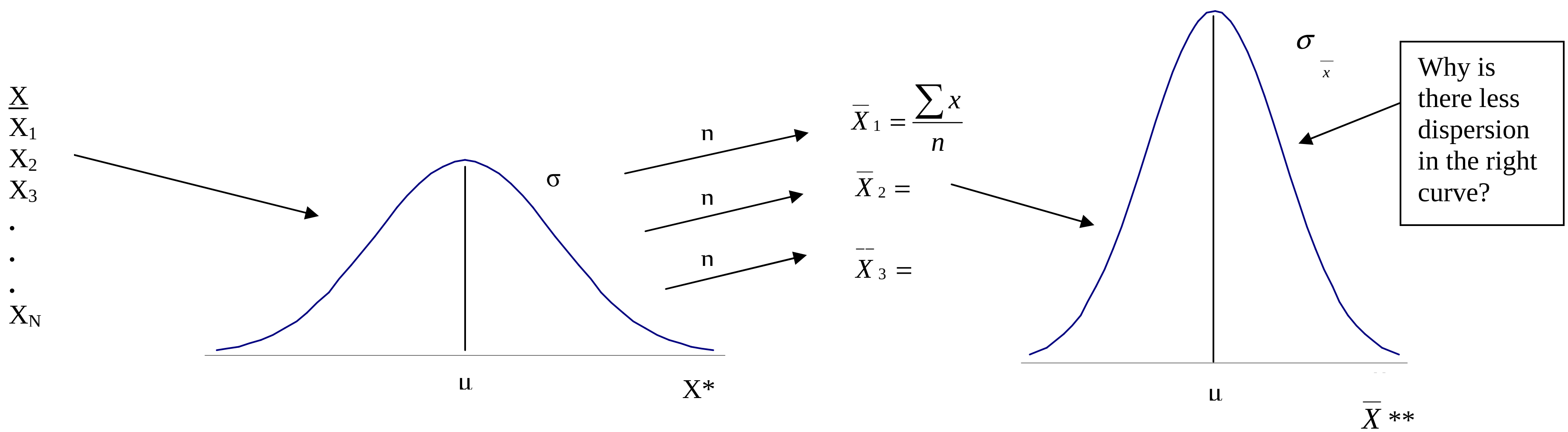
8.1 Sampling Distributions; The Central Limit Theorem

------------------------- Population Distribution (7.3) ------------------------ -------- Sampling Distribution of the Mean (8.1) ----------



* X represents individual observations from the population.

** $\bar{X}$ represents means of random samples of n observations.

Sampling Distribution of the Mean ($\bar{X}$): Probability distribution of ALL possible values of the random variable $\bar{X}$ computed from a sample of size n from a population with mean μ and standard deviation σ.

Theorem (p. 319): If a random variable X is normally distributed with mean μ and standard deviation σ, then the distribution of the sample mean, $\bar{X}$, is normally distributed with mean $\mu_{\bar{x}} = \mu$ and standard deviation $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.

Sampling Distribution of the Mean ($\bar{X}$):

(1) **Shape**—If the population is normally distributed, the distribution of $\bar{X}$ will be normal.

(2) **Mean**—The mean of the sampling distribution of $\bar{X}$ is $\mu_{\bar{x}}$, which is equal to the mean of the population, μ. Based on this result, $\bar{X}$ is termed an unbiased statistic (or estimate).

Definition: A statistic ($\bar{X}$) is an **unbiased estimate** of the population parameter (μ) if the mean of ALL possible values of the statistic is equal to the corresponding parameter.

If $\mu_{\bar{x}} = \frac{\sum \bar{x}_i}{\#} = \mu$, then $\bar{X}$ is an unbiased statistic.

An unbiased statistic, *on the average*, is equal to the corresponding parameter. *Individually*, a specific value of the statistic may (and generally is) different from the parameter.

Review the example on the next page.

(3) **Standard Deviation**—is equal to the population standard deviation (σ) divided by the square root of the number of sample observations (n).

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

The standard deviation of the sampling distribution of $\bar{X}$ is called the **standard error of the mean**.

See the example on p. 3.

Sampling Distribution of the Mean, $\overline{X}$:

Population: {2, 4, 6, 8, 4, 3, 7} with μ=4.9

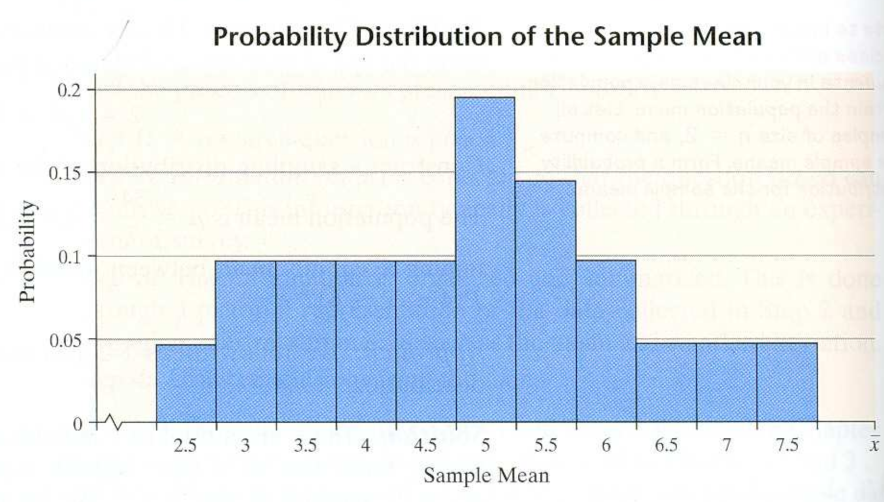
TABLE 7

Sample	Sample Mean	Sample	Sample Mean	Sample	Sample Mean
2, 4	3	4, 8	6	6, 7	6.5
2, 6	4	4, 4	4	8, 4	6
2, 8	5	4, 3	3.5	8, 3	5.5
2, 4	3	4, 7	5.5	8, 7	7.5
2, 3	2.5	6, 8	7	4, 3	3.5
2, 7	4.5	6, 4	5	4, 7	5.5
4, 6	5	6, 3	4.5	3, 7	5

What is the mean of ALL the 21 sample means?

$$\mu_{\bar{x}} = \frac{\sum \bar{x}}{\#} = \frac{3+4+5+...+5.5+5}{21} = \frac{102}{21} = 4.9$$

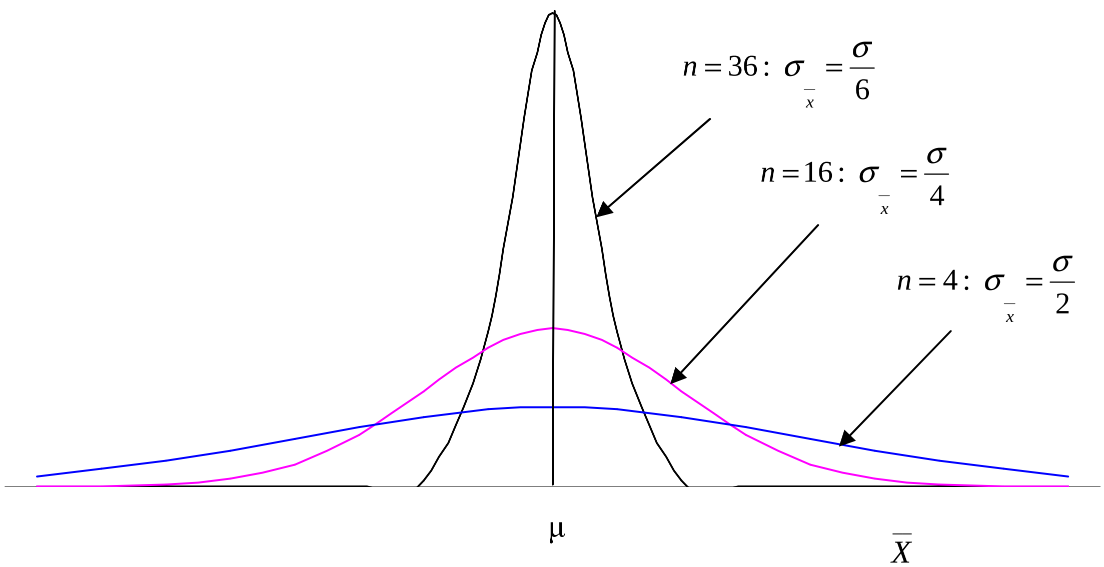
Figure 48



Sampling Distribution of the Mean—Standard Deviation:

n	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$
4	$\frac{\sigma}{\sqrt{4}} = \frac{\sigma}{2}$
16	$\frac{\sigma}{\sqrt{16}} = \frac{\sigma}{4}$
36	$\frac{\sigma}{36} = \frac{\sigma}{6}$

Sampling Distribution of the Mean



As the sample size increases, the sampling distribution of $\bar{X}$ becomes taller and thinner and more concentrated around the mean. That is, as n increases, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ decreases.

Problems on the Sampling Distribution of the Mean:

#1. For a normally distributed population with μ=100 and σ=8, draw a diagram of the distribution of the sample mean for n=25. Be sure to show the mean and standard deviation of this distribution.

#2. Assume a normally distributed population with μ=100 and σ=8. What is the probability of randomly selecting a sample of 25 observations with a mean between 100 and 103.2?—i.e., find P(100< $\bar{X}$ <103.2). Include a diagram (with labels) to illustrate your answer.

#3. For a normally distributed population with μ=100 and σ=8, find P(100< X <103.2). Include a diagram (with labels) to illustrate your answer.

Do you see the difference between #2 and #3? Explain.

#4. Assume a normally distributed population with μ=400 and σ=60. A random sample of 36 observations is drawn from the population. What proportion of sample means (for n=36) will be greater than 415? Include a diagram (with labels) to illustrate your answer.

#5. A normally distributed population with μ=400 and σ=60. A random sample of 36 observations is drawn from the population. Find the two limits ($\bar{X}_1$ and $\bar{X}_2$) that include the middle 95% of sample means.

#6. Continuing problem #5 but increasing the sample size to n=100, find the two limits ($\bar{X}_1$ and $\bar{X}_2$) that include the middle 95% of sample means.

Comparing the answers to #5 and #6, what can one say about the effect of sample size (n) on how accurate $\bar{X}$ estimates μ? Explain.

The Central Limit Theorem

Suppose a random variable X has population mean μ and standard deviation σ and that a random sample of size n is taken from this population. Then the sampling distribution of $\overline{X}$ becomes approximately normal as the sample size n increases. The mean of the distribution is $\mu_{\bar{x}} = \mu$ and the standard deviation is $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.

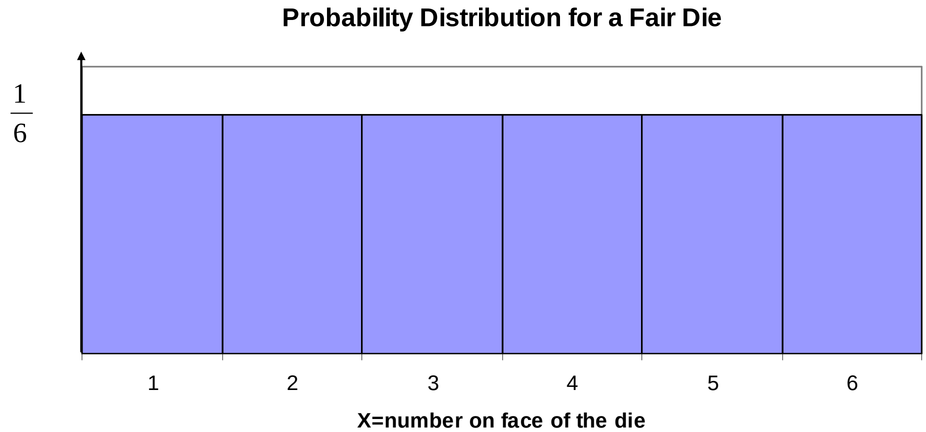
Summary of the Central Limit Theorem:

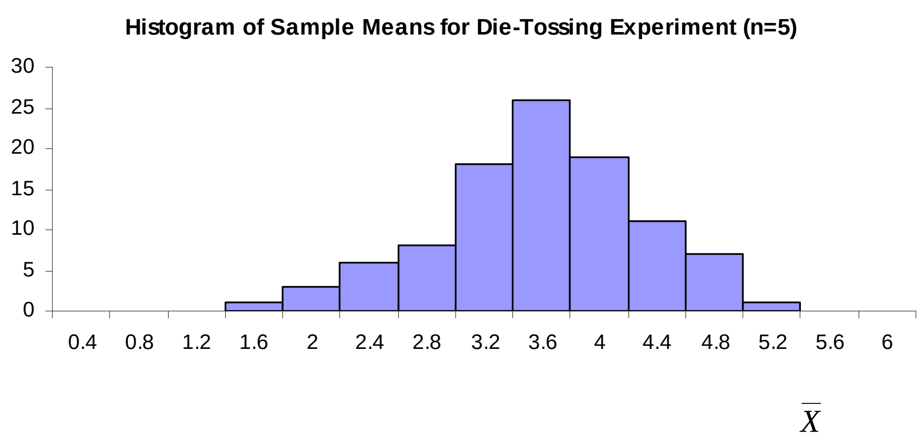
- The population distribution can be non-normal and have any shape.
- As the sample size increases, the distribution of $\overline{X}$ approaches a normal distribution.
- For "large n" (which is assumed to mean n $\geq$ 30), $\overline{X}$ will be approximately normally distributed even if the population does not satisfy a normal distribution.

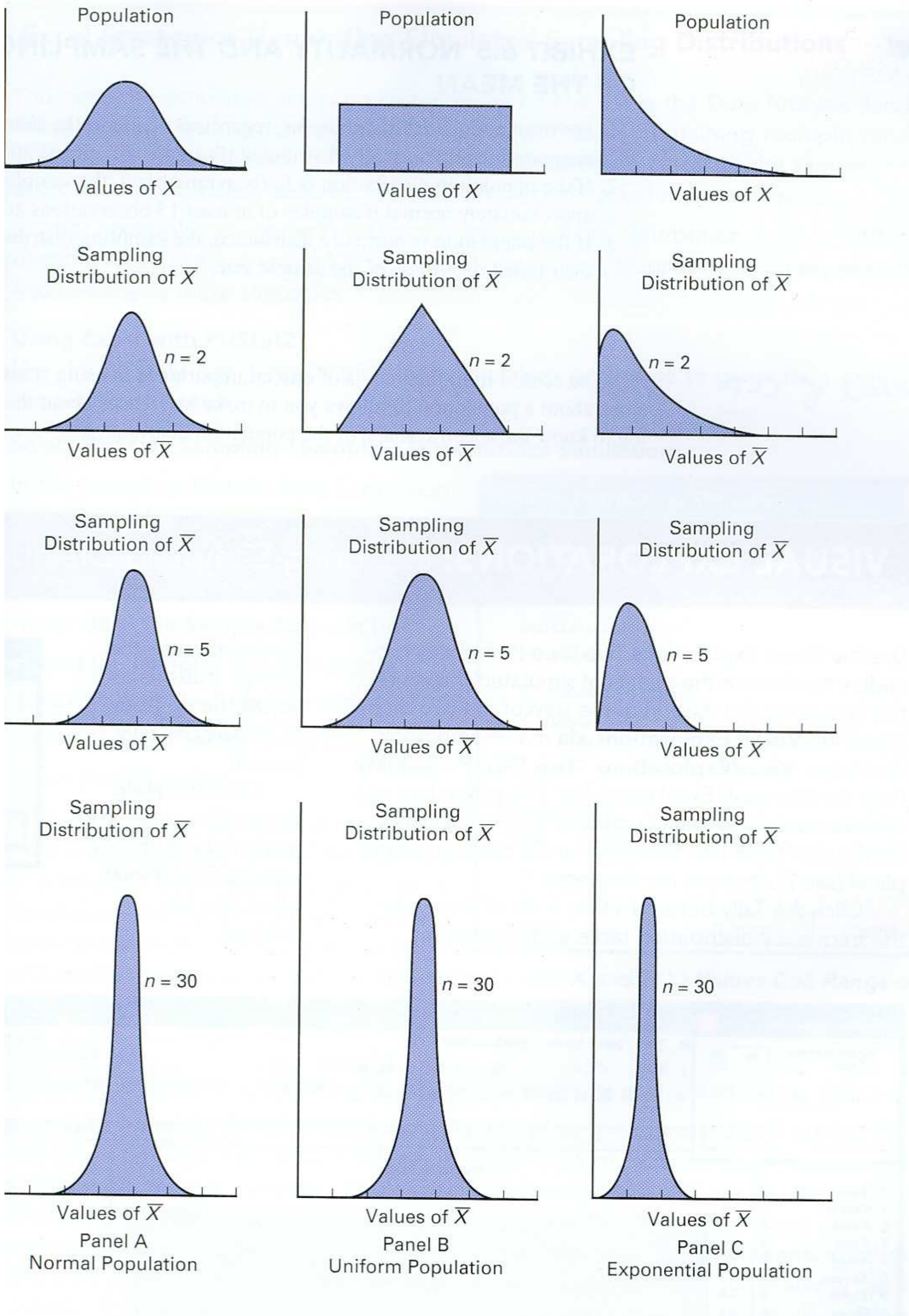
*****Note:** The Central Limit Theorem focuses on the **<u>***shape***</u>** of the distribution of the sample mean, i.e., is the shape normal? (The mean and spread (standard deviation) of the distribution take back stage in the Central Limit Theorem. Regardless of the size of the sample (n), the mean of the distribution of $\overline{X}$ is μ and the standard deviation is $\frac{\sigma}{\sqrt{n}}$.)

Die-Tossing Experiment (n=5):

Sample Number	Sample Observations					$\overline{X}$	Sample Number	Sample Observations					$\overline{X}$
1	5	4	5	2	6	4.4	51	3	5	2	1	6	3.4
2	2	2	6	1	4	3	52	4	1	4	4	3	3.2
3	6	1	3	5	5	4	53	4	6	2	1	1	2.8
4	2	1	6	1	3	2.6	54	5	5	5	6	5	5.2
5	3	1	1	1	6	2.4	55	2	6	6	2	5	4.2
6	2	3	2	4	1	2.4	56	1	6	1	3	1	2.4
7	4	1	1	4	1	2.2	57	2	2	1	5	5	3
8	6	2	3	6	6	4.6	58	6	1	6	1	2	3.2
9	2	2	5	2	2	2.6	59	4	2	2	5	3	3.2
10	1	5	4	1	1	2.4	60	2	2	3	5	5	3.4
11	4	3	6	6	1	4	61	5	2	2	3	5	3.4
12	6	3	4	3	6	4.4	62	5	4	3	3	1	3.2
13	6	4	5	5	5	5	63	2	1	5	5	4	3.4
14	2	4	5	6	5	4.4	64	1	3	5	6	6	4.2
15	1	5	6	4	4	4	65	3	1	1	2	6	2.6
16	3	3	6	6	2	4	66	1	2	3	4	3	2.6
17	4	2	6	6	2	4	67	2	1	1	4	4	2.4
18	3	1	4	3	4	3	68	6	5	6	4	2	4.6
19	1	3	6	2	3	3	69	5	2	5	4	4	4
20	2	6	5	5	3	4.2	70	4	3	1	2	4	2.8
21	4	4	6	1	2	3.4	71	4	6	1	5	3	3.8
22	6	3	3	6	5	4.6	72	3	3	6	4	1	3.4
23	2	2	6	6	2	3.6	73	4	3	4	6	5	4.4
24	2	2	2	5	3	2.8	74	1	1	4	1	3	2
25	1	4	4	1	3	2.6	75	2	5	2	3	6	3.6
26	3	6	6	3	4	4.4	76	3	4	6	6	2	4.2
27	6	4	4	3	1	3.6	77	3	1	6	3	4	3.4
28	1	2	3	1	2	1.8	78	4	6	2	3	5	4
29	1	4	6	6	2	3.8	79	4	4	5	4	1	3.6
30	2	4	2	4	6	3.6	80	6	3	2	2	6	3.8
31	4	4	4	4	5	4.2	81	4	1	4	1	2	2.4
32	3	2	3	4	2	2.8	82	5	2	3	5	2	3.4
33	5	1	4	3	3	3.2	83	2	5	3	6	6	4.4
34	3	1	4	2	3	2.6	84	2	3	1	5	1	2.4
35	2	1	4	3	6	3.2	85	4	1	1	3	3	2.4
36	3	1	1	1	6	2.4	86	5	5	5	4	6	5
37	5	2	5	5	6	4.6	87	5	4	5	6	3	4.6
38	6	3	4	3	6	4.4	88	6	5	1	6	5	4.6
39	6	3	1	6	1	3.4	89	6	1	1	3	1	2.4
40	4	3	5	5	4	4.2	90	1	2	2	4	3	2.4
41	5	4	1	2	2	2.8	91	3	6	3	2	6	4
42	3	2	5	2	6	3.6	92	3	5	6	6	6	5.2
43	3	2	3	2	2	2.4	93	6	4	4	4	5	4.6
44	5	3	1	6	1	3.2	94	2	3	3	4	3	3
45	5	6	6	5	2	4.8	95	2	3	1	5	1	2.4
46	6	1	2	2	6	3.4	96	2	1	1	2	5	2.2
47	6	6	3	3	1	3.8	97	4	4	6	5	6	5
48	4	1	6	2	2	3	98	1	2	3	2	2	2
49	2	4	5	4	5	4	99	2	3	1	3	1	2
50	4	2	1	1	5	2.6	100	5	5	4	6	1	4.2


Probability Distribution for a Fair Die
1/6
1
2
3
4
5
6
X=number on face of the die


Histogram of Sample Means for Die-Tossing Experiment (n=5)
30
25
20
15
10
5
0
0.4
0.8
1.2
1.6
2
2.4
2.8
3.2
3.6
4
4.4
4.8
5.2
5.6
6
$\overline{X}$



Source: *Statistics for Managers Using Microsoft Excel* by Levine, Stephan, Krehbiel, and Berenson (Pearson, Prentice-Hall, 4^{th} edition).