Warm-Up (10/16/17)

Use the following equation to complete the table below:

$$y = 1514 + 52.7x$$

Time (minutes)	Cost ($)	Predicted Value ($)	Residual ($)
14	1,510	$2251.8	-741.8
80	6,178	$5730	448
84	5,912	$5940.8	-28.8
118	9,184	$7732.6	1451.4
149	8,855	$9366.7	-511.3
192	11,023	$11632.4	-609.4

Recall: **Actual Value** - **Predicted Value** = **Residual**

Actual Values are given y-values in the table

Predicted values are y-values that are obtained by substituting the x-value into the given equation

Module 2: Lesson 16

More on Modeling Relationships with a Line

Lesson Summary

- The predicted y-value is calculated using the equation of the least squares line.
- The residual is calculated using

$$\text{residual} = \text{actual } y\text{-value} - \text{predicted } y\text{-value}.$$

- To make a residual plot, plot the x-values on the horizontal axis and the residuals on the vertical axis.

The curb weight of a car is the weight of the car without luggage or passengers. The table below shows the curb weights (in hundreds of pounds) and fuel efficiencies (in miles per gallon) of five compact cars.
Using a calculator, the least squares line for this data set was found to have the equation:

$$y = 78.62 - 1.5290x,$$

where x is the curb weight (in hundreds of pounds), and y is the predicted fuel efficiency (in miles per gallon).

Use the least squares line to calculate the residual values for the following chart:

Actual y-value

Predicted y-Value

Curb Weight (hundreds of pounds)	Fuel Efficiency (mpg)	Predicted Fuel Efficiency	Residual Value
25.33	43	39.89	3.11
26.94	38	37.43	.57
27.79	30	36.13	−6.13
30.12	34	32.57	1.43
32.47	30	28.97	1.03

Calculating Residual Values

$43 - 39.89 = 3.11$

$38 - 37.43 = 0.57$

$30 - 36.13 = -6.13$

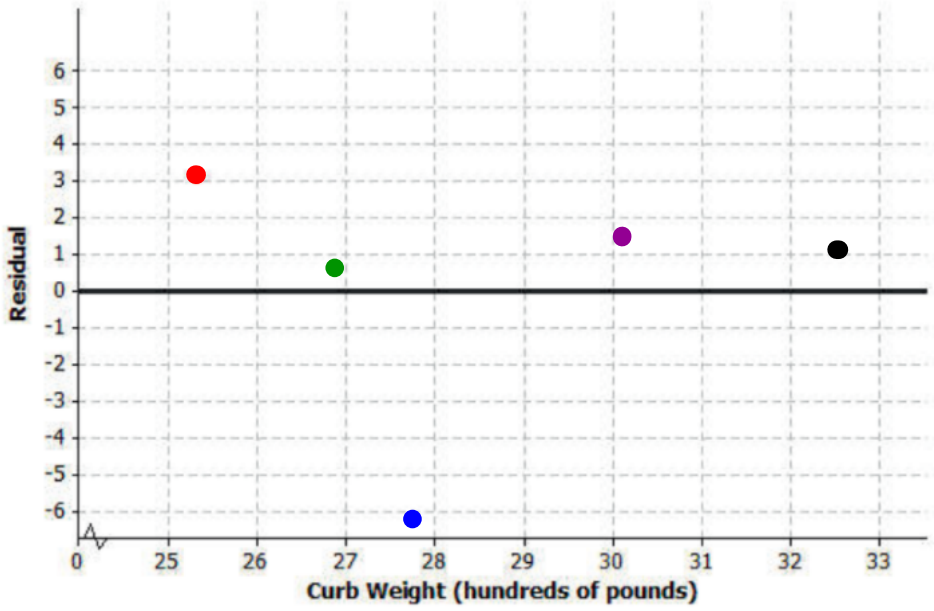
$34 - 32.57 = 1.43$

$30 - 28.97 = 1.03$

We can plot the residual values we previously calculated on a graph where our y-values are our residual values. This type of graph is called a residual graph.

(X-values are the same as before)

Curb Weight (hundreds of pounds)	Fuel Efficiency (mpg)	Predicted Fuel Efficiency	Residual Value
25.33	43		3.11
26.94	38		.57
27.79	30		-6.13
30.12	34		1.43
32.47	30		1.03



Example

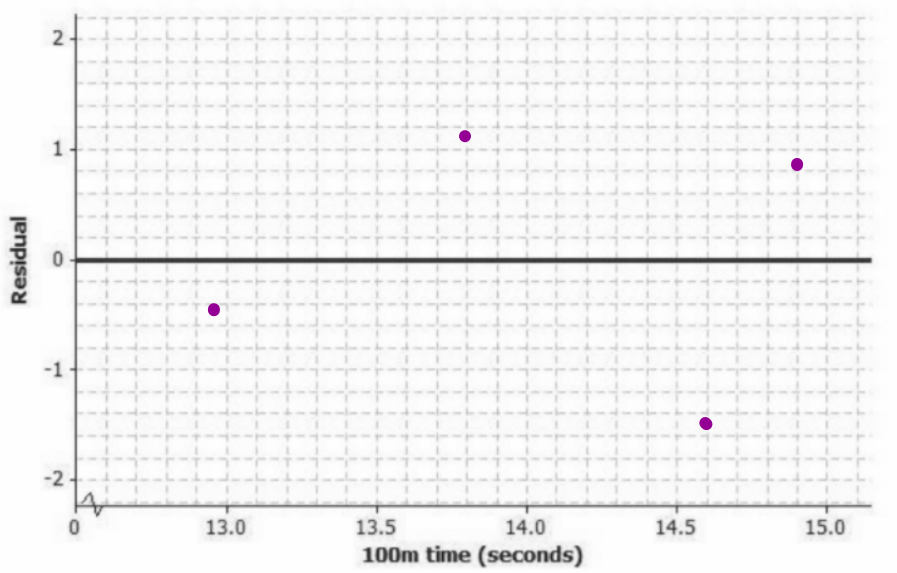
Residual = Actual − Prediction

Four athletes on a track team are comparing their personal bests in the 100 meter and 200 meter events. A table of their best times is shown below.

Athlete	100 m time (seconds)	200 m time (seconds)	Residual (seconds)	Predicted 200 m Time
1	12.95	26.68	-.42	27.10
2	13.81	29.48	1.08	28.40
3	14.66	28.11	-1.57	29.68
4	14.88	30.93	0.91	30.02

The least squares line is represented by the equation $y = 7.526 + 1.5115x$ where x represents 100-meter time and y represents 200-meter time. Calculate the residual values plot those points on a graph.

Using the axes provided below, construct a residual plot for this data set.



Module 2: Lesson 17 and Lesson 18

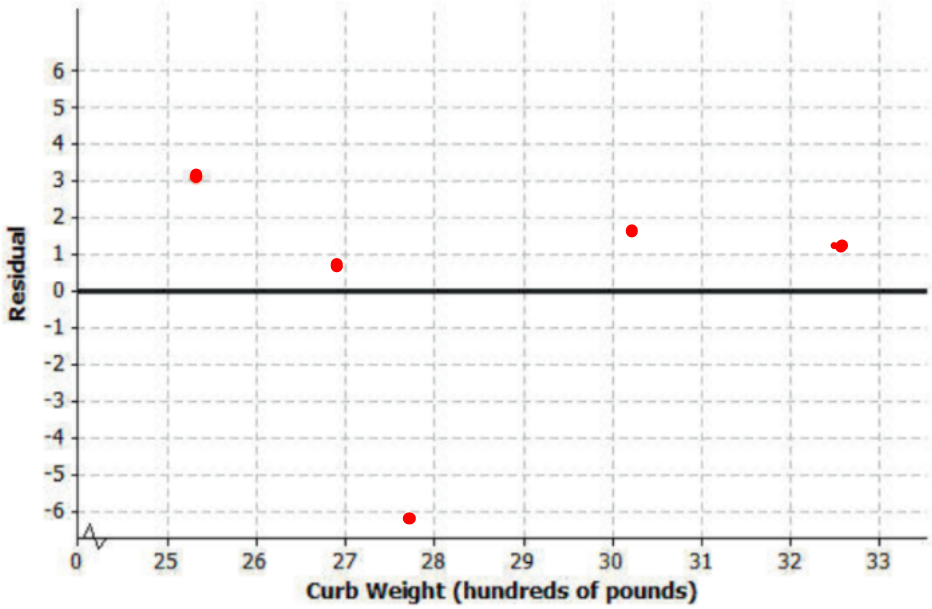
Analyzing Residuals

Lesson Summary

- After fitting a line, the residual plot can be constructed using a graphing calculator.
- A pattern in the residual plot indicates that the relationship in the original data set is not linear.
- A curve or pattern in the residual plot indicates a nonlinear relationship in the original data set.
- A random scatter of points in the residual plot indicates a linear relationship in the original data set.

Recall this example in from Lesson 16.

Curb Weight (hundreds of pounds)	Fuel Efficiency (mpg)	[redacted]	Residual Value
25.33	43		3.11
26.94	38		-.57
27.79	30		-6.13
30.12	34		1.43
32.47	30		1.03



What kind of pattern does this make?

No pattern

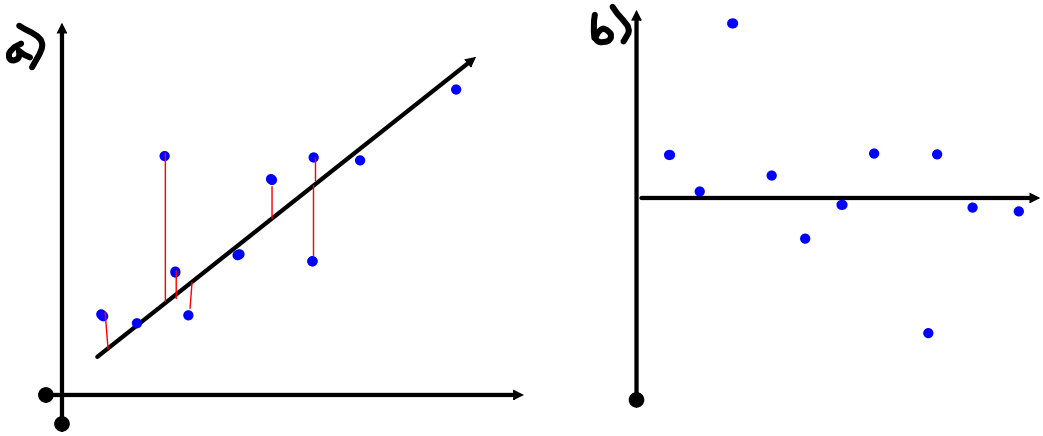
What does this kind of pattern tell us about the relationship that we used?

The scatter plot from the residual plot is Linear.

Example:

(a) Shows a linear scatter plot

(b) Shows the residual plot of that scatter plot



Recall that the residual is just the distance between the actual y-value (a given point on the scatter plot) and the predicted value on the line of best fit. This is indicated on graph (a) with a red line from the point to the line of best fit.

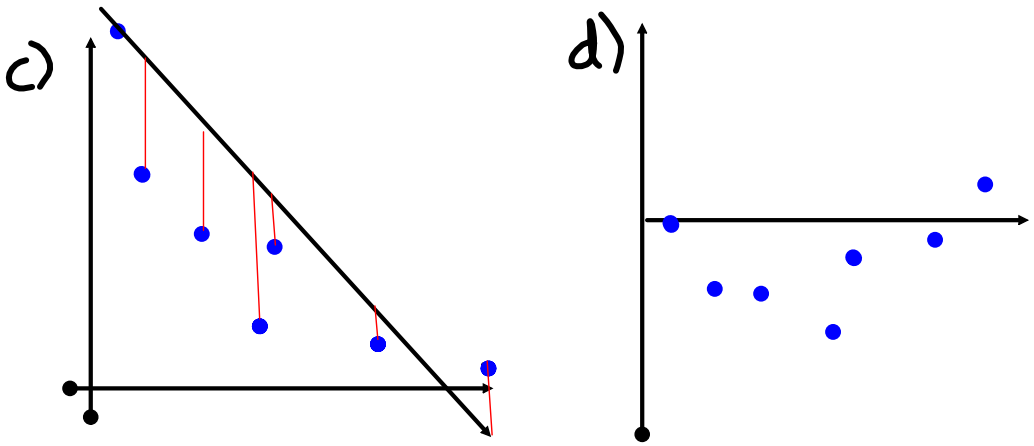
Graph (b) is made based on the residual value. If the point on graph (a) is above the line of best fit, then the point on the residual plot will be positive. If the point on graph (a) is below the line of best fit, then it will be negative on the residual graph. The x-values are the same in both graphs however.

Our result is a scattered residual plot for graph (b).

Example:

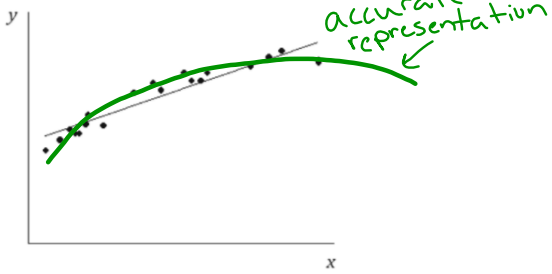
(c) Shows an exponential scatter plot fitted with a linear correlation

(d) Shows the residual plot for graph (c)

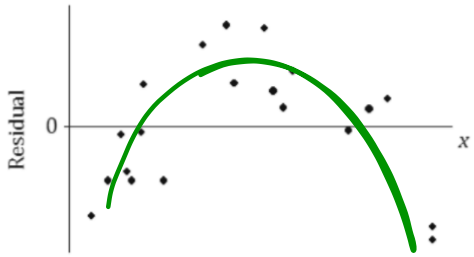


The shape of the residual plot does matter

Suppose you are given a scatter plot and least squares line that looks like this:



The residual plot has an arch shape like this:



What kind of shape does this residual plot make?

This residual plot makes a parabola

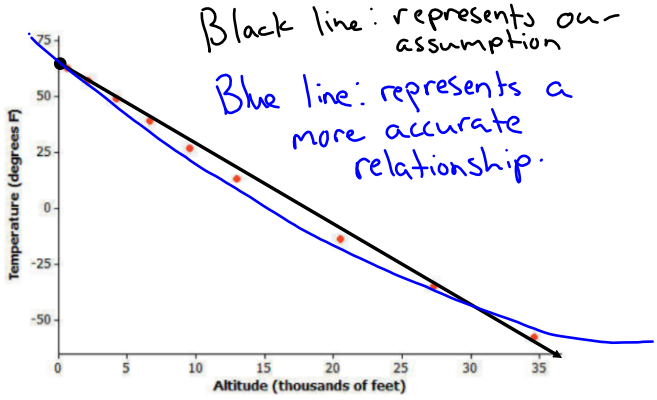
In general, if the residual plot of a linear relationship is scattered then we know that it's indeed linear. If there appears to be a pattern such as an exponential or a parabola then the relationship is instead nonlinear.

Is it even necessary to view these residual plots? Can't we just look at the regular graphs like we've been doing?

No. There will be cases where a nonlinear scatterplot will be mistaken for a linear relationship.

Example

The temperature (in degrees Fahrenheit) was measured at various altitudes (in thousands of feet) above Los Angeles. The scatter plot (below) seems to show a linear (straight-line) relationship between these two quantities.

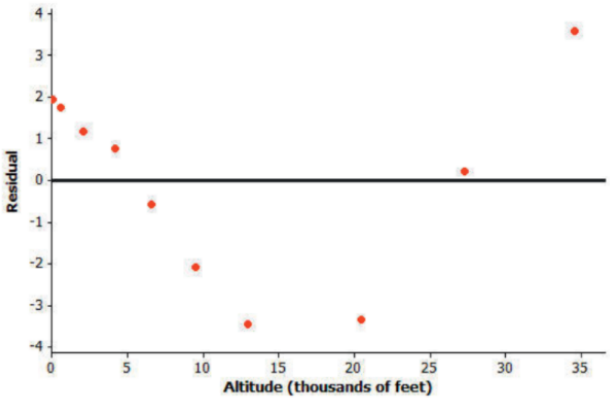


Data source: *Core Math Tools*, http://nctm.org

Since we have no other data to help us analyze this, what type of relationship should we expect from this type of scatter plot?

We expect this scatter plot to represent a Linear relationship.

However, look at the residual plot:



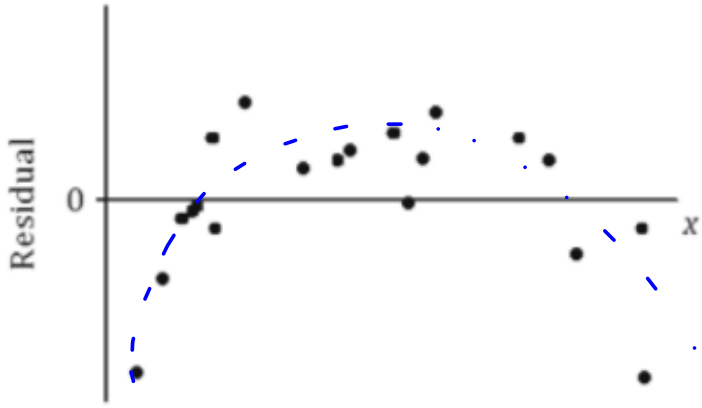
Based on what we previously learned about shapes of a residual plot, would we still think that the original scatter plot would be best represented by a linear relationship?

No. Since there's a clear pattern in the residual plot we know that the original scatter plot is non linear.

Practice Analyzing Residual Graphs

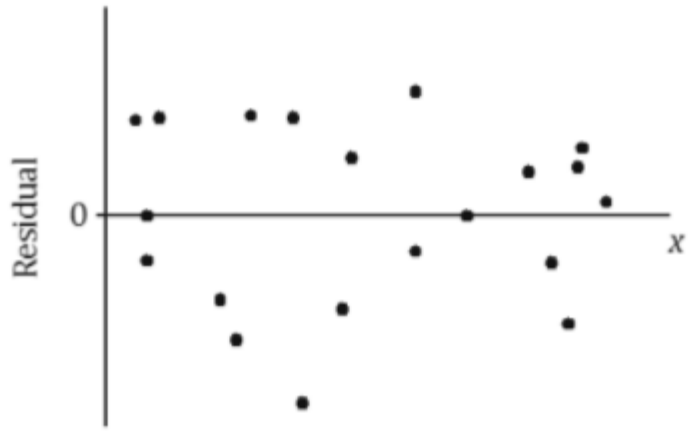
For each of the following residual plots, what conclusion would you reach about the relationship between the variables in the original data set? Indicate whether the values would be better represented by a linear or a nonlinear relationship.

a.



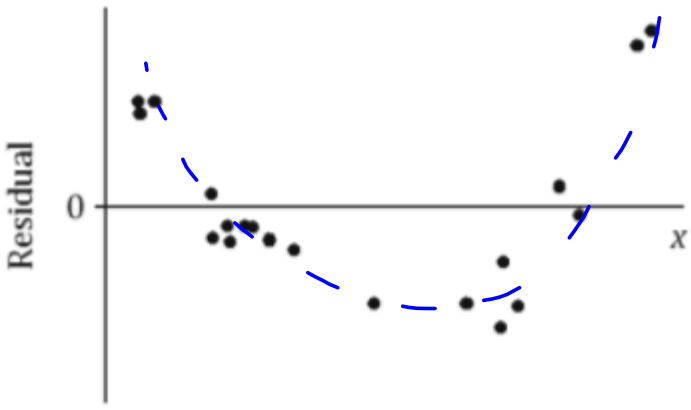
Non linear. The residuals make the shape of a quadratic.

b.



Linear because the residual plot looks completely scattered.

c.



Non linear because the residual points outline a parabola that opens up.

Consider again a data set giving the shoe lengths and heights of 10 adult men. This data set is shown in the table below.

The equation for the least squares line is $y = 3.6570x + 25.2769$.

x (Shoe Length) inches	y (Height) inches	Predicted Height	Residual Value
12.6	74	71.35	2.65
11.8	65	68.43	-3.43
12.2	71	69.89	1.11
11.6	67	67.70	- 0.70
12.2	69	69.89	- 0.89
11.4	68	66.97	1.03
12.8	70	72.09	- 2.09
12.2	69	69.89	- 0.89
12.6	72	71.35	0.65
11.8	71	68.43	2.57

1. Complete the table of values above.

All of the values were found by substituting the shoe length into our equation using a calculator

2. Graph those residual values on a residual plot. What kind of shape does the graph make? Does it indicate a linear or a nonlinear relationship?

Since the residual plot shows no pattern (or no correlation) then the scatter plot that it belongs to has a LINEAR RELATIONSHIP.