

Chapter 9: Hypothesis Testing with One Sample

What is a hypothesis?

- A **hypothesis** is a statement (or claim) about a property/characteristic of a population (a population parameter).
- **Hypothesis testing** is a procedure, based on sample evidence and probability, for testing claims about a population parameter. It is a way to test the results of a survey or experiment to see if you have statistically significant results.
- **Hypothesis testing can be done on any population parameter**

There are six basic steps for any hypothesis test:

- Step 1: Determine the null and alternative hypotheses.
- Step 2: Verify that all conditions have been met and state the level of significance (α).
- Step 3: Summarize the data into an appropriate test statistic.
- Step 4: Find the p-value by comparing the test statistic to the possibilities expected if the null hypothesis were true **OR** determine the critical value.
- Step 5: Decide whether the result is statistically significant based on the p-value **OR** the critical value.
- Step 6: Report the conclusion in the context of the situation.

Because we use sample data to test hypotheses, we cannot state with 100% certainty that the statement is true; we can only determine whether the sample data support the statement or not. In fact, because the statement can be either true or false, hypothesis testing is based on two types of hypotheses. They are called the **null hypothesis** and the **alternative hypothesis**. These hypotheses contain opposing viewpoints.

9.1: Null and Alternative Hypotheses

Null Hypothesis (H_0):

- A statement of no change, no effect, or no difference.
- Assumed true until evidence indicates otherwise.
- We either reject H_0 or fail to reject H_0 . **WE NEVER 'ACCEPT' H_0 TO BE TRUE!!!**

Alternative Hypothesis (H_a or H_1):

- A statement we are trying to find evidence to support
- Contradictory to H_0 .

Note: H_0 **always** has a symbol with an equal in it. H_1 **never** has a symbol with an equal in it. The choice of the symbols depends on the wording of the hypothesis test.

Symbol choices for H_0 : \leq , \geq , $=$

Symbol choices for H_1 : $>$, $<$, \neq

***Sometimes the researcher will use only $=$ in the null, even though it means \leq or \geq . This is acceptable because we only make the decision to reject or not reject the null hypothesis.

**the math dept at TWU will use only the $=$ in the null hypothesis (H_0)

Three ways to set up the null and alternative hypotheses are:

1. Equal hypothesis versus not equal hypothesis (**two-tailed test**)

H_0 : parameter = null value

H_1 : parameter \neq null value

2. Equal hypothesis versus less than (**left-tailed test**)

H_0 : parameter = null value

H_1 : parameter $<$ null value

3. Equal hypothesis versus greater than (**right-tailed test**)

H_0 : parameter = null value

H_1 : parameter $>$ null value

- It is important to know the type of test you are performing. Is it right-tailed, left-tailed, or two-tailed? **This is answered by the alternative hypothesis** ($<$, $>$, \neq)
- Left- and right-tailed tests are referred to as **one-tailed tests**.

*Remember that the null and the alternative hypotheses are contradictory statements.

Example 1: Determine the null and alternative hypothesis, and the tail-ness.

Reminder: Hypothesis testing can be done on **any population parameter**

(a) The Medco pharmaceutical company has just developed a new antibiotic for children. Among the competing antibiotics, 2% of children who take the drug experience headaches as a side effect. A researcher for the Food and Drug Administration wishes to know if the percentage of children taking the new antibiotic who experience headaches as a side effect is more than 2%.

(b) The blue book value of a used 3-year-old Chevy Corvette is \$37,500. Grant wonders if the mean price of a used 3-year-old Chevy Corvette in the Miami metropolitan area is different from \$37,500.

(c) The standard deviation of the content of a 64-ounce bottle of detergent using an old filling machine was known to be 0.23 ounces. The company purchased a new filling machine and wants to know if there is less variability with the new filling machine.

(d) In an issue of U.S. News and World Report, an article on school standards stated that about half of all students in France, Germany, and Israel take advance placement exams and a third pass. The same article stated that 6.6% of U.S. students take advanced placement exams and 4.4% pass. Test if the percentage of U.S. students who take advanced placement exams is more than 6.6%.

(e) Suppose a consumer group suspects that the proportion of households that have three cell phones is 30%. A cell phone company has reason to believe that the proportion is not 30%. Before they start a big advertising campaign, they conduct a hypothesis test. Their marketing people survey 150 households with the result that 43 of the households have three cell phones.

(f) The packaging on a light bulb states that the bulb will last 500 hours under normal use. A customer advocate would like to know if the mean lifetime of a bulb is less than 500 hours.

9.2: Outcomes and the Type I and Type II Errors

Every time we make a decision (either **Reject H_0** or **Fail to reject H_0**), we have made the correct decision OR we have made an error.

Four Outcomes from Hypothesis Testing

1. Reject the null hypothesis when the alternative one is true.

Reject H_0 when H_1 true

This decision would be correct; also known as **Power**

2. We do not reject the null hypothesis when the null is true.

Fail to Reject H_0 when H_0 true

This decision would be correct.

3. We reject the null hypothesis when the null hypothesis is true.

Reject H_0 when H_0 true

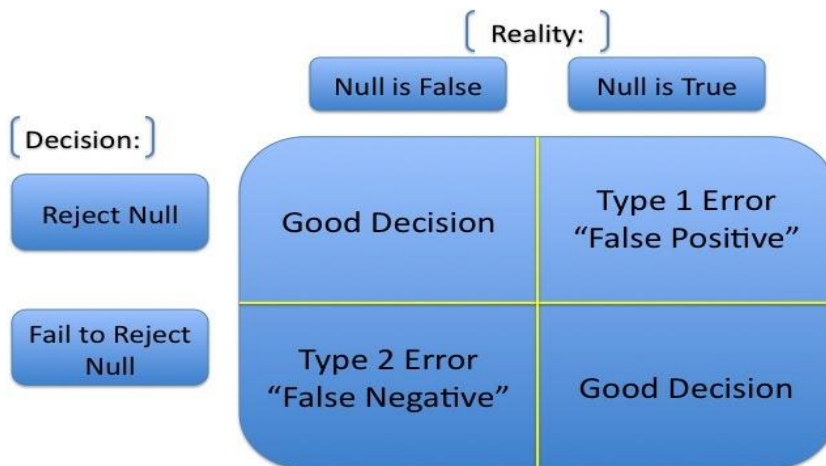
This decision would be incorrect (an error) – known as a **Type I error**.

4. We do not reject the null hypothesis when the alternative hypothesis is true.

Fail to Reject H_0 when H_1 true

This decision would be incorrect (an error) – known as a **Type II error**.

Type 1 and Type 2 Errors



α (alpha), the significance level, is the probability of making a Type I error.

β (beta) is the probability of making a Type II error.

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

$$\beta = P(\text{Type II Error}) = P(\text{Failing to Reject } H_0 \text{ when } H_1 \text{ is true})$$

α and β should be as small as possible because they are probabilities of errors.

Power is the probability we correctly reject the null hypothesis when the alternative is true.

This occurs with a probability $1 - \beta$. **Power** = $1 - \beta = P(\text{Rejecting } H_0 \text{ when } H_1 \text{ is true})$

The logic of statistical hypothesis testing is similar to the “presumed innocent until proven guilty” principle of the U.S. judicial system.

- We assume the null hypothesis is a possible truth until the sample data conclusively demonstrate otherwise.
- The alternative hypothesis is chosen only when the data shows us that we can reject the null hypothesis.

Example 2: Suppose you have the following hypothesis

H_0 : Frank’s rock climbing equipment is safe.

H_1 : Frank’s rock climbing equipment is not safe

1. What is the Type I error?

2. What is the Type II error?

Example 3: Determine the null and alternative hypotheses; explain what it would mean to make a Type I error; explain what it would mean to make a Type II error.

According to Giving and Volunteering in the United States, 2001 Edition, the mean charitable contribution per household in the United States in 2000 was \$1,623. A researcher believes that the level of giving has changed since then.

Example 4: According to the Centers for Disease Control and Prevention, 15.2% of tenth-grade students had tried marijuana. The Drug Abuse and Resistance Education (DARE) program underwent several major changes to keep up with technology and issues facing students in the 21st century. After the changes, a school resource officer (SRO) thinks that the proportion of tenth-grade students who have tried marijuana has decreased from the previous level.

(a) Determine the null and alternative hypotheses.

(b) Assume you fail to reject H_0 , and suppose, in fact, that the proportion of tenth-grade students who have tried marijuana is 14.7%. Was a Type I or Type II error committed?

Power of a Hypothesis Test

Rather than focusing on the risk of making a mistake, many investigators prefer to focus on the chance that their sample will provide the evidence necessary to make the right choice.

The **power** of a hypothesis test is the probability $(1 - \beta)$ of rejecting a false null hypothesis. That is, we decide in favor of the alternative hypothesis given a specific truth about the population. ****Power** is the probability that we **do not** make a Type II error.

There are two features of power that apply to hypothesis tests and that researchers should keep in mind when they plan a study.

- ***The power increases when the sample size is increased.*** The sample statistic is a more accurate estimate of the population value, making it easier to detect a difference between the true population value and the null value (The Law of Large Numbers).
- ***The power increases when the difference between the true population value and the null hypothesis value increases.*** This makes sense because the probability of detecting a large difference is higher than the probability of detecting a small difference.

9.3: Distribution Needed for Hypothesis Testing

Earlier in the course, we discussed sampling distributions of \bar{x} and \hat{p} . There are particular distributions that are associated with hypothesis testing.

- When performing a hypothesis test of a **population mean**, use the **t-distribution**. (Remember, the t-distribution is used when the population standard deviation σ is unknown and the distribution of the sample mean is approximately normal). $\bar{X} \sim t_{df}$
The population parameter is μ . The estimated value (point estimate) for μ is \bar{x} , the sample mean.

- When performing a hypothesis test of a **population proportion**, use the **z-distribution**. (z-distribution is normal and the n is usually large).

$$\hat{p} \sim N \left(p, \sqrt{\frac{p(1-p)}{n}} \right)$$

The population parameter is p . The estimated value (point estimate) for p is \hat{p} , the sample proportion. $\hat{p} = \frac{x}{n}$ where x is the number of successes and n is the sample size.

Step 2: Verify that all conditions have been met

- Always check to see if the sampling method is sound.

For means: $\bar{X} \sim N$ if $X \sim N$
 $\bar{X} \sim N$ if the Central Limit Theorem holds – check sample size, $n \geq 30$

For proportions: X must follow a binomial distribution
 np and nq must be at least 5

9.4: Rare Events, the Sample, Decision and Conclusion

Step 3: Test Statistic

After determining the null and alternative hypotheses, the next step is to calculate the data summary called a *test statistic* that measures the difference between the sample result and the null value.

Test Statistic for a single proportion: $\mathbf{z} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$

Test Statistic for a single mean: $\mathbf{t} = \frac{\bar{x} - \mu}{s / \sqrt{n}}$

We then compute a probability, called a **p-value**.

- A **p-value** is the probability of getting a value of the test statistic that is at least extreme as the test statistic obtained from the sample data, assuming the null hypothesis is true.

Step 4a: Computing the p-Value for the Test

In hypothesis testing, the objective is to decide if we should reject the null hypothesis in favor of the alternative.

We do this by comparing the p-value to a designated standard called the *level of significance* (α) for the test.

If p-value $< \alpha$, then we can reject H_0
If p-value $> \alpha$, then we fail to reject H_0

*The following memory device is useful when using p-value for hypothesis testing.

If the p-value is low, then the null must go. (Reject Null)

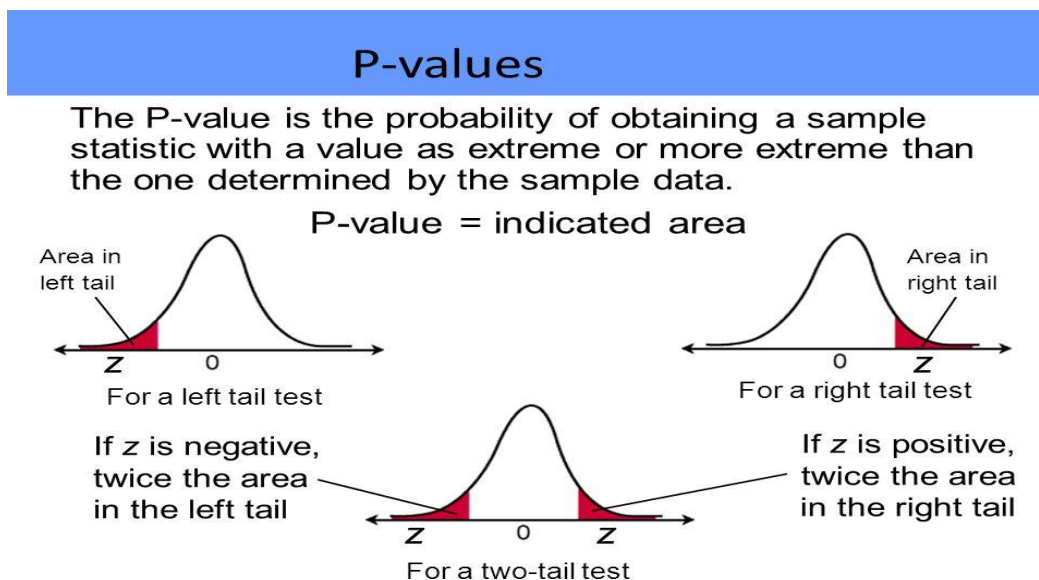
If the p-value is high, then the null must fly. (Fail to Reject Null)

The details of how to find the p-value (the probability of a test statistic as extreme or more extreme than the one observed under the assumption that the null hypothesis is true) depends on the direction specified in the alternative hypothesis ($<$, $>$, \neq) left-tailed, right-tailed, or two-tailed

- For a *less than* alternative hypothesis, find the probability that the test statistic z could have been *equal to or less than* what it is.
- For a *greater than* alternative hypothesis, find the probability that the test statistic z could have been *equal to or greater than* what it is.
- For a *two-tailed* alternative hypothesis, the p -value includes the probability areas in both extremes of the distribution of the test statistic z .

When you calculate the p-value and draw the picture, the p-value is the area in the left tail, area in the right tail, or area split evenly between the two tails. For this reason, we call the hypothesis test left, right, or two-tailed (direction specified in the alternative hypothesis H_1)

A visual is provided below.



Step 5a: Make a decision based on the p -value

The **level of significance**, α (alpha), is the value that is the borderline between when a p -value is small enough to choose the alternative hypothesis, and when it is not small enough.

- If $p\text{-value} \leq \alpha$, we reject the null hypothesis. ("If the P is low, the null must go!")
- If $p\text{-value} > \alpha$, we fail to reject the null hypothesis.

The level of significance (α) is chosen by the researcher BEFORE the experiment/study begins. If the α is not given, use the default $\alpha = 0.05$

The phrase **statistically significant** is used to describe the results when the researcher has decided that the p -value is small enough to decide in favor of the alternative hypothesis.

Step 4b: Computing the Critical Value for the Test

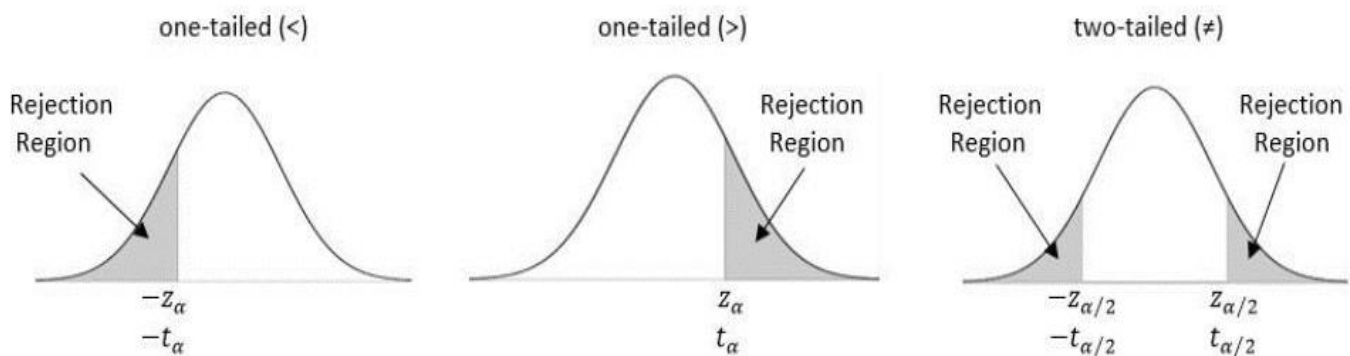
Before technology became so widespread, it was difficult or impossible to compute the p -value in many circumstances, so researchers used a method called the *rejection region approach*.

The **critical region**, or **rejection region**, in a hypothesis test is the region of possible values for the test statistic that would lead to rejection of the null hypothesis.

A boundary of a rejection region is called a **critical value**. The critical value is denoted as z^* or t^* (similar to what we found when computing confidence intervals), while the rejection region is the area more extreme than this value.

If the test statistic is in the shaded region (rejection region), **reject H_0**

If the test statistic is NOT in the shaded region (rejection region), **fail to reject H_0**



- The p -value method and the rejection region method give equivalent conclusions.

Step 5b: Make a decision based on the rejection region

Possible decisions:

- We reject the null hypothesis (Reject H_0)
- We fail to reject the null hypothesis (Fail to Reject H_0)

We never 'ACCEPT' the null hypothesis.

Recall our criminal trial. The trial is the process whereby the jury obtains information (sample data). The jury then deliberates about the evidence (the data analysis). Finally, the jury either convicts the defendant (rejects the null hypothesis) or declares the defendant not guilty (fails to reject the null hypothesis).

Note: the defendant is never declared innocent. That is, we never conclude that the null hypothesis is true.

Step 6: State your conclusion in terms of the problem

If you Reject H_0 : There is sufficient evidence to conclude [statement in H_1].

If you Fail to Reject H_0 : There is not sufficient evidence to conclude [statement in H_1].

9.5: ADDITIONAL INFORMATION AND FULL HYPOTHESIS TEST

Notes:

Remember that the hypotheses are always statements about a parameter, not about a statistic. The whole point is to see what you can infer about an unknown parameter value based on a sample statistic.

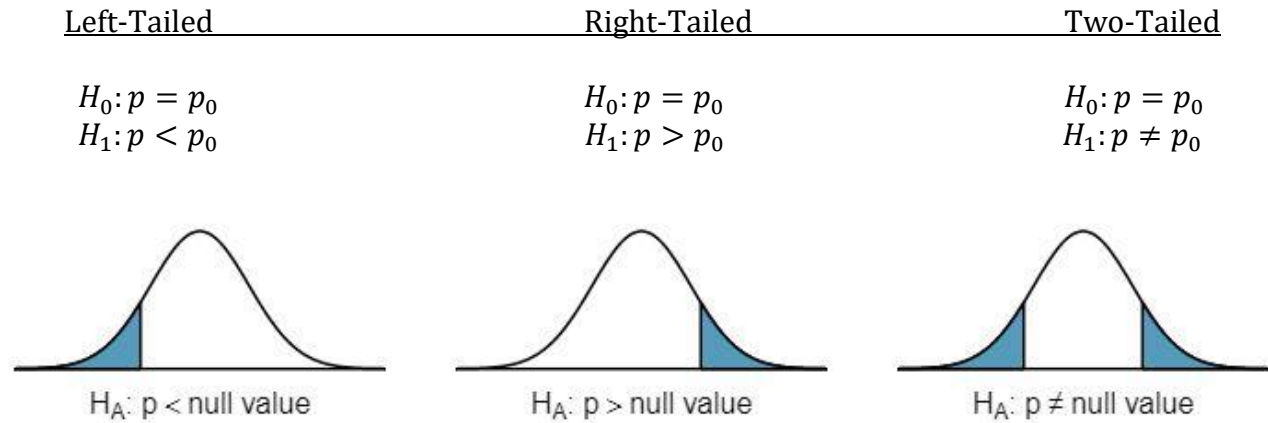
The p -value is NOT the probability that the null hypothesis is true. Rather, it is the probability of obtaining such an extreme sample result (or one even more extreme) if the null hypothesis were true.

Remember we do NOT *accept* the null hypothesis, we *fail to reject* the null.

9.6a: Hypothesis Testing for a Single Proportion

Testing Hypotheses Regarding a Single Population Proportion, p

1. Set up hypothesis:



2. Verify that the following three **requirements** are satisfied:

- (a) The sample is a representative sample
- (b) X must follow a binomial distribution
- (c) np_0 and $n(1 - p_0)$ must be at least 5

3. Compute the **test statistic**:
$$Z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

4a. Find the **p -Value**. Use the z-table.

- The p-value is a probability bounded by the test statistic

4b. OR determine the **critical value**, z^*

- The rejection region (or critical region) is bounded by the critical value(s)

5a. **Decision Rule** for p -Value: If the p -value $< \alpha$, then reject the null hypothesis.

5b. **Decision Rule** for critical value, z^* : If the test statistic is **MORE EXTREME** than the critical value, then reject the null hypothesis. In other words, if the test statistic falls inside a rejection region, then reject the null hypothesis.

6. Conclusion: State your decision and your conclusion in terms of the problem.

If you reject H_0 – There **is** sufficient evidence to conclude [statement in the H_1]

If you Fail to Reject H_0 – There is **not** sufficient evidence to conclude [statement in the H_1]

Example 5: Medics and Teachers want to know if a new vitamin supplement will help decrease the sick absentees of students during the winter. They took a sample of 742 students. They gave the vitamin supplement to the students for the months of August through December and calculated the Proportion of students who were absent due to sickness was close to 8%. **In general**, a given student is absent about 10% of the time. Is the difference statistically large enough ($\alpha = 0.05$) to say that the vitamin supplement reduces absentee sickness?

Step 1: Set up the Hypothesis

Step 2: Verify Requirements

Step 3: Compute Test Statistic

Step 4a: Find the p-value

Step 5a: Decision

Step 6: Conclusion

Step 4b: Find the critical value

Step 5b: Decision

9.6b: Hypothesis Testing for a Single Mean

Testing Hypotheses Regarding a Single Population Mean μ , using the t -distribution

1. Set up **hypothesis**:

<u>Two-Tailed</u>	<u>Left-Tailed</u>	<u>Right-Tailed</u>
$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$

2. Verify that the following **requirements** are satisfied:

- The sample is a representative sample
- One of the following:
 - The population must be normally distributed, OR
 - The sample size needs to be large enough, $n \geq 30$

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

4a. Find the **p –Value**. Use the t -table.

4b. OR determine the **critical value**, t^*

5a. **Decision Rule** for p –Value: If the p –value $< \alpha$, then reject the null hypothesis.

5b. **Decision Rule** for critical value, z^* : If the test statistic is MORE EXTREME than the critical value, then reject the null hypothesis. In other words, if the test statistic falls inside a rejection region, then reject the null hypothesis.

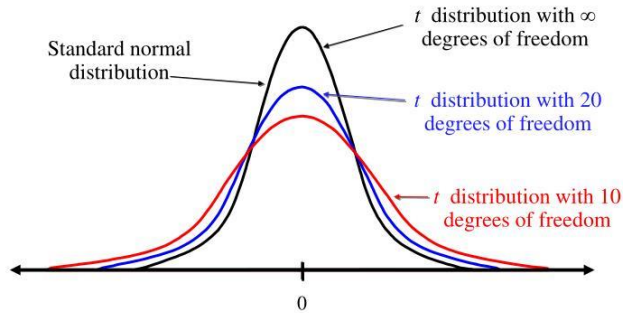
6. **Conclusion**: State your decision and your conclusion in terms of the problem.

If you reject H_0 – There **is** sufficient evidence to conclude [statement in the H_1]

If you Fail to Reject H_0 – There is **not** sufficient evidence to conclude [statement in the

t Distribution

The t -distribution is used when n is **small** and σ is **unknown**.

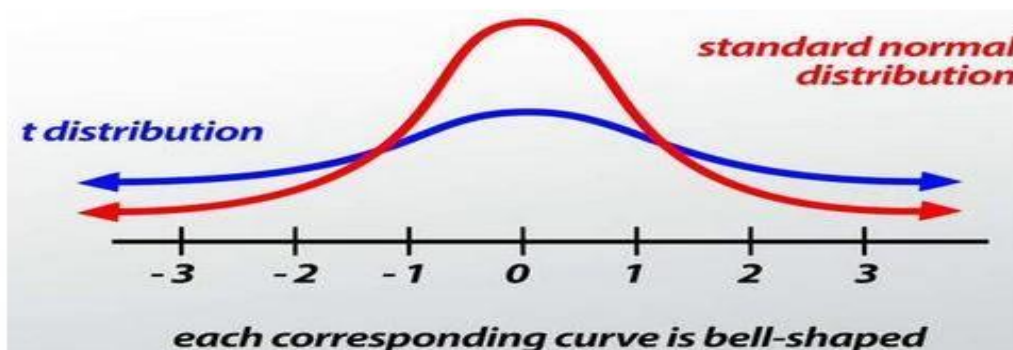


Example 6: Find the t -value for a right-tailed hypothesis test if we have 15 degrees of freedom and $\alpha = 0.10$. What if the alternative hypothesis is left tailed? Two-tailed?

Notice from the example above that the critical value for t ($t^* = 1.341$) is bigger than the corresponding critical value of z ($z^* = 1.28$) – both being the area to the right of 0.10.

This is because the t -distribution has **more spread** than the z -distribution.

If the degrees of freedom we desire are not available in the t -table, we follow the practice of choosing the closest number of degrees of freedom available in the table. In addition, the last row of the t -table provides the z -values from the standard normal distribution. We use these values for situations where the degrees of freedom are more than 2000. This is acceptable because the t -distribution starts to behave like the standard normal distribution as n increases.



Example 7: According to the Centers of Disease Control, the mean number of cigarettes smoked per day by individuals who are daily smokers is 18.1. Do retired adults who are daily smokers smoke less than the general population of daily smokers? To answer this question, we obtain a random sample of 40 retired adults who currently smoke daily and record the number of cigarettes smoked on a randomly selected date. The data result in a sample mean of 16.8 and a standard deviation of 4.7 cigarettes. Is there sufficient evidence at the $\alpha = 0.1$ level of significance to conclude that retired adults who are daily smokers smoke less than the general population of daily smokers?

Step 1: Set up the Hypothesis

Step 2: Verify Requirements

Step 3: Compute Test Statistic

Step 4a: Find the p-value

Step 5a: Decision

Step 6: Conclusion

Step 4b: Find the critical value

Step 5b: Decision

Mixed Problems

Example 8: To test $H_0: \mu = 100$ versus $H_1: \mu \neq 100$, a simple random sample of size $n = 23$ is obtained from a population that is known to be normally distributed.

(a) If $\bar{x} = 104.8$ and $s = 9.2$, compute the test statistic.

(b) If the researcher decides to test the hypothesis at the $\alpha = 0.01$ level of significance, determine the critical value(s).

(c) What is your decision? Why?

(d) State your conclusion in terms of the problem.

(e) Construct a 99% confidence interval to test the hypothesis.

(f) Interpret the interval in part (e).

Example 9: *Time* magazine, in a survey of 507 randomly selected adult American Catholics, 59% answered yes to the question “Do you favor allowing women to be priests?” Test the hypothesis that more than 55% of adult Catholics said yes.
(*Time*, 26 December – 2 January 1995, pp.74-76)

1. State your hypotheses.
2. What conditions are necessary for testing this hypothesis? Are those conditions met?
3. Compute the test statistic.
4. Calculate the p-value.
5. What is the decision using 0.05 as your level of significance?
6. State your conclusion in terms of the problem.
7. Find the rejection region and make a decision.