Chapters covered: Chapter 1, 2, 4

Show your work to receive full credit.

Textbook Problems : (Statistics, the art and science of learning from data, 4th)

- **Problem 1 Exercise 1.12 At what age did women marry?** (Page 16)
- **Problem 2 Exercise 1.35 Multiple Choice: Use of inferential statistics?** (Page 24) (No explanation required)
- **Problem 3 Exercise 2.4 Modified** Identify each of the following variables as either categorical or quantitative. For quantitative variable, identify whether it is discrete or continuous.
  a. Choice of diet (vegan, vegetarian, neither)
  b. Time spent shopping online per week
  c. Ownership of a tablet (yes, no)
  d. Number of siblings
  e. Distance of commute to work

- **Problem 4 Exercise 2.16 Modified** For a trip to Miami, Florida, over spring break 2014, the data below (obtained from travelocity.com) show the price per night (in U.S. dollars) for various hotel rooms. 239, 237, 245, 310, 218, 175, 330, 196, 178, 245, 255, 190, 330, 124, 162, 190, 386, 145
  a. Enter numbers above in R using the following R command. Use built-in R command `stem()` to construct a stem-and-leaf plot. Copy and paste your R output. How many hotel rooms cost more than $300?

     ```
     price<-c(239, 237, 245, 310, 218, 175, 330, 196, 178, 245, 255, 190, 330,
     124, 162, 190, 386, 145)
     ```

  b. Use built-in R command `hist()` to construct a histogram. Mention where most prices tend to fall and comment about the shape of the distribution. Include the plot in your assignment submission.
- **Problem 5 Refer to Problem 4 (Hotel Price)**
  a. Use R to calculate mean and standard deviation. Interpret the results in the context of problem.
  b. The most expensive hotel room in this problem is $386. How many standard deviations above the mean is it? According to three-standard-deviation-rule, is this price extremely unusual?
- **Problem 6 Exercise 2.46 Sick leave** (page 65) **Use R to find standard deviation in b) and c)**;
- **Problem 7 : Exercise 4.29 Identify the bias** (page 172)
- **Problem 8: Exercise 4.68 Aspirin prevents heart attacks?** (Page 192)

R Problem

In this problem we will use data set named 'Hurricane' available at `http://sites.williams.edu/bklingen/files/2015/05/hurricanes.csv`. It contains the damage (in billion dollars) of the 30 most costly hurricaines hitting the U.S. mainland between 1900 and 2010. (Numbers are inflation adjusted and in 2010 dollars). Import data set into R using the following R command:

```
dat<-read.csv("http://sites.williams.edu/bklingen/files/2015/05/hurricanes.csv")
```

- **R Problem 1**

  a. Construct a histogram of variable 'Damage' with a main title "Hurricane Damage" and x-axis title "Damage (in billion US Dollars)". Describe the shape of the distribution and comment if there is any outlier. Include the plot in your homework submission.

  b. Identify the 5-number-summary of the variable 'Damage' using R.

  c. Would you use the mean or median to describe the center of variable "Damage"? Why?

In the following problems we will study another variable 'Category' in the same data set. Go to (`https://www.nhc.noaa.gov/aboutsshws.php`) and learn about hurricane category index if you are not familiar with it.

- **R Problem 2**

  a. Construct a frequency table of `Category` using `table()` command in R. Copy and paste the output.

  b. A major hurricane is one that is category 3 or more. Based on the frequency table from above, find the **percent** of major hurricanes.

  c. Does either a bar graph or pie chart make sense for these data? Explain. (No need to make one)