Learning Materials for R and Statistics and Mixed Stuff

by Julius Fenn (julius.fenn@psychologie.uni-freiburg.de). If you have any additional materials I could add, feel free to write me an E-Mail.You will find explanations of the materials in the respective sections.

Possible Learning Process

Note on difficulty to start learning programming / statistics. It is often very difficult to teach statistics or programming, because quite often you are faced - in my personal experience - with negative beliefs (*Glaubensätze*), that "I cannot understand such a difficult subject" or most students experience statistics anxiety, and "statistics is one of the most anxiety-inducing courses in their programmes of study" (Macher et al., 2012, p. 195). explain further blub Julius Fenn

This statistic anxiety (in my personal opinion) is even exacerbated by a culture that not values failures and do not see the power of vulnerability[1]. Learning programming or statistics is a journey of constant failures (you will not be able to debug your code, you will not be able to solve a statistical exercise), but if you appreciate this process of failing and standing up again, you will become great in these subjects.

Focus on the essential. Before you start reading the this paragraph ask yourself the following two questions:

1. Are you investing in the right activities?
2. Are you choosing only those options, which are truly essential for your own development?

By no means I promote here that you should become egoistic or should lack solidarity, but you should truly focus on "The Disciplined Pursuit of Less" (see McKeown, 2014). If you want to become a great therapist (in my opinion) you should focus on clinical literature, therapeutic manuals and inspiring social experiences and not on programming or statistics. But if you want to become a great researcher (or data scientist), who is able to analyze complex data sets and communicate her / his results to the broader public, you should focus on learning programming and statistics as early as possible.

Possible Learning Process. To learn programming / statistics you should focus on flow experiences and choose tasks which are not too easy (you might experience apathy or boredom), and not too hard (you might feel anxious)[2]. So no one, not knowing you, can tell

[1] See YouTube Video "The power of vulnerability by Brené Brown": https://www.youtube.com/watch?v=iCvmsMzlF7o; last accessed on Dec. 20th, 2022
[2] See Flow Theory, e.g. https://spencerauthor.com/flow-theory/; last accessed on Dec. 20th, 2022

you what you should focus on or what is your learning speed, and that's why (in my opinion) **becoming a self-taught person (autodidact)** is the only way to really learn this skills: **set up your learning plan as a project**, which defines the sequence of next steps to learn a certain topic (see Allen, 2015).

As an example, imagine your "project" is to learn fundamentals in statistics: before you start your project you think about how to organize your learning process and decide to use Anki and Obsidian as learning tools (see section "Knowledge management and learning"). As the next step you want to get an overview about the principles of a general data analysis pipeline and the classical statistical methods taught in Psychology (see section "Data Analysis Pipelines"). Thereby you follow the principle of the inverted pyramid, whereby you start learning the most substantial / abstract ("getting big picture") and later focusing on more specific content. Now you are ready to start learning fundamentals in statistics, a reasonable working process is to work through classical text books and their accompanying workbooks (see section "Introductory Books to Statistics"). Almost always start learning statistics, you will learn something about univariate distributions, multivariate distributions, probabilities, discrete and continuous random variables and so on (just look at the table of contents of one of the recommended introductory books). If you want to additionally learn the statistical software R to apply your knowledge (e.g. compute a multiple linear regression) you can have a look in the "Introductory books to R" and the "Self-learning courses for R".

Learning R

Remark: the following sections contain materials / links to resources to teach yourself R (statistical software). R can be downloaded here:

- *base R: https://cran.r-project.org/*
- *R Studio (interface): https://posit.co/*

Introductory books to R

Base R / Fundamentals

- easy readable introduction R and statistics: Field, Andy, Jeremy Miles, and Zoë Field. *Discovering Statistics Using R*. SAGE, 2012.
- Resource for improving coding skills and deepening (technical) understanding of R:Wickham, Hadley. *Advanced R, Second Edition*. CRC Press, 2019. https://adv-r.hadley.nz/.

tidyverse (set of packages)

Tidyverse is a collection of R packages designed for data science:

- Wickham, Hadley, and Garrett Grolemund. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. O'Reilly Media, Inc., 2017. https://r4ds.had.co.nz/.

R Markdown (package)

An R Markdown file enables to save and execute code, and generate high quality reports that can be shared with an audience:

- Xie, Yihui, J. J. Allaire, and Garrett Grolemund. *R Markdown: The Definitive Guide*. New York: Chapman and Hall/CRC, 2018. https://doi.org/10.1201/9781138359444.

Self-learning courses for R

- Free R introduction self learning course:

  https://www.datacamp.com/courses/free-introduction-to-r

  - to get free datacamp account:

    https://www.youtube.com/watch?v=sVNOXvE6UoI

- Codecademy: https://www.codecademy.com/catalog

Learning Statistics

Remark: the following sections contain materials / links to resources to teach yourself statistics.

Introductory Books to Statistics

German:

- Fahrmeir, Ludwig, Christian Heumann, Rita Künstler, Iris Pigeot, and Gerhard Tutz. *Statistik: Der Weg zur Datenanalyse*. Springer-Verlag, 2016. https://link.springer.com/book/10.1007/978-3-662-50372-0.
    - accompanying workbook: Caputo, Angelika, Ludwig Fahrmeir, Rita Künstler, Stefan Lang, Iris Pigeot-Kübler, and Gerhard Tutz. *Arbeitsbuch Statistik*. Springer-Verlag, 2009. https://link.springer.com/book/10.1007/978-3-540-85083-0.
- Zucchini, Walter, Andreas Schlegel, Oleg Nenadic, and Stefan Sperlich. *Statistik für Bachelor- und Masterstudenten: Eine Einführung für Wirtschafts- und Sozialwissenschaftler*. Springer-Verlag, 2009. https://link.springer.com/book/10.1007/978-3-540-88987-8.
    - accompanying workbook: Böker, Fred, Stefan Sperlich, and Walter Zucchini. *Statistikübungen für Bachelor- und Masterstudenten: Ein Arbeitsbuch mit einer Einführung in R*. Springer-Verlag, 2013. https://link.springer.com/book/10.1007/978-3-642-34788-7.

Self-learning courses for Statistics

- Free Online Courses by Harvard, MIT, & more: https://www.edx.org/
- Stanford Online: https://online.stanford.edu/
    - Course “Statistical Learning”: Links:
        - https://online.stanford.edu/courses/sohs-ystatslearning-statistical-learning

        - https://www.edx.org/course/statistical-learning?index=product&queryID=d62708c3eacd3c6e877519ecb03b9451&position=1
        - accompanying book: https://hastie.su.domains/ElemStatLearn/
- MIT OpenCourseWare (> 2200 courses): https://ocw.mit.edu/index.htm
    - possible useful courses in descending difficulty:
        - Introduction to Probability and Statistics
        - Probability and Random Variables
        - Statistics for Applications
- Kahn Academy: https://www.khanacademy.org/
    - Statistics and probability

Advances Books to Statistics

Remark: the following sections contain materials / links to resources to teach yourself advanced statistical models / theories. Regression models are a general class of models to analyze linear relationships between predictor and outcome variables. Test theory is the fundamental theory to analyze survey data.

Regression Models

General books:

- Fahrmeir, Ludwig, Thomas Kneib, Stefan Lang, and Brian Marx. *Regression: Models, Methods and Applications*. Springer Science & Business Media, 2013.
- Fox, John. *Applied Regression Analysis and Generalized Linear Models*. SAGE Publications, 2016.

Specific books:

- for mixed models / longitudinal data:

- Hoffman, Lesa. *Longitudinal Analysis: Modeling Within-Person Fluctuation and Change*. New York: Routledge, 2015. https://doi.org/10.4324/9781315744094.
- Hox, Joop J. *Multilevel Analysis: Techniques and Applications*. 2. ed. Quantitative Methodology Series. New York: Routledge, Taylor & Francis, 2010.

Test Theory / Models for Analyzing Survey Data

Test Theory

General books:

- Moosbrugger, Helfried, and Augustin Kelava, eds. *Testtheorie und Fragebogenkonstruktion*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2020. https://doi.org/10.1007/978-3-662-61532-4.
- Schmidt-Atzert, Lothar, Stefan Krumm, and Manfred Amelang, eds. *Psychologische Diagnostik*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2021. https://doi.org/10.1007/978-3-662-61643-7.
- Keith, Markus A., and Denny Borsboom. *Frontiers of Test Validity Theory : Measurement, Causation, and Meaning*. Routledge, 2013. https://doi.org/10.4324/9780203501207.

Structural Equation Models

General books:

- Bollen, Kenneth A. *Structural Equations with Latent Variables*. John Wiley & Sons, 1989.
- Hoyle, Rick H. *Handbook of Structural Equation Modeling*. Guilford Press, 2012.
- Kline, Rex B. *Principles and Practice of Structural Equation Modeling, Fourth Edition*. Guilford Publications, 2015.

Latent Class / Profile Analysis

General books:

- Collins, Linda M., and Stephanie T. Lanza. *Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences*. John Wiley & Sons, 2010.

Item Response Theory

General books:

Mixed

Remark: the following sections contain materials / links for three topics:

- *Data Analysis Pipelines: fundamentals how data is generally processed and overview article of "classical" statistical models*
- *Knowledge management and learning: recommendations how to set up your projects, organize your knowledge and learn*
- *Learn other programming languages:*
    - *to collect data you could use lab.js or SoSci Survey*
    - *to program web pages you should learn HTML, CSS, Java Script*
    - *to run statistical analysis you could also use Mplus (Latent Variable Modeling) and Python (Big Data, neural networks, text data, ...)*

Data Analysis Pipelines

- Peng, Roger D., and Elizabeth Matsui. *The Art of Data Science: A Guide for Anyone Who Works with Data*. Lulu.com, 2016. https://bookdown.org/rdpeng/artofdatascience/.

Knowledge management and learning

knowledge management:

- Microsoft OneNote: note-taking software developed by Microsoft, part of the Microsoft Office suite
- open source alternative: Obsidian (https://obsidian.md/): knowledge base that works on local Markdown files and you can represent your notes as a graph; for tutorials see YouTube

To learn stuff:

- Anki (https://apps.ankiweb.net/): free and open-source flashcard program using spaced repetition, a technique from cognitive science for fast and long-lasting memorization

Learn other programming languages

Remark: blub Julius Fenn

to collect data

lab.js

- Henninger, Felix, Yury Shevchenko, Ulf K. Mertens, Pascal J. Kieslich, and Benjamin E. Hilbig. "Lab.Js: A Free, Open, Online Study Builder." *Behavior Research Methods* 54, no. 2 (April 1, 2022): 556–73. https://doi.org/10.3758/s13428-019-01283-5.

SoSci Survey

- tutorial videos: https://www.soscisurvey.de/de/screencast

to program web pages

- freeCodeCamp: https://www.freecodecamp.org/
- codecademy: https://www.codecademy.com/catalog
- Learn Java Script: https://eloquentjavascript.net/

to run statistical analysis

Mplus:

- Hallquist, Michael N., and Joshua F. Wiley. "MplusAutomation: An R Package for Facilitating Large-Scale Latent Variable Analyses in Mplus." *Structural Equation Modeling: A Multidisciplinary Journal* 25, no. 4 (July 4, 2018): 621–38. https://doi.org/10.1080/10705511.2017.1402334.
- Muthén, Linda K, and Bengt O Muthen. *Mplus User's Guide. Eighth Edition.* Muthén & Muthén, 2017. https://www.statmodel.com/html_ug.shtml.
- Wang, Jichuan, and Xiaoqian Wang. *Structural Equation Modeling: Applications Using Mplus*. John Wiley & Sons, 2020.

Python:

- learn Python online course on YouTube:

  https://www.youtube.com/watch?v=rfscVS0vtbw

Additional References

Allen, D. (2015). *Getting Things Done: The Art of Stress-Free Productivity*. Penguin.

Macher, D., Paechter, M., Papousek, I., & Ruggeri, K. (2012). Statistics anxiety, trait anxiety, learning behavior, and academic performance. *European Journal of Psychology of Education*, *27*(4), 483–498. https://doi.org/10.1007/s10212-011-0090-5

McKeown, G. (2014). *Essentialism: The Disciplined Pursuit of Less*. Random House.