1. Suppose a data set is represented by a normal distribution with a mean of 125 and a standard deviation of 7.

(a) What data value is 2 standard deviations above the mean? What is the z-score for this data value? Use the formula, and show your work!

To get the data value 2 standard deviations above the mean, we add $2\sigma$ to the mean, $\mu$, and get $\mu + 2\sigma = 125 + 2 \cdot 7 = 139$. Now, we'll compute the z-score of this data point:

$$z = \frac{x - \mu}{\sigma} = \frac{139 - 125}{7} = \frac{14}{7} = 2$$

(b) What data value is $\frac{1}{5}$ of a standard deviation below the mean? What is the z-score for this value? Use the formula, and show your work!

To get the data value $\frac{1}{5}$ of a standard deviations above the mean, we subtract $\frac{1}{5}\sigma$ from the mean, $\mu$, and get $\mu - \frac{1}{5}\sigma = 125 - \frac{1}{5} \cdot 7 = 123.6$. Now, we'll compute the z-score of this data point:

$$z = \frac{x - \mu}{\sigma} = \frac{123.6 - 125}{7} = \frac{-1.4}{7} = -0.2 = -\frac{1}{5}$$

2. You'll make some observations about z-scores in this problem.

(a) How do the z-scores from Problem 1 correspond to the standard deviations given in the problem? What does this tell you about the relationship between z-scores for a normal distribution and standard deviation?

The z-score indicates how many standard deviations a value is above the mean. Hence, if the z-score being negative indicates that the value is that many standard deviations below the mean.

(b) Suppose a normal distribution has mean 10 and standard deviation 2. Find the z-scores of the measurements 9, 10, 11, 14, and 17 using your observation from the previous problem. Do not use the formula!

The value 9 is $1 = \frac{1}{2} \cdot 2 = \frac{1}{2}\sigma$ below the mean. Hence, $z = -\frac{1}{2}$.

The value 10 is the mean. Hence, $z = 0$.

The value 11 is $1 = \frac{1}{2} \cdot 2 = \frac{1}{2}\sigma$ above the mean. Hence $z = \frac{1}{2}$.

The value 14 is $4 = 2 \cdot 2 = 2\sigma$ above the mean. Hence $z = 2$.

The value 17 is $7 = 3.5 \cdot 2 = 3.5\sigma$ above the mean. Hence $z = 3.5$.

3. In a normally distributed data set, find the value of the standard deviation if the following additional information is given.

(a) The mean is 226.2 and the z-score for a data value of 230 is 0.5.

The formula gives $z = \frac{x - \mu}{\sigma}$. For this data, we have $0.5 = \frac{230 - 226.2}{\sigma}$. Multiplying both sides by $\sigma$ and dividing by 0.5 gives

$$\sigma = \frac{230 - 226.2}{0.5} = \frac{3.8}{0.5} = 7.6.$$

(b) The mean is 9.8 and $Q_3$ is 10.61.

The formula here gives $Q_3 = \mu + 0.675\sigma$. For this data we have $10.61 = 9.8 + 0.675\sigma$. To solve for $\sigma$ we subtract 9.8 from both sides, then divide by 0.675. This then gives $\sigma = 1.2$.

(c) The inflection points of the corresponding bell curve are 29 and 54.

The inflection points of a normal distribution are located at $\mu - \sigma$ and $\mu + \sigma$. Hence, our points are equidistant from the mean, and we can take their average to find $\mu$. So,

$$\mu = \frac{29 + 54}{2} = \frac{83}{2} = 41.5.$$
4. For each of the following situations, write out what X represents, and find the 95% and 99.7% confidence intervals for the value \( \mu \) that you find for X.

(a) Of 900 randomly selected people who are surveyed, 324 respond “yes” to the survey question.
Here, \( X \) represents the number of people who say “yes” to the survey question out of 900 people surveyed. We then do the following computations:

\[
\mu = \frac{324}{900} = 0.36 \rightarrow 36\%; \sigma = \sqrt{900 \cdot 0.36 \cdot 0.64} = 14.4
\]

The standard error is \( \frac{14.4}{900} = 0.016 \rightarrow 1.6\% \). So we are 95\% confident that the actual percentage of people who would say “yes” is between 36\% - 2 \cdot 1.6\% = 32.8\% and 36\% + 2 \cdot 1.6\% = 39\%.
Similarly, we are 99.7\% confident that the actual percentage of people who would say “yes” is between 36\% - 3 \cdot 1.6\% = 31.5\% and 36\% + 3 \cdot 1.6\% = 40.8\%.

(b) Of 1500 randomly selected candy bars that are tested coming off the production line, 900 are at least 0.2 grams above the weight indicated on the packaging.
Here, \( X \) represents the number of candy bars that are at least 0.2 grams above the weight indicated on the packaging out of 1500 candy bars. We then do the following computations:

\[
\mu = \frac{900}{1500} = 0.6 \rightarrow 60\%; \sigma = \sqrt{1500 \cdot 0.60 \cdot 0.40} = 19
\]

The standard error is \( \frac{19}{1500} = 0.013 \rightarrow 1.3\% \). So we are 95\% confident that the actual percentage of candy bars that weigh at least 0.2 grams above the weight indicated on the packaging is between 60\% - 2 \cdot 1.3\% = 57.4\% and 60\% + 2 \cdot 1.3\% = 62.6\%.
Similarly, we are 99.7\% confident that the actual percentage of heavier candy bars is between 60\% - 3 \cdot 1.3\% = 56.1\% and 60\% + 3 \cdot 1.3\% = 63.9\%.

(c) Of 2412 randomly selected high school students who are surveyed, 1624 say at least one of their current high school teachers has met their parents.
Here, \( X \) represents the number of high school students that say at least one of their teachers has met their parents out of 2412 students. We then do the following computations:

\[
\mu = \frac{1624}{2412} = 0.67 \rightarrow 67\%; \sigma = \sqrt{2412 \cdot 0.67 \cdot 0.33} = 23
\]

The standard error is \( \frac{23}{2412} = 0.01 \rightarrow 1\% \). So we are 95\% confident that the actual percentage of students who have a teacher that have met their parents is between 67\% - 2 \cdot 1\% = 65\% and 67\% + 2 \cdot 1\% = 69\%.
Similarly, we are 99.7\% confident that the actual percentage of students whose teachers have met their parents is between 67\% - 3 \cdot 1\% = 64\% and 67\% + 3 \cdot 1\% = 70\%.

(d) A consumer testing agency tracks 325 randomly selected television sets of a certain brand and finds that 52 of them develop performance defects in the first year.
Here, \( X \) represents the number of television sets which will develop performance defects in the first year out of 325 students. We then do the following computations:

\[
\mu = \frac{52}{325} = 0.16 \rightarrow 16\%; \sigma = \sqrt{325 \cdot 0.16 \cdot 0.84} = 7
\]

The standard error is \( \frac{7}{325} = 0.022 \rightarrow 2.2\% \). So we are 95\% confident that the actual percentage of defective televisions is between 16\% - 2 \cdot 2.2\% = 11.6\% and 16\% + 2 \cdot 2.2\% = 20.4\%.
Similarly, we are 99.7\% confident that the actual percentage of defective televisions is between 16\% - 3 \cdot 2.2\% = 9.4\% and 16\% + 3 \cdot 2.2\% = 22.6\%.

(e) A radio station conducts a poll of randomly selected adults in its viewing area and finds that 1074 out of 1224 cannot recall any commercials that they heard running on that station in the last month.
Here, \( X \) represents the number of people who cannot recall commercials from the radio station out of 1224 people. We then do the following computations:

\[
\mu = \frac{1074}{1224} = 0.88 \rightarrow 88\%; \sigma = \sqrt{1224 \cdot 0.88 \cdot 0.12} = 11
\]

The standard error is \( \frac{11}{1224} = 0.009 \rightarrow 0.9\% \). So we are 95\% confident that the actual percentage of people who cannot recall any commercials is between 88\% - 2 \cdot 0.9\% = 86.2\% and 88\% + 2 \cdot 0.9\% = 89.8\%.
Similarly, we are 99.7\% confident that the actual percentage of people who cannot recall any commercials is between 88\% - 3 \cdot 0.9\% = 85.3\% and 88\% + 3 \cdot 0.9\% = 90.7\%.
5. Suppose a snack food manufacturer claims their boxes of crackers are filled to a mean weight of 1.3 pounds with a standard deviation of 0.2 pounds. In an ad campaign, they promise to reimburse customers if the actual weight of the box of crackers is less than 1 pound. If a box of crackers costs $2.50 and there are 1 million boxes sold in a year, what is the expected cost of such a program to the manufacturer?

This problem requires understanding what happens at one and a half standard deviations below the mean. We could look that information up, but we don’t have enough of a table of such information in our book. Oops.

So I’m going to change the problem slightly, and do the slightly different problem. Let’s assume instead that the mean weight is 1.4 pounds, but the other numbers remain the same.

The value of 1 pound is 2 standard deviations below the mean. We know 95% of the boxes are between 1 and 1.8 pounds (by the 68–95–99.7 rule). By the fact that the curve is symmetric, and the fact that 5% of the boxes are less than 1 or more than 1.8 pounds, then we know that 2.5% of the boxes weigh less than 1 pound. That’s 2.5% of 1 million, or

$$0.25 \cdot 1,000,000 = 250,000$$

boxes each year. At $2.50 per box, that adds up to 250,000 \cdot $2.50 = $625,000.