regression models – HW 2

chapter 4-5 excercises

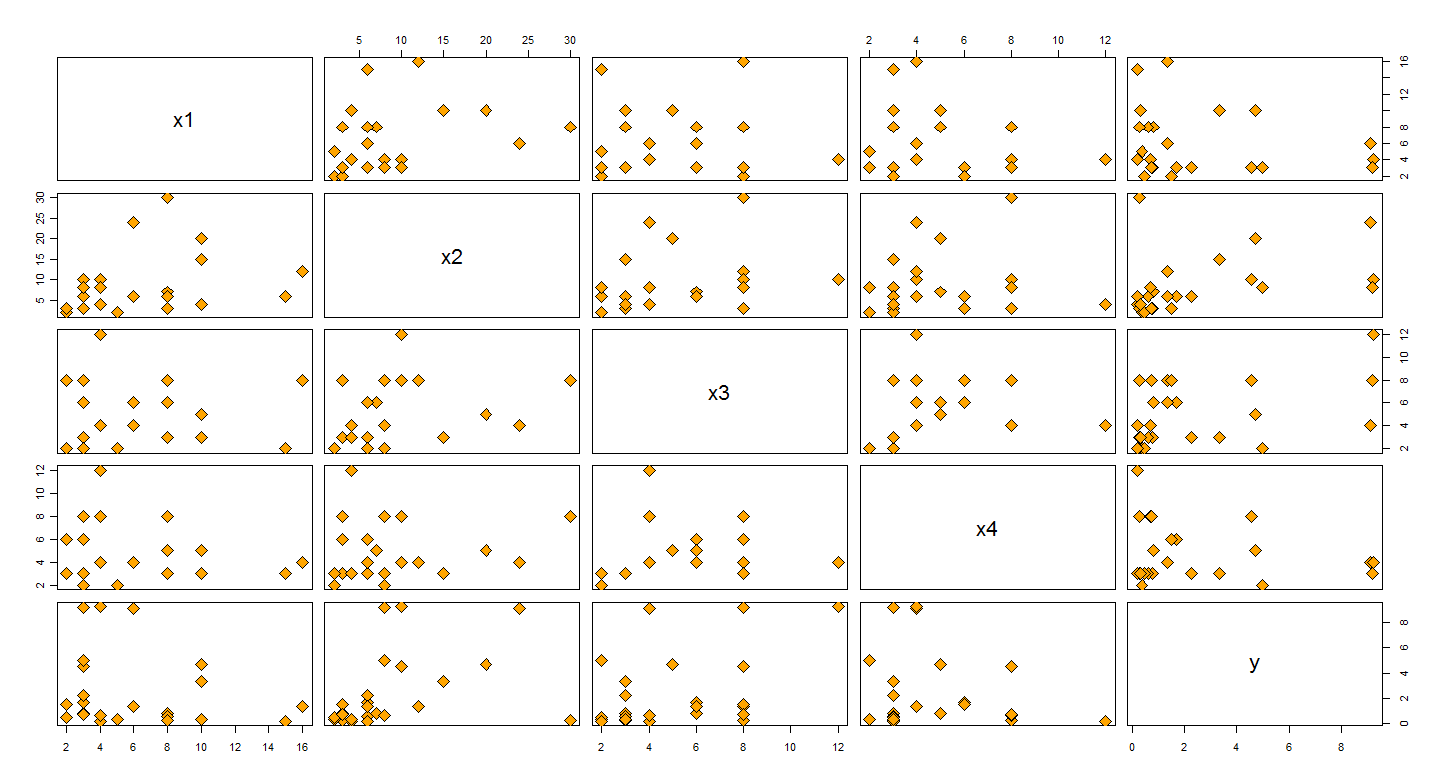
**Thanh Doan – Student ID 0159701**

|  |  |
| --- | --- |
| exercise 4.22 |  |

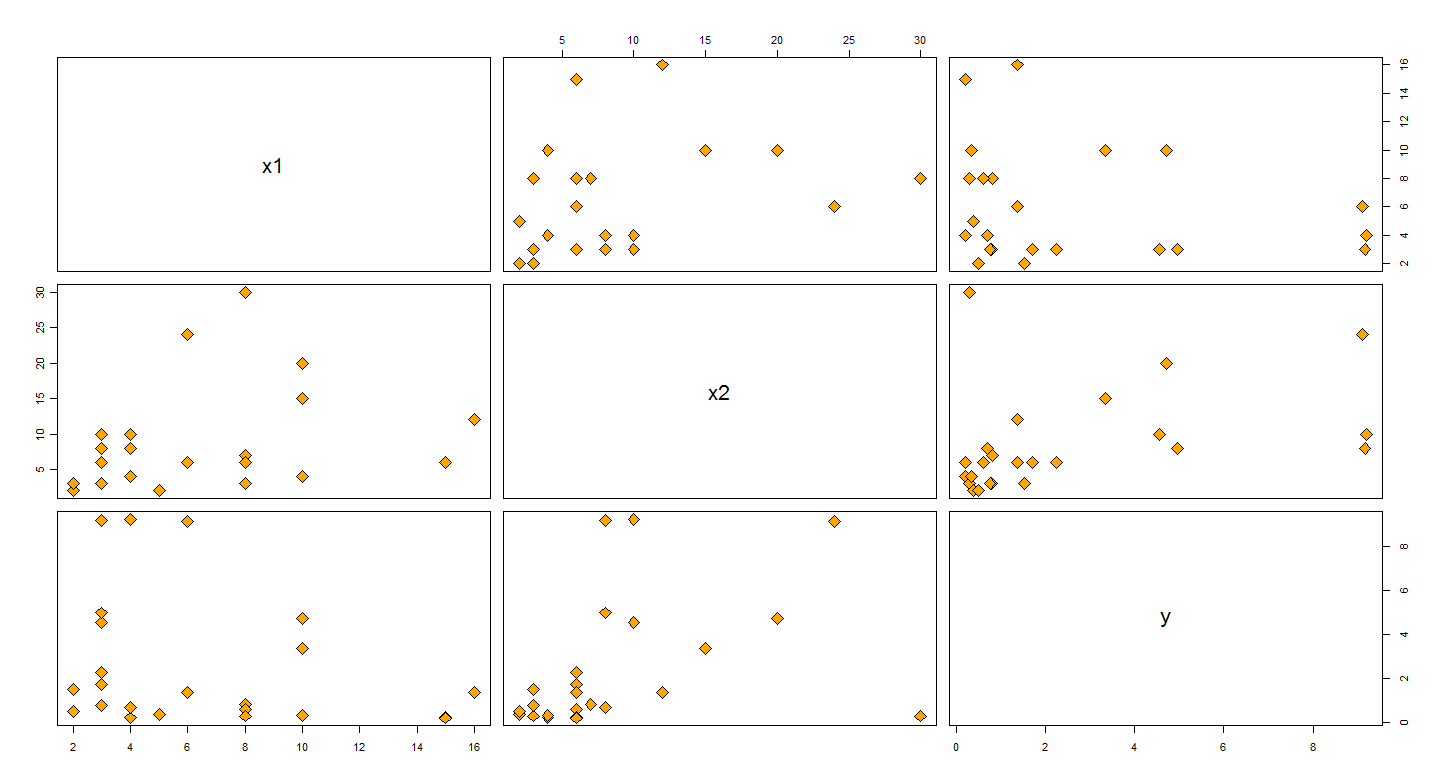
**a - Fit a multiple regression model**

and investigate the adequacy of the model

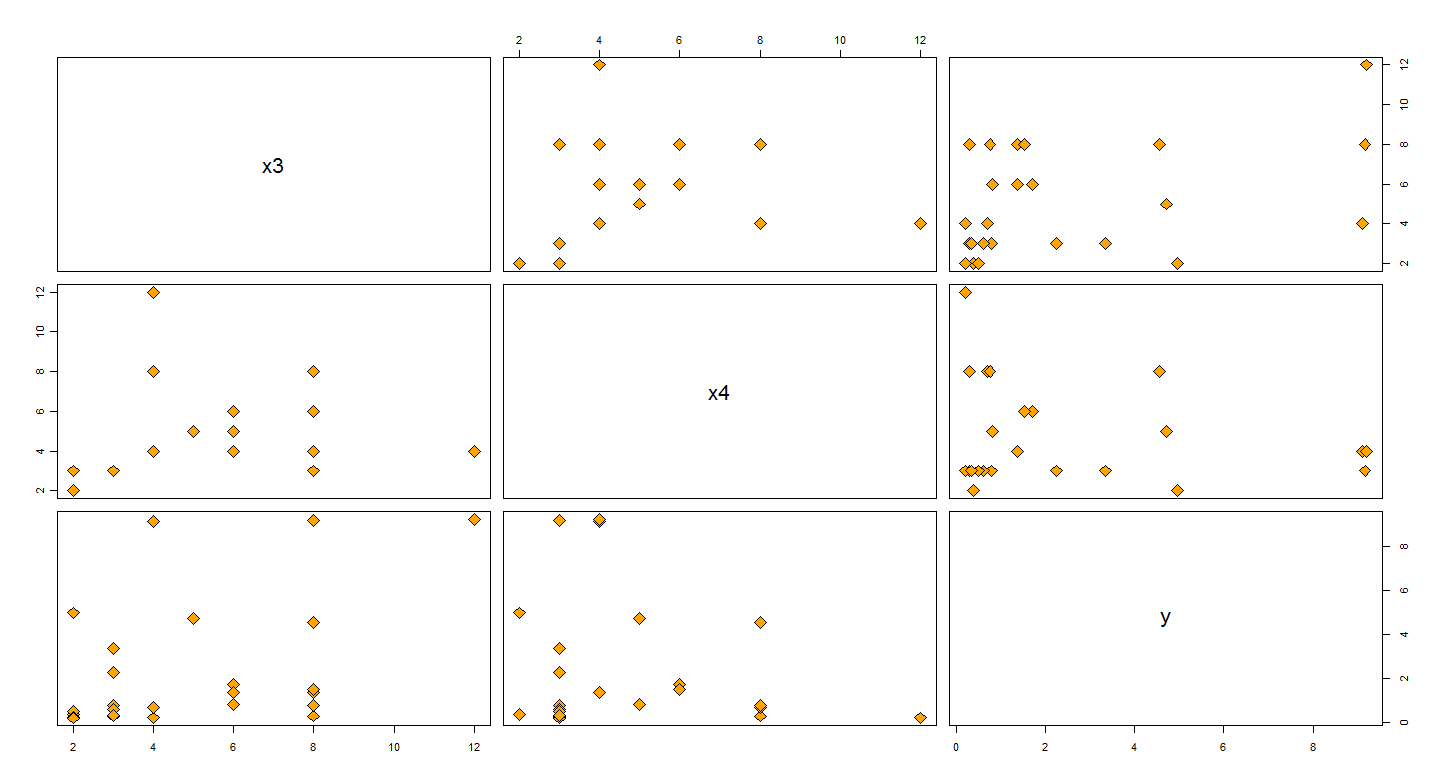
* Plot the scatter plot to have a bird-eye view on the relationship between response and explanatory variables



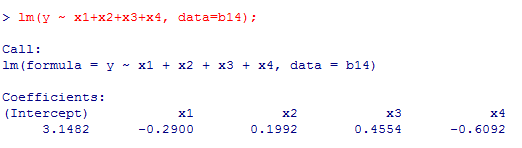
* Let us zoom in by creati8ng a scatter plot between response and x1, x2 variables



* Scatter plot between response and x3, x4 variables



* Fit a multiple regression model



* Normal Probability Plot of the residuals

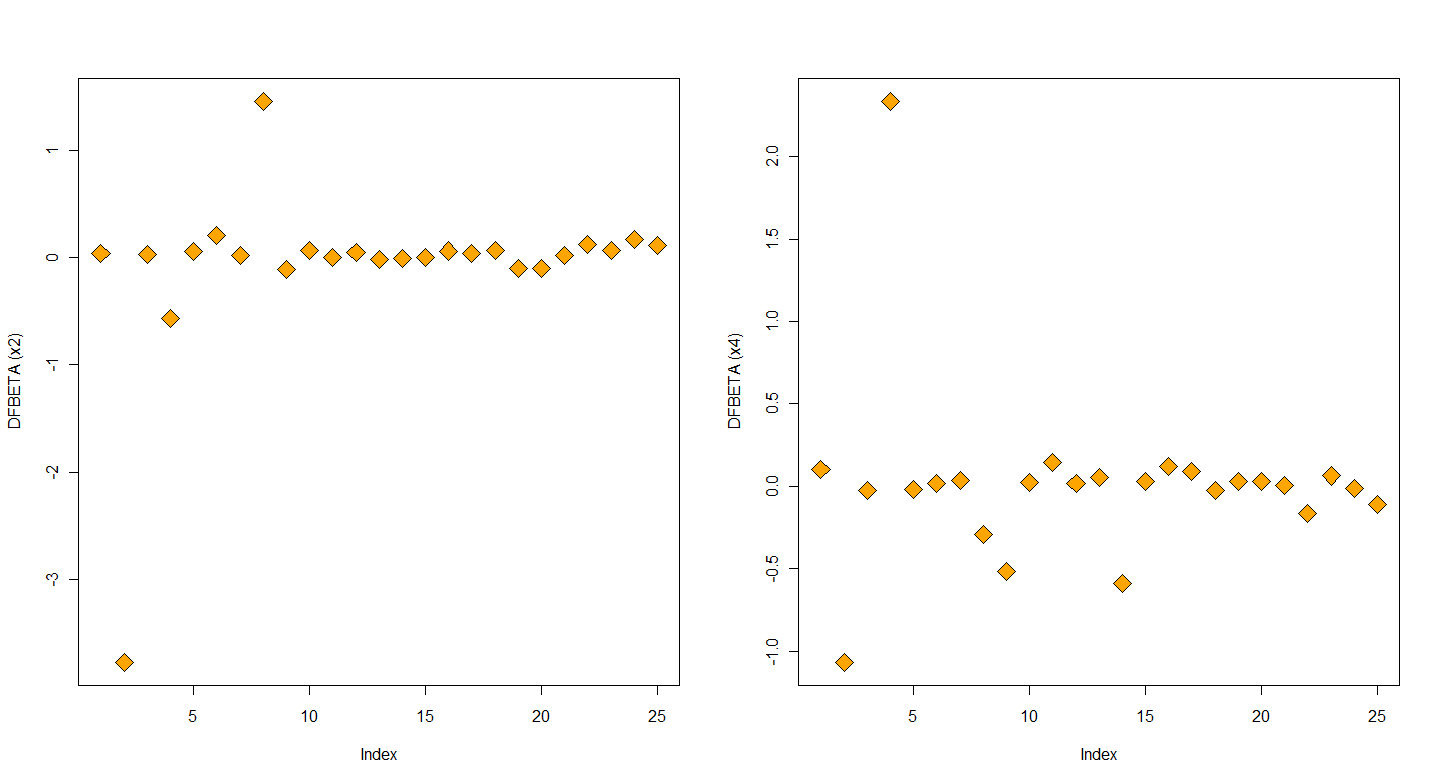


* Plot of the residuals against fitted values

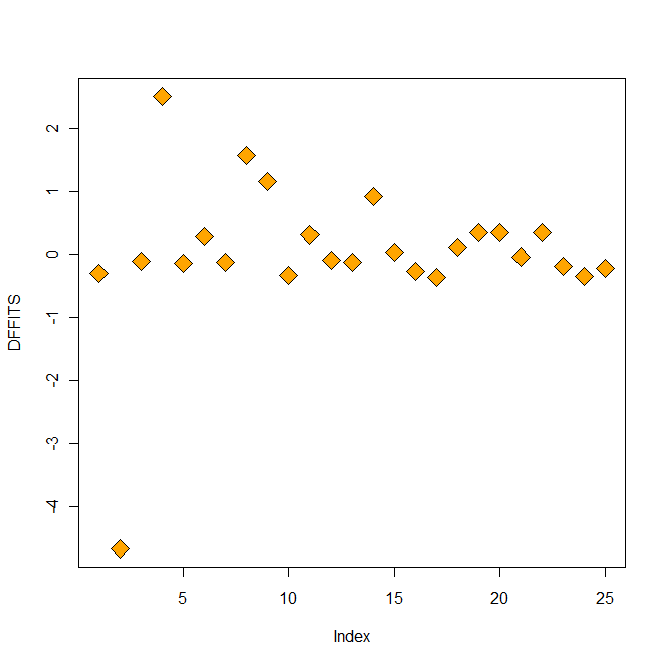




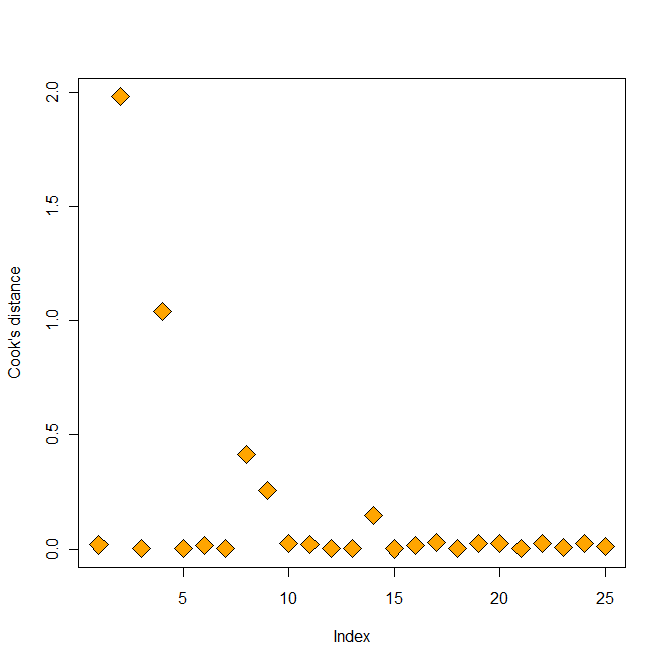
* DFBETA Plots



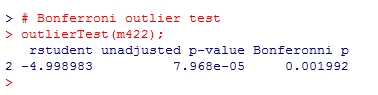
* DFFITS Plot



* Cook’s distance Plot



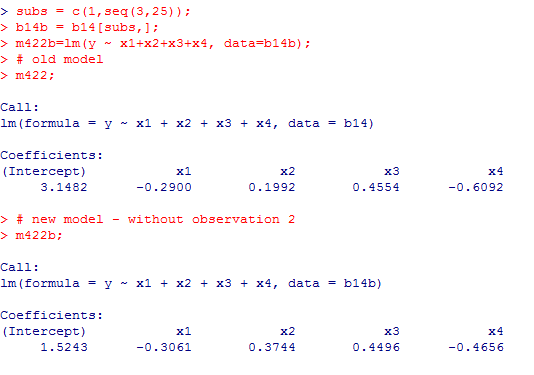
* Bonferroni outlier test



**Comment:**

**Diagnostic plots & Bonferroni outlier test suggest *Observation 2* is a potential outlier and the model does not fit very well**

**b – Delete observation 2 and refit the model, perform residual analysis.**

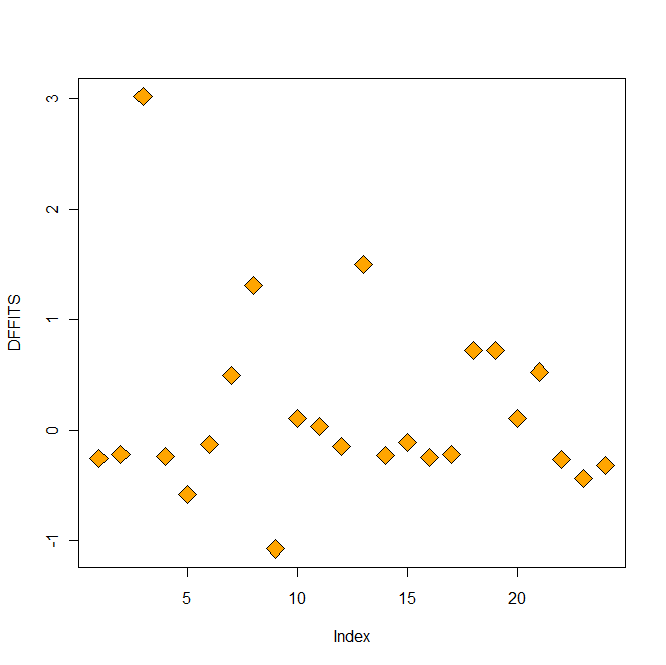
****

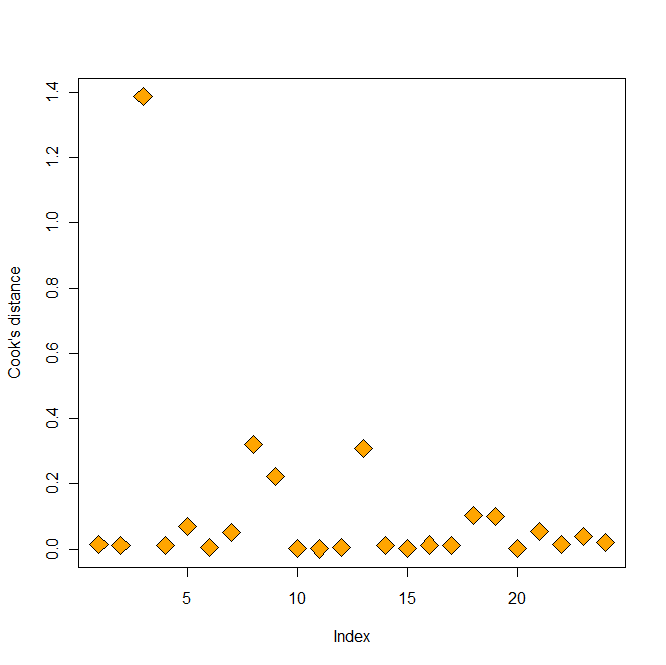
****

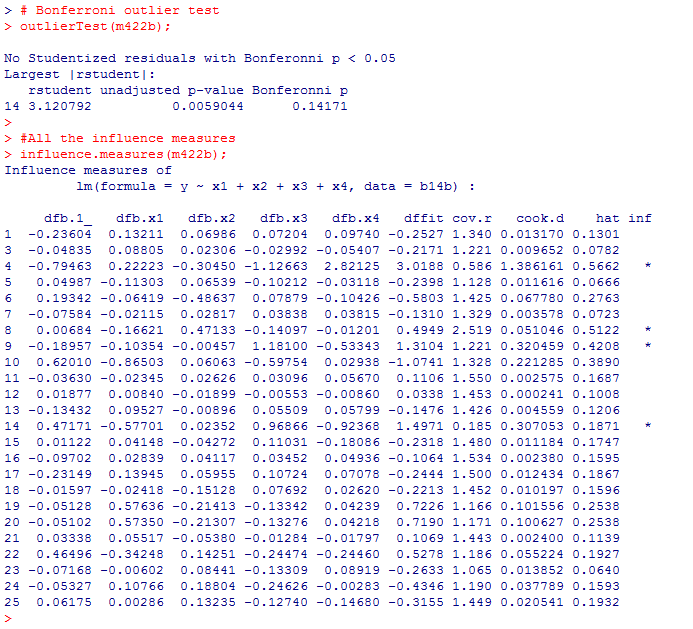
****

****

****

****

****

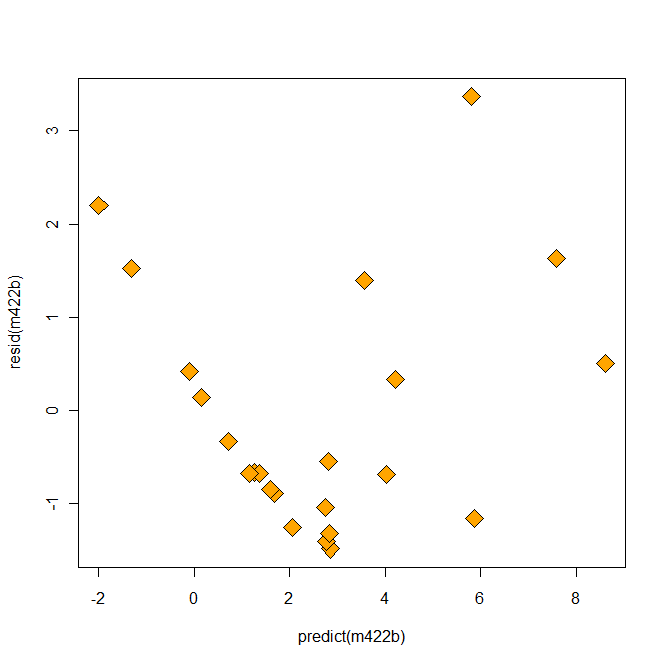
****

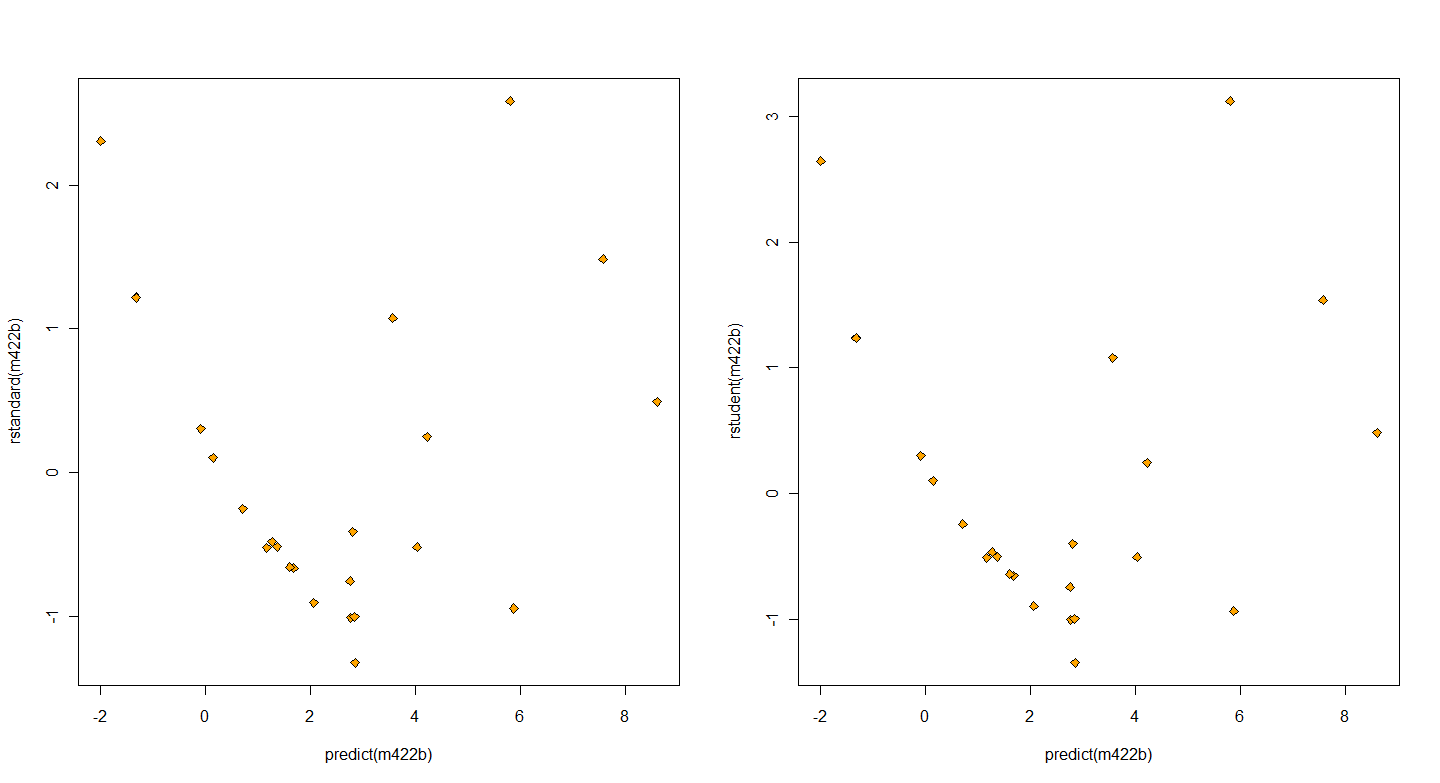
**Comment:**

**Residual analysis plots and other diagnostic plot suggest the model still does not fit very well. Residuals still do not follow normality and there seems to be non linear patterns. Observations 4, 8, 9, 14 seem to have high influences.**

|  |  |
| --- | --- |
| exercise 5.16 |  |

**a – Plot the original residuals, the studentized residuals and R-student vs the predicted response**



****

**Comment:**

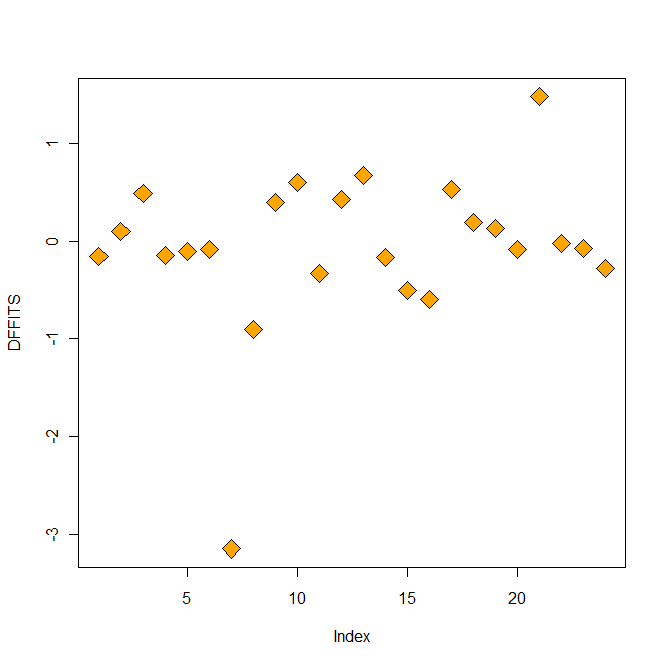
**Patterns for residual plots are not satisfactory. There is a non-linear pattern to the residuals.**

**b – From the (somewhat) parabola shape in the residual plot, let us try to transform the response y using log transformation**









**c –using partial regression or partial residuals plots to aid the study of suitable transformation on both response and repressors variables.**









**Now using log transformation on each of repressors and also use log transformation on response variable.**







**Comment:**

**After using log transformation both repressors and response variables, the final model looks satisfactory.**