Chapter 21

Correlation

In Social Study as well as Psychology there are times where it is needed to know whether there exists any relationship between the different abilities of the individual or they are independent of each other. Consequently, there are numerous questions like the following which, have to be answered.

(*i*) Does scholastic achievement depend upon the general intelligence of a child?

(*ii*) Is it true that the height of the children increases with the increase in their age?

(*iii*) Is there are relationship between the size of the skull and general intelligence of the individuals?

(*iv*) Is it true that Dull children tend to be more neurotic than the bright children?

The questions and problems like the above in which there is a need of finding out the relationship between two variables (Age and Height Intelligence and Achievement etc.) can be tackled properly by the method of correlation.

There are many types of correlation like Linear, Curvilinear, Biserial, Partial or Multiple correlation that are computed in Statistics. As we, in this text, aim to have an elementary knowledge of the statistical methods we will take only the Linear correlation in the following pages.

LINEAR CORRELATION

This the simplest kind of correlation to be found between the two sets of scores or variables. Actually when the relationship between two sets of scores or variables can be represented graphically by a straight line, it is known as Linear Correlation. Such type of correlation clearly reveals how the change in one variable is accompanied by a change or to what extent increase or decrease in one is accompanied by the increase or decrease in order.

The correlation between two sets of measures of variables can be positive or negative. It is said to be positive when an increase (or decrease) in the corresponds to an increase (or decrease) in the other. It is negative when increase corresponds to decrease and decrease corresponds with increase. There is also possibility of third type of correlation i.e. zero correlation between the two sets of measures of variables if there exists no relationship between them.

COEFFICIENT OF CORRELATION

For expressing the degree of relationship quantitatively between two sets of measures of variables we usually take the help of an index that is known as coefficient of correlation. It is a kind of ratio which expresses the extent to which changes in one variable are accompanied with changes in the other variable. It involves no units and varies from -1 (indicating perfect negative correlation) to + 1 (indicating perfect positive correlation). In case the coefficient of correlation is zero it indicates zero correlation between two sets of measures.

COMPUTATION OF COEFFICIENT OF CORRELATION

There are two different methods of computing coefficient of correlation (linear).

There are–

(*a*) Rank Difference Method

(*b*) Product Moment Method

(*a*) Rank Difference Method of Computing Coefficient of Correlation

In computing coefficient of correlation between two sets of scores achieved by the individuals, with the help of this method we require ranks i.e. positions of merits of these individuals in the possession of certain characteristics. The coefficient of correlation computed by this method as it considers only the ranks of the individuals in the characteristics A and B is known as Rank correlation coefficient and is designated by Greek letter (Rho). Some times it is also known as Spearman's coefficient of correlation after the name of its inventor.

In case where we do not have scores and have to work with data in which differences between the individuals in the possession of certain characteristics can be expressed only by ranks. Rank correlation coefficient is the only correlation coefficient that can be computed. But this does not mean that it cannot be computed from the usual data given in scores. In case the data contain scores of individuals, we can compute by converting them into ranks. For example, if the marks of a group of 5 students are given as 17, 25, 9, 35,18, we will rank them as 4, 2, 5, 1 and 3. We determine the rank of position of the individuals in both the given sets of scores. These ranks are then subjected to further calculation for the determination of the coefficient of correlation.

How it is done can be understood properly through the following illustration-

Example 22.1

Individual	*Marks in the subject of History*	*Marks in the subject of Civics y signs +ve or –ve* $d = R_1 - R_2$	*Rank in History* R_1	*Rank in Civics* R_2	*Difference in Rank irrespective of*	*Difference squared* D^2
A	80	82	2	3	1	1
B	45	86	11	2	9	81
C	55	50	10	10	0	0
0	56	48	9	191	2	4
E	58	60	8	9	1	12
F	60	62	7	8	1	1
G	65	64	6	7	1	1
H	68	65	5	6	1	1
I	20	70	4	5	1	1
J	75	74	3	4	1	1
K	85	90	1	1	0	1

$N = 11, \Sigma d^2 = 92$

Formula
$$p = 1-\frac{6\Sigma d^2}{N(N^2-1)}$$
$$= 1-\frac{6\times 92}{11(11^2-1)} = 1-\frac{6\times 92}{11\times 120}$$
$$= 1-\frac{33}{55} = 1-42 = 0.58$$

Example 22.2

Individuals	*Scores in Test X*	*Scores in Test Y*	*Rank in X_1R_1*	*Rank in X_2R_2*	R_1–R_2 = *d*	D^2
A	12	21	8	6	2	4
B	15	25	6.5	3.5	3	9
C	24	35	2	2	0	0
D	20	24	4	5	1	1
E	8	16	10	9	1	1
F	15	18	6.5	7	0.5	0.25
G	20	25	4	3.5	0.5	0.25
H	20	16	4	9	5	25
I	11	16	9	9	0	0
J	26	38	1	1	0	0
N = 10						Σd^2 = 40.5

Formula
$$p = 1-\frac{6\Sigma d^2}{N(N^2-1)}$$
$$= 1-\frac{6\times 40.5}{10(10^2-1)} = 1-\frac{6\times 40.5}{10\times 99}$$
$$= 1-\frac{8.1}{33} = 1-0.245 = 0.755$$

Steps for Calculation of *p*

1. First of all it is required to assign position of merit or rank to each individual join either test. These ranks are put under column 3. (designated as R_1) and 4 (designated as R_2) respectively. The task of assigning ranks in the cases like example 1st is not difficult. But in the cases, like

example 2nd, where two or more individuals are found to achieve same score, some difficulty arises. In the above example, in the first test *X*, *B* and *F* are two individuals who have the same score i.e. 15. Therefore, score 15 occupies 6th position in order of merit. But now the question arises which one of the two individuals *B* and *F* should be ranked as 6th or 7th. In order to overcome this difficulty we equally share the rank 5th and 7th between them and thus rank each one of them as 6.5.

Similarly, if there are three persons who have the same score and share the same ranks, we take the average of the ranks claimed by these persons. For example, we can take the score 20 in the 2nd example which is shared by three individuals *D*, *G* and *H*. It is ranked third in the whole series and therefore the ranks 3, 4 and 5 are shared equally by *D*, *G* and *H* and hence we attribute rank to each of them.

2. After writing down the allotted rank to all the individuals 0 either of the two tests, the differences in these ranks are calculated. In doing so we do not 'consider the algebraic signs +ve or –ve of the difference. This difference is written under column 5th (designated as $|d|$.
3. In the next column (designated as d^2) we square up the Rank difference or the values of *d* written in the column 5th.
4. Now we calculate the total of all the values of d^2 and this sum is designated as Σd^2 .
5. Now the value of *p* is calculated by the formula $\rho = 1 - \frac{6\Sigma d^2}{N(N^2 - 1)}$, where d^2 stands for the sum of the squares of differences between the ranks of the scores on two different tests and N for the number of individuals whose scores are under consideration for computing.

(*b*) Product Moment Method of Computing Coefficient of Correlation

This method is also known as Pearson Moment method in the honour of the English statesman Karl Pearson who is said to be the inventor of this method. The coefficient of correlation computed by this method is known as Product Moment coefficient of correlation symbolically represented by '*r*'.

(*a*) The calculation of '*r*' from ungrouped data

The basic formula for the computation of '*r*' for the ungrouped data by this method is $r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}}$

Where *x* and *y* represent the deviation of scores in the tests *X* and *Y* from the means of each distribution.

The procedure of calculation *r* by this formula can be understood by the following illustration-

Individuals	*Scores in*	*Scores in*	*X*	*Y*	*Xy*	X^2	Y^2
A	15	60	–10	10	–100	100	100
B	25	70	0	20	0	0	400
C	20	40	–5	–10	50	25	190
D	30	50	5	0	0	25	0
E	35	30	–10	–20	–200	100	400
					$\Sigma xy = -250$	$\Sigma x^2 = 250$	$\Sigma y^2 = 1000$

Mean of series X, $(M)_x = 25$
Mean of series Y $(M)_y = 50$

Formula $$r = 1 - \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}}$$

$$= \frac{-250}{\sqrt{250 \times 1000}} = \frac{-250}{\sqrt{250000}}$$

$$= \frac{-250}{500} = \frac{1}{2} = -0.5$$

(*b*) Computation of *r* directly from raw scores when deviations are taken from zero (without calculating deviations from the means. Here we apply the formula.

Subject	*Scores in 1st Test (X)*	*Scores in 2nd Test (Y)*		X^2	Y^2
A	5	12	60	25	144
B	3	15	45	2	225
C	2	11	22	4	121
D	8	10	80	64	100
E	6	18	108	36	324

$$r = \frac{5 \times 15 - 24 - 66}{\sqrt{(5 \times 138 - 576)} \times \sqrt{(5 + 914 - 66 \times 66)}}$$

$$= \frac{1575 - 1650}{\sqrt{(690 - 576)} \times \sqrt{(1470 - 4356)}}$$

$$= \frac{-75}{\sqrt{24396}} = \frac{-75}{1562} = -0.48$$

Ans. R = – 0.48

EXERCISES

Essay Type Questions

1. What is correlation in Statistics? Discuss its types. How is it useful in the field of education?
2. Find the correlation between the following two sets of scores using product moment method.

Subiect	A	B	C	D	E	F	G	H
Test X	15	18	22	17	19	20	16	21
Test Y	40	42	50	45	43	46	41	41

Ans. $r = 0.65$

3. Find the correlation between the following two set of raw scores without computing deviation from the mean.

4.

Subiect	A	B	C	0	E	F	G	H	I	T
Test X	13	12	10	8	7	6	6	4	3	1
Test Y	7	11	3	7	2	12	6	2	9	6

Ans. $r = 0.14$

4. Compute the coefficient of correlation between the following two series of test scores by Rank difference method.

(a) Pupils	*Test X*	*Test Y*	*(b) Pupils*	*Test X*	*Test Y*
	40	42		12	16
B	36	35	B	26	25
C	27	28	C	21	15
D	18	27	D	23	21
E	13	15	E	25	22
F	48	48	F	15	21
G	43	50	G	18	27
H	25	27	H	22	30
I	29	32	I	18	28
T	17	21	J	19	23

Ans. (*a*) 0.985, (*b*) 0.188

5. Find the correlation coefficient between the following set of scores using product moment method.

Subject	A	B	C	D	E	F	G	H	I	J
Test X	13	12	10	10	8	6	6	5	3	2
Test Y	11	14	11	7	9	11	3	7	6	1

Ans. 0.76

6. Find the Rank correlation coefficient from the following data.

Subject	A	B	C	D	E	F	G	H
Test *X*	30	40	50'	20	10	45	22	18
I Test *Y*	55	75	60	12	11	38	25	15

Ans. 0.86

7. What is coefficient of correlation? Calculate the coefficient of correlation by Rank order Method in the following groups.

Group (X)	50	62	68	69	73	73	78	81
Group (Y)	24	20	22	18	18	18	19	10

Ans. – 0.756

8. Calculate the coefficient of correlation by Rank difference method from the following data.

Students	A	B	C	0	E	F	G
Test X	32	28	35	26	22	20	30
Test Y	27	25	26	22	15	18	24

Subject	A	B	C	D	E	F	G	H
Test X	20	22	24	18	27	30	28	23
Test Y	35	40	32	30	38	39	34	33

Ans. 0.357

10. What is coefficient of correlation? Discuss in brief about the Rank difference and product moment methods of computing coefficient of correlation.
11. What is Rank difference method of computing coefficient of correlation? Discuss its procedure with the help of hypothetical example.
12. What is product moment method of computing coefficient of correlation? Discuss its procedure with the help of hypothetical example.

Short Answer Type Question (Answer in 100-120 words)

1. What is correlation? Explain the need of finding correlation between two variables in the field of education.
2. What is linear correlation? Mention its various types.
3. What is coefficient of correlation? Name the different methods of computing coefficient of correlation.
4. What is the Rank difference method of computing coefficient of correlation?
5. What is the product moment method of computing coefficient of correlation?
6. Mention and explain the formula for computing Rank correlation coefficient.
7. Mention and explain the formula for computing correlation coefficient by Product Moment method.