

# **Statistik und Datenanalyse**

Christina Bogner

2022-08-21T18:31:25+02:00

# Table of contents

<b>Vorwort</b>	<b>3</b>
Kurs . . . . .	3
Lernziele . . . . .	3
Sinn und Unsinn dieses Skripts . . . . .	3
Inspiration, Quellen und Danksagung . . . . .	4
Reproduzierbarkeit . . . . .	4
<b>1 Einführung</b>	<b>6</b>
1.1 Was ist Statistik? . . . . .	6
<b>I DATEN ERKUNDEN</b>	<b>7</b>
<b>2 Daten</b>	<b>8</b>
2.1 Was sind Daten? . . . . .	8
2.2 Wo kommen Daten her? . . . . .	8
2.2.1 Datenerhebung . . . . .	8
2.3 Testen Sie sich selbst . . . . .	8
2.4 Übungen . . . . .	8
<b>3 Daten visualisieren</b>	<b>9</b>
<b>Bibliographie</b>	<b>10</b>

# Vorwort

## Kurs

In dieser Veranstaltung werden wir folgende Werkzeuge verwenden:

**ILIAS:** die Online-Lernplattform der UzK. Entweder sind Sie bereits automatisch in dem Kurs registriert oder werden von mir per Hand angemeldet.

**Campuswire:** die Chatplattform dient der allgemeinen Kommunikation und der Selbstorganisation des Lernens. Verwenden Sie diese, um Fragen mit Ihren Kommilitonen\*innen und mir zu diskutieren. Sie sollten eine Einladungsmail zu Campuswire erhalten haben.

**Zoom:** die Videokonferenz-Software bleibt unser Notfall-Werkzeug, falls keine Präsenz möglich ist.

## Lernziele

### Lernziele des Kurses

- Einfache Datenanalysen reproduzierbar durchführen
- Daten und Ergebnisse der Analysen darstellen
- Ergebnisse interpretieren und berichten

## Sinn und Unsinn dieses Skripts

Dieses Skript ist ein lebendiges Begleitdokument des Kurses. Es wird laufend angepasst und aktualisiert.

## Inspiration, Quellen und Danksagung

Dieses Skript baut stark auf folgenden freien Quellen auf:

- r4ds: Wickham and Grolemund (2021)
- ggplot2: Wickham (2020)
- ModernDive: Ismay and Kim (2021)
- Introduction to Modern Statistics: Çetinkaya-Rundel and Hardin (n.d.)

Den Autoren dieser Bücher gilt ein großer Dank für Ihren Beitrag zur -Community !

## Reproduzierbarkeit

Dieses Skript wurde in [RStudio](#) mit [Quarto](#) geschrieben und in R version 4.2.1 (2022-06-23) gebaut. Folgende Pakete werden für die Beispiele und Übungen benötigt:

package	version	source
dabestr	0.3.0	Github (ACCLAB/dabestr@8775899f7eba743a6a32bd2fdab5f57e79401fd6)
emojifont	0.5.5	CRAN (R 4.2.0)
fontawesome	0.3.0	CRAN (R 4.2.1)
gapminder	0.3.0	CRAN (R 4.2.0)
infer	1.0.0	CRAN (R 4.2.0)
lubridate	1.8.0	CRAN (R 4.2.0)
moderndive	0.5.3	CRAN (R 4.2.0)
tidyverse	1.3.1	CRAN (R 4.2.0)

Die komplette Information zur Session lautet:

```
R version 4.2.1 (2022-06-23)
Platform: x86_64-pc-linux-gnu (64-bit)
Running under: Ubuntu 22.04.1 LTS
```

```
Matrix products: default
BLAS:   /usr/lib/x86_64-linux-gnu/atlas/libblas.so.3.10.3
LAPACK: /usr/lib/x86_64-linux-gnu/atlas/liblapack.so.3.10.3
```

```
locale:
 [1] LC_CTYPE=de_DE.UTF-8      LC_NUMERIC=C
 [3] LC_TIME=de_DE.UTF-8      LC_COLLATE=de_DE.UTF-8
```

```
[5] LC_MONETARY=de_DE.UTF-8    LC_MESSAGES=de_DE.UTF-8
[7] LC_PAPER=de_DE.UTF-8       LC_NAME=C
[9] LC_ADDRESS=C               LC_TELEPHONE=C
[11] LC_MEASUREMENT=de_DE.UTF-8 LC_IDENTIFICATION=C
```

attached base packages:

```
[1] stats      graphics  grDevices  utils      datasets  methods    base
```

loaded via a namespace (and not attached):

```
[1] rstudioapi_0.13    knitr_1.39          magrittr_2.0.3      R6_2.5.1
[5] rlang_1.0.2        fastmap_1.1.0       fansi_1.0.3         highr_0.9
[9] stringr_1.4.0      tools_4.2.1         xfun_0.31           sessioninfo_1.2.2
[13] rsvg_2.3.1         utf8_1.2.2          cli_3.3.0           htmltools_0.5.2
[17] ellipsis_0.3.2     yaml_2.3.5          digest_0.6.29       rprojroot_2.0.3
[21] tibble_3.1.7       fontawesome_0.3.0   lifecycle_1.0.1     crayon_1.5.1
[25] vctrs_0.4.1        glue_1.6.2          evaluate_0.15       rmarkdown_2.14
[29] stringi_1.7.6      compiler_4.2.1      pillar_1.7.0        desc_1.4.1
[33] jsonlite_1.8.0     pkgconfig_2.0.3
```



Dieses Skript ist lizenziert unter [Creative Commons Namensnennung - Nicht-kommerziell - Weitergabe unter gleichen Bedingungen 4.0 International](https://creativecommons.org/licenses/by-nc-sa/4.0/).

# **1 Einführung**

## **1.1 Was ist Statistik?**

## **Part I**

# **DATEN ERKUNDEN**

## 2 Daten

### 2.1 Was sind Daten?

### 2.2 Wo kommen Daten her?

#### 2.2.1 Datenerhebung

Of the 224 patients in the treatment group, 45 had a stroke by the end of the first year. Using these two numbers, compute the proportion of patients in the treatment group who had a stroke by the end of their first year. (Note: answers to all Guided Practice exercises are provided in footnotes!)<sup>1</sup>

### 2.3 Testen Sie sich selbst

### 2.4 Übungen

---

<sup>1</sup>The proportion of the 224 patients who had a stroke within 365 days:  $45/224 = 0.20$ .



## **3 Daten visualisieren**

# Bibliographie

Çetinkaya-Rundel, Mine, and Johanna Hardin. n.d. *Introduction to Modern Statistics*. <https://openintro-ims.netlify.app/>.

Ismay, Chester, and Albert Y. Kim. 2021. *ModernDive: Statistical Inference via Data Science*. <https://moderndive.com/>.

Wickham, Hadley. 2020. *Ggplot2: Elegant Graphics for Data Analysis*. 3rd, in progress.

Wickham, Hadley, and Garrett Golemund. 2021. *R for Data Science*. <https://r4ds.had.co.nz/>.