Statistics Part 2

Qualitative Vs Quantitative

Qualitative:

- Description of data based on observations
- Involving the 5 senses

Quantitative:

- Numerical description of data
- Discrete (Counting) – whole numbers only
- Continuous (Measuring) – can be decimals

Scales of Measurement

Scales of measurement are the different ways to categorize and quantify data.

Categories of scales of measurement:

1. Nominal Scale Data:
   - Qualitative form of measurement
   - Order doesn't matter
   - Can't be used in calculation

2. Ordinal Scale Data:

   - Order matters
   - Differences can't be measured

3. Interval Scale Data:

- Order matters
- Differences can be measured
- No true 0 starting point

4. Ratio Scale Data:

- Order matters
- Differences can be measured
- Has a true 0 starting point
- Can measure ratio

Hypothesis Testing

Hypothesis Testing - is a statistical method used to make inferences about a population based on sample data. It involves formulating two opposing hypotheses and using statistical techniques to determine which one is more supported by the data. The process includes defining the hypotheses, selecting a significance level, choosing an appropriate test, calculating the test statistic, and making a decision based on the results.

The **Null Hypothesis (H_0)** is the default assumption that there is no significant effect or difference in the data. It represents the status quo and is tested to determine whether it can be rejected.

The **Alternative Hypothesis (Ha)** is the opposing statement that suggests a significant effect or difference exists. It is what the researcher aims to prove. If statistical analysis provides sufficient evidence against the null hypothesis, it is rejected in favor of the alternative hypothesis.

P-Values

A **p-value** is the probability of obtaining a result as extreme as (or more extreme than) the observed data, assuming the **null hypothesis** is true. It helps determine statistical significance:

- Its value ranges between 0 and 1
- A **low p-value** (typically ≤ 0.05) suggests strong evidence against H_0, leading to its rejection.
- A **high p-value** (> 0.05) indicates weak evidence against H_0, meaning it is not rejected.

How to Calculate a p-value (two-sided):

1. The probability of an event you want
2. The probability of an event with the same rarity
3. The probability of a rarer event
4. Add them together

Confidence Intervals

A **confidence interval** is a range of values used to estimate a population parameter with a certain level of confidence. It provides an interval within which the true mean or proportion is likely to fall.

Key Points:

- A 95% confidence interval means that if we repeated the sampling many times, 95 for every 100 tests would contain the true population value.

Regression Analysis

Regression analysis is a statistical method used to examine the relationship between one or more independent variables (predictors) and a dependent variable (outcome). It helps in making predictions, identifying trends, and understanding associations between variables.

Types of Variables in Regression Analysis:

1. Categorical Variables - Represent groups or categories
    - Example: gender, colors, brands.
2. Continuous Variables - Numeric values that can take an infinite number of possibilities within a range.
    - Example: Age, Salary, Temperature

Predictor Variables (Independent Variables) - The factors that influence or predict changes in the outcome.

Outcome Variable (Dependent Variable) - The result that depends on predictor variables.