Minitab

LESSON: POPULATION MEAN HYPOTHESIS TESTING FOR LARGE SAMPLES
This lesson includes an overview of the subject, instructor notes, and example exercises using Minitab.

Population Mean Hypothesis Testing for Large Samples

Lesson Overview

Given a large sample size n (where we will assume $n > 30$), we will conduct hypothesis tests on the true mean μ of a population.

Prerequisites

This lesson requires knowledge from the ***Normal Distribution*** and ***Sampling Distribution of*** $\bar{X}$ (i.e. Central Limit Theorem) lessons. Students need to understand that the distribution of $\bar{X}$ is approximately normally distributed for a large sample size n. Otherwise, the process that we use in this lesson for performing a hypothesis test with a large sample will not make sense.

Learning Targets

This lesson teaches students to:

- Learn how to set up a two-sided hypothesis test (i.e. H_1: $\mu \neq$ value) and a one-sided hypothesis test (i.e. H_1: $\mu >$ value or H_1: $\mu <$ value).
- Perform a hypothesis test by hand and use a standard normal table for a population mean given a large sample size n.
- Perform a hypothesis test in Minitab for a population mean given a large sample size n.
- Learn what a p-value is and how it is calculated.
- Interpret the results of a hypothesis test based on the p-value.

Time Required

It will take the instructor at least 60 minutes in class to introduce this lesson since there is a lot of new terminology (e.g. p-values, significance level) and intricate details in hypothesis testing (e.g. greater than versus less than, alternative hypotheses). Students will need a variety of practice examples to understand the basic hypothesis testing methodology, and they will need to practice setting up a hypothesis test and drawing conclusions on their own. The exercises on the activity sheet will take at least 60 minutes, and they can be used as homework or quiz problems. Some of the examples and exercises in the ***Population Mean Confidence Intervals for Large Samples*** lesson will be repeated in this lesson to show how the two methods are related.

Materials Required

- Minitab 17 or Minitab Express
- Minitab worksheet of sample data, entitled ***HypTestForMean_LargeSample_Lesson.mtw***
- Double-sided coin. You can purchase these at game stores or online. Alternatively, instructions for making your own double-sided coin are available at http://www.instructables.com/id/How-to-make-a-magical-two-headed-coin/.
- Activity Sheet: Jigsaw puzzles (all identical) consisting of 25 to 40 pieces. Alternatively, you can find free online puzzles at http://www.jigzone.com. If you do not have time or internet access in class, then you can use the puzzle data provided in ***HypTestForMean_LargeSample_Lesson.mtw***.

Assessment

The activity sheet contains exercises for students to assess their understanding of the learning targets for this lesson.

Possible Extensions

This lesson introduces students to performing hypothesis tests for a population mean with a large sample size. The instructor may want to cover the ***Population Mean Confidence Intervals for Large Samples*** lesson first since students often find confidence intervals easier to understand than hypothesis testing.

References

Persi Diaconis and coin flipping: *http://news.stanford.edu/news/2004/june9/diaconis-69.html*

Cartoon images: Free images: *http://pixabay.com/en/photos/soda/*

Miscellaneous

Cartoon images: FreePik: *http://www.freepik.com/index.php?goto=2&searchform=1&k=teacher*

Instructor Notes with Examples

Hypothesis Testing

> The process of using probability and statistics to set up an experimental situation and decide whether or not to reject the "status quo" hypothesis based on sample data is called ***hypothesis testing***.

Recall that the **mean of a population**, or the average of an entire set of data, is **μ**. The **sample mean** $\bar{x}$ is the mean of a sample from that population. Often for a given population, we don't know μ, but we can determine $\bar{x}$ by selecting a random sample of data from the population. This lesson shows how to use $\bar{x}$ to make a claim about μ.

Scenario: We have a population, and we want to make a claim regarding the **population parameter** μ, such as "is it larger than a given value?"

- You might ask "Why don't we just compute the population mean rather than calculate an estimate for it?" Good question! Suppose your population of interest is Canada, and you want to know the mean age of the population.
    - Due to lack of time, energy, and money, you cannot obtain the age of every person in Canada.
    - You can select a sample (e.g. a simple random sample – see the ***Sampling*** lesson) and calculate the mean of that sample, $\bar{x}$.

- Your next question may be, "Why don't we just use the sample mean $\bar{x}$ to estimate the population mean μ?" Another good question!
    - We can – but the sample mean $\bar{x}$ may be *quite* different from the population mean μ, even if we obtained the sample correctly.
    - In addition, a single number estimate by itself, such as $\bar{x}$, provides no information about the precision and reliability of the estimate with respect to the larger population.

- Statisticians use the sample statistic $\bar{x}$ and the population or sample standard deviation to provide evidence in support of or against a given value for the population parameter μ. Unlike a **confidence interval** for μ, in which we try to capture the value of μ within some interval, we are trying to show with some certainty whether μ is equal to or not equal to (or greater than, or less than) some value.

Concepts and Terminology

In many experimental situations, the conclusion to be drawn from an experiment is a choice between ***two contrary hypotheses***. Here are examples:

- In a court trial, either the accused is found to be ***innocent*** or ***guilty***.
- When a pharmaceutical company develops a new medication, either the medication is ***effective*** or ***ineffective***.
- When you purchase a 12 oz. can of soda, either it ***contains 12 oz.*** of soda or it ***does not contain 12 oz.*** of soda; or it has ***at least 12 oz.*** or ***less than 12 oz.***

Formally, the two contrary hypotheses are:

- ***Null hypothesis*** **(H_0):** It is the "status quo" or "prior belief" hypothesis. Similar to the U.S. judicial system, the null hypothesis is assumed to be true unless proven otherwise.
- ***Alternative hypothesis*** **(H_1 or H_a):** It is the alternative or contradictory hypothesis to H_0. We reject the null hypothesis in favor of the alternative hypothesis only if there is convincing statistical evidence against H_0. The alternative hypothesis is sometimes referred to as the research hypothesis.

The two types of hypothesis tests, based on the alternative hypothesis **(H_1)**, are:

- ***Two-sided, or two-tailed, tests:*** When you want to detect a difference on either side of the mean, the test is said to be two-tailed and takes the form H_1: $\mu \neq$ value.
- ***One-sided, or one-tailed, tests:*** When you want to detect a difference on only one side of the mean, the test is said to be one-sided and takes the form H_1: $\mu >$ value or H_1: $\mu <$ value.

The hypothesis test for the population mean μ is one of the following:

(1) H_0: $\mu = \mu_0$ versus H_1: $\mu < \mu_0$ (2) H_0: $\mu = \mu_0$ versus H_1: $\mu > \mu_0$ (3) H_0: $\mu = \mu_0$ versus H_1: $\mu \neq \mu_0$

Example 1

When you purchase a 12 oz. can of soda, the "status quo" hypothesis is that it actually contains 12 oz. of soda. The null hypothesis is that the can contains 12 oz. of soda, and the alternative hypothesis is that it does not contain 12 oz. of soda. We write the null and alternative hypotheses as:

$$H_0: \mu = 12 \text{ oz.}$$
$$H_1: \mu \neq 12 \text{ oz.}$$

Additional hypothesis testing facts:

- A ***statistical hypothesis*** is typically a statement about a **population parameter**, such as μ or σ. We do not test sample statistics, such as $\bar{x}$, in a hypothesis test.
- In making a choice to reject or not to reject the assumed null hypothesis H_0, the probability of committing an error is inevitable. There is always a chance that we will make an incorrect decision, even with a reliable sample.
- Unless we are able to collect data from an entire population, which is often impossible or impractical, the truth of a statistical hypothesis is never known with absolute certainty. In **Example 1**, if we reject or do not reject H_0, there is a chance in either case that we have made a mistake.
- We take ***random samples*** from the population of interest to draw conclusions about its population parameters, such as μ.

Example 2

Let the random variable X represent the amount of soda in a 12 oz. can. We are told that X has a normal distribution with mean μ = 12 oz. (or more) and standard deviation σ = 0.1 oz. I have friends who drink 3 or more cans of soda in a day. They have told me that many of the cans seem to contain less than 12 oz.

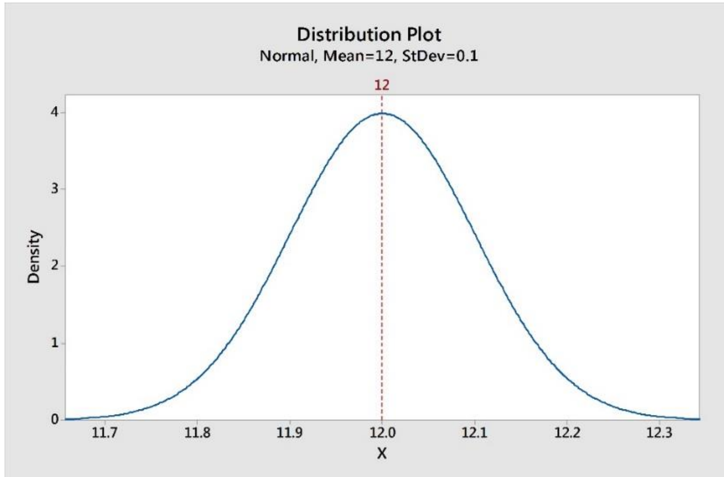
Suppose we have doubts that the "true" mean amount of soda is 12 oz., so we set up the following hypothesis test to investigate whether the mean amount is actually less than 12 oz.:

H_0: $\mu = 12$ (the "status quo" where the population average is 12 oz. or *at least* 12 oz.)
H_1: $\mu < 12$ (the population average is less than 12 oz.)

Think for a moment – how would you go about "proving" that the mean amount of soda in these cans is less than 12 oz.? Any suggestions?

In order to either reject or not reject H_0: $\mu = 12$, we need ***evidence*** that supports one of the two choices. In hypothesis testing, our evidence comes from collecting random samples of data from the population. Suppose we buy 100 cans of soda (from multiple stores) and measure the amount of soda in each can.

After sampling 100 cans, we compute the sample mean $\bar{x}$. Supposing the standard deviation is 0.1 oz., here are three possible sample means to consider and the hypothesis testing conclusions for each of them.



$\bar{x} \cong \mathbf{11.5}$

- Think about the likelihood of obtaining 100 cans of soda with an average of 11.5 oz. Without any statistical calculations, it is reasonable to assume that the null hypothesis is not true.
- Statistically speaking, how likely is it to obtain a sample mean of $\bar{x} \cong 11.5$ if the true population mean is μ = 12? The probability of this happening is extremely small.
- We would reject the null hypothesis in this case, in favor of the alternative hypothesis that the true population mean is less than 12 oz. In rejecting H_0, we could be making an incorrect decision; but if we are, the probability is extremely small.

$\bar{x} \cong \mathbf{12.01}$

- Without any statistical calculations, it is reasonable to assume the null hypothesis is true since the sample mean is greater than 12 oz.
- Statistically speaking, how likely is it to obtain a sample mean of $\bar{x} \cong 12.01$ if the true population mean is μ = 12? The probability is fairly high.
- We would not reject the null hypothesis. This does not mean that we "accept" the null hypothesis as true; we just do not have enough evidence to reject it based on our sample.
- In not rejecting H_0, we could be making an incorrect decision; but if we are, the probability is small.

$\bar{x} \cong \mathbf{11.975}$

- Now we are in a grey area – the sample mean is less than 12 oz., but does not provide strong evidence for rejecting H_0.
- Unlike the other two cases, we now need to conduct statistical calculations to determine how likely it is to obtain a sample mean of $\bar{x} \cong 11.975$ if the true population mean is μ = 12.

- If μ is really 12, there is a chance that we could obtain a sample mean of 11.975 oz. Are we willing to take that chance of incorrectly rejecting the null hypothesis?

> If we reject the null hypothesis, we **do not prove** the alternative hypothesis is true. We merely state there **is sufficient evidence** to reject the null hypothesis.
>
> If we do not reject (i.e. fail to reject) the null hypothesis, we **do not prove** the null hypothesis is true. We merely state there **is not sufficient evidence** to reject the null hypothesis.
>
> Unfortunately, whatever the decision, there is **always a chance we made an error**!

Example 3

Instructor's note: Purchase or construct a double-sided coin (see lesson overview). Before class, secretly choose a student to help you conduct this experiment. Give the double-sided coin to this student and when you come to this example in class, ask for a volunteer to give you a coin to use in the following hypothesis test. Choose the coin that your student is secretly holding. The class is more likely to believe it's a fair coin going into this experiment if it comes from a student.

For students: Suppose I choose a coin from someone in this class. There are two contrary hypotheses about this coin – it is either a fair coin (unbiased) or not a fair coin (biased). What is the null hypothesis? Is the status quo hypothesis that the coin is fair or not fair? In general, when you get a coin as change – do you think it is fair or not fair?

Can someone in class lend me a coin for the following hypothesis test?

H_0: Coin is fair
H_1: Coin is not fair

If we intend to reject the null, we better have some evidence to support such a conclusion. What would be "enough" evidence for you to reject the null hypothesis above?

- 2 heads out of 3 tosses? At 38%, not unusual.
- 4 heads out of 5 tosses? At 16%, not that unusual.
- 5 heads out of 6 tosses? At 9%, a little bit unusual, but certainly possible. Is anyone willing to reject H_0 yet?
- 8 heads of 10 coin tosses? At 4% is anyone willing to reject H_0?

Note that you can calculate the percentages above using Minitab:

- Minitab 17: **Calc > Probability Distributions > Binomial**
- Minitab Express: **Statistics > Probability Distributions > Probability Density Function**

Instructor's note: When I toss the coin in the air and catch it in my hand, I walk around and show the students the outcome (heads). After I toss 4 or 5 heads in a row, the students start to realize something may be amiss. After all, according to Stanford News "Persi Diaconis trained his thumb to flip a coin and make it come up heads 10 out of 10 times." In order to appease them, I ask the student who supplied the coin to flip it also. After 7 or 8 heads, someone in class usually asks to inspect the coin. It is a great exercise in considering the amount of evidence needed, with respect to the probability of the event, to reject the null hypothesis. Plus, it is easy to calculate the probability of getting *n* heads in a row – it's just 0.5^n. For example, $0.5^4 = 0.063$, $0.5^5 = 0.031$, $0.5^6 = 0.016$, and $0.5^7 = 0.008$.

Errors in Hypothesis Testing

There is ALWAYS a chance that the results of a hypothesis test will be incorrect!

- In the U.S. judicial system, the following errors can occur:
  - Type I Error: The accused is not guilty, but the jury says the person guilty.
  - Type II Error: The accused is guilty, but the jury says the person is not guilty.

		Truth: Not Guilty	**Truth: Guilty**
Jury's Decision Based on the Evidence	**Not Guilty**	Correct Decision	***Type II Error***
	Guilty	***Type I Error***	Correct Decision

- In statistical hypothesis testing, the following errors can occur:
  - Type I Error: The null hypothesis H_0 is true, but the researcher rejects H_0.
  - Type II Error: The null hypothesis H_0 is false, but the researcher does not reject H_0.

		Truth	
		H_0 true	**H_0 false**
Experimenter's Decision Based on the Sample	**Do not reject H_0**	Correct Decision	***Type II Error***
	Reject H_0	***Type I Error***	Correct Decision

α = *P*(Type I Error) = *P*(Reject H_0 when H_0 is actually true)

β = *P*(Type II Error) = *P*(Do not reject H_0 when H_0 is actually false)

- We would like the probability of committing either one of these errors to be as small as possible. Unfortunately, decreasing the probability of committing one type of error only increases the probability of committing the other type of error.
- In an introductory statistics course, we typically focus on the **Type I Error** – rejecting the null hypothesis when the null hypothesis is actually true.
- The ***significance level α*** of a hypothesis test is the probability of committing a **Type I Error**.
- What's the "ideal" value of α? Since we do not want to commit a Type I Error, ideally we want $\alpha = 0$. Why can't we allow α to be 0; what will happen?
- A commonly used value for **α = 0.05.**

What does a significance level of α = 0.05 mean? It means that if H_0 is actually true and the hypothesis test is repeated on different random samples of data from the same population, then we would expect H_0 to be incorrectly rejected 5% of the time.

Calculating the z-score

From the random sample of data, we can determine $\bar{x}$, which we **standardize** in the following manner to determine a "***standardized"* test statistic** for the hypothesis test. The resulting standardized test statistic is a ***z*-score**:

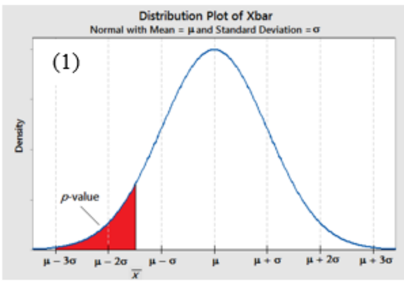
$$z_0 = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

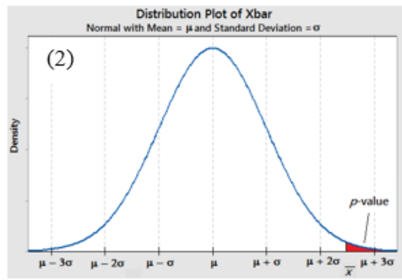
Calculating the *p*-value

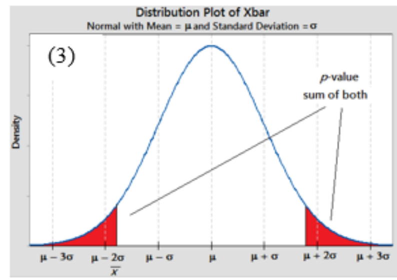
(1) H_0: $\mu = \mu_0$ versus **H_1: $\mu < \mu_0$** – the *p*-value is the region to the **left of the test statistic**

(2) H_0: $\mu = \mu_0$ versus **H_1: $\mu > \mu_0$** – the *p*-value is the region to the **right of the test statistic**

(3) H_0: $\mu = \mu_0$ versus **H_1: $\mu \neq \mu_0$** – the *p*-value is sum of **both regions**







Example 4

A quality control engineer works at a cereal company, and her job is to confirm that the **true mean fill weight μ** of cereal boxes filled on a given day is $\mu = 12$ oz. If the company puts more than 12 oz. of cereal in a box, then it is losing money on overfilled boxes. Alternatively, if the company does not fill a box with at least 12 oz., then customers may become upset and discontinue buying this brand of cereal. First and foremost, the quality engineer wants to ensure the boxes are not underfilled.

The engineer draws a simple random sample of 100 boxes from the population of all boxes filled that day. The **population standard deviation** is known to be **$\sigma = 0.1$ oz.**

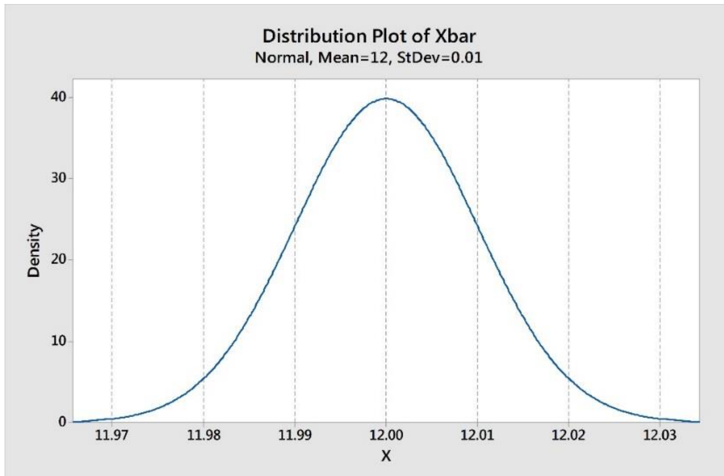
- Let $x_1, x_2, \ldots, x_{100}$ be the individual weights of the simple random sample of 100 cereal boxes.
- The value $\bar{x}$ is a sample average weight of a random sample of 100 cereal boxes. Every random sample of 100 boxes will produce a different sample average weight.
- Let $\bar{X}$ represent the distribution of all $\bar{x}$'s from different random samples from the population of all cereal boxes produced that day.
- The desired hypothesis test for this example is:

$$H_0: \mu = 12 \text{ oz.}$$
$$H_1: \mu < 12 \text{ oz.}$$

- From the Central Limit Theorem, we know that distribution of the sample mean $\bar{X}$:

- Is *approximately* normally distributed since the sample size is large (n = 100).
- Has mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$.

- **Assuming H_0 is true**, then:
  - The mean and standard deviation of $\bar{X}$ are $\mu_{\bar{X}}$ = 12 oz**.** and $\sigma_{\bar{X}} = \frac{0.1}{\sqrt{100}} = 0.01$ oz., respectively.
  - The approximate distribution of $\bar{X}$ for sample size n = 100 is:



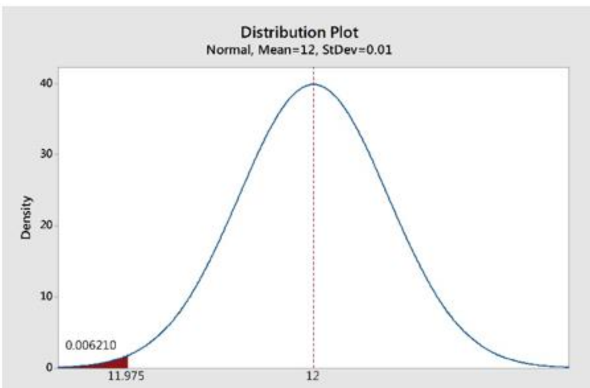
Let's consider one sample of 100 cereal box weights: $x_1 \approx$ 12.006 oz., $x_2 \approx$ 11.998 oz., ..., $x_{100} \approx$ 12.014 oz.

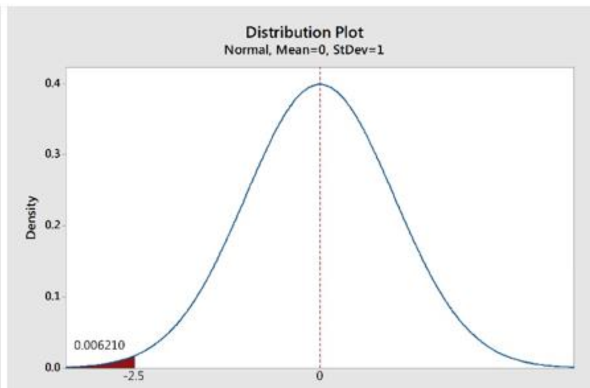
- How do we decide whether or not to reject the null hypothesis **H_0: μ = 12**? We start by determining a ***test statistic*** with our sample data.
  - A **test statistic** is the ***evidence*** in hypothesis testing. The most natural choice for a test statistic of the population mean μ is the sample mean $\bar{x}$.
  - The average of the 100 cereal box fill weights is approximately $\bar{x} = 11.975$ oz.
- If the evidence, $\bar{x}$, is **greater than, the same as, or even "close to" 12 oz.**, then we **would not reject the null hypothesis H_0: μ = 12.**
- Alternatively, if the evidence, $\bar{x}$, is **"much less than" 12 oz.**, then we would **reject the null hypothesis H_0: μ = 12**, in favor of the alternative hypothesis **H_1: μ < 12**.
- How do we determine **"close to"** or **"much less than?"** We haven't defined this numerically yet. Basing our decision on the test statistic alone does not give us enough information to make a decision.
- We need to use the standard deviation $\sigma_{\bar{X}}$ of the sample distribution of the mean $\bar{X}$ to convert the test statistic to a ***standardized test statistic***, which is a ***z-score***.

  The standardized test statistic for $\bar{x} = 11.975$ oz. is:

$$\boldsymbol{z} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{11.975 - 12}{\frac{0.1}{\sqrt{100}}} \cong \mathbf{-2.5}$$

- **Assuming that H_0 is true,** the *z*-score tells us that the test statistic $\bar{x}$ is 2.5 standard deviations to the left of the hypothesized population mean of μ = 12.
- Again, **assuming that H_0 is true and μ = 12**, what's the probability of obtaining **$\bar{x}$ = 11.975** for the random sample of 100 data points? That is, if μ = 12, what is the probability of obtaining a sample mean from the distribution of $\bar{X}$ that is 2.5 standard deviations away from μ = 12? Would you agree that a distance of 2.5 standard deviations makes the null hypothesis, H_0: μ = 12, seem unlikely?
- The actual probability of obtaining a *z*-score of -2.5 is $P(Z < -2.5) \approx 0.0062$. This probability is called a ***p*-value**. Thus, the likelihood of obtaining **$\bar{x}$ = 11.975** given **μ = 12** is **0.0062**. This probability is too low to believe that H_0 is actually μ = 12.





The ***p*-value**, or **observed significance level**, is the probability of obtaining a test statistic that is at least as extreme as the calculated value if H_0 is true.

The **smaller the *p*-value**, the **more contradictory the test statistic is to H_0**. The null hypothesis should be rejected if the *p*-value is "sufficiently small," using the given significance level α. If no significance level is given, most statisticians assume α = 0.05.

- ✓ Reject H_0 if the *p*-value is LESS THAN α.
- ✓ Do not reject H_0 if the *p*-value is GREATER THAN α.

A hypothesis testing saying to help you remember what to do: "**If *p*-value is low, then reject H-0**."

"Not rejecting H_0 " can also be referred to as "failing to reject H_0."

- Given a probability for committing a Type I Error of less than α = 0.05, can we reject H_0? The answer is YES. The actual probability of committing a Type I Error given μ = 12 with test statistic $\bar{x} = 11.975$ is 0.0062, which is less than α = 0.05.
- Given a probability for committing a Type I Error of less than α = 0.01, can we reject H_0? The answer is again YES since the *p*-value is less than α.
- Given a probability for committing a Type I Error of less than α = 0.001, can we reject H_0? The answer is NO. The *p*-value 0.0062 is greater than the allowable probability of a Type I Error α = 0.001.

We can determine the test statistic and accompanying *p*-value in Minitab:

1 Open the 1-sample Z dialog box.
   - Minitab 17: **Stat > Basic Statistics > 1-Sample Z**
   - Minitab Express Mac: **Statistics > 1-Sample Inference > Z**
   - Minitab Express PC: **STATISTICS > One Sample > Z**

2 From the drop-down list, select **Summarized data**.

3 In **Sample size**, enter *100*.

4 In **Sample mean**, enter *11.975*.

5 In **Known standard deviation**, enter *0.1*.

6 Select **Perform hypothesis test**. In **Hypothesized mean**, enter *12*.

7 Select **Options**.

8 Choose the desired **Alternative hypothesis**.
   - Minitab 17: Choose **Mean < hypothesized mean**
   - Minitab Express Mac: Choose **Mean < hypothesized value**
   - Minitab Express PC: Choose **Mean < hypothesized value**

9 Click **OK** in each dialog box.

Minitab returns the z-score and *p*-value as previously determined by hand. In addition, the Minitab output allows us to see if we chose the correct alternative hypothesis.

One-Sample Z

```
Test of μ = 12 vs < 12
The assumed standard deviation = 0.1

  N     Mean  SE Mean  95% Upper Bound      Z      P
100  11.9750   0.0100          11.9914  -2.50  0.006
```

Conclusion: Since the *p*-value is less than α, we **reject H_0**, in favor of $H_1 = \mu < 12$ oz.

Example 5

Let's use the same cereal box scenario as in **Example 4**, but assume the sample mean of data values $x_1, x_2, \ldots, x_{100}$ is approximately $\boldsymbol{\bar{x} = 11.989}$ **oz**.

First, the standardized test statistic is:

$$\boldsymbol{z} = \frac{11.989 - 12}{\frac{0.1}{\sqrt{100}}} \cong \boldsymbol{-1.10}$$

Thus, 11.989 oz. is -1.10 standard deviations to the left of μ = 12. It is not highly unlikely to obtain $\bar{x} = 11.989$ given the true mean is μ = 12.

The probability of obtaining $\bar{x} = 11.989$ is the *p*-value:

$$\boldsymbol{p} = P(Z < -1.10) \cong 0.1357$$

In Minitab, we obtain:

One-Sample Z

```
Test of μ = 12 vs < 12
The assumed standard deviation = 0.1

  N     Mean  SE Mean  95% Upper Bound      Z      P
100  11.9890   0.0100          12.0054  -1.10  0.136
```

Conclusion: Using α = 0.05, we would **NOT reject the null hypothesis** H_0: μ = 12 since the *p*-value = 0.1357 is greater than α.

Example 6

Every child loves getting his or her first bicycle. However, many parents dread the painful task of reading through confusing instructions to put that bike together. To make sure a parent leaves enough time for bike building before its exciting reveal, wouldn't it be nice if the instruction manual provided a reliable estimate for the average time it takes to assemble a bike?

Suppose the instruction manual claims that the true mean time to assemble the bike is μ = 10 minutes with standard deviation σ = 1.2 minutes.

You personally assemble the bike, and it takes you 10 ½ minutes. Although this time is "close" to $\mu = 10$ minutes (given that the standard deviation is 1.2 minutes), you start to wonder if the true average time is really 10 minutes or not? You then set up the following hypothesis test:

$$H_0: \mu = 10$$
$$H_1: \mu > 10$$

Suppose you obtain a random sample of 50 customer assembly times, with an average of 10.4 minutes. Based on this evidence, can you reject H_0 at the $\alpha = 0.01$ level of significance?
Solution: This is a 1-sample Z hypothesis test since the sample size $n = 50$ is large, and the Central Limit Theorem guarantees us that $\bar{X}$ is normally distributed.

Assuming H_0 is true, the **standardized test statistic** is:

$$\boldsymbol{z} = \frac{10.4 - 10}{\frac{1.2}{\sqrt{50}}} \cong \mathbf{2.36}$$

The test statistic's ***p*-value** is:

$$\boldsymbol{p} = P(Z > 2.36) \cong \mathbf{0.0092}$$

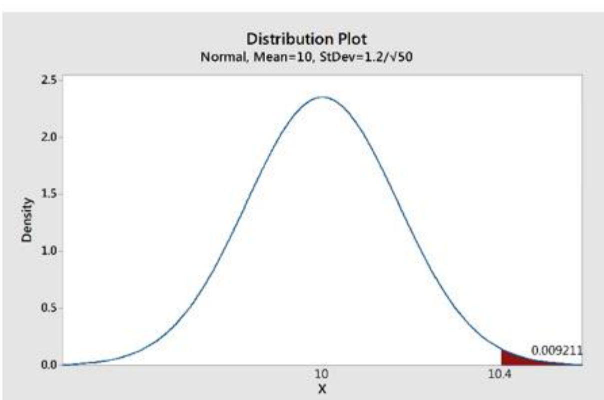
In Minitab, we obtain:

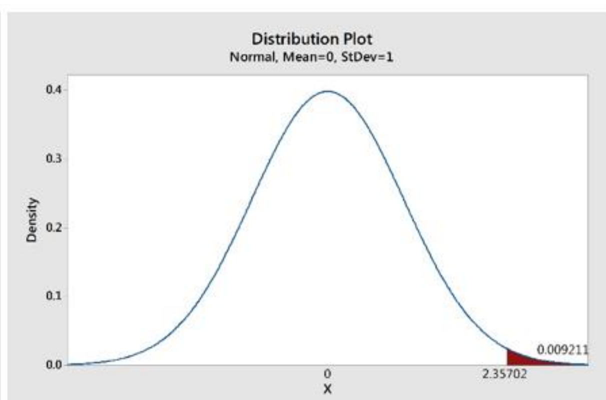
One-Sample Z

```
Test of μ = 10 vs > 10
The assumed standard deviation = 1.2

 N    Mean  SE Mean  95% Lower Bound     Z      P
50  10.400    0.170           10.121  2.36  0.009
```

Conclusion: Using the $\alpha = 0.01$ level of significance, we can reject H_0 in favor of $H_1: \mu > 10$ since 0.0092 is less than α.





Example 7

A bike manufacturer receives many calls per day about the assembly instructions for its popular bike. The main operator working the phone lines suggests to her boss that the true mean time between calls is $\mu = 3$ minutes. The back-up operator disagrees and suggests the mean time is actually greater than 3 minutes, so they want to test the following:

$$H_0: \mu = 3$$
$$H_1: \mu > 3$$

In order to determine whether the mean time is 3 minutes or greater than 3 minutes, the time between phone calls is recorded for $n = 64$ calls and entered into the Minitab worksheet in the column marked "Time Btw Calls."

Solution: According to the Central Limit Theorem, we can assume that $\bar{X}$ is approximately normally distributed since the sample size is $n = 64$. Unlike the previous examples, we are provided with a Minitab column of actual data values, rather than summary statistics. We also are not given a population standard deviation σ, so we will need to estimate σ using the sample standard deviation s.

We can use Minitab (see the ***Describing Data Numerically*** lesson) to calculate the standard deviation:

Descriptive Statistics: Time Btw Calls (mins)

Variable	N	Mean	SE Mean	StDev	Minimum	Maximum
Time Btw Calls (mins)	64	3.176	0.336	2.687	0.103	13.132

Assuming H_0 is true, the **standardized test statistic** is:

$$\boldsymbol{z} = \frac{3.176 - 3}{\frac{2.687}{\sqrt{64}}} \cong \frac{3.176 - 3}{0.336} \cong \mathbf{0.52}$$

The ***p*-value** for the test statistic is:

$$\boldsymbol{p} = P(Z > 0.52) \cong \mathbf{0.30}$$

We can determine the test statistic and accompanying p-value in Minitab:

1 Open the 1-sample Z dialog box.
   - Minitab 17: **Stat > Basic Statistics > 1-Sample Z**
   - Minitab Express Mac: **Statistics > 1-Sample Inference > Z**
   - Minitab Express PC: **STATISTICS > One Sample > Z**

2 In **One or more samples, each in a column**, enter *Time b/w Calls (mins)*.
3 In **Known standard deviation**, enter *2.687*.
4 Select **Perform hypothesis test**. In **Hypothesized mean**, enter *3*.
5 Select **Options**.
6 Choose the desired **Alternative hypothesis**.
   - Minitab 17: Choose **Mean > hypothesized mean**
   - Minitab Express Mac: Choose **Mean > hypothesized value**
   - Minitab Express PC: Choose **Mean > hypothesized value**
7 Click **OK** in each dialog box.

One-Sample Z: Time Btw Calls (mins)

```
Test of μ = 3 vs > 3
The assumed standard deviation = 2.687


Variable              N   Mean  StDev  SE Mean  95% Lower Bound     Z      P
Time Btw Calls (mins)  64  3.176  2.687    0.336            2.623  0.52  0.300
```

Conclusion: Since no significance level is given, we can assign $\alpha = 0.05$. Since the *p*-value is greater than α, then we **do NOT reject H_0**.

Example 8

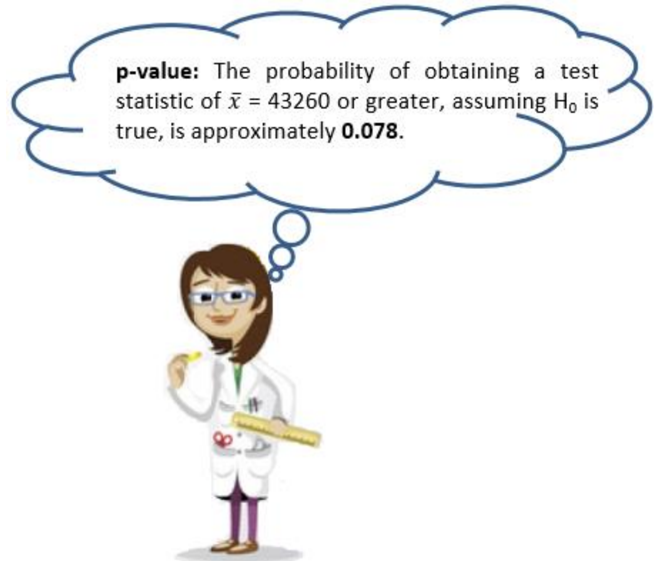
The local newspaper reports that the average salary of all assistant professors at College U is \$42,000. The president of College U believes that the average salary of assistant professors is more than \$42,000. The president takes a random sample of $n = 36$ assistant professors and determines that the mean salary for this sample is $\bar{x} = \$43,260$. At the $\alpha = 0.05$ level of significance, can the president refute the newspaper's claim? The *sample* standard deviation of the $n = 36$ salaries is $s = \$5,320$.

Solution: This is a 1-sample *Z* hypothesis test since the sample size is large, $n = 36$, and the Central Limit Theorem guarantees us that $\bar{X}$ is normally distributed. The null and alternative hypotheses are:

$$H_0: \mu = \$42,000$$
$$H_1: \mu > \$42,000$$

Assuming H_0 is true, the **standardized test statistic** is:

$$z = \frac{43260 - 42000}{\frac{5320}{\sqrt{36}}} \cong 1.42$$



The ***p*-value** for the test statistic is:

$\boldsymbol{p} = P(Z > 1.42) \cong \mathbf{0.078}$

In Minitab, we obtain:

One-Sample Z

```
Test of μ = 42000 vs > 42000
The assumed standard deviation = 5320


 N   Mean  SE Mean  95% Lower Bound     Z      P
36  43260      887            41802  1.42  0.078
```

Conclusion: Assuming H_0 is true, the probability of obtaining a test statistic value of $\bar{x}$ = \$43,260 or greater is approximately p=0.078. Since the *p*-value is greater than $\alpha = 0.05$, then we **do NOT reject H_0**. Therefore, there is not sufficient evidence at the $\alpha = 0.05$ level of significance to reject the null hypothesis that the average salary of all assistant professors at College U is \$42,000 (or less).

Example 9

DVD's are manufactured at a given company with a hole of mean diameter $\mu = 3$ cm. Production history suggests that the population standard deviation is $\sigma = 0.05$ cm and the diameter of these holes is normally distributed. If the hole is too small or too large, the DVD won't snap properly into the DVD case. Determine whether the true mean diameter μ is 3 cm or not at the $\alpha = 0.10$ level of significance using a random sample of $n = 20$ disks with a sample mean of $\bar{x} = 3.015$ cm.

Solution: This is a 1-sample *Z* hypothesis test since the distribution of hole diameters is normally distributed, and this is a ***two-sided*** hypothesis test since the alternative hypotheses is "not equal to."

H_0: $\mu = 3$
H_1: $\mu \neq 3$

Assuming H_0 is true, the **standardized test statistic** is:

$$\boldsymbol{z} = \frac{3.015 - 3}{\frac{0.05}{\sqrt{20}}} \cong \mathbf{1.34}$$

The test statistic's ***p*-value** is:

$$\boldsymbol{p} = 2 * P(Z > 1.34) \cong 2 * 0.09 \cong 0.180$$

We multiply $P(Z > 1.34)$ by 2 since the null hypothesis could be rejected by obtaining an extreme test statistic in either the right or left tail.

In Minitab, we obtain:

One-Sample Z

```
Test of μ = 3 vs ≠ 3
The assumed standard deviation = 0.05


 N    Mean  SE Mean       95% CI          Z      P
20  3.0150   0.0112  (2.9931, 3.0369)  1.34  0.180
```

Conclusion: The probability of obtaining a sample mean of $\bar{x} = 3.015$ cm or greater, or 2.985 or less, given that the true population mean $\mu = 3$ cm is 0.180. Since the *p*-value of 0.180 is greater than $\alpha = 0.10$, then we **do NOT reject H_0**.

Note: If the alternative hypothesis was H_1: $\mu > 3$ instead of H_1: $\mu \neq 3$, then we would have **rejected H_0** at $\alpha = 0.10$ since the one-sided *p*-value is 0.09.