test <- read.csv(file.choose(), header=TRUE)

attach(test)

model1= lm(Test3~Test1+Test2 +Gender+Year + GPA + CrHrs + Stick + ClassRow + CokePepsi + siblings + countries + jobs + DogCat)

summary(model1)

Call:

lm(formula = Test3 ~ Test1 + Test2 + Gender + Year + GPA + CrHrs +

Stick + ClassRow + CokePepsi + siblings + countries + jobs +

DogCat)

Residuals:

Min 1Q Median 3Q Max

-11.050 -2.495 0.000 2.845 11.142

Coefficients:

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

(Intercept) -1.80611 35.30568 -0.051 0.959764

Test1 0.49884 0.12460 4.004 0.000833 \*\*\*

Test2 0.60059 0.26400 2.275 0.035370 \*

GenderMale 5.06640 3.74706 1.352 0.193089

YearSenior -0.77521 5.85190 -0.132 0.896081

YearSophomore 0.40434 3.76516 0.107 0.915668

GPA -2.01104 4.91344 -0.409 0.687153

CrHrs -0.32922 1.10961 -0.297 0.770090

Stickno 0.01962 8.75519 0.002 0.998237

Stickyes 3.73659 2.91513 1.282 0.216180

Stickyes 6.77639 8.63500 0.785 0.442802

ClassRow 1.66398 1.06137 1.568 0.134348

CokePepsiNeither -3.82840 4.04018 -0.948 0.355891

CokePepsiPepsi 1.70388 4.34926 0.392 0.699833

siblings 1.32330 1.53466 0.862 0.399879

countries 0.21406 0.27399 0.781 0.444801

jobs -0.50202 0.96263 -0.522 0.608371

DogCatDog -3.26425 3.86530 -0.845 0.409473

---

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

Residual standard error: 6.677 on 18 degrees of freedom

Multiple R-squared: 0.7721, Adjusted R-squared: 0.5569

F-statistic: 3.587 on 17 and 18 DF, p-value: 0.005071

#Hypotheses:

#Test1 and Test2 have positive effects on predicting Test3

#while CrHrs has a negative effect

#Above showed Test1 and Test2 both have positive estimate coefficient and are significant

#though they could be further improved.

#Meanwhile, CrHrs have negative estimate coefficient and are insignificant

#since its p-value is high

At the first one I build,I change nothing but remove outliers and do backward regression

Model A:

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

plot.new()

hist(Test1)

hist(Test2)

hist(GPA)

hist(CrHrs)

hist(jobs)

hist(Test3)

plot(Test1,Test3)

plot(Test2,Test3)

plot(GPA,Test3)

plot(CrHrs,Test3)

plot(jobs,Test3)



> rstandard = rstandard(model1)

> leverages = hatvalues(model1)

> par(mfrow=c(1,2))

> hist(rstandard)

> hist(leverages)

> dim(test)

[1] 36 16

> rstandard[order(rstandard)]

36 25 23 27 28

-2.22195276 -1.92381738 -1.44108052 -1.38451614 -1.17300893

2 34 22 21 26

-0.78443066 -0.66124045 -0.53079605 -0.50458817 -0.49193298

11 16 15 7 9

-0.46958921 -0.33971258 -0.30034926 -0.26078807 -0.25160858

19 3 14 33 13

-0.15047122 -0.03212281 0.05119376 0.24481801 0.25081943

35 5 30 6 18

0.25160858 0.28627654 0.37140493 0.42013243 0.60382096

31 10 1 4 32

0.61238223 0.64524185 0.74148645 0.81653444 0.99220827

29 17 8 24 12

1.03760578 1.59025049 2.02598100 2.16811280 NaN

20

NaN

> #Thus we exclude 36,8,24,12,20

> #Since they are either above 2 or below -2

> 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

> #So there is no high leverage point

> 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

> #df1=16+1=17;df2=36-17=19;

> qf(.95, 17, 19)

[1] 2.197729

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

model2=lm(Test3~Test1+Test2 +Gender+Year + GPA + CrHrs + Stick + ClassRow + CokePepsi + siblings + countries + jobs + DogCat,subset=-c(36,8,24,12,20))

summary(model2)

Call:

lm(formula = Test3 ~ Test1 + Test2 + Gender + Year + GPA + CrHrs +

Stick + ClassRow + CokePepsi + siblings + countries + jobs +

DogCat, subset = -c(36, 8, 24, 12, 20))

Residuals:

Min 1Q Median 3Q Max

-10.5649 -1.0995 0.4255 2.2522 8.1130

Coefficients:

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

(Intercept) 3.9450 33.3432 0.118 0.90739

Test1 0.4154 0.1309 3.173 0.00631 \*\*

Test2 0.6519 0.2848 2.289 0.03700 \*

GenderMale 2.5634 3.1196 0.822 0.42411

YearSenior -1.1347 4.8044 -0.236 0.81649

YearSophomore -1.0089 3.4278 -0.294 0.77255

GPA -2.6897 4.1906 -0.642 0.53066

CrHrs -0.3098 0.9079 -0.341 0.73768

Stickyes 3.7663 2.4210 1.556 0.14064

ClassRow 1.2759 0.9990 1.277 0.22094

CokePepsiNeither -1.4738 3.3601 -0.439 0.66718

CokePepsiPepsi 2.3020 4.0734 0.565 0.58034

siblings 1.7711 1.3166 1.345 0.19854

countries 0.2274 0.2286 0.995 0.33564

jobs -0.5953 0.7752 -0.768 0.45449

DogCatDog -3.2156 3.2258 -0.997 0.33464

---

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

Residual standard error: 5.315 on 15 degrees of freedom

Multiple R-squared: 0.7426, Adjusted R-squared: 0.4851

F-statistic: 2.885 on 15 and 15 DF, p-value: 0.02419

> step(model2,direction='backward',criterion='AIC')

Start: AIC=113.07

Test3 ~ Test1 + Test2 + Gender + Year + GPA + CrHrs + Stick +

ClassRow + CokePepsi + siblings + countries + jobs + DogCat

Df Sum of Sq RSS AIC

- Year 2 3.011 426.70 109.28

- CokePepsi 2 25.935 449.63 110.91

- CrHrs 1 3.288 426.98 111.31

- GPA 1 11.637 435.33 111.91

- jobs 1 16.654 440.35 112.26

- Gender 1 19.072 442.76 112.43

- countries 1 27.950 451.64 113.05

- DogCat 1 28.069 451.76 113.05

<none> 423.69 113.07

- ClassRow 1 46.079 469.77 114.27

- siblings 1 51.115 474.81 114.60

- Stick 1 68.357 492.05 115.70

- Test2 1 148.007 571.70 120.35

- Test1 1 284.315 708.01 126.98

Step: AIC=109.29

Test3 ~ Test1 + Test2 + Gender + GPA + CrHrs + Stick + ClassRow +

CokePepsi + siblings + countries + jobs + DogCat

Df Sum of Sq RSS AIC

- CrHrs 1 4.13 430.83 107.58

- CokePepsi 2 33.21 459.92 107.61

- jobs 1 15.08 441.78 108.36

- Gender 1 17.00 443.70 108.50

- GPA 1 19.18 445.88 108.65

- DogCat 1 26.31 453.02 109.14

<none> 426.70 109.28

- countries 1 51.02 477.72 110.79

- siblings 1 51.42 478.13 110.81

- ClassRow 1 65.74 492.45 111.73

- Stick 1 66.11 492.81 111.75

- Test2 1 173.88 600.59 117.88

- Test1 1 324.47 751.18 124.82

Step: AIC=107.58

Test3 ~ Test1 + Test2 + Gender + GPA + Stick + ClassRow + CokePepsi +

siblings + countries + jobs + DogCat

Df Sum of Sq RSS AIC

- CokePepsi 2 42.61 473.44 106.51

- GPA 1 21.93 452.76 107.12

- jobs 1 24.69 455.52 107.31

- DogCat 1 27.98 458.81 107.53

<none> 430.83 107.58

- Gender 1 29.44 460.28 107.63

- countries 1 52.28 483.11 109.13

- Stick 1 67.08 497.91 110.07

- siblings 1 67.40 498.23 110.09

- ClassRow 1 69.35 500.18 110.21

- Test2 1 215.50 646.33 118.16

- Test1 1 368.74 799.58 124.75

Step: AIC=106.51

Test3 ~ Test1 + Test2 + Gender + GPA + Stick + ClassRow + siblings +

countries + jobs + DogCat

Df Sum of Sq RSS AIC

- GPA 1 14.53 487.96 105.44

- Gender 1 30.16 503.60 106.42

<none> 473.44 106.51

- countries 1 32.28 505.72 106.55

- Stick 1 43.00 516.44 107.20

- jobs 1 49.68 523.12 107.60

- DogCat 1 52.11 525.55 107.74

- siblings 1 55.72 529.16 107.96

- ClassRow 1 75.67 549.11 109.10

- Test2 1 225.93 699.37 116.60

- Test1 1 338.34 811.78 121.22

Step: AIC=105.44

Test3 ~ Test1 + Test2 + Gender + Stick + ClassRow + siblings +

countries + jobs + DogCat

Df Sum of Sq RSS AIC

- countries 1 25.59 513.55 105.03

- Gender 1 31.17 519.13 105.36

<none> 487.96 105.44

- Stick 1 37.82 525.79 105.76

- siblings 1 50.52 538.49 106.50

- jobs 1 54.61 542.58 106.73

- ClassRow 1 61.54 549.50 107.13

- DogCat 1 81.42 569.38 108.23

- Test2 1 218.32 706.28 114.91

- Test1 1 368.29 856.26 120.88

Step: AIC=105.03

Test3 ~ Test1 + Test2 + Gender + Stick + ClassRow + siblings +

jobs + DogCat

Df Sum of Sq RSS AIC

- Gender 1 29.05 542.60 104.73

<none> 513.55 105.03

- Stick 1 38.32 551.87 105.26

- jobs 1 59.28 572.83 106.42

- siblings 1 61.23 574.78 106.52

- ClassRow 1 74.08 587.63 107.20

- DogCat 1 87.86 601.41 107.92

- Test2 1 223.67 737.22 114.24

- Test1 1 407.60 921.15 121.14

Step: AIC=104.73

Test3 ~ Test1 + Test2 + Stick + ClassRow + siblings + jobs +

DogCat

Df Sum of Sq RSS AIC

- Stick 1 32.13 574.74 104.52

<none> 542.60 104.73

- siblings 1 56.93 599.53 105.83

- jobs 1 69.74 612.34 106.48

- DogCat 1 73.23 615.84 106.66

- ClassRow 1 92.47 635.07 107.61

- Test2 1 240.10 782.71 114.09

- Test1 1 379.94 922.54 119.19

Step: AIC=104.52

Test3 ~ Test1 + Test2 + ClassRow + siblings + jobs + DogCat

Df Sum of Sq RSS AIC

<none> 574.74 104.52

- siblings 1 42.15 616.88 104.71

- DogCat 1 58.96 633.70 105.55

- jobs 1 81.21 655.95 106.61

- ClassRow 1 91.76 666.49 107.11

- Test2 1 209.39 784.12 112.15

- Test1 1 379.55 954.29 118.24

Call:

lm(formula = Test3 ~ Test1 + Test2 + ClassRow + siblings + jobs +

DogCat, subset = -c(36, 8, 24, 12, 20))

Coefficients:

(Intercept) Test1 Test2 ClassRow siblings

14.7092 0.3135 0.5310 1.4891 1.3892

jobs DogCatDog

-0.9798 -3.6334

> model6 =lm(Test3 ~ Test1 + Test2 + ClassRow + siblings + jobs + DogCat, subset = -c(36, 8, 24, 12, 20))

> summary(model6)

Call:

lm(formula = Test3 ~ Test1 + Test2 + ClassRow + siblings + jobs +

DogCat, subset = -c(36, 8, 24, 12, 20))

Residuals:

Min 1Q Median 3Q Max

-11.9173 -1.9323 0.6734 2.2759 10.7547

Coefficients:

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

(Intercept) 14.70919 15.99995 0.919 0.367073

Test1 0.31347 0.07874 3.981 0.000553 \*\*\*

Test2 0.53104 0.17959 2.957 0.006872 \*\*

ClassRow 1.48911 0.76073 1.957 0.062019 .

siblings 1.38918 1.04712 1.327 0.197106

jobs -0.97975 0.53203 -1.842 0.077938 .

DogCatDog -3.63337 2.31548 -1.569 0.129702

---

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

Residual standard error: 4.894 on 24 degrees of freedom

Multiple R-squared: 0.6508, Adjusted R-squared: 0.5635

F-statistic: 7.455 on 6 and 24 DF, p-value: 0.0001376

> cor(cbind(Test1 , Test2 , ClassRow , siblings , jobs , DogCat))

Test1 Test2 ClassRow siblings jobs DogCat

Test1 1.00000000 0.46852900 -0.067989188 0.075140599 0.23846410 -0.03819603

Test2 0.46852900 1.00000000 -0.251076503 0.046020498 0.08577836 0.00000000

ClassRow -0.06798919 -0.25107650 1.000000000 -0.005066166 0.16000485 0.18806932

siblings 0.07514060 0.04602050 -0.005066166 1.000000000 0.16901905 0.23461061

jobs 0.23846410 0.08577836 0.160004851 0.169019051 1.00000000 0.06423932

DogCat -0.03819603 0.00000000 0.188069323 0.234610606 0.06423932 1.00000000

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

plot(residuals(model6) ~ fitted.values(model6), main="Residuals vs.Fitted Value")



Here in B,I log up some variables shown a below

Model B:

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)

Model C:

#realizing GPA may do little and we need to improve CrHrs significance and also notice job’s plot with Test3 has a quadratic feature.

SqrJobs=jobs\*jobs

lnGPA=log(GPA)

> model15= lm(Test3~Test1+Test2 +Gender+Year + lnGPA + CrHrs +sqrCrHrs + Stick + ClassRow + CokePepsi + siblings + countries + sqrJobs +jobs+ DogCat)

> rstandard = rstandard(model15)

> leverages = hatvalues(model15)

> cooks = cooks.distance(model15)

> rstandard[order(rstandard)]

36 23 25 15 27 22 7

-2.50702842 -1.96009251 -1.71241591 -0.93636357 -0.93450031 -0.81444747 -0.55423126

2 28 34 11 19 21 26

-0.55398582 -0.41698215 -0.40312759 -0.39873843 -0.32270350 -0.31629533 -0.23843069

9 6 35 5 33 31 3

-0.10141604 -0.01168035 0.10141604 0.10528500 0.16929613 0.19113415 0.23963294

13 30 16 14 10 4 1

0.24599635 0.25182738 0.27880182 0.41166414 0.46261636 0.56314691 0.86556144

17 18 29 32 8 24 12

1.09966503 1.25909337 1.31715707 1.46785327 1.72729231 1.88442876 NaN

20

NaN

> leverages[order(leverages)]

25 4 23 22 3 18 7 26

0.3065547 0.3523467 0.3835984 0.3840451 0.3918877 0.3975792 0.4249472 0.4296137

31 29 24 14 33 1 13 21

0.4311925 0.4487724 0.4504502 0.4643788 0.4787908 0.4867327 0.4872349 0.4967646

17 15 2 11 30 16 36 5

0.5147099 0.5213397 0.5255145 0.5332198 0.5418834 0.5502876 0.5711024 0.5740181

34 19 27 8 32 10 28 6

0.5880871 0.5942190 0.6116289 0.6595708 0.6712533 0.6930204 0.6991648 0.7227345

35 9 12 20

0.8066780 0.8066780 1.0000000 1.0000000

> cooks[order(cooks)]

6 5 33 31 3 26

1.778134e-05 7.468571e-04 1.316430e-03 1.384690e-03 1.850294e-03 2.140935e-03

35 9 13 30 16 21

2.145864e-03 2.145864e-03 2.875063e-03 3.750643e-03 4.755718e-03 4.937818e-03

14 19 4 11 7 34

7.346333e-03 7.624867e-03 8.626627e-03 9.081134e-03 1.134957e-02 1.160089e-02

2 28 22 10 1 15

1.699531e-02 2.020486e-02 2.067900e-02 2.415730e-02 3.552326e-02 4.774770e-02

18 17 25 27 29 23

5.231298e-02 6.412863e-02 6.481632e-02 6.876540e-02 7.062206e-02 1.195463e-01

24 32 8 36 12 20

1.455356e-01 2.199684e-01 2.890256e-01 4.184551e-01 NaN NaN

#Like the other 3 models I built I found there is no cutoff points other than #36,12,20

#as it has 1.417 as the high leverage cutoff and 2.198 as the Cooks cutoff.

> step(model15,direction='backward',criterion='AIC')

Start: AIC=146.52

Test3 ~ Test1 + Test2 + Gender + Year + lnGPA + CrHrs + sqrCrHrs +

Stick + ClassRow + CokePepsi + siblings + countries + sqrJobs +

jobs + DogCat

Df Sum of Sq RSS AIC

- Stick 3 34.59 728.57 142.27

- Year 2 35.54 729.53 144.32

- siblings 1 0.60 694.59 144.55

- DogCat 1 12.84 706.83 145.18

- lnGPA 1 37.43 731.42 146.41

<none> 693.99 146.52

- ClassRow 1 47.77 741.76 146.92

- jobs 1 58.78 752.77 147.45

- countries 1 67.11 761.10 147.84

- sqrCrHrs 1 69.90 763.88 147.98

- CrHrs 1 70.28 764.27 147.99

- sqrJobs 1 80.25 774.23 148.46

- CokePepsi 2 133.72 827.71 148.87

- Gender 1 151.75 845.74 151.64

- Test2 1 218.87 912.86 154.39

- Test1 1 733.32 1427.30 170.48

Step: AIC=142.27

Test3 ~ Test1 + Test2 + Gender + Year + lnGPA + CrHrs + sqrCrHrs +

ClassRow + CokePepsi + siblings + countries + sqrJobs + jobs +

DogCat

Df Sum of Sq RSS AIC

- Year 2 42.04 770.61 140.29

- siblings 1 1.06 729.64 140.32

- DogCat 1 11.30 739.87 140.83

- lnGPA 1 34.89 763.46 141.96

<none> 728.57 142.27

- ClassRow 1 54.54 783.12 142.87

- countries 1 73.37 801.94 143.73

- CokePepsi 2 121.45 850.02 143.82

- jobs 1 81.79 810.36 144.10

- sqrJobs 1 113.70 842.27 145.49

- sqrCrHrs 1 114.59 843.17 145.53

- CrHrs 1 115.93 844.51 145.59

- Gender 1 152.96 881.53 147.13

- Test2 1 219.00 947.57 149.73

- Test1 1 724.42 1452.99 165.12

Step: AIC=140.29

Test3 ~ Test1 + Test2 + Gender + lnGPA + CrHrs + sqrCrHrs + ClassRow +

CokePepsi + siblings + countries + sqrJobs + jobs + DogCat

Df Sum of Sq RSS AIC

- siblings 1 1.50 772.11 138.36

- lnGPA 1 9.55 780.16 138.74

- DogCat 1 20.27 790.89 139.23

- CokePepsi 2 87.13 857.74 140.15

- countries 1 42.01 812.62 140.20

<none> 770.61 140.29

- jobs 1 44.21 814.83 140.30

- ClassRow 1 75.37 845.99 141.65

- sqrJobs 1 77.05 847.67 141.72

- sqrCrHrs 1 89.67 860.28 142.25

- CrHrs 1 90.83 861.45 142.30

- Gender 1 124.09 894.70 143.67

- Test2 1 260.46 1031.07 148.77

- Test1 1 690.38 1461.00 161.32

Step: AIC=138.36

Test3 ~ Test1 + Test2 + Gender + lnGPA + CrHrs + sqrCrHrs + ClassRow +

CokePepsi + countries + sqrJobs + jobs + DogCat

Df Sum of Sq RSS AIC

- lnGPA 1 10.52 782.63 136.85

- DogCat 1 18.78 790.89 137.23

- CokePepsi 2 86.26 858.38 138.18

<none> 772.11 138.36

- countries 1 45.43 817.54 138.42

- jobs 1 47.30 819.41 138.50

- ClassRow 1 73.98 846.10 139.66

- sqrJobs 1 80.75 852.86 139.94

- sqrCrHrs 1 102.45 874.56 140.85

- CrHrs 1 104.18 876.30 140.92

- Gender 1 124.42 896.53 141.74

- Test2 1 259.16 1031.27 146.78

- Test1 1 696.67 1468.79 159.51

Step: AIC=136.85

Test3 ~ Test1 + Test2 + Gender + CrHrs + sqrCrHrs + ClassRow +

CokePepsi + countries + sqrJobs + jobs + DogCat

Df Sum of Sq RSS AIC

- CokePepsi 2 79.23 861.87 136.32

- countries 1 39.48 822.12 136.62

- DogCat 1 40.24 822.87 136.65

- jobs 1 41.27 823.90 136.70

<none> 782.63 136.85

- ClassRow 1 64.96 847.59 137.72

- sqrJobs 1 73.25 855.89 138.07

- sqrCrHrs 1 93.96 876.60 138.93

- CrHrs 1 95.73 878.36 139.00

- Gender 1 115.41 898.05 139.80

- Test2 1 258.46 1041.10 145.12

- Test1 1 938.51 1721.15 163.22

Step: AIC=136.32

Test3 ~ Test1 + Test2 + Gender + CrHrs + sqrCrHrs + ClassRow +

countries + sqrJobs + jobs + DogCat

Df Sum of Sq RSS AIC

- jobs 1 15.91 877.77 134.98

- countries 1 17.52 879.38 135.04

- sqrJobs 1 37.86 899.73 135.87

<none> 861.87 136.32

- DogCat 1 58.89 920.75 136.70

- sqrCrHrs 1 62.55 924.42 136.84

- CrHrs 1 65.06 926.92 136.94

- Gender 1 70.77 932.63 137.16

- ClassRow 1 134.19 996.06 139.53

- Test2 1 285.69 1147.55 144.63

- Test1 1 860.08 1721.94 159.24

Step: AIC=134.98

Test3 ~ Test1 + Test2 + Gender + CrHrs + sqrCrHrs + ClassRow +

countries + sqrJobs + DogCat

Df Sum of Sq RSS AIC

- countries 1 14.04 891.81 133.55

- DogCat 1 45.51 923.29 134.80

<none> 877.77 134.98

- sqrCrHrs 1 58.14 935.91 135.29

- Gender 1 60.41 938.18 135.38

- CrHrs 1 60.62 938.40 135.38

- sqrJobs 1 91.72 969.50 136.56

- ClassRow 1 178.76 1056.54 139.65

- Test2 1 278.70 1156.48 142.91

- Test1 1 945.28 1823.06 159.29

Step: AIC=133.55

Test3 ~ Test1 + Test2 + Gender + CrHrs + sqrCrHrs + ClassRow +

sqrJobs + DogCat

Df Sum of Sq RSS AIC

- DogCat 1 47.53 939.34 133.42

<none> 891.81 133.55

- Gender 1 56.98 948.80 133.78

- sqrCrHrs 1 59.21 951.02 133.87

- CrHrs 1 61.68 953.49 133.96

- sqrJobs 1 92.12 983.93 135.09

- ClassRow 1 204.15 1095.97 138.97

- Test2 1 287.93 1179.74 141.62

- Test1 1 1002.96 1894.77 158.68

Step: AIC=133.42

Test3 ~ Test1 + Test2 + Gender + CrHrs + sqrCrHrs + ClassRow +

sqrJobs

Df Sum of Sq RSS AIC

- Gender 1 46.19 985.54 133.15

<none> 939.34 133.42

- sqrCrHrs 1 75.77 1015.12 134.21

- CrHrs 1 78.60 1017.95 134.31

- sqrJobs 1 98.90 1038.25 135.02

- ClassRow 1 172.66 1112.01 137.50

- Test2 1 270.69 1210.04 140.54

- Test1 1 1003.58 1942.92 157.58

Step: AIC=133.15

Test3 ~ Test1 + Test2 + CrHrs + sqrCrHrs + ClassRow + sqrJobs

Df Sum of Sq RSS AIC

- sqrCrHrs 1 45.69 1031.23 132.78

- CrHrs 1 48.47 1034.01 132.88

<none> 985.54 133.15

- sqrJobs 1 76.67 1062.20 133.84

- ClassRow 1 194.93 1180.47 137.65

- Test2 1 258.56 1244.10 139.53

- Test1 1 970.68 1956.22 155.83

Step: AIC=132.78

Test3 ~ Test1 + Test2 + CrHrs + ClassRow + sqrJobs

Df Sum of Sq RSS AIC

- sqrJobs 1 41.09 1072.3 132.19

- CrHrs 1 51.89 1083.1 132.55

<none> 1031.2 132.78

- ClassRow 1 216.98 1248.2 137.65

- Test2 1 285.32 1316.5 139.57

- Test1 1 985.09 2016.3 154.92

Step: AIC=132.19

Test3 ~ Test1 + Test2 + CrHrs + ClassRow

Df Sum of Sq RSS AIC

<none> 1072.3 132.19

- CrHrs 1 92.93 1165.2 133.18

- ClassRow 1 200.19 1272.5 136.35

- Test2 1 265.18 1337.5 138.14

- Test1 1 948.68 2021.0 153.00

Call:

lm(formula = Test3 ~ Test1 + Test2 + CrHrs + ClassRow)

Coefficients:

(Intercept) Test1 Test2 CrHrs ClassRow

30.5305 0.3706 0.4716 -1.2490 1.7333

> model19=lm(formula = Test3 ~ Test1 + Test2 + CrHrs + ClassRow)

> summary(model19)

Call:

lm(formula = Test3 ~ Test1 + Test2 + CrHrs + ClassRow)

Residuals:

Min 1Q Median 3Q Max

-13.367 -1.792 1.245 3.138 9.661

Coefficients:

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

(Intercept) 30.53054 21.08274 1.448 0.15762

Test1 0.37058 0.07076 5.237 1.09e-05 \*\*\*

Test2 0.47159 0.17032 2.769 0.00941 \*\*

CrHrs -1.24902 0.76201 -1.639 0.11130

ClassRow 1.73329 0.72049 2.406 0.02230 \*

---

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

Residual standard error: 5.881 on 31 degrees of freedom

Multiple R-squared: 0.6955, Adjusted R-squared: 0.6562

F-statistic: 17.7 on 4 and 31 DF, p-value: 1.166e-07

plot(residuals(model19) ~ fitted.values(model19), main="Residuals vs.Fitted Value")



> cor(cbind(Test1,Test2,CrHrs,ClassRow))

Test1 Test2 CrHrs ClassRow

Test1 1.000000000 0.4685290 0.008831843 -0.06798919

Test2 0.468528996 1.0000000 -0.251231740 -0.25107650

CrHrs 0.008831843 -0.2512317 1.000000000 0.11957516

ClassRow -0.067989188 -0.2510765 0.119575165 1.00000000

> #There is no high Collinearity

#I pick model19 as the best because it has the best F p-value almost equal to 0

#It also have a comparatively larger Rsqr and Rsqr adj, and they has narrower difference

#Also Test1 and Test2’s p-values are very small, though CrHrs was insignificant and still negative.

#And the residuals vs fitness plot does not look bad plus,no collinearity exists.