MATH1401

Lecture 19

A/B Testing

Class Checklist

- **Lab 6 – Due Date** : Friday 11/5 – 9 PM

Review: Statistical Significance

Conventions About Inconsistency

- **"Inconsistent with the null":** The test statistic is in the tail of the empirical distribution under the null hypothesis

- **"In the tail," first convention:**
  - The area in the tail is less than 5%
  - The result is "statistically significant"

- **"In the tail," second convention:**
  - The area in the tail is less than 1%
  - The result is "highly statistically significant"

The P-Value as an Area

Empirical distribution of the test statistic under the null hypothesis

The red dot is the observed statistic.
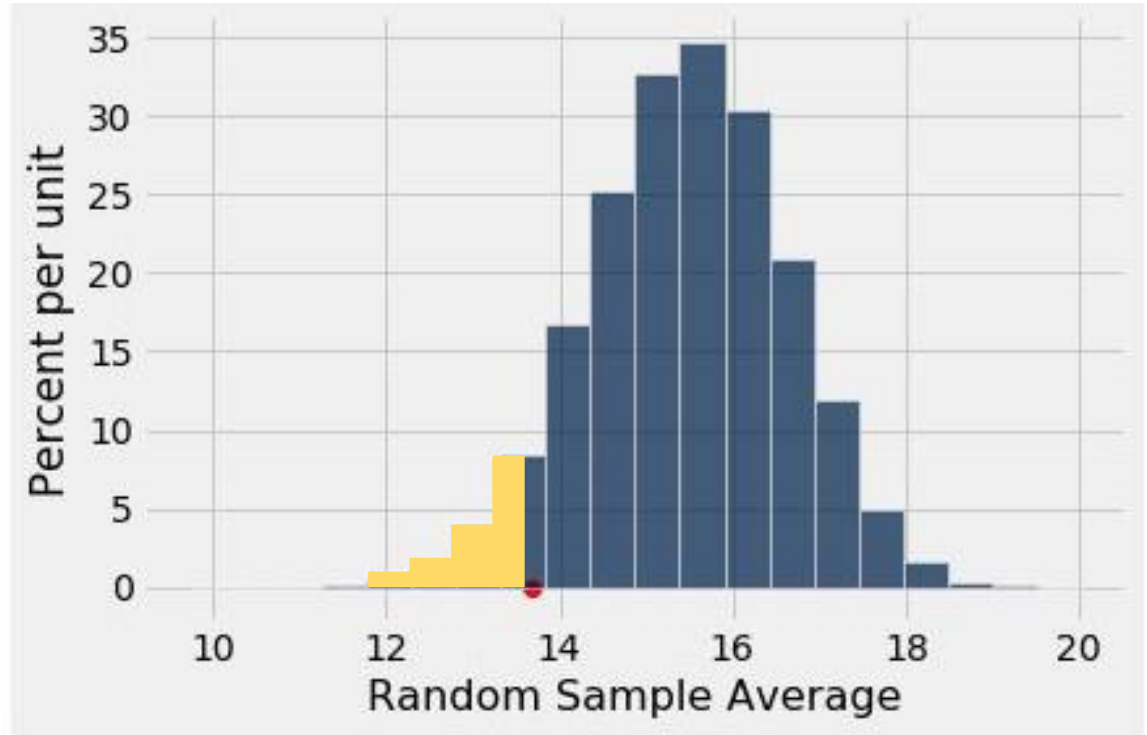
Definition of the *P*-value

Formal name: **observed significance level**

The *P*-value is the chance,

- if the null hypothesis is true,
- 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

120 oz

Non-smoker

113 oz

Smoker

128 oz

Smoker

108 oz

Non-smoker

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)