6.1 Graphs of Normal Probability Distributions

6.2 Standard Units and Areas Under the Standard Normal Distribution

6.3 Areas Under Any Normal Curve

6.4 Normal Approximation to the Binomial Distribution

One cannot escape the feeling that these mathematical formulas have an independent existence and an intelligence of their own, that they are wiser than we are, wiser even than their discoverers, that we get more out of them than was originally put into them.

—Heinrich Hertz

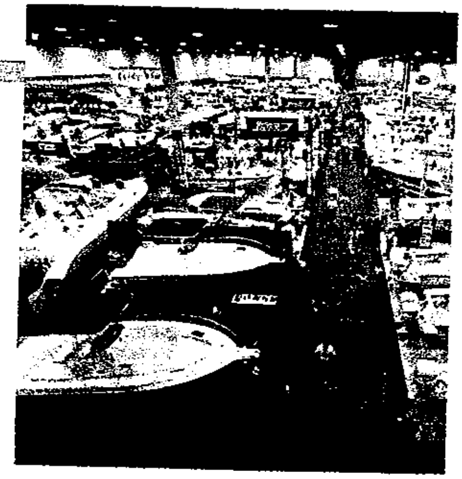
How can it be that mathematics, a product of human thought independent of experience, is so admirably adapted to the objects of reality?

—Albert Einstein

Heinrich Hertz (1857–1894) was a pioneer in the study of radio waves. His work and the later work of Maxwell and Marconi led the way to modern radio, television, and radar. Albert Einstein is world renowned for his great discoveries in relativity and quantum mechanics. Everyone who has worked in both mathematics and real-world applications cannot help but marvel at how the "pure thought" of the mathematical sciences can predict and explain events in other realms. In this chapter, we will study the most important type of probability distribution in all of mathematical statistics: the normal distribution. Why is the normal distribution so important? Two of the reasons are that it applies to a wide variety of situations and that other distributions tend to become normal under certain conditions.

For on-line student resources, visit the Brase/Brase, Understandable Statistics, 9th edition web site at **college.hmco.com/pic/braseUS9e**.

NORMAL DISTRIBUTIONS



PREVIEW QUESTIONS

What are some characteristics of a normal distribution? What does the empirical rule tell you about data spread around the mean? How can this information be used in quality control? (SECTION 6.1)

Can you compare apples and oranges, or maybe elephants and butterflies? In most cases, the answer is no—unless you first standardize *your measurements. What are a* standard normal distribution *and a* standard z score? (SECTION 6.2)

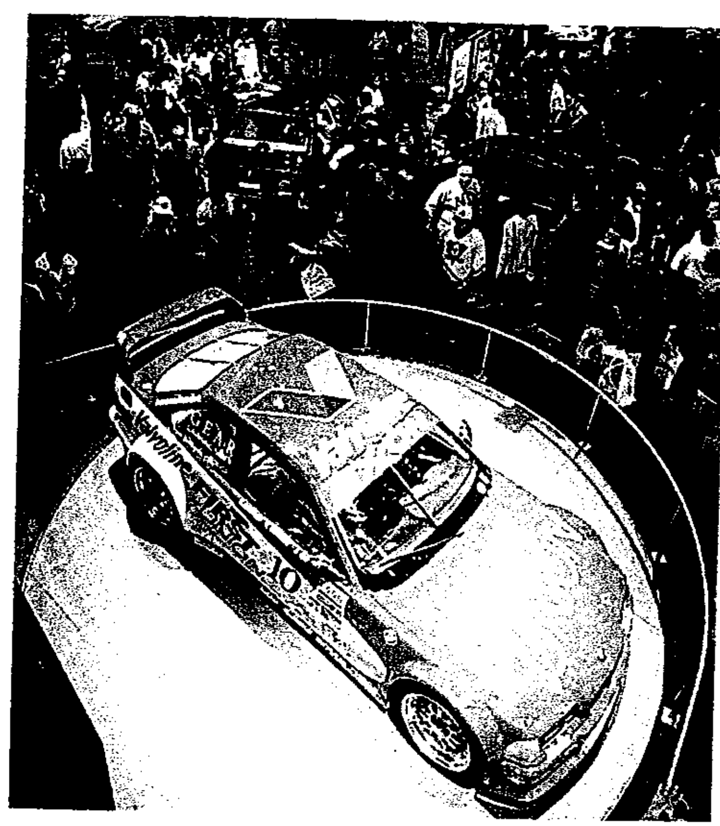
How do you convert any normal distribution to a standard normal distribution? How do you find probabilities of "standardized events"? (SECTION 6.3)

The binomial and normal distributions are two of the most important probability distributions in statistics. Under certain limiting conditions, the binomial can be thought to evolve (or envelope) into the normal distribution. How can you apply this concept in the real world? (SECTION 6.4)

FOCUS PROBLEMS

Large Auditorium Shows: How Many Will Attend?

1. For many years, Denver, as well as most other cities, has hosted large exhibition shows in big auditoriums. These shows include house and gardening shows, fishing and hunting shows, car shows, boat shows, Native American powwows, and so on. Information provided by Denver exposition sponsors indicates that most shows have an average attendance of about 8000 people per day with an estimated standard deviation of about 500 people. Suppose that the daily attendance figures follow a normal distribution.
   a) What is the probability that the daily attendance will be fewer than 7200 people?
   b) What is the probability that the daily attendance will be more than 8900 people?
   c) What is the probability that the daily attendance will be between 7200 and 8900 people?



2. Most exhibition shows open in the morning and close in the late evening. A study of Saturday arrival times

showed that the average arrival time was 3 hours and 48 minutes after doors open, and the standard deviation was estimated at about 52 minu Suppose that the arrival times follow a normal distribution.

(a) At what time after the doors open will 90% of the people who are com to the Saturday show have arrived?

(b) At what time after the doors open will only 15% of the people who coming to the Saturday show have arrived?

(c) Do you think the probability distribution of arrival times for Friday mig be different from the distribution of arrival times for Saturday? Expla (See Problems 36 and 37 of Section 6.3.)

SECTION 6.1

Graphs of Normal Probability Distributions

FOCUS POINTS

- Graph a normal curve and summarize its important properties.
- Apply the empirical rule to solve real-world problems.
- Use control limits to construct control charts. Examine the chart for three possible out-of-control signals.

One of the most important examples of a continuous probability distribution is the *normal distribution*. This distribution was studied by the French mathematician Abraham de Moivre (1667–1754) and later by the German mathematician Carl Friedrich Gauss (1777–1855), whose work is so important that the normal distribution is sometimes called *Gaussian*. The work of these mathematicians provided a foundation on which much of the theory of statistical inference is based.

Applications of a normal probability distribution are so numerous that some mathematicians refer to it as "a veritable Boy Scout knife of statistics." However, before we can apply it, we must examine some of the properties of a normal distribution.

A rather complicated formula, presented later in this section, defines a normal distribution in terms of μ and σ, the mean and standard deviation of the population distribution. It is only through this formula that we can verify if a distribution is normal. However, we can look at the graph of a normal distribution and get a good pictorial idea of some of the essential features of any normal distribution.

Normal curve

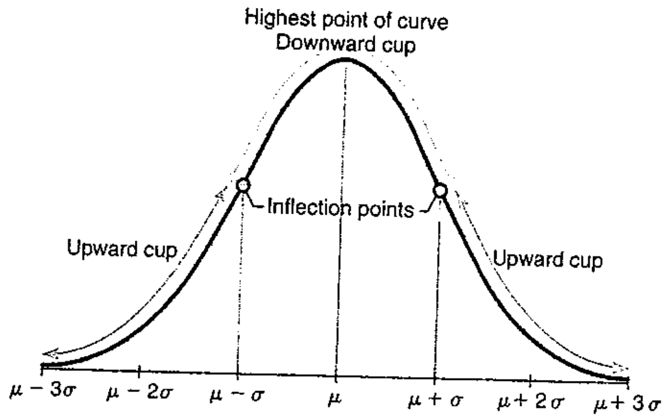
The graph of a normal distribution is called a *normal curve*. It possesses a shape very much like the cross section of a pile of dry sand. Because of its shape, blacksmiths would sometimes use a pile of dry sand in the construction of a mold for a bell. Thus the normal curve is also called a *bell-shaped curve* (see Figure 6-1).

We see that a general normal curve is smooth and symmetrical about the vertical line extending upward from the mean μ. Notice that the highest point of the curve occurs over μ. If the distribution were graphed on a piece of sheet metal, cut out, and placed on a knife edge, the balance point would be at μ. We also see that the curve tends to level out and approach the horizontal (x axis) like a glider making a landing. However, in mathematical theory, such a glider would never quite finish its landing because a normal curve never touches the horizontal axis.

The parameter σ controls the spread of the curve. The curve is quite close to the horizontal axis at $\mu + 3\sigma$ and $\mu - 3\sigma$. Thus, if the standard deviation σ is large, the curve will be more spread out; if it is small, the curve will be more peaked. Figure 6-1 shows the normal curve cupped downward for an interval on either side of the mean μ. Then it begins to cup upward as we go to the lower part of the bell. The exact places where the *transition* between the upward and downward cupping occur are above the points $\mu + \sigma$ and $\mu - \sigma$. In the terminology of calculus, transition points such as these are called *inflection points*.

FIGURE 6-1

A Normal Curve



Important properties of a normal curve

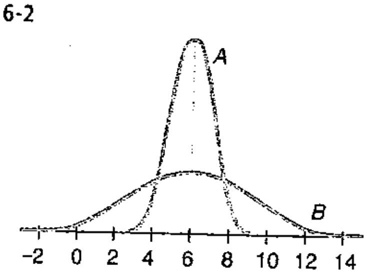
1. The curve is bell-shaped, with the highest point over the mean μ.
2. The curve is symmetrical about a vertical line through μ.
3. The curve approaches the horizontal axis but never touches or crosses it.
4. The inflection (transition) points between cupping upward and downward occur above $\mu + \sigma$ and $\mu - \sigma$.

The parameters that control the shape of a normal curve are the mean μ and the standard deviation σ. When both μ and σ are specified, a specific normal curve is determined. In brief, μ locates the balance point and σ determines the extent of the spread.

GUIDED EXERCISE 1 *Identify μ and σ on a normal curve*

Look at the normal curves in Figure 6-2.

FIGURE 6-2



(a) Do these distributions have the same mean? If so, what is it?

→ The means are the same, since both graphs have the high point over 6. $\mu = 6$.

(b) One of the curves corresponds to a normal distribution with $\sigma = 3$ and the other to one with $\sigma = 1$. Which curve has which σ?

→ Curve A has $\sigma = 1$ and curve B has $\sigma = 3$. (Since curve B is more spread out, it has the larger σ value.)

COMMENT The normal distribution curve is always above the horizontal axis. The area beneath the curve and above the axis is exactly 1. As such, the normal distribution curve is an example of a *density curve*. The formula used to generate the shape of the normal distribution curve is called the standard deviation σ, the formula for the

Normal density function if x is a normal random variable with mean μ and

$$f(x) = \frac{e^{(-1/2)((x-\mu)/\sigma)^2}}{\sigma\sqrt{2\pi}}$$

In this text, we will not use this formula explicitly. However, we will use tables of areas based on the normal density function.

The total area under any normal curve studied in this book will *always* be 1. The graph of the normal distribution is important because the portion of the *area* under the curve above a given interval represents the *probability* that a measurement will lie in that interval.

Empirical rule

In Section 3.2, we studied Chebyshev's theorem. This theorem gives us information about the *smallest* proportion of data that lies within 2, 3, or k standard deviations of the mean. This result applies to *any* distribution. However, for normal distributions, we can get a much more precise result, which is given by the *empirical rule.*

Empirical rule

For a distribution that is symmetrical and bell-shaped (in particular, for a normal distribution):

Approximately 68% of the data values will lie within one standard deviation on each side of the mean.

Approximately 95% of the data values will lie within two standard deviations on each side of the mean.

Approximately 99.7% (or almost all) of the data values will lie within three standard deviations on each side of the mean.

The preceding statement is called the *empirical rule* because, for symmetrical, bell-shaped distributions, the given percentages are observed in practice. Furthermore, for the normal distribution, the empirical rule is a direct consequence of the very nature of the distribution (see Figure 6-3). Notice that the empirical rule is a stronger statement than Chebyshev's theorem in that it gives *definite percentages,* not just lower limits. Of course, the empirical rule applies only to normal or symmetrical, bell-shaped distributions, whereas Chebyshev's theorem applies to all distributions.

FIGURE 6-3

Area Under a Normal Curve

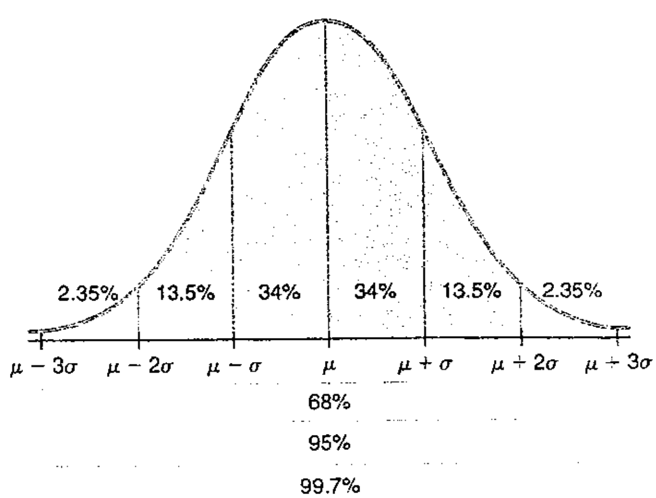
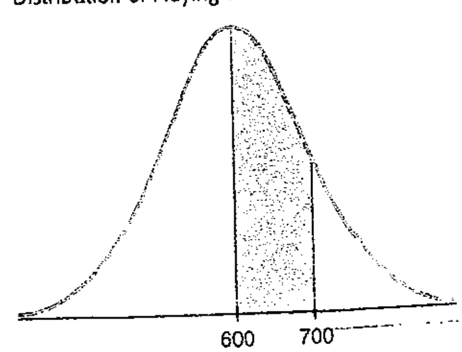


FIGURE 6-4

Distribution of Playing Times



for μ and σ should be reasonably close to actual data taken when the process was operating at a satisfactory production level. In Example 2, we will make a control chart; then we will discuss ways to analyze it to see if a process or service is "in control."

EXAMPLE 2

Control chart

Susan Tamara is director of personnel at the Antlers Lodge in Denali National Park, Alaska. Every summer Ms. Tamara hires many part-time employees from all over the United States. Most are college students seeking summer employment. One of the biggest activities for the lodge staff is that of "making up" the rooms each day. Although the rooms are supposed to be ready by 3:30 P.M., there are always some rooms not made up by this time because of high personnel turnover.

Every 15 days Ms. Tamara has a general staff meeting at which she shows a control chart of the number of rooms not made up by 3:30 P.M. each day. From extensive experience, Ms. Tamara is aware that the distribution of rooms not made up by 3:30 P.M. is approximately normal, with mean $\mu = 19.3$ rooms and standard deviation $\sigma = 4.7$ rooms. This distribution of x values is acceptable to the top administration of Antlers Lodge. For the past 15 days, the housekeeping unit has reported the number of rooms not ready by 3:30 P.M. (Table 6-1). Make a control chart for these data.

Mt. McKinley, Denali National Park

Number of Rooms x Not Made Up by 3:30 P.M.

Day	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x	11	20	25	23	16	19	8	25	17	20	23	29	18	14	10

SOLUTION: A control chart for a variable x is a plot of the observed x values (vertical scale) in time sequence order (the horizontal scale represents time). Place horizontal lines at

the mean $\mu = 19.3$

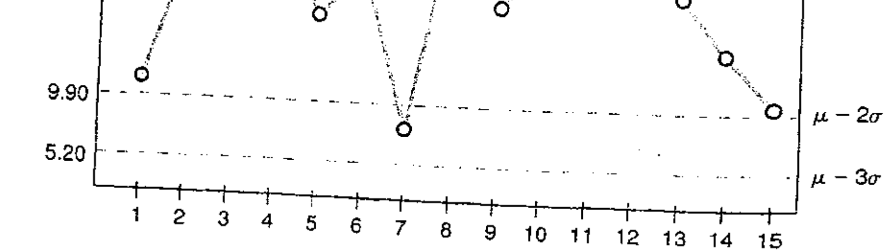
the control limits $\mu \pm 2\sigma = 19.3 \pm 2(4.7)$, or 9.90 and 28.70

the control limits $\mu \pm 3\sigma = 19.3 \pm 3(4.7)$, or 5.20 and 33.40

Then plot the data from Table 6-1. (See Figure 6-7.)

FIGURE 6-7

Number of Rooms Not Made Up by 3:30 P.M.



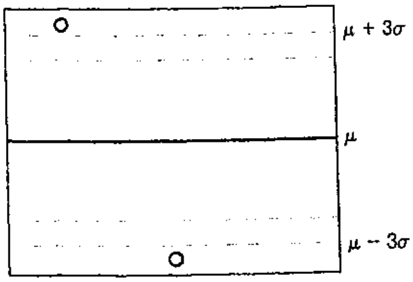
Once we have made a control chart, the main q…
As time goes on, is the x variable continuing in this same distribution, or is the distribution of x values changing? If the x distribution is continuing in more or less the same manner, we say it is *in statistical control*. If it is not, we say it is *out of control*.

Out-of-control warning signals

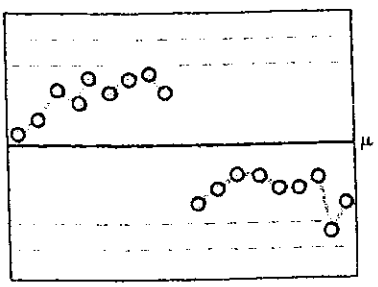
Many popular methods can set off a warning signal that a process is out of control. Remember, a random variable x is said to be *out of control* if successive time measurements of x indicate that it is no longer following the target probability distribution. We will assume that the target distribution is (approximately) normal and has (user-set) target values for μ and σ.

Three of the most popular warning signals are described next.

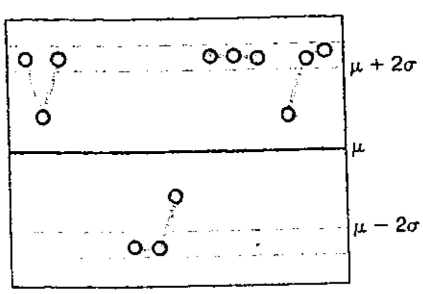
Out-of-Control Signal I



Out-of-Control Signal II



Out-of-Control Signal III



Out-of-control signals

1. **Out-of-Control Signal I: One point falls beyond the 3σ level**
   What is the probability that signal I will be a false alarm? By the empirical rule, the probability that a point lies within 3σ of the mean is 0.997. The probability that signal I will be a false alarm is $1 - 0.997 = 0.003$. Remember, a false alarm means that the x distribution is really on the target distribution, and we simply have a very rare (probability of 0.003) event.
2. **Out-of-Control Signal II: A run of nine consecutive points on one side of the center line (the line at target value μ)**
   To find the probability that signal II is a false alarm, we observe that if the x distribution and the target distribution are the same, then there is a 50% chance that the x values will lie above or below the center line at μ. Because the samples are (time) independent, the probability of a run of nine points on one side of the center line is $(0.5)^9 = 0.002$. If we consider both sides, this probability becomes 0.004. Therefore, the probability that signal II is a false alarm is approximately 0.004.
3. **Out-of-Control Signal III: At least two of three consecutive points lie beyond the 2σ level on the same side of the center line**
   To determine the probability that signal III will produce a false alarm, we use the empirical rule. By this rule, the probability that an x value will be above the 2σ level is about 0.023. Using the binomial probability distribution (with success being the point is above 2σ), the probability of two or more successes out of three trials is

$$\frac{3!}{2!1!}(0.023)^2(0.997) + \frac{3!}{3!0!}(0.023)^3 \approx 0.002$$

Taking into account *both* above and below the center line, it follows that the probability that signal III is a false alarm is about 0.004.

Remember, a control chart is only a warning device, and it is possible to get a false alarm. A false alarm happens when one (or more) of the out-of-control signals occurs, but the x distribution is really on the target or assigned distribution. In this case, we simply have a rare event (probability of 0.003 or 0.004). In practice, whenever a control chart indicates that a process is out of control, it is usually a good precaution to examine what is going on. If the process is out of control, corrective steps can be taken before things get a lot worse. The rare false alarm is a small price to pay if we can avert what might become real trouble.

3. *Critical Thinking* Look at the two normal curves in Figures 6-12 and 6-13. Which has the larger standard deviation? What is the mean of the curve in Figure 6-12? What is the mean of the curve in Figure 6-13?

FIGURE 6-12

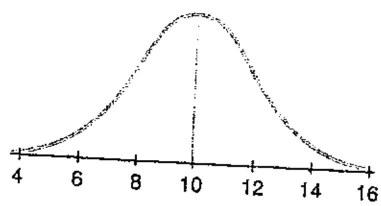
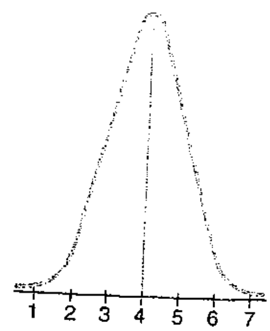


FIGURE 6-13



4. *Critical Thinking* Sketch a normal curve
   (a) with mean 15 and standard deviation 2.
   (b) with mean 15 and standard deviation 3.
   (c) with mean 12 and standard deviation 2.
   (d) with mean 12 and standard deviation 3.
   (e) Consider two normal curves. If the first one has a larger mean than the second one, must it have a larger standard deviation as well? Explain your answer.

5. *Critical Thinking* What percentage of the area under the normal curve lies
   (a) to the left of μ?
   (b) between $\mu - \sigma$ and $\mu + \sigma$?
   (c) between $\mu - 3\sigma$ and $\mu + 3\sigma$?

6. *Critical Thinking* What percentage of the area under the normal curve lies
   (a) to the right of μ?
   (b) between $\mu - 2\sigma$ and $\mu + 2\sigma$?
   (c) to the right of $\mu + 3\sigma$?

7. *Distribution: Heights of Coeds* Assuming that the heights of college women are normally distributed with mean 65 inches and standard deviation 2.5 inches (based on information from *Statistical Abstract of the United States,* 112th Edition), answer the following questions. (*Hint:* Use Problems 5 and 6 and Figure 6-3.)
   (a) What percentage of women are taller than 65 inches?
   (b) What percentage of women are shorter than 65 inches?
   (c) What percentage of women are between 62.5 inches and 67.5 inches?
   (d) What percentage of women are between 60 inches and 70 inches?

8. *Distribution: Rhode Island Red Chicks* The incubation time for Rhode Island Red chicks is normally distributed with a mean of 21 days and standard deviation of approximately 1 day (based on information from *World Book Encyclopedia*). Look at Figure 6-3 and answer the following questions. If 1000 eggs are being incubated, how many chicks do we expect will hatch
   (a) in 19 to 23 days?
   (b) in 20 to 22 days?
   (c) in 21 days or fewer?
   (d) in 18 to 24 days? (Assume all eggs eventually hatch.)

   (*Note:* In this problem, let us agree to think of a single day or a succession of days as a continuous interval of time.)

Excavation Project: Summer 1989 Excavations at Burnt Mesa Pueblo, edited by Kohler, Washington State University Department of Anthropology). The distribution of dates was more or less mound-shaped and symmetrical about the mean. Use the empirical rule to

(a) estimate a range of years centered about the mean in which about 68% of the data (tree-ring dates) will be found.

(b) estimate a range of years centered about the mean in which about 95% of the data (tree-ring dates) will be found.

(c) estimate a range of years centered about the mean in which almost all the data (tree-ring dates) will be found.

10. *Vending Machine: Soft Drinks* A vending machine automatically pours soft drinks into cups. The amount of soft drink dispensed into a cup is normally distributed with a mean of 7.6 ounces and standard deviation of 0.4 ounce. Examine Figure 6-3 and answer the following questions.

(a) Estimate the probability that the machine will overflow an 8-ounce cup.

(b) Estimate the probability that the machine will not overflow an 8-ounce cup.

(c) The machine has just been loaded with 850 cups. How many of these do you expect will overflow when served?

11. *Pain Management: Laser Therapy* "Effect of Helium-Neon Laser Auriculotherapy on Experimental Pain Threshold" is the title of an article in the journal *Physical Therapy* (Vol. 70, No. 1, pp. 24–30). In this article, laser therapy was discussed as a useful alternative to drugs in pain management of chronically ill patients. To measure pain threshold, a machine was used that delivered low-voltage direct current to different parts of the body (wrist, neck, and back). The machine measured current in milliamperes (mA). The pretreatment experimental group in the study had an average threshold of pain (pain was first detectable) at $\mu = 3.15$ mA with standard deviation $\sigma = 1.45$ mA. Assume that the distribution of threshold pain, measured in milliamperes, is symmetrical and more or less mound-shaped. Use the empirical rule to

(a) estimate a range of milliamperes centered about the mean in which about 68% of the experimental group will have a threshold of pain.

(b) estimate a range of milliamperes centered about the mean in which about 95% of the experimental group will have a threshold of pain.

12. *Control Charts: Yellowstone National Park* Yellowstone Park Medical Services (YPMS) provides emergency health care for park visitors. Such health care includes treatment for everything from indigestion and sunburn to more serious injuries. A recent issue of *Yellowstone Today* (National Park Service Publication) indicated that the average number of visitors treated each day by YPMS was 21.7. The estimated standard deviation was 4.2 (summer data). The distribution of numbers treated is approximately mound-shaped and symmetrical.

(a) For a 10-day summer period, the following data show the number of visitors treated each day by YPMS:

Day	1	2	3	4	5	6	7	8	9	10
Number treated	25	19	17	15	20	24	30	19	16	[illegible]

Make a control chart for the daily number of visitors treated by YPMS, and plot the data on the control chart. Do the data indicate that the number of visitors treated by YPMS is "in control"? Explain your answer.

(b) For another 10-day summer period, the following data were obtained:

Day	1	2	3	4	5	6	7	8	9	10
Number treated	20	15	12	21	24	28	32	36	35	[illegible]

Make a control chart, and plot the data on the chart. Do the data indicate that the number of visitors treated by YPMS is "in control" or "out of control"? Explain your answer. Identify all out-of-control signals by type (I, II, or III). If you were the park superintendent, do you think YPMS might need some (temporary) extra help? Explain.

13. *Control Charts: Bank Loans* Tri-County Bank is a small independent bank in central Wyoming. This is a rural bank that makes loans on items as small as horses and pickup trucks to items as large as ranch land. Total monthly loan requests are used by bank officials as an indicator of economic business conditions in this rural community. The mean monthly loan request for the past several years has been 615.1 (in thousands of dollars) with a standard deviation of 11.2 (in thousands of dollars). The distribution of loan requests is approximately mound-shaped and symmetrical.

(a) For 12 months, the following monthly loan requests (in thousands of dollars) were made to Tri-County Bank:

Month	1	2	3	4	5	6
Loan request	619.3	625.1	610.2	614.2	630.4	615.9

Month	7	8	9	10	11	12
Loan request	617.2	610.1	592.7	596.4	585.1	588.2

Make a control chart for the total monthly loan requests, and plot the preceding data on the control chart. From the control chart, would you say the local business economy is heating up or cooling down? Explain your answer by referring to any trend you may see on the control chart. Identify all out-of-control signals by type (I, II, or III).

(b) For another 12-month period, the following monthly loan requests (in thousands of dollars) were made to Tri-County Bank:

Month	1	2	3	4	5	6
Loan request	608.3	610.4	615.1	617.2	619.3	622.1

Month	7	8	9	10	11	12
Loan request	625.7	633.1	635.4	625.0	628.2	619.8

Make a control chart for the total monthly loan requests, and plot the preceding data on the control chart. From the control chart, would you say the local business economy is heating up, cooling down, or about normal? Explain your answer by referring to the control chart. Identify all out-of-control signals by type (I, II, or III).

14. *Control Charts: Motel Rooms* The manager of Motel 11 has 316 rooms in Palo Alto, California. From observation over a long period of time, she knows that on an average night, 268 rooms will be rented. The long-term standard deviation is 12 rooms. This distribution is approximately mound-shaped and symmetrical.

(a) For 10 consecutive nights, the following numbers of rooms were rented each night:

Night	1	2	3	4	5	6
Number of rooms	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Number of rooms	290	286	263	240

Make a control chart for the number of rooms rented each night, and plot the preceding data on the control chart. Looking at the control chart, would you say the number of rooms rented during this 10-night period has been unusually low? unusually high? about what was expected? Explain your answer. Identify all out-of-control signals by type (I, II, or III).

(b) For another 10 consecutive nights, the following numbers of rooms were rented each night:

Night	1	2	3	4	5	6
Number of rooms	238	245	261	269	273	250

Night	7	8	9	10
Number of rooms	241	230	215	217

Make a control chart for the number of rooms rented each night, and plot the preceding data on the control chart. Would you say the room occupancy has been high? low? about what was expected? Explain your answer. Identify all out-of-control signals by type (I, II, or III).

15. *Control Chart: Air Pollution* The visibility standard index (VSI) is a measure of Denver air pollution that is reported each day in the *Rocky Mountain News*. The index ranges from 0 (excellent air quality) to 200 (very bad air quality). During winter months, when air pollution is higher, the index has a mean of about 90 (rated as fair) with a standard deviation of approximately 30. Suppose that for 15 days, the following VSI measures were reported each day:

Day	1	2	3	4	5	6	7	8	9
VSI	80	115	100	90	15	10	53	75	80

Day	10	11	12	13	14	15
VSI	110	165	160	120	140	195

Make a control chart for the VSI, and plot the preceding data on the control chart. Identify all out-of-control signals (high or low) that you find in the control chart by type (I, II, or III).

SECTION 6.2

Standard Units and Areas Under the Standard Normal Distribution

FOCUS POINTS

- Given μ and σ, convert raw data to z scores.
- Given μ and σ, convert z scores to raw data.
- Graph the standard normal distribution, and find areas under the standard normal curve.

z Scores and Raw Scores

Normal distributions vary from one another in two ways: The mean μ may be located anywhere on the x axis, and the bell shape may be more or less spr according to the size of the standard deviation σ. The differences among

6. *Critical Thinking* Raul received a score of 80 on a history test for which the class mean was 70 with standard deviation 10. He received a score of 75 on a biology test for which the class mean was 70 with standard deviation 2.5. On which test did he do better relative to the rest of the class?

7. *z Scores: First Aid Course* The college Physical Education Department offered an Advanced First Aid course last semester. The scores on the comprehensive final exam were normally distributed, and the z scores for some of the students are shown below:

Robert, 1.10	Juan, 1.70	Susan, −2.00
Joel, 0.00	Jan, −0.80	Linda, 1.60

(a) Which of these students scored above the mean?
(b) Which of these students scored on the mean?
(c) Which of these students scored below the mean?
(d) If the mean score was $\mu = 150$ with standard deviation $\sigma = 20$, what was the final exam score for each student?

8. *z Scores: Fawns* Fawns between 1 and 5 months old in Mesa Verde National Park have a body weight that is approximately normally distributed with mean $\mu = 27.2$ kilograms and standard deviation $\sigma = 4.3$ kilograms (based on information from *The Mule Deer of Mesa Verde National Park*, by G. W. Mierau and J. L. Schmidt, Mesa Verde Museum Association). Let x be the weight of a fawn in kilograms. Convert each of the following x intervals to z intervals.
(a) $x < 30$ (b) $19 < x$ (c) $32 < x < 35$
Convert each of the following z intervals to x intervals.
(d) $-2.17 < z$ (e) $z < 1.28$ (f) $-1.99 < z < 1.44$
(g) If a fawn weighs 14 kilograms, would you say it is an unusually small animal? Explain using z values and Figure 6-15.
(h) If a fawn is unusually large, would you say that the z value for the weight of the fawn will be close to 0, −2, or 3? Explain.

9. *z Scores: Red Blood Cell Count* Let x = red blood cell (RBC) count in millions per cubic millimeter of whole blood. For healthy females, x has an approximately normal distribution with mean $\mu = 4.8$ and standard deviation $\sigma = 0.3$ (based on information from *Diagnostic Tests with Nursing Implications*, edited by S. Loeb, Springhouse Press). Convert each of the following x intervals to z intervals.
(a) $4.5 < x$ (b) $x < 4.2$ (c) $4.0 < x < 5.5$
Convert each of the following z intervals to x intervals.
(d) $z < -1.44$ (e) $1.28 < z$ (f) $-2.25 < z < -1.00$
(g) If a female had an RBC count of 5.9 or higher, would that be considered unusually high? Explain using z values and Figure 6-15.

10. *Normal Curve: Tree Rings* Tree-ring dates were used extensively in archaeological studies at Burnt Mesa Pueblo (*Bandelier Archaeological Excavation Project: Summer 1989 Excavations at Burnt Mesa Pueblo*, edited by Kohler, Washington State University Department of Anthropology). At one site on the mesa, tree-ring dates (for many samples) gave a mean date of μ_1 = year 1272 with standard deviation $\sigma_1 = 35$ years. At a second, removed site, the tree-ring dates gave a mean of μ_2 = year 1122 with standard deviation $\sigma_2 = 40$ years. Assume that both sites had dates that were approximately normally distributed. In the first area, an object was found and dated as x_1 = year 1250. In the second area, another object was found and dated as x_2 = year 1234.
(a) Convert both x_1 and x_2 to z values, and locate both of these values under the standard normal curve of Figure 6-15.
(b) Which of these two items is the more unusual as an archaeological find in its location?

In Problems 11–28, sketch the areas under the standard normal curve over the indicated intervals, and find the specified areas.

11. To the right of $z = 0$
12. To the left of $z = 0$
13. To the left of $z = -1.32$
14. To the left of $z = -0.47$
15. To the left of $z = 0.45$
16. To the left of $z = 0.72$
17. To the right of $z = 1.52$
18. To the right of $z = 0.15$
19. To the right of $z = -1.22$
20. To the right of $z = -2.17$
21. Between $z = 0$ and $z = 3.18$
22. Between $z = 0$ and $z = -1.93$
23. Between $z = -2.18$ and $z = 1.34$
24. Between $z = -1.40$ and $z = 2.03$
25. Between $z = 0.32$ and $z = 1.92$
26. Between $z = 1.42$ and $z = 2.17$
27. Between $z = -2.42$ and $z = -1.77$
28. Between $z = -1.98$ and $z = -0.$[illegible]

In Problems 29–48, let z be a random variable with a standard normal distribution. Find the indicated probability, and shade the corresponding area under the standard normal curve.

29. $P(z \le 0)$
30. $P(z \ge 0)$
31. $P(z \le -0.13)$
32. $P(z \le -2.15)$
33. $P(z \le 1.20)$
34. $P(z \le 3.20)$
35. $P(z \ge 1.35)$
36. $P(z \ge 2.17)$
37. $P(z \ge -1.20)$
38. $P(z \ge -1.50)$
39. $P(-1.20 \le z \le 2.64)$
40. $P(-2.20 \le z \le 1.04)$
41. $P(-2.18 \le z \le -0.42)$
42. $P(-1.78 \le z \le -1.23)$
43. $P(0 \le z \le 1.62)$
44. $P(0 \le z \le 0.54)$
45. $P(-0.82 \le z \le 0)$
46. $P(-2.37 \le z \le 0)$
47. $P(-0.45 \le z \le 2.73)$
48. $P(-0.73 \le z \le 3.12)$

SECTION 6.3

Areas Under Any Normal Curve

FOCUS POINTS

- Compute the probability of "standardized events."
- Find a z score from a given normal probability (inverse normal).
- Use the inverse normal to solve guarantee problems.

Normal Distribution Areas

In many applied situations, the original normal curve is not the standard normal curve. Generally, there will not be a table of areas available for the original normal curve. This does not mean that we cannot find the probability that a measurement x will fall into an interval from a to b. What we must do is *convert* the original measurements x, a, and b to z values.

PROCEDURE

How to work with normal distributions

To find areas and probabilities for a random variable x that follows a normal distribution with mean μ and standard deviation σ, convert x values to z values using the formula

$$z = \frac{x - \mu}{\sigma}$$

Then use Table 5 of Appendix II to find corresponding areas and probabilities.

EXAMPLE 7

Normal distribution probability

Let x have a normal distribution with $\mu = 10$ and $\sigma = 2$. Find the probability that an x value selected at random from this distribution is between 11 and 14. In symbols, find $P(11 \le x \le 14)$.

SOLUTION: Since probabilities correspond to areas under the distribution curve, we want to find the area under the x curve above the interval from $x = 11$ to $x = 14$. To do so, we will convert the x values to standard z values and then use Table 5 of Appendix II to find the corresponding area under the standard curve.

We use the formula

$$z = \frac{x - \mu}{\sigma}$$

to convert the given x interval to a z interval.

$$z_1 = \frac{11 - 10}{2} = 0.50 \qquad \text{(Use } x = 11, \mu = 10, \sigma = 2.\text{)}$$

$$z_2 = \frac{14 - 10}{2} = 2.00 \qquad \text{(Use } x = 14, \mu = 10, \sigma = 2.\text{)}$$

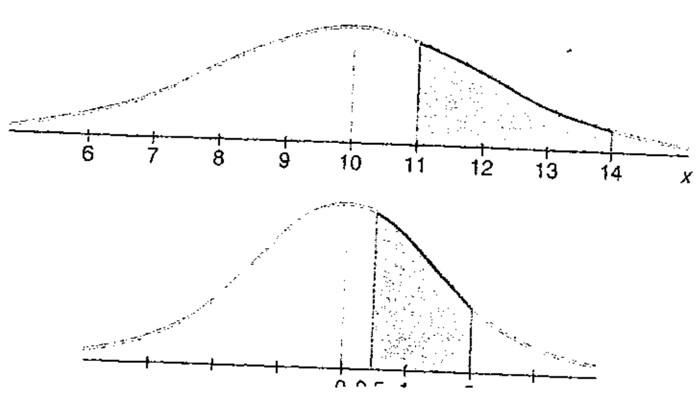
The corresponding areas under the x and z curves are shown in Figure 6-23. From Figure 6-23 we see that

$$\begin{aligned} P(11 \le x \le 14) &= P(0.50 \le z \le 2.00) \\ &= P(z \le 2.00) - P(z \le 0.50) \\ &= 0.9772 - 0.6915 \qquad \text{(From Table 5, Appendix II)} \\ &= 0.2857 \end{aligned}$$

The probability is 0.2857 that an x value selected at random from a normal distribution with mean 10 and standard deviation 2 lies between 11 and 14.

FIGURE 6-23

Corresponding Areas Under the x Curve and z Curve



GUIDED EXERCISE 7

Sunshine Stereo cassette decks have a deck life that is normally distributed with a mean of 2.3 years and a standard deviation of 0.4 year. What is the probability that a cassette deck will break down during the guarantee period of 2 years?

(a) Let x represent the life of a cassette deck. The statement that the cassette deck breaks during the 2-year guarantee period means the life is less than 2 years, or $x \le 2$. Convert this to a statement about z.

⇨ $z = \dfrac{x - \mu}{\sigma} = \dfrac{2 - 2.3}{0.4} = -0.75$

So, $x \le 2$ means $z \le -0.75$.

(b) Indicate the area to be found in Figure 6-24. Does this area correspond to the probability that $z \le -0.75$?

⇨ See Figure 6-25.
Yes, the shaded area does correspond to the probability that $z \le -0.75$.

FIGURE 6-24

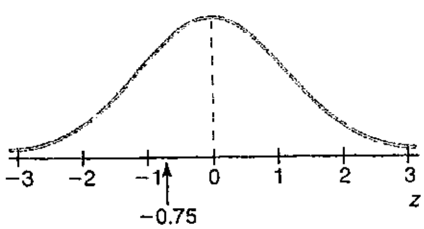
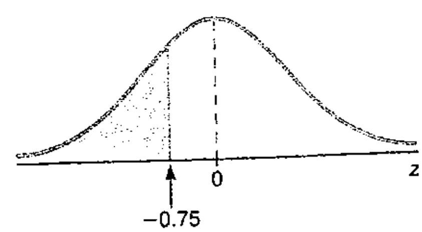


FIGURE 6-25 $z \le -0.75$



(c) Use Table 5 of Appendix II to find $P(z \le -0.75)$.

⇨ 0.2266

(d) What is the probability that the cassette deck will break before the end of the guarantee period? [*Hint:* $P(x \le 2) = P(z \le -0.75)$.]

⇨ The probability is

$$P(x \le 2) = P(z \le -0.75) = 0.2266$$

This means that the company will repair or replace about 23% of the cassette decks.

TECH NOTES

The TI-84Plus and TI-83Plus calculators, Excel, and Minitab all provide areas under any normal distribution. Excel and Minitab give the left-tail area to the left of a specified x value. The TI-84Plus/TI-83Plus has you specify an interval from a lower bound to an upper bound and provides the area under the normal curve for that interval. For example, to solve Guided Exercise 7 regarding the probability a cassette deck will break during the guarantee period, we find $P(x \le 2)$ for a normal distribution with $\mu = 2.3$ and $\sigma = 0.4$.

TI-84Plus/TI-83Plus Press the DISTR key, select 2:normalcdf (lower bound, upper bound, μ, σ) and press Enter. Type in the specified values. For a left-tail area, use a lower bound setting at about 4 standard deviations below the mean. Likewise, for a right-tail area, use an upper bound setting about 4 standard deviations above the mean. For our example, use a lower bound of $\mu - 4\sigma = 2.3 - 4(0.4) = 0.7$.

```
normalcdf(.7,2,2.3,
.4)
           .2265955934
```

COMMENT When we use Table 5 of Appendix II to find a z value corresponding to a given area, we usually use the nearest area value rather than interpolating between values. However, when the area value given is exactly halfway between two area values of the table, we use the z value halfway between the z values of the corresponding table areas. Example 9 demonstrates this procedure. However, this interpolation convention is not always used, especially if the area is changing slowly, as it does in the tail ends of the distribution. *When the z value corresponding to an area is smaller than -2, the standard convention is to use the z value corresponding to the smaller area. Likewise, when the z value is larger than 2, the standard convention is to use the z value corresponding to the larger area.* We will see an example of this special case in Guided Exercise 1 of the next chapter.

EXAMPLE 9

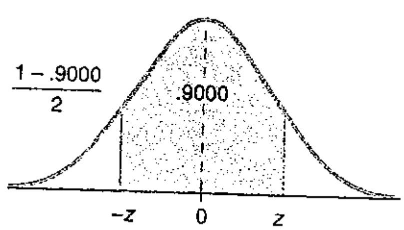
FIND z

Find the z value such that 90% of the area under the standard normal curve lies between $-z$ and z.

SOLUTION: Sketch a picture showing the described area (see Figure 6-28).

FIGURE 6-28

Area Between $-z$ and z Is 90%



We find the corresponding area in the left tail.

$$\text{(Area left of } -z) = \frac{1 - 0.9000}{2}$$
$$= 0.0500$$

Looking in Table 6-6, we see that 0.0500 lies exactly between areas 0.0495 and 0.0505. The halfway value between $z = -1.65$ and $z = -1.64$ is $z = -1.645$. Therefore, we conclude that 90% of the area under the standard normal curve lies between the z values -1.645 and 1.645.

Excerpt from Table 5 of Appendix II

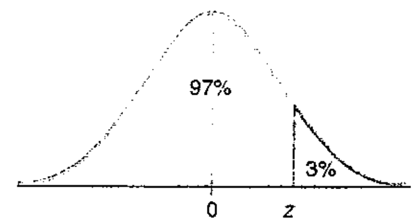
z	...	.04		.05
:				
−1.6		.0505	↑ 0.0500	.0495

GUIDED EXERCISE 8

Find the z value such that 3% of the area under the standard normal curve lies to the right of z.

(a) Draw a sketch of the standard normal distribution showing the described area (Figure 6-29).

⟹ FIGURE 6-29 3% of Total Area Lies to the Right of z



(b) Find the area to the left of z.

⟹ Area to the left of $z = 1 - 0.0300 = 0.9700$.

(c) Look up the area in Table 6-7 and find the corresponding z.

⟹ The closest area is 0.9699. This area is to the left of $z = 1.88$.

TABLE 6-7 **Excerpt from Table 5 of Appendix II**

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767

(d) Suppose the time to complete a test is normally distributed with $\mu = 40$ minutes and $\sigma = 5$ minutes. After how many minutes can we expect all but about 3% of the tests to be completed?

⟹ We are looking for an x value such that 3% of the normal distribution lies to the right of x. In part (c), we found that 3% of the standard normal curve lies to the right of $z = 1.88$. We convert $z = 1.88$ to an x value.

$$x = z\sigma + \mu$$
$$= 1.88(5) + 40 = 49.4 \text{ minutes}$$

All but about 3% of the tests will be complete after 50 minutes.

(e) Use Table 6-8 to find a z value such that 3% of the area under the standard normal curve lies to the left of z.

⟹ The closest area is 0.0301. This is the area to the left of $z = -1.88$.

TABLE 6-8 **Excerpt from Table 5 of Appendix II**

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294

(f) Compare the z value of part (c) with the z value of part (e). Is there any relationship between the z values?

⟹ One z value is the negative of the other. This result is expected because of the symmetry of the normal distribution.

TECH NOTES When we are given a z value and we find an area to the left of z, we are using a normal distribution function. When we are given an area to the left of z and we find the corresponding z, we are using an inverse normal distribution function. The TI-84Plus and TI-83Plus calculators, Excel, and Minitab all have inverse normal distribution

We see that $\bar{x} = 19.46$, median $= 19.5$, and $s = 2.705$.

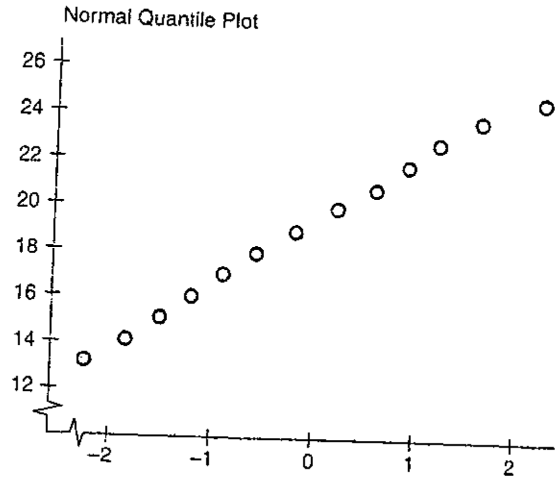
$$\text{Pearson's index} = \frac{3(19.46 - 19.5)}{2.705} \approx -0.04$$

Since the index is between -1 and 1, we detect no skewness. The data appear to be symmetric.

(c) Look at the normal quantile plot in Figure 6-31 and comment on normality.

FIGURE 6-31

Normal Quantile Plot



SOLUTION: The data fall close to a straight line, so the data appear to come from a normal distribution.

(d) Interpret the results.

SOLUTION: The histogram is roughly bell-shaped, there is only one outlier, Pearson's index does not indicate skewness, and the points on the normal quantile plot lie fairly close to a straight line. It appears that the data are likely from a distribution that is approximately normal.

Want to Be an Archaeologist?

Each year about 4500 students work with professional archaeologists in scientific research at the Crow Canyon Archaeological Center, Cortez, Colorado. In fact, Crow Canyon was included in The Princeton Review Guide to America's Top 100 Internships. *The nonprofit, multidisciplinary program at Crow Canyon enables students and laypeople with little or no background to get started in archaeological research. The only requirement is that you be interested in Native American culture and history. By the way, a knowledge of introductory statistics could come in handy for this internship. For more information about the program, visit the Brase/Brose statistics site at* **college.hmco.com/pic/braseUS9e** *and find the link to Crow Canyon.*

SECTION 6.3 PROBLEMS

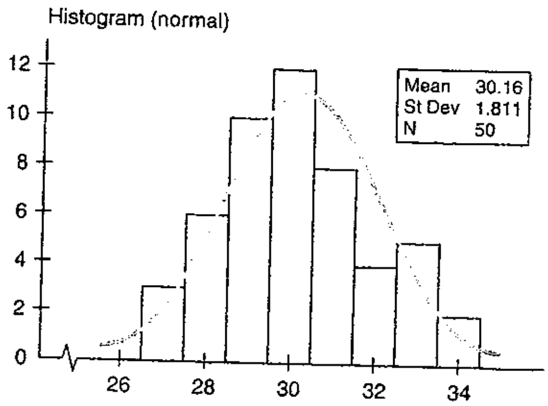
1. *Statistical Literacy* Consider a normal distribution with mean 30 and standard deviation 2. What is the probability a value selected at random from this distribution is greater than 30?

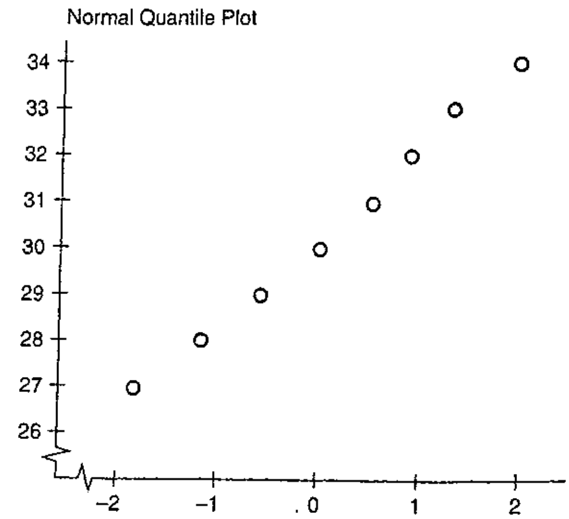
2. *Statistical Literacy* Suppose 5% of the area under the standard normal curve lies to the right of z. Is z positive or negative?

3. *Statistical Literacy* Suppose 5% of the area under the standard normal curve lies to the left of z. Is z positive or negative?

4. *Critical Thinking: Normality* Consider the following data. The summary statistics, histogram, and normal quantile plot were generated by Minitab.

27	27	27	28	28	28	28	28	28	29	29	29	29	29	29
29	29	29	29	30	30	30	30	30	30	30	30	30	30	30
30	31	31	31	31	31	31	31	31	32	32	32	32	33	33
33	33	33	34	34										





Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3
Data	50	0	30.160	0.256	1.811	27.000	29.000	30.000	31.000

Variable	Maximum
Data	34.000

(a) Does the histogram indicate normality for the data distribution? Explain.
(b) Does the normal quantile plot indicate normality for the data distribution? Explain.
(c) Compute the interquartile range and check for outliers.
(d) Compute Pearson's index. Does the index value indicate skewness?
(e) Using parts (a) through (d), would you say the data are from a normal distribution?

In Problems 5–14, assume that x has a normal distribution with the specified mean and standard deviation. Find the indicated probabilities.

5. $P(3 \le x \le 6)$; $\mu = 4$; $\sigma = 2$
6. $P(10 \le x \le 26)$; $\mu = 15$; $\sigma = 4$
7. $P(50 \le x \le 70)$; $\mu = 40$; $\sigma = 15$
8. $P(7 \le x \le 9)$; $\mu = 5$; $\sigma = 1.2$
9. $P(8 \le x \le 12)$; $\mu = 15$; $\sigma = 3.2$
10. $P(40 \le x \le 47)$; $\mu = 50$; $\sigma = 15$
11. $P(x \ge 30)$; $\mu = 20$; $\sigma = 3.4$
12. $P(x \ge 120)$; $\mu = 100$; $\sigma = 15$
13. $P(x \ge 90)$; $\mu = 100$; $\sigma = 15$
14. $P(x \ge 2)$; $\mu = 3$; $\sigma = 0.25$

In Problems 15–24, find the z value described and sketch the area described.

15. Find z such that 6% of the standard normal curve lies to the left of z.
16. Find z such that 5.2% of the standard normal curve lies to the left of z.
17. Find z such that 55% of the standard normal curve lies to the left of z.
18. Find z such that 97.5% of the standard normal curve lies to the left of z.
19. Find z such that 8% of the standard normal curve lies to the right of z.
20. Find z such that 5% of the standard normal curve lies to the right of z.
21. Find z such that 82% of the standard normal curve lies to the right of z.

(b) What is the probability that someone will keep a refrigerator fewer than 11 years before replacement?

(c) What is the probability that someone will keep a refrigerator more than 18 years before replacement?

(d) [illegible] Assume that the average life of a refrigerator is 14 years, with the standard deviation given in part (a) before it breaks. Suppose that a company guarantees refrigerators and will replace a refrigerator that breaks while under guarantee with a new one. However, the company does not want to replace more than 5% of the refrigerators under guarantee. For how long should the guarantee be made (rounded to the nearest tenth of a year)?

33. [illegible] The resting heart rate for an adult horse should average about $\mu = 46$ beats per minute with a (95% of data) range from 22 to 70 beats per minute, based on information from *The Merck Veterinary Manual* (a classic reference used in most veterinary colleges). Let x be a random variable that represents the resting heart rate for an adult horse. Assume that x has a distribution that is approximately normal.

(a) Estimate the standard deviation of the x distribution. *Hint:* See Problem 31.

(b) What is the probability that the heart rate is less than 25 beats per minute?

(c) What is the probability that the heart rate is greater than 60 beats per minute?

(d) What is the probability that the heart rate is between 25 and 60 beats per minute?

(e) *Inverse Normal Distribution* A horse whose resting heart rate is in the upper 10% of the probability distribution of heart rates may have a secondary infection or illness that needs to be treated. What is the heart rate corresponding to the upper 10% cutoff point of the probability distribution?

34. *Estimating the Standard Deviation: Veterinary Science* How much should a healthy kitten weigh? A healthy 10-week-old (domestic) kitten should weigh an average of $\mu = 24.5$ ounces with a (95% of data) range from 14 to 35 ounces. (See reference in Problem 33.) Let x be a random variable that represents the weight (in ounces) of a healthy 10-week-old kitten. Assume that x has a distribution that is approximately normal.

(a) Estimate the standard deviation of the x distribution. *Hint:* See Problem 31.

(b) What is the probability that a healthy 10-week-old kitten will weigh less than 14 ounces?

(c) What is the probability that a healthy 10-week-old kitten will weigh more than 33 ounces?

(d) What is the probability that a healthy 10-week-old kitten will weigh between 14 and 33 ounces?

(e) *Inverse Normal Distribution* A kitten whose weight is in the bottom 10% of the probability distribution of weights is called *undernourished*. What is the cutoff point for the weight of an undernourished kitten?

35. *Insurance: Satellites* A relay microchip in a telecommunications satellite has a life expectancy that follows a normal distribution with a mean of 90 months and a standard deviation of 3.7 months. When this computer-relay microchip malfunctions, the entire satellite is useless. A large London insurance company is going to insure the satellite for 50 million dollars. Assume that the only part of the satellite in question is the microchip. All other components will work indefinitely.

(a) *Inverse Normal Distribution* For how many months should the satellite be insured to be 99% confident that it will last beyond the insurance date?

(b) If the satellite is insured for 84 months, what is the probability that it will malfunction before the insurance coverage ends?

(c) If the satellite is insured for 84 months, what is the expected loss to the insurance company?

(d) If the insurance company charges \$3 million for 84 months of insurance, how much profit does the company expect to make?

36. *Focus Problem: Exhibition Show Attendance* The Focus Problem at the beginning of the chapter indicates that attendance at large exhibition shows in Denver averages about 8000 people per day, with standard deviation of about 500. Assume that the daily attendance figures follow a normal distribution.
(a) What is the probability that the daily attendance will be fewer than 7200 people?
(b) What is the probability that the daily attendance will be more than 8900 people?
(c) What is the probability that the daily attendance will be between 7200 and 8900 people?

37. *Focus Problem: Inverse Normal Distribution* Most exhibition shows open in the morning and close in the late evening. A study of Saturday arrival times showed that the average arrival time was 3 hours and 48 minutes after the doors opened, and the standard deviation was estimated at about 52 minutes. Assume that the arrival times follow a normal distribution.
(a) At what time after the doors open will 90% of the people who are coming to the Saturday show have arrived?
(b) At what time after the doors open will only 15% of the people who are coming to the Saturday show have arrived?
(c) Do you think the probability distribution of arrival times for Friday might be different from the distribution of arrival times for Saturday? Explain.

38. *Budget: Maintenance* The amount of money spent weekly on cleaning, maintenance, and repairs at a large restaurant was observed over a long period of time to be approximately normally distributed, with mean $\mu = \$615$ and standard deviation $\sigma = \$42$.
(a) If \$646 is budgeted for next week, what is the probability that the actual costs will exceed the budgeted amount?
(b) *Inverse Normal Distribution* How much should be budgeted for weekly repairs, cleaning, and maintenance so that the probability that the budgeted amount will be exceeded in a given week is only 0.10?

39. *Expand Your Knowledge: Conditional Probability* Suppose you want to eat lunch at a popular restaurant. The restaurant does not take reservations, so there is usually a waiting time before you can be seated. Let x represent the length of time waiting to be seated. From past experience, you know that the mean waiting time is $\mu = 18$ minutes with $\sigma = 4$ minutes. You assume that the x distribution is approximately normal.
(a) What is the probability that the waiting time will *exceed* 20 minutes, given that it has exceeded 15 minutes? *Hint:* Compute $P(x > 20 | x > 15)$.
(b) What is the probability that the waiting time will exceed 25 minutes, given that it has exceeded 18 minutes? *Hint:* Compute $P(x > 25 | x > 18)$.
(c) *Hint for solution:* Review item 6, conditional probability, in the summary of basic probability rules at the end of Section 4.2. Note that

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

and show that in part (a),

$$P(x > 20 | x > 15) = \frac{P((x > 20) \text{ and } (x > 15))}{P(x > 15)} = \frac{P(x > 20)}{P(x > 15)}$$

40. *Conditional Probability: Cycle Time* A cement truck delivers mixed cement to a large construction site. Let x represent the cycle time in minutes for the truck to leave the construction site, go back to the cement plant, fill up, and return to the construction site with another load of cement. From past experience, it is

known that the mean cycle time is $\mu = 45$ minutes with $\sigma = 12$ minutes. The x distribution is approximately normal.

(a) What is the probability that the cycle time will *exceed* 60 minutes, given that it has exceeded 50 minutes? *Hint:* See Problem 39, part (c).

(b) What is the probability that the cycle time will exceed 55 minutes, given that it has exceeded 40 minutes?

SECTION 6.4

Normal Approximation to the Binomial Distribution

FOCUS POINTS

- State the assumptions needed to use the normal approximation to the binomial distribution.
- Compute μ and σ for the normal approximation.
- Use the continuity correction to convert a range of r values to a corresponding range of normal x values.
- Convert the x values to a range of standardized z scores and find desired probabilities.

The probability that a new vaccine will protect adults from cholera is known to be 0.85. The vaccine is administered to 300 adults who must enter an area where the disease is prevalent. What is the probability that more than 280 of these adults will be protected from cholera by the vaccine?

This question falls into the category of a binomial experiment with the number of trials n equal to 300, the probability of success p equal to 0.85, and the number of successes r greater than 280. It is possible to use the formula for the binomial distribution to compute the probability that r is greater than 280. However, this approach would involve a number of tedious and long calculations. There is an easier way to do this problem, for under the conditions stated below, the normal distribution can be used to approximate the binomial distribution.

Criteria $np > 5$ and $nq > 5$

Normal approximation to the binomial distribution

Consider a binomial distribution where

n = number of trials

r = number of successes

p = probability of success on a single trial

$q = 1 - p$ = probability of failure on a single trial

If $np > 5$ and $nq > 5$, then r has a binomial distribution that is approximated by a normal distribution with

$$\mu = np \quad \text{and} \quad \sigma = \sqrt{npq}$$

Note: As n increases, the approximation becomes better.

Example 11 demonstrates that as n increases, the normal approximation to the binomial distribution improves.

EXAMPLE 11

BINOMIAL DISTRIBUTION GRAPHS

Graph the binomial distributions for which $p = 0.25$, $q = 0.75$, and the number of trials is first $n = 3$, then $n = 10$, then $n = 25$, and finally $n = 50$.

SOLUTION: The authors used a computer program to obtain the binomial distributions for the given values of p, q, and n. The results have been organized and graphed in Figures 6-32, 6-33, 6-34, and 6-35.

FIGURE 6-32

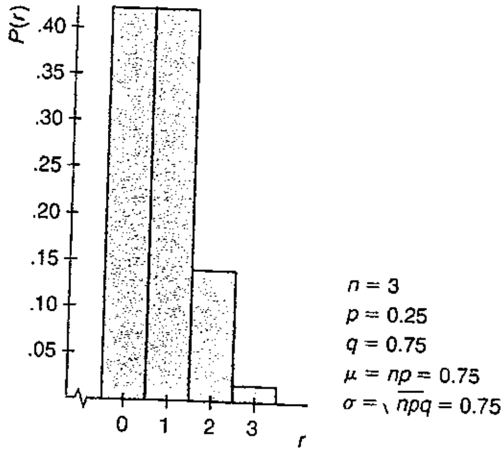


FIGURE 6-33

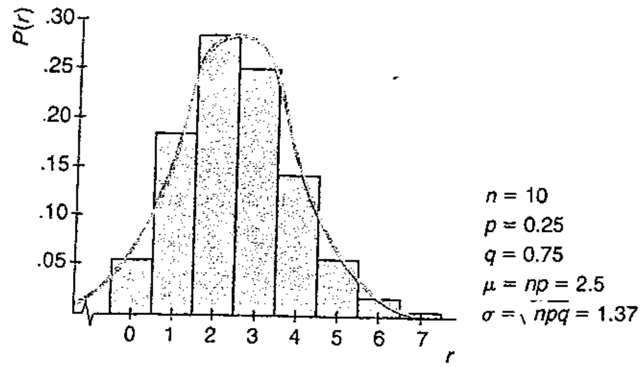


FIGURE 6-34 Good Normal Approximation; $np > 5$ and $nq > 5$

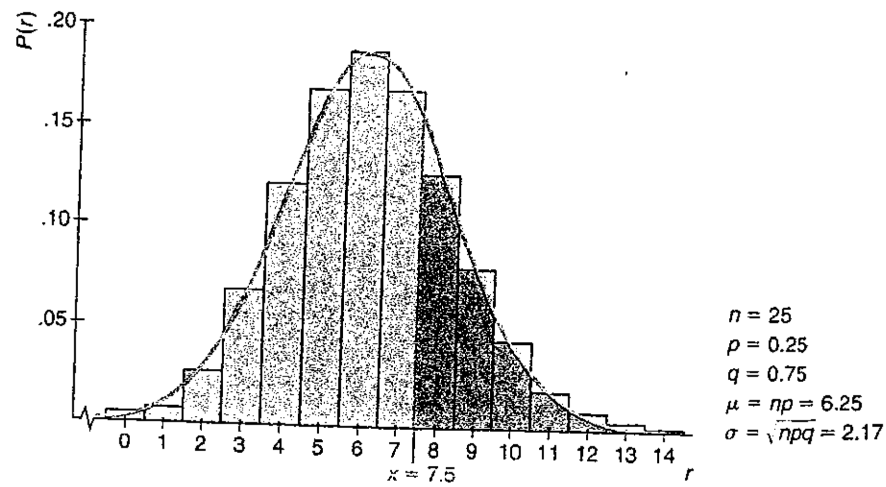
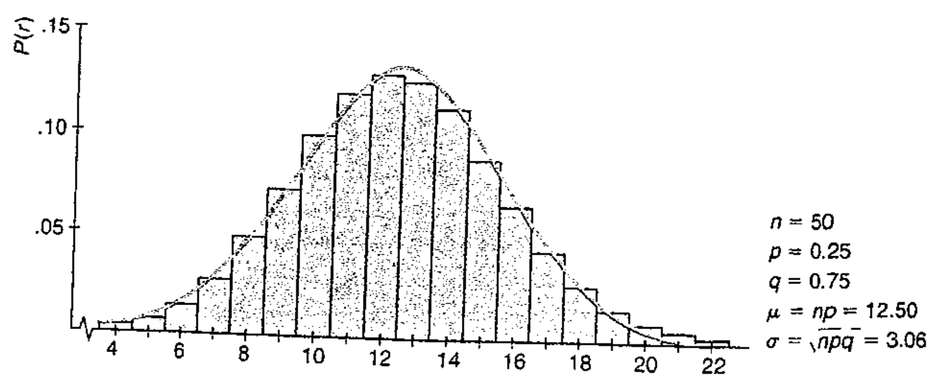
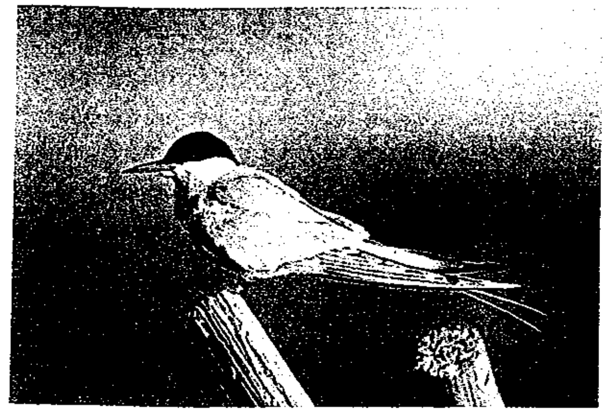


FIGURE 6-35 Good Normal Approximation; $np > 5$ and $nq > 5$



GUIDED EXERCISE 9 Continuity correction

From many years of observation, a biologist knows that the probability is only 0.65 that any given Arctic tern will survive the migration from its summer nesting area to its winter feeding grounds. A random sample of 500 Arctic terns were banded at their summer nesting area. Use the normal approximation to the binomial and the following steps to find the probability that between 310 and 340 of the banded Arctic terns will survive the migration. Let r be the number of surviving terns.



Arctic tern

(a) To approximate $P(310 \le r \le 340)$, we use the normal curve with $\mu =$ ______ and $\sigma =$ ______.

→ We use the normal curve with

$\mu = np = 500(0.65) = 325$ and

$\sigma = \sqrt{npq} = \sqrt{500(0.65)(0.35)} \approx 10.67$

(b) $P(310 \le r \le 340)$ is approximately equal to $P($______ $\le x \le$ ______$)$, where x is a variable from the normal distribution described in part (a).

→ Since 310 is the left endpoint, we subtract 0.5, and since 340 is the right endpoint, we add 0.5. Consequently,

$P(310 \le r \le 340) \approx P(309.5 \le x \le 340.5)$

(c) Convert the condition $309.5 \le x \le 340.5$ to a condition in standard units.

→ Since $\mu = 325$ and $\sigma \approx 10.67$, the condition $309.5 \le x \le 340.5$ becomes

$$\frac{309.5 - 325}{10.67} \le z \le \frac{340.5 - 325}{10.67}$$

or

$$-1.45 \le z \le 1.45$$

(d) $P(310 \le r \le 340) = P(309.5 \le x \le 340.5)$
$= P(-1.45 \le z \le 1.45)$
$=$ ______

→ $P(-1.45 \le z \le 1.45) = P(z \le 1.45) - P(z \le -1.45)$
$= 0.9265 - 0.0735$
$= 0.8530$

(e) Will the normal distribution make a good approximation to the binomial for this problem? Explain your answer.

→ Since

$np = 500(0.65) = 325$

and

$nq = 500(0.35) = 175$

are both greater than 5, the normal distribution will be a good approximation to the binomial.

Sunspots, Tree Rings, and Statistics

Ancient Chinese astronomers recorded extreme sunspot activity, with a peak around 1200 A.D. Mesa Verde tree rings in the period between 1276 and 1299 were unusually narrow, indicating a drought and/or a severe cold spell in the region at that time. A cooling trend could have narrowed the window of frost-free days below the approximately 80 days needed for cultivation of aboriginal corn and beans. Is this the reason the ancient Anasazi dwellings in Mesa Verde were abandoned? Is there a connection to the extreme sunspot activity? Much research and statistical work continues to be done on this topic.

Reference: *Prehistoric Astronomy in the Southwest*, by J. McKim Malville and C. Putnam, Department of Astronomy, University of Colorado.

SECTION 6.4 PROBLEMS

Note: When we say *between a* and *b*, we mean every value from *a* to *b*, *including a* and *b*. Due to rounding, your answers might vary slightly from answers given in the text.

1. *Statistical Literacy* Binomial probability distributions depend on the number of trials n of a binomial experiment and the probability of success p on each trial. Under what conditions is it appropriate to use a normal approximation to the binomial?

2. *Statistical Literacy* When we use a normal distribution to approximate a binomial distribution, why do we make a continuity correction?

3. *Critical Thinking* You need to compute the probability of 5 or fewer successes for a binomial experiment with 10 trials. The probability of success on a single trial is 0.43. Since this probability of success is not in the table, you decide to use the normal approximation to the binomial. Is this an appropriate strategy? Explain.

4. *Critical Thinking* Consider a binomial experiment with 20 trials and probability 0.45 of success on a single trial.
   (a) Use the binomial distribution to find the probability of exactly 10 successes.
   (b) Use the normal distribution to approximate the probability of exactly 10 successes.
   (c) Compare the results of parts (a) and (b).

In the following problems, check that it is appropriate to use the normal approximation to the binomial. Then use the normal distribution to estimate the requested probabilities.

5. *Health: Lead Contamination* More than a decade ago, high levels of lead in the blood put 88% of children at risk. A concerted effort was made to remove lead from the environment. Now, according to the *Third National Health and Nutrition Examination Survey* (NHANES III) conducted by the Centers for Disease Control, only 9% of children in the United States are at risk of high blood-lead levels.
   (a) In a random sample of 200 children taken more than a decade ago, what is the probability that 50 or more had high blood-lead levels?
   (b) In a random sample of 200 children taken now, what is the probability that 50 or more have high blood-lead levels?

6. *[illegible]* Do you try to *pad* an insurance claim to cover your deductible? About 40% of all U.S. adults will try to pad their insurance claims! (Source: *Are You Normal?*, by Bernice Kanner, St. Martin's Press.) Suppose that you are the director of an insurance adjustment office. Your office has just

received 128 insurance claims to be processed in the next few days. What is the probability that
(a) half or more of the claims have been padded?
(b) fewer than 45 of the claims have been padded?
(c) from 40 to 64 of the claims have been padded?
(d) more than 80 of the claims have *not* been padded?

7. *Longevity: 90th Birthday* It is estimated that 3.5% of the general population will live past their 90th birthday (*Statistical Abstract of the United States*, 112th Edition). In a graduating class of 753 high school seniors, what is the probability that
(a) 15 or more will live beyond their 90th birthday?
(b) 30 or more will live beyond their 90th birthday?
(c) between 25 and 35 will live beyond their 90th birthday?
(d) more than 40 will live beyond their 90th birthday?

8. *Fishing: Billfish* Ocean fishing for billfish is very popular in the Cozumel region of Mexico. In *World Record Game Fishes* (published by the International Game Fish Association), it was stated that in the Cozumel region about 44% of strikes (while trolling) resulted in a catch. Suppose that on a given day a fleet of fishing boats got a total of 24 strikes. What is the probability that the number of fish caught was
(a) 12 or fewer?
(b) 5 or more?
(c) between 5 and 12?

9. *Grocery Stores: New Products* The *Denver Post* stated that 80% of all new products introduced in grocery stores fail (are taken off the market) within 2 years. If a grocery store chain introduces 66 new products, what is the probability that within 2 years
(a) 47 or more fail?
(b) 58 or fewer fail?
(c) 15 or more succeed?
(d) fewer than 10 succeed?

10. *Crime: Murder* What are the chances that a person who is murdered actually knew the murderer? The answer to this question explains why a lot of police detective work begins with relatives and friends of the victim! About 64% of people who are murdered actually knew the person who committed the murder (*Chances: Risk and Odds in Everyday Life*, by James Burke). Suppose that a detective file in New Orleans has 63 current unsolved murders. What is the probability that
(a) at least 35 of the victims knew their murderers?
(b) at most 48 of the victims knew their murderers?
(c) fewer than 30 victims did *not* know their murderers?
(d) more than 20 victims did *not* know their murderers?

11. *Supermarkets: Free Samples* Do you take the free samples offered in supermarkets? About 60% of all customers will take free samples. Furthermore, of those who take the free samples, about 37% will buy what they have sampled. (See reference in Problem 6.) Suppose you set up a counter in a supermarket offering free samples of a new product. The day you were offering free samples, 317 customers passed by your counter.
(a) What is the probability that more than 180 will take your free sample?
(b) What is the probability that fewer than 200 will take your free sample?
(c) What is the probability that a customer will take a free sample *and* buy the product? *Hint:* Use the multiplication rule for *dependent* events. Notice that we are given the conditional probability $P(\text{buy}|\text{sample}) = 0.37$, while $P(\text{sample}) = 0.60$.
(d) What is the probability that between 60 and 80 customers will take the free sample *and* buy the product? *Hint:* Use the probability of success calculated in part (c).

12. *Ice Cream: Flavors* What's your favorite ice cream flavor? For people who bu ice cream, the all-time favorite is still vanilla. About 25% of ice cream sales ar vanilla. Chocolate accounts for only 9% of ice cream sales. (See reference in Problem 6.) Suppose that 175 customers go to a grocery store in Cheyenne Wyoming, today to buy ice cream.
   (a) What is the probability that 50 or more will buy vanilla?
   (b) What is the probability that 12 or more will buy chocolate?
   (c) A customer who buys ice cream is not limited to one container or one flavor. What is the probability that someone who is buying ice cream will buy chocolate or vanilla? *Hint:* Chocolate flavor and vanilla flavor are not mutually exclusive events. Assume that the choice to buy one flavor is independent of the choice to buy another flavor. Then use the multiplication rule for independent events, together with the addition rule for events that are not mutually exclusive, to compute the requested probability. (See Section 4.2.)
   (d) What is the probability that between 50 and 60 customers will buy chocolate or vanilla ice cream? *Hint:* Use the probability of success computed in part (c).

13. *Airline Flights: No-Shows* Based on long experience, an airline found that about 6% of the people making reservations on a flight from Miami to Denver do not show up for the flight. Suppose the airline overbooks this flight by selling 267 ticket reservations for an airplane with only 255 seats.
   (a) What is the probability that a person holding a reservation will show up for the flight?
   (b) Let $n = 267$ represent the number of ticket reservations. Let r represent the number of people with reservations who show up for the flight. Which expression represents the probability that a seat will be available for everyone who shows up holding a reservation?

   $$P(255 \le r); \quad P(r \le 255); \quad P(r \le 267); \quad P(r = 255)$$

   (c) Use the normal approximation to the binomial distribution and part (b) to answer the following question: What is the probability that a seat will be available for every person who shows up holding a reservation?

14. *General: Approximations* We have studied *two* approximations to the binomial, the normal approximation and the Poisson approximation (Section 5.4). Write a brief but complete essay in which you discuss and summarize the *conditions* under which each approximation would be used, the *formulas* involved, and the *assumptions* made for each approximation. Give details and examples in your essay. How could you apply these statistical methods in your own everyday life?

Chapter Review

SUMMARY

In this chapter, we examined properties and applications of the normal probability distribution.

- A normal probability distribution is a distribution of a continuous random variable. Normal distributions are bell-shaped and symmetric around the mean. The high point occurs over the mean, and most of the area occurs within 3 standard deviations of the mean. The mean and median are equal.
- The empirical rule for normal distributions gives areas within 1, 2, and 3 standard deviations of the mean. Approximately

  68% of the data lie within the interval $\mu \pm \sigma$

  95% of the data lie within the interval $\mu \pm 2\sigma$

  99.7% of the data lie within the interval $\mu \pm 3\sigma$

- For symmetric, bell-shaped distributions,

$$\text{standard deviation} \approx \frac{\text{range of data}}{4}$$

- A z-score measures the number of standard deviations a raw score x lies from the mean.

$$z = \frac{x - \mu}{\sigma} \quad \text{and} \quad x = z\sigma + \mu$$

- For the standard normal distribution, $\mu = 0$ and $\sigma = 1$.
- Table 5 of Appendix II gives areas under a standard normal distribution that are to the left of a specified value of z.
- After raw scores x have been converted to z scores, the standard normal distribution table can be used to find probabilities associated with intervals of x values from any normal distribution.
- The inverse normal distribution is used to find z values associated with areas to the left of z. Table 5 of Appendix II can be used to find approximate z values associated with specific probabilities.
- Tools for assessing the normality of a data distribution include:

  Histogram of the data. A roughly bell-shaped histogram indicates normality.

  Presence of outliers. A limited number indicates normality.

  Skewness. For normality, Pearson's index is between -1 and 1.

  Normal quantile plot. For normality, points lie close to a straight line.
- Control charts are an important application of normal distributions.
- The binomial distribution can be approximated by a normal distribution with $\mu = np$ and $\sigma = \sqrt{npq}$ provided

  $np > 5$ and $nq > 5$, with $q = 1 - p$

  and a continuity correction is made.

Data from many applications follow distributions that are approximately normal. We will see normal distributions used extensively in later chapters.

IMPORTANT WORDS & SYMBOLS

Section 6.1
Normal distributions
Normal curves
Upward cup and downward cup on normal curves
Symmetry of normal curves
Empirical rule
Control chart
Out-of-control signals

Section 6.2
z value or z score
Standard units
Standard normal distribution ($\mu = 0$ and $\sigma = 1$)
Raw score, x
Area under the standard normal curve

Section 6.3
Areas under any normal curve
Normality indicators

Section 6.4
Normal approximation to the binomial distribution
Continuity correction

Nenana Ice Classic

The Nenana Ice Classic is a betting pool offering a large cash prize to the lucky winner who can guess the time, to the nearest minute, of the ice breakup on the Tanana River in the town of Nenana, Alaska. Official breakup time is defined as the time when the surging river dislodges a tripod on the ice. This breaks an attached line and stops a clock set to Yukon Standard Time. The event is so popular that the first state legislature of Alaska (1959) made the Nenana Ice Classic an official statewide lottery. Since 1918, the earliest breakup was April 20, 1940, at 3:27 P.M., and the latest recorded breakup was May 20, 1964, at 11:41 A.M. Want to make a statistical guess predicting when the ice will break up? Breakup times from 1918 to 1996 are recorded in The Alaska Almanac, *published by Alaska Northwest Books, Anchorage.*

CHAPTER REVIEW PROBLEMS

1. *Statistical Literacy* Describe a normal probability distribution.

2. *Statistical Literacy* According to the empirical rule, approximately what percentage of the area under a normal distribution lies within 1 standard deviation of the mean? within 2 standard deviations? within 3 standard deviations?

3. *Statistical Literacy* Is a process in control if the corresponding control chart for data having a normal distribution shows a value beyond 3 standard deviations from the mean?

4. *Statistical Literacy* Can a normal distribution always be used to approximate a binomial distribution? Explain.

5. *Statistical Literacy* What characteristic of a normal quantile plot indicates that the data follow a distribution that is approximately normal?

6. *Statistical Literacy* For a normal distribution, is it likely that a data value selected at random is more than 2 standard deviations above the mean?

7. Given that z is the standard normal variable (with mean 0 and standard deviation 1), find
   (a) $P(0 \le z \le 1.75)$ (b) $P(-1.29 \le z \le 0)$
   (c) $P(1.03 \le z \le 1.21)$ (d) $P(z \ge 2.31)$
   (e) $P(z \le -1.96)$ (f) $P(z \le 1.00)$

8. Given that z is the standard normal variable (with mean 0 and standard deviation 1), find
   (a) $P(0 \le z \le 0.75)$ (b) $P(-1.50 \le z \le 0)$
   (c) $P(-2.67 \le z \le -1.74)$ (d) $P(z \ge 1.56)$
   (e) $P(z \le -0.97)$ (f) $P(z \le 2.01)$

9. Given that x is a normal variable with mean $\mu = 47$ and standard deviation $\sigma = 6.2$, find
   (a) $P(x \le 60)$ (b) $P(x \ge 50)$ (c) $P(50 \le x \le 60)$

10. Given that x is a normal variable with mean $\mu = 110$ and standard deviation $\sigma = 12$, find
    (a) $P(x \le 120)$ (b) $P(x \ge 80)$ (c) $P(108 \le x \le 117)$

11. Find z such that 5% of the area under the standard normal curve lies to the right of z.

12. Find z such that 1% of the area under the standard normal curve lies to the left of z.

13. Find z such that 95% of the area under the standard normal curve lies between $-z$ and z.

14. Find z such that 99% of the area under the standard normal curve lies between $-z$ and z.

15. *Nursing: Exams* On a practical nursing licensing exam, the mean score is 79 and the standard deviation is 9 points.
    (a) What is the standardized score of a student with a raw score of 87?
    (b) What is the standardized score of a student with a raw score of 79?
    (c) Assuming the scores follow a normal distribution, what is the probability that a score selected at random is above 85?

16. *Aptitude Tests: Mechanics* On an auto mechanic aptitude test, the mean score is 270 points and the standard deviation is 35 points.
    (a) If a student has a standardized score of 1.9, how many points is that?
    (b) If a student has a standardized score of -0.25, how many points is that?
    (c) Assuming the scores follow a normal distribution, what is the probability that a student will get between 200 and 340 points?

17. *Recycling: Aluminum Cans* One environmental group did a study of recycling habits in a California community. It found that 70% of the aluminum cans sold in the area were recycled.
    (a) If 400 cans are sold today, what is the probability that 300 or more will be recycled?
    (b) Of the 400 cans sold, what is the probability that between 260 and 300 will be recycled?

18. *Guarantee: Disc Players* Future Electronics makes compact disc players. Its research department found that the life of the laser beam device is normally distributed, with mean 5000 hours and standard deviation 450 hours.
    (a) Find the probability that the laser beam device will wear out in 5000 hours or less.
    (b) *Inverse Normal Distribution* Future Electronics wants to place a guarantee on the players so that no more than 5% fail during the guarantee period. Because the laser pickup is the part most likely to wear out first, the guarantee period will be based on the life of the laser beam device. How many playing hours should the guarantee cover? (Round to the next playing hour.)

19. *Guarantee: Package Delivery* Express Courier Service has found that the delivery time for packages is normally distributed, with mean 14 hours and standard deviation 2 hours.
    (a) For a package selected at random, what is the probability that it will be delivered in 18 hours or less?
    (b) *Inverse Normal Distribution* What should be the guaranteed delivery time on all packages in order to be 95% sure that the package will be delivered before this time? (*Hint:* Note that 5% of the packages will be delivered at a time beyond the guaranteed time period.)

20. *Control Chart: Landing Gear* Hydraulic pressure in the main cylinder of the landing gear of a commercial jet is very important for a safe landing. If the pressure is not high enough, the landing gear may not lower properly. If it is too high, the connectors in the hydraulic line may spring a leak.

    In-flight landing tests show that the actual pressure in the main cylinders is a variable with mean 819 pounds per square inch and standard deviation 23 pounds per square inch. Assume that these values for the mean and standard deviation are considered safe values by engineers.

    (a) For nine consecutive test landings, the pressure in the main cylinder is recorded as follows:

Landing number	1	2	3	4	5	6	7	8	9
Pressure	870	855	830	815	847	836	825	810	792

    Make a control chart for the pressure in the main cylinder of the hydraulic landing gear, and plot the data on the control chart. Looking at the control chart, would you say the pressure is "in control" or "out of control"? Explain your answer. Identify any out-of-control signals by type (I, II, or III).

    (b) For 10 consecutive test landings, the pressure was recorded on another plane as follows:

Landing number	1	2	3	4	5	6	7	8	9	10
Pressure	865	850	841	820	815	789	801	765	730	725

    Make a control chart and plot the data on the chart. Would you say the pressure is "in control" or not? Explain your answer. Identify any out-of-control signals by type (I, II, or III).

21. *Electronic Scanners: Errors* Instead of hearing the jingle of prices being rung up manually on cash registers, we now hear the beep of prices being scanned by electronic scanners. How accurate is price scanning? There are errors, and according to a *Denver Post* article, when the error occurs in the store's favor, it is larger than when it occurs in the customer's favor. An investigation of large discount stores by the Colorado state inspectors showed that the average error in the store's favor was \$2.66. Assume that the distribution of scanner errors is more or less mound-shaped. If the standard deviation of scanner errors (in the store's favor) is \$0.85, use the empirical rule to
   (a) estimate a range of scanner errors centered about the mean in which 68% of the errors will lie.
   (b) estimate a range of scanner errors centered about the mean in which 95% of the errors will lie.
   (c) estimate a range of scanner errors centered about the mean in which almost all the errors will lie.

22. *Customer Complaints: Time* The Customer Service Center in a large New York department store has determined that the amount of time spent with a customer about a complaint is normally distributed, with a mean of 9.3 minutes and a standard deviation of 2.5 minutes. What is the probability that for a randomly chosen customer with a complaint, the amount of time spent resolving the complaint will be
   (a) less than 10 minutes?
   (b) longer than 5 minutes?
   (c) between 8 and 15 minutes?

23. *Medical: Flight for Life* The Flight for Life emergency helicopter service is available for medical emergencies occurring from 15 to 90 miles from the hospital. Emergencies that occur closer to the hospital can be handled effectively by ambulance service. A long-term study of the service shows that the response time from receipt of the dispatch call to arrival at the scene of the emergency is normally distributed, with a mean of 42 minutes and a standard deviation of 8 minutes. For a randomly received call, what is the probability that the response time will be
   (a) between 30 and 45 minutes?
   (b) less than 30 minutes?
   (c) longer than 60 minutes?

24. *Unlisted Phones: Sacramento* How easy is it to contact a person in Sacramento, California? If you don't know the telephone number, it could be difficult. The data from Survey Sampling of Fairfield, Connecticut, reported in *American Demographics*, show that 68% of the telephone-owning households in Sacramento have unlisted numbers. For a random sample of 150 Sacramento households with telephones, what is the probability that
   (a) 100 or more have unlisted numbers?
   (b) fewer than 100 have unlisted numbers?
   (c) between 50 and 65 (including 50 and 65) have *listed* numbers?

25. *Medical: Blood Type* Blood type AB is found in only 3% of the population (*Textbook of Medical Physiology*, by A. Guyton, M.D.). If 250 people are chosen at random, what is the probability that
   (a) 5 or more will have this blood type?
   (b) between 5 and 10 will have this blood type?

DATA HIGHLIGHTS: GROUP PROJECTS

Break into small groups and discuss the following topics. Organize a brief outline in which you summarize the main points of your group discussion.

1. Examine Figure 6-37. Government documents and the Census Bureau show that the age at first marriage for U.S. citizens is approximately normally distributed for both men and women. For men, the average age is about 27 years, and for

women, the average is about 24 years. For both sexes, the standard deviation is about 2.5 years.

(a) If the distribution is symmetrical and mound-shaped (such as the normal distribution), why would you expect the median, mean, and mode to be equal?

(b) Consider the age at first marriage in 1995. What is the probability that a man selected at random was over age 30 at the time of his first marriage? What is the probability that he was under age 20? What is the probability that he was between 20 and 30?

(c) Consider the age at first marriage in 1995. What is the probability that a woman selected at random was over age 28 at the time of her first marriage? What is the probability that she was under age 18? What is the probability that she was between 18 and 28?

(d) At what age were only 10% of eligible men (who had never been married before) unmarried? At what age were only 5% unmarried?

(e) At what age were only 10% of eligible women (who had never been married before) unmarried? At what age were only 5% unmarried?

(f) The Census Bureau tracks data about marriages and age at first marriage. The *Statistical Abstract of the United States* is published each year and contains tables giving this information. For more information, visit the Brase/Brase statistics site at **college.hmco.com/pic/braseUS9e** and find the link to the Census Bureau. Look for marriage under the subject index, and follow the links to the table. Using either the Abstract or the Census Bureau web site, find the most recent information about median age at first marriage. Although the median age changes from year to year, the standard deviation usually does not change much. Under the assumption that the age at first marriage follows a distribution that is approximately normal, the mean age at first marriage equals the median age. Using a standard deviation of about 2.5 years, repeat parts (b) through (e) for the most recent year for which median age at first marriage is available.

2. Examine Figure 6-38, "Time Shopping."

(a) Notice that 52% of the people who go to a shopping center spend less than 1 hour. However, the figure also states that people spend an average of 69 minutes on each visit to a shopping center. How could *both* these claims be correct? Write a brief, complete essay in which you discuss mean, median, and mode in the context of a symmetrical distribution. Also discuss mean, median, and mode for distributions that are skewed left, for those skewed right, and for general distributions. Then answer the following question: How could 52% of the people spend less than 1 hour in the shopping center if the average (we don't know which average was used) time spent was 69 minutes?

FIGURE 6-37

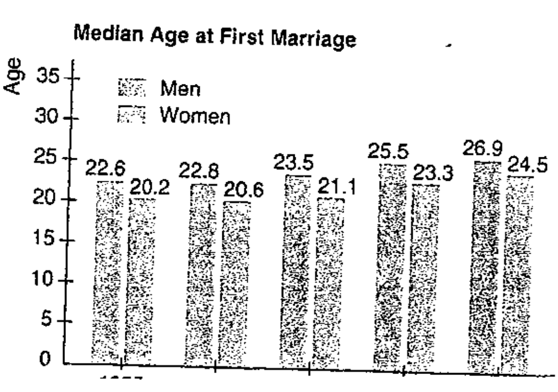
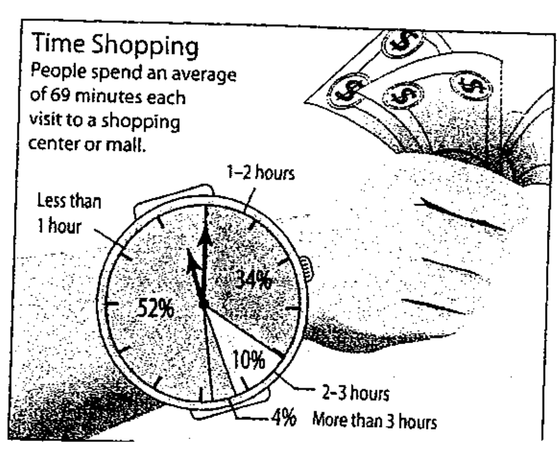


FIGURE 6-38



(b) Ala Moana Shopping Center in Honolulu is sometimes advertised as the largest shopping center for 2500 miles and the best shopping center in the middle of the Pacific Ocean. There are many interesting Hawaiian, Asian, and other ethnic shops in Ala Moana, so this center is a favorite of tourists. Suppose a tour group of 75 people has just arrived at the shopping center. The International Council of Shopping Centers indicates that 86% (52% plus 34%) of the people will spend less than 2 hours in a shopping center. Let us assume this statement applies to our tour group.

(i) For the tour group of 75 people, what is the expected number who finish shopping on or before 2 hours? What is the standard deviation?

(ii) For the tour group of 75 people, what is the probability that 55 or more will finish shopping in 2 hours or less?

(iii) For the tour group, what is the probability that 70 or more will finish shopping in 2 hours or less?

(iv) What is the probability that between 50 and 70 (including 50 and 70) people in the tour group will finish shopping in 2 hours or less?

(v) If you were a tour director, how could you use this information to plan an appropriate length of time for the stop at Ala Moana shopping center? You do not want to frustrate your group by allowing too little time for shopping, but you also do not want to spend too much time at the center.

LINKING CONCEPTS: WRITING PROJECTS

Discuss each of the following topics in class or review the topics on your own. Then write a brief but complete essay in which you summarize the main points. Please include formulas and graphs as appropriate.

1. If you look up the word *normal* in a dictionary, you will find that it is synonymous with the words *standard* or *usual*. Consider the very wide and general applications of the normal probability distribution. Comment on why good synonyms for *normal probability distribution* might be *standard probability distribution* or *usual probability distribution*. List at least three random variables from everyday life to which you think the normal probability distribution could be applicable.

2. Why are standard z values so important? Is it true that z values have no units of measurement? Why would this be desirable for comparing data sets with *different* units of measurement? How can we assess differences in quality or performance by simply comparing z values under a standard normal curve? Examine the formula for computing standard z values. Notice that it involves *both* the mean and standard deviation. Recall that in Chapter 3 we commented that the mean of a data collection is not entirely adequate to describe the data; you need the standard deviation as well. Discuss this topic again in light of what you now know about normal distributions and standard z values.

3. If you look up the word *empirical* in a dictionary, you will find that it means "relying on experiment and observation rather than on theory." Discuss the empirical rule in this context. The empirical rule certainly applies to the normal distribution, but does it also apply to a wide variety of other distributions that are not *exactly* (theoretically) normal? Discuss the terms *mound-shaped* and *symmetrical*. Draw several sketches of distributions that are mound-shaped *and* symmetrical. Draw sketches of distributions that are not mound-shaped or symmetrical. To which distributions will the empirical rule apply?

4. Most companies that manufacture a product have a division responsible for quality control or quality assurance. The purpose of the quality-control division is to make reasonably certain that the products manufactured are up to company standards. Write a brief essay in which you describe how the statistics you have learned so far could be applied to an industrial application (such as control charts and the Antlers Lodge example).

USING TECHNOLOGY

Application 1

How much money do people earn in a month? One way to answer this question is to look at government employees. The average earnings of a city government employee for the month of March are given for a sample of 40 large cities in the United States (*Statistical Abstract of the United States*, 120th Edition).

(a) To compare the earnings from one city to another, we will look at z values for each city. Use the sample mean and sample standard deviation to compute the z values. You may do this "by hand" using a calculator, or you may use a computer software package.

(b) Look at the z values for each salary. Which are above average? Which are below average?

(c) Which salaries are within 1 standard deviation of the mean?

Technology Hints: Standardizing Raw Scores

TI-84Plus/TI-83Plus

On the TI-84Plus and TI-83Plus calculators, enter the salary data in list L_1. Then use ➤ STAT ➤ CALC ➤ **1-Var Stats** to compute the sample mean $\bar{x}$ and the standard deviation s. Then go back to ➤ STAT ➤ EDIT. Arrow up to the header label L_2. At the L_2 = prompt, type in $(x - \bar{x})/s$ and press Enter. You will find the $\bar{x}$ and s symbols under the ➤ VARS ➤ Statistics ... keys. The z values will then appear in list L_2.

City	Average Earnings ($) for March	City	Average Earnings ($) for March
Albuquerque	2494	Memphis	2973
Anchorage	3571	Miami	3654
Arlington	3069	Milwaukee	3636
Atlanta	2494	Mobile	3756
Baltimore	3229	New York	3694
Birmingham	3961	Newark	2507
Boston	3526	Norfolk	2577
Cincinnati	3399	Oakland	5084
Cleveland	2899	Philadelphia	3511
Dallas	3326	Phoenix	3909
Denver	3324	Portland	2899
Detroit	3301	Raleigh	3026
Honolulu	3479	San Antonio	2427
Houston	2813	San Diego	4072
Indianapolis	2756	San Francisco	4487
Kansas City	3023	San Jose	5227
Lincoln	2882	Seattle	4462
Las Vegas	4512	St. Louis	3033
Los Angeles	4534	Washington, D.C.	3725
Louisville	2674	Wichita	2928

Excel

In Excel, you need to compute the sample mean $\bar{x}$ and standard deviation s ahead of time. To do so, enter the salary data in column A. To compute $\bar{x}$, use the menu choices ➤ **Paste function** (fx) ➤ **Statistical** ➤ **AVERAGE.** To compute s, use the menu choices ➤ **Paste function** (fx) ➤ **Statistical** ➤ **STDEV.** Finally, to generate the z values, use the menu choices ➤ **Paste function** (fx) ➤ **Statistical** ➤ **STANDARDIZED.** Remember to highlight the cell in which you want each value to appear before you use the commands.

Minitab

In Minitab, enter the salaries in column C1. To generate z values and store the results in column C2, use the following menu choices: ➤ **Calc** ➤ **Standardize.** In the dialogue box, select C1 for the input column and C2 for the output column. Then choose the option "subtract mean and divide by standard deviation." Minitab will compute the sample mean $\bar{x}$ and standard deviation s for the salaries and then compute z using the formula $z = (x - \bar{x})/s$.

SPSS

Enter your data in one column. Use the menu choices **Analyze** ➤ **Descriptive Statistics** ➤ **Descriptives.** In the dialogue box, move the variable containing data into the variables box. Check the option "Save standardized values as variables." Standardized values will appear in a column of the data editor.

Application 2

How can we determine if data originated from a normal distribution? We can look at a stem-and-leaf plot or histogram of the data to check for general symmetry, skewness, clusters of data, or outliers. However, a more sensitive way to check that a distribution is normal is to look at a special graph called a *normal quantile plot* (or a variation of this plot called a *normal probability plot* in some software packages). It really is not feasible to make a normal quantile plot by hand, but statistical software packages provide such plots. A simple version of the basic idea behind normal quantile plots involves the following process:

(a) Arrange the observed data values in order from smallest to largest, and determine the percentile occupied by each value. For instance, if there are 20 data values, the smallest datum is at the 5% point, the next smallest is at the 10% point, and so on.

(b) Find the z values that correspond to the percentile points. For instance, the z value that corresponds to the percentile 5% (i.e., percent in the left tail of the distribution) is $z = -1.645$.

(c) Plot each data value x against the corresponding percentile z score. If the data are close to a normal distribution, the plotted points will lie close to a straight line. (If the data are close to a standard normal distribution, the points will lie close to the line $x = z$.)

The actual process that statistical software packages use to produce the z scores for the data is more complicated.

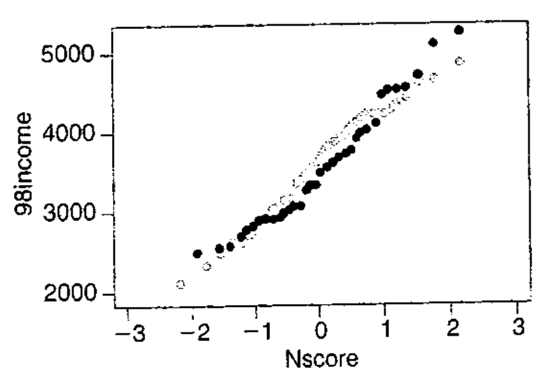
Interpreting normal quantile plots

If the points of a normal quantile plot lie close to a straight line, the plot indicates that the data follow a normal distribution. Systematic deviations from a straight line or bulges in the plot indicate that the data distribution is not normal. Individual points off the line may be outliers.

Consider Figure 6-39. This figure shows Minitab-generated quantile plots for two data sets. The black dots show the normal quantile plot for the salary data of the first application. The red dots show the normal quantile plot for a random sample of 42 data values drawn from a theoretical normal distribution with the same mean and standard deviation as the salary data ($\mu \approx 3421$, $\sigma \approx 709$).

FIGURE 6-39 Normal Quantile Plots

- Salary data for (city) government employees
- A random sample of 42 values from a theoretical normal distribution with the same mean and standard deviation as the salary data



(a) Do the black dots lie close to a straight line? Do the salaries appear to follow a normal distribution? Are there any outliers on the low or high side? Would you say that any of the salaries are "out of line" for a normal distribution?

(b) Do the red dots lie close to a straight line? We know the red dots represent a sample drawn from a normal distribution. Is the normal quantile plot for the red dots consistent with this fact? Are there any outliers shown?

Technology Hints

TI-84Plus/TI-83Plus

Enter the data. Press STATPLOT and select one of the plots. Highlight ON. Then highlight the sixth plot option. To get a plot similar to that of Figure 6-39, choose Y as the data axis.

Minitab

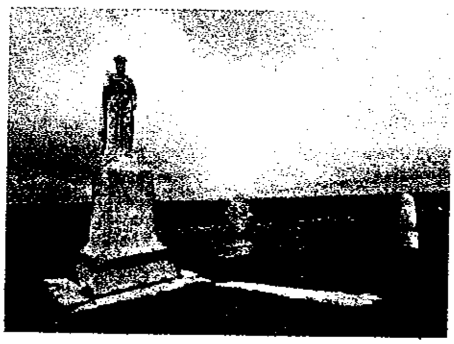
Minitab has several types of normal quantile plots that use different types of scales. To create a normal quantile plot similar to that of Figure 6-39, enter the data in column C1. Then use the menu choices **Calc ➤ Calculator**. In the dialogue box listing the functions, scroll to **Normal Scores**. Use **NSCOR(C1)** and store the results in column C2. Finally, use the menu choices **Graph ➤ Plot**. In the dialogue box, use C1 for variable *y* and C2 for variable *x*.

SPSS

Enter the data. Use the menu choices **Analyze ➤ Descriptive Statistics ➤ Explore**. In the dialogue box, move your data variable to the dependent list. Check **Plots. . . .** Check "Normality plots with tests." The graph appears in the output window.

CHAPTERS 4–6

The Hill of Tara is located in south central Meath, not far from Dublin, Ireland. Tara is of great cultural and archaeological importance, since it is by legend the seat of the ancient high kings of Ireland. For more information, see *Tara: An Archaeological Survey*, by Conor Newman, Royal Irish Academy, Dublin.



Magnetic surveying is one technique used by archaeologists to determine anomalies arising from variations in magnetic susceptibility. Unusual changes in magnetic susceptibility might (or might not) indicate an important archaeological discovery. Let x be a random variable that represents a magnetic susceptibility (MS) reading for a randomly chosen site on the Hill of Tara. A random sample of 120 sites gave the readings shown in Table A below.

TABLE A **Magnetic Susceptibility Readings, centimeter-gram-second $\times 10^{-6}$ (cmg $\times 10^{-6}$)**

Comment	Magnetic Susceptibility	Number of Readings	Estimated Probability
"cool"	$0 \le x < 10$	30	$30/120 = 0.25$
"neutral"	$10 \le x < 20$	54	$54/120 = 0.45$
"warm"	$20 \le x < 30$	18	$18/120 = 0.15$
"very interesting"	$30 \le x < 40$	12	$12/120 = 0.10$
"hot spot"	$40 \le x$	6	$6/120 = 0.05$

1. *Statistical Literacy: Sample Space* What is a statis[tical experiment?] cal experiment? How could the magnetic susceptibil[ity] intervals $0 \le x < 10$, $10 \le x < 20$, and so on be considered events in the sample space of all possible readings?

2. *Statistical Literacy: Probability* What is probability? What do we mean by relative frequency as a probability estimate for events? What is the law of large numbers? How would the law of large numbers apply in this context?

3. *Statistical Literacy: Probability Distribution* Do the probabilities shown in Table A add up to 1? Why should they total to 1?

4. *Probability Rules* For a site chosen at random, estimate the following probabilities.
   (a) $P(0 \le x < 30)$ (b) $P(10 \le x < 40)$
   (c) $P(x < 20)$ (d) $P(x \ge 20)$
   (e) $P(30 \le x)$ (f) $P(x$ *not* less than $10)$
   (g) $P(0 \le x < 10$ *or* $40 \le x)$
   (h) $P(40 \le x$ *and* $20 \le x)$

5. *Conditional Probability* Suppose you are working in a "warm" region in which all MS readings are 20 or higher. In this same region, what is the probability that you will find a "hot spot" in which the readings are 40 or higher? Use conditional probability to estimate $P(40 \le x \mid 20 \le x)$. *Hint*: See Problem 39 of Section 6.3.

6. *Discrete Probability Distribution* Consider the midpoint of each interval. Assign the value 45 as the midpoint for the interval $40 \le x$. The midpoints constitute the sample space for a discrete random variable. Using Table A, compute the expected value μ and the standard deviation σ.

Midpoint x	5	15	25	35	45
$P(x)$					

7. *Binomial Distribution* Suppose a reading between 30 and 40 is called "very interesting" from an archaeological point of view. Let us say you take readings at

$n = 12$ sites chosen at random. Let r be a binomial random variable that represents the number of "very interesting" readings from these 12 sites.

(a) Let us call "very interesting" a binomial success. Use Table A to find p, the probability of success on a single trial, where $p = P(\text{success}) = P(30 \le x < 40)$.

(b) What is the expected value μ and standard deviation σ for the random variable r?

(c) What is the probability that you will find *at least* one "very interesting" reading in the 12 sites?

(d) What is the probability that you will find *fewer than* three "very interesting" readings in the 12 sites?

8. Suppose a "hot spot" is a site with a reading of 40 or higher.

(a) In a binomial setting, let us call success a "hot spot." Use Table A to find $p = P(\text{success}) = P(40 \le x)$ for a single trial.

(b) Suppose you decide to take readings at random until you get your *first* "hot spot." Let n be a random variable representing the trial on which you get your first "hot spot." Use the geometric probability distribution to write out a formula for $P(n)$.

(c) What is the probability that you will need more than four readings to find the first "hot spot"? Compute $P(n > 4)$.

9. Suppose an archaeologist is looking for geomagnetic "hot spots" in an unexplored region of Tara. As in Problem 8, we have a binomial setting where success is a "hot spot." In this case, the probability of success is $p = P(40 \le x)$. The archaeologist takes $n = 100$ magnetic susceptibility readings in the new, unexplored area. Let r be a binomial random variable representing the number of "hot spots" in the 100 readings.

(a) We want to approximate the binomial random variable r by a Poisson distribution. Is this appropriate? What requirements must be satisfied before we can do this? Do you think these requirements are satisfied in this case? Explain. What is the value of λ?

(b) What is the probability that the archaeologists will find six or fewer "hot spots?" *Hint*: Use Table 4 of Appendix II.

(c) What is the probability that the archaeologists will find more than eight "hot spots"?

10. Consider a binomial setting in which "neutral" is defined to be a success. So $p = P(\text{success}) = P(10 \le x < 20)$. Suppose $n = 65$ geomagnetic readings are taken. Let r be a binomial random variable that represents the number of "neutral" geomagnetic readings.

(a) We want to approximate the binomial random variable r by a normal variable x. Is this appropriate? What requirements must be satisfied before we can do this? Do you think these requirements are satisfied in this case? Explain.

(b) What is the probability that there will be at least 20 "neutral" readings out of these 65 trials?

(c) Why would the Poisson approximation to the binomial *not* be appropriate in this case? Explain.

11. *Oxygen demand* is a term biologists use to describe the oxygen needed by fish and other aquatic organisms for survival. The Environmental Protection Agency conducted a study of a wetland area in Marin County, California. In this wetland environment, the mean oxygen demand was $\mu = 9.9$ mg/L with 95% of the data ranging from 6.5 mg/L to 13.3 mg/L (Reference: EPA Report 832-R-93-005). Let x be a random variable that represents oxygen demand in this wetland environment. Assume x has a probability distribution that is approximately normal.

(a) Use the 95% data range to estimate the standard deviation for oxygen demand. *Hint:* See Problem 31 of Section 6.3.

(b) An oxygen demand below 8 indicates that some organisms in the wetland environment may be dying. What is the probability that the oxygen demand will fall below 8 mg/L?

(c) A high oxygen demand can also indicate trouble. An oxygen demand above 12 may indicate an overabundance of organisms that endanger some types of plant life. What is the probability that the oxygen demand will exceed 12 mg/L?

12. Write a brief but complete essay in which you describe the probability distributions you have studied so far. Which apply to discrete random variables? Which apply to continuous random variables? Under what conditions can the binomial distribution be approximated by the normal? by the Poisson?