Data descriptions

Statistic: Characteristic or measure obtained from a Sample.

Parameter: Characteristic or measure obtained from a Population.

Mean: The sum of all values divided by number of values. This can be either a Population Mean (denoted by mu) or a Sample mean (denoted by x bar).

Median: The midpoint of the data after it is ranked (sorted in ascending order). There are as many numbers below the median as above the median.

Mode: The most frequent number.

Skewed Distribution: The majority of the values lie together on one side with very few values (the tail) to the other side. In positively skewed distribution, the tail lies on the right side with the *median* is lesser than the *mean*. In negatively skewed distribution, the tail lies on the left side with the *median* is greater than the *mean*.

Symmetric Distribution: The data values are evenly distributed on both sides of the mean. In this, *mean* and *median* are same.

Weighted Mean: The mean when each value is multiplied by its weight and summed. This sum is divided by the total of the weights.

Midrange: The mean of the highest and the lowest values. (Max + Min)/2

Range: The difference between the highest and the lowest values. (Max - Min)

Population variance: The average of the squares of the distances from the population mean. The units on the variance are the units of the population squared.

Sample variance: Unbiased estimator of the population variance. The sum of the squares of the deviations from the sample mean is divided by one less than the sample size. The units on the variance are the units of the population squared.

Standard deviation:

- The square root of the variance.
- The *population standard deviation* is the square root of the population variance.
- The *sample standard deviation* is the square root of the sample variance.
- The sample standard deviation is not unbiased estimator of the population standard deviation.
- The units on the standard deviation is same as of the population/sample.

Empirical or Normal Rule: Only valid when a distribution in bell-shaped(normal distribution). Approximately,

- 68% lies within 1 standard deviation of the mean

- 95% lies within 2 standard deviations;
- 99.7% lies within 3 standard deviations

Standard Score or Z-score: The value obtained by subtracting the mean and divided by the standard deviation. When all values are transformed to their Standard Scores, the mean is 0 and the Standard deviation is 1.

Percentile: The percent of the population which lies below that value. The data must be ranked to find percentiles.

Quartile: Either the 25th, 50th, or 75th percentiles. The 50th percentile is also called as the *Median*.

Decile: Either the 10th, 20th, ..., or 90th percentiles.

Lower Hinge: The median of the lower half of the numbers (up to and including the median).

Upper Hinge: The median of the upper half of the numbers (including the median).

Box plot: A graphical representation of the *minimum value*, *lower hinge*, *median*, *upper hinge*, and, *maximum value*.

Five Number Summary: *minimum value*, *lower hinge*, *median*, *upper hinge*, and, *maximum value*

InterQuartile Range: The difference between the 3rd and 1st quartiles.

Outlier: An extremely high or low values compared to the rest of the values.

Mild outliers: Values which lie between 1.5 and 3.0 times IQR below the 1st quartile or above the 3rd quartile.

Extreme outliers: Values which lie more than the 3.0 times IQR below the 1st quartile or above the 3rd quartile.