

Lesson 03: Descriptive Statistics

References

- Black, Chapter 3 Descriptive Statistics (pp. 45-80)
- Kabakoff, Chapter 5.2 Numerical and Character Functions (pp. 91-93), Chapter 6.5 Box Plots (pp. 129)
- Davies, Chapter 13 Elementary Statistics (pp. 261-279)
- Stowell, Chapter 5 Summary Statistics for Continuous Variables (pp. 59-62)

Data sets: `mileage.csv`, `shoppers.csv`, `pontus.csv`, `geyser.csv`

Exercises:

Description: `mileage.csv` is derived from a 1991 U.S EPA study of passenger car mileage. This file includes information on sixty cars: HP (engine horsepower), MPG (average miles per gallon) WT (vehicle weight in 100 lb units) and CLASS (vehicle weight class C1,,C6).

- 1) For each weight class determine the mean and standard deviation of MPG. What can you conclude from these calculations?
- 2) For each weight class determine the mean and standard deviation of HP. What can you conclude from these calculations?

Description: `shoppers.csv` contains the dollar amounts spent in a store by individual shoppers during one day.

Find the mean, median, range, standard deviation, variance, Q1, Q3 and P10. Plot the histogram and describe the distribution.

Description: `pontus.csv` lists the ages of USA Presidents at the time of their inauguration. Also listed are the heights of the Presidents and their opponents.

- 1) Find the mean, median, range, standard deviation, Q1, Q3 and P10 of the Presidents' ages.
- 2) Find the mean, median, range, standard deviation, Q1, Q3 and P10 of the heights of the Presidents and also their opponents.
- 3) Calculate the difference between each President's height and that of his opponent. Plot a histogram of these differences. Construct a boxplot. What do you conclude from your calculations? Why is the difference of average heights calculated in (2) different from the average of the pairwise differences calculated in (3)?

Description: `geyser.csv` contains the intervals (in minutes) between eruptions of Old Faithful Geyser in Yellowstone National Park. The data were taken on two consecutive weeks: WEEK1 and WEEK2.

Compare the two sets of data using `summary()`, `hist()` and `boxplot()`. What do you conclude?