CHAPTER **12** CHAPTER

Hypothesis Testing

Chapter Outline

12.1 Hypothesis Testing

Learning Objectives

- Develop null and alternative hypotheses to test for a given situation.
- Understand the difference between one- and two-tailed hypothesis tests.
- Understand Type I and Type II errors

Introduction

In everyday life, we often have to make decisions based on incomplete information. These may be decisions that are important to us such as, "Will I improve my biology grades if I spend more time studying vocabulary?" or "Should I become a chemistry major to increase my chances of getting into med school?" This section is about the use of hypothesis testing to help us with these decisions. **Hypothesis testing** is a kind of statistical inference that involves asking a question, collecting data, and then examining what the data tells us about how to procede.

In a formal hypothesis test, hypotheses are always statements about the population. The hypothesis tests we will examine in this chapter involve statements about the average values (means) of some variable in the population. For example, we may want to know if the average time that college freshmen spend studying each week is really 20 hours per week. We may want to compare this average time spent studying for freshmen that earned a GPA of 3.0 or higher and those that did not. In later chapters, we will be able to test if the average time studying differs for four groups: freshmen, sophomores, juniors and seniors.

Developing Null and Alternative Hypotheses

In statistical hypothesis testing, there are always two hypotheses. The hypothesis to be tested is called the **null hypothesis** and given the symbol H_0. The null hypothesis states that there is no difference between a hypothesized population mean and a sample mean. It is the status quo hypothesis.

For example, if we were to test the hypothesis that college freshmen study 20 hours per week, we would express our null hypothesis as:

$$H_0 : \mu = 20$$

We test the null hypothesis against an **alternative hypothesis,** which is given the symbol H_a. The alternative hypothesis is often the hypothesis that you believe yourself! It includes the outcomes not covered by the null hypothesis. In this example, our alternative hypothesis would express that freshmen *do not* study 20 hours per week:

$$H_a : \mu \neq 20$$

Example A

We have a medicine that is being manufactured and each pill is supposed to have 14 milligrams of the active ingredient. What are our null and alternative hypotheses?

Solution

$$H_0 : \mu = 14$$
$$H_a : \mu \neq 14$$

Our null hypothesis states that the population has a mean equal to 14 milligrams. Our alternative hypothesis states that the population has a mean that is different than 14 milligrams.

Example B

The school principal wants to test if it is true what teachers say – that high school juniors use the computer an average 3.2 hours a day. What are our null and alternative hypotheses?

Solution

$$H_0 : \mu = 3.2$$
$$H_a : \mu \neq 3.2$$

Our null hypothesis states that the population has a mean equal to 3.2 hours. Our alternative hypothesis states that the population has a mean that differs from 3.2 hours.

Deciding Whether to Reject the Null Hypothesis: One and Two-Tailed Hypothesis Tests

The alternative hypothesis can be supported only by *rejecting* the null hypothesis. To reject the null hypothesis means to find a large enough difference between your sample mean and the hypothesized (null) mean that it raises real doubt that the true population mean is 20. If the difference between the hypothesized mean and the sample mean is very large, we reject the null hypothesis. If the difference is very small, we do not. In each hypothesis test, we have to decide in advance what the magnitude of that difference must be to allow us to reject the null hypothesis. Below is an overview of this process. Notice that if we fail to find a large enough difference to reject, we **fail to reject** the null hypothesis. Those are your only two alternatives.
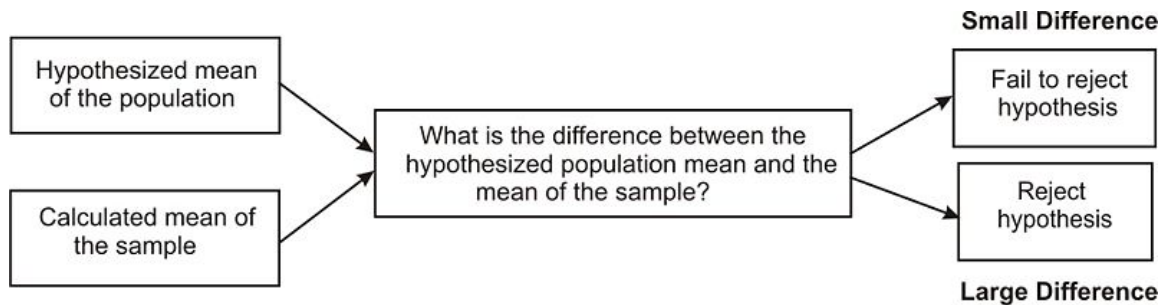
When a hypothesis is tested, a statistician must decide on how much of a difference between means is necessary in order to reject the null hypothesis.

Statisticians first choose a **level of significance** or **alpha** (α) **level** for their hypothesis test. Similar to the significance level you used in constructing confidence intervals, this alpha level tells us how improbable a sample mean must be for it to be deemed "significantly different" from the hypothesized mean. The most frequently used levels of significance are 0.05 and 0.01. An alpha level of 0.05 means that we will consider our sample mean to be significantly different from the hypothesized mean if the chances of observing that sample mean are less than 5%. Similarly, an alpha level of 0.01 means that we will consider our sample mean to be significantly differnet from the hypothesized mean if the chances of observing that sample mean are less than 1%.

Two-tailed Hypothesis Tests

A hypothesis test can be one-tailed or two-tailed. The examples above are all two-tailed hypothesis tests. We indicate that the average study time is either 20 hours per week, or it is not. Computer use averages 3.2 hours per week, or it does not. We do not specify whether we believe the true mean to be higher or lower than the hypothesized mean. We just believe it must be different.

In a two-tailed test, you will reject the null hypothesis if your sample mean falls in either tail of the distribution. For this reason, the alpha level (let's assume .05) is split across the two tails. The curve below shows the **critical regions** for a two-tailed test. These are the regions under the normal curve that, together, sum to a probability of 0.05. Each tail has a probability of 0.025. The z-scores that designate the start of the critical region are called the **critical values**.

If the sample mean taken from the population falls within these critical regions, or "rejection regions," we would conclude that there was too much of a difference and we would reject the null hypothesis. However, if the mean from the sample falls in the middle of the distribution (in between the critical regions) we would fail to reject the null hypothesis.
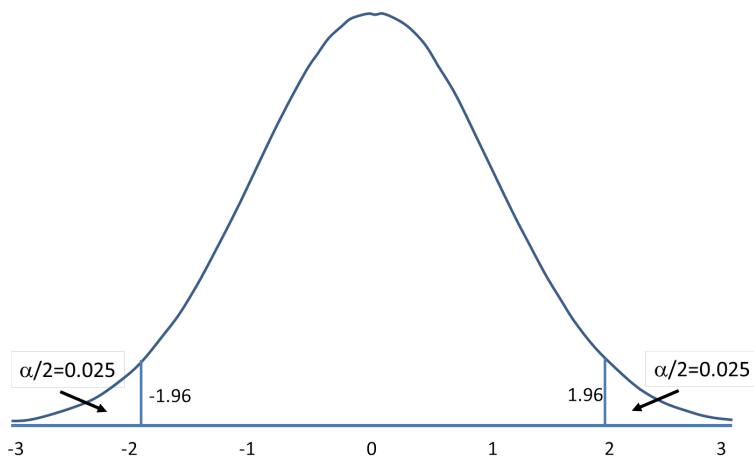
FIGURE 12.1

One-Tailed Hypothesis Test

We would use a single-tail hypothesis test when the direction of the results is anticipated or we are only interested in one direction of the results. For example, a single-tail hypothesis test may be used when evaluating whether or not to adopt a new textbook. We would only decide to adopt the textbook if it *improved* student achievement relative to the old textbook.

When performing a single-tail hypothesis test, our alternative hypothesis looks a bit different. We use the symbols of greater than or less than. For example, let's say we were claiming that the average SAT score of graduating seniors was GREATER than 1,110. Remember, our own personal hypothesis is the alternative hypothesis. Then our null and alternative hypothesis could look something like:

$$H_0 : \mu \leq 1100$$
$$H_a : \mu > 1100$$

In this scenario, our null hypothesis states that the mean SAT scores would be less than or equal to 1,100 while the alternate hypothesis states that the SAT scores would be greater than 1,100. A single-tail hypothesis test also means

that we have only one critical region because we put the entire critical region into just one side of the distribution. When the alternative hypothesis is that the sample mean is greater, the critical region is on the right side of the distribution (see below). When the alternative hypothesis is that the sample is smaller, the critical region is on the left side of the distribution.
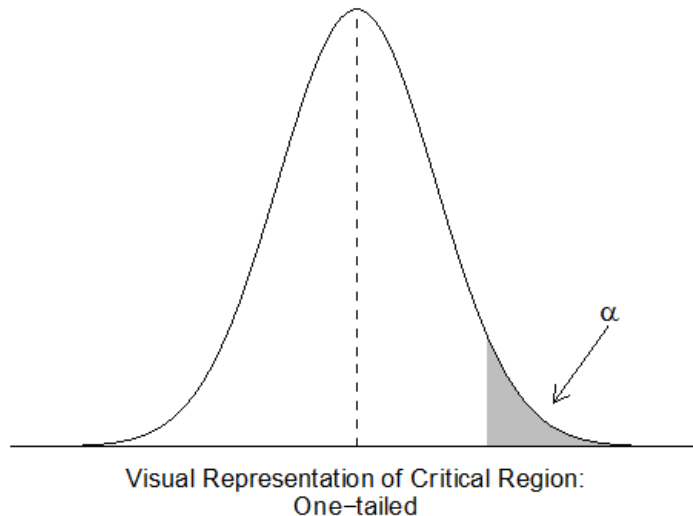

FIGURE 12.2

Visual Representation of Critical Region:
One–tailed

To calculate the critical regions, we must first find the **critical values** or the cut-offs where the critical regions start. This will be covered in the next section.

Type I and Type II Errors

Remember that there will be some sample means that are extremes – that is going to happen about 5% of the time, since 95% of all sample means fall within about two standard deviations of the mean. What happens if we run a hypothesis test and we get an extreme sample mean? It won't look like our hypothesized mean, even if it comes from that distribution. We would be likely to reject the null hypothesis. But we would be wrong.

When we decide to reject or not reject the null hypothesis, we have four possible scenarios:

   a. A true hypothesis is rejected.
   b. A true hypothesis is not rejected.
   c. A false hypothesis is not rejected.
   d. A false hypothesis is rejected.

If a hypothesis is true and we do not reject it (Option 2) or if a false hypothesis is rejected (Option 4), we have made the correct decision. But if we reject a true hypothesis (Option 1) or a false hypothesis is not rejected (Option 3) we have made an error. Overall, one type of error is not necessarily more serious than the other. Which type is more serious depends on the specific research situation, but ideally both types of errors should be minimized during the analysis.

TABLE 12.1: **The Four Possible Outcomes in Hypothesis Testing**

Decision Made	Null Hypothesis is True	Null Hypothesis is False
Reject Null Hypothesis	Type I Error	Correct Decision
Do not Reject Null Hypothesis	Correct Decision	Type II Error

The general approach to hypothesis testing focuses on the **Type I** error: rejecting the null hypothesis when it may be true. Guess what? The level of significance, also known as the alpha level, *IS* the probability of making a Type I error. At the 0.05 level, the decision to reject the hypothesis may be incorrect 5% of the time. Calculating the probability of making a **Type II** error is not as straightforward as calculating a Type I error, and we won't discuss that here.

You should be able to recognize what each type of error looks like in a particular hypothesis test. For example, suppose you are are testing whether listening to rock music helps you improve your memory of 30 random objects. Assume further that it doesn't. A Type I error would be concluding that listenting to rock music *did* help memory (but you are wrong). A Type I error will only occur when your null hypothesis is false. Let's assume that listening to rock music does improve memory. In this scenario, if you concluded that it didn't, you would be wrong again. But this time you would be making a Type II error — failing to find a significant difference when one in fact exists.

It is also important that you realize that the chance of making a Type I error is under our direct control. Often we establish the alpha level based on the severity of the consequences of making a Type I error. If the consequences are not that serious, we could set an alpha level at 0.10 or 0.20. In other words, we are comfortable making a decision where we could falsely reject the null hypothesis 10 to 20% of the time. However, in a field like medical research, we would set the alpha level very low (at 0.001 for example) if there was potential bodily harm to patients.

Lesson Summary

a. Hypothesis testing involves making educated guesses about a population based on a sample drawn from the population. We generate null and alternative hypotheses based on the mean of the population to test these guesses.

b. We establish critical regions based on level of significance or alpha (α) levels. If the value of the test statistic falls in these critical regions, we are able to reject it.

c. When we make a decision about a hypothesis, there are four different outcome and possibilities and two different types of errors. A Type I error is when we reject the null hypothesis when it is true and a Type II error is when we do not reject the null hypothesis, even when it is false.

Review Questions

1. If the difference between the hypothesized population mean and the mean of the sample is large, we ____ the null hypothesis. If the difference between the hypothesized population mean and the mean of the sample is small, we ____ the null hypothesis.

2. At the Chrysler manufacturing plant, there is a part that is supposed to weigh precisely 19 pounds. The engineers take a sample of parts and want to know if they meet the weight specifications. What are our null and alternative hypotheses?

3. In a hypothesis test, if difference between the sample mean and the hypothesized mean divided by the standard error falls in the middle of the distribution and in between the critical values, we ____ the null hypothesis. If this number falls in the critical regions and beyond the critical values, we ____ the null hypothesis.

4. Sacramento County high school seniors have an average SAT score of 1,020. From a random sample of 144 Sacramento High School students we find the average SAT score to be 1,100 with a standard deviation of 144. We want to know if these high school students are representative of the overall population. What are our

hypotheses ?
5. Please fill in the types of errors missing from the table below:

TABLE 12.2:

Decision Made	Null Hypothesis is True	Null Hypothesis is False
Reject Null Hypothesis	(1) ____	Correct Decision
Do not Reject Null Hypothesis	Correct Decision	(2) ____

Review Answers

1. Reject, Fail to Reject
2. $H_0 : \mu = 19, H_a : \mu \neq 19$
3. Fail to Reject, Reject
4. $H_0 : \mu = 1020, H_a : \mu \neq 1020, Z = 6.67$
5. Type I error, Type II error

12.2 Critical Values

Learning Objective

- Understand the relationship between the critical value and the critical region.
- To accurately identify the critical region for a one- or two-tailed test.
- To locate the critical value for a given significance level.

Critical Values

Critical values are the values that indicate the edge of the critical region. **Critical regions** describe the entire area of values that indicate you reject the null hypothesis. In other words, the **critical region** is the area encompassed by the values *not* included in the initial claim - the area of the 'tails' of the distribution.

The "tails" of a test are the values outside of the **critical values**. In other words, the tails are the ends of the distribution, and they begin at the greatest or least value included in the alternative hypothesis (the critical values).

In the graph below, the tails are in red and the rest of the distribution is in green. The critical values of the test in the image are -8 and 28, as these are the dividers between values supporting the alternative and null hypothesis. The area in red can also be seen as the **rejection region**, since an observed value in this region indicates that the null hypothesis should be rejected.

Hypothesis Tests and Their Tails

There are three types of test from a "tails" standpoint:

- A **left-tailed test** only has a tail on the left side of the graph:

- A **right-tailed test** only has a tail on the right side of the graph:

- A **two-tailed test** has tails on both ends of the graph. This is a test where the null hypothesis is a claim of a specific value. For example: $H_0 : X = 5$

Values supporting the null hypothesis occur in the green region.

Values supporting the alternative hypothesis occur in the red region.

-20   -10   0   10   20   30   40

Example A

A researcher claims that black horses are, on average, more than 30 lbs heavier than white horses, which average 1100 lbs. What is the null hypothesis, and what kind of test is this?

Solution

The null hypothesis would be notated $H_0 : \mu \leq 1130 \ lbs$

This is a right-tailed test, since the tail of the graph would be on the right. Recognize that values above 1130 would indicate that the null hypothesis be *rejected,* and the red area represents the *rejection region*.

Example B

A package of gum claims that the flavor lasts more than 39 minutes. What would be the null hypothesis of a test to determine the validity of the claim? What sort of test is this?

Solution

The null hypothesis would by notated as $H_0 : \mu \leq 39$.

This is a right-tailed test, since the rejection region would consist of values greater than 39.

Example C

An ice pack claims to stay cold between 35 and 65 minutes. What would be the null hypothesis of a test to determine the validity of the claim? What sort of test would it be?

Solution

The null hypothesis would be $H_0 : 35 > \mu \cup \mu > 65$.

This is a two-tailed test, since the null hypothesis contains the values above and below a given range.

Locating a Critical Value

In this lesson, we will use a table to find critical values. For the moment, we will be dealing only with the *Z*-score critical values of a normal distribution. (NOTE: The z-table is an appropriate resource for conducting tests when we know what the parameters are of our population. However, we will turn to a new table when we discuss how to conduct your typical hypothesis test with a sample.)

Example D

What is the critical value $\left(Z_{\frac{\alpha}{z}} \right)$ for a 95% confidence level, assuming a two-tailed test?

Solution

A 95% confidence level means that a total of 5% of the area under the curve is considered the critical region.

Since this is a two-tailed test, $\frac{1}{2}$ of 5% = 2.5% of the values would be in the left tail, and the other 2.5% would be in the right tail. Looking up the *Z*-score associated with 0.025 on a reference table, we find 1.96. Therefore, +1.96 is the critical value of the right tail and -1.96 is the critical value of the left tail.

**The critical value for a 95% confidence level is $Z = +/-1.96$.**

Example E

Sketch the *Z*-score critical region for Example D.

Solution

Sketch the graph of the normal distribution with the given values and mark the critical values from Example D, then shade the area from the critical values away from the center. The shaded areas are the critical regions.

Critical Regions for a Two-Tailed z Test
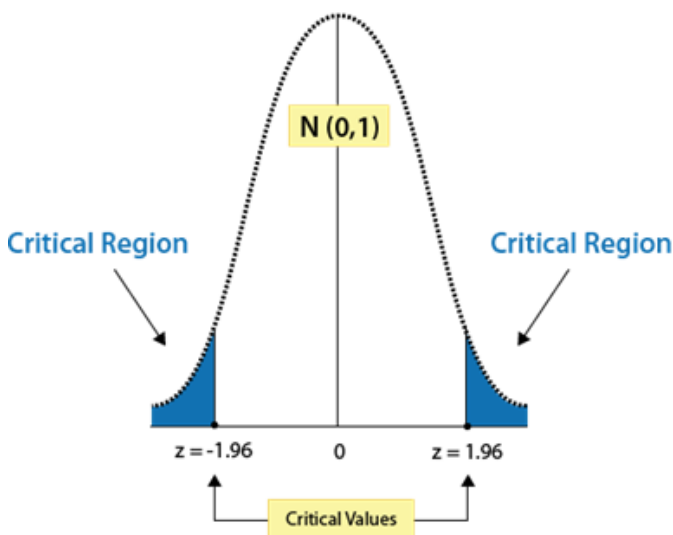
Example F

What would be the critical value for a right-tailed test with $\alpha = 0.01$?

Solution

If $\alpha = 0.01$, then the area under the curve representing H_1, the alternative hypothesis, would be 99%, since α (alpha) is the same as the area of the rejection region. Using the Z-score reference table above, we find that the Z-score associated with 0.9900 is approximately 2.33.

It appears that the critical value is $Z = 2.33$.

Let's see if that answer makes sense. Since this is a right-tailed test, α is on the right end of the graph, and $1 - \alpha$ is on the left. A Z-score of +2.33 is well to the right of the center of the graph, and describes the area under the curve from that point to the left, nearly the entire graph. Since the initial question specified $\alpha = 0.01$, indicating that only 1% of the area is in the critical region, $Z = +2.33$ is quite reasonable.

Z=+2.33

Vocabulary

Critical values are values separating the values that support or reject the null hypothesis.

Critical regions are the areas under the distribution curve representing values that support the null hypothesis.

Guided Practice

1. What would be the critical value for a left-tailed test with $\alpha = 0.01$?
2. What would be the critical *region* for a two-tailed test with $\alpha = 0.08$?
3. What would be the α for a right-tailed test with a critical value of $Z = 1.76$?

Solutions

1. A left-tailed test with $\alpha = 0.01$ would have 99% of the area under the curve outside of the critical region. If we use a reference to find the Z-score for 0.99, we get approximately 2.33. However, a Z-score of 2.33 is significantly to the right of the center of the distribution, including all the area to the left and only leaving a very small alpha value on the right. While we are indeed looking for a critical value with only a very small alpha, this is a *left-tailed* test, so the critical value we need is *negative*. The solution is z = -2.33.

2. We are looking for the critical region here, but let's start by finding the critical values. This is a two-tailed test, so half of the alpha will be in the left tail, and half in the right. That means that we are looking for a positive/negative critical value associated with an alpha of 0.04, which indicates that we need to find the

Z-score for $(1-0.04) = 0.96$. Referring to the Z-score table, we see that 0.96 corresponds to approximately 1.75. The critical values, then are +/- 1.75, and the critical region would be $Z < -1.75 \cup Z > 1.75$.

3. The area under the curve associated with a Z-score of 1.76, according to the reference table above, is 0.9608. Since 96.08% of the area is to the left of $Z = 1.76$, that leaves approximately $1 - 0.9608 = 0.0392$ as the area in the critical region. The solution is alpha = 0.0392.

Additional Guided Practice

For each question:

a) State the null and alternative hypothesis for a test to determine the validity of the claim

b) Identify left, right, or two-tailed test type

1. The average Siberian Husky weighs more than 35 pounds.
2. The United Parcel Service claims less than 5-day shipping times for ground packages.

Solutions

1. (a) $H_1 : \mu > 35\ lbs$ and $H_0 : 35\ lbs \geq \mu$
   (b) This is a right-tailed test, since the rejection region is located on the right end of the graph.
2. (a) $H_1 : \mu < 5\ days$ and $H_0 : \mu \geq 5\ days$
   (b) This is a left-tailed test, since the 'reject' values are on the left end of the graph.
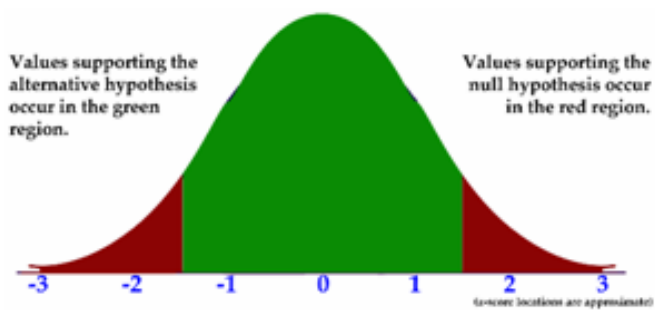
More Practice

For each question:

a) State the null and alternative hypothesis for a test to determine the validity of the claim
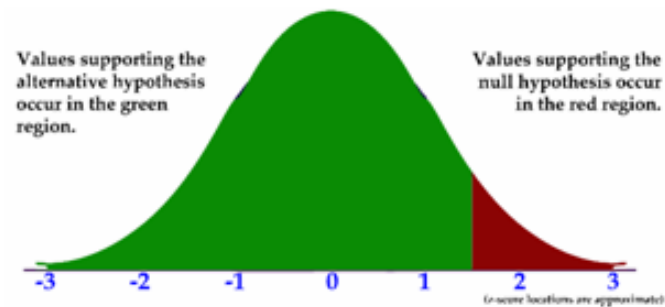
b) Identify left, right, or two-tailed test type

1. The average miniature horse is less than 34 inches at the shoulder. Tiny Equines claims that their miniature horses are an average of 3 inches less than the breed average.
2. Speedy Solutions is a delivery company that claims all deliveries average less than 72 hours.
3. Terrific Textiles claims that 45% to 55% of t-shirts sold in CO are red.
4. Gas Hater Car Sales claims that 70% or more of the cars on the lot average more than 40 mpg.
5. The 2011 average service time for Fast Fatz Burgers was 56 seconds. The store manager claims that the 2012 average is more than 5 seconds faster.
6. Colorful Cavity Causers sells multi-colored candies by the bag. Each bag is claimed to average between 5.5 and 6.5 oz.
7. Aaron works for an electronics company, quality-testing batteries. The company claims a minimum of 5 hrs battery life.
8. The lessons in this Probability and Statistics text average 13 practice problems each.
9. The outside temperature in Maui averages between 72 and 83 degrees year-round.
10. 72% of elementary school teachers are female.

11.

Values supporting the
alternative hypothesis
occur in the green
region.

Values supporting the
null hypothesis occur
in the red region.

-3   -2   -1   0   1   2   3
(z-score locations are approximate)

12.

Values supporting the
alternative hypothesis
occur in the green
region.

Values supporting the
null hypothesis occur
in the red region.

-3   -2   -1   0   1   2   3
(z-score locations are approximate)

13.

Values supporting the
alternative hypothesis
occur in the green
region.

Values supporting the
null hypothesis occur
in the red region.

-3   -2   -1   0   1   2   3
(z-score locations are approximate)

12.3 One-Sample T Test

Learning Objectives

- Understand the steps of hypothesis testing.
- Know when to use a z-test and when to use a t-test to test a single-sample hypothesis.
- Compare the t-distribution to the Normal distribution.
- Understand degrees of freedom.

Steps of Hypothesis Testing

Recall that hypothesis testing is a form of statistical inference. Previously, we inferred about a population by calculating a confidence interval. We estimated the *true* mean of a population from a *sample* mean and created a *confidence interval* provided a *margin of error* for our estimate. In this chapter, we also use sample data to help us make decisions about what is true about a population.

When conducting a hypothesis test, we are asking ourselves whether the information in the sample is **consistent**, or **inconsistent**, with the null hypothesis about the population. We follow a series of four basic steps:

1. State the null and alternative hypotheses.
2. Select the appropriate significance level and check the test assumptions.
3. Analyze the data and compute the test statistic.
4. Interpret the result

If we *reject* the null hypothesis we are saying that the difference between the observed sample mean and the hypothesized population mean is too great to be attributed to chance. When we **fail to reject** the null hypothesis, we are saying that the difference between the observed sample mean and the hypothesized population mean is probable if the null hypothesis is true. Essentially, we are willing to attribute this difference to sampling error.

Conducting a Hypothesis Test on One Sample Mean When the Population Parameters are Known

Although this is rarely the case, we can use our familiar z-statistic to conduct a hypothesis test on a single sample mean. In short, we find the z-statistic of our sample mean in the sampling distribution and determine if that z-score falls within the critical (rejection) region or not. This test is only appropriate when you know the true mean and standard deviation of the population.

Example A

The school nurse thinks the average height of 7th graders has increased. The average height of a 7th grader five years ago was 145 cm with a standard deviation of 20 cm. She takes a random sample of 200 students and finds that the average height of her sample is 147 cm. Are 7th graders now taller than they were before? Conduct a single-tailed hypothesis test using a .05 significance level to evaluate the null and alternative hypotheses.

First, we develop our null and alternative hypotheses:

$$H_0 : \mu \leq 145$$
$$H_a : \mu > 145$$

Choose $\alpha = .05$. The critical value for this one tailed test is z=1.64. This is a one-tailed test, and a z-score of 1.64 cuts off 5% in the single tail. Any test statistic greater than 1.64 will be in the rejection region.

Next, we calculate the test statistic for the sample of 7th graders.

$$z = \frac{147 - 145}{\frac{20}{\sqrt{200}}} \approx 1.414$$

The calculated $z-$score of 1.414 is smaller than 1.64 and thus does not fall in the critical region. Our decision is to fail to reject the null hypothesis and conclude that the probability of obtaining a sample mean equal to 147 is likely to have been due to chance.

Example B

A farmer is trying out a planting technique that he hopes will increase the yield on his pea plants. The average number of pods on one of his pea plants is 145 pods with a standard deviation of 100 pods. This year, after trying his new planting technique, he takes a random sample of his plants and finds the average number of pods to be 147. He wonders whether or not this is a statistically significant increase. What are his hypotheses and the test statistic?

1. First, we develop our null and alternative hypotheses:

$$H_0 : \mu \leq 145$$
$$H_a : \mu > 145$$

This alternative hypothesis is >since he believes that there might be a gain in the number of pods.

2. Next, we calculate the test statistic for the sample of pea plants.

$$z = \frac{(\bar{x} - \mu)}{\frac{\sigma}{\sqrt{n}}} = \frac{(147 - 145)}{\frac{100}{\sqrt{144}}} \approx 0.24$$

3. If we choose $\alpha = .05$
4. The critical value will be 1.645. We will reject the null hypothesis if the test statistic is greater than 1.645. The value of the test statistic is 0.24.
5. This is less than 1.645 and so our decision is to fail to reject H_0.
6. Based on our sample we believe the mean is equal to 145.

Reporting P-Values as an Alternative to Looking at Critical Regions

We can also evaluate a hypothesis by asking, "What is the probability of obtaining the value of the test statistic we did if the null hypothesis is true?" This is called the $p-value$.

Example C

Let's use the example about the pea farmer. As we mentioned, the farmer is wondering if the number of pea pods per plant has gone up with his new planting technique and finds that out of a sample of 144 peas there is an average number of 147 pods per plant (compared to a previous average of 145 pods). To determine the $p-$ *value* we ask what is $P(z > .24)$? That is, what is the probability of obtaining a z value greater than .24 is the null hypothesis is true? Using technology, we find this probability to be .49. This indicates that there is a 49% chance that under the null hypothesis the peas will produce more than 145 pods. Alternatively, we can just indicate that p >.05. Since we set alpha at .05, we won't reject if the probability of observing that sample mean is >.05.

When you use technology to conduct a hypothesis test, as in lab, you will receive a p-value as part of your output. The p-values is the likelihood of observing that particuluar sample value *if the null hypothesis were true*. Therefore, if the p-value is smaller than your significance level, you can reject the null hypothesis.

Hypothesis Tests When You Don't Know Your Population Parameters

A z-test makes for an easy hypothesis test, but most of the time we can't use it. The reality is that most of our analyses are done when we *don't* know what is true about a population. We don't know the true population mean, and we certainly don't know the true population standard deviation. Instead, we want to conduct a hypothesis test using only our sample and the information it can provide us. How can we do this?

How We Learned to Test Hypotheses on Samples, without Knowledge of the Population

Back in the early 1900's, a chemist at a brewery in Ireland discovered that when he was working with very small samples, the distributions of the mean differed significantly from the normal distribution. He noticed that as his sample sizes changed, the shape of the distribution changed as well. He published his results under the pseudonym 'Student' and this concept and the distributions for small sample sizes are now known as "Student's $t-$distributions."

The Student's **t-distribution** is similar to the normal distribution, except it is more spread out and wider in appearance, and has thicker tails. As the number of observations gets larger, the t-distribution shape becomes more and more like the shape of the normal distribution. In fact, if we had an infinite number of observations, the t distribution would perfectly match the normal distribution. It is the t-distribution that allows us to test hypotheses when we don't know the true population standard deviation, as we'll see below.
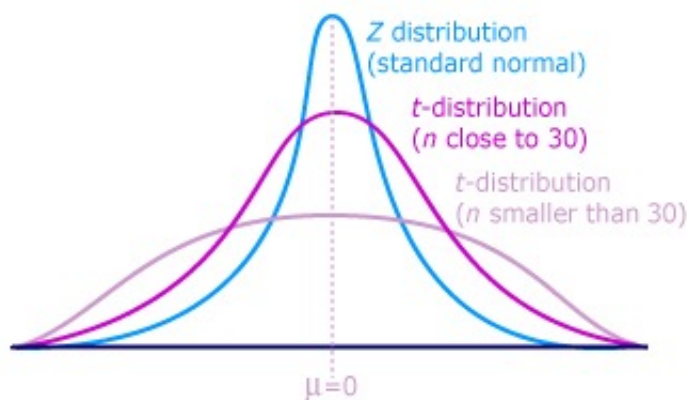

FIGURE 12.3

The differences between the t-distribution and the normal distribution are more exaggerated when there are fewer data points, and therefore fewer **degrees of freedom**. Degrees of freedom are essentially the number of samples that have the 'freedom' to change without affecting the sample mean. A clear description of degrees of freedom is beyond the scope of this lesson, but you can find many online lessons describing them if you are interested. For our

purposes, all you really need to know about degrees of freedom is that there is always one less degree of freedom than the number of data points:

$$df = n - 1$$

The reason you need to know how to find the number of degrees of freedom is quite simple: when you use a t-distribution for a hypothesis test, there is a *different* critical value for each number of degrees of freedom. The larger your sample, the closer the critical value gets to the z-score for your alpha level. Below is the t-table. It shows you the critical value for the selected alpha level (across the top) for the degrees of freedom of your test. If you were conducting a two-tailed hypothesis test on a sample of 25 students, your df = 24 and your critical value at α=0.05 is ±2.064 because there is 0.025 in each tail.

Conditions for Using T-Test

The $t-$distribution can be used with any statistic having a bell-shaped distribution. The Central Limit Theorem states the sampling distribution of a statistic will be close to normal with a large enough sample size. As a rough estimate, the Central Limit Theorem predicts a roughly normal distribution under any of the following conditions:

- The population distribution is normal; or
- The sampling distribution is symmetric and the sample size is ≤ 15;' or
- The sampling distribution is moderately skewed and the sample size is $16 \leq n \leq 30$; or
- The sample size is greater than 30, without outliers.

T Table

TABLE 12.3: Upper-Tail Probability

DF	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001
1	1.000	1.376	1.963	3.078	6.314	12.710	31.820	63.660	318.300
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.330
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.210
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527

TABLE 12.3: (continued)

22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	3.261
60	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232
80	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	3.160
z*	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090

T-Test for One Sample Mean

So when do we use the t-distribution and when do we use the normal distribution? It's simple: When we *know* the population standard deviation we use the normal distribution. When we *don't know* the population standard deviation (so we need to use our sample standard deviation), we use the t-distribution.

We use the Student's $t-$distribution in hypothesis testing the same way that we use the normal distribution. Each row in the t distribution table (see above) represents a different $t-$distribution. Each distribution is associated with a unique number of degrees of freedom (the number of observations minus one). The column headings in the table represent the portion of the area in the tails of the distribution –we use the numbers in the table just as we used the $z-$scores.

In calculating the $t-$test statistic, we use the formula:

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

where:

t is the test statistic and has $n - 1$ degrees of freedom.

$\bar{x}$ is the sample mean

μ_0 is the population mean under the null hypothesis.

s is the sample standard deviation

n is the sample size

$\frac{s}{\sqrt{n}}$ is the estimated standard error

Assumptions of the single sample t-test:

- A random sample is used.
- The random sample is made up of independent observations
- The population distribution must be nearly normal, or the size of the sample is large.

Example D

The high school athletic director is asked if football players are doing as well academically as the other student athletes. We know from a previous study that the average GPA for the student athletes is 3.10. After an initiative to help improve the GPA of student athletes, the athletic director randomly samples 20 football players and finds that the average GPA of the sample is 3.18 with a sample standard deviation of 0.54. Is there a significant improvement? Use a 0.05 significance level.

- **Hypothesis Step 1:** Cleary state the null and alternative hypotheses.

$$H_0 : \mu = 3.10$$
$$H_a : \mu \neq 3.10$$

- **Hypothesis Step 2:** Identify the appropriate significance level and confirm the test assumptions.

  We were told that we should use a 0.05 significance level. We assume that each football player is independently tested –that their GPA is not related to another football player's GPA. Without the data, we have to assume that the sample is nearly normal (the sample is an indication of the population shape). The size of the sample also helps here, as we have 20 players. So, we can conclude that the assumptions for the single sample t-test have been met.

- **Hypothesis Step 3**: Analyze the data

  We use our t-test formula:

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{3.18 - 3.10}{\frac{0.54}{\sqrt{20}}} = 0.66$$

  We know that we have 20 observations, so our degrees of freedom for this test is 19. Nineteen degrees of freedom at the 0.05 significance level gives us a critical value of $\pm\, 2.093$.

- **Hypothesis Step 4**: Interpret your results

  Since our calculated t-test value is lower than our t-critical value, we fail to reject the Null Hypothesis. Therefore, the average GPA of the sample of football players is not significantly different from the average GPA of student athletes. Therefore, we can conclude that the difference between the sample mean and the hypothesized value is not sufficient to attribute it to anything other than sampling error. Thus, the athletic director can conclude that the mean academic performance of football players does not differ from the mean performance of other student athletes.

Example E

Duracell manufactures batteries that the CEO claims will last an average of 300 hours under normal use. A researcher randomly selected 20 batteries from the production line and tested these batteries. The tested batteries had a mean life span of 270 hours with a standard deviation of 50 hours. Do we have enough evidence to suggest that the claim of an average lifetime of 300 hours is false?

- **Hypothesis Step 1**: Clearly state the Null and Alternative Hypothesis

$H_0 : \mu = 300$

$H_A : \mu \neq 300$

- **Hypothesis Step 2**: Identify the appropriate significance level and confirm the test assumptions.

We'll use the standard significance level of 0.05, and we assume a normal population distribution.

- **Hypothesis Step 3**: Analyze the data and compute the test statistic.

First, we need to calculate the Standard Error:

$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$

$SE_{\bar{x}} = \frac{50}{\sqrt{20}} = 11.18$

Now, we'll use that value in out t-test formula:

$t = \frac{\bar{x} - \mu}{SE_{\bar{x}}} = \frac{270 - 300}{11.18} = -2.68$

We know that we have 20 batteries, so our degrees of freedom for this test is (20-1)= 19. Nineteen degrees of freedom at the 0.05 significance level gives us a critical value of ± 2.093.
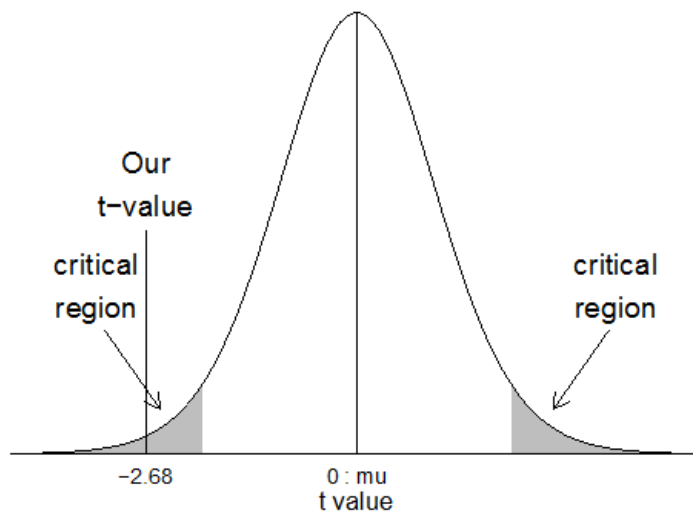
What does this look like visually?

FIGURE 12.4

- **Hypothesis Step 4**: Interpret your results

Since our calculated t-test value is outside of our t-critical value –it lies in the critical region –we reject the Null Hypothesis. The average battery life of the sample is significantly different from the average battery life claim by the CEO.

Estimation as a follow-up to a Hypothesis Test

When a hypothesis is rejected, it is often useful to turn to estimation to try to capture the true value of the population mean. Now that we have rejected the claim that the average lifetime of a battery is 300 hours, we want to know how long these batteries *do* last. Let's take a look at how we can use estimation to answer this question.

We will choose a 95% confidence interval. This is essentially the same degree of confidence that we had in our t-test, since confidence = 1 - α. To calculate the confidence interval, we need to know three things: the mean of our sample, the standard error, and the critical value. The formula for the confidence interval is:

$$\bar{x} \pm \text{Margin of Error}$$

Where the Margin of Error is defined as:

$$ME = \text{t-critical} \cdot SE_{\bar{x}}$$

Note that was are using the t-critical value (rather than z) because we are working only with sample data. We do not know the true population mean or standard deviation. We have to use both our sample mean and our sample standard error to create the confidence interval.

So, for our above Durablast example, the Margin of Error would be (remember we have 19 degrees of freedom):

$$ME = 2.093 \cdot 11.18 = 23.40$$

Therefore, our 95% confidence *interval* is from (270 –23.40) to (270 + 23.40). That interval of (246.6, 293.4) shows us something: the confidence interval for the population mean does *not* include 300. This is what we would expect, since we rejected the null hypothesis in our earlier hypothesis test.

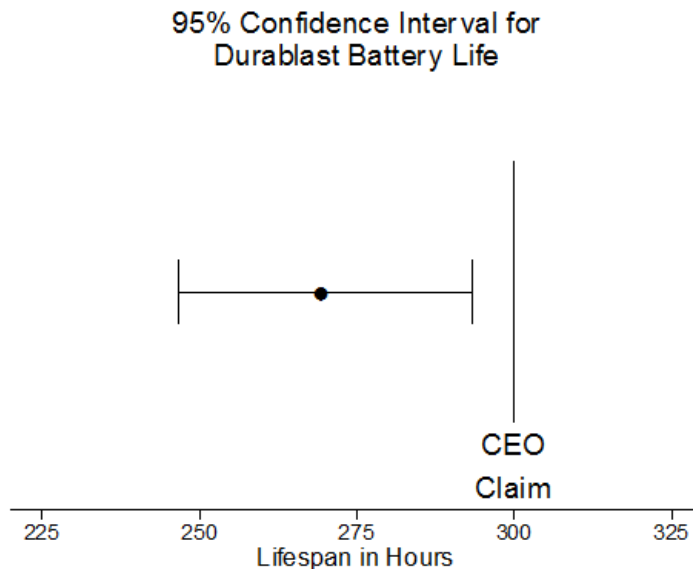The visual representation of this confidence interval would be:

FIGURE 12.5

The central dot is the mean value of the sample (270 hours), with the left and right whiskers representing the lower

and upper values of the 95% confidence interval.

If we were to interpret the 95% confidence interval we would say: "I am 95% confident that the true population mean of Durablast battery lifespan is between 246.6 and 293.4 hours."

Example F

You have just taken ownership of a pizza shop. The previous owner told you that you would save money if you bought the mozzarella cheese in a 4.5 pound slab. Each time you purchase a slab of cheese, you weigh it to ensure that you are receiving 72 ounces of cheese. The results of 7 random measurements are 70, 69, 73, 68, 71, 69 and 71 ounces. Are these differences due to chance or is the distributor giving you less cheese than you deserve?

    a. State the hypotheses.
    b. Calculate the test statistic.
    c. Would the null hypothesis be rejected at the 10% level? The 5% level? The 1% level?

Solution

    a. For H_0 the mean weight of cheese $\mu = 72$; and for $H_a : \mu \neq 72$.
    b. Begin by determining the mean of the sample and the sample standard deviation. This can be done using the graphing calculator. $\bar{x} = 70.143$ and $s = 1.676$.

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$
$$t = \frac{70.143 - 72}{\frac{1.676}{\sqrt{7}}}$$
$$t \approx -2.9315$$

    c. The test statistic computed in part b) was -2.9315.
    d. The null hypothesis would be rejected at the .10 and the .05 levels, but not at the .01 level.

Lesson Summary

A test of significance is done when a claim is made about the value of a population parameter. The test can only be conducted if the random sample taken from the population came from a distribution that is normal or approximately normal. When you use s to estimate σ, you must use t instead of z to complete the significance test for a mean.

More Practice

1. In hypothesis testing, when we know the population standard deviation it is appropriate to use the _____- distribution. When we do not know the population standard deviation, we should use the _____ distribution
2. True or False: When we fail to reject the null hypothesis, we are saying that the difference between the observed sample mean and the hypothesized population mean is probable if the null hypothesis is true.
3. The dean from UCLA is concerned that the student's grade point averages have changed dramatically in recent years. The graduating seniors' mean GPA over the last five years is 2.75. The dean randomly samples 256 seniors from the last graduating class and finds that their mean GPA is 2.85, with a sample standard deviation

of 0.65.

    a. What would the null and alternative hypotheses be for this scenario?

    b. What would the standard error be for this particular scenario?

    c. Describe in your own words how you would set the critical regions and what they would be at an alpha level of .05.

    d. Test the null hypothesis and explain your decision

4. For each of the following scenarios, state which one is more likely to lead to the rejection of the null hypothesis?

    a. A one-tailed or two-tailed test

    b. .05 or .01 level of significance

    c. A sample size of $n = 144$ or $n = 444$