

## STAT651 2018, Homework 1

To do the HWs you will need to either use JMP or another statistical software (I discourage using Excel). Do not give JMP commands in your homework. Any JMP output included in your HW should be placed within the text.

- (1) What sort of variable is
  - (i) The type of cycle a person owns.
  - (ii) The height of a person.
  - (iii) The gender of a person.
- (2) What is the mean and standard deviation of the following samples:
  - (a) 1, 1, 1, 1, 1.
  - (b) -99, -49, 1, 51, 101.
- (3) For a given data set the median is 10, the first quartile is 4 and the third quartile is 10. What percentage of this data set must be the same?
- (4) Sidney observes the data

1, 1.5, 2, 3, 3.5, 3.8, 6, 6.5, 7, 8.

The sample mean and standard deviation of this sample is 4.23 and 2.47 respectively. What happens to the mean, standard deviation and IQR of the sample if the last observation 8 is replaced with a larger number?

- (5) Input the Old Faithful Geyser (a Geyser in Yellowstone National Park) data into JMP ([https://www.stat.tamu.edu/~sahasini/teaching651/old\\_faithfuldat.txt](https://www.stat.tamu.edu/~sahasini/teaching651/old_faithfuldat.txt)). The second column corresponds to the time (in minutes) of each eruption and third column corresponds to the time (in minutes) between eruptions (waiting time). The third column succeeds/follows the second column; i.e. the time between eruptions *follows/succeeds* the time of each eruption.
  - (i) Make a relative frequency histogram of the waiting times.  
Describe the most interesting features of the histogram.
  - (ii) Does the boxplot convey the main features seen in the distribution?
  - (iii) Make a scatter plot of waiting time against eruption time.  
The JMP commands are Analyze > Fit Y by X >). Place waiting times in Y, response and eruption times in X, factor. Press Okay.
  - (iv) Explain how Yellowstone National Park could use this information to predict the time until the next eruption (by the geyser there is a clock which predicts the time).

(6) Upload the undergraduate height data

(<https://www.stat.tamu.edu/~suhasini/teaching301/heights301.txt>) and the average undergraduate height data (average over 5 students) data

(<https://www.stat.tamu.edu/~suhasini/teaching301/heightsAve.txt>) into JMP.

Using JMP make relative frequency histograms of both data sets (note that the summary statistics are given with the plot).

Compare the mean and standard deviations of data sets. How different are the standard deviations?

(7) Upload the M&M data into JMP

([https://www.stat.tamu.edu/~suhasini/teaching651/MandMs\\_2013.csv](https://www.stat.tamu.edu/~suhasini/teaching651/MandMs_2013.csv))

Using JMP make comparative boxplots for the number of M&Ms in for the different types (P, PB, M). Describe what you observe.