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

> attach(gc)

> model1=lm(Value~Location+Property+House+Age+Rooms+Baths+Garage)

> summary(model1)

Call:

lm(formula = Value ~ Location + Property + House + Age + Rooms +

Baths + Garage)

Residuals:

Min 1Q Median 3Q Max

-140.621 -35.229 5.011 34.904 134.388

Coefficients:

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

(Intercept) 49.43142 40.11913 1.232 0.22147

LocationGlen Cove 56.43512 17.69159 3.190 0.00202 \*\*

LocationRoslyn 210.40575 18.18675 11.569 < 2e-16 \*\*\*

Property 342.95198 68.73001 4.990 3.39e-06 \*\*\*

House 0.11480 0.01545 7.430 9.87e-11 \*\*\*

Age -0.58526 0.38984 -1.501 0.13717

Rooms -8.23645 5.25209 -1.568 0.12073

Baths 26.91846 12.46469 2.160 0.03376 \*

Garage 5.01186 14.10327 0.355 0.72324

---

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

Residual standard error: 61.94 on 81 degrees of freedom

Multiple R-squared: 0.9009, Adjusted R-squared: 0.8911

F-statistic: 92.02 on 8 and 81 DF, p-value: < 2.2e-16

> step(model1)

Start: AIC=751.23

Value ~ Location + Property + House + Age + Rooms + Baths + Garage

Df Sum of Sq RSS AIC

- Garage 1 485 311246 749.37

<none> 310762 751.23

- Age 1 8647 319409 751.70

- Rooms 1 9435 320197 751.92

- Baths 1 17893 328655 754.27

- Property 1 95525 406287 773.35

- House 1 211785 522547 796.00

- Location 2 575594 886356 841.56

Step: AIC=749.37

Value ~ Location + Property + House + Age + Rooms + Baths

Df Sum of Sq RSS AIC

<none> 311246 749.37

- Rooms 1 9771 321017 750.15

- Age 1 13709 324955 751.25

- Baths 1 19400 330646 752.81

- Property 1 98385 409632 772.09

- House 1 227544 538790 796.75

- Location 2 577148 888394 839.76

Call:

lm(formula = Value ~ Location + Property + House + Age + Rooms +

Baths)

Coefficients:

(Intercept) LocationGlen Cove LocationRoslyn Property

52.4687 56.9588 210.7564 345.7529

House Age Rooms Baths

0.1160 -0.6504 -8.3625 27.6472

#we have Garage dropped

> model2=lm(Value~Location+Property+House+Age+Rooms+Baths)

> summary(model2)

Call:

lm(formula = Value ~ Location + Property + House + Age + Rooms +

Baths)

Residuals:

Min 1Q Median 3Q Max

-140.497 -35.974 7.079 36.072 130.553

Coefficients:

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

(Intercept) 52.46867 38.98879 1.346 0.18210

LocationGlen Cove 56.95883 17.53592 3.248 0.00168 \*\*

LocationRoslyn 210.75640 18.06296 11.668 < 2e-16 \*\*\*

Property 345.75289 67.91187 5.091 2.23e-06 \*\*\*

House 0.11602 0.01498 7.743 2.25e-11 \*\*\*

Age -0.65039 0.34224 -1.900 0.06089 .

Rooms -8.36245 5.21211 -1.604 0.11247

Baths 27.64724 12.22915 2.261 0.02642 \*

---

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

Residual standard error: 61.61 on 82 degrees of freedom

Multiple R-squared: 0.9007, Adjusted R-squared: 0.8922

F-statistic: 106.3 on 7 and 82 DF, p-value: < 2.2e-16

> model2=lm(Value~Location+Property+House+Age+Baths)

> summary(model2)

Call:

lm(formula = Value ~ Location + Property + House + Age + Baths)

Residuals:

Min 1Q Median 3Q Max

-150.098 -37.497 6.978 37.459 135.667

Coefficients:

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

(Intercept) 23.39542 34.84783 0.671 0.5039

LocationGlen Cove 57.73716 17.69466 3.263 0.0016 \*\*

LocationRoslyn 213.45859 18.15401 11.758 < 2e-16 \*\*\*

Property 347.01493 68.54826 5.062 2.45e-06 \*\*\*

House 0.10614 0.01379 7.697 2.60e-11 \*\*\*

Age -0.79206 0.33377 -2.373 0.0200 \*

Baths 26.38375 12.31895 2.142 0.0351 \*

---

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

Residual standard error: 62.19 on 83 degrees of freedom

Multiple R-squared: 0.8976, Adjusted R-squared: 0.8902

F-statistic: 121.3 on 6 and 83 DF, p-value: < 2.2e-16

We dropped garage ( 0.72324 ) then room and its p-value is .1127

> modelA=model2

plot(Property,Value)



rstandard = rstandard(modelB)

leverages = hatvalues(modelB)

par(mfrow=c(1,2))

hist(rstandard)

hist(leverages)



The 1st one appears roughly normal,while the other one appears skewed right.

rstandard = rstandard(modelA)

leverages = hatvalues(modelA)

> rstandard[order(rstandard)]

62 40 88 33 14

-2.536012805 -2.433274698 -2.166858243 -2.074233335 -1.954354089

35 26 75 48 54

-1.917245736 -1.889514395 -1.418507848 -1.417969584 -1.310842128

61 39 37 49 42

-1.148245203 -1.100121025 -1.000451248 -0.951419422 -0.945535170

13 50 46 25 41

-0.844459847 -0.816305408 -0.815351520 -0.791699597 -0.738899601

55 21 11 19 77

-0.725837214 -0.648175391 -0.630530534 -0.594821389 -0.558887066

9 4 17 2 71

-0.549608357 -0.536712107 -0.486765716 -0.409465377 -0.361614024

24 1 65 44 67

-0.281306274 -0.179142605 -0.167586716 -0.136497687 -0.128828202

80 69 74 32 64

-0.073277327 -0.055173678 -0.034036266 0.002578859 0.047720277

82 87 66 7 60

0.050707893 0.055791033 0.065925073 0.070490116 0.102430866

89 6 83 47 43

0.129397253 0.145690222 0.147851915 0.173798943 0.185005658

63 27 18 70 3

0.185465192 0.188949431 0.227114770 0.243914926 0.252528238

76 5 23 12 16

0.269583007 0.270739431 0.285993180 0.288156223 0.330619040

20 22 51 52 84

0.347321553 0.379041913 0.386111805 0.479019639 0.485316886

72 30 85 59 73

0.493053803 0.571165460 0.638242997 0.671271913 0.674253450

28 79 78 81 15

0.691823185 0.708559623 0.733742360 0.753306854 0.786229203

90 38 68 86 29

0.820352581 0.851402978 0.906111304 0.957045283 0.998850568

58 53 31 57 36

1.242385998 1.425791271 1.481179411 1.554047554 1.585195615

34 45 8 10 56

1.786378053 1.964689766 2.030301872 2.036288367 2.306053813

> leverages[order(leverages)]

13 58 90 31 38 53

0.03453314 0.03502530 0.03643075 0.03785108 0.03864404 0.04035308

41 27 87 52 2 7

0.04131754 0.04224452 0.04305985 0.04371765 0.04522154 0.04557749

84 11 5 85 76 70

0.04604721 0.04626872 0.04679678 0.04722718 0.04745093 0.04832408

68 29 24 47 67 43

0.04866150 0.04892785 0.04925059 0.04972142 0.05018994 0.05188973

81 73 23 72 59 45

0.05226542 0.05264651 0.05346571 0.05427465 0.05471582 0.05502874

57 86 82 15 16 89

0.05529214 0.05545301 0.05793351 0.05803267 0.05925454 0.05947337

36 80 69 55 71 77

0.05995382 0.06005266 0.06081543 0.06188785 0.06222704 0.06267958

37 17 63 83 35 78

0.06312401 0.06327913 0.06328807 0.06467005 0.06600695 0.06740883

60 42 79 9 28 20

0.06767464 0.06779888 0.06955080 0.06980056 0.07003120 0.07120861

33 6 25 12 3 75

0.07230188 0.07537056 0.07545537 0.07563533 0.07571930 0.07786446

49 74 64 39 46 21

0.08576724 0.08637770 0.08762068 0.09211148 0.09227523 0.09371264

62 1 19 4 30 56

0.09426939 0.09887103 0.09948009 0.10062796 0.10503149 0.10513176

14 32 65 66 54 48

0.10714308 0.10790900 0.10892129 0.11472634 0.11600831 0.11740665

22 40 44 26 34 8

0.11848397 0.11968627 0.12504078 0.13354286 0.14239768 0.14627420

50 88 10 61 51 18

0.15581423 0.15989592 0.16782924 0.17337901 0.18439505 0.20349648

>

62 40 88 33 `8 10 56

Above points are all out by 2 and -2 in rstandard’s.

> dim(gc)

[1] 90 9

> 3\*10/90

There is no high leverage point

> modelC=lm(Value~Location+Property+House+Age+Baths,subset=-c(62,40,88,33,8,10,56)

+ )

> summary(modelC)

Call:

lm(formula = Value ~ Location + Property + House + Age + Baths,

subset = -c(62, 40, 88, 33, 8, 10, 56))

Residuals:

Min 1Q Median 3Q Max

-113.856 -29.025 -0.614 31.770 118.451

Coefficients:

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

(Intercept) 42.57914 29.64877 1.436 0.155072

LocationGlen Cove 34.24530 14.77891 2.317 0.023189 \*

LocationRoslyn 207.21355 15.86334 13.062 < 2e-16 \*\*\*

Property 416.32296 58.95160 7.062 6.71e-10 \*\*\*

House 0.08360 0.01309 6.387 1.22e-08 \*\*\*

Age -0.81098 0.29020 -2.795 0.006575 \*\*

Baths 38.13901 10.48501 3.637 0.000499 \*\*\*

---

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

Residual standard error: 49.79 on 76 degrees of freedom

Multiple R-squared: 0.9327, Adjusted R-squared: 0.9274

F-statistic: 175.5 on 6 and 76 DF, p-value: < 2.2e-16

#Age Baths at p-value get smaller and Rsqr Rsqr adj get larger and F-pvalue remains small where has improved the modelA

3.

> model1=lm(Wins~League + ERA + Runs + Hits + Walks +Saves + Errors)

> summary(model1)

Call:

lm(formula = Wins ~ League + ERA + Runs + Hits + Walks + Saves +

Errors)

Residuals:

Min 1Q Median 3Q Max

-4.8384 -0.9533 -0.0105 1.8841 3.4440

Coefficients:

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

(Intercept) 7.739e+01 2.460e+01 3.146 0.00469 \*\*

League -8.646e-01 1.128e+00 -0.766 0.45173

ERA -1.021e+01 3.756e+00 -2.719 0.01252 \*

Runs 7.169e-02 7.876e-03 9.102 6.49e-09 \*\*\*

Hits -1.469e-02 1.965e-02 -0.747 0.46268

Walks 1.156e-06 1.827e-02 0.000 0.99995

Saves 6.694e-01 8.813e-02 7.596 1.38e-07 \*\*\*

Errors -1.150e-01 3.313e-02 -3.471 0.00217 \*\*

---

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

Residual standard error: 2.507 on 22 degrees of freedom

Multiple R-squared: 0.9634, Adjusted R-squared: 0.9518

F-statistic: 82.82 on 7 and 22 DF, p-value: 2.495e-14

> model2=lm( Wins~ League + ERA + Runs + Hits + Walks +Saves + Errors +Hits\*Saves)

> summary(model2)

Call:

lm(formula = Wins ~ League + ERA + Runs + Hits + Walks + Saves +

Errors + Hits \* Saves)

Residuals:

Min 1Q Median 3Q Max

-4.7508 -1.0465 0.0046 1.8243 3.5895

Coefficients:

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

(Intercept) 4.667e+01 7.563e+01 0.617 0.5438

League -7.417e-01 1.185e+00 -0.626 0.5381

ERA -1.041e+01 3.854e+00 -2.701 0.0134 \*

Runs 7.154e-02 8.034e-03 8.905 1.42e-08 \*\*\*

Hits 6.631e-03 5.341e-02 0.124 0.9024

Walks -9.735e-04 1.875e-02 -0.052 0.9591

Saves 1.413e+00 1.730e+00 0.817 0.4232

Errors -1.132e-01 3.402e-02 -3.327 0.0032 \*\*

Hits:Saves -4.968e-04 1.154e-03 -0.431 0.6712

---

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

Residual standard error: 2.554 on 21 degrees of freedom

Multiple R-squared: 0.9638, Adjusted R-squared: 0.95

F-statistic: 69.8 on 8 and 21 DF, p-value: 2.176e-13

Ho: Beta8=0;

H1: Beta8!=0;

> dim(model2)

NULL

> dim(bb)

[1] 30 9

> #n=30,k=8

> qf(.95,9,21)

[1] 2.366048

>#our tscore = 0.431 <2.37,we can’t reject the null

#Hits\*Saves is insignificant

b)

c)

The following pairs are high in corr:

Wins and ERA -.63699,Runs and Wins:.606,

Hits and ERA =.8747 Wins and Saves= .784

> cor(cbind(Wins, ERA , Runs, Hits , Walks , Saves,Errors),use="pairwise.complete.obs")

Wins ERA Runs Hits Walks Saves

Wins 1.0000000 -0.63699126 0.60649067 -0.525254672 -0.3875021 0.7836745

ERA -0.6369913 1.00000000 0.06695488 0.874711236 0.2759977 -0.4372881

Runs 0.6064907 0.06695488 1.00000000 0.167872822 -0.3123794 0.3166633

Hits -0.5252547 0.87471124 0.16787282 1.000000000 -0.1001811 -0.4396077

Walks -0.3875021 0.27599772 -0.31237942 -0.100181131 1.0000000 -0.2287696

Saves 0.7836745 -0.43728815 0.31666333 -0.439607686 -0.2287696 1.0000000

Errors -0.4048837 0.18081365 -0.25508144 -0.002847303 0.2825716 -0.1441390

Errors

Wins -0.404883690

ERA 0.180813653

Runs -0.255081444

Hits -0.002847303

Walks 0.282571584

Saves -0.144138992

Errors 1.000000000

> library(Rcmdr)

Loading required package: splines

Loading required package: RcmdrMisc

Loading required package: car

Loading required package: sandwich

Rcmdr Version 2.1-3

> vif(model1)

Wins ERA Runs Hits Walks Saves Errors

26.640007 14.481367 7.529029 11.567022 2.709723 5.810076 1.885637

Thus,we have Wins ERA,Hits and Walks as high var inflation factor as their value above exceeds 10

> step(model1)

Start: AIC=61.83

Wins ~ League + ERA + Runs + Hits + Walks + Saves + Errors

Df Sum of Sq RSS AIC

- Walks 1 0.00 138.24 59.834

- Hits 1 3.51 141.75 60.586

- League 1 3.69 141.93 60.624

<none> 138.24 61.834

- ERA 1 46.47 184.71 68.527

- Errors 1 75.72 213.96 72.937

- Saves 1 362.52 500.75 98.448

- Runs 1 520.61 658.85 106.679

Step: AIC=59.83

Wins ~ League + ERA + Runs + Hits + Saves + Errors

Df Sum of Sq RSS AIC

- League 1 4.10 142.34 58.711

- Hits 1 7.75 145.99 59.470

<none> 138.24 59.834

- Errors 1 76.07 214.31 70.987

- ERA 1 111.87 250.11 75.621

- Saves 1 391.68 529.92 98.146

- Runs 1 521.10 659.34 104.701

Step: AIC=58.71

Wins ~ ERA + Runs + Hits + Saves + Errors

Df Sum of Sq RSS AIC

- Hits 1 7.02 149.36 58.155

<none> 142.34 58.711

- Errors 1 74.93 217.27 69.399

- ERA 1 109.46 251.80 73.823

- Saves 1 388.82 531.16 96.216

- Runs 1 656.44 798.78 108.457

Step: AIC=58.16

Wins ~ ERA + Runs + Saves + Errors

Df Sum of Sq RSS AIC

<none> 149.36 58.155

- Errors 1 67.92 217.28 67.400

- Saves 1 440.74 590.10 97.373

- ERA 1 593.28 742.65 104.271

- Runs 1 660.84 810.20 106.883

Call:

lm(formula = Wins ~ ERA + Runs + Saves + Errors)

Coefficients:

(Intercept) ERA Runs Saves Errors

62.10408 -12.25499 0.07238 0.68802 -0.10376

By backward reg:

> summary(model1)

Call:

lm(formula = Wins ~ League + ERA + Runs + Hits + Walks + Saves +

Errors)

Residuals:

Min 1Q Median 3Q Max

-4.8384 -0.9533 -0.0105 1.8841 3.4440

Coefficients:

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

(Intercept) 7.739e+01 2.460e+01 3.146 0.00469 \*\*

League -8.646e-01 1.128e+00 -0.766 0.45173

ERA -1.021e+01 3.756e+00 -2.719 0.01252 \*

Runs 7.169e-02 7.876e-03 9.102 6.49e-09 \*\*\*

Hits -1.469e-02 1.965e-02 -0.747 0.46268

Walks 1.156e-06 1.827e-02 0.000 0.99995

Saves 6.694e-01 8.813e-02 7.596 1.38e-07 \*\*\*

Errors -1.150e-01 3.313e-02 -3.471 0.00217 \*\*

---

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

Residual standard error: 2.507 on 22 degrees of freedom

Multiple R-squared: 0.9634, Adjusted R-squared: 0.9518

F-statistic: 82.82 on 7 and 22 DF, p-value: 2.495e-14

> model3=lm(Wins~League + ERA + Runs + Hits +Saves + Errors)

> summary(model3)

Call:

lm(formula = Wins ~ League + ERA + Runs + Hits + Saves + Errors)

Residuals:

Min 1Q Median 3Q Max

-4.8384 -0.9532 -0.0105 1.8841 3.4440

Coefficients:

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

(Intercept) 77.395744 14.037376 5.514 1.32e-05 \*\*\*

League -0.864581 1.046737 -0.826 0.417303

ERA -10.212705 2.367245 -4.314 0.000257 \*\*\*

Runs 0.071691 0.007699 9.311 2.89e-09 \*\*\*

Hits -0.014688 0.012938 -1.135 0.267942

Saves 0.669402 0.082923 8.073 3.67e-08 \*\*\*

Errors -0.115003 0.032325 -3.558 0.001675 \*\*

---

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

Residual standard error: 2.452 on 23 degrees of freedom

Multiple R-squared: 0.9634, Adjusted R-squared: 0.9539

F-statistic: 101 on 6 and 23 DF, p-value: 2.351e-15

> model4=lm(Wins~ ERA + Runs + Hits +Saves + Errors)

> summary(model4)

Call:

lm(formula = Wins ~ ERA + Runs + Hits + Saves + Errors)

Residuals:

Min 1Q Median 3Q Max

-4.5138 -1.1360 -0.1843 1.5008 3.5743

Coefficients:

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

(Intercept) 73.478527 13.124224 5.599 9.21e-06 \*\*\*

ERA -10.078327 2.345962 -4.296 0.000249 \*\*\*

Runs 0.074158 0.007049 10.521 1.81e-10 \*\*\*

Hits -0.013949 0.012821 -1.088 0.287388

Saves 0.666241 0.082284 8.097 2.55e-08 \*\*\*

Errors -0.114066 0.032091 -3.554 0.001610 \*\*

---

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

Residual standard error: 2.435 on 24 degrees of freedom

Multiple R-squared: 0.9624, Adjusted R-squared: 0.9545

F-statistic: 122.7 on 5 and 24 DF, p-value: 2.787e-16

> model5=lm(Wins~ ERA + Runs +Saves + Errors)

> summary(model5)

Call:

lm(formula = Wins ~ ERA + Runs + Saves + Errors)

Residuals:

Min 1Q Median 3Q Max

-5.2310 -1.1732 0.1601 1.5932 3.3885

Coefficients:

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

(Intercept) 62.104076 7.963579 7.799 3.73e-08 \*\*\*

ERA -12.254986 1.229788 -9.965 3.44e-10 \*\*\*

Runs 0.072381 0.006882 10.517 1.15e-10 \*\*\*

Saves 0.688019 0.080105 8.589 6.28e-09 \*\*\*

Errors -0.103761 0.030774 -3.372 0.00243 \*\*

---

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

Residual standard error: 2.444 on 25 degrees of freedom

Multiple R-squared: 0.9605, Adjusted R-squared: 0.9542

F-statistic: 152 on 4 and 25 DF, p-value: < 2.2e-16

Ended with the same model:

lm(formula = Wins ~ ERA + Runs + Saves + Errors)

b) > model7=lm(Wins~ ERA + Walks +Saves )

> summary(model7)

Call:

lm(formula = Wins ~ ERA + Walks + Saves)

Residuals:

Min 1Q Median 3Q Max

-9.6662 -4.2195 -0.5382 3.8981 12.4223

Coefficients:

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

(Intercept) 100.89971 21.71228 4.647 8.53e-05 \*\*\*

ERA -8.72088 2.98157 -2.925 0.00706 \*\*

Walks -0.04099 0.02690 -1.524 0.13955

Saves 1.00831 0.18581 5.427 1.09e-05 \*\*\*

---

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

Residual standard error: 6.1 on 26 degrees of freedom

Multiple R-squared: 0.7441, Adjusted R-squared: 0.7146

F-statistic: 25.2 on 3 and 26 DF, p-value: 7.373e-08

F-p value is smaller than alpha=.05

#Thus,they are jointly significant.