Calculating Correlations

In this *Guide*, let's go over how to calculate the correlation between two variables (variable 1, x, and variable 2, y). The correlation, known as a standard Pearson's correlation, is represented by r_{xy}.

Let's take a look at the formula for r_{xy}:

Pearon's Correlation

$$r_{xy} = \frac{n\sum XY - \sum X \sum Y}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}}$$

This formula is admittedly ugly and may look daunting, but in reality it is only made up of 6 values that are relatively easy to calculate with the right process. Here are the six values you'll need to find:

parameter	definition
n	sample size
$\sum X$	sum of X
$\sum Y$	sum of Y
$\sum X^2$	sum ofX^2
$\sum Y^2$	sum of Y^2
$\sum XY$	sum of XY

Here's some data for us to work with:

x	y
13	6
9	1
5	15
5	5
2	14
9	2
6	15
7	4
8	2
3	8
2	7
1	12

Step 1: Create a table

Let's start by creating a table that will, as is often the case, naturally lead us to the values we need. Since we already have x and y represented here, we'll need to create columns for x^2, y^2, and xy.

Here's what this will look like:

x	y	x^2	y^2	xy
13	6			
9	1			
5	15			
5	5			
2	14			
9	2			
6	15			
7	4			
8	2			
3	8			
2	7			
1	12			

Step 2: Square each x and y value

Great! Let's begin to fill out these values. To fill out the x^2 column, simply square each value of x and place each squared value in the table. To fill out the y^2 column, do the same for y (that is, square each value of y and place it in the new table). Like so:

x	y	x^2	y^2	xy
13	6	169	36	
9	1	81	1	
5	15	25	225	
5	5	25	25	
2	14	4	196	
9	2	81	4	
6	15	36	225	
7	4	49	16	
8	2	64	4	
3	8	9	64	
2	7	4	49	
1	12	1	144	

Step 3: Multiply x and y values

We're getting close! At this point, we only need to worry about the xy column. These are also easy to get! You simply need to multiply across (i.e., multiple each participant's–or row's–x value by its y value).

Once you do this, we'll have our full table:

x	y	x^2	y^2	xy
13	6	169	36	78
9	1	81	1	9
5	15	25	225	75
5	5	25	25	25
2	14	4	196	28
9	2	81	4	18
6	15	36	225	90
7	4	49	16	28
8	2	64	4	16
3	8	9	64	24
2	7	4	49	14
1	12	1	144	12

Step 4: Take the sum of each column

In order to get the values we need for the r_{xy} formula (e.g., $\sum X$, $\sum Y^2$; see above), we'll need to take the sum ($\sum$) of each column. Let's do that now.

$\sum X$	$\sum Y$	$\sum x^2$	$\sum y^2$	$\sum xy$
70	91	548	989	417

Step 5: Complete the equation

We're almost there! All we need to do now is to plug these values into the formula for r_{xy}, like so:

$$r_{xy} = \frac{(12)(417) - (70)(91)}{\sqrt{[(12)(548) - (70)^2][(12)(989] - (91)^2]}} \tag{1}$$

$$= \frac{5004 - 6370}{\sqrt{[6576 - 4900][11868 - 8281]}} \tag{2}$$

$$= \frac{-1366}{\sqrt{[1676][3587]}} \tag{3}$$

$$= \frac{-1366}{2451.9} \tag{4}$$

$$= -0.56 \tag{5}$$

So, the correlation, r_{xy}, between these two variables is -0.56.

A visual insepction of the data is a good idea. Here is a scatterplot of the data, indicating a negative correlation: