Practice Run

1. The heights of 1567 Douglas fir trees were measured to have a mean of 126.8 ft and a standard deviation of 43.6 ft.

© *Thinkstock*

a. Assuming this data is normal, determine the percent of trees that you would expect to be shorter than 24 ft.

b. Assuming the data is normal, determine the percentage of trees you expect to be taller than 241 ft.

c. The shortest and tallest trees measured were 24 ft and 241 ft. Based on the information provided, is it reasonable to treat this data as normal? Explain.

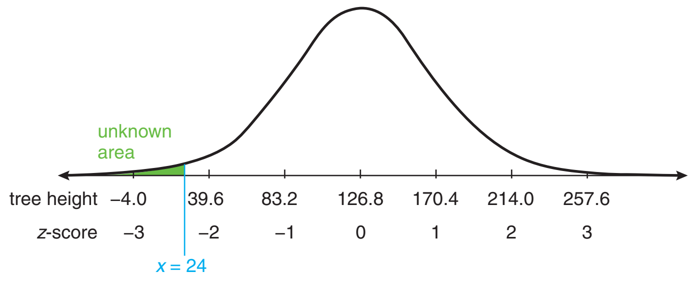
2. Suppose a set of data is normally distributed with a mean of 43 and a standard deviation of 12. Determine an upper limit and a lower limit that would encompass the **middle** 90% of the data.

Compare your answers.

1. The heights of 1567 Douglas fir trees were measured to have a mean of 126.8 ft and a standard deviation of 43.6 ft.

   a. Assuming this data is normal, determine the percent of trees that you would expect to be shorter than 24 ft.

   Sketch a normal curve to represent the problem.

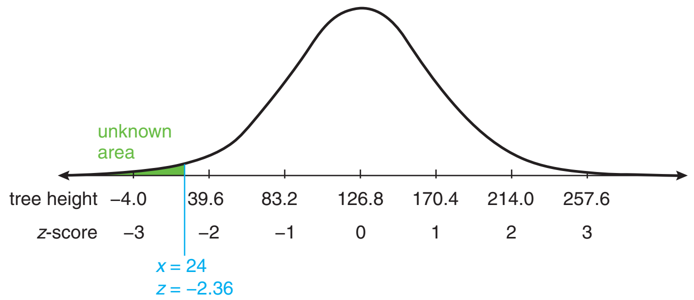


The unknown area represents the percentage of trees that are shorter than 24 ft. Determining the *z*-score will help you to determine the unknown area.

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

$$z=\frac{24-126.8}{43.6}$$

$$z=-2.36$$



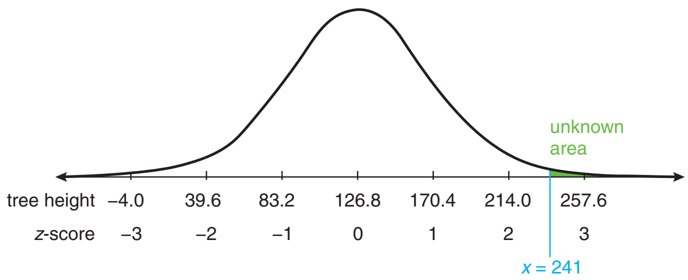
Now use technology or a *z*-score table to determine the area.

z	**0.09**	**0.08**	**0.07**	**0.06**	**0.05**	**0.04**	**0.03**	**0.02**	**0.01**	**0.00**
–2.4	0.0064	0.0066	0.0068	0.0069	0.0071	0.0073	0.0075	0.0078	0.0080	0.0082
–2.3	0.0084	0.0087	0.0089	**0.0091**	0.0094	0.0096	0.0099	0.0102	0.0104	0.0107
–2.2	0.0110	0.0113	0.0116	0.0119	0.0122	0.0125	0.0129	0.0132	0.0136	0.0139

The area to the left of a *z*-score of –2.36 is 0.0091, so approximately 0.91% of the trees will be shorter than 24 ft.

b. Assuming the data is normal, determine the percentage of trees you would expect to be taller than 241 ft.

Sketch a normal curve to represent the problem.

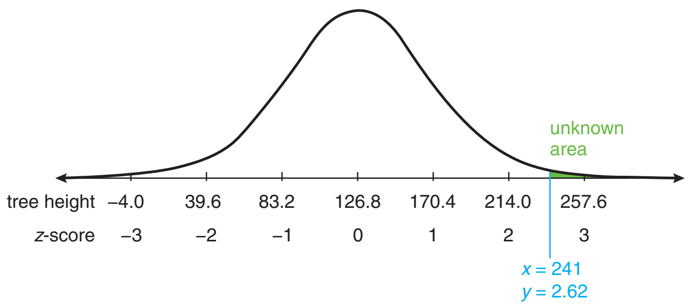


The unknown area represents the percentage of trees that are taller than 241 ft. Determining the *z*-score will help you to determine the unknown area.

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

$$z = \frac{241 - 126.8}{43.6}$$

$$z = 2.62$$



Now use a *z*-score table or technology to determine the unknown area. Using technology, the area is 0.0044. Approximately 0.44% of the trees will be taller than 241 ft.

Many computer programs and calculator applications will ask for an upper and a lower *z*-score when determining the area under a normal curve. There is essentially no data past a *z*-score of 5, so entering an upper limit *z*-score larger than 5 will give an extremely close approximation of the area you are looking for.

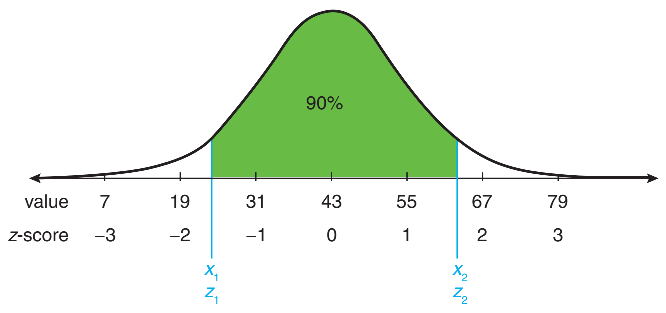
c. The shortest and tallest trees measured were 24 ft and 241 ft. Based on the information provided, is it reasonable to treat this data as normal? Explain.

The data is spread from nearly three standard deviations below the mean to nearly three standard deviations above the mean since 24 ft corresponds to a *z*-score of –2.36 and 241 ft corresponds to a *z*-score of 2.62. This is consistent with data that is normally distributed. Also, the mean is approximately half way between the maximum and minimum values, suggesting the data is fairly symmetrical. This data appears to be fairly normal.

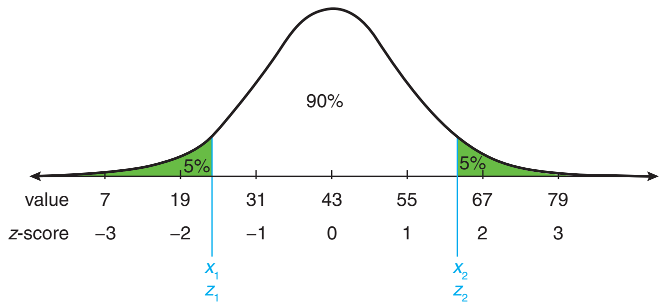
Note that no data set is exactly normal and that care needs to be taken when interpreting data as normal. For example, the normal model of the Douglas fir trees suggests there is a small probability that a tree will have a negative height since a *z*-score of –3 corresponds to a tree height of –4 ft. Clearly this is not possible and illustrates a difference between the normal curve and an actual distribution.

2. Suppose a set of data is normally distributed with a mean of 43 and a standard deviation of 12. Determine an upper limit and a lower limit that would encompass the **middle** 90% of the data.

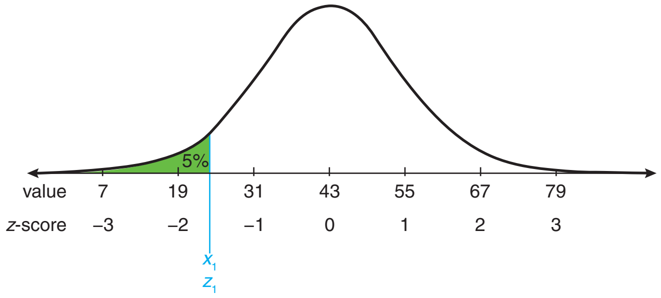
Draw a diagram to represent the problem.

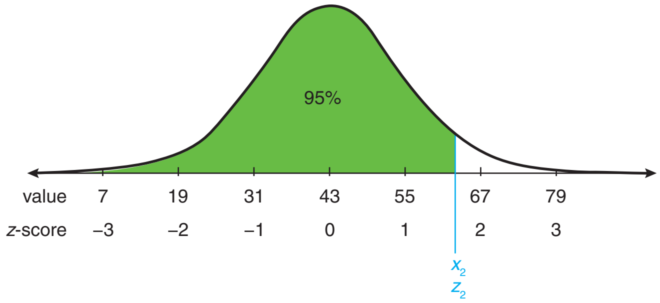


The goal is to determine values x_1 and x_2 such that 90% of the data lies between them. The question asks for the **middle** 90%, so 5% of the data should lie below x_1 and 5% of the data should lie above x_2.



In other words, x_1 has 5% of the data below it and x_2 has 95% of data below it.





Use technology or the *z*-score tables to determine the corresponding *z*-scores.

z	0.09	0.08	0.07	0.06	0.05	0.04	0.03	0.02	0.01	0.00
–1.7	0.0367	0.0375	0.0384	0.0392	0.0401	0.0409	0.0418	0.0427	0.0436	0.0446
–1.6	0.0455	0.0465	0.0475	0.0485	0.0495	0.0505	0.0516	0.0526	0.0537	0.0548
–1.5	0.0559	0.0571	0.0582	0.0594	0.0606	0.0618	0.0630	0.0643	0.0655	0.0668

The areas 0.0495 and 0.0505 both appear in the table and are equidistant from 0.05 or 5%. Using either z-score or averaging the two is reasonable.

$z_1 = -1.645$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633

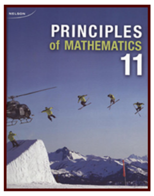
Similarly, $z_2 = 1.645$

These z-scores can now be used to determine the x-values.

$$\begin{aligned} z_1 &= \frac{x_1 - \mu}{\sigma} \\ -1.645 &= \frac{x_1 - 43}{12} \\ -19.74 &= x_1 - 43 \\ 23.26 &= x_1 \end{aligned}$$

$$\begin{aligned} z_2 &= \frac{x_2 - \mu}{\sigma} \\ 1.645 &= \frac{x_2 - 43}{12} \\ 19.74 &= x_2 - 43 \\ 62.74 &= x_2 \end{aligned}$$

The middle 90% of data will fall between data values of approximately 23.3 and 62.7.



For further information about *z*-scores, see pp. 283 – 291 of *Principles of Mathematics 11*.

Normal distributions contain data that is symmetrical, the data has an equal mean, median, and mode, and almost all of the data lies within three standard deviations of the mean. Real world data is never exactly normal, but there are many instances where the normal distribution is a useful model for making predictions. In the next lesson you will make further predictions using confidence levels, confidence intervals and margins of error.