AMS 315 Project 1 Part B
Professor: Stephen Finch
Group 87: Ivan Tinov, Lipeng Feng

Introduction:

In part B of this project, our goal is to take an excel data sheet with Y-values(DV) and X-values(DV) and transform either or both the independent variable (IV) and the dependent variable (DV) in order to find a best fitted linear model. Lack of fit tests were conducted on the data to see if the fit was adequate.

Methodology:

In order to solve this part of the project, we used Minitab to conduct the transformations, lack of fit tests, and creating the fitted linear model. We tried many transformations for both the X-axis (IV) and the Y-axis (DV) and found that the transformation that had the highest $R^2$ value was a transformation where we squared the independent variable (X). We left the dependent variable (Y) untouched and achieved an $R^2$ value of 55.8%. We also tried It is important to note that our data had no repeats but had near repeats instead which resulted in no lack of fit test needed. Similarly, there were no outliers in our data set. To transform the data, we used the calculator function in Minitab and edited the independent variable by squaring it ($X \rightarrow X^2$). This was done by going to **Calc -> Calculator**. After that we performed a fitted line plot by going to **Stat -> Regression -> Fitted Line Plot**. Likewise, we performed a fit regression model by going to **Stat -> Regression -> Regression -> Fit Regression Model**. After that, we were able to gather all of the data described below in the Result section of this report.

Result:

The original fitted linear equation for our untransformed data was $Y = 129986 + 4.63X$ that had an $R^2$ value of 0.0%. The regression line was basically horizontal across the graph (see below). After performing a transformation on the X-Axis (Independent variable), we observed a much better $R^2$ value of 55.8% which lead to a more diagonal line with the fitted linear equation of $Y = 103957 + 30.49X^2$. We observed value of $\beta_1 = 30.49$, and an F-value = 2518.10 for the null hypothesis, $H_0$. Likewise, the association between the independent and dependent variables was p = 0.0. The tables for the transformed data with more detail can be seen below:

The ANOVA table is:

*Table 1*
**Analysis of Variance (ANOVA) Table**
Regression Analysis: DV vs. IV
(n = 1999)

Analysis of Variance

| Source | DF | SS | MS | F | P |
|---|---|---|---|---|---|
| Regression | 1 | 1.03796E+12 | 1.03796E+12 | 2518.10 | 0.000 |
| Error | 1998 | 8.23579E+11 | 4.12201E+08 | | |
| Total | 1999 | 1.86154E+12 | | | |

The Coefficients table is:

*Table 2*
**Coefficients[a]**

Coefficients

| Term | Coef | SE Coef | T-Value | P-Value | VIF |
|---|---|---|---|---|---|
| Constant | 103957 | 689 | 150.81 | 0.000 | |
| X2 | 30.492 | 0.608 | 50.18 | 0.000 | 1.00 |

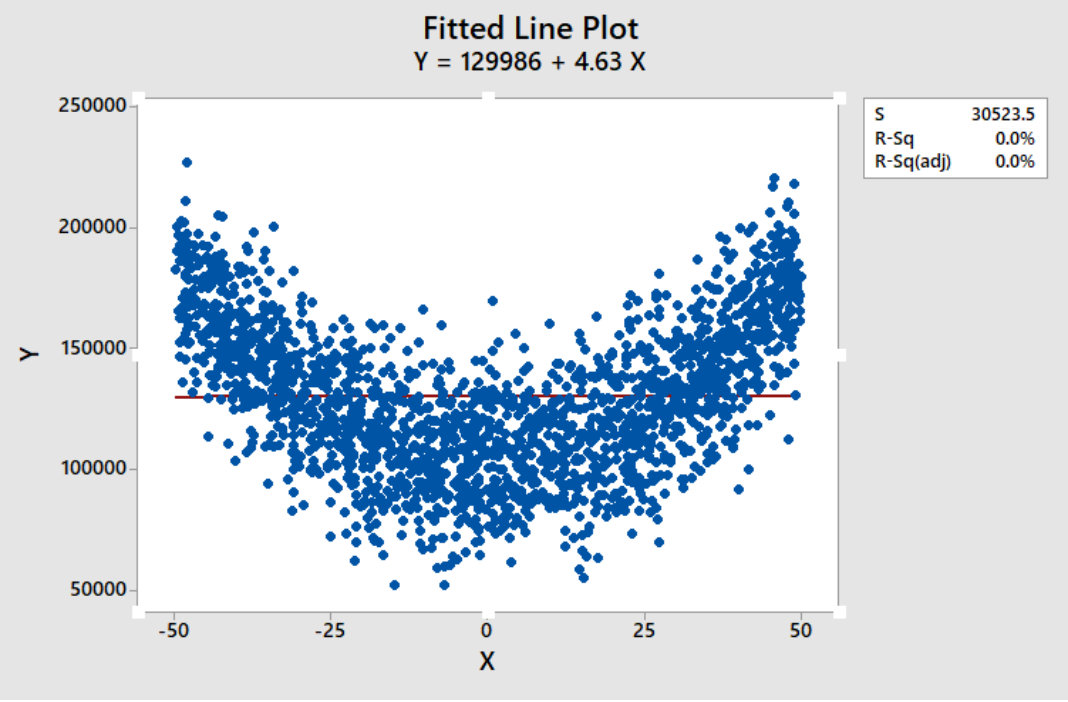
The Model Summary table is:

*Table 3*
**Model Summary[b]**

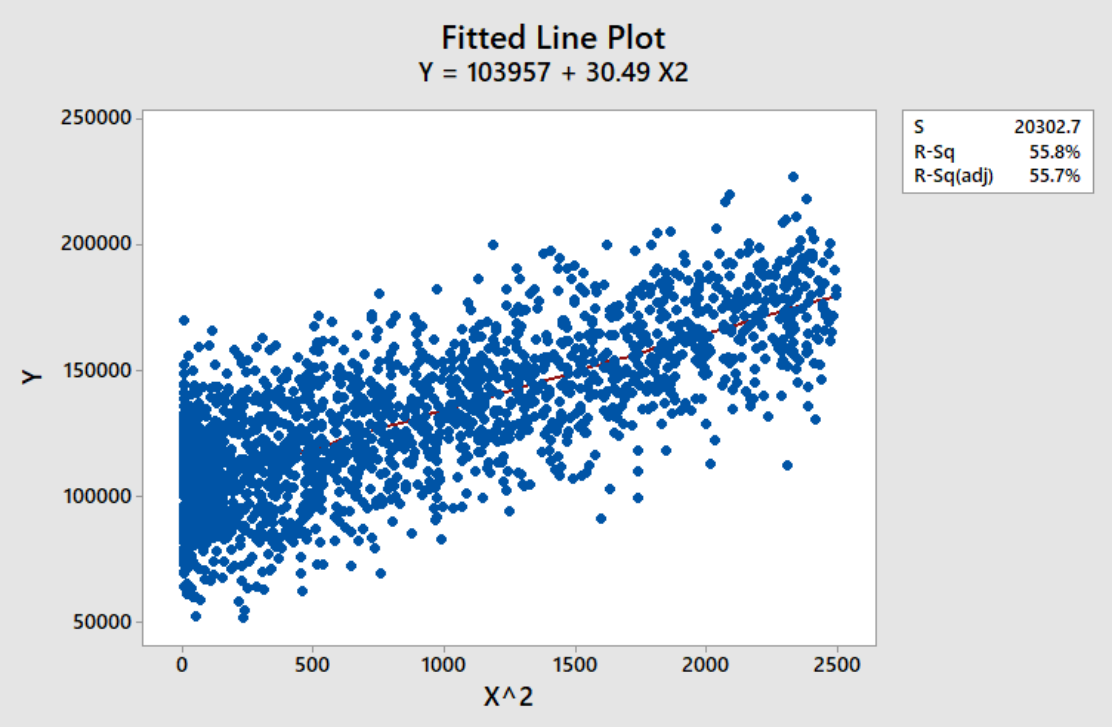
Model Summary

| S | R-sq | R-sq(adj) |
|---|---|---|
| 20302.7 | 55.76% | 55.74% |

**<u>Graph for the Fitted Line Plot:</u> (Before Transformation)**



**<u>Graph for the Fitted Line Plot:</u> (After Transformation)**




We managed to obtain a best possible $R^2$ value of 55.8% when performing a transformation on the X-axis by squaring the X-axis (hence, $X^2$).

<u>Conclusion:</u>

Based on the data gathered above in our fitted linear model, we will not reject the null hypothesis and state that there is not enough evidence to support a lack of fit. After doing the transformation on the X-axis (IV), we obtained a much better $R^2$ value (55.8%) as opposed to before the transformation where we had a $R^2$ value of 0.0%. Thus, we can conclude that the slope is nonzero.

*End of Report*