Kungliga Tekniska Högskolan

SF2930 Regression Analysis

Report I

Isac Karlsson
Ludvig Wärnberg Gerdin

Examiner
Tatjana Pavlenko

February 13, 2020

Contents

1 Introduction and Project Goals

2 Analyses and Model Development

2.1 Residual analysis

2.1.1 Normality of residuals

2.1.2 Fitted Against Residuals

2.1.3 Added Variable Analysis

2.2 Diagnostics and handling of Outliers

2.3 Transformations of variables

2.4 Diagnostics and handling of Multicolinearity

3 Results

3.1 Residual analysis

3.1.1 Normality of residuals

Figure 1 illustrates QQ plot of the model residuals. The observer may say that the points exhibit a pattern that indicates that the residuals come from a distribution with heavier tails than that of a normal distribution. [1]. Still, the deviations from the diagonal line is relatively small, and hence we conclude that the first Gauss-Markov condition is fulfilled. That is, the model errors seem to be normally distributed.

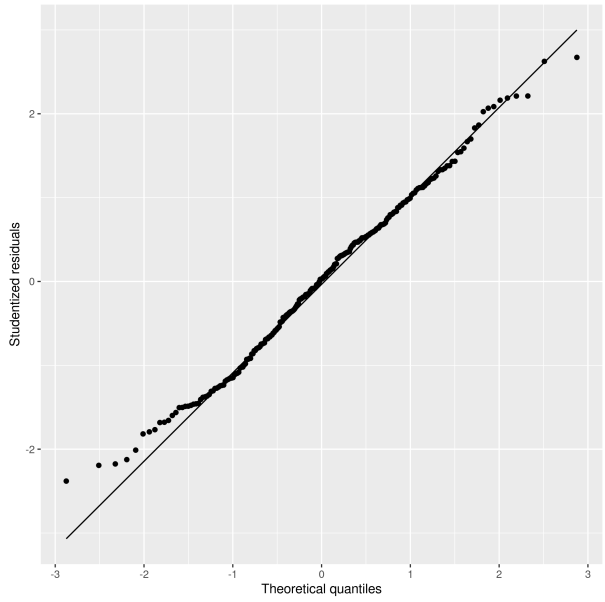

Figure 1: Normality plot of residuals.

3.1.2 Fitted Against Residuals

Figure 2 illustrates the fitted values $\hat{y}_j$ against the R-student residuals. No apparent pattern is formed by the points, i.e. the points seem to be randomly scattered along the horizontal line. Hence we conclude that the second Gauss-Markov condition is fulfilled, that is the errors have a constant variance.

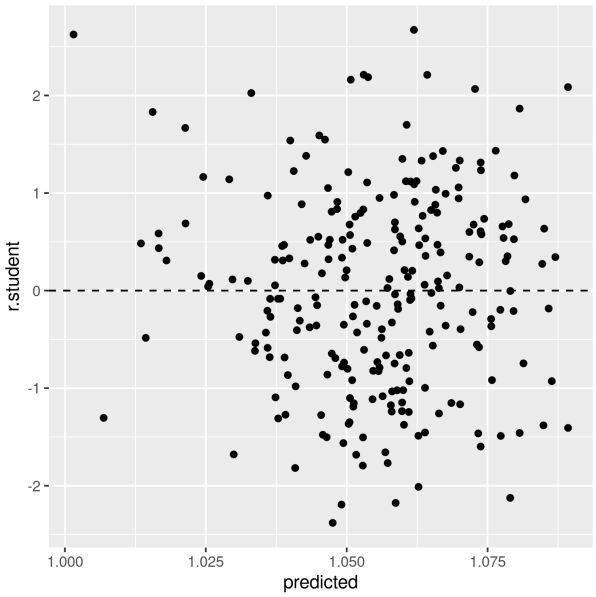

Figure 2: Fitted values against R-student residuals.

3.1.3 Added Variable Analysis

Partial regression plots are found in figure 3, 4, 5, and 6. All figures exhibits potential outliers (which will be further considered in section 2.2). More specifically, in figure 3 we note a few potential outliers on the right hand side of the plot for the `biceps` regressor, and on the right and left hand side for the `forearm` regressor. Moreover, in figure 4, we notice outliers on the right hand side of the `ankle` plot, and a group of potential outliers on the `thigh` plot. Finally, we notice a few potential outliers in figure 5 and 6.

Figure 4, 5, and 6 conveys important information about the information that `knee`, `height`, and `chest` adds to the model. These regressors seem to follow a horizontal band along a fitted line from the origin, which may suggest that none of the regressors adds additional information to the predictions.

3.2 Transformations of variables

Figure 7 displays the values of λ to be used in a potential Box-Cox transformation of the dependent variable `density`. The λ that maximized the log-likelihood is 0.9 (0.7-1.1 95% CI).

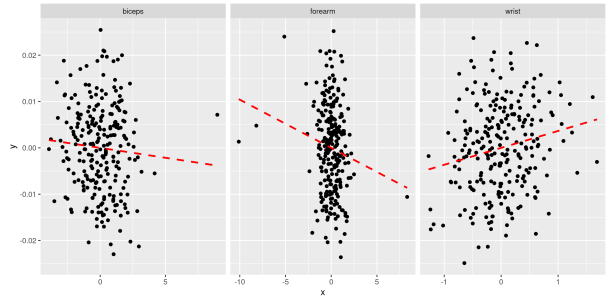

Figure 3: Partial regression plots of regressors `biceps`, `forearm`, and `wrist`.

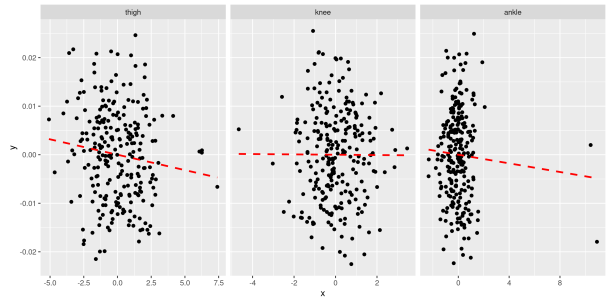

Figure 4: Partial regression plots of regressors `thigh`, `knee`, and `ankle`.

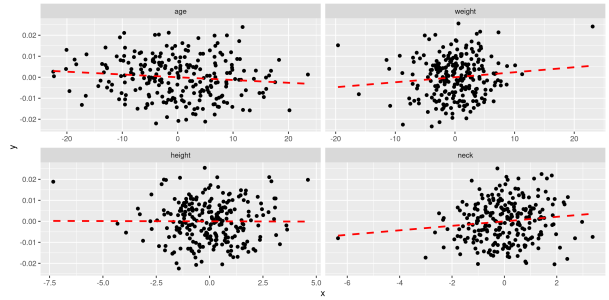

Figure 5: Partial regression plots of regressors `age`, `weight`, `height`, and `neck`.

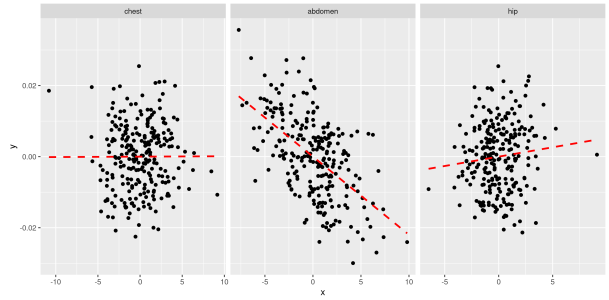

Figure 6: Partial regression plots of regressors `chest`, `abdomen`, and `hip`.

Using $\lambda = 0.9$ gives us the normal probability plot displayed on the right hand side in figure

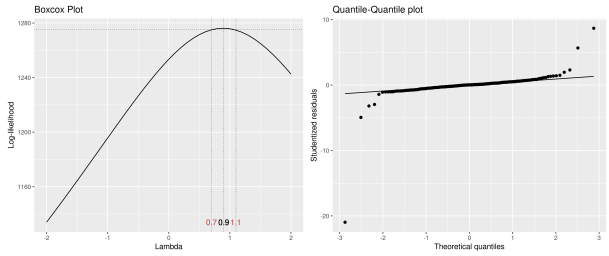

Figure 7: Values for lambda against the log-likelihood of `density` for Box-Cox transformations.

3.3 Diagnostics and handling of Outliers

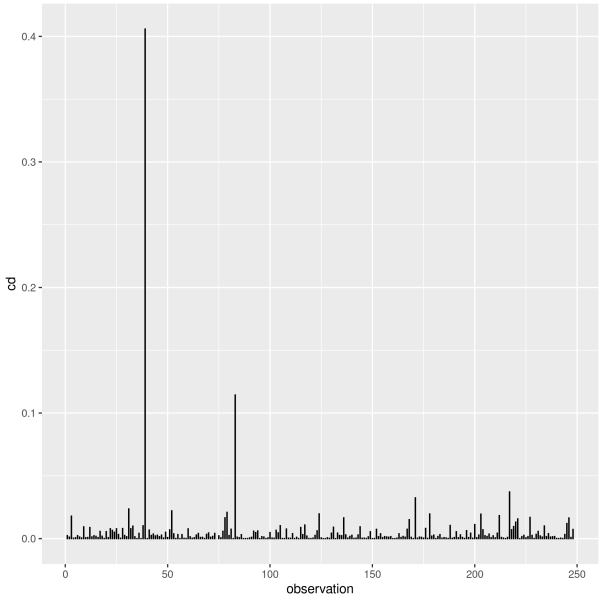

Figure 8: Plot of Cook's distance for all observations.

4 Conclusion

References

[1] Douglas C Montgomery, Elizabeth A Peck, and G Geoffrey Vining. *Introduction to Linear Regression Analysis*. Wiley-Interscience, 5 edition, 2012.

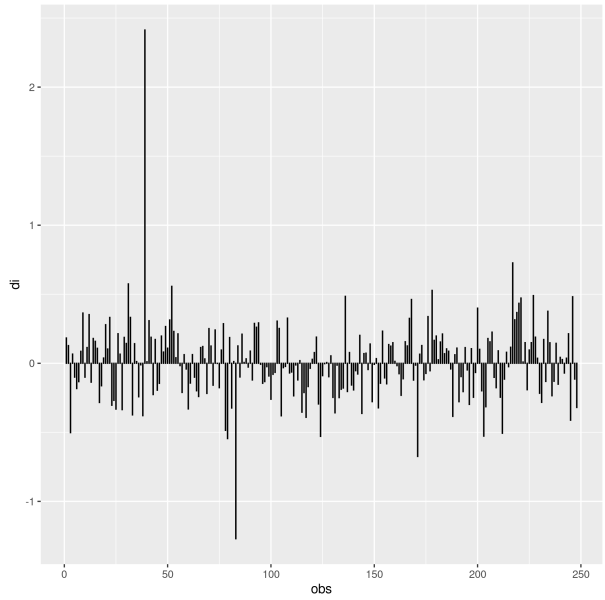

Figure 9: $DFFITS$ for all observations.

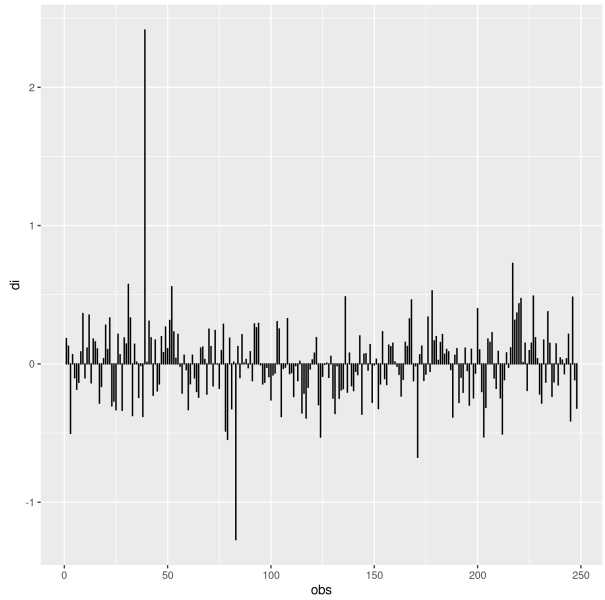

Figure 10: $DFFITS$ for all observations.

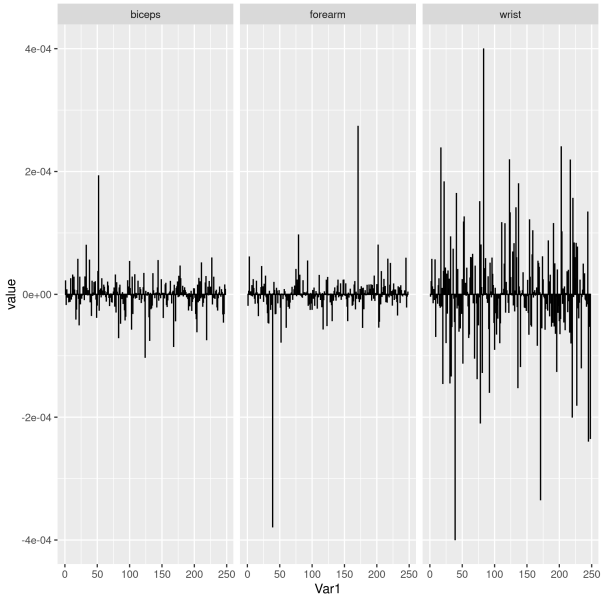

Figure 11: `DFBETA` for regressors `biceps`, `forearm`, and `wrist`.

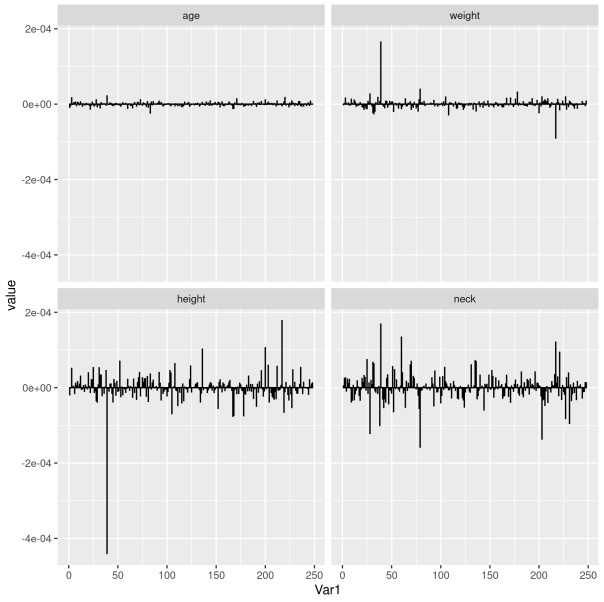

Figure 12: `DFBETA` for regressors `age`, `weight`, `height` and `neck`.

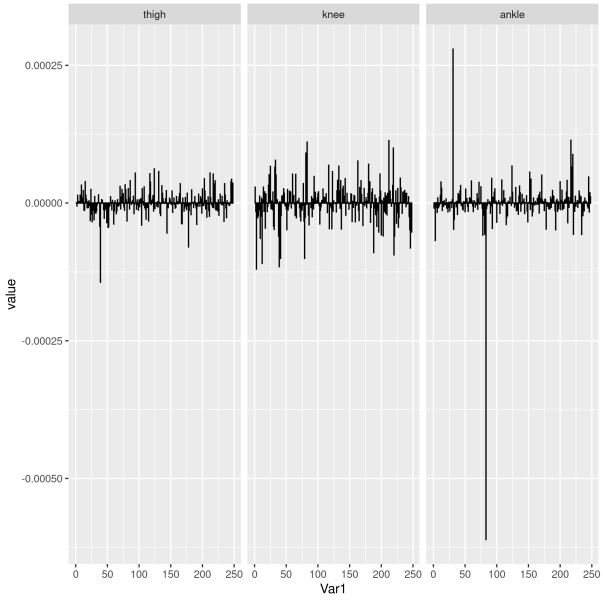

Figure 13: `DFBETA` for regressors `thigh`, `knee`, and `ankle`.

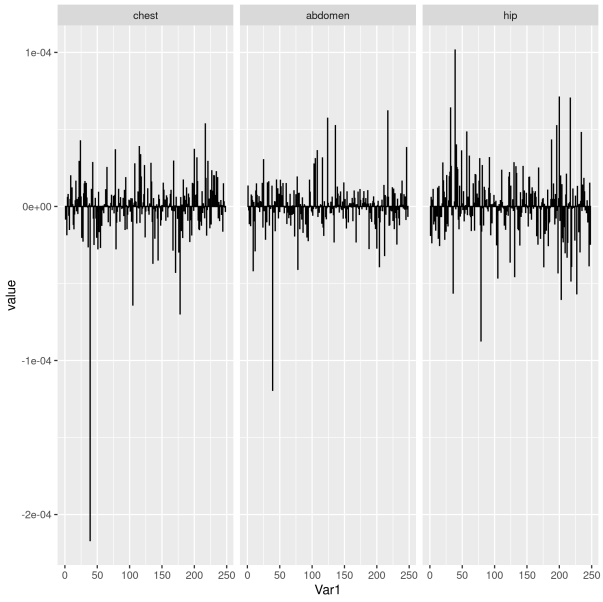

Figure 14: `DFBETA` for regressors `chest`, `abdomen`, and `hip`.