

Correlation as a Measure of Association

Name

Institution

Correlation is an essential statistical concept that denotes the association or interrelationship between variables. The purpose of studying correlation is for individuals to establish a relationship, control and plan input (independent variables) and output (dependent variables) (Sasvári, 2013).

There are two methods that can be employed to establish or measure the degree of correlation and are represented by "R" and "r" respectively.

(i) Coefficient of correlation

This method is denoted by "r". It provides a measure of strength of association between two variables, one independent and the other the dependent variable. The relationship between these variables can range between +1 and -1 for perfect positive correlation and perfect negative correlation. Zero indicates no relationship between variables (Poonia, Poonia & Poonia, 2011).

X	Y	X ²	Y ²	XY
15	60	225	3600	900
24	45	576	2025	1080
25	50	625	2500	1250
30	35	900	1225	1050
35	42	1225	1764	1470
40	46	1600	2116	1840
45	28	2025	784	1260
65	20	4225	400	1300
70	22	4900	484	1540

75	15	5625	225	1125
$\sum X=424$	$\sum Y=363$	$\sum X^2=21,926$	$\sum Y^2=15,123$	$\sum XY=12,815$

Therefore r

$$r = \frac{n\sum XY - \sum X \sum Y}{\sqrt{n\sum X^2 - (\sum X)^2} \sqrt{n\sum Y^2 - (\sum Y)^2}}$$

$$r = \frac{10 \times 12,815 - 424 \times 363}{\sqrt{10 \times 21,926 - 179,776} \sqrt{10 \times 15,123 - 131,768}}$$

$$= \frac{-25762}{198.76 \times 139.50}$$

Hence r = -0.93

The correlation, therefore, indicates a strong negative linear association between variable X and Y.

Advantage

It provides the causal or non-causal relationships hence appropriate to make critical decisions.

Disadvantage

A low negative correlation may be misinterpreted as lack of relationship between variables.

(ii) Rank correlation

This method is denoted by R, it is used to measure strength between two sets of ordered or ranked data. R, can range from +1 and -1 (Mari & Kotz, 2001). The range mainly represents perfect positive and negative rank correlation while zero means there is no correlation.

Example

Items	X	Y	D= x-y	D ²
A	6	5	1	1
B	1	3	-2	4
C	3	4	-1	1
D	7	6	1	1
E	8	7	1	1
F	2	1	1	1
G	4	8	-4	16
H	5	2	3	9
I	10	9	+1	1
J	9	10	-1	1
				$\sum d^2 = 36$

$$R = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

$$= 1 - \frac{6 \times 36}{10(10^2-1)}$$

$$= 1 - \frac{216}{990}$$

$$= 1 - 0.22$$

$$R = 0.78$$

Since correlation is 0.78 it implies that there is high positive correlation between the ranks.

Advantages

Its helps research to rank variables against their order of needs.

Disadvantages

The ranking may be bias in case of spurious relationships or correlation (Bobko, 2001).

Application of correlation

- Establishment between individual experience and corresponding performance
- The amount spent on an advertisement and the anticipated amount of income after the sales of goods and services.

References

Bobko, P. (2001). *Correlation and regression*. Thousand Oaks [Calif.]: Sage Publications.

Mari, D., & Kotz, S. (2001). *Correlation and dependence*. London: Imperial College Press.

Poonia, V., Poonia, V., & Poonia, M. (2011). *Advanced statistics*. New Delhi: Vishvabharti Publications.

Sasvári, Z. (2013). *Multivariate Characteristic and Correlation Functions*. Berlin: De Gruyter.