

Kapitel 4.5

Einführung in die Statistik

Thorsten Dickhaus
Universität Bremen
Institut für Statistik

Mathematik 3: Stochastik
Universität Bremen, Fachbereich 03, SoSe 2025

Übersicht

- 1 Was ist Statistik?
- 2 Statistik und Wahrscheinlichkeitstheorie
- 3 Das Statistik-Softwaresystem R

Übersicht

- 1 Was ist Statistik?
- 2 Statistik und Wahrscheinlichkeitstheorie
- 3 Das Statistik-Softwaresystem R

Statistik ist die **Lehre von den Daten.**

A Definition of Statistics

Statistics is the science of collecting, describing, and interpreting data, i.e., the tool box underlying empirical research.

In analyzing data, scientists aim to describe our perception of the world. Descriptions of stable relationships among observable phenomena in the form of theories are sometimes referred to as being explanatory. (Though one could argue that science merely describes *how* things happen rather than *why*.) Inventing a theory is a creative process of restructuring information embedded in existing (and accepted) theories and extracting exploitable information from the real world. (We are abstracting from purely axiomatic theories derived by logical deduction.)

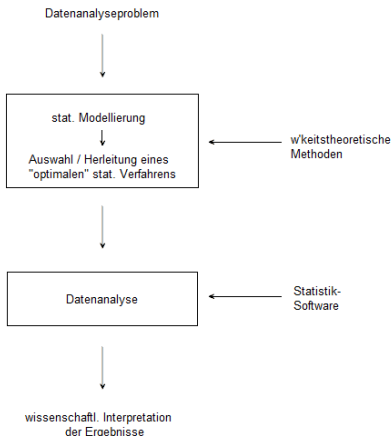
A first exploratory approach to groups of phenomena is typically carried out using methods of *statistical description*.

Härdle et al. (2015): Introduction to Statistics. Springer-Verlag.

Aufgaben der Statistik

- Planung der Datenerhebung, Studiendesign
- Beschreibung und Visualisierung von Daten
- Datenanalyse (schließende Statistik)
- Interpretation von Daten
(im fachwissenschaftlichen Kontext)

Was ist Statistikkforschung?



Übersicht

- 1 Was ist Statistik?
- 2 Statistik und Wahrscheinlichkeitstheorie
- 3 Das Statistik-Softwaresystem R

Modellierung von Daten (I)

Sei y_1, \dots, y_n eine **Datenreihe**,
d. h., eine Abfolge von n Messwerten
(in der Folge meistens: reelle Zahlen).

Da an jeder Datenerhebung **Zufallseinflüsse** beteiligt sind,
denken wir uns y_i als die **Realisierung einer Zufallsvariable** Y_i ,
also als den Wert, den Y_i angenommen hat, $1 \leq i \leq n$.

Die Zahl $n \in \mathbb{N}$ nennt man in der Statistik den
Stichprobenumfang.

Das Tupel $\mathbf{y} = (y_1, \dots, y_n)^\top$ nennt man auch **Stichprobe**.

Modellierung von Daten (II)

Sei $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$.

Die Menge aller Werte, die \mathbf{Y} annehmen kann, bezeichnen wir als den **Stichprobenraum** \mathcal{Y} .

Wir versehen \mathcal{Y} mit einer σ -Algebra $\mathcal{B}(\mathcal{Y})$
(in aller Regel: die Borel'sche σ -Algebra auf \mathcal{Y}).

Modellierung von Daten (III)

Ein **statistisches Modell** besteht aus $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$ sowie einer **Familie von W'keitsverteilungen** $\mathcal{P} = (\mathbb{P}_\theta : \theta \in \Theta)$.

Dabei heißt θ der **Parameter des Modells** und Θ der **Parameterraum des Modells**.

Wir interpretieren θ als den **unbekannten und unbeobachtbaren Zustand der Natur**. Es herrscht also **Parameter-Unsicherheit**.

Ist der wahre Wert des Parameters gleich θ^* , so gilt:

$$\mathbf{Y} \sim \mathbb{P}_{\theta^*}$$

Empirische Verteilung (I)

Ziel der **Inferenzstatistik (schließende Statistik)** ist es, auf der Basis der Daten $\mathbf{y} = (y_1, \dots, y_n)^\top$ Information über den unbekannten Wert von θ bzw. über die Verteilung von \mathbf{Y} zu gewinnen.

Viele statistische Verfahren basieren dabei auf der **empirischen Verteilung**, die durch \mathbf{y} induziert wird.

Dabei wird eine **Punktmasse von $1/n$** in jeden beobachteten Datenpunkt y_1, \dots, y_n gelegt.

(Diskrete Gleichverteilung auf den Daten)

Die empirische Verteilung ist ein **W'keitsmaß!**

Daher lassen sich die Konzepte aus Kapitel 4 auf die empirische Verteilung anwenden.

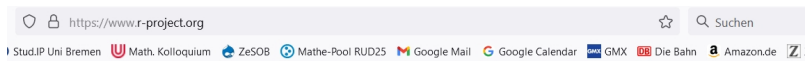
Empirische Verteilung (II)

Viele statistische Verfahren (sowohl beschreibende als auch schließende) basieren auf der empirischen Verteilung:

- Empirische Verteilungsfunktion
- Histogramm
- Empirische Momente
- Empirischer Median, empirische Quantile
- Momentenschätzer
- Parametrische und nichtparametrische Hypothesentests
- etc.

Übersicht

- 1 Was ist Statistik?
- 2 Statistik und Wahrscheinlichkeitstheorie
- 3 Das Statistik-Softwaresystem R



[Home]

Download

[CRAN](#)

R Project

[About R](#)

[Logo](#)

[Contributors](#)

[What's New?](#)

[Reporting Bugs](#)

[Conferences](#)

[Search](#)

[Get Involved: Mailing Lists](#)

[Get Involved: Contributing](#)

[Developer Pages](#)

[R Blog](#)

The R Project for Statistical Computing

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To [download R](#), please choose your preferred [CRAN mirror](#).

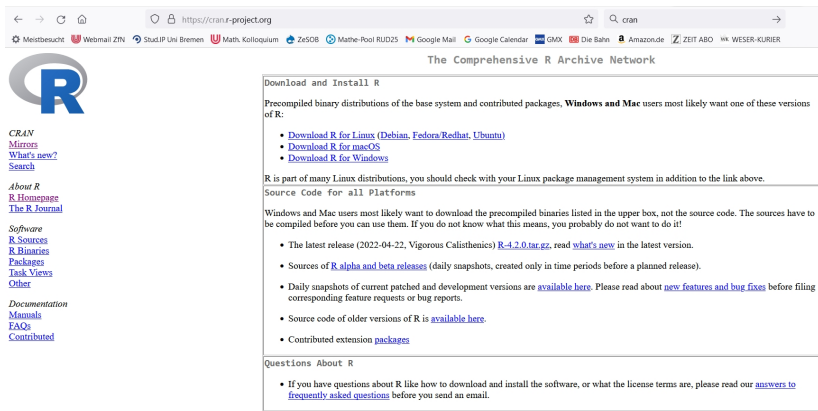
If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

News

- [R version 4.2.0 \(Vigorous Calisthenics\)](#) has been released on 2022-04-22.
- [R version 4.1.3 \(One Push-Up\)](#) was released on 2022-03-10.
- Thanks to the organisers of useR! 2020 for a successful online conference. Recorded tutorials and talks from the conference are available on the [R Consortium YouTube channel](#).
- You can support the R Foundation with a renewable subscription as a [supporting member](#)

News via Twitter

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



The screenshot shows a web browser window with the address bar displaying `https://cran.r-project.org`. The page title is "The Comprehensive R Archive Network". On the left side, there is a large blue "R" logo and a list of links under the heading "CRAN". The main content area is titled "Download and Install R" and contains information about precompiled binary distributions for Windows and Mac. It lists three download links: "Download R for Linux (Debian, Fedora/Redhat, Ubuntu)", "Download R for macOS", and "Download R for Windows". Below this, it states that R is part of many Linux distributions and provides a link to the source code for all platforms. It also mentions that Windows and Mac users should download the precompiled binaries. A list of links is provided for the latest release, sources of alpha and beta releases, daily snapshots, source code of older versions, and contributed extension packages. At the bottom, there is a section titled "Questions About R" with a link to frequently asked questions.

CRAN
[Mirrors](#)
[What's new?](#)
[Search](#)

About R
[R Homepage](#)
[The R Journal](#)

Software
[R Sources](#)
[R Binaries](#)
[Packages](#)
[Task Views](#)
[Other](#)

Documentation
[Manuals](#)
[FAQs](#)
[Contributed](#)

The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux \(Debian, Fedora/Redhat, Ubuntu\)](#)
- [Download R for macOS](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2022-04-22, Vigorous Calisthenics) [R-4.2.0.tar.gz](#), read [what's new](#) in the latest version.
- Sources of [R alpha and beta releases](#) (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are [available here](#). Please read about [new features](#) and [bug fixes](#) before filing corresponding feature requests or bug reports.
- Source code of older versions of R is [available here](#).
- Contributed extension [packages](#)

Questions About R

- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

`https://cran.r-project.org/`