Summary

Types of Data

- Quantitative: dealing with numbers
- Qualitative: dealing with categorical data. e.g. colors, odors.

Quantitative data has two types:

- Discrete: things that are generally counted not measured like people, number of games in a league, number of goals scored.
- Continuous: Things that get measured like height, weight, and forces.

Scales of Measurement

- Nominal: categorical, order doesn't matter, couldn't measure difference.
- Ordinal: order does matter, could not measure difference. e,g: ranking
- Interval: order does matter, could measure difference, no true zero. e.g: temperature in celesius or Fahrenheit
- Ratio: order does matter, could measure difference, there exists a true zero. e.g: grades,scores.

Hypothesis Testing

A hypothesis is a supposition made on little or no evidence. A null hypothesis is a hypothesis made based on no data. In hypothesis testing, we gather data to prove our primary hypothesis, null hypothesis, wrong. If the data gathered is strong enough to prove our null hypothesis wrong, we reject the null hypothesis, otherwise we fail to reject our null hypothesis. An alternative hypothesis is an opposing hypothesis to our null hypothesis. An alternative hypothesis is tested after our null hypothesis is proven wrong,rejected.

P-Values

P-values quantify how confident we should be in our hypothesis. A value of zero signifies that we should be positive about our hypothesis. A

value of one signifies that our hypothesis is likely to be wrong. Getting a small p-value when our hypothesis is wrong is called a *false positive*. A small p-value does not imply that our hypothesis is definitely right. However, it is a measure of the level of confidence we should have over our hypothesis.

A p-value consists of three parts:

1- The probability of an observation.
2- The probability that something equally likely should occur.
3- The probability of something rare should occur.

Two sided p-values consist of these three parts. One-sided p-values consist of an observation or something in one side of our probability distribution should occur. Two sided p-values are better at detecting rare events vs normal ones.

Confidence Intervals

An interval which a percentage of multiple means acquired from multiple statistical tests fall under. A confidence interval displays the probability that a parameter will fall between a pair of values around the mean.

Regression Analysis

Regression analysis is a method in which we identify the relationships between independent variables and outcomes, dependent variables.