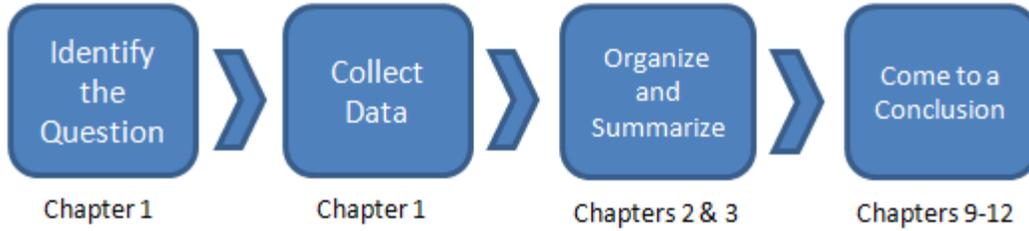


Chapter 10: Testing Claims Regarding a Parameter

- [10.1 The Language of Hypothesis Testing](#)
- [10.2 Hypothesis Tests for a Population Proportion](#)
- [10.3 Hypothesis Tests for a Population Mean](#)
- [10.4 Hypothesis Tests for a Population Standard Deviation](#)
- [10.5 Putting It Together: Which Method Do I Use?](#)

Let's take a minute to refresh the process of statistics we introduced earlier this semester.



Last chapter, we finally entered the "meat" of the course, where we use our knowledge of calculated statistics (like the sample mean and standard deviation) and our knowledge of probability (the normal distribution, et al) to come to conclusions - confidence intervals.

This chapter, we'll be expanding that knowledge and answering questions about population parameters based on sample data. For example:

- Do online students work more than traditional students?
- Are online students younger than the average ECC student?
- Do ECC students vote at a higher rate than the general population?

:: start ::



Section 10.1: The Language of Hypothesis Testing

10.1 The Language of Hypothesis Testing

[10.2 Hypothesis Tests for a Population Proportion](#)

[10.3 Hypothesis Tests for a Population Mean](#)

[10.4 Hypothesis Tests for a Population Standard Deviation](#)

[10.5 Putting It Together: Which Method Do I Use?](#)

Objectives

By the end of this lesson, you will be able to...

1. determine the null and alternative hypotheses from a claim
2. explain Type I and Type II errors
3. identify whether an error is Type I or Type II
4. state conclusions to hypothesis tests

The Nature of Hypothesis Testing

According to the [US Census Bureau](#), 64% of US citizens age 18 or older voted in the 2004 election. Suppose we believe that percentage is higher for ECC students for the 2008 presidential election. To determine if our suspicions are correct, we collect information from a random sample of 500 ECC students. Of those, 460 were citizens and 18 or older in time for the election. (We have some students still in high school, and some who do not yet have citizenship.) Of those who were eligible to vote, 326 (or about 71%) say that they did vote.

Problem: Based on this random sample, do we have enough evidence to say that the percentage of ECC students who were eligible to vote and did vote in the 2008 presidential election was higher than the proportion of US citizens who voted in the 2004 election?

Solution: 71% of the students in our sample voted. Obviously, this is higher than the national average. The thing to consider, though, is that this is just a *sample* of ECC students - it isn't every student. It's possible that the students who just happened to be in our sample were those who voted. Maybe this could just happen randomly. In order to determine if it really is that different from the national proportion, we need to find out how probable a sample proportion of 71% would be if the true proportion was really 64%.

To answer this, we consider the distribution of the proportion of eligible voters who did vote. From [Section 8.2](#), we know that this proportion is approximately normally distributed if $np(1-p) \geq 10$. Using the techniques from that section, we also know that the mean would be 64%, with a standard deviation of about 2.2%. With this information, the probability of observing a random sample with a proportion of 71% if the true proportion is 64% is about 0.001.

This means that about 1 in 1000 random samples will have a proportion that large. So we have two conclusions:

1. We just observed an *extremely* rare event, or
2. The proportion of ECC students who vote is actually higher than 64%.

This is the idea behind hypothesis testing. The general process is this:

Steps in Hypothesis Testing

1. A claim is made. (More than 64% of ECC students vote, in our case.)
2. Evidence is collected to test the claim. (We found that 326 of 460 voted.)
3. The data are analyzed to assess the plausibility of the claim. (We determined that the proportion is most likely higher than 64% for ECC students.)

Null and Alternative Hypotheses

In statistics, we call these claims *hypotheses*. We have two types of hypotheses, a **null hypothesis** and an **alternative hypothesis**.

The **null hypothesis**, denoted H_0 ("H-naught"), is a statement to be tested. It is usually the status quo, and is assumed true until evidence is found otherwise.

The **alternative hypothesis**, denoted H_1 ("H-one"), is a claim to be tested. We will try to find evidence to support the alternative hypothesis.

There are three general ways in this chapter that we'll set up the null and alternative hypothesis.

1. **two tailed**

H_0 : parameter = some value

H_1 : parameter \neq the value

2. **left-tailed**

H_0 : parameter = some value

H_1 : parameter < the value

3. **right-tailed**

H_0 : parameter = some value

H_1 : parameter > the value

Let's look an example of each.

Example 1

In the introduction of this section, we were considering the proportion of ECC students who voted in the 2008 presidential election. We assumed that it was the same as the national proportion in 2004 (64%), and tried to find evidence that it was higher than that. In that case, our null and alternative hypotheses would be:

H_0 : $p = 0.64$

H_1 : $p > 0.64$

Example 2

According to the [Elgin Community College website](#), the average age of ECC students is 28.2 years. We might claim that the average is less for online Mth120 students. In that case, our null and alternative hypotheses would be:

H_0 : $\mu = 28$

H_1 : $\mu < 28$

Example 3

It's fairly standard knowledge that IQ tests are designed to be normally distributed, with an average of 100. We wonder whether the IQ of ECC students is different from this average. Our hypotheses would then be:

H_0 : $\mu = 100$

H_1 : $\mu \neq 100$

The Four Outcomes of Hypothesis Testing

Unfortunately, we never know with 100% certainty what is true in reality. We always make our decision based on sample data, which may or may not reflect reality. So we'll make our decision, but it may not always be correct.

In general, there are four possible outcomes from a hypothesis test when we compare our decision with what is true in reality - which we will never know!

- We could decide to not reject the null hypothesis when in reality the null hypothesis was true. This would be a correct decision.
- We could reject the null hypothesis when in reality the alternative hypothesis is true. This would also be a correct decision.
- We could reject the null hypothesis when it really is true. We call this error a **Type I error**.
- We could decide to not reject the null hypothesis, when in reality we should have, because the alternative was true. We call this error a **Type II error**.

		reality	
		H_0 true	H_1 true
decision	do not reject H_0	correct decision	Type II error
	reject H_0	Type I error	correct decision

To help illustrate the idea, let's look at an example.

Example 4

Let's consider a [pregnancy test](#). The tests work by looking for the presence of the hormone human chorionic gonadotropin (hCG), which is secreted by the placenta after the fertilized egg implants in a woman's uterus.

If we consider this in the language of a hypothesis test, the null hypothesis here is that the woman is not pregnant - this is what we assume is true until proven otherwise.

The corresponding chart for this test would look something like this:

		reality	
		not pregnant	pregnant
pregnancy test	not pregnant	correct decision	Type II error
	pregnant	Type I error	correct decision

In this case, we would call a Type I error a "false positive" - the test was positive for pregnancy, when in reality the mother was not pregnant.

The Type II error in this context would be a "false negative" - the test did not reveal the pregnancy, when the woman really was pregnant.

When a test claims that it is "[99% Accurate at Detecting Pregnancies](#)", it is referring to Type II errors. The tests claim to detect 99% of pregnancies, so it will make a Type II error (not detecting the pregnancy) only 1% of the time.

Note: The "99% Accurate" claim is not entirely correct. Many tests do not

have this accuracy until a few days after a missed period. (Source: [US Dept. of Health and Human Services](#))

Level of Significance

As notation, we assign two Greek letters, α ("alpha") and β ("beta"), to the probability of Type I and Type II errors, respectively.

$$\alpha = P(\text{Type I error}) = P(\text{rejecting } H_0 \text{ when } H_0 \text{ is true})$$

$$\beta = P(\text{Type II error}) = P(\text{not rejecting } H_0 \text{ when } H_1 \text{ is true})$$

Unfortunately, we can't control *both* errors, so researchers choose the probability of making a Type I error *before* the sample is collected. We refer to this probability as the **level of significance** of the test.

Choosing α , the Probability of a Type I Error

One interesting topic is the choice of α . How do we choose? The answer to that is to consider the consequences of the mistake. Consider two examples:

Example 5

Let's consider again a murder trial, with a possibility death penalty if the defendant is convicted. The null and alternative hypotheses again are:

H_0 : the defendant is innocent

H_1 : the defendant is guilty

In this case, a Type I error (rejecting the null when it really is true) would be when the jury returns a "guilty" verdict, when the defendant is actually innocent.

In our judicial system, we say that we must be sure "beyond a reasonable doubt". And in this case in particular, the consequences of a mistake would be the death of an innocent defendant.

Clearly in this case, we want the probability of making this error very small, so we might assign to α a value like 0.00001. (So about 1 out of every 100,000 such trials will result in an incorrect guilty verdict. You might choose an even smaller value - especially if you are the defendant in question!)

Example 6

As an alternative, suppose we're considering the average age of ECC students. As stated in [Example 2](#), we assumed that average age is 28, but we think it might be lower for online Mth120 students.

$H_0: \mu = 28$

$H_1: \mu < 28$

In this example, $\alpha = P(\text{concluding that the average age is less than 28, when it really is 28})$. The consequences of that mistake are... that we're wrong about the average age. Certainly nothing as dire as sending an innocent defendant to jail - or worse.

Because the consequences of a Type I error or not nearly as dire, we might assign a value of $\alpha = 0.05$.

Example 7

Let's look at a different example. Suppose we have a test for the deadly disease statisticitis, which affects approximately 5% of the general population.

Problem: If a test is developed that claims to be 90% accurate, what is the probability that an individual with a positive test result for statisticitis actually has the disease?

Solution: One good way to do this is to start with 1000 theoretical individuals. Since we know 5% of the population has the disease, 5% of the 1000, or 50, must have the disease. We can illustrate this in a chart like the ones we did earlier:

		reality		1000
		no disease	disease	
test result	negative			
	positive			
		950	50	

We're also told that the test is "90% accurate", so 90% of those 50 with the disease will have positive test results. That gives us something like the following:

		reality		1000
		no disease	disease	
test result	negative		$0.1 * 50 = 5$	
	positive		$0.9 * 50 = 45$	
		950	50	

Now, we also have to consider the test being "90% accurate" and that it should return a negative result for 90% of the 950 who do *not* have the disease.

		reality		1000
		no disease	disease	
test result	negative	$0.9 * 950 = 855$	5	
	positive	$0.1 * 950 = 95$	45	
		950	50	

So in summary, we have the following table:

		reality		1000
		no disease	disease	
test result	negative	855	5	860
	positive	95	45	140
		950	50	

We were asked originally for the probability that an individual with a positive test actually has the disease. In other words, we want:

$$P(\text{has disease} \mid \text{positive test}) = 45/140 \approx 0.32$$

What does this mean? Well, if this test is "90% accurate", then a positive test result actually only means you have the disease about 32% of the time!

On the other hand, it's interesting to note that a negative result is fairly accurate - $855/860 \approx 0.994 = 99.4\%$. So you can feel pretty confident of a negative result.

This may seem unreal, but some tests really do work this way. What is done to balance this out is to treat a "positive" test result as not definitive - it simply warrants more study.

The [test for gestational diabetes](#) is a perfect example. The test is performed in two stages - an initial "screening", which is extended to a second test if the result is positive.

Stating Conclusions to Hypothesis Tests

We have to be very careful when we state our conclusions. The way this type of hypothesis testing works, we look for evidence to support the alternative claim. If we find it, we say we have enough to support the alternative. If we don't find that evidence, we say just that - we don't have enough evidence to support the alternative hypothesis.

We never support the null hypothesis. In fact, the null hypothesis is probably not *exactly* true. There are some very interesting discussions of this type of hypothesis testing (see links at the end of this page), and one criticism is that large enough sample will probably show a difference from the null hypothesis.

That said, we look for evidence to support the alternative hypothesis. If we don't find it, then we simply say that we don't have enough evidence to support the alternative hypothesis.

To illustrate, consider the three examples from earlier this section:

Example 8

Let's consider [Example 1](#), introduced earlier this section. The null and alternative hypotheses were as follows:

$$H_0: p = 0.64$$

$$H_1: p > 0.64$$

Suppose we decide that the evidence supports the alternative claim (it does). Our conclusion would then be:

There is enough evidence to support the claim that the proportion of ECC students who were eligible to vote and did was more than the national proportion of 64% who voted in 2004.

Example 9

In [Example 2](#), we stated null and alternative hypotheses:

$$H_0: \mu = 28$$

$$H_1: \mu < 28$$

Suppose we find an average age of 20.4, and we decide to reject the null hypothesis. Our conclusion would then be:

There is enough evidence to support our claim that the average age of online Mth120 students is less than 28 years.

Example 10

In [Example 3](#), we stated null and alternative hypotheses:

$$H_0: \mu = 100$$

$$H_1: \mu \neq 100$$

Suppose we find an average IQ from a sample of 30 students to be 101, which we determine is not significantly different from the assumed mean. Our conclusion would then be:

There is not enough evidence to support the claim that the average IQ of ECC students is not 100.

The Controversy Regarding Hypothesis Testing

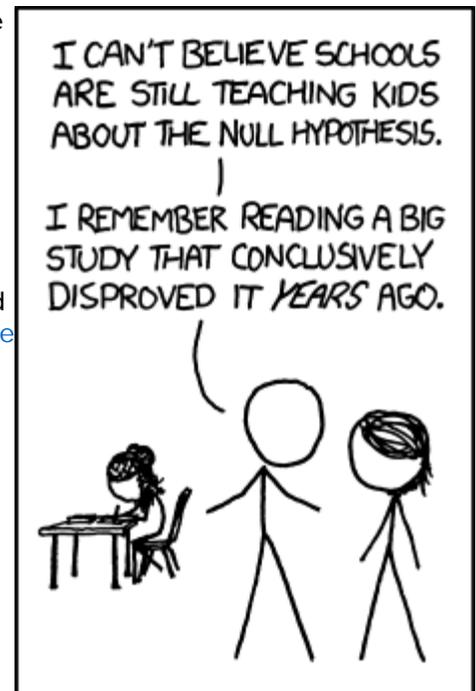
Unlike other fields in mathematics, there are many areas in statistics which are still being debated. Hypothesis Testing is one of them. There are several concerns that any good statistician should be aware of.

- Tests are significantly affected by sample size
- Not rejecting the null hypothesis does not mean it is true
- Statistical significance does not imply practical significance
- Repeated tests can be misleading (see [this illuminating comic](#) from XKCD)

Much more subtly, the truth is likely that the null hypothesis is *never* true, and rejecting it is only a matter of getting a large enough sample. Consider [Example 6](#), in which we assumed the average age was 28. Well... to how many digits? Isn't it likely that the average age is actually 27.85083 or something of the sort? In that case, all we need a large enough sample to get a sample mean that is statistically less than 28, even though the difference really has no practical meaning.

If you're interested in reading more about some of the weaknesses of this method for testing hypotheses, visit these links:

- [Commentaries on Significance Testing](#)
- [The Concept of Statistical Significance Testing](#)
- [Null-Hypothesis Controversy](#)
- [The Controversy Over How to Present Research Findings](#)



So Why Are We Studying Hypothesis Testing?

The reality is that we need some way to analytically make decisions regarding our observations. While it is simplistic and certainly contains errors, hypothesis testing still does have value, provided we understand its limitations.

This is also an *introductory* statistics course. There are limits to what we can learn in a single semester. There are more robust ways to perform a hypothesis test, including effect size, power analysis, and Bayesian inference. Unfortunately, many of these are simply beyond the scope of this course.

The point of this discussion is to be clear that hypothesis testing has weaknesses. By understanding them, we can make clear statements about the results of a hypothesis test and what we can actually conclude.



This work is licensed under a Creative Commons License.

Section 10.2: Hypothesis Tests for a Population Proportion

- 10.1 The Language of Hypothesis Testing
- 10.2 Hypothesis Tests for a Population Proportion**
- 10.3 Hypothesis Tests for a Population Mean
- 10.4 Hypothesis Tests for a Population Standard Deviation
- 10.5 Putting It Together: Which Method Do I Use?

Objectives

By the end of this lesson, you will be able to...

1. explain the logic of hypothesis testing
2. test hypotheses about a population proportion
3. test hypotheses about a population proportion using the binomial probability distribution

The Logic of Hypothesis Testing

Once we have our null and alternative hypotheses chosen, and our sample data collected, how do we choose whether or not to reject the null hypothesis? In a nutshell, it's this:

If the observed results are unlikely assuming that the null hypothesis is true, we say the result is **statistically significant**, and we reject the null hypothesis. In other words, the observed results are so unusual, that our original assumption in the null hypothesis must not have been correct.

Your textbook references three different methods for testing hypotheses:

- the classical approach
- **P-values**
- confidence intervals

Because *P*-values are so much more widely used, we will be focusing on this method. **You will be required to include *P*-values for your homework and exams.**

If you're interested in learning any of these other methods, feel free to read through the textbook.

P-Values

In general, we define the *P*-value this way:

The **P-value** is the probability of observing a sample statistic as extreme or more extreme than the one observed in the sample assuming that the null hypothesis is true.

The Sample Proportion

In Section 8.2, we learned about the distribution of the sample proportion, so let's do a quick review of that now.

In general, if we let x = the number with the specific characteristic, then the **sample proportion**, \hat{p} , (read "*p*-hat") is given by:

$$\hat{p} = \frac{x}{n}$$

Where \hat{p} is an estimate for the population proportion, p .

We also learned some information about how the sample proportion is distributed:

Sampling Distribution of \hat{p}

For a simple random sample of size n such that $n \leq 0.05N$ (in other words, the sample is less than 5% of the population),

- The shape of the sampling distribution of \hat{p} is approximately normal provided $np(1-p) \geq 10$
- The mean of the sampling distribution of \hat{p} is $\mu_{\hat{p}} = p$.
- The standard deviation of the sampling distribution of \hat{p} is

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

Why are these important? Well, suppose we take a sample of 100 online students, and find that 74 of them are part-time. You might recall that based on data from elgin.edu, 68.5% of ECC students in general are part-time. So is observing 74% of our sample unusual? How do we know - **we need the distribution of \hat{p} !**

So what we do is create a **test statistic** based on our sample, and then use a table or technology to find the probability of what we observed. Here are the details.

Testing Claims Regarding the Population Proportion Using P -Values

In this first section, we assume we are testing some claim about the population proportion. As usual, the following two conditions must be true:

1. $np(1-p) \geq 10$, and
2. $n \leq 0.05N$

Step 1: State the null and alternative hypotheses.

Two-Tailed Left-Tailed Right-Tailed

$$H_0: p = p_0 \quad H_0: p = p_0 \quad H_0: p = p_0$$

$$H_1: p \neq p_0 \quad H_1: p < p_0 \quad H_1: p > p_0$$

Step 2: Decide on a level of significance, α , depending on the seriousness of making a Type I error. (α will often be given as part of a test or homework question, but this will not be the case in the outside world.)

Step 3: Compute the test statistic,
$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

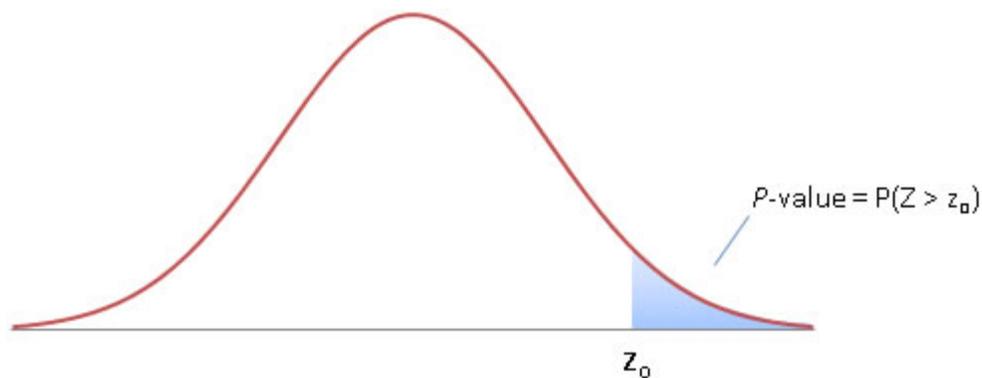
Step 4: Determine the P -value.

Step 5: Reject the null hypothesis if the P -value is less than the level of significance, α .

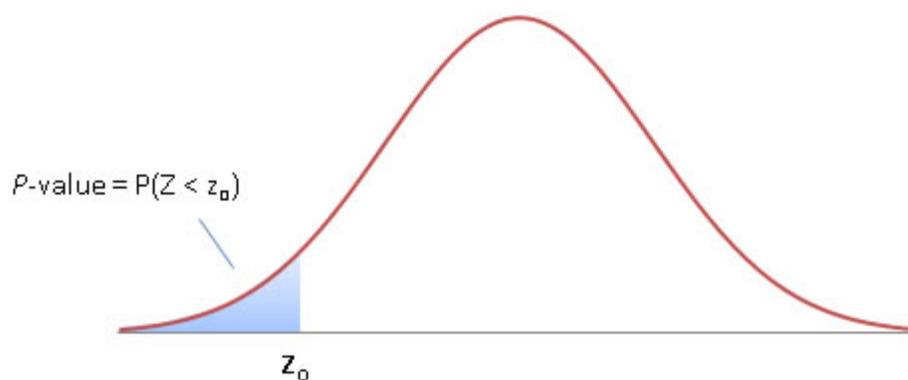
Step 6: State the conclusion. (You should also include a measure of the *strength* of the results, based on the P -value.)

Calculating P-Values

Right-Tailed Tests

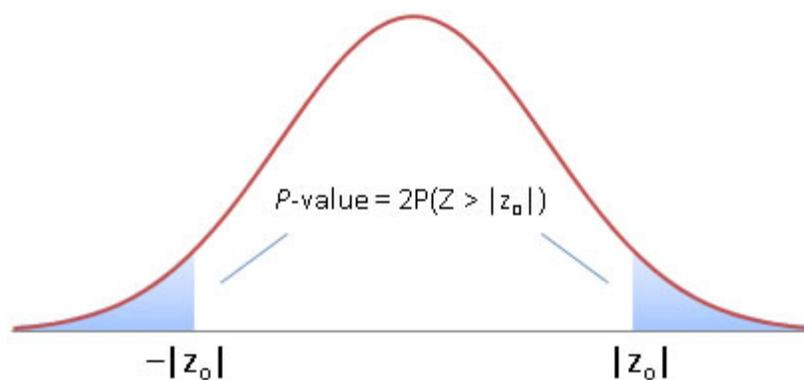


Left-Tailed Tests



Two-Tailed Tests

In a two-tailed test, the $P\text{-value} = 2P(Z > |z_0|)$.



It may seem odd to multiply the probability by two, since "or more extreme" seems to imply the area in the tail only. The reason why we do multiply by two is that even though the result was on one side, we didn't know *before collecting the data*, on which side it would be.

The Strength of the Evidence

Since the P -value represents the probability of observing our result or more extreme, the smaller the P -value, the more unusual our observation was. Another way to look at it is this:

The smaller the P -value, the stronger the evidence supporting the alternative hypothesis. We can use the following guideline:

- P -value < 0.01 : **very strong** evidence supporting the alternative hypothesis
- $0.01 \leq P$ -value < 0.05 : **strong** evidence supporting the alternative hypothesis
- $0.05 \leq P$ -value < 0.1 : **some** evidence supporting the alternative hypothesis
- P -value ≥ 0.1 : **weak to no** evidence supporting the alternative hypothesis

These values are not hard lines, of course, but they can give us a general idea of the strength of the evidence.

But wait! There is an important caveat here, which was mentioned earlier in the section about [The Controversy Regarding Hypothesis Testing](#). The problem is that it's relatively easy to get a large p -value - just get a really large sample size! So the chart above is really with the caveat "**assuming equal sample sizes in comparable studies, ...**"

This isn't something every statistics text will mention, nor will every instructor mention, but it's important.

Example 1

According to the [Elgin Community College website](#), approximately 56% of ECC students are female. Suppose we wonder if the same proportion is true for math courses. If we collect a sample of 200 ECC students enrolled in math courses and find that 105 of them are female, do we have enough evidence at the 10% level of significance to say that the proportion of math students who are female is different from the general population?

Note: Be sure to check that the conditions for performing the hypothesis test are met.

[[reveal answer](#)]

Hypothesis Testing Regarding p Using StatCrunch

1. Go to **Stat > Proportions > One sample > with summary**.
2. Enter the number of successes and the number of observations.
3. Enter p_0 and H_1 , then press **Calculate**.

The results should be displayed.

Example 2

Consider the excerpt shown below (also used in [Example 1](#), in Section 9.3) from a poll conducted by Pew Research:

Stem cell, marijuana proposals lead in Mich. poll

*A recent poll shows voter support leading opposition for ballot proposals to loosen Michigan's restrictions on embryonic stem cell research and allow medical use of marijuana. The EPIC-MRA poll conducted for The Detroit News and television stations WXYZ, WILX, WOOD and WJRT found **50 percent of***

likely Michigan voters support the stem cell proposal, 32 percent against and 18 percent undecided. The telephone poll of 602 likely Michigan voters was conducted Sept. 22 through Wednesday. It has a margin of sampling error of plus or minus 4 percentage points. (Source: Associated Press)

Suppose we wonder if the percent of Elgin Community College students who support stem cell research is different from this. If 61 of 100 randomly selected ECC students support stem cell research, is there enough evidence at the 5% level of significance to support our claim?

Note: Be sure to check that the conditions for performing the hypothesis test are met.

[[reveal answer](#)]

One question you might have is, "What do we do if the conditions for the hypothesis test about p aren't met?" Great question!

In that case, we can no longer say that sample proportion, \hat{p} , is approximately normally distributed. What we do instead is return to the binomial distribution, and just consider x , the number of successes. Let's do a quick review of binomial probabilities.

A Binomial Refresher

The Binomial Probability Distribution Function

The probability of obtaining x successes in n independent trials of a binomial experiment, where the probability of success is p , is given by

$$P(x) = {}_n C_x p^x (1-p)^{n-x}$$

Where $x = 0, 1, 2, \dots, n$

Using Technology to Calculate Binomial Probabilities

Here's a quick overview of the formulas for finding binomial probabilities in StatCrunch.

Click on **Stat > Calculators > Binomial**

Enter n , p , the appropriate equality/inequality, and x . The figure below shows $P(X \geq 3)$ if $n=4$ and $p=0.25$.

The screenshot shows the StatCrunch Binomial calculator interface. It has input fields for 'n' (4) and 'p' (0.25). Below these is a dropdown menu for the probability operator, currently set to '=>', and an input field for 'x' (3). To the right of the 'x' field is an equals sign and a text box containing the result '0.05078125'. At the bottom are two buttons: 'Close' and 'Compute'.

Hypothesis Testing Using the Binomial Distribution

Example 3

Traditionally, about 70% of students in a particular Statistics course at ECC are successful. If only 15 students in a class of 28 randomly selected

students are successful, is there enough evidence at the 5% level of significance to say that students of that particular instructor are successful at a rate less than 70%?

[\[reveal answer \]](#)

[<< previous section](#) | [next section >>](#)

[1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) **[10](#)** [11](#) [12](#) [13](#)



This work is licensed under a Creative Commons License.

Section 10.3: Hypothesis Tests for a Population Mean

- 10.1 The Language of Hypothesis Testing
- 10.2 Hypothesis Tests for a Population Proportion
- 10.3 Hypothesis Tests for a Population Mean**
- 10.4 Hypothesis Tests for a Population Standard Deviation
- 10.5 Putting It Together: Which Method Do I Use?

Objectives

By the end of this lesson, you will be able to...

1. test hypotheses about a mean
2. understand the difference between statistical significance and practical significance

In Section 10.2, we tested hypotheses regarding a population proportion. In this section, we'll consider claims regarding μ , the population mean.

As we did in the previous section, we have some conditions that need to be true in order to perform the test (based on the [Central Limit Theorem](#) from Chapter 8).

1. the sample is obtained using simple random sampling, and
2. the sample has no outliers and the population from which the sample is drawn is normally distributed, or the sample size is large ($n \geq 30$)

The steps in performing the hypothesis test are nearly identical, with only a couple changes.

Performing a Hypothesis Test Regarding μ

Step 1: State the null and alternative hypotheses.

Two-Tailed Left-Tailed Right-Tailed

$H_0: \mu = \mu_0$ $H_0: \mu = \mu_0$ $H_0: \mu = \mu_0$

$H_1: \mu \neq \mu_0$ $H_1: \mu < \mu_0$ $H_1: \mu > \mu_0$

Step 2: Decide on a level of significance, α , depending on the seriousness of making a Type I error. (α will often be given as part of a test or homework question, but this will not be the case in the outside world.)

Step 3: Compute the test statistic, $t_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$.

Step 4: Determine the P -value (see below).

Step 5: Reject the null hypothesis if the P -value is less than the level of significance, α .

Step 6: State the conclusion. (You should also include a measure of the *strength* of the results, based on the P -value.)

Finding P -values using the t -distribution

Because the t -table does not give probabilities, we will need to use technology to find probabilities in the t -distribution.

Probabilities in the t -Distribution Using StatCrunch

We can also use either of the software packages we've been using so far to find probabilities in the t -distribution. Just click on the name of the software you plan to use.

Click on **Stat > Calculators > T**

Enter the degree of freedom, the direction of the inequality, and x . Then press **Compute**.

Example 1

Problem: In [Example 1](#), in Section 10.2, we considered a survey about student work habits. The students who performed the survey found that ECC transfer students work, on average, about 17 hours per week.

In that example, we theorized that Mth120 online students chose the online format because they tend to work full-time, or at least more than the average student. We also assumed that the standard deviation for the hours worked per week was 5 hours.

Suppose instead that we don't have any prior information, and we simply use the sample data. If we collect a random sample of 30 Mth120 online students, and find a sample mean of 19.3 hours per week with a standard deviation of 6.2 hours, is there enough evidence at the 5% level of significance to support our claim that Mth120 online students tend to work more than the average ECC transfer student? (Note: Assume that the sample data are approximately normally distributed with no outliers.)

Solution:

Step 1:

$$H_0: \mu = 17$$

$$H_1: \mu > 17$$

Step 2: $\alpha = 0.05$ (given)

$$\text{Step 3: } t_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{19.3 - 17}{6.2/\sqrt{30}} \approx 2.03.$$

Step 4: $P\text{-value} = P(t > 2.03, df=29) \approx 0.0258$.

Step 5: Since the $P\text{-value} < \alpha$, we reject the null hypothesis.

Step 6: Yes, there is enough evidence at the 5% level of significance to support our claim that the average hours worked for Mth120 online students is more than 17 hours per week.

Hypothesis Testing Using StatCrunch

1. Enter the data.
2. Go to **Stat > t-Statistics > One Sample**
3. Select **with data** if you have the data, or **with summary** if you only have the summary statistics.
4. If you chose **with data**, click on the variable that you want for the confidence interval and enter the population standard deviation. Otherwise, enter the sample statistics.
5. Click on **Next**.

6. Enter μ_0 and the alternative hypothesis.
7. Click on **Calculate**.

The results should be displayed.

Example 2

In [Example 2](#) in Section 10.2, we looked the resting heart rates of 25 Statistics students.

heart rate				
61	63	64	65	65
67	71	72	73	74
75	77	79	80	81
82	83	83	84	85
86	86	89	95	95

([Click here](#) to view the data in a format more easily copied.)

In that problem, we assumed that the standard deviation of all resting heart rates was 12 beats per minute. Let's suppose we don't have information about the standard deviation.

According to the National Library of Medicine, the average resting heart rate for Americans is about 72 bpm. Is there evidence at the 5% level of significance to support a claim that this particular class of Statistics students has a heart rate different from the national average?

Be sure to check that the conditions for performing hypothesis tests are met.

[\[reveal answer \]](#)

Statistical Significance vs. Practical Significance

There's one last point we need to discuss in this section, and it's an important one. We need to talk about the difference between a result being **statistically significant** versus a result that is **practically significant**.

The idea is this - suppose we're comparing the performance of men and women in mathematics. To compare, we use a particular exam, given randomly to a certain number of students from across the country. We assume the difference between the two groups is zero, but we wonder whether it might be different. In that case, our null and alternative hypotheses might look something like this:

$$H_0: \mu = 0$$

$$H_1: \mu \neq 0$$

Let's suppose that women are only *slightly* better than men on this exam - say a difference of 0.5 (if we assume the scale is 0-100). If we have a sample of 30 students - 15 men and 15 women - it'll be very hard to get a sample mean very far from zero, and so it'd be pretty rare that we'd actually find enough evidence to reject the null hypothesis (meaning we'd be making Type II error - not rejecting H_0 when we should).

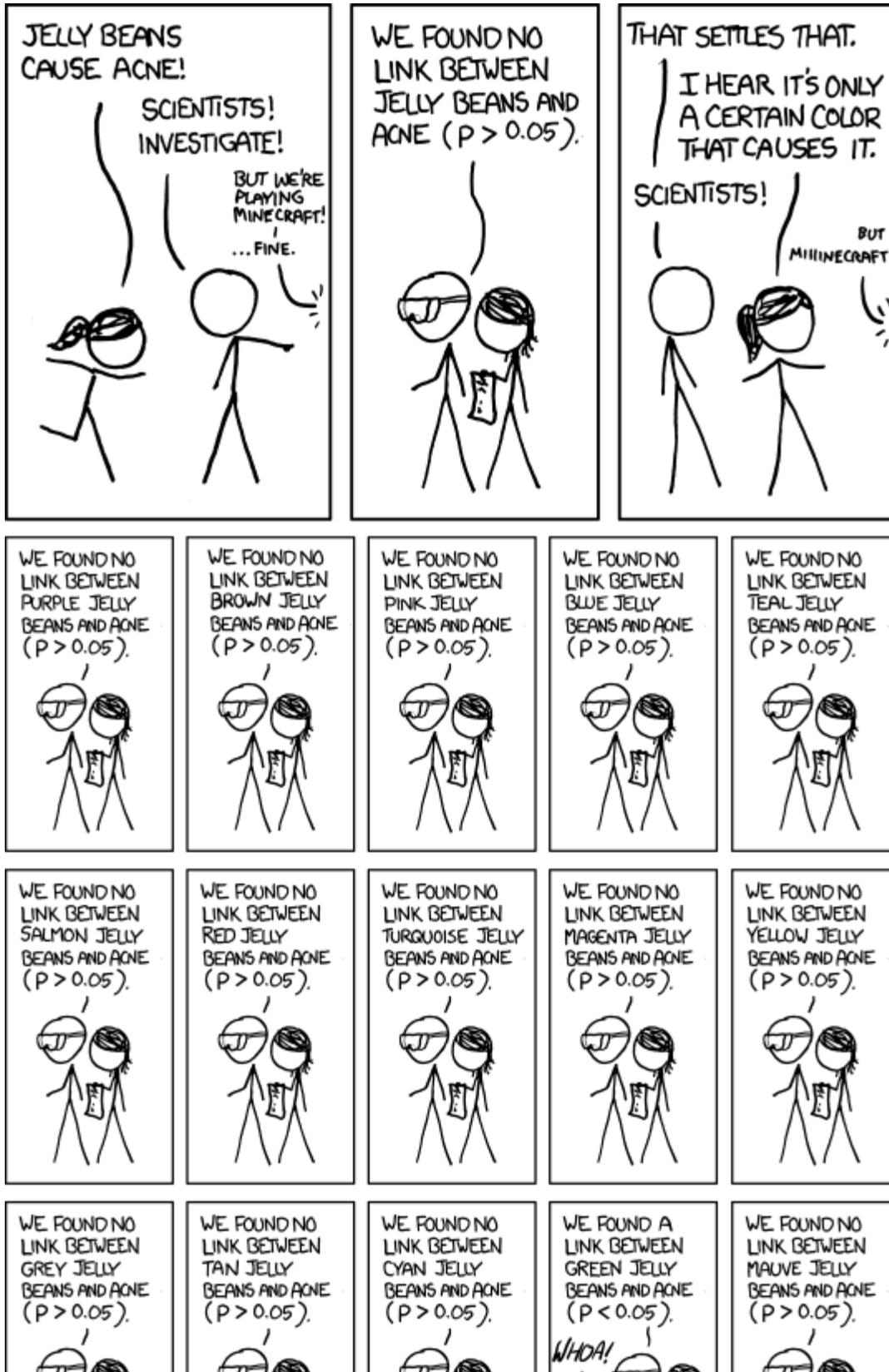
On the other hand, if we had 10,000 students, it'd be very likely to get sample mean very close to the real mean, and a sample size of 10,000 would give us an *extremely* small P -value - definitely small enough to reject H_0 .

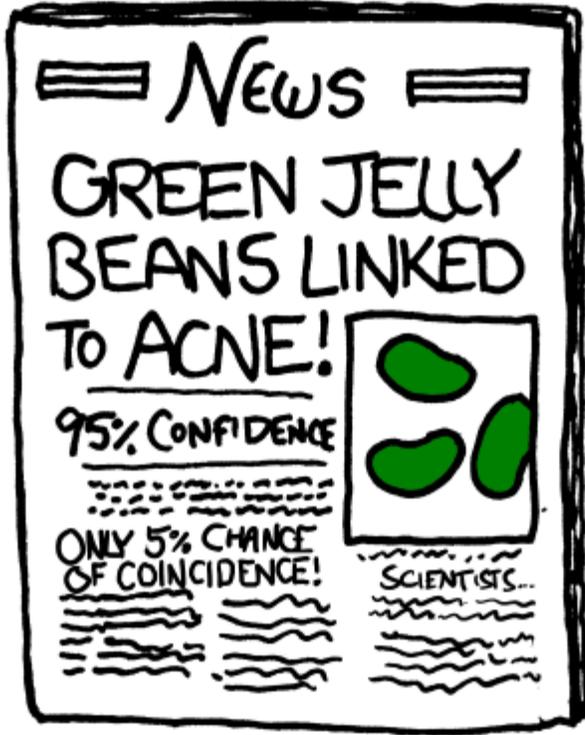
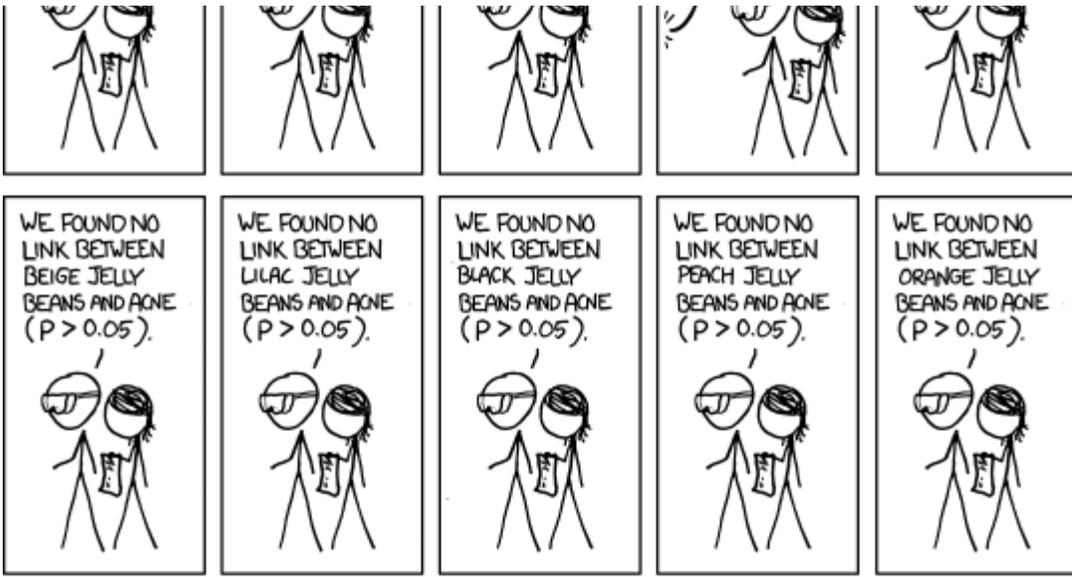
Here's the thing - in that second case, we'd reject the null hypothesis and say that the evidence supports the claim that men and women are different, but the difference in practical terms is insignificant. So *statistical*

significance (enough evidence to reject the null hypothesis) doesn't necessarily imply *practical significance*.

Beware of studies with very large sample sizes that claim statistical significance. It may be that the differences have no practical meaning.

As an illustration, check out this comic from [XKCD](#):





[<< previous section](#) | [next section >>](#)

1 2 3 4 5 6 7 8 9 **10** 11 12 13

Section 10.4: Hypothesis Tests for a Population Standard Deviation

- 10.1 The Language of Hypothesis Testing
- 10.2 Hypothesis Tests for a Population Proportion
- 10.3 Hypothesis Tests for a Population Mean
- 10.4 Hypothesis Tests for a Population Standard Deviation**
- 10.5 Putting It Together: Which Method Do I Use?

Objectives

By the end of this lesson, you will be able to...

1. test hypotheses about a population standard deviation

Before we begin this section, we need a quick refresher of the X^2 distribution.

The Chi-Square (X^2) distribution

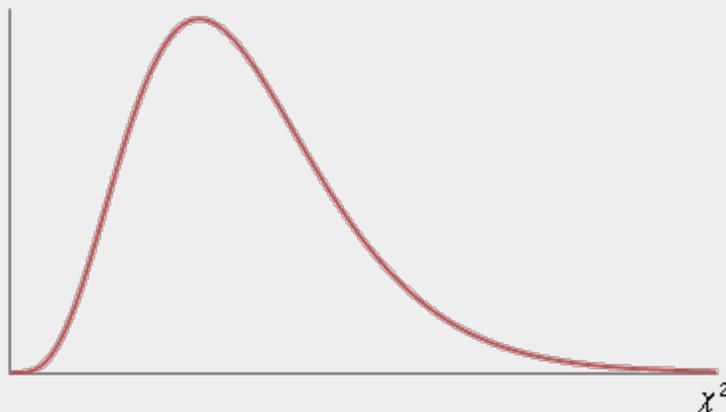
Reminder: "chi-square" is pronounced "kai" as in sky, not "chai" like [the tea](#).

The Chi-Square (X^2) distribution

If a simple random sample size n is obtained from a normally distributed population with mean μ and standard deviation σ , then

$$\chi^2 = \frac{(n - 1)s^2}{\sigma^2}$$

has a **chi-square distribution** with $n-1$ degrees of freedom.



Properties of the X^2 distribution

1. It is *not* symmetric.
2. The shape depends on the degrees of freedom.
3. As the number of degrees of freedom increases, the distribution becomes more symmetric.
4. $X^2 \geq 0$

Finding Probabilities Using StatCrunch

Click on **Stat** > **Calculators** > **Chi-Square**

Enter the degrees of freedom, the direction of the inequality, and X.
Then press **Compute**.

Performing a Hypothesis Test Regarding σ

Step 1: State the null and alternative hypotheses.

Two-Tailed Left-Tailed Right-Tailed

$H_0: \sigma = \sigma_0$ $H_0: \sigma = \sigma_0$ $H_0: \sigma = \sigma_0$

$H_1: \sigma \neq \sigma_0$ $H_1: \sigma < \sigma_0$ $H_1: \sigma > \sigma_0$

Step 2: Decide on a level of significance, α .

Step 3: Compute the test statistic, $\chi_0^2 = \frac{(n-1)s^2}{\sigma^2}$.

Step 4: Determine the P -value.

Step 5: Reject the null hypothesis if the P -value is less than the level of significance, α .

Step 6: State the conclusion.

Example 1

In [Example 2](#), in Section 10.2, we assumed that the standard deviation for the resting heart rates of ECC students was 12 bpm. Later, in [Example 2](#) in Section 10.3, we considered the actual sample data below.

heart rate				
61	63	64	65	65
67	71	72	73	74
75	77	79	80	81
82	83	83	84	85
86	86	89	95	95

([Click here](#) to view the data in a format more easily copied.)

Based on this sample, is there enough evidence to say that the standard deviation of the resting heart rates for students in this class is different from 12 bpm?

Note: Be sure to check that the conditions for performing the hypothesis test are met.

[[reveal answer](#)]

Hypothesis Testing Regarding σ Using StatCrunch

1. Go to **Stat > Variance > One sample**
2. Select **with data** if you have the data, or **with summary** if you only have the summary statistics.
3. If you chose **with data**, click on the variable that you want for the hypothesis test and press **Next**. Otherwise, enter the sample statistics and press **Next**.
4. Enter the population *variance* (not standard deviation!) and the H_1
5. Press **Calculate**.

The results should be displayed.

Example 2

Let's look at Example 1 again, and try the hypothesis test with technology.

[\[reveal answer \]](#)

[<< previous section](#) | [next section >>](#)

[1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) **10** [11](#) [12](#) [13](#)



This work is licensed under a Creative Commons License.

Section 10.5: Putting It Together: Which Method Do I Use?

- 10.1 The Language of Hypothesis Testing
- 10.2 Hypothesis Tests for a Population Proportion
- 10.3 Hypothesis Tests for a Population Mean
- 10.4 Hypothesis Tests for a Population Standard Deviation
- 10.5 Putting It Together: Which Method Do I Use?**

Objectives

By the end of this lesson, you will be able to...

1. determine the appropriate hypothesis test to perform

Hypothesis Test Summary

So far this semester, we've learned three different hypothesis tests, based on the parameter of interest, and what information is given. Those three are:

Tests Regarding the Population Proportion

In order to perform a hypothesis test regarding the population proportion, *all* of the following must be true:

1. the sample is a simple random sample, and
2. the sample is less than 5% of the population ($n \leq 0.05N$), and
3. $np(1-p) \geq 10$

The sample statistic for this test is:

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Tests Regarding the Population Mean

In order to perform a hypothesis test regarding the population mean, the sample must be a simple random sample, and one of the following must be true:

1. the sample comes from a normally distributed population, or
2. the sample size is more than 30 and there sample contains no outliers.

Two-Tailed Left-Tailed Right-Tailed

$$\begin{array}{lll}
 H_0: \mu = \mu_0 & H_0: \mu = \mu_0 & H_0: \mu = \mu_0 \\
 H_1: \mu \neq \mu_0 & H_1: \mu < \mu_0 & H_1: \mu > \mu_0
 \end{array}$$

The sample statistic for this test is:

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

with n-1 degrees of freedom.

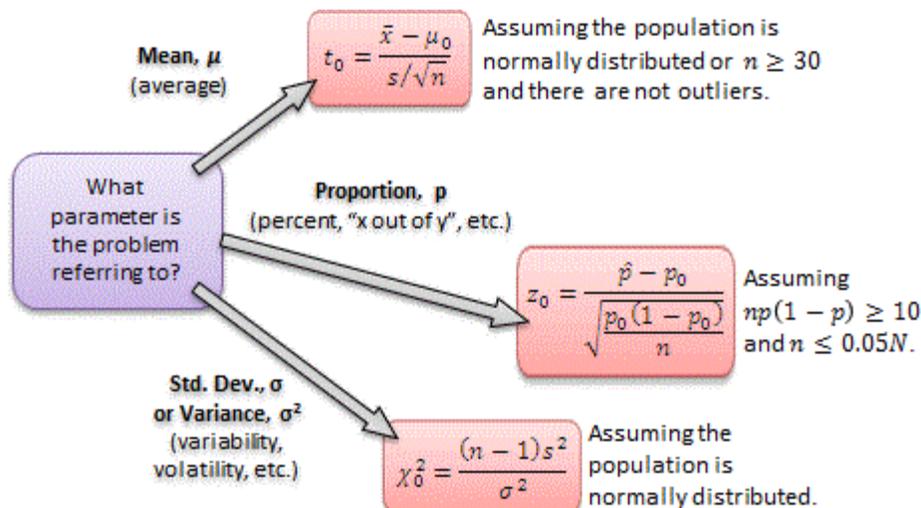
Tests Regarding the Population Standard Deviation

In order to perform a hypothesis test regarding the population standard deviation, the sample *must* come from a normally distributed population. In this case, the sample *must* come from a normally distributed population. In this case, the sample statistic is:

$$\chi_0^2 = \frac{(n-1)s^2}{\sigma^2}$$

Choosing the Appropriate Hypothesis Test

So now the big question is which test to apply. Here's a simplified version of the flowchart in your textbook:



Unfortunately, like most problems where you need to choose which technique to apply, there's no handy blueprint that you can always follow. The key, then, is to try as many examples as possible (like those on the next page) and doing all the assigned homework as soon as possible.

For each example, state null and alternative hypotheses and choose the appropriate test statistic.

Example 1

According to the [US Census Bureau](#), approximately 42% of Americans ages 18-24 voted in the 2004 presidential election. We wonder if the percentage of ECC students is different from this. If we collect a simple random sample of 200 ECC students and find that 103 of them voted in the 2008 presidential election, is there enough evidence at the 10% level of significance to support our claim?

[[reveal answer](#)]

Example 2

According to the ECC website, the average age of ECC students is 28.2 years. Suppose we collect data from 30 Mth120 online students and find an average age of 20.3 years, with a standard deviation of 2.3 years.

Is there enough evidence at the 5% level of significance to support the claim that the average age of Mth120 online students is less than 28.2 years?

[[reveal answer](#)]

Example 3

We know from previous examples that the standard deviation of IQs is normally distributed with a standard deviation of 15. Suppose we wonder if the IQs of ECC students have more variation. To answer this question, we collect the IQs from a random sample of ECC students and find a standard deviation of 16.2.

Based on this information, is there enough evidence at the 5% level of significance to say that the IQs of ECC students have more variation than the general population?

[\[reveal answer \]](#)

[<< previous section](#) | [next section >>](#)

[1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [11](#) [12](#) [13](#)



This work is licensed under a Creative Commons License.