**Statistics Assignment 4**

**Name = Khushal tondon**

**Ans no. 1**

Covariance is a measure of the relationship between two random variables and to what extent, they change together. Or we can say, in other words, it defines the changes between the two variables, such that change in one variable is equal to change in another variable.

Types of Covariance

Covariance can have both positive and negative values. Based on this, it has two types:

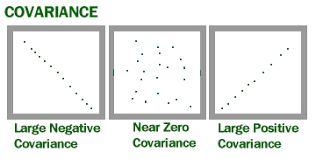
1. Positive Covariance
2. Negative Covariance

Positive Covariance

If the covariance for any two variables is positive, that means, both the variables move in the same direction. Here, the variables show similar behaviour. That means, if the values (greater or lesser) of one variable corresponds to the values of another variable, then they are said to be in positive covariance.

Negative Covariance

If the covariance for any two variables is negative, that means, both the variables move in the opposite direction. It is the opposite case of positive covariance, where greater values of one variable correspond to lesser values of another variable and vice-versa.



Covariance Formula

Covariance formula is a statistical formula, used to evaluate the relationship between two variables. It is one of the statistical measurements to know the relationship between the variance between the two variables.

Let us say X and Y are any two variables, whose relationship has to be calculated. Thus the covariance of these two variables is denoted by Cov(X,Y). The formula is given below for both population covariance and sample covariance.

Population Covariance Formula

Cov(X,Y)= ∑ ( x i − x― ) ( y i − y― ) / N

Sample Covariance Formula

Cov(X,Y)= ∑ ( x i − x― ) ( y i − y― ) / n − 1

Where,

xi = data value of x

yi = data value of y

x̄ = mean of x

ȳ = mean of y

N = number of data values.

**Ans no. 2**

Correlation is a step ahead of covariance as it quantifies the relationship between two random variables. In simple terms, it is a unit measure of how these variables change concerning each other (normalized covariance value).

The correlation has an upper and lower cap on a range, unlike covariance. It can only take values between +1 and -1. A correlation of +1 indicates that random variables have a direct and strong relationship.

On the other hand, the correlation of -1 indicates a strong inverse relationship, and an increase in one variable will lead to an equal and opposite decrease in the other variable. 0 means that the two numbers are independent

**Ans no. 3**

Pearson Correlation Coefficient

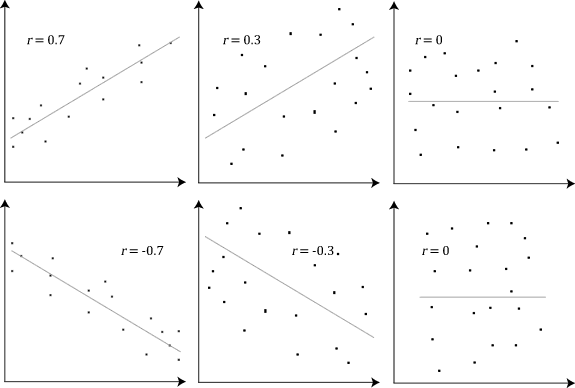
In statistics, the Pearson correlation coefficient also referred to as Pearson’s r or the bivariate correlation is a statistic that measures the linear correlation between two variables X and Y. It has a value between +1 and −1. A value of +1 is a total positive linear correlation, 0 is no linear correlation, and −1 is a total negative linear correlation.

Important Inference to keep in mind: The Pearson correlation can evaluate ONLY a linear relationship between two continuous variables (A relationship is linear only when a change in one variable is associated with a proportional change in the other variable)

Example use case: We can use the Pearson correlation to evaluate whether an increase in age leads to an increase in blood pressure.

Pearson correlation coefficient of X and Y = COV( X,Y ) / S.D of X \* S.D of Y

Below is an example of how the Pearson correlation coefficient (r) varies with the strength and the direction of the relationship between the two variables. Note that when no linear relationship could be established (refer to graphs in the third column), the Pearson coefficient yields a value of zero.



Spearman Correlation Coefficient

In statistics, Spearman’s rank correlation coefficient or Spearman’s ρ, named after Charles Spearman is a nonparametric measure of rank correlation (statistical dependence between the rankings of two variables). It assesses how well the relationship between two variables can be described using a monotonic function.

Important Inference to keep in mind: The Spearman correlation can evaluate a monotonic relationship between two variables — Continous or Ordinal and it is based on the ranked values for each variable rather than the raw data.

Rs  = COV ( Rx , Ry ) / S.D Rx \* S.D Ry

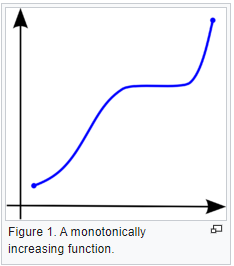
What is a monotonic relationship?

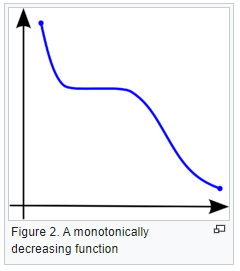
A monotonic relationship is a relationship that does one of the following:

(1) as the value of one variable increases, so does the value of the other variable, OR,

(2) as the value of one variable increases, the other variable value decreases.

BUT, not exactly at a constant rate whereas in a linear relationship the rate of increase/decrease is constant.





**Ans no. 4**

The Spearman rank correlation coefficient is only to be used to describe the relationship between linear data.​ Also, it can be used for data at the ordinal level and it is easier to calculate by hand than the Pearson correlation coefficient.

The Spearman rank correlation coefficient is only to be used to describe the relationship between nonlinear data.​ Also, it can be used for data at the ordinal level and it is easier to calculate by hand than the Pearson correlation coefficient.

The Spearman rank correlation coefficient can be used to describe the relationship between linear or nonlinear data.​ Also, it can be used for data at the ordinal level and it is easier to calculate by hand than the Pearson correlation coefficient.

**Ans no. 5**

* The central limit theorem (CLT) states that the distribution of sample means approximates a normal distribution as the sample size gets larger, regardless of the population's distribution.
* Sample sizes equal to or greater than 30 are often considered sufficient for the CLT to hold.
* A key aspect of CLT is that the average of the sample means and standard deviations will equal the population mean and standard deviation.
* A sufficiently large sample size can predict the characteristics of a population more accurately.
* CLT is useful in finance when analyzing a large collection of securities to estimate portfolio distributions and traits for returns, risk, and correlation.

Basically , if there is any distribution shape , if we pick up some samples from distributions and suppose size of sample is greater than or equal to 30 .

And if we calculate sample mean , and when we try to plot it then distribution of this sample will be the normal distributed .

Same for all other distributions which are different in shapes .