SOC 5050/5050: Lab-03 - Descriptive Statistics

Christopher Prener, Ph.D.

Fall 2018

Directions

Complete all of the following questions. Your answers "by hand" should be scanned and submitted as a `pdf` image along with your well-formatted R Notebook source (the `.Rmd` file) and `html` output for Part 2. This assignment should be uploaded to your Assignments Repository by 4:15PM on Monday, September 17th, 2017.

Analysis Development: Create a Project Folder System

This initial section follows the project workflow that was handed out during this week's lecture and is available in the `lecture-03` repo!

1. Using RStudio, add an R Project to the *existing* directory in your assignments repository named `Lab-02`. To do this, you will want to go to: File ▷ New Project ▷ Existing Directory and find your *existing* `Lab-02` folder.

2. In the Files tab on the lower righthand side of RStudio's screen, add a New Folder using the New Folder button right below Files. Name this new folder `docs`.

3. Create a new text file for your `README.md`. To do this, you will want to go to: File ▷ New File ▷ Text File. Save it in the top level of your `Lab-02` folder as `README.md`. You will need to add the `.md` file extension yourself - RStudio will not add it for you.

4. In the body of your `README.md` file, use Markdown formatting to write a sentance or two describing the purpose of this project.

5. Create a new notebook by going to File ▷ New File ▷ R Notebook. Save it within that `docs/` subdirectory you just created.

6. Expand the YAML heading as you did last week. Remember that a starter notebook was posted in the `lecture-02` repository that contains syntax to get you going!

7. Use RMarkdown syntax to create your first assignment notebook! Make sure it has an introductory section, a section for loading packages, a section for loading data, and a section for part 2 below. These sections should be second-level headings (e.g. `## Introduction`). Within Part 2, use third level headings to designate question numbers (e.g. `### Question 6`).

8. When you are done, "knit" your document by clicking the Knit button in the toolbar at the top of the notebook.

Part 1: Descriptive Statistics by Hand

Anscombe's Quartet

Dataset 1		Dataset 2		Dataset 3		Dataset 4	
x_1	y_1	x_2	y_2	x_3	y_3	x_4	y_4
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

Vector assignments:

1. Pair 1 ($x1$ and $y1$): Nadia, Zen, Jin, Raymond, Carissa

2. Pair 2 ($x2$ and $y2$): Mustafa, Nick, Jeanna, Janaé, Logan

3. Pair 3 ($x3$ and $y3$): Addie, Branson, Meadow, Bobby

4. Pair 4 ($x4$ and $y4$): Caroline, Carter, Mae, Via

For your given x and y vectors:

1. Calculate the median

2. Calculate the mean

3. Calculate the standard deviation

4. Calculate the range

When you calculate these statistics, use the intentional layout demoed during the lecture and included on the equations handout. You can check your work using the datasets::anscombe data if you wish.

Part 2: Descriptive Statistics in R

Use the `auto17` data frame saved in the `testDriveR` package to produce the following:

5. For the variable `driveStr2`, produce the following:

   (a) Produce an appropriate plot for some initial exploratory data analysis

   (b) A frequency table using the `janitor` package

   (c) What is the mode?

6. For the variable `cityFE`, produce the following:

   (a) Produce an appropriate plot for some initial exploratory data analysis

   (b) Calculate the median

   (c) Calculate the mean

   (d) Calculate the standard deviation

   (e) Calculate the range

   (f) Calculate the inter-quartile range

7. Calculate summary descriptive statistics for the entire `auto17` data set using `skimr::skim()`.