Business Research Methods 9e

Zikmund
Babin
Carr
Griffin

Chapter 17
Determination of Sample Size: A Review of Statistical Theory

©2013 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

LEARNING OUTCOMES

1. Understand basic statistical terminology
2. Interpret frequency distributions, proportions, and measures of central tendency and dispersion
3. Distinguish among population, sample, and sampling distributions
4. Explain the central-limit theorem
5. Summarize the use of confidence interval estimates
6. Discuss major issues in specifying sample size

©2013 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Federal Reserve Finds Cards are Replacing Cash

- Payment options have gone high-tech and research supports that.
- The Federal Reserve conducted surveys of depository institutions, asking them to report the number of each type of payment the institutions processed.
- A sample of 2,700 was drawn from the 14,117 institutions in the population, with a 95% confidence interval and accuracy of ±5 percent.

©2013 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Introduction

- Descriptive Statistics
  - *Describe characteristics of populations or samples.*
- Inferential Statistics
  - *Make inferences about whole populations from a sample.*
- Sample Statistics
  - *Variables in a sample or measures computed from sample data.*
- Population Parameters
  - *Variables in a population or measured characteristics of the population.*

©2013 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Making Data Usable

- Frequency Distribution
  - *A set of data organized by summarizing the number of times a particular value of a variable occurs.*
- Percentage Distribution
  - *A frequency distribution organized into a table (or graph) that summarizes percentage values associated with particular values of a variable.*
- Probability
  - *The long-run relative frequency with which an event will occur.*

©2013 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

EXHIBIT 17.1 Frequency Distribution of Deposits

Amount	Frequency (Number of People Who Hold Deposits in Each Range)
Under $3,000	499
$3,000–$5,999	530
$6,000–$8,999	562
$9,000–$11,999	718
$12,000 or more	811
	3,120

EXHIBIT 17.2 Percentage Distribution of Deposits

Amount	Percent (Percentage of People Who Hold Amount Deposits in Each Range)
Under $3,000	16%
$3,000–$5,999	17%
$6,000–$8,999	18%
$9,000–$11,999	23%
$12,000 or more	26%
	100%

EXHIBIT 17.3 Probability Distribution of Deposits

Amount	Probability
Under $3,000	0.16
$3,000–$5,999	0.17
$6,000–$8,999	0.18
$9,000–$11,999	0.23
$12,000 or more	0.26
	1.00

The Well-Chosen Average

- "Average" pay could mean there's a highly paid executive and many low-paid employees.
- Median may be more informative.

©2013 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Making Data Usable (cont'd)

- Proportion
  - *The percentage of elements that meet some criterion*
- Measures of Central Tendency
  - ***Mean:*** *the arithmetic average.*
  - ***Median:*** *the midpoint; the value below which half the values in a distribution fall.*
  - ***Mode:*** *the value that occurs most often.*

Population Mean

$$\mu = \frac{\sum_{i=1}^{n} X}{N}$$

Sample Mean

$$\overline{X} = \frac{\sum_{i=1}^{n} X}{n}$$

©2013 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

EXHIBIT 17.4

Number of Sales Calls Per Day by Salesperson

Salesperson	Number of Sales Calls
Mike	4
Patty	3
Billie	2
Bob	5
John	3
Frank	3
Chuck	1
Samantha	5
Total	26

Measures of Dispersion

- The Range
  - *The distance between the smallest and the largest values of a frequency distribution.*

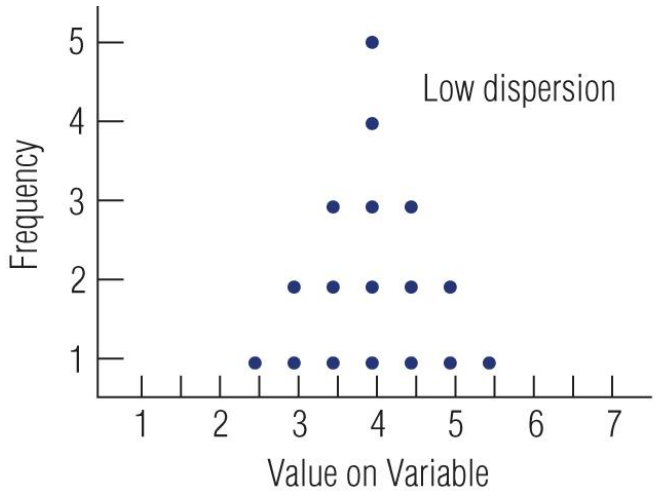
©2013 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

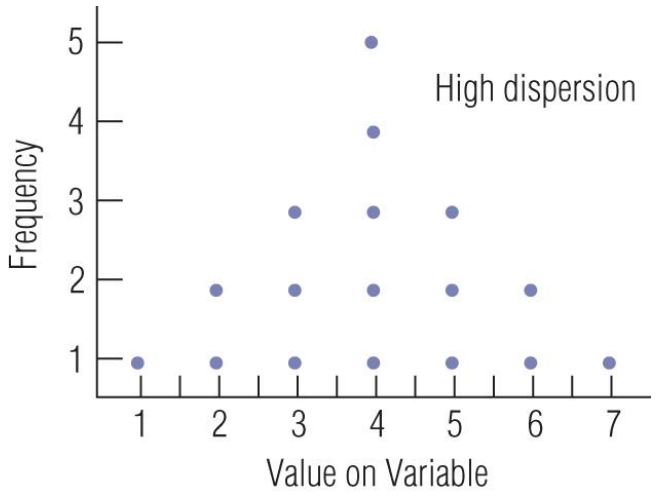
EXHIBIT 17.5

Sales Levels for Two Products with Identical Average Sales

	Units Product A	Units Product B
January	196	150
February	198	160
March	199	176
April	200	181
May	200	192
June	200	200
July	200	201
August	201	202
September	201	213
October	201	224
November	202	240
December	202	261
Average	**200**	**200**

EXHIBIT 17.6 Low Dispersion versus High Dispersion





Measures of Dispersion (cont'd)

- Why Use the Standard Deviation?
  - *Variance*
    - A measure of variability or dispersion.
    - Its square root is the standard deviation.
  - *Standard deviation*
    - A quantitative index of a distribution's spread, or variability; the square root of the variance for a distribution.
    - The average of the amount of variance for a distribution.
    - Used to calculate the likelihood (probability) of an event occurring.

©2013 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Calculating Deviation

$$\text{Average deviation} = \frac{\sum (X_i - \bar{X})}{n}$$

$$\text{Mean absolute deviation} = \frac{\sum |X_i - \bar{X}|}{n}$$

$$\text{Mean squared deviation} = \frac{\sum (X_i - \bar{X})^2}{n}$$

$$\text{Variance} = S^2 = \frac{\sum (X_i - \bar{X})^2}{n - 1}$$

$$\text{Standard Deviation} = S = \sqrt{S^2} = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n - 1}}$$

©2013 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

EXHIBIT 17.7 Calculating a Standard Deviation

X	$(X - \bar{X})$	$(X - \bar{X})^2$
4	$(4 - 3.25) = .75$	.5625
3	$(3 - 3.25) = -.25$	.0625
2	$(2 - 3.25) = -1.25$	1.5625
5	$(5 - 3.25) = 1.75$	3.0625
3	$(3 - 3.25) = -.25$	.0625
3	$(3 - 3.25) = -.25$	.0625
1	$(1 - 3.25) = -2.25$	5.0625
5	$(5 - 3.25) = 1.75$	3.0625
Σ^a	a^0	13.5000

$n = 8 \qquad \bar{X} = 3.25$

$$S = \sqrt{\frac{\Sigma(X - \bar{X})^2}{n - 1}} = \sqrt{\frac{13.5}{8 - 1}} = \sqrt{\frac{13.5}{7}} = \sqrt{1.9286} = 1.3887$$

[a]The summation of this column is not used in the calculation of the standard deviation.

The Normal Distribution

- Normal Distribution
  - *A symmetrical, bell-shaped distribution (normal curve) that describes the expected probability distribution of many chance occurrences.*
  - *99% of its values are within ± 3 standard deviations from its mean.*
    - Example: IQ scores
- Standardized Normal Distribution
  - *A purely theoretical probability distribution that reflects a specific normal curve for the standardized value, z.*

©2013 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

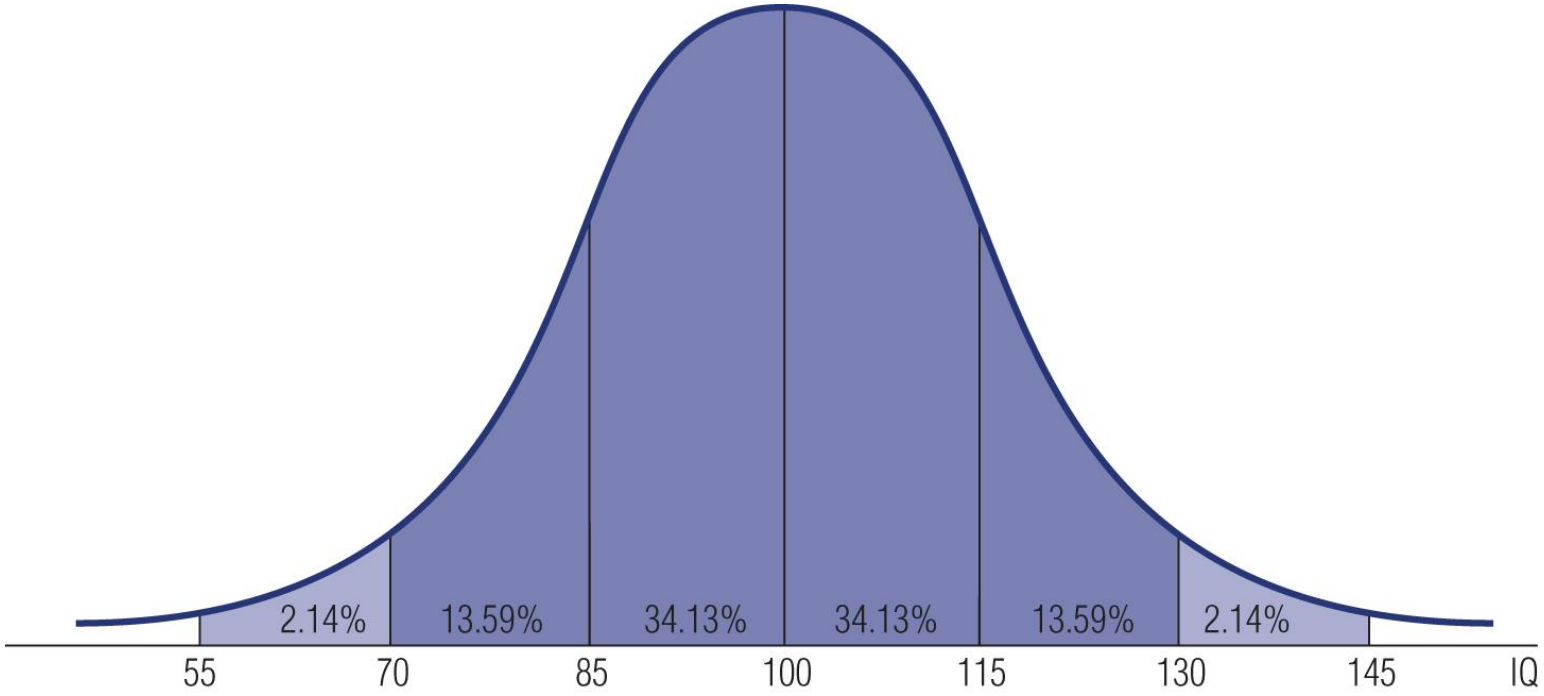
The Normal Distribution (cont'd)

- Characteristics of a Standardized Normal Distribution

1. *It is symmetrical about its mean.*
2. *The mode identifies the normal curve's highest point, which is also the mean and median, and the vertical line about which this normal curve is symmetrical.*
3. *The normal curve has an infinite number of cases (it is a continuous distribution), and the area under the curve has a probability density equal to 1.0.*
4. *The standardized normal distribution has a mean of 0 and a standard deviation of 1.*

©2013 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

EXHIBIT 17.8 Normal Distribution



The Normal Distribution (cont'd)

- Standardized Values
  - *Used to compare an individual value to the population mean in units of the standard deviation*

$$\text{Standardized value} = \frac{\text{Value to be transformed} - \text{Mean}}{\text{Standard deviation}}$$

where

$$Z = \frac{X - \mu}{\sigma}$$

μ = hypothesized or expected value of the mean

©2013 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

EXHIBIT 17.9 Standardized Normal Distribution

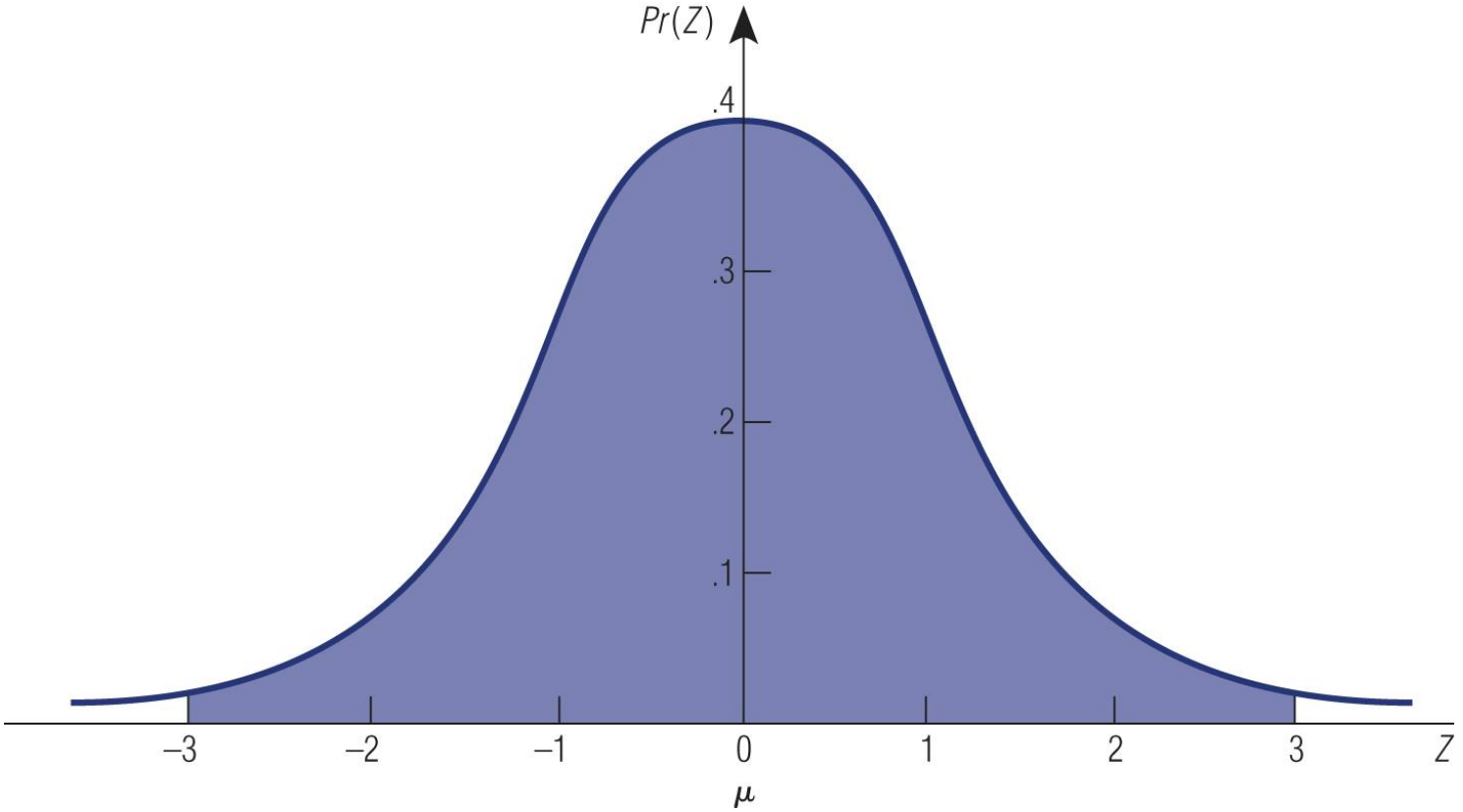


EXHIBIT 17.11

Standardized Values Can Be Computed from Flat or Peaked Distributions Resulting in a Standardized Normal Curve

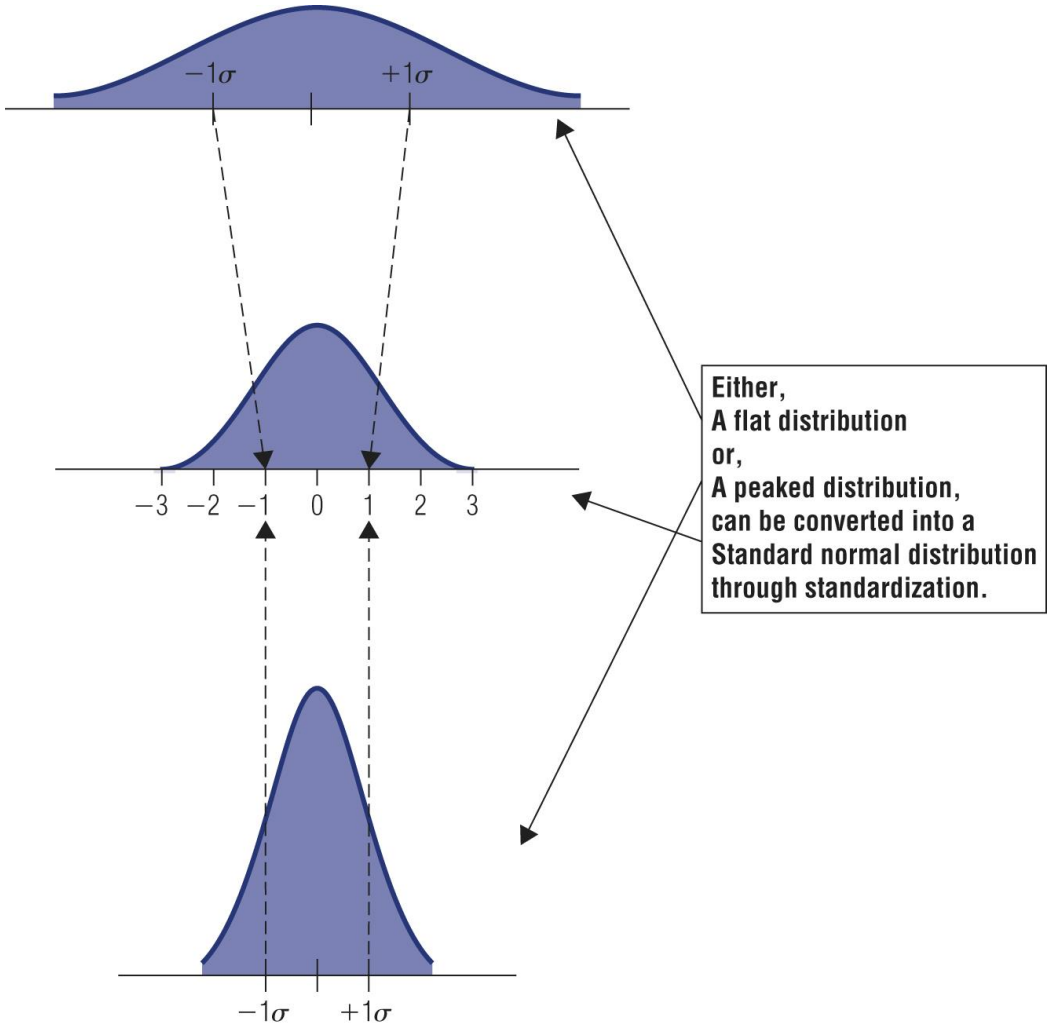
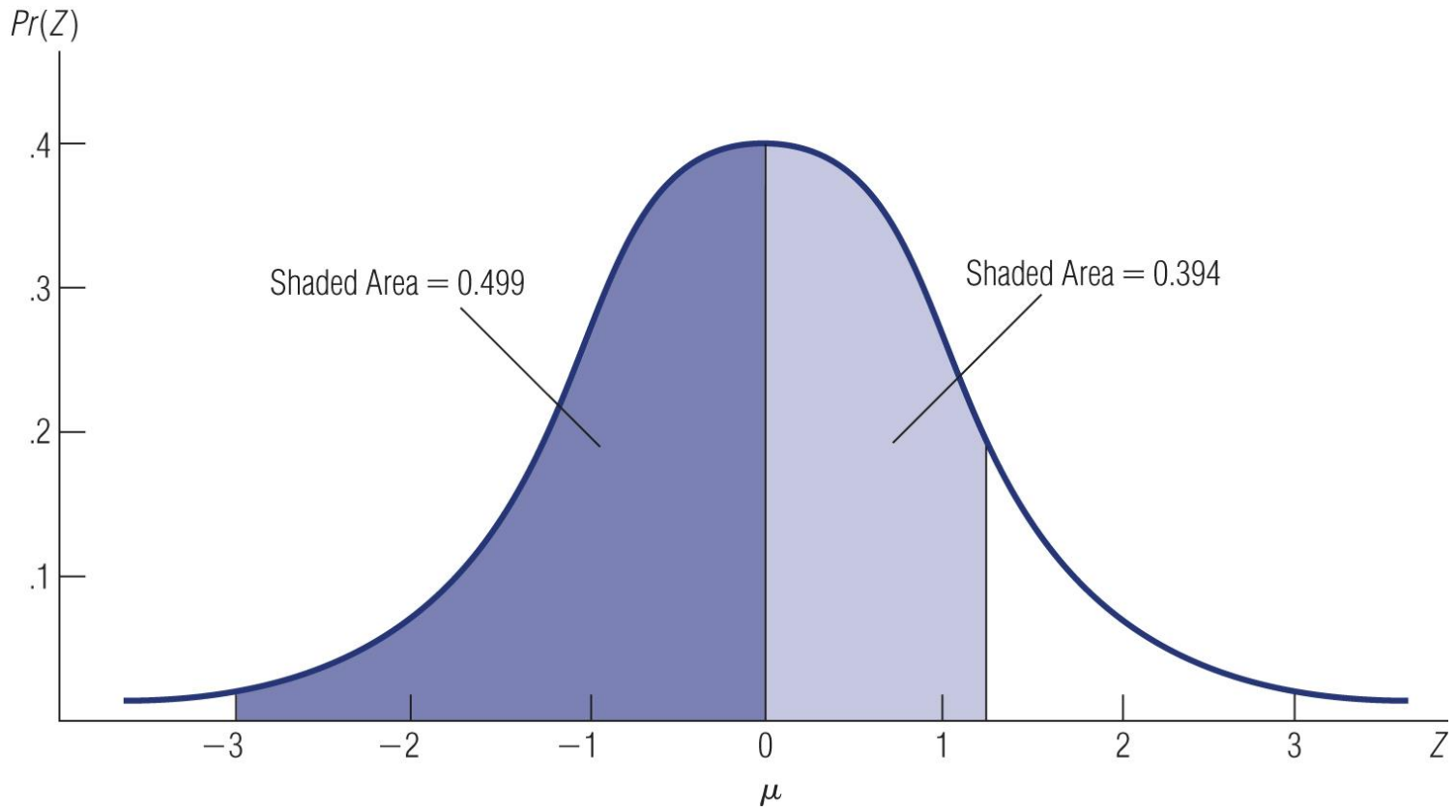


EXHIBIT 17.12 Standardized Distribution Curve



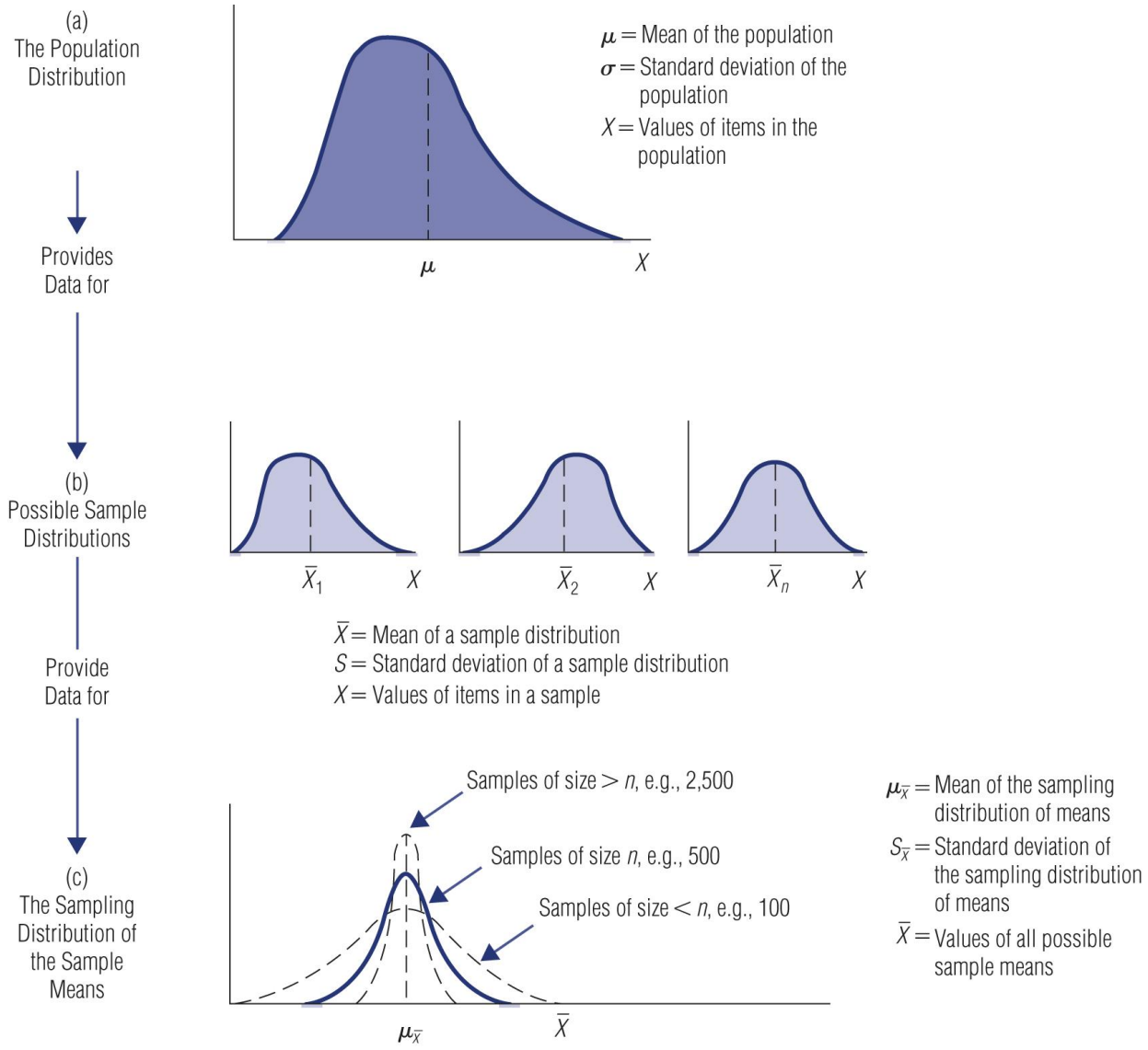
Population Distribution, Sample Distribution, and Sampling Distribution

- Population Distribution
  - *A frequency distribution of the elements of a population.*
- Sample Distribution
  - *A frequency distribution of a sample.*
- Sampling Distribution
  - *A theoretical probability distribution of sample means for all possible samples of a certain size drawn from a particular population.*
- Standard Error of the Mean
  - *The standard deviation of the sampling distribution.*

$$S_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

©2013 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

EXHIBIT 17.13 Fundamental Types of Distributions



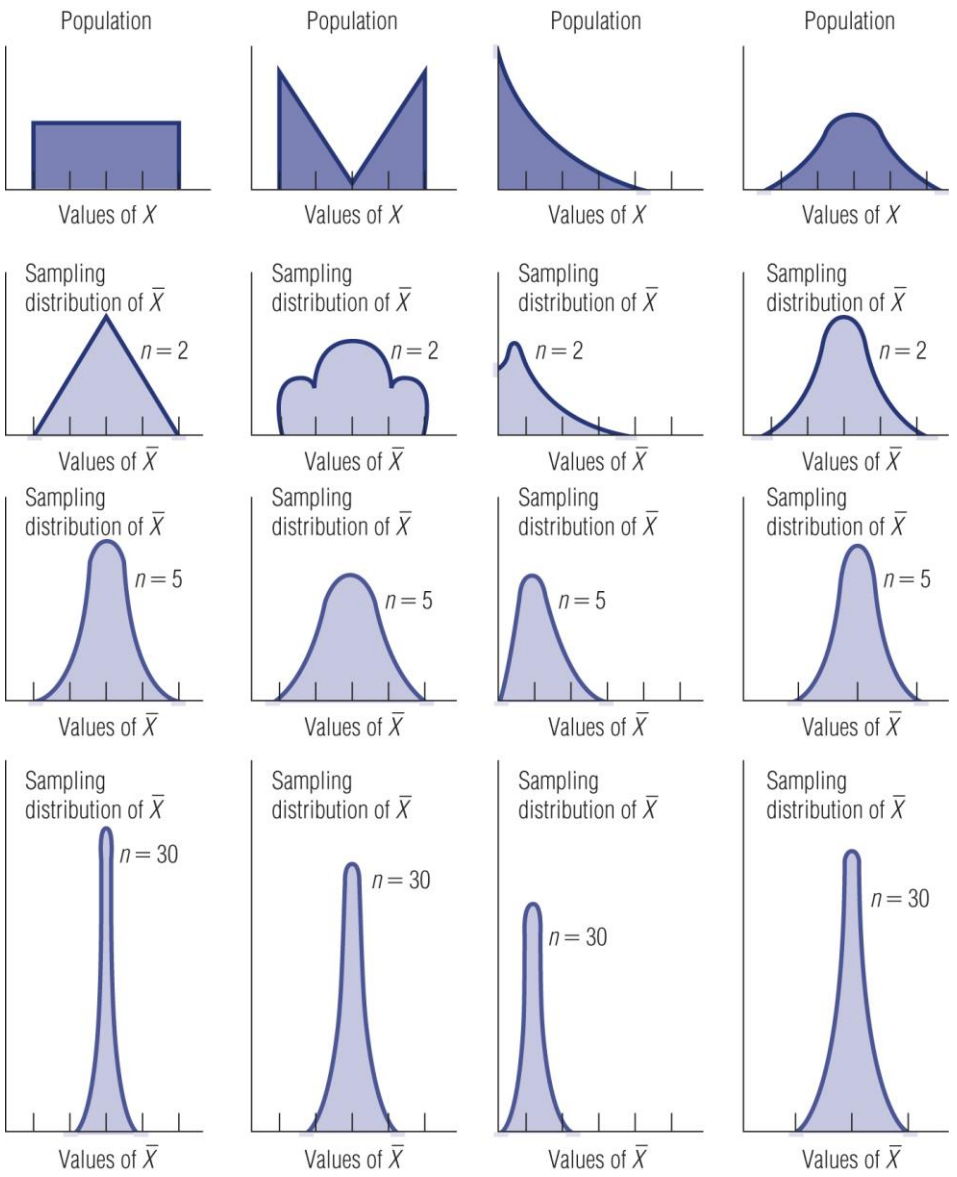
Central-limit Theorem

- Central-limit Theorem
  - *The theory that, as sample size increases, the distribution of sample means of size n, randomly selected, approaches a normal distribution.*

©2013 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

EXHIBIT 17.14

Distribution of Sample Means for Samples of Various Sizes and Population Distributions



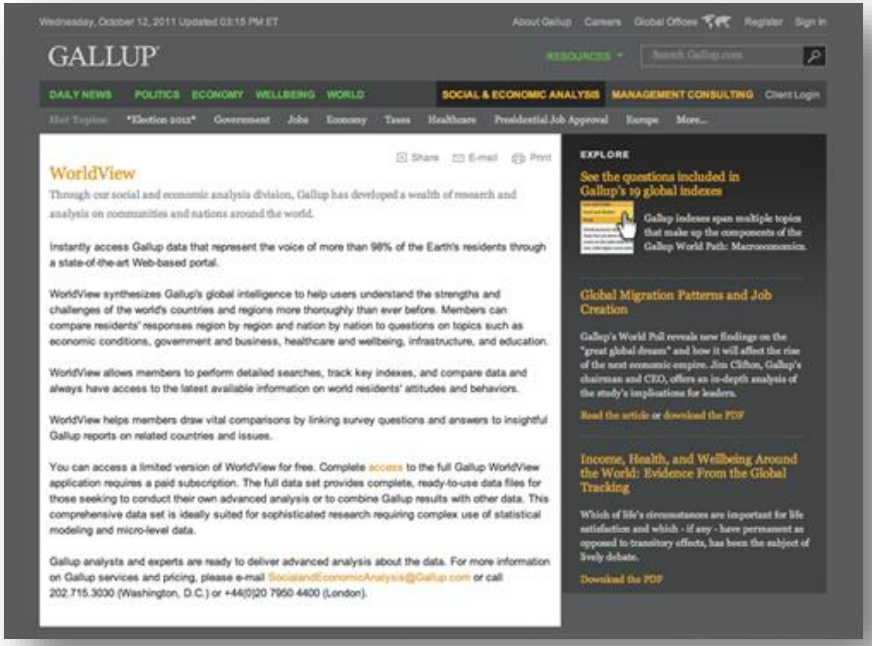
Estimation of Parameters and Confidence Intervals

- Point Estimates
  - *An estimate of the population mean in the form of a single value, usually the sample mean.*
    - Gives no information about the possible magnitude of random sampling error.
- Confidence Interval Estimate
  - *A specified range of numbers within which a population mean is expected to lie.*
  - *An estimate of the population mean based on the knowledge that it will be equal to the sample mean plus or minus a small sampling error.*

©2013 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Sampling the World

- Gallup's WorldView generates a comprehensive snapshot of the opinions of people from across the globe.
- Collects data from over 150 countries, generally through samples of 1,000 individuals in each.
- Provides valuable information on poorer and more rural populations.



©2013 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Confidence Intervals

- Confidence Level
  - *A percentage or decimal value that tells how confident a researcher can be about being correct.*
  - *It states the long-run percentage of confidence intervals that will include the true population mean.*
  - *The crux of the problem for a researcher is to determine how much random sampling error to tolerate.*
  - *Traditionally, researchers have used the 95% confidence level (a 5% tolerance for error).*

©2013 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Calculating a Confidence Interval

Approximate location (value) of the population mean

$$\mu = \overline{X} \pm \text{a small sampling error}$$

Estimation of the sampling error

$$\text{Small sampling error} = Z_{c.l.} S_{\overline{X}}$$

$$\overline{X} = \text{sample mean}$$

$$\mu = \overline{X} \pm Z_{c.l.} S_{\overline{X}}$$

©2013 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Calculating a Confidence Interval (cont'd)

1. Calculate $\overline{X}$ from the sample.
2. Assuming σ is unknown, estimate the population standard deviation by finding S, the sample standard deviation.
3. Estimate the standard error of the mean, using the following formula:

$$S_{\overline{X}} = \frac{S}{\sqrt{n}}$$

4. Determine the Z-value associated with the desired confidence level. The confidence level should be divided by 2 to determine what percentage of the area under the curve to include on each side of the mean.
5. Calculate the confidence interval.

©2013 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Target and Walmart Shoppers Really Are Different

- 40% of respondents named both Target and Walmart as places where they shop.
- But 30% shopped at Walmart and not Target and 12% shopped at Target but not Walmart.
- Target shoppers who shun Walmart shop at more upscale stores, while Walmart shoppers who shun Target shop at discounters.

©2013 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

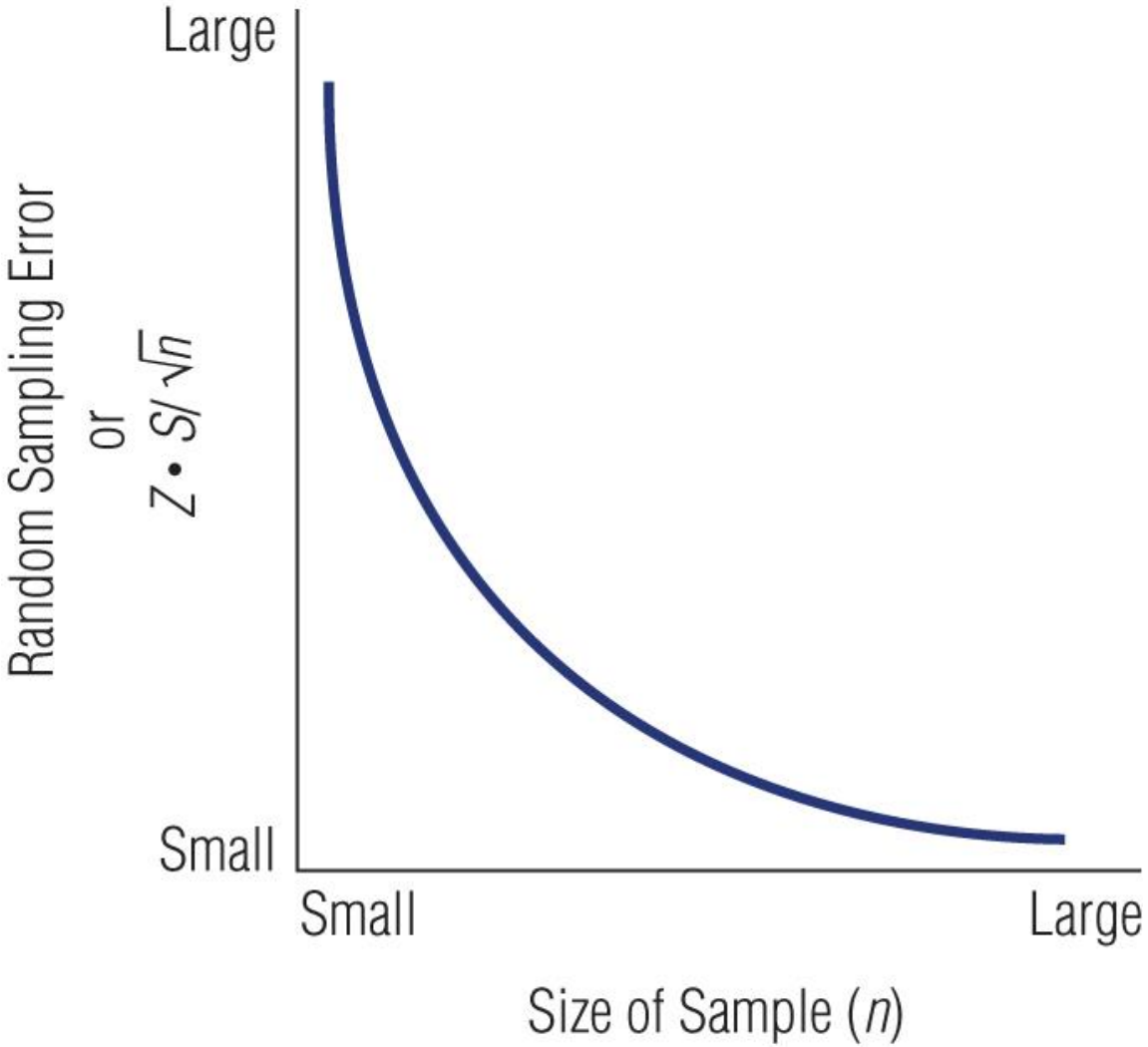
Sample Size

- Random Error and Sample Size
  - *Random sampling error varies with samples of different sizes.*
  - *Increases in sample size reduce sampling error at a decreasing rate.*
    - Diminishing returns - random sampling error is inversely proportional to the square root of *n*.

©2013 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

EXHIBIT 17.18

Relationship between Sample Size and Error



Factors of Concern in Choosing Sample Size

- Variance (or Heterogeneity)
  - *A heterogeneous population has more variance (a larger standard deviation) which will require a larger sample.*
  - *A homogeneous population has less variance (a smaller standard deviation) which permits a smaller sample.*
- Magnitude of Error (Confidence Interval)
  - *How precise must the estimate be?*
- Confidence Level
  - *How much error will be tolerated?*

©2013 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Estimating Sample Size for Questions Involving Means

- Sequential Sampling
  - *Conducting a pilot study to estimate the population parameters so that another, larger sample of the appropriate sample size may be drawn.*
- Estimating sample size:

$$n = \left(\frac{zs}{E}\right)^2$$

where

Z = standardized value that corresponds to the confidence level

S = sample standard deviation or estimate of the population standard deviation

E = acceptable magnitude of error, plus or minus error factor (range is one-half of the total confidence interval)[7]

©2013 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Sample Size Example

- Suppose a survey researcher, studying expenditures on lipstick, wishes to have a 95 percent confident level (Z) and a range of error (E) of less than \$2.00. The estimate of the standard deviation is \$29.00. What is the calculated sample size?

$$n = \left(\frac{zs}{E}\right)^2 = \left[\frac{(1.96)(29.00)}{2.00}\right]^2$$

$$= \left[\frac{56.84}{2.00}\right]^2 = (28.42)^2 = 808$$

©2013 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Sample Size Example

- Suppose, in the same example as the one before, the range of error (E) is acceptable at \$4.00. Sample size is reduced.

$$n = \left(\frac{zs}{E}\right)^2 = \left[\frac{(1.96)(29.00)}{4.00}\right]^2$$

$$= \left[\frac{56.84}{4.00}\right]^2 = (14.21)^2 = 202$$

©2013 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Calculating Sample Size at the 99 Percent Confidence Level

$$n = \left[\frac{(2.57)(29)}{2}\right]^2$$

$$= \left[\frac{74.53}{2}\right]^2$$

$$= [37.265]^2$$

$$= 1389$$

$$n = \left[\frac{(2.57)(29)}{4}\right]^2$$

$$= \left[\frac{74.53}{4}\right]^2$$

$$= [18.6325]^2$$

$$= 347$$

©2013 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Determining Sample Size for Proportions

$$\text{Confidence interval} = p \pm Z_{c.l.} S_p$$

$$S_p = \sqrt{\frac{pq}{n}} \text{ or } S_p = \sqrt{\frac{p(1-n)}{n}}$$

where

S_p = estimate of the standard error of the proportion

p = proportion of successes

$q = 1 - p$, or proportion of failures

©2013 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Determining Sample Size for Proportions (cont'd)

$$n = \frac{Z^2_{c.l.}pq}{E^2}$$

where

n = number of items in sample

$Z^2_{c.l.}$ = square of the confidence level in standard error units

p = estimated proportion of successes

q = $1 - p$, or estimated proportion of failures

E^2 = square of the maximum allowance for error between the true proportion and the sample proportion, or $Z_{c.l.}S_p$ squared

©2013 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Calculating Example Sample Size at the 95 Percent Confidence Level

$p = .6$

$q = .4$

$$n = \frac{(1.96)^2(.6)(.4)}{(.035)^2}$$

$$= \frac{(3.8416)(.24)}{001225}$$

$$= \frac{.922}{.001225}$$

$$= 753$$

©2013 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

EXHIBIT 17.6 Low Dispersion versus High Dispersion