Numerical Summarization and Correlation Practice Problems

1. Ever wonder how many words are in a typical inaugural address? The data in the next column represent the length of all the presidents up to Donald Trump.

1425	1125	1128	5433	2242	2283
135	1172	1337	1802	2446	1507
2308	3838	2480	1526	2449	2170
1729	8445	2978	3318	1355	1571
2158	4776	1681	4059	1437	2073
1175	996	4388	3801	2130	2406
1209	3319	2015	1883	1668	2137
3217	2821	3967	1807	1087	
4467	3634	2217	1340	2463	
2906	698	985	559	2546	

Source: infoplease.com

You will get the data On Canvas as data_csv3.

- a. Determine the mean and median number of words in a presidential inaugural address.
- b. Determine and interpret the quartiles for the number of words in a presidential inaugural address.
- c. Determine the five number summary of words of a presidential inaugural address.
- d. Determine the standard deviation and Interquartile range for the words in a presidential inaugural address.
- e. Are there any outliers in the dataset?

d. Draw a boxplot of the data.

e. Describe the shape of the data. Support your position using the boxplot and quartiles.

f. Which measure of central tendency you think better describes the typical number of words in a presidential inaugural address.

g. Which measure of dispersion do you think would better describe the spread of the typical number of words in a presidential inaugural address?

2. Consider the following data set:

X 5 6 7 7 8 8 8 8 9 9 10 10 11 11 12 12

Y 4.2 5 5.2 5.9 6 6.2 6.1 6.9 7.2 8 8.3 7.4 8.4 7.8 8.5 9.5

(a) Draw a scatter diagram with the *x*-axis starting at 0 and ending at 30 and with the *y*-axis starting at 0 and ending at 20.

(b) Compute the linear correlation coefficient.

(c) Now multiply both *x* and *y* by 2.

(d) Draw a scatter diagram of the new data with the *x*-axis starting at 0 and ending at 30 and with the *y*-axis starting at 0 and ending at 20. Compare the scatter diagrams.

(e) Compute the linear correlation coefficient.

(f) Conclude that multiplying each value in the data set by a nonzero constant does not affect the correlation between the variables. Explain why this is the case.

3. Lyme disease is an inflammatory disease that results in a skin rash and flulike symptoms. It is transmitted through the bite of an infected deer tick. The following data represent the number of reported cases of Lyme disease and the number of drowning deaths for a rural county in the United States.

Month	J	F	M	A	M	J	J	A	S	O	N	D
Cases of Lyme Disease	3	2	2	4	5	15	22	13	6	5	4	1
Drowning Deaths	0	1	2	1	2	9	16	5	3	3	1	0

(a) Draw a scatter diagram of the data using cases of Lyme disease as the explanatory variable.

(b) Compute the correlation coefficient for the data.

(c) Based on your results from parts (a) and (b), what type of relation exists between the number of reported cases of Lyme disease and drowning deaths? Do you believe that an increase in cases of Lyme disease causes an increase in drowning deaths.

4. Researchers wondered whether the size of a person's brain was related to the individual's mental capacity. They selected a sample of right-handed introductory psychology students who had SAT scores higher than 1350. The subjects took the Wechsler Adult Intelligence Scale-Revised to obtain their IQ scores. MRI scans were performed at the same facility for the subjects. The scans consisted of 18 horizontal MR images. The computer counted all pixels with a nonzero gray scale in each of the 18 images, and the total count served as an index for brain size.

Gender	MRI Count	IQ
Female	816,932	133
Female	951,545	137
Female	991,305	138
Female	833,868	132
Female	856,472	140
Female	852,244	132
Female	790,619	135
Female	866,662	130
Female	857,782	133
Female	948,066	133
Male	949,395	140
Male	1,001,121	140
Male	1,038,437	139
Male	965,353	133
Male	955,466	133
Male	1,079,549	141
Male	924,059	135
Male	955,003	139
Male	935,494	141
Male	949,589	144

a) Draw a scatter diagram treating MRI count as the explanatory variable and IQ as the response variable. Comment on what you see.

(b) Compute the linear correlation coefficient between MRI count and IQ. Are MRI count and IQ linearly related?

(c) Compute the linear correlation coefficient between MRI count and IQ for females. Compute the linear correlation coefficient between MRI count and IQ for males. Are MRI count and IQ linearly related? What is the moral?