9 Linear and Quadratic Regressions

In general, data obtained from real life events, do not match perfectly simple functions. Very often, scientists, engineers, mathematicians and business experts can model the data obtained from their studies, with simple linear functions. Even if the function does not reproduce the data exactly, it is possible to use this modeling for further analysis and predictions. This makes the linear modeling extremely valuable.

Let’s try to fit a set of data points from a crankcase motor oil producing company. They want to study the correlation between the number of minutes of TV advertisement per day for their product, and the total number of oil cases sold per month for each of the different advertising campaigns. The information is given in the following table:

<table>
<thead>
<tr>
<th>x: TV ads (min/day)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>3.5</th>
<th>5.5</th>
<th>6.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>y: units sold (in millions per month)</td>
<td>1</td>
<td>2.5</td>
<td>3.7</td>
<td>4.2</td>
<td>7</td>
<td>8.7</td>
</tr>
</tbody>
</table>

Using TI-83 we obtain the scatter plot of this given data (See Figure 9.1) by using the following steps:

I. Enter the data into two lists L1 and L2.

a. Push the STAT key and select the Edit key from the Edit menu.
b. Up arrow to move to the Use the top of the list L1.
c. Clear the list by hitting CLEAR ENTER.
d. Type in the x values of the data. Type in the number and hit enter.
e. Move to the list L2, clear it and enter the y data in this list.

II. Graph the data as a scatterplot.

a. Hit 2nd STAT PLOT. (upper left)
b. Move to plot 1 and hit enter.
c. Turn the plot on by hitting enter on the ON option.
d. Move to the TYPE option and select the “dot” graph type. Hit enter to select it.
e. Move to the Xlist and enter in 2nd L1.
f. Move to the Ylist and enter in 2nd L2.
g. Move to Mark and select the small box option.
h. Hit ZOOM key and select ZOOMSTAT.

Figure 9.1

The next question is the question of finding a line that best “fit” the given data. This line is called the **best fitting line** and has been derived with a very commonly used statistical technique called the **method of least squares**. The best fit line was chosen to minimize the sum of the squares of the vertical distances between the data points and the line.

The measure of how well this linear function fits the experimental points, is called **regression analysis**.

Graphic calculators, such as the TI-83, have built in programs which allow us to find the slope and the \( y \) intercept of the best fitting line to a set of data points. That is, the equation of the best linear fit. The calculators also give as a result of their procedure, a very important value called the **correlation coefficient**. This value is in general represented by the letter \( r \) and it is a measure of how well the best fitting line fits the data points. Its value varies from \(-1\) to \(1\). The TI-83 prompts the correlation coefficient \( r \) as a result of the linear regression. If it is negative, it is telling us that the line obtained has negative slope. Positive values of \( r \) indicate a positive slope in the best fitting line. A perfect fit gives a coefficient of \(1.0\) or \(-1.0\). Thus the higher the correlation coefficient the better. If \( r \) is close to 0 then the data may be completely scattered, or there may be a non-linear relationship between the variables.

A correlation greater than 0.8 and less than \(-0.8\) is generally described as strong, whereas a correlation between \(-0.5\) and 0.5 is generally described as
weak.
The square value of the correlation coefficient $r^2$ is generally used to determine if the best fitting line can be used as a model for the data. For that reason, $r^2$ is called the coefficient of determination. In most cases, a function is accepted as the model of the data, if this coefficient of determination is greater than 0.5.

A coefficient of determination represents the percent of the data that is the closest to the line of best fit. The closer the coefficient of determination is to 1 the better the fit.

The following are the steps required to find the equation of the best fit, its graph, the coefficient of correlation and the coefficient of determination using a TI-83 graphing calculator.

III. **Fit a line to the data.**

a. Turn on the option to display the correlation coefficient, $r$. This is accomplished by hitting 2nd Catalog (lower left). Scroll down the list until you find Diagnostic On, hit enter for this option and hit enter a second time to activate this option. The correlation coefficient will be displayed when you do the linear regression.

b. Hit STAT, CALC, and select option 8 LinReg.

c. Enter “2nd 1, 2nd 2”, with a comma in between.

d. Press ENTER. The equation for a line through the data is shown. The slope is “b”, the intercept is “a”, the correlation coefficient is “r”, and the coefficient of determination is “$r^2$”.

IV. **Graph the best fit line with the data.**

a. Press Y=, then press VARS to open the Variables window.

b. Arrow down to select 5: Statistics then press ENTER.

c. Right arrow over to select EQ and press ENTER. This places the formula for the regression equation into the $Y =$ window.

d. Press GRAPH to graph the equation. Your window should now show the graph of the regression equation as well as each of the data points.

Figure 9.2 below shows the scatter plot and the optimum linear function that describes the data.
Remark 9.1
The above discussion works as well for quadratic regression where QuadReg is used instead of LinReg.