Linear and Quadratic Least Squares

Prepared by

Stephanie Quintal, graduate student

Dept. of Mathematical Sciences, UMass Lowell

in collaboration with

Marvin Stick

Dept. of Mathematical Sciences, UMass Lowell

Summer 2015

Preface

The following least squares material shows the development of linear and quadratic least squares models. Examples are selected with TI-84 tables and graphs to exhibit the agreement between actual and modeled data.

The mathematical derivation for both linear and quadratic models is included so that the reader will have a general idea of the complexity behind the modeling procedures. Fortunately, algorithms have been developed and are easily accessible with a TI-84 calculator or Excel to generate these models.

Stephanie Quintal and Marvin Stick

Linear and Quadratic Least Squares

Although other models will be introduced later, let's consider linear and quadratic functions. When given real-world applications, we are often left to decide which type of model, if any, fits the data best. The question is often how to get a model from the given data points and how we can use the developed model to predict future outcomes.

The simple approach to curve fitting is to examine the scatterplot of data. In a scatterplot, we can look for a general pattern. For example, in a linear function with positive slope we look for an ascending line. In a linear function with negative slope, we look for a descending line. Finally, when looking for a quadratic function, we look for a data pattern that resembles a curve or parabola.

Let's look at a simple example involving the following data points $(-2, 1), (1, 2)$ and $(6, 3)$. We can find an equation of a line that passes through these points. To begin, we find the slope of the line containing two of the points. Let's choose the points $(-2, 1)$ and $(6, 3)$. Recall that the slope of the line containing points (x_1, y_1) and (x_2, y_2) is $m = \dfrac{y_2 - y_1}{x_2 - x_1} = \dfrac{\text{change in } y}{\text{change in } x}$.

So, the slope of this line is $m = \dfrac{y_2 - y_1}{x_2 - x_1} = \dfrac{3-1}{6-(-2)} = \dfrac{2}{8} = \dfrac{1}{4}$. Now, we use point-slope form to find the equation of a line.

$$y - y_1 = m(x - x_1)$$

$$y - 3 = \frac{1}{4}(x - 6)$$

Finally, we distribute one-forth and simplify the equation to be in $y = mx + b$ form.

$$y - 3 = \frac{1}{4}(x - 6)$$

$$y - 3 = \frac{1}{4}x - \frac{6}{4}$$

$$y = \frac{1}{4}x + \frac{6}{4}$$

$$y = \frac{1}{4}x + \frac{3}{2}$$

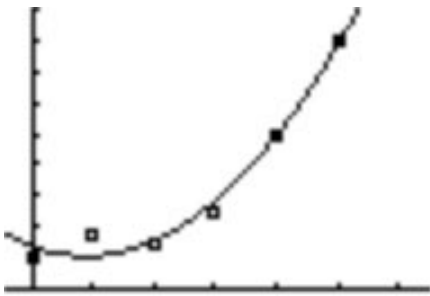The graph of this line and the initial data points are shown in figure 1.

| Figure 1 | Figure 1A |

Notice that the line does not pass through the point $(1,2)$ in figure 1. We will use linear regression to minimize this error. Later we will describe how to implement linear regression on your TI-84 calculator, but accurate to three decimal places. The R^2 value (a measure of the error) is $R^2 = 0.980$ and the graph of the actual data and linear model appear in figure 1A. While the model does not go through 2 points as in figure 1, the overall errors (sum of [actual data minus modeled result]²) is less than in figure 1.

      Next, we introduce linear regression, or the method of least squares with a bit more detail. This is the most common and preferred method to fitting a curve to a given data set. An advantage to the method of least squares is that this method uses all data points, as opposed to two data points when trying to find the equation of a straight line. As a result, we get a more accurate model for the data set. Let's look at an example.

| Figure 2 | Figure 3 |

      Using two decimal accuracy, the linear model (figure 2) for the data in figure 3 is $y = 0.99x + 1.45$. The R-squared, which is also produced by the calculator, is 0.91. In linear regression, the R-squared is a statistical measure for how accurate the model describes the given data. The range for R-squared is $0 \leq R^2 \leq 1$. An R-squared value close to one indicates a good-fitting model. An R-squared close to zero indicates that the model does not fit the data well. Statistically, R^2 represents the amount of variance in the model.

      Now let's look at an example of a quadratic model. Consider the example with the data in figure 5.

2

Figure 4

Figure 5

An excellent fit for the model shown in figure 4 is $y = 0.40x^2 - 0.72x + 1.38$ with R-squared equal to $0.98.$ Notice that the slight outliers (errors) occur at the points $(1,1.8)$ and $(3,2.4).$ The two outliers do not cause much error since the R-squared value is so close to one.

Your next question might be "How do we generate the models?". The method of least squares requires that we minimize the sum of the squares of the errors between the actual data and the predicted model results. Although the derivation of least squares can get complicated, it is important to have a basic understanding of where things are coming from. Let's look at the least squares derivation.

Derivation of Linear Least Squares Regression Model

To begin, let's take the difference of the estimate, or $I_n = \sum_{i=1}^{n}\left[y_i - (ax_i + b)\right]^2.$

Next, let's find the partial derivative of $I_n,$ with respect to a. This is done by treating b, x_i and y_i as constants. To find this particular derivative, we use the chain rule.

$$\frac{\partial I_n}{\partial a} = 2\sum_{i=1}^{n}\left[y_i - (ax_i + b)\right]^1(-x_i)$$

Distributing and setting the equation equal to zero, we get: $2\sum_{i=1}^{n}\left(-x_iy_i + ax_i^2 + bx_i\right) = 0.$

We set the partial derivative equal to zero to minimize the sum of the errors squared, and will follow a similar procedure when taking other partial derivatives to find values of least squares coefficients.

Next, let's divide by 2 and distribute the summation sign.

$$\sum_{i=1}^{n}-x_iy_i + \sum_{i=1}^{n}ax_i^2 + \sum_{i=1}^{n}bx_i = 0.$$

Let's bring the negative term to the right side of the equation.

3

$$\sum_{i=1}^{n} ax_i^2 + \sum_{i=1}^{n} bx_i = \sum_{i=1}^{n} x_i y_i$$

Finally, we can pull the constants out of the summation symbol to simplify.

$$a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i$$

Next, let's find the partial derivative of I_n, with respect to b. This is done by treating a, x_i and y_i as constants. Recall that we use the chain rule again.

$$I_n = \sum_{i=1}^{n} \left[y_i - \left(ax_i + b \right) \right]^2$$

$$\frac{\partial I_n}{\partial b} = 2\sum_{i=1}^{n} \left[y_i - \left(ax_i + b \right) \right]^1 (-1)$$

$$2\left(\sum_{i=1}^{n} -y_i + \sum_{i=1}^{n} ax_i + \sum_{i=1}^{n} b \right) = 0$$

$$\sum_{i=1}^{n} -y_i + \sum_{i=1}^{n} ax_i + \sum_{i=1}^{n} b = 0$$

$$a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} (1) = \sum_{i=1}^{n} y_i$$

We know that $b\sum_{i=1}^{n} 1 = bn$, so we get the following equation:

$$a\sum_{i=1}^{n} x_i + bn = \sum_{i=1}^{n} y_i$$

Now, we are left with a system of two equations and two unknowns:

$$a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i$$

$$a\sum_{i=1}^{n} x_i + bn = \sum_{i=1}^{n} y_i$$

To solve the 2x2 system of equations, let's use Cramer's Rule. In the system above, the variables we will be solving for are a and b. We begin by solving for a.

$$a = \frac{\begin{vmatrix} \sum\limits_{i=1}^{n} x_i y_i & \sum\limits_{i=1}^{n} x_i \\ \sum\limits_{i=1}^{n} y_i & n \end{vmatrix}}{\begin{vmatrix} \sum\limits_{i=1}^{n} x_i^2 & \sum\limits_{i=1}^{n} x_i \\ \sum\limits_{i=1}^{n} x_i & n \end{vmatrix}} = \frac{n\sum\limits_{i=1}^{n} x_i y_i - \sum\limits_{i=1}^{n} x_i \sum\limits_{i=1}^{n} y_i}{n\sum\limits_{i=1}^{n} x_i^2 - \left(\sum\limits_{i=1}^{n} x_i\right)^2} \qquad \textbf{(Equation 1)}$$

In order to use Cramer's Rule, we must find the determinant of a 2x2 matrix. In general, when finding the determinant of a 2x2 matrix, take the product of the top-left-bottom-right diagonal and subtract the product of the top-right-bottom-left diagonal.

Using the same method, let's solve for *b*.

$$b = \frac{\begin{vmatrix} \sum\limits_{i=1}^{n} x_i^2 & \sum\limits_{i=1}^{n} x_i y_i \\ \sum\limits_{i=1}^{n} x_i & \sum\limits_{i=1}^{n} y_i \end{vmatrix}}{\begin{vmatrix} \sum\limits_{i=1}^{n} x_i^2 & \sum\limits_{i=1}^{n} x_i \\ \sum\limits_{i=1}^{n} x_i & n \end{vmatrix}} = \frac{\sum\limits_{i=1}^{n} x_i^2 \sum\limits_{i=1}^{n} y_i - \sum\limits_{i=1}^{n} x_i \sum\limits_{i=1}^{n} x_i y_i}{n\sum\limits_{i=1}^{n} x_i^2 - \left(\sum\limits_{i=1}^{n} x_i\right)^2} \qquad \textbf{(Equation 2)}$$

Reader's Note:

An alternate form for the solution of *b* is $\quad b = \frac{1}{n}\left(\sum\limits_{i=1}^{n} y_i - a\sum\limits_{i=1}^{n} x_i\right).$ **(Equation 3)**

This can be verified by substituting the value of *a* in equation 1 into equation 3 and the result will yield equation 2. Utilizing equation 3 to do numerical data analysis is easier and more efficient than using equation 2.

Let's start with an example using your TI calculator. We will perform a linear least squares regression and generate a best-fit line in the form of $y = mx + b.$

Consider the following data points: $(1, 2), (2, 4), (3, 3) \text{ and } (4, 5).$
1. STAT, EDIT and enter the (x, y) observations into L1 and L2, respectively.

2. Put the cursor on L1 and press CLEAR in order to clear the list, if needed. Do likewise for L2. This may not be necessary if the lists are already cleared.
3. Put the cursor below L1 list and enter all x values. Note that the ↓ (down arrow) is used to move down the list of values. Next, click the → (right arrow) to move to the L2 column.
4. Enter the values for y in the L2 list.

Figure 6

5. STAT, CALC, option 4 (LinReg $(ax+b)$). If we use this approach, we have to type the resulting equation into y_1. In order to have the resulting equation appear in y_1 and be able to specify that the x data is stored in list1 and the y data is stored in list2, enter on your TI-84 the commands: STAT, CALC, option 4 $(\text{LinReg}(ax+b))$, L_1, L_2, VARS→Y-VARS→Option 1(function) → Option 1 (y_1). The ordering is important. Note that L_1 is the default for x data and L_2 is the default for y data and are assumed if omitted. However, using the approach just mentioned gives the user more flexibility regarding where the x and y data are stored.

6. The following will appear on your TI screen:

Figure 7

7. If we use the approach in step 5, the equation $y=0.8x+1.5$ is automatically entered in Y_1. The correlation coefficient, r, is a statistical measure of agreement between the x and y values. This particular model shows an r-value equal to 0.8. Examine that the R-squared is merely the r-value, squared. The r-value will only display for

linear regression models and has a range from negative one to one. A positive value for r implies that as x increases, y also increases (positive slope). A negative value for r implies that as x increases, y decreases (negative slope).

8. If your calculator displays the R-squared and r-values, skip this step. Note that the TI does not turn on the R^2 feature when the calculators are shipped. To turn it on, press 2nd CATALOG, D (automatically in alpha mode and allows quick access to the functions starting with the letter D), use ↓ (down arrow) to access DiagnosticsOn, and press enter twice. The screen should display done. You will not have to do this again.

9. To graph the linear model $y = 0.8x + 1.5$ along with the actual data points, get to the $Y =$ screen, highlight Plot1 with the cursor. Next, press ZOOM and scroll down to option 9 (ZoomStat). This option automatically sets the appropriate window and generates the picture of the model and actual data set. This appears in the figure below.
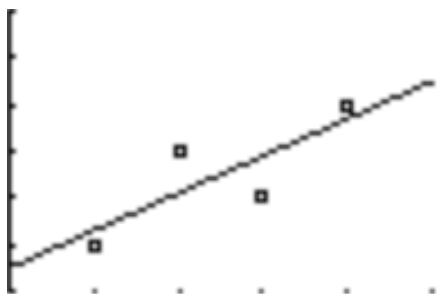
Figure 8

10. Be sure to turn off Plot1 when done, otherwise the actual data will continue to appear on all future graphs. Plot1 is merely a toggle switch, so highlight it and press the enter key to turn it off.

Let's consider a real-world application to further learn least squares linear regression. Consider an example involving studying time and test scores.

Example 1
The data in the following table relate study time and test scores. [Bittinger, CALCULUS and ITS APPLICATIONS, 10th ed, page 73]

Study Time (in hours)	Test Grade (in percent)
7	83
8	85
9	88
10	91
11	?

Table 1

First, let's find a regression line to fit the data. Then we'll graph the scatterplot of the data, along with the regression line. Lastly, we will use the model to predict the test score grade of a student that studies 11 hours. To fit a linear function to the data using regression, let's select the EDIT option of the STAT menu. Next, we enter the data.

Figure 9

Let's graph the data points. This is done by highlighting PLOT1 on the $y_1 =$ screen and pressing GRAPH. You may want to adjust your window, but be sure to make it large enough to see all the data points. Be sure to turn off PLOT1 when done.

Figure 10

To find the line that best fits the data, follow the procedure mentioned previously, i.e. STAT, CALC, option 4 ($\text{LinReg}(ax+b)$), L_1, L_2, VARS→Y-VARS→Option 1(function) → Option 1 (y_1).

Figure 11

8

The regression equation is $y = 2.7x + 63.8$. The graph appears in figure 12.
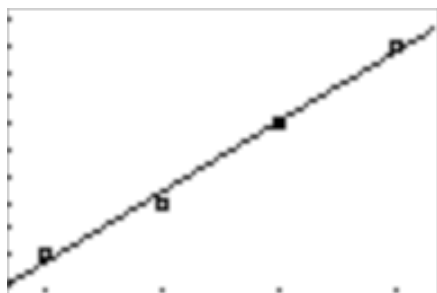
Figure 12

Next, we will predict the test grade for a student that studies 11 hours. On your TI, press 2nd, CALC, option 1: VALUE and enter 11. Note that you will have to adjust the window of your graph and make both the *x* max and *y* max values larger to accommodate the value $x = 11$ and the newly generated *y* value when $x = 11$. Of course, you can analytically solve for *y* when $x = 11$ by $y(11) = 2.7(11) + 63.8 = 93.5$.
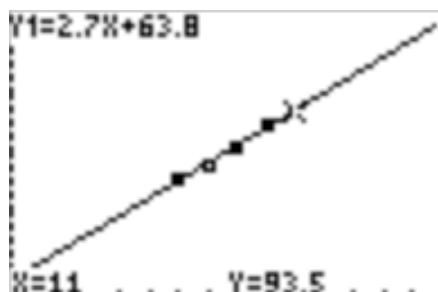
Figure 13

Example 2
Next, let's introduce an example similar to example 1, but slightly different. Consider the following example given years as *x* values and set $x = 0$ as the base value.

The following table shows the number of U.S. households with cable television. [Bittinger, CALCULUS and ITS APPLICATIONS, 10th ed, page 71]

Notice that the year 1999 is the base year and we can assign it the value of $x = 0$. Another and perhaps easier approach might be to assign the base year 1999 as $x = -1$ or $x = 9$. The choice is up to you.

Year, x	Number of Cable Subscribers (in millions), y
1999, 0	76.4
2000, 1	78.6
2001, 2	81.5
2002, 3	87.8
2003, 4	88.4
2004, 5	92.4
2005, 6	94.0
2006, 7	95.0

Table 2

Let's begin by finding an equation of a line using the equation for slope and point-slope form. We can use any two of the data points to determine an equation. To determine the slope, let's use the points $(2, 81.5)$ and $(6, 94.0)$.

$$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{94.0 - 81.5}{6 - 2} = \frac{12.5}{4} = 3.125$$

Then, we substitute 3.125 for m and either $(2, 81.5)$ or $(6, 94.0)$ for (x_1, y_1) into the point-slope equation.

$$y - y_1 = m(x - x_1)$$
$$y - 94.0 = 3.125(x - 6)$$
$$y - 94 = 3.125x - 18.75$$
$$y = 3.125x + 75.25$$

So, we have found that the line between the points $(2, 81.5)$ and $(6, 94.0)$ is $y = 3.125x + 75.25$, where x is the number of years passed 1999 and y is in millions. Let's graph the data points and the line. This is shown in figure 14.
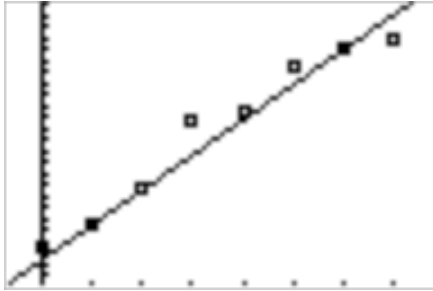
Figure 14

Notice the error in the line when using two data points.

Using the same process used in example 1, let's use the method of least squares to find a more accurate line of best fit. Let's begin by graphing the data in the table in a scatterplot.
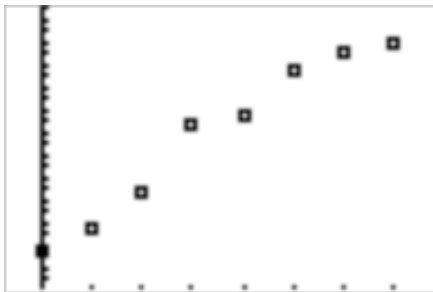

Figure 15

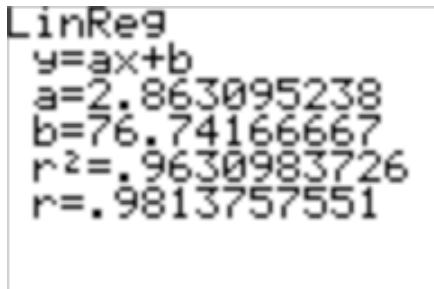Like before, using the CALC option in the STAT menu, find the line that best fits the data.

Figure 15

We find the line of best fit with rounding to 2 decimal places is $y = 2.86x + 76.74$. Notice that the r-squared value is high, i.e. pretty close to 1, which means the model provides a good fit. Although this line is close to the line that we found using two points, the least squares model provides a more accurate model for the data. Let's graph it along with the scatterplot. This is shown in figure 16.
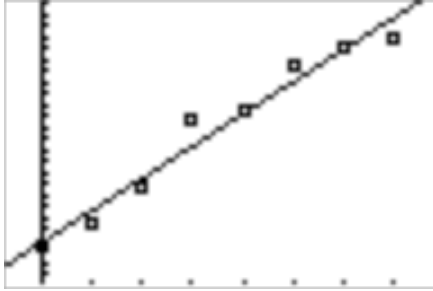
Figure 16

Example 3

Let's consider another example involving hours of sleep and death rate. In a study conducted by Dr. Harold J. Morowitz of Yale University, data were gathered to examine the relationship of hours of sleep and death rate. The following table shows the average hours of sleep per night and the death rate per 100,000 males. [Bittinger, CALCULUS and ITS APPLICATIONS, 10th ed, page 73]

Average number of hours of sleep, x	Death Rate per 100,000 males, y
5	1121
6	805
7	626
8	813
9	967

Table 3

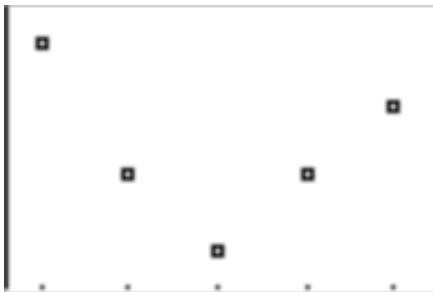Let's start by graphing the data in the scatterplot.

Figure 17

The data shown in figure 17 does not follow a general straight line. The data drops and rises, which resembles a parabola. This pattern indicates that instead of a linear function, we may need a quadratic function. Recall that a quadratic function has the form of

$y = ax^2 + bx + c$. So, to find the quadratic function that fits the scatterplot, let's use the points $(5, 1121), (7, 626)$ and $(9, 967)$. These points are solutions to the equation $y = ax^2 + bx + c$, so we obtain the following equations:

$$1121 = a(5)^2 + b(5) + c \rightarrow 1121 = 25a + 5b + c$$
$$626 = a(7)^2 + b(7) + c \rightarrow 626 = 49a + 7b + c$$
$$967 = a(9)^2 + b(9) + c \rightarrow 967 = 81a + 9b + c$$

Now, we have a system of three equations and three unknowns, so we can solve the system of equations. As a result, we find $a = 104.5$, $b = -1501.5$ and $c = 6016$. Substituting these values into $y = ax^2 + bx + c$, we find the equation $y = 104.5x^2 - 1501.5x + 6016$. Now that we have an equation that models the data, how can we find the death rate for males who sleep 2 hours? Let's plug 2 in for x and see what we get.

$$y = 104.5x^2 - 1501.5x + 6016$$
$$y(2) = 104.5(2)^2 - 1501.5(2) + 6016$$
$$y = 3431$$

So, the death rate for males who sleep 2 hours is 3431 (per 100,000 males). Solving for the unknowns in $y = ax^2 + bx + c$ is not an easy task and involves this 3x3 system of equations described previously. We could also use the TI approach described in the link http://faculty.uml.edu/mstick and directly at http://faculty.uml.edu/mstick/Links/TI_Systems_of_Eqs_and_Least_Squares_PDF.pdf.

Like linear regression, the model we found only uses three points on the parabola. How can we get a more accurate model for the data? Let's use a calculator to find the least squares regression for a quadratic function. This is a very similar process to the process that is outlined in linear regression. Using the same methods as linear regression, enter the data in the scatterplot and graph the data points. Then, using the CALC option in the STAT menu, choose option 5: QuadReg.
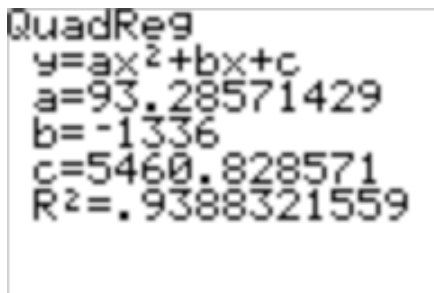


```
QuadReg
 y=ax²+bx+c
 a=93.28571429
 b= -1336
 c=5460.828571
 R²=.9388321559
```

Figure 18

The values for a, b and c display on the TI screen, along with a value for r-squared. So, the quadratic least squares suggests the model accurate to 2 decimal places to be $y = 93.29x^2 - 1336x + 5460.83$. Notice this model is similar to the model found using three data points and provides a more accurate model for the data. Make sure you can graph the function, along with the scatterplot of data.

Example 4

The following example is a life science application, relating the average number of live births to woman of a particular age. The data is shown in table 4. [Bittinger, CALCULUS and ITS APPLICATIONS, 10th ed, page 74]

Age, x	Average Number of Live Births per 1000 woman
16	34
18.5	86.5
22	111.1
27	113.9
32	84.5
37	35.4
42	6.8

Table 4

Let's start by finding a quadratic function using the least squares QuadReg option. After entering the data in the table and using the QuadReg option, we get the following results.
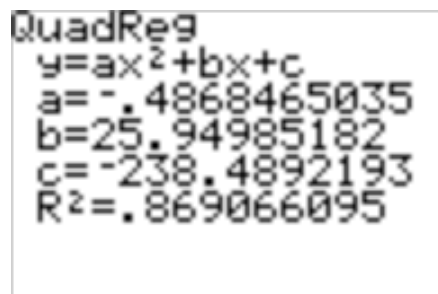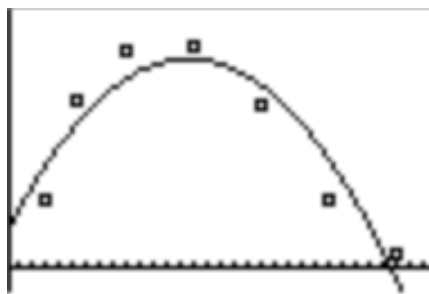

Figure 19

Figure 20

Figure 19 suggests the model $y = -0.49x^2 + 25.95x - 238.49$ accurate to 2 decimal places and an r-squared of 0.869, which means the model is not as good as models we have seen in prior examples. In figure 20, we can see there is some error between the data points and the model. Let's look at a cubic model.

14

Keeping the same data in the table, let's use the cubic regression command in the TI. Press STAT, move the arrow to the CALC option and select option 6, or CubicReg. The calculator calculates a value for *a*, *b*, *c* and *d*. Also, we graph the new model and the data.
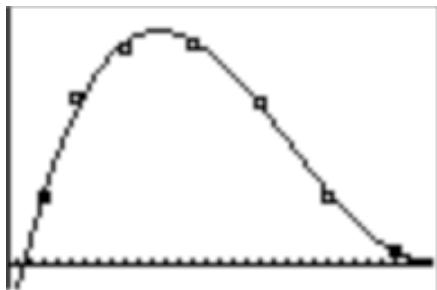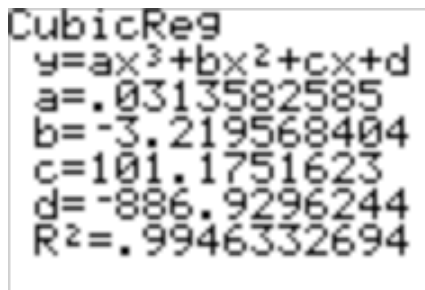

Figure 21

Figure 22

Figure 21 produces accurate to 3 decimal places the model
$y = 0.031x^3 - 3.220x^2 + 101.175x - 886.930.$ Now, the *r*-squared for the model is now 0.9946, which suggests this model is much better than the quadratic model.

Derivation of Quadratic Least Squares Regression Model

Similar to the linear regression, the method of least squares requires that we minimize the sum of the squares of the errors between the actual data and the predicted model results. Again, this math gets complicated but let's look at where things are coming from.

Just like before, we begin by taking the difference of the estimate, or
$I_n = \sum \left(y_i - \left(ax_i^2 + bx_i + c \right) \right)^2.$ Next, we take the partial derivative of I_n with respect to *a*, and set the equation equal to zero.

$$I_n = \sum_{i=1}^{n} \left(y_i - \left(ax_i^2 + bx_i + c \right) \right)^2$$

$$\frac{\partial \left(I_n \right)}{\partial a} = 2\sum_{i=1}^{n} \left(y_i - \left(ax_i^2 + bx_i + c \right) \right) \left(-x_i \right)^2$$

$$2\sum_{i=1}^{n} \left(y_i - \left(ax_i^2 + bx_i + c \right) \right) \left(-x_i \right)^2 = 0$$

Dividing by two gives us

$$\sum_{i=1}^{n} \left(y_i - \left(ax_i^2 + bx_i + c \right) \right) \left(-x_i \right)^2 = 0$$

Next, let's distribute the $\left(-x_i \right)^2$ term, along with the summation symbol.

15

$$\sum_{i=1}^{n}\left(-y_i x_i^2 + ax_i^4 + bx_i^3 + cx_i^2\right) = 0$$

$$-\sum_{i=1}^{n}y_i x_i^2 + \sum_{i=1}^{n}ax_i^4 + \sum_{i=1}^{n}bx_i^3 + \sum_{i=1}^{n}cx_i^2 = 0$$

$$a\sum_{i=1}^{n}x_i^4 + b\sum_{i=1}^{n}x_i^3 + c\sum_{i=1}^{n}x_i^2 = \sum_{i=1}^{n}y_i x_i^2 \qquad \textbf{(Equation 4)}$$

Next, let's use these same steps to find the partial derivative of derivative of I_n with respect to b.

$$I_n = \sum_{i=1}^{n}\left(y_i - \left(ax_i^2 + bx_i + c\right)\right)^2$$

$$\frac{\partial(I_n)}{\partial b} = 2\sum_{i=1}^{n}\left(y_i - ax_i^2 - bx_i - c\right)\left(-x_i\right)$$

$$2\sum_{i=1}^{n}\left(y_i - ax_i^2 - bx_i - c\right)\left(-x_i\right) = 0$$

Dividing by two gives us

$$\sum_{i=1}^{n}\left(y_i - ax_i^2 - bx_i - c\right)\left(-x_i\right) = 0$$

Next, let's distribute the $-x_i$ term, along with the summation symbol.

$$\sum_{i=1}^{n}\left(-y_i x_i + ax_i^3 + bx_i^2 + cx_i\right) = 0$$

$$-\sum_{i=1}^{n}y_i x_i + \sum_{i=1}^{n}ax_i^3 + \sum_{i=1}^{n}bx_i^2 + \sum_{i=1}^{n}cx_i = 0$$

$$\sum_{i=1}^{n}ax_i^3 + \sum_{i=1}^{n}bx_i^2 + \sum_{i=1}^{n}cx_i = \sum_{i=1}^{n}y_i x_i$$

$$a\sum_{i=1}^{n}x_i^3 + b\sum_{i=1}^{n}x_i^2 + c\sum_{i=1}^{n}x_i = \sum_{i=1}^{n}y_i x_i \qquad \textbf{(Equation 5)}$$

Using the same method, we find the partial derivative of I_n with respect to c.

$$I_n = \sum_{i=1}^{n}\left(y_i - \left(ax_i^2 + bx_i + c\right)\right)^2$$

$$\frac{\partial(I_n)}{\partial c} = 2\sum_{i=1}^{n}\left(y_i - ax_i^2 - bx_i - c\right)(-1)$$

$$2\sum_{i=1}^{n}\left(y_i - ax_i^2 - bx_i - c\right)(-1) = 0$$

$$\sum_{i=1}^{n}\left(-y_i + ax_i^2 + bx_i + c\right) = 0$$

$$\sum_{i=1}^{n}-y_i + \sum_{i=1}^{n}ax_i^2 + \sum_{i=1}^{n}bx_i + \sum_{i=1}^{n}c = 0$$

$$a\sum_{i=1}^{n}x_i^2 + b\sum_{i=1}^{n}x_i + \sum_{i=1}^{n}c = \sum_{i=1}^{n}y_i$$

Notice that we have $\sum_{i=1}^{n}c = cn$.

So, $a\sum_{i=1}^{n}x_i^2 + b\sum_{i=1}^{n}x_i + cn = \sum_{i=1}^{n}y_i$ **(Equation 6)**

Now, equations 4, 5 and 6 make up a 3x3 system of equations.

$$a\sum_{i=1}^{n}x_i^4 + b\sum_{i=1}^{n}x_i^3 + c\sum_{i=1}^{n}x_i^2 = \sum_{i=1}^{n}x_i^2 y_i$$

$$a\sum_{i=1}^{n}x_i^3 + b\sum_{i=1}^{n}x_i^2 + c\sum_{i=1}^{n}x_i = \sum_{i=1}^{n}x_i y_i$$

$$a\sum_{i=1}^{n}x_i^2 + b\sum_{i=1}^{n}x_i + cn = \sum_{i=1}^{n}y_i$$

We have a system of three equations with three unknowns. Like the system of two equations and two unknowns, we can use Cramer's Rule to solve. We let

$$\Delta = \begin{vmatrix} \sum_{i=1}^{n}x_i^4 & \sum_{i=1}^{n}x_i^3 & \sum_{i=1}^{n}x_i^2 \\ \sum_{i=1}^{n}x_i^3 & \sum_{i=1}^{n}x_i^2 & \sum_{i=1}^{n}x_i \\ \sum_{i=1}^{n}x_i^2 & \sum_{i=1}^{n}x_i & n \end{vmatrix}$$

So,

$$a = \dfrac{\begin{vmatrix} \sum\limits_{i=1}^{n} x_i^2 y_i & \sum\limits_{i=1}^{n} x_i^3 & \sum\limits_{i=1}^{n} x_i^2 \\ \sum\limits_{i=1}^{n} x_i y_i & \sum\limits_{i=1}^{n} x_i^2 & \sum\limits_{i=1}^{n} x_i \\ \sum\limits_{i=1}^{n} y_i & \sum\limits_{i=1}^{n} x_i & n \end{vmatrix}}{\Delta}, \; b = \dfrac{\begin{vmatrix} \sum\limits_{i=1}^{n} x_i^4 & \sum\limits_{i=1}^{n} x_i^2 y_i & \sum\limits_{i=1}^{n} x_i^2 \\ \sum\limits_{i=1}^{n} x_i^3 & \sum\limits_{i=1}^{n} x_i y_i & \sum\limits_{i=1}^{n} x_i \\ \sum\limits_{i=1}^{n} x_i^2 & \sum\limits_{i=1}^{n} y_i & n \end{vmatrix}}{\Delta}, \; c = \dfrac{\begin{vmatrix} \sum\limits_{i=1}^{n} x_i^4 & \sum\limits_{i=1}^{n} x_i^3 & \sum\limits_{i=1}^{n} x_i^2 y_i \\ \sum\limits_{i=1}^{n} x_i^3 & \sum\limits_{i=1}^{n} x_i^2 & \sum\limits_{i=1}^{n} x_i y_i \\ \sum\limits_{i=1}^{n} x_i^2 & \sum\limits_{i=1}^{n} x_i & \sum\limits_{i=1}^{n} y_i \end{vmatrix}}{\Delta}$$

In order to further solve for a, b and c we need to know how to solve a three by three determinant.

Consider the determinant $\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix}$. The formula to solve a three by three determinant is as follows:

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix}$$

$$= a(ei - hf) - b(di - gf) + c(dh - ge)$$

Note that we solved the three by three determinant by expanding about elements in the first row. We could have also solved the determinant by expanding about elements in another row or column. Some people like to expand about elements in the first column, but regardless, results should numerically be the same.

So, let's find the determinant of Δ.

$$\Delta = \begin{vmatrix} \sum_{i=1}^{n} x_i^4 & \sum_{i=1}^{n} x_i^3 & \sum_{i=1}^{n} x_i^2 \\ \sum_{i=1}^{n} x_i^3 & \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i & n \end{vmatrix}$$

$$= \sum_{i=1}^{n} x_i^4 \begin{vmatrix} \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & n \end{vmatrix} - \sum_{i=1}^{n} x_i^3 \begin{vmatrix} \sum_{i=1}^{n} x_i^3 & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i^2 & n \end{vmatrix} + \sum_{i=1}^{n} x_i^2 \begin{vmatrix} \sum_{i=1}^{n} x_i^3 & \sum_{i=1}^{n} x_i^2 \\ \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i \end{vmatrix}$$

$$= \sum_{i=1}^{n} x_i^4 \left(n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2 \right) - \sum_{i=1}^{n} x_i^3 \left(n\sum_{i=1}^{n} x_i^3 - \sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} x_i \right) + \sum_{i=1}^{n} x_i^2 \left(\sum_{i=1}^{n} x_i^3 \sum_{i=1}^{n} x_i - \left(\sum_{i=1}^{n} x_i^2\right)^2 \right)$$

Let's label the solution for Δ as **Equation 7**.

Similarly to finding the determinant of Δ, let's find equations for the determinants of a, b and c.

$$a = \frac{\begin{vmatrix} \sum_{i=1}^{n} x_i^2 y_i & \sum_{i=1}^{n} x_i^3 & \sum_{i=1}^{n} x_i^2 \\ \sum_{i=1}^{n} x_i y_i & \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} y_i & \sum_{i=1}^{n} x_i & n \end{vmatrix}}{\Delta} \qquad \text{(Equation 8)}$$

$$= \frac{\sum_{i=1}^{n} x_i^2 y_i \left(n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2 \right) - \sum_{i=1}^{n} x_i^3 \left(n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i \right) + \sum_{i=1}^{n} x_i^2 \left(\sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i \right)}{\Delta}$$

$$b = \frac{\begin{vmatrix} \sum_{i=1}^{n} x_i^4 & \sum_{i=1}^{n} x_i^2 y_i & \sum_{i=1}^{n} x_i^2 \\ \sum_{i=1}^{n} x_i^3 & \sum_{i=1}^{n} x_i y_i & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} y_i & n \end{vmatrix}}{\Delta} \qquad \text{(Equation 9)}$$

$$= \frac{\sum_{i=1}^{n} x_i^4 \left(n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i \right) - \sum_{i=1}^{n} x_i^2 y_i \left(n\sum_{i=1}^{n} x_i^3 - \sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} x_i \right) + \sum_{i=1}^{n} x_i^2 \left(\sum_{i=1}^{n} x_i^3 \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} x_i y_i \right)}{\Delta}$$

$$c = \frac{\begin{vmatrix} \sum\limits_{i=1}^{n} x_i^{4} & \sum\limits_{i=1}^{n} x_i^{3} & \sum\limits_{i=1}^{n} x_i^{2} y_i \\ \sum\limits_{i=1}^{n} x_i^{3} & \sum\limits_{i=1}^{n} x_i^{2} & \sum\limits_{i=1}^{n} x_i y_i \\ \sum\limits_{i=1}^{n} x_i^{2} & \sum\limits_{i=1}^{n} x_i & \sum\limits_{i=1}^{n} y_i \end{vmatrix}}{\Delta} \qquad \text{(Equation 10)}$$

$$= \frac{\sum\limits_{i=1}^{n} x_i^{4}\left(\sum\limits_{i=1}^{n} x_i^{2}\sum\limits_{i=1}^{n} y_i - \sum\limits_{i=1}^{n} x_i y_i \sum\limits_{i=1}^{n} x_i\right) - \sum\limits_{i=1}^{n} x_i^{3}\left(\sum\limits_{i=1}^{n} x_i^{3}\sum\limits_{i=1}^{n} y_i - \sum\limits_{i=1}^{n} x_i y_i \sum\limits_{i=1}^{n} x_i^{2}\right) + \sum\limits_{i=1}^{n} x_i^{2} y_i \left(\sum\limits_{i=1}^{n} x_i^{3}\sum\limits_{i=1}^{n} x_i - \left(\sum\limits_{i=1}^{n} x_i^{2}\right)^{2}\right)}{\Delta}$$

Notice that solving **Equations 7**, **8**, **9** and **10** to find the least squares coefficients a, b and c for a quadratic model $y = ax^2 + bx + c$ is quite cumbersome. What is presented here is the mathematics behind a quadratic least squares regression model. Fortunately, algorithms have been created to do a quadratic least squares on the TI calculators and in spreadsheets such as Excel.