Statistic Model to Analyze Student's Performance - Group 8

Xuchuan Zheng, Sungki Park, Prashant Sharma

02 April, 2024

1. ***Introduction***

In out final project for Data 603 - Statistical Modelling with Data, we have tried to develop a model to analyze the impact of various demographic and social factors on the performance of students. Academic performance, though it is not the only factor but is one of the crucial factors in shaping a student's future. To get into a good collage/university, student must score grades in school, a good college ca lead a better future and economic stability. So in order to secure good grades, getting into a great school is enough? Is there something more than a great school that can help a student to perform better? Do the social and demographic factors plays any role in student's performance? In our project we are trying to answer these question.

To answer these questions we are working with a dataset that is collected at 2 Portuguese schools for Mathematics and Portuguese subject. This data is collected by using school reports and questionnaires. The data attribute include students grades, family size information, education level of parents, free time of student, any many other factors. By working on this project we are hoping to develop more understanding about the factors which can impact the performance of a student.

2. ***Data***

This data is from UC Irvine Machine Learning Repository. There are 649 rows instances and 30 features in the dataset. Below are details of each feature

1. school - student's school (binary: 'GP' - Gabriel Pereira or 'MS' - Mousinho da Silveira) [Qualitative]
2. sex - student's sex (binary: 'F' - female or 'M' - male) [Qualitative]
3. age - student's age (numeric: from 15 to 22)
4. address - student's home address type (binary: 'U' - urban or 'R' - rural) [Qualitative]
5. famsize - family size (binary: 'LE3' - less or equal to 3 or 'GT3' - greater than 3) [Qualitative]
6. Pstatus - parent's cohabitation status (binary: 'T' - living together or 'A' - apart) [Qualitative]
7. Medu - mother's education (numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education) [Qualitative]
8. Fedu - father's education (numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education) [Qualitative]
9. Mjob - mother's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other') [Qualitative]

10. Fjob - father's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other') [Qualitative]
11. reason - reason to choose this school (nominal: close to 'home', school 'reputation', 'course' preference or 'other') [Qualitative]
12. guardian - student's guardian (nominal: 'mother', 'father' or 'other') [Qualitative]
13. traveltime - home to school travel time (numeric: 1 - <15 min., 2 - 15 to 30 min., 3 - 30 min. to 1 hour, or 4 - >1 hour) [Qualitative]
14. studytime - weekly study time (numeric: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours) [Qualitative]
15. failures - number of past class failures (numeric: n if 1<=n<3, else 4) [Qualitative]
16. schoolsup - extra educational support (binary: yes or no) [Qualitative]
17. famsup - family educational support (binary: yes or no) [Qualitative]
18. paid - extra paid classes within the course subject (Math or Portuguese) (binary: yes or no) [Qualitative]
19. activities - extra-curricular activities (binary: yes or no) [Qualitative]
20. nursery - attended nursery school (binary: yes or no) [Qualitative]
21. higher - wants to take higher education (binary: yes or no) [Qualitative]
22. internet - Internet access at home (binary: yes or no) [Qualitative]
23. romantic - with a romantic relationship (binary: yes or no) [Qualitative]
24. famrel - quality of family relationships (numeric: from 1 - very bad to 5 - excellent) [Qualitative]
25. freetime - free time after school (numeric: from 1 - very low to 5 - very high) [Qualitative]
26. goout - going out with friends (numeric: from 1 - very low to 5 - very high) [Qualitative]
27. Dalc - workday alcohol consumption (numeric: from 1 - very low to 5 - very high) [Qualitative]
28. Walc - weekend alcohol consumption (numeric: from 1 - very low to 5 - very high) [Qualitative]
29. health - current health status (numeric: from 1 - very bad to 5 - very good) [Qualitative]
30. absences - number of school absences (numeric: from 0 to 93) [Quantitative]
31. G1 - first period grade (numeric: from 0 to 20) [Quantitative]
32. G2 - second period grade (numeric: from 0 to 20) [Quantitative]
33. G3 - final grade (numeric: from 0 to 20, output target) [Quantitative]

G3 (Final grade) is the dependent variable for our model.

NOTE: We need to convert the qualitative variable from numeric to string

3. *Methodology*

Below is the outline of the steps we are going to perform in our analysis:

1. First Build a full additive model.
2. We will apply some model selection technique to come up with the best additive model.
3. Based on p-value (assuming $\alpha = 0.05$) we will drop variable which are non-significant.

4. Perform partial F-test to verify that dropped variables are indeed non-significant.
5. Provide interpretation for the best additive model to predict our dependent variable (G3 - Final Grades).
6. Based on our best additive model, we will check for interaction between the variables.
7. Using p-value (assuming $\alpha = 0.05$), we will drop the non-significant interaction terms.
8. Use partial F-test and analysis of Variance to verify the usability of our best interaction model.
9. Provide interpretation of our best interaction model.
10. Then we will check if we can include any higher order term in our model (Moving towards Higher order multiple regression model).
11. Verify the significance of higher order terms using p-value (assuming $\alpha = 0.05$).
12. Once we have done all our analysis we will try to define our best regression model (liner or higher order) to predict our dependent variable (G3 - Final Grades).
13. Using our final regression model, we will start checking the regression assumptions.
14. Linearity Assumption.
15. Independence Assumption.
16. Normality Assumption.
17. Multi-collinearity.
18. Outliers.

Starting our analysis with building additive model, then we will try to include interaction terms and higher order terms in our model.

3.1 *Full Additive Model*

Creating full additive model:

```
studentPerformance_fm = lm(G3 ~(school+sex+age+address+famsize+Pstatus+Medu+
                                  Fedu+Mjob+Fjob+reason+guardian+traveltime+studytime+
                                  failures+schoolsup+famsup+activities+nursery+higher+internet+
                                  romantic+famrel+freetime+
                                  goout+Dalc+Walc+health+absences),
                           data = studentDataset)

#summary(studentPerformance_fm)
```

> **NOTE:** Refer Appendix 1 for summary of full addtive model

Comments: From the summary of the full model we can see that many of the variables are non-significant. We can apply some techniques for model selection to get significant parameters.

3.2 *Model Selection*

Using Stepwise forward selection procedure to get the significant parameters:

```
studentPerformance_Forward_subsets = ols_step_forward_p(studentPerformance_fm,p_val = 0.05, details = FA
student_performance_forwardMdl = studentPerformance_Forward_subsets$model
summary(student_performance_forwardMdl)
```

```
##
## Call:
## lm(formula = paste(response, "~", paste(preds, collapse = " + ")),
##     data = l)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.0125  -1.4391   0.3944   2.1255   8.2166
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 11.70133    1.38394   8.455  < 2e-16 ***
## failures     -1.76047    0.17299 -10.177  < 2e-16 ***
## higheryes     1.20109    0.41419   2.900  0.00381 **
## schoolsupyes -1.08027    0.34127  -3.165  0.00159 **
## schoolMS     -0.52091    0.27020  -1.928  0.05415 .
## romanticyes  -0.57404    0.22855  -2.512  0.01217 *
## studytime2    0.33184    0.26152   1.269  0.20476
## studytime3    1.19724    0.35335   3.388  0.00073 ***
## studytime4    0.53768    0.48696   1.104  0.26979
## addressU      0.45472    0.25698   1.769  0.07712 .
## Medu1        -1.68348    1.17642  -1.431  0.15273
## Medu2        -1.64019    1.16935  -1.403  0.16103
## Medu3        -1.22194    1.17050  -1.044  0.29676
## Medu4        -0.73276    1.16900  -0.627  0.53092
## Dalc2        -0.75785    0.28374  -2.671  0.00769 **
## Dalc3        -0.05715    0.45502  -0.126  0.90007
## Dalc4        -1.74603    0.68863  -2.536  0.01138 *
## Dalc5         0.04561    0.70762   0.064  0.94862
## famsizeLE3    0.51236    0.23591   2.172  0.03010 *
## freetime2     0.84199    0.50684   1.661  0.09697 .
## freetime3    -0.15376    0.46711  -0.329  0.74208
## freetime4     0.25650    0.48784   0.526  0.59915
## freetime5     0.87384    0.56062   1.559  0.11938
## goout2        1.20786    0.46546   2.595  0.00960 **
## goout3        0.82742    0.45858   1.804  0.07148 .
## goout4        0.40653    0.48356   0.841  0.40071
## goout5        0.01356    0.50561   0.027  0.97861
## health2      -0.71212    0.42745  -1.666  0.09603 .
## health3      -1.11917    0.37841  -2.958  0.00317 **
## health4      -0.62825    0.39712  -1.582  0.11396
## health5      -1.03268    0.34383  -3.003  0.00273 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 3.392 on 1013 degrees of freedom
## Multiple R-squared:  0.252,  Adjusted R-squared:  0.2298
## F-statistic: 11.38 on 30 and 1013 DF,  p-value: < 2.2e-16
```

```r
#studentPerformance_Forward_subsets$metric
```

Using the result of forward selection process and dropping non-significant variables, the best additive model is:

```r
student_performance_addMdl = lm(G3 ~ (failures+higher+studytime+schoolsup+Dalc+health+romantic+famsize+
                                      +goout), data = studentDataset)
summary(student_performance_addMdl)
```

```
##
## Call:
## lm(formula = G3 ~ (failures + higher + studytime + schoolsup +
##     Dalc + health + romantic + famsize + +goout), data = studentDataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.1848  -1.5137   0.3029   2.1295   8.6106
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10.68826    0.63642  16.794  < 2e-16 ***
## failures    -1.82887    0.17261 -10.595  < 2e-16 ***
## higheryes    1.54720    0.41456   3.732 0.000200 ***
## studytime2   0.26960    0.26170   1.030 0.303179
## studytime3   1.18603    0.35442   3.346 0.000849 ***
## studytime4   0.62526    0.49164   1.272 0.203738
## schoolsupyes -1.16127    0.34238  -3.392 0.000721 ***
## Dalc2       -0.73721    0.28608  -2.577 0.010107 *
## Dalc3       -0.14301    0.45454  -0.315 0.753108
## Dalc4       -1.65491    0.69909  -2.367 0.018107 *
## Dalc5        0.11089    0.71654   0.155 0.877041
## health2     -0.66440    0.43152  -1.540 0.123946
## health3     -1.21726    0.38251  -3.182 0.001505 **
## health4     -0.74820    0.40008  -1.870 0.061751 .
## health5     -0.99991    0.34699  -2.882 0.004038 **
## romanticyes -0.59306    0.23040  -2.574 0.010190 *
## famsizeLE3   0.48561    0.23806   2.040 0.041625 *
## goout2       1.15597    0.46996   2.460 0.014070 *
## goout3       0.67978    0.45855   1.482 0.138527
## goout4       0.38442    0.48055   0.800 0.423919
## goout5       0.08698    0.50253   0.173 0.862617
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.449 on 1023 degrees of freedom
## Multiple R-squared:  0.2189, Adjusted R-squared:  0.2036
## F-statistic: 14.33 on 20 and 1023 DF,  p-value: < 2.2e-16
```

Our best additive model:

$$\widehat{G3} = 10.68826\text{-}1.82887(\text{failures})+1.54720(\text{higheryes})+0.26960(\text{studytime2})+1.18603(\text{studytime3})$$
$$+0.62526(\text{studytime4})\text{-}1.16127(\text{schoolsupyes})\text{-}0.73721(\text{Dalc2})\text{-}0.14301(\text{Dalc3})$$
$$\text{-}1.65491(\text{Dalc4})+0.11089(\text{Dalc5})\text{-}0.66440(\text{health2})\text{-}1.21726(\text{health3})\text{-}0.74820(\text{health4})$$
$$\text{-}0.99991(\text{health5})\text{-}0.59306(\text{romanticyes})+0.48561(\text{famsizeLE3})+1.15597(\text{goout2})+0.67978(\text{goout3})+0.38442(\text{goout4})$$
$$+0.08698(\text{goout5})$$

Add some comment

3.3 ***Interaction Model***

Using our best additive model, we can check for interaction term:

```
student_performance_intMdl1 = lm(G3 ~ (failures+higher+studytime+schoolsup+Dalc+health+romantic+famsize
                                       +goout)^2, data = studentDataset)
#summary(student_performance_intMdl1)
```

NOTE: Refer Appendix 2 for summary of the interaction model

In the summary of our full interaction model the interaction term *Dacl:goout and studytime:Dacl* has some NA entries, it simply mean that there is not enough data to generate values for those interaction. So we will drop these interaction completely.

Our interaction model will be:

```
student_performance_intMdl2 = lm(G3 ~ (failures+higher+studytime+schoolsup+Dalc+health+romantic+famsize
                                       +goout+failures:schoolsup+studytime:health+studytime:goout
                                       +schoolsup:health+schoolsup:goout),
                                 data = studentDataset)
#summary(student_performance_intMdl2)
```

NOTE: Refer Appendix 3 for summary of final interaction model.

3.4 ***Comparing Additive and Interaction model***

Comparing our additive and interaction models to see which one is better. To verify this we can use partial F-test using below hypothesis:

H_0 : All the interaction coefficients are not significant H_a : At least one of the interaction coefficient is significant

```
## Analysis of Variance Table
##
## Model 1: G3 ~ (failures + higher + studytime + schoolsup + Dalc + health +
##     romantic + famsize + +goout)
## Model 2: G3 ~ (failures + higher + studytime + schoolsup + Dalc + health +
##     romantic + famsize + +goout + failures:schoolsup + studytime:health +
##     studytime:goout + schoolsup:health + schoolsup:goout)
##   Res.Df   RSS Df Sum of Sq      F   Pr(>F)
## 1   1023 12169
## 2    990 11429 33    740.32 1.9433 0.001228 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

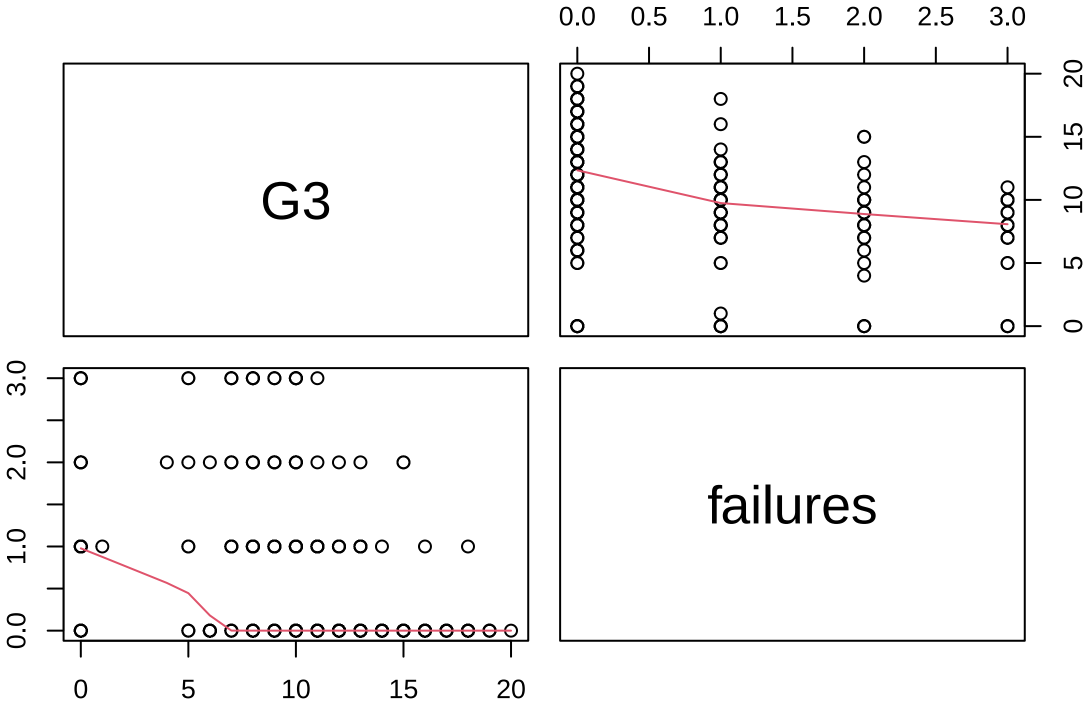
Source of variation	DF	Sum of Square	Mean Square	F-Statistic
Regression	33	740.32	22.43394	1.9433
Residual	990	11429	11.54444	
Total	1023	12169		

The p-value from our ANOVA test is **0.001228** which is less than our assumed $\alpha = 0.05$ hence we can reject our H_0 and conclude that at least one of our interaction coefficient is significant, and

3.5. *Higher Order Model*

Since we have finalized our interaction model, we can check if there are any variable (quantitative variable) for which we need to add higher order terms. For this analysis we will use pairs plot. In our final interaction model we have only one quantitative variable (i.e. failures).

```r
pairs(~G3 + failures, data = studentDataset, panel = panel.smooth)
```



Though there no clear visual indication that we should add higher order term for our variable, but we can still try by adding the higher order term and check the significance.

```
student_performance_intMdl3 = lm(G3 ~ (failures+I(failures^2)+higher+studytime+schoolsup+Dalc+health+ro
                                  +goout+failures:schoolsup+studytime:health+studytime:goout
                                  +schoolsup:health+schoolsup:goout),
                             data = studentDataset)
summary(student_performance_intMdl3)
```

```
##
## Call:
## lm(formula = G3 ~ (failures + I(failures^2) + higher + studytime +
##     schoolsup + Dalc + health + romantic + famsize + +goout +
##     failures:schoolsup + studytime:health + studytime:goout +
##     schoolsup:health + schoolsup:goout), data = studentDataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.5025  -1.5588   0.3212   2.1742   8.6013
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)             8.6707     0.9755   8.888  < 2e-16 ***
## failures               -3.8453     0.4898  -7.851 1.07e-14 ***
## I(failures^2)           0.7576     0.1914   3.958 8.09e-05 ***
## higheryes               1.4940     0.4105   3.639 0.000288 ***
## studytime2              2.5596     1.1200   2.285 0.022504 *
## studytime3              7.4010     2.1062   3.514 0.000462 ***
## studytime4             -1.7682     2.2031  -0.803 0.422394
```

```
## schoolsupyes           0.4547     1.4882   0.306 0.760012
## Dalc2                 -0.7863     0.2884  -2.727 0.006511 **
## Dalc3                  0.1068     0.4559   0.234 0.814862
## Dalc4                 -1.3768     0.6970  -1.975 0.048525 *
## Dalc5                  0.2097     0.7405   0.283 0.777071
## health2                0.7798     0.8517   0.916 0.360150
## health3               -0.9041     0.7513  -1.203 0.229152
## health4                0.2467     0.7676   0.321 0.747981
## health5               -0.4384     0.6495  -0.675 0.499822
## romanticyes           -0.5402     0.2329  -2.319 0.020589 *
## famsizeLE3             0.5953     0.2386   2.495 0.012744 *
## goout2                 2.6289     0.8202   3.205 0.001393 **
## goout3                 2.0995     0.8009   2.622 0.008886 **
## goout4                 2.3283     0.8351   2.788 0.005404 **
## goout5                 1.7670     0.8432   2.096 0.036370 *
## failures:schoolsupyes   1.4629     0.5276   2.773 0.005665 **
## studytime2:health2    -2.2138     1.0147  -2.182 0.029361 *
## studytime3:health2    -1.4987     1.6251  -0.922 0.356618
## studytime4:health2     0.2277     2.0525   0.111 0.911696
## studytime2:health3    -0.5703     0.9098  -0.627 0.530916
## studytime3:health3    -1.9426     1.4285  -1.360 0.174180
## studytime4:health3    -1.1523     1.5514  -0.743 0.457822
## studytime2:health4    -1.4005     0.9384  -1.492 0.135926
## studytime3:health4    -2.7377     1.4432  -1.897 0.058132 .
## studytime4:health4    -2.1322     2.0521  -1.039 0.299055
## studytime2:health5    -0.6203     0.7963  -0.779 0.436179
## studytime3:health5    -2.3509     1.3492  -1.742 0.081736 .
## studytime4:health5    -0.2144     1.5684  -0.137 0.891296
## studytime2:goout2     -1.1239     1.0541  -1.066 0.286582
## studytime3:goout2     -4.2356     1.8065  -2.345 0.019244 *
## studytime4:goout2      4.9803     2.1777   2.287 0.022411 *
## studytime2:goout3     -1.3213     1.0263  -1.287 0.198229
## studytime3:goout3     -3.9000     1.7812  -2.190 0.028790 *
## studytime4:goout3      3.3769     2.2263   1.517 0.129630
## studytime2:goout4     -1.8888     1.0534  -1.793 0.073252 .
## studytime3:goout4     -5.1949     1.9109  -2.719 0.006670 **
## studytime4:goout4      1.8847     2.7072   0.696 0.486475
## studytime2:goout5     -1.7584     1.0871  -1.618 0.106088
## studytime3:goout5     -3.2727     2.0008  -1.636 0.102214
## studytime4:goout5      2.3971     2.1921   1.094 0.274432
## schoolsupyes:health2   2.3211     1.4139   1.642 0.100989
## schoolsupyes:health3   2.3842     1.2604   1.892 0.058844 .
## schoolsupyes:health4   2.9360     1.2744   2.304 0.021438 *
## schoolsupyes:health5   2.9259     1.1839   2.471 0.013625 *
## schoolsupyes:goout2   -5.5980     1.3439  -4.165 3.38e-05 ***
## schoolsupyes:goout3   -4.5044     1.3586  -3.316 0.000948 ***
## schoolsupyes:goout4   -3.8287     1.3607  -2.814 0.004993 **
## schoolsupyes:goout5   -5.5483     1.5527  -3.573 0.000370 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.373 on 989 degrees of freedom
## Multiple R-squared:  0.2778, Adjusted R-squared:  0.2384
## F-statistic: 7.047 on 54 and 989 DF,  p-value: < 2.2e-16
```

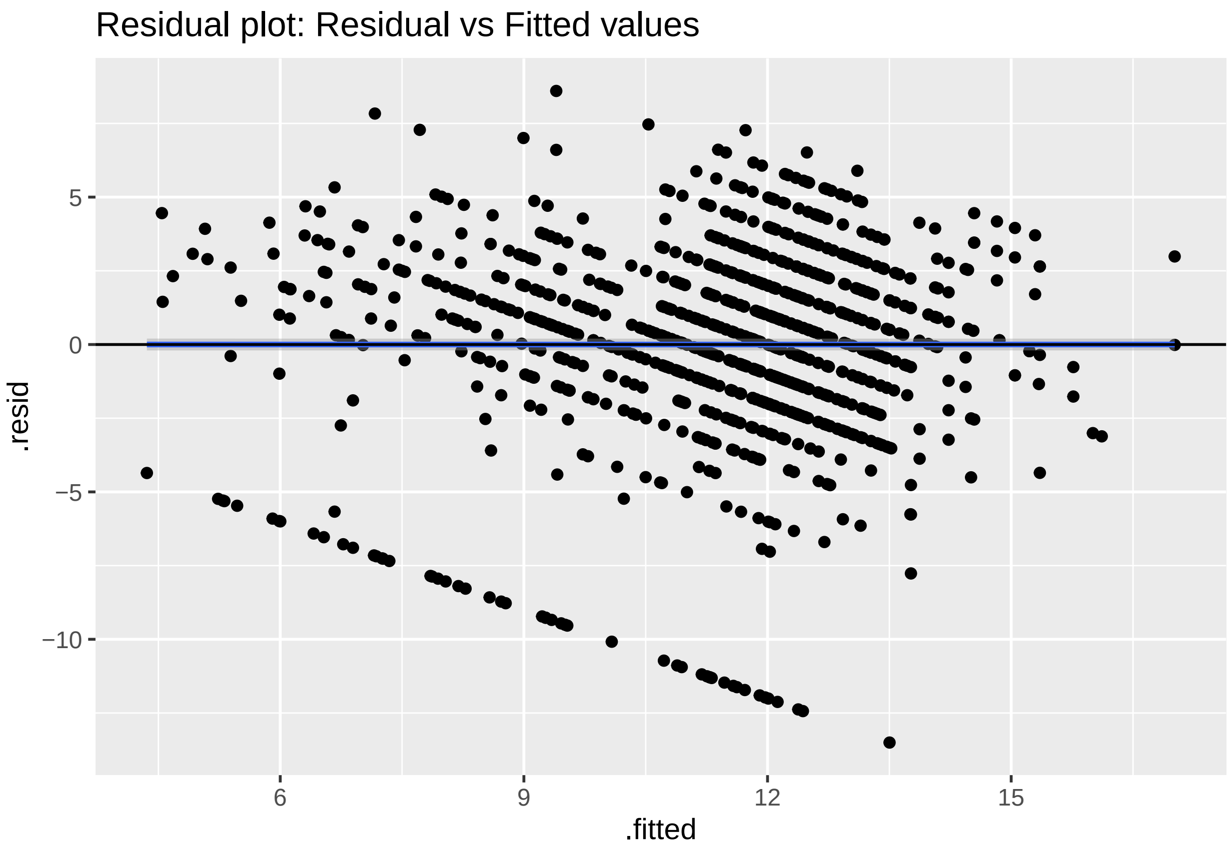
$$\widehat{G3} = 8.67068((\text{Intercept})) -3.84533(\text{failures}) +0.757609(\text{I(failures\^{}2)}) +1.49398(\text{higheryes})$$
$$+2.5596(\text{studytime2}) +7.40096(\text{studytime3}) -1.7682(\text{studytime4}) +0.454711(\text{schoolsupyes})$$
$$-0.786275(\text{Dalc2}) +0.106775(\text{Dalc3}) -1.37676(\text{Dalc4}) +0.209721(\text{Dalc5}) +0.779774(\text{health2})$$
$$-0.904056(\text{health3}) +0.246689(\text{health4}) -0.438419(\text{health5}) -0.540209(\text{romanticyes})$$
$$+0.595347(\text{famsizeLE3}) +2.62894(\text{goout2}) +2.09955(\text{goout3}) +2.32829(\text{goout4})$$
$$+1.767(\text{goout5}) +1.46286(\text{failures:schoolsupyes}) -2.21384(\text{studytime2:health2})$$
$$-1.49872(\text{studytime3:health2}) +0.227675(\text{studytime4:health2}) -0.570268(\text{studytime2:health3})$$
$$-1.94258(\text{studytime3:health3}) -1.15228(\text{studytime4:health3}) -1.40046(\text{studytime2:health4})$$
$$-2.73769(\text{studytime3:health4}) -2.13216(\text{studytime4:health4}) -0.62032(\text{studytime2:health5})$$
$$-2.35092(\text{studytime3:health5}) -0.214397(\text{studytime4:health5}) -1.12391(\text{studytime2:goout2})$$
$$-4.23562(\text{studytime3:goout2}) +4.98029(\text{studytime4:goout2}) -1.32132(\text{studytime2:goout3})$$
$$-3.90003(\text{studytime3:goout3}) +3.37694(\text{studytime4:goout3}) -1.88884(\text{studytime2:goout4})$$
$$-5.1949(\text{studytime3:goout4}) +1.88472(\text{studytime4:goout4}) -1.75839(\text{studytime2:goout5})$$
$$-3.27269(\text{studytime3:goout5}) +2.39711(\text{studytime4:goout5}) +2.3211(\text{schoolsupyes:health2})$$
$$+2.38419(\text{schoolsupyes:health3}) +2.93603(\text{schoolsupyes:health4}) +2.92593(\text{schoolsupyes:health5})$$
$$-5.59801(\text{schoolsupyes:goout2}) -4.50439(\text{schoolsupyes:goout3}) -3.8287(\text{schoolsupyes:goout4})$$
$$-5.54827(\text{schoolsupyes:goout5})$$

3.6 *Assumption Verification*

Our model building process is based on some assumption about the data. Hence it is very important to verify that these assumption are met or not. If our data do not fit the assumptions then the model we developed cannot be used for prediction.

3.6.1 Linearity Assumption

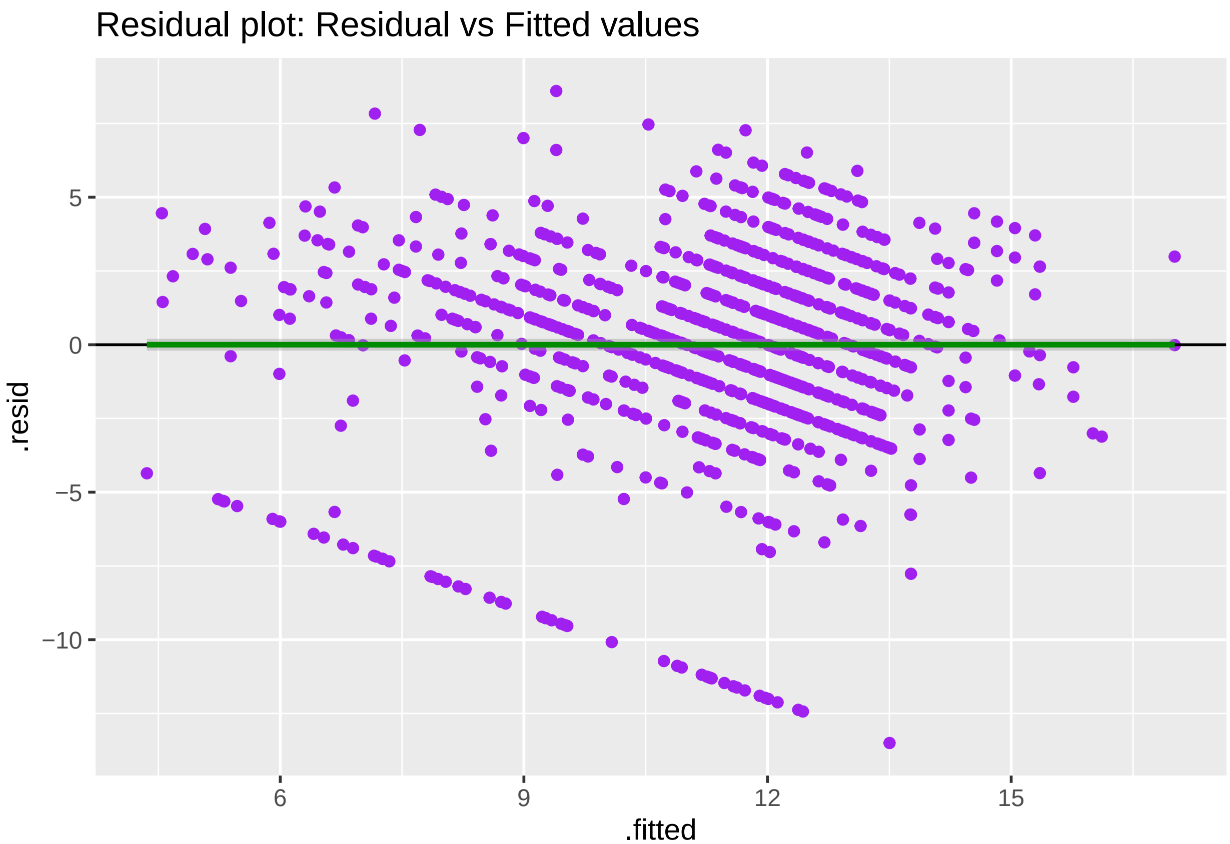
The linear regression model we build is based on the assumption that their is a linear relation between predictors and response variable. To confirm the linear relation we can use Residula plot.



Comment: The residual vs fitted plot show a straight line which indicate that our linearity assumption is met.

3.6.2 Equal Variance Assumption

Another important assumption for our liner regression model is that the error term has a constant variance (**homoscedasticity**). To verify the homoscedasticity assumption we can again use the residual vs fitted plot and check if there is any patter.



To confirm homoscedasticity we can perform the Breusch-Pagan Test (bptest) using below hypothesis:

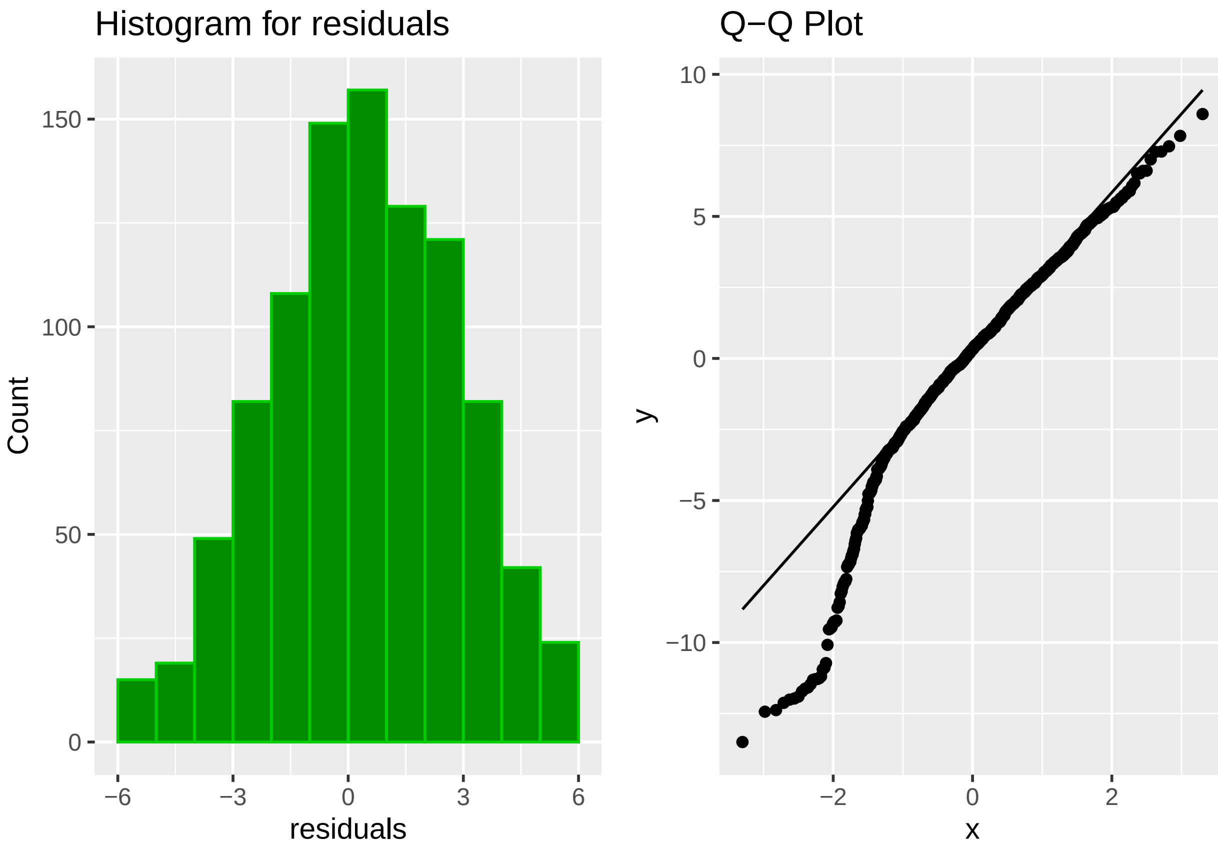
H_0 : Hetroscedasticity is not present (error term has comman variance) H_a : Hetroscedasticity is present (error term do not have comman variance)

```r
bptest(student_performance_intMdl3)  # It doesn't have heteroscedasticity.
```

```
##
##  studentized Breusch-Pagan test
##
## data:  student_performance_intMdl3
## BP = 67.517, df = 54, p-value = 0.1023
```

> Comment: Since the p-value is 0.3235 higher then our assumed $\alpha = 0.05$ we cannot reject the H_0 and conclude that our model meets the assumption of common variance.

3.6.3 Normality Assumption




```
shapiro.test(residuals(student_performance_intMdl3))  # We don't have normality
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(student_performance_intMdl3)
## W = 0.94218, p-value < 2.2e-16
```

3.6.4 Independence Assumption

Check if we need to add anything.

3.6.5 Multicollinearity

```
#finalMdl = lm(G3 ~ (failures+higher+studytime+schoolsup+Dalc+health+romantic+famsize+
#                                    +goout+failures:schoolsup+studytime:health+studytime:goout
#                                    +schoolsup:health+schoolsup:goout),
#                               data = studentDataset)

#finalMdl = lm(G3 ~ (failures+higher+schoolsup+Dalc+health+romantic+famsize+
#                                    +goout+failures:schoolsup+schoolsup:health+schoolsup:goout),
#                              data = studentDataset)

finalMdl = lm(G3 ~ (failures+higher+schoolsup+Dalc+health+romantic+famsize+
                                  +goout+studytime),
```

```r
                                    data = studentDataset)
imcdiag(finalMdl, method="VIF")
```

```
##
## Call:
## imcdiag(mod = finalMdl, method = "VIF")
##
##
##  VIF Multicollinearity Diagnostics
##
##                  VIF detection
## failures     1.1247         0
## higheryes    1.1762         0
## schoolsupyes 1.0391         0
## Dalc2        1.0953         0
## Dalc3        1.1192         0
## Dalc4        1.0416         0
## Dalc5        1.0943         0
## health2      1.6986         0
## health3      2.0999         0
## health4      1.9511         0
## health5      2.4854         0
## romanticyes  1.0672         0
## famsizeLE3   1.0306         0
## goout2       3.5109         0
## goout3       4.0215         0
## goout4       3.4486         0
## goout5       2.9202         0
## studytime2   1.5008         0
## studytime3   1.4453         0
## studytime4   1.1850         0
##
## NOTE:  VIF Method Failed to detect multicollinearity
##
##
## 0 --> COLLINEARITY is not detected by the test
##
## ===================================
```

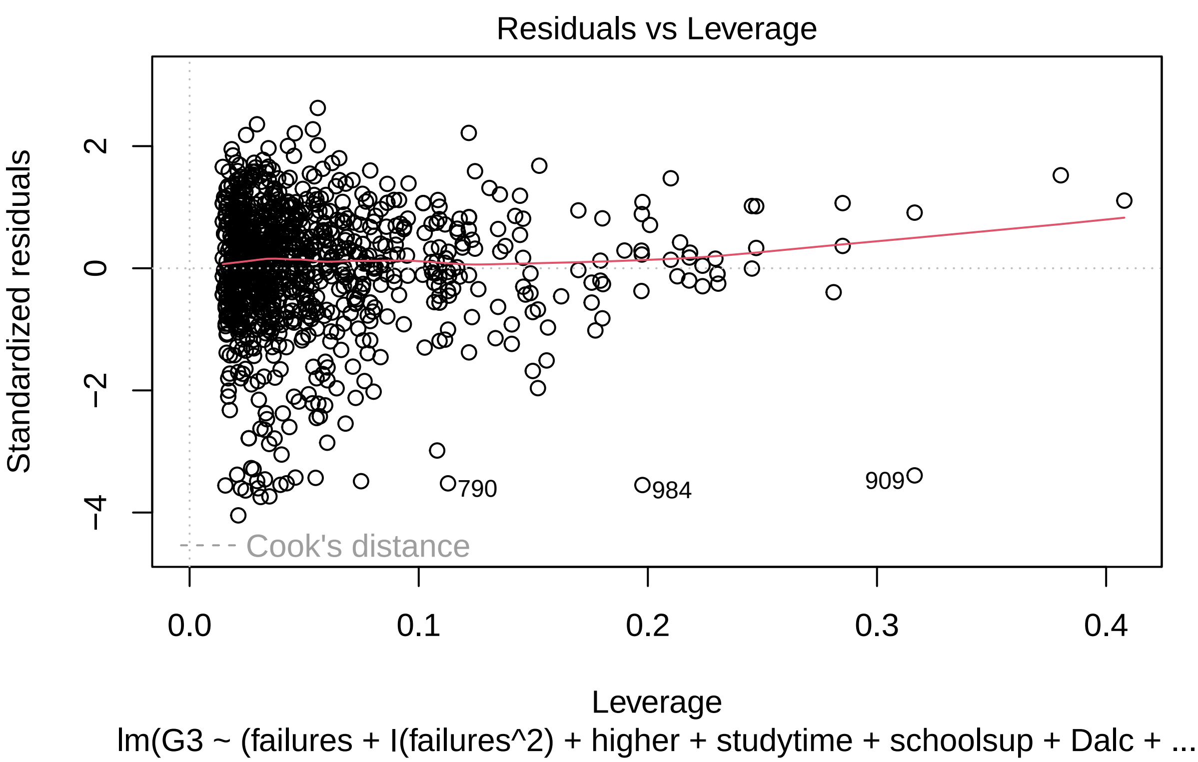
```r
summary(finalMdl)
```

```
##
## Call:
## lm(formula = G3 ~ (failures + higher + schoolsup + Dalc + health +
##     romantic + famsize + +goout + studytime), data = studentDataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.1848  -1.5137   0.3029   2.1295   8.6106
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  10.68826    0.63642  16.794  < 2e-16 ***
## failures     -1.82887    0.17261 -10.595  < 2e-16 ***
```

```
## higheryes      1.54720    0.41456   3.732 0.000200 ***
## schoolsupyes -1.16127    0.34238  -3.392 0.000721 ***
## Dalc2        -0.73721    0.28608  -2.577 0.010107 *
## Dalc3        -0.14301    0.45454  -0.315 0.753108
## Dalc4        -1.65491    0.69909  -2.367 0.018107 *
## Dalc5         0.11089    0.71654   0.155 0.877041
## health2      -0.66440    0.43152  -1.540 0.123946
## health3      -1.21726    0.38251  -3.182 0.001505 **
## health4      -0.74820    0.40008  -1.870 0.061751 .
## health5      -0.99991    0.34699  -2.882 0.004038 **
## romanticyes  -0.59306    0.23040  -2.574 0.010190 *
## famsizeLE3    0.48561    0.23806   2.040 0.041625 *
## goout2        1.15597    0.46996   2.460 0.014070 *
## goout3        0.67978    0.45855   1.482 0.138527
## goout4        0.38442    0.48055   0.800 0.423919
## goout5        0.08698    0.50253   0.173 0.862617
## studytime2    0.26960    0.26170   1.030 0.303179
## studytime3    1.18603    0.35442   3.346 0.000849 ***
## studytime4    0.62526    0.49164   1.272 0.203738
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.449 on 1023 degrees of freedom
## Multiple R-squared:  0.2189, Adjusted R-squared:  0.2036
## F-statistic: 14.33 on 20 and 1023 DF,  p-value: < 2.2e-16
```

3.6.6 Outlier

```
plot(student_performance_intMdl3,which=5)
```



```r
studentDataset[cooks.distance(student_performance_intMdl3)>0.5,]
```

```
##  [1] school     sex        age        address    famsize    Pstatus
##  [7] Medu       Fedu       Mjob       Fjob       reason     guardian
## [13] traveltime studytime  failures   schoolsup  famsup     paid
## [19] activities nursery    higher     internet   romantic   famrel
## [25] freetime   goout      Dalc       Walc       health     absences
## [31] G1         G2         G3
## <0 rows> (or 0-length row.names)
```

4. *Conclusion*

state final model give interpretation

5. *Discussion*

Future course of action

Appendix

1. Summary Full additive Model:

```r
summary(studentPerformance_fm)
```

```
##
## Call:
## lm(formula = G3 ~ (school + sex + age + address + famsize + Pstatus +
##     Medu + Fedu + Mjob + Fjob + reason + guardian + traveltime +
##     studytime + failures + schoolsup + famsup + activities +
##     nursery + higher + internet + romantic + famrel + freetime +
##     goout + Dalc + Walc + health + absences), data = studentDataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.2538  -1.4424   0.4217   2.0952   7.7895
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      11.481745   2.696281   4.258 2.26e-05 ***
## schoolMS         -0.309885   0.300992  -1.030 0.303479
## sexM             -0.104478   0.256079  -0.408 0.683368
## age               0.050072   0.104884   0.477 0.633179
## addressU          0.444262   0.278425   1.596 0.110896
## famsizeLE3        0.483175   0.250921   1.926 0.054443 .
## PstatusT         -0.110171   0.364903  -0.302 0.762778
## Medu1            -1.599656   1.214228  -1.317 0.188005
## Medu2            -1.633354   1.218213  -1.341 0.180304
## Medu3            -1.271212   1.233317  -1.031 0.302925
## Medu4            -0.693972   1.270242  -0.546 0.584964
## Fedu1            -1.170353   1.209054  -0.968 0.333289
## Fedu2            -1.083472   1.218187  -0.889 0.374000
## Fedu3            -1.324331   1.231681  -1.075 0.282541
```

```
## Fedu4            -1.153623   1.260240  -0.915 0.360208
## Mjobhealth        0.644770   0.566474   1.138 0.255309
## Mjobother        -0.097260   0.329016  -0.296 0.767593
## Mjobservices      0.573808   0.390234   1.470 0.141770
## Mjobteacher      -0.537369   0.535929  -1.003 0.316260
## Fjobhealth       -0.024084   0.756051  -0.032 0.974594
## Fjobother         0.146518   0.488355   0.300 0.764223
## Fjobservices     -0.169258   0.510364  -0.332 0.740231
## Fjobteacher       1.199449   0.693844   1.729 0.084178 .
## reasonhome       -0.118871   0.286125  -0.415 0.677903
## reasonother      -0.209278   0.384312  -0.545 0.586186
## reasonreputation  0.194450   0.300371   0.647 0.517546
## guardianmother   -0.329956   0.275944  -1.196 0.232089
## guardianother     0.118881   0.525016   0.226 0.820911
## traveltime2      -0.112876   0.256404  -0.440 0.659870
## traveltime3       0.332622   0.457432   0.727 0.467308
## traveltime4      -0.530444   0.776076  -0.683 0.494456
## studytime2        0.345119   0.274679   1.256 0.209255
## studytime3        1.189312   0.378589   3.141 0.001732 **
## studytime4        0.563988   0.520239   1.084 0.278590
## failures         -1.838328   0.185251  -9.923  < 2e-16 ***
## schoolsupyes     -1.114261   0.361618  -3.081 0.002119 **
## famsupyes        -0.296530   0.234051  -1.267 0.205476
## activitiesyes    -0.009298   0.228325  -0.041 0.967526
## nurseryyes       -0.061422   0.281003  -0.219 0.827021
## higheryes         1.190993   0.428916   2.777 0.005595 **
## internetyes       0.363775   0.293088   1.241 0.214836
## romanticyes      -0.648752   0.239575  -2.708 0.006889 **
## famrel2           0.297864   0.847050   0.352 0.725177
## famrel3           0.544597   0.722939   0.753 0.451445
## famrel4           0.956470   0.687110   1.392 0.164234
## famrel5           0.729534   0.701134   1.041 0.298363
## freetime2         0.974211   0.522148   1.866 0.062372 .
## freetime3        -0.021338   0.483631  -0.044 0.964818
## freetime4         0.332933   0.512091   0.650 0.515751
## freetime5         0.968995   0.585733   1.654 0.098382 .
## goout2            1.335736   0.480183   2.782 0.005511 **
## goout3            0.876640   0.473569   1.851 0.064452 .
## goout4            0.278556   0.501895   0.555 0.579016
## goout5           -0.079134   0.538137  -0.147 0.883121
## Dalc2            -0.709801   0.332067  -2.138 0.032803 *
## Dalc3            -0.092302   0.529267  -0.174 0.861590
## Dalc4            -2.000157   0.759479  -2.634 0.008582 **
## Dalc5            -0.617704   0.890539  -0.694 0.488080
## Walc2            -0.304902   0.309526  -0.985 0.324838
## Walc3             0.126708   0.350062   0.362 0.717461
## Walc4            -0.126274   0.442092  -0.286 0.775224
## Walc5             0.944524   0.659359   1.432 0.152324
## health2          -1.037651   0.444244  -2.336 0.019705 *
## health3          -1.300660   0.395549  -3.288 0.001044 **
## health4          -0.853415   0.411859  -2.072 0.038518 *
## health5          -1.249195   0.365081  -3.422 0.000648 ***
## absences          0.005013   0.018571   0.270 0.787265
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.393 on 977 degrees of freedom
## Multiple R-squared:  0.2781, Adjusted R-squared:  0.2293
## F-statistic: 5.703 on 66 and 977 DF,  p-value: < 2.2e-16
```

2. Summary of interaction model:

```r
summary(student_performance_intMdl1)
```

```
##
## Call:
## lm(formula = G3 ~ (failures + higher + studytime + schoolsup +
##     Dalc + health + romantic + famsize + +goout)^2, data = studentDataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.6059  -1.3914   0.1717   1.9203   6.6996
##
## Coefficients: (4 not defined because of singularities)
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 6.19551    2.81649   2.200  0.02809 *
## failures                   -2.75207    1.23301  -2.232  0.02587 *
## higheryes                   4.44810    2.38735   1.863  0.06278 .
## studytime2                  4.39372    1.64797   2.666  0.00782 **
## studytime3                  7.01518    3.50387   2.002  0.04559 *
## studytime4                  3.97662    5.20509   0.764  0.44509
## schoolsupyes                0.10943    4.34254   0.025  0.97990
## Dalc2                      -2.26256    2.33965  -0.967  0.33379
## Dalc3                      -0.80477    3.91410  -0.206  0.83715
## Dalc4                      -9.95939    8.68152  -1.147  0.25162
## Dalc5                      -2.63007    5.98232  -0.440  0.66031
## health2                    -0.35899    2.71301  -0.132  0.89476
## health3                     2.56397    2.58846   0.991  0.32219
## health4                     3.97467    2.79449   1.422  0.15530
## health5                     2.52100    2.41385   1.044  0.29660
## romanticyes                -2.61645    1.58806  -1.648  0.09981 .
## famsizeLE3                  0.10877    1.69830   0.064  0.94895
## goout2                      3.90749    2.56623   1.523  0.12821
## goout3                      3.76392    2.43585   1.545  0.12266
## goout4                      1.21308    2.45073   0.495  0.62074
## goout5                      1.70105    2.51234   0.677  0.49854
## failures:higheryes         -0.82183    0.52111  -1.577  0.11515
## failures:studytime2        -0.34101    0.45599  -0.748  0.45476
## failures:studytime3         0.06996    0.83190   0.084  0.93300
## failures:studytime4         1.31009    2.96755   0.441  0.65898
## failures:schoolsupyes       1.45452    0.73978   1.966  0.04960 *
## failures:Dalc2             -0.39955    0.47394  -0.843  0.39945
## failures:Dalc3              0.40039    0.70669   0.567  0.57116
## failures:Dalc4              0.13264    1.56996   0.084  0.93269
## failures:Dalc5              1.51596    1.20670   1.256  0.20935
## failures:health2            0.48151    1.04401   0.461  0.64477
## failures:health3            1.07117    0.83315   1.286  0.19890
## failures:health4           -0.46263    0.86941  -0.532  0.59478
```

```
## failures:health5           0.59094    0.79540   0.743  0.45772
## failures:romanticyes      -0.67921    0.48000  -1.415  0.15742
## failures:famsizeLE3        0.45925    0.47859   0.960  0.33753
## failures:goout2            0.31027    0.99617   0.311  0.75552
## failures:goout3            1.83641    0.99821   1.840  0.06616 .
## failures:goout4            1.46403    1.07575   1.361  0.17389
## failures:goout5            1.28334    0.99637   1.288  0.19809
## higheryes:studytime2      -2.03805    1.09505  -1.861  0.06307 .
## higheryes:studytime3       1.63302    2.59693   0.629  0.52963
## higheryes:studytime4      -8.78982    5.35875  -1.640  0.10131
## higheryes:schoolsupyes     0.71248    4.37144   0.163  0.87057
## higheryes:Dalc2            1.88685    1.19508   1.579  0.11474
## higheryes:Dalc3            2.64174    2.20154   1.200  0.23049
## higheryes:Dalc4            4.25797    5.00007   0.852  0.39468
## higheryes:Dalc5            4.75483    4.24757   1.119  0.26327
## higheryes:health2         -1.00878    2.15258  -0.469  0.63945
## higheryes:health3         -1.39330    2.04161  -0.682  0.49514
## higheryes:health4         -2.61984    2.12349  -1.234  0.21764
## higheryes:health5         -2.32303    1.90791  -1.218  0.22372
## higheryes:romanticyes      0.31145    1.02087   0.305  0.76038
## higheryes:famsizeLE3      -0.27038    1.17042  -0.231  0.81736
## higheryes:goout2          -1.19710    1.93327  -0.619  0.53594
## higheryes:goout3          -3.37541    1.79200  -1.884  0.05996 .
## higheryes:goout4           1.16620    1.90917   0.611  0.54147
## higheryes:goout5           0.51551    1.74393   0.296  0.76761
## studytime2:schoolsupyes   -0.27602    1.37089  -0.201  0.84048
## studytime3:schoolsupyes   -0.78356    1.66954  -0.469  0.63896
## studytime4:schoolsupyes   -2.24127    2.01097  -1.115  0.26537
## studytime2:Dalc2          -1.03569    0.78399  -1.321  0.18684
## studytime3:Dalc2          -0.83481    1.03042  -0.810  0.41807
## studytime4:Dalc2          -0.37492    2.42822  -0.154  0.87733
## studytime2:Dalc3          -0.35496    1.20886  -0.294  0.76911
## studytime3:Dalc3          -3.06285    3.23996  -0.945  0.34475
## studytime4:Dalc3          -1.18069    2.98535  -0.395  0.69258
## studytime2:Dalc4           1.69346    2.78591   0.608  0.54344
## studytime3:Dalc4           7.96165    7.45785   1.068  0.28602
## studytime4:Dalc4                NA         NA      NA       NA
## studytime2:Dalc5          -4.55563    4.39151  -1.037  0.29985
## studytime3:Dalc5         -16.46327    7.80422  -2.110  0.03519 *
## studytime4:Dalc5         -14.57926    6.00065  -2.430  0.01532 *
## studytime2:health2        -1.68591    1.19312  -1.413  0.15801
## studytime3:health2        -1.68791    1.83688  -0.919  0.35841
## studytime4:health2         3.11430    2.72881   1.141  0.25408
## studytime2:health3        -0.93001    1.07452  -0.866  0.38700
## studytime3:health3        -2.61377    1.63669  -1.597  0.11064
## studytime4:health3        -0.86750    1.78248  -0.487  0.62661
## studytime2:health4        -2.73272    1.12655  -2.426  0.01548 *
## studytime3:health4        -3.82429    1.65255  -2.314  0.02089 *
## studytime4:health4        -1.99454    2.37549  -0.840  0.40135
## studytime2:health5        -0.38549    0.90176  -0.427  0.66913
## studytime3:health5        -2.31601    1.52067  -1.523  0.12812
## studytime4:health5         3.05423    1.87299   1.631  0.10333
## studytime2:romanticyes     0.95350    0.63427   1.503  0.13313
## studytime3:romanticyes     0.67832    0.82952   0.818  0.41374
```

```
## studytime4:romanticyes     1.76484    2.12974   0.829  0.40753
## studytime2:famsizeLE3      0.12089    0.66683   0.181  0.85619
## studytime3:famsizeLE3      0.51127    0.96248   0.531  0.59542
## studytime4:famsizeLE3      2.35353    1.54717   1.521  0.12858
## studytime2:goout2         -1.46852    1.23334  -1.191  0.23410
## studytime3:goout2         -5.21845    2.22681  -2.343  0.01933 *
## studytime4:goout2          6.10169    3.11451   1.959  0.05042 .
## studytime2:goout3         -0.85596    1.22958  -0.696  0.48653
## studytime3:goout3         -4.20232    2.24467  -1.872  0.06153 .
## studytime4:goout3          5.90070    3.23892   1.822  0.06883 .
## studytime2:goout4         -1.96383    1.29053  -1.522  0.12845
## studytime3:goout4         -7.51204    2.39313  -3.139  0.00175 **
## studytime4:goout4          1.87345    3.74103   0.501  0.61665
## studytime2:goout5         -2.35473    1.34618  -1.749  0.08062 .
## studytime3:goout5         -4.89583    2.48231  -1.972  0.04890 *
## studytime4:goout5          1.24967    3.16468   0.395  0.69303
## schoolsupyes:Dalc2         1.82186    1.24149   1.467  0.14261
## schoolsupyes:Dalc3         0.10663    2.67987   0.040  0.96827
## schoolsupyes:Dalc4         5.99897    4.34381   1.381  0.16763
## schoolsupyes:Dalc5        -1.24564    3.72701  -0.334  0.73830
## schoolsupyes:health2       2.09829    1.62032   1.295  0.19567
## schoolsupyes:health3       2.14663    1.44409   1.486  0.13752
## schoolsupyes:health4       3.67709    1.43966   2.554  0.01082 *
## schoolsupyes:health5       1.83856    1.38313   1.329  0.18411
## schoolsupyes:romanticyes   0.40506    1.03780   0.390  0.69641
## schoolsupyes:famsizeLE3   -0.06095    1.12973  -0.054  0.95699
## schoolsupyes:goout2       -4.66081    1.77450  -2.627  0.00878 **
## schoolsupyes:goout3       -4.98956    1.84686  -2.702  0.00704 **
## schoolsupyes:goout4       -4.56433    1.86739  -2.444  0.01472 *
## schoolsupyes:goout5       -5.27996    2.10258  -2.511  0.01222 *
## Dalc2:health2             -1.28246    1.26426  -1.014  0.31068
## Dalc3:health2             -0.14983    2.21054  -0.068  0.94598
## Dalc4:health2             11.53218    6.58596   1.751  0.08030 .
## Dalc5:health2                   NA         NA      NA       NA
## Dalc2:health3             -0.69766    1.30187  -0.536  0.59217
## Dalc3:health3              0.27260    2.31822   0.118  0.90642
## Dalc4:health3             -7.49135    4.35025  -1.722  0.08542 .
## Dalc5:health3             -0.49121    3.62666  -0.135  0.89229
## Dalc2:health4              1.08093    1.18367   0.913  0.36139
## Dalc3:health4              2.77572    4.11811   0.674  0.50048
## Dalc4:health4             -2.47874    3.45081  -0.718  0.47276
## Dalc5:health4             -7.00544    5.08326  -1.378  0.16852
## Dalc2:health5             -0.07002    1.02722  -0.068  0.94567
## Dalc3:health5             -0.13611    1.84256  -0.074  0.94113
## Dalc4:health5              1.82151    4.05185   0.450  0.65315
## Dalc5:health5             -0.94226    4.35674  -0.216  0.82882
## Dalc2:romanticyes         -0.18658    0.69807  -0.267  0.78931
## Dalc3:romanticyes          2.56848    1.34287   1.913  0.05612 .
## Dalc4:romanticyes          3.06578    3.47672   0.882  0.37813
## Dalc5:romanticyes          4.21448    4.09434   1.029  0.30361
## Dalc2:famsizeLE3          -0.92768    0.69149  -1.342  0.18009
## Dalc3:famsizeLE3           0.17557    1.18086   0.149  0.88184
## Dalc4:famsizeLE3           7.40395    3.82153   1.937  0.05302 .
## Dalc5:famsizeLE3          -1.58703    2.65124  -0.599  0.54960
```

```
## Dalc2:goout2              0.98252    1.89981   0.517  0.60517
## Dalc3:goout2             -2.36548    2.68578  -0.881  0.37870
## Dalc4:goout2             -7.90805    3.48983  -2.266  0.02370 *
## Dalc5:goout2              8.17369    5.93387   1.377  0.16873
## Dalc2:goout3              1.44074    1.88254   0.765  0.44429
## Dalc3:goout3             -3.95701    2.69239  -1.470  0.14201
## Dalc4:goout3              6.94833    7.71471   0.901  0.36802
## Dalc5:goout3              9.46036    4.46007   2.121  0.03420 *
## Dalc2:goout4              0.33667    1.89862   0.177  0.85930
## Dalc3:goout4             -3.56439    2.57833  -1.382  0.16720
## Dalc4:goout4             -1.71477    3.94362  -0.435  0.66380
## Dalc5:goout4             -3.84918    3.93697  -0.978  0.32850
## Dalc2:goout5             -0.31882    1.97374  -0.162  0.87171
## Dalc3:goout5             -0.89956    2.62620  -0.343  0.73203
## Dalc4:goout5                   NA         NA      NA       NA
## Dalc5:goout5                   NA         NA      NA       NA
## health2:romanticyes       0.21476    1.04665   0.205  0.83747
## health3:romanticyes       0.70414    0.93910   0.750  0.45358
## health4:romanticyes      -0.70717    0.99705  -0.709  0.47835
## health5:romanticyes      -0.53777    0.83714  -0.642  0.52080
## health2:famsizeLE3       -1.88495    1.04453  -1.805  0.07149 .
## health3:famsizeLE3       -0.60730    0.93707  -0.648  0.51710
## health4:famsizeLE3       -0.32377    0.97530  -0.332  0.74000
## health5:famsizeLE3       -1.19048    0.83867  -1.419  0.15612
## health2:goout2            1.39577    2.08052   0.671  0.50248
## health3:goout2           -2.21354    1.85301  -1.195  0.23259
## health4:goout2           -1.23126    2.32097  -0.530  0.59591
## health5:goout2           -1.69178    1.66702  -1.015  0.31046
## health2:goout3            3.10719    1.96421   1.582  0.11404
## health3:goout3           -1.38610    1.81554  -0.763  0.44540
## health4:goout3            0.44052    2.18944   0.201  0.84059
## health5:goout3            0.84908    1.63247   0.520  0.60312
## health2:goout4            3.08032    2.00483   1.536  0.12480
## health3:goout4           -2.21220    1.88143  -1.176  0.24000
## health4:goout4           -0.12331    2.20087  -0.056  0.95533
## health5:goout4           -0.52818    1.64521  -0.321  0.74826
## health2:goout5            1.70747    2.11700   0.807  0.42015
## health3:goout5           -3.81315    2.07869  -1.834  0.06694 .
## health4:goout5            0.17455    2.38932   0.073  0.94178
## health5:goout5           -1.16655    1.78371  -0.654  0.51329
## romanticyes:famsizeLE3    0.19945    0.57899   0.344  0.73057
## romanticyes:goout2        1.71975    1.16526   1.476  0.14035
## romanticyes:goout3        0.56134    1.15334   0.487  0.62659
## romanticyes:goout4        1.52420    1.19973   1.270  0.20427
## romanticyes:goout5        1.53200    1.24633   1.229  0.21933
## famsizeLE3:goout2         2.00716    1.20129   1.671  0.09512 .
## famsizeLE3:goout3         1.33000    1.17687   1.130  0.25874
## famsizeLE3:goout4         0.68444    1.22336   0.559  0.57598
## famsizeLE3:goout5         1.39645    1.30170   1.073  0.28367
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.357 on 858 degrees of freedom
## Multiple R-squared:  0.3792, Adjusted R-squared:  0.2454
```

```
## F-statistic: 2.833 on 185 and 858 DF,  p-value: < 2.2e-16
```

3. Summary of final interaction model:

```r
summary(student_performance_intMdl2)
```

```
##
## Call:
## lm(formula = G3 ~ (failures + higher + studytime + schoolsup +
##     Dalc + health + romantic + famsize + +goout + failures:schoolsup +
##     studytime:health + studytime:goout + schoolsup:health + schoolsup:goout),
##     data = studentDataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.5009  -1.5919   0.2916   2.1990   7.9895
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)             8.498798   0.981759   8.657  < 2e-16 ***
## failures               -2.046312   0.183899 -11.127  < 2e-16 ***
## higheryes               1.571610   0.413108   3.804 0.000151 ***
## studytime2              2.629073   1.128137   2.330 0.019982 *
## studytime3              7.110347   2.120459   3.353 0.000829 ***
## studytime4             -1.775067   2.219330  -0.800 0.424006
## schoolsupyes            0.776582   1.496920   0.519 0.604026
## Dalc2                  -0.818135   0.290378  -2.817 0.004937 **
## Dalc3                   0.008952   0.458562   0.020 0.984428
## Dalc4                  -1.642254   0.698916  -2.350 0.018983 *
## Dalc5                   0.318271   0.745456   0.427 0.669510
## health2                 0.533736   0.855741   0.624 0.532960
## health3                -0.862240   0.756791  -1.139 0.254838
## health4                 0.376706   0.772522   0.488 0.625919
## health5                -0.468239   0.654243  -0.716 0.474348
## romanticyes            -0.601559   0.234131  -2.569 0.010335 *
## famsizeLE3              0.611245   0.240307   2.544 0.011123 *
## goout2                  2.701405   0.826061   3.270 0.001112 **
## goout3                  2.120678   0.806759   2.629 0.008705 **
## goout4                  2.270820   0.841135   2.700 0.007058 **
## goout5                  1.910583   0.848629   2.251 0.024580 *
## failures:schoolsupyes  1.470464   0.531499   2.767 0.005769 **
## studytime2:health2    -1.995736   1.020678  -1.955 0.050828 .
## studytime3:health2    -1.342237   1.636566  -0.820 0.412326
## studytime4:health2     0.633159   2.065026   0.307 0.759204
## studytime2:health3    -0.720734   0.915681  -0.787 0.431411
## studytime3:health3    -2.050107   1.438783  -1.425 0.154504
## studytime4:health3    -1.209162   1.562797  -0.774 0.439283
## studytime2:health4    -1.586649   0.944165  -1.680 0.093180 .
## studytime3:health4    -2.868793   1.453510  -1.974 0.048694 *
## studytime4:health4    -2.212189   2.067158  -1.070 0.284808
## studytime2:health5    -0.677095   0.802073  -0.844 0.398772
## studytime3:health5    -2.416723   1.359039  -1.778 0.075668 .
## studytime4:health5    -0.130447   1.579813  -0.083 0.934209
## studytime2:goout2     -1.117962   1.061877  -1.053 0.292682
```

```
## studytime3:goout2     -3.996336   1.818851  -2.197 0.028239 *
## studytime4:goout2      4.912056   2.193727   2.239 0.025368 *
## studytime2:goout3     -1.319228   1.033860  -1.276 0.202247
## studytime3:goout3     -3.495151   1.791390  -1.951 0.051329 .
## studytime4:goout3      3.472643   2.242616   1.548 0.121827
## studytime2:goout4     -1.842089   1.061063  -1.736 0.082861 .
## studytime3:goout4     -4.748364   1.921606  -2.471 0.013640 *
## studytime4:goout4      1.986642   2.727053   0.728 0.466483
## studytime2:goout5     -1.770224   1.095119  -1.616 0.106312
## studytime3:goout5     -2.911700   2.013425  -1.446 0.148454
## studytime4:goout5      2.181123   2.207602   0.988 0.323391
## schoolsupyes:health2   2.214681   1.424101   1.555 0.120232
## schoolsupyes:health3   2.450585   1.269637   1.930 0.053874 .
## schoolsupyes:health4   2.768818   1.283108   2.158 0.031176 *
## schoolsupyes:health5   2.849506   1.192475   2.390 0.017054 *
## schoolsupyes:goout2   -5.901575   1.351639  -4.366  1.4e-05 ***
## schoolsupyes:goout3   -4.808884   1.366416  -3.519 0.000452 ***
## schoolsupyes:goout4   -4.135247   1.368517  -3.022 0.002578 **
## schoolsupyes:goout5   -5.866249   1.562109  -3.755 0.000183 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.398 on 990 degrees of freedom
## Multiple R-squared:  0.2664, Adjusted R-squared:  0.2271
## F-statistic: 6.783 on 53 and 990 DF,  p-value: < 2.2e-16
```

END

```
#studentPerformance_p1 = lm(G3 ~ (sex+age+address+famsize+Pstatus+Medu+Fedu+traveltime+studytime+
#                                  failures+schoolsup+famsup+activities+higher+internet+romantic+
#                                  famrel+freetime+goout+Dalc+Walc+health+absences),
#                            data = studentDataset)

#summary(studentPerformance_p1)
```

```
#subsets = ols_step_forward_p(studentPerformance_fm,p_val = 0.05, details = FALSE)
#forwardMdl = subsets$model
#summary(forwardMdl)
```

```
#bestSubsetMdl = ols_step_best_subset(forwardMdl, details = FALSE)
#(bestSubsetMdl$metrics)
```

```
#studentPerformance_p2 = lm(G3 ~ (failures+higher+schoolsup+school+romantic+studytime+address+Medu+Dalc
 #                           data = studentDataset)

#summary(studentPerformance_p2)
```

```
#studentPerformance_p3 = lm(G3 ~ (failures+higher+schoolsup+school+romantic+studytime+address+Dalc+fams
#                            data = studentDataset)

#summary(studentPerformance_p3)
```

```
#studentPerformance_p4 = lm(G3 ~ (failures+higher+schoolsup+school+romantic+studytime+address+Dalc+fams
#                              data = studentDataset)

#summary(studentPerformance_p4)
```

```
#forwardMdl_int = lm(G3 ~ (failures+higher+schoolsup+school+romantic+studytime+address+Dalc+famsize+goo
#summary(forwardMdl_int)
```

```
#forwardMdl_int1 = lm(G3 ~ (failures+higher+schoolsup+school+romantic+studytime+address+Dalc+famsize+go
#summary(forwardMdl_int1)
```

```
#anova(studentPerformance_p4,forwardMdl_int1)
```

Assumption code

```
#Linearity Assumption
#library(ggplot2)
#ggplot(forwardMdl_int1, aes(x=.fitted, y=.resid)) +
#geom_point() +geom_smooth()+
#geom_hline(yintercept = 0)
```

```
# Equal Variance Assumption

#ggplot(forwardMdl_int1, aes(x=.fitted, y=.resid)) +
#geom_point(colour = "purple") +
#geom_hline(yintercept = 0) +
#geom_smooth(colour = "green4")+
#ggtitle("Residual plot: Residual vs Fitted values")
```

```
#ggplot(forwardMdl_int1, aes(x=.fitted, y=sqrt(abs(.stdresid)))) +
#geom_point(colour = "purple") +
#geom_hline(yintercept = 0) +
#geom_smooth( colour = "green4")+
#ggtitle("Scale-Location plot : Standardized Residual vs Fitted values")
```

```
#library(lmtest)

#bptest(forwardMdl_int1)  # It doesn't have heteroscedasticity.
```

```
# Normality Assumption

#ggplot(data=studentDataset, aes(residuals(forwardMdl_int1))) +
#geom_histogram(breaks = seq(-6,6,by=1), col="green3", fill="green4") +
#labs(title="Histogram for residuals") +
#labs(x="residuals", y="Count")
```

```
#ggplot(studentDataset, aes(sample=forwardMdl_int1$residuals)) +
#stat_qq() +
#stat_qq_line()
```

```
#Testing for Normality
#shapiro.test(residuals(forwardMdl_int1))  # We don't have normality
```

```
#Multicollinearity
#library(mctest)
#pairs(~G3+studytime+goout+health,data=studentDataset)
```

```
#imcdiag(studentPerformance_fm, method="VIF")
```

```
#forwardMdl_int2 = lm(G3 ~ (failures+higher+schoolsup+school+romantic+studytime+address+Dalc+famsize+go
#summary(forwardMdl_int2)
```

```
#imcdiag(forwardMdl_int2, method="VIF")
```

```
#Linearity Assumption
#library(ggplot2)
#ggplot(forwardMdl_int2, aes(x=.fitted, y=.resid)) +
#geom_point() +geom_smooth()+
#geom_hline(yintercept = 0)
```

```
#ggplot(forwardMdl_int2, aes(x=.fitted, y=sqrt(abs(.stdresid)))) +
#geom_point(colour = "purple") +
#geom_hline(yintercept = 0) +
#geom_smooth( colour = "green4")+
#ggtitle("Scale-Location plot : Standardized Residual vs Fitted values")
```

```
#library(lmtest)

#bptest(forwardMdl_int2)  # It doesn't have heteroscedasticity.
```

```
# Normality Assumption

#ggplot(data=studentDataset, aes(residuals(forwardMdl_int2))) +
#geom_histogram(breaks = seq(-5,5,by=1), col="green3", fill="green4") +
#labs(title="Histogram for residuals") +
#labs(x="residuals", y="Count")
```

```
#ggplot(studentDataset, aes(sample=forwardMdl_int2$residuals)) +
#stat_qq() +
#stat_qq_line()
```

```
#Testing for Normality
#shapiro.test(residuals(forwardMdl_int2))  # We don't have normality
```

```
#qqnorm(studentDataset$G3)
#qqline(studentDataset$G3, col="red")
```

```
# Outlier

#plot(forwardMdl_int2,which=5)
```

```
#studentDataset[cooks.distance(forwardMdl_int2)>0.5,]
```

```
#plot(forwardMdl_int2,pch=18,col="red",which=c(4))
```

```
#fanyi <- function(model,dep) {
#    coe=coefficients(model)
#    for(a in names(coe)){
#      if (coe[a]>0){
#        print(paste('For one unit increases in ',a,' ',dep,' increases by ', signif(coe[a])))
#      }else{
#        print(paste('For one unit increases in ',a,' ',dep,' decreases by ', signif(-1*coe[a])))
#      }
```

```
#
#    }
#}
#fanyi(studentPerformance_fm,'the grade of student')
#formula <- function(model,dep){
#  coe=coefficients(model)
#  b=paste(dep,'=')
#  for(a in names(coe)){
#    if (coe[a]>0){
#      b=paste(b,' +',signif(coe[a]),'(',a,')',sep="")
#    }else{
#      b=paste(b,' ',signif(coe[a]),'(',a,')',sep="")
#    }

#  }
#  print(b)
#}
#formula(student_performance_intMdl1,'G3')




#```{r}

#fanyi <- function(model,dep) {

#    coe=coefficients(model)

#    for(a in names(coe)){

 #     if (coe[a]>0){

#        print(paste('For one unit increases in ',a,' ',dep,' increases by ', signif(coe[a])))
#
#      }else{

#        print(paste('For one unit increases in ',a,' ',dep,' decreases by ', signif(-1*coe[a])))

#      }

#    }

#}

#fanyi(studentPerformance_fm,'the grade of student')

#formula <- function(model,dep){

#  coe=coefficients(model)

#  b=paste(dep,'=')

#  for(a in names(coe)){
```

```
#    if (coe[a]>0){
#      b=paste(b,' +',signif(coe[a]),'(',a,')',sep="")
#    }else{
#      b=paste(b,' ',signif(coe[a]),'(',a,')',sep="")
#    }
#  }
#  print(b)
#}
#formula(student_performance_intMdl3,'G3')
```