



Lecture 18

A/B Testing

Slides created by John DeNero (denero@berkeley.edu) and Ani Adhikari (adhikari@berkeley.edu)
Contributions by Fahad Kamran (fhdkmrn@berkeley.edu) and Vinitra Swamy (vinitra@berkeley.edu)

Announcements

Review

Testing Hypotheses

- Select a Null and Alternative Hypothesis
 - The null is a fully specified chance model we can simulate under
 - The alternative is some other viewpoint of the world
 - Choose a Test Statistic
 - This test statistic should help us determine between our two viewpoints
 - **Either** large values of the test statistic or small values should be evidence for our alternative
 - Simulate the test statistic under the null hypothesis to create an empirical distribution
 - Approximates the probability distribution of the statistic under the null
 - Calculate our observed test statistic
 - Compare our observed test statistic with values our null predicted
-

Definition of the P -value

Formal name: **observed significance level**

The P -value is the chance,

- under the null hypothesis,
 - that the test statistic
 - is equal to the value that was observed in the data
 - or is even further in the direction of the alternative.
-

Using the P -value

- If the P -value is small, that is evidence against the null hypothesis
 - Conventions about “small”:
 - Less than 5% (result is called statistically significant)
 - Less than 1% (result is called highly statistically significant)
-

A/B Testing

Comparing Two Samples

- Compare values of sampled individuals in Group A with values of sampled individuals in Group B.
- Question: Do the two sets of values come from the same underlying distribution?
- Answering this question by performing a statistical test is called **A/B testing**.

(Demo)

The Groups and the Question

- Random sample of mothers of newborns. Compare:
 - (A) Birth weights of babies of mothers who smoked during pregnancy
 - (B) Birth weights of babies of mothers who didn't smoke
 - Question: Could the difference be due to chance alone? Or do birth weights for mothers who smoked come from a distribution with a smaller average?
-

Hypotheses

- Null:
 - In the population, the distributions of the birth weights of the babies in the two groups are the same. (They are different in the sample just due to chance.)
 - Alternative:
 - In the population, the babies of the mothers who smoked were lighter, on average, than the babies of the non-smokers.
-

Test Statistic

- Group A: smokers
 - Group B: non-smokers

 - Statistic: Difference between average weights
Group A average - Group B average

 - Small values of this statistic favor the alternative
-

Simulating Under the Null

- If the null is true, all rearrangements of the birth weights among the two groups are equally likely
- Plan:
 - Shuffle all the birth weights
 - Assign some to “Group A” and the rest to “Group B”, maintaining the two sample sizes
 - Find the difference between the averages of the two shuffled groups
 - Repeat

(Demo)

Deflategate

2015 AFC Championship Game



Deflategate

Wikipedia:

The 2015 AFC Championship Game football tampering scandal, commonly referred to as Deflategate, or Ballghazi

...

'Deflategate' returns, focus on Tom Brady's destroyed cellphone

POSTED 9:54 AM, MARCH 5, 2016, BY [CNN WIRE](#), UPDATED AT 10:33AM, MARCH 5, 2016

(Demo)

Null hypothesis

The 4 Colts footballs are like a sample drawn at random without replacement from all 15 balls.

- To test this hypothesis, repeat this process:
 - Randomly permute all 15 balls
 - Label 11 of them “Patriots” and the remaining 4 “Colts”
 - Compare the averages of the two groups

(Demo)
