Coefficient of Correlation (Spearman)

Correlation (Spearman)

- **Correlation** is a bivariate analysis that measures the strength of association between two variables and the direction of the relationship, strength of relationship.

- In terms of the strength of relationship, the value of the correlation coefficient varies between +1 and -1.

A correlation coefficient is a numerical measure or index of the amount of association between two sets of scores. It ranges in size from a maximum of +1.00 through 0.00 to -1.00

THE'+' / **POSITIVE CORRELATION**

The + sign indicates a positive correlation (the scores on one variable increase as the scores on the other variable increase)

THE '-'/ **NEGATIVE CORRELATION**

The - sign indicates a negative correlation (the scores on one variable increase, the scores on the other variable decrease)

Spearman Correlation Coefficient

- Spearman Correlation Coefficient is also referred to as Spearman Rank Correlation or Spearman's rho.
- Its named after *Charles Spearman*.
- It is typically denoted either with the *Greek letter* **rho (ρ)**, or r_s.
- It is one of the few cases where a Greek letter denotes a value of a sample and not the characteristic of the general population.
- Like all correlation coefficients, Spearman's rho measures the strength of association of two variables.
- Spearman's rho is given the symbol $\mathbf{r_s}$, with r used to denote that it is a *correlation coefficient*, and the subscript s to denote that it is named after the *statistician Spearman*.

When to use it..

- Spearman rank correlation is used when you have two ranked variables, and you want to see whether the two variables covary; whether, as one variable increases, the other variable tends to increase or decrease.
- Spearman rank correlation is used if you have one measurement variable and one ranked variable; in this case, you convert the measurement variable to ranks and use Spearman rank correlation on the two sets of ranks.
- The Spearman correlation evaluates the monotonic relationship between two continuous or ordinal variables. A monotonic relationship is a relationship that does one of the following:
  1. as the value of one variable increases, so does the value of the other variable; or
  2. as the value of one variable increases, the other variable value decreases.

  Spearman correlation is often used to evaluate relationships involving ordinal variables.(1st, 2nd, 3rd,/Small, Medium, Large, XL,) For example, you might use a Spearman correlation to evaluate whether the order in which employees complete a test exercise is related to the number of months they have been employed.
- Spearman correlation is used when the sample size is small. It’s a tedious job to give ranks to large sample (manually).

Merits/Demerits

Merits

- It is easy to calculate.
- It is simple to understand.
- Shows the significance of the data
- It can be applied to any type of data. qualitative or quantitative.
- Proves/disproves correlation
- Allows for further analysis
- Doesn't assume normal distribution
- This is most suitable in case there are two attributes.

Demerits

- It is only an approximately calculate measure as actual values are not used for calculations.
- For large samples, it is not a convenient method.
- Combined r of different series cannot be obtained as in case of mean and S.D.
- It cannot be treated further algebraically

The Spearman's Rank difference method / Spearman's Rank Correlation

- Coefficient is used to discover the strength of a link between two sets of data. It is a technique which can be used to summarise the strength and direction (negative or positive) of a relationship between two variables.
- The result will always be between 1 and minus 1.

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

- $\rho \;= rho\ coffecient\ of\ correlation$
- $\sum D^2$ = the sum of the difference between two rank orders
- n = the number of paired observations

Method/ Steps - calculating the coefficient

- Create a table from your data.
- **Rank** the two data sets. Ranking is achieved by giving the ranking '1' to the biggest number in a column, '2' to the second biggest value and so on.
  - The smallest value in the column will get the lowest ranking. This should be done for both sets of measurements.
- Tied scores are given the **mean (average) rank**. Notice if there is two or more identical values in the data (called a "tie"), you need to take the average of the ranks that they would have otherwise occupied.
  - For example we give a joint rank 6.5 for two positions because, in this condition, we have no way of knowing which score should be put in rank 6 and which score should be ranked 7. Therefore, the ranks of 6 and 7 do not exist these two ranks have been averaged ($(6 + 7)/2 = 6.5$) and assigned to each of these "tied" scores.
- Find the difference in the ranks **(d)**: This is the difference between the ranks of the two values on each row of the table. The rank of the second value is subtracted from the rank of the first. **(Rank1-Rank 2)**
- **Square the differences (d^2)** To remove negative values and then sum them ($\sum D^2$).

For example

To find the significant correlation between marks of English and Maths of students

step one make table

S.no	*English*	*Maths*
1	4	16
2	4	8
3	7	8
4	25	20
5	7	16
6	17	15
7	16	12
8	9	20
9	21	25
10	7	8

Assign rank to both set

S.No.	English	Maths	Rank 1	Rank 2
1	4	16	9.5	4.5
2	4	8	9.5	9
3	7	8	7	9
4	25	20	1	2.5
5	7	16	7	4.5
6	17	15	3	6
7	16	12	4	7
8	9	20	5	2.5
9	21	25	2	1
10	7	8	7	9

Where d = difference between ranks and
d^2 = difference squared.
Σd^2 = sum of difference squared

S.No	English	Maths	Rank1	Rank2	D (R1-R2)	
1	4	16	9.5	4.5	5	25
2	4	8	9.5	9	-0.5	0.25
3	7	8	7	9	-2	4
4	25	20	1	2.5	-1.5	2.25
5	7	16	7	4.5	2.5	6.25
6	17	15	3	6	-3	9
7	16	12	4	7	-3	9
8	9	20	5	2.5	2.5	6.25
9	21	25	2	1	1	1
10	7	8	7	9	-2	4
Total	**n=10**					

put all the values in formula and calculate

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

$$\rho = 1 - \frac{6\sum 67}{10(10^2-1)}$$

$$\rho = 1 - \frac{6X67}{10(100-1)}$$

$$\rho = 1 - \frac{402}{990}$$

$$\rho = 1 - 0.406(0.41)$$

$$\rho = 0.59$$

Hence, we have a **ρ** (or r_s) of 0.59. This indicates a **strong positive relationship** between the rank's individuals obtained in the English and in Maths.

Spearman Foot Rule/ Gain Method

- Spearman's footrule, a well-known measure of rank correlation method.
- It measures the relationship between two sets of ranks is shown to be a chance-corrected measure of agreement.
- The footrule is generalized to include tied ranks and a comparison with Spearman's rank-order correlation coefficient is provided. Procedures to determine the non asymptotic probability of the footrule with tied ranks are presented
- The result will always be between 1 and minus 1.
- $\boldsymbol{r = 1 - \frac{6\sum G}{(N^2 - 1)}}$
- r = correlation coefficient
- $\sum G$ = total gain
- n = the number of paired observations

Method/ Steps - calculating the coefficient

- Create a table from your data.
- **Rank** the two data sets. Ranking is achieved by giving the ranking '1' to the biggest number in a column, '2' to the second biggest value and so on. If data contains same numbers then give them joint/tied ranks.
- Find the difference in the ranks **(d)**: This is the difference between the ranks of the two values on each row of the table. The rank of the second value is subtracted from the rank of the first. (**Rank1-Rank** 2)
- **Square the differences (d^2)** To remove negative values. **(DxD)**
- Find the gain G_1 only for those ranks where rank 2 > rank 1 the rank of the first value is subtracted from the rank of the second. (Rank 2-Rank 1)
- Find the gain G_2 only for those ranks where rank1 > rank 2 the rank of the second value is subtracted from the rank of the first. (Rank 1-Rank 2). **ΣG** The sum of all the value of G_1 **&** G_2 are same.
- Put all the values in formula.

For example

To find the significant correlation between marks of English and Maths of students

S.no	*English*	*Maths*
1	4	16
2	4	8
3	7	8
4	25	20
5	7	16
6	17	15
7	16	12
8	9	20
9	21	25
10	7	8

Assign ranks to both set

S. No.	*English*	*Maths*	*Rank 1*	*Rank 2*
1	4	16	9.5	4.5
2	4	8	9.5	9
3	7	8	7	9
4	25	20	1	2.5
5	7	16	7	4.5
6	17	15	3	6
7	16	12	4	7
8	9	20	5	2.5
9	21	25	2	1
10	7	8	7	9

Where d = difference between ranks and d^2 = difference squared.

S.No	English	Maths	Rank1	Rank2	D (R1-R2)	
1	4	16	9.5	4.5	5	25
2	4	8	9.5	9	-0.5	0.25
3	7	8	7	9	-2	4
4	25	20	1	2.5	-1.5	2.25
5	7	16	7	4.5	2.5	6.25
6	17	15	3	6	-3	9
7	16	12	4	7	-3	9
8	9	20	5	2.5	2.5	6.25
9	21	25	2	1	1	1
10	7	8	7	9	-2	4

G_1 only for those ranks where rank2 > rank1 (R2-R1)
G_2 only for those ranks where rank1 > rank2 (R1-R2)
ΣG The sum of all the value of G_1 & G_2 are same.

S.No	*English*	*Maths*	*Rank1*	*Rank2*	*D (R1-R2)*		*G_1 (R2-R1)*	*G_2 (R1-R2)*
1	4	16	9.5	4.5	5	25		5
2	4	8	9.5	9	0.5	0.25		0.5
3	7	8	7	9	-2	4	2	
4	25	20	1	2.5	-1.5	2.25	1.5	
5	7	16	7	4.5	2.5	6.25		2.5
6	17	15	3	6	-3	9	3	
7	16	12	4	7	-3	9	3	
8	9	20	5	2.5	2.5	6.25		2.5
9	21	25	2	1	1	1		1
10	7	8	7	9	-2	4	2	
Total	**N=10**						**ΣG=11.5**	**ΣG =11.5**

put all the values in formula and calculate

- $\boldsymbol{r = 1 - \frac{6\sum G}{(N^2-1)}}$
- $r = 1 - \frac{6\sum 11.5}{(10^2-1)}$
- $r = 1 - \frac{6X11.5}{99}$
- $r = 1 - \frac{69}{99}$

- $r = 1 - 0.696$

- $r = 1 - 0.70$

- $r = 0.30$

- Hence, we have a **ρ** (or r_s) of 0.30. This indicates a *Low degree correlation relationship* between the ranks individuals obtained in the English and Maths exam.

THANK YOU..