Extract from csv or other sources, Format as table, data cleanup, formatting (dates, currency, etc.), scatter lines (inside a column), filtering data using either column header filters or by slicer, various BI charts (scatter, line, column, pie), statistics charts (histogram, box and whisker chart), pivot table (along with pivot charts) used to create multidimensional data summaries

Difference between population(total dataset) vs sample.

Mean: average (sum of all the observations divided by total number of observations). μ = ( Σ Xi ) / N

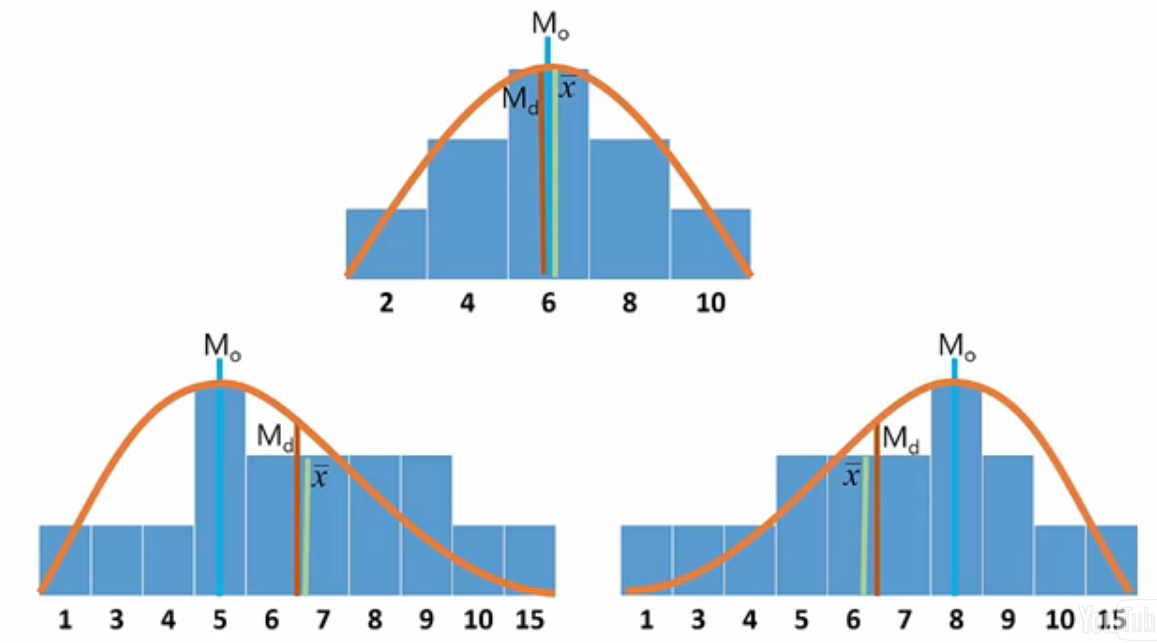
Median: middle value (after sorting the values). If the total number of observations is even, then we get 2 middle values. The average of those 2 values is median

Mode: most frequent (easily seen if you do a histogram)

A normal distribution is one in which mean, median and mode are same.

If mode and median are less than mean, then it is called right skewed distribution.

If mode and median are more than mean, then it is called left skewed distribution.



Range: max observation value – min observation value

Variance: sum of squares of each individual observation value subtracted from mean divided by number of observations.

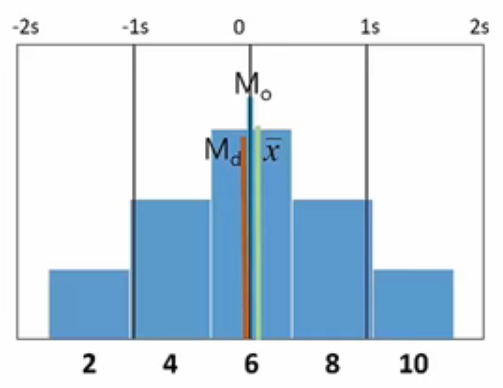
Variance for population: σ2 = Σ ( Xi - μ )2 / N where N is the population size

Variance for population: [s2 = Σ ( xi - x )2 / ( n - 1 )](http://stattrek.com/AP-Statistics-1/Variability.aspx?Tutorial=stat) where n is the sample size and x is the sample mean

Here we are dividing my n-1 to correct some bias that might be inherent in the sample.

Standard deviation: σ = sqrt(σ2) and s = sqrt( [s2](http://stattrek.com/AP-Statistics-1/Variability.aspx?Tutorial=stat) )

For normal distribution, the percentage of data that falls between +1 and -1 of standard deviation is 68.2% , that falls between +2 and -2 is 95.4% and that between +3 and -3 is 99.7%



Standard Error: SE = s/sqrt(n) and that is because we are dealing with samples and not the complete dataset

Descriptive statistics: use the data analysis add-in and select the columns you want stats on. Kurtosis is an estimate of normality of the data. The closer it is to 0, the more likely the data is normally distributed. Positive Skewness means data is right skewed and negative skewness means data is left skewed

Correlation: +1 to -1. Again use the data analysis add-in. Pick correlation. Highlight all the columns in which you are interested in finding a correlation. Correlation is not causation.

Hypothesis: null hypothesis, alternate hypothesis

One-sample t-test or z-test: