Notes MTH 332:Mathematical Statistics for Week of 1/25/2020

Monday 1/25:

Completed Steps 1-4 of the analysis

Tuesday 1/26:

Read Chapter 1 to understand data and study that is being analyzed

Wednesday 1/26:

Analyzed data thus far.
Compared mean and median of each data set to determine skewness and then found skewness found from Mathematica matches with this value
Use standard deviation to compare the dispersion of each data set (smokers and non-smokers)
Looked into the Kurtosis Further to understand what it was:
1 st moment : mean
2 nd (centralized) moment : variance
3 rd (centralized) moment : skew
4 th (centralized) moment : Kurtosis

Kurtosis of 3 : Mesokurtic
Kurtosis > 3 : Leptokurtic : higher peak and fatter (longer tails)
Kurtosis < 3 : Platykurtic :
flatter peak and shorter tails -- > Lowest kurtosis is 1
It is with the outliers that kurtosis is formed because
of $(x-\mu)^4$ is so strong and dictates conclusion

Friday 1/29:

Z – score : tells us, for any data point,
how many standard deviations that data point lies above or below the mean
Scales the data according to the standard deviation and
the difference in the sample vs. population mean

$Z = \frac{x_i - \mu}{\sigma}$

$E(z) = 0$

$E(z^2) = 1$

Standard deviation formula :

$\sigma = \sqrt{E((X-\mu)^2)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2}$

Skewness is related to the cube of the z – score
If z – score is less than Abs[1], the skewness contracting back to 0

Then z score is greater than Abs[1], the skewness is increasing exponentially

If there is much more data above the mean than below, the average value of z^3 will be highly positive
If there is much more data below the mean than above, the average value of z^3 will be highly negative

Skewness will tell us if there is more data above the mean or below the mean
If the numbers are very close to the mean, the skewness will approach 0
Balance of z^3 above and below the mean (z = 0)

Skewness is related to the 4 th power of the z – score
The closer the z – score is to 0, the closer the $\mathbf{z}^4$ wil be to 0
If a lot of data is above r below the mean (aka if a lot of the data is in the tail of the distribution),
thetr will be a larger kurtosis

If there was a massive amount of data near the center, the z^4 value will
compile and therefore there will be a larger kurtosis for this case too
The data way above or below the mean are the tails.

Saturday 1/30:

Finished all steps of analysis. And noted what each result meant relative to the other data set.