Shri Vile Parle Kelavani Mandal's
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)
NAAC Accredited with "A" Grade (CGPA : 3.18)

Department of Computer Science and Engineering (Data Science)

Subject: Applied Data Science (DJ19DSL703)

Experiment - 8

(A/B Testing)

Bhuvi Ghosh
60009210191

Aim: To prepare a case study on A/B Testing.

Theory:

A/B Testing:

A/B testing is one of the most important concepts in data science and in the tech world in general because it is one of the most effective methods in making conclusions about any hypothesis one may have. It's important that you understand what A/B testing is and how it generally works.

A/B testing is a common methodology to test new products or new features, especially regarding user interface, marketing and e-commerce. The main principle of an A/B test is to split users into two groups; showing the existing product or feature to the control group and the new product or feature to the experiment group. Finally, evaluating how users respond differently in two groups and deciding which version is better. Even though A/B testing is a common practice of online businesses, a lot can easily go wrong from setting up the experiment to interpreting the results correctly.

A/B testing is a form of statistical and two-sample hypothesis testing. Statistical hypothesis testing is a method in which a sample dataset is compared against the population data. Two-sample hypothesis testing is a method in determining whether the differences between the two samples are statistically significant or not.

Formulate your hypothesis:

Before conducting an A/B testing, you want to state your null hypothesis and alternative hypothesis:
The null hypothesis is one that states that sample observations result purely from chance. From an A/B test perspective, the null hypothesis states that there is no difference between the control and variant group.

The alternative hypothesis is one that states that sample observations are influenced by some non-random cause. From an A/B test perspective, the alternative hypothesis states that there is a difference between the control and variant group.

When developing your null and alternative hypotheses, it's recommended that you follow a PICOT format.

PICOT stands for:

- Population: the group of people that participate in the experiment
- Intervention: refers to the new variant in the study

- Comparison: refers to what you plan on using as a reference group to compare against your intervention
- Outcome: represents what result you plan on measuring
- Time: refers to the duration of the experience (when and how long the data is collected)

Example: "Intervention A will improve anxiety (as measured by the mean change from baseline in the HADS anxiety subscale) in cancer patients with clinical levels of anxiety at 3 months compared to the control intervention."

Does it follow the PICOT criteria?
Population: Cancer patients with clinical levels of anxiety
Intervention: Intervention A
Comparison: the control intervention
Outcome: improve anxiety as measured by the mean change from baseline in the HADS anxiety subscale
Time: at 3 months compared to the control intervention.

Yes, it does — therefore, this is an example of a strong hypothesis test.

Create your control group and test group

Once you determine your null and alternative hypothesis, the next step is to create your control and test (variant) group. There are two important concepts to consider in this step, random samplings and sample size.

Random Sampling
Random sampling is a technique where each sample in a population has an equal chance of being chosen. Random sampling is important in hypothesis testing because it eliminates sampling bias, and it's important to eliminate bias because you want the results of your A/B test to be representative of the entire population rather than the sample itself.
Sample Size
It's essential that you determine the minimum sample size for your A/B test prior to conducting it so that you can eliminate under coverage bias, bias from sampling too few observations.

Conduct the test, compare the results, and reject or do not reject the null hypothesis:

Once you conduct your experiment and collect your data, you want to determine if the difference between your control group and variant group is statistically significant. There are a few steps in determining this:
First, you want to set your alpha, the probability of making a type 1 error. Typically, the alpha is set at 5% or 0.05
Next, you want to determine the probability value (p-value) by first calculating the t-statistic using the formula above.
Lastly, compare the p-value to the alpha. If the p-value is greater than the alpha, do not reject the null!

Lab Assignment:

Prepare a case study based on data science project to formulate an A/B testing.

Shri Vile Parle Kelavani Mandal's
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)
NAAC Accredited with "A" Grade (CGPA : 3.18)

Department of Computer Science and Engineering (Data Science)

Case Study: A/B Testing for an E-commerce Website

Aim

To evaluate the impact of a new product page design on user engagement and sales conversion rates.

Theory

A/B Testing

A/B testing is a statistical method used to compare two versions of a product to determine which performs better. In this case, we aim to understand whether a redesigned product page (Variant A) increases user engagement and sales conversion compared to the existing design (Control B).

Hypothesis Formulation

- Null Hypothesis (H0): There is no significant difference in user engagement and conversion rates between the existing product page design and the new design.
- Alternative Hypothesis (H1): The new product page design significantly improves user engagement and conversion rates compared to the existing design.

PICOT Format

- Population: Users visiting the e-commerce website.
- Intervention: The new product page design (Variant A).
- Comparison: The existing product page design (Control B).
- Outcome: Measure user engagement (average time spent on the page) and conversion rates (percentage of users making a purchase).
- Time: Data will be collected over a two-week period.

Creating Control and Test Groups

Random Sampling

Users will be randomly assigned to either the control group (existing design) or the test group (new design) to eliminate bias.

Sample Size

The minimum sample size will be calculated using power analysis to ensure statistical significance. For this case study, we aim for a sample size of at least 1,000 users per group.

Conducting the Test

Steps to Analyze Results

1. Set Alpha Level: The significance level (alpha) will be set at 0.05.
2. Collect Data: User engagement metrics and conversion rates will be collected during the two-week testing period.
3. Calculate T-statistic and P-value:
    - Calculate the means and standard deviations for both groups.
    - Use the t-test formula to calculate the t-statistic.
    - Determine the p-value associated with the t-statistic.

Decision Making

- If the p-value is less than 0.05, we reject the null hypothesis, indicating that the new design has a significant impact on user engagement and conversion rates.
- If the p-value is greater than 0.05, we do not reject the null hypothesis, indicating no significant difference between the two designs.

Expected Outcomes

The expected outcome is to determine whether the new product page design leads to higher user engagement and conversion rates compared to the existing design. This information will guide future design decisions and improvements on the e-commerce platform.

Conclusion

By conducting this A/B test, we aim to provide data-driven insights into the effectiveness of design changes in e-commerce, ultimately improving user experience and sales performance.