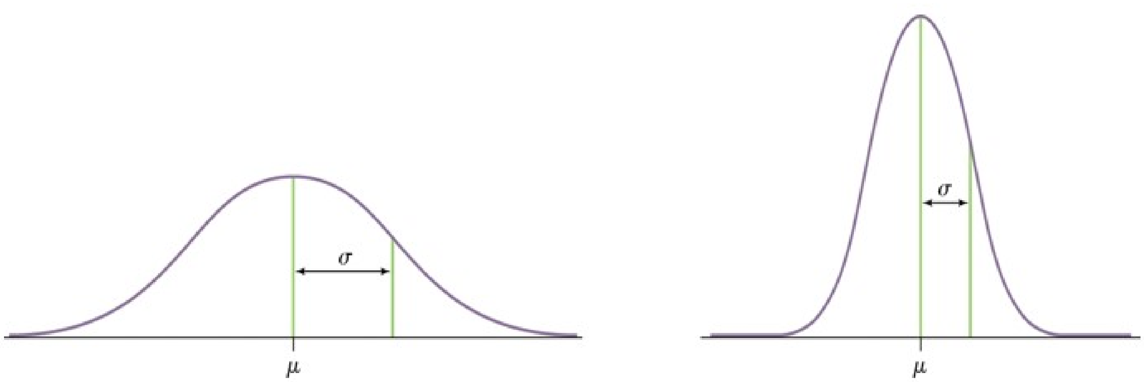
Ch 2.2 Density Curves and Normal Distributions

*The usual notation for the mean of a density curve is μ (the Greek letter mu). We write the standard deviation of a density curve as σ (the Greek letter sigma).

*All Normal curves have the same shape: symmetric, single-peaked, and bell-shaped

*Any specific Normal curve is completely described by giving its mean μ and its standard deviation σ.



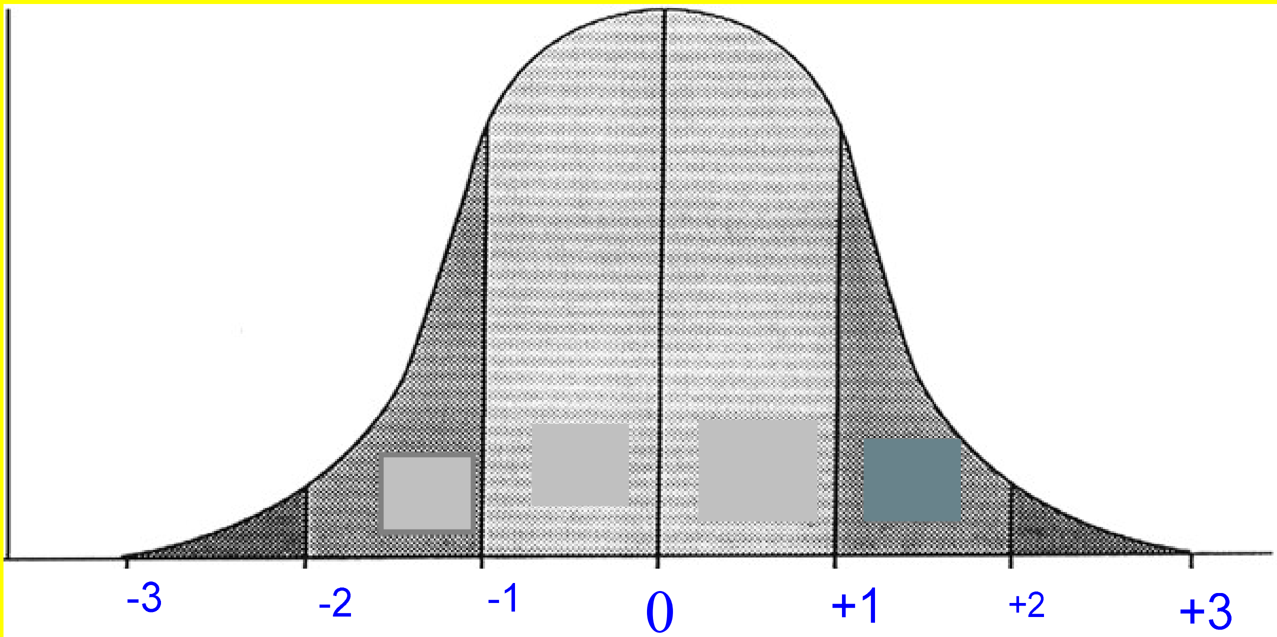
Two Normal curves, showing the mean μ and standard deviation σ.

*We abbreviate the Normal distribution with mean μ and standard deviation σ as $N(\mu,\sigma)$.

1. The total area under the curve must equal 1.

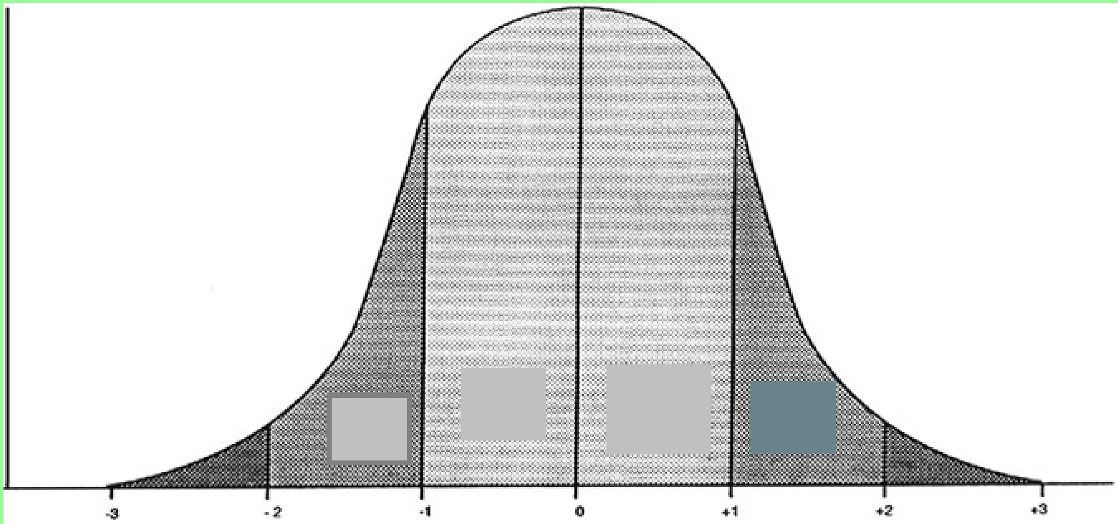
2. Every point on the curve must have a vertical height that is 0 or greater.

Normal Distribution



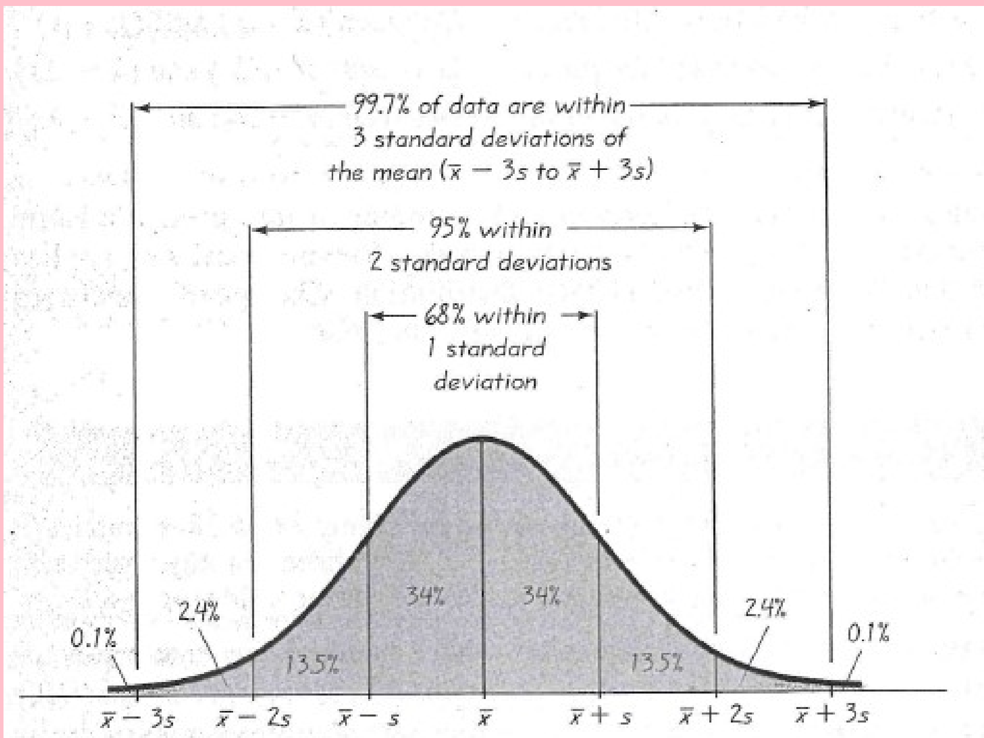
Ex1) mean = 80 and s = 2

+/- 1 s.d = 78 - 82
+/- 2 s.d. = 76 - 84
+/- 3 s.d. = 74 - 86



Try 1.)How would the data look if the mean was 75 and s = 5

<u>Empirical Rule</u> - if the data has a ***normal*** distribution or ***bell shaped curve*** the scores fall under:



Ex5) If men's height has a normal distribution with a mean = 69.0 and s = 2.8in, what percent of men have heights between 60.6 and 77.4

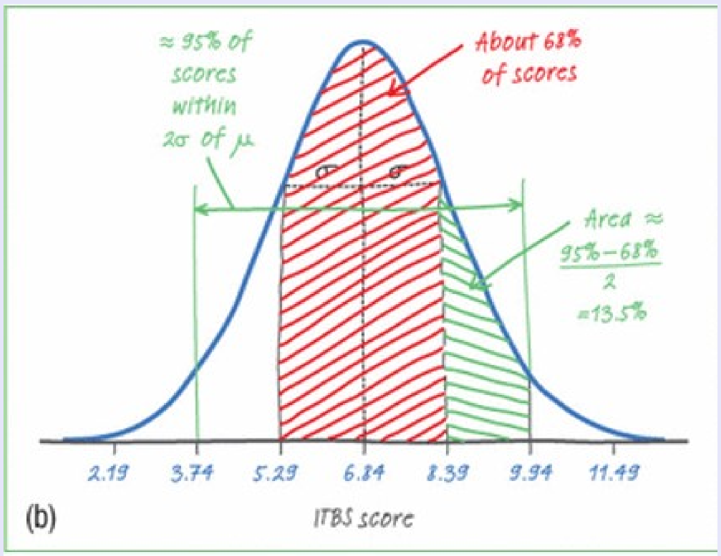
Pull

Example 1: The distribution of Iowa Test of Basic Skills (ITBS) vocabulary scores for 7th grade students in Gary, Indiana, is close to Normal. Suppose the distribution is $N(6.84, 1.55)$.
(mean, st.dev)

a) What percent of the ITBS vocabulary scores are less than 3.74? Show your work.

The other 5% *of scores are outside* this range. Because Normal distributions are symmetric, half of these scores are lower than 3.74 and half are higher than 9.94. That is, about 2.5% of the ITBS scores are below 3.74.

b) What percent of the scores are between 5.29 and 9.94? Show your work.



Let's start with a picture. Figure **(b)** shows the area under the Normal density curve between 5.29 and 9.94. We can see that about 68% + 13.5% = 81.5% of ITBS scores are between 5.29 and 9.94.

The Standard Normal Table
Because all Normal distributions are the same when we standardize, we can find areas under any Normal curve from a single table.

Page T1 and T2

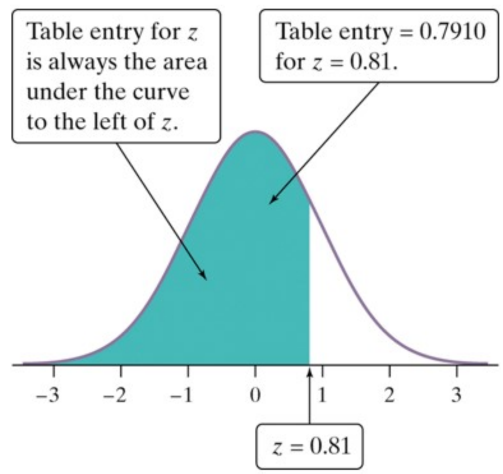
The **standard Normal Table (Table A)** is a table of areas under the standard Normal curve. The table entry for each value z is the area under the curve to the left of z.

A common student mistake is to look up a z-value in Table A and report the entry corresponding to that z-value, regardless of whether the problem asks for the area to the left or to the right of that z-value. To prevent making this mistake, always sketch the standard Normal curve, mark the z-value, and shade the area of interest. And before you finish, make sure your answer is reasonable in the context of the problem.

Suppose we want to find the proportion of observations from the standard Normal distribution that are less than 0.81.
We can use Table A:

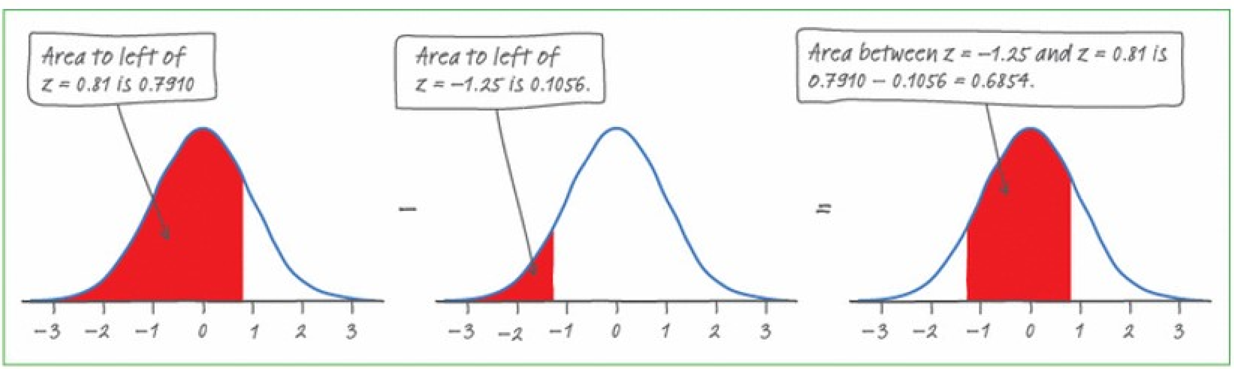
$P(z < 0.81) =$

Z	.00	.01	.02
0.7	.7580	.7611	.7642
0.8	.7881	.7910	.7939
0.9	.8159	.8186	.8212



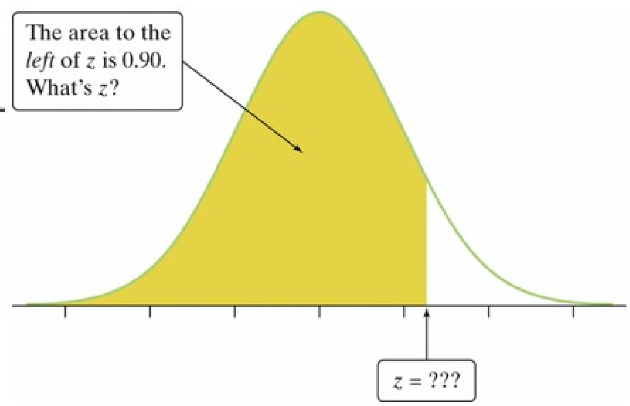
Example 2: Find the proportion of observations from the standard Normal distribution that are between −1.25 and 0.81.

From Table A, the area to the left of $z = 0.81$ is 0.7910 and the area to the left of $z = -1.25$ is 0.1056. So the area under the standard Normal curve between these two z-scores is $0.7910 - 0.1056 = 0.6854$.



Because Table A gives areas to the left of a specified z-score, all we need to do is find the value closest to 0.90 in the middle of the table. From the reproduced portion of Table A, you can see that the desired z-score is $z = 1.28$. That is, the area to the left of $z = 1.28$ is approximately 0.90.

z	.07	.08	.09
1.1	.8790	.8810	.8830
1.2	.8980	.8997	.9015
1.3	.9147	.9162	.9177



CHECK YOUR UNDERSTANDING

Use Table A in the back of the book to find the proportion of observations from a standard Normal distribution that fall in each of the following regions. In each case, sketch a standard Normal curve and shade the area representing the region.

1. $z < 1.39$

Show Answer

2. $z > -2.15$

Show Answer

3. $-0.56 < z < 1.81$

Show Answer

Use Table A to find the value z from the standard Normal distribution that satisfies each of the following conditions. In each case, sketch a standard Normal curve with your value of z marked on the axis.

4. The 20th percentile

Show Answer

5. 45% of all observations are greater than z

Show Answer

1. $z < 1.39$

Hide Answer

Correct Answer
The proportion is 0.9177.

2. $z > -2.15$

Hide Answer

Correct Answer
The proportion is 0.9842.

3. $-0.56 < z < 1.81$

Hide Answer

Correct Answer
The proportion is 0.9649 − 0.2877 = 0.6772.

Use **Table A** to find the value *z* from the standard Normal distribution that satisfies each of the following conditions. In each case, sketch a standard Normal curve with your value of *z* marked on the axis.

4. The 20th percentile

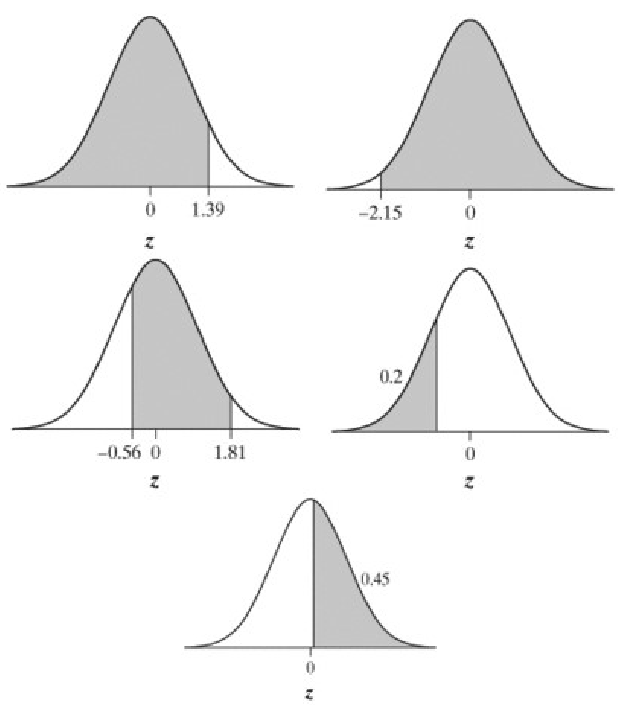
Hide Answer

Correct Answer
The *z*-score for the 20th percentile is $z = -0.84$.

5. 45% of all observations are greater than *z*

Hide Answer

Correct Answer
45% of the observations are greater than $z = 0.13$.



Day #2

We can answer a question about areas in *any* Normal distribution by standardizing and using Table A or by using technology.

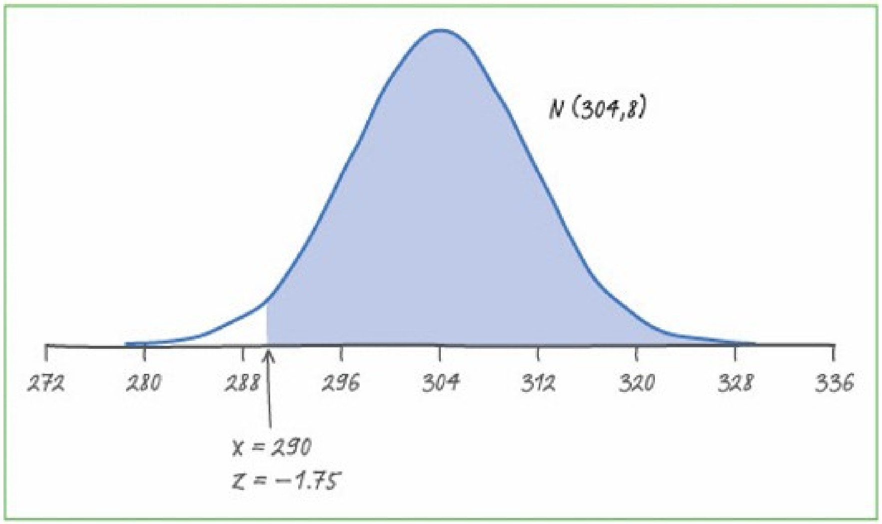
How To Find Areas In Any Normal Distribution

Step 1: State the distribution and the values of interest. Draw a Normal curve with the area of interest shaded and the mean, standard deviation, and boundary value(s) clearly identified.

Step 2: Perform calculations—show your work! Do one of the following: (i) Compute a *z*-score for each boundary value and use Table A or technology to find the desired area under the standard Normal curve; or (ii) use the normalcdf command and label each of the inputs.

Step 3: Answer the question.

Example 3: On the driving range, Tiger Woods practices his swing with a particular club by hitting many, many balls. Suppose that when Tiger hits his driver, the distance the ball travels follows a Normal distribution with mean 304 yards and standard deviation 8 yards. What percent of Tiger's drives travel at least 290 yards?



$$z = \frac{x-\mu}{\sigma} = \frac{290-304}{8} = -1.75$$

1 − 0.0401 = 0.9599. This is about 0.96,or 96%.

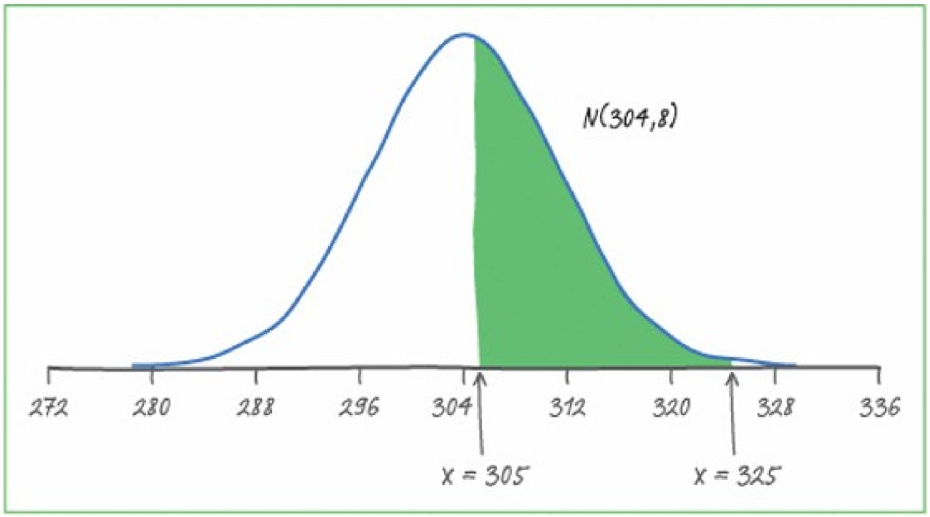
Step 3: Answer the question. About 96% of Tiger Woods's drives on the range travel at least 290 yards.

You can find the proportional area by using the graphing calculator.

- 2nd vars (distr)
- #2 normalcdf(lower limit bound, upperbound, mean, standard deviation)

Using technology: The command normalcdf(lower:2 90, upper:100000, μ:304, σ:8)also gives an area of 0.9599.

Example 4: What percent of Tiger's drives travel between 305 and 325 yards?



Step 2: Perform calculations–show your work! For the boundary value $x = 305$,

$$z = \frac{305 - 304}{8} = 0.13$$

The standardized score for $x = 325$ is

$$z = \frac{325 - 304}{8} = 2.63$$

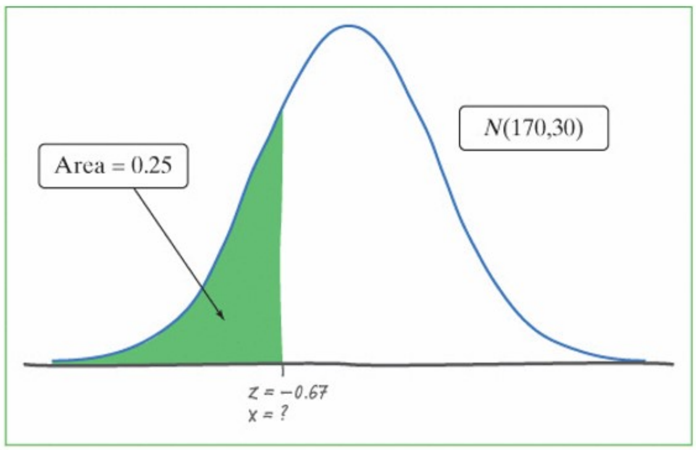
From Table A, we see that the area between $z = 0.13$ and $z = 2.63$ under the standard Normal curve is the area to the left of 2.63 minus the area to the left of 0.13. Look at the picture below to check this. From Table A, area between 0.13 and 2.63 = area to the left of 2.63 − area to the left of 0.13 = 0.9957 − 0.5517 = 0.4440.

Using technology: The command normalcdf(lower: 305, upper: 325, μ: 304, σ: 8) gives an area of 0.4459.

Step 3: Answer the question. About 45% of Tiger's drives travel between 305 and 325 yards.

Example 5: High levels of cholesterol in the blood increase the risk of heart disease. For 14-year-old boys, the distribution of blood cholesterol is approximately Normal with mean $\mu = 170$ milligrams of cholesterol per deciliter of blood (mg/dl) and standard deviation $\sigma = 30$ mg/dl. What is the first quartile of the distribution of blood cholesterol?

Step 1: State the distribution and the values of interest. The cholesterol level of 14-year-old boys follows a Normal distribution with $\mu = 170$ and $\sigma = 30$. The 1st quartile is the boundary value x with 25% of the distribution to its left. The figure below shows a picture of what we are trying to find.



Step 2: Perform calculations–show your work! Look in the body of Table A for the entry closest to 0.25. It's 0.2514. This is the entry corresponding to $z = -0.67$. So $z = -0.67$ is the standardized score with area 0.25 to its left. Now unstandardize. We know that the standardized score for the unknown cholesterol level x is $z = -0.67$.

So x satisfies the equation $\dfrac{x - 170}{30} = -0.67$.

Solving for x gives

$$x = 170 + (-0.67)(30) = 149.9$$

Step 3: Answer the question. The 1st quartile of blood cholesterol levels in 14-year-old boys is about 150 mg/dl.

TI -84:When given the area:

- 2nd vars (distr)
- #3 invNorm(area, mean, standard deviation)

Using technology: The command invNorm(area:0.25, μ:170, σ:30) gives $x = 149.8$.

CHECK YOUR UNDERSTANDING

Follow the method shown in the examples to answer each of the following questions. Use your calculator or the *Normal Curve* applet to check your answers.

1. Cholesterol levels above 240 mg/dl may require medical attention. What percent of 14-year-old boys have more than 240 mg/dl of cholesterol?

Show Answer

2. People with cholesterol levels between 200 and 240 mg/dl are at considerable risk for heart disease. What percent of 14-year-old boys have blood cholesterol between 200 and 240 mg/dl?

Show Answer

3. What distance would a ball have to travel to be at the 80th percentile of Tiger Woods's drive lengths?

Show Answer

Follow the method shown in the examples to answer each of the following questions. Use your calculator or the *Normal Curve* applet to check your answers.

1. Cholesterol levels above 240 mg/dl may require medical attention. What percent of 14-year-old boys have more than 240 mg/dl of cholesterol?

Hide Answer

Correct Answer
For 14-year-old boys, the amount of cholesterol follows a N(170, 30) distribution and we want to find the percent of boys with cholesterol of more than 240 (see graph below). $z = \dfrac{240 - 170}{30} = 2.33.$ From **Table A**, the proportion of *z*-scores above 2.33 is 1 – 0.9901 = 0.0099. *Using technology*: normalcdf (lower:240 upper:1000,μ:170,σ:30) = 0.0098. About 1% of 14-year-old boys have cholesterol above 240 mg/dl.

2. People with cholesterol levels between 200 and 240 mg/dl are at considerable risk for heart disease. What percent of 14-year-old boys have blood cholesterol between 200 and 240 mg/dl?

Hide Answer

Correct Answer
For 14-year-old boys, the amount of cholesterol follows a *N*(170, 30) distribution and we want to find the percent of boys with cholesterol between 200 and 240 (see graph below). $z = \dfrac{200 - 170}{30} = 1$ and $z = \dfrac{240 - 170}{30} = 2.33.$ From **Table A**, the proportion of *z*-scores between 1 and 2.33 is 0.9901 – 0.8413 = 0.1488. *Using technology*: normalcdf(lower:2 00,upper: 240,μ:170,σ:30) = 0.1488. About 15% of 14-year-old boys have cholesterol between 200 and 240 mg/dl.

3. What distance would a ball have to travel to be at the 80th percentile of Tiger Woods's drive lengths?

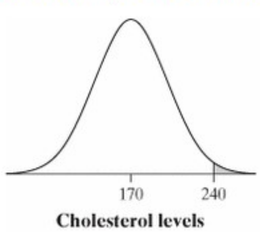
Hide Answer

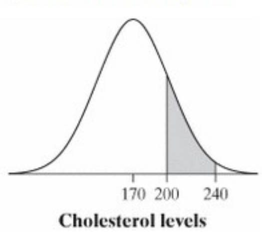
Correct Answer
For 14-year-old boys, the amount of cholesterol follows a *N*(170, 30) distribution and we want to find the percent of boys with cholesterol between 200 and 240 (see graph below). $z = \dfrac{200 - 170}{30} = 1$ and $z = \dfrac{240 - 170}{30} = 2.33.$ From **Table A**, the proportion of *z*-scores between 1 and 2.33 is 0.9901 – 0.8413 = 0.1488. *Using technology*: normalcdf(lower:2 00,upper: 240,μ:170,σ:30) = 0.1488. About 15% of 14-year-old boys have cholesterol between 200 and 240 mg/dl.

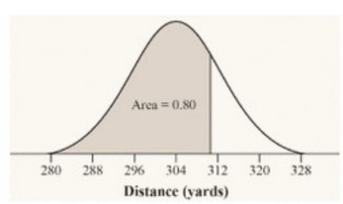
3. What distance would a ball have to travel to be at the 80th percentile of Tiger Woods's drive lengths?

Hide Answer

Correct Answer
For Tiger Woods, the distance his drives travel follows an *N* (304, 8) distribution and the 80th percentile is the boundary value *x* with 80% of the distribution to its left (see graph below). A *z*-score of 0.84 gives the area closest to 0.80 (0.7995). Solving $0.84 = \dfrac{x - 304}{8}$ gives *x* = 310.7. *Using technology*: invNorm(area:0.8,μ:304,σ:8) = 310.7. The 80th percentile of Tiger Woods's drive lengths is about 310.7 yards.







Assessing Normality Day # 3

The Normal distributions provide good models for some distributions of real data. Many statistical inference procedures are based on the assumption that the population is approximately Normally distributed. Consequently, we need a strategy for assessing Normality.

Plot the data.

*Make a dotplot, stemplot, or histogram and see if the graph is approximately symmetric and bell-shaped.

Check whether the data follow the 68-95-99.7 rule.

*Count how many observations fall within one, two, and three standard deviations of the mean and check to see if these percents are close to the 68%, 95%, and 99.7% targets for a Normal distribution.

Example 16 Unemployment in the States

Are the data close to Normal?

Let's start by examining data on unemployment rates in the 50 states. Here are the data arranged from lowest (North Dakota's 4.1%) to highest (Michigan's 14.7%).[10]

4.1	4.5	5.0	6.3	6.3	6.4	6.4	6.6	6.7	6.7	6.7	6.9	7.0
7.0	7.2	7.4	7.4	7.4	7.8	8.0	8.0	8.2	8.2	8.4	8.5	8.5
8.6	8.7	8.8	8.9	9.1	9.2	9.5	9.6	9.6	9.7	10.2	10.3	10.5
10.6	10.6	10.8	10.9	11.1	11.5	12.3	12.3	12.3	12.7	14.7		

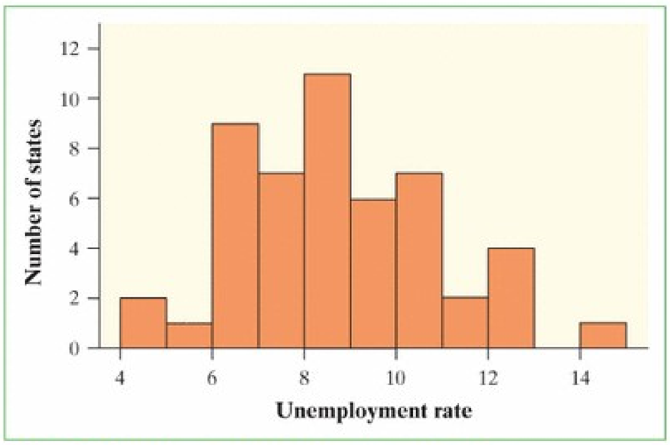


Figure 2.23 Histogram of state unemployment rates.

- *Check whether the data follow the 68–95–99.7 rule.*

We entered the unemployment rates into computer software and requested summary statistics. Here's what we got:

Mean = 8.682 Standard deviation = 2.225.

Now we can count the number of observations within one, two, and three standard deviations of the mean.

Mean ± 1 SD:	6.457 to 10.907	36 out of 50 = 72%
Mean ± 2 SD:	4.232 to 13.132	48 out of 50 = 96%
Mean ± 3 SD:	2.007 to 15.357	50 out of 50 = 100%

These percents are quite close to the 68%, 95%, and 99.7% targets for a Normal distribution.

Normal Probability Plots

Most software packages can construct Normal probability plots. These plots are constructed by plotting each observation in a data set against its corresponding percentile's *z*-score.

AP EXAM TIP: Normal probability plots are not included on the AP Statistics course outline. However, these graphs are very useful tools for assessing Normality. You may use them on the AP exam if you wish—just be sure that you know what you're looking for (linear pattern).

A **Normal probability plot** provides a good assessment of whether a data set follows a Normal distribution.

Interpreting Normal Probability Plots

If the points on a Normal probability plot lie close to a straight line, the plot indicates that the data are Normal.

Systematic deviations from a straight line indicate a non-Normal distribution.

Outliers appear as points that are far away from the overall pattern of the plot.

On the graphing calc:

- Find the z scores for each set of data
- (L1 - mean)/standard deviation
- store L2
- 2nd y =
- select the last graph
- zoom #9

Mean = 8.682 Standard deviation = 2.225.

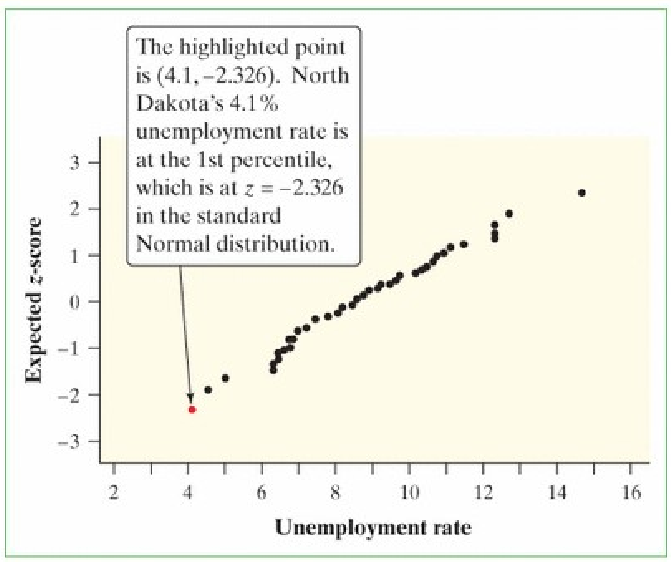


Figure 2.24 Normal probability plot of the percent of unemployed individuals in each of the 50 states.

How can we determine shape from a Normal probability plot?
Look at the Normal probability plot of the guinea pig survival data. Imagine drawing a line through the leftmost points, which correspond to the smaller observations. The larger observations fall systematically to the right of this line. That is, the right-of-center observations have much larger values than expected based on their percentiles and the corresponding *z*-scores from the standard Normal distribution. This Normal probability plot indicates that the guinea pig survival data are strongly right-skewed. *In a right-skewed distribution, the largest observations fall distinctly to the right of a line drawn through the main body of points*. Similarly, left-skewness is evident when the smallest observations fall to the left the line.