist fix und sollte nicht verändert werden.

Mit diesen Befehlen kann man die Daten wiederfinden und auch gezielt sichern. Über die Befehle „save.image“ und „load“ lassen sich auch unterschiedliche Workspaces anlegen und gezielt laden, also zwischen verschiedenen Themen und Problemen wechseln.

#### 15.4.6 Vektoren

Unser Start-Beispiel von Variablen und Wertezuweisungen zu diesen Variablen ist reichlich künstlich, weil diese Variablen nur einen Wert haben. Normalerweise haben wir eine Variable, die viele unterschiedliche Werte annimmt. Das, was in einer Tabellenkalkulation üblicherweise eine (senkrecht geschriebene) Spalte ist, bezeichnet man in **R** als Vektor und schreibt diesen als Zeile, wobei ein Komma als Trennzeichen zwischen den einzelnen Werten dient; diese Einzelwerte bilden eine Folge von mehreren Ausprägungen einer Variablen. Wir bleiben beim Beispiel Body-Mass-Index und haben Daten für acht Personen:

```
> vKörperhöhe <- c(1.55, 1.60, 1.65, 1.70, 1.75, 1.80, 1.85, 1.90)
> vGewicht <- c(70, 70, 70, 70, 90, 90, 90, 90)
```

Mit dem (freiwillig) vorangestellten „v“ zu „Körperhöhe“ merken wir uns, dass wir nun einen Vektor anlegen wollten. Mit dem der Wertereihe vorangestellten „c“ (für „concatenate“ or „combine“) signalisieren wir, dass die Zahlen gemeinsam als Vektor behandelt werden sollen. Ein „Vektor“ ist eine Kette gleichartiger Informationen, also z. B. von Zahlen oder Namen. Mit der Eingabe

```
> vKörperhöhe
```

rufen wir die Datenreihe ab und lesen:

```
[1] 1.55, 1.60, 1.65, 1.70, 1.75, 1.80, 1.85, 1.90
```

Die oben bereits verwendete Formel kann, mit den neuen Namen, wiederverwendet werden:

```
> BMI <- vGewicht/vKörperhöhe^2
> BMI
```

listet auf:

```
[1] 29.13632 27.34375 25.71166 24.22145 29.38776 27.77778 26.29657
24.93075
```

Wobei BMI jetzt ebenfalls einen Vektor bildet. Wir haben also für unsere acht fiktiven Menschen acht BMI-Werte errechnet – und Vektoren und das Arbeiten mit Vektoren kennengelernt. Nur als Ausblick auf Späteres schreiben wir noch schnell:

```
> mean(BMI)
```

und erhalten:

```
[1] 26.85075
```

mit 26,85 das arithmetische Mittel des BMI für diese acht Menschen. **R** beinhaltet (sehr) viele fertige statistische Funktionen wie z. B. „mean()“, die man wie aufgezeigt auf die Vektoren anwenden kann.

### 15.4.7 Matrizen

Tabellen, wie wir sie aus einer Tabellenkalkulation kennen, sind für **R** „Matrizen“, d. h. miteinander verbundene Vektoren. Wir legen unsere Beispieldaten als Matrix ab, indem wir die beiden Vektoren miteinander verbinden:

```
> Beispiel <- cbind(vKörperhöhe, vGewicht)
> Beispiel
```

und erhalten:

```
      vKörperhöhe vGewicht
[1,]      1.55      70
[2,]      1.60      70
[3,]      1.65      70
[4,]      1.70      70
[5,]      1.75      90
[6,]      1.80      90
[7,]      1.85      90
[8,]      1.90      90
```

Nun geben wir diesen Zeilen und Spalten noch Namen, damit wir sie besser lesen und verstehen können:

```
> colnames(Beispiel) = c("m", "kg")
> rownames(Beispiel) = c("Person A", "Person B",
"Person C", "Person D", "Person E", "Person F",
```

```
"Person G", "Person H")  
> Beispiel
```

und erhalten:

```
      m  kg  
Person A 1.55 70  
Person B 1.60 70  
Person C 1.65 70  
Person D 1.70 70  
Person E 1.75 90  
Person F 1.80 90  
Person G 1.85 90  
Person H 1.90 90
```

Auch die Namen werden also als Vektoren verstanden – siehe „c(...)“ – und die Tatsache, dass es sich nun um Zeichen statt Zahlen handelt, wird mit den An- und Abführungszeichen definiert. Keine Sorge, so aufwändig wie jetzt geht die Datenerfassung und -verwaltung in der Praxis nur selten vonstatten. Aber die Beispiele führen in grundlegende Strukturen von **R** ein. Die Sache mit den Zahlen und Zeichenketten (m, kg) führt zum nächsten und vorletzten Thema dieser ersten Kontaktaufnahme: Datentypen.

### 15.4.8 Datentypen

Da wir eingangs schon das Thema Skalenniveau besprochen haben, lassen die verschiedenen Datentypen, die man in **R** vereinbaren kann, recht schnell erläutern (**Abb. 24**). Im Kern geht es darum, zwischen Text/Zeichenketten und Zahlen zu unterscheiden. Text (Zeichenkette, *string*) wird – wir haben das schon gesehen – durch An- und Abführungszeichen markiert. Der Datentyp „numeric“ wird quasi erwartet, es wird nicht zwischen Ganzen Zahlen (ohne Komma; *integer*) und Reellen Zahlen unterschieden. Codierte Werte, wo die einzelnen Werte (Zahlen oder Zeichenketten) für eine Nominalskala stehen, werden als Faktoren bezeichnet.

Datentyp	Beschreibung	Beispiel
character	Zeichenkette	„dies ist ein Text“
numeric	Ganze und Reelle Zahl	4.07
complex	Komplexe Zahl <sup>(35)</sup>	2.8 -3i
logical	Logischer Wert	TRUE (1), FALSE (0)
factor	Faktor	1 = rot, 2 = blau, 3 = grün, ...

**Abb. 24** Datentypen in R.

### 15.4.9 Funktionen

Der bereits benutzte Befehl „mean()“ ist eine Funktion, und **R** enthält viele fertige Funktionen. Man kann Funktionen aber auch selbst schreiben, und das wollen wir uns am Beispiel des BMI kurz ansehen. Funktionen in **R** bestehen aus einem Namen, der Deklaration der Eingabe (auch Argument genannt) und dem Rumpf der Funktion, in dem fixiert ist, was mit dem Argument geschehen soll. Funktionen werden wie folgt geschrieben – wobei die Klammern und ihre unterschiedliche Art sehr wichtig sind:

```
Funktionsname <- function(Argument) {
  (Funktionsrumpf)
}
```

Setzen wir dieses allgemeine Gerüst konkret für die Berechnung des BMI um und geben statt einem Argument gleich zwei Argumente ein:

```
fBMI <- function(kg,m) {
  kg/(m*m)
}
```

Die Argumente für die Berechnung sind „kg“ und „m“ (Meter), und die Berechnung selbst besagt: teile kg durch Meter zum Quadrat. Soweit die Funktion – die man nun anwenden kann auf eine Tabelle „daten“, in der die Variablen „Gewicht“ und „Khoehe“ stehen:

```
fBMI(daten$Gewicht, daten$Khoehe)
```

In unserem Übungsbeispiel ist der Vorteil, die Berechnung hier statt als direktes Skript als Funktion zu schreiben, nur bedingt ersichtlich. Immerhin wird deutlich, dass man unserer Funktion `fBMI` nun auch andere resp. anders benannte Variablen aus anderen Datensätze würde anvertrauen können. Deutlicher wird der Vorteil, wenn wir bei-

---

<sup>35</sup> Wer seit der Schulzeit vergessen hat, was „komplexe Zahlen“ sind, wikipedierte das Thema, oder vergesse es, weil wir sie hier nur der Vollständigkeit halber anführen und ansonsten in diesem Buch nicht mehr benötigen werden.

spielsweise aus Daten mit Durchmessern und Höhen von Gefäßen deren Fassungsvermögen errechnen wollen. Da unsere Keramikgefäße meist rotationssymmetrisch sind, muss man dazu im Kern das Fassungsvermögen ( $V$ ) eines Kegelstumpfes berechnen. Die Lehrbuchformel dazu lautet:

$$V = ((\text{Höhe} * \pi) / 3) * (R^2 + R*r + r^2)$$

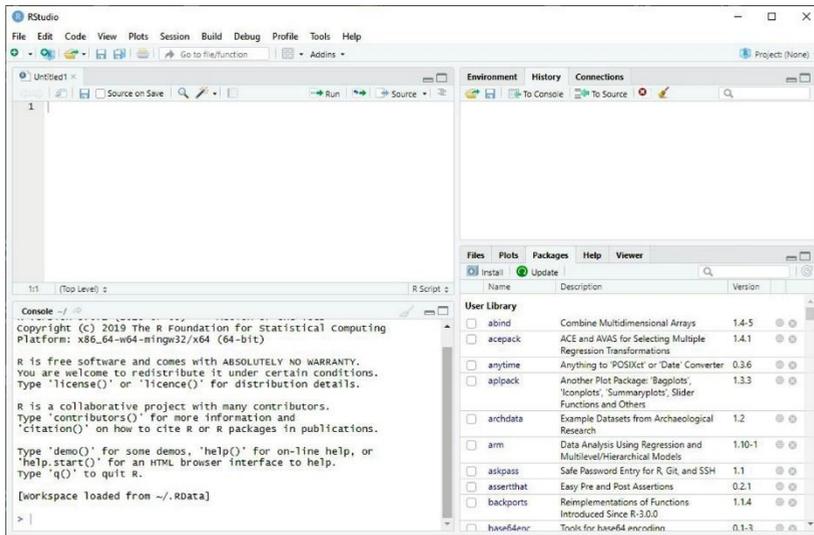
wobei „ $R$ “ für den größeren Radius und „ $r$ “ für den kleineren Radius steht. Bei offenen Gefäßen ist dies ein Kegelstumpf mit halbem Bodendurchmesser als „ $r$ “ und halbem Randdurchmesser als „ $R$ “. Bei zweigliedrigen Gefäßen sind es zwei, bei dreigliedrigen Gefäßen jeweils drei Kegelstümpfe (mit Boden-, Bauch-, Hals- und Randdurchmesser), die zu berechnen sind und addiert werden. In diesem Fall würden wir die Funktion für die Berechnung des Kegelstumpfes einmal schreiben, und ihr je nach Fall die passenden Durchmesser und Höhen übergeben und anschließend die Teilergebnisse addieren. In diesem Fall wäre das Arbeiten mit einer Funktion erheblich übersichtlicher (und damit auch weniger fehleranfällig) als das Hintereinanderstellen der einzelnen Formeln in einem Skript.

Ziehen wir eine Zwischenbilanz: Die Bedienung des rohen **R** ist ungewohnt, wir vermissen allerlei ansonsten üblichen Komfort und Anschaulichkeit. Andererseits: das Berechnen des BMI und das Bilden des Mittelwertes war konkurrenzlos schnell, mit jeder Tabellenkalkulation hätten wir dafür etwas mehr Zeit gebraucht. Doch an dieser Stelle beenden wir den Ausflug ins rohe **R** wieder und sehen uns mit dem RStudio und anschließend dem R-Commander zwei Werkzeuge an, die das Arbeiten mit **R** einfacher machen.

## 15.5 RSTUDIO

Das rohe **R** wird – wie soeben gesehen – via Befehlskonsole und Kommandozeile ohne eine grafische Benutzeroberfläche bedient, was für viele Anwender zunächst eine hohe Hürde darstellt. Viele Nutzer verwenden daher eine helfende Umgebung. Ich empfehle Ihnen sehr, dies ebenfalls zu tun. Eine sehr häufig gewählte Kombination mit **R** ist die Oberfläche „RStudio“, ergänzt um den „R-Commander“. RStudio ist nicht Teil der bereits installierten Distribution von **R**, sondern wird von einer eigenen Website bezogen: <https://www.rstudio.com/> [24.1.2020]. Im Downloadbereich findet man dort neben dem freien (kostenlosen) Basispaket für den Desktop-PC weitere RStudio-Varianten, die bezahlt werden müssen. Für uns ist das freie Basispaket völlig

hinreichend („Open Source License, free“). Laden Sie wiederum die für Ihren Computer bzw. Ihr Betriebssystem passende Version herunter und installieren Sie RStudio wie üblich.<sup>36</sup> Mit einem Doppelklick auf das Icon startet RStudio (**Abb. 25**).



**Abb. 25** Der Bildschirm von RStudio (Vers. 1.2.5).

Mit dem Doppelklick auf das Programmsymbol von RStudio öffnet sich die Benutzeroberfläche, die aus vier Fenstern besteht. Sieht man weniger als vier Fenster, sind Fenster in der linken oder rechten Spalte zusammengeklappt resp. übereinandergelegt. Mit einem Klick auf das von der Windows-Bedienung gewohnte Fenster- resp. Fenster-in-Fenster-Symbol lässt sich das beheben. In der Standardkonfiguration der RStudio-Ansicht erkennt man links unten „Console“, die Konsole. Der Text dort kommt Ihnen bekannt vor aus unserem Üben mit dem rohen **R** – es ist die Konsole von **R**, die in die Oberfläche von RStudio eingebettet ist. Wenn Sie hier testweise eine kleine Rechenaufgabe wie z. B. `2 + 3 & ENTER` eingeben, erscheint sofort das Gewohnte `„[1] 5“`, d. h. das linke (untere) Fenster ist das rohe **R** und funktioniert auch so.

<sup>36</sup> Möglicherweise „vergisst“ RStudio es, ein Programm-Icon auf Ihrem Desktop abzulegen. Sie können das händisch nachholen: In das Programmverzeichnis gehen (vermutlich: `c:/Programe/RStudio/bin`) und dort „`rstudio`“ suchen. Am besten eine Verknüpfung dieser Datei „`rstudio`“ erstellen und diese auf den Desktop legen.

Bitte beachten Sie das Fenster rechts oben, dessen Reiter auf „History“ steht (oder nun von Ihnen dort per Klick aktiviert wird), und das Ihre soeben erfolgte Eingabe protokolliert; hier sollte jetzt in der obersten Zeile des Fensters „2 + 3“ stehen – aber eben nur das Protokoll Ihrer Eingabe, ohne die von **R** in der Konsole ausgegebenen Ergebnisse.

Links zeigt Ihre Benutzeroberfläche oberhalb der R-Konsole ein weiteres Fenster. Wenn nicht, legen wir es jetzt an: unter „File“ links oben sehen Sie einen kleinen grünen Kreis mit weißem Kreuz darinnen: draufklicken, und ganz oben „R Script“ anklicken/auswählen. Nun sollten Sie auf der linken Seite von RStudio unten die R-Konsole sehen, und oben das Skriptfenster: eine Schreibfläche zum Schreiben von Skripten, d. h. Abfolgen von Befehlszeilen. Lassen Sie uns unser kleines Beispiel, mit dem wir erste Erfahrungen im rohen **R** gesammelt haben, hier wiederholen. Wir schreiben in dieses Skriptfenster links oben (oder copy-pasten dort hinein):

```
vKörperhöhe <- c(1.55, 1.60, 1.65, 1.70, 1.75, 1.80, 1.85, 1.90)
vGewicht <- c(70, 70, 70, 70, 90, 90, 90, 90)
BMI <- vGewicht/vKörperhöhe^2
BMI
mean(BMI)
Beispiel <- cbind(vKörperhöhe, vGewicht, BMI)
Beispiel
colnames(Beispiel) = c("m", "kg", "BMI")
rownames(Beispiel) <- c("Person A", "Person B", "Person C", "Person D",
"Person E", "Person F", "Person G", "Person H")
Beispiel
```

Das Skriptfenster setzt diesen Zeilen eine Nummerierung der Zeilen voran, sonst passiert nichts. Bis auf dem Umstand, dass die automatisch als Daten erkannten Zahlen im Skriptfenster nun in blauer statt schwarzer Schrift erscheinen. Nun kann man das Skript ausführen: Den ganzen Text mit dem Mauszeiger markieren (Hintergrund wird blau) und die kleine Schaltfläche unmittelbar über dem markierten Bereich anschauen, wo sich u. a. etwas rechts von der Mitte eine Schaltfläche „Run“ findet. Draufklicken. Jetzt wird das Skript ausgeführt: im Fenster links unten, d. h. in der R-Konsole, sieht man das Protokoll der Befehlszeilen und die von **R** ausgeworfenen Ergebnisse. Gleichzeitig taucht unser Skript jetzt als Protokoll im Fenster rechts oben auf. Aha, „History“ protokolliert alle unsere Eingaben.

Nun, das Ganze ist noch nicht vollends elegant, aber wir verstehen: Man kann seine Arbeit protokollieren (rechts oben, „History“), und man kann im Skriptfenster links oben – ggf. im Verfahren Versuch & Irrtum – Befehlsfolgen entwickeln. Die

dortige Schaltfläche „Sichern“ signalisiert, dass man Skripte speichern kann. Das Fenster links oben erlaubt es also, ein zuletzt sinnvoll durchgelaufenes Skript zu sichern, es später wieder in eben dieses Fenster hineinzuladen, durch Editieren an neue Aufgaben anzupassen und die neuen Ergebnisse in der R-Konsole anzuschauen. Für das Editieren gibt es oberhalb des Skriptfensters kleine Helferlein. Hinter der Lupe verbirgt sich eine Suche & Ersetze-Funktion. Hinter dem Zauberstab rechts davon öffnet sich ein ganzes Menü mit Unterstützungsangeboten – die wir jetzt nicht weiter austesten wollen.

Lassen Sie uns die beiden Fenster rechts noch etwas studieren. Rechts oben ist der Reiter „History“ aktiviert. Aktivieren Sie bitte mit einem Mausklick den Reiter „Environment“: Wir sehen „Data“ und „Values“, unsere Übungsdaten. Während wir beim Üben im rohen **R** in der Gefahr waren, den Überblick zu verlieren, sind hier die Daten wiedergegeben. Bei den Values zeigt uns „num“ die Art der Variable an (*numerical*), die Zahlen in den eckigen Klammern – hier [1 : 8] – signalisieren, dass es acht Werte gibt vom Wert 1 bis zum Wert 8, und anschließend lesen wir die ersten konkreten Daten jeder Variable. Bei „Data“ lesen wir, dass es sich um eine Datenmatrix handelt aus „num“, also Zahlen, von 1 bis 8 Zeilen und 1 bis 3 Spalten.

Lassen Sie uns nun noch die Sache mit dem Sichern erledigen: In der Befehlsleiste von RStudio ganz oben mit „File“, „Edit“, „Code“... finden wir „Session“. Das Menü dort enthält u. a. die Fläche „Set Working directory“ mit der nachfolgenden Option „Choose directory“. Ich wähle wieder die schon beim Einarbeiten ins rohe **R** gesetzte Option „D:/Rdata“, was mit einem Klick auf die Schaltfläche „Open“ bestätigt wird. Im Feld links unter, der R Konsole, lesen wir nun den Befehl in der **R** Sprache als „setwd(\"D:/RData\")“. Nochmals in die oberste Befehlsleiste zu „Session“, dann „Save Workspace As“, und unseren Daten einen Namen geben, z. B. „Startübung“, und das Speichern bestätigen. Jetzt wissen wir, wo die Daten sind und wie sie heißen, d. h. können sie morgen auch wieder neu hereinladen.

Aktivieren Sie nun im RStudio-Fenster rechts unten nun den ersten Reiter, „Files“. Im Fenster sollte nun die eben vorgenommene Einstellung sichtbar sein, d. h. der Ordner mit den Daten und der soeben gesicherte Workspace. Klar, hier bestünde nun auch die Option, an einem fortgeschrittenen Arbeitsstadium mehrere unterschiedliche Workspaces vorzufinden und ggf. den aktiven zu schließen und einen anderen zu laden, also zwischen Projekten, Datensätzen oder Fragestellungen zu wechseln.

Erinnern Sie sich noch an das Beispiel oben, bei dem wir für die Zahlen 1, 10, 100 und 1.000 jeweils den Zehnerlogarithmus berechnet haben und die Ergebnisse

anschließend auch rücktransformieren wollten? Mit den Werkzeugen, die wir nun kennen gelernt haben, geht das flüssiger und – weil mit weniger Tipparbeit verbunden – auch fehlersicherer:

```
# Eingabe von 4 Zahlen, zum Vektor „Eingabe“ zusammengefasst:
Eingabe <- c(1, 10, 100, 1000)
Eingabe                                     # Kontrolle, ob korrekt erfolgt
Ergebnis <- log10(Eingabe)                 # Logarithmus 10 für den Vektor "Eingabe",
                                             # herausschreiben in den Vektor "Ergebnis"
Ergebnis                                   # Ausgabe des Ergebnisses
Rückrechnung <- 10^Ergebnis                # Rücktransformation (zur Kontrolle)
Rückrechnung                               # Ausgabe der Rückrechnung, Soll:
                                             # Zahlen wie bei Eingabe. Stimmt's?
```

Bitte geben Sie das Beispiel in das Skriptfenster von RStudio, also das Fenster links oben ein, und setzen es mit „Run“ in Gang, schauen Sie die Abläufe und Ergebnisse links unten an.

Wir erweitern die Erfahrungen:

```
Ergebnis.log10 <- log10(Eingabe) # Logarithmus 10 für "Eingabe"
Ergebnis.ln <- log(Eingabe)      # Logarithmus naturalis für "Eingabe"
Ergebnis.cubic <- Eingabe^(1/3) # Kubikwurzeltransformation, 3.Wurzel
Ergebnis.sqrt <- sqrt(Eingabe)  # Wurzeltransformation für "Eingabe"
Ergebnis.log10
Ergebnis.ln
Ergebnis.cubic
Ergebnis.sqrt
```

... und sehen an den Ausgaben, wie die vier Transformationen wirken: der Zehnerlogarithmus staucht die Zahlenspannen von 1 bis 1.000 erheblich stärker als der Logarithmus naturalis (i.e. zur Basis „e“, Eulersche Zahl), während die Wurzeltransformation die am wenigsten stauchende Transformation ist.

Im Fenster rechts unten hatten wir zuletzt den Reiter „Files“ aktiviert. Aktivieren Sie nun bitte den Reiter „Packages“ (wie in **Abb. 25**): die Liste zeigt Ihnen, welche Pakete aktuell in Ihrem R installiert sind. Sobald Sie in das graue Kästchen vor einem Paketnamen klicken, wird das bereits installierte Paket auch geladen. Praktisch: nach Klicken auf das grüne Symbol „Update“ wird für alle installierten Pakete geprüft, ob es für sie Aktualisierungen gibt. Wenn ja, lässt sich dieses Aktualisieren sehr bequem über die aufpoppende Schaltfläche vornehmen.

Hier unterbrechen wir das Kennenlernen von RStudio und widmen uns den weiteren Ergänzungen, die wir anschauen wollen. Im Gegensatz zum selbstständigen

RStudio sind die nächsten Erweiterungen, die wir benötigen, sog. „Pakete“. Mit diesem Namen werden alle Erweiterungen der Basisausstattung von **R** bezeichnet, mit denen man es um Funktionalitäten (ggf. auch um als Paket deponierte Datensätze) erweitert. Der R-Commander ist ein solches Paket.

## 15.6 DAS INSTALLIEREN VON „PAKETEN“ IN R

Das Installieren der „Pakete“ genannten Erweiterungen der Basisfunktionalitäten von **R** geschieht immer in zwei Schritten: (a) erstmalig Installieren, und (b) bei jedem Start von **R** neu laden. Beides ist sowohl über das rohe **R** wie auch mit Hilfe von RStudio kein Hexenwerk, sondern schnell erledigte Routine.

### 15.6.1 Ein Paket erstmals beziehen

Beginnen wir mit (a), dem Beziehen von Paketen. Im Falle dieser Einführung nenne ich ggf. den Namen sowie die Quelle der Pakete, die wir benutzen wollen. Wie aber soll es weitergehen, wenn Sie selbstständig arbeiten wollen, welche Schätze warten dort, und vor allem: wo? Nun, (fast) alle Pakete gibt es via CRAN, und CRAN listet alle dort hinterlegten Pakete sorgfältig auf. CRAN? – ist die Abkürzung für „The Comprehensive R Archive Network“, das 1997 gegründete und frei zugängliche Archiv, in dem (fast) alle Pakete deponiert werden, die nicht zum Basisumfang von R dazugehören, aber bei Bedarf hinzugeladen werden können. Anfang 2020 waren es mehr als 15.300 Pakete, beim Start von CRAN Mitte 2001 waren es ca. 110. Eine ähnliche, aber deutlich kleinere Plattform ist „Bioconductor“: <https://www.bioconductor.org/> [24.1.2020] mit mehr als 1.800 Paketen vor allem zu biologischen Themen (u. a. der Analyse von aDNA-Untersuchungen). Das Einbringen eines Pakets auf CRAN ähnelt der Publikation eines wissenschaftlichen Aufsatzes, d. h. es gibt Redaktionsrichtlinien, an die sich Einreicher orientieren müssen, und anschließend eine Qualitätssicherung durch Kollegen (Peer Review), erst danach werden Pakete einschließlich einer Dokumentation auf CRAN den Nutzern bereitgestellt. Aus Nutzersicht bedeutet dies: Pakete für **R**, die man bei CRAN herunterlädt, können – wie auch wissenschaftliche Aufsätze – trotz ihrer Veröffentlichung weiterhin Fehler und Irrtümer enthalten; Nutzer dürfen aber sicher sein, dass Paket-Autoren einen beträchtlichen Ehrgeiz haben, dass dies nicht der Fall ist, und dass unabhängige Dritte dies vor der Veröffentlichung nochmals ernsthaft geprüft haben.

Zurück zur Praxis mit CRAN: Gehen Sie die Seite von CRAN, die wir schon kennen (<https://cran.r-project.org/> [24.1.2020]) und klicken in der Navigationsleiste links (Mitte, unter „Software“) auf „Packages“. Wenn man nicht durch Tipps, Lehr-

bücher oder wissenschaftliche Publikationen auf ein konkretes Paket aufmerksam geworden ist, sondern etwas sucht, kann man die alphabetische Auflistung öffnen und dort (mit der Suchfunktion Ihres Browsers) nach geeigneten Stichworten suchen. So spült etwa das Suchwort „correspondence“ für die in der Archäologie wichtige Korrespondenzanalyse mindestens zehn Treffer empor. Was genau die Pakete tun, sieht man dann nach einem Klick auf das konkrete Paket. So liest man z. B. hinter „ca – Simple, Multiple and Joint Correspondence Analysis“, dass es sich um ein am 14.12.2016 publiziertes Paket handelt, bei dem u. a. Michael Greenacre Autor ist, von dem man zugleich zwei wichtige Bücher zur Korrespondenzanalyse kennt, die heute als Grundlagenwerke gelten. Sprich: u. a. ein renommierter Experte und Buchautor steht hinter diesem Paket. Während diese erste Seite jedes Pakets nur Übersichtsinformationen liefert, ergibt sich Vertiefendes aus dem meist vorhandenen „Reference manual“, hier z. B. „ca.pdf“. Ein Klick dorthin zeigt, was diese Handbücher sind: sehr kurze, sehr technische Darlegungen der Funktionalitäten der Pakete, der Eingabe-Parameter, mit denen man sie steuert und bedient – keine leichte Lektüre. Im Optimalfall gehört ein Paket zu einem Buch oder einem Aufsatz und wird dort auch über das rein Technische hinaus näher beschrieben. Suchen Sie ggf. nach einer solchen erweiterten Beschreibung, für populäre Pakete finden sich meist auch Einführungen bei YouTube.

Im rohen **R** (bitte starten) gibt sich über die oberste Befehlsleiste „Pakete“ und „Lade Pakete“ ein schmales Menü, das alle Pakete auflistet, die Sie bereits auf Ihrem PC/Laptop installiert haben. Suchen Sie dort einmal nach „Rcmdr“. Steht der Rcmdr, der R-Commander in dieser Liste, wurde er bereits installiert. Wenn nicht:

=> „Pakete“; => „Setze CRAN-Mirror“ und wählen Sie dort einen (relativ) nahegelegenen Standort, z. B. „Germany (Göttingen)“ oder „Germany (Münster)“ und bestätigen Sie mit OK.

=> „Pakete“; => „Installiere Paket(e)“; es klappt das schmale Menü mit einer (sehr langen) alphabetisch sortierten Liste aller auf dem gewählten CRAN-Mirror deponierten Pakete auf. Wir suchen „Rcmdr“, klicken diesen an und bestätigen mit OK. **R** protokolliert in der R-Konsole die Geschehnisse, d. h. berichtet über das Herunterladen, den erfolgreichen Abgleich der „MD5 Summen“ (was mit Sicherheit und Datenintegrität zu tun hat), usw. Sollte beim Installieren irgendetwas schief laufen, stecken in diesem Protokoll die nötigen Fehlermeldungen. Damit ist die reine Installation, die nur beim ersten Mal nötig ist, abgeschlossen.

### 15.6.2 Installierte Pakete bei Bedarf laden

Nun zu Teil (b), dem Laden eines Pakets im rohen **R**: => „Pakete“; „Lade Paket“: die Liste der bei Ihnen installierten Pakete klappt auf, wo nun auch „Rcmdr“ sichtbar sein

müsste. Anklicken und mit OK bestätigen. Nach kurzem Warten öffnet sich ein bzw. zwei neue Fenster: die Oberfläche des R-Commanders. Sehen Sie bitte noch kurz in die R-Konsole, in das Protokoll. Dort wird u. a. darüber berichtet, dass der R-Commander zwingend weitere Komponenten (Pakete) benötigt und er sich diese selbstständig hineingeladen hat – weshalb der Start auch ein paar Sekunden länger dauerte.

Wenn Paket *x* für sein Funktionieren auch die Pakete *y* und *z* benötigt (ein recht häufiger Fall), nennt man dies Abhängigkeiten (*dependencies*). Sofern das Nötige automatisch geschehen ist (wie hier beim Laden des R-Commanders), muss nichts unternommen werden. So das nicht geschehen ist, muss (a) das Installieren wie (b) auch das Laden solcher nötigen Zusatzpakete händisch nachgeholt werden – was, wie Sie nun wissen, schnell erledigt ist. Solches Nacharbeiten ist selten nötig und wenn, dann meist nur bei der ersten Installation. Daher ist die Lektüre der Installations-Protokolle in der R-Konsole sinnvoll.

Wir sind jetzt eigentlich fertig, aber lassen Sie uns das ganze Thema noch kurz unter der Oberfläche von RStudio anschauen (bitte **R** verlassen und RStudio starten). Das Fenster unten rechts zeigt mehrere Reiter, klicken Sie bitte auf „Packages“. Nun sind Ihre installierten Pakete sichtbar, etwas übersichtlicher als zuvor im rohen **R**. Der Schalter „Install“ führt Sie bei Bedarf zur Installation weiterer Pakete. Der Schalter „Update“ veranlasst RStudio zu einer Überprüfung, ob alle Ihre Pakete aktuell sind oder ob es für einzelne Pakete neuere Versionen gibt. Klicken Sie in größeren Abständen immer mal wieder auf „Update“; Sie erhalten eine Meldung, ob alles aktuell ist und ggf. Unterstützung, um ein Update noch bequemer als im rohen **R** einzuspielen.

Ein Klick auf den Namen eines Paketes in diesem Menü springt in den Hilfetext dieses Paketes (bei CRAN). Wenn Sie ein Paket nur laden wollen, setzen Sie per Klick ein Häkchen in das graue Kästchen vor dem Paketnamen. Sie lesen dann links in der R-Konsole das Ihnen schon vom vorigen Mal bekannte Protokoll des Ladens, hier des R-Commanders, der sich zudem auch mit eigenem Fenster als gestartet meldet. Das war's mit der Technik, ab jetzt können wir uns der inhaltlichen Arbeit mit **R** widmen.

## 15.7 DEN R-COMMANDER KENNENLERNEN

Arbeitet man mit dem rohen R und will den R-Commander von dort aus starten, gibt man in der R-Konsole einfach

```
> library(Rcmdr)
```

ein. Der R-Commander startet dann in einem separaten, zweiteiligen Fenster. Die Oberfläche des R-Commanders wirkt recht leer, doch der R-Commander ermöglicht

für eine längere Zeit und für viele Aufgaben das Arbeiten mit **R** ohne Bedienung einer Kommandozeile. Fangen wir doch einfach an. Folgen Sie diesem Arbeitspfad:

=> „Datenmanagement“; => „Importiere Daten“; => „aus Excel-Datei“; in dem aufpoppenden Fenster geben Sie unserem Datensatz einen passenden Namen, z. B. „Geld“. Lassen Sie uns das neue Fenster prüfen: das Häkchen vor „Variablennamen in erster Zeile der Spreadsheet“ ist richtig; „Zeilennamen in erster Zeile der Spreadsheet“: wohl nicht, also kein Häkchen; „Konvertiere Zeichen-Variablen in Faktoren“: ja, hier ein Häkchen setzen.<sup>37</sup> Wir bestätigen mit OK – worauf sich das übliche Fenster des Dateimanagers öffnet und uns das Suchen und Finden unseres Datensatzes ermöglicht. Anklicken, ggf. mit „Öffnen“ bestätigen. Die R-Konsole sollte nun das Lesen des Datensatzes protokolliert haben. Klicken wir im R-Commander auf die Schaltfläche „Datenmatrix betrachten“, können wir das Ergebnis sehen, d. h. prüfen, ob unsere Daten korrekt eingelesen wurden.

### 15.7.1 ... und wenn's nicht klappt mit dem Datenimport?

Oft funktioniert der Datenimport wie beschrieben klick – klick – fertig. Manchmal aber nicht, und die Frage „warum?“ führt in Abgründe. Sprich: in Konstellationen, die individuell sind und für die hier kein allgemeingültiges einfaches Rezept gegeben werden kann. Was weniger an **R** oder dem R-Commander liegt, sondern daran, dass diverse Excel-Versionen sich sehr voneinander unterscheiden und .xls allzu oft eben nicht gleich .xls ist. Sehr hinderlich können z. B. (meist unbewusste) Voreinstellungen betreffs ASCII- oder UniCode-Kodierung samt resultierender Probleme mit deutschen Umlauten sein (Empfehlung: immer und auf beiden Seiten gezielt UniCode-Kodierung verwenden), mehr noch können es die länderspezifischen Einstellungen bei den Zahlenformaten sein, die – ja nach Ländereinstellung unterschiedlich – Punkte oder Kommas als Tausenderzeichen oder Dezimalzeichen verwenden. Wer in solchen Konstellationen festhängt, kann verschiedene Abspeicher-Arten von \*.xls durchprobieren – am besten stets mit Notizen, damit man weiß, wann genau etwas gescheitert resp. gut gegangen ist – und hoffen, dass einer der probierten Wege glatt funktioniert. Wenn's dauerhaft klemmt, hilft nur, \*.xls resp. \*.xlsx zu vermeiden und zu umgehen. Denn: Über den R-Commander lassen sich verschiedene Dateitypen (eigentlich) relativ unkompliziert einlesen. Es werden verarbeitet: SAS-Dateien, SPSS-Dateien, Excel-

---

<sup>37</sup> Bitte jetzt nicht fragen „warum?“ – der Sachverhalt ist komplexer, stört hier und wird etwas später gründlich erläutert.

Dateien (\*.xls und \*.xlsx), \*.csv- und \*.txt-Dateien und Textdateien. Auch ein schlichtes Copy-and-paste ist möglich. Als sehr stabil haben sich die Importe aus dem SPSS-Format erwiesen sowie solche, die als „Tab-delimited“ geschrieben wurden, bei denen also das Tabulator-Zeichen als Trenner zwischen den Variablen dient.<sup>38</sup> Das Hauptfeld des ggf. nötigen Experimentierens an dieser Stelle liegt also zunächst weniger am Import mit dem R-Commander, sondern mehr am geeigneten Export resp. Abspeichern aus der Ursprungsdatei.

Wer hier festhängt, geht beim R-Commander anstelle des oben beschriebenen Wegs über \*.xls besser über: => Datenmanagement; => Importiere Daten; => „aus Textdatei, Zwischenablage oder URL ...“, und findet hier ein neues, mächtigeres Fenster vor, wo u. a. exakt angegeben werden kann, welches Zeichen die einzelnen Variablen trennt und welches Zeichen als Dezimaltrennzeichen dient. Ggf. hilft nur geduldiges Durchprobieren und dann für spätere Importe den Weg notieren, der erfolgreich war.

Die R-Commander-Schaltfläche „Datenmatrix betrachten“ zeigt die Datentabelle, die im R-Commander angekommen ist: unbedingt anschauen und prüfen, ob das Bild dem Erwarteten entspricht. Sonst den Import mit anderen Einstellungen wiederholen, bis es passt.

Die R-Commander-Schaltfläche „Datenmatrix bearbeiten“ ermöglicht uns, einzelne Zeilen und Spalten zu löschen, zu verändern (um z. B. fehlerhafte Daten zu korrigieren) und auch weitere Daten einzugeben. Hier z. B. ist die Spalte „rownames“ überflüssig, wir klicken auf den Spaltenkopf, => „Bearbeiten“; => „Aktuelle Spalte löschen“ – und weg ist sie. Mit OK bestätigen, um die Bearbeitung wieder zu verlassen.

### 15.7.2 Hinweis: Editor Notepad++

Steckt man völlig verzweifelt an der Baustelle Datenexport/-import fest, kann es hilfreich sein, die Daten in der Datei anzuschauen, insbesondere bei \*.csv- und \*.txt/\*.dat-Dateien, und zwar unabhängig von dem Programm, das sie exportiert hat. Da gibt die Chance, z. B. zu erkennen, dass – wiewohl anders erwartet – das Komma statt des Semikolons (oder umgekehrt) als Trennzeichen verwendet wurde. Dazu benötigt man einen sog. Editor. Der bei Windows als Bestandteil des Betriebssystems integrierte „Editor“ kann dabei gute Dienste leisten. Mächtiger ist das freie Programm Notepad++, das ich Ihnen zu installieren anempfehle. Mit seiner Hilfe kann man auch

---

<sup>38</sup> Die Dateierweiterung für solche Dateien ist leider etwas ungenormt: Tab-delimited-Daten werden oft als \*.txt-Dateien benannt, aber auch als \*.dat- oder \*.tab-Dateien geschrieben.

Zeilenumbrüche oder Tabulatoren suchen und ersetzen, Zeilen sortieren, Spalten und Blöcke ausschneiden, zwischen den diversen Zeichensätzen (ASCII, ANSI, UTF-8) konvertieren, usw. Man konsultiere die einschlägige Website: <https://notepad-plus-plus.org/> [24.1.2020].

### 15.7.3 Erste Schritte jenseits des Datenimports im R-Commander

Lassen Sie uns neugierig sein und der verlockenden Befehlszeile ganz oben nachgeben. => „Grafiken“; => „Histogramm“, dort Prägejahr auswählen und mit OK bestätigen. Hm, wo ist jetzt mein Histogramm? Arbeitet man – u.a. wg. der höheren Stabilität – in der Kombination rohes **R** & R-Commander, öffnet sich für die Grafiken automatisch und deutlich sichtbar ein neues Fenster, das oben links die üblichen Weiterverarbeitungsoptionen anbietet, insbes. zu => Dateien; => Speichern als... die nötige Auswahl. Hat man hingegen den R-Commander aus RStudio heraus gestartet, gibt es kein weiteres Fenster, vielmehr liegen die Grafiken nun im R-Studio-Fenster unten rechts: dort statt des Reiters „Packages“, den wir eingangs benutzt haben, nun den Reiter „Plots“ aktivieren: unser Histogramm. Mehr? Ja:

=> „Grafiken“; => „Zeichne diskrete numerische Variable“; „Nominal“ aussuchen und mit OK bestätigen.

=> „Grafiken“; => „Balkendiagramm“; „Land“ auswählen und mit OK bestätigen.

=> „Grafiken“; => „Kreisdiagramm“; „Prägeort“ auswählen und mit OK bestätigen.

Schnell noch Statistiken: => „Statistik“; => „Deskriptive Statistik“; „Aktive Datenmatrix“. Links in der R-Konsole lesen wir die Ergebnisse, d. h. Häufigkeiten wurde ausgezählt und für metrische Variablen werden angegeben: Minimum und Maximum, der Mittelwert, der Median und die 25%- und 75%-Quantile sowie die Anzahl der NA's (*not available*) – der fehlenden Werte.

Was wir hier sehen, ist noch nicht perfekt und bedarf des Feinschliffs. Gleichwie: Die Grafiken waren nun sehr schnell erzeugt. Anders als beim Auswerten via LO-Calc bedurfte es nicht des komplizierten Umwegs über die Pivot-Tabellen. Hoffentlich haben Sie nebenbei den Eindruck gewonnen, dass die Oberflächen RStudio und R-Commander die Bedienung von **R** erheblich erleichtern. Nicht vergessen: Im Skriptfenster wurden alle zusammengedrückten Befehle in der Sprache von **R** protokolliert. Bei den Grafiken stand dort z. B. „col="darkgray"“ – eine Voreinstellung der Farbe, die man nun offensichtlich leicht verändern könnte (z. B. in „green“, „red“ oder „pink“), um

mit dem Markieren des Befehls und der Schaltfläche „Befehl ausführen“ (unten) die gewünschte andere Farbe ins Histogramm zu zaubern. Im Plot-Fenster von RStudio rechts unten haben Sie oberhalb der Grafiken sicherlich schon die Schaltfläche „Export“ entdeckt, die tut, was sie andeutet: Man kann der aktuellen Grafik einen Namen geben und sie in eine externe Datei abspeichern, als \*.pdf oder auch als \*.jpg, \*.tiff, \*.eps-Datei usw.

#### 15.7.4 Hinweis: Stabilität des R-Commanders

Der Verfasser arbeitet gerne mit RStudio und startet den R-Commander von dort, ein Arbeitsansatz, der m. E. vielerlei Vorteile hat und vor allem sukzessive auf das spätere Arbeiten an komplexeren Analysen ohne den R-Commander hinführen kann. Leider ist die Kombination von RStudio und R-Commander manchmal etwas instabil, immer mal wieder stürzt der R-Commander ab und muss dann neu gestartet werden. Wen das nervt, der kann den R-Commander auch vom „rohen“ **R** her starten und benutzen, was nach meinen Erfahrungen etwas schneller und vor allem deutlich stabiler ist. Wie das geht?

=> **R** starten; in der Kommandozeile eintippen `> library(Rcmdr)` [ENTER]  
– das war's, der R-Commander startet.

Das gelegentliche Abstürzen des R-Commanders unter RStudio ist ein bekanntes Problem (liegt also weder an Ihnen noch Ihrem Computer) und hängt wohl vor allem damit zusammen, dass der aktive Workspace des R-Commanders und jener von RStudio – sagen wir – sich nicht mögen. So empfiehlt der Paketautor der R-Commanders, John Fox, ganz grundsätzlich aus dem R-Commander heraus nie einen Workspace zu speichern, sondern die Daten und ggf. auch Skripte stets *ab ovo* einzulesen.

#### 15.7.5 Hinweis: R-Commander um weitere Funktionen erweitern

Der R-Commander kann, wie z.B. auch JAMOVI, durch Hinzuladen weiterer Pakete um Funktionen erweitert werden. Die Entwicklergemeinschaft, die solche Plug-ins für den R-Commander schreibt, ist vergleichsweise groß, unter den hier vorgestellten grafischen Benutzeroberflächen ist der R-Commander diejenige, für die am meisten Plug-ins verfügbar sind. Eine Liste aller Plug-ins samt Installationshinweisen findet man auf CRAN mit dem Suchwort „RcmdrPlugin“. Ich empfehle einen Blick beispielsweise auf `RcmdrPlugin.KMggplot2`; es ermöglicht, innerhalb des R-Commanders auch Grafiken mit dem als modern geltenden Grafiksystem `ggplot2` anzufertigen. Für (sehr) Fortgeschrittene könnten die im Plug-in `RcmdrPlugin.FactoMineR` enthaltenen multivariaten Verfahren von Interesse sein.

### 15.7.6 Nachtrag: was bedeutet „in Faktoren wandeln“?

„Faktoren“ sind eine mögliche spezielle, für **R** typische Umgehensweise mit Zeichenketten (*strings*) sowie speziellen Zahlenarten. Denn Zeichenketten wie z. B. Namen können, so das Konzept, ganz unterschiedliche Informationsträger sein. Im Falle von beispielsweise Personennamen oder Fallbezeichnungen wie Grabnummern, Inventarnummern, Literaturreferenzen o. ä. bleiben sie individuelle Informationen, d. h. werden im Grunde meist nicht weiter ausgewertet. Andere Zeichenketten stehen jedoch für Informationen, die weiterverarbeitet werden wollen. So wird man die Zeichenketten „infans“, „juvenil“, „adult“, „matur“, „senil“ als anthropologische Altersbestimmungen auffassen, die Gruppen benennen, die man z. B. weiter nutzen und auswerten möchte, etwa indem man die Beigabenausstattung dieser Gruppen miteinander vergleicht. In diesem Fall kann man die Gruppen schon bei der Erfassung als Zahlen codieren, also mit 1 = infans, 2 = juvenil, 3 = adult usw. (die bei Bedarf ebenfalls zum „Faktor“ gewandelt werden können), oder eben das Angebot von **R** nutzen, die Zeichenketten in „Faktoren“ zu verwandeln. Stößt die Import-Routine etwa des R-Commanders auf Zeichenketten, fragt das Programm, wie es diese verstehen soll, als Zeichenketten im engeren Sinne oder als „Faktoren“. Oft ist das Wandeln in „Faktoren“ ein für das weitere Prozedere notwendiger Schritt. Indes: eine einmal getroffene Entscheidung ist keine Einbahnstraße, denn „Faktoren“ können auch nachträglich in Zeichenketten i.e.S. gewandelt werden und umgekehrt.

### 15.8 BLUESKY STATISTICS – EINE GUTE ALTERNATIVE ZUM R-COMMANDER

BlueSky Statistics – bisweilen auch BlueΣky geschrieben und im Folgenden kurz BlueSky genannt – ist wie der R-Commander eine grafische Benutzeroberfläche von **R**, aber deutlich jünger (ca. 2014/15 ff.) und noch weniger verbreitet. Man sieht ihr die Ausrichtung auf **R** weniger an als dem R-Commander, vielmehr tritt BlueSky wie ein selbstständiges Programm mit einer Point-and-Click-Oberfläche auf und versucht in vielerlei Hinsicht, dem altbekannten und weit verbreiteten Bezahlprogramm SPSS zu ähneln. Seine Idee ähnelt dem Ansatz von JAMOVI, vermittels einer modernen und einfach zu bedienenden grafischen Benutzeroberfläche das R-typische Arbeiten mit Kommandozeilen weitgehend überflüssig zu machen. Im Vergleich zu JAMOVI wirkt BlueSky im Zustand 2020 noch weniger elegant, ist jedoch weitaus umfangreicher und kann als ernsthafte Alternative zum R-Commander betrachtet werden. Anders als der R-Commander, der hinsichtlich der Grafiken ganz auf die traditionelle Grafik von **R** setzt (*base*), verwendet BlueSky im Standard das Grafikpaket *ggplot2*, das derzeit bei vielen R-Nutzern als moderner gilt. Aus meiner Sicht ist Bluesky derzeit

vor allem dann interessant, wenn Datenerfassung und Datenverwaltung in einem Projekt eine wichtige Rolle spielen, da es dazu im Ist-Zustand 2020 in Relation zu JAMOVI und auch zum R-Commander mehr Möglichkeiten umfasst und leichter bedienbar ist.

Wie z. B. der R-Commander und JAMOVI erlaubt es auch BlueSky, die zunächst per Point-and-Click zusammengestellten Prozeduren und Statistiken in einem eigenen Fenster auch im R-Code zu sehen und diesen Code auch zu editieren und auszuführen. Doch im Gegensatz zum R-Commander ist dieser Code in einen BlueSky-Rahmen eingebaut und isoliert nicht ohne Weiteres lauffähig. Gleichwie eignet sich diese Option ebenfalls zum sukzessiven Erlernen des Codens in **R**.

Neben der gänzlich kostenlosen Version von BlueSky (Anfang 2020: Version 6.20) gibt es kommerzielle Versionen, die sich jedoch nicht in der Software, sondern in dem im Kauf enthaltenen Supportangebot unterscheiden. Das Programm steht auf der Website von BlueSky Statistics zum freien Download zur Verfügung und kann wie heute üblich installiert werden (<https://www.blueskystatistics.com/> [24.1.2020]). Der bekannte R-Experte und Statistikberater Bob Muenchen hat BlueSky eingehend untersucht und mit anderen Benutzeroberflächen von **R** verglichen.<sup>39</sup> S. E. ist es die derzeit mächtigste und hinsichtlich ihrer Weiterentwicklung vielversprechendste grafische Oberfläche zu **R**. Muenchen hebt die vergleichsweise umfangreichen Möglichkeiten der Datenverwaltung hervor, die in BlueSky integriert sind. Neben dem R-eigenen Format importiert BlueSky Daten der Formate \*.csv, \*.txt, \*.xls, \*.xlsx, \*.sav (SPSS), \*.sas7bdat (SAS) und \*.dbf (dBase) und kann aus den heute gängigsten Datenbanken importieren (Microsoft Access; Microsoft SQL Server; MySQL; PostgreSQL; SQLite).

### 15.8.1 BlueSky installieren

BlueSky setzt nicht auf ein zuvor zu installierendes **R** auf, sondern bringt sein eigenes **R** gleich mit. Einziger Nachteil: Man braucht ggf. zwei Mal Platz für **R** auf seiner Festplatte. Die Installation erfolgt daher wie heute üblich: Programm von der Herstellerseite herunterladen und installieren; ggf. selbst eine Verknüpfung zur zentralen Startdatei „BlueSky.exe“ auf den Desktop o. ä. legen – das war’s. Wenn dabei oder bei dem ersten Arbeiten mit BlueSky Defizite gemeldet werden, liegt das an einzelnen

---

<sup>39</sup> Bob Muenchen (2018a). A Comparative Review of the BlueSky Statistics GUI for R. *Blog r4stats*, 21.6.2018: <http://r4stats.com/2018/06/21/bluesky/> [24.1.2020]. Bob Muenchen (2018b). BlueSky Statistics 5.40 GUI for R Update. *Blog r4stats*, 1.10.2018: <https://r4stats.com/2018/10/01/bluesky-statistics-6-04-gui-for-r-update/> [24.1.2020].

fehlenden oder inkompatiblen R-Paketen. Man sollte ggf. die Warnungen oder Fehlermeldungen sehr sorgfältig lesen und die in den Meldungen benannten Pakete in **R** (d. h. über das rohe **R** oder mit Hilfe von RStudio) nachinstallieren resp. aktualisieren. Beim Verfasser war dies mit Ruhe und Gelassenheit und ca. 20 Minuten Zeitaufwand erledigt, wonach BlueSky fehlerfrei und stabil lief.

### 15.8.2 BlueSky: erste Schritte

Mit einem Doppelklick auf das Programm-Icon – eine blaue Wolke – öffnen sich zwei Fenster: das Arbeitsfenster mit einer noch leeren Datentabelle und zwei Steuerleisten oben mit den vorhandenen Optionen und Prozeduren (sowie einem kleinen Icon „B<sub>A</sub>“ in der linken oberen Ecke) und das Output-und-Syntax-Fenster (mit einem kleinen Icon „B<sub>O</sub>“ links oben) (Abb. 26).

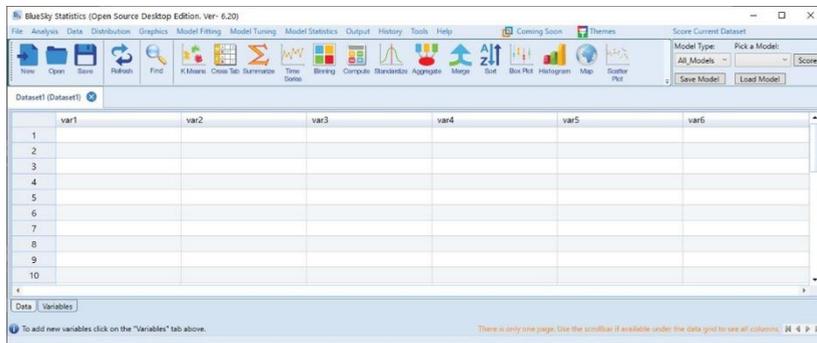


Abb. 26 Der leere Startbildschirm von BlueSky.

Das Einlesen von Daten findet sich ganz oben hinter „File“; => „Open“ usw., oder über die Schaltfläche „Open“ in der waagerechten Leiste von Symbolen (Schaltflächen). Wir laden die Dateien „Geldboersen.csv“ ein. Im Erfolgsfall (ansonsten s. oben Abschnitt „... und wenn’s nicht klappt mit dem Datenimport“ zum R-Commander) sieht man die Datentabelle nun, resp. die obersten ca. 10 Fälle. Die Datentabelle weist unten links zwei Reiter auf, „Data“ und „Variables“, zwischen denen man per Draufklicken hin- und herschalten kann. „Data“ sieht der Tabelle einer Tabellenkalkulation ähnlich, zeigt als oberste Zeile die Namen der Spalten und dann die Daten. Nach Wechsel zum Reiter „Variables“ zeigt das Fenster die Definition der einzelnen Datenspalten der Tabelle und erkennt „Name“ (den Variablennamen), „Label“ (die ggf. längere Beschreibung des Variablennamens), „Data Type“ usw. „Data Class“ ist wichtig,

hier wird die Art der Daten festgelegt – wir erinnern uns: z. B. die Frage, ob Zeichenketten als „string“ oder als „factor“ verstanden werden. Und ja: genau das kann man hier auch konvertieren, wie man es eben für die Analysen braucht.

Gönnen wir uns auch hier spielerisch ein paar schnelle Erfolgserlebnisse. Obere Symbol-Zeile „Histogramm“ – oder ganz oben: => „Graphics“; => „Histogramm“ – führt zum Aufpoppen eines Steuerfensters für Histogramme. Man schiebt „Jahr“ von links in das rechte Fenster, klickt unten „ok“ und erhält nach wenigen Sekunden im Output-Fenster das bestellte Histogramm. Wir wiederholen das Ganze, klicken dieses Mal aber im Steuerfenster zusätzlich rechts oben auf die Fläche „Options“: Es öffnet sich ein weiteres Steuerfenster, in dem man z. B. die Achsenbeschriftung festlegen kann. Man kann nichts kaputtmachen, daher ist es nützlich, sich im System Try-and-Error die diversen Möglichkeiten selbstständig zu verdeutlichen.

Zahlen statt Bildchen? Dafür ist die Schaltfläche „Summarize“ zuständig, resp. ganz oben „Analysis“; => „Summary Analysis“; => „Summary Statistics for All Variables“: Im Protokollfenster sehen wir jetzt (ähnlich zum analogen Wunsch an den R-Commander) die wichtigsten Statistiken zum gesamten Datensatz, u. a. welche Währungen wie oft vertreten sind, Minimum, Maximum und Mittelwert des Stichjahres, usw., usf.

Ähnlich wie beim R-Commander brechen wir an dieser Stelle nach einem ersten Kennenlernen von BlueSky ab, denn wir müssen jetzt die Ebene von Software und Skills wieder mit der Statistik und der Archäologie rückkoppeln und verbinden.

## 15.9 DIE R-INTERNE HILFE

Bevor wir in die produktive Arbeit eintauchen, ist es nützlich, sich die internen Hilfsangebote anzuschauen, die **R** mitbringt. Zunächst die Angebote des rohen **R**. Eher klassische Hilfen bietet das, was man nach Starten der R Console über die oberste Befehlsleiste angeboten bekommt: => „Hilfe“; => „Handbücher (PDF)“; ... das kostenfreie Angebot spricht für sich selbst, es kann das Selbststudium ohne das Kaufen von Büchern wesentlich stützen. Sodann gibt **Abb. 27** eine Übersicht über die Hilfen, die man aus der R Console heraus abrufen kann. Ein Teil der Hilfen wird gleich in der Konsole selbst ausgegeben, für die meisten Aufrufe mit etwas gründlicheren Texten öffnet sich aus **R** heraus ein eigenes Hilfefenster in Ihrem Browser.

Sucht man aus RStudio heraus nach Hilfe, stehen alle o. g. Funktionen ebenfalls zur Verfügung. RStudio hat das Hilfefenster fest in seiner Oberfläche rechts unten verankert (ggf. den Reiter „help“ aktivieren), was den Aufruf der „externen“ Texte erheblich beschleunigt. Aus dem Angebot der Befehlszeile „Help“ von RStudio ist

besonders erwähnenswert die Liste aller Tastenkombinationen („Shortcuts“), die später bei der direkten Kommandoingabe das Arbeiten sehr beschleunigen können. Besonders hilfreich ist die Befehlsergänzung und -kurzerläuterung, die RStudio einblendet, sobald man in der R Console Eingaben macht. Tippt man dort z. B. „mean“ ein, tauchen alle Funktionen auf, die mit „mean“ beginnen. Zusätzlich wird zu diesen jeweils ein Kurztext eingespielt, aus dem heraus man mit „F1“ weitere Hilfe aufrufen kann, die dann im Hilfefenster rechts unten auftaucht.

Befehlseingabe „Konsole“	Funktion
?help	öffnet eigenes Hilfefenster, gibt Überblick über die gesamten Hilfsfunktionen in <b>R</b> .
?mean	öffnet eigenes Hilfefenster, ruft Hilfe über die Funktion „mean“ auf: nennt alle Eingabeparameter etc.; funktioniert so mit allen anderen Funktionen in <b>R</b> .
help(mean) [oder] help(„mean“)	alternative Art des Aufrufs der Hilfsfunktion, führt auf die gleiche Seite wie „?mean“.
example(mean)	spielt ohne eigenes Hilfefenster in der R Console ein Beispiel für die Funktion mean aus.
apropos(„mea“)	hilft bei ungefähigem Wissen: Nennt innerhalb der R Console kurz alle Funktionen, in denen die Zeichenkette „mae“ vorkommt; beim Aufruf die An- und Abführungen beachten, die hier nötig sind.
help.search(„mea“)	öffnet eigenes Hilfefenster, nennt alle Funktionen in R-Paketen, in denen „mea“ vorkommt; Ausgabe zumeist (sehr) umfangreich, aber nützlich zum Stöbern.

**Abb. 27** Übersicht über die R-internen Hilfeangebote und -funktionen (R Console, ohne RStudio).

### 15.10 ZWISCHENBILANZ R, RSTUDIO, R-COMMANDER, BLUESKY

Das Arbeiten mit einer Befehlskonsole statt des gewohnten Auswählens per Mausclick aus sich sukzessive aufklappenden Menüs resp. Menüfolgen ist anfangs gewiss ungewohnt. **R** – vor allem in Verbindung mit der Oberfläche RStudio – bietet jedoch wirklich gute und schnell funktionierende Hilfen an, die auch Anfänger sicher durch das Arbeiten mit der Kommandozeile tragen. Klar ist, dass Anwender nach dem ersten Einüben in diesen Bedienmodus schneller zum Ziel kommen werden als über das gewohnte Point-and-Click Arbeiten. Indes: Die Point-and-Click-Oberfläche des R-Commanders oder von BlueSky liegt nahe an den Bediengewohnheiten und -erfahrungen der meisten heutigen PC-Benutzer, sie sind der m. E. niederschwelligere Einstieg in

das Arbeiten mit **R**. Um Frust zu verringern und schneller produktiv zu werden, empfehle ich für den Anfang diesen Weg, das Arbeiten mit einer graphischen Benutzeroberfläche.

### 15.11 SWIRL, YOUTUBE U. A.

Diejenigen, die sich aufgrund ihrer persönlichen Bedürfnisse entschieden haben, **R** zunächst nicht einzusetzen, haben nun genug Eindrücke von **R** gesammelt, sie können diesen Abschnitt überspringen. Für diejenigen, die ihre ersten Kenntnisse in **R** vertiefen möchten, mache ich auf „Swirl“ aufmerksam: die Möglichkeit, **R** in **R** interaktiv zu erlernen. Kein Muss, vielmehr eine Option, je nach Lerntyp. Denn wie man am liebsten lernt, ist individuell verschieden. Dieses Buch zielt auf Leser; es entspricht meinen persönlichen Bedürfnissen, weil ich nachschlagen, voraus- und zurückblättern können möchte und mir ein Buch neben Tastatur und Bildschirm guttut. Wegen der Verschiedenheit der Lerntypen möchte ich jedoch die vielen Einführungen zu **R** und einzelnen Komponenten von **R** nicht unerwähnt lassen, die man bei YouTube findet. Ich persönlich tue mich schwer mit dem Lernen anhand von Videos, mir fehlt das Vor- und Zurückblättern, das gezielte Nachschlagen-Können. Aber: Die Videos bieten visuell mehr als „nur“ mühsame textliche Beschreibungen, wo man jetzt hineinklicken sollte (samt Suchen, denn manchmal findet man die vom Autor gemeinte Stelle ja nicht sofort). Denn bei den Videos führen die Lehrenden die Aktionen einfach vor und man sieht die Wirkungen direkt am Bildschirm. Zwischen Buch und YouTube siedle ich den Weg von Swirl an: Man lädt Swirl als Paket resp. eine Folge von Paketen in **R** und wird dann in einem Frage- und Antwortspiel per Mitmachen in den jeweiligen Bereich eingeführt. Wer sich ernsthafter mit **R** beschäftigen möchte, sollte sich Swirl zumindest einmal anschauen und den ersten Schritt ausprobieren, um für sich persönlich zu prüfen, ob nicht auch das ein möglicher Weg des Lernens und Einarbeitens sein könnte.

„Swirl“ ist ein **R** Paket, es wird also so wie jedes andere Paket auch zuerst installiert, dann geladen. RStudio ist keine Voraussetzung, Swirl funktioniert auch im rohen **R**, aber die Swirl-Lehrenden empfehlen, es von RStudio aus zu installieren und zu starten. Mehr muss ich nicht erklären, denn wie das Installieren und Laden von Paketen geht, haben wir bereits anhand des R-Commanders gelernt.

## 16 TEIL III: EINFÜHRUNG IN DIE ANGEWANDTE STATISTIK

---

Den zweiten Teil dieser Einführung können wir nun abschließen, die wesentlichen Werkzeuge sind kennengelernt. Nun wenden wir uns im Schwerpunkt dem Thema angewandte Statistik an archäologischen und anthropologischen Problemen zu.

## 17 DIE GRAFISCHE DARSTELLUNG VON DATEN

---

Statistische Grafiken werden überschätzt, statistische Grafiken werden unterschätzt. Gerade Anfänger neigen dazu, bei ihren Auswertungen vor allem auf Grafiken zu setzen. Blickt man hingegen in seriöse, vor allem naturwissenschaftliche Publikationen, sieht man, dass überwiegend auf Kennzahlen und Tabellen gesetzt wird und Grafiken eher selten sind. Das hat gute Gründe. Harte Fakten, statistische Parameter und Tests vermittelt man besser in Zahlen, nicht in Bildern. Andererseits können Grafiken helfen, komplexe Sachverhalte tatsächlich auf einen Blick treffend zu vermitteln, weshalb sie nützlich sind, auch in der Wissenschaftskommunikation. Zudem helfen Grafiken gerade am Anfang einer Analyse, schnell in die Eigenschaften der untersuchten Daten hineinzufinden. Sie können ein ideenbringendes Kreativwerkzeug sein. Aber: Es sollten die richtigen, zu den Daten und Fragestellungen passenden Grafiken sein, nicht ein beliebiger Griff in den Fundus des Repertoires Ihres Office-Programms. Lassen Sie mich an den Anfang ein paar Empfehlungen setzen.

Wissenschaftliche Grafiken sind, mehr als statistische Methoden und Maßzahlen, auch von sich weiterentwickelnden technischen Möglichkeiten sowie Zeitgeschmack und Moden abhängig. In der Literaturliste am Ende des Bandes sind vier Titel genannt, die sich speziell mit dem Thema grafische Darstellungen befassen.

### 17.1 EMPFEHLUNGEN ZU STATISTISCHEN GRAFIKEN

*Empfehlung 1:* Grafiken am Computerbildschirm, an dem man arbeitet, sehen anders aus als gedruckte Bilder oder solche, die in einer Präsentation verwendet werden – allzu oft stimmen die Proportionen zwischen Schriftgrößen und Bildelementen bei einer Präsentation nicht mehr, Farben passen nicht, usw. Daher: Grafiken, die „fertig“ sind, immer jenseits des Arbeitsplatzbildschirmes testen: für den Druck auf Papier geeignet, für eine Präsentation geeignet?

*Empfehlung 2:* Setzen Sie in Ihren Publikationen Grafiken sparsam und sehr gezielt ein. Oft sind Zahlen im Text oder als Tabelle hinreichend und exakter. Verwenden Sie möglichst die üblichen Standardgrafiken, seien Sie sparsam mit Extravaganzen.

*Empfehlung 3:* Eine Abbildung (und auch eine Tabelle) sollte zusammen mit ihrer Abbildungsunterschrift für sich stehen, d. h. in ihrer Grundaussage auch ohne Lektüre des Fließtextes verständlich sein. Eine treffende Abbildungsunterschrift (Legende) ist daher ein wesentlicher Bestandteil jeder grafischen Darstellung. Dazu sind auch die Darstellungen selbst hinreichend (und lesbar) zu beschriften (Achsentitel etc.). Statt die Grafik mit Beschriftungen und anderen Informationen zu überlasten, ist es meist geschickter, viele oder einige der unabdingbar nötigen Zusatzinformationen nicht direkt in die Grafik zu packen, sondern in die Abbildungsunterschrift.

*Empfehlung 4:* Auf Manipulationen an den Skalen (Achsen) verzichten. Skalen sollten einen echten Nullpunkt haben und ihn zeigen. Skalen sollten, bis auf explizite und begründbare Ausnahmen, linear skaliert sein.

*Empfehlung 5:* Wählen Sie Darstellungen, die dem Skalenniveau der Daten angemessen sind. Zu Nominalskalen gehören Stab-/Balkendiagramme, zu Intervallskalen Histogramme. Sind die Daten einer Intervallskala stetig (wie z. B. Körperhöhe), stoßen die einzelnen Säulen eines Histogramms dicht an dicht aneinander; liegt eine diskrete Skala vor (es gibt Abstände zwischen den Werten), haben die Balken Abstand zueinander (Stab-/Balkendiagramm). Aus jeder Grafik oder ihrer Abbildungsunterschrift sollte die Grundgesamtheit ablesbar sein.

*Empfehlung 6:* Verzichten Sie, wiederum bis auf Ausnahmen, auf die sog. Kreis-/Torten-/Kuchendiagramme, sie sind nur selten angemessen. Informationen auf dem Niveau einer Nominalskala, die zu einem Kreisdiagramm zusammengefasst werden, sollten sich tatsächlich zu einem gemeinsamen Sachverhalt auf 100 % ergänzen. Bitte denken Sie jeweils darüber nach, welcher Umgang mit fehlenden Werten angemessen ist. Bei Kuchendiagrammen dient bereits die Größe der Tortenstücke zu Visualisierung der relativen Anteile; wählen Sie daher keine Prozentzahlen zu deren Beschriftung, sondern die tatsächlich beobachteten Häufigkeiten.

*Empfehlung 7:* Bitte verzichten Sie gänzlich auf 3D-Effekte von tatsächlich ein- oder zweidimensionalen Aussagen. „Dreidimensionale“ Kreisdiagramme, Stäbe/Blöcke etc. verwirren mehr als sie klären, und sie geben Raum für allerlei Suggestion (durch die willkürlich ansetzbare Tiefe des 3D-Effekts). In wissenschaftlichen Publikationen sind sie zu vermeiden.

*Empfehlung 8:* Bei Histogrammen werden intervallskalierte Daten – wie auch ordinalskalierte Daten mit vielen Zuständen – für das Zeichnen von Histogrammen zu Klassen zusammengefasst. Das Zusammenfassen zu Klassen hat keine festen Regeln und bietet viel Spielraum. Die wesentlichen flexiblen Parameter sind (a) die Klassenweite und (b) der Start der Klassensetzungen (der „Nullpunkt“). „Klassenweite“ bedeutet z. B. beim Thema Körperhöhe, dass man alle Angaben von z. B. 150-154 cm, 155-159 cm, 160-164 cm zu jeweils einer Säule zusammenfasst (i.e. Klassenweite 5 cm), oder etwa alle Angaben von 150-159 cm, 160-169 cm, usw. zu einer Säule (i.e. Klassenweite 10 cm). Mit der Klassenweite legt man also fest, ob das Bild mehr detailliert oder stärker generalisierend ausfällt.

Man vermeide – bis auf begründete Ausnahmen – innerhalb eines Diagramms unterschiedliche Klassenweiten. Dieser (schlechte) Weg wird gerne gewählt, wenn es z. B. einige wenige Beobachtungen mit hohen Werten gibt, weshalb dann manchmal – Beispiel Körperhöhe – eine letzte Klasse mit beispielsweise 200 – 225 cm (Klassenweite 25 cm) gesetzt wird. Aus Statistikersicht unfein!

Neben der Wahl der Klassenweite bestimmt der Startpunkt das Aussehen der Histogramme stark, also der Wert, mit dem die erste (kleinste) Klasse beginnt. Oft ergibt der kleinste beobachtete Wert – auf eine ganze Zahl/einen glatten Wert abgerundet – einen guten Startwert des Histogramms, und der größte beobachtete Wert – auf eine ganze Zahl/einen glatten Wert aufgerundet – einen guten Endwert des Histogramms. Als Faustregel gilt: Die Wurzel aus der Menge an Beobachtungen (Grundgesamtheit) ergibt eine sinnvolle Anzahl an Klassen, woraus sich dann die Klassenweite ergibt. Trotz dieser „guten Praxis“ gibt es bei Histogrammen einen nicht vermeidbaren subjektiven Moment.

Indes: Es kann gute Gründe geben, von diesen „goldenen Regeln“ abzuweichen, z. B. wenn man zwei oder drei Histogramme so skalieren will, dass man sie optimal nebeneinanderstellen und vergleichen kann. Dann wird man eine übergreifend gute Klasseneinteilung wählen, die für diese Histogramme gleich ist, wiewohl sie eventuell für ein einzelnes Bild nicht optimal ist.

### 17.1.1 Schätzung der empirischen Dichtefunktion

Eine sinnvolle Ergänzung oder Alternative zu den notwendigen Klasseneinteilungen von Histogrammen ist die sog. „Schätzung der empirischen Dichtefunktion“ mit Hilfe von Kernschätzern. Das ergibt hübsche Kurven und ist in der Praxis nicht ganz so kompliziert wie es klingt. Wir haben dieses Verfahren hier für die Erzeugung von **Abb.**

84 eingesetzt. Aber: Auch bei diesen scheinbar „objektiveren“ empirischen Dichtefunktionen werden im Hintergrund subjektive Entscheidungen getroffen (ggf. automatisch von der gewählten Software), und zwar (a) über die Art des zugrunde gelegten Kerns und (b) über die Kernweite. Die Wahl der Kernweite bestimmt, ob der Kurvenverlauf unruhiger ausfällt (geringe Kernweite) oder glatter, stärker generalisierend (größere Kernweite). Sprich: Diese Kurven wirken objektiver als Histogramme, sie sind es in Wirklichkeit nicht (völlig). Dennoch einen Versuch wert. Für an der Theorie Interessierte: Silverman (1986) erklärt alles Nötige und ist m. E. gut verständlich – und wir kommen später noch einmal auf dieses Thema zurück.

## 17.2 BEISPIEL-GRAFIKEN

Praxis statt Theorie: Im folgenden Abschnitt werden ein paar wichtige Typen seriöser wissenschaftlicher Grafiken vorgestellt, und zwar ausgehend von unseren Beispieldaten. Welche darin vorhandenen Informationen werden am besten wie dargestellt? Ein Anspruch auf Vollständigkeit ist nicht gegeben, alle vorgestellten Lösungen sind gute Lösungen, doch möglicherweise gibt es auch bessere Lösungen – ein Schuss Subjektivität bleibt. Wichtig ist auch zu sehen, dass es zwar einige DOs und DON'Ts gibt, aber zugleich ein weites Feld subjektiver Spielräume. Schriften sollten klar und lesbar sein, aber welche Schrift man wählt, ist meist offen. Manche Wissenschaftler und manche Zeitschrift mögen Farbe, andere weniger. Mancher denkt bei seiner Farbvergabe auch an typische Leser-Probleme, wie z. B. an die vor allem unter Männern verbreitete Rot-Grün-Blindheit, und kombiniert daher Farben stets mit einprägsamen Mustern, andere haben dies weniger im Blick oder bevorzugen Lösungen im Milieu von Schwarz, Weiß und Grau. Beim Blättern durch die Literatur wird zudem deutlich: Es gibt sich im Laufe der Zeit wandelnde Moden auch in der Art der Grafiken resp. ihrer Gestaltung. Weitverbreitete Statistik-Programme haben ihren spezifischen Grafik-Stil und schaffen dadurch – bei aller wissenschaftlichen Nüchternheit sowie Steuerbarkeit der Ausgaben – auch einen Wiedererkennungswert der Software; so haben statistische Grafiken von SPSS ihren spezifischen Stil, solche von **R** einen ebenso gut erkennbaren eigenen Stil.

Das Anfertigen von Grafiken ist stets ein guter persönlicher Einstieg in die Datenanalyse: sich einen ersten Eindruck vom Material verschaffen. Man nähert sich unmittelbar nach der Erfassung der Daten etwas spielerisch – oder vornehmer ausgedrückt: exploratorisch – dem Material, den Fragestellungen und auch dem Umgang mit der jeweiligen Software. Oft stößt man bei diesem Sichten noch auf Eingabefehler

und nutzt beispielsweise die gerade anhand von Grafiken schnell erkennbaren Ausreißer zur Korrektur von fehlerhaften Daten. Für viele Menschen eignen sich Grafiken mehr als Tabellen und Kennzahlen für eine erste Annäherung an die Daten und deren Auswertung. Mancher entdeckt hier leichter auch unerwartete Ergebnisse als später an den stark fragestellungsgeleitet erzeugten Kennzahlen und nutzt das Spielen mit Grafiken als eine Art Kreativabteilung. Obwohl ich öfter betone, dass für die spätere Analyse und auch Publikation eher Kennzahlen und Tabellen im Vordergrund stehen und bei den Grafiken die reflektierte Wahl der angemessenen Grafik wichtig ist, die zudem gut gestaltet sein sollte, darf und sollte man anfangs etwas mit den Grafiken spielen. Es muss ja nicht alles davon auch in die spätere Schriftfassung übernommen werden.

### 17.3 BALKENDIAGRAMM UND KREISDIAGRAMM

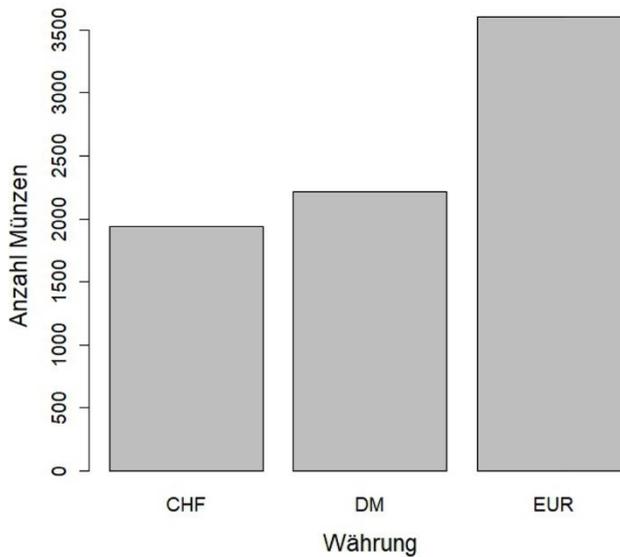
Für das praktische Nachvollziehen des Folgenden kann man z. B. RStudio starten, dort via Reiter „Packages“ im Feld rechts den R-Commander laden (einfach Häkchen an das Feld „Rcmdr“) und dort den Datensatz „1\_Muenzen“ einlesen: => Datenmanagement; => Importiere Daten; => aus Text, Zwischenablage oder URL; in dem neu aufklappenden Fenster die passenden Angaben machen, d.h. die Voreinstellungen wo nötig verändern: Datensatz als (z. B.) „Geldboersen“ bezeichnen, als Datenfeldtrennzeichen statt der Voreinstellung Leerzeichen richtig „Strichpunkt [;]“ anklicken, und mit OK [ENTER] plus anschließender Dateiauswahl einlesen. Stets eine Kontrolle machen, z. B. mit „Datenmatrix betrachten“, ob der R-Commander den Datensatz auch richtig verstanden hat. Als zweite Prüfung könnte man z. B. via => Statistik; => Deskriptive Statistik; => Aktive Datenmatrix ganz einfache Übersichtsstatistiken bestellen. Sie werden in der R Console, d. h. im Feld unten links von RStudio ausgespielt. Sie müssen nicht (noch nicht) vertiefend gelesen werden, sondern zeigen an, dass/ob der Datensatz sinnvoll gelesen wurde.

Der Beispieldatensatz „1\_Muenzen“ enthält verschiedene nominalskalierte Variablen wie Währung, Land und Prägeort. Sie sollten zunächst als Tabelle ausgegeben werden mit einer Spalte, welche die beobachteten Häufigkeiten nachweist und einer Spalte, die diese Häufigkeiten in Prozentzahlen umsetzt sowie unten einer Summenzeile, welche die Gesamtsumme der Beobachtungen ausweist (**Abb. 28**). Im R-Commander also => Statistik; => Deskriptive Statistik; => Häufigkeitsverteilung; „Währung“ auswählen & OK [Enter]. Eine zusätzliche grafische Darstellung hat m. E. nur einen geringen Mehrwert. Wenn, kommen ein Balkendiagramm (*barchart*) (**Abb. 29**) oder ein Torten-/Kreisdiagramm (*pie chart*) (**Abb. 30**) in Frage. Bei einem Balkendiagramm haben die einzelnen Säulen, ganz dem Skalenniveau entsprechend,

einen Abstand zueinander; grafisch sind die Abstände gleich, wiewohl sie von der Sache her nicht definiert sind. Bei einem Balkendiagramm steht jedes Merkmal für sich, ein Kreisdiagramm betont, dass sich alle Merkmale gemeinsam auf eine Hundertsumme ergänzen.

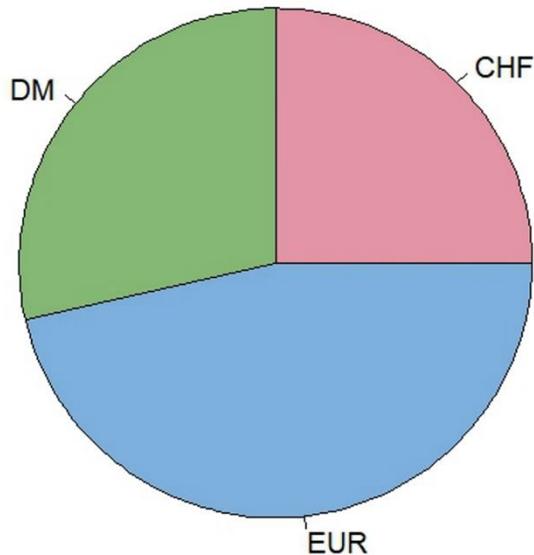
Währung	n	%
DM	2.218	28,6
CHF	1.937	25,0
Euro	3.605	46,5
Gesamt	7760	100,0

**Abb. 28** Häufigkeit der Währungen im Datensatz „1\_Muenzen“ (n=7.760).



**Abb. 29** Balkendiagramm (*bar chart*) der Währungen im Datensatz „1\_Muenzen“ (n=7.760).

## Anteile Währungen



**Abb. 30** Kreisdiagramm (Tortendiagramm, *pie chart*) der Währungen im Datensatz „1\_Muenzen“ (n=7.760).

Persönlich bevorzuge ich das Balkendiagramm **Abb. 29** gegenüber dem Kreisdiagramm **Abb. 30**. In dem Balkendiagramm wurde auf die Einspiegelung der beobachteten Häufigkeiten oder der prozentualen Anteile verzichtet, weil diese Information über die y-Achse recht gut ersichtlich ist. Sollen die Informationen exakter nachgewiesen werden, braucht man eine Tabelle wie **Abb. 28** an Stelle der Grafik oder ergänzend zu ihr.

*Praxistipp.* Beim Arbeiten mit dem R-Commander werden die per Point-&-Click generierten Codezeilen in der Konsole protokolliert. Die Zeile dürfte in etwa so aussehen:

```
with(muenzen, Barplot(Waehrung, xlab = "Waehrung", ylab = "Frequency"))
```

Der R-Commander bietet die Möglichkeit, nach => Grafiken; => Balkendiagramm, und Auswählen von „Währung“, den Reiter „Optionen“ zu bedienen, wo sich u. a. die Beschriftungen („Label“) der x- und y-Achse setzen lassen. Auf diese Weise wurden in unserem Beispiel die Beschriftungen mit deutschen Umlauten generiert:

```
with(muenzen, Barplot(Waehrung, xlab = "Währung", ylab = "Anzahl Münzen"))
```

Mir persönlich sind jedoch die Beschriftungen relativ zum Diagramm zu klein. Im R-Commander kann dies nicht per Point-&-Click geändert werden, aber händisch, indem man die Befehlszeile manuell erweitert, dann komplett mit dem Mauszeiger markiert und über die Schaltfläche (links unten) „Befehl ausführen“ aktiviert. Das Balkendiagramm **Abb. 29** beruht auf folgenden Spezifikationen:

```
with(muenzen, Barplot(Waehrung, xlab="Währung", ylab="Anzahl Münzen",  
cex.axis=1.2, cex.names=1.2, cex.lab=1.4))
```

Per Versuch & Irrtum lassen sich so schnell die Proportionen finden, die zu Ihrem Wunschiagramm passen. Die Befehls-elemente „cex.“ steuern die Textgrößen relativ zu den Standardvorgaben, d. h. sind ein Vergrößerungs- und ggf. auch ein Verkleinerungsfaktor. Dieses und Weiteres erfährt man z. B., wenn man in RStudio im Fenster rechts unten den Reiter „Help“ aktiviert und „barplot“ ins Suchfenster eingibt.

Das Kreisdiagramm **Abb. 30** entsteht wieder mit Hilfe des R-Commanders: => Grafiken; => Kreisdiagramm, usw., und anschließender manueller Anpassung der resultierenden Code-Zeile:

```
library(colorspace, pos = 17)  
with(muenzen, pie(table(Waehrung), labels = levels(Waehrung),  
xlab = "", ylab = "", main = "Anteile Währungen",  
col = rainbow_hcl(3), cex=1.3))
```

Hier sind die relativen Anteile der drei Währungen durch den Umfang der Kreissegmente visualisiert. Statt, wie leider oft üblich, als Beschriftung der Segmente die Prozentzahl anzugeben (also eine redundante Information), sollte man ggf. die tatsächlich beobachteten Häufigkeiten als Beschriftung einsetzen – was jedoch mit dem R-Commander nicht resp. nur mit beträchtlichem Aufwand möglich ist. Gerade im Falle weniger Beobachtungen bevorzuge ich diese Lösung statt der Prozentzahlen, weil dann neben dem klaren visuellen Eindruck von den Anteilen eben stets auch die ggf. begrenzte Basis der Aussage sichtbar bleibt. Versucht man eine (zu) kleine Datenbasis

mit Hilfe des Einspiegeln von Prozentzahlen statt absoluter Häufigkeiten zu kaschieren, werden die Profis dies schnell erkennen und beginnen, Ihren Ausführungen ganz allgemein zu misstrauen. Doch solange man die Daten zusätzlich in einer Tabelle wie **Abb. 28** dokumentiert, kann auf Beschriftungen im Sinne eines Datennachweises verzichtet werden.

*Einschub: Etwas Datenmanagement wegen der Prägeorte, „Faktoren“, ...*

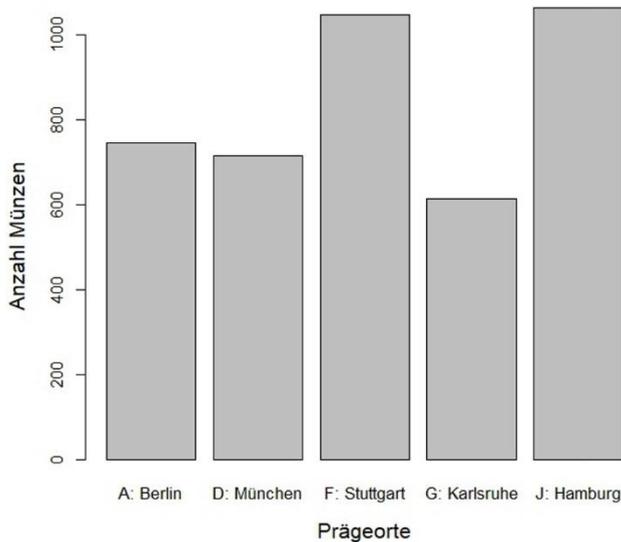
Will man ähnliche Statistiken und Grafiken für die Prägeorte erzeugen, stößt man im R-Commander auf das Problem, dass die Variable „Ort“ gar nicht im Angebot der entsprechenden Menüs auftaucht. Warum? Weil wir diese Information als Zahl abgelegt hatten, weshalb **R** und der R-Commander sie jetzt auch als Zahl verstehen und man daraus z. B. einen Mittelwert rechnen könnte, was indes inhaltlich wenig Sinn macht. Daher muss die Variable zunächst in einen „Faktor“ umgewandelt werden, wie wir es oben bereits erklärt hatten. Wie? => Datenmanagement; => Variablen bearbeiten; => Konvertiere numerische Variablen in Faktoren; „Ort“ auswählen und bestätigen; in dem aufpoppenden Fenster nun die Bedeutung der Zahlen eintragen, d. h. für 1 „A: Berlin“, 2 „D: München“, 3 „F: Stuttgart“, 4 „G: Karlsruhe“, 5 „J: Hamburg“ & bestätigen. Danach steht „Ort“ für das Gewünschte zur Verfügung: => Statistik; => Deskriptive Statistik; => „Häufigkeitsverteilungen“; „Ort“ aussuchen und mit OK bestätigen.

Danach stehen uns die nötigen Zahlen in der R Console (links unten) zur Verfügung (**Abb. 31**). Wie stets muss frühzeitig, d. h. nach einem ersten Sichten der Daten, das Thema der fehlenden Werte (*missing values*; in **R**: NA) beachtet und oft auch entschieden werden. Im vorliegenden Fall haben nur die in Deutschland geprägten Münzen auch Prägeorte, weshalb diese Angaben für knapp die Hälfte der Münzen unbekannt ist. Es ist zu klären, was sinnvollerweise die Bezugssumme der Prozentwerte ist. In der Tabelle **Abb. 31** haben wir zu Veranschaulichung des Themas beides aufgeführt: die Prozentwerte bezogen auf die Gesamtzahl der Münzen, und die Prozentwerte bezogen auf die Zahl der gültigen Beobachtungen. Im vorliegenden Fall ist es inhaltlich sinnvoller, die Prozentzahlen auf die Münzen mit bestimmtem Prägeort zu beziehen ( $n=4.215$ ) und nachfolgend mit der Spalte „gültige %“ zu operieren.

Prägeort	n	% (total)	gült. %
A: Berlin	745	9,6	17,8
D: München	716	9,2	17,1
F: Stuttgart	1046	13,5	25,0
G: Karlsruhe	615	7,9	14,7
J: Hamburg	1064	13,7	25,4
unbek./andere	3574	46,1	
Gesamt	7760	100,0	100,0

**Abb. 31** Tabelle mit den Häufigkeiten der Prägeorte im Datensatz "1\_Muenzen".

Solche Entscheidungen sind immer wieder notwendig, explizit zu begründen und dann auch auf die Grafiken zu übertragen. Die Entscheidung für die Einschränkung auf die gültigen Prozente ist nicht immer sinnvoll, oft z. B. soll die inkludierende Darstellung der Unbestimmten – und deren Aufnahme in die Hundertsomme – auch dazu dienen, Unsicherheiten in den Daten deutlich offen zu legen.



**Abb. 32** Balkendiagramm zu den Prägeorten (n=4.186).

Nun liefern wir mit **Abb. 32** noch das gewünschte Balkendiagramm nach. **R** hat uns die Entscheidung zum Thema der fehlenden Werte abgenommen und bezieht sich

beim Balkendiagramm nur auf die gültigen Beobachtungen (was allerdings mit etwas Aufwand auch anders dargestellt werden könnte).

Mit ähnlichen Tabellen und Grafiken würde man die nominalskalierten Variablen in den Datensätzen „Zwentendorf“ und „Howells-World-Data“ umsetzen, z. B. die Geschlechtsbestimmungen der Skelette und deren Altersklassen, z. B. die Variablen Geschlecht, Population und Gruppe bei Howells World Data.

### 17.3.1 Einschub: die gleichen Prozeduren mit BlueSky

BlueSky starten; Daten einlesen: => File; => Open; Datei aussuchen, hier besser das Format Tab-delimited wählen (\*.dat); im aufpoppenden Fenster das Tab als Trennzeichen auswählen & OK.

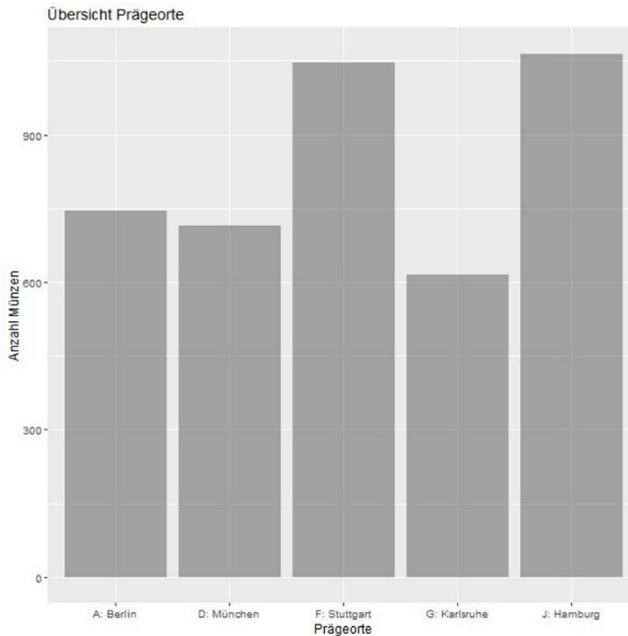
Die Daten sollten nun eingespiegelt werden, ggf. unten links Reiter „Data“ klicken. Wieder verwandeln wir die Variable Ort in einen „Faktor“: => Reiter „Variables“ anklicken, => Variable Ort mit rechter Maustaste anklicken und =>Make a factor“ wählen. Danach bei „Values“ von Ort auf das graue Kästchen klicken und in dem aufpoppenden Fenster die Werte („Label“) für die Nummern angeben. OK & fertig.

Einfache Statistiken: => Analysis; => Summary Analysis; => Frequency Table; im aufpoppenden Fenster „Ort“ in das rechte Fenster schieben und bestätigen. Im Output-Fenster werden die Zahlen und Prozente ausgespielt, hier zunächst einschließlich der fehlenden Werte, dann unter „Valid Percent“ ohne die fehlenden Werte berechnet.

Nun das Balkendiagramm: => Graphics; => Bar Chart, Counts; in dem aufpoppenden Fenster „Ort“ nach rechts schieben als „X variable“ (i.e. das Feld in der Mitte, nicht oben), mit OK bestätigen, fertig. Wenn man zuvor rechts oben „Options“ klickt und z. B. die Felder zur x- und y-Achse ausfüllt z. B. mit „Prägeorte“ und „Anzahl Münzen“, erhält man auch sinnvolle Achsenbeschriftungen. Gleichwie: Das fertige Balkendiagramm inkludiert die NA – fehlenden Werte, was im gegebenen Fall nicht erwünscht ist. Lösung? Ad hoc eine Untermenge der Daten anlegen, die nur die gültigen Fälle enthält, und für diese die gewünschte Grafik erstellen: => Data; => Missing Values; => Remove NAs; im Fenster rechts oben „New Dataset“ wählen und einen treffenden Namen eintragen, z. B. „temporaer“, dann „Ort“ von links in das rechte Fenster schieben und mit OK bestätigen. Im Hauptfenster ist nun ein zweiter Datensatz angelegt, den man sich anschauen kann: er enthält nur die Variable „Orte“, und zwar ohne fehlende Werte. Via => Analysis; => Summary Analysis; => Frequency Table kann man dies kontrollieren, und mit => Graphics; => Bar Chart,

Counts; in dem aufpoppersnden Fenster „Ort“ nach rechts schieben, mit OK bestätigen, die gewünschte Grafik gewinnen (**Abb. 33**). Der Stil unterscheidet sich von den Grafiken zuvor; während der R-Commander die Basis-Grafiken von R zu Grunde legt, arbeitet BlueSky mit dem Grafikpaket *ggplot2*, das einen eigenen Stil hat. Inhaltlich zeigen beide Optionen jedoch das Gleiche.

Ihnen erscheint die Beschriftung zu klein? Mir auch – da ist händische Nacharbeit am Code selbst nötig, denn weder der R-Commander noch BlueSky enthalten dafür einfache Point-&-Click-Lösungen. Weil das etwas komplizierter ist, verschieben wir das Thema „auf später“.



**Abb. 33** Balkendiagramm der Prägeorte, erstellt via BlueSky mit dem R-Grafikpaket *ggplot2*.

## 17.4 HISTOGRAMM

Die Variable Prägejahr bildet eine quantitative Skala, genauer eine diskrete Intervallskala. Hier ist eine Tabelle der beobachteten Häufigkeiten pro Prägejahr in der Regel nicht sinnvoll, sie wäre sehr unübersichtlich, während als angemessene Art von Grafik ein Histogramm unmittelbar eine gute Übersicht über die Daten verschafft. In

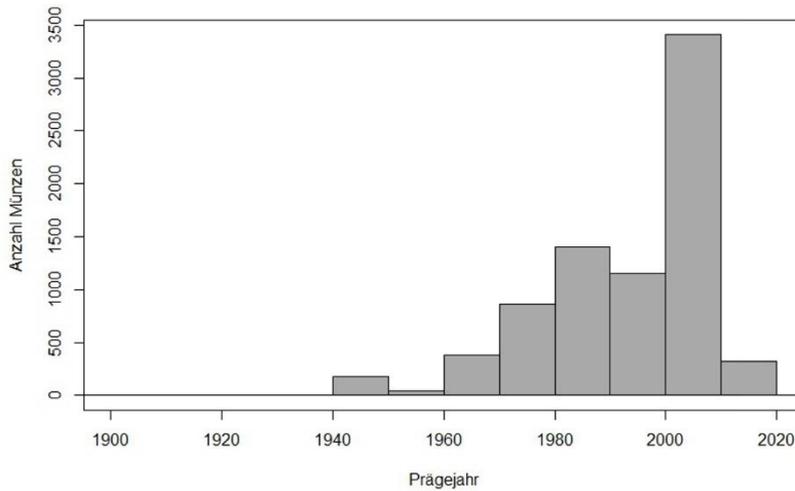
der Abbildungsunterschrift können dann geeignete statistische Kennzahlen (dazu später) die Übersicht vertiefen. Bei einem Histogramm stehen die einzelnen Balken dicht an dicht, weil bei der zugrunde liegenden Intervallskala alle Werte besetzt sind resp. besetzt sein könnten. Die einzelnen Beobachtungen werden jeweils innerhalb der zu gewählten Klassenweite zu einem Wert, zu einem Balken zusammengefasst.

Der Unterschied zwischen Balkendiagramm und Histogramm wird am besten an einer Zahlenreihe deutlich. Liegen Häufigkeitsbeobachtungen vor für 1, 2, 3 sowie 6, 7 und 8, werden im Balkendiagramm eben diese sechs Balken gesetzt, d. h. die Variablen als Nominaldaten aufgefasst. Der Abstand zwischen den Balken für 3 und 6 wäre wie so groß jener zwischen 2 und 3 oder für 6 und 7. Ein Histogramm hingegen nähme diese Daten als Teil einer quantitativen Skala wahr und würde auch die Ausprägungen 4 und 5 wiedergeben, eben mit der Häufigkeit Null; in Konsequenz sähe man eine deutliche Lücke zwischen den Balken für 3 und 6. Wenn beides technisch möglich ist, denke man vorab darüber nach, was den Daten und Fragen entspricht: Balkendiagramm oder Histogramm.

Im gegebenen Fall ist das Histogramm durch die Tatsache geprägt, dass es in unserer Serie einige wenige recht alte Münzen gibt, die älteste aus dem Jahr 1903. Sie führen dazu, dass die Skala für die Masse der jüngeren Münzen relativ stark gestaucht, d. h. zusammengefasst wird (

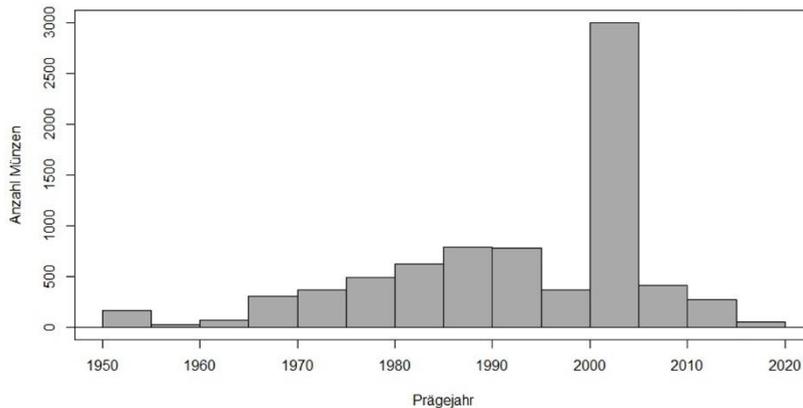
**Abb. 34**). Man kann (besser: sollte) daher erwägen, die Ausnahmen, in der Statistik „Ausreißer“ genannt, aus der Grafik auszuklammern, um die Masse der Daten angemessener darzustellen. Andererseits ist möglicherweise genau dies, nämlich das Vorhandensein einiger relativ alter Prägungen, auch eine erwünschte Botschaft und das Histogramm wie in

**Abb. 34** daher sinnvoll. Daher neige ich im vorliegenden Fall dazu, mit zwei Histogrammen zu operieren, einem, das den gesamten Datensatz umfasst, und einem zweiten, das – je nach Fragestellung – wie eine Ausschnittsvergrößerung die Münzen erst ab ca. 1950 (wg. Währungsreform in Deutschland im Jahr 1949) oder ca. 2000 (Einführung des Euro im Jahr 2002) umfasst.



**Abb. 34** Histogramm der erfassten Prägejahre (n=7.757). Minimum 1903, Maximum 2019, Mittelwert 1992,7, Median 1999.

In dieser Ausschnittsvergrößerung (**Abb. 35**) dominiert das Jahr 2002 (also die Säule 2000-2004), in dem der Euro emittiert wurde, die übrigen Zeitabschnitte – aber dies ist eine historische Tatsache und daher kein Fall für eine weitere geschickte Fallauswahl oder Skalenmodifikation.

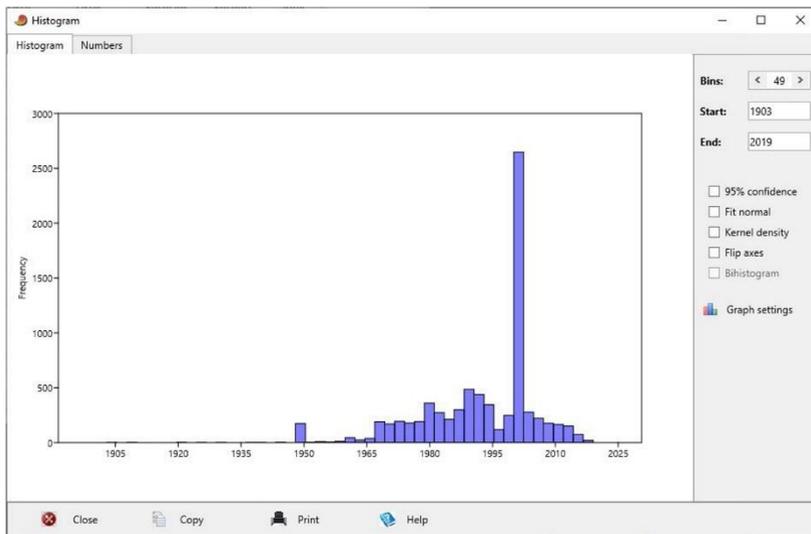


**Abb. 35** Histogramm der erfassten Prägejahre ab 1950 (n=7725, d. h. ohne 32 Münzen aus den Jahren 1903-1949).

Gerade für das Beispiel der Prägejahre sei an die oben angeführte Empfehlung 7 erinnert: Das Aussehen von Histogrammen ist stark von der gewählten Klassenbreite (hier: 5 Jahre) und dem Startwert (hier: 1950) abhängig, und dadurch auch in gewissem Maße manipulierbar. Tut man nichts, hat die Automatik der Software eine hoffentlich gute Entscheidung getroffen, d. h. man ersetzt das bewusste eigene Eingreifen durch die unkontrollierte Entscheidung der Software. Also: Nach einer ersten Sichtung der Daten sollte man sich bewusst für die Setzung der genannten Parameter entscheiden und über Startwert und Klassenweite bewusst entscheiden.

### 17.4.1 Praxis PAST: Einstellen von Klassenweite und Startwert bei Histogrammen

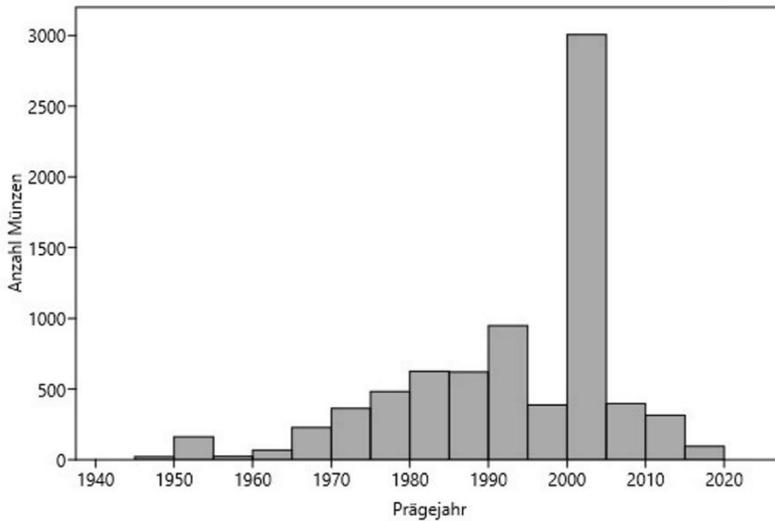
Wie bereits erlernt, bestelle man ein Histogramm in PAST: Datenspalte markieren, z. B. Prägejahr, => Plot; => Histogramm. Wir erhalten ein Histogramm, so wie es PAST uns nach den Daten vorschlagen möchte (**Abb. 36**).



**Abb. 36** Unbearbeitetes Histogramm aus PAST für die Datenspalte Prägejahr.

PAST zeigt rechts oben das Minimum („Start: 1903“) und das Maximum („End: 2019“) unserer Datenreihe an; das unbearbeitete Histogramm und das uns schon bekannte Problem, dass die sehr wenigen sehr alten Münzen das Histogramm in jenem Bereich, der uns mehr interessiert, arg stauchen. Also greifen wir ein: Startwert auf 1945, Endwert (um eine „runde“ Zahl zu erhalten) 2020. PAST behält die Anzahl der

Klassen (*bins*) jedoch bei, was im vorliegenden Fall zu kleinteilig ist. Ansatz: Die Spannweite zwischen 1945 und 2020 umfasst 75 Jahre; schweben mir Klassenweiten von jeweils 5 Jahren vor, sind dies 15 Klassen; also stellen wir *Bins* auf 15 ein, und erhalten nach einem ENTER das Histogramm **Abb. 37**.

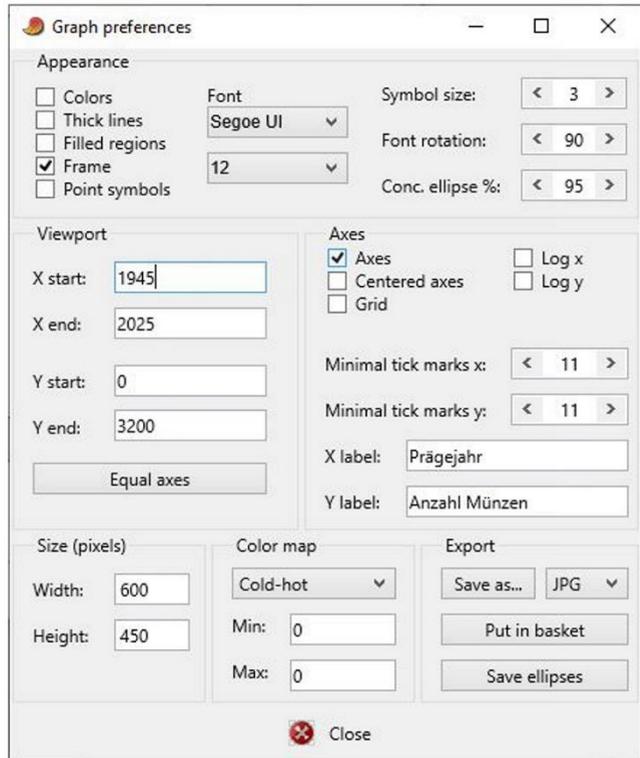


**Abb. 37** Mit PAST erzeugtes Histogramm der Prägejahre, nun mit bewusst gesetztem Startwert und Klassenweite von fünf Jahren. Wollte man das Bild publizieren, müsste man in der Abbildungslegende ausweisen, dass man Münzen älter als 1945 aus dem Bild ausgeklammert hat.

In der Forschungspraxis ist es sinnvoll mit diesen Möglichkeiten etwas zu „spielen“, um ein den Daten und der Fragestellung optimal angepasstes Bild zu gewinnen. „Der Appetit kommt beim Essen“: Bitte klicken Sie bei PAST auf die Schaltfläche „Graph settings“ (unterhalb des Histogramms), womit sich ein weiteres Menü öffnet (**Abb. 38**). Es zeigt erheblich mehr und tiefgreifendere Steuerungsmöglichkeiten für das Aussehen der Histogramme.

Es lohnt, die angebotenen Möglichkeiten einmal spielerisch auszutesten, weil man auf diesem Weg am schnellsten sieht und lernt, was PAST macht. Die beiden Abbildungen (**Abb. 37**, **Abb. 38**) zeigen ein bewusst stark überarbeitetes Beispiel von Steuerung und Ergebnis; man erkennt, wie flexibel man mit PAST die Histogramme so einstellen kann, wie man es sowohl inhaltlich als auch im Layout möchte. Nun ist klar: Die automatisch gesetzten Voreinstellungen sind nur ein Anfang, ein professionell gestaltetes Histogramm bedarf der Nachjustierung.

**Abb. 38** Die Histogrammsteuerung bei PAST. Beispiel einer Steuerfläche mit zahlreichen bewussten Setzungen zum Layout eines Histogramms; u.a. wurden hier die Anfangs- und Endpunkte der Skalen gesetzt, oben die Farben ausgeschaltet, die Schriftgröße auf „12“ angehoben. Rechts unten bei „Export“ lässt sich das gewünschte Dateiformat einstellen (hier „JPG“, ev. besser z. B. „TIF“) sowie der Dateiname („save as...“).

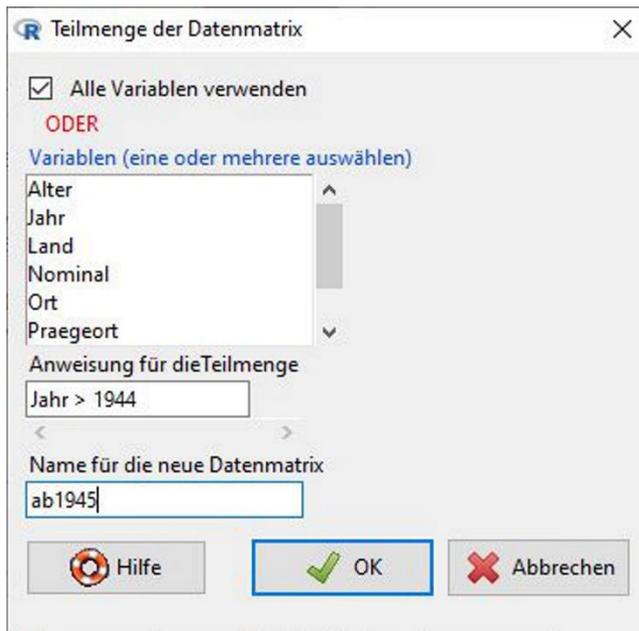


#### 17.4.2 Praxis R-Commander: Einstellen von Klassenweite und Startwert bei Histogrammen

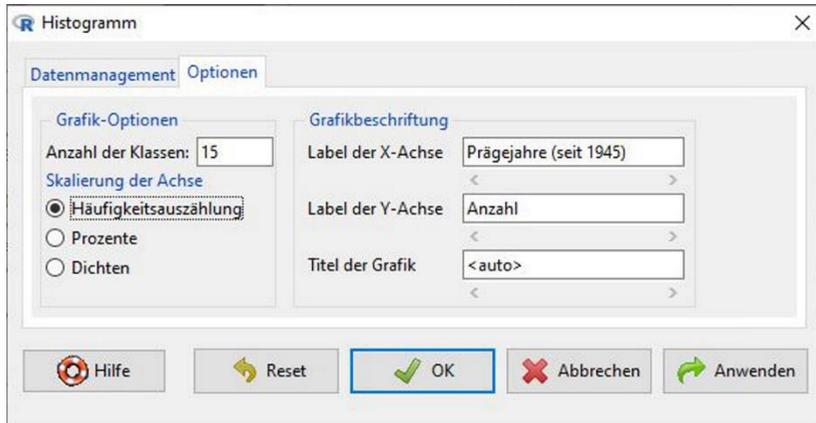
Das Anlegen eines Histogramms mit dem R-Commander ist nicht minder einfach: Nach dem Laden der Daten: => Grafiken; => Histogramm, Variable auswählen & OK. Das resultierende Histogramm ist von **R** schon automatisch recht passend skaliert, eigentlich bedarf es kaum weiterer Eingriffe.

Die (beim R-Commander: wenigen) Steuerungsmöglichkeiten liegen hinter dem Reiter „Optionen“, also: => Grafiken; => Histogramm; beim Reiter „Datenmanagement“ (wie bisher) die Variable auswählen, dann den Reiter „Optionen“ aktivieren. Hier gibt es Fenster, um die Beschriftung der Grafik und der Achsen anzupassen, z. B. X-Achse mit „Prägejahr“ und Y-Achse mit „Anzahl“, sowie (links) die Anzahl der Klassen zu steuern.

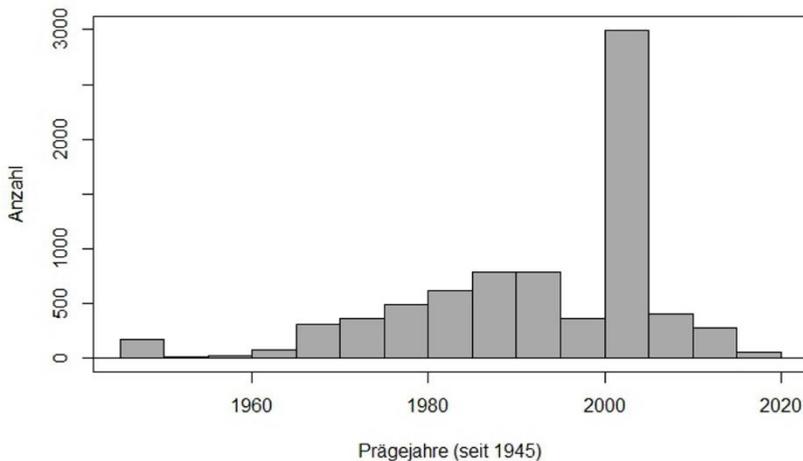
Nun, der Weg zum Ziel war das Setzen des Startwertes und das Setzen der Klassenweite. Wie geht das mit dem R-Commander? In zwei Schritten: (a) => Datenmanagement; => Aktive Datenmatrix => Teilmenge der aktiven Datenmatrix; in dem aufklappenden Menü setzen wir „Anweisungen für die Teilmenge“ auf „Jahr > 1944“ (d. h. alle Fälle ab 1945), geben dem neuen Teildatensatz einen Namen (**Abb. 39**) & OK. (b) Dann bestellen wir das Histogramm: => Grafiken; => Histogramm; wählen „Jahr“ aus und setzen unter „Optionen“ das Passende ein: bei Label der X-Achse z. B. „Münzen ab Prägejahr 1950“, bei Label der Y-Achse z. B. „Anzahl“ und bei Anzahl der Klassen z. B. 15 (**Abb. 40**), und erhalten das angestrebte Ergebnis (**Abb. 41**).



**Abb. 39** R-Commander: Schaltfläche zur Eingrenzung des Datensatzes auf Münzen ab Prägejahr 1945.



**Abb. 40** R-Commander: Schaltfläche zur Setzung der Klassenanzahl und der Achsenbeschriftung des Histogramms.



**Abb. 41** Das aus den Eingaben Abb. 39 und Abb. 40 via R-Commander resultierende Histogramm.

Ja, wegen der zwei Prozeduren ist der Weg im Vergleich zu PAST etwas holpriger, in der Praxis pröbelt man einige Zeit herum, bis man sein Ergebnis so eingestellt hat, wie man es sich vorstellt. Aber... es gibt ja noch eine andere Möglichkeit, nämlich die Kommandozeile zu nutzen.

Im oberen Protokollfenster der R-Commanders „R Script“ wurde protokolliert, was die Point-&-Click-Bedienung zuvor getan hat: ein paar Befehle in **R** aufgesetzt, die dann ausgeführt wurden:

```
ab1945 <- subset(Boersen, subset=Jahr > 1944)
with(ab1945, Hist(Jahr, scale="frequency", breaks=15, col="darkgray",
xlab="Präcejahre (seit 1945)", ylab="Anzahl"))
```

Man lese die erste Zeile: Aus dem Datensatz „Geldbörsen“ wurde ein *subset* gebildet mit allen Münzen „Jahr größer 1944“, das Ganze in einen Datensatz „ab1945“ gespeichert. Die zweite Zeile enthält die Angaben für das Histogramm mit all den Angaben, die zuvor gemacht wurden. Diese Zeile R-Code kann man verändern und dann ausführen lassen. Versuchsweise ersetze man die 15 Klassen (*breaks*) durch 30 Klassen, und den Eintrag „darkgray“ durch „red“. Wenn man jetzt die veränderte Zeile mit dem Cursor markiert und dann die Schaltfläche darunter „Befehl ausführen“ anklickt, wird das Histogramm neu gezeichnet. Wie zuvor bei den Kreisdiagrammen neige ich dazu, die Beschriftungen etwas zu vergrößern, d. h. füge noch *cex.axis* und *cex.lab* in diese Befehlszeile ein und erhalte dann ein besser für Publikationen und Präsentationen geeignetes Bild:

```
with(ab1945, Hist(Jahr, scale="frequency", breaks=30, col="red", xlab="Prä-
cejahre (seit 1945)", ylab="Anzahl Münzen", cex.axis=1.2, cex.lab=1.3))
```

Das Vorgehen soll hier nicht weiter vertieft werden, aber deutlich wird: Man kann den R-Commander nutzen, um einen grundsätzlich funktionierenden R-Code zu gewinnen, und diesen dann händisch gezielt anpassen. Diese Arbeitsweise beschleunigt das Arbeiten im Falle von wiederkehrenden Versuchen mit kleinen Änderungen oder beispielsweise beim Bereinigen von Tippfehlern ungemein. Zudem lassen sich erfolgreiche Kommandozeilen und -folgen als Datei abspeichern, so dass man sie für weitere Grafiken nach Anpassung wiederverwenden kann.

Kurz noch ein Blick in das rechte untere Fenster von RStudio, wo man den Reiter „Help“ aktiviert. In unserem Beispiel haben wir im Code-Text des R-Commanders gesehen, dass das Histogramm mit der Funktion „hist“ erzeugt wurde. Geben wir im besagten Fenster von RStudio im Suchfenster rechts oben „hist“ ein, wird der Hilfetext zu dieser Funktion eingespiegelt, zuoberst die Struktur, nachfolgend die Detailerläuterungen. Daraus lassen sich weitere Möglichkeiten gewinnen, durch Anpassen den Codes das Histogramm noch mehr so zu gestalten, wie man es publizieren möchte.

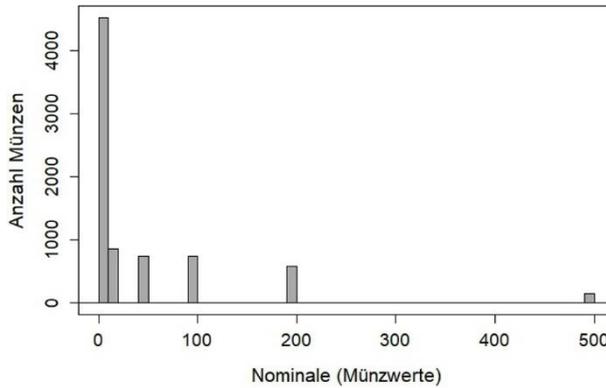
#### 17.4.2.1 *Hinweis: Grafiken nachbearbeiten mit Inkscape*

Man kann die Grafiken innerhalb von R (erheblich) weiter perfektionieren. Manchmal kommt man jedoch schneller zum Ziel, wenn man Hinzufügungen, Beschriftungen etc. extern vornimmt. Hierzu eignet sich das freie Programm Inkscape gut. Es bearbeitet Vektorgrafiken. Statt (wie hier) die Grafiken aus R als Bitmaps rauszuschreiben, kann man sie aus R auch als Vektordatei (\*.eps) herauschreiben, der man dann mit Inkscape den letzten Schliff gibt.

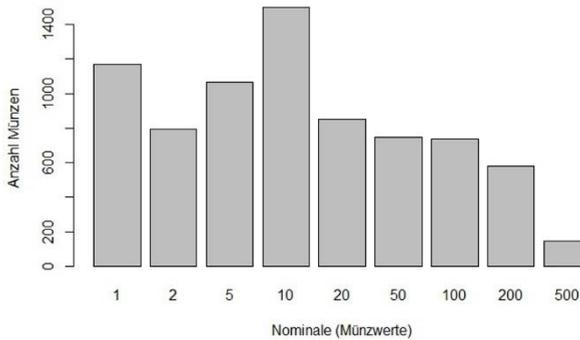
#### 17.4.3 Weitere Histogramm-Varianten

Zurück zum Thema „angemessene Grafiken“. Ein besonders reizvolles Problem für die Suche nach einer passenden Grafik sind die Nominalen der Münzen. Die Nominalen beruhen auf einer quantitativen stetigen Skala, weil mit einer Staffellung von 1 Cent alle Werte möglich sind. Daher ist ein Histogramm wie **Abb. 42** eine theoretisch berechnete Art der grafischen Darstellung. Als Münzen ausgeprägt werden hingegen nur die Werte 1, 2, 5, 10, 20, 50, 100, 200 und z. T. auch 500, sodass man die Prägewerte auch wie nominalskaliert behandeln könnte. Daher wäre hier auch ein Balkendiagramm angemessen. Wegen der Eigenart der tatsächlich geprägten Werte wird sich nämlich bei einem Histogramm immer der in **Abb. 42** deutlich werdende Effekt ergeben, dass bei den kleinen Nominalen viele Werte zu einer Klasse zusammengefasst werden müssen, während zwischen den höheren Werten weite Lücken klaffen. Dies wird in einem Balkendiagramm wie **Abb. 43** aufgefangen. Um ein solches Anliegen zu können, muss die (numerische) Variable „Nominal“ zuvor zum „Faktor“ umgewandelt werden.

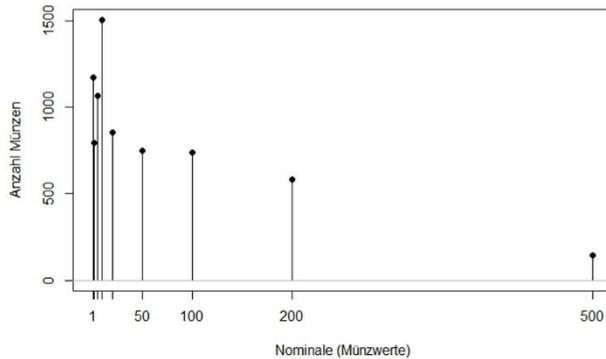
Als Alternative zu dem Balkendiagramm wäre auch ein Kreisdiagramm denkbar (bitte ausprobieren), m. E. ergibt jedoch das Balkendiagramm das angemessenere Bild. Der R-Commander bietet aber eine Alternative, die für unseren Datensatz angemessener ist (**Abb. 44**): statt der breiten Balken werden dünne Linien benutzt, so dass die Darstellung in diskreter Skala erhalten bleibt und eine Klassenbildung vermieden wird. Einerseits werden die Abstände verdeutlicht und zugleich bleiben auch die Häufigkeiten der kleinen Nominalen einzeln erkennbar. Der Weg dazu: => Grafiken; => Zeichne diskrete numerische Variable; Variable auswählen und per „Optionen“ ggf. die Achsentitel anpassen.



**Abb. 42** Histogramm der Nominalwerte (ausgedrückt in Cent, Rappen, Pfennigen) der Münzserie (n=7.760). Um (immerhin) diese Auflösung zu erreichen, wurde im Code die Zahl der „breaks“ auf 50 gesetzt.



**Abb. 43** Balkendiagramm der Nominalwerte der Münzserie (n=7.760); vgl. oben das Histogramm **Abb. 42** auf Basis der gleichen Daten.

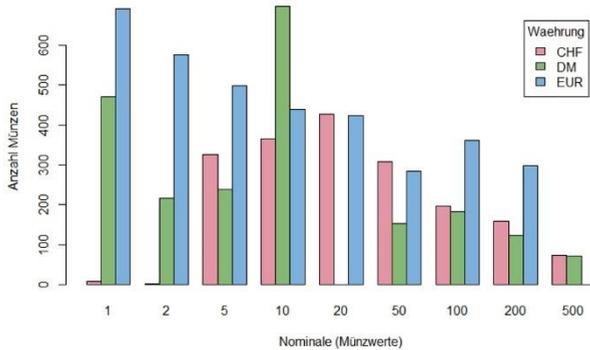


**Abb. 44** Häufigkeit der Nominale (in Cent) (n=7.760); gleiche Daten wie **Abb. 42** und **Abb. 43**.

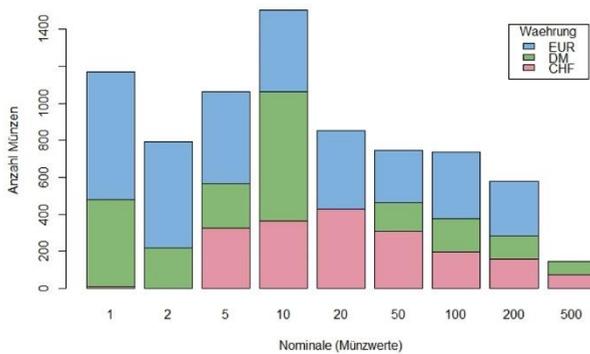
Bei den anderen Beispieldatensätzen Zwentendorf und Howells-World-Data würde man alle quantitativ skalierten Variablen in Histogramme übertragen, beispielsweise die Messstrecken an den Langknochen und die Maße der Schädel. Dies gelegentlich auch praktisch zu tun ist eine gute Übung, das bisher erlernte in der Software Ihrer Wahl umzusetzen und dabei auch die Datensätze kennen zu lernen.

#### 17.4.4 Datenreihen im Balkendiagramm vergleichen

Balkendiagramme werden, wie auch Kreisdiagramme, immer wieder benutzt, um verschiedene Datenreihen miteinander zu vergleichen. Bei unserem Beispiel der Nominale könnte man etwa die Häufigkeit in den Währungen Euro, Franken und DM miteinander vergleichen. Meines Erachtens ist dies mit einer Kreuztabelle – ein Thema das später vertieft werden wird – gut und erheblich klarer als mit Grafiken leistbar. Als Beispiel sollen dennoch zwei Optionen von Balkendiagrammen dargestellt und diskutiert werden. **Abb. 45** zeigt ein „gestaffeltes“ Balkendiagramm, in dem jede der drei Währungen mit einer Säule pro Nominal dargestellt ist. Meines Erachtens sieht man das Wesentliche erst „auf den zweiten Blick“ – weshalb sich durch die Grafik kein Vorteil gegenüber einer Tabelle ergibt, aber das kann für andere Wahrnehmungsgewohnheiten anders sein. Der Weg im R-Commander dorthin: = Grafiken; => Balkendiagramm ...; „Nominal“ als Variable auswählen, bei „Grafik für die Gruppen“ „Währung“ einsetzen, und beim Reiter Optionen „Gestapelt“ auswählen (sowie ggf. die Beschriftung der Achsen setzen und die Legende nach rechts bestellt).



**Abb. 45** Vergleich der Nominalen bei den Währungen Franken, DM und Euro (n=7.760) durch ein gestaffeltes Balkendiagramm.



**Abb. 46** Vergleich der Nominalen bei den Währungen Euro, Franken und DM (n=7.760) durch ein gestapeltes Balkendiagramm. Vgl. dazu **Abb. 43** mit Darstellung der gleichen Daten, bei ähnlicher Fragestellung.

**Abb. 46** bietet einen minim anderen Ansatz: Der Nominalanteil ist für jede Währung anteilig innerhalb der Säulen dargestellt. Diese Darstellung als „Stapeldiagramm“ ähnelt in ihrer Idee dem möglichen Nebeneinanderstellen von drei Kreisdiagrammen, je eins für jede Währung, wobei die Lösung Fehler! Verweisquelle konnte nicht gefunden

werden. **Abb. 46** die tatsächliche Datenlage wie auch die angestrebte Aussage besser wiedergibt. Der Weg dorthin mit dem R-Commander ist identisch wie zuvor für **Abb. 45**, nur dass bei den Optionen statt „gestapelt“ nun „parallel nebeneinander“ gewählt wurde.

Kleines PS: Solche Grafiken brauchen Farbe! Bei Präsentationen ist das in der Regel technisch kein Problem, im Gegenteil, Farbigkeit ist willkommen. Aber wer an einem künftigen Druckwerk arbeitet, das eventuell schwarz-weiß druckt, könnte die Farben durch Muster ersetzen oder um Muster ergänzen.

Inhaltlich erkennt man an diesen Grafiken, dass es beim Euro keine 500-Cent-Münze gibt, dass in der Schweiz die 1- und 2-Rappen-Münzen, wiewohl formal noch gültig, im Alltag praktisch keine Rolle mehr spielen, und dass die DM keine 20-Pfennig-Münze kannte, weshalb dort die 10-Pfennig-Münzen erheblich häufiger waren als sie es heute im Euro- und Frankenraum sind.

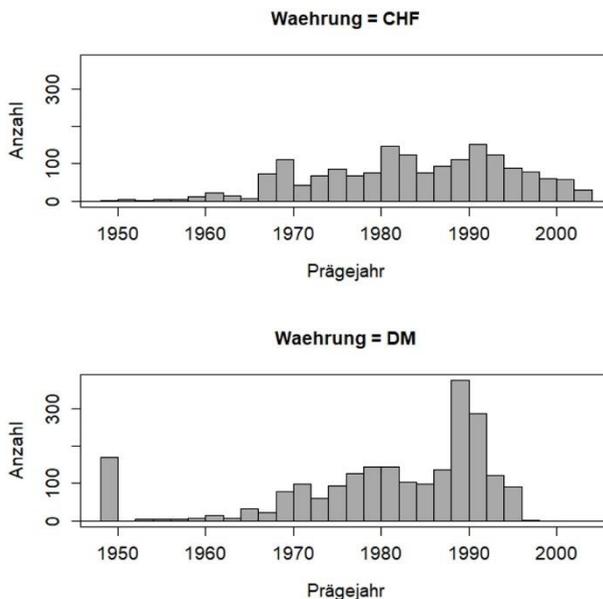
Arbeitet man mit dem R-Commander, ist es durchaus sinnvoll – z. B. via Hilfsfunktion von RStudio – in die Details des Codes hineinzuschauen. Zum Balkendiagramm z. B. findet man unter „barplot“ die Grammatik und Möglichkeiten der Fein-Steuerung derselben und kann diese im Try-and-Error Verfahren (oder in sorgfältiger Lektüre der Hilfe-Texte) durch Ergänzung des im R-Commander sichtbaren Codes durchprobieren. Hier ein Beispiel, was man aus den Standardeinstellungen (*defaults*) so machen kann:

```
with(Dataset, Barplot(Waehrung, # Basisangaben
  xlab="Währung", # setzen der Beschriftung der x-Achse
  ylab="Anzahl Münzen", # setzen der Beschriftung der y-Achse
  border=TRUE, # macht feinen Rahmen um die Säulen
  density=15, # Schraffuren statt Flächen, Zahl=Linien-dichte
  angle=0, # dreht Richtung d.Schraffuren, hier waagrecht
  cex.axis=1.3, # vergrößert die Beschriftung der y-Achse
  cex.names=1.4, # vergrößert die Beschriftung der x-Achse
  font=3, # setzt den Font auf kursiv
  col=palette()[2])) # wählt Farbpalette 2 = rot aus
```

#### 17.4.5 Datenreihen vergleichen durch verbundene Histogramme (*facets*)

Mehrere quantitativ skalierte Datenreihen wie z. B. die Prägejahre lassen sich mit übereinander gestapelten, gleich skalierten Histogrammen gut vergleichen; alternativ ist auch ein Nebeneinanderstellen von Histogrammen möglich, was aber für unser Beispiel nicht passt. Bei **R** wird dieses Über- oder Nebeneinanderstellen von Grafiken als

Facetten resp. Facettieren bezeichnet. Unser Münzbeispiel beruht auf echten Portemonnaies von Studierenden und Kollegen in den Jahren 1989-2019, spiegelt also den tatsächlichen Münzumschlag dieser Jahre wider.



**Abb. 47** Die Häufigkeit der Prägejahre 1945-2004 in den Geldbörsen der Jahre 1989-2018 für die Währungen CHF und DM (n=3.949).

**Abb. 47** zeigt in unserer Stichprobe für die Prägejahre 1945 bis 2004 einen Vergleich der im Umlauf befindlichen Mengen der Währungen DM und CHF. Der Vergleich zeigt, dass in den 1989 bis 2019 gezogenen Münzstichproben gerade im Vergleich zur Schweiz die Währungsgeschichte Deutschlands gut ablesbar ist: Mit der Währungsreform 1949 wurden in Deutschland alle älteren Münzen wertlos und verschwanden aus dem Umlauf. Münzen des Jahres 1949 sind später bei Sammlern attraktiv geworden und aus dem Münzumschlag verschwunden, aber von dem frühen Emissionspeak ist das Jahr 1950 noch nicht in Sammlungen weggehörtet, sondern auch 40 Jahre später noch nennenswert im Umlauf vertreten – anders als in der Schweiz, wo so alte Münzen relativ seltener sind. Die deutsche Wiedervereinigung am 3. Oktober 1990 war mit einer ungewöhnlich starken Münzemission verbunden, weil nun zusätzlich zu den Bürgern der alten Bundesrepublik (damals ca. 61 Mio. Einwohner) die Bürger der Ex-

DDR (ca. 17 Mio. Einwohner) neu mit DM-Hartgeld zu versorgen waren. Umgekehrt erkennt man, dass im Hinblick auf die Einführung des Euro im Januar 2002 wohl schon ab ca. 1997 keine frischen DM-Münzen mehr herausgegeben wurden.

Auch die Schweizer Datenreihe spiegelt bis heute einige schon lange zurückliegende Emissionsereignisse wider. Weil es in der Schweiz im 20. Jahrhundert keine Währungsreform gab, sind – anders als in Deutschland – vereinzelt noch Münzen aus der ersten Hälfte des 20. Jahrhunderts im alltäglichen Münzumsatz vorhanden, was wir aber für **Abb. 47** weggefiltert hatten. Münzen ab 50 Rappen waren m. W. ehemals aus einer stark silberhaltigen Legierung geprägt, was ab 1967 zugunsten einer preiswerteren Legierung geändert wurde, weil der Materialwert der bisherigen Münzen ihren Nominalwert überstieg. In Folge verschwanden Prägungen vor 1967 in Sammlungen und zugleich gab es 1967 einen Peak durch die neu herausgegebenen Münzen.

Technisch erfordert das gestapelte Histogramm **Abb. 47**, dass beide Bilder eine identische x-Skala erhalten. Ein zusätzliches Einspiegeln auch der Euro-Währung in dieses Bild würde erfordern, dort auf der y-Achse für die Häufigkeiten eine andere Skala zu verwenden, weil bei dieser massiven Emission für die ersten Jahre die y-Skala nicht, wie hier, bis ca. 250, sondern bis ca. 2.500 gehen müsste. Setzte man alle drei Histogramme auf eine gleich skalierte y-Achse, würde man umgekehrt die Unterschiede bei DM und CHF kaum mehr sehen können.

Wie **Abb. 47** entstand? In zwei Schritten: (1) Mit dem R-Commander erst einmal den Datensatz gezielt zuschneiden, d. h. die Währung Euro ausklammern und alle Münzen 1945-2004 herausuchen: => Datenmanagement; => aktive Datenmatrix; => Teilmenge der aktiven Datenmatrix. In dem aufpoppenden Fenster ganz unten einen neuen Namen für diesen Datensatz eingeben, z. B. „Auswahl“; in dem Fenster darüber den passenden Filter definieren. In diesem Fall:

```
Waehrung != "EUR" & Jahr >1944 & Jahr <2005
```

In Worten: Suche alle Münzen heraus, die nicht zur Währung Euro gehören, und deren Prägejahr größer (i. e. jünger) als 1944 ist (also ab 1945) und die kleiner (also älter) als 2005 sind (d. h. 2004). Keine Sorge, das Fenster in der betreffenden Schaltfläche des R-Commanders wirkt nur klein, kann aber lange Befehle aufnehmen. (2) Für diese „Auswahl“ wurde dann ein Histogramm bestellt: => Grafiken; => Histogramm; „Jahr“ ausgewählt, und die unter dem Variablenfenster liegende Taste „Grafik für die Gruppen“ bedient und dann „Waehrung“ ausgesucht; 2x mit OK bestätigen, usw. Der R-Code sieht dann z.B. wie folgt aus:

```
Auswahl <- subset(Dataset,
  subset=Waehrung != "EUR" & Jahr > 1944 & Jahr < 2005)
with(Auswahl, Hist(Jahr, groups=Waehrung, scale="frequency",
  breaks=30, col="darkgray", xlab="Prägejahr", ylab="Anzahl",
  cex.axis=1.2, cex.lab=1.2))
```

Wer den Stil von *ggplot2* bevorzugt, kann solche Histogramme im R-Commander über das Plugin *KMggplot2*<sup>40</sup> oder – m. E. etwas besser – via *BlueSky* recht leicht umsetzen: (1) über => Data; => Subset (usw.) wird die Auswahl getroffen, (2) über => Graphics; => Histogramm das Passende bestellt, indem man „Jahr“ als zu untersuchende Variable einsetzt („x-Achse“) und (ganz unten) „Waehrung“ bei „create facets“ verwendet, in unserem Fall für die y-Achse, sodann die „Bin width“ auf „1“, d. h. 1 Jahr setzt, und abschließend das Histogramm über die „Options“ noch etwas aufhübscht.

Wir wollen das Münzbeispiel nicht zu weit und detailliert auswalzen. Die Diskussion zeigt, dass es in der Wahl und Gestaltung der Grafiken jenseits einiger Standards, die man beherzigen sollte, einen erheblichen Gestaltungsraum gibt. Je nach Datenlage und Fragestellung sind flexibel angemessene Entscheidungen zu treffen.

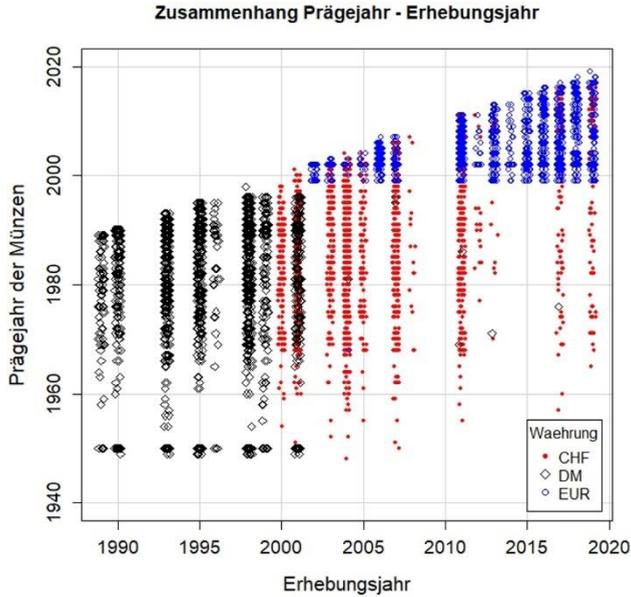
## 17.5 STREUUNGSDIAGRAMM: ZWEI VARIABLEN IM ZUSAMMENHANG

Den Zusammenhang von zwei (oder mehr) nominalskalierten Variablen wird man jenseits der in Kap. 17.4.4. aufgezeigten Möglichkeiten nur in Ausnahmefällen grafisch darstellen wollen, hierzu eignen sich Tabellen m. E. besser. Unter dem Suchwort „mosaic plot“ findet man eine neu aufkommende Option zur Darstellung des Zusammenhangs von zwei oder drei nominalskalierten Variablen, aber die praktische Umsetzung mit dem R-Commander ist nicht ganz trivial.

Bei quantitativen Skalen sind Streuungsdiagramme (auch: Streudiagramm; engl. *scatter plot*, *scatterplot*) die Grafik der Wahl. In dem Datensatz *Geldbörsen* bietet nur die Zeitachse eine quantitative Intervallskala. Also könnte man den Zusammenhang zwischen den Prägejahren der Münzen einerseits und dem Jahr die Ziehung der Stichprobe in einem Streuungsdiagramm darstellen (**Abb. 48**).

---

<sup>40</sup> In der im Januar 2020 aktuellen Fassung scheint das Plugin *KMggplot2* nur das Nebeneinanderstellen, nicht jedoch das Übereinanderstellen umsetzen zu können, obwohl das Steuerungsmenü beide Optionen anbietet.



**Abb. 48** Zusammenhang zwischen dem Prägejahr der Münzen (nur Prägungen ab 1945) und dem Jahr der Stichprobenziehung, d. h. dem Jahr, in dem die jeweiligen Geldbörsen erfasst wurden (n=7.743).

Das Streudiagramm **Abb. 48** veranschaulicht das erwartete Bild, dass die Prägejahre insgesamt einerseits mit jedem späteren Erhebungsjahr tatsächlich später enden, der Münzspiegel andererseits aber dennoch jeweils viele alte Münzen umfasst. Wir werden dieses Phänomen später anhand der Darstellung in statistischen Maßzahlen exakter beschreiben können. Auf dem zweiten Blick erkennt man, dass die Schweizer Portemonnaies immer wieder auch recht alte Münzen enthalten, während man bei der DM wie beim Euro klar das Jahr ihrer Einführung ausmachen kann. Der Euro wurde am 1. Jan. 2002 physisch eingeführt, in der Praxis der Banken jedoch bereits per 1. Jan. 1999. An der entsprechenden Stelle im Streudiagramm wird deutlich, dass in unserer Stichprobe die DM-Serien schon kurz vor dem Jahr 2002 nicht mehr (wie 1989 ff.) jünger werden, sondern die 2001er-Serien so alt bleiben wie die 1999er-Serie: Offenbar wurden nach dem politischen Beschluss des Euro und schon vor der Einführung des neuen Hartgeldes keine frischen DM-Münzen mehr geprägt.

Technisch zeigt das Streuungsdiagramm **Abb. 48** eine häufig gewählte Option: Zusätzlich zur Darstellung zweier quantitativer Variablen (Prägejahr, Erhebungsjahr) ist hier eine nominalskalierte Zusatzinformation (Währung) eingespielt, indem die einzelnen Punkte entsprechend dieser Eigenschaft markiert wurden. Durch die Wahl von Farbe und Symbolform für diese Markierung sollte die Grafik auch für Menschen mit Farbfehlsichtigkeit lesbar sein. Die Farben wiederum wurden so gewählt, dass man sie sich leicht merken kann, blau für Europa, rot für die Schweiz, schwarz für Deutschland.

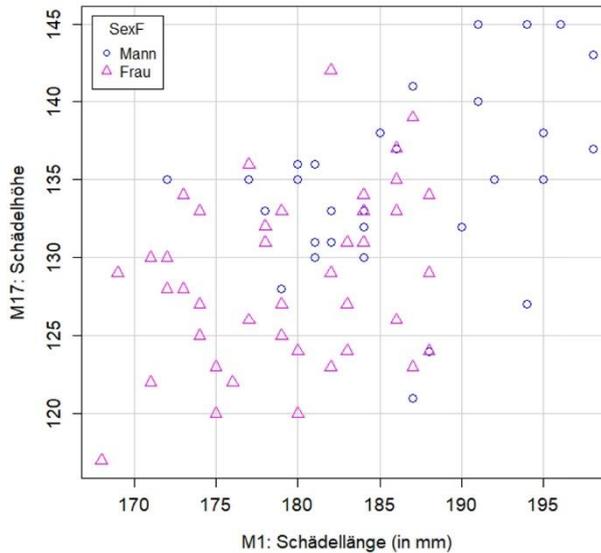
Wie das Bild erzeugt wurde? Zunächst mit den Möglichkeiten des R-Commanders: => Grafiken; => Streuungsdiagramm; „Stichjahr“ als x-Achse gewählt, „Jahr“ als y-Achse gewählt, „Währung“ als Gruppierungsvariable gesetzt; bei „Optionen“ die Achsenbeschriftungen eingesetzt, die Punkt- und Schriftgrößen angepasst, mit „x-Variable verrauschen“ den sog. Jitter-Effekt gewählt, die Position der Legende nach unten rechts gewählt. Das Ergebnis ist nahe dem, was **Abb. 48** zeigt. Zusätzlich wurden Optionen über die Help-Funktion gesucht und demgemäß händisch in den Code eingesetzt die Zeilen mit „col“, welche die Farbe der drei Symbole setzt, sowie „pch“, womit die Form der drei Symbole gewählt wird. Der resultierende Code:

```
scatterplot(Jahr~Stichjahr | Waerung,
  regLine=FALSE,          #schaltet Einspiegelung d.Regressionlinie aus
  smooth=FALSE,          #schaltet Einspiegelung von Glättungskurve aus
  boxplots=FALSE,        #schaltet Boxplots an den Achsen aus
  jitter=list(x=1),       #"jitter" verrauscht Punkte entlang x-Achse
  xlab="Erhebungsjahr",
  ylab="Prägejahr der Münzen",
  main="Zusammenhang Prägejahr - Erhebungsjahr",
  cex=0.8,                # Größenanpassung Punkte
  cex.axis=1.1, cex.lab=1.1, # Größenanpassung Achsen
  ylim=c(1940, 2020),     # begrenzt auf Münzen 1940ff.
  by.groups=TRUE,
  legend=list(coords="bottomright"), # setzt Legende rechts unten
  pch=c(20,5,1),         # setzt Form der Symbole
  col=c("red", "black", "blue"),   # setzt Farbe der Symbole
  grid=TRUE, data=Dataset) # ev. "Dataset" anpassen!
```

Wenn man entlang dieser Anleitung oder dieses Codebeispiels rumprobiert, z. B. einzelne Elemente weglässt, lernt man schnell, wo und wie diese Eingriffe wirken.

Eine üblichere Art von Daten als unser Münzbeispiel und ein üblicheres Bild eines Streuungsdiagramms bietet **Abb. 49**, wo für das Gräberfeld von Zwentendorf die beiden Schädelmaße Länge („Größte Hirnschädellänge“, M1) und Höhe („Basion-Bregma-Höhe“, „Basalhöhe“, M17) dargestellt sind. Das Diagramm zeigt eine Streu-

ung der Punkte eher als (leicht gestreckte) rundliche Wolke denn als Anordnung entlang einer Linie, d. h. es scheint nur ein loser Zusammenhang zwischen beiden Strecken zu bestehen. Zusätzlich sind in die Legende die zugehörigen statistischen Maßzahlen (Korrelationskoeffizient und Signifikanz) eingebracht, wonach es einen schwachen signifikanten Zusammenhang gibt (dazu später gründlich im Kapitel „Korrelation“).



**Abb. 49** Gräberfeld Zwentendorf: Zusammenhang zwischen Schädelhöhe (M1, in mm) und Schädelhöhe (M17, in mm) bei den Erwachsenen (n=94), differenziert nach Geschlecht. Es besteht nur ein schwacher, aber hochsignifikanter Zusammenhang zwischen beiden Maßen (Korr.-koeff. Pearson 0.490, Signifikanz <.0001, d. h. hochsignifikant).

Maßzahlen und Abbildung zusammen ergeben die Aussage, dass es einen kleinen, jedoch signifikanten Größeneffekt gibt (je größer M1, desto größer M17), daneben jedoch ein erhebliches Maß an individueller Streuung jenseits des reinen Größenfaktors. Denn bei einem 1 : 1-Zusammenhang nähme die Kennzahl – der Korrelationskoeffizient – den Wert +1.00 an, und davon ist der beobachtete Wert 0.361 sehr weit entfernt. Da an der Grafik auch deutlich wird, dass bei Männern M1 und M17 gemeinsam tendenziell größer sind als bei Frauen, würde man die genannten Maßzahlen zusätzlich

auch nach Geschlechtern getrennt berechnen. Sie betragen für die erwachsenen Männer 0.254, sign. 0.191 (n=28),<sup>41</sup> für die erwachsenen Frauen 0.314, sign. 0.058 (n=37), d. h. innerhalb der Geschlechter gibt es keinen signifikanten Zusammenhang. Sprich: Es dominiert die individuelle Variabilität. Wir wollen auch dieses Beispiel inhaltlich nicht weiter ausführen, sondern hier nur exemplarisch verdeutlichen, dass Grafiken und Maßzahlen eng zusammengehören. Die Maßzahlen dienen dazu, visuelle Eindrücke, die sich aus den Grafiken ergeben, statistisch klar und eindeutig zu quantifizieren, umgekehrt führen die an den Grafiken gewonnenen Eindrücke oft dazu, dass man Phänomene klarer sieht und Fragen exakter stellen kann, für die man dann wiederum präzise Antworten anhand der Maßzahlen finden wird.

### 17.5.1 Praxis PAST: Anlegen eines Streuungsdiagramms

Die Option, ein Streuungsdiagramm (*scatterplot*) anzulegen, findet sich bei PAST im Plotmenü: Daten laden; im Datensatz die beiden Zahlenreihen, die in das Streuungsdiagramm aufgenommen werden sollen, markieren (1. Reihe anklicken, STRG-Taste drücken & gedrückt halten, 2. Zahlenreihe anklicken); => Plot; => XY Graph; dort die Schaltfläche rechts oben ggf. auf „Points“ setzen. Weitere Steuerungsmöglichkeiten finden sich wie gewohnt hinter den „Graph settings“ (unten).

Nun der Weg, ähnlich wie in **Abb. 48** für die Währungen oder in **Abb. 49** für das biologische Geschlecht die Symbole und deren Farbe zu beeinflussen – hier am Beispiel der Währungen: Datensatz in PAST laden. Spalte „Währung“ markieren und oben die Option „Column attributes“ anklicken. Dort die Variable „Währung“ auf die Option „Group“ setzen, d. h. Währung zur Gruppierungsvariable erklären. „Column attributes“ ausschalten. Spalte „Währung“ markieren, oben => Edit; => Row colours/Symbols wählen; in dem neu aufpoppenden Fenster kann man Farbe und Symbol für jede Gruppe setzen. Danach schließen, und ein Streuungsdiagramm anlegen.

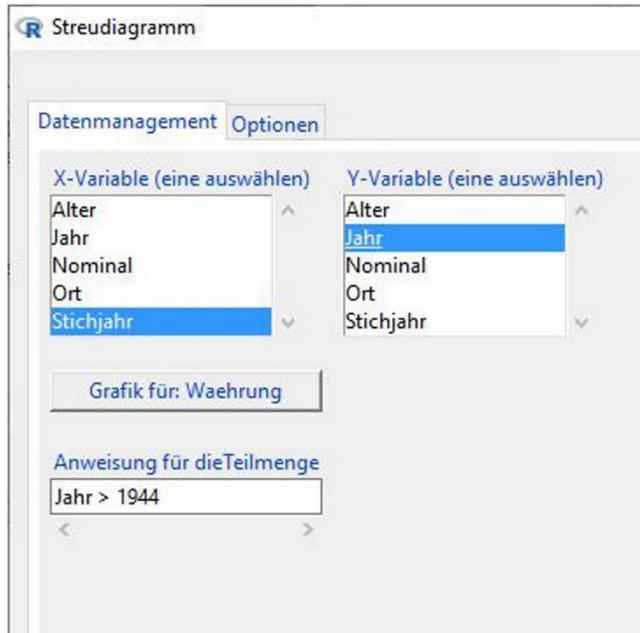
### 17.5.2 Praxis R-Commander: Anlegen eines Streuungsdiagramms

Das Aufsetzen eines Streuungsdiagramms (*scatterplot*) mit dem R-Commander ist recht einfach: Nach dem Laden des Datensatzes => Grafiken; => Streudiagramm; es öffnet sich ein Menü für die notwendigen Angaben. Unter dem Reiter „Datenmanagement“

---

<sup>41</sup> Bitte Geduld: näheres zum Thema Korrelationskoeffizient und zu „sign.“ – der Angabe seiner Signifikanz – später im Kapitel „Korrelation“. Doch weil wir von Anfang an ordentlich sein wollen, erfolgen diese (wichtigen) Angaben bereits hier, auch um zu zeigen, wie solche Informationen später in den Text Ihrer Publikationen eingebunden werden könnten.

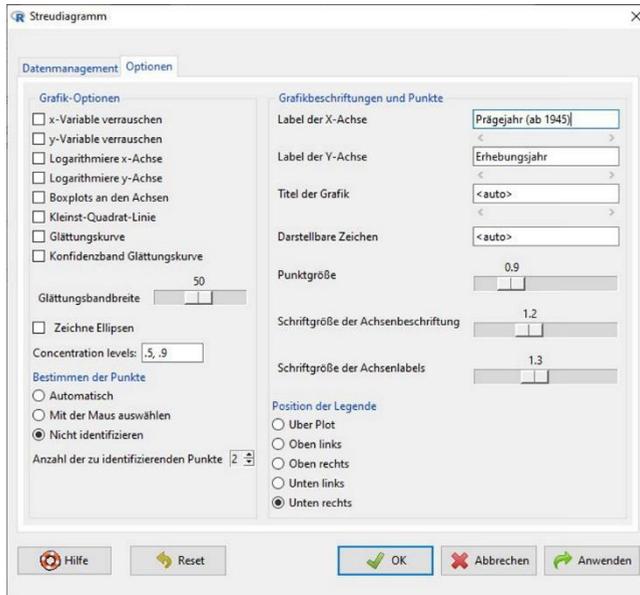
bestellt man als X-Variable „Jahr“, als Y-Variable „Stichjahr“, drückt darunter die Schaltfläche „Grafik für Gruppen...“, um „Währung“ auszuwählen; zudem können wir die Prägejahre wieder um die wenigen frühen Münzen beschneiden und tragen bei „Anweisung für die Teilmenge“ ein: „Jahr > 1944“, d. h. stellen nur die Prägejahre ab 1945 dar (Abb. 50). Hinter dem Reiter „Optionen“ lässt sich das Bild weiter spezifizieren (Abb. 51).



**Abb. 50** Menü „Datenmanagement“ (Ausschnitt) bei Aufsetzen des Streudiagramms mit dem R-Commander.

Nach dem OK erscheint die bestellte Grafik und das obere Fenster des R-Commanders hat zugleich die Befehlsfolge protokolliert, die wir mit dem Point-and-Click-Verfahren zusammengestellt haben.

```
scatterplot(Stichjahr~Jahr | Waehrung, reg.line=FALSE, smooth=FALSE,
spread=FALSE, boxplots=FALSE, span=0.5, ellipse=FALSE, levels=c(.5, .9),
xlab="Prägejahr (ab 1945)", ylab="Erhebungsjahr", by.groups=FALSE, le-
gend.coords="bottomright", data=Geldbörsen, subset=Jahr > 1944)
```



**Abb. 51** Menü „Optionen“ beim Aufsetzen des Streudiagramms mit dem R-Commander.

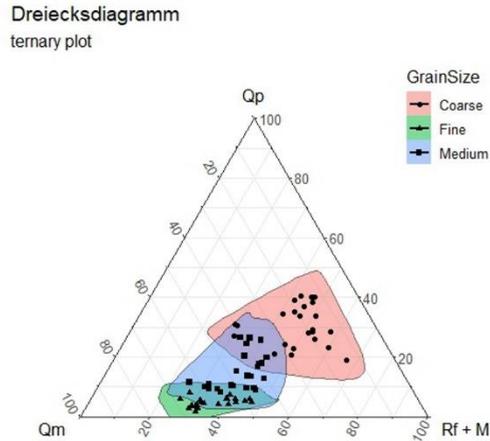
Damit ist das Prinzip erfasst; zum weiteren Kennenlernen ist es lohnend, auf Basis eines konkreten Datensatzes die verschiedenen Optionen einfach einmal durchzuprobieren.

## 17.6 GRAFIKEN ZUM ZUSAMMENHANG VON MEHR ALS ZWEI VARIABLEN: DREIECKSDIAGRAMM, STREUUNGSDIAGRAMM-MATRIX

Dreiecksdiagramme (*ternary plots*) dienen dazu, drei Variablen zugleich in einer zweidimensionalen Grafik darzustellen. So veranschaulicht man z. B. Prozentanteile mit ihnen wie etwa Grundformenanteile bei Silexinventaren, Tierknochenanteile etc., oder beispielsweise Elementanteile bei Metallanalysen. Dreiecksdiagramme gehören leider nicht zum Optionsumfang von **R base**, dem R-Commander oder BlueSky. Das R-Paket „Ternary“ erzeugt Dreiecksdiagramme im Grafikstil *base*, das Paket „ggtern“<sup>42</sup> im Stil von *ggplot2* (Abb. 52) – in beiden Fällen erfordert dies das Arbeiten mit dem rohen **R**

<sup>42</sup> Hamilton, N. E. & Ferry, M. (2018). ggtern: Ternary Diagrams Using ggplot2. *Journal of Statistical Software* 87, 3. doi: 10.18637/jss.v087.c03

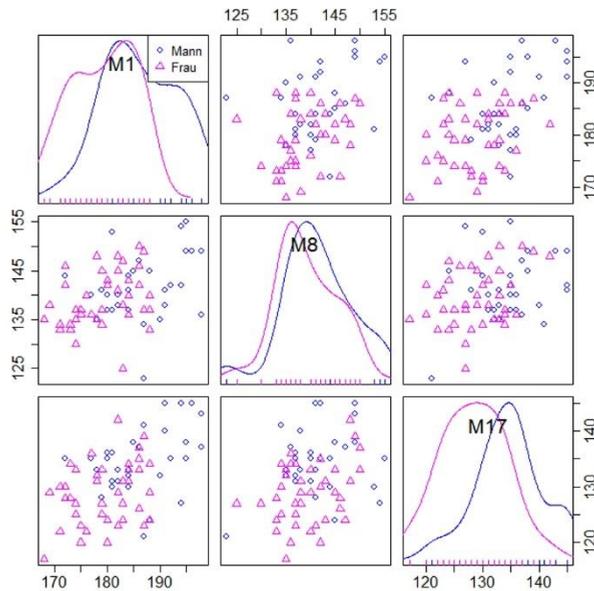
resp. RStudio. Nicht zuletzt sind Dreiecksdiagramme Teil des Grafikangebotes von PAST.



**Abb. 52** Beispiel eines Dreiecksdiagramms, hier nach Hamilton & Ferry (2018).

Der Grund für ihr Fehlen in **R** *base* und den üblichen GUIs liegt wohl darin, dass Dreiecksdiagramme bei Profi-Statistikern relativ unbeliebt sind und nur selten verwendet werden. Doch was ist die Alternative? Im Falle von drei Variablen wären auch sog. 3D-Streuungsdiagramme denkbar, die z. B. mit dem R-Commander oder BlueSky ohne besondere Mühen angelegt werden können. Beim R-Commander z. B. via => Grafiken; => 3D-Grafik; => 3D-Streudiagramm, usw. Doch so ansehnlich deren interaktives Bewegen am Bildschirm auch sein mag, so schwierig ist der Umgang mit ihnen, wenn die Bilder weitergenutzt werden sollen, also im Druck oder für eine Präsentation, da dann deren Interaktivität entfällt. Daher ist der übliche Weg im Grunde ähnlich wie beim Verbinden von Histogrammen (Kap. 17.4.5): man verbindet mehrere Streuungsdiagramme zu einer Matrix, was auch das Arbeiten mit mehr als drei Variablen ermöglicht. Bildlich gesprochen wird hier der sich z. B. aus drei Variablen aufspannende 3D-Würfel jeweils von exakt einer Seite des Würfels aus betrachtet: einmal exakt vor vorne, einmal exakt von der Seite und einmal exakt von oben. Mit etwas Eindenken sind dies verständliche Darstellungen, die auch auf Papier gedruckt oder in Präsentationen eingebunden werden können. Das Beispiel **Abb. 53** wurde mit dem R-Commander angelegt: => Grafiken; => Streudiagramm Matrix; dort drei Variablen

ausgewählt; bei „Grafik für die Gruppen“ Sex eingesetzt, und später händisch die Größen der Labels und Achsenbeschriftungen etwas angepasst.

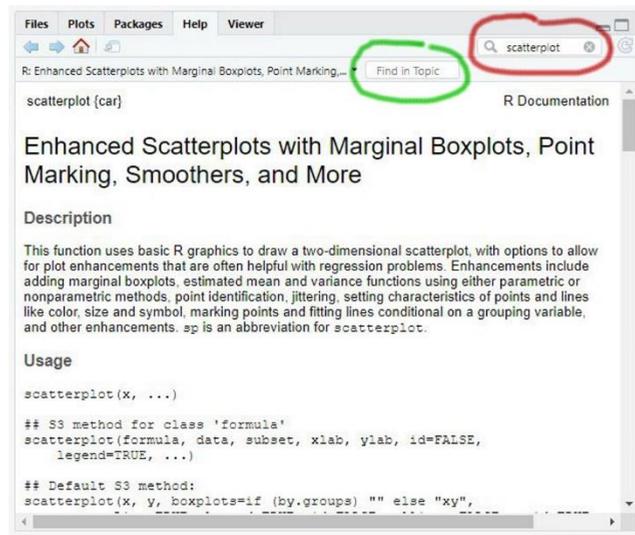


**Abb. 53** Beispiel einer Darstellung von drei Variablen als Matrix von Streuungsdiagrammen. Hier: Schädelmaße (in mm) der Serie Zwentendorf mit M1 (Größte Hirnschädellänge), M8 (Größte Hirnschädelbreite) und M17 (Basion-Bregma-Höhe) und Vergleich Frauen – Männer (n=149; nur Schädel mit M1 >150 mm).

Das Resultat zeigt den Zusammenhang der Länge (M1), Breite (M8) und Höhe (M17) der Zwentendorfer Schädel. In der Diagonalen wird jeweils die Variable selbst eingespiegelt, dargestellt als Dichtefunktion für die Gruppen Männer und Frauen. Die jeweiligen Streuungsdiagramme zeigen dann die Kombination der Variablen, in der obersten Reihe also M1 solo, dann M1 mit M8, dann M1 mit M17; in der zweiten Reihe M1 mit M8, M8 solo, und M8 mit M17 usw. Solche Streuungsdiagramm-Matrizen werden gerne auch für mehr als drei Variablen angelegt – eine Option, die Dreiecksdiagramme oder 3D-Streuungsdiagramme nicht bieten.

## 17.7 ALLGEMEINES ZU DEN GRAFIK-MÖGLICHKEITEN IN R

**R** beinhaltet drei verschiedene Grafikpakete: das Basispaket (*base*), das Paket *lattice* und das Paket *ggplot2*. Der R-Commander bedient sich des Basispakets von **R** mit seinem für **R** typischen, sehr auf das Wesentliche reduzierten Stil. Das Paket *ggplot2* bietet mehr Steuerungsmöglichkeiten und im Ergebnis bisweilen auch etwas ansehnlichere, moderner wirkende Grafiken; es kann aber nicht – zur Ausnahme der nächste Absatz – über den R-Commander gesteuert werden, sondern bleibt denjenigen vorbehalten, die statt des R-Commanders direkt mit **R** und RStudio arbeiten oder z. B. die Oberfläche BlueSky einsetzen. Weil wir hier in dieser Einführung soweit als möglich auf das selbstständige Arbeiten mit **R** im Kommandozeilenmodus verzichten wollen, sondern vor allem mit Hilfe des R-Commanders arbeiten, stehen uns vor allem die grafischen Möglichkeiten des Basispakets von **R** zur Verfügung.



**Abb. 54** Das Hilfe-Fenster von RStudio (rechts unten), über das sich die diversen Optionen der Befehle leicht finden lassen. Rot umkringelt: ein Thema finden, hier zur Funktion „scatterplot“; grün umkringelt: ein Stichwort innerhalb eines Themas finden.

Da der R-Commander oder auch RStudio die Point-and-Click-Eingaben protokollieren und in Kommandozeilen übersetzen, die man ggf. verändern und weiterverwenden kann, lassen sich auf dieser Basis sukzessive weitere Möglichkeiten erproben. Der

bequemste Weg führt über das Hilfefenster von RStudio (**Abb. 54**): Im Fenster rechts unten den Reiter „Help“ aktivieren. Aus dem Protokoll haben wir ersehen, dass die Streudiagramme mit der Funktion „scatterplot“ erstellt wurden. Gibt man „scatterplot“ in das obere Suchfenster ein, erscheint die Beschreibung dieser Funktion mit all ihren Optionen im Hilfefenster. Macht man eine Eingabe in dem Suchfenster darunter („Find in topic“), lassen sich einzelne Stichwörter innerhalb dieser Hilfe zu „scatterplot“ finden. Das dort gewonnene Wissen kann man nutzen, die protokollierten Befehle zu verändern. Auf diese Weise kann man seine grafischen Möglichkeiten schrittweise erweitern und sich zugleich allmählich in das Arbeiten mit dem Kommandozeilen-Modus hineinfinden.

### 17.7.1 ggplot2-Graphiken mit dem R-Commander

Mit dem Plugin „RcmdrPlugin.KMggplot2“ lassen sich innerhalb des R-Commanders auch *ggplot2*-Graphiken anfertigen. Zur erstmaligen Verwendung muss das Plugin, wie ein gewöhnliches Paket, erst installiert werden, später jeweils nur noch hinzugeladen werden. Das Laden übernimmt beim Arbeiten mit RStudio ein Klick auf dieses Plugin. Arbeitet man im rohen **R** mit dem R-Commander, ist die Abfolge => Extras; => Lade Rcmdr Plugin(s) ...; „RcmdrPlugin.KMggplot2“ auswählen & OK; den erforderlichen Neustart annehmen. Danach erscheint das Plugin als „KMggplot2“ oben im R-Commander in der obersten Menüzeile.

Hinter der Schaltfläche „KMggplot2“ findet man diverse Graphikmöglichkeiten, die nunmehr bereits grundsätzlich bekannt sind – oder die man nach Laden eines Datensatzes im Try-and-Error-Verfahren erst einmal spielerisch erkundet. Wer die beschriebenen Möglichkeiten des R-Commanders nutzt, kann sich auf diesem Wege über schnell verständliche Menüs *ggplot2*-Grafiken im Point-and-Click-Verfahren erzeugen. Ist man mit dem Resultat wie gegeben zufrieden, bietet das Plugin also wertvolle Hilfe. Möchte man die Resultate weiter an die eigenen Bedürfnisse und Vorstellungen anpassen, kann man zusätzlich den vom R-Commander erzeugten Code studieren, ggf. anpassen und erneut ausführen. Hilfreich hierbei sind m. E. die sog. „Spickzettel“ – auch: „Schummelzettel“, engl. *cheat sheets* genannt (Suchwörter) –, die zu diversen Elementen von RStudio bzw. des sog. „tidyverse“ verfügbar sind, u.a. eben auch zu *ggplot2*. Diese Spickzettel fassen kompakt die Grammatik, das Vokabular und die Möglichkeiten der jeweiligen R-Komponenten zusammen. Sie sind auf der Seite „RStudio Cheat Sheets“ zusammengestellt: <https://www.rstudio.com/resources/cheatsheets/> [24.1.2020].

So schön dies klingt: m. E. ist der vom Plugin KMggplot2 erzeugte Code, an dem man sich die Syntax von *ggplot2* erarbeiten würde, nicht sonderlich transparent,

d. h. er erfüllt seinen Zweck, erscheint mir aber schwer verständlich. Wer ähnliches via BlueSky versucht, erhält einen m. E. einfacheren, erheblich besser lesbaren Code für *ggplot2*-Graphiken. Wer also über das hinauskommen möchte, was die Point-and-Click-Menüs des Plugins *KMggplot2* bieten, wird vermutlich andere, effektivere Wege suchen.

## 18 HÄUFIGKEITEN, PROZENTZAHLEN, HÄUFIGKEITEN VERGLEICHEN ...

---

Die in der Archäologie wie Anthropologie übliche Art von Beobachtungen ergeben am Ende oft Häufigkeiten: Die Anzahl von Gräbern samt Zahlen zu deren Untermen- gen wie Kindergräber, Frauengräber, Männergräber, oder Beobachtungen zur Häufig- keit von Beigabengattungen und -arten wie z. B. Waffen und Waffenarten, oder die Auszählungen von zu Typen zusammengefassten Keramikgefäßen resp. deren Frag- menten aus Siedlungskontexten. Hinsichtlich des Skalenniveaus handelt es sich bei diesen Beispielen um nominalskalierte Variablen. Das Analysieren von solchen Häu- figkeiten wird hier ob seiner Bedeutung ausführlich behandelt und dient zugleich dazu – Stichwort Chi-Quadrat-Test ( $\chi^2$ -Test) –, nicht nur ein wichtiges statistisches Test- verfahren eingehend kennenzulernen, sondern daran exemplarisch statistische Tests (samt deren Fallen) überhaupt einzuführen.

Oft werden auch ordinalskalierte Variablen oder gar solche einer quantitativen Skala als Häufigkeiten ausgewertet, wie beispielsweise Altersbestimmungen an Skelet- ten. Bei den Altersbestimmungen folgen die Beobachtungen zwar theoretisch einer stetigen Intervallskala, also einem erheblich höherwertigen Skalenniveau. Die den anthropologischen Bestimmungen innewohnenden Unsicherheiten werden von den Spezialisten jedoch traditionell dadurch signalisiert und aufgefangen, dass man die Er- gebnisse als Klassen (Altersspannen) ausgibt, die dann oft als Häufigkeiten ausgewer- tet werden.

Nominalskalierte Informationen werden üblicherweise pro Merkmalsausprä- gung ausgezählt, sie bilden Häufigkeiten. Wir sind es gewohnt, solche Zahlen zusätz- lich in Prozentwerte umzusetzen. Denn die einzelnen Häufigkeiten sind erst verständ- lich vor dem Hintergrund der Gesamtzahl der Fälle. Diese „Grundgesamtheit“ (auch: Bezugssumme, Hundertsumme) ist stets zur Einschätzung der Einzelinformation im

Hinterkopf zu behalten. Die Tabelle (auch: Kontingenztafel; Häufigkeitstabelle) **Abb. 55** zeigt ein typisches Beispiel, hier anhand unseres Datensatzes Geldbörsen.

In diese Tabelle sind einige Entscheidungen und Momente statistischer Etikette eingeflossen:

- Die Zahlen innerhalb der Zellen sind rechtsbündig gesetzt, Text wie üblich linksbündig.
- Die Prozentzahlen sind durch Kursivsetzung in der Tabelle optisch unterschieden von den beobachteten Häufigkeiten. Dieses Gestaltungsmittel wird unterschiedlich gehandhabt, es ist nicht verpflichtend. Ich empfehle es so, da es die schnelle Lesbarkeit der Tabellen erhöht.
- Prozentzahlen bei Grundgesamtheiten unter 100 werden bis auf zu begründende Ausnahmen ohne Zahlen hinter dem Komma ausgewiesen, d. h. auf ganze Prozente gerundet (s. unten). Bei Grundgesamtheiten über 100 können die Prozente mit einer Ziffer hinter dem Komma, also Zehntelprozenten ausgewiesen werden. Dem entsprechend sollte man Prozentzahlen mit zwei Ziffern hinter dem Komma allenfalls dann ausgeben, wenn tatsächlich mehr als 1.000 Beobachtungen vorliegen, wobei sich auch dann die Frage stellt, inwieweit diese Genauigkeit der Übersichtlichkeit zuträglich ist.
- Alle Prozentzahlen sind gleich formatiert, d. h. der errechnete Wert 19,0 ist auch als „19,0“ ausgewiesen, nicht als „19“ (wie es leider allzu oft zu sehen ist).

Prägeort	Anzahl	%
Berlin (A)	123	7,5
München (D)	312	19,0
Stuttgart (F)	432	26,3
Karlsruhe (G)	277	16,9
Hamburg (J)	496	30,2
Summe	1640	100,0

**Abb. 55** Häufigkeit der Prägeorte der DM-Münzen in unseren Geldbörsen. Aufgenommen nur Münzen, bei denen der Prägeort bestimmt wurde ( $n=1.640$ ; Anzahl der unbestimmten, hier nicht aufgenommenen Münzen = 578).

Wie Tabellen gesetzt werden, bleibt Geschmackssache oder ist Teil von Redaktionsrichtlinien. Das ehemals übliche Anzeigen von feinen Linien rundum, d. h. fixen Kästchen, wird heute zunehmend von Tabellen mit möglichst wenigen Linien abgelöst. Ich persönlich bevorzuge Tabellen mit weniger Linien, möchte aber zumindest die Kopfzeile abgesetzt sehen. Hat eine Tabelle unten und rechts eine Summenzeile, sollte auch diese von den angrenzenden Zeilen bzw. Spalten mit einer Linie angesetzt sein. Viel

befolgt wird heute der APA-Standard,<sup>43</sup> für dessen praktische Umsetzung unter **R** es sogar ein eigenes Paket gibt, „apaTables“ (Suchwort).

### 18.1.1 Rundungsregeln

Wie rundet man? Alles unter  $xx,50$  auf die ganze Zahl abrunden ( $xx$ ), alles ab  $xx,50$  auf die nächste ganze Zahl aufrunden ( $xx+1$ ). Dabei ergeben sich beim Addieren der Einzelwerte in den Tabellen gelegentlich Gesamtsummen von etwas unter oder etwas über 100 %. Die Antwort auf die Frage „was tun?“ ist Geschmackssache: (a) man kann mit Prozentsummen „um 100 %“ leben, (b) man kann bei den einzelnen Werten „zufällig“ auch einmal passend ab- bzw. aufrunden, damit sich insgesamt glatt 100 Prozent ergeben. Ich präferiere Option (a).

## 18.2 PROZENTZAHLEN UND IHRE ZUVERLÄSSIGKEIT

Ein wichtiger Grund für die Verwendung von Prozentzahlen besteht in ihrer Eigenschaft, uns Vergleiche mit anderen Datensätzen zu erleichtern. Nur selten haben wir – wenn wir vergleichen wollen – zwei Erhebungen vor uns mit der identischen Anzahl von Beobachtungen. Vielmehr arbeiten wir in der Regel mit unterschiedlich großen Grundgesamtheiten. In solchen Fällen ist es schwierig, die Rohdaten direkt zu vergleichen; bilden wir für beide Gruppen Prozentwerte, fällt uns der Vergleich erheblich leichter. Bei unserem Geldbörsen-Beispiel wurden die Stichproben zu Zeiten der DM vor allem in Göttingen und in Köln erhoben. Wir wollen vergleichen, ob die Häufigkeit der Prägeorte an beiden Standorten gleich ist. Schließlich wäre es ja alternativ denkbar, dass die geografische Lage der Erhebungsorte und die Lage der Prägeorte miteinander in Beziehung stehen, d. h. Münzen von räumlich näher gelegenen Prägeorten stärker vertreten sind. In **Abb. 56** sind die entsprechenden Informationen zusammengestellt. Die Zahlen lassen uns etwas ratlos zurück, denn man kann eine Tendenz erkennen: In Göttingen sind die Prägestätten Berlin und Hamburg etwas stärker vertreten, in Köln sind es die Prägestätten Stuttgart und Karlsruhe, während München in beiden Stichproben in etwa gleich häufig vorkommt. In Summe Beobachtungen, die einen (schwachen) Bezug zwischen Erhebungsort und Prägestätte andeuten. Doch kann man bei so geringen prozentualen Unterschieden und angesichts einer Erhebung von 585 und 945 Münzen die geringen Unterschiede wirklich ernst nehmen?

---

<sup>43</sup> APA = American Psychological Association, deren Style auch hinsichtlich des Zitierens etc. viel verwendet wird. Siehe: <https://apastyle.apa.org/> [24.1.2020].

Prägeort	Köln	%	Göttingen	%	Summe	%
Berlin (A)	31	5,3	87	9,2	118	7,7
München (D)	108	18,5	182	19,3	290	19,0
Stuttgart (F)	165	28,2	229	24,2	394	25,8
Karlsruhe (G)	116	19,8	138	14,6	254	16,6
Hamburg (J)	165	28,2	309	32,7	474	31,0
Summe	585	100	945	100	1530	100

**Abb. 56** Häufigkeit der Prägeorte der DM-Münzen in den Geldbörsen aus Köln und Göttingen. Aufgenommen nur Münzen, bei denen der Prägeort bestimmt wurde (n=1.530).

Aus der Sicht eines Archäologen dürften Stichproben von insgesamt 1.530 Fällen schon als relativ umfangreiche Stichproben gelten, meist kämpft man mit einer erheblich geringeren Anzahl an Beobachtungen. Daher ist die Frage, ob und wie man Unterschiede in den Prozentzahlen zwischen zwei Datenreihen interpretieren kann, ein geläufiges Problem. Leider wird die Antwort oft „nach Gefühl und Wellenschlag“ gewählt, meist entscheidet das Naturell des Bearbeiters: Skeptiker formulieren „mit so wenigen Beobachtungen kann man statistisch nichts machen“ und resümieren ihre Ergebnisse im Sinne eines „nichts Genaues weiß man nicht“, während Optimisten keine Scheu zeigen, die geringen Unterschiede inhaltlich/historisch zu interpretieren, denn die Unterschiede sind in den Zahlen ja tatsächlich belegt. Richtig ist, dass die Statistik genau für dieses Problem Hilfe anbietet und den Entscheidungen, die hier anstehen, eine zuverlässige Grundlage bietet. Die erste Lösung lautet „Konfidenzintervalle“, die zweite – weiterreichende und erheblich bessere – Lösung lautet Chi-Quadrat-Test. Wir gehen im Folgenden beide Optionen gründlich durch, weil das hier behandelte Problem in der Archäologie ein häufiges und wichtiges ist.

### 18.3 KONFIDENZINTERVALLE

Zunächst sei das Problem genauer formuliert: Wollen wir die beobachtete Gruppe lediglich beschreiben („deskriptive Statistik“), spielt ihre absolute Größe keine Rolle, die errechneten Prozentzahlen sind auf jeden Fall valide. Die Aussage „die Kölner Geldbörsen unserer Stichprobe umfassen 5,3 % Berliner Prägungen“ ist eine treffende und korrekte Beschreibung, selbst wenn uns die Stichprobe mit  $n = 585$  als nicht sonderlich groß erscheint. Ein ganz anderer Fall liegt vor, wenn wir aufgrund einer begrenzten Erhebung Rückschlüsse auf eine dahinterstehende, uns jedoch unbekanntes Allgemeinheit treffen wollen („schließende Statistik“), im vorliegenden Fall im Grunde über den Münzumsatz in Deutschland generell. Oder genauer: den Vergleich des Münzumsatzes in Köln, einer Stadt mit ca. 1 Mio. Einwohnern, und Göttingen, einer Stadt mit ca. 120.000 Einwohnern. Legen wir die Zahlen der Wikipedia zum Münzbestand

in Deutschland aus dem Jahr 2000 zugrunde, kamen auf jeden Bundesbürger circa 600 Münzen,<sup>44</sup> woraus sich ableiten lässt, dass die Münzmenge in Köln und Göttingen, zu der hier Aussagen getroffen werden, zusammengenommen bei circa 672 Mio. Stück lag. Kann man auf Basis von gerade mal 1.500 Beobachtungen valide Aussagen über einige hundert Millionen treffen? Ja, man kann, mit abschätzbaren (Un-)Sicherheiten. Im Alltag erlebt man dies immer wieder bei Verbraucher- oder Wahlumfragen, bei denen meist nur eine relativ geringe Anzahl an Befragungen zur Verfügung steht, während die Zahl der Verbraucher oder Wähler, über die Aussagen getroffen werden, ungleich größer ist. In diesem Fall zielt die Beobachtung der Stichprobe auf eine Schätzung der tatsächlichen Prozentzahl in der Grundgesamtheit. Intuitiv wissen wir, dass die Schätzung bei einer kleinen Erhebung eher unsicher ist und bei größeren Stichproben vermutlich zuverlässiger sein dürfte.

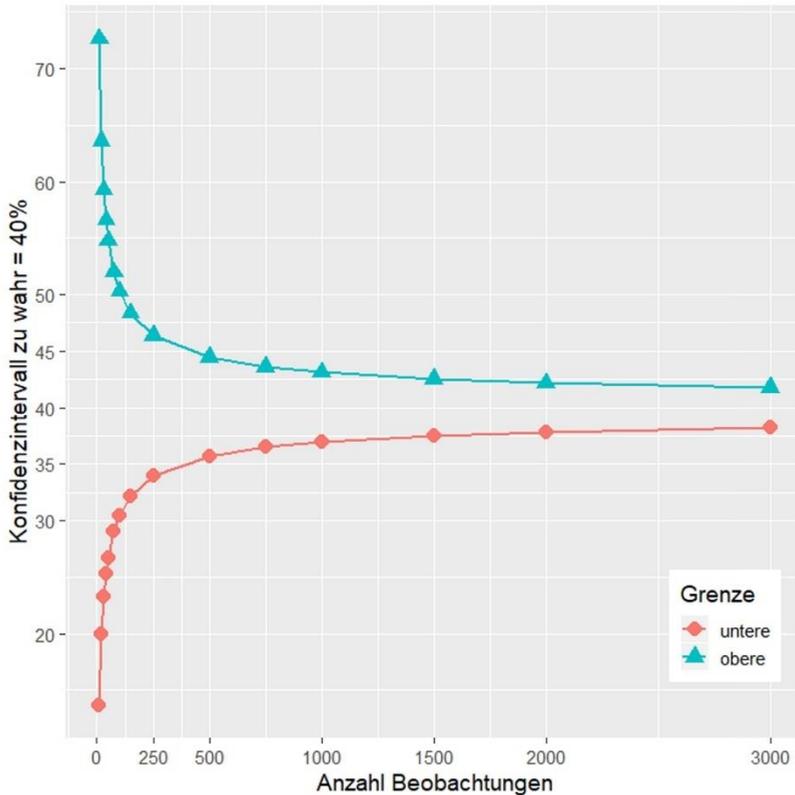
Eine erste Hilfe seitens der Statistik ergibt sich aus dem Fachwort „Vertrauensbereich (Konfidenzintervall) einer beobachteten Häufigkeit aus einer binomialverteilten Grundgesamtheit“. Mit dem Stichwort „Binomialverteilung“ und Verteilungen insgesamt werden wir uns später noch vertiefend beschäftigen; für den Einstieg mag der Hinweis genügen, dass Prozentzahlen typischerweise einer Binomialverteilung folgen. Das Errechnen der Konfidenzintervalle beruht auf der Binomialverteilung.

Der Vertrauensbereich gibt an, in welcher Spanne der Prozentwert in der Grundgesamtheit tatsächlich liegt, wenn wir ihn aufgrund einer konkreten Stichprobe schätzen. Besser als Zahlen verdeutlicht das Schaubild **Abb. 57** die Sachlage.

In Worten: In einer (hypothetischen) Stichprobe von zehn Fällen beobachte ich vier „Blau“, d. h. 40 %. Das Konfidenzintervall liegt bei 12,2 – 73,8 %, d. h. in einer dahinterstehenden (unendlich großen) Grundgesamtheit wird der tatsächliche Anteil der „Blauen“ sicher zwischen etwa 12 und 74 % liegen. Ist meine Stichprobe 50 Beobachtungen groß mit 20 „Blauen“ gleich 40 %, liegt das Konfidenzintervall bei 26,4 – 54,5 %. Mit zunehmender Stichprobengröße wird das Konfidenzintervall zunehmend enger, d. h. die Schätzung des tatsächlichen Anteils in der Grundgesamtheit immer genauer.

---

<sup>44</sup> Wikipedia.de (25.11.2019). Bargeld der Deutschen Mark. [https://de.wikipedia.org/wiki/Bargeld\\_der\\_Deutschen\\_Mark](https://de.wikipedia.org/wiki/Bargeld_der_Deutschen_Mark) [24.1.2020]. Die zunächst sehr hoch wirkende Zahl pro Bundesbürger dürfte damit zusammenhängen, dass um das Jahr 2000 sehr viel deutsches Bargeld auch außerhalb von Deutschland im Umlauf war.



**Abb. 57** Darstellung der Konfidenzintervalle zu einer Stichprobe, in der die untersuchte Beobachtung jeweils einen errechneten Anteil von 40 % hat. Das Konfidenzintervall (zwischen Ober- und Untergrenze) zeigt an, dass mit zunehmender Stichprobengröße (x-Achse) der Bereich, in dem der tatsächliche Anteil in der Grundgesamtheit liegen kann, zunehmend enger wird.

Marktforscher und Wahlbeobachter stützen auf diese statistische Theorie ihre ökonomische Kosten-Nutzen-Abwägung bei der Versuchsplanung. Ab einer Stichprobengröße von 1.000 ist der weitere Genauigkeitsgewinn bei einer Verdoppelung der Stichprobe nicht mehr sehr groß, der Aufwand aber viel höher. Daher sind bei Wahl- und Marktforschern Stichproben von 1.000 oder 2.000 üblich, mehr gilt als unökonomisch.

Die Beobachtung der Einschaltquoten beim Fernsehen, die ja heute mit erheblichen Einnahmen aus der Werbung verbunden ist, beruhte in Deutschland am

Anfang auf 650 Haushalten. In den Jahren 1975 bis 1984 wurden 1.200 Fernsehhaushalte beobachtet, 1985 bis 1990 waren es ca. 2.700 Haushalte.<sup>45</sup> Danach wurde die Stichprobe weiter aufgestockt, wobei sie vor allem angesichts der neuen Bundesländer für eine bessere regionale Abdeckung sowie genauere Erfassung von EU-Ausländern gezielt vergrößert wurde (sog. geschichtete Stichprobe), aber auch die Anzahl der zu beobachteten Sender sehr groß geworden ist. Seit dem 1.1.2019 werden die Daten von 5.200 Fernsehhaushalten abgerufen.<sup>46</sup>

### 18.3.1 Berechnen von Konfidenzintervallen mit R

Konfidenzintervalle präzise zu errechnen, ist aufwendig. Daher benutzte man ehemals üblicherweise entweder etwas vereinfachte Näherungsformeln oder Tabellenwerke. Da die Berechnungsgrundlagen der verschiedenen Tabellenwerke leicht differieren, sollte man in diesen Fällen z. B. in einer Fußnote immer angeben, welchem Tabellenwerk man die Zahlen entnommen hat. Zu den „Werkzeugen“ dieses Buchs gehört die LO-Calc-Tabelle „Konfidenz“, die nach einem gängigen Verfahren solche Konfidenzintervalle berechnet. Genauer sind die Werte z. B. bei Weber (1986), oder in den viel verwendeten Tafeln „Wissenschaftliche Tabellen Geigy (1980)“.<sup>47</sup> Besonders komfortabel und anschaulich kann man die Berechnung mit **R** durchführen, genauer der Funktion „prop.test“. Die Befehlszeile(n) für unser Beispiel mit 40 % und 10, 100, 1.000 oder 2.000 Beobachtungen sähe wie folgt aus:

```
prop.test(4, 10, conf.level=0.95)
prop.test(40, 100, conf.level=0.95)
prop.test(400, 1000, conf.level=0.95)
prop.test(800, 2000, conf.level=0.95)
```

Man setzt also in die Funktion prop.test nicht den Prozentwert ein, sondern die beobachtete Häufigkeit (Anzahl) und die Bezugssumme (Hundertsumme). Die (einzelne) Ausgabe von **R** lautet z. B. (unten):

---

<sup>45</sup> Im Jahr 1984 begann in Deutschland neben dem staatlich finanzierten Fernsehen auch das sog. Privatfernsehen, 1985 starteten u.a. Sat.1 und RTL als Fernsehsender. Mehr Sender bedeutet statistisch ein Mehr an zu beobachtenden Kategorien, weshalb eine Erhöhung der Stichprobengröße nötig wurde.

<sup>46</sup> Ausführlicher mit Literaturnachweisen: Siegmund 2000, 110-114; AGF Videoforschung GmbH, *Fernsehpanel: Die Entwicklung und Zusammensetzung des Fernsehpanels*. <https://www.agf.de/forschung/methode/fernsehpanel/> [24.1.2020].

<sup>47</sup> Wissenschaftliche Tabellen Geigy (8. Aufl. 1980), Seite 89-102 mit Werten für Stichprobengrößen von 1-100. Man wähle die mittlere Spalte mit alpha 0,05 bzw. 95 %. Sodann: Weber, Erna (1968). *Grundriß der biologischen Statistik*. 9. Aufl. (p. 612-624 Taf. 22). Stuttgart: Gustav Fischer.

```
95 percent confidence interval:  
 0.3047801 0.5029964  
sample estimates:  
 p  
0.4
```

... was besagt, dass der beobachtete Wert von 40 % („ $p = 0.4$ “) mit einer 95%-igen Sicherheit auf eine Grundgesamtheit schließen lässt, in der die betreffende Größe einen tatsächlichen Anteil von ca. 30,5 % bis 50,3 % hat.

Der Vergleich der Ergebnisse verdeutlicht nochmals, wie stark die Schritte von 10 zu 100 zu 1.000 Beobachtungen das Konfidenzintervall (d. h. die Unsicherheit) verkleinert und zeigt zugleich, dass die Verdoppelung des Aufwandes von 1.000 Beobachtungen zu 2.000 Beobachtungen nur relativ wenig mehr Sicherheit erwirkt. Wenn Sie diese Rechnungen mit **R** durchführen, können Sie durch Modifizierung der Eingaben zudem spielerisch erfahren, was geschieht, wenn man das Irrtumsrisiko von 0.95 (5 von 100 Fällen dürfen außerhalb dieser Schrankenwerte liegen) auf 0.99 verringert (1 von 100 Fällen darf außerhalb dieser Schranken liegen): nicht sonderlich viel. Der absolute Umfang der Stichprobe ist der wesentlich wirksame Parameter.

In den Sozialwissenschaften oder in der Medizin, beispielsweise, werden diese und ähnliche Tatsachen genutzt, um Untersuchungen vorab sorgfältig zu planen, d. h. unter Vorgabe einer zu erreichenden Genauigkeit den nötigen Umfang einer Erhebung festzulegen. Das ist jedoch nicht trivial, da es dazu zuverlässigen Vorwissens aus älteren Untersuchungen bedarf insbesondere zur Frage, welcher Verteilung die Daten unterliegen. Vor allem ist dies jedoch eine in der Archäologie eher selten gegebene Situation, denn meist können wir – auch wenn es wünschenswert wäre – den Umfang einer Stichprobe nicht vergrößern (Endlichkeit der Quellen), und würden umgekehrt die Bearbeitung einer Fundstelle wohl auch kaum einstellen, wenn punktuell ein hinreichendes Niveau erreicht ist.

### 18.3.2 Asymmetrie der Konfidenzintervalle

Wer **Abb. 57** und die errechneten Werte der Konfidenzintervalle sorgfältig studiert hat, wird gesehen haben, dass diese unsymmetrisch sind. Ja! Das hängt mit ihrer „Natur“ zusammen, d. h. der Tatsache, dass sie einer sog. Binomialverteilung unterliegen. Relative Häufigkeiten nehmen Werte zwischen 0 und 1 an, Prozentzahlen Werte zwischen 0 und 100. Bei einem Wert von 0,5 resp. 50 % haben zufällige Abweichungen gleiche Möglichkeiten, größer oder kleiner auszufallen, bei 50 % ist das Konfidenzintervall symmetrisch. Bei einem Wert von z. B. 5 % sind zufällige Abweichungen in Richtung 100 % gut möglich, in Richtung auf 0 jedoch weniger, bildlich gesprochen:

haben weniger Platz. Daher fallen Konfidenzintervalle, je näher sie an 0 oder an 100 % liegen, desto unsymmetrischer aus.

### 18.3.3 Berechnen von Konfidenzintervallen mit PAST

Hierzu ist in PAST keine bereits vorliegende Datentabelle nötig, vielmehr wird ein Berechnungsfenster angeboten. => Univariate; => Single proportion; es poppt ein neues Menü auf, in dem die Eingabe wie die Berechnung erfolgt. Bei „observed proportion“ den beobachteten Prozentwert eingeben („proportion“! – hier also den Prozentwert, nicht die Zahl der beobachteten Fälle), z. B. 0,4 (für 40 %); „Total sample size N“ meint die Größe der Gesamtstichprobe, d. h. die Anzahl der Beobachtungen insgesamt, eingeben. Dann „Compute“.

Gibt man in dem Feld darunter zusätzlich noch die „Hypothetical proportion“ ein (treffender wäre der Begriff „expected proportion“), wird zusätzlich ein statistischer Test durchgeführt, ob der beobachtete Anteil von dem erwarteten Anteil signifikant abweicht. Ein praktischer Anwendungsfall wäre z. B. das beobachtete Männer-Frauen-Verhältnis auf einem Gräberfeld, wo man meist ein 50 : 50 %-Verhältnis erwartet und nun prüfen möchte, ob die beobachtete Relation wirklich signifikant von dieser Erwartung abweicht. Doch da wir solche Fragen, die im Themenbereich Teststatistik liegen, erst später besprechen, hier nur der Hinweis ohne tiefere Erläuterung.

Statt der Einzelberechnungen kann man in PAST auch eine ganze Wertereihe eingeben und gleichzeitig berechnen lassen. Nun also zwei Spalten in der Tabelle anlegen: erste Spalte mit Prozentwert (Achtung: hier tatsächlich „40“ für 40 %), zweite Spalte mit der Größe der Gesamtstichprobe. Nach Eingabe der Reihe Bereich markieren, dann: => Univariate; => Multiple proportion CIs. Das aufpoppende Menü spielt die Ergebnisse aus, als Plot wie in Zahlen. Man kann den Plot sehr schön nutzen, um nach eigenen Vorstellungen und Zahlen eine Grafik ähnlich **Abb. 57** zu gewinnen.

### 18.3.4 Caveat: endliche Grundgesamtheiten

Genau genommen besteht ein Unterschied zwischen dem Rückschluss auf „endlich große Grundgesamtheiten“ und „unendlich große Grundgesamtheiten“ – also bei der Frage, ob tatsächlich allgemeingültige Aussagen für eine unendlich große Gruppe angestrebt sind, oder ob die Schätzung auf eine begrenzt große Gruppe zielt. Aussagen über alle Menschen wären der Kategorie „unendlich groß“ zuzuordnen; zielt die Untersuchung hingegen auf „französisch sprechende Schweizerinnen im Alter zwischen 8 und 10 Jahren“, müssten die Formeln für endlich große Grundgesamtheiten angewendet werden. Wie ein praktischer Vergleich beider unterschiedlicher Formeln

schnell zeigt, ist der Unterschied zwischen beiden Ansätzen jedoch gering, sodass üblicherweise die Formeln resp. Tabellen für unendlich große Grundgesamtheiten verwendet werden können.

Schauen wir uns die Wirkung einer Berechnung für endliche Grundgesamtheiten an einem Beispiel an, nämlich 400 Treffern aus 1000 Fällen, also 40 %. Mit dem Befehl:

```
prop.test(400, 1000, conf.level=0.95)
```

gibt **R** folgendes Konfidenzintervall für unendliche Grundgesamtheiten aus:

36,96 – 43,12 %.

Wendet man die bei Sachs & Hedderich (2006, 257 Formel 6.32) angegebenen Formeln für endliche Grundgesamtheiten an, ergeben sich die in **Abb. 58** tabellierten 95 %-Konfidenzintervalle, für jeweils endliche Grundgesamtheiten ab 1.000. Im Vergleich zu den Konfidenzintervallen für unendliche Grundgesamtheiten fallen jene für endliche Grundgesamtheiten also zunächst etwas enger aus, d. h. sicherer. Bei endlichen Grundgesamtheiten spätestens ab etwa 10.000 unterscheidet sich das Ergebnis jedoch nicht wesentlich von jenem für unendliche Grundgesamtheiten. Wir erkennen an diesem Beispiel, dass (a) das „caveat“ nur in Ausnahmefällen berücksichtigt werden muss, und (b) die Konfidenzintervalle bei endlichen Grundgesamtheiten zunächst etwas enger ausfallen.

Grundgesamtheit	% Untergrenze	% Obergrenze	% Spanne
1.000	39,95	40,05	0,10
2.000	37,80	42,20	4,40
5.000	37,23	42,77	5,54
10.000	37,07	42,93	4,86
50.000	36,94	43,06	6,12
100.000	36,93	43,07	6,14
500.000	36,92	43,08	6,16
1.000.000	36,92	43,08	6,16

**Abb. 58** Konfidenzintervalle für 400 aus 1.000 Beobachtungen, bezogen auf unterschiedlich große, endliche Grundgesamtheiten.

### 18.3.5 Konfidenzintervall: Anwendungsbeispiel

Der Vergleich der Konfidenzintervalle für die Gesamtstichprobe (Köln und Göttingen) in **Abb. 59** mit den für Köln und Göttingen beobachteten Prozentwerten in **Abb. 56** zeigt, dass nicht mehr alle der Werte in den Bereich zwischen die Intervalle fallen:

In der Kölner Stichprobe liegen die Werte für Berlin und Hamburg unterhalb der Untergrenze, für Stuttgart und Karlsruhe oberhalb der Obergrenze, und in der Göttinger Stichprobe liegen die Werte für Karlsruhe knapp unterhalb der Untergrenze. Liegt ein beobachteter Prozentwert außerhalb der Grenzen eines Konfidenzintervalls seiner Vergleichsstichprobe, liegt der Verdacht nahe, dass er nicht zu gleichen Population wie die Vergleichsstichprobe gehört. Folglich erhärtet die Betrachtung über die Konfidenzintervalle unseren ersten Verdacht, dass die Geldbörsen in Köln und jene aus Göttingen unterschiedlich sind.

Dies hätten wir gerne konkreter gefasst. Zudem überdenken wir gerade, dass in der Archäologie einerseits Fragen dieser Art üblich sind, andererseits Beobachtungen von immerhin 1.500 Fällen oft nicht vorliegen, wir also normalerweise mit erheblich breiter gefassten Konfidenzintervallen als in unserem Beispiel operieren (müssen). Benötigt man also stets eine so große Stichprobe, oder gibt es einen statistischen Ansatz, der auch für geringere Stichprobengrößen taugt? Diese Frage leitet über zum nächsten Thema, das in der Tat hilft, diesen wichtigen Sachverhalt schärfer zu fassen: Vier- oder Mehrfeldertafeln, Chi-Quadrat-Wert ( $\chi^2$ ) und Chi-Quadrat-Test.

Prägeort	Summe	%	untere Grenze	obere Grenze
Berlin (A)	118	7,7	6,4	9,2
München (D)	290	19,0	17,0	21,0
Stuttgart (F)	394	25,8	23,6	28,0
Karlsruhe (G)	254	16,6	14,8	18,6
Hamburg (J)	474	31,0	28,7	33,4
Summe	1530	100		

**Abb. 59** Beobachteter Prozentwert und untere wie obere Schranke des zugehörigen Konfidenzintervalls zu den Häufigkeiten der Prägeorte der DM-Münzen in den Geldbörsen aus Köln und Göttingen (n=1.530).

#### 18.4 KONTINGENZTAFELN (ALIAS KREUZTABELLEN, MEHRFELDERTAFELN) UND DAS CHI-QUADRATVERFAHREN

Das folgende Kapitel führt grundlegend in ein wichtiges statistisches Testverfahren ein und braucht etwas Geduld, weil man eigentlich alles zugleich wissen resp. erklären muss, dies aber kaum möglich ist, sprich: offene Fragen nicht stets an dem Punkt beantwortet werden können, wo sie sich stellen, sondern nacheinander abgearbeitet werden. Ich hoffe aber, dass am Ende der folgenden sieben Seiten alles Relevante besprochen ist.

Fangen wir mit dem Namen an. Wenn man zwei verschiedene Kategorien miteinander verbindet, wie hier die Prägeorte und die Orte, an denen Stichproben erhoben wurden, ergeben sich zweidimensionale Bezüge, die wir auch im Alltag gewöhnlich als Tabelle darstellen. Solche Tabellen, in deren Zellen Häufigkeiten dargestellt sind, nennt man in der Statistik ebenfalls Tabellen (*tables*), alternativ aber auch Häufigkeitstabellen, Kreuztabellen (*cross tabulation; crosstabs; 2-way table; two-way table*) oder Kontingenztabelle (*contingency table*); oft wird je nach Zahl der Zellen auch speziell von einer Vierfeldertafel ( $2 \times 2$  *contingency table*) oder einer Mehrfeldertafel gesprochen. Wissenschaftlich korrekt wäre es, durchgängig von „Kontingenztabelle“ zu sprechen, da die erste moderne wissenschaftliche Behandlung solcher Tabellen von Karl Pearson stammt, der sie eben *contingency table* nannte.<sup>48</sup> Üblicherweise zeigen die Kopfzeile dieser Tabellen und die linke Spalte die inhaltliche Bedeutung der Zeilen (*rows*) und Spalten (*columns*) an (**Abb. 60**). Meist wird rechts eine Spalte mit den Zeilensummen angefügt und unten eine Zeile mit den Spaltensummen, zusammenfassend auch Randsummen (*marginal totals*) genannt, sodass aus der Zelle rechts unten die Gesamtzahl der Beobachtungen (*grand total*) hervorgeht. Eine gängige statistische Prozedur zur Analyse dieser Tabellen ist das Chi-Quadrat-Verfahren (weil etwas über Kreuz gerechnet wird), das oft auch als „ $\chi^2$ -Verfahren“ geschrieben wird, bzw. man spricht von  $\chi^2$ -Test. Wir überführen also die Informationen aus **Abb. 59** in eine etwas modifizierte Tabelle mit den beobachteten Häufigkeiten; unten wie am rechten Rand der Tabelle werden die Spalten- und die Zeilensummen ausgewiesen (**Abb. 60**).

Prägeort	Köln	Göttingen	Summe
Berlin (A)	31	87	118
München (D)	108	182	290
Stuttgart (F)	165	229	394
Karlsruhe (G)	116	138	254
Hamburg (J)	165	309	474
Summe	585	945	1530

**Abb. 60** Wiederholung der Tabelle **Abb. 59** nun mit Spalten und Zeilensummen, allerdings ohne Eintragung der Prozentwerte.

Oft trägt man in die Zellen zusätzlich zu den beobachteten Häufigkeiten auch Prozentwerte ein. Dabei gibt es drei sehr verschiedene Modi, zwischen denen man je nach

<sup>48</sup> Pearson, Karl (1904). On the theory of contingency and its relation to association and normal correlation. In *Mathematical contributions to the theory of evolution*, 13. London: Dulau & Co.

Fragestellung wählen kann und sollte. Selbstverständlich ist es auch möglich, alle drei Prozentwerte in die Tabelle einzutragen, aber wirklich lesbar sind solche Tabellen nicht. Drei Prozentwerte? – ja: Spaltenprozent, Zeilenprozent und Gesamtprozent, je nachdem, auf welche Summe die Prozentwerte bezogen werden. Nimmt man die Gesamtsumme in einer Spalte als Bezugssumme für die Prozentwerte in dieser Spalte, lassen sich verschiedenen Spalten gut vergleichen; ähnlich führt die Gesamtsumme der Häufigkeiten in einer Zeile zu Zeilenprozenten, die wiederum den Vergleich zweier Zeilen erleichtert. Bei Gesamtprozenten bildet die Gesamtzahl aller Beobachtungen die Hundertsumme. Entsprechend unserem zuvor entwickelten Verdacht, dass die Erhebungsorte Köln und Göttingen nicht die gleichen Prägeorte aufweisen, haben wir uns im Beispiel **Abb. 61** für Spaltenprozent entschieden; diese sind durch das Prozentzeichen und den Kursivsatz markiert, sodass man beide Zahlenreihen optisch gut unterscheiden und die Werte, auf denen das Augenmerk liegt, gut vergleichen kann.

Prägeort	Köln	Göttingen	Summe
Berlin (A)	31 <i>5,3 %</i>	87 <i>9,2 %</i>	118
München (D)	108 <i>18,5 %</i>	182 <i>19,3 %</i>	290
Stuttgart (F)	165 <i>28,2 %</i>	229 <i>24,2 %</i>	394
Karlsruhe (G)	116 <i>19,8 %</i>	138 <i>14,6 %</i>	254
Hamburg (J)	165 <i>28,2 %</i>	309 <i>32,7 %</i>	474
Summe	585 <i>100 %</i>	945 <i>100 %</i>	1530

**Abb. 61** Wiederholung der Tabelle **Abb. 60**, nun mit Eintragung der Spaltenprozent.

Gleichwie: Für das nachfolgend entwickelte Chi-Quadrat-Verfahren verlassen wir diese Prozentzahlen wieder, denn sie führen bald schon in die schon zuvor erkannte Problematik der Konfidenzintervalle zurück, wo wir nicht sicher waren, ob sich die Stichproben Köln und Göttingen nun eindeutig unterscheiden oder nicht. Das Chi-Quadrat-Verfahren setzt unabhängig von Prozentwerten an: Ausgangspunkt ist die im Grunde naheliegende Hypothese: Die Stichproben Köln und Göttingen stammen aus einem großen, übergreifend gleichen gemeinsamen Versorgungspool mit Bargeld in Deutschland – eine Annahme, die man „Nullhypothese“ nennt. Wenn dem so ist, sind

die Häufigkeiten in den einzelnen Zellen allein von den Randsummen abhängig. Konkret: Weil Berlin wenig prägt, gibt es sowohl in der Zelle Berlin/Köln wie Berlin/Göttingen relativ kleinere Zahlen. Weil die Göttinger Stichprobe insgesamt jedoch circa 1,6 Mal so viele Münzen als die Kölner Serie umfasst, müssten die Häufigkeiten für Göttingen allesamt entsprechend etwas höher ausfallen als für Köln. Aus dieser Überlegung ergibt sich das Konzept und die Berechnung eines „Erwartungswertes“. Er weist in den einzelnen Zellen konkret aus, welchen Wert man erwarten würde, wenn er allein von den Randsummen abhinge. Die Formel lautet:

$$\text{Erwartungswert} = \text{Spaltensumme} \times \text{Zeilensumme} / \text{Gesamtsumme}.$$

In diesem Sinne fügen wir der Ausgangstabelle **Abb. 60** die Erwartungswerte hinzu (**Abb. 62**).

Prägeort	Köln	Göttingen	Summe
Berlin (A)	a	f	118
München (D)	b	g	290
Stuttgart (F)	c	h	394
Karlsruhe (G)	d	i	254
Hamburg (J)	e	k	474
Summe	585	945	1530

a: $585 \times 118 / 1.530$	f: $945 \times 118 / 1.530$
b: $585 \times 290 / 1.530$	g: $945 \times 290 / 1.530$
c: $585 \times 394 / 1.530$	h: $945 \times 394 / 1.530$
d: $585 \times 254 / 1.530$	i: $945 \times 254 / 1.530$
e: $585 \times 474 / 1.530$	k: $945 \times 474 / 1.530$

**Abb. 62** Wiederholung der Tabelle **Abb. 60** mit Eintragung der Formel für die jeweiligen Erwartungswerte. Vgl. **Abb. 63** mit den Ergebnissen.

Prägeort	Köln	Göttingen	Summe
Berlin (A)	31	87	118
	45,1	72,9	
München (D)	108	182	290
	110,9	179,1	
Stuttgart (F)	165	229	394
	150,6	243,4	
Karlsruhe (G)	116	138	254
	97,1	156,9	
Hamburg (J)	165	309	474
	181,2	292,8	
Summe	585	945	1530

**Abb. 63** Die Häufigkeiten (schwarz) aus Tabelle **Abb. 60** mit Eintragung der Erwartungswerte (darunter; rot). Zugunsten der Übersichtlichkeit sind die Erwartungswerte nur mit einer Stelle hinter dem Komma berechnet, tatsächlich rechnet man üblicherweise auf mind. drei Dezimalstellen genau.

Nach Ausführung der Berechnungen sehen wir, dass die Erwartungswerte (rot) recht nah bei den beobachteten Werten (schwarz) liegen, aber nicht identisch sind. Wie lässt sich dieser Eindruck objektivieren, d. h. die Ähnlichkeit oder die Differenz messen? Indem man die beobachteten Werte mit den Erwartungswerten vergleicht und die sog. (normierten) „Abweichungsquadrate“ ermittelt,<sup>49</sup> nach der Formel:

$$(\text{Beobachtet} - \text{Erwartet}) \text{ quadriert} / \text{Erwartet.}$$

---

<sup>49</sup> Der Formelteil „(Beobachtet - Erwartet) quadriert“ ist das Abweichungsquadrat. Dieses wird anschließend wieder durch den Erwartungswert dividiert, womit die absolute Größe „weggerechnet“ wird, weshalb man von „normiert“ spricht.

Prägeort	Köln	Göttingen	Summe
Berlin (A)	31	87	118
	45,1: 4,42	72,9: 2,73	
München (D)	108	182	290
	110,9: 0,07	179,1: 0,05	
Stuttgart (F)	165	229	394
	150,6: 1,37	243,4: 0,85	
Karlsruhe (G)	116	138	254
	97,1: 3,67	156,9: 2,27	
Hamburg (J)	165	309	474
	181,2: 1,45	292,8: 0,90	
Summe	585	945	1530

**Abb. 64** Die Häufigkeiten (schwarz) aus Tabelle **Abb. 60** mit Eintragung der Erwartungswerte (darunter; rot) und daneben den normierten Abweichungsquadraten (blau). Zugunsten der Übersichtlichkeit sind die Abweichungsquadrate nur mit zwei Stellen hinter dem Komma berechnet, tatsächlich rechnet man üblicherweise genauer.

Wir fügen die berechneten (normierten, relativierten) Abweichungsquadratrate in die Zellen hinzu (blau) (**Abb. 64**). Beim Vergleich der normierten Abweichungsquadrate / Chi-Werte der einzelnen Zellen – also der blauen Werte – erkennt man, dass die Zahlen da kleiner sind, wo wir bereits anhand der Prozente keine Besonderheiten identifiziert hatten. Da, wo wir zuvor eventuell abweichende Prozentzahlen identifiziert hatten, wie z. B. bei den in Berlin geprägten Münzen, sind die Abweichungsquadrate größer. Sprich: Diese Zahlen objektivieren und messen den Befund, dass einige Häufigkeiten dem gemeinsamen Muster gänzlich folgen, andere stärker davon abweichen. Addiert man die Werte der Abweichungsquadrate für die ganze Tabelle, erhält man den sog. Chi-Quadrat-Wert. Er beträgt hier 17,786.

#### 18.4.1 Kritischer Wert von Chi-Quadrat und Chi-Quadrat-Test

Die weitere Idee des Verfahrens ist die Annahme, dass zufällige Abweichungen vom idealen Wert (Erwartungswert) immer wieder vorkommen werden, sodass der Chi-Quadrat-Wert einer ganzen Tabelle (der ja diese Abweichungen zusammenfassend misst) fast nie den Wert Null annehmen wird. Die erwarteten zufälligen Abweichungen folgen – so das Modell des Verfahrens – der sog. Chi-Quadrat-Verteilung, einer theoretischen Wahrscheinlichkeitsverteilung, anhand derer sich ermitteln lässt, wie weit die Spanne des Zufalls geht und ab wann sich Abweichungen im Bereich des Nicht-mehr-Zufälligen bewegen. Denn irgendwann, so das Modell, ist das Ausmaß an Abweichung, das allein durch den Zufall erklärt werden kann, überschritten. Ist das an

einer Tabelle tatsächlich berechnete Chi-Quadrat größer als der sich aus der Chi-Quadrat-Verteilung ergebende „kritische Wert“ (auch Schwellenwert genannt), sind die Abweichungen nicht mehr allein durch den Zufall erklärbar, sprich: Es liegen tatsächlich statistisch signifikante Abweichungen vor.

Tun wir das: Im vorliegenden Fall beträgt Chi-Quadrat 17,786; demgegenüber liegt der kritische Wert (Schwellenwert) bei unserer Tabelle von 5 x 2 Zellen bei 9,488 (die Erklärung folgt sogleich), das beobachteten Chi-Quadrat unserer Tabelle ist also größer. Ergebnis: In der vorliegenden Tabelle sind die Abweichungen zwischen den beobachteten Werten und den Erwartungswerten insgesamt größer als es im Rahmen des Zufälligen möglich ist. Die Unterschiede sind signifikant und dürfen historisch-inhaltlich interpretiert werden. Die hier erfassten Geldbörsen aus Köln und die Geldbörsen aus Göttingen entstammen nicht einem gemeinsamen übergreifenden Pool an Münzgeld, sondern zwei verschiedenen Pools.

Ein Blick weg vom Gesamt-Chi-Quadrat der ganzen Tabelle in die einzelnen Chi-Werte in den Zellen – also die blauen Zahlen in **Abb. 64** – in den Zellen signalisiert uns, wo die Abweichungen besonders groß sind, d. h. in stärkerem Maße zum Gesamtergebnis der Tabelle beitragen. Es sind dies vor allem die Berliner Prägungen (insbes. Zelle Köln & Berlin mit gegenüber dem Erwartungswert niedrigem beobachtetem Wert) und die Karlsruher Prägungen (insbes. Zelle Karlsruhe & Köln mit gegenüber dem Erwartungswert erhöhtem beobachtetem Wert). Wir kommen auf dieses Thema (Stichwort Konfigurationsfrequenzanalyse, KFA) später noch einmal zurück.

Kurzes Nachdenken: Denkt man in den vertrauteren Kategorien von Prozentzahlen, stehen bei solchen Tabellen stets drei Betrachtungsoptionen zur Verfügung: (a) Spaltenprozent, die Prozentwerte werden über die Gesamtsumme einer Spalte berechnet, hier also „alle Kölner“ resp. „alle Göttinger“ Geldbörsen jeweils gleich 100 %; (b) Zeilenprozent, die Prozentwerte werden über die Gesamtsumme einer Zeile berechnet, also z. B. alle Berliner Prägungen bilden 100 Prozent, usw.; (c) Gesamtprozent, d. h. die Summe der Gesamthäufigkeiten (hier: 1.530) bildet die Grundlage der Berechnung aller Zellenprozent.<sup>50</sup> Das Chi-Quadrat-Verfahren verwendet die Prozentrechnung nicht, da diese – wie wir später noch sehen werden – statistisch gesehen erheblich Schwächen aufweist. Aber im Grunde wird durch die Chi-

---

<sup>50</sup> Weshalb es im Übrigen – je nach Frageperspektive – auch jeweils drei verschiedene Konfidenzintervalle gäbe.

Quadratrechnung gleichzeitig über die Spalten wie über die Zeilen gedacht und geprüft, ob die Werte in den einzelnen Zellen auffällig hoch oder auffällig niedrig ausfallen.

#### 18.4.2 Zwei wichtige Begriffe: Freiheitsgrad und Irrtumswahrscheinlichkeit

Der „kritische Wert“ ergibt sich aus der Chi-Quadrat-Verteilung; es ist aber nicht eine einzige fixe Zahl, vielmehr ist der jeweilige kritische Wert von zwei Parametern abhängig, vom „Freiheitsgrad“ und vom „Irrtumsrisiko“ – zwei Begriffe, die wir uns noch erarbeiten müssen.

Lassen wir die rechnerische Seite zunächst einmal beiseite und versuchen zu verstehen, welches inhaltliche Problem der Begriff „Freiheitsgrad“ löst. Berechnet man den Chi-Quadrat-Wert für eine kleine Tabelle von z. B. zwei Spalten und zwei Zeilen – eine sog. „Vierfeldertafel“ – wird der kritische Wert, bis zu dem man Abweichungen für zufällig nehmen kann, klein ausfallen. Schließlich tragen nur vier Zellen zum Gesamtwert der möglichen zufälligen Abweichungen bei. Bei unserer Münztabelle mit ihren zwei Spalten und immerhin fünf Prägeorten/Zeilen tragen bereits zehn Zellen zum Gesamtwert von Chi-Quadrat bei. Räumt man jeder einzelnen Zelle das gleiche Maß an zufälliger Abweichung ein, ist bei einer größeren Tabelle der kritische Wert, bis zu dessen Erreichen man noch von Zufall sprechen kann, entsprechend höher. Genau dieses Phänomen quantifiziert man mit dem Begriff „Freiheitsgrad“. Die Zahl der Freiheitsgrade (FG) – oft auch mit griech.  $\nu$  („nü“) bezeichnet – steht für die Anzahl der „freien Zellen“ einer Tabelle. Ihre Zahl ergibt sich aus:

(Anzahl der Zeilen minus 1) mal (Anzahl der Spalten minus 1).

In unserer Tabelle **Abb. 64** sind es also  $(5 - 1) \times (2 - 1) = 4 \times 1 = 4$  Freiheitsgrade. Bei einer  $2 \times 2$ -Feldertafel beträgt der Freiheitsgrad demnach 1.

Das Irrtumsrisiko  $\alpha$  („alpha“) formuliert, wie sicher ein Schluss sein soll resp. welche Menge an Irrtümern man zulassen möchte. Ein Irrtumsrisiko von  $\alpha$  0,05 besagt, dass in 5 von 100 Fällen eines solchen Tests die Schlussfolgerung falsch sein kann (nicht: muss). Weiterhin übliche Werte sind  $\alpha$  0,01 (1 von 100 Fällen möglicherweise irrig) und  $\alpha$  0,001 (1 von 1.000 Fällen möglicherweise irrig). In der Archäologie, wo es nicht um das Leben von Patienten oder die Flugsicherheit von Flugzeugen o. ä. geht, operiert man normalerweise mit einem  $\alpha$  von 0,05.

Das Zusammenspiel von Freiheitsgrad und Irrtumsrisiko wird verständlicher, wenn man sich einen Ausschnitt aus der Tabelle der kritischen Werte von Chi-Quadrat ansieht:

Freiheitsgrade	alpha 0,05	alpha 0,01	alpha 0,001
1	3,841	6,635	10,828
2	5,991	9,210	13,816
3	7,815	11,345	16,266
4	9,488	13,277	18,467
5	11,070	15,086	20,515

**Abb. 65** Kritische Werte von Chi-Quadrat (Auswahl), wie sie sich aus der Chi-Quadrat-Verteilung ableiten.

Der kritische Wert wächst mit der Zahl der Freiheitsgrade (FG): Je mehr Zellen eine Tabelle enthält, desto höhere Abweichungen sind rein zufällig möglich. Weiterhin lesen wir ab: Je niedriger „alpha“, das Irrtumsrisiko, desto höher der Schwellenwert. Dies führt uns zum Begriff der „Nullhypothese“. Es ist der letzte neue Begriff, den wir benötigen, um das ganze Verfahren zu verstehen – dann ist's geschafft!

### 18.4.3 Nullhypothese

Die Nullhypothese ( $H_0$ ) lautet: Die einzelnen Stichproben stammen aus einer gemeinsamen Grundgesamtheit, ihre Parameter (beobachteten Werte) unterscheiden sich nur im Rahmen des Zufälligen von den Erwartungswerten. Die Gegenhypothese/Alternativhypothese ( $H_1$ ) lautet: Die Stichproben unterscheiden sich statistisch signifikant, sie stammen nicht aus der gleichen Grundgesamtheit.

Der Chi-Quadrat-Test gilt als „konservativ“: Die Nullhypothese wird recht lange beibehalten, bis wegen Überschreitens des kritischen Wertes statt  $H_0$  richtig  $H_1$  gilt. Wegen des konservativen Charakters des Chi-Quadratverfahrens, weil in der Archäologie an einem Irrtum kaum jemand stirbt und weil in der Archäologie die Anzahl der Beobachtungen oft relativ niedrig ist, setzt man das Irrtumsrisiko üblicherweise mit alpha 0,05 relativ hoch an. Braucht man höhere Sicherheit, dann setzt man das Irrtumsrisiko niedriger an mit z. B. 0,01. Dann müssen die Abweichungen erheblich stärker sein, bis man von einem signifikanten Unterschied spricht. Das Vorgehen ist also hier – wie bei allen statistischen Tests –, dass grundsätzlich die Nullhypothese ( $H_0$ ) angenommen wird. Erst wenn diese sicher abgelehnt werden kann, gilt die Alternativhypothese ( $H_1$ ).

#### 18.4.4 Zwischenbilanz Chi-Quadrat-Test

Das Ausgangsproblem war die Frage nach der Einschätzung von Prozentwerten, die miteinander verglichen werden sollten. Sind Abweichungen so klein, dass sie zufällig sind, oder sind sie so stark, dass wir sie als nicht-zufällig interpretieren können? Das

Chi-Quadrat-Verfahren bietet einen soliden Ausweg aus der subjektiven Einschätzung von Prozentzahlen und deren Vergleich. Dazu ist es wichtig, die Daten und die Fragestellung sauber zu formulieren und entsprechend zu formalisieren. Das klare Formulieren einer Nullhypothese ist wichtig. Die Zahl der Freiheitsgrade kann der Leser aus den Tabellen selbst ersehen, nicht aber, welche Irrtumswahrscheinlichkeit vorausgesetzt wurde; zumindest „alpha“ ist also als zusätzliche Angabe (Fußnote, Abb.-Unterschrift o. ä.) nötig.

Noch Unerfahrene haben bisweilen zwar Informationen, finden aber den Hebel nicht, um überhaupt in ein Testverfahren einzusteigen. Das liegt nicht selten daran, dass man zwar positive Informationen vorliegen hat, aber der Vergleichsfall fehlt resp. man sich dessen Existenz nicht bewusst ist. Um bei unserem Beispiel **Abb. 60** zu bleiben: Man hat die Spalte Köln, aber den Rest nicht. Dann gilt es, sich den Vergleich zu suchen, z. B. aus der Literatur Vergleichsdaten zusammen zu tragen, oder grundsätzlich nachzudenken, wie es denn eigentlich sein sollte. Beispielsweise: Zu einem Gräberfeld liegen Geschlechtsbestimmungen vor. Und nun? Man kann andere Gräberfelder zusammensuchen und damit vergleichen, oder auch den Weg gehen: Wie ist denn die allgemeine, biologische Erwartung an das numerische Geschlechterverhältnis bei Menschen, und dann seine eigene Stichprobe gegen diese Erwartung testen.<sup>51</sup>

Ja, der Abschnitt war in Summe etwas länglich. Wir haben den Chi-Quadrat-Test hier ausführlich eingeführt, weil das Vergleichen von Häufigkeiten in der Archäologie wichtig ist und oft benötigt wird, und zugleich, weil dieser Test hier als Modell dient, wie statistische Test grundsätzlich funktionieren: Nullhypothese formulieren, darauf aufbauend ergeben sich Erwartungswerte, es wird eine Prüfgröße berechnet und das Ergebnis mit einem kritischen Wert (Schwellenwert) verglichen. Danach wird eine eindeutige Bilanz gezogen: Die Nullhypothese kann aufrechterhalten werden, oder sie ist abzulehnen.

#### 18.4.5 Chi-Quadrat-Test praktisch rechnen

*Mit einer Tabellenkalkulation:* Das Berechnen eines Chi-Quadrat-Tests ist mit einer Tabellenkalkulation eigentlich recht einfach. Wenn Sie sich mit etwas Energie an LO-Calc setzen und unser Beispiel **Abb. 63** bis **Abb. 64** nachbauen, hätte das den Vorteil,

---

<sup>51</sup> Siegmund, F. (2009). Der Maskulinitätsindex und statistische Verfahren zur Prüfung auf Unterschiede in den Geschlechteranteilen frühmittelalterlicher Populationen. *Bulletin der Schweizerischen Gesellschaft für Anthropologie*, 15, 5-17. (online via academia.edu).

dass Sie die Rechnereien und damit den Test wirklich verstanden hätten. Eine Beispieltabelle füge ich den Übungsdaten für dieses Buch bei. Der Vorteil dieses Vorgehens ist seine Anschaulichkeit, der Nachteil ist, dass man für jede Zeilen- und Spaltenzahl sich eigens eine passende Tabelle bauen muss. Gleichwie: Es ist wirklich lohnend, auch ganz spielerisch einmal Zahlen anzunehmen und in solch' eine Tabelle einzusetzen, um Erfahrungen mit dem Verfahren zu gewinnen.

*Mit dem R-Commander:* Normalerweise arbeitet man gleich aus der Datentabelle heraus, d. h. erzeugt aufgrund des vorhandenen Datensatzes mit seinen einzelnen Fällen und Merkmalen eine Vier- oder Mehrfeldertafel und bestellt bei der Software die Details, z. B. das Einspielen von Erwartungswerten, Chi-Werten und das Ausführen des Chi-Quadrattests. Manchmal liegen die Rohdaten jedoch nicht vor, man möchte vielmehr eine bereits publizierte Tabelle überprüfen, für die man die Rohdaten nicht vorliegen hat. Der R-Commander hat dazu ein wirklich gutes Werkzeug – leider etwas versteckt.

Der Weg: R-Commander starten, dann => Statistik; => Kontingenztabellen; => Analyse einer selbst eingegebenen Kreuztabelle. Es öffnet sich ein neues Menü. Unter dem Reiter „Tabelle“ kann man die Tabelle eingeben, die beiden Schieberegler erlauben das Anpassen der Zeilen- und Spaltenzahl. Unter dem Reiter „Statistik“ bestellt man, was außer den beobachteten Häufigkeiten in die Tabelle eingespielt werden soll. Nach „OK“ werden die Berechnungen ausgeführt und die Ergebnisse ausgegeben.

*Mit PAST:* PAST bietet ähnlich wie der R-Commander die Option, eine bereits fertige Tabelle zu berechnen. Nach dem Eingeben der Zahlen den aktiven Bereich (d. h. die Tabelle) markieren (Klick in die Zelle links oben; Shift-Taste drücken und halten; Klick in die Zelle rechts unten = die ganze Tabelle erscheint hellblau hinterlegt), dann: => Univariate; => Contingency table; [den ggf. aufpoppenden Protest, dass Fishers Exakter Test nicht berechnet werden könne, ignorieren]. Es klappt ein neues Menü auf, in dem unter dem Reiter „Test“ die entsprechenden Kennzahlen stehen und unter dem Reiter „Residuals“ einige der berechneten Zellenwerte abgelesen werden können, z. B. die „expected values“ (Erwartungswerte).

Drei unterschiedliche, kostenlose und einfach zu bedienende Werkzeuge für die Berechnung der Chi-Quadrat-Statistik: Es gibt keine Ausrede, diese Analyse nicht durchzuführen und statt dessen darüber zu sinnieren, ob man „mit so wenig Beobachtungen“ etwas machen kann.

### 18.4.6 Der p-Wert, oder: die Sache mit der Signifikanz

Beim Chi-Quadrat-Test hatten wir eine Prüfgröße berechnet (Chi-Quadrat) und gemäß der Randbedingungen (Freiheitsgrade) und einem vorab definierten Irrtumsrisiko (*alpha*) einen „kritischen Wert“ gesetzt; im Vergleich zwischen berechneter Prüfgröße und „kritischem Wert“ wird deutlich, ob die Abweichungen zwischen beobachteten und erwarteten Häufigkeiten noch mit dem Zufall vereinbar sind oder signifikant vom Zufälligen abweichen. Das Verfahren ermöglicht also eine klare Ja/Nein-Entscheidung.

Wenn Sie die Berechnungen wie zuvor beschrieben z. B. mit dem R-Commander oder mit PAST durchgeführt haben, werden Sie in den resultierenden Ausgaben zusätzlich auf den „p-Wert“ gestoßen sein, der ebenfalls protokolliert wurde. Der p-Wert besagt, mit welcher Wahrscheinlichkeit die beobachtete Prüfgröße es erlaubt, die Nullhypothese zu verwerfen, d. h. die Signifikanz einer Abweichung von der Nullhypothese erreicht ist. Der p-Wert nimmt Werte zwischen 1 und Null an, ab  $p < 0.05$ , oder kleiner 0.01, oder kleiner 0.001 ist die Signifikanzschwelle erreicht und die Nullhypothese wird verworfen. Ergänzend zum bislang verfolgten Weg einer klaren Ja/Nein-Entscheidung erlaubt der p-Wert eine metrisch skalierte Darstellung der Signifikanz – man kann sehen, wie weit ein p-Wert (in der einen oder anderen Richtung) von den gesetzten Signifikanzschranken entfernt ist.

In der Archäologie wird üblicherweise mit einem Schwellwert von  $p \leq 0.05$  operiert, d. h. p-Werte kleiner/gleich 0.05 gelten als signifikant. Bisweilen wird auch auf die Angabe eines exakten p-Wertes verzichtet und die Prüfgröße lediglich mit einem bis drei Sternen markiert. Dabei gilt:

p-Wert  $> 0.05$  (d.h. größer 0.05): „nicht signifikant“  
\* p-Wert 0.05 bis  $> 0.01$ : „signifikant“  
\*\* p-Wert 0.01 bis  $> 0.001$ : „sehr signifikant“  
\*\*\* p-Wert 0.001 bis 0.000: „hoch signifikant“

**Abb. 66** Die Begrifflichkeit zu „signifikant“.

Während Prüfgrößen (wie hier z. B. das beobachtete Chi-Quadrat) unterschiedlicher Test-Verfahren unterschiedliche, nicht direkt vergleichbare Werte annehmen und folglich unterschiedliche Schwellenwerte haben, kann der p-Wert in gleicher Weise für alle Testverfahren berechnet werden, er ist also als übergreifendes Maß für eine Signifikanz resp. Nicht-Signifikanz verwendbar.

#### 18.4.6.1 Kritik am p-Wert

Nicht verschwiegen sei, dass eine vielfach zu beobachtende, unkritische Überbewertung des p-Wertes sich in den 2010er Jahren in zunehmender Kritik befindet.<sup>52</sup> Diese Kritik betrifft jedoch nicht die Berechnung des p-Wertes, sondern dessen bisweilen leichtfertige oder auch falsche Interpretation. Dabei wendet sich die Kritik an den Interpretationen nicht gegen die Aussage, dass ein Befund bei  $p < 0.05$  signifikant sei! Sie wendet sich vielmehr gegen die Tendenz, alle Ergebnisse mit p-Werten größer 0.05 als wertlos zu verwerfen. Tatsächlich könnten in diesen verworfenen Studien auch wertvolle Ergebnisse stecken, die „noch nicht signifikant“ seien; in solchen Fällen könnte sich z. B. nach einer Vergrößerung der Stichprobe oder nach Anhebung der Messgenauigkeit durchaus ein signifikanter Effekt mit p-Wert kleiner 0.05 ergeben. Wer sich von dieser Debatte betroffen fühlt, möge statt der einschlägigen Zeitungsmeldungen, die solche Kritik gerne hochspielen, die wissenschaftliche Diskussion dazu studieren. Jahrgang 73(1), 2019 der Fachzeitschrift „The American Statistician“ vereint auf ca. 400 Seiten die Debatte der Fachwelt zum Thema. Deren Lektüre bringt zwar viel Erkenntnis, jedoch keine Lösung für die Frage, was an die Stelle des p-Wertes treten solle. Man soll halt klug und behutsam mit der Interpretation seiner Daten und Kennzahlen umgehen.

Ist das neu? Archäologen kennen Probleme dieser Art und wissen damit umzugehen, beispielsweise bei der ihnen grundlegenden Stratigrafie. Gewiss, eine ursprünglich am Hang sauber abgelagerte Schichtenfolge von A über B über C kann nach Mobilisierung der Oberfläche ins Tal erodieren und würde dort als C über B über A in einer Umkehrstratigrafie abgelagert werden. Solch eine Einzelbeobachtung hebt aber nicht das Prinzip aus, sondern fordert lediglich zum sorgfältigen Umgang mit dem Prinzip Stratigrafie auf und zur Berücksichtigung der Nebenumstände, oder zeitgeistig gesagt: des Kontexts. So ist es m. E. auch mit dem p-Wert.

#### 18.4.7 Chi-Quadrat-Verfahren: Schwach besetzte Tafeln (I)

Mit allzu kleinen Häufigkeiten sollte auch der Chi-Quadrat-Test nicht durchgeführt werden! Die Mindestforderung lautet, dass eine erwartete Häufigkeit von 5 in maximal zwanzig Prozent der Zellen unterschritten werden darf;<sup>53</sup> bei Tafeln mit  $2 \times 2$  Feldern – auch „Vierfeldertafeln“ genannt – besteht die Forderung, dass die kleinste erwartete

---

<sup>52</sup> Wasserstein & Lazar, 2016; Amrhein et al., 2019. Vgl. auch Ironisierungen des Themas, wie z.B. „The p-value song“ (14.7.2015): <https://www.youtube.com/watch?v=yy4nsEvKh2E> [24.1.2020].

<sup>53</sup> Auf die Frage „was tun, wenn doch?“ kommen wir für größere Tabellen noch zurück.

Häufigkeit in allen Zellen mindestens 5 betragen soll. In der Praxis wird diese Forderung meist reduziert auf die Forderung, dass die kleinste beobachtete Häufigkeit mindestens 5 betragen soll. Ist dies nicht der Fall, kann Chi-Quadrat gegenüber dem Wert der exakten Verteilung zu groß werden, d. h. man muss mit einer erhöhten Wahrscheinlichkeit rechnen,  $H_0$  fälschlicherweise zu verwerfen.

Was tun, wenn die kleinste beobachtete Häufigkeit unter 5 liegt? Die klassische Technik lautet „Korrektur nach Yates“. Leider wird das Verfahren in der Literatur unterschiedlich geschildert. Weber (1986, 194, 13.4) und Ihm (1978, 218, 8.6) gehen wie folgt vor: Bei unveränderten Randsummen werden die kleinsten Zellenhäufigkeiten um 0,5 vergrößert. Dies erfordert, dass die Nachbarzellen so verändert werden (minus 0,5), dass die Randsummen gleich bleiben. Mit den veränderten Häufigkeiten wird dann der Test durchgeführt. **Abb. 67** zeigt das Vorgehen an einer frei erfundenen Tabelle, links die Ausgangstabelle, rechts die Tabelle mit der Korrektur nach Yates.

Ausgangstabelle			korrigierte Tabelle		
4	6	10	4,5	5,5	10
7	8	15	6,5	8,5	15
11	14	25	11	14	25

**Abb. 67** Beispiel für eine „Korrektur nach Yates“ bei einer schwach besetzten Vierfeldertafel.

In PAST und **R** wird die Korrektur nach Yates nicht automatisch durchgeführt, ggf. müsste man die Eingabetabelle händisch anpassen.

#### 18.4.8 Exakter Test nach Fisher

Wenn (1.) nach dem einfachen Verfahren mit Yates-Korrektur Zweifel an der Entscheidung bestehen, und (2.) das Minimum der Erwartungshäufigkeiten kleiner 5 ist, und/oder der Stichprobenumfang insgesamt kleiner als 20 ist, wird bei Vierfeldertafeln ( $2 \times 2$  Zellen) der sog. „exakte Test nach Fisher“ bevorzugt (z. B. Ihm 1978, 220-222). Berechnungen und Prüfgröße sind anders als beim Chi-Quadrat-Test, das Test-Prinzip bleibt ähnlich. Zum „exakten Test nach Fisher-Yates“ siehe z. B. Bortz & Lienert (2003, 82 ff., Kap. 2.3.1., oder Kap. 32).

Der „exakte Test nach Fisher“ – gemeint ist der britische Statistiker Ronald Aylmer Fisher (1890-1962) – dient zur Beurteilung von  $2 \times 2$ -Felder-Tafeln, bei denen

der Stichprobenumfang klein ist, d. h. unter 20 liegt, oder bei denen das Minimum der Erwartungshäufigkeiten kleiner als 5 ist.

Seine Nullhypothese  $H_0$  lautet: Die Zelle mit der scheinbar auffallend geringsten beobachteten Häufigkeit liegt noch im Bereich zufällig möglicher Abweichungen bei zwei Stichproben aus einer gemeinsamen Grundgesamtheit. Die Alternativhypothese  $H_1$  ist: Die beiden Stichproben stammen nicht aus einer gemeinsamen Grundgesamtheit mit gleichen Wahrscheinlichkeiten. Das Vorgehen verdeutlichen wir hier anhand der frei erfundenen Tabelle **Abb. 68**.

(1) Wir haben Keramik nach Grob- und Feinkeramik klassifiziert und müssen entscheiden, ob sich zwei Gruben A und B in ihrer Zusammensetzung signifikant unterscheiden. Dazu verfertigen wir folgende Ausgangstabelle:

Ausgangstabelle	Grobkeramik	Feinkeramik	Summe
Grube A	4	2	6
Grube B	2	7	9
Summe	6	9	15

**Abb. 68** Ausgangstabelle für ein Beispiel des „exakten Tests nach Fisher“.

(2) Obwohl technisch nicht wirklich nötig, ist es beim exakten Test nach Fisher üblich, die Tabelle so zu organisieren, dass jene Zelle mit der geringsten relativen Häufigkeit links oben steht. Also prüfen wir dies und stellen die Tabelle ggf. um (**Abb. 69**):

Ausgangstabelle	Grobkeramik	Feinkeramik	Summe
Grube A	4 von 6 = 0,66	2 von 9 = 0,22	6
Grube B	2 von 6 = 0,33	7 von 9 = 0,78	9
Summe	6	9	15

**Abb. 69** Beispieltabelle für den „exakten Test nach Fisher“, nun mit den Erwartungswerten (Daten wie **Abb. 68**).

(3) Die Zelle Feinkeramik in Grube A ist jene mit der geringsten relativen Häufigkeit, also stellen wir die Spalten um. Hinzugefügt ist hier in der Tabelle (**Abb. 70**) zum besseren Verständnis die übliche Notation (**blaue Schrift**) zur Benennung der Zellen:

( neu sortiert )	Feinkeramik	Grobkeramik	Zeilensummen
Grube A	2 a	4 b	6 a + b
Grube B	7 c	2 d	9 c + d
Spaltensummen	9 a + c	6 b + d	15 a + b + c + d = N

**Abb. 70** Umgeordnete Beispieltabelle für den „exakten Test nach Fisher“, mit den bei diesem Verfahren üblichen Zellenbezeichnungen (Daten wie **Abb. 68**).

Die Hypothese  $H_0$  lautet: Die geringe relative Häufigkeit in der Zelle (a) Feinkeramik/Grube A mit 0,33 ist noch zufällig möglich. Die nun notwendigen konkreten Berechnungen sind – auch mit einer Tabellenkalkulation – etwas aufwendig, weshalb der exakte Test nach Fisher ehemals auch eher selten berechnet wurde; vielfach benutzte man Tabellenwerke. In die modernen Programme ist das Verfahren jedoch fest eingebettet, sodass diese Zurückhaltung heute nicht mehr geboten ist. Berechnet wird die Prüfgröße „gamma“ („Überschreitungswahrscheinlichkeit“). Ist diese Prüfgröße größer als die üblicherweise gesetzte Schranke vom 0.01, kann  $H_0$  beibehalten werden. In unserem Fall ergibt sich ein „gamma“ von 0.119, d. h. größer  $\alpha = 0.01$ :  $H_0$  kann beibehalten werden.

*Praktische Berechnung mit PAST:* Wie oben für den Chi-Quadrat-Test schon beschrieben: Tabelle eingeben, den aktiven Bereich markieren, dann => Univariate; => Contingency table. In der neu aufklappenden Fläche wird in der Mitte unter „Fischer’s exact“ nicht die Prüfgröße selbst ausgespielt, aber der p-Wert, inwieweit die Abweichungen nicht-zufällig sind. Im vorliegenden Fall ist  $p = 0,13566$ , d. h. größer als eine Signifikanz von 0.05 oder 0.01: Die Abweichungen sind nicht signifikant.

*Praktische Berechnung mit dem R-Commander:* Wie oben für den Chi-Quadrat-Test bereits beschrieben: => Statistik; => Kontingenztabelle; => Analyse einer selbst eingegebenen Kreuztabelle. Daten eingeben. Dann beim Reiter „Statistik“ zusätzlich zu den Voreinstellungen auch ein Häkchen an „Fishers exakter Test“ setzen (unten), OK. Ausgegeben wird auch hier nicht die Prüfgröße selbst, sondern der p-Wert. Im vorliegenden Beispiel beträgt er 0.1357, d. h. ist deutlich größer als die Signifikanzschranke von 0.01, also Befund nicht signifikant.

(4) Ergebnis: Die geringe beobachtete Häufigkeit von Feinkeramik in der Grube A ist noch mit der Annahme von zufälligen Abweichungen innerhalb einer gemeinsamen Grundgesamtheit verträglich.

*Fisher-Erwin-Test.* Für Vierfeldertafeln, in denen eine Zelle einen Erwartungswert unter 5, aber über 1 hat, wird in der theoretischen Literatur die „N-1 Korrektur nach Egon Pearson“ empfohlen, jedoch m.W. in der Praxis selten angewendet. Die Berechnung von  $\chi^2$  erfolgt zunächst nach dem üblichen  $\chi^2$ -Verfahren; abschließend wird das berechnete  $\chi^2$  mit dem Ergebnis von  $(N-1)/N$  multipliziert, und dann wiederum mit den üblichen kritischen Werten verglichen.<sup>54</sup> Angewendet auf unsere Tabelle **Abb. 68** ergibt sich ein korrigiertes  $\chi^2$  von 2.77, d. h. wie auch nach dem exakten Test nach Fisher der Befund „nicht signifikant“ ( $p=0.10$ ).

#### 18.4.9 Chi-Quadrat-Verfahren: Schwach besetzte Tafeln (II)

In Mehrfeldertafeln – also Tabellen mit mehr als  $2 \times 2$  Zellen – gilt die Forderung, dass maximal 20 % der Zellen erwartete (resp. beobachtete) Häufigkeiten unter 5 haben dürfen. Was tun, wenn mehr als 20 % der Zellen eine zu geringe Häufigkeit aufweisen? Wirklich gänzlich befriedigende Lösungen gibt es kaum. Der m.E. beste Weg ist, die Tabelle und die Fragestellung zu überdenken und demgemäß mehrere Zellen geeignet zusammen zu fassen, also etwa fragestellungsorientiert zwei Spalten zu einer Spalte zusammen zu ziehen, oder dies für Fälle (Zeilen) zu erwägen.

Die moderne Literatur (z. B. Bortz, Lienert & Boehnke, 2000) operiert nicht mehr mit der „Korrektur nach Yates“. Man erwägt stattdessen je nach Lage der Dinge folgende Verfahren (**Abb. 71**):

---

<sup>54</sup> Campbell I. (2007). Chi-squared and Fisher-Irwin tests of two-by-two tables with small sample recommendations. *Statistics in Medicine*, 26, 3661 – 3675. Die Berechnung kann z. B. mit Hilfe eines Online-Tools erfolgen: „Chi squared test calculator for two-by-two tables“: <http://www.iancampbell.co.uk/two-by-two/calculator.htm> [24.1.2020]. Für Berechnungen mit R steht ein Skript zur Verfügung: „EPCS test“: [https://rpubs.com/serious-tats/epcs\\_test](https://rpubs.com/serious-tats/epcs_test) [24.1.2020].

wenn:	dann:
die Tafel klein (2 x 3 oder 3 x 3 Felder) ist, alle Felder schwach besetzt sind,	exakter Kontingenztest (Bortz, Lienert & Boehnke 2000, 140-142);
die schwach besetzte Tafel höchstens 5 x 5 Felder hat, alle Erwartungswerte größer 1 sind,	empirisch approximierter Chi-Quadrat-Test von Craddock & Flood 1970 (s. Bortz, Lienert & Boehnke 2000, 139-140);
die schwach besetzte Tafel deutlich mehr als 5 x 5 Felder hat, es Erwartungswerte kleiner als 1 gibt,	modifizierter Chi-Quadrat-Test: sog. Haldane-Dawsen-Test (s. Bortz, Lienert & Boehnke 2003, 137-139).

**Abb. 71** Alternativen zum einfachen Chi-Quadrat-Test bei schwach besetzten Mehrfeldertafeln.

Bewertung: Die hier in **Abb. 71** genannten Tests können in speziellen Fällen sinnvoll herangezogen werden. Alternativ sollte jedoch unbedingt geprüft werden, ob die ggf. aus dem Testergebnis abgeleiteten Aussagen in der Praxis von archäologischer Relevanz wären. Dies ist oft nicht der Fall, selbst wenn die Tests dann „richtige“ Ergebnisse erbrächten. Daher empfehle ich, statt über die Anwendung eher wenig bekannter statistischer Verfahren vor allem über das Zusammenfassen von Zellen nachzudenken, d. h. die Fragestellung der Datenlage entsprechend anzupassen. Eine stets gute Alternative ist es, sich nach einer Vergrößerung der Stichprobe umzuschauen, beispielsweise indem man zusätzlich zu den „eigenen“ Daten solche aus der Literatur hinzugewinnt.

„Zellen zusammenfassen?“ Hätten wir in unserem Münzbeispiel nur eine kleine Stichprobe, könnte man etwa erwägen, die Nominale 1, 2 und 5 Cent zu „Kupfermünzen“ zusammenzufassen, die Nominale 10, 20 und 50 Cent zu „Messingmünzen“ und die Nominale 100, 200 und 500 Cent zu „Bimetallmünzen“ (o.ä.). Dann lägen nur mehr 3 statt 9 Münzkategorien vor, dem entsprechend wären die Häufigkeiten in den verbliebenen Zellen deutlich höher und es könnte vermutlich mit dem bewährten gewöhnlichen Chi-Quadrat-Verfahren gearbeitet werden.

#### 18.4.10 Testvoraussetzung Binomial- oder Poisson-Verteilung

Die Anwendung des Chi-Quadrat-Tests ist nur erlaubt, wenn die Häufigkeiten einer Poissonverteilung, einer Binomialverteilung oder einer Multinomialverteilung unterliegen. Keine Sorge, in unseren Beispielen ist dies der Fall. Aber das Ganze ist ein komplexes und wichtiges Thema, das wir im Kapitel 18.5 „Chi-Quadrat-Krise, ...“ ausführlich behandeln werden.

### 18.4.11 Informationsstatistik

Ein anderes Verfahren, die sog. Informationsstatistik (I) nach Solomon Kullback, bietet statistisch einige Vorteile gegenüber dem Chi-Quadrat-Verfahren (u. a. höhere Genauigkeit der ermittelten Prüfgröße). Allerdings bestehen die verteilungstheoretischen Forderungen des Chi-Quadrat-Verfahrens fort und ebenso das Problem eventuell zu gering besetzter Zellen. Daher hat sich dieses bei Ihm (1978, 211 ff.) propagierte Verfahren aus der Gruppe der „Likelihood-Verhältnis-Tests“ in der Praxis nicht breit durchgesetzt.<sup>55</sup>

### 18.4.12 Chi-Quadrat signifikant: und nun? – Konfigurationsfrequenzanalyse (KFA)

Der Befund „signifikant“ bedeutet: Es gibt signifikante Unterschiede, d. h. die Abweichungen zwischen den beobachteten und erwarteten Häufigkeiten sind nicht mehr zufällig. In der Regel stellt sich bei Mehrfeldertafeln anschließend die Frage, wo diese Unterschiede liegen. Hilfreich ist die systematische Durchsicht aller Zellen nach signifikanten Abweichungen: Man beobachtet in der Mehrfeldertafel, in welchen Zellen das errechnete (Zellen-) „Chi“ größer ist als der „kritische Wert“ zu  $\alpha = 0.05$  (3.84)<sup>56</sup> oder – selten so streng gesetzt – zu  $\alpha 0.01$  (6.63) bei einem Freiheitsgrad. Diese anscheinend signifikant überrepräsentierten Beobachtungen nennt man „Typ“, und die anscheinend signifikant unterrepräsentierten Zellen „Antityp“.

Meist gibt es in größeren Tabellen Zellen (oder Zeilen oder Spalten) ohne auffallende Effekte, während wiederum andere Zellen deutliche Abweichungen aufweisen und wesentlich zu dem Gesamt-Chi-Quadrat beitragen. In diesen Fällen ist ein theoretisch nicht ganz sauberer, aber effizient zielführender Weg, nach der Erstanalyse bei einer größeren Tabelle ihre Zerlegung in zwei neue Tabellen vorzunehmen, (a) diejenige, die alle nicht signifikanten Zeilen und Spalten aufnimmt, und (b) jene, welche die signifikant unterschiedlichen Zeilen und Spalten aufnimmt. Daraus resultiert eine exaktere Beschreibung der unspezifischen Unterschiede einerseits und der interpretierenswürdigen Unterschiede andererseits.

Eine Alternative dazu und der theoretisch korrektere Weg ist die Einzelprüfung der „Typen“ auf Signifikanz; hierzu erstellt man eine Vierfeldertafel mit den drei Zeilen Stichprobe, Spaltensumme minus Stichprobe, Spaltensumme, sowie den drei

---

<sup>55</sup> Zur Informationsstatistik z. B. Ihm (1978), 211 ff.; vgl. unter der Bezeichnung „G-Test (Anpassungstest)“ auch bei Weber (1986, 194-196).

<sup>56</sup> Siehe oben die Tabelle **Abb. 65** mit den kritischen Werten der Chi-Quadrat-Verteilung, der Wert 3,84 ist der kritische Wert zum Freiheitsgrad 1.

Spalten „Typ“, Zeilensumme minus Typ, Zeilensumme, und testet mit dem „exakten Test nach Fisher“. Zumindest im Falle von Tabellen mit geringen Häufigkeiten ist dieser Weg anzuraten.

Wem diese Hinweise nicht reichen, z. B. weil eine solche Tabelle eine ausnehmend hohe Bedeutung für die anliegende Studie hat, findet unter einem Verfahren namens „Konfigurationsfrequenzanalyse“ (KFA) sicheren Boden. Wobei es sich dabei nicht um spezielle Berechnungen handelt, sondern um einen klaren, statistisch ausgearbeiteten Weg zur vertieften Analyse der Tabellen (Bortz, Lienert & Boehnke 2008, 155-158 mit weiterer Lit.).

#### **18.4.13 Chi-Quadrat-Test und große Grundgesamtheiten**

Wir haben in den vorangehenden Abschnitten einen in den Lehrbüchern meist stark betonten Aspekt eingehend beleuchtet, der auch in der Archäologie eine wichtige Rolle spielt: das Arbeiten mit Häufigkeiten im Falle von wenigen Beobachtungen. Wenn man die Wahl hat, ist in solchen Fällen der bessere Weg, zu prüfen, ob man nicht anstelle der beschriebenen Verfahren durch eine gezielte Vergrößerung der Stichprobe eine umfangreichere Datenbasis gewinnen kann und damit in den Bereich des Standardverfahrens kommt – aber dies ist eben nicht immer möglich.

Einer besonderen Erwähnung wert ist aber auch das Gegenteil: das Vorhandensein einer großen Stichprobe mit zahlreichen Beobachtungen. Unser Geldbörsen-Beispiel mit insgesamt mehr als 7.700 Münzen etwa bewegt sich in diesem Bereich. In solchen Fällen neigt das Chi-Quadrat-Verfahren dazu, recht bald eine statistische Signifikanz anzuzeigen. Da das Verfahren klar und eindeutig ist, sollte man auch hier nicht an dem Befund „signifikant“ rütteln. Stets zu stellen ist jedoch zusätzlich die Frage, ob ein signifikanter Unterschied auch relevant ist. Eine Antwort darauf kann nicht die Statistik geben, sie muss aus der Archäologie heraus entwickelt werden. So kommt es in solchen Fällen nicht selten zum Befund, dass ein Unterschied zwar signifikant ist, jedoch in der Sache so gering ausfällt, dass er für die Lebenswirklichkeit vergangener Menschen keine Rolle spielte. Diese Frage nach der Relevanz sollte am Ende der statistischen Prozeduren unbedingt gestellt und eigens beantwortet werden.

Das beschriebene Phänomen, nämlich dass bei großen Fallzahlen auch relativ kleine Unterschiede eine statistische Signifikanz erreichen können, ist kein Spezifikum des Chi-Quadrat-Verfahrens. Es betrifft beispielsweise auch die Analyse von Messwerten und Mittelwertunterschieden, die z. B. mit dem T-Test als signifikant erwiesen wurden. Die geisteswissenschaftlich zu klärende Relevanzfrage stellt sich also auch jenseits des Chi-Quadratverfahrens.

#### 18.4.14 Praxistipp: Tests

Es kann vorkommen, dass man sich trotz sorgsamem Einarbeiten in ein Testverfahren unsicher ist, wie die Ergebnisse zu interpretieren sind. Schnelle Selbsthilfe: Man arbeitet parallel mit künstlichen („simulierten“) Zahlen. Im Falle des Chi-Quadrat-Test: mit einer Tabelle mit selbst erfundenen Zahlen, bei denen es mit Sicherheit keine Effekte gibt, und zum Vergleich mit einer solchen, in denen es mit Sicherheit starke Effekte gibt. So lernt man, Testergebnisse sicherer zu interpretieren. Eine Empfehlung, die auch für andere Testverfahren gilt.

### 18.5 CHI-QUADRAT-KRISE, MINDESTINDIVIDUENZAHLE, UND WARUM DIE PRÜFUNG DER VERTEILUNG SO WICHTIG IST

Der folgende Abschnitt adaptiert einen lehrreichen Abschnitt bei Ihm (1978) und zielt darauf, spielerisch einen möglichen Anwendungsfehler des Chi-Quadrat-Verfahrens zu erfahren. Am Ende wird deutlich geworden sein, weshalb in der Archäologie, der Anthropologie und auch der Archäozoologie das Thema Mindestindividuenzahlen (MIZ) so wichtig ist, und statistisch abstrakt sieht man exemplarisch, was geschieht, wenn die theoretischen Voraussetzungen für ein Testverfahren nicht beachtet werden.

Wir kommen auf den  $\chi^2$ -Test zurück. Wir vergleichen zwei Gruben (Grube 1 und Grube 2) mit zwei unterschiedlichen Keramiktypen (Typ A, Typ B), die in etwas unterschiedlicher Häufigkeit in beiden Gruben vorkommen (**Abb. 72**).

	Typ A	Typ B	Spaltensumme
Grube 1	6	6	12
Grube 2	5	8	13
Zeilensumme	11	14	25

**Abb. 72** Hypothetische Vierfeldertafel mit zwei archäologischen Inventaren „Grube 1“ und „Grube 2“ und der Häufigkeit von zwei unterschiedlichen Typen, z. B. von Keramikgefäßen, in ihnen.

Die Typen gelten als leicht unterschiedlich alt, wir verfolgen die Frage, ob die beiden Gruben eine signifikant unterschiedliche Typen-Zusammensetzung aufweisen. Dazu berechnen wir nach dem üblichen Verfahren  $\chi^2$ , und gewinnen ein beobachtetes  $\chi^2$  von 0.34. Der Vergleich mit dem kritischen Wert – hier 3.84 – zeigt, dass die Abweichungen zwischen beiden Gruben mit dem Zufall vereinbar sind, also nicht signifikant

unterschiedlich. Es ist nützlich, das hier verfolgte Beispiel mit PAST, dem R-Com-mander o. ä. jeweils selbst nachzurechnen.

Ein anderer, in Fragen der Statistik wenig versierter Bearbeiter dieses Fundmaterials hat für den gleichen Fall eine andere Darstellung gewählt und statt der Häufigkeiten deren Prozentwerte in die Tabelle eingesetzt. Danach sieht seine Tabelle so aus (**Abb. 73**).

	Typ A	Typ B	Spaltensumme
Grube 1	24 %	25 %	48 %
Grube 2	20 %	32 %	52 %
Zeilensumme	44 %	56 %	100 %

**Abb. 73** Die Vierfeldertafel wie **Abb. 72**, jedoch statt der gezählten Häufigkeiten mit den entsprechenden Prozentwerten.

Wie zuvor wird rasch noch  $\chi^2$  berechnet, wir gewinnen den Wert 1.35. Er liegt deutlich unter dem kritischen Schwellenwert von 3.84, also sind die Unterschiede zwischen den Gruben nicht signifikant.

Aber halt: Das beobachtete  $\chi^2$  ist größer geworden! Ursprünglich in **Abb. 72** bei den gezählten Häufigkeiten betrug es 0.34, nun (**Abb. 73**) bei den Prozentwerten 1.35. Was ist da los? Folgen wir doch dem Vorschlag eines Scherzbolds und drücken das Ganze in Promille statt Prozent aus! (**Abb. 74**).

	Typ A	Typ B	Spaltensumme
Grube 1	240 ‰	250 ‰	480 ‰
Grube 2	200 ‰	320 ‰	520 ‰
Zeilensumme	440 ‰	560 ‰	1000 ‰

**Abb. 74** Die Vierfeldertafel wie **Abb. 72**, jedoch statt der gezählten Häufigkeiten mit den entsprechenden Promillewerten.

Wir berechnen rasch  $\chi^2$  – und erhalten ein beobachtetes  $\chi^2$  von 13.49 (!), weit größer als der kritische Wert von 3.84, sogar größer als der kritische Wert von 10,83 für das Irrtumsrisiko  $\alpha$  (alpha) = 0.001. Das „Ergebnis“ bei der Formulierung des Problems in Promille ist: Beide Gruben unterscheiden sich hochsignifikant! Sind wir glücklich (denn ein signifikanter Unterschied ist ja ein endlich mal ein Ergebnis) – oder haben

wir haben ein Problem? Anscheinend kann ich die These „beide Gruben sind nicht unterschiedlich“ statistisch ebenso gut belegen ebenso wie die These „beide Gruben unterscheiden sich hochsignifikant voneinander“. Liest man die einschlägige Literatur gründlicher, stolpert man über Sätze wie bei Ihm (1978, 223): „Bei Kontingenztafeln setzt man Poisson- oder Multinomialverteilung voraus. Ist diese Voraussetzung nicht erfüllt, kann das  $\chi^2$ -Verfahren zu fehlerhaften Ergebnissen führen“. Später heißt es (Ihm 1978, 223 ff.) „... wurde der Test falsch angewendet, da keine Poissonverteilung vorliegt“ (ebd. p. 234). Offenbar haben statistische Verfahren Ansprüche nicht nur an die Menge, sondern auch an die Qualität von Daten. Die Verfahren und insbesondere die Tests dürfen nur angewendet werden, wenn die Daten diese Voraussetzungen erfüllen. Offenbar spielen „theoretische Verteilungen“ (wie Poissonverteilung, Multinomialverteilung etc.) eine wichtige Rolle. Kurz: Wir müssen uns um theoretische Verteilungen kümmern, bevor wir auf dem Gebiet der Testverfahren weiterkommen.

Vorab zu unserem  $\chi^2$ -Test: Ausschließlich der Test mit den Rohdaten (**Abb. 72**) ist richtig durchgeführt, denn nur diese unterliegen tatsächlich einer sog. Poissonverteilung. Das Umsetzen in Prozentwerte und Promillewerte ist falsch, es führt – wie Ihm (1978, 236) schön aufzeigt – zu einer sog. „Anstückungsverteilung“ (auch: „Ansteckungsverteilung“, engl. *contagious distribution*): Wo eines ist, da liegen auch mehr. Solche Phänomene folgen oft einer sog. „negativen Binomialverteilung“ (dazu später), auf die das  $\chi^2$ -Verfahren nicht angewendet werden darf. Und wir beginnen zu verstehen, weshalb statistikbewusste Archäologen vor der Auswertung von Grubeninventaren mit Scherben die einzelnen Scherben zu „Gefäßeinheiten“ zusammensetzen (müssen): Die „Gefäßeinheiten“ unterliegen (vermutlich) einer Poissonverteilung – die einzelnen Scherben nicht! Wird nämlich ein bestimmtes Gefäß („Gefäßeinheit“) in den Abfall geworfen, ist – wegen der Wahrscheinlichkeit des Zerbrechens von Gefäßen – die Wahrscheinlichkeit deutlich erhöht, dass weitere Scherben eben dieses Gefäßes in der Grube liegen. Weitere Beispiele: Die Anzahl der Kastanien pro Fläche im Park folgt einer Ansteckungsverteilung. Denn wo ein Kastanienbaum steht, liegen meist viele Kastanien. Die Pflanzenausbreitung in der Fläche bei Pioniervegetationen folgt einem ähnlichen Muster: Wo eine erste Pflanze dieser Art steht, ist die Wahrscheinlichkeit erhöht, dass (bald) auch weitere Pflanzen dieser Art in der Nachbarschaft stehen. Ähnlich steht es um Abfallspektren z. B. mit Tierknochen: Enthält eine Abfallgrube einen Rinderknochen, der anzeigt, dass hier ein Rind geschlachtet wurde, ist die Wahrscheinlichkeit erhöht, dass man weitere Rinderknochen findet, denn ein Rind hat

viele Knochen. Daher wäre die Anwendung des  $\chi^2$ -Verfahrens auf die gezählten Häufigkeiten der einzelnen Knochen falsch, statistisch richtig ist einzig der  $\chi^2$ -Test an Mindestindividuenzahlen (MIZ).<sup>57</sup>

Nicht verzweifeln: Das Thema „theoretische Verteilungen“ ist ebenso wichtig wie kompliziert, weshalb es in vielen Lehrbüchern schlicht umgangen wird; eine lehrreiche Ausnahme mit zahlreichen Beispielen auch zum Nachrechnen ist Ihm u. a. 1978. Wir wollen uns dem Thema Stück für Stück nähern, in der Reihenfolge seiner Wichtigkeit. Weshalb wir uns im folgenden Kapitel um die „Normalverteilung“ kümmern.

## 19 ANALYSE VON MESSWERTEN: MITTELWERT, STANDARDABWEICHUNG ...

---

In diesem Kapitel verlassen wir das Thema Häufigkeiten und wenden uns Messwerten zu, d. h. Daten, die mindestens Ordinalskalenniveau haben (Ränge, wie z. B. Noten), besser einer quantitativen Skala folgen (wie z. B. die Prägejahre in unserem Münzbeispiel). Die nächsten Abschnitte zielen darauf, solche Daten zusammenfassend zu beschreiben und zu analysieren, wozu zunächst einmal Kennzahlen gehören wie z. B. Zentral- und Streuungsmaße. Doch zuvor sollte man die Daten gesichtet haben, d. h. sich einen ersten visuellen Eindruck von ihnen verschafft haben. Also zurück zur Aufgabe „schnell mal eine Grafik“, denn es ist vorteilhaft, sich die Daten z. B. in einem Histogramm angeschaut zu haben.

### 19.1.1 Begriffsklärung unimodal, bimodal, multimodal

Zentral- und Streuungsmaße dienen dazu, Wertereihen abstrakt zu beschreiben und das wesentliche in ein oder zwei Zahlenwerten zusammenzufassen, z. B. um diese mit anderen Wertereihen zu vergleichen. Ihre Angabe ist jedoch nur dann sinnvoll, wenn die Datengruppe grundsätzlich auch durch ein Zentralmaß beschreibbar ist, d. h. es tatsächlich sinnvoll ist, einen „mittleren Wert“ anzugeben. Dies ist z. B. nicht der Fall

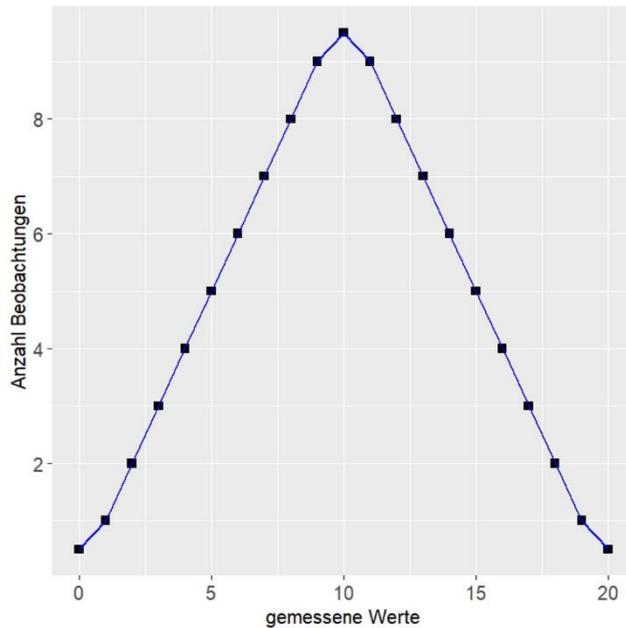
---

<sup>57</sup> Zum Thema „Knochenzahl, Knochengewicht, MIZ“: Friederike Johansson, Zusammensetzung des Knochenmaterials. In Becker, C. & Johansson, F. (1981). *Tierknochenfunde 2. Bericht: Mittleres und oberes Schichtpaket (MS und OS) der Cortaillod-Kultur*. (Die neolithischen Ufersiedlungen von Twann, 11) (p. 16-24, insbes. 17 Tab. 3). Bern: Staatlicher Lehrmittelverlag.

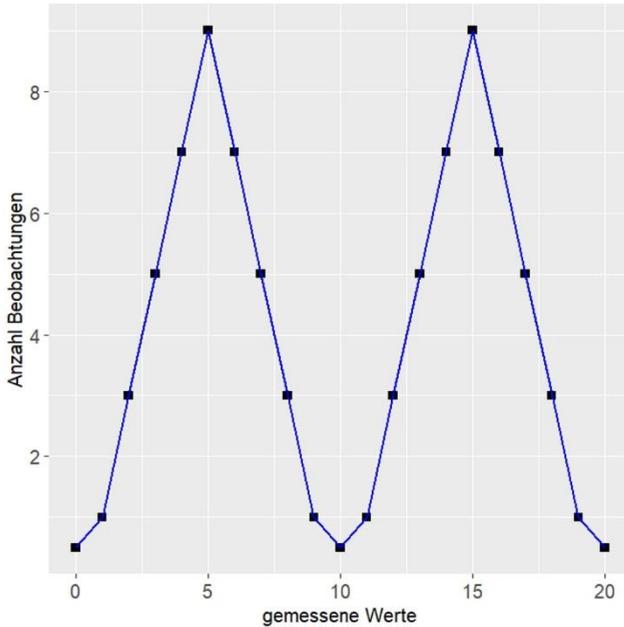
bei nominalskalierten Variablen. Wie sollte ein „mittleres Studienfach“ aussehen? Dies ist ebenfalls nicht der Fall bei Wertereihen, die erkennbar bi- oder multimodal sind. Was heißt das? Statt vieler Worte erklärt die Bildfolge **Abb. 75** bis **Abb. 77** mit den jeweiligen Legenden, worum es hier geht.

### *Empfehlung*

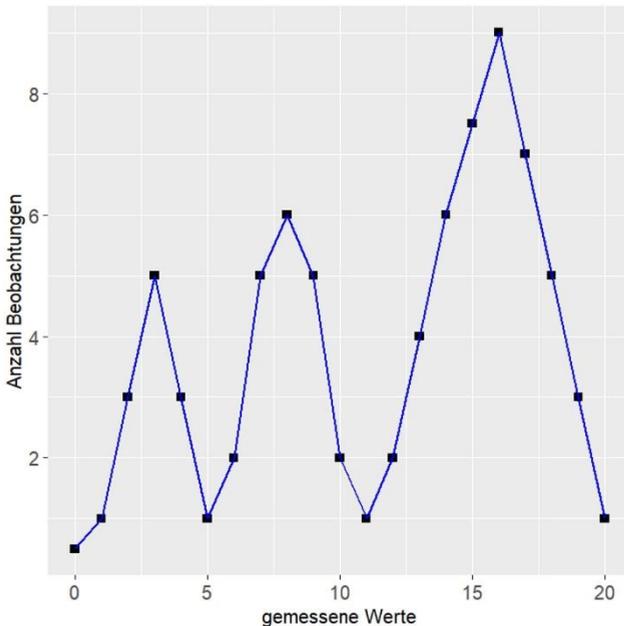
Weil die Angabe eines Zentralmaßes nicht immer sinnvoll ist, sollte man die Sinnhaftigkeit immer als erstes überprüfen. Der einfachste Weg, dies zu tun, ist eine grafische Darstellung der Daten als Histogramm. Diese Histogramme muss man später nicht zwingend auch publizieren. Im Gegenteil: In den Publikationen sollte man mit Grafiken sparsam umgehen. Aber im Prozess der Datenanalyse ist es stets sinnvoll, sich frühzeitig mit solchen Bildern einen visuellen Eindruck von den Daten zu verschaffen.



**Abb. 75** Unimodale Verteilung. Das Bild zeigt eine eingipflige / unimodale Verteilung. Sie kann durch ein Zentralmaß sinnvoll beschrieben werden, der mittlere Wert liegt bei 10.



**Abb. 76** Bimodale Verteilung. Das Bild zeigt eine zweigipflige / bimodale Verteilung. Sie hat – wie die Verteilung im Bild oben (**Abb. 75**) – rein rechnerisch einen mittleren Wert von etwa 10, was in diesem Fall jedoch sicherlich kein sinnvolles Zentralmaß ist. Richtiger wäre es – wenn man ein Zentralmaß benötigt – diese Wertereihe in zwei Gruppen zu teilen und jede für sich mit einem Zentralmaß zu beschreiben (hier ca. 5 und ca. 15).



**Abb. 77** Multimodale Verteilung. Dieses Bild zeigt eine mehrgipflige / multimodale Verteilung. Für solche Verteilungen ist die Angabe eines Zentralmaßes nicht sinnvoll.

## 19.2 ZENTRALMAßE: MODALWERT, MEDIAN, MITTELWERT

Sind Wertereihen durch ein Zentralmaß sinnvoll beschreibbar, stehen dazu je nach den Details unterschiedliche Maße zur Verfügung (s. Übersichtstabelle **Abb. 78**). Welches Zentralmaß sinnvoll verwendet wird, hängt zunächst einmal vom Skalenniveau ab.

- Bei nominalskalierten Variablen kann man die Häufigkeiten auszählen, und den häufigsten Wert bestimmen; man nennt ihn den Modalwert (engl. *mode*).
- Bei ordinalskalierten Variablen dient der Median als Zentralmaß: Er gibt den mittleren Wert einer geordneten Reihe an (mittlerer Rang). Der Median (auch: Zentralwert; engl. *median*) teilt eine geordnete Wertereihe in der Mitte. Bei ungerader Zahl von Beobachtungen ( $n$ ) ist der Median der Beobachtungswert des mittleren Ranges. Bei einer geraden Zahl von Beobachtungen ( $n$ ) ist der Median das arithmetische Mittel der beiden mittleren Beobachtungen. Bei mehrfach besetzten Klassen einer Ordinalskala ist der Median diejenige Klasse, in die der Beobachtungswert mit dem Rangplatz  $[(n+1) / 2]$  bei ungeradem  $n$ , bzw.  $[n / 2]$  bei geradem  $n$  fällt. „ $n$ “ ist die in der Statistik übliche Notation für Anzahl der Fälle.
- Bei quantitativ skalierten Variablen kann der Mittelwert als Zentralmaß dienen. Bisweilen ist es richtiger, stattdessen den Median zu wählen (dazu später). Der Mittelwert (auch: arithmetisches Mittel; engl. *mean*) ist die Summe der Werte, geteilt durch deren Anzahl.

### 19.2.1 Zentralmaße selbst gerechnet

Weil das Thema wichtig ist und weil das händische Rechnen die Sachverhalte einprägsam verdeutlicht, wollen wir eingangs an konkreten Beispielen die drei Zentralmaße selbst ermitteln. Klar, später nimmt Ihnen die Statistiksoftware Ihres Vertrauens diese Arbeit ab, aber lassen Sie uns den Weg zunächst als Fußgänger beschreiten.

- Modalwert. Die betreffende Datenspalte sortieren, den häufigsten Wert ermitteln. Dies ist der Modalwert.

Beispiel: Die Teilnehmer eines Kurses machten folgende Gewichtsangaben (geordnete Liste): 52; 53; 53; 55; 56; 64; 66; 70; 75; 75; 77; 96 kg. Die häufigsten Werte sind "53" und "75", demnach sind die Modalwerte 53 und 75, die Zahlenreihe ist „bimodal“.

- Median. Berechnung z. B. mit Hilfe einer Tabellenkalkulation. Schritt (1): Die betreffende Wertereihe (Spalte) wird sortiert. Schritt (2): Die Anzahl der Werte wird bestimmt. Ist die Anzahl der Werte gerade, ergibt sich eine minimal andere Berechnung

als für eine ungerade Anzahl von Werten. Schritt (3), ungerade Anzahl von Fällen: Der Median ist jener Wert, der sich in der mittleren Zelle befinden. Also: Anzahl der Fälle plus 1, dividiert durch 2 = Nr. der Zelle, die den Median enthält. Alternativer Schritt (3), bei gerader Anzahl von Fällen: Der Median ist jener Wert, der sich aus dem arithmetischen Mittel der beiden mittleren Zellen ergibt. Also: Anzahl der Fälle dividiert durch 2, Wert aus der Zelle unmittelbar über und unter dieser Zahl addieren und durch 2 dividieren.

Beispiel: Die Teilnehmer eines Kurses machten folgende Gewichtsangaben (geordnete Liste): 52; 53; 53; 55; 56; 64; 66; 70; 75; 75; 77; 96 kg. Es sind zwölf Angaben, d. h. der Median entspricht dem Mittel aus dem 6. und dem 7. Wert:  $64 + 66 / 2 = 65$  kg. Der Median beträgt 65 kg.

- Mittelwert (arithmetisches Mittel). Man bestimme die Anzahl der Werte; man berechne die Summe der Werte. Summe dividiert durch Anzahl gleich Mittelwert.

Beispiel: Die Teilnehmer eines Kurses machten folgende Gewichtsangaben (geordnete Liste): 52; 53; 53; 55; 56; 64; 66; 70; 75; 75; 77; 96 kg. Die Summe der Gewichte beträgt 792 kg; es sind zwölf Angaben. Das arithmetische Mittel beträgt  $792 / 12 = 66$  kg.

Skala	Zentralmaß	Streuungsmaß	Beispiele von Daten
Verhältnis~	Mittelwert	Standardabweichung	Körpergröße, Gewicht, Bargeld
Intervall~			Grad Celsius, schwimmende Dendrokurve
Ordinal~	Median	(Inter-) Quartilabstand	Medaillenränge, Noten
Nominal~	Modalwert	Werteliste	Geschlecht, Studienfach

**Abb. 78** Übersichtstabelle über die Art von Daten und die zugehörigen Zentral- und Streuungsmaße.

### 19.3 STREUUNGSMAßE: INTERQUARTILABSTAND, STANDARDABWEICHUNG

Zum Zentralmaß fehlt noch eine wichtige Information: Wie „typisch“ ist dieses Zentralmaß für eine beobachtete Datenreihe? Dazu ein Beispiel (**Abb. 79**). Es liegen uns drei unterschiedliche Datensätze (Wertereihen) mit je neun Beobachtungen vor.

Datensatz A	3; 4; 4; 5; 5; 5; 6; 6; 7	Median 5, Mittelwert 5
Datensatz B	1; 2; 3; 4; 5; 6; 7; 8; 9	Median 5; Mittelwert 5
Datensatz C	1; 1; 1; 1; 5; 9; 9; 9; 9	Median 5; Mittelwert 5

**Abb. 79** Beispiele unterschiedlicher Zahlenreihen, die den gleichen Wert des Medians und des Mittelwerts ergeben.

Für alle drei Reihen in **Abb. 79** ergeben sich als Zentralmaß der Median 5 und ein arithmetisches Mittel von 5. Ein Blick auf die Wertereihen verdeutlicht, dass die Reihe „A“ durch dieses Zentralmaß tatsächlich gut beschrieben wird, die Werte in der Reihe „B“ weniger gut durch das Zentralmaß repräsentiert sind, und für die Reihe „C“ beide Zentralmaße offenbar eine ungenügende zusammenfassende Beschreibung darstellen. Wir benötigen eine Quantifizierung der Streuung um das Zentralmaß, um diesen Sachverhalt treffend mit einem Kennwert darstellen zu können; dazu dienen die sog. Streuungsmaße. Streuungsmaße geben an, wie weit die Werte um das Zentralmaß herum streuen, d. h. ob die Beobachtungen stark auf das Zentralmaß hin genormt sind oder in einem sehr weiten Bereich um es herum „streuen“. Auch hier wird zunächst nach den Skalenniveaus differenziert.

Für nominalskalierte Variablen gibt es im Grunde kein sinnvolles Streuungsmaß. Man kann die Liste der vorkommenden Werte heranziehen. Bei ordinal- oder quantitativ skalierten Variablen wird gelegentlich die Spannweite (engl. *range*) angegeben, die Differenz zwischen dem kleinsten und dem größten Wert. Bei ordinalskalierten Variablen kann es sinnvoll sein, den Interquartilbereich anzugeben (näheres unten).

#### 19.3.1 Minimum, Maximum, Spannweite

In den Naturwissenschaften (z. B. Medizin, biolog. Anthropologie) werden häufig auch Minimum, Maximum und Spannweite (engl. *span*) als Streuungsmaße verwendet. Das Minimum ist der beobachtete kleinste Wert, das Maximum der beobachtete höchste Wert einer Datenreihe; die Spannweite ist die Differenz zwischen Maximum und Minimum, sie könnte ebenfalls als Streuungsmaß verwendet werden. In der Ar-

chäologie ist die praktische Bedeutung dieser drei Kennzahlen gering, da sie naheliegenderweise in besonderem Maße von – möglicherweise vom „typischen“ weit entfernten – extremen Werten beeinflusst werden, d. h. nicht geeignet sind, die gesamte Datenreihe treffend zu charakterisieren und eine nützliche Aussage über sie zu machen. Minimum, Maximum und Spannweite sind sehr stark von zufällig vorhandenen – oder auch eben nicht vorhandenen – Extremwerten abhängig.

Bitte selbst ein Beispiel rechnen: Die Teilnehmer machten folgende Gewichtsangaben (geordnete Liste): 52; 53; 53; 55; 56; 64; 66; 70; 75; 75; 77; 96 kg. Die Angaben streuen zwischen 52 (i. e. Minimum) und 96 kg (i. e. Maximum), die Spannweite beträgt  $96 - 52 = 44$  kg.

### 19.3.2 Interquartilabstand (IQR)

Bei Rangdaten und bei quantitativ skalierten Variablen, die bestimmte Voraussetzungen (dazu später) nicht erfüllen, dient der Interquartilabstand (engl. IQR, *interquartile range*) als Streuungsmaß: Er beschreibt in einer geordneten Reihe jene Spanne („mittlerer“ Bereich), in die – symmetrisch um den Median herum – die Hälfte aller Beobachtungen fällt. Ein Viertel der Werte liegt unterhalb des Medians, ein Viertel der Werte oberhalb des Medians.

Die oben in **Abb. 79** als Beispiel aufgeführten Wertereihen sind symmetrisch. Ist das in starkem Maße nicht der Fall, ist die Angabe von Zentralmaßen problematisch, z. B. bei Wertereihen, die sehr „schief“ sind. Beispiel einer sehr schiefen Verteilung: 1; 5; 7; 8; 8; 8; 8; 9.

Selbst rechnen: Das Vorgehen entspricht dem der Ermittlung des Medians. Nun zielt das Verfahren nicht auf dem mittleren Wert ab, sondern auf jenen Wert zwischen dem untersten und dem nächsten Viertel (Q25) und dem Wert zwischen dem dritten und dem letzten Viertel (Q75). Beispiel: Die Teilnehmer eines Kurses machten folgende Gewichtsangaben (geordnete Liste): 52; 53; 53; 55; 56; 64; 66; 70; 75; 75; 77; 96 kg. Es sind zwölf Angaben. Das Quartil Q25 entspricht dem Mittel aus dem 3. und dem 4. Wert:  $53 + 55 / 2 = 54$  kg. Das Quartil Q75 entspricht dem Mittel aus dem 9. und dem 10. Wert:  $75 + 75 / 2 = 75$  kg. Der Interquartilabstand beträgt  $75 - 54$  kg = 21 kg. Aussage: 50 % aller Beobachtungen liegen in einer Spanne von 21 kg.

In der Praxis ist es anschaulicher, zum Median statt der Spanne (IQR) die beiden Quartile mitzuteilen, d.h. die ermittelten Werte von Q25 und Q75. Die hierzu adäquaten Grafiken sind Boxplots, sie zeigen als Box den Bereich Q25 und Q75 und in deren Mitte als Balken den Median (**Abb. 81**).

Da in der Archäologie viele Daten entsprechend ihrer Skala oder Verteilung die Angabe von Mittelwert und Standardabweichung nicht rechtfertigen, sind Median und Interquartilabstand (bzw. Q25 und Q75) ein Paar, das in vielen Fällen als Maßzahl angemessen ist.

### 19.3.3 Standardabweichung

Bei quantitativ skalierten Variablen kann die Standardabweichung (Std.abw.; engl. *sd* oder *st.dev.*, *standard deviation*) als Streuungsmaß dienen. Oft ist es richtiger, statt ihrer ebenfalls den zuvor erläuterten Interquartilabstand zu wählen. Die Standardabweichung (*sd*) gibt – bei normalverteilten Variablen (dazu später) – jene Spanne ( $\pm sd$ ) an, in die symmetrisch um den Mittelwert circa zwei Drittel (genauer: 68,27 %) aller Beobachtungen fallen. In die doppelte Standardabweichung ( $\pm 2sd$ ) fallen 95,45 % aller Werte, in die dreifache Standardabweichung ( $\pm 3sd$ ) fallen 99,73 % aller Werte der Wertereihe.

Standardabweichung selbst rechnen: Die Ermittlung der Standardabweichung erfolgt in mehreren Schritten, die wir nachfolgend Schritt für Schritt nachvollziehen, weil daran wiederum eine in der Statistik typische Vorgehensweise und Begrifflichkeit verständlich wird. Als Beispieldaten wählen wir wieder die Gewichtsangaben der zwölf Teilnehmer eines Kurses: 52; 53; 53; 55; 56; 64; 66; 70; 75; 75; 77; 96 kg.

(a) Mittelwert. Wir berechnen den Mittelwert: Summe der Werte dividiert durch Anzahl der Werte = 66.00 kg.

(b) Berechnung der Abweichungsquadrate: Für jede Beobachtung der Wertereihe berechnen wir das „Abweichungsquadrat“ = (Beobachtung minus Mittelwert) quadriert:

(52 - 66) quadriert: 196  
(53 - 66) quadriert: 169  
(53 - 66) quadriert: 169  
(55 - 66) quadriert: 121  
(56 - 66) quadriert: 100  
(64 - 66) quadriert: 4  
(66 - 66) quadriert: 0  
(70 - 66) quadriert: 16  
(75 - 66) quadriert: 81  
(75 - 66) quadriert: 81  
(77 - 66) quadriert: 121  
(96 - 66) quadriert: 900

(c) Summe dieser Abweichungsquadrate (SAQ):  $196 + 169 + 169 + \dots + 81 + 121 + 900 = 1.958$ .

(d)  $s^2$  Varianz:  $SAQ / (\text{Anzahl der Fälle})^{58} 1.958 / (12) = 163,17$

(e) sd Standardabweichung: Wurzel aus der Varianz:  $\text{Wurzel}(163,17) = 12,77$

Wie eingangs gesagt, all diese Rechnereien nimmt Ihnen künftig die Software ihres Vertrauens ab. Dennoch empfehle ich sehr, diese Zentral- und Streuungsmaße wenigstens einmal anhand einer kleinen Datenreihe selbst zu berechnen, wobei man dies gerne auch mit Hilfe einer Tabellenkalkulation tun kann – wobei man dann dort eben zu Übungszwecken selbst rechnen sollte und nicht die fertigen Funktionen aufruft. Die „Summe der Abweichungsquadrate“ ist etwas, was auch bei vielen anderen statistischen Verfahren berechnet wird.

### 19.3.4 Quantile als robustere Alternativen zu Minimum und Maximum

Besteht das Bedürfnis, die Spannweite einer Datenreihe zu beschreiben, ohne allzu sehr von einzelnen Extremwerten beeinflusst zu sein, bietet sich das Arbeiten mit geeigneten „Quantilen“ an, z. B. Q10 und Q90. Quantile werden aus geordneten Datenreihen abgeleitet; Q10 beschreibt die Grenze zwischen den kleinsten 10 % der Fälle einer Datenreihe und dem Rest; Q90 die Grenze zwischen 90 % der Fälle dieser Datenreihe und den 10 % größten Werten. Zwischen Q10 und Q90 liegen also 80 % aller Fälle einer Datenreihe, während die kleinsten Werte (10 %) und die größten Werte (10 %) ausgeklammert bleiben. Bei umfangreichen Datenserien werden oft Q5 und Q95 gewählt, bei kleineren Datenreihen mit weniger als einhundert Fällen sind Q10 und Q90 geeigneter.

Richtig! Der Median ist identisch mit Q50, und aus Q25 und Q75 ergibt sich der Interquartilabstand (IQR).

In Publikationen werden Q5, Q10, ... Q90, Q95 eher selten verwendet, und auch Sie sollten dies nur in Ausnahmefällen tun. Wenn es aber darum geht, beim Sichten und Kennenlernen der Daten einen guten, stabilen Überblick über die Streuungsweite zu gewinnen, können diese Quantile hilfreich sein. Es gibt auch spezielle Zentralmaße (Mittelwerte), die diesem Ansatz folgen, z. B. der „5 %-trimmed mean“, also ein

---

<sup>58</sup> Anmerkung zur Varianz: Der hier verfolgte Weg mit dem Ausdruck „Division durch (Anzahl der Fälle)“ ist in der deskriptiven Statistik richtig. In der schließenden Statistik wird die Varianz durch Division mit der „(Anzahl der Fälle minus 1)“ berechnet. Die einschlägigen fertigen Funktionen der gängigen Tabellenkalkulationsprogramme operieren hier leider nicht alle gleich und nicht immer richtig.

Mittelwert (*mean*), der eben die höchsten und die kleinsten 5 % einer Datenreihe aus der Berechnung des arithmetischen Mittels ausschließt. Hier gilt das Gleiche: Für die Phase der ersten Datensichtung ein nützlicher Wert, ansonsten eher ungebräuchlich.

Abschließend noch etwas zu Begriffen und „Etikette“: Quartil / Quantil. Die Ränge 25 % und 75 % nennt man Quartil (mit  $r$ ), die weiteren Ränge wie z.B. 10 %, 5 % Quantile (mit  $n$ ) – weshalb IQR der Interquartilabstand ist.

### 19.3.5 Kurze Zusammenschau

Alle Ergebnisse für unsere Beispielreihe von Körpergewichten:

- Minimum 52 kg, Maximum 96 kg, Spanne 44 kg;
- Median 65 kg; IQR:  $Q_{25} = 54$  kg,  $Q_{75} = 75$  kg; d. h. 50 % aller Fälle liegen in der Spanne 54 – 75 kg;
- Mittelwert 66 kg, Standardabweichung 12,77 (13,34)<sup>59</sup>; d. h. 68,27 % aller Fälle liegen in der Spanne 52.7 – 79.3 kg.

### 19.3.6 grafische Darstellungen, u. a. als Selbstkontrolle

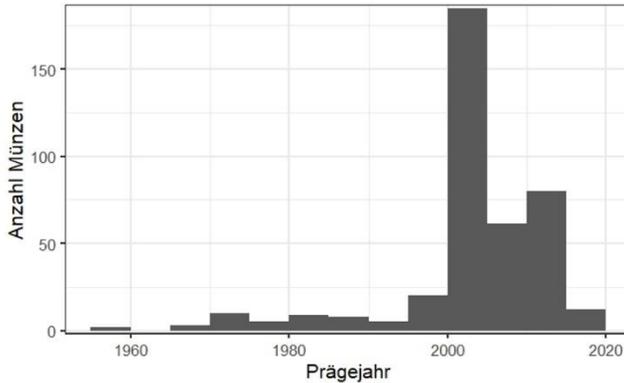
Es ist sinnvoll, sich während der Arbeit zu den Kennzahlen immer auch eine grafische Darstellung anzufertigen. Wenn Sie später Ihre Ergebnisse präsentieren, werden Sie sparsam mit solchen Bildern umgehen und nur das wirklich Wichtige zeigen. Doch während der Arbeit bilden die Grafiken eine wichtige (Selbst-) Kontrolle, z. B. für die Frage: Liegt tatsächlich ein unimodales Bild vor, d. h. sind die Kennzahlen auch sinnvoll?

Die beiden Bilder **Abb. 80** und **Abb. 81** geben die im Jahr 2017 gezogenen Stichproben unseres Geldbörsen-Datensatzes wieder. Das Histogramm **Abb. 80** könnte man – z. B. in der Abbildungsunterschrift – ergänzen um die Angabe der passenden Kennzahlen, in diesem Fall:  $n = 400$  (d. h. Anzahl der Beobachtungen = 400); Mittelwert Prägejahr 2003,2, Std.abw. 9,69 Jahre.

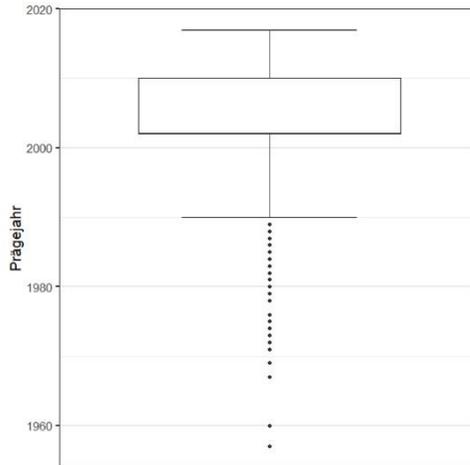
Der Grafiktyp **Abb. 81** heißt „Boxplot“ (auch: *box-and-whisker plot*). Boxplots zeigen nicht Mittelwert und Standardabweichung an, sondern den Median (dicke waagerechte Linie) und den Interquartilbereich (Kasten), sowie (in diesem Fall) auch die (übliche) Spanne der Daten (Linien mit Querstrichen am Ende, „Antennen“). Wollte man hierzu geeignete Kennzahlen angeben (z. B. in der Abbildungsunterschrift), könnten diese lauten: Median: 2002,  $Q_{25} - Q_{75}$ : 2002 – 2010, IQR: 8 Jahre.

---

<sup>59</sup> Der Wert 12,77 ergibt sich aus der dargelegten, für die beschreibende Statistik korrekten Berechnung, der Wert 13,34 aus der in der schließenden Statistik üblichen Berechnung der Varianz.



**Abb. 80** Prägejahre der im Jahr 2017 ausgezählten Geldbörsen (n=400), hier als Histogramm dargestellt.



**Abb. 81** Prägejahre der im Jahr 2017 ausgezählten Geldbörsen (n=400), hier als Boxplot dargestellt.

Der Boxplot **Abb. 81** wurde mit Hilfe von **R** erzeugt. Die in einem Querstrich (englisch *hinge* genannt), also T-förmig endenden „Antennen“ (engl. „*whisker*“ – Barthaare), zeigen den Bereich an, wo die extremen „Ausreißer“ liegen. Als Ausreißer werden gemäß J. W. Tukey (1977) Beobachtungen bezeichnet, die jeweils innerhalb einer Spanne liegen, die um den Faktor 1,5 des IQR dies- und jenseits des IQR liegen. Die kleinen Punkte darunter oder darüber sind dann die einzelnen Ausreißer. Caveat: Manche anderen Statistikprogramme definieren die Bedeutung der T-förmigen Linien und die

Begriffe Ausreißer oder Extremwert etwas anders, d. h. man sollte dies ggf. (a) dort in der Dokumentation nachlesen und (b) in seriösen Arbeiten den Lesern auch bei der ersten Verwendung einer solchen Grafik mitteilen. PAST und **R** folgen hinsichtlich der Antennenlänge dem Ansatz von Tukey.

Sie kennen nun den Unterschied zwischen Mittelwert und Median, und werden zu den Grafiken stets auch die passenden Kennzahlen verwenden. Vertauschungen zwischen „Mittelwert“ und „Median“ sind unprofessionell.

#### 19.4 PRAXIS PAST: ZENTRAL- UND STREUUNGSMAßE MIT PAST

Abstrakt lautet der Weg: Daten in PAST in eine Spalte eingeben oder aus Tabelle importieren. Spalte als aktiv markieren/selektieren. => Univariate; => Summary statistics. Das aufpoppende Fenster gibt alle nötigen Werte aus; die zusätzlichen Angaben können Sie einstweilen ignorieren. Achtung, der „Std. error“ = *standard error* (des Mittelwerts) ist etwas anderes als die „Std. dev.“ = *standard deviation* (Standardabweichung).

Als Beispiel wollen wir die Femurlängen im Gräberfeld von Zwentendorf untersuchen. Wir gehen davon aus, dass die Datenerfassung und Verwaltung zunächst in LO-Calc vorgenommen wurde. Daher müssen die Daten zunächst von dort nach PAST überführt werden. Vernachlässigt man die Option Copy-&Paste, ist der normale Weg, die Daten in LO-Calc als \*.cvs-Datei abzuspeichern und sie von dort in PAST einzulesen. PAST starten (am besten liegen Daten und PAST im gleichen Ordner), dann => File, => Open, und die entsprechende Datei aussuchen. In dem aufklappenden Menü sind vermutlich folgende Einstellungen zielführend: „Rows contain“: Only data cells; „Columns contain“: Names, data; „Separator“ Comma; dann „Import“. Manchmal muss man etwas probieren, bis man die passenden Einstellungen gefunden hat.

Möglicherweise sind die relativ vielen leeren Felder (fehlenden Werte) noch in einem unpassenden Zustand, weil LO-Calc dort vermutlich „#NULL!“ eingetragen hat, womit PAST nicht zurechtkommt. Also gilt es, diesen Mangel zu bereinigen: Spalte oder einige Spalten markieren, dann: => Edit; => Replace; in das Menü bei „Find what“ eintragen „#NULL!“, in „Replace with“ nichts eintragen (d. h. leer lassen), dann „Replace“, und etwas Geduld: bei größeren Datensätzen ist diese Operation nicht schnell. Danach am besten den Datensatz (zwischen-) sichern: => File; „Save as...“, und das PAST-Format (\*.dat) wählen. Ist der Datensatz auf diese Weise bereinigt und arbeitsfähig, kann die eigentliche Auswertung beginnen.

Abweichend vom in der Anthropologie Üblichen hat der Bearbeiter von Zwentendorf auch die Femurlängen der Kinder und Jugendlichen erfasst. Da wir uns im Folgenden (wie üblich) auf die Erwachsenen beschränken wollen, muss der Datensatz zunächst auf die Erwachsenen begrenzt werden. Das hätte man auch in LO-Calc vor dem Export der Daten machen können, aber zur Übung sollen hier die Möglichkeiten von PAST eingesetzt werden:

Spalte „Alter“ markieren; => Edit; => Rearrange; => Sort descending: Nun wird die gesamte Tabelle absteigend nach dem Alter der Toten bestimmt. Gemäß der Codierung wollen wir uns mit 4 adult, 5 matur und 6 senil beschäftigen und alle anderen Fälle verwerfen. Also die Tabelle herunterscrollen, bis man auf den Wechsel von Alter = 4 zu Alter = 3 stößt. Die erste Zeile mit Alter = 3 ganz links am Zeilenanfang markieren, sie sollte danach in PAST hellblau hervorgehoben sein. Dann die Shift-Taste drücken und gedrückt halten (das ist die Taste für Groß-/Klein-Schreibung), und weiter bis ans untere Tabellenende scrollen. Jetzt sollten alle Felder dieses unteren Tabellenteils hellblau markiert sein. Dann => Edit; => Remove: Die hellblau markierten Felder sind gelöscht. Datensatz unter neuem Namen sichern, z. B. „Erwachsene“. Es sollten jetzt noch 241 altersbestimmte erwachsene Individuen im Datensatz vorhanden sein. Damit sind die Datenvorbereitungen erstmal abgeschlossen.

Nun zur Statistik: Die Spalte mit F1 markieren (Spaltenkopf anklicken, Spalte wird hellblau), => Univariate; => Summary statistics: Die Ergebnisse werden ausgespielt, u. a. erfahren wir:

Anzahl der Beobachtungen (N) 89; Mittelwert (mean) 438,2; Stdabw. (std.dev) 27,4; Median 436, Q25 – Q75: 418,5 – 454,5.

Zur Veranschaulichung ist es sinnvoll, auch eine Grafik anzufertigen; angemessen ist – bei biologisch gewachsenen Daten als erste Arbeitshypothese – vermutlich ein Histogramm: => Plot; => Histogram. Die aufpoppende neue Fläche zeigt das Histogramm; sie beinhaltet zusätzlich die Möglichkeit, das Histogramm näher zu steuern. Empfehlung: „Start“ auf 380 setzen, „End“ auf 520 setzen (d. h. auf den jeweils nahegelegenen runden Wert), „Bins“ (Anzahl der Klassen) auf „7“ setzen. Warum 7? – die Spanne zwischen 380 und 520 beträgt 140, d. h. 7 Klassen á 20 mm, oder 14 Klassen á 10 mm. Mit 7 resp. 14 ergeben sich auf der Strecke 380 mm bis 520 mm gut brauchbare Klassenweiten für das Histogramm. Probieren Sie beide Einstellungen aus.

Das Histogramm erweist die Datenreihe als unimodal, einigermaßen symmetrisch und glockenkurven-ähnlich. Bei feinerer Auflösung (bins 14) deutet sich allerdings eine Zweigipfligkeit an, was mit dem Geschlechtsdimorphismus von (tendenziell

kleineren) Frauen und (tendenziell größeren) Männern zusammenhängen könnte. Übrigens, hinter der Schaltfläche „Graph settings“ (unten) verbirgt sich ein weiteres Menü mit zahlreichen sehr nützlichen Einstellungsoptionen für das Histogramm: Man kann das Layout des Histogramms mit PAST recht gut steuern.

Dem soeben entdeckten Frauen-Männer-Aspekt gehen wir nach und streben getrennte Statistiken für Frauen und Männer an. Man erreicht dieses Ziel wieder mit Sortieren, Markieren und Statistiken bestellen: Die Tabelle nach Geschlecht sortieren; die Spalte F1 vom ersten bis zum letzten Mann – und in einem Folgeschritt von der ersten bis letzten Frau – (1) markieren; via „Univariate“ und „Summary Statistics“ die Daten bestellen. Danach ergibt sich:

38 Männer, Mittelwert 457,7; Std.abw. 24,8; Median 454,5; Q25-Q75 441,8 – 471,5.  
51 Frauen, Mittelwert 423,6; Std.abw. 18,9; Median 420,0; Q25-Q75 412,0 – 436,0.

Auf ähnliche Weise erzeugen wir je ein Histogramm. Um die beiden Histogramme besser vergleichen zu können, setzen wir die gleichen Start- und Endwerte und Klassenweiten für beide Geschlechter ein, also 380 – 520 bei z. B. 14 Klassen. Damit man die Bilder vergleichen kann, muss man sie jeweils in eine externe Datei abspeichern: Auf der Fläche Histogramm unten die Schaltfläche „Graph settings“ klicken und in dem neuen Fenster unten rechts bei „Export“ die nötigen Einstellungen vornehmen (ggf. zuvor links bei „Size“ eine geeignete Größe der Grafik bestellen).

## **19.5 PRAXIS R: ZENTRAL- UND STREUUNGSMAßE – AM BEISPIEL ZWENTENDORF**

Abstrakt lautet der Weg: RStudio (oder das rohe **R**) und den R-Commander starten; Daten mit dem R-Commander aus einer bestehenden Tabelle importieren oder im R-Commander selbst erfassen: => Datenmanagement; => neue Datenmatrix; im aufpoppenden Fenster dem Datensatz einen geeigneten Namen geben & OK; die aufpoppende neue Datentabelle ausfüllen & OK. Dann in die Analyse einsteigen: => Statistik; => Deskriptive Statistik; => Zusammenfassungen numerischer Variablen; im aufpoppenden Fenster die gemeinte Variable(n) auswählen & OK. Die Statistiken werden im Fenster links unten von RStudio angezeigt. Wenn man im R-Commander bleibt, kann man auch schnell die nötigen Grafiken abrufen: => Grafiken; > Histogramm; ggf. die passende Variable aussuchen & OK: Im Fenster rechts unten erscheint das Histogramm der Datenreihe.

Als Beispiel wollen wir die Femurlänge F1 im Gräberfeld von Zwentendorf untersuchen. Dazu haben wir mit RStudio und R-Commander die Datenmatrix „Zwentendorf“ geladen. Abweichend vom in der Anthropologie Üblichen hat der Bearbeiter von Zwentendorf auch die Femurlängen der Kinder und Jugendlichen erfasst, daher muss ein Altersfilter gesetzt werden, denn wir wollen uns auf die Erwachsenen („age“ ab 4) konzentrieren.

=> Datenmanagement; => aktive Datenmatrix; => Teilmenge der aktiven Datenmatrix; in dem neu aufklappenden Fenster ist die mittlere Zeile relevant „Anweisung für die Teilmenge“, wo eingegeben wird: „age > 3“; im Fenster darunter „Name für die aktive Datenmatrix“ geben wir ein „Erwachsene“ – so wird die bestehende Datenmatrix „Zwentendorf“ nicht überschrieben, sondern zusätzlich eine zweite Matrix angelegt; OK; ein Klick auf die Schaltfläche „Datenmatrix betrachten“ im R-Commander spielt links oben in RStudio die aktive Datenmatrix ein; wir kontrollieren visuell, dass die Spalte „age“ nur Werte ab 4 aufführt. Dann holen wir die gebrauchten Statistiken:

=> Statistik; => deskriptive Statistik; => Zusammenfassungen numerischer Variablen; in der aufklappenden Schaltfläche nach „F1“ suchen (die Variablen erscheinen hier alphabetisch sortiert), anklicken & OK. Das linke untere Fenster von RStudio spielt das Ergebnis aus:

mean	sd	IQR	0%	25%	50%	75%	100%	n	NA
437.3667	27.9544	34.5	378	418.125	435.5	452.625	514.5	90	159

Bei 159 fehlenden Fällen („NA“ – not available) haben wir 90 auswertbare Beobachtungen („n“), die einen Mittelwert von 437,4 mm („mean“) bei einer Std.abw. („sd“) von 29.95 ergeben. Der Median beträgt 435,5 mm („50 %“), bei einem IQR von 418,1 – 452,6 („25 %“, „75 %“). Mittelwert und Median liegen relativ nahe beieinander, was für eine annähernd symmetrische Verteilung spricht.

Zur Kontrolle erzeugen wir ein Histogramm: => Grafiken; => Histogramm; in der Schaltfläche nach F1 suchen & draufklicken & OK. Das rechte untere Fenster von RStudio (ggf. Reiter „Plot“ anklicken) zeigt das Histogramm. Es ist unimodal und annähernd glockenförmig, weshalb (zunächst) das Weiterarbeiten mit Mittelwert und Std.abw. berechtigt ist.

Nur der Ergänzung halber erzeugen wir noch einen einfachen Boxplot: => Grafiken; => Boxplot; => F1 aussuchen & OK; wiederum rechts unten in RStudio anschauen.

Wir untersuchen den Geschlechtsdimorphismus. Dazu muss die bisherige Variable „sex“ zu einem „Faktor“ gemacht werden – so heißen in **R** Variablen, die Gruppen/Gruppierungen anzeigen.

=> Datenmanagement; => Variablen bearbeiten; => Konvertiere numerische Variablen in Faktoren; Variable „sex“ aussuchen; das voreingestellte Häkchen bei „verwende Etiketten“ belassen und in dem aufklappenden Fenster zur Zahl 1 „Mann“ eingeben, zur Zahl 2 „Frau“ & OK. Jetzt sind „Mann“ und „Frau“ als Gruppen definiert. Erneut lassen wir uns die Statistiken und ein Histogramm ausgeben:

=> Statistik; => Deskriptive Statistik; => Zusammenfassung für numerische Variablen, dort als Variable F1 aussuchen, nun aber zusätzlich auch die Schaltfläche „Zusammenfassung für Gruppen“ bedienen und in dem Fenster „Sex“ aussuchen, OK & OK. Das linke untere RStudio-Fenster weist nun die Statistiken für F1 getrennt nach Mann und Frau aus:

	mean	sd	IQR	0%	25%	50%	75%	100%	F1:n	F1:NA
Mann	457.5658	24.81259	27.500	394	442	454.25	469.500	514.5	38	69
Frau	422.6058	19.77083	23.125	378	412	419.75	435.125	480.5	52	60

Nicht überraschend sind die Femora der Männer (n=38) mit Mittelwert 457,6 mm um etwa 3,5 cm länger als die der Frauen (n=52) mit Mittelwert 422,6 mm.

Nun noch das zugehörige Histogramm: => Grafiken; => Histogramm; in der neu aufklappenden Schaltfläche die Variable F1 aussuchen und wie zuvor nun zusätzlich die Schaltfläche „Grafiken für die Gruppen“ bedienen; „sex“ aussuchen, OK & OK. Das rechte untere Fenster von RStudio (ggf. Reiter Plot aktivieren) zeigt ein gestapeltes Histogramm, das uns die o. g. Eindrücke aus den Zahlen nochmals veranschaulicht. Wir werden später erforschen, wie wir diese Darstellung verschönern können und inwieweit es noch passendere Grafiken gibt, das hier untersuchte Phänomen darzustellen. Wenn Sie selbst schon etwas vorab probieren möchten: hinter => Grafiken stecken bei Dichteschätzung, Stamm-Blatt-Diagramm, Plot für arithmetische Mittel und Strip Chart weitere interessante Möglichkeiten.

## 19.6 „JACKKNIFING“

Nehmen Sie sich etwas Zeit für ein lehrreiches (und seriöses) „Spielen“: mit PAST oder mit **R** ein paar weitere Berechnungen für Mittelwert, Standardabweichung, Median und IQR. Aber mit System, denn das spielerische Experimentieren soll dazu dienen, Erfahrungen zu gewinnen – wir führen ein „Jackknifing“ durch (statistischer Fachbegriff, abgeleitet von *jackknife*, Taschenmesser).

Wählen Sie einen Ihnen gut bekannten Datensatz, z. B. unseren Beispieldatensatz Geldbörsen, und konzentrieren sich dort auf eine Variable, z. B. auf das Prägejahr der Münzen. Der Datensatz ist recht umfangreich, für unsere Übung schlage ich vor, daraus 50 oder 100 beliebige Datensätze auszuwählen. Lassen Sie uns systematisch Median und die Quartile Q25 und Q75 sowie den Mittelwert und die Standardabweichung berechnen. Protokollieren Sie das Ergebnis in einer Tabelle. Nun das ein Experiment: Löschen Sie die Datensätze Nr. 1 bis 5 und berechnen die fünf Maßzahlen für die verbliebenen Werte. Danach die gelöschten Datensätze 1 bis 5 wiederum zurück zum kompletten Datensatz, zugleich die Datensätze Nr. 6 bis 10 löschen und die Kennzahlen berechnen, usw. usf. Notieren Sie jeweils die Ergebnisse und vergleichen Sie diese nach einer Weile miteinander.

Diskussion: Das genaue Ergebnis, das Sie erzielen, lässt sich nicht exakt voraussagen, wenn man statt Daten vorzugeben mit Realbeobachtungen arbeitet. Aber meistens zeigt sich beim Jackknifing folgendes Bild: Mittelwert und Standardabweichung sind empfindlicher als Median und IQR, d. h. es ergeben sich gerade bei kleinen Fallzahlen stärkere Schwankungen, wenn man Daten streicht. Besonders empfindlich ist der Mittelwert bei „Ausreißern“, d. h. einzelnen, extrem hohen oder extrem niedrigen Werten. Eine relativ kleine Serie mit einem oder wenigen Ausreißern liefert ganz andere Resultate als dann, wenn man die Ausreißer (z. B. beim Jackknifing zufällig) streicht. Median und IQR hingegen sind gegenüber einzelnen extremen Beobachtungen ziemlich unempfindlich.

Wenn man ein solches Jackknifing, d. h. das temporäre Löschen von einzelnen Fällen in einem Datensatz, systematisch betreibt, spricht man vom *Sampling* (systematisches Stichproben ziehen); geschieht dieses Sampling geordnet und mit Zurücklegen für den ganzen Datensatz (*resampling*), spricht man von *Bootstrapping* (statistischer Fachbegriff, abgeleitet von *bootstrap*, Stiefelriemen/Schnürsenkel; i. e. „sich an den eigenen Schnürsenkeln aus dem Sumpf ziehen“) und betritt das Feld der ebenso komplizierten wie derzeit beliebten Bayes-Statistik. Keine Angst, so mühsam wie gerade eben, d. h. manuell, erfolgt das in der Praxis nicht; das Bootstrapping ist eine Aufgabe für die Statistik-Programme. Bei PAST z. B. ist es bereits bequem eingebaut: => Univariate; => Summary statistics; und Häkchen an „Bootstrap“ setzen. Achtung: das anschließende Rechnen kann dauern! Die Sampling- (oder Resampling-) Verfahren sind zeitaufwändig, trotz moderner Computer. Ergebnis: Aus den sehr vielen Stichproben errechnet PAST u. a. ein Konfidenzintervall für den Mittelwert. In der hinter unserer Stichprobe liegenden Grundgesamtheit liegt er mit 95%-iger Wahrscheinlichkeit zwischen der angegebenen Unter- und Obergrenze.

Caveat: Mittelwert und Standardabweichung sind nur unter bestimmten Voraussetzungen vernünftige Angaben: Die Variablen sollten „annähernd normalverteilt sein“. Wir werden erst etwas später erlernen, was dies bedeutet. Einstweilen halten wir fest: Wir geben Mittelwert und Standardabweichung nur dann an, wenn die Variablen quantitativ skaliert sind, unimodal verteilt sind und die Verteilung im Histogramm annähernd symmetrisch aussieht. Daher sind wir grundsätzlich bestrebt, die Daten sowohl zu visualisieren (Histogramm o. ä.) als auch durch statistische Maßzahlen zu charakterisieren.

## **20 STETIGE WAHRSCHEINLICHKEITSVERTEILUNGEN, NORMALVERTEILUNG, VERGLEICH VON MITTELWERTEN**

---

Zum Themenfeld gezählte Häufigkeiten und Chi-Quadratverfahren gehört auch das Thema „diskrete Wahrscheinlichkeitsverteilungen“. Es ist recht kompliziert und wurde hintangestellt, damit wir schnell zu wichtigen Arbeitstechniken gelangen, wie man mit quantitativen Daten umgehen kann, wo uns allerdings das Thema „stetige Wahrscheinlichkeitsverteilungen“ wiederbegegnet. Hier geht es vor allem um die zentrale Eingangsfrage, ob die Daten einer sog. Normalverteilung unterliegen, weil sich daran entscheidet, welche Kennzahlen und Test-Verfahren geeignet sind, wenn man verschiedene Stichproben beschreiben und vergleichen will. Doch – ohne Formeln – was sind und wozu dienen diese „Wahrscheinlichkeitsverteilungen“? Wahrscheinlichkeitsverteilungen sind mathematische Modelle, die eine Annahme über die Wirklichkeit machen, d. h. ein hypothetisches Modell in abstrakte Zahlen und Formeln übersetzt haben. Danach wiederum ist es möglich, zu untersuchende reale Serien mit diesen theoretischen Modellen zu vergleichen und dabei zu prüfen, ob die Wirklichkeit einem bestimmten Modell nahekommt. Nicht zuletzt erfordern viele statistische Verfahren, dass die untersuchten Daten gewissen Mindeststandards genügen und nicht nur ein bestimmtes (Mindest-)Skalenniveau erreichen, sondern auch einer bestimmten theoretischen Verteilung unterliegen resp. nicht signifikant davon abweichen.

Ein Beispiel: Zwei Lehrer haben gänzlich unterschiedliche Vorstellungen von ihren Schülern und bauen darauf entsprechend unterschiedliche Unterrichtsmodelle auf, woraus letztlich auch unterschiedliche Parameter für eine Erfolgsmessung resultieren. Lehrer A behauptet „Alle Schüler sind gleich“. Dem entspreche in der Statistik das Modell einer „Gleichverteilung“. Folglich erwartet Lehrer A, dass alle Schüler am Ende gleich gute Noten erreichen und zuvor im konkreten Anwendungsfall, dass sie beim Diktat, bei der Lösung eines Mathe-Tests o. ä. auf die Dauer gesehen die gleiche

Menge an Leistungspunkten erwerben resp. Zahl an Fehlern machen. Der Statistiker nimmt dieses Modell ernst: Leistungen bei den diversen Schüler-Tests erheben und prüfen, ob die Abweichungen zwischen dem individuell Beobachteten und dem nach dem Modell „Gleichverteilung“ Erwarteten (i. e. die mittlere Zahl an Fehlern in der gesamten Klasse) noch mit dem Zufall vereinbar sind.

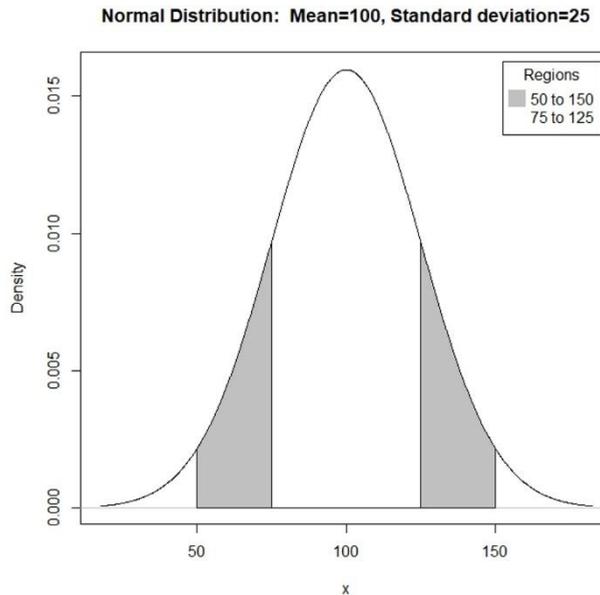
Lehrer B hingegen denkt anders, seine Grundannahme – nennen wir es Überzeugung, Ideologie o. ä. – ist: „Alle Schüler sind normal“. Er meint, es gebe quasi einen „mittleren Schüler“ und natürlicherweise eben einige andere Schüler, die talentierter, eifriger, intelligenter seien als dieser hypothetische normale Schüler und leider auch solche, die weniger fleißig und begabt seien als „normal“. Dem entspräche in der Statistik das Modell einer „Normalverteilung“. Der Statistiker nimmt dieses Modell ernst und formuliert unter der Annahme einer Normalverteilung aus dem klassenweit beobachteten Mittelwert an Fehlern beim Diktat, Mathe-Test o. ä. wiederum Erwartungswerte und vergleicht diese mit den tatsächlich beobachteten Fehlerhäufigkeiten. Ein gewisses Maß an zufälligen Abweichungen zwischen dem Ideal und der Wirklichkeit wird man einräumen, doch wie groß ist die Toleranz, ab wann wird man eingestehen müssen, dass die tatsächlichen Leistungen der Schüler dem hypothetischen Modell nicht entsprechen?

Die Antwort auf die Frage „Gleichverteilung oder Normalverteilung?“ ist einerseits für die beiden Lehrer zunächst von rein ideologischem Interesse, denn beide möchten recht haben und ihre Ideologie empirisch bestätigt sehen. Allerdings ergäben sich aus dem Befund für ihre Zukunft auch unterschiedliche Ansätze, wie sie ihre eigenen Fortschritte messen würden, woran sie z. B. festmachen werden, ob versuchsweise eingeführte Änderungen in ihrer Didaktik erfolgreich waren oder nicht. Das theoretische Modell liefert also auch die modell-adäquate, jeweils unterschiedliche „Benchmark“.

Nun ersetzen Sie das Bild von Lehrern und Schülern gedanklich beispielsweise durch die Begriffe Museumsleitung und Besucherzufriedenheit und denken an Zielvorgaben zur Verbesserung eines Museums: Sind die Besucher gleich oder normal? Ist folglich die Steigerung der „mittleren Besucherzufriedenheit“ die adäquate Benchmark („*Unzufriedene wird es immer geben*“), oder die Frage, ob alle Besucher gleichermaßen zufrieden(er) das Haus verlassen? Das Thema „Wahrscheinlichkeitsverteilungen“ hat somit eine (hohe) statistik-interne Bedeutung, weil es über die Anwendbarkeit oder Nicht-Anwendbarkeit weitergehender Verfahren entscheidet, hat aber eben auch ganz praktische Interpretierbarkeiten und Konsequenzen. Schauen wir uns also die wichtige Normalverteilung näher an.

## 20.1 NORMALVERTEILUNG

Die Normalverteilung ist eine eingipflige, symmetrische Verteilung (**Abb. 82**);<sup>60</sup> sie wird beschrieben durch den Mittelwert und die Standardabweichung (**Abb. 83**). Bei Vorliegen einer Normalverteilung sind Mittelwert, Median und Modalwert identisch.



**Abb. 82** „Glockenkurve“: Bild einer Normalverteilung mit Mittelwert 100 und Standardabweichung 25. In der mittleren weißen Zone (mean  $\pm$ std.dev:  $100 \pm 25$ , d.h. 125-175) liegen ca. 2/3 aller Fälle. Nimmt man die beiden grauen Zonen rechts und links hinzu ( $100 \pm 50$ , d. h. mean  $\pm 2$  std.dev.), sind 96,45% aller Fälle eingeschlossen.

---

<sup>60</sup> Die Kurve **Abb. 82** wurde mit Hilfe des R-Commanders gezeichnet: => Verteilungen; => Stetige Verteilungen; => Normalverteilung; => Graphik der Normalverteilung... In dem aufklappenden Fenster Werte für Mittelwert und Std.abw. eingeben.

## Standardabweichung ( $\sigma$ , sd, std.dev., Std.abw.)

---

Bei einer normalverteilten Variablen gibt die Standardabweichung  $\sigma$  jenen Bereich an, innerhalb dessen die Werte symmetrisch um den Mittelwert streuen.

Im Bereich  $\pm 1 \sigma$  liegen 68,27 % aller Beobachtungen („ca. zwei Drittel“).

Im Bereich  $\pm 2 \sigma$  liegen 96,45 % aller Beobachtungen.

Im Bereich  $\pm 3 \sigma$  liegen 99,73 % aller Beobachtungen.

---

**Abb. 83** Beziehung zwischen der Anzahl der Beobachtungen, dem Mittelwert und der Standardabweichung bei Vorliegen einer Normalverteilung.

Häufige Abweichungen von der Normalverteilung sind:

- Asymmetrie, „Verteilung linkssteil“ (und nach rechts flach verlaufend): Mittelwert größer Median, Kennzahl Schiefe größer 0.

Derartige Verteilungen „mit wenigen Ausreißern nach oben“ lassen sich oft durch geeignete Transformationen der Variablen in eine Normalverteilung überführen. Geläufig sind die Wurzeltransformationen und die logarithmische Transformation.

- Asymmetrie, „Verteilung rechtssteil“ (und von links her nur flach ansteigend): Mittelwert kleiner Median, Kennzahl Schiefe kleiner 0.

Solche Verteilungen können z. B. beim Gewicht von Münzserien durch unterschiedlich starke Abnutzung aus den ehemals normalverteilten Gewichten entstehen.

- Die Werte sind stärker um den Mittelwert gestreut, d. h. weniger als „normal“ auf diesen hin genormt; Kennzahl Wölbung deutlich kleiner als 0.

- Die Werte streuen sehr gering um den Mittelwert, d. h. sie sind stärker als „normal“ auf diesen hin normiert; Kennzahl Wölbung deutlich größer als 0.

Dies ist z. B. bei vielen modernen Industrieprodukten der Fall, wie etwa den Maßen von Schrauben und Ähnlichem, weil diese heute extrem exakt gefertigt werden.

Eine besondere Form der Normalverteilung ist die „standardisierte Normalverteilung“ (kurz: „Standardnormalverteilung“); sie hat den Mittelwert 0 und die Standardabweichung 1. Mithilfe einer sog. „z-Transformation“ wird eine vorliegende empirische Normalverteilung in eine standardisierte Normalverteilung überführt. Wie das geht? – ziemlich einfach: Man berechnet Mittelwert und Standardabweichung der vorliegenden Stichprobe; dann wird von jeder einzelnen Beobachtung der Mittelwert der ge-

samten Stichprobe abgezogen und dieser Wert durch die Standardabweichung der gesamten Stichprobe dividiert. Resultat: die Werte sind „z-transformiert“ und folgen einer standardisierten Normalverteilung.

## 20.2 SCHIEFE UND WÖLBUNG

### 20.2.1 Schiefe

Bei Vorliegen des Idealfalls streuen die Daten bei einer Normalverteilung symmetrisch um den Mittelwert. Mit „Schiefe“ (engl. *skewness*) benennt man die Abweichung von dieser Symmetrie, sie bewegt sich zwischen -1 und +1. Eine gänzlich symmetrische Kurve hat die Schiefe 0.

- Eine Kurve, deren Modalwert links des Medians und deren Median links des Mittelwerts liegt, d. h. bei der also der Median kleiner als der Mittelwert ist, hat eine positive Schiefe (größer Null) und wird „linkssteil“ oder „rechtsschief“ (i. e. rechts flach) genannt.
- Eine Kurve, deren Median und Modalwert rechts des Mittelwerts liegen, d. h. bei der der Median größer als der Mittelwert sind, hat eine negative Schiefe (kleiner Null) und wird „rechtssteil“ oder linksschief (i.e. links flach) genannt.

### 20.2.2 Wölbung und Exzess

Mit „Wölbung“ (engl. *kurtosis*) beschreibt man die Frage, ob eine Verteilung stärker oder schwächer als „normal“ auf den Mittelwert hin zentriert ist. So nützlich diese Kennzahl grundsätzlich ist, so verwirrend sind einige praktische Details. Bei einer ideal normalverteilten Variablen ist die Wölbung (*kurtosis*) gleich 3.<sup>61</sup> Höhere Werte beschreiben eine spitzere Glockenkurve, die stärker als normal auf den Mittelwert hin zentriert ist – ein Bild, das sich etwa bei stark auf ein Ideal normierten Industrieprodukten ergibt. Niedrigere Werte beschreiben eine flachere Kurve, bei der die Werte stärker streuen, also breiter um den Mittelwert herum streuen als bei einer Normalverteilung.

Da diese Kennzahl zumeist in Bezug auf eine Normalverteilung verwendet wird, hat man zusätzlich den Begriff „Exzess“ (engl. *excess*) eingeführt, der schlicht als „Wölbung minus 3“ definiert ist, sodass eine Normalverteilung den Exzess 0 hat, negative Werte für flachere, positive Werte für steilere Kurven stehen. Tückischerweise

---

<sup>61</sup> Diese Zahl leitet sich aus dem theoretischen Modell einer Normalverteilung und der Berechnungsweise der Wölbung ab.

unterscheiden sich die Berechnungen der Wölbung bei den einzelnen Statistikprogrammen leicht,<sup>62</sup> und tückischerweise sagt **R** „*kurtosis*“, meint aber „*excess*“ 😊. Jetzt, da wir dies wissen, können wir die diesbezüglichen Ausgaben von **R** auch mit Gewinn lesen.

### 20.2.3 Praxis: Schiefe und Wölbung

Wir hatten beide Parameter im Grunde bei unseren praktischen Beispielen zu Mittelwert, Standardabweichung, Median und IQR schon gesehen. Der Weg dazu in PAST ist wie schon beschrieben: Datenspalte(n) markieren; => Univariate; => Summary statistics.

Für ein praktisches Beispiel mit RStudio & R-Commander laden wir den Datensatz Zwentendorf (Achtung: Dezimaltrennzeichen ist Komma; Variablentrennung bei \*.dat ist Tab, bei \*.csv ist es Semikolon) und reduzieren ihn danach auf die Erwachsenen („Age >3“). Danach betrachten wir die Femurlänge F1: => Statistik; => Deskriptive Statistik; => Zusammenfassungen für numerische Variablen (usw.), wählen die Variable F1 (markiert=blau), klicken jedoch zusätzlich auf den Reiter „Statistik“ und bestellen in der aufklappenden Schaltfläche zusätzlich „Schiefe“ und „Kurtosis“ (Wölbung).

Schauen wir uns das Beispiel der Daten aus Zwentendorf an, die Femurlänge F1 für alle Erwachsenen.

mean	sd	IQR	skewness	kurtosis	0%	25%	50%	75%	100%
437.3667	27.9544	34.5	0.5596156	0.1244277	378	418.125	435.5	452.625	514.5

Der Median – hier abzulesen bei „50 %“ – ist mit 435,5 mm kleiner als der Mittelwert von 437,4 mm. Die Schiefe weicht mit 0,56 von Null ab und ist etwas größer, d. h. die Kurve ist (etwas) „linkssteil“. Die Wölbung liegt mit 0,12 etwas über Null, d. h. die Streuung der Werte ist etwas geringer als bei einer Normalverteilung.

Wir erinnern uns, dass wir bereits auf den Geschlechtsdimorphismus gestoßen waren, wonach Frauen im Mittel etwas kleiner und Männer im Mittel etwas größer sind und der Verdacht besteht, dass die F1-Werte bei hinreichend detaillierter Auflösung leicht zweigipflig sind. Also wiederholen wir die Analyse getrennt nach Männern und Frauen (ggf.: konvertiere numerische Variable in Faktoren, für Sex, wobei 1 = Mann und 2 = Frau) und gewinnen folgende Werte:

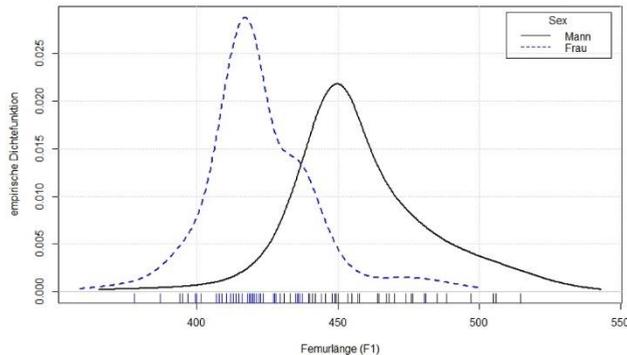
---

<sup>62</sup> Joanes, D. N. & Gill, C. A. (1998). Comparing measures of sample skewness and kurtosis. *The Statistician*, 47, 183-189.

	mean	sd	IQR	skewness	kurtosis	0%	25%	50%	75%	100%
für 38 Männer	457.5658	24.81259	27.500	0.2667551	0.5995957	394	442	454.25	469.500	514.5
für 52 Frauen	422.6058	19.77083	23.125	0.7321193	1.4660634	378	412	419.75	435.125	480.5

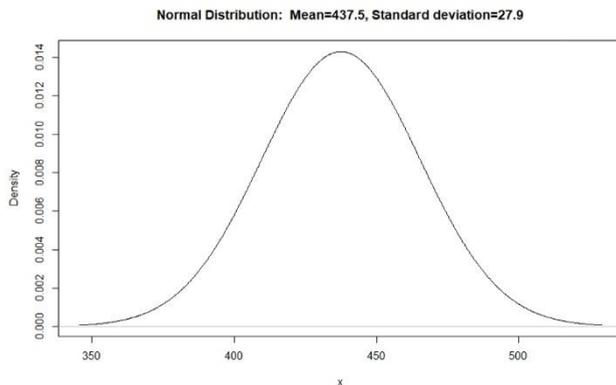
Bei Männern (n=38) wie Frauen (n=52) ist der Median (etwas) kleiner als der Mittelwert und die Schiefe größer null, d. h. die Kurve ist linkssteil. Die Wölbung liegt in beiden Fällen oberhalb von 0; beide Abweichungen von 0 sind bei den Frauen deutlicher: Die Zwentendorfer Männer wie Frauen sind stärker als bei einer Normverteilung auf den Mittelwert hin zentriert, wobei der Effekt bei den Frauen erheblich stärker ist als bei den Männern.

Man kann das sehen, wenn man für Männer und Frau je ein gleich skaliertes Histogramm anliegt (also Start- und Endpunkt gleich, Klassenweite gleich) und diese miteinander vergleicht. Schneller und visuell einleuchtender geht es im R-Commander so: => Grafiken; => Dichteschätzung, dort Variable „F1“ auswählen und als Gruppierungsvariable „Sex“ auswählen, dann OK. Bitte lassen Sie uns hintanstellen, was „Dichteschätzung“ ist (eine kluge Variante von Histogrammen), und konzentrieren wir uns auf die resultierende Darstellung **Abb. 84**: Beide Kurven sind durch einige wenige Individuen geprägt, die besonders groß sind, beide Kurven zeigen einen relativ spitzen Gipfel (spitzer als eine Normalverteilung), und die Kurve der Frauen ist höher (=stärkere Wölbung) als die der Männer.



**Abb. 84** Femurlängen der Zwentendorfer Männer (n=38) und Frauen im Vergleich (n=52); dargestellt ist die jeweilige empirische Dichtefunktion. Die kurzen Striche unten im Diagramm zeigen die einzelnen Fälle an.

Wie sähe denn eine Normalverteilung aus? Wir können uns das mit Hilfe des R-Commanders schnell zeichnen lassen: => Verteilungen; => Stetige Verteilungen; => Normalverteilung; => Grafik der Normalverteilung: In das aufpoppende Menü gibt man nun jene Parameter ein, die man darzustellen wünscht. In unserem Fall wären die Parameter von F1 für die Gesamtpopulation interessant, wie wir sie oben ermittelt hatten: arithmetisches Mittel: 437.5, Standardabweichung 27.9 (ggf. Häkchen an „Grafik der Dichtefunktion“) & OK. Man sieht an **Abb. 85**, dass diese ideale Kurve etwas flacher und breiter verläuft als die F1-Kurven der Männer und der Frauen (**Abb. 84**).



**Abb. 85** Zum Vergleich: Bild einer Normalverteilung mit Mittelwert 437,5 und Std.abw. 27,9.

Mit diesem Wissen und diesen Anschauungsbeispielen nähern wir uns einer Frage, die wir schon von den Häufigkeiten, den Erwartungswerten und dem Chi-Quadratverfahren her kennen: Sind diese Abweichungen, die wir hier konstatieren, noch mit dem Zufall vereinbar, oder weichen sie bereits signifikant vom Idealmodell einer Normalverteilung ab?

### 20.3 IST EINE VARIABLE NORMALVERTEILT?

Zur Beantwortung dieser wichtigen Frage sollte man zunächst die Variable grafisch prüfen. Also ein Histogramm erzeugen und es prüfen: Ist die Verteilung einigermaßen eingipflig (unimodal), symmetrisch etc.

Dann folgt eine erste Datenprüfung anhand der üblichen Kennzahlen für quantitative Skalenniveaus: Mittelwert und Median sollten zusammenfallen bzw. recht nahe beieinander liegen. Schiefe (engl. *skewness*) und Exzess (in R-Sprache: „*kurtosis*“) nahe 0 liegen.

### 20.3.1 Simulierte Daten als Vergleich

Eine über das bereits Gelernte hinausgehend nützliche Möglichkeit ist es, sich einen richtigen Datensatz mit normalverteilten Zufallszahlen zu generieren, den man zu Vergleichszwecken nutzen kann. Der R-Commander macht das wie folgt:

=> Verteilungen; => stetige Verteilungen; => Normalverteilung; => Zufallsstichprobe aus einer Normalverteilung; das aufklappende Menü z. B. für unser Beispiel füllen mit Arithmetisches Mittel: 437.4; Standardabweichung: 27.9; Anzahl der Stichproben (Zeilen): z. B. 500 (d. h. ich bestelle 500 Zufallszahlen); Anzahl der Beobachtungen (Spalten): 1; Häkchen an „arithmetisches Mittel der Stichprobenwerte“. Nach OK erhalten wir einen Datensatz mit 500 normalverteilten Zufallszahlen mit besagten Kennzahlen, den man wiederum untersuchen kann, z. B. die univariaten Statistiken und ein Histogramm davon abrufen kann.<sup>63</sup> Solche simulierten Zufallszahlen bekannter Eigenschaften bieten eine lehrreichere Anschauung im Vergleich mit den echten Daten.

Das grafische Mustern der Daten, die ersten Eindrücke via Mittelwert, Median etc., sowie das Spielen mit Zufallszahlen sind ein erster guter Einstieg. Statistisch weiterführend ist das Vorgehen gemäß dem folgenden Kapitel.

### 20.3.2 Test auf Vorliegen einer Normalverteilung

Sehr ähnlich zum Chi-Quadrat-Verfahren gibt es Verfahren zum Test auf Vorliegen einer Normalverteilung. Und ja, im Hintergrund rechnen sie ähnlich wie das Chi-Quadratverfahren: An der beobachteten Datenserie werden die nötigen Kennzahlen wie Mittelwert und Std.abw. ermittelt, daraus wiederum werden unter Annahme, es liege eine Normalverteilung vor, Erwartungswerte formuliert und diese wiederum mit den beobachteten Werten verglichen. Die Abweichungen beobachtet zu erwartet werden gemessen, aufsummiert und anschließend im Vergleich zu einer Testgröße geprüft, inwieweit diese Abweichungen noch mit dem Zufall vereinbar sind, oder ob diese bereits jenseits des Zufälligen angesiedelt sind. Man benennt diesen Ansatz auch mit dem Überbegriff „goodness-of-fit-Test“ (Anpassungstest).

Für den Test auf Normalverteilung gibt es mehrere, leicht unterschiedliche Tests. Der gerade beschriebene klassische Goodness-of-fit-Test (hier: Chi-Quadrat-Anpassungstest) gilt im vorliegenden Fall als nicht perfekt und nur für große Datenreihen geeignet. Ich empfehle, sofern man nicht gute Gründe für eine andere Wahl hat,

---

<sup>63</sup> Mit der R-Commander Option „Datenmatrix betrachten“ kann man sich diese Datenliste auch anschauen (was allerdings wenig erhellend ist).

den auch für kleine Serien guten und teststarken „Kolmogorov-Smirnov-Anpassungstest“ (kurz: KS-Test, auch „Kolmogoroff-Smirnoff-Test“ geschrieben) resp. den Lilliefors-Test, der eine verbesserte Variante des KS-Tests ist. Diejenigen, die Details nachlesen möchten, seien verwiesen z. B. auf Sachs & Hedderich (2006, 327-347), Bortz & Schuster (2010, 143-146) oder Bortz, Lienert & Boehnke (2008, 319-323). Die Nullhypothese ( $H_0$ ) ist, dass die vermutete Verteilung gegeben ist, die Alternativhypothese ( $H_1$ ) ist, dass eine signifikante Abweichung von der Normalverteilung vorliegt.

### 20.3.3 Praxis: KS-Test auf Vorliegen einer Normalverteilung mit R

Aus guten Gründen, die später deutlicher werden, beginnen wir den Praxisteil mit **R** und dem R-Commander statt wie bisher mit PAST, und stellen den PAST-Teil nach. Im R-Commander wählen wir die Erwachsenen aus Zwentendorf als Datensatz und Testen die Femurlänge F1 aller Individuen auf Normalverteilung: => Statistik; => Deskriptive Statistik; => Test auf Normalverteilung; in dem aufklappenden Menü wählen wir F1 als Variable aus, setzen das Häkchen an „Lilliefors (Kolmogorov-Smirnov)“ & OK. Wir erhalten:

```
Lilliefors (Kolmogorov-Smirnov) normality test
data:  F1
D = 0.080352, p-value = 0.1626
```

Lassen wir die Prüfgröße D beiseite und konzentrieren uns auf den p-Wert, der deutlich größer als 0.05 ausfällt. Die Zwentendorfer Femurlängen sind mit der Hypothese einer Normalverteilung vereinbar, die Abweichungen sind nicht signifikant, d. h. noch mit dem Zufall vereinbar.

Man wiederhole den Test, setze aber neu bei der Schaltfläche „Test für die Gruppen...“ die Variable Sex ein, d. h. teste für Männer und Frauen getrennt. Wir erhalten:

- für die Männer:  $D = 0.13264$ ,  $p\text{-value} = 0.08952$
- für die Frauen:  $D = 0.11752$ ,  $p\text{-value} = 0.07035$

In beiden Fällen nähert sich der Test deutlich der 0.05-Schranke des p-Wertes, liegt jedoch weiterhin darüber.

Wir bilanzieren: Wir haben zuvor die Abweichungen anhand der Kennzahlen und Histogramme recht eng gelesen, tatsächlich sind solche – wie wir jetzt lernen: kleinen – Abweichungen noch mit der Hypothese „Normalverteilung“ vereinbar. Das Ergebnis ist für das weitere Arbeiten über Zwentendorf wichtig, weil man nach diesem Befund – zumindest hinsichtlich der Femurlängen – mit statistischen Verfahren weiterarbeiten

darf, die eine Normalverteilung voraussetzen. Andernfalls müsste man zu non-parametrischen Verfahren wechseln (s. folgenden Abschnitt „Zurückgefragt: wozu der Stress...“).

Um zum Vergleich am vertrauten Material einmal eine schöne (und auch sachlogisch erwartbare) Nicht-Normalverteilung zeigen zu können, kommen wir auf den ursprünglichen Datensatz Zwentendorf zurück, der alle Individuen enthält, also auch die Femora von Kindern und Jugendlichen. Die univariate Statistik gibt folgendes aus:

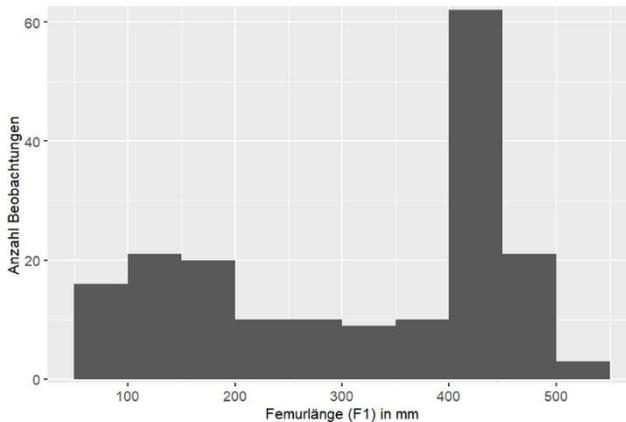
mean	sd	IQR	skewness	kurtosis
313.7088	141.619	273.25	-0.4198742	-1.417936

0%	25%	50%	75%	100%	n
55.5	162.25	390.5	435.5	514.5	182

Bei den insgesamt 182 Individuen ist der Mittelwert mit 313,7 mm erheblich kleiner als der Median (390,5 mm) und die Schiefe ist negativ, d. h. die Verteilung ist „rechtssteil“; die Wölbung weist mit -1,4 einen stark von 0 abweichenden Wert auf, was eine erheblich breitere Streuung als normal um den Mittelwert anzeigt. Das Abrufen eines Histogramms (**Abb. 86**) visualisiert die an den Kennzahlen ablesbaren Tatsachen: Die Femurlängen der Gesamtpopulation sind zwar unimodal, aber stark asymmetrisch und keinesfalls einer Glockenkurve ähnlich. Kommen wir zum Ergebnis eines KS/Lilliefors-Tests auf Normalverteilung für die Gesamtserie:

$D = 0.21463$ ,  $p\text{-value} < 2.2e-16$

Ergebnis: Ein sehr kleiner p-Wert, die beobachtete Verteilung weicht signifikant von der Hypothese einer Normalverteilung ab. Ja, während die Erwachsenen eine in sich halbwegs ähnliche Gruppe bilden, führen die Kinder und Jugendlichen dazu, dass unterschiedlich kleine Individuen in die Betrachtung eingeflossen sind und die gemeinsame Kurve der Femurlängen stark linksschief und rechtsteil ausfällt.



**Abb. 86** Histogramm der Femurlängen (F1, in mm) für die gesamte Population aus Zwentendorf, einschließlich der Kinder und Jugendlichen (n=182, Mittelwert 313.7, Std.abw. 141.6, Median 390.5).

Hm – „ein sehr kleiner p-Wert“? Ja, eine leidige Stolperstelle, hier wie auch immer wieder in anderen Statistiken und Statistikprogrammen: Der R-Commander gibt kleine Werte in der sog. wissenschaftlichen Zahlennotation aus. Die nachfolgende Tabelle **Abb. 87** zeigt ein paar Beispiele, wie unsere übliche Schreibweise von Zahlen in der „wissenschaftlichen Notation“ aussieht. Daran wird schnell deutlich, wie wir die Ausgabe des R-Commander lesen sollten: nach dem Komma folgen 15 Nullen und dann eine 2 – ein wirklich sehr kleiner p-Wert, weshalb wir bei der Abweichung auch von „hochsignifikant“ sprechen können.

im Alltag	wiss. Notation
10.000	1,0e+04
1.000	1,0e+03
100	1,0e+02
10	1,0e+01
1	1,0e+00
0,5	5,0e-01
0,05	5,0e-02
0,01	1,0e-02
0,001	1,0e-03
0,0001	1,0e-04
0,00001	1,0e-05

**Abb. 87** Tabelle mit Zahlen in der üblichen Schreibweise (links) und der sog. wissenschaftlichen Notation (rechts).

Wenn man unsicher ist, wie das Ergebnis eines solchen Tests zu lesen ist, helfen gezielt generierte Zufallszahlen (wie oben erläutert) weiter. Man generiert sich eine größere Menge (z. B. 500 oder 1.000 Fälle) an Zufallszahlen bekannter Eigenschaften (z. B. Normalverteilung, Gleichverteilung etc.) und führt parallel zum eigentlichen Material auch an den Zufallszahlen die gleichen Tests durch. An solchen künstlichen Daten bekannter Eigenschaften lassen sich die Tests und ihre Sensibilität spielerisch besser kennenlernen.

### 20.3.4 Praxis PAST: KS-Test auf Vorliegen einer Normalverteilung

Kommen wir zum Test auf Normalverteilung mit Hilfe von PAST. PAST bietet die o. g. bewährten Standardverfahren nicht an, sondern den Shapiro-Wilk-Test, den Anderson-Darling-Test und den Jarque-Bera-Test. Letzteren bezeichnet der Programmator nur als Referenz zu Vergleichszwecken, während Shapiro-Wilk- oder Anderson-Darling-Test für die Aufgabe besser geeignet seien (Hammer, 2019, 82-84). Im Falle signifikanter Abweichungen von einer Normalverteilung werden p-Werte kleiner 0.05 farbig hinterlegt, um das Test-Ergebnis augenfälliger zu machen. Folgen wir diesen Spuren.

Im Datensatz die zu testende Datenspalte markieren (hellblau hinterlegt); => Univariate; => Normality Tests; nach dem Klick wird sogleich gerechnet; eine neue Fläche klappt auf, wo die Ergebnisse ausgespielt sind. Für unsere Stichprobe ergibt Shapiro-Wilk einen p-Wert von 0.02238, Anderson-Darling einen p-Wert von 0.03374, in beiden Fällen also kleiner 0.05. Beide Tests lehnen also die Hypothese Normalverteilung ab.

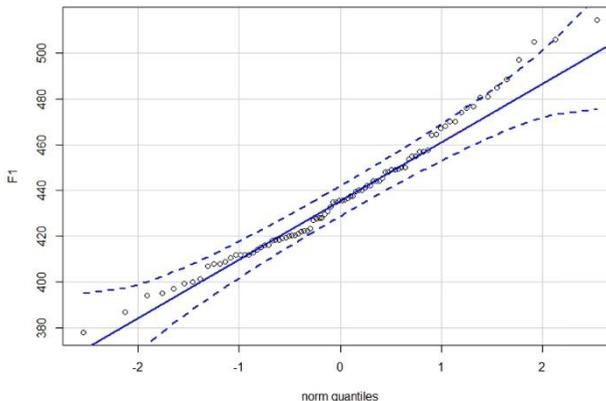
Ergebnis? Unterschiedliche Test ergeben unterschiedliche Ergebnisse und bei grenzfälligen Wertereihen (was im vorliegenden Fall nicht gegeben ist), die knapp diesseits oder jenseits einer Signifikanz liegen, kann das Resultat dann anders ausfallen. Für uns bedeutet dies zweierlei: (a) Shapiro-Wilk-Test und Anderson-Darling-Test sind sensibler gegenüber den Verletzungen der Normalverteilung als die von mir präferierte Lilliefors-Variante des KS-Tests; (b) es reicht nie, zu schreiben, man habe auf Vorliegen einer Normalverteilung getestet, sondern es ist stets exakt anzugeben, welcher Test genau durchgeführt wurde.

### 20.3.5 QQ-Plot – wie sehen die Abweichungen von der Normalverteilung aus?

Eine hilfreiche Visualisierung dessen, was wir bislang mit Kennzahlen gemessen haben, nämlich die Abweichungen einer beobachteten Wertereihe von einer normalverteilten Wertereihe, sind die sog. QQ-Plots. Da diese Streuungsdiagramme schneller gemacht als erklärt sind, erzeugen wir ein solches Bild mit PAST und **R**.

*Der Weg mit PAST:* Spalte mit der zu untersuchenden Wertereihe markieren (hellblau unterlegt); => Plot; => Normal probability plot. Das aufklappende Streuungsdiagramm zeigt als rote Linie den Verlauf, den eine ideal normalverteilte Variable hätte (mit den Parametern Mittelwert und Std.abw. der beobachteten Wertereihe), und in schwarzen Punkten drumherum, wie sich die beobachteten Werte dazu verhalten.

Im Falle unserer Zwentendorfer Femora sehen wir, dass die langen Femora über ca. 490 mm und die kurzen Femora unter etwa 405 mm stärker abweichen, während die Masse der Beobachtungen in dem weiten Bereich dazwischen recht nah am Ideal liegt.



**Abb. 88** Mit **R** gezeichneter QQ-Plot der Femurlängen für Zwentendorf (nur Erwachsene, n=57).

*Der Weg mit dem R-Commander:* => Grafiken; => QQ-Diagramm; im aufklappenden Fenster die darzustellende Variable herausuchen, hier „F1“ & OK. (Die Möglichkeiten hinter dem Reiter „Optionen“ brauchen wir noch nicht). (**Abb. 88**).

Im Unterschied zu PAST werden hier als zwei leicht gebogene Linien zusätzlich die Bereiche eingeschlielt, die bei Vorliegen einer Normalverteilung innerhalb der Zufallsstreuung liegen. Etwas deutlicher erkennen wir in dieser Darstellung, dass es

vor allem einige langen Femora sind (ab ca. 470 mm), die knapp an der Grenze der Normalverteilung resp. knapp jenseits dieser liegen.

Der analytische Blick anhand der QQ-Diagramme ist exakter als das Einspielen der Kurve einer Normalverteilung in das Histogramm, wie es ebenfalls möglich ist. Ich rate von Letzterem ab; wenn die Information wirklich benötigt wird, sollte man zu den QQ-Diagrammen greifen.

### 20.3.6 Option Variablen-Transformationen

Oft liegen Abweichungen von der Normalverteilung vor, die mit Hilfe geeigneter Transformationen in eine Normalverteilung überführt werden können – was der Grund dafür ist, dass wir das Thema Transformation und Rücktransformation rein technisch oben um **Abb. 22** herum bereits eingeführt hatten. Die Schätz- und Prüfverfahren werden dann auf die transformierten Variablen angewendet, z. B. die Frage, ob zwei Serien aus einer gemeinsamen Grundgesamtheit mit gleichem Mittelwert stammen, oder ob signifikante Mittelwertunterschiede vorliegen.

Die an den transformierten Variablen gewonnenen statistischen Aussagen wollen anschließend wieder in die Archäologie zurückübersetzt werden. Der Anschaulichkeit halber ist es zumeist sinnvoll, die Transformation „zurück zu rechnen“, d. h. eine Rücktransformation vorzunehmen. Beispiel: Wir haben eine Wurzel-Transformation durchgeführt, Statistiken gewonnen, und führen nun die geeignete Rücktransformation durch: quadrieren. Die an den transformierten Variablen, d. h. vor der Rücktransformation gemachten Signifikanz-Aussagen gelten auch für die ursprünglichen Variablen; allerdings sind die durch Rücktransformation erhaltenen Mittelwerte und Varianzen meist leicht verfälscht (Ihm 1978, 239).

Archäologisches Beispiel: Das Fassungsvermögen von Tongefäßen, gemessen in  $\text{cm}^2$ . Nach Versuchen stellen wir fest, dass eine Wurzeltransformation zu besser geeigneten Variablen führt. So wird aus  $5.000 \text{ cm}^2$  beobachtet transformiert (ca.) „70“. In diesem Sinne stelle ich fest, dass meine gesamte Beobachtungsreihe z. B. den (transformierten) Mittelwert 65 und die Std.abw. 26 hat. Um die Verbindung mit den realen Beobachtungen wieder zu gewinnen, führe ich abschließend eine geeignete Rücktransformation durch. In unserem Fall  $65 \times 65$  („quadrieren“ als Umkehrung des Wurzelziehens) und gewinne  $4.225 \text{ ccm}$  als Mittelwert, und  $26 \times 26 = 676 \text{ ccm}$  als Std.abw.

Flächen, Gewichte, Volumina (und kleine gezählte Häufigkeiten) gehen oft nach einer Quadratwurzel-Transformation (oder einer Kubikwurzel-Transformation) in eine Normalverteilung über; alternativ probiere man ggf. auch die stärker wirkende logarithmische Transformation ( $I_n$  „logarithmus naturalis“ oder  $\log_{10}$ , „Logarithmus

zur Basis 10<sup>64</sup>). Für gezählte Häufigkeiten (Poisson-Verteilung) werden oft spezielle Quadratwurzeltransformation verwandt:  $\sqrt{x} + \sqrt{(x+1)}$ ; oder:  $\sqrt{(x + 3/8)}$ .<sup>64</sup>

Rechts flach auslaufende Verteilungen (also mit wenigen relativ hohen Werten), die links durch den Wert Null begrenzt sind, gehen häufig nach einer logarithmischen Transformation in eine Normalverteilung über. Da der Logarithmus 0 nicht definiert ist, sollte vorher (wenn der Wert 0 als Beobachtung im Datensatz vorkommt bzw. erlaubt sein soll) eine kleine Konstante wie z. B. 3/8 zu allen Werten addiert werden. Bei sehr großen Zahlen ist auch die doppelt logarithmische Transformation zu erwägen.

Stark rechtsgipflige Verteilungen (wie z. B. die Prägejahre bei den Münzen) können mit einer Potenztransformation normalisiert werden (ggf. auch Winkeltransformation); je nach Stärke der Rechtsgipfligkeit wählt man  $x1,5$  oder  $x2$ .

Relative Häufigkeiten (Binomialverteilung) lassen sich durch die Winkeltransformation (auch: „Arcus-Sinus-Transformation“, inverse Sinus-Transformation) normalisieren:

transformierte Variable =  $\arcsin\sqrt{x}$ , bzw. transformierte Variable =  $\arcsin\sqrt{(x + 3/8)}$ .

Dabei werden die relativen Häufigkeiten zwischen 0 und 1 in Altgrad zwischen 0 und 90 umgewandelt:

<b>Arcus-Sinus-Transformation</b>					
relative Häufigkeit	0,00	0,25	0,50	0,75	1,00
(Alt-) Grad	0	30	45	60	90

**Abb. 89** Ausgewählte typische Wertepaare bei einer Arcus-Sinus-Transformation.

## 20.4 ZURÜCKGEFRAGT: UND WOZU DER STRESS MIT DER NORMALVERTEILUNG?

Die Normalverteilung ist eine Idee, ein Ideal, mit dem viele statische Prozeduren operieren. Man nennt diesen Bereich der Statistik auch „parametrische Statistik“. In der Archäologie sind normalverteilte Variablen eher selten, in der Biologie hingegen nicht:

<sup>64</sup> Dieses „plus 3/8“ ist ein Statistikertrick: Da „Wurzel aus Null“ nicht definiert ist und umgangen werden möchte, addiert man einfach 3/8 zu allen Werten dazu. Hinter „3/8“ steckt kein tiefes Geheimnis, man hat sich einfach drauf geeinigt, dass „man das so macht“.

Viele Resultate eines natürlichen Wachstums unterliegen einer Normalverteilung resp. können mit kleinen Transformationen dorthin überführt werden. Die Anthropologen unter den Lesern werden folglich häufiger mit normalverteilten Variablen (und parametrischer Statistik) arbeiten, die Archäologen unter den Lesern hingegen seltener.

Die Antwort auf die Frage „normalverteilt ja oder nein?“ hat zwei unterschiedliche Bedeutungen, von denen die wichtigere statistik-interner Natur ist. An der Antwort hängt nämlich die Entscheidung, welchen weiteren Weg eine Untersuchung geht. Normalverteilung heißt, dass man mit Mittelwert und Standardabweichung operiert, als Darstellung der Variablen Histogramme verwendet, sich bei Tests aus dem Fundus der sog. „parametrischen Verfahren“ bedient. Nicht-normalverteilt heißt, dass Median und IQR die angemesseneren Kennzahlen sind, man die Histogramme häufig durch Boxplots ersetzt und sich bei Tests aus dem Fundus der sog. „nicht-parametrischen Verfahren“ bedient. Letztere sind nicht schlechter als die parametrischen Verfahren, aber eben anders. Daher sollte bei vielen Untersuchungen metrischer Variablen die Frage nach der Verteilung früh gestellt und klar beantwortet werden, damit man für das Weitere in den richtigen Werkzeugkasten greift. Es wie mit Nägeln, Schlitzschrauben und Kreuzschrauben: Drei berechnete und nützliche, doch unterschiedliche Konzepte, und jedes für sich eine eigene Welt, denn Hammer und Schrauben passen ebenso wenig zueinander wie Schlitzschrauben und Kreuzschraubenzieher. Wir erinnern uns zum Thema Chi-Quadrat-Test und unserem Experiment mit beobachteten Häufigkeiten, Prozentwerten und Promillewerten daran was geschieht, wenn ein statistisches Verfahren angewendet wird, obwohl dessen verteilungstheoretische Voraussetzungen nicht erfüllt sind.

Selbstverständlich hat die Aussage „normalverteilt“ resp. die Abweichung davon auch eine inhaltliche Bedeutung. Wenn ein klares Konzept mit den üblichen kleinen Zufallsabweichungen realisiert wird, entstehen oft Normalverteilungen, gleich, ob es sich um eine biologische Art und deren Größenwerte handelt oder Artefakte wie beispielsweise das ideale Steinbeil zum Baumfällen. Wenn Objektgruppen dann stärker „als normal“ auf einen Mittelwert hin normiert sind (Wölbung), ist dies ein spannender Befund, ebenso, wenn ein Ideal zwar besteht, aber die Normierung schwächer ausgeprägt ist. Insofern geht die Deutbarkeit des Befundes „Normalverteilung ja/nein“ durchaus über die Binnensphäre der Statistik hinaus.

#### **20.4.1 Vorab: Prozentwerte und Verhältniszahlen sind nicht normalverteilt**

Bevor man nun seine Prozentzahlen auf Normalverteilung testet: Prozentwerte können nicht normalverteilt sein. Sie bewegen sich zwischen 0 und 100. Bei dem theore-

tischen Modell einer Normalverteilung sind die zufälligen Abweichungen vom Mittelwert symmetrisch, d. h. in beide Richtungen gleichermaßen möglich und wahrscheinlich. Das ist bei Prozentwerten naturgemäß nicht der Fall. Je näher der mittlere Wert an 0 % oder an 100 % heranreicht, desto „schiefer“ wird die Verteilung zufälliger Abweichungen, weil Prozente keinen geringeren Wert als 0 oder höheren Wert als 100 einnehmen können, während eine Streuung in die jeweils andere Richtung sehr wohl möglich ist.

Folglich sind auch Indizes alias Verhältniszahlen nicht normalverteilt. Denn bei der Berechnung – beispielsweise – einer relativen Bauchhöhe an Gefäßen (z. B. „Bauchhöhe / Gesamthöhe \* 100“) erfolgt, statistisch gesehen, der gleiche Bezug auf eine fixe Bezugsgröße wie bei Prozentzahlen, und es resultieren Verteilungen, die notwendigerweise nahe der Extreme zunehmend schief, d. h. unsymmetrisch werden. Das ist der Grund, weshalb die für unser Vorstellungsvermögen bisweilen so hilfreichen Verhältniszahlen bei Statistikern eher unbeliebt sind: Sie können nicht – zumindest nicht ohne weitere Vorkehrungen – in all jene weitergehenden Verfahren eingebracht werden, die Normalverteilung voraussetzen.

## **21 DER VERGLEICH VON MITTELWERTEN: VORÜBERLEGUNGEN UND AUSWAHL DES GEEIGNETEN TESTS**

---

Will man von verschiedenen Stichproben überprüfen, ob sie aus einer gemeinsamen Grundgesamtheit stammen – sich also nur im Rahmen des Zufalls ein wenig unterscheiden –, oder ob sie aus unterschiedlichen Grundgesamtheiten stammen – sich also über-zufällig stark voneinander unterscheiden –, stehen verschiedene Testverfahren zur Verfügung. Die Wahl des Verfahrens ist von der Art der Daten und des Problems abhängig. Folgende Begriffe sind bei der Auswahl geeigneter Testverfahren wichtig und müssen vorab geklärt werden:

### **21.1.1 parametrische / nicht-parametrische Verfahren**

Liegt Normalverteilung bzw. eine annähernde Normalverteilung vor, wählt man „parametrische Verfahren“; sie testen auf Mittelwert- (und ggf. auch Varianz-) Unterschiede. Liegt keine Normalverteilung vor, wählt man „nicht-parametrische Verfahren“; sie beruhen auf der Überführung der Beobachtungen in Ränge und testen in der

Regel auf Median-Unterschiede. Nicht-parametrische Verfahren können auch auf ordinalskalierte Variablen angewendet werden.

Was heißt „Überführung der Beobachtungen in Ränge“? Hier am Beispiel der Körperhöhe von Kursteilnehmern: Die beobachteten Daten waren 160, 165, 170, 172, 172, 173, 175, 175, 178, 181, 184, 185, 186, 187 cm. Daraus werden: 1., 2., 3., 4., 5., 6., 7., 8., 9., 10., 11., 12., 13., 14. Der Fall ursprünglich „184 cm“ hat nun den Rang 11. Die nicht-parametrischen Tests beziehen sich auf diese Ränge.

### **21.1.2 unabhängige Stichproben / abhängige Stichproben**

Abhängige Stichproben liegen vor, wenn wir zwei Beobachtungen am gleichen Objekt vergleichen wollen. Ein typisches Beispiel sind Wiederholungsmessungen, etwa eine „Vorher-nachher-Untersuchung“ beim Vergleich von Konservierungsmethoden an Holzobjekten; ebenfalls in diese Kategorie fallen Vergleiche rechter und linker Knochenmaße vom jeweils gleichen Individuum. Unabhängige Stichproben hingegen vergleichen Beobachtungen an unterschiedlichen Fällen.

### **21.1.3 einseitige oder zweiseitige Fragestellung**

Bei der Anlage der Tests und der Deutung der Ergebnisse müssen wir bisweilen unterscheiden zwischen „einseitiger Fragestellung“ und „zweiseitiger Fragestellung“ (auch: „ein- oder zweiseitige Signifikanz“): Erwarten wir Unterschiede zwischen den Mittelwerten ausschließlich in einer Richtung (einseitige Fragestellung), oder sind Unterschiede in beide Richtungen möglich (sowohl größer als auch kleiner, d. h. zweiseitige Fragestellung)? Die zweiseitige Fragestellung ist der übliche Fall. Daher sollte die Auswahl der einseitigen Fragestellung gut überlegt und stets explizit begründet sein.

### **21.1.4 $n = 2$ oder $n > 2$**

Üblicherweise werden für den Vergleich zweier Stichproben andere Verfahren gewählt als für den simultanen Vergleich mehrerer Stichproben.

Nach Beantwortung dieser Fragen kann das geeignete Testverfahren nach dem in **Abb. 90** zusammengefassten Raster ausgewählt werden.

welcher Test?	parametrisch (normalverteilt)	nicht-parametrisch
unabhängige Stichproben	n=2: t-Test f. unabhängige Stichproben	n=2: U-Test Wilcoxon, Mann & Whitney
	n>2: einfache Varianzanalyse (ANOVA)	n>2: Rangvarianzanalyse n. Kruskal & Wallis
abhängige Stichproben	n=2: t-Test für abhängige Stichproben	n=2: Wilcoxon-Test f. Paardifferenzen
	n>2: doppelte Varianzanalyse	n>2: Friedman-Test

**Abb. 90** Tabelle zur Auswahl des geeigneten Test-Verfahrens auf Mittelwertunterschiede <sup>(65)</sup>.

## 21.2 TESTVERFAHREN AUF MITTELWERTUNTERSCHIEDE

### 21.2.1 t-Test

Voraussetzung: (annähernde) Normalverteilung; Ziel: Vergleich von zwei unabhängigen Stichproben.

Ausgegeben werden Mittelwert und Standardabweichung pro Gruppe, der t-Wert (i. e. die Prüfgröße dieses Tests), die Freiheitsgrade und die entsprechende Signifikanz. Kleinere t-Werte und Signifikanzen von 1 bis ca. 0.05 bedeuten, dass keine signifikanten Mittelwertunterschiede vorliegen. Größere t-Werte und Signifikanzen kleiner 0.05 (ideal: 0.000) bedeuten, dass signifikante Mittelwertunterschiede vorliegen.

Bei Mittelwert-Ähnlichkeit ist denkbar, dass sich die Varianzen der Gruppen durchaus signifikant unterscheiden. Auch darüber informiert der t-Test (hier: Levene-Test). Liegt der F-Wert (i. e. die Prüfgröße des Levene-Tests) nahe bei 1, ist auch der Unterschied der Varianzen gering; der zusätzlich dazu ausgegebene Signifikanzwert ist

---

<sup>65</sup> Das Kleingedruckte zu **Abb. 90**: Varianzanalysen/multiple Tests, also Test mit  $n > 2$ , sind ein komplexes Thema. Denn es ist zu unterscheiden, ob letztlich paarweise Unterschiede getestet werden (können), was meist zu klareren Ergebnissen führt, ob tatsächlich die Testfrage global an alle Stichproben gleichzeitig gestellt wird, oder ob gegen eine Kontrolle resp. einen Standard getestet werden kann. Als Literatur ziehe man z. B. heran: Horn, Manfred & Vollandt, Rüdiger (1995). *Multiple Tests und Auswahlverfahren*. Stuttgart: G. Fischer; vgl. Lozán, José L. & Hartmut Kausch, Hartmut (2007). *Angewandte Statistik für Naturwissenschaftler*. 4. Aufl. Hamburg: Wiss. Auswertungen.

größer 0.05 (bis max. 1). Ist der F-Wert groß und ist der ausgegebene Signifikanzwert kleiner 0.05 (bis min. 0.000), liegen deutliche Unterschiede der Varianzen vor.

### **21.2.2 einfache / einfaktorielle Varianzanalyse**

Voraussetzung: (annähernde) Normalverteilung; Ziel: Vergleich von drei und mehr unabhängigen Stichproben.

Lesen der Ausgabe wie zuvor beim t-Test: große F-Werte und Signifikanz kleiner 0.05 zeigen deutliche Unterschiede an.

### **21.2.3 U-Test (nach Wilcoxon, Mann & Whitney)**

Voraussetzung: Variable mindestens ordinalskaliert, annähernd unimodal, nicht normalverteilt; Ziel: Vergleich von zwei unabhängigen Stichproben.

Es handelt sich um einen Rangtest: Die tatsächlichen Messwerte werden durch Ränge ersetzt, getestet wird der Median-Unterschied der Ränge. Voraussetzung ist, dass die zu vergleichenden Verteilungen annähernd formgleich sind. Der Test setzt keine Normalverteilung voraus, es ist aber auch nicht schädlich, wenn die zu vergleichenden Daten normalverteilt sind. Ausgegeben wird der mittlere Rang je Gruppe, anschließend die Testgröße U, aus der sich die Prüfgröße z ergibt, dazu die Signifikanz von z. Größere Werte von z und Signifikanz nahe 0 bis maximal 0.05 zeigen deutliche Unterschiede an; kleine Werte von z und Signifikanz zwischen 1 und 0.05 zeigen, dass es kaum Unterschiede gibt. Der U-Test gilt als sehr effizienter Test. Deshalb sollte er statt des t-Tests stets verwendet werden, sobald auch nur geringe Zweifel am Vorliegen einer Normalverteilung bestehen.

### **21.2.4 Kolmogorov-Smirnov-Test für zwei Stichproben**

Voraussetzung: Variable mindestens ordinalskaliert, annähernd unimodal, nicht normalverteilt; Ziel: Vergleich von zwei unabhängigen Stichproben.

Der KS-Test ist dem U-Test ähnlich, berücksichtigt jedoch Unterschiede nicht nur im Zentralmaß, sondern auch in der Streuung und der Schiefe der Verteilung. Der KS-Test gilt wie der U-Test als sehr effizient. Die Testtheorie empfiehlt, bei kleinen Stichproben (n kleiner 20) den KS-Test, bei großen Stichproben (n grösser 40) eher den U-Test anzuwenden.

Für spezielle Anwendungen vgl. auch den Wald-Wolfowitz-Test und den Moses-Test. Als weiterführende Literatur zum Thema nenne ich den „Klassiker“: Siegel (1987); wer's moderner und statistischer braucht, zieht Bortz, Lienert & Boehnke (2008) zu Rate.

### 21.2.5 Rangvarianzanalyse nach Kruskal & Wallis (auch: H-Test)

Voraussetzung: Variablen mindestens ordinalskaliert, annähernd unimodal, nicht normalverteilt; Ziel: Vergleich von drei und mehr unabhängigen Stichproben.

Es handelt sich um einen Rangtest. Er gilt als sehr geeignetes Verfahren bei nicht-normalverteilten Daten, effizienter als der ebenfalls mögliche Median-Test.

### 21.2.6 Praxis R: Mittelwertvergleich zweier Stichproben

Die zuvor dargelegten diversen Optionen, zwei oder mehrere Datenserien bzw. Gruppen zu vergleichen, sollen nun an einigen konkreten Anwendungen umgesetzt werden. Als erster Fall sollen die Femora (F1) der erwachsenen Zwentendorfer Frauen und Männer geprüft werden (d. h. filtern mit „Age >3“). Wir erinnern uns, im Mittel waren die Femora der Frauen um etwa 3,5 cm kürzer als die der Männer, wobei die Std.abw. bei jeweils um 2 cm liegt. Wir hatten zuvor festgestellt, dass diese Daten annähernd normalverteilt sind; es handelt sich um zwei Gruppen und um voneinander unabhängige Beobachtungen. Demnach liegt ein t-Test nahe. Zuvor haben wir darauf geachtet, dass das Geschlecht „Sex“ ein „Faktor“ ist resp. die Variable zum Faktor konvertiert. Sicherheitshalber haben wir uns das zu untersuchende Thema noch schnell in einer Grafik angeschaut, z.B. => Grafiken; => Histogramm; in dem kleinen Menü als Variable „F1“ ausgesucht, bei „Grafik für die Gruppen“ die Variable „Sex“ gewählt, dann OK & Enter. Bei starker Überlappung der beiden Wertebereiche erkennt man unterschiedliche „Peaks“ für Männer (um 450 mm) und Frauen (um 420 mm).

Der eigentliche t-Test ist im R-Commander schnell umgesetzt: => Statistik; => Mittelwerte vergleichen; => t-Test für unabhängige Stichproben...; im neuen Menü die Gruppierungsvariable auswählen (Sex) und „abhängige Variable“ (F1) & OK. Die wesentliche Ausgabezeile besagt:

$$t = 7.1784, \text{ df} = 68.587, \text{ p-value} = 6.494\text{e-}10$$

Der t-Wert (d. h. die Prüfgröße) ist hoch und der resultierende p-Wert liegt deutlich unter 0.05 (signifikant), sogar unter 0.001, so, dass wir von einem hochsignifikanten Unterschied zwischen beiden Gruppen sprechen können.

Möglicherweise ist man bei den „Optionen“ an der Frage „Von gleichen Varianzen ausgehen?“ „ja/nein“ hängen geblieben. Ja, wenn die Varianzen beider Gruppen erheblich unterschiedlich sind (*in praxi*: beide Gruppen eine deutlich unterschiedliche Standardabweichung aufweisen), oder beide Gruppen sehr unterschiedlich viele Fälle umfassen und die untersuchte Population insgesamt eher klein ist, bedarf es dafür einer Korrektur, um zu validen t-Werten zu gelangen. Das ist der sog. „Welch-t-Test“, den **R** per Voreinstellung ausführt, d. h. die nötige Korrektur vornimmt.

Bleiben wir bei dem Beispiel und führen noch einen „abhängigen“ t-Test durch. „Abhängig“ wären z. B. die jeweils rechte und linke Körperseite eines Individuums, denn Menschen, die ein recht langes rechtes Bein haben, dürften auch ein recht langes linkes Bein haben. Schauen wir uns also die Frage rechter vs. linker Femur an. Die einfache Statistik gibt uns aus, dass in Zwentendorf bei den Erwachsenen der Mittelwert rechts bei ca. 441 mm liegt, links bei 437 mm. Ist der Unterschied von ca. 4,6 mm signifikant? Lösung: => Statistik; => Mittelwerte vergleichen; => t-Test für gepaarte Stichproben; in dem Menü F1re und F1li aussuchen, OK & Enter. Wir erhalten:

```
Paired t-test
data: F1re and F1li
t = -1.2351, df = 47, p-value = 0.2229
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.4239449  0.3406116
sample estimates:
mean of the differences
 -0.5416667
```

In den Test gehen naheliegenderweise nur die Individuen ein, für die beide Messungen vorliegen. Unten in der Ausgabe lesen wir, dass bei ihnen der mittlere Unterschied beider Körperseiten bei 0.541667 mm liegt. Oben lesen wir, dass der t-Wert relativ klein ist und der p-Wert deutlich größer 0.05 ist, d. h. der (kleine) Unterschied ist statistisch nicht signifikant.

Um einmal einen Fall mit mehr als zwei Gruppen durchzuspielen, vergleichen wir die Femora der drei erwachsenen Altersgruppen adult, matur und senil miteinander und denken (eher im Scherz) an den Spruch „im Alter wird man wieder kleiner“. Vorbereitend muss die Kodierung wieder zum „Faktor“ gemacht werden. Im R-Commander also: => Datenmanagement; => Variablen bearbeiten; => Konvertiere numerische Variablen in Faktoren...; in der neuen Menüfläche wählen wir „Age“, wählen einen neuen Variablennamen wie z. B. Altersgruppe und machen 4 zu adult, 5 zu matur und 6 zu senil. Die Umwandlung zum Faktor ist hier nötig, damit „Age“ als Gruppierungsvariable verwendet werden kann.

Nun => Statistik; => Mittelwerte vergleichen; => Einfaktorielle (alias „einfache“) Varianzanalyse: Wir wählen „Altersgruppe“ als Gruppierungsvariable, „F1“ als abhängige Variable. Im Ergebnisfenster lesen wir:

	mean	sd	data:n	data:NA
adult	433.7576	31.64405	33	95
matur	442.9457	26.66442	46	55
senil	424.8636	13.89261	11	9

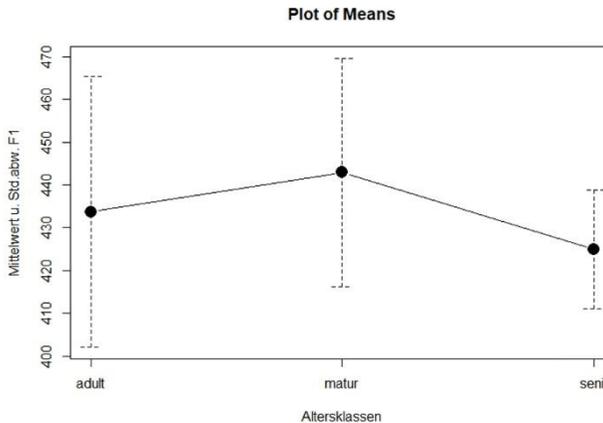
Die ausgegebenen Mittelwerte liegen nah beieinander, insbesondere, wenn man die jeweilige Standardabweichung von um 14–32 mm beachtet. Der Test ergibt folgendes Resultat:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Altersgruppe	2	3581	1790.6	2.361	0.1
Residuals	87	65968	758.2		

159 observations deleted due to missingness.

Der F-Wert, d. h. die Prüfgröße ist nicht groß, der p-Wert wird mit 0.1 angegeben: Die Unterschiede zwischen den drei Gruppen sind nicht signifikant.

Wir schauen uns parallel dazu die Daten als Grafik an: => Grafiken; => Plot für arithmetische Mittel...; in das neue Fenster setzen wir „Altersgruppe“ als Gruppierungsvariable ein und „F1“ als abhängige Variable. Hinter dem Reiter „Optionen“ wird beim „Fehlerbalken“ das Häkchen neu auf „Standardabweichungen“ gesetzt, dann mit OK die Statistik abgerufen.



**Abb. 91** Vergleich der Mittelwerte und Standardabweichungen der Femurlängen (in mm) für die drei Altersgruppen der Erwachsenen in Zwentendorf (n=90). Die kräftigen Punkte zeigen die Lage des Mittelwerts an, die gestrichelten Linien die Standardabweichung; deren Spannen überlappen einander deutlich.

Die Grafik macht sehr anschaulich (**Abb. 91**), was schon die sorgfältige Lektüre der Kennzahlen Mittelwert und Standardabweichung der drei Altersgruppen nahelegte und der Test bestätigt hat: Die drei Gruppen unterscheiden einander nicht signifikant,

vielmehr überlappen sich die Wertereihen stark. Ein in der Sache nicht überraschendes Ergebnis, aber es ging hier ja auch mehr um das Training des Verfahrens, weniger um tiefe historische Einsichten.

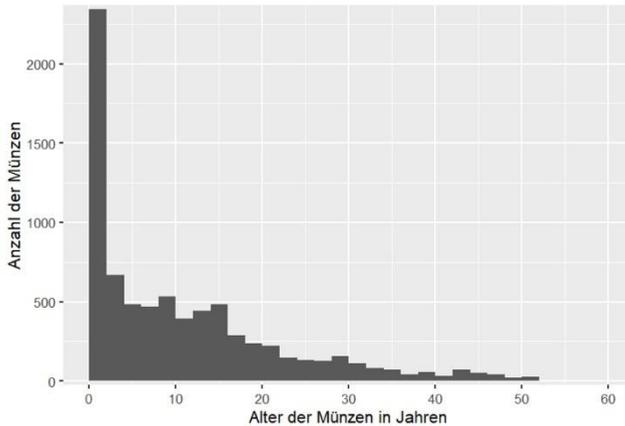
Übrigens: Wir argumentieren hier wegen gegebener Normalverteilung mit parametrischen Verfahren und wählen deshalb für die Darstellung **Abb. 91** als „Plot of Means“ auch Mittelwert und Standardabweichung. Im Falle einer nicht gegebenen Normalverteilung wären andere Verfahren zu wählen und auch andere Grafiken, z. B. Boxplots. Die „Antennen“ in **Abb. 91** bedeuten etwas anderes als die Antennen der Boxplots.

Ein kleines Caveat ist notwendig. Der Vergleich von mehr als zwei Mittelwerten leitet in das Thema ANOVA „Analyse of Variance“ über, was weitaus komplexer ist, u. a. da Varianzanalysen wiederum zu prüfende Voraussetzungen haben und auch unterschiedliche Auslegungen haben können, etwa, wenn man lineare Effekte zwischen den Gruppen vermutet – in unserem Beispiel etwa im Sinne: die Menschen werden mit zunehmendem Alter immer größer resp. immer kleiner. Wer also Probleme der hier skizzierten Art näher untersucht, kommt um einen Blick in vertiefende Literatur nicht herum.

Als zweites Beispiel greifen wir den Datensatz „Geldbörsen“ wieder auf und fragen nach dem Alter der Münzen, also der Differenz zwischen Erhebungsjahr und Prägejahr. Um die Analyse etwas zu fokussieren, beschränken wir uns auf Münzen der drei Währungen DM, CHF und EUR. Als ersten Schritt verschafft man sich einen Überblick über die Gesamtheit, d. h. die gängigen Kennzahlen und ein Histogramm.

```
mean      sd skewness kurtosis 0% 25% 50% 75% 100%    n
11.2304 12.0035 1.47388 2.781374 0 1 8 17 104 7756
```

Der Mittelwert liegt bei gut elf Jahren, die Std.abw. bei zwölf Jahren, der Median ist mit acht Jahren deutlich kleiner, Wölbung und Schiefe weisen relativ hohe, positive Werte auf. All das spricht für eine stark schiefe, linkssteile Verteilung. Das Histogramm **Abb. 92** bestätigt dies: Die meisten Münzen sind nur wenige Jahre alt, aber offenbar enthält der Datensatz auch einige wenige Münzen von mehr als 30 Jahren Alter. Bei einer so schiefen Verteilung sind Mittelwert und Std.abw. keine adäquaten Kennzahlen, weshalb wir auf den Median (8 Jahre) und die Q25–Q75-Spanne als Streuungsmaß zurückgreifen: 50 % aller Münzen sind zwischen 1 und 17 Jahren alt.



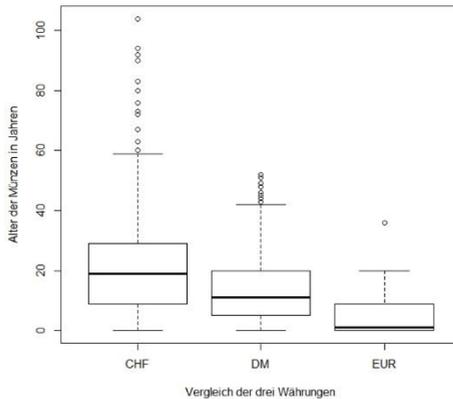
**Abb. 92** Alter der Münzen in Jahren aus unserer Serie „Geldbörsen“, begrenzt auf die Währungen DM, CHF und EUR und Münzen bis 60 Jahre (n=7.742).

Für eine vertiefende Analyse ist nun die Frage nach dem weiteren Vorgehen zu stellen: parametrisch oder nicht-parametrisch. Man könnte versuchen, z. B. durch eine Wurzeltransformation oder (besser, weil stärker) eine logarithmische Transformation die Alterswerte „zu normalisieren“. Von solchen Transformationen wären niedrige Alter wenig, hohe Alter zunehmend stark betroffen, d. h. die hohen Werte würden stauend an die niedrigen Werte herangerückt. Doch diese Idee eines weiteren Vorgehens befriedigt nicht, weil den transformierten Werten die Anschaulichkeit fehlen würde. Daher entscheiden wir, die untransformierten Altersangaben beizubehalten und mit nicht-parametrischen Verfahren fortzufahren. In diesem Sinne ein Vergleich der drei Währungen:

	mean	0%	25%	50%	75%	100%	Alter :n
CHF	19.808368	0	9	19	29	104	1936
DM	14.446799	0	5	11	20	52	2218
EUR	4.639367	0	0	1	9	36	3602

Der Median des Euro liegt bei 1 Jahr, 50 % der Euro-Münzen sind zwischen 0 und 9 Jahre alt: Klar, das spiegelt unsere Währungsgeschichte wider; der Euro ist jung, im Vergleich zu DM und CHF kann es gar keine wirklich alten EUR-Münzen geben. Interessant allerdings, dass auch zwischen DM und CHF ein Unterschied andeutet, denn nach allen Kennzahlen sind die CHF-Münzen deutlich älter als die DM-Münzen, also länger im Umlauf. Als adäquate Grafik bietet sich ein Boxplot an, resp. eine Folge von

drei Boxplots. Also R-Commander, => Grafiken; => Boxplots; in dem Menü „Alter“ als Variable aussuchen, bei der Schaltfläche „Grafik für die Gruppen...“ Währung aussuchen, OK; beim Reiter „Optionen“ zu „Identifiziere Ausreißer“ noch das Häkchen auf „Nein“ setzen und eine Achsenbeschriftung eingeben: x-Achse „Vergleich der drei Währungen“ und y-Achse „Alter der Münzen in Jahren“; OK (Abb. 93).



**Abb. 93** Vergleich des üblichen Alters der Münzen bei den Währungen CHF, DM und Euro in einem Boxplot (n=7.756).

**Abb. 93** verdeutlicht, dass die Währung CHF mehr noch als DM durch einzelne besonders alte Münzen geprägt ist, doch die Lage der Boxen – die ja die 50 %-Spanne abbilden – zeigt an, dass auch diesseits der Ausreißer die Masse der CHF-Münzen etwas älter ist. Wir wollen diesem ersten Eindruck von den Daten mit einem statistischen Test nachgehen. Die Sonderstellung des Euro ist offensichtlich, das bedarf keines Tests. Angesicht der Verteilung greifen wir also zu einem nicht-parametrischen Test (alias Rangtest) für die beiden Gruppen CHF und DM. Die einzelnen Schritte mit dem R-Commander:

Daten filtern: => Datenmanagement; => Aktive Datenmatrix; => Teilmenge der aktiven Datenmatrix; dort bei „Anweisungen für die Teilmenge“ eingeben:

Währung != „EUR“

(übersetzt: nimm bitte alle Fälle außer jenen, wo die Währung „EUR“ lautet; das „!“ steht für „nicht gleich“, die An- und Abführungen zu EUR sind nötig, da es sich hier um eine alphanumerische Variable handelt). Dann dem Datensatz in der Zelle darunter einen passenden Namen geben, z. B. „CHFundDM“, OK. Ich persönlich bin mir selbst gegenüber bei allen Filterungen vorsichtig, also prüfe ich stets vor allen weiteren Schritten erst einmal, ob das Filtern wie bestellt gelungen ist, z. B. durch: => Grafiken; => Balkendiagramm, „Währung“ aussuchen, OK: Das Bild zeigt nur mehr DM und

CHF an, jeweils circa 2.000 Fälle, das ist plausibel, also war das Filtern erfolgreich und korrekt. Bereit für den Test.

=> Statistik; => Nichtparametrische Tests; => Kruskal-Wallis-Test; bei dem aufklappenden Menü setze man „Währung“ als gruppierende Variable und „Alter“ als abhängige Variable, dann OK.

```
Kruskal-Wallis rank sum test
data:  Alter by Waehrung
Kruskal-Wallis chi-squared = 224.41, df = 1, p-value < 2.2e-16
```

Wir erkennen einen sehr kleinen p-Wert, der sogar weit jenseits der Schwelle von 0.001 liegt, d. h. der Unterschied im Alter der Münzen zwischen CHF und DM ist hochsignifikant.

### 21.2.7 Rekapitulation Mittelwertvergleiche

In allen Fällen stand eine erste Sichtung der Daten am Anfang, und zwar anhand der statistischen Kennzahlen und einfacher Grafiken. Auf dieser Grundlage wurde die Fragestellung präzisiert und über das weitere Vorgehen entschieden: parametrische Statistik oder nicht-parametrische Statistik. Danach wurden fokussiert die passenden Kennzahlen für die deskriptive Statistik ermittelt, begleitet von angemessenen Grafiken, wobei stets die Kennzahlen im Vordergrund stehen. Wichtige sich andeutende Eindrücke (Ergebnisse) resp. Fragen werden durch adäquate Tests geprüft und ggf. verifiziert. Am Ende stehen eindeutige Antworten auf klare Fragen.

## 22 EINSCHUB: SAMPLING UND BAYES-STATISTIK

---

Eine aktuell als modern geltende Schule der Statistik neigt dazu, die Fragen nach der zugrunde liegenden Verteilung resp. die Annahme, dass eine bestimmte Verteilung vorliege, für einen Irrweg zu halten. Der hier empfohlene Weg, im Zweifel statt der parametrischen Verfahren dann eben in den Werkzeugkasten der grundsoliden nicht-parametrische Statistik zu greifen, misshagt ihr. Stattdessen werden von dieser Schule Sampling und Bayes-Statistik propagiert.

## 22.1 SAMPLING (BOOTSTRAPPING)

Was „Sampling“ ist, hatten wir oben schon unter dem Thema „Jackknifing“ erfahren: Das systematische „Löschen mit Zurücklegen“ von Fällen samt Protokollieren der jeweils leicht unterschiedlichen Ergebnisse. Archäologen ist dieser Ansatz im Grunde sehr vertraut, nur umgekehrt: Nicht selten kommt es im Verlauf einer größeren Studie dazu, dass während des Arbeitens unerwartet noch etwas hinzukommt – man z. B. noch ein paar wichtige Neufunde aus einem frisch erschienenen Buch oder einem „Magazinfund“ berücksichtigen und integrierten möchte. Also flugs den Datensatz erweitert und kurz die ganzen Rechnungen noch mal durchlaufen. Wenn die Ergebnisse dabei grundsätzlich ähnlich bleiben: prima. Wenn die Ergebnisse plötzlich ganz anders ausfallen, sollte man das naheliegende Fluchen vermeiden und begreifen, dass der bisherige Datensatz möglicherweise nicht umfangreich genug war für die Aussagen, die man aus ihm ableiten wollte. Denn gute Statistiken und Thesen sollten in der Lage sein, relevante Neufunde zu integrieren. Oder andersrum: die Kraft zur Integration von Neufunden ist ein guter Test für die Stabilität von Ergebnissen. „Sampling“ ist nichts anderes als die Umkehrung dieses Ansatzes: Wenn man die Serie nicht erweitern kann (was ja in der Archäologie oft der Fall ist), testet man die Stabilität der Resultate eben durch systematisches Löschen am vorhandenen Datenbestand.

Praktischerweise ist der Sampling-Ansatz in PAST recht umfassend integriert – was zu beleuchten wir jetzt nachholen. Man lade z. B. den Datensatz „Rollet 1888“ und mache eine univariante Statistik für die Femurlänge F1 und erfährt u. a.: Mittelwert 431,945 mm, Std.abw. 29,59 mm. Die betreffende Schaltfläche von PAST „Univariate Statistics“ weist rechts Schaltflächen auf, die wir noch nicht benutzt haben. Man aktiviere „Bootstrap“ (=Häkchen in die Schaltfläche klicken), belasse es bei der Voreinstellung „Bootstrap type: Simple“ und auch bei der vorgeschlagenen Anzahl der Löschungen und Neuberechnungen „Bootstrap N: 9999“. **Abb. 94** zeigt das Ergebnis:

	All	Lower conf.	Upper conf.
N	100	100	100
Min	372,5		
Max	500		
Sum	43194,5	42607,5	43779
Mean	431,945	426,075	437,79
Std. error	2,95908	2,609862	3,279396
Variance	875,6156	681,1378	1075,444
Stand. dev	29,5908	26,46932	33,1857
Median	431,75	425	444
25 prcntil	410,625	403,625	417,25
75 prcntil	451,875	439,75	461,125
Skewness	0,1363743	-0,1618228	0,4227335
Kurtosis	-0,6166204	-1,130475	-0,2380427
Geom. mean	430,9434	425,0569	436,7236
Coeff. var	6,850595	6,133744	7,680592

Bootstrap  
 Bootstrap type: Simple  
 Bootstrap N: 9999

**Abb. 94** Femurlänge F1 in der Serie Rollet (1888): Ergebnisse eines Bootstrappings mit PAST.

– ein Ergebnis, das sich nach erneutem Drücken der Schaltfläche „Recompute“ wieder leicht verändert, weil eben neu wieder zufällig 9.999 Stichproben durchgerechnet werden. Neben den konventionellen (und unveränderten) univariaten Statistiken (Spalte „All“), die wir schon kennen, ist nun deren Unteres (5 %) und Oberes (95 %) Konfidenzintervall ausgewiesen: Sie zeigen jenen Wertebereich auf, in den 90 % der Ergebnisse der 9.999 Stichprobenziehungen fielen. Erschreckend? Nun, der konventionelle Mittelwert liegt bei ca. 43,2 cm mit einer Standardabweichung von knapp 3,0 cm, das Bootstrap-Konfidenzintervall (CI) bei ca. 42,6 – 43,8 cm. Das mag jeder für sich bewerten, mir erscheint das CI mit einer Weite von knapp 1,2 cm angesichts einer Grundgesamtheit von 100 zu Grunde liegenden Skeletten als Bestätigung des konventionellen Ansatzes.

PAST bietet dieses Bootstrapping = Sampling zu mehreren unterschiedlichen statistischen Prozeduren an, wobei im Hintergrund stets das gleiche geschieht: ein häufiges Wiederholen der Berechnungen anhand von Stichproben samt Protokoll der Ergebnisse.

Mit **R** ist Sampling zwar (hervorragend) möglich, doch sind diese Optionen nicht in den R-Commander integriert. Demgegenüber bietet BlueSky diesbezüglich so viele Optionen, dass für einen soliden Umgang damit mehr Wissen über Bayes-Statistik erforderlich ist, als es hier vermittelt werden soll.

## 22.2 BAYES-STATISTIK

Wiewohl Sampling und Bayes-Statistik gerne in einem Atemzug genannt werden, unterscheidet sich Bayes-Statistik vom Sampling-Ansatz deutlich. Deren Ansatz ist das Arbeiten mit a-priori-Wahrscheinlichkeiten, d. h. mit Wissen, das man schon vor der eigenen Untersuchung hatte resp. nun aufgrund der Literatur erst einmal zusammengestellt hat. Dieses explizit dargelegte (a priori) Vorwissen wird dann mit den Resultaten der aktuell untersuchten Serie verglichen und geprüft, ob die neuen Daten mit dem Vorwissen vereinbar sind, woraus sich anschließend ein neues, gemeinsames Wissen ableiten lässt, die a-posteriori-Wahrscheinlichkeiten. In der Praxis erfordert dies viel Wissen über Statistik und Wahrscheinlichkeitsrechnung. Wer sich darin vertiefen möchte: Gill (2015) und Kruschke (2015) werden als Lehrbücher oft empfohlen. Ich persönlich habe jedoch eine nicht geringe Skepsis, ob die Anwendung von Bayes-Statistik für die üblichen archäologischen Fragestellungen erhebliche Vorteile gegenüber den klassischen Wegen (inkl. Bootstrapping) bringt.

## 23 ÜBER DEN ZUSAMMENHANG VON ZWEI MESSWERTEN: KORRELATION UND REGRESSION

---

### 23.1 BEGRIFFSKLÄRUNG KORRELATION UND REGRESSION

Beim eingangs eingeführten Thema Vier- oder Mehrfeldertafel und Chi-Quadrat-Test wurde der Zusammenhang zwischen zwei Häufigkeiten untersucht, also nominalskalierten Daten. Im Folgenden geht es darum, den Zusammenhang zwischen zwei Messwerten zu untersuchen, die mindestens Ordinalskalenniveau aufweisen. Dazu gehören die Begriffe „Korrelation“ und „Regression“, die eine leicht unterschiedliche Bedeutung haben.

- Frage: Besteht ein Zusammenhang zwischen zwei Variablen? => Korrelation.
- Frage: Wie genau sieht der Zusammenhang zwischen den beiden Variablen aus? Wie gut lässt sich aus der einen Beobachtung die andere Beobachtung prognostizieren? => Regression.

Die erste Frage ist also stets die nach dem Vorliegen einer Korrelation. Sofern ein Zusammenhang besteht (und nur dann), macht es Sinn, den Zusammenhang näher beschreiben zu wollen, z. B. um auch Prognosen für weitere Beobachtungen ableiten zu können.

## 23.2 KORRELATION

Die Stärke des Zusammenhangs zwischen zwei metrisch- oder ordinal-skalierten Variablen wird mit einem Korrelationskoeffizienten ausgedrückt. Der geeignete Koeffizient richtet sich nach der Art der Variablen. Die nachfolgende Tabelle **Abb. 95** gibt eine Übersicht über geeignete Verfahren. Sie soll dazu dienen, dass Sie aus der Fülle der Möglichkeiten das richtige Verfahren auswählen. Die Tabelle enthält allerlei Fachbegriffe, die näher zu beleuchten sind.

Art der Variablen x und y	Verteilung der Daten	Koeffizient des Zusammenhangs	
metrisch	singulär	symmetrisch & unimodal	Maßkorrelationskoeffizient $r$
		asymmetrisch & unimodal	Rangkorrelationskoeffizient $r_s$ ( $\rho$ , $\rho$ ) oder $\tau$ (tau)
		bi- / multimodal	Kontingenzkoeffizient C bzw. K
	gruppiert	symmetrisch & unimodal	Maßkorrelationskoeffizient für Häufigkeiten $r$
		asymmetrisch & unimodal	Rangkorrelationskoeffizient für Häu- figkeiten $r_s$ ( $\rho$ , $\rho$ ) oder $\tau$ (tau)
		bi- / multimodal	Kontingenzkoeffizient C bzw. K
ordinal	singulär	–	Rangkorrelationskoeffizient $r_s$ ( $\rho$ , $\rho$ ) oder $\tau$ (tau)
	singulär mit Bin- dungen	–	korr. Rangkorrelationskoeffizient $r_{\text{korr}}$ oder $\tau$ (tau)
	gruppiert	–	Rangkorrelationskoeffizient $r_s$ ( $\rho$ , $\rho$ ) oder $\tau$ (tau)
kategorial	–	–	Kontingenzkoeffizient C bzw. K
alternativ	–	–	Phi-Koeffizient $\Phi$ bzw. $\Phi_{\text{cole}}$

**Abb. 95** Tabelle zur Auswahl des geeigneten Verfahrens für eine Korrelationsrechnung.

### 23.2.1 Begriffsklärungen

„Bindungen“ liegen vor, wenn die Daten solche Werte haben, dass sich mehrfach gleiche Ränge ergeben. Wenn ca. 1/5 oder mehr gleiche Ränge vorliegen, sollten die Verfahren „mit Bindungen“ gewählt werden. Unklar? Ein Beispiel: Die Beobachtungsreihe studentischer Körperhöhen in einer Statistik-Übung „160, 165, 170, 172, 172, 173, 175, 175, 178, 181, 184, 185, 186, 187 cm“ macht folgende Ränge: 1. – 160, 2. – 165, 3. – 170, 4.–5. 172, 6. – 173, 7.–8. – 175, 9. – 178, 10. – 181, 11. – 184, 12. – 185, 13. – 186, 14. – 187. Befund: Die meisten Daten sind singular, es liegen (so gut wie) keine Bindungen vor. Die Beobachtungsreihe Semesterzahl zur selben Gruppe „1, 1, 2, 3, 4, 4, 4, 4, 5, 5, 5, 6, 8, 10, 11“ macht die Ränge: 1.–2.: 1; 3.: 2; 4.: 3; 5.–8.: 4; 9.–11.: 5; 12.: 6; 13.: 8; 14.: 10; 15.: 11. Befund: Die Semesterzahlen sind „gruppiert“ bzw. bilden eine Datenreihe „mit Bindungen“.

„Alternativ“ meint Variablen mit 2 Merkmalszuständen, z. B. ja/nein, vorhanden/nicht vorhanden, oder auch männlich/weiblich.

### 23.2.2 Der Wert des Korrelationskoeffizienten

Die Werte eines Korrelationskoeffizienten (z. B. von  $r$ ,  $\rho$ ,  $\tau$ ) schwanken zwischen -1 und +1. Dabei bedeutet:

$r = 0$  Es besteht kein Zusammenhang zwischen beiden Größen.

$r = +1$  Es besteht ein enger positiver Zusammenhang zwischen beiden Größen, und zwar der Art „je größer  $x$ , desto größer  $y$ “.

$r = -1$  Es besteht ein enger negativer Zusammenhang zwischen beiden Größen, und zwar der Art „je größer  $x$ , desto kleiner  $y$ “ (und umgekehrt).

### 23.2.3 Korrelationskoeffizient $r$

Der sog. „Produkt-Moment-Korrelationskoeffizient  $r$  nach Bravais-Pearson“ oder auch „Maßkorrelationskoeffizient“ ist ein Maß für Stärke und Richtung eines linearen Zusammenhangs zwischen intervallskalierten Daten. Voraussetzungen:

- (1) es besteht ein linearer Zusammenhang zwischen
- (2) stetigen Zufallsvariablen, die
- (3) bivariat normalverteilt sind (gemeinsam normalverteilt sind) und
- (4) als unabhängige Beobachtungsdaten vorliegen.

Die Voraussetzungen (1) und (4) gelten als besonders wichtig. Die Forderung (4) nach Unabhängigkeit ist beispielsweise dann verletzt, wenn  $x$  und  $y$  sich zu 100 % ergänzende Prozentsätze sind („Scheinkorrelation“), oder Maße an ganzen Gefäßen ermittelt werden (verschiedene Maße am gleichen Gefäßindividuum).

### 23.2.4 Rangkorrelationskoeffizient ( $\rho$ , $\tau$ )

Bezweifelt man das Vorliegen einer gemeinsamen Normalverteilung von  $x$  und  $y$ , also der Voraussetzung (3), sollte man einen „Rangkorrelations-Koeffizienten“ errechnen; er betrachtet statt der Messwerte deren Ränge.

Rangkorrelationskoeffizient: welcher? Man verwendet zwei leicht unterschiedliche Koeffizienten: Kendall's  $\tau$  (tau) und Spearman's  $\rho$  (rho; auch:  $r_s$ ).

- Spearman's  $\rho$  (rho) ist der klassischerweise ermittelte Koeffizient.
- Kendall's  $\tau$  (tau) hat statistisch gewisse Vorteile gegenüber Spearman's  $\rho$  (rho), ist aber aufwändiger in der Berechnung und war daher früher weniger beliebt.  $\tau$  sollte dann bevorzugt werden, wenn Bindungen vorliegen, d. h. viele gleiche Ränge. Für große Stichproben führen  $\rho$  (rho) und  $\tau$  (tau) zu sehr ähnlichen Ergebnissen. Da die Frage des Aufwands für die Berechnungen heute keine Rolle mehr spielt, empfehle ich, in der Regel Kendall's tau den Vorzug zu geben.

Wichtig: Stets sorgfältig angeben, welchen Koeffizienten man gerechnet hat. Es sind nur Werte (Effektstärken) von  $r$ ,  $\tau$  und  $\rho$  untereinander vergleichbar, nicht aber miteinander. Man kann also schreiben, dass in einem Fall  $\rho$  höher ausfällt als in einem anderen Fall, aber nicht den  $\tau$ -Wert von zwei untersuchten Wertereihen mit dem  $\rho$ -Wert zweier anderer Reihen vergleichen.

### 23.2.5 Kontingenzkoeffizienten **C** und **K**

Die Kontingenzkoeffizienten bieten ein Maß für den Zusammenhang zwischen nominalskalierten Daten. „K“ ist eine Verallgemeinerung des  $\Phi$ -Koeffizient (s. unten); er ist abhängig auch von der Größe der Zwei- oder Mehrfeldertafel. Daher ist der Kontingenzkoeffizient **C** zu bevorzugen, der unabhängig von der Tafelgröße ist.

Beide Koeffizienten nehmen Werte zwischen 0 und 1 an, **K** und **C** = 1 stehen für einen vollständigen Zusammenhang.

### 23.2.6 $\Phi$ -Koeffizient

Maß für den Zusammenhang bei alternativen Daten. Der Koeffizient erreicht wie der Korrelationskoeffizient Werte zwischen +1 und -1.

### 23.2.7 Yule-Koeffizient

An Stelle von  $\Phi$  wurde in der älteren Literatur auch der Assoziationskoeffizient **Q** (auch: „Yule-Koeffizient“; z. B. Neuffer 1965; Gebühr 1983) errechnet. Er hat statistisch vielerlei Nachteile und sollte nicht mehr verwendet werden.

### 23.2.8 Korrelationskoeffizient und Signifikanz

Die Ausgaben der Programme umfassen stets zwei Werte, den Koeffizienten und seine Signifikanz. Zwei Zahlen? – ein Faktum, das immer wieder zu Verwirrungen führt. Aber richtig, man braucht beides. Der Koeffizient beschreibt die Stärke eines Zusammenhangs. Die Signifikanz beschreibt, ob der berechnete Zusammenhang auch statistisch signifikant ist. Man braucht beide Informationen. Liegen beispielsweise nur zwei oder drei Zahlenpaare vor, d. h. sehr wenige Beobachtungen, kann sich eine hohe Korrelation ergeben, die zwar rechnerisch richtig ist, aber – wegen der geringen Fallzahl – nicht signifikant.

### 23.2.9 Praxis R: Korrelation

Wir wollen an unseren bereits bekannten Datensätzen einige Variablen auf das Vorliegen eines Zusammenhangs miteinander prüfen. An den Toten aus Zwentendorf hatten wir uns bislang auf den Oberschenkelknochen (Femur) konzentriert. Nun interessiert die Frage, ob und ggf. wie stark die Länge der vier wesentlichen Langknochen eines Menschen miteinander in Zusammenhang steht, Femur (F1) und Tibia (T1b) als Beinknochen und Humerus (H1) und Radius (R1) als Armknochen. Wie gewohnt schauen wir uns zunächst die Kennzahlen an und testen auf das Vorliegen einer Normalverteilung.

	mean	sd	skewness	kurtosis	25%	50%	75%	100%	n
H1	314.3544	21.16452	0.2030534	-0.2644030	299.0	315.0	331.0	368.0	79
R1	239.1493	17.82767	0.3118104	-0.5107135	222.5	238.5	253.3	284.0	67
F1	437.3667	27.95440	0.5596156	0.1244277	418.1	435.5	452.6	514.5	90
T1b	355.3636	25.95922	0.2950386	0.1615635	338.5	357.0	366.8	424.0	66

```
Lilliefors (Kolmogorov-Smirnov) normality test
data: H1      D = 0.062516, p-value = 0.6253
data: R1      D = 0.08615,  p-value = 0.2506
data: F1      D = 0.080352, p-value = 0.1626
data: T1b     D = 0.11787,  p-value = 0.02354
```

Wir stellen fest – gestützt auch durch Histogramme, die wir hier nicht zeigen – dass alle Verteilungen unimodal sind; der Vergleich Mittelwert und Median zeigt zusammen mit einem Blick auf Schiefe und Wölbung, dass die Verteilungen relativ symmetrisch sind. Der Test auf Normalverteilung ergibt für H1, R1 und F1 keinen Grund, die Annahme einer Normalverteilung zu verwerfen. Der p-Wert für die Tibia (T1b) ist kleiner 0.05, d. h. hier wird die Annahme einer Normalverteilung signifikant, aber nicht sehr signifikant zurückgewiesen. Nach dem Gesamtbild entscheiden wir uns für die Untersuchung der Korrelation nach dem parametrischen Korrelationskoeffizienten  $r$  (Pearson-Produkt-Moment-Korrelation).

R-Commander: => Statistik; => Deskriptive Statistik; => Korrelationsmatrix; in dem aufklappenden Menü wählen wir die vier Variablen aus, setzen ein Häkchen an „Pearsons Produkt-Moment“, ein Häkchen an „Paarweise vollständige Beobachtungen“ und ein Häkchen an „paarweise p-Werte“.

Kurze Erklärung: Würden wir „vollständige Beobachtungen“ wählen, würden nur die Individuen in die Berechnung eingehen, bei denen alle vier Messstrecken vorhanden sind; unser Häkchen „paarweise vollständige Beobachtungen“ berechnet die Korrelation jedes Mal dann, wenn das betreffende Knochenpaar bei einem Individuum vorhanden ist. Sprich: Unsere Entscheidung ist wegen der fehlenden Werte etwas inkludierender. Die p-Werte brauchen wir, weil erst das Doppelpack aus Koeffizient und Signifikanz weiterführend ist.

Nach dem OK erhalten wir folgendes Bild:

```

Pearson correlations:
      F1      H1      R1      T1b
F1  1.0000  0.9132  0.8795  0.9488
H1  0.9132  1.0000  0.9046  0.8777
R1  0.8795  0.9046  1.0000  0.8825
T1b 0.9488  0.8777  0.8825  1.0000

Number of observations:
      F1  H1  R1  T1b
F1   90  54  45  52
H1   54  79  54  51
R1   45  54  67  39
T1b  52  51  39  66

Pairwise two-sided p-values:
      F1      H1      R1      T1b
F1          <.0001  <.0001  <.0001
H1  <.0001          <.0001  <.0001
R1  <.0001  <.0001          <.0001
T1b <.0001  <.0001  <.0001

```

**R** macht eine einfarbige Ausgabe der Resultate. Um das Lesen zu erleichtern und die Struktur der Ausgabe zu verdeutlichen, sind hier Farben eingesetzt worden. Die obere der drei Tabellen ist die „Korrelationsmatrix“ mit der Angabe der Korrelationskoeffizienten. Jede Variable ist vollständig mit sich selbst korreliert, daher geht eine diagonal verlaufende Achse von +1.000-Werten quer durch die Tabelle: F1 mit F1, H1 mit H1 usw., die jeweils eine „+1“ für die maximale positive Korrelation ausweist: richtig, und zugleich eine weitgehend überflüssige Information. Zu dieser diagonalen Achse ist die

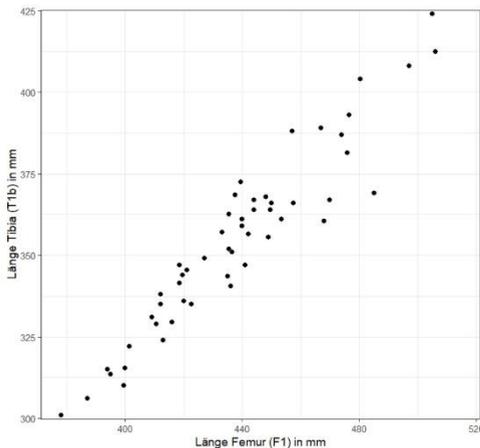
Tabelle spiegelsymmetrisch, unten links liest man den Wert von F1 mit T1b ab, oder oben rechts für T1b mit F1, beide Werte sind (wenig überraschend) identisch. Eigentlich würde es also reichen, eine Seite der Tabelle abzubilden (was manche Statistikprogramme auch tun).

Die zweite Tabelle darunter gibt – bei gleichem Tabellenaufbau – an, wie viele Pärchen jeweils zur Berechnung des Koeffizienten zur Verfügung standen, sie dient also der Quellenkritik. Die dritte Tabelle weist die p-Werte aus, hier in einer verkürzten Schreibweise: Alle p-Werte sind kleiner als 0.001, d. h. alle Koeffizienten (oben) sind jeweils hochsignifikant.

Also schauen wir zurück auf die oberste Tabelle. Alle Werte des Korrelationskoeffizienten sind hoch. Der Zusammenhang für die beiden Beinknochen (F1, T1b) ist mit 0.9488 besonders stark, der zweitstärkste Zusammenhang mit 0.9046 besteht zwischen den beiden Armknochen (H1, R1), während die Beziehung zwischen einem Arm- und einem Beinkochen minim schwächer ausfällt. Nicht überraschend, aber gut, dass wir uns dies einmal angeschaut haben.

Da wir zwar stets die statistischen Kennzahlen in den Vordergrund rücken als die entscheidenden harten Fakten, uns aber routinemäßig immer auch durch eine adäquate Grafik kontrollieren wollen, suchen wir nun nach einer geeigneten Grafik: ein Streudiagramm.

R-Commander: => Grafiken; => Streudiagramm; als x-Variable (waagrecht) wählen wir F1 aus, als y-Variable T1b, OK.



**Abb. 96** Streudiagramm von Femurlänge (F1) gegen Tibiallänge (T1b) für die Population aus Zwentendorf (n=52).

Wir erkennen in **Abb. 96** eine schlanke Punktwolke von links unten nach rechts oben: Wenn F1 länger wird, wird tendenziell auch T1b länger. Aber die Punkte liegen auch nicht auf einer Linie (dann wäre der Korr.-koeff. +1.0), sondern sie streuen um diese herum.

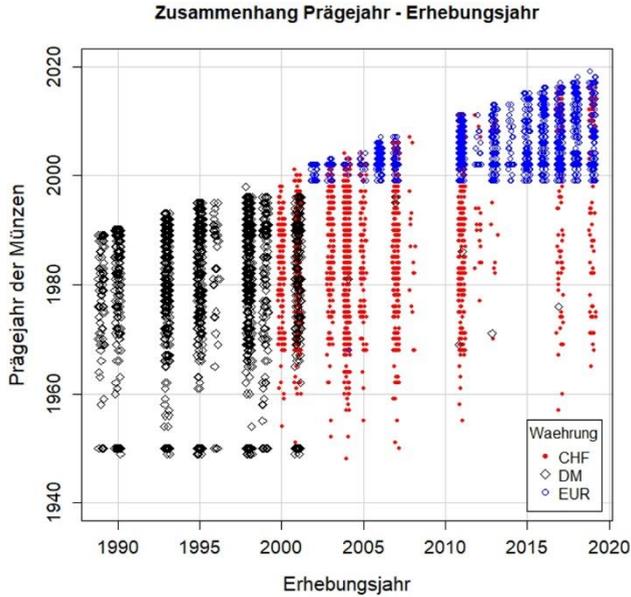
Für ein zweites Beispiel wechseln wir zu den Geldbörsen. Frage: Wir kennen das Erhebungsjahr jeder Münze, wir haben das Prägejahr jeder Münze, wie eng hängen beide Zahlen eigentlich zusammen? Wir erinnern uns: Die Daten waren zwar unimodal, aber alles andere als normalverteilt. Daher wird dieser Zusammenhang nicht-parametrisch untersucht, mit dem Rangkorrelationskoeffizienten  $\rho$  nach Spearman.

R-Commander: => Statistik; => Deskriptive Statistik; => Korrelationsmatrix; in dem Menü die beiden Variablen anwählen, Häkchen an „Spearman Rangkorrelation“ und „paarweise p-Werte“. Das Ergebnis:

```
Spearman correlations:
      Jahr Stichjahr
Jahr      1.0000   0.5612
Stichjahr 0.5612   1.0000
Number of observations: 7331

Pairwise two-sided p-values:
      Jahr Stichjahr
Jahr              <.0001
Stichjahr <.0001
```

Das Lesen ergibt: Beide Größen hängen hochsignifikant zusammen, der Korrelationskoeffizient liegt mit ca. 0.56 jedoch nicht sehr hoch, d. h. wir erwarten eine gegenüber dem vorangehenden Bild zwischen F1 und T1b (**Abb. 96**) deutlich weniger gestreckte, sondern rundere Wolke in einem Streudiagramm (**Abb. 97**). Das Bild illustriert nochmals das bereits vorhandene Wissen zum Thema: Der Münzfundus wird zwar durch Neuemissionen von Jahr zu Jahr jünger, aber es sind eben weiterhin viele ältere Münzen im Umlauf, sodass nur ein loser Zusammenhang zwischen dem Jahr der Stichprobenziehung und dem Prägejahr der einzelnen Münzen besteht.



**Abb. 97** Zusammenhang zwischen Erhebungsjahr der Münzstichprobe und dem Prägejahr der Münzen, differenziert nach den Währungen DM, CHF und EUR (n=7.743).

Übrigens: Wenn wir den p-Wert für die Korrelation genauer benötigen oder als Rangkorrelationskoeffizienten Kendall's  $\tau$  einsetzen wollen: das ist im R-Commander an anderer Stelle angesiedelt: => Statistik; => Deskriptive Statistik; => Test auf Signifikanz der Korrelation, usw.

### 23.3 REGRESSION

Die Regression untersucht den Zusammenhang zwischen metrisch skalierten Variablen. Kurz zum Namen Regression: „Ungewöhnlich große Väter haben eher kleinere Söhne – ungewöhnlich kleine Väter haben eher größere Söhne“ – diese Erfahrung einer allgemeinen Tendenz zum Zurückgehen auf den Mittelwert („Regression“) gab dem Verfahren seinen Namen.

Vor allem Weiteren sollte man sich die Beziehung zwischen zwei metrisch skalierten Variablen zuerst mit Hilfe eines Streudiagramms visualisieren. An der Punktwolke kann man erkennen resp. die Vermutung begründen, ob ein Zusammenhang besteht, und wie dieser Zusammenhang beschaffen sein könnte. Die Frage, ob

ein Zusammenhang besteht, kann mit dem Korrelationskoeffizienten näher untersucht werden. Die Frage, wie der Zusammenhang beschaffen ist, wird mit der Regression beschrieben. Dabei unterscheidet man eine „lineare Regression“ von einer „nicht-linearen Regression“.

### 23.3.1 Lineare oder nicht-lineare Regression?

- Lineare Regression. Optisch: Die Punktwolke streut um eine Gerade. Der Zusammenhang ist linear: je größer, desto schwerer; je mehr, desto teurer, usw.
- Nicht-lineare Regression. Optisch: Die Punktwolke streut um eine Kurve, z. B. um eine Parabel. Der Zusammenhang ist nicht-linear. Beispiel: Lebensalter und Körpergewicht; anfangs – als Babies und Kinder – sind Menschen jung und leicht, werden zunehmend älter und schwerer, aber die meisten Menschen werden im hohen Alter wieder leichter. Auch dieser Zusammenhang lässt sich statistisch gut fassen und beschreiben, aber es liegt kein linearer Zusammenhang vor, sondern ein nicht-linearer.

### 23.3.2 Lineare Regression

Die Regressionsgerade ist die Gerade, die so durch die Punktwolke führt, dass die Summe der Distanzen zwischen den Punkten und der Geraden minimal ist.<sup>66</sup>

Die Regressionsgerade wird beschrieben durch die Geradengleichung:  $y = b * x + a$ .

- b: „Steigung“ (engl. *slope*, manchmal auch: *factor*)
- a: Konstante = Schnittpunkt mit der y-Achse (engl. *intercept*)

Die Regressionsgerade erlaubt es, aus dem bekannten Wert x einen erwarteten Wert y zu schätzen.

Vorsicht: Je nach regionaler Tradition der Notation werden a und b auch exakt andersherum verstanden; eine weitere Notation ist:  $y = \alpha + \beta * x$ . Man muss also erst verstehen, wie das Statistikprogramm seines persönlichen Vertrauens die Begriffe bzw. Buchstaben besetzt hat.

### 23.3.3 Wie jetzt: zwei Regressionsgeraden?

Will man aus einer „unabhängigen“ Variablen x auf eine „abhängige“ Variable y schätzen, gibt es eine Regressionsgerade. Möchte man zugleich auch von y auf x schätzen,

---

<sup>66</sup> Genau genommen betrachtet die klassische Regressionsrechnung nicht die Summe der Distanzen senkrecht zu Regressionsgeraden (d. h. gleichermaßen in x- und y-Richtung), sondern sie minimiert die Summe der vertikalen Distanzen (d. h. allein in y-Richtung, resp. in z-Richtung bei der zweiten Geraden).

ergibt sich hierfür eine zweite Regressionsgerade; nur im Falle eines gänzlichen Zusammenhangs von  $x$  und  $y$  sind beide Regressionsgeraden deckungsgleich. Daher: ja, zwei Regressionsgeraden.

Wenn die Korrelation zwischen  $x$  und  $y$  nicht ideal  $+1.0$  oder  $-1.0$  ist, gibt es zwei unterschiedliche Regressionsgeraden, je nachdem, welche Variable man als die abhängige, welche man als die unabhängige setzt. Will man von der „Anzahl der Äpfel“ auf den „zu zahlenden Gesamtpreis“ schätzen, ist der Preis die „abhängige Variable“; will man von „Summe des eingesetzten Geldes“ auf „Anzahl der Äpfel, die man dafür bekommt“ schätzen, sind die Äpfel die abhängige Variable.

Voraussetzung für die lineare Regression ist das Vorliegen einer „bi-variaten Normalverteilung“. Was heißt das in der Praxis? Die Daten sollten mindestens das Niveau einer Ordinalskala haben, besser metrisch skaliert sein. Ggf. werden ungünstige Verteilung durch geeignete Transformation normalisiert; das kann heißen: Extreme Ausreißer werden gelöscht (da die Normalverteilung sehr empfindlich gegen Ausreißer ist), oder es wird z. B. eine Wurzeltransformation, eine logarithmische Transformation<sup>67</sup> o. ä. angewendet.

### 23.3.4 Praktisches Vorgehen und Ausblick

Das praktische Durchführen und Auswerten einer Regressionsrechnung erfolgt in acht Schritten:

1. Streuungsdiagramm ansehen: Lässt die Punktwolke die Vermutung eines linearen Zusammenhangs zu?
2. Mit dem Korrelationskoeffizienten ermitteln, ob tatsächlich ein Zusammenhang besteht.
3. Wenn ein Zusammenhang besteht: Die Frage der abhängigen/unabhängigen Variablen klären, oder im Folgenden jeweils zwei Ergebnisse/Regressionsgeraden ausweisen.
4. Regressionsgerade ermitteln; sie gibt die statistische Beschreibung des Zusammenhangs. Die „Konstante“ beschreibt den Schnittpunkt der Regressionsgeraden mit der  $y$ -Achse.
5. Das Bestimmtheitsmaß  $r^2$  angeben; es ist ein Maß für die Stärke des Zusammenhangs, d. h. für die Frage, wie gut aus „ $x$ “ die abhängige Variable „ $y$ “ geschätzt werden kann.

---

<sup>67</sup> Kleine Erinnerung: immer daran denken, dass der Logarithmus von Null nicht definiert ist, d. h. ggf. z.B. mit „ $+3/8$ “ operieren.

Rechnerisch ist  $r^2$  das Quadrat des Maßkorrelationskoeffizienten  $r_s$ . Das Bestimmtheitsmaß  $r^2$  schwankt zwischen den Werten 0 und 1;  $r^2 = 1$  bedeutet, dass die abhängige Variable vollständig aus der unabhängigen Variable abgeleitet werden kann;  $r^2 = 0.75$  hieße, dass 75 % der Varianz der abhängigen Variablen linear aus der unabhängigen Variable erklärbar ist.

6. Ein t-Wert testet das Ganze und gibt eine Signifikanz. Ein kleiner t-Wert signalisiert eine Streuung der Beobachtungen um die Regressionsgerade im Bereich des Zufälligen, ein großer t-Wert signalisiert eine über-zufällige Streuung. Eine nicht-zufällige Streuung besagt, dass die abhängige Variable tatsächlich gut aus der unabhängigen Variablen vorhersagbar ist. Die kritische Größe von t ergibt sich aus dem gesetzten Irrtumsrisiko und den Freiheitsgraden (hier: Zahl der Beobachtungen minus 2).

7. Ein Blick auf den mittleren Schätzfehler werfen, den *standard error of estimation*, kurz SEE. Er ermittelt (s. Kap. 23.3.5) die mittlere Differenz zwischen den nach der Regressionsgleichung geschätzten Werten und den tatsächlich beobachteten Werten. In den Ausgaben von **R** taucht der SEE als *Residual standard error* auf. Er sollte im Hinblick auf das untersuchte Phänomen eine hinreichend geringe Größe haben.

8. Nachdenken, ob das statistische Ergebnis auch sachlich sinnvoll ist! Weil der Nachweis einer Korrelation noch nicht besagt, dass ein tatsächlicher, ein ursächlicher Zusammenhang zwischen zwei Größen besteht. Man spricht dann von einer Scheinkorrelation (*spurious correlation*). Das immer wieder dazu angeführte Beispiel ist eine vor allem in Osteuropa nachgewiesene Korrelation: Mit der Anzahl der Störche pro Quadratkilometer wächst die Anzahl der Geburten pro Frau.<sup>68</sup> Erklärt diese – statistisch starke und signifikante Korrelation – irgendwas, beweist sie, dass „der Storch die Kinder bringt“? Das Vorliegen einer Korrelation sowie deren genauere Beschreibbarkeit durch eine Regression besagt – ohne Weiteres – nicht, dass auch ein ursächlicher Zusammenhang oder ein direkter Wirkungszusammenhang zwischen zwei Größen besteht.

### 23.3.5 „Residuen“: der Vergleich beobachtet vs. erwartet

Anhand der Regressionsgleichung kann man zu den beobachteten x-Werten die erwarteten y-Werte errechnen und speichern, um anschließend die beobachteten y-Werte mit diesen erwarteten y-Werten zu vergleichen. Die Divergenz „beobachtet vs.

---

<sup>68</sup> Matthews, R. (2000). Storks deliver babies ( $p = 0.008$ ). *Teaching Statistics*, 22(2), 36-38.

erwartet“ nennt man „Residuum“, pl. „Residuen“ (*residual*, pl. *residuals*). Aus dem Mittelwert der Residuen ergibt sich der Standardschätzfehler SEE. Es kann sehr gewinnbringend sein, nach der Feststellung eines Zusammenhangs und seiner genauen Beschreibung via Regression in einem zweiten Schritt die Residuen genauer zu untersuchen.

Wie das geht? Indem man die Parameter der Regressionsgeraden nimmt, eine neue Variable berechnet und deren Werte entsprechend als Erwartungswert einsetzt; anschließend kann man die Differenz beobachtet vs. erwartet bilden und diese analysieren. Bei einer gelungenen linearen Regression sollten die Residuen in etwa normalverteilt sein; ist das deutlich nicht der Fall, könnte das ein Hinweis darauf sein, dass die Verfahrensvoraussetzungen für eine lineare Regression nicht erfüllt sind.

### 23.3.6 Praxis R: lineare Regression

Wir beginnen wieder mit dem Beispiel Zwentendorf. Da es einen engen Zusammenhang zwischen den beiden Bein Knochen Femur und Tibia gibt, wollen wir prüfen, ob man aus einer bekannten Femurlänge schätzen kann, wie lang die zugehörige Tibia vermutlich ausfällt. Was beispielsweise helfen könnte, die Zusammengehörigkeit von Knochen in einem Fundarchiv zumindest mit einer gewissen Wahrscheinlichkeit wiederherzustellen, wenn es zu einem Durcheinander gekommen ist.

R-Commander: => Statistik; => Modelle anpassen; => Lineare Regression; in dem aufklappenden Menü setzen wir F1 als unabhängige Variable und T1b als abhängige Variable, dann OK.

Wir erhalten ca. folgende Ausgabe:

```
lm(formula = T1b ~ F1, data = Zwentendorf)
Residuals:
    Min       1Q   Median       3Q      Max
-26.574  -5.552  -0.266   5.126  16.977

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -29.68235   18.13303  -1.637   0.108
F1           0.87682    0.04129  21.235 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.819 on 50 degrees of freedom
(197 observations deleted due to missingness)
Multiple R-squared:  0.9002,    Adjusted R-squared:  0.8982
F-statistic: 450.9 on 1 and 50 DF, p-value: < 2.2e-16
```

Viele Zahlen, die man Schritt für Schritt lesen muss. Den beiden untersten Zeilen entnehmen wir, dass  $r^2$  (sog. Bestimmtheitsmaß) bei 0.9 liegt, d. h. die Gesamtheit der Varianz von T1b kann zu 90 % aus F1 abgeleitet werden (was sehr viel ist), und das Ergebnis einen hohen F-Wert und einen p-Wert nahe Null hat, also hochsignifikant ist.

In dem mittleren Informationsblock „Coefficients“ findet sich die mathematische Beschreibung des Zusammenhangs; demnach kann (siehe Spalte „Estimate“) aus  $F1 * 0.87682$  [i.e. „Faktor“, „Steigung“] + (-29.68235) [i.e. Konstante] die Länge von T1b geschätzt werden. Die zu F1 angegebene Steigung von 0.87682 hat (weiter rechts) einen hohen t-Wert und ist (p-Wert) hochsignifikant.

Die Residuen „beobachtet zu erwartet“ sind im oberen Informationsblock ausgegeben: 50 % aller Schätzfehler liegen in der Spanne -5.6 und +5.1 (1Q = 1. Quartil, 3Q = 3. Quartil). Es handelt sich hier entsprechend der Skala der eingegebenen Werte um Millimeter, d. h. die Schätzung der Tibialänge ist in 50 % aller Fälle auf etwa 1,1 cm genau erfolgt.

Wollten wir die Residuen, d. h. die Differenzen beobachtet/erwartet näher untersuchen, könnten wir anhand der Regressionsgleichung die nötigen Rechnereien selbst ausführen. Doch der R-Commander nimmt uns diese Arbeit auf Wunsch ab, die Parameter wurden vom R-Commander automatisch als „Modell“ gespeichert. Hat man selbst keinen Namen vergeben, zeigt die ansonsten stets leere Schaltfläche rechts oben

nun (in blauer Schrift) den Namen „RegModel.1“. Sind im Laufe des Arbeitens mehrere Modelle berechnet worden, kann man dort – ähnlich wie man bei der Schaltfläche „Datenmatrix“ zwischen Datensätzen hin- und herschalten kann – zwischen den verschiedenen Modellen auswählen. Also:

=> Modelle; => Füge Regressionsstatistiken zu den Daten hinzu; in dem aufgeklappten Menü kann man die Zahlen bestellen, z. B. „Vorhergesagte Werte“ und „Residuen“. Danach sind sie Variablen des bearbeiteten Datensatzes und können dargestellt und ggf. auch weiter untersucht werden. So könnten wir z. B. mit

=> Statistik; => Deskriptive Statistik; => Test auf Normalverteilung ...; Variable „residuals.RegModel.1“ auswählen, „Lilliefors (Kolmogorov-Smirnov)“ auswählen, & OK

prüfen, ob die Residuen signifikant von einer Normalverteilung abweichen. (Was sie in unserem Beispiel mit einem p-Wert von 0.8113 nicht tun). Oder auch ein Histogramm der Residuen zeichnen lassen.

Wir belassen es dabei und gehen hier nicht weiter in die Tiefe – werden das Thema aber im Teil IV anhand einer praktischen Anwendung nochmals erheblich vertiefen. An dieser Stelle ist nur noch ein Hinweis zur Begrifflichkeit von Nöten: Das hier zu Grunde gelegte Standardverfahren der linearen Regression nennt man in naturwissenschaftlichen Kontexten gerne auch OLS-Regression (auch: LSR), *ordinary least squares regression*.

### 23.3.7 Erweiterung der linearen Regression

Die nicht-lineare Regression untersucht die Frage nach dem Vorliegen von anderen, eben nicht-linearen Zusammenhängen: nicht-lineare Regression oder „*curvefit*“ (Kurven-Anpassung). Übliche Modelle für nicht-lineare Zusammenhänge sind Gleichungen zweiten Grades oder dritten Grades. Regressionsgleichungen zweiten Grades rechnen:  $y = c \cdot x^2 + b \cdot x + a$  (quadratischer Zusammenhang; Bild: Parabel), solche dritten Grades:  $y = d \cdot x^3 + c \cdot x^2 + b \cdot x + a$  (kubischer Zusammenhang). Man kann mehrere Modelle durchprobieren und anhand des Bestimmtheitsmaßes ermitteln, welches Modell besser passt. Rein mathematisch betrachtet ist diejenige Regression passender, bei der das Bestimmtheitsmaß (i. e. der Anteil der erklärten Varianz) höher ausfällt, also näher an 1 liegt. Doch Vorsicht: mathematisch passender heißt nicht zwingend sachlich passender. Gerade Funktionen dritten Grades sind sehr flexibel, d. h. sie passen sich oft gut an einen Datenkörper an. Auch hier muss also nach dem Rechnen und der Statistik ein nachdenkliches Interpretieren erfolgen, denn eine bessere Anpassung z. B. nach kubischer Regression heißt noch nicht, dass der Zusammenhang tatsächlich ein kubischer ist.

Das Stichwort „Multiple Regression“ meint Regressionsverfahren für den Fall, dass mehr als eine Einflussgröße vorliegt. Hier ist das Modell nicht: „eine Größe hängt (linear) in zu erforschender Weise mit einer anderen Größe zusammen“, sondern: „viele Größen hängen (linear) in zu erforschender Weise mit einer (oder mehreren) nicht direkt beobachteten Größen zusammen“. Dieser Gedanke leitet über zur Faktorenanalyse und soll in dieser Einführung nicht mehr vertieft werden – wir verfolgen aber im Teil IV „Anwendungen“ ein Beispiel.

Insgesamt ist das Thema Regression „ein weites Feld“, das an dieser Stelle bewusst nicht vertieft wird, weil es kompliziert ist und in der Archäologie in vielen Bereichen keine große Bedeutung hat. Ein guter Ansatz für's eigenständige Weiterarbeiten ist (wie so oft) Ihm (1978, 279 ff.; 424 ff.) oder auch Field, Miles & Field (2012, 261 ff.), besonders lesenswert m. E. McCall (2018) 100-136. Das Thema „Multiple Regression“ ist jedoch beliebt und wichtig im Bereich von analytischen GIS-Anwendungen, Schlagwort „*predictive modelling*“. Wenn man z. B. erforschen will, ob und wie die Wahl von Siedlungsstandorten mit verschiedenen Umweltparametern zusammenhängt wie z. B. Bodengüte, Gewässernähe, Hangneigung, Hangexposition, Niederschlagsmenge usw. Denn hat man an den bereits bekannten Siedlungen Präferenzen und Muster erkannt, lassen sich daraus Prognosen ableiten, wo weitere, noch unentdeckte Siedlungen liegen könnten – Prognosen, die man dann wiederum via GIS in den Raum zurückrechnen kann, um besonders „höffige“ Flächen (alias: Verdachtsflächen, Funderwartungsgebiete) konkret zu umreißen.<sup>69</sup> Ein Thema, das für die Praxis der Bodendenkmalpflege (Bauleitplanung, Ausweisung von archäologischen Schutzzonen etc.) mehr Bedeutung hat, als man ihm z. Zt. beimisst.

### **23.3.8 RMA – Reduced major axis regression**

Aus der Archäologie sind mir keine Anwendungen der „Reduced Major Axis Regression“ (RMA) bekannt, in der physischen Anthropologie erlebt das Thema jedoch in den 2000er und 2010er Jahren einen gewissen Hype und soll daher genannt und kurz beschrieben werden. Das Verfahren wird auch als „Regression Modell II“ bezeichnet (Sokal & Rohlf 1995, 451 ff.; Legendre & Legendre 2012, 543-555). Ausgangspunkt der Debatte sind bekannte Schwächen bzw. Probleme der klassischen (OLS) Regression. Typische in der physischen Anthropologie angeführte Anwendungen sind die

---

<sup>69</sup> Dazu z. B. Wheatley, David & Gillings, Mark (2002). *Spatial technology and archaeology: The archaeological applications of GIS*. London: Taylor & Francis.

Schätzung der Körperhöhe aus Langknochen.<sup>70</sup> So ist bei biologischen Kenngrößen, die oft annähernd normalverteilt sind, die Steigung der Regressionsgeraden weniger durch die vielen „typischen“ Individuen in der Mitte der Streuung bestimmt, denn durch die wenigen „Ausreißer“ an den beiden Enden der Streuung – im Fall der Körperhöhe also durch die wenigen besonders kleinen und besonders großen Individuen. Daher sind die Ergebnisse in den Endbereichen der Achse unsicherer und ggf. stärker fehlerbehaftet als im Zentrum, weshalb wiederum die (nachvollziehbare) These im Raum steht, dass man durch eine OLS-Regression erarbeitete Regressionsformeln nur auf Zielpopulationen anwenden solle, die der Referenzpopulation, an der eine Regressionsformel erarbeitet wurde, ähneln (Stichwort Isometrie/Allometrie). Eine These, die beispielsweise dann relevant ist, wenn die Körperhöhe für Pygmäen anhand jener Regressionsformeln geschätzt werden soll, die auf modernen europäischen oder US-amerikanischen Referenzserien beruhen, deren mittlere Körperhöhe deutlich größer ist. Ein weiteres Thema ist die Gerichtetheit der OLS-Regression, da es (wie oben beschrieben) zwei unterschiedliche Formeln ergibt, wenn aus bekanntem A ein unbekanntes B geschätzt werden soll oder umgekehrt aus bekanntem B die Größe A. Zwar besteht dieses Problem bei der Körperhöhen-schätzung im Grunde nicht, weil hier *in praxi* stets von den Langknochen auf die Körperhöhe geschätzt wird, nie umgekehrt, aber es sind Fragestellungen denkbar, wo auch eine Umkehrung erwünscht ist – wie soll man hier mit der Tatsache von zwei unterschiedlichen Formeln umgehen?

Als möglicher Ausweg aus diesen Problemen wird die RMA-Regression debattiert, deren Regressionsgerade im Grunde genommen in etwa der Mittelachse zwischen den beiden OLS-Regressionsgeraden entspricht.<sup>71</sup> Ein Aufsatz von Richard J. Smith (2009) ordnet diese Debatten;<sup>72</sup> er zeigt auf, dass die Fehler der RMA-Regression in Art und Intensität denen der OLS-Regression ähneln und gibt klare Empfehlungen, wann man besser das eine, wann man besser das andere Verfahren anwenden solle. OLS-Regression sei vorteilhaft, wenn die Schätzung eine klare Richtung habe, also stets aus bekanntem A die unbekannte Größe B geschätzt werden solle, während für Schätzungen in beide Richtungen RMA vorzuziehen sei. Bei der OLS-Regression sei zuvor die Frage der Ausreißer zu debattieren, und anschließend auch

---

<sup>70</sup> Einen wichtigen Startpunkt dieser kritischen, gegen die OLS-Regression gewandten Debatte bildet ein Aufsatz von T. Sjøvold (1990). Für Interessierte: Sjøvold, T. (1990). Estimation of stature from long bones utilizing the line of organic correlation. *Human Evolution* 5, 431-447.

<sup>71</sup> Konigsberg, L. W., Hens, S. M., Jantz, L. M., Jungers, W. L. (1998). Stature estimation and calibration: bayesian and maximum likelihood perspectives in physical anthropology. *Yearbook of Physical Anthropology* 41, 65-92.

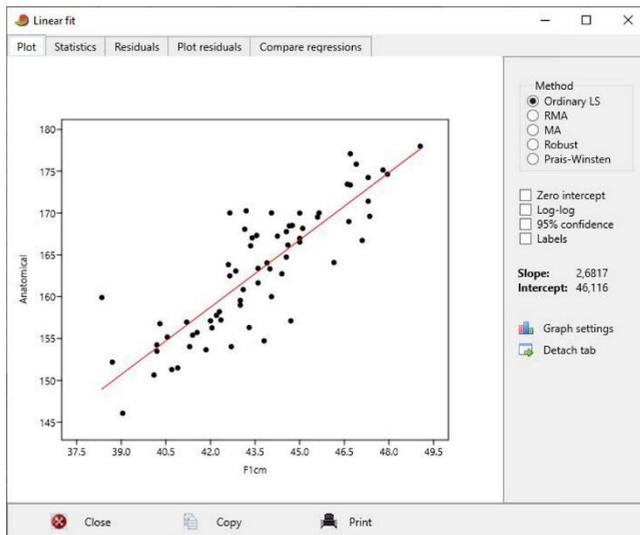
<sup>72</sup> Smith, Richard, J. (2009). Use and misuse of the Reduced Major Axis for line-fitting. *American Journal of Physical Anthropology* 140, 476-486. Auch: Legendre & Legendre 2012, 553 Tab. 10.4.

ein Weiterarbeiten mit den Residuen sinnvoll. In der Archäologie und Anthropologie sind daher die OLS-Regression in der Regel die angemessenere Wahl.

### 23.3.8.1 Praxis PAST: RMA- und OLS-Regression

Ein Vergleich der beiden Verfahren ist mithilfe von PAST besonders leicht möglich. Daten einlesen und die beiden relevanten Spalten markieren; dabei erwartet PAST die unabhängige Variable (x) als erste Spalte, die (zu schätzende) abhängige Variable (y) als zweite Spalte, d. h. bei Vorliegen eines vielspaltigen Datensatzes müssen zunächst die Spalten mit Markieren und „Drag rows/columns“ passend zurechtgeschoben werden. Danach erhält man mit

=> Model; => Linear; => Bivariate eine neue Schaltfläche, in der zunächst als Voreinstellung eine OLS-Regression berechnet wurde, zu der (unten) die Steigung (*slope*) und der Schnittpunkt mit der y-Achse (*intercept*) ausgegeben werden. Die Reiter „Statistics“, „Residuals“... geben weitere wichtige Einblick in die Resultate.



**Abb. 98** OLS-Regression mit Hilfe von PAST der Femurlänge F1 (in cm) auf die Körperhöhe („anatomical“, in cm) anhand der Referenzserie Mavropigi (n=68).

Setzt man das Häkchen in der Schaltfläche rechts oben von „Ordinary LS“ auf „RMA“ um, wird eine RMA-Regression berechnet und entsprechend deren Ergebnisse ausgegeben.

Bei **R** ist die RMA-Regression nicht im Umfang der vom R-Commander oder von BlueSky abgedeckten Verfahren enthalten. Arbeitet man selbstständig mit **R** und

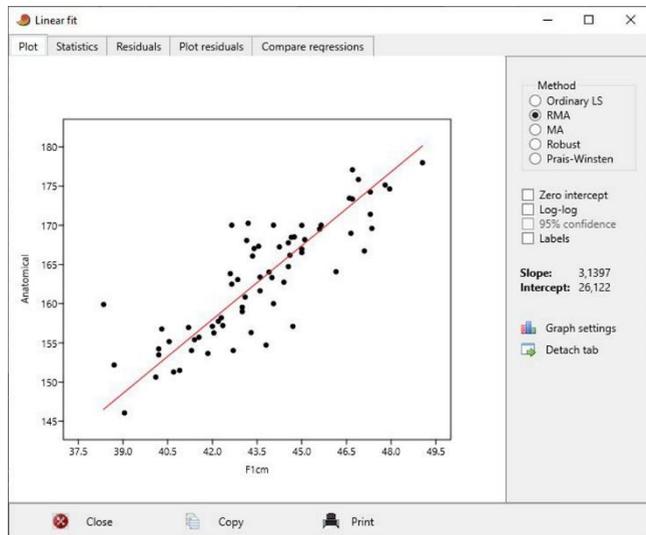
R-Studio, gibt u. a. die von Pierre Legendre zur Verfügung gestellte und gut dokumentierte Funktion „lmodel2“ alle nötigen Optionen und Ausgaben.

Für die konkrete Anschauung wähle ich eine kleine Population aus Nordgriechenland (n=68), zu der mir Femurlängen (F1, in cm) und die sog. anatomische Körperhöhe vorliegen; letztere gilt bei vollständig erhaltenen Skeletten als exakte Rekonstruktion der Lebendhöhe eines Individuums (Abb. 98). Die OLS-Regression ergibt die Schätzformel

$$F1 * 2.682 + 46.12 = \text{Körperhöhe},$$

wobei das Ergebnis (siehe Reiter „Statistics“) hochsignifikant ist,  $r^2$  bei 0.7295 liegt (also im positiven Sinne recht hoch ausfällt) und der mittlere Schätzfehler (siehe Reiter „Residuals“, hier „Std. error of estimate“) mit 3,87 berechnet wurde, also die wahre Körperhöhe mit einem mittleren Fehler von plus/minus 3,9 cm getroffen wird. Das Fenster „Plot residuals“ erlaubt einen schnellen Blick über die Residuen, der nichts Auffälliges zeigt. In Summe nach allen Parametern eine gute Schätzformel.

**Abb. 99** RMA-Regression mit PAST von der Femurlänge F1 auf die Körperhöhe anhand der Referenzserie Mavropigi (n=68).



Setzt man nun das Häkchen bei der Schätzung von F1 auf Körperhöhe auf „RMA“ (Abb. 99), erhält man die Schätzformel

$$F1 * 3.14 + 26.12 = \text{Körperhöhe},$$

ebenfalls hochsignifikant mit identischem  $r^2$  (letzteres ist zwischen OLS und RMA immer so) und einem Schätzfehler von 4,02 cm und ähnlichem Bild bei den Residuen.

Das Beispiel soll hier nicht weiterverfolgt werden, es dient nur dazu, die Ähnlichkeit resp. Unterschiedlichkeit der Resultate zu illustrieren. Nach meiner persönlichen Einschätzung (sprich: meiner Arbeitshypothese) ist das OLS-Resultat im vorliegenden Beispiel das Richtigere, doch dies plausibel zu machen, bedürfte es weiterer Schritte, entweder einer externen Testpopulation, an der man dies prüfen könnte, oder – sofern nur die Referenzpopulation selbst zur Verfügung steht – einem klug überlegten Resampling-Verfahren.

### 23.3.9 Logistische Regression

Mit den vorgestellten Regressionsverfahren wird, wie zuvor aufgezeigt, aus einer oder mehreren „unabhängigen“ quantitativen Variablen eine andere, „abhängige“ quantitative Variable geschätzt. Mit einer logistischen Regression lässt sich aus einer oder mehreren quantitativen Variablen die Zuordnung zu einer binären Klasse („abhängige Variable“) abschätzen, wie beispielsweise das biologische Geschlecht. Die Zielsetzung einer logistischen Regression ähnelt insofern stark der in dieser Einführung nicht weiter besprochenen Diskriminanzanalyse, doch das Verfahren der logistischen Regression hat diesem bekannteren Verfahren gegenüber einige statistisch-theoretische Vorteile.<sup>73</sup>

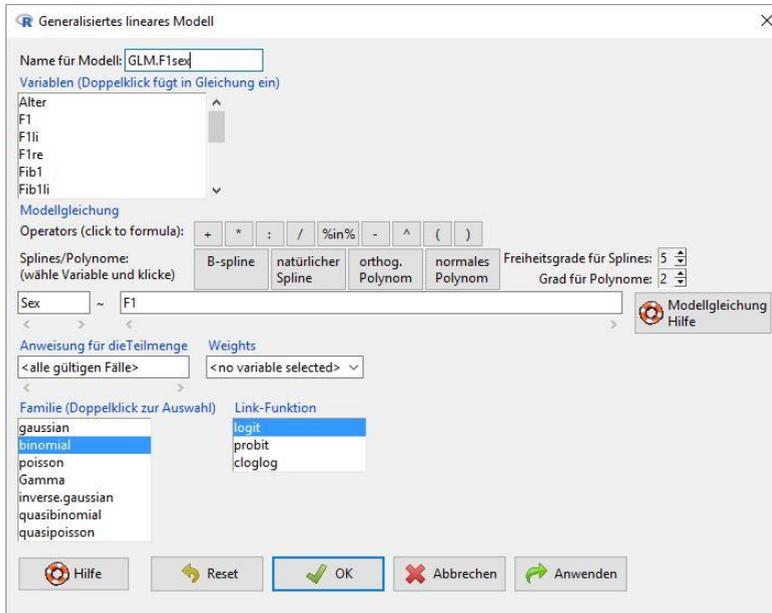
Eine praktische Anwendung ist z. B. das Schätzen des biologischen Geschlechts aus Messtrecken am Knochen, oder etwa das Abschätzen der Frage „Siedlungsstandort ja/nein“ aus diversen Lage- und Umweltparametern bekannter Siedlungen einer Epoche. Anhand einer Referenzserie, bei der die quantitativen unabhängigen Variablen wie auch die abhängige binäre Variable bekannt sind, ermittelt man, ob die in Frage stehende Schätzung möglich ist; wenn ja, können die resultierenden Parameter verwendet werden, um an anderen Serien, bei denen man die binäre Variable nicht kennt, diese zu schätzen.

Als (etwas künstliches) Anwendungsbeispiel sei hier am Datensatz „Rollet 1888“ der Versuch unternommen, aus der Länge des Oberschenkelknochens (Femur, F1) auf das biologische Geschlecht zu schließen. Die Knochenlängen erfüllen die Voraussetzungen des Verfahrens: sie sind annähernd normalverteilt. Die Umsetzung erfolgt nach Laden des Datensatzes in den R-Commander wie folgt:

---

<sup>73</sup> Z.B. Field, Miles & Field 2012, 312-342 u. ebd. 738-745 („DFA“, Diskriminanzanalyse).

=> Statistik; => Modelle anpassen; => Generalisiertes lineares Modell ... In dem dann aufklappenden Menü (**Abb. 100**) geben wir dem Modell (oben) einen Namen, z.B. „GLM.1“. Dann füge man bei „Splines/Polynome“ (Mitte) die Variable „Sex“ ein, und in das lange Feld rechts davon die Variable „F1“. In den beiden Steuerungsfelder unten sollte der blaue Markierungsbalken links („Familie“) auf „binomial“ gesetzt sein (oder werden) und rechts („Link-Funktion“) auf „logit“. Nach OK wird das Bestellte ausgeführt.



**Abb. 100** Menü GLM im R-Commander mit Eingaben zur Ausführung einer logistischen Regression von "F1" auf "Sex". Das Menü ist komplex, weil es das Aufsetzen weiterer komplexer Regressionsverfahren erlaubt. In dem Feld, in dem hier "F1" eingesetzt wurde, können ggf. auch weitere quantitative Variablen eingegeben werden – in unserem Fall z. B. um weitere Knochen für die Schätzung zu verwenden.

Das Ergebnis, das ähnlich zu lesen ist wie die Ergebnisse der zuvor besprochenen linearen Regression, weist aus, dass solch eine Schätzung mit hoher Erfolgsquote möglich ist. Unter „Coefficients“ lesen wir, dass die Schätzformel  $F1 * 0.06802 - 29.33189$  lautet, die ganz rechts ausgewiesenen p-Werte nahe Null signalisieren „hochsignifikant“.

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	-29.33189	5.85433	-5.010	0.00000543	***
F1	0.06802	0.01359	5.007	0.00000554	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 138.63 on 99 degrees of freedom

Residual deviance: 90.10 on 98 degrees of freedom

AIC: 94.1

Number of Fisher Scoring iterations: 5

```
Rcmdr> exp(coef(GLM.1)) # Exponentiated coefficients ("odds ratios")
```

```
(Intercept)      F1  
1.825244e-13 1.070386e+00
```

Die übrigen Werte der Ausgabe sind schwieriger zu interpretieren: Der AIC-Wert („*Akaike information criterion*“) – hier 94.1 – soll, ähnlich wie das  $r^2$  bei einer linearen Regression, die Güte der Schätzung anzeigen. Der AIC-Wert ist jedoch nicht absolut lesbar im Sinne AIC = 10 ist (zu) gering, AIC = 400 ist hoch, sondern nur innerhalb des gleichen Datensatzes relativ zu anderen Schätzungen, d. h. wir könnten ihn vergleichend nutzen, wenn wir die Schätzung zusätzlich anhand anderer Knochenmaße vornähmen – was Sie spielerisch gerne einmal tun könnten. Die „exponentiated coefficients“ dienen der Errechnung eines weiteren Gütemaßes, der Odds-Ratio – was hier aber nicht weiterverfolgt werden soll. Ein uns vertrauterer Gütemaß steckt hinter den beiden Zeilen „null deviances“ und „residual deviances“, welche den Erfolg der Schätzung ohne F1 (null dev.) und nach Verwendung von F1 (residual dev.) messen. Die Differenz beider Werte  $138.63 - 90.10 = 48.53$  entspricht dem Chi-Quadrat-Wert einer Tabelle „zufällige Zuordnung“ vs. „neue Zuordnung“. Ein beobachtetes Chi-Quadrat von 48.53 ist gegenüber dem Schrankenwert einer  $2 \times 2$  Tabelle von ca. 3.41 sehr hoch, wonach die hier erreichten Klassifikation hochsignifikant besser ist als eine zufällige Zuordnung.

Statt der vertieften Analyse dieser Parameter empfehle ich zunächst den Blick auf die Daten selbst. Im R-Commander: => Modelle; => Füge Regressionsstatistiken zu den Daten hinzu ... In dem aufklappenden Menü alle Optionen anwählen & OK. Danach sind die gewünschten Angaben jedem Fall in der Datentabelle hinzugefügt und können nun individuell (via „Datenmatrix betrachten“) oder als Ganze betrachtet werden. So zeigt z.B. => Grafiken; => Histogramm; „fitted.GLM.1“ für die Gruppen „Sex“ ein Histogramm der Zuordnungswahrscheinlichkeiten zur einen wie anderen Gruppe. Man erkennt, dass die Mehrheit der Fälle nahe Null (Frauen) resp. nahe 1

(Männer) angesiedelt ist und nur recht wenige Frauen Werte über 0.5 haben (d. h. als eher männlich eingeschätzt werden) resp. nur wenige Männer Werte unter 0.5 haben (also als Frauen eingeschätzt werden). Analog zum Vorgehen bei einer Diskriminanzanalyse würde man diese Wahrscheinlichkeiten nun nutzen, um die neue Klassifikation vorzunehmen. Unter Ausklammerung der sehr unsicheren Fälle beispielsweise nach dem Ansatz fitted.GLM.1 kleiner 0.4 = Frau, größer 0.6 = Mann, und 0.41-0.59 = unsicher. Der dann mögliche Vergleich zwischen Ausgangsklassifikation und Ergebnisklassifikation würde den Erfolg des Modells veranschaulichen.

Die Cook-Distanz beschreibt (wie auch bei der linearen Regression) in Bezug zum Regressionsmodell ungewöhnliche Fälle („Ausreißer“, zugleich extrem einflussreiche Fälle), die man nach Prüfung erwägen könnte, aus dem Datensatz zu entfernen, um ohne sie ein zweites, den Daten besser angepasstes Regressionsmodell zu rechnen. In unserem Beispiel zeichnet sich zwar ein Fall mit um 0,25 ab, der einen relativ hohen Wert erreicht, aber generell können Cook-Distanzen kleiner 1 als vernachlässigbar gering betrachtet werden. Wir werden auf dieses Thema in unserem Anwendungsfall in Kap. 4 noch zurückkommen.

Wichtiger als die Verfolgung der statistischen Parameter ist es m. E. auch hier (wie bei der linearen Regression), den Erfolg resp. das Leistungsvermögen eines Regressionsmodells durch Anwendung an einer externen Testserie zu prüfen, d. h. klar zwischen Referenzserie/-population und Testserie zu unterscheiden. Denn dass die Resultate an der Referenzserie selbst in der Regel „irgendwie gut“ sind, ist zu erwarten, sagt aber über die praktische Verwendbarkeit des Modells noch wenig aus. Auch dieses Vorgehen wird anhand der konkreten Anwendung in Kapitel 4 exemplarisch durchexerziert.

## 24 (DISKRETE) WAHRSCHEINLICHKEITSVERTEILUNGEN

---

Eine wichtige stetige Wahrscheinlichkeitsverteilung hatten wir schon kennengelernt: die Normalverteilung. Diskrete Wahrscheinlichkeitsverteilungen sind solche, bei denen die resultierenden Werte eine abzählbare Menge an Zuständen annehmen können. Das klassische Beispiel ist der Spielwürfel mit den Augenzahlen von 1 bis 6. Mit einem Wurf kann weder, beispielsweise, die Augenzahl 1,5 noch die Augenzahl 8 erreicht werden. Da es in der Archäologie vielfach um Häufigkeiten geht, spielen diskrete Wahrscheinlichkeitsverteilungen hier eine wichtige Rolle.

Auch die diskreten Wahrscheinlichkeitsverteilungen sind theoretische Modelle, die dazu dienen können, Erwartungen an Beobachtungen zu formulieren, um anschließend die Erwartungswerte und die beobachteten Werte miteinander vergleichen zu können. Die Tatsache, dass eine beobachtete Wertereihe einem bestimmten theoretischen Modell nahekommt, hat statistik-immanente Bedeutung insbes. bei der Wahl geeigneter weiterführender Methoden, aber auch praktisches Erkenntnispotential. So beruht die Idee des Würfelspiels wesentlich auf dem theoretischen Modell, dass alle Ereignisse (Augenzahlen) gleich wahrscheinlich sind. Die zu dieser Spielidee gehörende diskrete Wahrscheinlichkeitsverteilung ist die „Gleichverteilung“; der Abgleich der aus dieser Annahme abgeleiteten Erwartungen und mit den Beobachtungen kann z. B. zeigen, ob ein ganz konkret untersuchter Würfel ein „fairer Würfel“ ist oder ob er zugunsten eines bestimmten Ergebnisses manipuliert wurde.

Das Thema Wahrscheinlichkeitsverteilungen ist komplex, es gibt viele verschiedene solcher Verteilungen, wir begrenzen uns hier mit knappen Ausführungen auf die vier wichtigsten. Angesichts der mit dieser Einführung angestrebten Ziele ist eine intensive Beschäftigung mit dem Thema m. E. nicht notwendig. Die Stichworte sollen aber hier aufgenommen werden, da sie im Text immer wieder einmal zu erwähnen waren. Beim R-Commander findet man hinter der Schaltfläche => Verteilungen, stetige und diskrete Verteilungen; dort kann man unter Eingabe der Parameter u. a. Zufallszahlen erzeugen lassen, die der jeweiligen Verteilung folgen, sowie Grafiken, die – gemäß der eingegebenen Parameter – ein Bild dieser Verteilung zeigen.

## 24.1 GLEICHVERTEILUNG

Jedes der möglichen Ereignisse hat die gleiche konstante Wahrscheinlichkeit des Eintretens. Das typische Beispiel sind die Zahlen beim Würfeln: Bei einem Wurf hat jede der sechs möglichen Zahlen die konstante gleiche Wahrscheinlichkeit („p“) von  $1/6$ .

Ob – um das Eingangsbeispiel mit der Frage nach dem „fairen Würfel“ aufzugreifen – tatsächlich eine diskrete Gleichverteilung vorliegt, würde man mit einem sog. Chi-Quadrat-Anpassungstest untersuchen können. Wenn man, beispielsweise, einen realen Würfel 50 Mal geworfen und die Ergebnisse notiert hätte, wären für jede Augenzahl 1 bis 6 der Erwartungswert jeweils  $50/6$  („ $N/p$ “), d. h. 8,33. Also legt man eine Tabelle mit 6 Zeilen für jede der Augenzahlen an und 2 Spalten. In die erste Spalte trägt man für jede Augenzahl jeweils den Erwartungswert 8,33 ein, in die zweite Spalte die beobachtete Anzahl, wie oft die jeweilige Augenzahl auftrat. Alles Weitere wie schon für den Chi-Quadrattest besprochen.

## 24.2 POISSON-VERTEILUNG

Die Poissonverteilung ist die zufällige Verteilung seltener Ereignisse. Sie tritt auf, wenn die durchschnittliche Anzahl der Ereignisse das Ergebnis einer großen Menge an Ereignismöglichkeiten und einer sehr kleinen Ereigniswahrscheinlichkeit ist.

An Beispielen wird schnell greifbar, was damit gemeint ist: Druckfehler auf einer Buchseite – es gibt viele Buchstaben auf einer Seite und meist nur wenige Fehler; Tore beim Fußball – im Gegensatz zu vielen anderen Ballsportarten fallen beim Fußball eher selten Tore; Autounfälle – in Relation zur Menge an Verkehr sind Unfälle seltene Ereignisse; der radioaktive Zerfall folgt einer Poissonverteilung – relevant für die Praxis von  $^{14}\text{C}$ -Datierungen; Rosinen im Rosinenbrötchen.

Nun noch die gängige Beschreibung als „Urnenmodell Poissonverteilung“: Die Gesamtzahl  $\mathbf{N}$  der Kugeln ist sehr groß, die Anzahl der schwarzen Kugeln  $\mathbf{N} * \mathbf{p}$  sehr klein. Um überhaupt mit schwarzen Kugeln rechnen zu können, muss  $\mathbf{n}$  (i. e. die Zahl der Ziehungen mit Zurücklegen) genügend groß sein. Mathematisch:  $\mathbf{p} \rightarrow 0$ ,  $\mathbf{n} \rightarrow \infty$  und  $\lambda = \mathbf{n} * \mathbf{p} > 0$ . – " $\lambda$ " ist die mittlere Zahl der schwarzen Kugeln unter den herausgenommenen Kugeln.

Bei einer Poissonverteilung sind der Erwartungswert (alias Mittelwert) und die Varianz gleich und werden (traditionell) durch den Wert " $\lambda$ " beschrieben. Bei kleinen Werten von  $\lambda$  ist die Verteilung extrem linkssteil, mit wachsendem  $\lambda$  wird sie nahezu symmetrisch. Bei großem  $\lambda$  (d. h.  $\lambda > 9$ ) kann die Poissonverteilung durch die Normalverteilung approximiert werden.

Die Frage, wie man auf das Vorliegen einer Poissonverteilung testen kann, ist bei Ihm (1978, 223-230 mit Tab. 8-A) sorgfältig beschrieben und vorgerechnet. (a) Man „schätzt“ (errechnet)  $\lambda$  als das arithmetische Mittel, d. h. dividiert die Summe der Beobachtungen durch die Anzahl der Beobachtungen. (b) man berechnet für jede Beobachtung das Abweichungsquadrat, (c) berechnet die Summe der Abweichungsquadrate, (d) berechnet Chi-Quadrat als „Summe der Abweichungsquadrate dividiert durch das arithmetische Mittel“, und (e) vergleicht das hier beobachtete Chi-Quadrat mit dem „kritischen Wert“ von Chi-Quadrat, wobei die „Anzahl der Beobachtungen“ minus 1 den Freiheitsgrad ergibt. Im Falle des von Ihm (1978, Tab-8A) eingebrachten Beispiels ergab sich eine Ablehnung der Hypothese Poissonverteilung.

Um das Ganze handfester zu machen, gebe ich eine LO-Calc-Tabelle mit, die genau dieses Beispiel fertig umsetzt, sodass es besser nachvollzogen kann. Machen Sie doch selbst ein Experiment! Am Ende des Bandes finden Sie ein leeres Gitternetz

(zum Kopieren) vor – wie ein Schachbrett – mit  $8 \times 8$  Feldern. Nehmen Sie 30 Reiskörner und werfen diese zufällig auf die Fläche. Zählen Sie, wie viele Reiskörner in jedes Quadrat gefallen sind. Also „A1 x Körner, A2 y Körner, usw. Mit diesen 64 Beobachtungen von Häufigkeiten berechnen Sie, ob diese einer Poissonverteilung folgen – der zufälligen Verteilung seltener Ereignisse.

Kleines Caveat zu diesem Versuch: (1) Es sind Fälle denkbar, in denen die Anzahl der Beobachtungen pro Quadrat zwar poissonverteilt ist, aber dennoch keine Zufallsverteilung vorliegt. Denn der Test prüft nicht, wie die Quadrate angeordnet sind. Es könnten alle Quadrate mit wenigen oder keinen Reiskörnern konzentriert in einer Ecke des betrachteten Ausschnitts liegen. Medizin angesichts dieser Option: Neben einem statistischen Test ist immer auch eine Überprüfung durch eine Visualisierung notwendig. (2) Das Beispiel kann auch zeigen, dass der Test auf Vorliegen einer Poisson-Verteilung gegenüber dem gewählten Gitternetz sensibel ist. Nimmt man z.B. immer zwei Felder zusammen zu einem  $4 \times 4$ -Gitter, wird der Test ein wahrscheinlich anderes Ergebnis zeigen, ohne dass die Lage der Reiskörner verändert wurde. Sie sind dann eben nur nicht mehr „seltene Ereignisse“. Viel Spass!

### 24.3 BINOMIALVERTEILUNG

Definition: Die Binomialverteilung ist eine diskrete Wahrscheinlichkeitsverteilung für ein Alternativmerkmal (A, nicht-A) mit konstanter Erfolgswahrscheinlichkeit  $p$  für A.

Verständlicher wird der Satz am üblichen „Urnenmodell“ einer Binomialverteilung: Eine Urne enthält  $N \cdot p$  schwarze und  $N \cdot q$  weiße Kugeln, wobei  $p + q = 1$ . Die Wahrscheinlichkeit, eine schwarze Kugel zu ziehen, ist  $p = N \cdot p / (N \cdot p + N \cdot q)$ . Legt man nach dem Ziehen die Kugel wieder in die Urne zurück, bleibt die Wahrscheinlichkeit, eine schwarze oder weiße Kugel zu ziehen, beim nächsten Zug gleich. Zieht man in diesem Sinne jeweils mit Zurücklegen insgesamt  $n$  Kugeln, hat die Anzahl  $x$  der schwarzen Kugeln unter  $N$  eine Binomialverteilung. Das Vorliegen einer Binomialverteilung erwartet man z.B. bei Prozentwerten oder relativen Häufigkeiten.

Der Erwartungswert ("Mittelwert") einer Binomialverteilung ist:  $E_x = \mu = n \cdot p$ . Die Varianz ist:  $\text{var}(x) = \sigma^2 = n \cdot p \cdot q$ , oder in anderer Schreibweise:  $n \cdot p \cdot (1 - p)$ . Den Buchstaben „q“ benutzt man, um die Alternativwahrscheinlichkeit zu beschreiben, stattdessen könnte man auch von „nicht-p“ sprechen.

Bei einer Binomialverteilung ist das Mittel  $E_x$  stets größer als die Varianz  $\sigma^2$ . Für  $p = 1/2$  ist die Verteilung symmetrisch, für höhere oder niedrigere Werte ist sie schief. Mit wachsendem  $n$  nimmt die Binomialverteilung jedoch auch für  $p \neq 1/2$  die

für die Normalverteilung typische Glockenform an; die Konvergenz gegen die Normalverteilung erfolgt jedoch umso langsamer, je weiter  $p$  von  $1/2$  entfernt ist. - Für  $p \rightarrow 0$  und  $n \rightarrow \infty$  konvergiert die Binomialverteilung gegen die Poissonverteilung.

Werden im Urnenmodell die Kugeln nicht zurückgelegt, so ändert sich von Ziehung zu Ziehung die Wahrscheinlichkeit  $p$ ; die daraus resultierende Verteilung ist die hypergeometrische Verteilung; für große Stichproben konvergiert die hypergeometrische Verteilung gegen die Binomialverteilung. Eine Verallgemeinerung der Binomialverteilung ist die Multinomialverteilung (Polynomialverteilung), bei der mehr als zwei sich gegenseitig ausschließende Ereignisse betrachtet werden (Urnenmodell: schwarze, weiße und farbige Kugeln, mit Zurücklegen). Bei einer Mehrfeldertafel mit mehr als  $2 \times 2$  Zellen ist die Multinomialverteilung das zu erwartende Zufallsmodell.

Liegt eine Binomialverteilung vor? Der Weg zur Beantwortung ist wiederum bei Ihm 1978, 226 ff. mit Tab. 8-B sorgfältig beschrieben. Wir rekapitulieren seinem Weg (vgl. Tabelle „Verteilungen“):

Es liegen vor die Beobachtungen  $x_i$  in der Gesamtheit  $n_i$ . Wir errechnen: (a) die Summe aller  $x_i^2 / n_i$ , (b) bilden  $p = \text{Summe aller } x_i / \text{Summe aller } n_i$ , (c) berechnen  $q = 1 - p$ , (d) berechnen  $N \cdot p^2 = \text{Summe aller } n_i * p^2$ , (e) berechnen Chi-Quadrat =  $(1 / (p * q)) * (\text{Ergebnis [a]} \text{ minus Ergebnis [d]})$ , und vergleichen abschließend das beobachtete Chi-Quadrat mit dem Grenzwert.<sup>74</sup>

Übungsaufgabe: Unterliegt die Anzahl der 100-Cent-Stücke in den Portemonnaies unserer Datenserie „Muenzen“ einer Binomialverteilung?

#### 24.4 NEGATIVE BINOMIALVERTEILUNG

Definition: Die „Negative Binomialverteilung“ tritt auf, wenn  $x$  eine Poissonverteilung mit variablem Parameter  $\lambda$  hat und die Verteilung von  $\lambda$  eine Gammaverteilung ist.

Die „Negative Binomialverteilung“ kommt in der Archäologie häufig vor. Ein geläufiges Beispiel sind Typ-Häufigkeiten in Gruben, bei denen sich die Wahrschein-

---

<sup>74</sup> In dem Buch von Ihm (1978) finden sich gelegentlich Schreibfehler, seltener rechnerische oder inhaltliche Fehler. Bei Ihm 1978, 229 zu Beispiel 2 wird  $N$  als Summe aller  $x_i$  (im Beispiel: 792) berechnet; richtig steht in der Terminologie von Ihm  $N$  für die Summe aller  $n_i$  (im Beispiel: 1425). Daher beträgt das Ergebnis nicht 878,74, sondern richtig 87,77. Die hier verwendete Tabelle „Verteilung“ ist gegenüber Ihm 1978 entsprechend korrigiert. – Für diesen Hinweis danke ich I. Herzog (Bonn).

lichkeit  $\lambda$  von Grube zu Grube ändert, z. B. weil die Verfüllungszeiträume unterschiedlich sind und somit das Repertoire in der lebenden Kultur, aus dem die „Stichprobe“ Abfall für die Gruppe bezogen wurde, bereits wieder anders zusammengesetzt war. Eine weitere geläufige Ursache sind Ansteckungsverteilungen, die durch das Zerbrechen von Funden entstehen: Haben die Gefäßhäufigkeiten (im Sinne „Gefäßeinheiten“ / MIZ „Mindestindividuenzahlen“) in den Gruben selbst eine Poissonverteilung, liegen jedoch als Bruchstücke vor (Töpfe nicht zusammengesetzt; keine MIZ ermittelt), kann die Häufigkeit der Bruchstücke einer Negativen Binomialverteilung folgen.

Statistisch wird die Negative Binomialverteilung durch die Parameter „ $k$ “ und „ $p$ “ beschrieben: Der Erwartungswert ("Mittelwert") ist:  $E_x = \mu = k * p$ . Die Varianz ist:  $var(x) = \sigma^2 = k*p + k^2*p^2 = k * P * q$ . Demnach ist bei der Negativen Binomialverteilung die Varianz  $\sigma^2$  grösser als der Mittelwert  $E_x$ .

Die o. g. Tabelle „Verteilungen“ gibt das Beispiel bei Ihm (1978, 78 ff. Tab. 5-G) wieder und erleichtert das praktische Nachvollziehen. Bei Ihm (1980) wird beschrieben, wie in den Fällen vorgegangen kann, in denen die Häufigkeit der Null-Beobachtungen nicht gegeben ist – ein in der Archäologie oft vorkommender Fall.

## 25 TEIL IV: ANWENDUNGSBEISPIELE

---

Damit schließe ich den Teil Einführung in die Werkzeuge und die Theorie grundlegender Verfahren. Schwimmen lernt man nicht auf dem Trockenen, sondern im Wasser. Daher bin ich der Überzeugung, dass eine Fortführung hier – etwa im naheliegenden Sinne „multivariate Verfahren“ und Arbeiten mit **R** ohne Benutzeroberfläche, d. h. das direkte Coden – für die Mehrheit der Leser / Nutzer nicht zielführend wäre. Wichtiger ist es m. E., durch praktische Anwendungen an eigenen Fragestellungen und auch eigenen Daten nach bislang Erlernte zu festigen und Routine zu gewinnen. Denn der Kern kreativen selbständigen Arbeitens liegt nicht zuvorderst im Programmieren-Können mit **R** oder sicherer Theoriekenntnis auch sehr selten anwendbarer Verfahren, sondern darin, in diesem Dreikampf zwischen Archäologie, Statistik und Werkzeugen erfolgreich inhaltlich Denken zu können – Routine zu gewinnen im richtigen Stellen angemessener Fragen, im zielführenden Anpassen der Daten, und Sicherheit darin zu gewinnen, wie man das Errechnete inhaltlich und historisch deutet.

Daher möchte ich nun im vierten und letzten Teil dieser Einführung exemplarisch einige Anwendungsbeispiele durchgehen. In keinem Fall handelt es sich um Forschung im engeren Sinne, stets aber um Fragestellungen und Daten, die echt sind. Auch wenn es am Ende des einen oder anderen Beispiels seriöse Ergebnisse geben sollte: diese sind nicht als Forschungsbeiträge gedacht. Ziel dieser ausgewählten Fallstudien ist es vielmehr, mögliche „gute Wege“ aufzuzeigen, wie man von üblichen Fragestellungen und echten Daten ausgehend mit quantitativen Methoden zu seriösen Ergebnissen kommen kann. Im Zentrum der folgenden Beispiele steht nun nicht mehr die Bedienung der Software und bisweilen auch nicht die statistischen Verfahren, sondern das Erarbeiten eines besseren Verständnisses der Argumentationswege und Vorgehensweisen. Dabei ist es nicht notwendig, allen meinen dargelegten Wegen und Entscheidungen zuzustimmen; wichtig ist aber zu sehen, welche Überlegungen jeweils nötig sind, dass Entscheidungen zu treffen sind und dass man diese bei eigenen Arbeiten und Publikationen jeweils transparent darzulegen hat. Zum besseren Nachvollziehen ist es gewinnbringend, die folgenden Beispiele nicht einfach zu lesen, sondern sie anhand der mitgegebenen Daten und ihrer nunmehr gereiften PAST- oder R-Fähigkeiten (oder ihrer BlueSky, JAMOVI, JASP... -Kenntnisse) aktiv nachzuverfolgen.

## **25.1 MEHR GENAUGKEIT MIT EINFACHEM GPS**

Die erste Fallstudie gilt einem vielen Archäologen geläufigen Problem und setzt bewusst keine komplexe Software ein, sondern bedient sich allein der Fähigkeiten einer Tabellenkalkulation. Das Beispiel verdeutlicht, dass in manchen Fällen auch ein solch einfaches Instrument sehr zielführend sein kann, und dass Rechnen eine Stärke von Tabellenkalkulationen ist.

Mit der Technik des Jackknifing kann, die Fragestellung unterstützend, bei konkreten Projekten die Zuverlässigkeit der hier beschriebenen Ergebnisse weiter abgeschätzt werden.

Die Lektüre resp. das Durcharbeiten von Kap. 25.1 ist nur dann sinnvoll, wenn man z.B. als Archäologe in der Feldpraxis mit GPS-Positionierungen arbeitet. Wen das nicht betrifft oder näher interessiert, springe zu Kap. 25.2.

### **25.1.1 Anlass, Fragestellung**

Das inhaltliche Ausgangsproblem war die Mitwirkung des Autors an einem Survey- und Sondage-Projekt im außereuropäischen Ausland. Das Projekt war äußerst knapp finanziert, zeitlich auf ca. zwei Wochen Feldarbeit limitiert, teure technische Geräte standen selbst leihweise nicht zur Verfügung; vor Ort waren keine topographischen Festpunkte in akzeptabler Nähe zum Einsatzort verfügbar. So konnte im Feld – mit

den üblichen Instrumenten – nur ein exaktes lokales Messnetz eingerichtet werden. Dessen absolute Lokalisierung erfolgte mit einfachen Outdoor-GPS-Geräten. Bleibt die Frage: wie genau ist diese Einmessung?

Bei einer Beantwortung dieser Frage scheinen sich Alltagserfahrungen und technisch-wissenschaftliche Literatur zu widersprechen. Die auf manchen GPS-Geräten ablesbaren Messgenauigkeiten liegen oft im Bereich von z. B. 3 Metern und meine Autonavigation trifft zuverlässig sogar die richtige Straßenseite, auf der sich mein PKW befindet. Abweichend davon sind die Angaben in den Gebrauchsanweisungen der Geräte deutlich zurückhaltender. Diese beruhen, richtigerweise, auf der vom „Hersteller“ des GPS-Systems, dem US-amerikanischen Verteidigungsministerium, garantierten Leistung des Systems;<sup>75</sup> danach weisen bei den üblichen Messungen anhand des L1-Signals 95 % aller Ablesungen einen horizontalen Lage-Fehler von kleiner 7,8 m auf, was einem sog. RMS-Wert (Begriffsklärung in Kap. 25.1.5) von 4,0 m entspricht. Teure Geräte, die ein weiteres Signal der Satelliten auslesen und sog. dGPS-Systeme sind zweifellos genauer, doch die Ein-Frequenz-Geräte auf Basis des L1-Signals sind die übliche Technik im Outdoor-Bereich, in der Auto-Navigation sowie den Smartphones, d. h. im Kostensegment unter 1.000 Euro. Vom Betreiber der GPS-Satelliten unabhängige Messungen bestätigen das Einhalten der genannten Lagerichtigkeit.<sup>76</sup>

Manche Alltagserfahrung deutlich höherer Richtigkeit der Positionsbestimmung beruht u. a. auf Kniffs und (programmiertes) Wissen jenseits der reinen GPS-Messdaten, die z. B. in Autonavigationssystemen angewendet werden: Die Straßen auf den dort hinterlegten Karten sind nicht maßstäblich, sondern breiter; Fahrzeugdaten werden hinzugenommen und interpoliert, wie man z.B. bei jeder Fahrt durch einen Tunnel (in dem man ja keinen Empfang hat) erleben kann. Es ist erhellend, mit solch einem System eine Bahnfahrt zu unternehmen: Immer, wenn der Schienenweg nahe an einer Straße vorbeiführt, springt die Lokalisierung auf die Straße, schließlich „weiss“

---

<sup>75</sup> USA Department of Defense (2008). *GPS Standard Positioning Service (SPS): Performance Standard, 4th Edition, Sept. 2008*, table 3.4-1. <http://www.gps.gov/technical/ps/> [24.1.2020].

<sup>76</sup> z.B. Djebeniani, M. (2015). *Untersuchung der Positionsgenauigkeit von GPS mit preisgünstigen GPS-Empfängern und auf Basis der Open Source Software „RTKLIB“*. BA-Arbeit Informatik, Hochschule Bonn-Sieg-Kreis 2015. <http://mc-lab.inf.h-brs.de/projects/lokalisierung/2015-12-09-Djebeniani,Moetez.pdf> [24.1.2020]. – McNamara, J. (2008). *GPS for Dummies*. 2nd ed. Hoboken: Wiley.

das System doch, dass Autos auf Straßen fahren und dort in Fahrtrichtung stets auf der rechten Spur...

### 25.1.2 Erste Eindrücke

Also erst einmal ein kurzer Blick in die Theorie. Die einfachen GPS-Messungen beruhen auf der Bestimmung von Zeitdifferenzen zwischen Signalen mehrerer Satelliten. Viele Werte, die in diese Berechnungen eingehen, sind nicht fix, sondern leicht variabel: Die Uhren der Satelliten und die Bahnen der Satelliten am Himmel unterliegen Schwankungen. Wie viele Satelliten für das messende Gerät aktuell „sichtbar“ sind und wie diese Satelliten geometrisch in Bezug zum Gerät stehen (spitze Winkel vs. stumpfe Winkel), sind wichtige und zugleich variable Faktoren. Ab vier verfügbaren Satelliten sind Positionsbestimmungen möglich, je mehr Satelliten vom Gerät empfangen werden, desto besser. Satelliten, die – vom Messgerät aus „gesehen“ – einander scheinbar sehr nahe liegen (spitze Winkel), sind weniger nützlich als solche, die in guter Entfernung zueinanderstehen (stumpfe Winkel). Satelliten, die flach über dem Horizont stehen, liefern potenziell stärker fehlerbehaftete Werte. Diese Faktoren können während des Messens abgeschätzt werden, da das GPS-Gerät die Anzahl der Satelliten und ihre relative Lage zueinander wahrnimmt. Aus dieser Konstellation ergibt sich eine Abschätzung der aktuell theoretisch möglichen Messgenauigkeit, und diese wird bei vielen Geräten während der Nutzung angezeigt; häufig kann man hier Werte zwischen 3 und 8 Metern ablesen.

Daneben gibt es jedoch Störungen, die von einfachen Geräten nicht vorab oder während des Messens ermittelt werden können. Die Ausbreitung der Signale in der unteren Atmosphäre – der Bereich zwischen Erdoberfläche und etwa 11 km Höhe – schwankt mit dem Wetter, doch ist dessen Einfluss nicht groß. In der äußeren Atmosphäre – der sogenannten Ionosphäre – in 80 bis 400 km Höhe über der Erdoberfläche haben jedoch nicht vorhersehbare Schwankungen kräftige Einflüsse auf die Messungen. Diese Faktoren gehen in die Genauigkeitsabschätzung der GPS-Geräte während des Messvorgangs nicht ein. Daher ist der Lagefehler in einem während des Messens unkalkulierbaren Ausmaß größer als die Genauigkeitsanzeige der Geräte während des Messens.

Eine Möglichkeit, deutlich höhere Genauigkeit zu erreichen, beruht auf dem Empfang und der Auswertung eines zweiten Signals, dem sog. L2-Signal. Denn aus dem Abgleich des L1- und L2-Signals können Informationen über das Ausmaß der Signalstörungen in der Ionosphäre gewonnen werden. Doch dies ist den professionellen und teuren Geräten vorbehalten (Kosten m.W. >7.500 Euro), die damit Lagerichtigkeiten im Bereich von wenigen Dezimetern (wenigen Zentimetern?) erreichen.

Eine andere Möglichkeit beruht darauf, an festen Stationen exakt bekannter Lage auf der Erde fortlaufend die GPS-Daten zu erfassen und daraus auf die Fehler der Uhren und der Bahnen der Satelliten sowie die Störungen in der Atmosphäre zurück zu schließen. So können Korrekturfaktoren ermittelt werden, mit denen die tatsächlichen Messungen anderer, beweglicher Geräte („Rover“) aktuell – z.B. via Funkkontakt – oder nachträglich geeicht werden können. Darauf beruhen viele der sogenannten dGPS-Techniken (Differential Global Positioning System), die jedoch für mein o.g. Projekt ob der hohen Kosten nicht zur Verfügung standen. Derzeit werden diese Geräte, Stichwort RTK-Technik (real time kinematic), zunehmend handlicher, preiswerter und erreichen Lagerichtigkeiten in der Größenordnung von Zentimetern. Im Bereich der Landwirtschaft dienen sie zur Steuerung von Feldgeräten. Sehr preiswerte Lösungen scheinen aktuell (2019) gegen ca. 1.000 Euro käuflich zu sein, doch fehlt es mir an praktischen Erfahrungen damit.

### **25.1.3 Begriffsklärung Genauigkeit: Präzision und Richtigkeit**

Ein von vielen kleinen Störungen beeinflusstes dynamisches System wird nur selten den „wahren“ Wert liefern, sondern meist Beobachtungen, die um diesen Wert streuen. Die Beobachtung dieser Streuung kann sich auf zwei unterschiedliche Dinge beziehen, nämlich die systemimmanente Genauigkeit und die Genauigkeit in Bezug zur tatsächlichen Lage auf der Erde. So ist es durchaus vorstellbar, dass eine längere Folge von Messwerten untereinander eine geringe Streuung aufweist und damit eine hohe immanente Reproduzierbarkeit aufweist, der resultierende mittlere Wert jedoch kräftig von der tatsächlichen Lage abweicht. Daher ist die übliche Vereinfachung des Begriffs Genauigkeit unglücklich. In der Fachwelt wird unterschieden in Richtigkeit (*accuracy*) und Präzision (*precision*). Mit dem Begriff Präzision wird formuliert, wie homogen und reproduzierbar die Messergebnisse untereinander sind, d. h. wie weit sie um einen Mittelwert streuen, der aber selbst durchaus vom wahren Wert entfernt sein kann. In der englischsprachigen Literatur wird dies auch DOP (*dilution of precision*) genannt. Unter Richtigkeit hingegen wird formuliert, wie nah der Wert oder Mittelwert einer Positionsbestimmung via GPS der tatsächlichen Lage auf der Erde kommt. Die Angaben auf den Geräten während des Messens kalkulieren die Präzision (DOP), die vorsichtigeren Angaben der Hersteller in den Gebrauchsanweisungen beruhen auf der vom Betreiber der GPS-Satelliten garantierten und für den Anwender weitaus relevanteren Richtigkeit.

#### 25.1.4 praktische Anwendung der Präzision (DOP)

Die Konstellation der Satelliten ist auf eine gewisse Zeit vorhersehbar, so dass für jeden Ort der Welt eine systembedingte Erwartung an die Präzision – nicht an die Richtigkeit – errechnet werden kann. Im dichten Wald und in städtischen Straßenschluchten wird diese Erwartung nicht zu realisieren sein, zudem gehen unkalkulierbare Störungen etwa aus der Ionosphäre und durch Signalreflektionen (Häuserwände, Dächer) nicht in diese Kalkulation der DOP ein. Dennoch kann die vorausschauende Wahrnehmung der systembedingten Möglichkeiten nützlich sein. Der GPS-Gerätehersteller Trimble beispielsweise bietet frei ein gut bedienbares Programm an, das die nötigen Daten lädt und die Berechnungen für auswählbare Bereiche der Welt durchführt.<sup>77</sup> Dies kann zur Planung von Unternehmungen nützlich sein, bei denen man gute Bedingungen braucht; denn anhand der Voraussagen kann man seine Messungen gezielt in jene Zeitabschnitte legen, für die für einen bestimmten Ort relativ niedrige DOP-Werte vorausgesagt werden. Dies kann auch für das rückwärtsgewandte Verständnis von Aufzeichnungen (z. B. den Spuren / „tracks“ einer Reise) wertvoll sein, die eventuell teilweise irritierende Beobachtungen geliefert haben.

#### 25.1.5 Fehlerangaben: Standardabweichung und RMS

##### *Präzision & Standardabweichung*

Präzision und Richtigkeit werden statistisch beschrieben durch Kennzahlen, die angeben, wie groß der übliche Schwankungs- oder Fehlerbereich ist. Die Präzision, d. h. die Reproduzierbarkeit der Messungen, ergibt eine Streuung der Messwerte um eine ideal gedachte Glockenkurve (richtiger: Glocke, da räumlich zu denken); sie beschreibt den Umstand, dass die Mehrheit der Beobachtungen nahe am Mittelwert liegen und wenige Beobachtungen davon stärker abweichen. Man spricht dann von einer Normalverteilung. Die geeignete statistische Kennzahl für solche Phänomene ist die Standardabweichung (Std.abw.), innerhalb der 68,3 % aller Beobachtungen liegen („1-sigma-Bereich“). Man beachte, dass hierbei fast ein Drittel der Beobachtungen außerhalb dieser Schranke liegen. Benötigt man eine höhere Sicherheit der Aussage, wird die Spanne folglich vergrößert. Bei einer Normalverteilung liegen 95 % aller Beobachtungen im Bereich 1,96 mal 1-sigma, und 99 % aller Beobachtungen im Bereich 2,58 mal 1-sigma. Das Bedürfnis nach höherer Sicherheit der Aussage vergrößert die Spanne also beträchtlich. In der Statistik ist es üblich, mit der Standardabweichung (1-

---

<sup>77</sup> Trimble: “GPS Data Resources” (unter “Support”): [www.trimble.com](http://www.trimble.com) [24.1.2020].

sigma-Schranke; 68,3 %) zu arbeiten, häufig wird auch die 1,96-Schranke mit 95 %iger Wahrscheinlichkeit angegeben.

### *Richtigkeit & RMS-Wert*

Die Angaben auf den GPS-Geräten während des Messens beziehen sich auf die Abschätzung der Präzision der Messungen, denn diese ist systemimmanent möglich. Die Richtigkeit der Messungen hängt neben der systemimmanenten Präzision aber auch von anderen Faktoren ab, weshalb sich statistisch eine komplexere Wahrscheinlichkeit als beim einfachen Modell von Normalverteilung und Standardabweichung ergibt. Zudem ist der Fehler bei der Frage nach der Lagerichtigkeit – entgegen dem Konzept einer Normalverteilung / Glockenkurve – extrem asymmetrisch, denn ein Lagefehler kann zwar vereinzelt auch groß ausfallen (Abweichung nach oben), aber er kann nie kleiner als Null sein (Abweichung nach unten). Daher ist das (symmetrische) Maß „Standardabweichung“ für die Frage nach der Richtigkeit einer Lagebestimmung untauglich. Es muss ein anderes Streuungsmaß benutzt werden, für das es in der deutschen Sprache leider keinen passenden Fachbegriff gibt. Er beschreibt die Abweichung der einzelnen Werte von der tatsächlichen Lage (Richtigkeit). In der englischsprachigen Literatur spricht man vom *root mean squared error* (RMS). Der RMS-Wert berechnet sich als Wurzel aus dem Quadrat des mittleren (Lage-) Fehlers plus dem Quadrat der Standardabweichung des Fehlers. Ein alternativer, einfacherer Berechnungsweg lautet:

RMS = Wurzel aus (Summe aller quadrierten Fehler, dividiert durch Anzahl Beobachtungen).

Innerhalb des Bereichs 1 RMS liegen 68,3 % aller Beobachtungen, eine Sicherheit von 95 % erreicht man bei 1,96 RMS. Die Sicherheit von 95 % bei 1,96 RMS wird in der englischsprachigen Literatur gelegentlich verkürzend und nicht ganz exakt als ‘2rms’ oder ‘drms’ (*double rms*) notiert. Man beachte, dass bei der üblichen Angabe von 1 RMS knapp ein Drittel aller Bestimmungen außerhalb dieses Bereichs liegen.

Die Garantie der Betreiber des GPS-Systems bezieht sich auf die Richtigkeit (!) der Positionsbestimmung; sie besagt, dass seit Ende 2008 95 % aller Beobachtungen sicher horizontal innerhalb eines Umkreises von 7,8 m um den richtigen Wert liegen – es wird also der dRMS-Wert angegeben, entspricht ca. 4,0 m RMS. Die vertikale Richtigkeit ist in etwa um den Faktor 1,7 geringer, d. h. sie liegt bei etwa 13 m um den angezeigten Wert. Diese von den Betreibern des US-amerikanischen GPS-Systems garantierten Richtigkeiten spiegeln sich in den Gebrauchsanweisungen der Geräte-Hersteller wider, deren Angaben sind also weniger gerät-, sondern system-spezifisch. So gibt

das Handbuch meines „Garmin eTrex H“ (Baujahr 2008) als Richtigkeit (*accuracy*) an „<10 meters (33 ft) RMS“ und meint damit, dass 68,3 % aller Beobachtungen näher als 10 Meter am wahren Lagepunkt liegen; zum „Garmin eTrex Vista“ (Baujahr 2003) heißt es noch skeptischer „<15 meters (45 ft) RMS“.

### 25.1.6 Signale, Geräte und Prozessoren

Die Signale der GPS-Satelliten werden auf zwei Frequenzen ausgestrahlt. Auf der L1-Frequenz (1575,42 MHz) werden der C/A-Code („coarse/acquisition“) für die zivile Nutzung und versetzt dazu der nicht öffentlich bekannte P/Y-Code („precision/encrypted“) für die militärische Nutzung eingesetzt. Das L1-Signal enthält alle wichtigen Informationen und benötigt zur Übertragung eine halbe Minute. Zusätzlich werden zur Initialisierung der Geräte auch die so genannten Almanach-Daten übertragen, welche die voraussichtlichen Bahndaten aller Satelliten enthalten. Ihre Übertragung benötigt gut zwölf Minuten, weshalb ein Empfänger unter bestimmten Umständen (insbes. wenn längere Zeit nicht benutzt oder nach starkem Ortswechsel) deutlich längere Zeit braucht, um gute Lagebestimmungen zu liefern. Daher wird man in solchen Fällen exaktere Messungen erst nach ca. 12 Minuten erhalten. Die zweite Frequenz (L2-Frequenz; 1227,60 MHz), überträgt nur den P/Y-Code. Geräte, die zusätzlich diese zweite Frequenz empfangen, können aus dem Abgleich der L1- und L2-Signale die Störungen in der Ionosphäre ermitteln, was die Präzision der resultierenden Lagebestimmung erheblich steigert.

Die im Freizeitbereich üblichen Geräte beobachten durchweg nur die L1-Signale. Die in diesen Geräten verbauten verschiedenen Prozessoren unterscheiden einander in der Richtigkeit nur wenig. Im Vergleich zu älterem Gerät sind z.B. die modernen Chips deutlich schneller und empfindlicher, d. h. sie liefern nach dem Einschalten früher erste Positionsbestimmungen und sie empfangen schwächere Signale, was beispielsweise im Wald von Vorteil ist. Die höhere Empfindlichkeit führt jedoch tendenziell zu eher etwas unrichtigeren Ergebnissen. Denn außer den direkt von den GPS-Satelliten empfangenen Signale messen die empfindlicheren Geräte gerade im Bereich von Städten oder in den Bergen auch reflektierte Signale; es kommt zu sog. Mehrwegeeffekten, welche die Richtigkeit der Ergebnisse verringern.

Im Bereich des Vermessungswesens eingesetzte Geräte, die auch die L2-Signale empfangen und verarbeiten, kosten m. W. in einer Größenordnung von 5.000 bis 20.000 Euro und erreichen dann ohne weiteren (aber möglichen) Aufwand Richtigkeiten im Dezimeter- oder Zentimeterbereich. Ähnliche Ausstattungen werden zunehmend in der Landwirtschaft eingesetzt, um Ackergeräte präzise zu steuern; hier scheinen die Hardwarepreise in der Größenordnung von 10.000-20.000 Euro zu liegen, hinzu

kommen jährliche Lizenzgebühren für Korrektursignale (sog. RTK-Netzwerke) von etwa 2.000-3.000 Euro.

### 25.1.7 WAAS und EGNOS

Hinter den Abkürzungen WAAS (*wide area augmentation system*) und EGNOS (*European geostationary navigation overlay service*) verbergen sich regionale Dienstleistungen, die die Richtigkeit der Positionsbestimmungen erhöhen können. Beide beruhen auf der Bereitstellung von Korrektursignalen an die Nutzer. An Referenzstationen exakt bekannter Lage werden die Satelliten beobachtet und fortlaufend ein Soll-Ist-Vergleich durchgeführt. Auf dieser Grundlage werden Korrektursignale bereitgestellt, die über Funk oder Mobiltelefon verteilt werden. Im Empfänger entsteht daraus die Möglichkeit eines dGPS, die resultierenden Richtigkeiten von Geräten auf Basis des L1-Signals sollten dann bei 1 bis 3 m liegen. Der US-amerikanische Dienst WAAS steht nur in Nordamerika zur Verfügung. Das europäische EGNOS wurde 2009 im Testbetrieb gestartet und ist – bei andauernd weiterem Ausbau – seit März 2011 im Regelbetrieb. Die unter Nutzung der EGNOS-Signale erreichbare Lagerichtigkeit in Europa liegt – ähnlich wie beim WAAS-Dienst – nach der Selbstauskunft von EGNOS bei 3 m (European GNSS Agency: „Performance overview“). Höhere Richtigkeiten bis in den Zentimeter-Bereich sind möglich, aber um den Preis einer aufwändigeren Hardware und vor allem gegen Zahlung einer Gebühr für die nicht-öffentliche, präzisere Variante der EGNOS-Signale.

### 25.1.8 Empirie & Praxis: Wie genau misst mein GPS?

Es liegt nahe, für die Frage nach der Genauigkeit der GPS-Positionsbestimmungen von den Angaben der Hersteller auszugehen: ein RMS-Wert von kleiner 7,8 m (dRMS). Doch wer zweifelt, dass dies überhaupt erreicht wird, oder vermutet, ein überdurchschnittlich genaues Gerät zu besitzen, kann dies selbst prüfen: man legt das Gerät auf einen Punkt bekannter Lage, misst und ermittelt die Differenz. Genau das habe ich in den Jahren 2009-2010 systematisch getan. Der hier beigefügte Datensatz „6\_GPS-reihe24h“ zeigt ein Beispiel, bei dem das Gerät über 24 Stunden am gleichen Platz bekannter Koordinaten lag und alle 10 Sekunden einen Messwert aufzeichnete. Die Tabelle hält zugleich die nötigen Vorkehrungen und Rechenschritte zur Ermittlung von RMS bzw. dRMS fest. Wer dies am selbstgewählten Ort mit seinem persönlichen Gerät nachvollziehen möchte, orientiere sich am folgenden Rezept:

- Für die Messungen sollte man das vor Ort übliche rechtwinklige Koordinatensystem verwenden, d. h. das GPS-Gerät entsprechend einstellen. Denn mit rechtwinkligen

Koordinaten sind die nötigen Berechnungen leichter durchführbar als mit den Angaben in Grad, Minuten und Sekunden. In Deutschland sind das nicht mehr die früher üblichen Gauss-Krüger-Koordinaten, sondern das UTM-Gitter mit dem „Datum“ ETRS89 (was weitgehend WGS84 entspricht), auf das die Landesämter in den 1990er-Jahren schrittweise umgestellt haben. Dabei ist nicht die Wahl des Netzes entscheidend, sondern dass das Messgerät im gleichen System misst, in dem auch die bekannten Koordinaten des gewählten Festpunktes vorliegen, denn die andernfalls nötigen Umrechnungen sind eine mögliche Fehlerquelle.

- So möglich, liest man die Messungen nicht manuell ab und schreibt die Resultate auf, sondern stellt das Gerät so ein, dass es die Messungen aufzeichnet, damit man sie später auslesen kann. Wie das praktisch geht, ist gerätespezifisch und kann daher nicht allgemein beschrieben werden. Im Falle meiner Geräte konnten \*.csv-Tabellen ausgelesen werden, die wiederum mit MS-Excel oder LO-Calc gelesen und weiterverarbeitet werden können.
- In die resultierende Tabelle werden dann nachträglich auch die tatsächlichen Koordinaten des gemessenen Punktes eingetragen. Anschließend wird die Distanz beobachtet und erwartet berechnet, und zwar in zwei Schritten: quadrierter Lagefehler, dann Lagefehler.

$$\begin{aligned} \text{quadrierter Lagefehler} = \\ & ((\text{gemessenNord} - \text{Soll-Nord}) * (\text{gemessenNord} - \text{Soll-Nord}) + \\ & (\text{gemessenOst} - \text{Soll-Ost}) * (\text{gemessenOst} - \text{Soll-Ost})) \end{aligned}$$

Lagefehler = Wurzel aus quadriertem Lagefehler.

- Aus diesen Werten kann man weiteren Kennzahlen errechnen. Mit Hilfe der statistischen Funktionen z.B. Mittelwert und Standardabweichung des Lagefehlers, die im Grunde die Präzision der Messungen darlegen. Im gegebenen Fall ergaben sich (typische) Werte um 2,5 m – die aber kaum Aussagekraft über die Lagerichtigkeit haben.
- Lagerichtigkeit: Der RMS-Wert ergibt sich aus der Summe der Quadrierten Lagefehler, dividiert durch die Anzahl der dokumentierten Messungen. RMS mal 1,96 ergibt den dRMS-Wert, innerhalb dessen sich 95 % aller Beobachtungen bewegen.

Im hier dokumentierten Fall liegt dRMS bei 4,89 m, d.h. der tatsächliche Lagefehler war im Mittel etwas kleiner als die Herstellergarantie. Wie der ebenfalls ausgewiesene Maximalwert zeigt, wurde allerdings auch ein Fehler von 7,7 m erreicht.

Die hier verwendete Serie wurde unter optimalen Bedingungen ermittelt: in der Stadt auf einem relativ zur Umgebung hohen Flachdach mit freier Himmelssicht,

ohne dass es zu Verdeckungen oder starken Reflektionen von nahegelegenen höheren Gebäuden bzw. Dächern hätte kommen können.

Anhand von 40 solcher Serien, an denen jeweils zwischen 1 und 24 Stunden an verschiedenen Tagen in den Jahren 2009 und 2010 solche Messungen durchgeführt wurden, ergab sich ein mittlerer dRMS-Wert von 5,22 m (Std.abw. 1,55 m), in einem Fall allerdings auch eine Serie mit dRMS von 11,1 m. Meine Bilanz: die Herstellergarantie wird von den Einzelmessungen fast durchweg unterschritten, jedoch weniger stark, als oft vermutet wird. Die Messungen fanden mit vier verschiedenen Geräten statt. Die Fehler der vier Geräte unterschieden einander nicht signifikant, vielmehr sind die Ergebnisse der verschiedenen Geräte im Rahmen der Zufallsstreuung identisch.

### 25.1.9 Statische Messungen

Richtig: es geht genauer! Für archäologische Einmessungen z. B. bei einem Survey kann man sich den Umstand zunutze machen, dass viele Messungen an der gleichen Stelle möglich sind, ohne dass das Gerät bewegt wird. Aus vielen Einzelmessungen kann jeweils ein Mittelwert der Nord- und Ostkoordinate gebildet werden. So lässt sich an der hier dokumentierten Serie (nachrechnen!) z.B. ein Mittelwert über alle Messungen aus 24 Stunden bilden, der im Ergebnis einen Lagefehler von 1,89 m zu den tatsächlichen Koordinaten aufweist. Dieser Fehler von ca. 2 m ist erheblich kleiner als der sich aus der gleichen Serie, betrachtet als Einzelmessungen, ergebende Lagefehler von um 5 m dRMS.

Nun sind 24 Stunden eine lange Zeit, weshalb sich die Frage stellt, wie es um das Verhältnis von Aufwand und Nutzen steht. Aus den erwähnten 40 Serien von statischen Messungen ergab sich 2009 – 2010 der in Tabelle **Abb. 101** zusammengestellte Befund, wonach mit der Messdauer die Richtigkeit zunimmt, der Effekt jedoch nicht sehr stark ist. M. E. reicht für statische Messungen eine Dauer von 20, besser 30 Minuten aus, um mit ziemlicher Sicherheit eine Lagerichtigkeit von unter ca. 4 m zu erreichen. Wird mehr Sicherheit des Ergebnisses angestrebt, sind mit 4 Stunden Messdauer die Möglichkeiten dieses Verfahrens weitgehend erreicht, mehr Aufwand lohnt nach meinen Beobachtungen nicht. Nach dem hier vorgestellten Weg ist es jedem Interessierten möglich, mit seinen Instrumenten unter aktuellen Bedingungen vergleichbare Tests durchzuführen.

Messdauer	dRMS
5 min, 10 min, 20 min	ca. 4,3 m
30 min, 1 h, 2 h	ca. 4,1 m
4, 6, 8, 12, 24 h	ca. 3,7 m

**Abb. 101** Dauer statischer GPS-Messungen und resultierende Lagerichtigkeiten (dRMS) nach zahlreichen Messreihen des Autors in den Jahren 2009-10.

### 25.1.10 Lokale Messnetze mit Netzausgleichsrechnungen

Die eingangs geschilderte Herausforderung von „einfachem Gerät“ und „Survey / Sondage weitab erreichbarer Festpunkte“ beinhaltet indes eine Komponente, die mit einer gegenüber einzelnen statischen Messungen weiteren Steigerung der Lagerichtigkeit einhergehen kann. Denn vor Ort wurde ein in sich stimmiges und genaues lokales Messnetz eingerichtet. Folglich konnten mehrere lokal bekannte Punkte jeweils per statischen GPS-Messungen lokalisiert werden. Dies ermöglicht die Fixierung des Messnetzes als Ganzes, wobei die systemimmanente Lageungenauigkeit einzelner GPS-Lokalisierungen ausgeglichen werden kann.

Bei diesbezüglichen Versuchen mit exakt lagebekannten Punkten habe ich mit lokalen Messnetzen, deren sechs einzelne Punkte um 20 bis 25 m auseinander lagen, einen dRMS-Wert von um 3,0 m erreicht, und mit lokalen Netzen, deren drei (!) Punkte um 500 m auseinander lagen, einen dRMS-Wert von 2,2 m – wobei für die resultierende Lagerichtigkeit weniger die Anzahl der Punkte, sondern vor alle deren Entfernung zueinander relevant ist: weite Abstände sind vorteilhaft. Da in der archäologischen Feldpraxis, z. B. bei einem Survey, ein lokales Netz mit vier oder sechs Punkten mit sehr einfachen Mittel aufgebaut werden kann, ist auf diesem Wege (einfaches GPS, lokales Netz, 50 m Netzweite, statische Messungen á 30 min, Netzausgleichsrechnungen) eine Lagerichtigkeit von um 2,7 m, sicher unter 3,0 m erreichbar.

Interessiert, wie das geht? Beginnen wir mit dem einfachen, praktischen Teil: lokales Netz mit einfachsten Mitteln einrichten. In ebenem Gelände benötigt man nur zwei Maßbänder á 50 m und – als kleine Arbeitserleichterung – vier Ringnadeln. Die theoretische Grundlage ist der Satz des Pythagoras, nach dem in einem rechtwinkligen Dreieck (und nur in diesem) gilt:  $a^2 + b^2 = c^2$ . Wir stecken eine Strecke von exakt 50 m Länge ab und fixieren an deren Ende zwei Ringnadeln. In Anwendung der genannten Formel errechnen wir, dass – von diesen beiden Nadeln ausgehend – zwei gleich lange Dreiecks-Seiten á 35,355 m an ihrem Treffpunkt (Ringnadel setzen!) in einem rechten Winkel zueinander liegen. Wir wiederholen diese Übung in die andere Richtung der langen Seite des Dreiecks und setzen die vierte Ringnadel: die vier Punkte bilden jetzt ein exaktes Quadrat mit den lokalen Koordinaten (z. B.) 100 / 100 m, 100

/ 150 m, 150 / 150 m und 150 / 100 m. Bereits dieses einfache lokale Netz ermöglicht nun das Arbeiten mit Netzausgleichsrechnungen. Alternativ kann man statt des beschriebenen Quadrats auch ein exakt rechtwinkliges Rechteck aufbauen, indem man nach dem Abstecken der langen 50-m-Linie statt mit 2 Strecken á 35,255 m mit den Streckenlängen 30 m und 40 m arbeitet.

#### ***25.1.10.1 Tabelle „Netzausgleich\_Helmert“***

Der nachfolgend dargelegte Rechenweg ist etwas komplexer. In der zugehörigen, hier beigegefügteten Tabelle „Netzausgleich\_Helmert“ ist das Beschriebene fertig umgesetzt, so dass der Text besser nachvollzogen werden kann. Die Zahlen in dieser Tabelle sind gänzlich fiktiv. Die Tabelle kann für eigene Versuche und Anwendungsfälle angepasst werden. Die Umsetzung via Tabellenkalkulation hat den Vorteil, dass man in der Praxis z. B. einzelne Punkte oder Messungen, die ungewöhnlich große Fehler beibringen, löschen kann. Arbeitet man mit mehreren Punkten, kann man ggf. auch anhand eines systematischen und protokollierten Jackknifings die Stabilität der Resultate ersehen.

Die in der Tabelle in grauer Schrift abgesetzte Spalte „accuracy“ [Sp. R] greift nur dann, wenn die Messungen an lagebekannten Punkten erfolgen und beides – lokales Netz und GPS-Messungen – im gleichen Koordinatensystem vorliegen. So lässt sich durch eigene Versuche (eigene Netzweite, eigenes GPS-Gerät, örtliche Bedingungen) der RMS-Wert ermitteln, den man in etwa im echten Anwendungsfall in etwa erzielt.

Die untersten Zeilen in der Tabelle ermöglichen es, weitere einzelne Punkte im lokalen Netz, für die keine GPS-Daten zur Verfügung stehen, nach den nun ermittelten Parametern ebenfalls in das globale Netz zu rechnen.

#### ***25.1.10.2 Netzausgleichsrechnung***

Nun zum Netzausgleich, der mit Hilfe einer Tabellenkalkulation gut durchführbar ist. Der nachfolgend beschriebene Weg läuft in der Fachliteratur unter dem Begriff „Helmert-Transformation“. Als einfaches fiktives Beispiel nehmen wir ein rechtwinkliges lokales Koordinatennetz mit 4 Punkten, das wie zuvor beschrieben in ebenem Gelände konstruiert wurde. Von diesen vier Punkten ausgehend errechnen wir nachfolgend Schritt für Schritt die Transformation vom lokalen zum globalen System.

<i>Sp. A</i>	Lokal Ost	Lokal Nord	UTM Ost	UTM Nord
	<i>Sp. C</i>	<i>Sp. D</i>	<i>Sp. F</i>	<i>Sp. G</i>
Punkt 1	100	100	1001,1	998,0
Punkt 2	140	100	1038,3	1001,6
Punkt 3	140	130	1040,1	1029,7
Punkt 4	100	130	998,3	1033,6

(1) Man errechnet den Mittelwert jeder der vier Spalten B bis E, z. B. mit der Excel-Funktion MITTELWERT(C1:C4), usw. Für lokal Ost z.B. ergibt sich in unserem Beispiel 120,00.

(2) Es werden vier neue Spalten (I-L) für eine Zwischenrechnung angelegt. Hier wird jeweils errechnet, welche Differenz der jeweilige beobachtete Wert zu seinem Mittelwert hat. Für Punkt 1 z.B. ergibt sich 100 minus 120,00: -20,00 [F]; ... und so fort für alle 4 Punkte und 4 Spalten.

(3) Zur Kontrolle errechne man wie in Schritt 1 den Mittelwert dieser vier neuen Spalten. Er sollte jeweils bei 0,00 liegen.

(4) Es werden aus der Zwischenrechnung (2) drei Zahlen als Zwischenergebnisse errechnet, die im Weiteren benötigt werden.

(4.I) Die Differenzen für die beiden Werte der lokalen Koordinaten (Ost, Nord) werden quadriert und die Summe dieser Quadrate berechnet. Dazu dient die Funktion QUADRATESUMME(I1:I6;J1:J6). Ergebnis hier: 326,7.

(4.II) Aus den Spalten der Differenzen werden die Differenz für den lokalen Ostwert und den GPS-Ostwert miteinander multipliziert, und dann die Summe der Ergebnisse berechnet. Am einfachsten geht dies mit der Funktion SUMMENPRODUKT(I1:I6;K1:K6).

Aus den Spalten mit den Differenzen werden die Differenz für den lokalen Nordwert und den GPS-Nordwert multipliziert, und dann die Summe der Ergebnisse berechnet. Am einfachsten geht dies mit der Funktion SUMMENPRODUKT(J1:J6;L1:L6).

Die Summe beider Zahlen wird gebildet und ist das gesuchte zweite Zwischenergebnis. Ergebnis hier: 288,4.

(4.III) Aus den Spalten der Differenzen wird die Differenz des lokalen Ostwertes mit der Differenz des GPS-Nordwertes multipliziert. Am einfachsten geht dies mit der Funktion SUMMENPRODUKT(I1:I6;L1:L6).

Aus den Spalten der Differenzen wird die Differenz des lokalen Nordwertes mit der Differenz des GPS-Ostwertes multipliziert. Am einfachsten geht dies mit der Funktion SUMMENPRODUKT(J1:J6;K1:K6).

Die zweite Summe wird von der ersten Summe abgezogen und ergibt den dritten gesuchten Zwischenwert. Ergebnis hier: 169,6.

(5) Das Ergebnis (4.II) wird durch (4.I) geteilt und ergibt den für die Transformation benötigten Faktor „a“. Ergebnis hier: 0,882959.

(6) Das Ergebnis (4.III) wird durch (4.I) geteilt und ergibt den für die Transformation benötigten Faktor „b“. Ergebnis hier: 0,519286.

(7) Die globale Ostkoordinate für den lokalen Punkt 0,00 Ost, 0,00 Nord wird errechnet, wofür wir die Mittelwerte aus Schritt 1 benötigen:

$(\text{Mittelwert GPS-Ost}) - (\text{Mittelwert lokal Ost} * „a“) + (\text{Mittelwert lokal Nord} * „b“)$ .

Ergebnis hier: 1.180,426.

(8) Die globale Nordkoordinate für den lokalen Punkt 0,00 Ost, 0,00 Nord wird errechnet, wofür wir die Mittelwerte aus Schritt 1 benötigen:

$(\text{Mittelwert GPS-Nord}) - (\text{Mittelwert lokal Ost} * „b“) - (\text{Mittelwert lokal Nord} * „a“)$ .

Ergebnis hier: 1.507,702.

Damit sind alle nötigen Transformationsparameter errechnet, so dass sich nun für jeden Punkt des lokalen Netzes die geschätzten globalen Koordinaten errechnen lassen. Dazu werden zwei neue Spalten [N, O] angelegt und die Berechnungen (9-10) für jeden der sechs Netzpunkte durchgeführt.

(9) Geschätzter globaler Ostwert [K] = globaler Ostwert aus (7) + (lokaler Ostwert \* „a“) - (lokaler Nordwert \* „b“).

(10) Geschätzter globales Nordwert [L] = globaler Nordwert aus (7) + (lokaler Ostwert \* „b“) + (lokaler Nordwert \* „a“).

(11) Als Kontrolle kann der Mittelwert der globalen Ostwerte und der Mittelwert der globalen Nordwerte dienen (zu errechnen wie in 1). Er sollte identisch sein mit dem Mittelwert der GPS-Werte in Spalte F und G.

(12) Für jeden der Punkte wird der Abstand zwischen den beobachteten GPS-Koordinaten und den nun geschätzten globalen Koordinaten berechnet [Sp. Q]:

(12.1) (GPS-Ost - global Ostwert) quadriert.

(12.2) (GPS-Nord - global Nord) quadriert.

(12.3) Wurzel aus der Summe beider Werte (12.1, 12.2).

(13) Der Mittelwert dieser Zahlenreihe [Q] wird berechnet. Im hier verfolgten Beispiel ergibt sich ca. 0,44 m.

Das Ergebnis aus Schritt (13) zeigt den mittleren internen Fehler des Netzes an, der den Schätzungen von den lokalen Koordinaten auf die globalen Koordinaten innewohnt. Die einzelnen Werte von (12) geben den Beitrag der einzelnen Punkte zu diesem mittleren Gesamtfehler an. Hat man ein Netz mit einer hinreichend großen Anzahl an Punkten, ist es möglich, einzelne Punkte mit besonders großen Fehlern wegzulassen. Meine Erfahrung aus praktischen Versuchsreihen zeigt jedoch, dass so zwar der interne Netzfehler minimiert werden kann, damit jedoch keine / kaum eine zuverlässig eintretende Verbesserung der globalen Lagerichtigkeit einhergeht - jedenfalls nicht, wenn dank statischer Messungen geschätzte globale Koordinaten in die Rechnung eingehen, die keine wirklich starken Ausreißer mehr beinhalten.

## 25.2 KULTURMODELLE UND IDENTITÄTEN QUANTIFIZIERT

Das folgende Beispiel soll *pars pro toto* aufzeigen, dass es nicht immer aufwändiger oder komplexer statistischer Verfahren bedarf, um quantitative Methoden mit Gewinn einsetzen zu können. Vielmehr dient der hier herangezogene Fall als Beispiel dafür, in der Forschung bereits vertretene Gedanken und Hypothesen in klare Fragen an das Material zu übertragen, die dazu nötigen Daten systematisch zu erheben, um dann die bestehenden Thesen prüfen, schärfen und erweitern zu können – wobei die resultierenden Ergebnisse mit vergleichsweise einfachen, d.h. weithin verständlichen quantitativen Methoden erreicht wurden. Wiewohl die Praxis im vorgestellten Beispiel eine andere war, hätten die Kernthesen mit den Instrumenten Tabellenkalkulation und Prozentrechnung erarbeitet werden können.

Die Debatte um zeitlich und räumlich umgrenzte Gruppen ist in der Archäologie alt, sie setzt nicht erst mit den heute als klassisch geltenden, immer wieder kontrovers diskutierten Thesen von Gustaf Kossinna (1896) ein.<sup>78</sup> Die anhaltend strittige Debatte um Kossinnas Thesen und „Methode“ und um den Kulturbegriff ist m. E. davon geprägt, dass archäologische Kulturen seit Kossinna mit ethnischen Gruppen gleichgesetzt wurden resp. genau diese Möglichkeit bestritten wurde. Diese Debatte übersieht, dass archäologische Kulturen – schon aufgrund ihres räumlichen Umfangs<sup>79</sup> – in den meisten Fällen keine Ethnien im Sinne eigenständig agierender soziopolitischer Einheiten sind. Soweit wir aus historischer Überlieferung „Kulturen“ kennen, waren sie stets als Oberbegriff gemeint. Caesars „Kelten“ und „Germanen“ sind keine Völker, sondern ein zusammenfassender Oberbegriff für viele selbständig organisierte Völker, und die Texte z. B. von Caesar oder Tacitus nennen vor allem die Namen dieser Völker, während die Oberbegriffe Kelten und Germanen nur selten angeführt werden. Das 1969 von F. Barth eingeführte Stichwort „Ethnicity“ (Ethnizität) hat den Unterschied zwischen Kulturen und Ethnien scharf herausgearbeitet:<sup>80</sup> Völker / Ethnien sind (wie Kulturen) einander in vielen Merkmalen selbstähnliche Gruppen, die sich dessen – im Unterschied zu Kulturen – auch bewusst sind und die sich gegenüber benachbarten Völkern und Kulturen erkennbar abgrenzen.<sup>81</sup> Kulturen hingegen können an ihren Rändern „ausfransen“, d. h. sukzessive ihre Charakteristika verlieren.

---

<sup>78</sup> Kossinna, G. (1896). Ueber die vorgeschichtliche Ausbreitung der Germanen in Deutschland. *Correspondenzblatt der Deutschen Gesellschaft für Anthropologie, Ethnologie und Urgeschichte* 26, 109-112. Vgl. dazu z. B. Fontane, Th. (1878). *Vor dem Sturm. Roman aus dem Winter 1812 auf -13*. (Kap. 11-15). <https://gutenberg.spiegel.de/buch/vor-dem-sturm-4448/12> [24.1.2020]. - Montelius, O. (1888). Ueber die Einwanderung unserer Vorfahren in den Norden. *Archiv für Anthropologie – Zeitschrift für Naturgeschichte und Urgeschichte des Menschen* 17, 151-160. Umfassend: Jacob-Friesen, K.-H. (1928). *Grundfragen der Urgeschichtsforschung: Rassen, Völker und Kulturen*. Hannover: Helwingsche Verlagsbuchhandlung.

<sup>79</sup> Wotzka, H.-P. (1997). Maßstabsprobleme bei der ethnischen Deutung neolithischer „Kulturen“. *Das Altertum* 43, 163-176.

<sup>80</sup> Barth, F. (1969). Introduction. In: F. Barth (ed.), *Ethnic groups and boundaries: The social organisation of cultural difference*. (p. 9-38). Bergen: Universitetsforlaget.

<sup>81</sup> Begriffsklärung und Zusammenfassung der Debatte: Siegmund, F. (2014). Kulturen, Technokomplexe, Völker und Identitätsgruppen: eine Skizze der archäologischen Diskussion. *Archäologische Informationen* 37, 53-65. <https://journals.ub.uni-heidelberg.de/index.php/arch-inf/article/view/18191> – Ergänzend zur zeitlichen Dimension: Siegmund, F. (2012). Schnelle Zeiten – langsame Zeiten: Archäologische Chronologiesysteme als Geschichtsquelle. *Archäologische Informationen* 35, 259-270. <https://journals.ub.uni-heidelberg.de/index.php/arch-inf/article/view/10053>

In meiner Habilitationsschrift habe ich in einem iterativen Prozess für das frühe Mittelalter (5., 6. und 7. Jh. n.Chr.) mehrere „Kulturmodelle“ herausgearbeitet;<sup>82</sup> dabei wurden Merkmale der von lokalen Kollektiven ausgeübten Bestattungs- und Beigabensitte quantifizierend erfasst und beschrieben. Auf dieser Basis konnten einander selbstähnliche Gruppen detektiert werden. Dabei erwies sich das „Kulturmodell Nord“ als eine nach außen hin randunscharfe (sächsische) Kultur (*sic*), die während der Merowingerzeit aus mehreren, archäologisch derzeit kaum gut identifizierbaren Ethnien bestanden haben dürfte.<sup>83</sup> Demgegenüber zeigen die Kulturmodelle „West“, „Süd“ und „Ost“ ein quantitativ nachweisbares Abgrenzungsbedürfnis nach außen und können durch Hinzuziehung der Schriftquellen als Zeugnisse von drei frühmittelalterlichen Ethnien – Franken, Alemannen und Thüringern – gedeutet werden. Gegen Kritiker dieses Ansatzes und dessen Ergebnissen ist zu betonen, dass man sich von der im Grunde noch Kossinna verhafteten Idee, es gehe um räumliche Gruppen mit starren räumlichen Grenzen nach außen, lösen muss: Ethnizität kann von den Akteuren auch in räumlichen Gemengelage gelebt werden;<sup>84</sup> im frühen Mittelalter – beispielsweise – ist die Region des heutigen Franken im 6. Jahrhundert erkennbar ein Zuzugsgebiet, in dem Thüringer, Alemannen und Franken zusammenkommen und, je für sich, einander scharf unterscheiden wollende Siedlungsgemeinschaften gründen, ohne dass sich herbei konsistente Gebiete der ethnischen Gruppen abzeichneten. Im Gegenteil: gerade in der engen Begegnung mit Fremden – so das Modell Ethnicity – steigt das Bedürfnis von Kollektiven, ihre Andersartigkeit, ihre „Kultur“ demonstrativ hervorzukehren.

---

<sup>82</sup> Siegmund, F. (2000). *Alemannen und Franken. Archäologische Studie zu Ethnien und ihren Siedlungsräumen in der Merowingerzeit.* (Ergänzungsbände zum Reallexikon der Germanischen Altertumskunde 23). Berlin: de Gruyter.

<sup>83</sup> Siegmund, F. (1999). Sachsen und Franken - Ein Beitrag zur ethnischen Fragestellung. In U. von Freedon / U. Koch / A. Wiczorek (Hrsg.), *Völker an Nord- und Ostsee und die Franken. Akten des 48. Sachsensymposiums in Mannheim, 7.-11.9.1997.* (Kolloquien zur Vor- und Frühgeschichte 3). (S. 167-173). Bonn: Habelt. - Siegmund, F. (2003). Social relations among the Old Saxons. In: Dennis Howard Green & Frank Siegmund (eds.), *The Old Saxons.* (Studies in Historical Archaeoethnology). (p. 77-95). Boydell Press: San Marino.

<sup>84</sup> Siegmund, F. (2006). Commentarii: Anmerkungen zu dem Beitrag von S. Brather und H.-P. Wotzka. In: St. Burmeister & N. Müller-Scheeßel (Hrsg.), *Soziale Gruppen - kulturelle Grenzen: Die Interpretation sozialer Identitäten in der Prähistorischen Archäologie.* (Tübinger Archäologische Taschenbücher 5). (S. 225-232). Waxmann: Münster. – Vgl. in der Ethnologie: Hackstein, K. (1989). Ethnizität und Situation: Garas – eine vorderorientalische Kleinstadt. (Beihefte zum Tübinger Atlas des Vorderen Orients B 94). Wiesbaden: Reichert.

Im vorliegenden Fall stehen am Ende einer systematischen Evaluation aller Thesen und Merkmale „Kulturmodelle“ auf Basis jener ausgewählten Merkmale, die sich als tauglich erwiesen haben – insbesondere technische Merkmale sowie die Gefäß- und Waffenbeigabensitte. Aus der Summe der dem jeweiligen Kulturmodell (vorläufig) zugeordneten Gräberfelder lassen sich Gesamthäufigkeiten und Anteile für jedes Kulturmodell ermitteln und folglich konkrete Zahlen für einen idealen Vertreter dieses Kulturmodells angeben.<sup>85</sup> Anschließend kann für jedes einzelne Gräberfeld der euklidische Abstand (s.u.) berechnet werden, den es zu eben diesen Idealmodellen aufweist. Anschließend wiederum können diese Abstände für jedes Gräberfeld verglichen werden (geringe vs. große Unähnlichkeit?) sowie im Kartenbild verfolgt werden, wo diese Abstände größer sind oder kleiner sind. Das Verfahren ist auch auf andere Fragestellungen übertragbar, etwas nach sozialen Geschlechtern oder auch vertikalen sozialen Unterschieden (arm vs. reich, Adel?).

Da es sich statistisch gesehen um die Analyse von Häufigkeiten und deren Vergleich handelt, läge es nahe, mit Mehrfeldertafeln und Chi-Quadrat-Tests zu arbeiten. Im Laufe der archäologischen Argumentation zeigte sich jedoch, dass nur eine etwas komplexere Struktur geeignet ist, die Phänomene angemessen zu erfassen: Die Gefäßbeigabensitte kann auf die Gesamtzahl aller (zumindest grob datierbaren) Gräber bezogen werden. Bei der Waffenbeigabe lassen die Auswirkungen des Grabraubs dies nicht zu, weshalb hier nur die Summe der Waffen (nicht die der Gräber) als Bezugsrahmen dienen kann. Da es im Laufe des 7. Jahrhunderts zu einer Romanisierung der Waffenbeigabe kommt,<sup>86</sup> müssen dann wiederum die Saxe aus der Bezugssumme ausgeklammert werden. Für andere wichtige Phänomene wie z. B. das Verhältnis von Schlitz- zu Ganztüllen an den Lanzen spitzen des 6. Jahrhunderts oder von Drehscheibenkeramik zu handaufgebauter Keramik bildet naheliegenderweise die Fundgattung die Bezugssumme, nicht die Gräberzahl. Daher wurden entlang der archäologischen Argumentation jeweils auf spezifische Hundertsummen bezogene Prozentwerte ermittelt, um die Kulturmodelle angemessen zu beschreiben. Der „Normabstand“ eines einzelnen

---

<sup>85</sup> Siegmund 2000, p. 258-261 mit Abb. 138 (Kulturmodelle 5. Jh.); p. 266-271 mit Abb. 145 (Kulturmodelle 6. Jh.); p. 275-280 mit Abb. 154 (Kulturmodelle 7. Jh.); p. 284-286 mit Abb. 160 (Kulturmodell Nord, 5.-7.Jh.).

<sup>86</sup> „Romanisierung“ im Sinne Anpassung an die Beigabensitte der merowingerzeitlichen romanischen Kultur, die keine Waffenbeigabe ausübt. Der sukzessive Wegfall der Waffen erfolgt dabei nicht für alle Waffengattung gleichzeitig, sondern betrifft bestimmte Waffen zuerst und die Saxe zuletzt, was im 7. Jahrhundert mit einer steten Verschiebung der relativen Waffenhäufigkeiten einhergeht. Schön dargestellt bei: Ament, H. (1978). Franken und Romanen im Merowingerreich als archäologisches Forschungsproblem. *Bonner Jahrb.* 178, 377-394.

Gräberfeldes zu diesen Modellen errechnet sich jeweils in zwei Schritten als die euklidische Distanz zwischen den Modellwerten und dem einzelnen Gräberfeld:

- (1) Abstandsquadrat = Quadrat aus [beobachteter Wert minus („erwarteter“) Modell-Wert]
- (2) Normabstand = Wurzel aus der Summe aller Abstandsqurate.

In den hier beigefügten vier Tabellen „Modell...“ ist dieses Prinzip praktisch umgesetzt. Die Tabellen können dazu dienen, neue, in der Ausgangsstudie noch nicht berücksichtigte Gräberfelder in das System einzufügen. Die Tabellen enthalten die konkreten Parameter der jeweiligen Kulturmodelle und die Berechnung der Prozentzahlen und sodann (in roter Schrift) die Möglichkeit zur Eingabe eines konkreten Einzelfalles. Als Ergebnis (blaue Schrift) werden die jeweiligen Normabstände zu den Modellen ausgegeben. Eine Gräberfeldpopulation gehört jenem Modell an, zu dem es den geringsten Abstand aufweist. Je geringer der Abstand eines Gräberfeldes zu seinem Modell, desto höher die Selbstähnlichkeit.

Für Weiteres ist auf die Ausgangsstudie zu verweisen (Siegmond, 2000). Hier soll das Beispiel aufzeigen, dass der Kern des Einsatzes quantitativer Methoden stets in der Archäologie liegt und eine zielführende Statistik nicht notwendigerweise komplexe Berechnungen erfordert.

### 25.3 SAXE: BEISPIEL FÜR EINE METRISCH GESTÜTZTE TYPOLOGIE

Das folgende Anwendungsbeispiel verfolgt einen in der Archäologie häufigen Fall: die Typologie einer Fundgattung soll u. a. anhand ihrer metrischen Merkmale erfolgen. Für eine Fallstudie wählen wir die frühmittelalterlichen Saxe vom Niederrhein aus. Einerseits, weil ich darüber gearbeitet und publiziert habe,<sup>87</sup> andererseits, weil bei der Metrik der Saxe nur die Länge und die Breite der Klinge als relevant erachtet werden, der Fall also für Übungszwecke also – etwa im Vergleich zu Keramik mit deutlich mehr relevanten Messstrecken – vergleichsweise einfach ist. Das Beispiel wird zeigen, dass Zahlen und Grafiken für die Lösung der Aufgabenstellung wichtig sind, die Lösung aber im Wesentlichen mit Nachdenken, dem Sichten von Zahlen und anhand von Grafiken erarbeitet werden kann, während es statistischer Maßzahlen oder kom-

---

<sup>87</sup> Siegmund, F. (1998). *Merowingerzeit am Niederrhein*. (Rheinische Ausgrabungen, 34). Köln: Rheinland-Verlag. Zur Typologie der Saxe ebd. S. 87-94.

plexer Verfahren nicht bedarf. Indes bringt der quantitative Ansatz Macher wie Benutzer einer solchen Typologie weiter: Er zwingt dazu, subjektive Eindrücke, die jeder dicht am Material Arbeitende hat, systematisch zu objektivieren und in klaren Kriterien, Grenzwerten, Indizes o.ä. zu formulieren und deckt dabei ggf. vorhandene Schwächen und Lücken in zunächst eher subjektiv gefühlten Typologie-Modellen auf. Für die Anwender sind solche Topologien am Ende klarer und auch einfacher anwendbar als z.B. das Auflisten und das Zusammenmontieren „typischer Vertreter“ eines Typs auf einer Tafel.

Eine Sichtung der Literatur zeigt, dass im Frühmittelalter anhand der Klinglänge und -breite zwischen Schmal-, Breit- und Langsaxen unterschieden wird und dass Schmal- und Breitsaxe von manchen Bearbeitern in je zwei Untergruppen untergliedert werden. Zeitlich sind die Schmalsaxe geläufige Waffen des 6. Jh. n. Chr., werden zu Beginn des 7. Jh. von Breitsaxen abgelöst, im späten 7. Jh. dann von den Langsaxen, die bis weit in die Karolingerzeit hinein verwendet werden. Darüber verdeutlicht eine Literatursichtung, dass die konkreten metrischen Definitionen dieser Typen voneinander abweichen, d. h. die Grenze z. B. zwischen Schmal- und Breitsaxen von den verschiedenen Bearbeitern unterschiedlich gesetzt wird.<sup>88</sup> Die Aufgabenstellung des Folgenden ist es also, eine der untersuchten Region angepasste Gliederung zu schaffen, die eine möglichst eindeutige Klassifikation möglichst vieler Fälle ermöglicht.

Dazu enthält der Datensatz „8\_SaxeNRh“ alle Saxe des Arbeitsgebietes (Stand 1988). Zunächst einmal die üblichen Angaben zu Fundort (abgekürzt auf 8 Buchstaben), Grabnummer etc., wobei die Grabnummer 9999 für Einzelfunde ohne Grabkontext steht. Sodann ist die Erhaltung notiert, denn bei Eisenwaffen sind die Maße bisweilen ob der Korrosion nur mehr sehr unsicher erfassbar – das sollte für ggf. nötige Einzelfallprüfungen im Datensatz notiert sein. Mit drei weiteren Variablen werden nicht-metrische Merkmale an den Saxen erfasst, die für eine Typologie von Relevanz sind resp. es sein könnten:

- „Besonder“(-heiten): 1: keine; 2: Knauf; 3: Rillen. Insbesondere letztere sind von Interesse, da Rillen auf der Klinge bis auf seltene Ausnahmen erst an Breit- und Langsaxen vorkommen.
- „Nietform“: 1: keine Niete; 2: flache Niete; 3: hohe, hohle Niete; 4: kleine Niete. Dahinter steckt das Wissen der Frühmittelalterarchäologen, dass Schmalsaxe offenbar

---

<sup>88</sup> Siegmund (1998), 89 Tab. 12.

anders gehalten wurden als Breitsaxe: Zu Breitsaxen gehören kräftige lederne Scheiden, die oft mit ca. 4 – 5 Ziernieten sowie vielen kleinen Nieten zusammenhalten wurden. Dabei scheint es eine zeitliche Entwicklung von flachen Nieten zu größeren, hohen und innen hohlen Nieten mit kräftigem Perlrand zu geben.

- „Nietzier“: 1: keine Niete; 2: glatte, unverzierte Niete; 3: Niete mit 3 Löchern; 4: Niete mit Tierwirbel; 5: Niete mit kräftigem Perlrand. Neben der Form der Niete gibt es auch unterschiedliche Arten der Verzierung, die chronologisch relevant sein könnten.

Die Saxe waren – anders als Schwerter – fest am Hosengürtel befestigt und scheinen auch im Alltag mitgeführt worden zu sein. Daher wurde in stark verkürzter Form auch der mit einem Sax im Grab vergesellschaftet vorgefundene Gürtel resp. die Art seiner metallenen Beschläge erfasst:

- „Gürtel“: 1: kein Gürtel; 2: Schilddornschnalle ohne Gürtelhaften; 3: Schilddornschnalle mit Gürtelhaften; 4: Bronzeschnalle mit dreieckigem Beschlag; 5: Eisenschnalle mit halbrundem Beschlag; 6: eiserne Gürtelgarnitur, dreiteilig, untauschiert; 7: dreiteilige eiserne Gürtelgarnitur, (geometrisch) tauschiert; 8: drei- oder mehrteilige eiserne Gürtelgarnitur, plattiert (Tierstil); 9: vierteilige Gürtelgarnitur.<sup>89</sup>

Wie in dem meisten Publikationen üblich wird in dem Datensatz mit „lang“ die Gesamtlänge der Waffe angegeben, mit „griff“ die Grifflänge. Daraus kann die Klingenslänge („klinge“), um die es im Kern geht, errechnet werden:

=> Datenmanagement; => Variablen bearbeiten; => Erzeuge neue Variable; ins Feld „Neuer Variablenname“ „klinge“ eingeben; ins Feld „Anweisungen für die Berechnungen“ schreiben: „lang – griff“ (oder per Doppelklick auf den jew. Variablennamen einsetzen); OK.

Auch die Grifflänge ist ein verwertbares typologisches Merkmal: Während die Schmalsaxe und die meisten Langsaxe einhändig geführt wurden, gibt es einzelne (schwere) Breitsaxe mit so langen Griffen, dass sie vermutlich zweihändig geführt wurden. Weshalb ungewöhnlich lange Griffe wiederum als ein Indikator für die Klassifizierung als Breitsax angesehen werden können.

Obwohl es im Kern um die Klingenslänge und -breite der Saxe geht, wurden also recht viele weitere Merkmale erfasst. Dahinter steckt nicht überschüssige Energie des Bearbeiters, sondern die Fragestellung. Von den meisten Saxen gibt es nur die

---

<sup>89</sup> Dazu z.B. Siegmund, F. (1999). Stichwort Gürtel, Abschnitte Römische Kaiserzeit, Völkerwanderungszeit, Merowingerzeit, Karolingerzeit. *Reallexikon der Germanischen Altertumskunde (2. Aufl.) Band 13.* (S. 166-175). Berlin: de Gruyter.

Maße, so dass z. B. die Sax-Niete allein keine Typologie für die gesamte Gattung begründen können – zu nur 9 der 156 Datensätze sind Angaben zu Nieten vorhanden. Wichtiger sind die mit den Saxen kombinierten Gürtel, zunächst einmal, weil diese häufiger sind: zu 44 der 156 Datensätze ist ein klassifizierbarer Gürtel überliefert. Die Gürteltypen sind für die Frühmittelalterarchäologie zudem von besonderem Wert, da sie einem vergleichsweise schnellen zeitlichen Wandel unterliegen und dieser über einen weiten Raum hinweg (weitgehend) synchron verläuft, weshalb sich anhand der Gürtel verschiedene Lokal- und Regionalchronologien gut zusammenführen lassen.<sup>90</sup> Die gegebene Aufgabenstellung ist es, für die Saxe eine zeitlich relevante Typologie zu schaffen; da ist es hilfreich, anhand der mit einem Sax kombinierten Gürtel einen Indikator für die Zeit in die Datentabelle integriert zu haben. Die angeführten Zahlen verdeutlichen aber auch den Gewinn der Untersuchung der Saxe: man gewönne ggf. Datierungsanhaltspunkte für sehr viele Gräber, die z. B. anhand von Gürtelfunden nicht näher eingrenzbar sind.

Wir betten also gezielt chronologie-relevante Attribute und Informationen in den Datensatz ein, weil das explizite Ziel der Gliederung eine Sax-Typologie ist, die für die Datierung herangezogen werden kann. Es sind andere Fragestellungen denkbar! – die zu anderen Typologien führen würden. Die Saxklingen sind beispielsweise handwerklich unterschiedlich aufwändig gefertigt, eine Analyse etwa der Damaszierungen (insbes. bei Langsaxen) könnte für sozialgeschichtliche Fragestellungen interessant sein, führt aber ggf. zu einer Gliederung der gleichen Gattung in gänzlich andersartige Typen. Oft sind an den Saxen Lederreste der Scheide oder Holzreste der Griffe erhalten: Ansatzpunkte für wieder gänzlich andere Gliederungen dieser Fundgattung. Im Museumswesen sind Typisierungen üblich, die ebenfalls berechtigt sind, das Material jedoch wieder anders und unterschiedlich gliedern, z. B. im Hinblick auf die Archivierung nach Erhaltung im Sinne: Zustand stabil / unrettbar korrodiert / braucht Pflege des Restaurators, oder im Hinblick auf die Ausstellung: gut ausstellbar / Zustand stabil & o.k., aber nicht ausstellungsfähig. Jede dieser Gliederungen hat ihre Berechtigung. Daher: jede Typologie ist abhängig von ihrer Fragestellung, und je nach Fragestellung ist es hilfreich, passende Begleitinformationen in den untersuchten Datensatz miteinzubetten.

---

<sup>90</sup> Siegmund, 1998, 208-221. – Siegmund, F. (2012). Schnelle Zeiten - langsame Zeiten: archäologische Chronologiesysteme als Geschichtsquelle. *Archäologische Informationen*, 35, 259-270. <https://journals.ub.uni-heidelberg.de/index.php/arch-inf/article/view/10053> [24.1.2020].

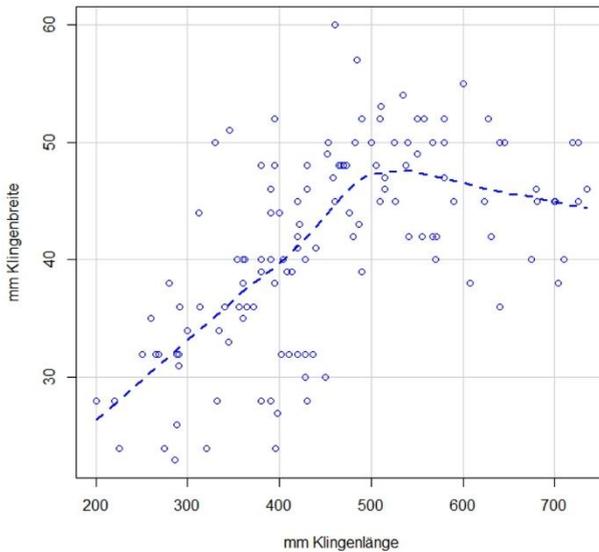
Für die weitere Auswertung wird man, wie üblich, die Daten zunächst univariat mit Hilfe von Histogrammen sichten. Angesichts der geschilderten Vorüberlegungen kann es hilfreich sein, nach ersten Histogrammen, die den Datensatz als Ganzes zeigen, die Histogramme für die beschriebenen Gruppen abzurufen, z. B. => Grafiken; => Histogramm; => Klinge, und unter „Für die Gruppen“ „Nietform“ auswählen, OK & Enter.<sup>91</sup> An dem Bild „keine Nieten“ ersieht man, dass es in unserem Datensatz Klingenslängen von ca. 15 – 65 cm gibt, während die wenigen Niete (d. h. Saxe in Saxscheideln mit Nieten) offenbar bei Klingenslängen im Bereich 25-40 cm auftreten, Niete also bei besonders kurzen wie auch besonders langen Klingen fehlen. Wiewohl es angesichts der wenigen Nieten durchaus problematisch ist, mit einem Nicht-Vorkommen zu argumentieren. Doch dieser Wertebereich zeichnet sich z. B. anhand der Variable „Besonderheiten“ bei den Rillen erneut ab: Obwohl es viele kurze Klingen gibt, setzt das Vorkommen von Rillen mit wenigen Ausnahmen erst ab ca. 30 cm ein.

Weiterführend ist die kombinierte Betrachtung von Klingenslänge und -breite in einem Streuungsdiagramm: => Grafiken; => Streuungsdiagramm; „Klinge“ auf die x-Achse und „breit“ auf die y-Achse legen, und bei Optionen ggf. auch die Achsen beschriften, z.B. mit „mm Klingenslänge“ und „mm Klingensbreite“. Durchaus nützlich für das Sichten und Ideen-Finden ist es, bei den Optionen auch (links) „Glättungskurve“ zu bestellen (**Abb. 102**). Das resultierende Streuungsdiagramm mit eben dieser Kurve zeigt nämlich die oben angedeutete typologische Entwicklung recht gut an: kurze und schmale Klingen werden länger und breiter, doch jenseits einer Länge von ca. 40 cm geht das Wachstum der Klingensbreite nicht weiter, sondern fällt mit weiter länger werdenden Klingen sogar wieder leicht ab.

Weiterführend ist es, wenn man nun statt der Glättungskurve (wieder ausblenden) die o. g. typologischen Merkmale wie z.B. die Nietform, die Rillen usw. in das Streuungsdiagramm ein spiegelt. Erst ab ca. 30 cm Länge und 4 cm Breite treten Nieten auf, und jenseits ca. 40/45 cm Länge nicht mehr. Die sprechendste Grafik erreicht man über die Gürtel, d. h. mit „Grafik für die Gruppen“ und „Gürtel“ (**Abb. 103**). Im Bereich der kurzen und schmalen Klingen findet man die Schilddornschnallen ohne und mit Gürtelhaften (mittl. Drittel 6. Jh.) sowie die (etwas späteren) bronzenen Gürtelschnallen mit dreieckigem Beschlag (2. H. 6.Jh.). Es ist jener Bereich, in dem Saxniete nicht auftreten.

---

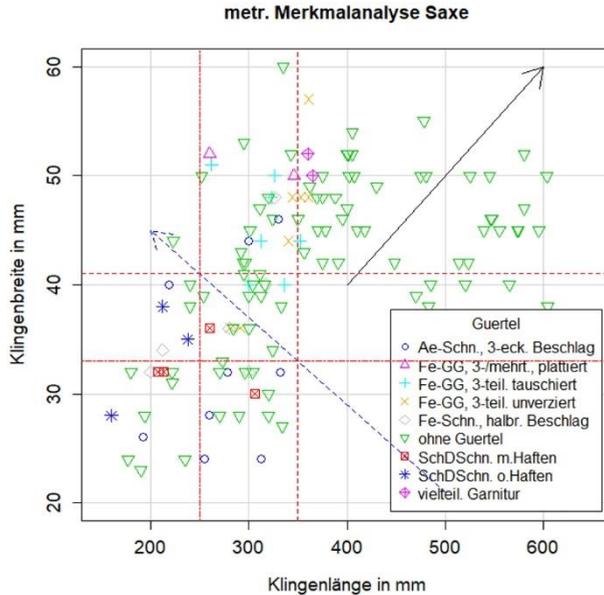
<sup>91</sup> Alternativ zu Histogrammen kann man als Darstellung auch „Strip charts“ ausprobieren.



**Abb. 102** Streudiagramm der Klingenslängen und -breiten der Saxe am Niederrhein (n=156). Eingespiegelt eine sog. Glättungskurve.

Mit den längeren und breiteren Klingen ab etwa 30/35 cm Länge und 4 cm Breite treten dann die verschiedenen Formen der eisernen Gürtelgarnituren auf, zunächst jene mit halbrundem Beschlag („um 600“), dann die tauschierten drei- und mehrteiligen Garnituren (1.-2. Drittel 7.Jh.), während Saxe jenseits von etwa 40 cm Klingenslänge nicht mehr mit Gürtelgarnituren kombiniert sind – weil sie zu Gräbern gehören, in denen ab dem letzten Drittel des 7. Jahrhunderts meist sehr einfache eiserne Schnallen als Gürtelschließe dienen, Schnallen, die sich leider bis auf wenige Ausnahmen einer weiteren Typisierung und Feinchronologie entziehen.

Das Streudiagramm **Abb. 103** ist etwas (über-) voll und so wohl kaum für eine Publikation geeignet – es ist nur ein Arbeitsinstrument. An ihm sollte aufgezeigt werden, dass man sich mit den Funktionen „arrow“ und „abline“, wenn man sie dem Plot-Befehl unmittelbar nachstellt, genau die Hilfslinien in solch‘ eine Grafik zeichnen lassen, die für’s Erarbeiten der Typologie wichtige Lesehilfen bieten.



**Abb. 103** Streuungsdiagramm der Saxklingen mit Eintragung der kombinierten Gürtelformen (n=156).

Hier der R-Code, der hinter **Abb. 103** steckt:

```
#
SaxeNRh$klinge <- with(SaxeNRh, lang - griff)
#
scatterplot(SaxeNRh$breit~SaxeNRh$klinge | SaxeNRh$Guertel,
  regLine=FALSE, smooth=FALSE, boxplots=FALSE,
  xlab="Klingenlänge in mm", ylab="Klingenbreite in mm",
  main="metr. Merkmalanalyse Saxe",
  xlim = c(150,650), ylim = c(20,60),
  cex=1.3, cex.axis=1.2, cex.lab=1.2, by.groups=TRUE,
  legend=list(coords="bottomright"), data=SaxeNRh)
arrows(500, 21, 200, 45, col="blue", lty=2)
arrows(400, 40, 600, 60)
abline(v=350, h=41, col="red", lty=2)
abline(v=250, h=33, col="red", lty=6)
#
```

Nach diesen ersten Einblicken folgt ein systematisches detaillierteres Sichten der Daten und Konstellationen. Das Sichten zielt einerseits auf das Erkennen und treffende Beschreiben „natürlicher Klassen“, d. h. Einteilungen, die sich aus Lücken resp.

schwach mit Daten besetzten Bereichen ergeben. „Natürliche Klassen“ fallen Bearbeitern in der Regel schon beim Umgang mit den Funden, d. h. während der Materialaufnahme auf, weil man sie „fühlt“, und hiermit sicherlich Klassen detektiert, die auch die damaligen Menschen so empfanden. So gibt es – als Beispiel – auch kleinere einschneidige Eisengeräte, nämlich Messer. Doch ein Messer ist funktional etwas anderes als ein Sax, für moderne Bearbeiter sind beide Gattungen in der Regel gefühlt gut voneinander unterscheidbar. Eine Grafik der Klingenslängen von Messern und Saxen zeigt, dass es bei um 15 cm Klingenslänge eine natürliche Klassengrenze gibt: kleiner = Messer, größer = (kleiner) Sax.<sup>92</sup> Trotz aller Unterschiede in der konkreten metrischen Definition haben praktische alle Bearbeiter merowingerzeitlicher Waffen das Material in Schmal-, Breit- und Langsaxe unterschieden, weil sich hier „natürliche Klassen“ andeuten. Persönlich bin ich sicher, dass auch die Damaligen diese drei Typen unterschieden haben, und dass diese drei Typen auch mit gänzlich unterschiedlichen Handhabungen, sprich: Kampfesweisen verbunden waren.

Doch typologische Gliederungen können nicht allein auf natürlichen Klassen beruhen. Einmal ist es gerade im vorliegenden Fall denkbar, dass diese durch die unterschiedlichen Erhaltungsbedingungen der Waffen so verschliffen sind, dass man sie nicht mehr erkennen kann. Sollten Waffen damals individuell für den Nutzer gefertigt worden sein, kommt der Faktor kleinere vs. größere Menschen hinzu, ein idealer leichter Breitsax für einen Mann von 1,5 m Körpergröße fiel ev. anders aus als die „gleiche“ Waffe für einen Mann von 1,8 m. Nicht zuletzt gibt es kontinuierliche Entwicklungen, bei denen es für unsere Forschungszwecke notwendig ist, sie entlang „künstlicher Klassen“ in Gruppen zu gliedern. Gerade hier ist es indes nützlich, wenn Klassengrenzen klar definiert sind, weil ein typologisches System dann eindeutiger ist und von Dritten besser angewendet und auf anderes Material übertragen werden kann.

In unserem Fall legt die nähere Sichtung von Grafiken wie z. B. **Abb. 103** nahe, erste Grenzziehungen vorzunehmen: „Schmalsaxe“ bis max. 35 cm Länge und 4,1 cm Breite – sie haben keine Rillen und keine Saxniete. „Breitsaxe“ ab 25 cm Länge und 3,3 cm Breite, bei ihnen treten Rillen und Saxniete auf. Irgendwie erkennt man im Streudiagramm in dem Überlappungsbereich beider Gruppen bei ca. 28 cm / 3,8 cm eine von links oben nach rechts unten verlaufende Zone, in der es kaum Waffen gibt, d. h. die beiden Gruppen („Punktwolken“) werden weniger entlang eines fixen Längen- und Breiten-Wertes getrennt, sondern eher durch eine Schräge. Ich nehme

---

<sup>92</sup> Siegmund 1998, 87 f. mit Abb. 18.-19 u. Tab. 11.

das als eine „natürliche“ Klassengrenze zwischen Schmal- und Breitsaxen wahr. Daher wurde an dieser Stelle eine Markierung in das Bild skizziert. Nach (vielen) weiteren Schritten, Sichtungen der Datenmatrix, Konsultation von Ausschnittvergrößerungen des Streudiagramms (Daten filtern) usw., worin auch das Sichten von Literatur einfließen kann (und floss), entsteht ein Regelwerk zur typologischen Zuordnung der Saxe (siehe Siegmund 1998, 90-92):

- Langsaxe sind ab 40 cm lang, das Verhältnis Klingenslänge/-breite liegt unter 10.
- Schmalsaxe sind kürzer 35 cm und schmaler 4,1 cm.
- Breitsaxe sind länger 25 cm und breiter 3,3 cm.
- In dem Überlappungsbereich dieser beiden Gruppen (25-35 cm lang, 33-41 cm breit) beschreibt eine von links oben nach recht unter verlaufende Grade die Trennung. Sie lässt sich mathematisch beschreiben als „61 mm – (Klingenslänge \* 0.08)“. Schmalsaxe unterschreiten diesen Trennwert, Breitsaxe überschreiten ihn.
- In der Gruppe der nunmehrigen Breitsaxe können besonders schwere Klingen als „schwere Breitsaxe“ (künstliche Klasse) ausgedontert werden: Klingenslänge über 36,5 cm oder Klingensbreite über 5 cm.
- In der Gruppe der Schmalsaxe kann man – U. Koch folgend – die besonders kleinen Kurzsaxe aussondern: bis 24 cm lang, bis 3,5 cm breit.

Mit dem bei den Langsaxen verwendeten Stichwort „Verhältnis“, auch Index genannt, kommt ein weiteres für Typologien häufig angewendetes Instrument zur Sprache. Oft sind Objekte unterschiedlich groß, so dass Formentwicklungen sich in den absoluten Maßen nicht gut widerspiegeln. Hier helfen Indexbildungen, etwa indem man den Bauchdurchmesser eines Gefäßes auf seine Gefäßhöhe bezieht oder gleich alle Masse eines Gefäßes durch seine Höhe dividiert, und damit alle Gefäße einer untersuchten Population gleich groß macht. Solche Indexzahlen sind bei Statistikern nicht beliebt, ähneln sie doch – rein statistisch gesehen – Prozentzahlen und teilen u. a. deren verteilungstheoretische Nachteile. Für Typologien sind sie indes oft unverzichtbar. Im gegebenen Fall geht es darum, bei relativ langen Waffen um 40-50 cm Klingenslänge eine Gruppe sehr schwerer Hieb Waffen mit sehr breiten Klingen zu unterscheiden von ähnlich langen, jedoch erheblich leichteren (Stich-?) Waffen mit schmaleren Klingen. Der Index Klingenslänge durch Klingensbreite ist gut geeignet, diesen subjektiv gefühlten Unterschied in Zahlen zu fassen, und dann auch klare typologische Grenzziehungen zu ermöglichen. Der schwarze Pfeil in **Abb. 103** verläuft entlang des Indexwertes 10 von Länge/Breite.

Nun überführen wir dieses Regelwerk in ein sog. Skript, d. h. eine Abfolge von Befehlen, die man z. B. extern abspeichern kann, ins Befehlsfenster von R kopieren und dann auch ausführen kann. Der Vorteil solcher Skripte ist vielfältig: Sie können schrittweise erarbeitet werden; beim Erarbeiten bemerkt man ggf. Fehler oder Lücken im typologischen Konzept; sie können auch auf andere Datensätze als den Ausgangsdatensatz angewendet werden, erlauben es z. B., erst nach Abschluss eines Projektes bekannte Neufunde nach dem einmal erarbeiteten Schema zu typisieren. Im vorliegenden Fall ist das Skript so geschrieben, dass es im R-Commander seine Arbeit tut. Hineinladen, markieren, „Befehl ausführen“. Wiewohl das Skript viele Zeilen aufweist, ist das Vorgehensschema ziemlich schlicht: Es wird eine Variable „SaxTyp“ kreiert und dann sukzessive nach den erarbeiteten metrischen Bedingungen mit Werten besetzt. Dabei kommt vor allem die bedingte Wertezuweisung mit der Funktion „ifelse“ zum Einsatz, nach folgendem Schema:

Zielvariable <- ((logische Bedingung), Wert wenn wahr, Wert wenn falsch).

Als Wert-wenn-falsch wird hier der zuvor bestehende Wert eingesetzt, so dass ggf. ein unklassifizierbares Objekt den Startwert 99 beibehält. Das Skript, das hereingeladen in den R-Commander laufen sollte, findet sich in den Übungsdateien („8\_Skript-f-Klassifikation“). Wer sich an bestimmten Namen stösst wie z. B. „SaxeNRh“, kann die Dateien in einen Editor laden (z.B. Notepad++) und dort mit Suche & Ersetze die gewünschten systematischen Veränderungen vornehmen. Auf diese Weise sind Skripte schnell zwischen verschiedenen Datensätzen und auch Projekten und Bearbeitern anpassbar. Zudem ist in unserem Beispiel die Möglichkeit genutzt, mit „#“ einen Kommentar einzuleiten: **R** überliest das in dieser Zeile Folgende. Mit solchen Kommentaren am Ende eines R-Befehls oder in Zwischenzeilen kann man für sich selbst ein Skript gliedern und dokumentieren. Klar, während des Arbeitsprozesses könnte das als überflüssig erschienen, hat man doch alles im Kopf, doch nach einer mehrwöchigen Unterbrechung wird man dankbar sein, ein paar Notizen im eigenen Skript vorzufinden, anhand derer man es leichter lesen kann.

```
#
# Saxtypologie nach Siegmund, F. (1998). Merowingerzeit
# am Niederrhein. (S. 87-94).
# neu zusaetzlich „Kurz sax“ nach Koch, U. (2001). Pleidelsheim
# (S. 61) „Code 5: bis 24 cm lg., bis 3,5 cm br.“
# Skript F. Siegmund, 1.8.2019
#
# Errechnen der Klingenlänge:
SaxeNRh$klinge <- (SaxeNRh$lang - SaxeNRh$griff)
# Errechnen des Trennwerts:
SaxeNRh$grenze <- with(SaxeNRh, 61 - (klinge*0.08))
```

```

# Setzen "SaxTyp" auf 99 = nicht klassifiziert:
SaxeNRh$SaxTyp <- 99
#
# Definition Langsaxe:
SaxeNRh$SaxTyp <- ifelse(((SaxeNRh$klinge >=400)&(SaxeNRh$klinge/SaxeNRh$breit
>=10)), 5, SaxeNRh$SaxTyp)
#
# nun Typzuweisung in Abhängigkeit von den Laenge und Breite, zunächst Schmal- vs.
Breitsaxe:
SaxeNRh$SaxTyp <- ifelse(((SaxeNRh$klinge <350)&(SaxeNRh$breit <41)), 2, Sa-
xeNRh$SaxTyp)
SaxeNRh$SaxTyp <- ifelse((SaxeNRh$SaxTyp !=5) & ((SaxeNRh$klinge >=350) | ((Sa-
xeNRh$klinge >=250)&(SaxeNRh$breit >=33))), 3, SaxeNRh$SaxTyp)
SaxeNRh$SaxTyp <- ifelse((((SaxeNRh$klinge >=250) & (SaxeNRh$klinge <350)) & ((Sa-
xeNRh$breit >=33) & (SaxeNRh$breit <41))) & (SaxeNRh$breit >SaxeNRh$grenze)), 3,
SaxeNRh$SaxTyp)
SaxeNRh$SaxTyp <- ifelse((((SaxeNRh$klinge >=250) & (SaxeNRh$klinge <350)) & ((Sa-
xeNRh$breit >33) & (SaxeNRh$breit <41))) & (SaxeNRh$breit <SaxeNRh$grenze)), 2,
=SaxeNRh$SaxTyp)
SaxeNRh$SaxTyp <- ifelse(((SaxeNRh$SaxTyp !=5) & (SaxeNRh$griff >=145)), 3, Sa-
xeNRh$SaxTyp)
#
# nun Kurzsaxe, Messer und schwere Breitsaxe
SaxeNRh$SaxTyp <- ifelse(((SaxeNRh$klinge <240)&(SaxeNRh$breit <35)), 1, Sa-
xeNRh$SaxTyp)
SaxeNRh$SaxTyp <- ifelse((SaxeNRh$klinge <150), 0, SaxeNRh$SaxTyp)
SaxeNRh$SaxTyp <- ifelse((SaxeNRh$SaxTyp ==3) & (SaxeNRh$SaxTyp !=5) & ((Sa-
xeNRh$klinge >=365) | (SaxeNRh$breit >=50)), 4, SaxeNRh$SaxTyp)
#
# abschliessend: numerische Variable in Faktor konvertieren,
# und Namen vergeben:
SaxeNRh <- within(SaxeNRh, {
  SaxTyp <- factor(SaxTyp, labels=c('Kurzsax', 'Schmalsax',
  'Breitsax', 'schw.Br.sax', 'Langsax', 'indet.')) })
# finis

```

Anregung: Das Skript mit den Übungsdaten schrittweise durchspielen, zwischendurch immer wieder den Datensatz anschauen, insbes. die letzte Spalte, wo der aktuelle Zustand der Variable SaxTyp sichtbar ist. Sofern man Dinge, die man berechnet hat, irritierenderweise (noch) nicht sieht: bisweilen kann es zwischendurch nötig sein, **R** offensiv zu ermuntern, den Datensatz zu aktualisieren: => Datenmanagement; => Aktive Datenmatrix; => Aktualisiere aktive Datenmatrix.

Zu Übungszwecken gibt es in den Datensätzen auch „8\_SaxeSchretzheim“, eine noch nicht perfekt geprüfte Datei, auf die die hier entwickelte Typologie samt Skript versuchsweise einmal angewendet werden kann. Kleine Herausforderung: der Schretzheim-Datensatz enthält auch Messer (SaxTyp 0), und in Schretzheim gibt es keine Langsaxe (SaxTyp 5), weshalb der obige Code zur Konvertierung der numerischen Variable in einen Faktor...

```
SaxeNRh <- within(SaxeNRh, {
  SaxTyp <- factor(SaxTyp, labels=c('Kurzsax', 'Schmalsax',
  'Breitsax', 'schw. Breitsax', 'Langsax', 'indet.')) })
```

... unbedingt angepasst werden muss, genauer: die Benennungen. Viel Freude mit dem Beispiel!

## 25.4 DEMOGRAPHIE ZWENTENDORF

In diesem Abschnitt untersuchen wir exemplarisch die demographischen Daten des Gräberfeldes von Zwentendorf, also Fragen nach der Geschlechterverteilung, dem Kinderanteil, der mittleren Lebenserwartung etc. Viele der Daten haben das Niveau einer Nominal- oder Ordinalskala, in der Praxis werden diese Daten oft als Häufigkeiten ausgewertet.

### 25.4.1 Datensichtung und -aufbereitung

Nach dem Einlesen der Daten „3\_Zwentendorf“ kann es notwendig sein, einige der Variablen aufzubereiten, z. B. den numerisch codierten Sex (1=Mann, 2=Frau) in einen „Faktor“ zu verwandeln. Bei einer ersten Datensichtung (z. B. => Statistik; => Deskriptive Statistik; => Häufigkeitsverteilung; für „Sex“ und „AlterKI“) stellen wir fest, dass der Datensatz 595 Zeilen hat, d. h. 595 Individuen umfasst, die Auszählung für Sex insgesamt 125 Männer und 129 Frauen ergibt, was in Summe 254 geschlechtsbestimmte Individuen sind. Wir prüfen nach: => Statistik; => Kontingenztabellen; => Kreuztabelle, und setzen „AlterKI“ in die Zeilen, „Sex“ in die Spalten. Ohne Details ausloten zu wollen, sehen wir, dass (naheliegenderweise) alle die Kinder betreffenden Altersklassen mit Null besetzt sind, d. h. geschlechtsunbestimmt, aber in der Tabelle neben „Mann“ und „Frau“ die zu erwartende Spalte „indet.“ fehlt.

Wir bereinigen diesen Mangel durch das Bilden einer neuen Variablen, die auch die Geschlechtsunbestimmten umfasst: => Datenmanagement; => Variablen bearbeiten; => Rekodiere Variablen; im Feld oben „Sex“ auswählen; im kleinen Feld rechts den neuen Variablennamen setzen, z. B. „SexAll“ und in dem Feld darunter die Recodierungsanweisung(en) eingeben (die „Grammatik“ des Rekodierens kann man via „Hilfe“ nachlesen):

```
1 = „Mann“
2 = „Frau“
0 = „indet.“
NA = „indet.“
```

...und mit OK bestätigen. Danach zeigt uns eine Häufigkeitsauszählung von „SexAll“ an, dass nun auch 341 „indet.“ im Datensatz enthalten sind. Eine für die spätere Übersichtlichkeit nützliche Maßnahme könnte zudem das Neuordnen der Variable „SexAll“ sein: => Datenmanagement; => Variablen bearbeiten; => Neuordnen der Faktorenstufen; „SexAll“ auswählen, Häkchen an „Erzeuge sortierte Faktorenvariable“, OK. In dem dann aufklappenden Fenster die Reihenfolge der Merkmalsausprägungen neu festlegen, z. B. – wie in der Anthropologie weithin üblich – 1 Mann, 2 Frau, 3 indet. & OK. Danach weist die Kreuztabelle AlterKl mit SexAll (=> Statistik; => Kontingenztabelle; => Kreuztabelle; ...) auch alle Geschlechtsunbestimmten aus. Anschließend führen wir diese Neuordnung der Faktorenstufen auch mit der Variablen „AlterKl“ durch, d. h. bringen die Faktoren in die entsprechende biologische Reihenfolge. Danach gibt uns eine Kreuztabelle (**Abb. 104**; => Statistik; => Kontingenztabelle; => Kreuztabelle; ...) eine Übersicht über die Alters- und Geschlechtsbestimmungen:

SexAll		Mann	Frau	indet.
AlterKl				
	foetal	0	0	37
	Inf. I/1a	0	0	73
	inf. I/1b	0	0	31
	inf. I/2	0	0	69
	inf. II	0	1	60
	juvenil	9	9	12
	f-ad	10	18	8
	s-ad	52	29	11
	f-mat	32	21	2
	s-mat	12	25	9
	sen	1	19	0
	erw.	9	7	29

**Abb. 104** Kreuztabelle Altersklassen und Geschlecht für Zwentendorf, unveränderte Ausgabe aus R.

Das ist zwar noch keine schöne, publikationsreife Tabelle, aber sie enthält die Informationen, die wir für die weitere demographische Auswertung benötigen. Die Bearbeiter von Zwentendorf haben für die Erwachsenen die übliche Klassifikation mit adult (20-40), matur (40-60) und senil (60 ff.) offenbar differenziert in f-ad, s-ad, f-mat und s-mat, was für frühadult, spätadult, frühmatur und spätmatur steht, d. h. für Altersklassen à 10 Jahre – was ein weithin verfolgter Standard ist. Die Alterseinteilung der Kinder hingegen z. B. mit „Inf. I/1a“ ist unüblich, wir müssen sie genauer verstehen. Man könnte die Information in der Publikation nachlesen (was gewiss der richtige

Weg ist!), statt dessen schauen wir im Datensatz selbst nach: Er enthält u. a. die Variablen „Jahre“ und „Monate“, in die Altersbestimmungen auch nach numerisch angelegt sind, wobei beide Variablen additiv gemeint sind, d. h. ein 1,5-jähriges Kind als 1 Jahr 6 Monate abgelegt ist. Dieses Ablegen der Information hängt mit unserem üblichen Notationssystem zusammen, nach dem wir Monate in einem Zwölfersystem denken, das numerisch wiederum als Zehnersystem zu notieren wäre, also z.B. 1 Jahr und 6 Monate als 1,5 Jahre, 1 Jahr und 9 Monate als 1,75 Jahre. Sprich: die Kollegen haben für das Ablegen der Daten eine angemessene, intuitiv verständliche Weise gewählt, deren weitere statistische Auswertung jedoch etwas holprig ist. Als eine der möglichen Lösungen fassen wir beide Variablen zusammen und bilden „Lebensmonate“. => Datenmanagement; => Variablen bearbeiten; => Erzeuge neue Variable“; dort setzen wir als „Neuer Variablenname“ „Lebensmonate“ ein und vereinbaren bei „Anweisung für die Berechnung“:

Jahre \* 12 + Monate. OK.

Routinemäßig überprüfen wir jede Transformation, z. B. mit einem Histogramm, und sehen, dass das erwünschte Ergebnis erreicht wurde: Für alle Alter bis einschließlich 14 Jahre (d. h. alle Kinder) steht nun das Alter in Lebensmonaten bereit. Anhand einer Kreuztabelle lässt sich nun ersehen, dass Inf. I/1a alle Kinder von 2 Monaten bis 2 Jahren meint (d. h. 1 Monat steht für „neonat.“ / Neugeborene), Inf. I/1b alle 2- bis 3-jährigen, Inf. I/2 alle 3- bis 7-jährigen und Inf. II (wie üblich) alle 7- bis 14-jährigen. Juvenil ist von 15 bis einschließlich 19 Jahre gemeint, die Altersklasse frühadult beginnt mit 20 Jahren.

Hinweis: Numerische Variablen können richtigerweise nicht in Kreuztabellen verwendet werden. Falls nötig, muss man dafür eine numerische Variable wie z. B. „Jahre“ zuvor in einen Faktor verwandeln, d.h. => Datenmanagement; => Variable bearbeiten; => Konvertiere numerische Variable in Faktoren; beispielsweise „Jahre“ mit „Verwende Ziffern“ in die neue Variable (z. B.) „JahreF“ hineintragen, weil man dann nach Bedarf sowohl die numerische Variable „Jahre“ als auch den Faktor „JahreF“ verwenden kann.

Praxistipp. Solche Datenaufbereitungen sind Teil nahezu jeder Analyse eines Datensatzes. Im Laufe eines Projekts kann dies insgesamt eine größere Menge an Arbeitsschritten sein, die zudem (tipp-) fehlerbehaftet sind. Daher lohnt es, sich alle Prozeduren, die erfolgreich und fehlerfrei waren und die man anschließend „immer wieder“ braucht, zu sammeln und in einem Skript abzuspeichern. Das vereinfacht das Wiederanknüpfen an das bisher Erreichte nach einer Arbeitsunterbrechung auf zwei Schritte:

(1) Daten einlesen, (2) Datenaufbereitungs-Skript in das Befehlsfenster einkopieren, markieren und ausführen. So hat man stets den aktuellen, sauberen Datensatz. Ein Beispiel für ein solches Skript – mit den zuvor für Zwentendorf beschriebenen Schritten – findet sich unter den Übungsdaten.

### 25.4.2 Analyse der Geschlechtszusammensetzung

Damit sind wir in der Lage, die Angaben in **Abb. 104** genauer zu verstehen und erstellen aus diesen Daten die zusammenfassende Übersicht **Abb. 105**, die man wohl auch publizieren würde.

	Mann	Frau	indet.	Summe
Kinder	0	1	270	271
Jugendliche	9	9	12	30
Erwachsene	116	119	59	294
Summe	125	129	341	595

**Abb. 105** Zusammenfassende Übersicht zu Alter und Geschlecht nach den Daten aus **Abb. 104**.

An der Zusammenfassung wird deutlich, dass der Anteil der Kinder und Jugendlichen in Zwentendorf mit 301 von 595 Individuen (50,6 %) sehr hoch ist, und dass bei den Erwachsenen ein ausgewogenes Männer-Frauen-Verhältnis besteht. Im gegebenen Fall ist das „ausgewogen“ offensichtlich, doch zu Übungszwecken haken wir nach.

Oft wird zu diesem Thema ein „Maskulinitätsindex“ (MI) berechnet, und zwar als  $(\text{Anzahl Männer} / \text{Anzahl Frauen}) * 100$ .<sup>93</sup> Wobei ein Ergebnis von 100 als „ausgewogen“ gelten würde, Werte über 100 stehen für „Männerüberschuss“, Werte unter 100 für „Frauenüberschuss“. So übersichtlich dieser Index auch sein mag – er ist statistisch weitgehend wertlos, da man die Abweichungen vom Idealwert im Grunde kaum auf eine Signifikanz einschätzen kann. Wollte man angemessen mit den Informationen umgehen, müsste man zumindest das Konfidenzintervall mitgeben, was in den einschlägigen Publikationen fast nie geschieht. Da diese Intervalle bei den üblichen Populationsgrößen in der Regel sehr weit ausfallen, dienten sie eher der Verunsicherung des Lesers als der Klärung. Hat man tatsächlich den Verdacht, dass eine Unausgewogenheit besteht, ist ein Chi-Quadrat-Anpassungstest die einzig angemessene Lösung.<sup>94</sup>

<sup>93</sup> Z. B. Herrmann, B. u. a. (1990). *Prähistorische Anthropologie. Leitfaden der Feld- und Labormethoden*. Berlin: Springer (p. 310). – Bisweilen wird statt 100 (i. e. Prozent) auch die Zahl 1000 eingesetzt (i. e. Promille), was eine höhere Genauigkeit suggeriert.

<sup>94</sup> Siegmund, F. (2009). Der Maskulinitätsindex und statistische Verfahren zur Prüfung auf Unterschiede in den Geschlechteranteilen frühmittelalterlicher Populationen. *Bulletin der Schweizerischen Gesellschaft für Anthropologie* 15(1-2), 5-17. Dort u. a. Belege für das Mädchen-Knaben-Verhältnis und Vergleichsdaten. – Mustergültig

Als Vergleich würde man entweder – wenn begründbar – ein anderes Gräberfeld einsetzen, besser jedoch die Tatsache, dass im (gut dokumentierten) neuzeitlichen Mitteleuropa auf 100 Mädchen-Geburten 105,55 Knabengeburt entfallen. Unter der Annahme, dass alle Geborenen irgendwann auch einmal unter den Gestorbenen auftauchen und das Sterberisiko unter den nicht geschlechtsbestimmten Kindern für Knaben wie Mädchen gleich ist, lassen sich daraus für jede Population Erwartungswerte formulieren. Mit dem Ansatz:

Geschlechtsbestimmte Population = 254 (im Falle Zwentendorf).

Faktor = geschlechtsbestimmte Population / 205.55;

erwartete Anzahl Frauen = Faktor \* 100;

erwartete Anzahl Männer = Faktor \* 105.55;

erhalten wir für Zwentendorf 123.6 Frauen und 130.4 Männer als die biologisch erwarteten Häufigkeiten für einen Test. Mit => Statistik; => Kontingenztabelle; => „Analyse einer selbst eingegebenen Kreuztabelle“ prüfen wir dies.

Pearson's Chi-squared test

data: .Table

X-squared = 0.01544, df = 1, p-value = 0.9011

Das resultierende Chi-Quadrat von 0,9 ist weit von jeder Signifikanz entfernt, d. h. das beobachtete quantitative Männer-Frauen-Verhältnis in Zwentendorf weicht nicht signifikant von den Erwartungen (eines ausgewogenen Geschlechterverhältnisses) ab – wie bei den meisten mir bekannten frühmittelalterlichen Gräberfeldpopulationen.

Gäbe es signifikante Abweichungen vom Erwarteten, müssten die Ursachen jenseits der Statistik gesucht und plausibel gemacht werden. Möglich ist, dass bei den Erwachsenen eine Selektion vorliegt (z. B. Männerkloster, Frauenkloster, Soldatenfriedhof u. a.). Möglich ist auch, dass bei den nicht-geschlechtsbestimmten Kindern und Jugendlichen ein sehr unterschiedliches Sterberisiko vorlag. Das kann an geschlechtsspezifisch unterschiedlicher Fürsorge für die Kinder liegen, oder an geschlechtsspezifisch unterschiedlichen Sterberisiken bei den Jugendlichen. So sind – man verzeihe den Vergleich – bei Schimpansen männliche Jugendliche untereinander sehr aggressiv, was die Herdendemographie von erwachsenen Schimpansen deutlich zuungunsten des Männeranteils beeinflusst. So ist z. B. auch denkbar, dass sehr junge

---

sauber durchgeführt für zu Zwentendorf zeitähnliche Gräberfelder bei: Caselitz, P. (1982). Die frühmittelalterlichen Skelettgräber vom Schölerberg in Osnabrück: Die Ergebnisse der anthropologischen Bearbeitung. *Nachrichten aus Niedersachsens Urgeschichte* 51, 127-169, hier: 146-147. Nur acht der dort zusammengetragenen 39 Gräberfelder weisen signifikante Abweichungen auf.

Frauen unmittelbar nach dem Fruchtbarwerden-Können auch in die Reproduktion eintreten, was ihre Gesundheitsrisiken erhöht und wiederum den Anteil erwachsener Frauen in einer Population stark beeinflussen kann.

### 25.4.3 Alterszusammensetzung I: der Kinderanteil

Der Anteil der Kinder und Jugendlichen in Zwentendorf liegt bei 50,6%. Im Vergleich zur Merowingerzeit, wo auf den Gräberfeldern Kinderanteile um 15% und um 25% üblich sind,<sup>95</sup> oder etwa zur Bandkeramik<sup>96</sup> mit Anteilen oft um 25 %, ist dies sehr viel. Die weitere Debatte um dieses Thema ist endlos.<sup>97</sup> Die Mehrheit der Kollegen hält einen Anteil von um 45 – 60 % verstorbener Kinder und Jugendlicher für „natürlich“, gerne werden für die Berechnung der Sterbetafeln (dazu weiter unten) sogar „fehlende Kinder“ hinzuinterpoliert. Andere Kollegen, zu denen der Verfasser gehört, können sich sehr wohl Populationen vorstellen, die bei passenden Lebensumständen (geringe Bevölkerungsdichte; stark ländliche statt städtische Lebensweise) und anderen Reproduktionsstrategien (hohes Elterninvestment in das einzelne Kind) auch mit einem tatsächlich geringen Kinderanteil erfolgreich Bestand haben. Diese Debatte sei hier nicht weitergeführt, sondern nur kurz die Konsequenzen hinsichtlich der Statistik herausgearbeitet: Das Arbeiten mit Prozentzahlen und Indices hilft, das vorliegende Material im Sinne einer rein beschreibenden Statistik darzustellen. Indes: Unterschiede zwischen Populationen wird es immer geben, sodass stets auch die Frage auftaucht, ob diese mehr oder weniger großen Unterschiede statistisch signifikant sind. Für eine sichere Beantwortung dieser Frage taugen Prozentwerte und Indizes nicht (gar nicht!) –

---

<sup>95</sup> Grundlegend: Acsádi, Györgi & Nemeskéri, János (1970). *History of human life span and mortality*. Budapest: Akadémia Kiadó. / Für das Frühmittelalter: Lohrke, Brigitte (2004). *Kinder in der Merowingerzeit: Gräber von Mädchen und Jungen in der Alemannia*. Rahden /Westf.: Leidorf. – Sehr gründliche, lesenswerte Debatte und Zusammenstellung von Vergleichszahlen bei: Kölbl, Stefanie (2004). *Das Kinderdefizit im frühen Mittelalter - Realität oder Hypothese? Zur Deutung demographischer Strukturen in Gräberfeldern*. Nat. Diss. Univ. Tübingen: <https://d-nb.info/970716850/34> [24.1.2020].

<sup>96</sup> Siemoneit, Beate (1997). *Das Kind in der Linienbandkeramik: Befunde aus Gräberfeldern und Siedlungen in Mitteleuropa*. (Internationale Archäologie 42). Rahden /Westf.: Leidorf.

<sup>97</sup> Beispiele: Ulrich-Bochsler, Susi (1997). *Anthropologische Befunde zur Stellung von Frau und Kind in Mittelalter und Neuzeit*. Bern: Berner Lehrmittel- und Medienverlag. / Ulrich-Bochsler, Susi (2006), in: Chr. Kissling, Chr. & S. Ulrich-Bochsler. *Kallnach-Bergweg: Das frühmittelalterliche Gräberfeld und das spätrömische Gebäude*. Bern: Archäologischer Dienst des Kantons Bern. / Lohrke, Brigitte & Cueni, Andreas (2010), in: K. Müller, *Gräber, Gaben, Generationen: Der frühmittelalterliche Friedhof (7. Jahrhundert) von der Frühberggasse in Baar (Kanton Zug)*. (Antiqua 48). Basel: Archäologie Schweiz. / Grefen-Peters, Silke (2013), in: M. C. Blaich, *Werla 2 – Die Menschen von Werlaburgdorf: Ein Beitrag zur Geschichte des Nordharzvorlandes im 8. bis 10. Jahrhundert*. (Monographien des Römisch-Germanischen Zentralmuseums, 114). Mainz: Verlag des Röm.-German. Zentralmuseums.

wie oben in den Abschnitten zu Konfidenzintervallen und „Chi-Quadrat-Krise“ dargelegt wurde. Wer den eigenen Befund mit anderen Serien zuverlässig in Beziehung setzen will, sollte die Rohdaten, d. h. die beobachteten Häufigkeiten zusammenstellen und diese – je nach Fragestellung: individuell oder als Sammelserie – den eigenen Daten gegenüberstellen. Allein die tatsächlichen Häufigkeiten und ihr Vergleich mit Hilfe des Chi-Quadratverfahrens samt eindeutigem Test sind seriös.

Nicht zuletzt: Das Ergänzen „fehlender“ Kinder ist grober Unfug!

#### 25.4.4 Alterszusammensetzung II: Demographie der Kinder

Für eine erste Vertiefung stellen wir die Nicht-Erwachsenen aus **Abb. 104** zu einer übersichtlichen Tabelle zusammen, die zum ersten Überblick und besseren Vergleich mit anderen Publikationen mit den üblichen Altersklassen operiert.

	Anzahl	%
Neonat.	37	12,3
Inf. I	173	57,9
Inf. II	60	18,0
Juv.	30	9,0
Summe	300	100

**Abb. 106** Auszug aus der Demographie von Zwentendorf: Häufigkeit der nicht-erwachsenen Toten.

Zudem ermitteln wir (anhand einer Tabelle der Lebensmonate) die für viel verwendete Testparameter benötigte Anzahl der gestorbenen Kinder im Alter von 5 bis einschl. 9 Jahren (in Zwentendorf: 50) und von 10 bis einschl. 14 Jahren (in Zwentendorf: 31). Daraus errechnen wir:  $D_{5-9}/D_{10-14} = 1.6$  und  $D_{5-14}/D_{erw.} = 0.28$ , was ganz grob den Verhältnissen der von G. Gruppe bearbeiteten Schleswiger Skelettserie und der von S. Greven-Peters publizierten Serie von Werlaburgdorf entspricht, jedoch von den laut Bocquet & Masset (1977) erwarteten Mindestwerten abweicht,<sup>98</sup> die laut Bocquet & Masset (1977) eine repräsentativ überlieferte Kinder- und Jugendlichen-Serie indizieren. Für einen Populationsvergleich samt Test verwenden wir das Gräberfeld von Werlaburgdorf (8.-10. Jh.).<sup>99</sup>

<sup>98</sup> Zusammenfassend samt Diskussion z.B. bei Herrmann u.a. (1990), 306-310; Greven-Peters (2013), 31-35.

<sup>99</sup> Greven-Peters (2013) 32 Tab. 3.

	Zwentendorf	<i>erm.</i>	Werlaburgdorf	<i>erm.</i>
Neonat	37	<i>33.5</i>	8	<i>11.5</i>
Inf. I	173	<i>163.7</i>	47	<i>56.2</i>
Inf. II	60	<i>61.0</i>	22	<i>21.0</i>
Juv.	30	<i>41.7</i>	26	<i>14.1</i>
Summe	300		103	

**Abb. 107** Vergleich der Kindersterblichkeit in Zwentendorf und Werlaburgdorf. Links jeweils die beobachteten Häufigkeiten, recht (kursiv) die Erwartungswerte, die sich ergäben, wenn beide Populationen dem gleichen Muster folgten. An größeren Differenzen beobachtet – erwartet wird deutlich, wo die Unterschiede liegen, die laut Chi-Quadrat hochsignifikant sind.

Ein Chi-Quadratstest mit  $\text{Chi-Quadrat} = 16.3$  und  $p=0.00095$  zeigt, dass sich beide Serien hochsignifikant unterscheiden. Die Erwartungswerte („*erm.*“) in **Abb. 107** zeigen an, wo die wesentlichen Unterschiede liegen: in Werlaburgdorf ist der Anteil von Inf. I geringer als (gegenüber Zwentendorf) erwartet und vor allem der Anteil der verstorbenen Juvenilen sehr hoch. Damit jedoch entspricht Zwentendorf stärker den Erwartungen an eine ungestörte repräsentative Serie – worauf im nächsten Kapitel beim Thema „life tables“ näher einzugehen sein wird.

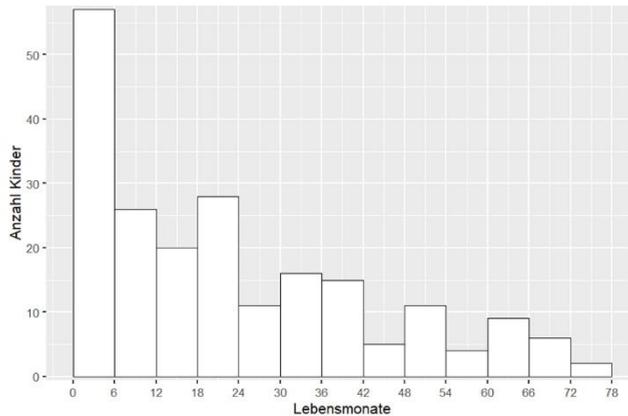
Wir vertiefen die Einsicht in die Verhältnisse weiter, indem wir allein für die jungen Kinder (bis Inf. II) die Lebensmonate näher studieren und z. B. als Histogramm ausgeben (**Abb. 108**).<sup>100</sup> Danach starben – wie üblich – die meisten Kinder in der ersten Lebensmonaten, danach sinkt die Sterblichkeit, ein zweites relatives Maximum liegt im Alter um 1,5 bis 2 Jahre.

Da das Abstillen der Kinder und die damit verbundene Nahrungsumstellung als eine Zeit erhöhten Risikos während der Lebensphase Infans gilt, könnte man das Histogramm als begründeten Anfangsverdacht darauf nehmen, dass das Abstillen in Zwentendorf oft im Alter von um 2 Jahren geschah – was wiederum Rückschlüsse auf den üblichen Intergeburtenabstand bei den Müttern zuließe. Ein Blick in die Rohdaten zeigt, dass die Häufung „um 2 Jahre“ genauer bei 24, 30 und 36 Monaten (Abstillalter?) liegt, und dass eine weitere Häufung davor bei 18 Monaten liegt (Kind lernt Laufen = erhöhtes Unfallrisiko?). Was an dieser Stelle nur als Hinweis darauf genommen werden soll, dass es hier um recht kleinteilig unterschiedliche Phänomene geht, bei denen stark

---

<sup>100</sup> **Abb. 108** entstand auf Basis der Kinder bis einschl. 83 Lebensmonaten (Datensatz „InfansI“) mit Hilfe des R-Commanders und dem Plugin KMGgplot2, die Befehlssequenz findet sich unter den Übungsdateien als Skript.

aggregierende Histogramme relevante Phänomene eher kaschieren denn herausarbeiten könnten. Wobei vorab quellenkritisch natürlich zu beleuchten wäre, welche Klassen und Häufungen sich allein aufgrund der üblichen Ansätze der anthropologischen Altersbestimmung ergeben. Wie gesagt: hier geht es nicht um Ergebnisse, sondern um typische Fragestellungen und Überlegungen, wie man diese löst.



**Abb. 108** Erreichtes Lebensalter in Monaten für die Zwentendorfer Kinder (nur Infans I, n=210).

### 25.4.5 Alterszusammensetzung III: Sterbetafeln (life tables)

Die hier in **Abb. 104** zusammengestellten Rohdaten werden traditionell in sog. Sterbetafeln überführt, die in der englischsprachigen Fachliteratur sinnigerweise als „life tables“ bezeichnet werden.<sup>101</sup> Zwar gibt es für deren Erstellung mit `mortAAR` sogar ein eigenes R-Paket,<sup>102</sup> doch ich persönlich ziehe für diese Aufgabe das Arbeiten mit einer Tabellenkalkulation deutlich vor. Während in solchen Tafeln für die moderne Demographie jedes Lebensjahr eine Datenzeile ausmacht, arbeitet man in der Paläodemographie fast durchweg mit sog. verkürzten Sterbetafeln, in denen mehrere Lebensjahre – entsprechend der Detailliertheit zu Grunde liegenden anthropologischen Altersbestimmung – zu Altersklassen zusammengefasst sind (**Abb. 109**).

---

<sup>101</sup> Grundlegend in archäologischem Kontext: Acsádi & Nemeskéri 1970. Kurz und gut: Herrmann u.a. 1990, 304-306 mit Tab. 5.2.1.

<sup>102</sup> <https://cran.r-project.org/web/packages/mortAAR/index.html> [24.1.2020].

Gehen wir die Bedeutung der einzelnen Spalten kurz durch:

a	Umfang der Altersklasse in Jahren
$D_x$	Anzahl der in dieser Altersklasse gestorbenen Individuen
$d_x$	relative Anzahl der in dieser Altersklasse gestorbenen Individuen
$l_x$	relative Anzahl der Überlebenden
$q_x$	Sterbewahrscheinlichkeit in dieser Altersklasse
$L_x$	Anzahl der in dieser Altersklasse insgesamt gelebten Jahre
$T_x$	Anzahl der in dieser Population insgesamt noch zu lebenden Jahre
$e_x$	mittlere Lebenserwartung zu Beginn dieser Altersklasse
$a_x$	mittlere Lebenserwartung derjenigen, die diese Altersklasse erreicht haben

Alter	a	$D_x$	$d_x$	$l_x$	$q_x$	$L_x$	$T_x$	$e_x$	$a_x$
infans I	7	210,0	382,5	1.000,0	382,5	5.661,2	22.404,4	22,4	22,4
infans II	7	60,0	109,3	617,5	177,0	3.939,9	16.743,2	27,1	34,1
juvenil	6	30,0	54,6	508,2	107,5	2.885,2	12.803,3	25,2	39,2
frühadult	10	36,0	65,6	453,6	144,6	4.207,7	9.918,0	21,9	41,9
spätadult	10	92,0	167,6	388,0	431,9	3.041,9	5.710,4	14,7	44,7
frühmatur	10	55,0	100,2	220,4	454,5	1.703,1	2.668,5	12,1	52,1
spät matur	10	46,0	83,8	120,2	697,0	783,2	965,4	8,0	58,0
Senil	10	20,0	36,4	36,4	1.000,0	182,1	182,1	5,0	65,0
<b>Summe</b>		<b>549,0</b>							

**Abb. 109** Sterbetafel (*life table*) für das Gräberfeld von Zwentendorf.

Die Summe der Verstorbenen in **Abb. 109** entspricht nicht ganz den Rohdaten (**Abb. 104**). Denn die nur als „erwachsen“ bestimmten Individuen wurden hier nicht aufgenommen. Denn wo sollte man sie zuordnen? Ein bisweilen gewählter Weg, sie anteilig auf alle Altersklassen frühadult bis senil zu verteilen, verzerrt das demographische Bild einer Population zu sehr – es ist besser, für solche Tafeln auf die allzu ungenau bestimmten Individuen zu verzichten. Dennoch: In geringem Umfang ist ein solches Aufteilen von Individuen üblich.<sup>103</sup> Gäbe es beispielsweise ein nur allgemein als

<sup>103</sup> Zum konkreten Vorgehen: Kemkes-Grottenthaler, Ariane (1993). *Kritischer Vergleich osteomorphologischer Verfahren zur Lebensalterbestimmung Erwachsener*. Phil. Diss. Mainz. S. 173-174.

„adult“ bestimmtes Individuum, würde man es im vorliegenden Fall als 0,5 frühadult und 0,5 spätadult in der Tabelle erfassen. Womit erklärt ist, weshalb solche Tabellen eher regelhaft als gelegentlich auch halbe und drittel Individuen enthalten. Als Faustregel sollte gelten: ein Verteilen sollte auf max. drei Altersklassen erfolgen, noch ungenauer bestimmte Individuen gehören nicht in eine solche Tabelle.

Indes: Man passt die Tabellen auch den anthropologischen Altersbestimmungen an. Erfolgen sie z. B. nach adult, matur, senil, also nach 20-Jahres-Klassen für die Erwachsenen, macht das in **Abb. 109** benutzte 10-Jahres-System keinen Sinn. Würden die Daten – wie in Mitteleuropa weithin üblich – nach der „komplexen Methode“ erhoben und auch jahrgenau oder mit engen Altersspannen publiziert,<sup>104</sup> können – zumindest technisch – auch die von Acsádi & Nemeskéri (1970) bevorzugten Sterbetafeln mit 5-Jahres-Klassen gebildet werden. In der archäologie-relevanten Anthropologie kommen für Erwachsene Tabellen mit 5-Jahres-Klassen, 10-Jahres-Klassen und 20-Jahres-Klassen vor, bisweilen wird auch mit 7er-Schritten (früh-, mittel-, spät-) gearbeitet. Hinsichtlich der Erwachsenen sollte man also frühzeitig prüfen, wie groß die Altersspannen üblicherweise sind. Hat man, wie eingangs beim Thema Datenerfassung diskutiert, die Information in zwei Spalten als „von“ und „bis“ abgelegt, lässt sich schnell sehen, ob in einer Serie für die Mehrheit der Erwachsenen ein 5er-, 7er-, 10er- oder 20er-System angemessen ist.

Den zu diesem Buch gehörenden Daten sind dem entsprechend mehrere Sterbetafeln als LO- und Excel-Tabellen beigefügt, die für einige übliche, unterschiedliche Systeme von Altersklassen angelegt sind und entweder so wie gegeben benutzt werden können, oder bei Bedarf nach etwas Eindenken selbst angepasst werden können.

Im vorliegenden Fall haben wir mit **Abb. 109** eine Tafel für die Gesamtpopulation mit 10-Jahres-Klassen (bei den Erwachsenen) erstellt, normalerweise wird zusätzlich eine eigene Tafel für die Männer und eine für die Frauen berechnet. Aus den Spalten  $d_x$  und  $q_x$  werden oft Graphiken abgeleitet – die mir persönlich als wenig aussagekräftig erscheinen. Eine regelhaft weiterverwendet Zahl ist  $e_0$ , die Lebenserwartung bei der Geburt, in unserem Fall 22,4 Jahre. Ich hoffe, Sie stutzen hier: 22,4? Die Kennzahl  $e_0$  ist in der Demographie ein viel verwendeter Wert, doch oft irritiert sie bei Nicht-Profis, insbesondere, wenn sie so tief ausfällt. Doch aus den Werten in der Spalte  $D_x$  ersehen wir, dass es durchaus eine nennenswerte Zahl älterer Menschen in Zwentendorf gab.

---

<sup>104</sup> Acsádi & Nemeskéri 1970; Herrmann u. a. 1990, 52-68. Einen wertvollen Methodenvergleich bietet eine Studie zur Abschätzung der Qualität der Altersbestimmungen: Kemkes-Grottenthaler, Ariane (1993). *Kritischer Vergleich osteomorphologischer Verfahren zur Lebensalterbestimmung Erwachsener*. Phil. Diss. Mainz.

Es lohnt, mit einer Sterbetafel „zu spielen“, um deren Berechnungen besser zu verstehen – weshalb es vorteilhaft, sie als Tabellenkalkulation aufgesetzt zu haben. Man lösche einmal alle Kinder aus der Tabelle und betrachte die Wirkung auf  $e_0$ . Man „ergänze“ einmal nicht vorhandene Kinder: Der Wert von  $e_0$  Fehler! Textmarke nicht definiert. ist wesentlich von der Zahl der verstorbenen Kinder abhängig – wie entweder schlichtes Nachdenken oder vergleichende Versuche via Tabellenkalkulation bei den Zahlen der Erwachsenen schnell verdeutlichen. Das Problem in der Archäologie ist – siehe oben die Debatte um das Kinderdefizit ist –, dass gerade die Anzahl der Kinder „unsicher“ ist: Es ist möglich, dass die Anzahl der auf einem Gräberfeld erfassten Kinder der tatsächlichen Zahl verstorbener Kinder entspricht. Es ist aber auch möglich, dass die Zahl der Kinder aufgrund schlechter Knochenhaltung oder einer Sonderbehandlung verstorbener Kinder etc. kleiner ausfällt. Was bedeutet, dass der Wert von  $e_0$  wesentlich von Faktoren abhängt, die bei archäologischen Serien als unsicher und schwer abschätzbar angesehen werden. Weshalb die in der Demographie viel verwendete Kennzahl  $e_0$  für den Vergleich archäologisch überlieferter Serien nur bedingt brauchbar ist. Daher wurde die Sterbetafel hier um die selten verwendete Spalte  $a_x$  erweitert. Sie beziffert die mittlere Lebenserwartung derjenigen, die diese Altersklasse erreichten. Ein demographisch gut verwendbarer Wert ist daraus  $a_{erw}$  (oder auch:  $a_{20}$ ), die mittlere Lebenserwartung derjenigen, die das Erwachsenenalter erreichten, hier 41,9 Jahre. Wie „Spielen“ schnell aufzeigt, ändert sich dieser Wert nicht, wenn man Kinder addiert oder löscht. Daher empfehle ich, zusätzlich zu  $e_0$  für die Gesamtpopulation, die Männer und die Frauen stets auch  $a_{20}$  für eine Population anzugeben. Populationsvergleiche auf Basis von  $a_{20}$  sind inhaltlich weitaus belastbarer als solche auf Basis von  $e_0$ .

Ein anderes Problem der Sterbetafeln lässt sich indes nicht beheben: Kennzahlen wie  $e_0$  oder  $a_{20}$  sind – Prozentzahlen ähnlich – zur Beschreibung einer Population gut geeignet. Doch wie steht es um die Frage, ob ein Unterschied zwischen zwei Populationen noch im Rahmen des Zufälligen liegt oder signifikant anders ist? Diese wichtige Frage lässt sich anhand der aus einer Sterbetafel abgeleiteten Kennzahlen in der Regel nicht beantworten, weshalb der folgende Abschnitt seltener begangene, jedoch wichtige alternative Zugriffe auf die Frage der Bevölkerungszusammensetzung und der mittleren Lebenserwartung eröffnet.

Ein letztes Caveat: Sterbetafeln und die resultierenden Kennzahlen setzen eine stationäre Bevölkerung voraus. Durch starke Ein- oder Auswanderung geprägte Populationen oder solche, die in sich sehr stark wachsen oder schwinden (z. B. Schwarze Pest),

werden durch die resultierenden Kennzahlen nicht angemessen charakterisiert. Ist gesichert, dass man einer solchen Serie arbeitet, sollte man auf das Erstellen einer konventionellen Sterbetafeln und das Berechnen von  $e_0$  oder  $a_{20}$  verzichten.

#### 25.4.6 Alterszusammensetzung IV: Demographie der Erwachsenen

Als erste Übersicht stellt man aus den Rohdaten **Abb. 104** wiederum eine Tabelle mit den Häufigkeiten und Prozentwerten zusammen (**Abb. 110**). Da die nur grob als „erwachsen“ Altersbestimmten sowie die Geschlechts-Unbestimmten dort sauber ausgewiesen sind, können sie in der Tabelle hier als nicht weiterführend ausgeklammert werden.

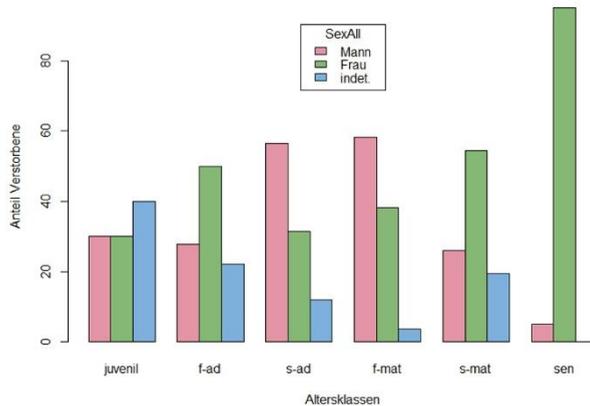
	Männer	%	erw. / Chi	Frauen	%	erw. / Chi
frühadult	10	9,3	13.7 / 1.0	18	16,1	14.3 / 0.9
spätadult	52	48,6	39.6 / 3.9	29	25,9	41.4 / 3.7
frühmatur	32	29,9	25.9 / 1.4	21	18,8	27.1 / 1.4
spät matur	12	11,2	18.1 / 2.0	25	22,3	18.9 / 2.0
senil	1	0,9	9.8 / 7.9	19	17,0	10.2 / 7.5
Summe	107	100		112	100	

**Abb. 110** Übersicht zum Sterbealter der Männer und Frauen in Zwentendorf (n=219). Die Prozentwerte sind als Spaltenprozent berechnet. Die zusätzliche Spalte zeigt „erw.“ den Erwartungswert und „chi“, den Chi-Wert der jeweiligen Zelle. Beobachtete Werte, die erheblich über dem Erwartungswert liegen, sind rot gefärbt, solche, die erheblich unter dem Erwartungswert liegen, grün gesetzt.

Das Balkendiagramm **Abb. 111**, in das wir die Jugendlichen und Unbestimmten bewusst eingeschlossen haben, visualisiert diese Daten, aber das wesentliche Ergebnis ist vor allem an der Tabelle **Abb. 110** ablesbar: Hier haben wir zusätzlich zu den üblichen Prozentzahlen in einer weiteren Spalte die Erwartungswerte eingetragen und das resultierende Zellen-Chi. Da der kritische Wert von Chi-Quadrat bei einer 2x2-Tafel 3,84 beträgt, indizieren Zellen-Chi-Werte von 3,84 oder höher eine sehr deutliche Abweichung zwischen beobachtet und erwartet, aber auch halb so hohe Chi-Werte (d. h. höher 1.92) verweisen auf auffällige Differenzen. Demnach unterscheiden sich die Sterblichkeiten für Männer und Frauen in Zwentendorf hochsignifikant, wobei Männer als spätadulte häufiger sterben und Frauen als spätmature und senile Individuen (d.h. die späte Adultas und frühe Maturitas rel. oft überlebend).

Die entsprechenden Kennzahlen aus der Sterbetafel lauten: Männer  $e_0$  19,9 Jahre,  $a_{20}$  36,8 Jahre, Frauen  $e_0$  22,7 Jahre,  $a_{20}$  41,9 Jahre – die Zwentendorfer Frauen wurden also im Mittel ca. 5 Jahre älter als die Männer, was ein für die Urgeschichte oder das Frühmittelalter ungewöhnlicher Befund ist. Doch die obige Darstellung als

Tabelle lässt einerseits besser erkennen, in welchem Alter was passiert, und vor allem erlaubt sie, die Unterschiede auch auf ihre Signifikanz zu testen. Es ergibt sich für die gesamte Tabelle ein Chi-Quadrat von 31.77 mit einem p-Wert von 0.000002132, d. h. die Unterschiede zwischen Frauen und Männern in Zwentendorf sind hochsignifikant. Das ist keinesfalls immer gegeben, die Möglichkeit, mit einer Teststatistik arbeiten zu können, ist sehr wertvoll.



**Abb. 111** Alters- und Geschlechtsverteilung der Zwentendorfer Jugendlichen und Erwachsenen (n=324).

Jenseits von Zwentendorf ist die im Vergleich zu Männern erhöhte Sterblichkeit junger Frauen ein nahezu regelhaft beobachtetes Phänomen. Man verbindet sie mit dem höheren Risiko für Frauen in der reproduktiven Phase. Frauen, die diese Lebensphase überlebt haben, hatten eine gute Chance, älter zu werden. Auch die anschließend erhöhte Sterblichkeit der Männer ist ein gängiges Phänomen. Relevant und je nach Population unterschiedlich ist vor allem der konkrete Altersbezug dieser „Peaks“: das ausgeglichene Bild bei den Zwentendorfer Jugendlichen zeigt an, dass die reproduktive Phase der Zwentendorfer Frauen wohl erst im Alter frühadult (i.e. 20-30) einsetzte und Jugendliche nicht betraf – das ist z. B. bei vielen bandkeramischen und frühbronzezeitlichen Serien anders. Das weitere Anwachsen der Sterblichkeit der Frauen in die Klasse spätadult könnte anzeigen, dass die reproduktive Lebensphase nicht auf ein Jahrzehnt begrenzt war – was wiederum in Zusammenhang zu setzen wäre mit den zuvor angeschnittenen Themen Intergeburtenabstand und Kindersterblichkeit. Relativ ungewöhnlich ist die (sehr) hohe Sterblichkeit der spätadulten Männer, die in diesem

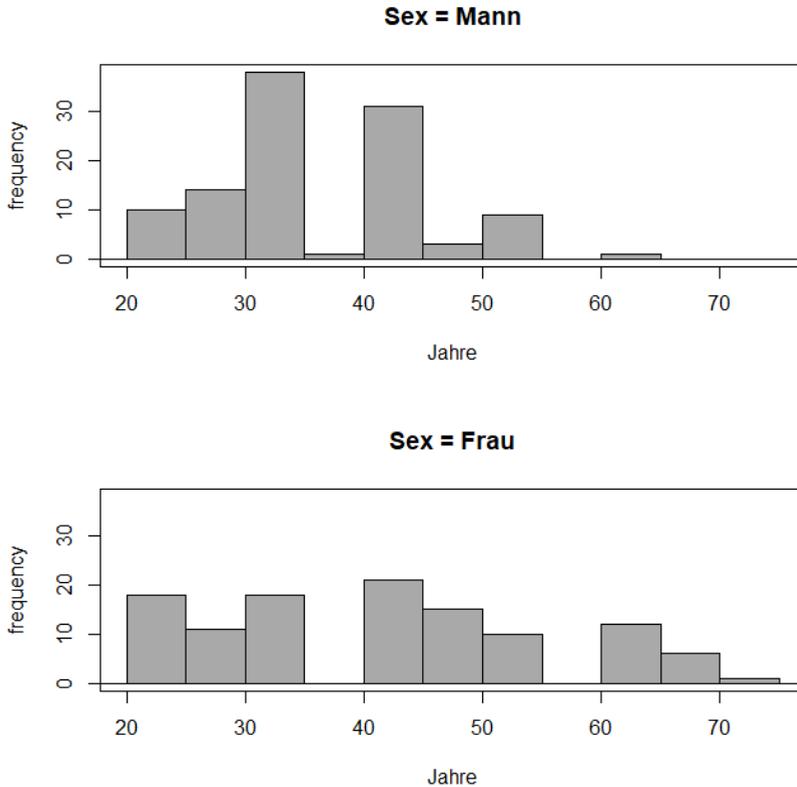
Ausmaß nur selten beobachtet wird – doch dies tiefergehend zu analysieren und zu interpretieren, ist nicht Aufgabe dieses Kapitels.

Noch kurz zum Praktischen: Für **Abb. 111** wurden aus dem Ausgangsdatensatz nur die Kinder weggefiltert, die Jugendlichen hingegen belassen (Filter: Age > 2). Weil der R-Commander in den Grafiken jedoch die leeren Klassen mitaufführt, was hier störend wäre, müssen diese zuvor noch entfernt werden: => Variablen bearbeiten; => Nicht verwendete Faktorstufen löschen; „AltersKI“ auswählen & OK. Dann: => Grafiken; => Balkendiagramm; „AlterKI“ ausgewählt, bei Grafiken für die Gruppen „SexAll“ eingesetzt; => beim Reiter „Optionen“ als Skalierung der Achse ausnahmsweise „Prozente“ gewählt, als „Art der Balkengrafik“ „Parallel nebeneinander“ gewählt, die „Position der Legende“ mit „Zentrum“ gesetzt und rechts die Achsen entsprechend beschriftet.

Inhaltlich wesentlich war hier darzulegen, dass die Zurückführung auf testbare Tabellen an Stelle oder ergänzend zu den Sterbetafeln wieder Sicherheit in der statistischen Bewertung von Unterschieden gibt. Was hier modellhaft am Vergleich zwischen Männern und Frauen innerhalb einer Population aufgezeigt wurde, eignet sich umso mehr für den Vergleich zwischen verschiedenen Populationen.

Die bei solchen Betrachtungen übliche Analyse mit Sterbetafeln und nach der (ggf. unterschiedlichen) Häufigkeit von Altersklassen wie hier in **Abb. 110** verstellt den Blick dafür, dass die zu Grunde liegenden anthropologischen Bestimmungen ein weitaus höheres Skalenniveau aufweisen als Nominaldaten. Die Altersbestimmungen folgen einer quantitativen Skala, genauer einer Verhältnisskala mit echtem Nullpunkt. Die Tatsache, dass an Stelle von Punktschätzungen, d. h. der Angabe eines wahrscheinlichen Sterbejahrs (wie bei Zwenkendorf) meist ein Intervall angegeben wird (von – bis, älter als & jünger als), um die Ungenauigkeit der Schätzungen zu signalisieren, ändert daran wenig. Denn zu den Intervallen kann ihre Mitte berechnet werden, und die Intervallmitten folgen einer Verhältnisskala. Dass in der Praxis dann meist nicht alle Werte des Zahlenstrahls besetzt sind, d. h. de facto keine stetige Skala vorliegt, tut dem keinen Abbruch. Folglich sind Statistiken jenseits von Häufigkeiten, Prozent und Chi-Quadrat-Tests anwendbar. Da die Annahme des Vorliegens einer Normalverteilung wiederum meist nicht berechtigt ist, können die Altersbestimmungen gut mit den Instrumenten nicht-parametrischer Verfahren beleuchtet werden. Median, Quartile und Rangtests sind adäquate Werkzeuge, sowohl für populationsinterne Vergleiche wie

etwa das hier verfolgte Frauen-Männer-Beispiel aus Zwentendorf, insbesondere aber auch für Vergleiche zwischen Populationen.



**Abb. 112** Histogramm der Sterbealterbestimmungen für die Frauen und Männer in Zwentendorf. Vgl. **Abb. 113**.

Gehen wir also für Zwentendorf kurz auch diesen Weg und vergleichen die Mittelwerte Männer / Frauen mit dem nicht-parametrischen Rangtest nach Kruskal und Wallis (**Abb. 113**).

	mean	sd	0%	25%	50%	75%	100%	n	NA
alle	40.86345	12.22098	20	35	35	50	75	249	45
Mann	38.69159	8.857798	20	35	35	45	65	107	9
Frau	43.83929	14.410053	20	30	45	55	75	112	7

Kruskal-Wallis rank sum test  
 data: Jahre by Sex  
 Kruskal-Wallis chi-squared = 6.4621, df = 1, p-value = 0.01102

**Abb. 113** Zusammenfassende Statistiken für das Sterbealter der Erwachsenen (n=249) in Zwentendorf samt Rangtest auf Signifikanz des Unterschieds Männer - Frauen.

Das mittlere Sterbealter der Erwachsenen – Kinder und Jugendliche sind hier aus bekannten Gründen wieder ausgeklammert – liegt demnach bei 40,9 Jahren, der Unterschied zwischen Männern (38,7 Jahre) und Frauen (43,8 Jahre) ist nach Ausweis des Kruskal-Wallis-Tests statistisch signifikant. Eine adäquate Grafik bietet das Histogramm **Abb. 112**.

Ich empfehle, für die Demografie zusätzlich zu den traditionellen Sterbetafeln auch diesen Weg der Häufigkeitstabellen mit Chi-Quadrat-Tests und der Analyse der Jahre als ganz gewöhnliche quantitative Skala zu verfolgen, denn die beiden letztgenannten Ansätze bieten erhebliche Vorteile beim Populationsvergleich.

#### 25.4.7 Schätzung der Lebendbevölkerung

Aus der Anzahl der Verstorbenen, ihrer mittleren Lebenserwartung und der untersuchten Zeit kann die Größe der Lebendbevölkerung geschätzt werden. Die Schätzformel lautet:<sup>105</sup>

$$\text{Lebendbevölkerung} = \text{Anzahl der Verstorbenen} * e_0 / \text{Zeitspanne}$$

In der ursprünglich publizierten Fassung dieser Formel wurde noch ein „Korrekturfaktor“ von plus 10 % hinzuaddiert zum Ausgleich von Abwanderung und andernorts erfolgten Bestattungen, doch dieser Wert ist willkürlich gesetzt und nicht begründbar: Man sollte auf diesen Faktor verzichten.

Nach dem oben Gesagten zum Thema  $e_0$  und der großen Abhängigkeit von der (unsicheren) Anzahl der Kinder könnte man dazu neigen, diese Schätzung für problematisch zu halten. Doch das ist nicht der Fall. Denn die Anzahl der Verstorbenen hängt mit  $e_0$  gegenläufig zusammen: Löscht man alle Kinder aus der Sterbetafel, steigt  $e_0$  an, aber die Anzahl der Verstorbenen sinkt, und vice versa. Es lohnt, all dies

---

<sup>105</sup> Herrmann u.a. 1990, 311-312.

einmal spielerisch zu simulieren und die Ergebnisse der Bevölkerungsschätzungen zu protokollieren: das Ergebnis ist weitgehend stabil.

Im Falle von Zwentendorf, dessen Belegungsdauer nach J. Giesler (1997) auf etwa 125 Jahre angesetzt werden kann, ergeben sich:

$595$  (alle Toten)  $\times$   $22.4$  ( $e_0$ ) /  $125 = 106$  gleichzeitig lebende Individuen, resp.

$294$  (Erwachsene)  $\times$   $41.9$  ( $a_{20}$ ) /  $125 = 99$  gleichzeitig lebende Individuen.

Die Anzahl der in Zwentendorf gleichzeitig Lebenden kann demnach auf etwa 100 Individuen geschätzt werden.

## **25.5 SOZIALE FRAGESTELLUNG AM BEISPIEL SANDHOFEN**

Frauen und Männer, Alte und Junge – wie spiegeln sich solche sozialen Gruppen in den Grabbeigaben wider, was kann man aus der Relation der Grabbeigaben zu den biologischen Parametern ablesen und daraus über die horizontalen Sozialstrukturen einer Gesellschaft ableiten? Arm und reich, Adel? Die Frage nach der vertikalen Differenzierung einer Gesellschaft beschäftigt uns auch in der aktuellen Politik, die in den 2010er Jahren die sich in vielen westlichen Gesellschaften wieder weitende Schere zwischen den ganz Reichen und der Normalbevölkerung beklagt (z. B. Atkinson 2015; Scheidel 2017). Diese Fragen können auch an vergangene Gesellschaften gestellt werden, wie z. B. gab es ererbten Status („Adel“); wurden Männer durchweg besser oder schlechter behandelt als Frauen, usw. Wir wollen dem Thema anhand eines frühmittelalterlichen Gräberfeldes im Raum Mannheim nachgehen, das in einem qualitätvollen Katalog gut publiziert ist (Koch 2018), zu dem jedoch eine ausführliche Auswertung bislang fehlt. Immerhin: die erfahrene Frühmittelalterarchäologin Ursula Koch, die das Material vorgelegt hat, hat zu wichtigen Themen kurze Einordnung in der Publikation vorgenommen, die hier als Grundlage verwendet werden können.

### **25.5.1 Theorie, Modelle, Ansatzpunkte, Gegenwartsbezug**

In den Anfängen der wissenschaftlichen Archäologie dominierten für sozialgeschichtlichen Fragestellungen Modelle, die eine direkte Gleichsetzung von Grabausstattung und Bestattungsaufwand einerseits und sozialen Gruppen andererseits vornahmen, wie es z. B. das beliebte Fachwort „Fürstengrab“ verdeutlicht. In der Frühmittelalterarchäologie wurden beispielsweise bestimmte Waffenausstattungen direkt mit Begriffen verknüpft, die sich in den zeitgenössischen Schriftquellen finden, wie z. B. frei / unfrei, Reiter, Adliger, oder auch Romane versus Franke (s. Lit.-Verweise bei Steuer 1994, Siegmund 2015). In den 1970er Jahren wollten (unabhängig voneinander) insbes. Michael Gebühr (1976) für die Römische Kaiserzeit und Rainer Christlein (1973) für die

Merowingerzeit diese alten Systeme ablösen; beiden schwebte vor, den Reichtum von Grabausstattungen zunächst beschreibend zu messen, d. h. objektiv zu quantifizieren, und erst danach soziale Deutungen zu entwickeln. Als mit ihrem Material erfahrene Archäologen definierten sie Werte einzelner Fundgattungen (Gebühr) bzw. Kriterien, um Grabfunde vier Reichtumsgruppen zuweisen zu können (Christlein: „Qualitätsgruppen“ A-D). In der Frühmittelalterarchäologie wird das System von Rainer Christlein – trotz aller richtigerweise an ihm geübten Kritik – bis heute weiterhin angewandt. Wegen des Fehlens von „Adelsgräbern“ in Sandhofen und um die lokalen Gegebenheiten besser fassen zu können, hat U. Koch für den Raum Mannheim ein an das Vorgehen von Christlein angelehntes, feinmaschigeres Kriteriensystem entwickelt und die Sandhofener Gräber danach zugeordnet und analysiert (Koch 2018, 278-301). Ihre Qualitätsgruppen von 1 bis 5 (reich -> arm) wurden in unseren Datensatz aufgenommen und dienen als Vergleich zu den beiden nachfolgend dargelegten Systemen.

Auf Kritik an den im System Gebührs vorgenommenen Setzungen reagierend, entstanden in Skandinavien für die Römische Kaiserzeit zwei unterschiedliche Systeme, die sich rein auf das Zählen und Messen beschränkten, d. h. ohne subjektive Faktoren auskommen wollten. Lotte Hedeager (1980) entwickelte als Maßzahl den – je nach Publikationssprache unterschiedlich abgekürzten – Begriff AAT (Anzahl von Artefakttypen) bzw. NAT (*number of artefact types*): sie summiert die Anzahl der in einem Grab vorkommenden unterschiedlichen Fundgattungen. Als Kontrollvariable ermittelte sie, ob darunter Objekte aus Gold sind, um diese Tatsache dem AAT-Wert gegenüberstellen zu können. Der Vergleich bei Hedeager (1980) zeigt, dass die Anwesenheit von Gold und hohe AAT-Werte in der Römischen Kaiserzeit Skandinaviens miteinander einhergehen.

Lars Jørgensen (1987) entwickelte eine feingliedrigere Messung des Werts der Fundgattungen. Seine Ausgangsthese ist ebenso einfach wie plausibel: Häufig vorkommende Objektgattung sind relativ wertärmer, selten vorkommende Gattungen sind relativ wertvoller. Also dividiert er die Anzahl der Gräber durch die Häufigkeit einer Fundgattung und nimmt das Ergebnis als Wert dieser Fundgattung. Demnach hätte ein Pferd als Beigabe, das nur einem der Sandhofener Gräber vorkam, den Wert  $281 / 1 = 281$ ; Keramikgefäße kommen in Sandhofen insgesamt 122 Mal vor, so dass ein Gefäß den Beigabewert  $281 / 122 = 1,48$  hat. Aus der Summe dieser einzelnen Beigabewerte (Jørgensen: *type value*) in einem Inventar ergibt sich dann der „Grabwert“ (Jørgensen: *grave value*).

Bei der Bearbeitung des kaiserzeitlichen Gräberfeldes von Costedt (Siegmond u. a. 1996, insbes. 100-105) wurden die drei Systeme von Gebühr, Hedeager und

Jørgensen parallel verfolgt und verglichen. Unterschiede sind gegeben, aber die Ergebnisse aller Systeme sind stark und hochsignifikant miteinander korreliert, sie führen zu sehr ähnlichen Einschätzungen der konkreten Gräber. Anlässlich einer Studie zu merowingerzeitlichen Gräbern mit besonderen Bügelfibeln konnte ich zeigen, dass das Konzept von Jørgensen auch gewinnbringend auf frühmittelalterliche Gräber übertragen werden kann (Siegmond 2015).

In den modernen Wirtschafts- und Sozialwissenschaften wird soziale Gleichheit oder Ungleichheit mit dem 1921 vorgeschlagenen Gini-Index formuliert, wobei Null der Zustand gänzlicher sozialer Gleichheit ist und 1 den extrem möglichsten Unterschied zwischen arm und reich formuliert. Da er auf Einkommensangaben von Bevölkerungsgruppen beruht, gibt es nicht nur methodische und praktische Probleme bei dessen Berechnung (was ist Einkommen? – man denke an die Berücksichtigung von Versicherungen, Rentenansprüchen und Immobilien, man denke an die Berücksichtigung von Kapital), sondern auch die Frage, inwieweit das Konzept überhaupt auf Gesellschaften, die keine Geldwirtschaft praktizieren, übertragen werden kann. Vielfach werden daher neben dem Gini-Index auch andere Kennzahlen herangezogen (Atkinson 2015, 17-62) – Kennzahlen, die man anhand des AAT-Systems oder der Grabwerte nach Jørgensen durchaus auch für urgeschichtliche Gesellschaften ermitteln kann: Wie hoch ist der Anteil am Ganzen (Gesamteinkommen; Bruttosozialprodukt), den die 1 % Reichsten oder auch die obersten 10 % einer Gesellschaft inne haben? Wie groß ist der Anteil an Armen in einer Population? – wobei der Median des Einkommens der relevanten Gesamtpopulation als Bezugspunkt gesetzt wird und diejenigen, die 60 % dieses Medians unterschreiten, als „arm“ klassifiziert werden.<sup>106</sup>

Anhand solcher Parameter wären prähistorische und frühgeschichtliche Populationen mit modernen Situationen vergleichbar. Einschränkend muss man jedoch

---

<sup>106</sup> Um diese abstrakte Zahl zur Armutsgrenze etwas anschaulicher zu machen: In Deutschland lag der Medianwert des monatlichen Nettoeinkommens einer einzelnen Person im Jahr 2019 bei ca. 1.870 €, 60 % davon sind ca. 1.120 €. Dieser Betrag liegt zwar deutlich über dem monatlichen Hartz-IV-Satz, der im Jahr 2019 für eine erwachsene alleinstehende Person monatlich 424 € betrug. Da in solchen Fällen aber eine als angemessen geltende Miete samt Strom- und Heizkosten (Größenordnung insges. 550 €) sowie die Krankenversicherung (ca. 85-90 €) ebenfalls direkt über die Sozialversicherung („Jobcenter“) getragen wird, ergibt sich in Summe ein (wenn auch nicht bar verfügbares) monatliches Einkommen von ca. 1.060 €. Diese Rechnung verdeutlicht, was die in den Wirtschafts- und Sozialwissenschaften oft angesetzte Armutsgrenze von 60 % des Medians aller Einkommen einer Gesamtpopulation bedeutet: Sie liegt in Deutschland im Jahr 2019 mit ca. 60 € Differenz ganz knapp über einem Leben auf Hartz-IV-Niveau.

– auch ohne die unten dargelegten eigenen Beobachtungen einzuflechten – nach Sichtung der Literatur feststellen, dass die sehr unterschiedlichen Grundlagen, die je nach Quellenlage bei einer historischen Betrachtung in solche Kennzahlen eingehen müssen, ihre globale Vergleichbarkeit (stark) einschränken. Die Aussagen z.B. bei Scheidel (2017) wirken dort besonders tragfähig und vergleichbar, wo innerhalb einer stärker konsistenten Informationslage Zeitreihen gelegt werden können, um Entwicklungen aufzuzeigen. Dies dürfte auch an prähistorischen Daten immer wieder möglich sein.

Für die Erfassung von Sandhofen mit sozialgeschichtlicher Fragestellung resultiert aus der Sichtung der gegebenen Modelle und Ansätze folgende Vorgehen:<sup>107</sup> erfasst werden die anthropologische Alters- und Geschlechtsbestimmung; die von U. Koch angegebenen Datierung des Grabes; die Größe und Eigenheiten des Grabes (Aufwand der Grablegung); die Anzahl der Gold und Silberobjekte im Grab (Kontrollvariable); alle vorkommenden Fundgattungen sowie deren jeweilige Häufigkeit im Grab; die Einordnung der Gräber nach den von U. Koch gesetzten Qualitätsgruppen.

### 25.5.2 Datensichtung und -aufbereitung

Wie üblich, muss am Beginn einer Auswertung der erfasste Datensatz gesichtet und z. T. aufbereitet werden. Für die biologische Alters- und Geschlechtsbestimmung liegen zwei unterschiedliche Bearbeitungen vor (vgl. Koch 2018, 10). Nach einer Sichtung erkennen wir, dass beide nicht selten zu unterschiedlichen Ergebnissen führen. Wir beschließen, die umfassendere anthropologische Untersuchung von Sv. Partheil zur Grundlage zu machen und die Ergebnisse von M. Teichmann, die im Grunde auf die Zähne zielte, nur da einzusetzen, wo erstere fehlende Werte aufweist. Danach beobachten wir 125 Männer zu 56 Frauen, was ein ungewöhnlich starkes Ungleichgewicht der Geschlechter ist. Auch wenn die biologischen Geschlechtsbestimmungen von Partheil („Sex“) in wiederum ungewöhnlicher Häufigkeit von den archäologischen Geschlechtsbestimmungen *sensu* soziales Geschlecht („Gender“) durch U. Koch abweichen (in 14 von 173 Fällen, i. e. 8 %), ergibt sich bei der Bestimmung des sozialen Geschlechts ein Verhältnis von 128 Männern zu 56 Frauen, d. h. eine ähnlich starke Abweichung. Statt der leider üblichen Berechnung des Maskulinitätsindex testen wir die Häufigkeiten des biologischen Geschlechts gegen die Erwartung von 105 : 100

---

<sup>107</sup> Der hier verwendete Datensatz wurde im Sommersemester 2019 anlässlich einer vom Verf. geleiteten Statistik-Übung an der Univ. Münster gemeinsam erarbeitet von: V. Becker, Chr. Dües, S. Geilhaupt L. Kopner, M.-M. Mancini und P. Tarner. Ich danke allen Teilnehmern des Kurses herzlich für ihr großes Engagement bei der gewissenhaften gemeinsamen Erarbeitung des Aufnahmeschemas und der Datenerfassung.

(hier also im Hinblick auf die Population von 181 bestimmten Individuen als 93 Männer zu 88 Frauen)<sup>108</sup> mit Hilfe von => Statistik; => Kontingenztabelle; => „Analyse einer selbst eingegebenen Kreuztabelle“ und erhalten ein Chi-Quadrat von 11.8 mit einem p-Wert von ca. 0.0058, d. h. hochsignifikant. Bilanz: Sandhofen ist signifikant ein ungewöhnliches, männerdominiertes Gräberfeld.

Die Altersbestimmungen wurden zunächst richtigerweise aus der Publikation heraus im System „von Alter“ / „bis Alter“ erfasst, d. h. die Unter- und die Obergrenze der angegebenen Spanne notiert. Für die sozialgeschichtliche Auswertung bedarf es einer gewissen Bereinigung und Vereinheitlichung, d. h. ein Überführen in Altersklassen – auch wenn man für Einzelfallprüfungen gewiss auch immer auf die Originaldaten zurückgreifen wollen. Dafür sind in der biologischen Anthropologie jenseits von Kindern und Jugendlichen – je nach Bearbeiter – Klassenweiten von 5 Jahren, 7 Jahren (früh-, mittel-, spätadult, usw.), 10 Jahren (früh-, spätadult, usw.) oder 20 Jahren (adult, matur, senil) üblich. Welche Klassenbildung ist angemessen? Wir errechnen die Altersspanne („bis“ minus „von“), filtern die Erwachsenen heraus (Alter „von“ größer gleich 18) und stellen fest: ca. 50 Mal beträgt sie 60 Jahre, wovon sich die lediglich als „erwachsen“ Bestimmten (20-80 Jahre) verbergen. Ca. 70 Mal beträgt die Spanne 10 Jahre, ca. 20 Mal 20 Jahre. Also war seitens der Bearbeiterin eine recht detaillierte Altersbestimmung gewünscht und oft auch möglich, so dass wir uns als Zielsystem für ein System mit 10-Jahres-Klassen entscheiden. Das nachfolgende Skript, das wir auch als Übungsdaten beilegen, erledigt diese Aufgabe. Am besten vollzieht man es Schritt für Schritt nach; zwischendurch jedes Mal die aktive Datenmatrix aktualisieren und die Matrix betrachten. Das resultierende System soll sein: 0 unbestimmt; 1 infans I; 2 infans II; 3 juvenil; 4 früh-adult; 5 spät-adult; 6 früh-matur; 7 spät-matur; 8 früh-senil; 9 spät-senil; 10 erwachsen.

```
#
# Überführung der Altersangaben von - bis
# in Altersklassen infans I, infans II, juvenil,
# und 10-Jahres-Klassen für Erwachsene, so möglich
#
# Als Namen des Datensatzes haben wir „sh“ für Sandhofen verwendet.
#
sh$alterskl <- with(sh, 0)
sh$alterskl <- (((sh$von + sh$bis)/2) %/%10) +2
#
```

---

<sup>108</sup> Siegmund, F. (2009). Der Maskulinitätsindex und statistische Verfahren zur Prüfung auf Unterschiede in den Geschlechteranteilen frühmittelalterlicher Populationen. *Bulletin der Schweizerischen Gesellschaft für Anthropologie* 15, 5-17.

```

# nun Kinder korrigieren:
sh$alterskl <- ifelse(((sh$von >=0) & (sh$bis <=7)), 1, sh$alterskl)
sh$alterskl <- ifelse(((sh$von >=7) & (sh$bis <=14)), 2, sh$alterskl)
sh$alterskl <- ifelse(((sh$von >=14) & (sh$bis <=18)), 3, sh$alterskl)
#
# nun bei Altersspanne grösser gleich 20 korrigieren auf "erwachsen"
sh$agespan <- (sh$bis - sh$von)
sh$alterskl <- ifelse((sh$agespan >=20), 10, sh$alterskl)
#
# nun zusätzlich Variable als Faktor ablegen
sh <- within(sh, {
  f.Altersklasse <- factor(alterskl, labels=c('inf. I','inf. II','juv.','f-ad.','s-
ad.','
'f-mat.','s-mat.','f-sen.','s-sen.','erw.'))
})
###

```

Abschließend wird die numerische Variable „alterskl“ noch in einen Faktor gewandelt, und zwar unter eigenem Namen („f.Altersklasse“), damit sie für das Folgende wahlweise numerisch oder als Faktor zur Verfügung steht. Im Ergebnis haben wir für Sandhofen 42 Kinder und Jugendliche (i. e. 16,6 %),<sup>109</sup> 127 genauer altersbestimmte Erwachsene und 84 als „erwachsen“ bestimmte Individuen. Die weitere Auswertung solcher Informationen hatte wir schon im Kapitel Zwentendorf eingehend besprochen, hier werden die Altersklassen nur für den Vergleich mit der Werteinschätzung der Grabausstattungen benötigt.

Als nächstes könnte man die von U. Koch vorgenommene Zuordnung der Gräber zu Qualitätsgruppen („Qualgr“) sichten: => Statistik; => Kontingenztabelle; => Kreuztabelle, Gender und Qualgr auswählen, OK (**Abb. 114**). Der Chi-Quadratstest zeigt, dass es nach dem Qualitätsmodell von U. Koch keine signifikanten Unterschiede zwischen Männern und Frauen gibt.

Wir werfen einen Blick in den Altersbezug der Qualitätsgruppen: => Statistik; => Mittelwerte vergleichen; => Einfaktorielle Varianzanalyse; setzen „f.Altersklasse“ als Grupperungsvariable ein und „Qualgr“ als abhängige Variable. Die resultierende Tabelle der jeweiligen Mittelwerte zeigt, dass für Kinder keine Daten vorliegen und der Mittelwert für alle Alter zwischen 2.3 und 3.1 recht dicht beieinander liegt, woraus sich angesichts der relativ großen Standardabweichungen keine signifikanten Unterschiede ergaben (p-Wert ca. 0.79). Nach diesem ersten Eindruck mit dem System der

---

<sup>109</sup> Das erscheint auf den ersten Blick wenig, doch siehe dazu das Kap. Demografie Zwentendorf: Kinderanteil.

Qualitätsgruppen gehen wir zunächst zum Thema AAT – Anzahl von Artefakttypen über.

Qual.-gruppe	Mann	Frau	alle	%
1	7	6	14	9,4
2	37	15	53	35,6
3	24	14	42	28,2
4	23	17	40	26,9
Summe	91	52	149	100,0

**Abb. 114** Häufigkeiten der Qualitätsgruppen nach U. Koch in Sandhofen.

### 25.5.2.1 Errechnen des AAT-Werts

Die diesbezügliche Datenaufbereitung ist etwas „tricky“. Es liegen vor 31 Variablen mit der jeweiligen Häufigkeit dieser Gattung in einem Grab, wobei die meisten Zellen leer sind, d. h. die betreffende Fundgattung in diesem Grab nicht vorkommt. Jede nicht-leere Zelle steht für 1 AAT, und wir müssen nun deren Summe pro Zeile zum AAT-Wert des Grabes aufaddieren. Wir benutzen dazu die Funktion „is.na“, die abfragt, ob es einen gültigen Wert gibt und – je nach Befund – eine „1“ ausgibt, sofern ja = fehlender Wert. Die Summe dieser Abfragen über alle Variablen ergibt demnach die Anzahl der leeren Zellen in einem Grab, die Differenz 31 minus dieser Summe ist der AAT-Wert. Ja, es mag auch andere Wege geben, mit Hilfe von **R** zum Ziel zu gelangen, der hier eingeschlagene ist nur ein möglicher Weg. Daraus resultiert folgenden Skript (als Übungsdatei beiliegend):

```
# Ermitteln des AAT-Werts für jedes Grab
sh$aat <- 31 - (is.na(sh$Muenze) + is.na(sh$Anhaenger) + is.na(sh$Perlen) +
  is.na(sh$Fibeln) + is.na(sh$Ringe) + is.na(sh$Guertel) + is.na(sh$GGH) +
  is.na(sh$FussSchn) + is.na(sh$Geraet) + is.na(sh$Kamm) + is.na(sh$Spatha) +
  is.na(sh$Spathagarn) + is.na(sh$Schild) + is.na(sh$Lanze) + is.na(sh$Sax) +
  is.na(sh$Axt) + is.na(sh$Pfeilspitze) + is.na(sh$GGunverz) + is.na(sh$GGtausch) +
  is.na(sh$Tasche) + is.na(sh$Pinzette) + is.na(sh$Schmied) + is.na(sh$Feuerstahl) +
  is.na(sh$Reitgeschirr) + is.na(sh$Pferd) + is.na(sh$Messer) + is.na(sh$Feile) +
  is.na(sh$Keramikgef) + is.na(sh$Glasgef) + is.na(sh$Holzgef) + is.na(sh$Bronze-
gef))
#
```

Ein Histogramm über „aat“ liefert uns einen ersten Einblick: Der maximale Wert in Sandhofen ist 12, doch die meisten Gräber weisen kleine Werte auf; der Median liegt bei 2,0, der Mittelwert bei 2.4. Für einen ersten Eindruck wiederholen wir die Fragen, die auch an die Koch’schen Qualitätsgruppen gestellt wurden: wie steht es um den

Geschlechts-, wie um den Altersbezug? Dazu führen wir wie zuvor jeweils eine „Einfaktorielle Varianzanalyse“ durch und sehen: Im Sinne der AAT-Werte sind Männer im Mittel mit 3.2 etwas reicher als Frauen mit 2.7, der Unterschied ist statistisch nicht signifikant. Doch halt! – gemäß dem Histogramm der AAT-Werte gab es viele Beobachtungen nahe Null und (erwartungsgemäß) sehr wenige hohe Werte, weshalb der Mittelwert zwar informativ ist, aber ein parametrischer Test gänzlich unangemessen. Daher: => Statistik; => Nichtparametrische Tests; => Kruskal-Walls-Test; dort f.Altersklasse & aat einsetzen, OK. Der p-Wert ist deutlich nicht-signifikant, d. h. auch nach dem Konzept der AAT-Werte gibt es keine besonderen Alterseffekte – was historisch gesehen ein bemerkenswerter Befund ist. Doch wir wollen die Analyse hier nicht weiter vertiefen, weil wir immer noch im Stadium der Datenaufbereitung sind, sondern gehen zum Errechnen des Grabwertes nach L. Jørgensen (1987) über.

### ***25.5.2.2 Errechnen des Grabwertes nach L. Jørgensen (1987)***

Den Grabwert errechnen wir in einem mehrstufigen Prozess. Zunächst muss für alle 31 Fundgattungen festgestellt werden, wie oft sie in Sandhofen vorkommen. Das geht z. B. mit diesem Befehl:

```
sum(sh$Muenze, na.rm=TRUE)
```

...der dann für alle 31 Fundgattungen durchzuspielen ist. Als nächstes kann der Wert ermittelt werden, den jede Fundgattung hat, der von Jørgensen so genannte „type value“. Wir berechnen also 31 neue Variablen, die den jeweiligen Wert festhalten und benutzen dazu wieder die Funktion „ifelse“: Ist der Wert der Ausgangsvariable „missing“, wird eine Null eingetragen, ansonsten der Wert der Gleichung: Häufigkeit der Fundgattung im Grab \* (Anzahl Gräber insgesamt / Häufigkeit der Fundgattung insgesamt). Das resultierende Skript (wiederum den Übungsdaten beiliegend) sieht so aus:

```
#
# Logik Zuweisung: wenn Variable "missing", dann Werte-Variable = Null,
# wenn nicht "missing", dann Werte-Variablen = Wert nach Joergensen
#
sh$j.muenze <- ifelse((is.na(sh$Muenze)==TRUE), 0, (sh$Muenze*(281/9)))
sh$j.anhaenger <- ifelse((is.na(sh$Anhaenger)==TRUE), 0, (sh$Anhaenger*(281/8)))
sh$j.perlen <- ifelse((is.na(sh$Perlen)==TRUE), 0, (sh$Perlen*(281/423)))
sh$j.fibeln <- ifelse((is.na(sh$Fibeln)==TRUE), 0, (sh$Fibeln*(281/12)))
sh$j.ringe <- ifelse((is.na(sh$Ringe)==TRUE), 0, (sh$Ringe*(281/23)))
sh$j.guertel <- ifelse((is.na(sh$Guertel)==TRUE), 0, (sh$Guertel*(281/81)))
sh$j.ggh <- ifelse((is.na(sh$GGH)==TRUE), 0, (sh$GGH*(281/3)))
sh$j.fusschn <- ifelse((is.na(sh$FussSchn)==TRUE), 0, (sh$FussSchn*(281/1)))
sh$j.geraet <- ifelse((is.na(sh$Geraet)==TRUE), 0, (sh$Geraet*(281/23)))
sh$j.kamm <- ifelse((is.na(sh$Kamm)==TRUE), 0, (sh$Kamm*(281/91)))
sh$j.spatha <- ifelse((is.na(sh$Spatha)==TRUE), 0, (sh$Spatha*(281/9)))
sh$j.spathagarn <- ifelse((is.na(sh$Spathagarn)==TRUE), 0,
(sh$Spathagarn*(281/11)))
```

```

sh$j.schild <- ifelse((is.na(sh$Schild)==TRUE), 0, (sh$Schild*(281/32)))
sh$j.lanze <- ifelse((is.na(sh$Lanze)==TRUE), 0, (sh$Lanze*(281/28)))
sh$j.sax <- ifelse((is.na(sh$Sax)==TRUE), 0, (sh$Sax*(281/20)))
sh$j.axt <- ifelse((is.na(sh$Axt)==TRUE), 0, (sh$Axt*(281/3)))
sh$j.pfeilspitze <- ifelse((is.na(sh$Pfeilspitze)==TRUE), 0, (sh$Pfeil-
spitze*(281/45)))
sh$j.ggunverz <- ifelse((is.na(sh$GGunverz)==TRUE), 0, (sh$GGunverz*(281/22)))
sh$j.ggtausch <- ifelse((is.na(sh$GGtausch)==TRUE), 0, (sh$GGtausch*(281/9)))
sh$j.tasche <- ifelse((is.na(sh$Tasche)==TRUE), 0, (sh$Tasche*(281/8)))
sh$j.pinzette <- ifelse((is.na(sh$Pinzette)==TRUE), 0, (sh$Pinzette*(281/3)))
sh$j.schmied <- ifelse((is.na(sh$Schmied)==TRUE), 0, (sh$Schmied*(281/2)))
sh$j.feuerstahl <- ifelse((is.na(sh$Feuerstahl)==TRUE), 0, (sh$Feuer-
stahl*(281/11)))
sh$j.reitgeschirr <- ifelse((is.na(sh$Reitgeschirr)==TRUE), 0, (sh$Reitge-
schirr*(281/11)))
sh$j.pferd <- ifelse((is.na(sh$Pferd)==TRUE), 0, (sh$Pferd*(281/4)))
sh$j.messer <- ifelse((is.na(sh$Messer)==TRUE), 0, (sh$Messer*(281/71)))
sh$j.feile <- ifelse((is.na(sh$Feile)==TRUE), 0, (sh$Feile*(281/2)))
sh$j.keramikgef <- ifelse((is.na(sh$Keramikgef)==TRUE), 0,
(sh$Keramikgef*(281/122)))
sh$j.glasgef <- ifelse((is.na(sh$Glasgef)==TRUE), 0, (sh$Glasgef*(281/25)))
sh$j.holzgef <- ifelse((is.na(sh$Holzgef)==TRUE), 0, (sh$Holzgef*(281/4)))
sh$j.bronzegef <- ifelse((is.na(sh$Bronzegef)==TRUE), 0, (sh$Bronzegef*(281/2)))
#
# alle Werte addieren
sh$j.oergensen <- with(sh, j.muenze + j.anhaenger + j.perlen +
j.fibeln + j.ringe + j.guertel + j.ggh + j.fussschn + j.geraet +
j.kamm + j.spatha + j.spathagarn + j.schild + j.lanze + j.sax +
j.axt + j.pfeilspitze + j.ggunverz + j.ggtausch + j.tasche + j.pinzette +
j.schmied + j.feuerstahl + j.reitgeschirr + j.pferd + j.messer + j.feile +
j.keramikgef + j.glasgef + j.holzgef + j.bronzegef)
#

```

Nachdem die Einzelwerte generiert sind, kann deren Summe gebildet werden, der Grabwert nach Jørgensen. Auch hier gilt: der hier eingeschlagene Rechenweg ist ein möglicher, sicherlich führen auch andere Wege zum Ziel. Ein Histogramm gibt einen ersten Einblick: der maximale Grabwert liegt in Sandhofen bei ca. 350, doch die überwiegende Mehrheit der Gräber liegt im Bereich Null bis 100 – wiederum eine stark linkssteile Verteilung.

Wiederum folgt eine erste Sichtung wie zuvor: Der Mittelwert des Grabwerts ist für Männer mit ca. 42.3 größer als für Frauen (ca. 34.6), doch ist der Unterschied nicht signifikant (Kruskal-Wallis-Test). Bei den Altersklassen liegt der Mittelwert (besser: Median) für Juvenile und für Spätmatüre am höchsten; klammert man die Kinder und Jugendlichen aus, zeigt der Kruskal-Wallis-Test jedoch auch hier, dass die Unterschiede entlang der Altersklassen statistisch nicht signifikant sind.

Ein Streudiagramm sowie eine Korrelationsrechnung (Rangkorrelation) zeigen, dass der AAT-Wert und der Grabwert einen starken Zusammenhang miteinander aufweisen (Rangkorr.-koeff. Spearman +0.84), während der Zusammenhang mit den Qualitätsgruppen nach Koch in beiden Fällen gegeben, aber deutlich schwächer ist (-0.41 resp. -0.50). Wie bereits anhand von Costedt erarbeitet (Siegmond u. a. 1996, 100-105), ergeben sich jeweils sehr ähnliche Befunde, gleich, ob man mit AAT-Werten nach Hedeager oder mit den Grabwerten nach Jørgensen operiert. Ich persönlich ziehe letztere vor, da sich eine breiter aufgefächerte Skala ergibt.

Soweit das Gewinnen und erste Sichten der Kennwerte Qualitätsgruppen (Koch), AAT (Hedeager) und Grabwert (Jørgensen). Die nächsten Analyseschritte könnten sein, diese Kennwerte mit der Frage des Aufwandes der Graberrichtung zu vergleichen (Grabtyp, Grablänge und -breite, ggf. auch Tiefe), um zu prüfen, ob man diese Aspekte miteinbetten kann resp. möchte. Die Datentabelle enthält, unabhängig von den Fundgattungen erfasst,<sup>110</sup> die Variablen „Au“ = Gegenstand aus Gold und „Ag“ = Gegenstand aus Silber; auch hier liegt es nahe, deren eventuelle Integration in den Grabwert zu prüfen, sowie deren Verhältnis zu den Qualitätsgruppen, AAT-Werten und dem Grabwert. Nicht zuletzt ist die Frage des Grabraubes und dessen Auswirkungen zu untersuchen, denn in Sandhofen liegt mit 91,5 % ein sehr hoher Anteil an gestörten Gräbern vor (Koch 2018, 277). Doch diese tiefergehenden (eentlichen) Analysen seien nun dem selbständigen Entdeckertrieb des interessierten Lesers überlassen.

Methodisch sei noch angemerkt, dass der Grabwert nach Jørgensen eine m. E. sehr geeignete relative Skala ergibt, die stärker differenziert als die Qualitätsgruppen nach Christlein (1973) oder Koch (2018) oder das AAT-System, seine absolute Größe jedoch stark von der Größe des Gräberfeldes abhängig ist. Der seltenste vorkommende Gegenstand dürfte meist die Häufigkeit 1 haben, so dass dieser Gegenstand als Einzelobjekt in Sandhofen mit seinen 281 Gräbern eben den Einzelwert 281 erhält, in dem kleinen Gräberfeld von Costedt mit seinen 44 Bestattungen hingegen den Einzelwert 44. Um diesen Effekt zu vermeiden und damit auch unterschiedliche Gräberfelder miteinander vergleichbar zu machen, sollte man die Berechnung gegenüber dem Originalvorschlag von Jørgensen leicht verändern und die Zahl der Gräber auf 100

---

<sup>110</sup> d. h. eine goldene Fibel wurde 1x als Fibel erfasst und 1x als Goldartefakt. Daher sind die Variablen Au und Ag bei der ersten Kalkulation der Grabwerte ausgeklammert worden, um nicht ein Objekt zwei Mal zu zählen. Doch dieses Vorgehen kann durchaus hinterfragt werden.

(oder 1.000) normieren. Die Formel für den Typwert der Münzen sähe dann wie folgt aus:

```
sh$j.muenze <-  
ifelse((is.na(sh$Muenze)==TRUE), 0, (sh$Muenze*((281/9)*(100/281))))
```

Ich empfehle sehr, eine solche Normierung vorzunehmen, um unterschiedliche Gräberfelder vergleichbar(er) zu machen.

Die unnormierten Grabwerte der 1% reichsten Gräber in Sandhofen liegen bei je ca. 300 – 350, nach einer solchen Normierung haben sie Werte von ca. 107-125. In Costedt hatten die drei reichsten Gräber Werte von 50-110, woraus sich nach einer solchen Normierung Grabwerte von 115-250 ergäben. Die Mediane der Grabwerte lägen nach dieser Normierung bei ca. 4.2 (Sandhofen) und 15.9 (Costedt).

Zudem ist offensichtlich, dass der Typwert nach Jørgensen gerade für seltene Objekte und kleine Gräberfelder stark vom Zufall abhängig sein kann. Daher würde ich für eine valide Untersuchung lieber auf die Daten und Fundhäufigkeiten einer ganzen Region zurückgreifen, anstatt ein einzelnes Gräberfeld zur Grundlage zu machen.

### ***25.5.2.3 Errechnen des Gini-Index***

Für das Berechnen des Gini-Index möchte ich es uns einfach machen und ein fertiges R-Paket nutzen. Man installiere von CRAN das Paket „reldist“ und lade es in **R** oder RStudio. Zur Dateneingabe sind zwei Modi möglich: (a) man hat Klassen und die Anzahl der Fälle in diesen Klassen, oder (b) schlicht eine Liste mit „Einkommen“. Das nachfolgende Skript, das man z. B. in das linke obere Fenster von RStudio kopieren und dort editieren kann, zeigt beide Optionen, zunächst (a), dann (b). Die Daten des Beispiels sind die AAT-Werte von Sandhofen.

```
# Eingabe von in Klassen vorliegenden Daten:  
klassen <- c(0,1,2,3,4,5,6,7,8,9,10,11,12)  
faelle <- c(61,47,60,38,30,21,12,6,2,3,0,0,1)  
#  
cumul.klassen = cumsum(klassen)/sum(klassen)  
cumul.faelle = cumsum(faelle)/sum(faelle)  
#  
# sofern nötig, hinzufügen der Klasse 0:  
# cumul.klassen <- c(0,cumul.klassen)  
# cumul.faelle <- c(0,cumul.faelle)  
#  
# Ausgabe der Lorenzkurve  
plot(cumul.faelle,cumul.klassen,type="l")  
#
```

```

# berechnen Gini-Koeffizient mit Paket "reldist"
# - es muss zuvor installiert worden sein! -
library(reldist)
gini(klassen, faelle)
#
# nun die Eingabe als Liste einzelner, nicht klassierter Werte:
liste <- c(0,6,6,0,0,4,5,0,0,0,4,4,6,3,7,0,2,0,2,2,2,8,1,1,0,2,0,
+ 4,1,0,2,0,4,3,3,1,1,4,7,5,6,4,2,9,0,0,3,0,4,1,1,0,5,2,
+ 2,1,5,3,3,5,3,5,0,2,0,0,2,3,2,1,4,2,4,3,5,1,1,2,3,2,1,
+ 1,1,0,0,1,0,0,4,0,1,1,3,2,3,0,0,2,2,0,0,3,1,0,6,3,3,1,
+ 9,0,1,0,4,4,0,2,1,0,3,1,2,2,0,1,0,4,1,3,6,5,6,7,0,4,1,
+ 1,2,2,2,6,0,2,1,4,0,3,2,2,1,0,1,1,2,1,1,4,3,1,2,3,2,5,
+ 2,2,0,2,5,2,2,0,5,8,0,2,3,2,3,1,2,4,1,0,1,3,1,5,4,0,4,
+ 4,2,3,4,5,4,2,2,7,4,0,3,1,2,2,4,0,1,5,4,0,1,2,3,1,2,0,
+ 3,0,3,0,2,3,2,4,5,5,2,0,5,4,2,12,3,2,0,3,1,4,5,3,7,2,0,
+ 1,2,2,1,1,1,0,0,3,3,2,2,5,0,2,5,0,6,6,6,3,3,0,3,2,9,0,
+ 6,4,2,3,5,0,7,2,2,1,0)
gini(liste)
#

```

Wenden wir diese Berechnungen auf Sandhofen an, erhalten wir den Wert 0.457 auf Grundlage der AAT-Werte, von 0.699 auf Grundlage der Grabwerte nach Jørgensen und von 0.227 auf der Basis der Koch'schen Qualitätsgruppen – d. h. der resultierende Wert hängt sehr stark von der Art ab, wie das „Einkommen“ bzw. verfügbare Kapital gemessen wurde.

### 25.5.3 Vergleich der Befunde mit urgeschichtlichen und modernen Gesellschaften

Lassen Sie uns abschließend noch, wie in der Einleitung angedeutet, den Vergleich mit modernen Gesellschaften ziehen. Legen wir die (unnormierten) Grabwerte nach Jørgensen zu Grunde, besitzen in Sandhofen die obersten 1 % der Erwachsenenbevölkerung (i. e. die drei reichsten Gräber: 3, 33 und 232) insgesamt 979,5 Einheiten des Erwachsenen-Gesamtkapitals in Sandhofen von 6.895, d. h. ca. 14 % des (Grab-) Kapitals. Die „Top Ten“, die reichsten 10 % Gräber (Grabwerte größer 85,6) besitzen insgesamt 3.372 Einheiten, also ca. 49 % des Populationsgesamtkapitals. Orientieren wir uns am Median von 14,3 und der genannten 60 %-Schwelle (= 8,6), liegen in Sandhofen 39,7 % der Erwachsenengräber unter dieser relativen Armutsgrenze, sie besitzen insgesamt nur 165 Einheiten, d. h. 2,4 % des Gesamtkapitals.

Der Publikation von Atkinson (2015, S. 29 f. Abb. 1.1; vgl. Scheidel 2017, 3 Fig. I.1) entnehmen wir die Vergleichswerte für die heutige USA und England: die 1 % Reichsten besitzen ca. 18 % resp. 13 % des gesamten Kapitals der Gesellschaft. Der

Anteil der Armen (Anzahl der Erwachsenen mit einem mittleren Einkommen von unter 60% des Einkommens-Medians der Gesamtpopulation) liegt in den USA und Großbritannien bei knapp unter 25 % (Atkinson 2015, 39 Abb. 1.4).<sup>111</sup>

Die hier verwendeten Kennzahlen – wie viel besitzen die 1% Reichsten, wie viel die 10% Reichsten, und wie hoch ist der Anteil der unter der Armutsgrenze Lebenden – sind in Summe komplex, weshalb das Bedürfnis besteht, das Ausmaß der sozialen Gleichheit oder Ungleichheit in eine einzige Kennzahl zu fassen. Dazu dient der Gini-Index. Bei ihm ergibt die völlige Gleichheit Aller einen Wert von 0, die totale Ungleichheit (1 Mensch besitzt alles) den Wert 1.<sup>112</sup> In Sandhofen beträgt der Gini-Index 0,475 (nach AAT) bzw. 0,699 (nach Jørgensen). Zum Vergleich sei zunächst der Wert für Costedt (nach den Jørgensen-Werten in Siegmund u. a. 1996, 104 Tab. 34) von 0,587 angeführt. Danach ist die Costedter Population etwas gleicher als die Sandhofener. Im heutigen Europa liegt der Gini-Index bei 0,25 bis 0,30, in den USA etwas über 0,35 und in einigen südamerikanischen Ländern, Indien und China mit sehr starken sozialen Unterschieden erreicht er Werte bis 0,60 (Atkinson 2015, 35 Abb. 1.3; vgl. Scheidel 2017, 98 Fig. 3.2 mit Werten >0,8 um 1800 in Europa). Doch wie Atkinson darlegt und auch unser eigenes Sandhofen-Beispiel zeigt, ist die Berechnungsgrundlage zentral, d. h. welche Informationen in die Berechnung eingehen. Für Sandhofen führt die Betrachtung nach AAT und nach Jørgensen zu deutlich unterschiedlichen Werten des Gini-Indexes; doch immerhin können wir mit dem Gini-Index Daten, die nach gleicher Methodik erhoben wurden, gut miteinander vergleichen.

Gewiss, man wird debattieren müssen, inwieweit der moderne Parameter Monatseinkommen vergleichbar ist mit dem in Gräbern repräsentierten Kapital, das u. a. den damals so wichtigen Land- und Viehbesitz nicht einschließt. Doch unsere Skizze sollte nur Wege und tiefergehende Fragestellungen aufzeigen, wie ein recht objektives Ermitteln des Grabwertes möglich ist und wie man über ein einzelnes Gräberfeld hinaus unterschiedliche Populationen vergleichbar machen könnte. Sie zeigt, dass die Kombination aus Grabwerten einerseits und den üblichen Parametern der modernen Sozial- und Wirtschaftswissenschaften andererseits gute Ansatzpunkte für eine tiefergehende Studie bietet.

---

<sup>111</sup> Weitere Vergleichsdaten und Beispiele für die moderne Betrachtungsweise findet man bequem in der „World Inequality Database“: [www.wid.world](http://www.wid.world) [24.1.2020].

<sup>112</sup> Wie Scheidel (2017, 14 f. u. insbes. 445 ff.) darlegt, kann der Gini-Index den höchsten Wert 1 in der Praxis nicht annehmen, da er, sobald die Bevölkerung mehr als 2 Personen umfasst, etwas kleiner ausfallen muss. Nach Scheidel (2017, 446 Fig. A.1) wäre ein Index von ca. 0.85-0.90 bereits ausnehmend hoch.

#### 25.5.4 Literatur zum Kapitel

Christlein, R. (1973). Besitzabstufungen zur Merowingerzeit im Spiegel reicher Grabfunde aus West- und Süddeutschland. *Jahrbuch des Römisch-Germanischen Zentralmuseums* 20, 147–180.

Gebühr, M. & Kunow, J. (1976). Der Urnenfriedhof von Kemnitz, Kr. Potsdam-Land. Untersuchungen zur anthropologischen Bestimmung, Fibeltracht, sozialen Gliederung und „Depot“-sitte. *Zeitschrift f. Archäologie* 10, 185-222.

Hedeager, L. (1980). Besiedlung, soziale Struktur und politische Organisation in der älteren und jüngeren Römischen Kaiserzeit Ostdänemarks. *Præhistorische Zeitschrift* 55, 38-109.

Jørgensen, L. (1987). Family burial practices and inheritance systems. The development of an Iron Age society from 500 BC to 1000 AD on Bornholm. *Acta Archaeologica (København)* 58, 17-53.

Koch, U. (2018). *Mannheim-Sandhofen: Das frühmittelalterliche Gräberfeld von Geroldisheim(?)*. (Mannheimer Geschichtsblätter, Sonderveröffentlichung 12). Mannheim: Verlag Regionalkultur.

Siegmund, F., u.a. (1996). *Das Gräberfeld der jüngeren Römischen Kaiserzeit von Costedt*. (Bodenaltertümer Westfalens, 32). Mainz: von Zabern.

Siegmund, F. (2015). „Jütländische“ Bügelfibeln mit Runenritzungen im südgermanischen Bereich: Zeitstellung und Herkunft der Fibeln, soziale Einordnung der Grabinventare. In: O. Grimm, A. Pesch (Hrsg.), *Archäologie und Runen: Fallstudien zu Inschriften im älteren Futhark. Beiträge zum Workshop am Zentrum für Baltische und Skandinavische Archäologie (ZBSA) in Schleswig am 3.-4. Februar 2011*. (Schriften des Archäologischen Landesmuseums, Ergänzungsreihe Bd. 11) (S. 475-486). Schleswig: Wachholtz.

Steuer, H. (1994). Archäologie und germanische Sozialgeschichte. Forschungstendenzen in den 1990er Jahren. In K. Düwel (Hrsg.), *Runische Schriftkultur in kontinental-skandinavischer und -angelsächsischer Wechselbeziehung. Internationales Symposium in der Werner-Reimers-Stiftung vom 24.–27. Juni 1992 in Bad Homburg*. (Reallexikon der Germanischen Altertumskunde, Ergänzungsband 10). (S. 10–55). Berlin: de Gruyter.

#### Gini-Koeffizient

Atkinson, Anthony B. (2015). *Inequality. What can be done?* Cambridge: Harvard Univ. Press. – deutsch (2016): *Ungleichheit: was wir dagegen tun können*. Stuttgart: Klett-Cotta.

Gini, Corrado (1921). Measurement of Inequality of Incomes. *The Economic Journal*, 31(121), 124–125. <https://doi.org/10.2307/2223319> [24.1.2020].

Juhi (6.3.2019). Gini coefficient and Lorenz curve explained. *Blog Towards Data Science*: <https://towardsdatascience.com/gini-coefficient-and-lorenz-curve-f19bb8f46d66> [24.1.2020].

Mühlenbrock, Hendrik (13.2.2017). Ungleichheit messen – der Gini Koeffizient. *Blog Geschichte als Film (Hypotheses.org)*: <https://gafprojekt.hypotheses.org/453> [24.1.2020].

Scheidel, Walter (2017). *Great Leveler: Violence and the History of Inequality from the Stone Age to the Twenty-First Century*. (The Princeton Economic History of the Western World, 74). Princeton: Princeton Univers. Press. – Hier zitiert nach der Taschenbuchausgabe 2018.

## 25.6 REGRESSION: KÖRPERHÖHENSCHÄTZUNG AUS LANGKNOCHEN

Den Datensatz „Rollet 1888“ hatten wir schon kennen gelernt. Er gilt in Fachkreisen als etwas Besonderes: Die einhundert im Jahr 1888 publizierten Anatomieleichen weisen ein mittleres Lebensalter von 58.1 Jahren auf, die Standardabweichung liegt des Alters bei 17.2 Jahren. Die Mehrheit der Toten ist demnach in den Jahren 1813 bis 1847 geboren: *vor* der Industriellen Revolution und auch *vor* dem massiven Einsatz von Mineraldünger in der Landwirtschaft. Somit bietet der Datensatz einerseits eine von Rollet sorgfältig dargelegte und bereits ziemlich moderne Dokumentationsweise und Messmethode der Knochen, andererseits aber Beobachtungen aus der vorindustriellen Neuzeit in Europa. Alle anderen für dieses Thema hinzugezogenen anthropologischen bzw. forensischen Referenzserien stammen frühestens aus der Mitte des 20. Jahrhunderts oder beruhen auf archäologisch gewonnenen Serien. Bei letzteren beruht die Angabe der „Lebendhöhe“ nie auf Lebendmessungen, sondern auf der Rekonstruktion der Körperhöhe mithilfe der sog. anatomischen Methode (auch „Fully-Methode“) an nahezu vollständig erhaltenen Skeletten, wo neben den Langknochen auch alle anderen körperhöhen-relevanten Skelettelemente wie z.B. der Schädel, alle Wirbel, das Becken und Fußknochen erhalten sind und vermessen werden können (näheres zum Thema: Siegmund 2010). Sollten in diesen Rekonstruktionen systemische Mängel liegen – so beruht die Fully-Methode allein auf männlichen Individuen – übertragen sie sich auch auf die Referenzdaten, weshalb die Rollet-Daten so kostbar sind.

### 25.6.1 Planen vor dem Handeln

*Zielsetzung.* Anhand der Referenzserie Rollet 1888 sollen Formeln zu Schätzung der Körperhöhe entwickelt werden, die auf Skelette anwendbar sind, von denen Langknochen vorliegen, an denen aber mangels vollständiger Erhaltung keine Rekonstruktion der anatomischen Körperhöhe möglich ist. Wiewohl das Verfahren in der Praxis auf alle Langknochen der Beine und Arme anwendbar sein sollte, beschränken wir uns

hier auf die Beinknochen Femur und Tibia, um das Beispiel übersichtlicher zu halten.<sup>113</sup>

*Ansatz:* Wir prüfen, ob eine Lösung mit Hilfe der Regression möglich ist, indem man anhand der Serie Rollet 1888 von den Langknochen auf die dort beobachtete Lebendhöhe schätzt, um im Erfolgsfall die anhand Rollet 1888 entwickelten Parameter auf andere Serien übertragen zu können. Anhand einer gründlichen Sichtung der zum Thema publizierten Literatur haben wir uns vorab davon überzeugt, dass (a) hierzu oft die lineare Regression eingesetzt wird, (b) ältere Studien stets die OLS-Regression verfolgt haben, während neuere Studien zwischen OLS- und RMA-Verfahren schwanken, und (c) ältere Studien in der Regel separate Formeln für Männer und Frauen ermittelt haben, während in jüngeren Studien die Notwendigkeit, geschlechtsdifferenziert zu schätzen, nicht mehr gesehen wird und oft auch übergreifende gemeinsame Formeln für beide Geschlechter angeboten werden. Im Sinne einer Ausgangshypothese soll daher der Weg lineare Regression gemeinsam für Männer und Frauen verfolgt werden.

Zudem haben wir uns anhand der Literatur zu den bestehenden Körperhöhenschätzungen sowie mit Blick auf die verfügbaren Daten in der Archäologie davon überzeugt, dass es wenig Sinn macht, körperseitenspezifische Schätzformeln anzustreben. Daher bilden wir aus einem vorliegenden Wert für die rechte und die linke Körperseite den Mittelwert, resp. nehmen im Falle des Fehlens der einen Information die andere für den Mittelwert und arbeiten im Folgenden mit F1, T1 etc.

### **25.6.2 Verfahren-Auswahl, Prüfen der Voraussetzungen, Vorgehen**

Hinsichtlich der bei diesem Thema viel debattierten Verfahren, welches Regressionsverfahren richtig(er) sei, halten wir uns an die theoretische Literatur:<sup>114</sup> OLS-

---

<sup>113</sup> Ebenso beschließen wir, zur Vereinfachung des Beispiels das komplexe Thema Leichenlänge / Lebendhöhe zu ignorieren. Denn Rollet (1888) gibt die Leichenlänge an, die – je nach Referenz – um ca. 2,0 bis 4,3 cm über der Lebendhöhe liegt (vgl. Trotter & Gleser 1952; zuletzt: Ferorelli, D., Dell’Erba, A. & Solarino, B. (2017). Body length estimation during the post mortem interval: preliminary study. *Romanian Journal of Legal Medicine*, 25, 369-372. DOI: 10.4323/rjlm.2017.369; Cardoso, H. F., Marinho, L. & Albanese, J (2016). The relationship between cadaver, living and forensic stature: A review of current knowledge and a test using a sample of adult Portuguese males. *Forensic Science International*, 258, 55-63. doi: 10.1016/j.forsciint.2015.10.012). Andererseits hat Rollet an den Anatomieleichen „frische“ Knochen vermessen, während der in der Archäologie übliche Fall trockene Knochen sind, die einer leichten Schrumpfung unterlegen sind. Die (wichtige) Debatte um einen Ausgleich dieser Faktoren klammern wir hier aus und nehmen für diese Übung die von Rollet aufgezeichnete „taille“ als Körperhöhe *sensu* Lebendhöhe.

<sup>114</sup> Smith, Richard, J. (2009). Use and misuse of the Reduced Major Axis for line-fitting. *American Journal of Physical Anthropology* 140, 476-486. Auch: Legendre & Legendre 2012, 553 Tab. 10.4.

Regression ist bei der gegebenen Fragestellung die richtige Lösung. Zudem konzentrieren wir uns auf den Fall, der in der Praxis tatsächlich vorkommt: das Schätzen der Körperhöhe (abhängige Variable) aus den Langknochenmaßen (unabhängige Variable), d. h. auf das Gewinnen einer OLS-Regressionsschätzung. Denn die umgekehrte Frage, d. h. die mögliche zweite Lösung, mit der man die Länge eines Knochens aufgrund einer bekannten Körperhöhe schätzen würde, kommt in der Praxis nicht vor.

*Prüfung der Voraussetzungen.* Wir schauen uns die relevanten Variablen mit ihren statistischen Kennzahlen und in Histogrammen an; sie sind unimodal und einigermaßen symmetrisch verteilt. Via Test auf Normalverteilung wird deutlich, dass die Länge des Oberschenkelknochens F1 und des Schienbeins T1 nicht signifikant von einer Normalverteilung abweicht und die Körperhöhe nur wenig. Da die Regression nicht die Normalverteilung der Variablen selbst erwartet, sondern annähernde Normalverteilung im Raum des Zusammenhangs der Variablen, nehmen wir die leichten Abweichungen bei der Körperhöhe als nicht gravierend – behalten aber das Thema „sind die Voraussetzungen erfüllt?“ im Kopf. Zudem vergewissern wir uns mit Sachverstand und Literaturrecherche, dass es keinen überzeugenden Grund gibt, einen nicht-linearen Zusammenhang anzunehmen – also etwa einen kubischen oder logarithmischen Zusammenhang zwischen der Länge der Beinknochen und Körperhöhe – und folglich in den Fundus der nicht-linearen Regressionsverfahren zu wechseln. Vielmehr ist der Ansatz lineare Regression vernünftig und man wird im Laufe des Verfahrens sehen, ob diese Eingangentscheidungen zutreffend waren. Sodann haben wir die Korrelation der Variablen überprüft (Produkt-Moment-Korrelationskoeffizient nach Pearson) und festgestellt, dass der Zusammenhang der Länge der Langknochen mit der Körperhöhe stark ist: Korr.-koeff. nahe oder über .9 und durchweg hochsignifikant.

*Vertiefter Versuchsplan.* Die Güte des Schätzergebnisses einer Regressionsrechnung kann man an den Residuen ablesen sowie am SEE „*Standard Error of Estimation*“, der mit den Berechnungen ausgegeben wird. Diesen wiederum könnte man vermittels Literaturstudien mit den SEE-Werten vergleichen, die zu anderen Schätzformeln für die Körperhöhe angegeben sind, und anhand dieses Vergleichs die Güte der eigenen Ergebnisse einordnen. Rein statistisch wäre dieses Vorgehen korrekt und zudem vergleichsweise unaufwändig, da der SEE von den einschlägigen Programmen ohne weiteres mitausgegeben wird. In der Sache führt dieses statistisch korrekte Vorgehen jedoch nur bedingt weiter. Denn die Regressionsrechnung ist darauf ausgerichtet, für den untersuchten Datensatz eine optimale Lösung zu ermitteln, inklusive eines kleinen SEE-Werts. Dass der SEE relativ klein ausfällt, besagt aber nur, dass die resultierende

Schätzung gut auf die untersuchte Serie passt, es sagt nichts darüber aus, was geschieht, wenn man die ermittelte Formel auf andere Serien anwendet. Es braucht also einerseits eine Referenzserie (hier: Rollet 1888), auf der die Regressionsrechnung und die resultierende Schätzformel beruht, und andererseits eine Testserie, an der man das Ergebnis unabhängig prüfen und das Potenzial der ermittelten Serie quantifizieren kann. Gäbe es passende Daten aus anderen Studien, am besten als Open Data, könnte man eine gänzlich externe Testserie heranziehen. Doch gehen wir zunächst einmal vom häufigen Fall aus, dass solch' eine externe Serie nicht zur Verfügung steht.<sup>115</sup> Also teilt man die eigene Erhebung auf in eine umfangreichere Referenzserie und eine kleinere Testserie. Im gegebenen Fall bietet sich eine 90 % : 10 %-Lösung an: 90 der 100 Rollet-1888-Individuen formen die Referenzserie, 10 von ihnen die Testserie. Freunde des Resamplings und der Bayes-Statistik würden das Problem vermutlich anders lösen, aber auf dem hier verfolgten Weg der klassischen Statistik bleibt keine andere Lösung, als einige Individuen in die Testserie zu geben und damit die quantitative Basis der eigentlichen Regressionsrechnung anhand der verbleibenden Referenzserie zu schmälern.

### 25.6.3 Zwei Zufallsstichproben bilden: Referenzserie und Testserie

Und wie? Indem man eine Zufallsstichprobe aus dem bestehenden Datensatz zieht. BlueSky bietet uns das pfannenfertig mit ein paar Klicks an: => Data; => Split; => For Partitioning; => Random Split; in der dann aufpoppenden Schaltfläche kann man die Namen von zwei neuen Datensätzen angeben sowie den Anteil, wie viele Fälle – zufällig ausgewählt – der jeweilige Datensatz enthalten soll.

Doch wir sind in diesem Band eher mit dem R-Commander unterwegs, wo das gleiche Ziel erreicht werden kann, aber ein paar mehr (lehrreiche) Überlegungen und Schritte nötig sind. Die R-Funktion „sample“ zieht zwar Zufallsstichproben, aber das ist nicht die Aufgabe, denn wir brauchen nicht eine Zufallsstichprobe, sondern das zufällige Teilen eines Datensatzes in zwei dann konstante Stichproben. Daher: Zufallszahlen zu jedem Fall des Ausgangsdatsatzes „Rollet 1888“ hinzufügen und dann anhand dieser Zufallszahlen zwei gezielt große Datensätze schaffen mit eindeutiger Zuweisung

---

<sup>115</sup> Nach dem Durchgang des hier skizzierten soliden Vorgehens werden wir einen (ungewöhnlichen) Weg finden, auch eine externe Testserie zu generieren. Doch der hier entwickelte Weg verläuft entlang gut etablierter Bahnen, der im Falle einer Publikation zweifellos auch die kritischen Augen der Reviewer passieren könnte, während der etwas später eingeführte Weg zu einer (zusätzlichen) externen Testserie noch nicht zum etablierten Methodenkanon gehört, sondern eine innovative Empfehlung des Autors ist.

jedes Falles in einen der beiden Datensätze. Mit den Werkzeugen des R-Commanders nach Laden des Ausgangsdatsatzes circa so:

(a) => Datenmanagement; => Variablen bearbeiten; => Erzeuge neue Variable...; in die Schaltfläche (links) einen Variablennamen eingeben, z.B. „dummy“, und bei „Anweisung für die Berechnung“ folgenden R-Code eingeben: `runif(100, min=0, max=10)`. Übersetzt: erzeuge Zufallszahlen (`r[andom]`), die gleichverteilt sind (`unif[orm]`), und zwar 100 mal, mit dem Minimum 0 und dem Maximum 10;<sup>116</sup> mit „OK“ ausführen. Wenn man mit „Datenmatrix betrachten“ in die Daten hineinschaut, ist am Ende jeden Falles „dummy“ eingefügt und hat unterschiedliche Werte – taucht aber möglicherweise für Statistiken oder Grafiken noch nicht in der Variablenliste auf. Daher: =>Datenmanagement; => Aktive Datenmatrix; => Aktualisiere aktive Datenmatrix; OK. Jetzt kann man sich „dummy“ auch in seinen Kennzahlen anschauen und z.B. als Histogramm.

(b) Ausgangsdatsatz geschickt sortieren. => Datenmanagement; => Aktive Datenmatrix; => Sortiere aktive Datenmatrix, im neuen Fenster „dummy“ und „sex“ auswählen.

Tipp: Die Strg-Taste (alias Control-Taste) nutzen, um einzelne Fälle herauszupicken. Mit der Shift-Taste (Umschalten Klein-Gross-Schreibung) hingegen lassen sich ganze Bereiche auswählen.

Nach Auswahl und Drücken von OK öffnet sich ein neues Fenster, das fragt, in welcher Reihenfolge die beiden Variablen als Sortierkriterien dienen sollen. Wir setzen „Sex“ auf 1 und „dummy“ auf 2, und bestätigen mit OK. Das Ergebnis mit „Datensatz bearbeiten“ anschauen: der Datensatz enthält erst 50 Frauen, dann 50 Männer, beide Kolonnen sind jeweils per Zufallszahl „dummy“ sortiert, vorne (zur Kontrolle) sind die ursprünglichen Fallnummern („rownames“) nicht mehr wie im Original von 1 bis 100 geordnet.

(c) 5 Frauen und 5 Männer händisch markieren – und zwar die jeweils 5 Fälle, die am Anfang der Liste stehen. Also in die erste Zeile gehen, zur Variable „dummy“, und dort händisch bei den obersten 5 Fällen eine Zahl größer 10, z.B. 20 eingeben. Das Gleiche bei den ersten 5 Männern (Datenzeile 50-54) wiederholen. Mit OK bestätigen. Sich selbst prüfen, z.B. indem man ein Histogramm zu „dummy“ abrufen, wo jetzt eine weit vom Rest entfernte Säule erkennbar sein sollte mit 10 Fällen. Ergebnis:

---

<sup>116</sup> Konkrete Parameter frei erfunden, geht selbstredend auch mit anderen Zahlen. Man könnte statt der Gleichverteilung auch eine andere Verteilung wählen wie z. B. eine Normalverteilung.

es sind jetzt 5 zufällig ausgewählte Frauen und 5 zufällig ausgewählte Männer bei „dummy“ mit dem Wert „20“ belegt, alle übrigen Fälle haben Werte von 0 bis 10.

(d) Datensatz in die beiden gewünschten Datensätze teilen. => Datenmanagement; => Aktive Datenmatrix; => Teilmenge der aktiven Datenmatrix; die beiden Fenster ausfüllen; bei „Anweisung für die Teilmenge“ schreibe man „dummy <15“, als „Name für die neue Datenmatrix“ trage man ein „Referenzserie“; mit OK bestätigen.

Sodann wählen wir als aktive Datenmatrix wieder „Rollet1888“ aus und wiederholen den Vorgang, indem wir mit „dummy >15“ und „Testserie“ die 10 Fälle der Testserie rausschreiben.

Abschließend für beide Serien prüfen, ob das Gewünschte entstanden ist, z.B. indem man einfache Statistiken aufruft und die Zahl der Fälle prüft – es müssten bei Referenzserie je 45 Männer und Frauen sein, bei Testserie je 5 Männer und Frauen.

Geschafft!

Jedenfalls fast: Da wir jetzt viel Arbeit in die Differenzierung in eine Referenzserie und eine Testserie investiert haben und ggf. auch morgen mit den Serien weiterarbeiten wollen anstatt das Stichprobenziehen (mit anderem Resultat) zu wiederholen, empfiehlt es sich, die Daten zu sichern. Am einfachsten den ganzen aktuellen Workspace, mit => Datei; =>Workspace speichern unter. So kann man am nächsten Arbeitstag seine Stichproben wieder schnell laden und weiterarbeiten.

#### 25.6.4 Lineare Regression anhand der Referenzserie

Nach diesen Vorbereitungen ist das Berechnen der Regressionsgleichungen nicht weiter kompliziert. Im R-Commander liegt das Gesuchte hinter => Statistik; => Modelle anpassen; => Lineare Regression. „Khoehc“ ist die abhängige Variable, „F1“ die unabhängige Variable. Da insgesamt drei verschiedene Modelle berechnet werden sollen, setzen wir oben auch einen Namen für das erste Modell, z.B. „regF1“, mit OK bestätigen. R gibt aus:<sup>117</sup>

```
Residuals:
  Min       1Q   Median       3Q      Max
-10.5894  -2.4769  -0.1354   2.2603  10.6078
```

---

<sup>117</sup> Es ist empfehlenswert, die Übung praktisch nachzuarbeiten. Da das Ziehen der Zufallsstichproben „Referenzserie“ und „Testserie“ zu jeweils unterschiedlichen Serien führt, werden sich beim Nacharbeiten jeweils leicht unterschiedliche Werte ergeben – wobei das Ergebnis grundsätzlich ähnlich sein sollte.

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	51.68190	5.96276	8.667	1.98e-13 ***
F1	0.25135	0.01377	18.259	< 2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.894 on 88 degrees of freedom  
Multiple R-squared: 0.7912, Adjusted R-squared: 0.7888  
F-statistic: 333.4 on 1 and 88 DF, p-value: < 2.2e-16

In der Mitte der Ausgabe ist unter „Estimate“ die Regressionsgerade angegeben mit der Steigung 0.25135 und der Konstante 51.68190. Sie lautet also:

$$\text{geschätzte Körperhöhe} = F1 * 0.25135 + 51.6810.$$

Am Ende der betreffenden Zeile sind für beide Parameter die p-Werte ausgegeben; sie sind sehr klein und hochsignifikant. In den Zeilen darüber stehen die Angaben zu den „Residuen“, d.h. das Ergebnis beobachtete Körperhöhe minus (per Regression) geschätzte Körperhöhe. Die Quartile liegen bei ca. -2.5 bis +2.3, d.h. 50 % der Schätzungen liegen in einer Spanne von ca. plus/minus 2.5 cm. Die maximalen Schätzfehler sind -10.6 resp. +10.6 cm. Die wichtigeren Parameter finden sich im unteren Block, die Kenngröße  $R^2$  beträgt 0.79, d.h. die Schätzung erklärt fast 80% der Varianz. In der Zeile darunter findet sich die zugehörige Teststatistik, wonach das Ergebnis (p-Wert) hochsignifikant ist.

Wir fügen einige Ergebnisse dem aktiven Datensatz hinzu: => Modelle; => Füge Regressionsstatistiken den Daten hinzu...; man wähle „Vorhergesagte Werte“ und „Residuen“. Das ermöglicht uns, Statistiken und Grafiken für die Residuen ("residuals.regF1") zu analysieren. Nach den Kennzahlen liegt deren Standardabweichung bei 3.87 (cm), alle Parameter passen zur Annahme einer Normalverteilung. Ein Histogramm bestätigt diese Eindrücke.

**R** gibt auf Anforderung weitere wertvolle diagnostische Graphiken ohne größeren Aufwand aus: => Modelle; => Grafiken; => grundlegende diagnostische Grafiken. Der Plot „Normal Q-Q“ zeigt, dass die Beobachtungen (Residuen) nahe der erwarteten Ideallinie liegen, am oberen Ende und am unteren Ende halten sich Abweichungen in Grenzen. Der Plot „Residuals vs. Fitted“ zeigt eine weitgehend ungeordnete Streuung der Residuen entlang des gesamten Wertebereichs. All dies zeigt, dass eine lineare Regression die passende Modellwahl war. Würde das lineare Modell nicht passen, würden die Residuen am einen Ende der Achse ganz anders sein als beim Gegenüber, oder einer Kurve folgen: all das ist nicht der Fall. Der zusätzlich zu den

Grafiken angebotenen „Numerische(n) Diagnose“ bedarf es hier nicht, sie zielt auf multiple lineare Regressionen (die wir später durchführen werden).

Bilanz: nach allen Bildern und Parametern wurde die lineare Regression richtig angewandt und führt zu Schätzungen mit einem vertretbaren Schätzfehler.

An dieser Stelle würde man jetzt im Falle realer Forschung die Berechnungen für T1 anschließen sowie eine multiple Regression an F1 und T1. Ersteres führen wir hier nicht aus, um den Text schlank zu halten – doch für Sie wäre es ein schöner Übungsfall. Letzteres behandeln wir unten separat im Kapitel „Fallstudie Multiple Lineare Regression“, damit wir im Folgenden einen Fall – den allein mit F1 – ungestört und konsequent durchdenken können.

### 25.6.5 Anwendung des Regressions-Ergebnisses auf die Testserie

Um im Sinne des traditionellen Verfahrens fortzufahren, prüfen wir nun die Ergebnisse der linearen Regression an der Testserie. Dazu errechnet man anhand der oben ermittelten Regressionsgerade für die „Testserie“ (a) die geschätzte Körperhöhe und (b) das Residuum, also die Differenz (beobachtete) Körperhöhe minus Schätzung. Anschließend betrachten wir die beobachtete Körperhöhe, die geschätzte Körperhöhe und die Residuen in ihren statistischen Kennzahlen:

	mean	sd	skewness	kurtosis
Khoehe	159.7000000	6.880730	0.2723339	-0.7483535
estF1	159.8243375	6.817732	-0.1889114	-0.5279279
resF1	-0.1243375	3.707157	-1.7271662	4.4797372

	25%	50%	75%	n
Khoehe	155.2500	158.5000000	165.500000	10
estF1	155.6771	159.0702875	164.882756	10
resF1	-0.7885	0.1918625	1.402712	10

Man erkennt, dass die Werte der beobachteten Körperhöhe nahe an den Kennzahlen der geschätzten Körperhöhe liegen. Die Residuen haben einen Mittelwert und Median sehr nahe Null und eine Standardabweichung von 3.7 (cm), was der Standardabweichung bei der Referenzserie entspricht. Die Anwendung der Schätzformel auf die externe Testserie bietet eine wichtige Einschätzung ihrer Güte – in unserem Fall eine Bestätigung ihrer Güte.

### 25.6.6 Regression: Ausreißer, Gütekriterien, Optimierungen

An diesem Punkt ist es lohnend, auf einige bislang nicht besprochene statistische Parameter der Güteabschätzung einzugehen. Das Thema „Residuen“ hatten wir schon

besprochen: die Differenz zwischen tatsächlichem Wert und dem via Regression geschätzten Wert. Man nennt diesen Wert genauer „unstandardisierte Residuen“. Bequemer als das Selbst-Rechnen ist die Nutzung des R-Commanders, der dies mit einer einfachen Befehlsfolge für uns übernimmt. Zunächst muss oben in der Zeile, in der links die aktive Datenmatrix angezeigt wird, außen rechts in der Schaltfläche „Modell“ das (oder ein) aktive Regressionsmodell ausgewählt sein oder werden. Die nächsten Schritte sind: => Modelle; => Füge Regressionsstatistiken zu den Daten hinzu ... In dem aufklappenden Menü setze man ein Häkchen an alle der aufgeführten Optionen, dann OK & Enter. Die neu hinzugefügten Werte kann man z.B. mit „Datenmatrix betrachten“ anschauen. Nehmen wir an, das aktive Regressionsmodell trägt den Namen „regF1“.

Die Spalte „fitted.regF1“ enthält nun den jeweiligen per Regression geschätzten Wert. Die Spalte „residuals.regF1“ zeigt das unstandardisierte Residuum, d. h. das Ergebnis der Rechnung: beobachteter Wert (i. e. Höhe) minus geschätzter Wert (i. e. fitted.RegrF1). Über => Statistik; => Deskriptive Statistik; „Zusammenfassungen numerischer Variablen“ und die Auswahl von „residuals.regF1“ kann man sich die Statistik dieser Größe nochmals vergegenwärtigen: der Mittelwert liegt ganz nahe Null, die Standardabweichung bei 3,7 cm. Über => Grafiken; => Histogramm, usw. kann man sich ein Bild von der Verteilung der Residuen machen: im vorliegenden Fall unimodal, symmetrisch und den Bereich -7,7 resp. +7,5 cm nur in Ausnahmefällen überschreitend.

Zusätzlich bestellt hatten wir „studentisierte Residuen“, in unserem Beispiel in der Matrix nun als „rstudent.regF1“ aufgeführt. Hier sind die Residuen („residuals.regF1“), welche noch die originale Skala der betrachteten Größe widerspiegeln, auf den Mittelwert Null und die Standardabweichung 1 standardisiert. Bei einer gelungenen linearen Regression sollten die Residuen annähernd normalverteilt sein, so dass nach dieser Normierung für die studentisierten Residuen entsprechend den Eigenschaften einer Normalverteilung gelten müsste: praktisch keine Fälle im Datensatz über- bzw. unterschreiten den Wert 3.29, und nur max. 1% der Fälle über- oder unterschreiten den Wert 2.58, während 95 % aller Beobachtungen im Bereich -1.96 und +1.96 liegen sollten. Man prüfe dies mit „Statistik“ und „Grafik“... und ggf. auch durch Sortieren der Daten und Anschauen der Datenmatrix. In unserem Fall überschreiten nur drei Fälle den Wert 1.96, und nur 1 Fall davon den Wert 2.58 – d.h. die Erwartungen der Theorie an eine gelungene lineare Regression sind bestens eingehalten.

Die „Cook-Distanz“ (hier: „cooks.distance.regF1“) und der „hat value“ (auch „leverage“ genannt) (hier: „hatvalues.regF1“) sind Maßzahlen zur Beschreibung, wie einflussreich ein einzelner Fall auf das Gesamtergebnis der Regressionsrechnung ist. Hat Values liegen zwischen Null (kein Einfluss) und 1 (diese eine Beobachtung bestimmt die gesamte Regressionsrechnung). Bleiben wir beim Hat Value: als üblicher Vergleichswert wird angesetzt:  $(k+1)/n$  – wobei „k“ die Anzahl der mit der Regression geschätzten Parameter ist und „n“ die Anzahl der Fälle. Da hier in einer linearen Regression die Steigung und die Konstante geschätzt wurde, ergibt sich  $(2+1)/100$ , d.h. 0.03. Je nach Literaturreferenz werden Fälle, deren Hat Value doppelt so groß oder dreimal so groß ist wie dieser Wert als auffällig betrachtet, d. h. gelten als besonders einflussreich auf das Gesamtergebnis. Bei der Cook-Distanz gelten Fälle mit Werten von größer 10 als auffallend einflussreich auf das Gesamtergebnis. In unserem Fall liegen die Cook-Distanzen und die Hat Values unter den genannten Grenzwerten.

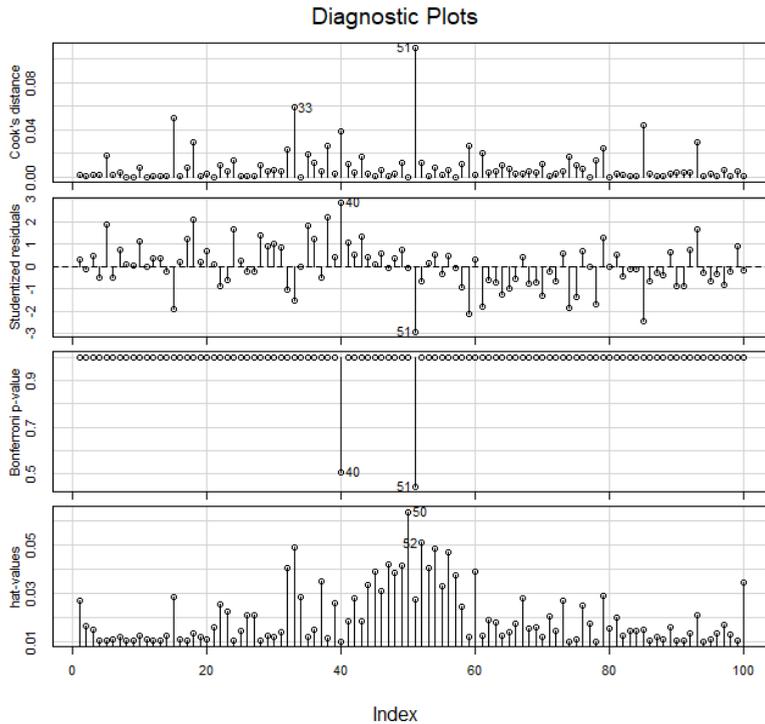
Caveat: einflussreich muss nicht heißen falsch! Gleichwie wird hier versucht zu prüfen, ob es in dem Datensatz einzelne Fälle – sog. „Ausreißer“ – gibt, die das Ergebnis sehr gravierend beeinflussen. So gravierend, dass es geboten wäre, sie aus dem Datensatz zu entfernen, um die Regressionsrechnung an dem um völlig untypische Fälle bereinigten Datensatz zu wiederholen, was ggf. treffendere Parameterschätzungen ergeben könnte. Doch Vorsicht: zwar ist ein solches Aussondern extrem ungewöhnlicher Fälle üblich und berechtigt, doch ich empfehle, dies nicht wiederholt zu tun, d. h. einer ersten Bereinigung weitere Bereinigungen folgen zu lassen, denn – wie einschlägigen Erfahrungen und auch Kennzahlen zeigen – die Ergebnisse werden am Ende nicht mehr wesentlich besser, jedoch die Referenzserie kleiner.

Was wir hier etwas händisch prüfen, kann eleganter und schneller (aber weniger anschaulich) mit dem sog. Bonferroni-Ausreißer-Test geprüft werden: => Modelle; => Numerische Diagnose; => Bonferroni-Ausreißer-Test. Ergebnis:

```
outlierTest(RegrF1)
No Studentized residuals with Bonferroni p < 0.05
Largest |rstudent|:
      rstudent unadjusted p-value Bonferroni p
51 -2.911925          0.0044572      0.44572
```

Der Test protokolliert (2. Zeile), dass es keine signifikanten Ausreißer gibt und führt unten Fall Nr. 51 mit einem studentisierten Residuum von -2.911925 als den Fall an, der am meisten als Ausreißer in Frage kommt. Der p-Wert des Tests für Fall Nr. 51 beträgt indes 0.44572 und ist weit von einer Signifikanz von unter 0.05 entfernt.

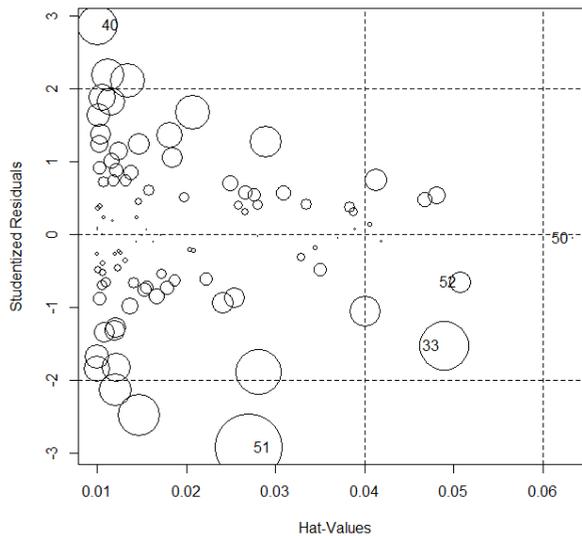
Die grafische Darstellung der beschriebenen Phänomene und Zahlen erhält man mit: => Modelle; => Grafiken; => Influence Index Plot; die voreingestellten vier Kennzahlen akzeptieren (oder Häkchen dran setzen), OK & Enter. Der Plot **Abb. 115** zeigt alle 100 in die Rechnung eingegangenen Fälle und hat auffallende Fälle automatisch markiert, insbes. die bereits besprochenen Fälle Nr. 51 und 40. Sie fallen auf, doch bleiben sie, wie zuvor besprochen, noch in den Grenzen des Üblichen.



**Abb. 115** "Diagnostic Plot" zu den Parametern, die den Einfluss einzelner Fälle auf das Regressionsergebnis schätzen.

Ergänzend ist sehr nützlich die „Einflussgrafik“: => Modelle; => Grafiken; => Einflussgrafik, Voreinstellungen belassen, OK & Enter. Das Streuungsdiagramm **Abb. 116** trägt die studentisierten Residuen gegen die Hat Values auf: Fälle mit hohem Einfluss auf das Gesamtergebnis haben höhere Hat Values. Beruhigenderweise erkennen wir hier, dass die Fälle mit relativ hohen studentisierten Residuen – also Werten größer

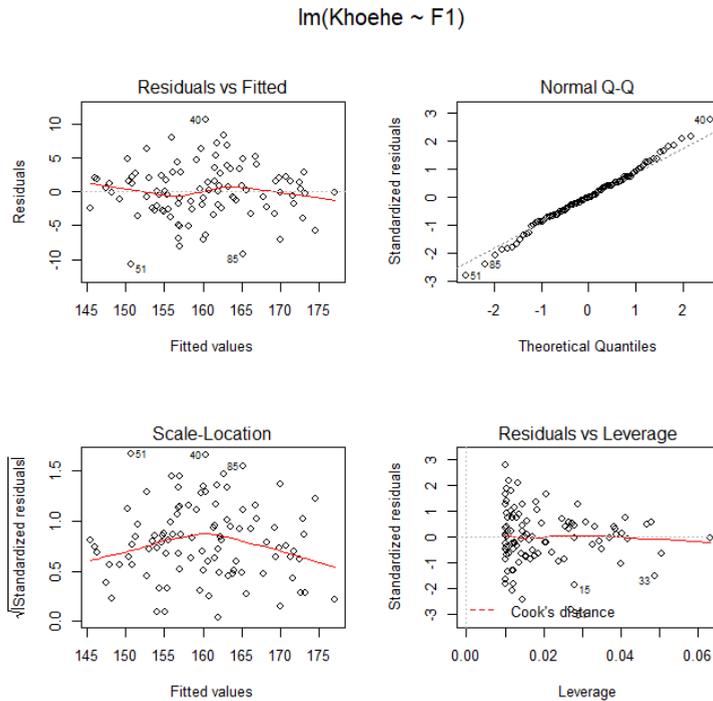
$\pm 2.58$  und bereits nahe der Grenze  $\pm 3.29$  – wie z. B. Fall Nr. 40 und 51 einen relativ geringen Hat Value aufweisen. Ihre Abweichung ist zwar relativ hoch, aber ihr Einfluss auf das Regressionsergebnis eher gering. Läge solch ein Punkt mit hohem Residuum weit rechts bei den hohen Hat Values, wäre es drängender, zu prüfen, ob er nicht aus der Referenzserie herausgenommen werden sollte. Daher wurden hier automatisch auch die Fälle 33 und 52 markiert mit relativ geringen Residuen, doch größerem Einfluss auf das Gesamtergebnis.



**Abb. 116** "Einfluss-Grafik". Sie setzt die Residuen (Schätzfehler) in Beziehung zum Hat Value, dem Einfluss eines Falls auf das Gesamtergebnis der Regression und markiert einzelne Fälle mit auffallenden Werten.

Diese geschilderte Staffel an Kriterien dient dazu, ggf. einzelne auffallende Fälle in der Referenzserie zu identifizieren und eine Prüfung nahezu legen. Mit anderen Grafiken werden – ergänzend zu den Maßzahlen – zumeist die Angemessenheit der Regressionsrechnung insgesamt bewertet. Die dazu üblicherweise hinzugezogenen Bilder sind via R-Commander fertig zusammengestellt: => Modelle; => Grafiken; => Grundlegende diagnostische Grafiken. In unserem Fall ergeben sich die erwarteten Bilder eines gelungenen Vorgehens (**Abb. 117**): Im Plot „Normal Q-Q“ liegen die Punkte dicht an der gestrichelten Geraden und weichen auch an den Enden nur relativ wenig davon

ab. Das Streuungsdiagramm „Residuals/Fitted“ zeigt eine unstrukturierte Wolke, d.h. die Residuen (Schätzfehler) sind unabhängig. Ähnlich das Bild mit den „standardized residuals“. Das Streuungsdiagramm Standardized Residuals vs. Leverage (alias Hat Value) zeigt eine breite Streuung der Residuen bei den Fällen geringen Einflusses, während die Fälle höheren Einflusses um die Null-Achse streuen.



**Abb. 117** "Grundlegende diagnostische Grafiken" für unser Regressionsbeispiel der Schätzung der Körperhöhe aus der Femurlänge (F1).

Man kann diese (und ähnliche) Bilder weitaus besser einschätzen, wenn man vergleichbare Bilder kennt, die auf ungeeigneten Daten beruhen. Ein Beispiel, das jeder selbst einmal durchspielen kann: man kreiere eine künstliche Variable, die stark nicht-linear ist. In unserem Datensatz „Rolle1888“ könnte man dies z. B. wie folgt tun:

```
Rolle1888$Kubik <- with(Rolle1888, ((Nr * Nr * Nr) + rnorm(100, mean=0,
sd=1000)))
```

Damit errechnen wir, inhaltlich sinnfrei, die Kunstvariable „Kubik“ aus der dritten Potenz der Fallnummer und addieren, um ein Mindestmaß an Streuung zu generieren, eine normalverteilte Zufallszahl mit dem Mittelwert Null und der Standardabweichung 1000 hinzu. Man schaue sich das Resultat in einem Streuungsdiagramm „Nr“ vs „Kubik“ an: es besteht ein sehr starker, jedoch deutlich nicht-linearer Zusammenhang zwischen beiden Variablen. Schätzt man nun aus „Nr“ (unabhängig) auf „Kubik“ (abhängige Variable), ergibt eine lineare Regression zunächst einmal signifikante Parameter und einen hohen  $R^2$ -Wert. Doch wie anders sehen nun die „Grundlegenden diagnostischen Grafiken“ aus! Der Normal-Q-Q-Plot zeigt an beiden Enden sehr starke Abweichungen von der Geraden; die Streuungsdiagramme Residuen vs. Fitted zeigen statt einer diffusen Wolke eine Parabel bzw. eine W-förmige Kurve, d. h. die Residuen (Schätzfehler) sind keinesfalls zufällig und unabhängig, sondern folgen einem System; bei der Cook-Distanz fallen hohe Residuenwerte mit starkem Einfluss auf das Ergebnis zusammen. In Summe Bilder, die (richtigerweise) stark vom Soll-Zustand für lineare Regressionen abweichen. Zugleich ein gutes Beispiel, wie man in solchen Fällen mit der Deutung von gegebenen Zahlen und Bildern weiterkommen kann: indem man künstliche Werte mit bekannten, ideal passenden und alternativ ideal unpassenden Werten in die Verfahren eingibt, um Ergebnisse vergleichen zu können.

Doch trotz aller Test und Grafiken: auf diesen Wegen prüft man vor allem, ob die Voraussetzungen für die Anwendung einer linearen Regression vorliegen und ob das Ergebnis den Ausgangsdatensatz (Referenzserie) gut beschreibt. Die Frage nach Ausreißern ermöglicht es, den Ausgangsdatensatz – so nötig – zu bereinigen und damit die aus der Regression hervorgehenden Schätzer noch zu optimieren. Ob diese Schätzer jedoch auch erfolgreich auf andere Datensätze anwendbar sind, d.h. inwieweit die Referenzserie repräsentativ ist, lässt sich auf diesem Wege nicht ermitteln.

### 25.6.7 Lineare Regression mit Resampling

Ein Weg, die Stabilität einer Regression zu prüfen, d. h. die Nicht-Zufälligkeit eines Ergebnisses, ist das Resampling: Wenn man aus der gegebenen Referenzserie immer wieder Stichproben zieht, kann jeweils neu eine Regression gerechnet und deren Parameter ermittelt werden. Wiederholt man diesen Vorgang sehr oft („Bootstrapping“), lassen sich die Ergebnisse protokollieren und wiederum statistisch darstellen. Keine Sorge, diese Arbeit übernimmt **R** für uns. In diesem Fall kehren wir zur Ausgangsstichprobe zurück, d. h. zum Datensatz „Rollet 1888“ mit 100 Fälle, und berechnen per Bootstrapping die Konfidenzintervalle der Parameter.

(a) Erst wieder die lineare Regression: => Statistik; => Modelle anpassen; => Lineare Regression; abhängige Variable „Khoehe“, Unabhängige Variable „F1“, OK.

(b) Nun – zum Vergleich – die konventionellen Konfidenzintervalle der Parameter: => Modelle“; => Konfidenzintervalle; bei der Voreinstellung 0.95 = 95 %-Konfidenzintervalle belassen; OK. Wir erhalten:

```
> Confint(RegModel.2, level=0.95)
      Estimate      2.5 %      97.5 %
(Intercept) 52.8076803 41.5527726 64.0625881
F1          0.2487176  0.2227216  0.2747136
```

...und sehen, dass sowohl die Konstante (jetzt ca. 52.8) mit ca. 41.6 bis 64.1 (cm!) eine recht weite Spanne aufweist, als auch die Steigung (alias Faktor) zu F1, die zwischen ca. 0.22 und 0.27 liegt.

(c) Nun mit Bootstrapping: => Modelle; => Bootstrap Konfidenzintervalle; in der Schaltfläche setzen wir die vorgeschlagene „Anzahl der Bootstrap Stichproben“ von 999 auf 5000 hoch und bestätigen mit OK. Wir üben uns jetzt in etwas Geduld, denn es werden tatsächlich hintereinander 5000 Stichproben mit Zurücklegen gezogen, jeweils die lineare Regression berechnet, die Ergebnisse protokolliert und anschließend verglichen. Das kann schon mal ein Minütchen oder auch zwei dauern. Wir erhalten:

```
Bootstrap bca confidence intervals
      2.5 %      97.5 %
(Intercept) 43.0832124 61.1910639
F1          0.2289365  0.2705948
```

Die mittlere Schätzung (oben) der Konstante lag also bei 52.8, beim Bootstrapping (unten) lagen 95 % der Resultate zwischen ca. 43.1 und 61.2. Für die Steigung ergab sich (oben) eine mittlere Schätzung von ca. 0.249, das 95 %-Konfidenzintervall (unten) beim Stichprobenziehen liegt zwischen 0.229 und 0.271. Da wir an den diesbezüglichen Voreinstellungen nichts geändert haben, hat **R** uns auch eine Grafik zum Bootstrapping ausgegeben, nämlich die „Bootstrap Distributions“ - eine Dichtekurve für die 5.000 Schätzungen für die beiden Parameter. Beide Kurven sind symmetrisch und sehr steil, d.h. stark auf den mittleren Wert hin zentriert: unsere Regressionsschätzungen sind sehr stabil.

Im Grunde genommen war unser konventioneller Ansatz, mit dem wir eingestiegen waren, eine einzige dieser Stichproben, und hatte die Formel ergeben:

$$\text{geschätzte Körperhöhe} = F1 * 0.25135 + 51.6810.$$

Wir sehen, dass diese ziemlich mittig in dem Intervall liegt.

Was ist das Ergebnis? Das Bootstrapping bestätigt, dass die Schätzungen recht stabil sind. Zudem haben wir bei der Anwendung der Formeln, die sich aus der Referenzserie ergaben, auf die Testserie gesehen, dass tatsächlich Schätzungen der vorhergesagten Richtigkeit erfolgt sind, also keine unvorhergesehen größeren Fehler entstanden. Mein persönlicher Weg wäre: einerseits das zu versuchen, was im nachfolgenden Abschnitt beschrieben wird, und andererseits zurück auf Anfang, zum Ausgangsdatensatz „Rollet 1888“: Lineare Regression auf Basis aller 100 Fälle rechnen; dann => Modelle; => Füge Regressionsstatistiken zu den Daten hinzu; „Vorhergesagte Werte“ und „Residuen“ auswählen, mit OK bestätigen. Dann => Modelle; => Graphiken; „Grundlegende diagnostische Graphiken; & OK. Dort erkennen wir, dass die Fälle Nr. 40 sowie 51 und 85 als ungewöhnlich markiert werden. Im Datensatz Rollet 1888 („Datenmatrix betrachten“), dass es die drei Fälle mit den höchsten Residuen sind (bei +10, resp. -10). Aber: wir hatten uns anhand der Kennzahlen Cook-Distanz und Hat Value davon überzeugt, dass das Maß ihre Ungewöhnlichkeit noch ihm Rahmen des Üblichen liegt, weshalb wir darauf verzichten, sie aus dem Datensatz der Referenzserie zu entfernen.

Um die zuvor erarbeiteten Parameter als Ergebnis und neue Schätzformel publizieren, aber realistische Schätzfehler angeben zu können, berechnen wir erneut Bootstrap-Intervalle – aber nicht die Standardeinstellung von 95 %, sondern wählen stattdessen den Wert 68.27 %, was der Angabe des 1-sigma-Bereichs bei Normalverteilungen entspricht, so dass der Fehlerbereich ca. 2/3 aller Beobachtungen umfasst:

```
Bootstrap bca confidence intervals
                15.9 %      84.1 %
(Intercept) 50.1700962 58.3795201
F1           0.2358396  0.2549801
```

Im Ergebnis würde man nun statt traditionell einer Schätzformel samt vermutlichem  $\pm$ Schätzfehler nun drei Schätzformeln (anhand F1) publizieren: abgeleitet aus der Linearen Regression eine Formel für den am meisten wahrscheinlichen Schätzwert, und abgeleitet aus dem oberen und unteren Konfidenzintervall des Bootstrappings je eine weitere Schätzformel mit der Aussage: „...und Individuum ist mit hoher Sicherheit größer als ... cm und kleiner als ... cm.“

$$\begin{aligned}\text{wahrscheinlicher Schätzwert (in cm)} &= 54.09 + F1 * 0.24597 (\pm 3.47 \text{ cm}) \\ \text{sehr wahrscheinlich größer als} &= 50.17 + F1 * 0.23584 \\ \text{sehr wahrscheinlich kleiner als} &= 58.38 + F1 * 0.25498\end{aligned}$$

### 25.6.8 Prüfung anhand externer Testserie

Das in der tatsächlichen Forschung bei Regressionsverfahren angewandte Verfahren, die Ausgangsdaten in eine Referenzserie und eine Testserie zu trennen, hat in unserem Fall letztlich keinen wesentlichen Erkenntnisgewinn erbracht; das Ergebnis liegt deutlich innerhalb des Bootstrap-Konfidenzintervalls und auch des Standardschätzfehlers der klassischen Regression. Das liegt letztlich daran, dass im gegebenen Fall (a) die lineare Regression gut verläuft, d. h. der Datensatz für die Fragestellung passend und gut ist, und (b) der Datensatz insgesamt eher klein ist, so dass am Ende eine empfindlich verkleinerte Referenzserie genutzt wird, während andererseits nur eine vergleichsweise kleine (zu kleine) Testserie zur Verfügung steht. Das kann unter anderen Konstellationen anders sein, weshalb ich den oben dargelegten Verfahrensansatz unter bestimmten Umständen weiterhin für sehr transparent und richtig halte.

Im gegebenen Fall ist jedoch noch ein zusätzlicher Umstand zu bedenken, der nicht statistischer, sondern sachlicher Natur ist. Eine Testserie zur Abschätzung und Darlegung der Leistungsfähigkeit einer Regressionsgleichung sollte eine externe Testserie sein, d. h. wirklich aus einer anderen Population stammen. Das ist im hier verfolgten Fall nicht gegeben, denn die Serie „Rollet 1888“ ist eine eng zusammengehörende, wie beschrieben spezifische Population. Sie ist im Grunde keine „externe“ Serie. Daher ist zu prüfen, ob eine solche anderweitig verfügbar ist.

Für unsere Fragestellung ist dies erfreulicherweise der Fall. Der US-amerikanische Anthropologe Christopher B. Ruff, der sich vielfach mit der Schätzung der Körperhöhe wie auch des Körpergewichts aus Knochenmaßen beschäftigt hat, hat 2012 zusammen mit einer internationalen Forschergruppe neue Schätzformeln für die Körperhöhe publiziert.<sup>118</sup> Deren Grundlage ist eine umfangreiche Sammlung von prähistorischen bis neuzeitlichen Skeletten, bei denen zwar die Lebendhöhe nicht bekannt ist, aber mit Hilfe der sog. anatomischen Methode (auch: Fully Methode) sehr zuverlässig rekonstruiert werden kann. Anlässlich der Veröffentlichung einer umfassenden

---

<sup>118</sup> Ruff, Chr. B., Holt, B. M., Niskanen, M., Sladek, V., Berner, M., Garofalo, E., Garvin, H. M., Hora, M., Majjanen, H., Niinimäki, S., Salo, K., Schuplerova, E. & Tompkins, D. (2012). Stature and body mass estimation from skeletal remains in the European Holocene. *American Journal of Physical Anthropology*, 148, 601-617.

Synthese zu diesem Thema, die auch die zeitliche wie räumliche Variabilität der Körperhöhe und Skelettproportionen im prähistorischen bis neuzeitlichen Europa darlegen will,<sup>119</sup> hat Ruff auch den zugrundeliegenden Datensatz auf seiner Website publiziert.<sup>120</sup> Interessierte können sich den Datensatz (\*.xlsx) dort herunterladen, in den R-Commander laden und die folgende Übung selbst nachvollziehen.

Es ist hier nicht nötig, den sehr variablenreichen Datensatz im Einzelnen nachzuvollziehen. Für unser kleines Experiment reicht es zu wissen, dass „STATURE“ die anatomische Körperhöhe meint und „FMAXLN“ dem Maß F1 entspricht, dass wir hier verwendet haben. Wie eine schnell abgerufene Statistik zeigt, enthält der Datensatz 2177 Fälle, zu 1130 von ihnen ist die anatomische Körperhöhe bekannt (Mittelwert 162.2 cm, Std.abw. 8.2 cm), zu 2050 von ihnen auch die maximale Femurlänge (Mittelwert 438.1 mm, Std.abw. 29.5 mm). Wir wollen jetzt die anhand der Serie Rollet 1888 erarbeitete Regressionsgleichung auf diese Serie anwenden. Dazu rechnen wir via => Datenmanagement; => Variable bearbeiten; => Erzeuge neue Variable, ..., wie oben anhand von Rollet 1888 ermittelt:

```
estStature = FMAXLN * 0.25135 + 51.6810
```

und dann

```
delta = estStature - STATURE
```

... womit wir die Differenz zwischen unserer Rollet-1888-Schätzung und der tatsächlichen Körperhöhe in diesem Datensatz ermittelt haben. Wenn wir eine Statistik für diese Werte abrufen, erhalten wir:

	mean	sd	n	NA
delta	-0.3425269	2.544696	2050	127
estStature	161.7959024	7.425504	2050	127

Der Mittelwert des Unterschieds liegt bei -0.3 cm, die Standardabweichung der Differenz bei 2.5 cm: unsere Regression nach Rollet 1888 unterschätzt die Population Ruff et al. 2018 im Mittel um etwa dreieinhalb Millimeter kleiner, bei einem mittleren Schätzfehler von 2,5 cm. Ein zur Kontrolle angefertigtes Streudiagramm STATURE vs. estStature unterstreicht die enge Übereinstimmung beider Größen.

---

<sup>119</sup> Ruff, Chr. B. (ed.) (2018). *Skeletal variation and adaptation in Europeans: Upper Paleolithic to the Twentieth Century*. Hoboken: Wiley Blackwell.

<sup>120</sup> John Hopkins School of Medicine, Center for Functional Anatomy and Evolution, Dr. Christopher B. Ruff: <https://www.hopkinsmedicine.org/fae/CBR.html> [24.1.2020]. – Ich möchte es vermeiden, durch eine mögliche direkte Weitergabe der Daten meinerseits eine sekundäre Datenquelle zu schaffen. Interessierte laden den Datensatz daher bitte selbst von der Originalquelle.

Nach dem SEE-Wert der Rollet-1888-Regression war ein mittlerer Fehler von 3.894 zu erwarten. Demgegenüber fällt die tatsächliche Differenz im Mittel kleiner aus.

Der Versuch zeigt, dass die Anwendung auf eine tatsächlich externe Testserie überzeugender und klarer darlegt, ob eine Regressionsrechnung erfolgreich war und deren Parameter gut verwendbar sind. Trotz des hier demonstrierten Erfolgs sei betont: dies ist kein Forschungsbetrag des Autors und kein Vorschlag einer neuen Schätzformel, sondern wirklich nur ein Beispiel für ein m. E. erstrebenswertes methodisches Vorgehen.

### **25.6.9 Fallstudie Multiple Lineare Regression**

Körperhöhenschätzungen werden üblicherweise anhand der Langknochen durchgeführt, wovon es beim Menschen sechs verschiedene gibt. In der Praxis werden indes (wegen ihrer schwierigen Geometrie) die Elle (Ulna) und (wegen seiner Fragilität) das Wadenbein (Fibula) zumeist nicht berücksichtigt. Die meisten Schätzformeln nutzen den Oberarmknochen (Humerus), die Speiche (Radius), den Oberschenkelkochen (Femur) und das Schienbein (Tibia). Wobei die Tibia in der Praxis etwas problematisch ist, da deren Länge nach mindestens drei verschiedenen Standards erhoben wird (T1, T1a, T1b, nach Martin 1914) und die jeweiligen Bearbeiter nur das von ihnen präferierte Maß mitteilen. Gleichwie: in der Regel stehen vier Langknochen für Schätzungen zur Verfügung. Eine Option ist es, für jeden einzelnen Knochen eine Schätzformel abzuleiten und im Anwendungsfall ggf. vier Schätzergebnisse zu berechnen und deren Mittelwert als Ergebnis auszugeben – so verfahren z. B. die in Mitteleuropa viel benutzten Schätzungen nach Breitingen (1938) und Bach (1965). Eine Alternative ist die Anwendung der multiplen linearen Regression, bei der mehrere unabhängige Variablen gemeinsam auf das Ergebnis einwirken. Diesen Fall wollen wir – hier vereinfachend nur für Femur, Tibia und Humerus – im Folgenden modellhaft durcharbeiten, wiederum am Datensatz Rollet 1888.

Nach Laden des Datensatzes: => Statistik; => Modelle anpassen; => Lineare Regression; Körperhöhe als abhängige Variable einsetzen; als unabhängige Variable F1, H1 und T1 auswählen (Strg- bzw. Ctrl-Taste gedrückt halten); mit OK bestätigen. Wir erhalten:

```
lm(formula = Khoehoe ~ F1 + H1 + T1, data = Rollet1888)
```

```
Residuals:
```

```
      Min       1Q   Median       3Q      Max
-8.9825 -2.1371 -0.0433  2.0831  9.5935
```

```
Coefficients:
```

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  50.82918    4.97472  10.218 < 2e-16 ***
F1            0.07255    0.03286   2.208 0.02966 *
H1            0.16066    0.04259   3.772 0.00028 ***
T1            0.08123    0.03761   2.160 0.03327 *
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 3.359 on 96 degrees of freedom
```

```
Multiple R-squared:  0.8412,    Adjusted R-squared:  0.8362
```

```
F-statistic: 169.5 on 3 and 96 DF,  p-value: < 2.2e-16
```

Die Ausgabe beschreibt das Ergebnis einer Regression, bei der H1, F1 und T1 gleichermaßen in das Ergebnis eingehen. Dabei ist, wie der Block „Coefficients“ der Ausgabe ausweist, H1 für das Ergebnis hochsignifikant, F1 und T1 signifikant. Das Bestimmtheitsmaß  $R^2$  fällt mit um 0.84 etwas höher aus als bei der Schätzung zuvor allein anhand des Femurs (F1: 0.79). Der Vergleich der Residuen zeigt, dass hier die Spanne, in die 50% aller Schätzungen fallen, mit ca. -2.1 bis +2.1 cm etwas enger ausfällt und auch die Maxima der Residuen mit ca. -9.0 bis +9,6 etwas näher beieinander liegen als im Falle der Regression allein anhand von F1. Die Hinzuziehung weiterer Messstrecken verbessert das Schätzergebnis demnach. Somit lautet die Schätzformel dieser multiplen Regression:

$$\text{geschätzte Körperhöhe} = 50.8292 + H1 * 0.16066 + F1 * 0.07255 + T1 * 0.08123$$

Man füge die Ergebnisse wiederum dem Datensatz hinzu. Der Bonferroni-Ausreißer-Test signalisiert: keine auffallenden Ausreißer. Mit vier geschätzten Parametern liegt der kritische Hat Value nun bei  $(4+1)/100 = 0.05$ . Er wird von Fall Nr. 15 und 27 überschritten, doch weisen beide eine noch vergleichsweise übliche Cook-Distanz auf. Die üblichen Grafiken (Histogramm der studentisierten Residuen, „Einflussgrafik“, „Influence Index Plot“) lassen keinen Fall-Ausschluss als zwingend notwendig erscheinen. Für das Gesamtverfahren zeigen die „Grundlegenden diagnostischen Grafi-

ken“ wiederum Bilder, die das weitgehende Einhalten der Voraussetzungen einer linearen Regression bestätigen,<sup>121</sup> insbesondere zeigen die Residuengrafiken wieder diffuse Wolken statt Strukturen. Neben der optischen Prüfung sind harte Tests sinnvoll: Die Residuen sollten normalverteilt sein. => Statistik; => Deskriptive Statistik; => Test auf Normalverteilung; „Shapiro-Wilk-Test“ auswählen und die Residuenvariable eingeben („residuals.regF1T1H1“ o.ä.). Wir erhalten:

```
Shapiro-Wilk normality test
data: residuals.regF1T1H1
W = 0.99363, p-value = 0.9229
```

... und sehen an dem hohen p-Wert (nahe 1), dass keine signifikante Abweichung von der Normalverteilung vorliegt. Erhielten wir einen kleinen p-Wert unter 0.05, wären die Residuen nicht normalverteilt, was wiederum darauf zurückschließen ließe, dass die Voraussetzungen für die Anwendung eines linearen Verfahrens verletzt wären.<sup>122</sup>

Die multiple lineare Regression macht einige weitere Voraussetzungen an die Daten, deren Erfüllung wir mit Tests prüfen. Wir beginnen mit der Autokorrelation resp. Unabhängigkeit: Für jedes Paar der eingegebenen Fälle sollten die Fehler (Residuen) unabhängig voneinander sein. Daher: => Modelle; => Numerische Diagnose; => Durbin-Watson-Test auf Autokorrelation, und erhalten:

```
Durbin-Watson test
data: Khoehue ~ dummy + F1 + H1 + T1
DW = 1.9559, p-value = 0.7637
alternative hypothesis: true autocorrelation is greater than 0
```

Der Durbin-Watson-Wert liegt bei 1.96. Werte unter 1 und über 3 wären Warnsignale, hier liegt das beobachtete 1.96 nahe am Idealwert 2.0 – entsprechend fällt der p-Wert mit 0.76 gänzlich unsignifikant aus. Die Voraussetzung der Unabhängigkeit ist erfüllt.

Als nächstes prüfen wir das Vorliegen von Homoskedasität resp. Heteroskedasität: Bei einer multiplen linearen Regression sollten die Abweichungen von der Regressionsgeraden (i.e. die Residuen) unabhängig von den eingegebenen Schätzparametern

---

<sup>121</sup> Suchstichworte zum Nachschlagen: Homoskedasität, Heteroskedasität.

<sup>122</sup> In diesem Fall wäre die Anwendung einer nicht-linearen Regression zu erwägen, oder eine geeignete Variablentransformation vorab durchzuführen.

sein; liegt Abhängigkeit vor, spricht man von Heteroskedasizität. Daher: => Modelle; => Numerische Diagnose; => Breusch-Pagan-Test auf Heteroskedasizität; Häkchen an „unabhängige Variablen“, OK & Enter. Wir erhalten:

```

Breusch-Pagan test
data:  Khoehe ~ dummy + F1 + H1 + T1
BP = 0.70375, df = 3, p-value = 0.8723

```

Der Testwert BP ist (p-Wert) nicht signifikant, d.h. die Verfahrens-Voraussetzung der Homoskedasizität ist nicht verletzt. Wir wiederholen den Test für die „vorhergesagten Werte“ und erhalten die Bestätigung: ein nicht signifikanter p-Wert.

```

Breusch-Pagan test
data:  Khoehe ~ dummy + F1 + H1 + T1
BP = 0.033353, df = 1, p-value = 0.8551

```

Als drittes prüfen wir die Frage der Multikollinearität: Die in eine multiple Regression eingegebenen Schätzer dürften keine enge Korrelation zueinander aufweisen. Denn wenn ein Schätzer de fakto 1 : 1 das Gleiche besagt wie ein zweiter Schätzer, sind Verfahrensvoraussetzungen nicht erfüllt. Also: => Modelle; Numerische Diagnose; => Varianzinflationsfaktor, OK & Enter. Wir erhalten:

```

Rcmdr> vif(regF1T1H1)
      F1      H1      T1
8.298703 8.260075 7.658881

```

```

Rcmdr> round(cov2cor(vcov(regF1T1H1)), 3) # Correlations of parameter
estimates
      (Intercept)      F1      H1      T1
(Intercept)      1.000 -0.268  0.028 -0.116
F1                -0.268  1.000 -0.518 -0.460
H1                0.028 -0.518  1.000 -0.456
T1                -0.116 -0.460 -0.456  1.000

```

Die drei VIF-Werte (i. e. „Varianzinflationsfaktor“; in der Ausgabe oben) liegen jeweils nahe 8, was hoch ist: Ist der größte beobachtete VIF-Wert größer 10, liegt sicher ein Problem vor; ist der Mittelwert aller VIF-Werte deutlich größer 1, könnte ein Problem vorliegen. Letzteres ist in unserem Fall gegeben. Zudem leitet man aus dem VIF-Wert die sog. Toleranz ab, sie berechnet sich als 1/VIF. Dazu heißt es: eine Toleranz kleiner 0.1 signalisiert das Vorliegen eines ernststen Problems, eine Toleranz kleiner 0.2 signalisiert den Verdacht auf Vorliegen eines Problems. In unserem Fall ergeben sich Werte

um 0.12 bzw. 0.13 – wir sind beunruhigt. Die von **R** darunter ausgegebene Korrelationsgrafik zeigt, dass mit Werten bis -0.518 eine vergleichsweise hohe Korrelation insbes. beim Humerus (H1) vorliegt, während die Beziehung zwischen F1 und T1 mit Werten unter 0.46 schwächer ausfällt.

Angesichts der guten Erfüllung aller anderen Voraussetzungen des Verfahrens ist an dieser Stelle also eine explizite Entscheidung zu treffen, ob man das erkannte Problem einer relativ großen Abhängigkeit der Eingangsvariablen untereinander ignoriert, oder beispielsweise unter Ausklammerung der Humerus nunmehr eine lineare Regression allein auf Basis Femur (F1) und Tibia (T1) rechnen möchte um zu prüfen, ob dieser Weg die genannten Voraussetzungen besser erfüllt. Der entsprechende Versuch sei dem Leser überlassen ...

Da die Körpergröße mit der Länge aller Langknochen eng zusammenhängt, was ja dem eigentlichen Ziel des Schätzens auch nur zu Gute kommt, ist das letztbeschriebene Problem von der Sache her im Grunde nicht zu umgehen. Daher sei der Hinweis erlaubt, dass Körperhöheschätzungen z. B. nach Bach (1938) / Breitinger (1965) zwar alle Langknochen berücksichtigen, aber nicht auf multiplen linearen Regressionen beruhen, sondern auf einfachen linearen Regressionsgleichungen für jede einzelne Knochenstrecke, aus denen anschließend ein Mittelwert gebildet wird. Doch in Summe gilt auch hier das bereits zuvor Konstatierte: im Grunde bemisst sich der Erfolg einer Regression weniger an den internen statistischen Parametern, sondern mehr an deren Erfolg, d. h. Brauchbarkeit für externe Schätzungen, die nur an einer Testserie validiert werden kann.

Lassen Sie uns abschließend noch wie bei der einfachen linearen Regression die 66,66 %-Konfidenzintervalle mit Hilfe eines Bootstrapping (5.000 Samplings) ermitteln, womit wir verteilungsunabhängig die beiden Regressionsparameter angeben können, innerhalb derer sich 2/3 aller Beobachtungen bewegen.

```

Bootstrap bca confidence intervals
                16.7 %      83.3 %
(Intercept) 46.56452417 55.0600792
F1          0.04563648  0.1056865
H1          0.11866151  0.2054775
T1          0.04688860  0.1190537

```

Bilanz insgesamt: Liegt die Annahme vor, dass mehrere Variablen die zu schätzende Größe beeinflussen, kann eine multiple (lineare) Regression durchgeführt werden. Das Verfahren zeigt an, ob dies tatsächlich der Fall ist und ob die Stärke des Einflusses ev.

unterschiedlich groß ist (p-Werte bei der Ausgabe der Koeffizienten). Anschließend sind die Ergebnisse wie bei der üblichen linearen Regression zu prüfen. Im vorliegenden Fall war die Anwendung des multiplen Verfahrens berechtigt. Allerdings ist das resultierende Bestimmtheitsmaß ( $R^2$ ) nur sehr wenig größer als beim vergleichbaren Fall der linearen Regression allein anhand des Femurs und auch der mittlere Schätzfehler (Std.abw. der Residuen) mit ca. 3.2 cm nur unwesentlich kleiner als bei der Femurschätzung allein (ca. 3.5) – d.h. der Zugewinn an Richtigkeit durch die Hinzuziehung von Tibia und Humerus ist gegeben, aber nicht sehr groß.

Das Thema Regression und Prüfung deren Voraussetzung ist auch deswegen wichtig, weil viele – nicht alle – multivariate Verfahren, insbesondere die Hauptkomponentenanalyse und die Faktorenanalyse, aus dem Ansatz der Regressionsrechnung erwachsen und ähnliche Voraussetzungen an die Natur und Qualität der Daten machen wie die Regressionsrechnung.

## **25.7 KLIMADATEN: EISBOHRKERNE U.A.**

Nein, wir wollen hier keine Klimaforschung betreiben! Der Anspruch dieses Abschnitts ist weitaus bescheidener: immer mal wieder möchten Archäologen die von ihnen untersuchten Zeiten und Kulturen mit der Klimageschichte vergleichen, und immer wieder blicken Klimaforscher auf die Geschichte menschlicher Gemeinschaften in Relation zu den von ihnen untersuchten Phänomenen. Alles, was es seitens der Archäologie dazu braucht ist: Wie komme ich an die Daten? Wie kann ich aus den Daten anschauliche Grafiken ableiten? Wohlan, genau das wollen wir uns in diesem Kapitel erarbeiten.

In der Praxis erzeugen wir mit Hilfe des R-Commanders samt kleinen Überarbeitungen seines Codes eine Folge von Streuungsdiagrammen – weshalb dieser Abschnitt den am Klima weniger interessierten Lesern auch dazu dienen kann, ihre Fertigkeiten im Anfertigen von Streuungsdiagrammen auszubauen. Die Befehlsfolgen sind wiederum als auf den R-Commander bezogene Skripte bei den Übungsdateien ablegt, so dass die Übungen statt durch Abschreiben auch per Copy & Paste nachgearbeitet werden können.

### **25.7.1 Woher die Daten kommen: Archive für Klimadaten**

Das Nachvollziehen der einschlägigen Forschungen der Klima-Profis und das Visualisieren der Daten ist erheblich weniger kompliziert als man eventuell erwartet, da es in der Klimaforschung seit langem üblich ist, die zu Grunde liegenden Daten zusammen mit ihrer Erstpublikation in öffentlich zugänglichen Archiven zu deponieren. Wer

in der aktuellen Debatte um „Big Data“ und nationale Datenarchive in der Archäologie aus den Erfahrungen anderer Wissenschaften profitieren möchte: die Klimaforschung und ihr Umgang mit Daten bietet dazu reiches Anschauungsmaterial.

Die beiden großen Datenarchive, welche vor allem die Daten zu den Eisbohrkernen auf Grönland und von der Antarktis bereitstellen, werden von der Universität Kopenhagen und der NOAA „National Oceanic and Atmospheric Administration“ (USA) betrieben (s. unten „Quellen“). Ein weiterer für die Archäologie wichtiger Datenpool liegt in den Niederlanden, wo der „KNMI Climate Explorer“ vor allem Datensätze zu jüngeren Zeitabschnitten sammelt und bereitstellt, d. h. Daten, die das jüngste Jahrtausend oder die Zeit ab Christi Geburt betreffen. In allen Fällen geht dem Bereitstellen der Datensätze in diesen Archiven eine Erstpublikation in wissenschaftlichen Zeitschriften voraus, einer seriösen Weiterverwendung dieser Datensätze sollte stets die Lektüre der Erstpublikation vorausgehen.

Ganz grob skizziert gibt es im archäologie-relevanten Zeitabschnitt drei unterschiedliche Kernbereiche, die auf unterschiedlichen Daten beruhen und ihre Schwerpunkte in unterschiedlichen Zeitfenstern haben. Die Historische Klimatologie setzt bei modernen Instrumentendaten für das Wetter an, die ab Mitte des 18. Jh. aufgezeichnet wurden. Für die Zeit davor versucht sie, aufgrund historischer Quellen Informationen über das Klima und Wetter in der Vergangenheit zu erhalten (Pfister 1999). Diese Forschung kann Klima (und z. T. Wetter) – bisweilen sehr detailreich – bis um 1500 n.Chr. rekonstruieren. Das Schweizer EU-Projekt Euro-Climhist (2010) sammelt in Fortführung von Ende der 1970er Jahren gestarteten Projekten alle relevanten Daten und stellt sie zur Verfügung.

Sehr viel weiter in die zeitliche Tiefe führen die Eisbohrkerne, die seit Ende der 1970er Jahre lange Ablagerungssequenzen aus dem grönländischen Festlandeis gewinnen. Viel genutzt werden die Bohrkerne GISP (1979-81), GRIP (1990-92), GISP2 (1990-93) und NGRIP (1996-2003), die zeitlich in Summe bis etwa 125.000 v.H. zurückreichen. Aus ihnen werden verschiedene Klimaschätzer (sog. Proxy-Daten) gewonnen, am bekanntesten ist wohl das Verhältnis der Sauerstoffisotopen  $^{16}\text{O}$  und  $^{18}\text{O}$  ( $\delta^{18}\text{O}$ ), das die Paläotemperatur über der nördlichen Halbkugel widerspiegelt. Sie werden ergänzt durch Bohrungen im Festlandeis der Antarktis, die zeitlich noch tiefer hinabreichende Sequenzen ergaben (Wostok, 1970er Jahre ff., bis ca. 420.000 v.H.;

EPICA, 1995 ff., „Dome C“ bis ca. 900.000 v.H.) und zudem ein Studium des Zusammenhangs der Entwicklungen von nördlicher und südlicher Hemisphäre ermöglichen.<sup>123</sup>

Bei den Eisbohrkernen sind die Daten aus den oberen, jüngsten Eislagen wegen der noch nicht erfolgten bzw. erst kurz währenden Sedimentation und Verfestigung problematisch, weshalb es schwer ist, z. B. aktuelle Temperaturdaten direkt mit den aus den Eiskernen gewonnenen Proxies zu verknüpfen. In diese Lücke springen Klimainformationen, die sich auf das jüngste Jahrtausend und seit den 2010er Jahren vermehrt auch auf die jüngsten 2000 Jahre fokussieren, weil es hier möglich ist, beide Beobachtungs- und Zeitebenen miteinander zu verknüpfen. Dabei spielen Baumringe als Klimaarchiv eine zunehmend wichtige Rolle: sie liefern hochauflösende, jahrgenau datierte Informationen und haben – anders als die Eiskerne – auch das Potenzial, diese Informationen in hoher räumlicher Auflösung zu bieten. Aus ihnen und anderen Informationen kann das Klimageschehen der jüngeren Zeit unter Angabe von Temperaturen in Grad Celsius recht genau rekonstruiert werden (z.B. Moberg et al. 2006; Projekt PAGES: <http://pastglobalchanges.org/> [24.1.2020]).

### 25.7.2 Das $\delta^{18}\text{O}$ -Verhältnis seit 12.000 v.H. als Schätzer der Paläotemperatur

Für unser Einüben benutzen wir die  $\delta^{18}\text{O}$ -Werte des Bohrkerns NGRIP nach der 2005 publizierten, mehrere grönländische Eiskerne vereinenden Chronologie „GICC05“ mit einer Auflösung von 20 Jahren: „NGRIP - GICC05 time scale and  $\delta^{18}\text{O}$  values“, die als maschinenlesbare Tabelle im Format \*.txt angeboten werden (Rasmussen et al. 2006). Diese Tabelle enthält, wie bei solchen Datenpublikationen üblich, zuoberst eine knappe Erklärung der Daten und die Verweise auf die Primärpublikationen. Nach dem Herunterladen editiert man die Tabelle, löscht die Textzeilen des Kopfes, reduziert die Tabelle auf die Daten und setzt (resp. erhält) die Spaltenbezeichnungen in der ersten Zeile. Man achte darauf, ggf. die in der Tabelle vereinzelt vorkommende „NA“ zu löschen, d. h. dort leere Zellen zu haben, weil ein „NA“ beim Einlesen von Excel-Tabellen zu **R** in der Regel dazu führt, dass die gesamte Spalte als Zeichenvariable gelesen wird.<sup>124</sup> Da wir hier das Holozän verfolgen wollen, kann man am unteren Ende

---

<sup>123</sup> Gründliche Übersicht: Jouzel, J. (1997). A brief history of ice core science over the last 50 yr. *Climate of the Past*, 9, 2525-2547. doi:10.5194/cp-9-2525-2013; <https://www.clim-past.net/9/2525/2013/cp-9-2525-2013.pdf> [24.1.2020].

<sup>124</sup> Auch hier füge ich den Datensatz nicht als Übungsdaten bei, um keine Sekundärdaten zu erzeugen. Jeder Interessierte möge das Herunterladen und Aufbereiten selbst durchführen – was gewiss auch eine gute Übung ist.

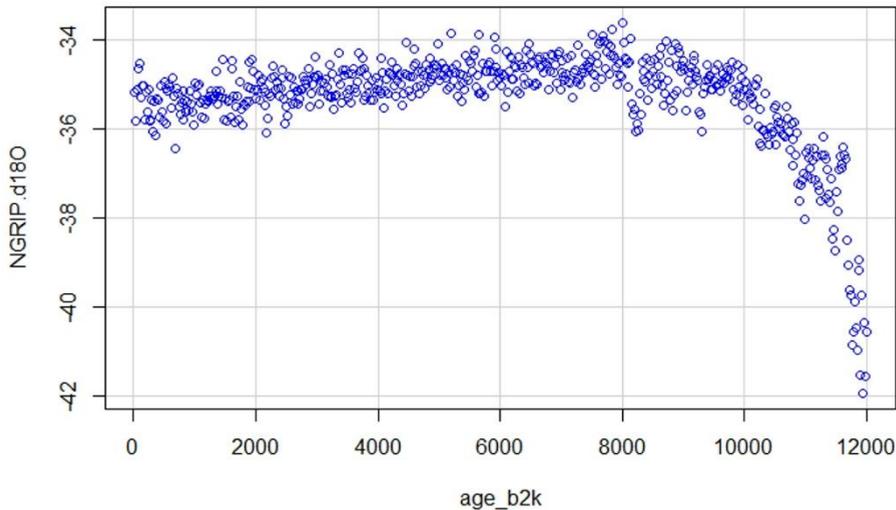
der Tabelle auch alle Daten älter als 12.000 v.H. löschen – und das Ergebnis unter neuem Namen als Arbeitstabelle speichern. Der Anfang der aufbereiteten Tabelle könnte wie folgt aussehen (**Abb. 118**).

age_b2k	NGRIP.depth	NGRIP.d18O	GRIP.depth	GRIP.d18O	yrs.max.error
20	7,35	-35,16	5,43		1
40	14,15	-35,82	13,60	-35,13	1
60	20,40	-35,11	21,43	-35,55	1
80	26,10	-34,65	28,23	-35,16	1
...	...	...			

**Abb. 118** Anfang der Datentabelle nach GICC05\_NGRIP\_GRIP\_20y (27nov2006).

Nach dem Einlesen der Datentabelle in den R-Commander ist mit => Graphik; => Streudiagramm; Zeit („age\_b2k“) auf die x-Achse,  $\delta^{18}\text{O}$  („NGRIP.d18O“) auf die y-Achse, usw. ein erstes Bild schnell gewonnen (**Abb. 119**). Man erkennt in ihm den sehr starken und schnellen Wiederanstieg der Temperaturen am Ende der Eiszeit, und beispielsweise auch das sog. „8200-Event“, einen kurzfristigen Kälterückschlag, an den sich das Wärmeoptimum des Holozäns ca. 8000 – 6.000 v.H. anschließt. Wie wir aus der Datenbeschreibung der Originaltabelle entnommen haben, ist die GICC05-Chronologie auf das Jahr 2000 als „Heute“ bezogen (ggf. Vorsicht: andere Chronologien halten dies anders!), weshalb man mit der Gleichung „year = 2000 – age\_b2k“ auch unsere gewohnten Kalenderjahre gewinnen kann (d. h. => Datenmanagement; => Variablen bearbeiten; => Erzeuge neue Variable; „year“ als neue Variable einsetzen und „2000 – age\_b2k“ als Berechnungsvorschrift.

Doch statt in unserem gewohnten System der R-Graphiken weiterzuarbeiten, wechseln wir zu Grafiken mit *ggplot2*, weil diese flexibler sind und weitergehende Möglichkeiten bieten. Dazu lade man in den R-Commander das Plugin „RcmdrPlugin.KMggplot2“ (=> Extras; => Lade Rcmdr Plugin(s) ...; usw.). Statt des Streudiagramms verfertigen wir ein Liniendiagramm: => KMggplot2; => Line chart; „year“ auf die x-Achse, „NGRIP.d18O“ auf die y-Achse, setzen Label für die Achsen und, & OK. Weil die Steuerungsmöglichkeiten an diesem Bild und dem Code im R-Commander, mit dem es erzeugt wurde, recht begrenzt, arbeite man anschließend mit Hilfe der verfügbaren Hilfen, Anleitungen etc. Schritt für Schritt weiter, um die Darstellung so zu verändern, dass sie den eigenen Vorstellungen entspricht.



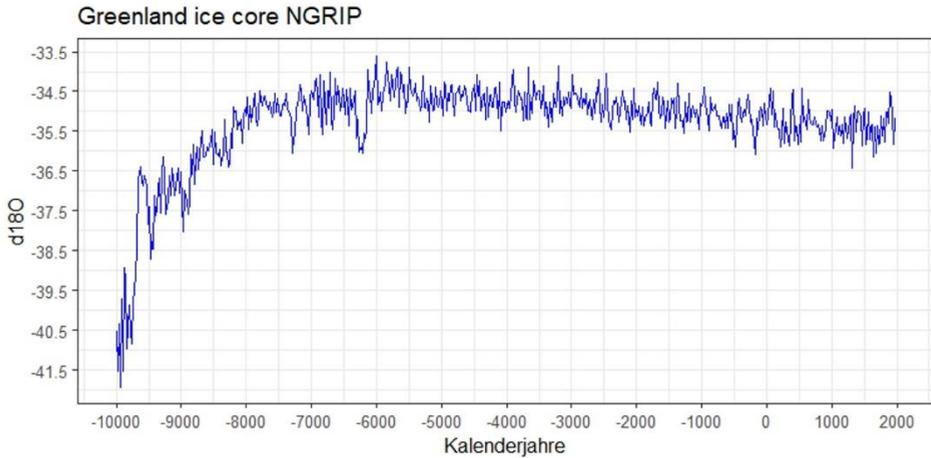
**Abb. 119** Erster, roher Plot der  $d^{18}O$ -Werte des Bohrkerns NGRIP.

Die via R-Commander und KMGgplot2 erzeugte Ausgangsversion sieht so aus (den Datensatz haben wir DTA genannt) – wobei wir uns die Abbildung der Grafik sparen:

```
# Variable "year" errechnen:
DTA$year <- with(DTA, 2000 - age_b2k)
#
require("ggplot2")
.df <- data.frame(x = DTA$year, y = DTA$NGRIP.d18O)
.df <- .df[order(.df$x), ]
.plot <- ggplot(data = .df, aes(x = x, y = y)) +
  geom_point() +
  geom_line(size = 1) +
  scale_y_continuous(expand = c(0.01, 0)) +
  xlab("Kalenderjahre") +
  ylab("d180") +
  labs(title = "Greenland ice core NGRIP") +
  theme_bw(base_size = 14, base_family = "sans")
print(.plot)
rm(.df, .plot)
#
```

Eine etwas ansehnlichere Fassung dieser Grafik ergibt sich nach einigen weiteren Spezifizierungen (**Abb. 120**) durch das nachfolgende Skript:

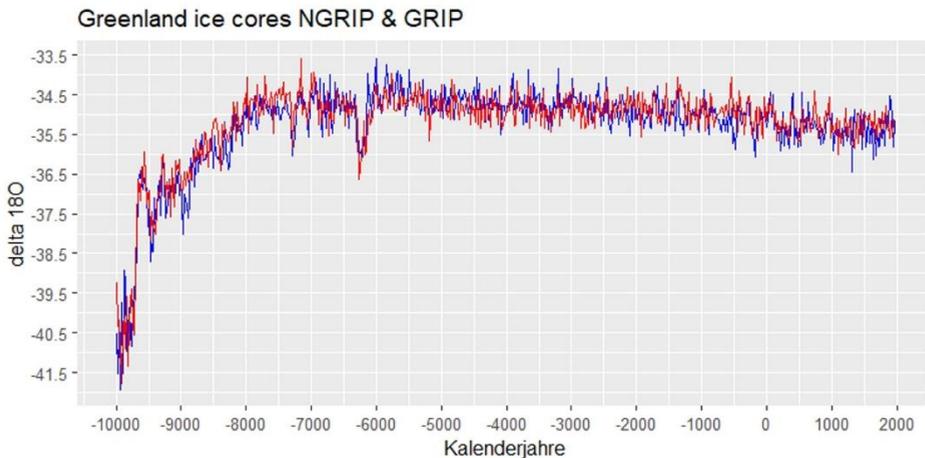
```
#
require("ggplot2")
# Größe des Plots festlegen:
windows(width=20/2.54, height=10/2.54)
#
.df <- data.frame(x = DTA$year, y = DTA$NGRIP.d180)
.df <- .df[order(.df$x), ]
.plot <- ggplot(data = .df, aes(x = x, y = y)) +
# Farbe und Linienstärke des Linienzuges vereinbaren:
  geom_line(color="blue",size=0.7) +
# Setzen von Min., Max., und Gitterweite der y- und x-Skala
  scale_y_continuous(breaks=seq(-41.5,-33.5,1)) +
  scale_x_continuous(breaks=seq(-10000,2000,1000)) +
# Beschriftung der Achsen, Vergabe eines Titels:
  xlab("Kalenderjahre") +
  ylab("d180") +
  labs(title = "Greenland ice core NGRIP") +
  theme_bw(base_size = 12, base_family = "sans")
print(.plot)
rm(.df, .plot)
#
```



**Abb. 120** Entwicklung des  $\delta^{18}\text{O}$ -Isotopenverhältnisses im Eisbohrkern NGRIP ab 10.000 v.Chr.

Nun eine Version dieser Darstellung mit zwei Datenserien in der gleichen Graphik, d. h. einer gemeinsamen x-Achse und zwei Variablen auf der y-Achse, und zwar den  $\delta^{18}\text{O}$ -Werten des Bohrkerns NGRIP, überlagert mit jenen des Bohrkerns GRIP (**Abb. 121**).

```
#
require("ggplot2")
# Größe des Plots festlegen:
windows(width=20/2.54, height=10/2.54)
#
Plot <- ggplot(data=DTA,
  aes(x=DTA$year, y1=DTA$NGRIP.d180, y2=DTA$GRIP.d180)) +
# Farbe und Linienstärke des Linienzuges vereinbaren:
  geom_line((aes(y=DTA$NGRIP.d180)),color="blue",size=0.7) +
  geom_line((aes(y=DTA$GRIP.d180)),color="red",size=0.7) +
# Setzen von Min., Max., und Gitterweite der y- und x-Skala
  scale_y_continuous(breaks=seq(-41.5,-33.5,1)) +
  scale_x_continuous(breaks=seq(-10000,2000,1000)) +
# Setzen der Achsenbeschriftungen und des Diagrammtitels
  xlab("Kalenderjahre") +
  ylab("delta 18O") +
  labs(title="Greenland ice cores NGRIP & GRIP") +
# Setzens des Layouts insgesamt auf das Thema "grau":
  theme_grey(base_size=12, base_family="sans")
print(Plot)
rm(Plot)
#
```



**Abb. 121** Entwicklung des  $\delta^{18}\text{O}$ -Isotopenverhältnisses in den Eisbohrkernen NGRIP (blau) und GRIP (rot) ab 10.000 v.Chr.

Nun eine Fassung des Streudiagramms, in der durch die Eintragung von Text, einem Pfeil und der Hinterlegung mit farbigen Arealen bestimmte Momente hervorgehoben sind (**Abb. 122**):

```
#
require("ggplot2")
# Größe des Plots festlegen:
windows(width=20/2.54, height=10/2.54)
#
Plot <- ggplot(data=DTA,
  aes(x=DTA$year, y=DTA$NGRIP.d180)) +
# Farbe und Linienstärke des Linienzuges vereinbaren:
  geom_line((aes(y=DTA$NGRIP.d180)),color="blue",size=0.7) +
# Einfügen von Text in die Grafik:
  geom_text(x=-5700,y=-36.5,label="8200-event",color="black") +
# mit Pfeil:
  geom_segment(aes(x=-5500,y=-36,xend=-6100,yend=-36),arrow=ar-
row(length=unit(0.5,"cm")), color="red") +
# Eintragen von Klimazonen:
  annotate("rect",xmin=-9700,xmax=-8700,ymin=-42.0,ymax=-33.5,alpha=.1,fill="red")
+
  geom_text(x=-9000,y=-41.0,label="Praeboreal",color="red") +
  annotate("rect",xmin=-8700,xmax=-7200,ymin=-42.0,ymax=-33.5,alpha=.1,fill="blue")
+
  geom_text(x=-8000,y=-41.5,label="Boreal",color="blue") +
  annotate("rect",xmin=-7200,xmax=-3700,ymin=-42.0,ymax=-33.5,alpha=.1,fill="red")
+
  geom_text(x=-5500,y=-41.0,label="Atlantikum",color="red") +
# Setzen von Min., Max., und Gitterweite der y- und x-Skala
  scale_y_continuous(breaks=seq(-41.5,-33.5,1)) +
  scale_x_continuous(breaks=seq(-10000,2000,1000)) +
# Setzen der Achsenbeschriftungen und des Diagrammtitels
  xlab("Kalenderjahre") +
  ylab("delta 180") +
  labs(title="Greenland ice core NGRIP") +
# Setzen des Layouts auf das einfache Thema "bw":
  theme_bw(base_size=12, base_family="sans")
print(Plot)
rm(Plot)
#
```

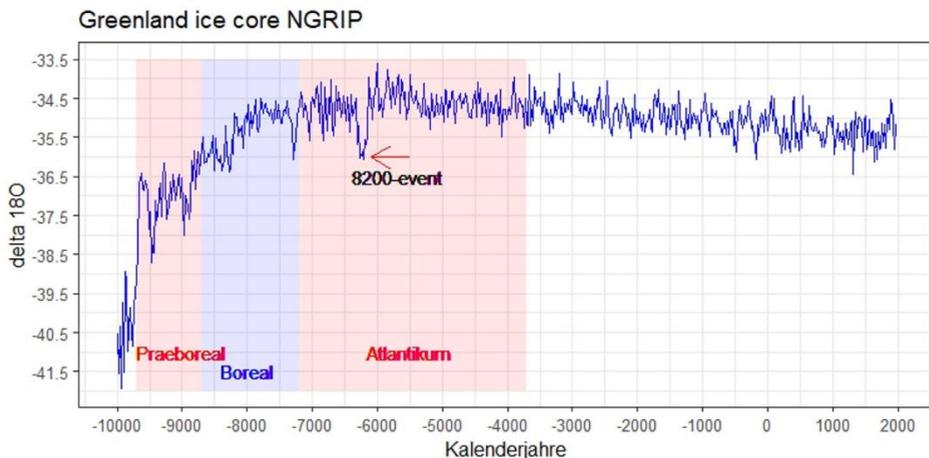


Abb. 122 Beispiel für ein Streuungsdiagramm mit erläuternden Eintragungen.

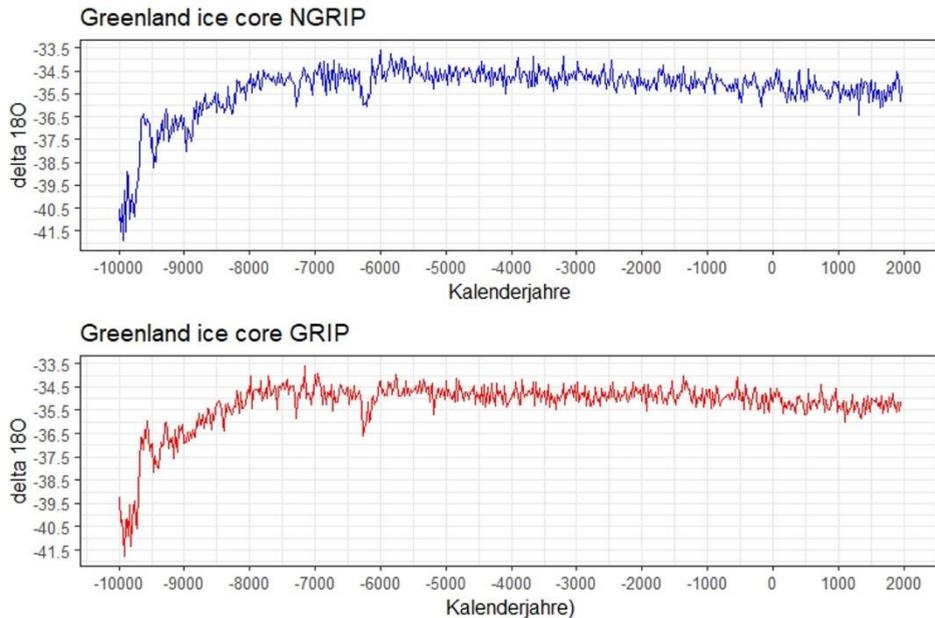
Nun statt der Darstellung von zwei Variablen in der y-Achse eines Diagramms zwei getrennte Diagramme mit gleich skalierter x-Achse exakt untereinander. Dazu wird zunächst die Funktion „Multiplot“ von W. Chang übernommen und genutzt:

```
#
require("ggplot2")
# ###
# Multiple plot function
# ###
# ggplot objects can be passed in ..., or to plotlist (as a list of ggplot objects)
# - cols: Number of columns in layout
# - layout: A matrix specifying the layout. If present, 'cols' is ignored.
#
# If the layout is something like matrix(c(1,2,3,3), nrow=2, byrow=TRUE),
# then plot 1 will go in the upper left, 2 will go in the upper right, and
# 3 will go all the way across the bottom.
#
# ###
# taken from: Chang, W. (n.d.). Website Cookbook for R. Website: http://www.cook-
book-r.com/ [24.1.2020].
# ###
multiplot <- function(..., plotlist=NULL, file, cols=1, layout=NULL) {
  library(grid)
  # Make a list from the ... arguments and plotlist
  plots <- c(list(...), plotlist)
  numPlots = length(plots)
  # If layout is NULL, then use 'cols' to determine layout
  if (is.null(layout)) {
    # Make the panel
    # ncol: Number of columns of plots
    # nrow: Number of rows needed, calculated from # of cols
    layout <- matrix(seq(1, cols * ceiling(numPlots/cols)),
                     ncol = cols, nrow = ceiling(numPlots/cols))
  }
}
```

```

if (numPlots==1) {
  print(plots[[1]])
} else {
# Set up the page
  grid.newpage()
  pushViewport(viewport(layout = grid.layout(nrow(layout), ncol(layout))))
# Make each plot, in the correct location
  for (i in 1:numPlots) {
# Get the i,j matrix positions of the regions that contain this subplot
  matchidx <- as.data.frame(which(layout == i, arr.ind = TRUE))
  print(plots[[i]], vp = viewport(layout.pos.row = matchidx$row,
                                layout.pos.col = matchidx$col))
  }
}
}
# #####
# Größe des Plots festlegen, dabei Einzelproportion vs. Gesamtproportion (=Größe)
bedenken:
windows(width=20/2.54, height=20/2.54)
#
plot1 <- ggplot(DTA, aes(x=DTA$year, y=DTA$NGRIP.d180)) +
  geom_line(color="blue",size=0.7) +
  scale_x_continuous(breaks=seq(-10000,2000,1000)) +
  scale_y_continuous(breaks=seq(-41.5,-33.5,1)) +
  xlab("Kalenderjahre") +
  ylab("delta 180") +
  labs(title="Greenland ice core NGRIP") +
  theme_bw(base_size=12, base_family="sans")
plot2 <- ggplot(DTA, aes(x=DTA$year, y=DTA$GRIP.d180)) +
  geom_line(color="red",size=0.7) +
  scale_x_continuous(breaks=seq(-10000,2000,1000)) +
  scale_y_continuous(breaks=seq(-41.5,-33.5,1)) +
  xlab("Kalenderjahre") +
  ylab("delta 180") +
  labs(title="Greenland ice core GRIP") +
  theme_bw(base_size=12, base_family="sans")
multiplot(plot1,plot2,rows=2)
rm(plot1,plot2)
#

```



**Abb. 123** Die beide Datenserien als gestapeltes Diagramm.

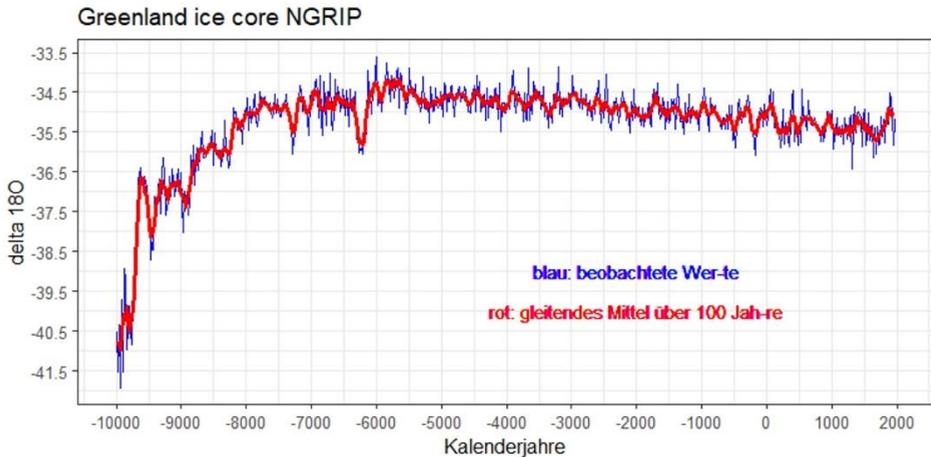
Abschließend verfertigen wir ein Diagramm, das zusätzlich zu den beobachteten Werten den gleitenden Mittelwert einspiegelt, für den hier ein Zeitfenster von 100 Jahren (d. h. jeweils 5 Beobachtungen) gewählt wurde. Dazu muss zuvor das Paket „caTools“ installiert werden, das dann vom R-Commander aus geladen wird, um die Berechnung des gleitenden Mittelwerts durchzuführen:

```
#
require("caTools")
#
DTA$d180.d100 <- runmean(DTA$NGRIP.d180, 5, endrule="constant", align="center")
#
require("ggplot2")
# Größe des Plots festlegen:
windows(width=20/2.54, height=10/2.54)
#
Plot <- ggplot(data=DTA,
  aes(x=DTA$year, y1=DTA$NGRIP.d180, y2=DTA$d180.d100)) +
# Farbe und Linienstärke des Linienzuges vereinbaren:
  geom_line((aes(y=DTA$NGRIP.d180)),color="blue",size=0.7) +
  geom_line((aes(y=DTA$d180.d100)),color="red",size=1.2) +
  geom_text(x=-2000,y=-39,label="blau: beobachtete Werte",color="blue") +
  geom_text(x=-2000,y=-40,label="rot: gleitendes Mittel über 100 Jahre",color="red")
+
# Setzen von Min., Max., und Gitterweite der y- und x-Skala
```

```

scale_y_continuous(breaks=seq(-41.5, -33.5, 1)) +
scale_x_continuous(breaks=seq(-10000, 2000, 1000)) +
# Setzen der Achsenbeschriftungen und des Diagrammtitels
xlab("Kalenderjahre") +
ylab("delta 18O") +
labs(title="Greenland ice core NGRIP") +
# Setzens des Layouts auf das Thema "bw":
theme_bw(base_size=12, base_family="sans")
print(Plot)
rm(Plot)
#

```



**Abb. 124** Einspielung eines gleitenden Mittelwerts in die Zeitreihe des Sauerstoffisotopenverhältnisses.

### 25.7.3 Quellen / Archive zum Kapitel

- Euro-Climhist (2010). *Euro-Climhist – Eine Klimadatenbank für die Schweiz und Europa*. <https://www.euroclimhist.unibe.ch/de/> [24.1.2020].
- KNMI Climate Explorer: <http://climexp.knmi.nl/start.cgi> [24.1.2020].
- NOAA National Centers for Environmental Information (USA): <https://www.ncdc.noaa.gov/paleo-search/reports/all> [24.1.2020]; hier ggf. „data type“ auswählen, z. B. „ice core“ und Search-Taste klicken, um die Auflistung einsehen zu können.
- PAGES – Past Global Changes: <http://pastglobalchanges.org/> [24.1.2020]. – Übersicht zu verfügbaren Daten: <http://pastglobalchanges.org/data/databases> [24.1.2020].

- Rasmussen, S. O., et al. (2006). A new Greenland ice core chronology for the last glacial termination. *Journal of Geophysical Research*, 111, D06102. – Datensatz: <https://www.ncdc.noaa.gov/paleo-search/study/2493> [24.1.2020]. – Vgl. Vinther, B. M., et al. (2006). A synchronized dating of three Greenland ice cores throughout the Holocene. *Journal of Geophysical Research*, 111, D13102. <https://www1.ncdc.noaa.gov/pub/data/paleo/icecore/greenland/summit/ngrip/gicc05-holocene-20yr.txt> [24.1.2020].
- Universität Kopenhagen, Former Center for Ice and Climate: <http://www.iceandclimate.nbi.ku.dk/data/> [24.1.2020]. Archäologierelevantes insbes. unter (rechts oben) “annual climate indices”.

#### 25.7.4 Literatur zum Kapitel

- Frank, D. Esper, J., Zorita, E. & Wilson, R. (2010). A noodle, hockey stick, and spaghetti plate: a perspective on high-resolution paleoclimatology (Opinion). *Wiley interdisciplinary reviews: Climate Change* 1(4), 507-516. <https://onlinelibrary.wiley.com/doi/epdf/10.1002/wcc.53> [24.1.2020].
- Moberg, A., Sonechkin, D. M., Holmgren, K., Datsenko, N. M. & Karlén, W. (2005). Highly variable Northern Hemisphere temperatures reconstructed from low- and high-resolution proxy data. *Nature*, 433, 613-617. doi:10.1038/nature03265 – Zugehörige Daten: <https://bolin.su.se/data/moberg-2005> [24.1.2020].
- Neukom, R., Steiger, N., Gómez-Navarro, J. J., Wang, J. & Werner, J. P. (2019). No evidence for globally coherent warm and cold periods over the preindustrial Common Era. *Nature*, 571, 550-554 (24.7.2019). – Zugehörige Daten total: <https://www1.ncdc.noaa.gov/pub/data/paleo/pages2k/neukom2019temp/> [24.1.2020], Daten des Ergebnisses: [https://www1.ncdc.noaa.gov/pub/data/paleo/pages2k/neukom2019temp/recons/Full\\_ensemble\\_median\\_and%2095pct\\_range.txt](https://www1.ncdc.noaa.gov/pub/data/paleo/pages2k/neukom2019temp/recons/Full_ensemble_median_and%2095pct_range.txt) [24.1.2020].
- Pfister, Chr. (1999). *Wetternachhersage. 500 Jahre Klimavariationen und Naturkatastrophen (1496-1995)*. Bern: Haupt.
- Sirocko, Fr. (Hrsg.) (2009). *Wetter, Klima, Menschheitsentwicklung von der Eiszeit bis ins 21. Jahrhundert*. Darmstadt: WBG.
- Wanner, H. (2016). *Klima und Mensch – eine 12'000-jährige Geschichte*. Bern: Haupt.

## 26 SCHLUSSWORT

---

Hier endet die Einführung (sic), und zwar aus zwei Gründen. Einerseits stößt man, wenn man das Thema Archäologie & Statistik weiter vertiefen möchte, nun allmählich an die Grenzen des Funktionsumfangs unserer grafischen Benutzeroberfläche R-Commander. Für die nächsten, sinnvollen Vertiefungen sollte man sich vom R-Commander lösen und z. B. mit dem umfangreicheren BlueSky weiterarbeiten, besser aber sich mit Hilfe von RStudio auf das Arbeiten mit **R** via Kommandozeile einlassen. Andererseits hat man als Archäologe mit dem bislang Erlernten einen Stand erreicht, mit dem sich sehr viele Probleme des Forschungsalltags bewältigen lassen. Meines Erachtens ist es sinnvoll, seine Fähigkeiten zunächst auf dieser Ebene zu stabilisieren, etwa durch ein erneutes Durcharbeiten dieses Buchs und vor allem viele weitere Praxiserfahrung anhand eigener Problemstellungen, um mehr Sicherheit zu gewinnen, als allzu forscht auf komplexere Statistik-Themen hinzuarbeiten. Man lasse sich nicht durch kompliziert wirkende Namen und trendige Themen blenden: Wer bis hierhin gut durchgekommen ist, kann sehr viel und ist für Alltag eines quantitativ arbeitenden und argumentierenden Archäologen gut gerüstet.

## 27 KOMMENTIERTE LITERATURLISTEN

---

Die Zitierweise der Literatur folgt dem Standard APA 6. Abweichend davon werden zur leichteren Bibliografierbarkeit die Vornamen der Autoren nach Möglichkeit nicht abgekürzt.

### 27.1 VERTIEFENDE LITERATUR ZU R

Anders als in der alphabetisch geordneten Liste der zitierten Literatur ist die Reihenfolge der Titelnennungen hier nicht zufällig, sie entspricht der von mir empfohlenen Abfolge der Lektüre resp. der typischen Leser-Bedürfnisse.

- Fox, John (2017). Using the R Commander: A point-and-click interface for R. Boca Raton: Chapman & Hall.

Fox ist der Programmautor des R-Commanders, dessen Ziel es ist, den Benutzern statt der Bedienung via Kommandozeile eine moderne grafische Benutzeroberfläche zu bieten. Fox stellt den R-Commander vor und geht anschließend exemplarisch verschiedene statistische Analysen mit **R** durch, die mit Hilfe des R-Commanders umgesetzt werden. Ein (sehr) guter Start für das Lernen von **R**.

- Kabacoff, Robert I. (2015). *R in action: Data analysis and graphics with R*. Second edition. Shelter Island: Manning.

Für viele Benutzer das R-Buch, an dem **R** gründlich erlernt werden kann. Es kann sehr gut auch als Nachschlagewerk verwendet werden. Kabacoff erwähnt zwar grafische Benutzeroberflächen wie den R-Commander, sein Buch jedoch vermittelt **R** in der Bedienung per Kommandozeile. Daher eindeutig Lese-Stoff für jene, die die ersten Schritte bereits hinter sich haben, dann aber der m. E. beste Einstieg in die Vertiefung.

- Matloff, Norm (2019). *fasteR: Fast Lane to Learning R!* GitHub. <https://github.com/matloff/fasteR> [24.1.2020].

Von der Idee her ähnlich wie Kabacoff (2015): pures **R** lernen ohne GUI. Hier jedoch statt des gedruckten Buches ein wirklich guter Kurs als frei zugängliches Online-Material. Gut geschrieben, zielt auf das handfeste praktische Einüben.

- Hinz, Martin (2019). vituR – a portal for video tutorials about R, especially for archaeologists. <http://vituR.archaeological.science/> [24.1.2020].

Eine von dem an den Universitäten Kiel und Bern tätigen Archäologen Martin Hinz erstellte Sammlung von Videos, welche die Umsetzung grundlegender statistischer Verfahren mit **R** und RStudio demonstriert – teils in deutscher, teils in englischer Sprache. Hinz arbeitet ohne graphische Benutzeroberfläche direkt in RStudio, d. h. seine Nutzer werden zugleich in das Coden eingeführt. Für alle, die gerne anhand von Videos lernen, eine wirklich gute Einführung und eine treffende Auswahl wichtiger Verfahren.

- Carlson, David L. (2017). *Quantitative Methods in Archaeology Using R*. (Cambridge Manuals in Archaeology). Cambridge: Cambridge University Press.

Wer als Archäologe statt Videos (siehe Hinz 2019) lieber mit einem Buch arbeitet, findet hier einen guten Startpunkt in englischer Sprache. Auch Carlson ist Purist: graphische Benutzeroberflächen zu **R** werden zwar erwähnt, das Buch arbeitet jedoch mit der Kombination von **R** und RStudio. Bei der Verfertigung von Graphiken beruht es auf dem R-Basispaket – die trendigen *ggplot2*-Graphiken bleiben außen vor. Innerhalb dieses Rahmens bietet Carlson eine sehr breite, viele unterschiedliche Themen behandelnde Einführung, die von einfachsten Anforderungen bis hin zu multivariaten Methoden reicht und diese stets an archäologischen Beispielen entwickelt. Das Buch bietet alle nötigen Code-Beispiele, die Beispieldaten stehen als R-Paket bereit, so

dass man entlang dieses Buchs die Analysen Schritt für Schritt nachvollziehen kann. Eine Stärke sind die konkreten Hinweise auch auf seltener angewandte R-Pakete, die zu den jeweils behandelten Themenfeldern nützlich sind. Eine Schwäche ist das weitgehende Fehlen des so wichtigen Themas Datenerfassung und -aufbereitung. Wenn man den Weg über die englische Sprache nicht scheut, bietet die Kombination von McCall 2018 und Carlson 2017 eine gute Grundlage für das Erlernen archäologierelevanter Statistik mit **R** im Selbststudium.

- Hain, Johannes (2011). *Statistik mit R: Grundlagen der Datenanalyse*. RRZN-Handbuch. RRZN: Hannover.

Wer für den Anfang etwas Ähnliches sucht wie Kabacoff (2015), aber lieber in deutscher Sprache, ist mit dem Heft von J. Hain gut versorgt, auch wenn es deutlich weniger Stoff bietet als Kabacoff. Die RRZN-Handbücher sind eine umfangreiche Reihe von Handbüchern für den Gebrauch von Studierenden; die Hefte sind sehr preiswert. Das Projekt des Regionalen Rechenzentrums für Niedersachsen (RRZN) sammelt gute Manuskripte von diversen Autoren an diversen Standorten und produziert sie als einfach gedruckte Reihe. Die Hefte können an nahezu allen Universitäten in Deutschland bezogen werden, üblicherweise zu einem Preis niedriger als etwa die Kopierkosten. Wie Kabacoff (2015) erklärt auch Hain (2011) den Umgang mit **R** auf Basis der Kommandozeilen.

- Chang, Winston (2019). *R Graphics Cookbook: Practical recipes for visualizing data*. 2<sup>nd</sup> edition. Sebastopol CA: O'Reilly.

Das Buch enthält, was der Titel verspricht: eine ausführliche Einführung in das Erzeugen statistischer Grafiken mit **R**. Wer tiefer einsteigen will resp. muss, als es mit dem R-Commander möglich ist, kann hier selbstständig weiterarbeiten und lernen, eine breite Vielfalt adäquater Grafiken zu erstellen. **R** enthält mehrere Grafik-Pakete. Nach einer Sichtung und Einführung in alle Pakete liegt der Schwerpunkt des Buches auf dem modernen Paket „*ggplot2*“, das besonders viele Gestaltungsmöglichkeiten einräumt. Der Aufbau als „Kochbuch“ führt allerdings dazu, dass es bisweilen etwas aufwändiger ist, die gesuchte Information auch zu finden. Neben einer gedruckten Ausgabe auch online im Open Access: <https://r-graphics.org/index.html> [24.1.2020].

- Wickham, Hadley & Grolemund, Garrett (2018). *R für Data Science: Daten importieren, bereinigen, umformen, modellieren und visualisieren*. Übersetzt von Frank Langenau. Heidelberg: O'Reilly.

Nicht die statistische Analyse steht im Vordergrund dieses Buches, sondern das Datenmanagement. Das Thema ist vor allem wichtig, wenn die Daten in komplizierten Strukturen vorliegen oder sie weitgehend unstrukturiert sind, d. h. eine Auswertbarkeit überhaupt erst geschaffen werden muss. Der ganze Themenbereich Data Mining/Text Mining fällt in diese Sphäre. Wenn Probleme dieser Art bestehen, ist Wickham & Grollemund der Klassiker zum Thema, was auch an der Tatsache einer (guten) Übersetzung des englischen Originals deutlich wird.

- Muenchen, Robert A. (2011). *R for SAS and SPSS Users*. Second edition. New York: Springer.

Das umfangreiche Handbuch tut genau das, was sein Titel sagt: im Umgang mit SAS oder SPSS erfahrenen Benutzern wird der Umstieg auf **R** unter Anknüpfung an ihr vorhandenes Wissen erklärt. Für diese Zielgruppe perfekt. Stellt eingangs auch diverse grafische Benutzeroberflächen vor, der Text insgesamt zielt aber auf Nutzer, die mit der Kommandozeile arbeiten, d. h. **R** plus RStudio.

- Hatzinger, Reinhold, Hornik, Kurt, Nagel, Herbert & Maier, Marco J. (2014). *R. Einführung in die angewandte Statistik*. (2. akt. Aufl.). Hallbergmoos: Pearson.

Deutschsprachig, gut geschrieben, eine moderne Einführung in R aus leicht sozialwissenschaftlicher Perspektive. Ein Lehrbuch, das sich gut zum Selbststudium eignet. **R** wird ohne grafische Benutzeroberfläche bedient, d. h. von Anbeginn an via Kommandozeile.

- Lovelace, Robin, Nowosad, Jakub & Muenchow, Jannes (2019). *Geocomputation with R*. Boca Raton: CRC Press, Taylor & Francis. Online im Open Access: <https://geocompr.robinlovelace.net/> [24.1.2020].

Das (hier exemplarisch ausgewählte) Buch nutzt **R** als Werkzeug für GIS-Analysen und zeigt damit Interessierten auf, ob und wie tauglich **R** als GIS-Software ist.

## 27.2 HINWEISE FÜR FREUNDE DES ONLINE-LESENS

### 27.2.1 Web 1.0

Awsome R: <https://awesome-r.com/> [24.1.2020].

Burk, T. & Anton, T. (2018). *R für Psychos*. Stand 17.11.2018. <https://r-intro.tadaa-data.de/book/> [24.1.2020].

Datascience+ An online community for showcasing R & Python tutorials. <https://datascienceplus.com/> [24.1.2020]. – Zahlreiche Anwendungsbeispiele mit Code in rohem R, meist sehr konkret und hilfreich.

Galili, Tal. (10.12.2015). Tutorials for learning R. *R-Bloggers*, 10. Dec. 2015: <https://www.r-bloggers.com/how-to-learn-r-2/> [24.1.2020].

Gosh, A. (25.4.2018). *Statistik-Software: R, Python, SAS, SPSS und STATA im Vergleich*. INWT Statistics: <https://www.inwt-statistics.de/blog-artikel-lesen/Statistik-Software-R SAS SPSS STATA im Vergleich.html> [24.1.2020].

Machlis, Sh. (11.7.2018). Top R language resources to improve your data skills. *Computerworld*, 11.7.2018: <https://www.computerworld.com/article/2497464/top-r-language-resources-to-improve-your-data-skills.html> [24.1.2020].

Navarro, D. J. (2019). *Learning statistics with R: a tutorial for psychology students and other beginners*. (Version 0.6). <https://learningstatisticswithr.com/> [24.1.2020].

R-bloggers (o.J.). *R news and tutorials by (750) R bloggers*: <https://www.r-bloggers.com/> [24.1.2020].

Rodrigues, Bruno (2019). *Modern R with the tidyverse*. Github, 17.8.2019. [https://b-rodrigues.github.io/modern\\_R/](https://b-rodrigues.github.io/modern_R/) [24.1.2020]. – Eine (gute!) Einführung in R, die auf's Coden setzt und das „Tidyverse“ lehrt, d. h. ein Bündel von R-Paketen, das insbes. auf dem Feld des Datenmanagements gegenüber den traditionellen Verfahren von **R** vorteilhaft sein soll.

RStudio Cheat Sheets: <https://www.rstudio.com/resources/cheatsheets/> [24.1.2020].

Stöcklin, M. & Brühlmann, F. (2019). *R-Toolbox*. <https://mmi.psych.unibas.ch/r-toolbox/index.html> [24.1.2020]. – Gute, auf's Wichtige reduzierte Einführung in das Themenfeld Statistik & R in deutscher Sprache.

SWIRL: <http://swirlstats.com/students.html> [24.1.2020].

Universität Hamburg (2017). *RLab – Skriptbasierte modulare Umweltstatistik*: <http://rlab.blogs.uni-hamburg.de/> [24.1.2020].

Williams, Lisa (2018). *RYouWithMe*. <https://rliessydney.org/ryouwithme> [24.1.2020]. (Ein guter Online-Kurs zu R mit RStudio als Folge von Video-Tutorials).

-----

Eine schöne Übersicht zu weiteren Materialien bietet: Northrop, Paul (2019). *CRAN Task View: Teaching Statistics*. CRAN, 14.3.2019: <https://CRAN.R-project.org/view=TeachingStatistics> [24.1.2020].

Eine gute Zusammenstellung von frei und online verfügbaren Lehrbüchern und Lehrmaterialien zu R samt kurzer Charakterisierung findet sich in einem Blog von David Smale:

Smale, D. (2019). *Free R reading material*. @committedtotape: <https://committedtotape.shinyapps.io/freeR/> [24.1.2020].

### 27.2.2 Twitter

#Rstats ist ein oft verwendeter Hashtag, @RLangTip = „One R Tip a Day“ ein interessanter Twitterer.

@Rbloggers ist der Twitter-Account des oben aufgeführten Blogs „R-bloggers“, der nahezu täglich auf wichtige Neueinträge in diesem Blog aufmerksam macht. Ein Muss, wenn man zum Thema **R** und neue Pakete „auf dem Laufenden“ bleiben will.

### 27.3 FOREN

Eine gute Möglichkeit, Hilfe und Anregungen zu finden, sind Foren. Dort stellen (angemeldete) Nutzer Fragen und andere erfahrene Nutzer geben Antworten. Man muss sich jedoch in der Regel weder anmelden noch selbst Fragen stellen, da fast alle Fragen schon irgendwann gestellt und beantwortet wurden. Der Nutzen der Foren ergibt sich daher vor allem dadurch, dass man in dem Pool bestehender Fragen und Antworten sucht. Im Milieu **R** sind zwei Foren besonders beliebt und reich: Stack Overflow: <https://stackoverflow.com/questions> [24.1.2020], wo man ohne Anmeldung suchen kann, und Cross Validated: <https://stats.stackexchange.com/> [24.1.2020], wo es vor allem um Statistik und Datenvisualisierung geht – weit über **R** hinaus. In beiden Foren sollte man möglichst spezifische Stichworte eingeben, um dem eigenen Problem oder Interessensbereich möglichst nahe zu kommen. In beiden Foren werden Lösungen oft als R-Code gegeben, d. h. man kann Beispiele und Lösungen vielfach mit Copy-&-Paste direkt nutzen resp. an das eigene Problem anpassen.

### 27.4 PERSÖNLICH, VON MENSCH ZU MENSCH

„AG Computeranwendungen und Quantitative Methoden in der Archäologie“ (kurz: AG CAA). 1981 unter dem Namen „Arbeitsgemeinschaft Quantitative Methoden in der Archäologie“ gegründet, bis heute die jährliche Zusammenkunft von einschlägig tätigen Archäologen. Muss man kennen. Die Teilnahme an einer Tagung der AG CAA sollte man zumindest ernsthaft erwägen. Näheres: <http://www.ag-caa.de/> [24.1.2020].

In meiner Stadt gibt es eine von vielen regionalen „R Users Group“ (ggf. so als Stichwort für die Suche): <https://jumpingrivers.github.io/meetingsR/r-user-groups.html> [24.1.2020]. Die von Zeit zu Zeit stattfindenden Treffen führen zu fruchtbaren Begegnungen mit Nicht-Archäologen: empfehlenswert.

## 27.5 NACHSCHLAGEWERKE STATISTIK SOWIE ARCHÄOLOGIE & STATISTIK

Auch hier werden wieder selektiv Titel genannt und eine Lese-Reihenfolge nahegelegt.

- Ihm, Peter, Lüning, Jens & Zimmermann, Andreas (1978). *Statistik in der Archäologie. Probleme der Anwendung, allgemeine Methoden, Seriation und Klassifikation.* (Archaeo-Physika 9). Köln: Rheinland-Verlag.

Trotz des Erscheinungsjahrs und des inzwischen altbacken wirkenden Äußeren: nützlich und lesenswert. Die besondere Stärke dieses umfangreichen Buches liegt im Brücken-Schlagen zwischen Statistik und Archäologie. Ihm (u. a., 1978) geht auf viele Themen ein, die in den üblichen Einführungswerken fehlen, jedoch gerade in der Archäologie wichtig sind. Meist werden wichtige Themen in der Sprache der mathematischen Formeln angeboten, doch anschließend wird Schritt für Schritt vorgerechnet, wie diese Formeln umzusetzen sind. Das Buch sollte man mit einem Taschenrechner oder einer Tabellenkalkulation wirklich durcharbeiten.

- Bortz, Jürgen & Schuster, Christof (2010). *Statistik für Human- und Sozialwissenschaftler.* 7., vollst. überarbeitete u. erweiterte Aufl. Berlin: Springer.

Ein bewährtes Nachschlagewerk in deutscher Sprache für die allgemeine Statistik. Eignet sich auch als Referenzwerk.

- Bortz, Jürgen, Lienert, Gustav A. & Boehnke, Klaus (2008). *Verteilungsfreie Methoden in der Biostatistik.* 3., korr. Aufl. Berlin: Springer.

Sehr umfassendes Nachschlagewerk für das weite Feld der in der Archäologie so wichtigen nicht-parametrischen Verfahren. Für dieses Themenfeld im deutschen Sprachraum ein klassisches Referenzwerk.

- Sachs, Lothar & Hedderich, Jürgen (2018). *Angewandte Statistik. Methodensammlung mit R.* 16., überarbeitete und erweiterte Auflage. Berlin: Springer. – (benutzt wurde hier die 12. Aufl., 2006).

„Der Sachs“ resp. die „Statistischen Auswertungsmethoden“ von Lothar Sachs (1968, 1969, 1972) sind seit langem ein bewährtes deutschsprachiges Nachschlagewerk zur Statistik. Es ist nunmehr – eher sparsam – um Hinweise zu **R** ergänzt

worden. Das Buch ist kein R-Lehrbuch, vielmehr ein (sehr) formel- und detailreiches, nicht immer übersichtliches Lehrbuch der angewandten Statistik. Wer von dort kommend den Weg zu einer praktischen Umsetzung eines Verfahrens mit **R** sucht und wiederum Hilfe beim Lesen und Verstehen der Ausgaben von **R** benötigt, ist hier richtig.

- Field, Andy; Miles, Jeremy & Field, Zoë (2012). *Discovering statistics using R*. London: Sage.

Andy Field hat sich einen Namen gemacht mit seinem unkonventionellen Buch *“Discovering statistics using SPSS for Windows: advanced techniques for the beginner“* (2002, 5. Aufl. 2018). In ihm steht die Statistik im Vordergrund, SPSS ist das Werkzeug, mit dem die Analysen durchgeführt werden. Field schreibt und erklärt anders als klassische Lehrbücher, sein lockerer Schreibstil ist Grundlage des Erfolgs seines Buches. Sein hier angeführtes Buch zu **R** ähnelt seinem Statistikbuch mit SPSS, benutzt aber eben **R** als Werkzeug. Im Vordergrund seiner Einführung steht jedoch nicht das Werkzeug, sondern die angewandte Statistik. Trotz seiner guten Lesbarkeit, hohen Qualität und des beträchtlichen Volumens: Für manches Archäologierelevante fehlt es bisweilen an Tiefe.

- McCall, Grant S. (2018). *Strategies for quantitative research. Archaeology by numbers*. Oxon: Routledge.

Eines der wenigen guten und modernen Werke zum Thema Archäologie & Statistik. Allerdings fokussiert McCall auf den Bereich der Theorie, alle Fragen zur praktischen Umsetzung mit Hilfe von Software klammert er aus. Eine Rezension findet sich in: *Archäologische Informationen* 42 (2019).

- Monka, Michael, Voß, Werner & Schöneck, Nadine M. (2008). *Statistik am PC: Lösungen mit Excel*. (5. Aufl.). München: Hanser.

Wem das alles zu viel und zu kompliziert ist mit **R** u. a. und weiß, dass er nie über das Potenzial einer Tabellenkalkulation hinauswachsen möchte, ist hier genau richtig. Das Buch erklärt Statistik verständlich (mit tendenziell sozialwissenschaftlicher Ausrichtung) und schöpft alle diesbezüglichen Möglichkeiten von Excel aus. Da LO-Calc in vielem sehr ähnlich funktioniert, passt das Buch m. E. auch für Anwender von LibreOffice.

Abschließend noch zwei wertvolle englischsprachige Nachschlage- und Referenzwerke für die Fraktion „nur die Harten komm‘n in den Garten“:

- Sokal, Robert R. & Rohlf, F. James (2012). *Biometry: The principles and practice of statistics in biological research*. 4<sup>th</sup> ed. New York: W. H. Freeman.
- Legendre, Pierre & Legendre, Louis (2012). *Numerical Ecology*. 3<sup>rd</sup> English Edition. Amsterdam: Elsevier.

Ausführliches, formelstarkes Lehrbuch, nicht zu verwechseln mit dem dünnen Auszug „Numerical ecology with R“, der ohne das hier gemeinte Werk wenig hilfreich ist.

## 27.6 LITERATUR ZUM THEMA GRAFISCHE DARSTELLUNGEN

- Riedwyl, Hans (1987). *Graphische Gestaltung von Zahlenmaterial*. 3. Auflage. (UTB 440). Bern: Haupt.

Der Titel ist inzwischen vergriffen, aber antiquarisch günstig zu erhalten. Auch wenn das schlanke Taschenbuch inzwischen etwas vergilbt daherkommen mag: die m. E. beste Einführung in das Thema mit vielen guten Empfehlungen und Beispielen.

- Wilke, Claus O. (2019). *Fundamentals of Data Visualization*. Printausgabe bei O'Reilley (2019), Pre-Print vollständig online: <https://serialmentor.com/dataviz/index.html> [24.1.2020] resp. <https://github.com/clauswilke/dataviz> [24.1.2020].

Sehr gründlich, umfassend und modern: eine Leseempfehlung. Arbeitet auf Basis von *ggplot2* mit vielen Code-Beispielen. Kann auch als Einführung in *ggplot2* genutzt werden.

- Burchell, Jodie & Vargas, Mauricio (2019). *The hitchhiker's guide to ggplot2. Don't panic and create beautiful plots with R*. Victoria: Leanpub. <https://leanpub.com/hitchhikers-ggplot2> [24.1.2020].

Eine gut verständliche, entlang eines klaren roten Fadens verfasste Schritt-für-Schritt Anleitung in die Praxis mit *ggplot2*: empfehlenswert.

- BBC Data Team (24.1.2019). *BBC Visual and Data Journalism cookbook for R graphics. How to create BBC style graphics*. (GitHub, rcookbook): <https://bbc.github.io/rcookbook/#how-to-create-bbc-style-graphics> [24.1.2020].

Der Titel spricht für sich: das Thema Grafik aus der Sicht aktueller Datenjournalisten, mit Beispielen samt deren Umsetzung via R-Code mit *ggplot2*. Eindeutig ein „*must see*“.

- Tufte, Edward Rolf (2001). *Visual Display of Quantitative Information*. (2<sup>nd</sup> ed.; 1<sup>st</sup> ed. 1983). Cheshire, Conn.: Graphics Press.

Ein immer wieder zitierter Klassiker, an dem nicht vorbeizukommen ist. Die hier genannte zweite Auflage existiert bis in die jüngsten Jahre in vielen unveränderten Nachdrucken, es ist jedoch stets das Buch von 2001. Einen guten ersten Eindruck – auch von der Geschäftstüchtigkeit des Autors – erhält man auf seiner Website: <https://www.edwardtufte.com/tufte/> [24.1.2020]. Trotz der weiten Verbreitung und Anerkennung “des Tufte“ vor allem im US-amerikanischen Raum bleiben Riedwyl 1987 und Wilke 2019 meine erste Wahl.

## 28 NACHWEIS DER ZITIERTEN LITERATUR

---

AMRHEIN, VALENTIN, GREENLAND, SANDER & MCSHANE, BLAKE (2019). Nature comment: Scientists rise up against statistical significance. *Nature* 567, 305-307 (20.3.2019): <https://www.nature.com/articles/d41586-019-00857-9> [24.1.2020].

BOCK, HANS HERMANN (1974). Automatische Klassifikation. Theoretische und praktische Methoden zur Gruppierung und Strukturierung von Daten (Cluster-Analyse). Göttingen: Vandenhoeck & Ruprecht.

BORTZ, JÜRGEN & SCHUSTER, CHRISTOF (2010). *Statistik für Human- und Sozialwissenschaftler*. 7. Vollständig überarbeitete und erweiterte Auflage. Berlin: Springer.

BUCK, CAITLIN E., CAVANAGH, WILLIAM & LITTON, CLIFFORD D. (1996). Bayesian approach to interpreting archaeological data. Chichester: Wiley.

DRENNAN, ROBERT D. (2004; 2009). *Statistics for Archaeologists. A common sense approach*. 2<sup>nd</sup> ed. New York: Springer.

GEBÜHR, MICHAEL (1983). Archäologischer und anthropologischer Befund der Beigaben und Leichenbrände aus Westholstein. In H. Hingst, *Die vorrömische Eisenzeit Westholsteins* (p. 183-192). Neumünster: Wachholtz.

GOLDMANN, KLAUS (1972). Zwei Methoden chronologischer Gruppierung. *Acta Praehistorica et Archaeologica* 3, p. 1-34.

HAIR, JOSEPH F. JUN., BLACK, WILLIAM C., BABIN, BARRY J. & ANDERSON, ROLPH E. (2010). *Multivariate data analysis*. 7<sup>th</sup> edition. Upper Saddle River, New Jersey: Prentice Hall.

DRENNAN, ROBERT D. (2010). *Statistics for archaeologists. A common sense approach*. 2<sup>nd</sup> edition. New York: Springer.

GILL, JEFF (2015). *Bayesian Methods: a social and behavioral sciences approach*. Third edition. Boca Raton: Chapman & Hall.

HAMMER, ØVIND (2020). PAST Paleontological STatistics Version 3.4 Reference manual. Oslo.

- IHM, PETER (1980). Poisson- und Negative Binomialverteilung. *Archaeo-Physika* 7. (p. 95-106). Bonn: Rheinland-Verlag.
- IHM, PETER (1983). Korrespondenzanalyse und Seriation. *Archäologische Informationen* 6, 8-21.
- KANDA, Y. (2013). Technical Report: Investigation of the freely available easy-to-use software ‚EZR‘ for medical statistics. *Bone Marrow Transplantation* 48, 452–458. doi:10.1038/bmt.2012.244 <https://www.nature.com/articles/bmt2012244.pdf> [24.1.2020].
- KRUSCHKE, JOHN K. (2015). *Doing Bayesian data analysis. A tutorial with R, JAGS and Stan*. Second edition. Amsterdam: Elsevier.
- LEGENDRE, PIERRE & LEGENDRE, LOUIS (2012). *Numerical ecology*. 3rd edition (Developments in Environmental Modelling 24). Amsterdam: Elsevier.
- MÜLLER, JOHANNES & ZIMMERMANN, ANDREAS (Hrsg.) (1997). *Archäologie und Korrespondenzanalyse: Beispiele, Fragen, Perspektiven*. (Internationale Archäologie, 23). Espelkamp: Leidorf.
- NEUFFER, EDUARD M. (1965). Eine statistische Bearbeitung von Kollektivfunden. *Bonner Jahrbücher* 165, 28-56.
- ORTON, CLIVE (1980). *Mathematics in Archaeology*. London: Collins. – Zweite Auflage 1982: Cambridge: Cambridge Univ. Press.
- SACHS, LOTHAR & HEDDERICH, JÜRGEN (2006). *Angewandte Statistik. Methodensammlung mit R*. Zwölfte, vollständig neu bearbeitete Auflage. Berlin: Springer.
- SHENNAN, STEPHEN (1988). *Quantifying Archaeology*. Edinburgh: University Press.
- SIEGEL, SIDNEY (1987). *Nicht-parametrische statistische Methoden*. 3. Auflage. (Methoden in der Psychologie 4). Eschborn: Verlag Fachbuchhandlung für Psychologie.
- SIEGMUND, FRANK (2009). Der Maskulinitätsindex und statistische Verfahren zur Prüfung auf Unterschiede in den Geschlechteranteilen frühmittelalterlicher Populationen (Masculinity index and the statistical improvement of a balanced or unbalanced secondary sex ratio in early medieval Europe). *Bulletin der Schweizerischen Gesellschaft für Anthropologie* 15, 5-17.
- SIEGMUND, FRANK (2010). *Die Körpergröße der Menschen in der Ur- und Frühgeschichte Mitteleuropas und ein Vergleich ihrer anthropologischen Schätzmethoden*. Norderstedt: BoD.
- SIEGMUND, FRANK (2015). *Genußt wie: Praxisleitfaden Seriation und Korrespondenzanalyse in der Archäologie*. Norderstedt: BoD. – Verfügbar auch in englischer und italienischer Sprache: *How to perform a correspondence analysis. A short guide to archaeological practice*. Charleston: Amazon Create Space. // Come realizzare un'analisi delle corrispondenze: guida breve per archeologi. In J. Pinar Gil (ed.) (2017), *'Small finds' e Cronologia (V-IX sec.)*. *Esempi, metodi e risultati*. (p. 31-70). Roma: BraDypUS.
- SOKAL, ROBERT R. & ROHLF, F. JAMES (1995). *Biometry. The principles and practice of statistics in biological research*. 3rd edition. New York: Freeman.
- TOUTENBURG, HELGE & HEUMANN, CHRISTIAN (2008). *Deskriptive Statistik. Eine Einführung in Methoden und Anwendungen mit R und SPSS*. 6. Aufl. Berlin: Springer.

ÜBERLA, KARL (1968). *Faktorenanalyse. Eine systematische Einführung für Psychologen, Mediziner, Wirtschafts- und Sozialwissenschaftler.* (2. Aufl. 1977). Berlin: deGruyter.

WASSERSTEIN, RONALD L. & LAZAR, NICOLE A. (2016). The ASA's Statement on p-Values: Context, Process, and Purpose. *The American Statistician* 70(2), 129-133. DOI: 10.1080/00031305.2016.1154108

## 29 FINIS

---

Dieses Buch ist als gedruckte Ausgabe beim BoD-Verlag erhältlich. Eine Vorabfassung wird via Academia.edu und über die Website des Autors im Open Access bereitgestellt, wo sich auch die Beispieldatensätze und einige Skripte finden. Es ist beabsichtigt, unvermeidlichen Fehler oder wichtigen Neuerungen von Zeit zu Zeit durch eine Überarbeitung nachzukommen.

## 30 STICHWORTVERZEICHNIS

---

---

### 3

3D-Effekt · 148

---

### 5

5 %-trimmed mean · 226

5,0e-02 · 246

---

### A

$a_{20}$  · 333

AAT · 340, 345

abhängige Stichprobe · 253

abhängige Variable · 275

Abhängigkeiten · 136

abline · 316

Abrunden · 187

Abweichungsquadrat · 199, 225

accuracy · 296

Richtigkeit · 296

AG CAA · 36

AIC · 286

alpha · 202

Alternativhypothese · 203, 244

Ansteckungsverteilung · 217

Antennen · 228

ARACHNE · 43

Archäologie · 21

arrow · 316

Aufrunden · 187

Ausreißer · 228

---

### B

Babies · 91

Balkendiagramm · 148, 151

barchart · 151

Bayes-Statistik · 27, 41, 265

Beigabewert · 340

Beispieldatensatz · 90

Benutzeroberfläche · 115

Beobachterfehler · 82

beschreibende Statistik · 23, 188

Bestimmtheitsmaß · 275, 278

Big Data · 45

bimodal · 218

Bindung · 267

Binomialverteilung · 290

bins · 162, 230

Bioconductor · 134

bi-variate Normalverteilung · 275

BlueSky · 50, 116, 141, 142, 157

Bodendenkmälerdatenbank · 35

Bonferroni-Ausreißer-Test · 362

Bonner Bodendenkmälerdatenbank · 35

Bootstrap · 263

Bootstrapping · 40, 234, 366

Boxplot · 227

break · 166

---

### C

CAA · 36

Camel case · 122

Cheat sheets · 184

Chi-Quadrat · 200, 202

Chi-Quadrat-Verfahren · 196

Christlein, R. · 339  
Cook-Distanz · 287, 362  
Costedt · 348  
CRAN · 114, 134  
CRAN-Mirror · 135

---

## *D*

Daten sichern · 85  
Datenbank · 111  
Datenbeschreibung · 89  
Datenerfassung · 53, 63, 112  
Datenfehler · 80, 81  
Datenformular · 56  
Datenimport · 137  
Datenprüfung · 79, 112  
Datentypen · 127  
Deducer · 115  
Demographie · 322  
dependencies · 136  
Diskrete Skala · 62  
Diskrete Wahrscheinlichkeitsverteilung ·  
287  
DOP · 296  
Dot Plot · 167  
Dreiecksdiagramm · 180  
dRMS · 301  
Durbin-Watson-Wert · 373

---

## *E*

$e_0$  · 333  
Easy R · 109  
EDA · 39  
Editor · 138  
EGNOS · 300  
einfache Varianzanalyse · 255  
einseitige Fragestellung · 253  
endliche Grundgesamtheit · 193

Epi Info · 59, 110  
Erwartungswert · 198  
ethnicity · 308  
exakter Test n. Fisher · 208  
Explorative Datenanalyse · 39  
EZR · 109

---

## *F*

Faktor · 141, 155, 278  
Farbbestimmung · 62  
fehlende Werte · 54, 139, 155  
Fehlersuche · 79  
Filtern · 67  
Formular · 56  
Freiheitsgrad · 202  
Funktion · 128  
Funktionen in R · 119  
fuzzy · 41

---

## *G*

Gebühr, M. · 37  
Geschlechtsdimorphismus · 233  
gestaffeltes Balkendiagramm · 169  
gestapeltes Balkendiagramm · 170  
ggplot2 · 183  
Gini-Index · 341, 349  
Gleichverteilung · 236, 288  
GPS · 293  
Grabwert · 340, 346  
Grafik · 147  
Grundgesamtheit · 185  
GUI · 115

---

## *H*

$H_0$  · 203  
 $H_1$  · 203

hat value · 362  
Häufigkeit · 185  
Hedeager, L. · 340  
Helmert-Transformation · 304  
Hilfe · 116, 144, 184  
Hilfe zu R · 116  
hinge · 228  
Histogramm · 149, 158  
Howells' world data · 32  
H-Test · 256  
Hundertsumme · 185

---

## *I*

IANUS · 43  
Ihm, P. · 37  
Index, Indices · 252  
Informationsstatistik · 213  
InkScape · 167  
Inter Observer Error · 82  
Interquartilabstand · 224  
Intervallskala · 61  
Intra Observer Error · 82  
IQR · 224  
Irrtumsrisiko · 202

---

## *J*

Jackknifing · 41, 233, 263  
JAMOVI · 50, 102, 103, 104  
JASP · 102, 103  
Jørgensen, L. · 340

---

## *K*

Kendall's  $\tau$  · 268, 273  
KFA · 213  
Kinderanteil · 327  
Klassenweite · 149, 161, 163

Kolmogorov-Smirnov-Anpassungstest ·  
244  
Kolmogorov-Smirnov-Test · 255  
Konfidenzintervall · 187, 188, 191  
Konfigurationsfrequenzanalyse · 213  
Konstante · 274, 278  
Kontingenzkoeffizient · 268  
Kontingenztabelle · 196  
Kopfzeile fixieren · 56  
Körperhöhenschätzung · 353  
Korrektur nach Yates · 208  
Korrelation · 265, 266  
Korrelationskoeffizient · 266, 267  
Korrelationsmatrix · 270  
Korrespondenzanalyse · 27  
Kreisdiagramm · 148, 151  
Kreuztabelle · 196  
kritischer Wert · 201, 202  
Kruskal-Wallis-Test · 262  
KS-Test · 244, 255  
Kulturmodell · 307  
kurtosis · 239

---

## *L*

Label · 87  
lattice · 183  
Lebendbevölkerung · 338  
Lektüreempfehlung · 15, 27, 265, 389,  
395  
Leseempfehlung · 397  
Levene-Test · 254  
LibreOffice Calc · 17  
life table · 330  
Lilliefors-Test · 244  
Lineare Regression · 274  
linkssteil · 238  
LO-Calc · 51, 65  
logarithmische Transformation · 132, 249  
logische Funktionen · 118

logistische Regression · 284

---

## **M**

Gebühr · 339  
Maske · 112  
Maskulinitätsindex · 325  
Maßkorrelationskoeffizient · 267  
mathematische Funktionen · 120  
Matrix · 126  
Maximum · 223  
Median · 221, 226  
Mehrfeldertafel · 196, 211  
Mindestindividuenzahl · 215  
Minimum · 223  
missing value · 54, 155  
Mittelwert · 221, 222  
Modalwert · 221  
Multinomialverteilung · 217  
multiple lineare Regression · 371  
Multiple Regression · 280  
Munsell Color Charts · 62

---

## **N**

NA · 54, 139, 155  
Namen · 87  
Namengebung · 122  
NAT · 340  
Negative Binomialverteilung · 291  
Netzausgleich · 304  
new archaeology · 34  
NFDI4objects · 44  
nicht-lineare Regression · 279  
Nicht-lineare Regression · 274  
nicht-parametrische Verfahren · 252  
Nominalskala · 61, 148  
Normabstand · 310  
Normalverteilung · 236, 237, 243

Notepad++ · 138  
Nullhypothese · 197, 203, 244

---

## **O**

observer error · 82  
OLS-Regression · 279  
Ordinalskala · 61

---

## **P**

Paket beziehen · 134  
Paket installieren · 135  
Pakete · 134  
parametrische Statistik · 250  
parametrische Verfahren · 38  
PAST · 49, 70, 96, 161  
pie chart · 151  
Pilotstudie · 85  
Pivot-Tabelle · 74, 76  
Plot für arithmetische Mittel · 258  
plot of means · 258  
Poissonverteilung · 217, 289  
postprozessuale Archäologie · 45  
Präzision · 297  
Präzision · 296  
Produkt-Moment-Korrelationskoeffizient  
· 267  
Prozentzahl · 186, 251  
Prozentzahlen runden · 186  
PSPP · 111  
p-Wert · 206, 246

---

## **Q**

QQ-Plot · 248  
Quantil · 226  
Quartil · 227

---

## R

R · 49, 113, 117  
R Console · 130, 131  
 $r^2$  · 278  
Randsumme · 196  
Rangkorrelationskoeffizient · 268  
Rangskala · 61  
Rangvarianzanalyse · 256  
R-bloggers · 116  
R-Commander · 50, 115, 136, 163  
    Abstürze · 140  
Rechnen mit R · 118  
rechteckige Datenmatrix · 112  
rechtssteil · 238  
Reduced Major Axis Regression · 280  
Referenserie · 356  
Regression · 265, 273  
regression modell II · 280  
Regressionsgerade · 274  
Reliabilitätskoeffizient R · 84  
Resampling · 234  
Residuen · 278  
Residuum · 277  
Richtigkeit · 298  
RMS · 298  
Rollet, E. (1888) · 94, 353  
RStudio · 114, 129  
rTEM · 83, 84  
Rücktransformation · 120  
Runden · 187  
Rundungsregeln · 187

---

## S

Sampling · 234, 263  
SAP · 34  
SAS · 39

Sax · 311  
scatterplot · 174, 178  
Schädelmaße · 92  
Schätzung der empirischen  
    Dichtefunktion · 149  
Scheinkorrelation · 276  
Schiefe · 239  
schließende Statistik · 188  
Schließende Statistik · 24  
Schwellenwert · 201  
Script · 166  
sd · 225  
SEE · 276, 355  
Seriation · 27  
Shapiro-Wilk-Test · 247  
signifikant · 206, 269  
Skalenniveau · 60, 185  
skewness · 239  
Snake case · 122  
SOFA · 108  
Sortieren · 67  
Sozialstrukturen · 339  
Spaltensumme · 196  
Spannweite · 223  
Spearman's  $\rho$  · 268  
Spickzettel · 184  
SPSS · 39, 110  
Stadler, P. · 37  
Standardabweichung · 225, 238  
standardisierte Normalverteilung · 238  
Stapeldiagramm · 170  
Statistik · 23  
statistische Grafik · 147  
Std.abw. · 225  
std.dev. · 225  
Steigung · 274, 278  
Sterbetafel · 330  
Stetige Skala · 62  
Streuungsdiagramm · 174, 178  
studentisierte Residuen · 361

subset · 166  
Swirl · 116, 146

---

## T

TEM · 83  
Ternary plot · 180  
Testserie · 356, 369  
Tortendiagramm · 148  
Transformation · 120  
t-Test · 254  
Typologie · 311  
Typwert · 349

---

## U

Umkehrfunktion · 120  
unabhängige Stichprobe · 253  
unimodal · 218  
U-Test · 255

---

## V

Variablenamen · 87  
Variablentransformation · 120, 249  
Varianz · 226  
Varianzanalyse · 255  
Vektor · 125  
Verhältnisskala · 61  
Verhältniszahl · 252  
Vierfeldertafel · 196

---

## W

WAAS · 300  
Wahrscheinlichkeitsverteilung, stetig · 235  
Werlaburgdorf · 328  
whisker · 228  
Winkeltransformation · 250  
*wiss. Notation* · 246  
wissenschaftlichen Zahlennotation · 246  
Wölbung · 239  
workspace · 124  
Wurzeltransformation · 133, 249

---

## X

**x-Achse** · 18

---

## Y

**y-Achse** · 18  
Yule-Koeffizient · 268

---

## Z

Zeilensumme · 196  
Zielsetzung · 16, 18, 47  
Zimmermann, A. · 36  
z-Transformation · 238  
zweiseitige Fragestellung · 253  
Zwentendorf · 91



8							
7							
6							
5							
4							
3							
2							
A 1	B	C	D	E	F	G	H

**Abb. 125** Leeres Gitternetz mit 8 x 8 Feldern ("Schachbrett").