1.

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

> attch(ts)

cor(cbind(PS,PA,FG,FGA,S,SA,R,RA),use="pairwise.complete.obs")

> attach(ts)

> model1=lm(Wins~PS+PA+FG+FGA+S+SA+R+RA)

> cor(cbind(PS,PA,FG,FGA,S,SA,R,RA),use="pairwise.complete.obs")

PS PA FG FGA S SA

PS 1.00000000 0.41346340 0.695919840 -0.07732249 0.14471627 0.07810202

PA 0.41346340 1.00000000 -0.020079568 0.70841024 0.12585481 0.62216384

FG 0.69591984 -0.02007957 1.000000000 -0.29853646 0.06703415 -0.11484072

FGA -0.07732249 0.70841024 -0.298536456 1.00000000 0.16214272 0.33992394

S 0.14471627 0.12585481 0.067034148 0.16214272 1.00000000 0.19367803

SA 0.07810202 0.62216384 -0.114840720 0.33992394 0.19367803 1.00000000

R 0.08878049 -0.22367048 -0.002085374 -0.50771237 -0.35077144 0.05153086

RA 0.22457412 0.69336499 -0.382980134 0.38035185 0.20053624 0.42652975

R RA

PS 0.088780486 0.2245741

PA -0.223670482 0.6933650

FG -0.002085374 -0.3829801

FGA -0.507712371 0.3803518

S -0.350771441 0.2005362

SA 0.051530859 0.4265298

R 1.000000000 -0.1965825

RA -0.196582494 1.0000000

a)

> summary(model1)

Call:

lm(formula = Wins ~ PS + PA + FG + FGA + S + SA + R + RA)

Residuals:

Min 1Q Median 3Q Max

-5.8000 -2.0019 0.0591 1.7143 6.5259

Coefficients:

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

(Intercept) -21.734177 89.350108 -0.243 0.810

PS 2.586712 0.382822 6.757 1.1e-06 \*\*\*

PA -2.794905 0.513028 -5.448 2.1e-05 \*\*\*

FG 0.467988 1.155195 0.405 0.689

FGA 0.464091 1.124596 0.413 0.684

S -0.006644 0.875005 -0.008 0.994

SA -0.454650 1.566086 -0.290 0.774

R 0.890540 0.663876 1.341 0.194

RA 0.162243 0.745372 0.218 0.830

---

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

Residual standard error: 3.235 on 21 degrees of freedom

Multiple R-squared: 0.9575, Adjusted R-squared: 0.9414

F-statistic: 59.19 on 8 and 21 DF, p-value: 1.128e-12

# Since the significancy focus on PA and PS,while the others are not significant.

And the above highlight indicate the model is significant as it has high R sqr and R sqr adj and low F p-value.

BELOW are the high correlation pairs

PS~PA 0.69591984

PA-fga 0.70841024

Sa-pa 0.62216384

Etc..Thus it could have high multicollearity problems.

b)

> library(Rcmdr)

> vif(model1)

PS PA FG FGA S SA R RA

6.797668 15.136656 8.706193 7.891327 1.483505 2.678605 2.810951 7.741816

As shown,PA has 15.1367 exceeding 10 thus this is the VARIANCE inflation factor.

c)

#everystep has the highest P-value row out.

> summary(model1)

Call:

lm(formula = Wins ~ PS + PA + FG + FGA + S + SA + R + RA)

Residuals:

Min 1Q Median 3Q Max

-5.8000 -2.0019 0.0591 1.7143 6.5259

Coefficients:

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

(Intercept) -21.734177 89.350108 -0.243 0.810

PS 2.586712 0.382822 6.757 1.1e-06 \*\*\*

PA -2.794905 0.513028 -5.448 2.1e-05 \*\*\*

FG 0.467988 1.155195 0.405 0.689

FGA 0.464091 1.124596 0.413 0.684

S -0.006644 0.875005 -0.008 0.994

SA -0.454650 1.566086 -0.290 0.774

R 0.890540 0.663876 1.341 0.194

RA 0.162243 0.745372 0.218 0.830

---

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

Residual standard error: 3.235 on 21 degrees of freedom

Multiple R-squared: 0.9575, Adjusted R-squared: 0.9414

F-statistic: 59.19 on 8 and 21 DF, p-value: 1.128e-12

> modelA=lm(Wins~PS+PA+FG+FGA+SA+R+RA)

> summary(modelA)

Call:

lm(formula = Wins ~ PS + PA + FG + FGA + SA + R + RA)

Residuals:

Min 1Q Median 3Q Max

-5.7939 -1.9992 0.0597 1.7092 6.5247

Coefficients:

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

(Intercept) -21.7205 87.2783 -0.249 0.806

PS 2.5860 0.3631 7.122 3.85e-07 \*\*\*

PA -2.7935 0.4653 -6.003 4.84e-06 \*\*\*

FG 0.4677 1.1281 0.415 0.682

FGA 0.4621 1.0669 0.433 0.669

SA -0.4595 1.3978 -0.329 0.745

R 0.8917 0.6303 1.415 0.171

RA 0.1612 0.7151 0.225 0.824

---

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

Residual standard error: 3.16 on 22 degrees of freedom

Multiple R-squared: 0.9575, Adjusted R-squared: 0.944

F-statistic: 70.87 on 7 and 22 DF, p-value: 1.276e-13

#getting rid of RA first

> modelB=lm(Wins~PS+PA+FG+FGA+R+SA)

> summary(modelB)

Call:

lm(formula = Wins ~ PS + PA + FG + FGA + R + RA)

Residuals:

Min 1Q Median 3Q Max

-5.8615 -1.7714 -0.0556 1.6468 6.5706

Coefficients:

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

(Intercept) -22.5683 85.5319 -0.264 0.794

PS 2.6299 0.3311 7.944 4.84e-08 \*\*\*

PA -2.8726 0.3905 -7.357 1.75e-07 \*\*\*

FG 0.4296 1.1001 0.391 0.700

FGA 0.5476 1.0144 0.540 0.595

R 0.8620 0.6115 1.410 0.172

RA 0.1641 0.7011 0.234 0.817

---

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

Residual standard error: 3.098 on 23 degrees of freedom

Multiple R-squared: 0.9573, Adjusted R-squared: 0.9462

F-statistic: 86 on 6 and 23 DF, p-value: 1.374e-14

> modelC=lm(Wins~PS+PA+FG+FGA+R)

> summary(modelC)

Call:

lm(formula = Wins ~ PS + PA + FG + FGA + R)

Residuals:

Min 1Q Median 3Q Max

-5.7779 -1.7974 -0.0372 1.6575 6.7557

Coefficients:

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

(Intercept) -6.0316 47.2283 -0.128 0.899

PS 2.6729 0.2699 9.904 5.93e-10 \*\*\*

PA -2.8042 0.2541 -11.036 6.93e-11 \*\*\*

FG 0.2162 0.6027 0.359 0.723

FGA 0.3910 0.7473 0.523 0.606

R 0.7708 0.4620 1.669 0.108

---

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

Residual standard error: 3.037 on 24 degrees of freedom

Multiple R-squared: 0.9572, Adjusted R-squared: 0.9483

F-statistic: 107.4 on 5 and 24 DF, p-value: 1.28e-15

> modelD=lm(Wins~PS+PA+FGA+R)

> summary(modelD)

Call:

lm(formula = Wins ~ PS + PA + FGA + R)

Residuals:

Min 1Q Median 3Q Max

-5.6722 -1.7803 -0.1918 1.5499 6.6647

Coefficients:

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

(Intercept) 2.1258 40.6643 0.052 0.959

PS 2.7428 0.1834 14.955 5.62e-14 \*\*\*

PA -2.8314 0.2383 -11.880 8.90e-12 \*\*\*

FGA 0.3760 0.7330 0.513 0.612

R 0.7279 0.4384 1.660 0.109

---

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

Residual standard error: 2.983 on 25 degrees of freedom

Multiple R-squared: 0.957, Adjusted R-squared: 0.9501

F-statistic: 139.1 on 4 and 25 DF, p-value: < 2.2e-16

> modelE=lm(Wins~PS+PA+R)

> summary(modelE)

Call:

lm(formula = Wins ~ PS + PA + R)

Residuals:

Min 1Q Median 3Q Max

-5.4436 -1.9400 -0.5152 1.7639 6.6991

Coefficients:

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

(Intercept) 19.3802 22.5297 0.860 0.398

PS 2.6901 0.1498 17.962 3.55e-16 \*\*\*

PA -2.7322 0.1375 -19.877 < 2e-16 \*\*\*

R 0.6183 0.3773 1.639 0.113

---

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

Residual standard error: 2.941 on 26 degrees of freedom

Multiple R-squared: 0.9565, Adjusted R-squared: 0.9515

F-statistic: 190.8 on 3 and 26 DF, p-value: < 2.2e-16

> model2=lm(Wins~PS+PA)

> summary(model2)

Call:

lm(formula = Wins ~ PS + PA)

Residuals:

Min 1Q Median 3Q Max

-5.0978 -1.6942 -0.4042 2.1241 6.5391

Coefficients:

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

(Intercept) 46.6301 15.6679 2.976 0.00609 \*\*

PS 2.7402 0.1511 18.133 < 2e-16 \*\*\*

PA -2.7969 0.1357 -20.609 < 2e-16 \*\*\*

---

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

Residual standard error: 3.031 on 27 degrees of freedom

Multiple R-squared: 0.9521, Adjusted R-squared: 0.9485

F-statistic: 268.1 on 2 and 27 DF, p-value: < 2.2e-16

d)

Ho:Beta0=Beta1 =0;

H1:at least one Beta!=0;

AS it has a almost 0 p-value<Alpha=.05

We should reject the null and believe model2 is significant,

e)

plot(Wins,S)



plot(Wins,SA)



plot(Wins,R)



plot(Wins,RA)



I think R need a sqr term

> model1=lm(Wins~PS+PA+FG+FGA+S+SA+R+RA)

> R2=R\*R

> model1=lm(Wins~PS+PA+FG+FGA+S+SA+R2+RA)

> modelF=lm(Wins~PS+PA+FG+FGA+S+SA+R2+RA)

> summary(modelF)

Call:

lm(formula = Wins ~ PS + PA + FG + FGA + S + SA + R2 + RA)

Residuals:

Min 1Q Median 3Q Max

-5.8018 -1.9802 0.0606 1.7029 6.5671

Coefficients:

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

(Intercept) -2.304886 78.067600 -0.030 0.977

PS 2.591339 0.381530 6.792 1.02e-06 \*\*\*

PA -2.794000 0.513072 -5.446 2.11e-05 \*\*\*

FG 0.451950 1.149313 0.393 0.698

FGA 0.461576 1.124772 0.410 0.686

S -0.019046 0.873199 -0.022 0.983

SA -0.445898 1.565224 -0.285 0.779

R2 0.010625 0.007945 1.337 0.195

RA 0.151048 0.741567 0.204 0.841

---

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

Residual standard error: 3.235 on 21 degrees of freedom

Multiple R-squared: 0.9575, Adjusted R-squared: 0.9413

F-statistic: 59.16 on 8 and 21 DF, p-value: 1.134e-12

> step(modelF)

Start: AIC=77.75

Wins ~ PS + PA + FG + FGA + S + SA + R2 + RA

Df Sum of Sq RSS AIC

- S 1 0.00 219.81 75.747

- RA 1 0.43 220.24 75.806

- SA 1 0.85 220.66 75.862

- FG 1 1.62 221.43 75.967

- FGA 1 1.76 221.57 75.986

<none> 219.81 77.747

- R2 1 18.72 238.53 78.199

- PA 1 310.40 530.20 102.162

- PS 1 482.85 702.66 110.610

Step: AIC=75.75

Wins ~ PS + PA + FG + FGA + SA + R2 + RA

Df Sum of Sq RSS AIC

- RA 1 0.43 220.24 73.806

- SA 1 1.08 220.89 73.894

- FG 1 1.61 221.43 73.967

- FGA 1 1.83 221.64 73.995

<none> 219.81 75.747

- R2 1 19.88 239.69 76.345

- PA 1 360.23 580.04 102.857

- PS 1 510.22 730.03 109.757

Step: AIC=73.81

Wins ~ PS + PA + FG + FGA + SA + R2

Df Sum of Sq RSS AIC

- SA 1 1.10 221.34 71.956

- FGA 1 1.47 221.71 72.005

- FG 1 1.64 221.88 72.028

<none> 220.24 73.806

- R2 1 26.73 246.98 75.243

- PA 1 589.24 809.48 110.856

- PS 1 708.41 928.65 114.976

Step: AIC=71.96

Wins ~ PS + PA + FG + FGA + R2

Df Sum of Sq RSS AIC

- FG 1 1.19 222.54 70.117

- FGA 1 2.59 223.93 70.304

<none> 221.34 71.956

- R2 1 25.63 246.98 73.243

- PS 1 904.42 1125.77 118.751

- PA 1 1122.97 1344.32 124.073

Step: AIC=70.12

Wins ~ PS + PA + FGA + R2

Df Sum of Sq RSS AIC

- FGA 1 2.40 224.94 68.439

<none> 222.54 70.117

- R2 1 24.49 247.03 71.250

- PA 1 1254.81 1477.35 124.904

- PS 1 1990.30 2212.84 137.025

Step: AIC=68.44

Wins ~ PS + PA + R2

Df Sum of Sq RSS AIC

<none> 224.9 68.439

- R2 1 23.1 248.1 69.374

- PS 1 2786.1 3011.0 144.265

- PA 1 3419.1 3644.0 149.989

Call:

lm(formula = Wins ~ PS + PA + R2)

Coefficients:

(Intercept) PS PA R2

32.305221 2.689618 -2.732533 0.007432

> model3=lm(Wins ~ PS + PA + R2)

> summary(model3)

Call:

lm(formula = Wins ~ PS + PA + R2)

Residuals:

Min 1Q Median 3Q Max

-5.4383 -1.9318 -0.5084 1.7618 6.7161

Coefficients:

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

(Intercept) 32.305221 17.548441 1.841 0.0771 .

PS 2.689618 0.149878 17.945 3.63e-16 \*\*\*

PA -2.732533 0.137453 -19.880 < 2e-16 \*\*\*

R2 0.007432 0.004546 1.635 0.1141

---

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

Residual standard error: 2.941 on 26 degrees of freedom

Multiple R-squared: 0.9565, Adjusted R-squared: 0.9515

F-statistic: 190.7 on 3 and 26 DF, p-value: < 2.2e-16

f)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Model 3 new |  | Model 1 |  |
| R² | .957 |  | .958 | B |
| R² adj | .952 | **B** | .9414 |  |
| s | 2.941 | B | 3.235 |  |
| F p-value | 0 | B | 1.128e-12 |  |
|  | Model 2 new |  |  |  |
| R² | .952 |  |  |  |
| R² adj | .949 |  |  |  |
| s | 3.03 |  |  |  |
| F p-value | 0 |  |  |  |

B=best

As what it appears above model3 appears the best

g)

> dim(ts)

[1] 30 10

rstandard = rstandard(model2)

leverages = hatvalues(model2)

> rstandard[order(rstandard)]

27 9 18 26 23 22

-1.750566634 -1.745659289 -1.437851062 -1.289126500 -1.085044672 -0.971653023

29 30 3 2 20 1

-0.917887062 -0.623723553 -0.474487768 -0.427626679 -0.410470590 -0.309303736

21 15 11 25 24 8

-0.208777415 -0.188660910 -0.161818583 -0.114385873 -0.045116556 -0.004880798

16 17 7 5 10 28

0.050172925 0.285196285 0.335234930 0.657816074 0.743475952 0.872165821

19 13 14 4 12 6

0.988039931 1.001133795 1.162742817 1.516847322 1.940043923 2.207837749

As shown,only #6 has exceeded the cutoff 2 as the outlier

> leverages[order(leverages)]

19 10 6 4 21 11 14

0.04460250 0.04485004 0.04520200 0.05182218 0.05244393 0.05404782 0.05555259

23 26 1 13 16 2 27

0.06073736 0.06798041 0.06817986 0.06910814 0.06979215 0.07377597 0.07695761

30 20 29 25 28 5 15

0.08286451 0.08355289 0.08420508 0.08643767 0.09424525 0.10225096 0.10425228

7 12 22 8 3 17 18

0.10856050 0.10884525 0.12204406 0.12464335 0.12488578 0.18515438 0.21423719

24 9

0.23012453 0.30864376

> 3\*(2+1)/30

[1] 0.3

Thus only #9 is the high leverage point who exceeds .3 cutoff

> model4=lm(Wins ~ PS + PA,subset=-c(6,9))

> summary(model4)

Call:

lm(formula = Wins ~ PS + PA, subset = -c(6, 9))

Residuals:

Min 1Q Median 3Q Max

-4.586 -1.787 -0.185 1.787 5.494

Coefficients:

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

(Intercept) 31.5156 16.0742 1.961 0.0612 .

PS 2.7666 0.1373 20.153 <2e-16 \*\*\*

PA -2.6729 0.1306 -20.472 <2e-16 \*\*\*

---

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

Residual standard error: 2.666 on 25 degrees of freedom

Multiple R-squared: 0.9626, Adjusted R-squared: 0.9596

F-statistic: 321.8 on 2 and 25 DF, p-value: < 2.2e-16

Thus,as shown above,s and Rsqr and Rsqr adj all has improved from model2 indicated as below

Residual standard error: 3.031 on 27 degrees of freedom

Multiple R-squared: 0.9521, Adjusted R-squared: 0.9485

F-statistic: 268.1 on 2 and 27 DF, p-value: < 2.2e-16

h)

Ho: Beta2 =0;

H1: Beta2!=0;

As we found in model2,PA has p-value=0 almost<alpha=.05,we should reject the null

And believe it is significant

i)

plot(residuals(model4) ~ fitted.values(model4), main="Residuals vs.Fitted Value")



As we hope,the plots scatter around the whole table and almost evenly spread,thus it doesn’t show model inequality.

j)

Residual standard error: 2.666 on 25 degrees of freedom

Multiple R-squared: 0.9626, Adjusted R-squared: 0.9596

F-statistic: 321.8 on 2 and 25 DF, p-value: < 2.2e-16

As the summary data shown above,

It has a almost 0 F-pvlaue and high in Rsqr and R sqr adj>0.9

Thus,we say y variable is well explained by Xs variables,so it doesn’t suffer omitted variables

k)

varb = data.frame(PS,PA,FG,FGA,S,SA,R,RA,Wins)

pairs(varb,upper.panel=NULL)



> #PA needs a sqr term

> PA2=PA\*PA

> modelK=lm(lm( Wins ~ PS + PA + PA2))

> summary(modelK)

Call:

lm(formula = lm(Wins ~ PS + PA + PA2))

Residuals:

Min 1Q Median 3Q Max

-4.9638 -1.2353 -0.2679 2.0608 5.7373

Coefficients:

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

(Intercept) -392.93636 227.26433 -1.729 0.0957 .

PS 2.80272 0.14751 19.000 <2e-16 \*\*\*

PA 5.77430 4.42379 1.305 0.2032

PA2 -0.04230 0.02182 -1.938 0.0635 .

---

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

Residual standard error: 2.887 on 26 degrees of freedom

Multiple R-squared: 0.9581, Adjusted R-squared: 0.9533

F-statistic: 198.2 on 3 and 26 DF, p-value: < 2.2e-16

#The modelK is still significant because F-pvalue is almost 0 <alpha

#the R sqr and R sqr adj are both still high and let the model variable Xs EXPlain Y

#Yet PA2 becomes insignificant by alpha =.05 as pvalue <.05

l)

cut=cut(ts$PS,br=c(0,98,104,1000))

table(cut)

model3 = lm(Wins ~ cut)

> summary(model3)

Call:

lm(formula = Wins ~ cut)

Residuals:

Min 1Q Median 3Q Max

-29.062 -6.852 5.938 8.903 16.938

Coefficients:

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

(Intercept) 33.222 4.251 7.815 2.1e-08 \*\*\*

cut(98,104] 10.840 5.314 2.040 0.0512 .

cut(104,1e+03] 11.978 7.113 1.684 0.1037

---

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

Residual standard error: 12.75 on 27 degrees of freedom

Multiple R-squared: 0.1513, Adjusted R-squared: 0.08841

F-statistic: 2.406 on 2 and 27 DF, p-value: 0.1092

#Point score is not a significant predictor,because F-pvalue is .1092 >.05 (alpha)

And the Rsqr and Rsqr adj are both very low,so the whole categorical variable model does n’t

Explain well on Wins,though the cut(0,98] is significant as its p-value is 2.1e-08 <.05(alpha)

m)

TukeyHSD(aov(PA ~ cut))

TukeyHSD(aov(PA ~ cut))

Tukey multiple comparisons of means

95% family-wise confidence level

Fit: aov(formula = PA ~ cut)

$cut

diff lwr upr p adj

(98,104]-(0,98] 1.284097 -3.0242471 5.592442 0.7427324

(104,1e+03]-(0,98] 6.246222 0.4788311 12.013613 0.0317891

(104,1e+03]-(98,104] 4.962125 -0.3355621 10.259812 0.0696377

when the p-value > alpha (0.05), we believe that there is not a significant difference between these two groups.

By the rule shown above,only (104,1e+03]-(0,98] shows a significant difference between these two groups.

there is not a significant difference between the other two groups.