1. Prepare a prediction model for profit of 50\_startups data.

Do transformations for getting better predictions of profit and

make a table containing R^2 value for each prepared model.

R&D Spend -- Research and devolop spend in the past few years

Administration -- spend on administration in the past few years

Marketing Spend -- spend on Marketing in the past few years

State -- states from which data is collected

Profit -- profit of each state in the past few years

**Answer :**

Here our target prediction is the profit

There are 4 column in the table.

Out of which one is state column

We have to prepare dummy values for such column ->

**The code->**

> # MultiLinear Reg prob 1

> st<- read.csv(file.choose())

> #install.packages('mlr')

> # Package to create dummy variables instantly

> library('mlr')

> stfinale<-createDummyFeatures(st, cols = "State")

> stf<-stfinale[,-4]

>

> #install.packages("GGally")

> library("GGally")

> #install.packages("stringi")

> cor(stf)

R.D.Spend Administration Marketing.Spend

R.D.Spend 1.00000000 0.241955245 0.72424813

Administration 0.24195525 1.000000000 -0.03215388

Marketing.Spend 0.72424813 -0.032153875 1.00000000

State.California -0.14316522 -0.015478106 -0.16887523

State.Florida 0.10571106 0.010493089 0.20568545

State.New.York 0.03906816 0.005145226 -0.03366980

State.California State.Florida State.New.York

R.D.Spend -0.14316522 0.10571106 0.039068162

Administration -0.01547811 0.01049309 0.005145226

Marketing.Spend -0.16887523 0.20568545 -0.033669800

State.California 1.00000000 -0.49236596 -0.515151515

State.Florida -0.49236596 1.00000000 -0.492365964

State.New.York -0.51515152 -0.49236596 1.000000000

> ggpairs(stf)

>

> #install.packages("corpcor")

> library('corpcor')

> cor2pcor(cor(stf))

[,1] [,2] [,3] [,4] [,5]

[1,] 1.00000000 0.38679805 0.75142791 -0.02117089 -0.07578548

[2,] 0.38679805 1.00000000 -0.31262870 -0.01162506 0.04561609

[3,] 0.75142791 -0.31262870 1.00000000 -0.09781395 0.19299288

[4,] -0.02117089 -0.01162506 -0.09781395 1.00000000 0.51430596

[5,] -0.07578548 0.04561609 0.19299288 0.51430596 1.00000000

[6,] 0.09862555 -0.03475398 -0.09792443 0.48571199 0.49984572

[,6]

[1,] 0.09862555

[2,] -0.03475398

[3,] -0.09792443

[4,] 0.48571199

[5,] 0.49984572

[6,] 1.00000000

> summary(stf)

R.D.Spend Administration Marketing.Spend State.California

Min. : 0 Min. : 51283 Min. : 0 Min. :0.00

1st Qu.: 39936 1st Qu.:103731 1st Qu.:129300 1st Qu.:0.00

Median : 73051 Median :122700 Median :212716 Median :0.00

Mean : 73722 Mean :121345 Mean :211025 Mean :0.34

3rd Qu.:101603 3rd Qu.:144842 3rd Qu.:299469 3rd Qu.:1.00

Max. :165349 Max. :182646 Max. :471784 Max. :1.00

State.Florida State.New.York

Min. :0.00 Min. :0.00

1st Qu.:0.00 1st Qu.:0.00

Median :0.00 Median :0.00

Mean :0.32 Mean :0.34

3rd Qu.:1.00 3rd Qu.:1.00

Max. :1.00 Max. :1.00

> #Marketng and R.D. have near to strong corelation

> attach(stf)

> model.stf <- lm(Profit~.,data=stfinale)

> summary(model.stf)

Call:

lm(formula = Profit ~ ., data = stfinale)

Residuals:

Min 1Q Median 3Q Max

-33504 -4736 90 6672 17338

Coefficients: (1 not defined because of singularities)

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

(Intercept) 5.008e+04 6.953e+03 7.204 5.76e-09 \*\*\*

R.D.Spend 8.060e-01 4.641e-02 17.369 < 2e-16 \*\*\*

Administration -2.700e-02 5.223e-02 -0.517 0.608

Marketing.Spend 2.698e-02 1.714e-02 1.574 0.123

State.California 4.189e+01 3.256e+03 0.013 0.990

State.Florida 2.407e+02 3.339e+03 0.072 0.943

State.New.York NA NA NA NA

---

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

Residual standard error: 9439 on 44 degrees of freedom

Multiple **R-squared: 0.9508**, Adjusted R-squared: 0.9452

F-statistic: 169.9 on 5 and 44 DF, **p-value: < 2.2e-16**

> #here we can see, which variable to eliminate if its convinient

> model.stf2 <- lm(Profit~ stfinale$R.D.Spend+stfinale$Administration+stfinale$Marketing,data=stfinale)

> summary(model.stf2)

Call:

lm(formula = Profit ~ stfinale$R.D.Spend + stfinale$Administration +

stfinale$Marketing, data = stfinale)

Residuals:

Min 1Q Median 3Q Max

-33534 -4795 63 6606 17275

Coefficients:

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

(Intercept) 5.012e+04 6.572e+03 7.626 1.06e-09 \*\*\*

stfinale$R.D.Spend 8.057e-01 4.515e-02 17.846 < 2e-16 \*\*\*

stfinale$Administration -2.682e-02 5.103e-02 -0.526 0.602

stfinale$Marketing 2.723e-02 1.645e-02 1.655 0.105

---

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

Residual standard error: 9232 on 46 degrees of freedom

Multiple **R-squared: 0.9507**, Adjusted R-squared: 0.9475

F-statistic: 296 on 3 and 46 DF, p-value: < 2.2e-16

>

> library(car)

**> vif(model.stf2)**

**stfinale$R.D.Spend stfinale$Administration stfinale$Marketing**

**2.468903 1.175091 2.326773**

> # gives degree of collinearity between the variables

> #coolinearity values are low thus no need to delete any variable

> avPlots(model.stf)

> # which variable is contributing least towards prediction

> #thus no one has plot with zero slope....thus no need to delete.

> #model is great in itself.

>

> influence.measures(model.stf)

Influence measures of

lm(formula = Profit ~ ., data = stfinale) :

dfb.1\_ dfb.R.D. dfb.Admn dfb.Mr.S dfb.St.C dfb.St.F dffit

1 1.53e-03 -0.000680 -0.00108 -0.002943 0.002293 0.003384 -0.00671

2 -8.24e-02 0.020841 0.04627 0.058280 0.070955 -0.014036 0.15256

3 6.59e-02 0.197811 -0.17477 -0.013702 0.028648 0.182661 0.40914

4 4.12e-02 0.110000 -0.04970 0.073783 -0.163879 -0.208123 0.37650

5 -9.51e-02 -0.153851 0.16342 0.047737 -0.017478 -0.134031 -0.28836

6 -1.00e-01 -0.081405 0.10566 -0.033055 0.115816 0.140425 -0.26364

7 2.59e-03 -0.095187 0.00468 0.082074 -0.050562 -0.011713 -0.12099

8 5.98e-02 -0.045130 -0.04118 0.001538 -0.007763 -0.087547 -0.15671

9 -5.46e-03 0.000992 0.00845 0.005091 -0.010515 -0.012304 0.02068

10 -1.92e-02 -0.087676 0.06482 0.009105 -0.114051 0.003328 -0.19312

11 1.11e-01 0.186811 -0.13740 -0.159583 0.009802 0.239030 0.36595

12 1.35e-01 0.152695 -0.19750 -0.063320 0.191722 0.004731 0.34058

13 -2.00e-02 0.069420 0.01009 -0.048411 0.005732 0.254906 0.36117

14 -7.29e-02 0.005568 0.05424 0.040526 0.139683 -0.009276 0.20968

15 2.75e-01 -0.221204 -0.25724 0.142195 -0.022591 -0.385100 -0.67871

16 -1.43e-01 -0.208289 0.06663 0.071114 0.300602 0.314939 -0.52394

17 -5.82e-02 -0.055403 0.02986 0.116725 0.202959 -0.021774 0.30836

18 3.85e-02 0.035255 -0.07538 -0.061166 0.089993 0.101438 -0.15985

19 -7.45e-03 -0.003027 0.01978 -0.017913 -0.002396 -0.093194 -0.14249

20 8.58e-02 0.252210 0.03934 -0.342025 -0.163205 -0.102348 0.44287

21 -8.79e-03 -0.018937 0.00127 0.034093 0.038229 -0.006160 0.06425

22 1.03e-01 0.134008 -0.16322 -0.156986 0.123303 0.145144 -0.26328

23 2.76e-02 0.057093 -0.02567 -0.065599 0.001184 -0.079703 -0.14418

24 -3.22e-03 0.017345 0.00492 -0.019954 0.000520 -0.024936 -0.04652

25 -1.31e-01 -0.077455 0.08926 0.087987 0.093171 0.077464 -0.17298

26 -3.41e-02 0.012139 0.04637 -0.028528 0.099577 0.005221 0.15683

27 1.76e-02 -0.047078 -0.04093 0.086363 0.002104 -0.119251 -0.18499

28 7.52e-02 0.271462 -0.14611 -0.339679 0.198665 0.248433 -0.44228

29 -4.02e-02 -0.004892 0.05451 -0.014043 -0.001794 0.036194 0.07940

30 9.31e-04 -0.001210 -0.01152 0.007045 0.014675 0.012422 -0.02593

31 4.58e-03 0.004911 -0.00318 -0.008237 -0.000222 0.007944 0.01251

32 7.47e-06 -0.000298 -0.00210 0.001610 0.002779 0.002275 -0.00500

33 -1.09e-02 -0.029851 0.00520 0.042128 -0.031218 -0.007219 -0.06200

34 -1.26e-02 0.005720 0.01076 -0.000214 0.001052 -0.026013 -0.03977

35 -2.09e-01 -0.196836 0.24231 0.173390 0.153489 -0.025833 0.35397

36 1.58e-01 -0.049889 -0.10222 0.034443 -0.126044 -0.121598 0.22205

37 -9.07e-02 -0.379353 0.18952 0.218405 -0.033735 0.304353 0.61014

38 1.44e-02 0.001322 -0.01609 -0.000850 0.008050 0.000157 0.02054

39 4.45e-01 -0.189819 -0.31345 0.109261 -0.273511 -0.255666 0.58408

40 -4.96e-02 0.011576 0.05092 -0.005414 -0.051171 0.000258 -0.09721

41 -8.69e-03 -0.063137 0.02014 0.044268 0.063116 -0.005637 0.11560

42 6.14e-02 -0.022659 -0.05101 -0.006525 -0.005162 0.062451 0.11358

43 1.05e-02 -0.010791 -0.00820 0.004902 0.016757 -0.000348 0.03122

44 1.27e-01 -0.090858 0.05891 -0.085052 -0.220181 -0.167084 0.37544

45 -7.16e-03 -0.005257 0.01339 -0.004732 0.011309 0.001304 0.02616

46 2.95e-01 -0.212843 0.09139 -0.189969 -0.428337 -0.308460 0.77900

47 9.37e-02 0.434369 -0.14265 -0.364064 0.025892 -0.120612 -0.50219

48 3.18e-03 0.046774 -0.04694 0.031389 -0.068742 -0.009544 -0.16247

49 -9.51e-01 -0.112734 0.70160 0.418630 0.373999 0.239438 -0.98871

50 -2.31e-01 0.578956 -0.11423 0.080954 -0.703325 -0.053488 -1.50721

cov.r cook.d hat inf

1 1.384 7.68e-06 0.1705

2 1.395 3.96e-03 0.1884

3 1.165 2.79e-02 0.1424

4 1.113 2.36e-02 0.1140

5 1.259 1.40e-02 0.1471

6 1.207 1.17e-02 0.1163

**7 1.445 2.49e-03 0.2119 \***

8 1.226 4.17e-03 0.0942

9 1.269 7.29e-05 0.0958

10 1.227 6.31e-03 0.1052

11 1.046 2.21e-02 0.0897

12 1.127 1.94e-02 0.1080

13 0.954 2.13e-02 0.0657

14 1.144 7.40e-03 0.0721

15 0.761 7.20e-02 0.1055

16 0.790 4.34e-02 0.0754

17 1.040 1.58e-02 0.0708

18 1.211 4.33e-03 0.0868

19 1.183 3.44e-03 0.0671

20 1.316 3.29e-02 0.2118

21 1.249 7.03e-04 0.0866

22 1.248 1.17e-02 0.1350

23 1.206 3.53e-03 0.0795

24 1.251 3.69e-04 0.0854

25 1.225 5.07e-03 0.0980

26 1.181 4.16e-03 0.0709

27 1.221 5.79e-03 0.0993

28 1.151 3.25e-02 0.1471

29 1.395 1.07e-03 0.1802

30 1.271 1.15e-04 0.0973

31 1.300 2.67e-05 0.1169

32 1.279 4.26e-06 0.1026

33 1.289 6.55e-04 0.1131

34 1.238 2.70e-04 0.0750

35 1.197 2.10e-02 0.1389

36 1.196 8.32e-03 0.0985

37 0.850 5.92e-02 0.1079

38 1.408 7.19e-05 0.1851

39 1.060 5.58e-02 0.1570

40 1.257 1.61e-03 0.0975

41 1.228 2.27e-03 0.0842

42 1.278 2.20e-03 0.1140

43 1.256 1.66e-04 0.0870

44 1.093 2.34e-02 0.1069

45 1.312 1.17e-04 0.1259

46 0.758 9.44e-02 0.1277

**47 1.419 4.23e-02 0.2654** \*

48 1.272 4.48e-03 0.1211

49 1.051 1.56e-01 0.2559

**50 0.128 2.64e-01 0.1015 \***

> influenceIndexPlot(model.stf,id.n=3)

There were 32 warnings (use warnings() to see them)

> influencePlot(model.stf,id.n=3)

StudRes Hat CookD

46 2.0357210 0.1277290 0.09439478

47 -0.8354542 0.2654200 0.04232333

49 -1.6860294 0.2558868 0.15637613

50 -4.4845939 0.1014896 0.26395944

> stfin<-stfinale[-c(50,49),]

>

>

> model.stf3 <- lm(Profit~., data=stfin)

> summary(model.stf3)

Call:

lm(formula = Profit ~ ., data = stfin)

Residuals:

Min 1Q Median 3Q Max

-16207 -4834 -1653 5728 14153

Coefficients: (1 not defined because of singularities)

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

(Intercept) 5.872e+04 6.364e+03 9.228 1.17e-11 \*\*\*

R.D.Spend 7.888e-01 3.728e-02 21.160 < 2e-16 \*\*\*

Administration -6.214e-02 4.499e-02 -1.381 0.174

Marketing.Spend 1.791e-02 1.405e-02 1.275 0.209

State.California 6.658e+02 2.679e+03 0.249 0.805

State.Florida -4.844e+02 2.682e+03 -0.181 0.858

State.New.York NA NA NA NA

---

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

Residual standard error: 7506 on 42 degrees of freedom

Multiple **R-squared: 0.9628**, Adjusted R-squared: 0.9584

F-statistic: 217.6 on 5 and 42 DF, p-value: < 2.2e-16

> sqrt(mean(model.stf3$residuals\*\*2))

**[1] 7021.34**

There were 21 warnings (use warnings() to see them)

>

> model.stf3\_1 <- lm (Profit ~ sqrt(R.D.Spend)+ sqrt(Administration)+sqrt(Marketing.Spend) + sqrt(State.California)

+ +sqrt(State.Florida)+ sqrt(State.New.York),data=stfin)

> summary(model.stf3\_1)

Call:

lm(formula = Profit ~ sqrt(R.D.Spend) + sqrt(Administration) +

sqrt(Marketing.Spend) + sqrt(State.California) + sqrt(State.Florida) +

sqrt(State.New.York), data = stfin)

Residuals:

Min 1Q Median 3Q Max

-17243 -9642 -4655 6110 34201

Coefficients: (1 not defined because of singularities)

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

(Intercept) 21626.841 19703.033 1.098 0.279

sqrt(R.D.Spend) 341.958 28.024 12.202 2.14e-15 \*\*\*

sqrt(Administration) -7.245 52.648 -0.138 0.891

sqrt(Marketing.Spend) 19.166 16.926 1.132 0.264

sqrt(State.California) -747.978 4659.701 -0.161 0.873

sqrt(State.Florida) -2136.959 4729.651 -0.452 0.654

sqrt(State.New.York) NA NA NA NA

---

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

Residual standard error: 13080 on 42 degrees of freedom

Multiple **R-squared: 0.8871**, Adjusted R-squared: 0.8737

F-statistic: 66.02 on 5 and 42 DF, p-value: < 2.2e-16

> sqrt(mean(model.stf3\_1$residuals\*\*2))

[1] **12236.83**

>

> model.stf3\_2 <- lm (Profit ~ (R.D.Spend)\*\*2 + (Administration)\*\*2 + (Marketing.Spend)\*\*2 + (State.California)\*\*2

+ +(State.Florida)\*\*2 + sqrt(State.New.York)\*\*2,data=stfin)

> summary(model.stf3\_2)

Call:

lm(formula = Profit ~ (R.D.Spend)^2 + (Administration)^2 + (Marketing.Spend)^2 +

(State.California)^2 + (State.Florida)^2 + sqrt(State.New.York)^2,

data = stfin)

Residuals:

Min 1Q Median 3Q Max

-16207 -4834 -1653 5728 14153

Coefficients: (1 not defined because of singularities)

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

(Intercept) 5.872e+04 6.364e+03 9.228 1.17e-11 \*\*\*

R.D.Spend 7.888e-01 3.728e-02 21.160 < 2e-16 \*\*\*

Administration -6.214e-02 4.499e-02 -1.381 0.174

Marketing.Spend 1.791e-02 1.405e-02 1.275 0.209

State.California 6.658e+02 2.679e+03 0.249 0.805

State.Florida -4.844e+02 2.682e+03 -0.181 0.858

sqrt(State.New.York) NA NA NA NA

---

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

Residual standard error: 7506 on 42 degrees of freedom

Multiple **R-squared: 0.9628**, Adjusted R-squared: 0.9584

F-statistic: 217.6 on 5 and 42 DF, p-value: < 2.2e-16

> sqrt(mean(model.stf3\_2$residuals\*\*2))

[1] **7021.34**

>

>

> #model.stf33 <- lm (Profit ~ log(R.D.Spend)+ log(Administration)+log(Marketing.Spend) + log(State.California)+log(State.Florida)+ (State.New.York),data=stfin)

> #there are zeroes present in the data thus log can't be applied

>

> model.stf3\_3 <- lm (log(Profit) ~ .,data=stfin)

> summary(model.stf3\_3)

Call:

lm(formula = log(Profit) ~ ., data = stfin)

Residuals:

Min 1Q Median 3Q Max

-0.35975 -0.03682 0.00562 0.05778 0.16754

Coefficients: (1 not defined because of singularities)

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

(Intercept) 1.113e+01 8.316e-02 133.882 <2e-16 \*\*\*

R.D.Spend 7.438e-06 4.871e-07 15.271 <2e-16 \*\*\*

Administration -7.486e-07 5.879e-07 -1.274 0.210

Marketing.Spend -1.100e-08 1.836e-07 -0.060 0.952

State.California -1.340e-02 3.500e-02 -0.383 0.704

State.Florida -1.207e-03 3.504e-02 -0.034 0.973

State.New.York NA NA NA NA

---

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

Residual standard error: 0.09808 on 42 degrees of freedom

Multiple **R-squared: 0.9252**, Adjusted R-squared: 0.9163

F-statistic: 103.9 on 5 and 42 DF, p-value: < 2.2e-16

> model.stf3\_3\_pred <- (exp(model.stf3\_3$fitted.values))

> model.stf3\_3\_err<- stfin$Profit - model.stf3\_3\_pred

> model.stf3\_3\_rmse <- sqrt(mean(model.stf3\_3\_err^2))

> model.stf3\_3\_rmse

[1] **8966.767**

**Following are the plots acquired from the R studios->**

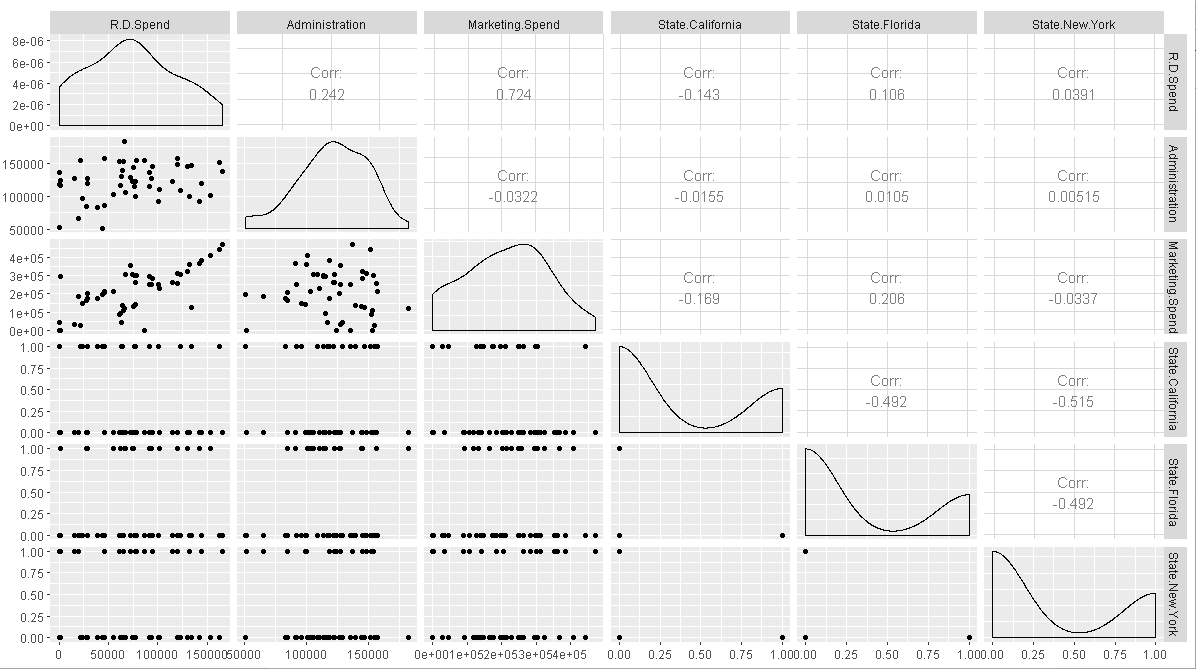


Figure 1. GGplot

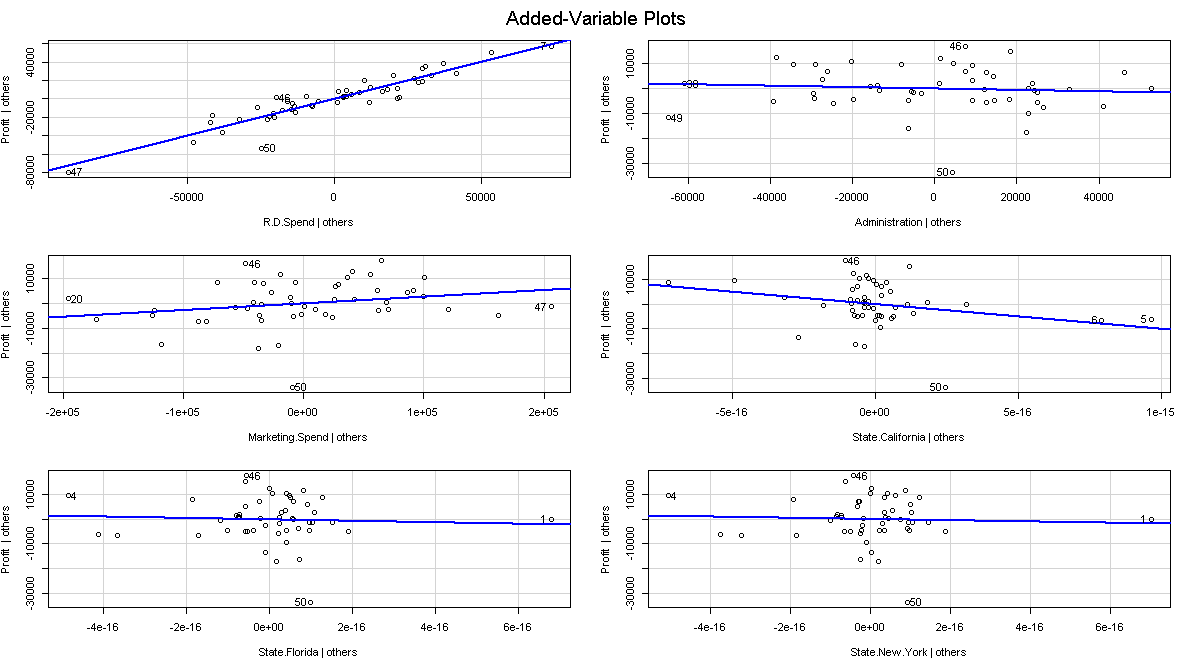


Figure 2. AV values for all variables

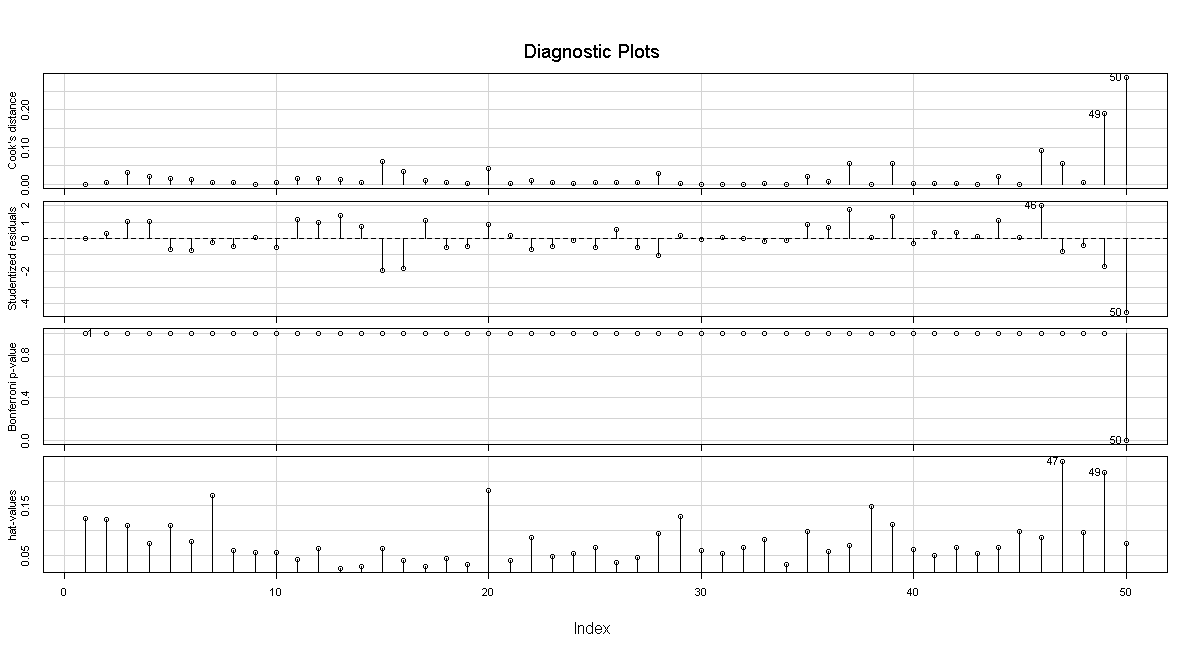


Figure 3. Influence measurement model 1

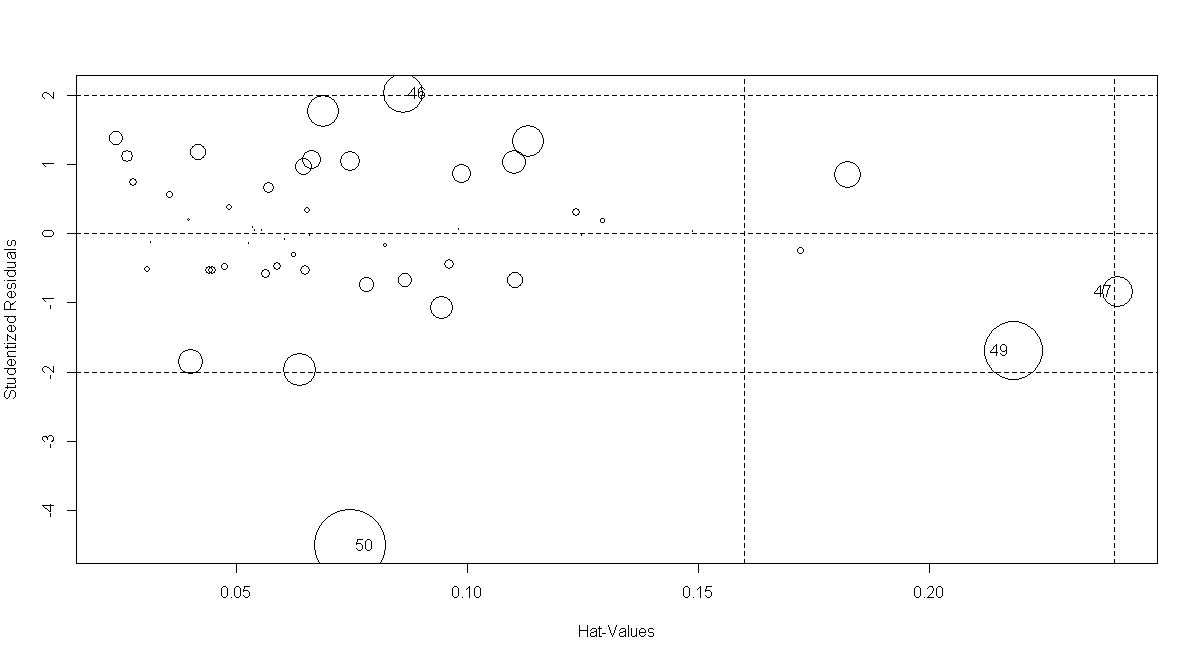


Figure 4.Influence Measurement model 2

1. Predict Price of the computer

A dataframe containing :

price : price in US dollars of 486 PCs

speed : clock speed in MHz

hd : size of hard drive in MB

ram : size of Ram in in MB

screen : size of screen in inches

cd : is a CD-ROM present ?

multi : is a multimedia kit (speakers, sound card) included ?

premium : is the manufacturer was a "premium" firm (IBM, COMPAQ) ?

ads : number of 486 price listings for each month

trend : time trend indicating month starting from January of 1993 to November of 1995.

Answer - >

**Code ->**

> # MultiLinear Reg prob 2

> pc<- read.csv(file.choose())

> ppc<-pc

> ppc$cd<-as.numeric(factor(as.numeric(pc$cd)-1))

> ppc$premium<-as.numeric(factor(as.numeric(pc$premium)-1))

> ppc$multi<-as.numeric(factor(as.numeric(pc$multi)-1))

>

> ppc <-ppc[,-1]

>

>

> #install.packages("GGally")

> library("GGally")

> #install.packages("stringi")

> cor(ppc)

price speed hd ram screen

price 1.00000000 0.30097646 0.43025779 0.62274824 0.296041474

speed 0.30097646 1.00000000 0.37230410 0.23476050 0.189074122

hd 0.43025779 0.37230410 1.00000000 0.77772630 0.232801530

ram 0.62274824 0.23476050 0.77772630 1.00000000 0.208953740

screen 0.29604147 0.18907412 0.23280153 0.20895374 1.000000000

cd 0.19734334 0.25825980 0.50357041 0.43850441 0.129487662

multi -0.01665139 0.08417193 0.09280483 0.04549689 -0.001740414

premium -0.08069636 0.11420791 0.19692359 0.19714459 0.018745223

ads 0.05454047 -0.21523206 -0.32322200 -0.18166971 -0.093919429

trend -0.19998694 0.40543833 0.57779013 0.27684384 0.188614445

cd multi premium ads trend

price 0.19734334 -0.016651388 -0.08069636 0.05454047 -0.19998694

speed 0.25825980 0.084171934 0.11420791 -0.21523206 0.40543833

hd 0.50357041 0.092804830 0.19692359 -0.32322200 0.57779013

ram 0.43850441 0.045496894 0.19714459 -0.18166971 0.27684384

screen 0.12948766 -0.001740414 0.01874522 -0.09391943 0.18861444

cd 1.00000000 0.432179298 0.21607660 -0.06109108 0.44578018

multi 0.43217930 1.000000000 0.12477474 -0.03039426 0.21090743

premium 0.21607660 0.124774741 1.00000000 -0.15202274 0.04210738

ads -0.06109108 -0.030394260 -0.15202274 1.00000000 -0.31855251

trend 0.44578018 0.210907431 0.04210738 -0.31855251 1.00000000

> ggpairs(ppc)

>

> #install.packages("corpcor")

> library('corpcor')

> cor2pcor(cor(ppc))

[,1] [,2] [,3] [,4] [,5]

[1,] 1.0000000 0.537326713 0.33716508 0.49690857 0.36279673

[2,] 0.5373267 1.000000000 -0.10433213 -0.28438702 -0.11491585

[3,] 0.3371651 -0.104332126 1.00000000 0.43156783 -0.10045528

[4,] 0.4969086 -0.284387018 0.43156783 1.00000000 -0.11960849

[5,] 0.3627967 -0.114915854 -0.10045528 -0.11960849 1.00000000

[6,] 0.0807190 -0.008311233 0.08722317 0.10493308 -0.02462381

[7,] 0.1148699 -0.073810359 -0.12420636 -0.10279078 -0.07426269

[8,] -0.4626969 0.292808267 0.20637502 0.25151338 0.14359050

[9,] 0.1599485 -0.140381854 -0.20186341 -0.04815123 -0.06683674

[10,] -0.7219155 0.504953018 0.55265421 0.15896659 0.30166432

[,6] [,7] [,8] [,9] [,10]

[1,] 0.080719005 0.11486991 -0.46269692 0.159948478 -0.721915473

[2,] -0.008311233 -0.07381036 0.29280827 -0.140381854 0.504953018

[3,] 0.087223167 -0.12420636 0.20637502 -0.201863412 0.552654205

[4,] 0.104933080 -0.10279078 0.25151338 -0.048151226 0.158966590

[5,] -0.024623811 -0.07426269 0.14359050 -0.066836740 0.301664320

[6,] 1.000000000 0.40003932 0.15563350 0.164413496 0.214237661

[7,] 0.400039320 1.00000000 0.10912193 -0.049839360 0.141574261

[8,] 0.155633504 0.10912193 1.00000000 -0.046911259 -0.423282587

[9,] 0.164413496 -0.04983936 -0.04691126 1.000000000 -0.001856949

[10,] 0.214237661 0.14157426 -0.42328259 -0.001856949 1.000000000

>

> #No strong corelation

> model.ppc <- lm(price~.,data=ppc)

> summary(model.ppc)

Call:

lm(formula = price ~ ., data = ppc)

Residuals:

Min 1Q Median 3Q Max

-1093.77 -174.24 -11.49 146.49 2001.05

Coefficients:

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

(Intercept) 651.97219 64.49224 10.109 < 2e-16 \*\*\*

speed 9.32028 0.18506 50.364 < 2e-16 \*\*\*

hd 0.78178 0.02761 28.311 < 2e-16 \*\*\*

ram 48.25596 1.06608 45.265 < 2e-16 \*\*\*

screen 123.08904 3.99950 30.776 < 2e-16 \*\*\*

cd 60.91671 9.51559 6.402 1.65e-10 \*\*\*

multi 104.32382 11.41268 9.141 < 2e-16 \*\*\*

premium -509.22473 12.34225 -41.259 < 2e-16 \*\*\*

ads 0.65729 0.05132 12.809 < 2e-16 \*\*\*

trend -51.84958 0.62871 -82.470 < 2e-16 \*\*\*

---

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

Residual standard error: 275.3 on 6249 degrees of freedom

Multiple R-squared: 0.7756, Adjusted R-squared: 0.7752

F-statistic: 2399 on 9 and 6249 DF, p-value: < 2.2e-16

>

>

> #here we can see, which variable to eliminate if its convinient

> library(car)

> vif(model.ppc)

speed hd ram screen cd multi premium ads

1.265364 4.207395 2.974628 1.081644 1.859370 1.290568 1.109388 1.217218

trend

2.022790

> # no model has high value so all are contributing

> # gives degree of collinearity between the variables

> #coolinearity values are low thus no need to delete any variable

> avPlots(model.ppc)

> # gives the contribution rate -> decide by th value of slope

>

> influence.measures(model.ppc)

Influence measures of

lm(formula = price ~ ., data = ppc) :

dfb.1\_ dfb.sped dfb.hd dfb.ram dfb.scrn dfb.cd dfb.mult

1 -2.44e-02 1.66e-02 1.15e-02 -0.000344 5.74e-03 -1.03e-02 3.44e-03

2 -9.45e-03 2.74e-03 2.88e-04 0.005890 2.27e-03 -4.86e-03 1.59e-03

3 -6.99e-04 2.51e-02 -5.91e-03 0.016960 -2.54e-02 -9.59e-03 1.08e-03

4 -1.05e-01 2.28e-02 1.95e-02 -0.035509 1.97e-02 -2.66e-02 -9.24e-03

5 2.44e-02 -7.32e-03 -2.21e-02 0.042043 -1.17e-02 -8.76e-03 4.73e-03

6 3.00e-02 3.23e-02 -3.10e-02 0.052347 -1.85e-02 -1.24e-02 6.37e-03

7 -2.28e-02 1.74e-02 1.93e-03 0.019249 5.39e-03 -6.12e-02 2.40e-02

8 -7.58e-03 -4.71e-03 8.98e-04 0.004546 2.52e-03 -3.60e-03 1.17e-03

9 -1.66e-02 -8.39e-03 7.04e-03 -0.007651 7.32e-03 -6.39e-04 -5.18e-04

10 1.83e-04 1.86e-03 1.76e-03 -0.003545 3.33e-03 1.08e-03 4.61e-05

11 -1.14e-03 5.61e-03 9.65e-03 -0.008433 -9.97e-03 -1.50e-03 -4.39e-04

12 -4.76e-03 -6.01e-03 2.36e-03 -0.002271 2.46e-03 -3.20e-05 -2.00e-04

13 -1.18e-02 -6.80e-03 2.39e-03 0.002790 4.38e-03 -3.71e-03 1.00e-03

14 -2.45e-03 1.73e-03 1.47e-04 -0.000611 7.96e-04 -1.41e-04 -8.98e-05

15 7.68e-03 3.87e-03 -3.14e-03 0.003453 -3.38e-03 2.82e-04 2.49e-04

16 -3.22e-03 -1.86e-03 6.52e-04 0.000762 1.20e-03 -1.01e-03 2.74e-04

17 -1.86e-02 5.37e-03 5.65e-04 0.011569 4.46e-03 -9.56e-03 3.12e-03

18 8.93e-06 -1.16e-04 1.16e-04 -0.000196 2.01e-04 6.70e-05 -6.39e-07

19 -5.69e-04 6.08e-03 -9.44e-03 0.012377 -9.81e-03 -2.85e-03 -3.10e-04

20 8.27e-02 8.56e-02 9.49e-02 -0.060737 -4.36e-02 -1.58e-02 1.55e-02

21 -2.08e-02 6.20e-03 2.47e-03 0.005382 5.91e-03 -7.34e-03 2.04e-03

22 1.69e-02 2.35e-02 -4.75e-03 -0.003653 -7.66e-03 4.72e-03 -1.23e-03

23 5.38e-03 -3.42e-03 -4.71e-03 0.009234 -2.40e-03 -1.86e-03 1.02e-03

24 -1.38e-02 1.28e-03 2.35e-03 -0.006720 2.11e-02 1.60e-03 6.33e-04

25 7.96e-02 2.55e-02 9.72e-02 -0.060024 -3.63e-02 -1.28e-02 1.41e-02

26 -3.22e-02 9.22e-03 2.28e-03 0.019176 7.72e-03 -1.68e-02 5.54e-03

27 -7.55e-03 2.25e-03 8.96e-04 0.001950 2.14e-03 -2.66e-03 7.40e-04

28 7.71e-02 -3.43e-02 1.00e-01 -0.059708 -2.93e-02 -1.00e-02 1.29e-02

29 1.27e-02 -3.81e-03 -1.15e-02 0.021904 -6.11e-03 -4.57e-03 2.46e-03

30 2.70e-02 3.32e-02 -6.65e-03 0.007917 -1.40e-02 -6.29e-04 1.73e-03

31 -1.62e-02 6.37e-03 1.95e-03 0.013512 4.43e-03 -4.32e-02 1.70e-02

32 -3.06e-03 2.18e-03 2.41e-04 0.000822 7.41e-04 -1.13e-03 3.19e-04

33 1.46e-02 5.21e-03 3.57e-03 -0.001772 -2.30e-02 -4.79e-04 -1.08e-03

34 -6.15e-06 -6.25e-05 -5.92e-05 0.000119 -1.12e-04 -3.62e-05 -1.55e-06

35 -7.98e-03 2.58e-03 -1.49e-03 0.003821 2.30e-03 -2.49e-03 5.52e-04

36 9.82e-05 2.78e-03 6.88e-04 -0.001701 1.45e-03 4.54e-04 4.96e-05

37 -1.92e-02 -2.42e-02 9.52e-03 -0.009163 9.92e-03 -1.29e-04 -8.08e-04

38 4.82e-02 4.78e-02 7.00e-02 -0.046221 -2.55e-02 -1.10e-02 1.03e-02

39 -7.36e-03 -3.72e-03 3.12e-03 -0.003389 3.24e-03 -2.83e-04 -2.30e-04

40 -6.33e-03 -3.20e-03 2.68e-03 -0.002916 2.79e-03 -2.44e-04 -1.98e-04

41 1.66e-02 7.90e-03 -2.88e-03 0.004571 -7.34e-03 1.34e-04 8.85e-04

42 7.20e-03 4.20e-02 6.70e-02 -0.048359 2.00e-02 -1.09e-02 1.16e-02

43 5.52e-03 -1.69e-03 -1.83e-03 0.002385 -1.98e-03 3.90e-04 1.15e-04

44 -6.90e-03 2.10e-03 2.39e-03 -0.003056 2.48e-03 -5.00e-04 -1.35e-04

45 2.69e-02 1.11e-02 -2.62e-02 0.046684 -1.48e-02 -1.04e-02 5.48e-03

46 1.44e-02 -4.68e-03 -1.39e-03 0.003702 -5.21e-03 5.99e-04 6.03e-04

47 2.89e-03 -8.58e-04 -3.42e-04 -0.000745 -8.19e-04 1.02e-03 -2.83e-04

48 -1.43e-02 1.01e-02 2.24e-03 0.003041 3.44e-03 -5.44e-03 1.59e-03

49 2.06e-03 1.04e-03 8.57e-04 -0.001404 -7.79e-04 4.79e-04 -5.57e-05

50 -1.14e-02 4.27e-03 -8.89e-03 0.010336 3.39e-03 -2.65e-03 1.51e-04

51 1.60e-02 -5.18e-03 2.99e-03 -0.007658 -4.61e-03 4.98e-03 -1.11e-03

52 1.24e-02 -3.68e-03 -1.47e-03 -0.003196 -3.51e-03 4.36e-03 -1.21e-03

53 4.71e-04 4.02e-03 1.50e-03 -0.001032 1.83e-03 -4.16e-04 6.10e-04

54 -3.25e-04 1.14e-04 -1.58e-04 0.000226 9.52e-05 -8.83e-05 1.33e-05

55 -7.87e-02 4.24e-03 5.05e-03 -0.018796 1.63e-02 -1.79e-02 -8.04e-03

56 -7.99e-03 -4.60e-03 1.62e-03 0.001888 2.97e-03 -2.51e-03 6.79e-04

57 -1.90e-02 -2.64e-02 5.33e-03 0.004103 8.60e-03 -5.30e-03 1.38e-03

58 -1.85e-02 9.61e-03 2.38e-03 -0.008805 2.75e-02 1.82e-03 9.61e-04

59 1.08e-02 -5.96e-03 3.59e-03 -0.001547 -1.61e-02 5.47e-05 -9.44e-04

60 -1.66e-02 -9.55e-03 3.35e-03 0.003917 6.15e-03 -5.21e-03 1.41e-03

61 6.82e-03 5.98e-02 1.82e-02 -0.012356 2.71e-02 -5.67e-03 8.65e-03

62 -5.98e-04 1.03e-02 -3.78e-03 0.013290 -1.93e-02 -7.24e-03 7.19e-04

63 -2.44e-04 1.10e-03 2.20e-04 -0.000394 -1.67e-03 -8.18e-05 -2.06e-04

64 -3.71e-03 1.63e-03 -1.57e-03 0.004516 1.05e-03 -9.46e-03 3.63e-03

65 -5.28e-03 -7.77e-03 1.06e-03 0.003046 2.20e-03 -2.31e-03 7.46e-04

66 -2.85e-02 2.09e-02 -6.47e-03 0.013961 7.05e-03 -9.39e-03 2.15e-03

dfb.prmm dfb.ads dfb.trnd dffit cov.r cook.d hat inf

1 -3.76e-03 0.061310 0.035506 -0.082350 0.998 6.78e-04 0.00188

2 -1.40e-03 0.023846 0.017060 -0.032659 1.003 1.07e-04 0.00187

3 -4.19e-03 0.069231 0.053820 -0.100041 0.995 1.00e-03 0.00197

4 1.51e-01 0.129619 0.083924 -0.212779 0.987 4.52e-03 0.00381 \*

5 -1.85e-03 -0.044858 -0.025406 0.077020 1.000 5.93e-04 0.00257

6 -4.70e-03 -0.052113 -0.039305 0.099078 0.999 9.81e-04 0.00284

7 5.02e-03 0.058373 0.048477 -0.091060 1.001 8.29e-04 0.00335

8 -7.08e-04 0.018619 0.015027 -0.026449 1.003 7.00e-05 0.00192

9 2.77e-04 0.034918 0.027179 -0.049914 1.001 2.49e-04 0.00180

10 2.20e-04 -0.010205 -0.010266 0.015361 1.003 2.36e-05 0.00190

11 -1.16e-03 0.032262 0.020659 -0.045346 1.002 2.06e-04 0.00189

12 2.93e-04 0.009738 0.008488 -0.015287 1.003 2.34e-05 0.00207

13 -6.16e-04 0.027393 0.022033 -0.038417 1.002 1.48e-04 0.00182

14 -9.43e-05 0.005229 0.003703 -0.007390 1.003 5.46e-06 0.00183

15 -1.35e-04 -0.016107 -0.012603 0.023035 1.003 5.31e-05 0.00180

16 -1.68e-04 0.007480 0.006016 -0.010490 1.003 1.10e-05 0.00182

17 -2.74e-03 0.046838 0.033509 -0.064149 1.000 4.11e-04 0.00187

18 2.46e-05 -0.000562 -0.000508 0.000827 1.004 6.85e-08 0.00191

19 -9.87e-04 0.027114 0.026812 -0.041452 1.002 1.72e-04 0.00198

20 -1.29e-02 -0.138310 -0.202502 0.265056 0.957 6.99e-03 0.00232 \*

21 -2.17e-03 0.049782 0.035420 -0.067353 0.999 4.54e-04 0.00178

22 6.13e-05 -0.038189 -0.034218 0.058868 1.001 3.47e-04 0.00210

23 -3.02e-04 -0.010016 -0.005244 0.017259 1.004 2.98e-05 0.00265

24 8.12e-04 -0.015752 -0.016981 0.031506 1.004 9.93e-05 0.00330

25 -9.04e-03 -0.137333 -0.184132 0.243651 0.960 5.91e-03 0.00210 \*

26 -4.85e-03 0.081682 0.057677 -0.111608 0.993 1.24e-03 0.00187 \*

27 -7.87e-04 0.018040 0.012835 -0.024407 1.003 5.96e-05 0.00178

28 -5.29e-03 -0.137262 -0.167030 0.237864 0.962 5.63e-03 0.00212 \*

29 -9.64e-04 -0.023370 -0.013236 0.040126 1.003 1.61e-04 0.00257

30 -2.05e-03 -0.053748 -0.050919 0.085255 0.998 7.27e-04 0.00203

31 3.92e-03 0.041015 0.035592 -0.063718 1.003 4.06e-04 0.00326

32 -3.94e-04 0.007405 0.004957 -0.010102 1.003 1.02e-05 0.00185

33 -1.03e-03 0.017940 0.014047 -0.033943 1.004 1.15e-04 0.00323

34 -7.39e-06 0.000342 0.000344 -0.000515 1.004 2.66e-08 0.00190

35 -6.76e-04 0.018477 0.014575 -0.025548 1.003 6.53e-05 0.00178

36 -1.68e-06 -0.004903 -0.005428 0.008006 1.004 6.41e-06 0.00213

37 1.18e-03 0.039285 0.034242 -0.061666 1.001 3.80e-04 0.00207

38 -8.33e-03 -0.077312 -0.124006 0.158932 0.988 2.52e-03 0.00251 \*

39 1.23e-04 0.015468 0.012040 -0.022112 1.003 4.89e-05 0.00180

40 1.06e-04 0.013308 0.010358 -0.019023 1.003 3.62e-05 0.00180

41 -5.20e-04 -0.033865 -0.028801 0.048929 1.001 2.39e-04 0.00177

42 -6.79e-03 -0.074828 -0.123477 0.154007 0.989 2.37e-03 0.00249 \*

43 1.67e-04 -0.011915 -0.008196 0.016472 1.003 2.71e-05 0.00177

44 -2.15e-04 0.014922 0.010206 -0.020618 1.003 4.25e-05 0.00177

45 -3.16e-03 -0.048076 -0.031767 0.085028 1.000 7.23e-04 0.00259

46 2.30e-04 -0.030213 -0.022737 0.042263 1.002 1.79e-04 0.00175

47 3.01e-04 -0.006894 -0.004905 0.009327 1.003 8.70e-06 0.00178

48 -1.91e-03 0.034885 0.022708 -0.047388 1.002 2.25e-04 0.00185

49 2.88e-05 -0.004476 -0.004362 0.006602 1.003 4.36e-06 0.00184

50 -5.36e-04 0.024785 0.023633 -0.036796 1.002 1.35e-04 0.00188

51 1.35e-03 -0.037033 -0.029211 0.051204 1.001 2.62e-04 0.00178

52 1.29e-03 -0.029563 -0.021034 0.039996 1.002 1.60e-04 0.00178

53 -3.27e-04 -0.007111 -0.008592 0.012014 1.004 1.44e-05 0.00205

54 -2.14e-05 0.000730 0.000634 -0.001041 1.003 1.08e-07 0.00182

55 1.14e-01 0.093649 0.070202 -0.156427 0.995 2.44e-03 0.00374 \*

56 -4.17e-04 0.018542 0.014914 -0.026004 1.003 6.76e-05 0.00182

57 -6.89e-05 0.042891 0.038431 -0.066115 1.000 4.37e-04 0.00210

58 6.20e-04 -0.020562 -0.024306 0.043206 1.004 1.87e-04 0.00346

59 -1.81e-04 0.012658 0.012384 -0.025570 1.005 6.54e-05 0.00333

60 -8.64e-04 0.038458 0.030934 -0.053935 1.001 2.91e-04 0.00182

61 -4.59e-03 -0.105579 -0.124948 0.176873 0.979 3.12e-03 0.00204 \*

62 -2.77e-03 0.054556 0.044840 -0.077892 0.998 6.07e-04 0.00188

63 -1.12e-04 0.005299 0.004219 -0.007593 1.003 5.77e-06 0.00181

64 1.00e-03 0.008903 0.008960 -0.014409 1.005 2.08e-05 0.00329 \*

65 -2.25e-04 0.012619 0.011325 -0.019652 1.004 3.86e-05 0.00220

66 -3.11e-03 0.066940 0.049863 -0.093406 0.996 8.72e-04 0.00185

[ reached 'max' / getOption("max.print") -- omitted 6193 rows ]

> influenceIndexPlot(model.ppc,id.n=3)

There were 50 or more warnings (use warnings() to see the first 50)

> influencePlot(model.ppc,id.n=3)

StudRes Hat CookD

1441 7.3058529 0.002228075 0.011819949

1701 7.1838002 0.002464463 0.012647347

3784 -0.8667018 0.020972880 0.001609237

4478 -1.3795547 0.020060286 0.003895407

Warning messages:

1: In plot.window(...) : "id.n" is not a graphical parameter

2: In plot.xy(xy, type, ...) : "id.n" is not a graphical parameter

3: In axis(side = side, at = at, labels = labels, ...) :

"id.n" is not a graphical parameter

4: In axis(side = side, at = at, labels = labels, ...) :

"id.n" is not a graphical parameter

5: In box(...) : "id.n" is not a graphical parameter

6: In title(...) : "id.n" is not a graphical parameter

7: In plot.xy(xy.coords(x, y), type = type, ...) :

"id.n" is not a graphical parameter

> ppcfin<-ppc[-c(1441,1701),]

>

>

> model.ppcfin1 <- lm(price~., data=ppcfin)

> summary(model.ppcfin1)

Call:

lm(formula = price ~ ., data = ppcfin)

Residuals:

Min 1Q Median 3Q Max

-1095.65 -172.78 -10.84 146.42 1510.65

Coefficients:

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

(Intercept) 681.7321 64.0272 10.648 < 2e-16 \*\*\*

speed 9.2992 0.1835 50.664 < 2e-16 \*\*\*

hd 0.7749 0.0274 28.286 < 2e-16 \*\*\*

ram 48.5222 1.0576 45.878 < 2e-16 \*\*\*

screen 121.0926 3.9714 30.492 < 2e-16 \*\*\*

cd 60.4964 9.4400 6.409 1.58e-10 \*\*\*

multi 104.7703 11.3195 9.256 < 2e-16 \*\*\*

premium -509.8352 12.2409 -41.650 < 2e-16 \*\*\*

ads 0.6510 0.0509 12.791 < 2e-16 \*\*\*

trend -51.6496 0.6238 -82.793 < 2e-16 \*\*\*

---

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

Residual standard error: 273.1 on 6247 degrees of freedom

Multiple **R-squared: 0.7777**, Adjusted R-squared: 0.7774

F-statistic: 2428 on 9 and 6247 DF, p-value: < 2.2e-16

> sqrt(mean(model.ppcfin1$residuals\*\*2))

[1] **272.8675**

>

> model.ppcfin2 <- lm (price ~ sqrt(speed)+ sqrt(hd)+sqrt(ram) + sqrt(screen)

+ +sqrt(cd)+ sqrt(multi)+sqrt(premium)+sqrt(ads)+sqrt(trend),data=ppcfin)

> summary(model.ppcfin2)

Call:

lm(formula = price ~ sqrt(speed) + sqrt(hd) + sqrt(ram) + sqrt(screen) +

sqrt(cd) + sqrt(multi) + sqrt(premium) + sqrt(ads) + sqrt(trend),

data = ppcfin)

Residuals:

Min 1Q Median 3Q Max

-1169.07 -171.16 -18.74 146.26 1319.34

Coefficients:

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

(Intercept) -1292.508 124.869 -10.351 <2e-16 \*\*\*

sqrt(speed) 128.454 2.631 48.815 <2e-16 \*\*\*

sqrt(hd) 37.038 1.217 30.424 <2e-16 \*\*\*

sqrt(ram) 298.980 6.791 44.027 <2e-16 \*\*\*

sqrt(screen) 854.947 30.386 28.136 <2e-16 \*\*\*

sqrt(cd) 68.742 22.343 3.077 0.0021 \*\*

sqrt(multi) 247.447 26.580 9.310 <2e-16 \*\*\*

sqrt(premium) -1274.358 28.850 -44.172 <2e-16 \*\*\*

sqrt(ads) 40.301 1.294 31.153 <2e-16 \*\*\*

sqrt(trend) -360.097 4.387 -82.090 <2e-16 \*\*\*

---

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

Residual standard error: 265.9 on 6247 degrees of freedom

Multiple **R-squared: 0.7892**, Adjusted R-squared: 0.7889

F-statistic: 2599 on 9 and 6247 DF, p-value: < 2.2e-16

> sqrt(mean(model.ppcfin2$residuals\*\*2))

[1] **265.6933**

>

> model.ppcfin3 <- lm (price ~ (speed)\*\*2+ (hd)\*\*2+(ram)\*\*2 + (screen)\*\*2

+ +(cd)\*\*2+ (multi)\*\*2+(premium)\*\*2+(ads)\*\*2+(trend)\*\*2,data=ppcfin)

There were 14 warnings (use warnings() to see them)

>

> summary(model.ppcfin3)

Call:

lm(formula = price ~ (speed)^2 + (hd)^2 + (ram)^2 + (screen)^2 +

(cd)^2 + (multi)^2 + (premium)^2 + (ads)^2 + (trend)^2, data = ppcfin)

Residuals:

Min 1Q Median 3Q Max

-1095.65 -172.78 -10.84 146.42 1510.65

Coefficients:

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

(Intercept) 681.7321 64.0272 10.648 < 2e-16 \*\*\*

speed 9.2992 0.1835 50.664 < 2e-16 \*\*\*

hd 0.7749 0.0274 28.286 < 2e-16 \*\*\*

ram 48.5222 1.0576 45.878 < 2e-16 \*\*\*

screen 121.0926 3.9714 30.492 < 2e-16 \*\*\*

cd 60.4964 9.4400 6.409 1.58e-10 \*\*\*

multi 104.7703 11.3195 9.256 < 2e-16 \*\*\*

premium -509.8352 12.2409 -41.650 < 2e-16 \*\*\*

ads 0.6510 0.0509 12.791 < 2e-16 \*\*\*

trend -51.6496 0.6238 -82.793 < 2e-16 \*\*\*

---

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

Residual standard error: 273.1 on 6247 degrees of freedom

Multiple **R-squared: 0.7777**, Adjusted R-squared: 0.7774

F-statistic: 2428 on 9 and 6247 DF, p-value: < 2.2e-16

> sqrt(mean(model.ppcfin3$residuals\*\*2))

[1] **272.8675**

>

>

> #there are zeroes present in the data thus log can't be applied

>

> model.ppcfin4 <- lm (log(price) ~ .,data=ppcfin)

> summary(model.ppcfin4)

Call:

lm(formula = log(price) ~ ., data = ppcfin)

Residuals:

Min 1Q Median 3Q Max

-0.52391 -0.07376 0.00267 0.07461 0.49147

Coefficients:

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

(Intercept) 6.973e+00 2.811e-02 248.019 <2e-16 \*\*\*

speed 4.246e-03 8.059e-05 52.682 <2e-16 \*\*\*

hd 3.387e-04 1.203e-05 28.153 <2e-16 \*\*\*

ram 2.092e-02 4.644e-04 45.057 <2e-16 \*\*\*

screen 5.373e-02 1.744e-03 30.811 <2e-16 \*\*\*

cd 4.937e-02 4.145e-03 11.911 <2e-16 \*\*\*

multi 4.749e-02 4.970e-03 9.555 <2e-16 \*\*\*

premium -2.272e-01 5.375e-03 -42.277 <2e-16 \*\*\*

ads 2.696e-04 2.235e-05 12.062 <2e-16 \*\*\*

trend -2.355e-02 2.739e-04 -85.979 <2e-16 \*\*\*

---

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

Residual standard error: 0.1199 on 6247 degrees of freedom

Multiple **R-squared: 0.7837**, Adjusted R-squared: 0.7833

F-statistic: 2514 on 9 and 6247 DF, p-value: < 2.2e-16

> model.ppcfin4\_pred <- (exp(model.ppcfin4$fitted.values))

> model.ppcfin4\_err<- ppcfin$price - model.ppcfin4\_pred

> model.ppcfin4\_rmse <- sqrt(mean(model.ppcfin4\_err^2))

> model.ppcfin4\_rmse

[1] **270.9619**

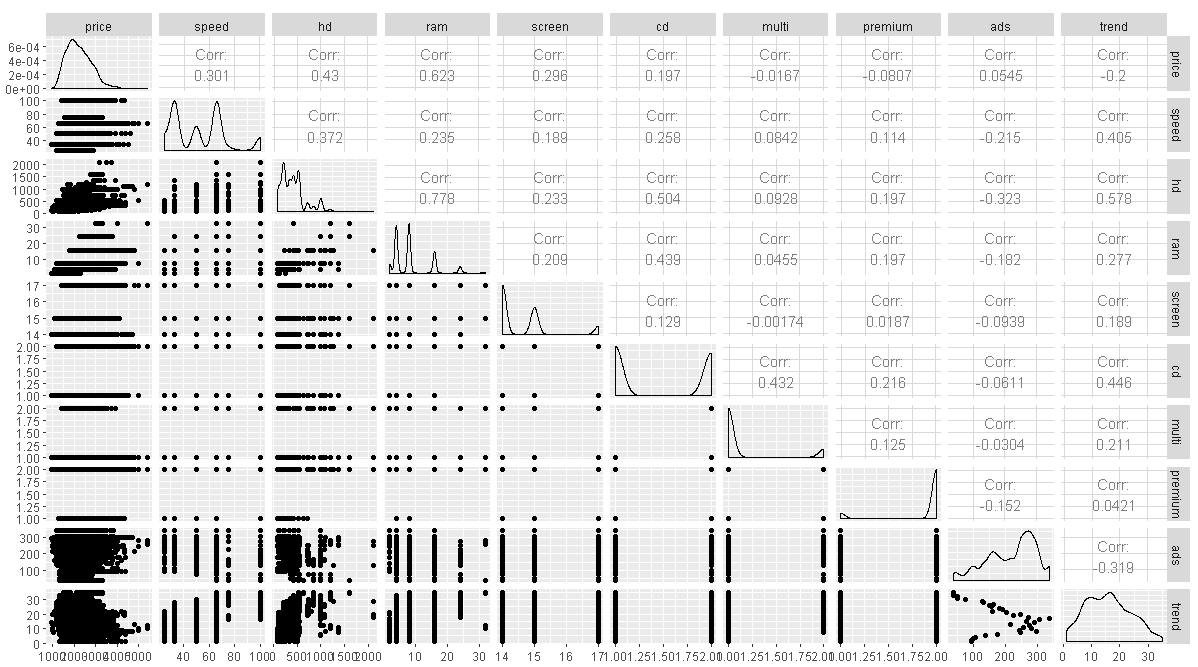


Figure 5. GG plot

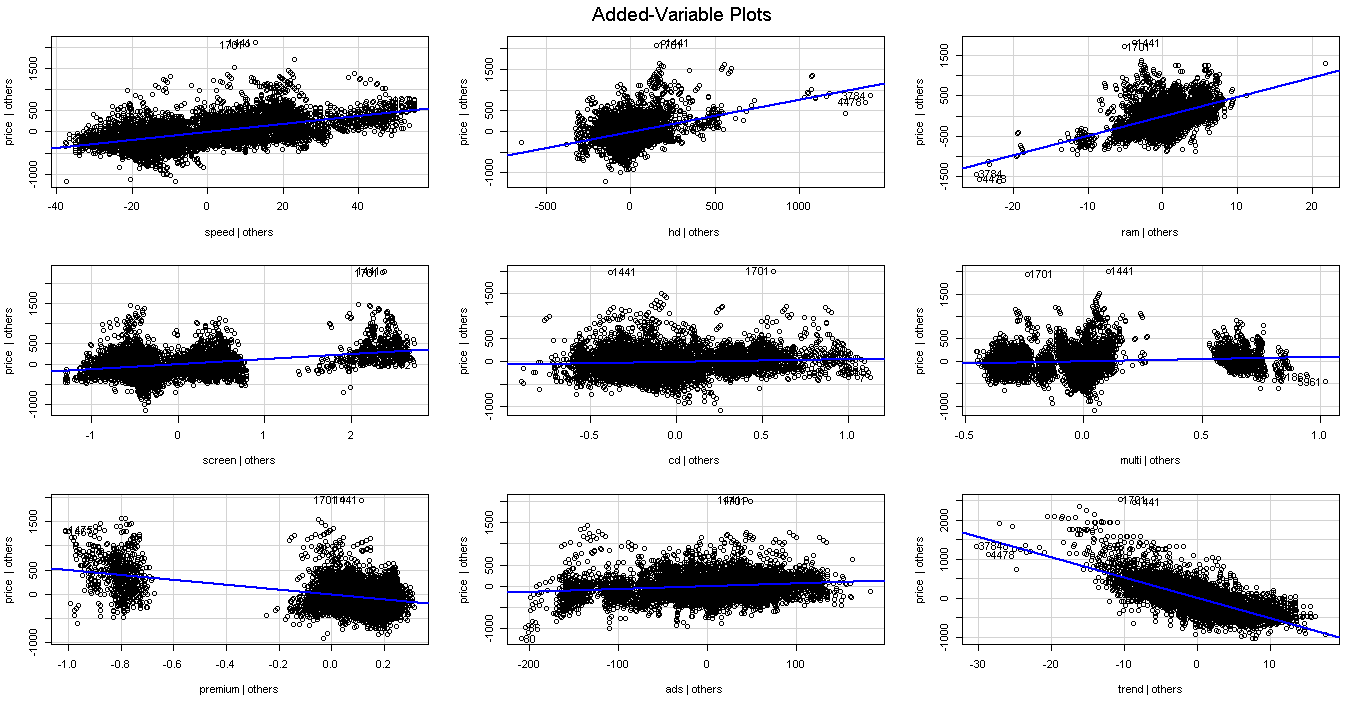


Figure 6. AV plot

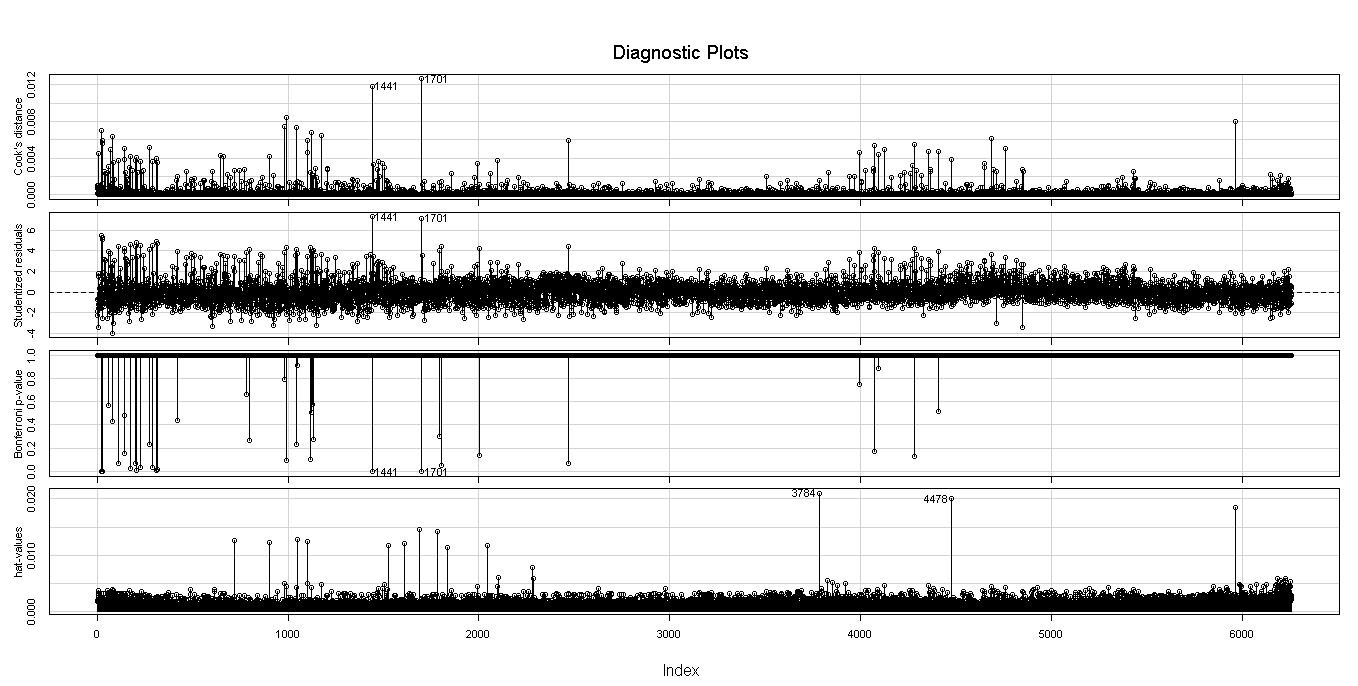


Figure 7. Influence measurement Model 1

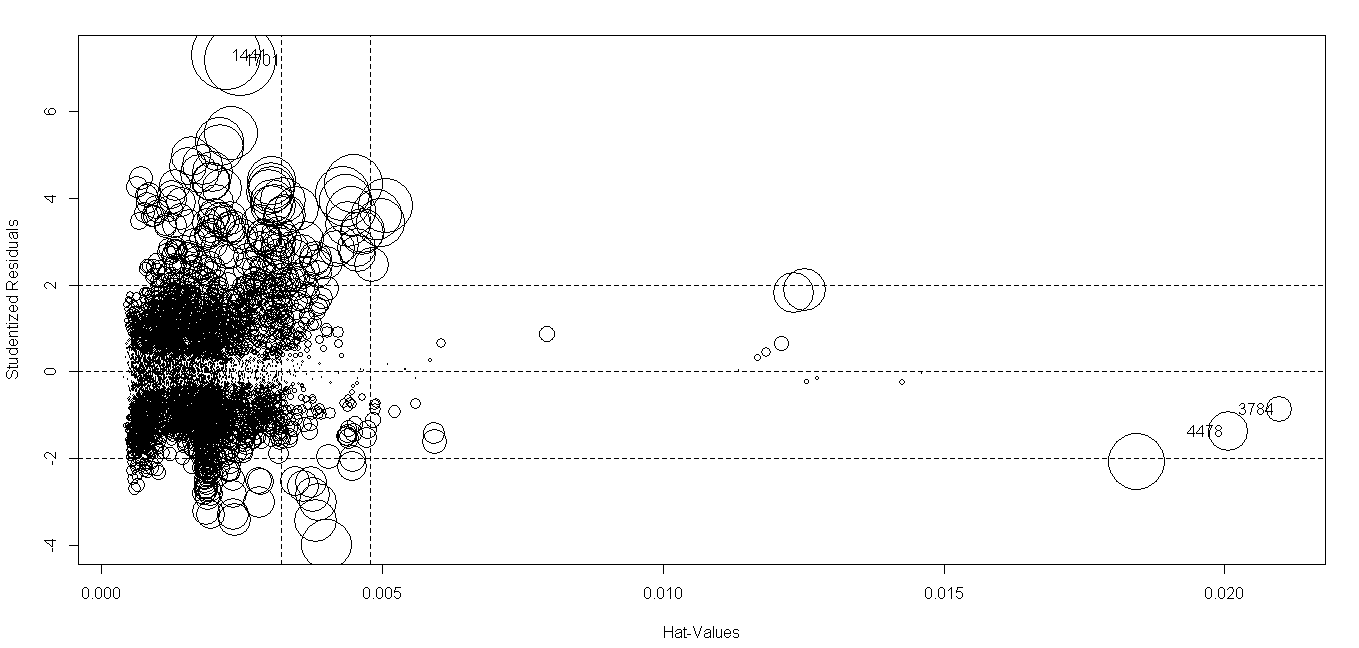


Figure 8. Influence measurement model 2

1. Consider only the below columns and prepare a prediction model for predicting Price.

Corolla<-Corolla[c("Price","Age\_08\_04","KM","HP","cc","Doors","Gears","Quarterly\_Tax","Weight")]

Model -- model of the car

Price -- Offer Price in EUROs

Age\_08\_04 -- Age in months as in August 2004

Mfg\_Month -- Manufacturing month (1-12)

Mfg\_Year -- Manufacturing Year

KM -- Accumulated Kilometers on odometer

Fuel\_Type -- Fuel Type (Petrol, Diesel, CNG)

HP -- Horse Power

Met\_Color -- Metallic Color? (Yes=1, No=0)

Color -- Color (Blue, Red, Grey, Silver, Black, etc.)

Automatic -- Automatic ( (Yes=1, No=0)

cc -- Cylinder Volume in cubic centimeters

Doors -- Number of doors

Cylinders -- Number of cylinders

Gears -- Number of gear positions

Quarterly\_Tax -- Quarterly road tax in EUROs

Weight -- Weight in Kilograms

Mfr\_Guarantee -- Within Manufacturer's Guarantee period (Yes=1, No=0)

BOVAG\_Guarantee -- BOVAG (Dutch dealer network) Guarantee (Yes=1, No=0)

Guarantee\_Period -- Guarantee period in months

ABS -- Anti-Lock Brake System (Yes=1, No=0)

Airbag\_1 -- Driver\_Airbag (Yes=1, No=0)

Airbag\_2 -- Passenger Airbag (Yes=1, No=0)

Airco -- Airconditioning (Yes=1, No=0)

Automatic\_airco -- Automatic Airconditioning (Yes=1, No=0)

Boardcomputer -- Boardcomputer (Yes=1, No=0)

CD\_Player -- CD Player (Yes=1, No=0)

Central\_Lock -- Central Lock (Yes=1, No=0)

Powered\_Windows -- Powered Windows (Yes=1, No=0)

Power\_Steering -- Power Steering (Yes=1, No=0)

Radio -- Radio (Yes=1, No=0)

Mistlamps -- Mistlamps (Yes=1, No=0)

Sport\_Model -- Sport Model (Yes=1, No=0)

Backseat\_Divider -- Backseat Divider (Yes=1, No=0)

Metallic\_Rim --Metallic Rim (Yes=1, No=0)

Radio\_cassette -- Radio Cassette (Yes=1, No=0)

Tow\_Bar -- Tow Bar (Yes=1, No=0)

Answer ->

**Code ->**

> # MultiLinear Reg prob 2

> tc<- read.csv(file.choose())

>

> tcp <- tc[,c(3,4,7,9,13,14,16,17,18)]

> #install.packages("GGally")

> library("GGally")

> #install.packages("stringi")

> cor(tcp)

Price Age\_08\_04 KM HP cc

Price 1.00000000 -0.876590497 -0.56996016 0.31498983 0.12638920

Age\_08\_04 -0.87659050 1.000000000 0.50567218 -0.15662202 -0.09808374

KM -0.56996016 0.505672180 1.00000000 -0.33353795 0.10268289

HP 0.31498983 -0.156622020 -0.33353795 1.00000000 0.03585580

cc 0.12638920 -0.098083739 0.10268289 0.03585580 1.00000000

Doors 0.18532555 -0.148359215 -0.03619661 0.09242450 0.07990330

Gears 0.06310386 -0.005363947 0.01502333 0.20947715 0.01462935

Quarterly\_Tax 0.21919691 -0.198430508 0.27816470 -0.29843172 0.30699580

Weight 0.58119759 -0.470253184 -0.02859846 0.08961406 0.33563740

Doors Gears Quarterly\_Tax Weight

Price 0.18532555 0.063103857 0.219196911 0.58119759

Age\_08\_04 -0.14835921 -0.005363947 -0.198430508 -0.47025318

KM -0.03619661 0.015023328 0.278164697 -0.02859846

HP 0.09242450 0.209477146 -0.298431717 0.08961406

cc 0.07990330 0.014629352 0.306995798 0.33563740

Doors 1.00000000 -0.160141430 0.109363225 0.30261764

Gears -0.16014143 1.000000000 -0.005451955 0.02061328

Quarterly\_Tax 0.10936323 -0.005451955 1.000000000 0.62613373

Weight 0.30261764 0.020613284 0.626133733 1.00000000

> ggpairs(tcp)

>

> #install.packages("corpcor")

> library('corpcor')

> cor2pcor(cor(tcp))

[,1] [,2] [,3] [,4] [,5]

[1,] 1.000000000 -0.776238352 -0.402745405 0.28521314 -0.03556185

[2,] -0.776238352 1.000000000 0.002383081 0.24531845 -0.02014628

[3,] -0.402745405 0.002383081 1.000000000 -0.06039653 0.05108725

[4,] 0.285213137 0.245318454 -0.060396533 1.00000000 0.09871851

[5,] -0.035561846 -0.020146283 0.051087249 0.09871851 1.00000000

[6,] -0.001069746 -0.002800916 0.026724172 0.06817527 -0.01606038

[7,] 0.079586710 0.051074865 0.100506331 0.20769268 -0.01198838

[8,] 0.079548117 0.015830863 0.261673195 -0.38254954 0.12380803

[9,] 0.387523482 0.094746528 0.187502181 0.12427899 0.16043171

[,6] [,7] [,8] [,9]

[1,] -0.001069746 0.07958671 0.07954812 0.38752348

[2,] -0.002800916 0.05107486 0.01583086 0.09474653

[3,] 0.026724172 0.10050633 0.26167319 0.18750218

[4,] 0.068175272 0.20769268 -0.38254954 0.12427899

[5,] -0.016060377 -0.01198838 0.12380803 0.16043171

[6,] 1.000000000 -0.18924933 -0.07482541 0.23196001

[7,] -0.189249333 1.00000000 0.03732241 -0.02325832

[8,] -0.074825415 0.03732241 1.00000000 0.51026027

[9,] 0.231960007 -0.02325832 0.51026027 1.00000000

>

> #No strong corelation

> model.tcp <- lm(Price~.,data=tcp)

> summary(model.tcp)

Call:

lm(formula = Price ~ ., data = tcp)

Residuals:

Min 1Q Median 3Q Max

-9366.4 -793.3 -21.3 799.7 6444.0

Coefficients:

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

(Intercept) -5.573e+03 1.411e+03 -3.949 8.24e-05 \*\*\*

Age\_08\_04 -1.217e+02 2.616e+00 -46.512 < 2e-16 \*\*\*

KM -2.082e-02 1.252e-03 -16.622 < 2e-16 \*\*\*

HP 3.168e+01 2.818e+00 11.241 < 2e-16 \*\*\*

cc -1.211e-01 9.009e-02 -1.344 0.17909

Doors -1.617e+00 4.001e+01 -0.040 0.96777

Gears 5.943e+02 1.971e+02 3.016 0.00261 \*\*

Quarterly\_Tax 3.949e+00 1.310e+00 3.015 0.00262 \*\*

Weight 1.696e+01 1.068e+00 15.880 < 2e-16 \*\*\*

---

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

Residual standard error: 1342 on 1427 degrees of freedom

Multiple **R-squared: 0.8638**, Adjusted R-squared: 0.863

F-statistic: 1131 on 8 and 1427 DF, p-value: < 2.2e-16

>

>

> #here we can see, which variable to eliminate if its convinient

> library(car)

> vif(model.tcp)

Age\_08\_04 KM HP cc Doors

1.884620 1.756905 1.419422 1.163894 1.156575

Gears Quarterly\_Tax Weight

1.098723 2.311431 2.516420

> # gives degree of collinearity between the variables

> #coolinearity values are low thus no need to delete any variable

> avPlots(model.tcp)

> # which variable is contributing least towards prediction

> #thus no one has plot with zero slope....thus no need to delete.

> #model is great in itself.

>

> influence.measures(model.tcp)

Influence measures of

lm(formula = Price ~ ., data = tcp) :

dfb.1\_ dfb.A\_08 dfb.KM dfb.HP dfb.cc dfb.Dors dfb.Gers

1 -4.12e-02 0.044717 0.059027 -1.17e-03 -0.014976 0.096060 2.24e-02

2 -3.45e-02 0.059102 -0.001105 -1.01e-02 -0.008099 0.074401 2.04e-02

3 -3.31e-02 0.028508 0.063641 1.27e-03 -0.013590 0.082311 1.84e-02

4 -1.60e-02 0.013970 0.028753 -4.75e-04 -0.006510 0.042743 9.91e-03

5 -1.52e-02 -0.002741 0.072923 5.05e-03 -0.010744 0.072989 1.52e-02

6 -1.58e-02 0.011740 0.039079 -2.56e-03 -0.008133 0.076683 1.86e-02

7 -4.33e-03 -0.001858 0.002233 -1.24e-03 -0.001166 -0.008777 -2.00e-03

8 -4.38e-02 0.005095 -0.009213 -1.65e-02 -0.008850 -0.078473 -1.60e-02

9 -4.82e-03 -0.007519 0.018792 1.85e-01 -0.016752 -0.072732 -6.05e-02

10 -4.18e-02 0.056381 0.001804 4.19e-02 -0.014742 0.031078 -1.63e-03

11 -2.34e-02 -0.003322 0.004450 3.12e-02 -0.003189 -0.008842 2.16e-02

12 6.08e-02 0.016464 -0.021051 -8.51e-02 0.008999 0.023878 -5.72e-02

13 7.98e-02 0.011624 -0.015717 -1.07e-01 0.010912 0.030160 -7.36e-02

14 -1.14e-01 0.002057 0.005068 1.43e-01 -0.014351 -0.040780 1.01e-01

15 -2.24e-01 -0.003950 0.029979 2.85e-01 -0.029841 -0.080774 1.97e-01

16 -1.16e-01 -0.002026 0.004349 1.47e-01 -0.014424 -0.042010 1.05e-01

17 -1.44e-02 -0.021249 0.070161 4.38e-01 -0.041670 -0.170348 -1.43e-01

18 9.83e-03 -0.035375 -0.006408 8.06e-03 -0.000768 -0.044804 -1.39e-02

19 7.65e-03 -0.049976 0.016069 -8.49e-03 0.012411 -0.036576 -9.43e-03

20 1.12e-02 -0.065751 0.045563 2.35e-02 -0.005941 -0.064418 -2.40e-02

21 6.62e-03 -0.039214 0.028962 1.40e-02 -0.003607 -0.036965 -1.40e-02

22 -8.58e-03 -0.004383 0.001515 8.53e-04 -0.004328 -0.017932 -5.07e-03

23 -1.92e-04 -0.013834 0.008533 3.81e-03 -0.001912 -0.016296 -5.52e-03

24 -1.56e-03 -0.017874 -0.000289 4.07e-03 -0.002538 -0.031518 -9.56e-03

25 -7.17e-04 -0.004141 -0.001673 9.24e-04 -0.000664 -0.009331 -2.74e-03

26 4.87e-05 0.008859 0.000228 -1.58e-03 0.000975 0.013365 4.01e-03

27 -1.32e-03 -0.029687 0.003168 6.54e-03 -0.003868 -0.046288 -1.42e-02

28 -2.05e-04 -0.002483 0.000673 6.61e-04 -0.000376 -0.003946 -1.25e-03

29 -1.21e-03 -0.027988 0.010135 7.17e-03 -0.003911 -0.039560 -1.27e-02

30 -6.59e-03 -0.012143 -0.030217 1.74e-03 -0.002701 -0.055746 -1.51e-02

31 3.82e-02 0.034096 0.017837 7.94e-02 0.004270 0.070814 2.95e-03

32 -8.05e-04 0.003949 -0.000468 1.48e-03 0.001129 0.003451 6.24e-04

33 1.87e-03 -0.011999 -0.002414 -7.09e-03 -0.004836 -0.015046 -2.50e-03

34 -1.98e-03 0.011862 0.000633 5.98e-03 0.004320 0.013257 2.29e-03

35 -1.69e-03 0.008503 -0.001982 2.76e-03 0.002342 0.006964 1.33e-03

36 2.35e-03 -0.014565 0.003079 -5.57e-03 -0.004827 -0.014102 -2.70e-03

37 9.40e-04 -0.005066 -0.001315 -2.95e-03 -0.001876 -0.005997 -9.58e-04

38 -6.04e-03 0.026327 0.019627 2.01e-02 0.010083 0.035176 4.70e-03

39 4.01e-04 -0.011017 -0.009881 -1.19e-02 -0.007492 -0.023547 -3.63e-03

40 -2.20e-03 0.014005 0.003877 8.78e-03 0.005787 0.018187 2.96e-03

41 -1.14e-02 0.059259 -0.021871 1.58e-02 0.015595 0.044677 9.19e-03

42 6.72e-03 -0.044813 0.009761 -1.76e-02 -0.015550 -0.045113 -8.68e-03

43 -1.15e-02 0.059488 -0.019985 1.67e-02 0.015834 0.045792 9.25e-03

44 -6.30e-03 -0.003720 0.005154 -6.01e-04 -0.002857 0.000786 -8.74e-04

45 8.43e-03 0.007321 -0.006244 1.47e-03 0.004023 -0.001281 1.12e-03

46 -2.68e-02 -0.004289 0.000106 -8.38e-03 -0.009548 0.000740 -1.59e-03

47 -1.42e-02 0.000349 -0.004099 -4.22e-03 -0.004893 0.002078 -7.09e-04

48 1.46e-03 -0.025633 0.017528 -6.29e-03 -0.006462 0.008986 8.55e-04

49 1.42e-03 0.000314 0.000542 5.31e-04 0.000518 -0.000244 6.07e-05

50 -2.18e-01 -0.032934 0.083345 2.64e-01 -0.032425 0.038219 2.01e-01

51 5.23e-03 0.000112 0.003809 2.28e-03 0.001731 -0.000934 7.85e-05

52 -3.02e-03 -0.028618 0.012486 -1.55e-02 -0.011643 0.017847 2.95e-03

53 -8.15e-02 -0.060152 0.046908 -4.03e-03 -0.037384 0.018114 -1.14e-02

54 -1.82e-02 -0.034702 0.077997 3.00e-01 -0.031489 -0.015175 -8.18e-02

55 1.94e-02 0.022882 -0.011226 -5.13e-04 0.010373 -0.009138 3.34e-03

56 -1.61e-02 0.030300 -0.006484 -1.11e-02 -0.001991 -0.022955 1.56e-03

57 -2.31e-04 -0.002337 0.000299 -1.58e-03 -0.000975 0.001637 3.35e-04

58 1.75e-04 0.003736 -0.000635 2.24e-03 0.001380 -0.002327 -4.70e-04

59 -3.11e-02 -0.021118 0.005222 -4.72e-03 -0.013535 0.007665 -3.37e-03

60 -2.23e-03 -0.004192 0.000355 2.86e-05 -0.000866 0.003301 -3.69e-04

61 -7.02e-04 0.030835 -0.003819 1.68e-02 0.009595 -0.016899 -3.60e-03

62 -2.45e-03 -0.007079 0.000916 1.08e-05 -0.001121 0.004483 -4.93e-04

63 -2.68e-02 -0.039307 -0.003720 -9.01e-04 -0.009325 0.036919 -3.64e-03

64 -1.75e-02 -0.027441 -0.005331 -1.39e-03 -0.006001 0.025297 -2.20e-03

65 -5.49e-03 -0.050207 -0.008726 -4.17e-02 -0.022202 0.040012 9.35e-03

66 -1.72e-03 -0.060191 0.023555 -1.81e-02 0.015055 0.029840 2.37e-03

67 -2.06e-02 -0.042119 -0.007579 -2.51e-03 -0.007553 0.033863 -2.70e-03

68 -3.96e-04 0.031936 0.001485 2.02e-02 0.010512 -0.019421 -4.54e-03

69 -5.65e-02 0.036023 -0.073514 3.72e-02 -0.022532 -0.002838 -1.50e-02

70 -2.04e-04 -0.003457 -0.000642 -2.69e-03 -0.001379 0.002547 6.11e-04

71 -1.50e-03 -0.008950 -0.003496 -8.91e-03 -0.004514 0.008324 2.05e-03

dfb.Qr\_T dfb.Wght dffit cov.r cook.d hat inf

1 -0.15549 0.023320 -0.25567 0.978 7.24e-03 0.01047 \*

2 -0.10665 0.019444 -0.18914 0.994 3.97e-03 0.00992

3 -0.13673 0.018183 -0.22205 0.988 5.46e-03 0.01068

4 -0.06938 0.008110 -0.11251 1.009 1.41e-03 0.01025

5 -0.11718 0.000423 -0.19281 0.995 4.12e-03 0.01071

6 -0.11158 -0.000719 -0.18731 0.993 3.89e-03 0.00944

7 0.00223 0.008727 0.01779 1.019 3.52e-05 0.01276 \*

8 0.02698 0.082656 0.15942 1.006 2.82e-03 0.01257

9 0.03886 0.020912 0.21168 1.036 4.98e-03 0.03591 \*

10 -0.05328 0.036831 -0.12339 1.008 1.69e-03 0.01062

11 0.00595 0.003699 0.04817 1.050 2.58e-04 0.04190 \*

12 -0.01386 -0.008326 -0.13015 1.049 1.88e-03 0.04273 \*

13 -0.02016 -0.012592 -0.16428 1.046 3.00e-03 0.04193 \*

14 0.03076 0.022039 0.22291 1.042 5.52e-03 0.04157 \*

15 0.05514 0.043955 0.43900 1.021 2.14e-02 0.04177 \*

16 0.03242 0.020591 0.23016 1.042 5.88e-03 0.04154 \*

17 0.08288 0.052794 0.49632 1.002 2.73e-02 0.03641 \*

18 -0.00758 0.011037 0.07777 1.002 6.72e-04 0.00443

19 -0.05500 0.015975 0.08981 1.007 8.96e-04 0.00751

20 -0.02552 0.019755 0.10940 0.994 1.33e-03 0.00449

21 -0.01534 0.011202 0.06388 1.006 4.54e-04 0.00468

22 -0.00946 0.019490 0.02920 1.012 9.48e-05 0.00684

23 -0.00812 0.008844 0.02681 1.010 7.99e-05 0.00483

24 -0.01167 0.018014 0.04995 1.007 2.77e-04 0.00435

25 -0.00310 0.005563 0.01479 1.010 2.43e-05 0.00431

26 0.00481 -0.007030 -0.02195 1.010 5.36e-05 0.00465

27 -0.01788 0.025555 0.07410 1.003 6.10e-04 0.00445

28 -0.00164 0.002277 0.00619 1.011 4.26e-06 0.00431

29 -0.01718 0.022042 0.06302 1.005 4.41e-04 0.00446

30 -0.01381 0.035204 0.09287 0.999 9.58e-04 0.00466

31 0.13013 -0.100056 -0.17749 0.997 3.50e-03 0.01010

32 0.00156 -0.001246 -0.00678 1.011 5.12e-06 0.00501

33 -0.00596 0.006998 0.02766 1.010 8.50e-05 0.00433

34 0.00557 -0.005854 -0.02457 1.010 6.71e-05 0.00443

35 0.00339 -0.002466 -0.01374 1.011 2.10e-05 0.00509

36 -0.00682 0.006038 0.02611 1.010 7.58e-05 0.00449

37 -0.00225 0.002592 0.01142 1.011 1.45e-05 0.00462

38 0.01015 -0.014521 -0.07204 1.005 5.77e-04 0.00522

39 -0.00825 0.013373 0.04161 1.008 1.92e-04 0.00397

40 0.00698 -0.008507 -0.03358 1.009 1.25e-04 0.00436

41 0.02394 -0.015377 -0.08941 1.002 8.88e-04 0.00530

42 -0.02194 0.020118 0.08227 1.001 7.52e-04 0.00436

43 0.02395 -0.015877 -0.09121 1.002 9.24e-04 0.00523

44 0.00828 0.008379 0.02692 1.018 8.06e-05 0.01197

45 -0.01314 -0.011509 -0.04065 1.018 1.84e-04 0.01195

46 0.02643 0.036154 0.09173 1.015 9.35e-04 0.01255

47 0.02099 0.018015 0.05522 1.016 3.39e-04 0.01160

48 -0.01131 0.002018 0.03364 1.011 1.26e-04 0.00589

49 -0.00243 -0.001849 -0.00628 1.018 4.39e-06 0.01176

50 0.03610 0.023334 0.41197 1.026 1.88e-02 0.04256 \*

51 -0.00920 -0.006719 -0.02282 1.018 5.79e-05 0.01201

52 -0.01734 0.009119 0.05043 1.005 2.83e-04 0.00354

53 -0.08500 0.128377 0.20708 0.963 4.74e-03 0.00544 \*

54 0.05001 0.019377 0.33250 1.020 1.23e-02 0.03345 \*

55 0.02066 -0.031619 -0.06294 1.005 4.40e-04 0.00440

56 -0.00650 0.021276 -0.04811 1.005 2.57e-04 0.00312

57 -0.00134 0.000753 0.00455 1.010 2.30e-06 0.00355

58 0.00192 -0.000922 -0.00670 1.010 4.99e-06 0.00380

59 -0.02992 0.049226 0.08144 1.004 7.37e-04 0.00535

60 -0.00352 0.003732 0.01138 1.009 1.44e-05 0.00298

61 0.01330 -0.004637 -0.05153 1.007 2.95e-04 0.00432

62 -0.00478 0.004491 0.01607 1.009 2.87e-05 0.00324

63 -0.03759 0.043350 0.12513 0.976 1.73e-03 0.00291 \*

64 -0.02473 0.028632 0.08650 0.994 8.30e-04 0.00299

65 -0.02797 0.017563 0.11166 0.989 1.38e-03 0.00367

66 -0.07346 0.012862 0.11079 1.002 1.36e-03 0.00676

67 -0.03234 0.035459 0.11845 0.982 1.56e-03 0.00315

68 0.01369 -0.005499 -0.05856 1.006 3.81e-04 0.00430

69 0.10463 0.068499 0.22238 1.002 5.49e-03 0.01534

70 -0.00172 0.000965 0.00732 1.010 5.96e-06 0.00392

71 -0.00541 0.003947 0.02295 1.009 5.86e-05 0.00362

[ reached 'max' / getOption("max.print") -- omitted 1365 rows ]

> influenceIndexPlot(model.tcp,id.n=3)

There were 50 or more warnings (use warnings() to see the first 50)

> influencePlot(model.tcp,id.n=3)

StudRes Hat CookD

81 8.164500 0.9182368 79.5201062

222 -7.673262 0.1397116 1.0210312

961 -5.456195 0.1572484 0.6049996

> tcpfin<-tcp[-c(81,222),]

> model.tcpfin1 <- lm(Price~., data=tcpfin)

There were 21 warnings (use warnings() to see them)

> summary(model.tcpfin1)

Call:

lm(formula = Price ~ ., data = tcpfin)

Residuals:

Min 1Q Median 3Q Max

-10451.2 -754.1 -22.2 743.2 6517.8

Coefficients:

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

(Intercept) -1.025e+04 1.396e+03 -7.346 3.44e-13 \*\*\*

Age\_08\_04 -1.166e+02 2.508e+00 -46.475 < 2e-16 \*\*\*

KM -1.728e-02 1.237e-03 -13.966 < 2e-16 \*\*\*

HP 3.929e+01 2.816e+00 13.953 < 2e-16 \*\*\*

cc -3.206e+00 3.050e-01 -10.511 < 2e-16 \*\*\*

Doors -6.537e+01 3.815e+01 -1.713 0.08685 .

Gears 4.850e+02 1.866e+02 2.599 0.00944 \*\*

Quarterly\_Tax 7.966e+00 1.382e+00 5.762 1.01e-08 \*\*\*

Weight 2.505e+01 1.186e+00 21.129 < 2e-16 \*\*\*

---

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

Residual standard error: 1270 on 1425 degrees of freedom

Multiple **R-squared: 0.8778**, Adjusted R-squared: 0.8772

F-statistic: 1280 on 8 and 1425 DF, p-value: < 2.2e-16

> sqrt(mean(model.tcpfin1$residuals\*\*2))

[1] **1265.72**

>

> model.tcpfin2 <- lm (Price ~ sqrt(Age\_08\_04)+ sqrt(KM)+sqrt(HP) + sqrt(cc)

+ +sqrt(Doors)+ sqrt(Gears)+sqrt(Quarterly\_Tax)+sqrt(Weight),data=tcpfin)

> summary(model.tcpfin2)

Call:

lm(formula = Price ~ sqrt(Age\_08\_04) + sqrt(KM) + sqrt(HP) +

sqrt(cc) + sqrt(Doors) + sqrt(Gears) + sqrt(Quarterly\_Tax) +

sqrt(Weight), data = tcpfin)

Residuals:

Min 1Q Median 3Q Max

-7962.1 -716.3 -44.0 731.8 7230.2

Coefficients:

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

(Intercept) -2.957e+04 2.741e+03 -10.787 < 2e-16 \*\*\*

sqrt(Age\_08\_04) -1.565e+03 3.428e+01 -45.651 < 2e-16 \*\*\*

sqrt(KM) -7.825e+00 6.411e-01 -12.205 < 2e-16 \*\*\*

sqrt(HP) 8.364e+02 5.269e+01 15.874 < 2e-16 \*\*\*

sqrt(cc) -2.288e+02 2.291e+01 -9.986 < 2e-16 \*\*\*

sqrt(Doors) -2.578e+02 1.464e+02 -1.761 0.078531 .

sqrt(Gears) 2.787e+03 8.120e+02 3.432 0.000616 \*\*\*

sqrt(Quarterly\_Tax) 1.687e+02 2.402e+01 7.022 3.37e-12 \*\*\*

sqrt(Weight) 1.439e+03 7.782e+01 18.496 < 2e-16 \*\*\*

---

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

Residual standard error: 1226 on 1425 degrees of freedom

Multiple **R-squared: 0.886**, Adjusted R-squared: 0.8854

F-statistic: 1385 on 8 and 1425 DF, p-value: < 2.2e-16

> sqrt(mean(model.tcpfin2$residuals\*\*2))

[1] **1222.632**

>

> model.tcpfin3 <- lm (Price ~ (Age\_08\_04)\*\*2+ (KM)\*\*2+(HP)\*\*2 + (cc)\*\*2

+ +(Doors)\*\*2+ (Gears)\*\*2+(Quarterly\_Tax)\*\*2+(Weight)\*\*2,data=tcpfin)

>

> summary(model.tcpfin3)

Call:

lm(formula = Price ~ (Age\_08\_04)^2 + (KM)^2 + (HP)^2 + (cc)^2 +

(Doors)^2 + (Gears)^2 + (Quarterly\_Tax)^2 + (Weight)^2, data = tcpfin)

Residuals:

Min 1Q Median 3Q Max

-10451.2 -754.1 -22.2 743.2 6517.8

Coefficients:

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

(Intercept) -1.025e+04 1.396e+03 -7.346 3.44e-13 \*\*\*

Age\_08\_04 -1.166e+02 2.508e+00 -46.475 < 2e-16 \*\*\*

KM -1.728e-02 1.237e-03 -13.966 < 2e-16 \*\*\*

HP 3.929e+01 2.816e+00 13.953 < 2e-16 \*\*\*

cc -3.206e+00 3.050e-01 -10.511 < 2e-16 \*\*\*

Doors -6.537e+01 3.815e+01 -1.713 0.08685 .

Gears 4.850e+02 1.866e+02 2.599 0.00944 \*\*

Quarterly\_Tax 7.966e+00 1.382e+00 5.762 1.01e-08 \*\*\*

Weight 2.505e+01 1.186e+00 21.129 < 2e-16 \*\*\*

---

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

Residual standard error: 1270 on 1425 degrees of freedom

Multiple **R-squared: 0.8778**, Adjusted R-squared: 0.8772

F-statistic: 1280 on 8 and 1425 DF, p-value: < 2.2e-16

> sqrt(mean(model.tcpfin3$residuals\*\*2))

[1] **1265.72**

>

>

> #there are zeroes present in the data thus log can't be applied

>

> model.tcpfin4 <- lm (log(Price) ~ .,data=tcpfin)

> summary(model.tcpfin4)

Call:

lm(formula = log(Price) ~ ., data = tcpfin)

Residuals:

Min 1Q Median 3Q Max

-0.70563 -0.06272 0.00066 0.07211 0.53026

Coefficients:

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

(Intercept) 8.328e+00 1.242e-01 67.045 < 2e-16 \*\*\*

Age\_08\_04 -1.010e-02 2.232e-04 -45.248 < 2e-16 \*\*\*

KM -1.868e-06 1.101e-07 -16.972 < 2e-16 \*\*\*

HP 2.954e-03 2.506e-04 11.786 < 2e-16 \*\*\*

cc -1.422e-04 2.714e-05 -5.241 1.84e-07 \*\*\*

Doors 7.952e-03 3.395e-03 2.342 0.019314 \*

Gears 6.370e-02 1.661e-02 3.836 0.000131 \*\*\*

Quarterly\_Tax 8.092e-04 1.230e-04 6.577 6.73e-11 \*\*\*

Weight 1.025e-03 1.055e-04 9.717 < 2e-16 \*\*\*

---

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

Residual standard error: 0.113 on 1425 degrees of freedom

Multiple **R-squared: 0.855**, Adjusted R-squared: 0.8542

F-statistic: 1050 on 8 and 1425 DF, p-value: < 2.2e-16

> model.tcpfin4\_pred <- (exp(model.tcpfin4$fitted.values))

> model.tcpfin4\_err<- tcpfin$Price - model.tcpfin4\_pred

> model.tcpfin4\_rmse <- sqrt(mean(model.tcpfin4\_err^2))

> model.tcpfin4\_rmse

[1] **1157.339**

**Plots ->**

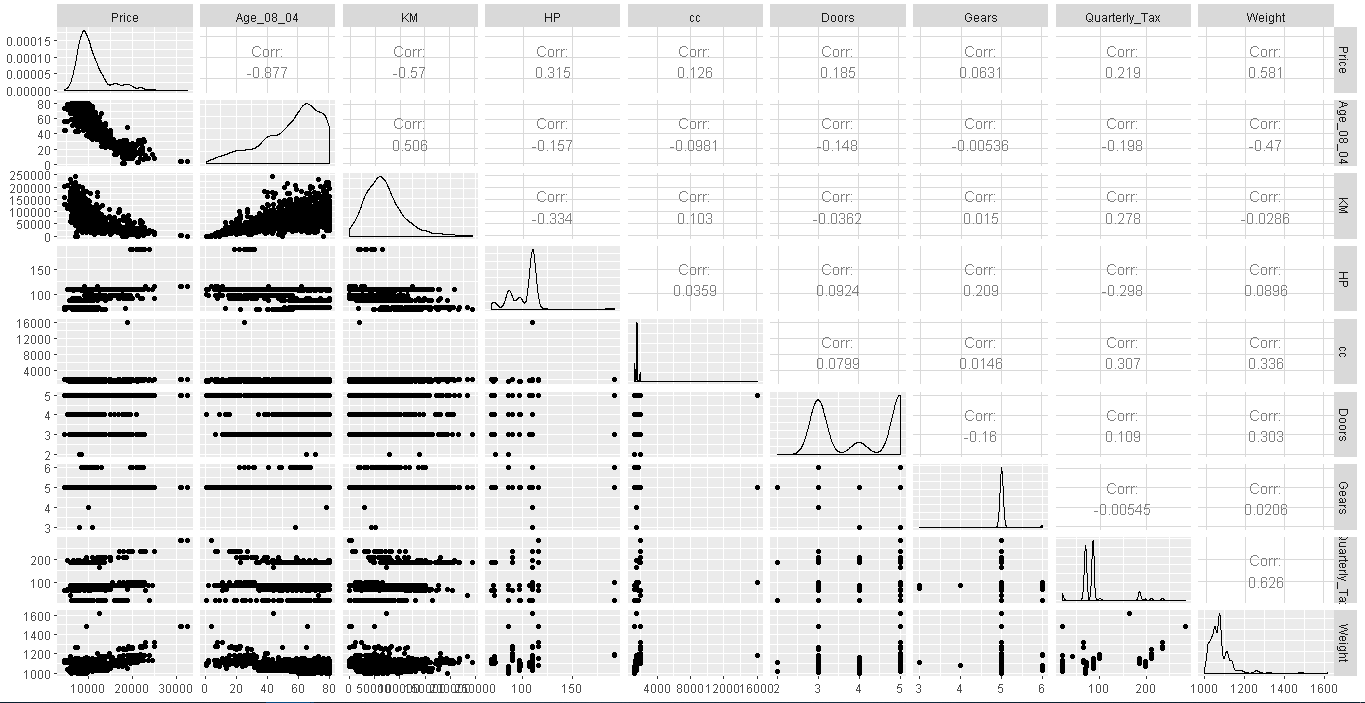


Figure 9. GG PLot

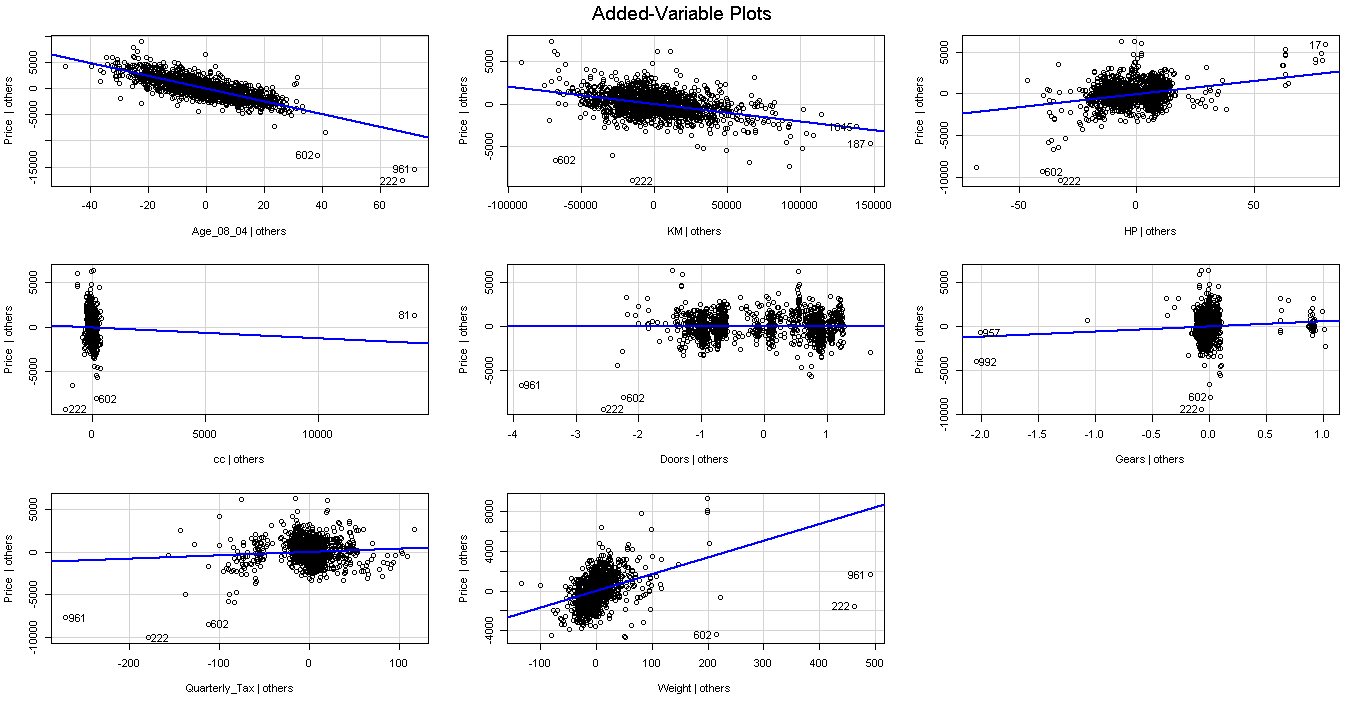


Figure 10. AV PLot

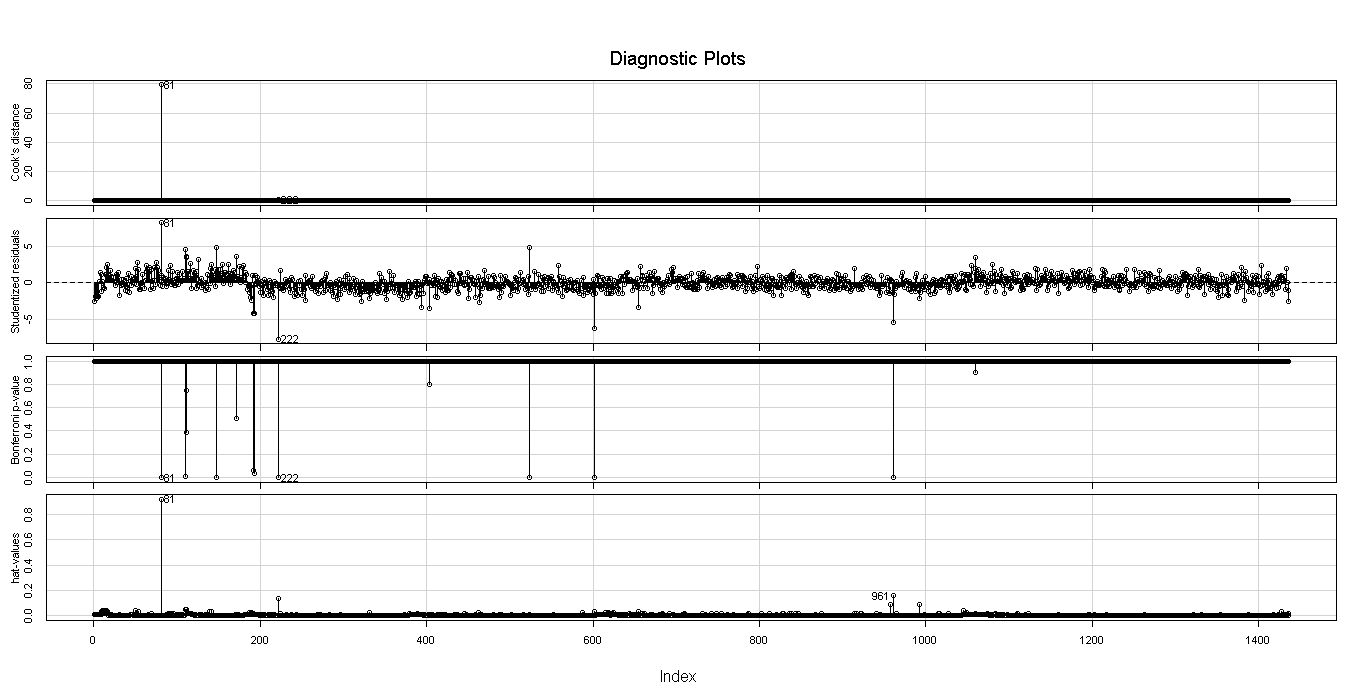


Figure 11. Influence Measurement Model 1

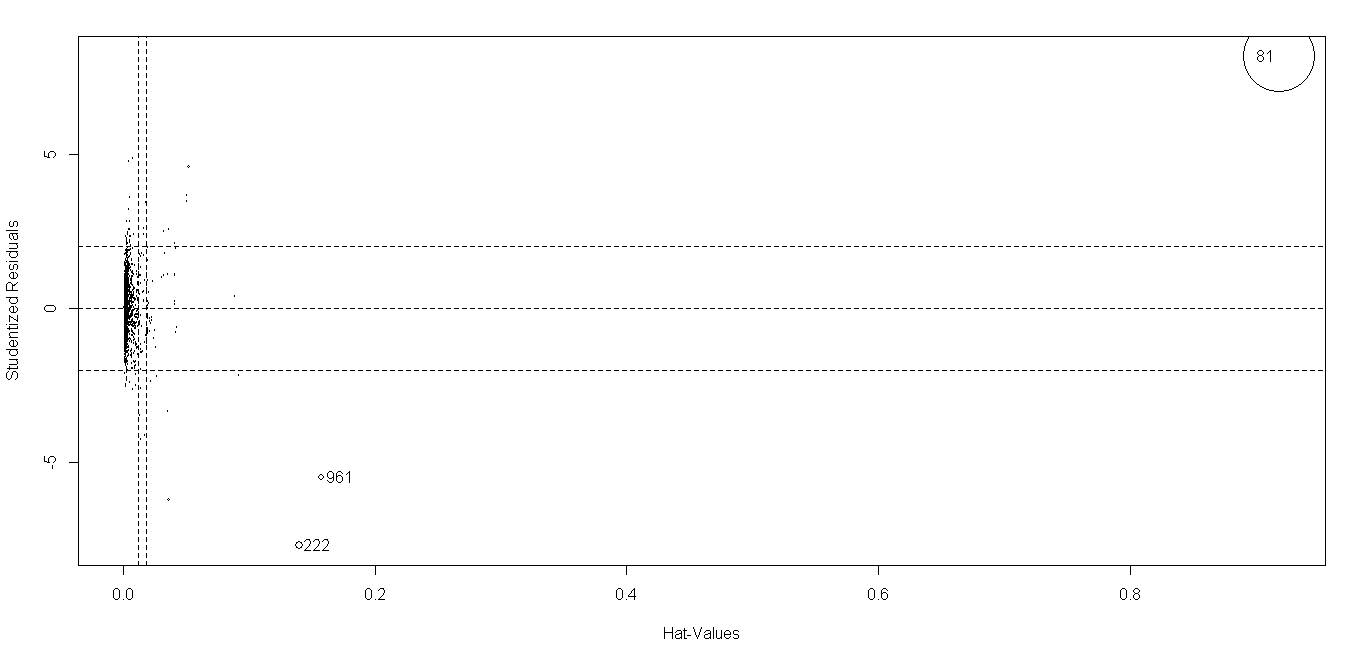


Figure 12. Influence Measurement Model 2