

Unit Vocabulary

- **Statistical Question** - A question that anticipates variability in the data that would be collected in order to answer the question.
- **Variability** - Variability in a data set occurs when the observations in the data set are not all the same.
- **Dot Plot** - A plot of numerical data along a number line.
- **Histogram** - A graphical representation of a numerical data set that has been grouped into intervals. Each interval is represented by a bar drawn above that interval that has a height corresponding to the number of observations in that interval.
- **Relative Frequency Histograms** - A histogram where relative frequency, rather than frequency, is used for the vertical scale.
Relative frequency is the value of the frequency in an interval divided by the total number of data values.

Measures of Center/Central Tendency

- **Median** - A measure of center appropriate for skewed data distributions. It is the middle value when the data are ordered from smallest to largest if there are an odd number of observations and half way between the middle two observations if the number of observations is even.
- **Mean** - A measure of center appropriate for data distributions that are approximately symmetric. It is the average of the values in the data set. Two common interpretations of the mean are as a "fair share" and as the balance point of the data distribution.
- **Mode** - A measure of center that describes the value that appears most often in a set of data.
- **Deviations from the Mean** - The differences calculated by subtracting the mean from the observations in a data set.
- **Mean Absolute Deviation (MAD)** - A measure of variability appropriate for data distributions that are approximately symmetric. It is the average of the absolute value of the deviations from the mean.
- **Box Plot** - A graph of five numerical summary measures: the minimum, lower quartile, median, upper quartile, and the maximum. It conveys information about center and variability in a data set.
- **Inter-quartile Range (IQR)** - A measure of variability appropriate for data distributions that are skewed. It is the difference between the upper quartile and the lower quartile of a data set and describes how spread out the middle 50% of the data is.