A/B Testing

Introduction

You've now seen all of the statistical techniques and background to design and conduct your own A/B tests in practice! To do this, you'll go through the process of stating the null hypothesis and the alternative hypothesis which will include some test statistic for comparison. For example, you might compare the average purchase price between customers on two different versions of your online store, or a pharmaceutical researcher might compare the blood pressure of patients before and after taking a prescription. You've also seen that good test design requires multiple decisions. Recall both the multiple comparisons problem and Goodheart's law: you can't effectively measure everything (without increasing the risk of mistakes), and incentivizing measurements can lead to unforeseen consequences. With that, this section will give you a chance to put your new statistical techniques into practical applications.

Objectives

You will be able to:

- List the steps required to design, structure, and run an A/B test

Choosing a Metric

Any hypothesis testing will start with a given scenario. This will determine everything from what metrics are deemed important to realistically obtainable sample sizes. Goodheart's law can also be an important consideration when performing ongoing tests in attempts to optimize performance metrics.

Defining the Null Hypothesis: H_0

Once an appropriate metric has been selected, it's time to formally define the experiment with a null hypothesis. Typically, the null hypothesis is the claim that a researcher is hoping to refute. For example, a medical researcher might hope to show that a new drug is more effective than a previous treatment option. A common practice is then to define the null hypothesis as the contrary: there is no difference between the two drugs. The researcher hopes to refute the null hypothesis thereby proving their claim by contradiction.

You might start with something like "drug_a is more effective than drug_b".

While this is a good start, proper formulation of the null hypothesis should ensure that it is written with a quantitative measurement. Perhaps the drugs are for high blood pressure and so the statement becomes, "drug_a at dose_a lowers blood pressure more than drug_b at dose_b".

Formulating the null-hypothesis like this is apt to lead you into conducting a paired t-test for the mean blood pressure of two groups: one representing drug_a, and another representing drug_b.

Alternatively, if one of these two medications were more heavily researched, one might wish to compare the effectiveness of the new medication with a predetermined metric such as the average drop in blood pressure from medication. This would lead to a 1-sample t-test, as opposed to a two-sample t-test.

Investigating α, power, effect size, and sample size

Finally, one must formulate the various surrounding parameters required to conduct the test. You've seen that there is an intimate relationship between α, power, sample size, and effect size. With that, questions such as "How costly is sample size?" are instrumental in experiment design. For example, in an online scenario, it might be quite easy to conduct experiments at scale. On the other hand, in medical research, larger sample sizes are apt to be extremely costly. Investigating the relationships between α, power, effect size, and sample size is important for all experiments, and a suitable combination will depend on these contextual factors regarding implementation.

Summary

When researching, you are often presented with two choices for stating a question. One is to estimate a parameter in question, such as the procedures previously examined for estimating the mean of a population. Alternatively, you may wish to test the validity of a claim—whether you can refute that claim, or whether you should withhold judgment. In practice, it is up to the practitioner to determine the appropriate alpha, beta, and sample size that is determined to be both satisfactory confidence and a viable sample

size to attain. In the upcoming labs, you'll get to practice this setup and design work-flow for a few scenarios.