



Lecture 19

Midterm Review

Contributions by Vinitra Swamy (vinitra@berkeley.edu) and Fahad Kamran (fndkmrn@berkeley.edu)
Slides created by John DeNero (denero@berkeley.edu) and Ani Adhikari (adhikari@berkeley.edu)

Announcements

Deflategate

2015 AFC Championship Game



Deflategate

Wikipedia:

The 2015 AFC Championship Game football tampering scandal, commonly referred to as Deflategate, or Ballghazi

...

'Deflategate' returns, focus on Tom Brady's destroyed cellphone

POSTED 9:54 AM, MARCH 5, 2016, BY [CNN WIRE](#), UPDATED AT 10:33AM, MARCH 5, 2016

(Demo)

Null hypothesis

The 4 Colts footballs are like a sample drawn at random without replacement from all 15 balls.

- To test this hypothesis, repeat this process:
 - Randomly permute all 15 balls
 - Label 11 of them “Patriots” and the remaining 4 “Colts”
 - Compare the averages of the two groups

(Demo)

Histograms

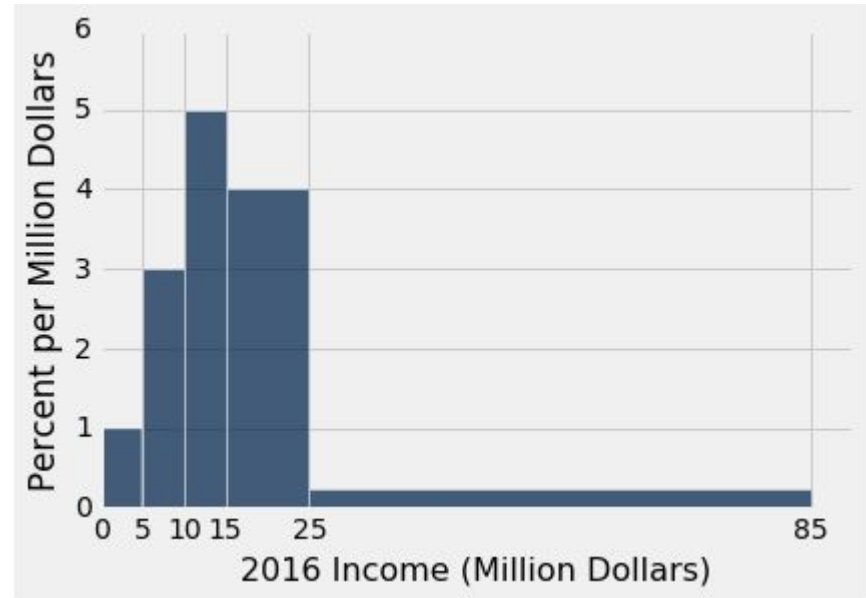
Using the Density Scale

(a) Which bin has more people: $[10, 15)$ or $[15, 25)$?

(b) What percent of incomes are in the $[25, 85)$ bin?

(c) If you draw one bar over $[10, 25)$, how tall will it be?

(d) Find (or give bounds for) the median income.



Answers

(a) $[15, 25)$

(b) 15%

(c) 4.33 percent per million dollars

(d) At least 15 and less than 25

Probability

Exercise 1

I pick one of the 12 months at random. Independently, you pick one of the 12 months at random.

What is the chance that we both pick the same month?

(i) $(1/12) * (1/12)$ (ii) $(1/12) + (1/12)$ (iii) $1/12$

(iii) = $(12/12) * (1/12)$ Also (iii) = $12 * (i)$

Exercise 2

Marbles: G, G, G, G, R, R, R, B, B, Y. Draw 4 at random.

$$P(\text{no G}) = ?$$

$$P(\text{all G}) = ?$$

If with replacement:

$$(6/10)^*(6/10)^*(6/10)^*(6/10)$$

If with replacement:

$$(4/10)^*(4/10)^*(4/10)^*(4/10)$$

If without replacement:

$$(6/10)^*(5/9)^*(4/8)^*(3/7)$$

If without replacement:

$$(4/10)^*(3/9)^*(2/8)^*(1/7)$$

Exercise 3

Marbles: G, G, G, G, R, R, R, B, B, Y. Draw 4 times at random with replacement.

$1 - (6/10)^4$ is the chance of:
at least one G

$(4/10)^4 + (3/10)^4 + (2/10)^4 + (1/10)^4$ is the chance of:
all four are the same color

Testing Hypotheses

Before You Compute Anything

- Figure out the viewpoint the question wants to test, and formulate:
 - Null hypothesis: Completely specified chance model under which you can simulate data
 - Alternative hypothesis: Viewpoint comes from the question
 - Test statistic: to help you choose one viewpoint
 - Say what kind of values of the statistic will make you lean towards each alternative
-

Categorical Data

Null Hypothesis

The sample is drawn at random from a specified categorical distribution.

- Swain's jury panel was drawn at random from a population that had 26% black men
 - Each pea plant has 75% chance of being purple flowering, regardless of other plants
 - The Alameda County jury panels were drawn at random from the specified distribution of eligible jurors
-

Swain v. Alabama

- **Null:** Swain's jury panel was drawn at random from a population that had 26% black men
 - **Alternative:** There were too few black men on the panel for it to look like a random sample
 - **Test statistic:**
Number of black men in panel
P-value direction: to the left
-

Mendel's Model

- **Null:** Each pea plant has 75% chance of being purple flowering, regardless of other plants
- **Alternative:** The model isn't good.
- **Test statistic:**

| percent purple in sample - 75 |

P-value direction: to the right

Could also have used TVD; direction is still to the right

$$\text{TVD} = (|\text{prop. purple} - 0.75| + |\text{prop. white} - 0.25|)/2$$

Alameda County Jury Panels

- **Null:** The Alameda County jury panels were drawn at random from the specified distribution of eligible jurors
 - **Alternative:** The panels were not drawn at random from the specified distribution.
 - **Test statistic:**
TVD
 P -value direction: to the right
-

Numerical Data

GSI's Defense

- **Null:** Section 3 scores are like a sample drawn at random without replacement from the whole class.
 - **Alternative:** The Section 3 average is too low for the section to be a random sample from the class.
 - **Test statistic:**
Section 3 average
 P -value direction: to the left
-

Comparing Two Samples

Birthweights

- **Null:** In the population, the distributions of the birth weights of the babies in the two groups are the same.
 - **Alternative:** In the population, the babies of the mothers who didn't smoke (B) were heavier, on average, than the babies of the smokers (A).
 - **Test statistic:**
Group B sample average - Group A sample average
P-value direction: to the right
-

Deflategate

- **Null:** Each group is like a sample drawn at random without replacement from all 15 footballs.
 - **Alternative:** The Colts' values are too low for them to look like a random sample from the 15 balls.
 - **Test statistic:**
Colts' average - Patriots' average
P-value direction: to the left
-

RCT

- **Null:** The distribution of all the potential control scores is the same as the distribution of all the potential treatment scores.
 - **Alternative:** The distribution of all the potential control scores is different from the distribution of all the potential treatment scores.
 - **Test statistic:**
| control group average - treatment group average |
P-value direction: to the right
-