Chapter 7

Sample Variability

Chapter 7 Overview

- 7-1 Sampling Distributions
- 7-2 The Sampling Distribution of Sample Means
- 7-3 Application of The Sampling Distribution of Sample Means

7-1 Sampling Distributions

- Samples are taken every day for many reasons. Industries monitor their products continually to be sure of their quality, agencies monitor our environment, medical professionals monitor our health; the list is limitless.

- Many of these samples are one-time samples, while many are samples that are repeated for ongoing monitoring.

Population Sampling

- A census, a 100% survey or sampling, in the United States is done only every 10 years. It is an enormous and overwhelming job, but the information that is obtained is vital to our country's organization and structure.

- Issues come up and times change; information is needed and a census is impractical. This is where representative and everyday samples come in.

Sample Variability

- Suppose you have the following population:

  {2, 3, 4, 5, 8, 11}
- What is this population's:

  μ =_______   σ =_______ N=_______
- Suppose you select a sample of size n=2 from this population. You might get the 2 and the 3 or maybe the 2 and the 4. In the table below, complete the list of all possible samples of size n=2 and their sample means:

Sample Variability

TABLE 1

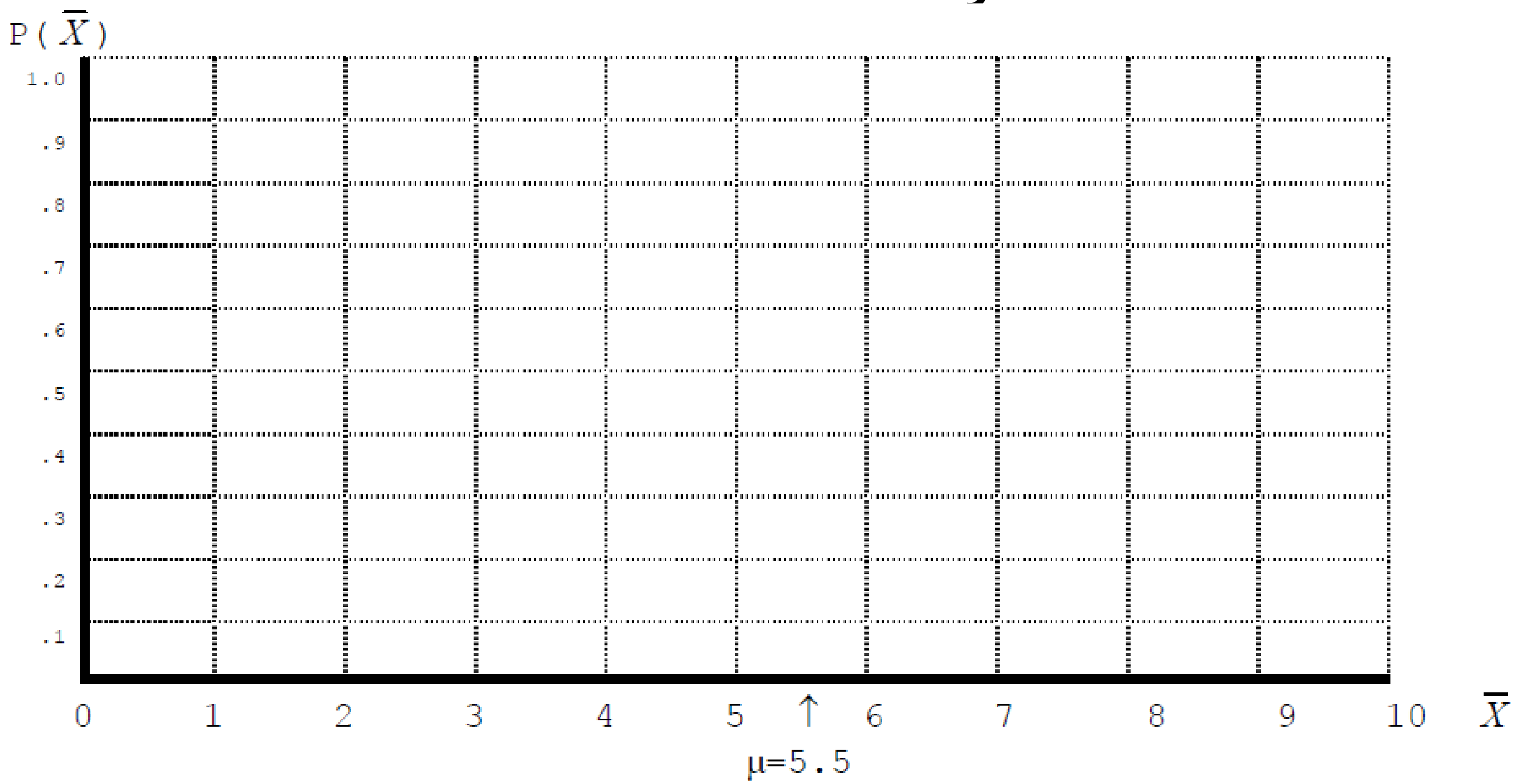
items in sample	$\overline{X}$
2, 3	2.5
2, 4	3.0

Sample Variability

- What is the probability of selecting the 2 and the 3?
- What is the probability of selecting the 2 and the 4?
- In fact, each of the sample combinations has what probability of being the one selected?

Sample Variability

- Graph the possible values for X when n=2. Notice the axes are already labeled.



Sample Variability

- Is it possible to select a sample of size 2 from the above population and obtain a sample mean whose value is close to the true value of the population mean?
- Is it possible to select a biased sample whose mean differs from the true value of the population mean?

Sample Variability

- Suppose a sample of size n=5 is to be taken instead. List all the possible samples of size n=5 and their respective sample means. [HINT: there will be 6.]

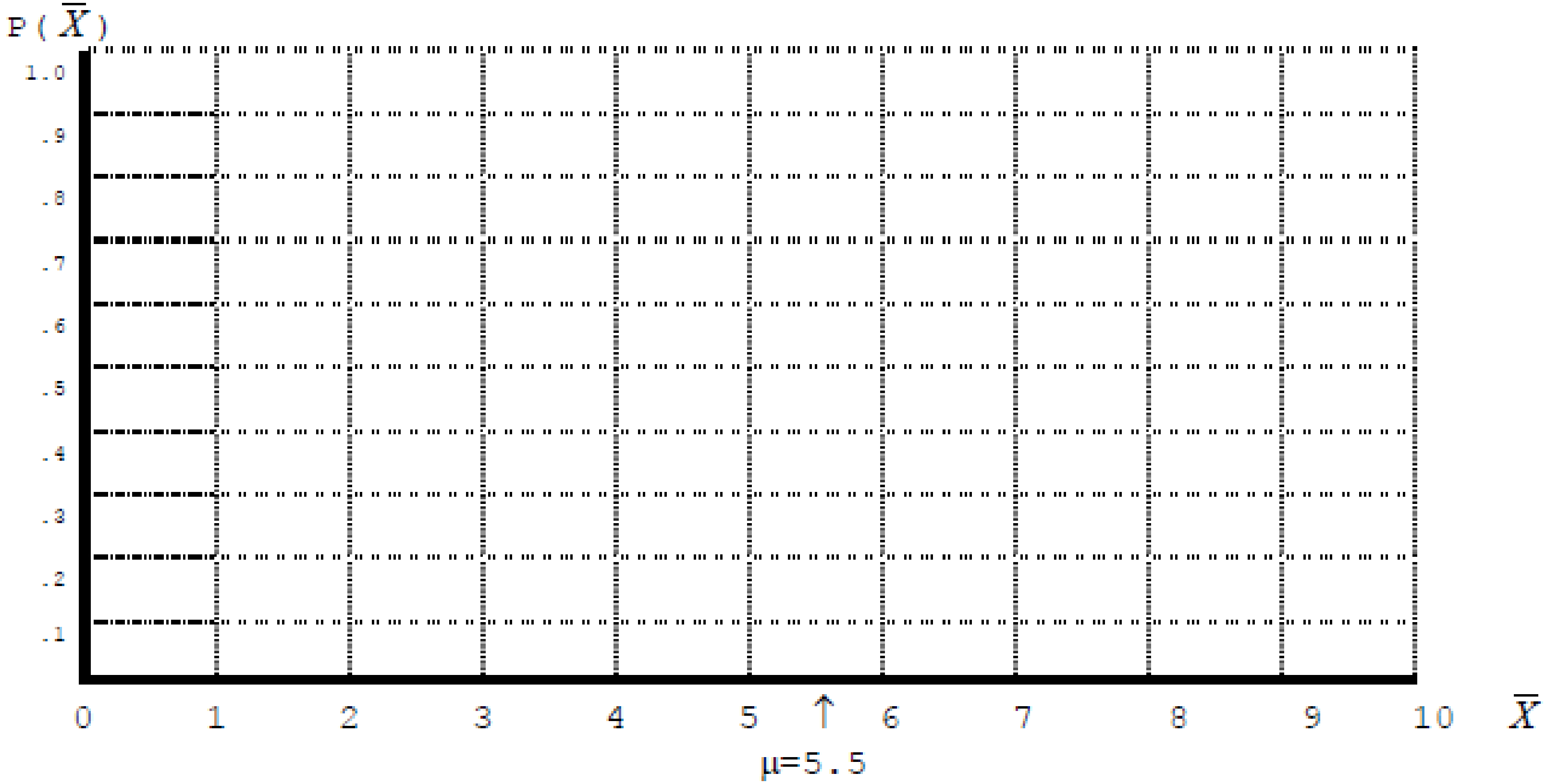
TABLE 2

items in sample	sample's mean

- Each of these sample combinations has what probability of being the one selected?

Sample Variability

- Graph the possible values for the sample mean when n=5.



Sample Variability

- Is it possible to select a sample of size 5 from the above population and obtain a sample mean whose value is close to the true value of the population mean?
- Is it possible to select a biased sample whose mean differs from the true value of the population mean?

Sample Variability

- The grand mean is defined as the mean of the possible values for the sample mean. Calculate the grand mean now by computing the mean of the values in the sample's mean column in Table 2 above.
- Compute the standard deviation of the values in the sample's mean column in Table 2 above. Use the formula for a population standard deviation and use the grand mean as the value for μ.

Sample Variability

- Suppose a sample of size n=6 is to be taken instead. List all the possible samples of size 6 and their respective sample means. ([HINT: there will be only 1.]

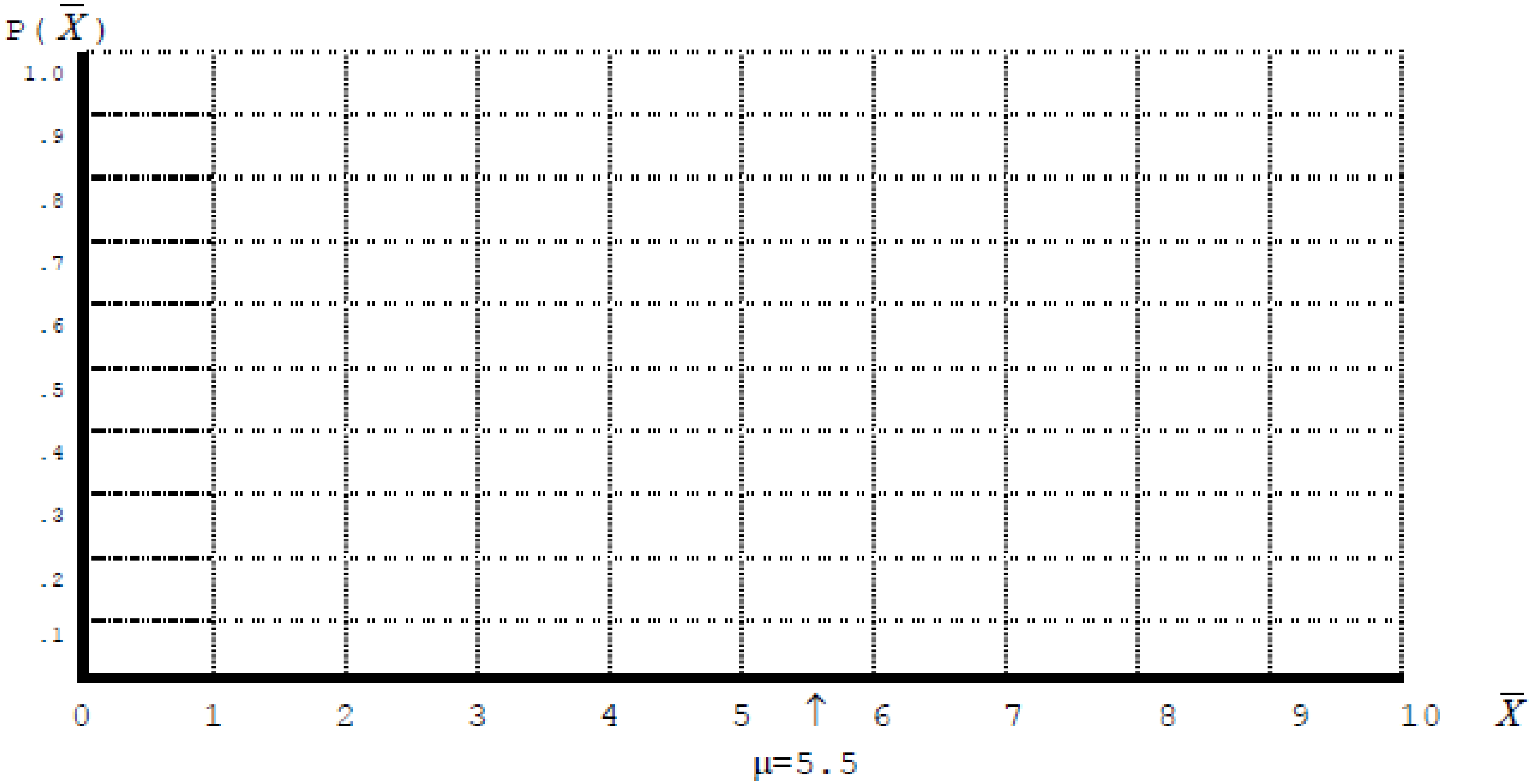
TABLE 3

items in sample	sample's mean

- This sample combination has what probability of being selected?

Sample Variability

- Graph the possible values for the sample mean when n=6.



Sample Variability

- Is it possible to select a sample of size 6 from the above population and obtain a sample mean whose value is close to the true value of the population mean?
- Is it possible to select a biased sample whose mean differs from the true value of the population mean?
- Which sample size (n=2, n=5 or n=6) has a higher risk of sampling error? Explain.

Sample Variability

- What could be done next is to make a *Sampling Distribution of Sample Means* from any of the samples we took (n=2, n=5, n=6) by recording all the possible means for each item in the sample verses the number of times they occurred.
- What happens when we do this?
  - We get a distribution that is very close to the normal distribution with a mean that is the same as the original populations mean.
  - In fact as we increase our sample size, the distribution becomes almost identically normal.

Sampling Distributions

- **Note**

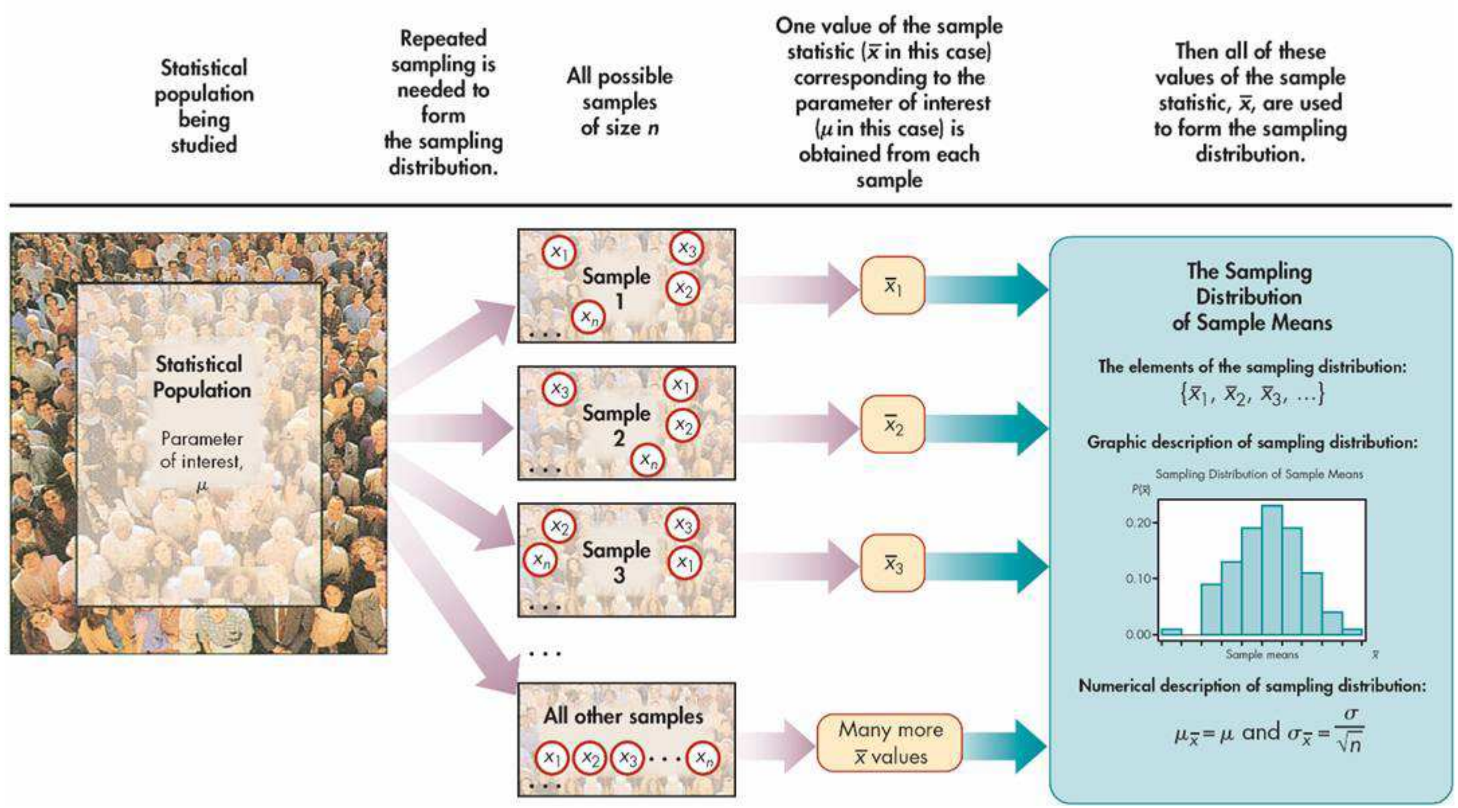
The variable for the sampling distribution is $\overline{x}$; therefore, the mean of the $\overline{x}$'s is $\overline{\overline{x}}$ and the standard deviation of $\overline{x}$ is $s_{\overline{x}}$.

- The theory involved with sampling distributions that will be described in the remainder of this chapter requires *random sampling*.

- Random sample A sample obtained in such a way that each possible sample of fixed size *n* has an equal probability of being selected. Figure 7.5 shows how the sampling distribution of sample means is formed.

Sampling Distributions

- Figure 7.5 shows how the sampling distribution of sample means is formed.



The Sampling Distribution of Sample Means
Figure 7.5

7-2 The Sampling Distribution of Sample Means

- Previously, we discussed the sampling distributions of sample means.

- Many others could be discussed; however, the only sampling distribution of concern to us at this time is the sampling distribution of sample means.

The Sampling Distribution of Sample Means

- Sampling distribution of sample means (SDSM) If all possible random samples, each of size *n*, are taken from any population with mean μ and standard deviation σ, then the sampling distribution of sample means will have the following:

- 1. A mean $\mu_{\overline{x}}$ equal to μ
- 2. A standard deviation $\sigma_{\overline{x}}$ equal to $\frac{\sigma}{\sqrt{n}}$

- Furthermore, if the sampled population has a normal distribution, then the sampling distribution of $\overline{x}$ will also be normal for samples of all sizes.

The Sampling Distribution of Sample Means

- Standard error of the mean ($\sigma_{\bar{x}}$) The standard deviation of the sampling distribution of sample means.

- Central limit theorem (CLT) The sampling distribution of sample means will more closely resemble the normal distribution as the sample size increases.

- If the sampled distribution is normal, then the sampling distribution of sample means (SDSM) is normal, as stated previously, and the central limit theorem (CLT) is not needed.

The Sampling Distribution of Sample Means

- But, if the sampled population is not normal, the CLT tells us that the sampling distribution will still be approximately normally distributed under the right conditions.

$\overline{x}$

- If the sampled population distribution is nearly normal, the distribution is approximately normal for fairly small *n* (possibly as small as 15).

- When the sampled population distribution lacks symmetry, *n* may have to be quite large (maybe 50 or more) before the normal distribution provides a satisfactory approximation.

The Sampling Distribution of Sample Means

- By combining the preceding information, we can describe the sampling distribution of $\overline{x}$ completely:

(1) The location of the center (mean),

(2) A measure of spread indicating how widely the distribution is dispersed (standard error of the mean), and

(3) an indication of how it is distributed.

The Sampling Distribution of Sample Means

- 1. $\mu_{\overline{x}} = \mu$; the mean of the sampling distribution ($\mu_{\overline{x}}$) is equal to the mean of the population (μ).
- 2. $\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$; the standard error of the mean ($\sigma_{\overline{x}}$) is equal to the standard deviation of the population (σ) divided by the square root of the sample size, *n*.
- 3. The distribution of sample means is normal when the parent population is normally distributed, and the CLT tells us that the distribution of sample means becomes approximately normal (regardless of the shape of the parent population) when the sample size is large enough.

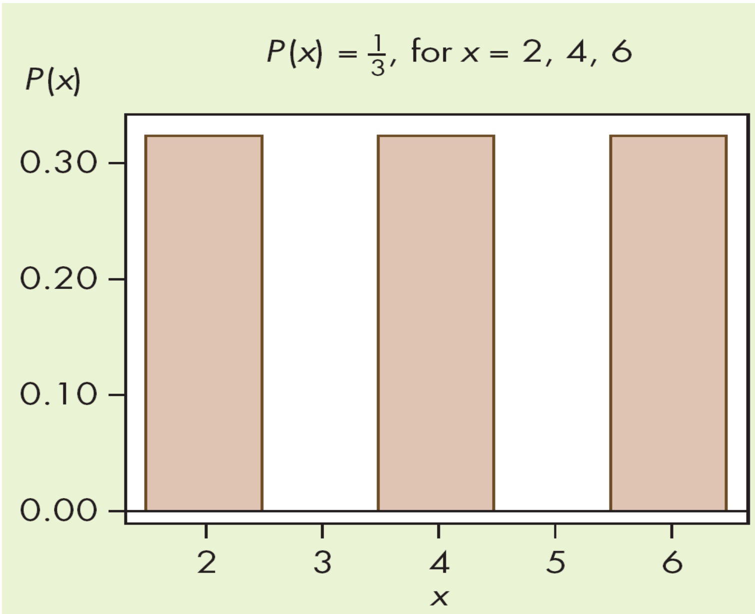
The Sampling Distribution of Sample Means

Note

- The *n* referred to is the size of each sample in the sampling distribution. (The number of repeated samples used in an empirical situation has no effect on the standard error.)

Example 2 – *Constructing a Sampling Distribution of Sample Means*

- Let's consider all possible samples of size 2 that could be drawn from a population that contains the three numbers 2, 4, and 6. First let's look at the population itself. Construct a histogram to picture its distribution, Figure 7.6;



Population

Figure 7.6

Example 2 – *Constructing a Sampling Distribution of Sample Means* cont'd

- Calculate the mean, μ , and the standard deviation, σ, Table 7.4.

x	$P(x)$	$xP(x)$	$x^2P(x)$
2	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{4}{3}$
4	$\frac{1}{3}$	$\frac{4}{3}$	$\frac{16}{3}$
6	$\frac{1}{3}$	$\frac{6}{3}$	$\frac{36}{3}$
Σ	$\frac{3}{3}$ ck	$\frac{12}{3}$	$\frac{56}{3}$
	1.0	4.0	$18.6\overline{6}$

$\mu = \mathbf{4.0}$

$$\sigma = \sqrt{18.6\overline{6} - (4.0)^2} = \sqrt{2.6\overline{6}} = \mathbf{1.63}$$

Extensions Table for *x*

Table 7.4

Example 2 – *Constructing a Sampling Distribution of Sample Means* cont'd

- Table 7.5 lists all the possible samples of size 2 that can be drawn from this population. (One number is drawn, observed, and then returned to the population before the second number is drawn.)

- Table 7.5 also lists the means of these samples.

Sample	$\bar{x}$	Sample	$\bar{x}$	Sample	$\bar{x}$
2,2	2	4,2	3	6,2	4
2,4	3	4,4	4	6,4	5
2,6	4	4,6	5	6,6	6

All Nine Possible Samples of Size 2

Table 7.5

Example 2 – *Constructing a Sampling Distribution of Sample Means* cont'd

- The sample means are then collected to form the sampling distribution.

- The distribution for these means and the extensions are given in Table 7.6 along with the calculation of the mean and the Standard error of the mean for the sampling distribution.

$\bar{x}$	$P(\bar{x})$	$\bar{x}P(\bar{x})$	$\bar{x}^2P(\bar{x})$
2	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{4}{9}$
3	$\frac{2}{9}$	$\frac{6}{9}$	$\frac{18}{9}$
4	$\frac{3}{9}$	$\frac{12}{9}$	$\frac{48}{9}$
5	$\frac{2}{9}$	$\frac{10}{9}$	$\frac{50}{9}$
6	$\frac{1}{9}$	$\frac{6}{9}$	$\frac{36}{9}$
Σ	$\frac{9}{9}$ ck	$\frac{36}{9}$	$\frac{156}{9}$
	1.0	4.0	$17.3\overline{3}$

$\mu_{\bar{x}} = \mathbf{4.0}$

$\sigma_{\bar{x}} = \sqrt{17.3\overline{3} - (4.0)^2} = \sqrt{1.3\overline{3}} = \mathbf{1.15}$

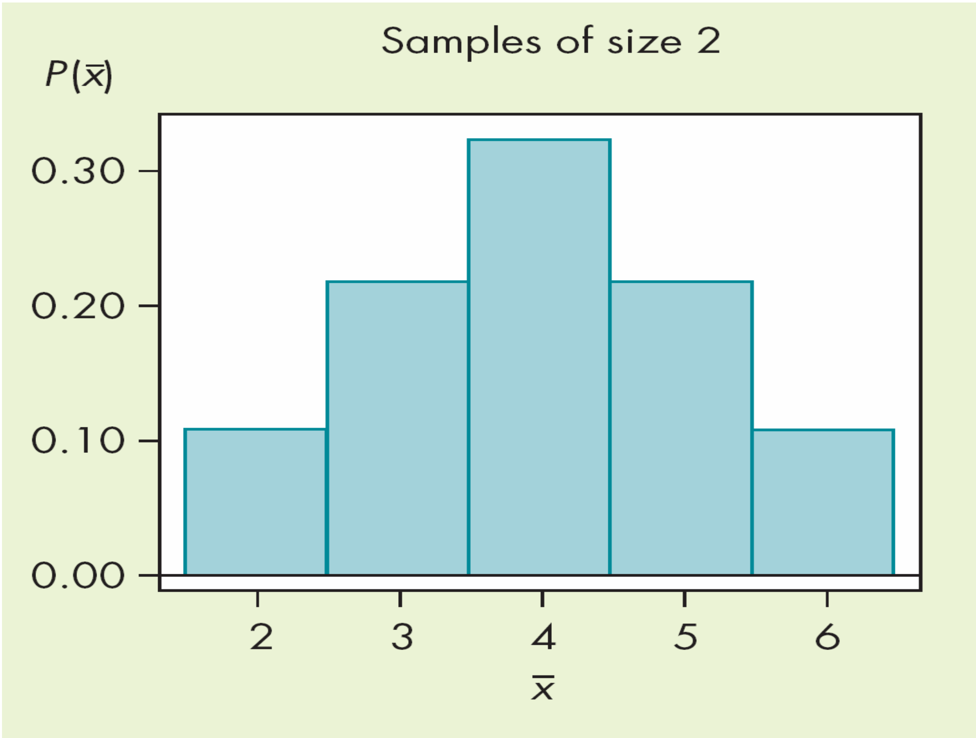
Extensions Table for $\bar{x}$

Table 7.6

Example 2 – *Constructing a Sampling Distribution of Sample Means*

cont'd

- The histogram for the sampling distribution of sample means is shown in Figure 7.7.



Sampling Distribution of Sample Means

Figure 7.7

Example 2 – *Constructing a Sampling Distribution of Sample Means* cont'd

- Let's now check the truth of the three facts about the sampling distribution of sample means:

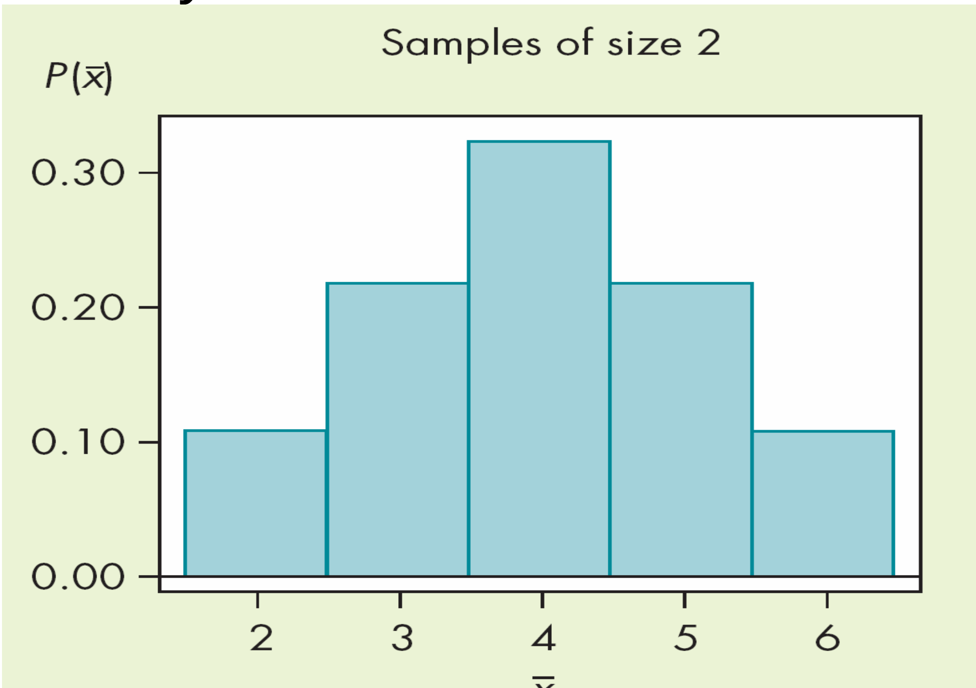
1. The mean of $\mu_{\overline{x}}$ the sampling distribution will equal the mean μ of the population: both μ and $\mu_{\overline{x}}$ have the value **4.0**.

2. The standard error of the mean $\sigma_{\overline{x}}$ for the sampling distribution will equal the standard deviation σ of the population divided by the square root of the sample size, n: $\sigma_{\overline{x}}$= **1.15** and σ = 1.63, n = 2, $\frac{\sigma}{\sqrt{n}} = \frac{1.63}{\sqrt{2}} = \mathbf{1.15}$; they are equal: $\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$.

Example 2 – *Constructing a Sampling Distribution of Sample Means*

cont'd

- 3. The distribution will become approximately normally distributed: the histogram in Figure 7.7 very strongly suggests normality.

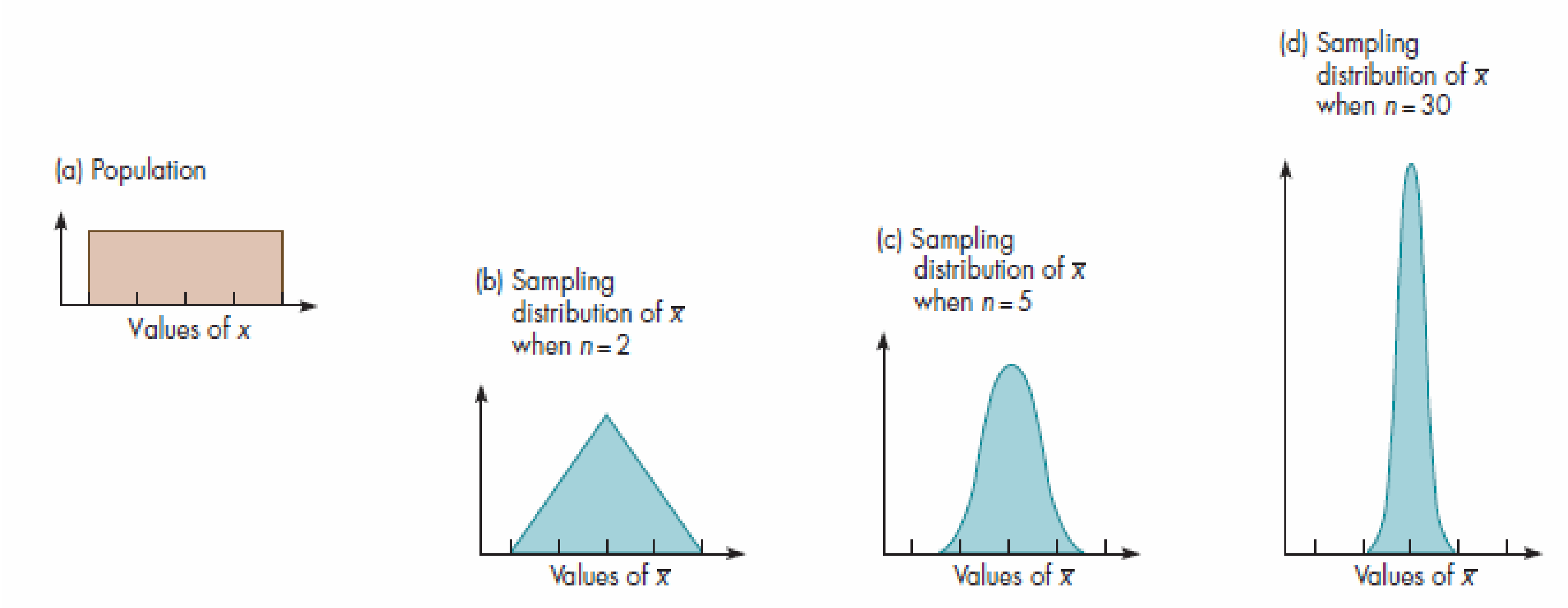


Sampling Distribution of Sample Means

Figure 7.7

The Sampling Distribution of Sample Means

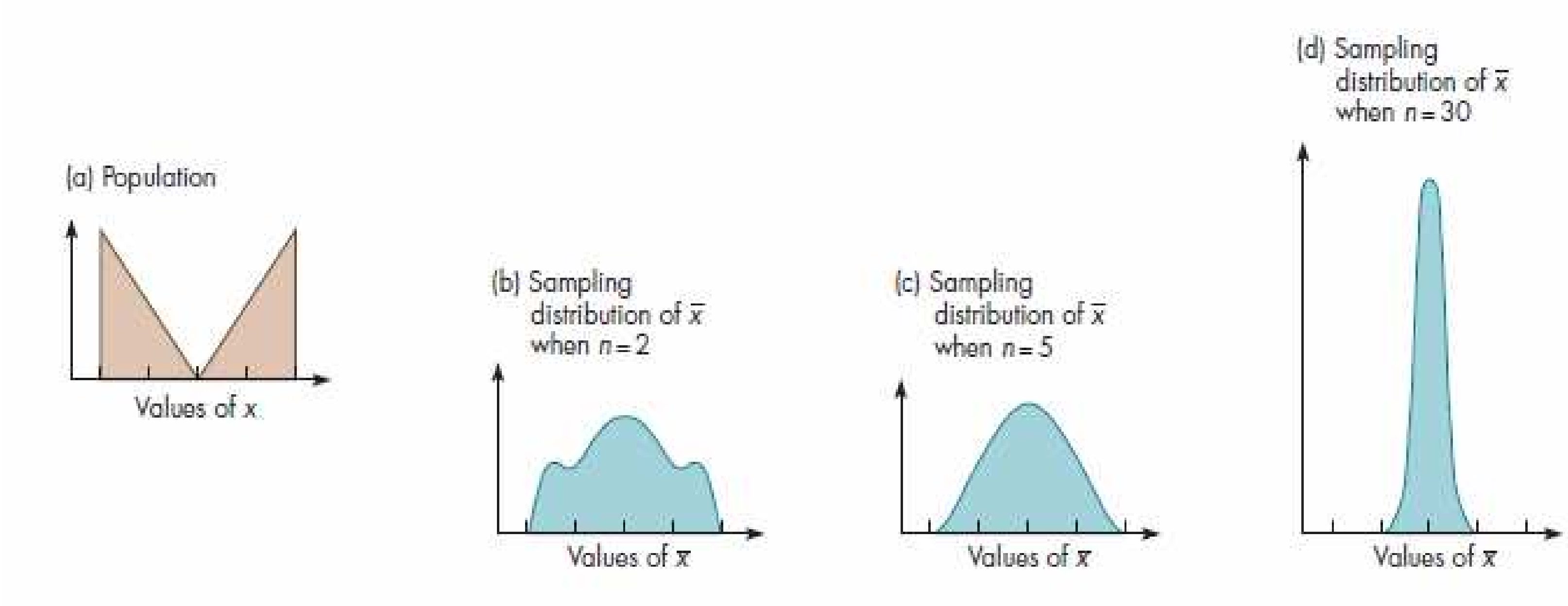
- In Figure 7.8 we have a uniform distribution, for the integer illustration, and the resulting distributions of sample means for samples of sizes 2, 5, and 30.



Uniform Distribution
Figure 7.8

The Sampling Distribution of Sample Means

- Figure 7.9 shows a U-shaped population and the three sampling distributions.

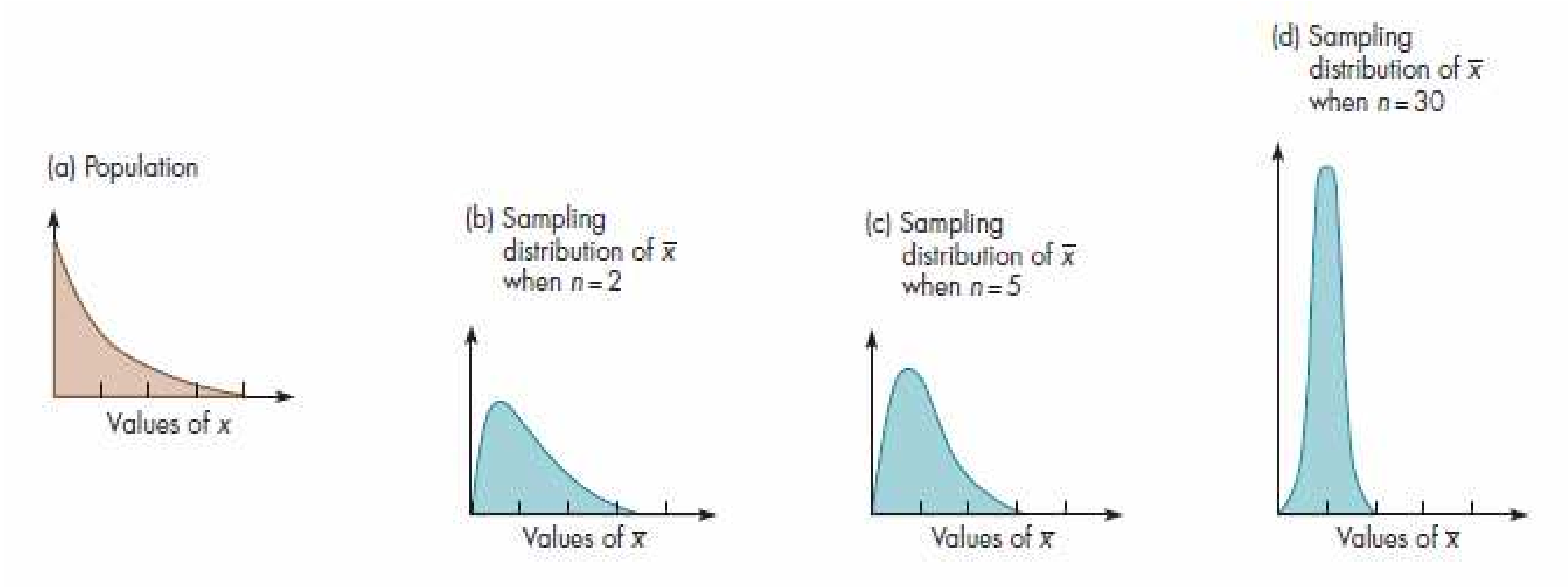


U-Shaped Distribution

Figure 7.9

The Sampling Distribution of Sample Means

- Figure 7.10 shows a J-shaped population and the sampling distributions.



J-Shaped Distribution

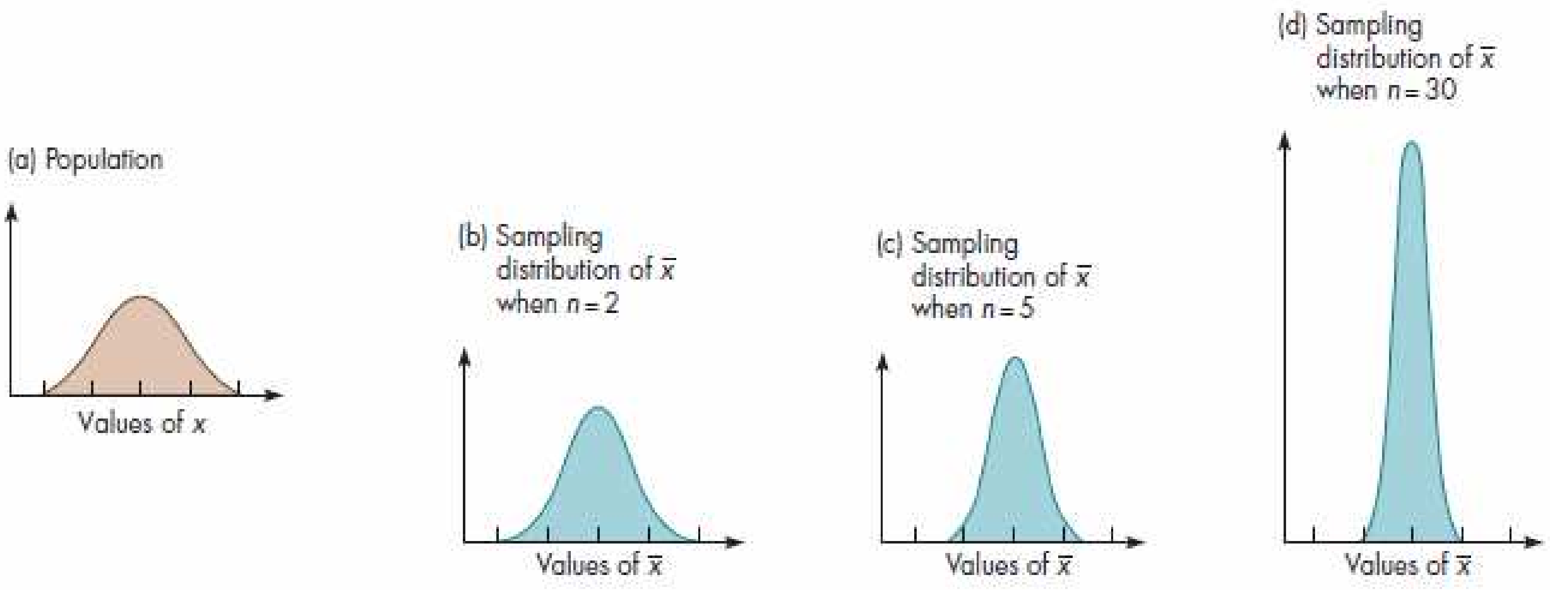
Figure 7.10

The Sampling Distribution of Sample Means

- All three nonnormal population distributions seem to verify the CLT; the sampling distributions of sample means appear to be approximately normal for all three when samples of size 30 are used.

The Sampling Distribution of Sample Means

- Now consider Figure 7.11, which shows a normally distributed population and the three sampling distributions.



Normal Distribution

Figure 7.11

The Sampling Distribution of Sample Means

- With the normal population, the sampling distributions of the sample means for all sample sizes appear to be normal.

- Thus, you have seen an amazing phenomenon: No matter what the shape of a population, the sampling distribution of sample means either is normal or becomes approximately normal when n becomes sufficiently large.

- You should notice one other point: The sample mean becomes less variable as the sample size increases. Notice that as n increases from 2 to 30, all the distributions become narrower and taller.

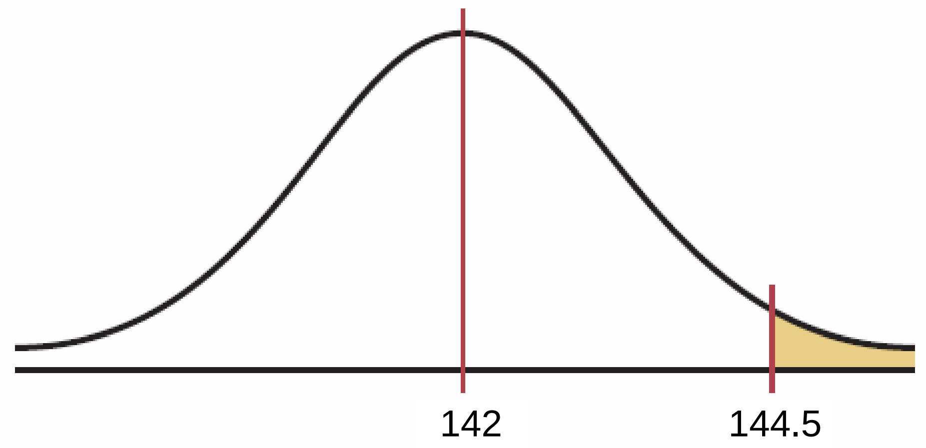
7-3 The Central Limit Theorem

- The central limit theorem can be used to answer questions about sample means in the same manner that the normal distribution can be used to answer questions about individual values.
- A new formula must be used for the *z* values:

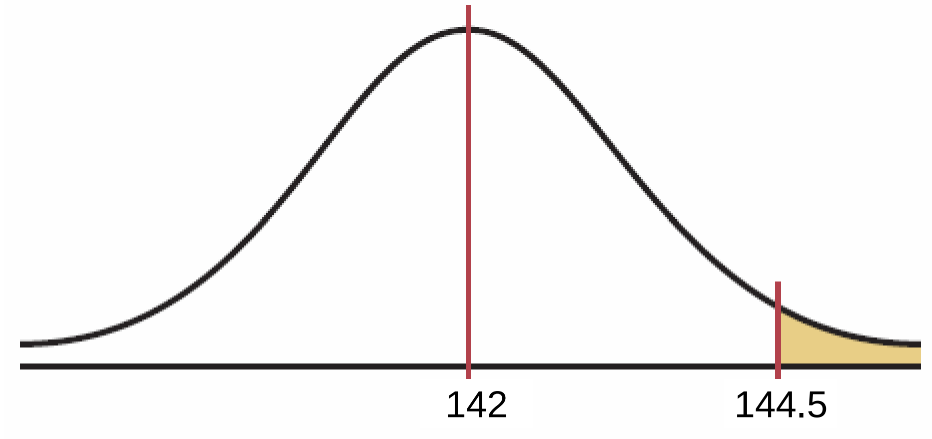
$$z = \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

Ex.) 3 – The Central Limit Theorem

Weights of 15-Year-Old Males The mean weight of 15-year-old males is 142 pounds, and the standard deviation is 12.3 pounds. If a sample of thirty-six 15-year-old males is selected, find the probability that the mean of the sample will be greater than 144.5 pounds. Assume the variable is normally distributed.



Ex.) 3 – The Central Limit Theorem



Since we are calculating probability for a sample mean, we need the Central Limit Theorem formula

$$z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = \frac{144.5 - 142}{12.3 / \sqrt{36}} = 1.22$$

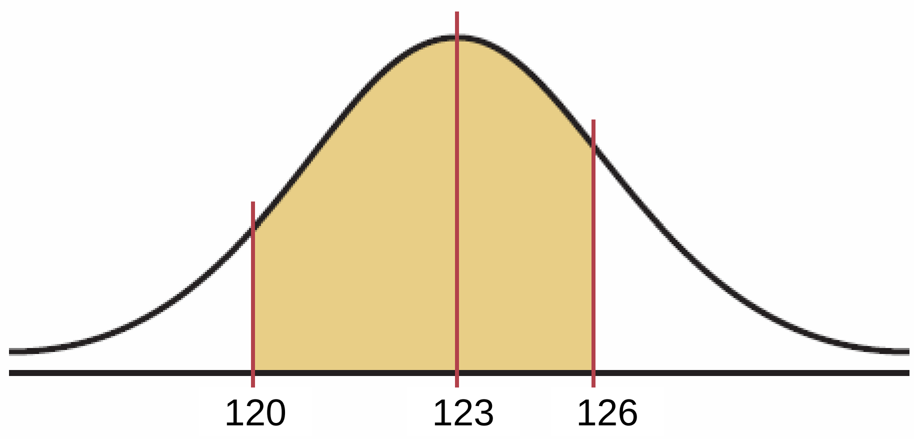
The area is 1.0000 – 0.8888 = 0.1112. The probability of obtaining a sample mean larger than 144.5 lbs is 1.11%.

The Central Limit Theorem

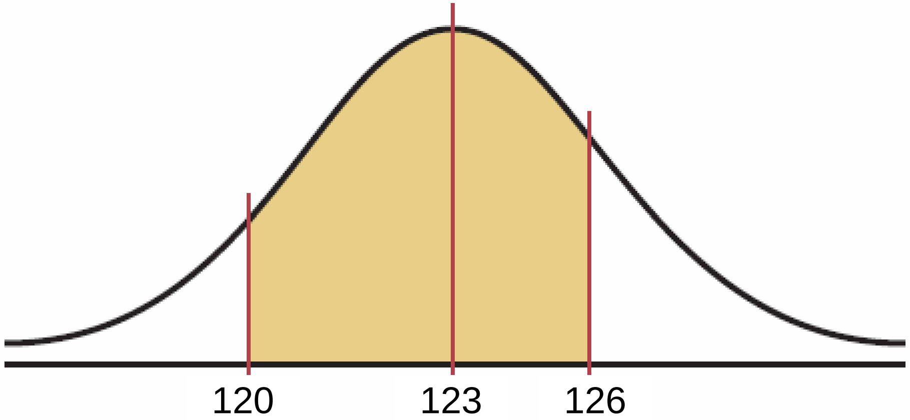
- The lifetime of light bulbs produced by a particular manufacturer have a mean of 1200 hours and a standard deviation of 400 hours. The population distribution can be regarded as normal. Suppose you purchase nine bulbs, which can be regarded as a random sample from the manufacturer’s output.
  - What is the mean of the sample means for light bulb lifetime?
  - What is the standard error of the sample mean?
  - What is the probability that, on average, those nine light bulbs have lives of less than 1150 hours?

Ex.) 4 – The Central Limit Theorem

Water Use The *Old Farmer's Almanac* reports that the average person uses 123 gallons of water daily. If the standard deviation is 21 gallons, find the probability that the mean of a randomly selected sample of 15 people will be between 120 and 126 gallons. Assume the variable is normally distributed.



Ex.) 4 – The Central Limit Theorem



$$z = \frac{120 - 123}{21/\sqrt{15}} = -0.55 \qquad z = \frac{126 - 123}{21/\sqrt{15}} = 0.55$$

Table 3 gives us areas 0.2912 and 0.7088, respectively.

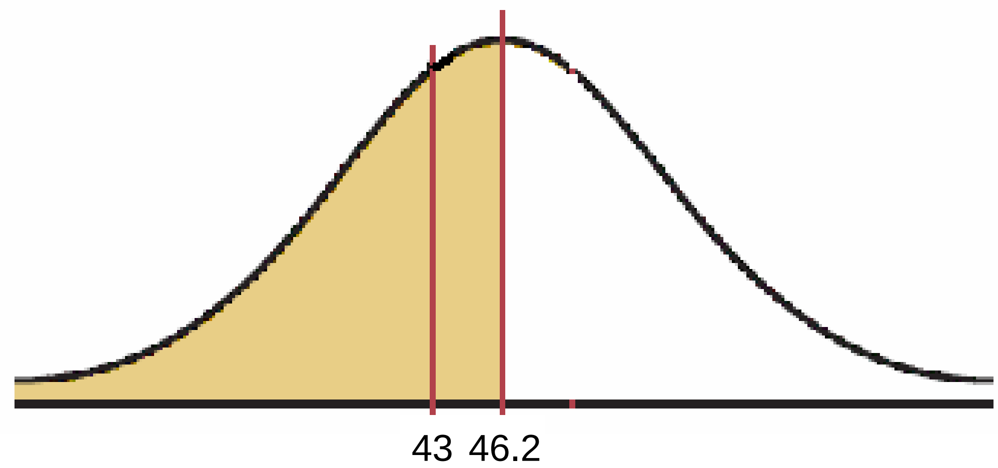
The desired area is 0.7088 - 0.2912 = 0.4176.

The probability of obtaining a sample mean between 120 and 126 months is 41.76%.

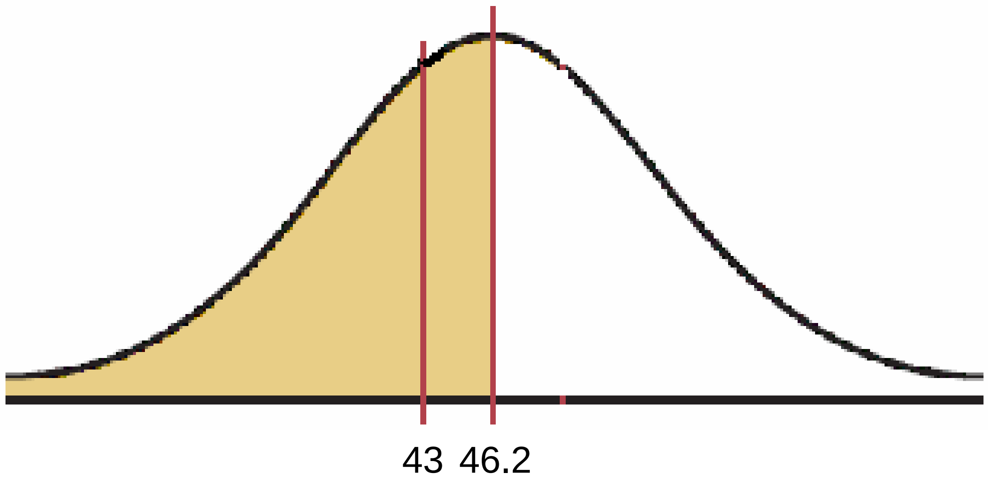
Ex.) 5 – The Central Limit Theorem

Time to Complete an Exam The average time it takes a group of adults to complete a certain achievement test is 46.2 minutes. The standard deviation is 8 minutes. Assume the variable is normally distributed.

b. Find the probability that if 50 randomly selected adults take the test, the mean time it takes the group to complete the test will be less than 43 minutes.



Ex.) 5 – The Central Limit Theorem



$$z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} = \frac{43 - 46.2}{8 / \sqrt{50}} = -2.83$$

The area to the left of *z* = -2.83 is 0.0023. Hence, the probability that the mean of a sample of 50 individuals is less than 43 minutes is 0.0023, or 0.23%.

Class Activity: The Central Limit Theorem

Breaking Strength of Steel Cable The average breaking strength of a certain brand of steel cable is 2000 pounds, with a standard deviation of 100 pounds. A sample of 20 cables is selected and tested. Find the sample mean that will cut off the upper 95% of all samples of size 20 taken from the population. Assume the variable is normally distributed.