

CK-12 Probability and Statistics - Advanced (Second Edition)

Ellen Lawsky
Larry Ottman
Raja Almukkahal
Brenda Meery
Danielle DeLancey

Chapter 8

Hypothesis Testing

Say Thanks to the Authors
Click <http://www.ck12.org/saythanks>
(No sign in required)

ck-12 **flexbook**
next generation textbooks

To access a customizable version of this book, as well as other interactive content, visit www.ck12.org

CK-12 Foundation is a non-profit organization with a mission to reduce the cost of textbook materials for the K-12 market both in the U.S. and worldwide. Using an open-content, web-based collaborative model termed the **FlexBook®**, CK-12 intends to pioneer the generation and distribution of high-quality educational content that will serve both as core text as well as provide an adaptive environment for learning, powered through the **FlexBook Platform®**.

Copyright © 2014 CK-12 Foundation, www.ck12.org

The names “CK-12” and “CK12” and associated logos and the terms “**FlexBook®**” and “**FlexBook Platform®**” (collectively “CK-12 Marks”) are trademarks and service marks of CK-12 Foundation and are protected by federal, state, and international laws.

Any form of reproduction of this book in any format or medium, in whole or in sections must include the referral attribution link <http://www.ck12.org/saythanks> (placed in a visible location) in addition to the following terms.

Except as otherwise noted, all CK-12 Content (including CK-12 Curriculum Material) is made available to Users in accordance with the Creative Commons Attribution-Non-Commercial 3.0 Unported (CC BY-NC 3.0) License (<http://creativecommons.org/licenses/by-nc/3.0/>), as amended and updated by Creative Commons from time to time (the “CC License”), which is incorporated herein by this reference.

Complete terms can be found at <http://www.ck12.org/terms>.

Printed: December 17, 2014

flexbook
next generation textbooks



AUTHORS

Ellen Lawsky
Larry Ottman
Raja Almukkahal
Brenda Meery
Danielle DeLancey

CHAPTER

8

Hypothesis Testing

Chapter Outline

- 8.1 HYPOTHESIS TESTING AND THE P-VALUE**
 - 8.2 TESTING A PROPORTION HYPOTHESIS**
 - 8.3 TESTING A MEAN HYPOTHESIS**
 - 8.4 STUDENT'S T-DISTRIBUTION**
 - 8.5 TESTING A HYPOTHESIS FOR DEPENDENT AND INDEPENDENT SAMPLES**
-

8.1 Hypothesis Testing and the P-Value

Learning Objectives

- Develop null and alternative hypotheses to test for a given situation.
- Understand the critical regions of a graph for one- and two-tailed hypothesis tests.
- Calculate a test statistic to evaluate a hypothesis.
- Test the probability of an event using the p -value.
- Understand Type I and Type II errors.
- Calculate the power of a test.

Introduction

In this chapter we will explore hypothesis testing, which involves making conjectures about a population based on a sample drawn from the population. Hypothesis tests are often used in statistics to analyze the likelihood that a population has certain characteristics. For example, we can use hypothesis testing to analyze if a senior class has a particular average SAT score or if a prescription drug has a certain proportion of the active ingredient.

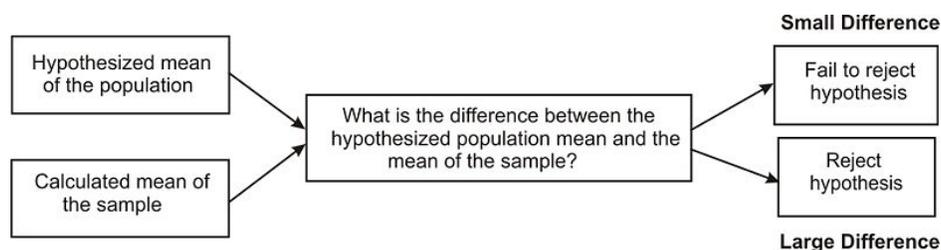
A hypothesis is simply a conjecture about a characteristic or set of facts. When performing statistical analyses, our hypotheses provide the general framework of what we are testing and how to perform the test.

These tests are never certain and we can never prove or disprove hypotheses with statistics, but the outcomes of these tests provide information that either helps support or refute the hypothesis itself.

In this section we will learn about different hypothesis tests, how to develop hypotheses, how to calculate statistics to help support or refute the hypotheses and understand the errors associated with hypothesis testing.

Developing Null and Alternative Hypotheses

Hypothesis testing involves testing the difference between a hypothesized value of a population parameter and the estimate of that parameter which is calculated from a sample. If the parameter of interest is the mean of the populations in hypothesis testing, we are essentially determining the magnitude of the difference between the mean of the sample and they hypothesized mean of the population. If the difference is very large, we reject our hypothesis about the population. If the difference is very small, we do not. Below is an overview of this process.



In statistics, the hypothesis to be tested is called the null hypothesis and given the symbol H_0 . The alternative hypothesis is given the symbol H_a .

The null hypothesis defines a *specific* value of the population parameter that is of interest. Therefore, the null hypothesis always includes the possibility of equality. Consider

$$H_0 : \mu = 3.2$$

$$H_a : \mu \neq 3.2$$

In this situation if our sample mean, \bar{x} , is very different from 3.2 we would reject H_0 . That is, we would reject H_0 if \bar{x} is much larger than 3.2 or much smaller than 3.2. This is called a *2-tailed test*. An \bar{x} that is very unlikely if H_0 is true is considered to be good evidence that the claim H_0 is not true. Consider $H_0 : \mu \leq 3.2$ $H_a : \mu > 3.2$. In this situation we would reject H_0 for very large values of \bar{x} . This is called a *one tail test*. If, for this test, our data gives $\bar{x} = 15$, it would be highly unlikely that finding \bar{x} this different from 3.2 would occur by chance and so we would probably reject the null hypothesis in favor of the alternative hypothesis.

Example: If we were to test the hypothesis that the seniors had a mean SAT score of 1100 our null hypothesis would be that the SAT score would be equal to 1100 or:

$$H_0 : \mu = 1100$$

We test the null hypothesis against an alternative hypothesis, which is given the symbol H_a and includes the outcomes not covered by the null hypothesis. Basically, the alternative hypothesis states that there is a difference between the hypothesized population mean and the sample mean. The alternative hypothesis can be supported only by rejecting the null hypothesis. In our example above about the SAT scores of graduating seniors, our alternative hypothesis would state that there is a difference between the null and alternative hypotheses or:

$$H_a : \mu \neq 1100$$

Let's take a look at examples and develop a few null and alternative hypotheses.

Example: 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. This is two tailed.

Example: 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?

$$H_0 : \mu = 3.2$$

$$H_a : \mu \neq 3.2$$

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. This is two tailed.

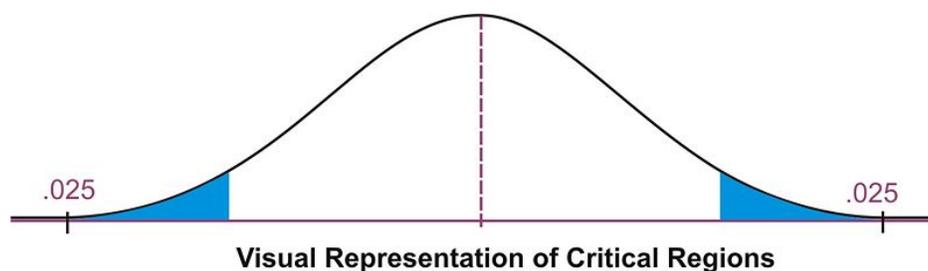
Deciding Whether to Reject the Null Hypothesis: One-Tailed and Two-Tailed Hypothesis Tests

When a hypothesis is tested, a statistician must decide on how much evidence is necessary in order to reject the null hypothesis. For example, if the null hypothesis is that the average height of a population is 64 inches a statistician wouldn't measure one person who is 66 inches and reject the hypothesis based on that one trial. It is too likely that the discrepancy was merely due to chance.

We use statistical tests to determine if the sample data give good evidence against the claim (H_0). The numerical measure that we use to determine the strength of the sample evidence we are willing to consider strong enough to reject H_0 is called the *level of significance* and it is denoted by α . If we choose, for example, $\alpha = .01$ we are saying that we would get data at least as unusual as the data we have collected no more than 1% of the time when H_0 is true.

The most frequently used levels of significance are 0.05 and 0.01. If our data results in a statistic that falls within the region determined by the level of significance then we reject H_0 . The region is therefore called the *critical region*. When choosing the level of significance, we need to consider the consequences of rejecting or failing to reject the null hypothesis. If there is the potential for health consequences (as in the case of active ingredients in prescription medications) or great cost (as in the case of manufacturing machine parts), we should use a more 'conservative' critical region with levels of significance such as .005 or .001.

When determining the critical regions for a two-tailed hypothesis test, the level of significance represents the extreme areas under the normal density curve. We call this a two-tailed hypothesis test because the critical region is located in both ends of the distribution. For example, if there was a significance level of 0.05 the critical region would be the most extreme 5 percent under the curve with 2.5 percent on each tail of the distribution.



Therefore, if the mean from the sample taken from the population falls within one of these critical regions, we would conclude that there was too much of a difference between our sample mean and the hypothesized population mean 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.

We calculate the critical region for the single-tail hypothesis test a bit differently. 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. A single-tail hypothesis simply states that the mean is greater or less than the hypothesized value.

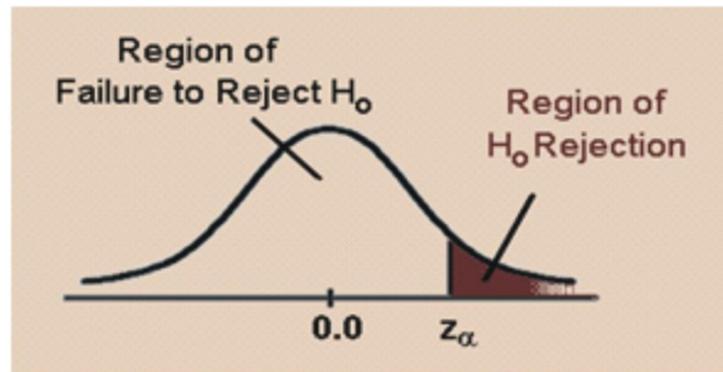
When performing a single-tail hypothesis test, our alternative hypothesis looks a bit different. When developing the alternative hypothesis in a single-tail hypothesis test we would use the symbols of greater than or less than. Using our example about SAT scores of graduating seniors, our null and alternative hypothesis could look something like:

$$H_0 : \mu = 1100$$

$$H_a : \mu > 1100$$

In this scenario, our null hypothesis states that the mean SAT scores would be equal to 1100 while the alternate hypothesis states that the SAT scores would be greater than 1100. A single-tail hypothesis test also means that we

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



To calculate the critical regions, we must first find the critical values or the cut-offs where the critical regions start. To find these values, we use the critical values found specified by the z -distribution. These values can be found in a table that lists the areas of each of the tails under a normal distribution. Using this table, we find that for a 0.05 significance level, our critical values would fall at 1.96 standard errors above and below the mean. For a 0.01 significance level, our critical values would fall at 2.57 standard errors above and below the mean. Using the z -distribution we can find critical values (as specified by standard z scores) for any level of significance for either single- or two-tailed hypothesis tests.

Example: Determine the critical value for a single-tailed hypothesis test with a 0.05 significance level.

Using the z distribution table, we find that a significance level of 0.05 corresponds with a critical value of 1.645. If alternative hypothesis is the mean is greater than a specified value the critical value would be 1.645. Due to the symmetry of the normal distribution, if the alternative hypothesis is the mean is less than a specified value the critical value would be -1.645.

Technology Note: Finding critical z values on the TI83/84 Calculator

You can also find this critical value using the TI83/84 calculator: 2nd [DIST] invNorm(.05,0,1) returns -1.64485. The syntax for this is invNorm (area to the left, mean, standard deviation).

Calculating the Test Statistic

Before evaluating our hypotheses by determining the critical region and calculating the test statistic, we need confirm that the distribution is normal and determine the hypothesized mean μ of the distribution.

To evaluate the sample mean against the hypothesized population mean, we use the concept of z -scores to determine how different the two means are from each other. Based on the Central Limit theorem the distribution of \bar{X} is normal with mean, μ and standard deviation, $\frac{\sigma}{\sqrt{n}}$. As we learned in previous lessons, the z score is calculated by using the formula:

$$z = \frac{(\bar{x} - \mu)}{\frac{\sigma}{\sqrt{n}}}$$

where:

z = standardized score

\bar{x} = sample mean

μ = the population mean under the null hypothesis

σ = population standard deviation. If we do not have the population standard deviation and if $n \geq 30$, we can use the sample standard deviation, s . If $n < 30$ and we do not have the population sample standard deviation we use a different distribution which will be discussed in a future lesson.

Once we calculate the z score, we can make a decision about whether to reject or to fail to reject the null hypothesis based on the critical values.

Following are the steps you must take when doing an hypothesis test:

1. Determine the null and alternative hypotheses.
2. Verify that necessary conditions are satisfied and summarize the data into a test statistic.
3. Determine the α level.
4. Determine the critical region(s).
5. Make a decision (Reject or fail to reject the null hypothesis)
6. Interpret the decision in the context of the problem.

Example: College A has an average SAT score of 1500. From a random sample of 125 freshman psychology students we find the average SAT score to be 1450 with a standard deviation of 100. We want to know if these freshman psychology students are representative of the overall population. What are our hypotheses and the test statistic?

1. Let's first develop our null and alternative hypotheses:

$$H_0 : \mu = 1500$$

$$H_a : \mu \neq 1500$$

2. The test statistic is $z = \frac{(\bar{x} - \mu)}{\frac{\sigma}{\sqrt{n}}} = \frac{(1450 - 1500)}{\frac{100}{\sqrt{125}}} \approx -5.59$

3. Choose $\alpha = .05$

4. This is a two sided test. If we choose $\alpha = .05$, the critical values will be -1.96 and 1.96. (Use invNorm (.025, 0,1) and the symmetry of the normal distribution to determine these critical values) That is we will reject the null hypothesis if the value of our test statistic is less than -1.96 or greater than 1.96.

5. The value of the test statistic is -5.59. This is less than -1.96 and so our decision is to reject H_0 .

6. Based on this sample we believe that the mean is not equal to 1500.

Example: A farmer is trying out a planting technique that he hopes will increase the yield on his pea plants. Over the last 5 years the average number of pods on one of his pea plants was 145 pods with a standard deviation of 100 pods. This year, after trying his new planting technique, he takes a random sample of 144 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:

$$H_0 : \mu = 145$$

$$H_a : \mu > 145$$

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. (Use `invNorm (.95, 0, 1)` to determine this critical value) 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 accept H_0 .

6. Based on our sample we believe the mean is equal to 145.

Finding the P-Value of an Event

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: 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, the null hypothesis). 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 if the null hypothesis is true? Using the calculator (`normcdf (.24, 99999999, 0, 1)`) we find this probability to be .405. This indicates that there is a 40.5% chance that under the null hypothesis the peas will produce 147 or more pods.

Type I and Type II Errors

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

- The null hypothesis is true and we reject it.
- The null hypothesis is true and we do not reject it.
- The null hypothesis is false and we do not reject it.
- The null hypothesis is false and we reject it.

Two of these four possible scenarios lead to correct decisions: accepting the null hypothesis when it is true and rejecting the null hypothesis when it is false.

Two of these four possible scenarios lead to errors: rejecting the null hypothesis when it is true and accepting the null hypothesis when it is false.

Which type of error is more serious depends on the specific research situation, but ideally both types of errors should be minimized during the analysis.

TABLE 8.1: Below is a table outlining the possible outcomes in hypothesis testing:

	H_0 is true	H_0 is false
Accept H_0	Good Decision	Error (type II)
Reject H_0	Error (type I)	Good Decision

The general approach to hypothesis testing focuses on the Type I error: rejecting the null hypothesis when it may be true. The level of significance, also known as the alpha level, is defined as the probability of making a Type I error

when testing a null hypothesis. For example, at the 0.05 level, we know that the decision to reject the hypothesis may be incorrect 5 percent of the time.

$$\alpha = P(\text{rejecting } H_0 | H_0 \text{ is true}) = P(\text{making a type I error})$$

Calculating the probability of making a Type II error is not as straightforward as calculating the probability of making a Type I error. The probability of making a Type II error can only be determined when values have been specified for the alternative hypothesis. The probability of making a type II error is denoted by β .

$$\beta = P(\text{accepting } H_0 | H_0 \text{ is false}) = P(\text{making a type II error})$$

Once the value for the alternative hypothesis has been specified, it is possible to determine the probability of making a correct decision ($1 - \beta$). This quantity, $1 - \beta$, is called the power of the test.

The goal in hypothesis testing is to minimize the potential of both Type I and Type II errors. However, there is a relationship between these two types of errors. As the level of significance or alpha level increases, the probability of making a Type II error (β) decreases and vice versa.

On the Web

<http://tinyurl.com/35zg7du> This link leads you to a graphical explanation of the relationship between α and β

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. 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. We can also attempt minimize the Type II errors by setting higher alpha levels in situations that do not have grave or costly consequences.

Calculating the Power of a Test

The power of a test is defined as the probability of rejecting the null hypothesis when it is false (that is, making the correct decision). Obviously, we want to maximize this power if we are concerned about making Type II errors. To determine the power of the test, there must be a specified value for the alternative hypothesis.

Example: Suppose that a doctor is concerned about making a Type II error only if the active ingredient in the new medication is greater than 3 milligrams higher than what was specified in the null hypothesis (say, 250 milligrams with a sample of 200 and a standard deviation of 50). Now we have values for both the null and the alternative hypotheses.

$$H_0 : \mu = 250$$

$$H_a : \mu = 253$$

By specifying a value for the alternative hypothesis, we have selected one of the many values for H_a . In determining the power of the test, we must assume that H_a is true and determine whether we would correctly reject the null hypothesis

Calculating the exact value for the power of the test requires determining the area above the critical value set up to test the null hypothesis when it is re-centered around the alternative hypothesis. If we have an alpha level of .05 our critical value would be 1.645 for the one tailed test. Therefore,

$$z = \frac{(\bar{x} - \mu)}{\frac{\sigma}{\sqrt{n}}}$$

$$1.645 = \frac{(\bar{x} - 250)}{\frac{50}{\sqrt{200}}}$$

Solving for \bar{x} we find: $\bar{x} = 1.645 \left(\frac{50}{\sqrt{200}} \right) + 250 \approx 255.8$

Now, with a new mean set at the alternative hypothesis $H_a : \mu = 253$ we want to find the value of the critical score when centered around this score when we center this \bar{x} around the population mean of the alternative hypothesis, $\mu = 253$. Therefore, we can figure that:

$$z = \frac{(\bar{x} - \mu)}{\frac{\sigma}{\sqrt{n}}} = \frac{(255.8 - 253)}{\frac{50}{\sqrt{200}}} \approx 0.79$$

Recall that we reject the null hypothesis if the critical value is to the right of .79. The question now is what is the probability of rejecting the null hypothesis when, in fact, the alternative hypothesis is true? We need to find the area to the right of 0.79. You can find this area using a z table or using the calculator with the Normcdf command (Invnorm (0.79, 9999999, 0, 1)). The probability is .2148. This means that since we assumed the alternative hypothesis to be true, there is only a 21.5% chance of rejecting the null hypothesis. Thus, the power of the test is .2148. In other words, this test of the null hypothesis is not very powerful and has only a 0.2148 probability of detecting the real difference between the two hypothesized means.

There are several things that affect the power of a test including:

- Whether the alternative hypothesis is a single-tailed or two-tailed test.
- The level of significance α
- The sample size.

On the Web

<http://intuitor.com/statistics/CurveApplet.html> Experiment with changing the sample size and the distance between the null and alternate hypotheses and discover what happens to the power.

Lesson Summary

Hypothesis testing involves making a conjecture about a population based on a sample drawn from the population.

We establish critical regions based on level of significance or alpha (α) level. If the value of the test statistic falls in these critical regions, we make the decision to reject the null hypothesis.

To evaluate the sample mean against the hypothesized population mean, we use the concept of z -scores to determine how different the two means are.

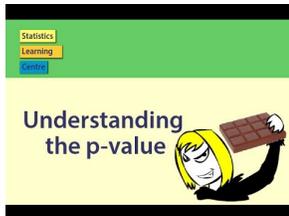
When we make a decision about a hypothesis, there are four different outcome and possibilities and two different types of errors. A Type I error is when we reject the null hypothesis when it is true and a Type II error is when we do not reject the null hypothesis, even when it is false. α , the level of significance of the test, is the probability of rejecting the null hypothesis when, in fact, the null hypothesis is true (an error).

The power of a test is defined as the probability of rejecting the null hypothesis when it is false (in other words, making the correct decision). We determine the power of a test by assigning a value to the alternative hypothesis and

using the z -score to calculate the probability of rejecting the null hypothesis when it is false. It is the probability of making a Type II error.

Multimedia Links

For an illustration of the use of the p -value in statistics (4.0) and how to interpret it (18.0), see [UCMSCI, Understanding the P-Value](#) (4:04)



MEDIA

Click image to the left or use the URL below.

URL: <http://www.ck12.org/flx/render/embeddedobject/21645>

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 the 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. Use the z -distribution table to determine the critical value for a single-tailed hypothesis test with a 0.01 significance level.
5. Sacramento County high school seniors have an average SAT score of 1020. From a random sample of 144 Sacramento High School students we find the average SAT score to be 1100 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 and the test statistic?
6. During hypothesis testing, we use the p -value to predict the ___ of an event occurring if the null hypothesis is true.
7. A survey shows that California teenagers have an average of \$500 in savings (standard error = 100). What is the probability that a randomly selected teenager will have savings greater than \$520?
8. Fill in the types of errors missing from the table below:

TABLE 8.2:

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

9. The ___ is defined as the probability of rejecting the null hypothesis when it is false (making the correct decision). We want to maximize___if we are concerned about making Type II errors.
10. The Governor's economic committee is investigating average salaries of recent college graduates in California. They decide to test the null hypothesis that the average salary is \$24,500 (standard deviation is \$4,800)) and is concerned with making a Type II error only if the average salary is less than \$25,000. $H_a : \mu = \$25,100$ For an $\alpha = .05$ and a sample of 144 determine the power of a one-tailed test.

8.2 Testing a Proportion Hypothesis

Learning Objectives

- Test a hypothesis about a population proportion by applying the binomial distribution approximation.
- Test a hypothesis about a population proportion using the P -value.

Introduction

In the previous section we studied the test statistic that is used when you are testing hypotheses about the mean of a population and you have a large sample (> 30).

Often statisticians are interested in making inferences about a population proportion. For example, when we look at election results we often look at the proportion of people that vote and who this proportion of voters choose. Typically, we call these proportions percentages and we would say something like “Approximately 68 percent of the population voted in this election and 48 percent of these voters voted for Barack Obama.”

So how do we test hypotheses about proportions? We use the same process as we did when testing hypotheses about populations but we must include sample proportions as part of the analysis. This lesson will address how we investigate hypotheses around population proportions and how to construct confidence intervals around our results.

Hypothesis Testing about Population Proportions by Applying the Binomial Distribution Approximation

We could perform tests of population proportions to answer the following questions:

- What percentage of graduating seniors will attend a 4-year college?
- What proportion of voters will vote for John McCain?
- What percentage of people will choose Diet Pepsi over Diet Coke?

To test questions like these, we make hypotheses about population proportions. For example,

H_0 : 35% of graduating seniors will attend a 4-year college.

H_0 : 42% of voters will vote for John McCain.

H_0 : 26% of people will choose Diet Pepsi over Diet Coke.

To test these hypotheses we follow a series of steps:

- Hypothesize a value for the population proportion P like we did above.
- Randomly select a sample.
- Use the sample proportion \hat{p} to test the stated hypothesis.

To determine the test statistic we need to know the sampling distribution of the sample proportion. We use the binomial distribution which illustrates situations in which two outcomes are possible (for example, voted for a

candidate, didn't vote for a candidate), remembering that when the sample size is relatively large, we can use the normal distribution to approximate the binomial distribution. The test statistic is

$$z = \frac{\text{sample estimate} - \text{value under the null hypothesis}}{\text{standard error under the null hypothesis}}$$

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

where:

p_0 is the hypothesized value of the proportion under the null hypothesis

n is the sample size

Example: We want to test a hypothesis that 60 percent of the 400 seniors graduating from a certain California high school will enroll in a two or four-year college upon graduation. What would be our hypotheses and the test statistic?

Since we want to test the proportion of graduating seniors and we think that proportion is around 60 percent, our hypotheses are:

$$H_0 : p = .6$$

$$H_a : p \neq .6$$

The test statistic would be $z = \frac{\hat{p} - .6}{\sqrt{\frac{.6(1-.6)}{n}}}$. To complete this calculation we would have to have a value for the sample size (n).

Testing a Proportion Hypothesis

Similar to testing hypotheses dealing with population means, we use a similar set of steps when testing proportion hypotheses.

- Determine and state the null and alternative hypotheses.
- Set the criterion for rejecting the null hypothesis.
- Calculate the test statistic.
- Decide whether to reject or fail to reject the null hypothesis.
- Interpret your decision within the context of the problem.

Example: A congressman is trying to decide on whether to vote for a bill that would legalize gay marriage. He will decide to vote for the bill only if 70 percent of his constituents favor the bill. In a survey of 300 randomly selected voters, 224 (74.6%) indicated that they would favor the bill. Should he or should he not vote for the bill?

First, we develop our null and alternative hypotheses.

$$H_0 : p = .7$$

$$H_a : p > .7$$

Next, we should set the criterion for rejecting the null hypothesis. Choose $\alpha = .05$ and since the null hypothesis is considering $p > .7$, this is a one tailed test. Using a standard z table or the TI 83/84 calculator we find the critical value for a one tailed test at an alpha level of .05 to be 1.645.

The test statistic is $z = \frac{.74-.7}{\sqrt{\frac{.7(1-.7)}{300}}} \approx 1.51$

Since our critical value is 1.645 and our test statistic is 1.51, we cannot reject the null hypothesis. This means that we cannot conclude that the population proportion is greater than .70 with 95 percent certainty. Given this information, it is not safe to conclude that at least 70 percent of the voters would favor this bill with any degree of certainty. Even though the proportion of voters supporting the bill is over 70 percent, this could be due to chance and is not statistically significant.

Example: Admission staff from a local university is conducting a survey to determine the proportion of incoming freshman that will need financial aid. A survey on housing needs, financial aid and academic interests is collected from 400 of the incoming freshman. Staff hypothesized that 30 percent of freshman will need financial aid and the sample from the survey indicated that 101 (25.3%) would need financial aid. Is this an accurate guess?

First, we develop our null and alternative hypotheses.

$$H_0 : p = .3$$

$$H_a : p \neq .3$$

Next, we should set the criterion for rejecting the null hypothesis. The .05 alpha level is used and for a two tailed test the critical values of the test statistic are 1.96 and -1.96.

To calculate the test statistic:

$$z = \frac{.25 - .3}{\sqrt{\frac{.3(1-.3)}{400}}} \approx -2.18$$

Since our critical values are ± 1.96 and $-2.18 < -1.96$ we can reject the null hypothesis. This means that we can conclude that the population of freshman needing financial aid is significantly more or less than 30 percent. Since the test statistic is negative, we can conclude with 95% certainty that in the population of incoming freshman, less than 30 percent of the students will need financial aid.

Lesson Summary

In statistics, we also make inferences about proportions of a population. We use the same process as in testing hypotheses about populations but we must include hypotheses about proportions and the proportions of the sample in the analysis. To calculate the test statistic needed to evaluate the population proportion hypothesis, we must also

calculate the standard error of the proportion which is defined as $s_p = \sqrt{\frac{p_0(1-p_0)}{n}}$

The formula for calculating the test statistic for a population proportion is

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

where:

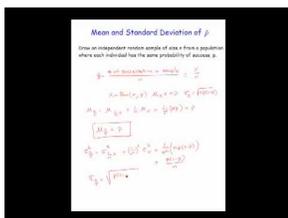
\hat{p} is the sample proportion

p_0 is the hypothesized population proportion

We can construct something called the confidence interval that specifies the level of confidence that we have in our results. The formula for constructing a confidence interval for the population proportion is $\hat{p} \pm z_{\frac{\alpha}{2}} \left(\frac{\hat{p}(1-\hat{p})}{n} \right)$.

Multimedia Links

For an explanation on finding the mean and standard deviation of a sampling proportion, p , and normal approximation to binomials **(7.0)(9.0)(15.0)(16.0)**, see [American Public University, Sampling Distribution of Sample Proportion](#) (8:24)



MEDIA

Click image to the left or use the URL below.

URL: <http://www.ck12.org/flx/render/embeddedobject/1100>

For a calculation of the z-statistic and associated P-Value for a 1-proportion test **(18.0)**, see [kbower50, Test of 1 Proportion: Worked Example](#) (3:51)



MEDIA

Click image to the left or use the URL below.

URL: <http://www.ck12.org/flx/render/embeddedobject/1101>

Review Questions

- The test statistic helps us determine ____.
- True or false: In statistics, we are able to study and make inferences about proportions, or percentages, of a population.
- A state senator cannot decide how to vote on an environmental protection bill. The senator decides to request her own survey and if the proportion of registered voters supporting the bill exceeds 0.60, she will vote for it. A random sample of 750 voters is selected and 495 are found to support the bill.
 - What are the null and alternative hypotheses for this problem?
 - What is the observed value of the sample proportion?
 - What is the standard error of the proportion?
 - What is the test statistic for this scenario?
 - What decision would you make about the null hypothesis if you had an alpha level of .01?

8.3 Testing a Mean Hypothesis

Evaluating Hypotheses for Population Means using Large Samples

When testing a hypothesis for the mean of a normal distribution, we follow a series of four basic steps:

1. State the null and alternative hypotheses.
2. Choose an α level
3. Set the criterion (critical values) for rejecting the null hypothesis.
4. Compute the test statistic.
5. Make a decision (reject or fail to reject the null hypothesis)
6. 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.

Example: The school nurse was wondering if the average height of 7th graders has been increasing. Over the last 5 years, the average height of a 7th grader was 145 cm with a standard deviation of 20 cm. The school nurse takes a random sample of 200 students and finds that the average height this year is 147 cm. 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:

$$H_0 : \mu = 145$$

$$H_a : \mu > 145$$

Choose $\alpha = .05$. The critical value for this one tailed test is 1.64. 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.

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

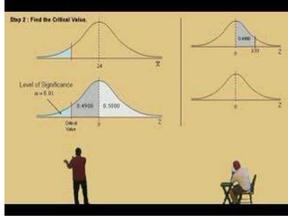
Since the calculated z -score of 1.414 is smaller than 1.64 and thus does not fall in the critical region. Our decision is to fail to reject the null hypothesis and conclude that the probability of obtaining a sample mean equal to 147 if the mean of the population is 145 is likely to have been due to chance.

When testing a hypothesis for the mean of a distribution, we follow a series of six basic steps:

1. State the null and alternative hypotheses.
2. Choose α
3. Set the criterion (critical values) for rejecting the null hypothesis.
4. Compute the test statistic.
5. Decide about the null hypothesis
6. Interpret our results.

Multimedia Links

For an step by step example of testing a mean hypothesis (4.0), see [MuchoMath, Z Test for the Mean](#) (9:34).



MEDIA

Click image to the left or use the URL below.

URL: <http://www.ck12.org/flx/render/embeddedobject/1102>

Review Questions

- 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.
- 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.
 - What would the null and alternative hypotheses be for this scenario?
 - What would the standard error be for this particular scenario?
 - Describe in your own words how you would set the critical regions and what they would be at an alpha level of .05.
 - Test the null hypothesis and explain your decision
- For each of the following scenarios, state which one is more likely to lead to the rejection of the null hypothesis?
 - A one-tailed or two-tailed test
 - .05 or .01 level of significance
 - A sample size of $n = 144$ or $n = 444$

8.4 Student's t -Distribution

Learning Objectives

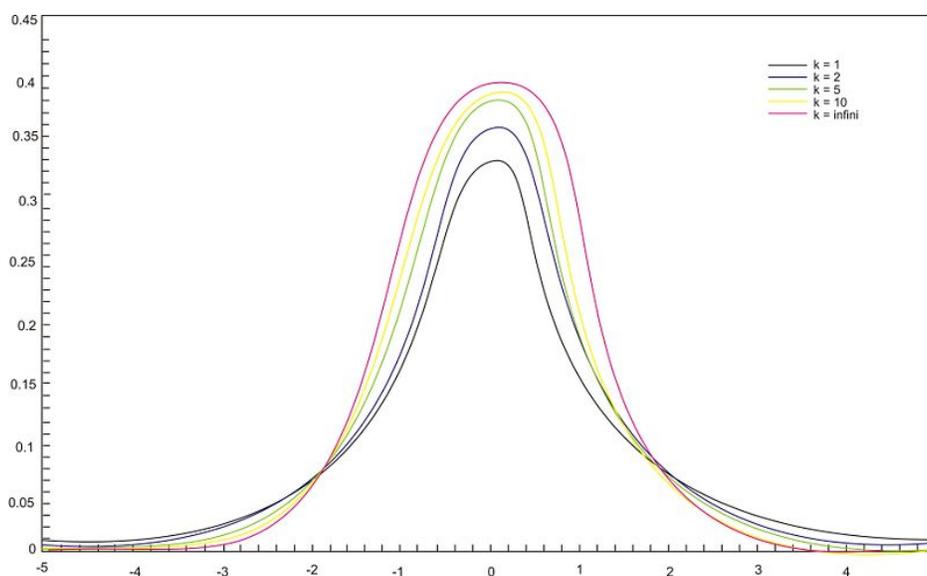
- Use Student's t -distribution to estimate population mean interval for smaller samples.
- Understand how the shape of Student's t -distribution corresponds to the sample size (which corresponds to a measure called the "degrees of freedom.")

Introduction

Hypothesis Testing with Small Populations and Sample Sizes

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."

T -distributions are a family of distributions that, like the normal distribution, are symmetrical and bell-shaped and centered on a mean. However, the distribution shape changes as the sample size changes. Therefore, there is a specific shape or distribution for every sample of a given size (see figure below; each distribution has a different value of k , the number of degrees of freedom, which is 1 less than the size of the sample).



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 link below) represents a different t -distribution and 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.

<http://tinyurl.com/ygcc5g9> Follow this link to the Student's t -table.

As the number of observations gets larger, the t -distribution approaches the shape of the normal distribution. In general, once the sample size is large enough - usually about 30 - we would use the normal distribution or the z -table instead. Note that usually in practice, if the standard deviation is known then the normal distribution is used regardless of the sample size.

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

Example: 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 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 .05 significance level.

First, we establish our null and alternative hypotheses.

$$H_0 : \mu = 3.10$$

$$H_a : \mu \neq 3.10$$

Next, we use our alpha level of .05 and the t -distribution table to find our critical values. For a two-tailed test with 19 degrees of freedom and a .05 level of significance, our critical values are equal to ± 2.093 .

In calculating the test statistic, we use the formula:

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{3.18 - 3.10}{\frac{.54}{\sqrt{20}}} \approx 0.66$$

This means that the observed sample mean 3.18 of football players is .66 standard errors above the hypothesized value of 3.10. Because the value of the test statistic is less than the critical value of 2.093, we fail to reject the null hypothesis.

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: The masses of newly produced bus tokens are estimated to have a mean of 3.16 grams. A random sample of 11 tokens was removed from the production line and the mean weight of the tokens was calculated as 3.21 grams with a standard deviation of 0.067. What is the value of the test statistic for a test to determine how the mean differs from the estimated mean?

Solution:

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

$$t = \frac{3.21 - 3.16}{\frac{0.067}{\sqrt{11}}}$$

$$t \approx 2.48$$

If the value of t from the sample fits right into the middle of the distribution of t constructed by assuming the null hypothesis is true, the null hypothesis is true. On the other hand, if the value of t from the sample is way out in the tail of the t -distribution, then there is evidence to reject the null hypothesis. Now that the distribution of t is known when the null hypothesis is true, the location of this value on the distribution. The most common method used to determine this is to find a p -value (observed significance level). The p -value is a probability that is computed with the assumption that the null hypothesis is true.

The p -value for a two-sided test is the area under the t -distribution with $df = 11 - 1 = 10$ that lies above $t = 2.48$ and below $t = -2.48$. This p -value can be calculated by using technology.

Technology Note: Using the `tcdf` command to calculate probabilities associated with the t distribution

Press **2ND** [**DIST**] Use \downarrow to select 5.tcdf (lower bound, upper bound, degrees of freedom) This will be the total area under both tails. To calculate the area under one tail divide by 2.

```

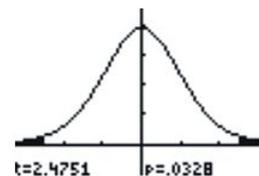
DISG DRAW
1:normalPdf(
2:normalcdf(
3:invNorm(
4:tpdf(
5:tcdf(
6:X²Pdf(
7↓X²cdf(

```

```

1-tcdf(-2.48,2.4
8,10)
      .0325428246
Ans/2
      .0162714123

```



There is only a .016 chance of getting an absolute value of t as large as or even larger than the one from this sample. The small p -value tells us that the sample is inconsistent with the null hypothesis. The population mean differs from the estimated mean of 3.16.

When the p -value is close to zero, there is strong evidence against the null hypothesis. When the p -value is large, the result from the sample is consistent with the estimated or hypothesized mean and there is no evidence against the null hypothesis.

A visual picture of the P -value can be obtained by using the graphing calculator.



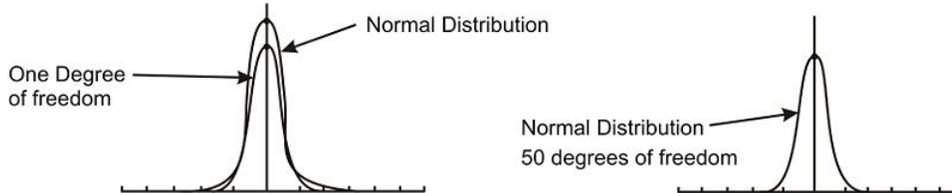
Enter

```
T-Test
Inpt:Data  Stats
μ₀:3.16
x̄:3.21
Sx: .067
n:11
μ:μ₀ <μ₀ >μ₀
Calculate Draw
```

Highlight Draw Enter



The spread of any t distribution is greater than that of a standard normal distribution. This is due to the fact that that in the denominator of the formula σ has been replaced with s . Since s is a random quantity changing with various samples, the variability in t is greater, resulting in a larger spread.



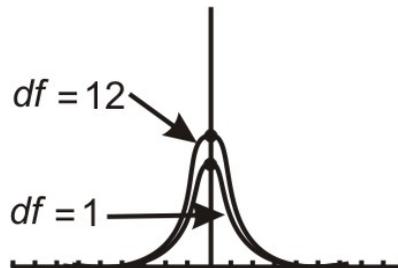
Notice in the first distribution graph the spread of the first (inner curve) is small but in the second one the both distributions are basically overlapping, so are roughly normal. This is due to the increase in the degrees of freedom.

Here are the t -distributions for $df = 1$ and for $df = 12$ as graphed on the graphing calculator

$$Y = 2^{\text{nd}} \text{ [Vars] (Dist) } \downarrow 4. \text{ tpdf(}$$

You are now on the $Y =$ screen.

$Y = \text{tpdf}(X, 1)$ [Graph]



Repeat the steps to plot more than one t -distribution on the same screen.

Notice the difference in the two distributions.

The one with 12 degrees of freedom approximates a normal curve.

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 the following conditions:

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

The t -distribution also has some unique properties. These properties are:

- The mean of the distribution equals zero.
- The population standard deviation is unknown.
- The variance is equal to the degrees of freedom divided by the degrees of freedom minus 2. This means that the degrees of freedom must be greater than two to avoid the expression being undefined.
- The variance is always greater than one, although it approaches 1 as the degrees of freedom increase. This is due to the fact that as the degrees of freedom increase, the distribution is becoming more of a normal distribution.
- Although the Student t -distribution is bell-shaped, the smaller sample sizes produce a flatter curve. The distribution is not as mounded as a normal distribution and the tails are thicker. As the sample size increases and approaches 30, the distribution approaches a normal distribution.
- The population is unimodal and symmetric.

Example: Duracell manufactures batteries that the CEO claims will last 300 hours under normal use. A researcher randomly selected 15 batteries from the production line and tested these batteries. The tested batteries had a mean life span of 290 hours with a standard deviation of 50 hours. If the CEO's claim were true, what is the probability that 15 randomly selected batteries would have a life span of no more than 290 hours?

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \text{ The degrees of freedom are } (n - 1) = 15 - 1. \text{ This means 14 degrees of freedom.}$$

$$t = \frac{290 - 300}{\frac{50}{\sqrt{15}}}$$

$$t = \frac{-10}{12.9099}$$

$$t = -.7745993$$

Using the graphing calculator or a table of values, the cumulative probability is 0.226, which means that if the true life span of a battery were 300 hours, there is a 22.6% chance that the life span of the 15 tested batteries would be less than or equal to 290 days. This is not a high enough level of confidence to reject the null hypothesis and count the discrepancy as significant.

$$Y = 2^{\text{nd}} \boxed{\text{Vars}} (\text{Dist}) \downarrow 4. \text{tcdf}($$

You are now on the $Y =$ screen.

$$Y = \text{tcdf}(-1E99, -.7745993, 14) = [0.226]$$

Example: 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?

- State the hypotheses.
- Calculate the test statistic.
- Find and interpret the p-value.
- Would the null hypothesis be rejected at the 10% level? The 5% level? The 1% level?

Solution:

- For H_0 the mean weight of cheese $\mu = 72$; and for $H_a : \mu \neq 72$.
- 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. Using technology, the p value is .0262. If the mean weight of cheese is 72 ounces, the probability that the weight of 7 random measurements would give a value of t greater than 2.9315 or less than -2.9315 is about 0.0262.

d) Because the p -value of 0.0262 is less than both .10 and .05, the null hypothesis would be rejected at these levels. However, the p -value is greater than .01 so the null hypothesis would not be rejected if this level of confidence was required.

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.

Points to Consider

- Is there a way to determine where the t -statistic lies on a distribution?
- If a way does exist, what is the meaning of its placement?

Multimedia Links

For an explanation of the T distribution and an example using it (7.0)(17.0), see [bionicturtle.com](http://www.bionicturtle.com), [Student's t distribution](#) (8:32).

13	Sample Mean	0.02%
14	Sample Std Dev	1.54%
15	Days (n-1)	10
16		
17	Confidence	95%
18	Significance (1-)	5%
19		
20	Critical t	2.262
21		
22	Lower limit	-1.08%
23	Upper limit	1.12%
24		
25		
26		

$\bar{X} \pm t_{\alpha} \frac{s}{\sqrt{n}}$

MEDIA

Click image to the left or use the URL below.

URL: <http://www.ck12.org/flx/render/embeddedobject/1103>

Review Questions

1. In hypothesis testing, when we work with large samples, we use the ___ distribution. When working with small samples (typically samples under 30), we use the ___ distribution.

2. You intend to use simulation to construct an approximate t -distribution with 8 degrees of freedom by taking random samples from a population with bowling scores that are normally distributed with mean, $\mu = 110$ and standard deviation, $\sigma = 20$.
 - a. Explain how you will do one run of this simulation.
 - b. Produce four values of t using this simulation.
3. The dean from UCLA is concerned that the students' 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 30 seniors from the last graduating class and finds that their mean GP is 2.85 with a sample standard deviation of 0.65. Suppose that the dean samples only 30 students. Would a t -distribution now be the appropriate sampling distribution for the mean? Why or why not?
4. Using the appropriate t -distribution, test the same null hypothesis with a sample of 30.
5. With a sample size of 30, do you need to have a larger or smaller difference between the hypothesized population mean and the sample mean to obtain statistical significance than with a sample size of 256? Explain your answer.

8.5 Testing a Hypothesis for Dependent and Independent Samples

Learning Objectives

- Identify situations that contain dependent or independent samples.
- Calculate the pooled standard deviation for two independent samples.
- Calculate the test statistic to test hypotheses about dependent data pairs.
- Calculate the test statistic to test hypotheses about independent data pairs for both large and small samples.
- Calculate the test statistic to test hypotheses about the difference of proportions between two independent samples.

Introduction

In the previous lessons we learned about hypothesis testing for proportion and means with both large and small samples. However, in the examples in those lessons only one sample was involved. In this lesson we will apply the principals of hypothesis testing to situations involving two samples. There are many situations in everyday life where we would perform statistical analysis involving two samples. For example, suppose that we wanted to test a hypothesis about the effect of two medications on curing an illness. Or we may want to test the difference between the means of males and females on the SAT. In both of these cases, we would analyze both samples and the hypothesis would address the difference between two sample means.

In this lesson, we will identify situations with different types of samples, learn to calculate the test statistic, calculate the estimate for population variance for both samples and calculate the test statistic to test hypotheses about the difference of proportions or means between samples.

Dependent and Independent Samples

When we are working with one sample, we know that we need to select a random sample from the population, measure that sample statistic and then make hypothesis about the population based on that sample. When we work with two independent samples we assume that if the samples are selected at random (or, in the case of medical research, the subjects are randomly assigned to a group), the two samples will vary only by chance and the difference will not be statistically significant. In short, when we have independent samples we assume that the scores of one sample do not affect the other.

Independent samples can occur in two scenarios.

Testing the difference of the means between two fixed populations we test the differences between samples from each population. When both samples are randomly selected, we can make inferences about the populations.

When working with subjects (people, pets, etc.), if we select a random sample and then randomly assign half of the subjects to one group and half to another we can make inferences about the populations.

Dependent samples are a bit different. Two samples of data are dependent when each score in one sample is paired with a specific score in the other sample. In short, these types of samples are related to each other. Dependent samples can occur in two scenarios. In one, a group may be measured twice such as in a pretest-posttest situation

(scores on a test before and after the lesson). The other scenario is one in which an observation in one sample is matched with an observation in the second sample.

To distinguish between tests of hypotheses for independent and dependent samples, we use a different symbol for hypotheses with dependent samples. For dependent sample hypotheses, we use the delta symbol δ to symbolize the difference between the two samples. Therefore, in our null hypothesis we state that the difference of scores across the two measurements is equal to 0; $\delta = 0$ or:

$$H_0 : \delta = \mu_1 - \mu_2$$

Calculating the Pooled Estimate of Population Variance

When testing a hypothesis about two independent samples, we follow a similar process as when testing one random sample. However, when computing the test statistic, we need to calculate the estimated standard error of the

difference between sample means, $s_{\bar{x}_1 - \bar{x}_2} = \sqrt{s^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$.

Where n_1 and n_2 are the sizes of the two samples s^2 is the pooled sample variance, which is computed as $s^2 = \frac{\sum(x_1 - \bar{x}_1)^2 + \sum(x_2 - \bar{x}_2)^2}{n_1 + n_2 - 2}$. Often, the top part of this formula is simplified by substituting the symbol SS for the sum of the squared deviations. Therefore, the formula often is expressed by $s^2 = \frac{SS_1 + SS_2}{n_1 + n_2 - 2}$

Example: Calculating s^2 Suppose we have two independent samples of student reading scores.

The data are as follows:

TABLE 8.3:

Sample 1	Sample 2
7	12
8	14
10	18
4	13
6	11
	10

From this sample, we can calculate a number of descriptive statistics that will help us solve for the pooled estimate of variance:

TABLE 8.4:

Descriptive Statistic	Sample 1	Sample 2
Number n	5	6
Sum of Observations $\sum x$	35	78
Mean of Observations \bar{x}	7	13
Sum of Squared Deviations $\sum_{i=1}^n (x_i - \bar{x})^2$	20	40

Using the formula for the pooled estimate of variance, we find that

$$s^2 = 6.67$$

We will use this information to calculate the test statistic needed to evaluate the hypotheses.

Testing Hypotheses with Independent Samples

When testing hypotheses with two independent samples, we follow similar steps as when testing one random sample:

- State the null and alternative hypotheses.
- Choose α
- Set the criterion (critical values) for rejecting the null hypothesis.
- Compute the test statistic.
- Make a decision: reject or fail to reject the null hypothesis.
- Interpret the decision within the context of the problem.

When stating the null hypothesis, we assume there is no difference between the means of the two independent samples. Therefore, our null hypothesis in this case would be:

$$H_0 : \mu_1 = \mu_2 \text{ or } H_0 : \mu_1 - \mu_2 = 0$$

Similar to the one-sample test, the critical values that we set to evaluate these hypotheses depend on our alpha level and our decision regarding the null hypothesis is carried out in the same manner. However, since we have two samples, we calculate the test statistic a bit differently and use the formula:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s.e.(\bar{x}_1 - \bar{x}_2)}$$

where:

$\bar{x}_1 - \bar{x}_2$ is the difference between the sample means

$\mu_1 - \mu_2$ is the difference between the hypothesized population means

$s.e.(\bar{x}_1 - \bar{x}_2)$ is the standard error of the difference between sample means

Example: The head of the English department is interested in the difference in writing scores between remedial freshman English students who are taught by different teachers. The incoming freshmen needing remedial services are randomly assigned to one of two English teachers and are given a standardized writing test after the first semester. We take a sample of eight students from one class and nine from the other. Is there a difference in achievement on the writing test between the two classes? Use a 0.05 significance level.

First, we would generate our hypotheses based on the two samples.

$$H_0 : \mu_1 = \mu_2$$

$$H_0 : \mu_1 \neq \mu_2$$

This is a two tailed test. For this example, we have two independent samples from the population and have a total of 17 students that we are examining. Since our sample size is so low, we use the t -distribution. In this example, we have 15 degrees of freedom (number in the samples minus 2) and with a .05 significance level and the t distribution, we find that our critical values are 2.131 standard scores above and below the mean.

To calculate the test statistic, we first need to find the pooled estimate of variance from our sample. The data from the two groups are as follows:

TABLE 8.5:

Sample 1	Sample 2
35	52
51	87
66	76
42	62
37	81
46	71
60	55
55	67
53	

From this sample, we can calculate several descriptive statistics that will help us solve for the pooled estimate of variance:

TABLE 8.6:

Descriptive Statistic	Sample 1	Sample 2
Number n	9	8
Sum of Observations $\sum x$	445	551
Mean of Observations \bar{x}	49.44	68.875
Sum of Squared Deviations $\sum_{i=1}^n (x_i - \bar{x})^2$	862.22	1058.88

Therefore:

$$s^2 = \frac{SS_1 + SS_2}{n_1 + n_2 - 2} = 128.07$$

and the standard error of the difference of the sample means is:

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{s^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} = \sqrt{128.07 \left(\frac{1}{9} + \frac{1}{8} \right)} \approx 5.50$$

Using this information, we can finally solve for the test statistic:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s.e.(\bar{x}_1 - \bar{x}_2)} = \frac{(49.44 - 68.875) - (0)}{5.50} \approx -3.53$$

Since -3.53 is less than the critical value of 2.13, we decide to reject the null hypothesis and conclude there is a significant difference in the achievement of the students assigned to different teachers.

Testing Hypotheses about the Difference in Proportions between Two Independent Samples

Suppose we want to test if there is a difference between proportions of two independent samples. As discussed in the previous lesson, proportions are used extensively in polling and surveys, especially by people trying to predict election results. It is possible to test a hypothesis about the proportions of two independent samples by using a similar method as described above. We might perform these hypotheses tests in the following scenarios:

- When examining the proportion of children living in poverty in two different towns.
- When investigating the proportions of freshman and sophomore students who report test anxiety.
- When testing if the proportion of high school boys and girls who smoke cigarettes is equal.

In testing hypotheses about the difference in proportions of two independent samples, we state the hypotheses and set the criterion for rejecting the null hypothesis in similar ways as the other hypotheses tests. In these types of tests we set the proportions of the samples equal to each other in the null hypothesis $H_0 : p_1 = p_2$ and use the appropriate standard table to determine the critical values (remember, for small samples we generally use the t distribution and for samples over 30 we generally use the z -distribution).

When solving for the test statistic in large samples, we use the formula:

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{se(p_1 - p_2)}$$

where:

\hat{p}_1, \hat{p}_2 are the observed sample proportions

p_1, p_2 are the population proportions under the null hypothesis

$se(p_1 - p_2)$ is the standard error of the difference between independent proportions

Similar to the standard error of the difference between independent samples, we need to do a bit of work to calculate the standard error of the difference between independent proportions. To find the standard error under the null hypothesis we assume that $p_1 = p_2 = p$ and we use all the data to estimate p .

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$$

Now the standard error of the difference is $\sqrt{\hat{p}(1 - \hat{p}) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$

The test statistic is now $z = \frac{(\hat{p}_1 - \hat{p}_2) - (0)}{\sqrt{\hat{p}(1 - \hat{p}) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$

Example: Suppose that we are interested in finding out which particular city is more is more satisfied with the services provided by the city government. We take a survey and find the following results:

TABLE 8.7:

Number Satisfied	City 1	City 2
Yes	122	84
No	78	66
Sample Size	$n_1 = 200$	$n_2 = 150$
Proportion who said Yes	0.61	0.56

Is there a statistical difference in the proportions of citizens that are satisfied with the services provided by the city government? Use a 0.05 level of significance.

First, we establish the null and alternative hypotheses:

$$H_0 : p_1 = p_2$$

$$H_a : p_1 \neq p_2$$

Since we have a large sample size we will use the z -distribution. At a .05 level of significance, our critical values are ± 1.96 . To solve for the test statistic, we must first solve for the standard error of the difference between proportions.

$$\hat{p} = \frac{200(.61) + 150(.56)}{350} = .589$$

$$se(\hat{p}_1 - \hat{p}_2) = \sqrt{0.589(.411) \left(\frac{1}{200} + \frac{1}{150} \right)} \approx 0.053$$

Therefore, the test statistic is:

$$z = \frac{(0.61 - 0.56) - (0)}{0.053} \approx 0.94$$

Since 0.94 does not exceed the critical value 1.96, the null hypothesis is not rejected. Therefore, we can conclude that the difference in the probabilities could have occurred by chance and that there is no difference in the level of satisfaction between citizens of the two cities.

Testing Hypotheses with Dependent Samples

When testing a hypothesis about two dependent samples, we follow the same process as when testing one random sample or two independent samples:

- State the null and alternative hypotheses.
- Choose the level of significance
- Set the criterion (critical values) for rejecting the null hypothesis.
- Compute the test statistic.
- Make a decision, reject or fail to reject the null hypothesis
- Interpret our results.

As mentioned in the section above, our hypothesis for two dependent samples states that there is no difference between the scores across the two samples $H_0 : \delta = \mu_1 - \mu_2 = 0$. We set the criterion for evaluating the hypothesis in the same way that we do with our other examples –by first establishing an alpha level and then finding the critical values by using the t -distribution table. Calculating the test statistic for dependent samples is a bit different since we are dealing with two sets of data. The test statistic that we first need calculate is \bar{d} , which is the difference in the means of the two samples. Therefore, $\bar{d} = \bar{x}_1 - \bar{x}_2$. We also need to know the standard error of the difference between the two samples. Since our population variance is unknown, we estimate it by first using the formula for the standard deviations of the samples:

$$s_d^2 = \frac{\sum(d - \bar{d})^2}{n - 1}$$

$$s_d = \sqrt{\frac{\sum d^2 - \frac{(\sum d)^2}{n}}{n - 1}}$$

where:

s_d^2 is the sample variance

d is the difference between corresponding pairs within the sample

\bar{d} is the difference between the means of the two samples

n is the number in the sample

s_d is the standard deviation

With the standard deviation, we can calculate the standard error using the following formula:

$$s_{\bar{d}} = \frac{s_d}{\sqrt{n}}$$

After we calculate the standard error, we can use the general formula for the test statistic:

$$t = \frac{\bar{d} - \delta}{s_d}$$

Example: The math teacher wants to determine the effectiveness of her statistics lesson and gives a pre-test and a post-test to 9 students in her class. Our hypothesis is that there is no difference between the means of the two samples and our alternative hypothesis is that the two means of the samples are not equal. In other words, we are testing whether or not these two samples are related or:

$$H_0 : \delta = \mu_1 - \mu_2 = 0$$

$$H_a : \delta = \mu_1 - \mu_2 \neq 0$$

The results for the pre-and post-tests are below:

TABLE 8.8:

Subject	Pre-test Score	Post-test Score	d difference	d^2
1	78	80	2	4
2	67	69	2	4
3	56	70	14	196
4	78	79	1	1
5	96	96	0	0
6	82	84	2	4
7	84	88	4	16
8	90	92	2	4
9	87	92	5	25
Sum	718	750	32	254
Mean	79.7	83.3	3.6	

Using the information from the table above, we can first solve for the standard deviation of the two samples, then the standard error of the two samples and finally the test statistic.

Standard Deviation:

$$s_d = \sqrt{\frac{\sum d^2 - \frac{(\sum d)^2}{n}}{n-1}} = \sqrt{\frac{254 - \frac{(32)^2}{9}}{8}} \approx 4.19$$

Standard Error of the Difference:

$$s_{\bar{d}} = \frac{s_d}{\sqrt{n}} = \frac{4.19}{\sqrt{9}} = 1.40$$

Test Statistic (t -Test)

$$t = \frac{\bar{d} - \delta}{s_{\bar{d}}} = \frac{3.6 - 0}{1.40} \approx 2.57$$

With 8 degrees of freedom (number of observations - 1) and a significance level of .05, we find our critical values to be ± 2.306 . Since our test statistic exceeds this critical value, we can reject the null hypothesis that the two samples are equal and conclude that the lesson had an effect on student achievement.

Lesson Summary

In addition to testing single samples associated with a mean, we can also perform hypothesis tests with two samples. We can test two independent samples (which are samples that do not affect one another) or dependent samples which assume that the samples are related to each other.

When testing a hypothesis about two independent samples, we follow a similar process as when testing one random sample. However, when computing the test statistic, we need to calculate the estimated standard error of the difference between sample means which is found by using the formula:

$$se(\bar{x}_1 - \bar{x}_2) = \sqrt{s^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \text{ with } s^2 = \frac{ss_1 + ss_2}{n_1 + n_2 - 2}$$

We carry out the test on the means of two independent samples in a similar way as the testing of one random sample. However, we use the following formula to calculate the test statistic:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s.e.(\bar{x}_1 - \bar{x}_2)} \text{ with the standard error defined above.}$$

We can also test the proportions associated with two independent samples. In order to calculate the test statistic associated with two independent samples, we use the formula:

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (0)}{\sqrt{\hat{p}(1 - \hat{p}) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \text{ with } \hat{p} = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2}$$

We can also test the likelihood that two dependent samples are related. To calculate the test statistic for two dependent samples, we use the formula:

$$t = \frac{\bar{d} - \delta}{s_d} \text{ with } s_d = \sqrt{\frac{\sum d^2 - \frac{(\sum d)^2}{n}}{n-1}}$$

Review Questions

1. In hypothesis testing, we have scenarios that have both dependent and independent samples. Give an example of an experiment with (1) dependent samples and (2) independent samples.
2. True or False: When we test the difference between the means of males and females on the SAT, we are using independent samples.
3. A study is conducted on the effectiveness of a drug on the hyperactivity of laboratory rats. Two random samples of rats are used for the study and one group is given Drug A and the other group is given Drug B and the number of times that they push a lever is recorded. The following results for this test were calculated:

TABLE 8.9:

	Drug A	Drug B
X	75.6	72.8
n	18	24
s^2	12.25	10.24
s	3.5	3.2

- (a) Does this scenario involve dependent or independent samples? Explain.
 - (b) What would the hypotheses be for this scenario?
 - (c) Compute the pooled estimate for population variance.
 - (d) Calculate the estimated standard error for this scenario.
 - (e) What is the test statistic and at an alpha level of .05 what conclusions would you make about the null hypothesis?
4. A survey is conducted on attitudes towards drinking. A random sample of eight married couples is selected, and the husbands and wives respond to an attitude-toward-drinking scale. The scores are as follows:

TABLE 8.10:

Husbands	Wives
16	15
20	18
10	13
15	10
8	12
19	16
14	11
15	12

- (a) What would be the hypotheses for this scenario?
- (b) Calculate the estimated standard deviation for this scenario.
- (c) Compute the standard error of the difference for these samples.
- (d) What is the test statistic and at an alpha level of .05 what conclusions would you make about the null hypothesis?

Keywords

Null hypothesis

Alternative hypothesis

One-tailed test

Two-tailed test

p -value

Power of a test

Level of significance

Critical region

Type I error

Type II error

α

β

Standard error

Dependent samples

t distribution