

AMA 2634

Introduction to Statistics

2024 Fall

Lecture 8

Hypothesis Testing

Researchers are interested in answering many types of questions. For example,

- Is the earth warming up?
- Does a new medication lower blood pressure?
- Does the public prefer a certain color in a new fashion line?
- Is a new teaching technique better than a traditional one?
- Do seat belts reduce the severity of injuries?

These types of questions can be addressed through statistical hypothesis testing, which is a decision-making process for evaluating claims about a population.

Three methods used to test hypotheses:

1. The traditional method
2. The P -value method
3. The confidence interval method



A **statistical hypothesis** is a conjecture about a population parameter. This conjecture may or may not be true.

The **null hypothesis**, symbolized by H_0 , is a statistical hypothesis that states that there is no difference between a parameter and a specific value, or that there is no difference between two parameters.

The **alternative hypothesis** is symbolized by H_1 .



The **alternative hypothesis**, symbolized by H_1 , is a statistical hypothesis that states the existence of a difference between a parameter and a specific value, or states that there is a difference between two parameters.

A medical researcher is interested in finding out whether a new medication will have any undesirable side effects. The researcher is particularly concerned with the pulse rate of the patients who take the medication. Will the pulse rate increase, decrease, or remain unchanged after a patient takes the medication? The researcher knows that the mean pulse rate for the population under study is 82 beats per minute.

The hypotheses for this situation are

$$H_0 : \mu = 82 \quad H_1 : \mu \neq 82$$

This is called a **two-tailed** hypothesis test.

A chemist invents an additive to increase the life of an automobile battery. The mean lifetime of the automobile battery without the additive is 36 months.

In this book, the null hypothesis is always stated using the equals sign. The hypotheses for this situation are

$$H_0 : \mu = 36 \quad H_1 : \mu > 36$$

This is called a **right-tailed** (upper-tailed, one-tailed, one-sided) hypothesis test.

A contractor wishes to lower heating bills by using a special type of insulation in houses. If the average of the monthly heating bills is \$78, her hypotheses about heating costs with the use of insulation are

$$H_0 : \mu = 78 \quad H_1 : \mu < 78$$

This is called a **left-tailed** hypothesis test.

When a researcher conducts a study, he or she is generally looking for evidence to support a **claim**. Therefore, the claim should be stated as the alternative hypothesis, or **research hypothesis**.

A claim, though, can be stated as either the null hypothesis or the alternative hypothesis; however, the statistical evidence can only *support* the claim if it is the alternative hypothesis. Statistical evidence can be used to *reject* the claim if the claim is the null hypothesis.

These facts are important when you are stating the conclusion of a statistical study.

After stating the hypotheses, the researcher's next step is to design the study. The researcher selects the **correct statistical test**, chooses an appropriate **level of significance**, and formulates a plan for conducting the study.

A **statistical test** uses the data obtained from a sample to make a decision about whether the null hypothesis should be rejected.

The numerical value obtained from a statistical test is called the **test value**.

In the hypothesis-testing situation, there are four possible outcomes.

In reality, the null hypothesis may or may not be true, and a decision is made to reject or not to reject it on the basis of the data obtained from a sample.

A **type I error** occurs if one rejects the null hypothesis when it is true.

A **type II error** occurs if one does not reject the null hypothesis when it is false.

Hypothesis Testing



	H_0 true	H_0 false
Reject H_0	Error Type I	Correct decision
Do not reject H_0	Correct decision	Error Type II

The **level of significance** is the maximum probability of committing a type I error. This probability is symbolized by α (alpha). That is,

$$P(\text{type I error}) = \alpha.$$

Likewise,

$$P(\text{type II error}) = \beta \text{ (beta).}$$

Typical significance levels are:

0.10, 0.05, and 0.01

For example, when $\alpha = 0.10$, there is a 10% chance of rejecting a true null hypothesis.

The **critical value, C.V.**, separates the critical region from the noncritical region.

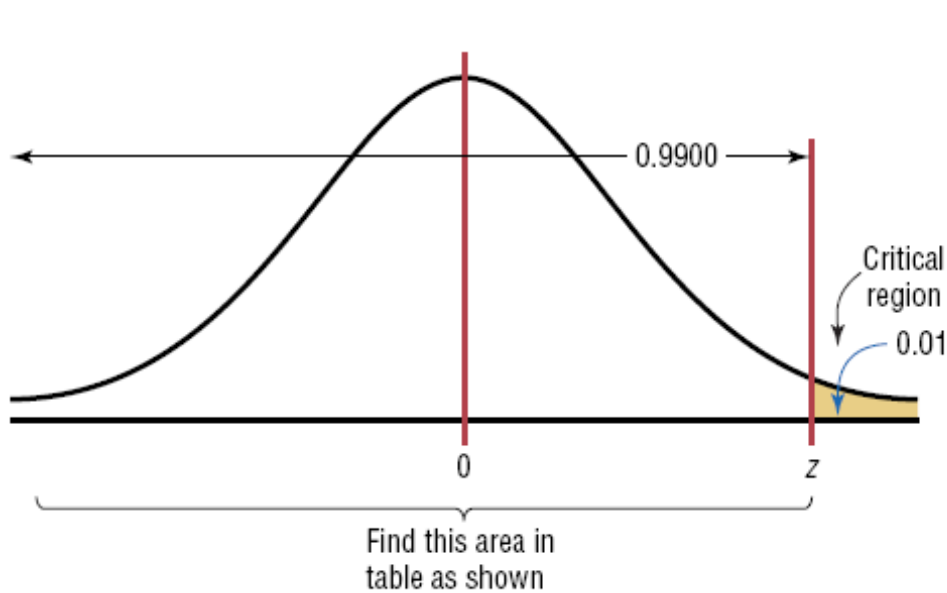
The **critical** or **rejection region** is the range of values of the test value that indicates that there is a significant difference and that the null hypothesis should be rejected.

The **noncritical** or **nonrejection region** is the range of values of the test value that indicates that the difference was probably due to chance and that the null hypothesis should not be rejected.

Hypothesis Testing



Finding the Critical Value for $\alpha = 0.01$ (Right-Tailed Test)

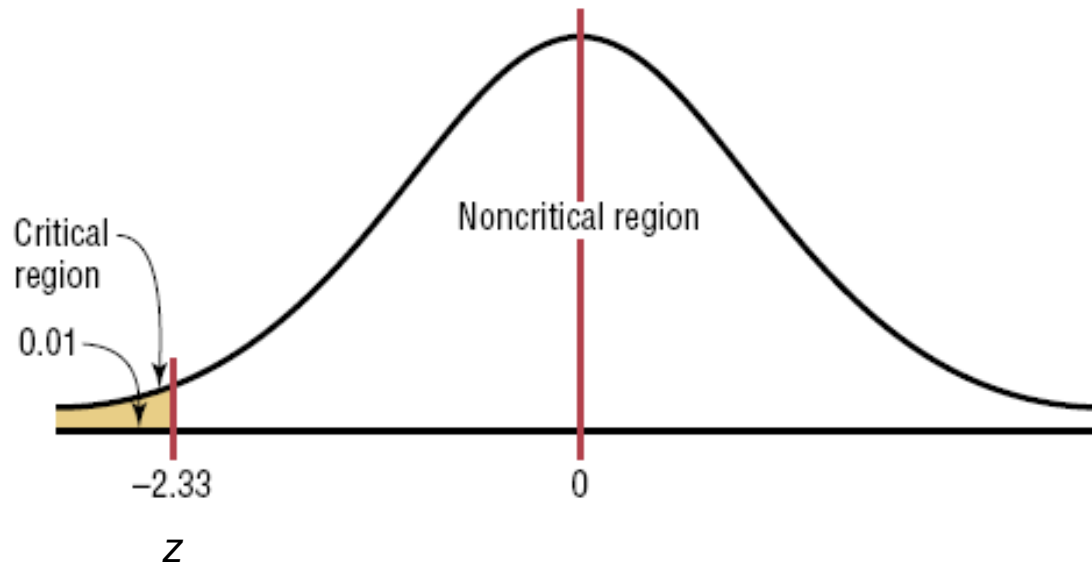


z	0.00	0.01	0.02	0.03	0.04	0.05	...
0.0							
0.1							
0.2							
0.3							
...							
2.1							
2.2							
2.3				0.9901			
2.4							
...							

Closest value to 0.9900

$z = 2.33$ for $\alpha = 0.01$ (Right-Tailed Test)

Finding the Critical Value for $\alpha = 0.01$ (Left-Tailed Test)

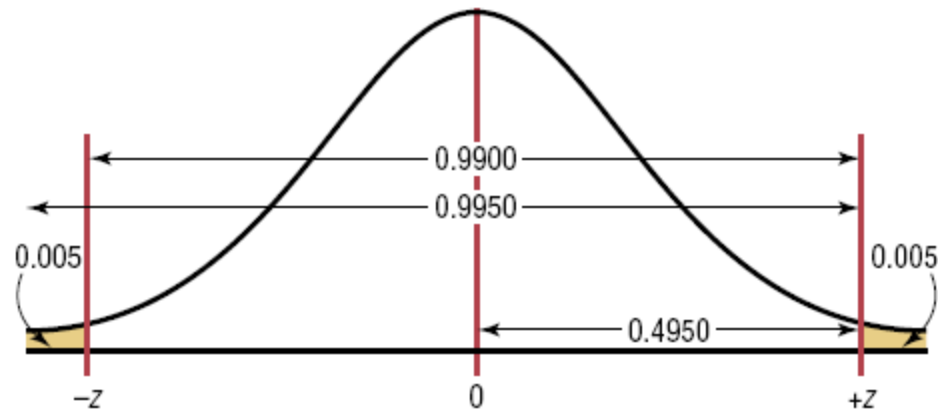


Because of symmetry,
 $z = -2.33$ for $\alpha = 0.01$ (Left-Tailed Test)

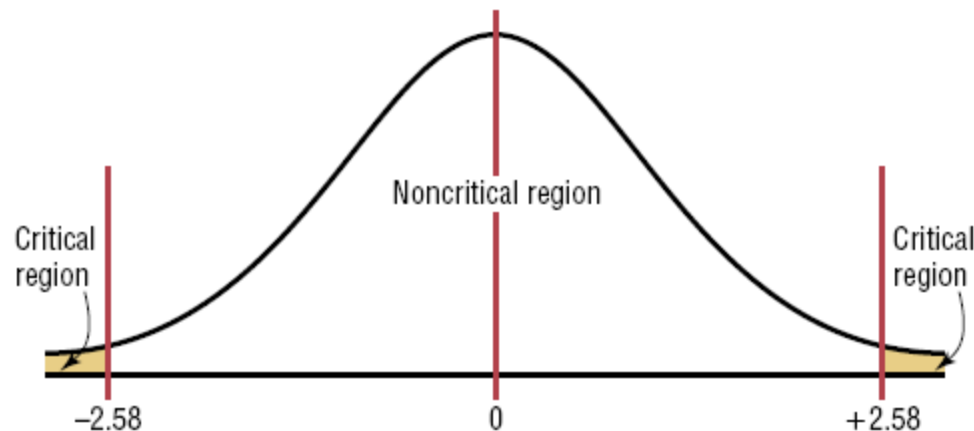
Hypothesis Testing



Finding the Critical Value for $\alpha = 0.01$ (Two-Tailed Test)



$$z = \pm 2.58$$



The z test is a statistical test for the mean of a population. It can be used when $n \geq 30$, or when the population is normally distributed and σ is known.

The formula for the *z test* is

$$z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

where

$\bar{X}$ = sample mean

μ = hypothesized population mean

σ = population standard deviation

n = sample size



A researcher wishes to see if the mean number of days that a basic, low-price, small automobile sits on a dealer's lot is 29. A sample of 30 automobile dealers has a mean of 30.1 days for basic, low-price, small automobiles.

At $\alpha = 0.05$, test the claim that the mean time is greater than 29 days. The standard deviation of the population is 3.8 days.

Step 1 State the hypotheses and identify the claim.

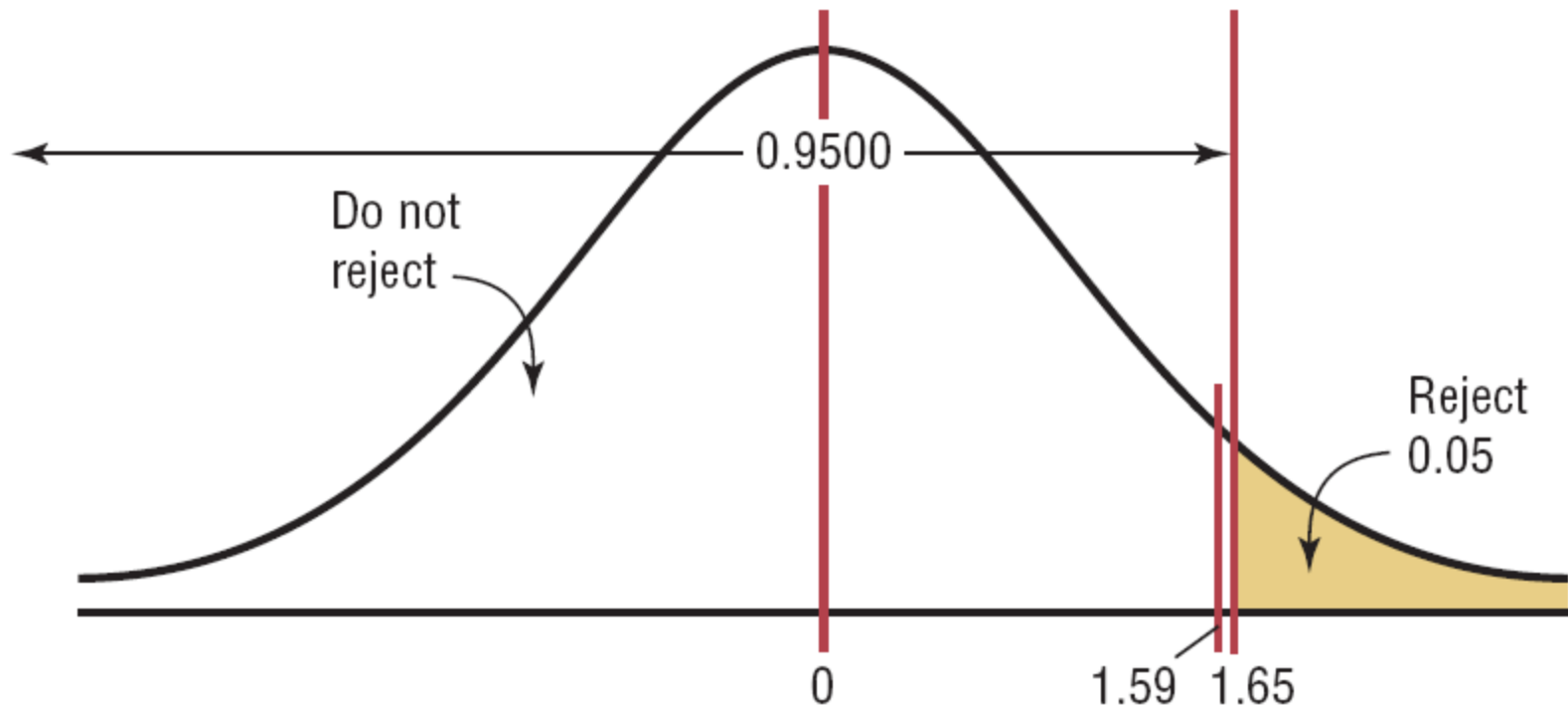
$$H_0: \mu = 29 \quad \text{and} \quad H_1: \mu > 29 \text{ (claim)}$$

Step 2 Find the critical value. Since $\alpha = 0.05$ and the test is a right-tailed test, the critical value is $z = +1.65$.

Step 3 Compute the test value.

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{30.1 - 29}{3.8/\sqrt{30}} = 1.59$$

- Step 4** Make the decision. Since the test value, $+1.59$, is less than the critical value, $+1.65$, and is not in the critical region, the decision is to not reject the null hypothesis.





Step 5 Summarize the results.

There is not enough evidence to support the claim that the mean time is greater than 29 days.



Even though in the above Example the sample mean of 30.1 is higher than the hypothesized population mean of 29, it is not significantly higher.

Hence, the difference may be due to chance.

When the null hypothesis is not rejected, there is still a probability of a type II error, i.e., of not rejecting the null hypothesis when it is false.

Example: Cost of Men's Shoes



A researcher claims that the average cost of men's athletic shoes is less than \$80. He selects a random sample of 36 pairs of shoes from a catalog and finds the following costs (in dollars). (The costs have been rounded to the nearest dollar.) Is there enough evidence to support the researcher's claim at $\alpha = 0.10$? Assume $\sigma = 19.2$.

60 70 75 55 80 55 50 40 80 70 50 95
120 90 75 85 80 60 110 65 80 85 85 45
75 60 90 90 60 95 110 85 45 90 70 70

Step 1: State the hypotheses and identify the claim.

$H_0: \mu = \$80$ and $H_1: \mu < \$80$ (claim)

Example: Cost of Men's Shoes



A researcher claims that the average cost of men's athletic shoes is less than \$80. He selects a random sample of 36 pairs of shoes from a catalog and finds the following costs (in dollars). (The costs have been rounded to the nearest dollar.) Is there enough evidence to support the researcher's claim at $\alpha = 0.10$? Assume $\sigma = 19.2$.

60 70 75 55 80 55 50 40 80 70 50 95
120 90 75 85 80 60 110 65 80 85 85 45
75 60 90 90 60 95 110 85 45 90 70 70

Step 2: Find the critical value.

Since $\alpha = 0.10$ and the test is a left-tailed test, the critical value is $z = -1.28$.

A researcher claims that the average cost of men's athletic shoes is less than \$80. He selects a random sample of 36 pairs of shoes from a catalog and finds the following costs (in dollars). (The costs have been rounded to the nearest dollar.) Is there enough evidence to support the researcher's claim at $\alpha = 0.10$? Assume $\sigma = 19.2$.

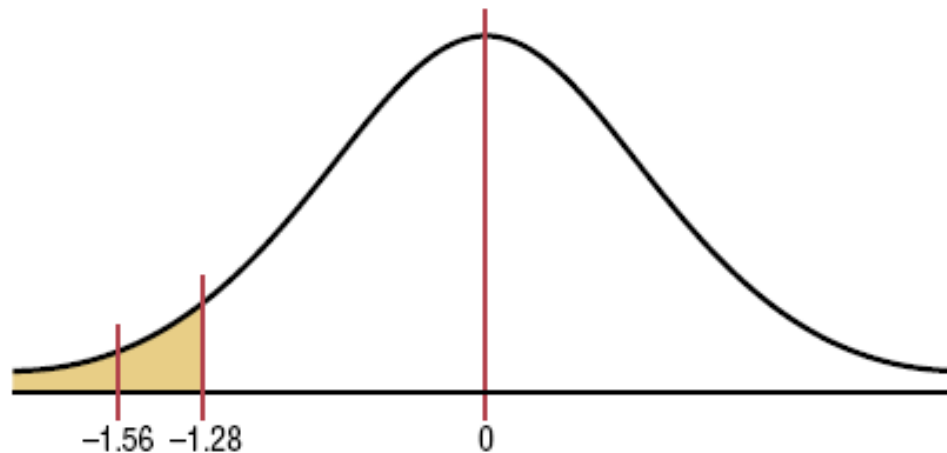
Step 3: Compute the test value.

Using technology, we find $\bar{X} = 75.0$ and $\sigma = 19.2$.

$$z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = \frac{75 - 80}{19.2 / \sqrt{36}} = -1.56$$

Step 4: Make the decision.

Since the test value, -1.56 , falls in the critical region, the decision is to reject the null hypothesis.



Step 5: Summarize the results.

There is enough evidence to support the claim that the average cost of men's athletic shoes is less than \$80.

The Medical Rehabilitation Education Foundation reports that the average cost of rehabilitation for stroke victims is \$24,672. To see if the average cost of rehabilitation is different at a particular hospital, a researcher selects a random sample of 35 stroke victims at the hospital and finds that the average cost of their rehabilitation is \$26,343. The standard deviation of the population is \$3251. At $\alpha = 0.01$, can it be concluded that the average cost of stroke rehabilitation at a particular hospital is different from \$24,672?

Step 1: State the hypotheses and identify the claim.

$$H_0: \mu = \$24,672 \quad \text{and} \quad H_1: \mu \neq \$24,672 \text{ (claim)}$$

Step 2: Find the critical value.

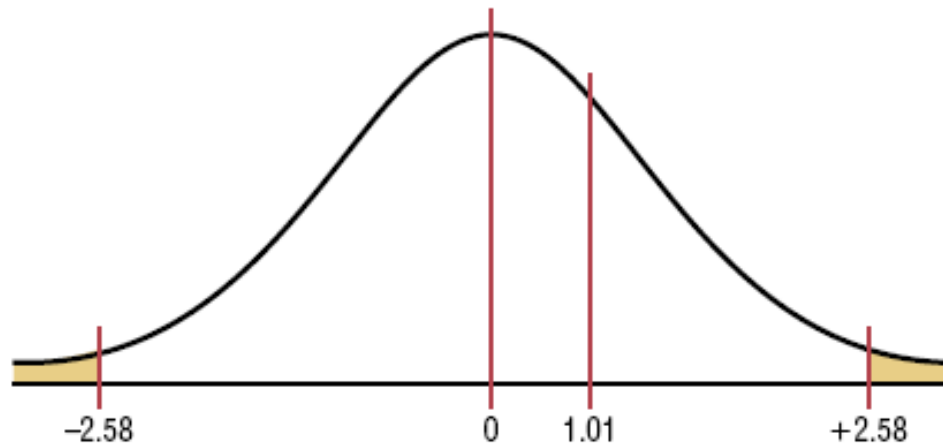
Since $\alpha = 0.01$ and a two-tailed test, the critical values are $z = \pm 2.58$.

Step 3: Find the test value.

$$z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = \frac{26,343 - 24,672}{3251 / \sqrt{35}} = 3.04$$

Step 4: Make the decision.

Reject the null hypothesis, since the test value falls in the critical region.



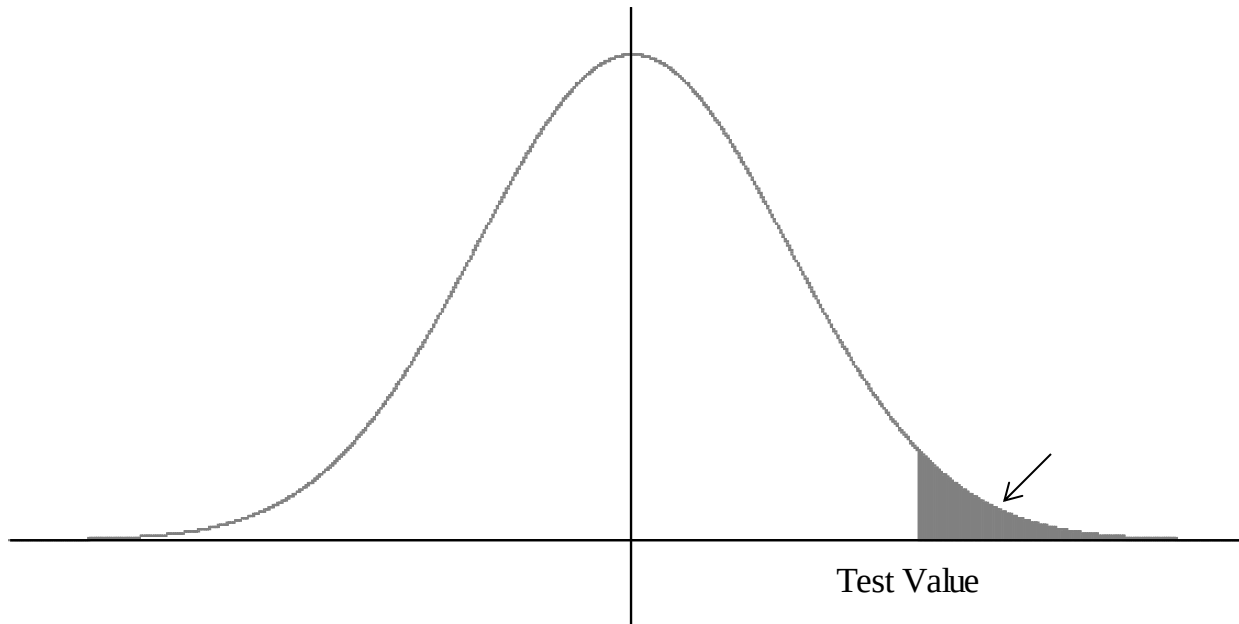
Step 5: Summarize the results.

There is enough evidence to support the claim that the average cost of rehabilitation at the particular hospital is different from \$24,672.

Hypothesis Testing



The **P-value** (or probability value) is the probability of getting a sample statistic (such as the mean) or a more extreme sample statistic in the direction of the alternative hypothesis when the null hypothesis is true.



In this section, the traditional method for solving hypothesis-testing problems compares **z-values**:

- critical value
- test value

The P -value method for solving hypothesis-testing problems compares **areas**:

- alpha
- P -value



A researcher wishes to test the claim that the average cost of tuition and fees at a four-year public college is greater than \$5700. She selects a random sample of 36 four-year public colleges and finds the mean to be \$5950. The population standard deviation is \$659. Is there evidence to support the claim at a 0.05? Use the P -value method.

Step 1: State the hypotheses and identify the claim.

$$H_0: \mu = \$5700 \text{ and } H_1: \mu > \$5700 \text{ (claim)}$$

Step 2: Compute the test value.

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{5950 - 5700}{659/\sqrt{36}} = 2.28$$

Step 3: Find the *P*-value.

Using normal table, find the area for $z = 2.28$.

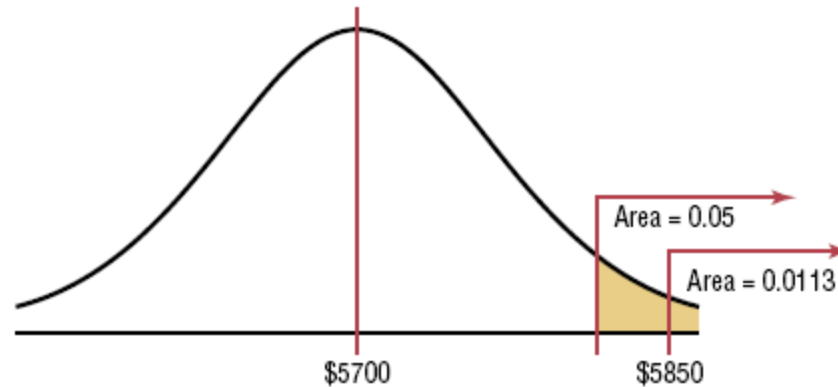
The area is 0.9887.

Subtract from 1.0000 to find the area of the tail.

Hence, the *P*-value is $1.0000 - 0.9887 = 0.0113$.

Step 4: Make the decision.

Since the P -value is less than 0.05, the decision is to reject the null hypothesis.



Step 5: Summarize the results.

There is enough evidence to support the claim that the tuition and fees at four-year public colleges are greater than \$5700.

Note: If $\alpha = 0.01$, the null hypothesis would not be rejected.

A researcher claims that the average wind speed in a certain city is 8 miles per hour. A sample of 32 days has an average wind speed of 8.2 miles per hour. The standard deviation of the population is 0.6 mile per hour. At $\alpha = 0.05$, is there enough evidence to reject the claim? Use the P -value method.

Step 1: State the hypotheses and identify the claim.

$$H_0: \mu = 8 \text{ (claim)} \quad \text{and} \quad H_1: \mu \neq 8$$

Step 2: Compute the test value.

$$z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = \frac{8.2 - 8}{0.6 / \sqrt{32}} = 1.89$$

Step 3: Find the P -value.

The area for $z = 1.89$ is 0.9706.

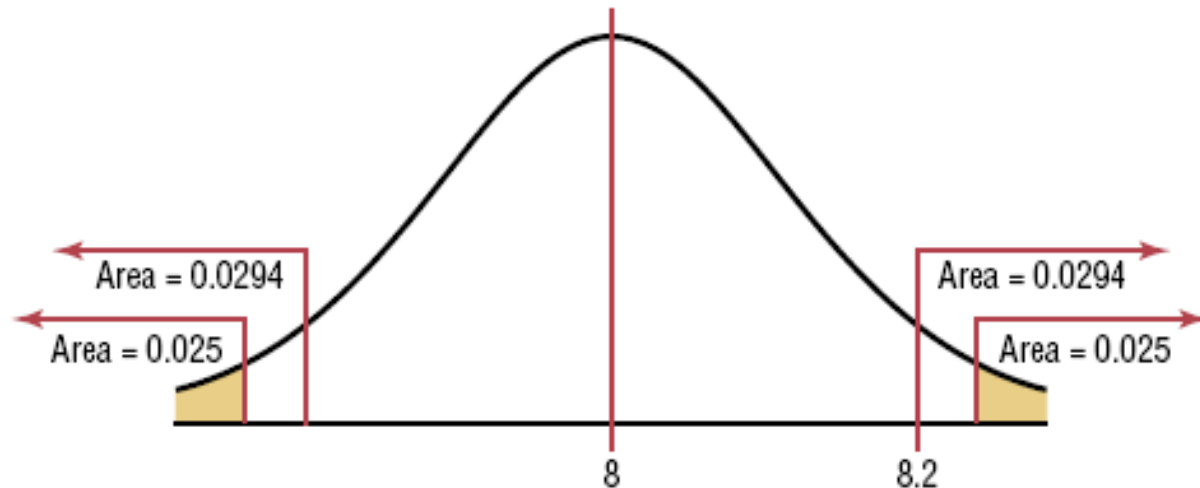
Subtract: $1.0000 - 0.9706 = 0.0294$.

Since this is a two-tailed test, the area of 0.0294 must be doubled to get the P -value.

The P -value is $2(0.0294) = 0.0588$.

Step 4: Make the decision.

The decision is to not reject the null hypothesis, since the P -value is greater than 0.05.



Step 5: Summarize the results.

There is not enough evidence to reject the claim that the average wind speed is 8 miles per hour.

If $P\text{-value} \leq 0.01$, reject the null hypothesis. The difference is highly significant.

If $P\text{-value} > 0.01$ but $P\text{-value} \leq 0.05$, reject the null hypothesis. The difference is significant.

If $P\text{-value} > 0.05$ but $P\text{-value} \leq 0.10$, consider the consequences of type I error before rejecting the null hypothesis.

If $P\text{-value} > 0.10$, do not reject the null hypothesis. The difference is not significant.

The researcher should distinguish between **statistical significance** and **practical significance**.

When the null hypothesis is rejected at a specific significance level, it can be concluded that the difference is probably not due to chance and thus is statistically significant. However, the results may not have any practical significance.

It is up to the researcher to use common sense when interpreting the results of a statistical test.

The *t* test is a statistical test for the mean of a population and is used when the population is normally or approximately normally distributed, σ is unknown.

The formula for the *t* test is

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

The degrees of freedom are d.f. = $n - 1$.

Example: Substitute Salaries



An educator claims that the average salary of substitute teachers in school districts in Allegheny County, Pennsylvania, is less than \$60 per day. A random sample of eight school districts is selected, and the daily salaries (in dollars) are shown. Is there enough evidence to support the educator's claim at $\alpha = 0.10$?

60 56 60 55 70 55 60 55

Step 1: State the hypotheses and identify the claim.

$$H_0: \mu = 60 \text{ and } H_1: \mu < 60 \text{ (claim)}$$

Step 2: Find the critical value.

At $\alpha = 0.10$ and d.f. = 7, the critical value is -1.415 .

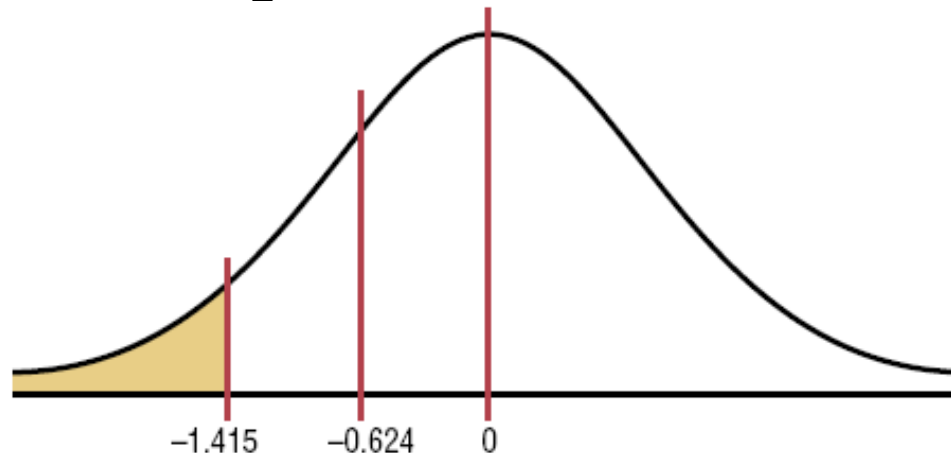
Step 3: Find the test value.

$\bar{X} = 58.88$ and $s = 5.08$.

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{58.88 - 60}{5.08/\sqrt{8}} = -0.624$$

Step 4: Make the decision.

Do not reject the null hypothesis since -0.624 falls in the noncritical region.



Step 5: Summarize the results.

There is not enough evidence to support the claim that the average salary of substitute teachers in Allegheny County is less than \$60 per day.

A physician claims that joggers' maximal volume oxygen uptake is greater than the average of all adults. A sample of 15 joggers has a mean of 40.6 milliliters per kilogram (ml/kg) and a standard deviation of 6 ml/kg. If the average of all adults is 36.7 ml/kg, is there enough evidence to support the physician's claim at $\alpha = 0.05$?

Step 1: State the hypotheses and identify the claim.

$$H_0: \mu = 36.7 \text{ and } H_1: \mu > 36.7 \text{ (claim)}$$

Step 2: Compute the test value.

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{40.6 - 36.7}{6/\sqrt{15}} = 2.517$$

Step 3: Find the P -value.

In the d.f. = 14 row, 2.517 falls between 2.145 and 2.624, corresponding to $\alpha = 0.025$ and $\alpha = 0.01$.

Thus, the P -value is somewhere between 0.01 and 0.025, since this is a one-tailed test.

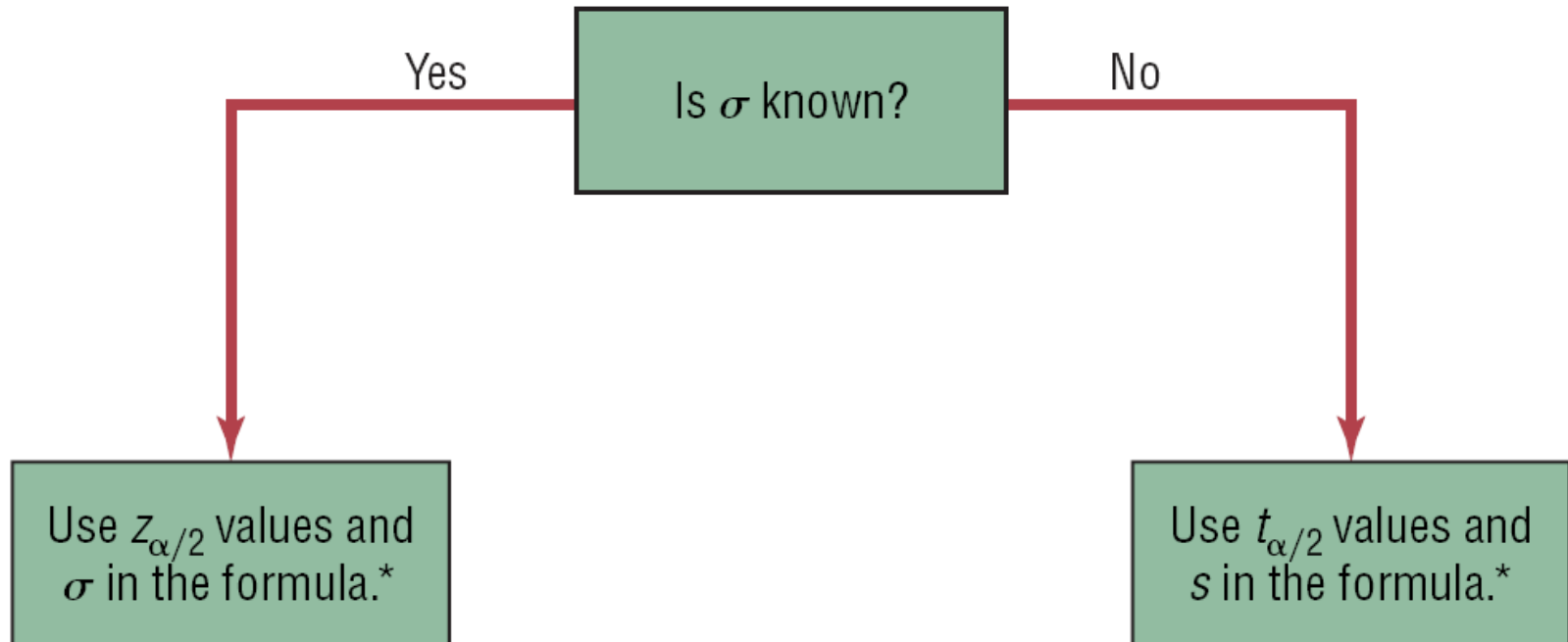
Step 4: Make the decision.

The decision is to reject the null hypothesis, since the P -value < 0.05 .

Step 5: Summarize the results.

There is enough evidence to support the claim that the joggers' maximal volume oxygen uptake is greater than 36.7 ml/kg.

Whether to use z or t



*If $n < 30$, the variable must be normally distributed.

z Test for a Proportion



Since a normal distribution can be used to approximate the binomial distribution when $np \geq 5$ and $nq \geq 5$, the standard normal distribution can be used to test hypotheses for proportions.

The formula for the z test for a proportion is

where

$$z = \frac{\hat{p} - p}{\sqrt{pq/n}}$$

$$\hat{p} = \frac{X}{n} \quad (\text{sample proportion})$$

p = population proportion

n = sample size

A dietitian claims that 60% of people are trying to avoid trans fats in their diets. She randomly selected 200 people and found that 128 people stated that they were trying to avoid trans fats in their diets. At $\alpha = 0.05$, is there enough evidence to reject the dietitian's claim?

Step 1: State the hypotheses and identify the claim.

$$H_0: p = 0.60 \text{ (claim)} \text{ and } H_1: p \neq 0.60$$

Step 2: Find the critical value.

Since $\alpha = 0.05$ and the test is a two-tailed test, the critical value is $z = \pm 1.96$.

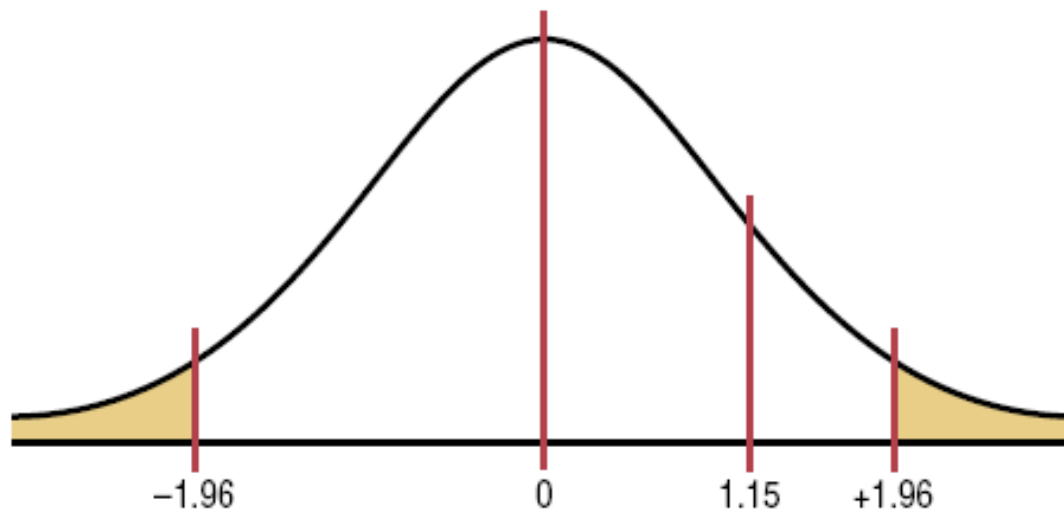
Step 3: Compute the test value.

$$\hat{p} = \frac{X}{n} = \frac{128}{200} = 0.64$$

$$z = \frac{\hat{p} - p}{\sqrt{pq/n}} = \frac{0.64 - 0.60}{\sqrt{(0.60)(0.40)/200}} = 1.15$$

Step 4: Make the decision.

Do not reject the null hypothesis since the test value falls outside the critical region.



Step 5: Summarize the results.

There is not enough evidence to reject the claim that 60% of people are trying to avoid trans fats in their diets.



The Family and Medical Leave Act provides job protection and unpaid time off from work for a serious illness or birth of a child.

In 2000, 60% of the respondents of a survey stated that it was very easy to get time off for these circumstances. A researcher wishes to see if the percentage who said that it was very easy to get time off has changed.

A sample of 100 people who used the leave said that 53% found it easy to use the leave. At $\alpha = 0.01$, has the percentage changed?



Step 1 State the hypotheses and identify the claim.

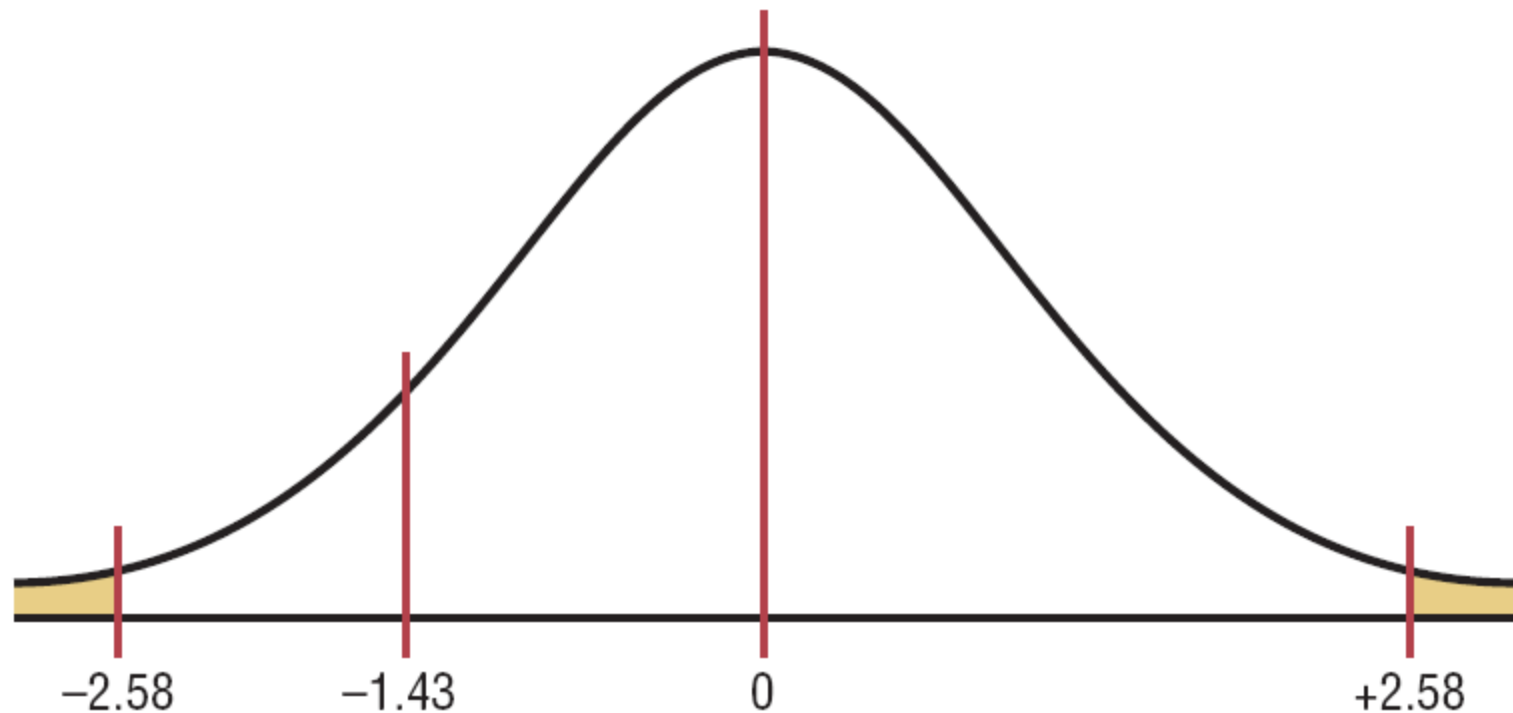
$$H_0: p = 0.60 \quad \text{and} \quad H_1: p \neq 0.60 \text{ (claim)}$$

Step 2 Find the critical value(s). Since $\alpha = 0.01$ and this test is two-tailed, the critical values are ± 2.58 .

Step 3 Compute the test value. It is not necessary to find $\hat{p}$ since it is given in the exercise; Substitute in the formula and evaluate. $\hat{p} = 53\%$.

$$z = \frac{\hat{p} - p}{\sqrt{pq/n}} = \frac{0.53 - 0.60}{\sqrt{(0.6)(0.4)/100}} = -1.43$$

- Step 4** Make the decision. Do not reject the null hypothesis, since the test value falls in the noncritical region.





Step 5 Summarize the results.

There is not enough evidence to support the claim that the percentage of those using the medical leave said that it was easy to get has changed.



The chi-square distribution is also used to test a claim about a single variance or standard deviation.

The formula for the chi-square test for a variance is

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

with degrees of freedom d.f. = $n - 1$ and

n = sample size

s^2 = sample variance

σ^2 = population variance



1. The sample must be randomly selected from the population.
2. The population must be normally distributed for the variable under study.
3. The observations must be independent of one another.



An instructor wishes to see whether the variation in scores of the 23 students in her class is less than the variance of the population. The variance of the class is 198. Is there enough evidence to support the claim that the variation of the students is less than the population variance ($\sigma^2 = 225$) at $\alpha = 0.05$? Assume that the scores are normally distributed.

Step 1: State the hypotheses and identify the claim.

$$H_0: \sigma^2 = 225 \text{ and } H_1: \sigma^2 < 225 \text{ (claim)}$$

Step 2: Find the critical value.

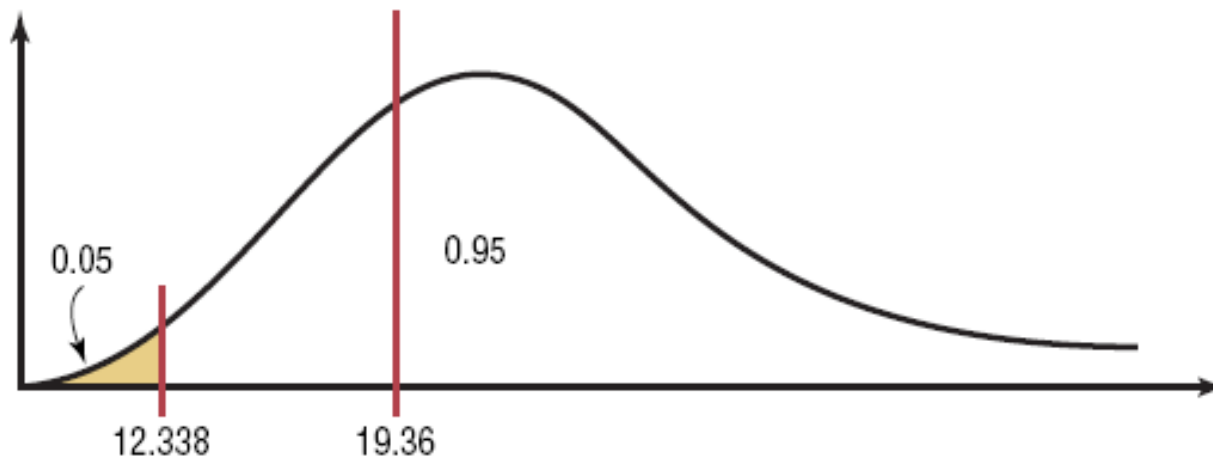
The critical value is $\chi^2 = 12.338$.

Step 3: Compute the test value.

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{(22)(198)}{225} = 19.36$$

Step 4: Make the decision.

Do not reject the null hypothesis since the test value 19.36 falls in the noncritical region.



Step 5: Summarize the results.

There is not enough evidence to support the claim that the variation in test scores of the instructor's students is less than the variation in scores of the population.

A cigarette manufacturer wishes to test the claim that the variance of the nicotine content of its cigarettes is 0.644. Nicotine content is measured in milligrams, and assume that it is normally distributed. A sample of 20 cigarettes has a standard deviation of 1.00 milligram. At $\alpha = 0.05$, is there enough evidence to reject the manufacturer's claim?

Step 1: State the hypotheses and identify the claim.

$$H_0: \sigma^2 = 0.644 \text{ (claim)} \text{ and } H_1: \sigma^2 \neq 0.644$$

Step 2: Find the critical value.

The critical values are 32.852 and 8.907.

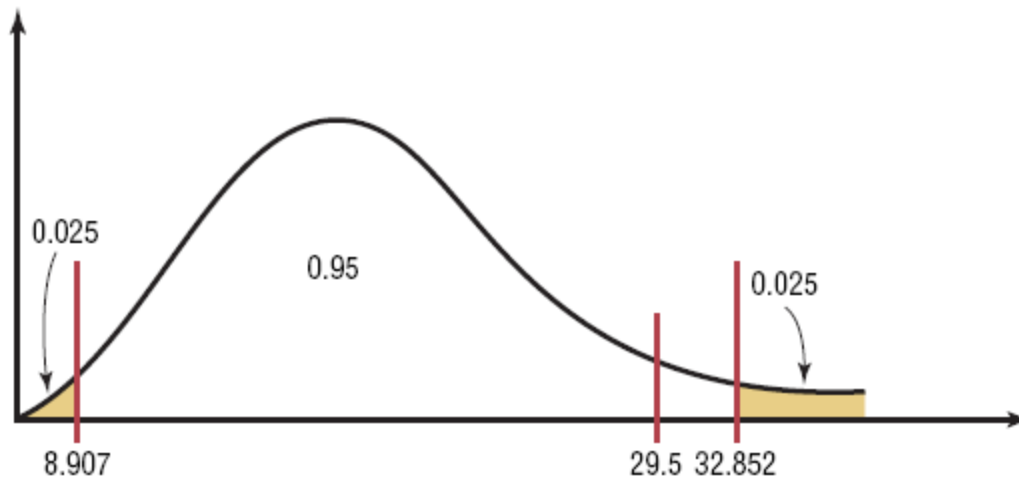
Step 3: Compute the test value.

The standard deviation s must be squared in the formula.

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{(19)(1.00)^2}{0.644} = 29.5$$

Step 4: Make the decision.

Do not reject the null hypothesis, since the test value falls in the noncritical region.



Step 5: Summarize the results.

There is not enough evidence to reject the manufacturer's claim that the variance of the nicotine content of the cigarettes is 0.644.



There is a relationship between confidence intervals and hypothesis testing.

When the null hypothesis is rejected in a hypothesis-testing situation, the confidence interval for the mean using the same level of significance *will not* contain the hypothesized mean.

Likewise, when the null hypothesis is not rejected, the confidence interval computed using the same level of significance *will* contain the hypothesized mean.

Sugar is packed in 5-pound bags. An inspector suspects the bags may not contain 5 pounds. A sample of 50 bags produces a mean of 4.6 pounds and a standard deviation of 0.7 pound. Is there enough evidence to conclude that the bags do not contain 5 pounds as stated at $\alpha = 0.05$? Also, find the 95% confidence interval of the true mean.

Step 1: State the hypotheses and identify the claim.

$$H_0: \mu = 5 \text{ and } H_1: \mu \neq 5 \text{ (claim)}$$

Step 2: Find the critical values.

The critical values are $t = \pm 2.010$.

Step 3: Compute the test value.

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{4.6 - 5.0}{0.7/\sqrt{50}} = -4.04$$

Step 4: Make the decision.

Since $-4.04 < -2.010$, the null hypothesis is rejected.

Step 5: Summarize the results.

There is enough evidence to support the claim that the bags do not weigh 5 pounds.

The 95% confidence interval for the mean is

$$\begin{aligned}\bar{X} - t_{\alpha/2} \frac{s}{\sqrt{n}} &< \mu < \bar{X} + t_{\alpha/2} \frac{s}{\sqrt{n}} \\ 4.6 - (2.010) \left(\frac{0.7}{\sqrt{50}} \right) &< \mu < 4.6 + (2.010) \left(\frac{0.7}{\sqrt{50}} \right) \\ 4.4 &< \mu < 4.8\end{aligned}$$

Notice that the 95% confidence interval of μ does *not* contain the hypothesized value $\mu = 5$.

Hence, there is agreement between the hypothesis test and the confidence interval.

The **power of a test** measures the sensitivity of the test to detect a real difference in parameters if one actually exists. The higher the power, the more sensitive the test. The power is $1 - \beta$.

	H_0 true	H_0 false
Reject H_0	Type I error α	Correct decision $1 - \beta$
Do not reject H_0	Correct decision $1 - \alpha$	Type II error β