Week Nine: Hypothesis Testing with Means

Week Nine Goals

- Vocabulary of Hypothesis Testing
- The Logic of Hypothesis Testing
- The Complete Testing Process ← you will be asked to show this process for each test.
- Population Step
    - Describe the Population and State the Goal of the Test
- Method Step
    - State the Null and Alternative Hypotheses (Left, Right, Two-Tailed)
    - State the Level of Significance
    - Indicate the Distribution that will be Used
- Sample Step
    - One Sample Mean (t) Test—with StatHelper
    - One Sample Mean (t) Test—with Minitab 19
    - Check Normality Assumption
- Results Step
    - The T - Test Statistic, df and T Interpretation
    - The P-value and Interpretation
- Conclusion Step
    - Make a Decision
    - Write a Statistical Conclusion
    - State an Everyday-English Conclusion
- Examples
- Type I Error and Type II Error

Vocabulary of Hypothesis Testing:

Parameter: A quantity that is calculated from data and describes a **population.**

Statistic: A quantity that is calculated from data and describes a **sample.**

Null Hypothesis: A general statement or default position that there is nothing new happening, like there is no association among groups, or no relationship between two measured phenomena or that nothing has changed…it is as we always thought it was. The status quo. The null hypothesis always includes an equal sign.

Alternative Hypothesis: A statement that indicates the change that the researcher is hoping to show. It is often the goal of the researcher. The alternative hypothesis is always written as either 'less than', 'greater than' or 'not equal'.

Hypothesized Value: The value that you are assuming is true in the null hypothesis.

Level of Significance: Also denoted as alpha or α, is the probability of rejecting the null hypothesis when it is true. For example, a **significance level** of 0.05 indicates a 5% risk of concluding that a difference exists when there is no actual difference.

Test Statistic: A numerical summary of a data-set that reduces the data to one value that can be used to perform the hypothesis test. The test statistic indicates the number of standard errors away that the sample statistic is from the hypothesized value.

Probability-value (P-value): The probability of a getting the statistic you get from sampling or more extreme under the assumption that the null hypothesis is true.

Reject the Null Hypothesis: When the sample data is sufficiently different than the hypothesized value, the calculated P-value will be smaller than or equal to alpha. As a result, we will decide that we CAN "reject the null hypothesis". In other words, we are no longer believing the statement in the null hypothesis; we are deciding that we have enough evidence to support the alternative hypothesis.

Fail to reject the null hypothesis or deciding we CANNOT reject the null hypothesis. In the event that our sample data is not sufficiently different from the hypothesized value, the calculated P-value will be greater than alpha. As a result, we will decide that we "CANNOT reject the null hypothesis". In other words, we do not have enough evidence to support the alternative hypothesis. We are not going to come right out and say we '*know that the null hypothesis is true*' but we certainly are not able to support a statement against the null.

Type I Error: Error that occurs when the sample data leads you to decide to reject the null, ***when the null is actually true***. The probability of Type I Error is alpha…the level of significance. The default alpha = 0.05.

Type II Error: Error that occurs when the sample data leads you to decide that we cannot reject the null, ***when the null should have been rejected***. The probability of Type II Error is beta. If beta decreases, alpha increases. If alpha decreases, beta increases; there is an inverse relationship between alpha and beta. We will not calculate beta in this course.

The Complete Testing Process: There are steps that Prof. Coffey wants to see for each hypothesis test. This process includes the following steps: Population Step, Method Step, Sample Step, Results Step and Conclusion Step. These are defined on the following pages and the process shown in all examples. When I ask for the complete testing process, I am asking for you to show these steps as taught. If you do not, then it is very clear to me that you have not read these typed notes.

Video Introducing you to the Vocabulary of hypothesis testing from Prof. Coffey

Here is a video about vocabulary for hypothesis testing: https://youtu.be/n5Z1WBEKYdo

Here is the same video with interpreting: https://youtu.be/yBCkkdddmVA

The Logic of Hypothesis Testing

In order to test for the alternative hypothesis, we will see whether or not the null hypothesis is ***plausible***.

- If the null hypothesis ***is not plausible***, we CAN reject the null hypothesis and conclude that there is sufficient evidence to support the alternative hypothesis.

- If the null hypothesis ***is plausible***, we CANNOT reject the null hypothesis and conclude that there is NOT sufficient evidence to support the alternative hypothesis.

Why do we check if H_o is plausible?

Think about the logic of jury trials:

- To prove someone is guilty, we start by assuming they are innocent.
- We retain that hypothesis of innocence until the facts make it unlikely beyond a reasonable doubt.
- Then, and only then, we decide we CAN reject the hypothesis of innocence and declare the person guilty.

The same logic used in jury trials is used in statistical tests of hypotheses:

- We begin by assuming that the null hypothesis, H_o, is true.

- Next we consider whether the sample data are consistent with this hypothesis (H_o).

- If the sample data are consistent with H_o, all we can do is retain the hypothesis we started with (H_o).

- If the sample data are not consistent with H_o, then, like a jury, we ask whether they are unlikely beyond a reasonable doubt. If they were very unlikely to have occurred (small probability), then the evidence raises more than a reasonable doubt in our minds about the null hypothesis (H_o) and we support the alternative hypothesis (H_a).

The Complete Testing Process

Population Step -- Describe the population and state the goal.

1. Describe the variable being studied (also indicate if the variable is numerical or categorical; this week it is numerical).
2. Define the parameter μ = The true mean___________
3. State the GOAL of the analysis. ***The Goal ALWAYS matches the Ha.***

We will be testing to see if the population mean, μ, has increased, decreased or is different from the hypothesized mean, $\mu_{0.}$

Method Step -- Indicate the method that will be used.

1. State the Ho and Ha.

$Ho: \mu = \mu_0$
$Ha: \mu < \mu_0$ Left-Tailed Test

$Ho: \mu = \mu_0$
$Ha: \mu > \mu_0$ Right-Tailed Test

$Ho: \mu = \mu_0$
$Ha: \mu \neq \mu_0$ Two-Tailed Test

2. State alpha (α). [α = 0.05 is the default; use it unless told otherwise]

NOTE: α is a level of significance. It is the probability of Type I error, i.e. the probability of rejecting the null (and concluding that the alternative hypothesis was true) when in fact, the null is true.

3. State the distribution with which you will work. When we study numerical data, we will work on the **t-curve** with df = n-1. [Next week, when we study categorical data, we will be on the Z-curve]

Sample Step -- Analyze the sdata by running the test with technology; also check the normality assumption.

Be sure you have the sample size, n, the sample mean, $\bar{x}$, and the sample standard deviation, s.
If you have raw data, the statistical technology will summarize the data while running the hypothesis test.

1. Run the hypothesis test using statistical technology and paste the output here.

2. Check to see if the normality assumption for using the Student's T-Distribution has been met. Recall that if n $\geq$ 30, then the Central Limit Theorem allows us to assume a normal sampling distribution. If n > 30, and we have raw data, then build a Normal Probability Plot (NPP) and check to see that the P-value > 0.05.

Results Step -- Test statistic, df and P-value results and Interpretations

1. State the T-test statistic, df = n -1 and interpret the T test statistic.
2. State and interpret the P-value.

Conclusion Step -- Make a decision and a statistical conclusion; everyday conclusion encouraged

Determine if your P-value is less than or equal to alpha? (P-value $\leq \alpha$?)	**Yes**	**No**
1. State your decision about the null hypothesis (Ho).	**Can Reject Ho**	**Cannot Reject Ho**

A statistical conclusion is always a statement indicating that we **HAVE ENOUGH** support or that we **DO NOT HAVE ENOUGH** support for the Ha/goal. You can use a template such as the one below

2. Write the statistical conclusion:
"At the ___% level of significance, the sample data **DOES** or **DOES NOT** provide sufficient evidence to state that the true mean …………….restate the goal here………………."

A conclusion in everyday English:
Often, I will ask you to also write an 'everyday' conclusion; a conclusion that is written casually, in the context of the problem, with no statistical language and gets right to the point of the conclusion.

Video discussing the COMPLETE TEST PROCESS from Prof. Coffey

Here is a video about the complete test process: https://youtu.be/ZDj1RNnDtBk

Here is the same video with interpreting: https://youtu.be/flbi5-DyG8E

Population Step

Describe the Population and State the Goal of the Test:

It is important that we can…

- State the variable (and units) being studied and indicate if it is numerical or categorical (this week they are numerical).
- Define the parameter μ = the true/population mean ....describe variable............
- State the GOAL of the hypothesis test: "Test to see if there is support for saying that μ has (increased, decreased, or changed) from ……describe μ_o $(the\ hypothesized\ mean)$…."

NOTE: Next week, we will learn to run a hypothesis test with proportions. In that case, this step will change. We will define p = the population proportion and the entire test will use the symbol p (instead of μ).

Example 1 – Population Step

It is believed that 72 is the average pulse rate for college-aged women. A group of students wonder if this long-held standard pulse rate value is still accurate (or has it changed) and gather a random sample of 35 pulse rates from college-aged women. The mean pulse rate was found to be $\bar{x}$ = 76.8 with a sample standard deviation of s = 11.62. Ultimately, we will show the complete testing process. For right now, show the population step.

Population Step

The variable is numerical and is the pulse rate for college-aged women.

μ = the population mean pulse rate for all college-aged women.

Goal: Test to see if there is support for saying that μ ≠ 72.

[NOTE: the hypothesized mean, μ_o, is 72 for this problem]

Method Step

State the Null and Alternative Hypotheses (Left, Right, Two-Tailed)

In Statistics, when testing **claims**, we use an objective method called hypothesis testing.
We call these **claims** hypotheses. Our starting point, the status quo, is called the **null hypothesis (H_o)** and the alternative claim is called the **alternative hypothesis (H_a).**
We may have some past information on a variable and wonder if the current data show that the value has statistically increased, decreased or is 'different' than the past information. Hypothesis testing can accomplish this.

The Null Hypothesis (H_o)

The Null Hypothesis, denoted by **H_o**, specifies a population parameter of interest (μ) and proposes a value for that parameter, called the hypothesized value (μ_0). This statement is what we are assuming is true as we begin testing. It is generally a statement of 'status quo', meaning it is what we have always believed μ to be. The null hypothesis always is a statement of 'equality'; meaning there is always an EQUAL SIGN in the null hypothesis. It is possible for the Ho to read μ ≤ ____ or μ ≥ ___ (as long as the equal sign is there, it is a fair statement for the null). However, it is not vital that the null be anything other than an equal sign, so we can agree to always make the null statement a statement of equality.

$H_o: \mu =$ ______   ← the value that is placed here is called the 'hypothesized mean, μ_0.

The Alternative hypothesis (H_a)

The Alternative hypothesis, denoted by **H_a or H_1**, is the claim we are testing for. It is what the researcher is interested in proving. The alternative hypothesis is a statement that μ has decreased or μ has increased or that μ has changed.

$\boldsymbol{H_a: \mu <}$ _____ OR	$\boldsymbol{H_a: \mu >}$ _____ OR	$\boldsymbol{H_a: \mu \neq}$ _____
If you are trying to prove that μ has decreased.	If you are trying to prove that μ has increased.	If you are trying to prove that μ is different (has changed).
This is a left-tailed test.	**This is a right-tailed test.**	**This is called a two-tailed test.**

In summary, these are the three possible hypotheses:

$\boldsymbol{H_o: \mu =}$ _____ $\boldsymbol{H_a: \mu <}$ _____	$\boldsymbol{H_o: \mu =}$ _____ $\boldsymbol{H_a: \mu >}$ _____	$\boldsymbol{H_o: \mu =}$ _____ $\boldsymbol{H_a: \mu \neq}$ _____

NOTE: When typing, you do not have to use subscripts, like I have done. I am fine with Ho and Ha.

State the Level of Significance

Every hypothesis test has a level of significance associated with it. Early on in the research process, before beginning to collect data, the researcher decides on this level of significance. It would be improper to change this level of significance after seeing your results! The Greek letter alpha α is used to represent level of significance. Most hypothesis tests use an alpha value of .05. When testing on life and death decisions is made, often an alpha value of .01 is chosen. Occasionally, we will see alpha = .10

Unless distinctly told otherwise, we will always choose an alpha = .05.

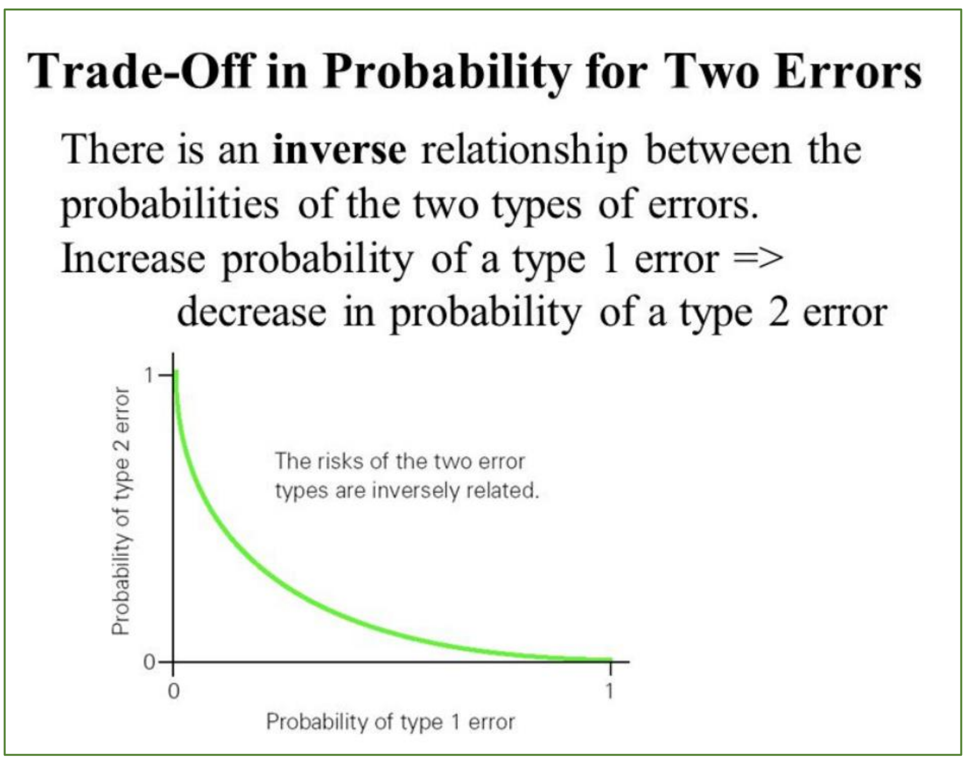
Alpha = α = level of significance/significance level = the probability of type 1 error

The level of significance, alpha, is the probability of Type 1 error. In other words, the level of significance, alpha, is the probability of rejecting the null hypothesis when it is true.

For example, a significance level of 0.05 indicates a 5% risk of concluding that a difference exists and we have enough evidence to support the alternative hypothesis (Ha), when there really is no difference.

Type I Error is the probability of deciding we **CAN** reject the H_o when, in fact, that decision to reject is actually a mistake. Alpha, α, is the probability of Type I Error.

Type II Error is the probability of deciding we **CANNOT** reject the H_o when, in fact, that decision is actually a mistake. Beta, β, is the probability of Type II Error. We will not learn to calculate beta in this course.



Indicate the Distribution that Will Be Used

When working with numerical data, we will study μ, the population mean; in other words, the true mean. We will always work on the students t-distribution to accomplish this and degrees of freedom df = n – 1.

NOTE: Degrees of freedom of an estimate is **the number of independent pieces of information that went into calculating the estimate**. It's not quite the same as the number of items in the sample. In order to get the df for the estimate, you have to subtract 1 from the number of items in the sample (df = n – 1).

SIDE NOTE: Next week, when we work with categorical data and study proportions, we will work on the Z-curve (the standard normal curve).

This is a left-tailed test.

$H_o: \mu =$ _____
$H_a: \mu <$ _____

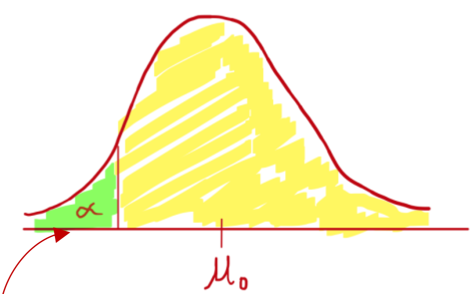
This is a right-tailed test.

$H_o: \mu =$ _____
$H_a: \mu >$ _____

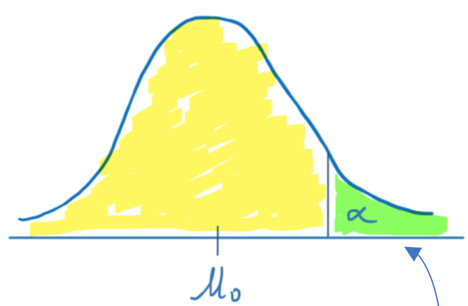
This is called a two-tailed test.

$H_o: \mu =$ _____
$H_a: \mu \neq$ _____

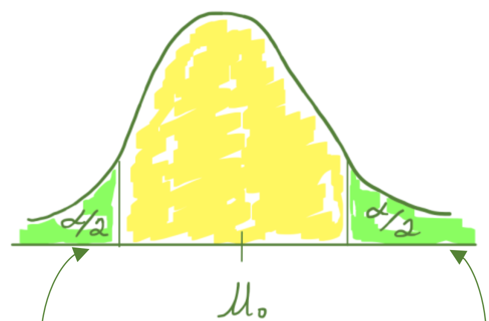
These are Student's T-curves depending on df = n -1. The hypothesized mean, μ_o, is at the center.



In a left-tailed test, alpha is in the left tail.



In a right-tailed test, alpha is in the right-tail.



In a two-tailed test, half of alpha is in each tail.

Here is the plan: We will analyze the sample data and determine if the sample is data is consistent with the hypothesis.

If the sample mean lands in the yellow highlighted region, then we conclude that the sample mean is consistent with the hypothesis; i.e., the sample mean is ***not statistically different*** from the hypothesized mean. This yellow-highlighted region is the region where we would say our sample data is really NOT THAT DIFFERENT from the hypothesized mean (the mean we were claiming in the null and at the center of the curve).

If the sample mean lands in the green highlighted region, then we have evidence that the sample mean is statistically different from the hypothesized mean (the mean we were claiming in the null and at the center of the curve). This green highlighted region is the region where we would say our sample data is ***statistically different*** from the hypothesized mean.

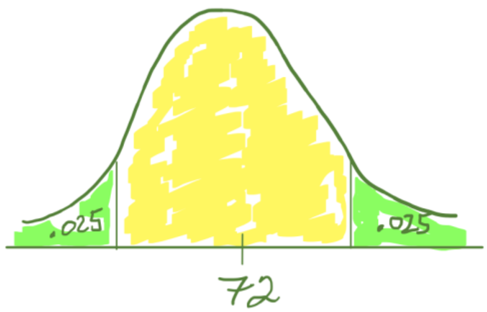
To determine where the sample mean lies relative to the hypothesized mean, we will find the corresponding T-value. This is called a T Test Statistic. I will define this soon.

Example 1 Continued – Method Step

It is believed that 72 is the average pulse rate for college-aged women. A group of students wonder if this long-held standard pulse rate value is still accurate (or has it changed) and gather a random sample of 35 pulse rates from college-aged women. The mean pulse rate was found to be $\bar{x}$ = 76.8 with a sample standard deviation of s = 11.62. Ultimately, we will show the complete testing process. For right now, show the method step (include Ho, Ha, alpha, which curve).

Method Step

H_o: μ = 72 ←72 is the hypothesized mean, μ_o. Notice there is always an equal sign (=) in Ho.
H_a: μ ≠ 72
Alpha = 0.05 ←this is the default; use it unless told otherwise
T-curve with df ←since we are studying means, we will be on the T-curve



← sketching the curve is optional; but may be useful.

Sample Step

Analyze the Sample Data using Technology

Use statistical technology to conduct the hypothesis test you stated in the Method Step; StatHelper instructions and Minitab 19 instructions can be found on the next pages. Paste the output in your work. The test statistic an d P-value outputs are needed.

Check to see if the Normality Assumption Has Been Met

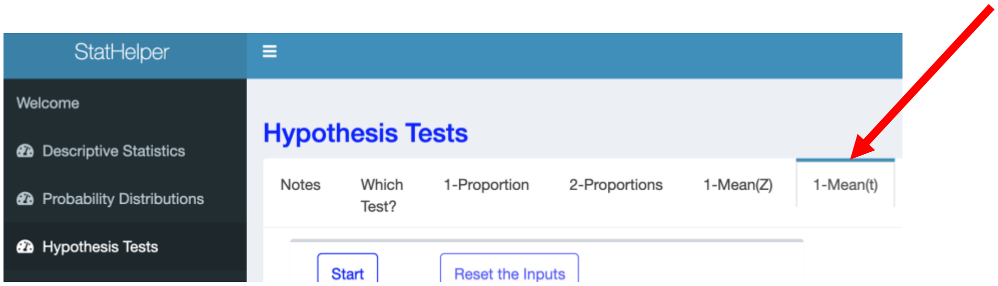
Summarized Data Provided

You will need to see if the sample size is large enough ($n \geq 30$) to assume a normal sampling distribution. The student's T-curve can be used if we have met the normality assumption.

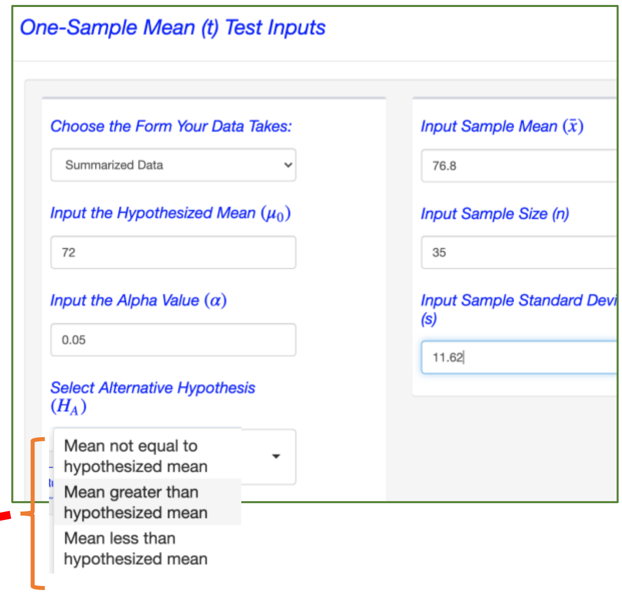
Raw Data Provided

You will need to see if the sample size is large enough ($n \geq 30$) to assume a normal sampling distribution. If it is not large enough, then build a normal probability plot (NPP) and see if the P-value is greater than .05. The student's T-curve can be used if we have met the normality assumption.

One Sample Mean (t) Test –with StatHelper



<u>StatHelper Instructions</u>



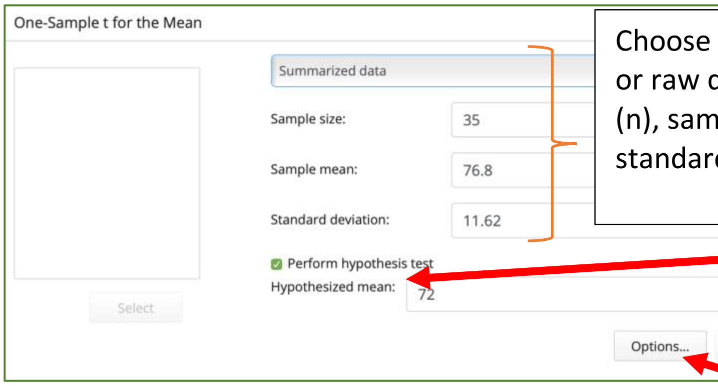
1. Choose Hypothesis Tests from the left side.
2. Choose **1-Mean(t)** from the options across the top.
3. Press Start.
4. Choose the form your data takes; Choose Summarized Data (input values) or Raw Data (choose file).
5. Input the Hypothesized Mean (μ_o)
6. Input the Alpha Value (α).
7. Select Alternative Hypothesis (H_A).
8. Click RUN. The work is under the WORK tab. The decision and conclusion are under the INTERPRETATION tab

Video Demonstrating Using StatHelper to conduct a hypothesis test from Prof. Coffey

Here is a video using StatHelper to conduct a hypothesis test:

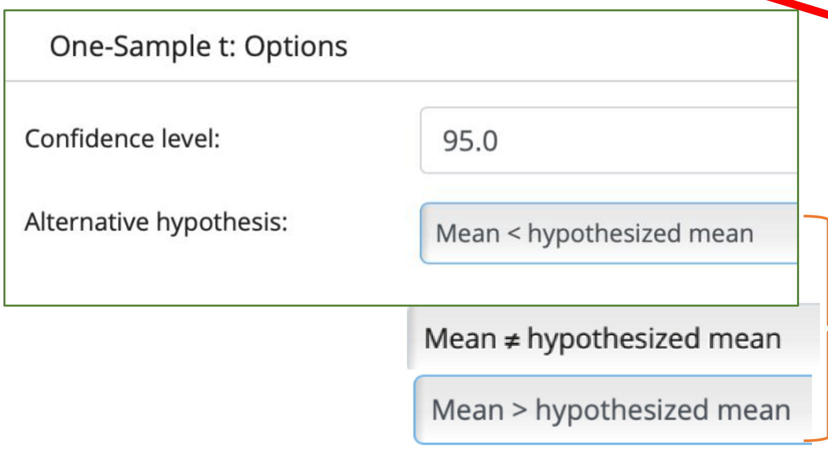
One-Sample Mean (t) Test –with Minitab 19

<u>Minitab 19 keystrokes</u>
Stat→Basic Statistics→ 1 Sample t



Choose between summarized data or raw data. Enter the sample size (n), sample mean ($\bar{x}$) and sample standard deviation (s).

Select: Perform hypothesis test
Enter μ_o, the hypothesized mean.



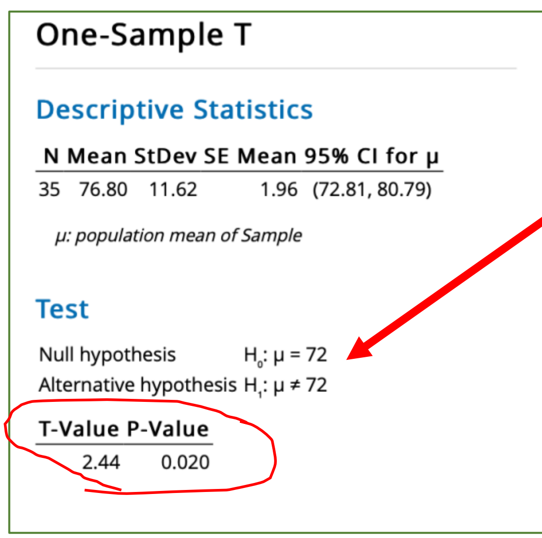
Select Options…

(note: confidence level does not need to be adjusted in testing)

Choose the correct Alternative Hypothesis.

Video Demonstrating Using Minitab 19 to conduct a hypothesis test from Prof. Coffey

Here is a video using Minitab to conduct a hypothesis test:



<u>Minitab One-Sample T Output</u>

Double check to be sure your Ho and Ha are as you intended.

The T-test statistic is labeled **T-Value**.
You will have to report df = n – 1.

The **P-Value** is how we will make a decision in this test.

Results Step

The T - Test Statistic

A test statistic is a measure of location and represents the number of standard errors (SE) away your sample mean is from what you are assuming in the null hypothesis (the hypothesized mean, μ_o).

The formula for the t test statistic is below. Always round the SE calculation to at least 3 decimal places. Round the test statistic calculation to 2 decimal places (round to the nearest hundredth)

$$t = \frac{\bar{x} - \mu_o}{s/\sqrt{n}} \qquad with\ df = n - 1$$

A negative test statistic means that your sample mean ($\bar{x}$) is ***<u>less than</u>*** the hypothesized mean (μ_0).

A positive test statistic means that your sample mean ($\bar{x}$) is ***<u>greater than</u>*** the hypothesized mean (μ_0).

Interpret the T - Test Statistic

It is important that you are able to interpret the t test statistic and state the degrees of freedom. You can use this template sentence (note that you will take the absolute value of the test statistic when you interpret it since the positive or negative sign indicates direction).

"My sample mean is ______ standard errors above/below __________ ."

State $|T\ test\ stat|$ — the hypothesized mean μ_0

- Leave off the +/- sign when interpreting the t test statistic.
- If it is a positive t test statistic, say "above".
- If it is a negative t test statistic, say "below".

If we were to try to make a decision at this point, we would need to know if the test statistic we obtained is statistically far enough away from the center of the curve (the hypothesized mean). This approach is called the 'Critical Value' approach. StatHelper calls this the 'Rejection Region' approach. You can learn more here, if interested. **NOTE** We will not use this approach to make a decision in hypothesis testing.

We will ALWAYS let technology obtain the Probability value or **P-value** associated with the test statistic we calculated. This is called the 'P-value' approach. Our e-text uses this approach, so you can learn more by reading the e-text or keep reading in these notes!

In this course, we will **<u>ALWAYS</u>** use the **P-value approach** to making decisions in Hypothesis Testing.

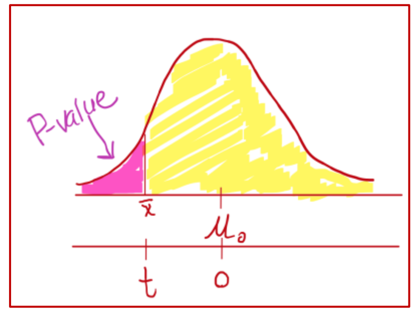
The P-value

A P-value is the probability of getting a statistic at least as extreme as the one that was obtained through sampling, assuming the null hypothesis is true. In other words, the P-value is the probability that, if the null hypothesis is true, the results from another randomly selected sample will be as extreme or more extreme as the results obtained from the given sample.

Here is the most casual way I can say it: The P-value is the area in the tail(s) of the curve beyond the test statistic that you found.

NOTE: Technology calculates the P-value for us. We *could* find it using Minitab's T-distribution or T-table. We do not have to. The P-value will be part of the hypothesis test output.

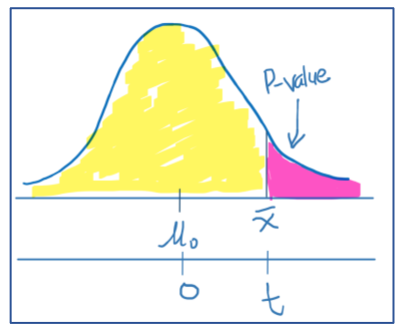
LEFT-TAILED TEST



If you are conducting a left-tailed test (the H_a has a 'less than' sign in it), then the P-value represents the area in the **left** tail beyond your test statistic.

To calculate a P-value on the T curve, the statistical technology will find the area beyond the t test statistic with df = n-1.

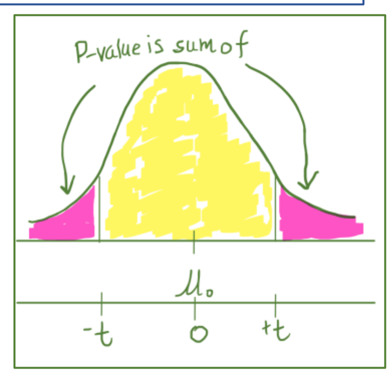
RIGHT-TAILED TEST



If you are conducting a right-tailed test (the H_a has a 'greater than' sign in it), then the P-value represents the area in the **right** tail beyond your test statistic.

To calculate a P-value on the T curve, statistical technology will find the area beyond the t test statistic with df = n-1.

TWO-TAILED TEST



If you are conducting a two-tailed test (the H_a has a 'NOT EQUAL TO' sign in it), then the P-value represents the area in the left and right tails beyond your test statistic. The areas are identical since the curve is symmetric. If you find one, just double it to get the P-value.

To calculate a P-value on the T curve, statistical technology will find the area beyond the test statistic **in both directions.**

Interpret the P-value

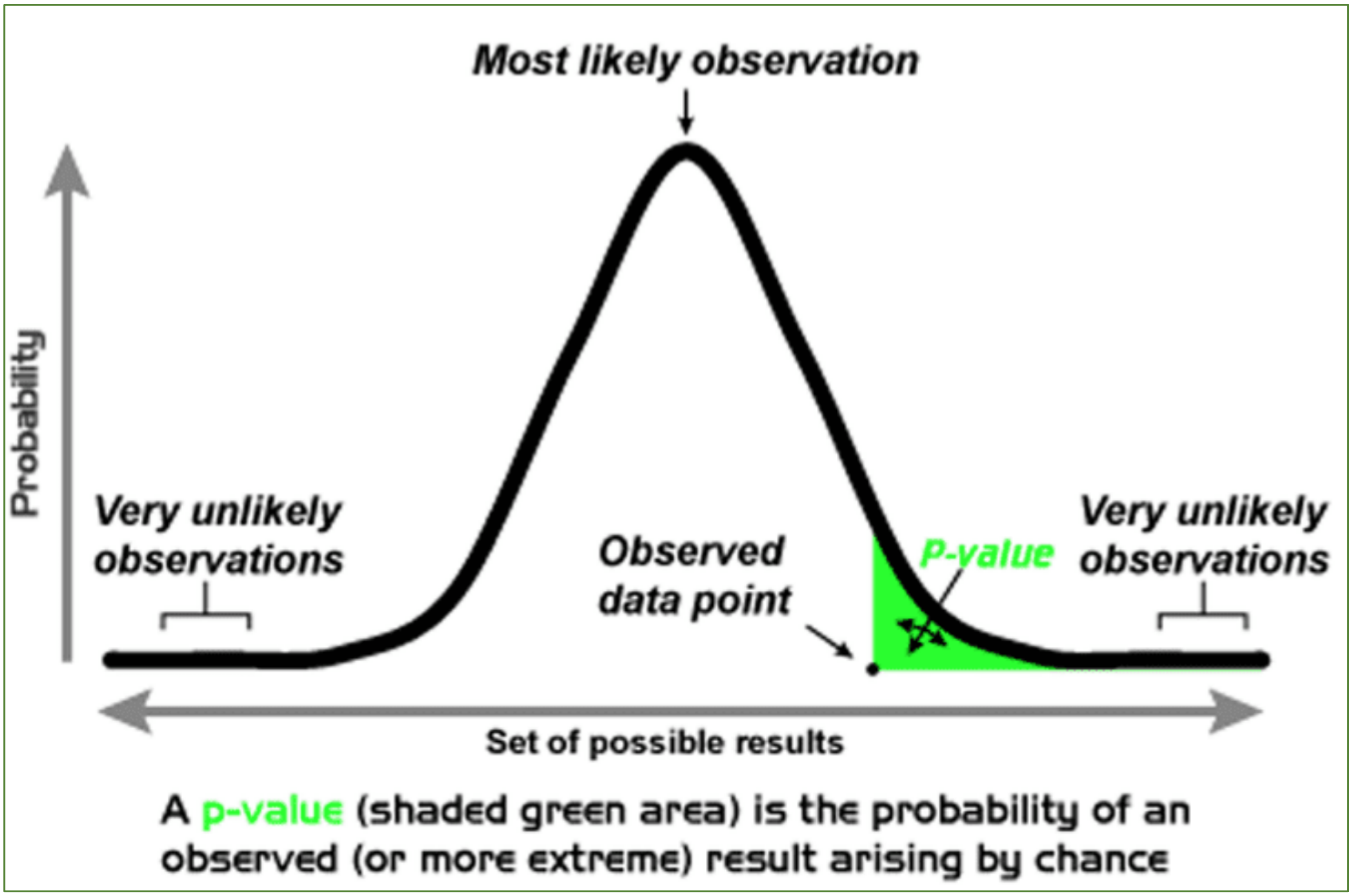
A P-value is the probability of obtaining a statistic (in this case, a sample mean) at least as extreme as the one that was obtained through sampling, ***assuming the null hypothesis is true***.

Our textbook describes it: The *p*-value is the probability that, if the null hypothesis is true, the results from another randomly selected sample will be as extreme or more extreme as the results obtained from the given sample.

NOTE: Technology calculates the P-value for us. P-value will be part of the hypothesis test output. **We need to be able to interpret the meaning behind the P-value.**

You may use the following P-value interpretation template:

"Assuming the true mean equals.....μ_o....., there is a ....P-value.... probability of getting a sample mean ($\bar{x}$) at least as extreme as the one we got from sampling."



Example 1 Continued – Sample and Results Steps

It is believed that 72 is the average pulse rate for college-aged women. A group of students wonder if this long-held standard pulse rate value is still accurate (or has it changed) and gather a random sample of 35 pulse rates from college-aged women. The mean pulse rate was found to be $\bar{x}$ = 76.8 with a sample standard deviation of s = 11.62. Ultimately, we will show the complete testing process. For right now, show the sample and results steps.

<u>Sample Step</u>

Since the sample size, n = 35, is large enough, we can assume a normal sampling distribution.

StatHelper: Hypothesis Tests → 1-Mean(t) → Enter information; be sure to select correct Ha; use P-value.

Minitab 10: Stat→Basic Statistics→1 Sample T → Select Perform Hypothesis Test; enter μ_o→Options; change alternative.

<u>StatHelper Output</u>

Null and Alternative Hypotheses

The null hypothesis is

$H_0 : \mu_0 = 72$

The alternative hypothesis is

$H_A : \mu_0 \neq 72$

The formula for the t-test statistic is

$$t_{Stat} = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

Using the data from the Descriptive Statistics tab, the test statistic becomes

$$t_{Stat} = \frac{(76.8 - 72)}{\frac{11.62}{\sqrt{35}}}$$

$$t_{Stat} = \frac{4.8}{1.964}$$

$$t_{Stat} = 2.444$$

For the p-value approach, the p-value is

$$\text{p-value} = 2P\left(t \geq |t_{Stat}|\right)$$
$$\text{p-value} = 2(1 - P(t \leq |2.444|))$$
$$\text{p-value} = 2\,(1 - 0.9901)$$
$$\text{p-value} = 2 * 0.0099$$
$$\text{p-value} = 0.0199$$

<u>Minitab 19 Output</u>

One-Sample T

Descriptive Statistics

N	Mean	StDev	SE Mean	95% CI for μ
35	76.80	11.62	1.96	(72.81, 80.79)

μ: population mean of Sample

Test

Null hypothesis H_0: μ = 72
Alternative hypothesis H_1: μ ≠ 72

T-Value	P-Value
2.44	0.020

<u>Results Step</u>

$$\boxed{t = \frac{\bar{x} - \mu_o}{s/\sqrt{n}} = \frac{76.8 - 72}{11.62/\sqrt{35}} = \frac{4.8}{1.964} = 2.44}$$

The test statistic is T = 2.44, df = 34

My sample mean is 2.44 standard errors <u>above</u> the hypothesized mean pulse of 72.

The P-value = 0.020

Assuming the true mean pulse is 72, there is a 0.02 probability of getting a sample mean at least as extreme as the one we got from sampling.

Conclusion Step

Make a Decision

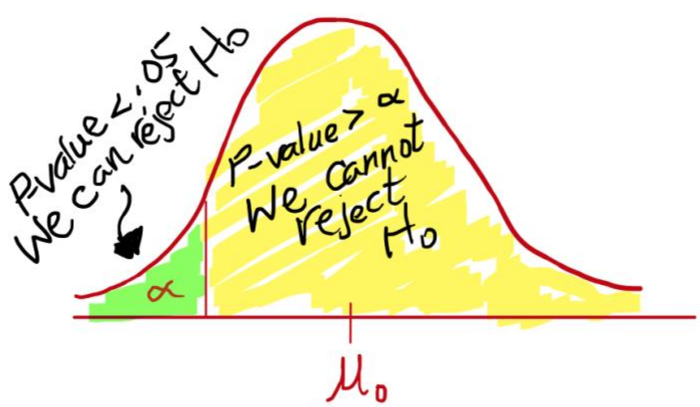
Statisticians developed the P-value approach to making a decision in hypothesis testing. [There is another method of making a decision. It is called the rejection region approach or critical value approach. We will not use this method in this course. NOTE: StatHelper does show both the P-value and rejection region approach.]

The P-value Approach

If the P-value $\leq \alpha$, then we will decide that we CAN REJECT the null hypothesis (in favor of the alternative hypothesis).

In other words:
<u>If the P-value is small</u>, then it would be <u>considered a rare event</u> to get a sample mean, $\bar{x}$, if the null were true. This is evidence that the null is not true (hence, we CAN reject the null). In other words, if the P-value is less than the significance level (**P-value < α**), then we <u>**can**</u> <u>**reject Ho**</u>.

If the P-value > α, then we will decide that we CANNOT reject the null hypothesis (in other words, we FAIL TO reject the null or DO NOT reject the null)
In other words:
<u>If the P-value is big enough</u>, then it would <u>**NOT** be considered a RARE event</u> to get a sample mean, $\bar{x}$, if the null were true. This is evidence that the null is probably true (hence, we cannot reject the null). In other words, if the P-value is greater than the significance level (**P-value > α**), then we <u>**CANNOT reject Ho**</u>.



To assist you with making a decision, you may do the following three things:

1. First ask and answer this question: Is your P-value ≤ α?

Example: In a test with α = 0.05 and P-value = 0.0397, we would say YES since 0.0397 < 0.05.

2. Make a Decision.
If your answer is **YES**, then the decision for this test is that **we CAN reject the null hypothesis (H_o)**.
If your answer is **NO**, then the decision for this test is that **we CANNOT reject the null hypothesis (H_o)**.

If P-value ≤ α, then we **CAN** reject Ho and, therefore, we will conclude that the sample **DOES** provide sufficient evidence to support the alternative hypothesis (Ha).

If P-value > α, then we **CANNOT** reject Ho and, therefore, we will conclude that the sample **DOES NOT** provide sufficient evidence to support the alternative hypothesis (Ha).

3. Write a Statistical Conclusion using this template.

The following template is a statistical conclusion written in the context of the problem and always indicating if we DO HAVE enough evidence to support the goal or if we DO NOT have enough evidence to support the goal.

"At the ____% level of significance, the sample data DOES / DOES NOT provide sufficient evidence to say that the true mean ……………{restate the goal in words}…………………".

Example 1 Continued –Conclusion Step

It is believed that 72 is the average pulse rate for college-aged women. A group of students wonder if this long-held standard pulse rate value is still accurate (or has it changed) and gather a random sample of 35 pulse rates from college-aged women. The mean pulse rate was found to be $\bar{x}$ = 76.8 with a sample standard deviation of s = 11.62. Ultimately, we will show the complete testing process. For right now, show the conclusion step.

Conclusion Step

Since P = 0.02 is less than α = 0.05, Yes, we can reject the null.

Statistical Conclusion:

At the 5% level of significance, the sample data DOES provide sufficient evidence to say that the true mean pulse rate for college-aged women is different from the pulse of 72.

Everyday Conclusion

We have evidence that the long-held standard pulse rate of 72 is not accurate. It has changed!

Example 1 –SHOWING THE COMPLETE TESTING PROCESS

It is believed that 72 is the average pulse rate for college-aged women. A group of students wonder if this long-held standard pulse rate value is still accurate (or has it changed) and gather a random sample of 35 pulse rates from college-aged women. The mean pulse rate was found to be $\bar{x}$ = 76.8 with a sample standard deviation of s = 11.62. Complete the appropriate hypothesis test at the 5% level of significance. Show the complete testing process.

Population Step

The variable is numerical and is the pulse rate for college-aged women.

μ = the population mean pulse rate for all college-aged women.

Goal: Test to see if there is support for saying that μ ≠ 72.

Method Step

H_o: μ = 72

H_a: μ ≠ 72

Alpha = 0.05

T-curve with df

Sample Step

Since the sample size, n = 35, is large enough, we can assume a normal sampling distribution.

StatHelper Output

Null and Alternative Hypotheses

The null hypothesis is

$H_0 : \mu_0 = 72$

The alternative hypothesis is

$H_A : \mu_0 \neq 72$

The formula for the t-test statistic is

$$t_{Stat} = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

Using the data from the Descriptive Statistics tab, the test statistic becomes

$$t_{Stat} = \frac{(76.8 - 72)}{\frac{11.62}{\sqrt{35}}}$$

$$t_{Stat} = \frac{4.8}{1.964}$$

$$t_{Stat} = 2.444$$

For the p-value approach, the p-value is

$$\text{p-value} = 2P\,(t \geq |t_{Stat}|)$$
$$\text{p-value} = 2(1 - P\,(t \leq |2.444|))$$
$$\text{p-value} = 2\,(1 - 0.9901)$$
$$\text{p-value} = 2 * 0.0099$$
$$\text{p-value} = 0.0199$$

Minitab 19 Output

One-Sample T

Descriptive Statistics

N	Mean	StDev	SE Mean	95% CI for μ
35	76.80	11.62	1.96	(72.81, 80.79)

μ: population mean of Sample

Test

Null hypothesis H_0: μ = 72

Alternative hypothesis H_1: μ ≠ 72

T-Value	P-Value
2.44	0.020

Results Step

The test statistic is T = 2.44, df = 34

My sample mean is 2.44 standard errors above the hypothesized mean pulse of 72.

The P-value = 0.020

Assuming the true mean pulse is 72, there is a 0.02 probability of getting a sample mean at least as extreme as the one we got from sampling.

Conclusion Step

Since P = 0.02 is less than α = 0.05, Yes, we can reject the null.

Statistical Conclusion:

At the 5% level of significance, the sample data DOES provide sufficient evidence to say that the true mean pulse rate for college-aged women is different from the pulse of 72.

Everyday Conclusion

We have evidence that the long-held standard pulse rate of 72 is not accurate. It has changed!

Video Showing Hypothesis Test Example 1 from Prof. Coffey

Here is a video showing hypothesis test example 1: https://youtu.be/5pyiNOTtudk

Here is the same video with interpreting: https://youtu.be/Nq8sjjdGwzE

Example 2 – Left-Tailed Test

A healthy body of water has 4.9 ppm or more of dissolved oxygen available for aquatic life. Local officials believe that the Genesee River is healthy. A RIT researcher disagrees and needs evidence that supports her theory so she can get local officials to take action. She tests water at 48 randomly selected locations in the Genesee River and finds the $\bar{x}$ = 4.85 ppm and the standard deviation **s** = .45 ppm. Show the complete testing process.

Population Step

The variable is numerical and is the amount of dissolved oxygen in the Genesee River

μ = the population mean amount of dissolved oxygen in the Genesee River

Goal: Test to see if there is support for saying that μ < 4.9.

[NOTE: the hypothesized mean, μ_o, is 4.9 ppm for this problem]

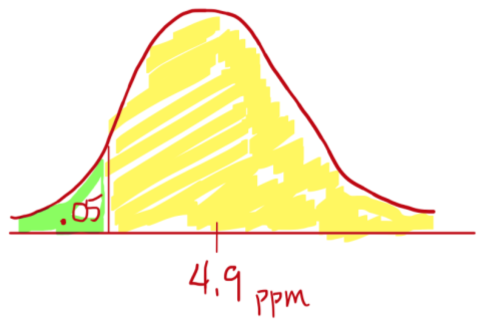
Method Step

H_o: μ = 4.9

H_a: μ < 4.9 left-tailed test

Alpha = 0.05

T-curve with df



Sample Step

Calculations for a 1-Sample t-test

Null and Alternative Hypotheses

The null hypothesis is

$$H_0 : \mu_0 = 4.9$$

The alternative hypothesis is

$$H_A : \mu_0 < 4.9$$

The formula for the t-test statistic is

$$t_{Stat} = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

Using the data from the Descriptive Statistics

$$t_{Stat} = \frac{(4.85 - 4.9)}{\frac{0.45}{\sqrt{48}}}$$

$$t_{Stat} = \frac{-0.05}{0.065}$$

$$t_{Stat} = -0.7698$$

For the p-value approach, the p-value is

$$\text{p-value} = P(t \leq t_{Stat})$$
$$\text{p-value} = P(t \leq -0.7698)$$
$$\text{p-value} = 0.2226$$

WORKSHEET 1

One-Sample T

Descriptive Statistics

N	Mean	StDev	SE Mean	95% Upper Bound for μ
48	4.8500	0.4500	0.0650	4.9590

μ: population mean of Sample

Test

Null hypothesis H_0: μ = 4.9
Alternative hypothesis H_1: μ < 4.9

T-Value	P-Value
-0.77	0.223

Since the sample size, n = 35, is large enough, we can assume a normal sampling distribution.

Results Step
T = -.77, df = 47
My sample mean is .75 standard errors below the hypothesized mean dissolved oxygen of 4.9 ppm.
P-value = .223
Assuming the true mean dissolved oxygen is 4.9 ppm, there is a .223 probability of getting a sample mean at least as extreme as the one we got from sampling.

Conclusion Step
Compare the P-value to alpha; Is the P-value $\leq$ alpha? No, it is not.
We cannot reject the null hypothesis.

At the 5% level of significance, the sample data DOES NOT provide sufficient evidence to say that the true mean dissolved oxygen is less than 4.9 ppm.

Everyday Conclusion
The RIT researcher does not have enough evidence to support her theory!

Some additional thoughts on this test from Prof. Coffey:
The researcher is comfortable with a level of significance of .05. This is decided before data is collected. She is hoping to get evidence to show local officials that the Genesee River is not as healthy as they believe it is and that improvements to environmental policies may be warranted! The sample data, however, resulted in a sample mean that was only .7698 standard errors below 4.9 ppm. Anytime you get a P-value greater than your alpha you know that your sample data is not statistically different than what was being claimed in the null. The sample mean for these water samples results in a sample mean that is literally less than 4.9. This, however, could happen by chance alone. Our test is a way of determining if the sample mean is statistically less than 4.9. In this case, it is not. We are UNABLE to show evidence that the true mean dissolved oxygen is less than 4.9 ppm.

Video Showing Hypothesis Test Example 2 from Prof. Coffey

Here is a video showing hypothesis test Example 2: https://youtu.be/cmy4Nby2n1s

Here is the same video with interpreting: https://youtu.be/EfRuMT--MgY

Example 3 – Right-Tailed Test

A 30-year career college football coach has always assumed that the mean weight that his players can bench press is 275 pounds. His new assistant coach believes that their players can bench press more. A random sample of 30 teammates recorded their estimated maximum lift on the bench press exercise. The data can be found in the data file: **9 Week 9 STAT 145 Data** under the sheet: **Bench Press**.

Conduct a hypothesis test using a 5% level of significance to determine if the bench press mean is more than 275 pounds. Show the complete testing process.

Population Step

The variable is the bench press weight, in pounds, It is numerical.

μ = the true mean bench press weight for this college's football players

Goal: test to see if $\mu > 275$ pounds.

Method Step

H_o: $\mu = 275$

H_a: $\mu > 275$ Right-tailed test

Alpha = 0.05

T-curve with df

Sample Step

Since the sample size is 30, the CLT allows us to assume a normal sampling distribution. We have met the normality assumption.

BENCH PRESS

One-Sample T: Bench Press Weight

Descriptive Statistics

N	Mean	StDev	SE Mean	95% Lower Bound for μ
30	287.3	57.6	10.5	269.4

μ: population mean of Bench Press Weight

Test

Null hypothesis H_0: $\mu = 275$

Alternative hypothesis H_1: $\mu > 275$

T-Value	P-Value
1.17	0.126

Results Step

T = 1.17, df = 29

My sample mean is 1.17 standard errors above the hypothesized mean weight of 275 pounds.

P-value = .126

Assuming the true mean bench press weight is 275 pounds, there is a .126 probability of getting a sample mean at least as extreme as the one we got from sampling.

Conclusion Step

Compare the P-value to alpha; Is the P-value $\leq$ alpha? No, it is not.

We cannot reject the null hypothesis.

At the 5% level of significance, the sample data DOES NOT provide sufficient evidence to say that the true mean bench press weight is greater than 275 pounds.

Everyday Conclusion

The assistant coach was not correct! The average bench press weight has not increased significantly.

Example 4 – Two-Tailed Test with Raw Data

In 1851, a German doctor established what is currently used as the 'normal' human body temperature of 98.6°F. Physician Assistant students at RIT believe the normal body temperature has changed. Core body temperatures (in degrees Fahrenheit) for a random sample of healthy, resting adults were collected and are given in the table below and in our Week 9 data set under the sheet: **Body Temp**.

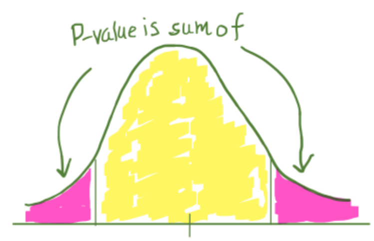
97.9	98.4	98.5	99.1	98.5	97.9	98.7	98.6	98.1
97.3	97.4	97.6	99.2	98.2	98.3	98.7	97.1	98.8

Conduct the appropriate hypothesis test using 5% level of significance. Show the complete testing process.

<u>Population Step</u>

The variable is body temperature, in Fahrenheit, and it is a numerical variable.
μ = the true mean body temperature of healthy resting adults
The goal of the test: To see if the true mean body temperature of healthy, resting adults has changed from 98.6°. [In other words, test to see if $\mu \neq 98.6$]

<u>Method Step</u>

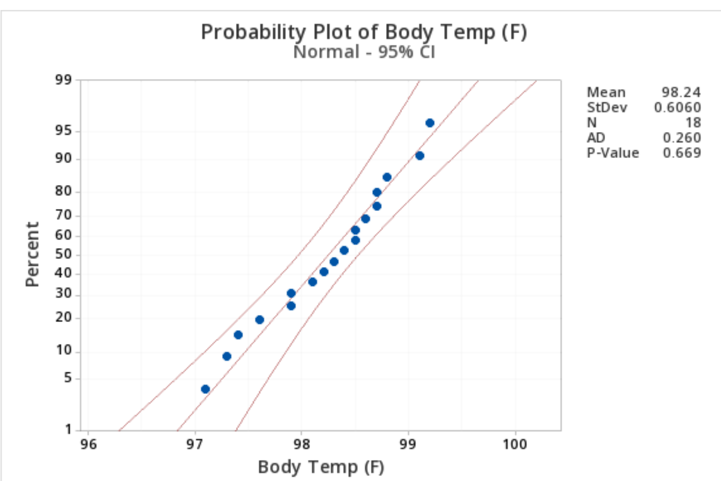


98.6°

H_o: $\mu = 98.6$
H_a: $\mu \neq 98.6$ Two-tailed test
Alpha = 0.05
T-curve with df

<u>Sample Step</u>

With a P-value = .669, we can assume a normal model.



BODY TEMP

One-Sample T: Body Temp (F)

Descriptive Statistics

N	Mean	StDev	SE Mean	95% CI for μ
18	98.239	0.606	0.143	(97.938, 98.540)

μ: population mean of Body Temp (F)

Test

Null hypothesis H_0: $\mu = 98.6$
Alternative hypothesis H_1: $\mu \neq 98.6$

T-Value	P-Value
-2.53	0.022

Results Step

T Test statistic = -2.53 with df = 17

The sample mean is 2.53 standard errors below the hypothesized mean of 98.6°F.

P-value = .022

Assuming the true mean body temperature is 98.6°F, there is a .022 probability of getting a sample mean at least as extreme as the one we got from sampling.

Conclusion Step

Compare the P-value to alpha; Is the P-value $\leq$ alpha? Yes, it is.

We can reject the null hypothesis.

At the 5% level of significance, the sample data DOES provide sufficient evidence to say that the true mean body temperature of healthy, resting adults is different than 98.6°F.

Everyday English---conclusion

The RIT students are correct…body temps are different than the doctor established in 1851. In fact, they are statistically less (as seen by the negative test statistic).

NOTE: A two-tailed test can be answered using another approach: The Confidence Interval Approach. Notice that the 2-sided confidence interval does not include 98.6°F. Since 98.6°F is not in the CI, we have evidence that the body temperature has changed.

BODY TEMP

One-Sample T: Body Temp (F)

Descriptive Statistics

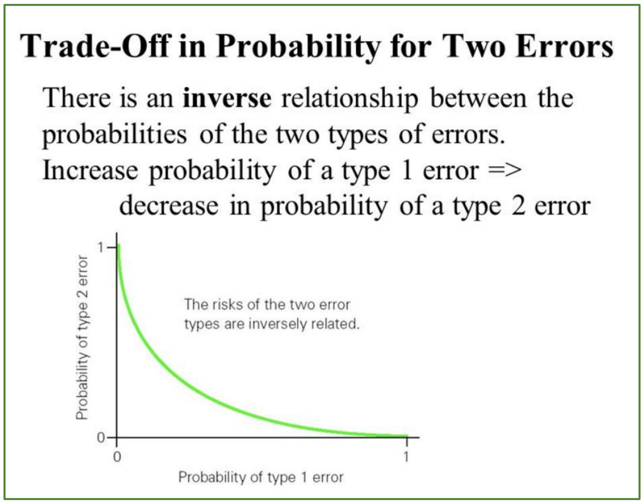
N	Mean	StDev	SE Mean	95% CI for μ
18	98.239	0.606	0.143	(97.938, 98.540)

μ: population mean of Body Temp (F)

Type I Error and Type II Error

Type I Error is the probability of deciding we **CAN** reject the H_o when, in fact, that decision to reject is actually a mistake. Alpha, α, is the probability of Type I Error.

Type II Error is the probability of deciding we **CANNOT** reject the H_o when, in fact, that decision is actually a mistake. Beta, β, is the probability of Type II Error. We will not learn to calculate beta in this course.



FOUR OUTCOMES

	The null is actually TRUE	The alternative is actually true
If P-value > α and we **CANNOT** Reject the null	CORRECT CONCLUSION	A **Type II Error** occurs when you decide you CANNOT reject the null (when you should have rejected the null) because the alternative is true. The probability of Type II error is beta (β), a value you can calculate but won't be calculated in this course.
If P-value < α and we **CAN** Reject the null	A **Type I Error** occurs when you… reject the null when the null is true. The probability of Type I error is alpha (α), the level of significance.	CORRECT CONCLUSION