DATA 8

Summer 2018

Lecture 17

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

Review

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.

Testing Hypotheses

- Select a Null and Alternative Hypothesis
  - The null is a fully specified chance model we can simulate under
  - The alternative is some other viewpoint of the world
- Choose a Test Statistic
  - This test statistic should help us determine between our two viewpoints
  - **Either** large values of the test statistic or small values should be evidence for our alternative
- Simulate the test statistic under the null hypothesis to create an empirical distribution
  - Approximates the probability distribution of the statistic under the null
- Calculate our observed test statistic
- Compare our observed test statistic with values our null predicted

Definition of the *P*-value

Formal name: **observed significance level**

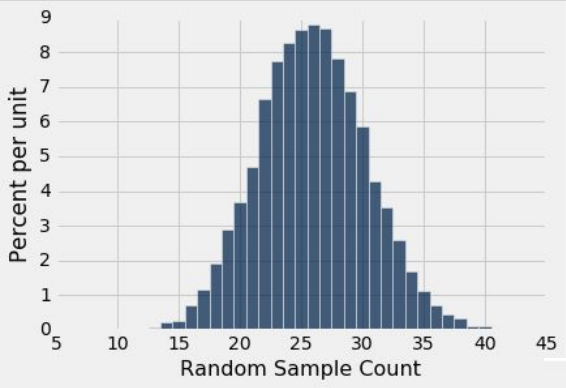
The *P*-value is the chance,

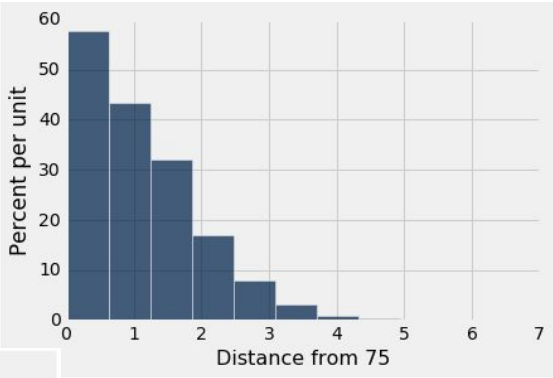
- under the null hypothesis,
- that the test statistic
- is equal to the value that was observed in the data
- or is even further in the direction of the alternative.

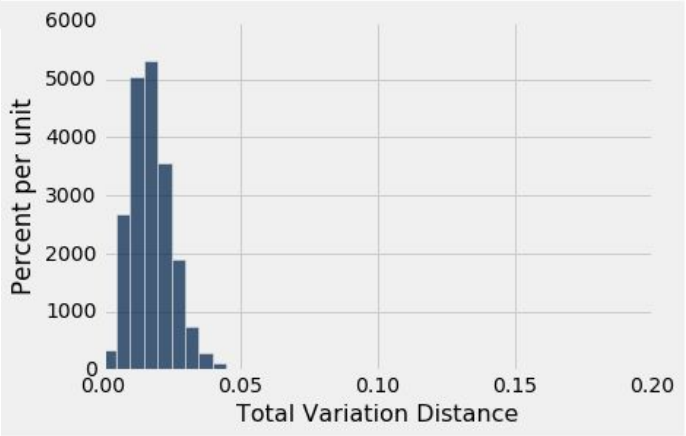
Connection with our simulation

- We simulated the test statistic under the assumption of the null hypothesis and created an empirical distribution
  - This is used an an *approximation* to the probability distribution of the test statistic under the null
- We now know what the chance of seeing different test statistics are, if our null hypothesis is True
- Use the above distribution to calculate our P-Value

Tail Areas


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


Percent per unit
Distance from 75
0
10
20
30
40
50
60
0
1
2
3
4
5
6
7


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

Using the *P*-value

- If the P-value is small, that is evidence against the null hypothesis

- Conventions about “small”:
  - Less than 5% (result is called statistically significant)
  - Less than 1% (result is called highly statistically significant)

Discussion Questions

Suppose the P-value of a test comes out to be about 0.5%.

(a) Fill in the blanks: The test supports the _____ hypothesis more than it supports the _____ hypothesis.

(a) True or false: There is about a 0.5% chance that the null hypothesis is true.

Error Probabilities

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**.

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

(Demo)

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

(Demo)

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

(Demo)

When to use TVD

- Use TVD to compare **two distributions**
  - Only use when you have categorical variables
  - Comparing two sets of proportions, so each distribution should add up to 1
- Examples
  - Compare the observed proportions of ethnicities in a jury panel to the expected proportions
  - Compare the observed proportions of values of 100 die rolls to the expected proportions
  - In both the above examples, we want to compare multiple proportions (proportion of 1s, proportion of 2s, proportion of 3s, etc)
  - TVD combines all of these comparisons into one number

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 were lighter, on average, than the babies of the non-smokers.

Test Statistic

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

- Statistic: Difference between average weights

  Group A average - Group B average

- Small values of this statistic favor the alternative

Simulating Under the Null

- If the null is true, all rearrangements of the birth weights among the two groups are equally likely
- Plan:
  - Shuffle all the birth weights
  - Assign some to “Group A” and the rest to “Group B”, maintaining the two sample sizes
  - Find the difference between the averages of the two shuffled groups
  - Repeat

(Demo)