Assignment 4

Chitresh Kumar

#if(!require("pacman")) install.packages("pacman")  
pacman::p\_load(tidyverse, reshape, gplots, RStata,haven)  
theme\_set(theme\_classic())  
options("RStata.StataPath")  
options("RStata.StataVersion" = 13)

getwd()

## [1] "C:/Users/chitr/OneDrive - The University of Texas at Dallas/Masters 1st sem/2nd Sem/Econometrics and Time Series/Assignments/Assignment 4"

df<- read\_dta(file = "cocaine.dta")  
df1 <- read\_dta(file = "br2.dta")

head(df)

## # A tibble: 6 x 4  
## price quant qual trend  
## <dbl> <dbl> <dbl> <dbl>  
## 1 57.5 1000 62.5 1  
## 2 77.2 454. 62.5 1  
## 3 77.6 28.4 19 1  
## 4 84.7 14.2 19 1  
## 5 77.6 7.09 19 1  
## 6 91.7 3.54 19 1

head(df1)

## # A tibble: 6 x 10  
## price sqft bedrooms baths age owner pool traditional fireplace waterfront  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 66500 741 1 1 18 1 1 1 1 0  
## 2 66000 741 1 1 18 0 1 1 0 0  
## 3 68500 790 1 1 18 1 0 1 1 0  
## 4 102000 2783 2 2 18 1 0 1 1 0  
## 5 54000 1165 2 1 35 0 0 1 0 0  
## 6 143000 2331 2 2 25 1 0 1 1 0

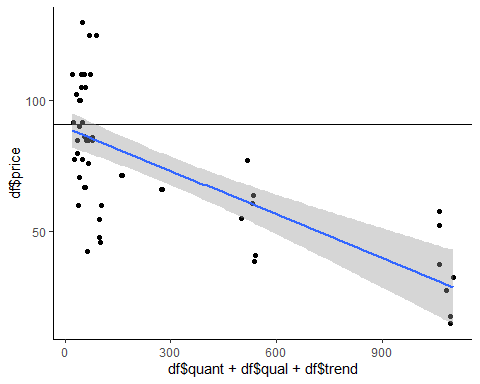
lm1<- lm(df$price~ df$quant + df$qual + df$trend,data=df)  
summary(lm1)

##   
## Call:  
## lm(formula = df$price ~ df$quant + df$qual + df$trend, data = df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -43.479 -12.014 -3.743 13.969 43.753   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 90.84669 8.58025 10.588 1.39e-14 \*\*\*  
## df$quant -0.05997 0.01018 -5.892 2.85e-07 \*\*\*  
## df$qual 0.11621 0.20326 0.572 0.5700   
## df$trend -2.35458 1.38612 -1.699 0.0954 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 20.06 on 52 degrees of freedom  
## Multiple R-squared: 0.5097, Adjusted R-squared: 0.4814   
## F-statistic: 18.02 on 3 and 52 DF, p-value: 3.806e-08

mean(df$price)

## [1] 75.4073

ggplot(df,aes(y=df$price,x=df$quant + df$qual + df$trend))+  
 geom\_point()+  
 geom\_hline(yintercept=90.84669)+  
 geom\_smooth(method="lm")



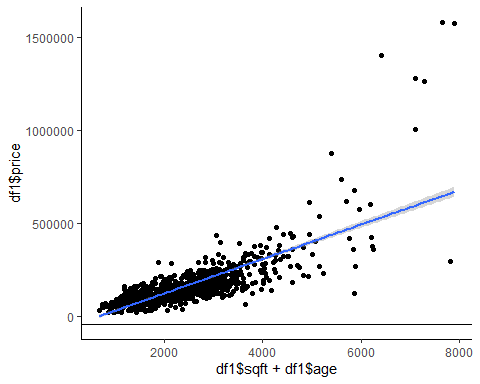
lm2<- lm(df1$price~ df1$sqft + df1$age ,data=df1)  
summary(lm2)

##   
## Call:  
## lm(formula = df1$price ~ df1$sqft + df1$age, data = df1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -358116 -33259 -6111 27242 936754   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -41947.696 6989.636 -6.001 2.67e-09 \*\*\*  
## df1$sqft 90.970 2.403 37.855 < 2e-16 \*\*\*  
## df1$age -755.041 140.894 -5.359 1.02e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 78810 on 1077 degrees of freedom  
## Multiple R-squared: 0.5896, Adjusted R-squared: 0.5888   
## F-statistic: 773.6 on 2 and 1077 DF, p-value: < 2.2e-16

mean(df$price)

## [1] 75.4073

ggplot(df1,aes(y=df1$price,x=df1$sqft + df1$age))+  
 geom\_point()+  
 geom\_hline(yintercept=-41947.696)+  
 geom\_smooth(method="lm")



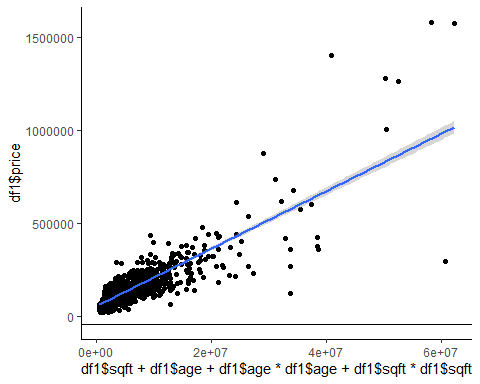
lm2<- lm(df1$price~ df1$sqft + df1$age + df1$age\*df1$age +df1$sqft\*df1$sqft ,data=df1)  
summary(lm2)

##   
## Call:  
## lm(formula = df1$price ~ df1$sqft + df1$age + df1$age \* df1$age +   
## df1$sqft \* df1$sqft, data = df1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -358116 -33259 -6111 27242 936754   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -41947.696 6989.636 -6.001 2.67e-09 \*\*\*  
## df1$sqft 90.970 2.403 37.855 < 2e-16 \*\*\*  
## df1$age -755.041 140.894 -5.359 1.02e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 78810 on 1077 degrees of freedom  
## Multiple R-squared: 0.5896, Adjusted R-squared: 0.5888   
## F-statistic: 773.6 on 2 and 1077 DF, p-value: < 2.2e-16

mean(df$price)

## [1] 75.4073

ggplot(df1,aes(y=df1$price,x=df1$sqft + df1$age + df1$age\*df1$age +df1$sqft\*df1$sqft))+  
 geom\_point()+  
 geom\_hline(yintercept=-41947.696)+  
 geom\_smooth(method="lm")



options(scipen = 999)  
lm2<- lm(df1$price~ df1$sqft + df1$age + poly(df1$age,2) +poly(df1$sqft,2) +df1$sqft\*df1$age ,data=df1)  
summary(lm2)

##   
## Call:  
## lm(formula = df1$price ~ df1$sqft + df1$age + poly(df1$age, 2) +   
## poly(df1$sqft, 2) + df1$sqft \* df1$age, data = df1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -796617 -21537 -439 17825 623609   
##   
## Coefficients: (2 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -76765.0183 7820.4123 -9.816 < 0.0000000000000002 \*\*\*  
## df1$sqft 108.2280 2.9664 36.485 < 0.0000000000000002 \*\*\*  
## df1$age 1023.8743 282.9441 3.619 0.00031 \*\*\*  
## poly(df1$age, 2)1 NA NA NA NA   
## poly(df1$age, 2)2 333720.7684 62148.2485 5.370 0.000000096609546031 \*\*\*  
## poly(df1$sqft, 2)1 NA NA NA NA   
## poly(df1$sqft, 2)2 1477502.0033 62772.5929 23.537 < 0.0000000000000002 \*\*\*  
## df1$sqft:df1$age -0.9306 0.1124 -8.277 0.000000000000000372 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 61470 on 1074 degrees of freedom  
## Multiple R-squared: 0.751, Adjusted R-squared: 0.7499   
## F-statistic: 648 on 5 and 1074 DF, p-value: < 0.00000000000000022

mean(df$price)

## [1] 75.4073

ggplot(df1,aes(y=df1$price,x=df1$sqft + df1$age + df1$age\*df1$age +df1$sqft\*df1$sqft +df1$sqft\*df1$age))+  
 geom\_point()+  
 geom\_hline(yintercept=-41947.696)+  
 geom\_smooth(method="lm")

