Residuals

- A **residual** value is the difference between an actual observed y value and the corresponding predicted y value, $\hat{y}$.

- Residuals are just errors.

- Residual = error = (observed – predicted) = $(y - \hat{y})$

1. A least-squares regression line was fitted to the weights (in pounds) versus age (in months) of a group of many young children. The equation of the line is $\hat{y} = 16.6 + 0.65t$, where $\hat{y}$ is the predicted weight and t is the age of the child. A 20-month old child in this group has an actual weight of 25 pounds. What is the residual weight, in pounds, for this child?
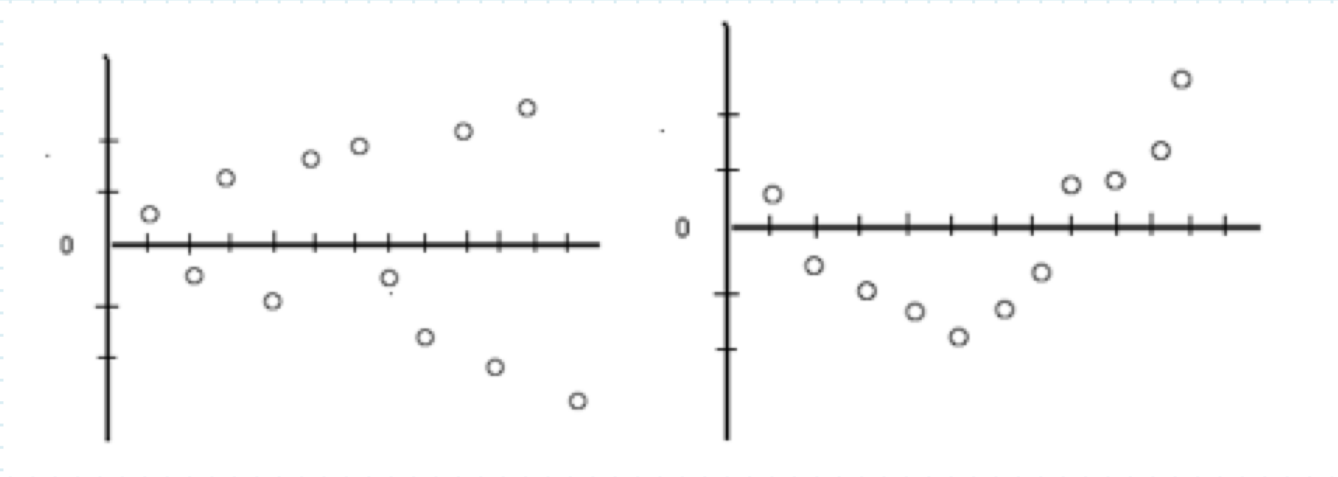
$$y(20) = 25 \qquad \hat{y}(20) = 16.6 + .65(20) = 29.6$$

$$\text{residual} = 25 - 29.6 = -4.6$$

- The plot of the residual values against the x values can tell us a lot about our LSRL model.
- Plots of residuals may display patterns that would give some idea about the appropriateness of the model.
- If the functional form of the regression model is incorrect, the residual plots constructed by using the model will often display a pattern.
- The pattern can then be used to propose a more appropriate model.
- When a residual plot shows no pattern, it indicates that the proposed model is a reasonable fit to a set of data.
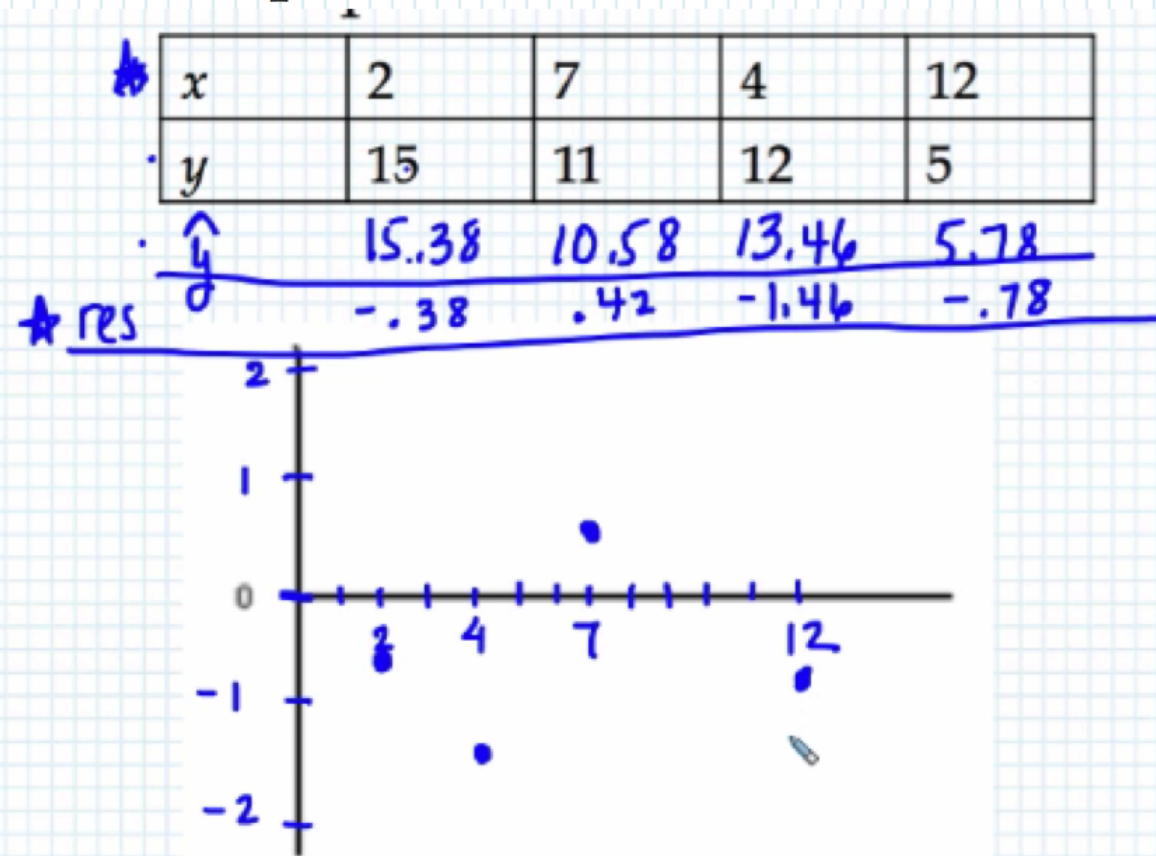
- Here are some examples of residual plots that show patterns:

Example:

2. A data set produced the regression equation $\hat{y} = 17.3 - 0.96x$
Below are four of the data points. Draw a residual plot for
these data points.

x	2	7	4	12
y	15	11	12	5

$\hat{y}$ 15..38   10.58   13.46   5.78

res $-.38$ .42 -1.46 $-.78$

Example:

3. The following data was collected comparing score on a measure of test anxiety and exam score.

Measure of test anxiety	23	14	14	0	7	20	20	15	21
Exam score	43	59	48	77	50	52	46	51	51

a. Construct a scatterplot.
b. Find the LSRL and fit it to the scatter plot. $\hat{y} = 68.838 - 1.0638x$
c. Find r and r^2. $r = .7877$ $r^2 = .6205$
d. Does there appear to be a linear relationship between the two variables? Based on what you found, would you characterize the relationship as positive or <u>negative</u>? Strong or weak?
e. What does the residual plot reveal?
f. Is it reasonable to conclude that anxiety <u>caused</u> poor exam performance?

no!
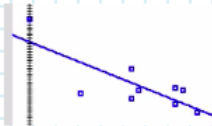
- Since the residuals show how far the data falls from the LSRL, examining the values of the residuals will help us to gauge how well the LSRL describes the data. The sum of the residuals is always 0 so the plot will always be centered around the x-axis.

- An **outlier** is a value that is well separated from the rest of the data set. An outlier will have a large absolute residual value.

- An observation that causes the values of the slope and the intercept in the line of best fit to be considerably different from what they would be if the observation were removed from the data set is said to be **influential.**

Example:

4. Johnny keeps track of his best swimming times for the 50 meter freestyle from each summer swim team season. Here is his data:

Age(years)	9	10	11	12	13	14	15	16
Time (sec)	34.8	34.2	32.9	29.1	28.4	22.4	25.2	24.9

influential

Investigate the LSRL model.

(see video for calculator steps)