Predict Price of the computer

#importing the data

library(readr)

data <- read.csv(file.choose())

View(data)

str(data)

attach(data1)

data1 <- data[,-1]##The first column is the index which would not effect the analysis of the data so we removed

attach(data1)

##First moment business model ##

summary(data1)

price speed hd ram screen cd multi premium

Min. : 949 Min. : 25.00 Min. : 80.0 Min. : 2.000 Min. :14.00 no :3351 no :5386 no : 612

1st Qu.:1794 1st Qu.: 33.00 1st Qu.: 214.0 1st Qu.: 4.000 1st Qu.:14.00 yes:2908 yes: 873 yes:5647

Median :2144 Median : 50.00 Median : 340.0 Median : 8.000 Median :14.00

Mean :2220 Mean : 52.01 Mean : 416.6 Mean : 8.287 Mean :14.61

3rd Qu.:2595 3rd Qu.: 66.00 3rd Qu.: 528.0 3rd Qu.: 8.000 3rd Qu.:15.00

Max. :5399 Max. :100.00 Max. :2100.0 Max. :32.000 Max. :17.00

ads trend

Min. : 39.0 Min. : 1.00

1st Qu.:162.5 1st Qu.:10.00

Median :246.0 Median :16.00

Mean :221.3 Mean :15.93

3rd Qu.:275.0 3rd Qu.:21.50

Max. :339.0 Max. :35.00

##from this we can identify there three factor variable and which must be converted by taking the dummy variables

##we should normalized the data for aqurate model

##Second moment business model

var(speed)##447.6498 the data points are spread from the mean

var(hd)##66847 the data points are spread from the mean

var(ram)##31.70928 the data points are not that spread from the mean compare other variables

var(screen)## 0.8192336 the data points are near to mean

var(ads)##5600.32 data point are spread from the mean

var(trend)##61.99 data point less spread from the mean compare to other variables

sd(speed)##21.1577 speed variable is 21 times deviated from the mean

sd(hd)##258.5484 hd variable has the higest sd compare other variables in the data set

sd(ram)##5.63 less sd compare to other variables

sd(screen)##0.9051 the data is more normalized

sd(ads)##74 the data is spread from the mean

sd(trend)##7.87 the data is spred from the mean

##Third moment business model

skewness(speed)##0.668505 moderately skewed right

skewness(hd)## 1.377689 storngly skewed right

skewness(ram)##1.377689 storngly skewed right

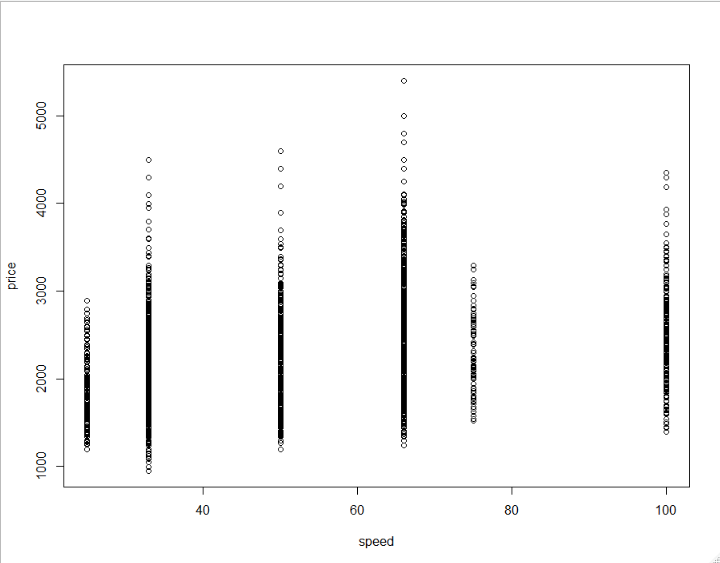
skewness(screen)##1.633616 storngly skewed right

skewness(ads)##-0.5531955 negatively skewed left

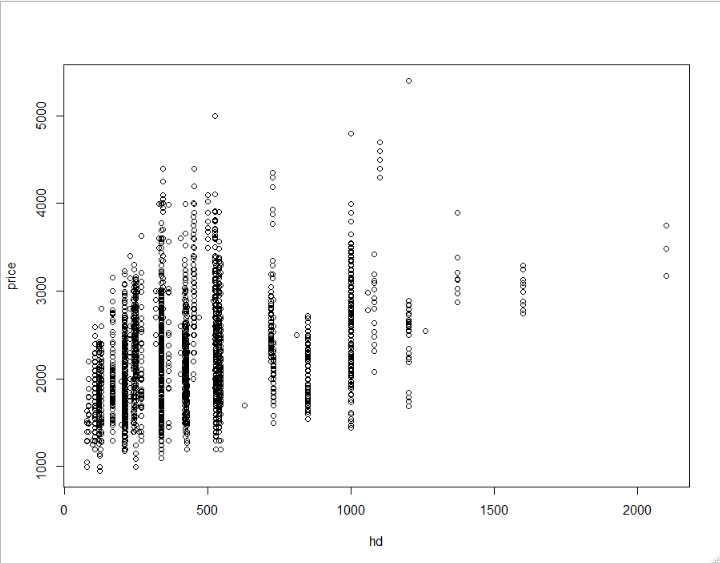
skewness(trend)##0.236617 positively skewed right

##fourth moment business model

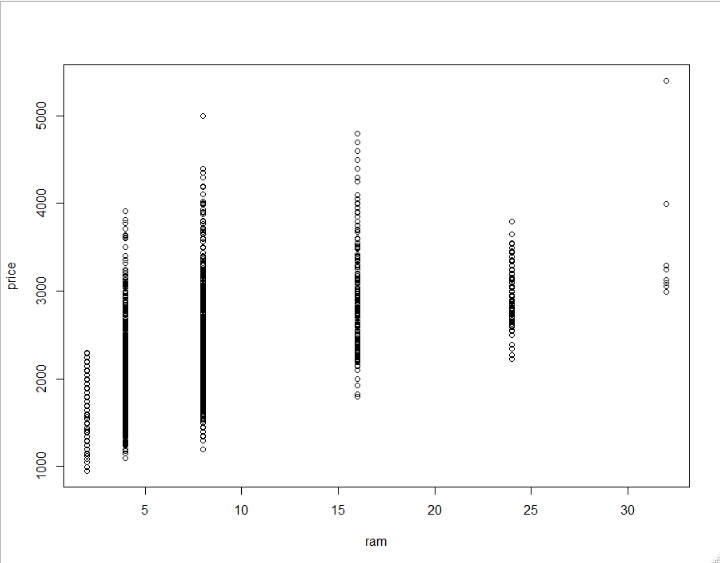
plot(speed,price)



plot(hd,price)## the hd does not effect the price



plot(ram,price)## more data point are with ram less then the 10



plot(screen,price)

plot(ads,price)

plot(trend,price)

#no individually variable is effecting the price variable

##data preprocessing

data1$cd=factor(data1$cd,

levels =c('no','yes'), labels =c(0,1))

data1$multi=factor(data1$multi,

levels =c('no','yes'), labels =c(0,1))

data1$premium=factor(data1$premium,

levels =c('no','yes'), labels =c(0,1))

pairs(data1)

View(data1)

## model buliding##

model <- lm(price~.,data = data1)

summary(model)

##multiple r square value .7756##

model2 <- lm(log(price)~.,data = data1)

summary(model2)

##multipler-squared value .7832

##prediction of model##

pre <- predict(model2)

pre

plot(model2)

