

4h – A Guide to Pearson’s Product Moment

Pearson’s Product Moment Correlation Coefficient measures the degree of correlation there may be between two variables.

Why would we use Pearson’s product moment correlation coefficient?

This test is best used by a researcher when their results have already been plotted on a scatter graph and there is an indication of a linear relationship between the two factors. In order for Pearson’s Product Moment to be useful, both sets of data have to be continuous and there should be no significant outliers (anomalies) in the manner in which the data is distributed.

Analysis of data using Pearson’s Product Moment gives a numerical value (r) between -1 and +1 to show the strength of any correlation.

Worked Example:

In the following worked example, the researcher is trying to find the strength of correlation between the measure of the velocity of the water in a river channel and the gradient of the river at various points.

1. Label the paired data sets x and y for ease of use.

Gradient (x)	8.5	7.1	5.2	5.5	2.0	3.5	4.3	4.7	6.0	5.5	1.1	7.4	3.2	2.4	6.4
Velocity (y)	0.51	0.11	0.24	0.33	0.40	0.35	0.27	0.24	0.15	0.18	0.48	0.08	0.31	0.38	0.16

2. Work out the following figures for each piece of data and then sum the figures. It is best to do this in a spreadsheet so the researcher does not have to undertake hundreds of calculations.

																Σ
x	8.5	7.1	5.2	5.5	2.0	3.5	4.3	4.7	6.0	5.5	1.1	7.4	3.2	2.4	6.4	72.8
y	0.51	0.11	0.24	0.33	0.40	0.35	0.27	0.24	0.15	0.18	0.48	0.08	0.31	0.38	0.16	4.19
x^2	72.3	50.4	27.0	30.3	4.0	12.3	18.5	22.1	36.0	30.3	1.21	54.8	10.2	5.76	41.0	416
y^2	0.26	0.01	0.06	0.11	0.16	0.12	0.07	0.06	0.02	0.03	0.23	0.01	0.10	0.14	0.03	1.41
xy	4.34	0.78	1.25	1.82	0.80	1.23	1.16	1.13	0.90	0.99	0.53	0.59	0.99	0.91	1.02	18.4

3. Using this summed data, the following calculations should then be worked out:

(n = the number of pairs of data)

$$ssxy = \Sigma xy - \frac{(\Sigma x \Sigma y)}{n}$$

$$ssxy = 18.4 - 20.34$$

$$ssxy = 18.4 - \frac{(72.8 \times 4.19)}{15}$$

$$ssxy = -1.94$$

$$ssxy = 18.4 - \frac{(305.032)}{15}$$

$$ssxx = \Sigma x^2 - \frac{(\Sigma x \Sigma x)}{n}$$

$$ssxx = 416 - 353.32$$

$$ssxx = 416 - \frac{(72.8 \times 72.8)}{15}$$

$$ssxx = 62.68$$

$$ssxx = 416 - \frac{(5299.84)}{15}$$

$$ssyy = \Sigma y^2 - \frac{(\Sigma y \Sigma y)}{n}$$

$$ssyy = 1.41 - 1.17$$

$$ssyy = 1.41 - \frac{(4.19 \times 4.19)}{15}$$

$$ssyy = 0.24$$

$$ssyy = 1.41 - \frac{(17.56)}{15}$$

4. To work out the Pearson's Product Moment correlation coefficient (r) use the above boxed results in the following equation:

$$r = \frac{ssxy}{ssxx \times ssyy}$$

$$r = \frac{-1.94}{62.68 \times 0.24}$$

$$r = \frac{-1.94}{15.04}$$

$$r = -0.13$$

A value of -1 indicates a perfect negative correlation while a value of +1 indicates a perfect positive correlation. A value of zero suggests no correlation at all.

Therefore, in this example, the results show a negative correlation between gradient and river velocity, which is what we might expect, but the results only show a very weak correlation. The next role of the geographical researcher is to decide whether there are enough samples in the data for the result to be valid or whether there are other, site specific, reasons why the data does not match the theory.