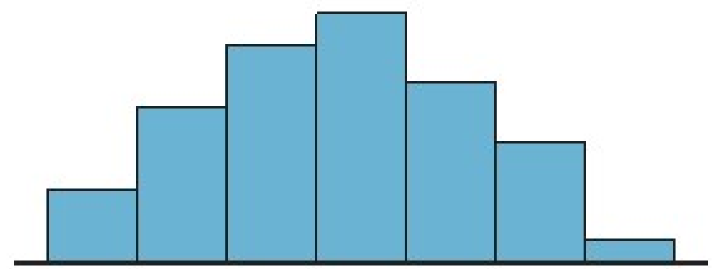
Normal Distribution

Random variables can be either discrete or continuous.

Many continuous variables have distributions that are bell – shaped, and these are called approximately normally distributed.

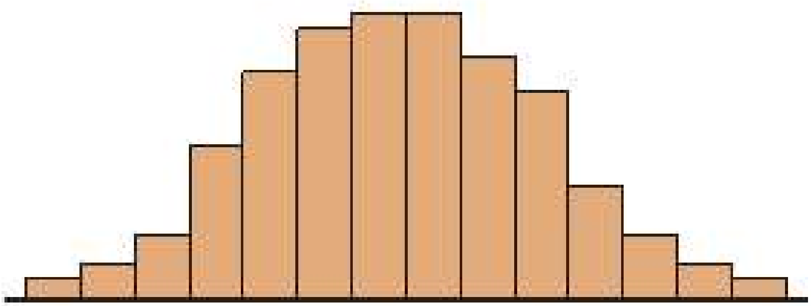
For example, if a researcher selects a random sample of 100 adult women, measures their heights, and constructs a histogram, the researcher get a graph similar to the one shown in Figure below:



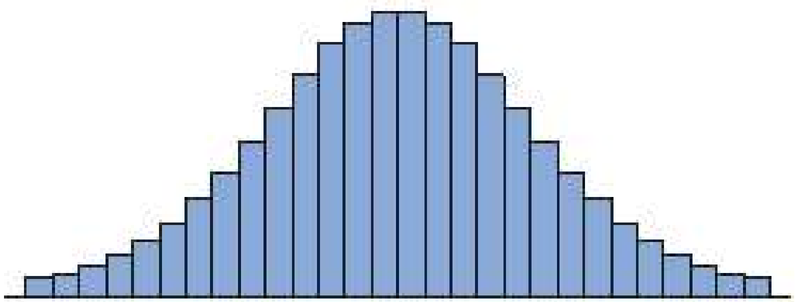
(a) Random sample of 100 women

Normal Distribution

Now, if the researcher increases the sample size and decreases the width of the classes, the histograms will look like the ones shown here



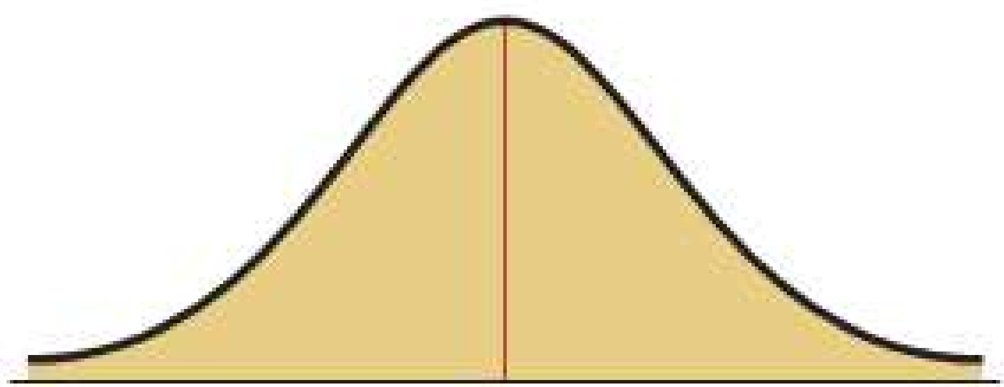
(b) Sample size increased and class width decreased



(c) Sample size increased and class width decreased further

Normal Distribution

Finally if it were possible to measure exactly the heights of all adult females in the Iraq Universities and plot them, the histogram would approach what is called a normal distribution, shown in Figure below. This distribution is also known as a bell curve or a Gaussian distribution.



(d) Normal distribution for the population

Normal Distribution

No variable fits a normal distribution perfectly, since a normal distribution is a theoretical distribution.

However, a normal distribution can be used to describe many variables, because the deviations from a normal distribution are very small.

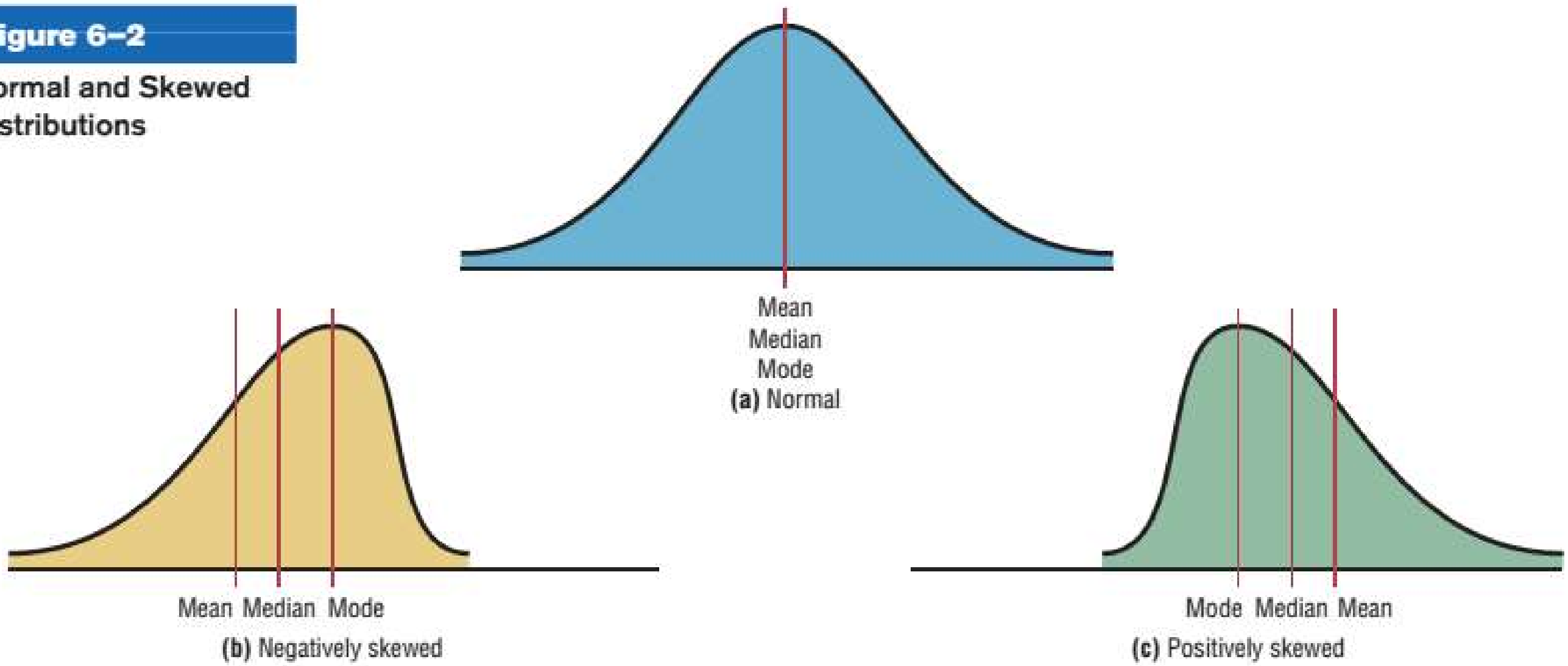
When the data values are evenly distributed about the mean, a distribution is said to be a **symmetric distribution**.

When the majority of the data values fall to the left or right of the mean, the distribution is said to be skewed.

Normal Distribution

Figure 6–2

Normal and Skewed Distributions



Properties of normal distribution

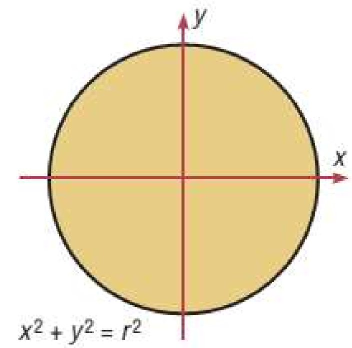
In mathematic, curve can be represented by equations

For example, the equation of the circle is $x^2 + y^2 = r^2$, where r is the circle radius

A circle can be used to represent many physical objects, such as a wheel or a gear.

Even though it is not possible to manufacture a wheel that is perfectly round, the equation and the properties of a circle can be used to study many aspects of the wheel, such as area, velocity, and acceleration.

Circle



Wheel

Properties of normal distribution

In a similar manner, the theoretical curve, called a normal distribution curve, can be used to study many variables that are not perfectly normally distributed but are nevertheless approximately normal.

The mathematical equation for a normal distribution is

$$y = \frac{e^{-(X-\mu)^2/(2\sigma^2)}}{\sigma\sqrt{2\pi}}$$

where

$e \approx 2.718$ ($\approx$ means "is approximately equal to")

$\pi \approx 3.14$

$\mu =$ population mean

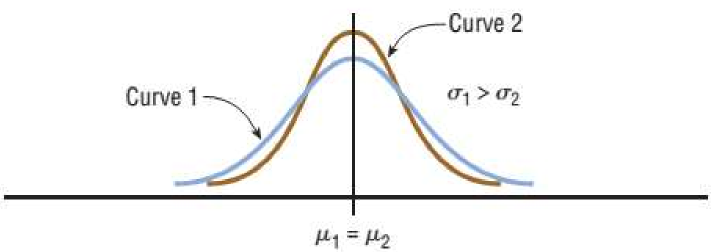
$\sigma =$ population standard deviation

This equation may look formidable, but in applied statistics, tables or technology is used for specific problems instead of the equation.

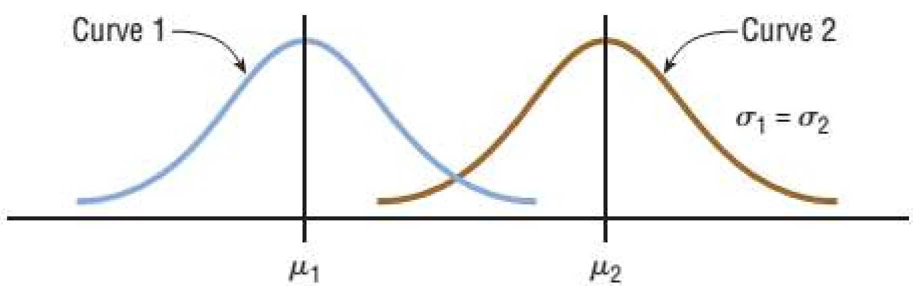
Properties of normal distribution

The shape and position of a normal distribution curve depends on two parameters, the μ and the σ .

Each normally distributed variable has its own normal distribution curve, which depends on the values of the variable mean and standard deviation.

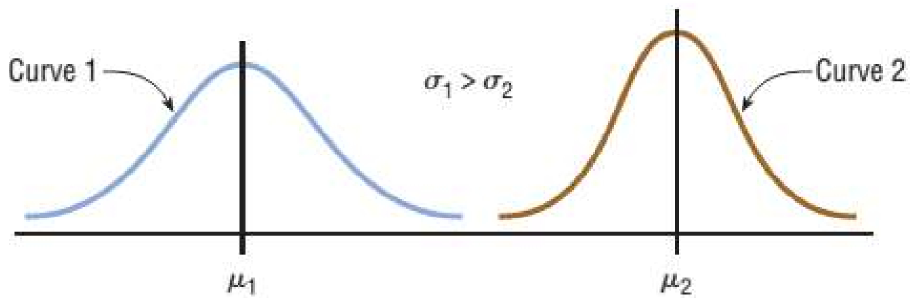


(a) Same means but different standard deviations



(b) Different means but same standard deviations

Properties of normal distribution



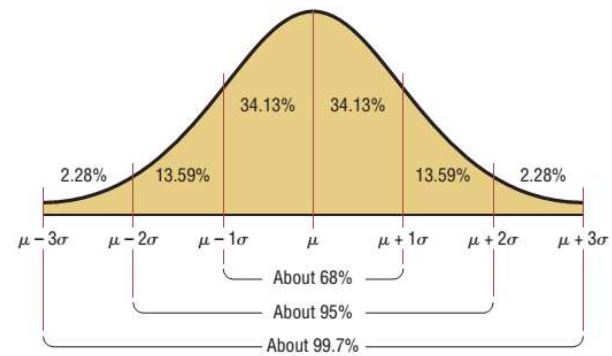
(c) Different means and different standard deviations

A **normal distribution** is a continuous, symmetric, bell-shaped distribution of a variable.

Properties of normal distribution

Summary of the Properties of the Theoretical Normal Distribution

1. A normal distribution curve is bell-shaped.
2. The mean, median, and mode are equal and are located at the center of the distribution.
3. A normal distribution curve is unimodal (i.e., it has only one mode).
4. The curve is symmetric about the mean, which is equivalent to saying that its shape is the same on both sides of a vertical line passing through the center.
5. The curve is continuous; that is, there are no gaps or holes. For each value of X, there is a corresponding value of Y.
6. The curve never touches the x axis. Theoretically, no matter how far in either direction the curve extends, it never meets the x axis—but it gets increasingly closer.
7. The total area under a normal distribution curve is equal to 1.00, or 100%. This fact may seem unusual, since the curve never touches the x axis, but one can prove it mathematically by using calculus. (The proof is beyond the scope of this textbook.)
8. The area under the part of a normal curve that lies within 1 standard deviation of the mean is approximately 0.68, or 68%; within 2 standard deviations, about 0.95, or 95%; and within 3 standard deviations, about 0.997, or 99.7%. See Figure 6–5, which also shows the area in each region.



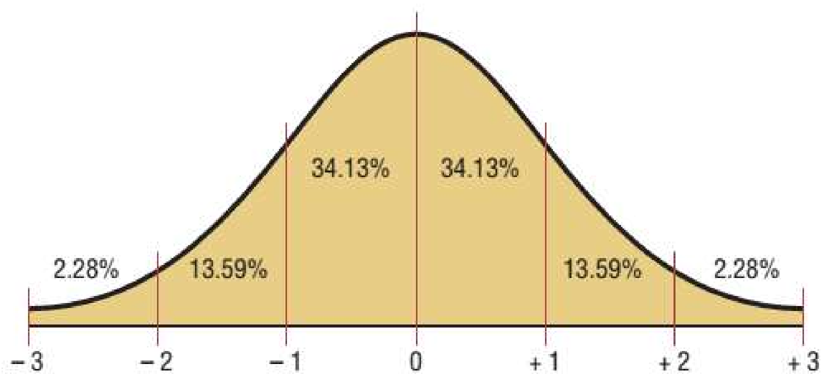
The standard normal distribution

Since each normally distributed variable has its own mean and standard deviation, the shape and location of these curves will vary.

In particular applications, you would have to have a table of areas under the curves for each variable.

To simplify this situation, statisticians use what is called the standard normal distribution.

The **standard normal distribution** is a normal distribution with a mean of 0 and a standard deviation of 1.



The standard normal distribution

The values under the curve indicate the proportion of area in each section.

The formula for the standard distribution is

$$y = \frac{e^{-z^2/2}}{\sqrt{2\pi}}$$

All normally distributed variables can be transformed into the standard normally distributed variable by using the formula for the standard score:

$$z = \frac{\text{value} - \text{mean}}{\text{standard deviation}} \qquad \text{or} \qquad z = \frac{X - \mu}{\sigma}$$

The table used to give the area (to four decimal places) under the standard normal curve for any z value from -3.49 to 3.49

Table E The Standard Normal Distribution

Cumulative Standard Normal Distribution

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

For z values less than −3.49, use 0.0001.

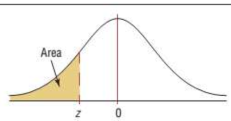
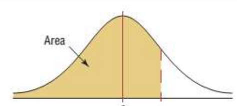


Table E (*continued*)

Cumulative Standard Normal Distribution

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

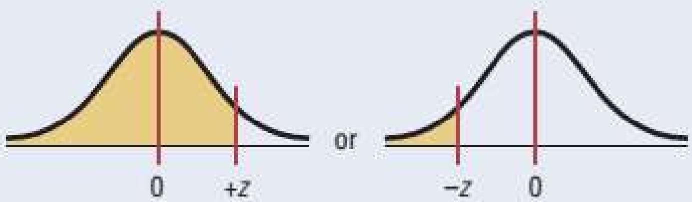
For z values greater than 3.49, use 0.9999.



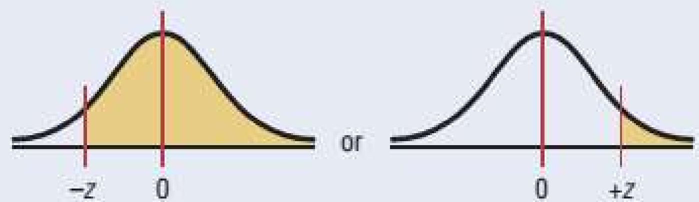
Procedure Table

Finding the Area Under the Standard Normal Distribution Curve

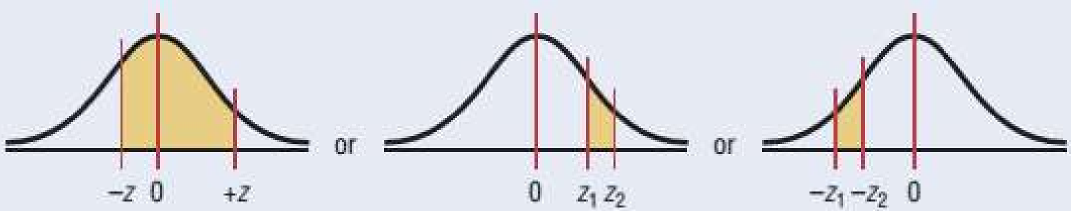
1. To the left of any z value:
   Look up the z value in the table and use the area given.



2. To the right of any z value:
   Look up the z value and subtract the area from 1.



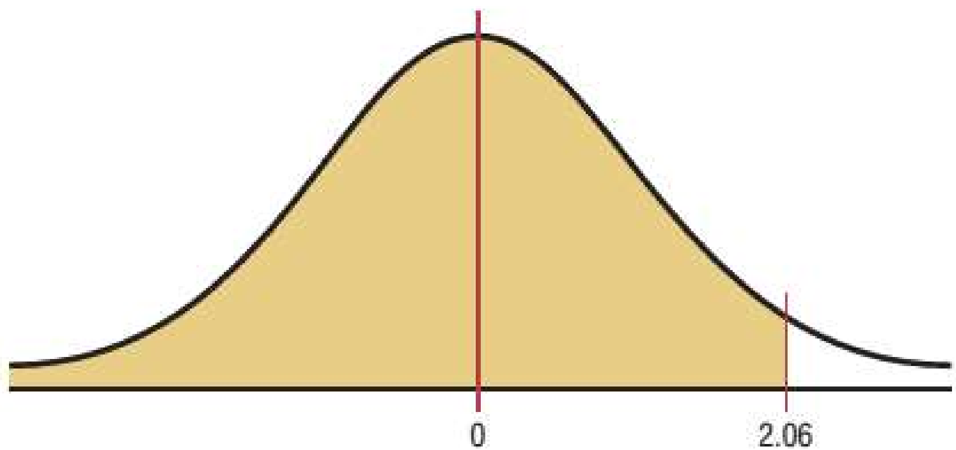
3. Between any two z values:
   Look up both z values and subtract the corresponding areas.



Find the area to the left of $z = 2.06$.

Solution

Step 1 Draw the figure. The desired area is shown in Figure 6–8.

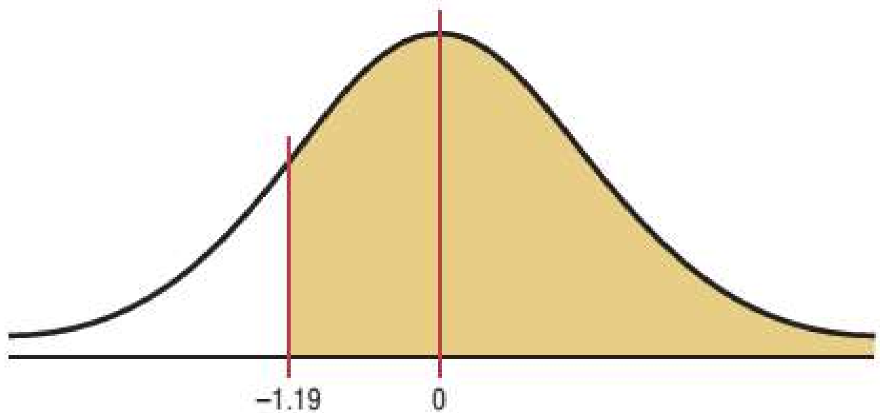


Step 2 We are looking for the area under the standard normal distribution to the left of $z = 2.06$. Since this is an example of the first case, look up the area in the table. It is 0.9803. Hence, 98.03% of the area is less than $z = 2.06$.

Find the area to the right of $z = -1.19$.

Solution

Step 1 Draw the figure. The desired area is shown in Figure 6–9.

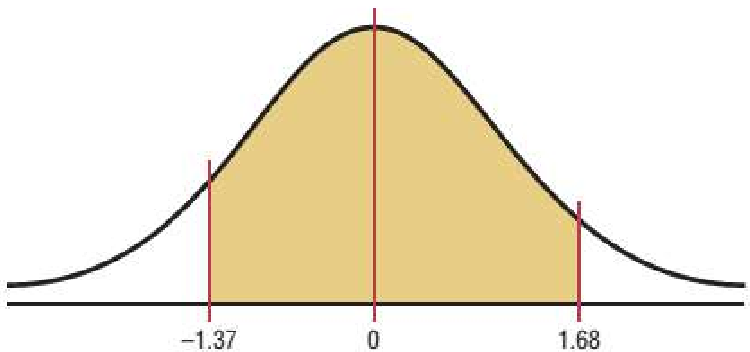


Step 2 We are looking for the area to the right of $z = -1.19$. This is an example of the second case. Look up the area for $z = -1.19$. It is 0.1170. Subtract it from 1.0000. $1.0000 - 0.1170 = 0.8830$. Hence, 88.30% of the area under the standard normal distribution curve is to the left of $z = -1.19$.

Find the area between $z = +1.68$ and $z = -1.37$.

Solution

Step 1 Draw the figure as shown. The desired area is shown in Figure 6–10.



Step 2 Since the area desired is between two given z values, look up the areas corresponding to the two z values and subtract the smaller area from the larger area. (Do not subtract the z values.) The area for $z = +1.68$ is 0.9535, and the area for $z = -1.37$ is 0.0853. The area between the two z values is $0.9535 - 0.0853 = 0.8682$ or 86.82%.

Find the probability for each.

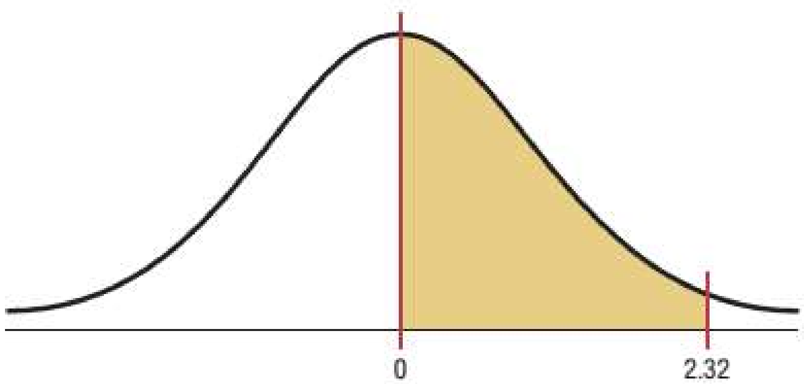
a. $P(0 < z < 2.32)$

b. $P(z < 1.65)$

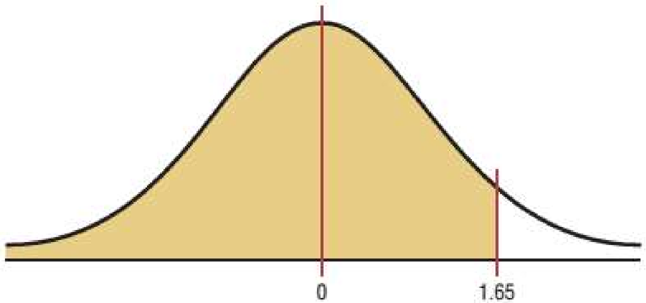
c. $P(z > 1.91)$

Solution

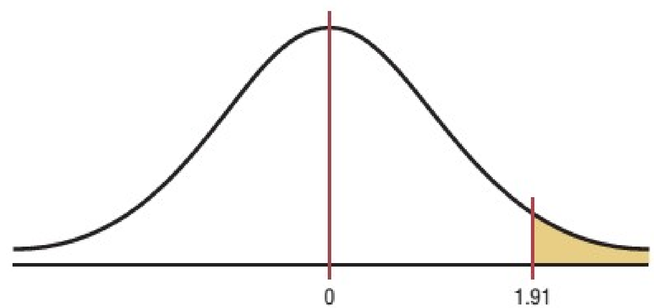
a. $P(0 < z < 2.32)$ means to find the area under the standard normal distribution curve between 0 and 2.32. First look up the area corresponding to 2.32. It is 0.9898. Then look up the area corresponding to $z = 0$. It is 0.500. Subtract the two areas: $0.9898 - 0.5000 = 0.4898$. Hence the probability is 0.4898, or 48.98%. This is shown in Figure 6–11.



b. $P(z < 1.65)$ is represented in Figure 6–12. Look up the area corresponding to $z = 1.65$ in Table E. It is 0.9505. Hence, $P(z < 1.65) = 0.9505$, or 95.05%.



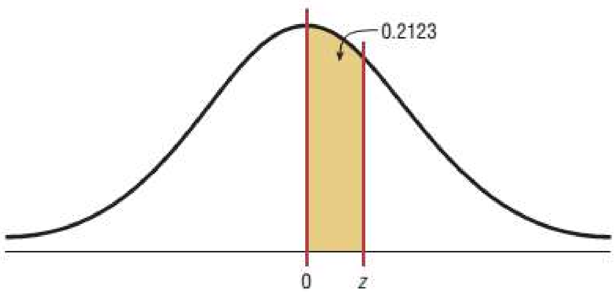
c. $P(z > 1.91)$ is shown in Figure 6–13. Look up the area that corresponds to $z = 1.91$. It is 0.9719. Then subtract this area from 1.0000. $P(z > 1.91) = 1.0000 - 0.9719 = 0.0281$, or 2.81%.



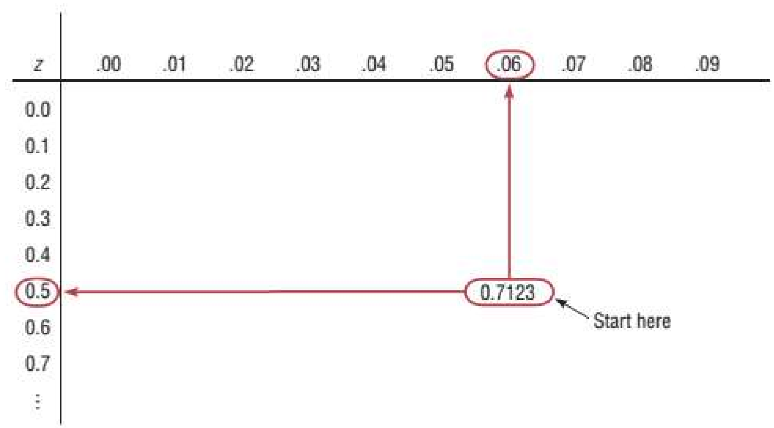
Find the z value such that the area under the standard normal distribution curve between 0 and the z value is 0.2123.

Solution

Draw the figure. The area is shown in Figure



In this case it is necessary to add 0.5000 to the given area of 0.2123 to get the cumulative area of 0.7123. Look up the area in Table E. The value in the left column is 0.5, and the top value is 0.06. Add these two values to get $z = 0.56$. See Figure



Application of the normal distribution

The standard normal distribution curve can be used to solve a wide variety of practical problems. The only requirement is that the variable be normally or approximately normally distributed.

To solve problems by using the standard normal distribution, transform the original variable to a standard normal distribution variable by using the formula

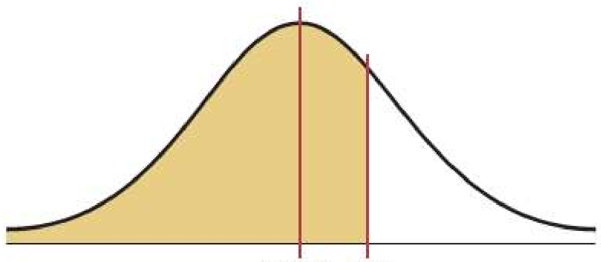
$$z = \frac{value - mean}{standard\ deviation} \quad or\ z = \frac{X - \mu}{\sigma}$$

Summer Spending

A survey found that women spend on average \$146.21 on beauty products during the summer months. Assume the standard deviation is \$29.44. Find the percentage of women who spend less than \$160.00. Assume the variable is normally distributed.

Solution

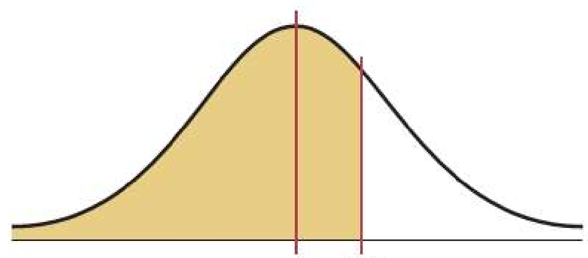
Step 1 Draw the figure and represent the area as shown in Figure



Step 2 Find the z value corresponding to \$160.00.

$$z = \frac{X - \mu}{\sigma} = \frac{\$160.00 - \$146.21}{\$29.44} = 0.47$$

Hence \$160.00 is 0.47 of a standard deviation above the mean of \$146.21, as shown in the z distribution in Figure



Step 3 Find the area, using Table E. The area under the curve to the left of $z = 0.47$ is 0.6808.

Therefore 0.6808, or 68.08%, of the women spend less than \$160.00 on beauty products during the summer months.

Monthly Newspaper Recycling

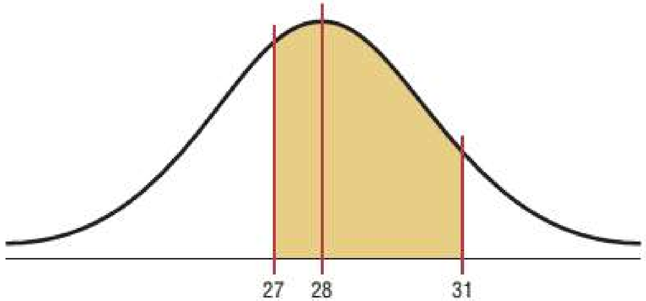
Each month, an American household generates an average of 28 pounds of newspaper for garbage or recycling. Assume the standard deviation is 2 pounds. If a household is selected at random, find the probability of its generating

a. Between 27 and 31 pounds per month

b. More than 30.2 pounds per month

Assume the variable is approximately normally distributed.

Step 1 Draw the figure and represent the area. See Figure 6–20.

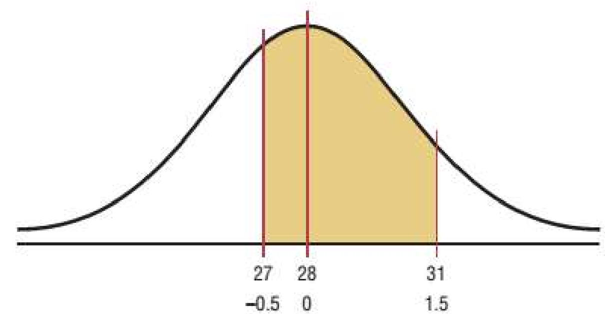


Step 2 Find the two z values.

$$z_1 = \frac{X - \mu}{\sigma} = \frac{27 - 28}{2} = -\frac{1}{2} = -0.5$$

$$z_2 = \frac{X - \mu}{\sigma} = \frac{31 - 28}{2} = \frac{3}{2} = 1.5$$

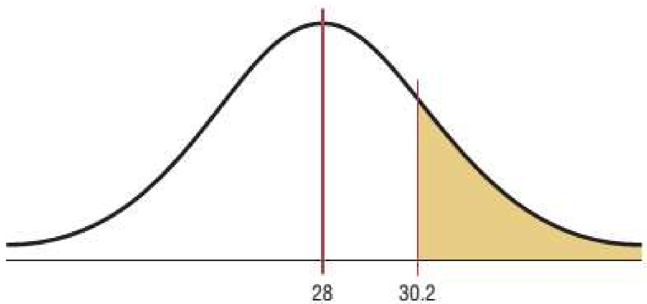
Step 3 Find the appropriate area, using Table E. The area to the left of z_2 is 0.9332, and the area to the left of z_1 is 0.3085. Hence the area between z_1 and z_2 is $0.9332 - 0.3085 = 0.6247$. See Figure 6–21.



Hence, the probability that a randomly selected household generates between 27 and 31 pounds of newspapers per month is 62.47%.

Solution *b*

Step 1 Draw the figure and represent the area, as shown in Figure 6–22.



Step 2 Find the z value for 30.2.

$$z = \frac{X - \mu}{\sigma} = \frac{30.2 - 28}{2} = \frac{2.2}{2} = 1.1$$

Step 3 Find the appropriate area. The area to the left of $z = 1.1$ is 0.8643. Hence the area to the right of $z = 1.1$ is $1.0000 - 0.8643 = 0.1357$.

Hence, the probability that a randomly selected household will accumulate more than 30.2 pounds of newspapers is 0.1357, or 13.57%.

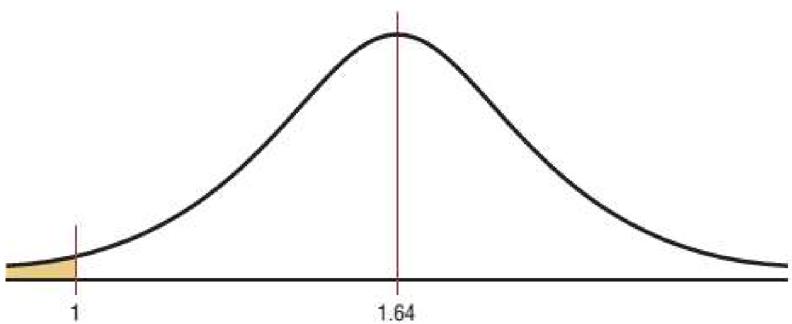
Coffee Consumption

Americans consume an average of 1.64 cups of coffee per day. Assume the variable is approximately normally distributed with a standard deviation of 0.24 cup. If 500 individuals are selected, approximately how many will drink less than 1 cup of coffee per day?

Source: *Chicago Sun-Times.*

Solution

Step 1 Draw a figure and represent the area as shown in Figure 6–23.



Step 2 Find the z value for 1.

$$z = \frac{X - \mu}{\sigma} = \frac{1 - 1.64}{0.24} = -2.67$$

Step 3 Find the area to the left of $z = -2.67$. It is 0.0038.

Step 4 To find how many people drank less than 1 cup of coffee, multiply the sample size 500 by 0.0038 to get 1.9. Since we are asking about people, round the answer to 2 people. Hence, approximately 2 people will drink less than 1 cup of coffee a day.

Determining Normality

A normally shaped or bell-shaped distribution is only one of many shapes that a distribution can assume; however, it is very important since many statistical methods require that the distribution of values (shown in subsequent chapters) be normally or approximately normally shaped.

There are several ways statisticians check for normality. The easiest way is to draw a histogram for the data and check its shape. If the histogram is not approximately bell-shaped, then the data are not normally distributed.

Skewness can be checked by using the Pearson coefficient of skewness (PC) also called Pearson's index of skewness. The formula is

$$PC = \frac{3(\overline{X} - \text{median})}{s}$$

If the index is greater than or equal to $+1$ or less than or equal to -1, it can be concluded that the data are significantly skewed.

Technology Inventories

A survey of 18 high-technology firms showed the number of days' inventory they had on hand. Determine if the data are approximately normally distributed.

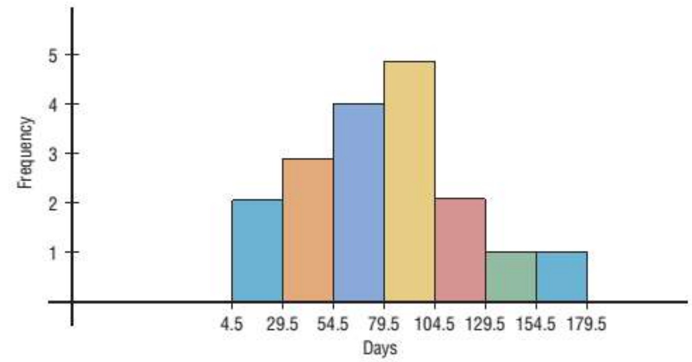
5	29	34	44	45	63	68	74	74
81	88	91	97	98	113	118	151	158

Source: *USA TODAY.*

Solution

Step 1 Construct a frequency distribution and draw a histogram for the data, as shown in Figure 6–27.

Class	Frequency
5–29	2
30–54	3
55–79	4
80–104	5
105–129	2
130–154	1
155–179	1



Since the histogram is approximately bell-shaped, we can say that the distribution is approximately normal.

Step 2 Check for skewness. For these data, $\overline{X} = 79.5$, median $= 77.5$, and $s = 40.5$. Using the Pearson coefficient of skewness gives

$$PC = \frac{3(79.5 - 77.5)}{40.5}$$
$$= 0.148$$

In this case, the PC is not greater than $+1$ or less than -1, so it can be concluded that the distribution is not significantly skewed.

Number of Baseball Games Played

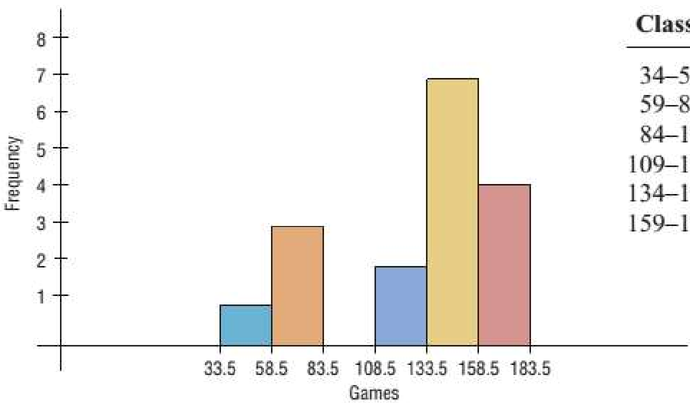
The data shown consist of the number of games played each year in the career of Baseball Hall of Famer Bill Mazeroski. Determine if the data are approximately normally distributed.

81	148	152	135	151	152
159	142	34	162	130	162
163	143	67	112	70	

Source: *Greensburg Tribune Review.*

Solution

Step 1 Construct a frequency distribution and draw a histogram for the data. See Figure 6–28.



Class	Frequency
34–58	1
59–83	3
84–108	0
109–133	2
134–158	7
159–183	4

The histogram shows that the frequency distribution is somewhat negatively skewed.

Step 2 Check for skewness; $\bar{X} = 127.24$, median $= 143$, and $s = 39.87$.

$$\text{PC} = \frac{3(\bar{X} - \text{median})}{s}$$
$$= \frac{3(127.24 - 143)}{39.87}$$
$$= -1.19$$

Since the PC is less than -1, it can be concluded that the distribution is significantly skewed to the left.

In summary, the distribution is somewhat negatively skewed.