YaleNUSCollege

YSC2239 Lecture 8

Recap

- Population, Sample, Parameter, Statistic
- Probability Distribution, Empirical Distribution
- Case study: jury selection (recommended)
- sample_proportions(sample_size, probabilities)

Today’s class

- Decisions and Uncertainty
- p-values
- A/B testing

- Reading: Chapter 11, 12

Decisions and Uncertainty

Incomplete Information

- We are trying to choose between two views of the world, based on data in a sample.
- It is not always clear whether the data are consistent with one view or the other.
- Random samples can turn out quite extreme. It is unlikely, but possible.

Terminology

Testing Hypotheses

- A test chooses between two views of how data were generated
- E.g. Alameda jury panel in the previous lecture: truly random panel or is it a biased panel?
- The views are called **hypotheses**
- The test picks the hypothesis that is better supported by the observed data

Null and Alternative

The method only works if we can simulate data under one of the hypotheses.

- **Null hypothesis**
  - A well defined chance model about how the data were generated
  - We can simulate data under the assumptions of this model – “under the null hypothesis”
- **Alternative hypothesis**
  - A different view about the origin of the data

Test Statistic

- The statistic that we choose to simulate, to decide between the two hypotheses

- E.g. total variation distance between the proportion of a randomly generated jury panel and the population proportion in Alameda

Questions before choosing the statistic:

- What values of the statistic will make us lean towards the null hypothesis?

- What values will make us lean towards the alternative?
  - Preferably, the answer should be just “high”. Try to avoid “both high and low”.

Prediction Under the Null Hypothesis

- Simulate the test statistic under the null hypothesis; draw the histogram of the simulated values
- This displays the **empirical distribution of the statistic under the null hypothesis**
- It is a prediction about the statistic, made by the null hypothesis
  - It shows all the likely values of the statistic
  - Also how likely they are **(if the null hypothesis is true)**
- The probabilities are approximate, because we can't generate all the possible random samples

Conclusion of the Test

Resolve choice between null and alternative hypotheses

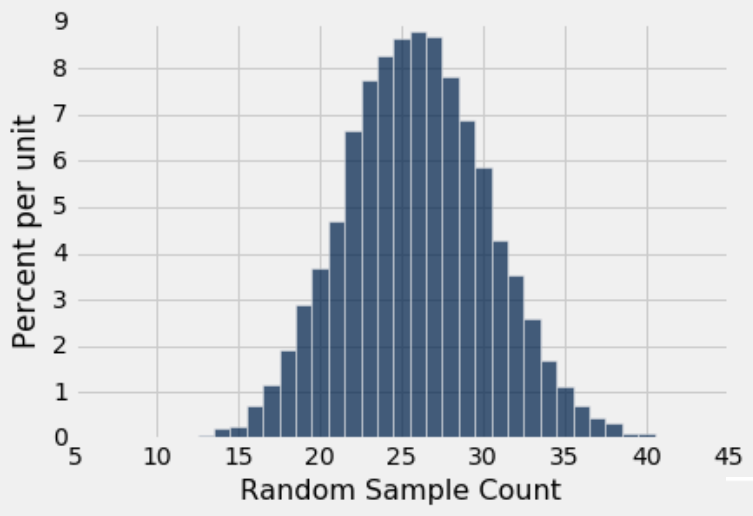
- Compare the **observed test statistic** and its empirical distribution under the null hypothesis
- If the observed value is **not consistent** with the distribution, then the test favors the alternative – “rejects the null hypothesis”

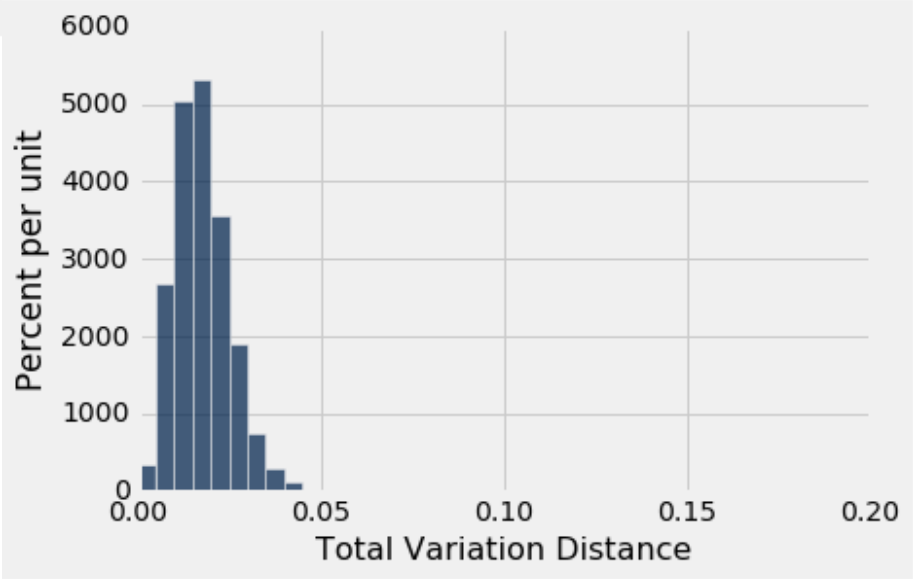
Whether a value is consistent with a distribution:

- A visualization may be sufficient
- If not, there are conventions about “consistency”

Statistical Significance

Tail Areas


Percent per unit
0
1
2
3
4
5
6
7
8
9
5
10
15
20
25
30
35
40
45
Random Sample Count


Percent per unit
0
1000
2000
3000
4000
5000
6000
0.00
0.05
0.10
0.15
0.20
Total Variation Distance

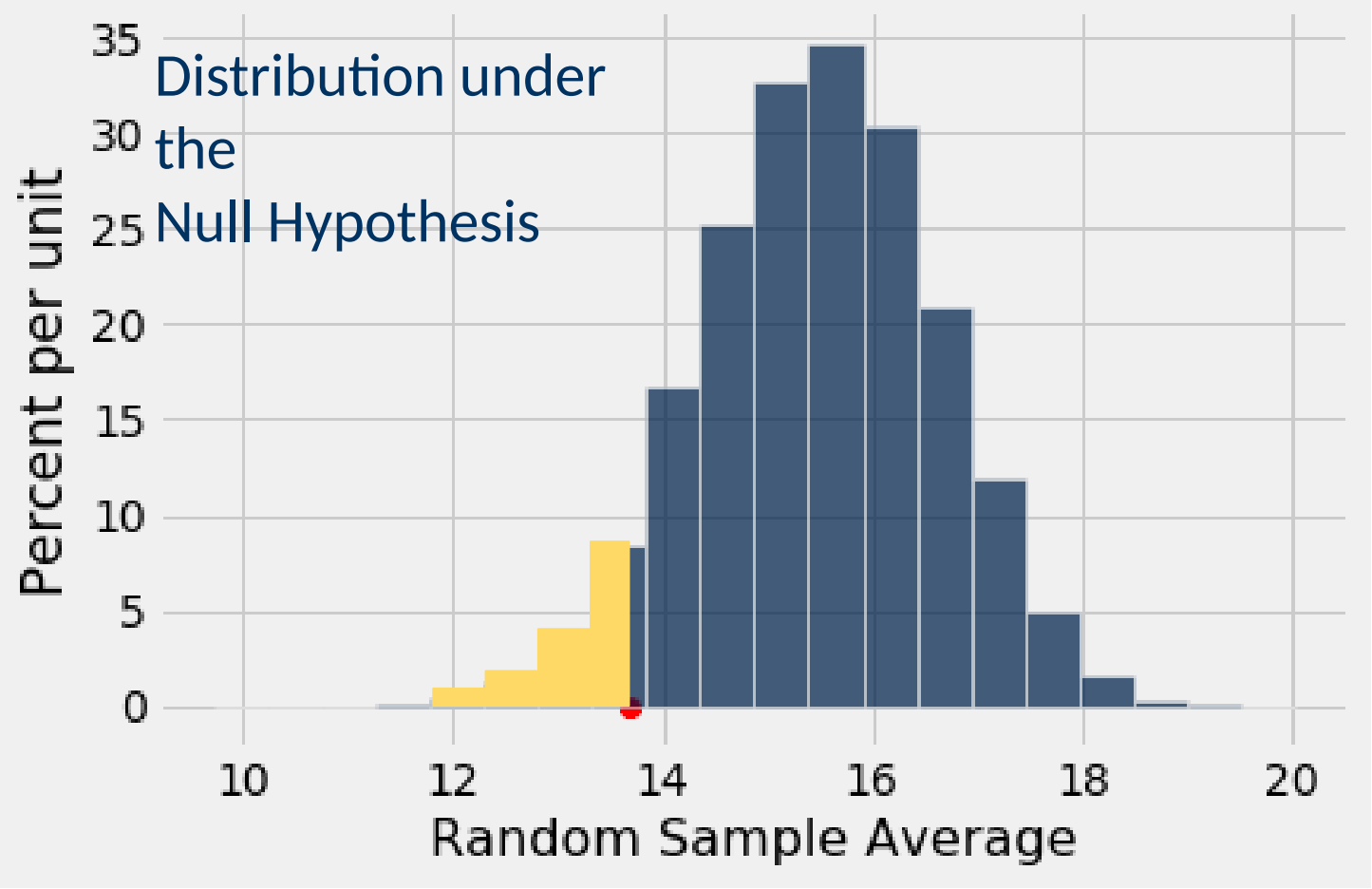
Conventions About Inconsistency

- **"Inconsistent":** 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"

(Demo)

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).



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.
- Last two bullets mean: “test statistic is <u>at least as extreme</u> as the observed value.”

Origin of the Conventions

Sir Ronald Fisher, 1890-1962

"We have the duty of formulating, of summarizing, and of communicating our conclusions, in intelligible form, in recognition of the right of other free minds to utilize them in making their own decisions."

Ronald Fisher

Sir Ronald Fisher, 1925

"It is convenient to take this point [5%] as a limit in judging whether a deviation is to be considered significant or not."

–– *Statistical Methods for Research Workers*

Sir Ronald Fisher, 1926

“If one in twenty does not seem high enough odds, we may, if we prefer it, draw the line at one in fifty (the 2 percent point), or one in a hundred (the 1 percent point). Personally, the author prefers to set a low standard of significance at the 5 percent point …”

To-do

- Lab 4
- Assignment 4