Hypothesis Testing

Instructor, Nero Chan Zhen Yu

Forward School

Orange Juice

- A bottled juice manufacturer's most popular product is 355ml bottle. Since this info is on the label, we assume it is true.
- But is it?

Forward School

Orange Juice Problem

- As a customer, we are concerned that there is at least 355ml in a bottle. If it is little more that is okay. We get more for the money.

- However, if we are the manufacturer, we want the volume to be exactly 355ml. We do not want to upset customers (under filling) and we do not higher cost of production (overfilling)

Forward School

Orange Juice Assumption

Customer

- Assumes at least 355ml

$$Quantity\ Juice \geq 355ml$$

Manufacturer

- Assumes exactly 355ml

$$Quantity\ Juice = 355ml$$

Forward School

Orange Juice Questions

Customer

- Is there an average, at least 355ml of orange juice in each bottle?

$$Quantity\ Juice \geq 355ml$$

Manufacturer

- Is there an average, exactly 355ml of orange juice in each bottle?

$$Quantity\ Juice = 355ml$$

Forward School

Orange Juice Experiment

- So we can collect 50 bottles from all over the country (world) to randomize the sample in terms of location, time, manufacturing plant etc.
- We then assume the volume of each bottle in the sample and find the mean volume for all 50 bottles.
- Using those sample means, we can test the assumptions; the status quo

Forward School

Hybrid Engine

- An auto manufacturer has developed a new hybrid engine technology. It **claims** reduces fuel consumption while driving in the city.
- The **claim** is that the new technology improves fuel efficiency making it better than the old engine that produces 30mpg
- The company will run controlled tests to look for statistical evidence to support the **claim** that new engine offers better efficiency than the old model.

Forward School

Hybrid Engine

- Company claims

Fuel Efficiency > 30mpg

- The manufacturer is making a claim it wishes to test; it is NOT testing an assumption (status quo) that already exists.

Forward School

Thinking about hypothesis

- When trying to formulate a statistical hypothesis, always ask yourself the following question:

- *"Am I testing an assumption, or status quo, that already exists? (juice bottle). Or am I testing a claim or assertion beyond what I already know or can know? (hybrid engines)*

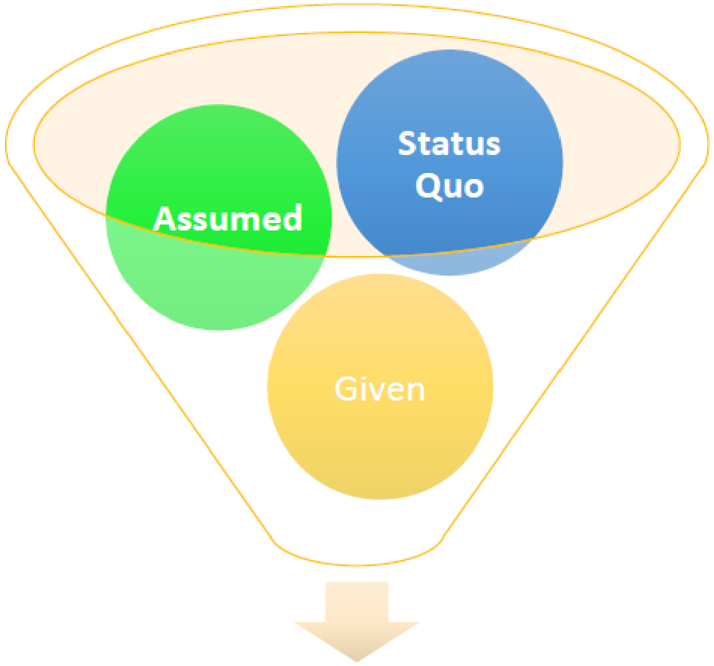
- If you wanted to test the gas mileage listed on a current car sticker, you would be testing an assumption, not a claim.

Forward School

Null and alternate hypothesis

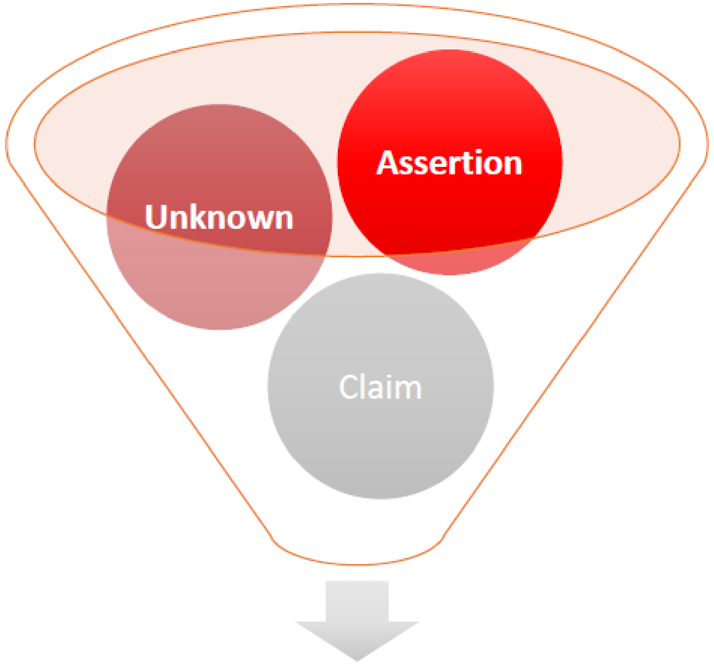
- Formulating hypothesis can be frustrating:
  - The reason is the null and alternative are two opposing roads that lead to the same place
- By definition, the null and alternative hypothesis are opposites; mutually exclusive
  - The null is either rejected or it is not. Only if null is rejected then you can proceed to the alternative
- Researchers can start with either the null or the alternative, and then form the other as a complement to the first
- Which to start with largely depends on the point of view of the researcher, the context of the problem, and what can or cannot be assumed to be known upfront.

Null and alternate hypothesis



Null

“This is accepted as true, let’s test it.”



Alternative

“This might be true, let’s test it.
If not, the truth is something else”

Forward School

Null and alternate meanings

Null hypothesis

H_0

Alternative hypothesis

H_a

"Null" means nothing new or different; assumption or status quo maintained.

The "Alternative" is simply 'the other option' when the null is rejected; nothing more.

Forward School

Null and alternate Properties

H_0	H_a
Assumption, status quo, nothing new	Rejection of an assumption
Assumed to be ‘true’; a given	Rejection of an assumption, or the given
Negation of the research question	Research question to be ‘proven’

Using the last property, you can logically derive the possible null/alternative pairs

$H_0 =$	$H_0 \leq$	$H_0 \geq$
$H_a \neq$	$H_a >$	$H_a <$

Forward School

Example 1: Orange Juice

- A bottled juice manufacturer states on the product label that each bottle contains 355ml of juice.
- You work for a government agency that protects consumers by testing product volumes. A sample of 50 bottles is tested.
- Is there anything we can assume to be true?
  - Yes. The355ml on the bottle is assumed to be true
- Which hypothesis pair seems to be appropriate?

$\boxed{H_0 =}$	$H_0 \leq$	$H_0 \geq$
$\boxed{H_a \neq}$	$H_a >$	$H_a <$

Forward School

Example 1: Orange Juice

- Is there anything we can assume to be true?
  - Yes. The355ml on the bottle is assumed to be true
- Which hypothesis pair seems to be appropriate?

$$\boxed{\begin{array}{c} H_0 = \\ \\ H_a \neq \end{array}}$$

$$H_0 = 355ml$$

$$H_a \neq 355ml$$

If the data indicates the bottles are being filled properly, then fail to reject the null; fail to reject our assumption.

It does not mean that the null proven but the assumption has held up.

Forward School

Example 1: Orange Juice

- Is there anything we can assume to be true?
  - Yes. The355ml on the bottle is assumed to be true
- Which hypothesis pair seems to be appropriate?

$$\boxed{\begin{array}{l} H_0 = \\ \\ H_a \neq \end{array}}$$

$$H_0 = 355ml$$

$$H_a \neq 355ml$$

If the data indicates the bottles are not being filled properly, then reject the null; reject our assumption.

The assumption has not held up under analysis. The statistical supports the validity of the alternative hypothesis.

Forward School

Example 2: Manchester United

- During the 2010-2011 English Premier League season, Manchester United home matches had an average attendance of 74,691. A club marketing analyst would like to see if attendance decreased during the most recent season. Establish a null and alternative for this analysis.
- What is our assumption?
  - We can only assume that the attendance remained the same.

Example 2: Manchester United

- What is our assumption?
  - We can only assume that the attendance remained the same at 74,691.
- Are we testing a preliminary claim?
  - The marketing analyst wishes to see if attendance has **DECREASED** since 2010-2011.
- Which hypothesis format should we choose?

$H_0 =$	$H_0 \leq$	$\boxed{H_0 \geq}$
$H_a \neq$	$H_a >$	$\boxed{H_a <}$

Forward School

Example 2: Manchester United

- What is our assumption?
  - We can only assume that the attendance remained the same at 74,691.
- Are we testing a preliminary claim?
  - The marketing analyst wishes to see if attendance has **DECREASED** since 2010-2011.
- Which hypothesis format should we choose?

$$\boxed{\begin{array}{c} H_0 \geq \\ \\ H_a < \end{array}}$$

$$H_0 \geq 74,691$$

$$H_a < 74,691$$

If the data indicates the attendance has decreased, then reject the null; reject our assumption.

The assumption has not held up under analysis. The statistical supports the validity of the alternative hypothesis.

Forward School

Null and alternative statements

- All statistical conclusions are made in reference to the null hypothesis
- As a researcher, you either **reject** the null hypothesis or **fail to reject** the null hypothesis, you do not **accept** the null
  - This is due to the fact that the null hypothesis is assumed to be true from the start; rejecting or failing to reject an assumption
- If you do **reject** the null hypothesis, then can conclude the data supports the alternative hypothesis
- However, if you **fail to reject** the null hypothesis, it does not mean we have proven that the null hypothesis to be "true"
  - Why? Because remember from that the outset we ASSUMED it was TRUE

Type I and Type II Errors

- In statistical hypothesis testing, a **type I error** is the incorrect rejection of a true null hypothesis (a "false positive"), while a **type II error** is incorrectly retaining a false null hypothesis (a "false negative").

Forward School

Fire Alarm Hypothesis

- Let’s say you are walking down the corridor; everything is normal. Suddenly you encounter a sudden smell of smoke.
- You know that mean a serious fire taking place. Or it could be nothing serious; maybe someone burned popcorn in a microwave.
- What do you do next?

Fire Alarm Hypothesis

- Let's take a look at what might happen:
  - If you think the smoky smell is nothing serious, you may decide your assumption that everything is normal is correct and you will not pull the fire alarm.
  - If you think that smoky smell is due to a serious fire, you may reject your assumption that everything is normal and you will pull the fire alarm

Forward School

Fire Alarm Hypothesis

- Now let's take a look what may go wrong:
  - You smell smoke
  - You think, "This is not normal" (reject the assumption that everything is ok). Reject your null hypothesis.
  - Therefore, you pull the fire alarm
  - The building is evacuated and the fire department arrives to investigate
  - After the investigation, it is determined there was no fire. You 'falsely' pulled the fire alarm.
  - When you rejected your assumption that everything was OK when it really was OK, hence you committed Type I error. A 'false alarm'

Type I Error

Rejection of the assumption (null hypothesis) when it should not have been rejected.

Incorrectly rejecting the null hypothesis.

In this case, a 'false alarm'

Forward School

Fire Alarm Hypothesis

- Now let’s take a look what may go wrong:
  - You smell smoke
  - You think, “It is probably someone who burned their lunch in the microwave. No big deal.”
  - Therefore, you did not reject your assumption (null hypothesis) that everything is OK. You uphold the null.
  - But there is indeed a fire, no one is injured, but the building burns to the ground.
  - When you failed to reject your assumption that everything was OK when it really was NOT OK, hence you committed Type II error.

Type II Error

Failure to reject of the assumption (null hypothesis) when it should have been rejected.

Incorrectly not rejecting the null hypothesis.

Forward School

Fire Alarm Hypothesis

$$H_0 = smoke\ is\ annoying\ but\ not\ serious; everything\ is\ OK\ as\ usual; no\ fire$$

$$H_a = smoke\ evidence\ of\ a\ serious\ fire; everything\ is\ NOT\ OK\ as\ usual; FIRE$$

		Actual Condition	
		No serious fire	Serious Fire
Conclusion	Do not reject H_0 (no serious fire)	Correct Conclusion	?
	Reject H_0 (serious fire)	?	Correct Conclusion

Where is the type I and Type II errors?

Forward School

Fire Alarm Hypothesis

$$H_0 = smoke\ is\ annoying\ but\ not\ serious; everything\ is\ OK\ as\ usual; no\ fire$$

$$H_a = smoke\ evidence\ of\ a\ serious\ fire; everything\ is\ NOT\ OK\ as\ usual; FIRE$$

		Actual Condition	
		No serious fire	Serious Fire
Conclusion	Do not reject H_0 (no serious fire)	Correct Conclusion	Type II Error
	Reject H_0 (serious fire)	Type I Error	Correct Conclusion

*In general, the real world consequences of a Type II error are much greater. In this case, a Type II error may mean loss of property or even lives.

Forward School

The Andon Hypothesis

- In a manufacturing systems such as the Toyota Production System (TPS) and other lean systems, line workers have the power and indeed the responsibility, to stop the entire production line when they notice a serious problem.
- This is done by pulling an andon cord.

Forward School

The Andon Hypothesis

- The assumption is that the production line is running correctly.
- This is the *minute by minute null hypothesis*
- When a worker notices a defect, he/she has a choice to make. To pull the cord or not?

Forward School

The Andon Hypothesis

- If the andon cord is pulled, the worker is rejecting the assumption that the process is error free; *rejecting the NULL hypothesis*.
- If the cord is not pulled, then the worker is NOT rejecting the assumption that the process is error free; thus maintaining the NULL hypothesis.

Forward School

The Andon Hypothesis

- What if the cord is pulled (rejection of the null hypothesis) but nothing actually went wrong with the production process? A false alarm.
- In this case, the worker “committed” a Type I error.
- Type I error is a wrongful rejection of the null hypothesis when it should not be rejected.

*In TPS and similar Lean manufacturing systems is perfectly OK

The Andon Hypothesis

- What if the cord is NOT pulled (thus supporting the null hypothesis) but something is actually wrong with the production process?
- In this case, the worker "committed" a Type II error.
- Type II error is a wrongful support of the null hypothesis when it should be rejected.

*In this case Type II error can be catastrophic. It may mean that hundreds or thousands of cars have the defect!

Forward School

The Andon Hypothesis

$$H_0 = problem\ is\ annoying\ but\ not\ serious;\ everything\ is\ OK\ as\ usual;\ no\ major\ defect$$

$$H_a = problem\ is\ serious;\ everything\ is\ NOT\ OK\ as\ usual;\ MAJOR\ defect$$

		Actual Condition	
		No serious defect	Serious defect
Conclusion	Accept H_0 (no serious defect)	Correct Conclusion	Type II Error
	Reject H_0 (serious defect)	Type I Error	Correct Conclusion

Forward School

Type I vs Type 2 Error

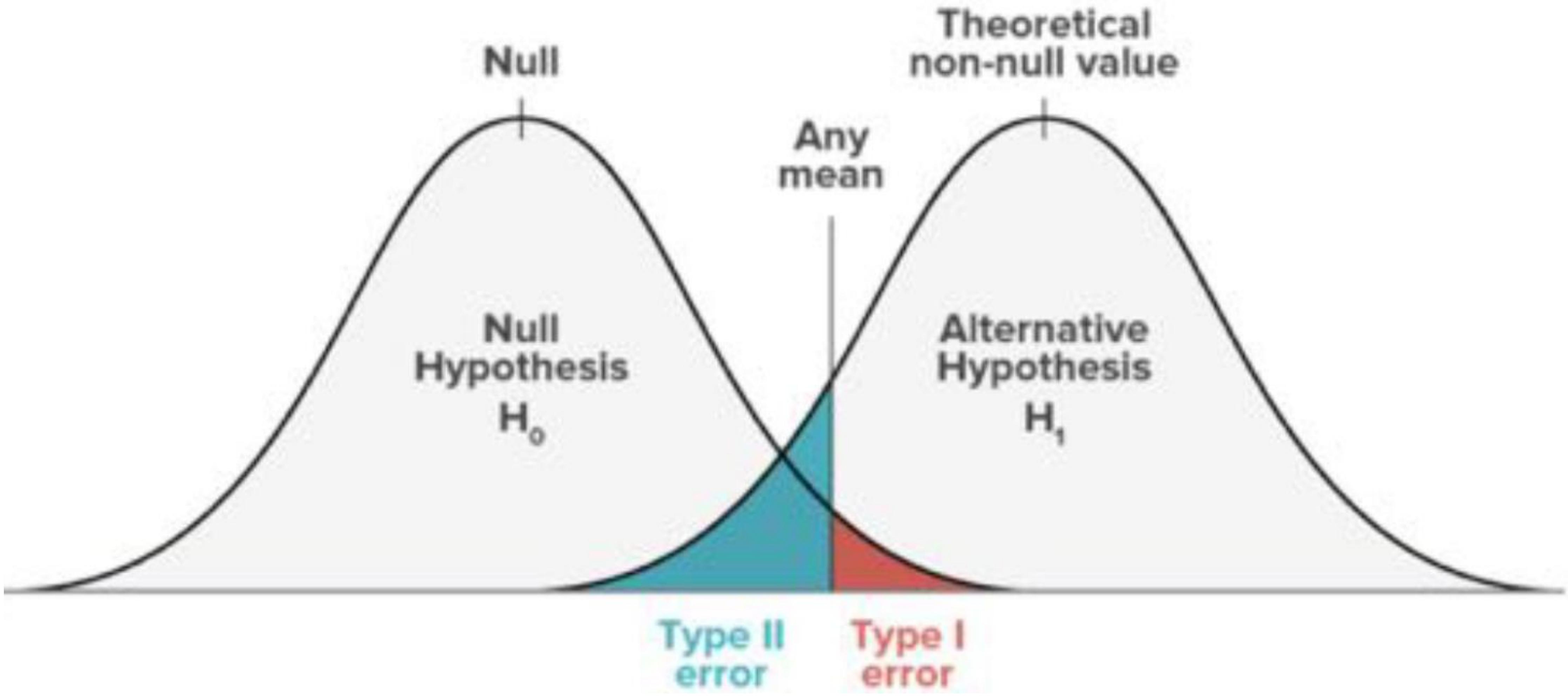
Type 1

- wrongly rejection of Null hypothesis
- False Positive
- To avoid:
  - high significance level leads to higher risk of Type 1 error
  - To reduce Type 1 error, use fairly low level of significance (0.05 - 0.01)

Type 2

- Did not reject the null hypothesis
- False Negative
- Can be avoided to select slighly higher level of significance

Forward School



Forward School

Hypothesis Testing Procedure

1. Start with a well developed, clear research problem or question
2. Establish hypothesis, both null and alternative
3. Determine appropriate statistical test and sampling distribution
4. Choose the Type I error rate
5. State the decision rule
6. Gather the sample data
7. Calculate test statistics
8. Make statistical conclusion
9. Make decision or inference based on conclusion

Forward School

σ Known or Unknown?

- As confidence intervals, there are two types of single-sample hypothesis tests:
  1. When the population standard deviation σ is known or given
  2. When the population standard deviation σ is not known or therefore we have to use an estimate, s
- When σ is known, we use the normal standard, or z-distribution
- When σ is not known, we use the t-distribution instead

Forward School

The hypothesized vs True Mean

$H_0: \mu = \mu_0$

$H_a: \mu \neq \mu_0$

μ is the true mean of the population under analysis

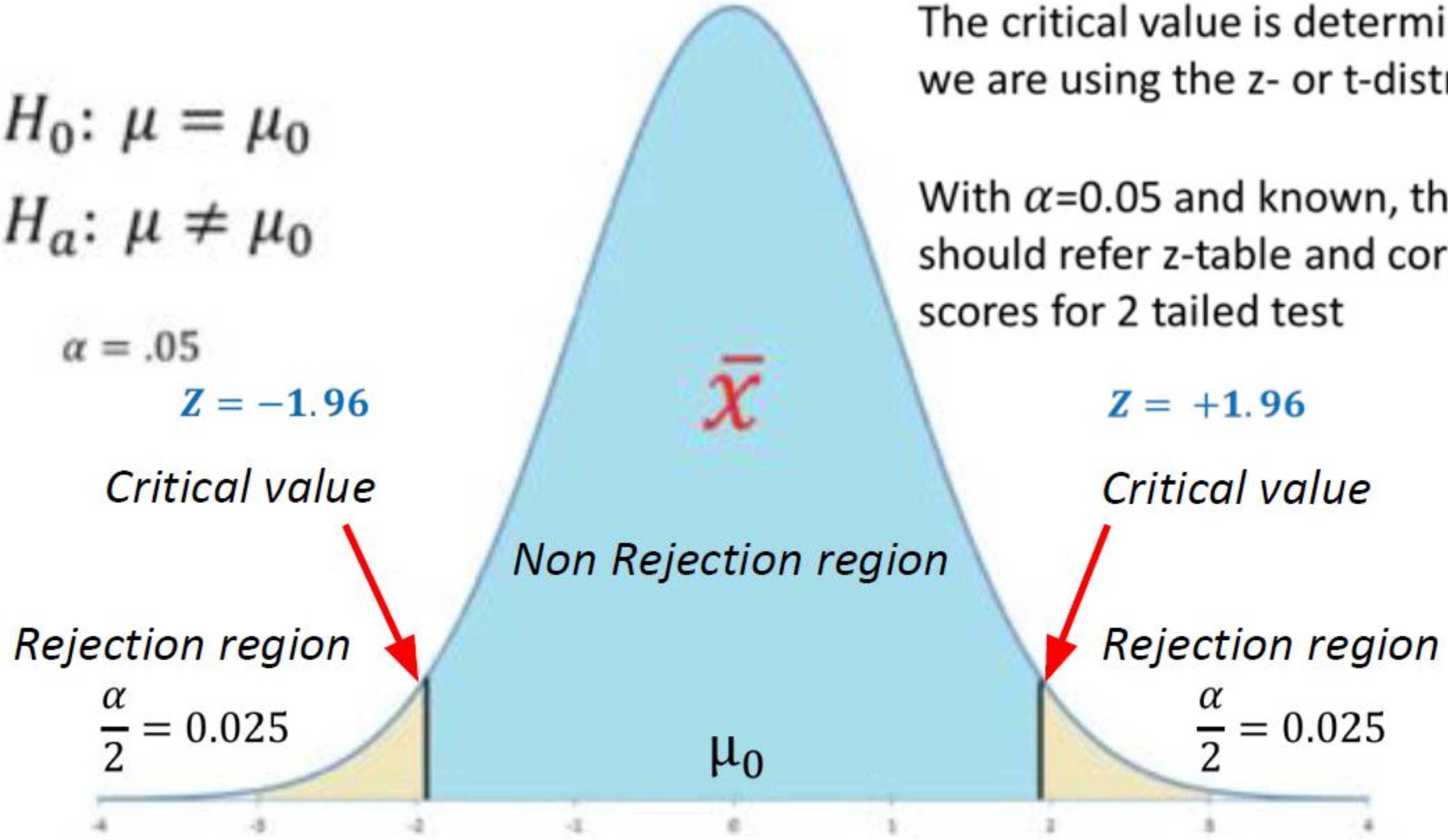
μ_0 is the hypothesized mean of the population under analysis

"Is the true mean same as the hypothesized mean?"

Test the question using sample means and confidence intervals.

Forward School

The two tailed Test Rejection Region



The critical value is determined by and if we are using the z- or t-distribution

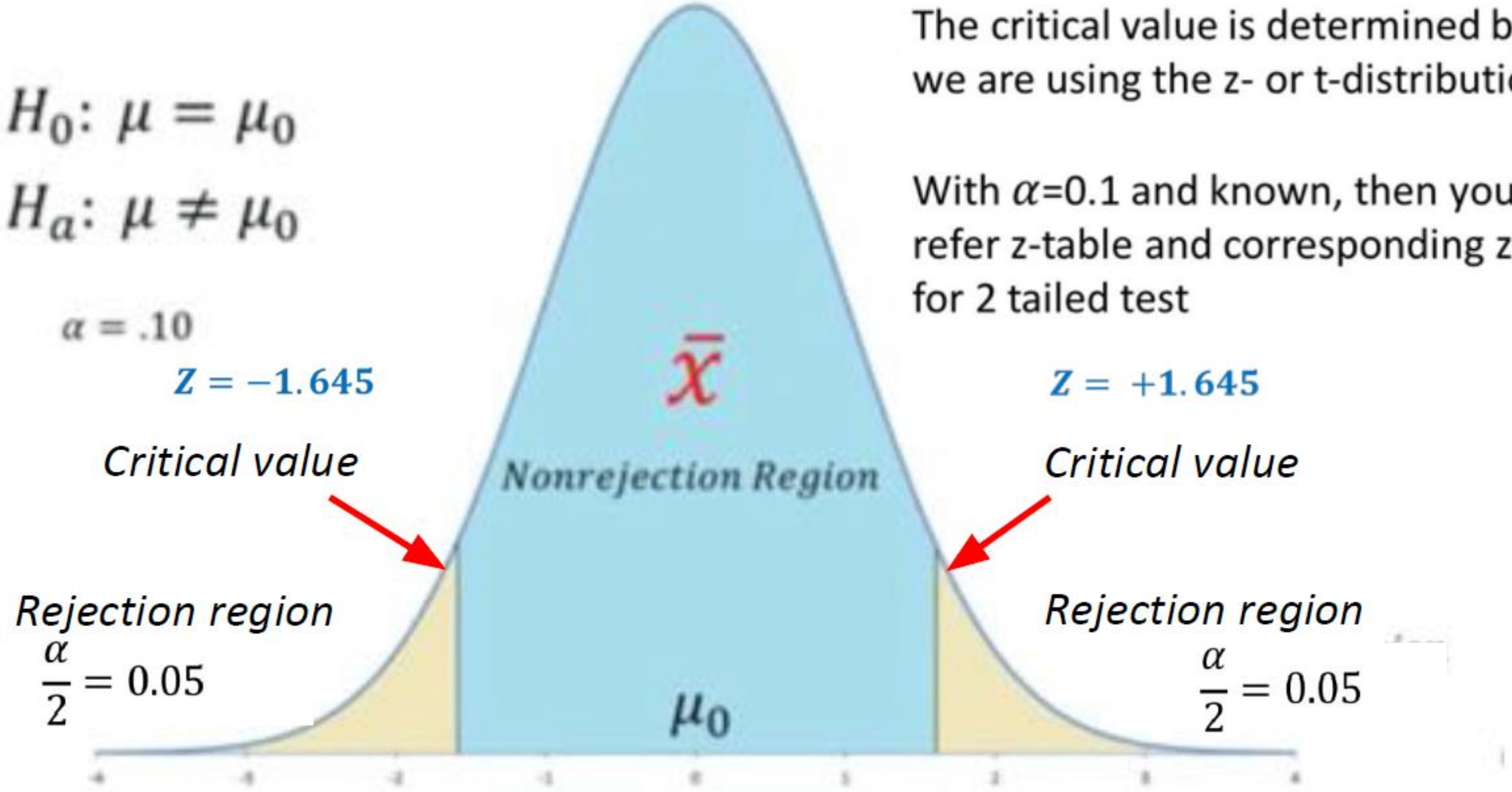
With α=0.05 and known, then you should refer z-table and corresponding z-scores for 2 tailed test

Forward School

The two tailed Test Rejection Region

The critical value is determined by and if we are using the z- or t-distribution

With α=0.1 and known, then you should refer z-table and corresponding z-scores for 2 tailed test



Forward School

The Two-tailed Test Rejection Region

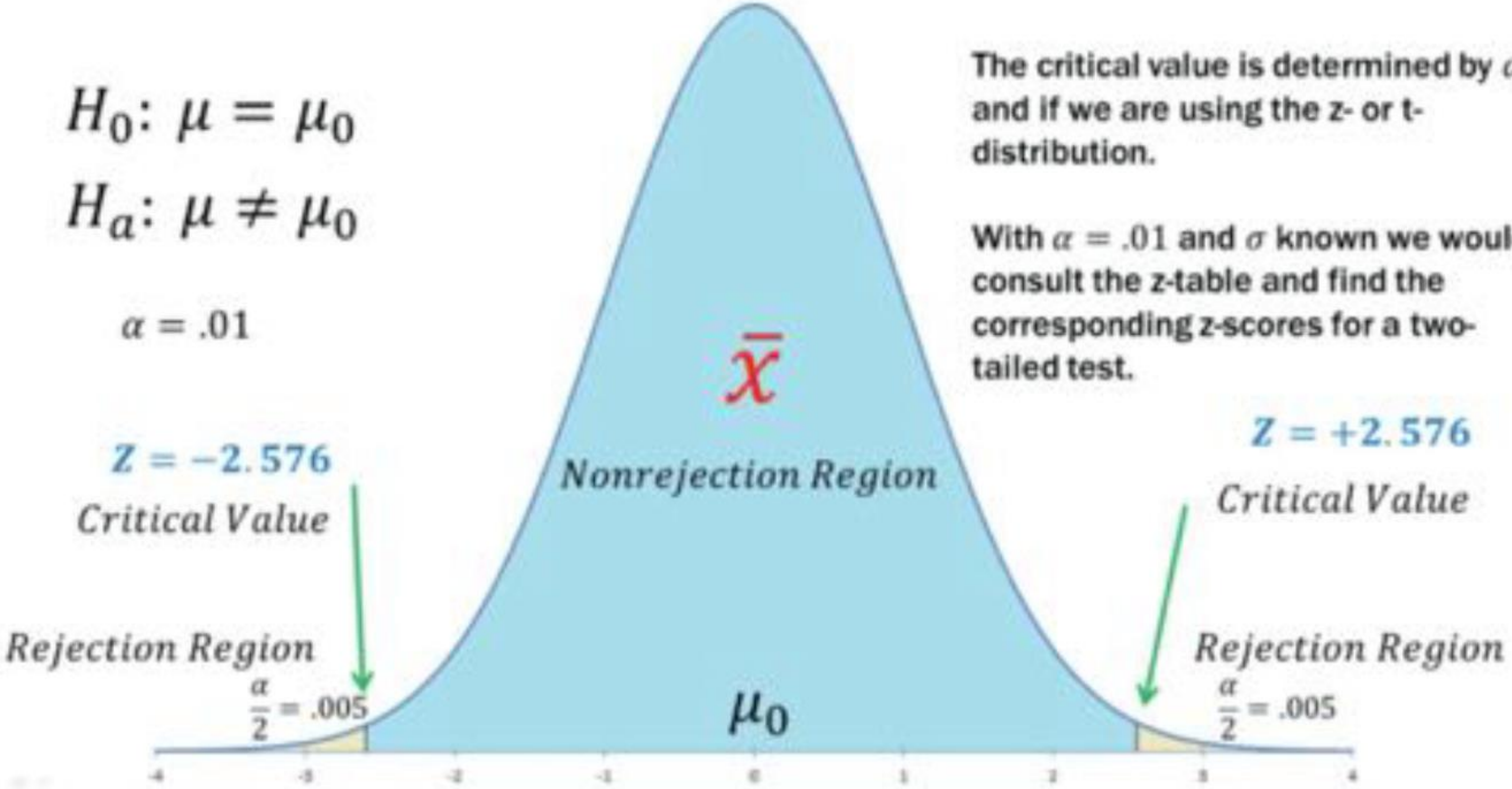
$H_0\colon \mu = \mu_0$

$H_a\colon \mu \neq \mu_0$

$\alpha = .01$

The critical value is determined by α and if we are using the z- or t-distribution.

With $\alpha = .01$ and σ known we would consult the z-table and find the corresponding z-scores for a two-tailed test.



The main question…

- Did our sample come from same population we assume is underlying the null hypothesis?
- If so, then we expect the same mean to be inside the critical region of 90% or 95% or 99% of the time depending on what we choose for α

Forward School

Testing One Population Proportion

Forward School

Research Question

In previous years, 52% of parents believed that electronics and social media was the cause of their teenager's lack of sleep. Do more parents today believe that their teenager's lack of sleep is caused due to electronics and social media?

Population: Parents with teenager (age 13 - 18)

Parameter of Interest: Population Proportion, p

Test of significant increase in the proportion of parents believed that electronics and social media was the cause of their teenager's lack of sleep

Forward School

Hypothesis

$H_0 : p = 0.52$

$H_a : p > 0.52$

“Significant Increase”

Where p is the population proportion of parents believed that electronics and social media was the cause of their teenager’s lack of sleep

$\alpha = 0.05$

Survey Result

A random sample of **1018** parents with teenager was taken and **56%** said they believe electronics and social media was the cause of their teenager's lack of sleep

Forward School

Assumption

- A **random sample** of parents
- A **large enough sample size** to ensure our distribution of sample proportions is normal
  - that is **n. p** be at least 10 ➡ $n \cdot p_0$ i.e. (1018) (0.52) = 529
  - **n . (1-p)** be at least 10 ➡ $n \cdot (1 - p_0)$ i.e (1018) (1-0.52) = 489

Forward School

Hypothesis

$H_0 : p = 0.52$

$H_a : p > 0.52$

Best Estimate of p is $\hat{p} = 0.56$

Where p is the population proportion of parents believed that electronics and social media was the cause of their teenager's lack of sleep

α = 0.05

Forward School

Test Statistic

$$\frac{\text{Best Estimate - Hypothesized Estimate}}{\text{Standard Error of Estimate}}$$

$$\frac{\hat{p}-p_0}{s.e.}$$

$$s.e.(\hat{p})=\sqrt{\frac{p.(1-p)}{n}} \quad \Rightarrow \quad \boxed{s.e.(\hat{p})=\sqrt{\frac{p_0.(1-p_0)}{n}}}$$

Forward School

Test Statistic

$$\frac{\hat{p}-p_0}{s.e.}$$

Null $s.e.(\hat{p}) = \sqrt{\frac{p_0.(1-p_0)}{n}}$

$$Z = \frac{0.56-0.52}{0.0157}$$

$$Z = 2.555$$

Forward School

Test Statistic Interpretation

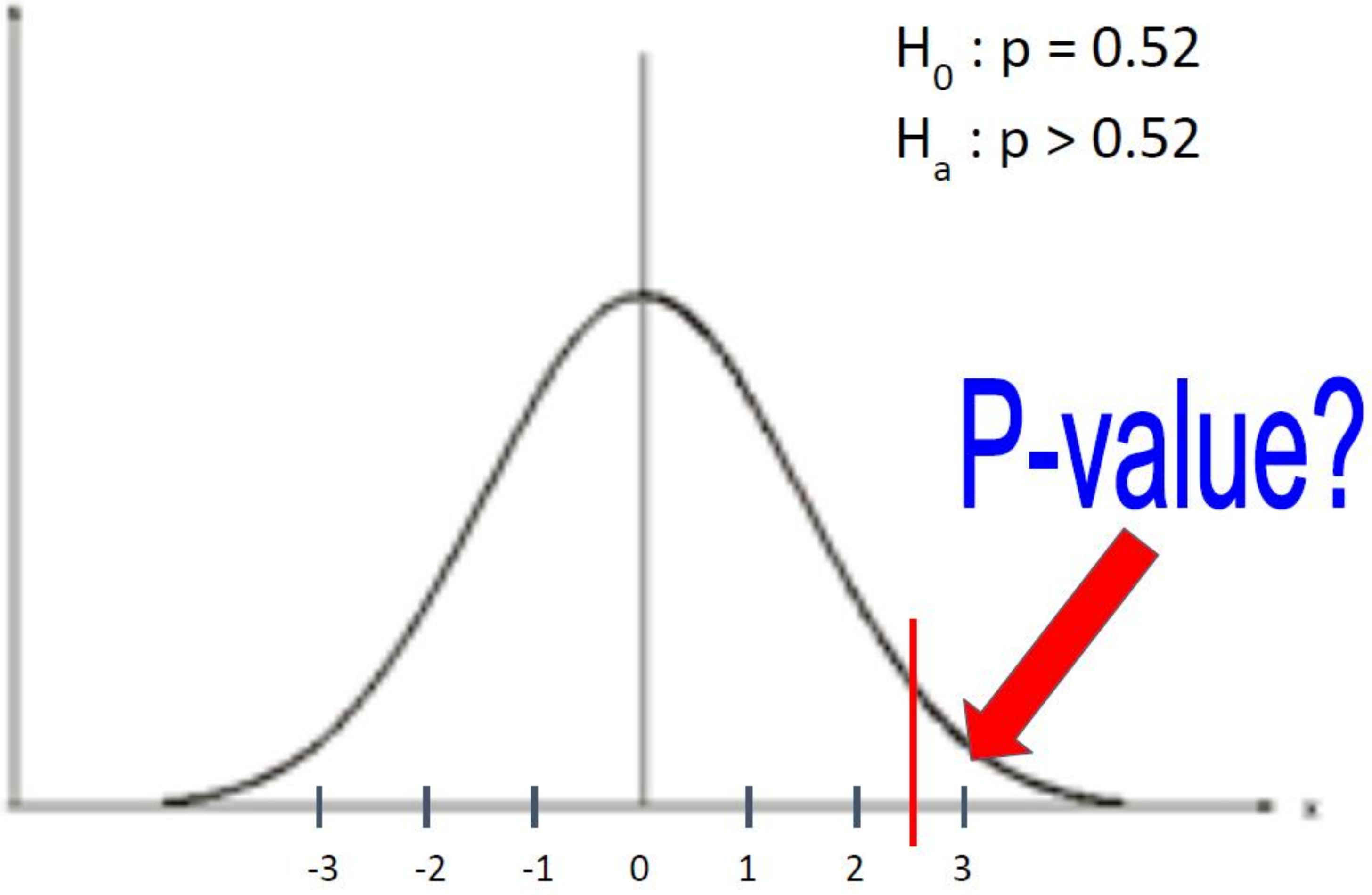
Z = 2.555

That means our observed sample proportion is 2.555 null standard errors above our hypothesized population proportion

Forward School

P- value

Z = 2.555



1-norm.cdf(2.555,0,1)= 0.0053

Forward School

Conclusion

p-value = 0.0053 < α = 0.05

Reject Null Hypothesis (H_0: p = 0.52)

There is sufficient evidence to conclude that the population proportion of parents with teenager who believe that electronics and social media is the cause for lack of sleep is greater than 52%.

Forward School

Testing a Different in Population Proportions

Forward School

Research Question

Is there a significant difference between population proportions of parents of black children and parents of hispanic children who report that their child has had some swimming lessons?

Population: All parents of black children age 6-18 and all parents of hispanic children age 6-18

We'll use 1 = Black 2 = Hispanic

Parameters of interest: p_1-p_2

Test for a significant difference in the population proportions of parents reporting that their child has had swimming lessons at the 10% significant level

Forward School

Hypothesis

$H_0 : p1 - p2 = 0$

$H_a : p1 - p2 \neq 0$

$\alpha = 0.10$

Survey Results

- A sample of 247 parents of black children age 6-18 was taken with 91 saying that their child has had some swimming lessons
- A sample of 308 parents of hispanic children age 6-18 was taken with 120 saying that their child has had some swimming lessons

Assumption

- We need to assume that we have two independent random samples
- We need a large enough sample size to assume that the distribution of our estimate is normal. That is, we need n_1.p-hat$_1$, n_1.(1-p-hat$_1$), n_2.p-hat$_2$, n_2.(1-p-hat$_2$) to be at least 10.

That is we need to estimate the common proportion, and then make sure that we would expect at least 10 yes's and 10 no's in each sample

Forward School

Best Estimate of the Parameter

$$\hat{p}_1 = \frac{91}{247} = 0.37$$

$$\hat{p}_2 = \frac{120}{308} = 0.39$$

$$\hat{p}_1 - \hat{p}_2 = 0.37 - 0.39 = -0.02$$

1 = Black
2 = Hispanic

Forward School

Hypothesis

$H_0 : p1 - p2 = 0$

$H_a : p1 - p2 \neq 0$

$\alpha = 0.10$

Forward School

Test Statistic

$$\frac{\text{Best Estimate - Hypothesised Estimate}}{\text{Standard error of estimate}}$$

$$\frac{\hat{p}_1 - \hat{p}_2 - 0}{se(\hat{p})}$$

where $se(\hat{p}) = \sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}$

$$Z = -\frac{0.02}{0.041} = -0.48$$

Forward School

Test Statistic Interpretation

$$z = -0.48$$

that means our observed difference in sample proportions is 0.48 estimated standard errors below our hypothesized mean of equal population proportions

Forward School

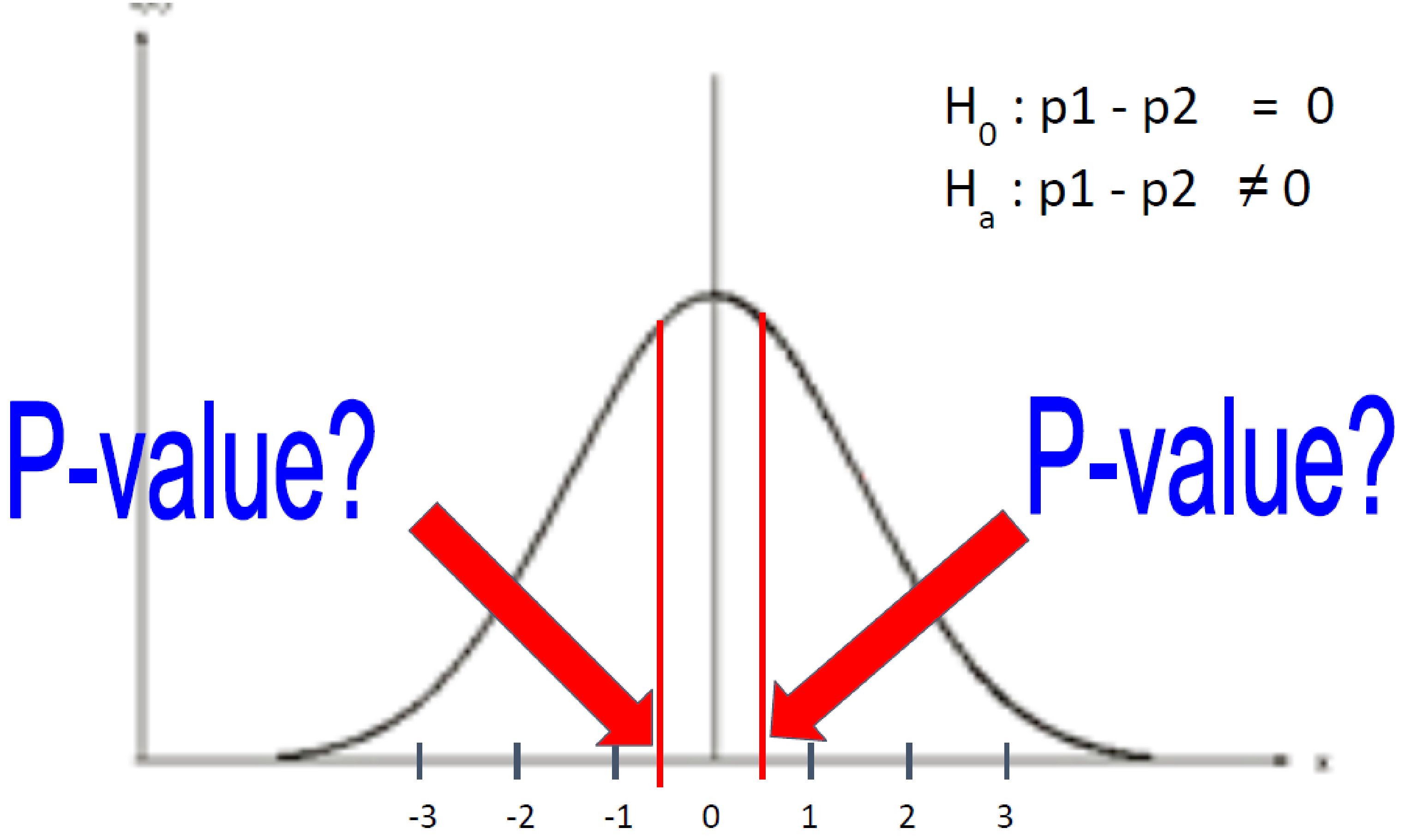
Test Statistic Distribution & p-value

$H_0: p1 - p2 = 0$

$H_a: p1 - p2 \neq 0$

$Z = -0.48$

$\alpha = 0.10$



2*(1-norm.cdf(0.48,0,1))= 0.63

Forward School

Decision & Conclusion

p-value = 0.63 > 0.10 = α → **fail to reject null hypothesis**

→ don't have evidence against equal population proportions

Formally, based on our sample and p-value, we fail to reject the null hypothesis. We conclude that there is no significant difference between the population proportion of parents of black and hispanic children who report their child has had swimming lessons.

Forward School

Alternative Approach

	Swim Lessons	No Swim Lessons	Total
Black	91	156	247
Hispanic	120	188	308
Total	221	344	555

Chi-Square (X^2) Test

Different hypothesis
require two-sided hypothesis
same conclusion*
* as two-sided hypothesis with proportions

Fisher's Exact Test

allows one sided hypothesis
typically for small sample sizes
calculate different p-values*
* compared to small setup for proportions

Forward School

One Mean: Testing Population Mean with Confidence

Forward School

Cartwheel study

25 team members/colleagues (all adults) asked to perform a cartwheel

Variable: Cartwheel distance (in inches)

Forward School

Research Question

Is the **average** cartwheel distance (in inches) for adults **more than 80 inches**

Population: All adults

Parameter of Interest: Population mean cartwheel distance, μ

Perform a one-sample test regarding the value for the mean cartwheel distance for the population of all such adults

Step 1: Define Null and Alternative

Null: Population mean CW distance, μ is 80 inches, H_0: $\mu = 80$

Alternative: Population mean CW distance, μ is greater than 80 inches, H_a: $\mu > 80$

Significant level, α = 0.05

Forward School

Step 2: Examine Results, Check Assumptions, Summarize Data via Test Statistic

```
df.describe()["CWDistance"]
```

```
count     25.000000
mean      82.480000
std       15.058552
min       63.000000
25%       70.000000
50%       81.000000
75%       92.000000
max      115.000000
Name: CWDistance, dtype: float64
```

n = 25 observations

Minimum = 63 inches
Maximum = 115 inches

Mean = 82.48 inches
Standard Deviation = 15.06 inches

Forward School

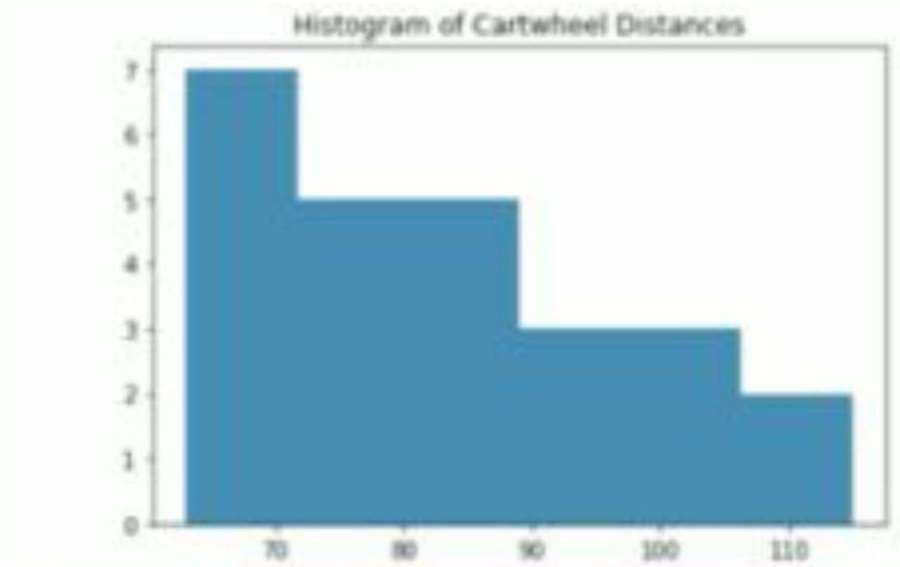
Step 2: Examine Results, Check Assumptions, Summarize Data via Test Statistic

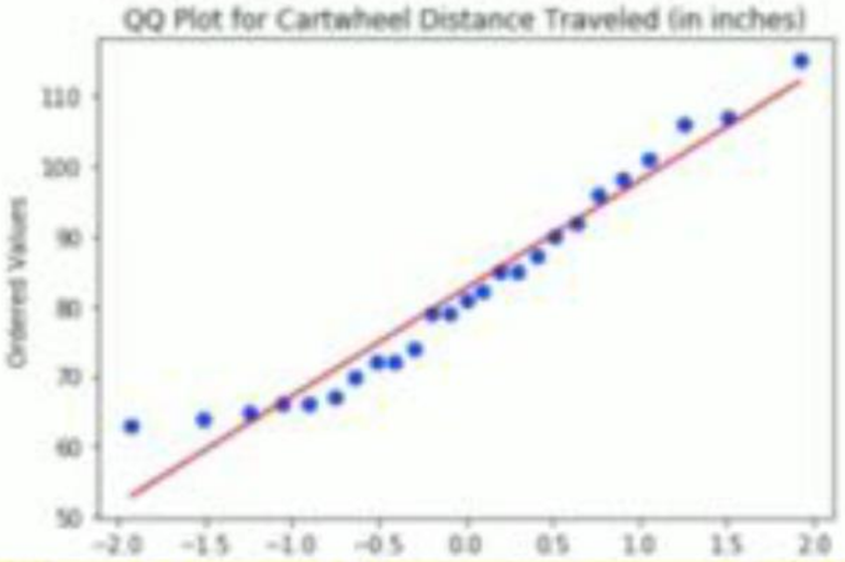
Assumptions:

- Sample of **CW Distance measurements** considered a **simple random sample**
- **Normal** distribution for CW distance in population (or with a large enough sample size)

Forward School

Step 2: Examine Results, Check Assumptions, Summarize Data via Test Statistic





Histogram and Normal Q-Q Plot suggest modest deviations from **normality**

Note: reasonable sample size + CLT … normality assumption not so crucial

Forward School

Step 2: Examine Results, Check Assumptions, Summarize Data via Test Statistic

H_0: $\mu = 80$

H_a: $\mu > 80$

Is sample mean of 82.48 inches *significantly greater* than hypothesized mean of 80 inches?

standard error of the sample mean $= \dfrac{\sigma}{\sqrt{n}}$

estimated standard error of the sample mean $= \dfrac{s}{\sqrt{n}}$

Forward School

Step 2: Examine Results, Check Assumptions, Summarize Data via Test Statistic

Test statistic: Assuming sampling distribution of sample mean is normal

$$t = \frac{best\ estimate - null\ value}{estimated\ standard\ error} = \frac{\bar{x} - 80}{\frac{s}{\sqrt{n}}}$$

$$= \frac{82.48 - 80}{\frac{15.06}{\sqrt{25}}} = \frac{2.48}{3.012} = \mathbf{0.82}$$

Our **sample** mean is only **0.82** (estimated) **standard errors above null value of 80 inches**

Forward School

Step 3: Determine p-value

- If null hypothesis was true, would a test statistic value of only t=0.82 be unusual enough to reject the null?
- p-value = probability of seeing test statistic of 0.82 or more extreme assuming the null hypothesis is true
- If null hypothesis was true, t statistic follows a Student t Distribution with degree freedom n-1 = 25-1 = 24
- One tailed test to the right →upper tail

Forward School

Z-test for single mean

- $$z = \frac{\bar{x} - \mu_0}{\underbrace{\frac{\sigma}{\sqrt{n}}}_{\text{Standard error of the mean (standard deviation of the sampling distribution)}}}$$

$$z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}}$$

- $\bar{x}$ = sample mean
- μ_0 = hypothesized population mean
- σ = population standard deviation (given)
- n = sample size

Forward School

Test Statistic

- Data summary used to evaluate the two hypotheses is called the **test statistic.**
- How extreme is the sample value from the null population value?

$$\text{Test statistic} = \frac{\text{sample value - null value}}{\text{standard deviation of sampling distribution}}$$

The Test Statistic For Testing a Single Mean

- The appropriate test statistic is a z-statistic if population SD, σ , is known:

$$\text{z-statistic} = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$$

- If σ unknown, then we use s to estimate σ and a t-statistic is used:

$$\text{t-statistic} = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$$

- t-statistic is also used for paired observation

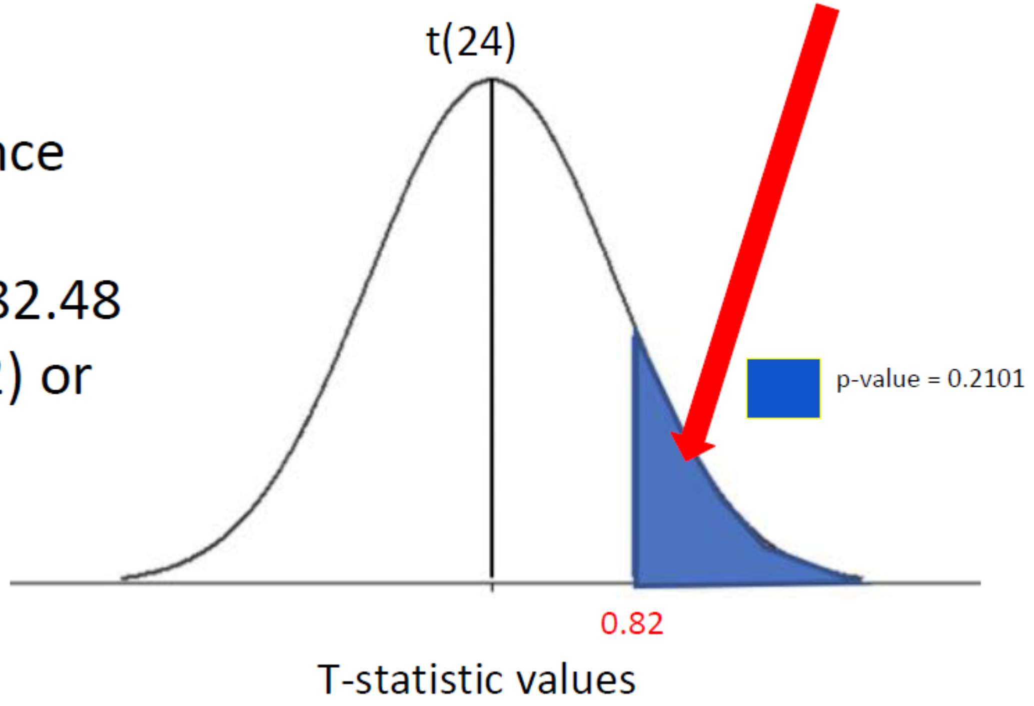
Forward School

Step 3: Determine p-value

P-value = 0.21

If population mean CW distance was really 80 inches, then observing a sample mean of 82.48 inches (i.e. a t statistic of 0.82) or larger is **quite likely**

```
from scipy.stats import t
1- t.cdf(0.82, 24)
> 0.2101
```



Forward School

Step 4: Make decision about the null

Since our p-value, 0.2101 is much bigger than 0.05 significance level, weak evidence against the null

→ **we fail to reject the null**

Based on estimated mean (82.48 inches), we **cannot support** the population mean CW distance is greater than 80 inches

Forward School

90% confidence interval estimate

Mean = 82.48 inches
Standard Deviation = 15.08 inches
n = 25 observation, t* →1.711

$$\bar{x} \pm t^{*}\left(\frac{s}{\sqrt{n}}\right)$$

$$82.48 \pm 1.711\left(\frac{15.06}{25}\right)$$

$$82.48 \pm 5.15$$

$$\boxed{(77.33\ inches, 87.63\ inches)}$$

Note: 80 inches is **<u>IN</u>** confidence interval of reasonable values for population mean CW distance

Forward School

Procedures of conducting the tests of hypothesis using critical region

1) Formulate hypothesis
   H_0:$\mu = \mu_0$ vs H_a:$\mu \neq \mu_0$ or H_a: $\mu > \mu_0$ or H_a: $\mu < \mu_0$
2) Specify α (usually is 5%, unless mention otherwise)
3) Determine a critical region of size α
4) Compute the value of the **test statistic** from the sample data
5) Conclusions : Reject Ho if the **test statistic** has a value in the **critical region**

Forward School

Summary

Hypothesis Tests are used to put theories about a parameter of interest to the test ~ parameter = population mean

Basic steps:

- State hypothesis (and select significance level)
- Examine results, check assumptions, summarize via test statistic
- Convert test statistic to p-value
- Compare p-value to significance level to make decision

Assumptions for one-sample mean (t) Test for population mean:

- Data considered a random sample
- Population of responses is normal (else large n helps)

Know how to interpret the p-value, decision and conclusion.

Forward School

Example

If five pieces randomly selected ribbon have a mean breaking strengths of 183.14 and standard deviation of 8.219 pounds, find the set of value of 0 that would fail to reject the null hypothesis H_0: μ = 0 in two sided 5% student t-test. ɑ= 0.05

$$t = \frac{\bar{x} - \mu_0}{\frac{s_d}{\sqrt{n}}}$$

Forward School

Example

1. Calculate the t-statistic

```
from scipy.stats import t
t.ppf(0.975, 4)
> 2.776
```

$$t = \frac{\bar{x} - \mu_0}{\frac{s_d}{\sqrt{n}}} \Rightarrow \mu_0 = \bar{x} - t(\frac{s}{\sqrt{n}})$$

```
183.14 + np.array([-1,+1]) * t.ppf
(0.975,4)*(8.219/np.sqrt(5))

> array([172.9347636, 193.3452364])
```

Reject H_0 if t-statistic < -2.78 or t-statistic > 2.78.

Fail to reject H_0 if t-statistic is between -2.78 to 2.78

Forward School

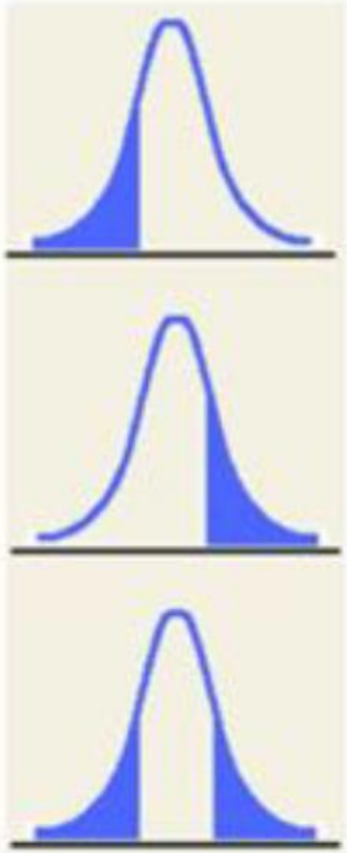
Procedures of conducting the tests of hypothesis using p-value

- p-value:
- The probability (or Likelihood) of observing a *test statistic* as extreme as what we did, or something even more extreme, if the null hypothesis is true

Procedures of conducting the tests of hypothesis using p-value

1) Formulate hypothesis
   $H_0 : \mu = \mu_0$ vs $H_a : \mu \neq \mu_0$ or $H_a : \mu > \mu_0$ or $H_a : \mu < \mu_0$
2) Specify α (usually is 5%, unless mention otherwise)
3) Compute the value of the **test statistic** from the sample data
4) Compute p-value
5) Reject H_0 if p-value $\leq \alpha$

How to find p-value



For $H_a: \mu < \mu_0$ (1-sided, *less than*)
the p-value is the less than area

For $H_a: \mu > \mu_0$ (1-sided, *greater than*)
the p-value is the more than area

For $H_a: \mu \neq \mu_0$ (2-sided, *different to*)
the p-value is 2×the area in tail

Forward School

Example

The specification for a certain kind of ribbon should have a mean breaking strength of 185 pounds. If five pieces randomly selected ribbon have a mean breaking strengths of 183.14 and standard deviation of 8.,219 pounds. Find the p-value for a two-sided student t-test.

Forward School

Example

The specification for a certain kind of ribbon should have a mean breaking strength of 185 pounds. If five pieces randomly selected ribbon have a mean breaking strengths of 183.14 and standard deviation of 8.219 pounds. Find the p-value for a two-sided student t-test.

$$H_0: \mu = 185 \text{ vs } H_a: \mu \neq 185$$

$$t = \frac{\bar{x} - \mu_0}{\frac{s_d}{\sqrt{n}}} \qquad t = \frac{183.14 - 185}{\frac{8.219}{\sqrt{5}}} = -0.51$$

P-value = 2P(T<0.51)

```
from scipy.stats import t

2*(t.cdf(-0.51,4))
> 0.6369
```

p-value = 0.64 > 0.05, α significance level, we **fail to reject** the NULL hypothesis

Forward School

Difference in Means for Paired Data

Forward School

Home Renovation

20 homes remodelling their kitchens, requesting cabinet quotes from 2 suppliers

Is there an **average** different in cabinet quotes from these two suppliers?

Variables: Difference in supplier quotes (Supplier A - Supplier B)

Forward School

Research Question

Is there an **average** different in cabinet quotes from these two suppliers?

Population : All houses

Parameter of Interest: Population mean difference of cabinets quotes, μ_d (Supplier A - Supplier B)

Test for a significant mean difference in cabinet quotes at the 5% significance level

Forward School

Hypothesis

$H_0: \mu_d = 0$

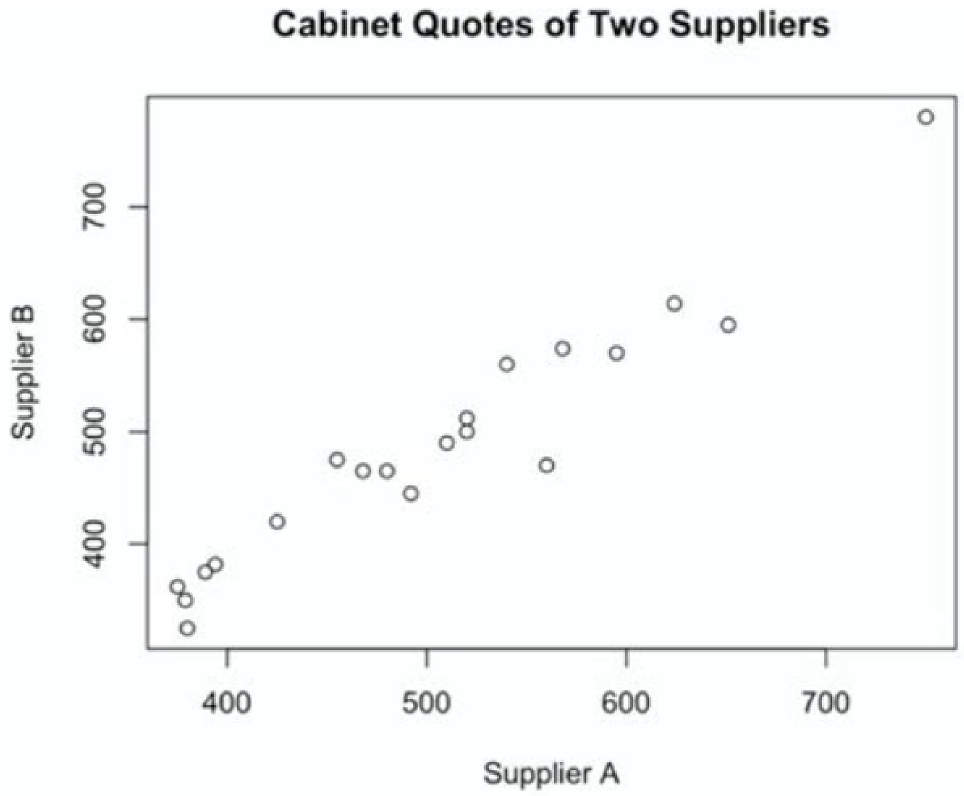
$H_a: \mu_d \neq 0$

$\alpha = 0.05$

Forward School

Cabinet Data

Supplier A	Supplier B	Difference
380	325	55
560	470	90
425	420	5
389	375	14
568	574	-6
651	595	56

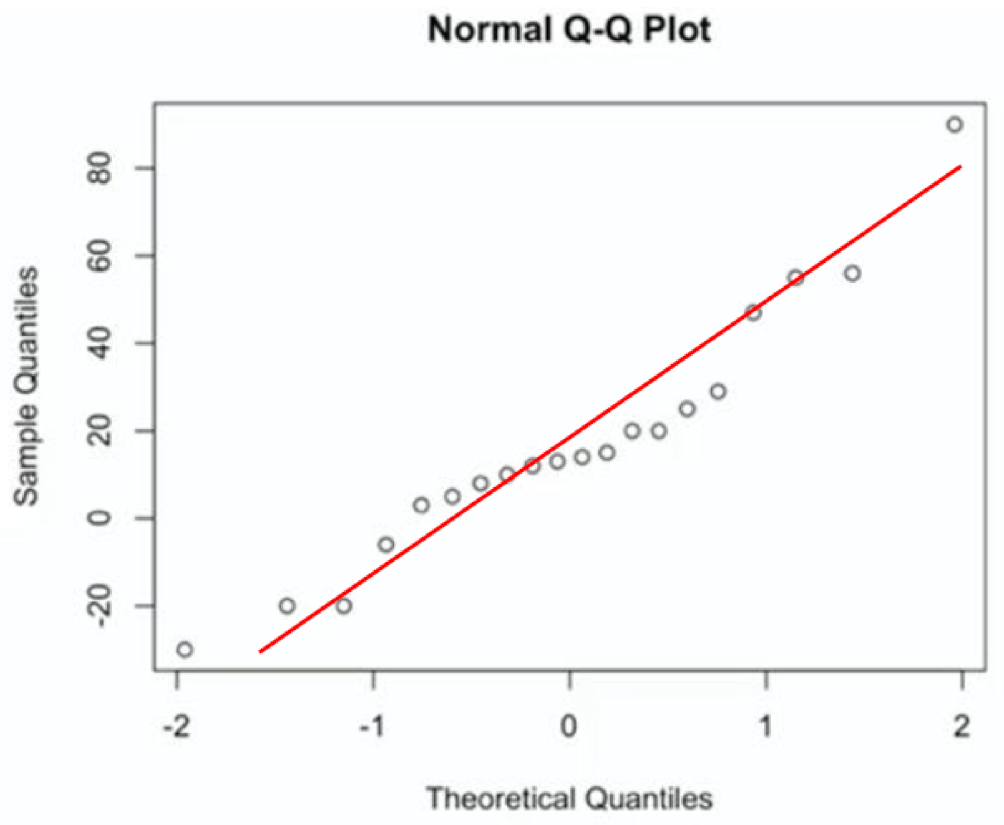


Forward School

Assumptions

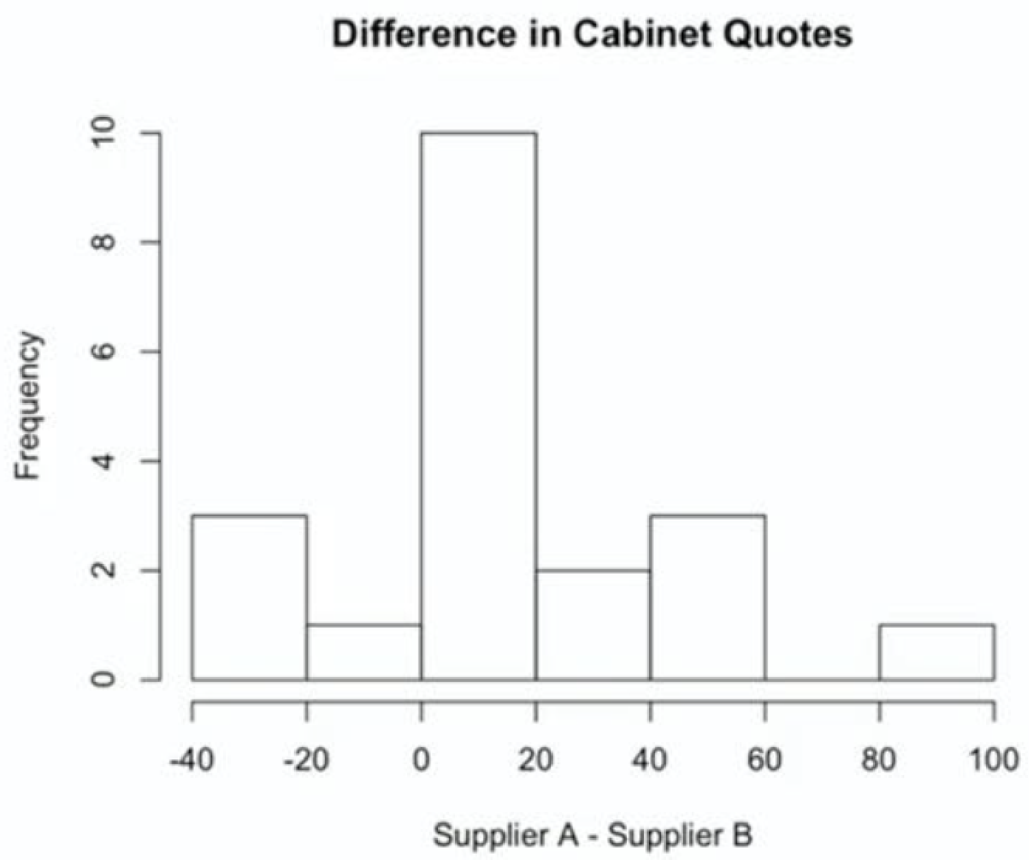
We need to assume that we have a **random sample of differences** i.e random sample of houses

Population of differences to be normally distributed or we have large enough sample size (about > 25)



Forward School

Summarize the data



n = 20 observations

minimum = -30

maximum = 90

median = 13.50

mean = 17.30

Standard Deviation = 28.49

Forward School

Test Statistic

Assuming the sampling distribution of the sample mean difference is normal

$$t = \frac{\text{best estimate - hypothesized mean}}{\text{estimated standard error of estimate}}$$

Forward School

Test Statistic

t = $\dfrac{\text{best estimate - hypothesized mean}}{\text{estimated standard error of estimate}}$

n = 20 observations

mean = 17.30

Standard Deviation = 28.49

Our observed mean difference is 2.72 (estimated) standard errors above our null value of 0

$$t = \frac{\bar{x}_d - 0}{\frac{s_d}{\sqrt{n}}}$$

$$t = \frac{17.30 - 0}{\frac{28.49}{\sqrt{20}}}$$

$$= 2.72$$

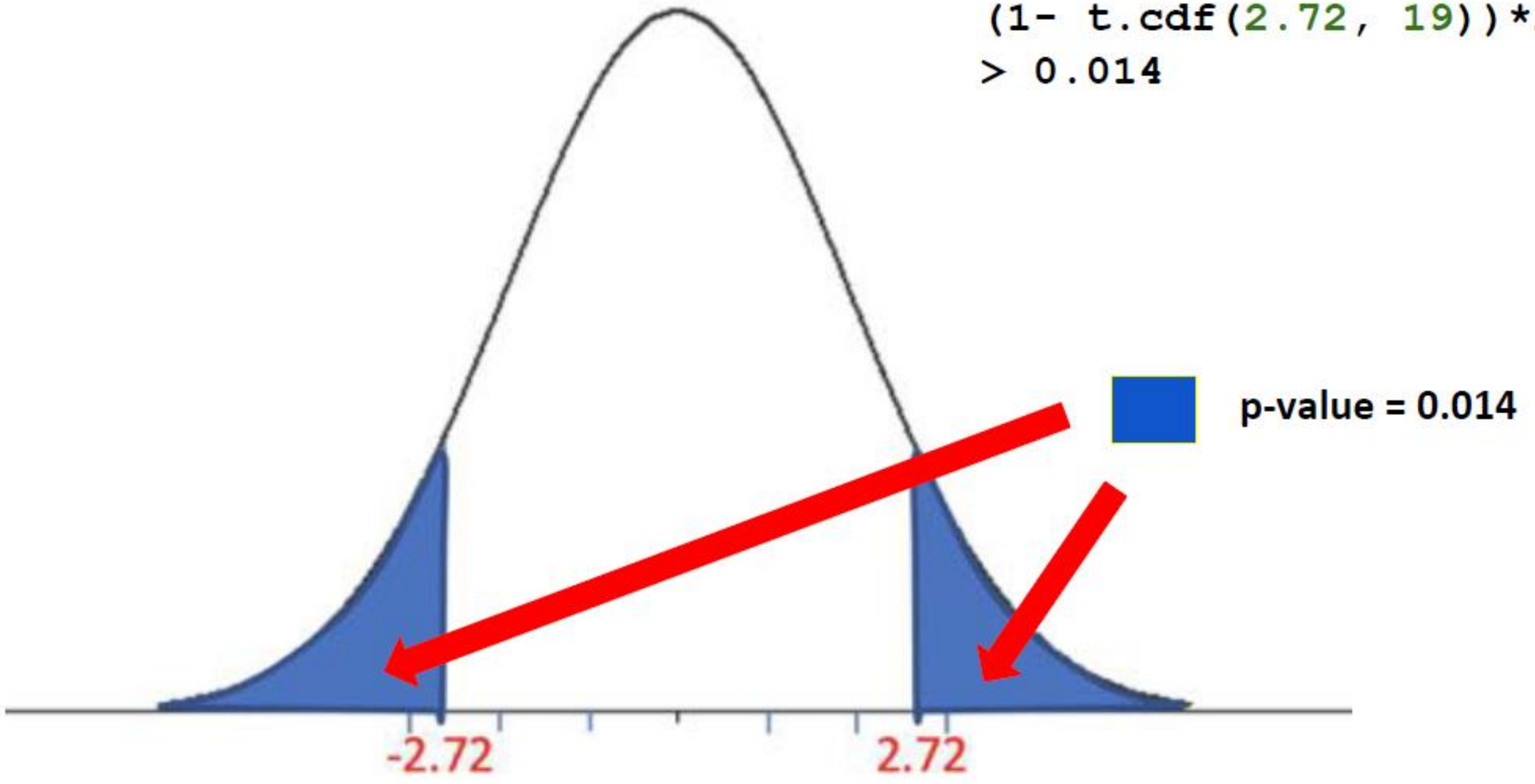
Forward School

Test Statistic Distribution & P-Value

Ha: $\mu_d \neq 0$

t = 19

```
from scipy.stats import t
(1- t.cdf(2.72, 19))*2
> 0.014
```



Forward School

Decision & Conclusion

p-value = 0.014 < 0.05 = $\alpha \rightarrow$ reject null hypothesis

$\rightarrow$have evidence against mean difference in cabinet quotes = 0

Formally, based on our sample and our p-value, we **reject** the null hypothesis. We conclude that the mean difference of cabinet quote prices for Suppliers A less B is **significantly different** from 0

Forward School

95% Confidence Interval

Mean =17.30
Standard Deviation = 28.49
n = 20 observation, t* →2.093

Note: 0 is **NOT** in the range of reasonable values for mean difference in cabinet prices

$$\bar{x}_d \pm t^* \left(\frac{s_d}{\sqrt{n}}\right)$$

$$17.30 \pm 2.093 \left(\frac{28.49}{\sqrt{20}}\right)$$

$$17.30 \pm 13.33$$

$$\boxed{(3.97, 30.63)}$$

Forward School

Summary

Hypothesis Test allows you to assess theories about a population parameter of interest

- parameter of interest = mean difference

Extension of one mean hypothesis test

- with difference variable
- collected on same individual (houses)

Forward School

Difference in Means for Independent Groups

Forward School

Research Question

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

Population: Mexican-American adults (age 18-29) in the US

Parameter of Interest: μ_a - μ_b - Body Mass Index (kg/m^2)

Task: Perform an independent sample t-test regarding the value of the difference in mean BMI between males and females

Forward School

Steps to Perform a Hypothesis Test

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

Steps 1: Define Hypothesis

Null: There is no difference in mean BMI

Alternate: There is a significant difference in mean BMI

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

$$H_a: \mu_1 \neq \mu_2 \quad (\text{or } H_0: \mu_1 - \mu_2 \neq 0)$$

Significant level, α = 0.05

Steps 2: Examine the data

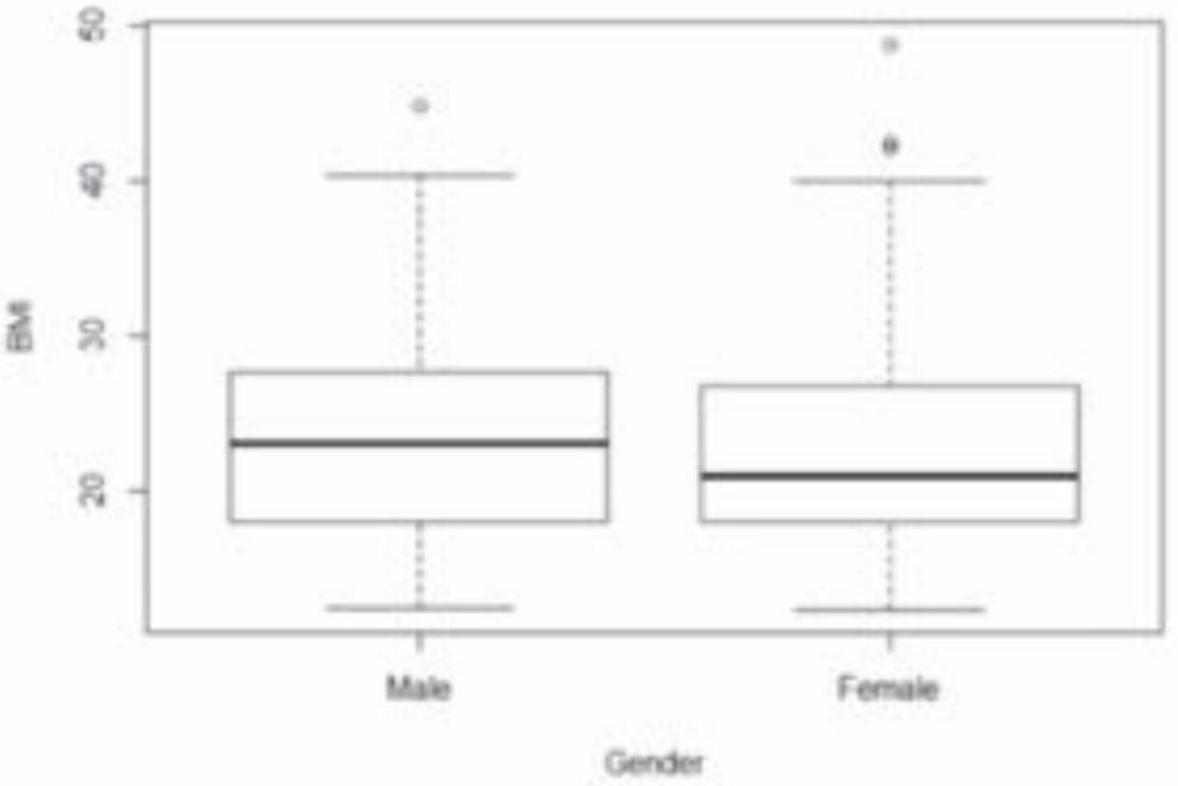
Gender	BMI	Race	Age 18-29
1	19.9	1	1
2	17.0	1	1
2	26.7	1	1
1	25.6	1	1
...	...	...	...

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

Forward School

Steps 2: Examine the 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



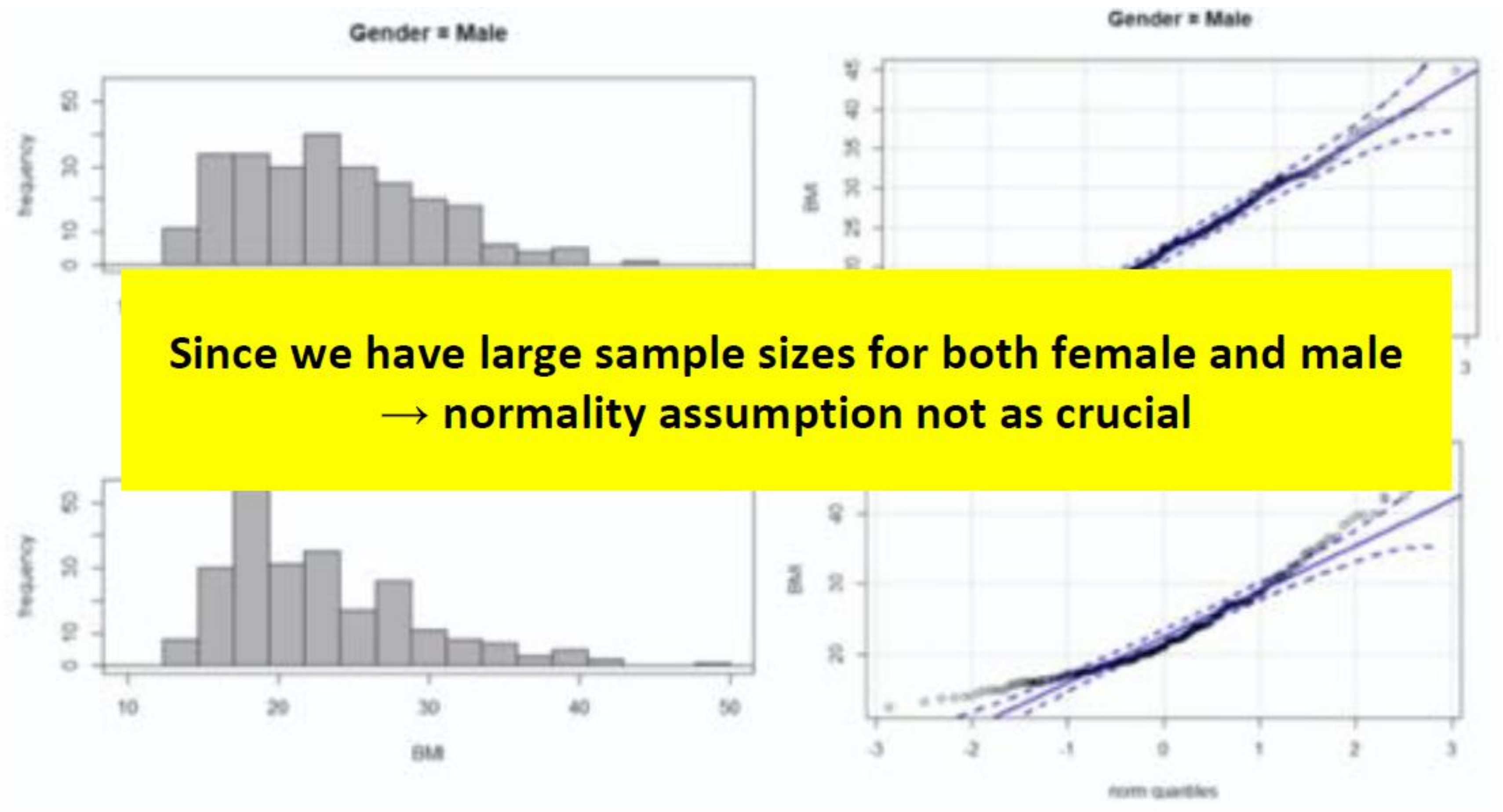
Forward School

Steps 2: Check Assumptions

- Samples are considered simple random samples
- Samples are independent from one another
- Both populations responses are approximately normal (large enough sample size)

Forward School

Steps 2: Check Assumptions



Since we have large sample sizes for both female and male
→ normality assumption not as crucial

Forward School

Steps 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/m^2 significantly different from 0?

Forward School

Steps 2: Calculate test statistic

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

Considerations:

Pooled Approach - The variance of two populations are assumed to be equal $(\sigma_1^2 = \sigma_2^2)$

Unpooled Approach - The assumption of equal variances is dropped

Forward School

Steps 2: Calculate test statistic

Pooled Approach

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

Forward School

Steps 2: Calculate test statistic

Unpooled Approach

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

Forward School

Two samples test for difference between 2 means (μ_a-μ_b)

$H_0: \mu_a - \mu_b = \mu_0$ vs $H_a: \mu_a - \mu_b \neq \mu_0$ or $H_a: \mu_a - \mu_b > \mu_0$ or $H_a: \mu_a - \mu_b < \mu_0$

Test statistic:

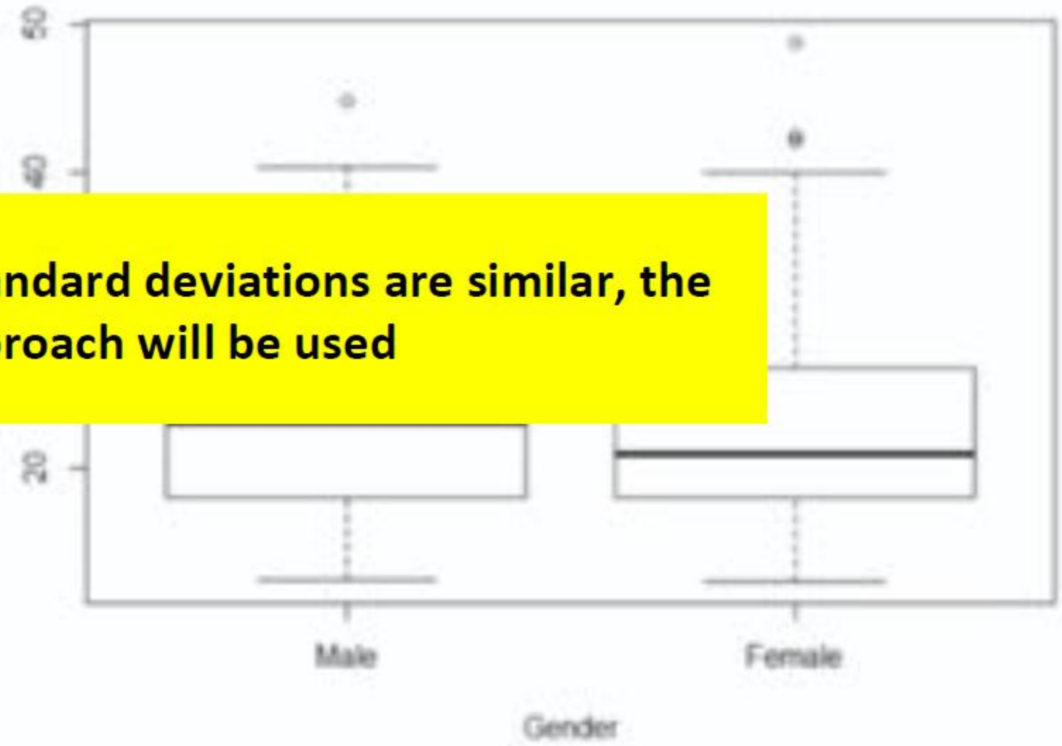
Common variance: $$t = \frac{(\overline{x}_a - \overline{x}_b) - (\mu_a - \mu_b)}{S_p \sqrt{\frac{1}{n_a} + \frac{1}{n_b}}}$$

Unequal variance: $$t = \frac{(\overline{x}_a - \overline{x}_b) - (\mu_a - \mu_b)}{\sqrt{\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}}}$$

Forward School

Pooled or Unpooled?

	Male	Female
Mean	23.57	22.83
St. Dev.	[illegible]	[illegible]
Min	[illegible]	[illegible]
Max	44.9	48.8
n	258	239



Because the IQR's and standard deviations are similar, the pooled approach will be used

Step 2: Calculate Test Statistic?

Pooled Approach

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

Our difference in sample mean is only 1.30 (estimated) standard errors above null difference of 0 kg/m^2

$$t = 1.30$$

Forward School

Step 3: Determine p-value

$t = 1.30$

If the null hypothesis (μ_1-μ_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

Forward School

Step 3: Determine p-value

$t = 1.30$

The degree of freedom, df $= n_1 + n_2 - 2 = 258 + 239 - 2 = 495$

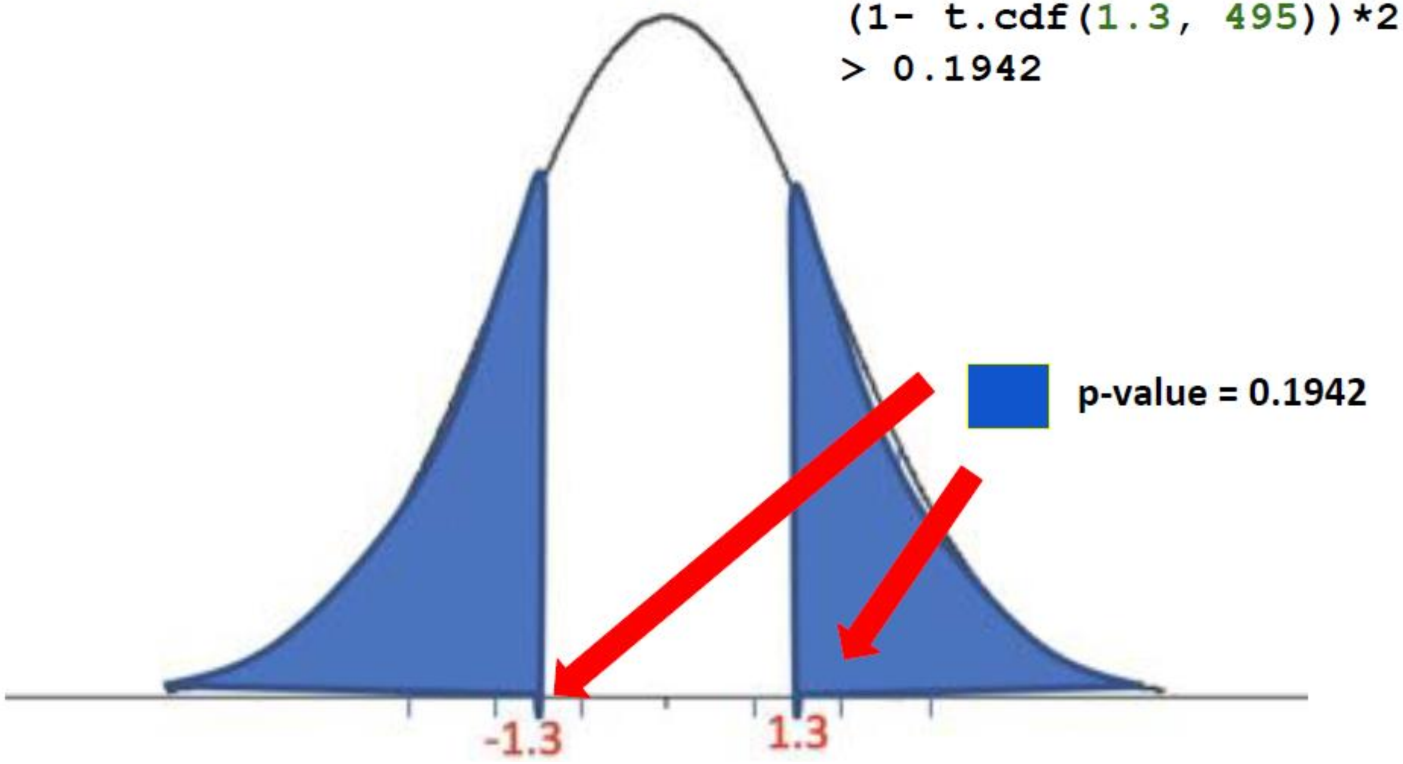
Our alternate hypothesis is two-sides ($\mu_1 - \mu_2 \neq 0$), so we will check for upper and lower tails

Step 3: Determine p-value

Ha: $\mu_1 - \mu_2 \neq 0$

t = 495

```
from scipy.stats import t
(1- t.cdf(1.3, 495))*2
> 0.1942
```



Forward School

Step 3: Determine p-value

$p = 0.19 > 0.05 =$ significance value

if the difference in population mean BMI between males and females was really 0 kg/m^2,

then observing a difference in sample means of 0.74 kg/m^2 (i.e. test statistic of 1.30) or more extreme is **fairly likely**

Step 4: Make Decision

$p = 0.19 > 0.05 =$ significance value

→ **we fail to reject null!**

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

Forward School