**Measure of Dispurtion**

* 1. **Range**
  2. **Variance**
  3. **Standard deviation**
  4. **Percentile, quartile and interquartile range (IQR)**

**Measure of dispersion:**

The measures of central tendency are not adequate to describe data. Two data sets can have the same mean but they can be entirely different. Thus to describe data, one needs to know the extent of variability. This is given by the measures of dispersion.

1. **Range**
2. **Variance**
3. **Standard deviation**
4. **Percentile and quartile**
5. **interquartile range (IQR)**

These are the commonly used measures of dispersion.

**Range:**

The range is the difference between the maximum and minimum values in a dataset. It provides a simple measure of the spread of the data. The range is sensitive to outliers because it depends only on two extreme values.

Mathematically, the range of a dataset is calculated as: Range=max(x)−min(x)

**Variance:**

Variance measures the average squared deviation of each data point from the mean.

Variance talks about the spread of data. There are two types of variance,

1. Population Variance
2. Sample Variance.

**Population Variance(**σ2**):**

When the data size is small then the population variance is calculated on the actual population dataset.

Calculate the Population variance for the below dataset.

X {1,2,3,4,5}

|  |  |
| --- | --- |
| Formula |  |
| Values | (1-3)2+(2-3)2+(3-3)2+(4-3)2+(5-3)2 |
| 5 |
| Population Variance(σ2) | 2 |
|

**Sample Variance(**s2**):**

When the number of observations increases then a few data points are selected that can represent the entire population. These specific data points form a sample and the variance calculated on this data is called the sample variance.

Calculate the Population variance for the below dataset.

X {1,2,3,4,5}

Consider the above(X) sample data is taken from the population data

|  |  |
| --- | --- |
| Formula | Data Analysis in the Geosciences |
|
|
|
| Values | (1-3)2+(2-3)2+(3-3)2+(4-3)2+(5-3)2 |
| 5-1 |
| Sample  Variance(S2) | 2.5 |
|

A quantity is expressed by how much the group members differ from the mean value.

The square root of the variance is known as the standard deviation i.e. S.D. = √σ. However standard deviation can be calculated by using the below formula.

**Standard Deviation(σ):**

Standard deviation is a statistical measure of the amount of dispersion in a set of values. It measures how spread out the values in a dataset are from the mean (average) value. A low standard deviation indicates that the values tend to be close to the mean, while a high standard deviation indicates that the values are spread out over a wider range.

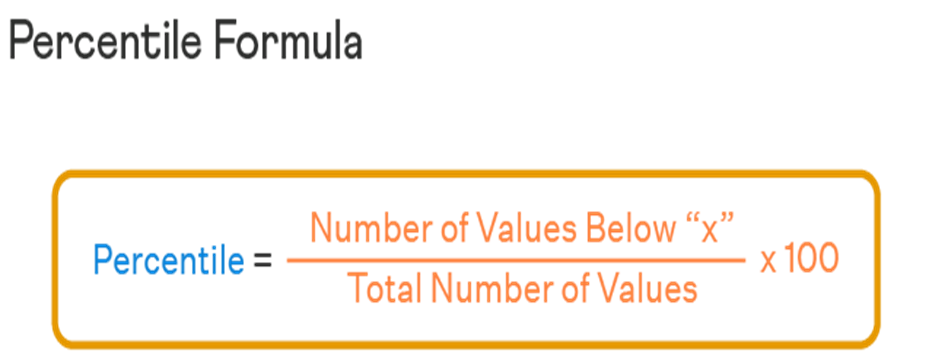
**Calculate Standard Deviation**

X {1,2,3,4,5}

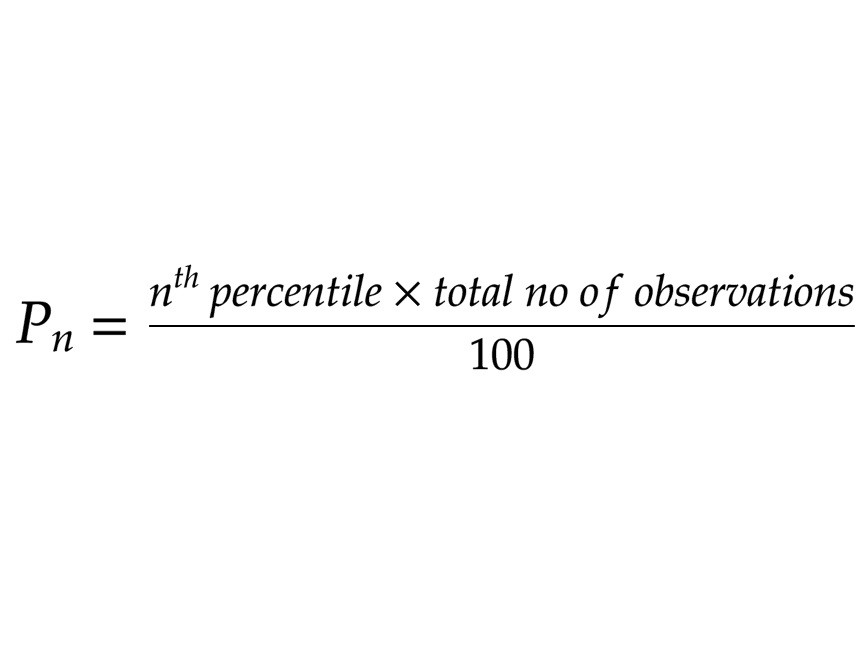
|  |  |  |
| --- | --- | --- |
| Formula | Standard Deviation  (σ)= |  |
| Values |  | |
| Standard Deviation  (σ) | 1.414 | |
|

**Percentile:**

A percentile is a statistical measure used to describe the relative position of a particular value in a dataset. It tells you what percentage of data points are below or equal to a given value. Percentiles are commonly used in various fields, including statistics, education, healthcare, and finance, to understand the distribution of data and make comparisons.

****

Percentile Rank Formula (Find the position of a given Percentile):



**Quartile:**

Quartiles are a specific type of percentile that divides a dataset into four equal parts or quarters. They are often used in statistics and data analysis to understand the distribution and spread of data. Quartiles are useful for identifying the central tendency and variability of a dataset, as well as for detecting potential outliers.

There are three quartiles commonly calculated:

**First Quartile (Q1):**

This is the 25th percentile, also known as the lower quartile. It divides the lowest 25% of the data from the rest. In other words, Q1 represents the value below which 25% of the data points fall.

**Second Quartile (Q2):**

This is the 50th percentile, which is also known as the median. Q2 divides the data into two equal halves, with 50% of the data below it and 50% above it. The median is often used to measure the central tendency of a dataset.

**Third Quartile (Q3):**

This is the 75th percentile, also known as the upper quartile. It divides the lowest 75% of the data from the highest 25%. Q3 represents the value below which 75% of the data points fall.