# Practice Problems 5: *Summary Statistics, Correlation & Regression Line*

## A. From your Book:

- Section 5.1-all examples and exercises, Section 9.1-Exercises: 1–4 , Page 348, Section 9.2-Exercises: 1(a,b), 2, 3a, Page 356, Section 9.3-Exercises: 1–5, Page 365, Section 10.1-Exercises: 1–3, Page 380

## B. Additional Problems:

1. For the following statements, state whether they are True/False and explain why:

   (a) If you subtract 3 from all the numbers of the list, this subtracts 3 from the median.

   (b) If you divide by 3 all the numbers of the list, this divides the median by 3 .

   (c) If you change the sign of each entry on a list, that changes the sign of the median.

   (d) If you change the sign of each entry on a list, that changes the sign of the mean.

   (e) Half of a list is always below the median.

   (f) If the histogram of your data is *right skewed*, then this means that you have outliers that take very large values compared to the mean.

   (g) If the histogram is left skewed, then you expect the median to be to the right of the mean.

   (h) In a symmetric dataset you expect the mean, the median and the mode to be always close together.

2. Suppose that you have collected the following data

$$\{12, 8, 9, 10, 9\}$$

   (a) Calculate the mean and median of the data.

   (b) Calculate the sample standard deviation s.

   (c) If we multiply this data by 100, what is the sample standard deviation of the new data?

3. Suppose that we measure the weight of 15 frogs in grams

$$\{35, 18, 28, 29, 38, 23, 29, 38, 38, 35, 33, 36, 38, 46, 22\}$$

   (a) Calculate the mean and median of the data to give an estimate of the average weight of this frog.

   (b) Calculate the sample standard deviation.

4. Every day for two weeks, Phil arrives at the bus stop at 4:30, and then uses a stop watch to measure how long until the bus comes. Here is the data he collected

$$\{0.4, 0.6, 1.3, 3.4, 3.7, 3.9, 3.9, 5.4, 7.6, 40.6\}$$

(a) Calculate the 5-number summary for this data.

(b) Calculate the IQR to give Phil an idea of how much his waiting time varies.

(c) Also calculate the sample standard deviation, s.

(d) Which measure of spread is more resistant to outliers? Which description of the spread of the data do you think is more valuable in this case?

5. (a) 10 people in a room have an average height of 5 feet 6 inches. An 11th person, who is 6 feet 5 inches tall, enters the room. Find the average height of all 11 people.

(b) 21 people in a room have an average height of 5 feet 6 inches. A 22nd person, who is 6 feet 5 inches tall, enters the room. Find the average height of all 22 people. Compare with (a).

(c) 21 people in a room have an average height of 5 feet 6 inches. A 22nd person enters the room. How tall would he have to be to raise the average height by 1 inch?

6. The following scatter plots have sample correlations of -0.997, -0.783, 0.0556 , and 0.847. Connect each plot with its appropriate correlation.
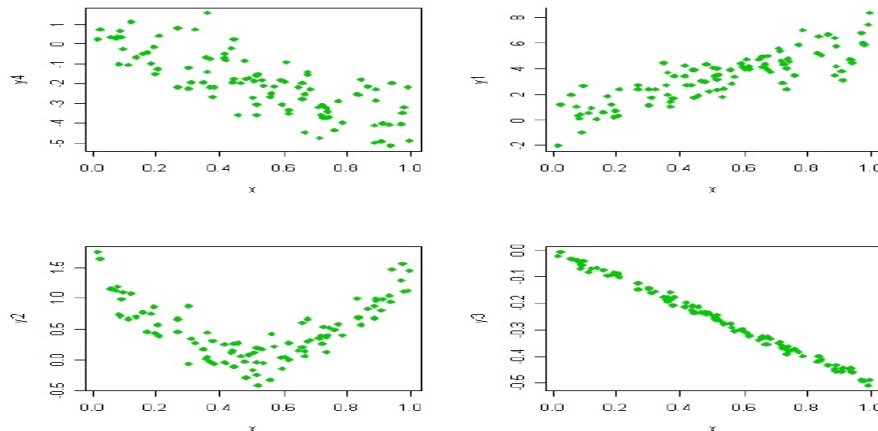


7. For the following data set

| x | 4  | 5  | 6 | 8 | 10 |
|---|----|----|---|---|----|
| y | -3 | -2 | 1 | 3 | 2  |

(a) Calculate $\bar{x}$ and $\bar{y}$ and the standard deviations of the x's and the y's.

(b) Calculate the correlation coeffcient between x and y?

8. *A study of elementary school children, ages 6 to 11, finds a high positive correlation between shoe size $X$ and score $Y$ on a test of reading comprehension. Suggest possible explanations for the association other than the causal relationship between $X$ and $Y$.*

9. A student wonders if people of similar heights tend to data each other. She measures herself, her dormitory roommate, and the women in the adjoining rooms; then she measures the next man each woman dates. Here are the data (height is in inches)

| Women | 66 | 64 | 66 | 65 | 70 | 65 |
|-------|----|----|----|----|----|----|
| Men   | 72 | 68 | 70 | 68 | 71 | 65 |

(a) Make the scatterplot of these data. Based on the scatterplot, do you expect the correlation to be positive or negative? Near $\pm 1$ or not?

(b) Compute the correlation coefficient $r$ between the height of men and women.

(c) How would $r$ change if all men were 12 inches shorter than the heights given in the table? Does the correlation help answer the question of whether women tend to dat men taller than themselves?

(d) If every woman dated a man exactly 3 inches taller than she, what would be the correlation between male and female?

(e) Suppose that the heights of men were measured in centimeters rather than inches, but that the heights of women remained in inches. (there are 2.54cm to an inch.) What would now be the correlation between male and female height?

10. I was wondering if I could predict the cost of pastries at the Arbor by the amount of sugar they contain. So, I took 15 pastries and found that they cost on average $1.93 with a standard deviation of $0.26. I looked at their ingredients and found that they had an average of 35.4 g of sugar with a standard deviation of 16 g. The correlation between sugar and price was 0.372.

(a) Which is the explanatory variable?

(b) Calculate $b_0$ and $b_1$ for the regression line.

(c) For a croissant that contains 7 g of sugar, use your model to predict its price.

11. A sociologists wants to find out if watching too much television has a detrimental effect at the attention span of children. First, 243 children were surveyed to find out how much television they watched. The children watched an average of 4.3 hours per day with a sample standard deviation of 2.84 hours. Then, each child was given a task and the sociologist measured how long it was until the child became distracted. She found that they became distracted in an average of 6.92 minutes with a standard deviation of 3.77 minutes.

(a) What is the response variable in this study?

(b) If $\sum x_i y_i = 4762.8$, what is the correlation coefficient?

(c) Describe what the correlation indicates about the relationship between watching television and the attention span of the children.

(d) Calculate the regression line.

(e) Predict the attention span for children who watch 4.5 hours of television per day.