

# Datenkompetenz für alle

Ein Werkstattbericht zum  
FOM-Statistik-Curriculum –  
**Sebastian Sauer**

- 1 Datenkompetenz
- 2 Inhalt: Statistical Thinking  
(Statistikverständnis)
- 3 Didaktik: Aktivierung
- 4 Projektmanagement 1: Markdown
- 5 Projektmanagement 2: Github
- 6 Abschluss

# 1 Datenkompetenz

Unter *Datenkompetenz* versteht man versteht man die Fähigkeit, Daten auf kritische Art und Weise zu sammeln, zu managen, zu bewerten und anzuwenden ([Ridsdale et al., 2015](#)).

## Warum ist Datenkompetenz wichtig?

---

*Work across nearly all domains is becoming more data driven, affecting both the jobs that are available and the skills that are required. As more data and ways of analyzing them become available, more aspects of the economy, society, and daily life will become dependent on data. It is imperative that educators, administrators, and students begin today to consider how to best prepare for and keep pace with this data-driven era of tomorrow<sup>1</sup>.*

---

<sup>1</sup>National Academies of Sciences, Engineering, and Medicine (2018), vgl. Kirchherr et al. (2018) oder Engel (2017).

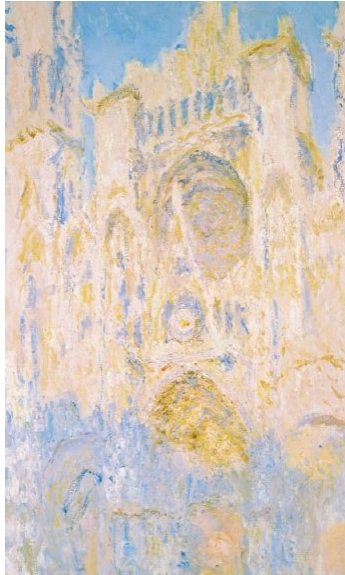


---

<sup>2</sup>w/m/d

<sup>3</sup>Chance and Tintle (2016): Teaching the statistical investigation process with simulation-based inference

## Eher ein “impressionistisches” Verständnis von Statistik?<sup>4</sup>



---

<sup>4</sup>Chance and Tittle (2016): Teaching the statistical investigation process with simulation-based inference

... ohne Überblick?<sup>5</sup>



---

<sup>5</sup>Chance and Tintle (2016): Teaching the statistical investigation process with simulation-based inference



### Inhalt: *statistical thinking*

- ▶ Modellierung
- ▶ (An)Erkennen von Unsicherheit
- ▶ Prinzipien statt Regelbaum

### Didaktik: *Aktivierung*

- ▶ Quizze
- ▶ Live-Coding
- ▶ Fallstudien
- ▶ ...
- ▶ ...

### Projektmanagement: *Git*

- ▶ Gemeinschaftsprojekt
- ▶ Änderungsmanagement mit Git
- ▶ Nutzerfreundliches Skript

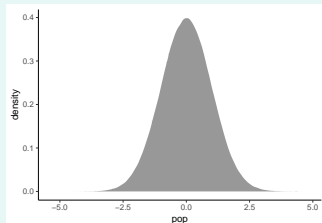
## **2 Inhalt: Statistical Thinking (Statistikverständnis)**

# Idee 1: Ungewissheit (an)erkennen

Kenntnis einer Stichprobe ist nicht identisch zu Kenntnis der Population

### Population

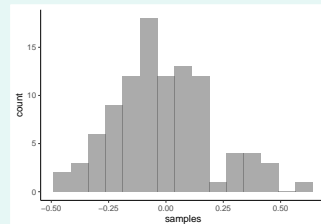
$$X \sim N(0, 1)$$



### 100 Stichproben (je $n = 20$ )

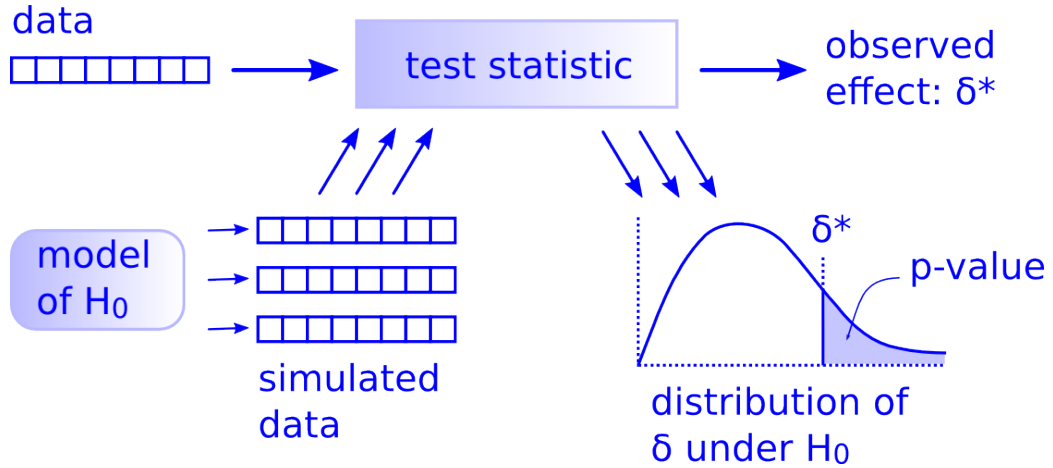
$$M_1 = 0.17, M_2 = 0.09, M_3 = 0$$

...



## Idee 2: Übergreifendes Prinzip

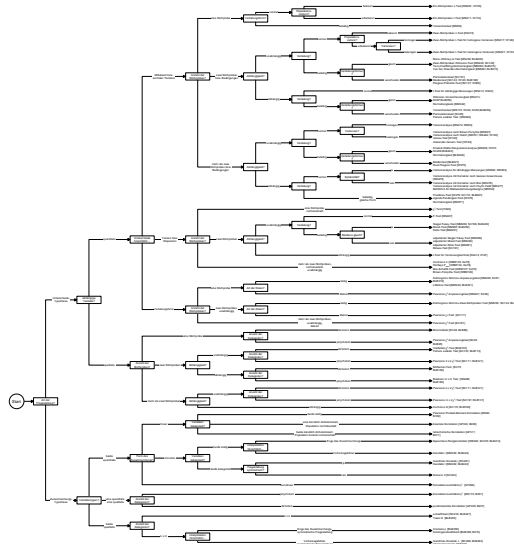
Nur **ein** statistischer Test: Simulation der Teststatistik



## Anstatt vieler Bäume

### Die Auswahl statistischer Tests und Maße

Sven Blankenberger, Institut für Psychologie der Martin-Luther-Universität Halle-Wittenberg  
Dirk Vorberg, Institut für Psychologie der Technischen Universität Braunschweig

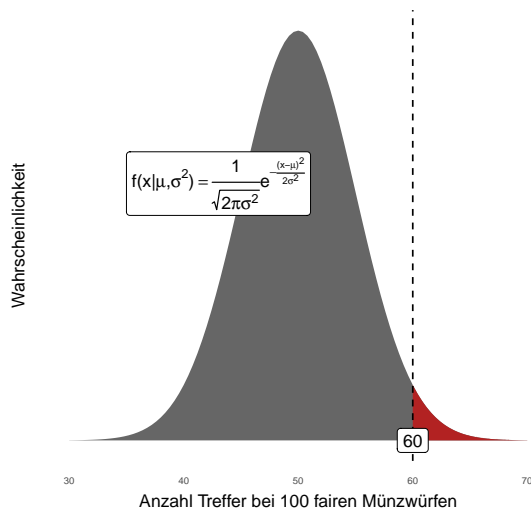


101. Blom, C. (1982). Statistik für Psychologen (2. Aufl.). Berlin: Springer.  
102. Blom, C. (1982). Statistik für Psychologen (2. Aufl.). Berlin: Springer.  
103. Blom, C. (1982). Statistik für Psychologen (2. Aufl.). Berlin: Springer.  
104. Blom, C. (1982). Statistik für Psychologen (2. Aufl.). Berlin: Springer.  
105. Blom, C. (1982). Statistik für Psychologen (2. Aufl.). Berlin: Springer.  
106. Blom, C. (1982). Statistik für Psychologen (2. Aufl.). Berlin: Springer.  
107. Blom, C. (1982). Statistik für Psychologen (2. Aufl.). Berlin: Springer.  
108. Blom, C. (1982). Statistik für Psychologen (2. Aufl.). Berlin: Springer.  
109. Blom, C. (1982). Statistik für Psychologen (2. Aufl.). Berlin: Springer.  
110. Blom, C. (1982). Statistik für Psychologen (2. Aufl.). Berlin: Springer.

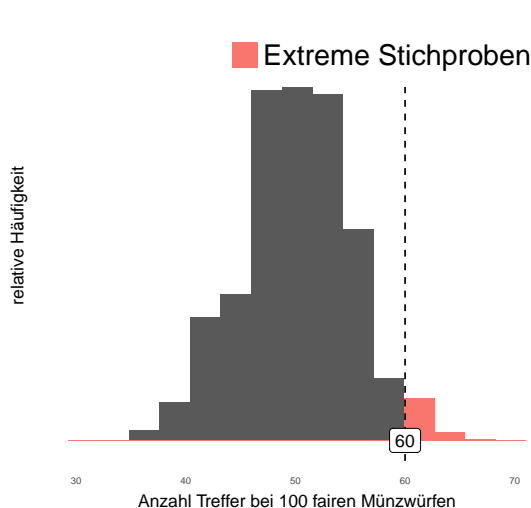
# Vergleich klassischer und simulationsbasierter Inferenz

- ▶ Klassisch: Leite Stichprobenverteilung aus Theorie ab
- ▶ Simulationsbasiert: Führe das Experiment laut  $H_0$  häufig aus

Berechne das Integral der Fläche unter der Kurve

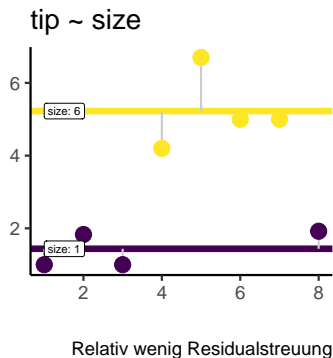
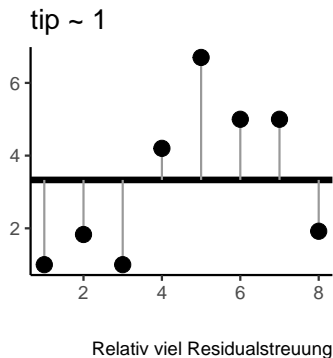


Führe Experiment oft aus; berechne jeweils Teststatistik



### Idee 3: Modellierung

Versuche, die Streuung zu erklären (verringern):



Geschicktes Modellieren verringert die Ungewissheit.<sup>6</sup>

<sup>6</sup>Quelle: <https://data-se.netlify.com/2019/03/26/reducing-residual-variance-in-modeling/>

# Mittelwert als “Modell”

### Idee:

$$\text{Daten} = \text{Modell} + \text{Rest.}$$

```
mean( ~ total_bill, data = tips)
```

```
## [1] 19.78594
```

- ▶ Für Beobachtung  $i$  gilt:  $x_i = \bar{x} + (x_i - \bar{x})$ .
- ▶ Der “Rest”  $x_i - \bar{x}$  beschreibt die Abweichung<sup>7</sup> der Beobachtung zum Mittelwert  $\bar{x}$  (hier: “Modell”).

---

<sup>7</sup>Zur Einschätzung, ob die Abweichung groß oder klein ist, wird die Streuung (s. u.) der Daten herangezogen.

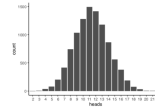


## **3** Didaktik: Aktivierung

```
analysiere( y # ggfs. abhängige Variable  
  ~ x # unabhängige Variable(n)  
  | z, # ggfs. bedingende (gruppierende) Variable(n)  
  Optionen, # ggfs. weitere Optionen  
  data = meine_daten ) # Datensatz
```

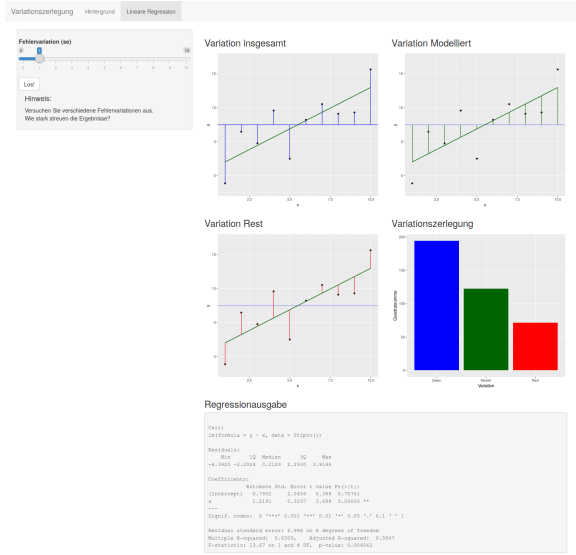
#### 8. Inferenzstatistik

#### Übung 106: Simulation (II / II)



Bei welchem Wert für  $x$  würden Sie bei  $n = 34$  am stärksten vermuten, dass ein Geschmacksunterschied vorliegt, d. h., dass  $\pi > \frac{1}{3}$  ist?

- A. Bei  $x = 5$ .
- B. Bei  $x = 10$ .
- C. Bei  $x = 15$ .
- D. Bei  $x = 20$ .



<sup>8</sup>z. B. <https://fomshinyapps.shinyapps.io/Modellierung/>

- ▶ Fallstudien-Webseite
- ▶ Auflockerung: Cartoons, Songs, Videos, ...
- ▶ ...

# 4 Projektmanagement 1: Markdown

# Kein Powerpoint<sup>9</sup>, sondern RMarkdown



---

<sup>9</sup>wirklich nicht

## 4. Projektmanagement 1: Markdown

# (R)Markdown ist nicht HTML

```
<ul class="nav-links">
  <li><a href="/about/">About</a></li>

  <li><a href="/">Blog</a></li>

  <li><a href="/privacy/">Data privacy</a></li>
</ul>
</nav>
</header>
<main class="content" role="main">
  <article class="article">
    <span class="article-duration">40 min read</span>
    <h1 class="article-title">
      Bayesian modeling of populist party success in German federal elections -
      A notebook from the lab</h1>
    <span class="article-date">2018/08/25</span>
    <div class="article-content">
      <p>Following up on an <a href="https://data-se.netlify.com/2017/10/10/afd-map/">earl1
we will model the voting success of the (most prominent) populist party,
AfD, in the recent federal elections. This time,
Bayesian modeling techniques will be used, drawing on the
<a href="https://xcelab.net/rm/statistical-rethinking/">excellent textbook</a> by McElreath.</p>
    <hr />
```



## 4. Projektmanagement 1: Markdown

# (R)Markdown ist nicht $\LaTeX$

```
\makeatletter
\newenvironment{kframe}{%
  \medskip{}
  \setlength{\fboxsep}{.8em}
  \def\at@end@of@kframe{}%
  \ifinner\ifhmode%
  \def\at@end@of@kframe{\end{minipage}}%
  \begin{minipage}{\columnwidth}%
  \fi\fi%
  \def\FrameCommand##1{\hskip\@totalleftmargin \hskip-\fboxsep
\colorbox{shadecolor}{##1}\hskip-\fboxsep
  % There is no \@totalrightmargin, so:
  \hskip-\linewidth \hskip-\@totalleftmargin \hskip\columnwidth}%
  \MakeFramed {\advance\hsize-\width
  \@totalleftmargin\z@ \linewidth\hsize
  \@setminipage}}%
{\par\unskip\endMakeFramed%
  \at@end@of@kframe}
\makeatother
```

# Markdown ist eine einfache Auszeichnungssprache

```
This is a sentence.
```

```
Now a list begins:
```

- no importance
- again
- repeat

```
A numbered list:
```

1. first
2. second
3. third

```
__bold__, _italic_,  
~~strike through~~
```

This is a sentence.

Now a list begins:

- ▶ no importance
- ▶ again
- ▶ repeat

A numbered list:

1. first
2. second
3. third

**bold**, *italic*,  
~~strike through~~

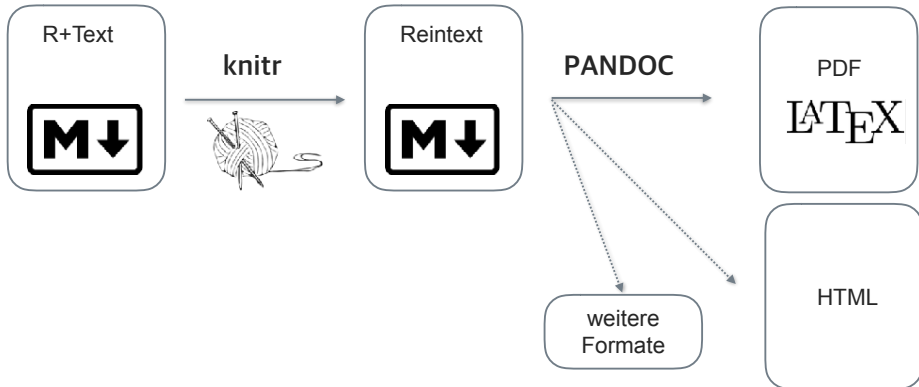
# Markdowns Philosophie

---

Ein Markdown kann in (fast) alle anderen Auszeichnungssprachen übersetzt werden:  $\LaTeX$ (PDF), HTML, Powerpoint, Word, ...

### Einfach. Einfach.

- ▶ Keine Formatierung. Nur Inhalt.
- ▶ Kaum Steuerzeichen-Müll
- ▶ Plattformunabhängig



# Einfacher Markdown-Code

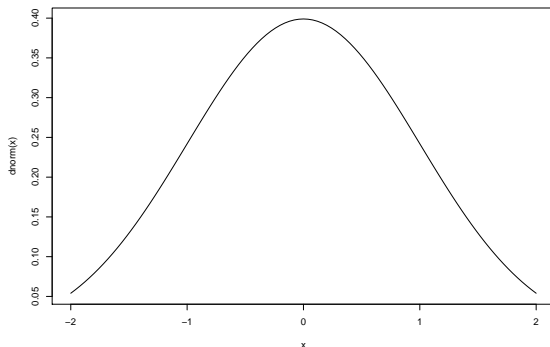
Sie kennen mich:

```
```
```

```
curve(dnorm(x), from = -2, to = 2)
```

```
```
```

Sie kennen mich:

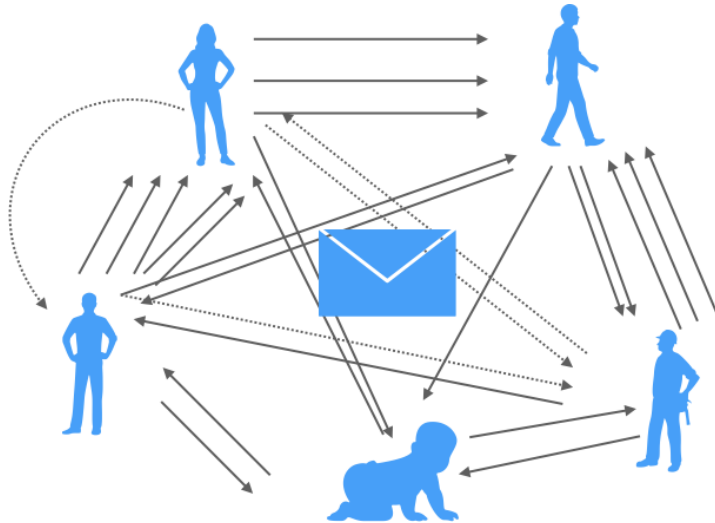


# 5 Projektmanagement 2: Github



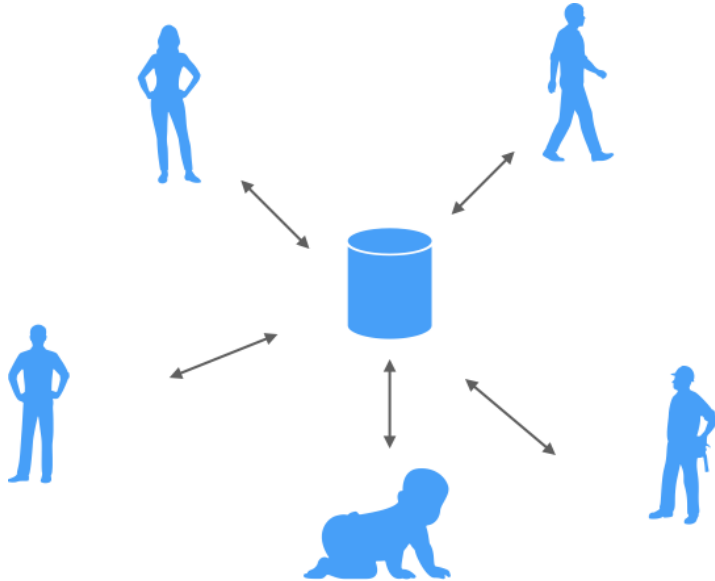
# Emails zum Austausch gemeinsamer Dateien?

Schicken Sie Ihre Dokument-Varianten *nicht* in der Arbeitsgruppe umher





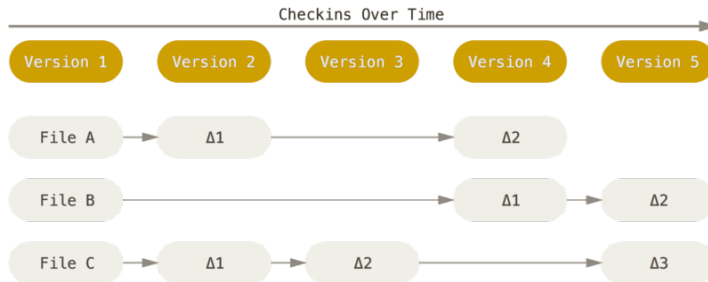
# Nutzen Sie eine gemeinsame Datei-Ablage



## ***Git* ist eine Software zur Versionsverwaltung**



# Git dokumentiert Änderungen im Text



# Live-Beispiel zu *Git*

---

<https://github.com/luebby/Vorlesungen>

Guter Einstieg: <https://happygitwithr.com/>

## 6 Abschluss

Entwicklungsteam: **Karsten Lübke**, **Norman Markgraf**, Oliver Gansser, Matthias Gehrke, Bianca Krol, Sebastian Sauer

Kontakt: [sebastian.sauer@fom.de](mailto:sebastian.sauer@fom.de)

Das Statistikskript ist verfügbar als CC-BY-SA-NC 3.0:

<https://github.com/luebby/Vorlesungsfolien>.

Die Beamer-Templates stammen aus dem [NPBT-Projekt von Norman Markgraf](#) und stehen wie der Source Code unter der Lizenz GNU General Public License v3.0.

Dieses Projekt baut auf vielen Open-Source-Projekten auf, z. B. R, RStudio, Markdown,  $\text{\LaTeX}$  und R-Paketen wie mosaic und ggplot2 – und vielen weiteren. Danke an alle Open-Source-Entwickler! `icon::fa("rocket")`

- Joachim Engel. Statistical literacy for active citizenship: A call for data science education. *Statistics Education Research Journal*, 16(1):44–49, 2017.
- Julian Kirchherr, Julia Klier, Cornels Lehmann-Brauns, and Matthias Winde. Future skill: Which skills are lacking in germany. Technical report, Stifterverband and McKinsey, 2018. URL <https://www.future-skills.net/download/file/fid/204>.
- National Academies of Sciences, Engineering, and Medicine. *Data Science for Undergraduates: Opportunities and Options*. The National Academies Press, 2018. doi:10.17226/25104.
- Chantel Ridsdale, James Rothwell, Michael Smit, Hossam Ali-Hassan, Michael Bliemel, Dean Irvine, Daniel Kelley, Stan Matwin, and Bradley Wuetherick. Strategies and best practices for data literacy education: Knowledge synthesis report. Technical report, Dalhousie University, 2015.