Week 2 Overview

Welcome to Week 2! Last week we focused on getting our environment prepared and installing the tools needed to conduct statistics for data science and had a brief refresher on some basic stats terms. This week, we are going to dig a little deeper into the basics of R to get you more comfortable in the environment. For example, how to do some basic math, defining the structures, and how-to bring data in to R. We are also going to read more about statistical models. Some of this might be a refresher, but a lot of this is likely to be new and it is important we all level set on the tools and terms this course will use. Next week, we will start to get more into the programing and "heavier" statistics topics.

Contents of the Week

Objectives

After completing this week, you should be able to:

- Basic R Functions
- Importing Data into R
- Understand R Data Structures
- Statistics Basics

Weekly Resources

- Comprehensive R Archive Network
- R Studio
- Sage Publications. (2021). Discovering Statistics Using R
- RStudio, PBC. (2021). RStudio Documentation
- RStudio, PBC. (2021). R Studio Cheatsheets

Week 2 Readings, Assignments, and Tasks

Here are your tasks for this week:

Read the following:

- *R for Everyone:* Chapters 4 – 6
- *Discovering Statistics for R*: Chapter 2 (Chapter 3 is optional, a lot of it will be repeats of *R for Everyone*: Chapters 4-6, but you are welcome to review it, as there are some additional tips/tricks! The only thing to note is that *R for Everyone* is a much newer text in terms of syntax)
- Base-r.pdf
- Data-import.pdf

Complete the following:

- 2.1 Discussion/Participation
- 2.2 Exercise

Helpful Sources

Luraschi, Javier. (2019). Importing Data with RStudio.

2.1 Discussion/Participation

Here are optional topics for discussion via Teams this week.

Remember, these topics aren't required, but if you are struggling to know what to post about, these can be used to initiate discussion!

1. What is a variable in R?
2. What are some of the different data types and what are examples of each?
3. What are vectors?
4. What is a function in a programming language? What is the point of a function?
5. What are the different types of missing data in R?
6. What are pipes and what are they used for in R?
7. What is a data.frame in R? Why would you use a data.frame?
8. What is a working directory in R?

9. What are lists?
10. What are matrices?
11. What are arrays?
12. What are some of the different types of data that can be brought into R?
13. What is a statistical model?
14. What does it mean where someone refers to the fit of a model?
15. What is a linear model?
16. What is the difference between population and sample? Why does this matter?
17. What is the sum of squared errors? What is it used for?
18. What is standard deviation? Can you give some examples to explain it?
19. What is standard error? What is standard error of the mean? Are they related to the central limit theorem?
20. What are confidence intervals?
21. What is a test statistic?
22. What is statistical power?
23. What are type I and type II errors?

2.2 Exercise

We will be utilizing GitHub for some of the exercises in this course. This is to give you more experience with interacting with this tool that will be used in most courses during the program. We will not be using all the features of GitHub, we will mostly be using it as a way to share files and establish a repository for keeping files. In Week 1 you should have forked and cloned the Bellevue Repository - which contains several exercises you will need to complete.

Complete the following 3 exercises that are available in your repository. These files are R script files - later in the course R Markdown will be introduced. After you have completed the code for these exercises, you will need to export the scripts to PDF.

The GitHub repository that will be used is:

1. Complete assigment00
2. Complete assignment01
3. Complete assignment02

Helpful resources:

1. Compiling Reports from R Scripts (R Studio)
2. Using R Studio's Compile Report Command
3. Class
   Function: https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/class
4. c()
   Function: https://www.rdocument

ation.org/packages/base/versions/3.6.2/topics/c

5. rbind
   Function: https://statisticsglobe.com/rbind-in-r-example-vector-data-frame-fill-columns/
6. cbind
   Function: https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/cbind
7. colSums
   Function: https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/colSums
8. rowSums
   Function: https://www.rdocumentation.org/packages/raster/versions/3.4-5/topics/rowSums
9. summary
   Function: https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/summary
10. head
    Function: https://www.rdocumentation.org/packages/utils/versions/3.6.2/topics/head
11. tail
    Function: https://www.rdocumentation.org/packages/rotations/versions/1.6.1/topics/tail