Sampling Methods and the Central Limit Theorem

Chapter 8

McGraw-Hill/Irwin

GOALS

1. Explain why a sample is the only feasible way to learn about a population.
2. Describe methods to select a sample.
3. Define and construct a sampling distribution of the sample mean.
4. Explain the *central limit theorem*.
5. Use the central limit theorem to find probabilities of selecting possible sample means from a specified 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

- A probability sample is a sample selected such that each item or person in the population being studied has a known likelihood of being included in the sample.

Most Commonly Used Probability Sampling Methods

- Simple Random Sample
- Systematic Random Sampling
- Stratified Random Sampling
- Cluster Sampling

Simple Random Sample

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.

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

Systematic Random Sampling

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.

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

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.

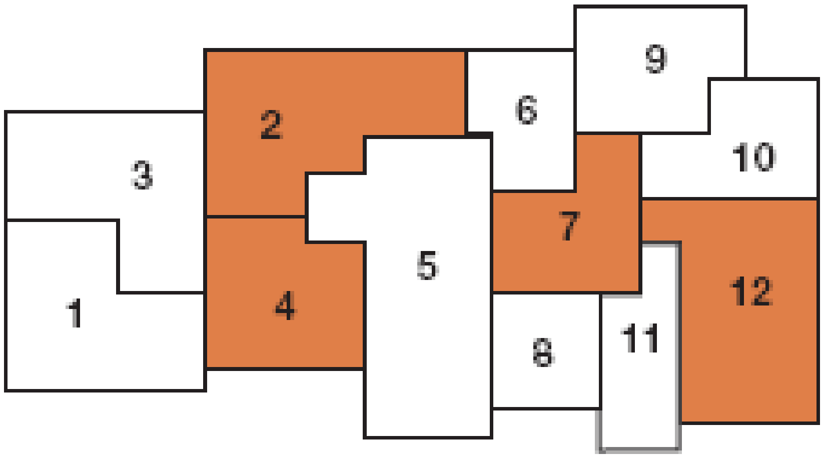
Stratum	Profitability (return on equity)	Number of Firms	Relative Frequency	Number Sampled
1	30 percent and over	8	0.02	1*
2	20 up to 30 percent	35	0.10	5*
3	10 up to 20 percent	189	0.54	27
4	0 up to 10 percent	115	0.33	16
5	Deficit	5	0.01	1
Total		352	1.00	50

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.

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.)



Methods of Probability Sampling

- In nonprobability sample inclusion in the sample is based on the judgment of the person selecting the sample.

- The sampling error is the difference between a sample statistic and its corresponding population parameter.

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.

Sampling Distribution of the Sample Means - Example

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

TABLE 8–2 Hourly Earnings of the Production Employees of Tartus Industries

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?

Sampling Distribution of the Sample Means - Example

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.

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

Sampling Distribution of the Sample Means - Example

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					

Sampling Distribution of the Sample Means - Example

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

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

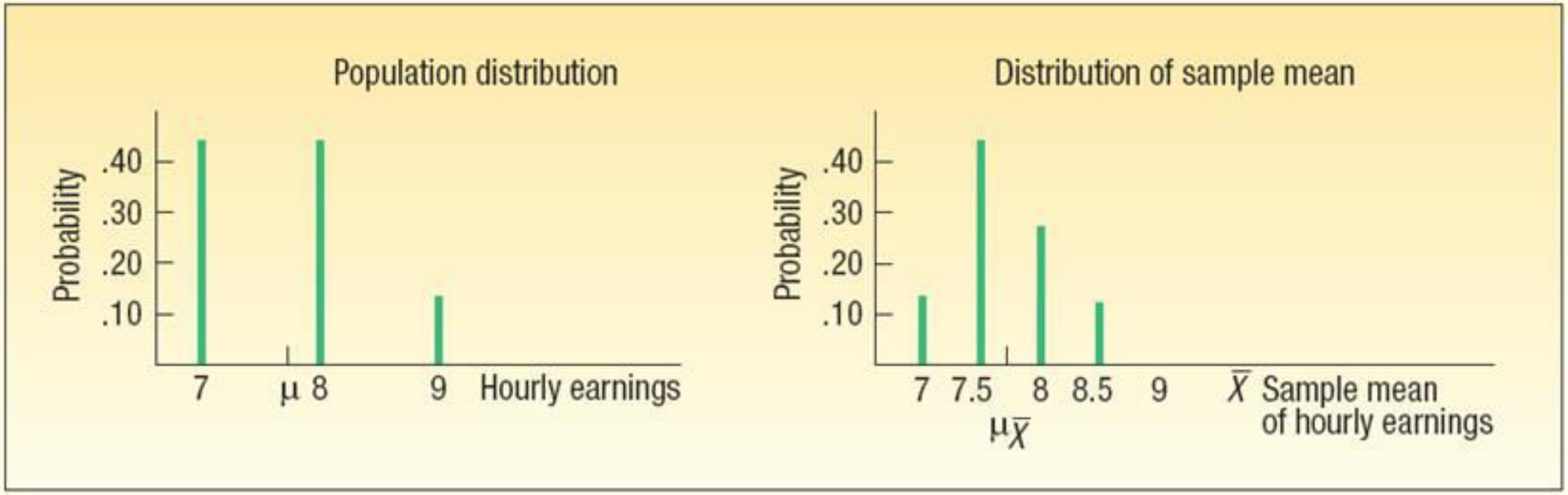


CHART 8–1 Distributions of Population Values and Sample Mean

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), the normal shape of the distribution of the sample mean emerge 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 σ^2/n.

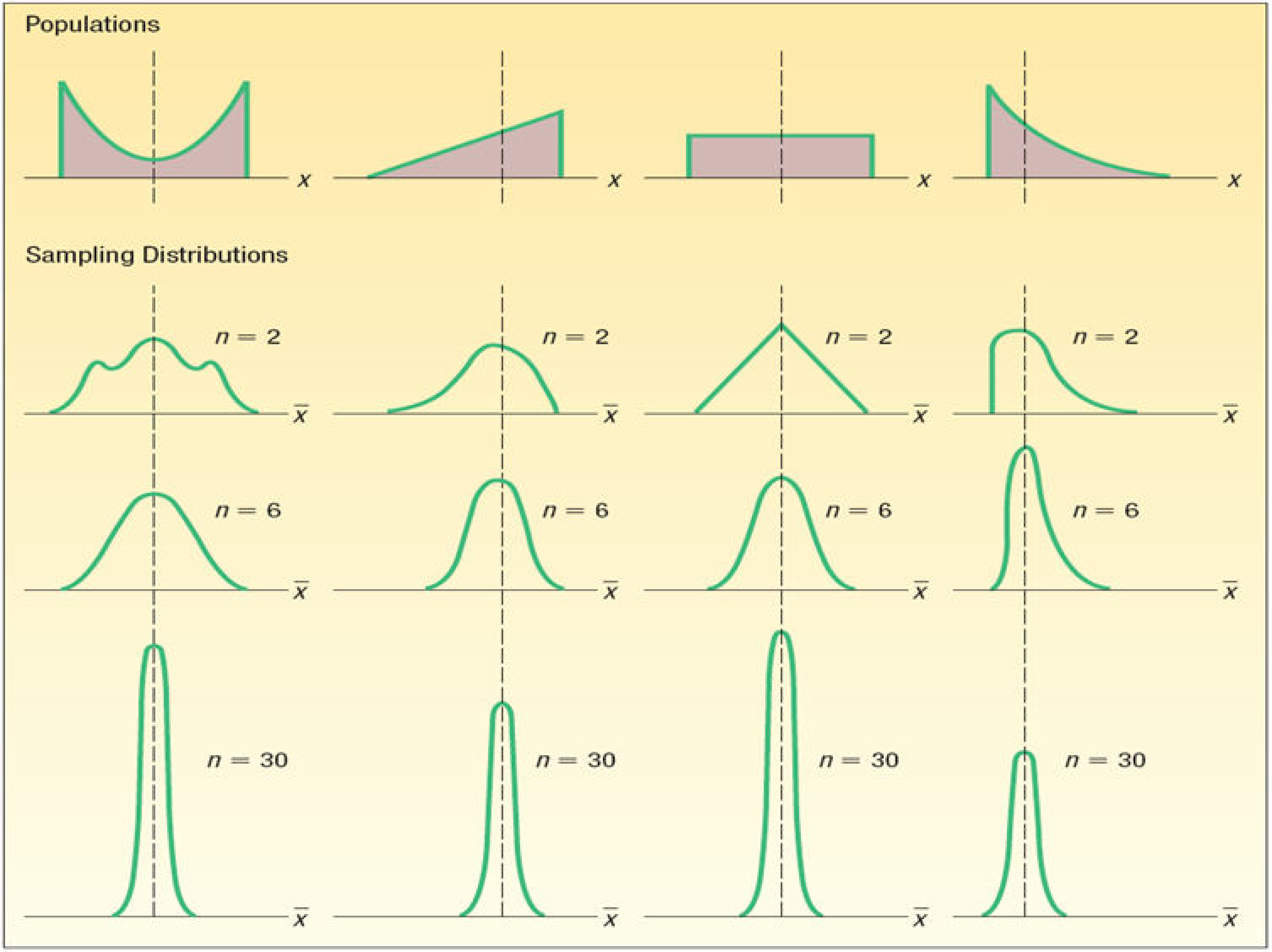


CHART 8–2 Results of the Central Limit Theorem for Several Populations

Using the Sampling Distribution of the Sample Mean (Sigma 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:

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

Using the Sampling Distribution of the Sample Mean (Sigma Unknown)

- 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:

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

Using the Sampling Distribution of the Sample Mean (Sigma Known) - 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. On the other hand, it cannot overfill each bottle. 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

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

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

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

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

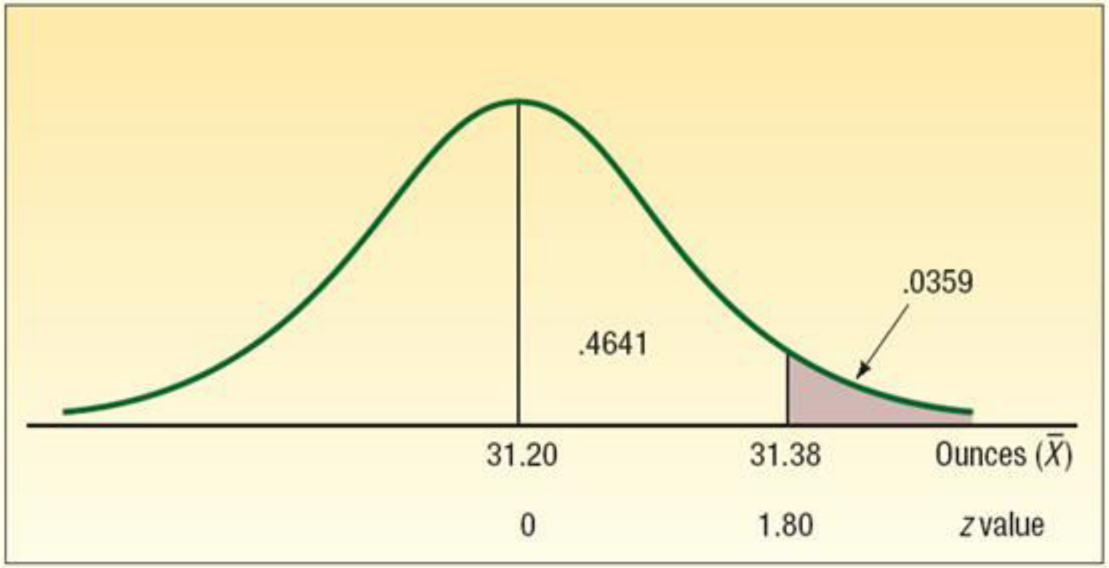


CHART 8–6 Sampling Distribution of the Mean Amount of Cola in a Jumbo Bottle

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

What do we conclude?

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.

We conclude the process is putting too much cola in the bottles.