GLOBAL EDITION

Statistics

THIRTEENTH EDITION

James McClave • Terry Sincich

Pearson

APPLET CORRELATION

Applet	Concept Illustrated	Description	Applet Activity
Sample from a population	Assesses how well a sample represents the population and the role that sample size plays in the process.	Produces random sample from population from specified sample size and population distribution shape. Reports mean, median, and standard deviation; applet creates plot of sample.	**4.4**, 240; **5.1**, 355; **5.3**, 279
Sampling distributions	Compares means and standard deviations of distributions; assesses effect of sample size; illustrates unbiasedness.	Simulates repeatedly choosing samples of a fixed size n from a population with specified sample size, number of samples, and shape of population distribution. Applet reports means, medians, and standard deviations; creates plots for both.	**6.1**, 330; **6.2**, 330
Random numbers	Uses a random number generator to determine the experimental units to be included in a sample.	Generates random numbers from a range of integers specified by the user.	**1.1**, 47; **1.2**, 48; **3.6**, 203; **4.1**, 221; **5.2**, 265
Long-run probability demonstrations illustrate the concept that theoretical probabilities are long-run experimental probabilities.			
Simulating probability of rolling a 6	Investigates relationship between theoretical and experimental probabilities of rolling 6 as number of die rolls increases.	Reports and creates frequency histogram for each outcome of each simulated roll of a fair die. Students specify number of rolls; applet calculates and plots proportion of 6s.	**3.1**, 157; **3.2**, 157; **3.3**, 168; **3.4**, 169; **3.5**, 183
Simulating probability of rolling a 3 or 4	Investigates relationship between theoretical and experimental probabilities of rolling 3 or 4 as number of die rolls increases.	Reports outcome of each simulated roll of a fair die; creates frequency histogram for outcomes. Students specify number of rolls; applet calculates and plots proportion of 3s and 4s.	**3.3**, 168; **3.4**, 169
Simulating the probability of heads: fair coin	Investigates relationship between theoretical and experimental probabilities of getting heads as number of fair coin flips increases.	Reports outcome of each fair coin flip and creates a bar graph for outcomes. Students specify number of flips; applet calculates and plots proportion of heads.	**4.2**, 221
Simulating probability of heads: unfair coin ($P(H) = .2$)	Investigates relationship between theoretical and experimental probabilities of getting heads as number of unfair coin flips increases.	Reports outcome of each flip for a coin where heads is less likely to occur than tails and creates a bar graph for outcomes. Students specify number of flips; applet calculates and plots the proportion of heads.	**4.3**, 239
Simulating probability of heads: unfair coin ($P(H) = .8$)	Investigates relationship between theoretical and experimental probabilities of getting heads as number of unfair coin flips increases.	Reports outcome of each flip for a coin where heads is more likely to occur than tails and creates a bar graph for outcomes. Students specify number of flips; applet calculates and plots the proportion of heads.	**4.3**, 239
Simulating the stock market	Theoretical probabilities are long run experimental probabilities.	Simulates stock market fluctuation. Students specify number of days; applet reports whether stock market goes up or down daily and creates a bar graph for outcomes. Calculates and plots proportion of simulated days stock market goes up.	**4.5**, 240
Mean versus median	Investigates how skewedness and outliers affect measures of central tendency.	Students visualize relationship between mean and median by adding and deleting data points; applet automatically updates mean and median.	**2.1**, 89; **2.2**, 89; **2.3**, 89

BIOGRAPHY

CARL F. GAUSS (1777–1855)

The Gaussian Distribution

The normal distribution began in the 18th century as a theoretical distribution for errors in disciplines in which fluctuations in nature were believed to behave randomly. Although he may not have been the first to discover the formula, the normal distribution was named the Gaussian distribution after Carl Friedrich Gauss. A well-known and respected German mathematician, physicist, and astronomer, Gauss applied the normal distribution while studying the motion of planets and stars. Gauss's prowess as a mathematician was exemplified by one of his most important discoveries: At the young age of 22, Gauss constructed a regular 17-gon by ruler and compasses—a feat that was the most major advance in mathematics since the time of the ancient Greeks. In addition to publishing close to 200 scientific papers, Gauss invented the heliograph as well as a primitive telegraph. ■

The normal distribution is perfectly symmetric about its mean μ, as can be seen in the examples in Figure 5.7. Its spread is determined by the value of its standard deviation σ. The formula for the normal probability distribution is shown in the next box. When plotted, this formula yields a curve like that shown in Figure 5.6.

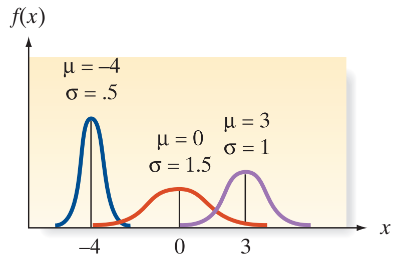


Figure 5.7
Several normal distributions with different means and standard deviations

Probability Distribution for a Normal Random Variable *x*

Probability density function: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(1/2)[(x-\mu)/\sigma]^2}$

where

μ = Mean of the normal random variable x
σ = Standard deviation
π = 3.1416 . . .
e = 2.71828 . . .

$P(x < a)$ is obtained from a table of normal probabilities.

Note that the mean μ and standard deviation σ appear in this formula, so that no separate formulas for μ and σ are necessary. To graph the normal curve, we have to know the numerical values of μ and σ. Computing the area over intervals under the normal probability distribution is a difficult task.* Consequently, we will use either technology or the computed areas listed in Table II of Appendix B. Although there are an infinitely large number of normal curves—one for each pair of values of μ and σ—we have formed a single table that will apply to any normal curve.

Table II is based on a normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$, called a *standard normal distribution*. A random variable with a standard normal distribution is typically denoted by the symbol z. The formula for the probability distribution of z is

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

Figure 5.8 shows the graph of a standard normal distribution.

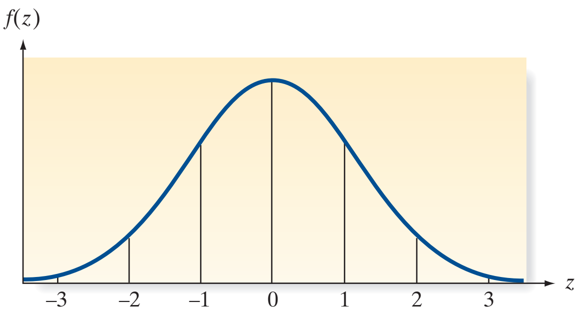


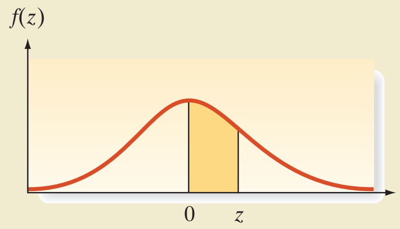
Figure 5.8
Standard normal distribution: $\mu = 0, \sigma = 1$

*The student with a knowledge of calculus should note that there is no closed-form expression for $P(a < x < b) = \int_a^b f(x)\,dx$ for the normal probability distribution. The value of this definite integral can be obtained to any desired degree of accuracy by numerical approximation. For this reason, it is tabulated for the user.

The **standard normal distribution** is a normal distribution with $\mu = 0$ and $\sigma = 1$. A random variable with a standard normal distribution, denoted by the symbol z, is called a **standard normal random variable**.

Since we will ultimately convert all normal random variables to standard normal variables in order to use Table II to find probabilities, it is important that you learn to use Table II well. A partial reproduction of that table is shown in Table 5.1. Note that the values of the standard normal random variable z are listed in the left-hand column. The entries in the body of the table give the area (probability) between 0 and z. Examples 5.3–5.6 illustrate the use of the table.

Table 5.1 Reproduction of Part of Table II in Appendix B



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441

Example 5.3

Using the Standard Normal Table to Find $P(-z_0 < z < z_0)$

Problem Find the probability that the standard normal random variable z falls between -1.33 and $+1.33$.

Solution The standard normal distribution is shown again in Figure 5.9. Since all probabilities associated with standard normal random variables can be depicted as areas under the standard normal curve, you should always draw the curve and then equate the desired probability to an area.

In this example, we want to find the probability that z falls between -1.33 and $+1.33$, which is equivalent to the area between -1.33 and $+1.33$, shown highlighted in Figure 5.9.

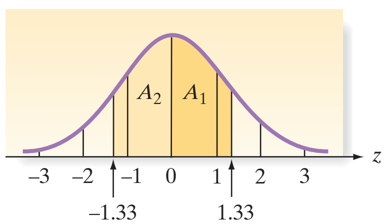


Figure 5.9
Areas under the standard normal curve for Example 5.3

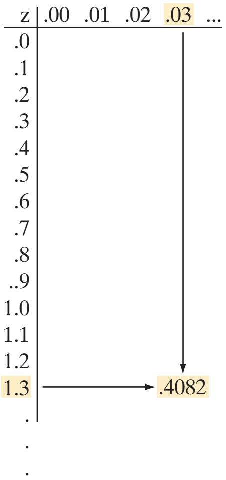


Figure 5.10
Finding $z = 1.33$ in the standard normal table, Example 5.3

Table II gives the area between $z = 0$ and any positive value of z, so that if we look up $z = 1.33$ (the value in the 1.3 row and .03 column, as shown in Figure 5.10), we find that the area between $z = 0$ and $z = 1.33$ is .4082. This is the area labeled A_1 in Figure 5.9. To find the area A_2 located between $z = 0$ and $z = -1.33$, we note that the symmetry of the normal distribution implies that the area between $z = 0$ and any point to the left is equal to the area between $z = 0$ and the point equidistant to the right. Thus, in this example the area between $z = 0$ and $z = -1.33$ is equal to the area between $z = 0$ and $z = +1.33$. That is,

$$A_1 = A_2 = .4082$$

The probability that z falls between -1.33 and $+1.33$ is the sum of the areas of A_1 and A_2. We summarize in probabilistic notation:

$$P(-1.33 < z < -1.33) = P(-1.33 < z < 0) + P(0 < z \leq 1.33)$$
$$= A_1 + A_2 = .4082 + .4082 = .8164$$

Look Back Remember that "<" and "≤" are equivalent in events involving z because the inclusion (or exclusion) of a single point does not alter the probability of an event involving a continuous random variable.

■ **Now Work Exercise 5.25e–f**

Example 5.4

Using the Standard Normal Table to Find $P(z > z_0)$

Problem Find the probability that a standard normal random variable exceeds 1.64; that is, find $P(z > 1.64)$.

Solution The area under the standard normal distribution to the right of 1.64 is the highlighted area labeled A_1 in Figure 5.11. This area represents the probability that z exceeds 1.64. However, when we look up $z = 1.64$ in Table II, we must remember that the probability given in the table corresponds to the area between $z = 0$ and $z = 1.64$ (the area labeled A_2 in Figure 5.11). From Table II, we find that $A_2 = .4495$. To find the area A_1 to the right of 1.64, we make use of two facts:

1. The standard normal distribution is symmetric about its mean, $z = 0$.
2. The total area under the standard normal probability distribution equals 1.

Taken together, these two facts imply that the areas on either side of the mean, $z = 0$, equal .5; thus, the area to the right of $z = 0$ in Figure 5.11 is $A_1 + A_2 = .5$. Then

$$P(z > 1.64) = A_1 = .5 - A_2 = .5 - .4495 = .0505$$

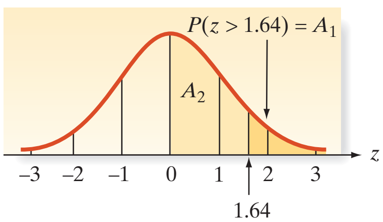


Figure 5.11
Areas under the standard normal curve for Example 5.4

Look Back To attach some practical significance to this probability, note that the implication is that the chance of a standard normal random variable exceeding 1.64 is only about .05.

■ **Now Work Exercise 5.26a**

Example 5.5

Using the Standard Normal Table to Find $P(z < z_0)$

Problem Find the probability that a standard normal random variable lies to the left of .67.

Solution The event sought is shown as the highlighted area in Figure 5.12. We want to find $P(z < .67)$. We divide the highlighted area into two parts: the area A_1 between $z = 0$ and $z = .67$, and the area A_2 to the left of $z = 0$. We must always make such a division when the desired area lies on both sides of the mean $(z = 0)$ because Table II contains areas between $z = 0$ and the point you look up. We look up $z = .67$ in Table II to find that $A_1 = .2486$. The symmetry of the standard normal distribution also implies that half the distribution lies on each side of the mean, so the area A_2 to the left of $z = 0$ is .5. Then

$$P(z < .67) = A_1 + A_2 = .2486 + .5 = .7486$$

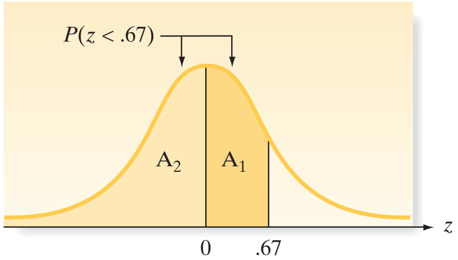


Figure 5.12
Areas under the standard normal curve for Example 5.5

Look Back Note that this probability is approximately .75. Thus, about 75% of the time, the standard normal random variable z will fall below .67. This statement implies that $z = .67$ represents the approximate 75th percentile (or upper quartile) of the standard normal distribution.

■ **Now Work Exercise 5.26h**

Example 5.6

Using the Standard Normal Table to Find $P(|z| > z_0)$

Problem Find the probability that a standard normal random variable exceeds 1.96 in absolute value.

Solution The event sought is shown highlighted in Figure 5.13. We want to find

$$P(|z| > 1.96) = P(z < -1.96 \text{ or } z > 1.96)$$

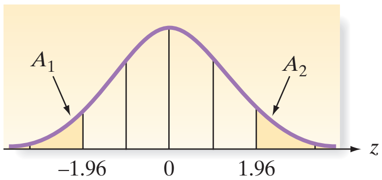


Figure 5.13
Areas under the standard normal curve for Example 5.6

Note that the total highlighted area is the sum of the two areas A_1 and A_2—areas that are equal because of the symmetry of the normal distribution.

We look up $z = 1.96$ and find the area between $z = 0$ and $z = 1.96$ to be .4750. Then A_2, the area to the right of 1.96, is $.5 - .4750 = .0250$, so that

$$P(|z| > 1.96) = A_1 + A_2 = .0250 + .0250 = .05$$

Look Back We emphasize, again, the importance of sketching the standard normal curve in finding normal probabilities. ■

To apply Table II to a normal random variable x with any mean μ and any standard deviation σ, we must first convert the value of x to a z-score. The population z-score for a measurement was defined in Section 2.6 as the *distance* between the measurement and the population mean, divided by the population standard deviation. Thus, the z-score gives the distance between a measurement and the mean in units equal to the standard deviation. In symbolic form, the z-score for the measurement x is

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

Note that when $x = \mu$, we obtain $z = 0$.

An important property of the normal distribution is that if x is normally distributed with any mean and any standard deviation, z is *always* normally distributed with mean 0 and standard deviation 1. That is, z is a standard normal random variable.

Converting a Normal Distribution to a Standard Normal Distribution

If x is a normal random variable with mean μ and standard deviation σ, then the random variable z defined by the formula

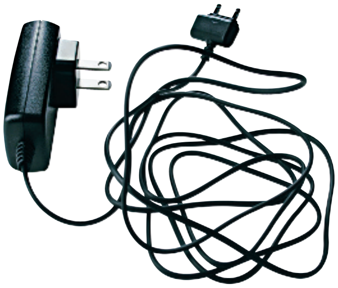
$$z = \frac{x - \mu}{\sigma}$$

has a standard normal distribution. The value z describes the number of standard deviations between x and μ.

Recall from Example 5.6 that $P(|z| > 1.96) = .05$. This probability, coupled with our interpretation of z, implies that any normal random variable lies more than 1.96 standard deviations from its mean only 5% of the time. Compare this statement to the empirical rule (Chapter 2), which tells us that about 5% of the measurements in mound-shaped distributions will lie beyond two standard deviations from the mean. The normal distribution actually provides the model on which the empirical rule is based, along with much "empirical" experience with real data that often approximately obey the rule, whether drawn from a normal distribution or not.

Example 5.7

Finding a Normal Probability—Cell Phone Application



Problem Assume that the length of time, x, between charges of a cellular phone is normally distributed with a mean of 10 hours and a standard deviation of 1.5 hours. Find the probability that the cell phone will last between 8 and 12 hours between charges.

Solution The normal distribution with mean $\mu = 10$ and $\sigma = 1.5$ is shown in Figure 5.14. The desired probability that the cell phone lasts between 8 and 12 hours is highlighted. In order to find that probability, we must first convert the distribution to a standard normal distribution, which we do by calculating the z-score:

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

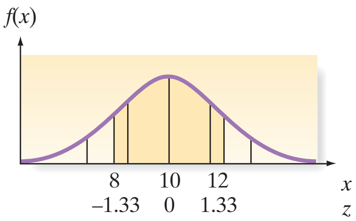


Figure 5.14
Areas under the normal curve for Example 5.7

The z-scores corresponding to the important values of x are shown beneath the x values on the horizontal axis in Figure 5.14. Note that $z = 0$ corresponds to the mean of $\mu = 10$ hours, whereas the x values 8 and 12 yield z-scores of -1.33 and $+1.33$, respectively. Thus, the event that the cell phone lasts between 8 and 12 hours is equivalent to the event that a standard normal random variable lies between -1.33 and $+1.33$. We found this probability in Example 5.3 (see Figure 5.9) by doubling the area corresponding to $z = 1.33$ in Table II. That is,

$$P(8 \le x \le 12) = P(-1.33 \le z \le 1.33) = 2(.4082) = .8164$$

■ **Now Work Exercise 5.31a**

Table II in Appendix B provides good approximations to probabilities under the normal curve. However, if you do not have access to a normal table, you can always rely on statistical software or a graphing calculator to compute the desired probability. With most statistical software, you will need to specify the mean and standard deviation of the normal distribution, as well as the key values of the variable for which you desire probabilities. In Example 5.7, we desire $P(8 \le x \le 12)$, where $\mu = 10$ and $\sigma = 1.5$. To find this probability using MINITAB's normal probability function, we enter 10 for the mean and 1.5 for the standard deviation, and then find two cumulative probabilities: $P(x \le 12)$ and $P(x < 8)$. These two probabilities are shown (shaded) on the MINITAB printout in Figure 5.15. The difference between the two probabilities yields the desired result:

$$P(8 \le x \le 12) = P(x \le 12) - P(x < 8) = .908789 - .0912112 = .8175778$$

Note that this probability agrees with the value computed using Table II to two decimal places. The difference is due to rounding of the probabilities given in Table II.

Cumulative Distribution Function

```
Normal with mean = 10 and standard deviation = 1.5

 x  P( X <= x )
12     0.908789
```

Cumulative Distribution Function

```
Normal with mean = 10 and standard deviation = 1.5

x  P( X <= x )
8   0.0912112
```

Figure 5.15
MINITAB output with cumulative normal probabilities

The steps to follow in calculating a probability corresponding to a normal random variable are shown in the following box:

Steps for Finding a Probability Corresponding to a Normal Random Variable

1. Sketch the normal distribution and indicate the mean of the random variable x. Then shade the area corresponding to the probability you want to find.
2. Convert the boundaries of the shaded area from x values to standard normal random variable z values by using the formula

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

Show the z values under the corresponding x values on your sketch.

3. Use technology or Table II in Appendix B to find the areas corresponding to the z values. If necessary, use the symmetry of the normal distribution to find areas corresponding to negative z values and the fact that the total area on each side of the mean equals .5 to convert the areas from Table II to the probabilities of the event you have shaded.

Example 5.8

Using Normal Probabilities to Make an Inference—Advertised Gas Mileage

Problem Suppose an automobile manufacturer introduces a new model that has an advertised mean in-city mileage of 27 miles per gallon. Although such advertisements seldom report any measure of variability, suppose you write the manufacturer for the details of the tests and you find that the standard deviation is 3 miles per gallon. This information leads you to formulate a probability model for the random variable x, the in-city mileage for this car model. You believe that the probability distribution of x can be approximated by a normal distribution with a mean of 27 and a standard deviation of 3.

a. If you were to buy this model of automobile, what is the probability that you would purchase one that averages less than 20 miles per gallon for in-city driving? In other words, find $P(x < 20)$.

b. Suppose you purchase one of these new models and it does get less than 20 miles per gallon for in-city driving. Should you conclude that your probability model is incorrect?

Solution

a. The probability model proposed for x, the in-city mileage, is shown in Figure 5.16. We are interested in finding the area A to the left of 20, since that area corresponds to the probability that a measurement chosen from this distribution falls below 20. In other words, if this model is correct, the area A represents the fraction of cars that can be