# Project 3

Cera Olson
Jacob Mastel
Robert Erick

May 25, 2015

# Contents

# Question 1

# Question 2

# 1 Part A

One alternative to the least squares line is the Least Absolute Deviations (LAD). Formulate a linear program whose optimal solution minimizes the sum of the absolute deviations of the data from the line. That is formulate

$$min \sum_{i=1}^{n} |y_i - (a_1 x_i + a_0)|$$

as an LP and solve for the $a_0$ and $a_1$ that minimize the sum of absolute deviations.

## 1.1 i: Write the linear program for the general problem written as an objective and set of constraints

The goal is to minimize $min \sum_{i=1}^{n} |y_i - (a_1 x_i + a_0)|$. In order to create an objective, we drop the sum and set it equal to $z_i$ for all values $i = 1, ..., n$. We can reduce that by dropping the absolute values and set it as an inequality.

$$-z_i \leq y_i - (a_1 x_i + a_0) \leq z_i$$

After that it gets simplified down to the following objectives and constraints.

$$y_i - (a_1 x_i + a_0) \leq z_i \text{ for all values } i = 1, ..., n$$

$$y_i - (a_1 x_i + a_0) \geq -z_i \text{ for all values } i = 1, ..., n$$

## 1.2 ii: Use the linear program to find the LAD regression line for the data set $(x, y) = (1, 5), (1, 3), (2, 13), (3, 8), (4, 10), (5, 14), (6, 18)$ What was the sum of absolute deviations?

The absolute deviation is calculated by taking the least squares values for y and finding the difference between that and the calculated actual value of y using the data. See the chart below. The trendline has an equation of $y = 2.315x + 2.875$

Table 1: Part A (ii)

| x | y: Data Points | Trendline | Differences | Squared |
|---|---|---|---|---|
| 1 | 5 | 3.93 | 1.07 | 1.15 |
| 1 | 3 | 3.93 | 0.93 | 0.87 |
| 2 | 13 | 5.99 | 7.01 | 49 |
| 3 | 8 | 8.07 | 0.07 | 0.01 |
| 4 | 10 | 10.14 | 0.14 | 0.02 |
| 5 | 14 | 12.21 | 1.79 | 3.2 |
| 6 | 18 | 14.29 | 3.72 | 13.84 |

Based on the chart above, the sum of the absolute deviations is 14.73.

### 1.3 iii: Plot the points and graph your LAD line and the least squares line. Comment.

The value for point 2 appears to be an outlier. The value of the data point at x = 2 falls outside the line of best fit the most.

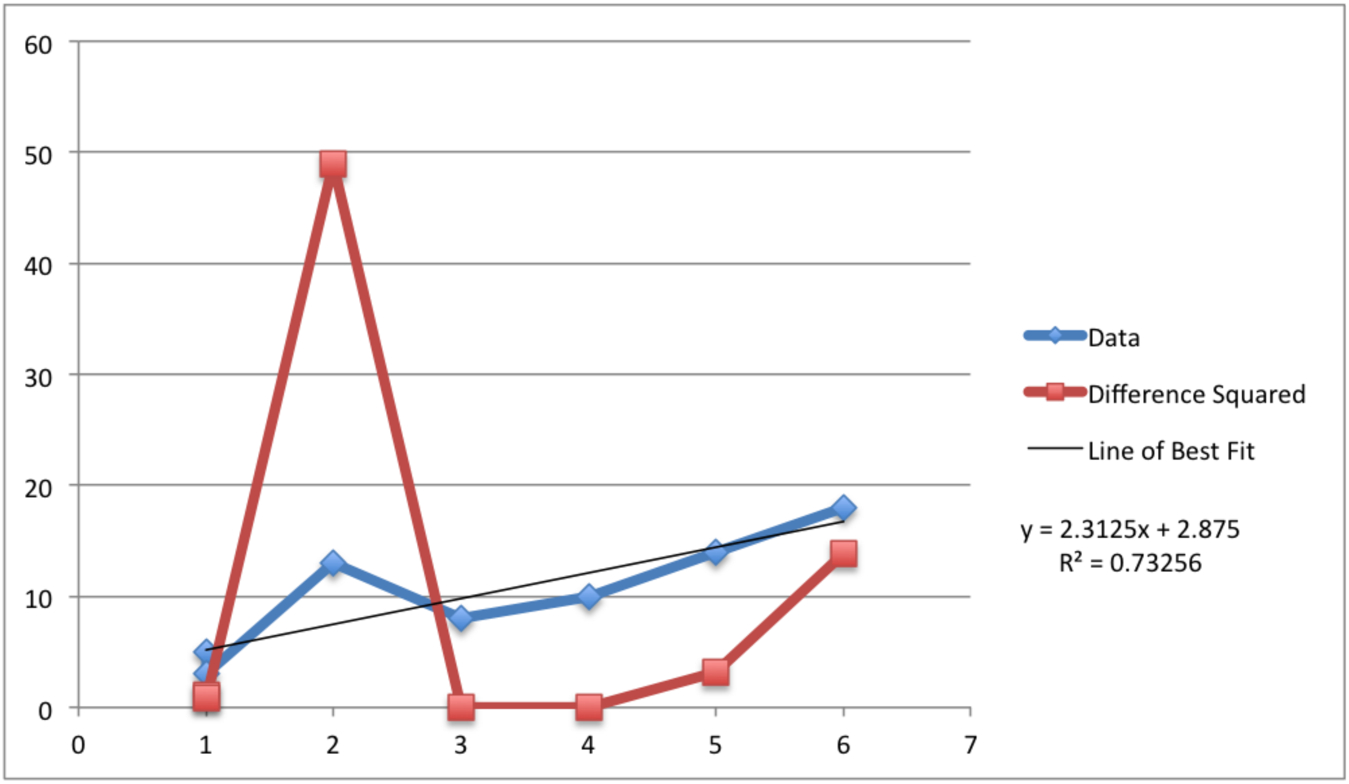


Figure 1: example caption

## 2 Part B

Another alternative to the least squares method is to find a line that minimizes of the maximum absolute deviation (MMAD). That is formulate

$$min\ max_i\ |y_i - (a_1 x_i + a_0)|$$

as an LP.

### 2.1 i: Write the linear program for the general problem written as an objective and set of constraints

Following the same procedures as in Part A, set the equation equal to z and try to minimize z for all values $i = 1, ..., n$. The resulting equations are:

$$y_i - (a_1 x_i + a_0) \geq z_i \text{ for all values } i = 1, ..., n$$

$$y_i - (a_1 x_i + a_0) \leq -z_i \text{ for all values } i = 1, ..., n$$

It is important to note that these are opposite of the solutions as found in part A.

### 2.2 ii: Use the linear program to find the MMAD regression line for the data set $(x, y) = (1, 5), (1, 3), (2, 13), (3, 8), (4, 10), (5, 14), (6, 18)$ What was the min of the max absolute deviations?

Minimize z subject to $z \geq max_i \ |y_i - (a_1 x_i + a_0)|$

### 2.3 iii: Plot the points and graph the MMAD line and the least squares line. Compare.

### 2.4 iv: Can you create a data set for which all three methods of regression (least squares, LAD, MMAD) compute the same line.

The only set that could have the same result is the empty set or a set of zero values. All three methods use different methodologies to calculate the line of best fit. They all use either minimization, maximization, or a combination there of, and as such, there will be minute differences between the calculations of the regression.

## 3 Part C

Multiple Linear Regression. Generalize the simple linear regression model to allow for two independent variables ($x_1$ and $x_2$). ?? =??$_2$??$_2$+??$_1$??$_1$+??$_0$ , The model is called multiple linear not because the result is a line but because all variables are 1st degree. Extend the techniques from Part A to find the least absolute deviations regression equation. Use linear programming to fit a LAD multiple linear regression model to the data set below:

| $x_1$ | $x_2$ | y |
|---|---|---|
| 1 | 1 | 5 |
| 1 | 2 | 9 |
| 2 | 2 | 12 |
| 0 | 1 | 3 |
| 0 | 0 | 0 |
| 1 | 3 | 11 |

Solving for $a_0$, $a_1$, and $a_2$ using a system of equations and the values in the table above. Using the above values, $a_0$ must be 0. It is the only way $x_1$ and $x_2$ could be zero if y is 0. The result is that $a_2$ equals 3. The final value, $a_1$, is 2 or 3 depending on the data points you use to calculate them. Using LAD, we minimize the values. making $a_1 = 2$.
The final estimation is $y = 3 \times x_2 + 2 \times x_1$.