

PMCC and Hypothesis Testing

Starter

1. **(Review of a previous lesson)** Find the PMCC for $\sum x_i = 367$, $\sum y_i = 270$,
 $\sum x_i^2 = 23845$, $\sum y_i^2 = 12976$, $\sum x_i y_i = 17135$ and $n = 6$.
 Comment on your findings.

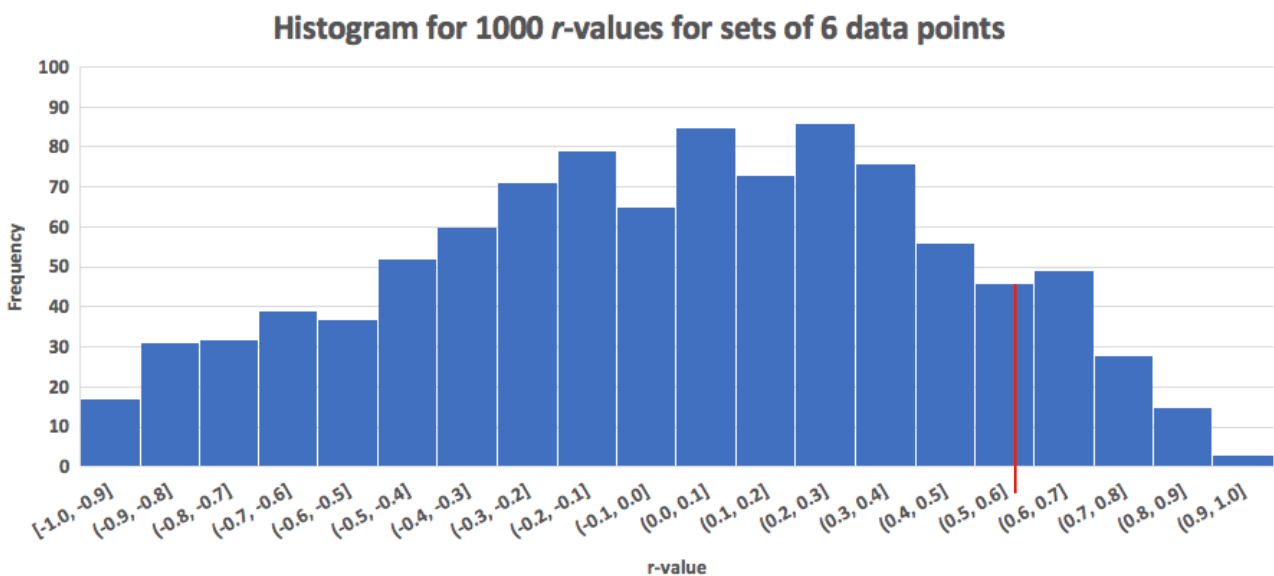
Notes

After getting our value for the PMCC (r) it would be useful to know how sure we are that it indicates correlation i.e. that it is not random.

To do this we consider a random set of values for x and y (we will use $n = 6$), which can be generated on Excel.

Please do not copy.

One thousand sets of random values for x and y were generated and the r -value for each set of data was calculated. These 1000 r -values were collated into a histogram using intervals $-1 \leq r < -0.9$, $-0.9 \leq r < -0.8$.



As can be seen, the majority of the values are close to $r = 0$. Since the values of x and y are supposed to be random this is what we would expect. From an earlier lesson, an r -value close to -1 or 1 indicates the two quantities are correlated.

To carry out a hypothesis test we need a **null** and **alternative** hypothesis. The **assumption** (the null hypothesis) is that the quantities are **not correlated**.

Null hypothesis: $H_0 : \rho = 0$

the quantities are not correlated

Alternative hypothesis: $H_1 : \rho \neq 0$

the quantities are correlated

N.B. ρ is the Greek letter “rho”

From the starter, the PMCC (r -value) is **0.577** for the 6 pairs of data points. The amount of doubt is to the right of this figure. In the case above the amount of doubt is about 12 %.

The question is, how sure do we want to be that the value is significant?

In general, statisticians want to be at least 95 % sure so there should only be 5 % doubt. Clearly 12 % doubt is too much

The critical value to decide whether an r -value is **significant** or not can be found in tables like the one on page 198 (also included in the formula booklet).

In our case of six pairs of data points, the critical value at the 5% level is 0.7293. Since our r -value of 0.577 is **smaller than the critical value**, we would say that the r -value is **not significant**.

Hence, there is not enough evidence to reject the null hypothesis i.e. the quantities x and y are not correlated. This does not mean they are not correlated, it means that the evidence provided is not strong enough to indicate that they are correlated.

Please copy.

Critical values for the PMCC

The critical value for the PMCC value defines the **smallest value** needed to suggest that there is evidence to **two quantities are correlated**.

For example, the 5% critical value for 20 pairs of data points is 0.4438. If the r -value is above 0.4438, it suggests that the two quantities are correlated. Below this value and we would say that the two quantities are not correlated.

As the level of doubt goes down, the critical value goes up i.e. the closer the r -value has to be to 1 (or -1).

The more data pairs of data points there are, the lower the critical value at a specific level of significance.

1-tailed vs. 2-tailed test

If we have no prior belief about the data, we carry out a 2-tailed test, where we apportion half the significant area to the left and right of the graph. If we believe there is a positive (or negative) correlation, we carry out a 1-tailed test i.e. all the doubt goes on one side of the graph.

If in doubt, calculate the r -value and then carry out an appropriate 1-tailed test.

Success criteria – how to carry out a significance test for the PMCC value

1. Calculate the r -value.
2. State the null hypothesis: it is always $H_0 : \rho = 0$ where ρ is the population correlation coefficient
3. Decide whether you will carry out a 1-tailed or 2-tailed test (it is usually given in the question).
4. State the alternative hypothesis, H_1 :
 - $H_1 : \rho \neq 0$ – two-tailed test
 - $H_1 : \rho > 0$ – one-tailed test, it is believed there could be a positive correlation
 - $H_1 : \rho < 0$ – one-tailed test, it is believed there could be a negative correlation
5. Decide on the level of significance (it is usually given in the question)
6. Use the tables to find the critical value based on the number of values, n , and the significance level
7. Compare your r -value to the critical value:
 - If $|r - \text{value}| > \text{critical value}$ then reject $H_0 \Rightarrow$ there is evidence to suggest there is a correlation between the two variables
 - If $|r - \text{value}| < \text{critical value}$ then do not reject $H_0 \Rightarrow$ there is no evidence to suggest there is a correlation between the two variables

N.B. A hypothesis test on the PMCC value can only be carried out on *bivariate data that is normally distributed*.

E.g. 1 The PMCC value for 12 sets of data points is found to be -0.593 . Carry out the following hypotheses tests and comment on your findings.

- (a) A 2-tailed test at a 5 % significance level.
- (b) A 1-tailed test at a 1 % significance level.

Working:

- (a) No information is given so carry out a two-tailed test.
 $H_0 : \rho = 0$ *the quantities are not correlated*
 $H_1 : \rho \neq 0$ *the quantities are correlated*
Critical value is 0.5760 *significance level is 2.5 % either side*
Since $|r| = |-0.593| > 0.5760$ there is evidence to reject H_0
i.e. there is evidence to suggest there is a correlation between the variables
- (b) $H_0 : \rho = 0$ *the quantities are not correlated*
 $H_1 : \rho > 0$
Critical value is 0.6581
Since $|r| = |-0.593| < 0.6581$ there is no evidence to reject H_0
i.e. there is evidence to suggest there is not a correlation between the variables

As can be seen, it can depend on the level of significance whether the null hypothesis is rejected or not. Notice the language of hypothesis testing “there is evidence to suggest that...”.

E.g. 2 The ages A (years) and heights H (cm) of 11 members of a football team were recorded and following statistics were found:

$$\sum A = 168, \sum H = 1275, \sum AH = 20704, \sum A^2 = 2585, \sum H^2 = 320019.$$

- (a) Calculate the product moment correlation coefficient.
- (b) Test at the 5 % level the assertion that age and height are positively correlated.

Video: [PMCC and hypothesis testing](#)

[Correlation EQ](#)

[Solutions to Starter and E.g.s](#)

Exercise

p75 5A Qu 3-8

Summary

Success criteria – how to carry out a significance test for the PMCC value

1. Calculate the r -value.
2. State the null hypothesis: it is always $H_0 : \rho = 0$ where ρ is the population correlation coefficient
3. Decide whether you will carry out a 1-tailed or 2-tailed test (it is usually given in the question).
4. State the alternative hypothesis, H_1 :
 - $H_1 : \rho \neq 0$ – two-tailed test
 - $H_1 : \rho > 0$ – one-tailed test, it is believed there could be a positive correlation
 - $H_1 : \rho < 0$ – one-tailed test, it is believed there could be a negative correlation
5. Decide on the level of significance (it is usually given in the question)
6. Use the tables to find the critical value based on the number of values, n , and the significance level
7. Compare your r -value to the critical value:
 - If $|r - \text{value}| > \text{critical value}$ then reject $H_0 \Rightarrow$ there is evidence to suggest there is a correlation between the two variables
 - If $|r - \text{value}| < \text{critical value}$ then do not reject $H_0 \Rightarrow$ there is no evidence to suggest there is a correlation between the two variables

N.B. A hypothesis test on the PMCC value can only be carried out on **bivariate data that is normally distributed**.