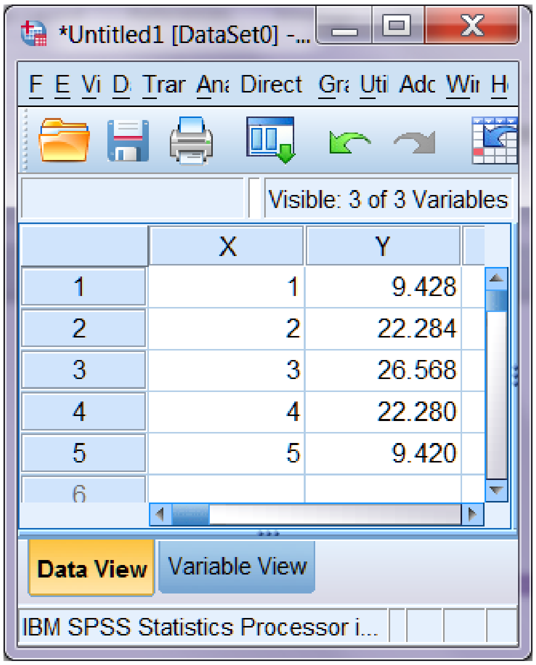
Linear *r* = 0. What Does That Mean?

Consider this small set of data:



	X	Y
1	1	9.428
2	2	22.284
3	3	26.568
4	4	22.280
5	5	9.420

Let's see how well they are correlated, using a linear model.

Model Summary

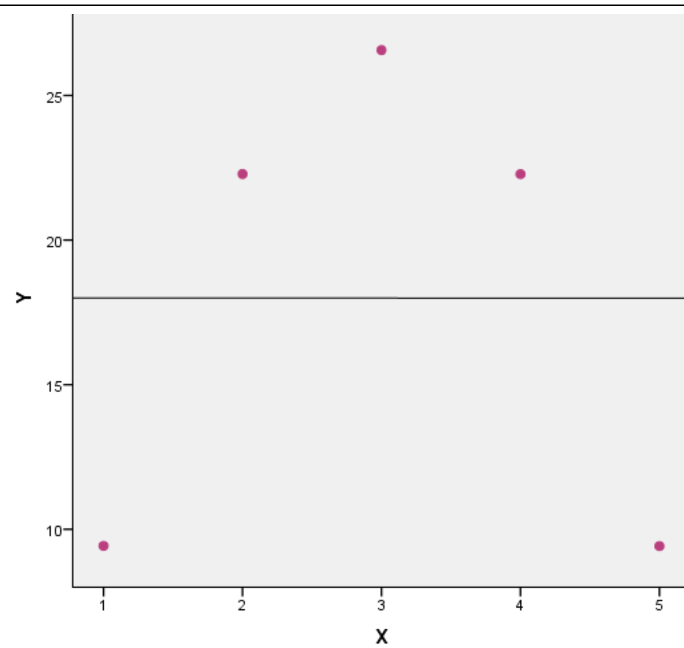
Model	**R**	R Square	Adjusted R Square	Std. Error of the Estimate
1	**.000**[a]	.000	-.333	9.259

The Pearson (linear) *r* = 0.

The prediction equation is $\hat{Y} = M_Y = 18$.

The *t* for testing the null that the rho is zero has value 0, the *p* value is 1.

To the right is a scatter plot with the regression line drawn in. The line is flat, slope = 0. Clearly these data cannot be fit well with a straight line model.



OK, I am going to try a quadratic model. In a quadratic model one predicts Y from both X and X^2. Here are the results of such an analysis,

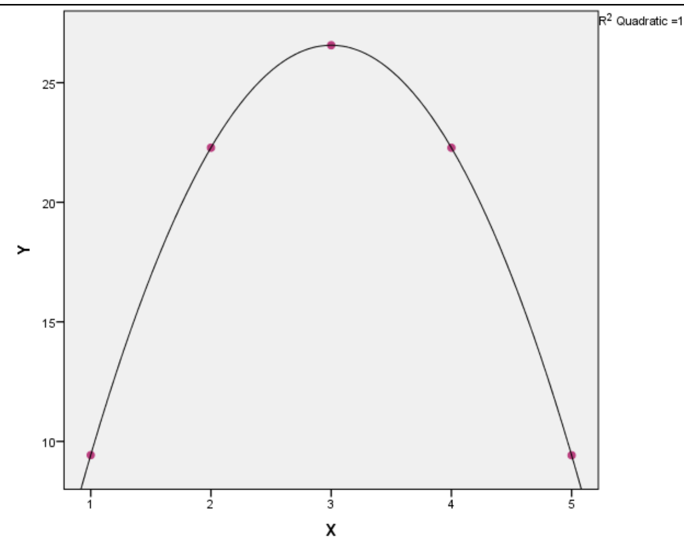
Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	**1.000**[a]	1.000	1.000	.000

Holy moly, the quadratic R = 1, a perfect correlation, absolutely no error in prediction.

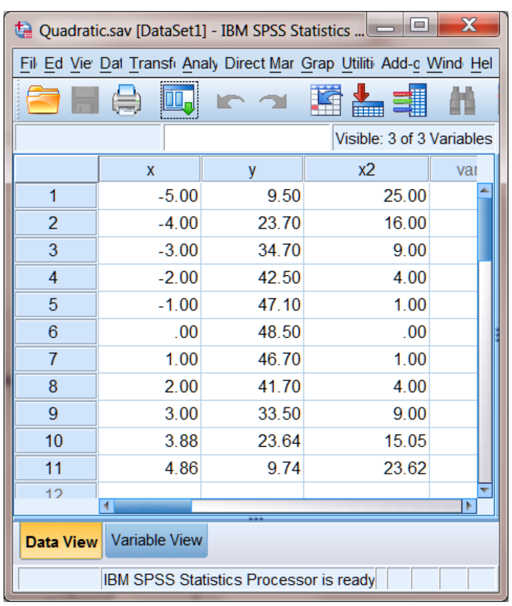
The prediction equation is $\hat{Y} = -12 + 25.714X - 4.286X^2$. The F statistic testing the null that the population quadratic R = 0 has a value that is infinitely larger, leading to a p that is infinitely small.

Take home message: Getting a value of r that is very small (close to zero) does not necessarily mean that the variables are not related to each other in any way – they could, in fact, be very highly related, just not in a linear fashion.



Here is another example.

An r of zero indicates that there is no <u>linear</u> relationship between the two variables. There may, however, be a strong nonlinear relationship between the two variables. Consider these data:



	x	y	x2	var
1	-5.00	9.50	25.00	
2	-4.00	23.70	16.00	
3	-3.00	34.70	9.00	
4	-2.00	42.50	4.00	
5	-1.00	47.10	1.00	
6	.00	48.50	.00	
7	1.00	46.70	1.00	
8	2.00	41.70	4.00	
9	3.00	33.50	9.00	
10	3.88	23.64	15.05	
11	4.86	9.74	23.62	

and this output:

Model 1 is a linear model – Y = *a* + *b*X.

Model 2 is a nonlinear model - $Y = a + b_1 X + b_2 X^2$ -- a quadratic model.

Model Summary

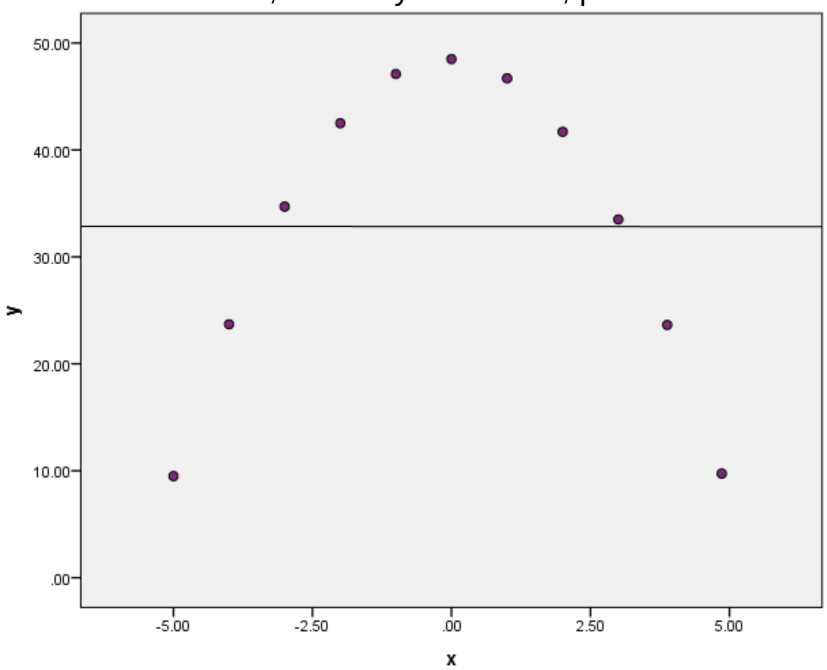
Model	R Square
1	.000
2	1.000

Coefficients[a]

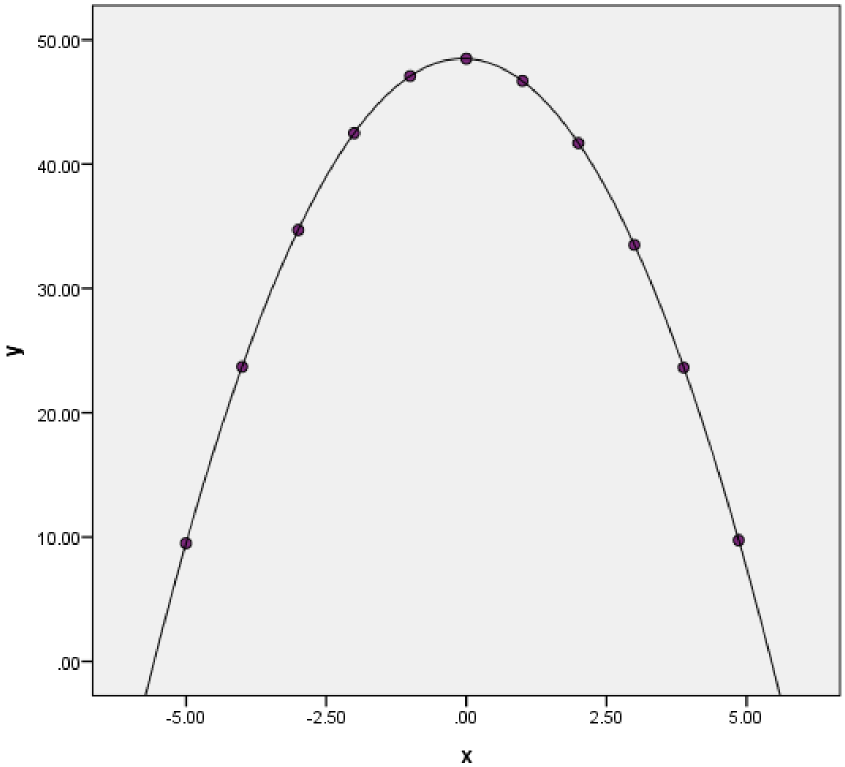
Model		Unstandardized Coefficients
		B
1	(Constant)	32.843
	x	-.003
2	(Constant)	48.500
	x	-.200
	x2	-1.600

a. Dependent Variable: y

Linear model: *r* = 0, for every value of X, predicted Y = the mean on Y



Quadratic model: *R* = 1. Every value of Y is perfectly predicted.



Karl L. Wuensch, 11-June-2014.