Estimation and Confidence Intervals

Chapter 9

Learning Objectives

LO1 Define a point estimate.

LO2 Define level of confidence.

LO3 Compute a confidence interval for the population mean when the population standard deviation is known.

LO4 Compute a confidence interval for a population mean when the population standard deviation is unknown.

LO5 Compute a confidence interval for a population proportion.

LO6 Calculate the required sample size to estimate a population proportion or population mean.

LO7 Adjust a confidence interval for finite populations

Sampling and Estimates

Why Use Sampling?
1. To contact the entire population is too time consuming.
2. The cost of studying all the items in the population is often too expensive.
3. The sample results are usually adequate.
4. Certain tests are destructive.
5. Checking all the items is physically impossible.

Point Estimate versus Confidence Interval Estimate
- A **point estimate** is a single value (point) derived from a sample and used to estimate a population value.
- A **confidence interval estimate** is 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.

What are the factors that determine the width of a confidence interval?
1. The **sample size**, *n.*
2. The **variability in the population**, usually σ estimated by *s.*
3. The desired **level of confidence**.

Point Estimate and Confidence Intervals for a Population Mean

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

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*.

To compute a confidence interval, we will consider two situations:

- We use sample data to estimate μ with $\bar{X}$ and the population standard deviation (σ) is known
- We use sample data to estimate μ with $\bar{X}$, and the population standard deviation is unknown. In this case, we substitute the sample standard deviation (s) for the population standard deviation (σ).

Interval Estimates - Interpretation

1) Ninety-five percent of the sample means selected from a population will be within 1.96 standard errors of the population mean μ.

2) Ninety-nine percent of the sample means will lie within 2.58 standard errors of the population mean
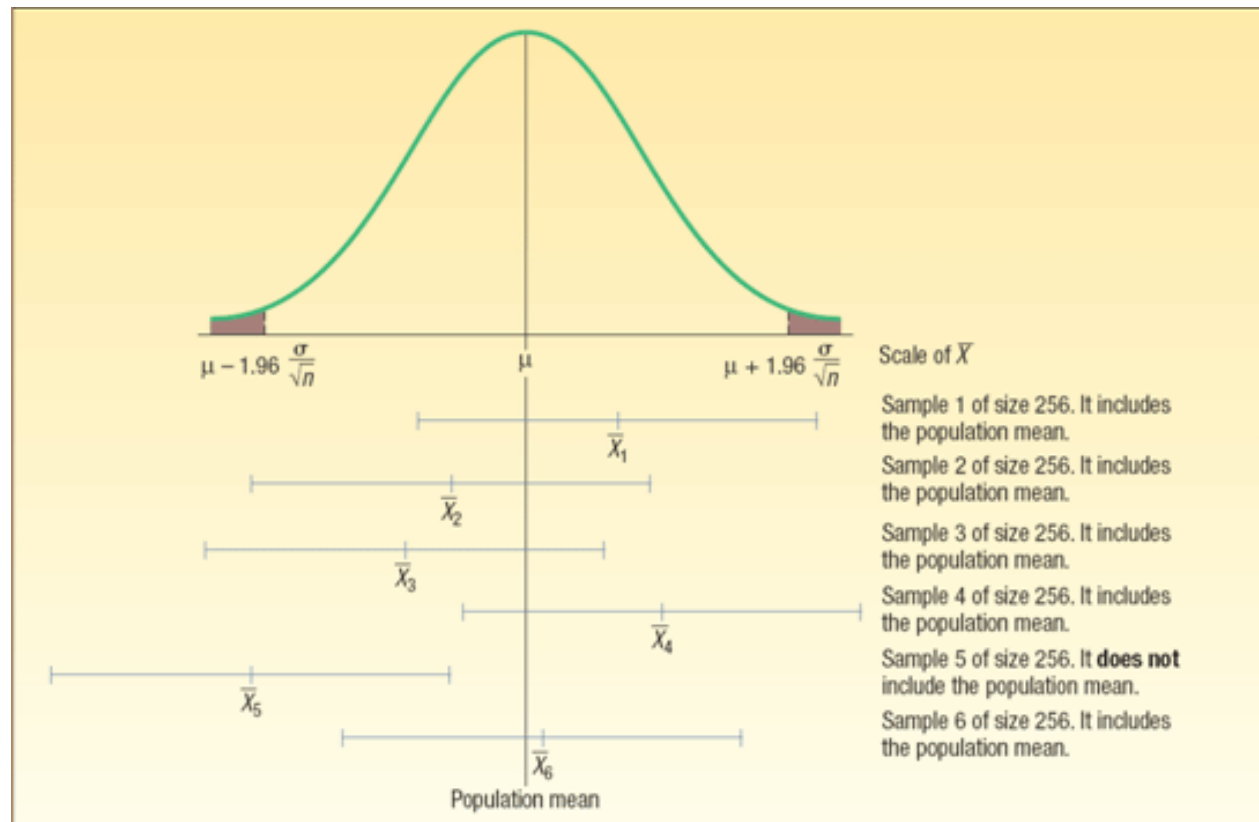
How to Obtain z value for a Given Confidence Level

The *95 percent confidence* refers to the middle 95 percent of the observations. Therefore, the remaining 5 percent are equally divided between the two tails.
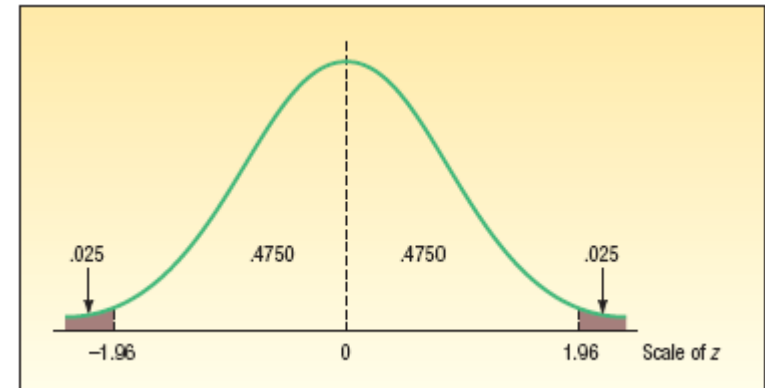
Following is a portion of Appendix B.1.

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

Point Estimates and Confidence Intervals for a Mean – σ Known

1. The width of the interval is determined by the level of confidence and the size of the standard error of the mean.
2. The standard error is affected by two values:
   - Standard deviation
   - Number of observations in the sample

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

where:

$\sigma_{\bar{x}}$ is the symbol for the standard error of the mean. We use a Greek letter because it is a population value, and the subscript $\bar{x}$ reminds us that it refers to a sampling distribution of the sample means.

σ is the population standard deviation.

n is the number of observations in the sample.

CONFIDENCE INTERVAL FOR POPULATION MEAN WITH σ KNOWN	$\bar{X} \pm z \dfrac{\sigma}{\sqrt{n}}$

The following calculations for a 95 percent confidence interval: $\bar{X} \pm 1.96 \dfrac{\sigma}{\sqrt{n}}$

Similarly, a 99 percent confidence interval is computed as follows. $\bar{X} \pm 2.58 \dfrac{\sigma}{\sqrt{n}}$

Example

- Del Monte Foods distributes diced peaches in 4-ounce plastic cups. To be sure each cup contains at least the required amount, Del Monte sets the filling operation to dispense 4.01 ounces of peaches and gel in each cup. So 4.01 is the population mean. Of course not every cup will contain exactly 4.01 ounces of peaches and gel. Some cups will have more and others less. Suppose the population standard deviation of the process is 0.04 ounces. Let's also assume that the process follows the normal probability distribution. Now, we select a random sample of 64 cups and determine the sample mean. It is 4.015 ounces of peaches and gel. The 95 percent confidence interval for the population mean of this particular sample is:

$$4.015 \pm 1.96 \left(\frac{0.04}{\sqrt{64}} \right) = 4.015 \pm 0.0098$$

- The 95 percent confidence interval is between 4.0052 and 4.0248

Point Estimates and Confidence Intervals for a Mean – σ Known

Example

The American Management Association wishes to have information on the mean income of store managers in the retail industry. A random sample of 256 managers reveals a sample mean of $45,420. The standard deviation of this population is $2,050. The association would like answers to the following questions:

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. The sample mean is $45,420 → the sample mean is a point estimate of the unknown population mean

2. The association decides to use the 95 percent level of confidence

$$\bar{X} \pm z \frac{\sigma}{\sqrt{n}} = \$45,420 \pm 1.96 \frac{\$2,050}{\sqrt{256}} = \$45,420 \pm \$251$$

Population Standard Deviation (σ) Unknown – The *t*-Distribution

In most sampling situations the population standard deviation (σ) is not known. Below are some examples where it is unlikely the population standard deviations would be known.

1. The Dean of the Business College wants to estimate the mean number of hours full-time students work at paying jobs each week. He selects a sample of 30 students, contacts each student and asks them how many hours they worked last week.

2. The Dean of Students wants to estimate the distance the typical commuter student travels to class. She selects a sample of 40 commuter students, contacts each, and determines the one-way distance from each student's home to the center of campus.

CHARACTERISTICS OF THE *t*-Distribution

1. It is, like the *z* distribution, a **continuous distribution**.
2. It is, like the *z* distribution**, bell-shaped** and **symmetrical**.
3. There is **not one *t* distribution**, but rather **a family of t distributions**. All *t* distributions have a mean of 0, but their standard deviations differ according to the sample size, *n.*
4. The ***t* distribution is more spread out** and *flatter at the center than the standard normal distribution* As the sample size increases, however, the *t* distribution approaches the standard normal distribution

t −distribution

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

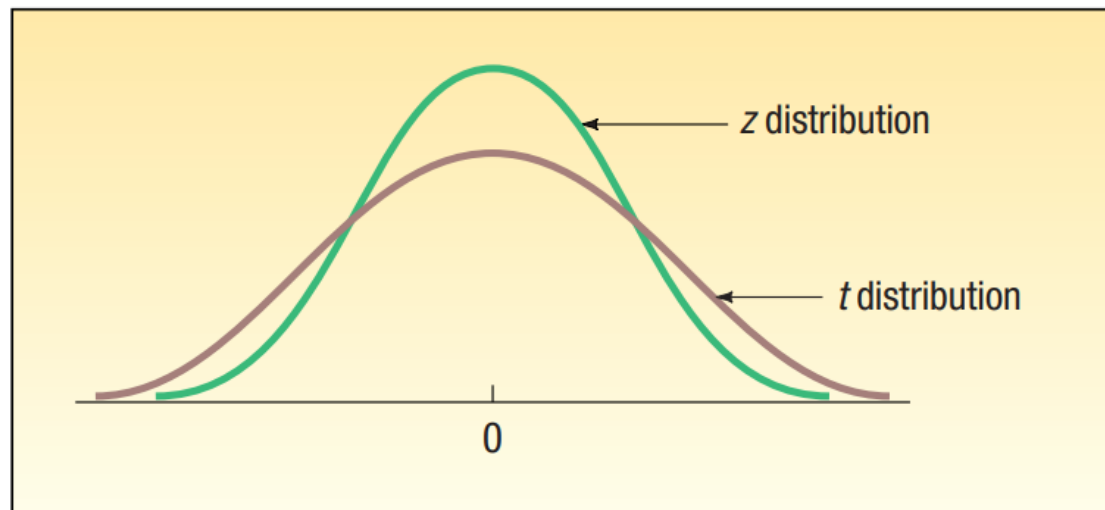■ Where s is an estimate of σ

CHART 9–1 The Standard Normal Distribution and Student's t Distribution

Confidence Interval Estimates for the Mean

<u>Use *Z*-distribution</u>

If the population standard deviation is known or the sample is greater than 30.

<u>Use *t*-distribution</u>

If the population standard deviation is unknown and the sample is less than 30.
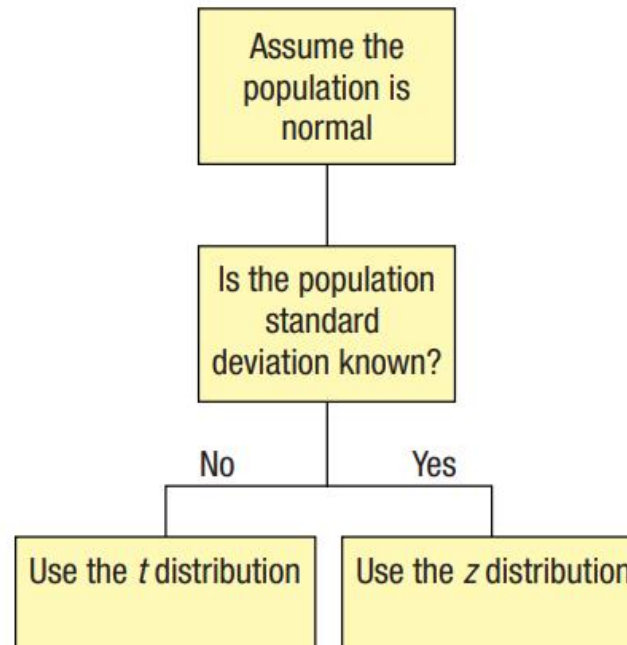
CONFIDENCE INTERVAL FOR POPULATION MEAN WITH σ KNOWN
$$\overline{X} \pm z\frac{\sigma}{\sqrt{n}}$$

CONFIDENCE INTERVAL FOR THE POPULATION MEAN, σ UNKNOWN
$$\overline{X} \pm t\frac{s}{\sqrt{n}}$$

Assume the population is normal

Is the population standard deviation known?

No   Yes

Use the *t* distribution   Use the *z* distribution

Determining When to Use the *z* Distribution or the *t* Distribution

t-distribution
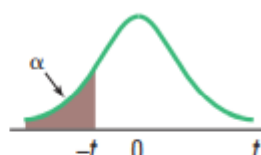
Confidence interval | Left-tailed test | Right-tailed test | Two-tailed test
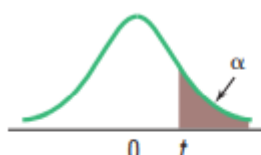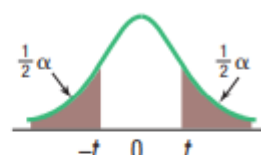
	Confidence Intervals, c							Confidence Intervals, c				
	80%	90%	95%	98%	99%	99.9%		80%	90%	95%	98%	99%	99.9%
	Level of Significance for One-Tailed Test, α							Level of Significance for One-Tailed Test, α					
df	0.10	0.05	0.025	0.01	0.005	0.0005	df	0.10	0.05	0.025	0.01	0.005	0.0005
	Level of Significance for Two-Tailed Test, α							Level of Significance for Two-Tailed Test, α					
	0.20	0.10	0.05	0.02	0.01	0.001		0.20	0.10	0.05	0.02	0.01	0.001
1	3.078	6.314	12.706	31.821	63.657	636.619	36	1.306	1.688	2.028	2.434	2.719	3.582
2	1.886	2.920	4.303	6.965	9.925	31.599	37	1.305	1.687	2.026	2.431	2.715	3.574
3	1.638	2.353	3.182	4.541	5.841	12.924	38	1.304	1.686	2.024	2.429	2.712	3.566
4	1.533	2.132	2.776	3.747	4.604	8.610	39	1.304	1.685	2.023	2.426	2.708	3.558
5	1.476	2.015	2.571	3.365	4.032	6.869	40	1.303	1.684	2.021	2.423	2.704	3.551
6	1.440	1.943	2.447	3.143	3.707	5.959	41	1.303	1.683	2.020	2.421	2.701	3.544
7	1.415	1.895	2.365	2.998	3.499	5.408	42	1.302	1.682	2.018	2.418	2.698	3.538
8	1.397	1.860	2.306	2.896	3.355	5.041	43	1.302	1.681	2.017	2.416	2.695	3.532
9	1.383	1.833	2.262	2.821	3.250	4.781	44	1.301	1.680	2.015	2.414	2.692	3.526
10	1.372	1.812	2.228	2.764	3.169	4.587	45	1.301	1.679	2.014	2.412	2.690	3.520

- For locating t is to move across the column identified for "Confidence Intervals" to the level of confidence requested.
- "df" refers to the number of degree of freedom that is the number of observations in the sample minus the number of samples $(n-1)$.

Confidence Interval for the Mean – Example using the *t*-distribution

EXAMPLE

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 of 0.09 inch.

Construct a 95 percent 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 inches?
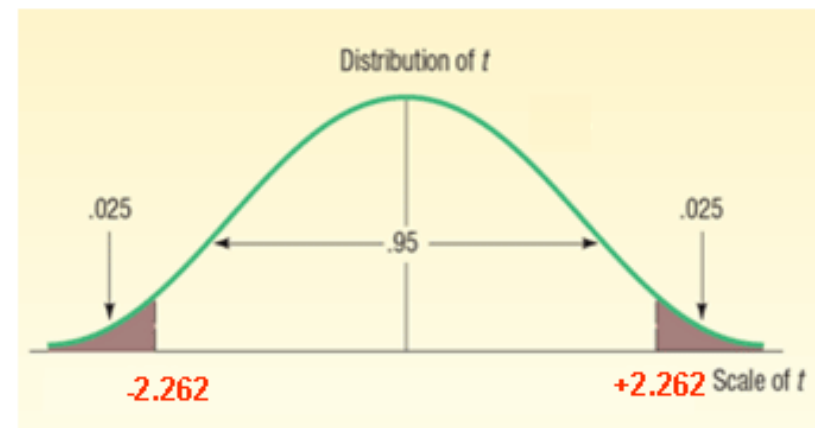
	Confidence Intervals				
	80%	90%	95%	98%	99%
	Level of Significance for One-Tailed Test				
df	0.100	0.050	0.025	0.010	0.005
	Level of Significance for Two-Tailed Test				
	0.20	0.10	0.05	0.02	0.01
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	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

Compute the C.I.

using the t - dist. (since σ is unknown)

$$\overline{X} \pm t_{\alpha/2, n-1} \frac{s}{\sqrt{n}}$$

$$= \overline{X} \pm t_{.05/2, 10-1} \frac{s}{\sqrt{n}}$$

$$= 0.32 \pm t_{.025, 9} \frac{0.09}{\sqrt{10}}$$

$$= 0.32 \pm 2.262 \frac{0.09}{\sqrt{10}}$$

$$= 0.32 \pm 0.064$$

$$= (0.256, 0.384)$$

Conclude : the manufacturer can be reasonably sure (95% confident) that the mean remaining tread depth is between 0.256 and 0.384 inches.

Distribution of *t*

.025   .95   .025

-2.262   +2.262 Scale of *t*

Confidence Interval Estimates for the Mean

EXAMPLE

The manager of the Inlet Square Mall, near Ft. Myers, Florida, wants to estimate the mean amount spent per shopping visit by customers. A sample of 20 customers reveals the following amounts spent.

$48.16	$42.22	$46.82	$51.45	$23.78	$41.86	$54.86
37.92	52.64	48.59	50.82	46.94	61.83	61.69
49.17	61.46	51.35	52.68	58.84	43.88	

Compute the C.I.

using the t - dist. (since σ is unknown)

$$\overline{X} \pm t_{\alpha/2, n-1} \frac{s}{\sqrt{n}}$$

$$= \overline{X} \pm t_{.05/2, 10-1} \frac{s}{\sqrt{n}}$$

$$= 49.35 \pm t_{.025, 19} \frac{9.01}{\sqrt{20}}$$

$$= 49.35 \pm 2.093 \frac{9.01}{\sqrt{20}}$$

$$= 49.35 \pm 4.22$$

The endpoints of the confidence interval are $45.13 and $53.57

Conclude : It is reasonable that the population mean could be $50. The value of $60 is not in the confidence interval. Hence, we conclude that the population mean is unlikely to be $60.

A Confidence Interval for a Proportion (π)

The examples below illustrate the **nominal scale** of measurement.

1. The career services director at Southern Technical Institute reports that 80 percent of its graduates enter the job market in a position related to their field of study.

2. A company representative claims that 45 percent of Burger King sales are made at the drive-through window.

3. A survey of homes in the Chicago area indicated that 85 percent of the new construction had central air conditioning.

4. A recent survey of married men between the ages of 35 and 50 found that 63 percent felt that both partners should earn a living.

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

Using the Normal Distribution to Approximate the Binomial Distribution

To develop a confidence interval for a proportion, we need to meet the following assumptions.

1. The binomial conditions, discussed in Chapter 6, have been met. Briefly, these conditions are:

a. The sample data is the result of counts.

b. There are only two possible outcomes.

c. The probability of a success remains the same from one trial to the next.

d. The trials are independent. This means the outcome on one trial does not affect the outcome on another.

2. The values $n\pi$ and $n(1-\pi)$ should both be greater than or equal to 5. This condition allows us to invoke the central limit theorem and employ the standard normal distribution, that is, z, to complete a confidence interval.

Let p represent the sample proportion, X the number of "successes", and n the number of items sampled.

SAMPLE PROPORTION $$p = \frac{X}{n}$$

CONFIDENCE INTERVAL FOR A POPULATION PROPORTION $$p \pm z\sqrt{\frac{p(1-p)}{n}}$$

Confidence Interval for a Population Proportion - Example

EXAMPLE

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

Develop a 95 percent confidence interval for the population proportion. Basing your decision on this sample information, can you conclude that the necessary proportion of BBA members favor the merger? Why?

First, compute the sample proportion :

$$p = \frac{x}{n} = \frac{1{,}600}{2000} = 0.80$$

Compute the 95% C.I.

$$\text{C.I.} = p \pm z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

$$= 0.80 \pm 1.96\sqrt{\frac{.80(1-.80)}{2{,}000}} = .80 \pm .018$$

$$= (0.782,\ 0.818)$$

Conclude : The merger proposal will likely pass because the interval estimate includes values greater than 75 percent of the union membership.

Selecting an Appropriate Sample Size

There are 3 factors that determine the size of a sample, none of which has any direct relationship to the size of the population.

- **The level of confidence desired,**
- **The margin of error the researcher will tolerate.**
- **The variation in the population being Studied.**

$$n = \left(\frac{z \cdot \sigma}{E} \right)^2$$

where:

   n is the size of the sample.
   z is the standard normal value corresponding to the desired level of confidence.
   σ is the population standard deviation.
   E is the maximum allowable error.

EXAMPLE

A student in public administration wants to determine the mean amount members of city councils in large cities earn per month as remuneration for being a council member. The error in estimating the mean is to be less than $100 with a 95 percent level of confidence. The student found a report by the Department of Labor that estimated the standard deviation to be $1,000. What is the required sample size?

Given in the problem:
- *E,* the maximum allowable error, is $100
- The value of *z* for a 95 percent level of confidence is 1.96,
- The estimate of the standard deviation is $1,000.

$$n = \left(\frac{z \cdot \sigma}{E} \right)^2$$
$$= \left(\frac{(1.96)(\$1{,}000)}{\$100} \right)^2$$
$$= (19.6)^2$$
$$= 384.16$$
$$= 385$$

Sample Size for Estimating a Population Proportion

$$n = \pi(1 - \pi)\left(\frac{z}{E}\right)^2$$

where:

n is the size of the sample
z is the standard normal value
   corresponding to the desired level of confidence
π is the population proportion.
E is the maximum allowable error

NOTE:
use $p = 0.5$ if no initial information on the probability of success is available

EXAMPLE 1

The American Kennel Club wanted to estimate the proportion of children that have a dog as a pet. If the club wanted the estimate to be within 3% of the population proportion, how many children would they need to contact? Assume a 95% level of confidence and that the club estimated that 30% of the children have a dog as a pet.

$$n = (.30)(.70)\left(\frac{1.96}{.03}\right)^2 = 897$$

EXAMPLE 2

A study needs to estimate the proportion of cities that have private refuse collectors. The investigator wants the margin of error to be within .10 of the population proportion, the desired level of confidence is 90 percent, and no estimate is available for the population proportion. What is the required sample size?

$$n = (.5)(1 - .5)\left(\frac{1.65}{.10}\right)^2 = 68.0625$$

$n = 69$ cities

Finite-Population Correction Factor

- A population that has a fixed upper bound is said to be finite.
- For a finite population, where the total number of objects is N and the size of the sample is n, the following adjustment is made to the standard errors of the sample means and the proportion:

Standard Error of the Mean   **Standard Error of the Proportion**

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \qquad \sigma_p = \sqrt{\frac{p(1-p)}{n}}\sqrt{\frac{N-n}{N-1}}$$

- However, if $n/N < .05$, the finite-population correction factor may be ignored. Why? See what happens to the value of the correction factor in the table below when the fraction n/N becomes smaller

Finite-Population Correction Factor for Selected Samples When the Population Is 1,000

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

- The FPC approaches 1 when n/N becomes smaller!

CI for Mean with FPC - Example

EXAMPLE

There are 250 families 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.

Could the population mean be $445 or $425?

What is the population mean? What is the best estimate of the population mean?

Given in Problem:

$N = 250$

$n = 40$

$s = \$75$

Since $n/N = 40/250 = 0.16$, the finite population correction factor must be used.

The population standard deviation is not known therefore use the *t*-distribution (may use the z-dist since n>30)

$$\overline{X} \pm t\, \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

$$= \$450 \pm t_{.10/2, 40-1} \frac{\$75}{\sqrt{40}} \sqrt{\frac{250-40}{250-1}}$$

$$= \$450 \pm 1.685 \frac{\$75}{\sqrt{40}} \sqrt{\frac{250-40}{250-1}}$$

$$= \$450 \pm \$19.98 \sqrt{.8434}$$

$$= \$450 \pm \$18.35$$

$$= (\$431.65, \ \$468.35)$$

It is likely that the population mean is more than $431.65 but less than $468.35. To put it another way, could the population mean be $445? Yes, but it is not likely that it is $425 because the value $445 is within the confidence interval and $425 is not within the confidence interval.