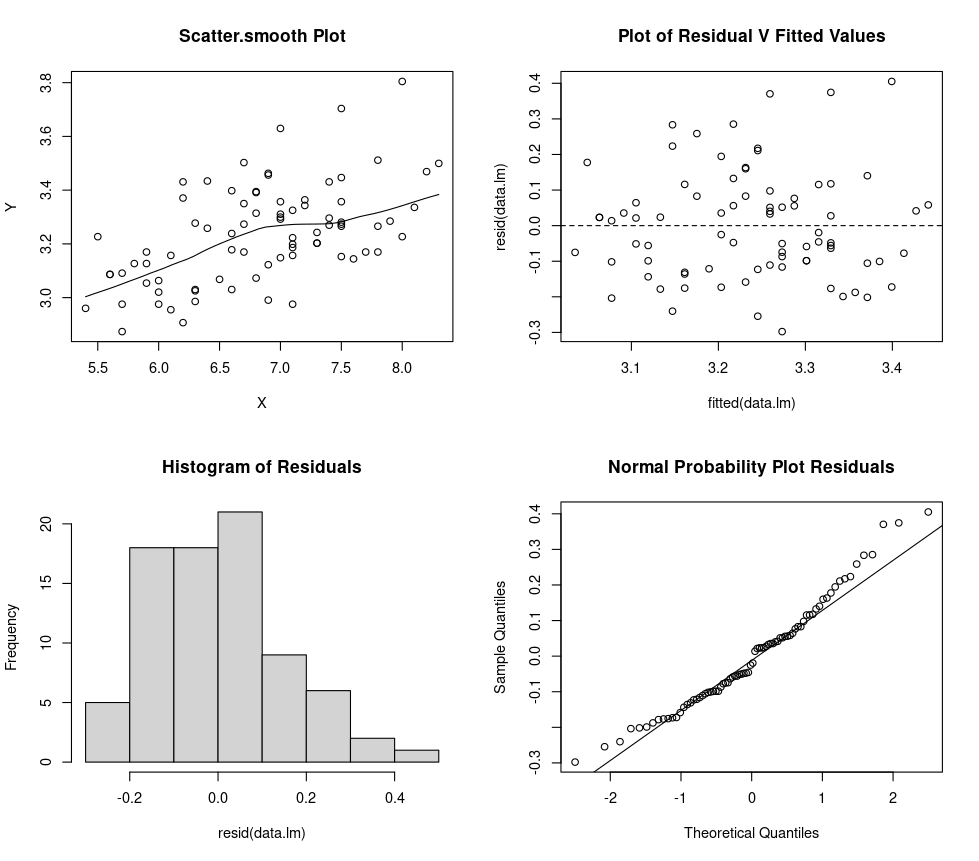
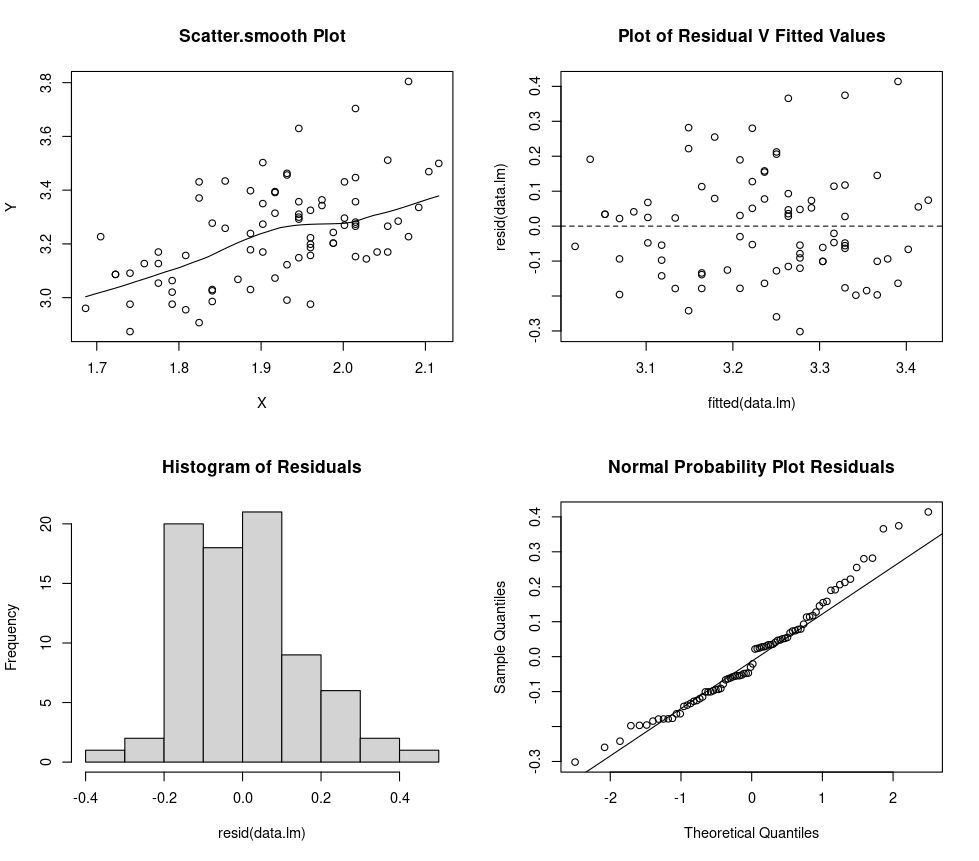
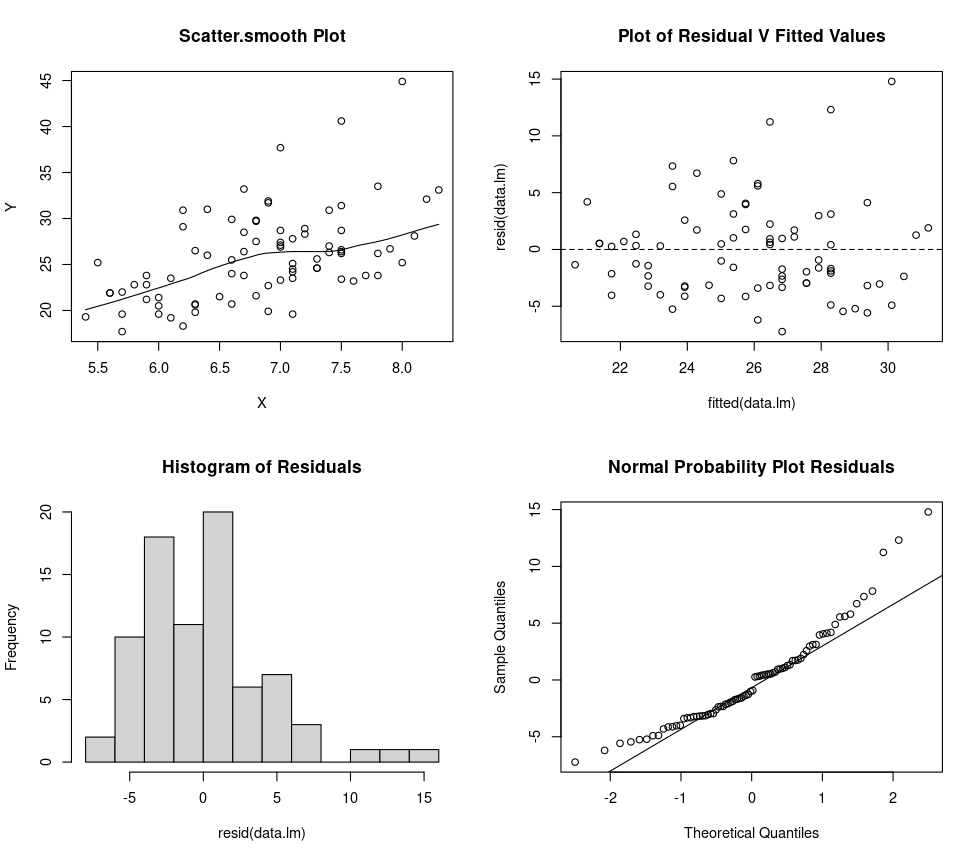
|  |  |
| --- | --- |
| **Module:** | ST2053 |
| **Name:** | Marcus Prunty |
| **Student Number:** | 118730509 |
| **Chapter:** | 4 |

**Maximum 2 pages! Do not delete the page number in the footer.**

Data were collected on a random sample of adults who were undergoing a physical examination. The data are stored in BMI.txt (on Canvas & P: drive).

For a simple linear regression of variable Y = BMI on variable X = Elbow, fit the following three models to the data:

**(a) Provide appropriate diagnostic plots for each model. Comment on each of the diagnostic plots.**



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Scatter-plot** | **Residuals vs Fitted** | **Histogram** | **Normal probability** |
|  | Linear relationship  Increasing Variance | Linear relationship  Increasing Variance | Right skewed Residuals | Non-Normal Residuals |
|  | Linear relationship  Increasing Variance | Linear relationship  Increasing Variance | Right skewed Residuals | Non-Normal Residuals |
|  | Linear relationship  Increasing Variance | Linear relationship  Increasing Variance | Approx Normal Residuals | Non-Normal Residuals |

**(b) Which of the models would you choose for these data? Explain.**

The 3rd Model (log(BMI) ~ log(Elbow)) is the better of the three.

It best satisfies the assumptions.

**(c) Explain why these particular transformations above (logarithmic) would have been considered.**

A logarithmic transformation can be useful on skewed datasets with positive values to normalise the data.

This can be seen in the 3rd chart(histogram of each model) where applying the logarithmic function brings the data closer to a normal distribution