Class 16: Inference and simulation II

June 12, 2018

General

Annoucements

- Questions for Reading 13 due on **June 13** by 9:00am: Introductory Statistics with Randomization and Simulation

    - From chapter 2: section 2.4 through to the end of section 2.5

    - From chapter 4: section 4.5 (skip 4.5.3)

- Homework 3 due by **11:59pm on Wednesday, June 13th**

- Midterm reports and presentation slides due at the start of class on Thursday, June 14th

Using R to analyze the gender discrimination study

Simulations in R

Introduce a new package: `infer`

Simulations in R

Introduce a new package: `infer`

Download by running:

```
install.packages("infer")
```

Simulations in R

Introduce a new package: `infer`

Download by running:

```
install.packages("infer")
```

Use gender discrimination dataset:

```
applicants_data <- data_frame(
  sex = combine(
    rep("Male", 24),
    rep("Female", 24)
  ),
  outcome = combine(
    rep("Promoted", 21),
    rep("Not Promoted", 3),
    rep("Promoted", 14),
    rep("Not Promoted", 10)
  )
)

experiment_result <- (21/24) - (14/24)
```

Recap of gender discrimination dataset

- Experiment involving 48 male bank supervisors that were each given the same personnel file and asked to judge whether the person should be promoted to a branch manager job that was described as "routine"

- The files were identical except that half of the supervisors had files showing the person was male while the other half had files showing the person was female

- It was randomly determined which supervisors got "male" applications and which got "female" applications

- **Result: 29.2% more men than women were recommended for promotion**

Is this result statistically significant?

Using `infer` to build the null distribution

- To test the alternative hypothesis that "women were less likely than men to be hired", we need to generate a null distribution using the `infer` package:

- To test the alternative hypothesis that "women were less likely than men to be hired", we need to generate a null distribution using the `infer` package:

```
simulation_results <- applicants_data %>%
  specify(outcome ~ sex, success = "Promoted") %>%
  hypothesize(null = "independence") %>%
  generate(reps = 10000, type = "permute") %>%
  calculate(stat = "diff in props", order = c("Male", "Female"))
```

Using `infer` to build the null distribution

- To test the alternative hypothesis that "women were less likely than men to be hired", we need to generate a null distribution using the `infer` package:

```
simulation_results <- applicants_data %>%
  specify(outcome ~ sex, success = "Promoted") %>%
  hypothesize(null = "independence") %>%
  generate(reps = 10000, type = "permute") %>%
  calculate(stat = "diff in props", order = c("Male", "Female"))
```

Using `infer` to build the null distribution

- To test the alternative hypothesis that "women were less likely than men to be hired", we need to generate a null distribution using the `infer` package:

```
simulation_results <- applicants_data %>%
  specify(outcome ~ sex, success = "Promoted") %>%
  hypothesize(null = "independence") %>%
  generate(reps = 10000, type = "permute") %>%
  calculate(stat = "diff in props", order = c("Male", "Female"))
```

- In `specify(outcome ~ sex, success = "Promoted")`, the first part `outcome ~ sex` is a formula where the lefthand variable `outcome` is the response and the righthand variable `sex` is explanatory.

Using `infer` to build the null distribution

- To test the alternative hypothesis that "women were less likely than men to be hired", we need to generate a null distribution using the `infer` package:

```
simulation_results <- applicants_data %>%
  specify(outcome ~ sex, success = "Promoted") %>%
  hypothesize(null = "independence") %>%
  generate(reps = 10000, type = "permute") %>%
  calculate(stat = "diff in props", order = c("Male", "Female"))
```

Using `infer` to build the null distribution

- To test the alternative hypothesis that "women were less likely than men to be hired", we need to generate a null distribution using the `infer` package:

```
simulation_results <- applicants_data %>%
  specify(outcome ~ sex, success = "Promoted") %>%
  hypothesize(null = "independence") %>%
  generate(reps = 10000, type = "permute") %>%
  calculate(stat = "diff in props", order = c("Male", "Female"))
```

- In `hypothesize(null = "independence")`, we specify that we will simulate what will happen if `outcome` and `sex` were independent.

- To test the alternative hypothesis that "women were less likely than men to be hired", we need to generate a null distribution using the `infer` package:

```
simulation_results <- applicants_data %>%
  specify(outcome ~ sex, success = "Promoted") %>%
  hypothesize(null = "independence") %>%
  generate(reps = 10000, type = "permute") %>%
  calculate(stat = "diff in props", order = c("Male", "Female"))
```

Using `infer` to build the null distribution

- To test the alternative hypothesis that "women were less likely than men to be hired", we need to generate a null distribution using the `infer` package:

```
simulation_results <- applicants_data %>%
  specify(outcome ~ sex, success = "Promoted") %>%
  hypothesize(null = "independence") %>%
  generate(reps = 10000, type = "permute") %>%
  calculate(stat = "diff in props", order = c("Male", "Female"))
```

- In `generate(reps = 10000, type = "permute")`, we specify that we will run 10,000 simulations by permuting the `outcome` and `sex` columns

Using `infer` to build the null distribution

- To test the alternative hypothesis that "women were less likely than men to be hired", we need to generate a null distribution using the `infer` package:

```
simulation_results <- applicants_data %>%
  specify(outcome ~ sex, success = "Promoted") %>%
  hypothesize(null = "independence") %>%
  generate(reps = 10000, type = "permute") %>%
  calculate(stat = "diff in props", order = c("Male", "Female"))
```

- In `generate(reps = 10000, type = "permute")`, we specify that we will run 10,000 simulations by permuting the `outcome` and `sex` columns

- To permute, we randomly shuffle the data in the `outcome` column, and then randomly shuffle the data in the `sex` column

- To test the alternative hypothesis that "women were less likely than men to be hired", we need to generate a null distribution using the `infer` package:

```
simulation_results <- applicants_data %>%
  specify(outcome ~ sex, success = "Promoted") %>%
  hypothesize(null = "independence") %>%
  generate(reps = 10000, type = "permute") %>%
  calculate(stat = "diff in props", order = c("Male", "Female"))
```

Using `infer` to build the null distribution

- To test the alternative hypothesis that "women were less likely than men to be hired", we need to generate a null distribution using the `infer` package:

```
simulation_results <- applicants_data %>%
  specify(outcome ~ sex, success = "Promoted") %>%
  hypothesize(null = "independence") %>%
  generate(reps = 10000, type = "permute") %>%
  calculate(stat = "diff in props", order = c("Male", "Female"))
```

- Using `calculate(stat = "diff in props", order = c("Male", "Female"))` means, after each simulation, compute:

$$\frac{\text{Promoted Men}}{\text{Total Men}} - \frac{\text{Promoted Men}}{\text{Total Men}}$$

Using `infer` to build the null distribution

- To test the alternative hypothesis that "women were less likely than men to be hired", we need to generate a null distribution using the `infer` package:

```
simulation_results <- applicants_data %>%
  specify(outcome ~ sex, success = "Promoted") %>%
  hypothesize(null = "independence") %>%
  generate(reps = 10000, type = "permute") %>%
  calculate(stat = "diff in props", order = c("Male", "Female"))
```

- Using `calculate(stat = "diff in props", order = c("Male", "Female"))` means, after each simulation, compute:

$$\frac{\text{Promoted Men}}{\text{Total Men}} - \frac{\text{Promoted Men}}{\text{Total Men}}$$

Note that this is exactly how `experiment_result` was calculated.

Visualizing the null distribution

```
simulation_results %>%
  ggplot() +
  geom_histogram(
    mapping = aes(x = stat, y = ..density..),
    center = 0,
    bins = 9
  ) +
  geom_vline(xintercept = experiment_result, color = "indianred3") +
  labs(
    title = "Gender discrimination null distribution",
    x = "difference in fraction of male and female promotions",
    y = "PMF"
  )
```

Visualizing the null distribution

Gender discrimination null distribution
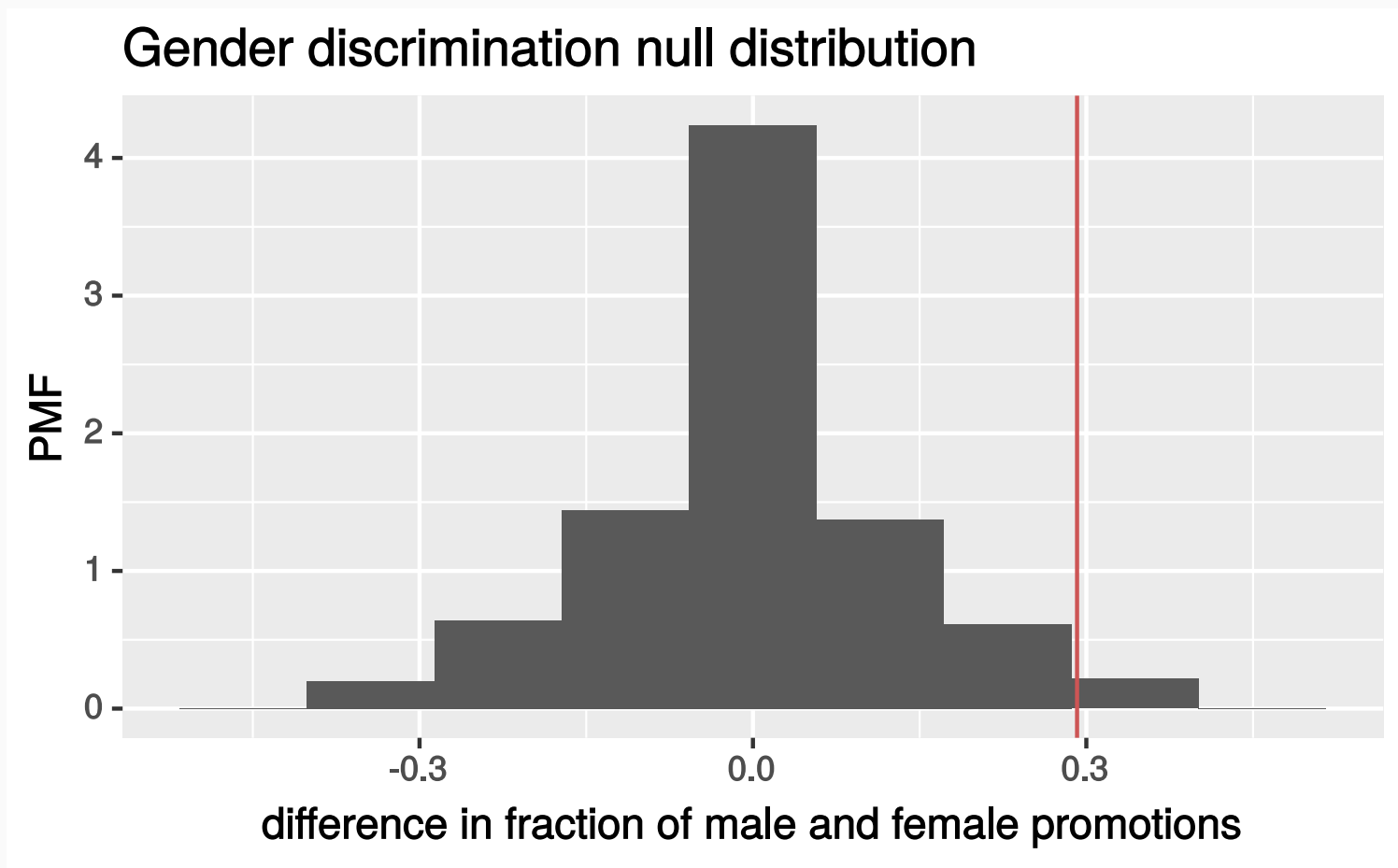
Probability of randomly getting result

One way we can quantify the probability that the experiment's result is due to statistical variance is to compute how much of the null distribution represents outcomes that are the same or more extreme than the actual experiment:

```
simulation_results %>%
  filter(stat >= 0.875 - 0.583) %>%
  summarize(random_result_probability = n() / 10000)
```

random_result_probability
0.0047

Conclusions from our simulation

Do the results of the simulation provide convincing evidence of gender discrimination against women, i.e. dependence between gender and promotion decisions?

1. No, the data do not provide convincing evidence for the alternative hypothesis, therefore we can't reject the null hypothesis of independence between gender and promotion decisions. The observed difference between the two proportions was due to chance.

2. Yes, the data provide convincing evidence for us to reject the null hypothesis in favor of the alternative hypothesis of gender discrimination against women in promotion decisions. The observed difference between the two proportions was due to a real effect of gender.

Conclusions from our simulation

Do the results of the simulation provide convincing evidence of gender discrimination against women, i.e. dependence between gender and promotion decisions?

1. No, the data do not provide convincing evidence for the alternative hypothesis, therefore we can't reject the null hypothesis of independence between gender and promotion decisions. The observed difference between the two proportions was due to chance.

2. Yes, the data provide convincing evidence for us to reject the null hypothesis in favor of the alternative hypothesis of gender discrimination against women in promotion decisions. The observed difference between the two proportions was due to a real effect of gender.

Constructing hypothesis tests

Number of college applications

A survey asked how many colleges students applied to, and 206 students responded to this question. This sample yielded an average of 9.7 college applications with a standard deviation of 7. College Board website states that counselors recommend students apply to roughly 8 colleges. Do these data provide convincing evidence that the average number of colleges all GMU students apply to is *higher* than recommended?

http://www.collegeboard.com/student/apply/the-application/151680.html

Setting the hypotheses

- The **parameter of interest** is the average number of schools applied to by *all* GMU students.

Setting the hypotheses

- The **parameter of interest** is the average number of schools applied to by *all* GMU students.

- There may be two explanations why our sample mean is higher than the recommended 8 schools.

    - The true population mean is different
    - The true population mean is 8, and the difference between the true population mean and the sample mean is simply due to natural sampling variability

Setting the hypotheses

- The **parameter of interest** is the average number of schools applied to by *all* GMU students.

- There may be two explanations why our sample mean is higher than the recommended 8 schools.

    - The true population mean is different
    - The true population mean is 8, and the difference between the true population mean and the sample mean is simply due to natural sampling variability

- We start with the assumption the average number of colleges GMU students apply to is 8 (as recommended)

$$H_0 : \mu = 8$$

Setting the hypotheses

- The **parameter of interest** is the average number of schools applied to by *all* GMU students.

- There may be two explanations why our sample mean is higher than the recommended 8 schools.

    - The true population mean is different
    - The true population mean is 8, and the difference between the true population mean and the sample mean is simply due to natural sampling variability

- We start with the assumption the average number of colleges GMU students apply to is 8 (as recommended)

$$H_0 : \mu = 8$$

- We test the claim that the average number of colleges GMU students apply to is greater than 8

$$H_A : \mu > 8$$

Formal testing using p-values

Statistical significance

Say that we conducted this study by polling an independent and representative sample of GMU students about how many colleges they applied to, and obtained a sample mean of 9.7.

Statistical significance

Say that we conducted this study by polling an independent and representative sample of GMU students about how many colleges they applied to, and obtained a sample mean of 9.7.

The national average is 8.

Statistical significance

Say that we conducted this study by polling an independent and representative sample of GMU students about how many colleges they applied to, and obtained a sample mean of 9.7.

The national average is 8.

Is this result statistically significant?

Statistical significance

Say that we conducted this study by polling an independent and representative sample of GMU students about how many colleges they applied to, and obtained a sample mean of 9.7.

The national average is 8.

Is this result statistically significant?

In order to evaluate if the observed sample mean is unusual for the hypothesized sampling distribution, we do the following:

- Choose a value for the significance level α (a common choice is 5%)

Statistical significance

Say that we conducted this study by polling an independent and representative sample of GMU students about how many colleges they applied to, and obtained a sample mean of 9.7.

The national average is 8.

Is this result statistically significant?

In order to evaluate if the observed sample mean is unusual for the hypothesized sampling distribution, we do the following:

- Choose a value for the significance level α (a common choice is 5%)

- Determine the percentile rank of the observed sample mean relative to the null distribution

p-values

- We then use the percentile to calculate the **p-value**, the probability of observing data at least as favorable to the alternative hypothesis as our current data set, if the null hypothesis were true.

p-values

- We then use the percentile to calculate the **p-value**, the probability of observing data at least as favorable to the alternative hypothesis as our current data set, if the null hypothesis were true.

- If the p-value is **lower** than the significance level α, we say that it would be very unlikely to observe the data if the null hypothesis were true, and hence **reject H_0**.

p-values

- We then use the percentile to calculate the **p-value**, the probability of observing data at least as favorable to the alternative hypothesis as our current data set, if the null hypothesis were true.

- If the p-value is **lower** than the significance level α, we say that it would be very unlikely to observe the data if the null hypothesis were true, and hence **reject H_0**.

- If the p-value is **higher** than α, we say that it is likely to observe the data even if the null hypothesis were true, and hence **do not reject H_0**.

p-value: probability of observing data at least as favorable to H_A as our current data set (a sample mean greater than 9.7), if in fact H_0 were true (the true population mean was 8).
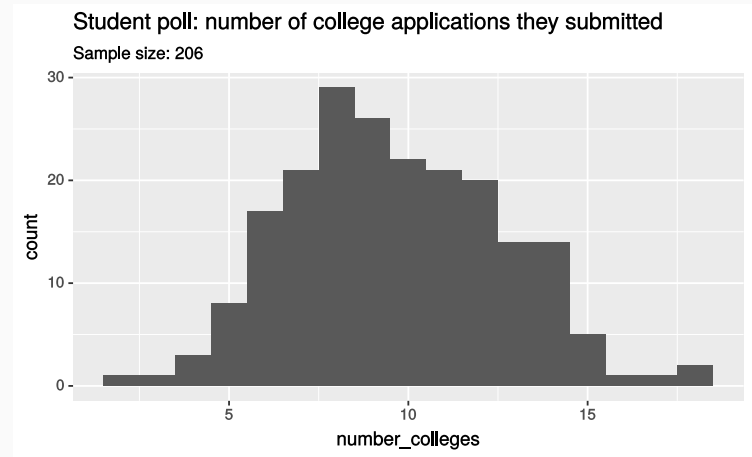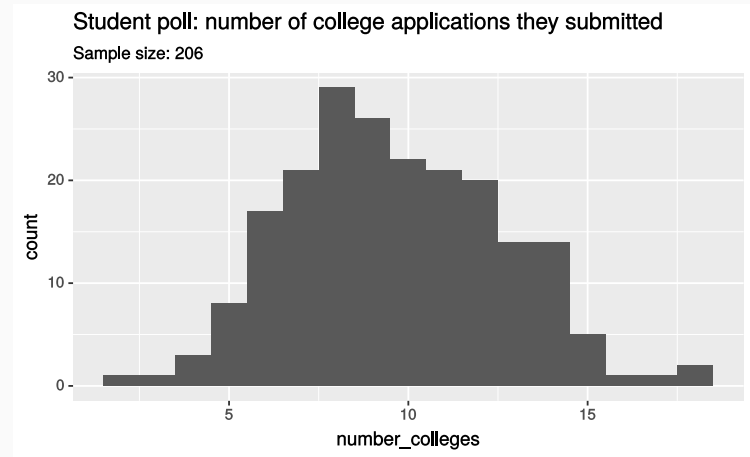
Number of college applications p-value

p-value: probability of observing data at least as favorable to H_A as our current data set (a sample mean greater than 9.7), if in fact H_0 were true (the true population mean was 8).

Student poll: number of college applications they submitted

Sample size: 206

Number of college applications p-value

p-value: probability of observing data at least as favorable to H_A as our current data set (a sample mean greater than 9.7), if in fact H_0 were true (the true population mean was 8).

Student poll: number of college applications they submitted
Sample size: 206

```r
college_apps_null <- college_apps %>%
  specify(formula = number_colleges ~ NULL) %>%
  hypothesize(null = "point", mu = 8) %>%
  generate(reps = 10000, type = "bootstrap") %>%
  calculate(stat = "mean")
```

```r
college_apps_p_value <- college_apps_null %>%
  filter(stat >= 9.7) %>%
  summarize(p_value = n() / 10000)
```

p-value: probability of observing data at least as favorable to H_A as our current data set (a sample mean greater than 9.7), if in fact H_0 were true (the true population mean was 8).
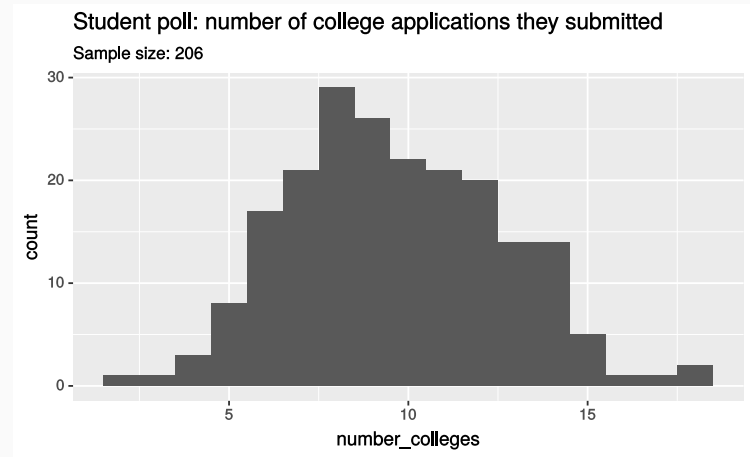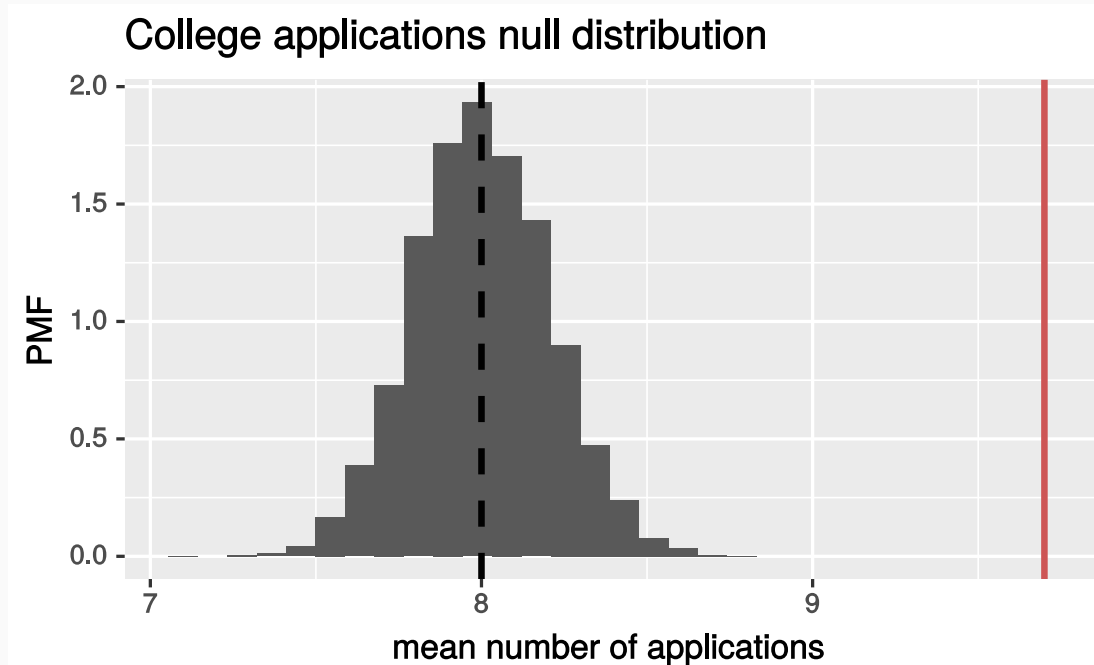

Student poll: number of college applications they submitted
Sample size: 206

```r
college_apps_null <- college_apps %>%
  specify(formula = number_colleges ~ NULL) %>%
  hypothesize(null = "point", mu = 8) %>%
  generate(reps = 10000, type = "bootstrap") %>%
  calculate(stat = "mean")
```

```r
college_apps_p_value <- college_apps_null %>%
  filter(stat >= 9.7) %>%
  summarize(p_value = n() / 10000)
```

p-value = 0

Number of applications p-value

```
ggplot(data = college_apps_null) +
  geom_histogram(mapping = aes(x = stat, y = ..density..)) +
  geom_vline(xintercept = 8, linetype = "dashed", size = 1) +
  geom_vline(xintercept = 9.7, color = "indianred3", size = 1) +
  labs(
    x = "mean number of applications",
    y = "PMF",
    title = "College applications null distribution"
  )
```


College applications null distribution

- p-value = 0

Number of college applications - Making a decision

- p-value = 0

- If the true average of the number of colleges GMU students applied to is 8, there is a 0% chance of observing a random sample of 206 GMU students who on average apply to 9.7 or more schools.

Number of college applications - Making a decision

- p-value = 0

- If the true average of the number of colleges GMU students applied to is 8, there is a 0% chance of observing a random sample of 206 GMU students who on average apply to 9.7 or more schools.

- This is a pretty low probability for us to think that a sample mean of 9.7 or more schools is likely to happen simply by chance.

Number of college applications - Making a decision

- p-value = 0

- If the true average of the number of colleges GMU students applied to is 8, there is a 0% chance of observing a random sample of 206 GMU students who on average apply to 9.7 or more schools.

- This is a pretty low probability for us to think that a sample mean of 9.7 or more schools is likely to happen simply by chance.

- Since p-value is **low** (lower than 5%) we **reject H_0.**

Number of college applications - Making a decision

- p-value = 0

- If the true average of the number of colleges GMU students applied to is 8, there is a 0% chance of observing a random sample of 206 GMU students who on average apply to 9.7 or more schools.

- This is a pretty low probability for us to think that a sample mean of 9.7 or more schools is likely to happen simply by chance.

- Since p-value is **low** (lower than 5%) we **reject H_0**.

- The data provide convincing evidence that GMU students apply to more than 8 schools on average.

Number of college applications - Making a decision

- p-value = 0

- If the true average of the number of colleges GMU students applied to is 8, there is a 0% chance of observing a random sample of 206 GMU students who on average apply to 9.7 or more schools.

- This is a pretty low probability for us to think that a sample mean of 9.7 or more schools is likely to happen simply by chance.

- Since p-value is **low** (lower than 5%) we **reject H_0**.

- The data provide convincing evidence that GMU students apply to more than 8 schools on average.

- The difference between the null value of 8 schools and observed sample mean of 9.7 schools is **not due to chance** or sampling variability.

Two-sided hypothesis testing with p-values

Two-sided hypothesis testing with p-values

- If the research question was "Do the data provide convincing evidence that the average amount of schools that GMU students apply to is **different** than the national average?", the alternative hypothesis would be different.

$$H_0 : \mu = 8$$
$$H_A : \mu \neq 8$$

Two-sided hypothesis testing with p-values

- If the research question was "Do the data provide convincing evidence that the average amount of schools that GMU students apply to is **different** than the national average?", the alternative hypothesis would be different.

$$H_0 : \mu = 8$$
$$H_A : \mu \neq 8$$

- Hence the p-value could change as well:

```
college_apps_p_value_two_sided <- college_apps_null %>%
  filter(stat >= 9.7 | stat <= 6.3) %>%
  summarize(p_value = n() / 10000)
```

Two-sided hypothesis testing with p-values

- If the research question was "Do the data provide convincing evidence that the average amount of schools that GMU students apply to is **different** than the national average?", the alternative hypothesis would be different.

$$H_0 : \mu = 8$$
$$H_A : \mu \neq 8$$

- Hence the p-value could change as well:

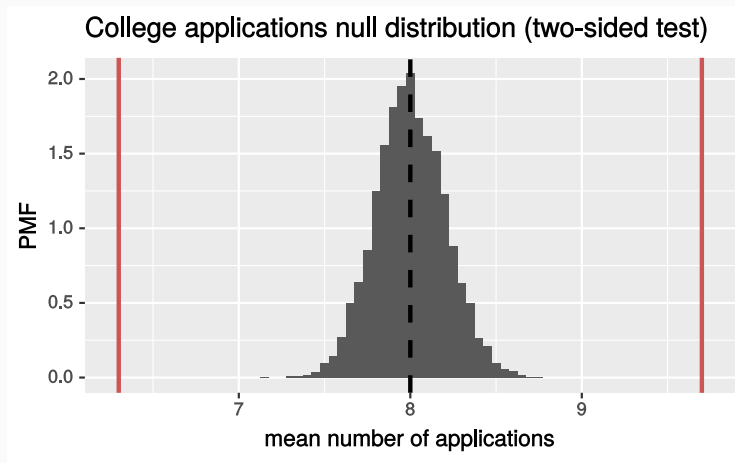College applications null distribution (two-sided test)
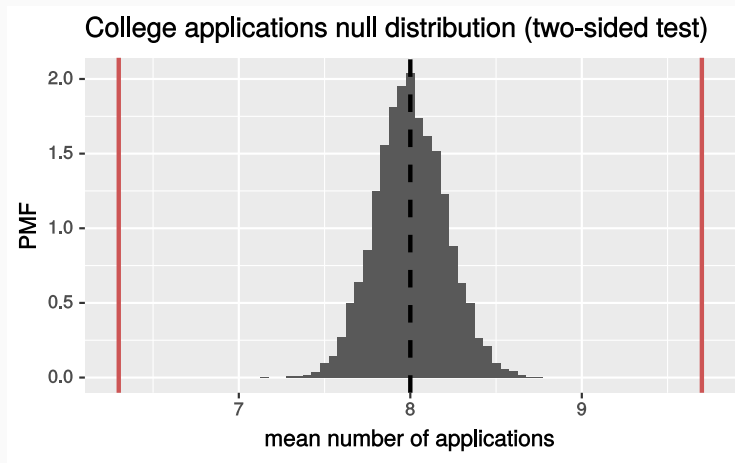
p-value = 0

Two-sided hypothesis testing with p-values

- If the research question was "Do the data provide convincing evidence that the average amount of schools that GMU students apply to is **different** than the national average?", the alternative hypothesis would be different.

$$H_0 : \mu = 8$$
$$H_A : \mu \neq 8$$

- Hence the p-value could change as well:

College applications null distribution (two-sided test)

p-value = 0

Although in this example, it does not change.

p-value for two-sided hypothesis test of gender discrimination dataset

Gender discrimination dataset: two-sided hypothesis test

- We can use the same null distribution that we generated earlier

Gender discrimination dataset: two-sided hypothesis test

- We can use the same null distribution that we generated earlier

- In the two-sided hypothesis test, we need to count when the difference in the men and women hiring fractions is **larger** than 0.292 and also when it is in the opposite extreme, which would be when the bias is towards hiring more women than men

- We can use the same null distribution that we generated earlier

- In the two-sided hypothesis test, we need to count when the difference in the men and women hiring fractions is **larger** than 0.292 and also when it is in the opposite extreme, which would be when the bias is towards hiring more women than men

- The opposite extreme corresponds to a difference in hiring fractions that is **less than** -0.292

Gender discrimination dataset: two-sided hypothesis test

- We can use the same null distribution that we generated earlier

- In the two-sided hypothesis test, we need to count when the difference in the men and women hiring fractions is **larger** than 0.292 and also when it is in the opposite extreme, which would be when the bias is towards hiring more women than men

- The opposite extreme corresponds to a difference in hiring fractions that is **less than** -0.292

- As before, we can filter the data just to keep these extreme outcomes, then divide the remaining rows and divide by 10,000

```
simulation_results %>%
  filter(stat >= experiment_result | stat <= -experiment_result) %>%
  summarize(pvalue = n() / 10000)
```

Gender discrimination dataset: two-sided hypothesis test

- We can use the same null distribution that we generated earlier

- In the two-sided hypothesis test, we need to count when the difference in the men and women hiring fractions is **larger** than 0.292 and also when it is in the opposite extreme, which would be when the bias is towards hiring more women than men

- The opposite extreme corresponds to a difference in hiring fractions that is **less than** -0.292

- As before, we can filter the data just to keep these extreme outcomes, then divide the remaining rows and divide by 10,000

```
simulation_results %>%
  filter(stat >= experiment_result | stat <= -experiment_result) %>%
  summarize(pvalue = n() / 10000)
```
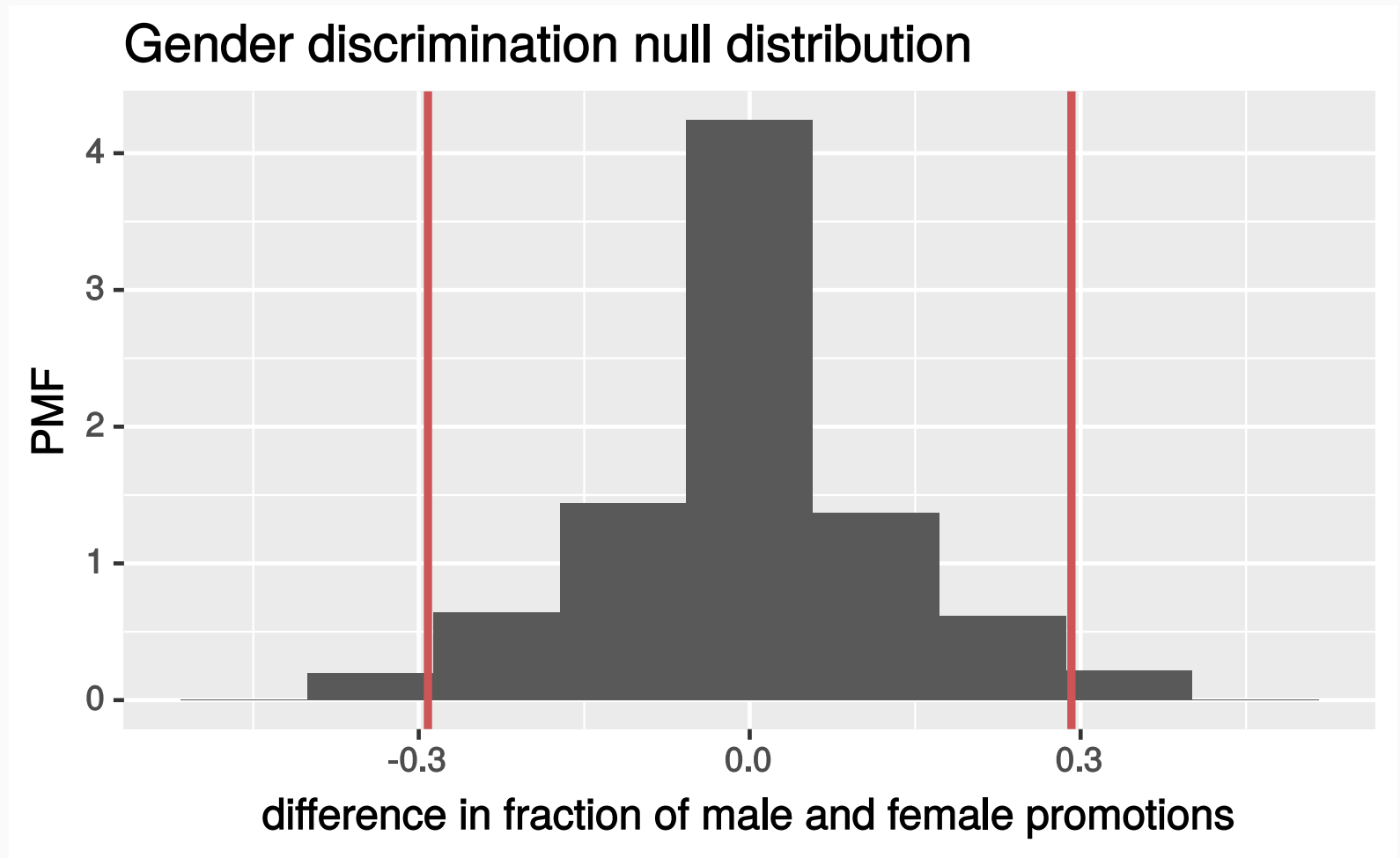
pvalue
0.0482

Visualization of null distribution (two-sided)

Gender discrimination null distribution

Mythbusting statistics

Mythbusters yawning experiment

An experiment conducted by the *MythBusters*, a science entertainment TV program that aired on the Discovery Channel, tested the following hypothesis:

> It becomes more likely for a person to yawn if another person near them yawns.

Episode 28: *Is Yawning Contagious?*, first aired on March 9, 2005

Mythbusters yawning experiment

An experiment conducted by the *MythBusters*, a science entertainment TV program that aired on the Discovery Channel, tested the following hypothesis:

> It becomes more likely for a person to yawn if another person near them yawns.

To conduct the experiment, the team drove 50 volunteers, one at a time, to a flea market in a van, and one of the team members, Kari, would either yawn (treatment) or not yawn (no treatment) during the ride.

Episode 28: *Is Yawning Contagious?*, first aired on March 9, 2005

Mythbusters yawning experiment

An experiment conducted by the *MythBusters*, a science entertainment TV program that aired on the Discovery Channel, tested the following hypothesis:

> It becomes more likely for a person to yawn if another person near them yawns.

To conduct the experiment, the team drove 50 volunteers, one at a time, to a flea market in a van, and one of the team members, Kari, would either yawn (treatment) or not yawn (no treatment) during the ride.

The 50 people were randomly assigned to two groups:

Episode 28: *Is Yawning Contagious?*, first aired on March 9, 2005

Mythbusters yawning experiment

An experiment conducted by the *MythBusters*, a science entertainment TV program that aired on the Discovery Channel, tested the following hypothesis:

> It becomes more likely for a person to yawn if another person near them yawns.

To conduct the experiment, the team drove 50 volunteers, one at a time, to a flea market in a van, and one of the team members, Kari, would either yawn (treatment) or not yawn (no treatment) during the ride.

The 50 people were randomly assigned to two groups:

- **Treatment group**: 34 people where a person near them yawned

- **Control group**: 16 people where there wasn't a person yawning near them

Episode 28: *Is Yawning Contagious?*, first aired on March 9, 2005

Mythbusters yawning experiment

group	yawns	total	fraction_yawned
Control	4	16	0.2500
Treatment	10	34	0.2941

On the show they presented the above results, concluding that the myth was confirmed because there was a 4% increase in yawns between the control and treatment groups.

Mythbusters yawning experiment

group	yawns	total	fraction_yawned
Control	4	16	0.2500
Treatment	10	34	0.2941

On the show they presented the above results, concluding that the myth was confirmed because there was a 4% increase in yawns between the control and treatment groups.

Having learned how to carry out a hypothesis test with `infer`, do the data allow us to draw the same conclusions as the *Mythbusters*?

Mythbusters yawning experiment

group	yawns	total	fraction_yawned
Control	4	16	0.2500
Treatment	10	34	0.2941

On the show they presented the above results, concluding that the myth was confirmed because there was a 4% increase in yawns between the control and treatment groups.

Having learned how to carry out a hypothesis test with `infer`, do the data allow us to draw the same conclusions as the *Mythbusters*?

Let's step through a series of guided questions to work this out.

```
library(tidyverse)
library(infer)
yawn <- read_csv(
  "http://summer18.cds101.com/files/datasets/yawn.csv"
)
experiment_result <- 10/34 - 4/16
```

The yawn dataset

subject_id	yawn	group
1	no	Control
2	no	Control
3	no	Treatment
4	no	Control
5	no	Treatment
…	…	…
46	no	Treatment
47	yes	Treatment
48	no	Control
49	yes	Treatment
50	yes	Control

Question 1

In this experiment, what is the **explanatory** variable and what is the **response** variable? What value in the **response** value should be classified as a success?

Question 1

In this experiment, what is the **explanatory** variable and what is the **response** variable? What value in the **response** value should be classified as a success?

- **explanatory** → `group`

- **response** → `yawn`

Question 2

To conduct the hypothesis test, we need to simulate the null distribution. What quantity will we compute so we can build the null distribution?

1. The mean number of `yawns` in the treatment group

2. The mean `fraction_yawned` in the treatment group

3. The mean difference in `yawns` between the treatment and control groups

4. The mean difference in `fraction_yawned` between the treatment and control groups

Question 2

To conduct the hypothesis test, we need to simulate the null distribution. What quantity will we compute so we can build the null distribution?

1. The mean number of `yawns` in the treatment group

2. The mean `fraction_yawned` in the treatment group

3. The mean difference in `yawns` between the treatment and control groups

4. The mean difference in `fraction_yawned` between the treatment and control groups

Question 3

When building up our `infer`-based simulation, what should the arguments for `specify()` be? Generally, the structure of specify is:

```
specify(formula = variable1 ~ variable2, success = "success_label")
```

The words `variable1`, `variable2`, and `"success_label"` are placeholders. Based on your response to question 1, tell me what to fill in for these three placeholders.

Question 3

When building up our `infer`-based simulation, what should the arguments for `specify()` be? Generally, the structure of specify is:

```
specify(formula = variable1 ~ variable2, success = "success_label")
```

The words `variable1`, `variable2`, and `"success_label"` are placeholders. Based on your response to question 1, tell me what to fill in for these three placeholders.

```
specify(formula = yawn ~ group, success = "yes")
```

Question 4

For a hypothesis test, `specify()` is piped into `hypothesize()`. What should the arguments for `hypothesize()` be?

1. `hypothesize(null = "independence")`

2. `hypothesize(null = "point")`

Question 4

For a hypothesis test, `specify()` is piped into `hypothesize()`. What should the arguments for `hypothesize()` be?

1. `hypothesize(null = "independence")`

2. `hypothesize(null = "point")`

Question 5

Next, we pipe `hypothesize()` into `generate()`, which is where we say how many simulations we want to run to generate the null distribution.

The structure of `generate()` is:

```
generate(reps = number,  type = "simulation_type")
```

`reps` is an integer, and type can be `"bootstrap"`, `"permute"`, or `"simulate"`.

`number` and `"simulation_type"` are placeholders.

If we want to run 10,000 simulations to create the null distribution for this dataset, what should I replace the placeholders with?

Question 5

Next, we pipe `hypothesize()` into `generate()`, which is where we say how many simulations we want to run to generate the null distribution.

The structure of `generate()` is:

```
generate(reps = number,  type = "simulation_type")
```

`reps` is an integer, and type can be `"bootstrap"`, `"permute"`, or `"simulate"`.

`number` and `"simulation_type"` are placeholders.

If we want to run 10,000 simulations to create the null distribution for this dataset, what should I replace the placeholders with?

```
generate(reps = 10000,  type = "permute")
```

Question 6

Next, we pipe `generate()` into `calculate()`, which is where we say what quantity is being measured.

The structure of `calculate()` is:

```
calculate(stat = "compute_stat",  order = c("level1", "level2"))
```

`"compute_stat"` and `c("level1", "level2")` are placeholders.

The `stat` argument can be one of the following: `"mean"`, `"median"`, `"sd"`, `"prop"`, `"diff in means"`, `"diff in medians"`, `"diff in props"`, `"Chisq"`, `"F"`, and `"slope"`.

The labels put in the `order` argument correspond to levels in the explanatory variable.

Based on your response to question 2, what should I replace the placeholders with?

Question 6

Next, we pipe `generate()` into `calculate()`, which is where we say what quantity is being measured.

The structure of `calculate()` is:

```
calculate(stat = "compute_stat",  order = c("level1", "level2"))
```

`"compute_stat"` and `c("level1", "level2")` are placeholders.

The `stat` argument can be one of the following: `"mean"`, `"median"`, `"sd"`, `"prop"`, `"diff in means"`, `"diff in medians"`, `"diff in props"`, `"Chisq"`, `"F"`, and `"slope"`.

The labels put in the `order` argument correspond to levels in the explanatory variable.

Based on your response to question 2, what should I replace the placeholders with?

```
calculate(stat = "diff in props",  order = c("Treatment", "Control"))
```

Question 7

Take your answers to questions 3 through 6 and write the `infer` code needed to simulate the null distribution for this experiment.

Question 7

Take your answers to questions 3 through 6 and write the `infer` code needed to simulate the null distribution for this experiment.

```r
yawn_null <- yawn %>%
  specify(yawn ~ group, success = "yes") %>%
  hypothesize(null = "independence") %>%
  generate(reps = 10000, type = "permute") %>%
  calculate(stat = "diff in props", order = c("Treatment", "Control"))
```

Question 8

Use the simulated null distribution you obtained in question 7 and find the p-value of a **one-sided hypothesis test**.

Question 8

Use the simulated null distribution you obtained in question 7 and find the p-value of a **one-sided hypothesis test**.

```
yawn_null %>%
  filter(stat >= experiment_result) %>%
  summarize(pvalue = n() / 10000)
```

pvalue
0.5198

Question 9

Use the simulated null distribution you obtained in question 7 and find the p-value of a **two-sided hypothesis test**.

Question 9

Use the simulated null distribution you obtained in question 7 and find the p-value of a **two-sided hypothesis test**.

```
yawn_null %>%
  filter(stat >= experiment_result | stat <= -experiment_result) %>%
  summarize(pvalue = n() / 10000)
```
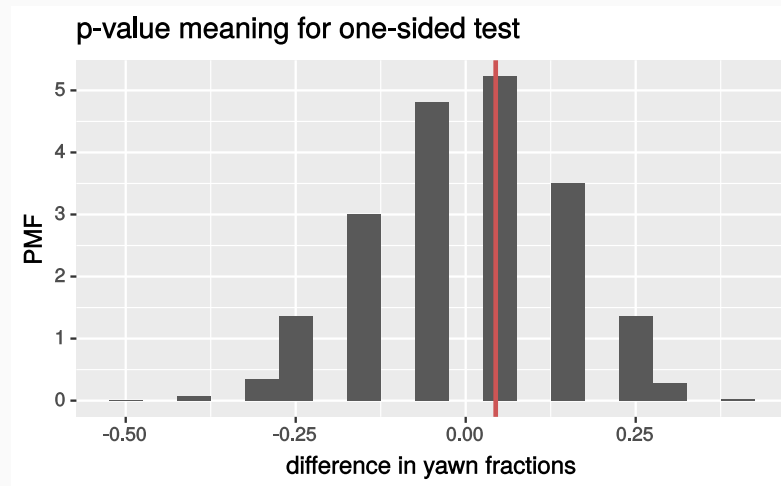
pvalue
1

Question 10

Create a plot that shows the meaning of the p-value for one-sided and two-sided hypothesis tests.

Question 10

Create a plot that shows the meaning of the *p*-value for one-sided and two-sided hypothesis tests.

```
ggplot(yawn_null) +
  geom_histogram(mapping = aes(x = stat, y = ..density..), binwidth = 0.05) +
  geom_vline(xintercept = experiment_result, color = "indianred3", size = 1) +
  labs(
    title = "p-value meaning for one-sided test",
    x = "difference in yawn fractions",
    y = "PMF"
  )
```

Question 10

Create a plot that shows the meaning of the *p*-value for one-sided and two-sided hypothesis tests.

```
ggplot(yawn_null) +
  geom_histogram(mapping = aes(x = stat, y = ..density..), binwidth = 0.05) +
  geom_vline(xintercept = experiment_result, color = "indianred3", size = 1) +
  geom_vline(xintercept = -experiment_result, color = "indianred3", size = 1) +
  labs(
    title = "p-value meaning for two-sided test",
    x = "difference in yawn fractions",
    y = "PMF"
  )
```
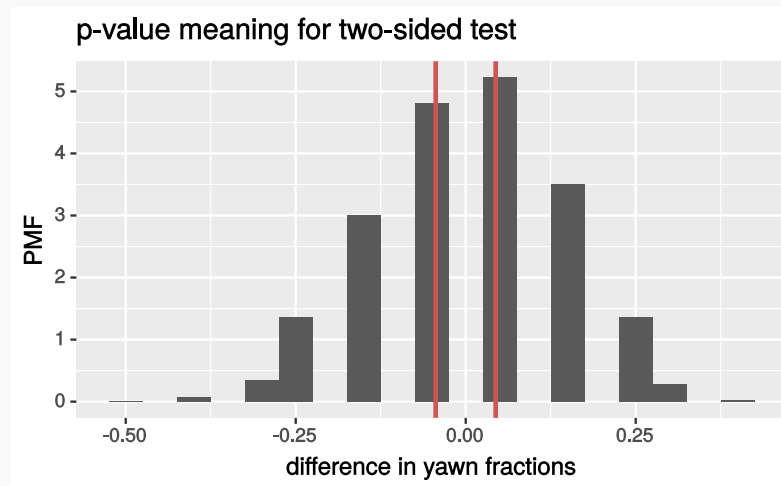
Question 11

For a significance level of $\alpha = 0.05$, are we able to reject the null hypothesis for either one of the tests?

Question 11

For a significance level of $\alpha = 0.05$, are we able to reject the null hypothesis for either one of the tests?

- *p-value of one-sided test: 0.5198*

- *p-value of two-sided test: 1*

Both are larger than $\alpha = 0.05$, hence we cannot reject the null hypothesis in either test.

Credits

Portions of these slides were adapted from the chapter 3 OpenIntro Statistics slides developed by Mine Çetinkaya-Rundel and made available under the CC BY-SA 3.0 license.