

## Auf einen Blick

1	Über dieses Buch .....	13
2	Einführung .....	19
3	R Base und RStudio: Installation und erste Schritte .....	25
4	Die Programmiersprache R .....	51
5	Grundlagen der statistischen Datenanalyse .....	73
6	Daten einlesen und für die Analyse vorbereiten .....	93
7	Daten analysieren mit einfacher Statistik .....	129
8	Umfassendes Praxisbeispiel .....	187
9	Abschluss .....	233

# Inhalt

Materialien zum Buch .....	11
<b>1 Über dieses Buch</b> .....	<b>13</b>
<b>1.1 Für wen ist dieses Buch? Für Sie?</b> .....	<b>13</b>
<b>1.2 Was sind die Ziele, was können Sie hier lernen?</b> .....	<b>13</b>
<b>1.3 Was Sie nicht lernen werden</b> .....	<b>15</b>
1.3.1 Verwandte Programmiersprachen .....	15
1.3.2 Big Data .....	15
1.3.3 Datenbankabfragen .....	16
1.3.4 Andere Datenformen .....	16
1.3.5 Kompliziertere Rechenverfahren .....	16
<b>1.4 Wie Sie mit diesem Buch arbeiten</b> .....	<b>17</b>
<b>2 Einführung</b> .....	<b>19</b>
<b>2.1 Statistik und Data Science im Vergleich</b> .....	<b>21</b>
2.1.1 Statistik .....	21
2.1.2 Data Science .....	22
<b>2.2 Was ist R, und warum sollten Sie das überhaupt lernen?</b> .....	<b>23</b>
<b>3 R Base und RStudio: Installation und erste Schritte</b> .....	<b>25</b>
<b>3.1 R Base</b> .....	<b>25</b>
3.1.1 Für Windows .....	27
3.1.2 Für Linux .....	28
3.1.3 Für macOS .....	29

<b>3.2</b>	<b>RStudio</b> .....	29
3.2.1	Für Windows .....	31
3.2.2	Für Linux .....	31
3.2.3	Für (Mac) OS X .....	31
<b>3.3</b>	<b>Wie sieht die Oberfläche aus, und was bedeuten die einzelnen Bereiche?</b> .....	31
<b>3.4</b>	<b>Die R-Konsole</b> .....	33
3.4.1	Erste Berechnungen in der R-Konsole .....	33
3.4.2	Skriptverlauf und Fehleranalyse .....	34
<b>3.5</b>	<b>Mein erstes Skript</b> .....	35
3.5.1	Skriptvorlage erstellen .....	35
3.5.2	Schreiben und Ausführen des ersten Skriptes .....	38
3.5.3	Abspeichern des erstellten Skripts .....	39
3.5.4	Laden des gespeicherten Skripts .....	41
<b>3.6</b>	<b>Hilfe!</b> .....	42
3.6.1	R-eigene Hilfe .....	42
3.6.2	Hilfe aus der riesigen R-Community .....	45
<b>4</b>	<b>Die Programmiersprache R</b> .....	51
<b>4.1</b>	<b>Objekte</b> .....	51
<b>4.2</b>	<b>Funktionen</b> .....	52
<b>4.3</b>	<b>Pakete (Packages)</b> .....	53
<b>4.4</b>	<b>Ein paar Vokabeln</b> .....	61
4.4.1	Die wichtigsten Funktionen und Befehle für Anfänger .....	61
4.4.2	Cheat-Sheets .....	67
<b>4.5</b>	<b>Kommentare</b> .....	68
<b>4.6</b>	<b>Groß- und Kleinschreibung und andere Syntaxregeln</b> .....	68
4.6.1	Benennung .....	68
4.6.2	Klammern .....	69
4.6.3	Syntax .....	69
<b>4.7</b>	<b>Computer sind dumm</b> .....	71

<b>5</b>	<b>Grundlagen der statistischen Datenanalyse</b> .....	73
<b>5.1</b>	<b>Fragestellung und Studiendesign</b> .....	73
<b>5.2</b>	<b>Von Daten und Datensätzen</b> .....	74
5.2.1	Querschnittsdaten .....	75
5.2.2	Längsschnitt- oder Paneldaten .....	78
5.2.3	Datenqualität .....	82
5.2.4	Datentypen in R (!) .....	82
5.2.5	Statistik .....	83
<b>5.3</b>	<b>Berechnung des Durchschnittswertes (Mittelwert, arithmetisches Mittel)</b> .....	84
<b>5.4</b>	<b>Wachstumsberechnung (Veränderungsberechnung)</b> .....	84
<b>5.5</b>	<b>Trend und lineare Regression</b> .....	86
<b>5.6</b>	<b>Beispieldatensatz »Zigarettenverbrauch«</b> .....	88
<b>6</b>	<b>Daten einlesen und für die Analyse vorbereiten</b> .....	93
<b>6.1</b>	<b>Daten aus Excel einlesen</b> .....	93
<b>6.2</b>	<b>Daten im .csv-Format einlesen</b> .....	98
6.2.1	Einlesen von .csv-Dateien mit RStudio .....	98
6.2.2	Umwandeln in Excel und weiter mit Abschnitt 6.1 .....	101
<b>6.3</b>	<b>Umgang mit Datumsangaben</b> .....	102
6.3.1	Datumsangaben als character-Strings .....	103
6.3.2	Datumsangaben aus Individualkomponenten .....	108
<b>6.4</b>	<b>Daten vorbereiten</b> .....	109
6.4.1	Excel .....	112
6.4.2	In R .....	114
<b>6.5</b>	<b>Not available! – Der Umgang mit fehlenden Werten</b> .....	122
6.5.1	NAs entdecken mit is.na() .....	122
6.5.2	Löschen von NAs .....	126

<b>7</b>	<b>Daten analysieren mit einfacher Statistik</b>	129
<b>7.1</b>	<b>Beispiel 1: Zigarettenkonsum</b>	129
7.1.1	Mittelwert	132
7.1.2	Wachstum	133
7.1.3	Trend/lineare Regression	136
<b>7.2</b>	<b>Beispiel 2: 100 Jahre Wohlstandsentwicklungen in Indien</b>	144
7.2.1	Zusammenstellung der benötigten Daten	144
7.2.2	Analyse	157
<b>7.3</b>	<b>Visualisierung</b>	166
7.3.1	Grundlagen und Funktionsweise von ggplot2	166
7.3.2	Einfaches Streudiagramm mit dem Datensatz Luftqualität	167
7.3.3	Barplot: Beispiel Zigarettendatensatz	170
7.3.4	Einfacher Lineplot: Beispiel Zigaretten- und Gapminder-Daten	179
<b>8</b>	<b>Umfassendes Praxisbeispiel</b>	187
<b>8.1</b>	<b>Fragestellung: Was will ich wissen?</b>	188
<b>8.2</b>	<b>Datenbeschaffung</b>	189
<b>8.3</b>	<b>Daten laden und Überblick verschaffen</b>	191
8.3.1	Daten laden	191
8.3.2	Überblick verschaffen	191
<b>8.4</b>	<b>Daten vorbereiten und bereinigen</b>	195
8.4.1	NAs entfernen	195
8.4.2	Unplausible Werte entfernen	196
8.4.3	Weitere Anpassungen	196
<b>8.5</b>	<b>Verarbeitung der Daten im Dreiklang</b>	196
<b>8.6</b>	<b>Kommunikation</b>	219
<b>8.7</b>	<b>Dokumentation</b>	219
<b>8.8</b>	<b>Bonus: Folgeanalyse in der Zeitreihe: Krankheitstage</b>	219

<b>9</b>	<b>Abschluss</b>	233
<hr/>		
	<b>Anhang</b>	235
<hr/>		
<b>A</b>	<b>Datenmaterial</b>	235
A.1	Zigaretten 1	235
A.2	Zensus Länder	236
A.3	Luftqualität	237
A.4	Zigaretten 2	243
A.5	Indien 1990–2000	244
A.6	Personaldaten	245
A.7	Personaldaten – Erweitert	252
Index		261