
PEARSON PRODUCT MOMENT
CORRELATION COEFFICIENT (PPMC)

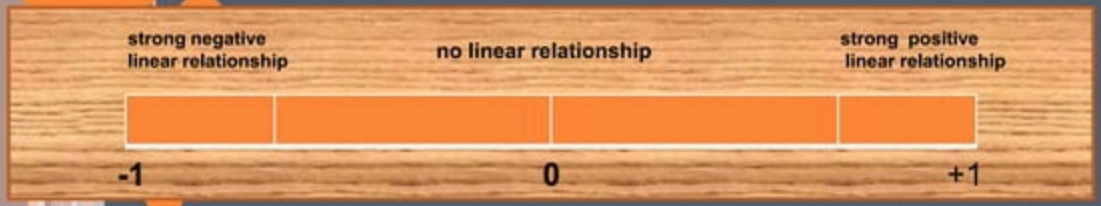
CORRELATION COEFFICIENT: - computed from the sample data measures the strength and direction of a linear relationship between two variables.

- The symbol for the sample Correlation Coefficient is "*r*".

-The symbol for the population Correlation Coefficient is "*ρ*" *(Greek letter rho).*

The range of the correlation coefficient is from -1 to 1.

- strong positive linear relationship between the variables = the value of r will be close to +1
- strong negative linear relationship between the variables = the value of r will be close to -1
- When there is no linear relationship between the variables or only weak relationship , the value of *r*= will be close to 0.



Formula for the correlation coefficient *r*

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

Where n is the number of data pairs.

Example.1 Compute the value of the correlation coefficient for the data obtained in the study of age and blood pressure given.

Solution:

Step 1 make a table.

subject	age x	pressure y	xy	x^2	y^2
A	43	128			
B	48	120			
C	56	135			
D	61	143			
E	67	141			
F	70	152			

STEP 2: find the values of xy, x^2, y^2 and place these values in the corresponding columns of the table.

subject	age x	pressure y	xy	x^2	y^2
A	43	128	5504	1849	16384
B	48	120	5760	2304	14400
C	56	135	7560	3136	18225
D	61	143	8723	3721	20449
E	67	141	9447	4489	19881
F	70	152	10640	4900	23104
	$\sum x = 3455$	$\sum y = 819$	$\sum xy = 47,634$	$\sum x^2 = 20,399$	$\sum y^2 = 112,443$

Step 3 : Substitute in the formula and solve for *r*.

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

$$r = \frac{6(47{,}643) - (345)(819)}{\sqrt{[(6)(20{,}399) - (345)^2][(6)(112{,}443) - (819)^2]}}$$

$$= 0.897$$

The correlation coefficient suggests a strong positive relationship between age and blood pressure.

POSSIBLE RELATIONSHIPS BETWEEN VARIABLES:

1. There is a direct cause- and – effect relationship between the variables.
2. There is a reverse cause- and – effect relationship between the variables.
3. The relationship between the variables may be caused by a third variable.
4. There may be complexity of interrelationships among many variables.
5. The relationship may be coincidental .

Formally defined, the population correlation coefficient ρ is the correlation computed by using all possible pairs of data values (x,y) taken from the population.

Several methods can be used to test the significance of the correlation coefficient. The first uses the t - test .

Formula for the t test for the correlation coefficient

$$t = r\sqrt{\frac{n-2}{1-r^2}}$$

Where degrees of freedom equal to n - 2.

Test the significance of the correlation coefficient found in the example 1. use $\alpha = 0.05$ and $r = 0.897$

Solution:

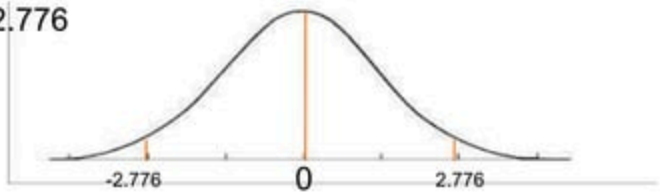
STEP 1: State the hypotheses:

$Ho : \rho = 0$ $\quad$ $H_1: \rho \neq 0$

(Ho means that there is no correlation between the x and y variables in the population.

H_1 means that there is a correlation between the x and y variables in the population.)

STEP 2 : $\alpha = 0.05$ df= 6-2= 4 the critical value obtained from table t distribution are +2.776 and – 2.776



Step 3: compute the test value.

$$t = r\sqrt{(n-2)/(1-r^2)}$$

$$= (0.897)\sqrt{(6-2)/(1-(0.897)^2)}$$

$$= 4.059$$

Step 4: Make the decision. Reject the null hypothesis, since the test value falls in the critical region.

Step 5: Summarize the results.

There is a significant relationship between the variables of age and blood pressure.

The second method: to test the significance of *r* is to use the Critical Values for PPMC. It uses the same steps except the step 3.

Referring to example 1.
Step 3.
df = n – 2 = 6- 2 = 4
α = 0.05
r = 0.897

➢ Locate the value to table of critical values for PPMC.

➢ The table gives a critical value of 0.811.

➢ 0.811 < *r* < −0. 811 will be significant, and the null hypothesis will be rejected.
(Reject H0: $\rho = 0$ if the absolute value of *r* is greater than the value given in the table for critical values for the PPMC)

Critical Values for PPMC		
Level of Significance (p) for Two-Tailed Test	.05	.01
DF (df= N- 2)		
1	0.997	0.9999
2	0.95	0.99
3	0.878	0.959
4	0.811	0.917
5	0.754	0.874
6	0.707	0.834
7	0.666	0.798
8	0.632	0.765
9	0.602	0.735
10	0.576	0.708
11	0.553	0.684
12	0.532	0.661

Activity:

1. A random sample of U.S. cities is selected to determine if there is a relationship between the population (in thousands) of people under 5 years of age and the population (in thousands) of those 65 years of age and older. The data for the sample are shown here. With test the significance of the correlation coefficient at α = 0.05. Usingt-test formula.

Under 5 x	178	27	878	314	322	143
65 and over y	361	72	1496	501	585	207

2. A rar to determine if there is a relationship between the population (in thousands) of people under 5 years of age and the population (in thousands) of those 65 years of age and older. The data for the sample are shown here. With test the significance of the correlation coefficient at α = 0.05. Using Critical Values for the PPMC

Test score x	98	105	100	100	106	95	116	112
GPA y	2.1	2.4	3.2	**2.7**	2.2	2.3	3.8	3.4

THANK YOU FOR LISTENING