

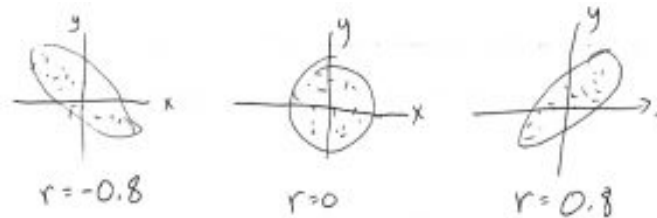
## 14.2: Correlation

The correlation coefficient we will use here is called the “Pearson product moment correlation coefficient” and will be represented by the following symbols :

$\rho$  — population correlation

$r$  — sample correlation

The correlation is always a number between  $-1$  and  $+1$  :  $-1 \leq r \leq +1$  and  $-1 \leq \rho \leq +1$ . If  $r$  (or  $\rho$ ) equals 0 then that means there is *no correlation* between  $x$  and  $y$ . A minus sign means a minus slope, a plus sign means a positive slope.



The formula for  $r$  is<sup>[1]</sup> :

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

**Example 14.1** : Compute the correlation between  $x$  and  $y$  for the data on [Section 14.1](#) used for the scatter plot.

**Solution** : To compute  $r$ , first make a table, fill in the data columns (on the right of the double vertical line below), fill in the other computed columns, sum the columns and finally plug the sums into the formula for  $r$  :

Subject	$x$	$y$	$xy$	$x^2$	$y^2$
A	6	82	492	36	6724
B	2	86	172	4	7396
C	15	43	645	225	1849
D	9	74	666	81	5476
E	12	58	696	144	3364
F	5	90	450	25	8100
G	8	78	624	64	6084
$n = 7$	$\sum x = 57$	$\sum y = 511$	$\sum xy = 3745$	$\sum x^2 = 579$	$\sum y^2 = 38993$

Plug in the numbers :

$$\begin{aligned}
 r &= \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n(\sum x^2) - (\sum x)^2][n(\sum y^2) - (\sum y)^2]}} \\
 &= \frac{7(3745) - (57)(511)}{\sqrt{[7(579) - (57)^2][7(38993) - (511)^2]}} \\
 &= -0.944
 \end{aligned}$$

Here there is a strong negative relationship between  $x$  and  $y$ . That is, as  $x$  goes up,  $y$  goes down with a fair degree of certainty. Note the  $r$  is **not** the slope. All we know here, from the correlation coefficient, is that the slope is negative and the scatterplot ellipse is long and skinny.

□

*Standard warning about correlation and causation* : If you find that  $x$  and  $y$  are highly correlated (i.e.  $r$  is close to  $+1$  or  $-1$ ) then you cannot say that  $x$  causes  $y$  or that  $y$  causes  $x$  or that there is and causal relationship between  $x$  and  $y$  at all. In other words, it is true that if  $x$  causes  $y$  or that  $y$  causes  $x$  then  $x$  will be correlated with  $y$  but the reverse implication does not logically follow. So beware of looking for relations between variables by looking at correlation alone. Simply finding correlations by themselves doesn't prove anything.

The significance of  $r$  is assessed by a **hypothesis test** of

$$H_0 : \rho = 0 \quad H_1 : \rho \neq 0 \quad (14.2.1)$$

To test this hypothesis, you need to convert  $r$  to  $t$  via:

$$t = r \sqrt{\frac{n-2}{1-r^2}} \quad (14.2.2)$$

and use  $\nu = n - 2$  to find  $t_{\text{crit}}$ . The [Pearson Correlation Coefficient Critical Values Table](#) offers a shortcut and lists critical  $r$  values that correspond to the critical  $t$  values.

**Example 14.2** : Given  $r = 0.897$ ,  $n = 6$  and  $\alpha = 0.05$ , test if  $r$  is significant.

*Solution* :

1. Hypothesis.  $H_0 : \rho = 0 \quad H_1 : \rho \neq 0$
2. Critical statistic.

From the [t Distribution Table](#) with  $\nu = n - 2 = 6 - 2 = 4$  and  $\alpha = 0.05$  for a two-tailed test find

$$t_{\text{crit}} = \pm 2.776 \quad (14.2.3)$$

As a short cut, you can also look in the [Pearson Correlation Coefficient Critical Values Table](#) for  $\alpha = 0.05$ ,  $\nu = 4$  to find the corresponding

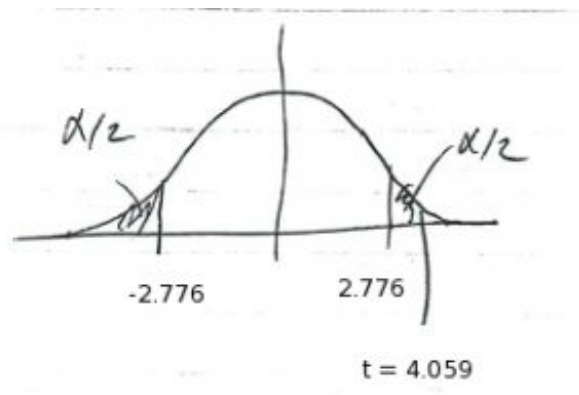
$$r_{\text{crit}} = \pm 0.811 \quad (14.2.4)$$

3. Test statistic.

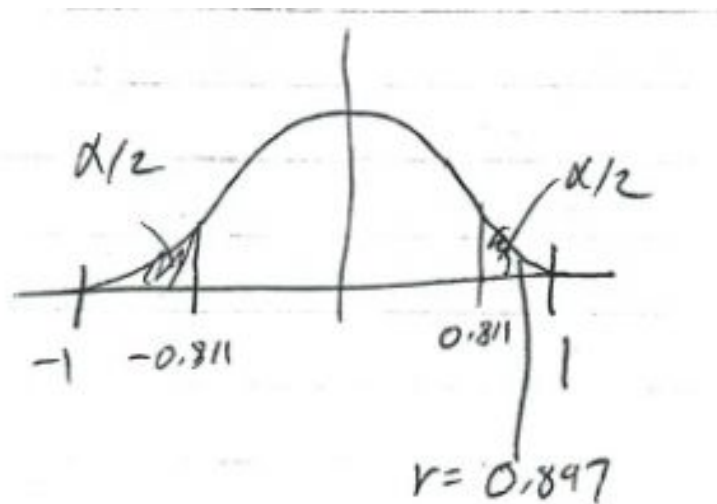
$$t_{\text{test}} = r \sqrt{\frac{n-2}{1-r^2}} = 0.897 \sqrt{\frac{6-2}{1-(0.897)^2}} = 4.059 \quad (14.2.5)$$

4. Decision.

Using the  $t$  :



or using the [Pearson Correlation Coefficient Critical Values Table](#) short cut :



we conclude that we can reject  $H_0$ .

5. Interpretation. The correlation is statistically significant at  $\alpha = 0.05$ .

□

1. The formula for  $\rho$  is the same with all  $x$  and  $y$  in the population used. ↩

This page titled [14.2: Correlation](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Gordon E. Sarty](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.