2

modelA=lm(Test3 ~ Gender+GPA+GradSchl+ClassRow+siblings+countries+jobs+tattoos + pets +HW3)

Call:

lm(formula = Test3 ~ Gender + GPA + GradSchl + ClassRow + siblings +

countries + jobs + tattoos + pets + HW3)

Residuals:

Min 1Q Median 3Q Max

-52.378 -3.718 -0.539 5.518 20.361

Coefficients:

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

(Intercept) 27.56042 12.27242 2.246 0.0283 \*

GenderMale -3.47887 2.79421 -1.245 0.2178

GPA 8.85384 3.43812 2.575 0.0124 \*

GradSchlyes 2.13222 2.75278 0.775 0.4415

ClassRow 0.54910 1.03831 0.529 0.5988

siblings -0.12145 1.28396 -0.095 0.9249

countries 0.08731 0.25797 0.338 0.7362

jobs 2.77797 1.65169 1.682 0.0976 .

tattoos -2.34708 2.02892 -1.157 0.2518

pets 0.09825 1.04416 0.094 0.9253

HW3 0.64052 0.11199 5.719 3.3e-07 \*\*\*

---

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

Residual standard error: 10.68 on 62 degrees of freedom

Multiple R-squared: 0.5373, Adjusted R-squared: 0.4627

F-statistic: 7.2 on 10 and 62 DF, p-value: 2.05e-07

> step(modelA)

Start: AIC=355.84

Test3 ~ Gender + GPA + GradSchl + ClassRow + siblings + countries +

jobs + tattoos + pets + HW3

Df Sum of Sq RSS AIC

- pets 1 1.0 7071.1 353.85

- siblings 1 1.0 7071.1 353.85

- countries 1 13.1 7083.2 353.98

- ClassRow 1 31.9 7102.0 354.17

- GradSchl 1 68.4 7138.5 354.54

- tattoos 1 152.6 7222.7 355.40

- Gender 1 176.8 7246.9 355.64

<none> 7070.1 355.84

- jobs 1 322.6 7392.7 357.10

- GPA 1 756.2 7826.3 361.26

- HW3 1 3730.1 10800.2 384.77

Step: AIC=353.85

Test3 ~ Gender + GPA + GradSchl + ClassRow + siblings + countries +

jobs + tattoos + HW3

Df Sum of Sq RSS AIC

- siblings 1 0.7 7071.8 351.86

- countries 1 12.7 7083.8 351.98

- ClassRow 1 31.1 7102.2 352.17

- GradSchl 1 67.9 7139.0 352.55

- tattoos 1 169.6 7240.7 353.58

- Gender 1 175.8 7246.9 353.64

<none> 7071.1 353.85

- jobs 1 321.9 7393.0 355.10

- GPA 1 757.5 7828.6 359.28

- HW3 1 3730.0 10801.1 382.78

Step: AIC=351.86

Test3 ~ Gender + GPA + GradSchl + ClassRow + countries + jobs +

tattoos + HW3

Df Sum of Sq RSS AIC

- countries 1 12.2 7084.0 349.98

- ClassRow 1 32.1 7103.9 350.19

- GradSchl 1 67.9 7139.7 350.56

- tattoos 1 168.9 7240.7 351.58

- Gender 1 176.6 7248.4 351.66

<none> 7071.8 351.86

- jobs 1 346.0 7417.8 353.35

- GPA 1 758.6 7830.4 357.30

- HW3 1 3733.6 10805.4 380.81

Step: AIC=349.98

Test3 ~ Gender + GPA + GradSchl + ClassRow + jobs + tattoos +

HW3

Df Sum of Sq RSS AIC

- ClassRow 1 27.3 7111.3 348.27

- GradSchl 1 60.2 7144.2 348.60

- tattoos 1 159.8 7243.9 349.61

- Gender 1 164.6 7248.6 349.66

<none> 7084.0 349.98

- jobs 1 334.3 7418.3 351.35

- GPA 1 766.4 7850.4 355.48

- HW3 1 3778.3 10862.3 379.19

Step: AIC=348.27

Test3 ~ Gender + GPA + GradSchl + jobs + tattoos + HW3

Df Sum of Sq RSS AIC

- GradSchl 1 71.0 7182.3 346.99

- Gender 1 147.5 7258.8 347.76

- tattoos 1 170.0 7281.3 347.99

<none> 7111.3 348.27

- jobs 1 318.6 7429.9 349.47

- GPA 1 800.3 7911.6 354.05

- HW3 1 3765.0 10876.3 377.28

Step: AIC=346.99

Test3 ~ Gender + GPA + jobs + tattoos + HW3

Df Sum of Sq RSS AIC

- Gender 1 147.0 7329.4 346.47

- tattoos 1 191.5 7373.9 346.91

<none> 7182.3 346.99

- jobs 1 363.6 7546.0 348.60

- GPA 1 894.8 8077.1 353.56

- HW3 1 3965.7 11148.0 377.08

Step: AIC=346.47

Test3 ~ GPA + jobs + tattoos + HW3

Df Sum of Sq RSS AIC

- tattoos 1 165.0 7494.4 346.10

<none> 7329.4 346.47

- jobs 1 364.4 7693.8 348.01

- GPA 1 1068.3 8397.6 354.40

- HW3 1 4253.8 11583.2 377.88

Step: AIC=346.1

Test3 ~ GPA + jobs + HW3

Df Sum of Sq RSS AIC

<none> 7494.4 346.10

- jobs 1 384.6 7879.0 347.75

- GPA 1 1054.1 8548.5 353.70

- HW3 1 4400.5 11894.9 377.82

Call:

lm(formula = Test3 ~ GPA + jobs + HW3)

Coefficients:

(Intercept) GPA jobs HW3

22.4196 10.0621 2.7695 0.6722

> modelB=lm(Test3 ~ GPA+jobs+HW3)

> summary(modelB)

> modelB=lm(Test3 ~ GPA+jobs+HW3)

> summary(modelB)

#modelA:we have Residual standard error: 10.68 on 62 degrees of freedom

Multiple R-squared: 0.5373, Adjusted R-squared: 0.4627

And F-pvalue =2.05e-07

#modelB:

Residual standard error: 10.42 on 69 degrees of freedom

Multiple R-squared: 0.5095, Adjusted R-squared: 0.4882

F-statistic: 23.89 on 3 and 69 DF, p-value: 1.026e-10

We could see Rsqr and R sqr adj has improved a lot.and F-pvalue become more significant

GPA 10.0621 3.2299 3.115 0.00268 \*\*

jobs 2.7695 1.4718 1.882 0.06409 .

HW3 0.6722 0.1056 6.365 1.84e-08 \*\*\*

All the above shows great improvement

> G2=GPA\*GPA

> J2=jobs\*jobs

> H2=HW3\*HW3

> modelB=lm(Test3 ~ GPA+jobs+HW3+G2+J2+H2)

> modelB=lm(Test3 ~ GPA+jobs+HW3)

> modelC=lm(Test3 ~ GPA+jobs+HW3+G2+J2+H2)

> summary(modelC)

Call:

lm(formula = Test3 ~ GPA + jobs + HW3 + G2 + J2 + H2)

Residuals:

Min 1Q Median 3Q Max

-33.263 -6.295 -0.503 6.247 25.428

Coefficients:

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

(Intercept) -66.41896 54.92054 -1.209 0.230837

GPA 56.19789 34.36654 1.635 0.106757

jobs -3.40416 3.94668 -0.863 0.391515

HW3 2.16773 0.37726 5.746 2.53e-07 \*\*\*

G2 -6.96113 5.28367 -1.317 0.192233

J2 2.09880 1.14720 1.829 0.071842 .

H2 -0.02542 0.00635 -4.003 0.000161 \*\*\*

---

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

Residual standard error: 9.325 on 66 degrees of freedom

Multiple R-squared: 0.6244, Adjusted R-squared: 0.5903

F-statistic: 18.29 on 6 and 66 DF, p-value: 2.218e-12

b)

Analysis of Variance Table

Model 1: Test3 ~ GPA + jobs + HW3

Model 2: Test3 ~ GPA + jobs + HW3 + G2 + J2 + H2

Res.Df RSS Df Sum of Sq F Pr(>F)

1 69 7494.4

2 66 5738.9 3 1755.5 6.7296 0.0004967 \*\*\*

---

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

Ho: Beta2=Beta4=Beta6=0;

H1:at least one Beta!=0;

#model2 is significant, because F-pvalue= 2.218e-12 <0.05 on 66 degrees of freedom,

c)

newdata=data.frame(GPA=3.23,jobs=2,HW3=85)

predict(modelB, newdata, interval="prediction", level=.95)

fit lwr upr

1 123.5728 100.5393 146.6062

If an individual has Xs values as described, we have 95% confidence that the predicted Y for that individual would have a value at (100.5393 146.6062) .

d)

rstandard = rstandard(modelB)

leverages = hatvalues(modelB)

par(mfrow=c(1,2))

hist(rstandard)

hist(leverages)



> dim(cl)

[1] 73 14

> k=3

> 3\*(3+1)/73

[1] 0.1643836

The 1st one skews left and there are outliers above 2 and below -2

when the 2nd one skews right,and there appears outliers above the cutoff points

e)

> rstandard[order(rstandard)]

20 26 7 32 69 45

-5.459639961 -2.076408742 -1.365640029 -1.217361067 -1.162624663 -1.121496035

72 3 2 67 61 46

-1.014844446 -1.002258318 -0.900221853 -0.843304694 -0.772876310 -0.770637288

54 48 70 58 6 47

-0.704858976 -0.654408516 -0.646141775 -0.516590375 -0.467150276 -0.424633487

49 31 9 8 30 21

-0.401505440 -0.391780824 -0.366628186 -0.335829299 -0.332696114 -0.328723956

63 36 71 59 38 18

-0.301073435 -0.293028287 -0.259220749 -0.219668107 -0.171999760 -0.120325169

57 55 33 27 65 34

-0.117143601 -0.111019777 -0.096242457 -0.094228807 -0.086392106 -0.082518179

53 56 43 42 13 12

-0.078857149 -0.036326242 -0.007304653 0.029386472 0.051673775 0.064428262

37 1 64 62 16 19

0.163454992 0.184445415 0.186856589 0.189693416 0.197963082 0.215672485

15 35 10 17 24 22

0.263032931 0.377602458 0.391470323 0.418408648 0.459984760 0.470884110

50 23 60 68 29 52

0.485086599 0.622655825 0.628409275 0.636736438 0.648960245 0.727227410

51 39 41 11 73 4

0.802425628 0.810977323 0.959603144 1.079248228 1.259554168 1.273804198

14 28 44 66 5 25

1.351193014 1.379230688 1.401690655 1.441777449 1.639357945 1.942063621

40

2.214607367

Thus #20,26,40 is outliers since it is larger than 2

> leverages[order(leverages)]

39 10 65 71 53 58 33 44

0.01664025 0.01693314 0.01725154 0.01894487 0.01942656 0.02093215 0.02138901 0.02240438

45 37 21 12 15 47 55 50

0.02240438 0.02268594 0.02365432 0.02407692 0.02412596 0.02649136 0.02681245 0.02773101

43 61 48 72 11 19 64 5

0.02863423 0.02922293 0.02967551 0.02987764 0.03022274 0.03125417 0.03132546 0.03302745

29 27 54 31 66 1 49 14

0.03326937 0.03365521 0.03393889 0.03506044 0.03547798 0.03560401 0.03719424 0.03725421

23 42 70 16 13 38 30 36

0.03730173 0.03872138 0.03874676 0.03996510 0.04079783 0.04099498 0.04164270 0.04539765

24 56 2 57 22 18 73 63

0.04559938 0.04574337 0.04694945 0.04747990 0.04904300 0.04931441 0.05468004 0.05622692

9 41 35 52 4 68 17 7

0.05636996 0.05794213 0.05939035 0.06051273 0.06174554 0.06329673 0.06423654 0.06523023

32 62 51 34 60 6 69 8

0.06812659 0.07199147 0.07274532 0.07427886 0.07558515 0.07794129 0.08507101 0.09428187

40 59 46 67 25 26 20 28

0.10507456 0.10636189 0.11723627 0.11858821 0.12506752 0.15206487 0.15330251 0.18368021

3

0.23467485

> dim(cl)

[1] 73 14

> k=3

> 3\*(3+1)/73

[1] 0.1643836

#as .1643836 is the cutoff

f)#we have #28 and #3 as high leverage point

modelB=lm(Test3 ~ GPA+jobs+HW3,subset=-c(20,26,40))

summary(modelB)

> modelB=lm(Test3 ~ GPA+jobs+HW3,subset=-c(20,26,40))

> summary(modelB)

Call:

lm(formula = Test3 ~ GPA + jobs + HW3, subset = -c(20, 26, 40))

Residuals:

Min 1Q Median 3Q Max

-21.0116 -4.2779 0.2742 4.2326 12.5264

Coefficients:

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

(Intercept) 38.84004 7.21026 5.387 1.03e-06 \*\*\*

GPA 8.42561 2.21737 3.800 0.000318 \*\*\*

jobs 2.36145 1.01193 2.334 0.022671 \*

HW3 0.41870 0.08345 5.017 4.21e-06 \*\*\*

---

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

Residual standard error: 6.975 on 66 degrees of freedom

Multiple R-squared: 0.4776, Adjusted R-squared: 0.4539

F-statistic: 20.11 on 3 and 66 DF, p-value: 2.273e-09

g)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Model B new |  | Model B |  |
| R² | .4776 |  | .5095 | B |
| R² adj | .4539 |  | 0.4882 | B |
| s | 6.975 | B | 10.42 |  |
| F p-value | 2.273e-09 |  | 1.026e-10 | B |

B = better

Thus this modified model has not improved

attach(cl)

cor(cbind(GPA,jobs,HW3),use="pairwise.complete.obs")

> cor(cbind(GPA,jobs,HW3),use="pairwise.complete.obs")

GPA jobs HW3

GPA 1.00000000 -0.01097167 0.3270477

jobs -0.01097167 1.00000000 -0.2689029

HW3 0.32704769 -0.26890287 1.0000000

#the cor above are small predicting the variables are not inflated

library(Rcmdr)

vif(modelB)

GPA jobs HW3

1.209737 1.037755 1.251228

There are no severe colleanearity,since they don’t exceed the 10 cut off point

> modelD=lm(Test3~factor(Laptop))

> summary(modelD)

Call:

lm(formula = Test3 ~ factor(Laptop))

Residuals:

Min 1Q Median 3Q Max

-60.500 -2.259 1.241 6.833 25.500

Coefficients:

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

(Intercept) 72.00 13.83 5.207 2.31e-06 \*\*\*

factor(Laptop)Apple 10.50 16.93 0.620 0.537

factor(Laptop)Asus 21.50 16.93 1.270 0.209

factor(Laptop)Compaq 24.00 19.55 1.227 0.224

factor(Laptop)Dell 16.68 14.31 1.165 0.248

factor(Laptop)HP 10.17 14.21 0.716 0.477

factor(Laptop)Mac 15.26 14.08 1.084 0.283

factor(Laptop)PC 14.00 19.55 0.716 0.477

factor(Laptop)Sony 2.00 19.55 0.102 0.919

factor(Laptop)Toshiba -11.50 15.46 -0.744 0.460

factor(Laptop)Vaio 11.00 16.93 0.650 0.518

---

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

Residual standard error: 13.83 on 62 degrees of freedom

Multiple R-squared: 0.2243, Adjusted R-squared: 0.09918

F-statistic: 1.793 on 10 and 62 DF, p-value: 0.08065

#above tells me that laptop choice IS not significant as F p-value is .08<.05 (alpha)

And Rsqr R sqr adj are both very small <.3

And none of the above factors are significant as p-value are all below .05\

k)

16.68

Since the Acer is the baseline at default

If someone owns a Dell laptop,they will score 16.78 points higher on average on Test3 than someone who owns an Acer laptop.

i)

> modelD=lm(Test3~HW3+GPA+factor(Laptop))

> summary(modelD)

Call:

lm(formula = Test3 ~ HW3 + GPA + factor(Laptop))

Residuals:

Min 1Q Median 3Q Max

-37.265 -4.375 -0.138 4.975 20.994

Coefficients:

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

(Intercept) 34.2635 13.7568 2.491 0.0155 \*

HW3 0.6102 0.1097 5.562 6.52e-07 \*\*\*

GPA 9.9997 3.2890 3.040 0.0035 \*\*

factor(Laptop)Apple -4.8586 12.5497 -0.387 0.7000

factor(Laptop)Asus -0.5628 12.8542 -0.044 0.9652

factor(Laptop)Compaq 7.7373 14.4616 0.535 0.5946

factor(Laptop)Dell -5.6505 10.9009 -0.518 0.6061

factor(Laptop)HP -4.1200 10.5492 -0.391 0.6975

factor(Laptop)Mac -5.1077 10.6420 -0.480 0.6330

factor(Laptop)PC 3.3811 14.5359 0.233 0.8169

factor(Laptop)Sony -5.5068 14.3190 -0.385 0.7019

factor(Laptop)Toshiba -23.9980 11.4258 -2.100 0.0399 \*

factor(Laptop)Vaio -9.3053 12.7588 -0.729 0.4686

---

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

Residual standard error: 10.1 on 60 degrees of freedom

Multiple R-squared: 0.5998, Adjusted R-squared: 0.5198

F-statistic: 7.494 on 12 and 60 DF, p-value: 3.295e-08

#now,the model becomes significant by F-pvalue =3.295e-08, and R sqr and Rsqr adj are both increased a lot.that is because the added 2 variables attribute more explanation to the model.They are more significant than the laptop choice though Toshiba turns significant now.