CHAPTER 6

Normal Random Variables

Among other peculiarities of the 19th century is this one, that by initiating the systematic collection of statistics it has made the quantitative study of social forces possible.

Alfred North Whitehead

CONTENTS

Introductory Statistics, DOI: 10.1016/B978-0-12-374388-6.00006-5

We introduce continuous random variables, which are random variables that can take on any value in an interval. We show how their probabilities are determined from an associated curve known as a *probability density function*. A special type of continuous random variable, known as a *normal random variable*, is studied. The standard normal random variable is introduced, and a table is presented that enables us to compute the probabilities of that variable. We show how any normal random variable can be transformed to a standard one, enabling us to determine its probabilities. We present the additive property of normal random variables. The percentiles of normal random variables are studied.

6.1 INTRODUCTION

In this chapter we introduce and study the normal distribution. Both from a theoretical and from an applied point of view, this distribution is unquestionably the most important in all statistics.

The normal distribution is one of a class of distributions that are called *continuous*. Continuous distributions are introduced in Sec. 6.2. In Sec. 6.3 we define what is meant by a normal distribution and present an approximation rule concerning its probabilities. In Sec. 6.4, we consider the standard normal distribution, which is a normal distribution having mean 0 and variance 1, and we show how to determine its probabilities by use of a table. In Sec. 6.5 we show how any normal random variable can be transformed to a standard normal, and we use this transformation to determine the probabilities of that variable. The additive property of normal random variables is discussed in Sec. 6.6, and in Sec. 6.7 we consider their percentiles.

The normal distribution was introduced by the French mathematician Abraham De Moivre in 1733.

6.2 CONTINUOUS RANDOM VARIABLES

Whereas the possible values of a discrete random variable can be written as a sequence of isolated values, a *continuous random variable* is one whose set of possible values is an interval. That is, a continuous random variable is able to take on any value within some interval. For example, such variables as the time it takes to complete a scientific experiment and the weight of an individual are usually considered to be continuous random variables.

Every continuous random variable X has a curve associated with it. This curve, formally known as a *probability density function*, can be used to obtain probabilities associated with the random variable. This is accomplished as follows. Consider any two points a and b, where a is less than b. The probability that X assumes a

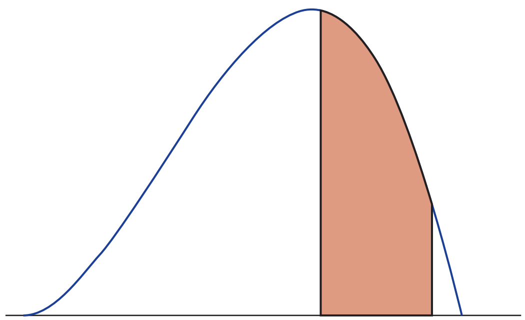


FIGURE 6.1
Probability density function of X.

value that lies between a and b is equal to the area under the curve between a and b. That is,

$$P\{a \leq X \leq b\} = \text{area under curve between } a \text{ and } b$$

Figure 6.1 presents a probability density function.

Since X must assume some value, it follows that the total area under the density curve must equal 1. Also, since the area under the graph of the probability density function between points a and b is the same regardless of whether the endpoints a and b are themselves included, we see that

$$P\{a \leq X \leq b\} = P\{a < X < b\}$$

Historical Perspective

Abraham De Moivre (1667–1754)

(North Wind Picture Archives)

Abraham De Moivre

Today there is no shortage of statistical consultants, many of whom ply their trade in the most elegant of settings. However, the first of their breed worked, in the early years of the 18th century, out of a dark, grubby betting shop in Long Acres, London, known as Slaughter's Coffee House. He was Abraham De Moivre, a Protestant refugee from Catholic France, and for a price he would compute the probability of gambling bets in all types of games of chance.

Although De Moivre, the discoverer of the normal curve, made his living at the coffee shop, he was a mathematician of recognized abilities. Indeed, he was a member of the Royal Society and was reported to be an intimate of Isaac Newton.

This is Karl Pearson imagining De Moivre at work at Slaughter's Coffee House:

> I picture De Moivre working at a dirty table in the coffee house with a broken-down gambler beside him and Isaac Newton walking through the crowd to his corner to fetch out his friend. It would make a great picture for an inspired artist.

That is, the probability that a continuous random variable lies in some interval is the same whether you include the endpoints of the interval or not.

The probability density curve of a random variable X is a curve that never goes below the x axis and has the property that the total area between it and the x axis is equal to 1. It determines the probabilities of X in that the area under the curve between points a and b is equal to the probability that X is between a and b.

PROBLEMS

1. Figure 6.2 is a probability density function for the random variable that represents the time (in minutes) it takes a repairer to service a television. The numbers in the regions represent the areas of those regions. What is the probability that the repairer takes
   (a) Less than 20
   (b) Less than 40
   (c) More than 50
   (d) Between 40 and 70 minutes to complete a repair?
2. A random variable is said to be a *uniform* random variable in the interval (a, b) if its set of possible values is this interval and if its density curve is a horizontal line. That is, its density curve is as given in Fig. 6.3.

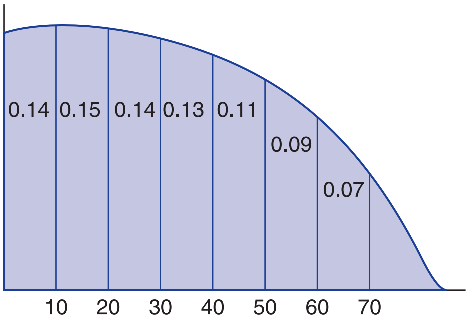


FIGURE 6.2
Probability density function of X.

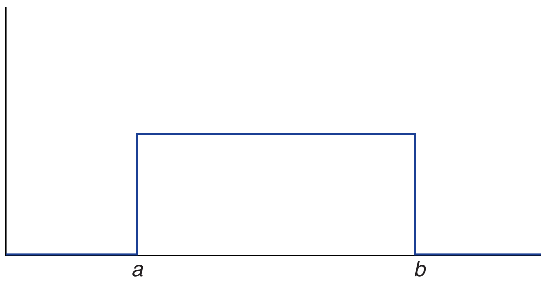
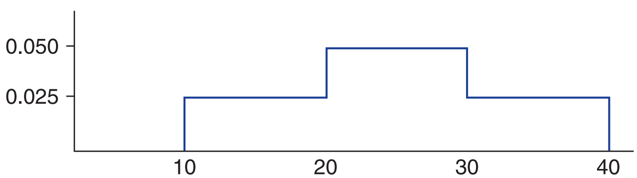


FIGURE 6.3
Density curve of the uniform (a, b) random variable.

(a) Explain why the height of the density curve is $1/(b-a)$. (*Hint*: Remember that the total area under the density curve must equal 1, and recall the formula for the area of a rectangle.)
(b) What is $P\{X \leq (a+b)/2\}$?

3. Suppose that X is a uniform random variable over the interval $(0, 1)$. That is, $a = 0$ and $b = 1$ for the random variable in Prob. 2. Find
   (a) $P\{X > 1/3\}$
   (b) $P\{X \leq 0.7\}$
   (c) $P\{0.3 < X \leq 0.9\}$
   (d) $P\{0.2 \leq X < 0.8\}$
4. You are to meet a friend at 2 p.m. However, while you are always exactly on time, your friend is always late and indeed will arrive at the meeting place at a time uniformly distributed between 2 and 3 p.m. Find the probability that you will have to wait
   (a) At least 30 minutes
   (b) Less than 15 minutes
   (c) Between 10 and 35 minutes
   (d) Less than 45 minutes
5. Suppose in Prob. 4 that your friend will arrive at a time that is uniformly distributed between 1:30 and 3 p.m. Find the probability that
   (a) You are the first to arrive.
   (b) Your friend will have to wait more than 15 minutes.
   (c) You will have to wait over 30 minutes.
6. Suppose that the number of minutes of playing time of a certain college basketball player in a randomly chosen game has the following density curve.



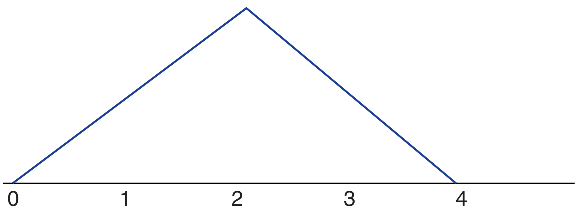
Find the probability that the player plays

(a) Over 20 minutes
(b) Less than 25 minutes
(c) Between 15 and 35 minutes
(d) More than 35 minutes

7. Let X denote the number of minutes played by the basketballer of Prob. 6. Find

(a) $P\{20 < X < 30\}$
(b) $P\{X > 50\}$
(c) $P\{20 < X < 40\}$
(d) $P\{15 < X < 25\}$

8. It is now 2 p.m., and Joan is planning on studying for her statistics test until 6 p.m., when she will have to go out to dinner. However, she knows that she will probably have interruptions and thinks that the amount of time she will actually spend studying in the next 4 hours is a random variable whose probability density curve is as follows:



(a) What is the height of the curve at the value 2? (*Hint*: You will have to recall the formula for the area of a triangle.)
(b) What is the probability she will study more than 3 hours?
(c) What is the probability she will study between 1 and 3 hours?

6.3 NORMAL RANDOM VARIABLES

The most important type of random variable is the normal random variable. The probability density function of a normal random variable X is determined by two parameters: the expected value and the standard deviation of X. We designate these values as μ and σ, respectively. That is, we will let

$$\mu = E[X] \quad \text{and} \quad \sigma = \mathrm{SD}(X)$$

The normal probability density function is a bell-shaped density curve that is symmetric about the value μ. Its variability is measured by σ. The larger σ is, the

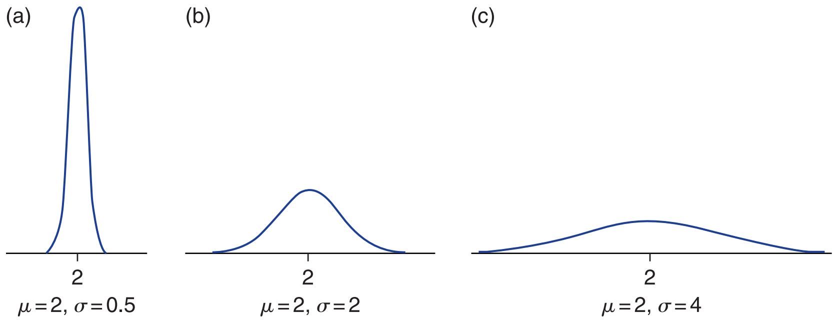


FIGURE 6.4
Three normal probability density functions.

more variability there is in this curve. Figure 6.4 presents three different normal probability density functions. Note how the curves flatten out as σ increases.

Because the probability density function of a normal random variable X is symmetric about its expected value μ, it follows that X is equally likely to be on either side of μ. That is,

$$P\{X < \mu\} = P\{X > \mu\} = \frac{1}{2}$$

Not all bell-shaped symmetric density curves are normal. The normal density curves are specified by a particular formula: The height of the curve above point x on the abscissa is

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

Although we will not make direct use of this formula, it is interesting to note that it involves two of the famous constants of mathematics: π (the area of a circle of radius 1) and e (which is the base of the natural logarithms). Also note that this formula is completely specified by the mean value μ and the standard deviation σ.

A normal random variable having mean value 0 and standard deviation 1 is called a *standard normal* random variable, and its density curve is called the *standard normal curve*. Figure 6.5 presents the standard normal curve. In this text we will use (and reserve) the letter Z to represent a standard normal random variable.

In Sec. 6.5 we will show how to determine probabilities concerning an arbitrary normal random variable by relating them to probabilities about the standard normal random variable. In doing so, we will show the following useful approximation rule for normal probabilities.

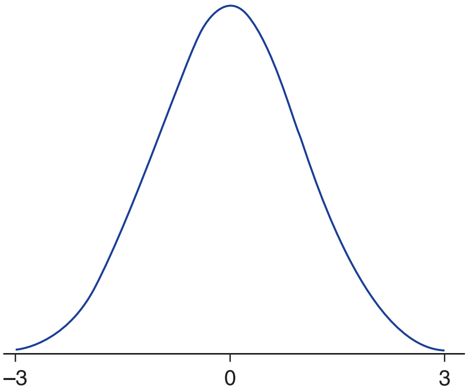


FIGURE 6.5
The standard normal curve.

Approximation Rule

A normal random variable with mean μ and standard deviation σ will be

Between $\mu - \sigma$ and $\mu + \sigma$ with approximate probability 0.68

Between $\mu - 2\sigma$ and $\mu + 2\sigma$ with approximate probability 0.95

Between $\mu - 3\sigma$ and $\mu + 3\sigma$ with approximate probability 0.997

This approximation rule is illustrated in Fig. 6.6. It often enables us to obtain a quick feel for a data set.

■ Example 6.1

Test scores on the Scholastic Aptitude Test (SAT) verbal portion are normally distributed with a mean score of 504. If the standard deviation of a score is 84, then we can conclude that approximately 68 percent of all scores are between $504 - 84$ and $504 + 84$. That is, approximately 68 percent of the scores are between 420 and 588. Also, approximately 95 percent of them are between $504 - 168 = 336$ and $504 + 168 = 672$; and approximately 99.7 percent are between 252 and 756. ■

The approximation rule is the theoretical basis of the empirical rule of Sec. 3.6. The connection between these rules will become apparent in Chap. 8, when we show how a sample mean and sample standard deviation can be used to estimate the quantities μ and σ.

By using the symmetry of the normal curve about the value μ, we can obtain other facts from the approximation rule. For instance, since the area between μ and $\mu + \sigma$ is the same as that between $\mu - \sigma$ and μ, it follows from this rule

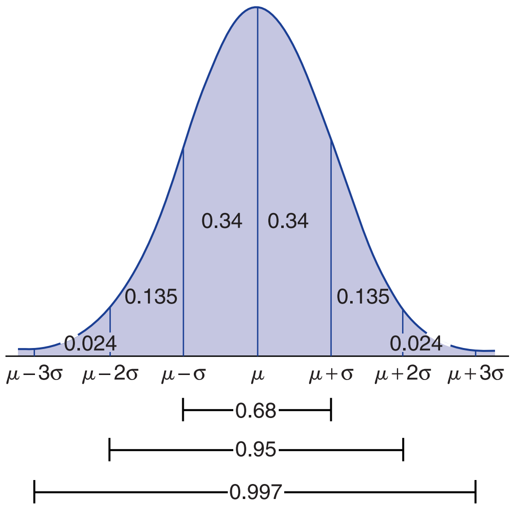


FIGURE 6.6

Approximate areas under a normal curve.

that a normal random variable will be between μ and $\mu + \sigma$ with approximate probability $0.68/2 = 0.34$.

PROBLEMS

1. The systolic blood pressures of adults, in the appropriate units, are normally distributed with a mean of 128.4 and a standard deviation of 19.6.
   (a) Give an interval in which the blood pressures of approximately 68 percent of the population fall.
   (b) Give an interval in which the blood pressures of approximately 95 percent of the population fall.
   (c) Give an interval in which the blood pressures of approximately 99.7 percent of the population fall.
2. The heights of a certain population of males are normally distributed with mean 69 inches and standard deviation 6.5 inches. Approximate the proportion of this population whose height is less than 82 inches.

Problems 3 through 16 are multiple-choice problems. Give the answer you think is closest to the true answer. Remember, Z always refers to a standard normal random variable. Draw a picture in each case to justify your answer.

3. $P\{-2 < Z < 2\}$ is approximately
   (a) 0.68 (b) 0.95 (c) 0.975 (d) 0.50

4. $P\{Z > -1\}$ is approximately
   (a) 0.50 (b) 0.95 (c) 0.84 (d) 0.16
5. $P\{Z > 1\}$ is approximately
   (a) 0.50 (b) 0.95 (c) 0.84 (d) 0.16
6. $P\{Z > 3\}$ is approximately
   (a) 0.30 (b) 0.05 (c) 0 (d) 0.99
7. $P\{Z < 2\}$ is approximately
   (a) 0.95 (b) 0.05 (c) 0.975 (d) 0.025

In Probs. 8 to 11, X is a normal random variable with expected value 15 and standard deviation 4.

8. The probability that X is between 11 and 19 is approximately
   (a) 0.50 (b) 0.95 (c) 0.68 (d) 0.34
9. The probability that X is less than 23 is approximately
   (a) 0.975 (b) 0.95 (c) 0.68 (d) 0.05
10. The probability that X is less than 11 is approximately
   (a) 0.34 (b) 0.05 (c) 0.16 (d) 0.50
11. The probability that X is greater than 27 is approximately
   (a) 0.05 (b) 0 (c) 0.01 (d) 0.32
12. Variable X is a normal random variable with standard deviation 3. If the probability that X is between 7 and 19 is 0.95, then the expected value of X is approximately
   (a) 16 (b) 15 (c) 14 (d) 13
13. Variable X is a normal random variable with standard deviation 3. If the probability that X is less than 16 is 0.84, then the expected value of X is approximately
   (a) 16 (b) 15 (c) 14 (d) 13
14. Variable X is a normal random variable with standard deviation 3. If the probability that X is greater than 16 is 0.975, then the expected value of X is approximately
   (a) 20 (b) 22 (c) 23 (d) 25
15. Variable X is a normal random variable with expected value 100. If the probability that X is greater than 90 is 0.84, then the standard deviation of X is approximately
   (a) 5 (b) 10 (c) 15 (d) 20
16. Variable X is a normal random variable with expected value 100. If the probability that X is greater than 130 is 0.025, then the standard deviation of X is approximately
   (a) 5 (b) 10 (c) 15 (d) 20
17. If X is normal with expected value 100 and standard deviation 2, and Y is normal with expected value 100 and standard deviation 4, is X or is Y more likely to
   (a) Exceed 104 (b) Exceed 96 (c) Exceed 100

18. If X is normal with expected value 100 and standard deviation 2, and Y is normal with expected value 105 and standard deviation 10, is X or is Y more likely to

(a) Exceed 105 **(b)** Be less than 95

19. The scores on a particular job aptitude test are normal with expected value 400 and standard deviation 100. If a company will consider only those applicants scoring in the top 5 percent, determine whether it should consider one whose score is

(a) 400 **(b)** 450 **(c)** 500 **(d)** 600

6.4 PROBABILITIES ASSOCIATED WITH A STANDARD NORMAL RANDOM VARIABLE

Let Z be a standard normal random variable. That is, Z is a normal random variable with mean 0 and standard deviation 1. The probability that Z is between two numbers a and b is equal to the area under the standard normal curve between a and b. Areas under this curve have been computed, and tables have been prepared that enable us to find these probabilities. One such table is Table 6.1.

For each nonnegative value of x, Table 6.1 specifies the probability that Z is less than x. For instance, suppose we want to determine $P\{Z < 1.22\}$. To do this, first we must find the entry in the table corresponding to $x = 1.22$. This is done by first searching the left-hand column to find the row labeled 1.2 and then searching the top row to find the column labeled 0.02. The value that is in both the row labeled 1.2 and the column labeled 0.02 is the desired probability. Since this value is 0.8888, we see that

$$P\{Z < 1.22\} = 0.8888$$

A portion of Table 6.1 illustrating the preceding is presented here:

x		**0.00**	**0.01**	**0.02**	**0.03**	**...**	**0.09**
0.0	0.5000	0.5040					
⋮							
1.1	0.8413						
1.2	0.8849	0.8869	0.8888				
1.3	0.9032						

We can also use Table 6.1 to determine the probability that Z is greater than x. For instance, suppose we want to determine the probability that Z is greater than 2. To accomplish this, we note that either Z is less than or equal to 2 or Z is greater than 2, and so

$$P\{Z \leq 2\} + P\{Z > 2\} = 1$$

Table 6.1 Standard Normal Probabilities

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Data value in table is $P\{Z < x\}$.

or

$$\begin{aligned}P\{Z > 2\} &= 1 - P\{Z \leq 2\}\\ &= 1 - 0.9772\\ &= 0.0228\end{aligned}$$

In other words, the probability that Z is larger than x can be obtained by subtracting from 1 the probability that Z is smaller than x. That is, for any x,

$$P\{Z > x\} = 1 - P\{Z \leq x\}$$

■ Example 6.2

Find

(a) $P\{Z < 1.5\}$
(b) $P\{Z \geq 0.8\}$

Solution

(a) From Table 6.1,

$$P\{Z < 1.5\} = 0.9332$$

(b) From Table 6.1, $P\{Z < 0.8\} = 0.7881$ and so

$$P\{Z \geq 0.8\} = 1 - 0.7881 = 0.2119$$ ■

While Table 6.1 specifies $P\{Z < x\}$ for only nonnegative values of x, it can be used even when x is negative. Probabilities for negative x are obtained from the table by making use of the symmetry about zero of the standard normal curve. For instance, suppose we want to calculate the probability that Z is less than -2. By symmetry (see Fig. 6.7), this is the same as the probability that Z is greater than 2; and so

$$\begin{aligned}P\{Z < -2\} &= 1 - P\{Z > 2\}\\ &= 1 - P\{Z < 2\}\\ &= 1 - 0.9772 = 0.0028\end{aligned}$$

In general, for any value of x,

$$P\{Z < -x\} = P\{Z > x\} = 1 - P\{Z < x\}$$

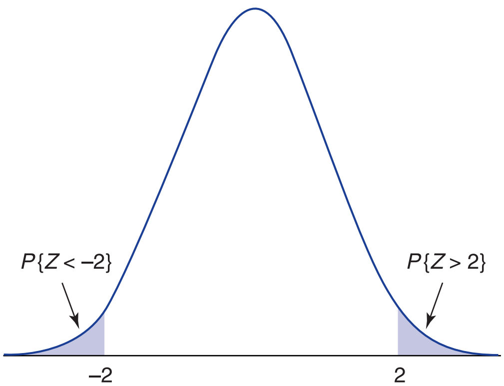


FIGURE 6.7
$P\{Z < -2\} = P\{Z > 2\}$.

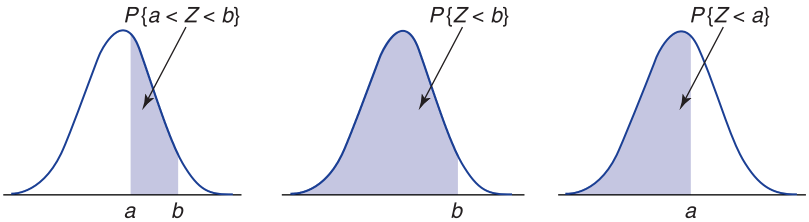


FIGURE 6.8
$P\{a < Z < b\} = P\{Z < b\} - P\{Z < a\}$.

We can determine the probability that Z lies between a and b, for $a < b$, by determining the probability that Z is less than b and then subtracting from this the probability that Z is less than a. (See Fig. 6.8.)

■ Example 6.3

Find

(a) $P\{1 < Z < 2\}$
(b) $P\{-1.5 < Z < 2.5\}$

Solution

(a) $$P\{1 < Z < 2\} = P\{Z < 2\} - P\{Z < 1\} = 0.9772 - 0.8413 = 0.1359$$

(b) $$\begin{aligned} P\{-1.5 < Z < 2.5\} &= P\{Z < 2.5\} - P\{Z < -1.5\} \\ &= P\{Z < 2.5\} - P\{Z > 1.5\} \\ &= 0.9938 - (1 - 0.9332) \\ &= 0.9270 \end{aligned}$$

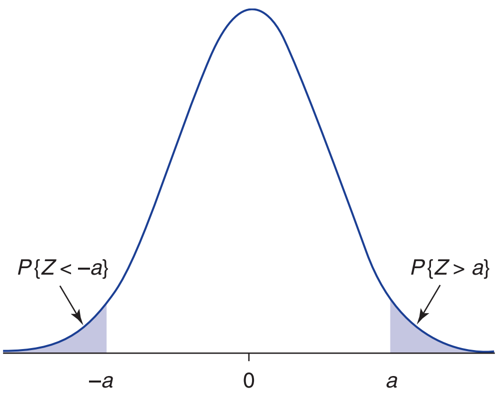


FIGURE 6.9
$P\{Z > a\} = P\{Z < -a\}$.

Let a be positive and consider $P\{|Z| > a\}$, the probability that a standard normal is, in absolute value, larger than a. Since $|Z|$ will exceed a if either $Z > a$ or $Z < -a$, we see that

$$\begin{aligned} P\{|Z| > a\} &= P\{Z > a\} + P\{Z < -a\} \\ &= 2P\{Z > a\} \end{aligned}$$

where the last equality uses the symmetry of the standard normal density curve (Fig. 6.9). ■

■ Example 6.4

Find $P\{|Z| > 1.8\}$.

Solution

$$\begin{aligned} P\{|Z| > 1.8\} &= 2P\{Z > 1.8\} \\ &= 2(1 - 0.9641) \\ &= 0.0718 \end{aligned}$$

■

Another easily established result is that for any positive value of a

$$P\{-a < Z < a\} = 2P\{Z < a\} - 1$$

The verification of this result is left as an exercise.

Table 6.1 is also listed as Table D.1 in App. D. In addition, Program 6-1 can be used to obtain normal probabilities. You enter the value x, and the program outputs $P\{Z < x\}$.

■ Example 6.5

Determine $P\{Z > 0.84\}$.

Solution

We can either use Table 6.1 or run Program 6-1, which computes the probability that a standard normal random variable is less than x. Running Program 6-1, we learn that if the desired value of x is 0.84, the probability is 0.7995459.

The desired probability is $1 - 0.80 = 0.20$. That is, there is a 20 percent chance that a standard normal random variable will exceed 0.84. ■

PROBLEMS

1. For a standard normal random variable Z find
   (a) $P\{Z < 2.2\}$
   (b) $P\{Z > 1.1\}$
   (c) $P\{0 < Z < 2\}$
   (d) $P\{-0.9 < Z < 1.2\}$
   (e) $P\{Z > -1.96\}$
   (f) $P\{Z < -0.72\}$
   (g) $P\{|Z| < 1.64\}$
   (h) $P\{|Z| > 1.20\}$
   (i) $P\{-2.2 < Z < 1.2\}$
2. Show that $-Z$ is also a standard normal random variable. *Hint*: It suffices to show that, for all x,

$$P\{-Z < x\} = P\{Z < x\}$$

3. Find the value of the question mark:

$$P\{-3 < Z < -2\} = P\{2 < Z <?\}$$

   Use a picture to show that your answer is correct.
4. Use a picture of the standard normal curve to show that

$$P\{Z > -2\} = P\{Z < 2\}$$

5. Argue, using either pictures or equations, that for any positive value of a,

$$P\{-a < Z < a\} = 2P\{Z < a\} - 1$$

6. Find
   (a) $P\{-1 < Z < 1\}$
   (b) $P\{|Z| < 1.4\}$
7. Find the value of x, to two decimal places, for which
   (a) $P\{Z > x\} = 0.05$
   (b) $P\{Z > x\} = 0.025$
   (c) $P\{Z > x\} = 0.005$
   (d) $P\{Z < x\} = 0.50$
   (e) $P\{Z < x\} = 0.66$
   (f) $P\{|Z| < x\} = 0.99$
   (g) $P\{|Z| < x\} = 0.75$
   (h) $P\{|Z| > x\} = 0.90$
   (i) $P\{|Z| > x\} = 0.50$

6.5 FINDING NORMAL PROBABILITIES: CONVERSION TO THE STANDARD NORMAL

Let X be a normal random variable with mean μ and standard deviation σ. We can determine probabilities concerning X by using the fact that the variable Z defined by

$$Z = \frac{X - \mu}{\sigma}$$

has a standard normal distribution. That is, if we *standardize* a normal random variable by subtracting its mean and then dividing by its standard deviation, the resulting variable has a standard normal distribution.

The value of the standardized variable tells us how many standard deviations the original variable is from its mean. For instance, if the standardized variable Z has value 2, then

$$Z = \frac{X - \mu}{\sigma} = 2$$

or

$$X - \mu = 2\sigma$$

That is, X is larger than its mean by 2 standard deviations.

We can compute any probability statement about X by writing an equivalent statement in terms of $Z = (X - \mu)/\sigma$ and then making use of Table 6.1 or Program 6-1. For instance, suppose we want to compute $P\{X < a\}$. Since $X < a$ is equivalent to the statement

$$\frac{X - \mu}{\sigma} < \frac{a - \mu}{\sigma}$$

we see that

$$P\{X < a\} = P\left\{\frac{X-\mu}{\sigma} < \frac{a-\mu}{\sigma}\right\}$$
$$= P\left\{Z < \frac{a-\mu}{\sigma}\right\}$$

where Z is a standard normal random variable.

■ Example 6.6

IQ examination scores for sixth-graders are normally distributed with mean value 100 and standard deviation 14.2.

(a) What is the probability a randomly chosen sixth-grader has a score greater than 130?

(b) What is the probability a randomly chosen sixth-grader has a score between 90 and 115?

Solution

Let X denote the score of a randomly chosen student. We compute probabilities concerning X by making use of the fact that the standardized variable

$$Z = \frac{X - 100}{14.2}$$

has a standard normal distribution.

(a) $P\{X > 130\} = P\{\frac{X-100}{14.2} > \frac{130-100}{14.2}\}$
$= P[Z > 2.1127]$
$= 0.017$

(b) The inequality $90 < X < 115$ is equivalent to

$$\frac{90-100}{14.2} < \frac{X-100}{14.2} < \frac{115-100}{14.2}$$

or, equivalently,

$$-0.7042 < Z < 1.056$$

Therefore,

$$P\{90 < X < 115\} = P\{-0.7042 < Z < 1.056\}$$
$$= P\{Z < 1.056\} - P\{Z < -0.7042\}$$
$$= 0.854 - 0.242$$
$$= 0.612$$

■

■ Example 6.7

Let X be normal with mean μ and standard deviation σ. Find

(a) $P\{|X - \mu| > \sigma\}$
(b) $P\{|X - \mu| > 2\sigma\}$
(c) $P\{|X - \mu| > 3\sigma\}$

Solution

The statement $|X - \mu| > a\sigma$ is, in terms of the standardized variable $Z = (X - \mu)/\sigma$, equivalent to the statement $|Z| > a$. Using this fact, we obtain the following results.

(a) $$\begin{aligned} P\{|X - \mu| > \sigma\} &= P\{|Z| > 1\} \\ &= 2P\{Z > 1\} \\ &= 2(1 - 0.8413) \\ &= 0.3174 \end{aligned}$$

(b) $$\begin{aligned} P\{|X - \mu| > 2\sigma\} &= P\{|Z| > 2\} \\ &= 2P\{Z > 2\} \\ &= 0.0456 \end{aligned}$$

(c) $$\begin{aligned} P\{|X - \mu| > 3\sigma\} &= P\{|Z| > 3\} \\ &= 2P\{Z > 3\} \\ &= 0.0026 \end{aligned}$$

Thus, we see that the probability that a normal random variable differs from its mean by more than 1 standard deviation is (to two decimal places) 0.32; or equivalently, the complementary probability that it is within 1 standard deviation of its mean is 0.68. Similarly, parts (b) and (c) imply, respectively, that the probability the random variable is within 2 standard deviations of its mean is 0.95 and the probability that it is within 3 standard deviations of its mean is 0.997. Thus we have verified the approximation rule presented in Sec. 6.3. ■

6.6 ADDITIVE PROPERTY OF NORMAL RANDOM VARIABLES

The fact that $Z = (X - \mu)/\sigma$ is a standard normal random variable follows from the fact that if one either adds or multiplies a normal random variable by a constant, then the resulting random variable remains normal. As a result, if X is normal with mean μ and standard deviation σ, then $Z = (X - \mu)/\sigma$ also will be normal. It is now easy to verify that Z has expected value 0 and variance 1.

An important fact about normal random variables is that the sum of independent normal random variables is also a normal random variable. That is, if X and Y are independent normal random variables with respective parameters μ_x, σ_x and

μ_y, σ_y, then $X + Y$ also will be normal. Its mean value will be

$$E[X + Y] = E[X] + E[Y] = \mu_x + \mu_y$$

Its variance is

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) = \sigma_x^2 + \sigma_y^2$$

That is, we have the following result.

Suppose X and Y are independent normal random variables with means μ_x and μ_y and standard deviations σ_x and σ_y, respectively. Then $X + Y$ is normal with mean

$$E[X + Y] = \mu_x + \mu_y$$

and standard deviation

$$\text{SD}(X + Y) = \sqrt{\sigma_x^2 + \sigma_y^2}$$

■ Example 6.8

Suppose the amount of time a light bulb works before burning out is a normal random variable with mean 400 hours and standard deviation 40 hours. If an individual purchases two such bulbs, one of which will be used as a spare to replace the other when it burns out, what is the probability that the total life of the bulbs will exceed 750 hours?

Solution

We need to compute the probability that $X + Y > 750$, where X is the life of the bulb used first and Y is the life of the other bulb. Variables X and Y are both normal with mean 400 and standard deviation 40. In addition, we will suppose they are independent, and so $X + Y$ is also normal with mean 800 and standard deviation $\sqrt{40^2 + 40^2} = \sqrt{3200}$. Therefore, $Z = (X + Y - 800)/\sqrt{3200}$ has a standard normal distribution. Thus, we have

$$\begin{aligned} P[X + Y > 750] &= P\left\{\frac{X + Y - 800}{\sqrt{3200}} > \frac{750 - 800}{\sqrt{3200}}\right\} \\ &= P\{Z > -0.884\} \\ &= P\{Z < 0.884\} \\ &= 0.81 \end{aligned}$$

Therefore, there is an 81 percent chance that the total life of the two bulbs exceeds 750 hours. ■

■ Example 6.9

Data from the U.S. Department of Agriculture indicate that the annual amount of apples eaten by a randomly chosen woman is normally distributed with a mean of 19.9 pounds and a standard deviation of 3.2 pounds, whereas the amount eaten by a randomly chosen man is normally distributed with a mean of 20.7 pounds and a standard deviation of 3.4 pounds. Suppose a man and a woman are randomly chosen. What is the probability that the woman ate a greater amount of apples than the man?

Solution

Let X denote the amount eaten by the woman and Y the amount eaten by the man. We want to determine $P\{X > Y\}$, or equivalently $P\{X - Y > 0\}$. Now X is a normal random variable with mean 19.9 and standard deviation 3.2. Also $-Y$ is a normal random variable (since it is equal to the normal random variable Y multiplied by the constant -1) with mean -20.7 and standard deviation $|-1|(3.4) = 3.4$. Therefore, their sum $X + (-Y)(= X - Y)$ is normal with mean

$$E[X - Y] = 19.9 + (-20.7) = -0.8$$

and standard deviation

$$\mathrm{SD}(X - Y) = \sqrt{(3.2)^2 + (3.4)^2} = 4.669$$

Thus, if we let $W = X - Y$, then

$$\begin{aligned} P\{W > 0\} &= P\left\{\frac{W + 0.8}{4.669} > \frac{0.8}{4.669}\right\} \\ &= P[Z > 0.17] \\ &= 1 - 0.5675 = 0.4325 \end{aligned}$$

That is, with probability 0.4325 the randomly chosen woman would have eaten a greater amount of apples than the randomly chosen man. ■

PROBLEMS

1. Explain carefully why the inequality

$$x > a$$

is equivalent to the inequality

$$\frac{x - \mu}{\sigma} > \frac{a - \mu}{\sigma}$$

What fact are we using about the quantity? (*Hint*: Would these inequalities be equivalent if σ were negative?)

2. If X is normal with mean 10 and standard deviation 3, find
   (a) $P\{X > 12\}$
   (b) $P\{X < 13\}$
   (c) $P\{8 < X < 11\}$
   (d) $P\{X > 7\}$
   (e) $P\{|X - 10| > 5\}$
   (f) $P\{X > 10\}$
   (g) $P\{X > 20\}$
3. The length of time that a new hair dryer functions before breaking down is normally distributed with mean 40 months and standard deviation 8 months. The manufacturer is thinking of guaranteeing each dryer for 3 years. What proportion of dryers will not meet this guarantee?
4. The scores on a scholastic achievement test were normally distributed with mean 520 and standard deviation 94.
   (a) If your score was 700, by how many standard deviations did it exceed the average score?
   (b) What percentage of examination takers received a higher score than you did?
5. The number of bottles of shampoo sold monthly by a certain discount drugstore is a normal random variable with mean 212 and standard deviation 40. Find the probability that next month's shampoo sales will be
   (a) Greater than 200
   (b) Less than 250
   (c) Greater than 200 but less than 250
6. The life of a certain automobile tire is normally distributed with mean 35,000 miles and standard deviation 5000 miles.
   (a) What proportion of such tires last between 30,000 and 40,000 miles?
   (b) What proportion of such tires last over 40,000 miles?
   (c) What proportion last over 50,000 miles?
7. Suppose you purchased such a tire as described in Prob. 6. If the tire is in working condition after 40,000 miles, what is the conditional probability that it will still be working after an additional 10,000 miles?
8. The pulse rate of young adults is normally distributed with a mean of 72 beats per minute and a standard deviation of 9.5 beats per minute. If the requirements for the military rule out anyone whose rate is over 95 beats per minute, what percentage of the population of young adults does not meet this standard?

9. The time required to complete a certain loan application form is a normal random variable with mean 90 minutes and standard deviation 15 minutes. Find the probability that an application form is filled out in
   (a) Less than 75 minutes
   (b) More than 100 minutes
   (c) Between 90 and 120 minutes
10. The bolts produced by a manufacturer are specified to be between 1.09 and 1.11 inches in diameter. If the production process results in the diameter of bolts being a normal random variable with mean 1.10 inches and standard deviation 0.005 inch, what percentage of bolts do not meet the specifications?
11. The activation pressure of a valve produced by a certain company is a normal random variable with expected value 26 pounds per square inch and standard deviation 4 pounds per square inch. What percentage of the valves produced by this company have activation pressures between 20 and 32 pounds per square inch?
12. You are planning on junking your old car after it runs an additional 20,000 miles. The battery on this car has just failed, and you must decide which of two types of batteries, costing the same amount, to purchase. After some research you have discovered that the lifetime of the first battery is normally distributed with mean life 24,000 and standard deviation 6000 miles, and the lifetime of the second battery is normally distributed with mean life 22,000 and standard deviation 2000 miles.
   (a) If all you care about is that the battery purchased lasts at least 20,000 miles, which one should you buy?
   (b) What if you wanted the battery to last for 21,000 miles?
13. Value at Risk (VAR) has become a key concept in financial calculations. The VAR of an investment is defined as that value v such that there is only a 1 percent chance that the loss from the investment will be greater than v.
   (a) If the gain from an investment is a normal random variable with mean 10 and variance 49 determine the VAR. (If X is the gain, then $-X$ is the loss.)
   (b) Among a set of investments all of whose gains are normally distributed, show that the one having the smallest VAR is the one having the largest value of $\mu - 2.33\sigma$, where μ and σ are the mean and variance of the gain from the investment.
14. The annual rainfall in Cincinnati, Ohio, is normally distributed with mean 40.14 inches and standard deviation 8.7 inches.
   (a) What is the probability that this year's rainfall exceeds 42 inches?
   (b) What is the probability that the sum of the next 2 years' rainfall exceeds 84 inches?

(c) What is the probability that the sum of the next 3 years' rainfall exceeds 126 inches?
(d) For parts (b) and (c), what independence assumptions are you making?

15. The height of adult women in the United States is normally distributed with mean 64.5 inches and standard deviation 2.4 inches. Find the probability that a randomly chosen woman is
    (a) Less than 63 inches tall
    (b) Less than 70 inches tall
    (c) Between 63 and 70 inches tall
    (d) Alice is 72 inches tall. What percentage of women are shorter than Alice?
    (e) Find the probability that the average of the heights of two randomly chosen women is greater than 67.5 inches.
16. The weight of an introductory chemistry textbook is a normal random variable with mean 3.5 pounds and standard deviation 2.2 pounds, whereas the weight of an introductory economics textbook is a normal random variable with mean 4.6 pounds and standard deviation 1.3 pounds. If Alice is planning on taking introductory courses in both chemistry and economics, find the probability that
    (a) The total weight of her two books will exceed 9 pounds.
    (b) Her economics book will be heavier than her chemistry book.
    (c) What assumption have you made?
17. The weekly demand for a product approximately has a normal distribution with mean 1,000 and standard deviation 200. The current on-hand inventory is 2200 and no deliveries will be occurring in the next two weeks. Assuming that the demands in different weeks are independent,
    (a) What is the probability that the demand in each of the next two weeks is less than 1100?
    (b) What is the probability that the total of the demands in the next two weeks exceeds 2200?

6.7 PERCENTILES OF NORMAL RANDOM VARIABLES

For any α between 0 and 1, we define z_α to be that value for which

$$P\{Z > z_\alpha\} = \alpha$$

In words, the probability that a standard normal random variable is greater than z_α is equal to α (see Fig. 6.10).

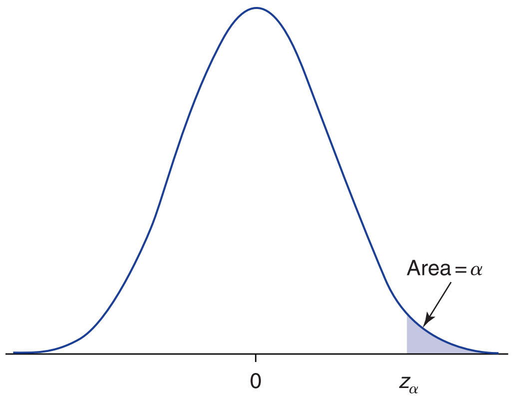


FIGURE 6.10
$P\{Z > z_\alpha\} = \alpha$.

We can determine the value of z_α by using Table 6.1. For instance, suppose we want to find $z_{0.025}$. Since

$$P\{Z < z_{0.025}\} = 1 - P\{Z > z_{0.025}\} = 0.975$$

we search in Table 6.1 for the entry 0.975, and then we find the value x that corresponds to this entry. Since the value 0.975 is found in the row labeled 1.9 and the column labeled 0.06, we see that

$$z_{0.025} = 1.96$$

That is, 2.5 percent of the time a standard normal random variable will exceed 1.96.

Since 97.5 percent of the time a standard normal random variable will be less than 1.96, we say that 1.96 is the 97.5 percentile of the standard normal distribution. In general, since $100(1 - \alpha)$ percent of the time a standard normal random variable will be less than z_α, we call z_α the $100(1 - \alpha)$ *percentile* of the standard normal distribution.

Suppose now that we want to find $z_{0.05}$. If we search Table 6.1 for the value 0.95, we do not find this exact value. Rather, we see that

$$P\{Z < 1.64\} = 0.9495$$

and

$$P\{Z < 1.65\} = 0.9505$$

Therefore, it would seem that $z_{0.05}$ lies roughly halfway between 1.64 and 1.65, and so we could approximate it by 1.645. In fact, it turns out that, to three decimal places, this is the correct answer, and so

$$z_{0.05} = 1.645$$

The values of $z_{0.10}$, $z_{0.05}$, $z_{0.025}$, $z_{0.01}$, and $z_{0.005}$, are, as we will see in later chapters, of particular importance in statistics. Their values are as follows:

$$z_{0.10} = 1.282 \quad z_{0.025} = 1.960 \quad z_{0.005} = 2.576$$
$$z_{0.05} = 1.645 \quad z_{0.01} = 2.326$$

For all other values of α, we can use Table 6.1 to find z_α by searching for the entry that is closest to $1 - \alpha$. In addition, Program 6-2 can be used to obtain z_α.

■ Example 6.10

Find

(a) $z_{0.25}$
(b) $z_{0.80}$

Solution

(a) The 75th percentile $z_{0.25}$ is the value for which

$$P\{Z > z_{0.25}\} = 0.25$$

or, equivalently,

$$P\{Z < z_{0.25}\} = 0.75$$

The closest entry to 0.75 in Table 6.1 is the entry 0.7486, which corresponds to the value 0.67. Thus, we see that

$$z_{0.25} \approx 0.67$$

A more precise value could be obtained by running Program 6-2. This gives the following: If a is equal to 0.25, we learn that the value of $z_{0.25}$ is 0.6744897.

(b) We are asked to find the value $z_{0.80}$ such that

$$P\{Z > z_{0.80}\} = 0.80$$

Now the value of $z_{0.80}$ will be negative (why is this?), and so it is best to write the equivalent equation (see Fig. 6.11)

$$P\{Z < -z_{0.80}\} = 0.80$$

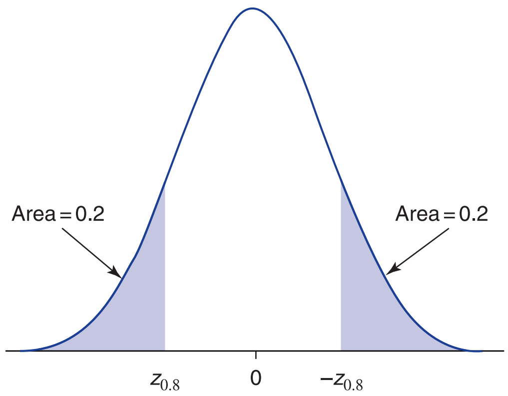


FIGURE 6.11
$P\{Z < -z_{0.8}\} = 0.80$.

From Table 6.1 we see that

$$-z_{0.80} \approx 0.84$$

and so

$$z_{0.80} \approx -0.84$$

We can easily obtain the percentiles of any normal random variable by converting to the standard normal. For instance, suppose we want to find the value x for which

$$P\{X < x\} = 0.95$$

when X is normal with mean 40 and standard deviation 5. By writing the inequality $X < x$ in terms of the standardized variable $Z = (X - 40)/5$, we see that

$$\begin{aligned} 0.95 &= P\{X < x\} \\ &= P\left\{\frac{X-40}{5} < \frac{x-40}{5}\right\} \\ &= P\left\{Z < \frac{x-40}{5}\right\} \end{aligned}$$

But $P\{Z < z_{0.05}\} = 0.95$, and so we obtain

$$\frac{x-40}{5} = z_{0.05} = 1.645$$

and so the desired value of x is

$$x = 5(1.645) + 40 = 48.225$$

■ Example 6.11

An IQ test produces scores that are normally distributed with mean value 100 and standard deviation 14.2. The top 1 percent of all scores is in what range?

Solution

We want to find the value of x for which

$$P\{X > x\} = 0.01$$

when X is normal with mean 100 and standard deviation 14.2. Now

$$P\{X > x\} = P\left\{\frac{X - 100}{14.2} > \frac{x - 100}{14.2}\right\}$$
$$= P\left\{Z > \frac{x - 100}{14.2}\right\}$$

Since $P\{Z > z_{0.01}\} = 0.01$, it follows that the preceding probability will equal 0.01 if

$$\frac{x - 100}{14.2} = z_{0.01} = 2.33$$

and so

$$x = 14.2(2.33) + 100 = 133.086$$

That is, the top 1 percent consists of all those having scores above 134. ■

Figure 6.12 illustrates the result

$$P\{X > \mu + \sigma z_\alpha\} = \alpha$$

when X is a normal random variable with mean μ and standard deviation σ.

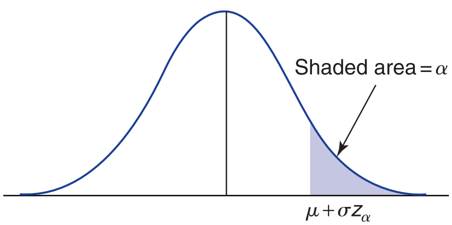


FIGURE 6.12
$P\{X > \mu + \sigma z_\alpha\} = \alpha$.

PROBLEMS

1. Find to two decimal places:
   (a) $z_{0.07}$
   (b) $z_{0.12}$
   (c) $z_{0.30}$
   (d) $z_{0.03}$
   (e) $z_{0.65}$
   (f) $z_{0.50}$
   (g) $z_{0.95}$
   (h) $z_{0.008}$
2. Find the value of x for which
   (a) $P\{|Z| > x\} = 0.05$
   (b) $P\{|Z| > x\} = 0.025$
   (c) $P\{|Z| > x\} = 0.005$
3. If X is a normal random variable with mean 50 and standard deviation 6, find the approximate value of x for which
   (a) $P\{X > x\} = 0.5$
   (b) $P\{X > x\} = 0.10$
   (c) $P\{X > x\} = 0.025$
   (d) $P\{X < x\} = 0.05$
   (e) $P\{X < x\} = 0.88$
4. Scores on an examination for real estate brokers are normally distributed with mean 420 and standard deviation 66. If the real estate board wants to designate the highest 10 percent of all scores as *excellent*, at what score should excellence begin?
5. Suppose in Prob. 4 that the board wants only the highest 25 percent to pass. What should be set as the passing score?
6. The time it takes for high school boys to run 1 mile is normally distributed with mean 460 seconds and standard deviation 40 seconds. All those falling in the slowest 20 percent are deemed to need additional training. What is the critical time above which one is deemed to need additional training?
7. In Prob. 6, the fastest 5 percent all ran the mile in less than x seconds. What is the smallest value of x for which the preceding is a true statement?
8. Repeat Prob. 7, replacing *fastest 5 percent* by *fastest 1 percent*.
9. The amount of radiation that can be absorbed by an individual before death ensues varies from individual to individual. However, over the entire population this amount is normally distributed with mean 500 roentgens and standard deviation 150 roentgens. Above what dosage level will only 5 percent of those exposed survive?

10. Transmissions on a new car will last for a normally distributed amount of time with mean 70,000 miles and standard deviation 10,000 miles. A warranty on this part is to be provided by the manufacturer. If the company wants to limit warranty work to no more than 20 percent of the cars sold, what should be the length (in miles) of the warranty period?
11. The attendance at home football games of a certain college is a normal random variable with mean 52,000 and standard deviation 4000. Which of the following statements are true?
    (a) Over 80 percent of the games have an attendance of over 46,000.
    (b) The attendance exceeds 58,000 less than 10 percent of the time.
12. Scores on the quantitative part of the Graduate Record Examination were normally distributed with a mean score of 510 and a standard deviation of 92. How high a score was necessary to be in the top
    (a) 10
    (b) 5
    (c) 1

    percent of all scores?
13. The fasting blood glucose level (per 100 milliliters of blood) of diabetics is normally distributed with mean 106 milligrams and standard deviation 8 milligrams. In order for the blood glucose level of a diabetic to be in the lower 20 percent of all diabetics, that person's blood level must be less than what value?

KEY TERMS

Continuous random variable: A random variable that can take on any value in some interval.

Probability density function: A curve associated with a continuous random variable. The probability that the random variable is between two points is equal to the area under the curve between these points.

Normal random variable: A type of continuous random variable whose probability density function is a bell-shaped symmetric curve.

Standard normal random variable: A normal random variable having mean 0 and variance 1.

100*p* percentile of a continuous random variable: The probability that the random variable is less than this value is p.

SUMMARY

A *continuous* random variable is one that can assume any value within some interval. Its probabilities can be obtained from its *probability density curve*.

Specifically, the probability that the random variable will lie between points a and b will equal the area under the density curve between a and b.

A *normal* random variable X has a probability density curve that is specified by two parameters, the mean μ and the standard deviation σ of X. The density curve is a bell-shaped curve that is symmetric about μ and spreads out more as the value of σ gets larger.

A normal random variable will take on a value that is within 1 standard deviation of its mean approximately 68 percent of the time; it will take on a value that is within 2 standard deviations of its mean approximately 95 percent of the time; and it will take on a value that is within 3 standard deviations of its mean approximately 99.7 percent of the time.

A normal random variable having mean 0 and standard deviation 1 is called a *standard normal* random variable. We let Z designate such a random variable. Probabilities of a standard normal random variable can be obtained from Table 6.1 (reprinted as Table D.1). For any nonnegative value x, specified up to two decimal places, this table gives the probability that a standard normal random variable is less than x. For negative x, this probability can be obtained by making use of the symmetry of the normal curve about 0. This results in the equality

$$P\{Z < x\} = P\{Z > -x\}$$

The value of $P\{Z > -x\} = 1 - P\{Z < -x\}$ can now be obtained from Table 6.1.

Program 6-1 can also be used to obtain probabilities of standard normal random variables.

If X is normal with mean μ and standard deviation σ, then Z, defined by

$$Z = \frac{X - \mu}{\sigma}$$

Historical Perspective

The Normal Curve

(Bettmann)

Karl F. Gauss

The normal distribution was introduced by the French mathematician Abraham De Moivre in 1733. De Moivre, who used this distribution to approximate probabilities connected with coin tossing, called it the *exponential bell-shaped curve*. Its usefulness, however, only became truly apparent in 1809 when the famous German mathematician K. F. Gauss used it as an integral part of his approach to predicting the location of astronomical entities. As a result, it became common after this time to call it the *Gaussian distribution*.

During the middle to late 19th century, however, most statisticians started to believe that the majority of data sets would have histograms conforming to the

Gaussian bell-shaped form. Indeed, it came to be accepted that it was "normal" for any well-behaved data set to follow this curve. As a result, following the lead of Karl Pearson, people began referring to the Gaussian curve as simply the *normal* curve. (For an explanation as to why so many data sets conform to the normal curve, the interested student will have to wait to read Secs. 7.3 and 12.6.)

Karl Friedrich Gauss (1777–1855), one of the earliest users of the normal curve, was one of the greatest mathematicians of all time. Listen to the words of the well-known mathematical historian E. T. Bell, as expressed in his 1954 book *Men of Mathematics*. In a chapter entitled "The Prince of Mathematicians," he writes:

> Archimedes, Newton, and Gauss; these three are in a class by themselves among the great mathematicians, and it is not for ordinary mortals to attempt to rank them in order of merit. All three started tidal waves in both pure and applied mathematics. Archimedes esteemed his pure mathematics more highly than its applications; Newton appears to have found the chief justification for his mathematical inventions in the scientific uses to which he put them; while Gauss declared it was all one to him whether he worked on the pure or on the applied side.

has a standard normal distribution. This fact enables us to compute probabilities of X by transforming them to probabilities concerning Z. For instance,

$$\begin{aligned} P\{X < a\} &= P\left\{\frac{X-\mu}{\sigma} < \frac{a-\mu}{\sigma}\right\} \\ &= P\left\{Z < \frac{a-\mu}{\sigma}\right\} \end{aligned}$$

For any value of α between 0 and 1 the quantity z_α is defined as that value for which

$$P\{Z > z_\alpha\} = \alpha$$

Thus, a standard normal will be less than z_α with probability $1-\alpha$. That is, $100(1-\alpha)$ percent of the time Z will be less than z_α. The quantity z_α is called the $100(1-\alpha)$ percentile of the standard normal distribution.

The values z_α for specified α can be obtained either from Table 6.1 or by running Program 6-2. The percentiles of an arbitrary normal random variable X with mean μ and standard deviation σ can be obtained from the standard normal percentiles by using the fact that $Z = (X-\mu)/\sigma$ is a standard normal distribution. For instance, suppose we want to find the value x for which

$$P\{X > x\} = \alpha$$

Thus, we want to find x for which

$$\alpha = P\left\{\frac{X-\mu}{\sigma} > \frac{x-\mu}{\sigma}\right\}$$
$$= P\left\{Z > \frac{x-\mu}{\sigma}\right\}$$

Therefore, since $P\{Z > z_\alpha\} = \alpha$, we can conclude that

$$\frac{X-\mu}{\sigma} = z_\alpha$$

or

$$x = \mu + \sigma z_\alpha$$

REVIEW PROBLEMS

1. The heights of adult males are normally distributed with a mean of 69 inches and a standard deviation of 2.8 inches. Let X denote the height of a randomly chosen male adult. Find
   (a) $P\{X > 65\}$
   (b) $P\{62 < X < 72\}$
   (c) $P\{|X - 69| > 6\}$
   (d) $P\{63 < X < 75\}$
   (e) $P\{X > 72\}$
   (f) $P\{X < 60\}$
   (g) x if $P\{X > x\} = 0.01$
   (h) x if $P\{X < x\} = 0.95$
   (i) x if $P\{X < x\} = 0.40$
2. Find
   (a) $z_{0.04}$
   (b) $z_{0.22}$
   (c) $P\{Z > 2.2\}$
   (d) $P\{Z < 1.6\}$
   (e) $z_{0.78}$
3. In tests conducted on jet pilots, it was found that their blackout thresholds are normally distributed with a mean of $4.5g$ and a standard deviation of $0.7g$. If only those pilots whose thresholds are in the top 25 percent are to be allowed to apply to become astronauts, what is the cutoff point?

4. In Prob. 3, find the proportion of jet pilots who have blackout thresholds
   (a) Above $5g$
   (b) Below $4g$
   (c) Between $3.7g$ and $5.2g$
5. The working life of a certain type of light bulb is normally distributed with a mean of 500 hours and a standard deviation of 60 hours.
   (a) What proportion of such bulbs lasts more than 560 hours?
   (b) What proportion lasts less than 440 hours?
   (c) If a light bulb is still working after 440 hours of operation, what is the conditional probability that its lifetime exceeds 560 hours?
   (d) Fill in the missing number in the following sentence. Ten percent of these bulbs will have a lifetime of at least _______ hours.
6. The American Cancer Society has stated that a 25-year-old man who smokes a pack of cigarettes a day gives up, on average, 5.5 years of life. Assuming that the number of years lost is normally distributed with mean 5.5 and standard deviation 1.5, find the probability that the decrease in life of such a man is
   (a) Less than 2 years
   (b) More than 8 years
   (c) Between 4 and 7 years
7. Suppose that the yearly cost of upkeep for condominium owners at a certain complex is normal with mean \$3000 and standard deviation \$600. Find the probability that an owner's total cost over the next 2 years will
   (a) Exceed \$5000
   (b) Be less than \$7000
   (c) Be between \$5000 and \$7000

   Assume that costs incurred in different years are independent random variables.
8. The speeds of cars traveling on New Jersey highways are normally distributed with mean 60 miles per hour and standard deviation 5 miles per hour. If New Jersey police follow a policy of ticketing only the fastest 5 percent, at what speed do the police start to issue tickets?
9. The gross weekly sales at a certain used-car lot are normal with mean \$18,800 and standard deviation \$9000.
   (a) What is the probability that next week's sales exceed \$20,000?
   (b) What is the probability that weekly sales will exceed \$20,000 in each of the next 2 weeks?
   (c) What is the probability that the total sales in the next 2 weeks exceed \$40,000?

   In parts (b) and (c) assume that the sales totals in different weeks are independent.

10. The yearly number of miles accumulated by an automobile in a large car rental company's fleet is normal with mean 18,000 miles and standard deviation 1700 miles. At the end of the year the company sells 80 percent of its year-old cars, keeping the 20 percent with the lowest mileage. Do you think a car whose year-end mileage is 17,400 is likely to be kept?
11. An analysis of the scores of professional football games has led some researchers to conclude that a team that is favored by x points will outscore its opponent by a random number of points that is approximately normally distributed with mean x and standard deviation 14. Thus, for instance, the difference in the points scored by a team that is favored by 5 points and its opponent will be a normal random variable with mean 5 and standard deviation 14. Assuming that this theory is correct, determine the probability that
    (a) A team that is favored by 7 points wins the game.
    (b) A team that is a 4-point underdog wins the game.
    (c) A team that is a 14-point favorite loses the game.
12. U.S. Department of Agriculture data for 1987 indicate that the amount of tomatoes consumed per year by a randomly chosen woman is a normal random variable with a mean of 14.0 pounds and a standard deviation of 2.7 pounds, while the amount eaten yearly by a randomly chosen man is a normal random variable with a mean of 14.6 pounds and a standard deviation of 3.0 pounds. Suppose a man and a woman are randomly chosen. Find the probability that
    (a) The woman ate more than 14.6 pounds of tomatoes in 1987.
    (b) The man ate less than 14 pounds of tomatoes in 1987.
    (c) The woman ate more and the man ate less than 15 pounds of tomatoes in 1987.
    (d) The woman ate more tomatoes in 1987 than did the man.
13. Suppose in Prob. 12 that a person, equally likely to be either a man or a woman, is chosen. Find the probability that this person is
    (a) A woman who ate less than 14 pounds of tomatoes in 1987.
    (b) A man who ate more than 14 pounds of tomatoes in 1987.
14. The salaries of physicians in a certain speciality are approximately normally distributed. If 25 percent of these physicians earn below 180,000 dollars and 25 percent earn above 320,000 what fraction earn
    (a) Below 250,000 dollars?
    (b) Between 260,000 and 300,000 dollars?
15. The monthly demand for a certain product is a normal random variable with mean 50 and standard deviation 10.
    (a) If you have 60 units on stock at the beginning of the month, what is the probability you will still have stock remaining at the end of the month?

(b) What is the probability that the total demand over the next three months exceeds 180?

(c) What assumption did you make in answering part (b)?

16. The sample mean of the scores on your economics exam was 60 with a sample standard deviation of 20 while the sample mean of the scores on your statistics exam was 55 with a sample standard deviation of 10. You scored 70 on the economics exam and 62 on the statistics exam. Assume that test scores on both exams approximately follow a normal histogram.

(a) On which exam is the percentage of scores that are below your score the highest?

(b) Approximate the percentage of scores on the economics exam that were below your score.

(c) Approximate the percentage of scores on the statistics exam that were below your score.