Lesson 22: Residual Plots

Daily Data Collection

Select two topics you think are correlated, make a hypothesis,

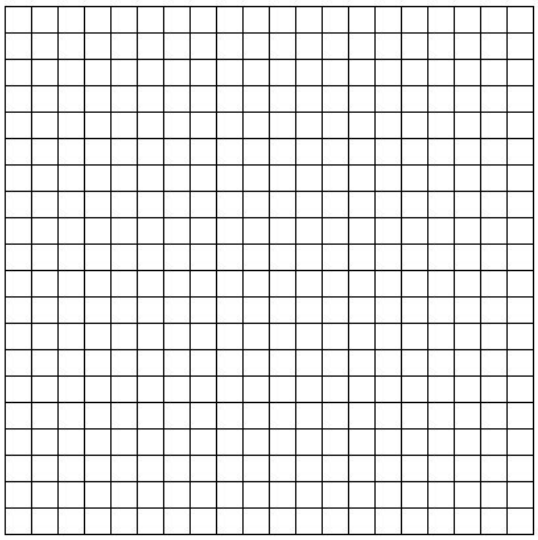
and run a test to see if your assumptions were true.

Class Data:

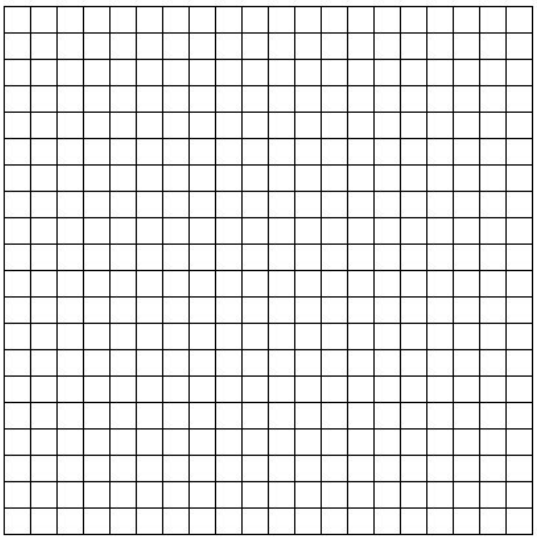
Explanatory Variable:

Response variable:

Create a scatterplot.



Create a residual plot.



Describe the Direction, Form, and Strength

Write an equation for the regression line

Describe the slope in the context of the situation

Find the residual value for your own data point

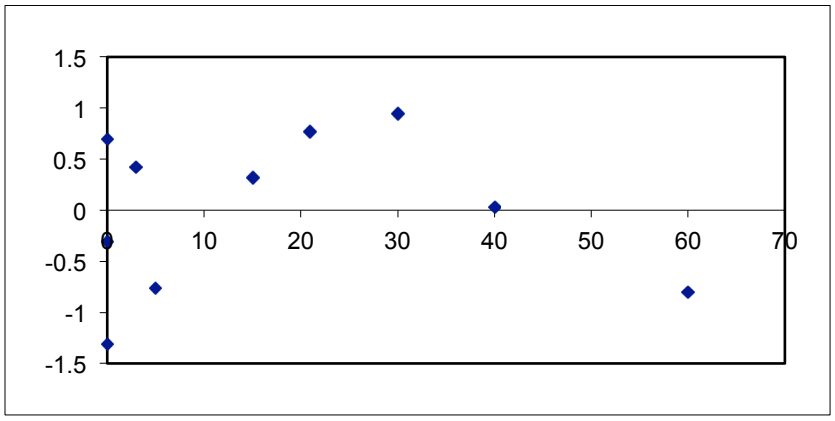
Conclusion/Analysis

Residual Plots:

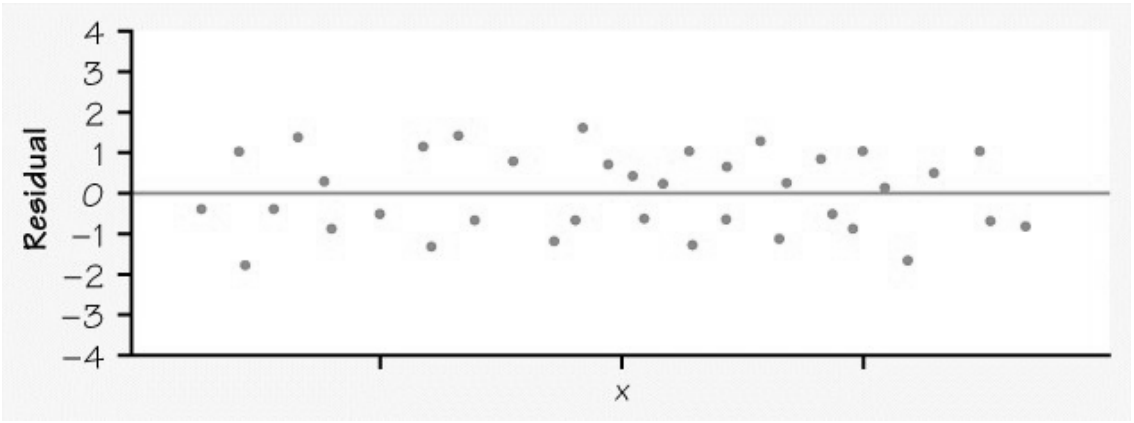
Residuals help us determine how well our data can be modeled by a straight line, by enabling us to construct a residual plot. A residual plot is a scatter diagram that plots the residuals on the y-axis and their corresponding x values on the x-axis. So for our example, the residual plot would be plotted as follows:

0	1	1.3069	-0.3069
3	2	1.5817	0.4183
21	4	3.2305	0.7695
15	3	2.6809	0.3191
30	5	4.0549	0.9451
5	1	1.7649	-0.7649
40	5	4.9709	0.0291
60	6	6.8029	-0.8029
0	2	1.3069	0.6931
0	0	1.3069	-1.3069
			0

Notice that the sum of the residuals is zero (any error is due to rounding)



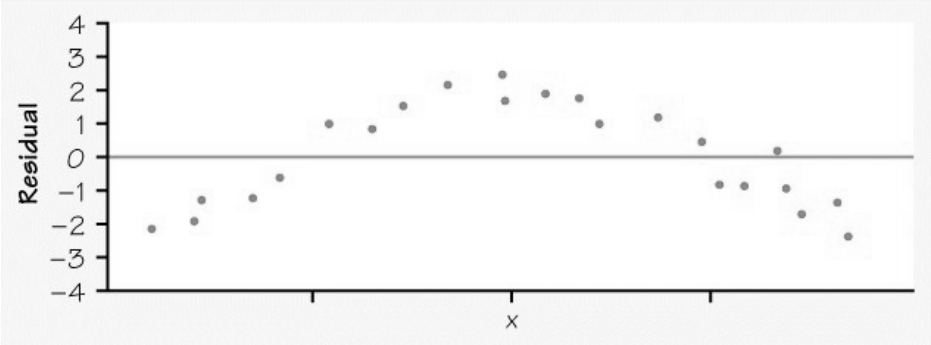
How does a residual plot help us? The pattern of the plot is the indicator of whether or not the data can be modeled linearly. The ideal plot is the one below:



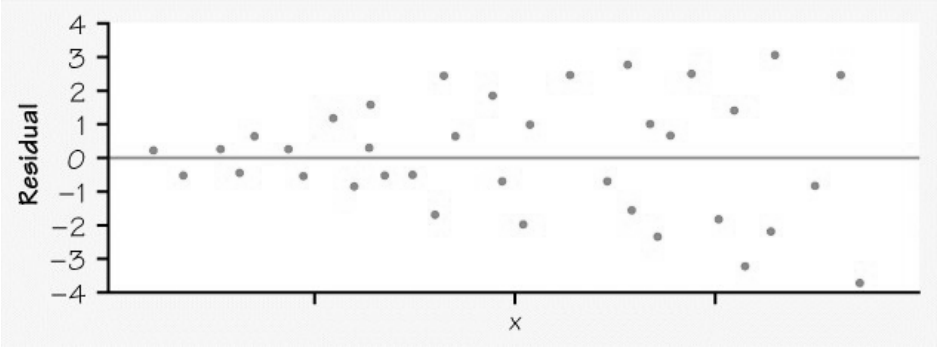
The plot shows a uniform scatter of the points above and below the fitted line with no unusual individual observations.

Here are a few things to watch out for when examining the pattern of a residual plot:

1) A curved pattern: Shows that the overall pattern of the data is not linear



2) Increasing or decreasing spread about the line as x increases: Predictions for y will be less accurate for larger values of x



3) Individual points with large residuals: indicate outliers from the overall pattern

Outlier: An observation that lies outside the overall pattern in the scatter plot (either in the x or y direction)

Influential point: A point is influential if removing it would markedly change the position of the regression line. Points that are outliers in the x direction are often influential. Influential points often have small residuals because they tend to pull the line towards themselves. Therefore, you might miss influential points if you only look at residuals.

HW 22 Section 3-2: 59-62