

Chapter 3: Introduction to Hypothesis Testing

Vocabulary

Population: The collection of all people or objects to be studied.

Sample: Collecting data from a small subgroup of the population.

Random Sample: Sample data collected in such a way that everyone in the population has an equal chance to be included.

Statistic: A number calculated from sample data in order to understand the characteristics of the data. For example, a sample mean average, a sample standard deviation, or a sample percentage.

Parameter: A number that describes the characteristics of a population like a population mean or a population percentage. Can be calculated from an unbiased census, but is often just a guess about the population.

Hypothesis Test: A procedure for testing a claim about a population.

Null Hypothesis (H_0): A statement about the population that involves equality. It is often a statement about “no change”, “no relationship” or “no effect”.

Alternative Hypothesis (H_A or H_1): A statement about the population that does not involve equality. It is often a statement about a “significant difference”, “significant change”, “relationship” or “effect”.

Population Claim: What someone thinks is true about a population.

Test Statistic: A number calculated in order to determine if the sample data significantly disagrees with the null hypothesis. There are a variety of different test statistics depending on the type of data.

One-Population Proportion Test Statistic (z): The sample proportion is this many standard errors above or below the population proportion in the null hypothesis.

One-Population Mean Test Statistic (t): The sample mean is this many standard errors above or below the population mean in the null hypothesis.

Critical Value: A number we compare our test statistic to in order to determine significance. In a sampling distribution or a theoretical distribution approximating the sampling distribution, the critical value shows us where the tail or tails are. The test statistic must fall in the tail to be significant.

Sampling Variability: Also called “random chance”. The principle that random samples from the same population will usually be different and give very different statistics. The random samples will usually be different than the population parameter.

P-value: The probability of getting the sample data or more extreme because of sampling variability (by random chance) if the null hypothesis is true.

Significance Level (α): Also called the Alpha Level. This is the probability of making a type 1 error. The P-value is compared to this number to determine significance and sampling variability. If the P-value is lower than the significance level, then the sample data significantly disagrees with the null hypothesis and is unlikely to have happened because of sampling variability.

Randomized Simulation: A technique for visualizing sampling variability in a hypothesis test. The computer assumes the null hypothesis is true, and then generates random samples. If the sample data or test statistic falls in the tail, then the sample data significantly disagrees with the null hypothesis. This technique can also calculate the P-value and standard error without a formula.



Type 1 Error: When biased sample data leads you to support the alternative hypothesis when the alternative hypothesis is actually wrong in the population.

Type 2 Error: When biased sample data leads you fail to reject the null hypothesis when the null hypothesis is actually wrong in the population.

Beta Level (β): The probability of making a type 2 error.

Conclusion: A final statement in a hypothesis test that addresses the claim and evidence.

Introduction: The goal of statistics is to learn about the world around us. While we may collect sample data, our goal is not just to analyze sample data. We want to know what is happening in the population. Sometimes, people make guesses about what they think is happening in the population. These guesses are often called “claims”. In this chapter, we will discuss the process of hypothesis testing. A hypothesis test is a scientific procedure for using representative random sample data to investigate claims about populations. A hypothesis test involves many difficult concepts and has many steps. It is difficult to learn all of hypothesis testing at once. For this reason, we will be learning the steps for hypothesis testing one at a time. Eventually we will put all of the parts together and complete the hypothesis test from start to finish.

Section 3A – Null and Alternative Hypothesis

Vocabulary

Population: The collection of all people or objects to be studied.

Census: Collecting data from everyone in a population.

Sample: Collecting data from a small subgroup of the population.

Statistic: A number calculated from sample data in order to understand the characteristics of the data.
For example, a sample mean average, a sample standard deviation, or a sample percentage.

Parameter: A number that describes the characteristics of a population like a population mean or a population percentage. Can be calculated from an unbiased census, but is often just a guess about the population.

Hypothesis Test: A procedure for testing a claim about a population.

Null Hypothesis (H_0): A statement about the population that involves equality. It is often a statement about “no change”, “no relationship” or “no effect”.

Alternative Hypothesis (H_A or H_1): A statement about the population that does not involve equality. It is often a statement about a “significant difference”, “significant change”, “relationship” or “effect”.

Population Claim: What someone thinks is true about a population.

Introduction:

Remember, a hypothesis test is a procedure for checking what someone has said about the population. This is often called the population “claim”. From this claim, we will need to identify two opposing views. The claim and the rival hypothesis. If the claim is not true, then what would be true? These two opposing views are referred to as the null hypothesis (H_0) and the alternative hypothesis (H_A or H_1). The symbol often used for “null hypothesis” is “ H_0 ”. The symbol used to represent the alternative hypothesis is “ H_A ” or “ H_1 ”. I prefer “ H_A ” for alternative hypothesis.



Null Hypothesis (H_0): A statement about the population that involves equality. It is often a statement about “no change”, “no relationship” or “no effect”.

Alternative Hypothesis (H_A): A statement about the population that does not involve equality. It is often a statement about a “significant difference”, “significant change”, “relationship” or “effect”.

Important Notes:

- In the last chapter, we talked about symbols that represent population parameters like the population mean (μ), the population proportion (π) or the population standard deviation (σ). The null and alternative hypothesis are competing ideas about the population and only involve population parameters like μ , π or σ . The sample data and sample statistics like the sample proportion (\hat{p}), the sample mean (\bar{x}) or the sample standard deviation (s) are never part of the null or alternative hypotheses.

A hypothesis test is a procedure for deciding between two opposing views about the population. Sometimes the person will tell you the two opposing views, but there will be one view that the person thinks is true or wants you as the data scientist to give evidence toward. This is called the “claim”.

Claim: What the person now thinks is true about the population. The claim can be a question that someone needs to figure out. It can also be an opinion about the population that they want you to investigate.

Let’s go over some basic steps to writing the null and alternative hypothesis.

Steps for finding the Null and Alternative Hypothesis

1. Write down the two competing views about the population in symbolic language. Make sure to determine if it is one-population, or two-population, or more and the correct letter (parameter) to use. If the person only gives you his or her claim, we will often use the opposite as the rival hypothesis or competing view.
2. Write the word “claim” next to what the person thinks is true or what they are asking you to provide evidence toward.
3. The statement that has “=” or “ \geq ” or “ \leq ” is the null hypothesis. It is usually “=”. Put an “ H_0 ” next to it. Remember, the null hypothesis is usually a statement about no change, no effect, or not related. That is why the null hypothesis is often given with “=”.
4. The statement that has “ \neq ” or “ $<$ ” or “ $>$ ” is the alternative hypothesis. Put an “ H_A ” next to it. This is usually a statement about something changing or being related or having a significant effect.
5. Determine what type of test you are dealing with. Is it right-tailed, left-tailed or two-tailed? In a hypothesis test, we are often attempting to “reject the null hypothesis”. This would happen if we believed the alternative hypothesis is correct. For this reason, the alternative hypothesis determines what type of test you are doing. In a one-population or two-population test, if H_A is less than ($<$), it is a left tailed test. (Notice “ $<$ ” points to the left). If H_A is greater than ($>$), it is a right tailed test. (Notice “ $>$ ” points to the right). If H_A is not equal (\neq), it is a two-tailed test. (not equal means less than or greater than.) Things get more complicated if we are dealing with 3 or more populations. We will deal with these cases in the next chapter.
 - Left-tailed test: H_A is less than ($<$)
 - Right-tailed test: H_A is greater than ($>$)
 - Two-tailed test: H_A is less than (\neq)

Symbols for population parameters:

μ (population mean)

π or p (population proportion/percentage)

σ (population standard deviation)

σ^2 (population variance)



Important Notes

- Never use a sample statistic (\bar{x}, \hat{p}, s). Remember a null and alternative hypothesis are statements about the population parameters (μ, π, σ).
- The symbol “<” means less than. Notice the symbol for “less than” points to the left. The opposite of less than is “≥” (greater than or equal to).
- The symbol “>” means less than. Notice the symbol for “greater than” points to the right. The opposite of less than is “≤” (less than or equal to).
- For one-population, always write the letter on the left side and the number on the right side. $\mu < 50$ (never as $50 > \mu$)
- For two-population, always put population 1 on the left side and population 2 on the right side. $\pi_1 > \pi_2$ (never as $\pi_2 < \pi_1$).

Three Types of Hypothesis Tests

- Hypothesis Tests are designated as one of three types. It is important to know what type of test you are doing.
- Note that the alternative hypothesis H_A decides the type of test. In a hypothesis test, we are often attempting to “reject the null hypothesis”. This would happen if we believed the alternative hypothesis is correct. For this reason, the alternative hypothesis determines what type of test you are doing. In a one-population or two-population test, if H_A is less than (<), it is a left tailed test. (Notice “<” points to the left). If H_A is greater than (>), it is a right tailed test. (Notice “>” points to the right). If H_A is not equal (\neq), it is a two-tailed test. (not equal means less than or greater than.)

Left-tailed test: H_A is less than (<) (points to the left)

Right-tailed test: H_A is greater than (>) (points to the right)

Two-tailed test: H_A is less than (\neq)

Note: If you have more than two populations, things become more complicated. For example, ten populations will often be condensed into one right-tailed test. We will explain more about these tests in the next chapter.

Example 1: Auto Magazine Article

“The population mean average weight of car transmissions used to be about 300 pounds. With more and more small car options, we think the population mean average weight of car transmissions has decreased.”

Step 1: The population views are about one population mean (μ). Notice there are two opposing views are the population mean given. Write down the two opposing views in symbolic language. Notice again that “decreased” means less than “<” which points to the left. Make sure to put the symbol on the left and the number on the right. (“ $\mu < 300$ ” NOT “ $300 > \mu$ ”). Notice also that the opposing view is that it used to be exactly 300 pounds. They never believed that the population mean average was more than 300 pounds. It seems more appropriate to use “=” as the opposing view instead of “≥”.

$$\mu = 300$$

$$\mu < 300$$

Step 2: Let’s identify the claim. It seems they think that it used to be 300 pounds, but now they think the population mean is really less than 300 pounds. This would be the claim.

$$\mu = 300$$

$$\mu < 300 \text{ (Claim)}$$



Step 3: The statement that has “=” or “≥” or “≤” is the null hypothesis. Put an “ H_0 ” next to it. Notice that “claim” has nothing to do with a statement being a null or alternative hypothesis. It is the symbol that decides H_0 .

$H_0: \mu = 300$
 $\mu < 300$ (Claim)

Step 4: The statement that has “≠” or “<” or “>” is the alternative hypothesis. Put an “ H_A ” next to it. Notice that “claim” has nothing to do with a statement being a null or alternative hypothesis. It is the symbol that decides H_A .

$H_0: \mu = 300$
 $H_A: \mu < 300$ (Claim)

Step 5: Determine the type of test you are doing. Remember, the alternative hypothesis determines the type of test. If H_A is less than (<), it is a left tailed test. (Notice “<” points to the left). If H_A is greater than (>), it is a right tailed test. (Notice “>” points to the right). If H_A is not equal (≠), it is a two-tailed test. (not equal means less than or greater than.)

In this problem, the alternative hypothesis is less than (<) which points to the left. So this is a left tailed test.

$H_0: \mu = 300$
 $H_A: \mu < 300$ (Claim)

Left Tailed Test

Example 2: Medication Side Effects

“The FDA says that about 2.5% of people that take this medicine will have serious side effects”.

Step 1: The population views are about one population proportion (π or p). Notice that only one view about the population proportion is given. In this case we will have to think about opposites. The FDA seems to be making a claim about equaling 2.5% (0.025). They did not specify higher or lower. The opposite of equal is not equal (≠) so we will use that as our opposing view. Write down the two opposing views in symbolic language. Population percentage claims are usually written as a decimal proportion. Make sure to put the symbol on the left and the number on the right. (“ $\pi = 0.025$ ” NOT “ $0.025 = \pi$ ”).

$\pi = 0.025$
 $\pi \neq 0.025$

OR

$p = 0.025$
 $p \neq 0.025$

Step 2: Let’s identify the claim. It seems that the FDA thinks it is 2.5%. Don’t let the word “about” change your mind about this. Remember, we never know anything definite about millions of people. Language like “about” or “around” is often used in population claims. They did not specify higher or lower, so it is still an equal to claim.

$\pi = 0.025$ (Claim)
 $\pi \neq 0.025$

OR

$p = 0.025$ (Claim)
 $p \neq 0.025$



Step 3: The statement that has “=” or “≥” or “≤” is the null hypothesis. Put an “ H_0 ” next to it. Notice that “claim” has nothing to do with a statement being a null or alternative hypothesis. It is the symbol that decides H_0 .

H_0 : $\pi = 0.025$ (Claim)
 $\pi \neq 0.025$

OR

H_0 : $p = 0.025$ (Claim)
 $p \neq 0.025$

Step 4: The statement that has “ \neq ” or “ $<$ ” or “ $>$ ” is the alternative hypothesis. Put an “ H_A ” next to it. Notice that “claim” has nothing to do with a statement being a null or alternative hypothesis. It is the symbol that decides H_A .

H_0 : $\pi = 0.025$ (Claim)
 H_A : $\pi \neq 0.025$

OR

H_0 : $p = 0.025$ (Claim)
 H_A : $p \neq 0.025$

Step 5: Determine the type of test you are doing. Remember, the alternative hypothesis determines the type of test. If H_A is less than ($<$), it is a left tailed test. (Notice “ $<$ ” points to the left). If H_A is greater than ($>$), it is a right tailed test. (Notice “ $>$ ” points to the right). If H_A is not equal (\neq), it is a two-tailed test. (not equal means less than or greater than.)

In this problem, the alternative hypothesis is not equal (\neq), so this is a two-tailed test.

H_0 : $\pi = 0.025$ (Claim)
 H_A : $\pi \neq 0.025$

OR

H_0 : $p = 0.025$ (Claim)
 H_A : $p \neq 0.025$

Two-Tailed Test

Example 3: Comparing female and male SAT scores

The school board claims that the average SAT score for female high school students is greater than the average SAT score for male high school students. If gender is not related to SAT scores, then the SAT scores should be the same.

Step 1: Since this is a two-population mean average problem we will need to decide what is population 1 and what is population 2 and the correct letter to use. I tend to write statements as they are claimed. In this problem, they said that the population mean average SAT for females is higher than males. The most straight forward way is to make females population 1 and males population 2 and then say females is higher than males. We could reverse it and put males as population 1, but then we would have to say that population 1 is lower than population 2. Remember that the population 1 parameter should always go on the left.

μ_1 : Female
 μ_2 : Male

Write down the two opposing views in symbolic language. Notice that the two opposing views indicate a gender / SAT relationship (the mean average for females is higher than males) verses no relationship (mean averages are the same.) Notice “equal to” goes with not related. This is why not related is always the null hypothesis. Even though the opposite of $>$ is “ \leq ”, this symbol does not seem appropriate for this test. If the female SAT’s were lower, that also would indicate a gender/SAT relationship.



$$\mu_1 > \mu_2$$

$$\mu_1 = \mu_2$$

Step 2: Decide the claim. It seems that while the school board longs for the day when the genders are the same, they do not believe that is true right now. They believe that gender does matter. They believe that the population mean average for females is higher than for males.

$$\mu_1 > \mu_2 \text{ (Claim)}$$

$$\mu_1 = \mu_2$$

Step 3: Decide which statement is the null hypothesis. Remember, the statement that has “=” or “≥” or “≤” is the null hypothesis. Put an “ H_0 ” next to it. Remember, claim has nothing to do with it.

$$\mu_1 > \mu_2 \text{ (Claim)}$$

$$H_0: \mu_1 = \mu_2$$

Step 4: Decide which statement is the alternative hypothesis. Remember, the statement that has “≠” or “<” or “>” is always the alternative hypothesis. Put an “ H_A ” next to it. Claim has nothing to do with a statement being the alternative hypothesis. It is the symbol.

$$H_A: \mu_1 > \mu_2 \text{ (Claim)}$$

$$H_0: \mu_1 = \mu_2$$

Step 5: Determine the type of test you are doing. Remember, the alternative hypothesis determines the type of test. If H_A is less than (<), it is a left tailed test. (Notice “<” points to the left). If H_A is greater than (>), it is a right tailed test. (Notice “>” points to the right). If H_A is not equal (≠), it is a two-tailed test. (not equal means less than or greater than.)

In this problem, the alternative hypothesis is greater than (>), which points to the right. So this is a right-tailed test.

Null Hypothesis Confusion

Some students get confused over the sign in the null hypothesis. It is important to pay attention to the language. Look at the following two examples.

Claim: “The population standard deviation (σ) used to be 2 inches but now we think it has increased.”

Since we have two opposing views, we can write them both and then chose the null and alternative hypothesis by the sign.

$$H_0: \sigma = 2$$

$$H_A: \sigma > 2 \text{ (Claim)}$$

Claim: “We think that the population standard deviation is more than 2 inches.”

This time, we do not have the opposing view, so we use opposites. The opposite of “>” is “≤”.

$$H_0: \sigma \leq 2$$

$$H_A: \sigma > 2 \text{ (Claim)}$$



These two examples illustrate a point of confusion. Sometimes you may see the null hypothesis of the same hypothesis test written as “=” and sometimes it may be written with “≤” or “≥”.

$$H_0: \sigma = 2$$
$$H_A: \sigma > 2 \text{ (Claim)}$$

OR

$$H_0: \sigma \leq 2$$
$$H_A: \sigma > 2 \text{ (Claim)}$$

Either answer is ok. Notice the parameter is still 2 inches and they are both right-tailed tests. In all practicality, they are the same test.

Many scientists prefer to write the null almost always as “=”. Remember the null in an experiment is usually “no change” or “no effect”. Change of any kind is usually denoted by the alternative hypothesis.

Confusion about “At Least” or “At Most”.

When we say we have at least \$20, we mean the amount of money is greater than or equal to \$20 (\geq).
When we say we have at most \$20, we mean the amount of money is less than or equal to \$20 (\leq).

If we stick to the language of “at least” or “at most” we would have to make them the null hypothesis.

If the claim was that the population mean amount of money is at least \$20, we would write the following null and alternative hypothesis. Remember at least means greater than or equal to “ \geq ”. The opposing view would be less than “ $<$ ”. Since the alternative is less than, this would be a left tailed test.

$$H_0: \mu \geq 20 \text{ (Claim)}$$
$$H_A: \mu < 20$$

This can create confusion. Does the person really want to check “at least” or do they really mean “more than”? It has been my experience that when people want to check an “at least” claim, they are better off changing the claim to “more than” or “increased” and doing a traditional right tailed test.

Similarly if someone wants to test an “at most” claim, they are better off changing the claim to “less than” or “decreased” and doing a traditional left tailed test.



Practice Problems Section 3A

For each of the following problems:

- a) Write the null and alternative hypothesis.
- b) Label whether the null or the alternative is the claim.
- c) Tell whether this is a left tail test, a right tail test, or a two tail test.

1. According to a CNN report, 93% of all Americans also own a traditional phone. We disagree with this report. We claim that the percentage has decreased as more and more Americans opt to only use a cell phone and throw away their traditional phones.
2. According to a recent Newspaper article, the population mean average amount of time people in California spend eating and drinking per day is 1.25 hours. Test the claim that the population mean average number of hours spent eating and drinking really is 1.25 hours.
3. According to an article in *USA Today*, 74% of Americans own a credit card. We disagree with the *USA Today* article and claim that more than 74% of Americans own a credit card.
4. It has long been thought that the population mean average body temperature for all humans is 98.6 degrees Fahrenheit. A recent scientific study is now claiming that the population mean average body temperature is actually lower than 98.6 degrees Fahrenheit.
5. The population standard deviation for the heights of adult men was thought to be 2.9 inches. We disagree with this. We claim that the population standard deviation for heights of men is not 2.9 inches.
6. Test the claim that more than 10% of the world population is left-handed.
7. The population percent of all women worldwide that hold CEO level jobs is lower than the population percent of all men worldwide that hold CEO level jobs. We long to eliminate this gender bias, so that the population percentages are the same.
8. The population variance for cholesterol levels is about the same for adult men and adult women in the U.S.
9. The majority of all Republicans support decreasing taxes.
10. We claim that the population correlation coefficient is zero.
11. While some think that the population slope is zero, we claim that the population slope is significantly positive.
12. According to the center for disease control (CDC), 9.4% of people in the U.S. have some form of diabetes.
13. According to an article in an automobile magazine, the population mean average price of a used mustang is 18 thousand dollars. We disagree with this article and claim that data indicates that the population mean is lower than 18 thousand dollars.
14. According to the center for disease control (CDC), the population percentage of rabies cases from wild animals in 2015 is different from the percentage of rabies cases from wild animals in 2017.
15. Some people believe that the population variance for human body temperature is about 0.5 degrees Fahrenheit squared. A recent scientific study is now claiming that the population variance for body temperature may actually be higher than 0.5.
16. Test the claim that less than 90% of the world population is right-handed.
17. We claim that the population mean average salary of female lawyers in New York is significantly lower than the population mean average salary of male lawyers in New York. We long to eliminate this gender bias, so that the population mean averages are the same.
18. The population variance for men's heights is not the same as for women's heights.
19. We claim that the population correlation coefficient is greater than zero.



20. While some think that the population slope is zero, we claim that the population slope is significantly negative.

Section 3B – Test Statistics and Critical Values

Vocabulary

Population: The collection of all people or objects to be studied.

Census: Collecting data from everyone in a population.

Sample: Collecting data from a small subgroup of the population.

Statistic: A number calculated from sample data in order to understand the characteristics of the data.
For example, a sample mean average, a sample standard deviation, or a sample percentage.

Parameter: A number that describes the characteristics of a population like a population mean or a population percentage. Can be calculated from an unbiased census, but is often just a guess about the population.

Hypothesis Test: A procedure for testing a claim about a population.

Null Hypothesis (H_0): A statement about the population that involves equality. It is often a statement about “no change”, “no relationship” or “no effect”.

Test Statistic: A number calculated in order to determine if the sample data significantly disagrees with the null hypothesis. There are a variety of different test statistics depending on the type of data.

One-Population Proportion Test Statistic (z): The sample proportion is this many standard errors above or below the population proportion in the null hypothesis.

One-Population Mean Test Statistic (t): The sample mean is this many standard errors above or below the population mean in the null hypothesis.

Two-Population Proportion Test Statistic (z): The sample proportion for group one is this many standard errors above or below the sample proportion for group two. Can also be thought of as how many standard errors the difference between the sample proportions is from zero or from some other specific difference.

Two-Population Mean Average Test Statistic (t): The sample mean for group one is this many standard errors above or below the sample mean for group two. Can also be thought of as how many standard errors the difference between the sample means is from zero or from some other specific difference.

Critical Value: A number we compare our test statistic to in order to determine significance. In a sampling distribution or a theoretical distribution approximating the sampling distribution, the critical value shows us where the tail or tails are. The test statistic must fall in the tail to be significant.



Introduction

A hypothesis test compares random sample data to the null and alternative hypothesis in order to determine the validity of a population claim. Remember, the null hypothesis is the population statement that involves equality, usually to some population parameter. The principle of sampling variability (random chance) tells us that random sample statistics will usually be very different than population parameters in the null hypothesis. We know the sample data will almost always disagree with the null hypothesis. The question is does it significantly disagree?

Significant differences are very difficult to determine. For example, a sample mean is 4.3 kg higher than the population mean in the null hypothesis, but is that significant? A sample percentage may be 3.5% below the population percentage in the null hypothesis, but is that significant? Statisticians and mathematicians put a lot of thought into figuring out ways to determine if a sample statistic significantly disagrees with a population parameter. The result was the invention of the “test statistic”.

Test Statistic: A number calculated in order to determine if the sample data significantly disagrees with the null hypothesis.

Data can take on many forms, categorical or quantitative, one-population or two-population or ten-populations. So there are many types of test statistics to deal with the various types of data and the number of populations. Test statistics have one thing in common. They are trying to see if the random sample data significantly disagrees with the null hypothesis or if it is not a significant disagreement.

Using Test Statistics and Critical Values to Judge Significance

Key question: How can we tell if the test statistic indicates a significant disagreement between the sample data and the null hypothesis?

The answer to this is to compare the test statistic to a critical value.

Critical Value: We compare a test statistic to this number to determine if the sample data significantly disagrees with the null hypothesis.

Critical values and test statistics are always different depending on the situation. Am I dealing with a left-tailed, right-tailed or two-tailed test? What confidence level (or significance level) should I use? Most computer programs calculate the test statistics and critical values for a situation. Your job as a data scientist is to be able to explain these numbers and judge significance.

The most important rule to remember about test statistics is that the following. You may not understand the test statistic calculation or how the critical value was found, but you can at least judge whether your random sample data significantly disagrees with the null hypothesis or not.

Critical values help us judge significance. To understand a critical value you have seen the sampling distribution or a theoretical distribution curve that estimates a sampling distribution. If the test statistic falls in one of the tails of the sampling distribution or in the tail of the theoretical curve, this indicates significance. If the test statistic does not fall in one of the tails of the theoretical distribution curve, the sample data does not significantly disagree with the null hypothesis. A key question is where do the tails start? That is what the critical value is calculating. The critical value is the cutoff for where the tails of the distribution begin. We just need to compare the test statistic to the critical value and see if the test statistic falls in the tail. When a computer program calculates the test statistic and critical value for you, it is always good to draw the theoretical curve and use the critical values to visualize the tails. Now ask yourself if the test statistic is in the tail or not.

Significance Rule for Test Statistics and Critical Values

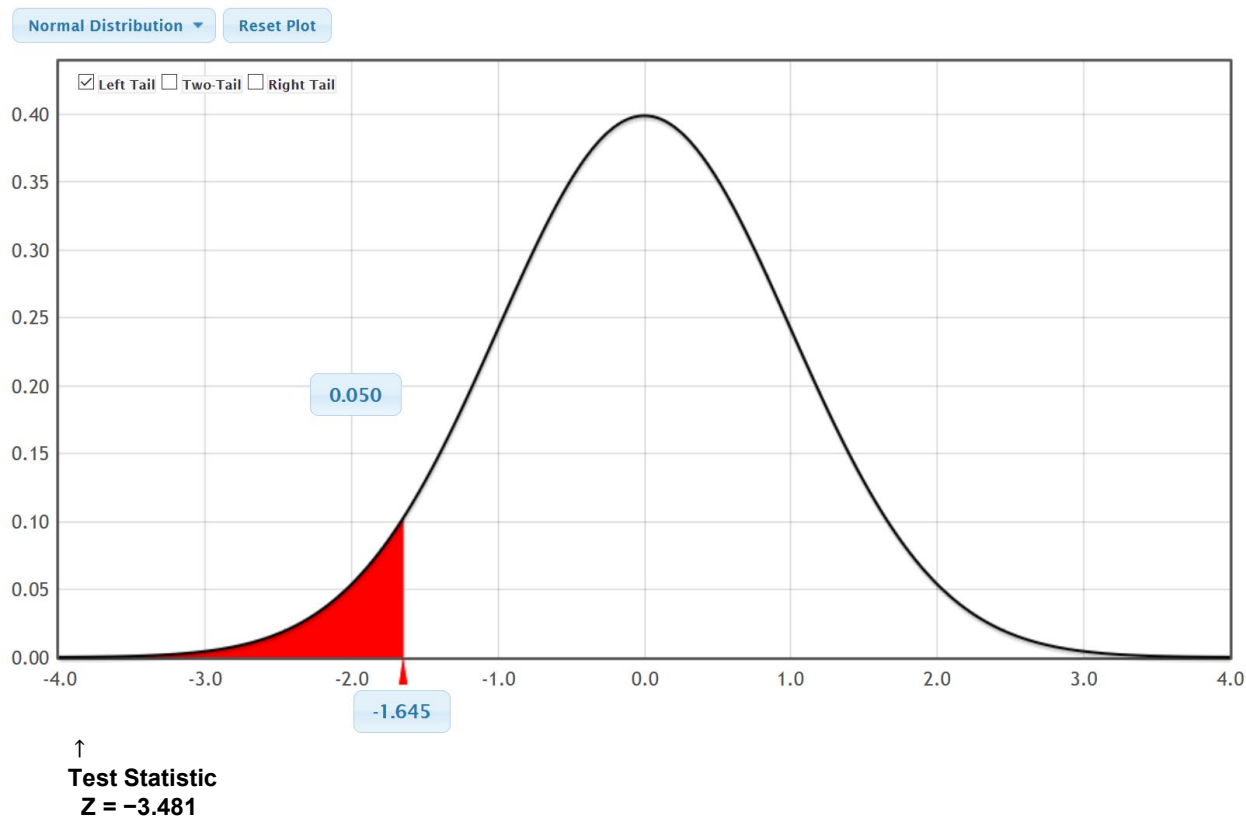
- **If the test statistic falls in the tail determined by the critical value (or values), then the sample data significantly disagrees with the null hypothesis.**
- **If the test statistic does NOT fall in the tail determined by the critical value (or values), then the sample data does NOT significantly disagree with the null hypothesis.**



Left-tailed test statistic example

Suppose we are doing a left tailed hypothesis test for proportions and the computer calculates a test statistic of $Z = -3.481$ and a critical value of -1.645 . The Z-distribution is a normal curve. So we need to see if the test statistic falls in the left tail. Negative numbers can be particularly challenging. Remember on the number line, -3.481 is less than -1.645 , so the test statistic does fall in the left tail. This means that the sample data significantly disagrees with the null hypothesis.

A nice way to visualize this principle is to go to the “Theoretical Distributions” menu in StatKey at www.lock5stat.com. Click on “normal”, then “Left Tail”. Put the critical value of -1.645 in the bottom box.

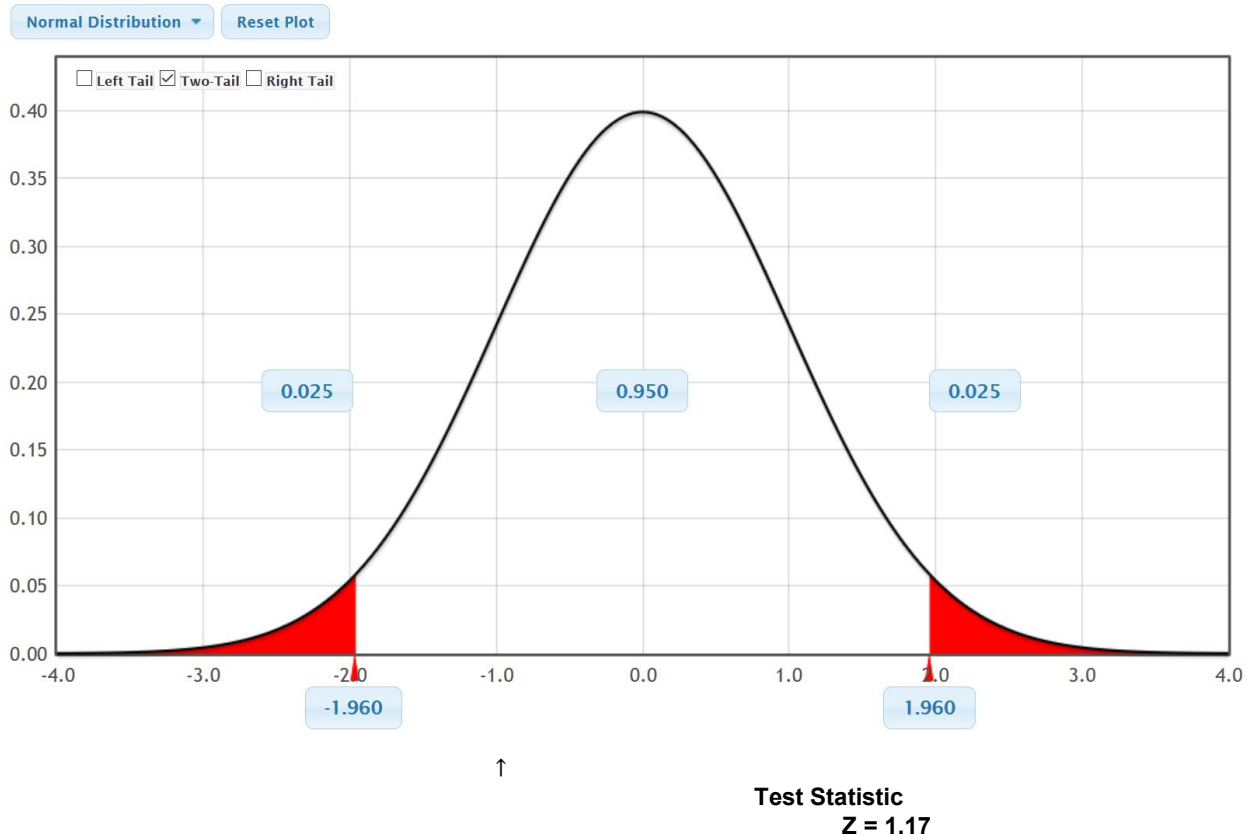


Two-tailed Test Statistic Example

Suppose we are doing a two-tailed hypothesis test for proportions. The computer calculated a Z-test statistic of $Z = +1.17$ and Critical Values of ± 1.96 .

Notice that in a two-tailed test, there are two critical values. In normal distributions like Z or T, two-tailed critical values are a certain number of standard errors above or below and so the positive and negative number of standard errors. In this example we see the “plus or minus” notation for the critical value ± 1.96 . This really means that the lower critical value for the left tail is -1.96 and the upper critical value for the right tail is $+1.96$. In non-normal distributions like F or χ^2 , the upper and lower critical values are not plus or minus. They all work the same way though. Is my test statistic in the tail. If we draw the normal Z-curve with the right tail starting at the upper critical value of $+1.96$ and left tail starting at -1.96 , we can see if the test statistic of $+1.17$ falls in the tail. Again, we can use StatKey to visualize the curve and the critical values. Go to the “Theoretical Distributions” menu in StatKey at www.lock5stat.com. Click on “normal”, then “Two Tail”. Put the critical value of -1.96 in the bottom box in the left tail and $+1.96$ in the bottom box of the right tail. Where does our test statistic fall?





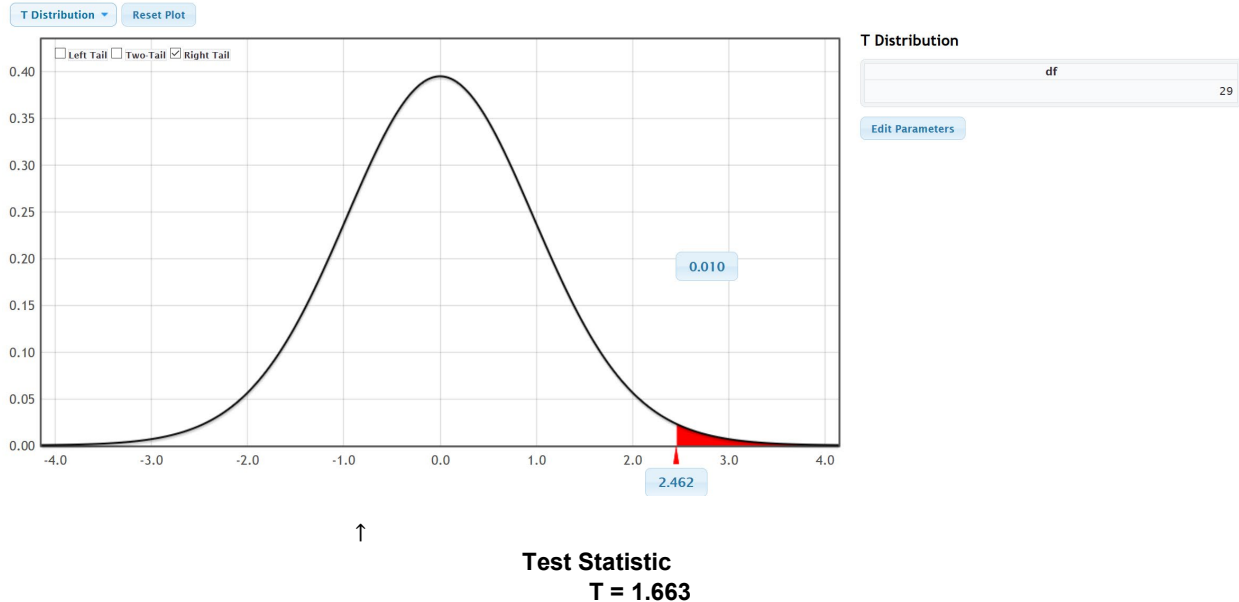
Notice our test statistic did not fall in one of the tails. So our test statistic and critical values are indicating that the sample data does not significantly disagree with the null hypothesis.

Right-tailed Test Statistic Example

Suppose we are doing a right tailed hypothesis test for means and a degrees of freedom of 29. The computer calculated a T-test statistic of $T = +1.663$ and a critical value of $+2.462$.

Like the Z-distribution, the T-distribution is also normal. So if this was a two-tailed test, we would see the “±” notation. This is a right tailed test though, so we see there is one upper critical value for the right tail. Draw the normal T-curve with the right tail starting at the upper critical value of $+2.462$. You can have StatKey draw it for you if you wish. Go to the “Theoretical Distributions” menu in StatKey at www.lock5stat.com. Click on “t”, then put in the degrees of freedom of 29. Now click “Right Tail” and put the critical value of $+2.462$ in the bottom box below the right tail. Does our test statistic fall in the tail?





Notice that our T-test statistic of 1.663 did not fall in the tail determined by the critical value. This again tells us that our sample data does not significantly disagree with the null hypothesis.

Significance Levels (Alpha Levels) (α)

The significance level (or alpha level) is an important number in a hypothesis test. It is often denoted by the Greek letter “alpha” or “ α ”. When scientists perform a hypothesis test, they choose a significance level for the test. This number is very important. It is tied to the critical value and is also very important in understanding the amount of sampling variability in a hypothesis test. Do you recall the confidence levels from the last chapter? Think of the significance level (α) as the opposite of the confidence level ($1 - \alpha$).

<u>Confidence Levels ($1 - \alpha$)</u>	<u>Significance Levels (α)</u>
90% (0.90)	10% (0.10)
95% (0.95)	5% (0.05)
99% (0.99)	1% (0.01)

If you want to be 95% confident, you will use a 5% significance level in your hypothesis test. Similarly, if you want to be 90% confident, you will use a 10% significance level. 99% confident corresponds to a 1% significance level.

Note: Remember that a 95% confidence level is the most common. Not surprisingly, a 5% significance level is also the most common. A good rule of thumb, is that if you do not know what significance level to use, use 5% ($\alpha = 0.05$).

One-population proportion Z-test statistic

The test statistic used for a one-population proportion (%) hypothesis test is the Z-test statistic. The Z-test statistic counts the number of standard errors that the sample proportion (\hat{p}) is above or below the population proportion (π) in the null hypothesis. If you recall from previous chapters, the number of standard errors or standard deviations is often called a “Z-score”. Not surprising, the one-population proportion test statistic is a Z-score.

Key question: How can we tell if the Z-test statistic indicates a significant disagreement between the sample proportion and the population proportion in the null hypothesis?



Remember, to know if a test statistic indicates a significant difference, compare the test statistic to a critical value. If the test statistic fall in the tail of the distribution determined by the critical value, then it is a significant difference.

If you remember, there are three famous critical value Z-scores for 90%, 95% and 99% confidence (± 1.645 , ± 1.96 or ± 2.576). These numbers are often used to see if our Z-test statistic is significant.

One-population proportion Z-test statistic example

Suppose we want to test the claim that the population percentage is 25%. Let's look at the percentage problem where the population proportion (π) is 0.25 (25%) and the sample proportion (\hat{p}) is 0.22 (22%). We know the sample value is % lower, but we do not know if that is significant. Another important bit of information is the sample size. In this case, it was 100.

$$H_0: \pi = 0.25 \text{ (Claim)}$$

$$H_A: \pi \neq 0.25$$

To find out if our sample data disagrees with the null hypothesis, let's calculate the test statistic.

Formulas for Test Statistics often follow general patterns. A one-population proportion or mean test statistic counts how many standard errors that the sample statistic is above or below the population parameter in the null hypothesis. Here is the general formula.

$$\frac{(\text{Sample Statistic} - \text{Population Parameter})}{\text{Standard Error}}$$

Remember in the last unit we saw that statisticians often used formulas to approximate the standard error. For one-proportion confidence intervals the standard error was $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. In a hypothesis test we have an idea of what the population proportion (π) is, so we prefer to use π instead of \hat{p} in the standard error estimation formula.

Here is the formula for a one-population proportion Z-test statistic.

$$Z = \frac{(\hat{p} - \pi)}{\sqrt{\frac{\pi(1-\pi)}{n}}}$$

Let's plug in our numbers. Remember that the sample proportion $\hat{p} = 0.22$, the population proportion $\pi = 0.25$ and the sample size $n = 100$.

$$Z = \frac{(\hat{p} - \pi)}{\sqrt{\frac{\pi(1-\pi)}{n}}} = \frac{(0.22 - 0.25)}{\sqrt{\frac{0.25(1-0.25)}{100}}} = \frac{(0.22 - 0.25)}{\sqrt{\frac{0.25(0.75)}{100}}} \approx \frac{(-0.03)}{0.0433} \approx -0.69$$

It is always very important to explain your test statistic. In this one population proportion hypothesis test, the Z-test statistic is counting how many standard errors the sample proportion (\hat{p}) is above or below the population proportion (π). If the test statistic is negative, it is "below". If the test statistic is positive, it is "above".

Z-test statistic sentence:

The sample proportion (0.22) was 0.69 standard errors below the population proportion (0.25).

OR

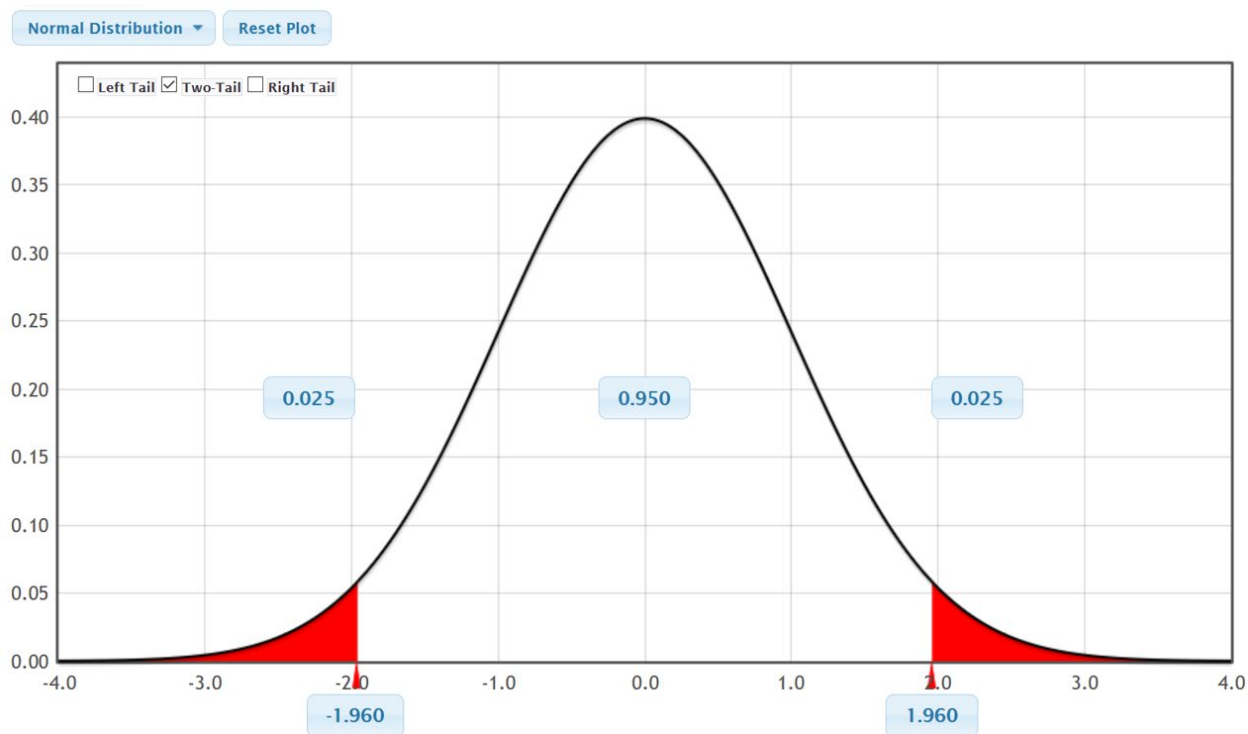
The sample percentage (22%) was 0.69 standard errors below the population percentage 25%.

Is it significant? To determine if this is significant, we should calculate the critical value. Computer programs usually calculate the test statistic and critical value for us. We also learned in the last chapter that we can use StatKey to calculate a critical value.



To calculate a critical value, you must first determine what significance level you plan to use and if your hypothesis test is left-tailed, right-tailed or two-tailed. In this example, the data scientist used a 5% significance level. Notice the alternative hypothesis was “ \neq ”, so this is a two-tailed test.

Go to www.lock5stat.com and click on “StatKey”. Under the theoretical distributions menu click on “normal”. This is the menu for calculating critical value Z-scores. Leave the mean at zero and the standard deviation as one. Click two tail. The significance level is the probability in the tail. Since there are two tails, the significance level will be broken in half with 2.5% in each tail. The numbers on the bottom are the critical values.



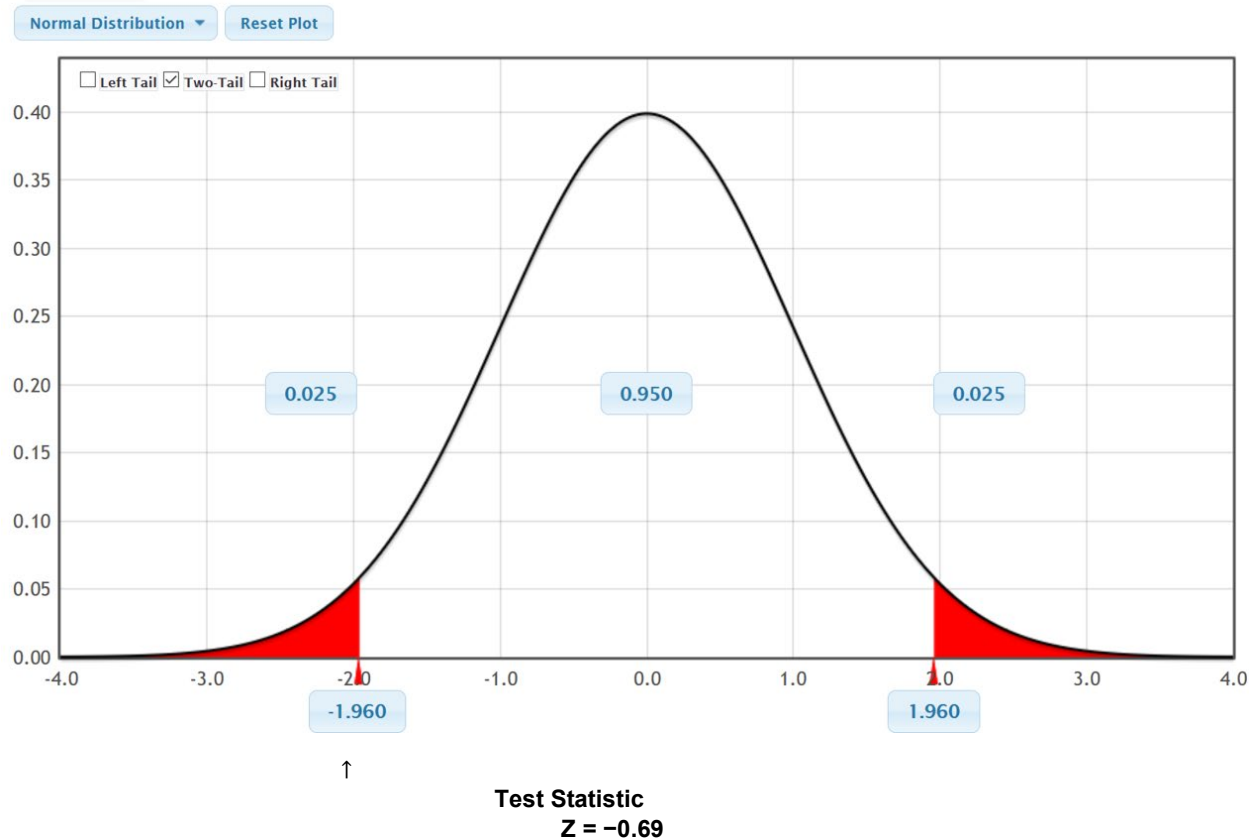
So we see that our critical values for this hypothesis test are ± 1.96 . Remember, our test statistic is $Z = -0.69$. Let's use our significance rule to determine if the sample data significantly disagreed with the null hypothesis.

Significance Rule for Test Statistics and Critical Values

- If the test statistic falls in the tail determined by the critical value (or values), then the sample data significantly disagrees with the null hypothesis.
- If the test statistic does NOT fall in the tail determined by the critical value (or values), then the sample data does NOT significantly disagree with the null hypothesis.

Did our Z-test statistic of $Z = -0.69$ fall in one of the tails?





Our test statistic did not fall in the tail, so is not significant. The sample proportion and population proportion in the null hypothesis are not significantly different. They are actually pretty close. This also tells us that the sample data does not significantly disagree with the null hypothesis.

Statcato: Most people working in statistics today do not calculate test statistics and critical values. Statistics computer programs can calculate them for us. In Statcato, go to the statistics menu, then click on hypothesis test. Click on "one-population proportion". Most computer programs require the number of success (events) to calculate proportions. In the last example, they did not tell us the number of successes. They just told us the total sample size (100) and the sample proportion (0.22). If you multiply them, you get the number of successes or events.

$$\text{Number of Successes (events)} = 100 \times 0.22 = 22$$

So in the one-population proportion hypothesis test menu in Statcato, go to the summarized data and enter 22 for the number of events and 100 for the total number of trials. Put in not equal for the alternative hypothesis, 0.25 for the hypothesized proportion, and a significance level of 0.05. Here is the Statcato printout. Notice the test statistic is close to our calculation and the critical values are the same as what we calculated with StatKey.



Inputs

Samples in column:

Summarized sample data:
 Number of events:
 Number of trials:

Significance

Significance Level: 0 - 1.00 (e.g. 0.05)
 Confidence Level: 0 - 1.00 (e.g. 0.95)

Alternative Hypothesis

Alternative Hypothesis:
 Hypothesized Proportion:

Hypothesis Test - One population proportion: confidence level = 0.95

Input: Summary data

Null hypothesis: $p = 0.25$

Alternative hypothesis: $p \neq 0.25$

N	Sample Proportion	Significance Level	Critical Value	Test Statistic Z	p-Value
100	0.22	0.05	-1.96, 1.96	-0.693	0.4884

Note about Sample Size

Sample size plays a key role in significance. Let's look at the previous example but with a sample size of 1000.

$$Z = \frac{(\hat{p} - \pi)}{\sqrt{\frac{\pi(1-\pi)}{n}}} = \frac{(0.22 - 0.25)}{\sqrt{\frac{0.25(1-0.25)}{1000}}} = \frac{(0.22 - 0.25)}{\sqrt{\frac{0.25(0.75)}{1000}}} = \frac{-0.03}{0.0137} \approx -2.19$$

If our sample size had been 1000, then the sample percentage of 22% is significantly lower than our population percentage of 25%. In fact, our sample percentage of 22% is 2.19 standard errors below the population percentage of 25%. A test statistic of $Z = -2.19$ would have fallen in the left tail since it is less than the lower critical value of -1.96 . At a sample size of 1000, 22% is significantly less than 25%.

One-population mean T-test statistic

The test statistic used for a one-population mean hypothesis test is the T-test statistic. Like the Z-test statistic, the one-population T-test statistic also counts the number of standard errors that the sample statistic is from the population parameter in the null hypothesis. In this case, it will count the number of standard errors that the sample mean (\bar{x}) is above or below the population mean (μ) in the null hypothesis.

Key question: How can we tell if the T-test statistic indicates a significant disagreement between the sample mean (\bar{x}) and the population mean (μ) in the null hypothesis?

Remember, to know if a test statistic indicates a significant difference, compare the test statistic to a critical values. For a significant difference, the test statistic should fall in the tail of the distribution determined by the critical values.

In our last chapter, we saw that T-score critical values are organized by degrees of freedom. For a one-population hypothesis test, the degrees of freedom is the sample size $- 1$ or "n - 1".



One-population mean T-test statistic example

An article in a health magazine claims that the mean average weight of all men is less than 180 pounds. A random sample of 40 men found that the sample mean was 172.55 pounds with a sample standard deviation of 26.327 pounds. Use a 10% significance level and the random sample data to test this claim.

$$H_0 : \mu = 180$$

$$H_A : \mu < 180 \text{ (Claim)}$$

Notice that the sample mean of 172.55 pounds is 7.45 pounds less than the population mean of 180 pounds. Is that significant?

The answer again is we don't know yet. We would need calculate a test statistic and a critical value to see if 7.45 pounds is a lot in this situation. Notice that since the alternative hypothesis is less than, this is a left-tailed test.

Here is the formula for calculating test statistics to compare sample and population means. Notice it follows the same general pattern and seeks to count how many standard errors the sample statistic is above or below the population parameter in the null hypothesis. In this case, the one-population T-test statistic calculates the number of standard errors that the sample mean (\bar{x}) is above or below the population mean (μ) in the null hypothesis. If the test statistic comes out negative it will be "below" and if the test statistic is positive, it will be "above". Notice also that we are using the same standard error estimation formula that we used for confidence intervals.

$$T = \frac{\text{(Sample Mean - Population Mean)}}{\text{Standard Error}} = \frac{(\bar{x} - \mu)}{\left(\frac{s}{\sqrt{n}}\right)}$$

To calculate the test statistic, plug in the sample mean $\bar{x} = 172.55$, the population mean $\mu = 180$, the sample standard deviation $s = 26.327$, and sample size $n = 40$.

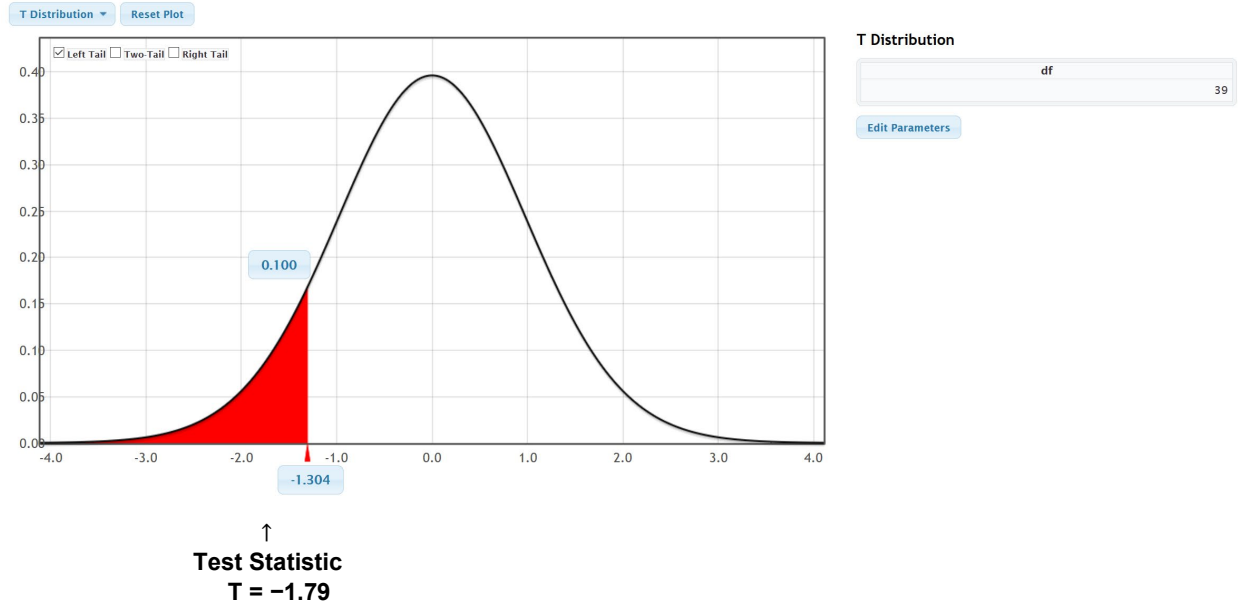
$$T = \frac{(\bar{x} - \mu)}{\left(\frac{s}{\sqrt{n}}\right)} = \frac{(172.55 - 180)}{\left(\frac{26.327}{\sqrt{40}}\right)} \approx \frac{(-7.45)}{4.1627} \approx -1.79$$

Test-statistic sentence: The sample mean of 172.55 pounds is 1.79 standard error below the population mean of 180 pounds.

Is it significant?

Again, we will use StatKey to calculate the critical value. Go to the "Theoretical Distributions" menu and click on "T". In this problem the degrees of freedom is $40 - 1 = 39$. We also used a 10% significance level for this left-tailed test. After entering the degrees of freedom into StatKey, click on left tail, and put in 10% (0.10) for the tail proportion. Does the test statistic of $T = -1.79$ fall in the tail?





Notice that since the test statistic $T = -1.79$ is less than the critical value of -1.304 , the T-test statistic does fall in the left tail. So our sample mean of 172.55 pounds is significantly lower than the population mean of 180 pounds and the sample data significantly disagrees with the null hypothesis.

Statcato: Again, we can calculate the test statistic and critical value with a program like Statcato. Go to the “statistics” menu again and click on “hypothesis tests”. Now click on “one-population mean”. Under summarized data, put in the sample size of 40, sample mean of 172.55, and the sample standard deviation of 26.327. Change the significance level to 0.10, the alternative hypothesis to less than, and the hypothesized mean as 180. Then push ok. The printout is given below. Notice the test statistic is the same as our calculation and the critical value is the same as what we got with StatKey.

Hypothesis Test: 1-Population Mean

Help

Inputs

Samples in column: Summarized sample data:

Size: 40

Mean: 172.55

Standard deviation: 26.327

Population Standard Deviation

Population standard deviation:

Known: - use z distribution

Unknown - use t distribution

Significance

Significance Level: 0.10 0 - 1.00 (e.g. 0.05)

Confidence Level: 0.90 0 - 1.00 (e.g. 0.95)

Alternative Hypothesis

Alternative Hypothesis: Less than

Hypothesized Mean: 180

OK Cancel



Hypothesis Test - One Population Mean: confidence level = 0.90

Input: Summary data

σ unknown (using t distribution)

Null hypothesis: $\mu = 180.0$

Alternative hypothesis: $\mu < 180.0$

N	Sample Mean	Stdev s	Significance Level	Critical Value	Test Statistic	p-Value
40	172.55	26.327	0.10	-1.304	-1.790	0.0406

One-population variance χ^2 -test statistic

When doing a hypothesis test about a population standard deviation (σ), we often prefer to use the population variance (σ^2). Remember, the variance is the square of the standard deviation. You can also think of the standard deviation as the square root of the variance. In a one population variance or standard deviation hypothesis test, we will compare the sample variance (s^2) to the population variance (σ^2).

One-population variance χ^2 -test statistic Example

While many people believe that the population standard deviation for men's weight is 20 pounds. This also implies that the population variance for men's weight is 400 square pounds. A random sample of 40 men has a sample standard deviation of 26.327 pounds and a sample variance is 693.111 square pounds. Let's use a 1% significance level for this problem. Notice that the test can be done using the standard deviation or the variance.

$$H_0 : \sigma = 20 \ (\sigma^2 = 400) \text{ (Claim)}$$

$$H_A : \sigma \neq 20 \ (\sigma^2 \neq 400)$$

Notice that the sample standard deviation of 26.327 pounds is 6.327 pounds more than the population standard deviation of 20 pounds.

Is that significant?

To determine this, we will need a test statistic. In our discussions about sampling distributions in the last chapter, we learned that the sampling distribution for variance is rarely normal and usually skewed to the right. For this reason we cannot use a Z or T-test statistic. These are based on normal sampling distributions. We saw in the last chapter that for one-population variance and standard deviation, we use the Chi-Squared distribution (χ^2) with a degrees of freedom $n - 1$.

Chi-Squared One-population Variance Test Statistic: $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$

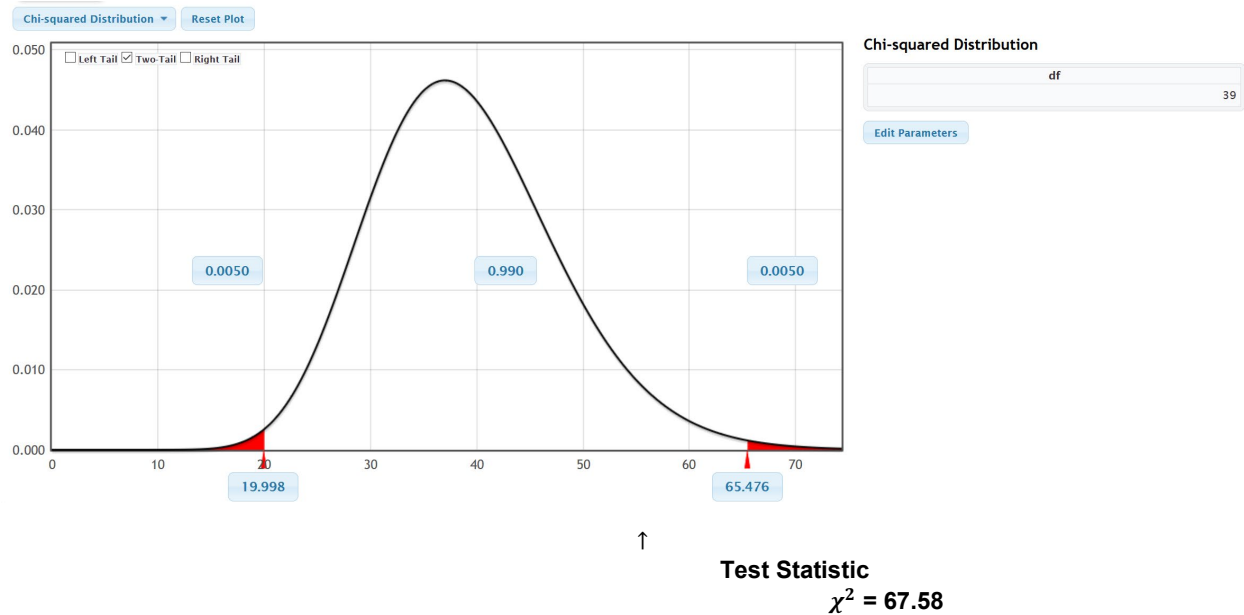
Let's calculate the Chi-Squared test statistic for this problem. The degrees of freedom for this problem is $df = n - 1 = 40 - 1 = 39$. The sample standard deviation is $s = 26.327$, so the sample variance is $s^2 = 693.111$. The population standard deviation in the null hypothesis is 20 pounds, so the population variance is $\sigma^2 = 400$.

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{(40-1)26.327^2}{20^2} = \frac{(39)693.111}{400} \approx 67.58$$

Is this significant? Again, we will need to calculate the critical values. Let's go to StatKey at www.lock5stat.com. Under the theoretical distributions menu click on " χ^2 ". Put in the degrees of freedom 39 and click "Two Tail".

Since this is a two-tailed test we will need to divide the significance level of 1% in half and put 0.5% (0.005) in each tail. What are the critical values? Does the test statistic fall in the tail?





First notice that the Chi-squared distribution is not perfectly normal. The F-distribution is very similar. Also the upper and lower critical values are not positive or negative like Z or T. It still works the same way though. We see that we have divided the significance level 0.01 in half and put 0.005 in each tail. The upper critical value is 65.476 and lower critical value is 19.998.

We also see that the χ^2 test statistic of 67.58 is higher than the upper critical value of 65.476, so falls in the right tail. The sample standard deviation significantly disagrees with the population standard deviation and the sample data significantly disagrees with the null hypothesis.

Statcato: Remember, it is always preferable to use technology to calculate test statistics. To calculate the χ^2 -test statistic and critical value with Statcato, go to the “statistics” menu, click on “hypothesis tests”, then click on “one-population variance”. Put in the sample size of 40 and either the sample standard deviation (26.327 pounds) or the sample variance. Change the significance to 1% (0.01). We also need to put in the population standard deviation (20) or the population variance and change the alternative hypothesis to “Not Equal to”. Now press “OK”.

Hypothesis Test: 1-Population Variance ×

Help F1

Inputs

Samples in column:

Summarized sample data:

Sample Size:

Variance:

Standard deviation:

Significance

Significance Level: 0 - 1.00 (e.g. 0.05)

Confidence Level: 0 - 1.00 (e.g. 0.95)

Alternative Hypothesis

Alternative Hypothesis:

Hypothesized variance:

Hypothesized standard deviation:



Hypothesis Test - One population variance: confidence level = 0.99

Input: Summary data

Null hypothesis: $\sigma^2 = 400.0$

Alternative hypothesis: $\sigma^2 \neq 400.0$

N	Sample Stdev s	Sample Var s^2	Significance Level	Critical Value	Test Statistic	p-Value
40	26.327	693.111	0.01	19.996, 65.476	67.578	0.0061

Notice the test statistic is close to what we calculated above and the critical values are very close to our StatKey calculation.

Important Notes:

- *It is important to understand how test statistics work and what the formulas represent. It is not important to calculate these by hand with a calculator. Statistics programs can calculate the test statistic quickly and with much better accuracy. It is important that you can explain the test statistic and what it tells us about significance.*
- *Whether or not a test statistic is significant can be difficult to interpret. Computer programs often give critical values to compare the test statistic to so that you can know if it was significant. It is important to draw a picture and visualize where the critical values are and where the tails begin. Here is the significance rule for test statistics and critical values.*

Significance Rule for Test Statistics and Critical Values

- **If the test statistic falls in the tail determined by the critical value (or values), then the sample data significantly disagrees with the null hypothesis.**
 - **If the test statistic does NOT fall in the tail determined by the critical value (or values), then the sample data does NOT significantly disagree with the null hypothesis.**
-



Practice Problems Section 3B

(#1-20) For each of the following, use the given test statistic and critical value or values to answer the following questions.

- Draw the indicated distribution and use the critical values to label the tails.
- Does the test statistic fall in one of the tails or not?
- Does the sample data significantly disagree with the null hypothesis? Explain how you know.

	Tail	Test Statistic	Critical Value
1	Two	$Z = 2.47$	± 1.96
2	Left	$T = -3.318$	-1.747
3	Right	$\chi^2 = 6.943$	12.33
4	Right	$F = 1.126$	3.881
5	Left	$Z = -1.33$	-1.645
6	Two	$T = 1.994$	± 2.738
7	Right	$\chi^2 = 18.441$	6.972
8	Right	$F = 7.509$	3.469
9	Two	$Z = -2.72$	± 2.576
10	Left	$T = -3.871$	-2.114
11	Left	$Z = -1.884$	-2.576
12	Right	$T = 0.472$	1.577
13	Two	$\chi^2 = 11.943$	2.346 & 9.841
14	Right	$F = 5.218$	2.791
15	Left	$Z = -2.712$	-1.96
16	Two	$T = 1.138$	± 2.005
17	Right	$\chi^2 = 38.644$	12.359
18	Right	$F = 1.528$	2.467
19	Left	$Z = -0.72$	-2.576
20	Two	$T = -2.871$	± 2.334

(#21-23) Use the “theoretical distributions” menu in StatKey at www.lock5stat.com to look up the following critical values. Click on the button that says “normal”. Then answer the questions.

- Z-test statistic = 2.36
Two-tailed test
Significance Level = 5% (0.025 in each tail)
Critical Values =
Does the sample data significantly disagree with the null hypothesis? Explain why.
- Z-test statistic = -1.48
Left-tailed test
Significance Level = 1% (0.01 in left tail)
Critical Value =
Does the sample data significantly disagree with the null hypothesis? Explain why.
- Z-test statistic = 2.02
Right-tailed test
Significance Level = 10% (0.10 in right tail)
Critical Value =
Does the sample data significantly disagree with the null hypothesis? Explain why.



(#24-26) Use the “theoretical distributions” menu in StatKey at www.lock5stat.com to look up the following critical values. Click on the button that says “t”. Then answer the questions.

24. T-test statistic = -1.773

Two-tailed test

Degrees of Freedom = 29

Significance Level = 1% (0.005 in each tail)

Critical Values =

Does the sample data significantly disagree with the null hypothesis? Explain why.

25. T-test statistic = 2.871

Right-tailed test

Degrees of Freedom = 34

Significance Level = 10% (0.10 in right tail)

Critical Value =

Does the sample data significantly disagree with the null hypothesis? Explain why.

26. T-test statistic = -1.144

Left-tailed test

Degrees of Freedom = 49

Significance Level = 5% (0.05 in left tail)

Critical Value =

Does the sample data significantly disagree with the null hypothesis? Explain why.

(#27-29) Use the “theoretical distributions” menu in StatKey at www.lock5stat.com to look up the following critical values. Click on the button that says “ χ^2 ”. Then answer the questions.

27. χ^2 -test statistic = 38.725

Right-tailed test

Degrees of Freedom = 29

Significance Level = 5% (0.05 in right tail)

Critical Value =

Does the sample data significantly disagree with the null hypothesis? Explain why.

28. χ^2 -test statistic = 15.846

left-tailed test

Degrees of Freedom = 39

Significance Level = 10% (0.10 in left tail)

Critical Value =

Does the sample data significantly disagree with the null hypothesis? Explain why.

29. χ^2 -test statistic = 5.119

two-tailed test

Degrees of Freedom = 19

Significance Level = 1% (0.005 in each tail)

Critical Value =

Does the sample data significantly disagree with the null hypothesis? Explain why.



(#30-32) Use the following one-population test statistic formula to calculate the one-population proportion Z-test statistic. Then write a sentence to explain the test statistic.

$$\text{One-Population Proportion Z-Test Statistic} = \frac{(\text{Sample Proportion} - \text{Population Proportion})}{\text{Standard Error}}$$

30. Sample Proportion (\hat{p}) = 0.317
Population Proportion (π) = 0.25
Standard Error = 0.031

Z-test statistic =

Test Statistic Sentence:

31. Sample Proportion (\hat{p}) = 0.835
Population Proportion (π) = 0.9
Standard Error = 0.053

Z-test statistic =

Test Statistic Sentence:

32. Sample Proportion (\hat{p}) = 0.112
Population Proportion (π) = 0.2
Standard Error = 0.047

Z-test statistic =

Test Statistic Sentence:

(#33-35) Use the following one-population test statistic formula to calculate the one-population mean T-test statistic. Then write a sentence to explain the test statistic.

$$\text{One-Population Mean T-Test Statistic} = \frac{(\text{Sample Mean} - \text{Population Mean})}{\text{Standard Error}}$$

33. Sample Mean (\bar{x}) = 135.7 mg
Population Mean (μ) = 100 mg
Standard Error = 23.9 mg

T-test statistic =

Test Statistic Sentence:

34. Sample Mean (\bar{x}) = 89.26 °F
Population Mean (μ) = 89.6 °F
Standard Error = 0.108 °F

T-test statistic =

Test Statistic Sentence:

35. Sample Mean (\bar{x}) = 52.71 thousand dollars
Population Mean (μ) = 60 thousand dollars
Standard Error = 6.42 thousand dollars

T-test statistic =

Test Statistic Sentence:



Section 3C – P-value and Significance Levels

Vocabulary

Population: The collection of all people or objects to be studied.

Sample: Collecting data from a small subgroup of the population.

Statistic: A number calculated from sample data in order to understand the characteristics of the data.
For example, a sample mean average, a sample standard deviation, or a sample percentage.

Parameter: A number that describes the characteristics of a population like a population mean or a population percentage. Can be calculated from an unbiased census, but is often just a guess about the population.

Hypothesis Test: A procedure for testing a claim about a population.

Null Hypothesis (H_0): A statement about the population that involves equality. It is often a statement about “no change”, “no relationship” or “no effect”.

Sampling Variability: Also called “random chance”. The principle that random samples from the same population will usually be different and give very different statistics. The random samples will usually be different than the population parameter.

P-value: The probability of getting the sample data or more extreme because of sampling variability (by random chance) if the null hypothesis is true. If the P-value is close to zero (lower than the significance level) is unlikely to have happened because of sampling variability. If the P-value is too large (higher than the significance level, then the sample could have occurred because of sampling variability.

Significance Level (α): Also called the Alpha Level. This is the probability of making a type 1 error. The P-value is compared to this number to determine significance and if sampling variability is likely to be involved in the hypothesis test.

Randomized Simulation: A technique for visualizing sampling variability in a hypothesis test. The computer assumes the null hypothesis is true, and then generates random samples. If the sample data or test statistic falls in the tail, then the sample data significantly disagrees with the null hypothesis. This technique can also calculate the P-value and standard error without a formula.

Test Statistic: A number calculated in order to determine if the sample data significantly disagrees with the null hypothesis. There are a variety of different test statistics depending on the type of data.

Critical Value: We compare a test statistic to this number to determine if the sample data significantly disagrees with the null hypothesis. If the absolute value of the test statistic is higher than the absolute value of the critical value, then the sample data significantly disagrees with the null hypothesis.

Introduction: There is a dilemma in hypothesis testing. In a hypothesis test, we want to determine if the sample data disagrees with the null hypothesis, but there is a problem. The principle of sampling variability (random chance) tells us that random samples will almost always be different than population parameters in the null hypothesis. So even if the population parameter in the null hypothesis is correct, my random sample data will still disagree with it. So how can we use random sample data to ever decide about the accuracy of a population parameter? Random samples almost always disagree with the null hypothesis.



The real question is why does the random sample data disagree? There are only two possible answers to that question and these are at the heart of the problem.

1. The random sample disagrees because the null hypothesis is wrong.

OR

2. The null hypothesis is correct and the random sample data disagrees because of sampling variability (random chance).

How do we know which option is correct in a situation? Does my random sample data disagree because all random samples disagree, or does my random sample data disagree because the null hypothesis is wrong?

To answer this question, statisticians invented the P-value.

P-value

P-value: The probability of getting the sample data or more extreme because of sampling variability (by random chance) if the null hypothesis is true.

We see from the definition, we see several important ideas addressed.

- The P-value is a conditional probability based on the null hypothesis being true. The P-value can only be calculated by assuming the null hypothesis was true.
- The P-value is a probability that our random sample data occurs. If the null hypothesis really is correct, then what is the probability of our random sample data occurring by random chance?
- The P-value helps us understand why our random sample data disagrees with the null hypothesis. Does it disagree because of sampling variability (random chance) or not? If it is not sampling variability, then the only other alternative is that the null hypothesis is wrong.
- The P-value not only takes into account the probability of the sample data occurring, but also any other samples that disagrees even more with the null hypothesis than our random sample data. This is what is meant by “or more extreme”.

Reading your P-value

P-value can help us with the dilemma discussed above, but only if you know how to interpret it correctly. Remember, the P-value is the probability of your random sample data occurring because of sampling variability (by random chance). Does my random sample data disagree with the null hypothesis just because of sampling variability? If so, then the population parameter in the null hypothesis might be correct. If my sample data does not disagree because of sampling variability, then the only other alternative is that the null hypothesis must be wrong. In that case we will say that we “reject the null hypothesis”.

Low P-value

Scientists like the P-value to be very close to zero. The lower the P-value, the better. Remember, the P-value is measuring the probability that the random sample data or more extreme occurred because of sampling variability. If the P-value is zero (or really close to zero), then the data probably did not occur because of sampling variability.

Think of sampling variability as a confounding variable that we need to control or at least make sure it is unlikely to be the reason the sample data disagrees. If the P-value is zero, then it is unlikely to be sampling variability (random chance).

Suppose you P-value is 0.013 (1.3%). If your car has only a 1.3% probability of starting, do you think your car will start or is it unlikely to start? If your car only has a 1.3% chance of starting, it is unlikely to start. That is a good way to think about P-value. If there is only a 1.3% probability of our random sample data disagreeing by random chance,



it is probably not random chance! It is unlikely to be sampling variability.

Remember our dilemma about the two options in a hypothesis test.

1. The random sample disagrees because the null hypothesis is wrong.

OR

2. The null hypothesis is correct and the random sample data disagrees because of sampling variability (random chance).

Low P-value Key Idea: If the P-value is really close to zero, it is ruling out option 2. At least we can say that it is very unlikely to be sampling variability (option 2). In that case, the only other alternative is option 1. The random sample data disagrees with the null hypothesis because those population parameters in the null hypothesis are wrong. In that case, we can reject the null hypothesis.

P-value close to zero → Unlikely to be sampling variability → Reject H_0

High P-value

Remember, the goal is to totally rule out sampling variability as the reason our random sample data disagrees. We need the P-value to be zero or at least as close to zero as possible. It doesn't take much for a P-value to be too high. For example, suppose our P-value was 0.15 (15%). Don't events with a 15% probability sometimes happen? While 15% may be a low probability, is it really low enough to totally rule out that the event will not happen? For this reason, we need the P-value to be extremely low and extremely close to zero.

So how can we know if the P-value is too high?

The answer to this is to compare the P-value to the significance level.

In the last section, we saw that scientists pick a significance level at the beginning of the hypothesis test. We also saw that the significance level is connected to the critical value and determining if the test statistic falls in the tail. The proportion in the tail is the significance level. The significance level is also called the alpha level (α) and can be thought of as the complement of the confidence levels ($1 - \alpha$). We saw in the last section, that the most common significance level chosen is 5% ($\alpha = 0.05$).

<u>Confidence Levels</u> ($1 - \alpha$)	<u>Significance Levels</u> (α)
90% (0.90)	10% (0.10)
95% (0.95)	5% (0.05)
99% (0.99)	1% (0.01)

So the P-value must be less than or equal to the significance level, to rule out sampling variability (or at least to ensure it is very unlikely to be sampling variability). If the P-value is higher than the significance level, then the sample data could have occurred because of sampling variability.

P-value less than or equal to the significance level → Unlikely to be sampling variability → Reject H_0

P-value higher than the significance level → Could be sampling variability → Fail to reject H_0

Let's talk about these rules some. Let's look again at the P-value of 15% (0.15). If we are using a 5% significance level then the rule would indicate that the random sample data could have occurred because of sampling variability. This implies that the null hypothesis could be correct, and my sample data might disagree because all samples disagree. Does this guarantee that the null hypothesis is correct? Absolutely not. Let's go back to the car starting analogy. If my car only has a 15% probability of starting, is it guaranteed to start? No. It still has a low probability of starting, but it might start. That is the point. A high P-value does not tell us that the null hypothesis is correct for sure. It tells us that it might be correct.

So what about our dilemma? What does a high P-value tell us about our two options?



1. The random sample disagrees because the null hypothesis is wrong.

OR

2. The null hypothesis is correct and the random sample data disagrees because of sampling variability (random chance).

High P-value Key Idea: If the P-value is too high, then we cannot rule out option 2 (sampling variability). The high P-value does not guarantee it is sampling variability though. It just might be. When the P-value is large, we will not be able to tell which option is correct. The null hypothesis might be wrong. Our sample data disagrees with it after all. On the other hand, the null hypothesis might be correct and our sample data disagrees because of sampling variability. In a sense, we cannot tell which option is correct. That is why we say “Fail to reject the null hypothesis”. This means that we do not have a low enough P-value to rule out sampling variability, so we cannot say for sure that the null hypothesis is wrong. It might be correct.

P-value less than or equal to the significance level → Unlikely to be sampling variability → Reject H_0

P-value higher than the significance level → Could be sampling variability → Fail to reject H_0

For this reason, high P-values are generally not preferred by data scientists. A low P-value rules out sampling variability (rules out random chance) and allows us to reject the null hypothesis and support the alternative hypothesis. A low P-value is also considered evidence. Scientific reports often require a low P-value as evidence to support their findings. When a scientist gets a high P-value, they do not have evidence. Sampling variability is involved and they cannot really say anything definitively. This does not mean that a high P-value has no value. A low P-value gives us evidence that the alternative hypothesis is probably correct. A high P-value indicates that the null hypothesis might be correct, but we do not have evidence.

Low P-value (Less than or equal to the significance level)

- Unlikely to be sampling variability
- Reject H_0
- H_A is probably correct
- We have significant evidence.

High P-value (Higher than the significance level)

- Could be sampling variability
- Fail to reject H_0
- H_0 is probably correct
- We do not have evidence.

Example 1 (Interpreting P-values)

Suppose we have a 5% significance level and a P-value = 0.0278. Convert the P-value into a percentage and write a sentence to explain the P-value. Compare the P-value to the significance level. Is this a low P-value or a high P-value. Could this be sampling variability or is it unlikely to be sampling variability? Explain your answer. Does the sample data significantly disagree with the null hypothesis or not? Explain your answer. Should we reject the null hypothesis or fail to reject the null hypothesis? Explain your answer.

P-value = 0.0278 = 2.78%

P-value Sentence: If the null hypothesis is true, there is a 2.78% probability of getting the sample data or more extreme by random chance (because of sampling variability).

P-value (2.78%) is lower than our significance level (5%), so this is a low P-value close to zero.

Since the P-value is very low (2.78%), it is unlikely to be sampling variability (unlikely to be random chance).



A low P-value means the sample data fell in the tail and has a large test statistic. That means that the sample data does significantly disagree with the null hypothesis.

Reject the null hypothesis, since the P-value is lower than the significance level and is unlikely to be sampling variability.

Example 2 (Interpreting P-values)

Suppose we have a 10% significance level and a P-value = 0.414. Convert the P-value into a percentage and write a sentence to explain the P-value. Compare the P-value to the significance level. Is this a low P-value or a high P-value. Could this be sampling variability or is it unlikely to be sampling variability? Explain your answer. Does the sample data significantly disagree with the null hypothesis or not? Explain your answer. Should we reject the null hypothesis or fail to reject the null hypothesis? Explain your answer.

P-value = 0.414 = 41.4%

P-value Sentence: If the null hypothesis is true, there is a 41.4% probability of getting the sample data or more extreme by random chance (because of sampling variability).

P-value (41.4%) is higher than our significance level (10%), so this is a high P-value.

Since the P-value is very high (41.4%), the sample data could have occurred because of sampling variability (could be random chance).

A high P-value means the sample data did not fall in the tail and has a small test statistic. That means that the sample data does NOT significantly disagree with the null hypothesis.

Fail to reject the null hypothesis, since the P-value is higher than the significance level and could be sampling variability.

Calculating P-values

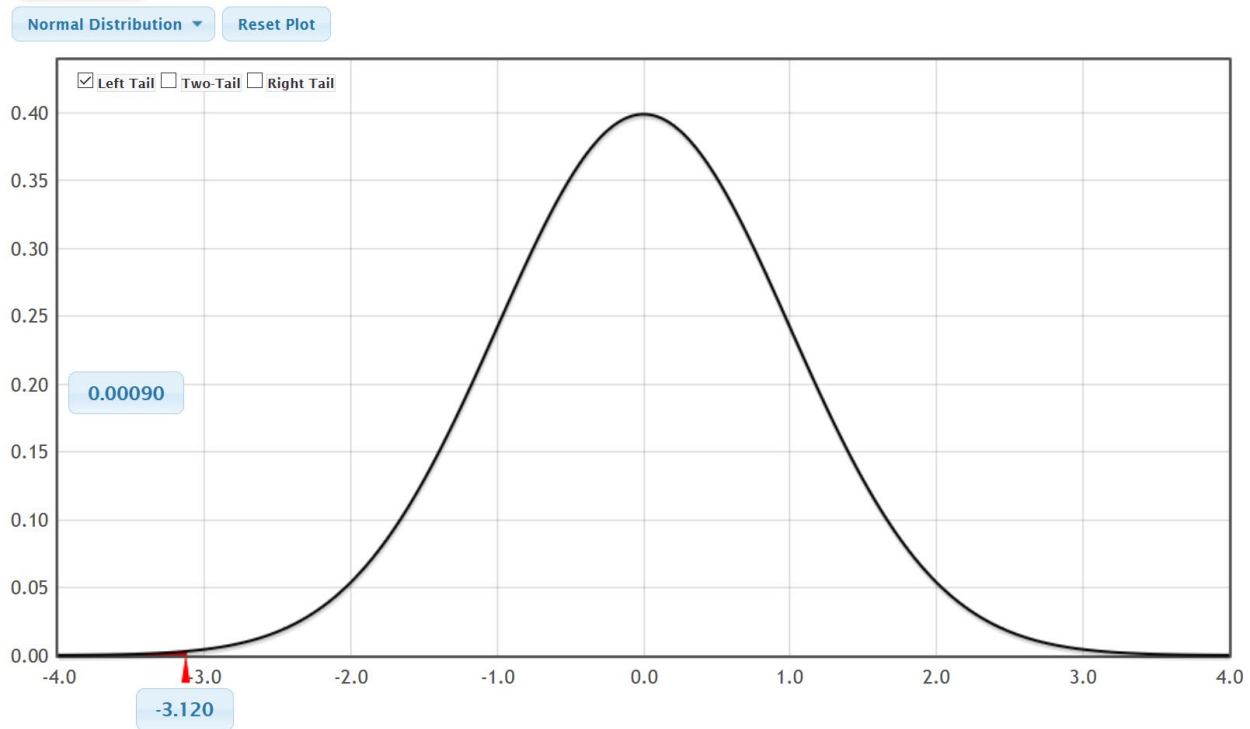
Method 1: Traditional Approach (Using the test statistic and a theoretical curve.)

One way to calculate P-value is with the test statistic and the theoretical curve that represents the sampling distribution. Remember, the P-value is the probability of getting the sample data or more extreme if the null hypothesis is true. Think of the test statistic as representing the sample data if the null hypothesis was true. "The probability of getting the sample data or more extreme" would be the proportion in the tail or tails using the test statistic as your cutoff and taking into account the type of test you are doing.

Example 1: Suppose we are doing a left tailed hypothesis test that uses the Z-test statistic for proportions. Our test statistic compares the sample data to the null hypothesis. In this example, our Z-test statistic was calculated to be $Z = -3.12$. What would the estimated P-value be?

Go to the "theoretical distributions" menu in StatKey at www.lock5stat.com and click on "normal". Click on "Left Tail". In the bottom box in the left tail, put in the test statistic of -3.12 . The proportion calculated above is the estimated P-value.



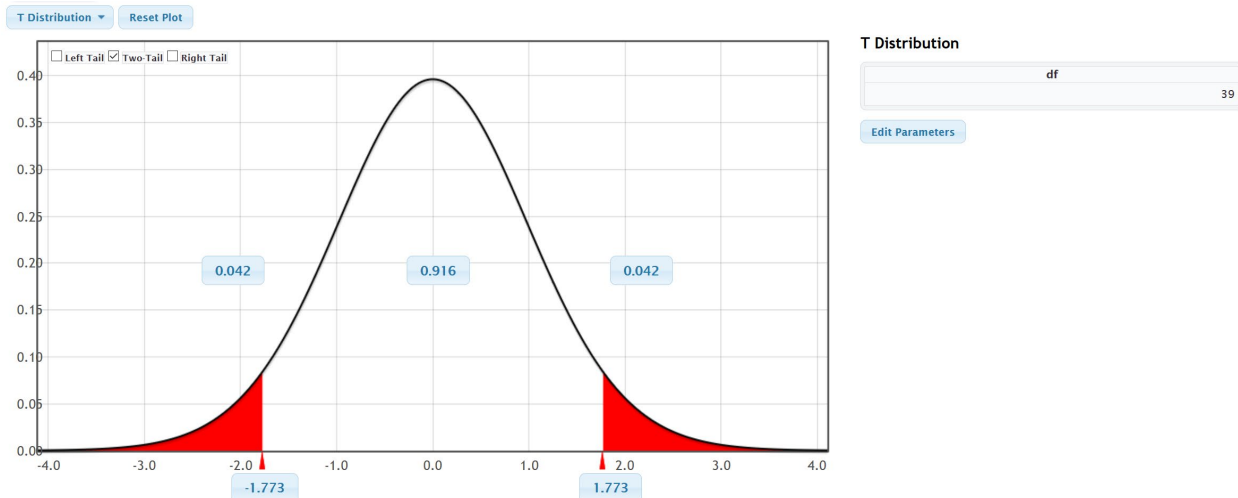


We see that the proportion in the left tail that corresponds to the test statistic cutoff is 0.0009 or 0.09%. This is the estimated P-value, the probability of getting the sample data or more extreme by random chance if the null hypothesis was true.

Example 2: Suppose we are doing a two-tailed hypothesis test that uses the T-test statistic. Remember, our test statistic compares the sample data to the null hypothesis. In this example, our T-test statistic was calculated to be $T=1.773$ and our degrees of freedom was 39. What would the estimated P-value be?

Go to the “theoretical distributions” menu in StatKey at www.lock5stat.com and click on “t”. Under degrees of freedom put in 39 and then click on “Two Tail”. A two-tailed P-value calculation takes a little thought. Notice that there are now two bottom boxes. One in the left tail and one in the right tail. If you T-test statistic is close to the left tail (negative) put in the bottom box in the left tail. If you T-test statistic is close to the right tail (positive) put in the bottom box in the right tail. Since our test statistic is closer to the right tail (positive), we will type in the test statistic of 1.773 into the right bottom box. You do not need to type the test statistic in both boxes. The left tail will automatically adjust to the number you typed in the right tail box. In a two-tailed hypothesis test, “or more extreme” could be any sample data that higher or lower than the parameter in the null hypothesis. So we need to include both of the proportions in the left and right tail. Add the two proportions calculated above the left and right tail. This is the estimated P-value.

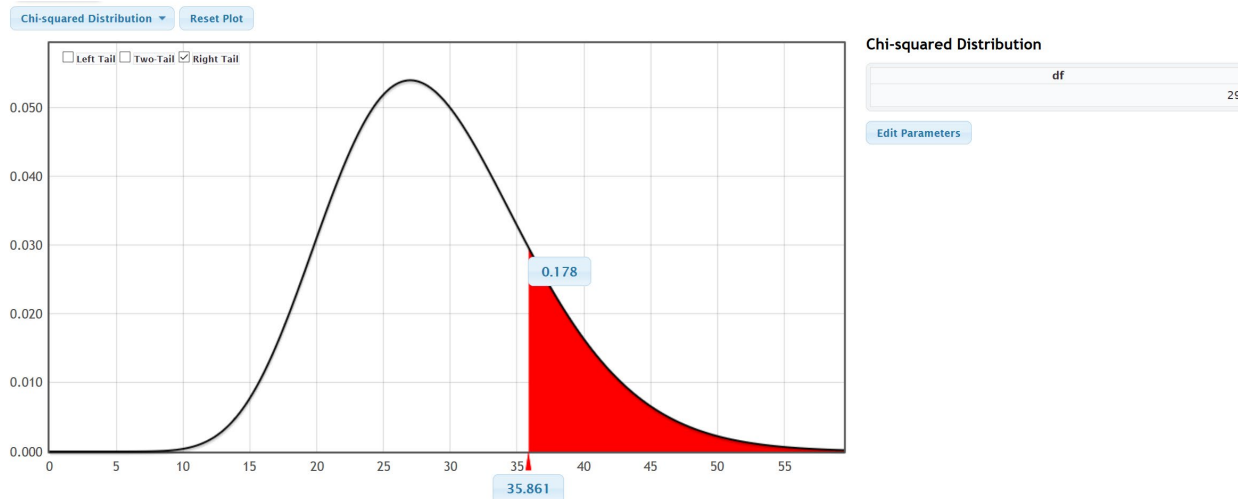




We see that the proportion in the right tail that corresponds to the T-test statistic is 0.042. We also see that the computer has calculated the left tail probability as well. Notice it is also 0.042. So the estimated P-value for this two-tailed hypothesis test is $0.042 + 0.042 = 0.084$ or 8.4%. This is the estimated P-value, the probability of getting the sample data or more extreme by random chance if the null hypothesis was true. Note that a two-tailed P-value is twice as large as a one-tailed P-value from the same data. A common formula that is often used is to take the proportion in the tail corresponding to the test statistic and multiply by two ($0.042 \times 2 = 0.084$).

Example 3: Suppose we are doing a right tailed hypothesis test that uses the chi-squared test statistic. Our test statistic compares the sample data to the null hypothesis. In this example, our chi-squared test statistic was calculated to be $\chi^2 = 35.861$ and our degrees of freedom was 29. What would the estimated P-value be?

Go to the “theoretical distributions” menu in StatKey at www.lock5stat.com and click on “ χ^2 ”. Under degrees of freedom put in 29 and then click on right tail. In the bottom box, put in the test statistic of 35.861. The proportion calculated above is the estimated P-value.



We see that the proportion in the right tail that corresponds to the test statistic cutoff is 0.178 or 17.8%. This is the estimated P-value, the probability of getting the sample data or more extreme by random chance if the null hypothesis was true.

Example 4: Most traditional statistics programs calculate the P-value with this approach. In the previous section on test statistics, we compared the number of alcoholic beverages per week that Math 140 statistics students drink and the number of alcoholic beverages per week that Math 075 pre-statistics students drink. Statcato is using the test



statistic, degrees of freedom, and theoretical T-curve to estimate the P-value. Notice that this is a two-tailed hypothesis test with a test statistic of $T=1.846$ and degrees of freedom of 800. We used these numbers with StatKey and got about the same result. In the StatKey printout, we added the tails $0.033 + 0.033 = 0.066$ to get our estimated P-value.

Hypothesis Test - Two population means: confidence level = 0.95

Samples of population 1 in Math 140 alcohol...

Samples of population 2 in Math 075 alcohol...

	N	Mean	Stdev
Population 1	322	2.224	4.684
Population 2	481	1.470	6.884

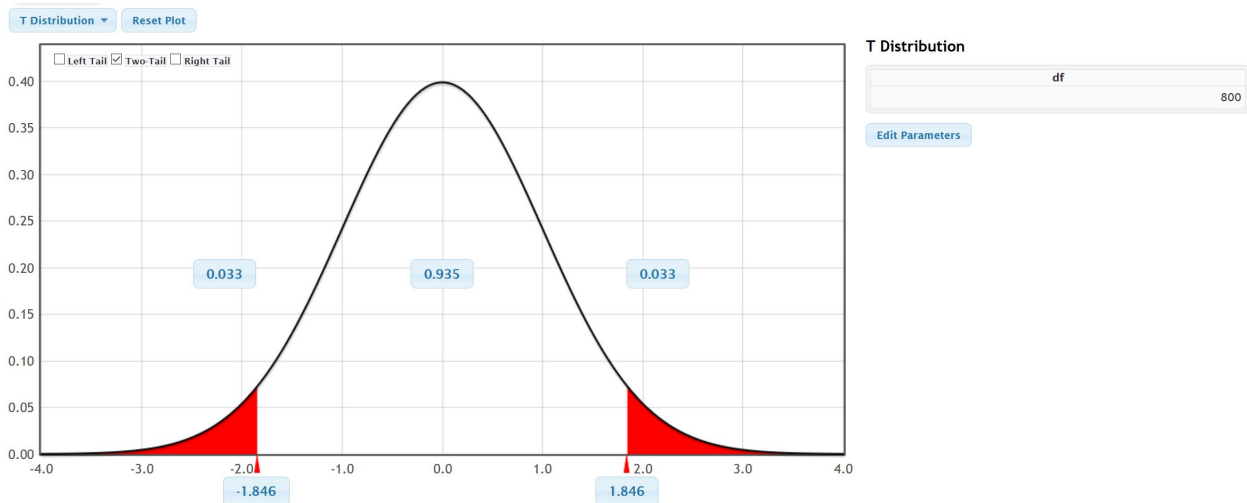
Null hypothesis: $\mu_1 - \mu_2 = 0.0$

Alternative hypothesis: $\mu_1 - \mu_2 \neq 0.0$

* Population standard deviations are unknown. *

DOF = 800

Significance Level	Critical Value	Test Statistic t	p-Value
0.05	-1.963, 1.963	1.846	0.0653



Method 2: Randomized Simulation (Randomization)

In the previous method, we saw that if we know the test statistic, we can estimate the P-value using a theoretical curve. There are many questions about the accuracy of calculating P-values in this way. For one, we need the data to meet certain assumptions to ensure that the curve is a good approximation of the sampling distribution.

Another approach that is sometimes used to calculate P-value is called “randomized simulation” or “randomization”. This is a more direct way of calculating the P-value. Let’s examine the P-value definition again.

P-value: The probability of getting the sample data or more extreme because of sampling variability (by random chance) if the null hypothesis is true.



What would sampling variability look like if the null hypothesis was true? The idea behind randomized simulation is to address this question directly. Computers can create a simulated sampling distribution under the premise that the null hypothesis is true. Once this is accomplished, we can calculate the probability of getting the sample data or more extreme directly. One and two-population hypothesis tests do not require the test statistic calculation and can calculate the P-value directly from either the sample statistic or the difference between the two sample statistics. This technique is also not tied to the accuracy of a theoretical curve, so it has other advantages as well.

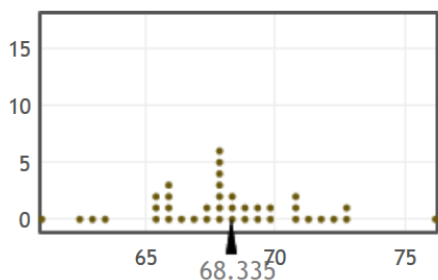
Example 1 (Simulation): Open the health data at www.matt-teachout.org and open the men's height data. Some believe that the population mean average height of all men is 68 inches. We want to test the claim that the population mean average height of men is now greater than 68 inches. Here is the null and alternative hypothesis.

$$H_0 : \mu = 68$$
$$H_A : \mu > 68 \text{ (claim)}$$

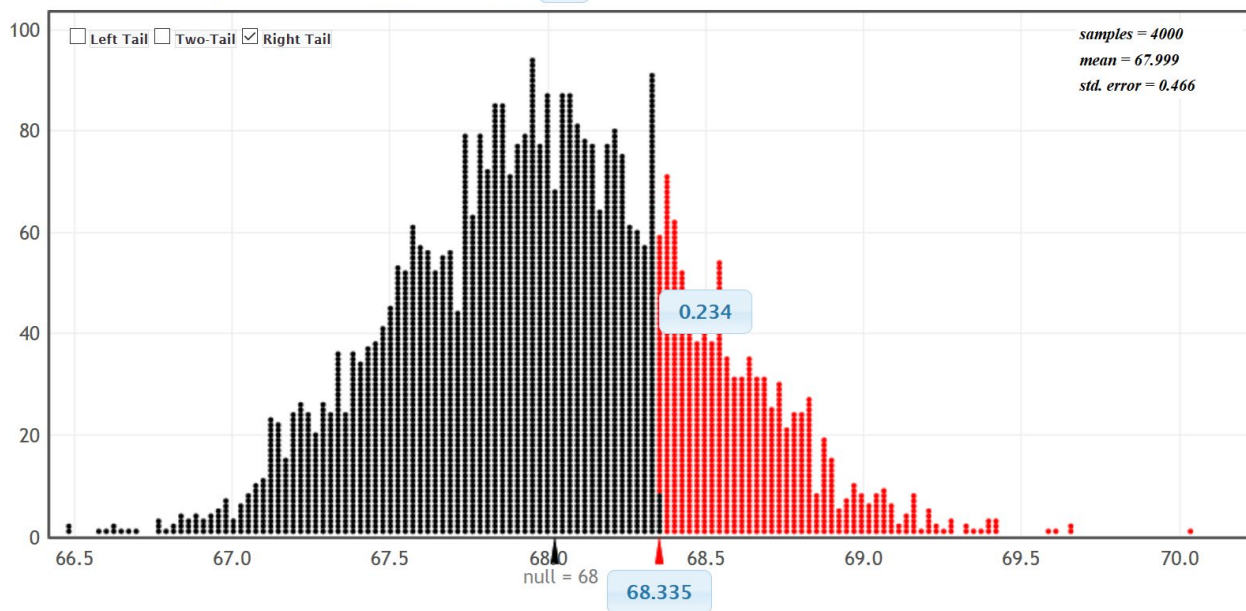
Notice this is a right tailed test. Go to the "Randomization Hypothesis Tests" menu in StatKey at www.lock5stat.com. Click on "Test for Single Mean". Click on the "Edit Data" button and copy and paste the men's height data into StatKey. Uncheck "Identifier" and check "Header Row". An identifier is a word next to every number that explains something about that value. A header row is a title. Push "OK". Change the null hypothesis to " $\mu = 68$ ". Now click the "Generate thousand samples" button a bunch of times. We are simulating what sampling variability looks like if the null hypothesis is true. In simulation, it is important to not confuse the simulated samples with the actual original random sample data. The sample mean for the original data is 68.335, so click on "Right Tail" and then put in 68.335 in the bottom box. The proportion above the sample mean is the P-value. Notice we have calculated the probability of getting the sample data or more extreme by sampling variability if the null hypothesis was true. We also did not require the test statistic to calculate it. This is called randomized simulation or randomization.

Original Sample

$n = 40$, mean = 68.335
median = 68.3, stdev = 3.02



Randomization Dotplot of \bar{x} . Null hypothesis: $\mu = 68$



We see that the sample mean of 68.335 inches is not in the tail. We also see that the estimated P-value is 0.234 or 23.4%. Both of these tell us that this sample data does not significantly disagree with the null hypothesis and could have occurred because of sampling variability. We will fail to reject the null hypothesis.

Example 2 (Simulation): Suppose we want to compare the percentage (proportion) of math 075 (pre-statistics) students that smoke cigarettes and the percentage of math 140 students that smoke cigarettes. Our claim is that the population proportions are the same. Here is the representative sample data from the Fall 2015 semester at COC and the null and alternative hypothesis.

Math 075 (pre-stat) students: 480 total students, 33 smoke cigarettes
 Math 140 (statistics) students: 330 total students, 30 smoke cigarettes

π_1 : Population proportion of pre-stat students that smoke cigarettes at COC.
 π_2 : Population proportion of statistics students that smoke cigarettes at COC.

H_0 : $\pi_1 = \pi_2$ (claim)
 H_A : $\pi_1 \neq \pi_2$

Notice this is a two-tailed two-population proportion hypothesis test. To use randomized simulation, go to the "Randomization Hypothesis Test" menu and click on "Test for Difference in Proportions". Under the "Edit Data" menu, put in the sample count and total sample size as follows. Since we designated the math 075 pre-stat students as group 1 and math 140 statistics students in group 2, we need to enter the data in that order. Now push "Ok".

Edit data ✕

Please select values for two categories of count and sample size.

Group 1 count:

Group 1 sample size:

Group 2 count:

Group 2 sample size:

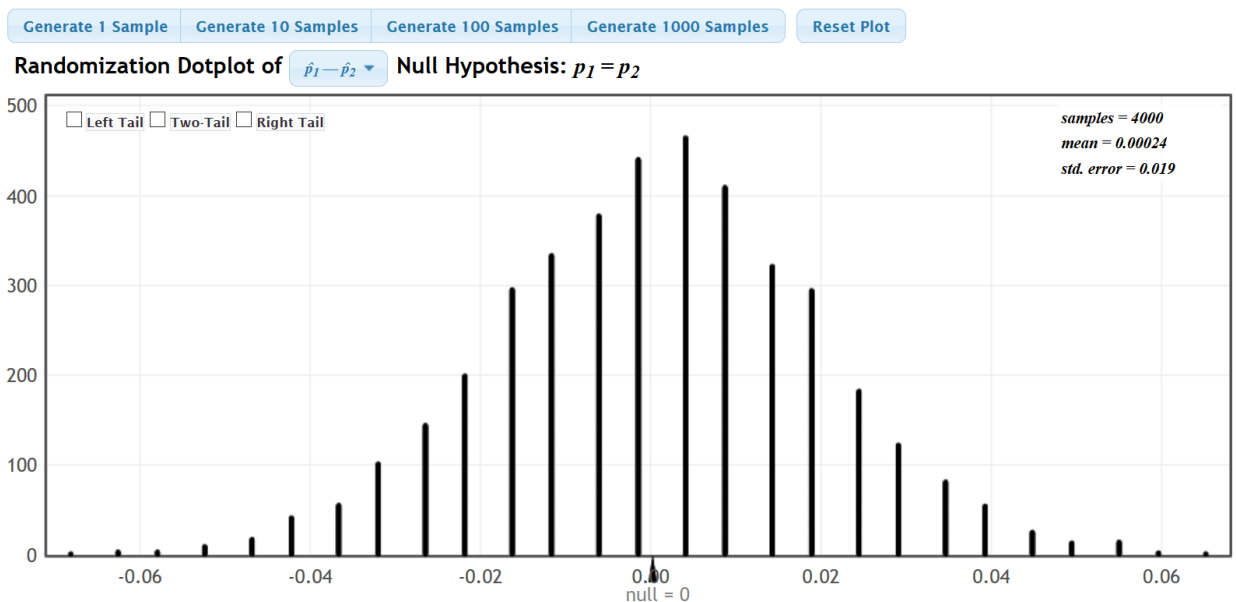


In simulation, it is important not to confuse the original real sample data with all of the simulated samples. Here is the original sample proportions. Notice that the sample proportion for the pre-stat students (\hat{p}_1) was 0.069 and the sample proportion for the stat students (\hat{p}_2) is 0.091. In two-population simulation we will be using the difference between the sample statistics ($\hat{p}_1 - \hat{p}_2$) = -0.022 to calculate the P-value.

Original Sample

Group	Count	Sample Size	Proportion
Group 1	33	480	0.069
Group 2	30	330	0.091
Group 1-Group 2	3	n/a	-0.022

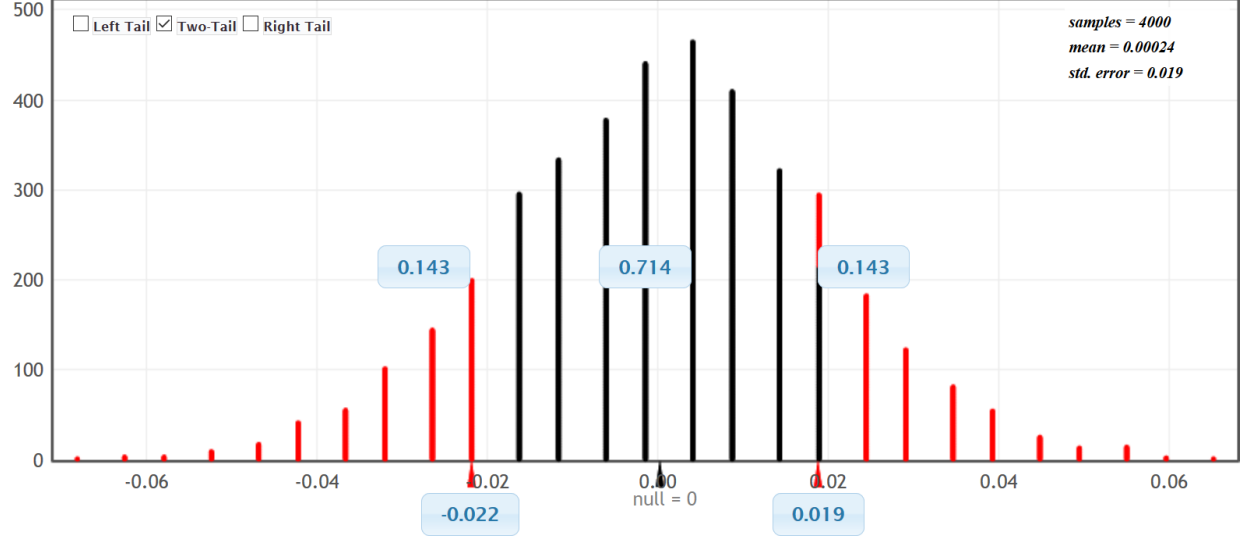
Let's simulate the null hypothesis. We are creating thousands of random samples from the premise that the populations are equal.



Click on "Two-Tail". Since the difference between the proportions was negative and in the left tail, we will put the difference -0.022 in the left tail. The right tail will automatically adjust. Remember, in a two-tailed hypothesis test, we will need to add the proportions in the top boxes of the two tails to get the estimated P-value. Notice the estimated P-value = 0.143 + 0.143 = 0.286 or 28.6%.



Randomization Dotplot of $\hat{p}_1 - \hat{p}_2$ Null Hypothesis: $p_1 = p_2$



Practice Problems 3C

For #1-32, fill out the following table to interpret the given P-value.

	P-value	P-value %	Significance Level %	Significance Level Proportion	Low P-value or High P-value?	Sample Data significantly disagree with H_0 ? (Yes or No)	Could be sampling variability or Unlikely?	Reject H_0 or fail to reject H_0 ?
1.	0.238		5%					
2.	0.0003		1%					
3.	5.7×10^{-6}		10%					
4.	0.441		5%					
5.	0.138		1%					
6.	0		10%					
7.	0.043		5%					
8.	0.085		1%					
9.	1.4×10^{-4}		10%					
10.	0.112		5%					
11.	0		1%					
12.	0.539		10%					
13.	0.0006		10%					
14.	2.5×10^{-7}		1%					
15.	0.861		5%					
16.	0.199		5%					
17.	0.034		5%					
18.	0.128		1%					
19.	8.6×10^{-4}		10%					
20.	0.0437		5%					
21.	0		1%					
22.	0.612		10%					
23.	0.087		5%					
24.	0.0048		1%					
25.	5.5×10^{-7}		10%					
26.	0.0216		5%					
27.	0.444		1%					
28.	0.0539		10%					
29.	0.722		10%					
30.	3.8×10^{-3}		1%					
31.	0.0823		5%					
32.	0.0227		5%					



33. According to a CNN report, 93% of all Americans also own a traditional phone. We disagree with this report. We think that the percentage has decreased as more and more Americans opt to only use a cell phone and throw away their traditional phones. A random sample of 106 Americans was taken and 92 of them owned a traditional phone. The p-value was found to be 0.0168. Use a 5% significance level. The null and alternative hypothesis are given below.

Ho: $\pi = 0.93$

Ha: $\pi < 0.93$ (claim)

- If the null hypothesis is true, what is probability that the sample data or more extreme occurred because of sampling variability?
- Write a standard p-value sentence to describe the true meaning of the p-value in the context of the problem.
- Was the sample data significantly different than the population parameter in H_0 ? Explain why.
- If H_0 was true, could the sample data have occurred because of sampling variability or is it unlikely to be sampling variability? (Could be or unlikely) Explain your answer.
- Use the p-value and the significance level to decide whether we should reject the null hypothesis or fail to reject the null hypothesis. Explain why.

34. According to a recent Newspaper article, the population mean average amount of time people in California spend eating and drinking per day is 1.25 hours. In order to test this claim, we take a random sample of 400 people in California. The average number of hours for the sample was 1.22 and a p-value of 0.248 was found. Use a 10% significance level. The null and alternative hypothesis are given below.

Ho: $\mu = 1.25$ hours (claim)

Ha: $\mu \neq 1.25$ hours

- If the null hypothesis is true, what is probability that the sample data or more extreme occurred because of sampling variability?
- Write a standard p-value sentence to describe the true meaning of the p-value in the context of the problem.
- Was the sample data significantly different than the population parameter in H_0 ? Explain why.
- If H_0 was true, could the sample data have occurred because of sampling variability or is it unlikely to be sampling variability? (Could be or unlikely) Explain your answer.
- Use the p-value and the significance level to decide whether we should reject the null hypothesis or fail to reject the null hypothesis. Explain why.

35. According to an article in *USA Today*, 74% of Americans own a credit card. We disagree with the *USA Today* article. We claim that more than 74% of Americans own a credit card. In order to verify the claim that more than 74% of Americans have a credit card, a random sample of 250 Americans was taken and 77.2% of them owned a credit card and a p-value of 0.1244 was found. Use a 5% significance level. The null and alternative hypothesis are given below.

Ho: $\pi = 0.74$

Ha: $\pi > 0.74$ (claim)

- If the null hypothesis is true, what is probability that the sample data or more extreme occurred because of sampling variability?
- Write a standard p-value sentence to describe the true meaning of the p-value in the context of the problem.
- Was the sample data significantly different than the population parameter in H_0 ? Explain why.



- d) If H_0 was true, could the sample data have occurred because of sampling variability or is it unlikely to be sampling variability? (Could be or unlikely) Explain your answer.
- e) Use the p-value and the significance level to decide whether we should reject the null hypothesis or fail to reject the null hypothesis. Explain why.

36. It has long been thought that the population mean average body temperature is 98.6 degrees Fahrenheit. A recent study is now claiming that the population mean average body temperature is really lower than 98.6 degrees. A random sample of 50 adults worldwide was conducted and the average temperature was 98.26 degrees with a p-value of 0.0014 was found. Use a 1% significance level. The null and alternative hypothesis are given below.

Ho: $\mu = 98.6$ degrees Fahrenheit
 Ha: $\mu < 98.6$ degrees Fahrenheit (claim)

- a) If the null hypothesis is true, what is probability that the sample data or more extreme occurred because of sampling variability?
- b) Write a standard p-value sentence to describe the true meaning of the p-value in the context of the problem.
- c) Was the sample data significantly different than the population parameter in H_0 ? Explain why.
- d) If H_0 was true, could the sample data have occurred because of sampling variability or is it unlikely to be sampling variability? (Could be or unlikely) Explain your answer.
- e) Use the p-value and the significance level to decide whether we should reject the null hypothesis or fail to reject the null hypothesis. Explain why.

37. It has been suggested that at least 10% of the world population is left handed. To test this claim, a sample of 77 randomly selected adults was taken and we found that 11 of them were left handed. A P-value of 0.895 was found. Use a 10% significance level. The null and alternative hypothesis are given below.

Ho: $\pi \geq 0.1$ (claim)
 Ha: $\pi < 0.1$

- a) If the null hypothesis is true, what is probability that the sample data or more extreme occurred because of sampling variability?
- b) Write a standard p-value sentence to describe the true meaning of the p-value in the context of the problem.
- c) Was the sample data significantly different than the population parameter in H_0 ? Explain why.
- d) If H_0 was true, could the sample data have occurred because of sampling variability or is it unlikely to be sampling variability? (Could be or unlikely) Explain your answer.
- e) Use the p-value and the significance level to decide whether we should reject the null hypothesis or fail to reject the null hypothesis. Explain why.

(#38-40) Use the “theoretical distributions” menu in StatKey at www.lock5stat.com to look up the P-value. Click on the button that says “normal”. Click on the tail and enter the test statistic in the bottom box below the tail. Remember in a two-tailed test, you will need to add the proportions in both tails to get the P-value.

38. Z-test statistic = 2.41
 Two-tailed test
 P-value =

39. Z-test statistic = -1.38
 Left-tailed test
 P-value =



40. Z-test statistic = 1.02
Right-tailed test
P-value =

(#41-43) Use the “theoretical distributions” menu in StatKey at www.lock5stat.com to look up the following critical values. Click on the button that says “ t ” and enter the given degrees of freedom. Click on the tail and enter the test statistic in the bottom box below the tail. Remember in a two-tailed test, you will need to add the proportions in both tails to get the P-value.

41. T-test statistic = -2.471
Two-tailed test
Degrees of Freedom = 29
P-value =

42. T-test statistic = 1.352
Right-tailed test
Degrees of Freedom = 34
P-value =

43. T-test statistic = -1.644
Left-tailed test
Degrees of Freedom = 49
P-value =

(#44-45) Use the “theoretical distributions” menu in StatKey at www.lock5stat.com to look up the following critical values. Click on the button that says “ χ^2 ” and enter the given degrees of freedom. Click on the tail and enter the test statistic in the bottom box below the tail. Remember in a two-tailed test, you will need to add the proportions in both tails to get the P-value.

44. χ^2 -test statistic = 38.724
Right-tailed test
Degrees of Freedom = 29
P-value =

45. χ^2 -test statistic = 12.551
left-tailed test
Degrees of Freedom = 39
P-value =



Section 3D – Conclusions

Vocabulary

Population: The collection of all people or objects to be studied.

Sample: Collecting data from a small subgroup of the population.

Statistic: A number calculated from sample data in order to understand the characteristics of the data.
For example, a sample mean average, a sample standard deviation, or a sample percentage.

Parameter: A number that describes the characteristics of a population like a population mean or a population percentage. Can be calculated from an unbiased census, but is often just a guess about the population.

Hypothesis Test: A procedure for testing a claim about a population.

Null Hypothesis (H_0): A statement about the population that involves equality. It is often a statement about “no change”, “no relationship” or “no effect”.

Alternative Hypothesis (H_A or H_1): A statement about the population that does not involve equality. It is often a statement about a “significant difference”, “significant change”, “relationship” or “effect”.

Population Claim: What someone thinks is true about a population.

Test Statistic: A number calculated in order to determine if the sample data significantly disagrees with the null hypothesis. There are a variety of different test statistics depending on the type of data.

Sampling Variability: Also called “random chance”. The principle that random samples from the same population will usually be different and give very different statistics. The random samples will usually be different than the population parameter.

P-value: The probability of getting the sample data or more extreme because of sampling variability (by random chance) if the null hypothesis is true.

Significance Level (α): Also called the Alpha Level. This is the probability of making a type 1 error. The P-value is compared to this number to determine significance and sampling variability. If the P-value is lower than the significance level, then the sample data significantly disagrees with the null hypothesis and is unlikely to have happened because of sampling variability.

Critical Value: We compare a test statistic to this number to determine if the sample data significantly disagrees with the null hypothesis. If the absolute value of the test statistic is higher than the absolute value of the critical value, then the sample data significantly disagrees with the null hypothesis.

Randomized Simulation: A technique for visualizing sampling variability in a hypothesis test. The computer assumes the null hypothesis is true, and then generates random samples. If the sample data or test statistic falls in the tail, then the sample data significantly disagrees with the null hypothesis. This technique can also calculate the P-value and standard error without a formula.

Conclusion: A final statement in a hypothesis test that addresses the claim and evidence.

Introduction: So far we have learned many of the key ideas used in hypothesis testing. We learned how to write a null and alternative hypothesis and how to determine if the sample data significantly disagrees with the null hypothesis by comparing the test statistic to the critical value. We also learned how to compare the P-value and significance level in order to judge if the sample data occurred by random chance and whether to reject the null hypothesis or fail to reject.



The last step in a hypothesis test is to write a formal conclusion. A conclusion is a statement about the claim the person made. Do we disagree with that claim (reject the claim) or do we agree with the claim (support the claim)? The conclusion is also a statement about evidence. Do we have evidence to back up our view about the population or not?

Conclusion: A final statement in a hypothesis test that addresses the claim and evidence.

Note: “Reject H_0 ” or “Fail to reject H_0 ” is NOT a conclusion. This is a simple statement of what the P-value tells us about the null hypothesis. It does not address evidence and the claim.

Writing conclusions can be difficult. Many statistics students struggle with the logic of the conclusion. Think of it this way.

Suppose we have a low P-value and H_0 is the claim. A low P-value means we have evidence and will reject H_0 . Since H_0 is our claim, then our conclusion will be that we should “reject the claim”. In other words, we have evidence that the random sample data significantly disagrees with the claim. In this case, our conclusion will be that “we have significant evidence to reject the claim”.

Suppose we have a low P-value and H_A is the claim. Again, a low P-value means that we have evidence and we will reject H_0 . Disagreeing with H_0 is equivalent to agreeing with or supporting H_A . So our conclusion will be that “we have evidence to support the claim”. In other words, random sample data significantly agrees with the claim.

Suppose we have a high P-value and H_0 is the claim. A high P-value means that we do not have evidence and we will fail to reject H_0 . Since H_0 is our claim, then we should think of our conclusion as “failing to reject the claim”. So our conclusion will be that “we do NOT have evidence to reject the claim”. In this case, the random sample data is close to H_0 and does not significantly disagree with the claim.

Suppose we have a high P-value and H_A is the claim. A high P-value is not evidence and means we will fail to reject H_0 . You have to be able to reject H_0 in order to support H_A . Since our random sample data does not significantly disagree with H_0 , it also does not significantly agree with H_A . So our conclusion is that we do not have significant evidence to support the claim. The random sample data does not significantly agree with the claim.

Let’s look at some steps for writing a formal conclusion to a hypothesis test.

Step 1: Address the Claim

If the null hypothesis (H_0) is the claim: *There are two possibilities.*

- Yes, we have evidence to reject the claim
OR
- No, we do not have evidence to reject the claim.

If the alternative hypothesis (H_A) is the claim: *There are two possibilities.*

- Yes, we have evidence to support the claim
OR
- No, we do not have evidence to support the claim.

Step 2: Address the evidence (Yes or No)

So how do we know if we have evidence or not?

We have already seen that a low P-value less than the significance level indicates significant evidence. Remember, scientists usually require a low P-value as evidence on their reports. If we have a high P-value greater than the significance level, then we do NOT have significant evidence.



Note: While P-value is the usual method for determining significance evidence, a test statistic in the tail also indicates significance evidence. For Z or T-test statistics, if the |test statistic| is larger than the |critical value| is considered significant. Remember a large test statistic in the tail will correspond to a small P-value. If the |test statistic| is smaller than the |critical value|, it is NOT considered significant.

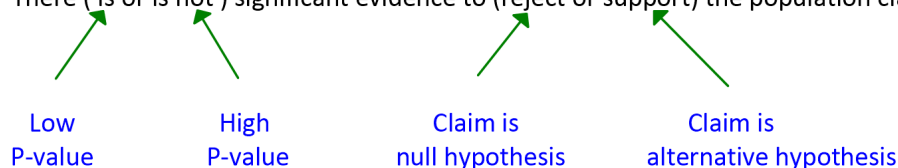
Low P-value (less than significance level) : We have significant evidence.

High P-value (higher than significance level) : We do NOT have significant evidence.

Step 3: Write the conclusion sentence

Remember a low P-value is considered significant statistical evidence but a high P-value is not evidence. When the claim is H_0 , we will either be rejecting or not rejecting the claim. When the claim is H_A , we will either be supporting or not supporting the claim.

There (is or is not) significant evidence to (reject or support) the population claim.



Step 4: Explain the conclusion sentence

Our job as data scientists, statisticians and data analysts is to explain. People rarely understand the language and difficult ideas in statistics. The conclusion is a summary of the hypothesis test, but is rarely understood. It is always good to explain the conclusion in plain language.

There is significant evidence to reject the population claim. *(Random sample data significantly disagrees with the population claim and we have a low P-value as evidence.)*

There is significant evidence to support the population claim. *(Random sample data significantly agrees with the population claim and we have a low P-value as evidence.)*

There is not significant evidence to reject the population claim. *(Unfortunately, the hypothesis test was inconclusive. Random sample data does not significantly disagree with the population claim. We do not have any statistical evidence to disagree with the population claim. The claim might be correct.)*

There is not significant evidence to support the population claim. *(Unfortunately, the hypothesis test was inconclusive. Random sample data does not significantly agree with the population claim. We do not have any statistical evidence to agree with the population claim. The claim might be incorrect.)*

Example 1

A nursing magazine recently claimed that the population mean average amount of a particular medicine that is being given to patients is about 100 milligrams. Looking at a large random sample, we found a P-value of 0.0041 and a 5% significance level ($\alpha = 0.05$) was used in the study. Assuming the data met all the assumptions, what would be the conclusion?

H_0 : $\mu = 100$ milligrams (Claim)

H_A : $\mu \neq 100$ milligrams

Step 1: Address the Claim



The claim is H_0 , so there are two possible conclusions.

- Yes, we have evidence to reject the claim
OR
- No, we do not have evidence to reject the claim.

Step 2: Address the Evidence

The P-value = 0.0041 = 0.41% is less than the 5% significance level. So this is a low P-value close to zero. This means that sampling variability (random chance) is very unlikely and tells us the sample data significantly disagrees with the null hypothesis. We would reject the null hypothesis. Since the claim is the null hypothesis, we have evidence to reject the claim.

Step 3: Writing the conclusion sentence:

“There (*is or is not*) significant evidence to (*reject or support*) the claim.”

Notice this problem has a low P-value (“is evidence”) and claim is null (“reject”).

Formal Statistics Conclusion: “There is significant evidence to reject the claim that the mean average amount of this medicine given to patients is 100 mg.”

Step 4: Explain the conclusion in plain language?

Significant statistical evidence disagrees with the claim that the population mean average is 100 mg.

Example 2

An online article is currently estimating that more than 35% of people in the U.S. voted in the last election. Looking at a large random sample, we found a P-value of 0.267 and a 10% significance level ($\alpha = 0.10$) was used in the study. Assuming the data met all the assumptions, what would be the conclusion?

$H_0: p = 0.35$
 $H_A: p > 0.35$ (claim)

Step 1: Address the Claim

The claim is H_A , so there are two possible conclusions.

- Yes, we have evidence to support the claim
OR
- No, we do not have evidence to support the claim.

Step 2: Address the Evidence

The P-value = 0.267 = 26.7% is more than the 10% significance level. So this is a high P-value and is not close to zero. This data could have happened because of sampling variability (random chance) and tells us the sample data does not significantly disagree with the null hypothesis. We would fail to reject the null hypothesis. We do not have evidence to reject the null hypothesis. You have to reject the null hypothesis to be able to support the alternative hypothesis. So we do not have evidence to support the claim.

Step 3: Writing the conclusion sentence:

“There (*is or is not*) significant evidence to (*reject or support*) the claim.”

Notice this problem has a high P-value (“NOT evidence”) and claim is null (“support”).

Formal Statistics Conclusion: “There is not significant evidence to support the claim that more than 35% of people in the U.S. voted in the last election.”



Step 4: Explain the conclusion in plain language?

The hypothesis test was inconclusive. Random sample data does not significantly agree with the population claim. We do not have significant statistical evidence to agree with the population claim.

Conclusion Summary (4 Possible Conclusions)

If the claim is H_0 , P-value is low (*Think “Yes Evidence Reject”*)

Conclusion Sentence: There is significant evidence to reject the claim.

(You are rejecting the null hypothesis and the null hypothesis is the claim.)

If the claim is H_0 , P-value is high (*Think “No Evidence Reject”*)

Conclusion Sentence: There is not significant evidence to reject the claim.

(You do not have evidence to reject the null hypothesis and the null hypothesis is the claim.)

If the claim is H_A , P-value is low (*Think “Yes Evidence Support”*)

Conclusion Sentence: There is significant evidence to support the claim.

(*You are rejecting the null hypothesis which means you think the alternative hypothesis is correct and the alternative hypothesis is the claim.*)

If the claim is H_A , P-value is high (*Think “No Evidence Support”*)

Conclusion Sentence: There is not significant evidence to support the claim.

(*You do not have evidence to reject the null hypothesis which means you do not know if the alternative hypothesis is correct and the alternative hypothesis is the claim.*)

Important Notes:

- The claim being the null hypothesis does not in itself mean that you will reject the claim. It may be a “not reject” situation. You would only reject the claim if the claim is H_0 and the P-value was low.
 - The claim being the alternative hypothesis does not in itself mean that you will support the claim. It may be a “not support” situation. You would only support the claim if the claim is H_A and the P-value was low.
 - Always address the “claim” in a conclusion. Never say that you support H_A or you reject H_0 . That is not a conclusion.
-



Practice Problems 3D

For #1-30, fill out the following table and write the formal conclusion. A “Low” P-value means that the P-value was lower than the significance level while a “High” P-value means that the P-value was higher than the significance level.

	Claim	P-value	Evidence (Yes or No)	Formal Hypothesis Test Conclusion
1.	H_0	Low		
2.	H_A	High		
3.	H_A	High		
4.	H_0	Low		
5.	H_0	High		
6.	H_A	High		
7.	H_A	Low		
8.	H_0	High		
9.	H_0	Low		
10.	H_A	Low		
11.	H_A	High		
12.	H_0	High		
13.	H_0	Low		
14.	H_A	High		
15.	H_A	Low		
16.	H_0	Low		
17.	H_0	High		
18.	H_A	Low		
19.	H_A	High		
20.	H_0	Low		
21.	H_0	Low		
22.	H_A	High		
23.	H_A	High		
24.	H_0	High		
25.	H_0	Low		
26.	H_A	Low		
27.	H_A	High		
28.	H_0	Low		
29.	H_0	High		
30.	H_A	Low		

31. The hospital claims that less than 4% of people who received the medication showed symptoms of side effects. Use a 1% significance level. (P-value = 0.0027)

Ho: $p \geq 0.04$

Ha: $p < 0.04$ (Claim)

- Should we reject or fail to reject the null hypothesis? Explain why.
- Do we have statistical evidence for our conclusion? Explain why.
- Write the formal hypothesis test conclusion sentence addressing the claim and evidence.
- Explain your conclusion in non-statistics, easy to understand language.



32. We think that the population mean average height of women is at most 63.5 inches.
Use a 5% significance level. (P-value = 0.1843)

Ho: $\mu \leq 63.5$ inches (Claim)

Ha: $\mu > 63.5$ inches

- Should we reject or fail to reject the null hypothesis? Explain why.
- Do we have statistical evidence for our conclusion? Explain why.
- Write the formal hypothesis test conclusion sentence addressing the claim and evidence.
- Explain your conclusion in non-statistics, easy to understand language.

33. Latest polls suggest that the candidate should receive about 54% of the vote.
Use a 10% significance level. (P-value = 0.0711)

Ho: $p = 0.54$ (Claim)

Ha: $p \neq 0.54$

- Should we reject or fail to reject the null hypothesis? Explain why.
- Do we have statistical evidence for our conclusion? Explain why.
- Write the formal hypothesis test conclusion sentence addressing the claim and evidence.
- Explain your conclusion in non-statistics, easy to understand language.

34. The population mean average weight of electrically powered car weighs is more than 2000 pounds. Use a 5% significance level. (P-value = 0.2682)

Ho: $\mu \leq 2000$ pounds

Ha: $\mu > 2000$ pounds (Claim)

- Should we reject or fail to reject the null hypothesis? Explain why.
- Do we have statistical evidence for our conclusion? Explain why.
- Write the formal hypothesis test conclusion sentence addressing the claim and evidence.
- Explain your conclusion in non-statistics, easy to understand language.

35. The medication Toprol is showing real promise in treating migraines. At least 50% of all patients taking Toprol have seen an improvement in their migraine symptoms.
Use a 1% significance level. (P-value = 0.0086)

Ho: $p \geq 0.5$ (Claim)

Ha: $p < 0.5$

- Should we reject or fail to reject the null hypothesis? Explain why.
- Do we have statistical evidence for our conclusion? Explain why.
- Write the formal hypothesis test conclusion sentence addressing the claim and evidence.
- Explain your conclusion in non-statistics, easy to understand language.



36. The population mean average cholesterol for men is different from the population mean average cholesterol for women. Use a 5% significance level. (P-value = 0.0391)

Ho: $\mu_1 = \mu_2$

Ha: $\mu_1 \neq \mu_2$ (Claim)

- Should we reject or fail to reject the null hypothesis? Explain why.
- Do we have statistical evidence for our conclusion? Explain why.
- Write the formal hypothesis test conclusion sentence addressing the claim and evidence.
- Explain your conclusion in non-statistics, easy to understand language.

37. In this state, the population percentage of people that will vote for the democratic candidate will be higher than the percentage of people that will vote for the republican candidate.

Use a 10% significance level. (P-value = 0.3144)

Ho: $p_1 \leq p_2$

Ha: $p_1 > p_2$ (Claim)

- Should we reject or fail to reject the null hypothesis? Explain why.
- Do we have statistical evidence for our conclusion? Explain why.
- Write the formal hypothesis test conclusion sentence addressing the claim and evidence.
- Explain your conclusion in non-statistics, easy to understand language.

Section 3E – Type 1 and Type 2 Errors

Vocabulary

Population: The collection of all people or objects to be studied.

Sample: Collecting data from a small subgroup of the population.

Random Sample: Sample data collected in such a way that everyone in the population has an equal chance to be included.

Bias: When sample data does not represent the population.

Statistic: A number calculated from sample data in order to understand the characteristics of the data.
For example, a sample mean average, a sample standard deviation, or a sample percentage.

Parameter: A number that describes the characteristics of a population like a population mean or a population percentage. Can be calculated from an unbiased census, but is often just a guess about the population.

Hypothesis Test: A procedure for testing a claim about a population.

Null Hypothesis (H_0): A statement about the population that involves equality. It is often a statement about “no change”, “no relationship” or “no effect”.

Alternative Hypothesis (H_A or H_1): A statement about the population that does not involve equality. It is often a statement about a “significant difference”, “significant change”, “relationship” or “effect”.

Population Claim: What someone thinks is true about a population.



Sampling Variability: Also called “random chance”. The principle that random samples from the same population will usually be different and give very different statistics. The random samples will usually be different than the population parameter.

P-value: The probability of getting the sample data or more extreme because of sampling variability (by random chance) if the null hypothesis is true.

Significance Level (α): Also called the Alpha Level. This is the probability of making a type 1 error. The P-value is compared to this number to determine significance and sampling variability. If the P-value is lower than the significance level, then the sample data significantly disagrees with the null hypothesis and is unlikely to have happened because of sampling variability.

Type 1 Error: When biased sample data leads you to support the alternative hypothesis when the alternative hypothesis is actually wrong in the population.

Type 2 Error: When biased sample data leads you fail to reject the null hypothesis when the null hypothesis is actually wrong in the population.

Beta Level (β): The probability of making a type 2 error.

Introduction

It is very difficult to understand large populations. Statisticians and data scientists spend years studying data, sampling variability, sample statistics and population parameters. Yet they can sometimes come to the wrong conclusion about the population. It has nothing to do with their knowledge or ability. The problem is sampling variability. In a sense, data can lead us to a wrong conclusion sometimes. Think about it this way. We are asked to try to understand millions of people in a population, but all we have is one random sample of 250 people. What if that one random sample is biased or is not reflective of what is going on in the population? That random sample can lead us to conclude something is true about the population that really is not true. When this happens, it is not the fault of the statistician or data scientist. They analyzed the data correctly and came to the correct conclusion about the population. The data has lead them astray. These wrong conclusions are called Type 1 or Type 2 errors.

Type 1 Error (False Positive)

A type 1 or type 2 error occurs because the random sample data does not reflect what is really going in the population. A type 1 error is when random sample data gives us a low P-value that is not reflective of the population. Maybe if we collected other random samples they would all have high P-values, but this one biased random sample has lead us to the wrong conclusion.

So biased sample data has given us a low P-value less than the significance level. From this, we rightly make the conclusion that we have evidence to reject the null hypothesis or support the alternative hypothesis. The problem is this the exact opposite conclusion from what is really going in the population. Type 1 errors can look bad on the statistician or data scientist. Remember a type 1 error involves a low P-value and having evidence. Later studies about the population may find that there is not significant evidence. Because the data scientist thinks they have evidence, a type 1 error is sometimes called a “false positive”.

So how can we limit the possibility of making a type 1 error? The probability of making a type 1 error is the significance level (or alpha level “ α ”). So if you want to decrease the chances of making a type 1 error, lower the significance level (lower the alpha level). This is the reason that significance levels are always set low (1%, 5%, or 10%). If the significance level is 5%, then we have only a 5% chance of making a type 1 error. The most common significance level is 5%, but when a statistician is really worried about making a type 1 error, they may lower the significance level to 1% (or a 99% confidence level). Now they have only a 1% probability of making a type 1 error. Making the wrong conclusion about a population can have serious consequences. You may see a statistician decrease the significance level to 0.5% even (99.5% confidence level). In a sense they are making sure the P-value is really close to zero before considering it evidence.



Type 1 Error Summary

- Biased random sample data leads to a low P-value.
- We support the alternative hypothesis when in the population the alternative hypothesis is wrong.
- The probability of a type 1 error is the significance level or alpha level (α)
- To decrease the probability of a type 1 error, decrease the significance level (decrease α).

So why don't statisticians always set the significance level at 1% or 0.5%? Why is the most common significance level 5%? To understand this, we need to take a look at type 2 errors.

Type 2 Error (False Negative)

Type 1 and type 2 errors are inversely related. As the probability of one decreases, the probability of the other tends to increase. So if we decrease the significance level to 1%, the probability of type 1 error will decrease, but the probability of type 2 error will increase. Let's take a look at type 2 errors so we can better understand this.

A type 2 error occurs when random sample data gives a high P-value. The statistician fails to reject the null hypothesis, but later it turns out that this is the wrong conclusion about the population. In the population, the null hypothesis is wrong. This one biased random sample indicated that they did not have evidence, but later studies about the population find that there is evidence. The null hypothesis is actually wrong and the alternative hypothesis is actually correct in the population. Since the data scientist did not have evidence, a type 2 error is sometimes referred to as a "false negative".

So how do we decrease the probability of making a type 2 error. The probability of a type 2 error is called a "Beta Level" (β). Beta levels are highly impacted by sample size. Remember the principle that more data = less error. This is particularly true for type 2 errors. A small sample size usually has a larger probability of making a type 2 error. A small sample size will have a larger beta level. So the main technique for decreasing the probability of making a type 2 error, is to simply collect more data. Increasing the sample size (n) will decrease the probability of making a type 2 error.

Notice that the best way to limit the chances of making a type 1 or type 2 error is to have a large random sample and a small significance level. This is the standard for collecting data.

Type 2 Error Summary

- Biased random sample data leads to a high P-value.
- We fail to reject the null hypothesis when in the population the null hypothesis is wrong.
- The probability of a type 2 error is called the beta level (β).
- To decrease the probability of a type 2 error, increase the sample size. Collect more data.

Setting your Significance and Confidence Levels

Before ever collecting data, a statistician thinks about the consequences of making a type 1 or type 2 error. They set the significance level for the test based on that assessment. Let's look at some of the situations that may lead to using a 1%, 5% or 10% significance level in a hypothesis test.

1% Significance Level ($\alpha = 0.01$): This corresponds to a 99% confidence level. Setting the significance level this low is trying to decrease the probability of type 1 error. At a 1% significance level, the probability of type 2 error will be higher. So this is usually a situation, where someone is trying to avoid a type 1 error, at the expense of allowing a higher probability of type 2 error. A good rule is that if you set the significance level at 1%, collect more random data so that the probability of type 2 error stays relatively low. Collecting more data is not always possible though.

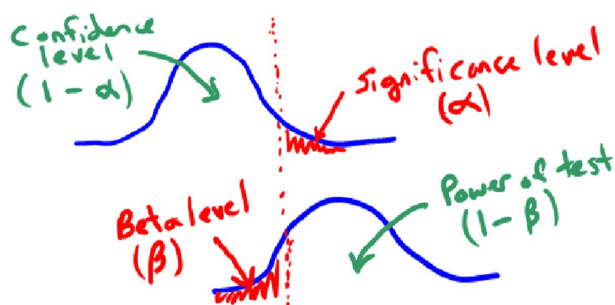
10% Significance Level ($\alpha = 0.10$): This corresponds to a 90% confidence level. Setting the significance level this high will increase the probability of type 1 error. Remember, setting the alpha level high will decrease the beta level. So the probability of type 2 error will be lower. 10% significance levels are sometimes used to decrease the probability of type 2 error when you are not able to collect more data. The scientist is willing to risk the type 1 error, but does not want a type 2 error.



5% Significance level ($\alpha = 0.10$): This corresponds to a 95% confidence level. This is the most common significance and confidence level for a good reason. At a 5% significance level, the alpha and beta levels are both relatively low. So setting the significance level at 5%, keeps the probabilities of type 1 and type 2 errors relatively low.

So why are alpha and beta levels inversely related? Think of the probability distributions associated with a type 1 error is the “alpha curve”. The probability of a type 1 error is the significance level and the complement of the significance level is the confidence level. The probability distribution associated with a type 2 error is the “beta curve”. The complement of the beta level is called the “power” of the hypothesis test. So high confidence will correspond to a low alpha level or a low probability of type 1 error. A hypothesis test with a high power has a low beta level or a low probability of type two error. Remember that larger sample sizes result in a lower probability of type 2 error. As the sample size increases, your hypothesis test power also increases.

Diagram of Alpha levels, Beta levels, confidence levels, and power.



A 1% significance level is like taking the line above and moving it to the right. In the alpha curve on top, the probability in the right tail (significance level) decreases, but look what happens to the beta curve on the bottom. As we pull the line to the right, the beta level increases. Similarly, if we set the significance level at 10% we are pulling the line to the left. The significance level is increasing, but the beta level is now getting smaller.

Example

A pharmaceutical company wants to sell a new medicine in the U.S. To get approval they need to convince the FDA that the medicine is safe and has few side effects. If side effects happen in 3% or more of the people taking the medicine, then the FDA may not approve sale of the medicine in the U.S. If side effects happen in less than 3% of people taking the medicine, then the FDA may approve sale of the medicine in the U.S.

What is the null and alternative hypothesis?

Ho: $p \geq 3\%$ (FDA does not allow medicine to be sold in U.S.)

Ha: $p < 3\%$ (FDA does allow medicine to be sold in U.S.)

Describe the consequences of a type 1 error and what we could do to limit the probability of a type 1 error.

Because of some biased sample data, we got a low P-value and think that the alternative hypothesis is correct when it is not. (In reality, the null hypothesis is correct.) That would mean that the FDA approved sale of the medicine by mistake (false positive). The medicine causes serious side effects in a lot of people. People could die or become very sick. They may sue the pharmaceutical company or the FDA.

This is about as bad of a type 1 error that we can possibly have. To make sure this doesn't happen, we will lower the significance level to 1% (or even lower).

Describe the consequences of a type 2 error and what we could do to limit the probability of a type 2 error.



Because of some biased sample data, we got a high P-value and failed to reject the null hypothesis when in the population, the null hypothesis is really wrong and the alternative hypothesis is correct. So we think that the null hypothesis may be correct when it is not. That would mean that the FDA blocked the sale of a good medicine that rarely causes any side effects. Patients will be deprived of a good medicine and the pharmaceutical company will lose a lot of money in potential profits.

To make sure a type 2 error does not happen, we can increase the sample size. We can collect more data before calculating the P-value and making a decision. We should not increase the significance level just to avoid this type 2 error. We should keep the significance level low because a type 1 error of deaths and side effects is much worse than the pharmaceutical company losing money is.

Practice Problems 3E

Directions: Answer the following questions.

1. Which error is often called a “false negative”?
2. Which error is often called a “false positive”?
3. What is the probability of a type 2 error occurring called?
4. What is the probability of a type 1 error occurring called?
5. What is the complement of the alpha level called?
6. What is the complement of the beta level called?
7. What could we do to decrease the probability that a type 2 error occurs?
8. What could we do to decrease the probability that a type 1 error occurs?
9. If we increase the significance level from 5% to 10% and keep the sample size the same, what will happen to the probabilities for type 1 and type 2 errors?
10. If we decrease the significance level from 5% to 1% and keep the sample size the same, what will happen to the probabilities for type 1 and type 2 errors?
11. What significance level achieves keeps both type 1 and type 2 errors relatively low?
12. Why do type 1 and type 2 errors sometimes occur?
13. If random sample data yielded a high P-value, what kind of error might occur?
14. If random sample data yielded a low P-value, what kind of error might occur?



15. A car company is debating whether to recall its vehicles because of a malfunction in its airbags. Executives think that the defect rate is probably low, but if the airbags malfunction and do not open in 2% or more of crashes, then they will need to put out a general recall. If there are relatively few defective airbags then the company prefers to fix them as needed and not put out a general recall. A random sample of vehicles have their airbags checked for defaults. (The study is currently using a 5% significance level.)

H_0 : $p \geq 0.02$ (Large amount of malfunctioning airbags resulting in recalling all of the vehicles for replacement airbags.)

H_A : $p < 0.02$ (Relatively few malfunctioning airbags resulting in not recalling all of the vehicles for replacement airbags.)

- Write a description of a type 1 error and possible consequences of that error in the context of the problem.
- Write a description of a type 2 error and possible consequences of that error in the context of the problem.
- Would you recommend any changes to the significance level or sample size based on what you know about the type 1 and type 2 errors in this problem? Explain.

16. Mike and his advertisement team have created an advertisement plan for a new flavor of soda. Right now, approximately 16% of soda drinkers are purchasing this flavor. Mike needs to show his bosses that his advertisement plan will increase the percentage of soda drinkers purchasing this new flavor. If Mike's advertising team succeeds in increasing the percentage of customers that prefer this new flavor, then the company will increase supply and make more of the soda to meet demand. If not, then the company will keep the supply as it currently is. After the advertising changes, Mike takes a random sample of customers to determine if the percentage of soda drinkers that like the new flavor has increased. (They are currently using a 5% significance level).

H_0 : $p = 0.16$ (The company will not increase production of the new flavor of soda.)

H_A : $p > 0.16$ (The company needs to increase production of the new flavor of soda to meet the increased demand.)

- Write a description of a type 1 error and possible consequences of that error in the context of the problem.
- Write a description of a type 2 error and possible consequences of that error in the context of the problem.
- Would you recommend any changes to the significance level or sample size based on what you know about the type 1 and type 2 errors in this problem? Explain.

17. Trisha is studying trends in a particular stock to determine if the price of the stock per week will increase or decrease. If the population slope (β_1) in dollars per week is negative, the stock price may decrease and she will recommend selling the stock. If the population slope (β_1) is zero or positive, the stock price may not decrease and she will recommend holding onto the stock.

H_0 : $\beta_1 \geq 0$ (Trisha recommends keeping the stock.)

H_A : $\beta_1 < 0$ (Trisha recommends selling the stock.)

- Write a description of a type 1 error and possible consequences of that error in the context of the problem.
- Write a description of a type 2 error and possible consequences of that error in the context of the problem.
- Would you recommend any changes to the significance level or sample size based on what you know about the type 1 and type 2 errors in this problem? Explain.



18. A global sportswear company is contemplating contributing money to a political candidate in the next election. The managers of the company do not want to contribute unless they are sure the candidate will get the majority of the population vote and win the election. Otherwise, the company will not contribute to the candidates' campaign.

H_0 : $p \leq 0.5$ (The company will not contribute to the political candidates' campaign.)

H_A : $p > 0.5$ (The company will contribute to the political candidates' campaign.)

- Write a description of a type 1 error and possible consequences of that error in the context of the problem.
- Write a description of a type 2 error and possible consequences of that error in the context of the problem.
- Would you recommend any changes to the significance level or sample size based on what you know about the type 1 and type 2 errors in this problem? Explain.

19. A professional basketball team wants to determine whether they should spend millions of dollars to sign a player to a new contract. The team determines that they will sign the player if the population mean average number of points scored per game by the player will be at least twenty. If not, then they will not sign the player.

H_0 : $\mu \geq 20$ (The team will sign the player to a new contract.)

H_A : $\mu < 20$ (The team will not sign the player to a new contract.)

- Write a description of a type 1 error and possible consequences of that error in the context of the problem.
- Write a description of a type 2 error and possible consequences of that error in the context of the problem.
- Would you recommend any changes to the significance level or sample size based on what you know about the type 1 and type 2 errors in this problem? Explain.

Section 3F – One-population Mean and Proportion Hypothesis Tests

Vocabulary

Population: The collection of all people or objects to be studied.

Sample: Collecting data from a small subgroup of the population.

Random Sample: Sample data collected in such a way that everyone in the population has an equal chance to be included.

Statistic: A number calculated from sample data in order to understand the characteristics of the data. For example, a sample mean average, a sample standard deviation, or a sample percentage.

Parameter: A number that describes the characteristics of a population like a population mean or a population percentage. Can be calculated from an unbiased census, but is often just a guess about the population.

Hypothesis Test: A procedure for testing a claim about a population.

Null Hypothesis (H_0): A statement about the population that involves equality. It is often a statement about "no change", "no relationship" or "no effect".

Alternative Hypothesis (H_A or H_1): A statement about the population that does not involve equality. It is often a statement about a "significant difference", "significant change", "relationship" or "effect".

Population Claim: What someone thinks is true about a population.



Test Statistic: A number calculated in order to determine if the sample data significantly disagrees with the null hypothesis. There are a variety of different test statistics depending on the type of data.

One-Population Proportion Test Statistic (z): The sample proportion is this many standard errors above or below the population proportion in the null hypothesis.

One-Population Mean Test Statistic (t): The sample mean is this many standard errors above or below the population mean in the null hypothesis.

Critical Value: A number we compare our test statistic to in order to determine significance. In a sampling distribution or a theoretical distribution approximating the sampling distribution, the critical value shows us where the tail or tails are. The test statistic must fall in the tail to be significant.

Sampling Variability: Also called “random chance”. The principle that random samples from the same population will usually be different and give very different statistics. The random samples will usually be different than the population parameter.

P-value: The probability of getting the sample data or more extreme because of sampling variability (by random chance) if the null hypothesis is true.

Significance Level (α): Also called the Alpha Level. This is the probability of making a type 1 error. The P-value is compared to this number to determine significance and sampling variability. If the P-value is lower than the significance level, then the sample data significantly disagrees with the null hypothesis and is unlikely to have happened because of sampling variability.

Randomized Simulation: A technique for visualizing sampling variability in a hypothesis test. The computer assumes the null hypothesis is true, and then generates random samples. If the sample data or test statistic falls in the tail, then the sample data significantly disagrees with the null hypothesis. This technique can also calculate the P-value and standard error without a formula.

Rejecting the null hypothesis: Random sample data significantly disagrees with the null hypothesis.

Failing to reject the null hypothesis: Random sample data does not significantly disagree with the null hypothesis. This does not prove that the null hypothesis is correct however.

Conclusion: A final statement in a hypothesis test that addresses the claim and evidence.

Introduction

So far, we have been learning all of the pieces and ideas behind a hypothesis test. In this section, we will finally start to put the pieces together and perform a hypothesis test from start to finish. While different tests have different test statistics, null and alternative hypotheses, and assumptions, the ideas and methods are very similar. Let's start by looking at the steps for performing a hypothesis test.

Steps to perform a Hypothesis Test

1. Write the null and alternative hypothesis.

Write down the population claim and the null and alternative hypothesis. Consider the type of data you will need and the number of populations to decide the appropriate test for the situation. Is this a right tailed, left tailed or two tailed test? What test statistic should we use? Consider the type 1 and type 2 errors and pick a significance level.



2. Collect random sample data and check the assumptions.

If the sample data has not been collected yet, collect random sample data that is appropriate. Once the sample data is collected, check the assumptions to make sure the data represents the population and can be used for the hypothesis test. You may need to create a histogram to check normality. If certain assumptions are not met, consider revising your hypothesis test method. You may need to move to a non-parametric test or use a randomization technique.

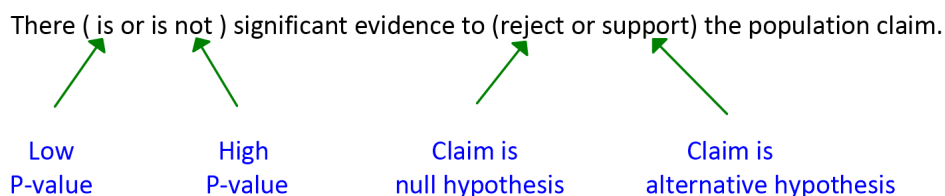
3. Use computer technology to calculate the test statistic, critical value, and P-value and use them to interpret significance and sampling variability.

Compare the test statistic to the critical value. If the test statistic falls in the tail, then the sample data significantly disagrees with the null hypothesis. Compare the P-value to your chosen significance level. If the P-value is less than the significance level, then it is unlikely to be sampling variability.

4. Determine if we should reject or fail to reject the null hypothesis.

Remember the P-value is calculated from the assumption that the null hypothesis is true. In a sense, this is about seeing what your P-value tells you about the null hypothesis. If the P-value is less than the significance level, then we will be rejecting the null hypothesis. If the P-value is higher than the significance level, then we will be failing to reject the null hypothesis. Failing to reject does not mean that the null hypothesis is correct.

5. Write the formal conclusion addressing the claim and evidence.



6. Explain the conclusion in plain language that non-statisticians can understand.

Does the random sample data disagree with claim or agree with the claim? Is it a significant disagreement or a significant agreement? Do we have evidence? What kind of evidence?

Assumptions

In this chapter, we have been discussing test statistics, critical values, theoretical distributions (curves) and P-values. These are the building blocks of hypothesis tests. A key question to think about is how accurate are these building blocks? The standard error is the standard deviation of a sampling distribution. What if that sampling distribution is not normal? The standard error may not be very accurate. To deal with theoretical curves and formulas that involve standard error, the sample data must meet certain assumptions. The good news is that the assumptions for one or two population hypothesis tests are the same as the assumptions we learned for confidence intervals. In our next chapter, we will look at more advanced hypothesis tests where the assumptions get more complicated. Remember the data must meet all of the assumptions for us to use the traditional hypothesis test methods.

Notice that hypothesis tests have some assumptions in common.

Random

One is that the data should be a random sample that is unbiased and represents the population. Statistics students often ask about whether the data can be an unbiased census. The answer is no, but this does not mean that a census is not great data. It is the best. In some ways it is too good. If you have an unbiased census then you know what is going in the population. There is no need for a hypothesis test to surmise what may be true about the population, you know the population parameters. A hypothesis test uses sample data to try to understand population when a census is not possible.



Independence

Another assumption that is particularly difficult to check is that the data values within each sample must be independent of each other. This means that the individuals do not have something in common, at least not for the variable you are studying. This usually rules out the same people measured twice or family members or people from the same classroom or on the same Facebook page. If you know the characteristics of one person, it should not mean that the next person in your data set has a higher probability of that characteristic. A simple random sample from a large population will usually pass the independence criteria. Data scientists usually like the population to be at least ten times larger than the sample size for the individuals to be independent. Think about it. If we collect a simple random sample of 350 people from a population of millions, it is very unlikely that I will accidentally get family members or people that know each other.

Sample Size Requirements

Most assumptions also have some sort of way of checking that the data set is large enough to ensure that it matches the theoretical curve. For proportions, this is usually having at least ten successes and at least ten failures. So ten or more people or objects in the sample data have the categorical criterion you are studying and ten or more do not. If the sample data has at least ten successes and failures, then the sampling distribution for sample proportions is likely to look normal, the standard error will be roughly accurate and the Z-distribution will be a good fit. For mean averages, this is usually having a sample size greater than or equal to 30 or nearly normal. If one of these two criteria is met, then the sampling distribution for sample means is likely to look normal, the standard error will be roughly accurate, and the T-distribution will be a good fit.

Key Question: If we have not collected the data yet, is there a way to know if the sample size is large enough for a given hypothesis test?

Absolutely. If the random sample data has not been collected yet, we can determine what sample sizes should be appropriate. It is generally at least 30 for mean average hypothesis tests, but is more difficult for proportions. We need to collect enough data to ensure that we will get at least ten successes and at least ten failures. Since we have an approximate population proportion (π) in the null hypothesis, we can use this to calculate how many success and failures I am likely to get.

Expected number of successes = $n \times \pi$

Expected number of failures = $n \times (1 - \pi)$

For example, suppose the null hypothesis is $\pi = 0.20$. Will a sample size of 120 randomly selected people be enough?

Expected number of successes = $n \times \pi = 120 \times 0.2 = 24$

Expected number of failures = $n \times (1 - \pi) = 120 \times (1 - 0.2) = 120 \times 0.8 = 96$

Since both of these are greater than ten, a random sample of 120 should be enough.

You can also turn these formulas around and calculate the minimum sample size to get ten. I don't like these too much because you want your expected to be significantly greater than ten. If you calculate for an expected number of success and failures of exactly ten, you may get 8. (We don't always get what we expect in random data.)

In the previous example, here are the minimum sample size calculations.

Minimum sample size (n) based on number of successes and population proportion = $\frac{10}{\pi} = \frac{10}{0.2} = 50$

Minimum sample size (n) based on number of failures and population proportion = $\frac{10}{(1-\pi)} = \frac{10}{0.8} = 12.5$

Always take the larger of the two calculations. In this example, the population proportion was 0.2. We will need a random sample of above 50 if we expect to get at least ten successes and failures.



One-Population Hypothesis Test Assumptions

One-population Mean Assumptions

- The quantitative sample data should be collected randomly or be representative of the population.
- Data values within the sample should be independent of each other.
- The sample size should be at least 30 or have a nearly normal shape.

One-population Proportion Assumptions

- The categorical sample data should be collected randomly or be representative of the population.
- Data values within the sample should be independent of each other.
- There should be at least ten successes and at least ten failures.

Example 1

Tim needed to collect some data to test a claim that the population mean average monthly cost of living in California is higher than in Nevada. He put a survey on Facebook. He received data from 217 people living in California and 168 people living in Nevada. The shape of both of the data sets was skewed right. Can we use this data to test perform a two-population mean T-test?

Are the data sets random or representative? *No. This was voluntary response data.*

Are both samples at least 30 or Normal? *Yes. The data sets were not bell shaped (skewed right) but since both of the sample sizes were over 30 (217 and 168) it does pass the 30 or normal requirement.*

Are the data values within the samples independent? *No. These people came from the same Facebook page. They may know each other and even have similar jobs, socioeconomic levels and salaries.*

Are the data values between the samples independent? *No. These California and Nevada samples came from the same Facebook page. They may know each other and even have similar jobs, socioeconomic levels and salaries.*

Since this data did not pass all of the assumptions, we should not use it to judge a population claim with a hypothesis test.

Example 2

Julie needs to test a claim about the percent of people in her hometown that will vote. She took a simple random sample of 45 people and found that 32 said they would vote and 13 said they would not. Can we use this data to test a claim about the population proportion (percentage)?

Is the sample data random or representative? *Yes. This was a simple random sample.*

Are there at least 10 success? *Yes. There are at least 10 success (32 said they will vote)*

Are there at least 10 failures? *Yes. There are at least 10 failures (13 said they will not vote)*

Are the data values within the samples independent? *Yes. There were 45 people out of a relatively large population. So it is unlikely that they would be family members. One person saying they vote will not affect the probability of other people voting. They did come from the same town, but that is the population of interest, so will not be a factor in independence.*

Since the data met the assumptions, we can use this data to test the claim about the population proportion with a hypothesis test and the Z-test statistic. We just need the number of events (voters) = 32 and the total number of trials = 45.



What if the data does not meet all of the traditional assumptions?

Are there ways of working with sample data that does not meet the traditional assumptions?

Absolutely. The one assumption that really needs to be there for any population hypothesis test is that the sample data should be representative, unbiased, and collected randomly. If we are conducting an experiment to prove cause and effect, we will need to separate our groups with random assignment. Besides that, there are many options.

Many advanced techniques in statistics were developed to deal with not meeting all of the traditional assumptions. Non-parametric hypothesis tests were invented for just this purpose. They generally have less assumptions and can be used in a variety of situations. I am a big fan of the non-parametric “Mann-Whitney” test when comparing population averages. Non-parametric hypothesis tests are usually covered in more advanced statistics courses.

One technique that we have discussed is randomized simulation or a randomization test. In some ways, this falls under the umbrella of non-parametric tests. In the section on P-value, we saw that we can create a simulated sampling distribution based on the premise that the null hypothesis is true. We can then calculate the P-value directly from the simulation. This technique does not require the data to match up with a theoretical distribution curve, so randomization often does not have as many requirements as traditional approaches to hypothesis testing. Notice that there is no requirements for sample size or a normal shape.

Assumptions for a Randomization (Randomized Simulation) Hypothesis Test

- The sample data should be collected randomly or be representative of the population.
- Data values within each sample should be independent of each other.
- If multiple samples were collected that were not matched pair, then the data values between the samples should be independent of each other.

Using Technology to Calculate Test Statistics, Critical Values and P-value

Always use technology to calculate test statistics, critical values and P-values. We may need to calculate something from time to time, but always use a computer to do the bulk of the calculations. Your job as a data analyst is to analyze and explain data, not to calculate.

In this book, we have been using two free programs, StatKey and Statcato. Here are some basic directions for using Statcato. Statcato printouts will usually be provided for you to analyze in the problems.

Statcato One-population percentage (proportion) Hypothesis Test with Z-test statistic

Statistics => Hypothesis Tests => 1-Population Proportion

- “Samples in Column” (If you have raw categorical data, enter the column. For example “C1”.)
- “Summarized Sample Data” (Use this option if you have the number of events, total number of trials).
- Put in significance level as a decimal proportion.
- “Hypothesized proportion” (This is asking for the population proportion (π) in H_0)
- “Alternative Hypothesis” (Less than, greater than, not equal. This tells the computer whether to do a left tail, right tail or two-tail hypothesis test.)



Statcato One-population mean average Hypothesis Test with T-test statistic

Statistics => Hypothesis Tests => 1-Population Mean

- “Samples in Column” (If you have raw quantitative data, enter the column. For example “C1”.)
- “Summarized Sample Data” (Use this option if you have the sample size, sample mean and sample standard deviation).
- Put in significance level as a decimal proportion.
- “Hypothesized mean” (This is asking for the population mean (μ) in H_0)
- “Alternative Hypothesis” (Less than, greater than, not equal. This tells the computer whether to do a left-tail, right-tail or two-tail hypothesis test.)
- Leave population standard deviation as “unknown”.

Example 1 (One Population Proportion Z-test)

Ex) A doctor thinks that the population percent of people in his city that have a certain infection is about 6%. He took a simple random sample of 175 people in the city and found that 13 of them had the infection. Use the following Statcato printout to test the doctor’s claim that exactly 6% have the infection. (Use a 5% significance level.)

Hypothesis Test - One population proportion: confidence level = 0.95

Input: Summary data

Null hypothesis: $p = 0.06$

Alternative hypothesis: $p \neq 0.06$

N	Sample Proportion	Significance Level	Critical Value	Test Statistic Z	p-Value
175	0.074	0.05	-1.96, 1.96	0.796	0.4262

Null and alternative hypothesis?

$H_0 : \pi = 0.06$ (claim)

$H_A : \pi \neq 0.06$

Type of hypothesis test? One population proportion Z-test (two tail)

Assumptions? The data did pass all of the assumptions, so we can proceed with the hypothesis test.

- The sample data was collected randomly.
- The sample data was sufficiently large since it had at least ten success and failures. The data had 13 success and $175-13 = 162$ failures.
- Individuals were likely to be independent since it was a simple random sample out of a large population. It is unlikely that the individuals in the sample data will be related.

Test stat Z = 0.796

Test Stat Sentence: The sample percent 7.4% was only 0.796 standard errors above the population value 6%.

- Z - Test Statistic does not fall in the tail and so is not significant. (The sample proportion needs to be higher than +1.96 or more to be in the right tail (significantly higher) or -1.96 or less to be in the left tail (significant lower.)
- This tells us that the sample value (7.4%) was not significantly different from the population value (6%)
- The sample data does not significantly disagree with the null hypothesis.



P-value = 0.426

P-Value Sentence: If H_0 is true, and the population percent really is 6%, we had a 42.6% probability of getting the sample percentage of 7.4% or more extreme by random chance.

- This is a high P-value. (The P-value of 42.6% is much larger than the 5% significance level.)
- If H_0 is true, this tells us that the sample data could have happened by random chance (sampling variability).
- Sample value of 7.4% is not significantly different from the population value of 6%.
- There is not a significant disagreement between the sample data and the null hypothesis.

Reject H_0 or Fail to reject H_0 ? Fail to reject H_0 since the P-value is larger than the significance level.

Conclusion?

There is not significant evidence to reject the claim that 6% of the population have the infection.

(The random sample data does not significantly disagree with the claim that 6% of the city is infected. The doctor might be correct, but we do not have evidence.)

Practice Problems Section 3F

1. Write out the traditional assumptions for a one-population proportion Z-test.
2. Write out the traditional assumptions for a one-population mean average T-test.
3. Write out the assumptions for a one-population randomized simulation hypothesis test.

(#4-7) Directions: Check the assumptions for a one-population proportion Z-test. You must check all of the assumptions. Explain your answers. Does the problem meet all the assumptions? Can we use this data to perform the hypothesis test?

4. Marsha works for the Republican Party and is asked to test a claim about the percentage of people in Sacramento will vote for the republican candidate in the next election. She has a computer randomly pick phone numbers with a Sacramento area code. She then calls the phone numbers and asks people if they would vote for the republican candidate. She spoke with 123 people and 37 said they would vote for the republican candidate.

5. A health organization is doing a study on smoking tobacco and need to test a claim about the population percentage of tobacco smokers use a pipe. They had people hang out in stores that sell tobacco and pipes and counted how many total people came to the store and how many of them used a pipe. They found that out of the 79 people, eight of them used a pipe.

6. The COC Admissions department needs to test a claim about the population percentage of students that would be in favor of using a new program to register for classes. They put a link on their website so that any students that want to try out the program can. The students can then take a survey and say how well they like the new system. A total of 247 students tried the system and 112 of them said they liked the new system.

7. Michelle, a teacher at Valencia High, wants to test a claim about the population percentage of students at Valencia High school will be attending COC. She gives the students in her English 1 class a questionnaire to fill out that asks where they will be attending college. Of the 34 students in her class, 26 are planning to attend COC.

(#8-11) Directions: Check the assumptions for traditional one-population mean average T-test. You must check all of the assumptions. Explain your answers. Does the problem meet all the assumptions? Can we use this data to perform the hypothesis tests?



8. A company wants to test a claim about the population mean average amount of alcohol drunk by people vacationing in Las Vegas per day. They also want to test a claim about the population standard deviation. They posted a survey on Facebook asking how much people drink when on vacation in Las Vegas. A total of 8,355 people responded and listed how much they drink. The sample mean was 3.5 drinks per day with a standard deviation of 1.2 drinks. A histogram of the alcohol consumption data was skewed to the right.

9. Jimmy works for a company that designs new homes and just moved to Oklahoma City. Jimmy's boss wants him to test a claim about the population mean average price of all homes in Oklahoma City and test a claim about the population variance. Jimmy had a computer randomly select 28 homes. He then found out the price paid for each home. A histogram of the home prices showed a normal distribution. The sample mean average price of the 28 homes was \$212.4 thousand dollars with a standard deviation of \$27.6 thousand dollars.

10. Rick works for a sports equipment manufacturing company. He needs to test a claim about the population mean average amount customers spend at his stores and test a claim about the population standard deviation. He went to the store by his house and kept track of how much people spent on a Tuesday. There were 45 customers and the histogram of the data showed a skewed right distribution. The mean average price of the sample was \$71 with a standard deviation of \$19.

11. Mike is trying to take an opinion poll and found out the average amount of money in thousands of dollars that people in Los Angeles would pay per year in order to have an NFL football team. He randomly selects three streets in Los Angeles and asks every person living on those streets. His sample size was 63 people, but the histogram was drastically skewed right. The mean average amount of money was 1.3 (thousand dollars) with a standard deviation of 0.6 (thousand dollars).

(#12-24) Directions: For each of the following problems answer the following questions.

a) Give the null and alternative hypothesis. Which is the claim? Is this a right-tailed, left-tailed, or two-tailed test? Explain how you know what tail to use.

b) Check the assumptions.

c) Give the test statistic and write the standard sentence to explain it. Compare your test statistic to the critical value. Did the sample data significantly disagree with the null hypothesis? Explain how.

d) Give the p-value and write the definition sentence to explain it. Could the sample data have happened because of sampling variability (random chance) or is it unlikely to be sampling variability? Explain why.

e) Compare the p-value to the significance level. State whether you reject the null hypothesis or fail to reject the null hypothesis. Explain your answer.

f) Write the standard conclusion.

g) Explain your conclusion in easy to understand language.

12. The United States has the highest teen pregnancy rate in the industrialized world. The Center for Disease control says that as of 2011, 33% of girls get pregnant before the age of 20. We are wondering if the teen pregnancy rate this year is even higher than 33%. A random sample of 400 girls is taken. Of the 400 girls randomly selected, 144 of them were pregnant before the age of 20. Use the following Statcato printout and a 5% significance level to test the claim that the teen pregnancy rate is higher than 33%.

N	Sample Proportion	Significance Level	Critical Value	Test Statistic Z	p-Value
400	0.36	0.05	1.645	1.276	0.1010



13. According to the National Association of College Stores, digital textbooks are projected to account for approximately 13% of course materials sold by the fall of 2012. In a random sample of 260 college course materials, 39 (or 15%) of the sample course materials were digital. Use the random sample data, a 10% significance level, and StatKey to test the claim that 13% of course materials are digital. Under the “Randomization Hypothesis Tests” menu, click on “Test for Single Proportion”. Under the “Edit Data” button put in the random sample data of 39 (count) and 260 (sample size). Change the null hypothesis to 13% (0.13). Determine if the sample proportion fell in the tail and use the sample proportion (0.15) to calculate the P-value. After performing the simulation, make note of the standard error and use the following formula to calculate the test statistic. Write a sentence to explain the test statistic.

$$Z\text{-test statistic} = \frac{\text{Sample Proportion} - \text{Population Proportion}}{\text{Standard Error}}$$

14. An online source suggests that one out of every three people in the U.S have high blood pressure and the population proportion of U.S. adults is 33.3%. Another website disagrees with this and claims that the true percentage of U.S. adults with high blood pressure is actually dramatically lower than 1 in 3 (33.3%). A random sample of 500 U.S. adults found that 165 of them had high blood pressure. Use the Statcato printout below and a 10% significance level to test the claim that less than 33% of U.S. adults have high blood pressure.

N	Sample Proportion	Significance Level	Critical Value	Test Statistic Z	p-Value
500	0.33	0.10	-1.282	-0.142	0.4434

15. Childhood obesity has more than tripled in the past 30 years. The percentage of children aged 6–11 years in the United States who were obese increased from 7% in 1980 to nearly 20% in 2008. If this trend continues, we can expect that the percent of young children that are obese today to be significantly greater than 20%. In order to test this claim, a random sample of 800 children in the U.S. was taken and 179 of them were found to be obese. Use StatKey and a 10% significance level to test the claim that the population proportion of obese children in the U.S. is higher than 0.20. Under the “Randomization Hypothesis Tests” menu, click on “Test for Single Proportion”. Under the “Edit Data” button put in the random sample data of 179 (count) and 800 (sample size). Change the null hypothesis to 20% (0.20). Determine if the sample proportion fell in the tail and use the sample proportion (0.22375) to calculate the P-value. After performing the simulation, make note of the standard error and use the following formula to calculate the Z-test statistic. Write a sentence to explain the test statistic.

$$Z\text{-test statistic} = \frac{\text{Sample Proportion} - \text{Population Proportion}}{\text{Standard Error}}$$

16. Test the claim that exactly 25% of Math 140 statistics students take their class at the COC Canyon Country Campus. A census of the fall 2015 semester indicated that of the 334 statistics students at COC, 111 of them took their class at the Canyon Country campus. Assume the population of interest is all COC statistics students from all semesters. Use the Statcato printout below and a 5% significance level to test the claim.

N	Sample Proportion	Significance Level	Critical Value	Test Statistic Z	p-Value
334	0.332	0.05	-1.96, 1.96	3.475	0.0005

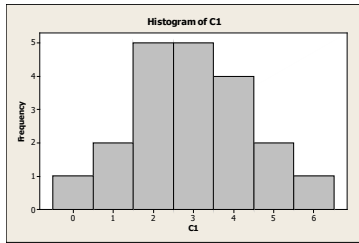
17. Use StatKey and a 5% significance level to test the claim that more than 50% of Math 075 pre-stat students identify as female. A census of the fall 2015 semester indicated that of the 481 pre-stat students at COC, 270 of them identified as female. Assume the population of interest is all COC statistics students from all semesters. Under the “Randomization Hypothesis Tests” menu, click on “Test for Single Proportion”. Under the “Edit Data” button put in the random sample data of 270 (count) and 481 (sample size). Change the null hypothesis to 50% (0.50). Determine if the sample proportion fell in the tail and use the sample proportion (0.5613) to calculate the P-value. After performing the simulation, make note of the standard error and use the following formula to calculate the Z-test statistic. Write a sentence to explain the test statistic.

$$Z\text{-test statistic} = \frac{\text{Sample Proportion} - \text{Population Proportion}}{\text{Standard Error}}$$



18. The manager at a local Starbucks wants to make sure that customers wait less than 4 minutes from the time they order to the time that they pick up their coffee. In order to test this, twenty random customers were selected and the staff measured the number of minute between when the person ordered and when their drink was ready. The sample mean was 2.870 minutes and the sample standard deviation was 1.379 minutes. Here is a histogram of the twenty wait times. Does this data meet the assumptions necessary to perform a hypothesis test? If so, use a 1% significance level to test the claim that the average wait time is less than 4 minutes.

N	Sample Mean	Stdev s	Significance Level	Critical Value	Test Statistic	p-Value
20	2.87	1.379	0.01	-2.540	-3.665	0.0008



19. Redwood trees are the tallest plants on Earth. California is famous for its giant Redwood trees. However, just how tall are they? A random sample of 47 California Redwood trees was taken and their heights measured. (This was not easy by the way.) The sample mean average height was 248 feet with a standard deviation of 26 feet. Does this data meet the assumptions necessary to perform a hypothesis test? If so, use a 5% significance level to test the claim that Redwood trees have an average height greater than 240 feet.

N	Sample Mean	Stdev s	Significance Level	Critical Value	Test Statistic	p-Value
47	248.0	26.0	0.05	1.679	2.109	0.0202

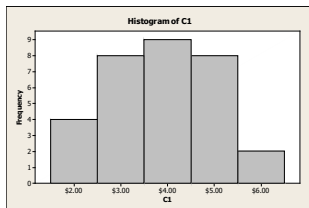
20. Maria is planning to attend UCLA. She is curious what the average age of UCLA students is. Since most students that attend UCLA are in their 20's yet there are also students up to 70 years old, the sample data is skewed right. The college conducted a random sample of 65 students and found that the sample mean was 29.0 years old with a standard deviation of 5.2 years. Does this data meet the assumptions necessary to perform a hypothesis test? If so, use a 10% significance level to test the claim that the average age of students at UCLA is 30 years old.

N	Sample Mean	Stdev s	Significance Level	Critical Value	Test Statistic	p-Value
65	29.0	5.2	0.10	-1.669, 1.669	-1.550	0.1260



21. Mike wants to know the average price of a hamburger. He randomly selected 24 restaurants and recorded the price of a regular hamburger. The sample mean price was \$3.88 and the sample standard deviation was \$1.14. A histogram of the data is below. Does this data set meet the assumptions necessary to perform a hypothesis test? If so, use a 10% significance level to test the claim that the average price of a hamburger is greater than \$3.50?

N	Sample Mean	Stdev s	Significance Level	Critical Value	Test Statistic	p-Value
24	3.88	1.14	0.10	1.320	1.633	0.0580



22. Use StatKey at www.lock5stat.com and the “Math 140 Survey Data Fall 2015” at www.matt-teachout.org to test the claim that the population mean average age of math140 students is higher than 21 years old. Check the assumptions and use a 1% significance level. Under the “Randomization Hypothesis Test” menu, click on “Test for Single Mean”. Under “Edit Data” copy and paste the age data into StatKey and click on “raw data”. Do not click the “identifier” button. Push “OK”. What is the sample mean age of the statistics students? Create the randomized simulation and make a note of the standard error. Use the significance level to determine the tail. Does the sample mean fall in the tail? Use the sample mean and the simulation to calculate the P-value. Finish the hypothesis test. After your test is complete, use the following formula to calculate the T-test statistic. Write a sentence to explain the T-test statistic.

$$\text{T-test statistic} = \frac{\text{Sample Mean} - \text{Population Mean}}{\text{Standard Error}}$$

23. Use StatKey at www.lock5stat.com and the “Math 140 Survey Data Fall 2015” at www.matt-teachout.org to test the claim that the population mean average weight of math140 students is less than 160 pounds. Check the assumptions and use a 5% significance level. Under the “Randomization Hypothesis Test” menu, click on “Test for Single Mean”. Under “Edit Data” copy and paste the weight data into StatKey and click on “raw data”. Do not click the “identifier” button. Push “OK”. What is the sample mean weight of the statistics students? Create the randomized simulation and make a note of the standard error. Use the significance level to determine the tail. Does the sample mean fall in the tail? Use the sample mean and the simulation to calculate the P-value. Finish the hypothesis test. After your test is complete, use the following formula to calculate the T-test statistic. Write a sentence to explain the T-test statistic.

$$\text{T-test statistic} = \frac{\text{Sample Mean} - \text{Population Mean}}{\text{Standard Error}}$$

24. Test the claim that the population mean average height of math140 students is 64 inches. Assume the data met the assumptions. Use a 10% significance level.

Use StatKey at www.lock5stat.com and the “Math 140 Survey Data Fall 2015” at www.matt-teachout.org to test the claim that the population mean average height of math140 students is 64 inches. Check the assumptions and use a 10% significance level. Under the “Randomization Hypothesis Test” menu, click on “Test for Single Mean”. Under “Edit Data” copy and paste the height data into StatKey and click on “raw data”. Do not click the “identifier” button. Push “OK”. What is the sample mean height of the statistics students? Create the randomized simulation and make a note of the standard error. Use the significance level to determine the tail. Does the sample mean fall in the tail? Use the sample mean and the simulation to calculate the P-value. Finish the hypothesis test. After your test is complete, use the following formula to calculate the T-test statistic. Write a sentence to explain the T-test statistic.

$$\text{T-test statistic} = \frac{\text{Sample Mean} - \text{Population Mean}}{\text{Standard Error}}$$



Chapter 3 Review Sheet

Topics:

- Use population claims to construct the null and alternative hypothesis. Also, know how to determine the type of tail for the test.
- Know the assumptions (conditions) necessary to do a one-population mean or proportion hypothesis test.
- Be able to explain the meaning of the Z-test statistic for one-population proportion hypothesis tests.
- Be able to explain the meaning of the T-test statistic for one-population mean average hypothesis tests.
- Be able to determine how likely it is for the sample data to have occurred by sampling variability (random chance).
- Know how to interpret significance by judging whether the test statistic falls in the tail corresponding to the critical value.
- Know how to interpret significance by judging whether the sample statistic falls in the tail corresponding to the significance level in a randomized simulation.
- Know how to use “Theoretical Distributions” menu in StatKey to use the significance level to look up critical values.
- Know how to use “Theoretical Distributions” menu in StatKey to use the test statistic to look up the P-value.
- Know how to create randomized simulations and use the significance level to judge the tails.
- Know how to create randomized simulations with StatKey and use the sample statistic to calculate the P-value.
- Know how to use the P-value and significance level to determine if the sample data could have happened because of sampling variability or if it was unlikely.
- Know the definition of P-value.
- Know how to use the P-value and significance level to determine if we should reject the null hypothesis or fail to reject the null hypothesis.
- Be able to write the formal conclusion for a hypothesis test and explain its meaning.
- Know the definitions and terms associated with Type 1 and Type 2 errors. Know how to decrease the chances of having a Type 1 or Type 2 error.

Chapter 3 Vocabulary Terms and Definitions

Population: The collection of all people or objects to be studied.

Sample: Collecting data from a small subgroup of the population.

Random Sample: Sample data collected in such a way that everyone in the population has an equal chance to be included.

Statistic: A number calculated from sample data in order to understand the characteristics of the data.
For example, a sample mean average, a sample standard deviation, or a sample percentage.

Parameter: A number that describes the characteristics of a population like a population mean or a population percentage. Can be calculated from an unbiased census, but is often just a guess about the population.

Hypothesis Test: A procedure for testing a claim about a population.

Null Hypothesis (H_0): A statement about the population that involves equality. It is often a statement about “no change”, “no relationship” or “no effect”.

Alternative Hypothesis (H_A or H_1): A statement about the population that does not involve equality. It is often a statement about a “significant difference”, “significant change”, “relationship” or “effect”.

Population Claim: What someone thinks is true about a population.



Test Statistic: A number calculated in order to determine if the sample data significantly disagrees with the null hypothesis. There are a variety of different test statistics depending on the type of data.

One-Population Proportion Test Statistic (z): The sample proportion is this many standard errors above or below the population proportion in the null hypothesis.

One-Population Mean Test Statistic (t): The sample mean is this many standard errors above or below the population mean in the null hypothesis.

Critical Value: A number we compare our test statistic to in order to determine significance. In a sampling distribution or a theoretical distribution approximating the sampling distribution, the critical value shows us where the tail or tails are. The test statistic must fall in the tail to be significant.

Sampling Variability: Also called “random chance”. The principle that random samples from the same population will usually be different and give very different statistics. The random samples will usually be different than the population parameter.

P-value: The probability of getting the sample data or more extreme because of sampling variability (by random chance) if the null hypothesis is true.

Significance Level (α): Also called the Alpha Level. This is the probability of making a type 1 error. The P-value is compared to this number to determine significance and sampling variability. If the P-value is lower than the significance level, then the sample data significantly disagrees with the null hypothesis and is unlikely to have happened because of sampling variability.

Randomized Simulation: A technique for visualizing sampling variability in a hypothesis test. The computer assumes the null hypothesis is true, and then generates random samples. If the sample data or test statistic falls in the tail, then the sample data significantly disagrees with the null hypothesis. This technique can also calculate the P-value and standard error without a formula.

Type 1 Error: When biased sample data leads you to support the alternative hypothesis when the alternative hypothesis is actually wrong in the population.

Type 2 Error: When biased sample data leads you fail to reject the null hypothesis when the null hypothesis is actually wrong in the population.

Beta Level (β): The probability of making a type 2 error.

Conclusion: A final statement in a hypothesis test that addresses the claim and evidence.



Chapter 3 Review Notes

Steps to Writing the Null and Alternative Hypothesis

1. **Write down the population claim in symbolic notation.**
(Note: This can be the null or the alternative hypothesis.)
(Note: This is the statement the scientist thinks is true.)
(Note: In one-population tests, always write the parameter on the left and the number on the right.)
2. **Write down the opposite of the population claim.**
(Note: This can be the null or the alternative hypothesis.)
(Note: This is the statement that would be true if the scientist is wrong.)
(Note: In one-population tests, always write the parameter on the left and the number on the right.)
3. **The statement with equality ($=, \leq, \geq$) is always the null hypothesis (H_0).**
(Note: The null hypothesis is usually " $=$ ". Statements with " \leq " or " \geq " are often changed to " $=$ ".)
4. **The statement that does NOT have equality ($\neq, <, >$) is always the alternative hypothesis (H_A).**

Which Tail should we use in a one-population Hypothesis Test?

- **Right-Tailed Test:** Alternative Hypothesis (H_A) is **GREATER THAN ($>$)**
- **Left-Tailed Test:** Alternative Hypothesis (H_A) is **LESS THAN ($<$)**
- **Two-Tailed Test:** Alternative Hypothesis (H_A) is **NOT EQUAL (\neq)**

One-Population Hypothesis Test Assumptions (Conditions)

Assumptions for One-Population Mean Hypothesis Test

- Random Sample Data
- Individuals Independent
- Sample data normal or sample size at least 30

Assumptions for One-Population Proportion (%) Hypothesis Test

- Random Sample Data
- Individuals Independent
- Sample data has at least 10 successes and at least 10 failures.

Assumptions for One-Population Randomized Simulation Hypothesis Test

- Random Sample Data
- Individuals Independent



Summary Table of P-Value, Test Statistic, Simulation, Significance, Sampling Variability and Evidence

	SIGNIFICANT	NOT SIGNIFICANT
	Test Statistic falls in tail determined by a critical value	Test Statistic does NOT fall in tail determined by a critical value
	OR	OR
	Low P-value ($P\text{-value} \leq \text{significance level}$)	High P-value ($P\text{-value} > \text{significance level}$)
	OR	OR
	Sample Statistic falls in tail determined by the significance level of randomized simulation.	Sample Statistic does NOT fall in tail determined by the significance level of randomized simulation.
Is the sample data significantly different than H_0 ?	Significantly Different	NOT Significantly Different
Could the sample data happen because of sampling variability (random chance) if H_0 is true?	Unlikely	Could Happen
Reject H_0 or Fail to Reject H_0 ?	Reject H_0	Fail to Reject H_0
Is there significant evidence?	Significant Evidence	NOT Significant Evidence



Summary Table: Conclusions

	Claim is H_0	Claim is H_A
Low P-value <i>(P-value \leq significance level)</i>	There is significant evidence to reject the claim. <i>(Evidence indicates that the claim may be wrong.)</i>	There is significant evidence to support the claim. <i>(Evidence indicates that the claim may be correct.)</i>
High P-value <i>(P-value $>$ significance level)</i>	There is NOT significant evidence to reject the claim. <i>(Random sample data does not disagree with the claim. The claim could be true, but we do not have evidence.)</i>	There is NOT significant evidence to support the claim. <i>(Random sample data disagrees with the claim. The claim could be wrong, but we do not have evidence.)</i>

Chapter 3 Review Practice Problems

- Write a definition for the following key terms.
 - hypothesis test
 - Null hypothesis
 - Alternative Hypothesis
 - Population Claim
 - Test statistic
 - one-population proportion Z test statistic
 - one-population mean T test statistic
 - Critical Value
 - Sampling Variability (Random Chance)
 - P-value
 - significance level (alpha level)
 - Randomized Simulation
 - beta level
 - type 1 error
 - type 2 error
 - Conclusion
- How is randomized simulation used in hypothesis testing and describe what it can tell us.
- How can we know if the sample data significantly disagrees with the null hypothesis?
- How can we determine how likely it is for the sample data to have occurred because of sampling variability?
- How do we know if we reject the null hypothesis or fail to reject the null hypothesis?
- What are the four steps to writing conclusions?
- What assumptions do we need to check for the following:
 - When testing a hypothesis about a one-population mean average?
 - When testing a hypothesis about a one-population proportion (percentage)?
 - When testing a hypothesis about one-population using randomized simulation?



8. Fill out the following table regarding test statistics and critical values.

Test Statistic	Critical Value	Does sample significantly disagree with H_0 or not?
T = +1.774	± 2.751	
Z = -2.481	-1.96	
T = -3.394	± 2.566	
Z = +1.362	+1.645	

9. Fill out the following table regarding P-value and Significance levels.

P-value	P-value %	Significance Level	Sampling Variability or Unlikely	Reject H_0 or Fail to reject H_0 ?
0.0002		5%		
0.3327		1%		
1.84×10^{-5}		10%		
0.0941		5%		

10. Fill out the following table to practice writing conclusions.

P-value	Claim	Write the Conclusion addressing Evidence and claim
Low	H_0	
High	H_A	
High	H_0	
Low	H_A	

11. Write the null and alternative hypotheses for the following. Which is the claim? Is this a left-tailed, right-tailed, or two-tailed test?

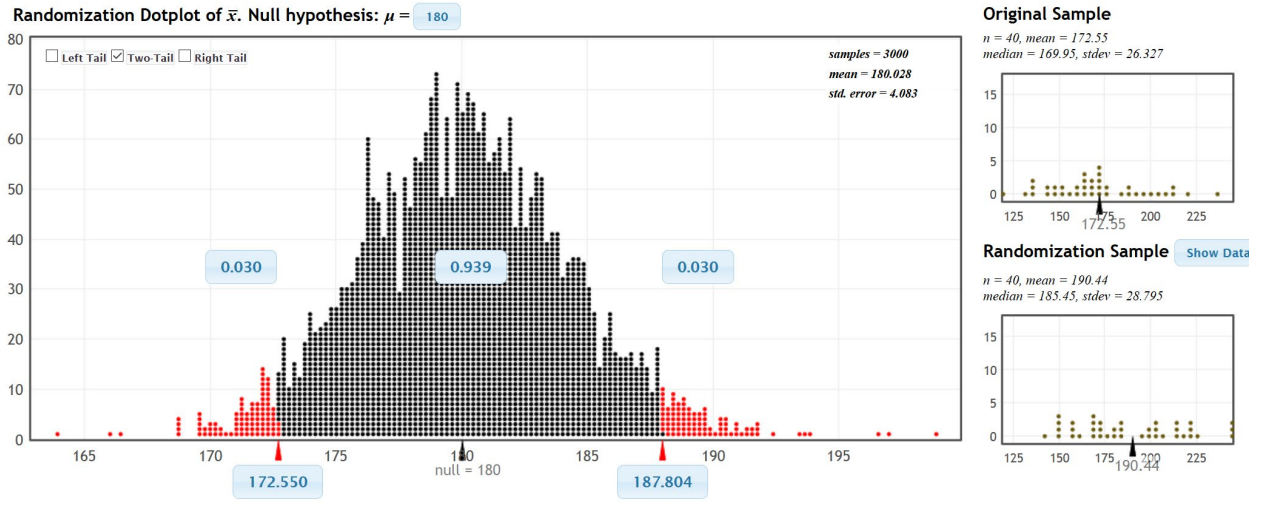
- "We used to think that the population mean average normal body temperature for all humans was exactly 98.6 degrees Fahrenheit. Now we think it is lower."
- "The percentage of all people in China with Typhoid (π_1) used to be higher than the percentage of all people in India with Typhoid (π_2). Now we think the percentage is about the same."
- "The population mean average age of students at UCLA (μ_1) is the same as the population mean average age of students at USC (μ_2)."

12. Answer the following questions about type 1 and type 2 errors.

- What is a type 1 error?
- What is a type 2 error?
- What is the probability of type 1 error called?
- What is the probability of type 2 error called?
- Why do type 1 and type 2 errors occur?
- What can we do to limit the chances of a type 1 error?
- What can we do to limit the chances of a type 2 error?
- Which significance level is best for keeping both type 1 and type 2 errors low?
- If the significance level is 1%, what happens to the probability of type 1 and type 2 errors?
- If the significance level is 10%, what happens to the probability of type 1 and type 2 errors?



13. An article states that the mean average weight of men in the U.S. is 180 pounds. The random health data at teachoutoc.org was pasted into StatKey and the following randomized simulation was created in order to test the article's claim. Use a 1% significance level. Make sure to check the assumptions. Give the null and alternative hypothesis, estimate the P-value, and the conclusion. Write a sentence to explain the P-value. Was there a significant difference between the sample mean and the population mean? How likely was it that the sample data occurred by random chance if the population mean really is 180 pounds?



14. Use the Statcato printout below to test the following claim. A speaker at a nursing convention said that the population mean average hourly salary of a registered nurse used to be \$25 per hour, but now is greater than \$25 per hour. A random sample of 28 nurses gave a sample mean of \$26.82 and a standard deviation of \$4.37. A histogram of the salaries showed a bell shaped distribution. Use a 5% significance level. Make sure to check the assumptions. Give the null and alternative hypothesis, test statistic, P-value, and the conclusion. Write a sentence to explain the test statistic. Write a sentence to explain the P-value. Was there a significant difference between the sample mean and the population mean? How likely was it that the sample data occurred by random chance if the population mean really is \$25?

N	Sample Mean	Stdev s	Significance Level	Critical Value	Test Statistic	p-Value
28	26.82	4.37	0.05	1.703	2.204	0.0181



15. Use simulation on www.lock5stat.com to test the following claim. The Harris Poll conducted a random survey in which they asked 1097 women "How many tattoos do you currently have?" Of the 1,097 females surveyed, 143 responded that they had at least one tattoo. A tattoo magazine claimed that more than 10% of women have at least one tattoo. Use a 5% significance level to test the magazine's claim. Give the null and alternative hypothesis, the estimated P-value, and the conclusion. Write a sentence to explain the P-value. Was there a significant difference between the sample percent and the population value? How likely was it that the sample data occurred by random chance from a 10% population.

