

# **STAT 8670 - Computational Methods in Statistics**

Chi-Kuang Yeh

2025-07-04

# Table of contents

<b>Preface</b>	<b>4</b>
Description . . . . .	4
Prerequisites . . . . .	4
Instructor . . . . .	4
Office Hour . . . . .	4
Assignment . . . . .	4
Midterm . . . . .	5
Topics and Corresponding Lectures . . . . .	5
Recommended Textbooks . . . . .	5
<b>1 Data Structure and R Programming</b>	<b>6</b>
1.1 Data type . . . . .	6
1.1.1 To change data type . . . . .	7
1.2 Operators . . . . .	7
1.2.1 Comparison Operator . . . . .	7
1.2.2 Logical Operator . . . . .	7
1.3 Indexing . . . . .	7
1.4 Naming . . . . .	7
1.5 Array and Matrix . . . . .	7
1.6 Key and Value Pair . . . . .	7
1.7 Data Frame . . . . .	7
1.8 Tidyverse . . . . .	7
<b>2 Summary</b>	<b>8</b>
<b>References</b>	<b>9</b>
<b>I Appendix</b>	<b>10</b>
<b>3 Appendix: Introduction to R?</b>	<b>11</b>
3.1 R . . . . .	11
3.2 IDE . . . . .	11
3.2.1 Rstudio . . . . .	11
3.2.2 Visual Studio Code (VS Code) . . . . .	11
3.2.3 Positron . . . . .	12

3.3	RStudio Layout . . . . .	12
3.4	R Scripts . . . . .	12
3.5	R Help . . . . .	12
3.6	R Packages . . . . .	12
3.7	R Markdown . . . . .	13
3.8	Vectors . . . . .	13
3.9	Data Sets . . . . .	13

# Preface

## Description

Topics included are optimization, numerical integration, bootstrapping, cross-validation and Jackknife, density estimation, smoothing, and use of the statistical computer package of S-plus/R.

## Prerequisites

MATH 4752/6752 – Mathematical Statistics II, and the ability to program in a high-level language.

## Instructor

[Chi-Kuang Yeh](#), I am a postdoctoral scholar at the Department of Statistics and Actuarial Science, McGill University.

- Office: 1216 Burnside Hall.
- Email: [chi-kuang.yeh@mail.mcgill.ca](mailto:chi-kuang.yeh@mail.mcgill.ca).

## Office Hour

[By appointment and a online link will be provided later]

## Assignment

- Assignment 1: Date and topics TBA

## Midterm

- Midterm 1: Date and topics TBA

## Topics and Corresponding Lectures

Those chapters are based on the lecture notes. This part will be updated frequently.

Topic	Lecture Covered
Optimization	TBA
Numerical integration	TBA
Jackknife	TBA
Bootstrap	TBA
Cross-validation	TBA
Smoothing	TBA
Density estimation	TBA
Monte Carlo Methods	TBA

## Recommended Textbooks

- Givens, G.H. and Hoeting, J.A. Computational Statistics. Wiley, 2012.
- Rizzo, M.L. Statistical Computing with R. Chapman & Hall/CRC, 2007.
- Hothorn, T. and Everitt, B.S. A Handbook of Statistical Analyses Using R. Chapman & Hall/CRC.

# 1 Data Structure and R Programming

Data types, operators, variables

Two basic types of objects: (1) data & (2) functions

- Data: can be a number, a vector, a matrix, a dataframe, a list or other datatypes
- Function: a function is a set of instructions that takes input, processes it, and returns output. Functions can be built-in or user-defined.

## 1.1 Data type

- Boolean/Logical: Yes or No, Head or Tail, True or False
- Integers: Whole numbers  $\mathbb{Z}$ , e.g., 1, 2, 3, -1, -2, -3
- Characters: Text strings, e.g., "Hello", "World"
- Floats:
- Missing data

```
day <- c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday")
weather <- c("Raining", "Sunny", NA, "Windy", "Snowing")
data.frame(rbind(day, weather))
```

	X1	X2	X3	X4	X5
day	Monday	Tuesday	Wednesday	Thursday	Friday
weather	Raining	Sunny	<NA>	Windy	Snowing

- Other more complex type

### **1.1.1 To change data type**

## **1.2 Operators**

- Unary: One argument
- Binary: Two arguments

### **1.2.1 Comparison Operator**

### **1.2.2 Logical Operator**

## **1.3 Indexing**

## **1.4 Naming**

## **1.5 Array and Matrix**

## **1.6 Key and Value Pair**

## **1.7 Data Frame**

## **1.8 Tidyverse**

---

Some of the materials are adapted from [CMU Stat36-350](#).

## 2 Summary

In summary, this book has no content whatsoever.

1 + 1

[1] 2



## References

**Part I**

**Appendix**

## 3 Appendix: Introduction to R?

### 3.1 R

For conducting analyses with data sets of hundreds to thousands of observations, calculating by hand is not feasible and you will need a statistical software. **R** is one of those. **R** can also be thought of as a high-level programming language. In fact, **R** is [one of the top languages](#) to be used by data analysts and data scientists. There are a lot of analysis packages in **R** that are currently developed and maintained by researchers around the world to deal with different data problems. Most importantly, **R** is free! In this section, we will learn how to use **R** to conduct basic statistical analyses.

### 3.2 IDE

#### 3.2.1 Rstudio

RStudio is an integrated development environment (IDE) designed specifically for working with the **R** programming language. It provides a user-friendly interface that includes a source editor, console, environment pane, and tools for plotting, debugging, version control, and package management. RStudio supports both R and Python and is widely used for data analysis, statistical modeling, and reproducible research. It also integrates seamlessly with tools like R Markdown, Shiny, and Quarto, making it popular among data scientists, statisticians, and educators.

#### 3.2.2 Visual Studio Code (VS Code)

VS Code is a versatile code editor that supports multiple programming languages, including R. With the R extension for VS Code, users can write and execute R code, access R's console, and utilize features like syntax highlighting, code completion, and debugging. While not as specialized as RStudio for R development, VS Code offers a lightweight alternative with extensive customization options and support for various programming tasks.

### 3.2.3 Positron

Positron IDE is the next-generation integrated development environment developed by Posit, the company behind RStudio. Designed to be a modern, extensible, and language-agnostic IDE, Positron builds on the strengths of RStudio while supporting a broader range of languages and workflows, including R, Python, and Quarto.

## 3.3 RStudio Layout

RStudio consists of several panes: - **Source**: Where you write scripts and markdown documents. - **Console**: Where you type and execute R commands. - **Environment/History**: Shows your variables and command history. - **Files/Plots/Packages/Help/Viewer**: For file management, viewing plots, managing packages, accessing help, and viewing web content.

## 3.4 R Scripts

R scripts are plain text files containing R code. You can create a new script in RStudio by clicking **File > New File > R Script**.

## 3.5 R Help

Use `?function_name` or `help(function_name)` to access help for any R function. For example:

```
?mean  
help(mean)
```

## 3.6 R Packages

Packages extend R's functionality. Install a package with:

```
install.packages("package_name")
```

Load a package with:

```
library(package_name)
```

## 3.7 R Markdown

R Markdown allows you to combine text, code, and output in a single document. Create a new R Markdown file in RStudio via **File > New File > R Markdown...**

Recently, the posit team has developed a new version of the R Markdown called quarto document, with the file extension `.qmd`. It is still under rapid development.

## 3.8 Vectors

Vectors are the most basic data structure in R.

```
x <- c(1, 2, 3, 4, 5)
x
```

```
[1] 1 2 3 4 5
```

You can perform operations on vectors:

```
x * 2
```

```
[1] 2 4 6 8 10
```

## 3.9 Data Sets

Data frames are used for storing data tables. Create a data frame:

```
df <- data.frame(Name = c("Alice", "Bob"), Score = c(90, 85))
df
```

```
  Name Score
1 Alice   90
2  Bob   85
```

You can import data from files using `read.csv()` or `read.table()`.

---

This appendix is adapted from [Why R?](#).