DATA 8

Fall 2022

Lecture 19

A/B Testing

Announcements

- **Homework 7** due Wednesday 10/12

- **Lab 6** due today

- **Midterm** is 10/14 at 7pm (one week!)
  - [Midterm Prep Post](#)

Weekly Goals

- Monday
    - Comparing distributions
    - Hypothesis tests

- Wednesday
    - Making decisions when visualizations don't suffice
    - Tail areas and p-values

- **Today**
    - A/B testing
    - Permutation Test

P-Value Review

The p-Value as an Area

- Empirical distribution of the test statistic **under the null hypothesis**.

- Red dot denotes the observed statistic.

- Yellow area denotes the tail probability (p-value).

Distribution under the Null Hypothesis
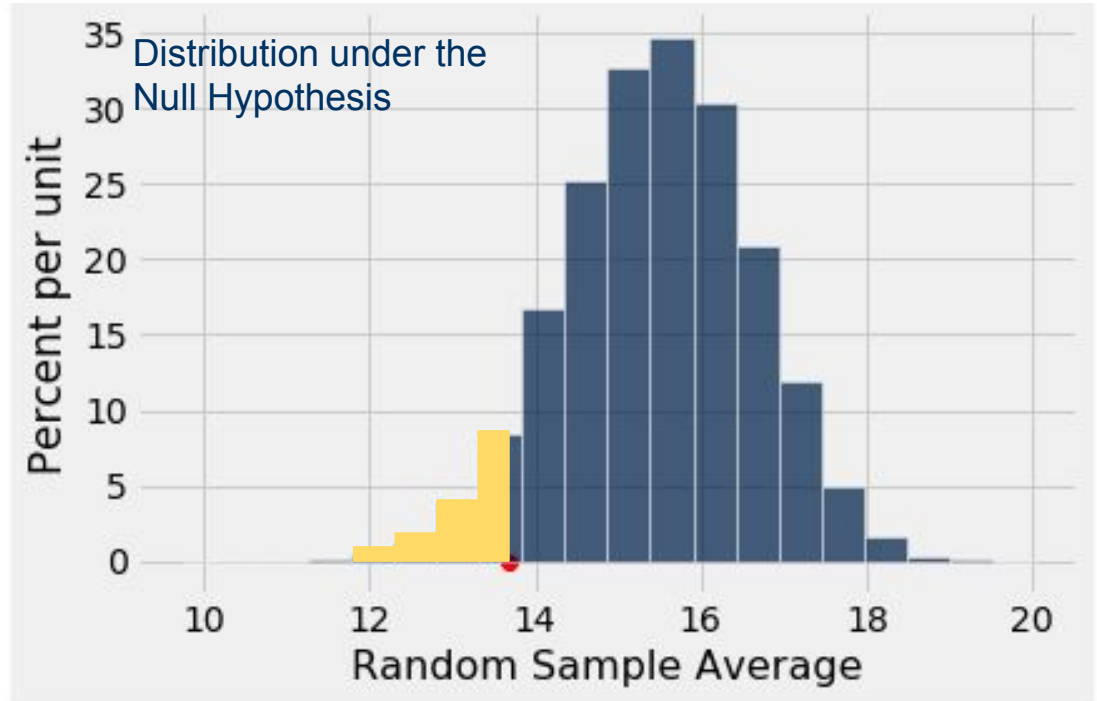
Definition of the *p*-value

Formal name: **observed significance level**

The *p*-value is the chance (probability),

- 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.

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?

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 weigh less, on average, than the babies of the non-smokers.

Test Statistic

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

- Statistic: Difference between average weights

    Group B average - Group A average

- Negative values of this statistic favor the alternative

(Demo)

The Data

Non-smoker	Non-smoker	Smoker	Smoker		Non-smoker
120 oz	113 oz	128 oz	108 oz	. . .	117 oz

Shuffling Labels Under the Null

Smoker	Non-smoker	Non-smoker	Smoker	...	Smoker
120 oz	113 oz	128 oz	108 oz		117 oz

Shuffling Rows

Random Permutation

- **`tbl.sample(n)`**
  - Table of n rows picked randomly with replacement
- **`tbl.sample()`**
  - Table with same number of rows as original **`tbl`**, picked randomly with replacement
- **`tbl.sample(n, with_replacement = False)`**
  - Table of n rows picked randomly without replacement
- **`tbl.sample(with_replacement = False)`**
  - All rows of tbl, in random order

(Demo)

Simulating Under the Null

- If the null is true, all rearrangements of labels are equally likely
- Plan:
  - Shuffle all group labels
  - Assign each shuffled label to a birth weight
  - Find the difference between the averages of the two shuffled groups
  - Repeat

(Demo)

How We've Tested Thus Far

Hypothesis Testing Review

- **1 Sample: One Category** *(e.g. percent of flowers that are purple)*
  - Test Statistic: `observed_proportion` , `abs(observed_proportion - null_proportion)`
  - How to Simulate: `sample_proportions(n, null_dist)`

- **1 Sample: More Than 2 Categories** *(e.g. ethnicity distribution of jury panel)*
  - Test Statistic: `tvd(observed_dist, null_dist)`
  - How to Simulate: `sample_proportions(n, null_dist)`

- **1 Sample: Numerical Data** *(e.g. scores in a lab section)*
  - Test Statistic: `observed_mean, abs(observed_mean - null_mean)`
  - How to Simulate: `population_data.sample(n, with_replacement=False)`

- **2 Samples: Underlying Values** *(e.g. birth weights of smokers vs. non-smokers)*
  - Test Statistic: `group_a_mean - group_b_mean`, `group_b_mean - group_a_mean`, `abs(group_a_mean - group_b_mean)`
  - How to Simulate: `observed_data.sample(with_replacement=False)`