

Unit 8: Normal Calculations

SUMMARY OF VIDEO

In this video, we continue the discussion of normal curves that was begun in Unit 7. Recall that a normal curve is bell-shaped and completely characterized by its mean, μ , and standard deviation, σ .

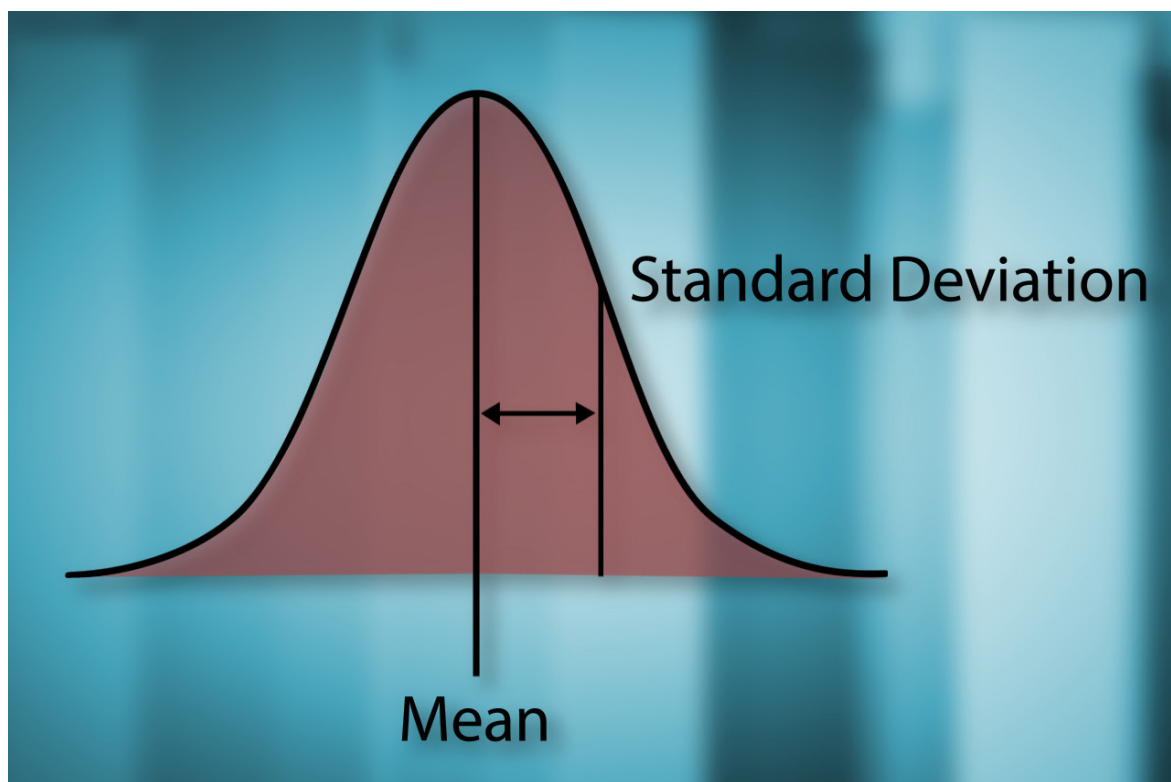


Figure 8.1. Normal curve specified by its mean and standard deviation.

Given the mean and standard deviation of a normal curve, we'd like to approximate the proportion of data that falls within certain intervals. The Empirical Rule or 68-95-99.7% Rule can give us a good starting point. This rule tells us that around 68% of the data will fall within one standard deviation of the mean; around 95% will fall within two standard deviations of the mean; and 99.7% will fall within three standard deviations of the mean. The standard deviation σ is a natural yardstick for any measurements that follow a normal distribution. For example, the distribution of the height of American women can be described by a normal curve with mean $\mu = 63.8$ inches and standard deviation $\sigma = 4.2$ inches. To illustrate how the standard deviation works as a yardstick, begin with a normal curve centered at 63.8. Now, using our

4.2-inch standard deviation yardstick, we measure one standard deviation on either side of the mean, giving 59.6 inches and 68.0 inches. It turns out that about 68% of the total area under the normal curve is over the interval from 59.6 to 68.0 as shown in Figure 8.2. This means that roughly 68% of women's heights are between 59.6 inches and 68.0 inches.

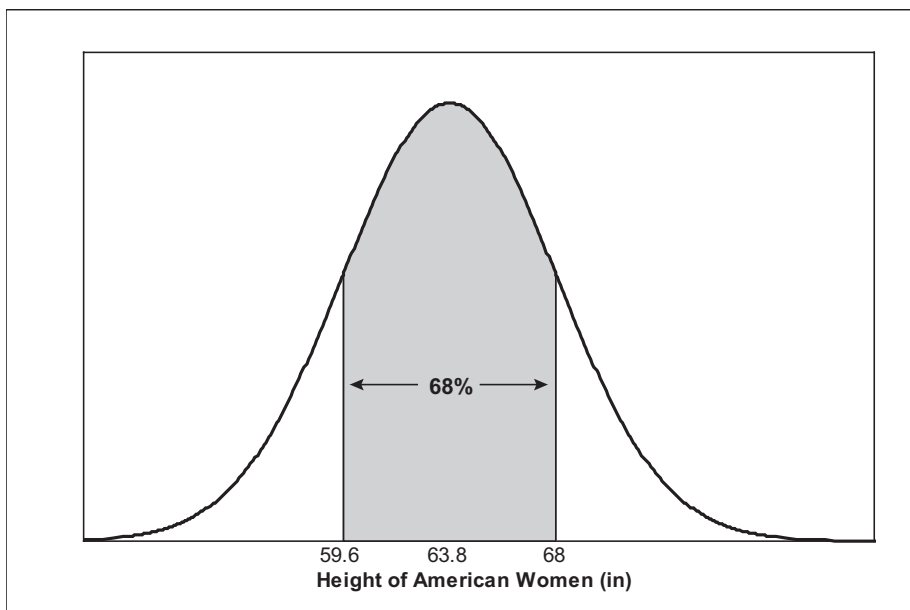


Figure 8.2. Measuring one standard deviation from the mean.

Now, measure out two standard deviations from the mean, from 55.4 inches to 72.2 inches. According to the Empirical Rule, roughly 95% of women's heights will fall within this interval (See Figure 8.3.).

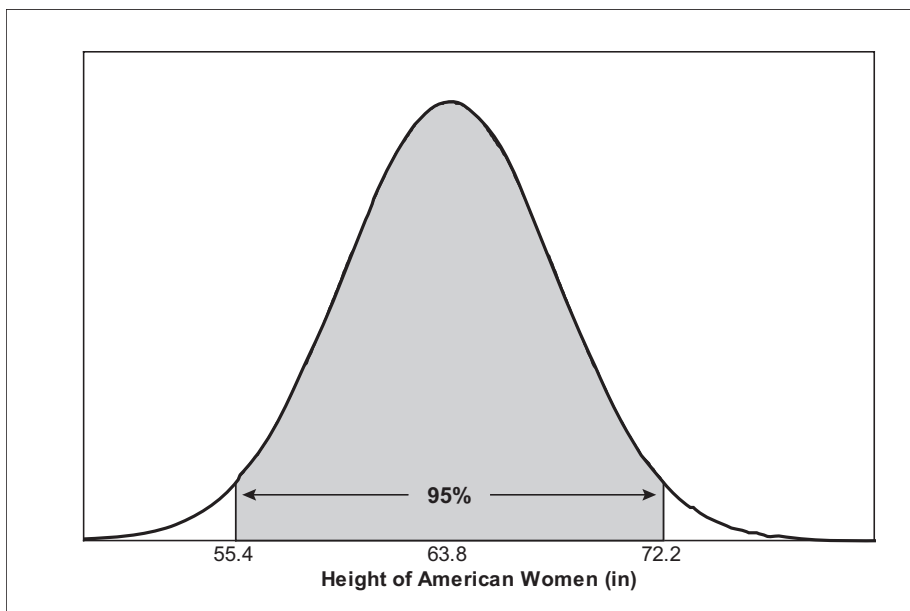


Figure 8.3. Measuring two standard deviations from the mean.

And finally, measure out three standard deviations from the mean, from 51.2 inches to 76.4 inches. About 99.7% of all women's heights will fall within this interval. That leaves only 0.15% of women who are shorter than 51.2 inches and 0.15% of women who are taller than 76.4 inches.

Next, we take a trip to a meeting of the Boston Beanstalks, a social club for tall people — women must be at least 5 feet 10 inches (70 inches) and men at least 6 feet 2 inches (74 inches). To see how far out (in terms of height) the Boston Beanstalks' women are, we convert the entry height of 70 inches into a standard unit called a z-score by subtracting the mean of 63.8 and then dividing the result by the standard deviation of 4.2:

$$z = \frac{70 - 63.8}{4.2} = 1.48$$

This tells us that in order for a woman to be a member of the Boston Beanstalks Club, her height must be at least 1.48 standard deviations above the mean. Heights of men can be described by a normal curve with $\mu = 69.4$ inches and $\sigma = 4.7$ inches. Converting the entry height for males into a z-score gives:

$$z = \frac{74 - 69.4}{4.7} = 0.98$$

Changing normal data into z-scores transforms the data into standard normal data. The standard normal curve has mean $\mu = 0$ and standard deviation $\sigma = 1$. Using the standard normal distribution, Figure 8.4 compares our z-scores for men's and women's entry height into the Boston Beanstalks. In order for men to join, they only need to be about one standard deviation above the men's mean height. The height requirements for women are more stringent than for men. Women need to be a half standard deviation further from the mean than their male counterparts.

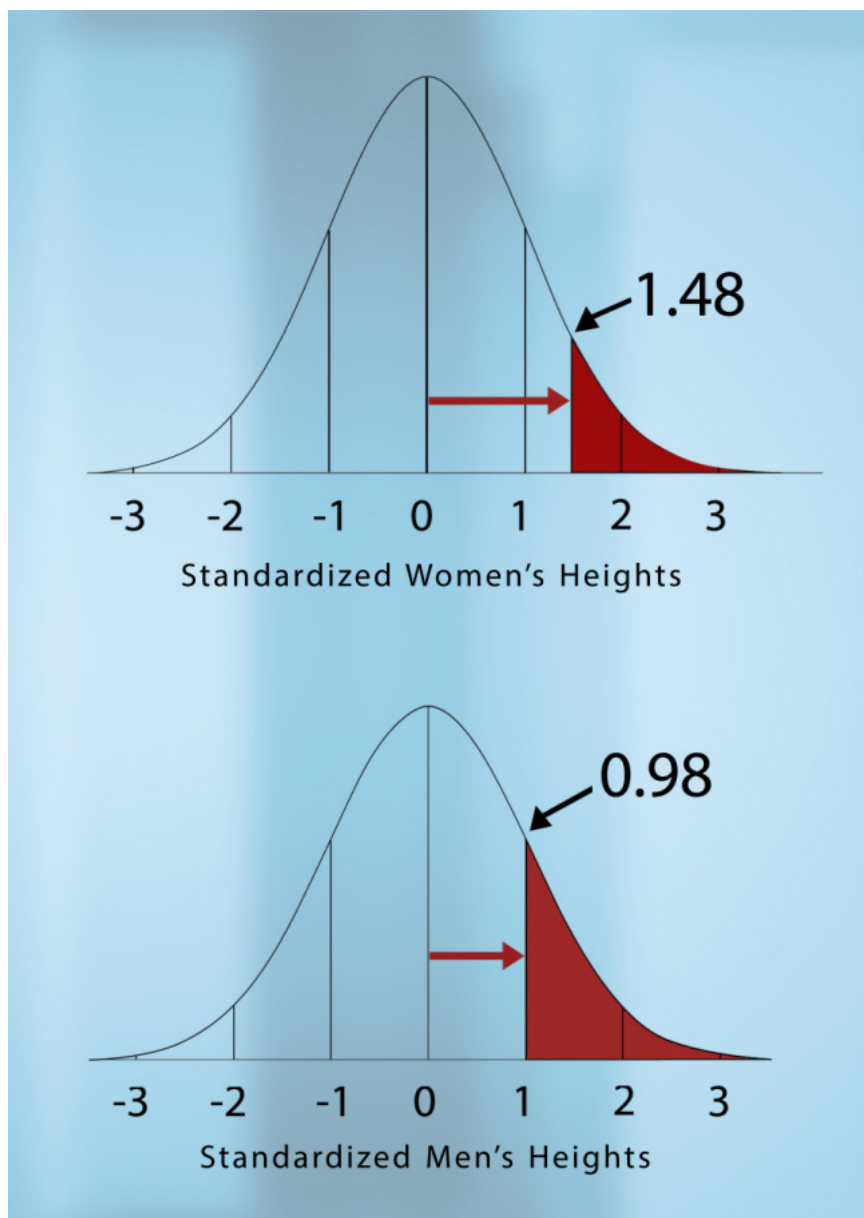
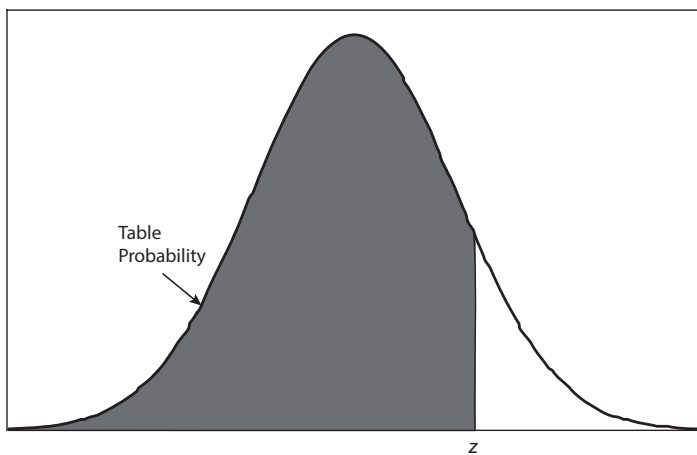


Figure 8.4. Comparing z-scores for entry into the Boston Beanstalks.

Based on the 68-95-99.7% Rule, approximately 16% of men are tall enough to join the club (leaving roughly 84% of the men who are too short to join). But what about women? Again using the 68-95-99.7% Rule, we know that the percent of women tall enough to join is somewhere between 2.5% and 16%. In order to get a more accurate answer, we can turn to z-tables for the standard normal distribution. A z-table tells us how much of the distribution falls below any z-value. Here's how to use the table. For a z-score of 1.48, we follow the z column in Figure 8.5 down to 1.4 and then move right until we are in the .08 column. We can now read off the proportion of women who are too short to join the Beanstalks, 0.9306 or 93.06%. That means that only about 7% of women are tall enough to join the Beanstalks compared to around 16% of men.



z	.00	.01	.02	.03	.04	.05	.06	.07	.08
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7175	.7190
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365
1.0	.8413	.8438	.8461	.8485	.8505	.8531	.8554	.8577	.8599
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306

Figure 8.5. Using the standard normal table for a z-score of 1.48.

We can also use z-scores to compute the percent of data that falls in an interval between two values. For example, suppose we want to know the proportion of American women who are taller than our host Pardis Sabeti, who is 64.5 inches tall, but still not tall enough to make it into the Boston Beanstalks. The z-score for a height of 64.5 inches is 0.16. Recall the z-score for women's entry into the Boston Beanstalks is 1.48. So, we need the area under the standard normal curve between 0.16 and 1.48, which is shown in Figure 8.6. We know the proportion of women shorter than the cutoff for the Beanstalks is 0.9306. Using the table in Figure 8.5, we find that the proportion of women shorter than Dr. Sabeti is 0.5636. We get the desired proportion by subtracting these two proportions:

$$0.9306 - 0.5636 = 0.3670$$

So, approximately 36.7% of American women are taller than Pardis Sabeti but too short to be accepted by the Boston Beanstalks.

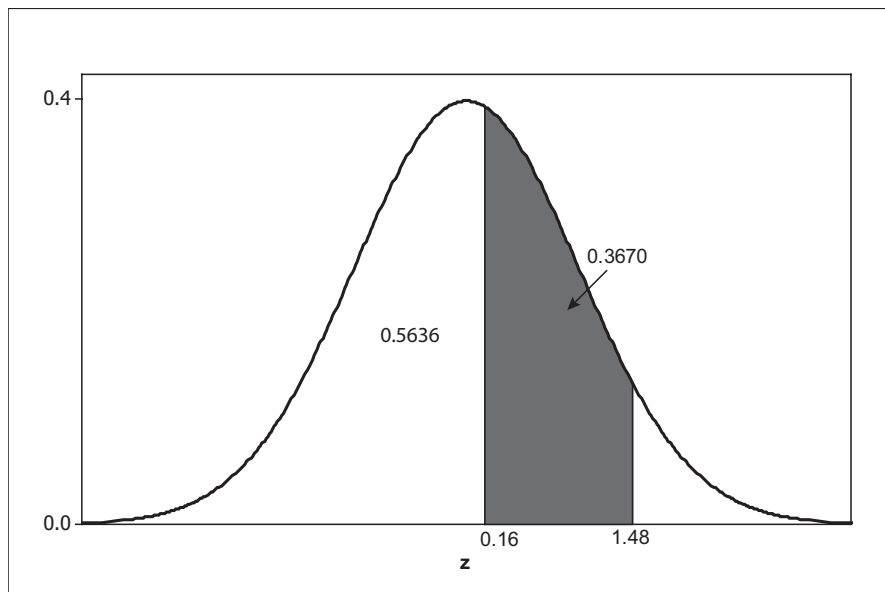


Figure 8.6. Computing the proportion between z-scores of 0.16 and 1.48.

STUDENT LEARNING OBJECTIVES

- A. Be able to use the Empirical Rule (68-95-99.7% Rule) to approximate the proportions of normal data falling in certain intervals.
- B. Understand that standardizing (by subtracting the mean and dividing by the standard deviation) allows us to compare observations from different normal distributions.
- C. Know that in order to use a standard normal table to do calculations involving normal distributions, we must first standardize measurements.
- D. Be able to use the standard normal table (z-table) to find the proportion of observations below any value of z .
- E. Be able to combine standardization with use of the z-table to find the proportion above, below, or between given values in any normal distribution.
- F. Be able to use software such as Excel, Minitab, or SPSS, or graphing calculators to perform the calculations in D and E without having to convert to z-scores.

CONTENT OVERVIEW

Normal density curves are bell-shaped curves, which have been scaled so that the area under the curve is 1. A normal density curve is completely determined by its mean, μ , and standard deviation σ . Figure 8.7 shows two normal density curves, the solid curve has mean $\mu = 5$ and standard deviation $\sigma = 3$ and the dashed curve has mean $\mu = 8$ and standard deviation $\sigma = 2$. Vertical line segments mark the means of the two normal distributions and horizontal line segments mark the standard deviations.

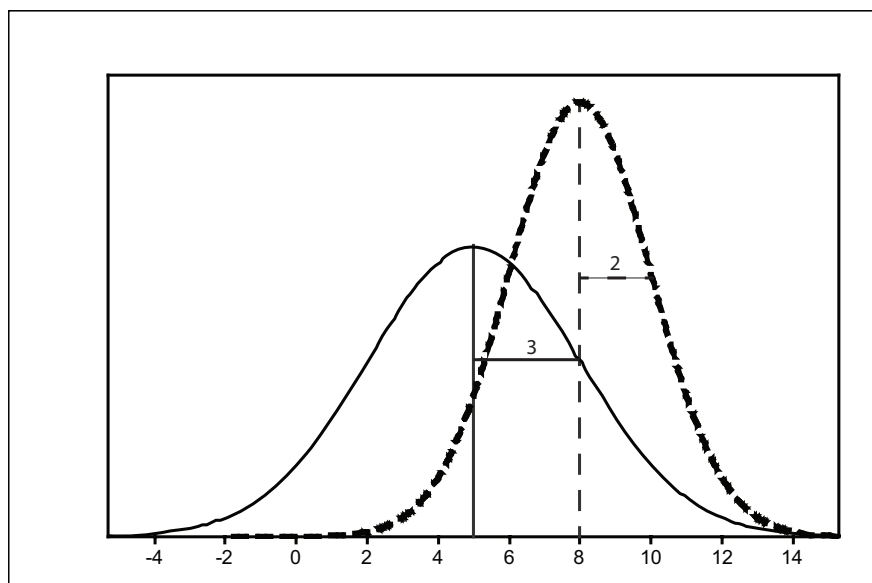


Figure 8.7. Comparing two normal density curves.

The standard deviation acts as a natural yardstick for any measurement that follows a normal distribution. We can summarize use of the standard deviation as a yardstick with the **Empirical Rule**, also known as the **68-95-99.7% Rule**.

Empirical Rule or 68-95-99.7% Rule

In any normal distribution with mean μ and standard deviation σ :

- Approximately 68% of the data fall within one standard deviation of the mean.
- Approximately 95% of the data fall within two standard deviations of the mean.
- Approximately 99.7% of the data fall within three standard deviations of the mean.

Figure 8.8 below shows the percentage of normal data falling within one, two, and three standard deviations from the mean.

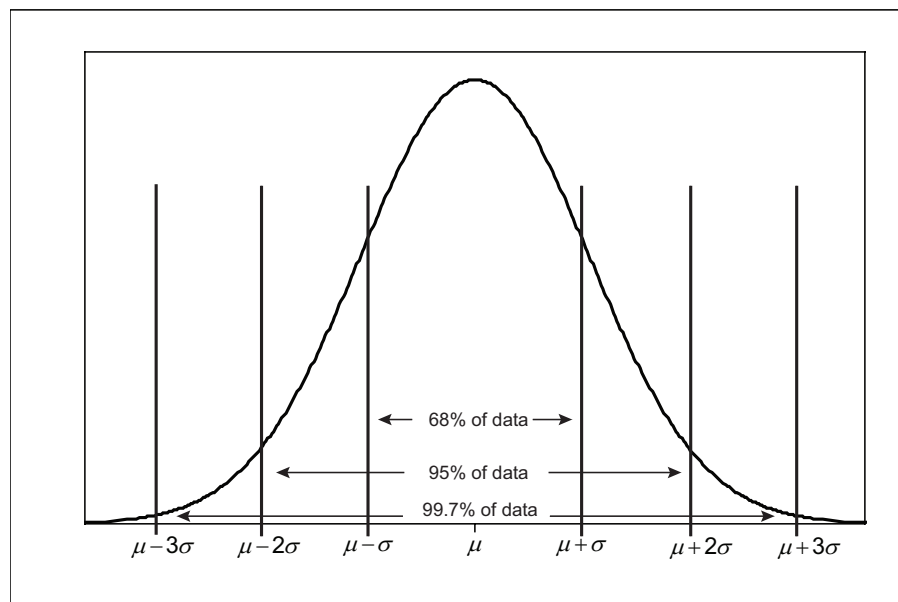


Figure 8.8. Graphic of 68-95-99.7% Rule.

We can apply the Empirical Rule to the normal distribution shown by the solid curve in Figure 8.7. Since this curve has mean 5 and standard deviation 3, the Empirical Rule tells us that roughly 68% of data from this distribution will fall between 2 and 8, roughly 95% of the data will fall between -1 and 11 and nearly all (99.7%) of the data will fall between -4 and 14. If, instead, we focus on the dashed curve in Figure 8.7, we get the same percentages if we form intervals one, two, and three standard deviations on either side of the mean.

The Empirical Rule suggests that the difference between two normal curves is only a matter of scaling. All normal distributions are the same when we measure how many standard deviations an observation x lies away from the mean, which we calculate as follows:

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

The **standardized value** for x , calculated by the formula above, is called its **z-score**. Observations from different normal distributions are best compared by comparing their standardized values, or z-scores. The z-score states how many standard deviations the original observation falls away from the mean and in which direction. Observations larger than the mean have positive z-scores, while observations smaller than the mean have negative z-scores.

Now, return to the dashed normal curve in Figure 8.7. When we transform the numbers on the horizontal scale into a z-score scale, we get the standard normal curve, with mean $\mu = 0$ and standard deviation $\sigma = 1$, as shown in Figure 8.9. If we switch to the solid curve in Figure 8.7 and transform its horizontal scale into a z-score scale (subtract 5 and divide the result by 3), we would also wind up with the standard normal curve. So, any two normal curves can be changed into the standard normal curve by transforming the scale on the horizontal axis into a z-score scale.

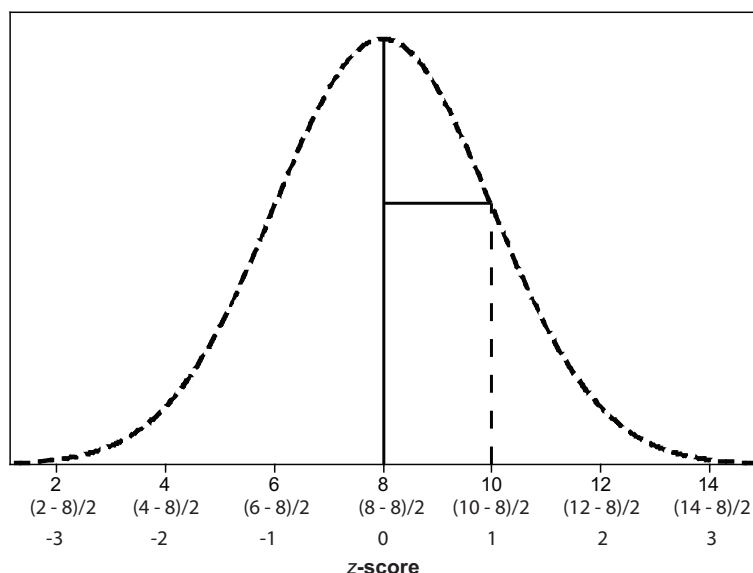


Figure 8.9. Transforming the horizontal scale to a z-score scale.

Converting to standardized values allows us to find proportions that we can't get from the Empirical Rule. For example, suppose we want to know the percentage of data from a normal distribution with mean $\mu = 8$ and standard deviation $\sigma = 2$ (Figure 8.7, original scale) that falls below $x = 9$. We can use the Empirical Rule to learn that the percentage is between 50% and 84%, but that is not a very accurate estimate. Instead, we convert $x = 9$ into a z-score:

$$z = \frac{9 - 8}{2} = 0.5$$

Now, we find the proportion of standard normal data that falls below 0.5. To find this proportion, we use a standard normal table, which gives the proportion of data that falls below any value for z . Using the portion of a z-table shown Figure 8.5, we look down the z -column to locate 0.5 and then move to the right under the .00 column. Our answer is 0.6915, or around 69.15%.

We can use the same technique to find the percentage of data that fall between two values such as $x = 9$ and $x = 13$. This proportion is represented by the shaded area in Figure 8.10. First, we need to convert $x = 13$ into a z-score:

$$z = \frac{13 - 8}{2} = 2.5$$

To find the proportion of standard normal data that falls below $z = 2.5$, we consult a z-table, which gives 0.9938. We already have determined that the proportion of standard normal data that falls below $z = 0.5$ is 0.6915. All that is needed to find the shaded gray area in Figure 8.10 is to subtract these two proportions:

$$\text{proportion} = 0.9938 - 0.6915 = 0.3023$$

So, we find that just over 30% of observations from a normal distribution with $\mu = 8$ and $\sigma = 2$ will fall between $x = 9$ and $x = 13$.

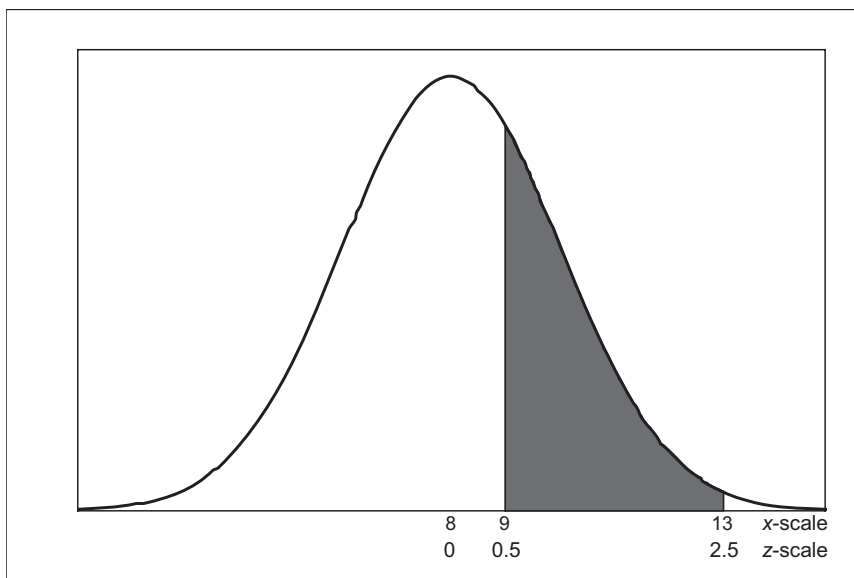


Figure 8.10. Finding the area under the curve from $x = 9$ to $x = 13$.

KEY TERMS

The Empirical Rule or **68-95-99.7% Rule** gives the approximate percentage of data that fall within one standard deviation (68%), two standard deviations (95%), and three standard deviations (99.7%) of the mean. This rule should be applied only when the data are approximately normal.

An observation x from a normal distribution with mean μ and standard deviation σ can be transformed into a **standardized value** called a **z-score** as follows:

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

A **standard normal distribution** is a normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$.

THE VIDEO:

NORMAL CURVES

Take out a piece of paper and be ready to write down answers to these questions as you watch the video.

1. What is another name for the Empirical rule?
2. How tall must a woman be to join the Boston Beanstalks Club?
3. How do you calculate a z-score?
4. Based on z-scores, are eligibility requirements to join the Boston Beanstalks more difficult to meet for men or for women?

UNIT ACTIVITY:

USING AREA TO ESTIMATE STANDARD NORMAL PROPORTIONS

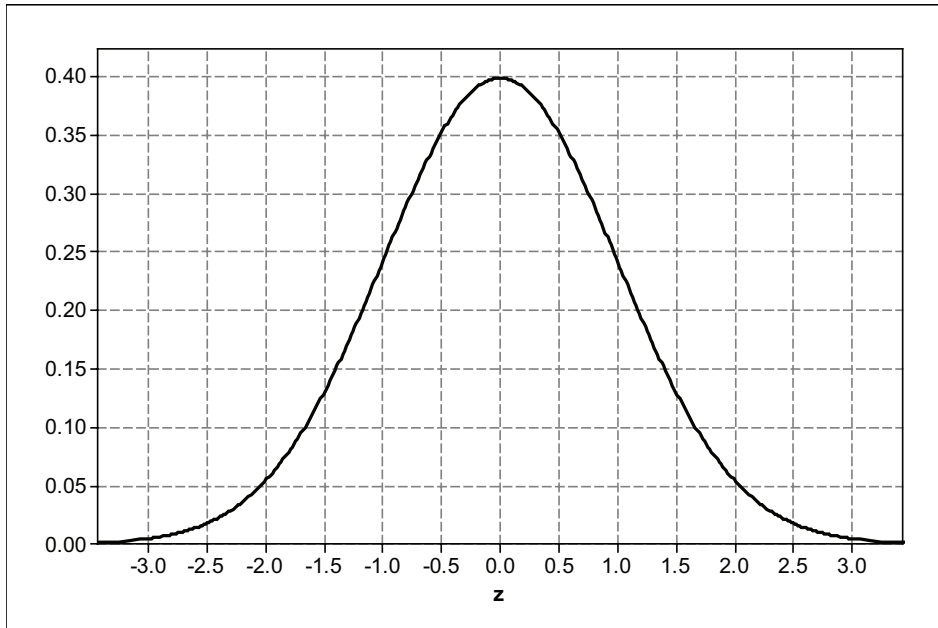
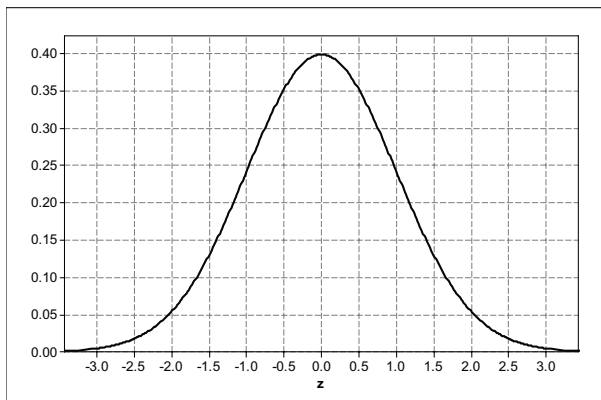


Figure 8.11. The standard normal density curve.

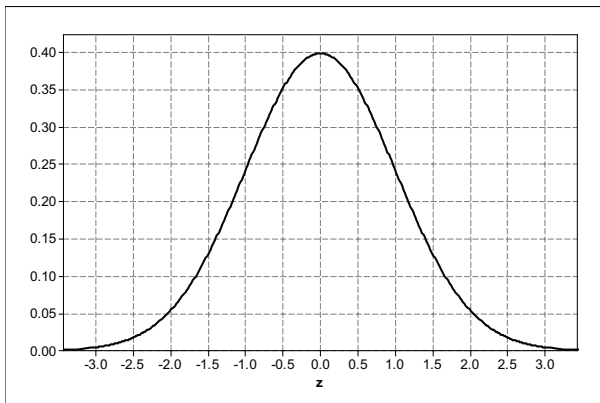
1. Estimate the number of rectangles that fit under the standard normal density curve in Figure 8.11. The area under any density curve is 1, which represents a proportion of all the data or 100% of the data.
2. Using the rectangles, you will estimate the proportion of data from a standard normal distribution that falls below 0. (You should already know the answer. So, this question will help you check the process.)
 - a. Shade the area under the curve that lies to the left of $z = 0$ on the normal curve below. Count the number of rectangles contained in the shaded region.



b. To find the proportion of data that falls below $z = 0$, divide the number of rectangles you got in (a) by the total number of rectangles counted for question 1. What is your estimate for this proportion?

3. Next, using the rectangles, you will estimate the proportion of data that falls below $z = -1$.

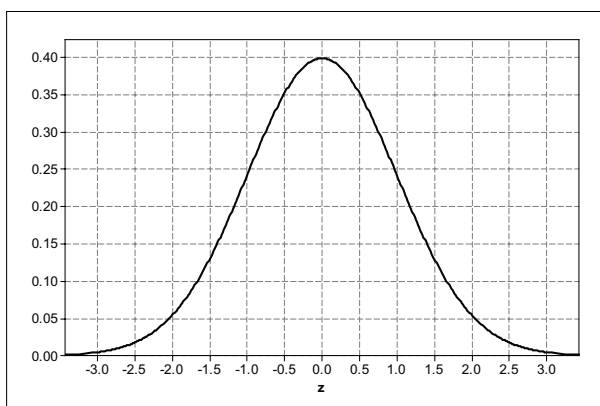
a. Shade the area under the curve that lies to the left of $z = -1$ on the normal curve below. Count the number of rectangles contained in the shaded region.



b. To find the proportion of data that falls below $z = -1$, divide the number of rectangles you got in (a) by the total number of rectangles from question 1. What is your estimate for this proportion?

4. Repeat question 3 using $z = -2$. Show your calculations.

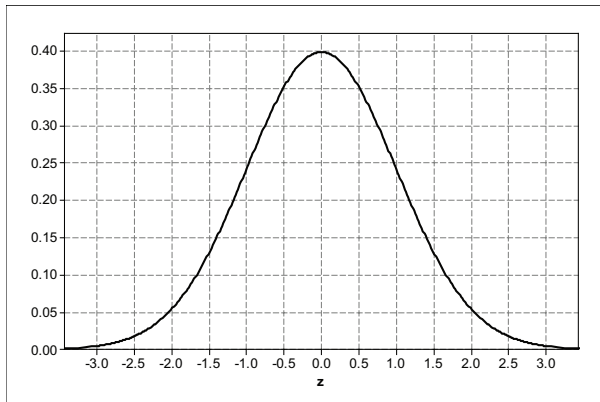
a.



b. Proportion = _____

5. Repeat question 3 using $z = 1$.

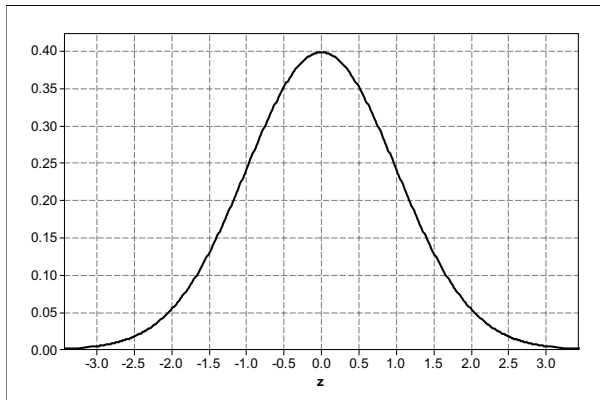
a.



b. Proportion = _____

6. Repeat question 3 with $z = 2$.

a.



b. Proportion = _____

7. Now it is time to compare your estimates with the proportions given in the standard normal table (z-table). Given any value of z , the standard normal table will tell you the proportion of data that falls at or below the given value.

a. Use a standard normal table to complete the entries in column two of Table 8.1.

b. Enter your estimates from questions 2 – 6 into column three of Table 8.1.

z	Proportion from z-table	Estimated Proportion
-2.00		
-1.00		
0.00		
1.00		
2.00		

Table 8.1. Comparisons of proportions from a z-table and estimates.

c. How close were your estimates of the area under the standard normal curve to the actual proportions from the z-table?

EXERCISES

1. The distribution of heights of young women (18 – 24 years old) is approximately normal with mean $\mu = 65.5$ inches and standard deviation $\sigma = 2.5$ inches.

a. Between what two values do the heights of the central 95% of young women lie? Draw a normal curve and shade this area under the normal curve.

b. Based on the 68-95-99.7% Rule, explain why only about 2.5% of young women are more than two standard deviations taller than the mean.

c. What is the z-score for a 20-year-old woman who is 6 feet tall? Interpret this value.

2. There are two national college-entrance examinations, the SAT and the American College Testing program (ACT). Scores on individual SAT exams are approximately normal with mean 500 and standard deviation 100. Scores on the ACT exams are approximately normal with mean 18 and standard deviation 6. (These are the distributions for a large reference population of students used to set standards that remain the same from year to year.)

a. What percent of all SAT scores are above 600? Explain how you arrived at your answer.

b. Julie's SAT Math score is 630. John's ACT Math score is 22. Find the standardized scores for both Julie and John. Assuming that both tests measure the same kind of ability, who has the higher score?

3. Do you think that the distribution of the selling price of houses in a city is close to normal? Why or why not? Would you be willing to apply the 68-95-99.7% Rule to house prices? Explain.

4. Find the percentage of observations from a standard normal distribution that satisfy the following.

a. $z < -1$

b. $z \geq 2.25$

c. $-1 < z < 2.25$

REVIEW QUESTIONS

1. A study investigated the effect of car speed on accident severity. As part of the study, the vehicle speed at impact was extracted from a sample of 6000 accident reports of fatal automobile accidents. Analysis revealed that vehicle speed at impact could be described by a normal distribution with mean $\mu = 44$ mph and standard deviation $\sigma = 14$ mph.

a. Draw a normal curve that represents the speed at impact for fatal accidents. Be sure to add appropriate scaling to the horizontal axis. (For example, use the information on mean and standard deviation to mark the center of the curve and indicate its spread.)

Use the Empirical Rule to answer parts (b) – (d).

b. Approximately what proportion of vehicle speeds were between 30 mph and 58 mph?

c. Approximately what proportion of vehicle speeds were less than 30 mph?

d. Approximately what proportion of vehicle speeds exceeded 72 mph?

2. A teacher gives two different statistics tests, but one is harder than the other. Scores on test A have mean 78 and standard deviation 6, and scores on test B have mean 65 and standard deviation 9. Carrie scored 79 on test B and Pat scored 85 on test A. Who had the higher standardized score?

3. The Army finds that the head sizes (forehead circumference) of soldiers vary according to the Normal distribution with mean $\mu = 22.8$ inches and standard deviation $\sigma = 1.1$ inches.

a. What proportion of soldiers have head size at least 21 inches? Explain how to use a standard normal table to answer this question. Then verify your answer using technology.

b. What proportion of soldiers have head size between 21 inches and 23 inches. Explain how to use a standard normal table to answer this question. Then verify your answer using technology.

4. Assume that in recent years the arrival time (in days since the spring equinox) for Blackpoll Warblers at Manomet Center for Conservation Sciences follows a normal distribution with mean $\mu = 67$ and standard deviation $\sigma = 5.5$.

a. Sketch a normal curve that represents the arrival time of Blackpoll Warblers passing

through Manomet. Add an appropriate scale to the horizontal axis that uses the mean and standard deviation.

- b. What percentage of Blackpoll Warblers passing through Manomet Center arrived before day 60?
- c. What percentage of Blackpoll Warblers arrived after day 70?
- d. What percentage of Blackpoll Warblers arrived between day 60 and day 70?