Descriptive Statistics With R Software

Association of Variables
::
Rank Correlation Coefficient

Shalabh
Department of Mathematics and Statistics
Indian Institute of Technology Kanpur

Association between Ranks
Spearman‘s Rank Correlation Coefficient

Two variables : *X*, *Y*

***n* observations on *X* and *Y* are available.**

***n* observations are ranked with respect to *X* and *Y*.**

Ranks of the *n* observations are recorded.

Association between Ranks
Spearman‘s Rank Correlation Coefficient

Example:

***n* candidates participate in a talent competition.**

Two judges *X* and *Y* judge their performances and give ranks to every participant.

Association between Ranks
Spearman's Rank Correlation Coefficient

Judge *X* gives ranks to *n* candidates as

- **Rank 1 to worst candidate with lowest score x_i**
- **Rank 2 to candidate with second lowest score x_i**
- **….**
- **Rank *n* to the best candidate with highest score x_i.**

Similarly, judge *Y* give ranks to *n* candidates and gives ranks 1,2,...,*n* based on scores y_1, y_2,..., y_n. In the same way as judge *X* gave.

Association between Ranks
Spearman‘s Rank Correlation Coefficient

Every participant has two ranks given by two different judges.

We expect that both the judges give

- **higher ranks to good candidates and**
- **lower ranks to bad candidates.**

We want to measure the degree of association between the two different judgements, i.e., the two different set of ranks.

Association between Ranks
Spearman‘s Rank Correlation Coefficient

Measure the degree of agreement between the ranks of two judges.

Use Spearman‘s rank correlation coefficient.

Uses ranks of the values and not the values themselves.

Association between Ranks
Spearman's Rank Correlation Coefficient

***Rank*(x_i) : Rank of i^{th} observation on *X*.**

: Rank of x_i among ordered values $x_1, x_2, \ldots, x_n$ of *X*.

***Rank*(y_i) : Rank of i^{th} observation on *Y*.**

: Rank of y_i among ordered values $y_1, y_2, \ldots, y_n$ of *Y*.

$d_i = Rank(x_i) - Rank(y_i)$

Spearman's rank correlation coefficient (*R*) is defined as

$$R = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \quad ; \quad -1 \le R \le 1$$

Association between Ranks
Spearman's Rank Correlation Coefficient

It does not matter whether the ascending or descending order of ranks is used.

When both the judges assign exactly the

- **same ranks to all the candidates then $R = +1$**

- **opposite ranks to all the candidates then $R = -1$**

Association between Ranks
Spearman's Rank Correlation Coefficient

Example: Scores given by two judges to 5 candidates are as follows:

Candidates	Judge1		Judge2		$d_i = Rank(x_i) - Rank(y_i)$
	Scores (x_i)	Rank(x_i)	Scores(y_i)	Rank(y_i)	
1	75	4	70	4	0
2	25	1	80	5	-4
3	35	2	60	3	-1
4	95	5	30	1	4
5	50	3	40	2	1

Association between Ranks
Spearman‘s Rank Correlation Coefficient

$$n = 5$$

$$R = \frac{1 - 6\sum_{i=1}^{n} d_i^{\,2}}{n(n^2 - 1)} = \frac{1 - 6\sum_{i=1}^{5} d_i^{\,2}}{5(5^2 - 1)} = -0.7$$

Association between Ranks
Spearman‘s Rank Correlation Coefficient

R Command

`cor(x,y)` **computes the correlation between** `x` **and** `y`

```
cor(x, y, use = "everything", method =
c("spearman"))
```

`x` **: a numeric vector, matrix or data frame.**

`y` **: a numeric vector, matrix or data frame with compatible dimensions to** `x`**.**

Association between Ranks
Spearman‘s Rank Correlation Coefficient

`use` : an optional character string giving a method for computing covariances in the presence of missing values. This must be (an abbreviation of) one of the strings `"everything"`, `"all.obs"`, `"complete.obs"`, `"na.or.complete"`, or `"pairwise.complete.obs"`.

Association between Ranks
Spearman's Rank Correlation Coefficient

Example: Scores given by two judges to 5 candidates are as follows:

Candidates	Judge1	Judge2
	Scores (x_i)	Scores(y_i)
1	75	70
2	25	80
3	35	60
4	95	30
5	50	40

```
> x = c(75, 25, 35, 95, 50)
> y = c(70, 80, 60, 30, 40)
```

Association between Ranks
Spearman‘s Rank Correlation Coefficient

```
> judge1 = c(75, 25, 35, 95, 50)
> judge2 = c(70, 80, 60, 30, 40)

> cor(judge1, judge2, use = "everything",
method = c("spearman"))
[1] -0.7
```