STATISTIC COURSE SUMMARIZATION PART 2

Made By
Rawan Hatem

Qualtitative and Quantitative data

- **Quantitative** data deals with numerical information ,**uses numbers**
  - Quantitative data can **divided into**:
    - **Discrete** : Based on counting, whole numbers , it can't have any number in between.
    - **Continuous** : Based on measurement, it can have any number in between.
- **Qualitative** data is descriptive data based on observations and **uses words** to describe data.
  - Qualitative data usually **involves** the **five senses**:
    - visual (color, shape), touch (soft, hard), taste (sweet, sour), hearing (loud, soft), and smell (pleasant, unpleasant).

Scales Of Measurements
Nominal & Ordinal

- **Nominal scale** data: **qualitative, categorical data, labeled**, **no** meaningful **order**, **cannot measure differences**
  - **example**: assigning numbers to colors without quantitative value (e.g., red = 1)
  - Can be used in calculations based on responses (e.g., percentage of favorites)
- **Ordinal scale** data: used for **ranking, order** matters but **differences cannot** be measured (e.g., race positions)
  - **example**: ranking experiences (**excellent, good, satisfactory, bad**)
  - Assigning **numerical values** to categories but **cannot measure differences** between them

Scales Of Measurements
Interval & Ratio

- **Interval scale** data: **order** and **differences** can be **measured** but **ratios cannot measure**, and **not** applicable for **true zero**
  - **example**: temperature (e.g., 30°F, 60°F, 90°F)
  - No true zero starting point (e.g., 0°F is not the lowest temperature)
- **Ratio scale** data: **order, differences, and ratios** can be **measured**, has a **true zero** starting point
  - **example:** grades (e.g., 30, 56, 70, 82, 90)
  - True zero starting point (e.g., **grade of zero** means no questions answered correctly).

Hypothesis Testing

Outlines:

- How can we create a hypothesis ?
- How do you deal with the hypothesis?
- What is the null hypothesis ?

- we can create a hypothesis if :
    a. data gives us strong evidence that the hypothesis is wrong then we can reject the hypothesis
    b. data that is similar to the hypothesis but not exactly the same then the best we can do is fail to reject the hypothesis because it’s unclear
- hypothesis should be based on this result or this other ,slightly different result
- thy hypothesis that there is no difference between things is called the null hypothesis
- The null hypothesis does not require preliminary data, as the **only value that represents no difference is zero.**

Alternative Hypothesis

Outlines:

- What is the alternative hypothesis ?
- What is the importance or alternative hypothesis ?
- When using the alternative hypothesis ?
- What is the three things needed for a statistical test ?

- The alternative hypothesis is the opposite of the null hypothesis.
- The alternative hypothesis is important in statistical tests because it determines whether to reject or fail to reject the null hypothesis.
- When there are three or more groups, the alternative hypothesis becomes more interesting and there are different options for it.
- The alternative hypothesis used in the statistical test can affect the decision about the null hypothesis.
- the three things needed for a statistical test: data, a null hypothesis, and an alternative hypothesis.

P-Values

Intro

- Determining Similarity between two things is called **Hypothesis Testing** when it the same it is called a **Null Hypothesis** and the **P-value** helps us decide if we should reject the Null Hypothesis or not

- **definition of P-values**
  - p-values are numbers between 0 and 1 quantify how confident we should be between two different experiments.

- **false positive is** Getting a small p-value when there is no difference
- using a threshold of 0.00001 mean we would only get a False Positive once every 100.000 experiments
- trying to reduce the number of False Positives below 5%_ if the most common threshold is 0.05 _often costs more than it is worth

P-value helps us decide similarity but it does not tell us how different they are, so pvalue does not imply that the effect size or difference between two things

calculating P-Value

- P-values are used to test hypotheses and are determined by adding up probabilities.
- The null hypothesis in statistics is that there is no difference between the variable being tested and a normal situation.
- A small p-value indicates that the observation is unlikely to have occurred by chance, and therefore the null hypothesis should be rejected.
- Two-sided p-values, which are the most common, test for differences in both directions from the null hypothesis.
- One-sided p-values are less common and can be potentially dangerous because they only test for differences in one direction.
- The p-value is calculated by adding up the probability of getting the observed result and the probability of getting results that are equally or more rare.
- The p-value is then compared to a significance level, typically 0.05, to determine if the null hypothesis should be rejected.
- A p-value is calculated by adding up the area under the curve that is more extreme than the observed result.

Confidence Intervals

- Bootstrapping is one way to calculate confidence intervals, and it is the easiest way to understand them according to the speaker.
- Bootstrapping involves randomly selecting weights from the original sample with replacement, calculating the mean of the new sample, and repeating the process to generate many mean values.
- A confidence interval is an interval that covers a certain percentage of the means.
- Confidence intervals are useful for statistical tests performed visually because they allow us to determine the p-value of a hypothesis.
- If the p-value is less than 0.05, the hypothesis is significantly different from the confidence interval.
- Confidence intervals can also be used to compare two samples. If the confidence intervals do not overlap, there is a statistically significant difference between the two means.

Regression Analysis

- Analysis of regression is a technique to analyze relationships between variables to better understand the world around us.
- Variables are data elements that describe an attribute or characteristic of an object or participant. They can be categorical, placing objects or participants in discrete groups, or continuous, placing an object or participants somewhere on a scale based on a specific attribute.
- In data analysis, variables can be affected as predictors (independent variables) or results (dependent variables).
- A line of regression can be drawn to estimate or model the relationship between independent and dependent variables.
- One advantage of regression analysis is that a simple model can be extended to include more independent variables, allowing for adjustment of confounding variables.
- Confounding variables are linked to both the independent and dependent variables but are not a factor intermediary.
- More variables can be added to regression models, making it a powerful method for understanding the relationship between variables.