CHAPTER 8

ESTIMATION OF THE MEAN AND PROPORTION

Opening Example

Imagine you are sleeping at night and need to get up early in the morning to go to work. Or, you are reading something with intense concentration, or watching your most favorite show on TV. Suddenly a loud noise disturbs you. You get angry. You feel like you have lost something. What is that frustrating noise that disturbs you? In a survey, 28% of adults said that the most frustrating sound is the jackhammer. (See Case Study 8-2).

8.1 Estimation: An Introduction

Definition

The assignment of value(s) to a population parameter based on a value of the corresponding sample statistic is called ***estimation***.

ESTIMATION: AN INTRODUCTION

Definition

The value(s) assigned to a population parameter based on the value of a sample statistic is called an ***estimate***.

The sample statistic used to estimate a population parameter is called an ***estimator***.

ESTIMATION: AN INTRODUCTION

The estimation procedure involves the following steps.

- Select a sample.
- Collect the required information from the members of the sample.
- Calculate the value of the sample statistic.
- Assign value(s) to the corresponding population parameter.

8.2 Point and Interval Estimates

- A Point Estimate
- An Interval Estimate

A Point Estimate

Definition

The value of a sample statistic that is used to estimate a population parameter is called a ***point estimate***.

A Point Estimate

- Usually, whenever we use point estimation, we calculate the margin of error associated with that point estimation.
- The margin of error is calculated as follows:

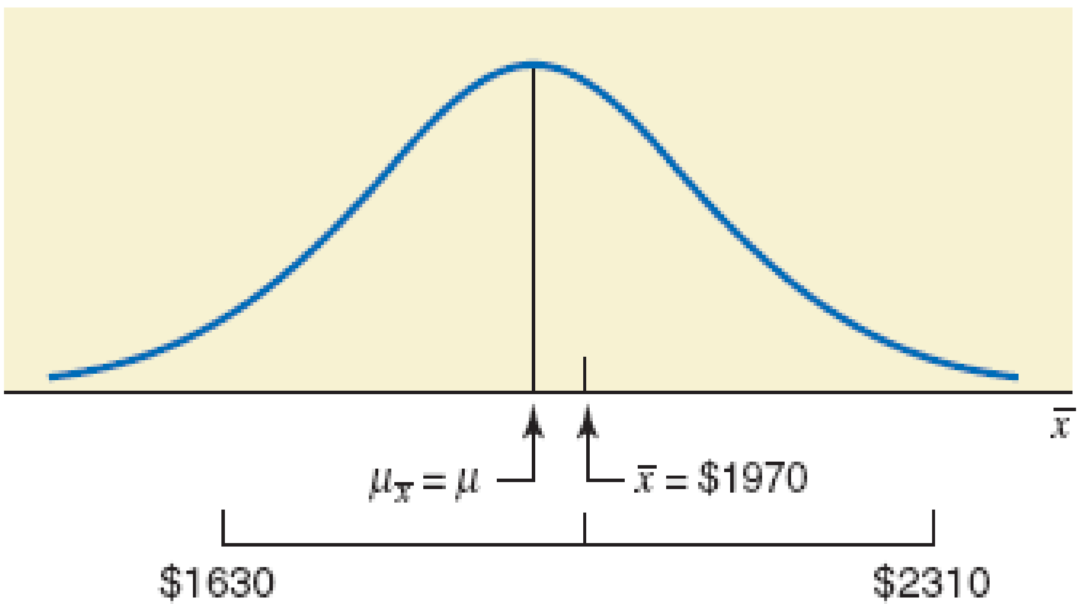
$$\textbf{Margin of error} = \pm 1.96\sigma_{\bar{x}} \quad \textbf{or} \quad \pm 1.96 s_{\bar{x}}$$

An Interval Estimation

Definition

In ***interval estimation***, an interval is constructed around the point estimate, and it is stated that this interval is likely to contain the corresponding population parameter.

Figure 8.1 Interval estimation.



Prem Mann, *Introductory Statistics*, *7/E*

Confidence Level and Confidence Interval

Definition

Each interval is constructed with regard to a given ***confidence level*** and is called a ***confidence interval***. The confidence level is given as

Point estimate $\pm$ Margin of error

The confidence less associated with a confidence interval states how much confidence we have that this interval contains the true population parameter. The confidence level is denoted by $(1 - \alpha)100\%$.

8.3 Estimation of a Population Mean: σ Known

Three Possible Cases

Case I. If the following three conditions are fulfilled:

1. The population standard deviation σ is known
2. The sample size is small (i.e., $n < 30$)
3. The population from which the sample is selected is normally distributed.

ESTIMATION OF A POPULATION MEAN: ? KNOWN

Three Possible Cases

Case II. If the following two conditions are fulfilled:

1. The population standard deviation σ is known
2. The sample size is large (i.e., $n \geq 30$)

ESTIMATION OF A POPULATION MEAN: ? KNOWN

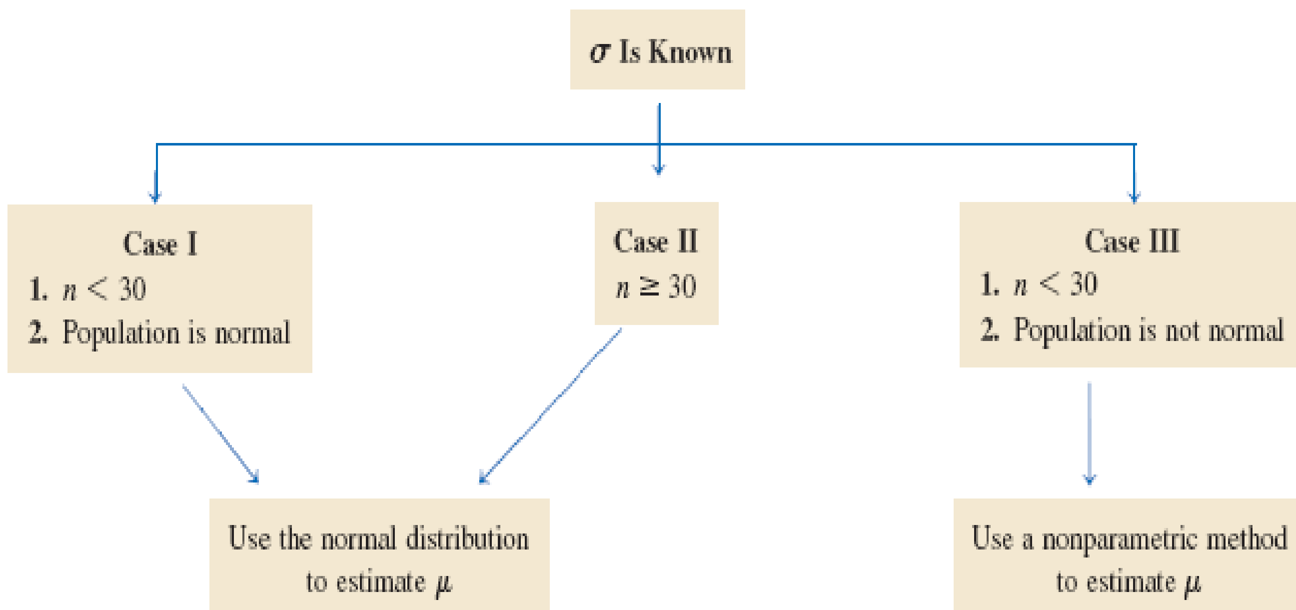
Three Possible Cases

Case III. If the following three conditions are fulfilled:

1. The population standard deviation σ is known
2. The sample size is small (i.e., n < 30)
3. The population from which the sample is selected is not normally distributed (or its distribution is unknown).

ESTIMATION OF A POPULATION MEAN: σ KNOWN

Three Possible Cases



ESTIMATION OF A POPULATION MEAN: σ KNOWN

Confidence Interval for μ

The $(1 - \alpha)100\%$ ***confidence interval for μ*** under Cases I and II is

$$\bar{x} \pm z\sigma_{\bar{x}}$$

$$\text{where} \quad \sigma_{\bar{x}} = \sigma / \sqrt{n}$$

The value of z used here is obtained from the standard normal distribution table (Table IV of Appendix C) for the given confidence level.

ESTIMATION OF A POPULATION MEAN: σ KNOWN

Definition

The ***margin of error for the estimate for μ***, denoted by E, is the quantity that is subtracted from and added to the value of $\bar{x}$ to obtain a confidence interval for μ. Thus,

$$E = z\sigma_{\bar{x}}$$

Figure 8.2 Finding z for a 95% confidence level.

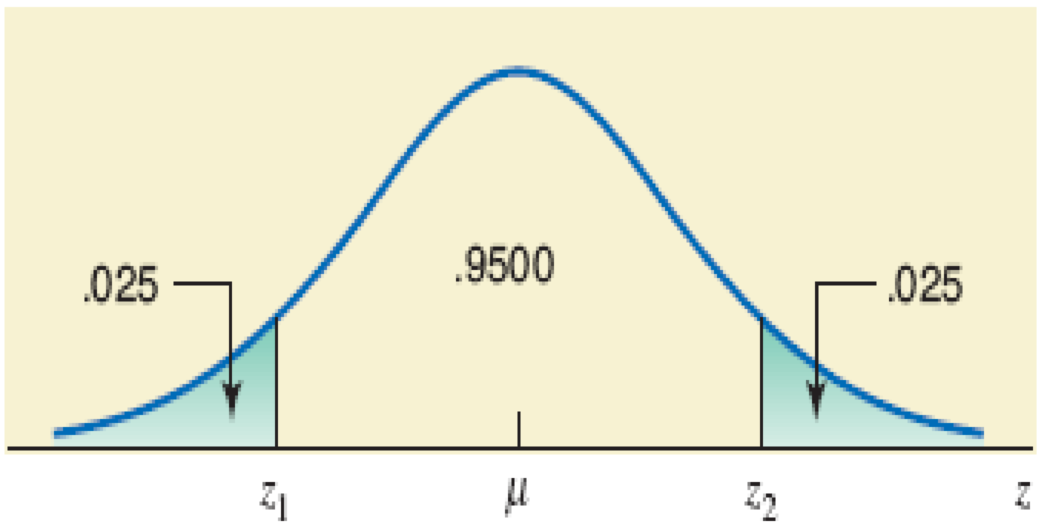


Figure 8.3 Area in the tails.

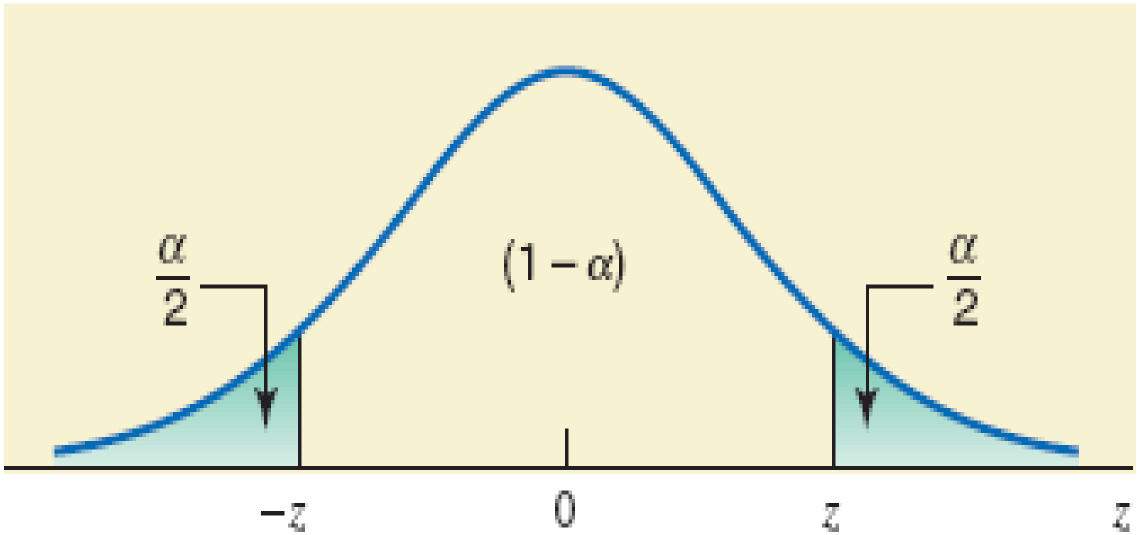


Table 8.1 *z* Values for Commonly Used Confidence Levels

Confidence Level	Areas to Look for in Table IV	*z* Value
90%	.0500 and .9500	1.64 or 1.65
95%	.0250 and .9750	1.96
96%	.0200 and .9800	2.05
97%	.0150 and .9850	2.17
98%	.0100 and .9900	2.33
99%	.0050 and .9950	2.57 or 2.58

Example 8-1

A publishing company has just published a new college textbook. Before the company decides the price at which to sell this textbook, it wants to know the average price of all such textbooks in the market. The research department at the company took a sample of 25 comparable textbooks and collected information on their prices. This information produces a mean price of $145 for this sample. It is known that the standard deviation of the prices of all such textbooks is $35 and the population of such prices is normal.

Example 8-1

(a) What is the point estimate of the mean price of all such textbooks?

(b) Construct a 90% confidence interval for the mean price of all such college textbooks.

Example 8-1: Solution

a)

n = 25, $\bar{x}$ = \$145, and σ = \$35

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{35}{\sqrt{25}} = \mathbf{\$7.00}$$

Point estimate of μ = $\bar{x}$ = **\$145**

Example 8-1: Solution

b)

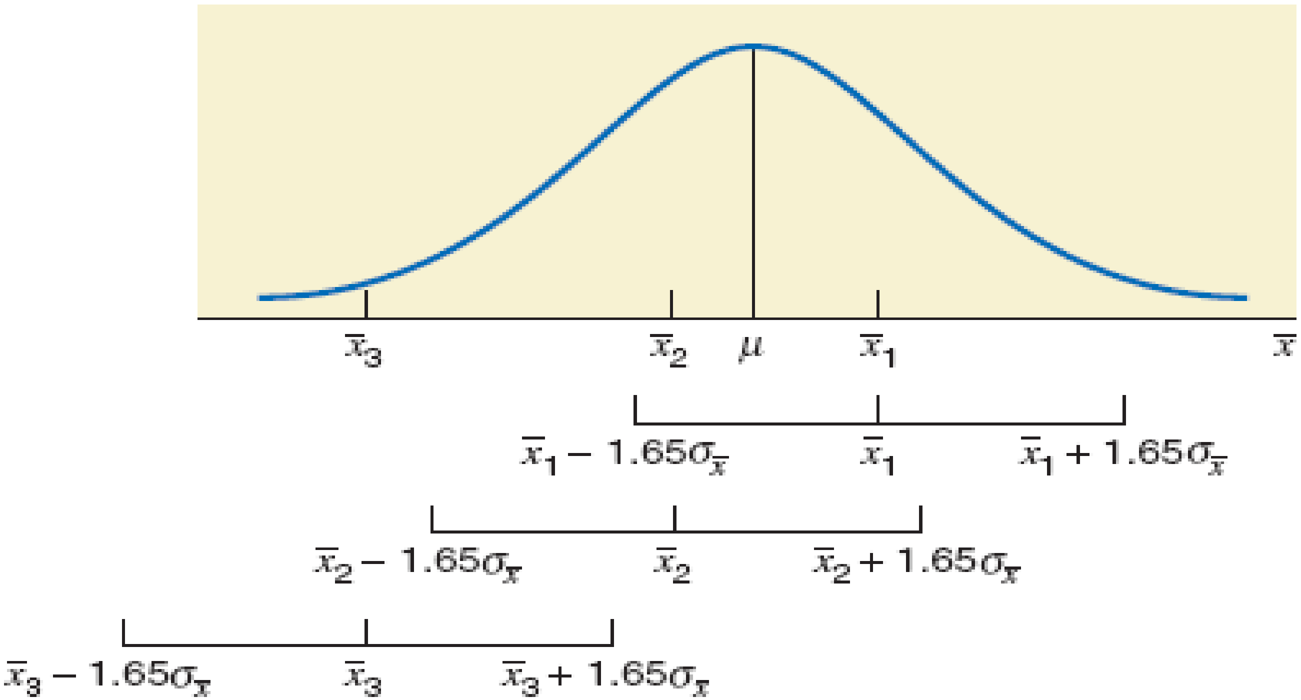
Confidence level is 90% or .90. Here, the area in each tail of the normal distribution curve is $\alpha/2=(1-.90)/2=.05$. Hence, z = 1.65.

$$\bar{x} \pm z\sigma_{\bar{x}} = 145 \pm 1.65(7.00) = 145 \pm 11.55$$
$$= (145-11.55) \text{ to } (145+11.55)$$
$$= \$133.45 \text{ to } \$156.55$$

Example 8-1: Solution

We can say that we are 90% confident that the mean price of all such college textbooks is between $133.45 and $156.55.

Figure 8.4 Confidence intervals.



Example 8-2

In a 2009 survey by I-Pension LLC, adults with annual household incomes of $50,000 to $125,000 were asked about the average time they spend reviewing their 401(k) statements. Of the adults surveyed, about 72% said that they spend less than 5 minutes, and 27% said that they spend 5 to 10 minutes to review their 401(k) statements (*USA TODAY*, February 16, 2009). Suppose a random sample of 400 adults of all income levels, who have 401(k) statements, were asked. The sample produced a mean of 8 minutes. Assume that the standard deviation of such times for all 401(k) account holders is 2.20 minutes. Construct a 99% confidence interval for the mean time spent by all 401(k) holders reviewing their 401(k) statements.

Example 8-2: Solution

- Confidence level 99% or .99
- The sample size is large (n > 30)
  - Therefore, we use the normal distribution
  - z = 2.58

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2.20}{\sqrt{400}} = .11$$

Example 8-2: Solution

$$\bar{x} \pm z\sigma_{\bar{x}} = 8 \pm 2.58(.11) = 8 \pm .28$$

$$= \mathbf{7.72 \quad to \quad 8.28 \ minutes}$$

Thus, we can state with 99% confidence that the mean time spent by all 401(k) account holders reviewing their 401(k) statements is between 7.72 and 8.28 minutes.

ESTIMATION OF A POPULATION MEAN: σ KNOWN

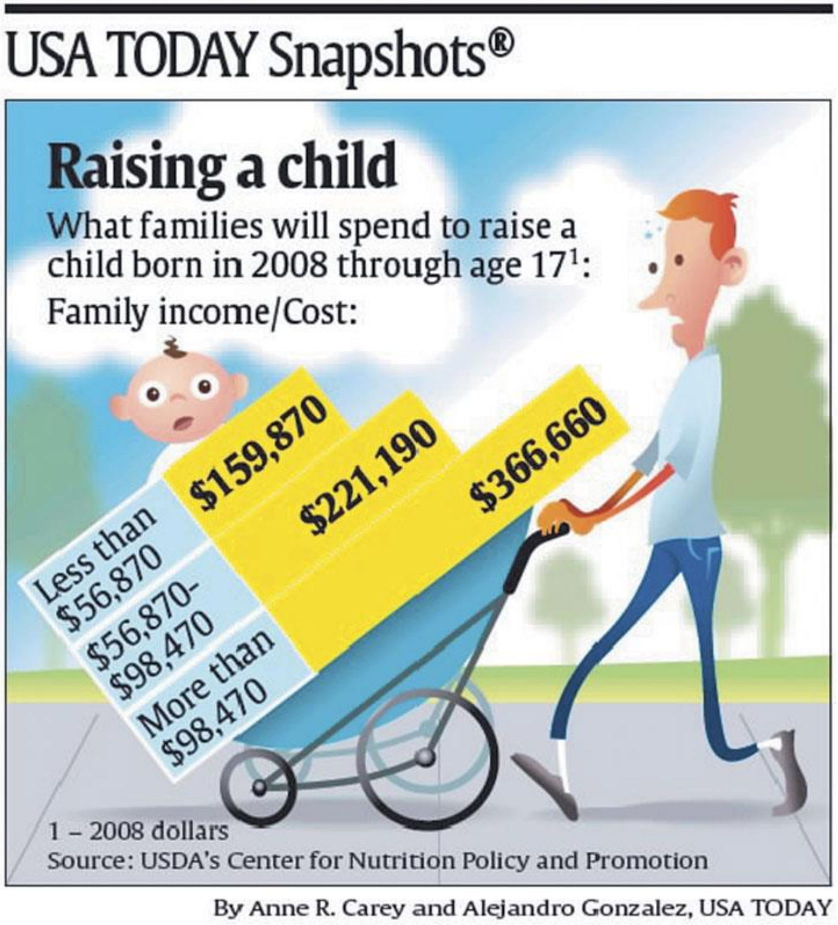
Width of a Confidence Interval

The width of a confidence interval depends on the size of the margin of error, $z\sigma_{\bar{x}}$. Hence, the width of a confidence interval can be controlled using

1. The value of z, which depends on the confidence interval

2. The sample size, n

Case Study 8-1 Raising a Child



Determining the Sample Size for the Estimation of Mean

Determining the Sample Size for the Estimation of μ

Given the confidence level and the standard deviation of the population, the sample size that will produce a predetermined margin of error *E* of the confidence interval ***estimate of μ*** is

$$n = \frac{z^2 \sigma^2}{E^2}$$

Example 8-3

An alumni association wants to estimate the mean debt of this year's college graduates. It is known that the population standard deviation of the debts of this year's college graduates is $11,800. How large a sample should be selected so that the estimate with a 99% confidence level is within $800 of the population mean?

Example 8-3: Solution

- The maximum size of the margin of error of estimate is to be \$800; that is, E = \$800.
- The value of z for a 99% confidence level is z = 2.58.
- The value of σ is \$11,800.

$$n = \frac{z^2\sigma^2}{E^2} = \frac{(2.58)^2(11,800)^2}{(800)^2} = 1448.18 \approx 1449$$

- Thus, the required sample size is 1449.

8.4 Estimation of a Population Mean: σ Not Known

Three Possible Cases

Case I. If the following three conditions are fulfilled:

1. The population standard deviation σ is not known
2. The sample size is small (i.e., $n < 30$)
3. The population from which the sample is selected is normally distributed.

ESTIMATION OF A POPULATION MEAN: σ NOT KNOWN

Three Possible Cases

Case II. If the following two conditions are fulfilled:

1. The population standard deviation σ is not known
2. The sample size is large (i.e., $n \geq 30$)

ESTIMATION OF A POPULATION MEAN: σ NOT KNOWN

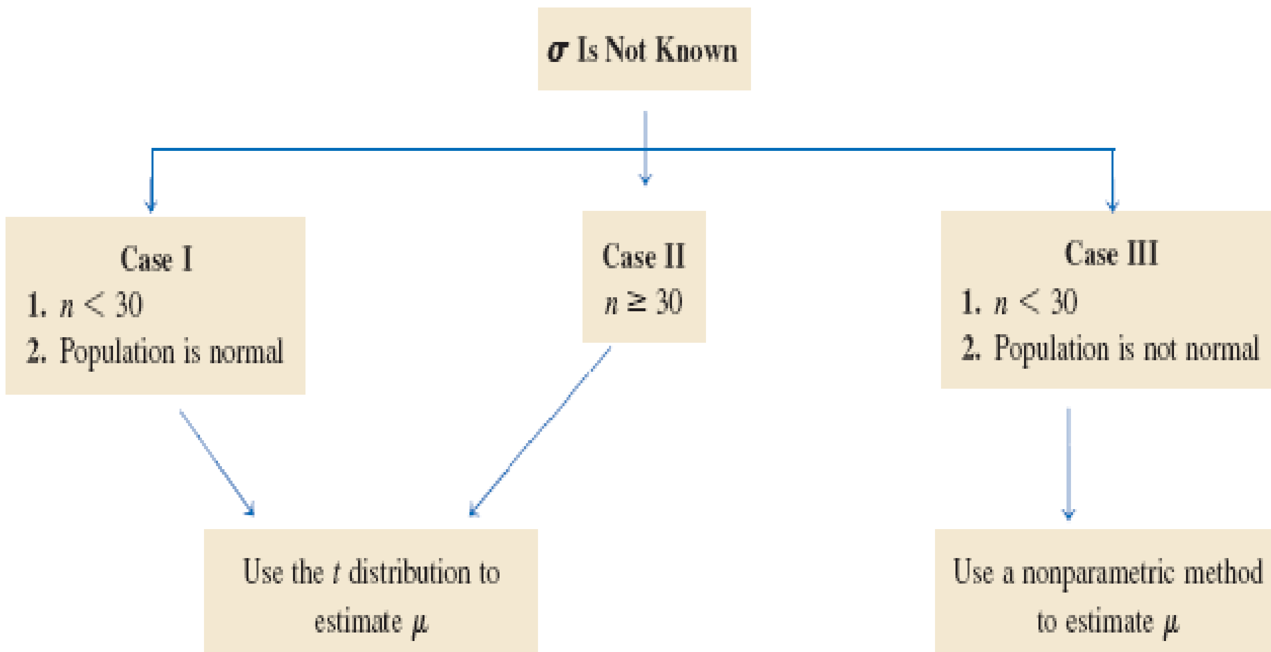
Three Possible Cases

Case III. If the following three conditions are fulfilled:

1. The population standard deviation σ is not known
2. The sample size is small (i.e., n < 30)
3. The population from which the sample is selected is not normally distributed (or its distribution is unknown).

ESTIMATION OF A POPULATION MEAN: σ NOT KNOWN

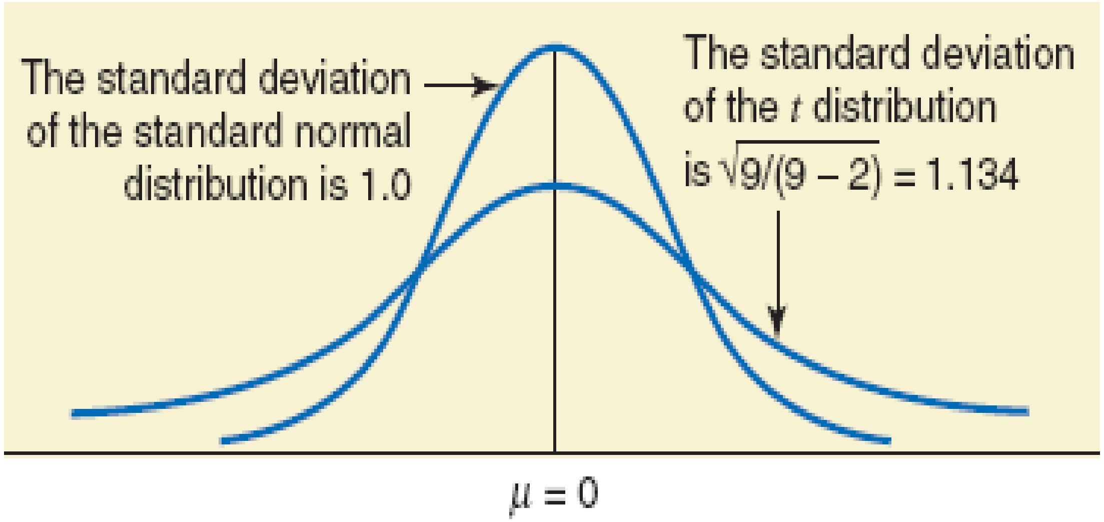
Three Possible Cases



The t Distribution

The ***t distribution*** is a specific type of bell-shaped distribution with a lower height and a wider spread than the standard normal distribution. As the sample size becomes larger, the t distribution approaches the standard normal distribution. The t distribution has only one parameter, called the degrees of freedom (df). The mean of the t distribution is equal to 0 and its standard deviation is $\sqrt{df/(df-2)}$.

Figure 8.5 The *t* distribution for $df = 9$ and the standard normal distribution.



Example 8-4

Find the value of t for 16 degrees of freedom and .05 area in the right tail of a t distribution curve.

Table 8.2 Determining t for 16 *df* and .05 Area in the Right Tail

Area in the right tail

	Area in the Right Tail Under the t Distribution Curve				
df	**.10**	**.05**	**.025**	**. . .**	**.001**
1	3.078	6.314	12.706	. . .	318.309
2	1.886	2.920	4.303	. . .	22.327
3	1.638	2.353	3.182	. . .	10.215
.	. . .	. . .	. . .	. . .	. . .
.	. . .	. . .	. . .	. . .	. . .
.	. . .	. . .	. . .	. . .	. . .
df → **16**	1.337	**1.746**	2.120	. . .	3.686
.	. . .	. . .	. . .	. . .	. . .
.	. . .	. . .	. . .	. . .	. . .
.	. . .	. . .	. . .	. . .	. . .
75	1.293	1.665	1.992	. . .	3.202
∞	1.282	1.645	1.960	. . .	3.090

The required value of t for 16 *df* and .05 area in the right tail

Figure 8.6 The value of *t* for 16 *df* and .05 area in the right tail.

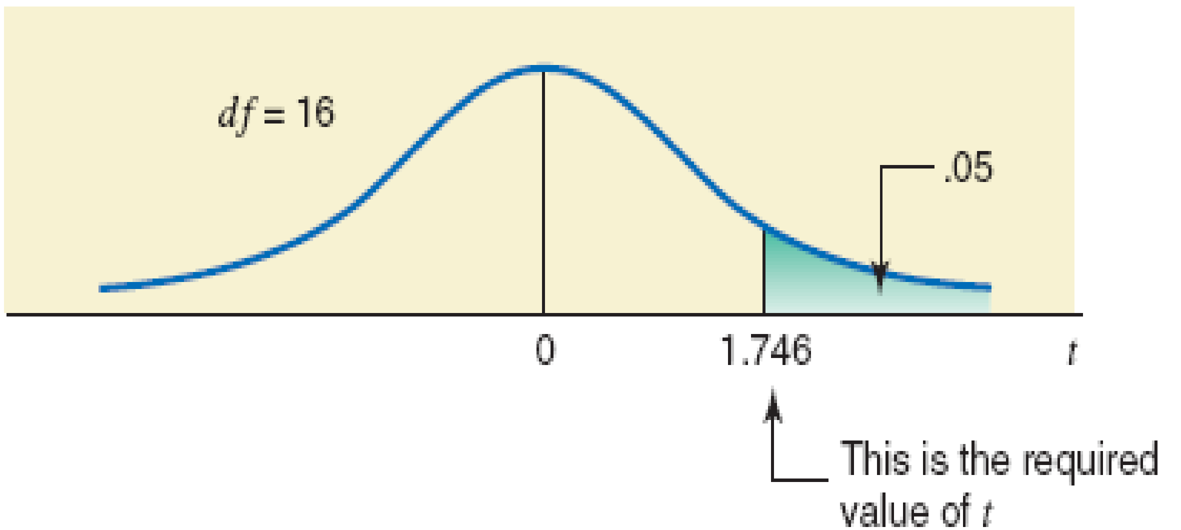
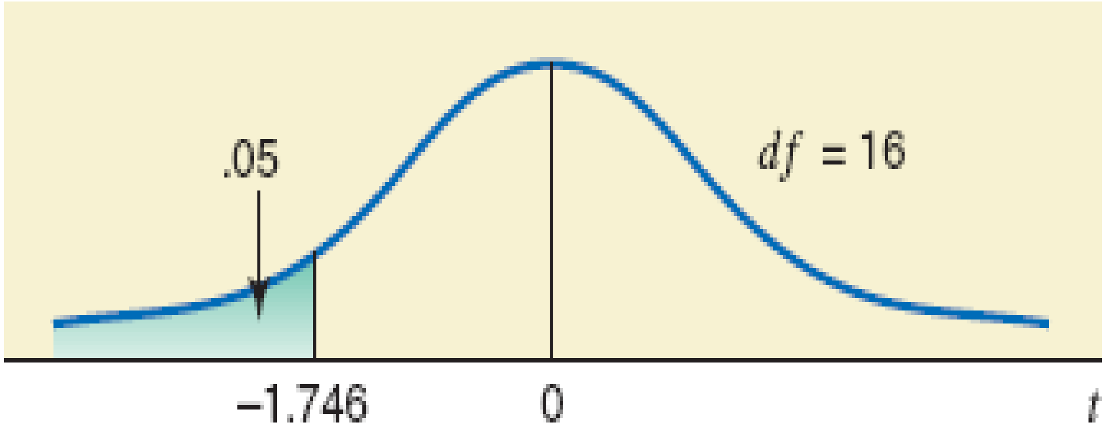


Figure 8.7 The value of *t* for 16 *df* and .05 area in the left tail.



Confidence Interval for μ Using the *t* Distribution

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

$$\bar{x} \pm t s_{\bar{x}} \quad \text{where} \quad s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

The value of t is obtained from the t distribution table for $n - 1$ degrees of freedom and the given confidence level. Here $t s_{\bar{x}}$ is the margin of error of the estimate; that is,

$$E = t s_{\bar{x}}$$

Example 8-5

Dr. Moore wanted to estimate the mean cholesterol level for all adult men living in Hartford. He took a sample of 25 adult men from Hartford and found that the mean cholesterol level for this sample is 186 mg/dL with a standard deviation of 12 mg/dL. Assume that the cholesterol levels for all adult men in Hartford are (approximately) normally distributed. Construct a 95% confidence interval for the population mean μ.

Example 8-5: Solution

- σ is not known, *n* < 30, and the population is normally distributed (Case I)
- Use the t distribution to make a confidence interval for μ
- **$n=25$, $\bar{x}=186$, $s=12$, and confidence *level* $= 95\%$**
- $$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{12}{\sqrt{25}} = 2.40$$

Example 8-5: Solution

- $df = n - 1 = 25 - 1 = 24$
- Area in each tail = .5 – (.95/2)
  = .5 - .4750 = .025
- The value of t in the right tail is 2.064

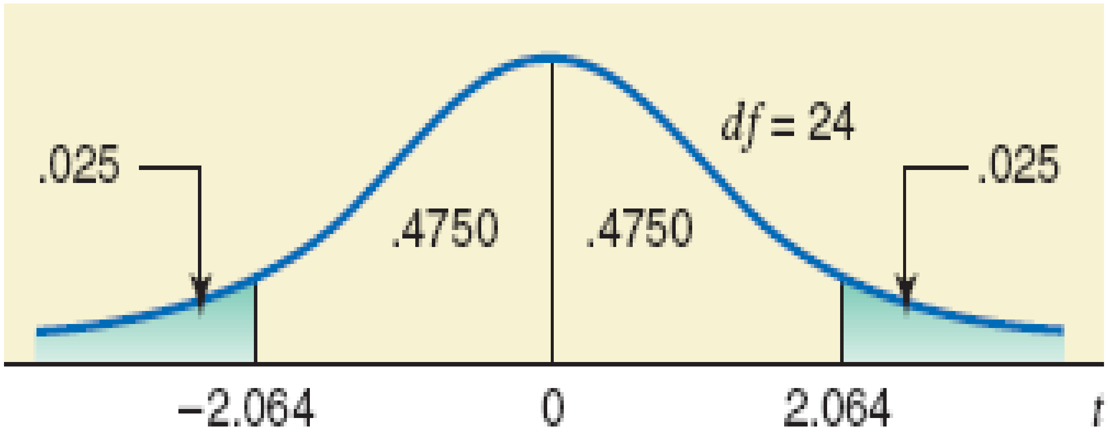
$$\bar{x} \pm t s_{\bar{x}} = 186 \pm 2.064(2.40) = 186 \pm 4.95$$

$$= 181.05 \quad \text{to} \quad 190.95$$

Example 8-5: Solution

- Thus, we can state with 95% confidence that the mean cholesterol level for all adult men living in Harford lies between 181.05 and 190.95 mg/dL.

Figure 8.8 The value of t.



Example 8-6

Sixty-four randomly selected adults who buy books for general reading were asked how much they usually spend on books per year. The sample produced a mean of $1450 and a standard deviation of $300 for such annual expenses. Determine a 99% confidence interval for the corresponding population mean.

Example 8-6: Solution

- σ is not known, n > 30 (Case II)
- Use the t distribution to make a confidence interval for μ
- **$n=64$, $\bar{x}=\$1450$, $s=\$300$, and confidence *level* $= 99\%$**

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{300}{\sqrt{64}} = \$37.50$$

Example 8-6: Solution

- $df = n - 1 = 64 - 1 = 63$
- Area in each tail = .5 – (.99/2)
  = .5 - .4950 = .005
- The value of t in the right tail is 2.656

$$\bar{x} \pm t s_{\bar{x}} = \$1450 \pm 2.656(37.50) = \$1450 \pm \$99.60$$
$$= \$1350.40 \quad \text{to} \quad \$1549.60$$

Example 8-6: Solution

- Thus, we can state with 99% confidence that based on this sample the mean annual expenditure on books by all adults who buy books for general reading is between $1350.40 and $1549.60.

Confidence Interval for μ Using the *t* Distribution

What If the Sample Size Is Too Large?

1. Use the *t* value from the last row (the row of ∞) in Table V.
2. Use the normal distribution as an approximation to the t distribution.

8.5 Estimation of a Population Proportion: Large Samples

Estimator of the Standard Deviation of $\hat{p}$

The value of $s_{\hat{p}}$, which gives a point estimate of $\sigma_{\hat{p}}$, is calculated as follows. Here, $s_{\hat{p}}$ is an estimator of $\sigma_{\hat{p}}$

$$s_{\hat{p}} = \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

ESTIMATION OF A POPULATION PROPORTION: LARGE SAMPLES

Confidence Interval for the Population Proportion, p

The $(1 - \alpha)100\%$ ***confidence interval for the population proportion***, p, is

$$\hat{p} \pm z s_{\hat{p}}$$

The value of z used here is obtained from the standard normal distribution table for the given confidence level, and $s_{\hat{p}} = \sqrt{\hat{p}\hat{q}/n}$. The term $\boldsymbol{z s_{\hat{p}}}$ is called the ***margin of error***, *E*.

Example 8-7

According to a survey by Pew Research Center in June 2009, 44% of people aged 18 to 29 years said that religion is very important to them. Suppose this result is based on a sample of 1000 people aged 18 to 29 years.

Example 8-7

a) What is the point estimate of the population proportion?
b) Find, with a 99% confidence level, the percentage of all people aged 18 to 29 years who will say that religion is very important to them. What is the margin of error of this estimate?

Example 8-7: Solution

- $n = 1000, \quad \hat{p} = .44, \quad \text{and,} \quad \hat{q} = .56$
- $$s_{\hat{p}} = \sqrt{\frac{\hat{p}\hat{q}}{n}} = \sqrt{\frac{(.44)(.56)}{1000}} = .01569713$$
- Note that $n\hat{p}$ and $n\hat{q}$ are both greater than 5.

Example 8-7: Solution

a)

Point estimate of p = $\hat{p}$ = .44

Example 8-7: Solution

b)

The confidence level is 99%, or .99. z = 2.58.

$$\hat{p} \pm z s_{\hat{p}} = .44 \pm 2.58(.01569713) = .44 \pm .04$$

$$= \ .40 \ \text{ to } \ .48 \ \text{ or } \ 40\% \ \text{ to } 48\%$$

Margin of error = $\pm 1.96\, s_{\hat{p}}$

= $\pm 1.96(.01569713)$

= $\pm .04$ or $\pm 4\%$

Example 8-8

According to a Harris Interactive survey of 2401 adults conducted in April 2009, 25% of adults do not drink alcohol. Construct a 97% confidence interval for the corresponding population proportion.

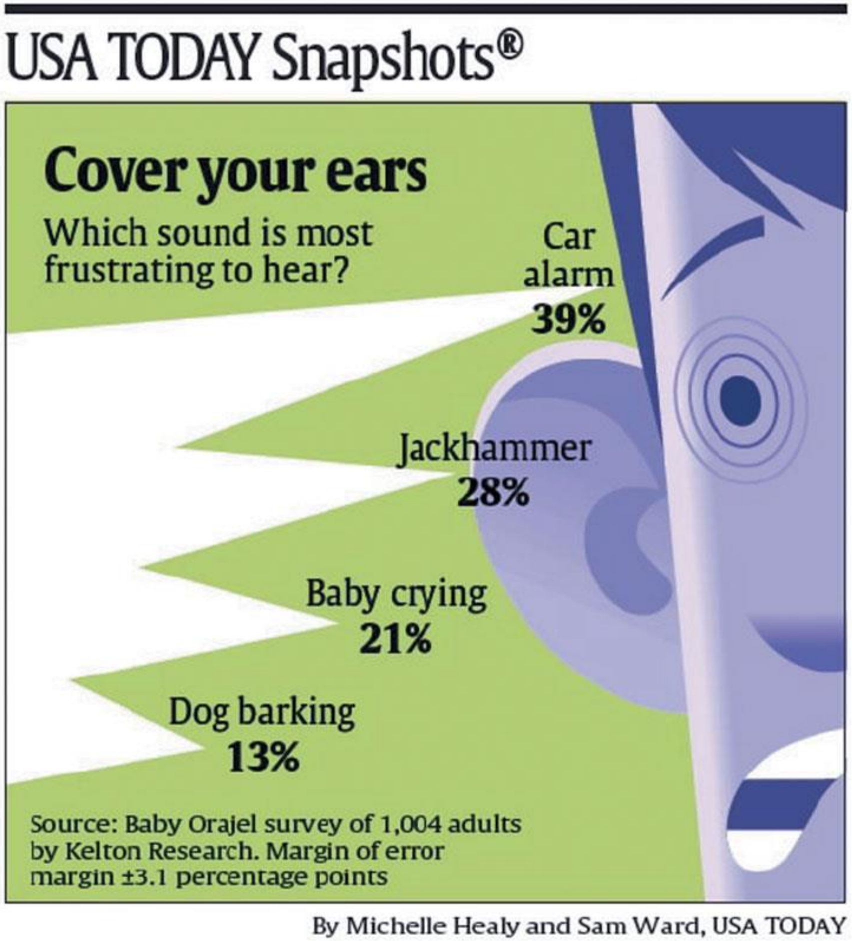
Example 8-8: Solution

- Confidence level = 97% or .97
- $s_{\hat{p}} = \sqrt{\dfrac{\hat{p}\hat{q}}{n}} = \sqrt{\dfrac{(.25)(.75)}{2401}} = .00883699$
- The value of z for .97 / 2 = .4850 is 2.17.

Example 8-8: Solution

$$\hat{p} \pm z s_{\hat{p}} = .25 \pm 2.17(.00883699)$$
$$= .25 \pm .019$$
$$= \ .231 \ \text{to} \ .269 \ \text{or} \ 23.1\% \ \text{to} \ 26.9\%$$

Case Study 8-2 Which Sound Is The Most Frustrating To Hear?



DETERMINING THE SAMPLE SIZE FOR THE ESTIMATION OF PROPORTION

Given the confidence level and the values of $\hat{p}$ and $\hat{q}$, the sample size that will produce a predetermined maximum of error E of the confidence interval ***estimate of p*** is

$$n = \frac{z^2 \hat{p}\hat{q}}{E^2}$$

DETERMINING THE SAMPLE SIZE FOR THE ESTIMATION OF PROPORTION

In case the values of $\hat{p}$ and $\hat{q}$ are not known

1. We make the most conservative estimate of the sample size n by using $\hat{p} = .5$ and $\hat{q} = .5$

2. We take a preliminary sample (of arbitrarily determined size) and calculate $\hat{p}$ and $\hat{q}$ from this sample. Then use these values to find n.

Example 8-9

Lombard Electronics Company has just installed a new machine that makes a part that is used in clocks. The company wants to estimate the proportion of these parts produced by this machine that are defective. The company manager wants this estimate to be within .02 of the population proportion for a 95% confidence level. What is the most conservative estimate of the sample size that will limit the maximum error to within .02 of the population proportion?

Example 8-9: Solution

- The value of z for a 95% confidence level is 1.96.
- $\hat{p} = .50 \ \ and \ \ \hat{q} = .50$
- $$n = \frac{z^2 \hat{p}\hat{q}}{E^2} = \frac{(1.96)^2(.50)(.50)}{(.02)^2} = 2401$$
- Thus, if the company takes a sample of 2401 parts, there is a 95% chance that the estimate of p will be within .02 of the population proportion.

Example 8-10

Consider Example 8-9 again. Suppose a preliminary sample of 200 parts produced by this machine showed that 7% of them are defective. How large a sample should the company select so that the 95% confidence interval for p is within .02 of the population proportion?

Example 8-10: Solution

$\hat{p} = .07$ and $\hat{q} = .93$

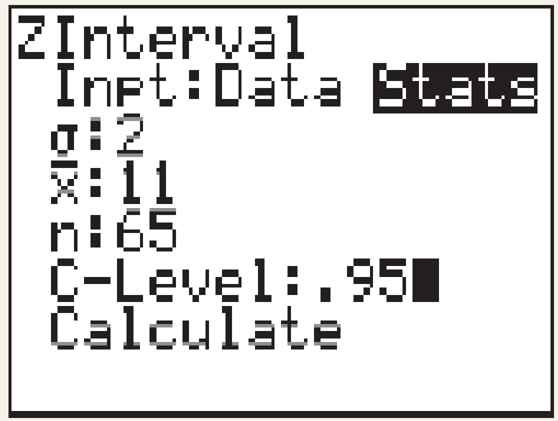
$$n = \frac{z^2 \hat{p}\hat{q}}{E^2} = \frac{(1.96)^2(.07)(.93)}{(.02)^2}$$

$$= \frac{(3.8416)(.07)(.93)}{.0004} = 625.22 \approx 626$$

TECHNOLOGY INSTRUCTION

Confidence Intervals for Population Means and Proportions

TI-84

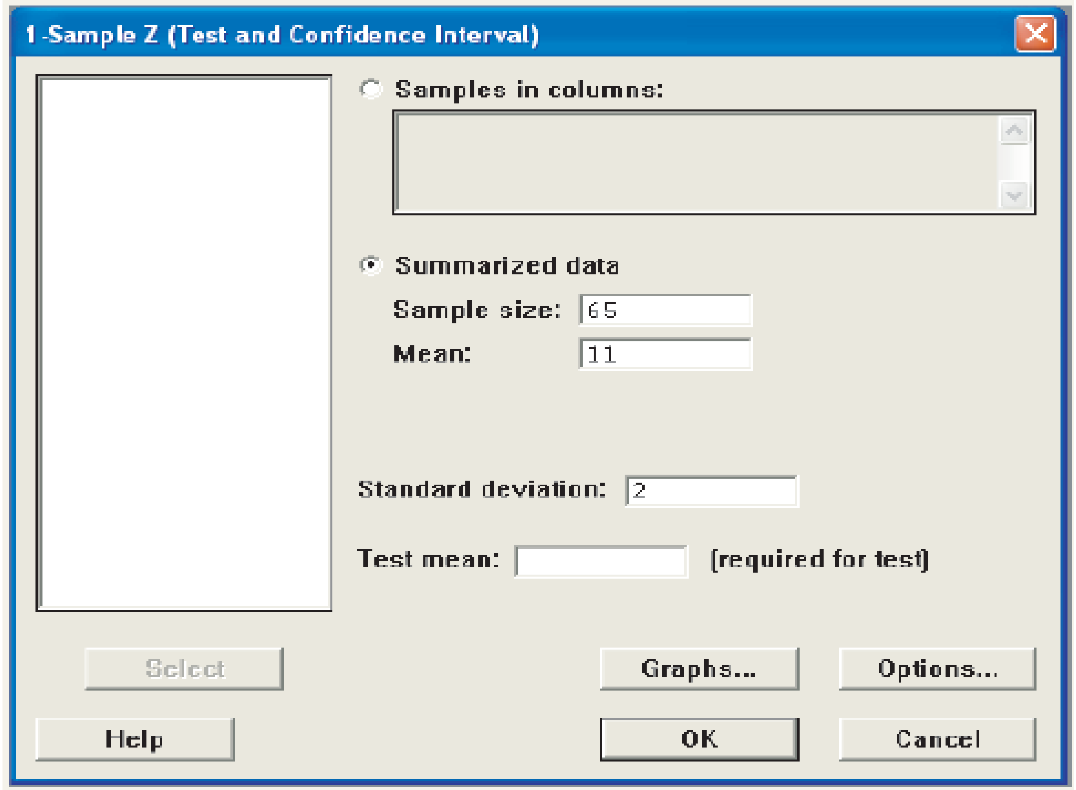


Screen 8.1

TECHNOLOGY INSTRUCTION

Confidence Intervals for Population Means and Proportions

Minitab

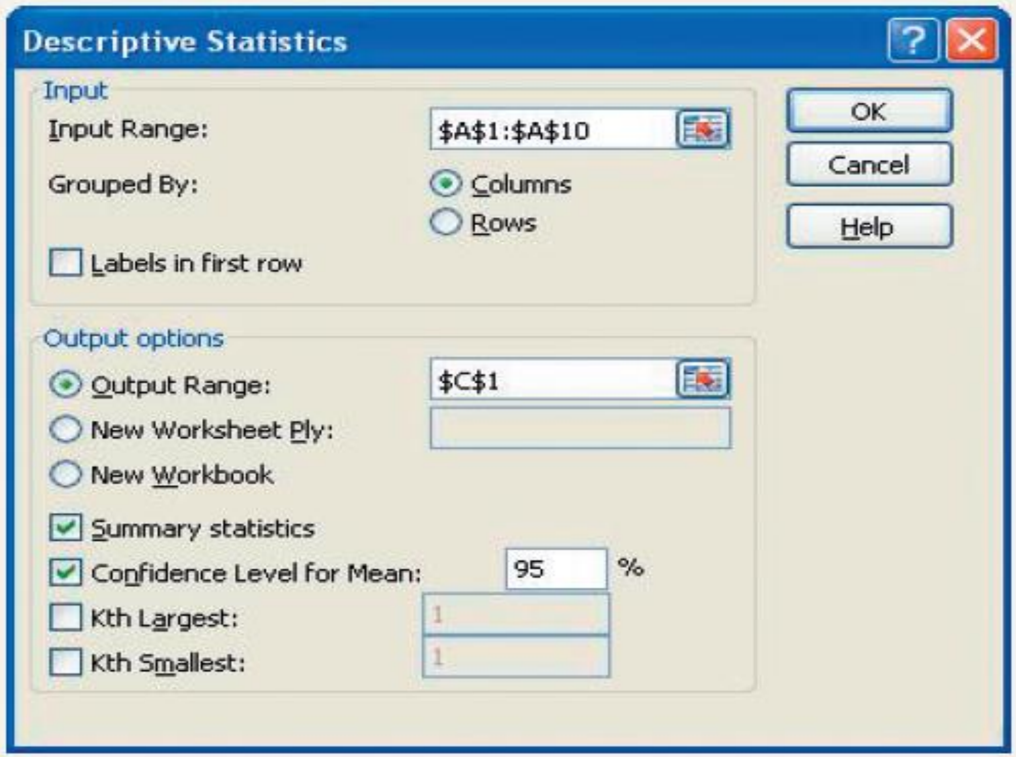


Screen 8.2

TECHNOLOGY INSTRUCTION

Confidence Intervals for Population Means and Proportions

Excel



Screen 8.3

TECHNOLOGY INSTRUCTION

Confidence Intervals for Population Means and Proportions

Excel

C	D
Column1	
Mean	34.4
Standard Error	9.443634
Median	25.5
Mode	#N/A
Standard Deviation	29.86339
Sample Variance	891.8222
Kurtosis	0.898782
Skewness	1.171464
Range	93
Minimum	5
Maximum	98
Sum	344
Count	10
Confidence Level(95.0%)	21.36298

Screen 8.4

TECHNOLOGY INSTRUCTION

Confidence Intervals for Population Means and Proportions

Excel

	A	B
1	Mean	11
2	Std. Dev.	2
3	Size	65
4	Alpha	0.05
5		
6	Margin of error	=CONFIDENCE(0.05,2,65)

Screen 8.5

TECHNOLOGY INSTRUCTION

Confidence Intervals for Population Means and Proportions

Excel

	A	B
1	Mean	11
2	Std. Dev.	2
3	Size	65
4	Alpha	0.05
5		
6	Margin of error	0.486207225

Screen 8.6