M.A-SEM III rd Paper- CC 10

Quantitative Techniques & Research Methodology

Karl Pearson's: Coefficient of Correlation

Dr. Supriya

Asst. Professor &

Post Doc. Fellow (ICSSR)

Department of Geography

Patna University, Patna

Contents

- ✓ Introduction to the methods of correlation
- ✓ Coefficient of Correlation
- ✓ Interpreting Correlation Coefficient (r)
- ✓ Coefficient of Determination
- ✓ Merits and Demerits of Pearson's method of studying correlation
- ✓ Suggested Readings

Introduction to the methods of correlation

- Correlation a LINEAR association between two random variables
- Correlation analysis show us how to determine both the nature and strength of relationship between two variables
- When variables are dependent on time correlation is applied
- Correlation lies between +1 to -1

A .Mathematical Methods	B. Non –Mathematical Methods
1. Karl Pearson's Method	Graphic Method
2. Spearman's Method of Rank	2. The Scatter Or Dot diagram
3. The Concurrent Deviation Method	3. Correlation table Method

Coefficient of Correlation

- A measure of the strength of the linear relationship between two variables that is defined in terms of the (sample) covariance of the variables divided by their (sample) standard deviations
- Represented by “r”
- r lies between +1 to -1
- Magnitude and Direction
- $-1 < r < +1$
- The + and – signs are used for positive linear correlations and negative linear correlations, respectively

Interpreting Correlation Coefficient (r)

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[\, n\Sigma x^2 - (\Sigma x)^2 \,][\, n\Sigma y^2 - (\Sigma y)^2 \,]}}$$

- Shared variability of X and Y variables on the top
- Individual variability of X and Y variables on the bottom
- strong correlation: r > .70 or r < –.70
- moderate correlation: r is between .30 & .70 or r is between –.30 and –.70
- weak correlation: r is between 0 and .30 or r is between 0 and – .30

Calculating Pearson's 'r'

Example 4. Compute coefficient of correlation by Karl Pearson Method for the following data

X :	1800	1900	2000	2100	2200	2300	2400	2500	2600
f :	5	5	6	9	7	8	6	8	9

Solution

Let the A.M.s A_x and A_y be 2200 and 6 for X and Y series respectively

X	Y	dx	(i=100) dx	dy	dx^2	dy^2	dxdy
1800	5	–400	–4	–1	16	1	4
1900	5	–300	–3	–1	9	1	3
2000	6	–200	–2	0	4	0	0
2100	9	–100	–1	3	1	9	–3
2200	7	0	0	1	0	1	0
2300	8	100	1	2	1	4	2
2400	6	200	2	0	4	0	0
2500	8	300	3	2	9	4	6
2600	9	400	4	3	16	9	12
N = 9			$\Sigma dx = 0$	$\Sigma dy = 9$	$\Sigma dx^2 = 60$	$\Sigma dy^2 = 29$	$\Sigma dx\,dy = 24$

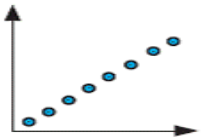
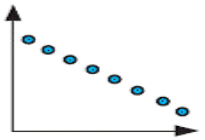
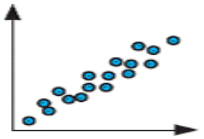
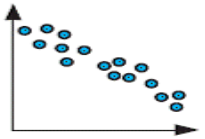
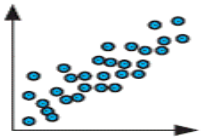
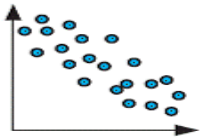
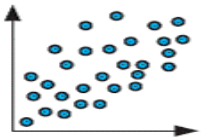
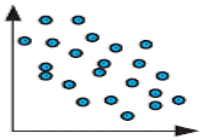
$$r = \frac{(9)(24) - (0)(9)}{\sqrt{(9)(60) - (0)^2}\sqrt{(9)(29) - (9)^2}} = \frac{216}{\sqrt{97200}} = .69$$

(**Note :** We can also proceed dividing X by 100)

Karl Pearson's Coefficient of Correlation explained

INTERPRETATION OF PEARSON'S CORRELATION COEFFICIENT

Pearson's correlation coefficient gives a measure of the relationship between two variables on a scale from −1 to 1. Word descriptors based on r-values seem doubtful at the best of times and the majority of texts on this subject do not include them. Many texts and Internet sites vary on the advice they give. Here is one possible interpretation.

r	Description	r	Description
1	perfect positive correlation 	−1	perfect negative correlation 
0.75 to 1	strong positive correlation 	−1 to −0.75	strong negative correlation 
0.50 to 0.75	moderate positive correlation 	−0.75 to −0.50	moderate negative correlation 
0.25 to 0.50	weak positive correlation 	−0.50 to −0.25	weak negative correlation 

Coefficient of Determination

- Coefficient of determination lies between 0 to 1, Represented by r2
- The coefficient of determination is a measure of how well the regression line represents the data
- If the regression line passes exactly through every point on the scatter plot, it would be able to explain all of the variation
- The further the line is away from the points, the less it is able to explain
- r 2, is useful because it gives the proportion of the variance (fluctuation) of one variable that is predictable from the other variable
- It is a measure that allows us to determine how certain one can be in making predictions from a certain model/graph

Coefficient of Determination…

- The coefficient of determination is the ratio of the explained variation to the total variation
- The coefficient of determination is such that 0 < r 2 < 1, and denotes the strength of the linear association between x and y
- The coefficient of determination represents the percent of the data that is the closest to the line of best fitv For example, if r = 0.922, then r 2 = 0.850
- Which means that 85% of the total variation in y can be explained by the linear relationship between x and y (as described by the regression equation)
- The other 15% of the total variation in y remains unexplained

Merits and Demerits of Pearson's method of studying correlation

Merits:

1. This method indicates the presence or absence of correlation between two variables and gives the exact degree of their correlation.
2. In this method, we can also ascertain the direction of the correlation; positive, or negative.
3. This method has many algebraic properties for which the calculation of co-efficient of correlation, and other related factors, are made easy.

Demerits:

1. It is more difficult to calculate than other methods of calculations.

2. It is much affected by the values of the extreme items.

3. It is based on a many assumptions, such as: linear relationship, cause and effect relationship etc. which may not always hold good.

4. It is very much likely to be misinterpreted in case of homogeneous data.

Suggested Readings

- Kothari, C.R. & Garg, G.; (2014) Research Methodology : Methods and Techniques, New Edge International Publisher, New Delhi.
- Mahmood, Aslam. (2008) Statistical Methods in Geographical Studies, Rajesh Publishing, New delhi
- Misra, R.P & Ramesh, A.; (1989) Fundamentals of Cartography, Concept publishing , New delhi
- Negi, B.S.; (2008) Statistical Geography, Kedar Nath ram Nath, Meerut

Contact me:
Mob/ WA: 9006640841
Emai:supriyavatsa52256@gmail.com
Webpage: http://bhuvankosh.com