Data 8, Lab 7

A/B Testing

Hubert Luo

Fall 2019

25 October 2019

Berkeley
UNIVERSITY OF CALIFORNIA

Why A/B Testing?

- Used to test if two observed distributions (sets of values) are from the same underlying distribution
- Example:
  - Distribution 1: Weights of babies of moms who smoke
  - Distribution 2: Weights of babies of moms who do not smoke
  - Question: Are these distributions from the same underlying distribution?
- An A/B test is an example of a permutation test!

A/B Testing Hypothesis

- **Null Hypothesis:** The two distributions are from the same underlying distribution, and any differences are due to chance
  - Example: The distribution of birth weights of babies is the same for mothers who don't smoke as for mothers who do. The difference in the sample is due to chance.
- **Alternative Hypothesis:** The two distributions are **not** from the same underlying distribution, and any differences are **not** due to chance
  - Example: The babies of the mothers who smoke have a lower birth weight, on average, than the babies of the non-smokers

A/B Test Statistic

- Main Idea: Get two representations of two distributions and find the difference between them
- Example:
  - Difference between the average weights of smoking and non-smoking babies
- Under the null hypothesis, there should be no difference between the two distributions

A/B Testing: Shuffle Labels

- Shuffle the **labels** of the table: if the data are from the same distributions, rearranging the labels won't matter!

Maternal Smoker	Birth Weight
False	120
False	113
True	128
True	108
False	136
False	138
False	132
False	120
True	143
False	140

Shuffled Label
False
False
True
False
True
True
True
False
True
True

A/B Testing: Shuffle Labels (Cont'd)

- How do we shuffle?
  - Sample the original labels of the table **without replacement**
  - Example: If the table has 500 rows, sample 500 labels without replacement
- Why without replacement?
  - This is so we keep the same proportions of labels each time!
- Command: tbl.sample(n, with_replacement = False)
  - Need to specify without replacement!! With replacement is True by default

A/B Testing: Conclusions

- The rest of the procedure is similar to other hypothesis tests
- For a large number of iterations:
  - Shuffle the labels of the data (assuming the null is true)
  - Get the test statistic from the shuffled data
  - Store the test statistic
- Compare the observed test statistic with all the simulated test statistic
- P-value is the proportion of simulated test statistics equal to the observed test statistic or further in the direction of the alternative, assuming the null hypothesis is true

A/B Testing vs. TVD

- A/B Testing: Test whether two samples of data come from the same/different distribution
- TVD: Compute distance between two different samples from the same distribution

Announcements

- Mid-Semester Feedback Form: https://forms.gle/kUcX6VVCfyrbD7GEA (Please fill out by Sunday, October 27th)
- Lab 7 is brand new because of the power outage, so please be patient with any errors!
- Homework 8 is due next Thursday (10/31)
- Project 2: first checkpoint is due next Friday (11/1)

Lab Notebook

TBD