

# Statistik\_\_21

*Sebastian Sauer*

*2016-11-21*



# Contents

<b>Vorwort</b>	<b>5</b>
<b>1 Einführung</b>	<b>7</b>
1.1 Rahmen . . . . .	7
1.2 Was ist Statistik? Wozu ist sie gut? . . . . .	7
<b>2 Trends</b>	<b>9</b>
<b>3 Unbehagen</b>	<b>11</b>
<b>4 Datenjudo</b>	<b>13</b>
<b>5 Visualisierung</b>	<b>15</b>
<b>6 Statistisches Modellieren</b>	<b>17</b>
<b>7 Numerische Modelle</b>	<b>19</b>
<b>8 Klassifizierende Modelle</b>	<b>21</b>

```
source("./source/libs.R")
#> Loading tidyverse: ggplot2
#> Loading tidyverse: tibble
#> Loading tidyverse: tidyr
#> Loading tidyverse: readr
#> Loading tidyverse: purrr
#> Loading tidyverse: dplyr
#> Conflicts with tidy packages -----
#> filter(): dplyr, stats
#> lag():    dplyr, stats
```



# Vorwort

- Worum geht es in diesem Buch
  - Einführung in moderne Verfahren der Statistik
  - Für Praktiker
  - Betonung liegt auf “modern” und “Praktiker”
- Ziel des Buches
  - Intuitives, grundlegendes Verständnis zu zentralen Konzepten
  - Handwerkszeug zum selber Anwenden
- Unterschied zu anderen Büchern
  - Wenig Formeln
  - Keine/weniger “typischen” klassischen Methoden wie ANOVA, Poweranalyse etc.
  - Aufzeigen von Problemen mit klassischen Verfahren
  - Kritik am Status-Quo
- Didaktik
  - Hands-on
  - R
  - Lernfragen
  - Fallstudien
  - Aktuelle Entwicklungen ausgerichtet



# Chapter 1

## Einführung

### 1.1 Rahmen

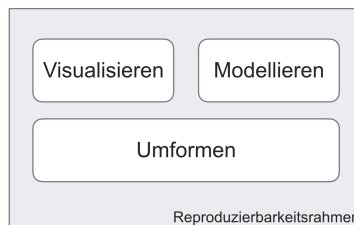
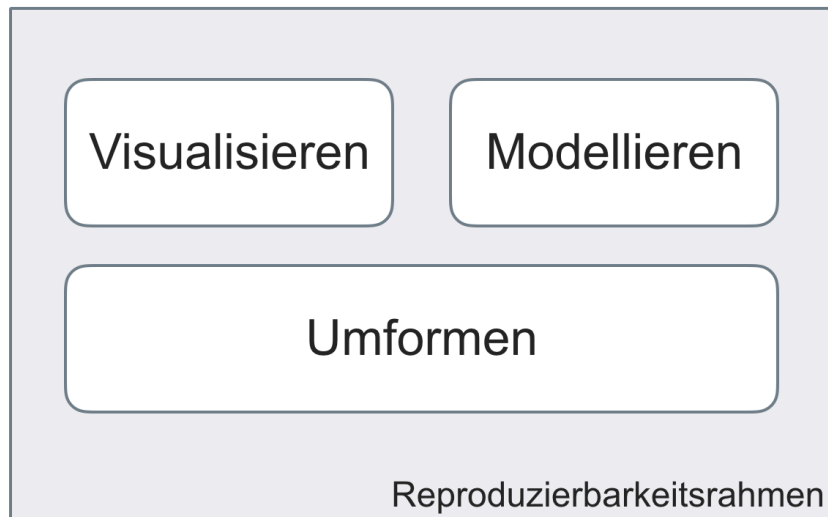


Figure 1.1:



### 1.2 Was ist Statistik? Wozu ist sie gut?

- Diese zwei Fragen sollte man sich am Anfang der Beschäftigung mit jedem Thema stellen.
- Statistik meint Methoden, die das Ziel haben, Ereignisse präzise vorherzusagen

- Statistik soll sich um Dinge dieser Welt drehen, nicht um Parameter
- Statt einer Frage “ist  $\mu_1$  größer als  $\mu_2$ ?” besser “Wie viel Umsatz erwarte ich von diesem Kunden?”, “Wie viele Saitensprünge hatte er wohl?”, “Wie groß ist die Wahrscheinlichkeit für sie zu überleben?” und dergleichen.
- Der Nutzen von Vorhersagen liegt auf der Hand: Vorhersagen sind praktisch; eine nützliche Angelegenheit (wenn auch schwierig).



## Chapter 2

# Trends

- Big Data
- Open Science
- Computerisierung
- Neue Methoden zur numerischen Vorhersage
- Textmining



## Chapter 3

# Unbehagen

- p-Werte
- Theorie der Wahrscheinlichkeit im Frequentismus
- Reproduzierbarkeitskrise
- Parameter
- Kausalität
- Übersicherheit



## Chapter 4

# Datenjudo

Daten umformen.

- dplyr
- Normalform



## Chapter 5

# Visualisierung

- Nutzen (Anscombe)
- Prinzipien nach Tufte
- Cleveland
- ggplot2





## Chapter 6

# Statistisches Modellieren

- Was sind Modelle?
- Überanpassung
- Prädiktion vs. Explanation
- Numerische vs. klassifizierende Modelle
- Geleitete vs. ungeleitete Modelle
- Parametrische vs. nichtparametrische Modelle
- Fehler- vs. Varianzreduktion
- Modellgüte



## Chapter 7

# Numerische Modelle

- Lineare Regression
  - Grundlagen
  - Multiple Regression
  - Interaktion
  - Eisberge
- Logistische Regression
- Penalisierende Regression
- Baumbasierte Verfahren
- Ausblick



## Chapter 8

# Klassifizierende Modelle

- Clusteranalyse
- Nächste-Nachbarn-Analyse



# Bibliography