Frequently Asked Questions

1. Define Product Moment Correlation Coefficient?

Answer:
Correlation coefficient between two random variables X and Y, usually denoted by r(X, Y) or r_{XY}, is a numerical measure of linear relationship between them and is defined as

$$r_{XY} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

2. Who developed the formula of product moment correlation coefficient?

Answer:
As a measure of intensity or degree of linear relationship between two variables, Karl Pearson, a British Biometrician, developed a formula called correlation coefficient which is based on moments. Hence, it is called product moment correlation coefficient.

3. What is another name of Product moment correlation coefficient?

Answer: Since this method is developed by Karl Pearson, it is also known as Karl Pearson's correlation coefficient.

4. Write the formula for finding product moment correlation coefficient.

Answer:
For raw data,

$$r_{XY} = \frac{n\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{\sqrt{[n\Sigma x_i^2 - (\Sigma x_i)^2][n\Sigma y_i^2 - (\Sigma y_i)^2]}}$$

For tabulated data,

$$r_{XY} = \frac{n\Sigma f_i x_i y_i - \Sigma f_i x_i \Sigma f_i y_i}{\sqrt{[n\Sigma f_i x_i^2 - (\Sigma f_i x_i)^2][n\Sigma f_i y_i^2 - (\Sigma f_i y_i)^2]}}$$

5. What are the characteristics of Product moment correlation coefficient?

Answer:
Following are the main characteristics of product moment correlation coefficient:

- Based on arithmetic mean and standard deviation – The formula is based upon arithmetic mean and standard deviation. The products of the corresponding values of the two series that is, co-variance is divided by the product of standard deviations of the two series to determine the formula
- Determines the direction of relationship - Karl Pearson's method establishes the direction of relationship of variables, namely positive or negative
- Establishes the size of relationship – This method also shows the size of relationship between variables of the two series. It ranges between plus one and minus one. Plus one means perfect positive correlation and minus one means

perfect negative relationship. In case the value is zero, then it means no relationship between the variables
- Ideal measure – This method is considered to be an ideal method of calculation of correlation coefficient. It is because of the covariance which is most reliable as a standard statistical tool

6. Write assumptions on which product moment correlation coefficients are bases.

Answer:
Coefficient of correlation of Product Moment is based on the following assumptions:
- The relationship is linear – We assume that there is a linear relationship between the two variables. That means if the two variables are plotted, we get a straight line
- Normal distribution – A large number of independent causes are operating in both the variables correlated so as to produce a normal distribution
- Related in a Casual Function – This is the forces so operating on each of the variable series is not independent of each other but are related in a causal fashion. In other words, cause and effect relationship exists between different forces operating on the items of the two variable series. These forces must be common to both the series. If the operating forces are entirely independent of each other and not related in any fashion, then there cannot be any correlation between the variables under study. For example, the correlation coefficient between,
  a. The series of heights and incomes of individuals over a period of time
  b. The series of marriage rate and the rate of agricultural production in a country over a period of time
  c. The series relating to the size of the shoe and intelligence of a group of individuals

  should be zero, since the forces affecting the two variable series in each of the above cases are entirely independent of each other.

7. What are the merits of the product moment correlation coefficient?

Answer:
The Product Moment correlation coefficient has the following merit:
- Counts all values: It takes into account all values of the given data of x and y. Therefore, it is based on all observations of the series
- More Practical and Popular: Product moment correlation coefficient ‘r’ is considered to be more practical method as compared to other mathematical methods used for ‘r’. It is also very popular and as such, commonly used method
- Numerical Measurement of ‘r’: It provides numerical measurement of coefficient of correlation
- Measures Degree and Direction: This method measures both degree and direction of the correlation between the variables at a time
- Facilitates Comparison: Product Moment coefficient of correlation is a pure number independent of units. Therefore, the comparison between the series can be done easily
- Algebraic treatment Possible: Product Moment coefficient of correlation techniques can easily be applied for higher algebraic treatment

8. Write the demerits of product moment correlation coefficient?

Answer:

The use of coefficient of correlation has certain limitations which are as under:

- Linear relationship: Coefficient of correlation assumes linear relationship between the variables regardless of the fact whether that assumption is correct or not
- More time consuming: Compared with some other methods, this method is more time consuming
- Affected by extreme items: Another drawback of coefficient of correlation is that it is affected by the extreme items
- Difficult to interpret: It is not easy to interpret the significance of correlation coefficient. It is generally misinterpreted

9. Show that two independent variables are uncorrelated.

Answer:
Proof:

If X and Y are independent variables, then Cov(X, Y)=0

Implies, $Cov(X, Y) = \frac{1}{n} \Sigma XY - \overline{X}\,\overline{Y} = 0$

Hence, two independent variables are uncorrelated.

10. Show that the converse of the above is not true.

Answer:
To prove that the converse is not true, that is uncorrelated variables may not be independent as the following example illustrates.

X	Y	XY
-3	9	-27
-2	4	-8
-1	1	-1
1	1	1
2	4	-8
3	9	27
0	**25**	**0**

$$\overline{X} = \frac{1}{n} \Sigma X = 0$$

$$r_{XY} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} = 0$$

$$r_{XY} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} = 0$$

Thus, in the above example, the variables X and Y are uncorrelated. But on careful examination we find that X and Y are not independent but they are connected by the relation $Y=X^2$. Hence, two uncorrelated variables need not necessarily be independent.

11. If each of the two variables X and Y takes two values, 0, 1 with positive probabilities, then r(X,Y)=0 implies X and Y are independent.

Answer:
Proof:
Let X take values 1 and zero with positive probabilities p1 and q1 respectively and let Y take the values 1 and zero with positive probabilities p_2 and q_2 respectively.
Then r(X,Y) = 0 implies, Cov(X,Y)=0
Implies, 0=E(XY)-E(X)E(Y)

=1.P(X=1 ∩Y=1)-[1.P(X=1).1.P(Y=1)]

$=P(X=1 \cap Y=1) - p_1 p_2$

Implies, $P(X=1 \cap Y=1)= p_1 p_2 =P(X=1)P(Y=1)$

Hence, X and Y are independent.

12. Show that if correlation is present, then coefficient of correlation would lie between ±1. If correlation is absent, then it is denoted by zero. That is $-1 \leq r \leq +1$.

Answer:

$$r_{XY}=\frac{\frac{1}{n}\Sigma(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\frac{1}{n}\Sigma(x_i-\overline{x})^2\frac{1}{n}\Sigma(y_i-\overline{y})^2}}$$

Where,

$$r_{XY^2}=\frac{\Sigma(a_i b_i)^2}{\Sigma a_{i^2}\Sigma b_{i^2}}$$

$$a_i=x_i-\overline{x}$$

$$b_i=y_i-\overline{y}$$

¿

righ

¿

¿

¿ (i) ¿

¿

¿

From Schwartz inequality, which states that if a_i b_i, i = 1, 2, . . ., n are real quantities, then,

$$(\Sigma a_i b_i)^2 \leq (\Sigma a_{i^2})(\Sigma b_{i^2})$$

The sign of equality holding iff, $a_1/b_1 = a_2/b_2, \ldots = a_n/b_n$.

Using Schwartz inequality, we get,

$r_{XY} \leq 1$, ie., $|r_{XY}| \leq 1$, implies, $-1 \leq r \leq +1$.

13. Correlation coefficient is independent of change of origin and scale.

Answer:

Let U=(X-a)/h, V=(Y-b)/k, so that X=a+hU and Y=b+kV, where a, b, h, k are constants; h>0, k>0.

We shall prove that, r(X,Y) = r(U,V)

Since X=a+hU and Y=b+kV.

On taking expectations, we get,

E(X)=a+hE(U) and E(Y)=b+kE(V)

Therefore, X-E(X)=h[U-E(U)] and Y-E(Y)=k[V-E(V)]

Implies,

Cov(X,Y)=E[{X-E(X)}{Y-E(Y)]

=E[h{U-E(U)}k{V-E(V)}]

$=hkE[\{U-E(U)\}\{V-E(V)\}] = hk\,Cov(U,V)$

$\sigma_X^2 = E[\{X-E(X)\}^2] = E[h^2\{U-E(U)\}^2] = h^2\sigma_U^2$

$\sigma_X = h\sigma_u$ (h>0)

$\sigma_Y^2 = E[\{Y-E(Y)\}^2] = E[k^2\{V-E(V)\}^2] = k^2\sigma_v^2$

$\sigma_Y = k\sigma_V$ (k>0)

Substituting in r(X,Y),

$$r_{XY} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \frac{hkCov(U,V)}{hk\,\sigma_U \sigma_V} = \frac{Cov(U,V)}{\sigma_U \sigma_V} = r_{UV}$$

14. If X and Y are random variables and a, b, c, d are any numbers provided only that $a \neq 0, c \neq 0$ then

$$r(aX+b, cY+d) = \frac{ac}{|ac|} r(X,Y)$$

Answer:

With usual notations, we have,

$Var(aX+b) = a^2\sigma_x^2$; $Var(cY+d) = c^2\sigma_Y^2$

$Cov(ax+b, cY+d) = ac\sigma_{xy}$

Therefore,

$$r(aX+b, cY+d) = \frac{Cov(aX+b, cY+d)}{\sqrt{Var(aX+b)Var(cY+d)}} = \frac{ac\,\sigma_{XY}}{|a||c|\sigma_X \sigma_Y} = \frac{ac}{|ac|} r(X,Y)$$

If ac>0, i.e., if a and c are of same signs, then ac/|ac|=+1.

If ac<0, i.e., if a and c are of opposite signs, then ac/|ac|=-1.

15. Mention some of the properties of product moment correlation coefficient.

Answer:

- Coefficient of correlation is based on a suitable measure of variation as it takes into account all items of the variable
- If there is accidental correlation, in that case the coefficient of correlation might lead to fallacious conclusions. It is known as non-sense or spurious correlation.
- Coefficient of correlation works both ways:

  i.e., $r_{XY} = r_{YX}$
- The coefficient of correlation does not prove causation but it is simply a measure of co-variation. It is because variations in X and Y series may be due to:

  i. Some common cause

  ii. Some mutual dependence

  iii. Some change and

iv. Some causation of the subject to be relative.

- Coefficient of correlation is independent of the unit of measurement