

Statistics

Lecture 8

Correlation analysis

By Dr.Chelli




Lecture objectives

- Training students in calculating Pearson correlation coefficient



Introduction

Correlation analysis is used to quantify the association between two continuous variables (between an independent and dependent variable or between two independent variables). In this kind of analysis, we estimate a sample correlation coefficient, denoted as **r**



The correlation coefficient ranges between -1 and $+1$ and quantifies the direction and strength between the two variables. The correlation may be positive or negative

The sign of the correlation coefficient indicates the direction of the association.

The magnitude of the correlation coefficient indicates the strength of the association

For example, a correlation of $r = 0.8$ reveals a strong, positive association, while $r = -0.3$ shows a weak negative association.

A correlation close to zero suggest no association between the two continuous variables

Correlation Strength vs. Axes

Perp. Distance	Conclusion
distance = 1	Perfect Correlation
$0.9 \leq \text{distance}$	Very Strong Correlation
$0.7 \leq \text{distance} < 0.9$	Strong Correlation
$0.5 \leq \text{distance} < 0.7$	Moderate Correlation
$0.3 \leq \text{distance} < 0.5$	Weak Correlation
$0 < \text{distance} < 0.3$	Very Weak Correlation
distance = 0	No Correlation



Pearson correlation coefficient

The Pearson correlation coefficient is one of many types of coefficient in the field of statistics.

It is a helpful statistical formula that measures the strength between variables and relationships.

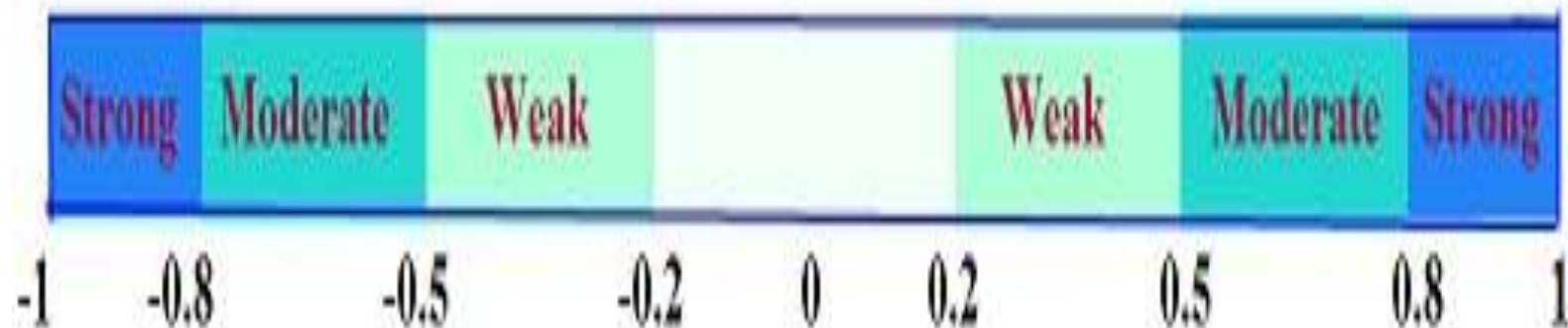
As stated previously the value of r ranges between -1.00 and $+1.00$. If the value is in the positive range, this means that the correlation is positive. But if it is in the negative range, it is negative.

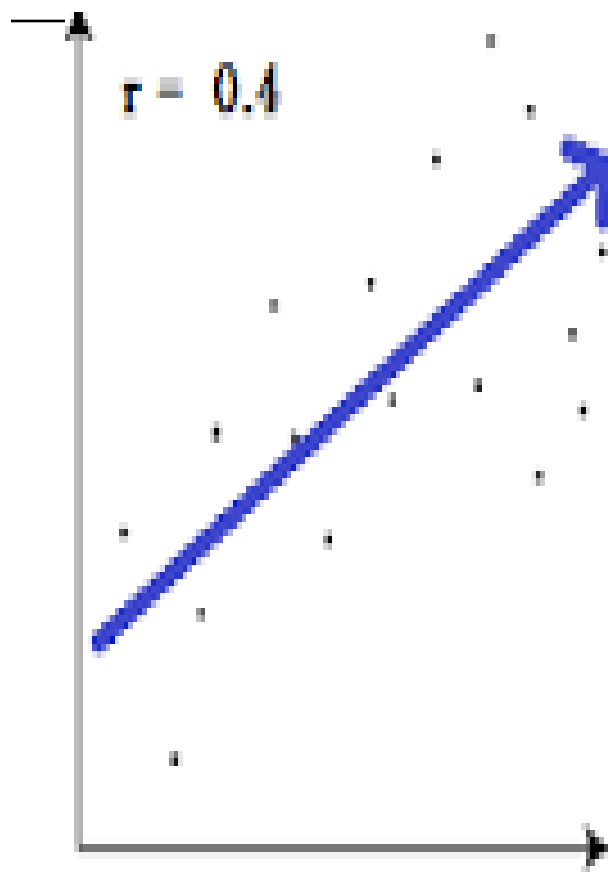
Pearson Correlation Coefficient

Negative Correlation

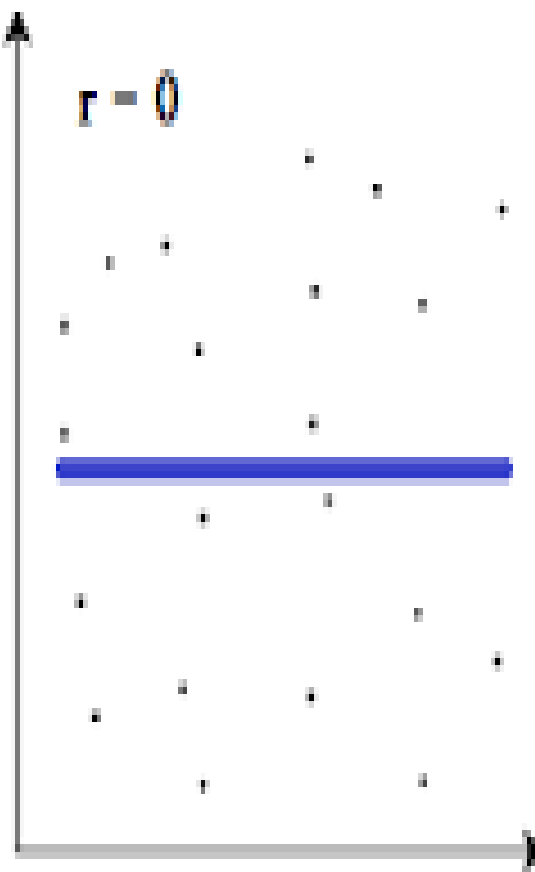
No Correlation

Positive Correlation

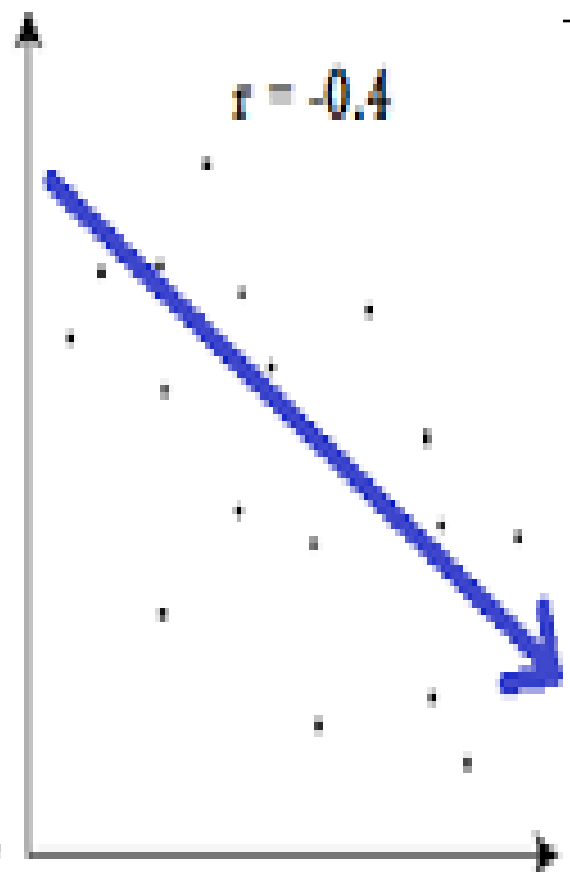




Positive Correlation



No correlation



Negative

Calculating the value of Pearson correlation coefficient (r)

- Step one: Make a table with your data for two variables, label the variable (x) and (y) and add three more columns labeled (xy), and (x^2) and y^2 like this

Sample	x	y	xy	x^2	y^2
1					
2					
3					

Step two: Complete the chart using basic multiplication of variable values

Sample	x	y	xy	x^2	y^2
1	20	30	600	400	900
2	24	20	480	576	400
3	17	27	459	289	729

Step 3: After the multiplications, find the sums

Sample	x	y	xy	x ²	y ²
1	20	30	600	400	900
2	24	20	480	576	400
3	17	27	459	289	729
Total	61	77	1539	1265	2029

Step 4: Use the formula to find Pearson correlation value

$$r = \frac{N\sum xy - (\sum x)(\sum y)}{\sqrt{[N\sum x^2 - (\sum x)^2][N\sum y^2 - (\sum y)^2]}}$$

Where:

- N = number of pairs of scores
- $\sum xy$ = sum of the products of paired scores
- $\sum x$ = sum of x scores
- $\sum y$ = sum of y scores
- $\sum x^2$ = sum of squared x scores
- $\sum y^2$ = sum of squared y scores

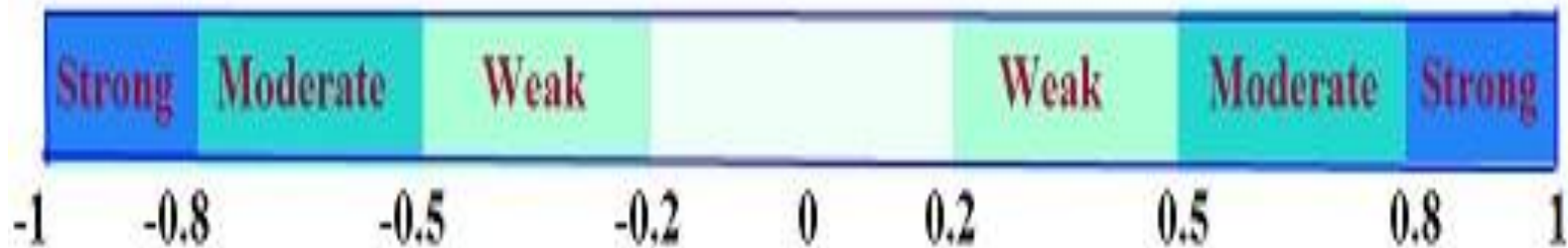
The value found indicates the magnitude of the correlation a shown in the diagram below

Pearson Correlation Coefficient

Negative Correlation

No Correlation

Positive Correlation



Pearson r hand calculation

After the calculation of the coefficient r

1. Choose alpha level 0.05
2. Calculate the degree of freedom: $N - 2$
3. Use the table of the critical value

Refer to :

Pearson r: Hand calculation

psc.dss.ucdavis.edu/sommerb/.../correlation/hand/pearson_hand.ht...