Kenya Medical Training College Department of Clinical Medicine Year Two Semester One Spearman's Correlation Coefficient: Worked Examples 26th November 2020

Willis J. Opalla

Learning Objective

- To apply Spearman's Correlation Coefficient in determining the relationship between independent and dependent variables.

Learning Outcomes

- By the end of this session, you should be able to
  1. Explain the Spearman's Correlation Coefficient formula.
  2. Use Spearman's Correlation Coefficient to make statistical inferences on relationship between independent and dependent variables.

Spearman's Correlation Coefficient

- For finding correlation between two variables by taking their ranks.
  - Useful for qualitative data.
  - Can be used when the actual magnitude of characteristics under consideration is not known, but relative position or rank of the magnitude is known.
  - Is the nonparametric version of the Pearson correlation coefficient.
  - The data must be ordinal, interval or ratio with ranks.

Spearman's Correlation Coefficient

- For summarizing the strength and direction (negative or positive) of a relationship between two variables.
- Will always be between +1 and -1, where:
- +1 = a perfect positive correlation between ranks.
- -1 = a perfect negative correlation between ranks.
- 0 = no correlation between ranks.

Spearman's Correlation Coefficient

- It is denoted by " rho" (ρ).
- There are two cases for calculating rank correlation.
- Case 1.
  - No tie of allotted rank
- Case 2.
  - There is a tie for two or more values or ranks in either "x" or "y" or in both "x" and "y".

Case 1: No Tie of Allotted Rank

- In this, none of the values/ranks of x and y are repeated.
- "*p*" can be calculated using the formula:

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

d = difference in the ranks of data set of 'x' and 'y'.

i.e. d = Rx – Ry (i.e. d = rank x – rank y)

Spearman's Correlation Coefficient

- The formula for the Spearman rank correlation coefficient when there are no tied ranks is:

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

- **Example Question:**

  Pathology and physiology scores for nine students are:

  - Pathology: 35, 23, 47, 17, 10, 43, 9, 6, 28
  - Physiology: 30, 33, 45, 23, 8, 49, 12, 4, 31

- Compute the student's ranks in the 2 subjects and compute the Spearman's rank correlation.

Spearman's Correlation Coefficient

- Step 1: Find the ranks for each subject.
  - To rank manually by hand, the scores are ordered from the largest to smallest then assigned the rank 1 to the highest score, 2 to the next highest etc:

Pathology	Rank	Physiology	Rank
35	3	30	5
23	5	33	3
47	1	45	2
17	6	23	6
10	7	8	8
43	2	49	1
9	8	12	7
6	9	4	9
28	4	31	4

Spearman's Correlation Coefficient

- Step 2: Add a third column, d.
  - The d is the difference between ranks.
  - e.g., the first student's pathology rank is 3 and physiology rank is 5, so the difference is 3 points.
- In the 6th column, square the values of d.

Pathology	Rank	Physiology	Rank	d	d^2
35	3	30	5	2	4
23	5	33	3	2	4
47	1	45	2	1	1
17	6	23	6	0	0
10	7	8	8	1	1
43	2	49	1	1	1
9	8	12	7	1	1
6	9	4	9	0	0
28	4	31	4	0	0

Spearman's Correlation Coefficient

- Step 3: Add up all values of d^2.
  $4 + 4 + 1 + 0 + 1 + 1 + 1 + 0 + 0 = 12$.

  This will be required for the factor $6\Sigma d^2$ of the formula.
- Step 4: Insert the values into the formula. These ranks are not tied (i.e. not similar) so the first formula is applied:

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

Spearman's Correlation Coefficient

- Substituting:

$$p = 1 - (6x12)/\left[9(81-1)\right]$$
$$= 1 - 72/720$$
$$= 1 - 0.1$$
$$= 0.9$$

The Spearman's Correlation for this set of data is 0.9, hence implying a strong positive correlation.

Worked Example

- Calculate the rank correlation of the marks for five students in Microbiology and Immunology.
  - Only the ranks should be arranged in ascending or descending order.
  - One data pair belongs to one student.
  - Prepare a table to calculate Σd^2

Microbiology 85 81 77 68 53

Immunology 78 70 72 62 67

Microbiology	Rank	Immunology	Rank	d	d^2
85	1	78	1	0	0
81	2	70	3	1	1
77	3	72	2	1	1
68	4	62	5	1	1
53	5	67	4	1	1

Worked Example

- Substituting in the equation:

$$\rho = 1 - 6x4 / 5 (25-1)$$

$$= 1 - 24/120$$

$$= 0.8$$

The marks of the two subjects are strongly positively correlated.

Exercise

- Calculate the Spearman's correlation coefficient for the temperatures (^oC) of two patients, Adan and Kadzo on different days in one week.

  Adan 20 28 25 23 22 30 31

  Kadzo 15 26 17 19 21 24 27

- First step:

  Feed the data into a 6 x 8 table.

Adan	Rank	Kadzo	Rank	d	d^2

Case 2: Tie of Allotted Rank

- i.e. more than one rank is present in either x or y or both x and y.
- "ρ" is calculated using the Spearman's formula and then adding CF, the Correlation Factor.

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} + \text{CF}.$$

- CF has to be calculated for each repeated ranks and then added.
- The CF is calculated using $CF = m (m^2 - 1)/12$, where m is the number of times the data repeats.
- d = difference in the ranks of data set of 'x' and 'y' (d = Rx - Ry).

Case 2: Tied Ranks

- Calculate the rank correlation of the following marks obtained by eight students in Medicine and Obstetrics.
  - Medicine 60 81 72 68 53 75 85 68
  - Obstetrics 78 70 72 62 67 70 70 61
  - Here Medicine (x) the value 68 is repeated twice and in Obstetrics (y) the value 70 is repeated thrice.
  - In the first series CF = 2x(4-1)/12 = 0.5
  - In the second series CF = 3x(9-1)/12 = 2

Case 2: Tied Ranks

- Tabulating:

Medicine	Rank, Rx	Obstetrics	Rank, Ry	d	d^2
60	6	78	1	5	25
81	2	70	3	-1	1
72	4	72	2	2	4
68	5	62	5	0	0
53	7	67	4	3	9
75	3	70	3	0	0
85	1	70	3	-2	4
68	5	61	6	-1	1

- $\Sigma d^2 = 44$
- $n = 8$

Case 2: Tied Ranks

- Substituting:

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)} + CF$$

$p = 1 – 6x\ 44 + 0.5 + 2 / 8\ (64\text{-}1)$

$= 1 – 266.5/504$

$= 1\text{-}\ 0.53$

$= 0.47$

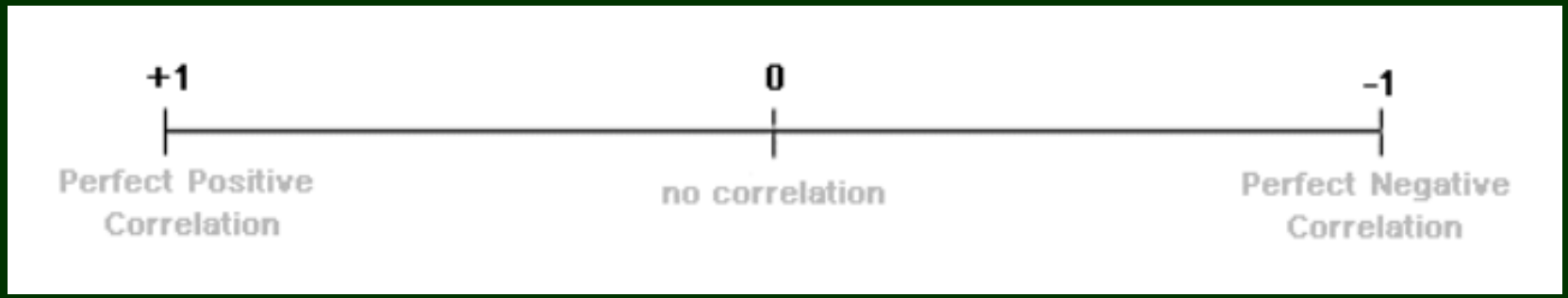
- The marks of the two subjects have a positive correlation.

Merits and Demerits of Spearman's Correlation Coefficient

- Merits
  - Can be used as a measure of degree of association between qualitative data.
  - Is very simple and easily understandable.
  - Can be used when the actual data is given or when only the ranks of the data are given.
- Demerits
  - Ranks coefficient for a frequency distribution, i.e., grouped data, can't be calculated.
  - Calculation gets tedious when the number of observations is large.

Meaning of p Value

- The closer p is to +1 or -1, the stronger the likely correlation.
- A perfect positive correlation is +1 and a perfect negative correlation is -1.
- A p value of -0.73 suggests a fairly strong negative relationship, i.e.



Summary

- Spearman's or Rank Correlation Coefficient is the nonparametric version of the Pearson correlation coefficient and is useful for determining the correlation between two variables by taking their ranks.
- Applicable for summarizing the strength and direction of a relationship between two variables in both tied and non-tied ranks.

References

- Glen, S..(2015) *Spearman's Rank Correlation (Spearman's Rho): Definition and How to Calculate it*, [Online] Available: https://www.statisticshowto.com/spearman-rank-correlation-definition-calculate/(Retrieved 23.11.2020)
- Joseph, J. K. (n.d) *Measures of Relationship*, [Online] Available: https://www.slideshare.net/JohnykuttyJoseph/measures-of-relationship, (Retrieved 26.11.2020)