> 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 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")

