**Dalton Anderson – anderson669@usf.edu**

**Dr. Ronald K. Satterfield -** [**rsatterf@usf.edu**](mailto:rsatterf@usf.edu)

**QMB6304**

**2021/09/03**

Module 5

**Preprocessing**

**R Code**

#Dalton Anderson

library(readxl)

library(dplyr)

library(ggplot2)

library(janitor)

#1.) Load in dataset

#Import data

cl\_master <- clean\_names(read\_excel("6304 Module 5 Assignment Data.xlsx"))

#2.) Create random sample sets by unumber

#Create a primary data set for your analysis of n=250 randomly selected cars. As always, use the numerical portion of your U number for the random number seed and apply the method we've demonstrated in class. The characteristics of this primary data set will be:

#a. Only vehicles with MAKE of “cadillac”.

#b. Only cars from the 2006 through 2011 model years (inclusive).

#c. Only cars with engines of 6 or 8 cylinders.

#Filter data

car\_make = filter(cl\_master, make == 'cadillac')

car\_year\_1 = filter(car\_make, year >= 2006)

car\_year = filter(car\_year\_1, year <= 2011)

car\_cyl = filter(car\_year, !cylinders == 4)

#sample set created

set.seed(59657076)

sample <- car\_cyl

cars = sample\_n(sample, 250)

#check population against sample set

head(cl\_master)

head(cars)

END R CODE

**R Console Output**

**> #Dalton Anderson**

**> library(readxl)**

**> library(dplyr)**

**> library(ggplot2)**

**> library(janitor)**

**>**

**> #1.) Load in dataset**

**> #Import data**

**> cl\_master <- clean\_names(read\_excel("6304 Module 5 Assignment Data.xlsx"))**

**>**

**> #2.) Create random sample sets by unumber**

**> #Create a primary data set for your analysis of n=250 randomly selected cars. As always, use the numerical portion of your U number for the random number seed and apply the method we've demonstrated in class. The characteristics of this primary data set will be:**

**> #a. Only vehicles with MAKE of “cadillac”.**

**> #b. Only cars from the 2006 through 2011 model years (inclusive).**

**> #c. Only cars with engines of 6 or 8 cylinders.**

**>**

**> #Filter data**

**> car\_make = filter(cl\_master, make == 'cadillac')**

**> car\_year\_1 = filter(car\_make, year >= 2006)**

**> car\_year = filter(car\_year\_1, year <= 2011)**

**> car\_cyl = filter(car\_year, !cylinders == 4)**

**>**

**>**

**> #sample set created**

**> set.seed(59657076)**

**> sample <- car\_cyl**

**> cars = sample\_n(sample, 250)**

**>**

**> #check population against sample set**

**> head(cl\_master)**

**# A tibble: 6 × 10**

**region price year make model condition cylinders fuel odometer paint\_color**

**<chr> <dbl> <chr> <chr> <chr> <chr> <dbl> <chr> <dbl> <chr>**

**1 albuquerque 15500 1965 ford musta… excellent 8 gas 4800 blue**

**2 albuquerque 17995 2015 ford trans… good 6 gas 71181 white**

**3 albuquerque 18995 2014 ram proma… good 6 gas 80483 white**

**4 albuquerque 8998 2012 volkswagen jetta… excellent 4 dies… 89000 white**

**5 albuquerque 22500 2003 ford mach1… excellent 8 gas 15700 white**

**6 albuquerque 18995 2014 chevrolet expre… good 8 gas 93187 white**

**> head(cars)**

**# A tibble: 6 × 10**

**region price year make model condition cylinders fuel odometer paint\_color**

**<chr> <dbl> <chr> <chr> <chr> <chr> <dbl> <chr> <dbl> <chr>**

**1 fayetteville, NC 12798 2011 cadi… srx … like new 6 gas 80812 silver**

**2 macon / warner robins 6995 2010 cadi… srx … excellent 6 gas 154192 silver**

**3 new river valley 4800 2006 cadi… dts excellent 8 gas 111200 blue**

**4 wausau, WI 7995 2008 cadi… cts4 good 6 gas 120801 red**

**5 texoma 7200 2008 cadi… dts good 8 gas 65500 white**

**6 mankato, MN 4000 2006 cadi… dts excellent 8 gas 114000 custom**

END R CONSOLE OUTPUT

**Analysis**

**R Code**

#1.) Use R to conduct a multiple linear regression on your random sample

#with PRICE as the dependent variable and ODOMETER, YEAR, and CYLINDERS

#as the independent variables. Make appropriate decisions as to

#whether any of the independent variables should be treated as factor variables.

#cars$price1 = cars$price/1000

cars$odometer1 = cars$odometer^2

# Converting Factor to numeric

cars$condition1 = cars$condition

cars$condition1=as.factor(cars$condition)

print(fact\_con)

cars.out = lm(price ~ odometer + year + condition + paint\_color, data = cars)

summary(cars.out)

cars9.out = lm(price ~ odometer + year + condition, data = cars)

summary(cars9.out)

cars2.out=lm(price ~ odometer + year + I(odometer^2)

,data=cars)

summary(cars2.out)

cars3.out=lm(price ~ odometer + year +I(year^2),data=cars)

summary(cars3.out)

# What about an interaction?

cars4.out=lm(price~odometer + year + I(odometer^2)+

odometer:cylinders,data=cars)

summary(cars4.out)

cars5.out=lm(price~odometer + year + cylinders, family = 'binomial')

summary(cars4.out)

cars6.out=lm(price~ odometer + odometer1 + year,data=cars)

summary(cars6.out)

glm.fit <- glm(price ~ odometer , data = cars, family = binomial)

#looks like car.out is the best model

#2.) Report the beta coefficients and associated p values and beta coefficient

#confidence intervals from your model.

summary(cars.out)

confint(cars.out)

#3. Conduct appropriate analyses and give interpretations to determine

#if your model is a good fit to the data in your primary data set.

plot(cars$price,cars.out$fitted.values,

pch=19,

main="Actuals v. Fitteds, Restaurant Revenue")

abline(0,1,col="red",lwd=3)

plot(cars$price,cars2.out$fitted.values,

pch=19,

main="Actuals v. Fitteds, Restaurant Revenue")

abline(0,1,col="red",lwd=3)

plot(cars$price,cars3.out$fitted.values,

pch=19,

main="Actuals v. Fitteds, Restaurant Revenue")

abline(0,1,col="red",lwd=3)

plot(cars$price,cars4.out$fitted.values,

pch=19,

main="Actuals v. Fitteds, Restaurant Revenue")

abline(0,1,col="red",lwd=3)

plot(cars$price,cars6.out$fitted.values,

pch=19,

main="Actuals v. Fitteds, Restaurant Revenue")

abline(0,1,col="red",lwd=3)

#4.) Assess your model’s conformance with the LINE assumptions of regression.

#code from the professor

# Combined Plot

par(mfrow=c(2,2))

# Linearity

plot(cars$price,cars.out$fitted.values,

pch=19,

main="Actuals v. Fitteds, Restaurant Revenue")

abline(0,1,col="red",lwd=3)

# Normality

qqnorm(cars.out$residuals,pch=19,

main="Normality Plot, Restaurant Revenue")

qqline(cars.out$residuals,lwd=3,col="red")

hist(cars.out$residuals,col="red",

main="Residuals, Restaurant Revenue",

probability = TRUE)

curve(dnorm(x,mean(cars.out$residuals),

sd(cars.out$residuals)),

from=min(cars.out$residuals),

to=max(cars.out$residuals)

,lwd=3,col="blue",add=TRUE)

# Equality of Variances

plot(cars.out$fitted.values,rstandard(cars.out),

pch=19,main="Equality of Variances, Restaurant Revenue")

abline(0,0,lwd=3,col="red")

par(mfrow=c(1,1))

# I think the data looks okay I see there is an outlier. I don't there is anything that

#stands out like what I saw in class. There is a bit of curve I am having trouble

#applying the code from class.

#4.) Assess your model’s conformance with the LINE assumptions of regression.

#Linearity

plot(crime12$reported\_crimes\_per\_million,simlin12$fitted.values,

pch=19,main="Crime Actual v. Fitted Values")

xlim=c(300,2000)ylim=c(300,2000)

abline(0,1,col="red",lwd=3)

#Does not pass linearity

#I am having a hard time with making a logistic regression.

#I need to over this with the professor.

#Normality

qqnorm(crime12$reported\_crimes\_per\_million,pch=19,main="Crime Normality Plot")

qqline(simlin12$fitted.values,col="red",lwd=3)

#There is a bit of curve on the right tail.

#Equality of Variances

plot(simlin12$fitted.values,scale(simlin12$residuals),

pch=19,main="Crime Standardized Residuals",ylim = c(-4,4))

abline(0,0,col="red",lwd=3)

#A bit of heteroscedasticity.

#5.) Throckmorton P. Gildersleeve of Summerfield, Tennessee would like to sell

#his 2011 Cadillac DTS pictured above. He says the vehicle is in "excellent"

#condition and has 175,757 miles on the odometer.

#Mr. Gildersleeve has not shared details of his Cadillac’s engine with you

#because he thinks you know that all 2011 DTS cars had the same famous engine.

#If you know nothing about Cadillac engines the point is easily researched online. Use R and your model to determine what price he should ask for the car.

#Do you believe your pricing advice to the Great Gildersleeve is accurate and usable?

#Give reasoning for your conclusion.

#odometer ==175757

#condition == 'excellent'

#year == 2011

throck = data.frame(odometer=175757,condition = 'excellent', year = '2011')

predict(cars9.out,throck,interval="predict")

#price of the car could be around $11,662

#I don't think the information is useful my model is bad as my squaring isn't working.

#then I had to take inputs out of the model because of I did know the 'famous engine'

#info. I would ask for more information.

END R CODE

**R Console Output**

**> #1.) Use R to conduct a multiple linear regression on your random sample**

**> #with PRICE as the dependent variable and ODOMETER, YEAR, and CYLINDERS**

**> #as the independent variables. Make appropriate decisions as to**

**> #whether any of the independent variables should be treated as factor variables.**

**> #cars$price1 = cars$price/1000**

**> cars$odometer1 = cars$odometer^2**

**> # Converting Factor to numeric**

**> cars$condition1 = cars$condition**

**> cars$condition1=as.factor(cars$condition)**

**> print(fact\_con)**

**Error in print(fact\_con) : object 'fact\_con' not found**

**>**

**> cars.out = lm(price ~ odometer + year + condition + paint\_color, data = cars)**

**> summary(cars.out)**

**Call:**

**lm(formula = price ~ odometer + year + condition + paint\_color,**

**data = cars)**

**Residuals:**

**Min 1Q Median 3Q Max**

**-10053.2 -2862.0 -626.2 2171.4 24195.3**

**Coefficients:**

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

**(Intercept) 1.150e+04 1.418e+03 8.114 2.86e-14 \*\*\***

**odometer -1.444e-02 8.347e-03 -1.729 0.085074 .**

**year2007 3.495e+03 1.058e+03 3.302 0.001110 \*\***

**year2008 1.721e+03 1.096e+03 1.570 0.117741**

**year2009 1.404e+03 1.459e+03 0.962 0.337129**

**year2010 3.606e+03 1.100e+03 3.280 0.001200 \*\***

**year2011 4.728e+03 1.143e+03 4.137 4.92e-05 \*\*\***

**conditionfair -4.496e+03 3.332e+03 -1.349 0.178530**

**conditiongood -3.545e+02 6.865e+02 -0.516 0.606042**

**conditionlike new 3.068e+03 1.068e+03 2.873 0.004449 \*\***

**paint\_colorblue -4.812e+03 1.340e+03 -3.590 0.000403 \*\*\***

**paint\_colorbrown -2.768e+03 1.497e+03 -1.849 0.065682 .**

**paint\_colorcustom 2.722e+02 1.243e+03 0.219 0.826876**

**paint\_colorgrey -1.965e+03 1.388e+03 -1.416 0.158084**

**paint\_colorpurple -3.491e+03 3.396e+03 -1.028 0.305032**

**paint\_colorred -3.103e+03 1.110e+03 -2.794 0.005638 \*\***

**paint\_colorsilver -4.268e+03 8.893e+02 -4.799 2.86e-06 \*\*\***

**paint\_colorwhite -6.460e+02 9.092e+02 -0.711 0.478057**

**---**

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

**Residual standard error: 4613 on 232 degrees of freedom**

**Multiple R-squared: 0.2689, Adjusted R-squared: 0.2154**

**F-statistic: 5.021 on 17 and 232 DF, p-value: 2.981e-09**

**>**

**> cars9.out = lm(price ~ odometer + year + condition, data = cars)**

**> summary(cars9.out)**

**Call:**

**lm(formula = price ~ odometer + year + condition, data = cars)**

**Residuals:**

**Min 1Q Median 3Q Max**

**-10094 -3155 -1492 3685 24888**

**Coefficients:**

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

**(Intercept) 9.740e+03 1.313e+03 7.416 2.07e-12 \*\*\***

**odometer -1.594e-02 8.516e-03 -1.871 0.062514 .**

**year2007 3.638e+03 1.086e+03 3.349 0.000942 \*\*\***

**year2008 1.891e+03 1.129e+03 1.674 0.095351 .**

**year2009 1.356e+03 1.523e+03 0.890 0.374250**

**year2010 4.000e+03 1.139e+03 3.512 0.000531 \*\*\***

**year2011 4.623e+03 1.165e+03 3.968 9.56e-05 \*\*\***

**conditionfair -2.976e+03 3.521e+03 -0.845 0.398920**

**conditiongood -1.072e+02 7.188e+02 -0.149 0.881570**

**conditionlike new 2.446e+03 1.113e+03 2.198 0.028907 \***

**---**

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

**Residual standard error: 4902 on 240 degrees of freedom**

**Multiple R-squared: 0.1462, Adjusted R-squared: 0.1142**

**F-statistic: 4.567 on 9 and 240 DF, p-value: 1.425e-05**

**>**

**> cars2.out=lm(price ~ odometer + year + I(odometer^2)**

**+ ,data=cars)**

**> summary(cars2.out)**

**Call:**

**lm(formula = price ~ odometer + year + I(odometer^2), data = cars)**

**Residuals:**

**Min 1Q Median 3Q Max**

**-10410 -3486 -1450 3585 26788**

**Coefficients:**

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

**(Intercept) 9.192e+03 2.364e+03 3.889 0.000130 \*\*\***

**odometer -3.484e-03 3.551e-02 -0.098 0.921925**

**year2007 3.659e+03 1.093e+03 3.348 0.000945 \*\*\***

**year2008 2.007e+03 1.136e+03 1.767 0.078420 .**

**year2009 1.489e+03 1.534e+03 0.971 0.332738**

**year2010 4.205e+03 1.144e+03 3.675 0.000292 \*\*\***

**year2011 5.061e+03 1.165e+03 4.346 2.04e-05 \*\*\***

**I(odometer^2) -5.872e-08 1.396e-07 -0.421 0.674438**

**---**

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

**Residual standard error: 4939 on 242 degrees of freedom**

**Multiple R-squared: 0.126, Adjusted R-squared: 0.1007**

**F-statistic: 4.982 on 7 and 242 DF, p-value: 2.749e-05**

**>**

**> cars3.out=lm(price ~ odometer + year +I(year^2),data=cars)**

**Error in year^2 : non-numeric argument to binary operator**

**> summary(cars3.out)**

**Error in summary(cars3.out) : object 'cars3.out' not found**

**>**

**> # What about an interaction?**

**> cars4.out=lm(price~odometer + year + I(odometer^2)+**

**+ odometer:cylinders,data=cars)**

**> summary(cars4.out)**

**Call:**

**lm(formula = price ~ odometer + year + I(odometer^2) + odometer:cylinders,**

**data = cars)**

**Residuals:**

**Min 1Q Median 3Q Max**

**-9951.4 -2428.9 258.8 2057.2 23917.5**

**Coefficients:**

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

**(Intercept) 8.903e+03 2.049e+03 4.344 2.06e-05 \*\*\***

**odometer -1.624e-01 3.550e-02 -4.576 7.61e-06 \*\*\***

**year2007 3.507e+03 9.479e+02 3.700 0.000267 \*\*\***

**year2008 2.571e+03 9.865e+02 2.606 0.009723 \*\***

**year2009 4.148e+03 1.362e+03 3.045 0.002584 \*\***

**year2010 5.685e+03 1.005e+03 5.654 4.41e-08 \*\*\***

**year2011 6.731e+03 1.027e+03 6.557 3.32e-10 \*\*\***

**I(odometer^2) -8.792e-08 1.211e-07 -0.726 0.468515**

**odometer:cylinders 2.193e-02 2.437e-03 8.998 < 2e-16 \*\*\***

**---**

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

**Residual standard error: 4282 on 241 degrees of freedom**

**Multiple R-squared: 0.3457, Adjusted R-squared: 0.324**

**F-statistic: 15.92 on 8 and 241 DF, p-value: < 2.2e-16**

**>**

**> cars5.out=lm(price~odometer + year + cylinders, family = 'binomial')**

**Error in eval(predvars, data, env) : object 'price' not found**

**> summary(cars4.out)**

**Call:**

**lm(formula = price ~ odometer + year + I(odometer^2) + odometer:cylinders,**

**data = cars)**

**Residuals:**

**Min 1Q Median 3Q Max**

**-9951.4 -2428.9 258.8 2057.2 23917.5**

**Coefficients:**

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

**(Intercept) 8.903e+03 2.049e+03 4.344 2.06e-05 \*\*\***

**odometer -1.624e-01 3.550e-02 -4.576 7.61e-06 \*\*\***

**year2007 3.507e+03 9.479e+02 3.700 0.000267 \*\*\***

**year2008 2.571e+03 9.865e+02 2.606 0.009723 \*\***

**year2009 4.148e+03 1.362e+03 3.045 0.002584 \*\***

**year2010 5.685e+03 1.005e+03 5.654 4.41e-08 \*\*\***

**year2011 6.731e+03 1.027e+03 6.557 3.32e-10 \*\*\***

**I(odometer^2) -8.792e-08 1.211e-07 -0.726 0.468515**

**odometer:cylinders 2.193e-02 2.437e-03 8.998 < 2e-16 \*\*\***

**---**

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

**Residual standard error: 4282 on 241 degrees of freedom**

**Multiple R-squared: 0.3457, Adjusted R-squared: 0.324**

**F-statistic: 15.92 on 8 and 241 DF, p-value: < 2.2e-16**

**>**

**> cars6.out=lm(price~ odometer + odometer1 + year,data=cars)**

**> summary(cars6.out)**

**Call:**

**lm(formula = price ~ odometer + odometer1 + year, data = cars)**

**Residuals:**

**Min 1Q Median 3Q Max**

**-10410 -3486 -1450 3585 26788**

**Coefficients:**

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

**(Intercept) 9.192e+03 2.364e+03 3.889 0.000130 \*\*\***

**odometer -3.484e-03 3.551e-02 -0.098 0.921925**

**odometer1 -5.872e-08 1.396e-07 -0.421 0.674438**

**year2007 3.659e+03 1.093e+03 3.348 0.000945 \*\*\***

**year2008 2.007e+03 1.136e+03 1.767 0.078420 .**

**year2009 1.489e+03 1.534e+03 0.971 0.332738**

**year2010 4.205e+03 1.144e+03 3.675 0.000292 \*\*\***

**year2011 5.061e+03 1.165e+03 4.346 2.04e-05 \*\*\***

**---**

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

**Residual standard error: 4939 on 242 degrees of freedom**

**Multiple R-squared: 0.126, Adjusted R-squared: 0.1007**

**F-statistic: 4.982 on 7 and 242 DF, p-value: 2.749e-05**

**>**

**> glm.fit <- glm(price ~ odometer , data = cars, family = binomial)**

**Error in eval(family$initialize) : y values must be 0 <= y <= 1**

**> #looks like car.out is the best model**

**>**

**>**

**>**

**>**

**>**

**>**

**> #2.) Report the beta coefficients and associated p values and beta coefficient**

**> #confidence intervals from your model.**

**> summary(cars.out)**

**Call:**

**lm(formula = price ~ odometer + year + condition + paint\_color,**

**data = cars)**

**Residuals:**

**Min 1Q Median 3Q Max**

**-10053.2 -2862.0 -626.2 2171.4 24195.3**

**Coefficients:**

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

**(Intercept) 1.150e+04 1.418e+03 8.114 2.86e-14 \*\*\***

**odometer -1.444e-02 8.347e-03 -1.729 0.085074 .**

**year2007 3.495e+03 1.058e+03 3.302 0.001110 \*\***

**year2008 1.721e+03 1.096e+03 1.570 0.117741**

**year2009 1.404e+03 1.459e+03 0.962 0.337129**

**year2010 3.606e+03 1.100e+03 3.280 0.001200 \*\***

**year2011 4.728e+03 1.143e+03 4.137 4.92e-05 \*\*\***

**conditionfair -4.496e+03 3.332e+03 -1.349 0.178530**

**conditiongood -3.545e+02 6.865e+02 -0.516 0.606042**

**conditionlike new 3.068e+03 1.068e+03 2.873 0.004449 \*\***

**paint\_colorblue -4.812e+03 1.340e+03 -3.590 0.000403 \*\*\***

**paint\_colorbrown -2.768e+03 1.497e+03 -1.849 0.065682 .**

**paint\_colorcustom 2.722e+02 1.243e+03 0.219 0.826876**

**paint\_colorgrey -1.965e+03 1.388e+03 -1.416 0.158084**

**paint\_colorpurple -3.491e+03 3.396e+03 -1.028 0.305032**

**paint\_colorred -3.103e+03 1.110e+03 -2.794 0.005638 \*\***

**paint\_colorsilver -4.268e+03 8.893e+02 -4.799 2.86e-06 \*\*\***

**paint\_colorwhite -6.460e+02 9.092e+02 -0.711 0.478057**

**---**

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

**Residual standard error: 4613 on 232 degrees of freedom**

**Multiple R-squared: 0.2689, Adjusted R-squared: 0.2154**

**F-statistic: 5.021 on 17 and 232 DF, p-value: 2.981e-09**

**> confint(cars.out)**

**2.5 % 97.5 %**

**(Intercept) 8.708985e+03 1.429506e+04**

**odometer -3.088214e-02 2.010708e-03**

**year2007 1.410027e+03 5.580562e+03**

**year2008 -4.384413e+02 3.879850e+03**

**year2009 -1.471606e+03 4.278945e+03**

**year2010 1.439733e+03 5.772879e+03**

**year2011 2.476418e+03 6.979460e+03**

**conditionfair -1.106175e+04 2.068853e+03**

**conditiongood -1.707061e+03 9.980126e+02**

**conditionlike new 9.638009e+02 5.172806e+03**

**paint\_colorblue -7.452757e+03 -2.171158e+03**

**paint\_colorbrown -5.717349e+03 1.809846e+02**

**paint\_colorcustom -2.177281e+03 2.721717e+03**

**paint\_colorgrey -4.699111e+03 7.689443e+02**

**paint\_colorpurple -1.018294e+04 3.200254e+03**

**paint\_colorred -5.290423e+03 -9.149167e+02**

**paint\_colorsilver -6.019991e+03 -2.515623e+03**

**paint\_colorwhite -2.437307e+03 1.145236e+03**

**>**

**> #3. Conduct appropriate analyses and give interpretations to determine**

**> #if your model is a good fit to the data in your primary data set.**

**>**

**> plot(cars$price,cars.out$fitted.values,**

**+ pch=19,**

**+ main="Actuals v. Fitteds, Restaurant Revenue")**

**> abline(0,1,col="red",lwd=3)**

**>**

**> plot(cars$price,cars2.out$fitted.values,**

**+ pch=19,**

**+ main="Actuals v. Fitteds, Restaurant Revenue")**

**> abline(0,1,col="red",lwd=3)**

**>**

**> plot(cars$price,cars3.out$fitted.values,**

**+ pch=19,**

**+ main="Actuals v. Fitteds, Restaurant Revenue")**

**Error in xy.coords(x, y, xlabel, ylabel, log) :**

**object 'cars3.out' not found**

**> abline(0,1,col="red",lwd=3)**

**>**

**> plot(cars$price,cars4.out$fitted.values,**

**+ pch=19,**

**+ main="Actuals v. Fitteds, Restaurant Revenue")**

**> abline(0,1,col="red",lwd=3)**

**>**

**> plot(cars$price,cars6.out$fitted.values,**

**+ pch=19,**

**+ main="Actuals v. Fitteds, Restaurant Revenue")**

**> abline(0,1,col="red",lwd=3)**

**>**

**>**

**> #4.) Assess your model’s conformance with the LINE assumptions of regression.**

**> #code from the professor**

**> # Combined Plot**

**> par(mfrow=c(2,2))**

**> # Linearity**

**> plot(cars$price,cars.out$fitted.values,**

**+ pch=19,**

**+ main="Actuals v. Fitteds, Restaurant Revenue")**

**> abline(0,1,col="red",lwd=3)**

**> # Normality**

**> qqnorm(cars.out$residuals,pch=19,**

**+ main="Normality Plot, Restaurant Revenue")**

**> qqline(cars.out$residuals,lwd=3,col="red")**

**> hist(cars.out$residuals,col="red",**

**+ main="Residuals, Restaurant Revenue",**

**+ probability = TRUE)**

**> curve(dnorm(x,mean(cars.out$residuals),**

**+ sd(cars.out$residuals)),**

**+ from=min(cars.out$residuals),**

**+ to=max(cars.out$residuals)**

**+ ,lwd=3,col="blue",add=TRUE)**

**> # Equality of Variances**

**> plot(cars.out$fitted.values,rstandard(cars.out),**

**+ pch=19,main="Equality of Variances, Restaurant Revenue")**

**> abline(0,0,lwd=3,col="red")**

**> par(mfrow=c(1,1))**

**>**

**> # I think the data looks okay I see there is an outlier. I don't there is anything that**

**> #stands out like what I saw in class. There is a bit of curve I am having trouble**

**> #applying the code from class.**

**>**

**>**

**>**

**>**

**>**

**> #4.) Assess your model’s conformance with the LINE assumptions of regression.**

**>**

**> #Linearity**

**> plot(crime12$reported\_crimes\_per\_million,simlin12$fitted.values,**

**+ pch=19,main="Crime Actual v. Fitted Values")**

**Error in plot(crime12$reported\_crimes\_per\_million, simlin12$fitted.values, :**

**object 'crime12' not found**

**> xlim=c(300,2000)ylim=c(300,2000)**

**Error: unexpected symbol in "xlim=c(300,2000)ylim"**

**> abline(0,1,col="red",lwd=3)**

**>**

**> #Does not pass linearity**

**> #I am having a hard time with making a logistic regression.**

**> #I need to over this with the professor.**

**>**

**> #Normality**

**> qqnorm(crime12$reported\_crimes\_per\_million,pch=19,main="Crime Normality Plot")**

**Error in qqnorm(crime12$reported\_crimes\_per\_million, pch = 19, main = "Crime Normality Plot") :**

**object 'crime12' not found**

**> qqline(simlin12$fitted.values,col="red",lwd=3)**

**Error in quantile(y, probs, names = FALSE, type = qtype, na.rm = TRUE) :**

**object 'simlin12' not found**

**> #There is a bit of curve on the right tail.**

**>**

**> #Equality of Variances**

**> plot(simlin12$fitted.values,scale(simlin12$residuals),**

**+ pch=19,main="Crime Standardized Residuals",ylim = c(-4,4))**

**Error in plot(simlin12$fitted.values, scale(simlin12$residuals), pch = 19, :**

**object 'simlin12' not found**

**> abline(0,0,col="red",lwd=3)**

**> #A bit of heteroscedasticity.**

**>**

**> #5.) Throckmorton P. Gildersleeve of Summerfield, Tennessee would like to sell**

**> #his 2011 Cadillac DTS pictured above. He says the vehicle is in "excellent"**

**> #condition and has 175,757 miles on the odometer.**

**> #Mr. Gildersleeve has not shared details of his Cadillac’s engine with you**

**> #because he thinks you know that all 2011 DTS cars had the same famous engine.**

**> #If you know nothing about Cadillac engines the point is easily researched online. Use R and your model to determine what price he should ask for the car.**

**> #Do you believe your pricing advice to the Great Gildersleeve is accurate and usable?**

**> #Give reasoning for your conclusion.**

**>**

**> #odometer ==175757**

**> #condition == 'excellent'**

**> #year == 2011**

**>**

**>**

**> throck = data.frame(odometer=175757,condition = 'excellent', year = '2011')**

**>**

**> predict(cars9.out,throck,interval="predict")**

**fit lwr upr**

**1 11562.44 1697.657 21427.22**

**>**

**>**

**>**

**>**

**> #price of the car could be around $11,662**

**> #I don't think the information is useful my model is bad as my squaring isn't working.**

**> #then I had to take inputs out of the model because of I did know the 'famous engine'**

**> #info. I would ask for more information.**

END R CONSOLE OUTPUT