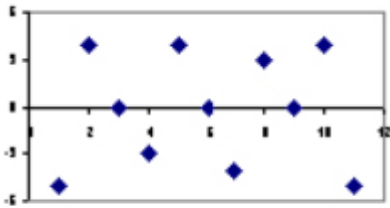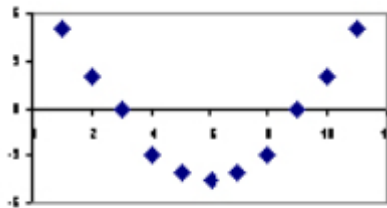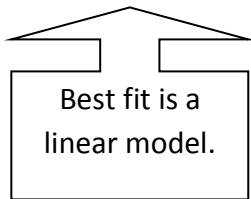# Unit 8 Investigation   Residual Plots

- A residual plot is a graph that shows the difference between the actual data (what is provided through a table or graph) and the predicted data (what the model says should happen).
- The independent variable is graphed on the horizontal axis and the residual value (actual – predicted) is graphed on the vertical axis.
- If the residual plots are randomly scattered around the horizontal axis, a linear model is the best choice to model the data.
- If the residual plot shows a pattern, and does not appear to random or scattered, a non-linear model would most likely be a better fit.
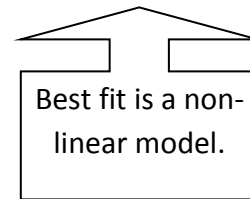
Random pattern	Non-random: U-shaped curve	Non-random: Inverted U
Best fit is a linear model.	Best fit is a non-linear model.	Best fit is a non-linear model.
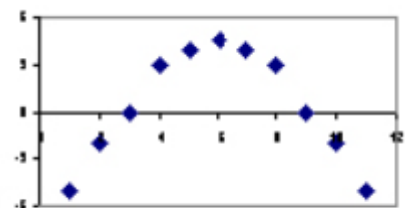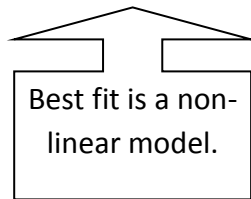
Let's start with some data!

The data below shows the number of active woodpecker clusters in the DeSoto National Forest.

Year	1992	1993	1994	1995	1996	1997	1998	1999	2000
Active Clusters	22	24	27	27	34	40	42	45	51

1. Enter the data into your calculator and find the line of best fit. Let x represent the number of years after 1992.

2. What is the correlation coefficient? Describe the goodness-of-fit.

3. Using your equation from question 1, find the predicted number of active woodpecker clusters. You can do this using the TABLE feature in your graphing calculator

Year	1992	1993	1994	1995	1996	1997	1998	1999	2000
Predicted Active Clusters									

4. To find the residual plots, we need to find the difference between what actually happened (original table) and what is predicted to happen (table from #3).

Year	1992	1993	1994	1995	1996	1997	1998	1999	2000
Residual Value									

5. Now let's construct a residual plot. On the horizontal axis will be our independent variable. On the vertical axis will the residual value.

6. Would you describe the residual plot as scattered and random or do you see a pattern? Do you think a linear model is best?

More data!

Below is population data for Jamestown, Virginia.

Year	2002	2004	2005	2007	2009
Population	5564	6121	6300	6812	7422

1. Find the line of best fit. Let x represent the number of years after 2000.

2. Recreate the table using the model from question 1 to find the predicted population.

Year	2002	2004	2005	2007	2009
Predicted Population					

3. Find the difference between actual and predicted and make a residual plot.

Year	2002	2004	2005	2007	2009
Residual Value					

4. Construct a residual plot.

5. Do you think a linear model is best? Why or why not?