**STAT 512**

**HW 6**

**Name: Haoran Zhang**

**Section: 010**

**Time: 1:30-2:45 pm Tuesday/Thursday**

An assistant in the district sales office of a national cosmetics firm obtained data on advertising expenditures and sales last year in the district’s 44 territories. Data is consmetics.csv

X1: expenditures for point-of-sale displays in beauty salons and department stores (X$1000).

X2: expenditures for local media advertising.

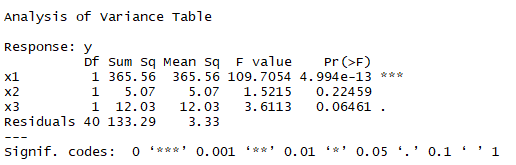
X3: expenditures for prorated share of national media advertising.

Y: Sales (X$1000).

1. (3) Test the regression relation between sales and the three predictor variables. State the hypotheses, test statistic and degrees of freedom, the p-value, the conclusion in words.

Assume the regression model is

F test:



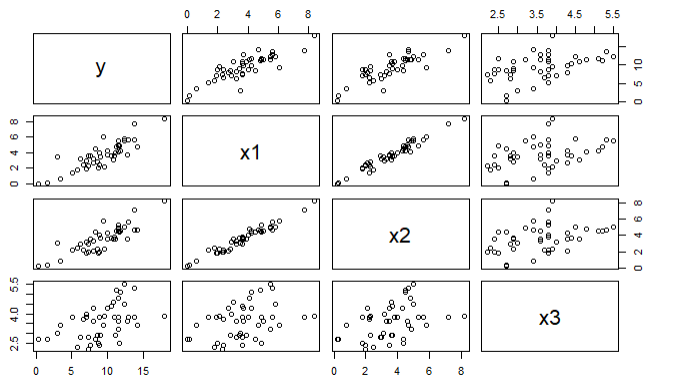
Since from F table

Therefore, the rejection region is:

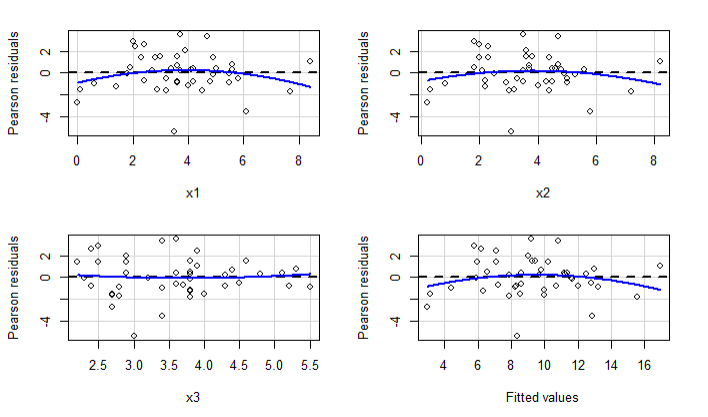
Besides, the corresponding p-value is <0.01.

As a result, since so it is evident that the null hypothesis should be rejected, not all of the predictors can be dropped from the regression model.

2. (3) Determine whether the linear regression model is appropriate by using the “usual” plots (scatterplot, residual plots, histogram/QQ plot). Explain in detail whether or not each assumption appears to be substantially violated.

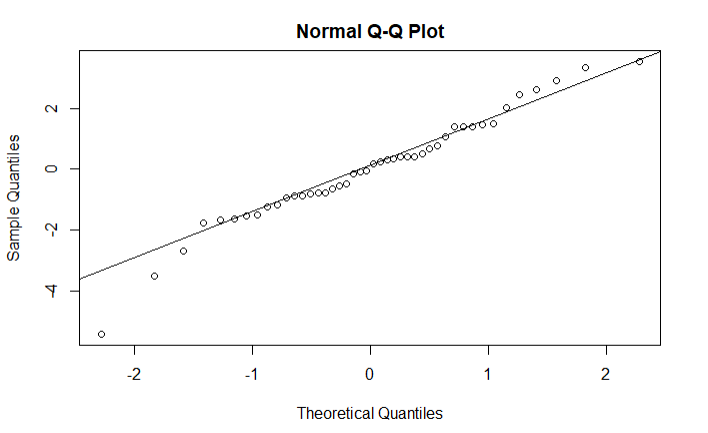


From the scatter plot, Y has a strong association with , but there is an obvious correlation between . Besides, the relationships between Y & is much weaker.



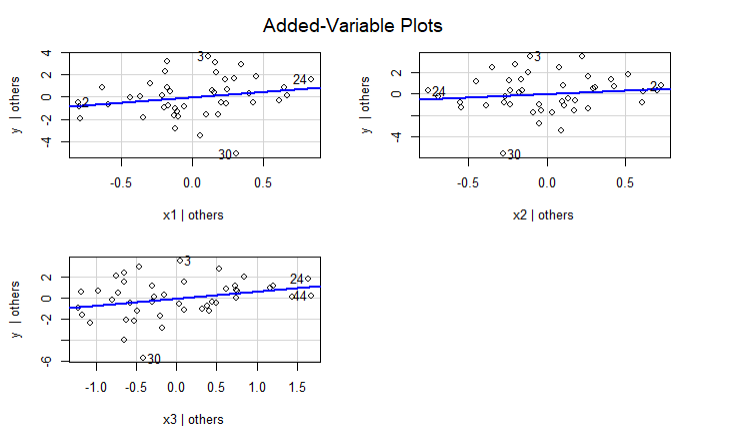
From the above residual plots, it can be concluded that the distribution of residuals does not have a certain pattern, which indicates the non-constant and non-normal variances.

Hypothesis test:



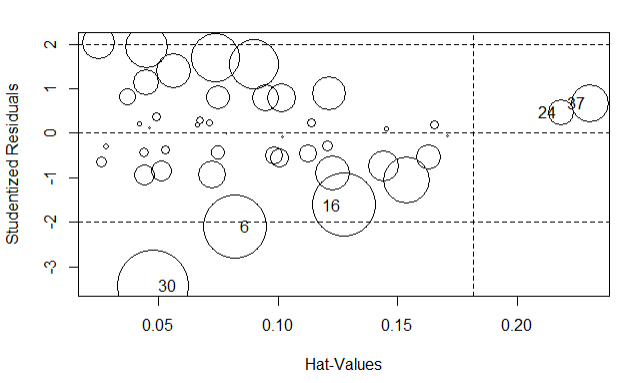
As shown in the above normal probability plot, the residual points fit the line closely. Therefore, the error is proved to be normally distributed.

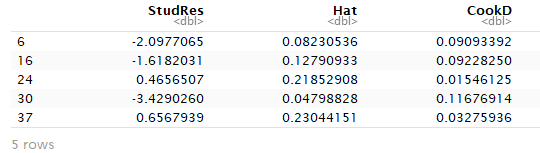
3. (3) Prepare a partial regression plot for each of the predictor variables. Do your plots suggest that the regression relationship in the fitted regression function are inappropriate for any of the predictor variables? Explain.



Those plots show that the regression relationship in the fitted regression function indicates that the linear term in each predictor may be a helpful addition to the regression model containing other predictors respectively, because It can be shown that the slope of the least squares line through the origin fitted to the plotted residuals is the , the regression coefficient of the predictor.

4. (3) Are there any outlying Y observations? (Do not include in your answer the values for all cases. Use plots and verbal summaries instead. You may include values for a few selected cases if you wish).

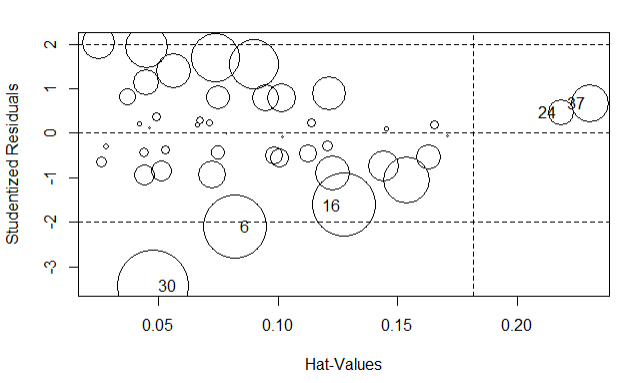


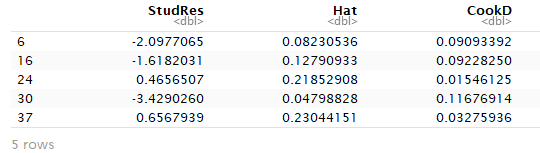


From the plot and the table, it is found that case 30 is the outlier regard to Y, since the absolute studentized residual value of case 30 .

5. (3) Are there any outlying X observations? (Do not include in your answer the values for all cases. Use plots and verbal summaries instead. You may include values for a few selected cases if you wish).

Therefore, the case with is considered to be the outlier regard to X.





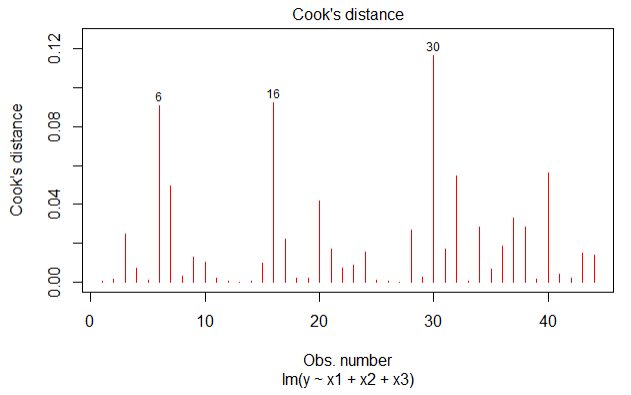
From the plot and the table, it is found that cases 24 and 37 are the outliers regard to Y, since their hat matrix values .

6. (3) Are there any influential points?

i) Cook’s distance

The case i is considered as major influential point when

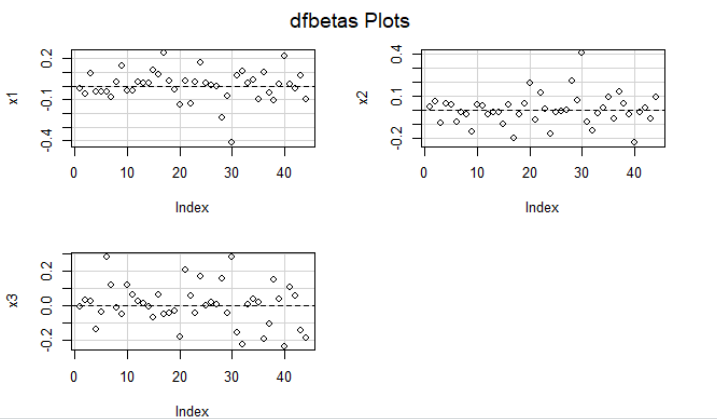
The case i is considered as minor influential point when , which is in the range of .



Therefore, there is no influential points found using Cook’s distance.

ii) DEBETAS method

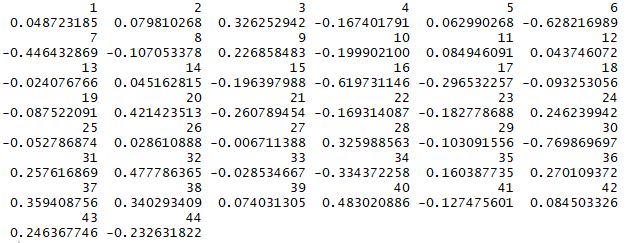
A case i is considered has a large impact on the regression coefficients when



From the DFBETAS plots, case 30 is considered has a large impact on the regression coefficients, since its absolute

iii) DFFITS:

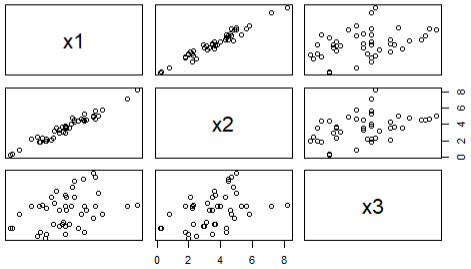
A case i is considered has a large impact on the regression coefficients when absolute value of

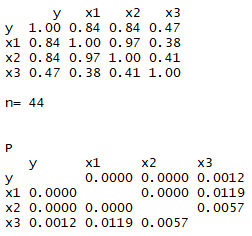


Therefore, cases 6 and 16 are considered as influential points.

7. Is there a serious multicollinearity problem?

1. (3) Include an appropriate scatterplot and correlation values between the explanatory variables.

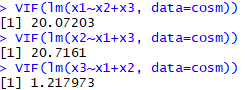




From the scatter plots and the correlation matrix, it is found that there is the correlation between .

1. (3) Judge by VIF, do you think there is a problem with multicollinearity? (Hint: VIP or tolerance)

If the maximum , there is a strong multicollinearity.





Therefore, it is found that .

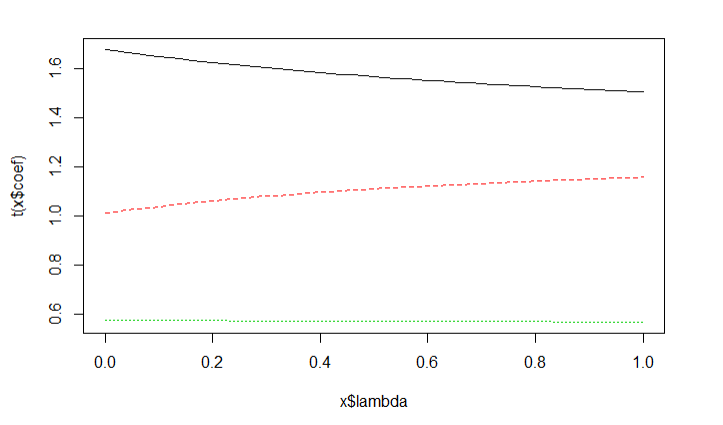
1. (2) Compare your answers in parts i and ii. Are your conclusions the same or different? Please explain your answer.

The conclusions are the same, the correlation matrix and scatter plots in part a) denotes that there is a strong correlation between , and the VIF offers the same results in part b).

8. Instead of removing variables, we are going to use the Ridge Regression to determine the parameter values.

a) (3) Make a ridge trace plot. What value of the parameter () do you believe is best? Explain your choice.

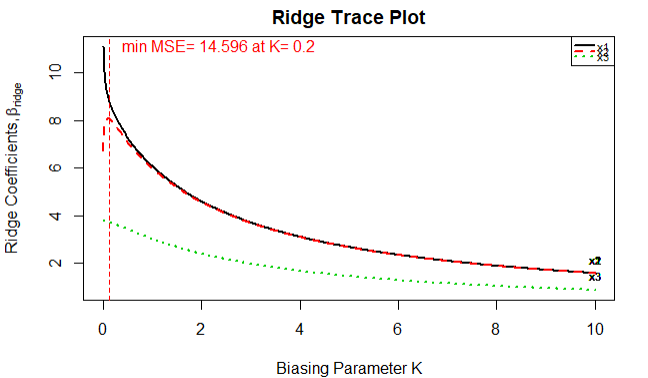
Ridge Trace Plot



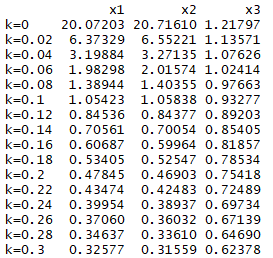


The model performs the best in the cross-validation using

b) (3) Using the VIF factors, what value of the parameter do you believe should be used? (Hint: Look at both the graph and the printed numbers.) Explain your choice.

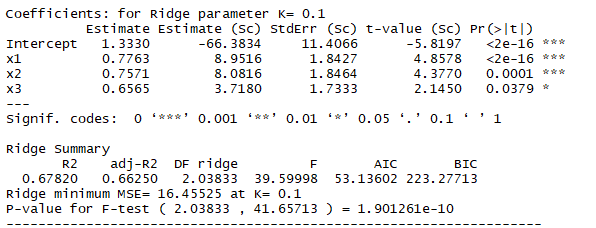


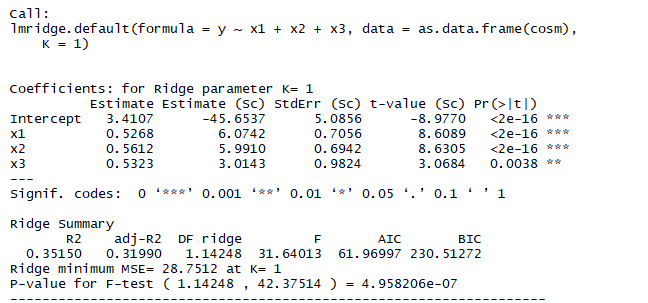
For reducing MSE, . However, in this question, this method is not used.



From this VIF factors table, we can conclude that since when k = 0.1, those three VIF values approaches 1.0, the best choice for

c) (2) Using your choice for the parameter, what are the standard regression parameters in this situation?

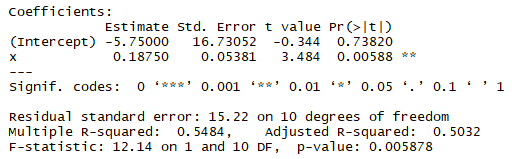


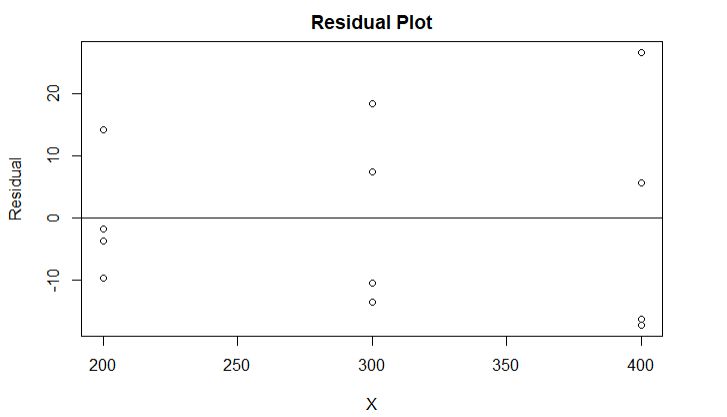


From those two summary tables of k = 0.1 and k = 1, the AIC and BIC values are smaller for k = 0.1, and adjusted is higher in the k = 0.1 case. Therefore, we choose k = 0.1.

9. The number of defective items produced by a machine produced by a machine (Y) is known to be linearly related to the speed setting of the machine (X). Data is in machine.csv.

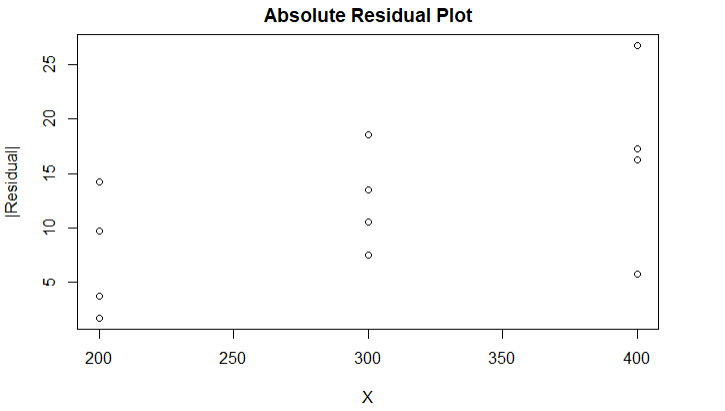
1. (2) Fit a linear regression function by ordinary least squares; obtain the residuals and plot the residuals against X. What does the residual plot suggest?

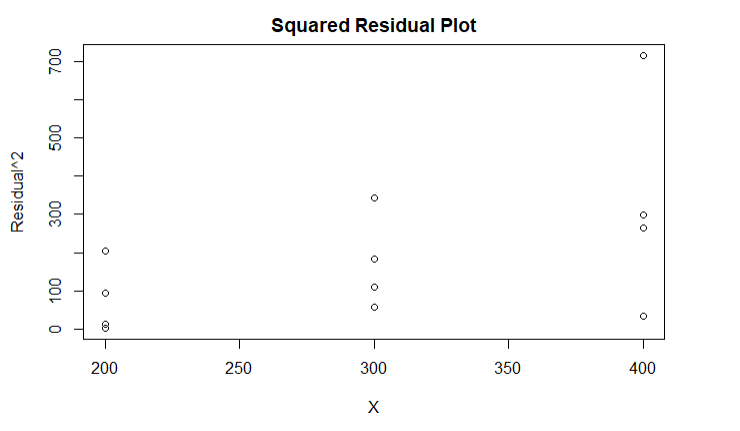




As shown in the residual plot, the residuals diverge from 0 when X increases, thus there is a pattern in this residual plot, which suggests the non-constant variance.

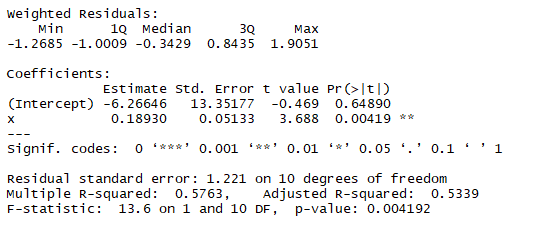
1. (3) Plot the absolute value of the residuals and the squared residuals vs. X. Which plot has a better line?





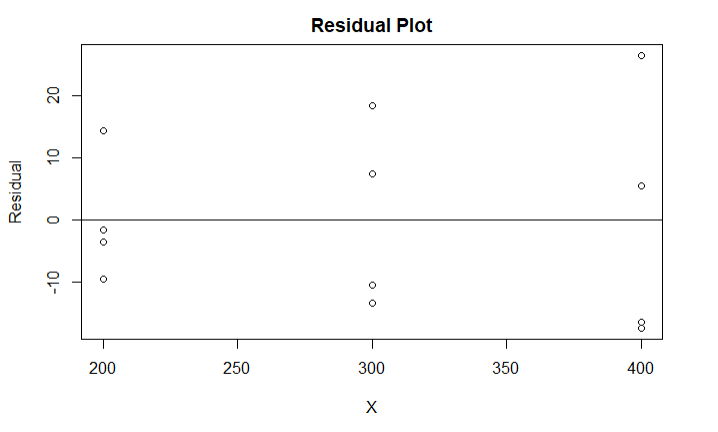
Comparing two plots, the absolute residual plot has a better line, because regressing absolute residuals is less affected by the outliers than regressing squared residuals.

1. (4) Perform a weighted least square using the squared residuals to compute the weights. Obtain the weighted least squares estimates for the estimated parameters and their standard errors. Are these values similar to the ones produced in a)? Which results are better, the ones generated in a) or c)? Please explain your answer.



The estimated parameters and their standard errors are different from those values in part a). The results in part c) is preferred, because the standard errors of regression coefficients are smaller than those values in part a) respectively; and the adjusted in part c) is also higher.

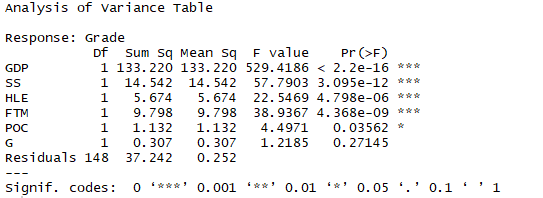
d) (3) Re-calculate the residuals for the weighted least squares and make a residual plot vs. X. Did this correct the problem that was seen in a).



Yes, the weighted least square method solves the problem caused by non-constant variance in part a).

10. (Independent assignment, do not discuss) Fit a linear regression function on your project data. Perform primary procedure of assumption checking, model selection and diagnostic.

1. Regression Test:

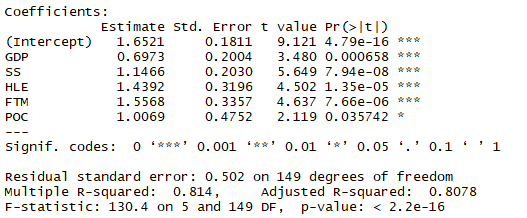


from F table.

When , reject null hypothesis; corresponding p-value < 0.01

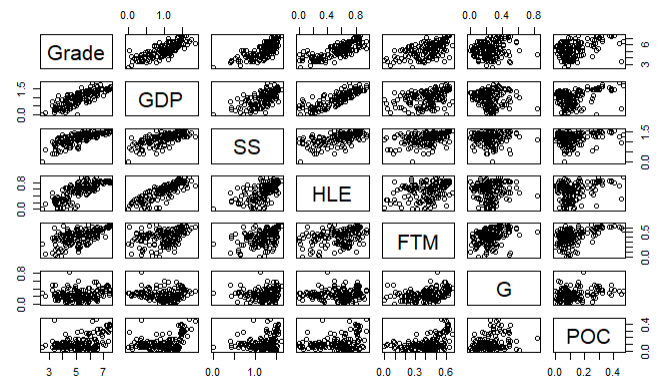
Since, not all of the predictors can be dropped from the regression model.

1. Primary model:

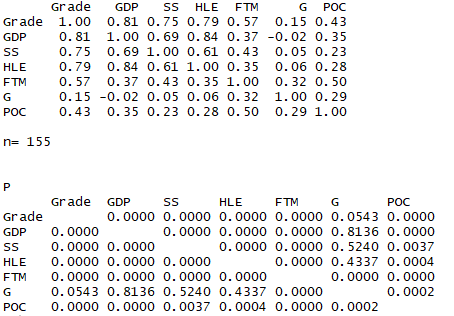


1. Correlation test:

Scatter plots:

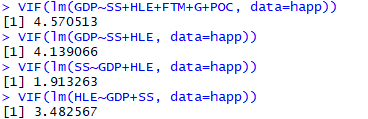


Correlation matrix:



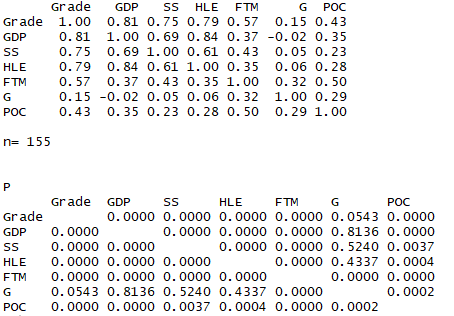
It is found that there is a correlation between GDP & social support, and GDP & HLE.

VIF values:



There is no obvious correlation found in VIF test.

1. After standardization (), no change in correlation matrix



1. Regression coefficients reduction:

Calculate marginal relative variance reduction:

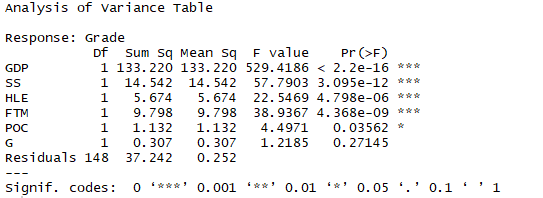
second lowest

lowest

F test:

Full model

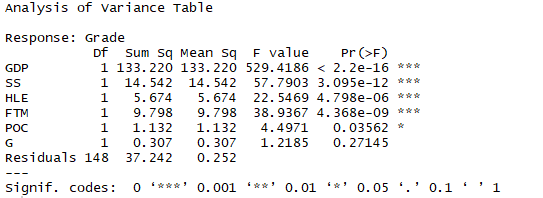
Reduced model



Since from F table

Since , .

F test:



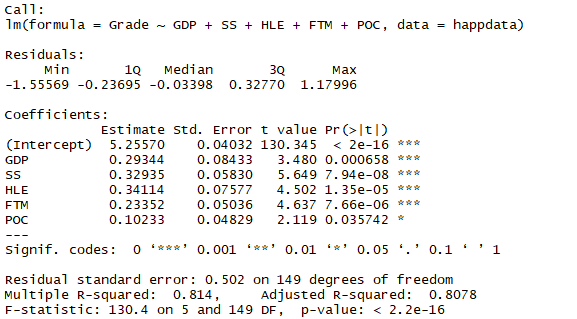
Since from F table, which is close to

Therefore, it is not confident to make a decision here.

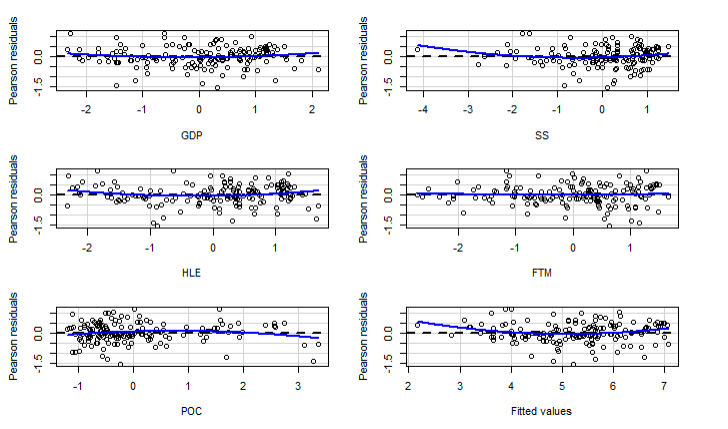
Bestsub:

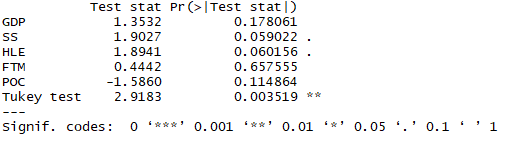
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | p | 1 | 2 | 3 | 4 | 5 | 6 | SSEp | r2 | r2.adj | Cp | AICp | SBCp | PRESSp |
| 1 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 68.69347 | 0.65978721 | 0.65756359 | 121.989477 | -122.1345 | -116.04765 | 70.56867 |
| 1 | 2 | 0 | 0 | 1 | 0 | 0 | 0 | 77.03301 | 0.61848464 | 0.61599107 | 155.130991 | -104.37463 | -98.287775 | 79.29513 |
| 1 | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 87.1803 | 0.56822892 | 0.56540688 | 195.456585 | -85.19424 | -79.107388 | 89.45355 |
| 1 | 2 | 0 | 0 | 0 | 1 | 0 | 0 | 136.34258 | 0.32474673 | 0.32033331 | 390.828649 | -15.87943 | -9.792584 | 139.71373 |
| 1 | 2 | 0 | 0 | 0 | 0 | 0 | 1 | 164.35982 | 0.18598791 | 0.18066757 | 502.169834 | 13.08811 | 19.174959 | 169.8037 |
| 1 | 2 | 0 | 0 | 0 | 0 | 1 | 0 | 197.06837 | 0.02399485 | 0.01761573 | 632.15437 | 41.21947 | 47.30632 | 202.64514 |
| 2 | 3 | 1 | 0 | 0 | 1 | 0 | 0 | 51.71168 | 0.74389162 | 0.74052177 | 56.503423 | -164.14994 | -155.019661 | 53.74846 |
| 2 | 3 | 0 | 1 | 1 | 0 | 0 | 0 | 53.06939 | 0.73716739 | 0.73370906 | 61.898999 | -160.13286 | -151.00258 | 55.16945 |
| 2 | 3 | 1 | 1 | 0 | 0 | 0 | 0 | 54.15146 | 0.73180829 | 0.72827945 | 66.199181 | -157.00423 | -147.873955 | 56.14518 |
| 2 | 3 | 0 | 0 | 1 | 1 | 0 | 0 | 57.05143 | 0.71744581 | 0.71372799 | 77.723749 | -148.91815 | -139.78787 | 59.55631 |
| 2 | 3 | 1 | 0 | 1 | 0 | 0 | 0 | 61.49015 | 0.69546254 | 0.69145547 | 95.363305 | -137.30496 | -128.174687 | 63.98477 |
| 2 | 3 | 1 | 0 | 0 | 0 | 1 | 0 | 62.8277 | 0.68883816 | 0.68474393 | 100.678758 | -133.96951 | -124.839234 | 65.39532 |
| 3 | 4 | 0 | 1 | 1 | 1 | 0 | 0 | 42.77099 | 0.78817146 | 0.78396295 | 22.972918 | -191.57259 | -179.398889 | 45.02872 |
| 3 | 4 | 1 | 1 | 0 | 1 | 0 | 0 | 43.50899 | 0.78451643 | 0.7802353 | 25.905748 | -188.92092 | -176.747224 | 45.63701 |
| 3 | 4 | 0 | 1 | 1 | 0 | 0 | 1 | 45.54102 | 0.77445255 | 0.76997147 | 33.981087 | -181.84581 | -169.672111 | 48.61169 |
| 3 | 4 | 1 | 0 | 1 | 1 | 0 | 0 | 46.06168 | 0.7718739 | 0.7673416 | 36.050212 | -180.08378 | -167.910076 | 48.69907 |
| 3 | 4 | 1 | 1 | 1 | 0 | 0 | 0 | 48.47789 | 0.75990734 | 0.75513729 | 45.652278 | -172.15918 | -159.985484 | 51.00115 |
| 3 | 4 | 1 | 1 | 0 | 0 | 0 | 1 | 48.98666 | 0.7573876 | 0.75256748 | 47.674142 | -170.54095 | -158.367254 | 51.9926 |
| 4 | 5 | 1 | 1 | 1 | 1 | 0 | 0 | 38.68008 | 0.80843218 | 0.80332371 | 8.715555 | -205.15556 | -189.938431 | 41.2582 |
| 4 | 5 | 0 | 1 | 1 | 1 | 0 | 1 | 40.5999 | 0.79892407 | 0.79356204 | 16.344944 | -197.64724 | -182.430118 | 43.78488 |
| 4 | 5 | 1 | 1 | 0 | 1 | 1 | 0 | 42.40603 | 0.78997897 | 0.78437841 | 23.52256 | -190.90086 | -175.683736 | 45.15863 |
| 4 | 5 | 0 | 1 | 1 | 1 | 1 | 0 | 42.60986 | 0.78896947 | 0.78334199 | 24.332593 | -190.15761 | -174.940486 | 45.49287 |
| 4 | 5 | 1 | 1 | 0 | 1 | 0 | 1 | 42.65705 | 0.78873575 | 0.78310204 | 24.520131 | -189.98604 | -174.768917 | 45.72234 |
| 4 | 5 | 1 | 1 | 1 | 0 | 0 | 1 | 42.96752 | 0.78719811 | 0.7815234 | 25.753941 | -188.862 | -173.644873 | 46.5343 |
| 5 | 6 | 1 | 1 | 1 | 1 | 0 | 1 | 37.54846 | 0.81403665 | 0.80779627 | 6.218486 | -207.75786 | -189.497308 | 41.092 |
| 5 | 6 | 1 | 1 | 1 | 1 | 1 | 0 | 38.08522 | 0.81137832 | 0.80504873 | 8.351551 | -205.55783 | -187.297283 | 41.18554 |
| 5 | 6 | 0 | 1 | 1 | 1 | 1 | 1 | 40.57666 | 0.79903912 | 0.79229546 | 18.252625 | -195.73596 | -177.475406 | 44.40969 |
| 5 | 6 | 1 | 1 | 1 | 0 | 1 | 1 | 41.86144 | 0.79267609 | 0.78571891 | 23.358372 | -190.90428 | -172.643732 | 45.92332 |
| 5 | 6 | 1 | 1 | 0 | 1 | 1 | 1 | 41.88511 | 0.7925589 | 0.78559779 | 23.452401 | -190.8167 | -172.556148 | 45.56512 |
| 5 | 6 | 1 | 0 | 1 | 1 | 1 | 1 | 45.20833 | 0.77610021 | 0.7685868 | 36.658987 | -178.98227 | -160.72172 | 49.72588 |
| 6 | 7 | 1 | 1 | 1 | 1 | 1 | 1 | 37.24185 | 0.81555519 | 0.80807769 | 7 | -207.02875 | -185.724775 | 41.34232 |

Conclusion: Choose model:



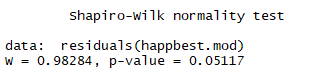
Residual plot:





From the above residual plots, it can be concluded that the distribution of residuals does not have a certain pattern, which indicates the non-constant and non-normal variances.

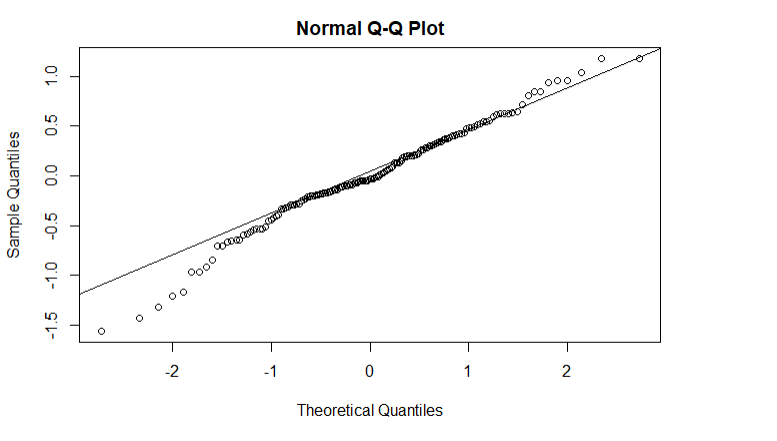
Shapiro-test:



p-value is slightly larger than 0.05, so the result is not perfect.

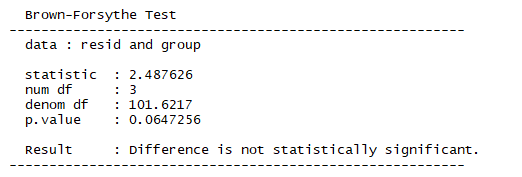
QQ-plot:

Hypothesis test:



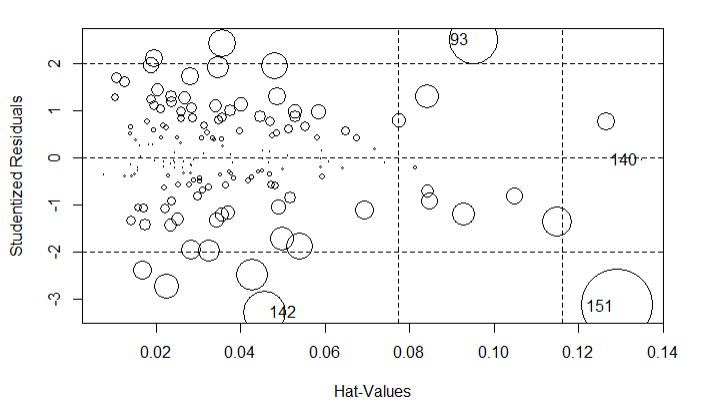
The results show that the qq plot slightly departures from the line, so the normality can still be hold.

BF-test:



Since p-value >0.05, it is evident that the departure from the model is not significant.

1. Outliers checking:





Therefore, the case with is considered to be the outlier regard to X.

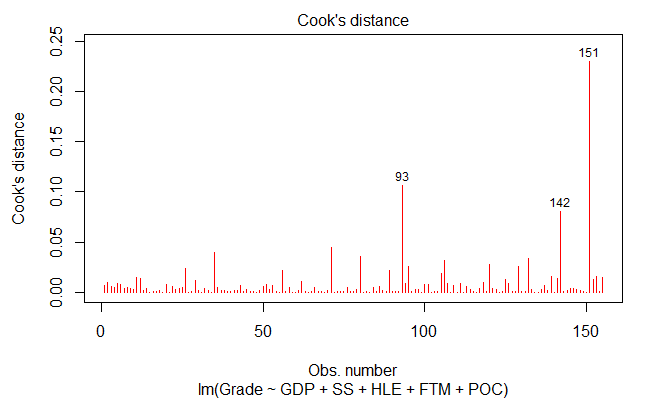
Therefore, it can be concluded from the table that the cases 93, 140 and 151 are considered as outliers regard to Y, since their hat matrix values are larger than .

1. Influential points:

i) Cook’s Distance:

The case i is considered as major influential point when

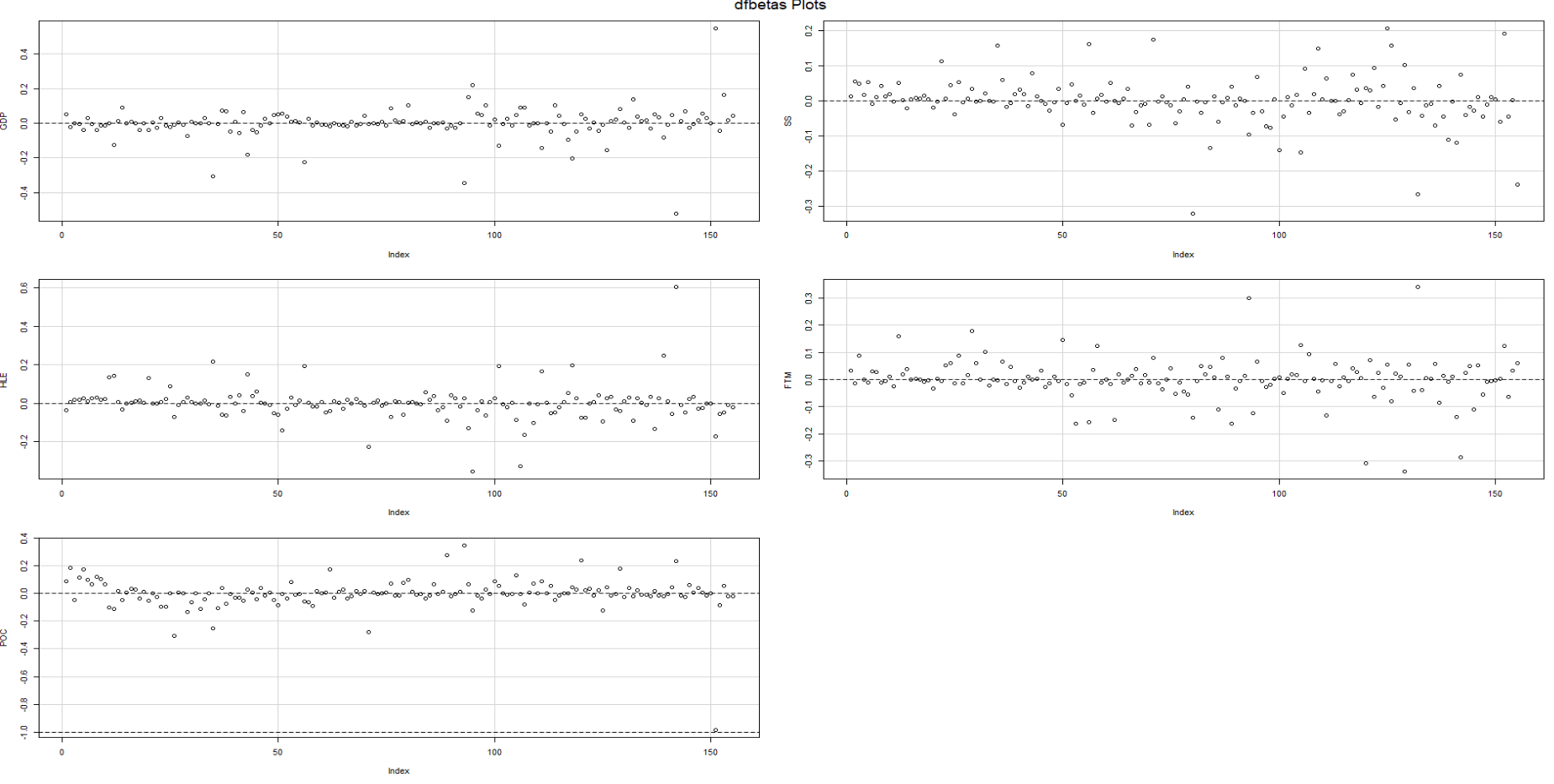
The case i is considered as minor influential point when , which is in the range of .



No influential points found using Cook’s Distance method.

ii) DEBETAS method:

A case i is considered has a large impact on the regression coefficients when



From the DFBETAS plots, there are many cases that are considered has a large impact on the regression coefficients in the data set, since its absolute

hw6

Haoran

11/09/2018

*#Problem 1***library**(ALSM)cosm <- **read.csv**(file="C:/Users/Haoran Zhang/Desktop/cosmetics.csv", header=TRUE,sep=",")cosm.mod <- **lm**(y**~**x1**+**x2**+**x3, data = cosm)**summary**(cosm.mod)

## ## Call:## lm(formula = y ~ x1 + x2 + x3, data = cosm)## ## Residuals:## Min 1Q Median 3Q Max ## -5.4217 -0.9115 0.0703 1.1420 3.5479 ## ## Coefficients:## Estimate Std. Error t value Pr(>|t|) ## (Intercept) 1.0233 1.2029 0.851 0.4000 ## x1 0.9657 0.7092 1.362 0.1809 ## x2 0.6292 0.7783 0.808 0.4237 ## x3 0.6760 0.3557 1.900 0.0646 .## ---## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1## ## Residual standard error: 1.825 on 40 degrees of freedom## Multiple R-squared: 0.7417, Adjusted R-squared: 0.7223 ## F-statistic: 38.28 on 3 and 40 DF, p-value: 7.821e-12

**anova**(cosm.mod)

## Analysis of Variance Table## ## Response: y## Df Sum Sq Mean Sq F value Pr(>F) ## x1 1 365.56 365.56 109.7054 4.994e-13 \*\*\*## x2 1 5.07 5.07 1.5215 0.22459 ## x3 1 12.03 12.03 3.6113 0.06461 . ## Residuals 40 133.29 3.33 ## ---## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

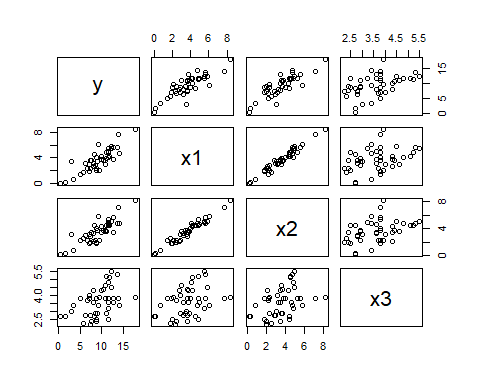
**anova**(**lm**(y**~**x1**+**x3**+**x2, data = cosm))

## Analysis of Variance Table## ## Response: y## Df Sum Sq Mean Sq F value Pr(>F) ## x1 1 365.56 365.56 109.7054 4.994e-13 \*\*\*## x3 1 14.93 14.93 4.4793 0.04058 \* ## x2 1 2.18 2.18 0.6535 0.42365 ## Residuals 40 133.29 3.33 ## ---## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

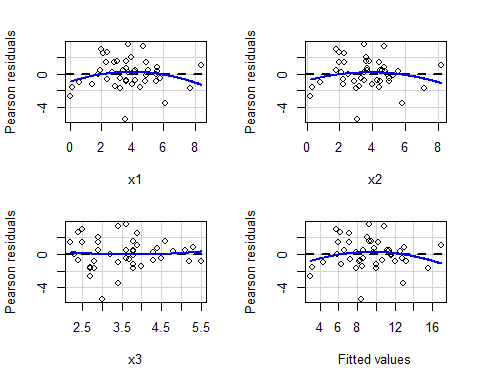
**anova**(**lm**(y**~**x2**+**x3**+**x1, data = cosm))

## Analysis of Variance Table## ## Response: y## Df Sum Sq Mean Sq F value Pr(>F) ## x2 1 366.21 366.21 109.901 4.864e-13 \*\*\*## x3 1 10.27 10.27 3.083 0.08677 . ## x1 1 6.18 6.18 1.854 0.18094 ## Residuals 40 133.29 3.33 ## ---## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##problem 2**plot**(cosm)

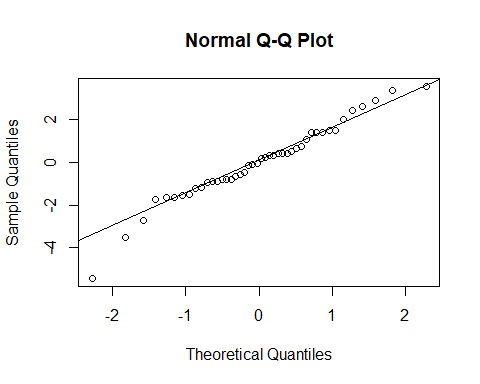


##residual plots##residual plot**library**(car)**residualPlots**(cosm.mod)

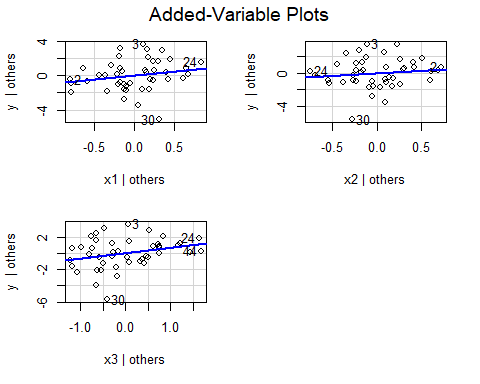


## Test stat Pr(>|Test stat|)## x1 -1.3151 0.1962## x2 -0.9895 0.3285## x3 0.3600 0.7208## Tukey test -1.1386 0.2549

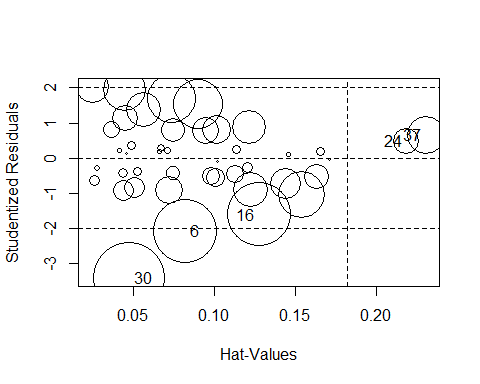
## qq plot**qqnorm**(**residuals**(cosm.mod))**qqline**(**residuals**(cosm.mod))



##Prolem 3**library**(car)**avPlots**(**lm**(y**~**x1**+**x2**+**x3, data=cosm))



##Problems 4 and 5 *#rstudent(cosm.mod)***influencePlot**(cosm.mod)



## StudRes Hat CookD## 6 -2.0977065 0.08230536 0.09093392## 16 -1.6182031 0.12790933 0.09228250## 24 0.4656507 0.21852908 0.01546125## 30 -3.4290260 0.04798828 0.11676914## 37 0.6567939 0.23044151 0.03275936

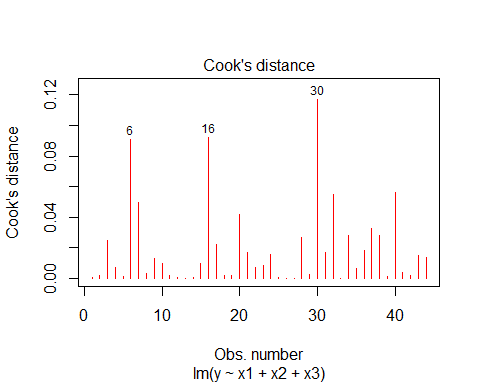
##Problem 6**qf**(0.5,4,40)

## [1] 0.8535658

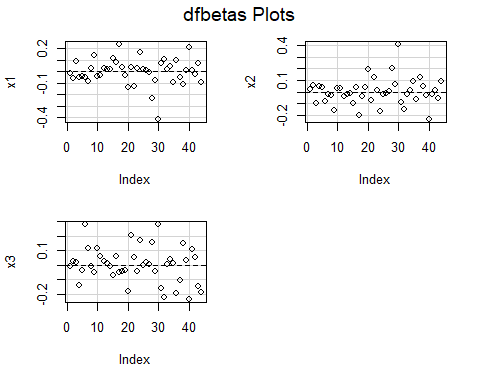
**qf**(0.2,4,40)

## [1] 0.4104948

**plot**(cosm.mod, pch=18, col="red", which=**c**(4))



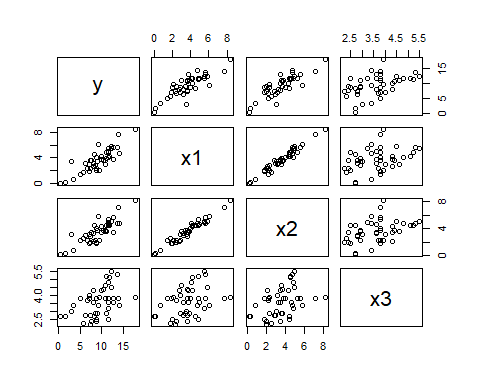
**dfbetasPlots**(cosm.mod)



**dffits**(cosm.mod)

## 1 2 3 4 5 ## 0.048723185 0.079810268 0.326252942 -0.167401791 0.062990268 ## 6 7 8 9 10 ## -0.628216989 -0.446432869 -0.107053378 0.226858483 -0.199902100 ## 11 12 13 14 15 ## 0.084946091 0.043746072 -0.024076766 0.045162815 -0.196397988 ## 16 17 18 19 20 ## -0.619731146 -0.296532257 -0.093253056 -0.087522091 0.421423513 ## 21 22 23 24 25 ## -0.260789454 -0.169314087 -0.182778688 0.246239942 -0.052786874 ## 26 27 28 29 30 ## 0.028610888 -0.006711388 0.325988563 -0.103091556 -0.769869697 ## 31 32 33 34 35 ## 0.257616869 0.477786365 -0.028534667 -0.334372258 0.160387735 ## 36 37 38 39 40 ## 0.270109372 0.359408756 0.340293409 0.074031305 0.483020886 ## 41 42 43 44 ## -0.127475601 0.084503326 0.246367746 -0.232631822

##Problem 7**plot**(cosm)## correlation matrix**library**("Hmisc")



**rcorr**(**as.matrix**(cosm))

## y x1 x2 x3## y 1.00 0.84 0.84 0.47## x1 0.84 1.00 0.97 0.38## x2 0.84 0.97 1.00 0.41## x3 0.47 0.38 0.41 1.00## ## n= 44 ## ## ## P## y x1 x2 x3 ## y 0.0000 0.0000 0.0012## x1 0.0000 0.0000 0.0119## x2 0.0000 0.0000 0.0057## x3 0.0012 0.0119 0.0057

##VIF**library**(fmsb)**VIF**(**lm**(x1**~**x2**+**x3, data=cosm))

## [1] 20.07203

**VIF**(**lm**(x2**~**x1**+**x3, data=cosm))

## [1] 20.7161

**VIF**(**lm**(x3**~**x1**+**x2, data=cosm))

## [1] 1.217973

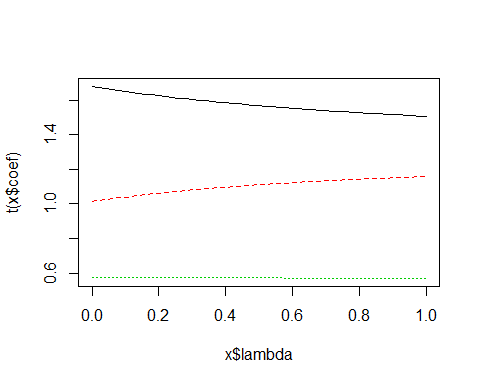
**VIF**(**lm**(x1**~**x2, data=cosm))

## [1] 19.80836

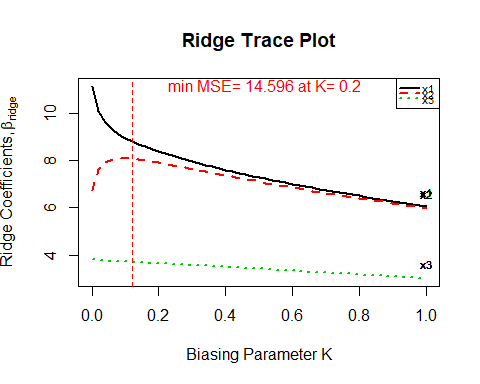
**VIF**(**lm**(x1**~**x3, data=cosm))

## [1] 1.164604

##Problem 8##Ridge Trace Plot**library**(MASS)**library**(lmridge)bridge.mod<-**lm.ridge**(y**~**x1**+**x2**+**x3, data=cosm, lamb=**seq**(0,1,0.02))**plot**(bridge.mod)



*#select(bridge.mod)*##VIF factorsmod2<-**lmridge**(y**~**x1**+**x2**+**x3,data=**as.data.frame**(cosm),K=**seq**(0,1,0.02))**plot**(mod2)



**vif**(mod2)

## x1 x2 x3## k=0 20.07203 20.71610 1.21797## k=0.02 6.37329 6.55221 1.13571## k=0.04 3.19884 3.27135 1.07626## k=0.06 1.98298 2.01574 1.02414## k=0.08 1.38944 1.40355 0.97663## k=0.1 1.05423 1.05838 0.93277## k=0.12 0.84536 0.84377 0.89203## k=0.14 0.70561 0.70054 0.85405## k=0.16 0.60687 0.59964 0.81857## k=0.18 0.53405 0.52547 0.78534## k=0.2 0.47845 0.46903 0.75418## k=0.22 0.43474 0.42483 0.72489## k=0.24 0.39954 0.38937 0.69734## k=0.26 0.37060 0.36032 0.67139## k=0.28 0.34637 0.33610 0.64690## k=0.3 0.32577 0.31559 0.62378## k=0.32 0.30802 0.29798 0.60191## k=0.34 0.29254 0.28268 0.58121## k=0.36 0.27890 0.26924 0.56159## k=0.38 0.26677 0.25733 0.54299## k=0.4 0.25588 0.24668 0.52532## k=0.42 0.24605 0.23709 0.50853## k=0.44 0.23711 0.22838 0.49255## k=0.46 0.22893 0.22044 0.47734## k=0.48 0.22141 0.21316 0.46285## k=0.5 0.21445 0.20644 0.44903## k=0.52 0.20800 0.20022 0.43584## k=0.54 0.20199 0.19443 0.42324## k=0.56 0.19636 0.18903 0.41119## k=0.58 0.19109 0.18396 0.39967## k=0.6 0.18612 0.17921 0.38865## k=0.62 0.18144 0.17472 0.37808## k=0.64 0.17701 0.17049 0.36796## k=0.66 0.17281 0.16647 0.35825## k=0.68 0.16882 0.16266 0.34893## k=0.7 0.16502 0.15904 0.33998## k=0.72 0.16139 0.15559 0.33138## k=0.74 0.15793 0.15229 0.32312## k=0.76 0.15462 0.14914 0.31517## k=0.78 0.15145 0.14612 0.30752## k=0.8 0.14841 0.14323 0.30015## k=0.82 0.14549 0.14045 0.29305## k=0.84 0.14268 0.13777 0.28621## k=0.86 0.13997 0.13520 0.27961## k=0.88 0.13737 0.13272 0.27325## k=0.9 0.13485 0.13033 0.26711## k=0.92 0.13242 0.12803 0.26118## k=0.94 0.13008 0.12580 0.25545## k=0.96 0.12781 0.12364 0.24991## k=0.98 0.12562 0.12155 0.24455## k=1 0.12349 0.11953 0.23938

##model summary**summary**(**lmridge**(y**~**x1**+**x2**+**x3,data=**as.data.frame**(cosm),K=0.1))

## ## Call:## lmridge.default(formula = y ~ x1 + x2 + x3, data = as.data.frame(cosm), ## K = 0.1)## ## ## Coefficients: for Ridge parameter K= 0.1 ## Estimate Estimate (Sc) StdErr (Sc) t-value (Sc) Pr(>|t|) ## Intercept 1.3330 -66.3834 11.4066 -5.8197 <2e-16 \*\*\*## x1 0.7763 8.9516 1.8427 4.8578 <2e-16 \*\*\*## x2 0.7571 8.0816 1.8464 4.3770 0.0001 \*\*\*## x3 0.6565 3.7180 1.7333 2.1450 0.0379 \* ## ---## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1## ## Ridge Summary## R2 adj-R2 DF ridge F AIC BIC ## 0.67820 0.66250 2.03833 39.59998 53.13602 223.27713 ## Ridge minimum MSE= 16.45525 at K= 0.1 ## P-value for F-test ( 2.03833 , 41.65713 ) = 1.901261e-10 ## -------------------------------------------------------------------

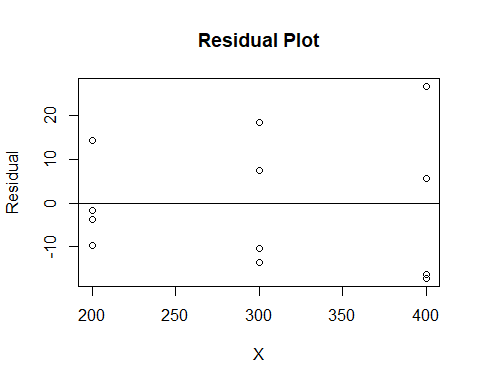
**summary**(**lmridge**(y**~**x1**+**x2**+**x3,data=**as.data.frame**(cosm),K=1))

## ## Call:## lmridge.default(formula = y ~ x1 + x2 + x3, data = as.data.frame(cosm), ## K = 1)## ## ## Coefficients: for Ridge parameter K= 1 ## Estimate Estimate (Sc) StdErr (Sc) t-value (Sc) Pr(>|t|) ## Intercept 3.4107 -45.6537 5.0856 -8.9770 <2e-16 \*\*\*## x1 0.5268 6.0742 0.7056 8.6089 <2e-16 \*\*\*## x2 0.5612 5.9910 0.6942 8.6305 <2e-16 \*\*\*## x3 0.5323 3.0143 0.9824 3.0684 0.0038 \*\* ## ---## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1## ## Ridge Summary## R2 adj-R2 DF ridge F AIC BIC ## 0.35150 0.31990 1.14248 31.64013 61.96997 230.51272 ## Ridge minimum MSE= 28.7512 at K= 1 ## P-value for F-test ( 1.14248 , 42.37514 ) = 4.958206e-07 ## -------------------------------------------------------------------

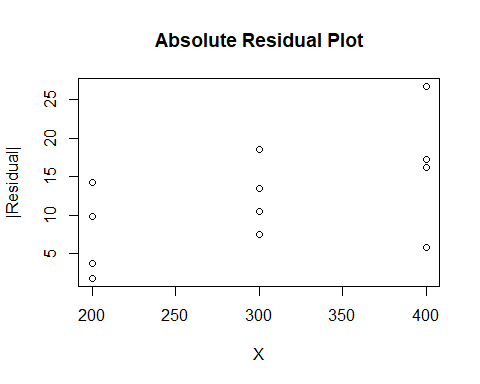
##Problem 9##a)**library**(ALSM)machine <- **read.csv**(file="C:/Users/Haoran Zhang/Desktop/machine.csv", header=TRUE,sep=",")machine.mod <- **lm**(y**~**x, data = machine)**summary**(machine.mod)

## ## Call:## lm(formula = y ~ x, data = machine)## ## Residuals:## Min 1Q Median 3Q Max ## -17.250 -11.250 -2.750 9.188 26.750 ## ## Coefficients:## Estimate Std. Error t value Pr(>|t|) ## (Intercept) -5.75000 16.73052 -0.344 0.73820 ## x 0.18750 0.05381 3.484 0.00588 \*\*## ---## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1## ## Residual standard error: 15.22 on 10 degrees of freedom## Multiple R-squared: 0.5484, Adjusted R-squared: 0.5032 ## F-statistic: 12.14 on 1 and 10 DF, p-value: 0.005878

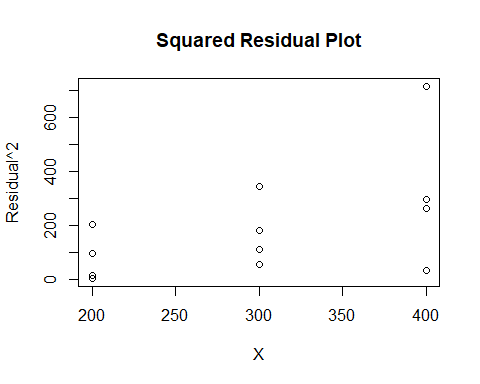
*# residual plot*resid<-**residuals**(machine.mod)**plot**(machine**$**x,resid, main="Residual Plot", xlab="X", ylab="Residual")**abline**(h=0)



abs\_resid<-**abs**(resid)**plot**(machine**$**x,abs\_resid, main="Absolute Residual Plot", xlab="X", ylab="|Residual|")**abline**(h=0)



sq\_resid<-resid**^**2**plot**(machine**$**x,sq\_resid, main="Squared Residual Plot", xlab="X", ylab="Residual^2")



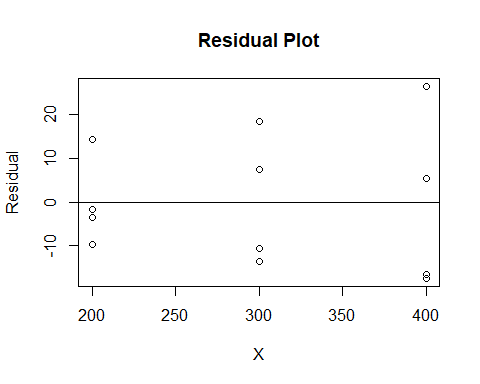
##c)wts<-1**/fitted**(**lm**(**abs**(**residuals**(machine.mod))**~**x,machine))**^**2wts

## 1 2 3 4 5 6 ## 0.017485144 0.003591017 0.006801998 0.003591017 0.017485144 0.006801998 ## 7 8 9 10 11 12 ## 0.006801998 0.003591017 0.017485144 0.003591017 0.017485144 0.006801998

machine2.mod<-**lm**(y**~**x,weight=wts,data=machine)**summary**(machine2.mod)

## ## Call:## lm(formula = y ~ x, data = machine, weights = wts)## ## Weighted Residuals:## Min 1Q Median 3Q Max ## -1.2685 -1.0009 -0.3429 0.8435 1.9051 ## ## Coefficients:## Estimate Std. Error t value Pr(>|t|) ## (Intercept) -6.26646 13.35177 -0.469 0.64890 ## x 0.18930 0.05133 3.688 0.00419 \*\*## ---## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1## ## Residual standard error: 1.221 on 10 degrees of freedom## Multiple R-squared: 0.5763, Adjusted R-squared: 0.5339 ## F-statistic: 13.6 on 1 and 10 DF, p-value: 0.004192

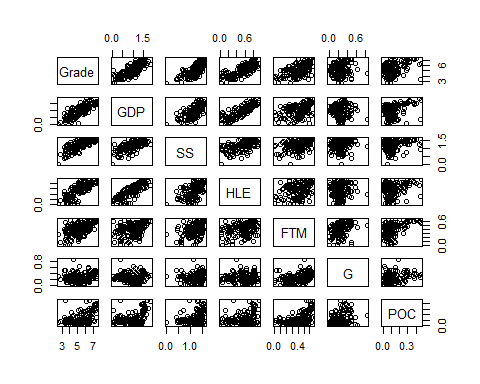
##d)resid<-**residuals**(machine2.mod)**plot**(machine**$**x,resid, main="Residual Plot", xlab="X", ylab="Residual")**abline**(h=0)



##Problem 10##Team Project *#Read the data***library**(ALSM)**library**(dplyr)happ<- **read.csv**(file="C:/Users/Haoran Zhang/Desktop/STAT512/happiness.csv", header=TRUE,sep=",")**colnames**(happ)<-**c**("WP5","Country","year","Grade","GDP", "SS","HLE","FTM","G","POC")##select useful datamyvars <- **c**("Grade","GDP","SS","HLE","FTM","G","POC")happdata <- happ[myvars]##linear modelhapp.mod <- **lm**(Grade**~**GDP**+**SS**+**HLE**+**FTM**+**POC, data = happdata)**summary**(happ.mod)

## ## Call:## lm(formula = Grade ~ GDP + SS + HLE + FTM + POC, data = happdata)## ## Residuals:## Min 1Q Median 3Q Max ## -1.55569 -0.23695 -0.03398 0.32770 1.17996 ## ## Coefficients:## Estimate Std. Error t value Pr(>|t|) ## (Intercept) 1.6521 0.1811 9.121 4.79e-16 \*\*\*## GDP 0.6973 0.2004 3.480 0.000658 \*\*\*## SS 1.1466 0.2030 5.649 7.94e-08 \*\*\*## HLE 1.4392 0.3196 4.502 1.35e-05 \*\*\*## FTM 1.5568 0.3357 4.637 7.66e-06 \*\*\*## POC 1.0069 0.4752 2.119 0.035742 \* ## ---## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1## ## Residual standard error: 0.502 on 149 degrees of freedom## Multiple R-squared: 0.814, Adjusted R-squared: 0.8078 ## F-statistic: 130.4 on 5 and 149 DF, p-value: < 2.2e-16

## correlation matrix**plot**(happdata)



**library**("Hmisc")**rcorr**(**as.matrix**(happdata))

## Grade GDP SS HLE FTM G POC## Grade 1.00 0.81 0.75 0.79 0.57 0.15 0.43## GDP 0.81 1.00 0.69 0.84 0.37 -0.02 0.35## SS 0.75 0.69 1.00 0.61 0.43 0.05 0.23## HLE 0.79 0.84 0.61 1.00 0.35 0.06 0.28## FTM 0.57 0.37 0.43 0.35 1.00 0.32 0.50## G 0.15 -0.02 0.05 0.06 0.32 1.00 0.29## POC 0.43 0.35 0.23 0.28 0.50 0.29 1.00## ## n= 155 ## ## ## P## Grade GDP SS HLE FTM G POC ## Grade 0.0000 0.0000 0.0000 0.0000 0.0543 0.0000## GDP 0.0000 0.0000 0.0000 0.0000 0.8136 0.0000## SS 0.0000 0.0000 0.0000 0.0000 0.5240 0.0037## HLE 0.0000 0.0000 0.0000 0.0000 0.4337 0.0004## FTM 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000## G 0.0543 0.8136 0.5240 0.4337 0.0000 0.0002## POC 0.0000 0.0000 0.0037 0.0004 0.0000 0.0002

## parametersmeanGDP<-**sum**(happdata**$**GDP)**/length**(happdata**$**GDP)meanSS<-**sum**(happdata**$**SS)**/length**(happdata**$**SS)meanHLE<-**sum**(happdata**$**HLE)**/length**(happdata**$**HLE)meanFTM<-**sum**(happdata**$**FTM)**/length**(happdata**$**FTM)meanG<-**sum**(happdata**$**G)**/length**(happdata**$**G)meanPOC<-**sum**(happdata**$**POC)**/length**(happdata**$**POC)sdGDP<-**sd**(happdata**$**GDP)sdSS<-**sd**(happdata**$**SS)sdHLE<-**sd**(happdata**$**HLE)sdFTM<-**sd**(happdata**$**FTM)sdG<-**sd**(happdata**$**G)sdPOC<-**sd**(happdata**$**POC)##standardarize the data pointshappdata**$**GDP<-(happdata**$**GDP**-**meanGDP)**/**sdGDPhappdata**$**SS<-(happdata**$**SS**-**meanSS)**/**sdSShappdata**$**HLE<-(happdata**$**HLE**-**meanHLE)**/**sdHLEhappdata**$**FTM<-(happdata**$**FTM**-**meanFTM)**/**sdFTMhappdata**$**G<-(happdata**$**G**-**meanG)**/**sdGhappdata**$**POC<-(happdata**$**POC**-**meanPOC)**/**sdPOC##check correlation again**library**("Hmisc")**rcorr**(**as.matrix**(happdata))

## Grade GDP SS HLE FTM G POC## Grade 1.00 0.81 0.75 0.79 0.57 0.15 0.43## GDP 0.81 1.00 0.69 0.84 0.37 -0.02 0.35## SS 0.75 0.69 1.00 0.61 0.43 0.05 0.23## HLE 0.79 0.84 0.61 1.00 0.35 0.06 0.28## FTM 0.57 0.37 0.43 0.35 1.00 0.32 0.50## G 0.15 -0.02 0.05 0.06 0.32 1.00 0.29## POC 0.43 0.35 0.23 0.28 0.50 0.29 1.00## ## n= 155 ## ## ## P## Grade GDP SS HLE FTM G POC ## Grade 0.0000 0.0000 0.0000 0.0000 0.0543 0.0000## GDP 0.0000 0.0000 0.0000 0.0000 0.8136 0.0000## SS 0.0000 0.0000 0.0000 0.0000 0.5240 0.0037## HLE 0.0000 0.0000 0.0000 0.0000 0.4337 0.0004## FTM 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000## G 0.0543 0.8136 0.5240 0.4337 0.0000 0.0002## POC 0.0000 0.0000 0.0037 0.0004 0.0000 0.0002

##build new modelhapp.mod <- **lm**(Grade**~**GDP**+**SS**+**HLE**+**FTM**+**POC, data = happdata)##summary(happ.mod)## find out any potential nonlinear relationship *# plot(happdata$GDP,happdata$Grade,main="Grade vs. GDP",ylab="Grade",xlab="Grade")  
# plot(happdata$SS,happdata$Grade,main="Grade vs. SS",ylab="Grade",xlab="SS")  
# plot(happdata$HLE,happdata$Grade,main="Grade vs. HLE",ylab="Grade",xlab="HLE")  
# plot(happdata$FTM,happdata$Grade,main="Grade vs. FTM",ylab="Grade",xlab="FTM")  
# plot(happdata$G,(happdata$Grade),main="Grade vs. G",ylab="Grade",xlab="G")  
# plot(happdata$POC,(happdata$Grade),main="Grade vs. POC",ylab="Grade",xlab="POC")*## Marginal variation reducion##POC**anova**(happ.mod)

## Analysis of Variance Table## ## Response: Grade## Df Sum Sq Mean Sq F value Pr(>F) ## GDP 1 133.220 133.220 528.6434 < 2.2e-16 \*\*\*## SS 1 14.542 14.542 57.7057 3.113e-12 \*\*\*## HLE 1 5.674 5.674 22.5139 4.845e-06 \*\*\*## FTM 1 9.798 9.798 38.8797 4.412e-09 \*\*\*## POC 1 1.132 1.132 4.4905 0.03574 \* ## Residuals 149 37.548 0.252 ## ---## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##GhappG.mod <- **lm**(Grade**~**GDP**+**SS**+**HLE**+**FTM**+**POC**+**G, data = happdata)**anova**(happG.mod)

## Analysis of Variance Table## ## Response: Grade## Df Sum Sq Mean Sq F value Pr(>F) ## GDP 1 133.220 133.220 529.4186 < 2.2e-16 \*\*\*## SS 1 14.542 14.542 57.7903 3.095e-12 \*\*\*## HLE 1 5.674 5.674 22.5469 4.798e-06 \*\*\*## FTM 1 9.798 9.798 38.9367 4.368e-09 \*\*\*## POC 1 1.132 1.132 4.4971 0.03562 \* ## G 1 0.307 0.307 1.2185 0.27145 ## Residuals 148 37.242 0.252 ## ---## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##FTMhappFTM.mod <- **lm**(Grade**~**GDP**+**SS**+**HLE**+**G**+**POC**+**FTM, data = happdata)**anova**(happFTM.mod)

## Analysis of Variance Table## ## Response: Grade## Df Sum Sq Mean Sq F value Pr(>F) ## GDP 1 133.220 133.220 529.419 < 2.2e-16 \*\*\*## SS 1 14.542 14.542 57.790 3.095e-12 \*\*\*## HLE 1 5.674 5.674 22.547 4.798e-06 \*\*\*## G 1 3.135 3.135 12.460 0.0005544 \*\*\*## POC 1 3.481 3.481 13.834 0.0002828 \*\*\*## FTM 1 4.620 4.620 18.358 3.280e-05 \*\*\*## Residuals 148 37.242 0.252 ## ---## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##HLEhappHLE.mod <- **lm**(Grade**~**GDP**+**SS**+**FTM**+**G**+**POC**+**HLE, data = happdata)**anova**(happHLE.mod)

## Analysis of Variance Table## ## Response: Grade## Df Sum Sq Mean Sq F value Pr(>F) ## GDP 1 133.220 133.220 529.4186 < 2.2e-16 \*\*\*## SS 1 14.542 14.542 57.7903 3.095e-12 \*\*\*## FTM 1 10.642 10.642 42.2934 1.135e-09 \*\*\*## G 1 1.103 1.103 4.3832 0.0380 \* ## POC 1 0.521 0.521 2.0702 0.1523 ## HLE 1 4.643 4.643 18.4524 3.139e-05 \*\*\*## Residuals 148 37.242 0.252 ## ---## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##SShappSS.mod <- **lm**(Grade**~**GDP**+**HLE**+**FTM**+**G**+**POC**+**SS, data = happdata)**anova**(happSS.mod)

## Analysis of Variance Table## ## Response: Grade## Df Sum Sq Mean Sq F value Pr(>F) ## GDP 1 133.220 133.220 529.4186 < 2.2e-16 \*\*\*## HLE 1 7.203 7.203 28.6262 3.271e-07 \*\*\*## FTM 1 15.428 15.428 61.3131 8.638e-13 \*\*\*## G 1 0.566 0.566 2.2474 0.1360 ## POC 1 0.288 0.288 1.1439 0.2866 ## SS 1 7.966 7.966 31.6590 8.924e-08 \*\*\*## Residuals 148 37.242 0.252 ## ---## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##GDPhappGDP.mod <- **lm**(Grade**~**SS**+**HLE**+**FTM**+**G**+**POC**+**GDP, data = happdata)**anova**(happGDP.mod)

## Analysis of Variance Table## ## Response: Grade## Df Sum Sq Mean Sq F value Pr(>F) ## SS 1 114.733 114.733 455.9514 < 2.2e-16 \*\*\*## HLE 1 34.111 34.111 135.5576 < 2.2e-16 \*\*\*## FTM 1 10.298 10.298 40.9261 1.959e-09 \*\*\*## G 1 0.161 0.161 0.6403 0.4248755 ## POC 1 2.033 2.033 8.0800 0.0051083 \*\* ## GDP 1 3.335 3.335 13.2526 0.0003756 \*\*\*## Residuals 148 37.242 0.252 ## ---## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##Bestsub**library**("ALSM")bs<-**BestSub**(happdata[,2**:**7],happdata**$**Grade,num=6)bs

## p 1 2 3 4 5 6 SSEp r2 r2.adj Cp AICp## 1 2 1 0 0 0 0 0 68.69347 0.65978721 0.65756359 121.989477 -122.13450## 1 2 0 0 1 0 0 0 77.03301 0.61848464 0.61599107 155.130991 -104.37463## 1 2 0 1 0 0 0 0 87.18030 0.56822892 0.56540688 195.456585 -85.19424## 1 2 0 0 0 1 0 0 136.34258 0.32474673 0.32033331 390.828649 -15.87943## 1 2 0 0 0 0 0 1 164.35982 0.18598791 0.18066757 502.169834 13.08811## 1 2 0 0 0 0 1 0 197.06837 0.02399485 0.01761573 632.154370 41.21947## 2 3 1 0 0 1 0 0 51.71168 0.74389162 0.74052177 56.503423 -164.14994## 2 3 0 1 1 0 0 0 53.06939 0.73716739 0.73370906 61.898999 -160.13286## 2 3 1 1 0 0 0 0 54.15146 0.73180829 0.72827945 66.199181 -157.00423## 2 3 0 0 1 1 0 0 57.05143 0.71744581 0.71372799 77.723749 -148.91815## 2 3 1 0 1 0 0 0 61.49015 0.69546254 0.69145547 95.363305 -137.30496## 2 3 1 0 0 0 1 0 62.82770 0.68883816 0.68474393 100.678758 -133.96951## 3 4 0 1 1 1 0 0 42.77099 0.78817146 0.78396295 22.972918 -191.57259## 3 4 1 1 0 1 0 0 43.50899 0.78451643 0.78023530 25.905748 -188.92092## 3 4 0 1 1 0 0 1 45.54102 0.77445255 0.76997147 33.981087 -181.84581## 3 4 1 0 1 1 0 0 46.06168 0.77187390 0.76734160 36.050212 -180.08378## 3 4 1 1 1 0 0 0 48.47789 0.75990734 0.75513729 45.652278 -172.15918## 3 4 1 1 0 0 0 1 48.98666 0.75738760 0.75256748 47.674142 -170.54095## 4 5 1 1 1 1 0 0 38.68008 0.80843218 0.80332371 8.715555 -205.15556## 4 5 0 1 1 1 0 1 40.59990 0.79892407 0.79356204 16.344944 -197.64724## 4 5 1 1 0 1 1 0 42.40603 0.78997897 0.78437841 23.522560 -190.90086## 4 5 0 1 1 1 1 0 42.60986 0.78896947 0.78334199 24.332593 -190.15761## 4 5 1 1 0 1 0 1 42.65705 0.78873575 0.78310204 24.520131 -189.98604## 4 5 1 1 1 0 0 1 42.96752 0.78719811 0.78152340 25.753941 -188.86200## 5 6 1 1 1 1 0 1 37.54846 0.81403665 0.80779627 6.218486 -207.75786## 5 6 1 1 1 1 1 0 38.08522 0.81137832 0.80504873 8.351551 -205.55783## 5 6 0 1 1 1 1 1 40.57666 0.79903912 0.79229546 18.252625 -195.73596## 5 6 1 1 1 0 1 1 41.86144 0.79267609 0.78571891 23.358372 -190.90428## 5 6 1 1 0 1 1 1 41.88511 0.79255890 0.78559779 23.452401 -190.81670## 5 6 1 0 1 1 1 1 45.20833 0.77610021 0.76858680 36.658987 -178.98227## 6 7 1 1 1 1 1 1 37.24185 0.81555519 0.80807769 7.000000 -207.02875## SBCp PRESSp## 1 -116.047650 70.56867## 1 -98.287775 79.29513## 1 -79.107388 89.45355## 1 -9.792584 139.71373## 1 19.174959 169.80370## 1 47.306320 202.64514## 2 -155.019661 53.74846## 2 -151.002580 55.16945## 2 -147.873955 56.14518## 2 -139.787870 59.55631## 2 -128.174687 63.98477## 2 -124.839234 65.39532## 3 -179.398889 45.02872## 3 -176.747224 45.63701## 3 -169.672111 48.61169## 3 -167.910076 48.69907## 3 -159.985484 51.00115## 3 -158.367254 51.99260## 4 -189.938431 41.25820## 4 -182.430118 43.78488## 4 -175.683736 45.15863## 4 -174.940486 45.49287## 4 -174.768917 45.72234## 4 -173.644873 46.53430## 5 -189.497308 41.09200## 5 -187.297283 41.18554## 5 -177.475406 44.40969## 5 -172.643732 45.92332## 5 -172.556148 45.56512## 5 -160.721720 49.72588## 6 -185.724775 41.34232

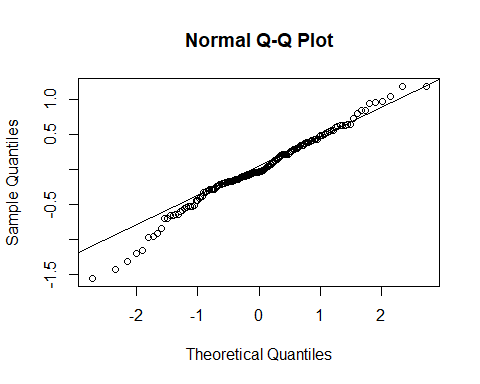
##best model?happbest.mod <- **lm**(Grade**~**GDP**+**SS**+**HLE**+**FTM**+**POC, data = happdata)**summary**(happbest.mod)

## ## Call:## lm(formula = Grade ~ GDP + SS + HLE + FTM + POC, data = happdata)## ## Residuals:## Min 1Q Median 3Q Max ## -1.55569 -0.23695 -0.03398 0.32770 1.17996 ## ## Coefficients:## Estimate Std. Error t value Pr(>|t|) ## (Intercept) 5.25570 0.04032 130.345 < 2e-16 \*\*\*## GDP 0.29344 0.08433 3.480 0.000658 \*\*\*## SS 0.32935 0.05830 5.649 7.94e-08 \*\*\*## HLE 0.34114 0.07577 4.502 1.35e-05 \*\*\*## FTM 0.23352 0.05036 4.637 7.66e-06 \*\*\*## POC 0.10233 0.04829 2.119 0.035742 \* ## ---## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1## ## Residual standard error: 0.502 on 149 degrees of freedom## Multiple R-squared: 0.814, Adjusted R-squared: 0.8078 ## F-statistic: 130.4 on 5 and 149 DF, p-value: < 2.2e-16

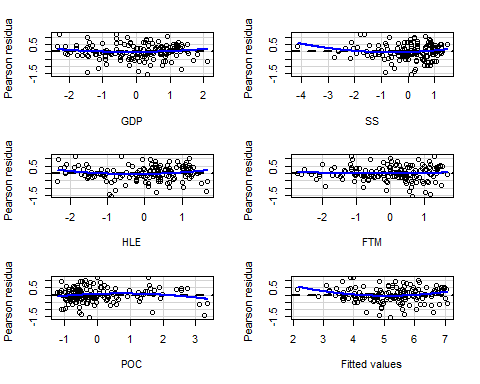
##F-test for lack of fit(data too large for computing) *# happbestR.mod<-happbest.mod  
# happbestF.mod<-lm(Grade~factor(GDP)\*factor(SS)\*factor(HLE)\*factor(FTM), data = happdata)  
# anova(happbestR.mod, happbestF.mod) # lack of fit test*## check random error## shapiro method**shapiro.test**(**residuals**(happbest.mod))

## ## Shapiro-Wilk normality test## ## data: residuals(happbest.mod)## W = 0.98284, p-value = 0.05117

## qq plot**qqnorm**(**residuals**(happbest.mod))**qqline**(**residuals**(happbest.mod))



##residual plot**library**(car)**residualPlots**(happbest.mod)



## Test stat Pr(>|Test stat|) ## GDP 1.3532 0.178061 ## SS 1.9027 0.059022 . ## HLE 1.8941 0.060156 . ## FTM 0.4442 0.657555 ## POC -1.5860 0.114864 ## Tukey test 2.9183 0.003519 \*\*## ---## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##bf-test**library**(onewaytests)happdata**$**fit<-happbest.mod**$**fitted.valueshappdata**$**resid<-happbest.mod**$**residualshappdata**$**group<-**cut**(happdata**$**fit, 4)**bf.test**(resid**~**group, happdata)

## ## Brown-Forsythe Test ## --------------------------------------------------------- ## data : resid and group ## ## statistic : 2.487626 ## num df : 3 ## denom df : 101.6217 ## p.value : 0.0647256 ## ## Result : Difference is not statistically significant. ## ---------------------------------------------------------

##VIF test**library**(fmsb)**VIF**(**lm**(GDP**~**SS**+**HLE**+**FTM**+**G**+**POC, data=happ))

## [1] 4.570513

**VIF**(**lm**(GDP**~**SS**+**HLE, data=happ))

## [1] 4.139066

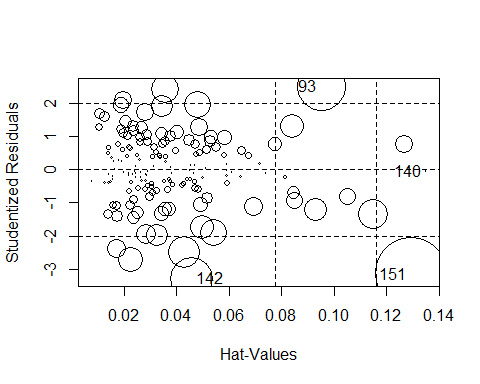
**VIF**(**lm**(SS**~**GDP**+**HLE, data=happ))

## [1] 1.913263

**VIF**(**lm**(HLE**~**GDP**+**SS, data=happ))

## [1] 3.482567

##Outliers *#rstudent(cosm.mod)***influencePlot**(happbest.mod)



## StudRes Hat CookD## 93 2.51466235 0.09506413 1.068959e-01## 140 -0.04415001 0.13501283 5.104986e-05## 142 -3.27427625 0.04576025 8.043822e-02## 151 -3.13876259 0.12929090 2.301427e-01

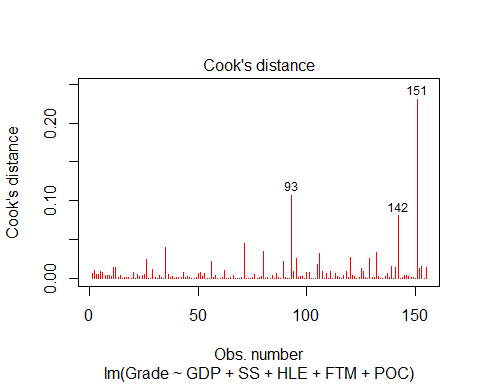
**qf**(0.5,6,149)

## [1] 0.8953978

**qf**(0.2,6,149)

## [1] 0.5101136

**plot**(happbest.mod, pch=18, col="red", which=**c**(4))



**dfbetasPlots**(happbest.mod)

