Business Mathematics & Statistics

NCWEB Hansraj Centre

E-Class on 13/03/2020

Topics discussed:

- Karl Pearson's Coefficient of Correlation
- Spearman's Rank Correlation

By Ms.Komal Chhikara
Bcom Sem II
Section A

Karl Pearson's Correlation Coefficient

Karl Pearson's coefficient of correlation (r) is one of the mathematical methods of measuring the degree of correlation between any two variables X and Y is given as:

$$r = \frac{\sum dxdy}{\sqrt{\sum dx^2}\sqrt{\sum dy^2}} \qquad (1)$$

Where $dx = X - \bar{X}$; $dy = Y - \bar{Y}$, $dx^2 = (X - \bar{X})^2$ and, $dy^2 = (Y - \bar{Y})^2$

The following is the alternative formula,

$$r = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{N}}{\sqrt{\sum X^2 - \frac{(\sum X)^2}{N}}\sqrt{\sum Y^2 - \frac{(\sum Y)^2}{N}}}$$

Before we proceed to take up an illustration for measuring the degree of correlation, it is worthwhile to note some of the following important points.

i) ‘r’ is a dimensionless number whose numerical value lies between +1 to –1. The value +1 represents a perfect positive correlation, while the value –1 represents a perfect negative correlation. The value 0 (zero) represents lack of correlation. Figure 15.1 shows a number of scatter plots with corresponding values for correlation coefficient.

ii) The coefficient of correlation is a pure number and is independent of the units of measurement of the variables.

iii) The correlation coefficient is independent of any change in the origin and scale of X and Y values.

Question. A teacher is interested in studying the relationship between the performance in Statistics and Economics of a class of 20 students. For this he compiles the scores on these subjects of the students in the last semester examination. Some data of this type are presented in Table. Calculate correlation coefficient for the data.

Serial	Score in		Serial	Score in	
Number	Statistics	Economics	Number	Statistics	Economics
1	82	64	11	76	58
2	70	40	12	76	66
3	34	35	13	92	72
4	80	48	14	72	46
5	66	54	15	64	44
6	84	56	16	86	76
7	74	62	17	&I	52
8	84	66	18	60	40
9	60	52	19	82	60
10	86	82	20	90	60

Observation No.	X	Y	X^2	Y^2	$\overline{XY}$
1	82	64	6724	4096	5248
2	70	40	4900	1600	2800
3	34	35	1156	1225	1190
4	80	48	6400	2304	3840
5	66	54	4356	2916	3564
6	84	56	7056	3136	4704
7	74	62	5476	3844	4588
8	84	66	7056	4356	5544
9	60	52	3600	2704	3120
10	86	82	7396	6724	7052
11	76	58	5776	3364	4408
12	76	66	5776	4356	5016
13	92	72	8464	5184	6624
14	72	46	5184	2116	3312
15	64	44	4096	1936	2816
16	86	76	7396	5776	6536
17	84	52	7056	2704	4386
18	60	40	3600	1600	2400
19	82	60	6724	3600	4920
20	90	60	8100	3600	5400
Total	**1502**	**1133**	**116292**	**67141**	**87450**

$$\bar{X} = \frac{\Sigma X}{N} = \frac{1502}{20} = 75.1;$$

$$\bar{Y} = \frac{\Sigma Y}{N} = \frac{1133}{20} = 56.65;$$

$$\sigma_{\overline{x}} = \frac{1}{N}\sqrt{\Sigma X^2 - \frac{(\Sigma X)^2}{N}} = \frac{1}{20}\sqrt{116292 - \frac{(1502)^2}{20}} = \sqrt{174.59}; = 13.21;$$

$$\sigma_{\overline{y}} = \frac{1}{N}\sqrt{\Sigma Y^2 - \frac{(\Sigma Y)^2}{N}} = \frac{1}{20}\sqrt{67141 - \frac{1133^2}{20}} = \sqrt{147.83}; = 12.16;$$

$$\sigma xy = \frac{1}{N}\left[\Sigma XY - \frac{(\Sigma X)(\Sigma Y)}{N}\right] = \frac{1}{20}\left[87450 - \frac{1502 \times 1133}{20}\right] = 118.09$$

Thus using formula i.e.

$$r = \frac{\sigma xy}{\sigma x \ \sigma y}$$

$$r = \frac{118.09}{13.21 \ \times 12.16} = 0.735$$

Spearman's Rank Correlation

The Karl Pearson's correlation coefficient is not applicable in cases where the direct quantitative measurement of a phenomenon under study is not possible. Sometimes we are required to examine the extent of association between two ordinally scaled variables such as two rank orderings. For example, we can study intelligence, efficiency, performance, competitive events, attitudinal surveys etc. In such cases, a measure to ascertain the degree of association between the ranks of two variables, X and Y, is called Rank Correlation. It was developed by Edward Spearman , its coefficient (R) is expressed by the following formula:

$R = 1 - \frac{6\sum D^2}{N^3 - N}$ where, N = Number of pairs of ranks, and $\sum D^2$= squares of difference between the ranks of two variables.

Question. Salesmen employed by a company were given one month training. At the end of the training, they conducted a test on 10 salesmen on a sample basis who were ranked on the basis of their performance in the test. They were then posted to their respective areas. After six months, they were rated in terms of their sales performance. Find the degree of association between them.

Salesmen:	1	2	3	4	5	6	7	8	9	10
Ranks in training (X):	7	1	10	5	6	8	9	2	3	4
Ranks on sales Peformance (Y):	6	3	9	4	8	10	7	2	1	5

Solution: Table : Calculation of Coefficient of Rank Correlation.

Sales men	Ranks Secured in Training X	Ranks Secured on Sales Y	Difference in Ranks D = (X–Y)	D2
1	7	6	1	1
2	1	3	–2	4
3	10	9	1	1
4	5	4	1	1
5	6	8	–2	4
6	8	10	–2	4
7	9	7	2	4
8	2	2	0	0
9	3	1	4	4
10	4	5	–1	1
				$\sum D^2 = 24$

Using the spearman's formula, we obtain

$$R = 1 - \frac{6\sum D^2}{N^3 - N} = 1 - \frac{6\sum 24}{10^3 - 10}$$

$$= 1 - \frac{144}{990} = 0.855$$

we can say that there is a high degree of positive correlation between the training and sales performance of the salesmen.

Like Karl Pearson's coefficient of correlation the Spearman's rank correlation has a value + 1 for perfect matching of ranks, -1 for perfect mismatching of ranks and 0 for the lack of relation between the ranks.

Sometimes the data, relating to qualitative phenomenon, may not be available in ranks, but only in values. In such a situation it is necessary to assign the ranks to the values. Ranks may be assigned by taking either from largest to the smallest or vice versa. But the same method must be followed in case of both the variables.

Sometimes there is a tie between two or more ranks in the first and/or second series. **For example**, if the values of two items are same and presume that the rank of one item may be 4th rank, then instead of awarding 4th rank to the respective two observations, we award 4.5 [(4+5)/2] for each of the two observations.

Now we will take up an illustration to understand how to award the ranks when the data is given in values and to calculate the rank, correlation. The illustration will also give clarity how to award the ranks when values of items in series are same.

Question: Calculate rank correlation from the following data related to a group of 10 students and percentage of marks secured.

Roll Nos. of the students	21	22	23	24	25	26	27	28	29	30
% of marks in statistics	45	66	55	45	80	75	50	55	60	45
% of marks in Accountancy	70	81	75	75	70	85	65	80	45	60

The above data was given in percentage of marks not in the ranks. Therefore, for calculation of rank correlation, first, we have to assign the ranks to the given values. As we discussed earlier the ranks may be assigned either from the largest value to smallest value or visa-versa. Here, we assign the ranks from largest to smallest value which is normally in practice.

Calculation of rank correlation:

Roll Nos.	% of marks in statistics	% of marks in Accountancy	Ranks of % of marks in Statistics	Ranks of Marks in Accountancy	Difference in Ranks D	D^2
21	45	70	9	6.5	2.5	6.25
22	66	81	3	2	1	1.00
23	55	75	5.5	4.5	1	1.00
24	45	75	9	4.5	4.5	20.25
25	80	70	1	6.5	−5.5	30.25
26	75	85	2	1	1	1.00
27	50	65	7	8	−1	1.00
28	55	80	5.5	3	2.5	6.25
29	60	45	4	10	−6	36.00
30	45	60	9	9	0	0
						ΣD^2= 103.00

$$r = 1 - \frac{6(\Sigma D^2)}{N^3 - N} = 1 - \frac{103}{10^3 - 10} = 1 - \frac{103}{990} = 1 - 0.10 = 0.90$$

Explanation of assigning ranks:

For the values of percentage of marks in statistics for 80, 75, 66, 60 there are only single values. Therefore, ranks have been assigned 1,2,3,4. Whereas the next value 55 repeated two times in the data, therefore, 5 + 6 ranks divided by 2 = 5.5 rank has been allotted to the value of 55 two times.

Similarly, the value of 45 repeated three times in the data, therefore the ranks 8+9+10 divided by 3 equal to 9. Accordingly, the rank 9 has been allotted to value of 45 (in between) value of 55, 45. There is a value of 50, hence rank seven has been allotted to 50. In the same manner, you may try to observe the assigning of ranks to the values of percentage of marks in accountancy.

Question for practice:

Ten competitors in a musical contest were ranked by 3 judges, A, B and C in the following order:

Competitors:	1	2	3	4	5	6	7	8	9	10
Rank by A	1	6	5	10	3	2	4	9	7	8
Rank by B	3	5	8	4	7	10	2	1	6	9
Rank by C	6	4	9	8	1	2	3	10	5	7

Using rank correlation method, discuss which pair of judges has the nearest approach to common liking in music.

Thank You