Sampling Methods and the Central Limit Theorem

Chapter 08

McGraw-Hill/Irwin

Copyright © 2013 by The McGraw-Hill Companies, Inc. All rights reserved.

LEARNING OBJECTIVES

LO 8-1 Explain why a sample is often the only feasible way to learn something about a population.

LO 8-2 Describe methods to select a sample.

LO 8-3 Define sampling error.

LO 8-4 Describe the sampling distribution of the sample mean.

LO 8-5 Explain the central limit theorem.

LO 8-6 Define the standard error of the mean.

LO 8-7 Apply the central limit theorem to find probabilities of selecting possible sample means from a specified population.

LO 8-1 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

- 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

LO 8-2

Simple Random Sample

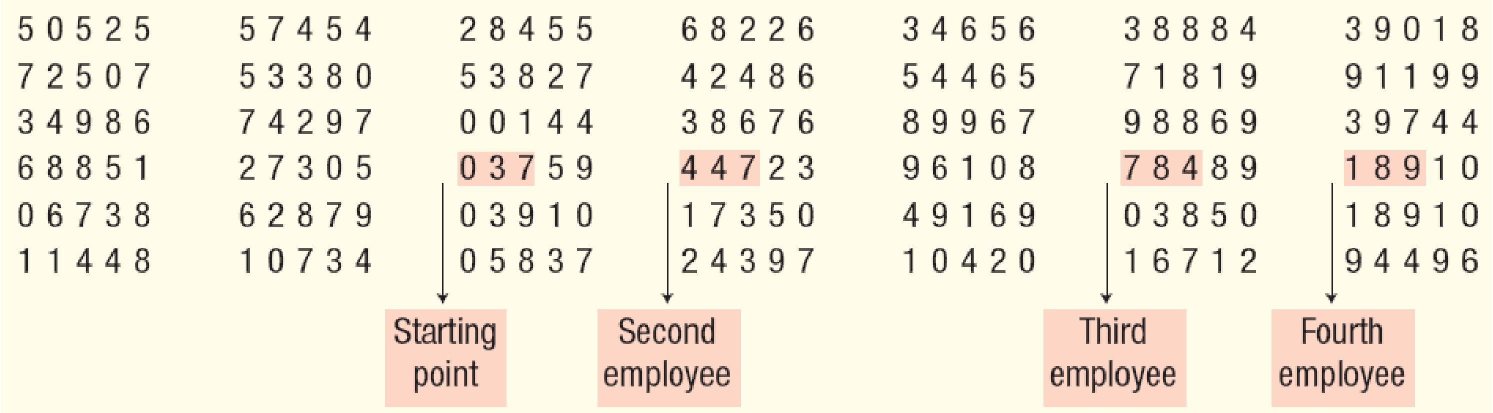
Simple Random Sample: A sample selected so that each item or person in the population has the same chance of being included.

LO 8-2

Simple Random Sample: Using Tables 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.

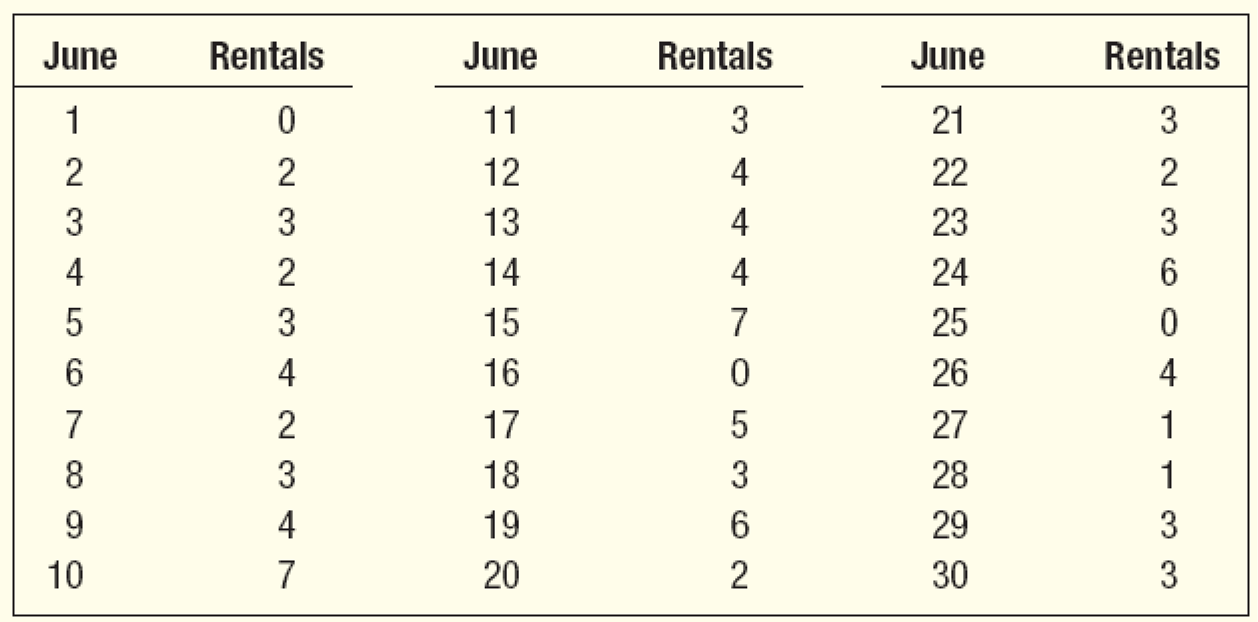


50525	57454	28455	68226	34656	38884	39018
72507	53380	53827	42486	54465	71819	91199
34986	74297	00144	38676	89967	98869	39744
68851	27305	03759	44723	96108	78489	18910
06738	62879	03910	17350	49169	03850	18910
11448	10734	05837	24397	10420	16712	94496

Starting point
Second employee
Third employee
Fourth employee

Simple Random Sample: Using Excel

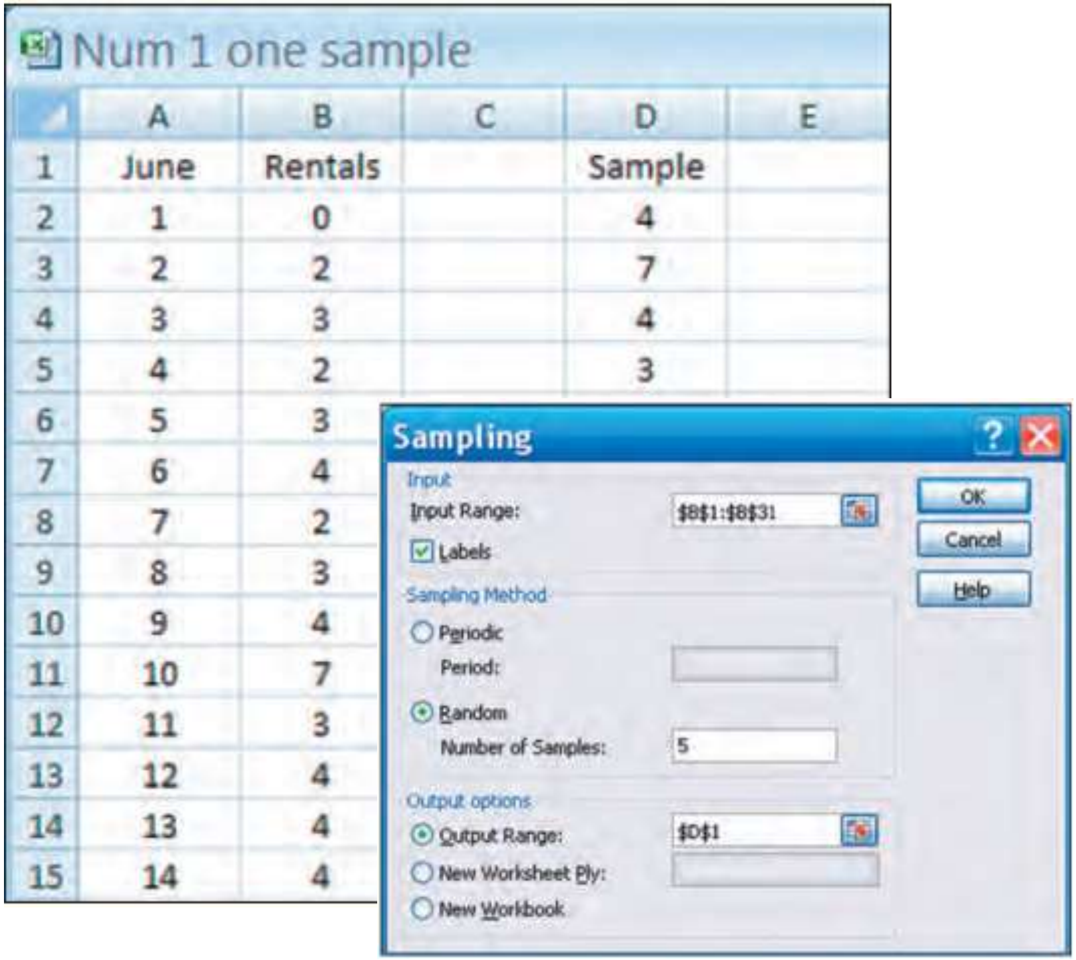
Jane and Joe Miley operate the Foxtrot Inn, a bed and breakfast in Tryon, North Carolina. There are eight rooms available for rent at this B&B. Listed below is the number of these eight rooms rented each day during June 2008. Use Excel to select a sample of five nights during the month of June.



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

Simple Random Sample: Using Excel

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 2008. Use Excel to select a sample of five nights during the month of June.



Num 1 one sample

	A	B	C	D	E
1	June	Rentals		Sample	
2	1	0		4	
3	2	2		7	
4	3	3		4	
5	4	2		3	
6	5	3			
7	6	4			
8	7	2			
9	8	3			
10	9	4			
11	10	7			
12	11	3			
13	12	4			
14	13	4			
15	14	4			

Sampling

Input

Input Range: $B$1:$B$31

☑ Labels

Sampling Method

○ Periodic

Period:

● Random

Number of Samples: 5

Output options

● Output Range: $D$1

○ New Worksheet Ply:

○ New Workbook

LO 8-2

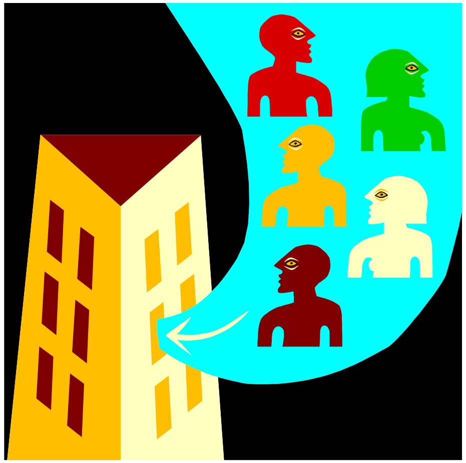
Systematic Random Sampling

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.

Use the systematic random sampling to select the samples.



Systematic Random Sampling

Given: *N=845, n=52*

Step 1: Calculate *k*

$k = N/n$

$= 845/52$ *(round down)*

$k = 16$

Step 2: Use simple random sampling to select the first sample.

Step 3: Select every *16*th element on the list after Step 2

LO 8-2

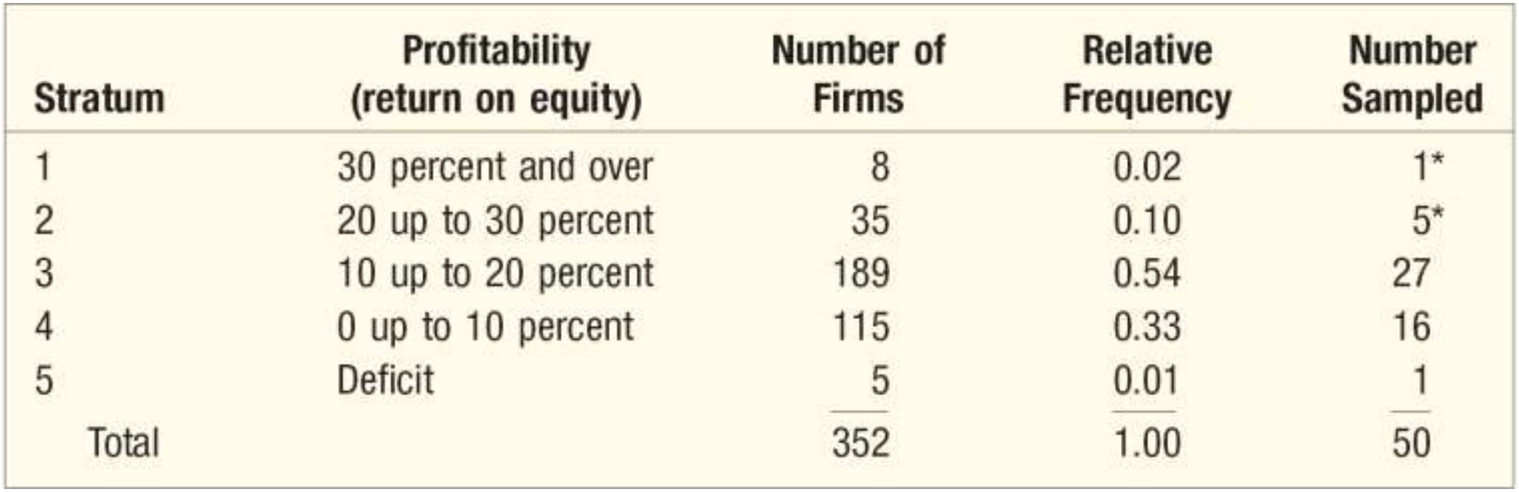
Stratified Random Sampling

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.

Stratified Random Sampling

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.

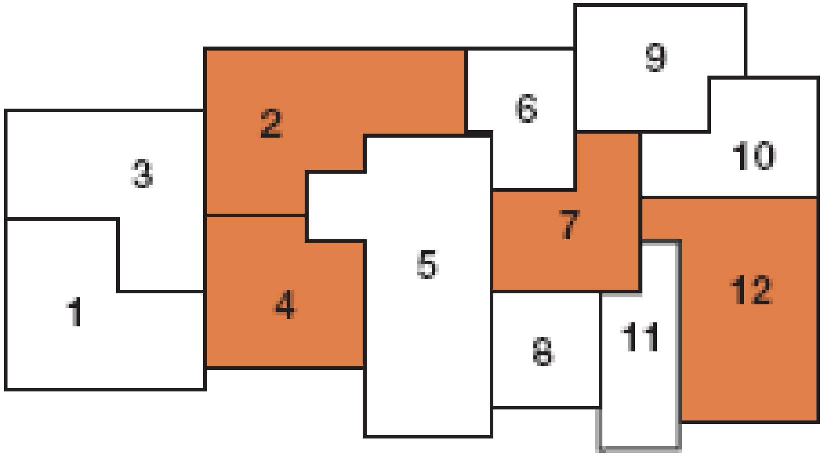


Stratum	Profitability (return on equity)	Number of Firms	Relative Frequency	Number Sampled
1	30 percent and over	8	0.02	1*
2	20 up to 30 percent	35	0.10	5*
3	10 up to 20 percent	189	0.54	27
4	0 up to 10 percent	115	0.33	16
5	Deficit	5	0.01	1
Total		352	1.00	50

LO 8-2

Cluster Sampling

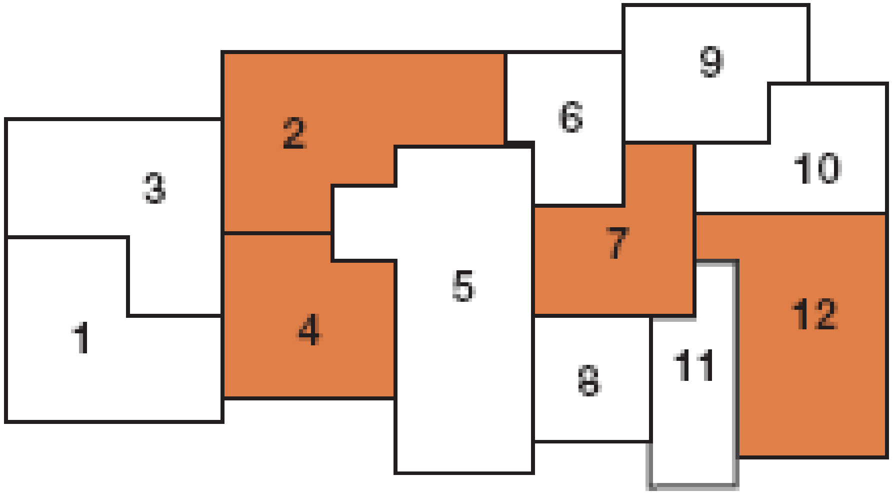
Cluster Sampling: A population is divided into clusters using naturally occurring geographic or other boundaries. Then, clusters are randomly selected and a sample is collected by randomly selecting from each cluster.



Cluster Sampling

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

Step 1: Subdivide the state into small units—either counties or regions,
Step 2: Select at random, say 4 regions, then
Step 3: Take samples of the residents in each of these regions and interview them.



Sampling Error

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

Examples:

$$\overline{X} - \mu$$

$$s - \sigma$$

$$s^2 - \sigma^2$$

$$p - \pi$$

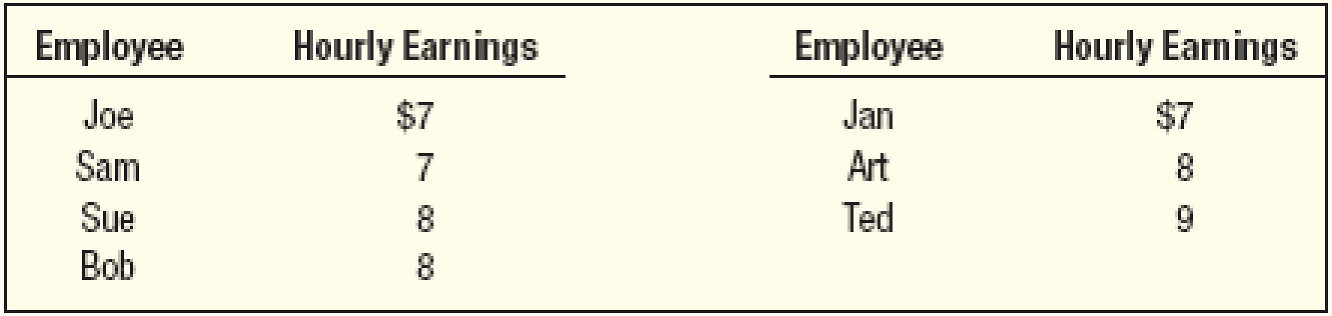
Sampling Distribution of the Sample Mean

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.

LO 8-4

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.

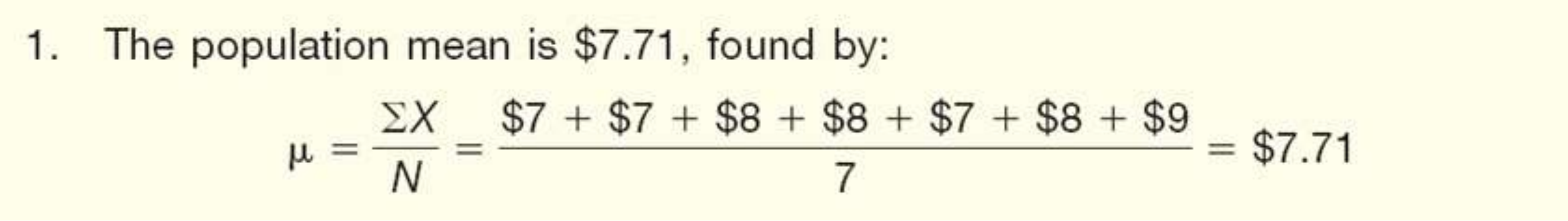


Employee	Hourly Earnings	Employee	Hourly Earnings
Joe	$7	Jan	$7
Sam	7	Art	8
Sue	8	Ted	9
Bob	8		

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?

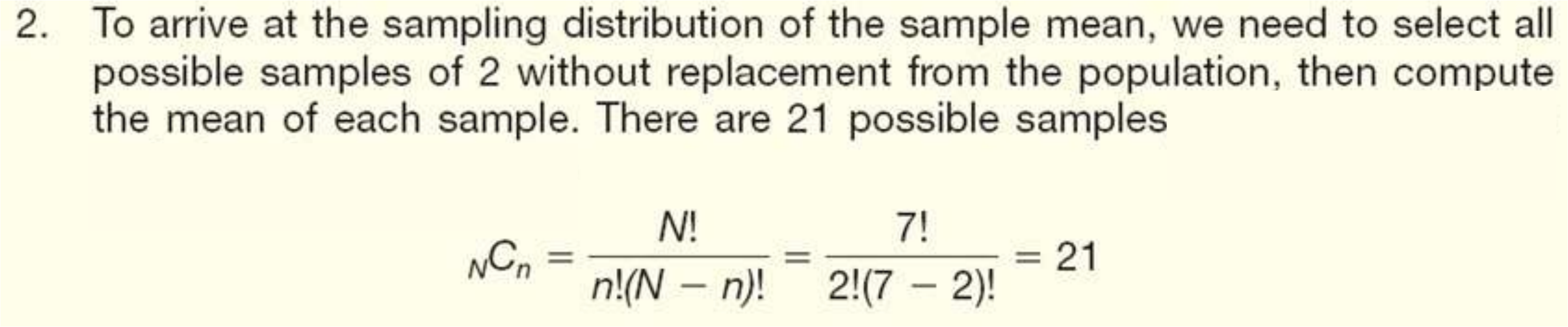
LO 8-4

Sampling Distribution of the Sample Means – Example



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

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

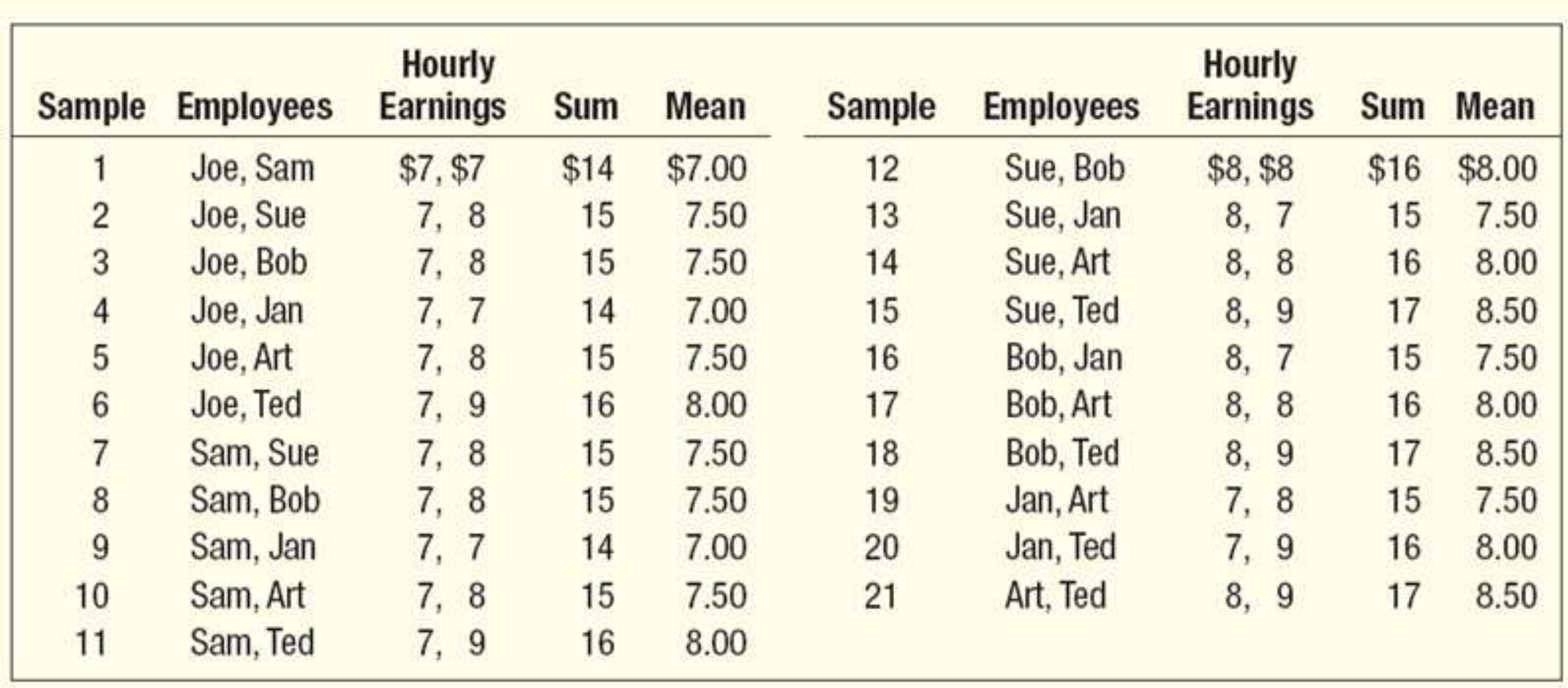


2. To arrive at the sampling distribution of the sample mean, we need to select all possible samples of 2 without replacement from the population, then compute the mean of each sample. There are 21 possible samples

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

LO 8-4

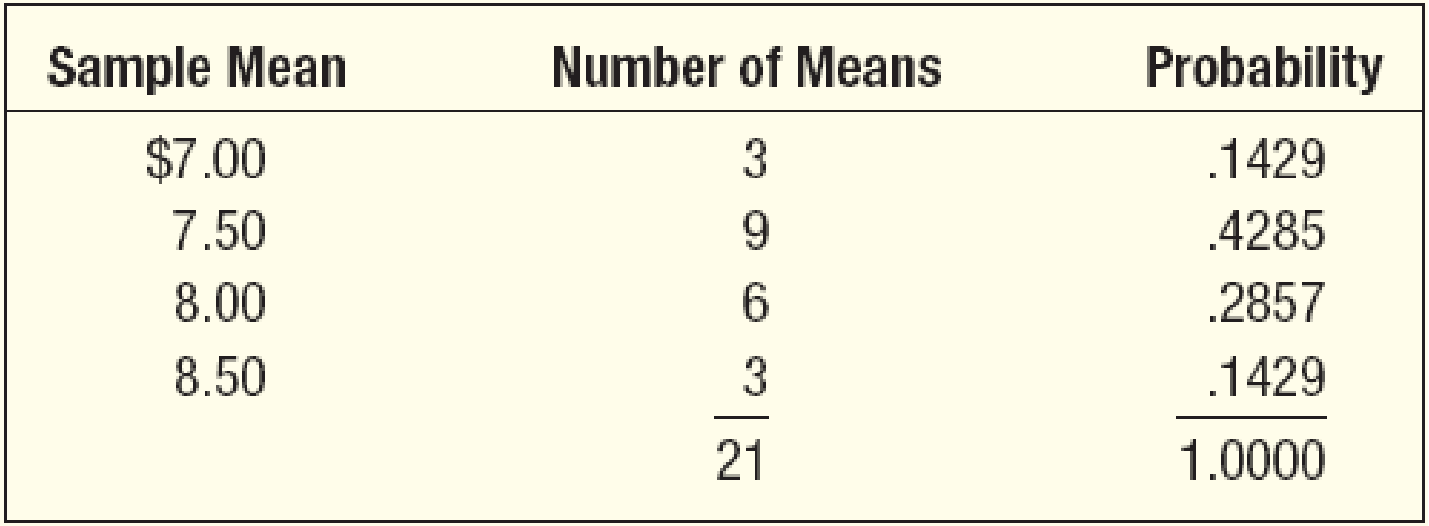
Sampling Distribution of the Sample Means – Example



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					

LO 8-4

Sampling Distribution of the Sample Means – Example

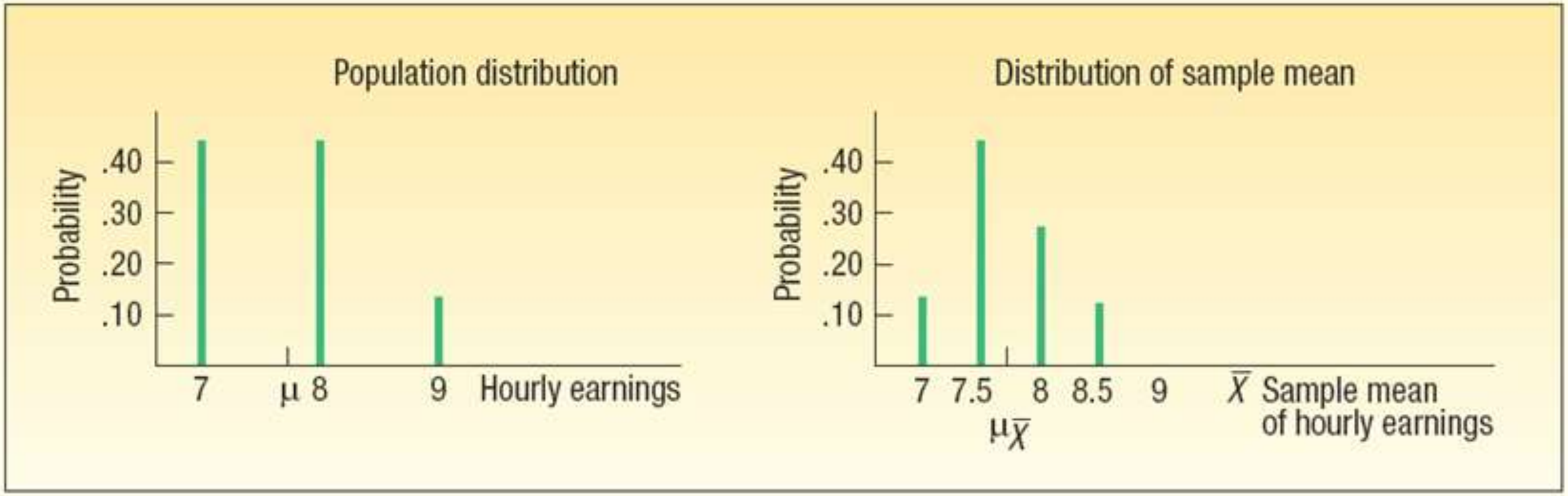


Sample Mean	Number of Means	Probability
$7.00	3	.1429
7.50	9	.4285
8.00	6	.2857
8.50	3	.1429
	21	1.0000

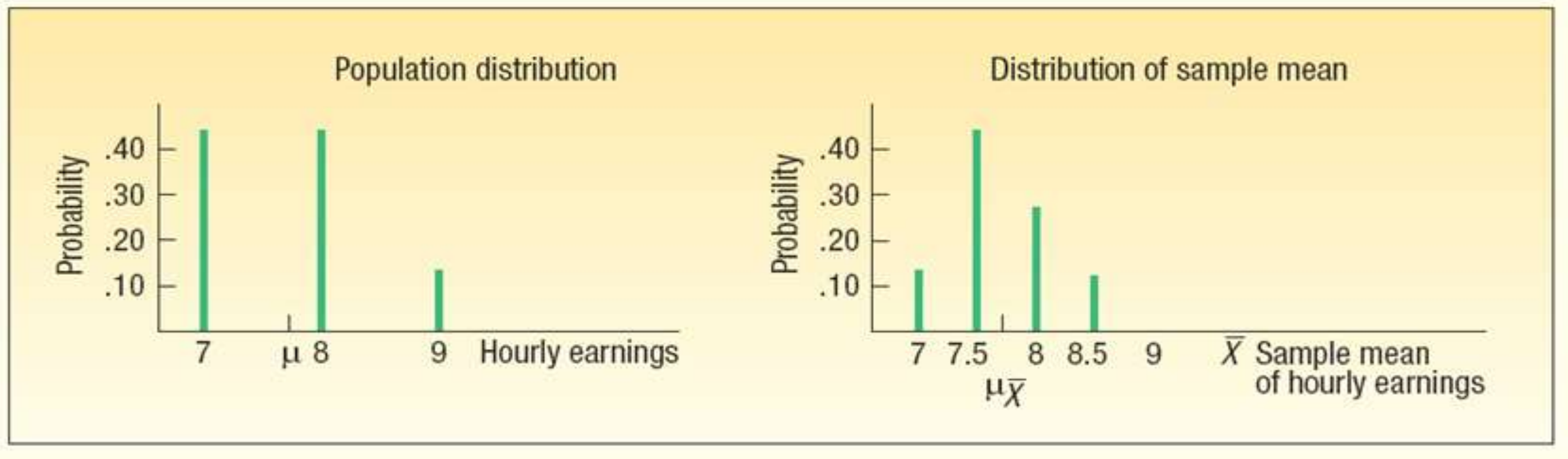
Sampling Distribution of the Sample Means – Example

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

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



Sampling Distribution of the Sample Means – Example



These observations can be made:

A. The mean of the distribution of the sample mean ($7.71) is equal to the mean of the population.

B. The spread in the distribution of the sample mean is less than the spread in the population values.

C. As the size of the sample is increased, the spread of the distribution of the sample mean becomes smaller.

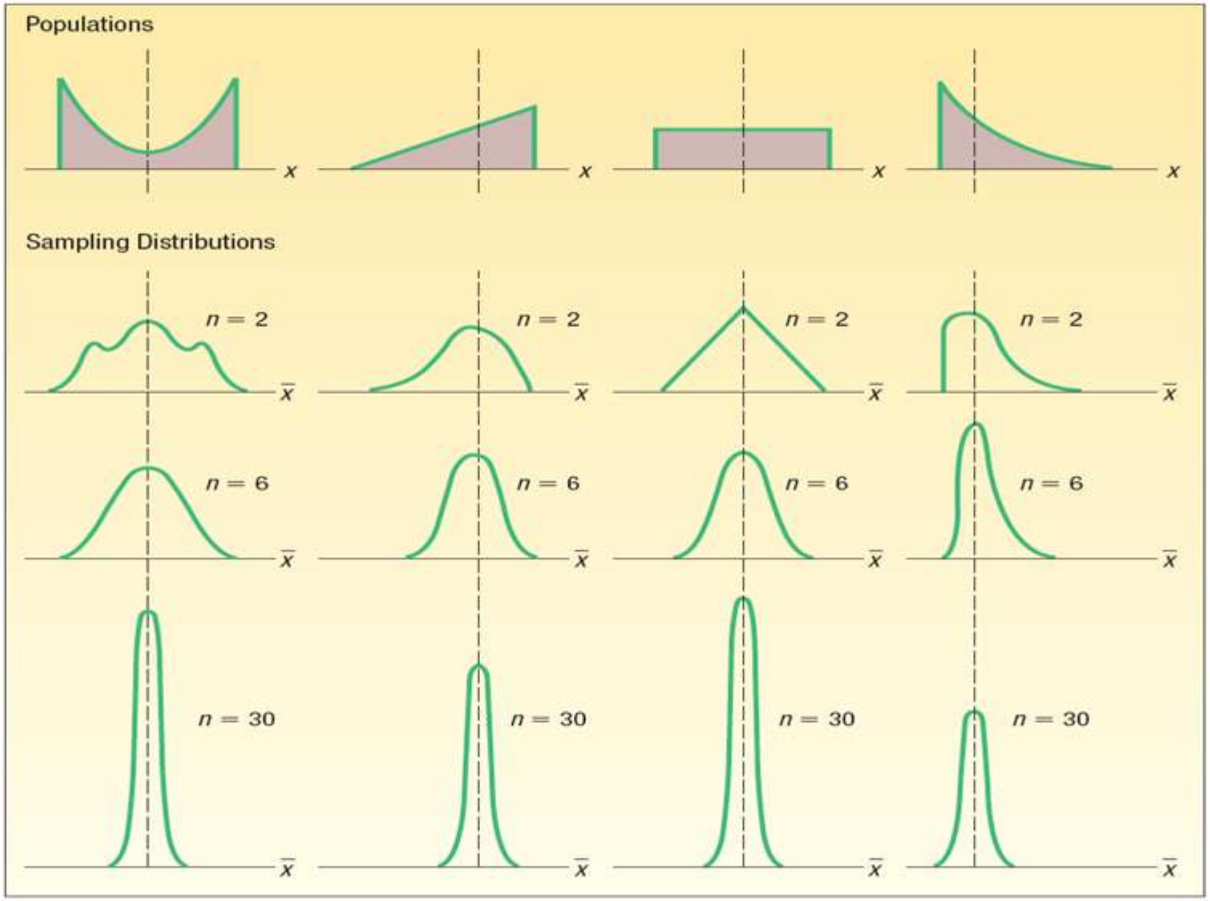
D. The shape of the sampling distribution of the sample mean and the shape of the frequency distribution of the population values are different. The distribution of the sample mean tends to be more bell-shaped and to approximate the normal probability distribution.

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.

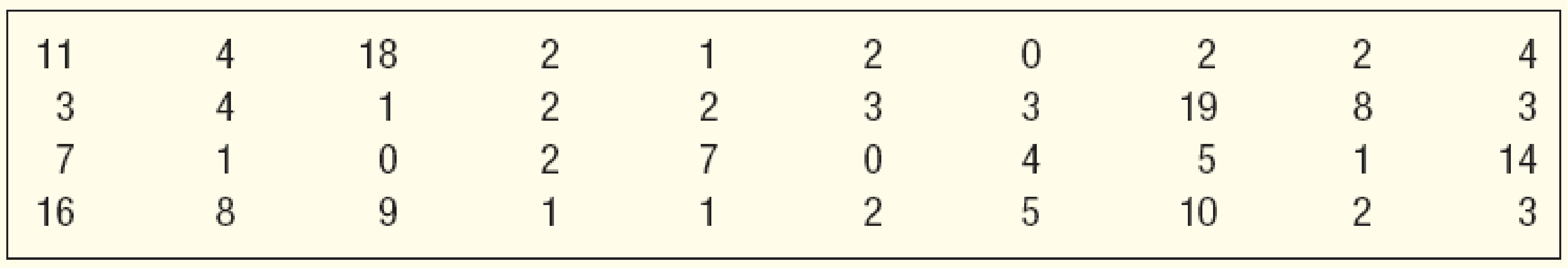
- For any sample size, the sampling distribution of the sample mean will also be normal if the population follows a normal probability distribution.
- If the population distribution is symmetrical (but not normal), the normal shape of the distribution of the sample mean emerges with samples as small as 10.
- If a distribution 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 is equal to μ and the variance is equal to σ^2/n.

Sampling Methods and the Central Limit Theorem


Populations
x
x
x
x
Sampling Distributions
$n = 2$
$n = 2$
$n = 2$
$n = 2$
$\bar{x}$
$n = 6$
$n = 6$
$n = 6$
$n = 6$
$\bar{x}$
$n = 30$
$n = 30$
$n = 30$
$n = 30$
$\bar{x}$

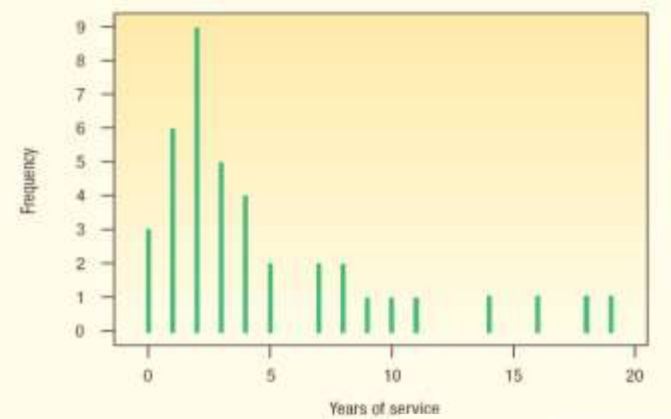
Central Limit Theorem – Example

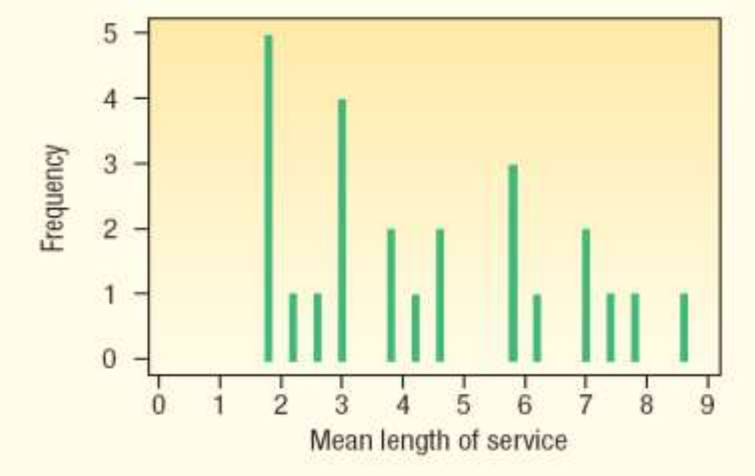
Spence Sprockets, Inc. employs 40 people and faces some major decisions regarding health care for these employees. Before making a final decision on what health care plan to purchase, Ed decides to form a committee of five representative employees. The committee will be asked to study the health care issue carefully and make a recommendation as to what plan best fits the employees’ needs. Ed feels the views of newer employees toward health care may differ from those of more experienced employees. If Ed randomly selects this committee, what can he expect in terms of the mean years with Spence Sprockets for those on the committee? How does the shape of the distribution of years of experience of all employees (the population) compare with the shape of the sampling distribution of the mean? The lengths of service (rounded to the nearest year) of the 40 employees currently on the Spence Sprockets, Inc., payroll are as follows.



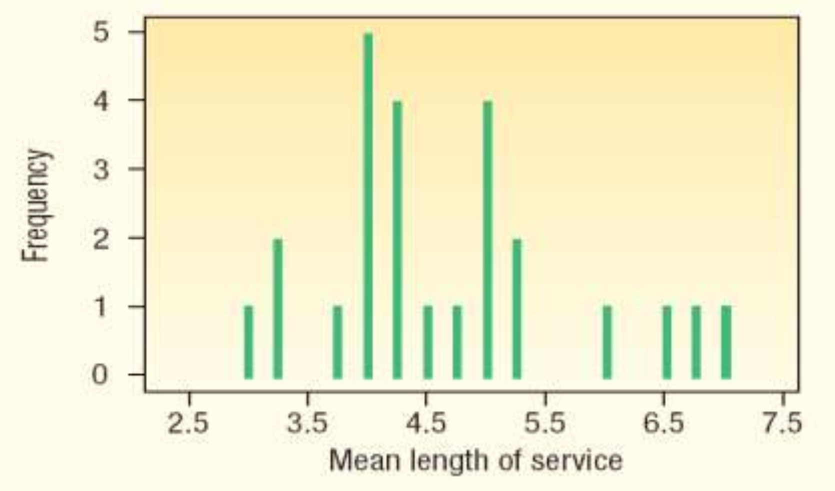
11	4	18	2	1	2	0	2	2	4
3	4	1	2	2	3	3	19	8	3
7	1	0	2	7	0	4	5	1	14
16	8	9	1	1	2	5	10	2	3

Central Limit Theorem – Example





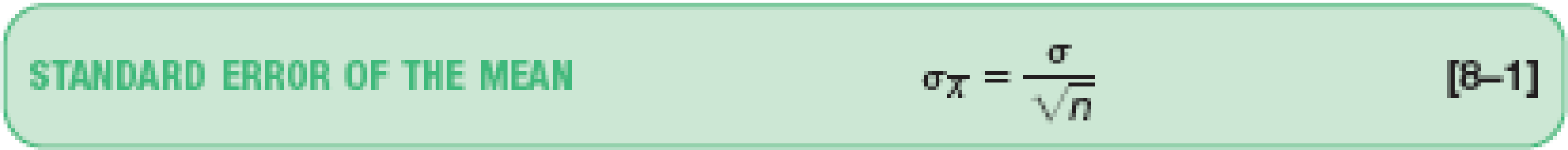
25 Samples of Five Employees



25 Samples of 20 Employees

LO 8-6 Define the standard error of the mean.

Standard Error of the Mean



STANDARD ERROR OF THE MEAN $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ [8–1]

1. The mean of the distribution of sample means will be *exactly equal to the population* mean if we are able to select all possible samples of the same size from a given population.

2. There will be less dispersion in the sampling distribution of the sample mean than in the population. As the sample size increases, the standard error of the mean decreases

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 non-normal, 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 distribution of the sample means will follow the normal distribution.
- To determine the probability a sample mean falls within a particular region, use:

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

LO7 8-7 Apply the central limit theorem to find probabilities of selecting possible sample means from a specified population.

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

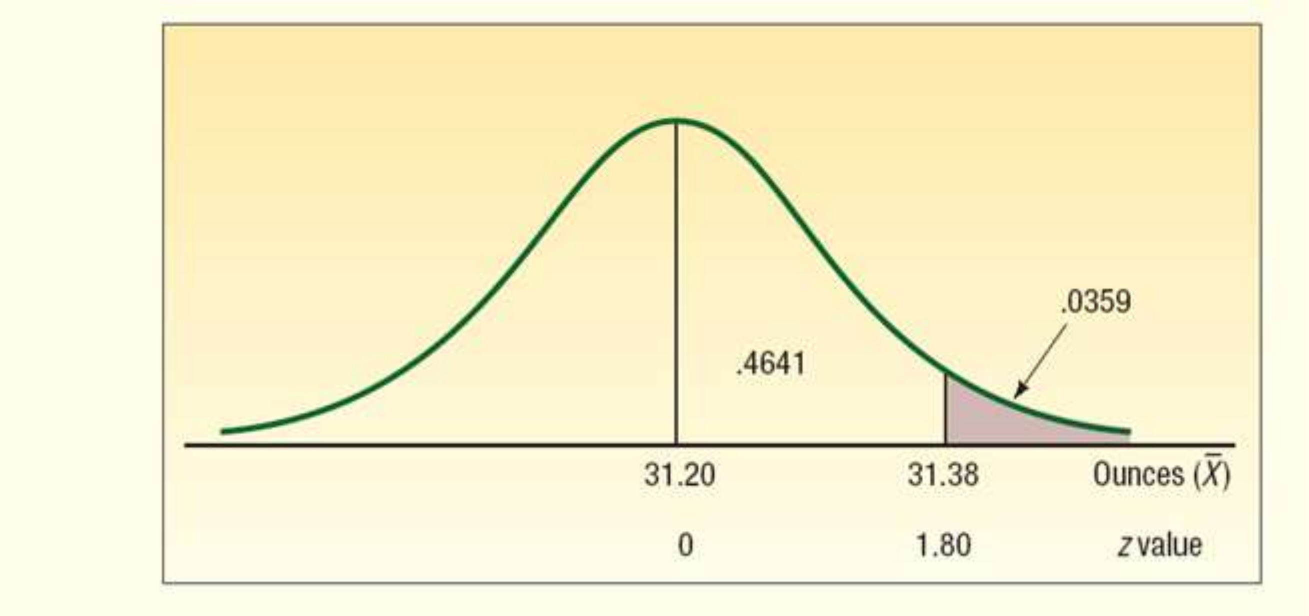
Step 1: Find the *z* value corresponding to the sample mean of 31.38.

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

LO 8-7

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

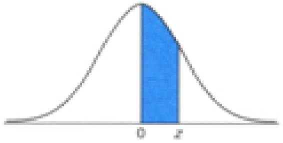
Step 2: Find the probability of observing a *z* equal to or greater than 1.80.

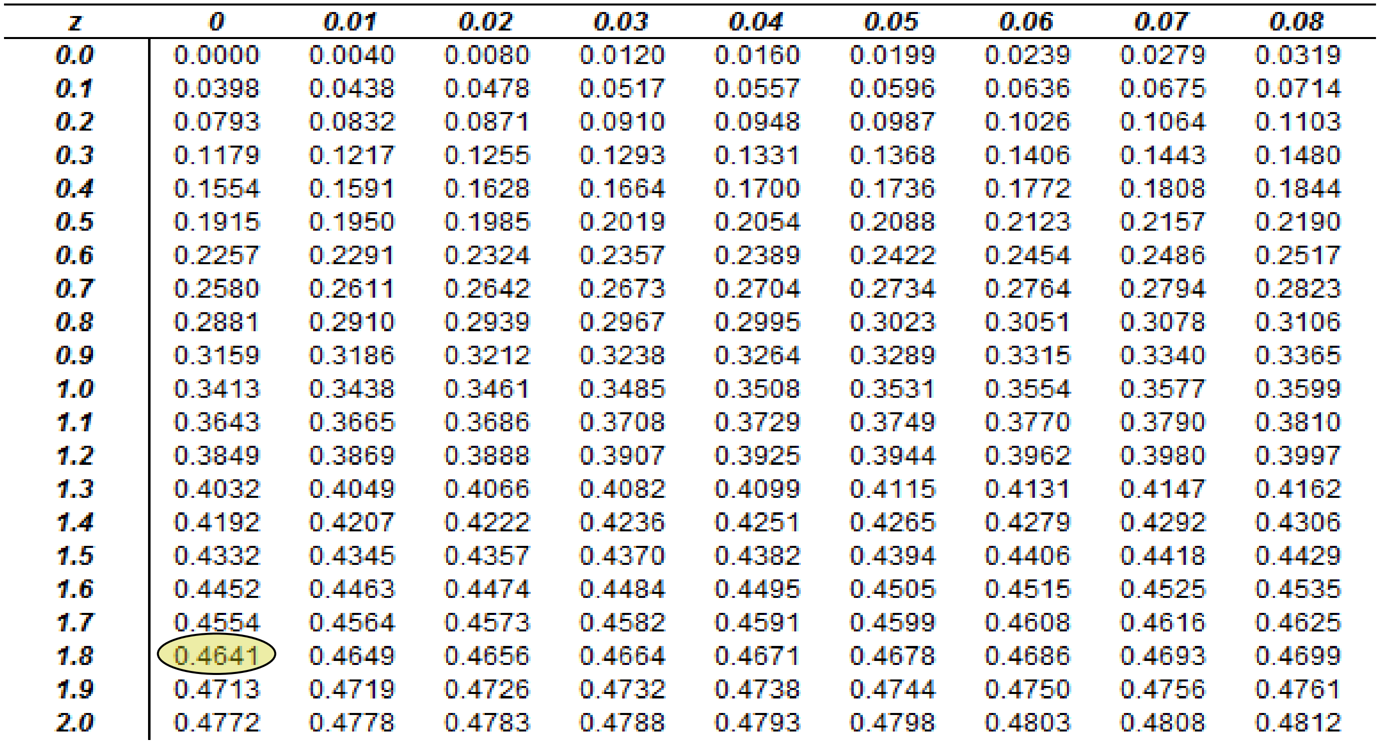


Using the Standard Normal Distribution Table in Finding Probability – Example

$P(z \geq 1.80) = ?$

Areas under the normal curve, 0 to z





z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812

LO 8-7

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.