

Sampling Methods and the Central Limit Theorem

Chapter 8

Learning Objectives

- LO8-1** Explain why populations are sampled and describe four methods to sample a population
- LO8-2** Define sampling error
- LO8-3** Demonstrate the construction of a sampling distribution of the sample mean
- LO8-4** Recite the central limit theorem and define the mean and standard error of the sampling distribution of the sample mean
- LO8-5** Apply the central limit theorem to calculate probabilities

Reasons for Sampling a Population

- ▶ The results of a sample may adequately estimate the value of the population parameter, saving time and money
- ▶ It may be too time-consuming to contact all members of the population
- ▶ It may be impossible to check or locate all the members of the population
- ▶ The cost of studying all the items in the population may be prohibitive
- ▶ Often testing destroys the sampled item and it cannot be returned to the population

Probability Sampling Methods

- ▶ In a simple random sample, all members of the population have the same chance of being selected for the sample
- ▶ In a systematic sample, a random starting point is selected, and then every k th item thereafter is selected for the sample
- ▶ In a stratified sample, the population is divided into several groups, called strata, and then a random sample is selected from each stratum
- ▶ In clustered sampling, the population is divided into primary units, then samples are drawn from the primary units

Simple Random Sampling

- ▶ The most widely used method of sampling is a simple random sample

SIMPLE RANDOM SAMPLE A sample selected so that each item or person in the population has the same chance of being selected.

- ▶ Example
- ▶ There were 750 Major League Baseball players at the end of the 2016 season. A committee of 10 players is to be formed to study the issue of concussions. To make sure every player has an equal chance of being selected, write each name on a piece of paper, place the names in a box and mix them up, then draw 10 names.

Using a Table of Random Numbers

Suppose the population of interest is the 750 Major League Baseball players on the active rosters of the 30 teams at the end of the 2016 season. A committee of 10 players is to be formed to study the issue of concussions. To make sure every player has an equal chance of being selected, use a table of random numbers.

1. Prepare a list of all the players and number them 1 through 750
2. Randomly pick a starting place in the random number table
3. Select 10 three-digit numbers between 1 and 750

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 player			Third player

Systematic Random Sampling

- ▶ If you do not have a list of the entire population to begin with, you can use the systematic random sample

SYSTEMATIC RANDOM SAMPLE A random starting point is selected, and then every k th member of the population is selected.

- ▶ **Example**
 - ▶ Stood's Grocery Store wants to study the length of time customers spend in their store
 - ▶ Randomly select the days of the week, the times, and the starting point of the study, then systematically select the customers and measure the time each spends in the store

Stratified Random Sampling

- ▶ When the population can be divided into groups based on some characteristic, use stratified random sampling

STRATIFIED RANDOM SAMPLE A population is divided into subgroups, called strata, and a sample is randomly selected from each stratum.

- ▶ Example
- ▶ A study of 50 of the 352 largest US firms' ad spending
- ▶ Begin by identifying the strata, then use random sampling within each group based on relative frequencies to collect the sample

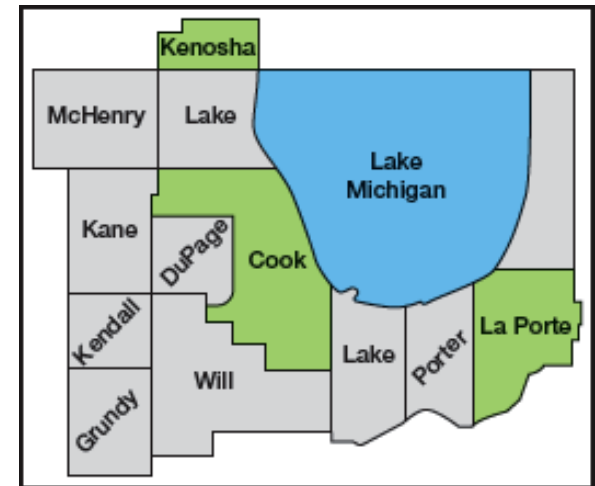
Stratum	Profitability (return on equity)	Number of Firms	Relative Frequency	Number Sampled
1	30 percent and over	8	0.02	1*
2	20 up to 30 percent	35	0.10	5*
3	10 up to 20 percent	189	0.54	27
4	0 up to 10 percent	115	0.33	16
5	Deficit	5	0.01	1
Total		352	1.00	50

Cluster Sampling

- ▶ Cluster sampling is a common type of sampling, used to reduce the cost of sampling over large geographic areas

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 we wish to sample residents of the 12 counties in the greater Chicago area about government policy. Randomly select 3 counties and then select a random sample of the residents in each of the 3 counties.



Exercise (1)

- ▶ Listed below are the 27 Nationwide Insurance agents in the Toledo, Ohio, metropolitan area. We would like to estimate the mean number of years employed with Nationwide

ID Number	Agent	ID Number	Agent	ID Number	Agent
00	Bly Scott 3332 W Laskey Rd	10	Heini Bernie 7110 W Centra	19	Riker Craig 2621 N Reynolds Rd
01	Coyle Mike 5432 W Central Av	11	Hinckley Dave 14 N Holland Sylvania Rd	20	Schwab Dave 572 W Dussel Dr
02	Denker Brett 7445 Airport Hwy	12	Joehlin Bob 3358 Navarre Av	21	Seibert John H 201 S Main
03	Denker Rollie 7445 Airport Hwy	13	Keisser David 3030 W Sylvania Av	22	Smithers Bob 229 Superior St
04	Farley Ron 1837 W Alexis Rd	14	Keisser Keith 5902 Sylvania Av	23	Smithers Jerry 229 Superior St
05	George Mark 7247 W Central Av	15	Lawrence Grant 342 W Dussel Dr	24	Wright Steve 105 S Third St
06	Gibellato Carlo 6616 Monroe St	16	Miller Ken 2427 Woodville Rd	25	Wood Tom 112 Louisiana Av
07	Glemser Cathy 5602 Woodville Rd	17	O'Donnell Jim 7247 W Central Av	26	Yoder Scott 6 Willoughby Av
08	Green Mike 4149 Holland Sylvania Rd	18	Priest Harvey 5113 N Summit St		
09	Harris Ev 2026 Albon Rd				

- a. We want to select a random sample of four agents. The random numbers are: 02, 59, 51, 25, 14, 29, 77, 69, and 18. Which dealers would be included in the sample?
- b. Use the table of random numbers to select your own sample of four agents.
- c. A sample is to consist of every seventh dealer. The number 04 is selected as the starting point. Which agents will be included in the sample?

Sampling Error

- ▶ It is unlikely the mean of a sample will be exactly equal to the mean of the population

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

- ▶ Example
- ▶ The Foxtrot Inn's number of rooms rented in June. The mean number of rooms rented, μ , is 3.13
- ▶ Taking three random samples of size 5, we find sample means, $\bar{x}$, of 3.80, 3.40 and 1.80. The sampling error is the difference between each $\bar{x}$ and μ

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

Sampling Distribution of the Sample Mean

- ▶ How can we determine how accurate the sample mean is?

SAMPLING DISTRIBUTION OF THE SAMPLE MEAN A probability distribution of all possible sample means of a given sample size.

- ▶ For a given sample size, the mean of all possible sample means selected from a population is equal to the population mean
- ▶ There is less variation in the distribution of the sample mean than in the population distribution
- ▶ The sampling distribution of the sample mean tends to become bell-shaped

Sampling Distribution Example

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

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			\$14				Jan			14	
	Sam			14				Art			16	
	Sue			16				Ted			18	
	Bob			16								

Sampling Distribution Example

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

I. What is the population mean?

$$\mu = \frac{\sum x}{N} = \$15.43$$

Employee			Hourly Earnings			Employee			Hourly Earnings		
	Joe			\$14			Jan			14	
	Sam			14			Art			16	
	Sue			16			Ted			18	
	Bob			16							

Sampling Distribution Example

2. What is the sampling distribution of the sample mean for samples of size 2?

Sample	Employees	Hourly Earnings	Sum	Mean	Sample	Employees	Hourly Earnings	Sum	Mean
1	Joe, Sam	\$14,\$14	\$28	\$14	12	Sue, Bob	16,16	32	16
2	Joe, Sue	14,16	30	15	13	Sue, Jan	16,14	30	15
3	Joe, Bob	14,16	30	15	14	Sue, Art	16,16	32	16
4	Joe, Jan	14,14	28	14	15	Sue, Ted	16,18	34	17
5	Joe, Art	14,16	30	15	16	Bob, Jan	16,14	30	15
6	Joe, Ted	14,18	32	16	17	Bob, Art	16,16	32	16
7	Sam, Sue	14,16	30	15	18	Bob, Ted	16,18	34	17
8	Sam, Bob	14,16	30	15	19	Jan, Art	14,16	30	15
9	Sam, Jan	14,14	28	14	20	Jan, Ted	14,18	32	16
10	Sam, Art	14,16	30	15	21	Art, Ted	16,18	34	17
11 8-15	Sam, Ted	14,16	32	16					

Sampling Distribution Example

Tartus Industries has seven production employees (the population).

3. What is the mean of the sampling distribution?

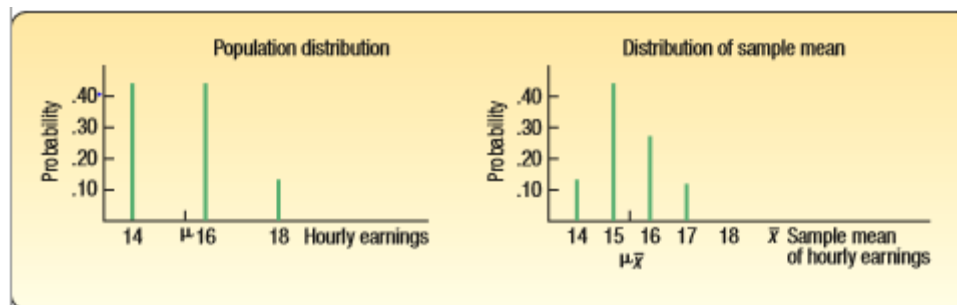
Sample Mean			Number of Means			Probability		
	\$14			3			.1429	
	15			9			.4285	
	16			6			.2857	
	17			3			.1429	
				21			1.0000	

$$\mu_{\bar{x}} = \frac{\text{Sum of all sample means}}{\text{Total number of samples}} = \frac{\$324}{21} = \$15.43$$

Sampling Distribution Example

4. What observations can be made about the population and the sampling distribution?

- ▶ The mean of the distribution of the sample mean (\$15.43) is equal to the mean of the population, $\mu = \mu_{\bar{X}}$
- ▶ The spread in the distribution of the sample mean is less than the spread in the population values
- ▶ The shapes of the population and sample distributions are different



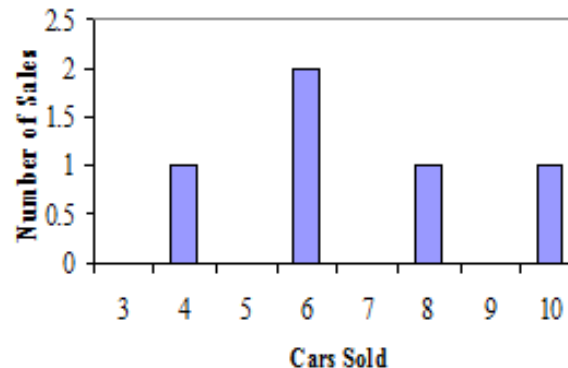
Exercise 2

- There are five sales associates at Mid-Motors Ford. The five representatives and the number of cars they sold last week are:

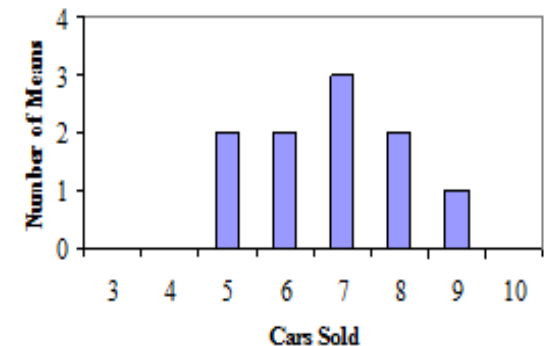
Sales Representative	Cars Sold
Peter Hankish	8
Connie Stallter	6
Juan Lopez	4
Ted Barnes	10
Peggy Chu	6

- How many different samples of size 2 are possible?
- List all possible samples of size 2, and compute the mean of each sample.
- Compare the mean of the sampling distribution of sample means with that of the population.
- On a chart, compare the dispersion in sample means with that of the population

Population Means



Sample Means



The number of combinations **C** is:

number = n

Number_chosen = k

$$\binom{n}{k} = \frac{P_{k,n}}{k!} = \frac{n!}{k!(n-k)!}$$

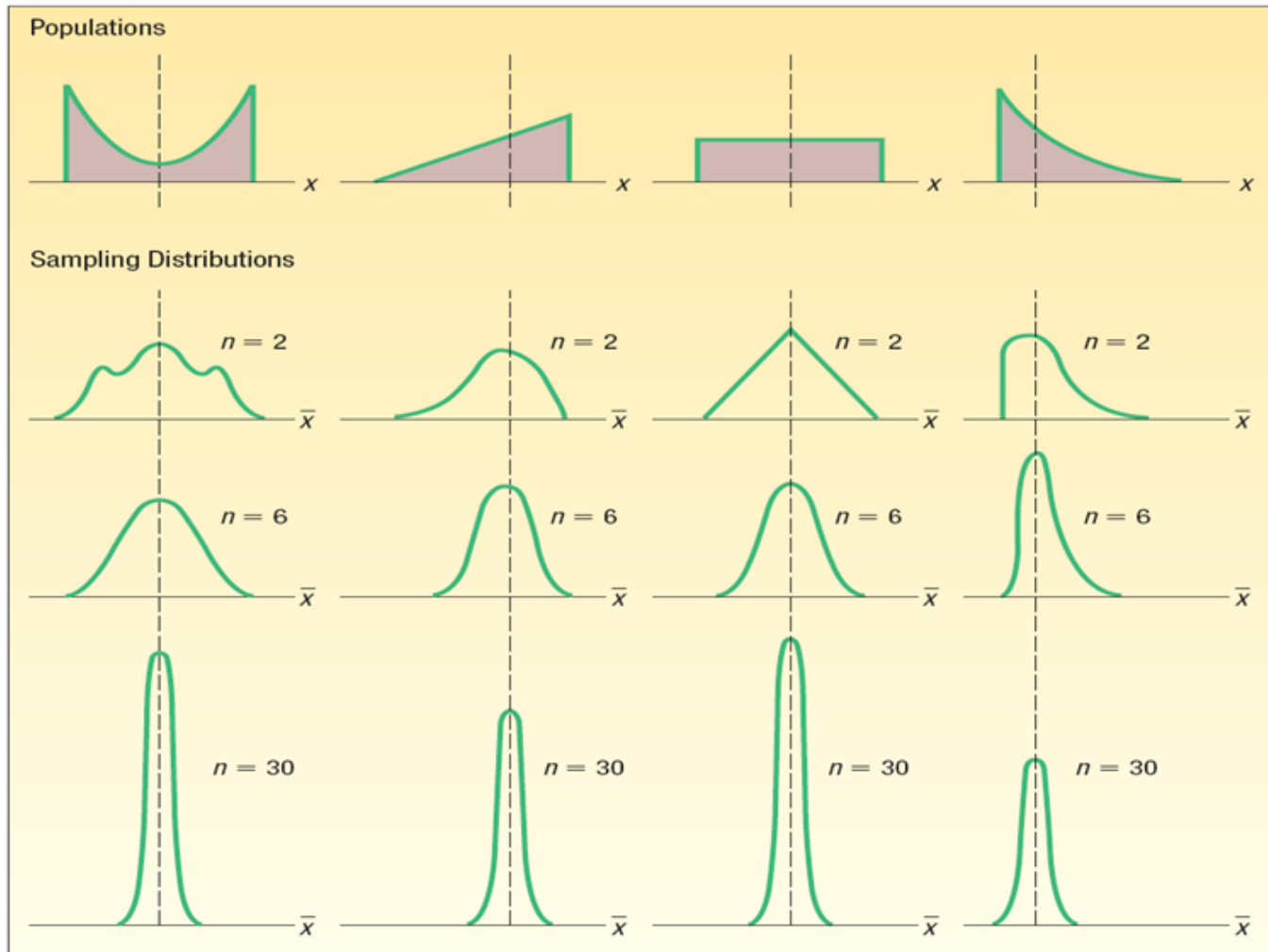
$$P_{k,n} = \frac{n!}{(n-k)!}$$

Central Limit Theorem

THE CENTRAL LIMIT THEOREM If samples of a particular size are selected from any population, the sampling distribution of the sample mean is approximately a normal distribution. The 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, you will see the normal shape of the distribution of the sample mean emerge with samples as small as 10
- ▶ If the distribution is skewed or has thick tails, it may require samples of 30 or more to observe the normality feature

Central Limit Theorem Results



CLT Conclusions

- ▶ The mean of the distribution of sample means will be exactly equal to the population mean, if we select all possible samples of same size from the population

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

- ▶ The standard deviation of the sampling distribution of the sample mean is also called the standard error of the mean

STANDARD ERROR OF THE MEAN

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

[8-1]

- ▶ There will be less dispersion in the sampling distribution of the sample mean, $\sigma/\sqrt{n}$, than in the population σ

Normal Distribution

- ▶ If the population follows a normal distribution, the sampling distribution of the sample mean will also follow the normal distribution for samples of any size
- ▶ If the population is not normally distributed, the sampling distribution of the sample mean will approach a normal distribution when the sample size is at least 30
- ▶ Assume the population standard deviation is known
- ▶ To determine the probability that a sample mean falls in a particular region, use the following formula

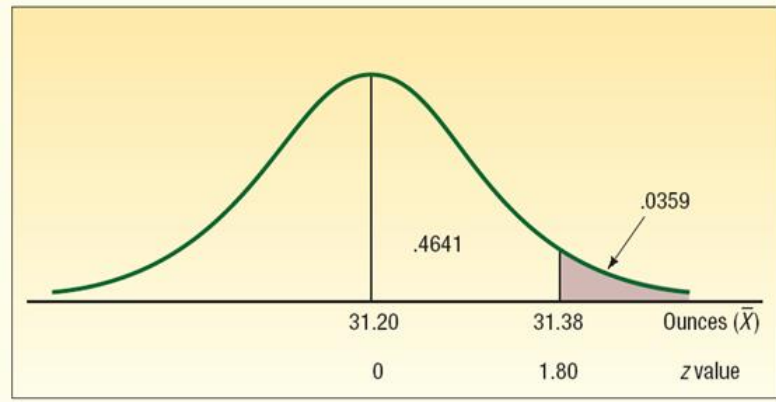
$$z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

Using the Sampling Distribution Example

The Quality Assurance Dept. for Cola, Inc. maintains records regarding the amount of cola in its jumbo bottle. The actual amount of cola in each bottle varies a small amount from one bottle to another. Records indicate the amounts of cola follow the normal distribution, the mean amount of cola in the bottles is 31.2 ounces, and the 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 was 31.38 ounces. Is this an unlikely result? Is it a likely the process is putting too much soda in the bottle? Is the sampling error of 0.18 ounce unusual?

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

We conclude that it is unlikely; there is less than a 4% chance. The process is putting too much soda in the bottles.

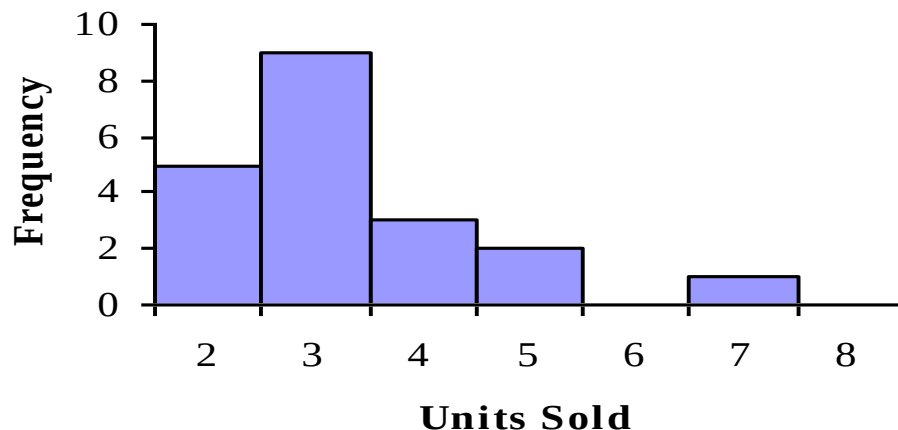


Exercise 3 (1 / 3)

Scraper Elevator Company has 20 sales representatives who sell its product throughout the United States and Canada. The number of units sold last month by each representative is listed below. Assume these sales figures to be the population values.

2 3 2 3 3 4 2 4 3 2 2 7 3 4 5 3 3 3 3 5

- Draw a graph showing the population distribution.
- Compute the mean of the population.



$$\mu = \frac{2 + 3 + \dots + 5}{20} = 3.3$$

Exercise 3 (2/3)

Scraper Elevator Company has 20 sales representatives who sell its product throughout the United States and Canada. The number of units sold last month by each representative is listed below. Assume these sales figures to be the population values.

2 3 2 3 3 4 2 4 3 2 2 7 3 4 5 3 3 3 3 5

- Select five random samples of 5 each. Compute the mean of each sample. Use the methods described in this chapter to determine the items to be included in the sample.
- Compare the mean of the sampling distribution of the sample means to the population mean. Would you expect the two values to be about the same?

Sample	Sample Values	Sum	$\bar{X}$
1	2,3,2,3,3	13	2.6
2	3,3,4,2,4	16	3.2
3	3,3,4,4,2	16	3.2
4	3,2,5,5,3	18	3.6
5	3,4,4,2,7	20	4.0

$$\mu_{\bar{X}} = \frac{2.6 + 3.2 + 3.2 + 3.6 + 4.0}{5} = 3.32$$

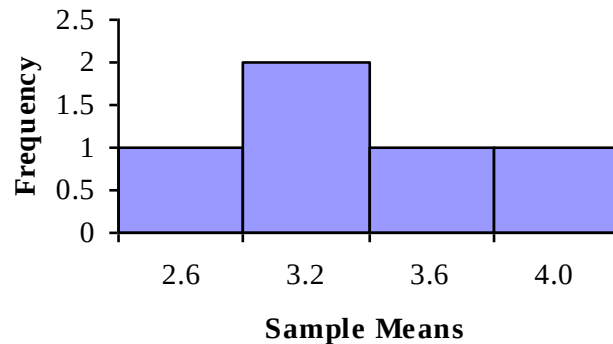
Sample mean is very close to the population mean. It is not to be expected that they are exact.

Exercise 3 (3/3)

Scrappier Elevator Company has 20 sales representatives who sell its product throughout the United States and Canada. The number of units sold last month by each representative is listed below. Assume these sales figures to be the population values.

2 3 2 3 3 4 2 4 3 2 2 7 3 4 5 3 3 3 3 5

- e. Draw a histogram of the sample means. Do you notice a difference in the shape of the distribution of sample means compared to the shape of the population distribution?



There is less dispersion in the sample means than the population

Exercise 4

According to an IRS study, it takes a mean of 330 minutes for taxpayers to prepare, copy, and electronically file a 1040 tax form. This distribution of times follows the normal distribution and the standard deviation is 80 minutes. A consumer watchdog agency selects a random sample of 40 taxpayers.

- a. What is the standard error of the mean in this example?
- b. What is the likelihood the sample mean is greater than 320 minutes?
- c. What is the likelihood the sample mean is between 320 and 350 minutes?
- d. What is the likelihood the sample mean is greater than 350 minutes?

$$s_{\bar{x}} = \frac{80}{\sqrt{40}} = 12.649$$

$$Z = \frac{320 - 330}{80 / \sqrt{40}} = -0.79$$

So probability is 0.7852, found by $0.2852 + 0.5$

$$Z = \frac{350 - 330}{80 / \sqrt{40}} = 1.58$$

So probability is 0.7281, found by $0.2852 + 0.4429$

0.0571, found by $0.5000 - 0.4429$