

# Project

## Instructions

Please visit the website <https://vincentarelbundock.github.io/Rdatasets/datasets.html>. Each row of this table corresponds to a dataset. For this project, you will perform some basic statistical analysis on one dataset of your choice. Try to find a dataset that interests you and has at least 30 rows. Browse the titles and read the descriptions in the DOC files to understand what the data is and how it was collected.

In the DOC file, read the section labeled Format to see what sort of data is recorded and what units are used. Most datasets have lots of different measurements (columns), and you will need to pick one to focus on. Once you have selected a dataset and a column within that data set, click on the CSV file. Open this file with any spreadsheet program/app (Microsoft Excel, Mac Numbers, Google Sheets, etc). The numbers of rows (minus the header row) is the sample size. The column that you selected contains all of the sample measurements. Use these sample measurements to answer the following questions.

Please write your answers in a word document and submit it on blackboard. This project is extra credit and worth 5% of extra credit added to your final grade. Please note that you need to answer all the questions completely and correctly to receive all 5%.

1. What is the title of the dataset you have selected?
2. What is the measurement (column) you have selected, and what is the unit for this measurement?
3. What is the sample size  $n$ ? That is, how many rows does your dataset have (not counting the header row)? Note: you must select a dataset with  $n \geq 30$ .
4. What is the sample mean  $\bar{x}$ ?
5. What is the sample standard deviation  $s$ ?
6. Create a histogram and describe the distribution of the measurement you have chosen.
7. Approximately what kind of distribution does the random variable  $\bar{x}$  have, and why? What is the name of the theorem that tells you so?
8. Estimate the population mean  $\mu$  in two ways.
  - Give a point estimate for  $\mu$  and a “margin of error”.
  - Construct a 95% confidence interval for  $\mu$ .

- Construct a 99% confidence interval for .
- 

### Hints

- Remember back to our first few lectures - we discussed how to calculate the mean, standard deviation using Excel.

This project is adapted from Professor Adamski's class summer 2020 class.