# Sampling Methods and the Central Limit Theorem

Chapter 8

McGraw-Hill/Irwin

# Learning Objectives

**LO1** Explain why a sample is often the only feasible way to learn something about a population.

**LO2** Describe methods to select a sample.

**LO3** Define sampling error.

**LO4** Describe the sampling distribution of the sample mean.

**LO5** Explain the central limit theorem.

**LO6** Define the standard error of the mean.

**LO7** Apply the central limit theorem to find probabilities of selecting possible sample means from a specified population.

**LO1** Explain why a sample is often the only feasible way to learn something about a population

# Why Sample the Population?

1. To contact the whole population would be **time consuming**.
2. The **cost** of studying all the items in a population **may be prohibitive**.
3. The **physical impossibility** of checking all items in the population.
4. The **destructive nature** of some tests.
5. The **sample results** are **adequate**.

# Probability Sampling and Sampling Methods

**What is a Probability Sample?**

A **probability sample** is a sample of items or individuals chosen so that each member of the population has a chance of being included in the sample.

**Four Most Commonly Used Probability Sampling Methods**

1. Simple Random Sample
2. Systematic Random Sampling
3. Stratified Random Sampling
4. Cluster Sampling

LO2

# Simple Random Sample and Systematic Random Sampling

Simple Random Sample: A sample selected so that each item or person in the population has the same chance of being included.

EXAMPLE:
A population consists of 845 employees of Nitra Industries. A sample of 52 employees is to be selected from that population. The name of each employee is written on a small slip of paper and deposited all of the slips in a box. After they have been thoroughly mixed, the first selection is made by drawing a slip out of the box without looking at it. This process is repeated until the sample of 52 employees is chosen.

Systematic Random Sampling: The items or individuals of the population are arranged in some order. A random starting point is selected and then every *k*th member of the population is selected for the sample.

EXAMPLE

A population consists of 845 employees of Nitra Industries.

A sample of 52 employees is to be selected from that population. First, *k* is calculated as the population size divided by the sample size. *For Nitra Industries*, we would select every 16th (845/52) employee list. If *k* is not a whole number, then round down. Random sampling is used in the selection of the first name. Then, select every 16$^{th}$ name on the list thereafter.

LO2

# Simple Random Sample: Using Table of Random Numbers

A population consists of 845 employees of Nitra Industries. A sample of 52 employees is to be selected from that population.

A more convenient method of selecting a random sample is to use the identification number of each employee and a **table of random numbers such as the** one in Appendix B.6.

| | | | | | | |
|---|---|---|---|---|---|---|
| 5 0 5 2 5 | 5 7 4 5 4 | 2 8 4 5 5 | 6 8 2 2 6 | 3 4 6 5 6 | 3 8 8 8 4 | 3 9 0 1 8 |
| 7 2 5 0 7 | 5 3 3 8 0 | 5 3 8 2 7 | 4 2 4 8 6 | 5 4 4 6 5 | 7 1 8 1 9 | 9 1 1 9 9 |
| 3 4 9 8 6 | 7 4 2 9 7 | 0 0 1 4 4 | 3 8 6 7 6 | 8 9 9 6 7 | 9 8 8 6 9 | 3 9 7 4 4 |
| 6 8 8 5 1 | 2 7 3 0 5 | 0 3 7 5 9 | 4 4 7 2 3 | 9 6 1 0 8 | 7 8 4 8 9 | 1 8 9 1 0 |
| 0 6 7 3 8 | 6 2 8 7 9 | 0 3 9 1 0 | 1 7 3 5 0 | 4 9 1 6 9 | 0 3 8 5 0 | 1 8 9 1 0 |
| 1 1 4 4 8 | 1 0 7 3 4 | 0 5 8 3 7 | 2 4 3 9 7 | 1 0 4 2 0 | 1 6 7 1 2 | 9 4 4 9 6 |

Starting point — Second employee — Third employee — Fourth employee

# Simple Random Sample: Example

**Example**

Jane and Joe Miley operate the Foxtrot Inn, a bed and breakfast in Tryon, North Carolina. There are eight rooms available for rent at this B&B. Listed below is the number of these eight rooms rented each day during June 2011. Use Excel to select a sample of five nights during the month of June.

| June | Rentals | June | Rentals | June | Rentals |
|---|---|---|---|---|---|
| 1 | 0 | 11 | 3 | 21 | 3 |
| 2 | 2 | 12 | 4 | 22 | 2 |
| 3 | 3 | 13 | 4 | 23 | 3 |
| 4 | 2 | 14 | 4 | 24 | 6 |
| 5 | 3 | 15 | 7 | 25 | 0 |
| 6 | 4 | 16 | 0 | 26 | 4 |
| 7 | 2 | 17 | 5 | 27 | 1 |
| 8 | 3 | 18 | 3 | 28 | 1 |
| 9 | 4 | 19 | 6 | 29 | 3 |
| 10 | 7 | 20 | 2 | 30 | 3 |

| June | Rentals | | Sample |
|---|---|---|---|
| 1 | 0 | | 4 |
| 2 | 2 | | 4 |
| 3 | 3 | | 4 |
| 4 | 2 | | 7 |
| 5 | 3 | | 1 |
| 6 | 4 | | |
| 7 | 2 | | |
| 8 | 3 | | |
| 9 | 4 | | |
| 10 | 7 | | |

# Stratified Random Sampling

**Stratified Random Sampling**: A population is first divided into subgroups, called strata, and a sample is selected from each stratum. Useful when a population can be clearly divided in groups based on some characteristics

**EXAMPLE**
Suppose we want to study the advertising expenditures for the 352 largest companies in the United States to determine whether firms with high returns on equity (a measure of profitability) spent more of each sales dollar on advertising than firms with a low return or deficit.

To make sure that the sample is a fair representation of the 352 companies, the companies are grouped on percent return on equity and a sample proportional to the relative size of the group is randomly selected.

| Stratum | Profitability (return on equity) | Number of Firms | Relative Frequency | Number Sampled |
|---|---|---|---|---|
| 1 | 30 percent and over | 8 | 0.02 | 1* |
| 2 | 20 up to 30 percent | 35 | 0.10 | 5* |
| 3 | 10 up to 20 percent | 189 | 0.54 | 27 |
| 4 | 0 up to 10 percent | 115 | 0.33 | 16 |
| 5 | Deficit | 5 | 0.01 | 1 |
| Total | | 352 | 1.00 | 50 |

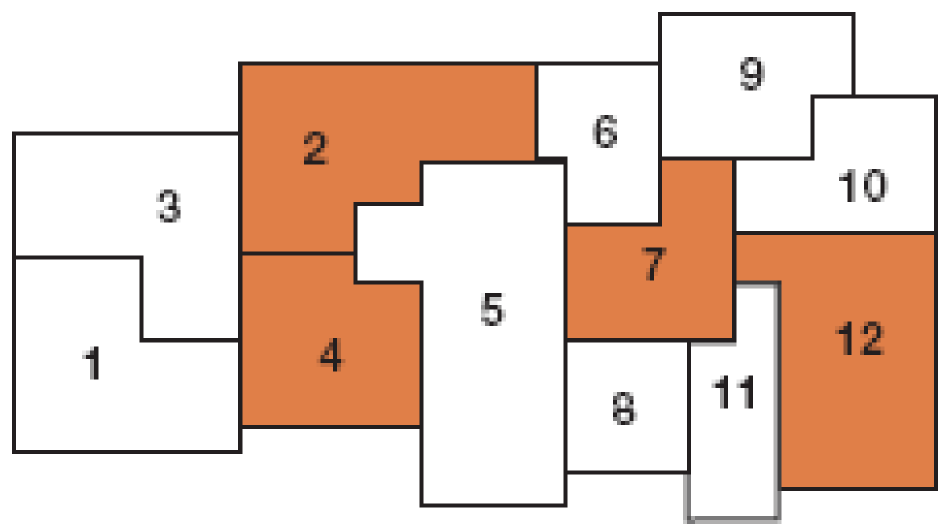
*0.02 of 50 = 1, 0.10 of 50 = 5, etc.

# Cluster Sampling

**Cluster Sampling**: A population is divided into clusters using naturally occurring geographic or other boundaries. Then, clusters are randomly selected and a sample is collected by randomly selecting from each cluster.

**EXAMPLE**
Suppose you want to determine the views of residents in Oregon about state and federal environmental protection policies.

Cluster sampling can be used by subdividing the state into small units—either counties or regions, select at random say 4 regions, then take samples of the residents in each of these regions and interview them. (Note that this is a combination of cluster sampling and simple random sampling.)



# Sampling "Error"

**SAMPLING ERROR** The difference between a sample statistic and its corresponding population parameter.

**Example**

Refer to the previous example on page 268, where we studied the number of rooms rented at the Foxtrot Inn bed and breakfast in Tryon, North Carolina. The population is the number of rooms rented each of the 30 days in June 2011. Find the mean of the population. Use Excel or other statistical software to select three random samples of five days. Calculate the mean of each sample and compare it to the population mean. What is the sampling error in each case?

- $\mu = \frac{\sum X}{N} = 3.13 \rightarrow$ the population mean
- The 1st random sample: 4,4,4,7 and 1. The mean of this example: $\overline{X_1} = \frac{20}{5} = 4.00$
- The sampling error for the 1st sample is $\overline{X_1} - \mu = 4.00 - 3.13 = 0.87$

| | Sample 1 | Sample 2 | Sample 3 |
|---|---|---|---|
| | 4 | 3 | 3 |
| | 4 | 2 | 4 |
| | 4 | 6 | 0 |
| | 7 | 0 | 4 |
| | 1 | 0 | 3 |
| Total | 20 | 11 | 14 |
| Sample means | 4 | 2.2 | 2.8 |
| Population mean | 3.133333 | | |
| The sampling error | 0.866667 | -0.93333 | -0.33333 |

# Sampling Distribution of the Sample Mean

The **sampling distribution of the sample mean** is a probability distribution consisting of all possible sample means of a given sample size selected from a population.

**EXAMPLE**

Tartus Industries has seven production employees (considered the population). The hourly earnings of each employee are given in the table below.

| Employee | Hourly Earnings | Employee | Hourly Earnings |
|---|---|---|---|
| Joe | $7 | Jan | $7 |
| Sam | 7 | Art | 8 |
| Sue | 8 | Ted | 9 |
| Bob | 8 | | |

1. What is the population mean?
2. What is the sampling distribution of the sample mean for samples of size 2?
3. What is the mean of the sampling distribution?
4. What observations can be made about the population and the sampling distribution?

LO4

# Sampling Distribution of the Sample Mean

**EXAMPLE**

Tartus Industries has seven production employees (considered the population). The hourly earnings of each employee are given in the table below.

1. What is the population mean?
2. What is the sampling distribution of the sample mean for samples of size 2?
3. What is the mean of the sampling distribution?
4. What observations can be made about the population and the sampling distribution?

| Employee | Hourly Earnings | Employee | Hourly Earnings |
|---|---|---|---|
| Joe | $7 | Jan | $7 |
| Sam | 7 | Art | 8 |
| Sue | 8 | Ted | 9 |
| Bob | 8 | | |

1. The population mean is $7.71, found by:

$$\mu = \frac{\Sigma X}{N} = \frac{\$7 + \$7 + \$8 + \$8 + \$7 + \$8 + \$9}{7} = \$7.71$$

2. To arrive at the sampling distribution of the sample mean, we need to select all possible samples of 2 without replacement from the population, then compute the mean of each sample. There are 21 possible samples, found by using formula (5–10) on page 173.

$$_NC_n = \frac{N!}{n!(N-n)!} = \frac{7!}{2!(7-2)!} = 21$$

**TABLE 8–3** Sample Means for All Possible Samples of 2 Employees

| Sample | Employees | Hourly Earnings | Sum | Mean | Sample | Employees | Hourly Earnings | Sum | Mean |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Joe, Sam | $7, $7 | $14 | $7.00 | 12 | Sue, Bob | $8, $8 | $16 | $8.00 |
| 2 | Joe, Sue | 7, 8 | 15 | 7.50 | 13 | Sue, Jan | 8, 7 | 15 | 7.50 |
| 3 | Joe, Bob | 7, 8 | 15 | 7.50 | 14 | Sue, Art | 8, 8 | 16 | 8.00 |
| 4 | Joe, Jan | 7, 7 | 14 | 7.00 | 15 | Sue, Ted | 8, 9 | 17 | 8.50 |
| 5 | Joe, Art | 7, 8 | 15 | 7.50 | 16 | Bob, Jan | 8, 7 | 15 | 7.50 |
| 6 | Joe, Ted | 7, 9 | 16 | 8.00 | 17 | Bob, Art | 8, 8 | 16 | 8.00 |
| 7 | Sam, Sue | 7, 8 | 15 | 7.50 | 18 | Bob, Ted | 8, 9 | 17 | 8.50 |
| 8 | Sam, Bob | 7, 8 | 15 | 7.50 | 19 | Jan, Art | 7, 8 | 15 | 7.50 |
| 9 | Sam, Jan | 7, 7 | 14 | 7.00 | 20 | Jan, Ted | 7, 9 | 16 | 8.00 |
| 10 | Sam, Art | 7, 8 | 15 | 7.50 | 21 | Art, Ted | 8, 9 | 17 | 8.50 |
| 11 | Sam, Ted | 7, 9 | 16 | 8.00 | | | | | |

| Sample Mean | Number of Means | Probability |
|---|---|---|
| $7.00 | 3 | .1429 |
| 7.50 | 9 | .4285 |
| 8.00 | 6 | .2857 |
| 8.50 | 3 | .1429 |
| | 21 | 1.0000 |

$$\mu_{\bar{X}} = \frac{\text{Sum of all sample means}}{\text{Total number of samples}} = \frac{\$7.00 + \$7.50 + \cdots + \$8.50}{21}$$

$$= \frac{\$162}{21} = \$7.71$$

LO4

# Sampling Distribution of the Sample Means - Example

$$\mu_{\bar{X}} = \frac{\text{Sum of all sample means}}{\text{Total number of samples}} = \frac{\$7.00 + \$7.50 + \cdots + \$8.50}{21}$$

$$= \frac{\$162}{21} = \$7.71$$

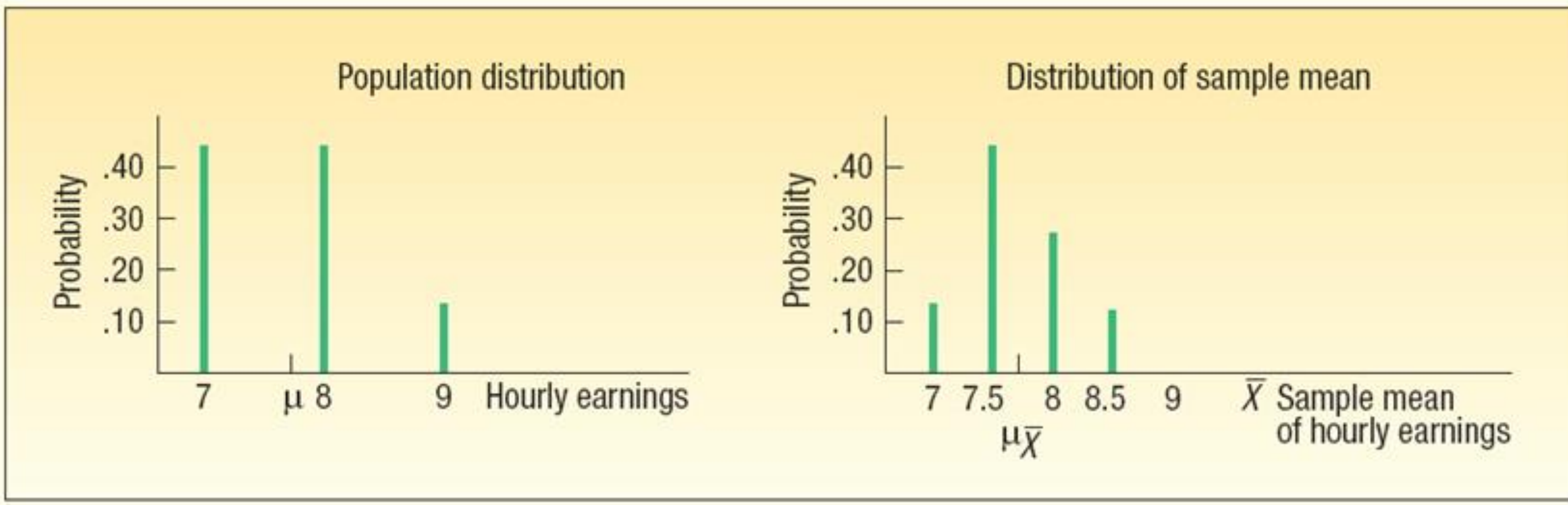


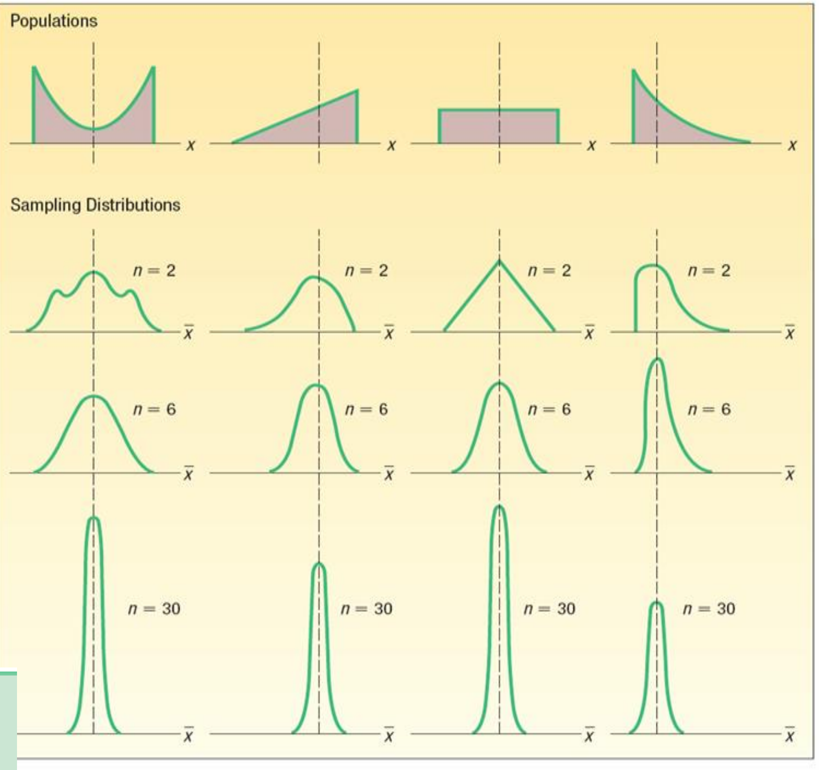

**CHART 8–1** Distributions of Population Values and Sample Mean

# Central Limit Theorem

**CENTRAL LIMIT THEOREM** If all samples of a particular size are selected from any population, the sampling distribution of the sample mean is approximately a normal distribution. This approximation improves with larger samples.

- If the population follows a normal probability distribution, then for any sample size the sampling distribution of the sample mean will also be normal.
- If the population distribution is symmetrical (but not normal), shape of the distribution of the sample mean will emerge as normal with samples as small as 10.
- If a distribution that is skewed or has thick tails, it may require samples of 30 or more to observe the normality feature.
- The mean of the sampling distribution equal to μ and the variance equal to $\sigma^2/n$.

$$\mu = \mu_{\overline{X}}$$




**STANDARD ERROR OF THE MEAN** $$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

# Using the Sampling Distribution of the Sample Mean

- The sampling distribution of the sample mean will follow the normal probability distribution under two conditions:
  1. When the samples are taken from populations known to follow the normal distribution. In this case, the size of the sample is not a factor.
  2. When the shape of the population distribution is not known or the shape is known to be nonnormal, but our sample contains at least 30 observations. We should point out that the number 30 is a guideline that has evolved over the years. In this case, the central limit theorem guarantees the sampling distribution of the mean follows a normal distribution.

LO5

# Using the Sampling Distribution of the Sample Mean

**IF SIGMA IS KNOWN**

- If a population follows the normal distribution, the sampling distribution of the sample mean will also follow the normal distribution.
- If the shape is known to be nonnormal, but the sample contains at least 30 observations, the central limit theorem guarantees the sampling distribution of the mean follows a normal distribution.
- To determine the probability a sample mean falls within a particular region, use:

FINDING THE $z$ VALUE OF $\bar{X}$ WHEN THE POPULATION STANDARD DEVIATION IS KNOWN

$$z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

**IF SIGMA IS UNKNOWN, OR IF POPULATION IS NON NORMAL**

- If the population does not follow the normal distribution, but the sample is of at least 30 observations, the sample means will follow the normal distribution.
- To determine the probability a sample mean falls within a particular region, use:

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

LO5

# Using the Sampling Distribution of the Sample Mean (Sigma Known) - Example

**Example**

The Quality Assurance Department for Cola Inc. maintains records regarding the amount of cola in its Jumbo bottle. The actual amount of cola in each bottle is critical, but varies a small amount from one bottle to the next. Cola Inc. does not wish to underfill the bottles, because it will have a problem with truth in labeling. On the other hand, it cannot overfill each bottle, because it would be giving cola away, hence reducing its profits. Its records indicate that the amount of cola follows the normal probability distribution. The mean amount per bottle is 31.2 ounces and the population standard deviation is 0.4 ounces. At 8 A.M. today the quality technician randomly selected 16 bottles from the filling line. The mean amount of cola contained in the bottles is 31.38 ounces. Is this an unlikely result? Is it likely the process is putting too much soda in the bottles? To put it another way, is the sampling error of 0.18 ounces unusual?

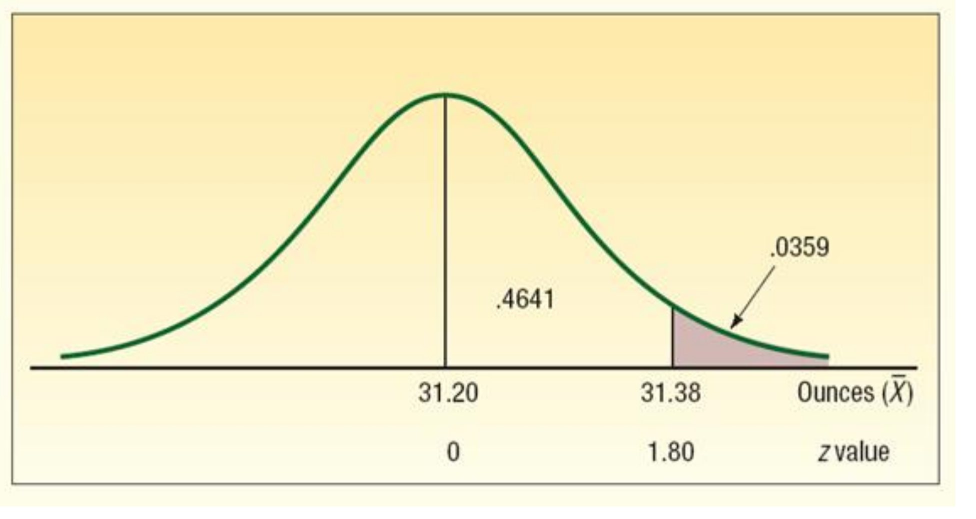
# Using the Sampling Distribution of the Sample Mean (Sigma Known) - Example

**Solution:**

Step 1: Find the *z*-values corresponding to the sample mean of 31.38

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{31.38 - 31.20}{\$0.4/\sqrt{16}} = 1.80$$

Step 2: Find the probability of observing a Z equal to or greater than 1.80




Conclusion: It is unlikely, less than a 4 percent chance, we could select a sample of 16 observations from a normal population with a mean of 31.2 ounces and a population standard deviation of 0.4 ounces and find the sample mean equal to or greater than 31.38 ounces. The process is putting too much cola in the bottles.