**Unit 2 Vocabulary: One Variable Statistics**

| **Concept/Vocabulary Word** | **Definition** |
| --- | --- |
| categorical data | Data that records qualities or characteristics of an individual, such as gender or eye color. (Also call qualitative data) |
| data | A collection of information in context. |
| dot plot | A graph of quantitative data using dots placed above a simple number line. If a data value occurs more than once, the dots are stacked on top of each other over that value on the number line. |
| frequency distribution | A table that summarizes the distribution of quantitative data by dividing the data into intervals (also called classes) and counting the number of data values within each interval. This count is called the frequency. |
| histogram | A graph of a frequency distribution of quantitative data where intervals of data are represented with bars. The widths of the bars are determined by the length of the intervals into which the data has been divided and the heights of the bars are proportional to the class frequencies. |
| individual | A member of a population. |
| measures of center | Numerical measures that describe the typical value of a quantitative data set. In this unit, we will be studying the mean and the median. |
| measures of spread | Numerical measures that describe how much values typically vary from the center in a quantitative data set. In this unit, we will be studying interquartile range and standard deviation. |
| population | In statistics, a set of individuals that we wish to describe and/or make predictions about, such as group of people, a row of plants, a set of batteries, a group of rats, or bacteria placed into test tubes. |
| quantitative data | Data that measures a characteristic of an individual, such as height, weight, or age. (Also called measurement data) |
| variable | Characteristic recorded about each individual in a data set. Examples: age, time since last online purchase, zip code, gender, height, hair color, etc. |
| mean | A numerical measure of center that is the arithmetic average of the data. |
| mean absolute deviation | A numerical measure of spread that shows how much data values vary from the mean for a quantitative data set. A low mean absolute deviation indicates that the data points tend to be very close to the [mean](http://en.wikipedia.org/wiki/Mean), whereas a high mean absolute deviation indicates that the data points are spread out over a large range of values. The process of calculating the mean absolute deviation involves taking the absolute value of the deviations from the mean. |
| median | A numerical measure of center that describes the middle value of a data set. Note that the median does not have to be one of the values in the data set, but a value that divides the data set in half so that 50% of the data values lie above the median and 50% of the data values fall below the median. |
| standard deviation | A numerical measure of spread that shows how much data values vary from the mean for a quantitative data set. A low standard deviation indicates that the data points tend to be very close to the [mean](http://en.wikipedia.org/wiki/Mean), whereas a high standard deviation indicates that the data points are spread out over a large range of values. The process of calculating the standard deviation involves squaring the deviations from the mean. |
| box plot | Also called a box-and-whisker plot. A graph of quantitative data built using the five number summary of the data. The graph is made by making vertical lines above the lower quartile, median, and upper quartile on a number line and connecting these vertical lines with horizontal segments to form the box. The minimum and maximum values are represented by dots and connected by segments, or "whiskers", to the box. Box plots can be drawn either horizontally or vertically. |
| interquartile range | A measure of the spread of the middle 50% of a set of quantative data; the difference between the upper and lower quartiles. IQR = Q3 − Q1 |
| modified box plot | A box plot that indicates which data values, if any, are outliers by representing them as dots separate from the box plot. The whisker(s) connect the box to the lowest and/or highest data values that are not outliers, instead of the minimum and/or maximum values. |
| outlier | A data value that does not fit the overall pattern of the data distribution. In the case of one-variable data, an outlier is a value that is more than 1.5 IQR above the third quartile or below the first quartile. |