

CASE COMPETITION TECHNICAL SKILLS: R

January 7, 2025



Agenda

- 1 Introduction to R
- 2 Basic Operations
- 3 Data Structures
- 4 R Packages
- 5 Useful Functions
- 6 Demonstration in RStudio

SECTION 01

INTRODUCTION TO R

SETTING UP R/RSTUDIO

- Download R
 - Install based on your operating system (Windows/Mac/Linux)
 - <https://cran.r-project.org/>
 - This installs the R console
- RStudio is an integrated development environment (IDE)
 - Allows you to more easily edit your programs
 - Saves your work in an organized manner
 - <https://posit.co/products/open-source/rstudio/>



BASIC SYNTAX

- R is an interpreted language, which means you can type commands directly into the console and see the results immediately
- Use the `#` symbol for comments
- Commands in R are typically written one per line, but you can also write multiple commands on the same line separated by a semicolon (;)
 - Personally, do not recommend this practice :(



DATA TYPES

Numeric

Numbers with
decimal points

3.1415, 0.99

Integer

Whole numbers

1, 2, 3

Character

Text or strings

“hello”, “world”

Logical

Boolean values

TRUE, FALSE



WHAT IS THE DATA TYPE?

10.5

“Hello world”

“False”

SECTION 02

BASIC OPERATIONS

BASIC OPERATIONS

- Assignment operator: `<-` (same as `=` in other languages)
- Arithmetic operators: `+`, `-`, `*`, `/`, `^` (exponentiation), `%%` (modulo)
- Comparison operators: `==`, `!=`, `<`, `>`, `<=`, `>=`
- Logical operators: `&&&` (AND), `|||` (OR), `!` (NOT)
- Pipe operator (tidyverse package) `%>%`
 - Command + shift + M for Mac
 - Ctrl + shift + M for Windows
 - ex. `pi %>% sin %>% cos` is equivalent to `cos(sin(pi))`

SECTION 03

DATA STRUCTURES

DATA STRUCTURES

- **Vectors:** One-dimensional arrays containing elements of the same data type. Create vectors using the `c()` function.
 - If the elements are not of the same type, it will be coerced to the highest mode
 - Logical < Numeric < Character
- **Matrices:** Two-dimensional arrays with rows and columns. Create matrices using the `matrix()` function.
 - All elements must be of the same data type



DATA STRUCTURES

- **Lists:** Ordered collections of objects of different data types. Create lists using the `list()` function.
 - Most flexible
 - Consists of any combination of data types
- **Data Frames:** Rectangular data structures consisting of rows and columns, similar to a table in a database. Create data frames using the `data.frame()` function.
 - Each column must be same data type
 - Often have column names
 - This is the one you would likely encounter the most!



SECTION 04

R PACKAGES

R PACKAGES

- R packages are collections of functions, data, and documentation
- You must install a package before using it
 - Type `install.packages("package_name")` in the console
 - Packages only need to be installed once per computer
- Packages need to be loaded in each session
 - Use `library(package_name)` to load
 - Use `package_name::function_name()` to access a function without loading the package
- `?` or `help()` shows documentation
 - Helps you understand how a function/package works



R PACKAGES

- **tidyverse** is a collection of R packages for data science tasks
 - **dplyr** – for data manipulation
 - `filter()`, `mutate()`, `group_by()`, `summarize()`
 - **ggplot2** – for data visualization
 - `ggplot()`
 - **readr** / **readxl** – for importing data from delimited files / Excel
 - `read_csv("file.csv")`, `read_excel("file.xlsx")`
- **actuar** contains tools specifically for actuarial science
 - Fitting loss data to distributions
 - Calculating risk measures
 - Modeling loss distributions

SECTION 05

USEFUL FUNCTIONS

LOADING DATA INTO R

- Read Excel sheet into a data frame in R
 - `read_excel(file_path, sheet)`
 - File path will be unique to where you save the file on your computer
 - Make sure the Excel sheet and R file are in the same folder
 - Check your working directory if the command errors
 - `getwd()` to output current working directory
 - `setwd("~/Documents/file_path")` to set to the correct directory

USEFUL OPERATIONS

- `$` to extract a column from a data frame
 - You can extract based on numeric or character index
- `set.seed(###)` for simulations to get reproducible results
- Repeat identical trials using a `for` loop or `apply()`
- Formatted printing – helpful for presenting your data
 - `format()` – typically used to format numeric vectors
 - Optional arguments: digits (number of significant figures), nsmall (number of digits after the decimal), scientific (scientific notation), etc.
 - `cat()` – concatenates and prints objects, does not return output
 - `paste()` – concatenates objects, can store output into a variable



SECTION 06

DEMONSTRATION

KEY TAKEAWAYS

- R is a powerful tool for data analysis
- It will be helpful for this year's case competition
- Apply what you have learned in your future endeavors!

ANNOUNCEMENTS

- Case Competition Introduction and Kickoff on Wednesday (1/8) 6:00-7:30 PM at MS 6627
- If you are looking for potential teammates, use this [spreadsheet](#)
- Register your team for the case competition using this [form](#) (Deadline: end of the kickoff meeting)



www.math.ucla.edu/~actuary/



bruinactuaries@gmail.com



[@bruinactuaries](https://www.instagram.com/bruinactuaries)



THANK YOU

Any questions?

