



Statistics Assignment 2

1. How can we figure out what the interquartile range is?

**Answer:** This is statistical dispersion which measures the spread in the middle half of your data. In this measurement dividing your data into four quarters from low to high as below

First Quartile 25% - Q1, Second Quartile 50% - Q2, Third Quartile 75% - Q3 and Upper Quartile 100% - Q4.

To calculate IQR = Q3-Q1

1. What exactly is the value of the 5-number theory?

**Answer:** This theory helps us to understand the center of our data in a data set, as well as how spread out the data points are. We use a five-number summary consisting of the following.

Minimum – This is the smallest value in our data set.

First Quartile – This number is denoted *Q*1 and 25% of our data falls below the first quartile.

Median – This is the midway point of the data. 50% of all data falls below the median.

Third Quartile – This number is denoted *Q*3 and 75% of our data falls below the third quartile.

Maximum – This is the largest value in our data set.

1. What is the relationship between standard deviation and variance?

**Answer:**

The variance is a measure of how far a set of data is dispersed out from their mean or average value. It is denoted as ‘σ2

Formula:

The spread of statistical data is measured by the standard deviation. Distribution measures the deviation of data from its mean position. The degree of dispersion is computed by the method of estimating the deviation of data points. It is denoted by the symbol, ‘σ’.

Formula:

Relation Between Variance and SD:

The standard deviation is a square root of variance. Both measures exhibit variability in distribution, but their units vary: Standard deviation is expressed in the same units as the original values, whereas the variance is expressed in squared units.

1. What does the difference between variance and standard deviation mean?

**Answer:**

|  |  |
| --- | --- |
| **Difference between Variance and Standard Deviation** | |
| **Variance** | **Standard Deviation** |
| It can simply be defined as the numerical value, which describes how variable the observations are. | It can simply be defined as the observations that get measured are measured through dispersion within a data set. |
| Variance is nothing but the average taken out of the squared deviations. | Standard Deviation is defined as the root of the mean square deviation |
| Variance is expressed in Squared units. | Standard deviation is expressed in the same units of the data available. |
| It is mathematically denoted as (σ2) | It is mathematically denoted as (σ) |
| Variance is a perfect indicator of the individuals spread out in a group. | Standard deviation is the perfect indicator of the observations in a data set. |

5. When is it appropriate to refer to a skewed data distribution?

**Answer:**

The Skewed distribution is non-symmetric and data values trail off on one side rather than centralized. If it's skewed to the right side it's called write skewed/positively skewed and if it's trail to the left side it's called left skewed/negative skewed.

The central tendency in skewed distribution is:

Right Skewed : Mod<Median<Mean

Left Skewed: Mean<Median<Mod

The most appropriate measure of central tendency for a skewed distribution is the Median because the median is more resistant to skewed values than the other measure of central tendency.

The median is less affected by outliers and skewed data than the mean, and is usually the preferred measure of central tendency when the distribution is not symmetrical.