I fit a model regressing glucose concentration levels on the number of pregnancies a woman had, their diastolic blood pressure, the thickness of skin folds, BMI, and age.

Chart

Description automatically generated

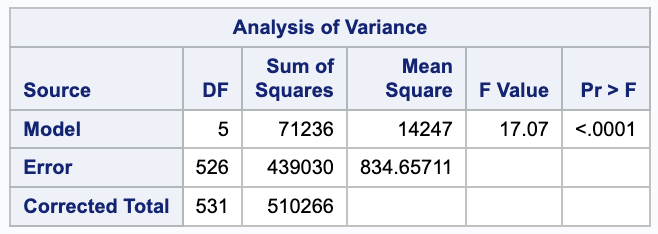
There is not much evidence of heteroscedasticity in the residuals vs predicted values plot. The qq-plot follows a 45-degree line for the most part, as well. There seems to be 3 very small residuals and 1 very large residual, relative to what would be expected from a normal distribution.

The residual versus X plots vary quite a bit, with numerous plots looking very heteroscedastic. I tried to take out outliers and refit my regression, but I was unable to solve this problem.

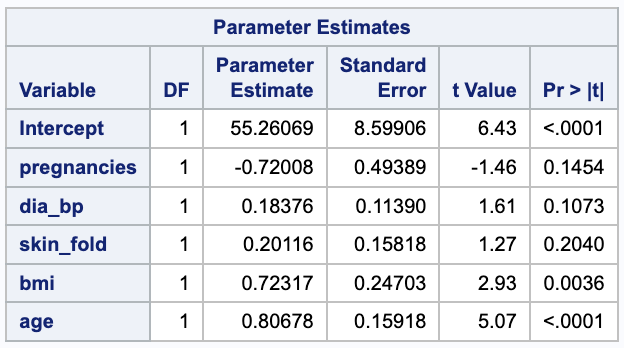
Scatter chart

Description automatically generated

This is the ANOVA table for the model taken as a whole. It is significant because p < 0.0001. However, because of how the residuals look, this must be taken under consideration and re-evaluated, as the heteroscedasticity can overestimate the significance of the model.



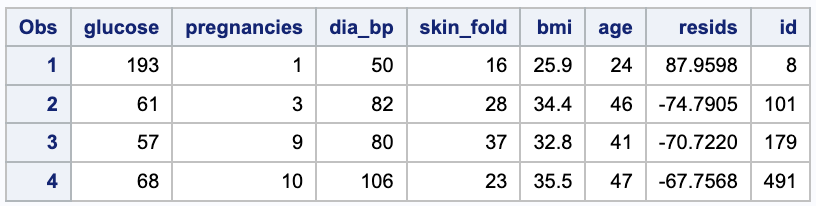
Here are the parameter estimates tables for our primary model, and the sensitivity analysis that   
constituted rerunning the model without the four potential outliers.



Table

Description automatically generated

The data for the four potential outliers are listed below:



We fit a model regressing glucose concentration levels on the number of pregnancies a woman had, their diastolic blood pressure, the thickness of skin folds, BMI, and age. The number of pregnancies was not significantly related to glucose concentration levels (β=-0.72, SE=0.49, p<0.1454), nor was diastolic blood pressure (β=0.18, SE=0.11, p<0.1073). The only data points that were significant were the intercept (β=55.26, SE=8.6, p<0.0001), although that was not of interest in our study, and age (β=0.81, SE=0.16, p<0.0001). We noticed four participants whose data did not fit the pattern seen in the rest of the data. However, their residuals did not appear to be very influential as measured by Cook’s D. Nonetheless, as a sensitivity analysis, we fit a second model without those data. The pattern of significances did not change and the parameters of that model were less than one standard error from the parameters of our initial model.