**Statistics Worksheet 6**

1. D
2. A
3. A
4. C
5. A
6. B
7. C
8. B
9. B
10. What is the difference between a boxplot and histogram?

Histograms and box plots are very similar in that they both help to visualize and describe numeric data. Although histograms are better in determining the underlying distribution of the data, box plots allow you to compare multiple data sets better than histograms as they are less detailed and take up less space.

1. How to select metrics?

prioritize objectives, examine which metric consistently predicts their achievement, and identify which activities influence predictors, in that order. And continuously re-evaluate this process to keep up with the times.

1. How do you assess the statistical significance of an insight?

To assess statistical significance, you would use hypothesis testing. The null hypothesis and alternate hypothesis would be stated first. Second, you’d calculate the p-value, which is the likelihood of getting the test’s observed findings if the null hypothesis is true. Finally, you would select the threshold of significance (alpha) and reject the null hypothesis if the p-value is smaller than the alpha — in other words, the result is statistically significant.

1. Give examples of data that does not have a Gaussian distribution, nor log normal.

Any type of categorical data won't have a gaussian distribution or lognormal distribution. Exponential distributions - eg. the amount of time that a car battery lasts or the amount of time until an earthquake occurs.

1. Give an example where the median is a better measure than the mean.

Income is the classic example of when to use the median instead of the mean because its distribution tends to be skewed. The median indicates that half of all incomes fall below 27581, and half are above it. For these data, the mean overestimates where most household incomes fall.

1. What is the Likelihood?

The likelihood is the probability that a particular outcome is observed when the true value of the parameter is, equivalent to the probability mass on; it is not a probability density over the parameter. The likelihood, should not be confused with, which is the posterior probability of given the data