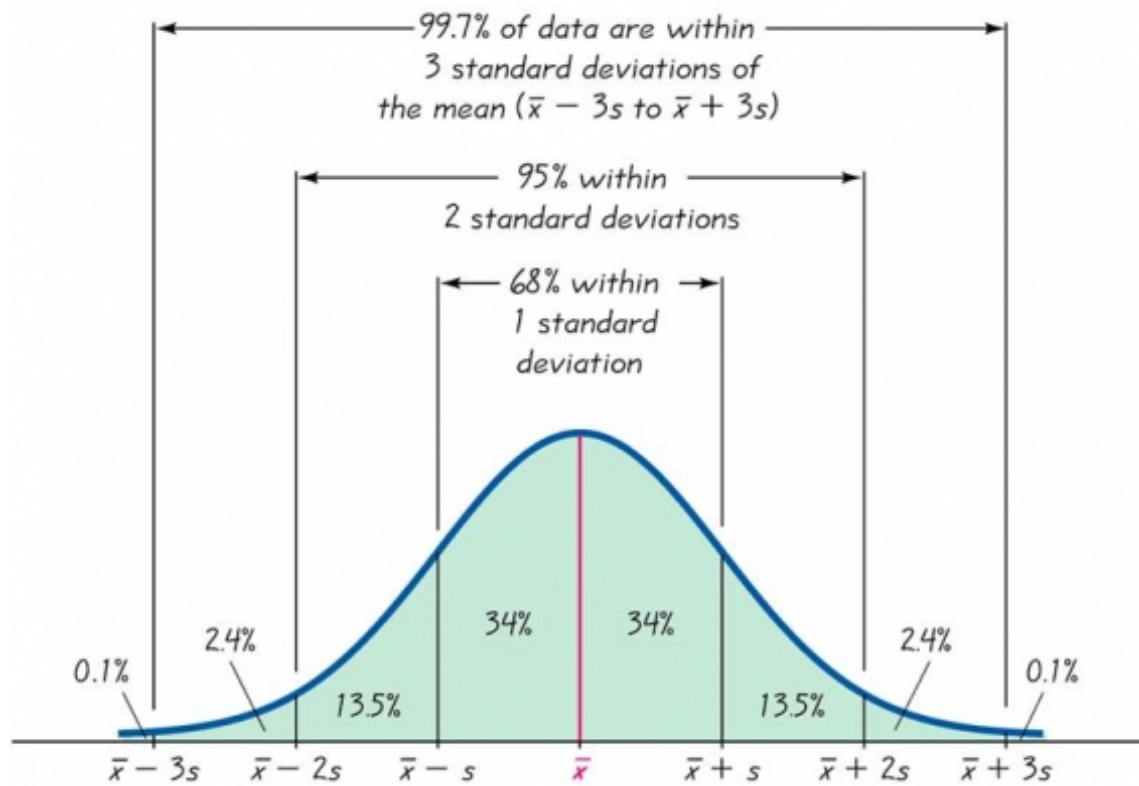


Problem set 8-1 and 8-2

1. Open up “National SAT Scores (CODAP)” and make a dot plot of the data.
 - a) What are the mean and standard deviation of the data?
 - b) Give the SAT scores that are one standard deviation above and one standard deviation below the mean.
 - c) What proportion of the data lies within one standard deviation of the mean? (Hint: put the scores in ascending order.)
 - d) Give the SAT scores that are two standard deviations above and two standard deviations below the mean.
 - e) What proportion of the data lies within two standard deviations of the mean?
 - f) Give the SAT scores that are three standard deviations above and three standard deviations below the mean.
 - g) What proportion of the data lies within three standard deviations of the mean?

The dot plots for both data sets in problem set 8-1 both had the same bell-shaped distribution. \wedge In addition, the answer to “What proportion of the data lies within one, two, or three standard deviations of the mean?” was about the same for both data sets. This is not a coincidence. That is because both data sets were approximately normally distributed. A normally distributed dot plot looks somewhat like a bell and colloquially is called the bell-curve. “Normally distributed” data conform to what statisticians call the Empirical Rule or the 68%-95%-99.7% Rule. That is, about 68% of the data lie within one standard deviation of the mean, about 95% of the data lie within two standard deviations of the mean, and about 99.7% of the data lie within three standard deviations of the mean.



In statistics, normalization of data means adjusting values measured on different scales to a common scale, often prior to averaging.

2. Open up again “National SAT Scores (CODAP)”.

Create a new attribute (column) called **zscore**.

Each case (row) will be the associated z-score if you make the formula of this new attribute

$$\frac{(\text{SATscore} - \text{mean}(\text{SATscore}))}{\text{stdDev}(\text{SATscore})}$$

- a) What is the mean of “zscore”?

- b) What is the standard deviation of “zscore”?

- c) Recall the SATscores that were ± 1 standard deviations from the mean from #1. What are the z-scores at these cases?

- d) Recall the SATscores that were ± 2 standard deviations from the mean from #1. What are the z-scores at these cases?

- e) Recall the SATscores that were ± 3 standard deviations from the mean from #1. What are the z-scores at these cases?

