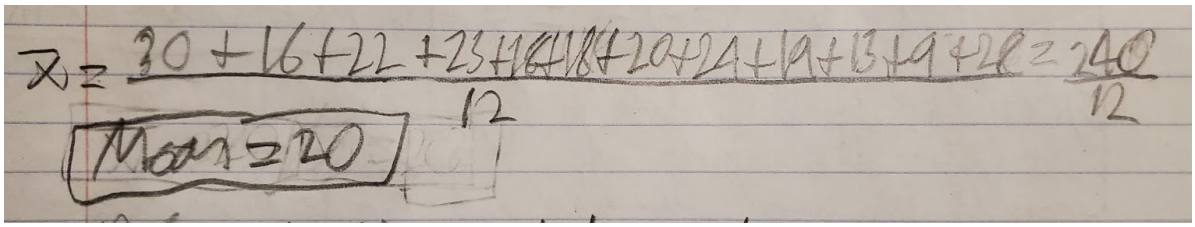
STAT 50 Quiz #1

1. Consider the following data set:

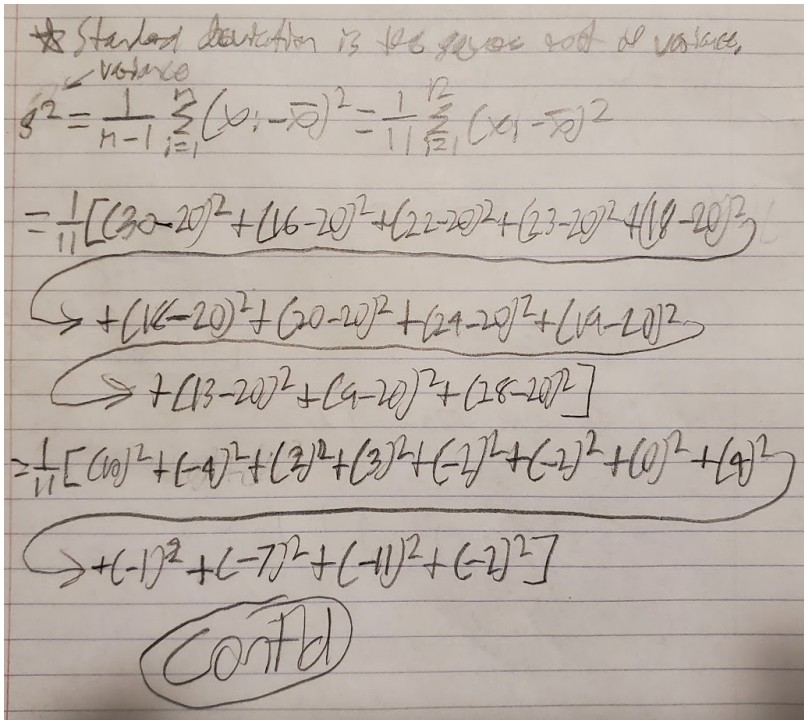
30	16	22	23	18	18
20	24	19	13	9	28

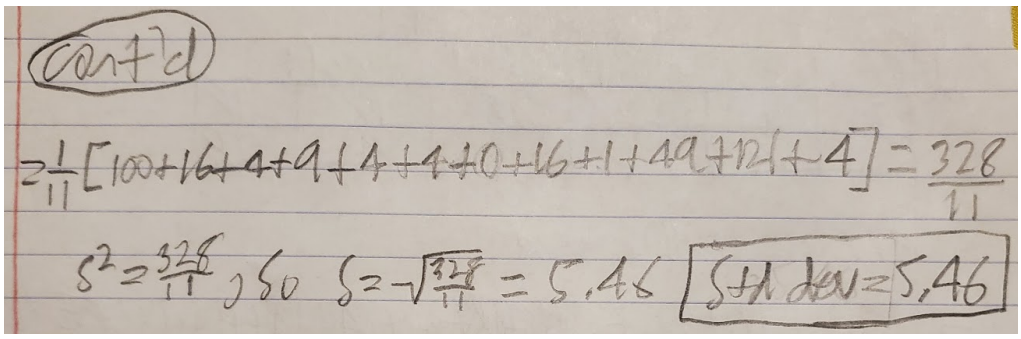
a) Compute the mean.



b) Compute the standard deviation.

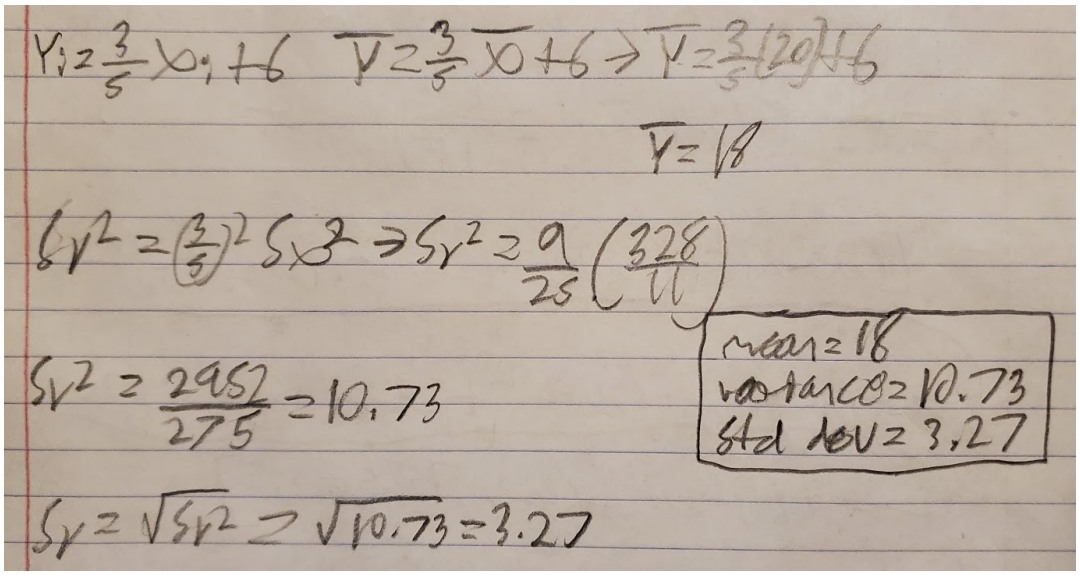
***There's two images showing my work for this part!**





c) Consider the above data to be the $X_i's$. Now, consider the transformation $Y_i = \frac{3}{5}X_i + 6$.

Compute the mean, variance, and standard deviation of the transformed data <u>without actually constructing the transformed data set.</u>

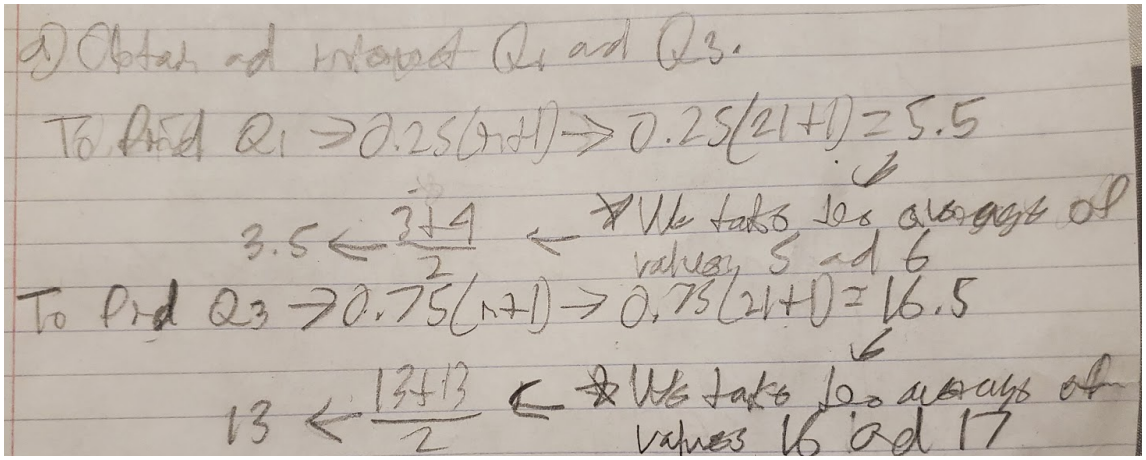


2. The U.S. National Center for Health Statistics compiles data on the length of stay by patients in short-term hospitals. A random sample of 21 patients yielded the following data on length of stay, in days:

1	1	3	3	3	4	5
5	5	5	7	9	9	10
12	13	13	15	18	42	55

a) Obtain and interpret Q_1 and Q_3.

***The work is shown in the image, but the interpretation is typed below. This is so you won't have to deal with my handwriting!**



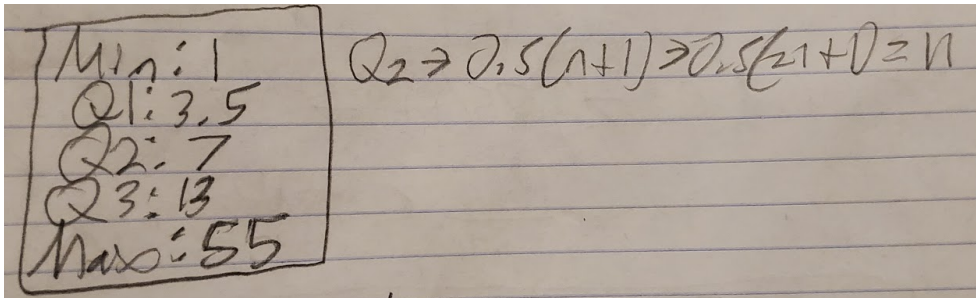
The 1st quartile is 3.5, which is the median of the lower half of the data. This is the 25th percentile; it can be said that 25% (*as closely as possible) of the data is below the value of 3.5 and that 75% (*again as closely as possible) is above it.
The 3rd quartile is 13, which is the median of the upper half of the data. This is the 75th percentile; it can be said that 75% (*as closely as possible) of the data is below the value of 13 and that 75% (*again as closely as possible) is above it.

b) Determine and interpret the interquartile range.

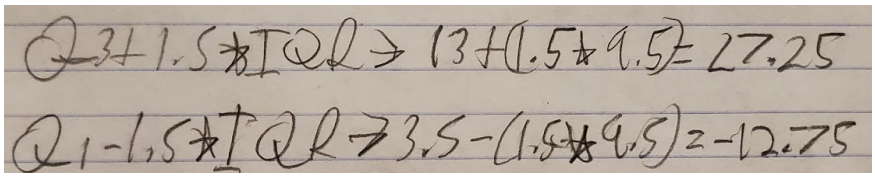
IQR = $Q_3 - Q_1 = 13 - 3.5 = 9.5$

The interquartile range of 9.5 is the distance needed to span the middle half (or middle 50%) of the data. In other words, the middle half of the data had a spread of roughly 9.5 days.

c) List the five-number summary.

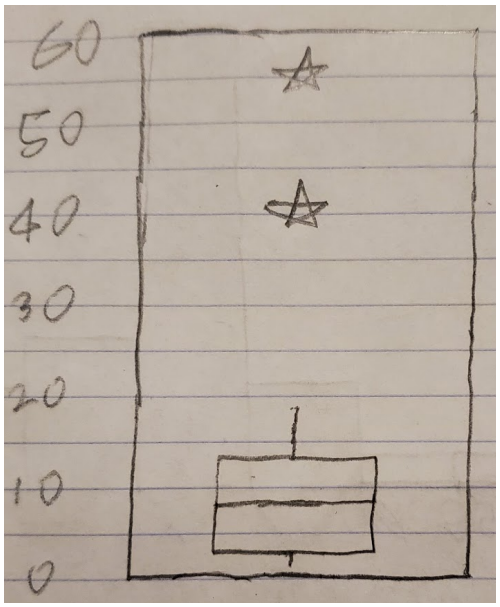


d) Identify potential outliers, if any.



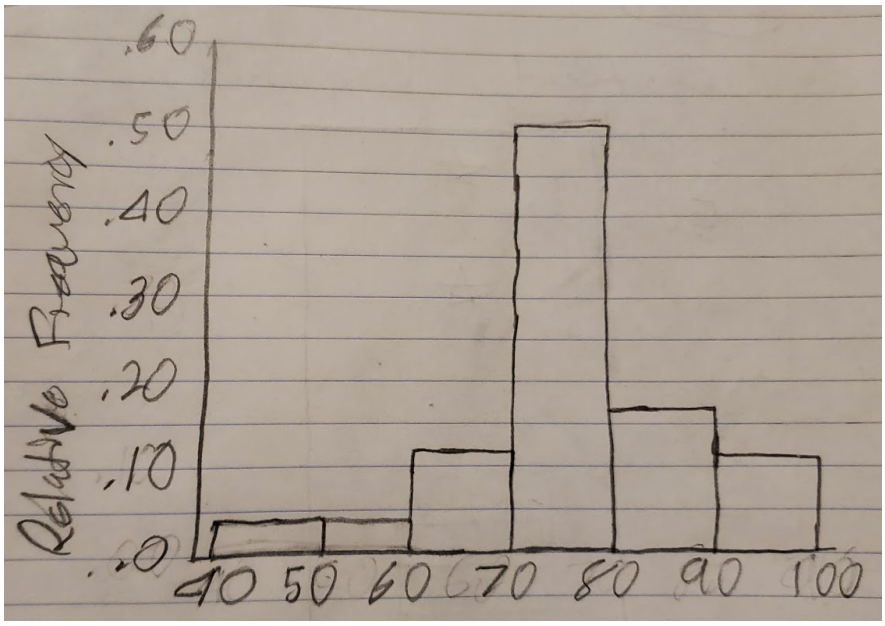
There are two potential outliers above the data, with both values being 42 and 55 days.

e) Construct a boxplot.



3. The following data represent scores of 25 students on a calculus test. Use the data to construct a relative frequency histogram. Use class widths of 40 - 49, 50 – 59, 60 – 69, etc.

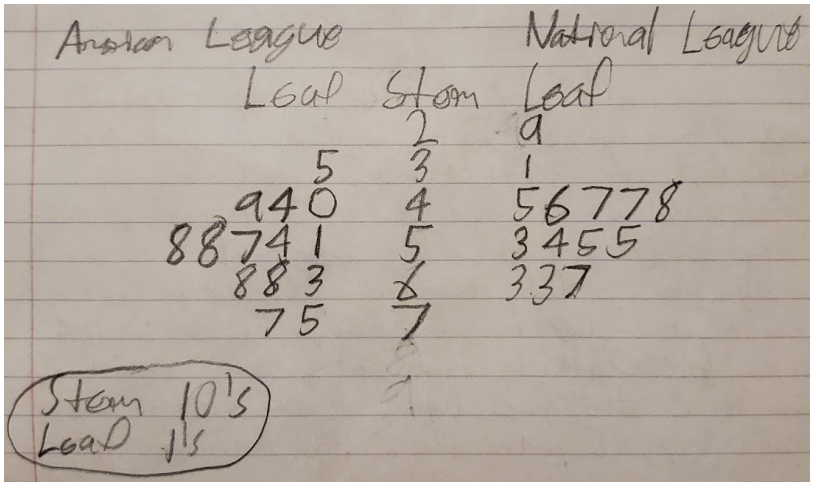
72	72	93	70	59	78	74	96	73	80
91	67	72	88	83	76	74	75	68	67
74	76	79	49	84					



4. During the early part of the 1994 baseball season, many sports fans and baseball players noticed that the number of home runs being hit seemed to be unusually large. Here are the data on the number of home runs hit by American and National League teams:

American League	35, 40, 44, 49, 51, 54, 57, 58, 58, 63, 68, 68, 75, 77
National League	29, 31, 45, 46, 47, 47, 48, 53, 54, 55, 55, 63, 63, 67

Construct back to back stem and leaf plots for the two data sets.



5. A large population has a mean of 27 and a standard deviation of 3.2. A sample of size 50 is taken from the population. The sample mean is 26.6 and the sample standard deviation is 3.5. The values of the population mean and standard deviation are called parameters, while the sample mean and sample standard deviation are called statistics. Cause and effect can sometimes be established in a controlled experiment, but generally not in an observational study. Explain.

The reason why it is easier to establish cause and effect in a controlled experiment compared to an observational study is because when the values of factors are in direct control of the experimenter, an experimenter would be able to conclude that the differences in the results are due to the differences in the factors. In an observational study, an experimenter doesn't control the values of the factors; he/she could only observe things as they are, making it more difficult to say that one thing happened because of a certain factor.

6. Define a Simple Random Sample (SRS) and give an example of a sampling scheme which does not produce an SRS.

A simple random sample of size n is a sample in such a way that each collection of n population items is equally likely to make up the sample; just like in a lottery. One sampling scheme that would not produce an SRS would be if you were to go to the park at 3 in the morning and ask people if they smoke. This wouldn't be a simple random sample because only a certain group of people would be awake and active at 3 in the morning(*especially hanging around in the park at that time). Not only that, but not everyone visits the park, so you aren't truly asking a random group for their opinion.

For questions 7 and 8, state whether the investigation is an observational study or a designed experiment. Explain your answer.

7. Researchers randomly assigned 39,876 initially healthy women 45 years of age or older to 100 mg of aspirin or placebo on alternative days and then monitored them for 10 years for a first major cardiovascular event (i.e. nonfatal myocardial infarction, non-fatal stroke, or death from cardiovascular causes).

This is a designed experiment because researchers themselves are directly assigning women to taking either aspirin or a placebo. Seeing how they are controlling the factors, they can more easily establish a cause-and-effect relationship with aspirin and cardiovascular events.

8. A study involving 3800 people over the age of 65 in East Boston. The study took place over a six year period and compared the age at which death occurred for right handed

people vs. left handed people among the 3800 people who died during the six year period. The study concluded there was no difference.

This is an observational study because the researchers observed things as they are without any interference. Not only that, but it's not like you can just make people be right-handed or left-handed for a study. You can try, but it would be difficult for them to adjust and assuming they've spent years writing with one hand, there's no valid way to make someone opposite-handed overnight.