Write Up Report

Our main topic is about Hypothesis Testing and Causal Inference, which is a powerful concept in data analysis and research in statistics. It allows us to explore cause-and-effect relationships between variables or events based on observed data in real life, including healthcare, education, economics, and public policy.

Hypothesis testing is more like a statistical tool to make inferences about population parameters based on sample data. It involves formulating a null hypothesis (H0) and an alternative hypothesis (H1), conducting a statistical test, and drawing conclusions about the null hypothesis's rejection or acceptance.

We talked about the two-sample t-test and its comparison with the paired t-test.

Two-sampled t-test

- Used to determine if the two-population means are the same.
- Common application: if a new process or treatment is superior to the current process or treatment.

Example: *body fat percentage among women and men who work at gym three times a week for a year*

Paired t-test (dependent sample t-test)

- Used to test whether the mean difference between pairs of measurements is zero or not
- each subject or entity is measured twice

Example: (*before-and-after weight for a smoker in the* example *above must be from the same person*

What are the assumptions for these tests?

Two-sampled t-test

- Data values must be **independent**, measurements for one observation do not affect the measurements for any other measurements
- Data in each group must be obtained via **a random** sample from the population
- **Normally distributed (**looking at histogram and eyeball, no outlier)
- Data values are continuous
- The variance of two groups is equal (check the sd, similar is fine, 5.32, 6.84)
- Sample size
-

Paired t-test (dependent sample t-test)

- Data values must be **independent**, measurements for one observation do not affect the measurements for any other measurements
- Each of the paired measurements must be obtained from the same subject

(*before-and-after weight for a smoker in the example above must be from the same person)*

- Normally distributed

A motivated question being asked related to causal inference is what do you think of causation?

- It marks the relationship between two variables and events. They are dependent on each other and if one happened, it would have an impact on the other one. Often talked with correlation, but causation does not only mean two things are related and they have a certain direction of causation.

Talking about why we care about casual inference can be illustrated particularly in the context of A/B testing. A/B testing allows tech companies to evaluate the impact of new features or changes objectively by randomly assigning users to different groups and comparing their outcomes. For instance, in the case of a social media platform introducing a new Story feature.

1. Users are divided into two groups: one group exposed to the new feature and another group serving as the control.
2. By analyzing the outcomes between the experimental and control groups, companies can determine whether the new feature causally influences user behavior and engagement.

However, it's essential to consider confounding variables—factors that are correlated with both the treatment (new feature) and the outcome. These variables can potentially bias the results if not properly accounted for, like device type, relevant content for users, and so on. To avoid this, randomized controlled trials (RCTs) help mitigate confounder bias by randomizing the assignment of participants to treatment groups, thus allowing researchers to estimate causal effects accurately.

In conclusion, causal inference through methods like A/B testing and RCTs empowers tech companies to make data-driven decisions, enhance user experiences, and drive platform growth by understanding the true impact of their interventions. It enables researchers to quantify uncertainty and draw reliable conclusions about cause-and-effect relationships, contributing to advancements in technology and innovation.