

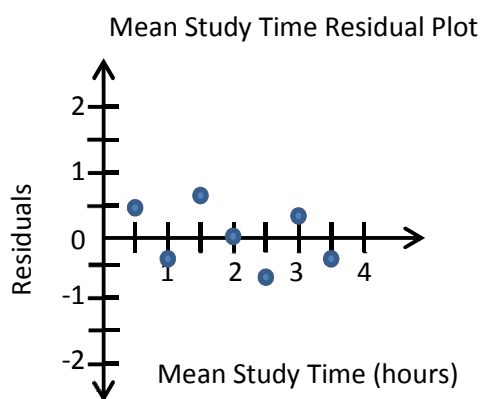
## COMPLETED

A scatter plot can be modeled as a linear function using the line of best fit, or a linear regression. This line then needs to be tested for a goodness of fit. If a graph is to be a good predictor of data, the line must be accurate in predicting data. This can be found by analyzing the residuals.

Residuals are **error distances**. They measure the goodness of a fit. To find the residuals of a regression model you find the difference between the observed value (from a table) and the models predicted value (from the equation of regression).

$$\text{Residual} = [\text{Observed Value}] - [\text{Predicted Value}]$$

Residuals are represented by graphing them. The plot resembles a coordinate plane but only utilizing the 1<sup>st</sup> and 4<sup>th</sup> quadrants. The residual values are the vertical axis (y-axis) and the independent variable (x) on the horizontal axis.



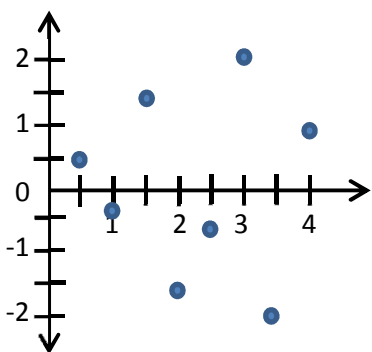
← Example

The residuals should add up to **0**. If they do add up to zero then the regression line was properly calculated. The goal is to minimize all of the residuals. We want the residuals to be “short” or **close** to the **x-axis**. The “longer” or more dispersed the residuals, the **further** they are from the **x-axis**.

If the residuals are clustered towards the x-axis, then the regression line is a **good** predictor of the data.

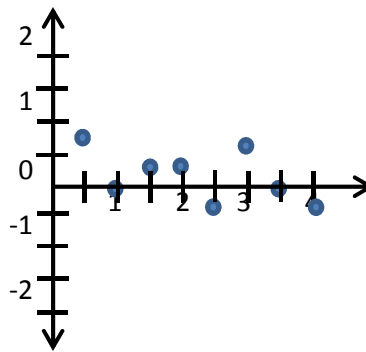
Identify if the linear regression line is a good predictor of the data based on the residuals plot. State why.

Ex1)



**No this is not a good predictor because the residuals are dispersed from the x-axis.**

Ex2)

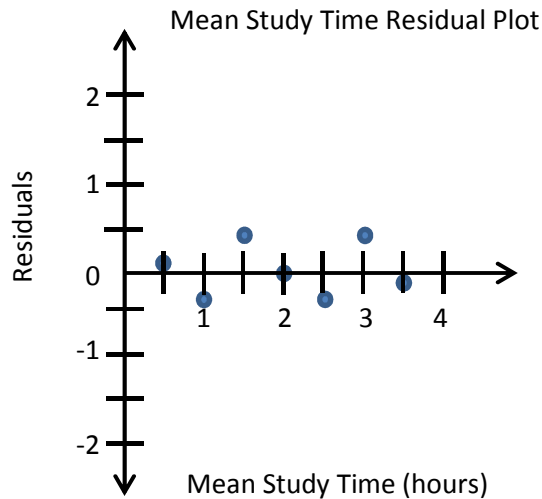


**Yes this is a good predictor because the residuals are short and closer to the x-axis.**

Ex3) Using the data that Melissa collected from Class 1, let's calculate the residuals.

Class 1 Test Score Analysis			
Mean Study Time (in hours)	Mean Test Score (Observed)	Regression Model (Predicted) $y = 8.7x + 58.6$	Residual
0.5	63	62.95	0.05
1	67	67.3	-0.3
1.5	72	71.65	0.35
2	76	76	0
2.5	80	80.35	-0.35
3	85	84.7	0.3
3.5	89	89.05	-0.05

- First calculate the predicted values by using the regression model and the mean study time.
- Subtract the Observed value and the Predicted Value
- Graph the residual



a) What do all the residuals add up to?

$$0.05 + -0.3 + 0.35 + 0 + -0.35 + 0.3 + -0.05 = 0$$

b) Would this regression line be a good predictor of the data

**Yes, this regression line would be a good predictor because the residuals are clustered towards to the x-axis and the residuals add up to zero.**