Arab World Edition

Statistics for Business

Farouk Benghezal

ALWAYS LEARNING

PEARSON

The variance of the sample means is

$$1.04/2$$
$$= 0.52$$

where 1.04 is the variance σ^2 of the population. *The variance of the sample means is equal to the variance of the population divided by the sample size n:* σ^2/n.

Characteristics of the sampling distribution of the sample are:

The mean of the sample means is

$$\mu_{\bar{X}} = \mu \qquad \textbf{6.1}$$

The variance of the sample means is

$$\sigma^2_{\bar{X}} = \sigma^2/n \qquad \textbf{6.2}$$

The standard deviation of the sample means is

$$\sigma_{\bar{X}} = \sigma/\sqrt{n} \qquad \textbf{6.3}$$

The spread in the distribution of the sample means is less than the spread in the population because of the sample size. *As n increases the variance of the sample means decreases*.

Using the data of Table 6.2, we obtain the grade distribution displayed in Table 6.5.

TABLE 6.5
Grade Distribution

Grade	Frequency	$P(X)$
1	1	0.2
2	1	0.2
3	2	0.4
4	1	0.2
Total	5	

The shape of the population distribution is displayed in Figure 6.1. The distribution of the sample means given in Table 6.4 is displayed in Figure 6.2.

FIGURE 6.1
Population Distribution

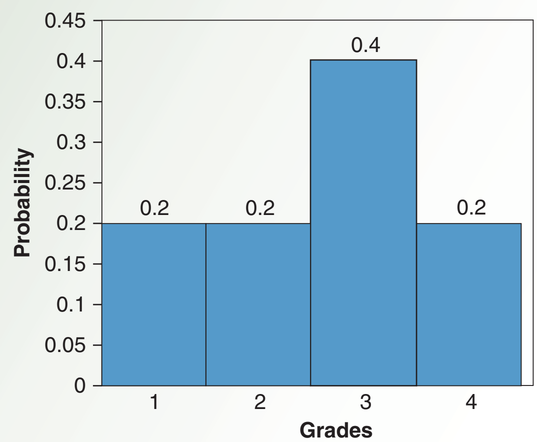
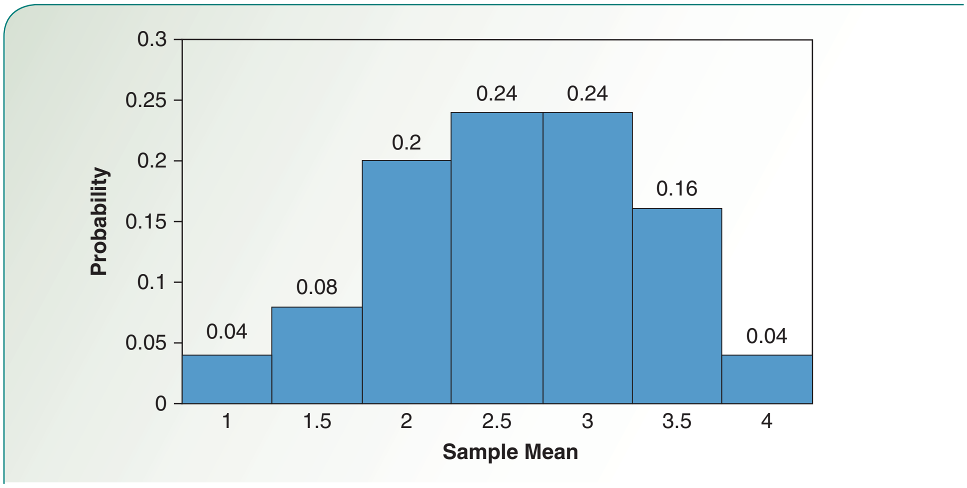


Figure 6.2

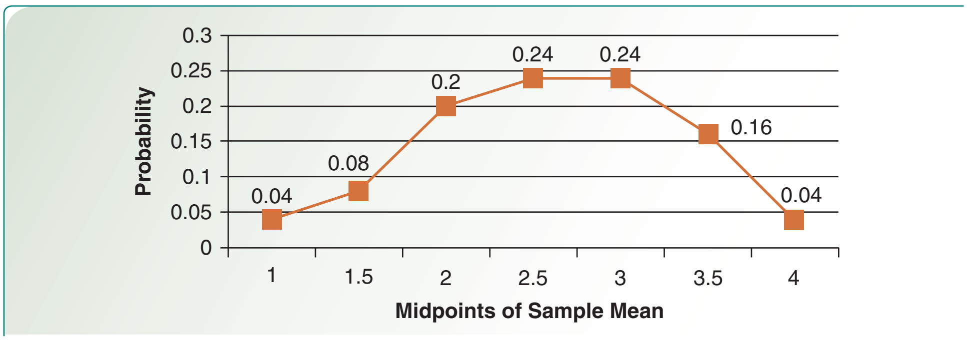
Distribution of the Sample Means



If we compare these two distributions, the sample mean distribution tends to be more *bell-shaped* and similar to the normal curve. Figure 6.3 displays the frequency polygon of the sample means. Indeed, the curve is near bell-shaped.

Figure 6.3

Frequency Polygon of Sample Means



In the next section, we will see that the distribution of sample means becomes more bell-shaped as n increases.

An important result is the following. If we sample from a *normal distribution* with mean μ and standard deviation σ, *the sample mean has a normal distribution with mean* μ *and standard deviation* $\sigma/\sqrt{n}$. So the standard deviation of the sample means is $1/\sqrt{n}$ times the size of the standard deviation of the population distribution.

The fact that the sampling distribution of $\overline{X}$ has mean μ indicates that, on average, the sample mean is equal to the population mean. The distribution of $\overline{X}$ is centered on the parameter (μ) to be estimated, which makes $\overline{X}$ a good estimator of μ.

If the sampled population is not normally distributed, what is the distribution of the sample means? As previously stated, Formulas 6.1–6.3 hold: $\mu_{\overline{X}} = \mu$ and $\sigma_{\overline{X}} = \sigma/\sqrt{n}$. For large random samples, the shape of the sampling distribution of the sample mean approaches a normal distribution. This result is formally stated as the *central limit theorem*.

The Central Limit Theorem

The result that we just stated – that the sampling distribution of $\overline{X}$ tends to the normal distribution as the sample size increases – is one of the most important results in Statistics. Most of the time, the population from which the samples are selected is not normally distributed. In such situations, the shape of the sampling distribution of $\overline{X}$ is inferred from a very important theorem called the **central limit theorem**.

CENTRAL LIMIT THEOREM

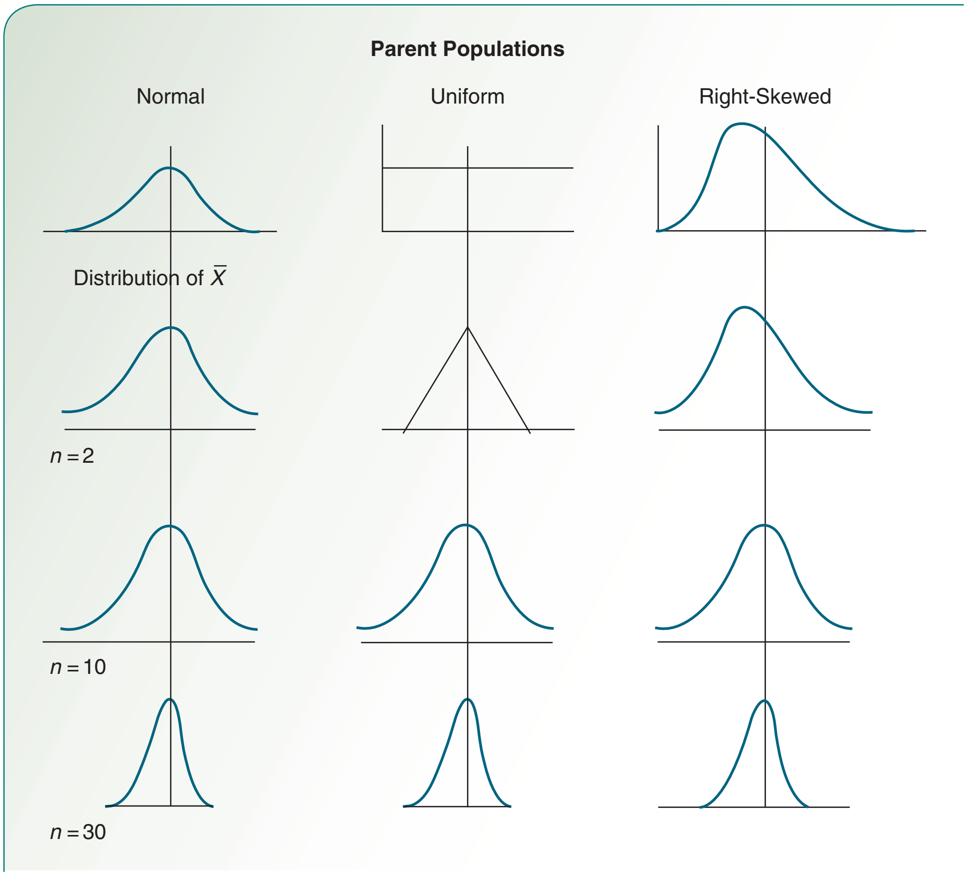
If samples of size n are drawn randomly from a population with mean μ and standard deviation σ, the sample means $\overline{X}$ are approximately normally distributed with mean μ and standard deviation $\sigma/\sqrt{n}$ for sufficiently large samples ($n \geq 30$), regardless of the shape of the population distribution.

The second part of the theorem is an extremely strong result: it states that we can assume that $\overline{X}$ follows an approximate normal distribution regardless of the shape of the population from which the sample was drawn, provided the sample is large enough, say 30 or more. For example, if we repeatedly sampled from a population that is exponentially distributed, the resulting sample means $\overline{X}$ would follow a normal distribution and not an exponential distribution. However, the real advantage of the central limit theorem is that sample data drawn from a population that is not normally distributed or from a population of unknown shape can also be analyzed by using the normal distribution because the sample means are normally distributed for sufficiently large sample sizes.

The idea that the distribution of the sample means from a population that is not normally distributed will converge to normality is illustrated in Figure 6.4. If we take the right-skewed distribution for example, as the sample size increases, the distribution becomes more bell-shaped.

FIGURE 6.4

Results of the Central Limit Theorem for Different Populations



As the sample size increases, the standard deviation of the sample means becomes smaller and smaller because the population standard deviation is being divided by larger and larger values of the square root of n. The benefit of the central limit theorem is a practical version of the z-score formula for sample means:

$$z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \quad (6.4)$$

Example 6.1

Suppose that during any hour in City Center Mall in Doha, the average number of shoppers is 664, with a standard deviation of 33. What is the probability that a random sample of 64 different shopping hours will result in a sample mean that is

a) less than 655?
b) greater than 670?
c) between 660 and 669?

Solution

For this problem, we have $\mu = 664, \sigma = 33$, and $n = 64$.

a) We compute the z-score:

$$z = \frac{655 - 664}{\frac{33}{\sqrt{64}}} = -2.18$$

Next, we use the standard normal table of Appendix C:

$$\begin{aligned} P(Z < -2.18) &= 0.5 - 0.4854 \\ &= 0.0146 \end{aligned}$$

b) We compute the z-score:

$$z = \frac{670 - 664}{\frac{33}{\sqrt{64}}} = 1.45$$

Next, we use the standard normal table of Appendix C:

$$\begin{aligned} P(Z > 1.45) &= 0.5 - 0.4265 \\ &= 0.0735 \end{aligned}$$

c) We compute the z-scores:

$$z_1 = \frac{660 - 664}{\frac{33}{\sqrt{64}}} = -0.97$$

$$z_2 = \frac{669 - 664}{\frac{33}{\sqrt{64}}} = 1.21$$

We use the standard normal table of Appendix C:

$$\begin{aligned} P(600 \leq \overline{X} \leq 669) &= P(-0.97 \leq Z \leq 1.21) = 0.3340 + 0.3869 \\ &= 0.7209 \end{aligned}$$

There is a 72.09% chance of randomly selecting 64 hourly periods for which the sample mean is between 660 and 669 shoppers.

EXAMPLE 6.2

Assume that patients at Frantz Fanon Hospital in Blida, Algeria, stay an average length of 4.8 days, with a standard deviation of 1.7 days. Dr. El Mahdi selects a sample of 46 patients.

a) What is the probability that the average length of stay does not exceed 5.2 days?

b) If the sample size had been 14 patients instead of 46, what further assumption would have been necessary to solve the above problem?

(*Continued*)

Solution

a) Since the sample size 46 is greater than 30 we can use the central limit theorem; the sample mean is approximately normally distributed:

$$\mu = 4.8,\ \sigma = 1.7, \text{ and } n = 46,\ X = 5.2$$

$$z = \frac{5.2 - 4.8}{\dfrac{1.7}{\sqrt{46}}} = 1.63$$

b) Since the sample size 14 is less than 30, we cannot apply the central limit theorem. To solve the problem we need to assume that the population of the lengths of stay is normally distributed.

$$P(\overline{X} \leq 5.2) = P(Z \leq 1.65) = 0.9484$$

TECHNOLOGY

TEMPLATE 6.1A

The Sampling Distribution of the Sample Mean

We enter the population mean and population standard deviation in cells C5 and C6 respectively. The sample size n is entered in cell F3. The standard deviation of the sample mean is automatically calculated in cell G6. Probabilities are computed in cells C9, F9, and I9.

	A	B	C	D	E	F	G	H	I	J
1	**Sampling Distribution of the Sample Mean**									
2	**Example 6.1**									
3	σ known				Sample Size	64				
4	Population Distribution				Sampling Distribution					
5		Mean	664			Mean	664			
6		St. Deviation	33			St. Deviation	4.125			
7										
8		x	P(X<x)		x	P(X>x)		x1	P(x1≤X≤x2)	x2
9		655	0.0146		670	0.0729		660	0.7212	669

CHECK YOUR UNDERSTANDING

Review Problems

6.1 The manager of Shifa Showroom, a branch of Al Jazirah Vehicles, a Ford dealer in Riyadh, finds that the four salespersons have sold 0, 1, 3, and 4 cars respectively in the past week. List the number of cars sold by two salespersons selected randomly with replacement for all possible samples of size two.

6.2 Consider Problem 6.1.

a) Compute the sample mean of each sample of size two.