Chapter 13

The Normal Distribution

Lecture Slides

Case Study: Normal Distributions 1

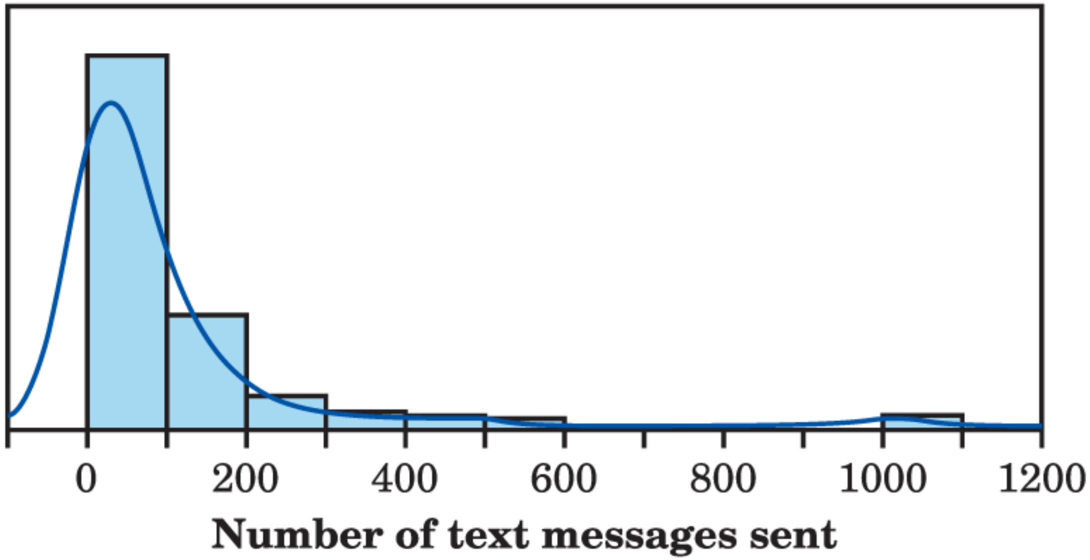
Using bars to display data goes back to William Playfair (1759–1823), an English economist who was an early pioneer of data graphics.

Histograms require that we choose classes, and their appearance can change with different choices.

Surely, modern software offers a better way to picture distributions?

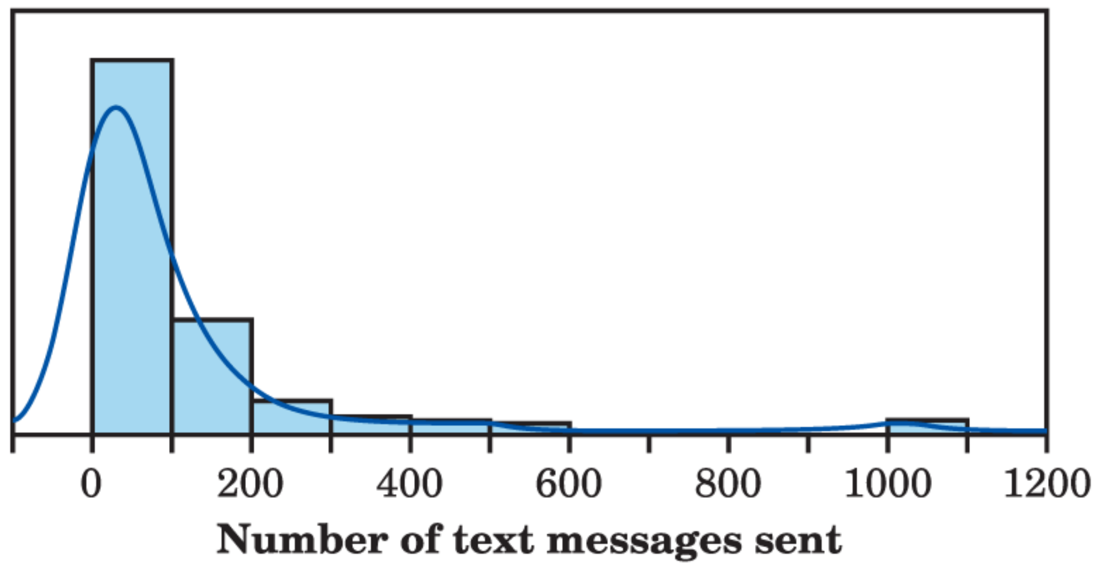
Software can replace the separate bars of a histogram with a smooth curve that represents the overall shape of a distribution.

Case Study: Normal Distributions 2



The data pictured here are the number of text messages reported as being sent on a particular day in 2017 by a random sample of 447 high school seniors. The curve is generated from statistical software as a way of replacing the histogram.

Case Study: Normal Distributions 3



Moore/Notz, *Statistics: Concepts and Controversies*, 10e,

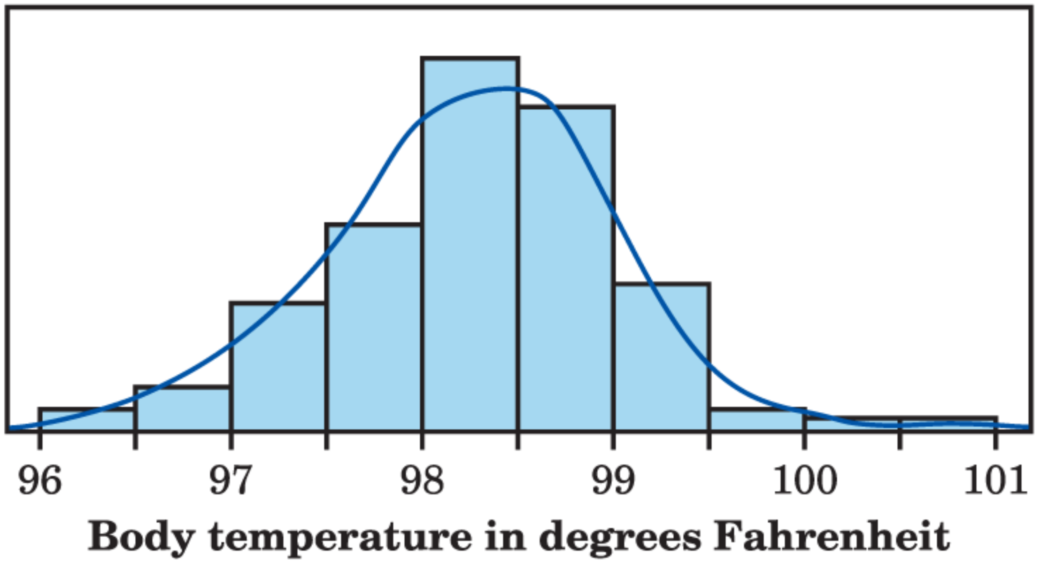
The software doesn’t start from the histogram. It starts with the actual observations and cleverly draws a curve to describe their distribution.

Case Study: Normal Distributions 4

In Figure 13.1, the software has caught the overall shape and shows the ripples in the long right tail more effectively than does the histogram.

It struggles a bit with the peak: it has extended the curve beyond zero in an attempt to smooth out the sharp peak.

Case Study: Normal Distributions 5

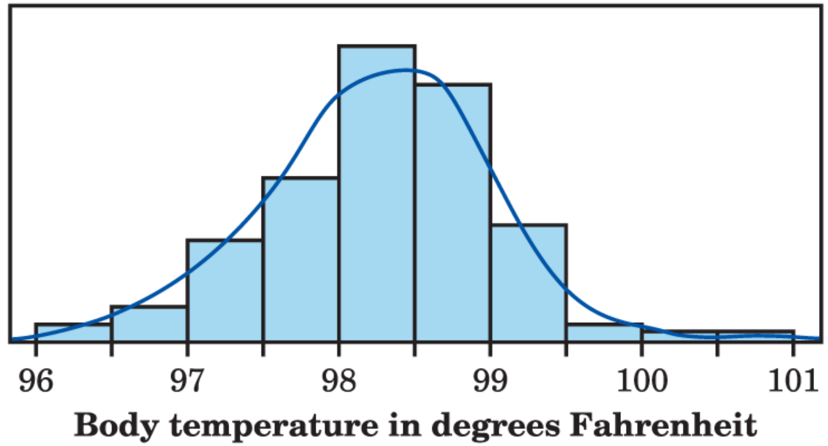


Moore/Notz, *Statistics: Concepts and Controversies*, 10e,

In Figure 13.2, we apply the same software to a set of data with a more regularly shaped distribution. These are the body temperatures from a sample of 130 healthy adults. The software draws a curve that shows a distinctive symmetrical, single-peaked, bell shape.

Case Study: Normal Distributions 6

Figure 13.3 shows the Normal curve for these data. The curve looks a lot like the one in Figure 13.2, but a close look shows that it is smoother. The Normal curve is much easier to work with and does not require clever software.



Case Study: Normal Distributions 7

In this chapter, we will learn that Normal curves have special properties that help us use them and think about them.

By the end of this chapter, you will be able to use these properties to answer questions about the underlying distributions represented in Figures 13.2 and 13.3 that cannot easily be determined from the histograms.

Density Curves 1

1. Always plot your data: make a graph, usually a histogram or a stemplot.
2. Look for the overall pattern (shape, center, variability) and for striking deviations such as outliers. Choose either the five-number summary or the mean and standard deviation to briefly describe center and variability in numbers.
3. Sometimes, the overall pattern of a large number of observations is so regular that we can describe it by a smooth curve.

Density Curves 2

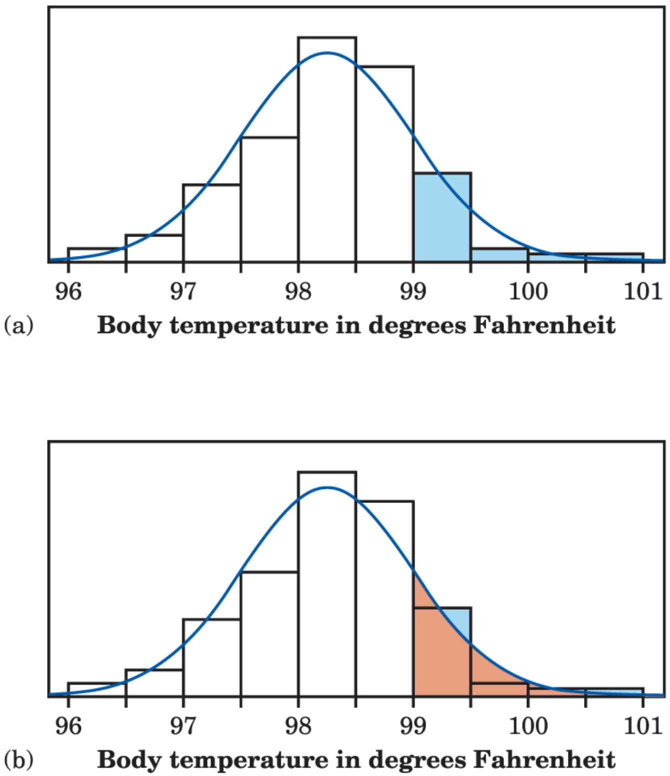
You can think of drawing a curve through the tops of the bars in a histogram, smoothing out the irregular ups and downs of the bars.

There are two important distinctions between histograms and these curves.

1. We set up curves to show the proportion of observations in any region by areas under the curve. To do that, we choose the scale so that the total area under the curve is exactly 1.

2. A histogram is a plot of data obtained from a sample. A density curve is intended to reflect the idealized shape of the population distribution.

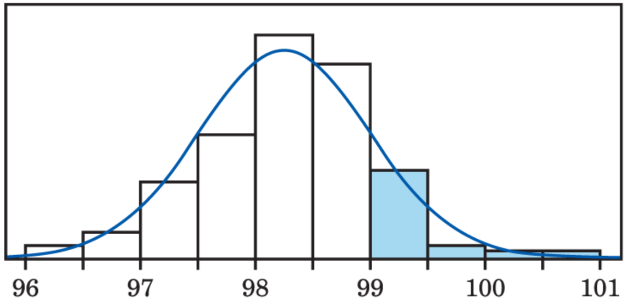
Example: Using a density curve 1

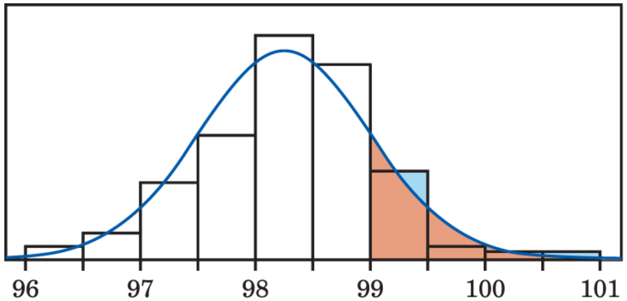


Moore/Notz, *Statistics: Concepts and Controversies*, 10e,

Figure 13.4 copies Figure 13.3, showing the histogram and the Normal density curve that describe this data set of 130 body temperatures.

Example: Using a density curve 2

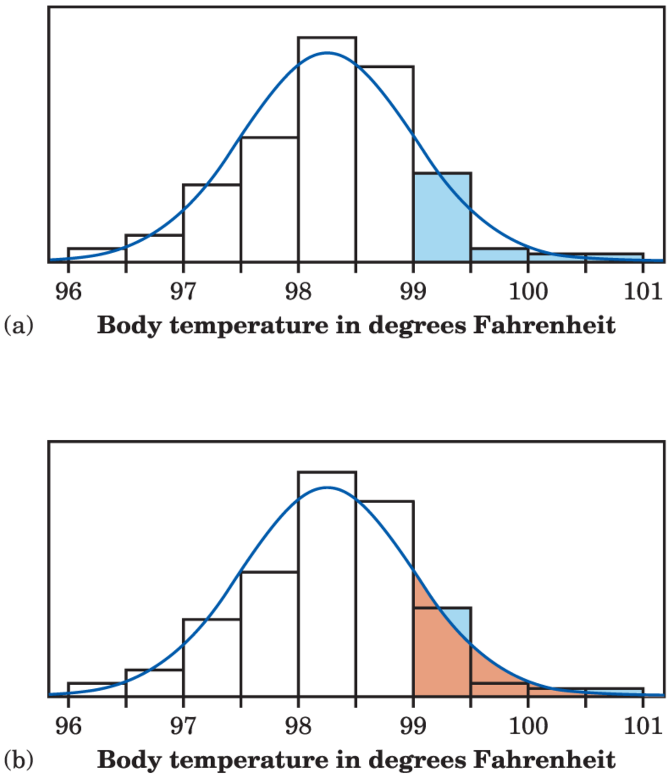




Moore/Notz, *Statistics: Concepts and Controversies*, 10e,

The proportion of observations greater than 99 is 19/130, or 0.146. Because 99 is one of the break points between the classes in the histogram, the area of the shaded bars in Figure 13.4(a) makes up 0.146 of the total area of all the bars.

Example: Using a density curve 3



Moore/Notz, *Statistics: Concepts and Controversies*, 10e,

The total area under the density curve is 1, and the shaded area in Figure 13.4(b) represents the proportion of observations that are greater than or equal to 99°F. This area is 0.1587. You can see that the density curve is a quite good approximation: 0.1587 is close to 0.146.

Center and Variability of a Density Curve 1

Areas under a density curve represent proportions of the total number of observations.

The median is the point with half the observations on either side. So the median of a density curve is the equal-areas point, the point with half the area under the curve to its left and the remaining half of the area to its right.

The median on a symmetric density curve is at is center.

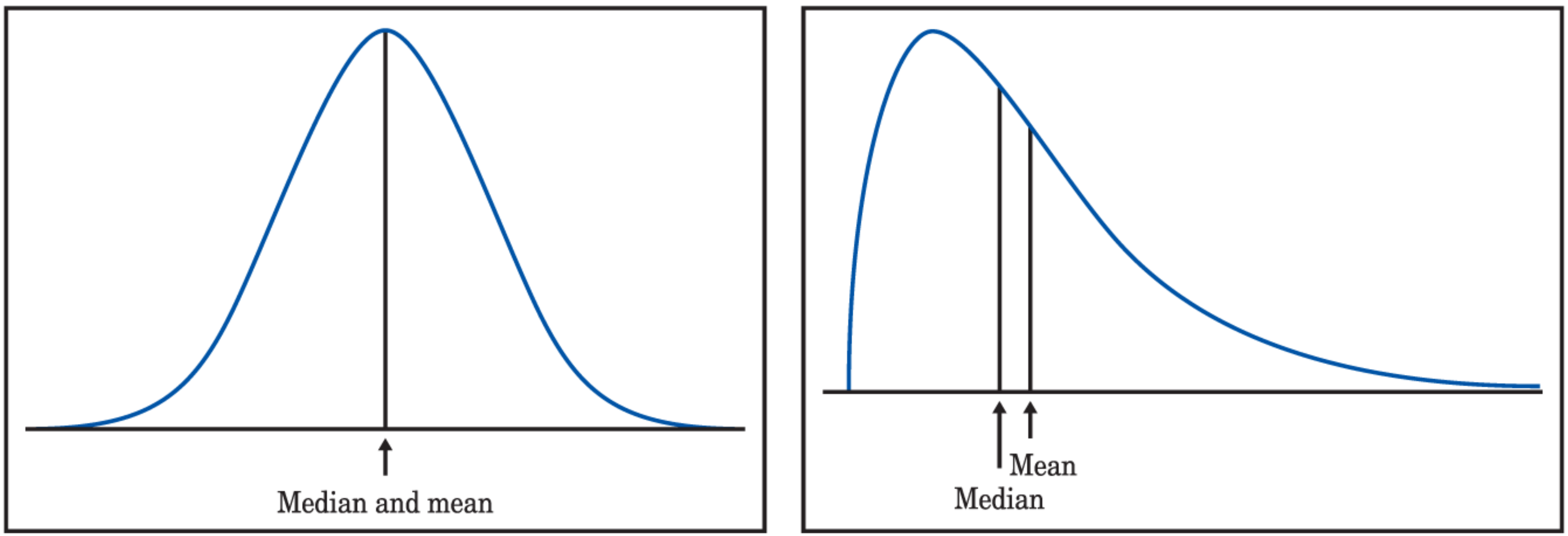
Center and Variability of a Density Curve 2

The mean of a set of observations is their arithmetic average. If we think of the observations as weights stacked on a seesaw, the mean is the point at which the seesaw would balance.

This fact is also true of density curves. The mean is the point at which the curve would balance if made of solid material. Figure 13.6 illustrates this fact about the mean:

A symmetrical curve balances at its center because the two sides are identical.

Center and Variability of a Density Curve 3



(a) (b)

Center and Variability of a Density Curve 4

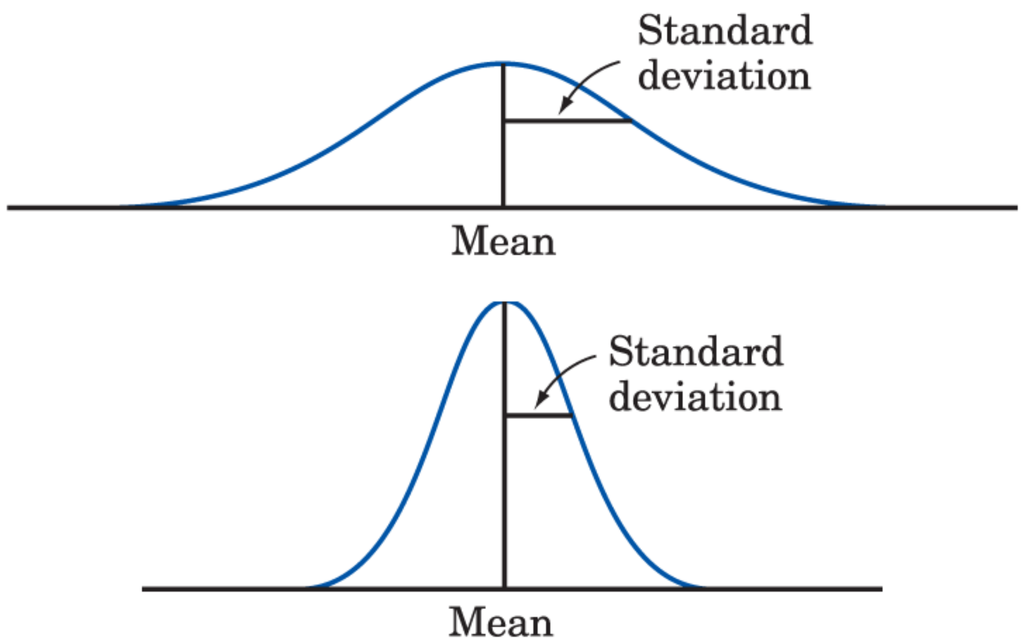
The **median** of a density curve is the equal-areas point, the point that divides the area under the curve in half.

The **mean** of a density curve is the balance point, or center of gravity, at which the curve would balance if made of solid material.

The median and mean are the same for a symmetrical density curve. They both lie at the center of the curve. The mean of a skewed curve is pulled away from the median in the direction of the long tail.

Normal Distributions 1

Figure 13.7 presents two Normal density curves. Normal curves are symmetrical, single-peaked, and bell-shaped.



Moore/Notz, *Statistics: Concepts and Controversies*, 10e,

Normal Distributions 2

Normal curves also have the special property that we can locate the standard deviation of the distribution by eye on the curve. Imagine that you are skiing down a mountain that has the shape of a Normal curve. At first, you descend at an ever-steeper angle as you go out from the peak:

Normal Distributions 3

Fortunately, before you find yourself going straight down, the slope begins to grow flatter rather than steeper as you go out and down:

Normal Distributions 4

The points at which this change of curvature takes place are located one standard deviation on either side of the mean. The standard deviations are marked on the two curves in Figure 13.7.

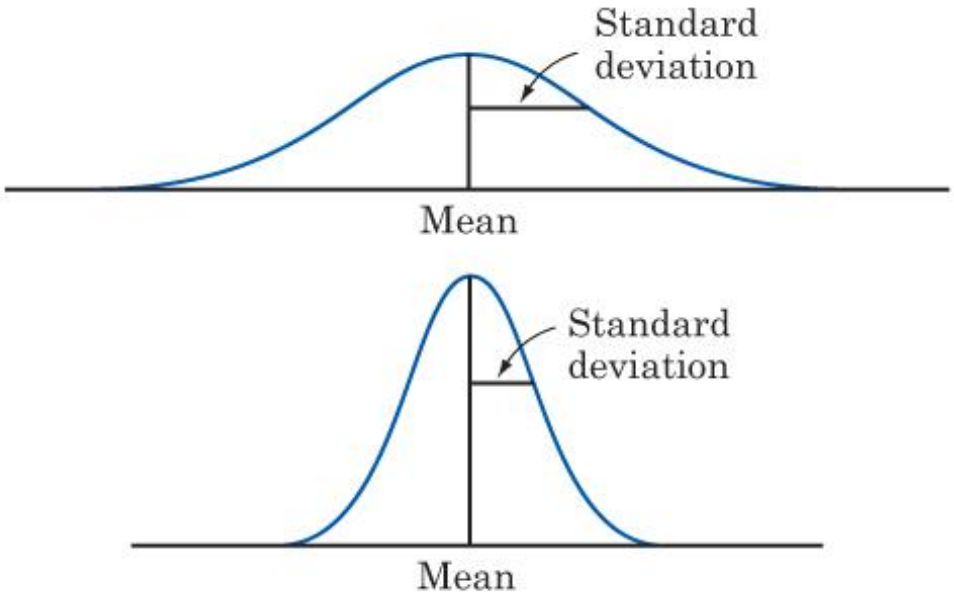


Figure 13.7
Moore/Notz, *Statistics: Concepts and Controversies*, 9e, © 2017 W. H. Freeman and Company

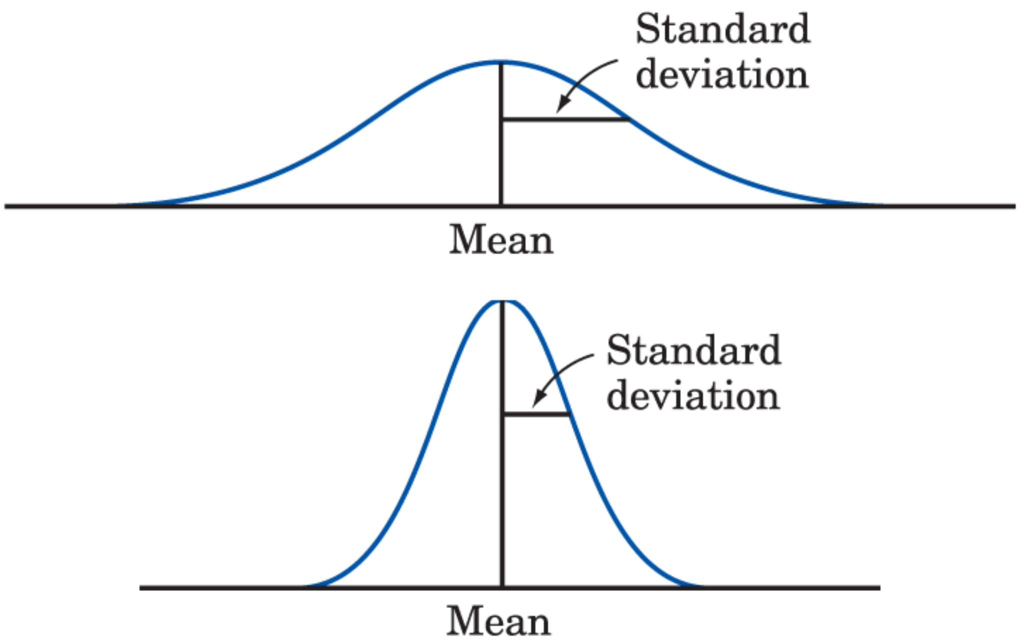
Normal Distributions 5

The mean determines the center of the curve, and the standard deviation determines its shape.

Changing the mean of a Normal distribution does not change its shape, only its location on the axis.

Changing the standard deviation does change the shape of a Normal curve.

Normal Distributions 6



The distribution with the smaller standard deviation is less variable and more sharply peaked. The mean determines location.

Normal Distributions 7

The **Normal curves** are symmetrical, bell-shaped curves that have these properties:

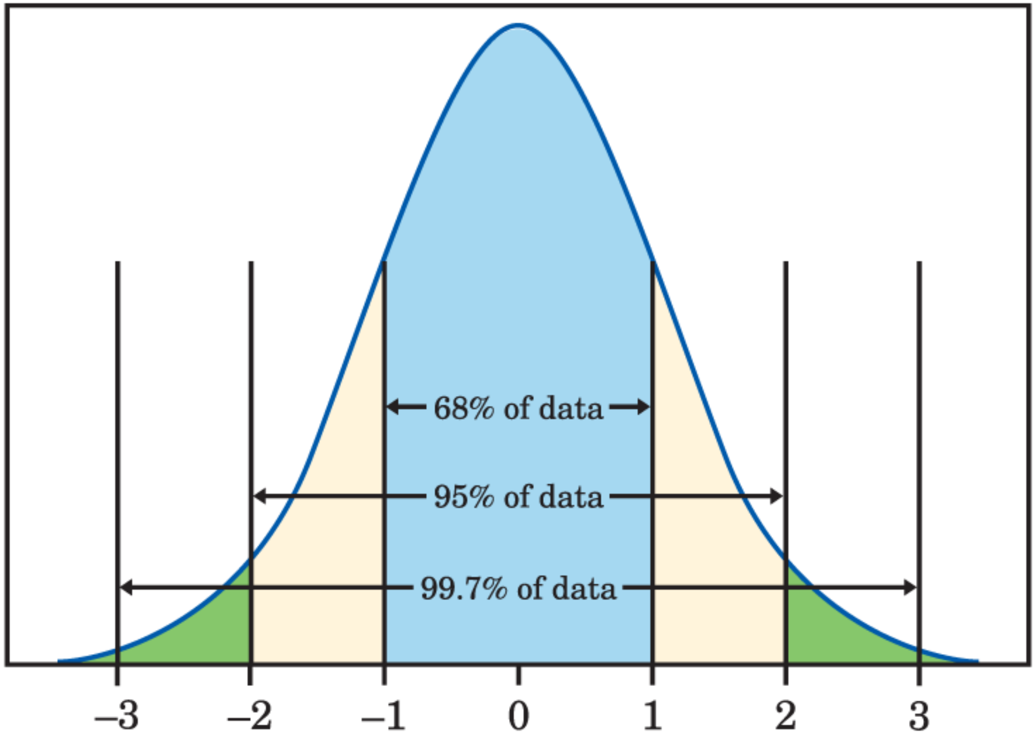
- A specific Normal curve is completely described by giving its mean and its standard deviation.
- The mean determines the center of the distribution. It is located at the center of symmetry of the curve.
- The standard deviation determines the shape of the curve. It is the distance from the mean to the change-of-curvature points on either side.

The 68–95–99.7% Rule

In any Normal distribution, approximately

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

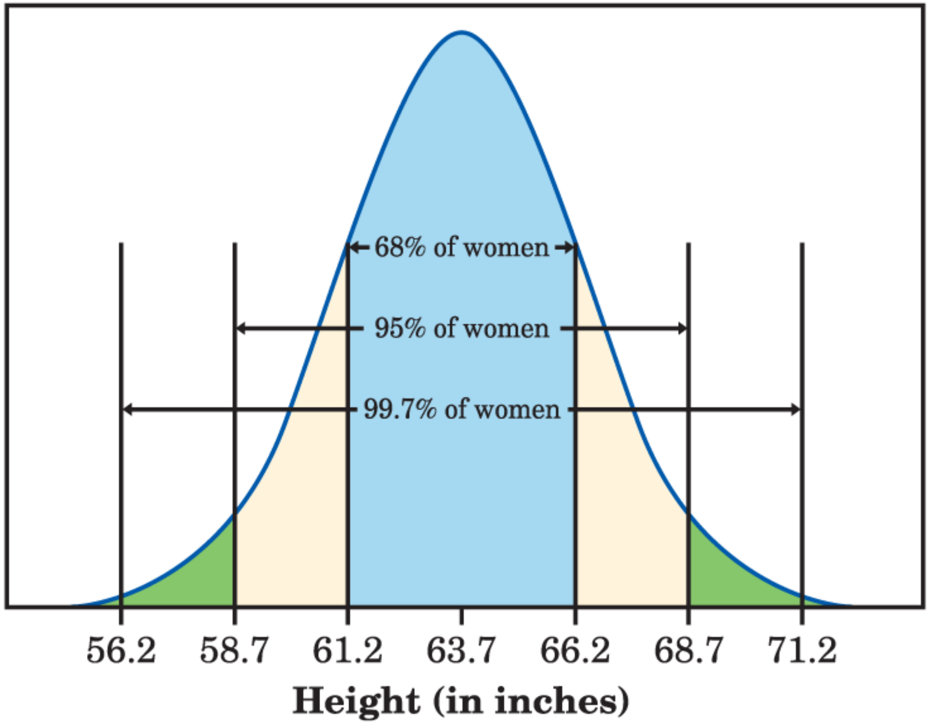
The 68–95–99.7% Rule (continued)



Moore/Notz, *Statistics: Concepts and Controversies*, 10e,

Example: Heights of young women 1

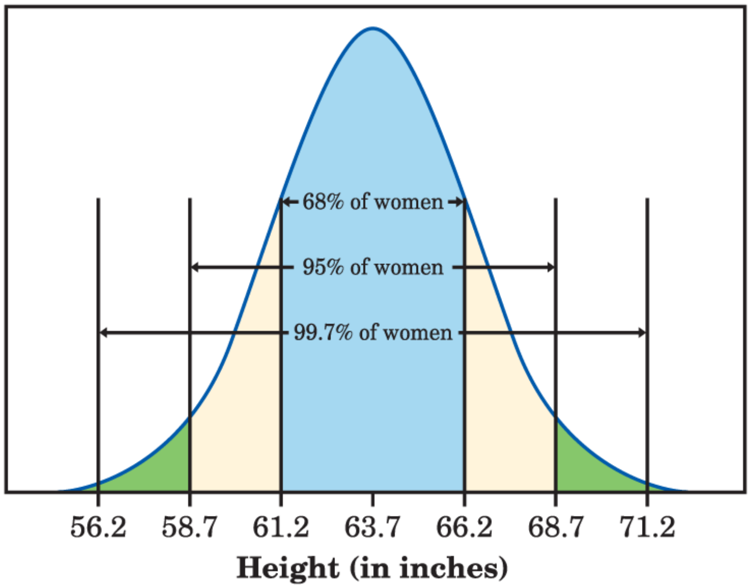
The distribution of heights of women aged 18 to 24 is approximately Normal with mean 63.7 inches and a standard deviation of 2.5 inches.



Moore/Notz, *Statistics: Concepts and Controversies*, 10e,

Example: Heights of young women 2

Half of the observations in any Normal distribution lie above the mean, so half of all young women are taller than 63.7 inches. The central 68% of any Normal distribution lies within one standard deviation of the mean. Half of this central 68%, or 34%, lies above the mean. So 34% of young women are between 63.7 inches and 66.2 inches.



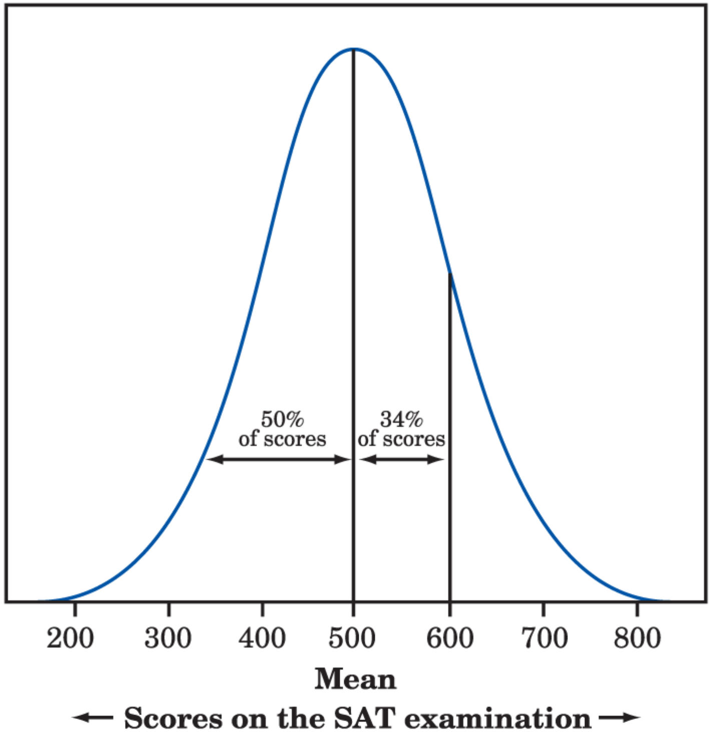
Example: Heights of young women 3

Half of the observations in any Normal distribution lie above the mean, so half of all young women are taller than 63.7 inches. The central 68% of any Normal distribution lies within one standard deviation of the mean. Half of this central 68%, or 34%, lies above the mean. So 34% of young women are between 63.7 inches and 66.2 inches tall. Adding the 50% who are shorter than 63.7 inches, we see that 84% of young women have heights of less than 66.2 inches. That leaves 16% who are taller than 66.2 inches.

Example: Heights of young women 4

The central 95% of any Normal distribution lies within two standard deviations of the mean. Two standard deviations is 5 inches here, so the middle 95% of young women's heights are between 58.7 inches (that's 63.7 − 5) and 68.7 inches (that's 63.7 + 5). The other 5% of young women have heights outside the range from 58.7 to 68.7 inches. Because the Normal distributions are symmetric, half of these women are on the short side. The shortest 2.5% of young women are less than 58.7 inches tall. Almost all (99.7%) of the observations in any Normal distribution lie within three standard deviations of the mean. Almost all young women are between 56.2 and 71.2 inches tall.

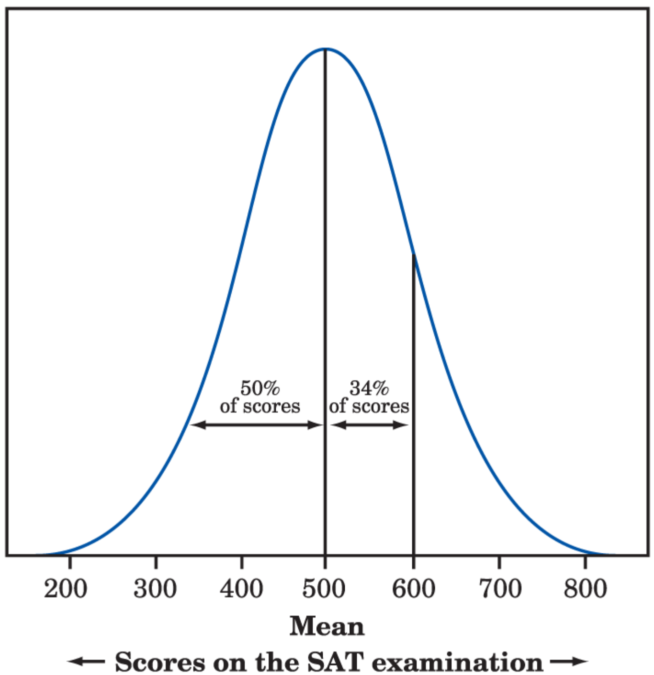
Standard Scores 1



Madison scored 600 on the SAT Mathematics college entrance exam. How good a score is this?

The SAT exams are scaled so that scores should roughly follow the Normal distribution with mean 500 and standard deviation 100. Madison's 600 is one standard deviation above the mean.

Standard Scores 2



Moore/Notz, *Statistics: Concepts and Controversies*, 10e,

Madison did better than 84% of the students who took the SAT. Her score report will say that she scored at the "84th percentile." That's statistics speak for "You did better than 84% of those who took the test."

Standard Scores 3

We restated Madison's score of 600 as "one standard deviation above the mean."

Observations expressed in standard deviations above or below the mean of a distribution are called standard scores.

Standard scores are also sometimes referred to as *z*-scores.

Standard Scores 4

The **standard score** for any observation is

$$\frac{observation - mean}{standard\ deviation}$$

A positive standard score indicates the observation is above the mean.
A negative standard score indicates the observation is below the mean.

- Use standard scores only for roughly symmetric distributions.

Example: ACT versus SAT scores 1

Madison scored 600 on the SAT Mathematics exam. Her friend Gabriel took the American College Testing (ACT) test and scored 21 on the math part. ACT scores are Normally distributed with mean 18 and standard deviation 6. Assuming that both tests measure the same kind of ability, who has the higher score?

Madison's standard score is $\frac{600-500}{100} = 1$

Gabriel's standard score is $\frac{21-18}{3} = 0.5$

Example: ACT versus SAT scores 2

Because Madison's score is one standard deviation above the mean and Gabriel's is only 0.5 standard deviation above the mean, Madison's performance is better.

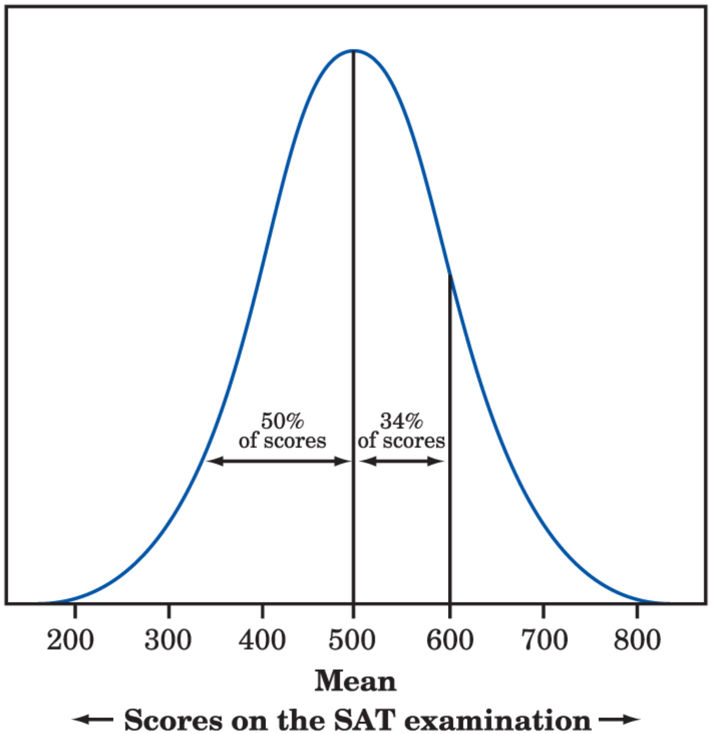
Percentiles of Normal Distributions 1

The ***c*th percentile** of a distribution is a value such that *c* percent of the observations lie below it and the rest lie above.

The median of any distribution is the 50th percentile, and the quartiles are the 25th and 75th percentiles.

Percentiles of Normal Distributions 2

In any Normal distribution, the point one standard deviation above the mean (standard score 1) is the 84th percentile. Figure 13.10 shows why.



Moore/Notz, *Statistics: Concepts and Controversies*, 10e, © 2020 W. H. Freeman and Company

Percentiles of Normal Distributions 3

Every standard score for a Normal distribution translates into a specific percentile, which is the same no matter what the mean and standard deviation of the original Normal distribution are.

Table B at the back of this book gives the percentiles corresponding to various standard scores. This table enables us to do calculations in greater detail than does the 68–95–99.7 rule.

Example: ACT versus SAT scores 3

Madison's score of 600 on the SAT translates into a standard score of 1.0. We saw that the 68–95–99.7 rule says that this is the 84th percentile.

Table B is a bit more precise: it says that standard score 1 is the 84.13 percentile of a Normal distribution.

Gabriel's 21 on the ACT is a standard score of 0.5. Table B says that this is the 69.15 percentile.

Gabriel did well, but not as well as Madison. The percentile is easier to understand than either the raw score or the standard score.

Statistics in Summary 1

- You can roughly locate the median (equal-areas point) and the mean (balance point) on a density curve, by eye.

- Stemplots, histograms, and boxplots are created from samples. Density curves are intended to display the idealized shape of the distribution of the population from which the samples are taken.

Statistics in Summary 2

- Stemplots, histograms, and boxplots all describe the distributions of quantitative variables.

- **Density curves** also describe distributions. A density curve is a curve with area exactly 1 underneath it, whose shape describes the overall pattern of a distribution.

- An area under the curve gives the proportion of the observations that fall in an interval of values.

Statistics in Summary 3

- **Normal curves** are a special kind of density curve that describes the overall pattern of some sets of data. Normal curves are symmetric and bell shaped. A specific Normal curve is completely described by its mean and standard deviation. You can locate the mean (center point) and the standard deviation (distance from the mean to the change-of-curvature points) on a Normal curve. All Normal distributions obey the **68–95–99.7 rule**

Statistics in Summary 4

- **Standard scores** express observations in standard deviation units about the mean, which has standard score 0. A given standard score corresponds to the same percentile in any Normal distribution. Table B gives percentiles of Normal distributions.