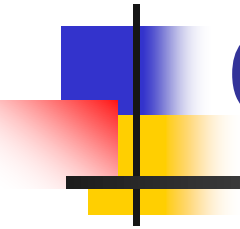


CHAPTER 6:

ESTIMATION AND HYPOTHESIS TESTING: TWO POPULATIONS

INFERENCES ABOUT THE DIFFERENCE BETWEEN TWO POPULATION MEANS FOR LARGE AND INDEPENDENT SAMPLES

- Independent versus Dependent Samples
- Mean, Standard Deviation, and Sampling Distribution of $\bar{x}_1 - \bar{x}_2$
- Interval Estimation of $\mu_1 - \mu_2$
- Hypothesis Testing About $\mu_1 - \mu_2$

Independent versus Dependent Samples

Definition

Two samples drawn from two populations are ***independent*** if the selection of one sample from one population does not affect the selection of the second sample from the second population. Otherwise, the samples are ***dependent***.

Example 10-1

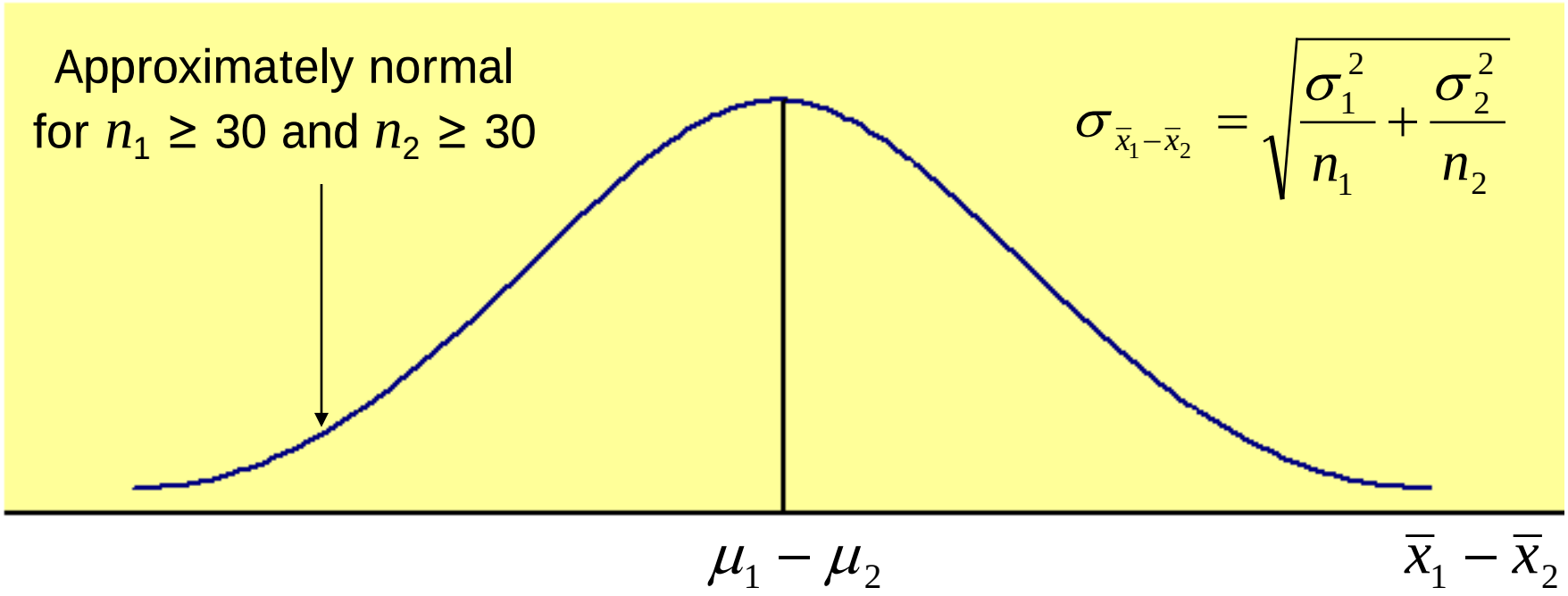
Suppose we want to estimate the difference between the mean salaries of all male and all female executives. To do so, we draw two samples, one from the population of male executives and another from the population of female executives. These two samples are ***independent*** because they are drawn from two different populations, and the samples have no effect on each other.

Example 10-2

Suppose we want to estimate the difference between the mean weights of all participants before and after a weight loss program. To accomplish this, suppose we take a sample of 40 participants and measure their weights before and after the completion of this program. Note that these two samples include the same 40 participants. This is an example of two ***dependent*** samples. Such samples are also called paired or matched samples.

Mean, Standard Deviation, and Sampling Distribution of $\overline{x}_1 - \overline{x}_2$

Figure 10.1



Sampling Distribution, Mean, and Standard Deviation of $\overline{x}_1 - \overline{x}_2$ cont.

For two large and independent samples selected from two different populations, the ***sampling distribution of*** $\overline{x}_1 - \overline{x}_2$ is (approximately) normal with its ***mean*** and ***standard deviation*** as follows:

$$\mu_{\overline{x}_1 - \overline{x}_2} = \mu_1 - \mu_2 \quad \text{and} \quad \sigma_{\overline{x}_1 - \overline{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Sampling Distribution, Mean, and Standard Deviation of $\overline{x}_1 - \overline{x}_2$ cont.

Estimate of the Standard Deviation of $\overline{x}_1 - \overline{x}_2$

The value $s_{\overline{x}_1 - \overline{x}_2}$, which gives an ***estimate*** of $\sigma_{\overline{x}_1 - \overline{x}_2}$, is calculates as

$$s_{\overline{x}_1 - \overline{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

where s_1 and s_2 are the standard deviation of the two samples selected from the two populations.

Interval Estimation of $\mu_1 - \mu_2$

The $(1 - \alpha)100\%$ ***confidence interval for*** $\mu_1 - \mu_2$ is

$$(\bar{x}_1 - \bar{x}_2) \pm z\sigma_{\bar{x}_1 - \bar{x}_2}$$ if σ_1 and σ_2 are known

$$(\bar{x}_1 - \bar{x}_2) \pm zs_{\bar{x}_1 - \bar{x}_2}$$ if σ_1 and σ_2 are not known

The value of z is obtained from the normal distribution table for the given confidence level. The values of $\sigma_{\bar{x}_1 - \bar{x}_2}$ and $s_{\bar{x}_1 - \bar{x}_2}$ are calculated as explained earlier.

Example 10-3

According to the U.S. Bureau of the Census, the average annual salary of full-time state employees was $49,056 in New York and $46,800 in Massachusetts in 2001 (*The Hartford Courant*, December 5, 2002). Suppose that these mean salaries are based on random samples of 500 full-time state employees from New York and 400 full-time employees from Massachusetts and that the population standard deviations of the 2001 salaries of all full-time state employees in these two states were $9000 and $8500, respectively.

Example 10-3

a) What is the point estimate of $\mu_1 - \mu_2$? What is the margin of error?

b) Construct a 97% confidence interval for the difference between the 2001 mean salaries of all full-time state employees in these two states.

Solution 10-3

a)

Point estimate of $\mu_1 - \mu_2 = \bar{x}_1 - \bar{x}_2$

= \$49,056 - \$46,800

= **\$2256**

Solution 10-3

a)

$$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} = \sqrt{\frac{(9000)^2}{500}+\frac{(8500)^2}{400}}$$

$$= \$585.341781$$

$$\text{Margin of error} = \pm 1.96\sigma_{\bar{x}_1-\bar{x}_2}$$

$$= \pm 1.96(585.341781) = \pm \$1147.27$$

Solution 10-3

b)

- 97% confidence level
- Thus, the value of z is 2.17

$$(\bar{x}_1 - \bar{x}_2) \pm z\sigma_{\bar{x}_1 - \bar{x}_2} = (49{,}056 - 46{,}800) \pm 2.17(585.341781)$$

$$= 2256 \pm 1270.19 = \$985.81 \text{ to } \$3526.19$$

- Thus, with 97% confidence we can state that the difference between the 2001 mean salaries of all full-time state employees in New York and Massachusetts is $985.81 to $3526.19.

Example 10-4

According to the National Association of Colleges and Employers, the average salary offered to college students who graduated in 2002 was $43,732 to MIS (Management Information Systems) majors and $40,293 to accounting majors (*Journal of Accountancy,* September 2002). Assume that these means are based on samples of 900 MIS majors and 1200 accounting majors and that the sample standard deviations for the two samples are $2200 and $1950, respectively. Find a 99% confidence interval for the difference between the corresponding population means.

Solution 10-4

$$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{(2200)^2}{900} + \frac{(1950)^2}{1200}} = \$92.447433$$

The value of z is 2.58.

$$(\bar{x}_1 - \bar{x}_2) \pm z s_{\bar{x}_1-\bar{x}_2} = (\$43{,}732 - \$40{,}293) \pm 2.58(92.447433)$$

$$= 3439 \pm 238.51$$

$$= \$3200.49 \quad \text{to} \quad \$3677.51$$

Hypothesis Testing About $\mu_1 - \mu_2$

1. Testing an alternative hypothesis that the means of two populations are different is equivalent to $\mu_1 \neq \mu_2$, which is the same as $\mu_1 - \mu_2 \neq 0$.

HYPOTHESIS TESTING ABOUT $\mu_1 - \mu_2$ cont.

2. Testing an alternative hypothesis that the mean of the first population is greater than the mean of the second population is equivalent to $\mu_1 > \mu_2$, which is the same as $\mu_1 - \mu_2 > 0$.

HYPOTHESIS TESTING ABOUT $\mu_1 - \mu_2$ cont.

3. Testing an alternative hypothesis that the mean of the first population is less than the mean of the second population is equivalent to $\mu_1 < \mu_2$, which is the same as $\mu_1 - \mu_2 < 0$.

HYPOTHESIS TESTING ABOUT $\mu_1 - \mu_2$ cont.

Test Statistic z for $\bar{x}_1 - \bar{x}_2$

The value of the ***test statistic z*** for $\bar{x}_1 - \bar{x}_2$ is computed as

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{x}_1 - \bar{x}_2}}$$

The value of $\mu_1 - \mu_2$ is substituted from H_0. If the values of σ_1 and σ_2 are not known, we replace $\sigma_{\bar{x}_1 - \bar{x}_2}$ with $s_{\bar{x}_1 - \bar{x}_2}$ in the formula.

Example 10-5

Refer to Example 10-3 about the 2001 average salaries of full-time state employees in New York and Massachusetts. Test at the 1% significance level if the 2001 mean salaries of full-time state employees in New York and Massachusetts are different.

Solution 10-5

- H_0: $\mu_1 - \mu_2 = 0$ (the two population means are not different)
- H_1: $\mu_1 - \mu_2 \neq 0$ (the two population means are different)

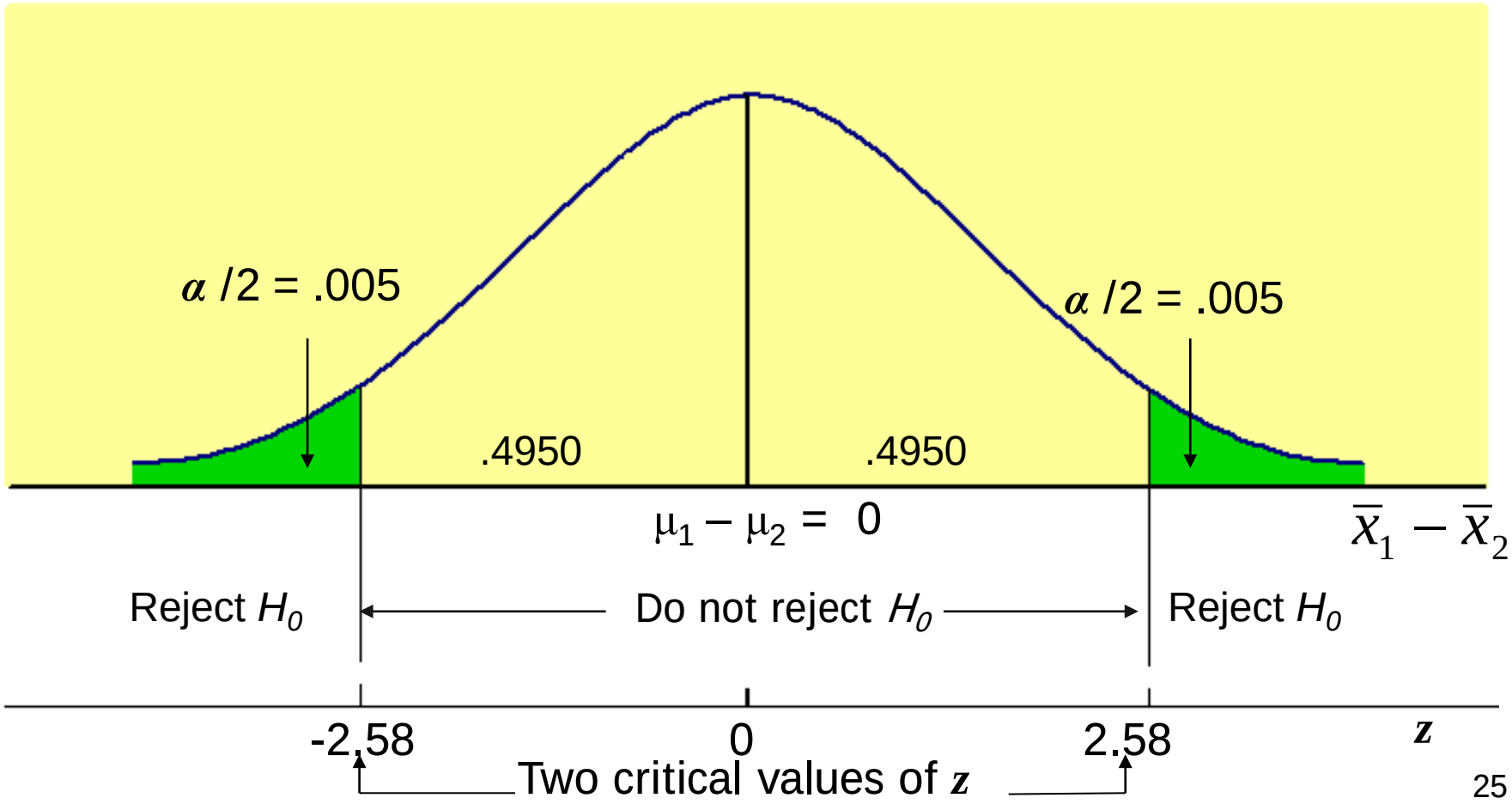
Solution 10-5

- Both samples are large; $n_1 > 30$ and $n_2 > 30$
- Therefore, we use the normal distribution to perform the hypothesis test

Solution 10-5

- α = .01.
- The ≠ sign in the alternative hypothesis indicates that the test is two-tailed
- Area in each tail = α / 2 = .01 / 2 = .005
- The critical values of z are 2.58 and -2.58

Figure 10.2


α /2 = .005
α /2 = .005
.4950
.4950
μ1 – μ2 = 0
$\bar{X}_1 - \bar{X}_2$
Reject H0
Do not reject H0
Reject H0
-2.58
0
2.58
z
Two critical values of z

Solution 10-5

$$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{(9000)^2}{500} + \frac{(8500)^2}{400}} = \$585.341781$$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{x}_1-\bar{x}_2}} = \frac{(49{,}056 - 46{,}800) - 0}{585.341781} = 3.85$$

From H_0

Solution 10-5

- The value of the test statistic z = 3.85
  - It falls in the rejection region.
- We reject the null hypothesis H_0.

Example 10-6

Refer to Example 10-4 about the mean salaries offered to college students who graduated in 2002 with MIS and accounting majors. Test at the 2.5% significance level if the mean salary offered to college students who graduated in 2002 with the MIS major is higher than that for accounting majors.

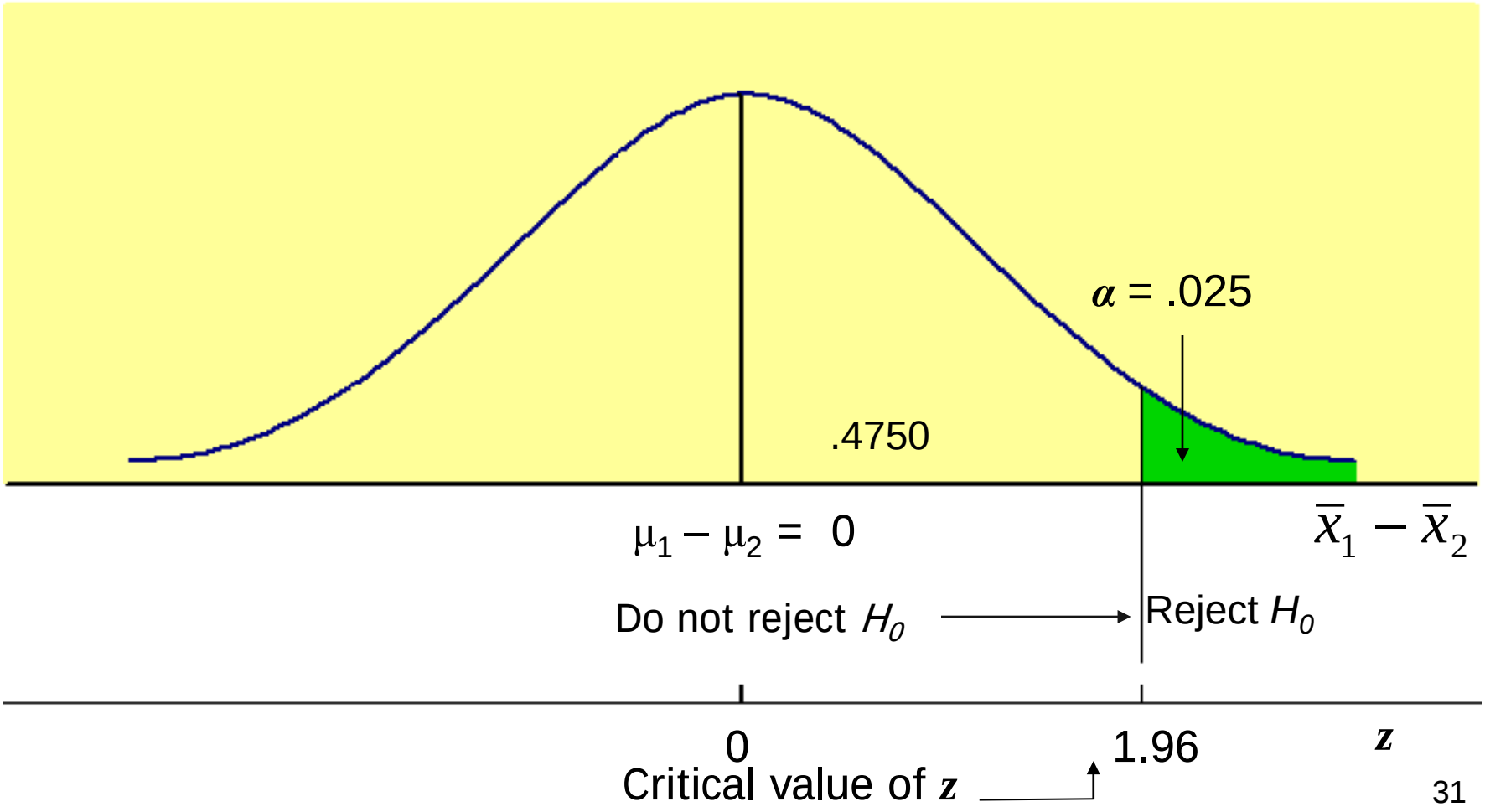
Solution 10-6

- H_0: $\mu_1 - \mu_2 = 0$
  - μ_1 is equal to μ_2
- H_1: $\mu_1 - \mu_2 > 0$
  - μ_1 is higher than μ_2

Solution 10-6

- $\alpha = .025$.
- The > sign in the alternative hypothesis indicates that the test is right-tailed
- The critical value of z is 1.96

Figure 10.3


α = .025
.4750
μ1 – μ2 = 0
X̄1 – X̄2
Do not reject H0
Reject H0
0
1.96
z
Critical value of z

Solution 10-6

$$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{(2200)^2}{900} + \frac{(1950)^2}{1200}} = \$92.447433$$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_{\bar{x}_1-\bar{x}_2}} = \frac{(43{,}732 - 40{,}293) - 0}{92.447433} = 37.20$$

From H_0

Solution 10-6

- The value of the test statistic $z = 37.20$
  - It falls in the rejection region.
- We reject H_0.

INFERENCES ABOUT THE DIFFERENCE BETWEEN TWO POPULATION MEANS FOR SMALL AND INDEPENDENT SAMPLES: EQUAL STANDARD DEVIATIONS

- Interval Estimation of $\mu_1 - \mu_2$
- Hypothesis Testing About $\mu_1 - \mu_2$

When to Use the t Distribution to Make Inferences About $\mu_1 - \mu_2$

The t distribution is used to make inferences about $\mu_1 - \mu_2$ when the following assumptions hold true:

1. The two populations from which the two samples are drawn are (approximately) normally distributed.
2. The samples are small ($n_1 < 30$ and $n_2 < 30$) and independent.
3. The standard deviations σ_1 and σ_2 of the two populations are unknown but they are assumed to be equal; that is, $\sigma_1 = \sigma_2$.

Pooled Standard Deviation for Two Samples

The ***pooled standard deviation for two samples*** is computed as

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

where n_1 and n_2 are the sizes of the two samples and s_1^2 and s_2^2 are the variances of the two samples. Hence s_p is an estimator of σ.

INFERENCES ABOUT THE DIFFERENCE BETWEEN TWO POPULATION MEANS FOR SMALL AND INDEPENDENT SAMPLES: EQUAL STANDARD DEVIATIONS cont.

Estimator of the Standard Deviation of $\bar{x}_1 - \bar{x}_2$

The ***estimator of the standard deviation of*** $\bar{x}_1 - \bar{x}_2$ is

$$s_{\bar{x}_1 - \bar{x}_2} = s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Interval Estimation of $\mu_1 - \mu_2$

Confidence Interval for $\mu_1 - \mu_2$

The $(1 - \alpha)100\%$ ***confidence interval*** for $\mu_1 - \mu_2$ is

$$(\bar{x}_1 - \bar{x}_2) \pm ts_{\bar{x}_1 - \bar{x}_2}$$

where the value of t is obtained from the *t* distribution table for a given confidence level and $n_1 + n_2 - 2$ degrees of freedom, and $s_{\bar{x}_1 - \bar{x}_2}$ is calculated as explained earlier.

Example 10-7

A consuming agency wanted to estimate the difference in the mean amounts of caffeine in two brands of coffee. The agency took a sample of 15 one-pound jars of Brand I coffee that showed the mean amount of caffeine in these jars to be 80 milligrams per jar with a standard deviation of 5 milligrams. Another sample of 12 one-pound jars of Brand II coffee gave a mean amount of caffeine equal to 77 milligram per jar with a standard deviation of 6 milligrams.

Example 10-7

Construct a 95% confidence interval for the difference between the mean amounts of caffeine in one-pound jars of these two brands of coffee. Assume that the two populations are normally distributed and that the standard deviations of the two populations are equal.

Solution 10-7

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{(15-1)(5)^2 + (12-1)(6)^2}{15+12-2}}$$

$$= 5.46260011$$

$$s_{\bar{x}_1 - \bar{x}_2} = s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = (5.46260011)\sqrt{\frac{1}{15} + \frac{1}{12}} = 2.11565593$$

Solution 10-7

$$\text{Area in each tail} = \alpha/2 = .5 - (.95/2) = .025$$

$$\text{Degrees of freedom} = n_1 + n_2 - 2 = 25$$

$$t = 2.060$$

$$(\bar{x}_1 - \bar{x}_2) \pm ts_{\bar{x}_1 - \bar{x}_2} = (80 - 77) \pm 2.060(2.11565593)$$

$$= 3 \pm 4.36$$

$$= -1.36 \text{ to } 7.36 \text{ milligrams}$$

Hypothesis Testing About $\mu_1 - \mu_2$

Test Statistic t for $\bar{x}_1 - \bar{x}_2$

The value of the ***test statistic t for $\bar{x}_1 - \bar{x}_2$*** is computed as

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

The value of $\mu_1 - \mu_2$ in this formula is substituted from the null hypothesis and $s_{\bar{x}_1 - \bar{x}_2}$ is calculated as explained earlier.

Example 10-8

A sample of 14 cans of Brand I diet soda gave the mean number of calories of 23 per can with a standard deviation of 3 calories. Another sample of 16 cans of Brand II diet soda gave the mean number of calories of 25 per can with a standard deviation of 4 calories.

Example 10-8

At the 1% significance level, can you conclude that the mean number of calories per can are different for these two brands of diet soda? Assume that the calories per can of diet soda are normally distributed for each of the two brands and that the standard deviations for the two populations are equal.

Solution 10-8

- H_0: $\mu_1 - \mu_2 = 0$
  - The mean numbers of calories are not different
- H_1: $\mu_1 - \mu_2 \neq 0$
  - The mean numbers of calories are different

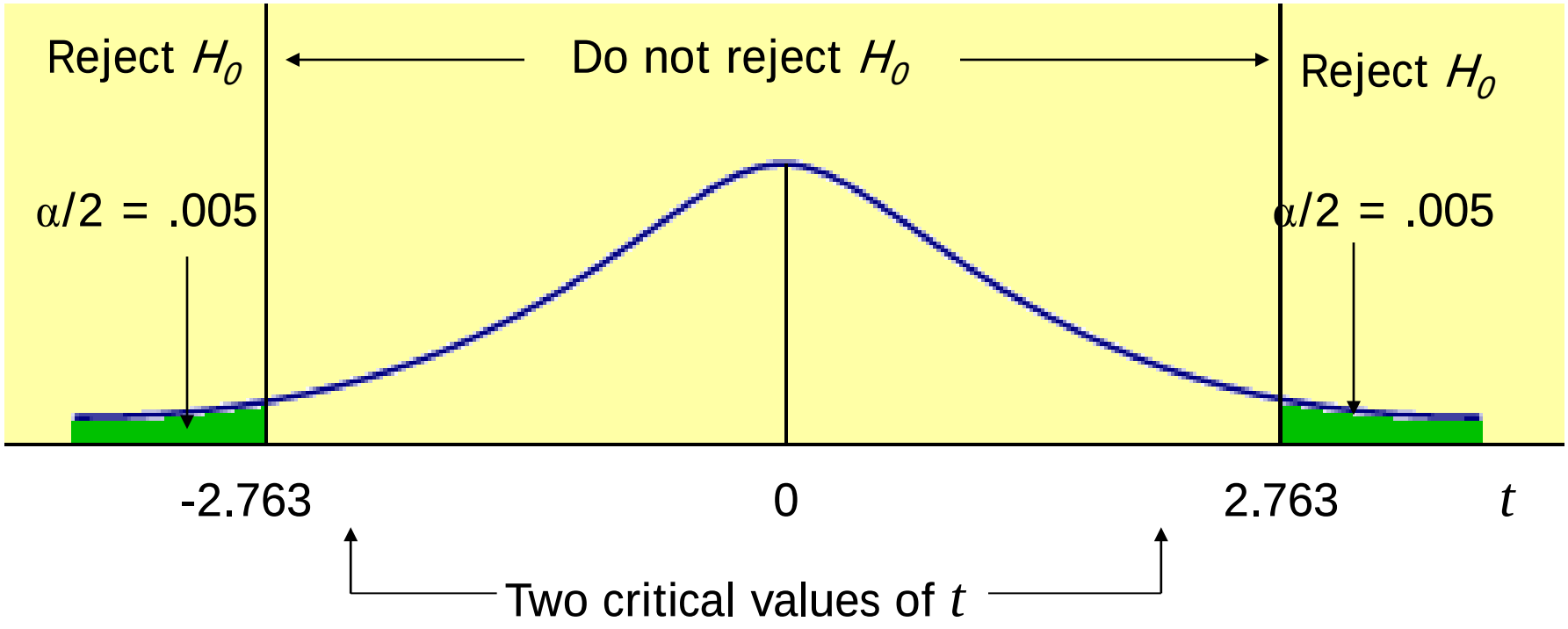
Solution 10-8

- The two populations are normally distributed
- The samples are small and independent
- Standard deviations of the two populations are unknown
- We will use the t distribution

Solution 10-8

- The ≠ sign in the alternative hypothesis indicates that the test is two-tailed.
- $\alpha = .01$.
- Area in each tail = $\alpha / 2 = .01 / 2 = .005$
- $df = n_1 + n_2 - 2 = 14 + 16 - 2 = 28$
- Critical values of t are -2.763 and 2.763.

Figure 10.4


Reject H_0
Do not reject H_0
Reject H_0
$\alpha/2 = .005$
$\alpha/2 = .005$
-2.763
0
2.763
t
Two critical values of t

Solution 10-8

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{(14-1)(3)^2 + (16-1)(4)^2}{16+16-2}} = 3.57071421$$

$$s_{\bar{x}_1 - \bar{x}_2} = s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = (3.57071421)\sqrt{\frac{1}{14} + \frac{1}{16}} = 1.30674760$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_{\bar{x}_1 - \bar{x}_2}} = \frac{(23-25) - 0}{1.30674760} = -1.531$$

From H_0

Solution 10-8

- The value of the test statistic t = -1.531
  - It falls in the nonrejection region
- Therefore, we fail to reject the null hypothesis

Example 10-9

A sample of 15 children from New York State showed that the mean time they spend watching television is 28.50 hours per week with a standard deviation of 4 hours. Another sample of 16 children from California showed that the mean time spent by them watching television is 23.25 hours per week with a standard deviation of 5 hours.

Example 10-9

Using a 2.5% significance level, can you conclude that the mean time spent watching television by children in New York State is greater than that for children in California? Assume that the times spent watching television by children have a normal distribution for both populations and that the standard deviations for the two populations are equal.

Solution 10-9

- H_0: $\mu_1 - \mu_2 = 0$
- H_1: $\mu_1 - \mu_2 > 0$

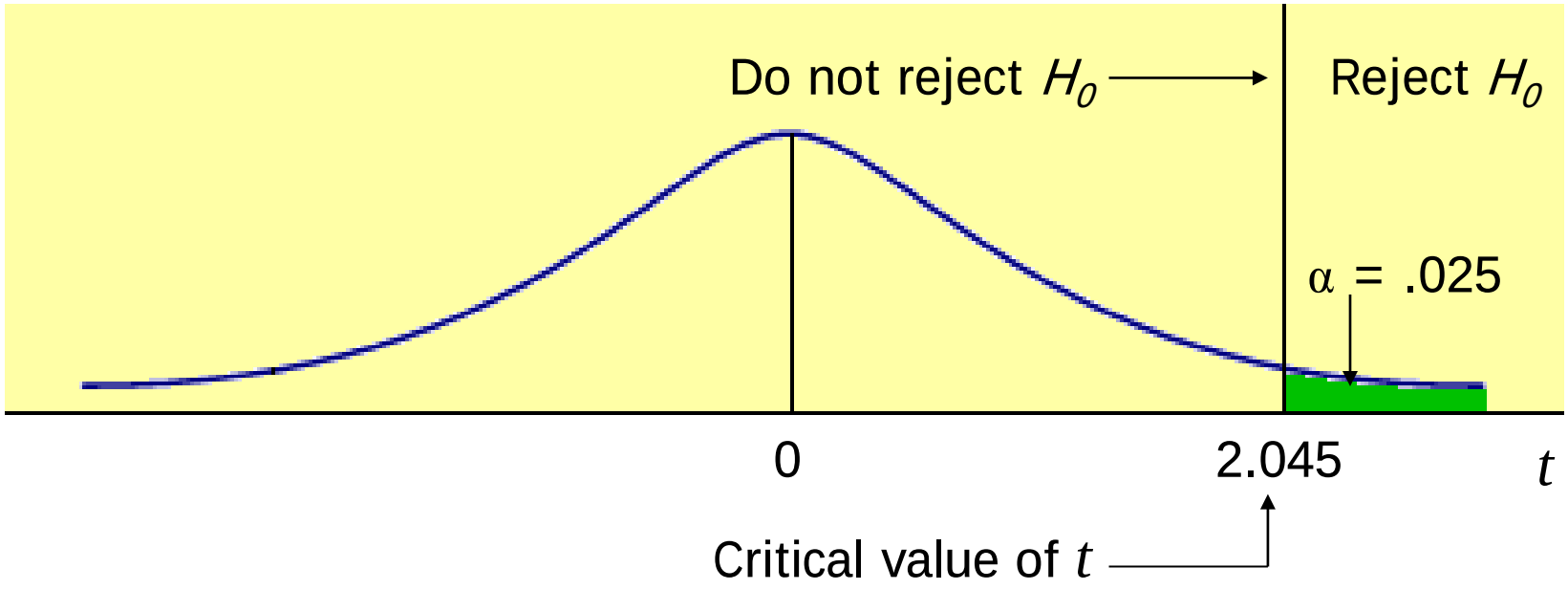
Solution 10-8

- The two populations are normally distributed
- The samples are small and independent
- Standard deviations of the two populations are unknown
- We will use the t distribution

Solution 10-9

- $\alpha = .025$
- Area in the right tail = $\alpha = .025$
- $df = n_1 + n_2 - 2 = 15 + 16 - 2 = 29$
- Critical value of t is 2.045

Figure 10.5



Solution 10-9

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{(15-1)(4)^2 + (16-1)(5)^2}{15+16-2}} = 4.54479619$$

$$s_{\bar{x}_1 - \bar{x}_2} = s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = (4.54479619)\sqrt{\frac{1}{15} + \frac{1}{16}} = 1.63338904$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_{\bar{x}_1 - \bar{x}_2}} = \frac{(28.50 - 23.25) - 0}{1.63338904} = 3.214$$

From H_0

Solution 10-9

- The value of the test statistic $t = 3.214$
  - It falls in the rejection region
- Therefore, we reject the null hypothesis H_0

INFERENCES ABOUT THE DIFFERENCE BETWEEN TWO POPULATION MEANS FOR SMALL AND INDEPENDENT SAMPLES: UNEQUAL STANDARD DEVIATIONS

- Interval Estimation of $\mu_1 - \mu_2$
- Hypothesis Testing About $\mu_1 - \mu_2$

INFERENCES ABOUT THE DIFFERENCE BETWEEN TWO POPULATION MEANS FOR SMALL AND INDEPENDENT SAMPLES: UNEQUAL STANDARD DEVIATIONS cont.

Degrees of Freedom

If

- the two populations from which the samples are drawn are (approximately) normally distributed
- the two samples are small (that is, $n_1 < 30$ and $n_2 < 30$) and independent
- the two population standard deviations are unknown and unequal

Degrees of Freedom cont.

then the t distribution is used to make inferences about $\mu_1 - \mu_2$ and the ***degrees of freedom*** for the t distribution are given by

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$

The number given by this formula is always rounded down for df.

INFERENCES ABOUT THE DIFFERENCE BETWEEN TWO POPULATION MEANS FOR SMALL AND INDEPENDENT SAMPLES: UNEQUAL STANDARD DEVIATIONS cont.

Estimate of the Standard Deviation of $\bar{x}_1 - \bar{x}_2$

The value of $s_{\bar{x}_1 - \bar{x}_2}$, is calculated as

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Confidence Interval for $\mu_1 - \mu_2$

The $(1 - \alpha)100\%$ ***confidence interval*** for $\mu_1 - \mu_2$ is

$$(\bar{x}_1 - \bar{x}_2) \pm t s_{\bar{x}_1 - \bar{x}_2}$$

Example 10-10

According to Example 10-7, a sample of 15 one-pound jars of coffee of Brand I showed that the mean amount of caffeine in these jars is 80 milligrams per jar with a standard deviation of 5 milligrams. Another sample of 12 one-pound coffee jars of Brand II gave a mean amount of caffeine equal to 77 milligrams per jar with a standard deviation of 6 milligrams.

Example 10-10

Construct a 95% confidence interval for the difference between the mean amounts of caffeine in one-pound coffee jars of these two brands. Assume that the two populations are normally distributed and that the standard deviations of the two populations are not equal.

Solution 10-10

$$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} = \sqrt{\frac{(5)^2}{15}+\frac{(6)^2}{12}} = 2.16024690$$

$$\text{Area in each tail} = \alpha/2 = .025$$

$$df = \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1}+\frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}} = \frac{\left(\frac{(5)^2}{15}+\frac{(6)^2}{12}\right)^2}{\frac{\left(\frac{(5)^2}{15}\right)^2}{15-1}+\frac{\left(\frac{(6)^2}{12}\right)^2}{12-1}} = 21.42 \approx 21$$

Solution 10-10

$$t = 2.080$$

$$(\bar{x}_1 - \bar{x}_2) \pm t s_{\bar{x}_1 - \bar{x}_2} = (80 - 77) \pm 2.080(2.16024690)$$

$$= 3 \pm 4.49 = -1.49 \text{ to } 7.49$$

Hypothesis Testing About $\mu_1 - \mu_2$

Test Statistic t for $\bar{x}_1 - \bar{x}_2$

The value of the ***test statistic t for $\bar{x}_1 - \bar{x}_2$*** is computed as

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

The value of $\mu_1 - \mu_2$ in this formula is substituted from the null hypothesis and $s_{\bar{x}_1 - \bar{x}_2}$ is calculated as explained earlier.

Example 10-11

According to example 10-8, a sample of 14 cans of Brand I diet soda gave the mean number of calories per can of 23 with a standard deviation of 3 calories. Another sample of 16 cans of Brand II diet soda gave the mean number of calories as 25 per can with a standard deviation of 4 calories.

Example 10-11

Test at the 1% significance level whether the mean numbers of calories per can of diet soda are different for these two brands. Assume that the calories per can of diet soda are normally distributed for each of these two brands and that the standard deviations for the two populations are not equal.

Solution 10-11

- H_0: $\mu_1 - \mu_2 = 0$
  - The mean numbers of calories are not different
- H_1: $\mu_1 - \mu_2 \neq 0$
  - The mean numbers of calories are different

Solution 10-11

- The two populations are normally distributed
- The samples are small and independent
- Standard deviations of the two populations are unknown
- We will use the t distribution

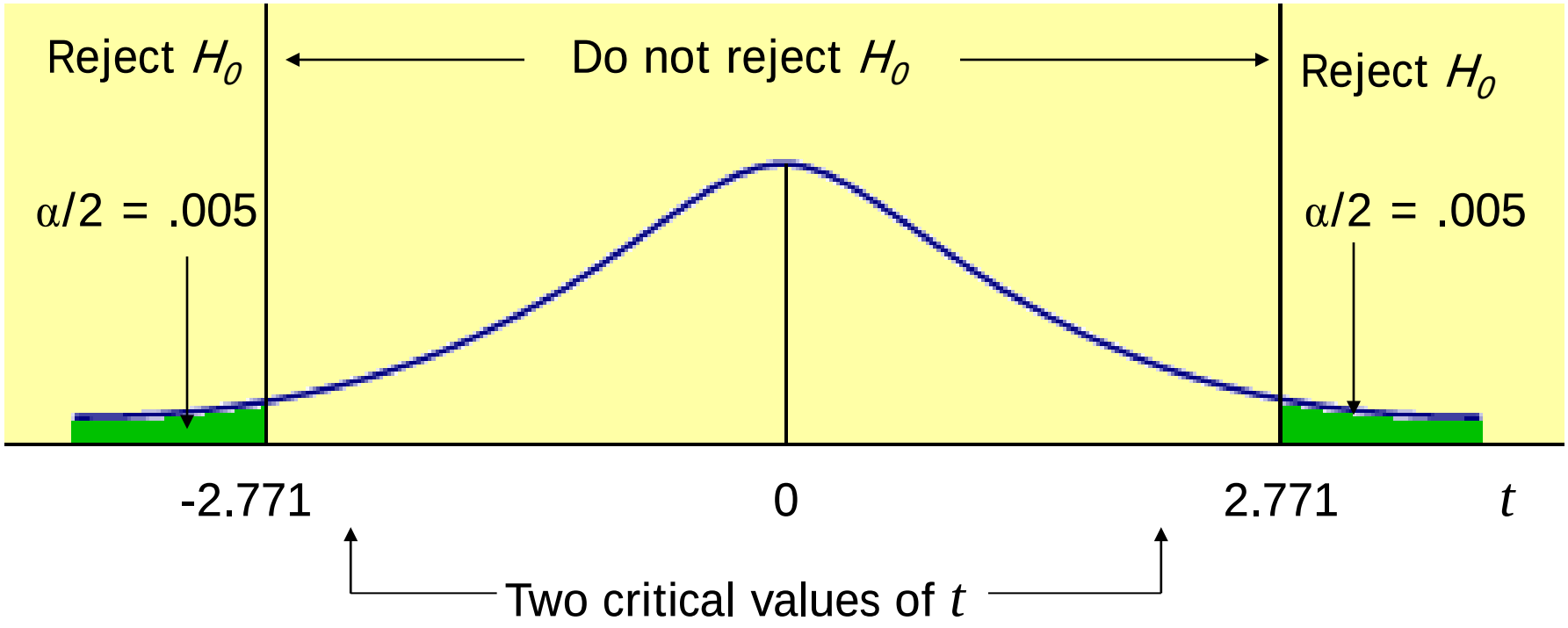
Solution 10-11

- The ≠ in the alternative hypothesis indicates that the test is two-tailed.
- $\alpha = .01$
- Area in each tail $= \alpha / 2 = .01 / 2 = .005$
- The critical values of t are -2.771 and 2.771

Solution 10-11

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}} = \frac{\left(\frac{(3)^2}{14} + \frac{(4)^2}{16}\right)^2}{\frac{\left(\frac{(3)^2}{15}\right)^2}{14-1} + \frac{\left(\frac{(4)^2}{12}\right)^2}{16-1}} = 27.41 \approx 27$$

Figure 10.6


Reject H_0
Do not reject H_0
Reject H_0
$\alpha/2 = .005$
$\alpha/2 = .005$
-2.771
0
2.771
t
Two critical values of t

Solution 10-11

$$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} = \sqrt{\frac{(3)^2}{14}+\frac{(14)^2}{16}} = 1.28173989$$

$$t = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{s_{\bar{x}_1-\bar{x}_2}} = \frac{(23-25)-0}{1.28173989} = -1.560$$

From H_0

Solution 10-11

- The test statistic t = -1.560
  - It falls in the nonrejection region
- Therefore, we fail to reject the null hypothesis

INFERENCES ABOUT THE DIFFERENCE BETWEEN TWO POPULATION MEANS FOR PAIRED SAMPLES

- Interval Estimation of μ_d
- Hypothesis Testing About μ_d

INFERENCES ABOUT THE DIFFERENCE BETWEEN TWO POPULATION MEANS FOR PAIRED SAMPLES cont.

Definition

Two samples are said to be ***paired*** or ***matched samples*** when for each data value collected from one sample there is a corresponding data value collected from the second sample, and both these data values are collected from the same source.

INFERENCES ABOUT THE DIFFERENCE BETWEEN TWO POPULATION MEANS FOR PAIRED SAMPLES cont.

Mean and Standard Deviation of the Paired Differences for Two Samples

The values of the mean and standard deviation, $\bar{d}$ and s_d, of paired differences for two samples are calculated as

$$\bar{d} = \frac{\sum d}{n}$$

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

INFERENCES ABOUT THE DIFFERENCE BETWEEN TWO POPULATION MEANS FOR PAIRED SAMPLES cont.

Sampling Distribution, Mean, and Standard Deviation of $\bar{d}$

If the number of paired values is large ($n \geq 30$), then because of the central limit theorem, the ***sampling distribution of*** $\bar{d}$ is approximately normal with its ***mean*** and ***standard deviation*** given as

$$\mu_{\bar{d}} = \mu_d \quad \text{and} \quad \sigma_{\bar{d}} = \frac{\sigma_d}{\sqrt{n}}$$

INFERENCES ABOUT THE DIFFERENCE BETWEEN TWO POPULATION MEANS FOR PAIRED SAMPLES cont.

Estimate of the Standard Deviation of Paired Differences

If

- n is less than 30
- σ_d is not known
- the population of paired differences is (approximately) normally distributed

Estimate of the Standard Deviation of Paired Differences cont.

then the t distribution is used to make inferences about μ_d. The standard deviation of $\sigma_{\bar{d}}$ of $\bar{d}$ is estimated by $s_{\bar{d}}$, which is calculated as

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

Interval Estimation of μ_d

Confidence Interval for μ_d

The $(1 - \alpha)100\%$ ***confidence interval*** for μ_d is

$$\bar{d} \pm ts_{\bar{d}}$$

where the value of t is obtained from the t distribution table for a given confidence level and $n - 1$ degrees of freedom, and $S_{\bar{d}}$ is calculated as explained earlier.

Example 10-12

A researcher wanted to find the effect of a special diet on systolic blood pressure. She selected a sample of seven adults and put them on this dietary plan for three months. The following table gives the systolic blood pressures of these seven adults before and after the completion of this plan.

Example 10-12

Before	210	180	195	220	231	199	224
After	193	186	186	223	220	183	223

Example 10-12

Let μ_d be the mean reduction in the systolic blood pressure due to this special dietary plan for the population of all adults. Construct a 95% confidence interval for μ_d. Assume that the population of paired differences is (approximately) normally distributed.

Solution 10-12

Table 10.1

Before	After	Difference d	d^2
210	193	17	289
180	186	-6	36
195	186	9	81
220	223	-3	9
231	220	11	121
199	183	16	256
224	233	-9	81
		$\Sigma d = 35$	$\Sigma d^2 = 873$

Solution 10-12

$$\bar{d} = \frac{\sum d}{n} = \frac{35}{7} = 5.00$$

$$s_d = \sqrt{\frac{\sum d^2 - \frac{(\sum d)^2}{n}}{n-1}} = \sqrt{\frac{873 - \frac{(35)^2}{7}}{7-1}} = 10.78579312$$

Solution 10-12

Hence,

$$s_{\bar{d}} = \frac{s_d}{\sqrt{n}} = \frac{10.78579312}{\sqrt{7}} = 4.07664661$$

$$\text{Area in each tail} = \alpha/2 = .5 - (.95/2) = .025$$

$$df = n - 1 = 7 - 1 = 6$$

$$t = 2.447$$

$$\bar{d} \pm ts_{\bar{d}} = 5.00 \pm 2.447(4.07664661) = 5.00 \pm 9.98$$

$$= -4.98 \text{ to } 14.98$$

Hypothesis Testing About μ_d

Test Statistic t for $\bar{d}$

The value of the ***test statistic t for*** $\bar{d}$ is computed as follows:

$$t = \frac{\bar{d} - \mu_d}{s_{\bar{d}}}$$

Example 10-13

A company wanted to know if attending a course on “how to be a successful salesperson” can increase the average sales of its employees. The company sent six of its salesperson to attend this course. The following table gives the one-week sales of these salespersons before and after they attended this course.

Example 10-13

Before	12	18	25	9	14	16
After	18	24	24	14	19	20

Using the 1% significance level, can you conclude that the mean weekly sales for all salespersons increase as a result of attending this course? Assume that the population of paired differences has a normal distribution.

Solution 10-13

Table 10.2

Before	After	Difference d	d^2
12	18	-6	36
18	24	-6	36
25	24	1	1
9	14	-5	25
14	16	-5	25
16	20	-4	16
		Σd = -25	Σd^2 = 139

Solution 10-13

$$\bar{d} = \frac{\sum d}{n} = \frac{-25}{6} = -4.17$$

$$s_d = \sqrt{\frac{\sum d^2 - \frac{(\sum d)^2}{n}}{n-1}} = \sqrt{\frac{139 - \frac{(-25)^2}{6}}{6-1}} = 2.63944439$$

$$s_{\bar{d}} = \frac{s_d}{\sqrt{n}} = \frac{2.63944439}{\sqrt{6}} = 1.07754866$$

Solution 10-13

- H_0: $\mu_d = 0$
  - $\mu_1 - \mu_2 = 0$ or the mean weekly sales do not increase
- H_1: $\mu_d < 0$
  - $\mu_1 - \mu_2 < 0$ or the mean weekly sales do increase

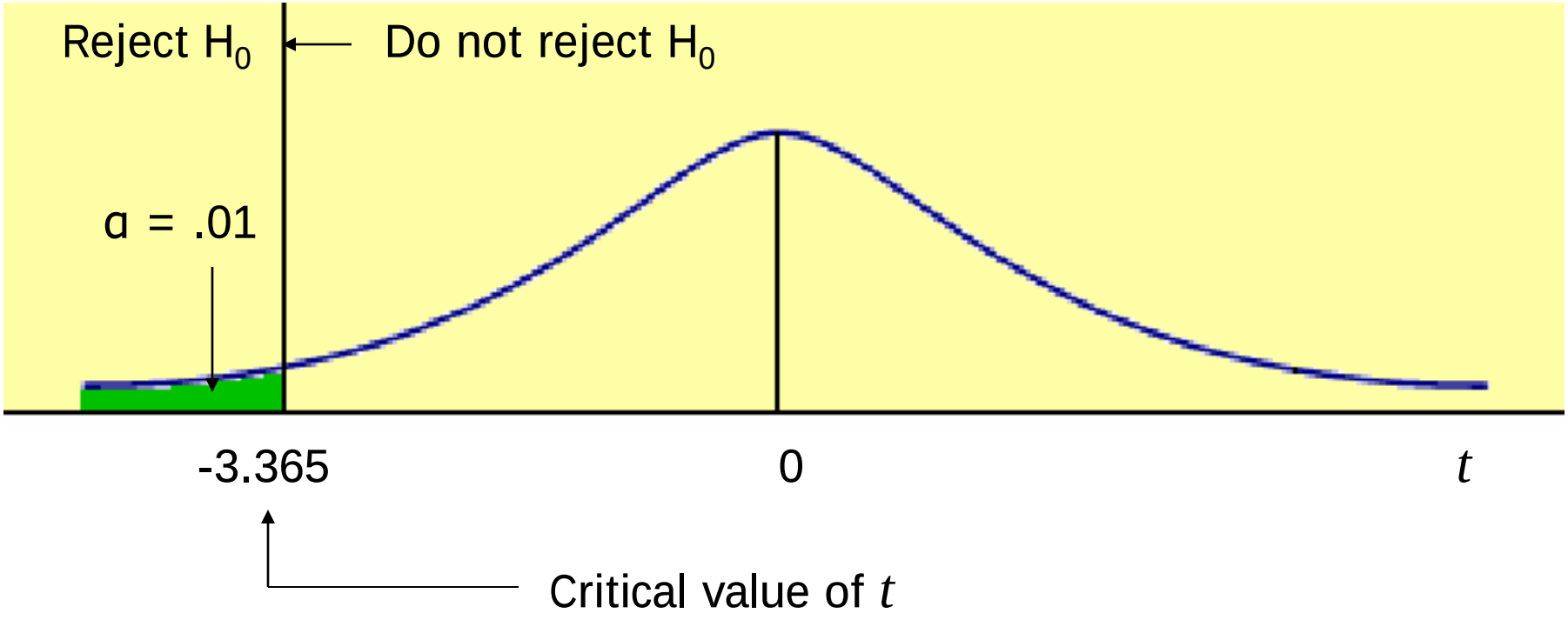
Solution 10-13

- The sample size is small (n < 30)
- The population of paired differences is normal
- σ_d is unknown
- Therefore, we use the t distribution to conduct the test

Solution 10-13

- α = .01.
- Area in left tail = α = .01
- df = n – 1 = 6 – 1 = 5
- The critical value of t is -3.365

Figure 10.7



Solution 10-13

From H_0

$$t = \frac{\bar{d} - \mu_d}{s_{\bar{d}}} = \frac{-4.17 - 0}{1.07754866} = -3.870$$

Solution 10-13

- The value of the test statistic t = -3.870
  - It falls in the rejection region
- Therefore, we reject the null hypothesis

Example 10-14

Refer to Example 10-12. The table that gives the blood pressures of seven adults before and after the completion of a special dietary plan is reproduced here.

Before	210	180	195	220	231	199	224
After	193	186	186	223	220	183	233

Example 10-14

Let μ_d be the mean of the differences between the systolic blood pressures before and after completing this special dietary plan for the population of all adults. Using the 5% significance level, can we conclude that the mean of the paired differences μ_d is difference from zero? Assume that the population of paired differences is (approximately) normally distributed.

Table 10.3

Solution 10-14

Before	After	Difference d	d^2
210	193	17	289
180	186	-6	36
195	186	9	81
220	223	-3	9
231	220	11	121
199	183	16	256
224	233	-9	81
		$\Sigma d = 35$	$\Sigma d^2 = 873$

Solution 10-13

$$\bar{d} = \frac{\sum d}{n} = \frac{35}{7} = 5.00$$

$$s_d = \sqrt{\frac{\sum d^2 - \frac{(\sum d)^2}{n}}{n-1}} = \sqrt{\frac{873 - \frac{(35)^2}{7}}{7-1}} = 10.78579312$$

$$s_{\bar{d}} = \frac{s_d}{\sqrt{n}} = \frac{10.78579312}{\sqrt{7}} = 4.07664661$$

Solution 10-14

- H_0: $\mu_d = 0$
  - The mean of the paired differences is not different from zero
- H_1: $\mu_d \neq 0$
  - The mean of the paired differences is different from zero

Solution 10-14

- Sample size is small
- The population of paired differences is (approximately) normal
- σ_d is not known
- Therefore, we use the t distribution

Solution 10-14

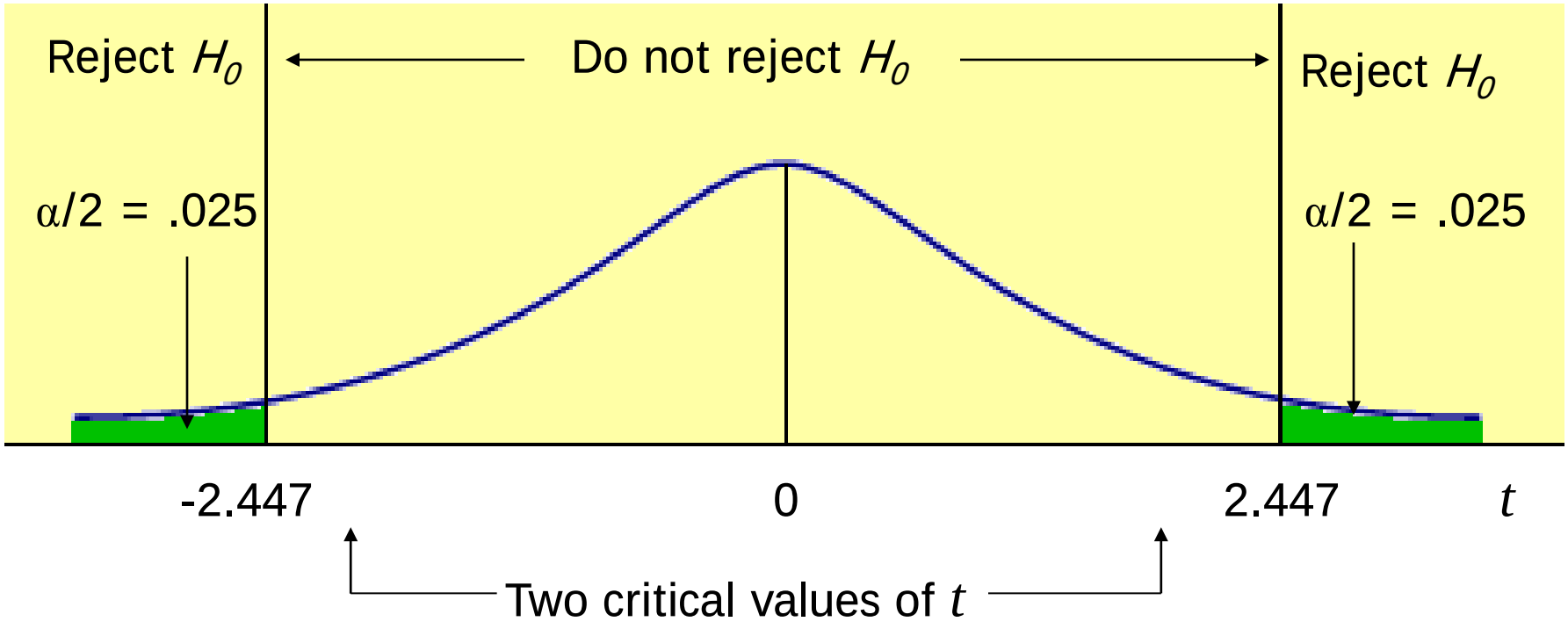
- $\alpha = .05$
- Area in each tail $= \alpha / 2 = .05 / 2 = .025$
- df $= n - 1 = 7 - 1 = 6$
- The two critical values of t are -2.447 and 2.447

Solution 10-14

From H_0

$$t = \frac{\bar{d} - \mu_d}{s_{\bar{d}}} = \frac{5.00 - 0}{4.07664661} = 1.226$$

Figure 10.8


Reject H_0
Do not reject H_0
Reject H_0
$\alpha/2 = .025$
$\alpha/2 = .025$
-2.447
0
2.447
t
Two critical values of t

Solution 10-14

- The value of the test statistic $t = 1.226$
- It falls in the nonrejection region
- Therefore, we fail to reject the null hypothesis

INFERENCES ABOUT THE DIFFERENCE BETWEEN TWO POPULATION PROPORTIONS FOR LARGE AND INDEPENDENT SAMPLES

- Mean, Standard Deviation, and Sampling Distribution of $\hat{p}_1 - \hat{p}_2$
- Interval Estimation of $p_1 - p_2$
- Hypothesis Testing About $p_1 - p_2$

Mean, Standard Deviation, and Sampling Distribution of $\hat{p}_1 - \hat{p}_2$

For two large and independent samples, the ***sampling distribution*** of $\hat{p}_1 - \hat{p}_2$ is (approximately) normal with its ***mean*** and ***standard deviation*** given as

$$\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$$

and

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

respectively, where $q_1 = 1 - p_1$ and $q_2 = 1 - p_2$.

Interval Estimation of $p_1 - p_2$

The $(1 - \alpha)100\%$ ***confidence interval*** for $p_1 - p_2$ is

$$(\hat{p}_1 - \hat{p}_2) \pm z s_{\hat{p}_1 - \hat{p}_2}$$

where the value of z is read from the normal distribution table for the given confidence level, and $s_{\hat{p}_1 - \hat{p}_2}$ is calculates as

$$s_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

Example 10-15

A researcher wanted to estimate the difference between the percentages of users of two toothpastes who will never switch to another toothpaste. In a sample of 500 users of Toothpaste A taken by this researcher, 100 said that they will never switch to another toothpaste. In another sample of 400 users of Toothpaste B taken by the same researcher, 68 said that they will never switch to another toothpaste.

Example 10-15

a) Let p_1 and p_2 be the proportions of all users of Toothpastes A and B, respectively, who will never switch to another toothpaste. What is the point estimate of $p_1 - p_2$?

b) Construct a 97% confidence interval for the difference between the proportions of all users of the two toothpastes who will never switch.

Solution 10-15

$$\hat{p}_1 = x_1 / n_1 = 100/500 = .20$$

$$\hat{p}_2 = x_2 / n_2 = 68/400 = .17$$

Then,

$$\hat{q}_1 = 1 - .20 = .80 \quad \text{and} \quad \hat{q}_2 = 1 - .17 = .83$$

Solution 10-15

a) Point estimate of $p_1 - p_2$

$= \hat{p}_1 - \hat{p}_2$

$= .20 - .17 = .03$

Solution 10-15

b)

$$n_1 \hat{p}_1 = 500(.20) = 100$$

$$n_1 \hat{q}_1 = 500(.80) = 400$$

$$n_2 \hat{p}_2 = 400(.17) = 68$$

$$n_2 \hat{q}_2 = 400(.83) = 332$$

Solution 10-15

b)

- Each of these values is greater than 5
- Both samples are large
- Consequently, we use the normal distribution to make a confidence interval for $p_1 - p_2$.

Solution 10-15

b) The standard deviation of $\hat{p}_1 - \hat{p}_2$ is

$$s_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} = .02593742$$

$z = 2.17$

$$(\hat{p}_1 - \hat{p}_2) \pm zs_{\hat{p}_1-\hat{p}_2} = (.20 - .17) \pm 2.17(.02593742)$$

$$= .03 \pm .056 = -.026 \text{ to } .086$$

Hypothesis Testing About $p_1 - p_2$

Test Statistic z for $\hat{p}_1 - \hat{p}_2$

The value of the ***test statistic z for*** $\hat{p}_1 - \hat{p}_2$ is calculated as

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

The value of $p_1 - p_2$ is substituted from H_0, which is usually zero.

Hypothesis Testing About $p_1 - p_2$ cont.

$$\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

$$\overline{p} = \frac{x_1 + x_2}{n_1 + n_2} \quad \text{or} \quad \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2}$$

$$s_{\hat{p}_1-\hat{p}_2} = \sqrt{\overline{p}\,\overline{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

where $\overline{q} = 1 - \overline{p}$

Example 10-16

Reconsider Example 10-15 about the percentages of users of two toothpastes who will never switch to another toothpaste. At the 1% significance level, can we conclude that the proportion of users of Toothpaste A who will never switch to another toothpaste is higher than the proportion of users of Toothpaste B who will never switch to another toothpaste?

Solution 10-16

$$\hat{p}_1 = x_1 / n_1 = 100/500 = .20$$

$$\hat{p}_2 = x_2 / n_2 = 68/400 = .17$$

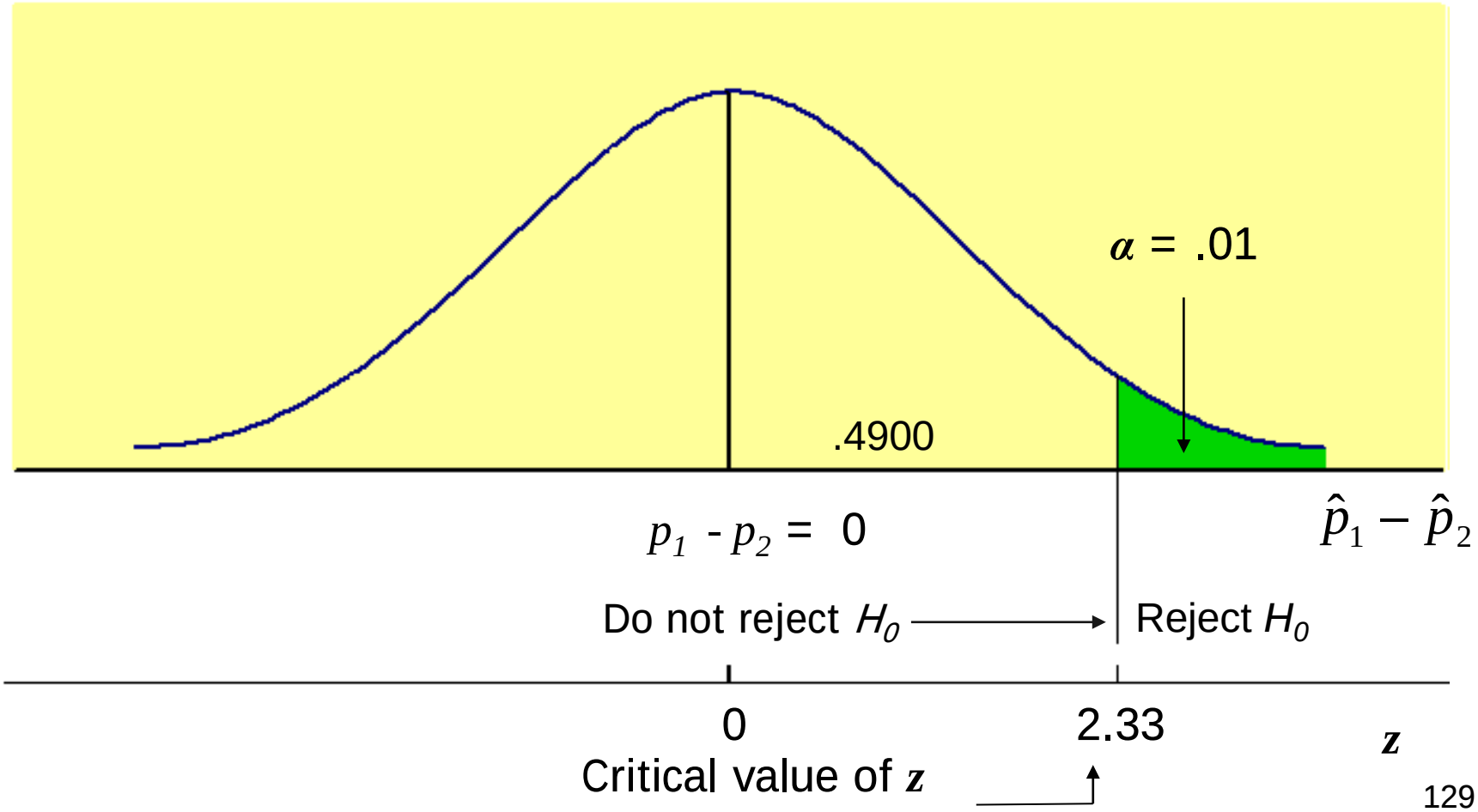
Solution 10-16

- H_0: $p_1 - p_2 = 0$
  - p_1 is not greater than p_2
- H_1: $p_1 - p_2 > 0$
  - p_1 is greater than p_2

Solution 10-16

- $n_1\hat{p}_1$, $n_1\hat{q}_1$, $n_2\hat{p}_2$, and $n_2\hat{q}_2$ are all greater than 5
- Samples sizes are large
- Therefore, we apply the normal distribution
- $\alpha = .01$
- The critical value of z is 2.33

Figure 10.9


α = .01
.4900
$p_1 - p_2 = 0$
$\hat{p}_1 - \hat{p}_2$
Do not reject H_0
Reject H_0
0
2.33
z
Critical value of z

Solution 10-16

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{100 + 68}{500 + 400} = .187$$

$$\bar{q} = 1 - \bar{p} = 1 - .87 = .813$$

$$s_{\hat{p}_1 - \hat{p}_2} = \sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \sqrt{(.187)(.813)\left(\frac{1}{500} + \frac{1}{400}\right)} = .02615606$$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{s_{\hat{p}_1 - \hat{p}_2}} = \frac{(.20 - .17) - 0}{.02615606} = 1.15$$

From H_0

Solution 10-16

- The value of the test statistic $z = 1.15$
  - It falls in the nonrejection region
- Therefore, we reject the null hypothesis

Example 10-17

According to a study by Hewitt Associates, the percentage of companies that hosted holiday parties was 67% in 2001 and 64% in 2002 (*USA TODAY*, December 2, 2002). Suppose that this study is based on samples of 1000 and 1200 companies for 2001 and 2002, respectively. Test whether the percentages of companies that hosted holiday parties in 2001 and 2002 are different. Use the 1% significance level.

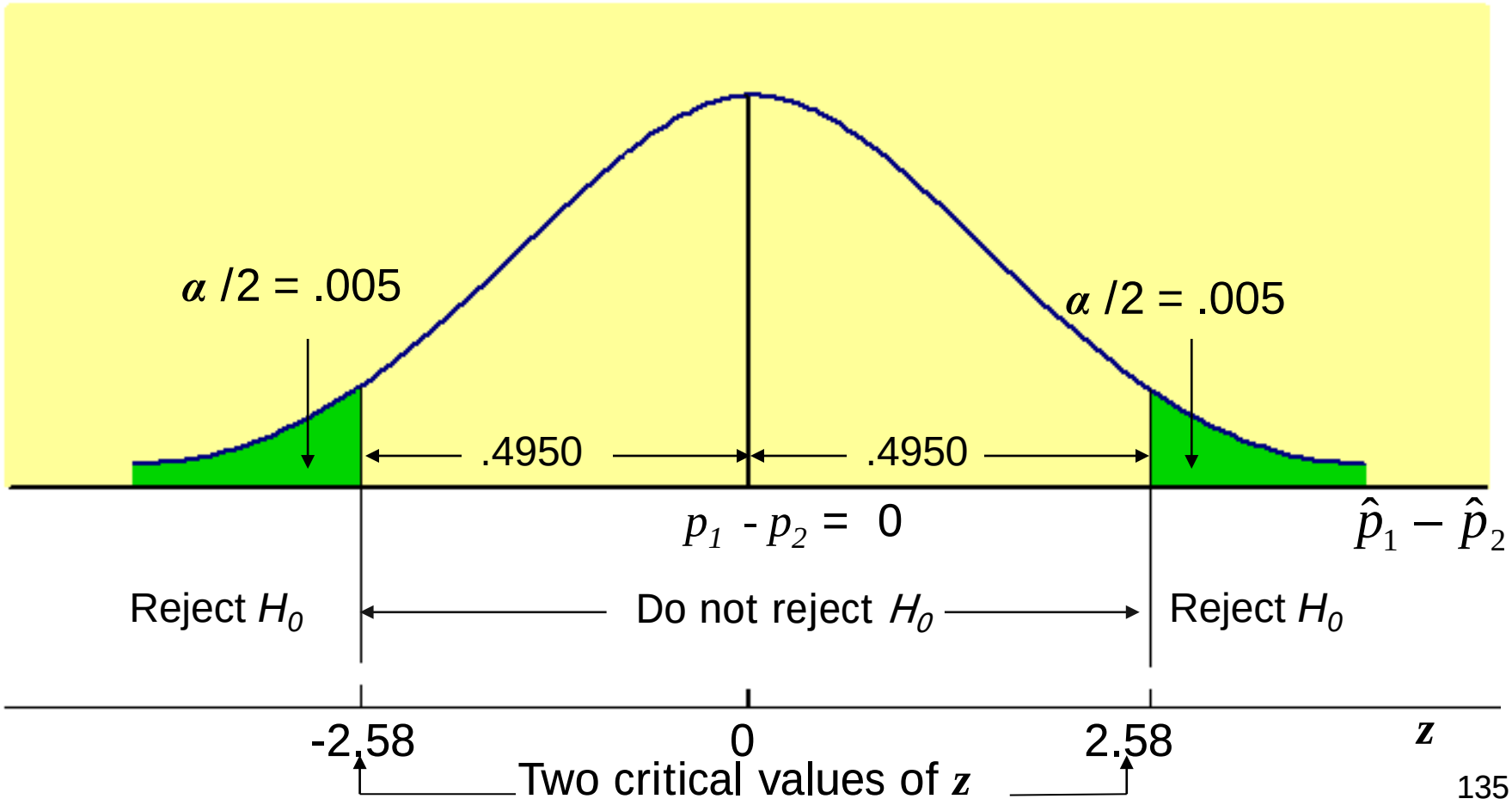
Solution 10-17

- H_0: $p_1 - p_2 = 0$
  - The two population proportions are not different
- H_1: $p_1 - p_2 \neq 0$
  - The two population proportions are different

Solution 10-17

- Samples are large and independent
- Therefore, we apply the normal distribution
- $n_1\hat{p}_1$,$n_1\hat{q}_1$,$n_2\hat{p}_2$ and $n_2\hat{q}_2$ are all greater than 5
- $\alpha = .01$
- The critical values of z are -2.58 and 2.58

Figure 10.10


α /2 = .005
α /2 = .005
.4950
.4950
$p_1 - p_2 = 0$
$\hat{p}_1 - \hat{p}_2$
Reject H_0
Do not reject H_0
Reject H_0
-2.58
0
2.58
z
Two critical values of z

Solution 10-17

$$\bar{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{1000(.67) + 1200(.64)}{1000 + 1200} = .654$$

$$\bar{q} = 1 - \bar{p} = 1 - .654 = .346$$

$$s_{\hat{p}_1 - \hat{p}_2} = \sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \sqrt{(.654)(.346)\left(\frac{1}{1000} + \frac{1}{1200}\right)} = .02036797$$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{s_{\hat{p}_1 - \hat{p}_2}} = \frac{(.67 - .64) - 0}{.02036797} = 1.47$$

From H_0

Solution 10-17

- The value of the test statistic $z = 1.47$
- It falls in the nonrejection region
- Therefore, we fail to reject the null hypothesis H_0