

### 12.2.3: Coefficient of Determination

The **coefficient of determination**  $R^2$  (or  $r^2$ ) is the fraction (or percent) of the variation in the values of  $y$  that is explained by the least-squares regression of  $y$  on  $x$ .  $R^2$  is a measure of how well the values of  $y$  are explained by  $x$ .

For example, there is some variability in the dependent variable values, such as grade. Some of the variation in student's grades is due to hours studied and some is due to other factors. How much of the variation in a student's grade is due to hours studied?

When considering this question, you want to look at how much of the variation in a student's grade is explained by the number of hours they studied and how much is explained by other variables. Realize that some of the changes in grades have to do with other factors. You can have two students who study the same number of hours, but one student may have a higher grade. Some variability is explained by the model and some variability is not explained. Together, both of these give the total variability.

$$\begin{aligned} (\text{Total Variation}) &= (\text{Explained Variation}) + (\text{Unexplained Variation}) \\ \sum (y - \bar{y})^2 &= \sum (\hat{y} - \bar{y})^2 + \sum (y - \hat{y})^2 \end{aligned}$$

#### Coefficient of Determination

The proportion of the variation that is explained by the model is

$$R^2 = \frac{\text{Explained Variation}}{\text{Total Variation}} = \frac{SSR}{SST}$$

Find and interpret the coefficient of determination for the hours studied and exam grade data.

Hours Studied for Exam 20 16 20 18 17 16 15 17 15 16 15 17 16 17 14 Grade on Exam 89 72 93 84 81 75 70 82 69 83 80 83 81 84 76

#### Solution

The coefficient of determination is this correlation coefficient squared. Note: when  $r$  is negative, then when you square  $r$  the answer becomes positive. For the hours studied and exam grade  $r = 0.825358$ , so  $r^2 = R^2 = 0.825358^2 = 0.6812$ .

Approximately 68% of the variation in a student's exam grade is explained by the least square regression equation and the number of hours a student studied.

**TI-84:** Press the [STAT] key, arrow over to the [TESTS] menu, arrow down to the option [LinRegTTest] and press the [ENTER] key. The default is Xlist:L1, Ylist:L2, Freq:1,  $\beta$  and  $\rho \neq 0$ . Arrow down to Calculate and press the [ENTER] key. The calculator returns the  $y$ -intercept =  $a = b_0$ , slope =  $b = b_1$ , the standard error of estimate =  $s$ , the coefficient of determination =  $r^2 = R^2$ , and the correlation coefficient =  $r$ .

<pre> EDIT CALC TESTS 1:Z-PropZInt... 2:X<sup>2</sup>-Test... 3:X<sup>2</sup>GOF-Test... 4:2-SampFTest... 5:LinRegTTest... 6:LinRegTInt... 7:ANOVA( </pre>	<pre> LinRegTTest Xlist:L1 Ylist:L2 Freq:1 8 &amp; ρ: [≠] &lt;0 &gt;0 RegEQ: Calculate </pre>	<pre> LinRegTTest y=a+bx 8≠0 and ρ≠0 t=5.270675167 p=1.5134617e-4 df=13 a=26.74198718 </pre>	<pre> LinRegTTest y=a+bx 8≠0 and ρ≠0 t=5.270675167 p=1.5134617e-4 df=13 a=26.74198718 </pre>
------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------

**TI-89:** In the Stats/List Editor select F6 for the Tests menu. Use cursor keys to select A:LinRegTTest and press [Enter]. In the "X List" space type in the name of your list with the  $x$  variable without spaces: for our example, "list1." In the "Y List" space type in the name of your list with the  $y$  variable without spaces: for our example, "list2." Under the "Alternate Hyp" menu, select the  $\neq$  sign that is the same as the problem's alternative hypothesis statement then press the [ENTER] key, arrow down to [Calculate] and press the [ENTER] key. The calculator returns the  $y$ -intercept of the regression line =  $a = b_0$ , the slope of the regression line =  $b = b_1$ , the correlation =  $r$ , and the coefficient of determination =  $r^2 = R^2$ .



### Excel:

Regression Statistics	
Multiple R	0.825358
R Square	0.681216
Adjusted R Square	0.656695
Standard Error	3.935892
Observations	15

Coefficient of Determination  $R^2$

### TI-84:

```
LinRegTTest
y=a+bx
8≠0 and P≠0
↑b=3.216346154
s=3.935892216
r²=.6812164131
r=.8253583544
```

### SPSS:

#### Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.825 <sup>a</sup>	.681	.657	3.93589

a. Predictors: (Constant), Hours Studied

The coefficient of determination can take on any value between 0 and 1, or 0% to 100%. The closer  $R^2$  is to 100%, the better the regression equation models the data. Unlike  $r$ , which can only be used for simple linear regression, we can use  $R^2$  for different types of regression. In more advanced courses, if we were to do non-linear or multiple linear regression we could compare different models and pick the one that has the highest  $R^2$ .

For instance, if we ran a linear regression the scatterplot that showed an obvious curve pattern we would get a regression equation with a zero slope and  $R^2 = 0$ . See Figure 12-16.

If we were to fit a parabola through the data we get a perfect fit and  $R^2 = 1$ . See Figure 12-17.

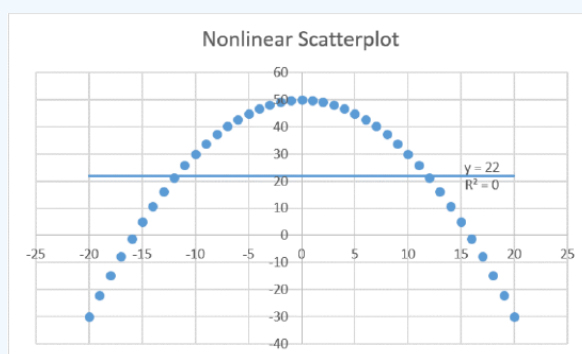


Figure 12-16: Linear regression on parabolic scatterplot.

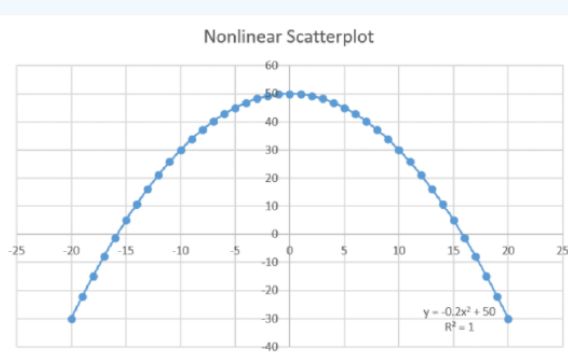


Figure 12-17: Nonlinear regression on parabolic scatterplot.