Sampling Methods and the Central Limit Theorem

Chapter 8

Dr. M R Karim (Associate Professor, Department of Statistics, Jahangirnagar University)

GOALS

- Explain why a sample is the only feasible way to learn about a population.

- Describe methods to select a sample.

- Define and construct a sampling distribution of the sample mean.

- Explain the central limit theorem.

- Use the Central Limit Theorem to find probabilities of selecting possible sample means from a specified population.

Dr. M R Karim (Associate Professor, Department of Statistics, Jahangirnagar University)

Why Sample the Population?

- The physical impossibility of checking all items in the population.

- The cost of studying all the items in a population.

- The sample results are usually adequate.

- Contacting the whole population would often be time-consuming.

- The destructive nature of certain tests.

Dr. M R Karim (Associate Professor, Department of Statistics, Jahangirnagar University)

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.

Dr. M R Karim (Associate Professor, Department of Statistics, Jahangirnagar University)

Methods of Probability Sampling

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

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

Dr. M R Karim (Associate Professor, Department of Statistics, Jahangirnagar University)

Methods of Probability Sampling

- Stratified Random Sampling: A population is first divided into subgroups, called strata, and a sample is selected from each stratum.

- Cluster Sampling: A population is first divided into primary units then samples are selected from the primary units.

Dr. M R Karim (Associate Professor, Department of Statistics, Jahangirnagar University)

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.

Dr. M R Karim (Associate Professor, Department of Statistics, Jahangirnagar University)

Sampling Distribution of the Sample Means

● 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.

Dr. M R Karim (Associate Professor, Department of Statistics, Jahangirnagar University)

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?

Dr. M R Karim (Associate Professor, Department of Statistics, Jahangirnagar University)

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.

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

Dr. M R Karim (Associate Professor, Department of Statistics, Jahangirnagar University)

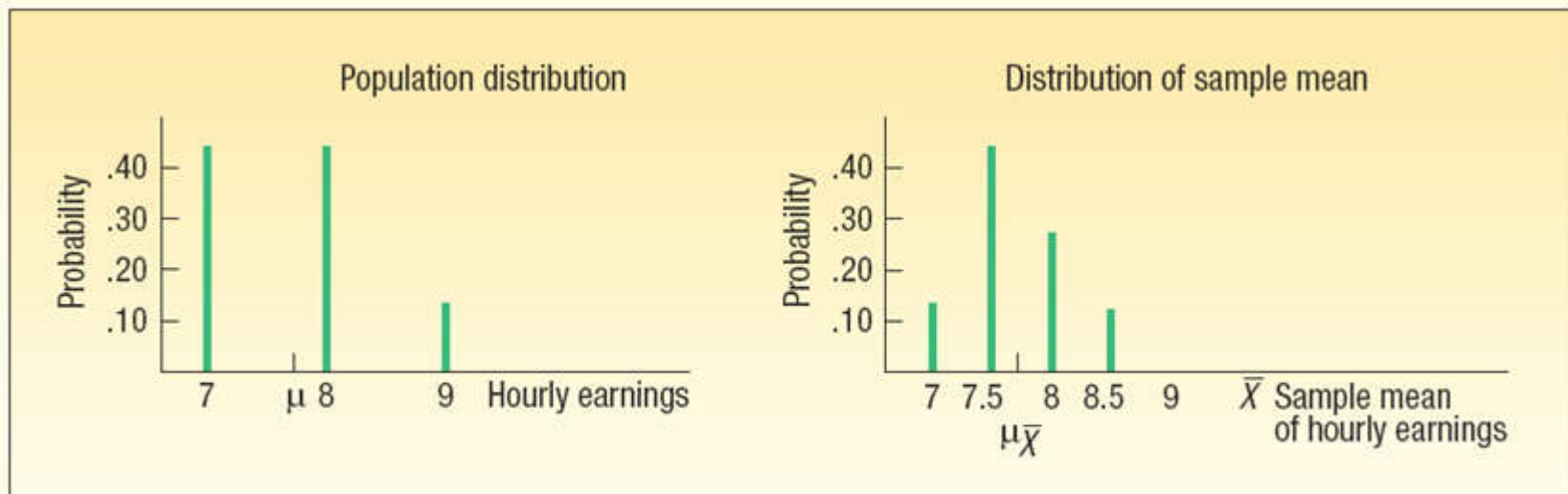# Sampling Distribution of the Sample Means - Example

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					

Dr. M R Karim (Associate Professor, Department of Statistics, Jahangirnagar University)

Sampling Distribution of the Sample Means - Example

$$\mu_{\overline{X}} = \frac{\text{Sum of all sample means}}{\text{Total number of samples}} = \frac{\$7.00 + \$7.50 + \cdots + \$8.50}{21}$$
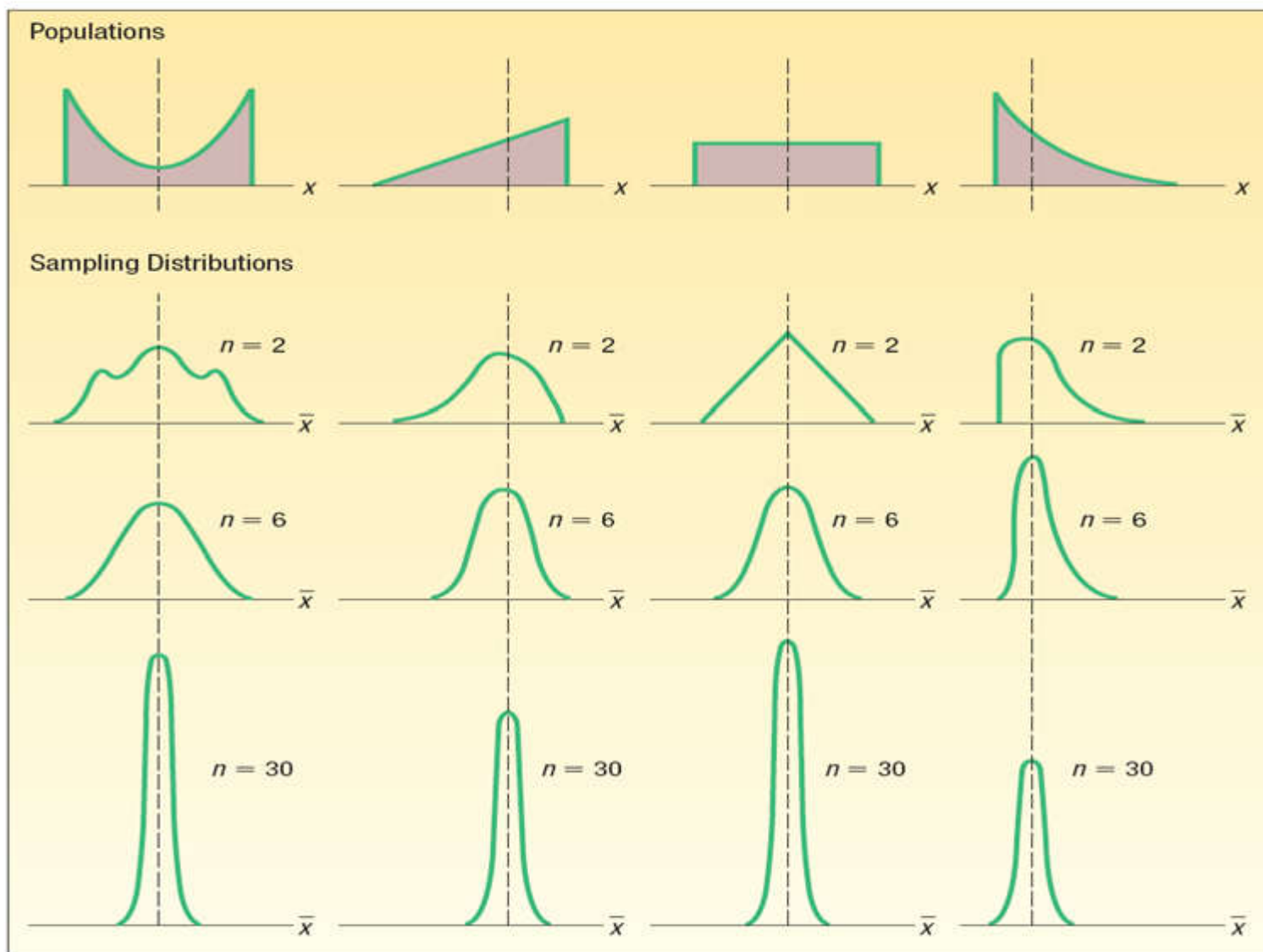
$$= \frac{\$162}{21} = \$7.71$$

CHART 8–1 Distributions of Population Values and Sample Mean

Dr. M R Karim (Associate Professor, Department of Statistics, Jahangirnagar University)

Central Limit Theorem

- For a population with a mean μ and a variance σ^2 the sampling distribution of the means of all possible samples of size *n* generated from the population will be approximately normally distributed.

- The mean of the sampling distribution equal to μ and the variance equal to σ^2/n.

Dr. M R Karim (Associate Professor, Department of Statistics, Jahangirnagar University)

Dr. M 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.

- To determine the probability a sample mean falls within a particular region, use:

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

Dr. M R Karim (Associate Professor, Department of Statistics, Jahangirnagar University)

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:

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

Dr. M R Karim (Associate Professor, Department of Statistics, Jahangirnagar University)

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?

Dr. M R Karim (Associate Professor, Department of Statistics, Jahangirnagar University)

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

$$z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{31.38 - 32.20}{\$0.2/\sqrt{16}} = 1.80$$

Dr. M R Karim (Associate Professor, Department of Statistics, Jahangirnagar University)

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

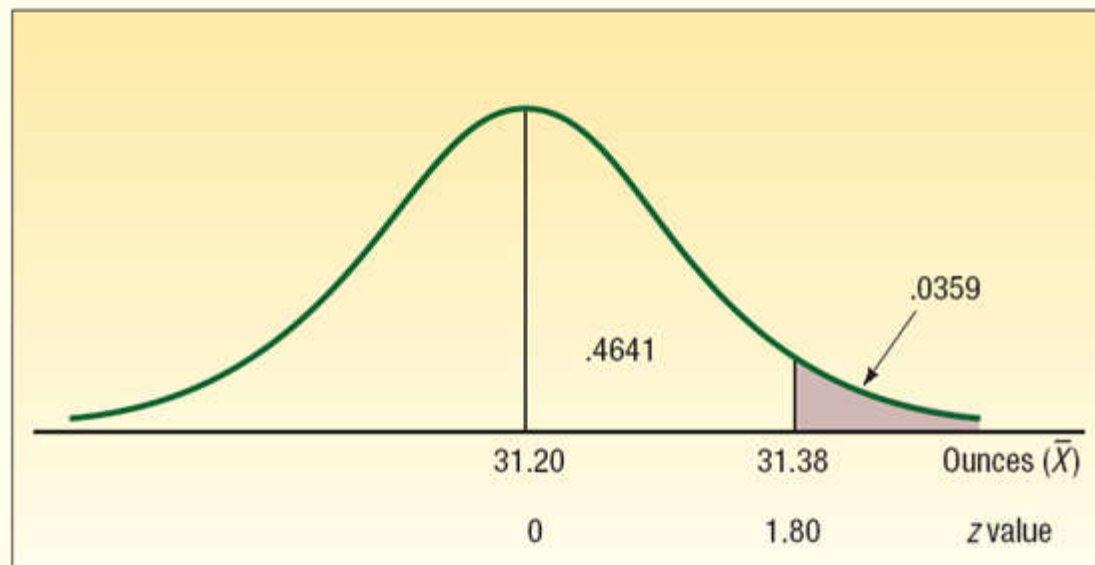Step 2: Find the probability of observing a Z equal to or greater than 1.80

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

Dr. M R Karim (Associate Professor, Department of Statistics, Jahangirnagar University)

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.

Dr. M R Karim (Associate Professor, Department of Statistics, Jahangirnagar University)

End of Chapter 8

Dr. M R Karim (Associate Professor, Department of Statistics, Jahangirnagar University)