A New Week

Unit 8 – Hypothesis Testing Day 3 and 4

Your Don’t-Iguana-Talk-About-It Professor Colton

Unit 8 Second Half - Outline

Unit 8 – Hypothesis Testing

8.4 Hypothesis Testing for Population Means (with known population standard deviation)

- Determine Ho and Ha
- Traditional Method: Critical Value Z, Test Statistic, Conclusion
- P-value Method

8.5 Hypothesis Testing for Population Means (with unknown sigma)

- Determine Ho and Ha
- Traditional Method: Critical Value t, Test Statistic, Conclusion
- Examples
- P-value Method

8.2 Hypothesis Testing Overview
- Type I and Type II Errors

P-Values

More Practice!

Hypothesis Test Steps – This is Your Life Now…

1. **State** the Hypotheses
   - Define parameter + context.

2. **Check** Assumptions.

3. **Determine** and **Sketch** Rejection Region based of Significance Level

4. **Compute** value of Test Statistic / P-value.

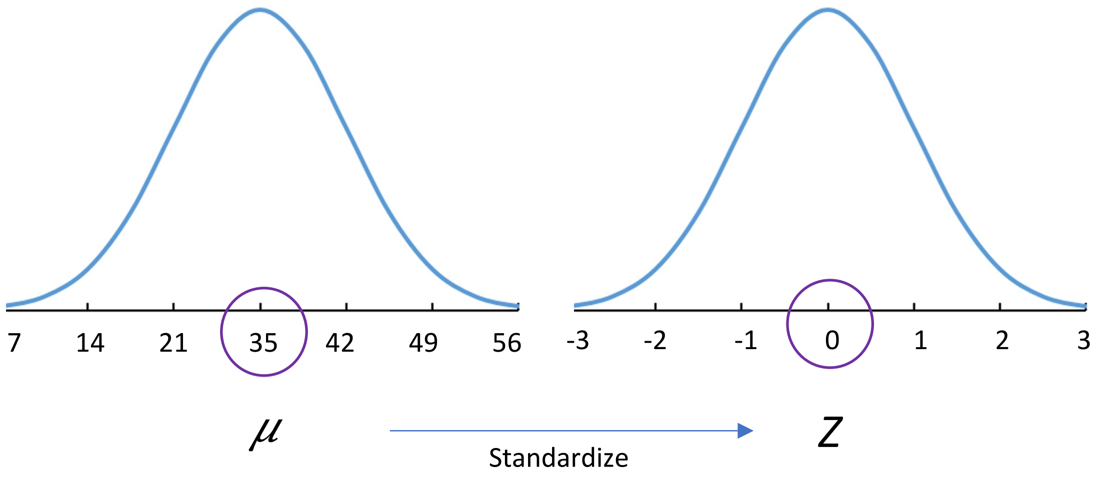
5. **Conclude** and **Interpret**
   - State whether you reject H_0 or fail to reject H_0 AND WHY!
   - Interpret your results in the context of the problem

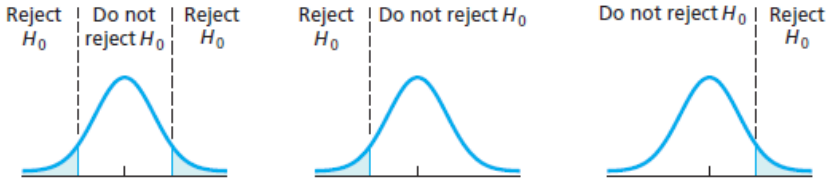
Hypothesis Tests for Means with KNOWN $\boldsymbol{\sigma}$!

- All of the previous Hypothesis tests overview applies, now we are just going to apply it specifically to a Means Test!
- And going back to the Confidence Interval unit, we have a known population standard deviation! So the same logic and implications of that apply here as well!
- We will be doing a **Z-Test**!

The Hypothesis Statements - Review

1. State the Hypotheses
   - **Define parameter + context.**





Type of Test:	(a) Two tailed	(b) Left tailed	(c) Right tailed
Sign in H_A:	$\neq$	$<$	$>$
Critical Values:	$-Z^*$ $+Z^*$	$-Z^*$	Z^*

Define Parameter

- Always define your **parameter** at the start!
- Think about the variable / quantity of interest!
  - Quantitative (numeric) → population mean μ

Null Hypothesis H_0

- This is the status quo, typically a *known* value of the parameter (μ_0)
  - **When written symbolically ALWAYS =**

Alternative Hypothesis H_A

- May be left-tailed (<), right-tailed (>), or two-tailed (≠).
  - Uses the same value of the parameter as in the Null hypothesis H_0

In general

- H_0: $\mu = \mu_0$ and H_A: $\mu \neq,<,> \mu_0$

Examples:

'Research from previous studies suggests the average number of people

- Equal to → H_0: $\mu = 7$

'The owner believes his average monthly profit is more than $50,000'

- In this case, greater than → H_A: $\mu > 50{,}000$

LCQ – Hypotheses

Problem: (1) Define the parameter of interest and (2) State the Null and Alternative Hypotheses and the directionality of the test (two-tailed, left-tailed or right-tailed) for the following scenarios:

a) A random sample of 15 human body temperatures were obtained. Assume that human body temperatures are normally distributed. Is there sufficient evidence to conclude that the true mean human body temperature **differs** from 98.6^{o}F?

b) In 2012, a large number of foreclosed homes in Washington, D.C. were sold. Real estate experts say the standard deviation for sales the past 10 years was $190,000. In one community, a random sample of 30 foreclosed homes sold for an average of $443,705. A prospective home-buyer wants to know if prices have **decreased** from the 2002 average of $450,000.

c) Test 1 grades on the most fun class you've ever taken averaged 80.76 with standard deviation 13.34 points. From a random sample of 19 Test 2 grades, there was a mean of 83.39. Your super cool instructor wants to know if the Test 2 grades **improved**.

LCQ – Hypotheses

Problem: (1) Define the parameter of interest and (2) State the Null and Alternative Hypotheses and the directionality of the test (two-tailed, left-tailed or right-tailed) for the following scenarios:

a) A random sample of 15 human body temperatures were obtained. Assume that human body temperatures are normally distributed. Is there sufficient evidence to conclude that the true mean human body temperature **differs** from 98.6^{o}F?

Let μ = the true mean of human body temperature → VERY GOOD!

First try:
$H_0 = 98.6$ → INCORRECT! MISSING μ!!!
$H_A \neq 98.6$

H_0: $\mu = 98.6$ and H_A: $\mu \neq 98.6$ → two-tailed → NOW IT'S PERFECT!

b) In 2012, a large number of foreclosed homes in Washington, D.C. were sold. Real estate experts say the standard deviation for sales the past 10 years was $190,000. In one community, a random sample of 30 foreclosed homes sold for an average of $443,705. A prospective home-buyer wants to know if prices have **decreased** from the 2002 average of $450,000.

First try:
H_0: $p = 450{,}000$ and H_A: $p < 450{,}000$ → INCORRECT!! NOT TALKING ABOUT PROPORTIONS!!! SHOULD BE NO Ps

Let μ = the true mean of the prices of foreclosed homes → PERFECT

H_0: $\mu = 450{,}000$ and H_A: $\mu < 450{,}000$ left-tailed → CORRECT!

c) Test 1 grades on the most fun class you've ever taken averaged 80.76 with standard deviation 13.34 points. From a random sample of 19 Test 2 grades, there was a mean of 83.39 Your super cool instructor wants to know if the Test 2 grades **improved**.

Both correct (I won't be too picky with the context, technically we are looking at our Test 2 grade average and comparing that to a value from the Test 1 (the null):
Let μ be the true mean of the test 2 grades.
Let μ = represent the true mean of test scores

H_0: $\mu = 80.76$ and H_A: $\mu > 8.076$ right-tailed → CORRECT!

Means Assumptions

2. Check Assumptions.

- EXACT same Assumptions as for Confidence Intervals that were based on the sampling distribution of $\bar{x}$ and the CLT!
- (Remember we need to know / be given the population standard deviation $\boldsymbol{\sigma}$)

Hypothesis Test Conditions

✓Randomization Condition

Need to have a random sample

✓Large Enough Sample Condition

Normal population OR

$n \geq 30$

LCQ – Assumptions

*** Note: If only a sample standard deviation is provided, then σ is unknown*

Problem: Check the conditions for a Hypothesis Test of the population mean for the following scenarios:

a) A random sample of 15 human body temperatures were obtained. Assume that human body temperatures are **normally distributed**. Is there sufficient evidence to conclude that the true mean human body temperature differs from 98.6°F?

b) In 2012, a large number of foreclosed homes in Washington, D.C. Real estate experts say the standard deviation for sales the past 10 years was $190,000. In one community, a random sample of 30 foreclosed homes sold for an average of $443,705. A prospective home-buyer wants to know if prices have decreased from the 2002 average of $450,000.

c) Test 1 grades on the most fun class you've ever taken averaged 80.76 with standard deviation 13.34 points. From a random sample of 19 Test 2 grades, there was a mean of 83.39. Your super cool instructor wants to know if the Test 2 grades improved.

LCQ – Assumptions

*** Note: If only a sample standard deviation is provided, then σ is unknown*

Problem: Check the conditions for a Hypothesis Test of the population mean for the following scenarios:

a) A random sample of 15 human body temperatures were obtained. Assume that human body temperatures are **normally distributed**. Is there sufficient evidence to conclude that the true mean human body temperature differs from 98.6°F?

INCORRECT to try and check $np \geq 5$ and $nq \geq 5$ because there is no success and failures anymore! We have a quantitative variable now!!

Random sample and normal distribution so conditions are both met! → VERY GOOD! We have a normal population, so don't have to look at the sample size for the condition!

b) In 2012, a large number of foreclosed homes in Washington, D.C. Real estate experts say the standard deviation for sales the past 10 years was $190,000. In one community, a random sample of 30 foreclosed homes sold for an average of $443,705. A prospective home-buyer wants to know if prices have decreased from the 2002 average of $450,000.

Random sample (YES) and normal distribution(???) → NOPE!! No mention of a normal population
So how do you know is normal distribution? We don't → have to look at n for the large enough sample condition
$n = 30 \geq 30$ → YES!!! So it still meets the large enough sample size condition
Both conditions are met! → notice that we had to do this differently than part (a)

c) Test 1 grades on the most fun class you've ever taken averaged 80.76 with standard deviation 13.34 points. From a random sample of 19 Test 2 grades, there was a mean of 83.39. Your super cool instructor wants to know if the Test 2 grades improved.

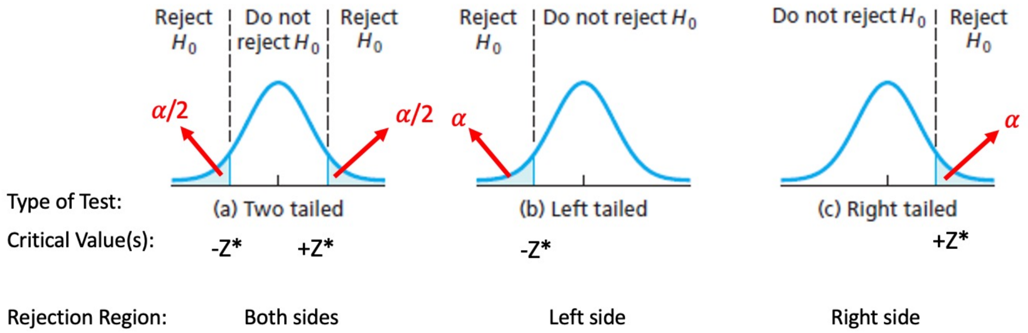
Random sample, YES!
Large enough sample??? → Does not say we have normally distributed population, and the sample size is only $n = 19$. So are the conditions met????
We don't have enough information to make a final conclusion → Would have to look at the distribution of the sample (symmetric / skewed, outliers, etc.) and make a judgement call if $n = 19$ is large enough so that the CLT results are applicable.

Rejection Region for Means with Known $\boldsymbol{\sigma}$

3. Determine and Sketch Rejection Region based of Significance Level

Rejection Region for Means Test with KNOWN $\boldsymbol{\sigma}$

- This is the SAME as for a Proportions Test!



Review

- We have to determine the the when there is or is not enough evidence against the Null.
- Our Rejection Region (RR) is based on whether we are doing a one or two tailed test (this is the direction from the H_A)!
- Critical Values that define the RR are based on the standard normal (Z) curve
  - So all of our CVs will be Z*s!
- Using calc:

Left-Tailed: Z* = invNorm(area =$\alpha, \mu = 0, \sigma = 1$)
Right-Tailed: Z* = invNorm(area = $1 - \alpha, \mu = 0, \sigma = 1$)
Two-Tailed: Z* = invNorm(area = $\frac{\alpha}{2}, \mu = 0, \sigma = 1$)

Using Calc - Test Statistic and P-Value for Means with Known σ

4. Compute value of Test Statistic / P-value.

Formula for Z_{stat} by hand:

Test for	H_0	Test statistic
Pop. mean μ	$\mu = \mu_0$	$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$

(Original) Setup
Scientists discovered a new mountain range under the sea. Lets assume the sea mountain heights are normally distributed with known standard deviation of 5000 ft.

From a random sample of 13 peaks, there was an average height of 12,000 ft. Is there enough evidence to conclude the average heights of these new sea mountains is different than the Rocky Mountains, which average 14,400 ft? Use α = 0.12

GOAL: Conduct a Hypothesis Test!

1. Z-Test
   a) Input = Stats
   b) μ_0 = the Null mean
   c) σ = population SD
   d) $\bar{x}$ = sample mean
   e) n = sample size
   f) μ: Alternative hypothesis

   Calculate or Draw

New Scenario
Three more sea mountains were discovered. The new sample mean is equal to 15,000 ft. Is there enough evidence to say the sea mountains are taller than the Rockies. Use α = 0.10

- Run another ZTest

Using Calc - Test Statistic and P-Value for Means with Known σ

4. Compute value of Test Statistic / P-value.

(Original) Setup
Scientists discovered a new mountain range under the sea. Lets assume the sea mountain heights are normally distributed with known standard deviation of 5000 ft.

From a random sample of 13 peaks, there was an average height of 12000 ft. Is there enough evidence to conclude the average heights of these new sea mountains is different than the Rocky Mountains, which average 14,400 ft? Use α = 0.12

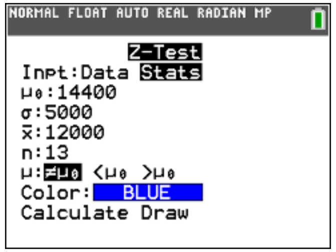
Formula for Z_{stat} by hand:

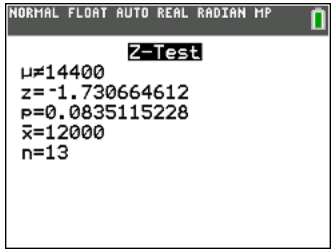
Test for	H_0	Test statistic
Pop. mean μ	$\mu = \mu_0$	$z = \dfrac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$

GOAL: Conduct a Hypothesis Test!

1. Z-Test
   a) Input = Stats
   b) μ_0 = the Null mean
   c) σ = population SD
   d) $\bar{x}$ = sample mean
   e) n = sample size
   f) μ: Alternative hypothesis

   Calculate or Draw

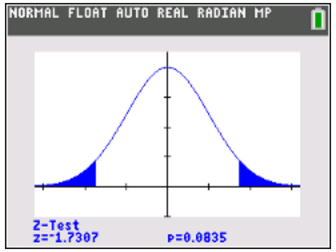




H_0: μ = 14,400
H_A: μ ≠ 14,400

Calculate Output
μ = Alternative hypothesis
z = Z_{stat}
p = p-value
$\bar{x}$ = sample proportion
n = sample size

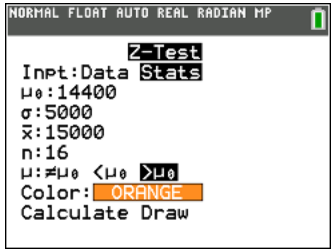
Again, this does NOT give us the Critical Value Z, we have to figure that out ourselves

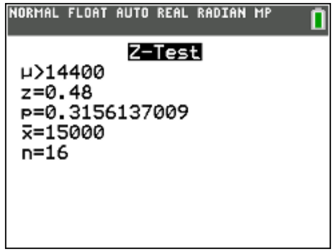


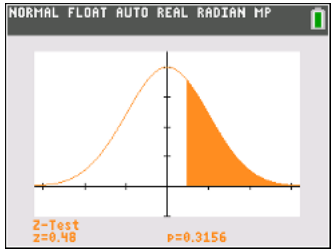
Draw Output
Plot (and displays values) of p = p-value and z = Z_{stat} on the standard normal curve

New Scenario
Three more sea mountains were discovered. The new sample mean is equal to 15,000 ft. Is there enough evidence to say the sea mountains are taller than the Rockies. Use α = 0.10
- Run another ZTest

H_0: μ = 14,400
H_A: μ > 14,400







LCQ – Conclusions and Interpretations

5. Conclude and Interpret

- State whether you reject H_0 or fail to reject H_0 AND WHY!
- Interpret your results in the context of the problem

Problem: Write the conclusions and interpretations for the previous scenarios using our results.

a) Use the Traditional Method

Original Setup: Scientists discovered a new mountain range under the sea. Lets assume the sea mountain heights are normally distributed with known standard deviation of 5000 ft. From a random sample of 13 peaks, there was an average height of 12000 ft. Is there enough evidence to conclude the average heights of these new sea mountains is different than the Rocky Mountains, which average 14,400 ft? Use α = 0.12

b) Use the P-Value Method

New Scenario: Three more sea mountains were discovered. The new sample mean is equal to 15,000 ft. Is there enough evidence to say the sea mountains are taller than the Rockies. Use α = 0.10

LCQ – Conclusions and Interpretations

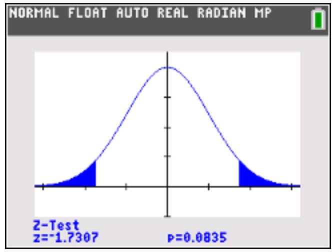
5. **Conclude** and **Interpret**

- State whether you reject H_0 or fail to reject H_0 AND WHY!
- Interpret your results in the context of the problem

Problem: Use the specified method to write the conclusions and interpretations for the previous scenarios using our results.

a) Use the Traditional Method

Original Setup: Scientists discovered a new mountain range under the sea. Lets assume the sea mountain heights are normally distribued with known standard deviation of 5000 ft. From a random sample of 13 peaks, there was an average height of 12000 ft. Is there enough evidence to conclude the average heights of these new sea mountains is different than the Rocky Mountains, which average 14,400 ft? Use α = 0.12



Need these…

H_0: μ = 14,400
H_A: μ ≠ 14,400
α = 0.12

Test Statistic

Traditional method, so have to find the critical value first based on the significance level and alternative hypothesis

Z = invNorm(area = 0.12/2, μ = 0, $\boldsymbol{\sigma}$ = 1) = -1.555 → two-tailed test, so will compare the absolute values of the TS and Z**

Z_{stat} = ZTest(μ_0 = 14400, $\boldsymbol{\sigma}$ = 5000, $\bar{x}$ = 12000, n = 13, $\mu \neq \mu_0$) = -1.7307

$|Z_{stat}|$ = |-1.7307| = 1.7307 > 1.555 = |-1.555| = |Z| → Reject H_0!*

Conclusion and Interpretation

Because the absolute value of our Test Statistic Z_{stat} = 1.7307 is greater than the absolute value of our Critical Value Z = 1.555, we reject the Null hypothesis. There IS sufficient evidence to conclude that the true mean height of the sea mountains are different than 14,400 ft, which is the average height of the Rocky Mountains.*

LCQ – Conclusions and Interpretations Cont…

5. Conclude and Interpret

- State whether you reject H_0 or fail to reject H_0 AND WHY!
- Interpret your results in the context of the problem

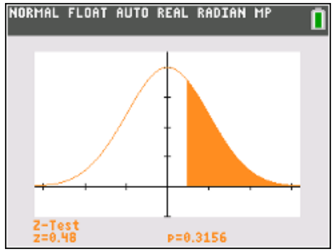
Problem: Use the specified method to write the conclusions and interpretations for the previous scenarios using our results.

b) Use the P-Value Method

New Scenario: Three more sea mountains were discovered. The new sample mean is equal to 15,000 ft. Is there enough evidence to say the sea mountains are taller than the Rockies. Use α = 0.10

Need these (new) …

H_0: μ = 14,400
H_A: μ > 14,400
α = 0.10



P-Value

Don't need to find the critical value, so can just run the test first

p-value = ZTest(μ_0 = 14400, σ = 5000, $\bar{x}$ = 15000, n = 16, $\mu > \mu_0$) = 0.3156
p-value = 0.3156 < 0.10 = α → Fail to Reject H_0

Conclusion and Interpretation

Because our p-value = 0.3156 is greater than the significance level 0.10, we fail to reject the Null hypothesis. There is NOT sufficient evidence to conclude that the true mean height of the sea mountains are greater than 14,400 ft, which is the average height of the Rocky Mountains.

Hypothesis Tests for Means with UNKNOWN **σ**!

- All of the previous Hypothesis tests overview and for Means applies, now we are just going to look at the situation when we have an unknown **σ** (unlike before)!

- And going back to the Confidence Interval unit, we have now have an unknown population standard deviation! So the same logic and implications of that apply here as well!

- We will be doing a **T-Test**!

T Test vs Z-Test - Step Similarities

1. State the Hypotheses
   - **Define parameter + context.**

2. Check Assumptions.

Hypothesis Test Conditions

✓Randomization Condition
Need to have a random sample
✓Large Enough Sample Condition
Normal population OR
$n \geq 30$

5. Conclude and **Interpret**
- State whether you reject H_0 or fail to reject H_0 AND WHY!
- Interpret your results in the context of the problem

Samesies

- EXACT same Hypotheses Statements and Assumptions as for testing a Mean with known **σ**!
- BUT NOW, we only have the sample standard deviation *s*, the population standard deviation **σ** is UNKNOWN…
- And the Conclusion / Interpretation is the same as well!

Different

- The ONLY steps that involve something sightly different are the Rejection Region and the Test Statistic / P-Value

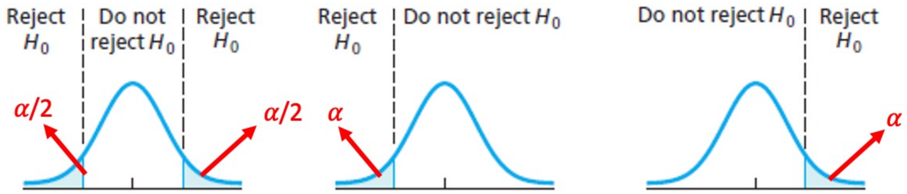
Rejection Region for Means with Unknown **σ**

3. Determine and Sketch Rejection Region based of Significance Level

Rejection Region for Means Test with UNKNOWN **σ**

- It is now based on the ***t*-distribution** rather than the standard normal distribution Z (just like it was for the Confidence Intervals).
  - So all of our CVs will be **t*s** with the **correct degrees of freedom** *df*!
- Using calc:

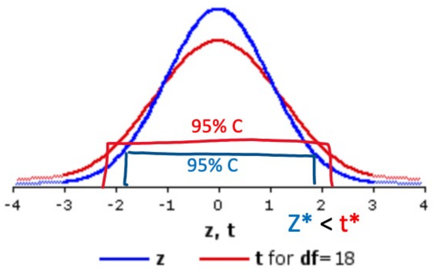
Left-Tailed: t* = invT(area = $\boldsymbol{\alpha, df = n-1}$***)***
Right-Tailed: t* = invT(area = $\boldsymbol{1-\alpha, df = n-1}$***)***
Two-Tailed: t* = invT(area = $\boldsymbol{\frac{\alpha}{2}, df = n-1}$***)***



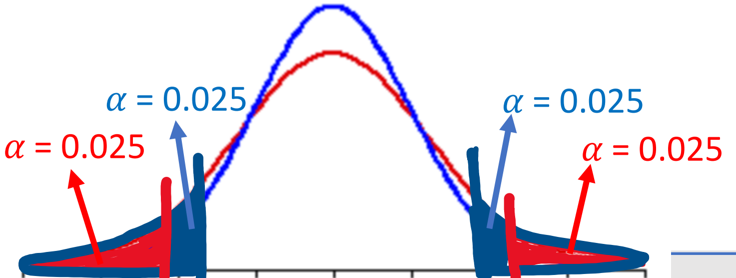
Type of Test:	(a) Two tailed	(b) Left tailed	(c) Right tailed
Critical Value(s):	-t* +t*	-t*	+t*
Rejection Region:	Both sides	Left side	Right side

Effect of Z vs T on Hypothesis Tests

- Recall from CI that because we have to **estimate σ** with ***s***, there is inherently more variability (uncertainty) which produces wider *t*-intervals compared to Z-intervals.



- In Hypothesis Tests, this translates to **Rejections Regions** being further away from the center for the same α!
- Which means we need to have more extreme results in order to reject when switching from Z to T Tests



Using Calc - Test Statistic and P-Value for Means with Unknown σ

4. Compute value of Test Statistic / P-value.

Formula for t_{stat} by hand:

Test for	H_0	Test statistic
Pop. mean μ	$\mu = \mu_0$	$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$

(ALMOST SAME Original) Setup
Scientists discovered a new mountain range under the sea. Lets assume the sea mountain heights are normally distributed.

From a random sample of 13 peaks, there was an average height of 11,308 ft and standard deviation of 5,287 ft. Is there enough evidence to conclude the average heights of these new sea mountains is different than the Rocky Mountains, which average 14,400 ft? Use α = 0.12

*** Only a sample standard deviation is provided, so σ is unknown → T-Test*

GOAL: Conduct a Hypothesis Test!

1. T-Test

- Option 1) Input = Stats

a) μ_0 = the Null mean

b) $\bar{x}$ = sample mean

c) Sx = sample SD

d) n = sample size

e) μ: Alternative hypothesis

Calculate or Draw

- Option 2) Input = Data
  - Enter raw data in L_1

a) μ_0 = the Null mean

b) List = L1

c) Freq = 1

d) μ: Alternative hypothesis

Calculate or Draw

Height
7700
11500
16000
5800
9000
5500
12400
22100
14200
19300
8300
9100
6100
Mean = 11308
SD = 5287

Using Calc - Test Statistic and P-Value for Means with Unknown σ

4. Compute value of Test Statistic / P-value.

Formula for t_{stat} by hand:

Test for	H_0	Test statistic
Pop. mean μ	$\mu = \mu_0$	$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$

<u>**(ALMOST SAME Original) Setup**</u>
Scientists discovered a new mountain range under the sea. Lets assume the sea mountain heights are normally distributed.

From a random sample of 13 peaks, there was an average height of 11,308 ft and standard deviation of 5,287 ft. Is there enough evidence to conclude the average heights of these new sea mountains is different than the Rocky Mountains, which average 14,400 ft? Use α = 0.12

<u>**GOAL**</u>: Conduct a Hypothesis Test!

*** Only a sample standard deviation is provided, so σ is unknown → T-Test*

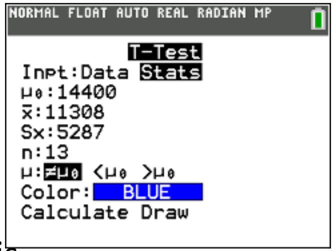
1. T-Test

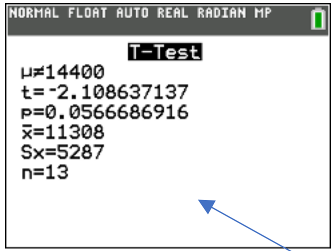
Height
7700
11500
16000
5800
9000
5500
12400
22100
14200
19300
8300
9100
6100

Mean = 11308
SD = 5287

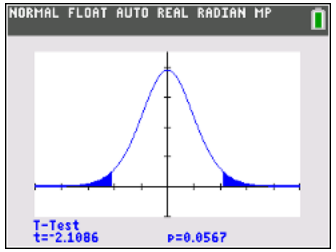
- Option 1) Input = Stats

a) μ_0 = the Null mean
b) $\bar{x}$ = sample mean
c) Sx = sample SD
d) n = sample size
e) μ: Alternative hypothesis
Calculate or Draw





<u>Calculate Output</u>
μ = Alternative hypothesis
t = t_{stat}
p = p-value
$\bar{x}$ = sample proportion
Sx = sample SD
n = sample size



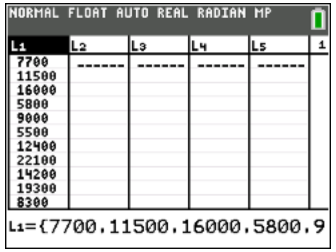
<u>Draw Output</u>
Plot (and displays values) of p = p-value and t = t_{stat} on the t curve with df = n -1

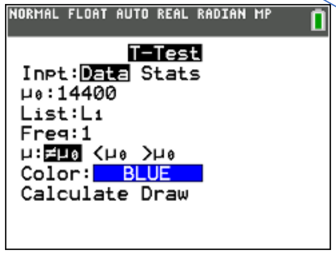
H_0: μ = 14,400
H_A: μ ≠ 14,400

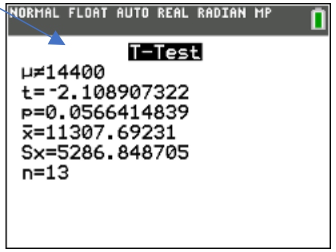
Same results! (maybe a little roundoff error from *$\bar{x}$ or s)*

- Option 2) Input = Data
  - Enter raw data in L_1

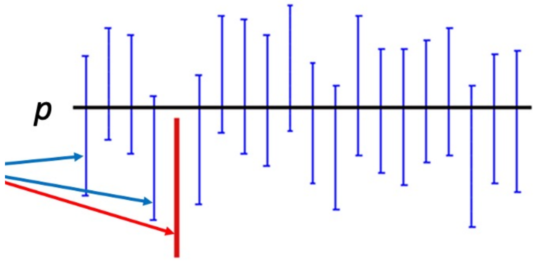
a) μ_0 = the Null mean
b) List = L1
c) Freq = 1
d) μ: Alternative hypothesis
Calculate or Draw







Decisions in Hypothesis Tests



A 95% confidence interval indicates that 19 out of 20 samples (95%) from the same population will produce confidence intervals that contain the population parameter.

- Recall from Confidence Intervals that it's not a guarantee that our interval captures the true population parameter!
    - As researchers we can do things to minimize the chances this happens such as having a high confidence level and a large sample size.
    - But we are working with real data when we take samples, and this is what your intervals are based on. There is always the possibility that our sample data leads us astray resulting in an interval that misses the population parameter ☹
- Of course, we never actually know if we capture or don't capture (because we don't know the truth). But it's something that we have to keep in mind when interpreting and making decisions based on our results.
- **This same dilemma is present in Hypothesis Tests as well!**
- There is the real possibility that we are making the WRONG conclusion to either Reject or Fail to Reject the Null hypothesis.
    - These are called **Type 1 and Type 2 Errors**!

Incorrect Decisions Example

Null Hypothesis: Not pregnant
Alternative Hypothesis: Pregnant

WRONG CONCLUSION!

Telling someone they <u>ARE pregnant</u> when in reality they are NOT.

ANOTHER WRONG CONCLUSION!

Telling someone they are <u>NOT pregnant</u> when in reality they ARE.

Type I and Type II Error

<u>Type I Error</u>

- This occurs when we incorrectly **reject** a TRUE Null hypothesis
    - False Positive → The test says you have COVID, but you actually don't
- *Reject the Null Hypothesis and conclude the alternative, when in reality the Null Hypothesis is actually correct.*
    - Recall α is the probability of rejecting the Null hypothesis. This also means that …
    - Probability of committing Type I error = α

<u>Type II error</u>

- This occurs when we **fail to (2) reject** a FALSE Null hypothesis
    - False Negative → The test says you don't have COVID, but you actually do
- *Fail to (2) Reject the Null Hypothesis, when in reality the the alternative is actually correct.*
    - Probability of committing Type II error = β

** If I ask you to describe a Type 1 or Type 2 error, these are the structures you should use + CONTEXT!!

There's two layers here:

1. The TRUTH (which we don't actually know)
2. Our DECISION (which he hope is correct)

Given the Null Hypothesis Is

Your Decision Based On a Random Sample		True	False
	Reject	Type I Error	Correct Decision
	Do Not Reject	Correct Decision	Type II Error

Two Types of Errors in Decision Making

https://home.ubalt.edu/ntsbarsh/Business-stat/opre504.htm

Relationship Between Type 1 and Type 2 Errors

Probabilities of Errors

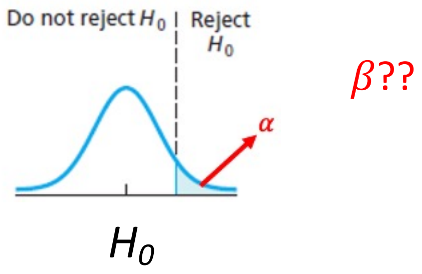
- When setting up a Hypothesis test, we need to think about the probabilities of committing each type of error!

- Depending on the context, the consequences of one type of error could be worse than the other!
  - Of course we don't want to commit an error and make the wrong conclusion, but this is always a possibility!

- P(Type 1 Error) = α → The **significance level** in of the hypothesis test! We control this!

  - This makes sense! We reject if our results land in the RR, which has size α. So if the Null is actually true, we make the wrong decision with α probability.

- P(Type 2 Error) = β → We cannot directly set this!



Relationship Between Alpha and Beta

- There is actually an inverse relationship between relationship α and β

  - As α ↓ decreases, because β ↑ increases
  - Why is this knowledge useful???

Consequences

- As the researcher, we can only control α! So this is how we have to manipulate the Test and control which type of error is less or more likely!

- When determining how to set the significance level, look at consequences of committing Type I and Type II error:

  - If Type I error is worse, minimize its probability α (i.e. decrease significance level) → for a stats class, this means setting α = 0.01
  - If Type II error is worse, minimize its probability (β) by increasing α (i.e. increase significance level) → for a stats class, this means setting α = 0.1
  - If you're unsure or errors are equally bad, stick with α = 0.05

- This strategy reduces the probability of the worse error!

Errors Example

Setup: All commercial elevators must pass yearly inspections. An inspector has to choose between certifying an elevator as safe (no repairs needed) or saying that the elevator is not safe (repairs are needed). There are two hypotheses:

H_0: The elevator is not safe (repairs are needed)

H_A: The elevator is safe (no repairs needed)

a) Describe Type I error and its consequences in context.

b) Describe Type II error and its consequences in context.

c) Which error is more serious? Explain.

Errors Example - Solution

https://home.ubalt.edu/ntsbarsh/Business-stat/opre504.htm

Given the Null Hypothesis Is

Your Decision Based On a Random Sample		True	False
	Reject	Type I Error	Correct Decision
	Do Not Reject	Correct Decision	Type II Error

Two Types of Errors in Decision Making

H_0: The elevator is not safe (repairs are needed)

H_A: The elevator is safe (no repairs needed)

a) Describe Type I error and its consequences in context → *Reject the Null Hypothesis and conclude the alternative, when in reality the Null Hypothesis is actually correct*

- *Description → A Type I error is wrongly concluding that the elevator is safe and no repairs are needed, when actually the elevator is not safe.*
- *Consequences → (snowballing effect) This would lead to not repairing an elevator that is in need of repairing (that is not safe), which could put people in danger.*

b) Describe Type II error and its consequences in context → *Fail to (2) Reject the Null Hypothesis, when in reality the the alternative is actually correct.*

- *Description → A Type II error is wrongly concluding that the elevator is not safe and repairs are needed, when actually the elevator is safe.*
- *Consequences → (snowballing effect) This would lead to repairing an elevator that does not need to be repaired (that is safe), which would cost some extra money*

c) Which error is more serious? Explain.

Committing a Type I error is more serious as the consequences would be not repairing an elevator that is in need of repairing (that is not safe), and people could be injured or die (which is always worse than unnecessarily spending money)

LCQ – Type 1 and Type 2 Errors

Your Decision Based On a Random Sample	Given the Null Hypothesis Is: True	False
Reject	Type I Error	Correct Decision
Do Not Reject	Correct Decision	Type II Error

Two Types of Errors in Decision Making

Setup: A restaurant got their supply of food (meat, veggies, etc.) a few days late and are trying to decide if they are still able to serve it or not. There are two hypotheses:

H_0: The food is still fresh

H_A: The food has gone bad

a) Describe Type I error and its consequences in context.

b) Describe Type II error and its consequences in context.

c) Which error is more serious? Explain.

LCQ – Type 1 and Type 2 Errors

Your Decision Based On a Random Sample			
	Reject	Type I Error	Correct Decision
	Do Not Reject	Correct Decision	Type II Error

Two Types of Errors in Decision Making

Setup: A restaurant got their supply of food (meat, veggies, etc.) a few days late and are trying to decide if they are still able to serve it or not. There are two hypotheses:

H_0: The food is still fresh

H_A: The food has gone bad

a) Describe Type I error and its consequences in context. → *Reject the Null Hypothesis and conclude the alternative, when in reality the Null Hypothesis is actually correct*

<u>Descriptions</u> - Options

1) The food still would be thought as fresh, when the food is not fresh → INCORRECT! This would mean concluding the Null is TRUE but in reality the Alternative is correct (this is actually Type 2!)

2) Type I error is wrongly concluding that the food is not fresh → ALMOST, this has the first part perfect! But MISSING the 'in reality the Null is TRUE'. So NEED to ADD 'but in reality it's good'

3) We wrongly conclude that the food has gone bad, when it is still good → Very good! Stating we wrongly conclude alternative in context but the Null is actually true context!

<u>Consequences</u> - Options

1) Someone gets sick → *INCORRECT! This would go with the Type 2 error*

2) The food was actually still fresh so you lost money by not using it → *Very good! Snowballing, good description of what's the worst case scenario if we make the above error*

3) We throw out good food and waste it → Also good!

4) Waste of money and food → Good again! Many ways to phrase it

b) Describe Type II error and its consequences in context. → *Fail to (2) Reject the Null Hypothesis, when in reality the the alternative is actually correct.*

<u>Descriptions</u> - Options

1) We wrongly conclude that the food is fresh → ALMOST again! MISSING the second key part of committing an error, the truth being different than our conclusion! NEED to HAVE 'when it actually it has gone bad'

2) We wrongly conclude that the food is still good when it has actually gone bad → PERFECT! Conclude H_A in context but actually H_0 context is true!

<u>Consequences</u> - Options

1) People may get sick → Good! Again snowballing to what could happen

2) We serve bad food and someone gets sick from it → Also great!

c) Which error is more serious? Explain.

A Type 2 Error would be worse because customers are at risk of getting sick from food poisoning. Even though the restaurant might lose money, people's safety is more important

Another Look at P-Values

P-Value REVIEW

- The **p-value** is the probability of getting a result as, or more extreme than, the result obtained from the sample given (assuming) the Null Hypothesis (H_0) is TRUE.
- In other terms, "The P-Value is a measure of how plausible the data are, given our null hypothesis."

LCQ: Which of the following are true? If false, explain briefly.

a) A very high P-value is strong evidence that the null hypothesis is false.

b) A very low P-value proves that the null hypothesis is false.

c) A high P-value shows that the null hypothesis is true.

d) A P-value below 0.05 is always considered sufficient evidence to reject a null hypothesis.

e) A P-value is the probability of getting a result equal to or more extreme than our sample results assuming the Null hypothesis is correct.

Another Look at P-Values

P-Value REVIEW

- The **p-value** is the probability of getting a result as, or more extreme than, the result obtained from the sample given (assuming) the Null Hypothesis (H_0) is TRUE.

- In other terms, “The P-Value is a measure of how plausible the data are, given our null hypothesis.”

LCQ: Which of the following are true? If false, explain briefly.

a) A very high P-value is strong evidence that the null hypothesis is false.
FALSE → A very low p-value is strong evidence against the null hypothesis (H_0), which is why we reject when it’s really small
- *It would mean that it is very unlikely we would get the result we did if the normal curve based on the null were correct*

b) A very low P-value proves that the null hypothesis is false.
FALSE → Does not prove, but is a strong evidence
- *We NEVER **prove** or **disprove** H_0 ; can never PROVE something when working with sample data (need fancy science or population data probably)*

c) A high P-value shows that the null hypothesis is true.
FALSE → We are trying to find evidence AGAINST H_0 ;
- *So a high p-value indicates a LACK of evidence **against** the Null (which is DIFFERENT than having evidence in **favor / support** of the Null)*
- *Think about decisions in court: either guilty or not guilty (NOT ‘innocent’).*
- *Not guilty does NOT mean innocent, we just can’t say that you did commit the crime*

d) A P-value below 0.05 is always considered sufficient evidence to reject a null hypothesis.
FALSE → Could be true, but it would depend on the significance level (which can change from test to test based on the ideas of the researcher)

e) A P-value is the probability of getting a result equal to or more extreme than our sample results assuming the Null hypothesis is correct.
TRUE!!! → Remember the whole test is set up with the assumption the Null is true from the start. And our p-values are tail probabilities, so it also includes

LCQ – Find all the Errors!!

Setup: Dr. Suess is trying interested in the colors of the fish in his pond. Originally there was 55% red and 45% blue. But now he suspects the mean blue fish are starting to take over and there is a higher proportion of blues. In order to check this, he randomly samples of 40 fish in which 23 were blue.

Is there sufficient evidence to conclude that the proportion of blue fish is greater than 0.45? Test this at the 10% significance level.

Solution

Hypotheses:

Let p = proportion of blue fish in Dr. Suess' pond

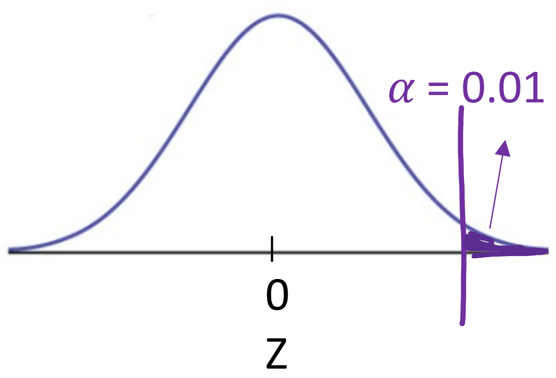
$H_0 = 0.45$
$H_A > 0.46$

Set $\alpha = 0.01$

Check assumptions:

- *Randomization: Random sample of students was taken*
- *Large enough sample:*
    - $40(.575) = 23 > 5$
    - $40(0.425) = 17 > 5$
- *Both conditions are met, appropriate to continue with test!*

Rejection Region:



P-value:

p-value = 1–PropZTest(p_0 = 0.45, x = 23, n = 40, prop > p_0) = 0.056
p-value = 0.056 > 0.01 = α → Fail to reject H_0

Conclusion:

Because our p-value = 0.056 is greater than the significance level 0.01, we fail to reject the Null hypothesis. There IS sufficient evidence to conclude that the true proportion of blue fish in Dr. Suess' pond is greater than 0.45.

LCQ – FOUND all the Errors!!

Setup: Dr. Suess is trying interested in the colors of the fish in his pond. Originally there was 55% red and 45% blue. But now he suspects the mean blue fish are starting to take over and there is a higher proportion of blues. In order to check this, he randomly samples of 40 fish in which 23 were blue.

Is there sufficient evidence to conclude that the proportion of blue fish is greater than 0.45? Test this at the 10% significance level.

Solution

Hypotheses:

Let p = proportion true of blue fish in Dr. Suess' pond

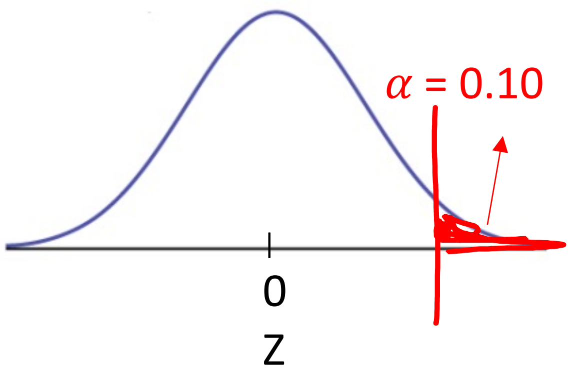
$H_0: p = 0.45$
$H_A: p > 0.45$

Set $\alpha = 0.10$

Check assumptions:

- *Randomization: Random sample of fish was taken*
- *Large enough sample:*
  - $np_0 = 40(0.45) = 18 > 5$
  - $n(1 - p_0) = 40(0.55) = 22 > 5$
- *Both conditions are met, appropriate to continue with test!*

Rejection Region:



P-value:

p-value = 1–PropZTest(p_0 = 0.45, x = 23, n = 40, prop > p_0) = 0.056

p-value = 0.056 < 0.10 = α → Reject H_0

Conclusion:

Because our p-value = 0.056 is less than the significance level 0.10, we reject the Null hypothesis. There (now) IS sufficient evidence to conclude that the true proportion of blue fish in Dr. Suess' pond is greater than 0.45.

LCQ – Find all the Errors Part 2!!

Setup: Dr. Suess is trying to figure out how big his fish are. From his previous random sample of 40 fish and there was an average of 35 lbs with standard deviation 5.5 lbs.

Is there sufficient evidence to conclude that the mean weight of fish is different than 37 lbs? Test this at the 5% significance level.

Solution

Hypotheses:

$H_0: p = 35$
$H_A: p \neq 35$

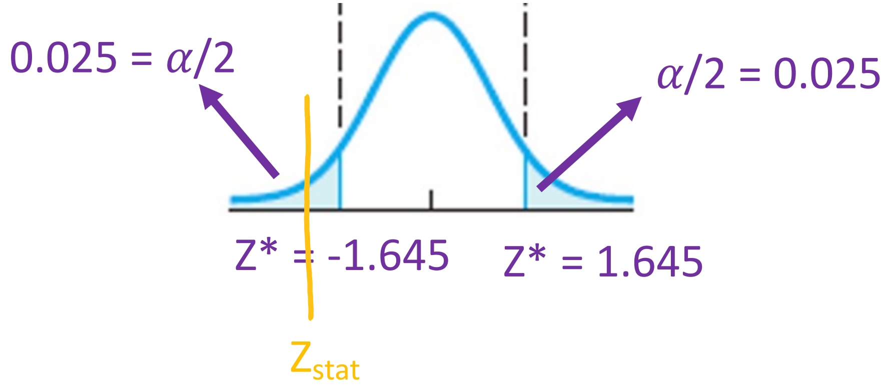
Set $\alpha = 0.05$

Check assumptions:

- *Large enough sample:*
    - $n > 30$

Rejection Region:

$Z^* = invNorm(area = 0.05, \mu = 0, \boldsymbol{\sigma} = 1) = -1.645$



Test Statistic:

$Z_{stat} = Z\text{–}Test(\mu_0 = 37, \boldsymbol{\sigma} = 5.5, \bar{x} = 35, n = 13, \mu \neq \mu_0) = -2.3$

$|Z_{stat}| = 2.3 > 1.645 = |Z^*| \rightarrow$ *Reject* $H_0!$

Conclusion:

Because the absolute value of our Test Statistic $Z_{stat} = 2.3$ *is greater than the absolute value of our Critical Value* $Z^* = 1.645$*, we reject the Null hypothesis. There IS sufficient evidence to conclude the alternative.*

LCQ – FOUND all the Errors Part 2!!

Setup: Dr. Suess is trying to figure out how big his fish are. From his previous random sample of 40 fish and there was an average of 35 lbs with standard deviation 5.5 lbs.

Is there sufficient evidence to conclude that the mean weight of fish is different than 37 lbs? Test this at the 5% significance level.

Solution

Hypotheses:

Let μ = true mean weight of Dr. Suess' fish

$H_0: \mu = 35$
$H_A: \mu \neq 35$

Set α = 0.05

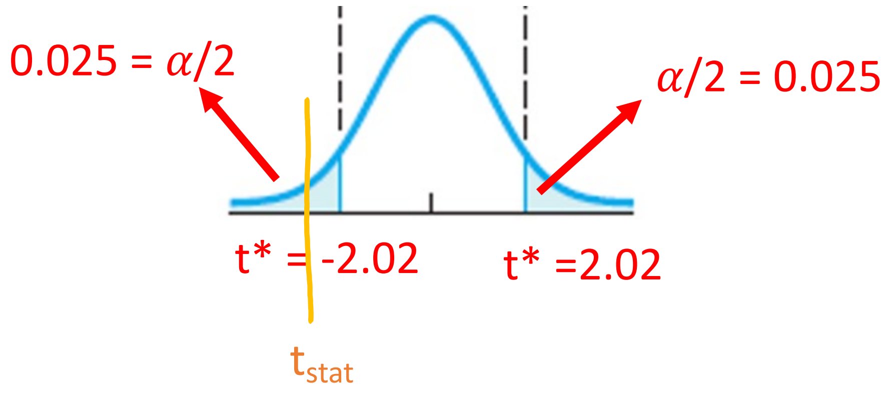
Check assumptions:

- *Random sample was taken!*
- *Large enough sample:*
    - *n = 40 > 30*

Both conditions are met, okay to continue!

Rejection Region:

t = invT(area = 0.025, df = 40-1) = -2.02*



Test Statistic:

t_{stat} = T–Test(μ_0 = 37, $\bar{x}$ = 35, Sx = 5.5, n = 40, $\mu \neq \mu_0$) = -2.3
$|t_{stat}|$ = 2.3 > 2.02 = |t| → Reject H_0!*

Conclusion:

Because the absolute value of our Test Statistic t_{stat} = 2.3 is greater than the absolute value of our Critical Value t = 2.02, we reject the Null hypothesis. There IS sufficient evidence to conclude the true mean weight of Dr. Suess' fish is greater than 35 lbs.*

Problem Session!!!

Human Body Temperatures Example

- A random sample of 15 human body temperatures were obtained. Assume that human body temperatures are normally distributed
- Is there sufficient evidence to conclude that the true mean human body temperature differs from 98.6^{o}F? Use $\alpha = 0.05$

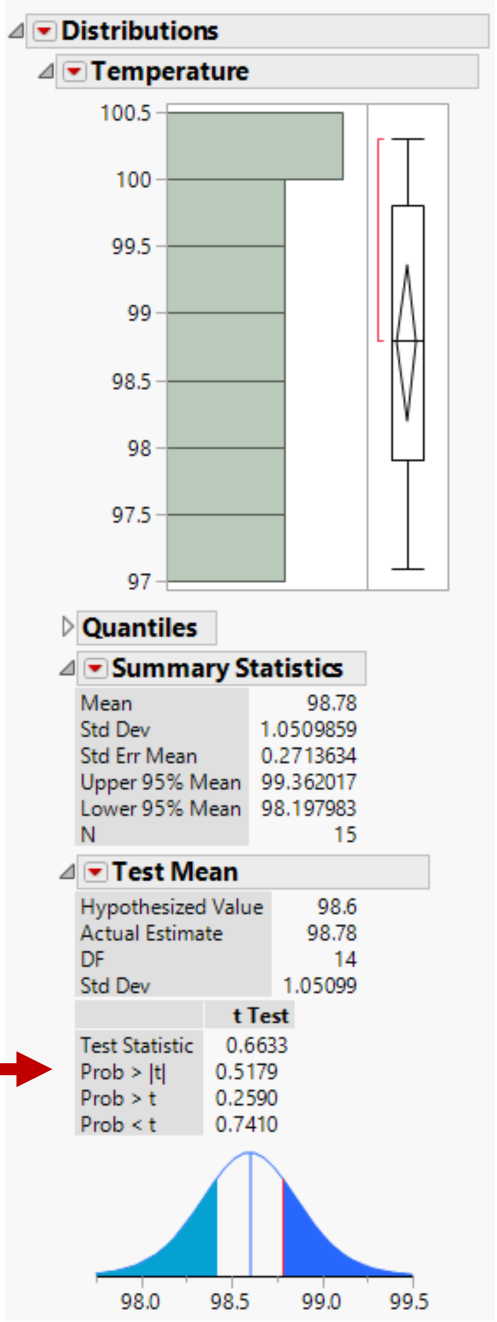
97.6	100.3	100	97.9	99.2
98.8	97.4	98	99.1	100.2
97.1	99.8	98.5	99.5	98.3

Human Body Temperatures Solution

- Let μ = true mean human body temperature
- Check Assumptions: We have a quantitative variable, temperature, we have a SRS of 15 human body temperatures as mentioned in the problem statement. We were also told to assume that human body temperatures are normally distributed.
- H_0: $\mu = 98.6^o$, H_a: $\mu \neq 98.6^o$, $\alpha = 0.05$

Human Body Temps Solution, p. 2

- H_0: μ = 98.6º, H_a: μ ≠ 98.6º, $\alpha = 0.05$
- T-test statistic = 0.6633, df = 14
- We are running a two-tailed test, and the t-test statistic is greater than 0. To find the correct p-value use **Prob >|t|** → 0.5179
- Since the p-value of 0.5179 is greater than our significance level of 0.05, we fail to reject H_0. There is not sufficient evidence to conclude that the true mean human body temperature differs from 98.6ºF.



Example

The mean weight of U.S. Fancy Grade watermelons is 22.0 pounds. Watermelons that are too light are dry and less flavorful. A large grocery chain receives a shipment of watermelons for sale, and the produce supervisor randomly selects 36 watermelons and weighs each. The sample mean is 21.5 pounds, and the sample standard deviation is 3.0 pounds. Should the shipment of watermelons be accepted?

Use $\alpha = 0.05$.

Example Solution

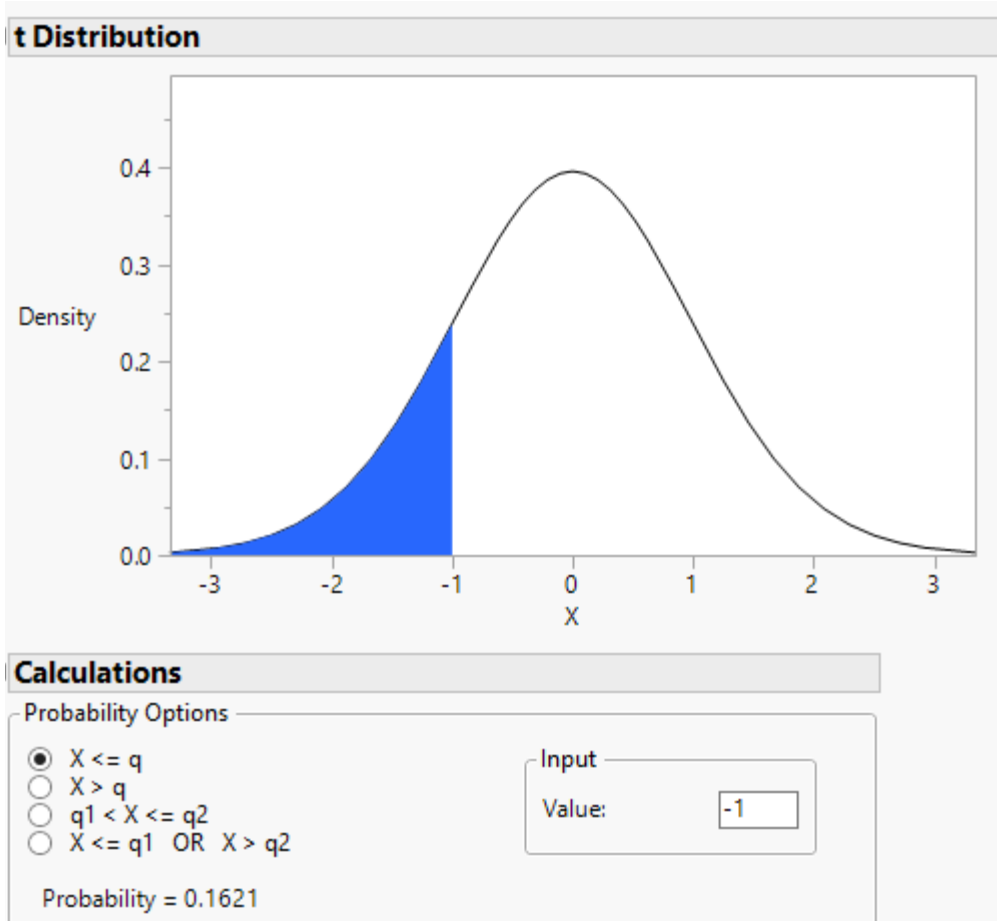
- Let μ = true mean weight of watermelons in the shipment
- Check Assumptions: We have a quantitative variable, the mean weight of watermelons in the shipment, we have a SRS as mentioned in the problem statement. We will assume that weight of watermelons is normally distributed.
- H_0: μ = 22.0, H_a: μ < 22.0, $\alpha = 0.05$

Type I and Type II Errors

- Type I error is rejecting a true H_0: concluding that mean weight of the watermelons is less than 22.0 pounds when it is actually 22.0 pounds (or more).
  - Consequence of Type I error, refusing a shipment of good watermelons. Supplier may refuse to sell product to us in the future. Or, we won't have watermelons to sell to customers
- Type II error is failing to reject a false H_0: concluding that the mean weight of the watermelons is 22.0 pounds (or more) when it is actually less than 22.0 pounds.
  - Consequence of Type II error, accepting a shipment of dry, tasteless watermelons. Customers will buy the watermelons and not be pleased with them. Could possibly loose future customers.
- Problem specified using $\alpha = 0.05$

Example Solution

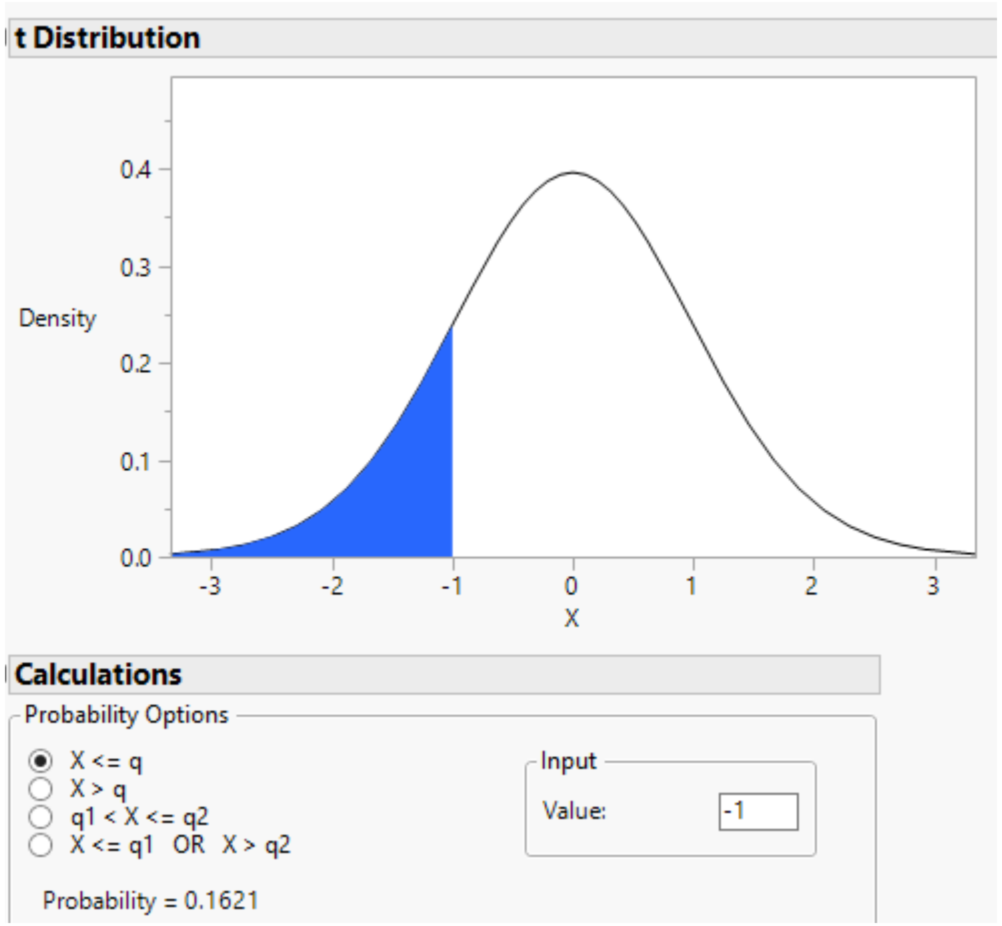
- H_0: μ = 22.0, H_a: μ < 22.0, $\alpha = 0.05$
- $\bar{x} = 21.5, s = 3.0, n = 36$
- Test statistic: $t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} = \frac{21.5-22.0}{3.0/\sqrt{36}} = -1$
- $df = n - 1 = 36 - 1 = 35$

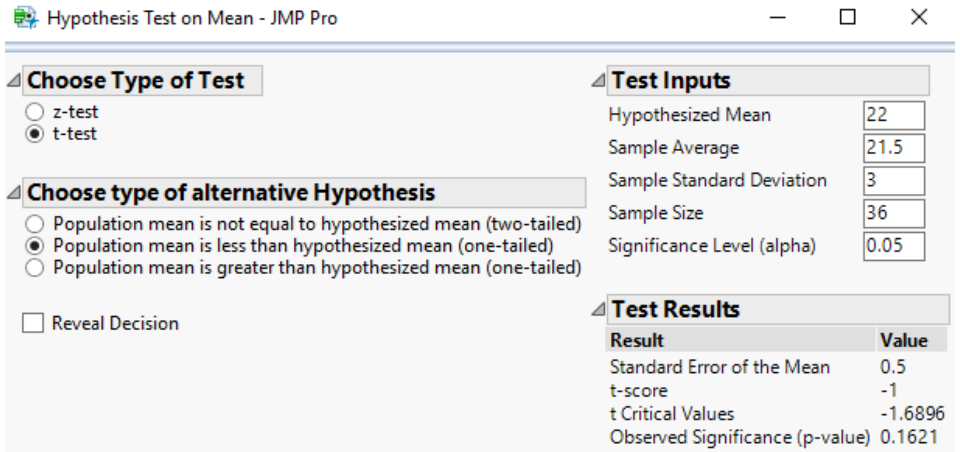


Example Solution, p. 2

- Since the p-value = 0.1621, which is greater than the significance level of 0.05, we fail to reject H_0.

- We conclude that the mean weight of the watermelons in the shipment is at least 22.0 pounds. The shipment should be accepted.





Practice Problem #1

A survey of 25 randomly selected customers found the following ages (in years):

20	32	34	29	30	25	22	32	44	48	26	32	35
30	30	14	29	11	35	42	48	38	22	44	44	

The mean was 31.84 years and the standard deviation 9.84 years. The owner of the store wants to know if the mean age of all customers is 25 years old.

a) What is the null hypothesis?
b) What is the alternative hypothesis? Is it one- or two-sided?
c) What is the value of the test statistic?
d) What is the P-value of the test statistic?
e) What do you conclude at alpha = 0.05?
f) What is a Type I error in this context?
g) What is a Type II error in this context?

Practice Problem #1 Solution

Let μ = the true mean age of all customers

a) H_0: μ = 25
b) Two-sided; H_a: $\mu \neq 25$
c) t = 3.4745, df = 24
d) P-value = 0.0019
e) Reject H_0. Since the p-value is less than the significance level of 0.05, there is sufficient evidence to reject the null hypothesis and conclude that the true mean age of all customers is not equal to 25 years of age.
f) A Type I error is wrongly rejecting a true null hypothesis; concluding that the true mean age of all customers is not equal to 25 years of age when it actually is.
g) A Type II error is wrongly failing to reject a false null hypothesis; concluding that the true mean age of all customers is equal to 25 years of age when it actually is not.

Practice Problem #2

The average age of online consumers a few years ago was 23.3 years. As older individuals gain confidence with the Internet, it is believed that the average age has increased. We would like to test this belief.

a) Write appropriate hypotheses.

b) We plan to test the null hypothesis by selecting a random sample of 40 individuals who have made an online purchase this year. Do you think the necessary assumptions for the inference are satisfied? Explain.

c) The online shoppers in our sample had an average age of 14.2 years, with a standard deviation of 5.3 years. What's the P-value for this result?

d) Explain (in context) what this P-value means.

e) What's your conclusion?

Practice Problem #2 Solution

Let μ = the true mean age of online customers

a) H_0: μ = 23.3 years vs. H_a: $\mu > 23.3$ years

b) The sample is random and should be large enough to ensure normality. We should check the distribution for serious skewness or outliers.

c) P-value = 1

d) If the mean age of online customers is 23.3 years, the chance of obtaining a sample mean of 14.4 years , or less due only to sampling error is 100%.

e) Fail to reject H_0. Since the p-value is larger than any reasonable significance level, there is not sufficient evidence to reject the null hypothesis and conclude that the true mean age of online customers has increased.

Practice Problem #3

A company with a large fleet of cars hopes to keep gasoline costs down and sets a goal of attaining a fleet average of at least 26 miles per gallon. To see if the goal is being met, they check the gasoline usage for 50 company trips chosen at random, finding a mean of 25.02 mpg and a standard deviation of 4.83 mpg. Is this strong evidence that they have failed to attain their fuel economy goal?

a) Conduct and interpret a hypothesis test to answer the research question. Use a significance level of 0.05.
b) Explain carefully what the P-value means in this context.
c) Explain what a Type I error is in this context.
d) Explain what a Type II error is in this context.

Practice Problem #3 Solution

Part A:

1. Let μ = the true mean gas mileage for the fleet of cars
2. H_0: μ = 26 mpg vs. H_a: $\mu < 26$ mpg
3. α = 0.05
4. Check the assumptions:
   - The 50 company trips were selected randomly.
   - The mileage on one trip should not affect the mileage on other trips. It is safe to assume that the mileage is independent for each trip.
   - A sample of 50 trips should be large enough to ensure normality, according to the CLT.
5. t = -1.4347, df = 49
6. P-value = 0.0789
7. Fail to reject H_0.
8. Since the p-value is larger than a 5% significance level, there is not sufficient evidence to reject the null hypothesis and conclude that the true mean gas mileage of the fleet is less than 26 mpg.

Practice Problem #3 Solution, p. 2

Part B: The p-value indicates that if the actual mean mileage of cars in the fleet is 26 mpg, the chance of obtaining a sample mean of 25.02 mpg , or less due only to sampling error is 7.89%.

Part C: A Type I error is wrongly rejecting a true null hypothesis; concluding that the true mean gas mileage for the fleet of cars is less than 26 mpg when it is actually at least 26 mpg.

Part D: A Type II error is wrongly failing to reject a false null hypothesis; concluding that the true mean gas mileage for the fleet of cars is at least 26 mpg when it is actually less than 26 mpg.

Practice Problem #4

Consumer Reports tested 11 brands of vanilla yogurt and found these numbers of calories per serving:

130	160	150	120	120	110	170	160	110	130	90

a) Check the assumptions and conditions for inference.

b) Create and interpret a 95% confidence interval for the average calorie content of vanilla yogurt.

c) A diet guide claims that you will get an average of 120 calories from a serving of vanilla yogurt. What does this evidence indicate? Use your confidence interval to test an appropriate hypothesis and state your conclusion.

Practice Problem #4 Solution

a) Assume a representative sample, the histogram is unimodal and symmetric.

b) I am 95% confident that the true mean amount of calories per serving in vanilla yogurt is between 114.87 and 148.77 calories.

c) Since 120 is contained in the interval there is not sufficient evidence to reject the claim that the mean number of calories per serving of vanilla yogurt is 120 calories; the claim seems reasonable.

Practice Problem #5

In 2012, a large number of foreclosed homes in the Washington, D.C., metro area were sold. In one community, a sample of 30 foreclosed homes sold for an average of $443,705 with a standard deviation of $196,196.

a) What assumptions and conditions must be checked before finding a confidence interval for the mean? How would you check them?
b) Find a 95% confidence interval for the mean value per home.
c) Interpret this interval and explain what 95% confidence means.
d) Suppose nationally, the average foreclosed home sold for $350,000. Do you think the average sale price in the sampled community significantly from the national average?

Practice Problem #5 Solution

a) Random sample, symmetric, unimodal distribution without outliers.
b) ($370,444, $516,966)
c) I am 95% confident that the true mean amount foreclosed homes in the DC area will sell for is between $370,444 and $516,966. The 95% confidence level means that if we were to take all possible samples of 30 foreclosed homes, 95% of them would capture the true mean selling price.
d) Yes, since $350,000 is below the confidence interval. I would conclude that houses in the DC area sell for more money.

Practice Problem #6

Police departments often try to control traffic speed by placing speed-measuring machines on roads that tell motorist how fast they are driving. Traffic safety experts must determine where the machines should be placed. In one recent test, police recorded the average speed clocked by cars driving on one busy street close to an elementary school. For a sample of 25 speeds, it was determined that the average amount over the speed limit for the 25 clocked speeds was 11.6 mph with a standard deviation of 8 mph. The 95% confidence interval estimate for this sample is 8.30 mph to 14.90 mph.

a) What is the margin of error for this problem?

b) The researchers commented that the interval was too wide. Explain specifically what should be done to reduce the margin of error to no more than +/- 2 mph.

Practice Problem #6 Solution

a) 3.3 mph

b) To reduce the margin of error we need to use a larger sample size.

$$1.96 \times \frac{8}{\sqrt{n}} = 2$$

$$1.96 \times 8 = 2\sqrt{n}$$

$$\left(\frac{1.96 \times 8}{2}\right)^2 = (\sqrt{n})^2$$

$$7.84^2 = n$$

$$62 = n$$