# Homework 2

**Name: Anisha Desai**

**PUID; 0032031106**

**Course: MS BAIM**

rm(list = ls())  
db = read.csv("carsdata.csv")  
  
## coerce model year into a factor variable, use 2006 as the reference level  
db$modelyear = factor(db$modelyear)  
db$modelyear = relevel(db$modelyear,"2006")  
## coerce month into a factor variable, use month 9 as the reference level  
db$month = factor(db$month)  
db$month = relevel(db$month,"9")  
summary(db)

## sold price mile engine\_vol   
## Min. :0.0000 Min. : 8.599 Min. : 3.10 Min. :1.700   
## 1st Qu.:0.0000 1st Qu.:14.998 1st Qu.: 19.55 1st Qu.:2.400   
## Median :0.0000 Median :15.998 Median : 28.01 Median :2.400   
## Mean :0.1251 Mean :16.562 Mean : 32.53 Mean :2.438   
## 3rd Qu.:0.0000 3rd Qu.:17.998 3rd Qu.: 40.53 3rd Qu.:2.500   
## Max. :1.0000 Max. :24.998 Max. :117.25 Max. :3.500   
##   
## wheelbase model month modelyear   
## Min. :102.0 Length:975 6 :298 2006: 97   
## 1st Qu.:107.0 Class :character 7 :130 2007:367   
## Median :109.0 Mode :character 5 :118 2008:304   
## Mean :107.7 2 :109 2009:156   
## 3rd Qu.:109.0 9 : 94 2010: 51   
## Max. :110.0 1 : 83   
## (Other):143

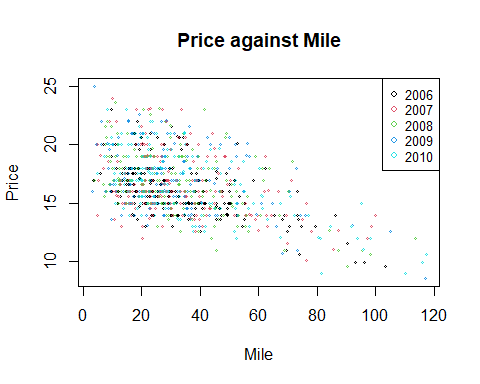
## decompose the mile  
db$mile10k = floor(db$mile/10)\*10  
db$mile1k = floor(db$mile - db$mile10k)  
db$milermd = db$mile - floor(db$mile)  
db$milermd = round(db$milermd,digits = 3)  
  
head(db[,c("mile","mile10k","mile1k","milermd")])

## mile mile10k mile1k milermd  
## 1 21.057 20 1 0.057  
## 2 39.445 30 9 0.445  
## 3 45.727 40 5 0.727  
## 4 20.251 20 0 0.251  
## 5 40.415 40 0 0.415  
## 6 50.365 50 0 0.365

## decompose the price   
db$price10k = floor(db$price/10)\*10  
db$price1k = floor(db$price - db$price10k)  
db$pricermd = db$price - floor(db$price)  
db$pricermd = round(db$pricermd,digits = 3)  
  
head(db[,c("price","price10k","price1k","pricermd")])

## price price10k price1k pricermd  
## 1 17.599 10 7 0.599  
## 2 14.998 10 4 0.998  
## 3 14.698 10 4 0.698  
## 4 14.998 10 4 0.998  
## 5 13.998 10 3 0.998  
## 6 14.998 10 4 0.998

## plot price against mile  
plot(db$mile,db$price,type="p",main="Price against Mile",xlab="Mile",ylab="Price",pch=1,col=rep(1:5,each=10),cex=0.4)  
legend("topright",legend=paste(2006:2010), col=1:5,pch=1,cex=0.8,bty="o")



## Question [1] Plot a scatterplot of price against mile. Briefly explain the major patterns in the price-mile relationship.

1) There is a slight negative correlation i.e., as the mileage is increasing, prices are decreasing.

2) However, there are many horizontal lines seen. Which means that as the miles are increasing, prices are staying constant and not decreasing at the same rate.

3) This means that there are a lot of cars with similar selling price even though the mileage is different and that’s because mileage is not the only factor affecting the price.

## linear price regression  
reg1 = glm(price ~ mile + modelyear + model + month, data = db)  
summary(reg1)  
## Call:  
## glm(formula = price ~ mile + modelyear + model + month, data = db)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.2508 -1.0844 -0.2450 0.9274 4.9922   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 17.788847 0.284916 62.436 < 2e-16 \*\*\*  
## mile -0.052406 0.003111 -16.847 < 2e-16 \*\*\*  
## modelyear2007 1.048105 0.186134 5.631 2.36e-08 \*\*\*  
## modelyear2008 2.237173 0.195578 11.439 < 2e-16 \*\*\*  
## modelyear2009 2.553833 0.215769 11.836 < 2e-16 \*\*\*  
## modelyear2010 -0.277939 0.275602 -1.008 0.313479   
## modelAltima -0.624304 0.146592 -4.259 2.26e-05 \*\*\*  
## modelCamry -1.119073 0.139001 -8.051 2.43e-15 \*\*\*  
## modelCivic -2.506974 0.159042 -15.763 < 2e-16 \*\*\*  
## modelCorolla -3.833755 0.191821 -19.986 < 2e-16 \*\*\*  
## modelSonata -3.251665 0.267446 -12.158 < 2e-16 \*\*\*  
## month1 0.742231 0.231156 3.211 0.001367 \*\*   
## month2 1.537395 0.217868 7.057 3.28e-12 \*\*\*  
## month3 0.962706 0.271792 3.542 0.000416 \*\*\*  
## month4 0.108542 0.258074 0.421 0.674153   
## month5 0.030208 0.221789 0.136 0.891692   
## month6 -0.028787 0.195051 -0.148 0.882699   
## month7 -0.341402 0.211501 -1.614 0.106818   
## month8 -0.173673 0.319780 -0.543 0.587187   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for gaussian family taken to be 2.339377)  
##   
## Null deviance: 6343.1 on 974 degrees of freedom  
## Residual deviance: 2236.4 on 956 degrees of freedom  
## AIC: 3616.4  
##   
## Number of Fisher Scoring iterations: 2

## linear price regression,with mile replaced with the decomposed mile digits  
reg2 = glm(price ~ mile10k + mile1k + milermd + engine\_vol + modelyear + model + month, data = db)  
summary(reg2)  
## Call:  
## glm(formula = price ~ mile10k + mile1k + milermd + engine\_vol +   
## modelyear + model + month, data = db)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.7010 -0.8611 -0.1553 0.8245 5.1999   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 11.862227 0.419176 28.299 < 2e-16 \*\*\*  
## mile10k -0.055302 0.002692 -20.547 < 2e-16 \*\*\*  
## mile1k -0.065550 0.015090 -4.344 1.55e-05 \*\*\*  
## milermd 0.154222 0.145726 1.058 0.290184   
## engine\_vol 2.285679 0.126593 18.055 < 2e-16 \*\*\*  
## modelyear2007 1.052141 0.160938 6.538 1.02e-10 \*\*\*  
## modelyear2008 2.268664 0.168932 13.429 < 2e-16 \*\*\*  
## modelyear2009 2.831005 0.187043 15.136 < 2e-16 \*\*\*  
## modelyear2010 -0.140476 0.238737 -0.588 0.556393   
## modelAltima -0.818766 0.127091 -6.442 1.86e-10 \*\*\*  
## modelCamry -1.143767 0.120153 -9.519 < 2e-16 \*\*\*  
## modelCivic -0.656356 0.171206 -3.834 0.000135 \*\*\*  
## modelCorolla -1.997748 0.194948 -10.248 < 2e-16 \*\*\*  
## modelSonata -3.720641 0.232408 -16.009 < 2e-16 \*\*\*  
## month1 0.651672 0.200217 3.255 0.001175 \*\*   
## month2 1.347402 0.188469 7.149 1.74e-12 \*\*\*  
## month3 1.049963 0.235786 4.453 9.47e-06 \*\*\*  
## month4 0.030028 0.223240 0.135 0.893028   
## month5 -0.067199 0.191676 -0.351 0.725977   
## month6 -0.051488 0.168729 -0.305 0.760316   
## month7 -0.378920 0.183120 -2.069 0.038793 \*   
## month8 0.014435 0.276416 0.052 0.958362   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for gaussian family taken to be 1.744392)  
##   
## Null deviance: 6343.1 on 974 degrees of freedom  
## Residual deviance: 1662.4 on 953 degrees of freedom  
## AIC: 3333.2  
##   
## Number of Fisher Scoring iterations: 2

## Question [2] Regress price on all car attributes (use decomposed mile) and month. How does the price-mile relationship here compare with that shown in the scatterplot?

Only mile10k and mile1k variables are significant as the p-value is less than 0.05. The other i.e. milermd decomposing variable is insignificant while determining the price.

For every 1 unit increase in price, the mile decreases by -0.052334, which is comparatively very less. This justifies the presence of horizontal lines in the plot i.e. a lot of cars have similar prices even though the mileage is different.

## logistic demand regression, with mile replaced with the decomposed price reg3 = glm(sold ~ price10k + price1k + pricermd + mile + modelyear + engine\_vol+model + month, data = db, family = binomial)  
summary(reg3)

##   
## Call:  
## glm(formula = sold ~ price10k + price1k + pricermd + mile + modelyear +   
## engine\_vol + model + month, family = binomial, data = db)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.2446 -0.5487 -0.4431 -0.3250 2.5904   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 3.710371 1.492546 2.486 0.01292 \*   
## price10k -0.404123 0.087502 -4.618 3.87e-06 \*\*\*  
## price1k -0.418397 0.096242 -4.347 1.38e-05 \*\*\*  
## pricermd 0.086713 0.435146 0.199 0.84205   
## mile -0.023793 0.007750 -3.070 0.00214 \*\*   
## modelyear2007 -0.141437 0.349213 -0.405 0.68546   
## modelyear2008 0.025659 0.402747 0.064 0.94920   
## modelyear2009 0.383110 0.452930 0.846 0.39764   
## modelyear2010 -0.146015 0.486002 -0.300 0.76384   
## engine\_vol 0.743527 0.351826 2.113 0.03457 \*   
## modelAltima -0.592642 0.307289 -1.929 0.05378 .   
## modelCamry -0.810200 0.301801 -2.685 0.00726 \*\*   
## modelCivic -0.899725 0.430406 -2.090 0.03658 \*   
## modelCorolla -1.037383 0.474580 -2.186 0.02882 \*   
## modelSonata -1.337210 0.580315 -2.304 0.02121 \*   
## month1 0.658561 0.468691 1.405 0.15999   
## month2 0.581488 0.478260 1.216 0.22405   
## month3 1.404006 0.505235 2.779 0.00545 \*\*   
## month4 -0.251075 0.565219 -0.444 0.65689   
## month5 -0.219554 0.464410 -0.473 0.63639   
## month6 -0.166033 0.414983 -0.400 0.68909   
## month7 0.009632 0.438625 0.022 0.98248   
## month8 -1.440724 1.082431 -1.331 0.18319   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 735.19 on 974 degrees of freedom  
## Residual deviance: 686.82 on 952 degrees of freedom  
## AIC: 732.82  
##   
## Number of Fisher Scoring iterations: 6

## Question [3] Fit a logistic regression for whether a car was sold on the first day to investigate the LDB of car buyers. Does car buyers show LDB in their attention to the digits of price? Briefly explain your answer

The p-value for price10k and price1k is less than 0.05. Therefore, price10k and price1k variables are significant with almost 100% confidence whereas the right ones aren’t significant enough.

As the first 2 digits from the left side are significant, the LDB bias is present for sold vs price.

## Question [4] Briefly discuss the implications of your findings above for the pricing of used cars

We can conclude that the left digit bias exists when a consumer is trying to purchase a used car. Also, the LDS can be across multiple variables of price and miles.

The store managers can manipulate the right digits as they are not considered as a significant factor by consumers while deciding on purchasing the car. This is the reason we always find prices of cars ending with .999 or something similar.