Pearson Product Moment Correlation Coefficient

The Pearson Product Moment Correlation Coefficient is a statistical test which is used to show the strength and direction of a correlation between two sets of data in which the individual values are paired against each other. In this sense, it is similar to the Spearman's Rank Correlation Coefficient, though the former uses real data values rather than the rankings of the data in sequence.

The Pearson Product Moment Correlation Coefficient can be used by researchers to test a variety of geographical variables against each other to see if there is a meaningful correlation. Here are some examples of possible geographical investigations that could use this test:

- Settlements - how **house prices** change as one moves away from the CBD.
- Rivers - the variance in **bedload size** with distance from the source.
- Development - how **life expectancy** changes with average calorie intake.
- Coasts - how **favourability scores** vary with the length of time sea defences have been in place.
- Weather - how **percentage cloud cover** and precipitation volume may be correlated.
- Industry - how the **air miles** of a garment vary with the recommended retail price.
- Tourism - how average **tourist spend** changes with distance to the home area or country.
- Demography - how the **birth rate** of a region or country changes with the average age of first time mothers.

How to carry out a Pearson Product Moment Correlation Coefficient calculation:

For this example, a geographical researcher is examining how air pollution levels vary with distance from the CBD of a major city. Nitrogen dioxide levels were recorded and mapped across the city at regular intervals along a transect. The researcher believed that as the centre of the city was quite old and contained few roads wide enough for large volumes of traffic, it would not have as high a level of nitrogen dioxide as the city outskirts. The researcher therefore formulated the following hypothesis:

> "There is a correlation between the level of nitrogen dioxide in the air and the distance from the CBD."

The researcher started by tabulating the observed data and calculating the mean values.

Distance from CBD (km)	Nitrogen dioxide level (μgm^{-3})
0	61
1	64
2	82
3	79
4	60
5	49
6	38
7	25
8	34
9	28
10	20
11	13
12	15
mean ($\bar{x}$) = 6.0	mean ($\bar{y}$) = 43.7

The deviation (dx and dy) was then calculated for each data set by subtracting the x or y value from its corresponding mean. These figures were squared to remove any negative values. Finally, the sum of the squared deviations was noted, as was dx multiplied by dy for each value. This produced the following results table:

x	dx	dx^2	y	dy	dy^2	dxdy
0	6	36	61	-17.3	299.6	-103.8
1	5	25	64	-20.3	412.4	-101.5
2	4	16	82	-38.3	1467.5	-153.2
3	3	9	79	-35.3	1246.6	-105.9
4	2	4	60	-16.3	265.9	-32.6
5	1	1	49	-5.3	28.2	-5.3
6	0	0	38	5.7	32.4	0
7	-1	1	25	18.7	349.4	-18.7
8	-2	4	34	9.7	93.9	-19.4
9	-3	9	28	15.7	246.2	-47.1
10	-4	16	20	23.7	561.3	-94.8
11	-5	25	13	30.7	942.0	-153.5
12	-6	36	15	28.7	823.2	-172.2
$\bar{x}$ = 6		Σ = 182	$\bar{y}$ = 43.7		Σ = 6768.8	Σ = -1008.0

The Pearson Product Moment Correlation Coefficient value (r) was then calculated using the formula:

$$r = \frac{\Sigma (dxdy)}{\sqrt{\Sigma (dx^2) \times \Sigma (dy^2)}}$$

$$r = \frac{-1008.0}{\sqrt{182 \times 6768.8}}$$

$$r = \frac{-1008.0}{\sqrt{1231921.6}}$$

$$r = \frac{-1008.0}{1109.9}$$

r = -1008.0

This is known as the calculated value.

The negative calculated value for r indicates a strong negative correlation. However, before one can fully accept the hypothesis, a significance test should also be carried out. This tells the researcher the extent to which one can be sure that the results are meaningful and the level to which one can be sure that the results did not occur by chance.

To calculate this, a significance table is required. The researcher compares the r value with the critical value for the appropriate number of sets of paired data. If the calculated value (regardless of the correlation direction) is greater than the critical value, the hypothesis can be accepted. A full significance table for the Pearson Product Moment Correlation Coefficient can be downloaded from the island geographer site.

The degrees of freedom (df) = n-1 where n is the number of paired observations. Therefore, in this case, the degrees of freedom is 12.

To a 95% significance level, the critical value in this example is 0.458. Therefore, the calculated value is greater than the critical value and the researcher can fully accept their hypothesis.