Chapter 12
Multiple Regression and Model Building

Model Building: Quadratic and Other Higher Order Models

Example 12.7 Analyzing a Quadratic Model: Predicting Electrical Usage

Problem: In all-electric homes, the amount of electricity expended is of interest to consumers, builders and groups involved with energy conservation. Suppose we wish to investigate the monthly electric usage y in all-electric homes. Moreover, suppose we think that monthly electrical usage in all-electric homes is related to the size of the home by a quadratic equation. To fit the model, the values of y and x are collected for 15 homes during a particular month. The data are shown in Table 12 – 1.

Table 12 – 1
Electrical Usage

x = Size of Home (sq. ft.)	y = monthly usage (in kilowatt hours)
1290	1182
1350	1172
1470	1264
1600	1493
1710	1571
1840	1711
1980	1804
2230	1840
2400	1956
2710	2007
2930	1984
3000	1960
3210	2001
3240	1928
3520	1945

Solution: We begin this problem by entering the data into the calculator. After which we will run the Quadratic regression.

1. Enter size of home (x) into List 1 and Monthly-usage (y) into List 2. See Figure 12 – 1. Remember: *For all data given in a frequency table or listed as two or more variables it is necessary to enter the data in the EXACT order it is given.*

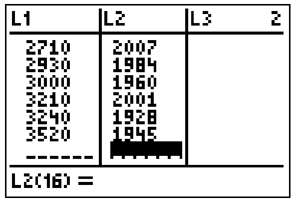

Figure 12 – 1

Note: It is good to glance at the two lists to see if they are the same length. If they are not the same length you probably missed a piece of data in the shorter list. Many mistakes occur at data entry. Make sure you have entered the data correctly. If you make a mistake here nothing else will be correct!

After you have entered all the data you will want to see a scatter plot of the data.

2. Before attempting to graph any statistical plot, be sure that your function plot screen is clear, or the plots are turned off. Enter the stat plot screen by pressing 2nd STATPLOT. Your screen will appear as in Figure 12 – 2.

3. Enter the set-up screen for plot 1 by pressing ENTER. Your screen should appear similar to Figure 12 – 3.

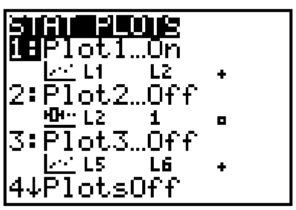

Figure 12 – 2

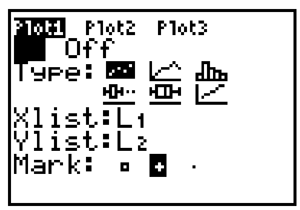

Figure 12 – 3

4. Turn Plot 1 on by highlighting On and pressing ENTER. See Figure 12 – 3.

5. Select the scatterplot by highlighting the scatterplot icon (the first one in the Type list) and pressing ENTER. See Figure 12 – 3.

6. Arrow down to Xlist and press 2nd L1 for List 1. Arrow down to Ylist and press 2nd L2 for List 2. See Figure 12 – 3.

7. Select the mark you would like to represent the points on the graph. Press ENTER. See Figure 12 – 3 for your final screen.

8. Press ZOOM 9 for ZoomStat and the graph will appear. See Figure 12 – 4. You may choose to adjust your window for a different view. Press WINDOW and change your values. Then press GRAPH.

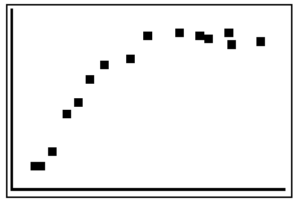

Figure 12 - 4

As can be seen from the screen in Figure 12 – 4 there is a noticeable curve. A linear regression would not be appropriate. However a quadratic regression may be appropriate.

To run the Quadratic regression

1. Press [STAT] and arrow over to the CALC menu.

2. Press [8] or arrow down to 5: QuadReg. See Figure 12 – 5. Press [ENTER]. **QuadReg** will appear at the top of your home screen. See Figure 12 – 6.

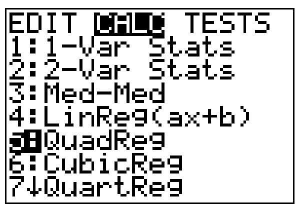

Figure 12 – 5

Figure 12 – 6

3. The format for this command is QuadReg xlist, ylist, Y-vars where the regression equation should be stored. Here we are using Lists 1 and 2 so press [2nd] L1 [,] [2nd] L2 for the lists. To get the Y where the equation will be stored, press [,] [VARS] arrow to Y-vars, press [ENTER] and then [ENTER] again to choose Y_1. Make sure your Y1 is clear of any old equations. See Figure 12 – 7.

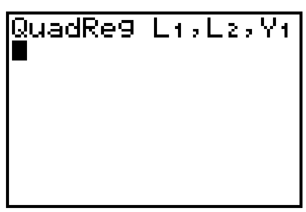

Figure 12 – 7

4. Press [ENTER] and your screen should look like Figure 12 – 8.

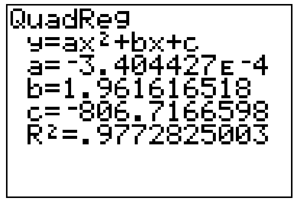

Figure 12 – 8

As can be seen from the screen in Figure 12 – 8, the Quadratic equation that best describes the data is: $Y = -0.00034x^2 + 1.9616x - 806.7167$

To see the graph of the regression equation along with the data, press GRAPH. See Figure 12 – 9

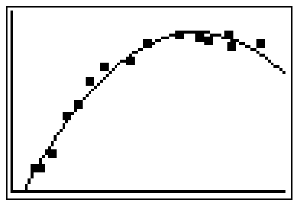

Figure 12 – 9

In Figure 12 – 8 we see that $R^2 = 0.9773$. Thus, we expect the close fit we see on the screen in Figure 12 – 9.

Residual Analysis: Checking the Regression Assumptions

Example 12.15 Using Residuals to Check Equal Variances

Problem: Consider the data on salaries, y, and years of experience, x, for the sample of 50 social workers. A first order model was fitted to the data. Construct a plot of the residuals vs. $\hat{y}$ and interpret the results. Make model modifications as necessary.

For this problem we will begin by looking at a plot of the residuals and then proceed to the data transformation, and finally we will determine the Line of best fit for the transformed data.

Plot of Residuals (Original, Untransformed Data)

1. Enter all data into its appropriate list. See Figure 12 – 10.

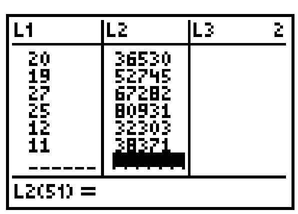

Figure 12 - 10

2. After you have entered all the data, press STAT and arrow over to the CALC menu.

3. Press [8] or arrow down to **8:LinReg(a + bx)**. See Figure 12 – 11. Press [ENTER] and **LinReg (a+bx)** will appear at the top of your home screen. See Figure 12 – 12.

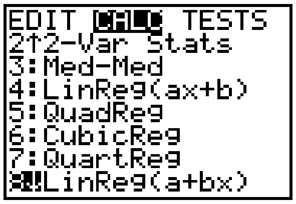

Figure 12 – 11

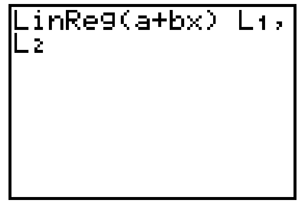

Figure 12 – 12

4. You must run the regression first in order to have the residuals to plot therefore press [ENTER] and your screen will look like Figure 12 – 13.

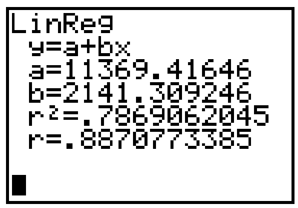

Figure 12 – 13

From Figure 12 – 13 we see that the R^2 value = 79% shows a moderately strong linear relationship between years of experience and salary.

To plot the residuals, go to the Stat Plot menu. Plot the residuals and the original x values using the scatter plot option.

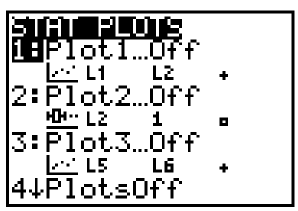

Figure 12 – 14

We assume that age is in List 1 and that the residuals are stored in the list RESID. (To accomplish this, run a linear regression on the years of experience and salaries and the residuals will be stored there.)

First we will plot the residuals against years of experience. We will use the scatterplot to do this. Be sure to clear the Y= screen or turn off any equations located there. Also, be sure all stat plots are turned off except Plot 1.

1. Press [2nd] STATPLOT [ENTER] to enter the Stat Plot setup screen.

2. Highlight On and press [ENTER].

3. Select the scatterplot by highlighting the icon and pressing [ENTER].

4. At the Xlist prompt press [2nd] L1 for List 1 (years of experience).

5. At the Ylist prompt press [2nd] LIST, arrow down to RESID and press [ENTER]. See Figure 12 – 15.

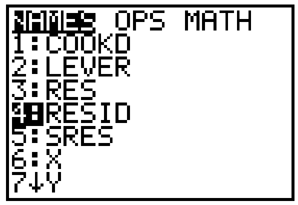

Figure 12 – 15

6. Choose your mark and press [ENTER]. Here the box is chosen. Your screen should appear as in Figure 12 – 16.

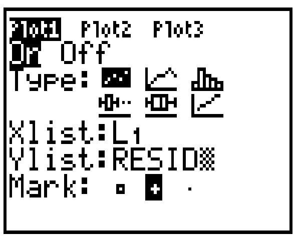

Figure 12 – 16

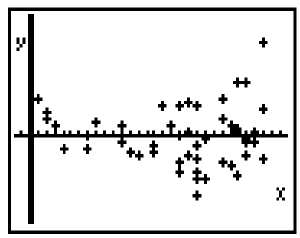

Figure 12 – 17

7. Use ZoomStat to set the window by pressing [ZOOM] [9]. Your screen should appear as in Figure 12 – 17.

The scatterplot of the residuals against $\hat{y}$ in Figure 12 – 17 shows a potential problem. The size of the residuals increases as the estimated mean salary increases, implying the constant variance assumption is violated.

Variance-Stabilizing Transformation of Data

We now want to transform the data to try to stabilize the variance.

1. Press [STAT] and then [ENTER]

2. Highlight the List 3, L3, all the way at the top of the column. See Figure 12 – 18.

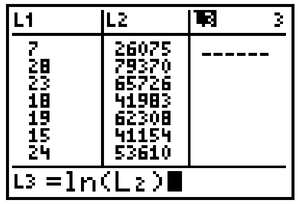

Figure 12 – 18

3. Press [LN] [2nd] L2. See Figure 12 – 18.

4. Press [ENTER]. The transformed data will appear in the List L3. See Figure 12 – 19.

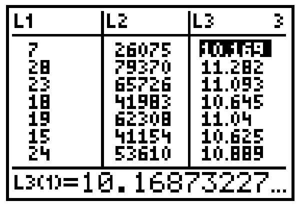

Figure 12 - 19

Now we want to plot the residuals using the newly created list 3.

5. Press [STAT] and arrow over to the CALC menu. Run the regression using L1 and L3. See Figure 12 – 20.

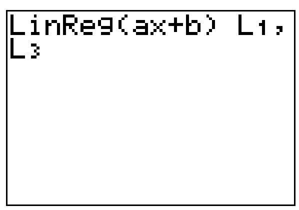

Figure 12 – 20

6. The previous residuals have been replaced with the residuals formed by using the newly transformed data. See Figure 12 – 21.

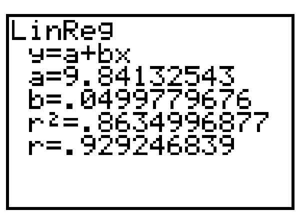

Figure 12 – 21

Note: The R^2 value has increased from 78.7% for the original data to 86.3% for the transformed data. This indicates a stronger relationship.

7. Now create the new plot of the residuals. Use ZoomStat to set the window by pressing [ZOOM] [9]. See Figure 12 – 22.

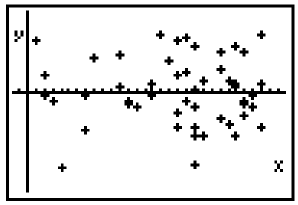

Figure 12 – 22

As you can see from the screen in Figure 12 – 22, the new residuals fall roughly in a horizontal band centered and symmetric about the x-axis. Thus, the transformation stabilized the error variances.