



Einführung in R

Statistik 2: Ziele

- Konzepte von einer breiten Auswahl von Methoden verstehen
- Umsetzung mit R: Daten einlesen, Daten analysieren, Grafiken erstellen und exportieren

Fahrplan

- 1: Einführung in R und RStudio
- 2-8: Erweiterungen der Linearen Regression (Faktoren, Interaktion, GLM, Mixed Effects, ANOVA)
- 9: Kategoriale Daten
- 10: Poweranalyse (Stichprobengrösse)
- 11-12: Design von Experimenten
- 13: Unsupervised Learning (PCA)
- 14: Wiederholung

Administration 1/3: Vorlesung & Übungen

- Homepage: <https://moodle-app2.let.ethz.ch/course/view.php?id=3508>
- Kein Skript; konkrete Buchempfehlungen pro Thema oder Folien
- Quiz: Konzeptfragen & Anwendung in R
- Stil der Übungsstunde: **Laptop** mitbringen, Quiz lösen, Assistent fragen
- R hat hohes Gewicht in dieser Vorlesung
- Zur Wiederholung: etutoR

Administration 2/3: Moodle

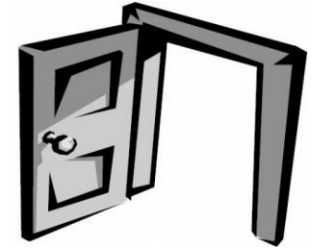
- Folien und Quiz über Moodle:
<https://moodle-app2.let.ethz.ch/course/view.php?id=3508>
Quiz: Konzeptfragen und R Aufgaben
→ sehr ähnlich wie Prüfungsaufgaben
- Wer ein Quiz zu mind. **~90%** richtig hat und innerhalb der Zeitfrist (Mittwoch nächste Woche, 8:00 Uhr) abgibt:
→ Badge
→ Bonus-Material: Prüfungsrelevante Variationen der R Übungen
- Fortschrittsbalken zum besseren Zeitmanagement

Administration 3/3: Prüfung

- 180 min schriftlich am Computer (es soll keine Zeitnot geben)
- Multiple Choice Fragen:
 - Konzepte
 - Datenanalyse
- R soll (wie ein Taschenrechner) verwendet werden um Aufgaben zu lösen
- Prüfungsaufgaben werden ähnlich sein wie Übungsaufgaben
- Hilfsmittel: Eine Zusammenfassung auf 5 Blättern (maximal DIN-A 4; beliebig beschrieben).

Literatur: ISL

“Türöffner“



Springer Texts in Statistics
Volume 103 2013

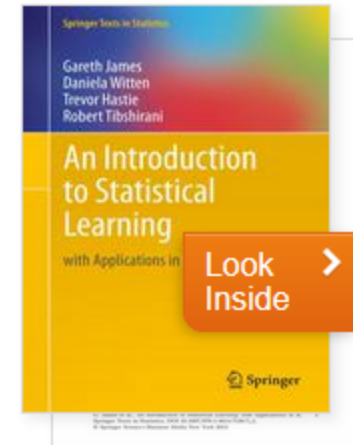
An Introduction to Statistical Learning with Applications in R

Authors: Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani

ISBN: 978-1-4614-7137-0 (Print) 978-1-4614-7138-7 (Online)



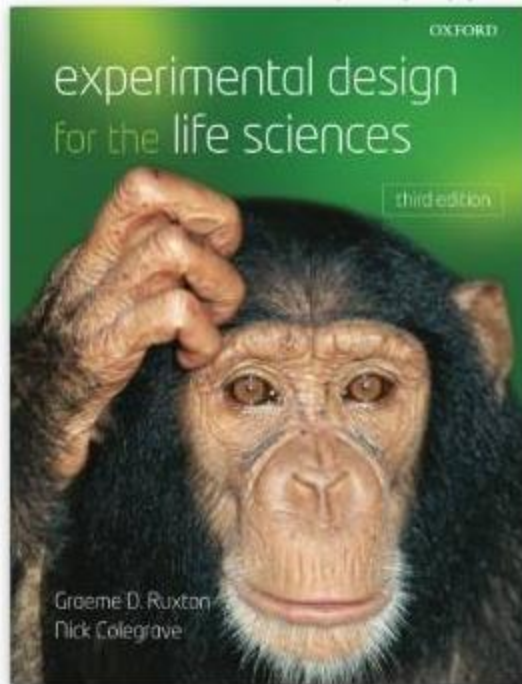
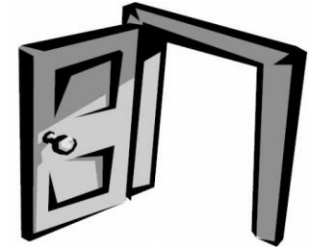
Download Book (10,958 KB)



Online erhältlich via ETH Bibliothek, wenn man im ETH Netzwerk ist.

Literatur: **Experimental Design**

“Türöffner“



Taschenbuch: 196 Seiten

Verlag: Oxford University Press, USA; Auflage: 0003 (28. November 2010)

Sprache: Englisch

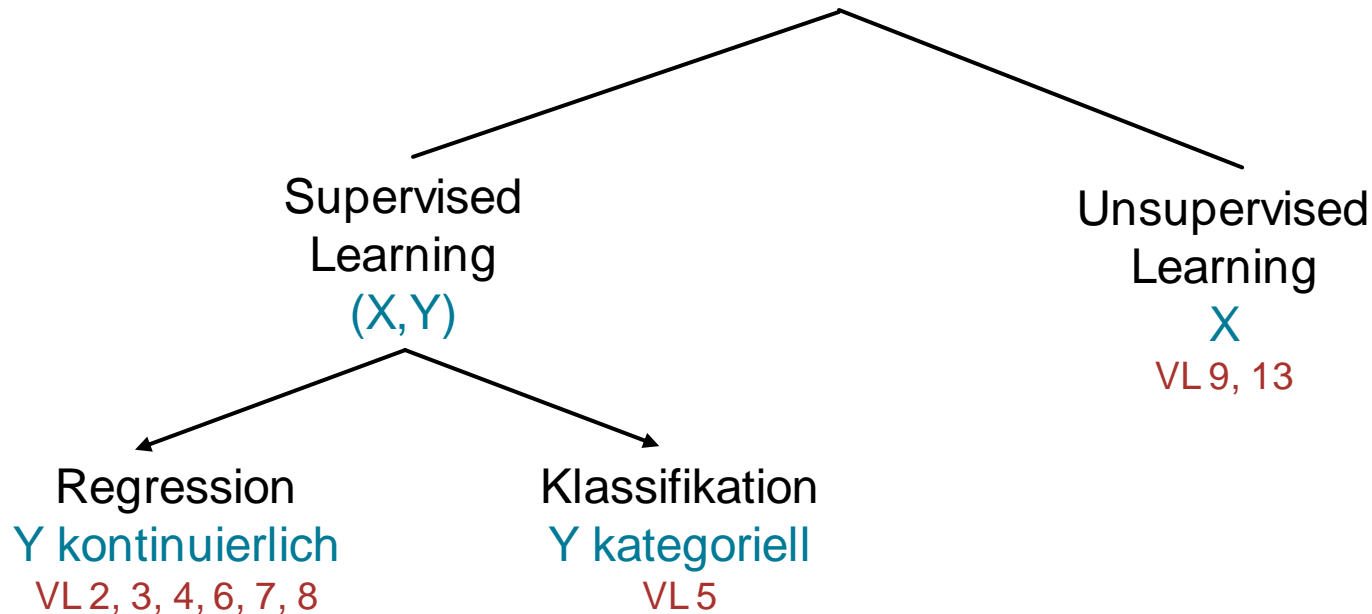
ISBN-10: 0199569126

ISBN-13: 978-0199569120

Buch nur als Ergänzung

Via ETH Bibliothek erhältlich (leider nicht online)

Big Picture: Statistisches Lernen



Exkurse: VL 10 (Poweranalyse), VL 11 & 12 (Experimentelles Design)

Einführung in R

- R: Programmiersprache
RStudio: Guter Editor für R
- Comprehensive R Archive Network (CRAN):
Pakete, Task Views, Manuals
- Daten einlesen (csv, rda, rds)
- Daten bearbeiten
- Grafiken erstellen und exportieren

R ≠ R-Studio

Software R: <https://cran.r-project.org/>



- Was sind R und CRAN?

R: “R is [‘GNU S’,] a freely available language and environment for statistical computing and graphics which provides a wide variety of statistical and graphical techniques”

CRAN: Comprehensive R Archive Network

“CRAN is a network of ftp and web servers around the world that store identical, up-to-date, versions of code and documentation for R. Please use the CRAN [mirror](#) nearest to you to minimize network load.”

- Basis Software + mehrere tausend “packages”
- Manual: “An Introduction to R” (<https://cran.r-project.org/manuals.html>)
- Task Views (<https://cran.r-project.org/web/views/>)



Editor RStudio

“RStudio IDE is a powerful and productive user interface for R. It’s free and open source, and works great on Windows, Mac, and Linux.”


Alternativen: Emacs, Tinn-R, ...

Wir verwenden in der Vorlesung RStudio.

Daten einlesen: Datentypen


- **CSV:**
Comma Separated Values für Daten
- **RDA oder Rdata:**
Komprimiertes R-Datenfile für mehrere R-Objekte mit Objektnamen
- **RDS:**
Komprimiertes R-Datenfile für EIN R-Objekt ohne Objektnamen

CSV: Comma Separated Values

- Funktionen: *read.csv()*, *write.csv()*
RStudio:  (verwendet Paket *readr*)
- Gut für Datenübertragung zwischen verschiedenen Programmen (z.B. von Excel nach R)
- + Flexibel, geht bei praktisch jedem Programm zur Datenverarbeitung
- Nur Datentabellen können einfach gespeichert werden (keine anderen komplexeren R Objekte)
- Nicht komprimiert, daher grosse Datenmenge und langsam zu laden

Bem.: Andere Trennzeichen sind auch möglich (; space tab)

RDA oder Rdata: All objects

- Funktionen: `load()`, `save()`, `save.image()`
RStudio: The icon shows a yellow folder with a green arrow pointing into it, and a blue floppy disk icon next to it, representing the save function in RStudio.
- Gut, wenn man alle aktuellen R-Objekte speichern will und man selbst oder ein Kollege weiterarbeiten soll
- + Komprimiert; speichert ALLE R-Objekte ab
- Andere Programme zur Datenverarbeitung können das Format nicht lesen
- Namen von Objekten können nicht einfach geändert werden

RDS: Single Objects

- Funktion: `readRDS()`, `saveRDS()`
RStudio: ??
- Gut um einzelne R-Objekte komprimiert zu speichern und beim Laden einen neuen Namen geben zu können
- + Komprimiert, schnell zu laden
- + Namen von Objekten können einfach geändert werden
- Andere Programme zur Datenverarbeitung können das Format nicht lesen

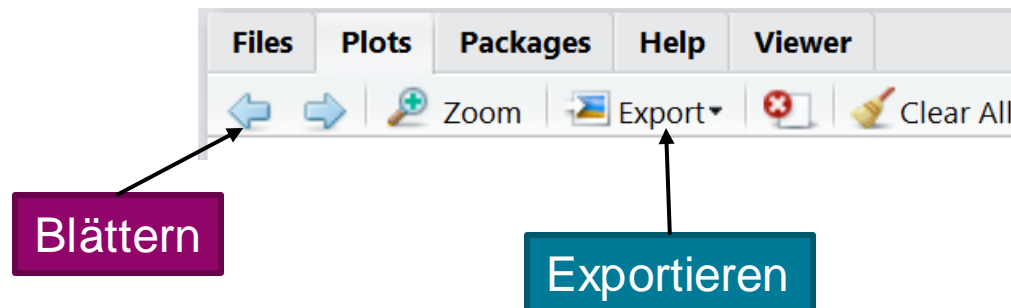
Ausblick: Pakete *foreign*; Manual *R Data Import/Export*

Daten bearbeiten

- Datentypen: numeric, character, logical, factor
- Container:
 - vector `v`, matrix `m` – überall gleicher Datentyp
 - list `l` – beliebige Datentypen pro Eintrag
 - `data.frame d` – Liste aus Vektoren
 - gleicher Datentyp pro Spalte = “**Datentabelle**”
- `str()`: Struktur des Objekts
- Zugriff auf Datenelemente
 - Index: `vec[3]`, `mat[2,3]`
 - Spaltenname: `d$name`
 - Auswahl mit logischem Vektor

Grafik erstellen und exportieren

- plot, points, line
- Per default: Ausgabe auf Bildschirm
- Ausgabe in File:
 - 1) File öffnen: bmp(), jpeg(), tiff(), pdf() oder postscript()
 - 2) Plot erzeugn
 - 3) File schliessen: dev.off() (“device off”)
- RStudio:



Nächstes Mal

- Lineare Regression für Fortgeschrittene

