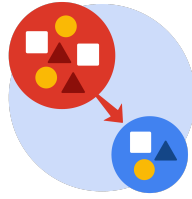


Course Four

From Data to Insight: The Power of Statistics



Instructions

Use this PACE strategy document to record decisions and reflections as you work through this end-of-course project. As a reminder, this document is a resource that you can reference in the future, and a guide to help you consider responses and reflections posed at various points throughout projects.

Course Project Recap

Regardless of which track you have chosen to complete, your goals for this project are:

- ☒ Complete the questions in the Course 4 PACE strategy document
- ☒ Answer the questions in the Jupyter notebook project file
- ☒ Compute descriptive statistics
- ☒ Conduct a hypothesis test
- ☒ Create an executive summary for external stakeholders

Relevant Interview Questions

Completing this end-of-course project will empower you to respond to the following interview topics:

- How would you explain an A/B test to stakeholders who may not be familiar with analytics?
- If you had access to company performance data, what statistical tests might be useful to help understand performance?
- What considerations would you think about when presenting results to make sure they have an impact or have achieved the desired results?
- What are some effective ways to communicate statistical concepts/methods to a non-technical audience?
- In your own words, explain the factors that go into an experimental design for designs such as A/B tests.

Reference Guide

This project has four tasks; the visual below identifies how the stages of PACE are incorporated across those tasks.



Data Project Questions & Considerations



PACE: Plan Stage

- What is the main purpose of this project?

The primary goal of this project is to develop a model that can identify and tag claims within user-reported content.

- What is your research question for this project?

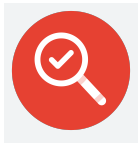
The research question examines whether there is a statistically significant difference in the number of views between TikTok videos posted by verified and unverified accounts.

- What is the importance of random sampling?

Random sampling is important because it minimizes unintentional bias, ensuring that the sample is fair and representative of the population. This leads to more accurate insights and reliable predictions.

- Give an example of sampling bias that might occur if you didn't use random sampling.

An example of sampling bias that can occur without random sampling is **undercoverage bias**, which happens when certain groups are underrepresented in the sample. This often occurs in **convenience sampling**, a non-random sampling technique. For instance, if a researcher collects data by surveying people near an educational institute, they may only capture responses from students and staff who are present at that time. If a particular class or group of employees has the day off, their perspectives will be missing from the sample, leading to biased results.



PACE: Analyze & Construct Stages

- In general, why are descriptive statistics useful?

Descriptive statistics provide a quick and convenient way to understand the structure of a dataset. They summarize key aspects such as central tendency, dispersion, and the distribution of data points, offering an overall view of patterns and trends within the dataset.

- How did computing descriptive statistics help you analyze your data?

Computing descriptive statistics provides an overall view of the data, serving as a starting point for analysis. It helps identify central tendency, dispersion, and positional measures, offering key insights into the structure and distribution of the dataset.

- In hypothesis testing, what is the difference between the null hypothesis and the alternative hypothesis?

The null hypothesis represents a statement of no effect, no difference, or equality in the population, suggesting that any observed differences in the sample data are due to random variability. It is not accepted but rather failed to be rejected if there isn't enough evidence against it. The alternative hypothesis contradicts the null hypothesis by stating that there is a statistically significant difference or effect in the population. If sufficient evidence is found, the null hypothesis is rejected in favor of the alternative hypothesis.

- How did you formulate your null hypothesis and alternative hypothesis?

I formulated the hypotheses by first identifying the claim that needs to be tested for its validity at the population level. This claim, which suggests a difference or effect, forms the basis of the alternative hypothesis. The null hypothesis is then formulated as its contradiction, stating that there is no difference or no statistically significant effect.

- What conclusion can be drawn from the hypothesis test?

The conclusion of a hypothesis test depends on the type of test conducted, such as a one-sample or two-sample z-test or t-test, and whether it is a one-tailed or two-tailed test. However, the general conclusion is whether the observed difference between the values (either between two sample values or between a sample and a benchmark value) is statistically significant or simply due to random sampling variability. If the null hypothesis is rejected, it suggests a statistically significant difference; otherwise, there is not enough evidence to conclude a significant difference.



PACE: Execute Stage

- What key business or organizational insight(s) emerged from your A/B test?

The A/B test revealed that the mean video view counts for verified and non-verified accounts have a statistically significant difference. This insight suggests that verified accounts might post content that naturally garners higher views, while non-verified accounts may rely on more engaging or clickbait-style content to attract viewers. This relationship is strong enough to explore further and consider verified status as a potential factor influencing video engagement. Understanding these patterns could help platforms refine their content recommendation algorithms and optimize strategies for audience engagement.

- What recommendations do you propose based on your results?

Based on the results, I recommend conducting further research to explore whether unverified accounts are more likely to post controversial or clickbait content. Gaining a deeper understanding of this behavior could provide valuable insights into content strategies for both verified and non-verified accounts. Additionally, this research can inform future steps, such as refining content creation strategies or enhancing algorithms for content recommendation based on verified status and engagement patterns.