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

Possible Learning Process. As with any new skill, when you 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). So no one (unless they know you pretty well maybe) can tell you what you should focus on or what your learning speed is, and that's why (in my opinion) **becoming a self-taught person (autodidact)** is the only way to really learn these skills: I suggest that you **set up your learning plan as a project**, which defines the sequence of 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 might want to think about how to organize your learning process and decide to use Anki and Obsidian as learning tools (see section "Knowledge management and learning"). The next step could be that you want to get an overview about the classical statistical methods taught in Psychology and the principles of a general data analysis pipeline (see section "Data Analysis Pipelines"). Therefore you might follow a principle called "the inverted pyramid", in which you start off by learning the most substantial / general content first ("getting the bigger picture") while focusing on more specific content later on. Even if you do not feel like you have a lot of prior knowledge in the field of statistics, you have nevertheless got everything it takes to start learning fundamentals. A reasonable working process is to work through classical text books and their accompanying workbooks (see section "Introductory Books to Statistics"). When starting to learn about statistics, you will almost always learn something about univariate distributions, multivariate distributions, probabilities, discrete and continuous random variables and so on (just take a 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".

Don't be afraid to learn statistics! It is often very difficult to teach statistics or programming, because quite often you are faced – at least in my personal experience - with negative beliefs (Glaubenssätze). You might think that "I cannot understand such a difficult subject", most students even experience statistics anxiety. You are not alone with this! Macher et al. (2012) found that "statistics is one of the most anxiety-inducing courses in their programmes of study" (Macher et al., 2012, p. 195). Anxiety can impede the ability to receive

and to learn statistical concepts or to concentrate on the subject, even memory processes can be impaired (Onwuegbuzie & Wilson, 2003). The most effective intervention is to address the **incorrect perception that strong mathematical ability is required** to do well in statistics and to emphasize that statistic teachers are not "abstract" or "inhuman" (Siew et al., 2019)

This statistics anxiety (in my personal opinion) is even exacerbated by a culture that does not value failures and does 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.

If you know what you want, focus on the essential. Nobody will learn advanced statistical methods overnight, but if you feel like you want to become better at it, focus on the essentials. Somebody who is aiming to become a clinical psychologist might invest a lot of time in reading clinical literature and therapeutic manuals, inspiring social experiences and broadening their communicative skills. If you want to become a great researcher or data scientist, who is able to analyze complex data sets and communicate your results to the broader public, you might want to focus on learning about statistics and how to program. By focusing on what truly matters to you and investing some time in the activities that will serve your personal development towards this goal, you might even find joy in reading the book „The Disciplined Pursuit of Less" (McKeown, 2014).

[1] See YouTube Video "The power of vulnerability by Brené Brown": https://www.youtube.com/watch?v=iCvmsMzlF7o; last accessed on Dec. 20th, 2022

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:

- Fischer, Gerhard H., and Ivo W. Molenaar. *Rasch Models: Foundations, Recent Developments, and Applications*. Springer Science & Business Media, 1995.
- Rost, Jürgen. *Lehrbuch Testtheorie - Testkonstruktion*. Huber, 2004.
- Koller, Ingrid, Rainer Alexandrowicz, and Reinhold Hatzinger. *Das Rasch Modell in der Praxis*. UTB GmbH, 2012.

Mixed

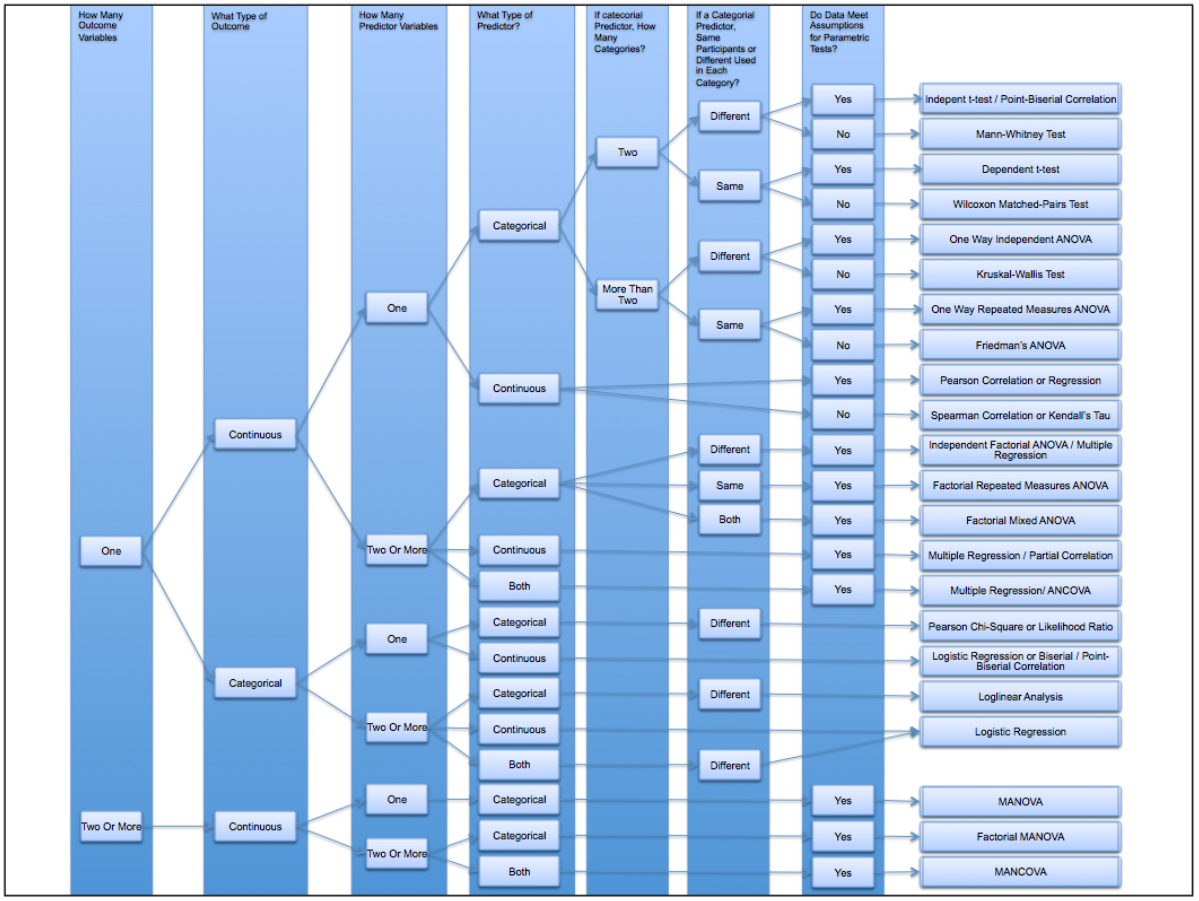
Remark: the following sections contain materials / links for three topics:

- *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/.
- Leek, Jeffery T., and Roger D. Peng. "What Is the Question?" *Science* 347, no. 6228 (March 20, 2015): 1314–15. https://doi.org/10.1126/science.aaa6146.

If you want to get an overview of possible statistical procedures you can have a look at "decision trees"[2]:



see also final pages in (Field et al., 2012) or in german "ETH Zürich Methodenberatung" (https://www.methodenberatung.uzh.ch/de/datenanalyse_spss.html)

Knowledge management and learning

knowledge management:

- Microsoft OneNote: note-taking software developed by Microsoft, part of the Microsoft Office suite

[2] from webpage: https://www.utwente.nl/en/bms/m-store/step-by-step-guide/statistical-tests/choice-test-2/; last accessed on Jan. 9th, 2023

- 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

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.

Field, A., Miles, J., & Field, Z. (2012). *Discovering Statistics Using R*. SAGE.

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.

Onwuegbuzie, A. J., & Wilson, V. A. (2003). Statistics Anxiety: Nature, etiology, antecedents, effects, and treatments--a comprehensive review of the literature. *Teaching in Higher Education*, *8*(2), 195. https://doi.org/10.1080/1356251032000052447

Siew, C. S. Q., McCartney, M. J., & Vitevitch, M. S. (2019). Using network science to understand statistics anxiety among college students. *Scholarship of Teaching and Learning in Psychology*, *5*(1), 75–89. https://doi.org/10.1037/stl0000133