GLOBAL EDITION

Basic Business Statistics

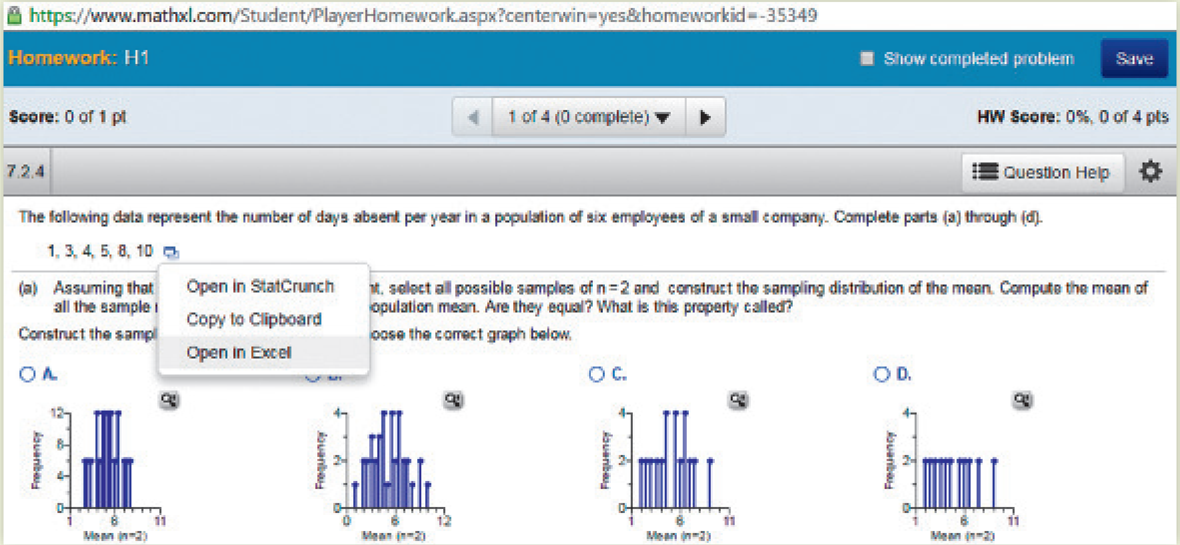
Concepts and Applications

FOURTEENTH EDITION

Mark L. Berenson • David M. Levine
Kathryn A. Szabat • David F. Stephan

Career Readiness

Preparing students for careers is a priority. Now with the Open in Excel functionality in homework problems, students can open data sets directly in Excel and gain experience with the tools they will use in their careers.



Bring Statistics to Life

Using the StatCrunch Twitter app you can see what people are tweeting about in real time. You can easily load tweets into StatCrunch and construct an interactive word wall showing the most commonly used words.

Visit pearson.com/myLab/statistics and click Get Trained to make sure you're getting the most out of MyLab Statistics.

The Minitab normal probability plot has the 3YrReturn variable on the X axis and the cumulative percentage for a normal distribution on the Y axis. In this plot, if the data are normally distributed, the points will plot along an approximately straight line. In Figure 6.19, most points, other than several extreme values, approximately follow a straight line, indicating an approximately normal distribution. Had the data been right-skewed, the curve would have risen more rapidly at first and then leveled off. Had the data been left-skewed, the data would have risen more slowly at first and then risen at a faster rate for larger values of the variable.

PROBLEMS FOR SECTION 6.3

LEARNING THE BASICS

6.14 For a sample of $n = 39$ elements, find the lower and upper values of Z and show that the middle value has a Z value of zero.

6.15 For a sample of $n = 6$, list the six Z values.

APPLYING THE CONCEPTS

SELF TEST **6.16** The FIFA World Cup was one of the biggest sporting events of 2018. The file **WC2018Players** contains data of the players of the 32 teams that qualified for the event. A dummy variable is included to indicate whether a player is also a captain.

Source: Data adapted from **https://bit.ly/2zGSWRD**.

Decide whether players' ages appear to be approximately normally distributed by

a. comparing data characteristics to theoretical properties.
b. constructing a normal probability plot.

6.17 The FIFA World Cup was one of the biggest sporting events of 2018. The file **WC2018TeamAge** contains average age of the players (years, in 2018) of the 32 teams that qualified for the event.

Source: Data adapted from **https://bit.ly/2zGSWRD**.

Decide whether the teams' mean ages appear to be approximately normally distributed by

a. comparing data characteristics to theoretical properties.
b. constructing a normal probability plot.

6.18 Unemployment is one of the major issues most governments of the world are faced with. The file **EuUnempl2017** contains employment data for 319 European regions in 2017.

Decide whether employment rates appear to be approximately normally distributed by

a. comparing data characteristics to theoretical properties.
b. constructing a normal probability plot.

6.19 Thirty companies comprise the DJIA. How big are these companies? One common method for measuring the size of a company is to use its market capitalization, which is computed by multiplying the number of stock shares by the price of a share of stock. On January 10, 2017 the market capitalization of these companies ranged from Traveler's \$33.3 billion to Apple's \$625.6 billion. The entire population of market capitalization values is stored in **DowMarketCap**.

Source: Data extracted from **money.cnn.com**, January 10, 2017.

Decide whether the market capitalization of companies in the DJIA appears to be approximately normally distributed by

a. comparing data characteristics to theoretical properties.
b. constructing a normal probability plot.
c. constructing a histogram.

6.20 One operation of a mill is to cut pieces of steel into parts that will later be used as the frame for front seats in an automotive plant. The steel is cut with a diamond saw, and the resulting parts must be within ± 0.005 inch of the length specified by the automobile company. The data come from a sample of 100 steel parts and are stored in **Steel**. The measurement reported is the difference, in inches, between the actual length of the steel part, as measured by a laser measurement device, and the specified length of the steel part. Determine whether the data appear to be approximately normally distributed by

a. comparing data characteristics to theoretical properties.
b. constructing a normal probability plot.

6.21 The file **IndexReturn** contains data about the performance of 38 indexes across the world as of July 2018.

Source: Data extracted from **https://bit.ly/2yS1QcS**.

Decide whether one-year and five-year returns appear to be approximately normally distributed by

a. comparing data characteristics to theoretical properties.
b. constructing a normal probability plot.

6.22 The file **Utility** contains the electricity costs, in dollars, during July of a recent year for a random sample of 50 one-bedroom apartments in a large city:

96	171	202	178	147	102	153	197	127	82
157	185	90	116	172	111	148	213	130	165
141	149	206	175	123	128	144	168	109	167
95	163	150	154	130	143	187	166	139	149
108	119	183	151	114	135	191	137	129	158

Decide whether the data appear to be approximately normally distributed by

a. comparing data characteristics to theoretical properties.
b. constructing a normal probability plot.

6.4 The Uniform Distribution

In the **uniform distribution**, the values are evenly distributed in the range between the smallest value, a, and the largest value, b. Selecting random numbers is one of the most common uses of the uniform distribution. When you use simple random sampling (see Section 1.3), you assume that each random digit comes from a uniform distribution that has a minimum value of 0 and a maximum value of 9.

Equation (6.4) defines the probability density function for the uniform distribution.

UNIFORM PROBABILITY DENSITY FUNCTION

$$f(X) = \frac{1}{b - a} \text{ if } a \leq X \leq b \text{ and 0 elsewhere} \tag{6.4}$$

where

a = minimum value of X
b = maximum value of X

Equation (6.5) defines the mean of the uniform distribution, and Equation (6.6) defines the variance and standard deviation of the uniform distribution.

MEAN OF THE UNIFORM DISTRIBUTION

$$\mu = \frac{a + b}{2} \tag{6.5}$$

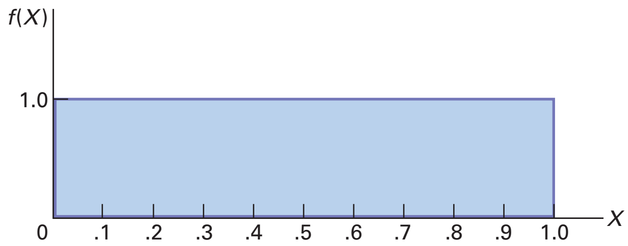
VARIANCE AND STANDARD DEVIATION OF THE UNIFORM DISTRIBUTION

$$\sigma^2 = \frac{(b - a)^2}{12} \tag{6.6a}$$

$$\sigma = \sqrt{\frac{(b - a)^2}{12}} \tag{6.6b}$$

Because of its shape, the uniform distribution is sometimes called the **rectangular distribution** (see Figure 6.1 Panel B on page 256). Figure 6.20 illustrates the uniform distribution with $a = 0$ and $b = 1$. The total area inside the rectangle is 1.0, equal to the base (1.0) times the height (1.0). Having an area of 1.0 satisfies the requirement that the area under any probability density function equals 1.0.

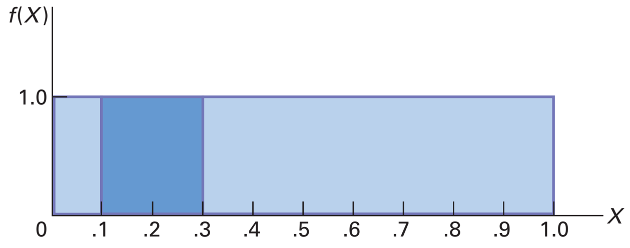
FIGURE 6.20
Probability density function for a uniform distribution with $a = 0$ and $b = 1$



In this uniform distribution, what is the probability of getting a random number between 0.10 and 0.30? The area between 0.10 and 0.30, depicted in Figure 6.21, is equal to the base (which is $0.30 - 0.10 = 0.20$) times the height (1.0). Therefore,

$$P(0.10 < X < 0.30) = (\text{Base})(\text{Height}) = (0.20)(1.0) = 0.20$$

FIGURE 6.21
Finding $P(0.10 < X < 0.30)$ for a uniform distribution with $a = 0$ and $b = 1$



From Equations (6.5) and (6.6), the mean and standard deviation of the uniform distribution for $a = 0$ and $b = 1$ are computed as follows:

$$\mu = \frac{a + b}{2} = \frac{0 + 1}{2} = 0.5$$

and

$$\sigma^2 = \frac{(b - a)^2}{12} = \frac{(1 - 0)^2}{12} = \frac{1}{12} = 0.0833$$

$$\sigma = \sqrt{0.0833} = 0.2887.$$

Thus, the mean is 0.5, and the standard deviation is 0.2887.

Example 6.6 provides another application of the uniform distribution.

EXAMPLE 6.6

Computing Uniform Probabilities

In the MyTVLab scenario on page 255, the load time of the new sales page was assumed to be normally distributed with a mean of 7 seconds. Suppose that the load time follows a uniform (instead of a normal) distribution between 4.5 and 9.5 seconds. What is the probability that a load time will take more than 9 seconds?

SOLUTION The load time is uniformly distributed from 4.5 to 9.5 seconds. The area between 9 and 9.5 seconds is equal to 0.5 seconds, and the total area in the distribution is $9.5 - 4.5 = 5$ seconds. Therefore, the probability of a load time between 9 and 9.5 seconds is the portion of the area greater than 9, which is equal to $0.5/5.0 = 0.10$. Because 9.5 is the maximum value in this distribution, the probability of a load time above 9 seconds is 0.10. In comparison, if the load time is normally distributed with a mean of 7 seconds and a standard deviation of 2 seconds (see Example 6.1 on page 261), the probability of a load time above 9 seconds is 0.1587.

PROBLEMS FOR SECTION 6.4

LEARNING THE BASICS

6.23 Suppose you select one value from a uniform distribution with $a = 0$ and $b = 10$. What is the probability that the value will be

a. between 5 and 7?
b. between 2 and 3?
c. What is the mean?
d. What is the standard deviation?

APPLYING THE CONCEPTS

SELF TEST **6.24** The time it takes for a plane to be cleaned and ready for the next flight is uniformly distributed between 35 and 45 minutes. What is the probability that the cleaning time will be

a. less than 37minutes?
b. between 35 and 40 minutes?
c. more than 38 minutes?
d. calculate the mean and the standard deviation for the cleaning time of an airplane.

6.25 A study of the time spent by visitors to finish the viewing of a marine life aquarium is uniformly distributed between 120 and200 minutes. What is the probability that the viewing time will be
a. between 150 and 190 minutes.
b. less than 160 minutes.
c. Calculate the mean and the standard deviation for the viewing time.

6.26 How long does it take to download a two-hour HD movie from the iTunes store? According to Apple's technical support site, **support.apple.com/en-us/HT201587**, downloading such a movie using a 15 Mbit/s broadband connection should take 29–43 minutes. Assume that the download times are uniformly distributed between 29 and 43 minutes. If you download a two-hour movie, what is the probability that the download time will be
a. less than 30 minutes?
b. more than 36 minutes?
c. between 30 and 40 minutes?
d. What are the mean and standard deviation of the download times?

6.27 The scheduled time for a flight between Kuwait city and the city of Dubai in the United Arab Emirates is 75 minutes. Assume that the actual flight time is uniformly distributed between 73 and 85 minutes. Find the probability that the flight time will be
a. less than 78 minutes.
b. between 75 and 80 minutes.
c. greater than 65 minutes.
d. Calculate the mean and standard deviation of the flight time between the two cities.

6.5 The Exponential Distribution

The **exponential distribution** is a continuous distribution that is right-skewed and ranges from 0 to positive infinity (see Figure 6.1 on page 256). The **Section 6.5 online topic** discusses this distribution and illustrates its application.

6.6 The Normal Approximation to the Binomial Distribution

In many circumstances, the normal distribution can be used to approximate the binomial distribution, discussed in Section 5.2. The **Section 6.6 online topic** discusses this technique and illustrates its use.

USING **STATISTICS**
Normal Load Times . . . , Revisited

In the Normal Downloading at MyTVLab scenario, you were the sales and marketing vice president for a web-based business. You sought to ensure that the load time for a new sales web page would be within a certain range. By running experiments in the corporate offices, you determined that the amount of time, in seconds, that passes from first pointing a browser to a web page until the web page is fully loaded is a bell-shaped distribution with a mean load time of 7 seconds and standard deviation of 2 seconds. Using the normal distribution, you were able to calculate that approximately 84% of the load times are 9 seconds or less, and 95% of the load times are between 3.08 and 10.92 seconds.

Now that you understand how to compute probabilities from the normal distribution, you can evaluate load times of similar sales web pages that use other designs. For example, if the standard deviation remained at 2 seconds, lowering the mean to 6 seconds would shift the entire distribution lower by 1 second. Thus, approximately 84% of the load times would be 8 seconds or less, and 95% of the load times would be between 2.08 and 9.92 seconds. Another change that could reduce long load times would be reducing the variation. For example, consider the case where the mean remained at the original 7 seconds but the standard deviation was reduced to 1 second. Again, approximately 84% of the load times would be 8 seconds or less, and 95% of the load times would be between 5.04 and 8.96 seconds.

SUMMARY

In this and the previous chapter, you have learned about mathematical models called probability distributions and how they can be used to solve business problems. In Chapter 5, you used discrete probability distributions in situations where the values come from a counting process such as the number of social media sites to which you belong or the number of tagged order forms in a report generated by an accounting information system. In this chapter, you learned about continuous probability distributions where the values come from a measuring process such as your height or the download time of a video.

Continuous probability distributions come in various shapes, but the most common and most important in business is the normal distribution. The normal distribution is symmetrical; thus, its mean and median are equal. It is also bell-shaped, and approximately 68.26% of its values are within ± 1 standard deviation of the mean, approximately 95.44% of its values are within ± 2 standard deviations of the mean, and approximately 99.73% of its values are within ± 3 standard deviations of the mean. Although many variables in business are closely approximated by the normal distribution, do not think that all variables can be approximated by the normal distribution.

In Section 6.3, you learned about various methods for evaluating normality in order to determine whether the normal distribution is a reasonable mathematical model to use in specific situations. In Section 6.4, you learned about another continuous distribution, the uniform distribution, that was not normal. Chapter 7 uses the normal distribution to develop the subject of statistical inference.

REFERENCES

1. Gunter, B. "Q-Q Plots." *Quality Progress* (February 1994): 81–86.
2. Hogg, R. V., J. T. McKean, and A. V. Craig. *Introduction to Mathematical Statistics*, 7th ed. New York: Pearson Education, 2013.
3. Kishnan, S. and R. Sitaraman. "Video stream quality impacts viewer behavior: inferring causality using quasi-experimental designs," in *Proceedings of the 2012 ACM conference on Internet measurement conference*: 211–224. New York: ACM.
4. Levine, D. M., P. Ramsey, and R. Smidt. *Applied Statistics for Engineers and Scientists Using Microsoft Excel and Minitab*. Upper Saddle River, NJ: Prentice Hall, 2001.
5. Miller, J. "Earliest Known Uses of Some of the Words of Mathematics." **jeff560.tripod.com/mathword.html**.
6. Pearl, R. "Karl Pearson, 1857–1936." *Journal of the American Statistical Association*, 31 (1936): 653–664.
7. Pearson, E. S. "Some Incidents in the Early History of Biometry and Statistics, 1890–94." *Biometrika* 52 (1965): 3–18.
8. Taleb, N. *The Black Swan*, 2nd ed. New York: Random House, 2010.
9. Walker, H. "The Contributions of Karl Pearson." *Journal of the American Statistical Association* 53 (1958): 11–22.

KEY EQUATIONS

Normal Probability Density Function

$$f(X) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(1/2)[(X-\mu)/\sigma]^2} \quad \textbf{(6.1)}$$

Z Transformation Formula

$$Z = \frac{X - \mu}{\sigma} \quad \textbf{(6.2)}$$

Finding an *X* Value Associated with a Known Probability

$$X = \mu + Z\sigma \quad \textbf{(6.3)}$$

Uniform Probability Density Function

$$f(X) = \frac{1}{b - a} \quad \textbf{(6.4)}$$

Mean of the Uniform Distribution

$$\mu = \frac{a + b}{2} \quad \textbf{(6.5)}$$

Variance and Standard Deviation of the Uniform Distribution

$$\sigma^2 = \frac{(b - a)^2}{12} \quad \textbf{(6.6a)}$$

$$\sigma = \sqrt{\frac{(b - a)^2}{12}} \quad \textbf{(6.6b)}$$

KEY TERMS

CHECKING YOUR UNDERSTANDING

6.28 How do you find the area between two values under the normal curve?

6.29 How do you find the *X* value that corresponds to a given percentile of the normal distribution?

6.30 What are some of the distinguishing properties of a normal distribution?

6.31 How does the shape of the normal distribution differ from the shapes of the uniform and exponential distributions?

6.32 How can you use the normal probability plot to evaluate whether a set of data is normally distributed?

CHAPTER REVIEW PROBLEMS

6.33 An industrial sewing machine uses ball bearings that are targeted to have a diameter of 0.75 inch. The lower and upper specification limits under which the ball bearings can operate are 0.74 inch and 0.76 inch, respectively. Past experience has indicated that the actual diameter of the ball bearings is approximately normally distributed, with a mean of 0.753 inch and a standard deviation of 0.004 inch. What is the probability that a ball bearing is

a. between the target and the actual mean?
b. between the lower specification limit and the target?
c. above the upper specification limit?
d. below the lower specification limit?
e. Of all the ball bearings, 93% of the diameters are greater than what value?

6.34 The fill amount in 2-liter soft drink bottles is normally distributed, with a mean of 2.0 liters and a standard deviation of 0.05 liter. If bottles contain less than 95% of the listed net content (1.90 liters, in this case), the manufacturer may be subject to penalty by the state office of consumer affairs. Bottles that have a net content above 2.10 liters may cause excess spillage upon opening. What proportion of the bottles will contain

a. between 1.90 and 2.0 liters?
b. between 1.90 and 2.10 liters?
c. below 1.90 liters or above 2.10 liters?
d. At least how much soft drink is contained in 99% of the bottles?
e. Ninety-nine percent of the bottles contain an amount that is between which two values (symmetrically distributed) around the mean?

6.35 In an effort to reduce the number of bottles that contain less than 1.90 liters, the bottler in Problem 6.34 sets the filling machine so that the mean is 2.02 liters. Under these circumstances, what are your answers in Problem 6.34 (a) through (e)?

6.36 *Webrooming*, researching products online before buying them in store, has become the new norm for some consumers and contrasts with *showrooming*, researching products in a physical store before purchasing online. A recent study by Interactions reported that most shoppers have a specific spending limit in place while shopping online. Findings indicate that men spend an average of $250 online before they decide to visit a store.

Source: Data extracted from **bit.ly/1JEcmqh**.

Assume that the spending limit is normally distributed and that the standard deviation is $20.

a. What is the probability that a male spent less than $210 online before deciding to visit a store?
b. What is the probability that a male spent between $270 and $300 online before deciding to visit a store?
c. Ninety percent of the amounts spent online by a male before deciding to visit a store are less than what value?
d. Eighty percent of the amounts spent online by a male before deciding to visit a store are between what two values symmetrically distributed around the mean?

Suppose that the spending limit follows a uniform distribution between $200 and $300.

e. What is the probability that a male spent less than $210 online before deciding to visit a store?
f. What is the probability that a male spent between $270 and $300 online before deciding to visit a store?
g. Compare the results of (a) and (b) to those of (e) and (f).

6.37 The file **RateBeerTop50** contains the percentage alcohol, alcohol by volume (abv), number of ratings (count), rank, and average score as of July 2018 for the top 50 beers of the world. Determine whether number of ratings, alcohol by volume and average scores appear to be approximately normally distributed. Support your decisions through the use of appropriate statistics and graphs.

Source: Data extracted from **https://bit.ly/2BcTBMo**.

6.38 The evening manager of a restaurant was very concerned about the length of time some customers were waiting in line to be seated. She also had some concern about the seating times—that is, the length of time between when a customer is seated and the time he or she leaves the restaurant. Over the course of one week, 100 customers (no more than 1 per party) were randomly selected, and their waiting and seating times (in minutes) were recorded in **Wait**.

a. Think about your favorite restaurant. Do you think waiting times more closely resemble a uniform, an exponential, or a normal distribution?
b. Again, think about your favorite restaurant. Do you think seating times more closely resemble a uniform, an exponential, or a normal distribution?
c. Construct a histogram and a normal probability plot of the waiting times. Do you think these waiting times more closely resemble a uniform, an exponential, or a normal distribution?
d. Construct a histogram and a normal probability plot of the seating times. Do you think these seating times more closely resemble a uniform, an exponential, or a normal distribution?

6.39 The major stock market indexes had strong results in 2016. The mean one-year return for stocks in the S&P 500, a group of 500 very large companies, was +9.54%. The mean one-year return for