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

Point Estimates

■ A **point estimate** is a single value (point) derived from a sample and used to estimate a population value.

$$\overline{X} \rightarrow \mu$$

$$s \rightarrow \sigma$$

$$s^2 \rightarrow \sigma^2$$

$$p \rightarrow \pi$$

Confidence Interval Estimates

■ 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.

C.I. = point estimate ± margin of error

Factors Affecting Confidence Interval Estimates

The width of a confidence interval are determined by:

1. The **sample size**, *n.*
2. The **variability in the population**, usually σ estimated by *s.*
3. The desired **level of confidence**.

Confidence Intervals for a Mean – σ Known

CONFIDENCE INTERVAL FOR POPULATION MEAN WITH σ KNOWN

$$\overline{X} \pm z \frac{\sigma}{\sqrt{n}} \qquad \text{[9–1]}$$

$\overline{x}$ – sample mean

z – z - value for a particular confidence level
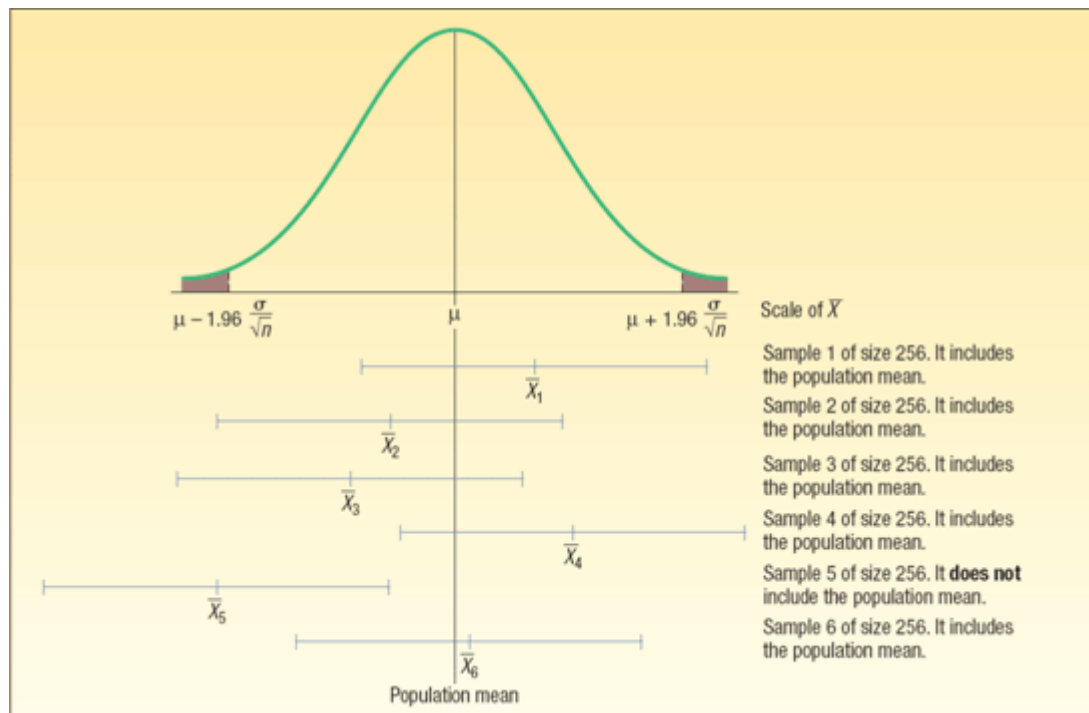
σ – the population standard deviation

n – the number of observations in the sample

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

Interval Estimates - Interpretation

For a 95% confidence interval about 95% of the similarly constructed intervals will contain the parameter being estimated. Also 95% of the sample means for a specified sample size will lie within 1.96 standard deviations of the hypothesized population

$\mu - 1.96 \frac{\sigma}{\sqrt{n}}$ μ $\mu + 1.96 \frac{\sigma}{\sqrt{n}}$ Scale of $\overline{X}$

Sample 1 of size 256. It includes the population mean.

$\overline{X}_1$

Sample 2 of size 256. It includes the population mean.

$\overline{X}_2$

Sample 3 of size 256. It includes the population mean.

$\overline{X}_3$

Sample 4 of size 256. It includes the population mean.

$\overline{X}_4$

Sample 5 of size 256. It **does not** include the population mean.

$\overline{X}_5$

Sample 6 of size 256. It includes the population mean.

$\overline{X}_6$

Population mean

Example: Confidence Interval for a Mean – σ Known

The American Management Association wishes to have information on the mean income of middle 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 with a 95% confidence?

3. What do these results mean?

Example: Confidence Interval for a Mean – σ Known

The American Management Association wishes to have information on the mean income of middle 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:

What is the population mean?

In this case, we do not know. We do know the **sample mean is $45,420. Hence, our best estimate of the unknown population value** is the corresponding **sample statistic**.

The sample mean of $45,420 is a *point estimate* of the unknown 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.
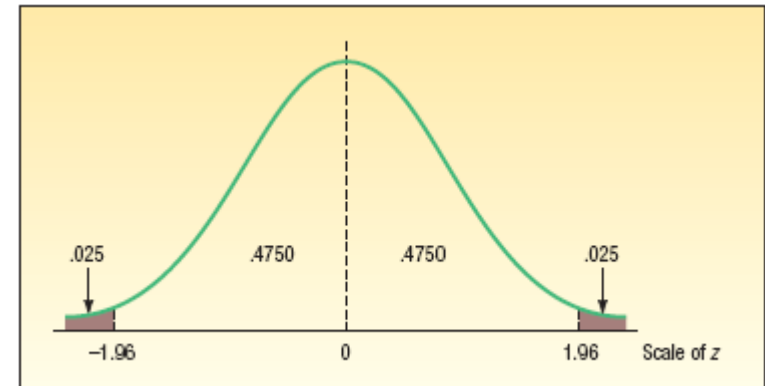
Following is a portion of Appendix B.1.

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

Example: Confidence Interval for a Mean – σ Known

The American Management Association wishes to have information on the mean income of middle 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:

What is a reasonable range of values for the population mean?

Suppose the association decides to use the 95 percent level of confidence:

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

The confidence limit are $45,169 and $45,671
The ±$251 is referred to as the margin of error

Example: Confidence Interval for a Mean - Interpretation

The American Management Association wishes to have information on the mean income of middle 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 confidence limit are $45,169 and $45,671

What is the interpretation of the confidence limits $45,169 and $45,671?

If we select many samples of 256 managers, and for each sample we compute the mean and then construct a 95 percent confidence interval, **we could expect about 95 percent of these confidence intervals to contain the *population* mean.** Conversely, about 5 percent of the intervals would not contain the population mean annual income, μ
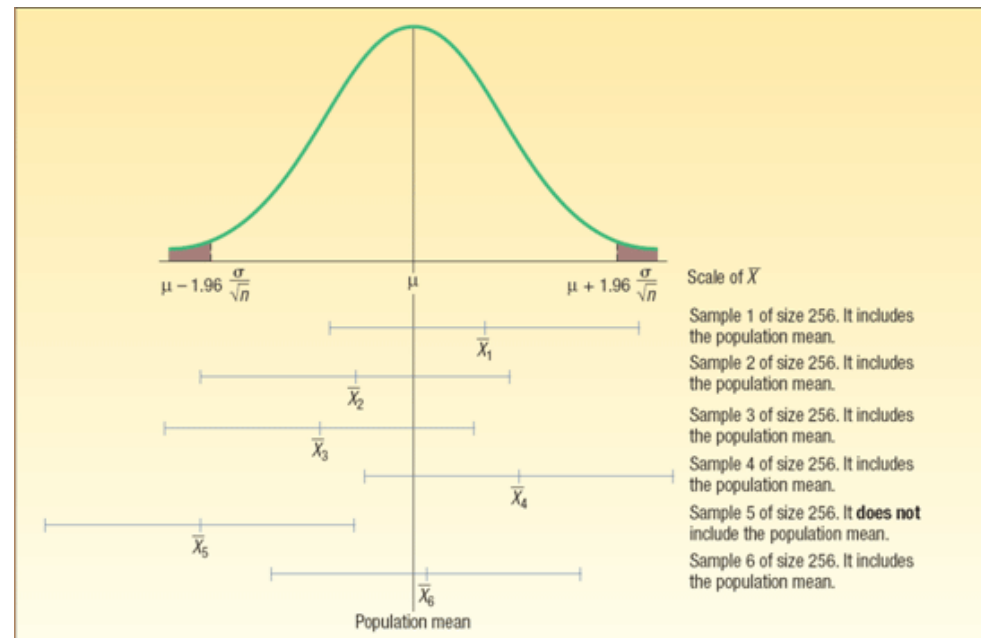
Population Standard Deviation (σ) Unknown

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.

3. The Director of Student Loans wants to know the mean amount owed on student loans at the time of his/her graduation. The director selects a sample of 20 graduating students and contacts each to find the information.

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
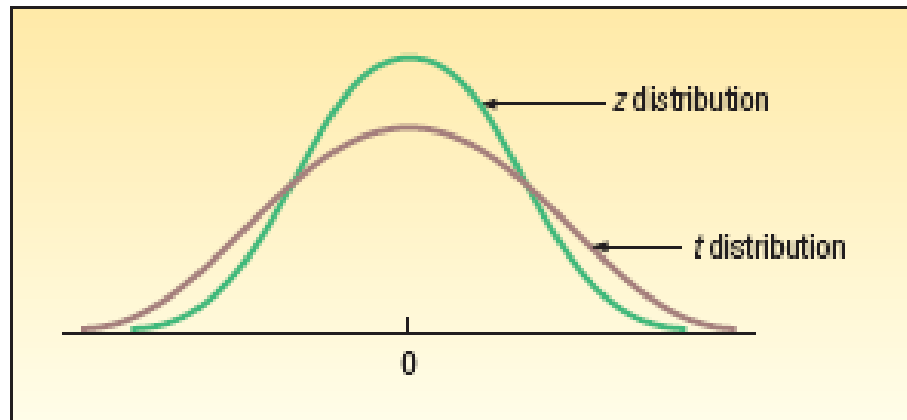
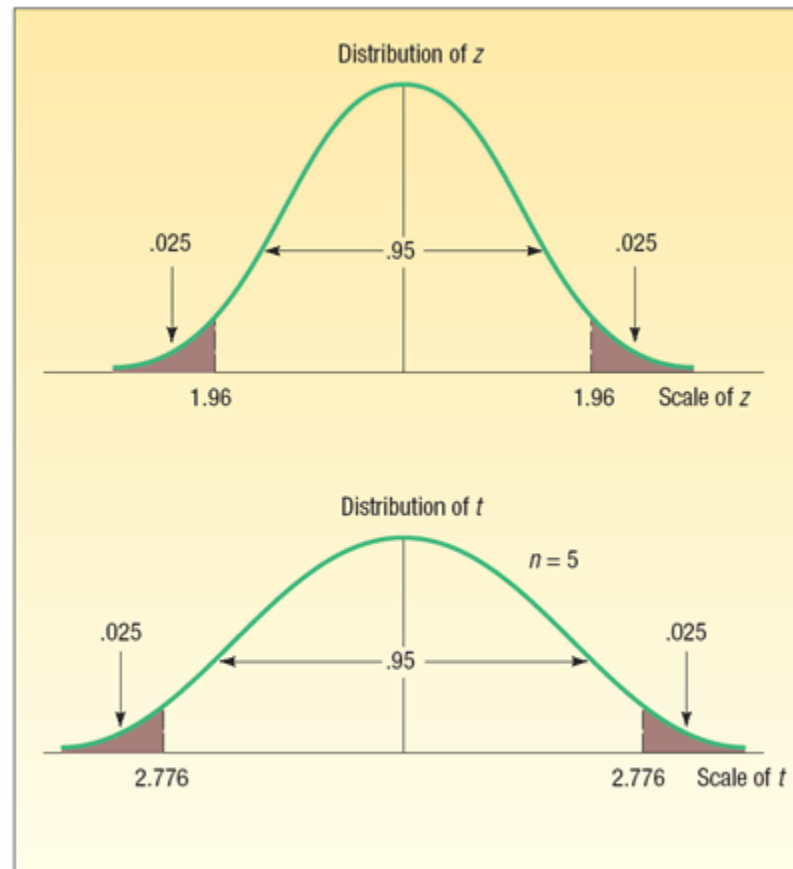# Comparing the z and *t* Distributions when *n* is small, 95% Confidence Level

CHART 9–2 Values of *z* and *t* for the 95 Percent Level of Confidence

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

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?

Given in the problem:

$$n = 10$$

$$\bar{x} = 0.32$$

$$s = 0.09$$

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

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

Student's *t*-distribution Table

Compute the C.I.

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

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

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

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

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

$$= 0.32 \pm 0.064$$

$$= (0.256, 0.384)$$

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

	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

Distribution of *t*

.025

.95

.025

-2.262

+2.262 Scale of *t*
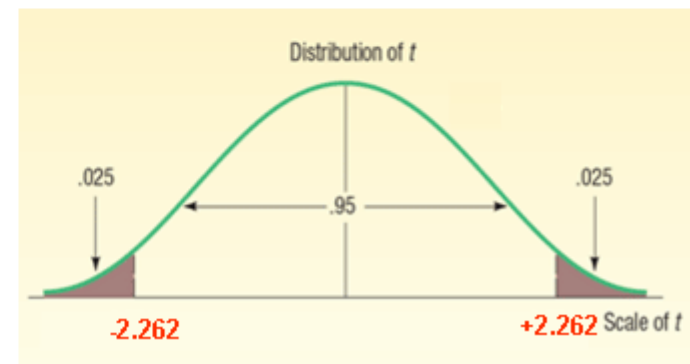
Confidence Interval Estimates for the Mean – Using Minitab

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	

Confidence Interval Estimates for the Mean – By Formula

Compute the C.I.

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

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

$$= \overline{X} \pm t_{.05/2, 20-1} \frac{s}{\sqrt{n}}$$

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

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

$$= 49.35 \pm 4.22$$

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.

Confidence Interval Estimates for the Mean – Using Minitab

Confidence Interval Estimates for the Mean – Using Excel

Confidence Interval Estimates for the Mean

Use Z-distribution

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

$$\overline{X} \pm z \frac{\sigma}{\sqrt{n}}$$

Use t-distribution

If the population standard deviation is unknown and the sample is less than 30.

$$\overline{X} \pm t \frac{s}{\sqrt{n}}$$

When to Use the *z* or *t* Distribution for Confidence Interval Computation
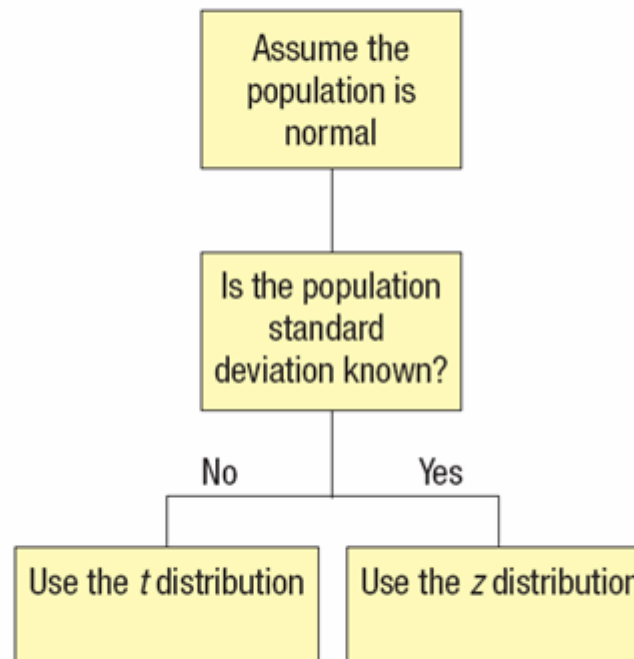
Estimation and Confidence Intervals

CHART 9–3 Determining When to Use the *z* Distribution or the *t* Distribution

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.

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.

Confidence Interval for a Population Proportion - Formula

SAMPLE PROPORTION $p = \dfrac{X}{n}$ [9–3]

CONFIDENCE INTERVAL FOR A
POPULATION PROPORTION $p \pm z \sqrt{\dfrac{p(1-p)}{n}}$ [9–4]

Confidence Interval for a Population Proportion - 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.**

What If Population Standard Deviation is not Known

1. **Conduct a Pilot Study**
2. **Use a Comparable Study**
3. **Use a Range-based approach**

Sample Size for Estimating the Population Mean

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

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.

Sample Size Determination for a Variable-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 Determination for a Variable- Another Example

A consumer group would like to estimate the mean monthly electricity charge for a single family house in July within $5 using a 99 percent level of confidence. Based on similar studies the standard deviation is estimated to be $20.00. How large a sample is required?

$$n = \left(\frac{(2.58)(20)}{5} \right)^2 = 107$$

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

E is the maximum allowable error

Sample Size Determination - Example

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.

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

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

Another Example

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?

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

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

$$n = 69 \, \text{cities}$$

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.

- However, if *n/N* < .05, the finite-population correction factor may be ignored.

Standard Error of the Mean   Standard Error of the Proportion

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

$$\sigma_p = \sqrt{\frac{p(1-p)}{n}} \sqrt{\frac{N-n}{N-1}}$$

Effects on FPC when *n/N* Changes

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

Observe that FPC approaches 1 when *n/N* becomes smaller

Confidence Interval Formulas for Estimating Means and Proportions with Finite Population Correction

C.I. for the Mean (μ)

$$\overline{X} \pm z \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

C.I. for the Mean (μ)

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

C.I. for the Proportion (π)

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

CI for Mean with FPC - 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?

1. What is the population mean? What is the best estimate of the population mean? Develop is 90% confidence interval.

2. Discuss why the finite-population correction factor should be used.

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)

Use the formula below to compute the confidence interval:

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

CI For Mean with FPC - Example

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

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

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

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

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

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

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.