SAMLab Tip Sheet #7

Computing a Pearson r

In this Tip Sheet, you will learn how to calculate a Pearson r. To better understand the factors that affect the sign and magnitude of r, we will calculate r using the definitional formula shown below and then we will look at Excel's paste functions for computing r. Input the data below and calculate the deviation scores as shown.

$$r = \frac{\sum(x - \bar{x})(y - \bar{y})/N}{S_x S_y}$$

Note that the descriptive standard deviation is used. For help creating deviation scores, see page 6 of Tip Sheet #2, and for more information on standard deviations, see Tip Sheet #6.

B13 fx =STDEVP(B2:B11)

	A	B	C	D	E
1		X	Y	(X-Xbar)	(Y-Ybar)
2		45	17	-58.7	7.5
3		15	15	-88.7	5.5
4		115	10	11.3	0.5
5		89	8	-14.7	-1.5
6		207	1	103.3	-8.5
7		135	7	31.3	-2.5
8		175	4	71.3	-5.5
9		60	14	-43.7	4.5
10		141	9	37.3	-0.5
11		55	10	-48.7	0.5
12	Mean	103.7	9.5		
13	Standard Deviation	58.36	4.67		
14					

The data (fictitious) in this example represent the number of college credits a student has earned (**X**) and the student's score on a metric of test anxiety (**Y**) administered before the final exam period.

Next we are going to create the **covariance**, which is the numerator of our definitional equation above. The covariance is actually just the average **cross-product**. To calculate the covariance, create a column of cross-products (labeled **(X – X)(Y – Y)** below), sum them, and divide by N.

A **cross-product** is the product of paired deviation scores. The average of a group of cross-products is the **covariance**.

fx =D2*E2

D	E	F
(X-Xbar)	(Y-Ybar)	(X-Xbar)(Y-Ybar)
-58.7	7.5	-440.25
-88.7	5.5	-487.85
11.3	0.5	5.65
-14.7	-1.5	22.05
103.3	-8.5	-878.05
31.3	-2.5	-78.25
71.3	-5.5	-392.15
-43.7	4.5	-196.65
37.3	-0.5	-18.65
-48.7	0.5	-24.35
	Sum	-2488.5
Covariance (Sum/N)		-248.85

Looking at the formula above, you can see that only the numerator (covariance) affects the sign of r. Also, note that because deviation scores are multiplied together, those data points that deviate greatly from one or both of the means can have a large impact on the magnitude of r.

SAMLab Tip Sheet #7
The next step is to compute the Pearson r. Divide the covariance by the product of the two standard deviations. Note that the formula used to compute r is displayed in the formula bar.

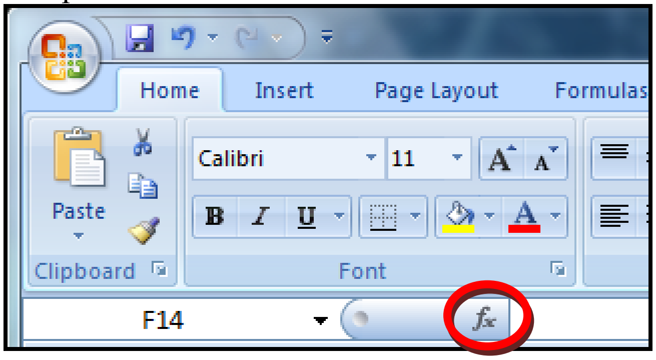
F14 | fx | =F13/(B13*C13)

	A	B	C	D	E	F	G
1		**X**	**Y**	**(X-Xbar)**	**(Y-Ybar)**	**(X-Xbar)(Y-Ybar)**	
2		45	17	-58.7	7.5	-440.25	
3		15	15	-88.7	5.5	-487.85	
4		115	10	11.3	0.5	5.65	
5		89	8	-14.7	-1.5	22.05	
6		207	1	103.3	-8.5	-878.05	
7		135	7	31.3	-2.5	-78.25	
8		175	4	71.3	-5.5	-392.15	
9		60	14	-43.7	4.5	-196.65	
10		141	9	37.3	-0.5	-18.65	
11		55	10	-48.7	0.5	-24.35	
12	**Mean**	103.7	9.5		**Sum**	-2488.5	
13	**Standard Deviation**	58.36	4.67	**Covariance (Sum/N)**		-248.85	
14					**r**	-0.91	
15							

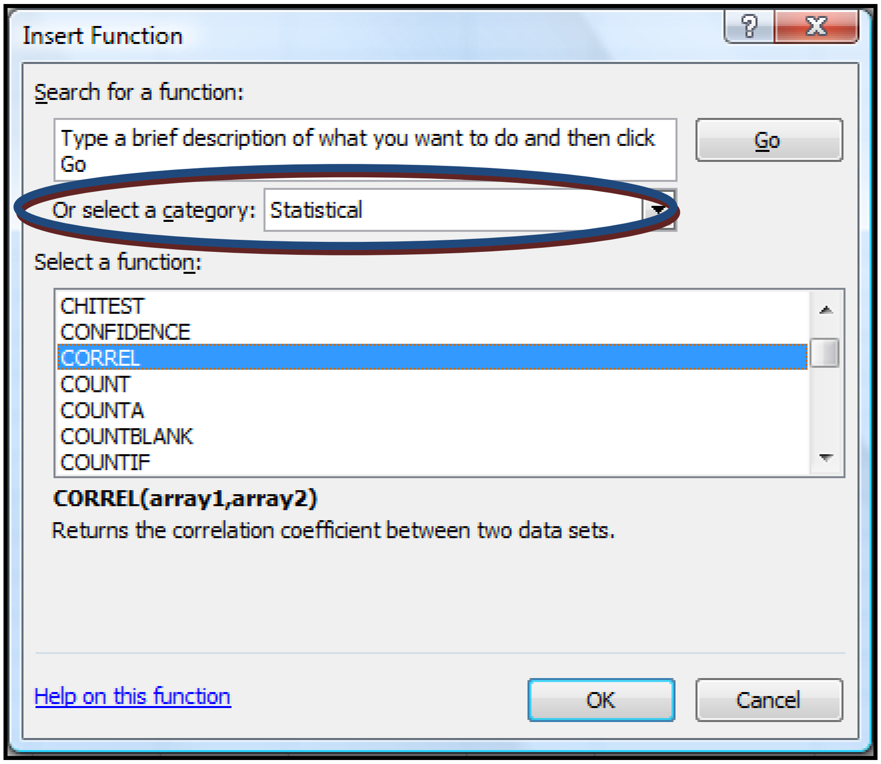
Using the Paste Functions

There are two paste functions that will calculate a Pearson r for you when given only the raw data. They are CORREL and PEARSON. We will use CORREL because it uses the definitional formula to compute r. PEARSON will usually work fine; however, if you are trying to find correlations for very large numbers PEARSON may return a value containing significant rounding errors because it uses a different formula, which is often referred to as the computational or calculator formula. The use of CORREL is illustrated below.

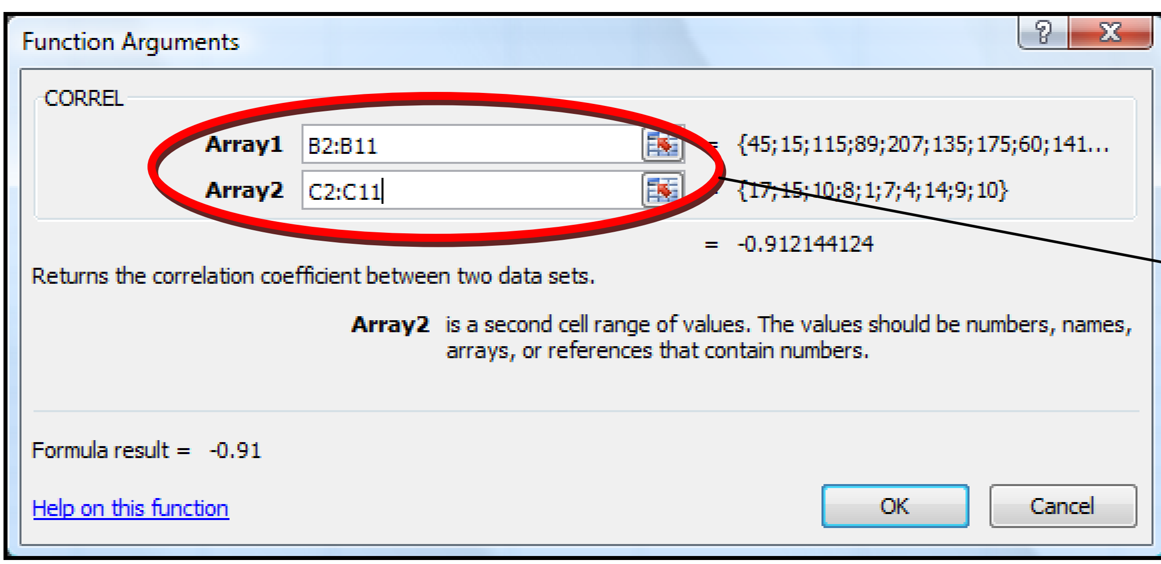
SAMLab Tip Sheet #7



First, click on the f_x symbol next to the function bar (circled in red to the left)



The "Insert Function" window will open. Select "Statistical" from the "Or select a category" menu (circled in blue to the left). Then, scroll down and highlight "CORREL" in the field below "Select a function" (highlighted to the left). Click "OK."



The two columns of data are selected in the "Array1" and "Array2" fields. Press OK and you should see output similar to that on the next page.

SAMLab Tip Sheet #7

=CORREL(B2:B11,C2:C11)

	B	C	D
	X	**Y**	
	45	17	
	15	15	
	115	10	
	89	8	
	207	1	
	135	7	
	175	4	
	60	14	
	141	9	
	55	10	
r	-0.91		

In the image below you can see that PEARSON gives the same answer in this case.

=PEARSON(B2:B11,C2:C11)

	B	C	D
	X	**Y**	
	45	17	
	15	15	
	115	10	
	89	8	
	207	1	
	135	7	
	175	4	
	60	14	
	141	9	
	55	10	
r	-0.91		