Variability in Statistics

Range: In statistics, the range is the smallest of all dispersion measures. It is the difference between the distribution's two extreme conclusions. In other words, the range is the difference between the distribution's maximum and minimum observations.

$$\text{Range} = X_{max} - X_{min}$$

where X_{max} represent the largest observation and X_{min} represent the smallest observation of the variable values.

Percentiles, Quartiles and Interquartile Range (IQR)

- Percentiles - It is a statistician's unit of measurement that indicates the value below which a given percentage of observations in a group of observation fall.

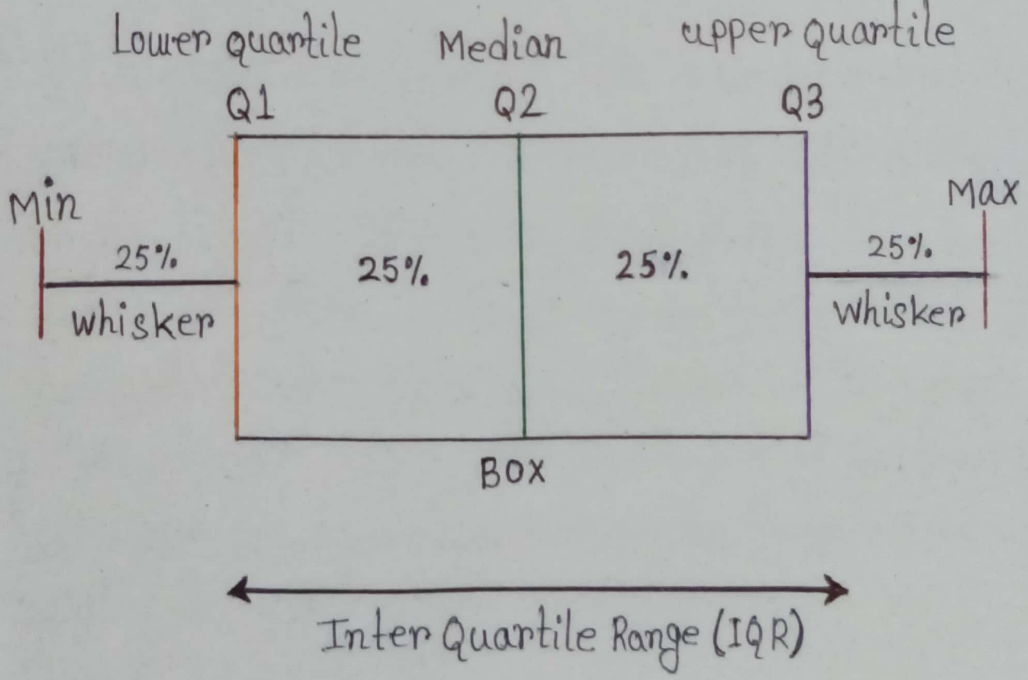
  For instance, the value Qx repersents the 40th percentile of xx (0.40)

- Quantiles - values that divide the number of data points into four more or less equal parts, or quarters. Quantiles are the 0th, 25th, 50th, 75th, and 100th percentile values or the 0th, 25th, 50th, 75th, and 100th percentile values.

Scanned with OKEN Scanner

- Interquartile Range (IQR) - The difference between the third and first quartiles is defined by the interquartile rang. The partitioned values that divide the entire series into four equal parts are known as quartiles. So, there are three quartiles. The first quartile, known as the lower quartile, is denoted by Q1, the second quartile by Q2, and third quartile by Q3, known as the upper quartile. As a result, the interquartile range equals the upper quartile minus the lower quartile.

IQR = Upper Quartile - Lower Quartile
= Q3 - Q1



- **Variance** – The dispersion of a data collection is measured by variance. It is defined technically as the average of squared deviations from the mean.

Population Variance	Sample Variance
$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$	$S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$
σ^2 = Population Variance x_i = Value of i^{th} element μ = Population mean N = Population Size	S^2 = Sample variance x_i = Value of i^{th} element $\bar{x}$ = Sample mean n = Sample Size

- **Standard Deviation** – The standard deviation is a measure of data dispersion WITHIN a single sample selected from the study population. The square root of the variance is used to compute it. It simply indicates how distant the individual values in a sample are from the mean. To put it another way, how dispersed is the data from the sample? As a result, it is a sample statistic.

Scanned with OKEN Scanner

Standard Deviation Formula

Population	Sample
$\sigma = \sqrt{\frac{\Sigma(X-\mu)^2}{N}}$	$s = \sqrt{\frac{\Sigma(X-\bar{X})^2}{n-1}}$
X - The value in the data distribution μ - The population mean N - Total Number of observations	X - The value in the data distribution $\bar{X}$ - The Sample Mean n - Total Number of observations

- Standard Error (SE) - The standard error indicate how close the mean of any given sample, from that population is to the true population mean. when the standard error rises, implying that the means are more dispersed, it becomes more likely that any given mean is an inaccurate representation of the true population mean. when the sample size is increased, the standard error decreases - as the sample size approaches the true population size, the sample means cluster more around the true population mean.

Standard Error Formula

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Scanned with OKEN Scanner

Relationship Between variables

- Causality: The term "causation" refers to a relationship between two events in which one is influenced by the other. There is causality in statistics when the value of one event, or variable, grows or decreases as a result of other events. Each of the events we just observed may be thought of as a variable, and as the number of hours worked grows, so does the amount of money earned. On the other hand, if you work fewer hours, you will earn less money.

statistic assesses how much- and how far-the variables change in tandem. To put it another way, it's a measure of the variance between two variables. The metric, on the other hand, does not consider the interdependence of factors. Any positive or negative value can be used for the variance.

The following is how the values are interpreted:

- Positive covariance: when two variables move in the same direction, this is called positive covariance.

- Negative Covariance: indicates that two variables are moving in opposite directions.

Scanned with OKEN Scanner

mean of all values of independent variable

Single observed value of dependent variable

Single observed value of independent variable

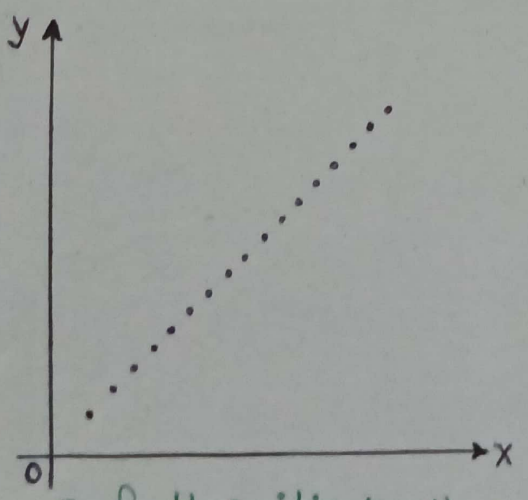
total count of Sample values

$$Cov(X,Y) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n - 1}$$

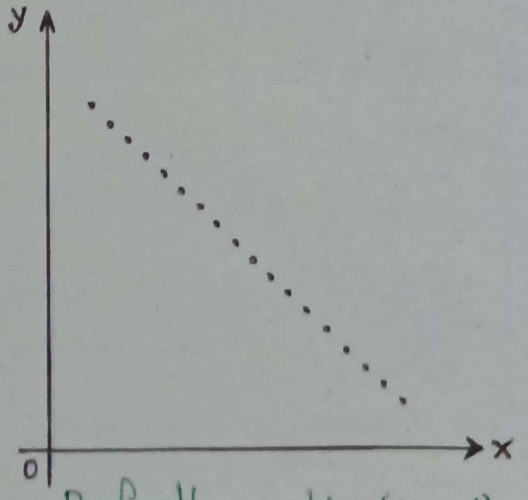
mean of all values of independent variable

Population count minus one (Bessel's Correction)

- **Correlation:** Correlation is a statistical method for determining whether or not two quantitative or categorical variables are related. To put it another way, it's a measure of how things are connected. Correlation analysis is the study of how variables are connected.

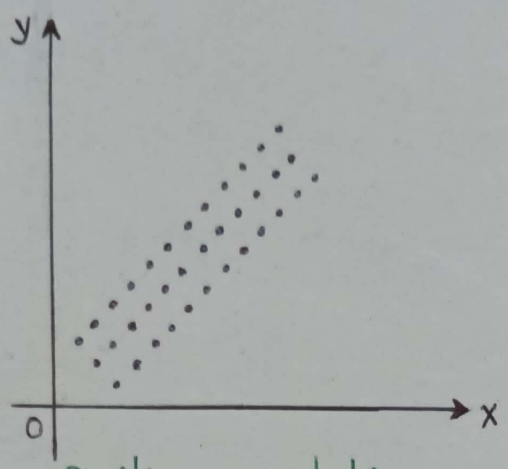


Perfectly Positive (r = 1)

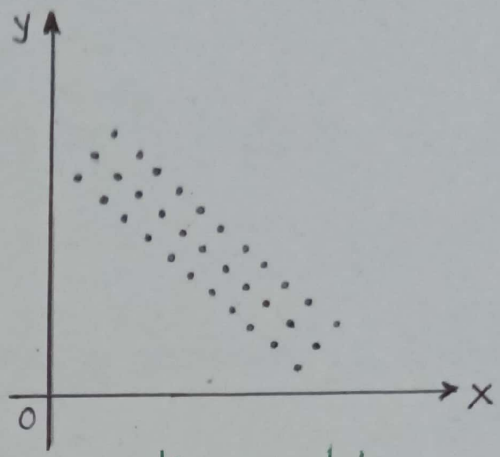


Perfectly negative (r = -1)

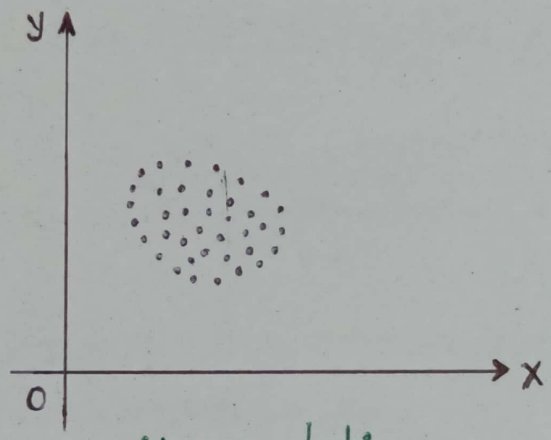
Scanned with OKEN Scanner



Positive correlation



Negative correlation



Uncorrelation

- Pearson's product moment correlation coefficient (r) :

$$r = \frac{cov(x,y)}{\sigma_x \cdot \sigma_y}$$, where $cov(x,y) = \frac{1}{n}\Sigma(x-\bar{x})(y-\bar{y})$

σ_x = standard Deviation of x

σ_y = standard Deviation of y

Scanned with OKEN Scanner

• Rank correlation :

- Spearman's rank correlation coefficient (R)

$$R = 1 - \frac{6\Sigma D^2}{N^3 - N}$$

where D = difference of the ranks of an individual in the two characters and

N = number of individuals in the group.

- Coefficient of rank correlation (R) in case of 'tied rank's'

$$R = 1 - \frac{6\left[\Sigma D^2 + \Sigma \frac{t^3 - t}{12}\right]}{N^3 - N}$$

where t = number of individuals involved in a tie in any of the two series.

Correlations are useful because they allow you to forecast future behaviour by determining what relationship variables exist. In the practical field, such as government and healthcare, knowing what the future holds is critical. Budgets and company plans are also based on these facts.

Scanned with OKEN Scanner