

# Math 21 – Pointers for Section 4.2

In this section we learn to find the equation that describes the relation between two numerical variables. We can use the equation to predict a person's height based on their shoe size.

## Least Squares Regression Line

This is also called the line of best fit. It gives us an equation that predicts values of the dependent/response variable ( $y$ ) in terms of the independent/explanatory variable ( $x$ ).

Before finding the equation, you are supposed to use the techniques from Section 4.1 to determine that there is indeed a linear relation or association between the two variables.

We will start with the procedure for calculating the equation of the line, as well as how to graph the line, with StatCrunch.

### StatCrunch

- Enter all of the ordered pairs in two columns, 1 ordered pair per line.  
Put the  $x$ -variable in the first column and the  $y$ -variable in the second column.
- Press the Stat button, and select Regression > Simple Linear.
- Select the correct columns for the  $x$ -column and  $y$ -column.
- Click on Compute!

I worked out an example for a scenario comparing hours worked ( $x$ ) by a student during the week before a final exam and that student's score ( $y$ ) on that final exam.

I have included the results screen that StatCrunch generates on the next page.

## StatCrunch Output Screen

**Options** (1 of 2)

**Simple linear regression results:**  
Dependent Variable: Score  
Independent Variable: Hours  
Score = 92.933824 - 0.81029412 Hours  
Sample size: 10  
R (correlation coefficient) = -0.8113687  
R-sq = 0.65831917  
Estimate of error standard deviation: 7.6113064

**Parameter estimates:**

Parameter	Estimate	Std. Err.	Alternative	DF	T-Stat	P-Value
Intercept	92.933824	3.6044555	≠ 0	8	25.783041	<0.0001
Slope	-0.81029412	0.20639052	≠ 0	8	-3.9260239	0.0044

**Analysis of variance table for regression model:**

Source	DF	SS	MS	F-stat	P-value
Model	1	892.94412	892.94412	15.413663	0.0044
Error	8	463.45588	57.931985		
Total	9	1356.4			

Click here to see fitted line / scatter plot

The first thing to look at is the correlation coefficient. In this example,  $r = -0.811$ .

To determine whether a linear relation exists, look at the circled P-value on the output screen. (It is on the line labeled "Slope".) If this P-value is less than 0.05 then we can conclude that there is a linear relation between the two variables.

The relation is a positive linear relation if the value of  $r$  is positive.

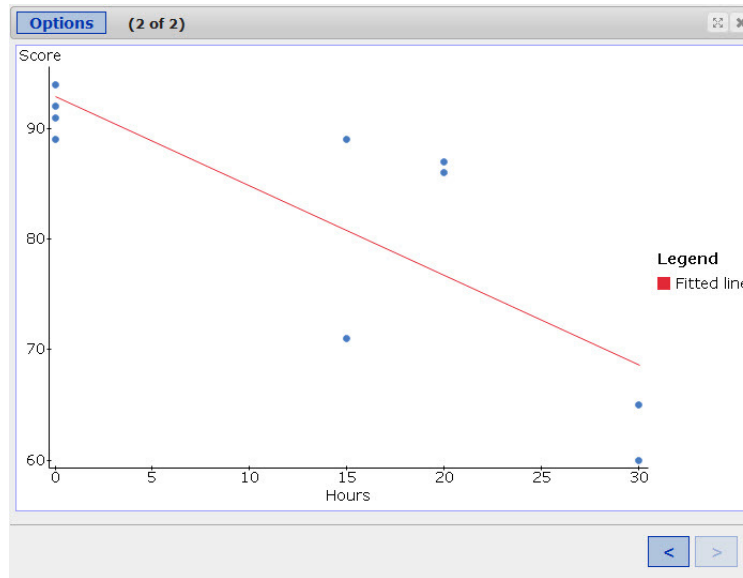
The relation is a negative linear relation if the value of  $r$  is negative.

Once you know there is a relation, look at the regression equation.

In this example, the equation is  $Score = 92.9 - 0.8Hours$ .

To predict the score of a student who worked for 10 hours, plug in 10 for  $x$  (Hours).

To see the fitted line plot, click the arrow in the lower left corner of the screen.



We can interpret the slope and y-intercept of the equation.

The slope is the second value listed in the equation, the one that is being multiplied by the independent variable (x).

In the example, the slope is -0.8:  $Score = 92.9 - 0.8Hours$

A slope of -0.8 means that for every extra hour worked, the student's score will drop by 0.8 points.

The y-intercept is the value listed first in the equation.

In this example, the y-intercept is 92.9:  $Score = 92.9 - 0.8Hours$

The y-intercept of 92.9 means that if a student worked 0 hours, their predicted score is 92.9.

### Residual

The residual for a value of x is the difference between the actual value of y from the data and the predicted value of y from the equation.

$$\text{Residual} = \text{Observed} - \text{Expected}$$