Mustansiriyah Uni.
College of science
Atmospheric Science Dept.

الجامعة المستنصرية
كلية العلوم
قسم علوم الجو

المرحلة ألرابعة

Lecture Title | عنوان المحاضرة

Correlation Coefficient | معامل الارتباط

Lecturer Name | اسم التدريسي

Dr. Ali Raheem Al-nassar | د. علي رحيم النصار

لجنة التعليم الالكتروني

Spearman’s correlation

Spearman's correlation is a statistical measure of rank correlation (monotonic relationship: statistical dependence between the rankings of two variables) that determines the strength and direction (whether linear or not) of the monotonic relationship between two variables (x and y) rather than the strength and direction of the linear relationship between two variables, which is what Pearson's correlation determines, and constrained as follows:

$$-1 < \rho < 1$$

And its interpretation is similar to that of Pearson’s, e.g. the closer ρ is to ±1 the stronger the monotonic relationship.

Correlation is an effect size and so we can verbally describe the strength of the correlation using the following guide for the absolute value of

0.01-0.19	“very weak”
0.20- 0.39	“weak”
0.40-0.59	“moderate”
0.60-0.79	“strong”
0 .80-1.0	very strong

Spearman’s rank correlation requires ordinal data.

Examples of ordinal data:

1st, 2nd, 3rd,

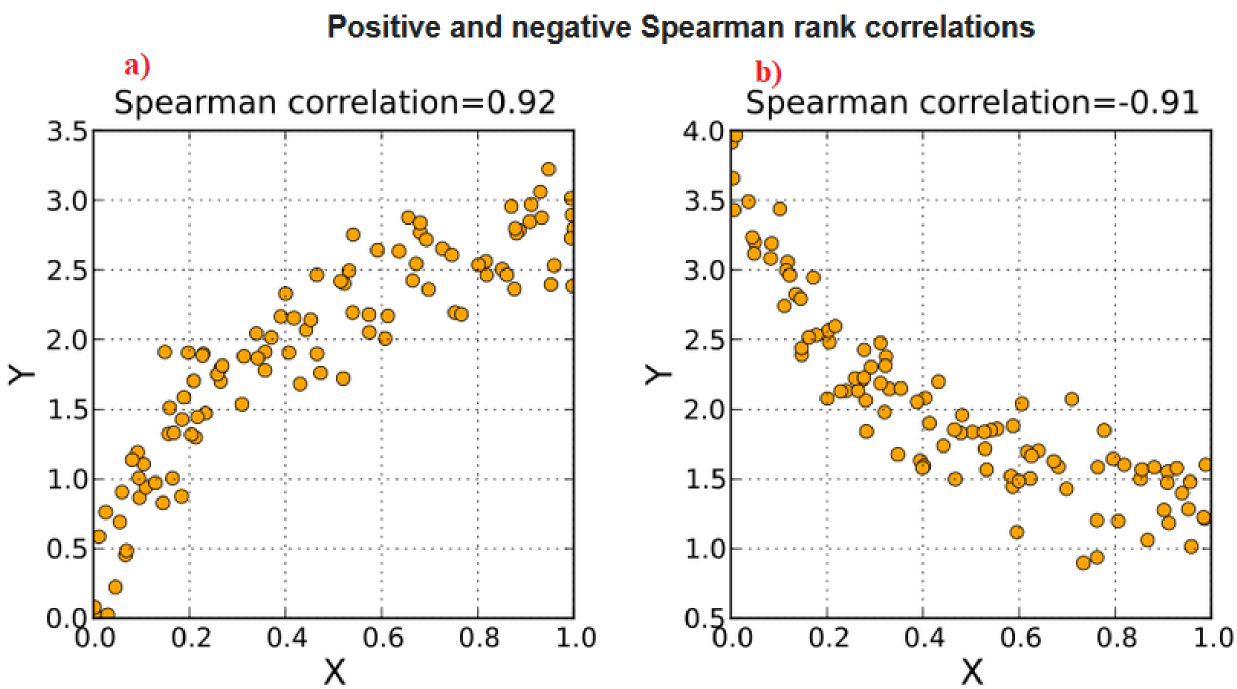
Small, Medium, Large, XL,

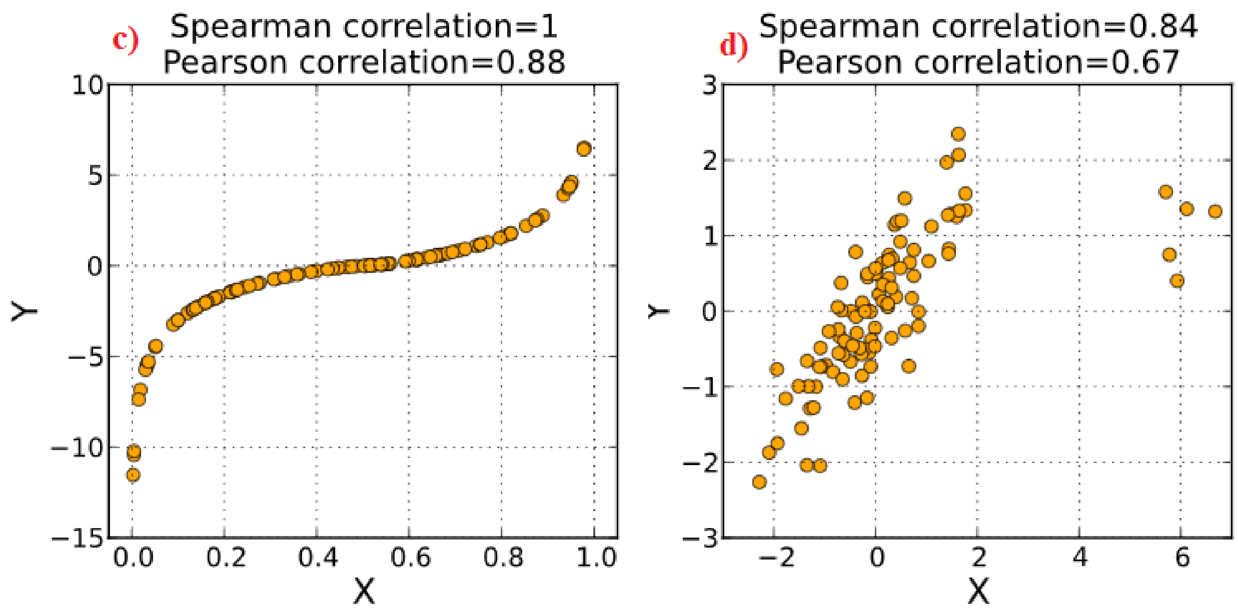
Strongly agree, Agree, Neutral, Disagree, Strongly Disagree

Very often, Often, Not Often, Not at all

What is a monotonic relationship?

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. Examples of monotonic and non-monotonic relationships are presented in the diagram below:





A Spearman's rank correlation between variables X and Y is calculated by:

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

Where d is the difference between ranks.

Example: The scores for nine students in physics and math are as follows:

Physics: 35, 23, 47, 17, 10, 43, 9, 6, 28

Mathematics: 30, 33, 45, 23, 8, 49, 12, 4, 31

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

Solution:

Step 1: Find the ranks for each individual subject: order the scores from greatest to smallest; assign the rank 1 to the highest score, 2 to the next highest and so on:

Step 2: Add a fourth column, d, to your data. The d is the difference between ranks. For example, the first student's physics rank is 3 and math rank is 5, so the difference is -2 points. In a sixth column, square your d values.

Physics(X)	Rank (X)	Math(Y)	Rank (Y)	d= (rank x – rank y)	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
					$\sum$=12

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

$$\rho = 1 - \frac{6 * 12}{9(81-1)}$$

$$\rho = 0.9$$

The Spearman Rank Correlation for this set of data is 0.9 (very strong)

Kendall's Tau (Kendall Rank Correlation Coefficient)

Kendall's Tau is a non-parametric measure of relationships between columns of ranked data. The Tau correlation coefficient returns a value of 0 to 1, where:

0 is no relationship,

1 is a perfect relationship.

A quirk of this test is that it can also produce negative values (i.e. from -1 to 0).

Kendall's Tau = (C – D / C + D)

Where C is the number of concordant pairs and D is the number of discordant pairs.

The Kendall formula is the difference between two probabilities: the probability that all pairs will be Concordant or Discordant:

$$(Concordant): \quad (x_i - x_j)(y_i - y_j) > 0$$

$$(Discordant): \quad (x_i - x_j)(y_i - y_j) < 0$$

The Partial Correlation:

A Partial correlation estimates the relationship between two variables while removing (holding constant) the influence of a third variable from the relationship.

(e.g., the relationship between the daily temperature with evaporation rate (irrespective of wind, when it is a really hot day).

like whether or not the sale value of a particular commodity is related to the expenditure on advertising when the effect of price is controlled.

For example, a researcher might want to calculate a partial correlation between height and weight while removing (holding constant) the effect of gender on the correlation.

Correlation between Height and Weight =0.805

However, when controlling for gender the correlation between height and weight is= 0.770

Multiple correlation coefficient :

Y , X_1 ,X_2

$$r_{YX_1} = \frac{n\sum YX_1 - \sum Y \sum X_1}{\sqrt{N\sum Y_I^2 - \left(\sum Y_I\right)^2}\sqrt{n\sum X_1^2 - \left(\sum X_1\right)^2}}$$

$$r_{YX_2} = \frac{n\sum YX_2 - \sum Y \sum X_2}{\sqrt{n\sum Y_1^2 - \left(\sum Y_1\right)^2}\sqrt{n\sum X_2^2 - \left(\sum X_2\right)^2}}$$

$$r_{X_1X_2} = \frac{n\sum X_1X_2 - \sum X_1 \sum X_2}{\sqrt{n\sum X_1^2 - \left(\sum X_1\right)^2}\sqrt{n\sum X_2^2 - \left(\sum X_2\right)^2}}$$

$$R_{YX_1X_2} = \sqrt{\frac{r^2{}_{YX_1} + r^2{}_{YX_2} - 2r_{YX_1} r_{YX_2} r_{X_1X_2}}{1 - r^2{}_{X_1X_2}}}$$

Example

Y	X_1	X_2
2	1	1
3	6	8
2	5	2
1	7	6
4	10	8

Y	X_1	X_2	YX_1	YX_2	X_1X_2	Y_i^2	X_1^2	X_2^2
2	1	1	2	2	1	4	1	1
3	6	8	18	24	48	9	36	64
2	5	2	10	4	10	4	25	4
1	7	6	7	6	42	1	49	36
4	10	8	40	32	80	16	100	64
$\sum 12$	29	25	77	68	181	34	211	169

$$r_{YX_1} = \frac{n\sum YX_1 - \sum Y \sum X_1}{\sqrt{n\sum Y_i^2 - \left(\sum Y_i\right)^2}\sqrt{n\sum X_1^2 - \left(\sum X_1\right)^2}}$$

$$= \frac{5(77)-(12)(29)}{\sqrt{5(34)-(12^2)}\sqrt{5(211)-(29)^2}}$$

$$= \frac{383-348}{\sqrt{(170-144)}\sqrt{(1055-841)}} = \frac{37}{\sqrt{26}\sqrt{214}} = \frac{37}{74.5} = 0.49$$

$$r_{YX_2} = \frac{n\sum YX_2 - \sum Y \sum X_2}{\sqrt{n\sum Y_i^2 - \left(\sum Y_i\right)^2}\sqrt{n\sum X_2^2 - \left(\sum X_2\right)^2}}$$

$$= \frac{5(68)-(12)(25)}{\sqrt{5(34)-(12)^2}\sqrt{5(169)-(25)^2}} = \frac{40}{}$$

$$\frac{340-300}{\sqrt{(170-144)}\sqrt{(845-625)}} = \frac{40}{\sqrt{26}\sqrt{220}} = \frac{40}{75.6} = 0.53$$

$$r_{X_1X_2} = \frac{n\sum X_1X_2 - \sum X_1 \sum X_2}{\sqrt{n\sum X_1^2 - \left(\sum X_1\right)^2}\sqrt{n\sum X_2^2 - \left(\sum X_2\right)^2}}$$

$$r_{X_1X_2} = \frac{5(181)-(29)(25)}{\sqrt{\sum 5(211)-(29)^2}\sqrt{5(169)-(25)^2}}$$

$$= \frac{905-725}{\sqrt{(1055-841)}\sqrt{842-625}} = \frac{180}{\sqrt{214}\sqrt{220}} = \frac{180}{217} = 0.83$$

$$R_{YX_1X_2} = \sqrt{\frac{r^2{}_{YX_1} + r^2{}_{YX_2} - 2r_{YX_1}r_{YX_2}r_{X_1X_2}}{1-r^2{}_{X_1X_2}}}$$

$$= \sqrt{\frac{(0.49)^2 + (0.53)^2 - 2(0.49)(0.53)(0.83)}{1 - (0.83)^2}}$$

$$= \sqrt{\frac{0.240 + 0.280 - 2(0.431)}{0.312}}$$

$$= \sqrt{\frac{0.52 - 0.431}{0.312}}$$

$$= \sqrt{\frac{0.089}{0.312}} = \sqrt{0.285} = 0.53$$