Chapter 3: Statistical Experiments and Statistical Testing

The classical statistical inference pipeline

- Data scientists often need to design an experiment in order to test an initial hypothesis.
- Hypothesis formulation and experiment design are part of the classical statistical inference pipeline.
- The classical statistical inference pipeline involves the steps to draw a conclusion (about the population) from a limited set of data.

Formulate hypothesis → Design Experiment → Collect Data → Inference / Conclusions

A/B testing | What it is

- An A/B test is an experiment with two (2) groups to establish which of two (2) treatments is superior.

- An A/B test has subjects (ex. customers, web site visitors etc.) that are assigned randomly to one treatment or another (randomization).

- The difference between the treatment groups are due to either the effect of different treatment or random chance.

A/B testing │ The test statistic

- *Test statistic*: The metric to compare group A to group B. It can be binary or continuous.

- When the test statistic is binary, the testing result is a 2x2 table (rows: 0 & 1 result, columns: A & B group).

- When the test statistic is continuous, the testing results are two distributions (the test statistic distribution for each of the two groups). Softwares usually describe the test statistic distributions by their mean and standard deviation.

A/B testing | The control group

- Often one of the two treatments is the existing treatment, or no treatment. A typical hypothesis is that the new treatment is better than the existing one.

- *Control group*: A group of subjects exposed to no (or standard) treatment; *Treatment group:* A group of subjects exposed to a specific treatment.

- The control group allows us to assume that the experiment conditions are the same (except the type of the treatment) on both groups. This way no other factor (beside the different treatment and random chance) can affect the results.

Hypothesis tests

- A hypothesis test is used to examine if the difference observed between group A and group B is due to random chance.

- If the observed difference is random, treatment A and treatment B are considered to be equivalent.

- The hypothesis test starts with the assumption that treatment A and treatment B are equivalent (null hypothesis). In the end, the null hypothesis is either rejected or failed to be rejected.

The alternative hypothesis

- The alternative hypothesis is the opposite of the null hypothesis. In other words, it assumes a difference between treatment A and treatment B.

- When there are more than two treatments to compare, there are more than one alternative hypothesis that we can make.

- Even though the result of the hypothesis test is about the null hypothesis (reject or fail to reject), the alternative hypothesis plays a role to this outcome.

StatQuest on Alternative Hypotheses

The null model | Getting a bit deeper to hypothesis tests

'The hypothesis test assumes that the null hypothesis is true, creates a "null model" (a probability model), and tests whether the effect you observe is a reasonable outcome of that model' - pg.96

- The null model assumes that group A and group B are subgroups of the same initial set.
- We can create a null model by shuffling the results of group A and group B and then resampling from the combined set (resampling permutation procedure).
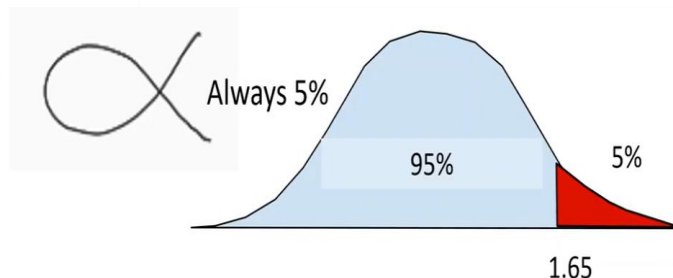
One–way vs two–way hypothesis tests

- When an A/B test examines if a new treatment is better than an established one, an one-way hypothesis test is used.

- When an A/B test examines if there is difference (either negative or positive) between two treatments, a two-way hypothesis test is used.

Two-way test

One-way test

Always 5%

2.5%   95%   2.5%

-1.96   1.96

Always 5%

95%   5%

1.65
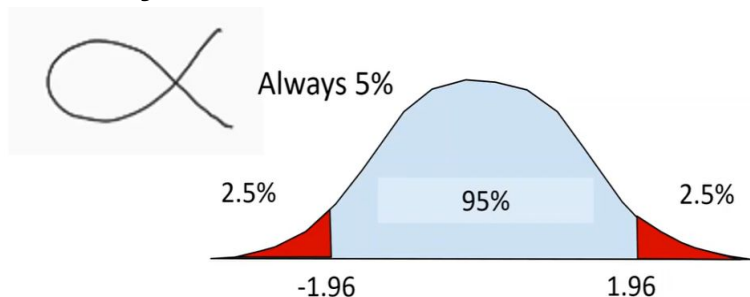
One tail vs Two tail hypothesis tests

This was a summary created by the members of the #66DaysOfData Study Group Rea Kalampaliki and William Guesdon.