

PH 712 Probability and Statistical Inference

Part 0: Syllabus Review, R Basic Syntax & Calculus Basics

Zhiyang Zhou (zhou67@uwm.edu, zhiyanggeezhou.github.io)

2025/09/07 22:32:17

Contact

- Instructor: Zhiyang Zhou
 - Email: zhou67@uwm.edu
 - Homepage: zhiyanggeezhou.github.io
- Lectures
 - Mon/Wed 11:30–12:45 at Zilber 227
- Office Hours
 - By appointment

Grading

- Assignments (60%)
 - Submitting digital copies
 - Attaching (if applicable) both outputs and source codes
 - Including necessary interpretation
 - Organized in a CLEAR and READABLE way
 - Accepting no late submission
- Final project (40%)
 - Refer to the project guideline
- Bonus points (TBD)
 - Irregular quiz held on Canvas

Materials

- Reading list (recommended but NOT required)
 - [HMC] R. Hogg, J. McKean, & A. Craig. (2018). *Introduction to Mathematical Statistics*, 8th Ed. Boston: Pearson.
 - [ModernDive] Chester Ismay, Albert Y. Kim, & Arturo Valdivia. (2025). *Statistical Inference via Data Science: A ModernDive into R and the Tidyverse*, 2nd Ed. Boca Raton: CRC Press. (Accessible at moderndive.com/v2)
- Lecture notes and beyond
 - Posted at Canvas and zhiyanggeezhou.github.io
 - Subject to update **without prior notice**

Why using R?

- Open source: Freely accessible to everyone.

- Powerful for data analysis: Extensive libraries created and maintained by statisticians; built-in methods for advanced statistical modeling.
- Data visualization: High-quality plots with libraries like *ggplot2*.
- Cross-platform: Working on Windows, macOS, and Linux.

Comparing R, JMP, and SPSS

- Cost
 - R: Free and open-source.
 - JMP: Proprietary software with a paid license.
 - SPSS: Proprietary software, requires a paid license.
- Ease of use
 - R: Requires programming; steep learning curve; graphical interface like RStudio improves usability.
 - JMP: User-friendly interface with drag-and-drop capabilities.
 - SPSS: Point-and-click interface for beginners.
- Statistical analysis capabilities
 - R: Extremely powerful for basic, advanced, and cutting-edge statistical methods; extensive free packages support complex modeling.
 - JMP: Robust statistical capabilities with a strong focus on exploratory data analysis; less effective for advanced machine learning compared to R.
 - SPSS: Suitable for traditional statistical methods; limited support for advanced analytics unless additional modules are purchased; less customizable compared to R unless using SPSS Syntax (a programming language that is unique to SPSS).
- Visualization
 - R: Best-in-class visualization capabilities via free packages; allows highly customized, publication-quality visualizations.
 - JMP: Excels in dynamic and interactive graphics; provides instant visual feedback as data and models are explored; limited customization options compared to R.
 - SPSS: Basic charting and visualization options; less customizable and visually appealing compared to R and JMP.

Be careful when using R

- NO quality control: Packages developed by small groups without extensive testing

How to learn R?

- Self-learning with regular practice
- Be sophisticated in statistics

How to find solutions when using R?

- Help manual (reliable)
- Online resources (not always reliable)
 - Search engines: Google, etc.
 - Q&A communities: Stack Overflow, Reddit, Posit Community, etc.
 - AI tools: ChatGPT, Claude, Grok, etc.

Installation

- Base R: <https://cran.r-project.org>
- RStudio: <https://posit.co/download/rstudio-desktop/>

- Old versions of RStudio: <https://dailies.rstudio.com/release/>
- Enable GitHub Copilot in RStudio: <https://docs.posit.co/ide/user/ide/guide/tools/copilot.html>
 - Requiring a Github account

R basic syntax

- Pls refer to R script UWM_PH712_2025Fall_Part00.r

Course expectations on R usage

1. (expected to) Understand given R code trunks and error messages (if any).
2. (not required to) Be able to modify existing R code to meet specific requirements.
3. (not required to) Develop R code from scratch, starting with basic structures and building up.

Univariate differentiation

- The process of finding the derivative of a univariate function
 - Interpretation for derivatives:
 - * Geometric: the derivative at a point gives the slope of the tangent line to the curve at that point
 - * Practical: measuring the rate of changes, i.e., how a function changes as its input changes
 - E.g. the velocity being the derivative of position with respect to time
- Basic rules of differentiation
 - Refer to sections “Rules for basic functions” and “Rules for combined functions” at https://en.wikipedia.org/wiki/Derivative#Rules_of_computation

Univariate integration

- The process of finding the integral of a univariate function, reversing the differentiation process
 - Interpretation for integrals:
 - * Geometric: representing the area under the curve of a function over a given interval
 - * Practical: accumulating tiny quantities
 - E.g. the traveled distance being the integral of the velocity function of time
- Indefinite integrals
 - Representing the general form of an antiderivative of a function
 - Including a constant C , as integrating a function reverses the differentiation process, but there could be multiple functions that differentiate to the same result (differing by a constant)
 - * E.g., $x^2 + 1$ and $x^2 + 2$ share the identical derivative $2x$ and hence $\int 2x dx = x^2 + C$ with indefinite C .
- Definite integrals
 - Calculating the net area under the curve of a function between two specific limits
 - Dropping the indefinite C in the corresponding definite integral and plugging in limits
 - * E.g., $\int_a^b 2x dx = x^2 \Big|_{x=a}^{x=b} = b^2 - a^2$
- Basic rules of indefinite integration
 - Refer to the sections “Rational functions”, “Exponential functions” and “Logarithms” at https://en.wikipedia.org/wiki/Lists_of_integrals