Frequently Asked Questions

1. When do you use Spearman's rank correlation coefficient?

Answer:
Spearman's formula is the only formula to be used for finding correlation coefficient if we are dealing with qualitative characteristics, which cannot be measured quantitatively but can be arranged serially. It can also be used where actual data are given. In case of extreme observations, Spearman's formula is preferred to Pearson's formula.

2. Write the formula used for calculating the spearman's rank correlation coefficient.

Answer:

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^{\,2}}{n(n^2-1)}$$

3. How do you interpret the Spearman's rank correlation coefficient?

Answer:
Since Spearman's rank correlation coefficient row is same as Pearsonian correlation coefficient between the ranks, it can be interpreted in the same way as the Karl Pearson's correlation coefficient.

4. Derive the formula for spearman's rank correlation coefficients when there are no ties.

Answer:
Assuming that no two individuals are bracketed equal in either classification, each of the variables X and Y takes the values 1, 2… n. Hence

$$\bar{x} = \bar{y} = \frac{1}{n}(1+2+3+...+n) = \frac{n+1}{2}$$

$$\sigma_x^{\,2} = \frac{1}{n}\sum_{i=1}^{n} x_i^{\,2} - \bar{x}^2 = \frac{1}{n}(1^2+2^2+...+n^2) - \left(\frac{n+1}{2}\right)^2 = \frac{n(n+1)(2n+1)}{6n} - \left(\frac{n+1}{2}\right)^2 = \frac{n^2-1}{12}$$

In general, $x_i \neq y_i$. Let $d_i = x_i - y_i$

$$\therefore d_i = (x_i - \bar{x}) - (y_i - \bar{y})$$

Squaring and summing over i from 1 to n we get,

$$\sum_{i=1}^{n} d_i^{\,2} = \sum_{i=1}^{n}\left[(x_i-\bar{x})-(y_i-\bar{y})\right]^2 = \sum_{i=1}^{n}(x_i-\bar{x})^2 + \sum_{i=1}^{n}(y_i-\bar{y})^2 - 2\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$$

Dividing both sides by n, we get,

$$\frac{1}{n}\sum_{i=1}^{n} d_i^{\,2} = \sigma_X^{\,2} + \sigma_Y^{\,2} - 2Cov(X,Y) = \sigma_X^{\,2} + \sigma_Y^{\,2} - 2\rho\sigma_X\sigma_Y$$

, where, ρ is the rank correlation coefficient between A and B.

$$\therefore \frac{1}{n}\sum d_i^{\,2} = 2\sigma_X^{\,2} - 2\rho\sigma_X^{\,2} \Rightarrow 1-\rho = \frac{\sum_{i=1}^{n} d_i^{\,2}}{2n\sigma_X^{\,2}} \Rightarrow \rho = 1 - \frac{\sum_{i=1}^{n} d_i^{\,2}}{2n\sigma_X^{\,2}} = 1 - \frac{6\sum_{i=1}^{n} d_i^{\,2}}{n(n^2-1)}$$

5. Derive the formula for spearman's rank correlation coefficient when there are tied ranks.

Answer:

If some of the individuals receive the same rank in a ranking of merit, they are said to be tied. Suppose m of the individuals say $(k+1)^{th}$, $(k+2)^{th}$, . . . $(k+m)^{th}$, are tied, then each of these m individuals is assigned a common rank, which is arithmetic mean of the ranks k+1, k+2, …,k+m.

We have

$$\rho(X,Y) = \frac{\Sigma(X-\overline{X})(Y-\overline{Y})}{[\Sigma(X-\overline{X})^2\Sigma(Y-\overline{Y})^2]^{1/2}} = \frac{\Sigma xy}{\Sigma x^2 \Sigma y^2}$$

$$x = (X-\overline{X}), y = (Y-\overline{Y})$$

Where

If X and Y each takes values 1, 2, …, n then

$$\overline{X} = \overline{Y} = \frac{1}{n}(1+2+3+...+n) = \frac{n+1}{2}$$

$$n\sigma_X^{\,2} = \Sigma x^2 = \frac{n(n^2-1)}{12}, \; n\sigma_Y^{\,2} = \Sigma y^2 = \frac{n(n^2-1)}{12}$$

And

$$\Sigma d^2 = \Sigma(X-Y)^2 = \Sigma[(X-\overline{X})-(Y-\overline{Y})]^2 = \Sigma(x-y)^2$$

Also

$$\Rightarrow \Sigma d^2 = \Sigma x^2 + \Sigma y^2 - 2\Sigma xy \Rightarrow \Sigma xy = \frac{1}{2}(\Sigma x^2 + \Sigma y^2 - \Sigma d^2)$$

We shall now investigate the effect of common ranking (in case of ties), on the sum of squares of the ranks. Let S^2 and S_1^2 denote the sum of squares of untied and tied ranks respectively. Then we have,

$$S^2 = (k+1)^2 + (k+2)^2 + ... + (k+m)^2$$

$$= mk^2 + (1^2 + 2^2 + ... + m^2) + 2k(1+2+...+m)$$

$$S_1^{\,2} = m(AverageRank)^2 = m\left[\frac{(k+1)+(k+2)+...(k+m)}{m}\right]^2$$

$$= mk^2 + \frac{m(m+1)(2m+1)}{6} + km(m+1)$$

$$= m\left[k + \frac{m+1}{2}\right]^2 = mk^2 + \frac{m(m+1)^2}{4} + mk(m+1)$$

$$\therefore S^2 - S_1{}^2 = \frac{m(m+1)}{12}[2(2m+1) - 3(m+1)] = \frac{m(m^2-1)}{12}$$

Thus the effect of typing m individuals (ranks) is to reduce the sum of squares by $m(m^2-1)/12$, though the mean value of the ranks remains the same, viz., (n+1)/2

$$\frac{1}{12}\sum_{i=1}^{s} m_i(m_i{}^2 - 1) = \frac{1}{12}\sum_{i=1}^{s}(m_i{}^3 - m_i) = T_X\,(Say)$$

Suppose there are s such sets of ranks to be tied in the X –series so that the total sum of squares due to them is,

$$\frac{1}{12}\sum_{j=1}^{t} m_j{}'(m_j{}'^2 - 1) = \frac{1}{12}\sum_{i=1}^{t}(m_j{}'^3 - m_j{}') = T_Y\,(Say)$$

Similarly, suppose that there are t such sets of ranks to be tied with respect to the other series Y so that sum of squares is due to them is,

Thus, in case of ties, the new sum of squares are given by,

$$nVar'(X) = \Sigma x^2 - T_X = \frac{n(n^2-1)}{12} - T_X$$

,

$$nVar'(Y) = \Sigma y^2 - T_Y = \frac{n(n^2-1)}{12} - T_Y$$

And

$$nCo\,\mathrm{var}'(X,Y) = \frac{1}{2}\left(\Sigma x^2 - T_X + \Sigma y^2 - T_Y - \Sigma d^2\right)$$

$$= \frac{1}{2}\left[\frac{n(n^2-1)}{12} - T_X + \frac{n(n^2-1)}{12} - T_Y - \Sigma d^2\right]$$

$$= \frac{n(n^2-1)}{12} - \frac{1}{2}\left[T_X + T_Y + \Sigma d^2\right]$$

$$= \frac{\frac{n(n^2-1)}{6} - \left[T_X + T_Y + \Sigma d^2\right]}{\left[\frac{n(n^2-1)}{6} - 2T_X\right]^{1/2}\left[\frac{n(n^2-1)}{6} - 2T_Y\right]^{1/2}}$$

$$\rho(X,Y) = \frac{\frac{n(n^2-1)}{12} - \frac{1}{2}\left[T_X + T_Y + \Sigma d^2\right]}{\left[\frac{n(n^2-1)}{12} - T_X\right]^{1/2}\left[\frac{n(n^2-1)}{12} - T_Y\right]^{1/2}}$$

6. Give the formula for finding the spearman's rank correlation coefficient when there are ties.

Answer:

$$\rho(X,Y) = 1 - \frac{6(\Sigma d^2 + T_X + T_y)}{n(n^2-1)}$$

7. How do you choose between product moment (Karl Pearson's) correlation coefficient and Spearman's rank correlation coefficient?

Answer:
Karl Pearson's correlation coefficient assumes that the parent population from which sample observations are drawn is normal. If this assumption is violated, then we need a measure, which is distribution free or non-parametric. A distribution free measure is one, which does not make any assumptions about the parameters of the population. Spearman's ρ is such a measure, since no assumptions are made about the form of the population from which sample observations are drawn.

8. What are the limitations of Spearman's rank correlation Coefficient?

Answer:
Spearman's formula has its limitations also. It is not practicable in the case of bivariate frequency distribution (correlation table). For n>30, this formula should not be used unless the ranks are given, since in the contrary case the calculations are quite time-consuming.

9. Who discovered the formula for finding rank correlation coefficient?

Answer:
Spearman had discovered the formula for finding rank correlation coefficient.

10. Among Karl Pearson's and Spearman's correlation coefficient, which is superior?

Answer:
Spearman's formula is easy to understand and apply as compared with product moment formula. The value obtained by the two formulae, viz., Pearsonian r and spearman's ρ, are generally different. The differences arise due to the fact that when ranking is used

instead of full set of observations, there is always some loss of information. Unless many ties exist, the coefficient of rank correlation should be only slightly lower than the Pearsonian Coefficient.

11. What are the limits for the rank correlation coefficient?

Answer:
The coefficient of rank correlation lies between -1 and +1.

12. Show that when the ranks of the individuals in the two characteristics are in the opposite direction, the value of spearman's rank correlation coefficient is -1

Answer:

X	1	2	3	…	…	n-1	n
Y	n	n-1	n-2	…	…	2	1

Case (i). Suppose n is odd. ie., n=2m+1, then the values of d are

$$\therefore \sum_{i=1}^{n} d_i^{\,2} = 2[(2m)^2 + (2m-2)^2 + ... + 4^2 + 2^2]$$

d: 2m, 2m-2, 2m-4, …, 2, 0, -2, -4, … , -(2m-2), -2m.

$$= \frac{8m(m+1)(2m+1)}{6}$$

$$= 8[m^2 + (m-1)^2 + ... + 2^2 + 1^2]$$

Hence

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^{\,2}}{n(n^2-1)} = 1 - \frac{8m(m+1)(2m+1)}{(2m+1)\{(2m+1)^2 - 1\}}$$

$$= 1 - \frac{8m(m+1)}{4m(m+1)} = -1$$

Case (ii): let n be even. ie., n=2m

Then the values of d are, (2m-1), (2m-3),…, 1, -1, -3, … , (2m-3), (2m-1).

$$\therefore \sum_{i=1}^{n} d_i^{\,2} = 2[(2m-1)^2 + (2m-3)^2 + ... + 3^2 + 1^2]$$

$$-\{(2m)^2 + (2m-2)^2 + ... + 4^2 + 2^2\}] = 2[\{(2m)^2 + (2m-1)^2 + (2m-2)^2 ... + 3^2 + 2^2 + 1^2\}$$

$$=2\left[\frac{2m(2m+1)(4m+1)}{6}-4\frac{m(m+1)(2m+1)}{6}\right]$$

$$=2[\{1^2+2^2+3^2+...+(2m)^2\}-\{2^2m^2+2^2(m-1)^2+...+2^2\}]$$

$$=\frac{2m}{3}(2m+1)(2m-1)=\frac{2m(4m^2-1)}{3}=\frac{2m}{3}\left[(2m+1)(4m+1)-2(m+1)(2m+1)\right]$$

Therefore, the expression for ρ becomes

$$\rho=1-\frac{6\sum_{i=1}^{n}d_i^{\,2}}{n(n^2-1)}=1-\frac{4m(4m^2-1)}{2m(4m^2-1)}=-1$$

13. When the spearman's rank correlation coefficient takes the maximum value?

Answer:

Spearman's Rank correlation coefficient is given by

$$\rho=1-\frac{6\sum_{i=1}^{n}d_i^{\,2}}{n(n^2-1)}$$

ρ is maximum, if $\sum_{i=1}^{n}d_i^{\,2}$ is minimum, .i.e. if each of the deviations d_i is minimum. However, the minimum value of d_i is zero in the particular case $xi=y_i$, i.e. if the ranks of the i^{th} individual in the two characteristics are equal. Hence the maximum value of ρ is+1.

14. Derive limits for Spearman's rank correlation coefficient.

Answer:

Spearman's Rank correlation coefficient is given by

$$\rho=1-\frac{6\sum_{i=1}^{n}d_i^{\,2}}{n(n^2-1)}$$

ρ is maximum, if $\sum_{i=1}^{n}d_i^{\,2}$ is minimum, i.e. if each of the deviations d_i is minimum. However, the minimum value of d_i is zero in the particular case $xi=y_i$, i.e. if the ranks of the i^{th} individual in the two characteristics are equal. Hence the maximum value of ρ is+1, i.e. ρ≤1. ρ' is minimum if $\sum_{i=1}^{n} d_i^2$ is maximum. i.e. if each of the deviations d_i is

maximum which is so if the ranks of the individuals in the two characteristics are in the opposite direction as follows.

X	1	2	3	…	…	n-1	n
Y	n	n-1	n-2	…	…	2	1

Case (i). Suppose n is odd. ie., n=2m+1, then the values of d are

$$\therefore \sum_{i=1}^{n} d_i^2 = 2[(2m)^2 + (2m-2)^2 + ... + 4^2 + 2^2]$$

d: 2m, 2m-2, 2m-4, …, 2, 0, -2, -4, … , -(2m-2), -2m.

$$= \frac{8m(m+1)(2m+1)}{6}$$

$$= 8[m^2 + (m-1)^2 + ... + 2^2 + 1^2]$$

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} = 1 - \frac{8m(m+1)(2m+1)}{(2m+1)\{(2m+1)^2 - 1\}}$$

$$= 1 - \frac{8m(m+1)}{4m(m+1)} = -1$$

Hence

Case (ii): let n be even. ie., n=2m

Then the values of d are, (2m-1), (2m-3),…, 1, -1, -3, … , (2m-3), (2m-1).

$$\therefore \sum_{i=1}^{n} d_i^2 = 2[(2m-1)^2 + (2m-3)^2 + ... + 3^2 + 1^2]$$

$$-\{(2m)^2+(2m-2)^2+...+4^2+2^2\}] = 2[\{(2m)^2+(2m-1)^2+(2m-2)^2...+3^2+2^2+1^2\}$$

$$= 2\left[\frac{2m(2m+1)(4m+1)}{6} - 4\frac{m(m+1)(2m+1)}{6}\right]$$

$$= 2[\{1^2+2^2+3^2+...+(2m)^2\} - \{2^2m^2+2^2(m-1)^2+...+2^2\}]$$

$$= \frac{2m}{3}(2m+1)(2m-1) = \frac{2m(4m^2-1)}{3} = \frac{2m}{3}\left[(2m+1)(4m+1) - 2(m+1)(2m+1)\right]$$

Therefore, the expression for ρ becomes

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} = 1 - \frac{4m(4m^2-1)}{2m(4m^2-1)} = -1$$

Thus, the limits for rank correlation coefficient are given by, -1≤ρ≤ 1

15. How to rank the individuals when more than one has same ranks?

Answer:

If two or more individual have same values in the series, common ranks are given to them. This common rank is the average of the ranks, which these items would have been assumed if they were slightly different from each other and the next item will get the rank next to the ranks already assumed.