

Quantitative Methods 1, ZHAW

Jürgen Degenfellner

2025-11-03

Contents

1	Introduction	5
1.1	Books I can recommend:	6
1.2	6
1.3	Additional Tools	7
1.4	Workflow suggestion	7
1.5	Orientation for the course and script	8
1.6	Warning of incompleteness	9

Chapter 1

Introduction

This script is a collection of notes and exercises for the course “Quantitative Methods 1” (Codename for: Statistical foundations) at ZHAW in Winterthur, Switzerland. It is permanently evolving and the github repository is public.

The **vision** is to write this script in **collaboration with students**. If you find any errors, have suggestions, or want to contribute, feel free to contact me. Which (online) content (video, book, blog...) helped you understand the topic at hand better? We should link those resources in the script!

The **goal** of this script is to provide a starting point for further reading and learning.

Feel free to use any content for your own purposes.

At the end of each chapter, you will find a list of exercises. A previous exam will be uploaded in time in the respective Moodle folder to help you prepare.

The most important lessons you'll learn are not part of the final exam: Intellectual honesty and humility.

A lot of content relevant for this course can be found in the ZHAW_teaching folder, which contains many R scripts (among other stuff).

We will focus a lot on descriptive statistics and basic concepts as it is a very important part of understanding data. Statistical modeling will be covered later.

One thing to consider for health sciences with respect to quantitative methods is the following: We do not *have* to use quantitative methods for each and every question arising in applied research (physiotherapy, midwifery, nursing, occupational therapy). Small sample sizes (e.g., $n = 10$) often do not warrant the use of inferential statistics or estimation or at least make it considerably harder to answer the question at hand. *If* one decides to answer questions in a statistical way, we are bound to the rules of the game.

GPT4o and Github Copilot were used for writing this script.

1.1 Books I can recommend:

- (Free, looks nice at first glance) Introduction to Probability
- (Free, purchase of the book is recommended; page numbers in our script refer to the printed version of the book) Statistical Rethinking , YouTube-Playlist: Statistical Rethinking 2023
- (Free with a bit of searching) Understanding Regression Analysis: A Conditional Distribution Approach
- Data Analysis Using Regression and Multilevel/Hierarchical Models
- (Free) Doing Bayesian Data Analysis
- Applied Regression Analysis and Generalized Linear Models

These books are well-written, approachable, and not overly technical.

For more theoretically advanced approaches, I can recommend:

- (Free) Bayesian Data Analysis
- (Free) Elements of Statistical Learning
- (Free) Understanding Advanced Statistical Methods

1.2

In this course, we use R as our main tool for data analysis. R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows, and Mac OS.

We are following a **two-tier** approach to learning R. The **first tier** jumps right into solving exercises and problems with R using GPT. This anticipates the capacities of the software without knowing the details yet. It is not expected that you understand every detail of the codes GPT might suggest. This comes with time.

The **second tier** is a more detailed approach to R, where we learn the basics of the language and how to use it effectively. This part will be done in extensive recorded eLearning sessions over the course.

You may want to give Hadley Wickham's book a try:

- (Free) R for Data Science

Especially, the chapter on exploratory data analysis is very helpful.

This free book seems also very helpful for beginners:

- (Free) R for non-programmers R4NP

Further resources:

(Example) Introduction to R

In this video, many basic commands are explained. There are many more R introductions. My tip: Watch the beginning of a few different ones and see which explanations work best for you individually.

Overview of introductory resources

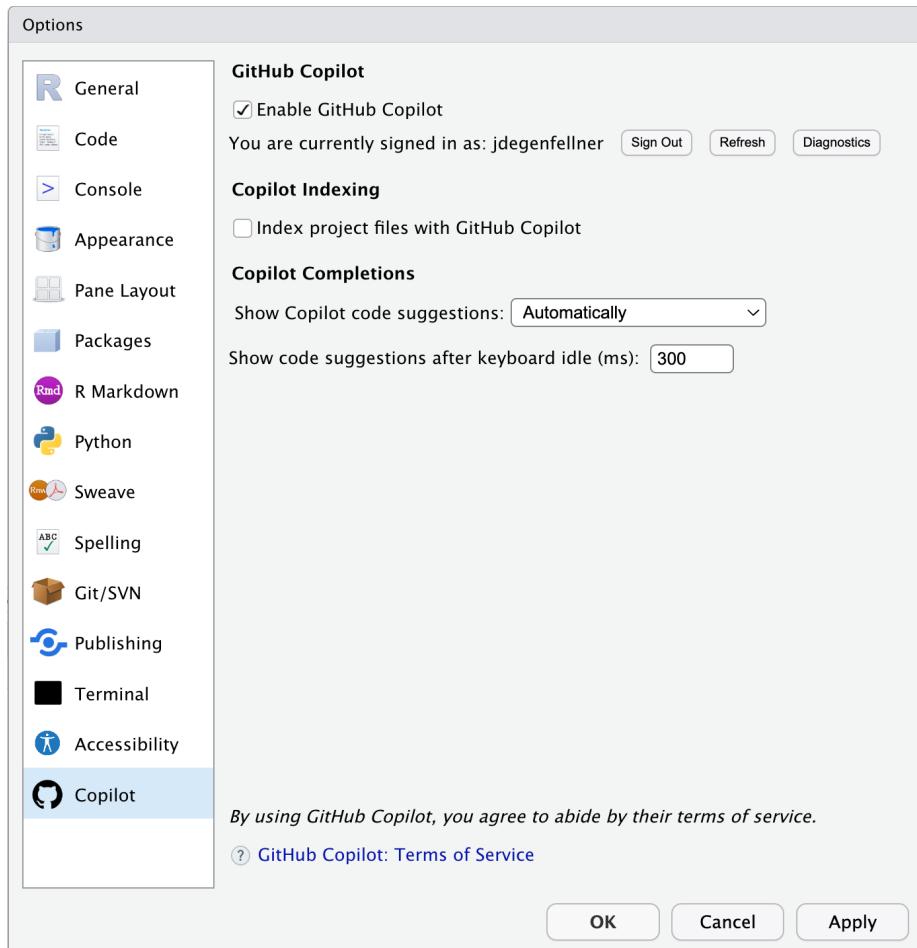
1.3 Additional Tools

Currently I use a combination of Github Copilot (paid), GPT5 (paid), and RStudio for writing code and this script. You can use Github Copilot directly in RStudio to make code suggestions, which is often increasing productivity. I can highly recommend using these tools in combination. GPT5 can process images as well, which is a game changer. As far as I know, there is no self-debugging version of GPT5 in combination with RStudio available yet (in the style of AutoGPT or similar approaches).

Large Language Models (LLM) like GPT5 enable you to write/adapt code using natural language. Among other tasks, they help you create complicated, aesthetic plots. Very often, debugging attempts get stuck. They are far from perfect yet, but an impressive feat of engineering.

1.4 Workflow suggestion

You could use RStudio to write and execute your code. In the global options of RStudio (below), you can add your github copilot account. Additionally, you can have your GPT4o(1) window open and copy-paste images of your plots or code in order to achieve your current objective.



1.5 Orientation for the course and script

Throughout the script it is important to keep a mental mind map of the topics we are covering and why. Typically, we start with a question like “Is this therapy more effective than usual care”, then we collect data, summarize it, and then we try to answer the question. Often, we already have the data (a so-called convenience sample) and we try to answer a question with it (exploratively).

By answering the question, we make an inference about the population respectively the process generating the data. So we want to conclude something about the population from the sample in front of us.

When we only **describe the data to better understand it**, we are in the realm of **descriptive statistics**.

When we try to make an inference about the population, we are in the realm of **inferential statistics**. Simply put, we want to find out if, for example, the treatment effect we are seeing in the sample is also present in the population or if what we are seeing was just by chance. We also add measures of uncertainty (for instance credible or confidence intervals) to our estimates.

We can do this inference in two ways: **Frequentist approach** using null hypothesis significance testing (NHST), or by using the **Bayesian approach** (Bayesian inference).

1.6 Warning of incompleteness

This script is not static but a continuously evolving document. It represents a snapshot of our understanding at the time of writing. Many of the topics touched upon or mentioned briefly here are vast research fields in their own right, and it is beyond the scope of this script to cover them comprehensively.

Our goal, however, is to create a resource that aligns closely with the needs of health sciences and reflects our needs in applied research. While it may not be exhaustive, we aim for it to be both practical and relevant, serving as a tailored guide for our specific context.