

Björn Walther

# Statistik mit R




## Schnelleinstieg

R einfach lernen in 14 Tagen

Mit  
praktischer  
Nachschlage-  
hilfe



# Inhalt

	<b>Nachschlagehilfe .....</b>	<b>13</b>
	<b>Einleitung .....</b>	<b>17</b>
	E.1 R lernen in 14 Tagen .....	17
	E.2 Der Aufbau des Buches .....	17
	E.3 Downloads zum Buch .....	18
	E.4 Fragen und Feedback .....	18
<b>Teil I</b>	<b>Einführung in die Arbeit mit R und RStudio .....</b>	<b>19</b>
	<b>1 Warum gerade R für statistische Analysen? .....</b>	<b>21</b>
	<b>2 R-Grundlagen in Kurzform .....</b>	<b>23</b>
	2.1 Syntax .....	23
	2.2 Objekttypen in R .....	24
	2.3 R-Pakete finden und verwenden .....	25
	2.3.1 Pakete installieren und laden .....	25
	2.3.2 Finden von Paketen .....	26
	2.4 Datenformate in R .....	28
	2.4.1 Wide-Format .....	28
	2.4.2 Long-Format .....	29
	2.4.3 Transformation der Formate .....	30
	2.5 Pipe-Operatoren .....	31
	<b>3 RStudio als hilfreiche Oberfläche .....</b>	<b>33</b>
	3.1 Layout von RStudio .....	33
	3.2 Empfohlene Einstellungen .....	35
	3.2.1 Dark Mode .....	35
	3.2.2 Tastatur-Shortcuts .....	36
	3.2.3 In Projekten arbeiten .....	36

**Teil II    Datenmanagement und deskriptive Statistiken ..... 39**



<b>Datenmanagement in R .....</b>	<b>41</b>
4.1    Datensätze in R einlesen .....	41
4.1.1    Nutzen des Importassistenten .....	41
4.1.2    Import über Code .....	43
4.2    Datensätze zusammenfügen .....	46
4.2.1    Fälle hinzufügen .....	46
4.2.2    Variablen hinzufügen .....	47
4.3    Teildatensätze erstellen .....	49
4.3.1    Auswahl bestimmter Variablen .....	49
4.3.2    Auswahl bestimmter Fälle .....	50
4.3.3    Auswahl bestimmter Fälle und Variablen .....	50
4.4    Datensätze exportieren .....	51
4.4.1    CSV- und TXT-Export .....	51
4.4.2    XLSX-Export .....	52
4.4.3    SAV-Export (SPSS) und DTA-Export (STATA) .....	52
4.5    Datensätze speichern und wieder laden .....	52
4.6    Fehlende Werte ausschließen .....	53
4.7    Variablen faktorisieren .....	53
4.8    Datumsvariablen als Datum formatieren .....	54
4.9    Dummycodierung von kategorialen Variablen .....	55
4.9.1    Das Prinzip einer Dummycodierung .....	55
4.9.2    Dummycodierung in R .....	56
4.10    Skalenbildung .....	56
4.10.1    Zweck einer Skalenbildung .....	56
4.10.2    Interne Konsistenz .....	57
4.10.3    Inverscodierung von Items .....	59
4.10.4    Skalenbildung .....	59



<b>Deskriptive Statistik von Stichproben .....</b>	<b>61</b>
5.1    Häufigkeiten .....	61
5.1.1    Absolute Häufigkeiten .....	61
5.1.2    Relative Häufigkeiten .....	62
5.1.3    Kumulierte relative Häufigkeiten .....	63
5.1.4    Übersichtstabelle .....	64

5.2	Lageparameter .....	65
5.3	Streuparameter .....	68
5.4	Schiefe und Kurtosis .....	70
5.5	Überblicksfunktionen für die deskriptive Statistik in R .....	71
5.5.1	Überblick mit describe() .....	71
5.5.2	Überblick mit Desc() .....	72
5.6	Deskriptive Statistiken für Untergruppen .....	73
5.6.1	Nutzen von tapply() .....	73
5.6.2	Nutzen von describeBy() .....	74
5.6.3	Nutzen des Pipe-Operators .....	75
5.7	Zusammenhänge .....	76
5.7.1	Kreuztabellen .....	76
5.7.2	Korrelation .....	77
<b>Teil III</b>	<b>Diagramme .....</b>	<b>79</b>



<b>Allgemeine Darstellungen von Verteilungen für eine oder mehrere Gruppen .....</b>	<b>81</b>
6.1 Histogramm .....	82
6.1.1 Histogramm mit der Basisversion von R .....	82
6.1.2 Einfaches Histogramm mit ggplot2 .....	85
6.1.3 Histogramm für Gruppen mit ggplot2 .....	89
6.2 Säulendiagramm .....	90
6.2.1 Säulendiagramm mit der Basisversion von R .....	90
6.2.2 Einfaches Säulendiagramm mit ggplot2 .....	93
6.2.3 Säulendiagramm für Gruppen mit ggplot2 .....	94
6.3 Balkendiagramm .....	95
6.3.1 Balkendiagramm mit der Basisversion von R .....	95
6.3.2 Balkendiagramm mit ggplot2 .....	96
6.4 Boxplot .....	98
6.4.1 Boxplot mit der Basisversion von R .....	98
6.4.2 Boxplot mit ggplot2 .....	100
6.5 Kreisdiagramm .....	103
6.6 Q-Q-Plot .....	104



<b>Veränderungen in Diagrammen darstellen .....</b>	<b>107</b>
7.1 Diagramme mit der Basisversion von R .....	108
7.1.1 Liniendiagramm für eine Variable .....	108
7.1.2 Liniendiagramm für zwei oder mehr Variablen .....	112
7.2 Diagramme mit ggplot2 .....	114
7.2.1 Liniendiagramm für eine Variable .....	114
7.2.2 Liniendiagramm für zwei oder mehr Variablen .....	116
7.2.3 Gestapeltes Flächendiagramm .....	119
7.2.4 Boxplots .....	121
7.2.5 Säulendiagramm mit Fehlerbalken .....	122
7.2.6 Liniendiagramm mit Fehlerbalken .....	123



<b>Zusammenhänge in Diagrammen darstellen .....</b>	<b>127</b>
8.1 Streudiagramm .....	127
8.1.1 Streudiagramm mit der Basisversion von R .....	127
8.1.2 Streudiagramm mit ggplot2 .....	130
8.2 Korrelationsdiagramm .....	133

<b>Teil IV Analytische Tests .....</b>	<b>137</b>
--	------------

---



<b>Stichprobe mit Population vergleichen – Einstichproben-Tests .....</b>	<b>141</b>
9.1 Einstichproben-t-Test für den Mittelwert .....	142
9.1.1 Voraussetzungen .....	142
9.1.2 Durchführung .....	142
9.1.3 Interpretation der Ergebnisse .....	144
9.1.4 Berechnung der Effektstärke .....	144
9.1.5 Reporting der Ergebnisse .....	145
9.2 Einstichproben-Wilcoxon-Test für den Median .....	146
9.2.1 Voraussetzungen .....	146
9.2.2 Durchführung .....	146
9.2.3 Interpretation der Ergebnisse .....	148
9.2.4 Berechnung der Effektstärke .....	148
9.2.5 Reporting der Ergebnisse .....	149

9.3	Chi <sup>2</sup> -Anpassungstest für die Verteilung .....	149
9.3.1	Voraussetzungen .....	150
9.3.2	Durchführung .....	150
9.3.3	Interpretation der Ergebnisse .....	151
9.3.4	Reporting der Ergebnisse .....	151



<b>Veränderungen zwischen Zeitpunkten nach Intervention prüfen .....</b>	<b>153</b>
10.1 Zwei Zeitpunkte .....	153
10.1.1 t-Test bei abhängigen Stichproben .....	154
10.1.2 Wilcoxon-Test bei abhängigen Stichproben .....	159
10.2 Mehr als zwei Zeitpunkte .....	165
10.2.1 ANOVA mit Messwiederholung .....	165
10.2.2 Friedman-ANOVA .....	174



<b>Unterschiede zwischen Gruppen prüfen .....</b>	<b>181</b>
11.1 Zwei Gruppen zu einem Zeitpunkt mit einem Einflussfaktor .....	181
11.1.1 t-Test bei unabhängigen Stichproben .....	181
11.1.2 Mann-Whitney-U-Test (Mann-Whitney-Wilcoxon-Test) .....	189
11.2 Mehr als zwei Gruppen zu einem Zeitpunkt mit einem Einflussfaktor .....	195
11.2.1 Einfaktorielle ANOVA .....	196
11.2.2 Kruskal-Wallis-Test .....	205



<b>Unterschiede zwischen Gruppen mit mehreren Einflussfaktoren sowie mit Messwiederholung (gemischte Modelle) .....</b>	<b>213</b>
12.1 Mehrere Gruppen infolge mehrerer Einflussfaktoren – Mehrfaktorielle ANOVA .....	213
12.1.1 Voraussetzungen .....	214
12.1.2 Durchführung .....	214
12.1.3 Interpretation der Ergebnisse .....	224
12.1.4 Reporting der Ergebnisse .....	225

12.2	Gemischte ANOVA als Sonderfall .....	226
12.2.1	Voraussetzungen .....	227
12.2.2	Durchführung .....	228
12.2.3	Interpretation der Ergebnisse .....	236
12.2.4	Reporting der Ergebnisse .....	237



<b>13</b>	<b>Ungerichtete Zusammenhänge – Korrelationsanalysen .....</b>	<b>239</b>
13.1	Pearson-Korrelation .....	240
13.1.1	Durchführung .....	241
13.1.2	Ergebnis und Interpretation .....	242
13.1.3	Reporting der Ergebnisse .....	242
13.2	Spearman-Korrelation .....	243
13.2.1	Durchführung .....	243
13.2.2	Ergebnis und Interpretation .....	244
13.2.3	Reporting der Ergebnisse .....	245
13.3	Kendall-Tau-Korrelation .....	245
13.3.1	Durchführung .....	246
13.3.2	Ergebnis und Interpretation .....	247
13.3.3	Reporting der Ergebnisse .....	248
13.4	Pearson-punktbiseriale Korrelation .....	248
13.4.1	Durchführung .....	249
13.4.2	Ergebnis und Interpretation .....	250
13.4.3	Exkurs: Interpretation einer signifikanten Korrelation .....	250
13.4.4	Reporting der Ergebnisse .....	251
13.5	Chi <sup>2</sup> -Test auf Unabhängigkeit .....	251
13.5.1	Durchführung .....	251
13.5.2	Ergebnis und Interpretation .....	253
13.5.3	Reporting der Ergebnisse .....	255
13.6	Kontingenzkoeffizient / Cramer V .....	255
13.7	Odds-Ratio .....	256
13.8	Zusatz: Partialkorrelation .....	257



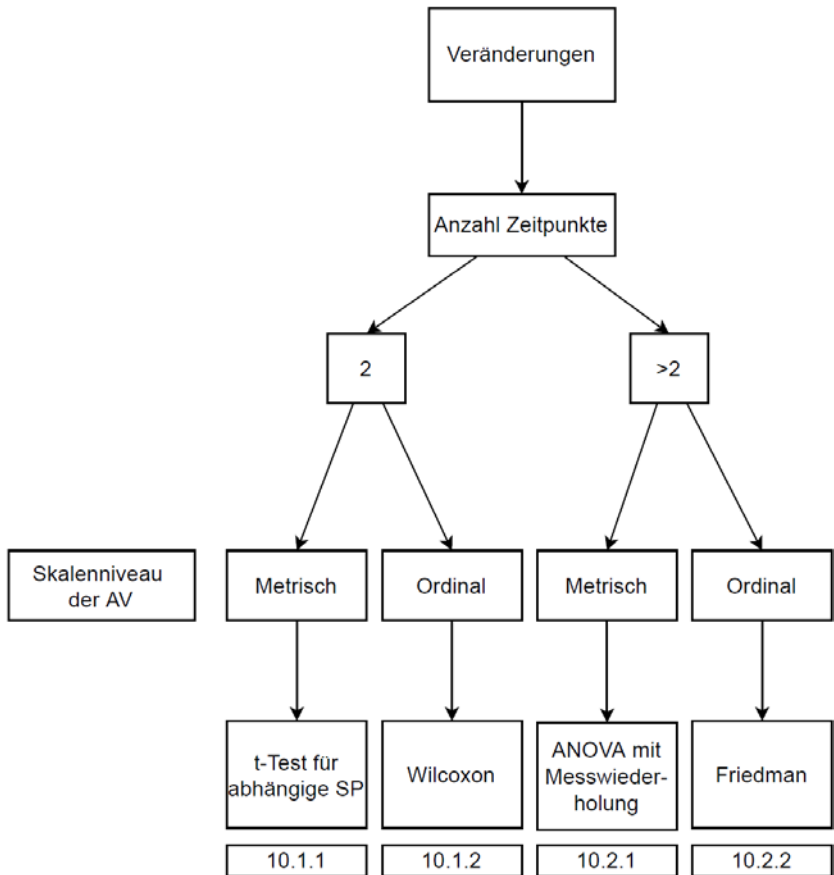
<b>14</b>	<b>Gerichtete Zusammenhänge – Regressionsanalysen .....</b>	<b>259</b>
14.1	Lineare Regression .....	259
14.1.1	Vorbemerkungen und Vorbereitungen .....	260
14.1.2	Voraussetzungen der linearen Regression .....	261

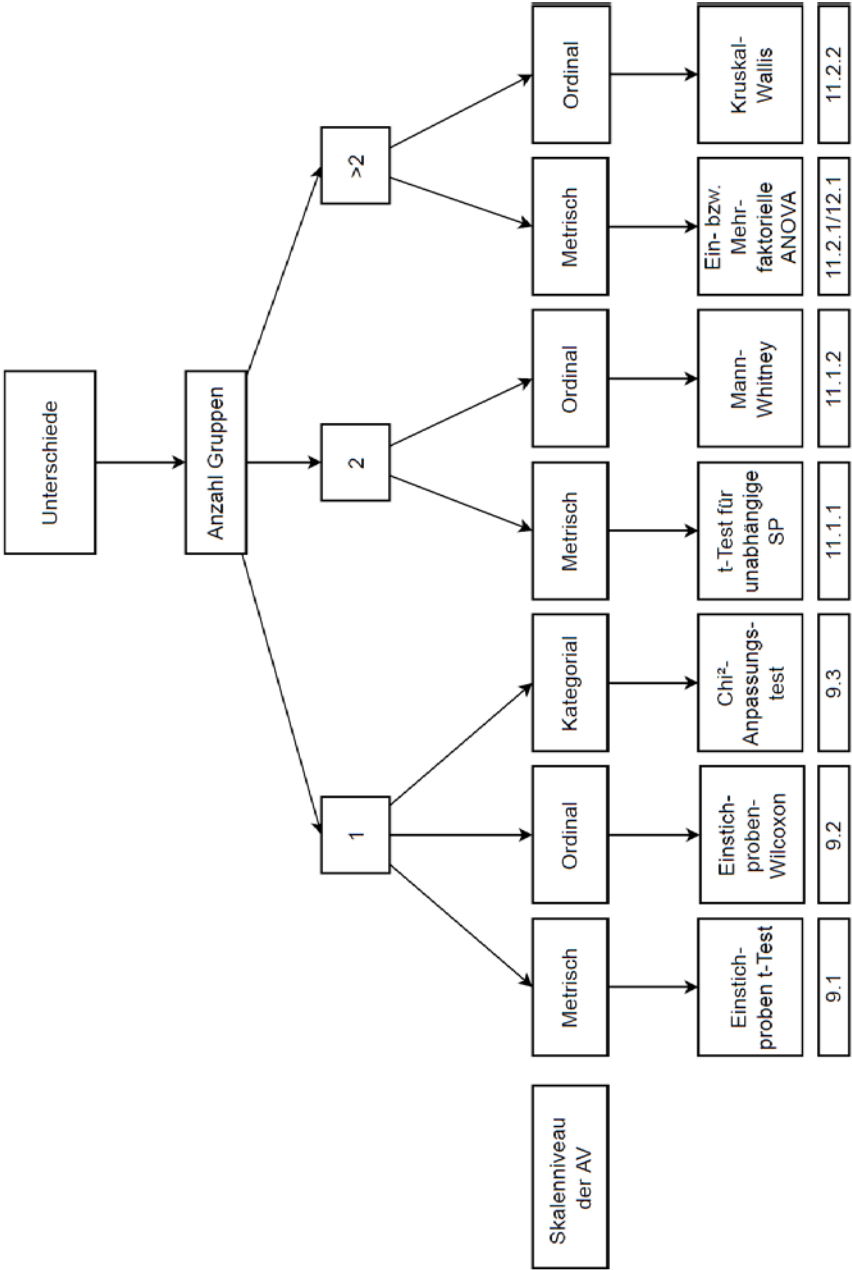
14.1.3	Durchführung .....	262
14.1.4	Ergebnis .....	269
14.1.5	Interpretation der Ergebnisse .....	271
14.1.6	Reporting der Ergebnisse .....	274
14.2	Moderation und Mediation im Rahmen der linearen Regression .....	276
14.2.1	Moderation .....	276
14.2.2	Mediation .....	279
14.3	Binär-logistische Regression .....	281
14.3.1	Voraussetzungen .....	282
14.3.2	Durchführung .....	282
14.3.3	Ergebnis .....	283
14.3.4	Interpretation .....	284
14.3.5	Reporting der Ergebnisse .....	289
14.4	Ordinal-logistische Regression .....	289
14.4.1	Voraussetzungen .....	290
14.4.2	Durchführung .....	290
14.4.3	Ergebnis .....	291
14.4.4	Interpretation .....	292
14.4.5	Reporting der Ergebnisse .....	295
<b>Anhang .....</b>		<b>297</b>
A.1	Übersicht der allgemeinen Befehle für Diagramme mit der Basisversion von R .....	297
A.1.1	Beschriftungen .....	297
A.1.2	Schriftarten, Schriftvariation, Schriftgröße, Schriftfarben .....	297
A.1.3	Achsenformatierung .....	298
A.1.4	Linienarten und Datenpunkteformate .....	299
A.1.5	Legende .....	300
A.2	Übersicht der allgemeinen Befehle für Diagramme mit ggplot2 .....	301
<b>Glossar .....</b>		<b>303</b>
<b>Stichwortverzeichnis .....</b>		<b>307</b>

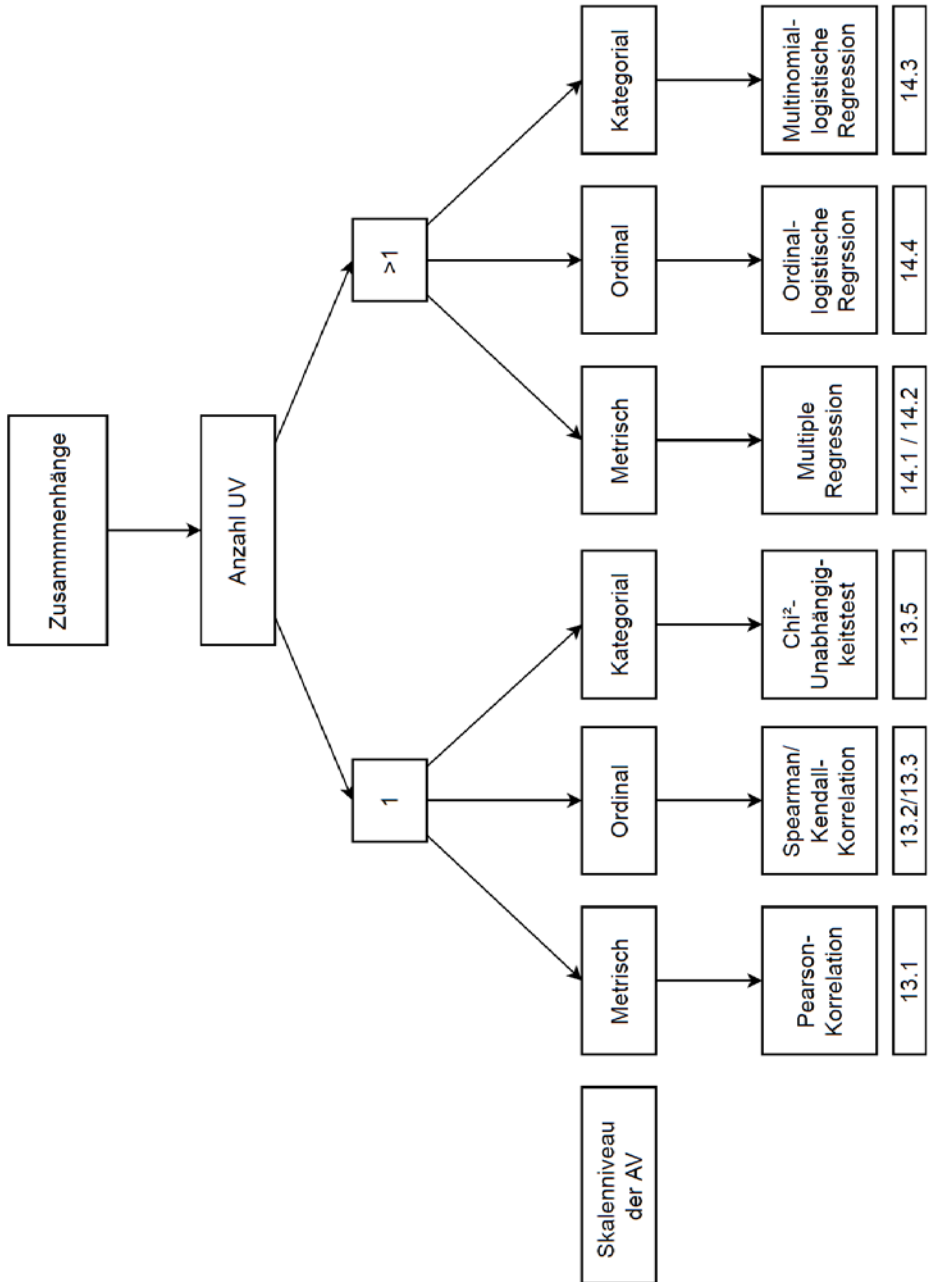


# Nachschlagehilfe

Mithilfe der unten abgebildeten Entscheidungsbäume können Sie die richtige statistische Testmethode finden und im jeweils darunter ausgewiesenen Abschnitt nachschlagen. Alsdann finden Sie im angegebenen Abschnitt stets den Vierklang aus 1) Voraussetzungsprüfungen, 2) Durchführung, 3) Interpretation der Ergebnisse und 4) Reporting.







# Einleitung

## E.1 R lernen in 14 Tagen

Mit diesem Buch haben Sie sich für einen einfachen, praktischen und fundierten Einstieg in die Welt der statistischen Analysen mit R entschieden. Sie lernen ohne unnötigen Ballast (in 14 Tagen oder Ihrem eigenen Tempo) alles, was Sie wissen müssen, um selbstständig statistische Analysen in R effektiv für Projekte in Ihrem Berufs-, Interessensgebiet oder Studienfach durchzuführen.

Alle Erklärungen sind leicht verständlich formuliert und setzen keine Vorkenntnisse in R voraus. Ein Grundverständnis von Statistik ist allerdings notwendig, da eine Erklärung jedes Fachbegriffes den Rahmen des Buches sprengen würde.

Dieses Buch ist als Nachschlagewerk konzipiert, welches Ihr Untersuchungsdesign in eine konkrete Analysemethode überführt. Hierbei helfen die Entscheidungsbäume, die Sie im Anschluss an das Inhaltsverzeichnis finden: Ausgehend vom Untersuchungsziel (Veränderung, Unterschiede, Zusammenhänge) und der Beschaffenheit der Testvariable(n) geben sie eine Entscheidungshilfe, um ein angemessenes Testverfahren auszuwählen.

## E.2 Der Aufbau des Buches

Dieses Buch ist kein klassisches Lehrbuch. Zur Geschichte und Entwicklung kann man sich – sofern man das möchte – ausführlich auf Wikipedia informieren. Vielmehr ist dieses Buch ein anwendungsorientiertes Nachschlagewerk. Es gliedert sich in vier Teile, beginnend mit einer Einführung in R und die grafische Benutzeroberfläche RStudio in **Teil I**. Anschließend stehen in **Teil II** das Datenmanagement in R und deskriptive Statistiken im Mittelpunkt. In **Teil III** werden verschieden Arten von Diagrammen gezeigt, die in R erstellt werden können. Schließlich werden in **Teil IV** des Buches statistische Analysemethoden gezeigt, die sich grob in Veränderungen, Unterschiede und Zusammenhänge unterteilen lassen.

Am Ende des Buches finden Sie ein praktisches Glossar mit den wichtigsten Fachbegriffen sowie ein Stichwortverzeichnis, das Ihnen hilft, bestimmte Themen im Buch schneller zu finden.

## E.3 Downloads zum Buch

Der Code aller Beispielprogramme steht Ihnen auf der Webseite des Verlags unter [www.mitp.de/0494](http://www.mitp.de/0494) zum Download zur Verfügung.

## E.4 Fragen und Feedback

Unsere Verlagsprodukte werden mit großer Sorgfalt erstellt. Sollten Sie trotzdem einen Fehler bemerken oder eine andere Anmerkung zum Buch haben, freuen wir uns über eine direkte Rückmeldung an [lektorat@mitp.de](mailto:lektorat@mitp.de).

Falls es zu diesem Buch bereits eine Errata-Liste gibt, finden Sie diese unter [www.mitp.de/0494](http://www.mitp.de/0494) im Reiter **DOWNLOADS**.

Wir wünschen Ihnen viel Erfolg und Spaß bei den statistischen Analysen mit R!

Björn Walther und das mitp-Lektorat

# Teil I

# Einführung in die Arbeit mit R und RStudio

Im ersten Teil dieses Buches geht es primär darum, Grundlagen im Umgang mit R und RStudio zu schaffen.

Gute Gründe, R für statistische Analysen zu nutzen, werden in **Kapitel 1** kurz dargelegt.

In **Kapitel 2** stehen die Grundprinzipien der R-Programmierung (Abschnitt 2.1) sowie die zur Verfügung stehenden Objekttypen im Fokus (Abschnitt 2.2). Hieran schließt sich das Management von Analysepaketen an (Abschnitt 2.3), bevor die für die in diesem Buch gezeigten Analyseverfahren notwendigen Analyseformate und die gegenseitige Überführung (Abschnitt 2.4) gezeigt werden. Den Abschluss des zweiten Kapitels bilden die zunächst noch etwas abstrakt anmutenden Pipe-Operatoren (Abschnitt 2.5). Diesen Abschnitt können Sie zunächst getrost überspringen und erst nach Verweis durch einen konkreten Anwendungsfall durcharbeiten.

Den Abschluss des ersten Teils dieses Buches bildet die Einführung in RStudio in **Kapitel 3**. Speziell wird das Layout erklärt (Abschnitt 3.1) und empfohlene Einstellungen gezeigt (Abschnitt 3.2).



# Warum gerade R für statistische Analysen?

Die Frage nach dem »Warum« ist auch in der Datenanalyse allgegenwärtig. Damit dieses Buch nicht zu philosophisch wird und seinem Versprechen eines anwendungsorientierten Nachschlagewerkes gerecht wird, werde ich hier nicht zu ausschweifend sein. So viel sei aber gesagt: Jede Person hat andere Präferenzen, **warum** gerade dieses eine Analyseprogramm das für sie beste ist. Zu den Kriterien zählen Einsteigerfreundlichkeit, Bedienbarkeit, Leistungsumfang, Updates, Preis – um nur ein paar zu nennen.

In den meisten o.g. Kategorien schneidet R sehr gut ab. Eigentlich in allen, außer der Einsteigerfreundlichkeit – aber dieses Buch ist ja dafür da, genau diesen Malus zu beheben. Eine gewisse Grundkenntnis statistischer Begriffe ist ohnehin bei allen Analyseprogrammen von Vorteil.

Zur Bedienung von R wird eine sog. *Syntax* verwendet. Sie beschreibt vereinfacht ausgedrückt das korrekte Kombinieren von Befehlen mit Objekten. Objekte können Variablen, Dataframes usw. sein. Diese Arbeitsweise zeichnet alle statistischen Analyseprogramme aus. Allerdings wurden im Laufe der Jahre aus Gründen der einfacheren Bedienbarkeit von manchen Herstellern (z.B. SPSS, inzwischen IBM) grafische Benutzeroberflächen mit Dialogfeldern aufgesetzt. Diese nehmen dem Nutzer das Eingeben der Syntax ab. Dies hat den Vorteil, dass man die Befehle nicht auswendig kennen muss und es nicht zu Tippfehlern kommen kann – allerdings zum Teil auf Kosten der Nachvollziehbarkeit und Reproduzierbarkeit der Analyseschritte.

Im Hinblick auf den Leistungsumfang ist R das »mächtigste« Analyseprogramm. Es werden standardmäßig sog. *Base packages* mitgeliefert, die aber nur einen Bruchteil der 19.000 existierenden Pakete darstellen. Diese Pakete beinhalten die von Nutzern verwendeten Analysefunktionen. Diese enorme Anzahl von Paketen wird größtenteils von Wissenschaftlern mit statistischem

Hintergrundwissen freiwillig erstellt und beständig mit Updates versorgt. Für jedes dieser R-Pakete existiert eine umfangreiche auf CRAN (Comprehensive R Archive Network) zugängliche Dokumentation.

Abschließend kann noch kurz der Preis erwähnt werden. R und sämtliche Pakete sind vollständig kostenlos herunterladbar. Es gibt auch kostenlose Zusatzprogramme, allen voran RStudio Desktop in der Open Source Edition. RStudio vereinfacht das Arbeiten erheblich, indem es die Übersichtlichkeit stark erhöht. Daher steht bereits an dieser Stelle meine klare Empfehlung, dieses Programm zu nutzen. Zu RStudio, dessen Installationen sowie Nutzung komme ich in Kapitel 3.





# R-Grundlagen in Kurzform

In den Grundlagen geht es nur um die rudimentärsten Dinge, die in R möglich sind und uns eine einfachere Auswertung ermöglichen. Dazu gehört das Verständnis der Syntax und deren Aufbau (Abschnitt 2.1), die Variablenformate (sog. *Objekttypen*, Abschnitt 2.2) sowie das Management der bereits erwähnten Pakete (Abschnitt 2.3). Dazu kommt das je nach Analysemethode unterschiedliche Datenformat (Abschnitt 2.4) und das Prinzip einer sehr eleganten Art der Schachtelung von Befehlen mittels Pipe-Operatoren (Abschnitt 2.5), die Ihnen später häufiger begegnen wird.

## 2.1 Syntax

Im vorangegangenen Kapitel wurde bereits kurz auf die Syntax eingegangen. Bisweilen liest man auch den Begriff »R-Programmiersprache«. An dieser Stelle werde ich mit den Begrifflichkeiten nicht zu genau sein – die kann man bei Bedarf (erneut) bei Wikipedia oder in diversen Büchern (à la »Einführung in R«) sehr detailliert nachlesen. Da es der Zweck des Buches ist, ein anwendungsorientiertes Nachschlagewerk zu sein, sei zum Thema Syntax nur so viel erwähnt, dass die Kombination von Befehlen mit Objekten für die Datenanalyse im Mittelpunkt steht. Der vom Nutzer eingegebene Quelltext wird nicht extra an einen Compiler übergeben, der dies dann in Maschinensprache übersetzen müsste, und dann erst zur Ausführung gebracht. Vielmehr wird durch die `Enter`-Taste die Ausführung direkt angestoßen.

Wichtige zu verinnerlichende Prinzipien beim »Programmieren« mit R sind die folgenden:

- R unterscheidet **Klein- und Großbuchstaben** (»case sensitive«).
- Das **Dezimaltrennzeichen** in R ist ein Punkt (z.B. 3.45 in R bedeutet 3,45).

- **Zuweisungen** (dazu später mehr) erfolgen über `<-`.  
In vielen Funktionen ist auch `=` nutzbar.
- Die **Bezeichnung** von Variablen bzw. Objekten allgemein darf nur alphanumerische Zeichen (A–Z, 0–9), Punkte und Unterstriche beinhalten, darf aber nicht mit einer Zahl beginnen (`data.2` wäre okay, `2.data` hingegen nicht).
- **Zeilenumbrüche** zur besseren Lesbarkeit sind mit `+` am Zeilenende möglich.
- **Abhängigkeiten** in Formeln werden mit `~` dargestellt. `y~x+z` bedeutet, dass die links stehende abhängige Variable »y« aus den rechts stehenden unabhängigen Variablen »x« und »z« geschätzt werden soll. Das `+` ist hier jedoch kein arithmetischer Operator und wird hier nur für die Aufnahme der Variablen verwendet.

## 2.2 Objekttypen in R

Die Arbeit in und mit R dreht sich um sog. **Objekte** bzw. **Objekttypen**. Dies sind Vektoren, Faktoren und Data Frames.

Im Gegensatz zur mathematischen Definition repräsentieren **Vektoren** in R *numerische* Variablen. Numerisch bedeutet Ordinal-, Intervall- und Verhältnisskalenniveau. Beispiel: Hat man die Körpergröße von Befragten (Verhältnisskalenniveau) erhoben, wird diese in einem beliebigen Vektor entsprechend der o.g. Namenskonvention gespeichert. Jede weitere *numerische* Variable (z.B. Alter, Einkommen) wird in einem extra Vektor gespeichert. Dies sind sog. **numeric-Vektoren**.

Ein Spezialfall eines Vektors ist der sog. **Faktor**. Faktoren enthalten Variablen auf Nominal- bzw. Kategorienskalenniveau. Hierzu zählen z.B. das Geschlecht von Befragten oder deren Lieblingsfarbe. Diese können entweder als Zahlen mit zusätzlicher Identifikation hinterlegt sein (z.B. 0-männlich, 1-weiblich) oder direkt als Wort (sog. **character-Vektoren**).

Eine Menge von Vektoren und Faktoren sind in einem sog. **Data Frame** zusammenfassbar. Sie können sich dies wie eine große Datentabelle (aus Excel oder SPSS) vorstellen, die zeilenweise die Befragten und spaltenweise die Variablen enthält:

ID	Geschlecht	Alter	Körpergröße	Einkommen
1	W	20	1,62	2100
2	M	21	1,78	2200
3	W	22	1,94	2300
4	...	...	...	...

In R ist die Arbeit mit Data Frames alltäglich, weil nach einem Datenimport die Speicherung der Daten in der Regel in einem Data Frame vorgenommen wird.

Der Vollständigkeit halber sei noch erwähnt, dass es drei weitere Objekte gibt, die aber im Rahmen der in diesem Buch gezeigten Analysemethoden praktisch keine Relevanz besitzen. **Matrizen** beinhalten wie Data Frames Objekte, allerdings können sie nur *entweder* numerische *oder* Textdaten beinhalten.

**Arrays** umfassen mehrere Matrizen und sind mehrdimensional. Sie können sie sich also wie eine Stapelung von Matrizen vorstellen.

**Listen** umfassen, ähnlich wie Data Frames oder Matrizen, mehrere Objekte. Der Unterschied ist, dass Listen Vektoren mit unterschiedlichen Längen (= Anzahl von Elementen) und Eigenschaften repräsentieren können.

## 2.3 R-Pakete finden und verwenden

### 2.3.1 Pakete installieren und laden

Es wäre logischer, an dieser Stelle mit dem Auffinden von Paketen zu beginnen. Allerdings findet man bestenfalls per Hörensagen heraus, welche Pakete für die eigenen Vorhaben taugen. Das Orientieren an Paketnamen schlägt leider ebenfalls fehl, da es z.T. sehr generische Namen sind und kaum Konventionen zu existieren scheinen. Und R wäre nicht R, wenn es kein Paket für das Finden von Paketen geben würde. ;-) Daher zeige ich zunächst anhand des Pakets **packagefinder** die Installation und das Aktivieren von Paketen.

Ein Paket wird stets mit der `install.packages()`-Funktion installiert. Hierbei ist es zwingend notwendig, das Paket mit exaktem Namen in Anführungszeichen in die Klammer zu setzen. Nach Ausführung dieser Codezeile werden benötigte Dateien bzw. Pakete, auf die das aktuell zu installierende Paket zugreift, automatisch heruntergeladen und installiert.

```
install.packages("packagefinder")
```

Nach erfolgreicher Installation (und bei jedem Start von RStudio, sofern kein Startskript existiert) muss das Paket zwingend geladen werden. Zum Laden wird die `library()`-Funktion verwendet. Hier ist das Paket erneut mit exaktem Namen, allerdings OHNE Anführungszeichen einzugeben. Zum Entladen wird die `detach()`-Funktion verwendet.

```
# Laden des Pakets packagefinder
library(packagefinder)

# Entladen des Pakets packagefinder
detach("package:packagefinder", unload = TRUE)
```

### 2.3.2 Finden von Paketen

Neben den Erfahrungen anderer, die Auswertungen vornehmen und hierfür für sie gut funktionierende Pakete gefunden haben, oder diesem Buch, wo ich auch diverse Pakete vorstelle, gibt es noch die Möglichkeit, über das Paket »packagefinder« Stichworte einzugeben.

Die `fp()`-Funktion erlaubt das gezielte Suchen nach Stichwörtern, die sich im Namen, der Kurz- und Langbeschreibung des Pakets befinden. Speziell in beiden Letzteren ist die Chance sehr gut, Treffer zu erzielen. Hierzu muss lediglich in Anführungszeichen das entsprechende Stichwort in die Klammer gesetzt und diese Codezeile ausgeführt werden.

Bei mehr als einem Stichwort wird in `fp()` zusätzlich `c()` eingefügt und die Stichwörter per Komma getrennt. Das Argument `mode=""` gibt mit einem logischen Operator an, ob ein oder mehrere der Stichwörter in der Suche vorkommen müssen. `or` verlangt *mindestens eins* der Stichwörter, and verlangt zwingend das Vorkommen *aller* Stichwörter.

```
001 fp("regression")
002 fp(c("regression", "interaction"), mode = "or")
```

Nachfolgend erhalten Sie im *Viewer* von RStudio (das Fenster unten rechts) eine Ergebnisübersicht (vgl. Abbildung 2.1), die einen kleinen Score am Anfang der jeweiligen Zeile hat, der als Indikator des Suchmatchings fungiert. Daneben stehen Paketname, die Kurzbeschreibung und der sog. **GO-Code**.

Score	Name	Short Description	GO
100.0	<b>SIMPLE.REGRESSION</b>	Multiple Regression and Moderated Regression Made Simple	15891
90.6	<b>fRegression</b>	Rmetrics - Regression Based Decision and Prediction	5637
85.6	<b>iRegression</b>	Regression Methods for Interval-Valued Variables	7878
84.4	<b>quickregression</b>	Quick Linear Regression	13120
83.6	<b>AnchorRegression</b>	Perform AnchorRegression	377
82.8	<b>mrregression</b>	Regression Analysis for Very Large Data Sets via Merge and Reduce	10284
74.1	<b>deepregression</b>	Fitting Deep Distributional Regression	3492
73.3	<b>safeBinaryRegression</b>	Safe Binary Regression	15104
71.9	<b>TwoRegression</b>	Process Data from Wearable Research Devices Using Two-Regression Algorithms	18106
69.8	<b>UnilsoRegression</b>	Unimodal and Isotonic L1, L2 and Linf Regression	18192
68.1	<b>MultipleRegression</b>	Multiple Regression Analysis	10453
67.7	<b>riskRegression</b>	Risk Regression Models and Prediction Scores for Survival Analysis with Competing Risks	14221

**Abb. 2.1:** Suchergebnisübersicht für das Stichwort »regression« im RStudio Viewer

Mit dem GO-Code, am Beispiel des SIMPLE.REGRESSION-Pakets 15891 arbeitet man wie folgt:

- `go(15891)` gibt die Kurzbeschreibung des Pakets in die R-Konsole aus.
- `go(15891, "manual")` ruft das Handbuch des Pakets im PDF-Format auf, während
- `go(15891, "website")` zur Homepage des Pakets führt, die in einem separaten Browserfenster geöffnet wird.

Wenn Sie die Arbeit im Browser bevorzugen, arbeiten Sie mit dem Zusatzargument `display = "browser"`.

```
fp("regression", display = "browser")
```

Im Ergebnis wird im Browser (vgl. Abbildung 2.2) zusätzlich eine Langbeschreibung, die Anzahl an Downloads, Links zur Beschreibung und zum Handbuch auf CRAN sowie der Installationscode angezeigt. Letzterer kann per Klick in die Zwischenablage kopiert und im R-Skript oder der R-Konsole eingefügt werden.



### Search Results

Original call: `findPackage(...)`  
[Copy to clipboard](#)

Search (keyword): **regression**

1098 of 19063 CRAN packages found in 2 seconds.

Score	Name	Short Description	Long Description	Total Downloads	Links	Install code
100.0	<b>SIMPLEREGRESSION</b>	Multiple Regression and Moderated Regression Made Simple	Provides SPSS- and SAS-like output for least squares multiple regression and moderated regression, as well as interaction plots and Johnson-Neyman regions of significance for interactions. The output includes standardized coefficients, partial and semi-partial correlations, collinearity diagnostics, plots of residuals, and detailed information about simple slopes for interactions. There are numerous options for designing interaction plots, including plots of interactions for both lin and lme models.	<a href="#">Downloads: 200</a>	<a href="#">R</a> <a href="#">G</a>	<a href="#">Copy</a>
90.6	<b>Regression</b>	Rmetrics - Regression Based Decision and Prediction	A collection of functions for linear and non-linear regression modeling. It implements a wrapper for several regression models available in the base and contributed packages of R.	<a href="#">Downloads: 200</a>	<a href="#">R</a> <a href="#">G</a>	<a href="#">Copy</a>
85.6	<b>Regression</b>	Regression Methods for Interval-Valued Variables	Contains some important regression methods for interval-valued variables. For each method, it is available the fitted values, residuals and some goodness-of-fit measures.	<a href="#">Downloads: 200</a>	<a href="#">R</a> <a href="#">G</a>	<a href="#">Copy</a>
84.4	<b>quickregression</b>	Quick Linear Regression	Helps to perform linear regression analysis by reducing manual effort. Reduces the	<a href="#">Downloads: 100</a>	<a href="#">R</a> <a href="#">G</a>	<a href="#">Copy</a>

Abb. 2.2: Suchergebnisübersicht für das Stichwort »regression« im Browser

## 2.4 Datenformate in R

Die aus der Statistik bekannten Datenformate *wide* und *long* können natürlich auch in R verwendet werden. Diese Unterscheidung ist essenziell, da je nach Analyseziel und anzuwendender Methode das eine oder andere Datenformat notwendig ist. Daher wird an dieser Stelle eine kurze Einordnung vorgenommen.

### 2.4.1 Wide-Format

Das Wide-Format ist das in den meisten Disziplinen häufiger anzutreffende Format. Es wird auch »ungestapelt« genannt und zeichnet sich dadurch aus, dass jedes Untersuchungsobjekt in einer separaten Zeile steht. Gleichzeitig stehen in den Spalten die Variablen, die für die Untersuchungsobjekte erhoben wurden. Ähnlich der Darstellung in Tabelle 2.1.

Sollte beispielsweise der BMI für die Probanden zu verschiedenen Zeitpunkten erhoben werden, wird für jeden Messzeitpunkt eine separate Variable angelegt, z. B. BMI\_t0, BMI\_t1 usw. Das ist zwar prinzipiell möglich und auch deutlich übersichtlicher, z.B. verlangt aber eine ANOVA mit Messwiederholung, dass die Daten im Long-Format vorliegen.

ID	Geschlecht	BMI_t0	BMI_t1	BMI_t2
1	w	21,72	21,48	21,65
2	m	30,41	30,00	29,48
3	w	24,05	24,18	23,82
...	...	...	...	...

Tab. 2.1: Beispiel für das Wide-Format

## 2.4.2 Long-Format

Im Long-Format (auch »gestapelt«) wird, um im Beispiel des BMI zu bleiben, für jede Messung des BMI eine separate Zeile erstellt. Zusätzlich bedarf es zweier Variablen: Zum einen muss erkennbar sein, um welche Messung bzw. welchen Zeitpunkt es sich handelt. Zum anderen ist das Untersuchungsobjekt mit einem **Identifier** (ID) eindeutig zuzuordnen.

ID	Geschlecht	Zeitpunkt	BMI
1	w	1	21,72
2	m	1	30,41
3	w	1	24,05
1	w	2	21,48
2	m	2	30,00
3	w	2	24,18
1	w	3	21,65
2	m	3	29,48
3	w	3	23,82
...	...	...	...

Tab. 2.2: Beispiel für das Long-Format

### 2.4.3 Transformation der Formate

Für die Transformationen vom einem zum anderen der beiden o.g. Formate kann das sog. **tidyr**-Paket verwendet werden. Die Installation und das Laden von Paketen kennen Sie bereits aus Abschnitt 2.3 und wenden dieses Wissen direkt an. Mit `install.packages()` wird es installiert und mit `library()` geladen:

```
001 install.packages("tidyr")
002 library(tidyr)
```

#### Transformation Wide-Format zu Long-Format

Aus dem `tidyr`-Paket wird die `pivot_longer()`-Funktion verwendet.

In die Funktion ist zu Beginn der Data Frame einzugeben, der transformiert werden soll. Anschließend sind die Variablen, in denen die Messwiederholungen stehen, anzugeben. Schließlich werden die Namen, die den Zeitpunkt (`names_to`) sowie die Messwerte (`values_to`) bezeichnen, vergeben.

Im Beispiel heißt der zu transformierende Data Frame `data_wide` und die Variablen `t0` bis `t20` aus ihm sollen transformiert werden. Die neue ZeitpunktvARIABLE wird schlicht mit `t` und die Wertevariable mit `v` abgekürzt und benannt.


```
001 data_long <- pivot_longer(data_wide, t0:t20,
002                             names_to = "t",
003                             values_to = "v")
```

#### Transformation Long-Format zu Wide-Format

Für das umgekehrte Prinzip wird die `pivot_wider()`-Funktion aus dem `tidyr`-Paket angewandt. Mit (`names_from`) werden die Variablennamen für die Variable, die die x Zeitpunkte (hier `t`) ausdrückt, erfasst. Mit (`values_from`) wird die Variable (hier `v`) benannt, aus der die Werte in die neuen x Spalten verschoben werden.

```
001 data_wide <- pivot_wider(data_long,
002                             names_from = t,
003                             values_from = v)
```





Proband	t0	t10	t20
A	28	42	48
B	5	10	17
C	23	26	28
D	20	30	25
E	16	24	27

Proband	t	v
A	t0	28
B	t0	5
C	t0	23
D	t0	20
E	t0	16
A	t10	42
B	t10	10
C	t10	26
D	t10	30
E	t10	24
A	t20	48
B	t20	17
C	t20	28
D	t20	25
E	t20	27

**Abb. 2.3:** Beispiel des Wide-Formats (links) und des Long-Formats (rechts) und die jeweilige Überführung in das andere Format

## 2.5 Pipe-Operatoren

Dieser Abschnitt behandelt ein recht kompliziertes Vorgehen und ist erst für spätere Kapitel relevant und kann und sollte daher zunächst übersprungen und erst im konkreten Anwendungsfall durchgearbeitet werden – für den Umgang mit R erachte ich ihn als unbedingt grundlegend. Speziell in späteren Kapiteln dient es zur erheblichen Reduktion des Arbeitsaufwands. Unscheinbar, aber sehr hilfreich, ist der sog. Pipe-Operator `%>%`, der ursprünglich aus dem `magrittr`-Paket stammt und inzwischen fester Bestandteil der Paketsammlung `tidyverse` ist.

Im Rahmen dieses Buches findet der Pipe-Operator vor allem in den Paketen `dplyr` und `rstatix` seine Anwendung, weil er eine sehr einfache Möglichkeit bietet, Funktionen zu schachteln, ohne sie paradoxerweise schachteln zu müssen. Beim Piping werden Befehle aneinandergereiht und mit dem Pipe-Operator `%>%` verbunden. Das Ergebnis sind flexiblere und weniger geschachtelte Funktionen. Dieses Vorgehen mutet zunächst kryptisch und kompliziert an, ist aber der einfachste Weg.

Dieses Vorgehen ist v.a. bei Auswertungen von Datensätzen im Long-Format (speziell bei mehr als zwei Beobachtungszeitpunkten, vgl. Abschnitt 10.2) sehr hilfreich. Die Simplität kann an einem einfachen Beispiel dargelegt werden:

Anstelle der `aggregate()`-Funktion und einer Schachtelung diverser Argumente und Dopplungen:

```
001 aggregate(data.a_l[, 3], list(data.a_l$Zeitpunkt), mean)
```

kann eine viel intuitivere Verkettung vorgenommen werden:

```
001 library(dplyr)
002 data.a_l %>%
003   group_by(Zeitpunkt) %>%
004   summarize(M = mean(Wert))
```

Die Ergebnisse sind identisch:

```
# aggregate()-Funktion
  Group.1      Wert
1      T0 24.48333
2      T1 27.40000
3      T2 29.01667
# Pipe Operator >%>
  Zeitpunkt      M
1      T0 24.48333
2      T1 27.40000
3      T2 29.01667
```

Mit einer zunehmend komplexeren Datenstruktur, speziell von im Long-Format vorliegenden Daten, ist die Auswertung mittels Pipe-Operatoren bequemer, teilweise auch nahezu alternativlos. Dies wird uns v.a. in den Kapiteln zur Untersuchung von Unterschieden und Veränderungen beugen.

# Stichwortverzeichnis

## A

Absolute Häufigkeit .....	61
Achsenformatierung .....	298
aes() .....	85
alpha() .....	57
Alphafehlerkumulierung .....	139, 171, 222
Alternativhypothese .....	142, 144, 147
ANOVA	
einfaktorielle .....	196
Friedman-ANOVA .....	174
gemischte .....	226
mehrfaktorielle .....	213
Welch-ANOVA .....	202
zweifaktorielle .....	213
ANOVA mit Messwiederholung .....	165
Ausreißer .....	168
Berechnung .....	170
deskriptive Voranalyse .....	166
Effektstärke .....	172
Interpretation .....	172
Normalverteilung .....	169
Reporting .....	173
Sphärizität .....	170
Voraussetzungen .....	166
anova_test() .....	170
Antennen .....	98
Argument .....	303
Arithmetischer Mittelwert .....	65
as.Date() .....	54
as.factor() .....	54
Assocs() .....	255
Auspartialisieren .....	257
Ausreißer .....	262, 264

## B

Balkendiagramm .....	95
barplot() .....	91
Baseline .....	165
Base packages .....	21
beside = TRUE .....	93
Binär .....	303
Binär-logistische Regression .....	281
Durchführung .....	282
Interpretation .....	284
Reporting .....	289
Voraussetzungen .....	282
Bindungen .....	162, 193, 245
Bootstrapping .....	280, 303
box_m() .....	232
boxplot() .....	98
Boxplot .....	98, 121
Erweiterung .....	101
ggplot .....	100
Lage- und Streuparameter .....	98
Teilgruppen .....	98
Box-Test .....	232
bptest() .....	268
Breusch-Pagan-Test .....	268

## C

car .....	59, 186
cbind() .....	64
Chancenverhältnis .....	256
character .....	54
character-Vektor .....	24
Chi <sup>2</sup> -Anpassungstest .....	149
Interpretation .....	151
Voraussetzungen .....	150

Chi <sup>2</sup> -Test .....	251
Chi <sup>2</sup> -Unabhängigkeitstest .....	251
Interpretation .....	253
chisq.test() .....	151
coefstest() .....	268
cohensD() .....	145, 188
Cohens d .....	144, 159, 173, 188, 204
Cohens W .....	254
cohenW() .....	254
col = c() .....	99
Comprehensive R Archive Network .....	22
Console .....	34
Cooks-Distanz .....	265
coord_flip() .....	96
corrplot .....	133
cor.test() .....	241, 243
CRAN .....	22
Cronbachs Alpha .....	57
CSV-Datei .....	44
cumsum() .....	63

### D

---

Dark Mode .....	35
Data Frame .....	24
Datenformat .....	28, 41
Datenpunkt .....	111
ggplot .....	114
Datenpunktformat .....	299
Datensatz .....	
einlesen .....	41
exportieren .....	51
speichern .....	52
zusammenfügen .....	46
Datumsformat .....	109
Datumsvariable .....	54
Dependent Variable .....	202
Desc() .....	72
describe() .....	70, 71
describeBy() .....	74

DescTools .....	72, 244
Deskriptive Statistik .....	61
Überblicksfunktionen .....	71
Untergruppen .....	73
detach() .....	26
Dichotom .....	181, 303
Dichtefunktion .....	86
dplyr .....	75, 161
DTA-Datei .....	46
Dummyscodierte Variable .....	272
Dummyscodierung .....	55, 303

### E

---

Effektgröße .....	144
Effektstärke .....	140, 144
Einfaktorielle ANOVA .....	196
Berechnung .....	201
Effektstärke .....	204
Interpretation .....	203
Prüfung auf Normalverteilung .....	198
Prüfung auf Varianzhomogenität .....	200
Voranalyse .....	197
Voraussetzungen .....	196
Einseitige Testung .....	142, 147, 157, 187
Einstichproben-t-Test .....	142
Effektstärke berechnen .....	144
Interpretation .....	144
Voraussetzungen .....	142
Einstichproben-Wilcoxon-Test .....	146
Effektstärke .....	148
Interpretation .....	148
Voraussetzungen .....	146
Eliminierungsverfahren .....	60
emmeans .....	223
emmeans_test() .....	220
Environment .....	34
Excel-Datei .....	45
Export .....	51
Exzess .....	70

## F

facet.by .....	169
Faktor .....	24
Faktorisierung .....	53
Fehlender Wert .....	53
Fehlerbalken .....	122, 123
file.choose() .....	44
Filter .....	50
kombinierter .....	50
Fisher-Test .....	251
Fit .....	264
Flächendiagramm	
gestapeltes .....	119
Follow-up-Beobachtung .....	165
Freiheitsgrad .....	171
Friedman-ANOVA .....	174
Ausreißer .....	176
Berechnung .....	176
deskriptive Voranalyse .....	175
Effektstärke .....	178
Interpretation .....	178
Reporting .....	179
Voraussetzungen .....	174
friedman-effsize() .....	178
friedman_test() .....	176
Funktion .....	303

## G

Gemischte ANOVA .....	226, 227
Berechnung .....	233
deskriptive Voranalyse .....	228
Gleichheit der Kovarianzmatrizen .....	232
Interpretation .....	236
Prüfung auf Ausreißer .....	232
Prüfung auf Normalverteilung .....	230
Prüfung auf Sphärizität .....	232
Prüfung auf Varianzhomogenität .....	231
Rechnung der Post-hoc-Tests .....	234, 236
geom_area() .....	120

geom_bar() .....	93
geom_boxplot() .....	100
geom_histogram() .....	85
geom_line() .....	114
geom_point() .....	114, 130
geometric.mean() .....	67
Geometrisches Mittel .....	67
Gepoolte Standardabweichung .....	303
Gerichtete Zusammenhänge	
<i>Siehe Regressionsanalyse</i>	
Gestapelt .....	303
Gestapeltes Flächendiagramm .....	119
get_anova_table() .....	171
Getrimmter Mittelwert .....	67
ggplot .....	85
Theme .....	88
ggplot2	
Befehlsübersicht .....	301
ggpubr .....	169, 184
ggqqplot() .....	169, 184
ggtitle() .....	85
glm() .....	287
GO-Code .....	26
Greenhouse-Geisser-Korrektur .....	233
Greenhouse-Geisser-Verfahren .....	171

## H

harmonic.mean() .....	68
Harmonisches Mittel .....	68
Häufigkeit	
absolute .....	61
kumulierte relative .....	63
relative .....	62
Häufigkeitstabelle .....	62
haven .....	45
Hebelwirkung .....	243, 264
Heteroskedastizität .....	267, 268
Histogramm .....	82, 85
für Gruppen .....	89

Homoskedastizität ..... 262, 267  
horiz = TRUE ..... 95

## I

---

Identifizier ..... 29  
identify\_outliers() ..... 168  
ifelse() ..... 56  
Importassistent ..... 41  
Imputation ..... 60  
Innersubjektfaktor ..... 226, 233  
interact\_plot() ..... 278  
Interaktionseffekt ..... 213, 234  
Interne Konsistenz ..... 57  
Interquartilsabstand ..... 69  
Inverscodierung ..... 57, 59  
IQR() ..... 69  
Item ..... 303

## K

---

Kausalität ..... 140  
Kendall's W ..... 178  
Kendall-Tau-Korrelation  
    Interpretation ..... 247  
Kendall-Tau-Rangkorrelation ..... 245  
Koeffiziententabelle ..... 285  
Konfidenzintervall ..... 143, 147  
Konsistenz  
    interne ..... 57  
Konstrukt  
    latentes ..... 56, 304  
Kontingenzkoeffizient ..... 255  
Korrelation ..... 77, 140, 242  
    negative ..... 138  
    positive ..... 138  
    punktbiserial ..... 248  
Korrelationsanalyse ..... 239  
    Voraussetzungen ..... 239  
Korrelationsdiagramm ..... 133  
Korrelationsmatrix ..... 133  
Kreisdiagramm ..... 103

Kreuztabelle ..... 76  
Kriterium ..... 138, 259  
kruskal-effsize() ..... 210  
kruskal\_test() ..... 208  
Kruskal-Wallis-Test ..... 205  
    Berechnung ..... 208  
    Effektstärke ..... 210  
    Interpretation ..... 209  
    Reporting ..... 211  
    Voranalyse ..... 206  
    Voraussetzungen ..... 206  
Kumulierte relative Häufigkeit ..... 63  
Kurtosis ..... 70

## L

---

label=..count.. ..... 86  
labels ..... 84  
Lageparameter ..... 65, 141  
Lang-Format ..... 303  
Latentes Konstrukt ..... 56, 304  
Legende ..... 112, 300  
leveneTest() ..... 186, 200  
Levene-Test ..... 186, 218  
Likelihood-Quotient-Chi-Quadrat-Test ..... 284  
Likelihood-Ratio-Test ..... 304  
Likert-skaliert ..... 189, 304  
Lineare Regression ..... 259  
    Durchführung ..... 262  
    Ergebnis ..... 269  
    Interpretation ..... 271  
    Mediation ..... 279  
    Moderation ..... 276  
    Reporting ..... 274  
    Voraussetzungen ..... 261  
    Vorbereitung ..... 260  
Linienart ..... 299  
Liniendiagramm ..... 108  
    Achsenbeschriftung ..... 109, 115  
    Achsen limitieren ..... 110  
    Fehlerbalken ..... 123

ggplot .....	114	Mittelwertscore .....	59, 304
Legende .....	117	Mixed ANOVA .....	226
Linie anpassen .....	111	Modellgüte .....	285
mehrere Variablen .....	112, 116	Moderation .....	276
Liniendiagramm mit Datenpunkten .....	108	Moderator .....	277
load.image() .....	53	Modus .....	66
Long-Format .....	29, 303	Multikollinearität .....	262, 263, 283
lm() .....	283	Multinomial logistische Regression .....	282
lsr .....	145		
<b>M</b>		<b>N</b>	
Mann-Whitney-U-Test		Nagelkerke $R^2$ .....	285
Berechnung .....	192	names = c() .....	99
Deskriptive Voranalyse .....	190	na.omit() .....	53
Effektstärke .....	194	Normalverteilungskurve .....	86
Interpretation .....	194	notch = TRUE .....	101
Voraussetzungen .....	190	Nullkorrelation .....	250
Mann-Whitney-Wilcoxon-Test		Nullmodell .....	284
<i>Siehe Mann-Whitney-U-Test</i>		numeric-Vektor .....	24
Mauchly-Test .....	170, 233		
mean() .....	65	<b>O</b>	
median() .....	65	Objekttyp .....	24
Median .....	65	OddsRadio() .....	257
mediate() .....	280	Odds-Ratio .....	256, 286
Mediation .....	276, 279	ols() .....	282, 290
Mehrfaktorielle ANOVA .....	213	openxlsx .....	52
Berechnung .....	219	Ordinal-logistische Regression .....	289
deskriptive Voranalyse .....	214	Interpretation .....	292
Interpretation .....	224	Reporting .....	295
Prüfung auf Normalverteilung .....	216	Voraussetzungen .....	290
Prüfung auf Varianzhomogenität .....	217	Ordinalskaliert .....	304
Reporting .....	225		
signifikante Interaktion .....	221	<b>P</b>	
Voraussetzungen .....	214	packagefinder .....	26
merge() .....	48	p.adjust.method .....	171
Mittel		pairwise_t_test() .....	171, 202
geometrisches .....	67	Paket .....	21
harmonisches .....	68	finden .....	26
Mittelwert		installieren .....	25
getrimmter .....	67	Partialkorrelation .....	138, 257
		partial.r() .....	258

Pearson-Korrelation .....	240
Interpretation .....	242
Pearson-Korrelationskoeffizient .....	241
pie() .....	103
Pipe-Operator .....	31
Piping .....	161, 304
pivot_longer() .....	30
pivot_wider() .....	30
plot() .....	108
Plotten .....	83
position = .....	89
Post-hoc-Test .....	171, 177, 202
Power .....	104, 139
statistische .....	160
Poweranalyse .....	139
p-Quantil .....	66
Prädiktor .....	138, 259
Prävalenz .....	257
pROC .....	287
prop.table() .....	62
Proxy .....	141, 304
Pseudobestimmungsmaß .....	304
Punktbiseriale Korrelation .....	248
Interpretation .....	250
Punktdiagramm .....	108

## Q

---

QQ-Diagramm .....	169
Q-Q-Plot .....	104, 155, 169
Quantil .....	66
Quartil .....	305
Quasi-metrisch .....	305
Quasimetrisches Skalenniveau .....	261
Quer-Format .....	304

## R

---

range() .....	69
raw_alpha .....	58
rbind() .....	46
rcompanion .....	149

read_dta() .....	46
read_excel() .....	45
read_sav() .....	45
read.table() .....	44
recode() .....	59
Regression	
binär-logistische .....	281
lineare	
<i>Siehe Lineare Regression</i>	
multinomial logistisch .....	282
ordinal-logistische .....	289
Regressionsanalyse .....	259
Regressionsgerade .....	262
Relative Häufigkeit .....	62
Relatives Risiko .....	257
Reliabilität .....	57
residuals() .....	199
Residuen .....	198, 305
normalverteilt .....	266
Residuendiagramm .....	200
Risiko	
relatives .....	257
rms .....	282
Robuster Standardfehler .....	267, 268
ROC-Curve .....	287
round() .....	62
rowMeans() .....	60
R-Script .....	33
rstatix .....	168
RStudio .....	33
Importassistent .....	41

## S

---

Säulendiagramm .....	90
Beschriftung .....	92
Fehlerbalken .....	122
für Gruppen .....	94
ggplot2 .....	93
Teilgruppen beschriften .....	92



SAV-Datei .....	45	stat_boxplot .....	101
save.image() .....	52	stat_function() .....	86
scale() .....	273	Statistik	
scale_color_manual() .....	119	deskriptive .....	61
scale_fill_manual() .....	89	Statistische Power .....	160
scale_linetype_manual() .....	119	Streudiagramm .....	127, 130
scale_x_continuous() .....	116	Legende .....	129, 132
scale_x_date() .....	115, 116	Teilgruppen .....	128, 131
Schiefe .....	70	Streuparameter .....	68
Schrift		subset() .....	49
-art .....	297	summary() .....	66
-farbe .....	298	Summenscore .....	59
-größe .....	298	Syntax .....	21, 23
-schnitt .....	297		
-variation .....	297	<b>T</b>	
Schweigeverzerrung .....	60	table() .....	62, 76
sd() .....	69	tapply() .....	73
Sensitivität .....	287	Tastatur-Shortcut .....	36
setwd() .....	43	Teildatensatz .....	49
Shortcut .....	36	Teststatistik	
Skala .....	56, 305	standardisierte .....	180
Skalenniveau		Testung	
quasimetrisches .....	261	einseitige .....	142, 147, 157, 187
sort() .....	66	Theme .....	88
Spannweite .....	69	tidyr .....	30
Spearman-Korrelation		t.test() .....	142, 186
Interpretation .....	244	t-Test bei abhängigen Stichproben .....	154
Spearman-Rangkorrelation .....	243	Effektstärke .....	158
Spezifität .....	287	Interpretation .....	158
Sphärizität .....	305	Voraussetzungen .....	154
SPSS .....	45	t-Test bei unabhängigen Stichproben .....	181
Standardabweichung .....	69, 122, 124	deskriptive Voranalyse .....	182
gepoolte .....	303	Effektstärke .....	188
Standardfehler		Interpretation .....	188
robuster .....	267, 268	Prüfung auf Normalverteilung .....	184
Standardisierte Teststatistik .....	180	Prüfung auf Varianzhomogenität .....	185
stargazer .....	274	Voraussetzungen .....	182
stat = .....	86	TXT-Datei .....	44
STATA .....	46		

## U

Übersichtstabelle .....	64
Ungerichtete Zusammenhänge	
<i>Siehe Korrelationsanalyse</i>	
Ungestapelt .....	304
unique .....	72
Unterschied	
gemischte ANOVA .....	226
mehr als zwei Gruppen .....	195
mehrere Gruppen infolge mehrerer	
Einflussfaktoren .....	213
t-Test bei unabhängigen Stichproben .	181
zwei Gruppen .....	181

## V

Validität .....	57
var() .....	69
Variable .....	305
bestimmte auswählen .....	49
dummycodierte .....	272
faktorisieren .....	53
kategoriale .....	55
Variablenname .....	43
Variance Inflation Factor .....	263
Varianz .....	69
Varianzhomogenität .....	185
Variationskoeffizient .....	69
Vektor .....	24
Veränderung	
mehr als zwei Zeitpunkte .....	165
zwei Zeitpunkte .....	153
Veränderungsdarstellung .....	107
vif() .....	263
VIF-Wert .....	263

## W

Welch-t-Test .....	186
Wert	
fehlender .....	53
Wertelabel .....	54
Whisker .....	98, 305
Wide-Format .....	28, 304, 305
wilcox_effsize() .....	179, 194
wilcoxonOneSampleR() .....	149
wilcoxonPairedR() .....	164
Wilcoxon-Test für abhängige Stich-	
proben .....	159
Effektstärke .....	164
Interpretation .....	163
Voraussetzungen .....	160
wilcox_test() .....	177, 208
wilcox.test() .....	146, 162, 192
Wölbung .....	70
<i>Siehe Kurtosis</i>	
write.csv() .....	51
write_sav() .....	52
write.table() .....	51
write.xlsx() .....	52

## Z

z-standardisiert .....	273
Zusammenhangsdarstellung .....	127
Zusammenhangsprüfung .....	76, 138
Zweifaktorielle ANOVA .....	213
Zweiseitige Testung .....	143, 147, 157
Zweistichproben-t-Test .....	181
z-Wert .....	195
Zwischensubjektfaktor .....	202, 226, 233