



Testing for a Difference in Population Means (for Independent Groups)

Mark Rulkowski

Lecturer, Department of Statistics



Research Question

Considering Mexican-American adults (ages 18 - 29) living in the United States, do males have a significantly higher mean Body Mass Index than females?

Research Question

Considering Mexican-American adults (ages 18 - 29) living in the United States, do males have a significantly higher mean Body Mass Index than females?

Population: Mexican-American adults (ages 18 - 29) in the U.S.

Research Question

Considering Mexican-American adults (ages 18 - 29) living in the United States, do males have a significantly higher mean Body Mass Index than females?

Population: Mexican-American adults (ages 18 - 29) in the U.S.

Parameter of Interest ($\mu_1 - \mu_2$): Body Mass Index or BMI (kg/m^2)

Research Question

Considering Mexican-American adults (ages 18 - 29) living in the United States, do males have a significantly higher mean Body Mass Index than females?

Task: Perform an independent samples t-test regarding the value for the difference in mean BMI between males and females.

Steps to Perform a Hypothesis Test

1. Define null and alternative hypotheses
2. Examine data, check assumptions, and calculate test statistic
3. Determine corresponding p-value
4. Make a decision about null hypothesis

Step 1: Define Hypotheses

Null: There is no difference in mean BMI

Alternative: There is a significant difference in mean BMI

(Both statements are for the specified populations)

Step 1: Define Hypotheses

Null: There is no difference in mean BMI

Alternative: There is a significant difference in mean BMI

(Both statements are for the specified populations)

$$H_0: \mu_1 = \mu_2$$

$$H_a: \mu_1 \neq \mu_2$$

Step 1: Define Hypotheses

Null: There is no difference in mean BMI

Alternative: There is a significant difference in mean BMI

(Both statements are for the specified populations)

$$H_0: \mu_1 = \mu_2 \text{ (or } \mu_1 - \mu_2 = 0 \text{)}$$

$$H_a: \mu_1 \neq \mu_2 \text{ (or } \mu_1 - \mu_2 \neq 0 \text{)}$$

Step 1: Define Hypotheses

Null: There is no difference in mean BMI

Alternative: There is a significant difference in mean BMI

(Both statements are for the specified populations)

$$H_0: \mu_1 = \mu_2 \text{ (or } \mu_1 - \mu_2 = 0 \text{)}$$

$$H_a: \mu_1 \neq \mu_2 \text{ (or } \mu_1 - \mu_2 \neq 0 \text{)}$$

Significance Level = 5%

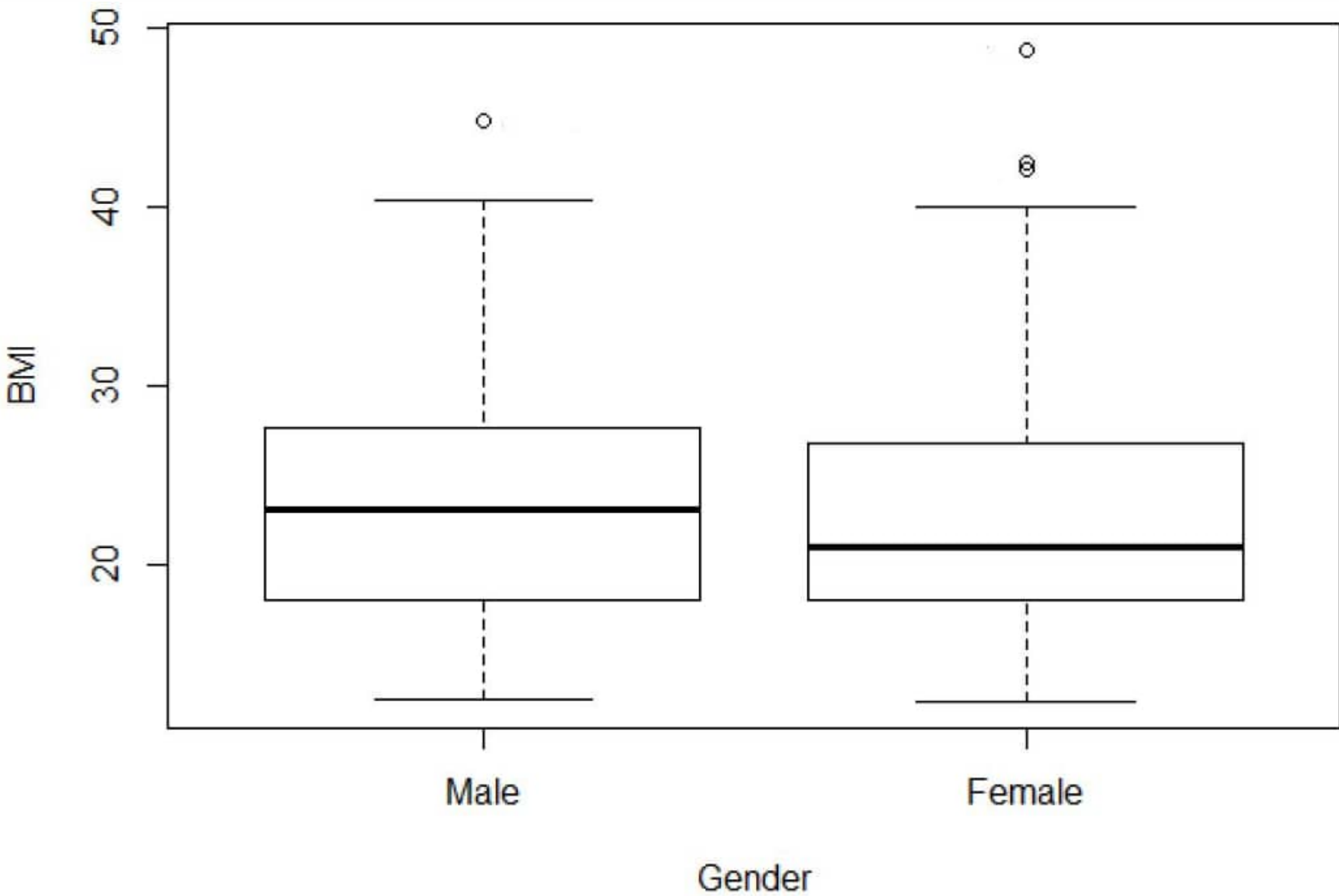
Step 2: Examine Data

Gender	BMI	Race	Age 18-29
1	19.9	1	1
2	17.0	1	1
2	26.7	1	1
1	25.6	1	1
...

The data was filtered to include only Mexican-American adults that were between the ages of 18 and 29.

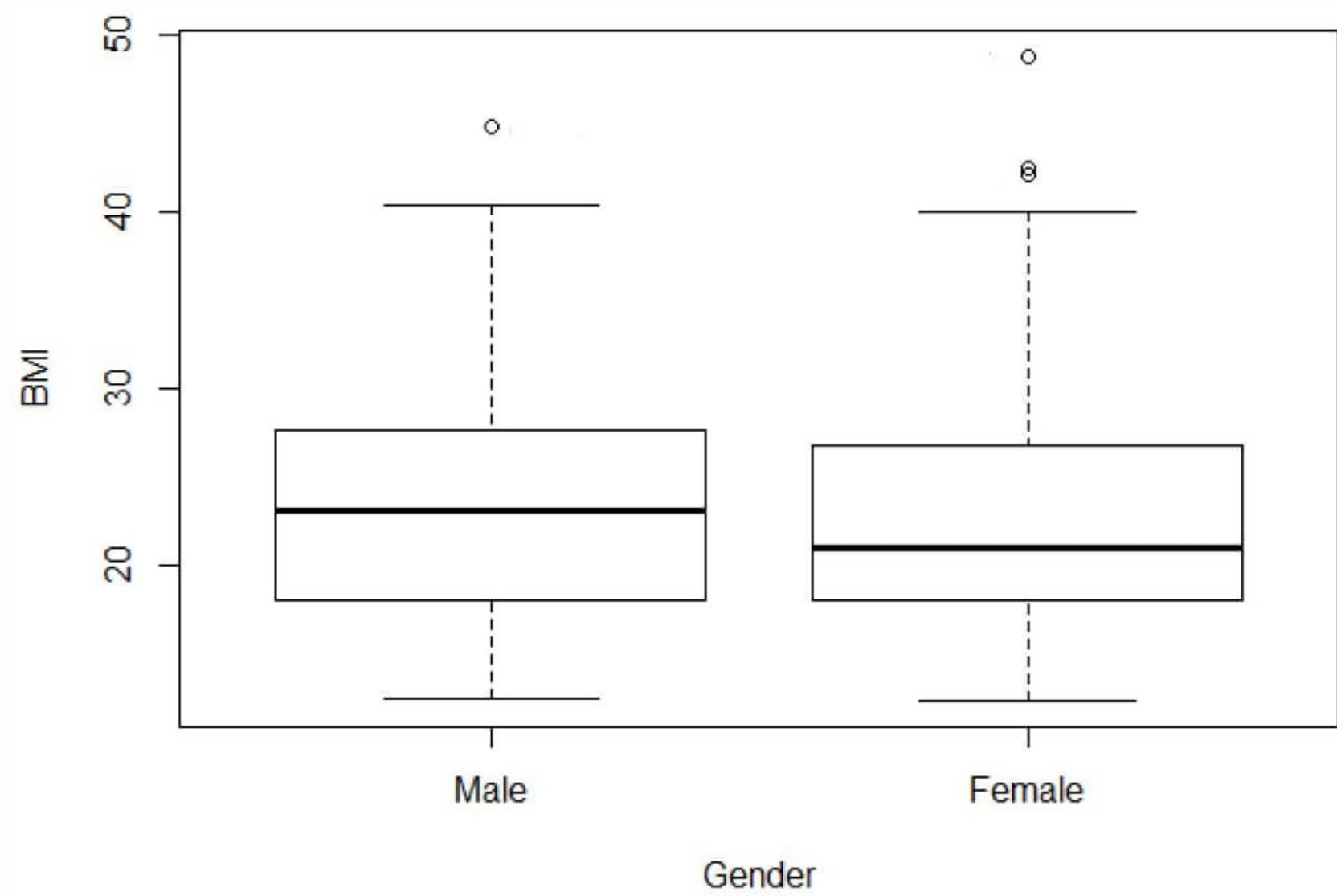
Step 2: Examine Data

	Male	Female
Mean	23.57	22.83
St. Dev.	6.24	6.43
Min	12.5	12.4
Max	44.9	48.8
n	258	239



Step 2: Examine Data

	Male	Female
Mean	23.57	22.83
St. Dev.	6.24	6.43
Min	12.5	12.4
Max	44.9	48.8
n	258	239



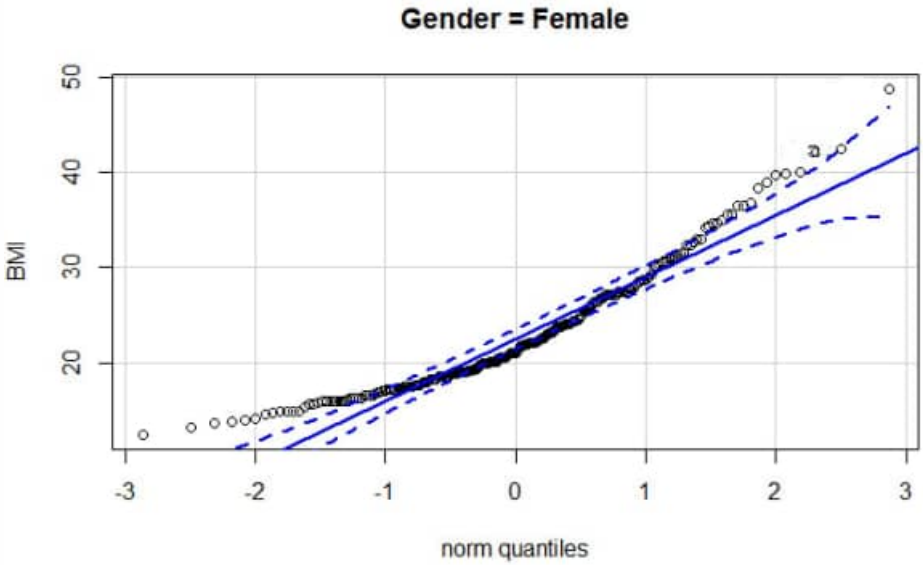
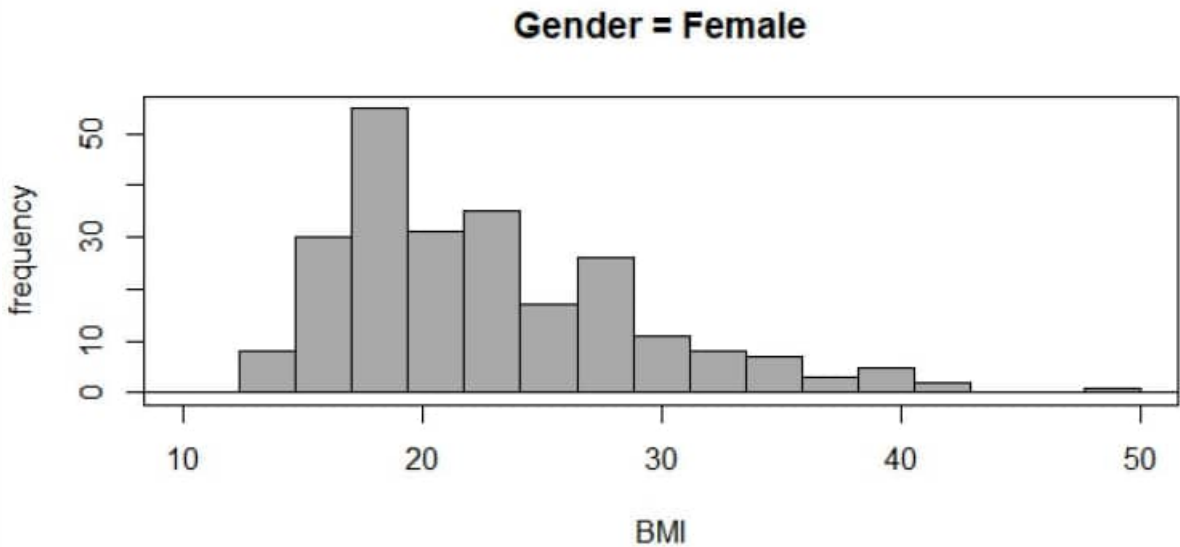
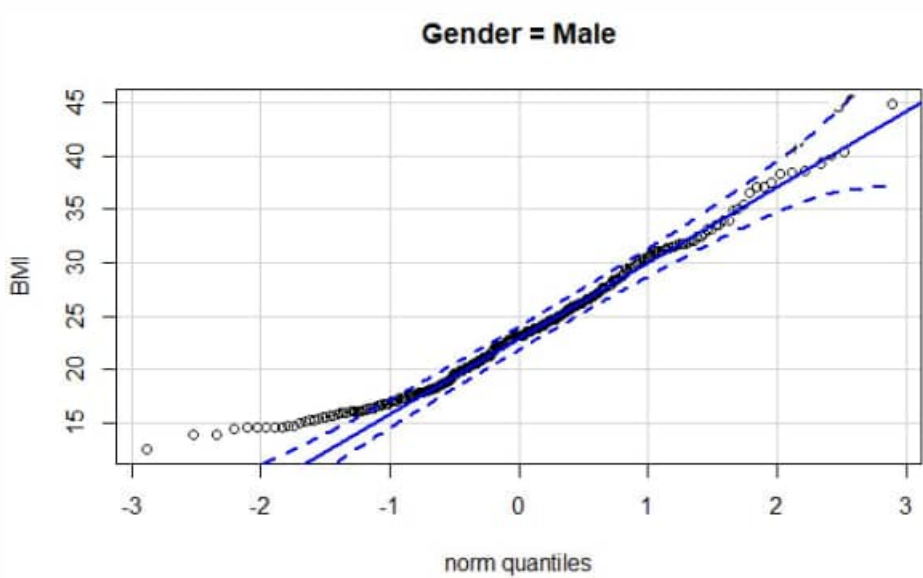
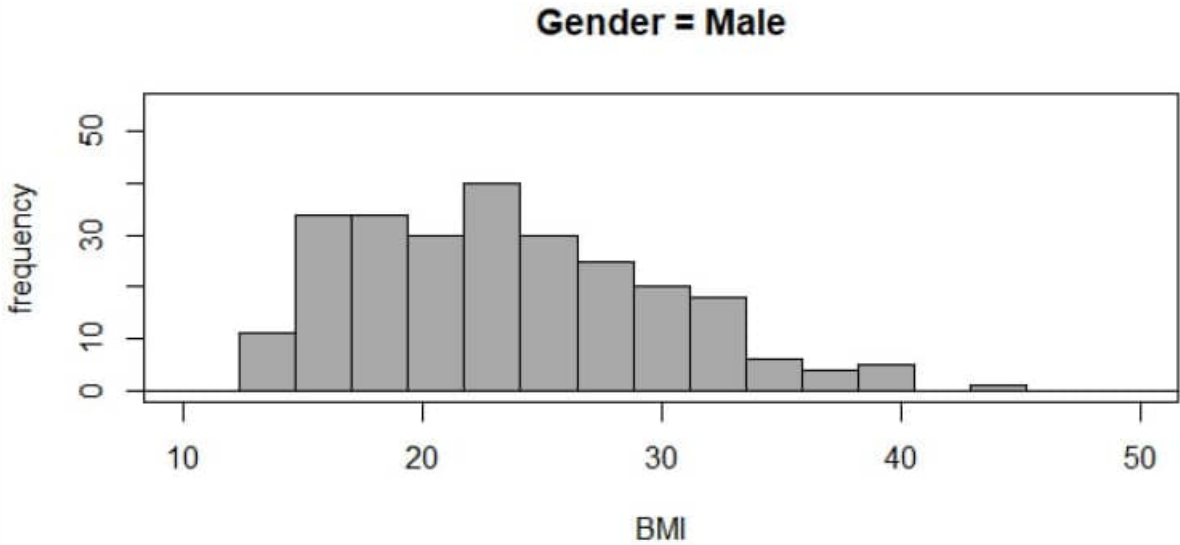
Step 2: Check Assumptions

Samples are considered simple random samples

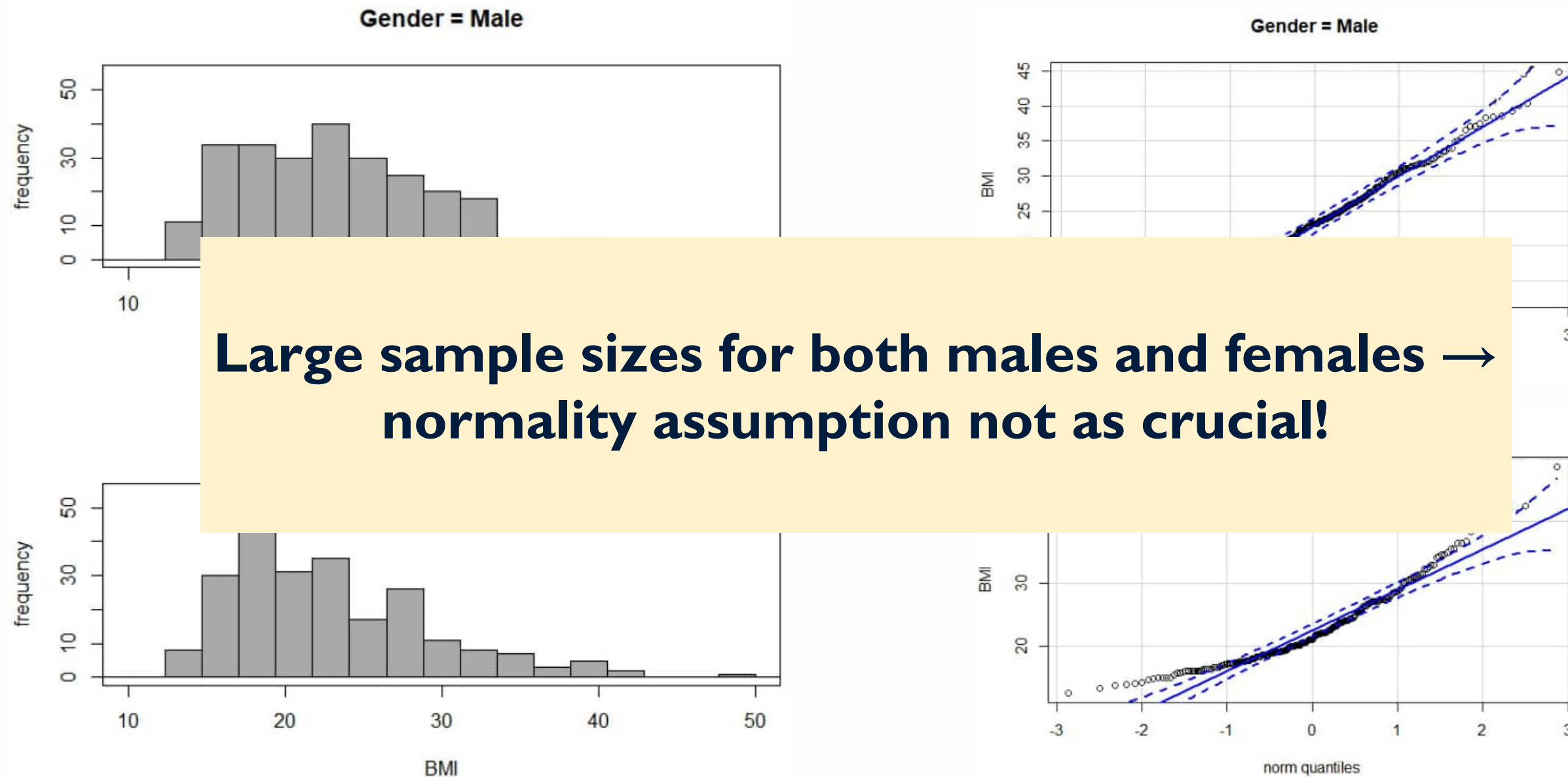
Samples are independent from one another

Both populations of responses are approximately normal (or sample sizes are both 'large' enough)

Step 2: Check Assumptions



Step 2: Check Assumptions



Step 2: Calculate Test Statistic

- $H_0: \mu_1 - \mu_2 = 0$ vs. $H_a: \mu_1 - \mu_2 \neq 0$

Best Estimate: $\bar{x}_1 - \bar{x}_2 = 23.57 - 22.83 = 0.74$

Is our sample mean difference of 0.74 kg/m² significantly different than 0?

Step 2: Calculate Test Statistic

- $H_0: \mu_1 - \mu_2 = 0$ vs. $H_a: \mu_1 - \mu_2 \neq 0$

Best Estimate: $\bar{x}_1 - \bar{x}_2 = 23.57 - 22.83 = 0.74$

Is our sample mean difference of 0.74 kg/m² significantly different than 0?

We'll use a test statistic to find out!

Test Statistic

A measure of how far our sample statistic is from our hypothesized population parameter, in terms of estimated standard errors

The further away our sample statistic is, the less confident we'll be in our null hypothesized value

Step 2: Calculate Test Statistic

-

$$t = \frac{\text{best estimate} - \text{null value}}{\text{estimated standard error}}$$

Step 2: Calculate Test Statistic

Pooled Approach

The variance of the two populations are assumed to be equal

$$(\sigma_1^2 = \sigma_2^2)$$

Unpooled Approach

The assumption of equal variances is dropped

Step 2: Calculate Test Statistic

Pooled Approach:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Step 2: Calculate Test Statistic

Pooled Approach:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

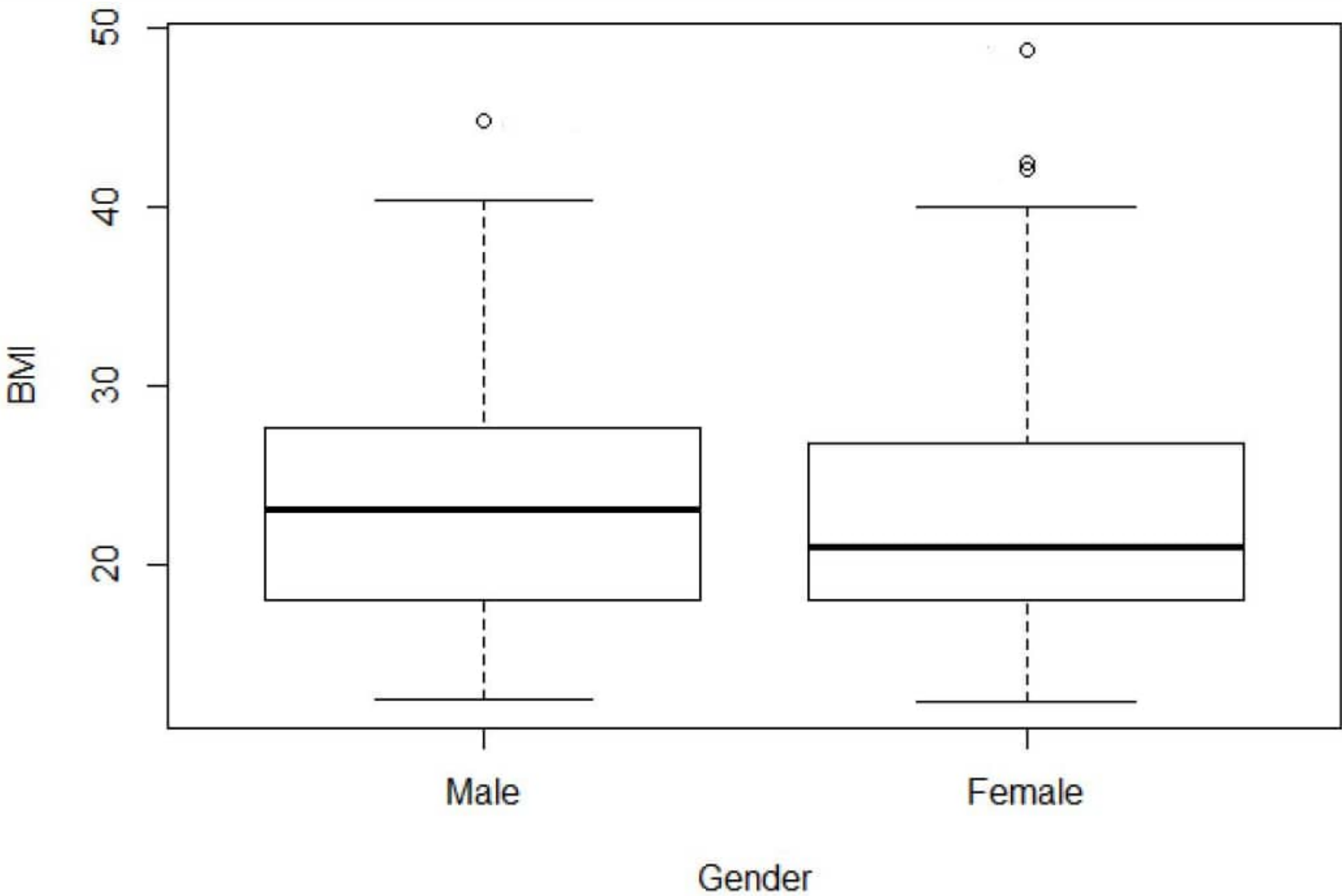
Step 2: Calculate Test Statistic

Unpooled Approach:

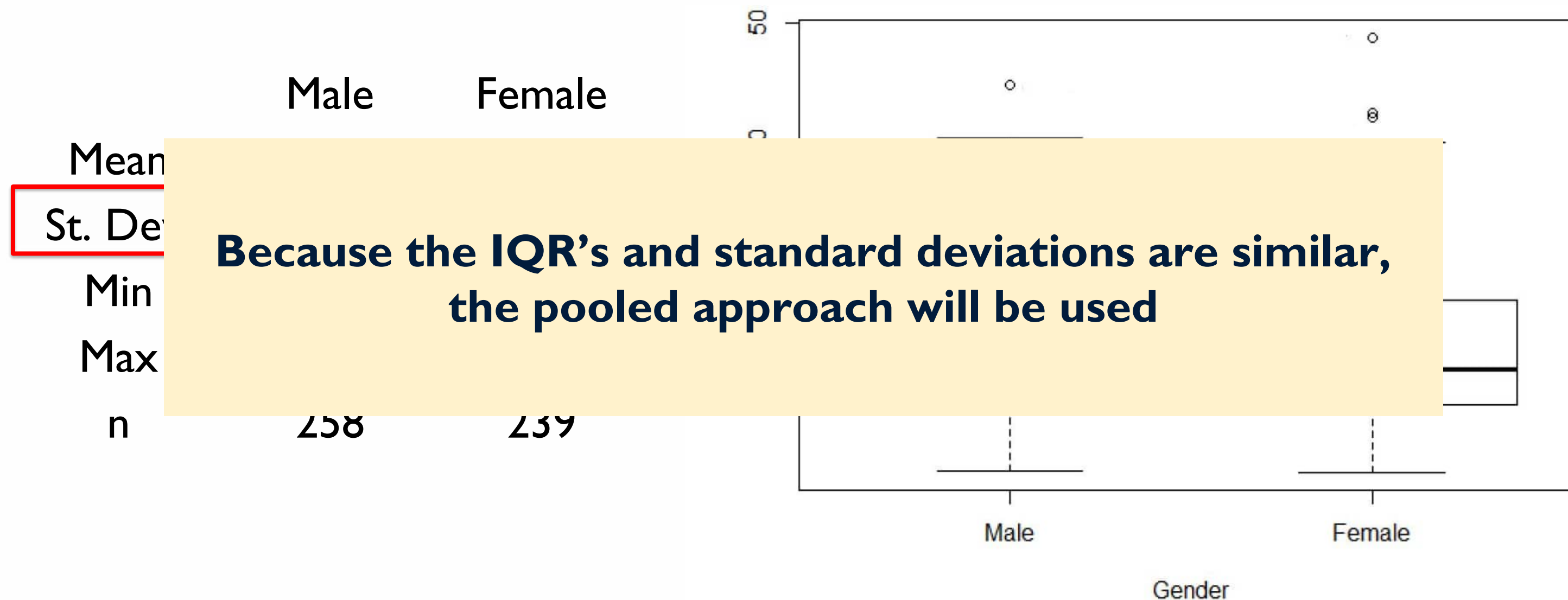
$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Pooled or Unpooled?

	Male	Female
Mean	23.57	22.83
St. Dev.	6.24	6.43
Min	12.5	12.4
Max	44.9	48.8
n	258	239



Pooled or Unpooled?



Step 2: Calculate Test Statistic

- $$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{(23.57 - 22.83)}{\sqrt{\frac{(257)6.24^2 + (238)6.43^2}{495}} \sqrt{\frac{1}{258} + \frac{1}{239}}}$$

Step 2: Calculate Test Statistic

- $$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{(23.57 - 22.83)}{\sqrt{\frac{(257)6.24^2 + (238)6.43^2}{495}} \sqrt{\frac{1}{258} + \frac{1}{239}}}$$

$$t = \frac{0.74}{0.0898 * 6.332} = \mathbf{1.30}$$

Step 2: Calculate Test Statistic

- $$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{(23.57 - 22.83)}{\sqrt{\frac{1}{258} + \frac{1}{239}}}$$

Our difference in sample means is only 1.30
 (*estimated*) standard errors
 above the null difference of 0 kg/m²

$$t = \frac{0.74}{0.0898 * 6.332} = 1.30$$

Step 3: Determine p-value

- $t = 1.30$

If the null hypothesis ($\mu_1 - \mu_2 = 0$) were true, would a test statistic value of 1.30 be unusual enough to reject the null?

Step 3: Determine p-value

- $t = 1.30$

If the null hypothesis ($\mu_1 - \mu_2 = 0$) were true, would a test statistic value of 1.30 be unusual enough to reject the null?

p-value: assuming the null hypothesis is true, it is the probability of observing a test statistic of 1.30 or more extreme

Step 3: Determine p-value

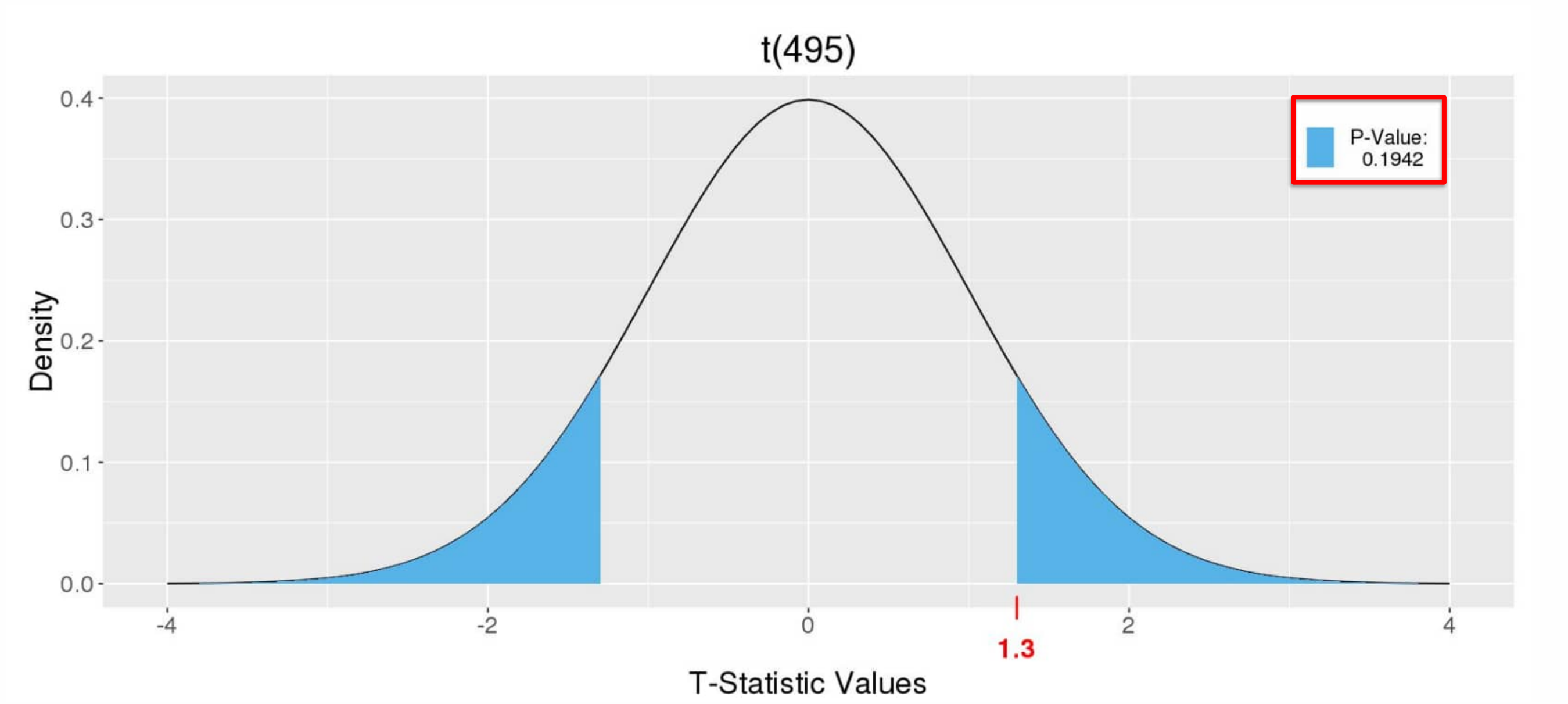
Using a $t(df)$ distribution where $df = n_1 + n_2 - 2$

Step 3: Determine p-value

Using a $t(df)$ distribution where $df = n_1 + n_2 - 2$

Our alternative hypothesis is two-sided ($\mu_1 - \mu_2 \neq 0$) so we will check both the upper and lower tail

Step 3: Determine p-value



Step 3: Determine p-value

$$\text{p-value} = 0.19$$

If the difference in population mean BMI between males and females was really 0 kg/m²,
then observing a difference in sample means of 0.74 kg/m² (i.e. a t-statistic of 1.30) or more extreme is **fairly likely**.

Step 4: Make a Decision

Our p-value is larger than the 0.05 significance level, which means there is weak evidence against the null.

Thus, we **fail to reject the null!**

Step 4: Make a Decision

Based on our estimated difference in sample means, we cannot support that there is a significant difference between the population mean BMI for males and the population mean BMI for females for the population of all Mexican-Americans adults (ages 18 - 29) living in the U.S.

95% Confidence Interval Results

In a previous lecture, we calculated the 95% CI for the difference in mean BMI between males and females

$$\left(-0.385 \text{ kg}/\text{m}^2, 1.865 \text{ kg}/\text{m}^2 \right)$$

Our test value of 0 kg/m² falls within our interval. This is a reasonable value for the difference in mean BMI.

Summary

- Hypothesis Tests are used to put theories about a parameter of interest to the test ~ **parameter = difference in population means**

Summary

- Hypothesis Tests are used to put theories about a parameter of interest to the test ~ **parameter = difference in population means**
- Basic Steps:
 - **Define hypotheses (and select significance level)**
 - **Examine data, check assumptions, and calculate test statistic**
 - **Determine corresponding p-value**
 - **Make a decision about null hypothesis**

Summary

- Hypothesis Tests are used to put theories about a parameter of interest to the test ~ **parameter = difference in population means**
- Basic Steps:
 - **Define hypotheses (and select significance level)**
 - **Examine data, check assumptions, and calculate test statistic**
 - **Determine corresponding p-value**
 - **Make a decision about null hypothesis**
- Assumptions for Two-Sample (t) Test for Population Means
 - ~ **data are two simple random sample, independent**
 - ~ **both populations of responses are normal (else n large helps)**

Summary

- Hypothesis Tests are used to put theories about a parameter of interest to the test ~ **parameter = difference in population means**
- Basic Steps:
 - **Define hypotheses (and select significance level)**
 - **Examine data, check assumptions, and calculate test statistic**
 - **Determine corresponding p-value**
 - **Make a decision about null hypothesis**
- Assumptions for Two-Sample (t) Test for Population Means
 - ~ **data are two simple random sample, independent**
 - ~ **both populations of responses are normal (else n large helps)**
- Know how to interpret the **p-value, decision, and conclusion**