DATA 8
Summer 2018

# Lecture 16

## Decisions and Uncertainty

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

# Assessing Models

# Steps in Assessing a Model

- Come up with a statistic that will help you decide whether the data support the model or an alternative view of the world.
- Simulate the statistic under the assumptions of the model.
- Draw a histogram of the simulated values. This is the model's prediction for how the statistic should come out.
- Compute the observed statistic from the sample in the study.
- Compare this value with the histogram.
- If the two are not consistent, that's evidence against the model.

# Choosing Statistic

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

Questions before choosing the statistic:

- What values will will be evidence against my model and for my alternative viewpoint
  - The answer should be **either** large values or small values, but not **both**

# 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

- 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

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” or just “low”. 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”

# Performing a Test

# The Problem

- Large Statistics class divided into 12 discussion sections
- Graduate Student Instructors (GSIs) lead the sections

- After the midterm, students in Section 3 notice that the average score in their section is lower than in others

# The GSI's Defense

**GSI's position (Null Hypothesis):**

- If we had picked my section at random from the whole class, we could have got an average like this one.

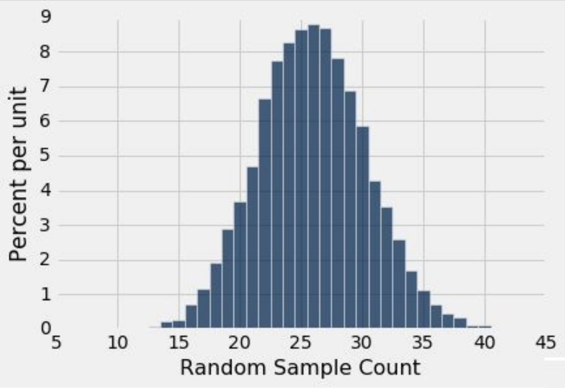
**Alternative:**

- No, the average score is too low. Randomness is not the only reason for the low scores.

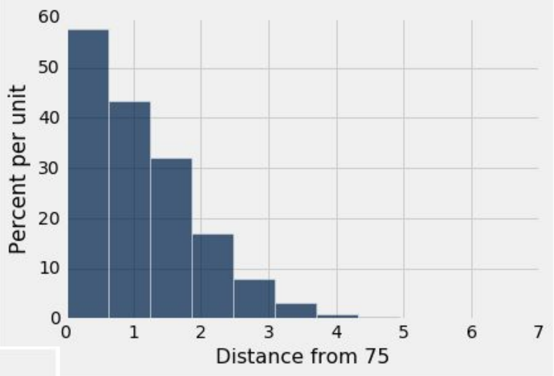
(Demo)

# Statistical Significance

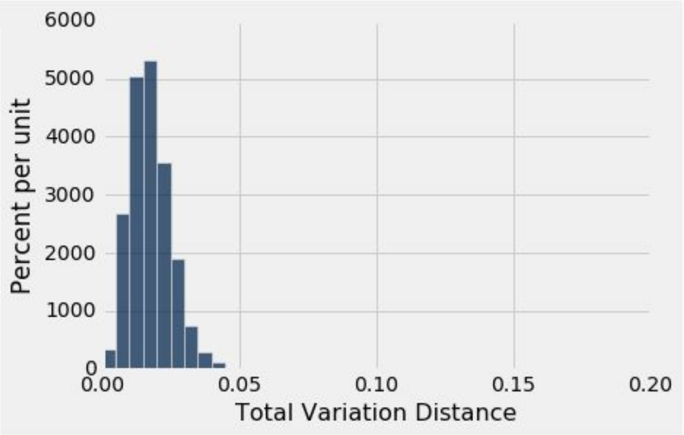
# Tail Areas



Percent per unit
Random Sample Count

Percent per unit
Distance from 75

Percent per unit
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)

# 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.

# An Error Probability

# Can the Conclusion be Wrong?

**Yes.**

| | Null is true | Alternative is true |
|---|---|---|
| **Test rejects the null** | ✗ | ✓ |
| **Test doesn't reject the null** | ✓ | ✗ |

# An Error Probability

- The cutoff for the *P*-value is an error probability.

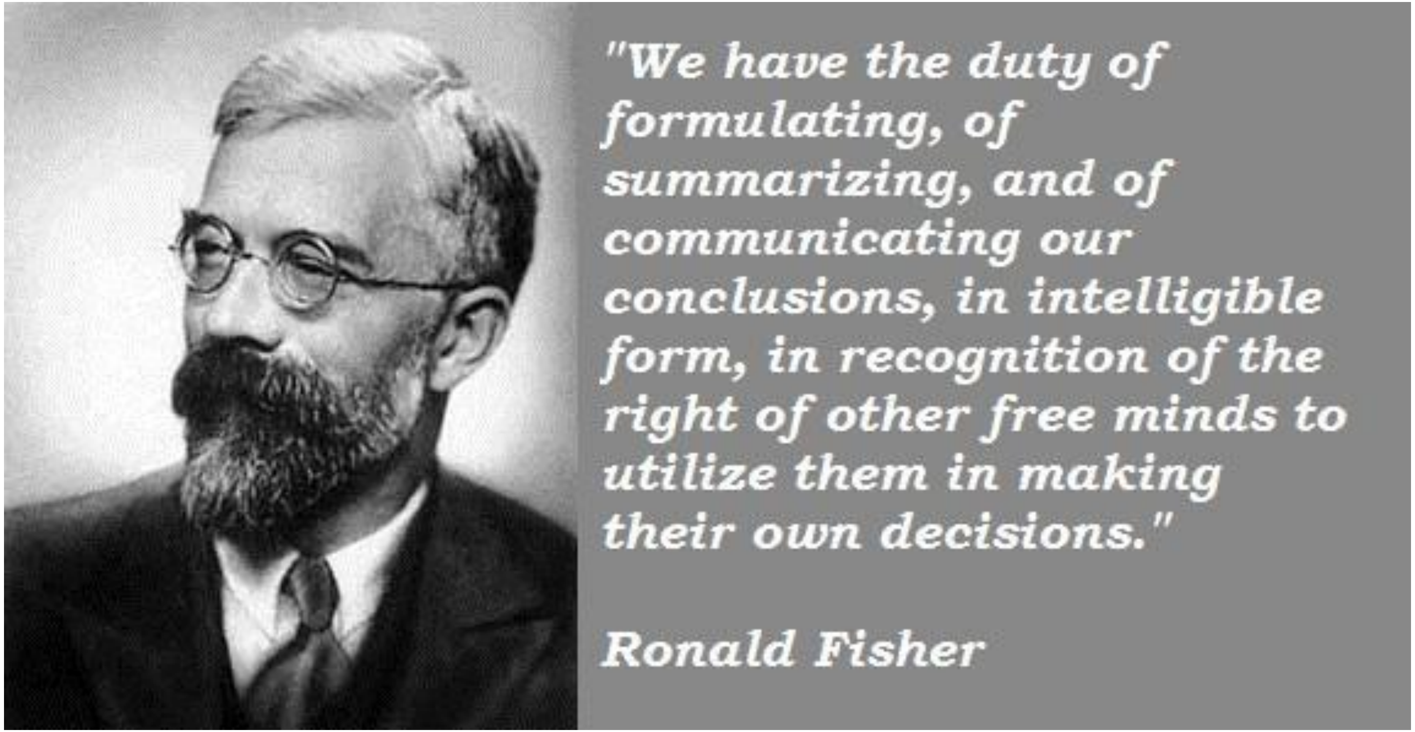
- If:
  - your **cutoff is 5%**
  - and the **null hypothesis happens to be true**

- then there is about a **5% chance** that **your test will reject the null hypothesis**.

(Demo)

# Origin of the Conventions

# Sir Ronald Fisher, 1890-1962



# 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 …”

# Revisiting Old Tests

# Swain vs Alabama

- **Null Hypothesis:** The number of black men comes randomly selected from a distribution with 26% black men and 74% other men. Any difference in our sample is due to chance
- **Alternative Hypothesis:** There is a bias against picking black men; the difference in our observed sample is not just due to chance
- **Test Statistic:** The number of black men
  - Small values of our test statistic point towards our alternative hypothesis

# Mendel's Pea Plants

- **Null Hypothesis:** There is a 75% chance of getting a purple-flowered plant. Any difference in our sample is due to chance.
- **Alternative Hypothesis:** There is not a 75% chance of getting a purple-flowered plant (The difference is systematic and not due to chance.)
- **Test Statistic:** Distance between 75 and the percentage of purple-flowered plants
    - Large values of our test statistic point towards our alternative hypothesis

# Ethnicities of Jury Panels

- **Null Hypothesis:** Our jury panel was picked from a distribution where there is a 15% chance of picking Asians, 18% chance of picking Black, 12% chance of picking Latino, 54% chance of picking White, and 1% chance of picking Other. Any difference is due to chance.
- **Alternative Hypothesis:** Our jury panel was not selected from the above distribution.
- **Test Statistic:** TVD between the distribution above and a sample distribution
  - Large values of our test statistic point towards our alternative hypothesis