Algebra 2
Lesson Plan on Fran Data

Goals:

- Describe graphically, algebraically and verbally real-world phenomena as functions; identify the independent and the dependent variable (3.01)
- Translate among graphic, algebraic, and verbal representations of relations (3.02)
- Graph relations and functions and find the zeros of functions (3.03)
- Write and interpret an equation of a curve (linear) which models a set of data (4.01)
- Find the equation of best fit (linear) for a set of data. Interpret the constants, coefficients, and based in the context of the data. Check the equation for goodness of fit and use equation for prediction (4.02)

Materials and Equipment Needed:

- Copy of handout for each student
- Graphing calculator
- Graph paper for each student

Activity One: Fit linear data with linear regression model, interpret meaning of slope and y-intercept, use linear function to predict outside given data. (Data is taken from Algebra 2 Indicators)

Customers without electricity after Fran data.

- Look at http://www.wral.com/hurricanes/ and find the Storm Tracker button. This will allow us to watch the path of Fran, which hit North Carolina on September 5, 1996 (5 years ago).
- Using data from indicator 4.02 C, students enter data in calculators. In list 1 (L1) put day of September and in list 2 (L2) put customers without power.
- Look at a scatter plot of the data. Discuss window settings. Talk about domain and range. Talk about ZoomStat on the calculator.
- Should we define the independent variable to be days since September 6 to give meaning to the $y-$intercept? Should this data be fit with a line?
- Discuss how to take **notes** on this process. Sketch the scatter plot on paper while looking at the graph on the calculator. Label axes—transfer window settings to sketch.
- If the independent variable is days since September 6, the linear regression model is $y = -151978.6x + 940000$. (How many decimal values do we keep?) Write equation in **notes** using the variable names that describe the data: $customers\ without\ electricity = -151978.6(days\ from\ Sept6) + 940000$. Sketch plot of line and data in **notes**.
- What is the meaning of the slope (including the expectation that it is negative) and the y-intercept? The slope represents the number of customers who get electricity fixed each day. The y-intercept shows the number of people without electricity on day 0, which is September 6. Write complete sentences in **notes** to describe the slope and the y-intercept.
- What is the residual associated with September 10? Find **Y(4).** This value of this residual is -136885.71. It is negative because the point is below the line. Write in **notes** the meaning of residual, the formula for finding the residual (for a given value of x from the data, the residual is $y_{data} - y_{line}$),

- What does this model forecast about September 16? Find **Y(10)**. This produces the value –579786. What does that mean? Reinforce meaning of independent and dependent variables. In **notes**, write about prediction equation and what the equation predicts for Sept. 16.
- Should we fit the data with a line? That is a major question. Introduce the idea of **goodness of fit**. Looking at the data there is a pattern in the **residuals** which implies we have not totally removed the message in the data with our fit. Write in **notes** about goodness of fit.
- What is the LinReg button on my calculator doing? Discussion of what least squares criteria means. Look at web site: http://www.keypress.com/sketchpad/java_gsp/squares.html that has an active illustration of to describe the procedure.

Follow-Up Activity: Percent of Food Expenditures Spent Away from Home Over Time.

- When students are using the calculator for graphs of scatter plots or least squares line, they should make sketches of what they see on the calculator on their paper.
- Be sure students first adjust data so that the independent variable is years since 1935.
- Each teacher needs to decide whether this should be done individually or with partners.

1. To make the y-intercept meaningful, I would shift the data back toward the origin by subtracting 1935 from each x-value. Here is the meaning of my variables:

   x = years since 1935

   y = % of food expenditures away from home

   The Linear Regression model (or best-fit line) is $y = 0.4324x + 11.9890$.

2. The slope of the model is approximately 0.4324. The slope is defined as

$$\frac{\Delta y}{\Delta x} = \frac{\Delta \text{\% of food expenditures away from home}}{\Delta \text{year since 1935}} = \frac{0.4324}{1}$$

   The slope says that for every year that passes, the percentage of food expenditures away from home increases by approximately 0.4324%.

   The y-intercept of the model is approximately 11.9890. The y-intercept is defined as the y-value when x is zero. In terms of the context of the problem, the y-intercept is the percentage of food expenditures away from home when the year is 1935. So, this y-intercept implies that in 1935 the percentage of food expenditures away from home was approximately 12%.

3. Remember I have shifted the data back toward the origin by subtracting 1935 from the x-value of each data point. Therefore, 1965 corresponds to $x = 1965 - 1935 = 30$. The data point is $(30, 22.8)$. According to our model, in 1965, or $x = 30$, the percentage of food expenditures away from home is approximately $y = 0.4324 * 30 + 11.9890 = 24.96\%$. The residual is the difference in these y-values: $|22.8 - 24.96| = 2.16$. Since the data point is below the line, the residual is negative. The residual value is -2.16. This value tells me that in 1965 the actual percentage of food expenditures away from home was approximately 2.16% less than my linear regression model would predict.

4. The residuals imply our model is a good fit. They are randomly scattered and fairly evenly distributed above and below the line. Their size is small relative to the y-values of the data set. The residuals range in value roughly from –4 to 4 and the data values are between 12.9 and 38.2

5. Again remembering I have shifted the data back toward the origin, the year 2010 corresponds to $x = 75$. The linear function implies that for $x = 75$, $y = 0.4324 * 75 + 11.9890 = 44.419$. This means that in 2010, approximately 44.419% of food expenditures will be away from home. This percentage is a little less than half, which is probably reasonable – maybe not too healthy, but reasonable in the sense of people eating more and more fast food.

Student Handout
Fran Data Problem
Algebra 2

When Hurricane Fran hit North Carolina on the evening of September 5, 1996, over one million homes and businesses were left without power. Repair crews began immediately restoring electrical service.

Date	Customers Without Power
Sept. 6	1,159,000
Sept. 7	804,000
Sept. 8	515,000
Sept. 9	340,500
Sept. 10	195,200
Sept. 11	136,300
Sept. 12	77,000
Sept. 13	37,600

This data is taken from the Algebra 2 Indicators developed by NC Department of Public Instruction.

1. Here is a sketch of the scatter plot of the data. Label axes with scale and with titles of meaning of the variables.

2. What is the linear regression model that fits this data? Write the equation using variable names appropriate for the data set.

3. Sketch the scatter plot with the line superimposed over the data. Label scale on axes.

4. What is the meaning of the slope and of the y-intercept in terms of the phenomena? Write in complete sentences.

5. What is a residual?

6. What is the residual associated with September 10 in this data set?

7. What does the prediction equation forecast for September 16?

8. Discuss the goodness of fit of this line for this data. (Or how good does this line fit the data?)

Follow-Up Activity
Fran Data
Algebra 2

Percent of Food Expenditures Away from Home

1935	12.9	1970	26.3
1940	15.2	1975	28.5
1945	19.6	1980	32.2
1950	17.8	1985	35.8
1955	18.6	1990	36.7
1960	19.9	1995	38.2
1965	22.8		

1. Using your calculator, do a scatter plot of the data above. What is the Linear Regression (Best fit line) for the data? What does each variable represent?

2. What is the meaning of the slope? What is the meaning of the *y*-intercept?

3. What is the residual associated with the year 1965? What does the value of the residual tell you?

4. Thinking about the residuals, discuss the goodness of fit of this line?

5. Using this linear function as an equation of prediction, find the percent of food expenditures spent away from home in the year 2010. What does this number mean? Do you think this is reasonable?