Axis 5: Measures of Dispersion

the dispersion measures help interpret data variability, to understand how homogeneous or not homogeneous the data is. In simple words, it indicates how concentrated or scattered the variable is. They are as below:

- Range
- Variance
- Standard deviation
- Quartiles and Quartile deviation
- Mean deviation

1. Range

In Statistics, you might have studied the methods of finding a representative value for the given data, i.e. the measure of central tendency. To recall, mean, median, and mode are three **measures of central tendency**. As we know, measuring central tendency gives us an idea of where data points are centered. However, to interpret the data thoroughly, we should also see how the data are scattered. And how much they have bunched about a measure of central tendency. The method of calculate is:

range is the difference between the highest and lowest observations in a given data.

$$\text{Range} = \text{Maximum value} - \text{Minimum value}$$

$$\mathbf{R=Max-Min}$$

Example: We have the weights of a group of people as follows (WITH KILOGRAM):
59; 66; 58.8; 61; 69; 77; 61; 85; 77; 55; 80.
What is the range of this series

$$R=Max-Min=85-55=30Kg$$

Note: we can interpret this result if we have other sample to compare with it.

2. InterQuartile Range

The interquartile range defines the difference between the third and the first quartile. Quartiles are the partitioned values that divide the whole series into 4 equal parts. So, there are 3 quartiles. First Quartile is denoted by Q_1 known as the lower quartile, the second Quartile is denoted by Q_2 and the third Quartile is denoted by Q_3 known as the upper quartile. Therefore.

$$\mathbf{IQ = Q_3 - Q_1}$$

where Q_1 is the first quartile and Q_3 is the third quartile.

Its features:

- It is contain the middle 50% of statistical population.
- Its uses are limited, but it is used to compare two or more distributions.

The ratio between the range and the interquartile range.

It measures the middle 50% of dispersion of the statistical population compared to the range, and its relationship is written as follows
It measures dispersion of the middlly 50% to the statistical population compared to the range, and its relationship is written as follows:

RIQ=(IQ*100)/R = ((Q3-Q1)*100)/R

if it was RIQ=50% the statistical distribution is symmetrical, but if it is RIQ>50%, the statistical distribution is highly dispersed, and if it is RIQ<50%, the distribution in this case is little or weakly dispersed with respect to the central values.

3. Variance and Standard deviation

Varianceis the expected value of the squared variation of a random variable from its mean value, in probability and statistics. Informally, variance estimates how far a set of numbers (random) are spread out from their mean value.

The value of variance is equal to the square of standard deviation, which is another central tool.

Variance is symbolically represented by σ^2, s^2, or **Var(X)**.

The formula for variance is given by 3 ways:

$$\mathrm{Var(x)} = \frac{\sum(x_i - \bar{x})^2}{N}$$

$$\mathrm{Var(x)} = \frac{\sum_{i=1}^{k} n_i (x_i - \bar{x})^2}{\sum_{i=1}^{k} n_i}$$

$$\mathrm{Var(x)} = \frac{\sum_{i=1}^{k} n_i {x_i}^2}{\sum_{i=1}^{k} n_i} - \bar{x}^2$$

As we know already, the variance is the square of standard deviation, i.e.,

Variance = (Standard deviation)^2= σ^2

Properties

The variance, var(X) of a random variable X has the following properties.

1. Var(X + C) = Var(X), where C is a constant.

1. Var(CX) = C^2.Var(X), where C is a constant.
2. Var(aX + b) = a^2.Var(X), where a and b are constants.
3. If X_1, X_2,……., X_n are n independent random variables, then

$Var(X_1 + X_2 + \ldots + X_n) = Var(X_1) + Var(X_2) + \ldots + Var(X_n)$.
Now let's have a look at the relationship between Variance and Standard Deviation.

Variance and Standard Deviation

Standard deviation is the positive square root of the variance. The symbols σ and s are used correspondingly to represent population and sample standard deviations.

Standard Deviation is a measure of how spread out the data is. Its formula is simple; it is the square root of the variance for that data set. It's represented by the Greek symbol sigma (σ).

How to Calculate Variance

Variance can be calculated easily by following the steps given below:

- Find the mean of the given data set. Calculate the average of a given set of values
- Now subtract the mean from each value and square them
- Find the average of these squared values, that will result in variance

Say if $x_1, x_2, x_3, x_4, \ldots, x_n$ are the given values.

Therefore, the mean of all these values is:

$\bar{x} = (x_1+x_2+x_3+\ldots+x_n)/n$

Now subtract the mean value from each value of the given data set and square them.

$(x_1-\bar{x})^2, (x_2-\bar{x})^2, (x_3-\bar{x})^2, \ldots, (x_n-\bar{x})^2$

Find the average of the above values to get the variance.

$Var(X) = [(x_1-\bar{x})^2 + (x_2-\bar{x})^2 + (x_3-\bar{x})^2 + \ldots + (x_n-\bar{x})^2]/n$

Hence, the variance is calculated.

Example of Variance

Let's say the heights (in mm) are 610, 450, 160, 420, 310.

Mean and Variance is interrelated. The first step is finding the mean which is done as follows,

Mean = (610+450+160+420+310)/ 5 = 390

So the mean average is 390 mm.

To calculate the Variance, compute the difference of each from the mean, square it and find then find the average once again.

So for this particular case the variance is :

$= (220^2 + 60^2 + (-230)^2 + 30^2 + (-80)^2)/5$

$= (48400 + 3600 + 52900 + 900 + 6400)/5$

Final answer : Variance = 22440

4. Mean deviation

The mean deviation is defined as a statistical measure that is used to calculate the average deviation from the mean value of the given data set. The mean deviation of the data values can be easily calculated using the below procedure.

Step 1: Find the mean value for the given data values

Step 2: Now, subtract the mean value from each of the data values given (Note: Ignore the minus symbol)

Step 3: Now, find the mean of those values obtained in step 2.

Mean Deviation Formula

The formula to calculate the mean deviation for the given data set is given below.

Mean Deviation =The mean deviation is defined as a statistical measure that is used to calculate the average deviation from the mean value of the given data set. The mean deviation of the data values can be easily calculated using the below procedure.

Step 1: Find the mean value for the given data values

Step 2: Now, subtract the mean value from each of the data values given (Note: Ignore the minus symbol)

Step 3: Now, find the mean of those values obtained in step 2.

Mean Deviation Formula

The formula to calculate the mean deviation for the given data set is given below.

$$\text{Mean Deviation} = \frac{\sum_{i=1}^{k} n_i |x_i - \bar{x}|}{\sum_{i=1}^{k} n_i}$$