6.1: USING THE STANDARDIZED NORMAL DISTRIBUTION TABLE

Any set of normally distributed data can be converted to its standardized form and the desired probabilities can then be determined from a table of the standardized normal distribution.

To see how the transformation Equation (6.2) on page 199 can be applied and how you can use the results to obtain probabilities from the table of the standardized normal distribution (Table E.14), consider the following problem.

Suppose a consultant was investigating the time it took factory workers in an automobile plant to assemble a particular part after the workers had been trained to perform the task using an individual learning approach. The consultant determined that the time in seconds to assemble the part for workers trained with this method was normally distributed with a mean μ of 75 seconds and a standard deviation σ of 6 seconds.

Transforming the Data

From Figure 6.4a observe that every measurement X has a corresponding standardized measurement Z obtained from the transformation formula (6.2). Hence from Figure 6.4a a time of 81 seconds required for a factory worker to complete the task is equivalent to 1 standardized unit (that is, 1 *standard deviation*) above the mean, since

$$Z = \frac{81 - 75}{6} = +1$$

and a time of 57 seconds required for a worker to assemble the part is equivalent to 3 standardized units (that is, 3 *standard deviations*) below the mean because

$$Z = \frac{57 - 75}{6} = -3$$

Thus, the standard deviation has become the unit of measurement. In other words, a time of 81 seconds is 6 seconds (i.e., 1 standard deviation) higher, or *slower* than the average time of 75 seconds and a time of 57 seconds is 18 seconds (i.e., 3 standard deviations) lower, or *faster* than the average time.

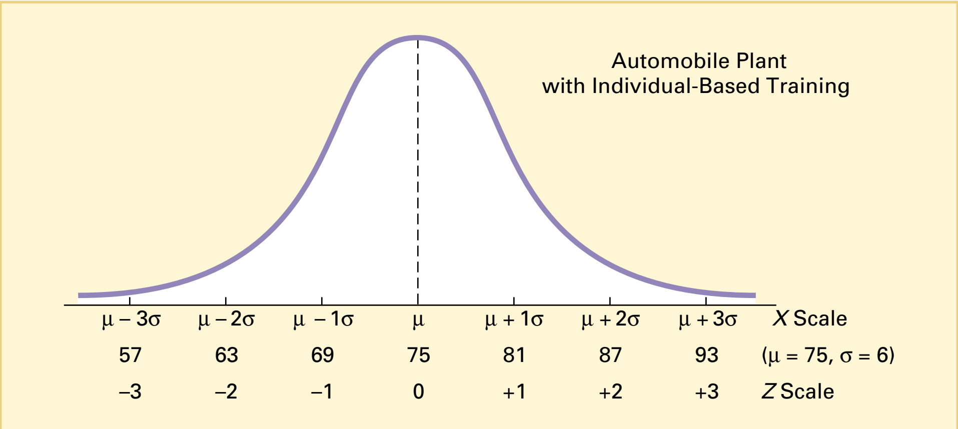

FIGURE 6.4a
Transformation of scales

Suppose now that the consultant conducted the same study at another automobile plant, where the workers were trained to assemble the part by using a team-based learning method. Suppose that at this plant the consultant determined that the time to perform the task was normally distributed with mean μ of 60 seconds and a standard deviation σ of 3 seconds. The data are depicted in Figure 6.5a. In comparison with the results for workers who had an individual learning method, note, for example, that at the plant where workers had team-based training, a time of 57 seconds to complete the task is only 1 standard deviation below the mean for the group, since

$$Z = \frac{57 - 60}{3} = -1$$

Note that a time of 63 seconds is 1 standard deviation above the mean time for assemblage, since

$$Z = \frac{63 - 60}{3} = +1$$

and a time of 51 seconds is 3 standard deviations below the group mean because

$$Z = \frac{51 - 60}{3} = -3$$

Using the Normal Probability Tables

The two bell-shaped curves in Figures 6.4a and 6.5a depict the relative frequency polygons for the normal distributions representing the time (in seconds) for all factory workers to assemble a part at two automobile plants, one that employed individual-based training and the other that employed team-based training. Since at each plant the times to assemble the part are known for every factory worker, the data represent the entire population at a particular plant, and therefore the *probabilities* or proportion of area under the entire curve must add up to 1. Thus, the area under the curve between any two reported times values represents only a portion of the total area possible.

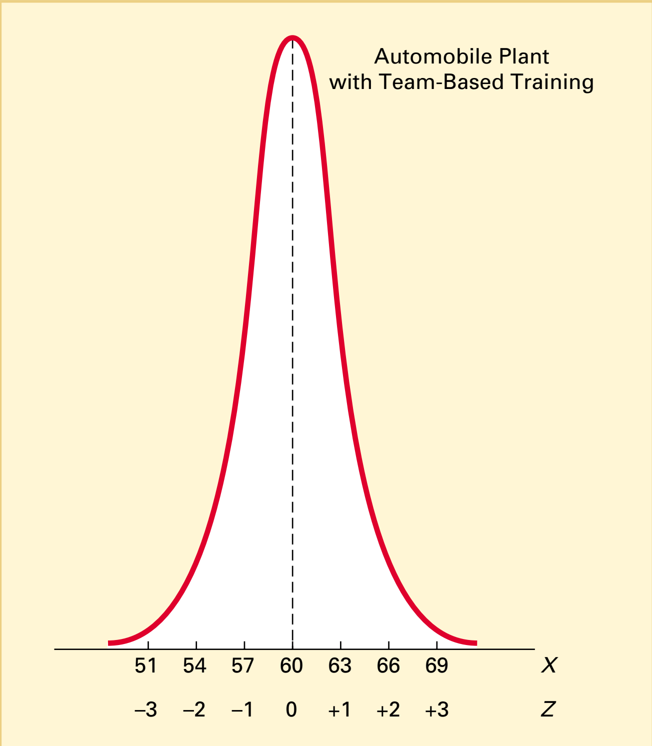

FIGURE 6.5a
A different transformation of scales

Suppose the consultant wishes to determine the probability that a factory worker selected at random from those who underwent individual-based training should require between 75 and 81 seconds to complete the task. That is, what is the likelihood that the worker's time is between the plant mean and one standard deviation above this mean? This answer is found by using Table E.14.

Table E.14 represents the probabilities or areas under the normal curve calculated from the mean μ to the particular values of interest X. Using Equation (6.2), this corresponds to the probabilities or areas under the standardized normal curve from the mean to the transformed values of interest Z. Only positive entries for Z are listed in the table, since for a symmetrical distribution having a mean of zero, the area from the mean to $+Z$ (that is, Z standard deviations above the mean) must be identical to the area from the mean to $-Z$ (that is, Z standard deviations below the mean).

To use Table E.14 note that all Z values must first be recorded to two decimal places. Thus the Z value of interest is recorded as +1.00. To read the probability or area under the curve from the mean to $Z = +1.00$, scan down the Z column from Table E.14 until you locate the Z value of interest (in tenths). Hence you stop in the row $Z = 1.0$. Next read across this row until you intersect the column that contains the hundredths place of the Z value. Therefore, in the body of the table the tabulated probability for $Z = 1.00$ corresponds to the intersection of the row $Z = 1.0$ with the column $Z = .00$ as shown in Table 6.2a (which is a portion of Table E.14). This probability is 0.3413. As depicted in Figure 6.6a, there is a 34.13% chance that a factory worker selected at random who has had individual-based training will require between 75 and 81 seconds to assemble the part.

TABLE 6.2a
Obtaining an area under the normal curve

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2518	.2549
0.7	.2580	.2612	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.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

Source: Extracted from Table E.14.

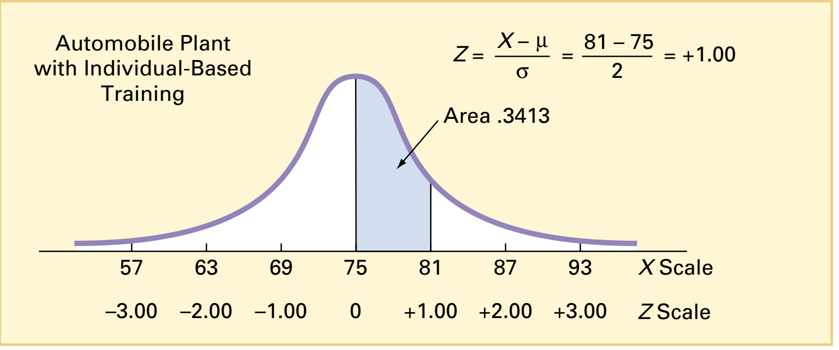

FIGURE 6.6a
Determining the area between the mean and Z from a standardized normal distribution

On the other hand, you know from Figure 6.5a that at the automobile plant where the workers received team-based training, a time of 63 seconds is 1 standardized unit above the mean time of 60 seconds. Thus the likelihood that a randomly selected factory worker who received team-based training will complete the assemblage in between 60 and 63 seconds is also 0.3413. These results are illustrated in Figure 6.7a, which demonstrates that regardless of the value of the mean μ and standard deviation σ of a particular set of normally distributed data, a transformation to a standardized scale can always be made from Equation (6.2), and, by using Table E.14, any probability or portion of area under the curve can be obtained. From Figure 6.7a you see that the probability or area under the curve from 60 to 63 seconds for the workers who had team-based training is identical to the probability or area under the curve from 75 to 81 seconds for the workers who had individual-based training.

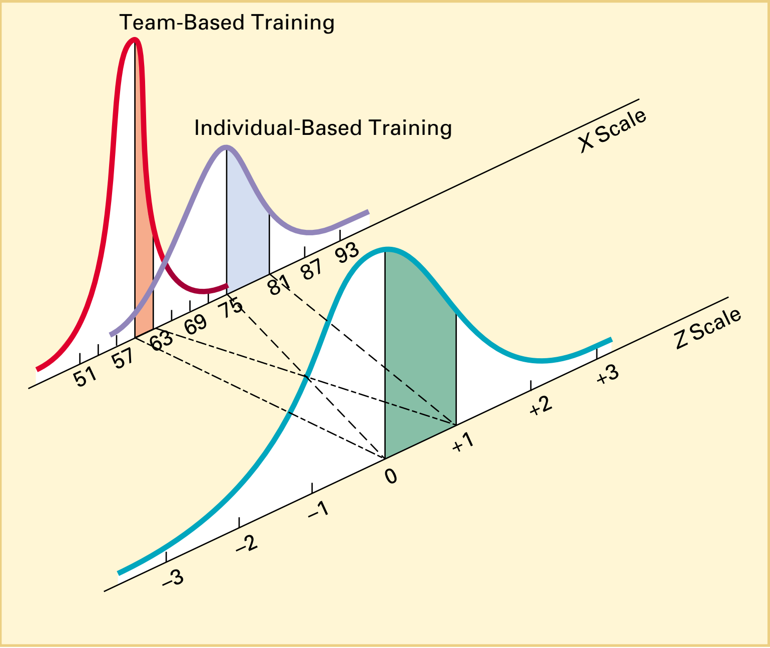

FIGURE 6.7a
Demonstrating a transformation of scales for corresponding portions under two normal curves

Now that you have learned to use Table E.14 in conjunction with Equation (6.2), many different types of probability questions pertaining to the normal distribution can be resolved. To illustrate, suppose that the consultant raises the following questions with regard to assembling a particular part by workers who had individual-based training:

1. What is the probability that a randomly selected factory worker can assemble the part in under 75 seconds or in over 81 seconds?
2. What is the probability that a randomly selected factory worker can assemble the part in 69 to 81 seconds?
3. What is the probability that a randomly selected factory worker can assemble the part in under 62 seconds?
4. What is the probability that a randomly selected factory worker can assemble the part in 62 to 69 seconds?
5. How many seconds must elapse before 50% of the factory workers assemble the part?
6. How many seconds must elapse before 10% of the factory workers assemble the part?
7. What is the interquartile range (in seconds) expected for factory workers to assemble the part?

Finding the Probabilities Corresponding to Known Values

Recall that for workers who had individual-based training the assembly time data are normally distributed with a mean μ of 75 seconds and a standard deviation σ of 6 seconds. In responding to questions 1 through 4, this information is used to determine the probabilities associated with various measured values.

Question 1: Finding $P(X < 75 \text{ or } X > 81)$ How can you determine the probability that a randomly selected factory worker will perform the task in under 75 seconds or over 81 seconds? Since you have already determined the probability that a randomly selected factory worker will need between 75 and 81 seconds to assemble the part, from Figure 6.6a observe that the desired probability must be its *complement*, that is, $1 - 0.3413 = 0.6587$.

Another way to view this problem, however, is to separately obtain both the probability of assembling the part in under 75 seconds and the probability of assembling the part in over 81 seconds and then use the *addition rule for mutually exclusive events* [Equation (4.4)] on page 145 to obtain the desired result. This is depicted in Figure 6.8a.

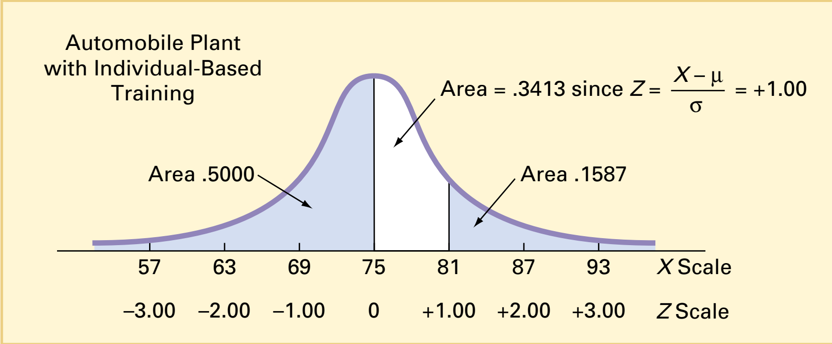

FIGURE 6.8a
Finding $P(X < 75 \text{ or } X > 81)$

Since the mean and median are theoretically the same for normally distributed data, it follows that 50% of the workers can assemble the part in under 75 seconds. To show this, from Equation (6.2)

$$Z = \frac{X-\mu}{\sigma} = \frac{75-75}{6} = 0.00$$

Using Table E.14, you see that the area under the normal curve from the mean to $Z = 0.00$ is 0.0000. Hence the area under the curve less than $Z = 0.00$ must be $0.5000 - 0.0000 = 0.5000$ (which happens to be the area for the entire left side of the distribution from the mean to $Z = -\infty$, as shown in Figure 6.8a).

Now you need to obtain the probability of assembling the part in over 81 seconds. But Equation (6.2) only gives the areas under the curve from the mean to Z, not from Z to $+\infty$. Thus you find the probability from the mean to Z and subtract this result from 0.5000 to obtain the desired answer. Since you know that the area or portion of the curve from the mean to $Z = +1.00$ is 0.3413, the area from $Z = +1.00$ to $Z = +\infty$ must be $0.5000 - 0.3413 = 0.1587$. Hence the probability that a randomly selected factory worker will perform the task in under 75 or over 81 seconds, $P(X < 75 \text{ or } X > 81)$, is $0.5000 + 0.1587 = 0.6587$.

Question 2: Finding $P(69 \le X \le 81)$ Suppose that you are now interested in determining the probability that a randomly selected factory worker can complete the part in 69 to 81 seconds, that is, $P(69 \le X \le 81)$. Note from Figure 6.9a that one of the values of interest is above the mean assembly time of 75 seconds and the other value is below it. Since the transformation Equation (6.2) on page 199 permits you only to find probabilities from a particular value of interest to the mean, you can obtain the desired probability in three steps:

1. Determine the probability from the mean to 81 seconds.
2. Determine the probability from the mean to 69 seconds.
3. Sum up the two mutually exclusive results.

For this example, you already completed step 1; the area under the normal curve from the mean to 81 seconds is 0.3413. To find the area from the mean to 69 seconds (step 2),

$$Z = \frac{X-\mu}{\sigma} = \frac{69-75}{6} = -1.00$$

Table E.14 shows only positive entries for Z. Because of symmetry, the area from the mean to $Z = -1.00$ must be identical to the area from the mean to $Z = +1.00$. Discarding the negative sign, then, you look up (in Table E.14) the value $Z = 1.00$ and find the probability to be 0.3413. Hence, from step 3, the probability that the part can be assembled in between 69 and 81 seconds is $0.3413 + 0.3413 = 0.6826$. This result is displayed in Figure 6.9a.

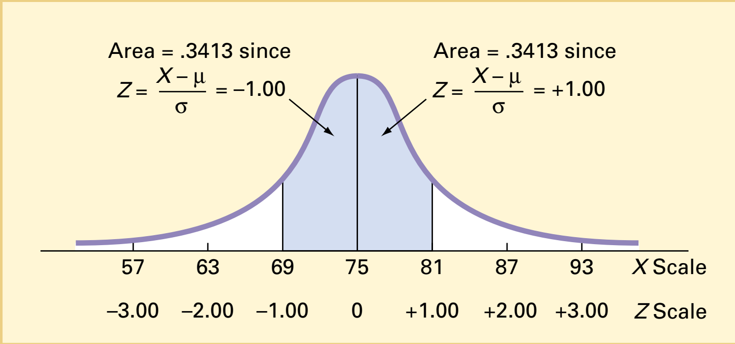

FIGURE 6.9a
Finding $P(69 \le X \le 81)$

The above result is rather important. You can see that for any normal distribution there is a 0.6826 chance that a randomly selected item will fall within ±1 standard deviation above or below the mean.

For the plant in which the workers received individual-based training, slightly more than two out of every three factory workers (68.26%) can be expected to complete the task within ±1 standard deviation from the mean. Moreover, from Figure 6.10a, slightly more than 19 out of every 20 factory workers (95.44%) can be expected to complete the assembly within ±2 standard deviations from the mean (that is, between 63 and 87 seconds), and, from Figure 6.11a, practically all factory workers (99.73%) can be expected to assemble the part within ±3 standard deviations from the mean (that is, between 57 and 93 seconds).

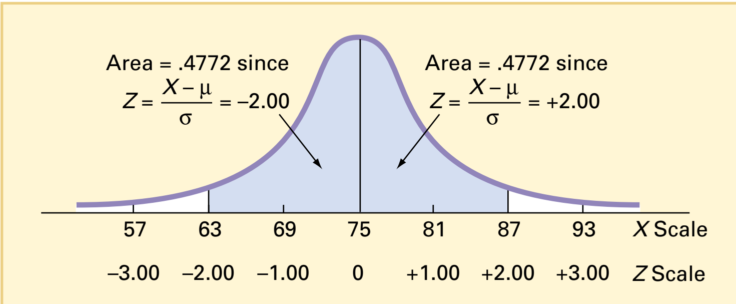

FIGURE 6.10a
Finding $P(63 \le X \le 87)$

From Figure 6.11a it is indeed quite unlikely (.0027 or only 27 factory workers in 10,000) that a randomly selected factory worker will be so fast or so slow that he or she could be expected to complete the assembly of the part in under 57 seconds or over 93 seconds. Thus it is clear why 6σ (that is, 3 standard deviations above the mean to 3 standard deviations below the mean) is often used as a *practical approximation of the range* for normally distributed data.

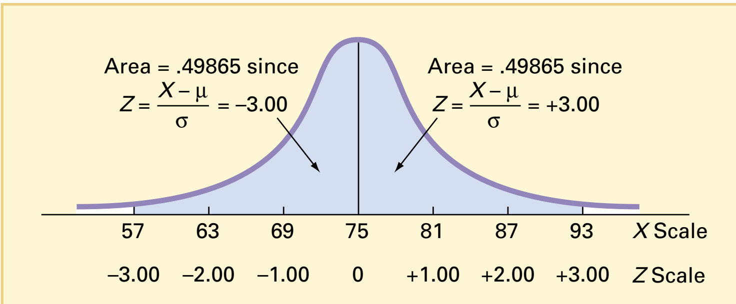

FIGURE 6.11a
Finding $P(57 \le X \le 93)$

Question 3: Finding P(X < 62) To obtain the probability that a randomly selected factory worker can assemble the part in under 62 seconds you examine the shaded lower left-tailed region of Figure 6.12a. The transformation Equation (6.2) on page 199 only permits you to find areas under the standardized normal distribution from the mean to Z, not from Z to $-\infty$. Thus, you must find the probability from the mean to Z and subtract this result from 0.5000 to obtain the desired answer.

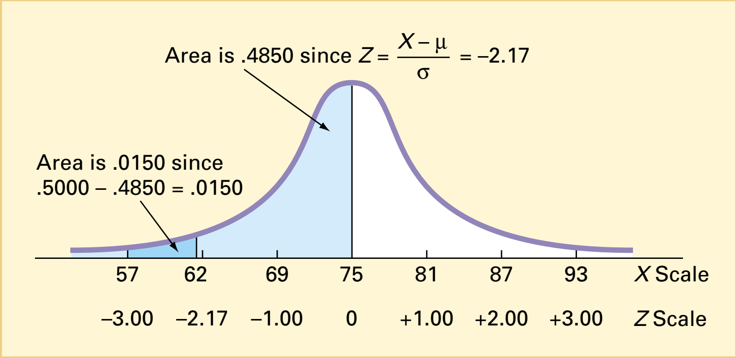

FIGURE 6.12a
Finding $P(X < 62)$

To determine the area under the curve from the mean to 62 seconds,

$$Z = \frac{X - \mu}{\sigma} = \frac{62 - 75}{6} = \frac{-13}{6} = -2.17$$

Neglecting the negative sign, you look up the Z value of 2.17 in Table E.14 by matching the appropriate Z row (2.1) with the appropriate Z column (.07) as shown in Table 6.3a (a portion of Table E.14). Therefore, the resulting probability or area under the curve from the mean to 2.17 standard deviations below it is 0.4850. Hence the area from $Z = -2.17$ to $Z = -\infty$ must be $0.5000 - 0.4850 = 0.0150$. This result is illustrated in Figure 6.12a.

TABLE 6.3a
Obtaining an area under the normal curve

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936

Source: Extracted from Table E.14.

Question 4: Finding P(62 ≤ X ≤ 69) As a final illustration of determining probabilities from the standardized normal distribution, suppose you wish to find how likely it is that a randomly selected factory worker can complete the task in 62 to 69 seconds. Since both values of interest are below the mean, from Figure 6.13a, the desired probability (or area under the curve between the two values) is less than 0.5000. Since the transformation Equation (6.2) on page 199 only permits you to find probabilities from a particular value of interest to the mean, you can obtain the desired probability in three steps:

1. Determine the probability or area under the curve from the mean to 62 seconds.
2. Determine the probability or area under the curve from the mean to 69 seconds.
3. Subtract the smaller area from the larger (to avoid double counting).

For this example, you have already completed steps 1 and 2 in answering questions 3 and 2, respectively. The area from the mean to 62 seconds is 0.4850, and the area from the mean to 69 seconds is 0.3413. Hence, from step 3, by subtracting the smaller area from the larger one you determine that there is only a 0.1437 probability of randomly selecting a factory worker who could be expected to complete the task in between 62 and 69 seconds. That is,

$$\begin{aligned} P(62 \le X \le 69) &= P(62 \le X \le 75) - P(69 \le X \le 75) \\ &= 0.4850 - 0.3413 = 0.1437 \end{aligned}$$

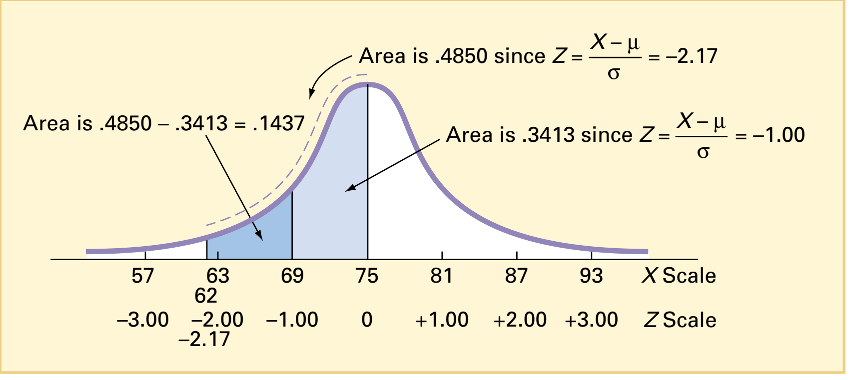

FIGURE 6.13a
Finding $P(62 \le X \le 69)$

Finding the Values Corresponding to Known Probabilities

In previous applications regarding normally distributed data, probabilities associated with various measured values have been determined. Now, however, suppose you wish to determine particular numerical values of the variables of interest that correspond to known probabilities. As examples, see questions 5 through 7.

Question 5 To determine how many seconds elapse before 50% of the factory workers assemble the part, examine Figure 6.14a. Since this time value corresponds to the median, and the mean and median are equal in all symmetric distributions, the median must be 75 seconds.

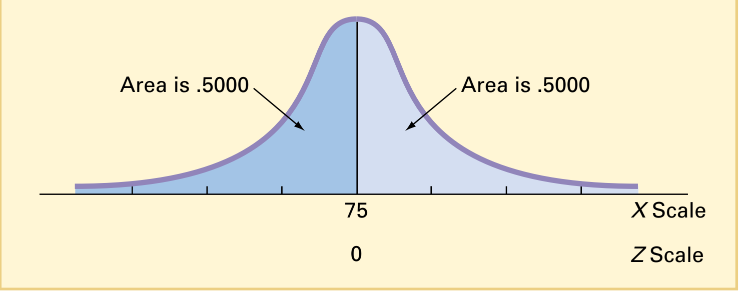

FIGURE 6.14a
Finding X

Question 6 To determine how many seconds elapse before 10% of the factory workers assemble the part, focus on Figure 6.15a. Since 10% of the factory workers are expected to complete the task in under X seconds, then 90% of the workers would be expected to require X seconds or more to do the job. From Figure 6.15a observe that this 90% can be broken down into two parts—times (in seconds) above the mean (that is, 50% of the workers) and times between the mean and the desired value X (that is, 40% of the workers). While you do not know X, you can determine the corresponding standardized value Z, since the area under the normal curve from the standardized mean 0 to this Z must be 0.4000. Using the body of Table E.14, you search for the area or probability 0.4000. The closest result is 0.3997, as shown in Table 6.4a (a portion of Table E.14).

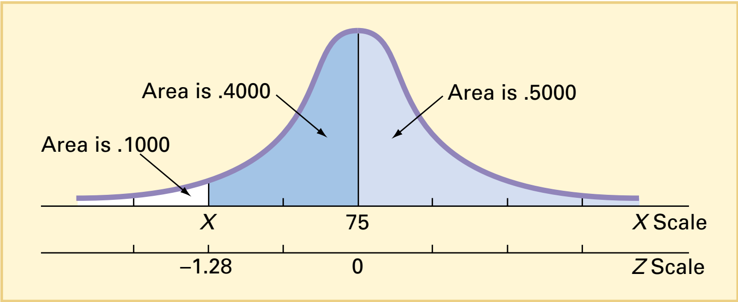

FIGURE 6.15a
Finding Z to determine X

TABLE 6.4a
Obtaining a Z value corresponding to a particular area under the normal curve

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
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

Source: Extracted from Table E.14.

Working from this area to the margins of the table, you see that the Z value corresponding to the particular Z row (1.2) and Z column (.08) is 1.28. However, from Figure 6.15a, the Z value must be recorded as a negative (that is, $Z = -1.28$), since it is below the standardized mean of 0.

Once Z is obtained, you can now use the transformation Equation (6.2) on page 199 to determine the value of interest, X. Since

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

then

$$X = \mu + Z\sigma \tag{6.3}$$

Substituting,

$$X = 75 + (-1.28)(6) = 67.32 \text{ seconds}$$

Thus you expect that 10% of the workers will be able to complete the task in less than 67.32 seconds.

As a review, to find a *particular* value associated with a known probability you must take the following steps:

1. Sketch the normal curve and then place the values for the means on the respective X and Z scales.
2. Split the appropriate half of the normal curve into two parts—the portion from the desired X to the mean and the portion from the desired X to the tail.
3. Shade the area of interest.
4. Using Table E.14, determine the appropriate Z value corresponding to the area under the normal curve from the desired X to the mean μ.
5. Using Equation (6.3), solve for X; that is,

$$X = \mu + Z\sigma$$

Question 7 To obtain the interquartile range you must first find the value for Q_1 and the value for Q_3; then you must subtract the former from the latter.

To find the first quartile value, you must determine the time (in seconds) for which only 25% of the factory workers can be expected to assemble the part faster. This is depicted in Figure 6.16a.

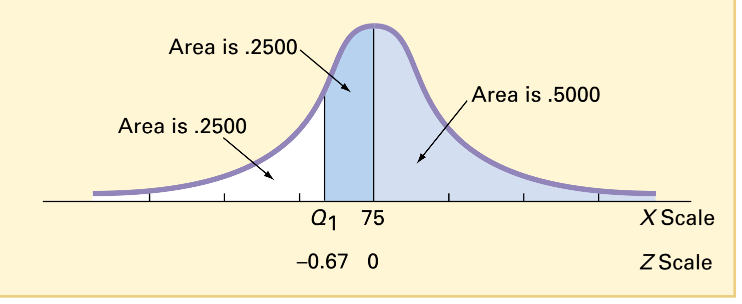

FIGURE 6.16a
Finding Q_1

Although you do not know Q_1, you can obtain the corresponding standardized value Z, since the area under the normal curve from the standardized mean 0 to this Z must be 0.2500. Using the body of Table E.14, you search for the area or probability 0.2500. The closest result is 0.2486, as shown in Table 6.5a (which is a portion of Table E.14).

TABLE 6.5a
Obtaining a Z value corresponding to a particular area under the normal curve

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2518	.2549
0.7	.2580	.2612	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.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

Source: Extracted from Table E.14.

Working from this area to the margins of the table, you see that the Z value corresponding to the particular Z row (0.6) and Z column (.07) is 0.67. However, from Figure 6.16a, the Z value must be recorded as a negative (that is, $Z = -0.67$), since it lies to the left of the standardized mean of 0.

Once Z is obtained, the final step is to use Equation (6.3). Hence,

$$\begin{aligned} Q_1 = X &= \mu + Z\sigma \\ &= 75 + (-0.67)(6) \\ &= 75 - 4 \\ &= 71 \text{ seconds} \end{aligned}$$

To find the third quartile, you must determine the time (in seconds) for which 75% of the factory workers can be expected to assemble the part faster (and 25% could complete the task slower). This is displayed in Figure 6.17a.

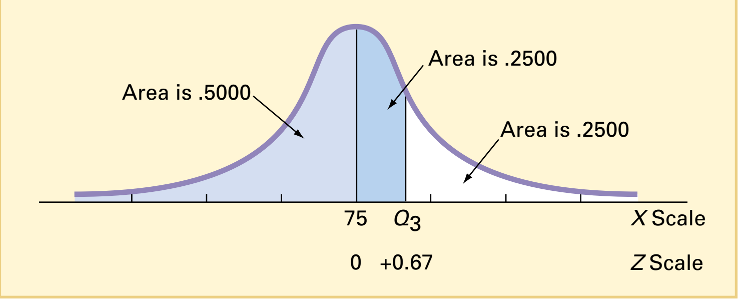

FIGURE 6.17a
Finding Q_3

From the symmetry of the normal distribution, the desired Z value must be +0.67 (since Z lies to the right of the standardized mean of 0). Therefore, using Equation (6.3),

$$\begin{aligned} Q_3 = X &= \mu + Z\sigma \\ &= 75 + (+0.67)(6) \\ &= 75 + 4 \\ &= 79 \text{ seconds} \end{aligned}$$

The interquartile range or middle spread of the distribution is

$$\begin{aligned} \text{interquartile range} &= Q_3 - Q_1 \\ &= 79 - 71 \\ &= 8 \text{ seconds} \end{aligned}$$