Applied Data Analysis (ADA)

Summer Term 2022

Maria Kateri

Institute of Statistics
Pontdriesch 14-16

maria.kateri@rwth-aachen.de
www.isw.rwth-aachen.de

General information (please read carefully!)

Preliminary note

- We will run the course in presence
  (provided it is allowed by the regulations for universities due to the pandemic).
- If needed, we will switch to an online format.
- The ADA lectures will also be provided as slidecasts on a weekly basis (recorded SoSe 2020) .

Course information will be continuously updated.

Course elements, dates & rooms

Lectures	Tu$^{10.30\text{-}12.00}$	Start: April, 5	Hörsaal I
	We$^{10.30\text{-}11.15}$		Hörsaal I
Tutorial	Th$^{10.30\text{-}12.00}$	Start: April, 14 (every two weeks)	Hörsaal I
R-lab	**Introduction to** R: Th$^{10.30\text{-}12.00}$		April, 7
	Mo$^{14.30\text{-}16.00}$, Tu$^{12.30\text{-}14.00}$, Tu$^{14.30\text{-}16.00}$, We$^{14.30\text{-}16.00}$		Start: April, 11
further details (e.g., time shifts)	see RWTHonline, RWTHmoodle		

General information

Contact

Please contact us by email:

- M. Kateri: maria.kateri@rwth-aachen.de
- W. Herff: herff@isw.rwth-aachen.de (Tutorial)
- L. Kaufmann: kaufmann@isw.rwth-aachen.de (R-Lab)
- C. Queckenberg: clemens.queckenberg@rwth-aachen.de (R-Lab)

Regular office hours (as Zoom meetings upon request)

M. Kateri	Tuesday	13.30-14.30
W. Herff	Tuesday	12.30-13.30
L. Kaufmann	Tuesday	09.00-10.00

Information & teaching material

A preliminary note

Self-organized and active learning is very important for successful course participation. However, online material cannot replace a lecture in presence. Thus, we strongly recommend to study the material carefully AND to attend the lectures, tutorials and R-Labs.

Access to teaching material

- **RWTHmoodle-class room** corresponding to the lecture (**https://moodle.rwth-aachen.de**)
  - Lecture slides, videos, literature, exercise sheets, updated information, etc.
  - R-lab sheets, R-code, etc.
  - communication (e.g., via email)
  - information on exams
- website of Institute of Statistics: **www.isw.rwth-aachen.de**

Access to RWTHmoodle

- Use registration to the lecture *Applied Data Analysis (Vorlesung)* via RWTHonline
- The registration is only open to students of MSc. Data Science & MSc. Mathematik!

Course concept

Prerequisites

In order to take successfully part in the course, we assume knowledge of

- probability & statistics (e.g., at least level of course *Mathematics of Data Science*),
- linear algebra & optimization,
- programming in R (on an introductory level).

Note: The lectures are of mathematical nature. Thus, a (profound) mathematical background is strongly recommended!

Teaching concept

- Contents will be provided in the lectures.
- Every two weeks, you will get an exercise sheet with theoretical problems (over all, there will be 6 sheets). A solution of the problems will be provided one week later.
- Each week, you will get programming problems (in R).

Exam & Exam admission

Admission to the exam

In order to be admitted to the exam,

1. you have to be enrolled in one of the programs
   - MSc. Data Science
   - MSc. Mathematik

   Students of other programs can not be admitted to the exam!
2. you have to **take part in electronic tests** which are offered on **May, 13; May, 27; July, 1; July, 15**. These e-tests are based on the lectures as well as on the problems given in both the theoretical and the programming part (R-lab) of the course. Details will be given in due time.
3. you need a **minimum score of 40%** of the potential points to be admitted to the exam. Furthermore, you can earn bonus points for the exam (see next slide).

Exam (Details will be given in due time)

- The written exam will be conducted either in a lecture hall or online (with Zoom proctoring) depending on the constraints valid at that time.
- The exam will be an open book exam.
- Tentative dates: PT1 (July, 28); PT2 (September, 15)

Bonus points

- Depending on the outcome of the e-tests you will earn the following bonus points for the exam:

result e-test (in %)	bonus points
$[40\%, 50\%)$	0
$[50\%, 60\%)$	2
$[60\%, 70\%)$	3
$[70\%, 80\%)$	4
$[80\%, 100\%]$	5

- The bonus points will only be added to the points obtained in the exam if you satisfy the minimum requirement to pass the exam.

Important Notice

The admission as well as the bonus points are only valid for the PT1 and PT2 exams of summer term 2022. They are not transferable to future semesters.

References Prerequisites & Data Analysis in R

Prerequisites

- Carlton, M.A., Devore, J.L. (2017). Probability with Applications in Engineering, Science, and Technology. 2nd edn., Springer, NY.
- Casella, G., Berger, R.L. (2002). Statistical Inference, 2nd edn, Duxbury Thomson Learning, Pacific Grove, CA.
- Cramer, E., Kamps, U. (2020). Grundlagen der Wahrscheinlichkeitsrechnung und Statistik. 5th edn, Springer, Heidelberg (in German).
- DasGupta, A. (2011). Probability for Statistics and Machine Learning. Springer, NY.
- Seber, G.A.F. (2008) A Matrix Handbook for Statisticians. John Wiley & Sons, Hoboken, NJ.
- Shorack, G.R. (2017). Probability for Statisticians. Springer, NY. (*advanced*, measure theoretic approach)

Data Analysis in R

- Agresti A., Kateri, M. (2022). Foundations of Statistics for Data Scientists: With R and Python, CRC Press, Boca Raton.
- Hothorn, T., Everitt, B. S. (2014). A Handbook of Statistical Analyses Using R, 3rd edn, CRC Press, Boca Raton.
- James, G., Witten, D., Hastie, T., Tibshirani, R. (2013). An Introduction to Statistical Learning with Applications in R, Springer, NY.
- Ugarte, M.D., Militino, A.F., Arnholt, A.T. (2015). Probability and Statistics with R, 2nd edn, CRC Press, Boca Raton.

References Part I: Linear Models & Part II: Generalized Linear Models

Linear Models

- Christensen, R. (2020). Plane Answers to Complex Questions: The Theory of Linear Models. 5th edn., Springer, Heidelberg.
- Härdle, W.K., Simar, L. (2015). Applied Multivariate Statistical Analysis. 4th edn., Springer, Heidelberg.
- Rencher, A.C., Christensen, W.F. (2012). Methods of Multivariate Analysis. John Wiley, Hoboken, NJ.
- Rencher, A.C., Schaalje, G.B. (2008). Linear Models in Statistics, 2nd edn., John Wiley, Hoboken, NJ.

Generalized Linear Models

- Agresti, A. (2015). Foundations of Linear and Generalized Linear Models, John Wiley, Hoboken, NJ.
- Dobson, A.J., Barnett, A.G. (2008). An Introduction to Generalized Linear Models, 3rd edn, Chapman and Hall /CRC Press, NY.
- Dunn, P.K., Smyth, G.K. (2018). Generalized Linear Models With Examples in R, Springer, NY.
- Kateri, M. (2014). Contingency Table Analysis: Methods and Implementation Using R, Birkhäuser/Springer, NY.
- McCullagh, P., Nelder, J.A. (1989). Generalized Linear Models, 2nd edn. Chapman and Hall, London.

Preliminary Table of Contents

(may be updated during the course)

Part I: Linear Models

Part II: Generalized Linear Models