STA 100 Fall 2018
Homework 1 - Due Friday, Oct 5^{th}, before 11am

Book Homework

These prompts correspond to "Book Homework" portion of the homework on Canvas. You turn in the answers to these questions online.

1. For the following random variables, specify if they are nominal, ordinal, continuous, or discrete.
   (a) Number of outbreaks of pneumonia at UC Davis.
   (b) The amount of money you can physically hand to another person.
   (c) The shape of a particular cell.
   (d) The width of a muskrats' snout in centimeters.

2. The number of leaves that a type of tree shed in a week was recorded, with the following results:

   $$21, 22, 22, 24, 27, 28, 30, 32, 33, 60$$

   Use this sample data to solve the following:
   (a) Calculate the mean.
   (b) Calculate the median.
   (c) Calculate the variance.
   (d) Calculate the standard deviation.
   (e) Interpret the standard deviation in terms of the problem.

3. Continue with the data in Problem 2.
   (a) Calculate the first quartile.
   (b) Calculate the third quartile.
   (c) Calculate the lower cutoff for outliers, using the boxplot method.
   (d) Calculate the upper cutoff for outliers, using the boxplot method.
   (e) Identify any outliers in the dataset.
   (f) Calculate the $30th^{th}$ percentile.

4. A random sample of 100 students was taken, and the number of times a week the student exercised was recorded:

# of Times Exercised	0	1	2	3	10
Freq	20	40	24	14	2

   i.e, 20 students did not exercise, 40 exercised 1 time a week, 24 exercised twice, etc.
   (a) Find the average number of times a student exercised.
   (b) Find the median of the number of times a student exercised.
   (c) Find the variance of the number of times a student exercised.
   (d) Find the standard deviation of the number of times a student exercised.

5. Continue with the data in Problem 4.
   (a) Calculate the first quartile for time number of times a student exercised.
   (b) Calculate the third quartile for time number of times a student exercised.
   (c) Calculate the lower cutoff for outliers, using the boxplot method.
   (d) Calculate the upper cutoff for outliers, using the boxplot method.
   (e) Identify all outliers in the dataset.

6. Answer the following questions with TRUE or FALSE. It is good practice to explain your answers.
   (a) The standard deviation must always be larger than the mean.
   (b) Outliers do not have a strong influence on the range of a dataset.
   (c) The 90^{th} percentile is the value for which 10% of the data lies above it.
   (d) Outliers have a strong influence on the mean of a dataset.

R Homework

These prompts correspond to "R Portion" of the homeworks on Canvas. You use R to find the answers to the following questions, and submit your answers online.

I. On Canvas, you will find the dataset "pain.txt" (in the folder Datasets). This dataset has the following columns:

Column 1: HairColor: The hair color of the subject

Column 2: Score: The pain tolerance of the subject

Each person in the experiment was given a pain threshold score based on his or her performance in a pain sensitivity test (the higher the score, the higher the person's pain tolerance).

(a) Find the overall mean of pain tolerance, regardless of hair color.

(b) Find the overall standard deviation of pain tolerance, regardless of hair color.

(c) Find the mean pain tolerance for the category DarkBrunette. *Hint: Use R to find the mean for each group.*

(d) Find the standard deviation of pain tolerance for the category LightBrunette. *Hint: Use R to find the standard deviation for each group.*

(e) Find the number of subjects that were in the category LightBlond. *Hint: Use R to find the number of subjects for each group.*

(f) Which hair color category had the highest average pain tolerance?

(g) Which hair color category had the lowest standard deviation of pain tolerance?

(h) Which hair color category had the most subjects?