# Data 8, Lab 6

## Testing Hypotheses

**Hubert Luo**
Fall 2019

11 October 2019

Berkeley
UNIVERSITY OF CALIFORNIA

# Agenda

1. Null/Alternative Hypotheses
2. Test Statistics
3. Multiple Categories
4. Simulations
5. P-Values
6. Decision-Making
7. Examples

# Assessing a Model

- Identifying a model (a set of assumptions)
  - Example: Jury panel selected at random from the population
- Choose a statistic to measure discrepancy between model and data
  - Example: Number of black men selected for jury panel
- Simulate statistic under the model
- Compare observed statistic with the distribution of simulated statistics

# Null Hypothesis

- Goal is to prove a model **does not** fit the observed data
- We call this model we are trying to disprove the **null hypothesis**
  - Should be clearly defined from the problem!
- The null hypothesis typically says some observed value was due to random chance alone
- Example: Jury panel selected at random; the observed number of black males on the panel was due to random chance alone
- By disproving the null hypothesis, we demonstrate that there is something other than random chance affecting what we observe

# Alternative Hypothesis

- An alternative model for the data
  - Based on what you observed: was it something weird or unexpected if the null hypothesis was true?
- Example: Jury panel was not selected at random
  - Both high and low number of Black jurors relative to the population would have supported this alternative hypothesis
- Example: Jury panel was biased against Black jurors
  - Only low number of Black jurors relative to the population would have supported this alternative hypothesis

Berkeley
UNIVERSITY OF CALIFORNIA

# Test Statistic

- Measures the discrepancy between a model and the data
- Depends on the Alternative Hypothesis:
- Example: (AH) Jury panel was not selected at random
  - | % eligible Black jurors in population - % Black jurors on panel |
- Example: (AH) Jury panel was biased against Black jurors
  - % Black jurors on panel
- Ask: What values of the statistic will support the alternative?
  - Ideally should be in just statistics in one direction

# Multiple Categories

- To assess models about multiple categories
- Example: We want to look at the ethnic composition of jury panels for all groups instead of just one
- New test statistic for multiple categories: **Total Variation Distance**
  - Find differences between theoretical and empirical proportions for each category
  - Take the absolute value of each difference
  - Sum these absolute differences and divide by 2

# TVD (cont'd)

| Ethnicity | Eligible | Panels | Difference | Absolute Difference |
|---|---|---|---|---|
| Asian | 0.15 | 0.26 | 0.11 | 0.11 |
| Black | 0.18 | 0.08 | -0.1 | 0.1 |
| Latino | 0.12 | 0.08 | -0.04 | 0.04 |
| White | 0.54 | 0.54 | 0 | 0 |
| Other | 0.01 | 0.04 | 0.03 | 0.03 |

Berkeley
UNIVERSITY OF CALIFORNIA

# Simulation: Sample Proportions

- The sample_proportions function takes in two arguments:
  - Sample size
  - **Theoretical probabilities** of each category
- Returns array consisting of the **empirical probabilities** of each category
  - Simulates drawing a sample of specified size with replacement
- **Common Mistake:** Don't forget the theoretical probabilities have to sum to 1, and the empirical probabilities also have to sum to 1

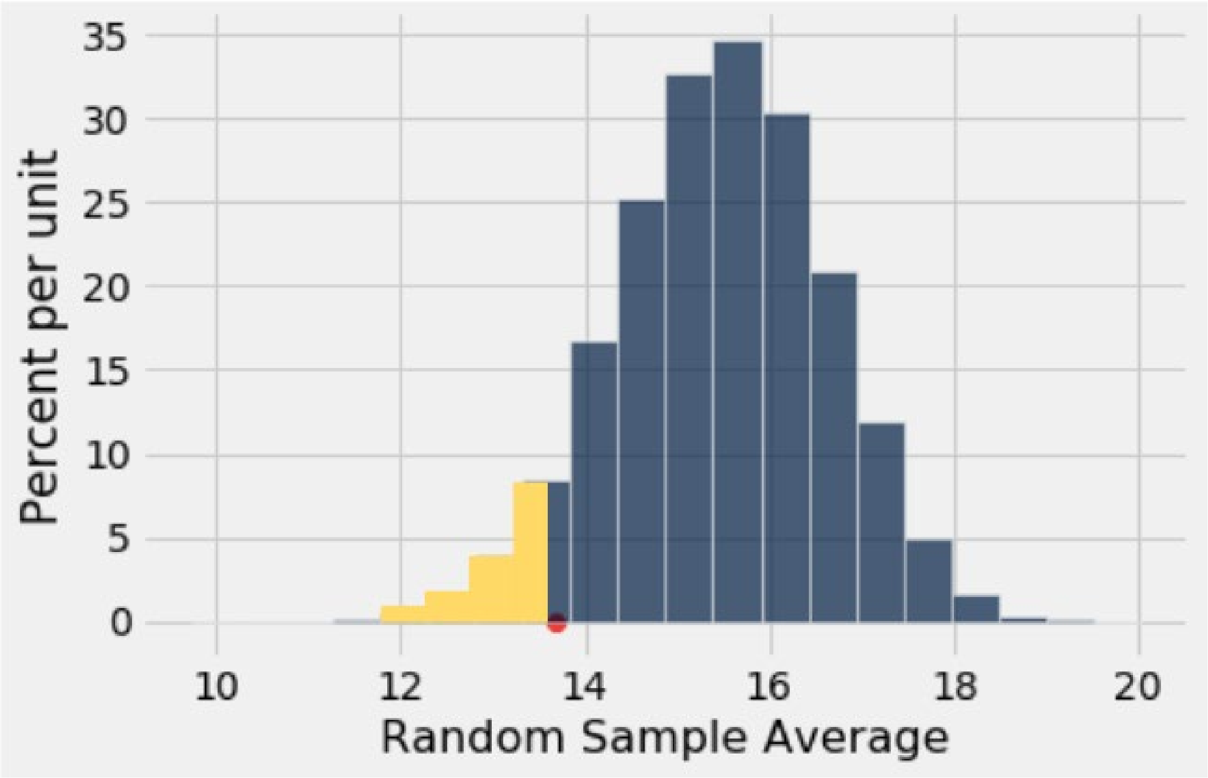
# Sample Proportions Examples

- Flip a fair coin 15 times
  - sample_proportions(15, make_array(0.5,0.5))
  - Hypothetically, suppose you get 10 heads and 5 tails: the output is an array [10/15, 5/15]
- Roll a fair die 30 times
  - sample_proportions(30, make_array(1/6,1/6,1/6,1/6,1/6,.1/6))
  - Hypothetically, suppose you get five 1's, seven 2's, four 3's, six 5's, and eight 6's: the output is an array [5/30, 7/30, 4/30, 6/30, 8/30]

# P-Value

- 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.
- Each one of these bullet points is necessary! Without it, this would not be the p-value so the full definition is important

# P-Value (cont'd)



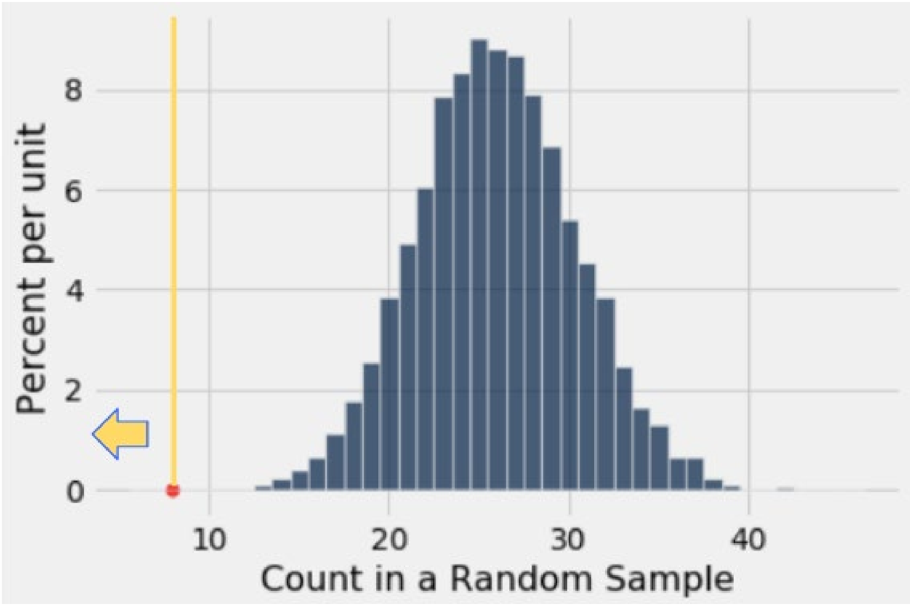

Berkeley
UNIVERSITY OF CALIFORNIA

# Making a Decision

- Compare the observed test statistic with the simulated test statistics to make a decision on whether or not to reject the null
- Traditionally people use a p-value cutoff of 5%
  – Any p-value less than or equal to 5% is treated as "statistically significant"
- If less than the cutoff, the null hypothesis is inconsistent with the data and we reject the null
  – Otherwise, we "fail to reject the null"

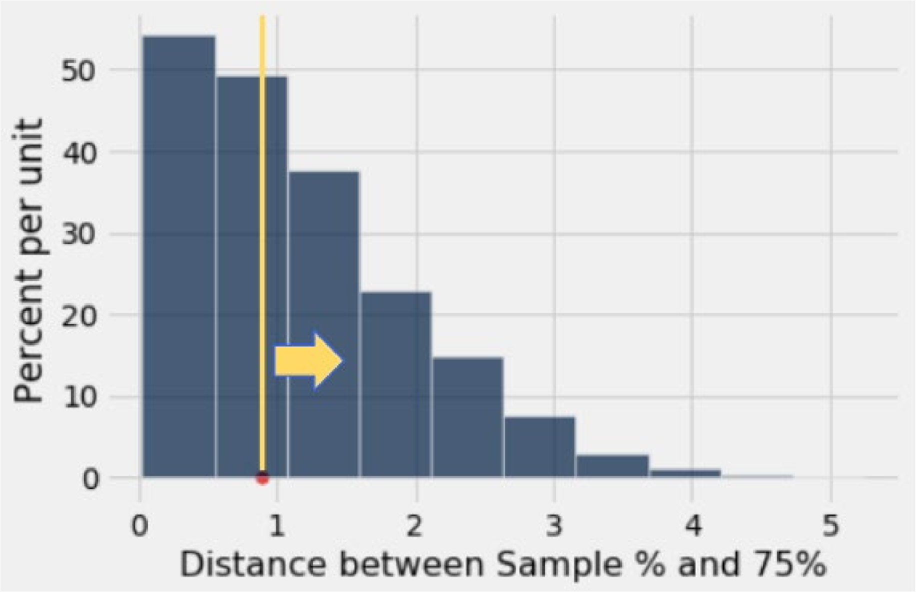
# Example: Swain v. Alabama

- Null: Jury panel drawn at random
- Alternative: Too few Black jurors
- Statistic: Number of black jurors on panel
- P-Value: Area at or to the left of the observed value (yellow line)
- Result: Test favours alternative



# Example: Mendel's Peas

- Null: Each pea plant has 75% chance of being purple flowering
- Alternative: Pea plant does not have 75% chance of being purple flowering
- Statistic: | % purple flowering peas in sample – 75 |
- P-Value: Area at or to the right of the observed value (yellow line)
- Result: Test favours null




Berkeley
UNIVERSITY OF CALIFORNIA

# Error Probabilities

- The p-value cutoff is also the probability of a Type I Error
- It is the probability of rejecting the null even when the null is true
- Example: Using a p-value cutoff of 5%, there is a 5% chance of rejecting the null even when the null is true

Berkeley
UNIVERSITY OF CALIFORNIA

# Announcements

- Next week is midterm review! Please go to a lab on Wednesday or Thursday to work on the problems
  - Lab next Friday will just be Office Hours
- Refer to your email for list of Resources!