The Pearson product-moment correlation coefficient

You need to set null and alternative hypotheses:
Null hypothesis: There is no correlation between ….

Alterative hypothesis: There is a correlation between ….

The test assumes that data is on a continuous scale and the values of both members of the pairs are normally distributed.

Calculations

1. Fill in your observations in the table below. Add up the x values and the y values and calculate the mean for each column
2. For each value of x calculate the difference between the value and the mean, $x-\bar{x}$. Then do the same for the y values, calculating $y-\bar{y}$.
3. Square each $x-\bar{x}$ value and $y-\bar{y}$ value and enter the results into the table in the $(x-\bar{x})^2$ and $(y-\bar{y})^2$ columns
4. Add up the values in each column. The sum of $x-\bar{x}$ values and the sum of $y-\bar{y}$ values should both be zero.
5. Finally multiply each $(x-\bar{x})^2$ value by the corresponding $(y-\bar{y})^2$ value and enter the results. You are now ready to calculate the Pearson correlation coefficient.

Observation no	**A x value**	**B $x-\bar{x}$**	**C $(x-\bar{x})^2$**	**D y value**	**E $y-\bar{y}$**	**F $(y-\bar{y})^2$**	**G $(x-\bar{x})^2\ (y-\bar{y})^2$**
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
Sums of columns							
No. of observations							
Means of x and y							
	A	**B**	**C**	**D**	**E**	**F**	**G**

The Pearson formula is:

$$r_{xy} = \frac{\Sigma(x-\bar{x})\,(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2\,\Sigma\,(y-\bar{y})^2}}$$

Fearsome as this looks, you have already done all the hard work!

The Pearson product-moment correlation coefficient

6. Carry forward your calculated values into the summary table below.

For the top row of the formula	What it is	Actual value	Row no.
$\Sigma(x - \bar{x})(y - \bar{y})$	(sum of figures in column **G**)		**1**
For the bottom row of the formula	What it is	Actual value	**2**
$\Sigma(x - \bar{x})^2$	(sum of figures in column C)		**3**
$\Sigma(y - \bar{y})^2$	(sum of figures in column F)		**4**
$\Sigma(x - \bar{x})^2 \Sigma(y - \bar{y})^2$	(multiply the sum of figures in column C by the sum of figures in column F)		**5**
$\sqrt{\Sigma(x - \bar{x})^2 \Sigma(y - \bar{y})^2}$	(Calculate the square root of the value above)		**6**
$r_{xy} =$	(divide the value in row 1 by the value in row 6): the answer should lie between +1 and -1		

Significance of the result

You now need to compare your result with a table of significant values of r at the 5% probability level

N = number of pairs of data	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
degrees of freedom = N-2			**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
Critical value of r			.997	.950	.878	.811	.754	.707	.666	.632	.602	.576	.553	.532	.514

Calculated value of r_{xy} =

Critical value of r from table above =-

If the calculated value (ignoring the sign if it is negative) is **equal to or greater than** the critical value then the correlation is significant at the 5% probability level, so you can reject your null hypothesis and accept your alternative hypothesis.

If the calculated value is less than the critical value then the correlation is not significant so you can fail to reject your null hypothesis and reject your alternative hypothesis.

Remember that this test is able to show whether two variables are connected. It is *not* able to show that the variables are *not* connected. If one variable depends on another, i.e., there is a causal relation, then it is always possible to find some kind of *correlation* between the two variables. However, if both variables depend on a third, they can show a sizable correlation without any causal dependency between them.

A famous example is the fact that the position of the hands of all clocks are correlated, without one clock being the cause of the position of the others. Another example is the significant correlation between human birth rates and stork population sizes.