

Estimation and Confidence Intervals

Chapter 9

Learning Objectives

- LO9-1** Compute and interpret a point estimate of a population mean
- LO9-2** Compute and interpret a confidence interval for a population mean
- LO9-3** Compute and interpret a confidence interval for a population proportion
- LO9-4** Calculate the required sample size to estimate a population proportion or population mean
- LO9-5** Adjust a confidence interval for finite populations

Point Estimate

- ▶ A point estimate is a single value (statistic) used to estimate a population value (parameter)

POINT ESTIMATE The statistic, computed from sample information, that estimates a population parameter.

- ▶ Example
- ▶ Suppose the Bureau of Tourism for Barbados wants an estimate of the mean amount spent by tourists visiting that country. They randomly select 500 tourists as they depart and ask these tourists about their spending while there. The mean amount spent by the sample of 500 tourists serves as an estimate of the unknown population parameter.

Confidence Intervals

- ▶ A confidence interval is a range of values within which the population parameter is expected to occur

CONFIDENCE INTERVAL A range of values constructed from sample data so that the population parameter is likely to occur within that range at a specified probability. The specified probability is called the level of confidence.

- ▶ The factors that determine the width of a confidence interval for a mean are
 - ▶ The number of observations in the sample, n
 - ▶ The variability in the population, usually estimated by the sample standard deviation, s
 - ▶ The desired level of confidence

Level of Confidence, σ Known

- ▶ To determine the confidence limits when the population standard deviation is known, we use the z distribution
- ▶ The formula is

**CONFIDENCE INTERVAL FOR A POPULATION
MEAN WITH σ KNOWN**

$$\bar{x} \pm z \frac{\sigma}{\sqrt{n}}$$

[9-1]

$\bar{x}$ – sample mean

z – z - value for a particular confidence level

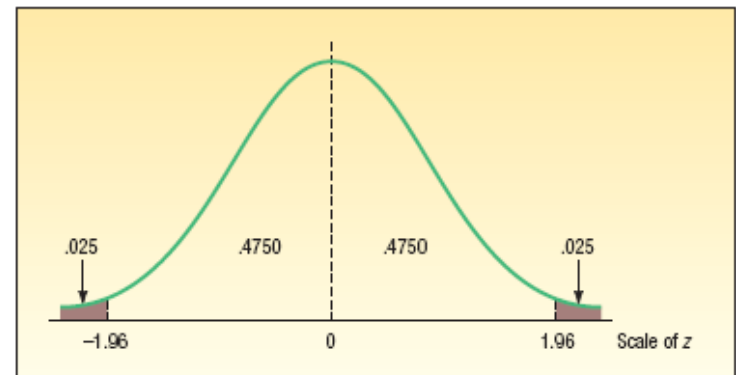
σ – the population standard deviation

n – the number of observations in the sample

Finding a Value of z

- ▶ The method for finding z for a 95% confidence interval is
 - ▶ Divide the confidence interval in half, $.9500/2 = .4750$
 - ▶ Find the value .4750 in the body of the table
 - ▶ Identify the row and column and add the values
 - ▶ The probability of finding a value between 0 and 1.96 is .4750
 - ▶ So the probability of finding a value between ± 1.96 is .9500

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936



Level of Confidence, z Example (1 / 2)

The American Management Association is studying the income of store managers in the retail industry. A random sample of 49 managers reveals a sample mean of \$45,420. The standard deviation of the population is \$2,050.

1. What is the population mean?
2. What is a reasonable range of values for the population mean?
3. How do we interpret these results?

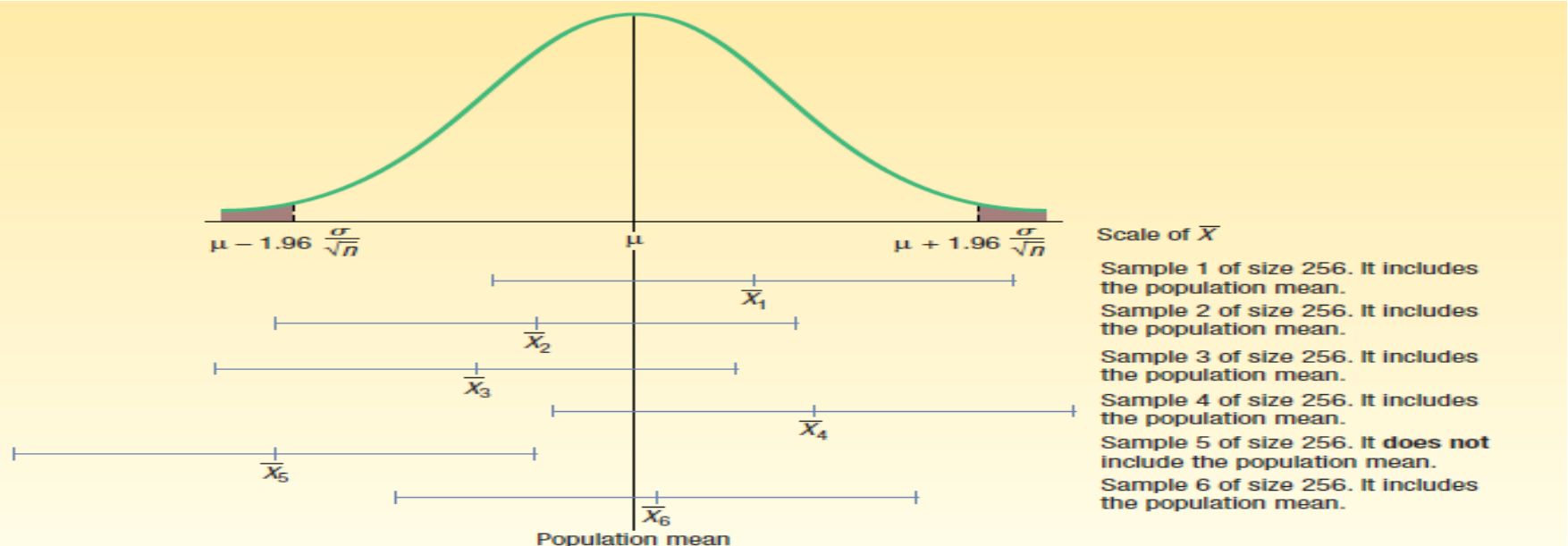
1. We do not know the population mean, so we can use the sample mean, \$45,420 as our best estimate.
2. The AMA decides to use a 95% level of confidence, so use equation (9-1),

$$\bar{x} \pm z \frac{\sigma}{\sqrt{n}} = \$45,420 \pm 1.96 \frac{\$2,050}{\sqrt{49}} = \$45,420 \pm \$574$$

3. The confidence interval is from \$44,846 and \$45,994, the value \$574 is called the margin of error.

Z example (2/2)

How do we interpret these results? Suppose we select many samples of 256 store managers, perhaps several hundred. For each sample, we compute the mean and then construct a 95 percent confidence interval, such as we did in the previous section. We could expect about 95 percent of these confidence intervals to contain the *population mean*. About 5 percent of the intervals would not contain the population mean annual income, which is . Note that not all intervals include the population mean. Both the endpoints of the fifth sample are less than the population mean. We attribute this to sampling error, and it is the risk we assume when we select the level of confidence.



Exercise

7. Bob Nale is the owner of Nale's Quick Fill. Bob would like to estimate the mean number of gallons of gasoline sold to his customers. Assume the number of gallons sold follows the normal distribution with a population standard deviation of 2.30 gallons. From his records, he selects a random sample of 60 sales and finds the mean number of gallons sold is 8.60.

- What is the point estimate of the population mean?
- Develop a 99 percent confidence interval for the population mean.
- Interpret the meaning of part (b).

$$\bar{x} \pm z \frac{\sigma}{\sqrt{n}}$$

b. 7.83 and 9.37, found by
 $8.60 \pm 2.58(2.30 / \sqrt{60})$

If 100 such intervals were determined, the population mean would be included in about 99 intervals

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936

Level of Confidence, σ Unknown

- ▶ To determine the confidence limits when the population standard deviation is unknown, we use the t distribution
- ▶ The formula is

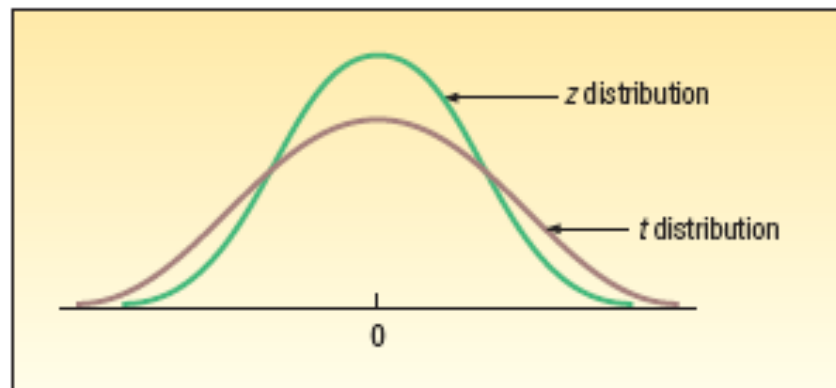
**CONFIDENCE INTERVAL FOR THE
POPULATION MEAN, σ UNKNOWN**

$$\bar{x} \pm t \frac{s}{\sqrt{n}} \quad (9-2)$$

- ▶ Example
- ▶ The Dean of the Business College wants to estimate the mean number of hours full-time students work at paying jobs each week. He randomly selects a sample of 30 students and asks them how many hours they worked last week. He can calculate the sample mean, but it is unlikely he would know the population standard deviation required for formula 9-1.

Characteristics of the t Distribution

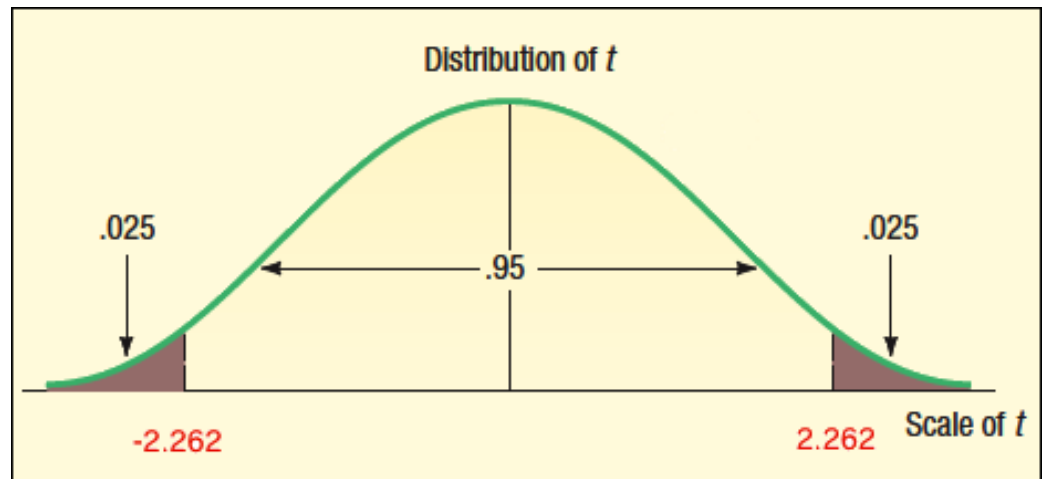
- ▶ The t distribution is a continuous distribution
- ▶ It is mound-shaped and symmetrical
- ▶ It is flatter, or more spread out, than the standard normal distribution
- ▶ There is a family of t distributions, depending on the number of degrees of freedom



Finding a Value of t

- ▶ First assume the population is normal
- ▶ Using Appendix B.5, move across the columns identified for confidence intervals
- ▶ In the next example, we want to use the 95% level of confidence, so move to that column
- ▶ Then find df , the degrees of freedom, $n-1$

Confidence Intervals					
	80%	90%	95%	98%	99%
Level of Significance for One-Tailed Test					
df	0.10	0.05	0.025	0.010	0.005
Level of Significance for Two-Tailed Test					
	0.20	0.10	0.05	0.02	0.01
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169



Level of Confidence, t Example (1 / 2)

A tire manufacturer wishes to investigate the tread life of its tires. A sample of 10 tires driven 50,000 miles revealed a sample mean of 0.32 inch of tread remaining with a standard deviation 0.09 inch. Construct a 95% confidence interval for the population mean.

Would it be reasonable for the manufacturer to conclude that after 50,000 miles the population mean amount of tread remaining is 0.30 inch?

$$\bar{x} \pm t \frac{s}{\sqrt{n}} = 0.32 \pm 2.262 \frac{0.09}{\sqrt{10}} = 0.32 \pm 0.064$$

The endpoints of the confidence interval are 0.256 and 0.384. The margin of error is 0.064. The manufacturer can be reasonably sure (95% confident) that the mean remaining tread depth is between 0.256 and 0.384 inch. Because the value 0.30 is in this interval, it is possible that the mean of the population is 0.30.

Level of Confidence, *t* Example (2/2)

df: # of degree of freedom =
 $n(\text{\# of observations in the sample}) - (\text{\# of samples})$

$$df = n - 1$$

If we repeated this study 200 times, calculating the 95 percent confidence interval with each sample's mean and the standard deviation, 190 of the intervals would include the population mean. Ten of the intervals would not include the population mean. This is the effect of sampling error. A further interpretation is to conclude that the population mean is in this interval.

df	Confidence Intervals				
	80%	90%	95%	98%	99%
	Level of Significance for One-Tailed Test				
	0.10	0.05	0.025	0.010	0.005
	Level of Significance for Two-Tailed Test				
	0.20	0.10	0.05	0.02	0.01
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169

Exercise

13. Merrill Lynch Securities and Health Care Retirement Inc. are two large employers in downtown Toledo, Ohio. They are considering jointly offering child care for their employees. As a part of the feasibility study, they wish to estimate the mean weekly child-care cost of their employees. A sample of 10 employees who use child care reveals the following amounts spent last week.

\$107	\$92	\$97	\$95	\$105	\$101	\$91	\$99	\$95	\$104
-------	------	------	------	-------	-------	------	------	------	-------

Develop a 90 percent confidence interval for the population mean. Interpret the result.

Between 95.39 and 101.81, found by

$$98.6 \pm 1.833 \left(\frac{5.54}{\sqrt{10}} \right)$$

df	Confidence Intervals				
	80%	90%	95%	98%	99%
	Level of Significance for One-Tailed Test				
	0.10	0.05	0.025	0.010	0.005
Level of Significance for Two-Tailed Test					
	0.20	0.10	0.05	0.02	0.01
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169

Confidence Intervals for Proportions

PROPORTION The fraction, ratio, or percent indicating the part of the sample or the population having a particular trait of interest.

- ▶ A sample proportion, p , is found by x , the number of successes, divided by n , the number of observations

SAMPLE PROPORTION

$$p = \frac{x}{n}$$

[9-3]

- ▶ Examples
- ▶ Southern Tech career services reports that 80% of its graduates enter the job market in a position related to their field of study
- ▶ A recent study of married men between the ages 35 and 50 found that 63% felt that both partners should earn a living

Confidence Intervals for Proportions

- ▶ A population proportion is identified by π
- ▶ Two requirements
 - ▶ The binomial conditions have been met
 - ▶ The values $n\pi$ and $n(1-\pi)$ should both be greater than or equal to 5
- ▶ We construct a confidence interval for a population proportion with the following formula

**CONFIDENCE INTERVAL FOR A
POPULATION PROPORTION**

$$p \pm z \sqrt{\frac{p(1-p)}{n}}$$

[9-4]

Confidence Interval, π Example

The union representing the Bottle Blowers of America (BBA) is considering a proposal to merge with the Teamsters Union. At least three-fourths of the BBA membership must approve any merger. A random sample of 2,000 current members reveals 1,600 plan to vote for the merger proposal. What is the estimate of the population proportion? Can you conclude that the necessary proportion of BBA members favor the merger? Why?

To interpret; if the poll was conducted 100 times with 100 different samples, we expect the confidence intervals constructed from 95 of the samples to contain the true population proportion

First, calculate the sample proportion, $p = \frac{X}{n} = \frac{1,600}{2,000} = .80$

Next, use formula (9-4) to determine the 95% confidence interval,

$$p \pm z \sqrt{\frac{p(1-p)}{n}} = .80 \pm 1.96 \sqrt{\frac{.80(1-.80)}{2,000}} = .80 \pm .018$$

The endpoints of the confidence interval are .782 and .818, so we conclude the merger will likely pass because the interval estimate includes values greater than 75% of the union membership.

Exercise

17. The Fox TV network is considering replacing one of its prime-time crime investigation shows with a new family-oriented comedy show. Before a final decision is made, network executives commission a sample of 400 viewers. After viewing the comedy, 250 indicated they would watch the new show and suggested it replace the crime investigation show.

- Estimate the value of the population proportion.
- Develop a 99 percent confidence interval for the population proportion.
- Interpret your findings.

a. 0.625, found by 250/400.

$$p = \frac{x}{n}$$

$$p \pm z \sqrt{\frac{p(1-p)}{n}}$$

b. Between 0.563 and 0.687, found by $0.625 \pm 2.58 \left(\sqrt{\frac{0.625(1-0.625)}{400}} \right)$

c. We are reasonably sure the population proportion is between 56 and 69 percent

Determining Sample Size for Means

- ▶ There are three factors that determine the sample size when we wish to estimate the mean
- ▶ The margin of error, E
- ▶ The desired level of confidence, for example 95%
- ▶ The variation or dispersion of the population being studied
- ▶ The formula to determine the sample size for the mean is

**SAMPLE SIZE FOR ESTIMATING
THE POPULATION MEAN**

$$n = \left(\frac{zs}{E} \right)^2 \quad (9-5)$$

- ▶ n: size of the sample – sigma σ : population standard deviation
- ▶ The result is not always a whole number; the usual practice is to round up any fractional result to the next whole number.

Sample Size to Estimate a Population Mean

Example

A student in public administration wants to estimate the mean monthly earnings of city council members in large cities. She can tolerate a margin of error of \$100 in estimating the mean. She would also prefer to report the interval estimate with a 95% level of confidence. The student found a report by the Department of Labor that reported a standard deviation of \$1,000. What is the required sample size?

$$n = \left(\frac{z\sigma}{E} \right)^2 = \left(\frac{(1.96)(\$1,000)}{\$100} \right)^2 = (19.6)^2 = 384.16$$

The computed value of 384.16 is rounded up to 385. A sample size of 385 is required to meet the specifications.

Determining Sample Size for Proportions

- ▶ There are three factors that determine the sample size when we wish to estimate a proportion
- ▶ The margin of error, E
- ▶ The desired level of confidence
- ▶ A value for π to calculate the variation in the population
- ▶ The formula to determine the sample size for a proportion is

**SAMPLE SIZE FOR THE
POPULATION PROPORTION**

$$n = p(1 - p) \left(\frac{z}{E} \right)^2 \quad (9-6)$$

Sample Size for the Population Proportion Example

The student in the previous example also wants to estimate the proportion of cities that have private refuse collectors. The student wants to estimate the population proportion with a margin of error of .10, prefers a level of confidence of 90%, and has no estimate for the population proportion. What is the required sample size?

$$n = (.50)(1-.50)\left(\frac{1.645}{.10}\right)^2 = 67.65$$

The student needs a random sample of 68 cities.

Exercise

25. Suppose the U.S. president wants an estimate of the proportion of the population who support his current policy toward revisions in the health care system. The president wants the estimate to be within .04 of the true proportion. Assume a 95 percent level of confidence. The president's political advisors estimated the proportion supporting the current policy to be .60.

a. How large of a sample is required?

b. How large of a sample would be necessary if no estimate were available for the proportion supporting current policy?

a. 577, found by $n = 0.60(0.40)\left(\frac{1.96}{0.04}\right)^2 = 576.24$

b. 601, found by $n = 0.50(0.50)\left(\frac{1.96}{0.04}\right)^2 = 600.25$

Finite Population Adjustment

- ▶ A population that has a fixed upper bound is finite
- ▶ For a finite population, the standard error is adjusted by the Finite-Population Correction (FPC) factor, $\sqrt{\frac{N-n}{N-1}}$
- ▶ This will make your estimate more precise by reducing the standard error and resulting in a smaller range of values in estimating the population mean

Multiplying this correction factor by the standard error *reduces* the standard error by about 5 percent (1 - .9492 = .0508). This reduction in the size of the standard error yields a smaller range of values in estimating the population mean or the population proportion. If the sample is 200, the correction factor is .8949, meaning that the standard error has been reduced by more than 10 percent. This table shows the effects of various sample sizes.

Sample Size	Fraction of Population	Correction Factor
10	.010	.9955
25	.025	.9879
50	.050	.9752
100	.100	.9492
200	.200	.8949
500	.500	.7075

Finite Population Adjustment Example

There are 250 families residing in Scandia, Pennsylvania. A random sample of 40 of these families revealed the mean annual church contribution was \$450 and the standard deviation of this was \$75.

1. What is the population mean? What is the best estimate of the population mean?
2. Develop a 90% confidence interval for the population mean.
3. Using the confidence interval, explain why the population mean could be \$445. Could the population mean be \$425? Why?

1. We do not know the population mean. The best estimate is \$450.
2. The formula is $df = 39$

$$\bar{x} \pm t \frac{s}{\sqrt{n}} \left(\sqrt{\frac{N-n}{N-1}} \right) = \$450 \pm 1.685 \frac{\$75}{\sqrt{40}} \left(\sqrt{\frac{250-40}{250-1}} \right) = \$450 \pm \$18.35$$

3. The endpoints are \$431.65 and \$468.35, so the population mean could be \$445. It is not likely the population mean is \$425 since \$425 is not within the confidence interval.

Exercise

28. Forty-nine items are randomly selected from a population of 500 items. The sample mean is 40 and the sample standard deviation 9. Develop a 99 percent confidence interval for the population mean.

$$36.722 \text{ and } 43.278, \text{ found by } 40 \pm 2.682 \left(\frac{9}{\sqrt{49}} \right) \sqrt{\frac{500 - 49}{500 - 1}}$$