lnTest3=log(Test3)

lnTest1=log(Test1)

lnTest2=log(Test2)

lnJobs=log(jobs)

> model3 = lm(lnTest3~lnTest1+lnTest2 +Gender+Year + GPA + CrHrs + Stick + ClassRow + CokePepsi + siblings + countries + lnJobs + DogCat)

> summary(model3)

Call:

lm(formula = lnTest3 ~ lnTest1 + lnTest2 + Gender + Year + GPA +

CrHrs + Stick + ClassRow + CokePepsi + siblings + countries +

lnJobs + DogCat)

Residuals:

Min 1Q Median 3Q Max

-0.120837 -0.030567 -0.008703 0.036776 0.147989

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -0.609449 1.409504 -0.432 0.6706

lnTest1 0.510082 0.102099 4.996 9.37e-05 \*\*\*

lnTest2 0.656373 0.288213 2.277 0.0352 \*

GenderMale 0.075906 0.045851 1.656 0.1152

YearSenior -0.030631 0.073936 -0.414 0.6836

YearSophomore -0.011788 0.047089 -0.250 0.8052

GPA -0.029660 0.059698 -0.497 0.6253

CrHrs -0.003457 0.013369 -0.259 0.7989

Stickno 0.011018 0.108914 0.101 0.9205

Stickyes 0.052118 0.035389 1.473 0.1581

Stickyes 0.088869 0.106996 0.831 0.4171

ClassRow 0.017306 0.013284 1.303 0.2091

CokePepsiNeither -0.054505 0.049163 -1.109 0.2822

CokePepsiPepsi 0.017488 0.054565 0.320 0.7523

siblings 0.017975 0.018863 0.953 0.3533

countries 0.002129 0.003456 0.616 0.5455

lnJobs -0.037230 0.047797 -0.779 0.4461

DogCatDog -0.043274 0.046359 -0.933 0.3629

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.08245 on 18 degrees of freedom

Multiple R-squared: 0.8135, Adjusted R-squared: 0.6374

F-statistic: 4.619 on 17 and 18 DF, p-value: 0.001168

> attach(model3)

> layout(matrix(c(1,2,3,4,5,6,7,8,9,10,11,12),byrow=TRUE,ncol=6))

> plot.new()

> hist(lnTest1)

> hist(lnTest2)

> hist(GPA)

> hist(CrHrs)

> hist(lnJobs)

> hist(lnTest3)

> text(lnTest1,lnTest3,labels=row.names(test),pos=1)

> plot(lnTest1,lnTest3)

> text(lnTest2,lnTest3,labels=row.names(test),pos=1)

> plot(lnTest2,lnTest3)

> text(GPA,lnTest3,labels=row.names(test),pos=1)

> plot(GPA,lnTest3)

> text(CrHrs,lnTest3,labels=row.names(test),pos=1)

> plot(CrHrs,lnTest3)

> text(lnJobs,lnTest3,labels=row.names(test),pos=1)

> plot(lnJobs,lnTest3)



> rstandard = rstandard(model3)

> rstandard[order(rstandard)]

36 25 23 28 27 2

-2.3205355 -1.7019673 -1.3610458 -1.2803149 -1.2552110 -0.8279250

34 11 21 22 16 9

-0.6719327 -0.5357281 -0.5220765 -0.4617988 -0.4279533 -0.3546848

15 7 26 19 14 3

-0.3403000 -0.3054213 -0.2543326 -0.2292508 -0.2083302 -0.1945691

33 30 6 5 35 13

-0.0788680 0.2428613 0.3000881 0.3532438 0.3546848 0.3941280

31 18 10 4 1 29

0.5448429 0.6858368 0.7239280 0.7848302 0.8665941 0.9839627

32 17 8 24 12 20

1.4230926 1.5731217 2.1061026 2.2947195 NaN NaN

> #36 12 and20 should be removed

> leverages = hatvalues(model1)

> leverages[order(leverages)]

25 18 31 23 4 22 3

0.2600099 0.2994886 0.3288605 0.3342199 0.3390779 0.3456842 0.3667841

14 1 24 26 7 28 33

0.3775466 0.3967148 0.4076856 0.4105836 0.4131272 0.4203401 0.4242876

29 17 15 16 21 2 30

0.4327955 0.4350301 0.4368414 0.4539963 0.4555418 0.4896144 0.5008589

13 11 36 8 32 5 27

0.5035896 0.5227079 0.5488438 0.5561148 0.5601393 0.5601865 0.5690577

19 34 6 10 9 35 12

0.5744037 0.5766673 0.6064701 0.6141480 0.7392910 0.7392910 1.0000000

20

1.0000000

#high leverage cut is 3(16+1)/36

#1.417

> #none is excluded

>

> cooks = cooks.distance(model1)

> cooks[order(cooks)]

3 14 19 33 7

3.320575e-05 8.831313e-05 1.697672e-03 2.453968e-03 2.659762e-03

13 15 16 5 30

3.545568e-03 3.887529e-03 5.330984e-03 5.799134e-03 7.689797e-03

22 18 26 9 35

8.269420e-03 8.659824e-03 9.365238e-03 9.973274e-03 9.973274e-03

31 21 11 6 4

1.020873e-02 1.183492e-02 1.341648e-02 1.511234e-02 1.900315e-02

1 2 34 10 29

2.008583e-02 3.279385e-02 3.308947e-02 3.681500e-02 4.563892e-02

28 23 32 25 17

5.543165e-02 5.791693e-02 6.964886e-02 7.224690e-02 1.081814e-01

27 24 8 36 12

1.406244e-01 1.797481e-01 2.856880e-01 3.336714e-01 NaN

20

NaN

>

> leverages = hatvalues(model3)

> leverages[order(leverages)]

25 18 4 31 23 22 28

0.2584376 0.2712856 0.3237278 0.3252416 0.3306543 0.3419996 0.3614822

24 14 26 3 1 7 33

0.3881446 0.3882948 0.3890348 0.4019760 0.4041002 0.4041361 0.4083602

29 15 21 17 16 2 13

0.4348101 0.4394299 0.4415906 0.4593865 0.4638654 0.4867396 0.5069989

30 27 11 19 5 34 6

0.5075253 0.5098067 0.5243946 0.5454469 0.5687183 0.5855562 0.5895334

8 32 10 36 9 35 12

0.5963004 0.6206456 0.6211870 0.6370001 0.7320946 0.7320946 1.0000000

20

1.0000000

> cooks = cooks.distance(model3)

> cooks[order(cooks)]

33 3 14 26 30

0.0002385147 0.0014136989 0.0015305635 0.0022882504 0.0033768969

19 7 15 22 6

0.0035036214 0.0035148482 0.0050432569 0.0061578990 0.0071854743

31 16 13 5 18

0.0079492803 0.0088031569 0.0088748541 0.0091413987 0.0097283315

21 4 11 35 9

0.0119746503 0.0163808859 0.0175803567 0.0190984584 0.0190984584

1 34 2 29 10

0.0282927256 0.0354390411 0.0361133970 0.0413799644 0.0477436690

23 28 25 27 17

0.0508389620 0.0515554974 0.0560838676 0.0910330735 0.1168270835

32 24 8 36 12

0.1840739712 0.1855800945 0.3639933244 0.5249729552 NaN

20

NaN

> qf(.95, 17, 19)

[1] 2.197729

> #Thus no observation exceeds the cutoff of 2.198 here for Cook’s Distance

> ###thus,we only cut off 12,20 and 36 from the previous

> model3 = lm(lnTest3~lnTest1+lnTest2 +Gender+Year + GPA + CrHrs + Stick + ClassRow + CokePepsi + siblings + countries + lnJobs + DogCat)

> model4 = lm(lnTest3~lnTest1+lnTest2 +Gender+Year + GPA + CrHrs + Stick + ClassRow + CokePepsi + siblings + countries + lnJobs + DogCat,subset=-c(36,12,20))

> summary(model4)

Call:

lm(formula = lnTest3 ~ lnTest1 + lnTest2 + Gender + Year + GPA +

CrHrs + Stick + ClassRow + CokePepsi + siblings + countries +

lnJobs + DogCat, subset = -c(36, 12, 20))

Residuals:

Min 1Q Median 3Q Max

-0.12424 -0.01431 0.00113 0.03341 0.09903

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 1.044118 1.360537 0.767 0.45335

lnTest1 0.330535 0.110353 2.995 0.00814 \*\*

lnTest2 0.479070 0.256852 1.865 0.07952 .

GenderMale 0.059683 0.039954 1.494 0.15355

YearSenior -0.042537 0.063844 -0.666 0.51418

YearSophomore -0.017942 0.040629 -0.442 0.66433

GPA 0.001284 0.052693 0.024 0.98085

CrHrs -0.009964 0.011767 -0.847 0.40890

Stickyes 0.045373 0.030588 1.483 0.15628

ClassRow 0.003682 0.012511 0.294 0.77208

CokePepsiNeither -0.040541 0.042667 -0.950 0.35534

CokePepsiPepsi -0.006787 0.047860 -0.142 0.88890

siblings 0.008790 0.016603 0.529 0.60334

countries 0.001634 0.002982 0.548 0.59079

lnJobs -0.024178 0.041458 -0.583 0.56742

DogCatDog -0.041651 0.039940 -1.043 0.31162

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.07102 on 17 degrees of freedom

Multiple R-squared: 0.6444, Adjusted R-squared: 0.3307

F-statistic: 2.054 on 15 and 17 DF, p-value: 0.07773

> step(model4,direction='backward',criterion='AIC')

Start: AIC=-164.44

lnTest3 ~ lnTest1 + lnTest2 + Gender + Year + GPA + CrHrs + Stick +

ClassRow + CokePepsi + siblings + countries + lnJobs + DogCat

Df Sum of Sq RSS AIC

- Year 2 0.002579 0.088331 -167.47

- CokePepsi 2 0.004826 0.090577 -166.64

- GPA 1 0.000003 0.085755 -166.44

- ClassRow 1 0.000437 0.086189 -166.28

- siblings 1 0.001414 0.087166 -165.90

- countries 1 0.001515 0.087267 -165.87

- lnJobs 1 0.001716 0.087467 -165.79

- CrHrs 1 0.003617 0.089368 -165.08

<none> 0.085752 -164.44

- DogCat 1 0.005486 0.091238 -164.40

- Stick 1 0.011099 0.096851 -162.43

- Gender 1 0.011256 0.097008 -162.37

- lnTest2 1 0.017548 0.103300 -160.30

- lnTest1 1 0.045254 0.131006 -152.46

Step: AIC=-167.46

lnTest3 ~ lnTest1 + lnTest2 + Gender + GPA + CrHrs + Stick +

ClassRow + CokePepsi + siblings + countries + lnJobs + DogCat

Df Sum of Sq RSS AIC

- CokePepsi 2 0.003758 0.092088 -170.09

- GPA 1 0.000004 0.088334 -169.46

- siblings 1 0.001254 0.089585 -169.00

- lnJobs 1 0.001597 0.089927 -168.87

- ClassRow 1 0.002260 0.090590 -168.63

- countries 1 0.003628 0.091959 -168.14

- CrHrs 1 0.004197 0.092528 -167.93

- DogCat 1 0.004703 0.093033 -167.75

<none> 0.088331 -167.47

- Stick 1 0.009383 0.097714 -166.13

- Gender 1 0.009580 0.097910 -166.07

- lnTest2 1 0.016601 0.104931 -163.78

- lnTest1 1 0.043366 0.131696 -156.28

Step: AIC=-170.09

lnTest3 ~ lnTest1 + lnTest2 + Gender + GPA + CrHrs + Stick +

ClassRow + siblings + countries + lnJobs + DogCat

Df Sum of Sq RSS AIC

- GPA 1 0.000066 0.092154 -172.07

- siblings 1 0.000548 0.092637 -171.89

- lnJobs 1 0.001041 0.093129 -171.72

- ClassRow 1 0.002099 0.094188 -171.35

- countries 1 0.002902 0.094991 -171.07

- DogCat 1 0.005010 0.097099 -170.34

<none> 0.092088 -170.09

- Gender 1 0.006454 0.098542 -169.85

- Stick 1 0.006921 0.099009 -169.70

- CrHrs 1 0.009652 0.101740 -168.80

- lnTest2 1 0.014094 0.106183 -167.39

- lnTest1 1 0.042121 0.134210 -159.66

Step: AIC=-172.07

lnTest3 ~ lnTest1 + lnTest2 + Gender + CrHrs + Stick + ClassRow +

siblings + countries + lnJobs + DogCat

Df Sum of Sq RSS AIC

- siblings 1 0.000557 0.092712 -173.87

- lnJobs 1 0.001010 0.093165 -173.71

- ClassRow 1 0.002456 0.094611 -173.20

- countries 1 0.003118 0.095273 -172.97

- DogCat 1 0.005319 0.097473 -172.22

<none> 0.092154 -172.07

- Gender 1 0.006562 0.098717 -171.80

- Stick 1 0.007002 0.099157 -171.65

- CrHrs 1 0.009647 0.101802 -170.78

- lnTest2 1 0.016414 0.108568 -168.66

- lnTest1 1 0.062500 0.154655 -156.98

Step: AIC=-173.87

lnTest3 ~ lnTest1 + lnTest2 + Gender + CrHrs + Stick + ClassRow +

countries + lnJobs + DogCat

Df Sum of Sq RSS AIC

- lnJobs 1 0.000782 0.093494 -175.59

- ClassRow 1 0.002259 0.094971 -175.07

- countries 1 0.003669 0.096381 -174.59

- DogCat 1 0.004762 0.097474 -174.21

<none> 0.092712 -173.87

- Gender 1 0.006117 0.098829 -173.76

- Stick 1 0.006494 0.099205 -173.63

- CrHrs 1 0.011218 0.103930 -172.10

- lnTest2 1 0.015873 0.108584 -170.65

- lnTest1 1 0.062022 0.154734 -158.96

Step: AIC=-175.59

lnTest3 ~ lnTest1 + lnTest2 + Gender + CrHrs + Stick + ClassRow +

countries + DogCat

Df Sum of Sq RSS AIC

- ClassRow 1 0.001698 0.095191 -177.00

- countries 1 0.003968 0.097462 -176.22

- DogCat 1 0.005450 0.098943 -175.72

<none> 0.093494 -175.59

- Gender 1 0.006389 0.099882 -175.41

- Stick 1 0.006623 0.100117 -175.33

- CrHrs 1 0.013707 0.107201 -173.07

- lnTest2 1 0.015220 0.108713 -172.61

- lnTest1 1 0.063597 0.157090 -160.47

Step: AIC=-177

lnTest3 ~ lnTest1 + lnTest2 + Gender + CrHrs + Stick + countries +

DogCat

Df Sum of Sq RSS AIC

- DogCat 1 0.004489 0.099680 -177.48

- countries 1 0.004993 0.100185 -177.31

<none> 0.095191 -177.00

- Gender 1 0.007124 0.102316 -176.62

- Stick 1 0.007536 0.102727 -176.48

- CrHrs 1 0.013344 0.108535 -174.67

- lnTest2 1 0.013558 0.108750 -174.60

- lnTest1 1 0.062242 0.157433 -162.39

Step: AIC=-177.48

lnTest3 ~ lnTest1 + lnTest2 + Gender + CrHrs + Stick + countries

Df Sum of Sq RSS AIC

- Gender 1 0.004923 0.10460 -177.88

- countries 1 0.005152 0.10483 -177.81

<none> 0.09968 -177.48

- Stick 1 0.006280 0.10596 -177.46

- lnTest2 1 0.012627 0.11231 -175.54

- CrHrs 1 0.013858 0.11354 -175.18

- lnTest1 1 0.059104 0.15879 -164.11

Step: AIC=-177.88

lnTest3 ~ lnTest1 + lnTest2 + CrHrs + Stick + countries

Df Sum of Sq RSS AIC

- countries 1 0.004500 0.10910 -178.50

- Stick 1 0.005679 0.11028 -178.14

<none> 0.10460 -177.88

- lnTest2 1 0.010959 0.11556 -176.60

- CrHrs 1 0.020243 0.12485 -174.05

- lnTest1 1 0.054244 0.15885 -166.10

Step: AIC=-178.49

lnTest3 ~ lnTest1 + lnTest2 + CrHrs + Stick

Df Sum of Sq RSS AIC

- Stick 1 0.005918 0.11502 -178.75

<none> 0.10910 -178.50

- lnTest2 1 0.011819 0.12092 -177.10

- CrHrs 1 0.020135 0.12924 -174.91

- lnTest1 1 0.059251 0.16835 -166.18

Step: AIC=-178.75

lnTest3 ~ lnTest1 + lnTest2 + CrHrs

Df Sum of Sq RSS AIC

<none> 0.11502 -178.75

- lnTest2 1 0.010029 0.12505 -177.99

- CrHrs 1 0.018039 0.13306 -175.94

- lnTest1 1 0.054219 0.16924 -168.01

Call:

lm(formula = lnTest3 ~ lnTest1 + lnTest2 + CrHrs, subset = -c(36,

12, 20))

Coefficients:

(Intercept) lnTest1 lnTest2 CrHrs

2.5937 0.2305 0.2611 -0.0178

model5=lm( lnTest3 ~ lnTest1 + lnTest2 + CrHrs, subset = -c(36,12,20))

summary(model5)

Call:

lm(formula = lnTest3 ~ lnTest1 + lnTest2 + CrHrs, subset = -c(36,

12, 20))

Residuals:

Min 1Q Median 3Q Max

-0.216257 -0.020607 0.008654 0.037819 0.096747

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 2.593696 0.734184 3.533 0.001399 \*\*

lnTest1 0.230536 0.062352 3.697 0.000904 \*\*\*

lnTest2 0.261094 0.164193 1.590 0.122641

CrHrs -0.017805 0.008349 -2.133 0.041549 \*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.06298 on 29 degrees of freedom

Multiple R-squared: 0.523, Adjusted R-squared: 0.4737

F-statistic: 10.6 on 3 and 29 DF, p-value: 7.14e-05

**> cor(cbind(lnTest1, lnTest2 , CrHrs))**

**lnTest1 lnTest2 CrHrs**

**lnTest1 1.00000000 0.496222 0.03528039**

**lnTest2 0.49622199 1.000000 -0.24330598**

**CrHrs 0.03528039 -0.243306 1.00000000**

**> #There is no high correlation figures.so I expect no big collinearity exists**

plot(residuals(model5) ~ fitted.values(model5), main="Residuals vs.Fitted Value")



# This shows a good scatter plot since most of points reflects a good spread around (-.1 , .1)