Data Basics & Sampling Principles

Mini-Assignment 2023-01-23 - MTH 361 - Spring 2023

Instructions:

- Please provide complete solutions for each problem. If it involves mathematical computations, explanations, or analysis, please provide your reasoning or detailed solutions.
- Note that some problems have multiple solutions or ways to solve it. Make sure that your solutions are clear enough to showcase your work and understanding of the material.
- Creativity and collaborations are encouraged. Use all of the resources you have and what you need to complete the mini-assignment. Each student must take personal responsibility and submit their work individually. Please abide by the University of Portland Academic Honor Principle.
- **Please save your work as one pdf file, don't put your name in any part of the document, and submit it to the Teams Assignments for this course. Your document upload will correspond to your name automatically in Teams.**
- If you have questions or concerns, please feel free to ask the instructor.

Libraries

```
library(tidyverse)
```

I. Iris Flowers

Materials

In this mini-assignment, you are going to use the `iris` dataset. This dataset is in the datasets package which is already included in the base R installation.

```
glimpse(iris)
```

```
## Rows: 150
## Columns: 5
## $ Sepal.Length <dbl> 5.1, 4.9, 4.7, 4.6, 5.0, 5.4, 4.6, 5.0, 4.4, 4.9, 5.4, 4.~
## $ Sepal.Width  <dbl> 3.5, 3.0, 3.2, 3.1, 3.6, 3.9, 3.4, 3.4, 2.9, 3.1, 3.7, 3.~
## $ Petal.Length <dbl> 1.4, 1.4, 1.3, 1.5, 1.4, 1.7, 1.4, 1.5, 1.4, 1.5, 1.5, 1.~
## $ Petal.Width  <dbl> 0.2, 0.2, 0.2, 0.2, 0.2, 0.4, 0.3, 0.2, 0.2, 0.1, 0.2, 0.~
## $ Species      <fct> setosa, setosa, setosa, setosa, setosa, setosa, setosa, s~
```

There are five columns and 150 rows in the iris dataset.

Exercises

1. Identify the types of variables in the `iris` dataset.
2. Sample the `iris` dataset in different ways.
    a. Sample 10 observations using simple random sampling.
    b. Using the species as strata, sampling 30% of observation from each species.
    c. Using the species as clusters, sample one species.
3. Using the `Petal.Length` variable in the `iris` dataset, perform a sampling procedure where you sample 10 observations and compute its mean. Repeat this procedure for 1000 trials while recording the means. Plot the distribution of the means. What is the shape of the distribution? Repeat the procedure but with 100 observations per sample and 2000 trials. Plot the distribution of the means. Is the distribution much more refined from the previous plot?