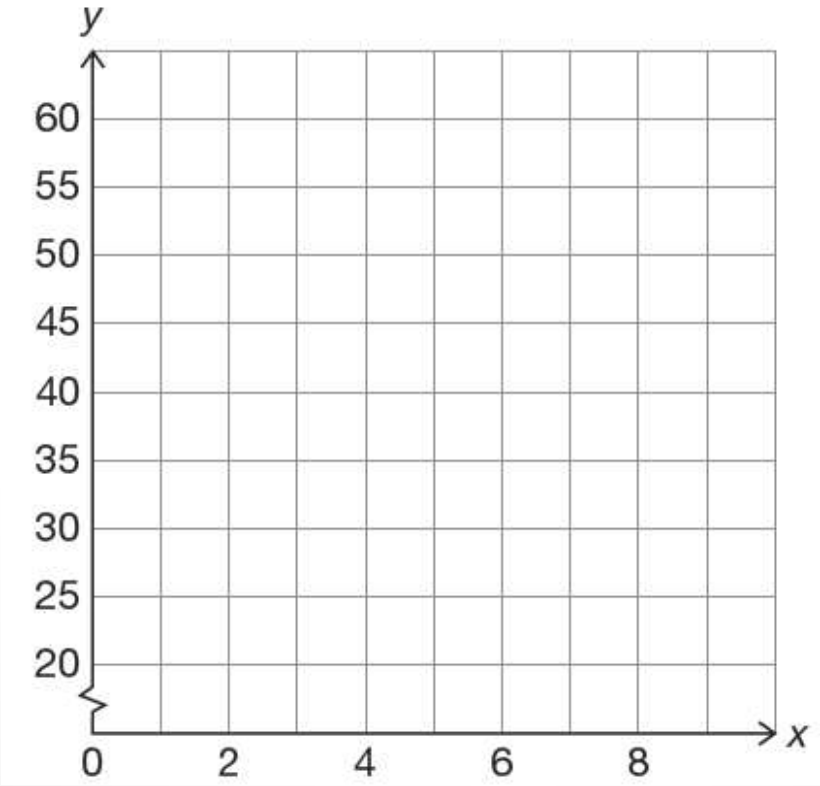
Get out the Notes from Monday

Feb. 4th, 2015

Example 2: Consider the table below displaying the percentage of recorded music sales coming from music stores from 1998 to 2004.

Year	1998	1999	2000	2001	2002	2003	2004
Percent of Total Sales from Music Stores	50.8	44.5	42.4	39.7	36.8	33.2	32.5

Let's create a scatter plot for this data,



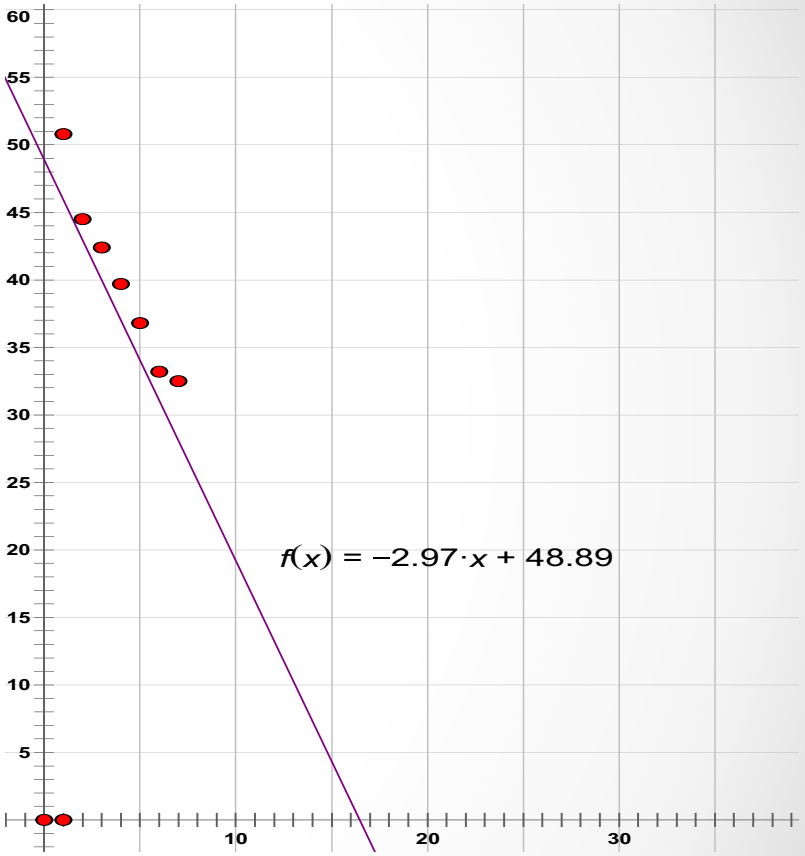
For our scatter plot in example 2, the linear function **$y = -2.97x + 48.89$** represents the linear function that best models our data. The graph below shows the scatter plot and function plotted together.

What does this equation mean in context of this problem?

What is the slope of the line? What does it mean in the context of the problem?

What is the y-intercept? What does it mean in the context of the problem?

What percentage of total sales came from music stores in 2008?

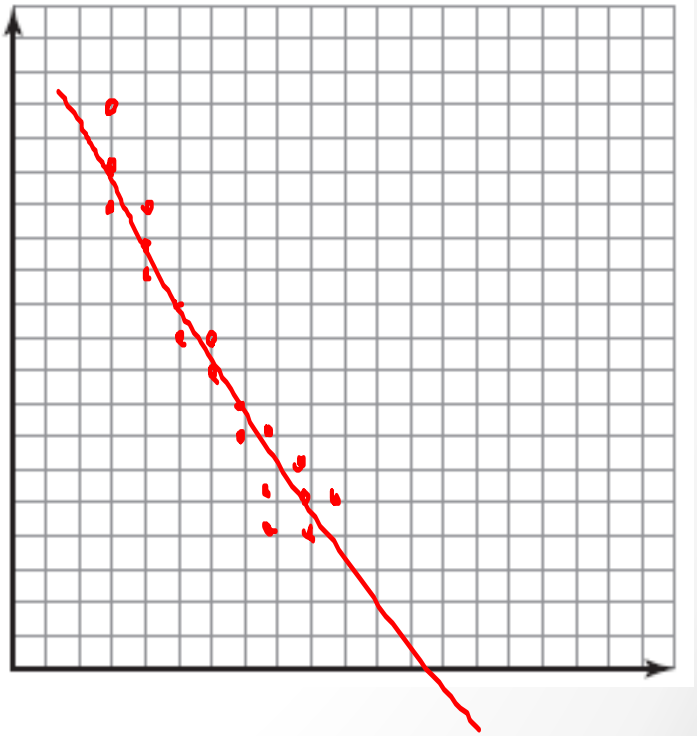


Example 3: The amount of antibiotic that remains in your body over a period of time varies from one drug to the next. The table given shows the amount of Antibiotic X that remains in your body over a period of two days.

Year	2005	2006	2007	2008	2009	2010	2011
Population (Thousands)	9.4	8.3	8.9	8.0	6.9	6.3	6.6

a). Create a scatter plot for the data.

b) Create the line of best fit for the data using the instructions from example 1.



c) The exact line of best fit is

$$y = -0.514x + 9.314$$

Use this function to determine the population in 2015.

d) What year will the population be below 3,000?

$$y = -0.514x + 9.314$$

e) What does the number -0.514 represent? What does it mean in the context of the problem?

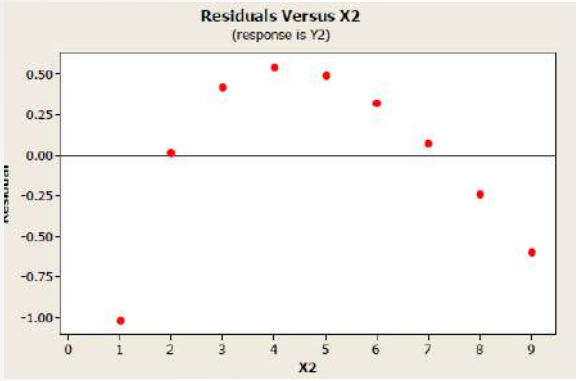
f) What does the number 9.314 represent? What does it mean in the context of the problem?

Residual (Error) – the _____exact_____ distance between a data point and the graph of the __ _____Line of Best Fit_____.
the difference between the observed value of y and the predicted value of y.
Your predicted values come from the equation for the line of best fit.

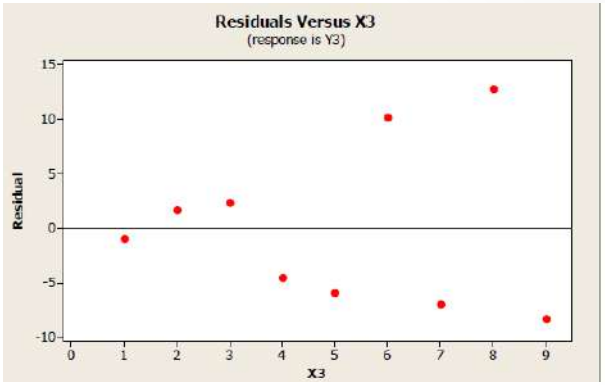
EXAMPLES:

Determine if the linear regression model is appropriate by looking at the graph of the residuals.

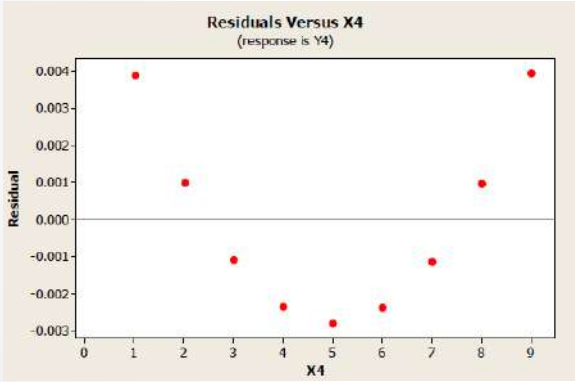
4.



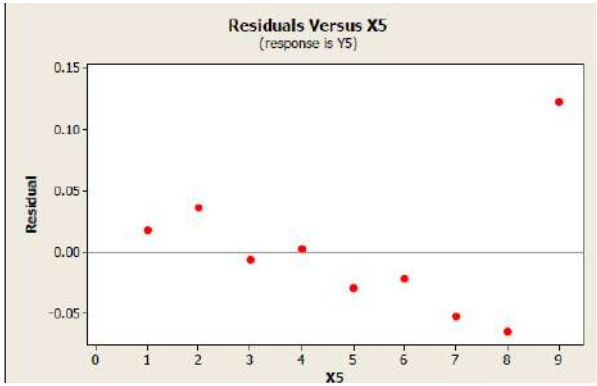
no, b/c its a pattern



no



no



yes

Step 1: Use the regression model to find the predicted values. (3rd Column)

$$y = 1.73x - 0.96$$

plug in x

X

observed y

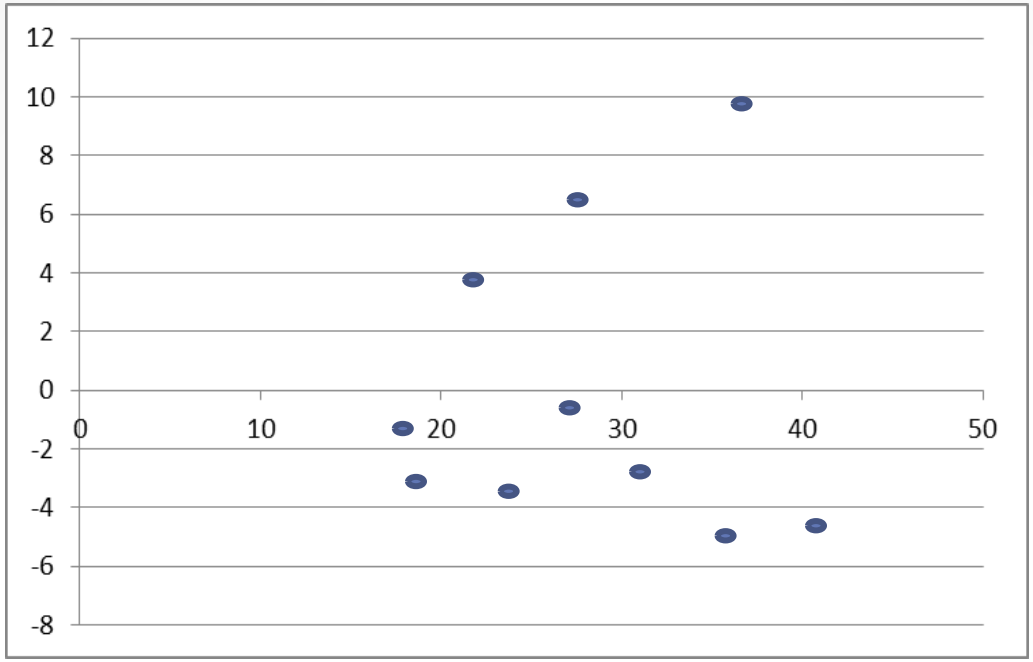
Total Time (minutes)	Total Distance (miles)	Predicted Total Distance	Residuals (observed – predicted)
32	51	54.4	-3.4
19	30	31.9	-1.9
28	47	47.5	-0.5
36	56	61.3	-5.3
17	27	28.5	-1.5
23	35	38.8	-3.8
41	65	70	-5
22	41	37.1	3.9
37	73	63.1	9.9
28	54	47.5	6.5

predicted y

This column - this column = last column

Step 2: Find the residuals: 2nd Column – 3rd Column = 4th Column

Step 3: Make a residual plot (model): Graph the points (1st Column, 4th Column).



A **residual plot** is another way to help us determine if a linear relationship exists between our variables. A residual plot is a scatter plot of the independent variable (y) and the residuals. For residual plots, the independent variable goes on the x-axis, and the residuals go on the y-axis.

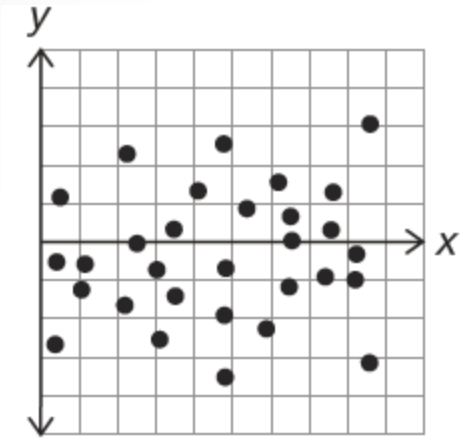
Step 4: Determine if the regression model is appropriate.

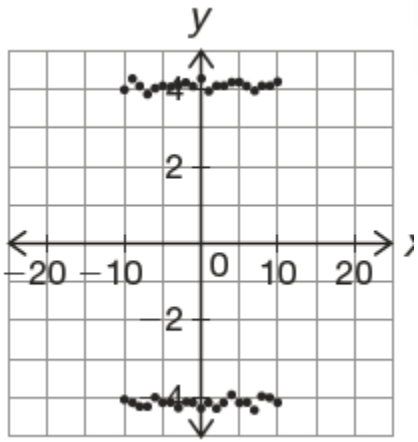
No pattern - yes

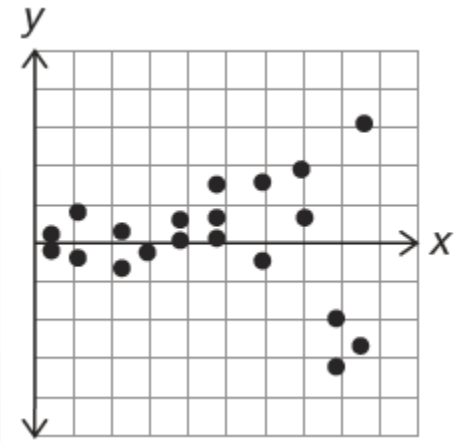
Discussion:

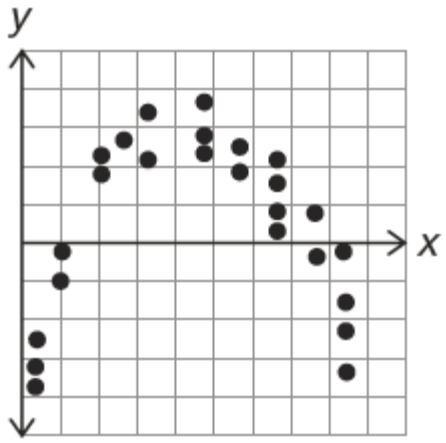
- What is this residual plot telling us about the relationship between speed and braking distance?

Let's now end by discussing how to interpret the residual plots below…









Discussion:

1. Can we come up with any general rules in regards to interpreting residual plots?