

# Statistik

CH.1 - Grundlagen

2024 || Prof. Dr. Buchwitz, Sommer, Henke

Wir geben Impulse

## Benjamin Buchwitz

- B.Sc. Business Informatics, Berlin School of Economics and Law.
- M.Sc. Management Science, University of Eichstätt-Ingolstadt.
- Ph.D. (Doctorate) at the Chair of Statistics and Quantitative Methods, University of Eichstätt-Ingolstadt.
- Work Experience in Consulting, Information Technology and Renewable Energy Sector, e.g. for IBM, Siemens and own Company.

## Prof. Dr. Benjamin Buchwitz

- Professur für Data Science und Data Engineering
- Raum 2.3.14, Lindenstr. 53, Meschede
- **Email:** buchwitz.benjamin@fh-swf.de



- 1 Deskriptive Statistik
- 2 Wahrscheinlichkeitsrechnung
- 3 Induktive/Analytische Statistik

- Häufigkeiten
- Graphische Darstellungen
- Tabellen
- Lagemaße
- Streuungsmaße
- Konzentrationsmaße
- Korrelationskoeffizient

- Angenommen Sie besitzen ein Wertpapier (z.B. Bitcoin), das in den letzten Jahren die folgenden Wertveränderungen durchlaufen hat. Ermitteln Sie die durchschnittliche Wertsteigerung dieses Wertpapiers.

| Jahr | Veränderung |
|------|-------------|
| 2019 | 12 %        |
| 2020 | 7 %         |
| 2021 | 1 %         |
| 2022 | 4 %         |
| 2023 | -10 %       |

# Beispiel: Deskriptive Statistik

- Bei einer Befragung von 1195 über 14jährigen Deutschen durch das Institut **promit** im Februar 2011 tippten 51,5 % auf Borussia Dortmund, 15,8 % auf Bayern München, 3,7 % auf FC Schalke 04, 1,5 % auf den Hamburger SV und 1,0 % auf Bayer 04 Leverkusen als Deutschen Fußballmeister 2011. 21,6 % machten keine Angaben, der Rest glaubte an einen anderen Verein als Meister. Stellen Sie die Daten graphisch dar.

Wer wird Deutscher Fußballmeister 2011



- Kombinatorik
- Zufallsexperimente
- Wahrscheinlichkeiten
- Bedingte Wahrscheinlichkeiten
- Zufallsvariablen
- Verteilungsfunktionen
- Spezielle Verteilungen (z.B. Normalverteilung)



- Ein Unternehmen erhält wiederholt Lieferungen von 800 Flaschen zur Verpackung von flüssigem Waschmittel. Mit dem Lieferanten ist vereinbart, dass Lieferungen mit mehr als 2% fehlerhaften Flaschen zurückgewiesen werden dürfen. Aus Zeit- und Kostengründen kann das Unternehmen nicht alle gelieferten Flaschen prüfen. Um zu entscheiden, ob es die Lieferungen zurückweist, verfährt es daher nach folgender Regel:
  - ▶ Der Lieferung werden 50 Flaschen zufällig entnommen und geprüft. Die Lieferung wird zurückgewiesen, wenn mehr als eine Flasche nicht dem vereinbarten Qualitätsstandard entspricht. Wie groß ist die Wahrscheinlichkeit, dass eine gerade noch zulässige Lieferung, d.h. genau 2% fehlerhafte Flaschen, zurückgewiesen wird.

- Stichproben
- Konfidenzintervalle
- Testen von Hypothesen
- Lineare Regression
- Multiple Regression

- Für ein Konzert in Meschede wurden die Karten im Radio verlost. Sie befragen alle 400 Besucher des Konzertes nach ihrem Schokoladenkonsum zu Weihnachten (Dezember 2023). Dabei ermitteln Sie einen durchschnittlichen Schokoladenkonsum von 455g bei einer Standardabweichung von 130g. Mit diesen Daten möchten Sie auf den Schokoladenkonsum der Bevölkerung im Hochsauerlandkreis schließen. Geben Sie an, in welchem Bereich der durchschnittlich Schokoladenkonsum eines HSK-Bürgers mit 90%iger Sicherheit lag.

- 1 VL: Intro, Begriffe, Häufigkeit
- 2 VL: Maßzahlen
- 3 VL: Zusammenhang
- 4 VL: Kombinatorik
- 5 VL: Wahrscheinlichkeitsrechnung
- 6 VL: Anwendungen Wahrscheinlichkeitsrechnung
- 7 VL: Wahrscheinlichkeitsverteilungen
- 8 VL: Spezielle Verteilungen
- 9 VL: Schätzung
- 10 VL: Hypothesentests
- 11 VL: Regression
- 12 VL: Multiple Regression

**Alle Unterlagen finden Sie hier:**

<https://bchwtz.github.io/bchwtz-stat/>

Als Hochschule für *angewandte Wissenschaften* lernen Sie im Modul *Statistik* nicht nur die theoretischen Grundlagen der Statistik kennen, sondern Sie wenden diese auch mit der Software *R* an. Dieses Semester werden die folgenden Veranstaltungen angeboten:

- **Vorlesung:** Vermittlung der theoretischen Grundlagen, erste exemplarische Anwendung der vorgestellten Verfahren inklusive Interpretationslogiken der Ergebnisse.
- **Seminar:** Festigung der Lerninhalte und vertiefte Anwendung der statistischen Verfahren, Arbeiten mit der Statistiksoftware *R*, Vorbereitung der semesterbegleitenden Gruppenarbeit.
- **R-Sprechstunde:** Freie Sprechstunde zur Unterstützung bei allen Aspekten die die Statistiksoftware *R* betreffen.
- **Tutorium:** Vertiefung der rechnerischen Anwendung der statistischen Verfahren. Rechnen von Übungsblättern.

Die Prüfungsleistung im Modul Statistik ist eine Klausur, die durch eine semesterbegleitende Gruppenarbeit ergänzt wird. Die Endnote im Modul Statistik setzt sich aus Gruppenarbeit (25%) und Klausur (75%) zusammen.

- **Klausur (75%):** Prüfung der theoretischen und rechnerischen Grundlagen der Statistik inkl. Interpretation von Ergebnissen. **Hilfsmittel: Taschenrechner + Studienbuch inkl. handschriftlicher Anmerkungen**
- **Gruppenarbeit (25% + PVL):** Prüfung der eigenständigen Anwendung und Umsetzung statistischer Analyselogiken. Für die Gruppenarbeit gibt es drei Abgaben: Exposé, Zwischenabgabe und Endabgabe. Für die Anwendung mit R müssen zudem Kurse der online Lernplattform *DataCamp* nachgewiesen werden. Für die **PVL** sind mindestens 5 Punkte in der Gruppenarbeit zum Stand der Zwischenabgabe zu erreichen und es müssen 3 der 4 vorgegebenen Datacamp-Zertifikaten vorgelegt werden.

## Abschluss des Moduls

Das Modul *Statistik* gilt als bestanden, wenn Sie mindestens 50 Punkte aus der Gruppenarbeit (max. 25 Punkte) und der Klausur (max. 75 Punkte) erreichen.

Für die Gruppenarbeit wenden Sie die erlernten Verfahren mit Hilfe von R an und erstellen mit Hilfe der Analyseergebnisse eine **wissenschaftliche Arbeit**.

- Die Gruppenarbeit besteht aus drei Abgaben: Exposé, Zwischenabgabe und Endabgabe. Jeder Gruppe wird ein Korrektor zugeordnet, der alle Abgaben betreut.
- Für das Erstellen der Gruppenarbeit gibt es eine Vorlage, die es erlaubt, Analyseprogramm (R) und Text in einer Datei zu kombinieren. Diese Vorlage enthält eine Ehrenwörtliche Erklärung und eine Checkliste für die Abgabe zur Vermeidung von häufigen Fehlern.
- Die erzielten Punkte für die Gruppenarbeit gelten für die Prüfung im aktuellen Semester sowie für die Wiederholungsprüfung; erzielte Punkte bleiben also *zwei Semester* lang bestehen. Die PVL bleibt nach einmaligem bestehen **immer** gültig.

**Alle Details zur Gruppenarbeit erfahren Sie im Seminar!**

## R Installieren

<https://cloud.r-project.org/>

## RStudio installieren

<https://www.rstudio.com/products/rstudio/download/#download>

## DataCamp Zugang

Über die Einschreibung bei Moodle erhalten Sie den Zugang zu DataCamp für die begleitenden Online Kurse.



- R ist eine kostenlose Software-Umgebung für statistische Datenanalyse und Graphiken. Es beruht auf einer Implementation der Sprache S. Anfänglich wurde R von Ross Ihaka und Robert Gentleman (Univ. Auckland) entwickelt und wird seit Mitte der 90er Jahre von einem Entwickler-Kollektiv (R-Core) betreut.
- Sehen Sie Statistiksoftware als Hilfsmittel, um dem Computer mitzuteilen, wie Sie Ihre Daten analysieren wollen.
- Für einfache Analysen reichen einfache Programme, aber gerade in Excel haben Sie irgendwann das letzte Untermenü oder die letzte Formel erreicht und können nicht mehr weiter arbeiten.
- R-Befehle sind Ihre Anweisungen an den Computer.

```
Sys.time()
```

```
## [1] "2024-12-31 22:01:53 UTC"
```

```
sessionInfo()
```

```
## R version 4.4.1 (2024-06-14)
## Platform: x86_64-pc-linux-gnu
## Running under: Ubuntu 22.04.5 LTS
##
## Matrix products: default
## BLAS: /usr/lib/x86_64-linux-gnu/openblas-pthread/libblas.so.3
## LAPACK: /usr/lib/x86_64-linux-gnu/openblas-pthread/libopenblaspr0.3.20.so; LAPACK version 3.10.0
##
## locale:
## [1] LC_CTYPE=en_US.UTF-8      LC_NUMERIC=C
## [3] LC_TIME=en_US.UTF-8      LC_COLLATE=en_US.UTF-8
## [5] LC_MONETARY=en_US.UTF-8  LC_MESSAGES=en_US.UTF-8
## [7] LC_PAPER=en_US.UTF-8     LC_NAME=C
## [9] LC_ADDRESS=C             LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
##
## time zone: Etc/UTC
## tzcode source: system (glibc)
##
## attached base packages:
## [1] grid      stats    graphics  grDevices  utils      datasets  methods
## [8] base
##
## other attached packages:
## [1] vcd_1.4-13      lubridate_1.9.3 forcats_1.0.0   stringr_1.5.1
## [5] dplyr_1.1.4     purrr_1.0.2     readr_2.1.5     tidyr_1.3.1
## [9] tibble_3.2.1    ggplot2_3.5.1   tidyverse_2.0.0 texreg_1.39.4
```