

Tests of Hypothesis: One Sample

8.1 Determining the Claim, Null and Alternative Hypotheses

In this chapter, we will begin to learn how to test a hypothesis. To demonstrate how this is done, we will need a hypothesis to test. We will form our hypotheses from a claim like the one below.

Example 117: The Federal Aviation Administration claims that the mean weight of an airline passenger (with carry-on baggage) is greater than the 185 lbs that it was 20 years ago, express this claim symbolically.



This claim can be reworked symbolically to form two competing hypotheses. The first hypothesis we will form is called the **null hypothesis** which usually expresses the status quo scenario. It is denoted by H_0 (read "H sub zero"). In a hypothesis test, we start out by assuming H_0 is true, and it is always H_0 that we are testing during the test. Finally, H_0 must always have an equal sign ($\leq, \geq, \text{ or } =$).

Null Hypothesis: H_0

- This will be supported unless the data provide evidence that it is false
- The *status quo*

The second hypothesis we will form from the claim is the **alternative hypothesis**—alternative because it is the alternative to the null hypothesis. It is also often referred to as the *research hypothesis*. It is denoted by H_A . H_A always has one of the symbols: $<$, $>$, or \neq . The H_A will determine the kind of test we conduct on the null hypothesis (more on that later).

Alternative Hypothesis: H_a

- This will be supported if the data provide sufficient evidence that it is true
- The *research hypothesis*

*note some texts use H_1 instead of H_A to denote the alternative hypothesis.

Example 118: Using the claim from the FAA above, create our set of competing hypotheses H_0 and H_a :

Solution: $H_0 : \mu \leq 185$ and $H_A : \mu > 185$

The way we will test this claim is to collect a sample, assume that the null hypothesis is true, and then determine if our sample supplies evidence that indicates that the null is not true.

Suppose our sample results seem very extreme (under the assumption that the null hypothesis is true), we would then be left with two possible explanations:

- 1) The null is really correct, and we just happened to select an unusual sample.
- 2) The sample isn't really that unusual. It just seems unusual because we are assuming the null hypothesis is correct.

The whole process relies on the idea that in order for us to reject the null hypothesis, we need to observe a sample that would only show up very rarely just by chance when the null is true.

But how do we know how rare our sample results are? The only way to know how rare the sample is would be for us to use a summary statistic to extract the information contained in the sample. After summarizing the information contained in the sample into a statistic, we will need to know the probability distribution for the statistic formed. This statistic will be called the **Test Statistic**. The **Test Statistic** will be a z-value that measures the distance as a number of standard errors between the value of \bar{X} and the mean μ specified under the null hypothesis.

Example 119: Suppose that after taking a random sample of airline passengers we have the following data: $n = 36$, $\bar{X} = 200\text{lbs}$, and $\sigma = 33.3$ (known from previous studies). Use the CLT, the null hypothesis from above, and the data above to create a test statistic that has a Z-distribution.



8.2 Critical Values for the Rejection Region

Since we know the distribution of the normal random variable z , we can determine how unusual our test statistic is under the assumption that the null hypothesis is correct. In fact, we might decide that if the chance of getting a statistic as extreme, or even more extreme, is less than some predetermined value, we will conclude that the null can be rejected. That predetermined value will be called the **critical value**. The **critical value** will be the boundary point between the **rejection region** and the rest of the number line. The **rejection region** refers to the values of the test statistic for which we will decide to reject the null hypothesis.

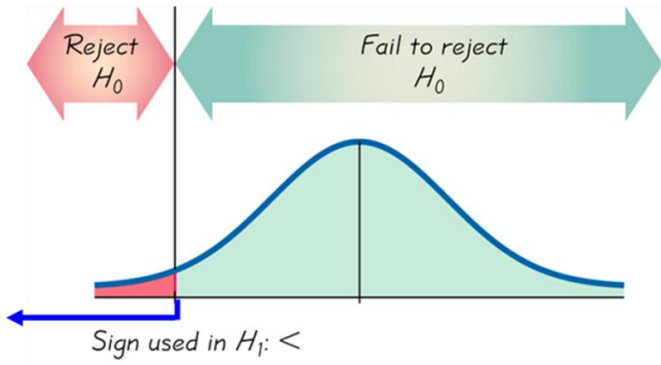
- If the *test statistic* has a high probability when H_0 is true, then H_0 is *not rejected*.
- If the *test statistic* has a (very) low probability when H_0 is true, then H_0 is *rejected*.

When testing a hypothesis we will need to know what kind of extreme test statistic would make us question the validity of our null hypothesis. In other words, will a test statistic that is abnormally large indicate the null might be false? Or perhaps, an abnormally small test statistic will indicate a false null? Or will a test statistic that is either too small or too large make us doubt the null? Abnormally small or large values in a distribution will fall into the tails of the distribution.

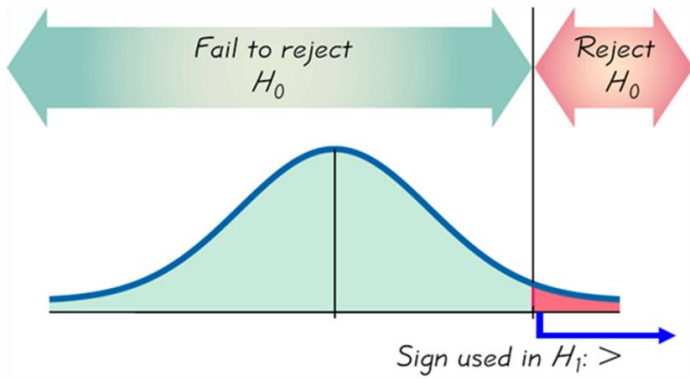
The critical value(s) will define the rejection region(s) of the curve. These regions will be located in the tails of the distribution. **The alternative hypothesis will indicate where our critical region should be located: the left tail, the right tail, or in both tails.**

Determining the Number of Tails in a Hypothesis Test:

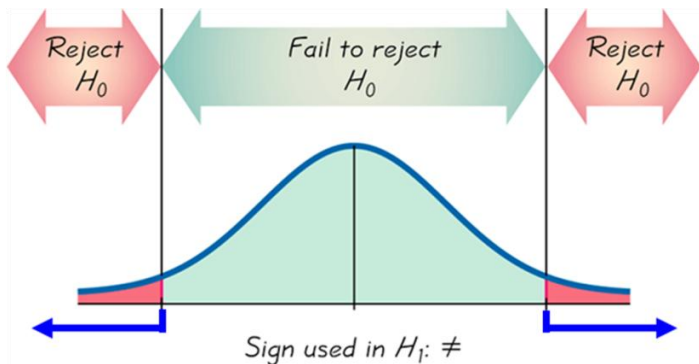
Alternative Hypothesis	Tails
$\mu < \mu_0$	Left-tailed
$\mu > \mu_0$	Right-tailed
$\mu \neq \mu_0$	Two-tailed



Notice the $<$ symbol is like the tip of an arrow pointing to the left.

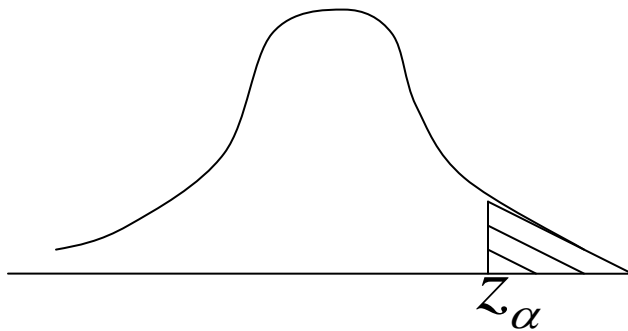


Notice the $>$ symbol is like the tip of an arrow pointing to the right.



Before we can determine the critical value for a hypothesis test, we need to decide the maximum amount of error we are willing to allow assuming the null hypothesis is true.

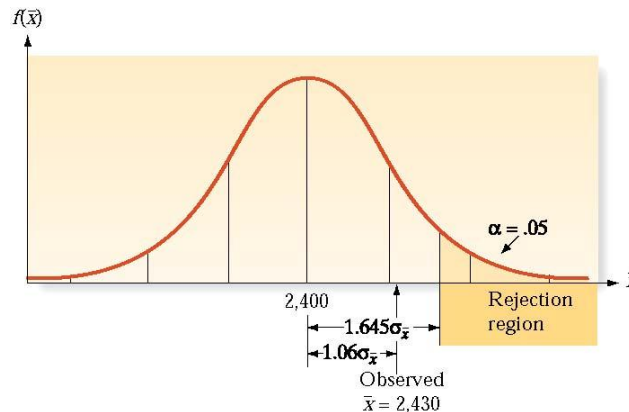
To understand this, consider the hypothesis we have been using in our examples above. What if in reality the true weight of airline passengers (including carry-on luggage) was 185 pounds? Is it impossible for us to have unusual samples even given that the null is true? Of course, it is possible to have extreme or unusual samples, but they would be rare occurrences. Using our z-chart, we can select the critical value such that when the null hypothesis is true our test statistic has only a very small chance of being more extreme than the chosen critical value. The **significance level** is a limit on the probability of committing the error of rejecting a true null hypothesis. It is denoted as α . This *critical value* might remind you of:



*Recall the area in the tail here is equal to alpha.

Conclusions	Reality	
	The Null Is True	The Null Isn't True
We Reject the Null	Type I Error	Correct Decision
We Do Not Reject Null	Correct Decision	Type II Error

In the diagram below, the researchers were testing a claim that said $\mu > 2400$, and they had sample data that produced a sample mean of $\bar{x} = 2430$, which was not far enough to the right on the number line to put it in the rejection region.



Example 120: Find the critical value for a test of the above FAA hypothesis ($H_A : \mu > 185$) at the 1% significance level.

If the test statistic we created earlier is greater than the above found critical value we will reject the null hypothesis.

Note: Null hypotheses are either rejected, or else there is insufficient evidence to reject them. (I.e., we don't *accept* null hypotheses.)

Example 121: Assume that the data have a normal distribution and the number of observations is greater than 50. Using $\alpha = 0.05$ for a left-tailed test, find the critical z value used to test the null hypothesis.

Solution: $z = -1.645$

8.3 Large-Sample Test of Hypothesis about a Population Mean

Let us summarize what we have done with the FAA example above:

Claim: The Federal Aviation Administration claims that the mean weight of an airline passenger (with carry-on baggage) is greater than the 185 lbs it was 20 years ago.

Hypotheses: $H_0 : \mu \leq 185$
 $H_A : \mu > 185$

Sample Data: $n = 36$, $\bar{X} = 200\text{lbs}$, and $\sigma = 33.3$ (known from previous studies).

Test Statistic: $z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = 2.703$

Significance Level: $\alpha = 0.01$

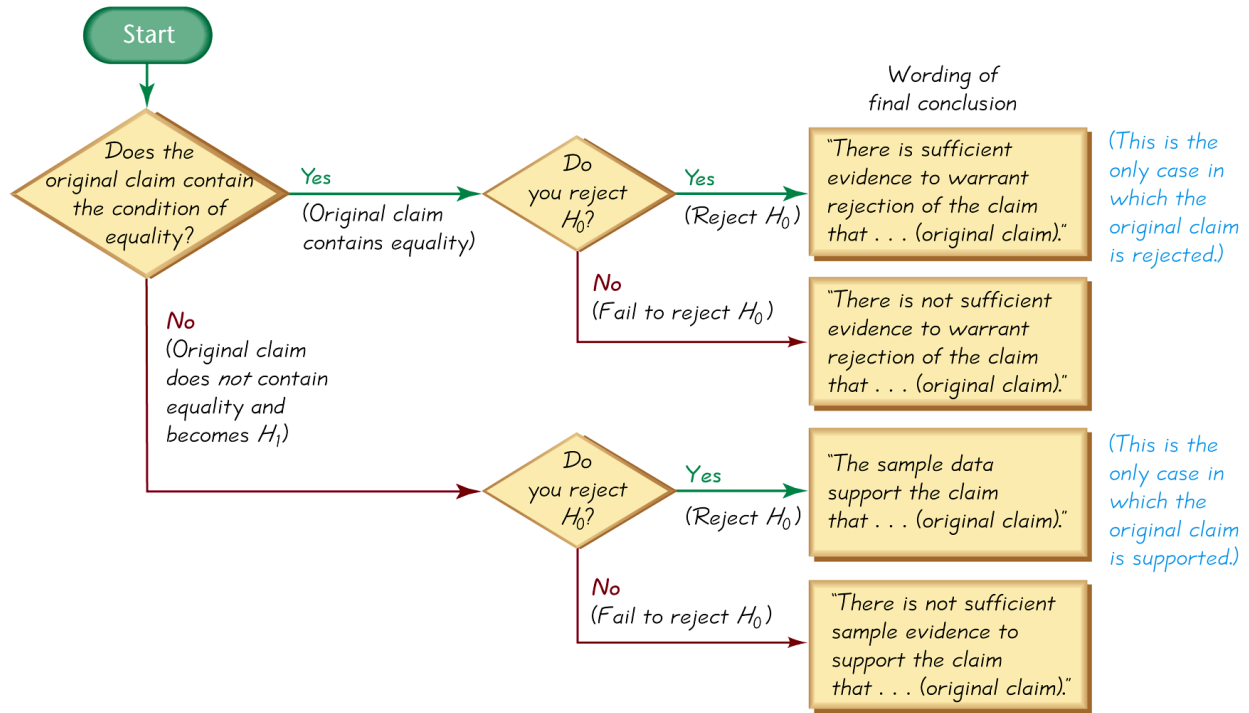
Type of Test: Right-tailed, $H_A : \mu > 185$

Critical Value: 2.326

Once we have our test statistic and we have determined our rejection region all that is left to do is to compare our test statistic to our critical value(s). If the test stat is more extreme (i.e.-farther away from the mean on the same side of the curve as the critical value(s)) we reject the null Hypothesis. This step is called the initial conclusion step.

Initial Conclusion: Since the test stat, $z = 2.703$, is greater than the critical value, $z = 2.326$, we will reject the null hypothesis.

It is important to word our final conclusion carefully. We want to make sure that we address the original claim, and we want to make sure that we do not say more than the evidence grants us to say. To learn how to word our conclusions properly look at the flow chart provided with the formula card on the web site (it has been reproduced below).



Finally, let's look at the four possible outcomes (in blue below) for our hypothesis test:

Conclusions	Reality	
	The Null Is True	The Null Isn't True
We Reject the Null	Type I Error	Correct Decision
We Do Not Reject Null	Correct Decision	Type II Error

- ❖ A Type I error is the mistake of rejecting the null hypothesis when it is true.
- ❖ A Type II error is the mistake of failing to reject the null hypothesis when it is false.

In hypothesis testing, we want to make sure the worst of the two possible errors is the type I error. The reason for this is that we design the test to control the probability of the type one error. The below table explains how the significance level is related to the type one error:

Probability of the type one error	
For a left-tailed test	At most α
For a right-tailed test	At most α
For a two-tailed test	Exactly equal to α

Note: We have notation for the probability of a type two error, $\beta = P(\text{Type II error})$.

To reduce the error rate for a type one error we lower the significance level, alpha (α), but this will increase the likelihood of committing a type II error.

To reduce the error rate for both a type one error and a type two error, we need to increase the sample size and lower the significance level (α)*.

To understand the two statements above consider a criminal trial, if a country decides to convict any person who ends up in court regardless of the evidence, they will not let any guilty people go free who end up in court, but as a result a lot of innocent people will end up in jail. If you decide to let people off the hook for a crime unless the evidence is overwhelmingly against them, you will end up letting many guilty people go free. This tug of war always exists. We cannot reduce both kinds of errors at once, guarding against one will produce more of the other, unless we can find more evidence. For us, this would mean taking a larger sample size.

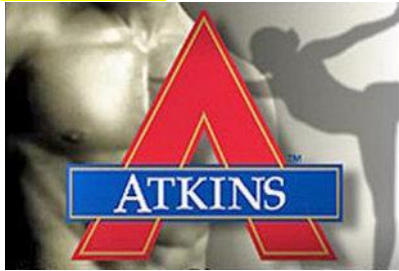
*When conducting a two-tailed test or when $\mu = \mu_0$ during a one-tailed test, increasing the sample size alone will not reduce the likelihood of a type I error, but it will reduce the likelihood of a type II error. That is why, we must also lower the significance level to be certain that we will lower the type I error too.

To summarize, the following set up steps can be used to conduct a test of hypothesis:

Steps to test a hypothesis:

1. Express the original claim symbolically
2. Identify the null and alternative hypothesis
3. Record the data from the problem
4. Calculate the test statistic
5. Determine your rejection region
6. Find the initial conclusion (reject the null hypothesis (with possible Type I error) or do not reject it (with possible Type II error))
7. Word your final conclusion

Example 122: In a 2007 study of different popular diets, 77 people used a modified* version of the Atkins Diet for one year. Their mean weight change was -10.34 lbs. Assume that the population standard deviation for all such weight changes is known to be 15.51 pounds. Use a significance level of 0.05 to test the claim that the mean weight change is equal to zero. Does the diet seem to be effective? Does the mean weight change seem substantial enough to justify the diet? What assumptions are necessary for the test we just conducted to be valid?



*The Atkins diet was modified to include higher levels of fiber and high quality complex carbohydrate.

Example 123: A sociologist claims that the average Hispanic teen male is engaging in intercourse before turning 17 years old. A 2008 study of 47 Hispanic males revealed that the mean age at which they had intercourse for the first time was 16.31 years old with a standard deviation of 1.78 years. Use a 0.01 significance level to test the sociologist's claim that the mean age for Hispanic males to first engage in intercourse is less than 17 years old.



The following assumptions need to be upheld in order for the results of the above tests to be valid:

- The sample was selected randomly
- σ known & normally distributed
or
- σ known & $n > 30$

8.4 Observed Significance Levels: p-Values

P-Value Method

The **Observed Significance Level** or **P-Value**, for a specific statistical test is the probability (assuming the null is true) of observing a value of the test statistic that is at least as extreme as the test stat computed from your sample data.

Recall that, we always test the null hypothesis, so the initial conclusion will always be one of the following:

1. Reject the null hypothesis.
2. Fail to reject the null hypothesis.

In the **Traditional method** of hypothesis testing, we:

Reject H_0 if the test statistic falls within the critical region.

Fail to reject H_0 if the test statistic does not fall within the critical region.

In this section we will learn the **P-value method**:

Reject H_0 if the P -value $< \alpha$ (where α is the significance level, such as 0.05).

Fail to reject H_0 if the P -value $> \alpha$.

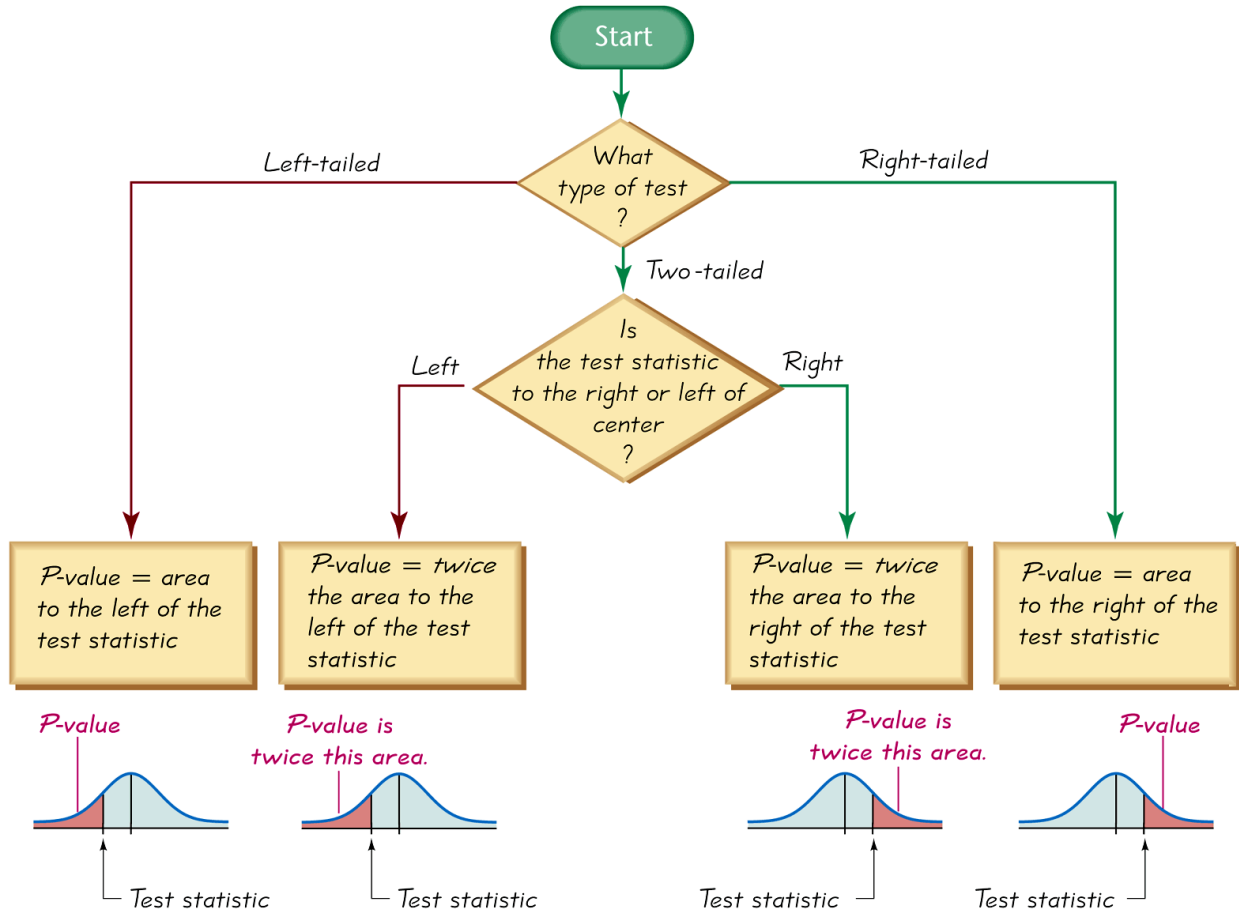
Example 124: In a study of the effects of prenatal cocaine use on infants, the following sample data on birth weights was obtained: $n = 36$, $\bar{X} = 2800$, $\sigma = 645$. Using a significance level, $\alpha = 0.01$, test the claim that the mean birth weight for children of cocaine users is less than 3103 grams, the mean weight for children who had mothers who did not use cocaine.



Claim: $\mu < 3103$ (mean for children who had mothers who did not use cocaine)

1. Calculate the test statistic.
2. Find the p-value

To help you with part two of the example question above you may bring the p-value flowchart contained in the website's formula card. It is recreated below:



Finding a p-value

First place the test statistic on the curve, then calculate the appropriate area according to the following:

For a left-tailed test, the p-value is the area to the left of the test statistic.

For a right-tailed test, the p-value is the area to the right of the test statistic.

For a two-tailed test, the p-value is **twice** the tail area beyond the test statistic.

To use the p-value to test a hypothesis, we simply need to compare it to our stated significance level.

If $p < \alpha$, we reject the null hypothesis

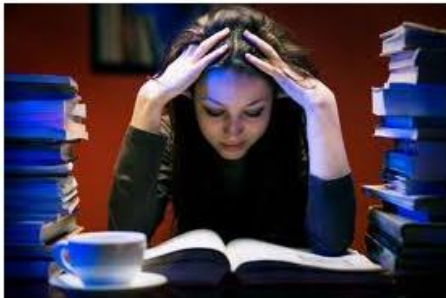
If $p > \alpha$, we do not reject the null hypothesis

- note we should not encounter too many situations where the p-value is equal to the significance level; however, if it happens it is up to the statistician to decide if the evidence warrants rejection of the null or not.

Example 125: A researcher predicts that a low carbohydrate diet will result in a loss of lean muscle mass of 3.5 pounds of muscle per ten pounds of overall weight loss. A recent study that looked at the effects of restricting carbohydrate intake on weight loss involved reducing total calorie intake by 600 calories per day while following a diet that had the following macro nutrient ratios: 38:30:32 (percent of carbohydrate to protein to fat). Thirty-two overweight men followed the diet for 16 weeks. The average loss of lean mass (for every ten pounds of overall weight loss) was 3.3 pounds. The standard deviation for the amount of lean mass lost per ten pounds of weight loss was 5.1 pounds. Use the p-value method to test the researcher’s prediction at the 2% significance level.



Example 126: In 1980, the average time to complete a “four-year” degree was 4.9 years. In 2006, a study of 31 randomly selected students had an average completion time for their four-year degree of 5.3 years with a population standard deviation of 1 year. Use the p-value method to test the claim that the mean time to complete a four-year degree is now more than 4.9 years.



8.5 Small-Sample Test of Hypothesis about a Population Mean

The t-test If you recall the situation we faced when constructing confidence intervals when the population standard deviation was unknown, it will come as no surprise to you that when testing hypotheses without knowledge of the population standard deviation we will need to use the t-distribution.

Method	Conditions
Z-distribution	σ known & normally distributed or σ known & $n > 30$
t-distribution	σ not known & normally distributed or σ not known & $n > 30$
Nonparametric	Population is not normally distributed and $n \leq 30$

Aside from the change from Z-distribution to t-distribution, the problems in this section are the same. The only step that changes is step 5 below. This change is minor because we will use the t-table provided on the web.

Steps to test a hypothesis:

1. Express the original claim symbolically
2. Identify the null and alternative hypothesis
3. Record the data from the problem
4. Calculate the test statistic
5. Determine your rejection region (don't forget to use degrees of freedom)
6. Find the initial conclusion
7. Word your final conclusion

Example 127: The Windsor bottling company received complaints that their 12oz root beer bottles contained less than 12 ounces in them. When 24 bottles are randomly selected and measured the amounts had a mean of 11.4 ounces and a standard deviation of 0.62 ounces. Test the claim that consumers are being cheated. If the company says the sample is too small for the results to be meaningful, is that reasoning valid here?



Example 128: In previous tests, baseballs were dropped 24ft onto a concrete surface, and they bounced an average of 92.84 inches. In a test of a sample of 28 new balls, the bounce heights had a mean of 92.67 inches and a standard deviation of 1.79 inches. Use a 5% level of significance to test the claim that the new balls have a mean bounce height different from 92.84.



8.6 Hypothesis about a Population Proportion

Testing a Claim about a Proportion

Many of the most interesting problems arise when we are looking at survey data. Usually the data generated from surveys consists of proportions. This means the distribution of the sample proportion will be important to us. The distribution of the proportion used in many studies is Binomial in nature; however, we can approximate the distribution of the sample proportion with the Normal distribution if the sample size is sufficiently large*.

*Note: Sufficiently large means that the interval $p_0 \pm 3\sqrt{\frac{p_0q_0}{n}}$ is entirely contained within $[0,1]$.

The main change in our steps to testing a hypothesis will be our test statistic:

Test Stat for Testing Claims About a Proportion:

$$z = \frac{\text{Sample Proportion} - \text{Null Hypothesized Proportion}}{\text{Standard Error for Sample Proportion}} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0q_0}{n}}}$$

Another change will be the parameter used in our hypotheses. It will be rho, the Greek symbol for the population proportion. For example, $H_0 : \rho = 0.56$ is the symbolic form of the claim that the population proportion is equal to 56%. Other than the two changes mentioned above, the seven steps given in earlier sections will work on these problems as well.

Steps to test a hypothesis:

1. Express the original claim symbolically
2. Identify the null and alternative hypothesis
3. Record the data from the problem
4. Calculate the test statistic
5. Determine your rejection region
6. Find the initial conclusion
7. Word your final conclusion

Example 129: An economist claims that less than two-thirds of married women spent over \$1000 on their wedding gown. Glamour magazine sponsored a survey of 2500 prospective brides and found that 65% of them spent more than \$1,000 dollars on their wedding gown. Use a 0.01 significance level to test the claim that less than two-thirds of married women spent over \$1000 on their wedding gown. If these results were obtained from internet users who voluntarily went to the web to answer the survey, does that affect the result of the survey in any way?



Example 130: An article distributed by the Associated Press included these results from a nationwide survey: Of 880 randomly selected drivers, 56% admitted that they run red lights. Test the claim that the majority of all Americans run red lights.



8.7 Type I and Type II Error Probabilities

The type I and type II errors were covered in section 8.3, but since these ideas are quite important to the topic of hypothesis testing, we will look at some key ideas again.

- ❖ **A Type I error is the mistake of rejecting the null hypothesis when it is true.**
- ❖ **A Type II error is the mistake of failing to reject the null hypothesis when it is false.**

Probability of the type one error	
For a left-tailed test	At most α
For a right-tailed test	At most α
For a two-tailed test	Exactly equal to α

To reduce the error rate for a type one error we lower the significance level, alpha (α), but this will increase the likelihood of committing a type II error.

To reduce the error rate for both a type one error and a type two error, we need to increase the sample size and lower the significance level (α).*

*When conducting a two-tailed test or when $\mu = \mu_0$ during a one-tailed test, increasing the sample size alone will not reduce the likelihood of a type I error, but it will reduce the likelihood of a type II error. That is why, we must also lower the significance level to be certain that we will lower the type I error too.

Example 130.5: A researcher wants to test the claim that the mean happiness score for married couples is greater than 2.5 on a three point scale (1 = *not very happy*, 2 = *pretty happy*, 3 = *very happy*). The significance level for the test is 0.01.

- a. What is the probability of the type one error?
- b. If the p-value for the test turned out to be 0.0004, what is the initial conclusion? What possible error (type I or II) could have been committed after forming that conclusion?
- c. If the p-value for the test turned out to be 0.0301, what is the initial conclusion? What possible error (type I or II) could have been committed after forming that conclusion?

