Quantitative Methods Edition

Timur Öztürk

Lecture 2: Basic Programming and Statistics

April 18, 2025

Introduction

We started with basics of R and programming. The best way to learn programming is to practice with some material you might use in the future. Hence, this lecture we will be working with basic statistics, and if time permits, some econometrics.

The lecture material follows `https://www.modernstatisticswithr.com/`.

I encourage you to have a look at the website and further resources you can find in my GitHub repository.

Functions I

Functions take inputs and return outputs. Almost everything you will do in R will be done by functions.

```
# If you would like to sum a list of numbers;
x <- c(45,3,9,99,0)
sum(x)

# Remember the "object orientedness" of R.
z <- sum(x)
z

# You can also multiply a list of numbers, but the question is, who comes
    up with these functions?
prod(x)

# Asking for help is quite easy.
?prod
```

Functions II

Here are some exercises to practice with functions.

```
# Calculate the median of 34, 16, 105, and 27.

# What does range() do?

# Is mean(4,5) the same as mean(c(4,5))?
```

MSR: Exercise 2.3 with a Function Twist

```
# Write a function that calculates your net income, given that your full
    income and the taxes you pay.
income <- 100
taxes <- 0.4

net_income <- function(income, taxes) {
  net_income <- income * (1 - taxes)
  return(net_income)
}
```

MSR: Exercise 2.4 – 2.7

```
# Before we proceed, we need to understand how data might look like in R.

# Create two vectors, height and weight, containing the heights and weights
    of five fictional people (i.e., just make up some numbers!).

# Combine your two vectors into a data frame.

# Compute the mean and median height.

# Compute the correlation between height and weight.

# Read the documentation for 'length', apply it to height. Read the
    documentation for 'sort', and apply it to weight.
```

Descriptive Statistics

```
# Before we proceed, we need to understand how data might look like in R.

# Create two vectors, height and weight, containing the heights and weights
    of five fictional people (i.e., just make up some numbers!).

# Combine your two vectors into a data frame.

# Compute the mean and median height.

# Compute the correlation between height and weight.

# Read the documentation for 'length', apply it to height. Read the
    documentation for 'sort', and apply it to weight.
```

Numerical and Categorical Variables

We will now follow the script. Then, with the knowledge you have gained, answer the following questions.

```
# Install and load the palmerpenguins package.

# View the documentation for the penguins data and read about its
    variables.

# Check the data structures: how many observations and variables
    and what type of variables (numeric, categorical, etc.) are
    there?

# Compute summary statistics (means, median, min, max, counts for
    categorical variables). Are there any missing values?
```