Advanced Linear Regression

library(readxl)  
library("car")

## Loading required package: carData

library(lmtest)

## Loading required package: zoo

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

library("ggpubr")

## Warning: package 'ggpubr' was built under R version 4.0.3

## Loading required package: ggplot2

library(stargazer)

## Warning: package 'stargazer' was built under R version 4.0.3

##   
## Please cite as:

## Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.

## R package version 5.2.2. https://CRAN.R-project.org/package=stargazer

library(data.table)

## Warning: package 'data.table' was built under R version 4.0.3

setwd("C:/Users/surya/Downloads")  
  
hs <- read\_excel("HuntersGreenHomeSales.xlsx", sheet = 'Data')  
  
#NA values column wise  
sapply(hs, function(x) sum(is.na(x)))

## slnoskm Status Address   
## 0 0 0   
## Beds bathsfull bathshalf   
## 0 0 0   
## bathstotal sqft garages   
## 0 0 4   
## Roof lotsqft yrblt   
## 0 0 0   
## Pool spa subdivn   
## 0 313 0   
## adom\_agentdaysonmarket cdom\_cumuldaysmls listprice   
## 0 0 0   
## lppersqft PendingDate pricesold   
## 0 0 0   
## sppersqft datesold splsale   
## 0 0 0

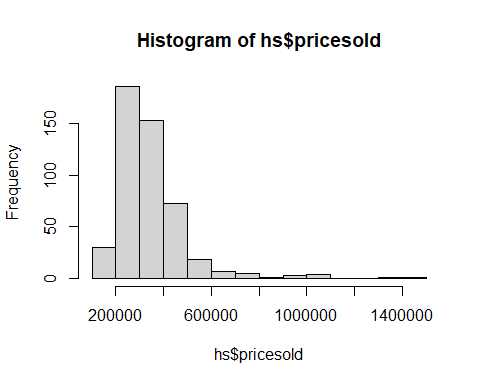
str(hs)

## tibble [482 x 24] (S3: tbl\_df/tbl/data.frame)  
## $ slnoskm : num [1:482] 1 2 3 4 5 6 7 8 9 10 ...  
## $ Status : chr [1:482] "Sold" "Sold" "Sold" "Sold" ...  
## $ Address : chr [1:482] "17711 ESPRIT DR" "17769 ESPRIT DR" "17622 ESPRIT DR" "18111 ASHTON PARK WAY" ...  
## $ Beds : num [1:482] 3 2 3 3 4 4 2 3 3 3 ...  
## $ bathsfull : num [1:482] 2 2 2 2 2 2 2 2 2 2 ...  
## $ bathshalf : num [1:482] 0 0 0 0 1 1 0 0 0 0 ...  
## $ bathstotal : num [1:482] 2 2 2 2 3 3 2 2 2 2 ...  
## $ sqft : num [1:482] 1695 1305 1420 1598 2302 ...  
## $ garages : num [1:482] 2 2 2 2 2 2 2 2 2 2 ...  
## $ Roof : chr [1:482] "Shingle" "Shingle" "Shingle" "Shingle" ...  
## $ lotsqft : num [1:482] 8694 6540 4845 5750 5390 ...  
## $ yrblt : num [1:482] 1992 1992 1991 1996 1992 ...  
## $ Pool : chr [1:482] "Community" "None" "Community" "Community" ...  
## $ spa : logi [1:482] NA NA NA NA NA TRUE ...  
## $ subdivn : chr [1:482] "HUNTERS GREEN PH 1B 02" "HUNTERS GREEN PH 1A" "HUNTERS GREEN PRCL 17B PH 1A" "HUNTERS GREEN" ...  
## $ adom\_agentdaysonmarket: num [1:482] 2 133 4 44 14 207 69 128 57 0 ...  
## $ cdom\_cumuldaysmls : num [1:482] 2 133 4 44 14 207 69 128 57 0 ...  
## $ listprice : num [1:482] 150000 179000 165000 184000 180000 ...  
## $ lppersqft : num [1:482] 88.5 137.2 116.2 115.1 78.2 ...  
## $ PendingDate : POSIXct[1:482], format: "2017-07-12" "2015-07-02" ...  
## $ pricesold : num [1:482] 141500 160000 161500 170000 170000 ...  
## $ sppersqft : num [1:482] 83.5 122.6 113.7 106.4 73.8 ...  
## $ datesold : POSIXct[1:482], format: "2018-10-06" "2015-07-30" ...  
## $ splsale : chr [1:482] "Short Sale" "None" "None" "None" ...

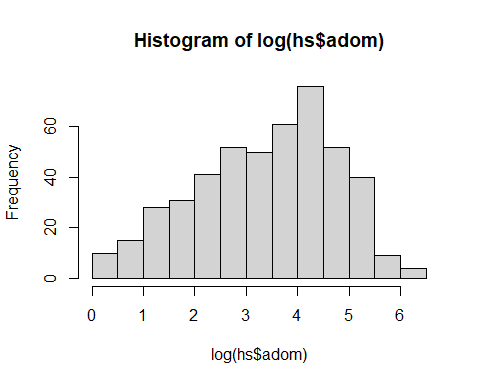
#Pre-processing  
#Converting categoricals into binary factor levels  
names(hs) <- tolower(colnames(hs))  
setnames(hs, c('adom\_agentdaysonmarket', 'cdom\_cumuldaysmls'), c('adom', 'cdom'))  
hs$spl\_sale <- ifelse(hs$splsale == 'Short Sale' | hs$splsale == 'Bank Owned/REO' |  
 hs$splsale == 'Auction, Bank Owned/REO', 1, 0)  
hs$baths\_total <- hs$bathsfull + hs$bathshalf\*0.5  
#hs$shingle\_roof <- ifelse(hs$roof == 'Shake, Shingle' | hs$roof == 'Shingle' | hs$roof == 'Shingle, Tile', 1, 0)  
hs$tile\_roof <- ifelse(hs$roof == 'Concrete, Tile' | hs$roof == 'Slate' | hs$roof == 'Slate, Tile' | hs$roof == 'Tile', 1, 0)  
hs$private\_pool <- ifelse(hs$pool == 'Private' | hs$pool == 'Private, Community', 1, 0)  
hs$community\_pool <- ifelse(hs$pool == 'Community' | hs$pool == 'Private, Community', 1, 0)  
  
hs$pendingdate <- as.numeric(format(hs$pendingdate,'%Y'))  
hs$datesold <- as.numeric(format(hs$datesold,'%Y'))  
str(hs)

## tibble [482 x 29] (S3: tbl\_df/tbl/data.frame)  
## $ slnoskm : num [1:482] 1 2 3 4 5 6 7 8 9 10 ...  
## $ status : chr [1:482] "Sold" "Sold" "Sold" "Sold" ...  
## $ address : chr [1:482] "17711 ESPRIT DR" "17769 ESPRIT DR" "17622 ESPRIT DR" "18111 ASHTON PARK WAY" ...  
## $ beds : num [1:482] 3 2 3 3 4 4 2 3 3 3 ...  
## $ bathsfull : num [1:482] 2 2 2 2 2 2 2 2 2 2 ...  
## $ bathshalf : num [1:482] 0 0 0 0 1 1 0 0 0 0 ...  
## $ bathstotal : num [1:482] 2 2 2 2 3 3 2 2 2 2 ...  
## $ sqft : num [1:482] 1695 1305 1420 1598 2302 ...  
## $ garages : num [1:482] 2 2 2 2 2 2 2 2 2 2 ...  
## $ roof : chr [1:482] "Shingle" "Shingle" "Shingle" "Shingle" ...  
## $ lotsqft : num [1:482] 8694 6540 4845 5750 5390 ...  
## $ yrblt : num [1:482] 1992 1992 1991 1996 1992 ...  
## $ pool : chr [1:482] "Community" "None" "Community" "Community" ...  
## $ spa : logi [1:482] NA NA NA NA NA TRUE ...  
## $ subdivn : chr [1:482] "HUNTERS GREEN PH 1B 02" "HUNTERS GREEN PH 1A" "HUNTERS GREEN PRCL 17B PH 1A" "HUNTERS GREEN" ...  
## $ adom : num [1:482] 2 133 4 44 14 207 69 128 57 0 ...  
## $ cdom : num [1:482] 2 133 4 44 14 207 69 128 57 0 ...  
## $ listprice : num [1:482] 150000 179000 165000 184000 180000 ...  
## $ lppersqft : num [1:482] 88.5 137.2 116.2 115.1 78.2 ...  
## $ pendingdate : num [1:482] 2017 2015 2015 2016 2015 ...  
## $ pricesold : num [1:482] 141500 160000 161500 170000 170000 ...  
## $ sppersqft : num [1:482] 83.5 122.6 113.7 106.4 73.8 ...  
## $ datesold : num [1:482] 2018 2015 2015 2016 2016 ...  
## $ splsale : chr [1:482] "Short Sale" "None" "None" "None" ...  
## $ spl\_sale : num [1:482] 1 0 0 0 1 1 0 0 1 0 ...  
## $ baths\_total : num [1:482] 2 2 2 2 2.5 2.5 2 2 2 2 ...  
## $ tile\_roof : num [1:482] 0 0 0 0 0 1 0 0 0 1 ...  
## $ private\_pool : num [1:482] 0 0 0 0 1 1 0 0 0 0 ...  
## $ community\_pool: num [1:482] 1 0 1 1 0 0 0 1 1 1 ...

#Removing NA columns  
hs <- hs[-c(1:3, 5:7, 10, 13:15, 24)]  
hist(hs$pricesold)



hist(log(hs$adom))



#Checking correlations  
library(PerformanceAnalytics)

## Warning: package 'PerformanceAnalytics' was built under R version 4.0.3

## Loading required package: xts

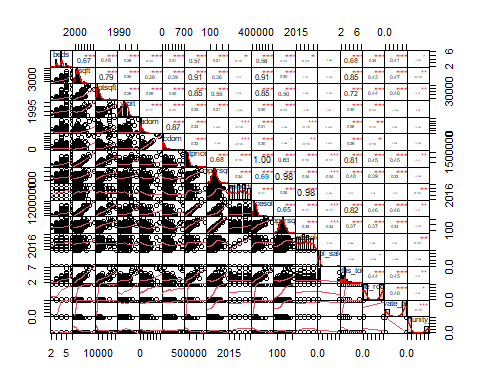
##   
## Attaching package: 'xts'

## The following objects are masked from 'package:data.table':  
##   
## first, last

##   
## Attaching package: 'PerformanceAnalytics'

## The following object is masked from 'package:graphics':  
##   
## legend

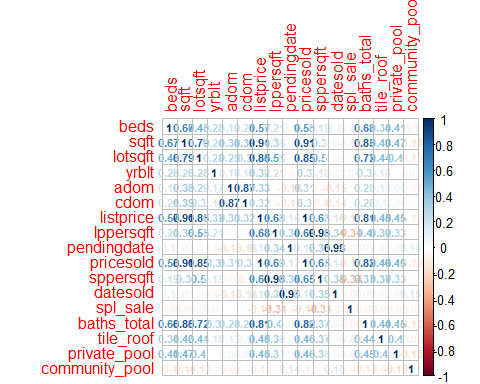
hs\_corr <- hs[, -c(3)]  
chart.Correlation(hs\_corr)



library(corrplot)

## corrplot 0.84 loaded

hs\_corplot <- cor(hs\_corr)  
corrplot(hs\_corplot, method = "number", number.cex= 12/ncol(hs))



#1st DV Regression models  
names(hs)

## [1] "beds" "sqft" "garages" "lotsqft"   
## [5] "yrblt" "adom" "cdom" "listprice"   
## [9] "lppersqft" "pendingdate" "pricesold" "sppersqft"   
## [13] "datesold" "spl\_sale" "baths\_total" "tile\_roof"   
## [17] "private\_pool" "community\_pool"

p1 <- lm(pricesold ~ pendingdate + spl\_sale, data = hs)  
summary(p1)

##   
## Call:  
## lm(formula = pricesold ~ pendingdate + spl\_sale, data = hs)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -178648 -95218 -24148 42194 1064395   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -33849515 9948011 -3.403 0.000723 \*\*\*  
## pendingdate 16957 4932 3.438 0.000637 \*\*\*  
## spl\_sale -103296 28462 -3.629 0.000315 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 148100 on 479 degrees of freedom  
## Multiple R-squared: 0.05379, Adjusted R-squared: 0.04984   
## F-statistic: 13.62 on 2 and 479 DF, p-value: 1.772e-06

p2 <- lm(pricesold ~ beds + sqft + garages + yrblt + spl\_sale + baths\_total   
 + tile\_roof + private\_pool + community\_pool, data = hs)  
summary(p2)

##   
## Call:  
## lm(formula = pricesold ~ beds + sqft + garages + yrblt + spl\_sale +   
## baths\_total + tile\_roof + private\_pool + community\_pool,   
## data = hs)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -157077 -30522 -115 28492 349599   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -9.306e+06 2.238e+06 -4.158 3.82e-05 \*\*\*  
## beds -2.423e+04 5.380e+03 -4.504 8.45e-06 \*\*\*  
## sqft 1.409e+02 6.509e+00 21.651 < 2e-16 \*\*\*  
## garages 1.121e+04 6.657e+03 1.684 0.09283 .   
## yrblt 4.636e+03 1.126e+03 4.119 4.50e-05 \*\*\*  
## spl\_sale -6.465e+04 1.136e+04 -5.690 2.24e-08 \*\*\*  
## baths\_total 3.080e+04 7.614e+03 4.045 6.11e-05 \*\*\*  
## tile\_roof 1.781e+04 6.136e+03 2.902 0.00388 \*\*   
## private\_pool 1.172e+04 7.171e+03 1.634 0.10298   
## community\_pool 9.573e+02 5.937e+03 0.161 0.87196   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 57560 on 468 degrees of freedom  
## (4 observations deleted due to missingness)  
## Multiple R-squared: 0.86, Adjusted R-squared: 0.8573   
## F-statistic: 319.3 on 9 and 468 DF, p-value: < 2.2e-16

p3 <- lm(pricesold ~ beds + sqft + garages + yrblt + spl\_sale + baths\_total   
 + tile\_roof + private\_pool + community\_pool + pendingdate, data = hs)  
summary(p3)

##   
## Call:  
## lm(formula = pricesold ~ beds + sqft + garages + yrblt + spl\_sale +   
## baths\_total + tile\_roof + private\_pool + community\_pool +   
## pendingdate, data = hs)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -154124 -26071 59 30413 374899   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -3.426e+07 4.573e+06 -7.493 3.41e-13 \*\*\*  
## beds -2.678e+04 5.194e+03 -5.157 3.72e-07 \*\*\*  
## sqft 1.409e+02 6.264e+00 22.494 < 2e-16 \*\*\*  
## garages 9.867e+03 6.411e+03 1.539 0.12445   
## yrblt 5.342e+03 1.089e+03 4.903 1.30e-06 \*\*\*  
## spl\_sale -6.168e+04 1.094e+04 -5.636 3.01e-08 \*\*\*  
## baths\_total 3.069e+04 7.327e+03 4.188 3.36e-05 \*\*\*  
## tile\_roof 1.690e+04 5.907e+03 2.861 0.00441 \*\*   
## private\_pool 1.176e+04 6.901e+03 1.704 0.08907 .   
## community\_pool -4.594e+03 5.783e+03 -0.794 0.42741   
## pendingdate 1.168e+04 1.888e+03 6.187 1.34e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 55400 on 467 degrees of freedom  
## (4 observations deleted due to missingness)  
## Multiple R-squared: 0.8706, Adjusted R-squared: 0.8678   
## F-statistic: 314.1 on 10 and 467 DF, p-value: < 2.2e-16

#1st DV Stargazer  
stargazer(p1, p2, p3, type='text', single.row = TRUE)

##   
## ===========================================================================================================================  
## Dependent variable:   
## -------------------------------------------------------------------------------------------------------  
## pricesold   
## (1) (2) (3)   
## ---------------------------------------------------------------------------------------------------------------------------  
## pendingdate 16,957.440\*\*\* (4,932.184) 11,683.520\*\*\* (1,888.458)   
## beds -24,228.040\*\*\* (5,379.690) -26,784.360\*\*\* (5,193.908)   
## sqft 140.920\*\*\* (6.509) 140.901\*\*\* (6.264)   
## garages 11,211.410\* (6,657.231) 9,866.753 (6,410.645)   
## yrblt 4,636.479\*\*\* (1,125.710) 5,341.608\*\*\* (1,089.369)   
## spl\_sale -103,296.400\*\*\* (28,461.800) -64,647.050\*\*\* (11,360.840) -61,683.560\*\*\* (10,944.230)   
## baths\_total 30,799.080\*\*\* (7,613.655) 30,688.760\*\*\* (7,327.450)   
## tile\_roof 17,808.110\*\*\* (6,135.807) 16,900.590\*\*\* (5,906.960)   
## private\_pool 11,715.010 (7,170.511) 11,758.430\* (6,900.947)   
## community\_pool 957.324 (5,936.581) -4,593.883 (5,783.429)   
## Constant -33,849,515.000\*\*\* (9,948,011.000) -9,306,190.000\*\*\* (2,238,211.000) -34,261,679.000\*\*\* (4,572,796.000)  
## ---------------------------------------------------------------------------------------------------------------------------  
## Observations 482 478 478   
## R2 0.054 0.860 0.871   
## Adjusted R2 0.050 0.857 0.868   
## Residual Std. Error 148,084.200 (df = 479) 57,564.820 (df = 468) 55,400.740 (df = 467)   
## F Statistic 13.616\*\*\* (df = 2; 479) 319.328\*\*\* (df = 9; 468) 314.114\*\*\* (df = 10; 467)   
## ===========================================================================================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

stargazer(p1, p2, p3, type='text', ci=TRUE, ci.level=0.95, single.row = TRUE)

##   
## ===================================================================================================================================================================================  
## Dependent variable:   
## ---------------------------------------------------------------------------------------------------------------------------------------------------------------  
## pricesold   
## (1) (2) (3)   
## -----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------  
## pendingdate 16,957.440\*\*\* (7,290.540, 26,624.350) 11,683.520\*\*\* (7,982.208, 15,384.830)   
## beds -24,228.040\*\*\* (-34,772.040, -13,684.040) -26,784.360\*\*\* (-36,964.240, -16,604.490)   
## sqft 140.920\*\*\* (128.163, 153.676) 140.901\*\*\* (128.624, 153.178)   
## garages 11,211.410\* (-1,836.527, 24,259.340) 9,866.753 (-2,697.879, 22,431.390)   
## yrblt 4,636.479\*\*\* (2,430.128, 6,842.830) 5,341.608\*\*\* (3,206.484, 7,476.732)   
## spl\_sale -103,296.400\*\*\* (-159,080.500, -47,512.340) -64,647.050\*\*\* (-86,913.890, -42,380.210) -61,683.560\*\*\* (-83,133.860, -40,233.260)   
## baths\_total 30,799.080\*\*\* (15,876.590, 45,721.570) 30,688.760\*\*\* (16,327.220, 45,050.300)   
## tile\_roof 17,808.110\*\*\* (5,782.146, 29,834.070) 16,900.590\*\*\* (5,323.160, 28,478.020)   
## private\_pool 11,715.010 (-2,338.930, 25,768.960) 11,758.430\* (-1,767.174, 25,284.040)   
## community\_pool 957.324 (-10,678.160, 12,592.810) -4,593.883 (-15,929.190, 6,741.430)   
## Constant -33,849,515.000\*\*\* (-53,347,258.000, -14,351,772.000) -9,306,190.000\*\*\* (-13,693,004.000, -4,919,377.000) -34,261,679.000\*\*\* (-43,224,194.000, -25,299,163.000)  
## -----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------  
## Observations 482 478 478   
## R2 0.054 0.860 0.871   
## Adjusted R2 0.050 0.857 0.868   
## Residual Std. Error 148,084.200 (df = 479) 57,564.820 (df = 468) 55,400.740 (df = 467)   
## F Statistic 13.616\*\*\* (df = 2; 479) 319.328\*\*\* (df = 9; 468) 314.114\*\*\* (df = 10; 467)   
## ===================================================================================================================================================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

#2nd DV Regression models  
a1 <- lm(adom ~ pendingdate, data = hs)  
summary(a1)

##   
## Call:  
## lm(formula = adom ~ pendingdate, data = hs)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -79.67 -47.61 -26.34 22.70 553.33   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 19054.871 5317.568 3.583 0.000374 \*\*\*  
## pendingdate -9.417 2.636 -3.572 0.000390 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 79.43 on 480 degrees of freedom  
## Multiple R-squared: 0.02589, Adjusted R-squared: 0.02386   
## F-statistic: 12.76 on 1 and 480 DF, p-value: 0.0003903

a2 <- lm(adom ~ yrblt + pendingdate + lppersqft + spl\_sale, data = hs)  
summary(a2)

##   
## Call:  
## lm(formula = adom ~ yrblt + pendingdate + lppersqft + spl\_sale,   
## data = hs)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -108.09 -46.49 -24.21 19.92 519.96   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 8832.52170 6691.28549 1.320 0.187467   
## yrblt 4.93781 1.44528 3.417 0.000689 \*\*\*  
## pendingdate -9.23382 2.80087 -3.297 0.001051 \*\*   
## lppersqft 0.05017 0.19109 0.263 0.793031   
## spl\_sale -5.86950 16.12239 -0.364 0.715975   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 78.59 on 477 degrees of freedom  
## Multiple R-squared: 0.05233, Adjusted R-squared: 0.04438   
## F-statistic: 6.585 on 4 and 477 DF, p-value: 3.65e-05

a3 <- lm(adom ~ yrblt + pendingdate + lppersqft + spl\_sale + baths\_total + private\_pool + community\_pool, data = hs)  
summary(a3)

##   
## Call:  
## lm(formula = adom ~ yrblt + pendingdate + lppersqft + spl\_sale +   
## baths\_total + private\_pool + community\_pool, data = hs)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -217.04 -46.25 -19.83 22.97 456.97   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 11461.7655 6639.1706 1.726 0.08493 .   
## yrblt 2.7513 1.4674 1.875 0.06142 .   
## pendingdate -8.3960 2.7430 -3.061 0.00233 \*\*   
## lppersqft -0.3426 0.2020 -1.696 0.09047 .   
## spl\_sale -8.8552 15.6506 -0.566 0.57179   
## baths\_total 35.5468 5.9077 6.017 3.56e-09 \*\*\*  
## private\_pool -8.1692 8.8837 -0.920 0.35827   
## community\_pool -4.2739 7.8468 -0.545 0.58624   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 75.7 on 474 degrees of freedom  
## Multiple R-squared: 0.1263, Adjusted R-squared: 0.1134   
## F-statistic: 9.79 on 7 and 474 DF, p-value: 2.039e-11

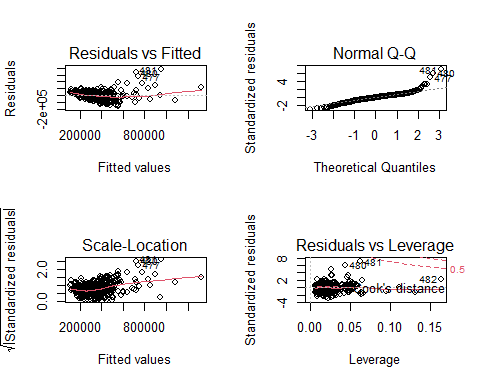
#2nd DV Stargazer  
stargazer(a1, a2, a3, type='text', single.row = TRUE)

##   
## ============================================================================================  
## Dependent variable:   
## ------------------------------------------------------------------------  
## adom   
## (1) (2) (3)   
## --------------------------------------------------------------------------------------------  
## yrblt 4.938\*\*\* (1.445) 2.751\* (1.467)   
## pendingdate -9.417\*\*\* (2.636) -9.234\*\*\* (2.801) -8.396\*\*\* (2.743)   
## lppersqft 0.050 (0.191) -0.343\* (0.202)   
## spl\_sale -5.870 (16.122) -8.855 (15.651)   
## baths\_total 35.547\*\*\* (5.908)   
## private\_pool -8.169 (8.884)   
## community\_pool -4.274 (7.847)   
## Constant 19,054.870\*\*\* (5,317.568) 8,832.522 (6,691.285) 11,461.770\* (6,639.171)  
## --------------------------------------------------------------------------------------------  
## Observations 482 482 482   
## R2 0.026 0.052 0.126   
## Adjusted R2 0.024 0.044 0.113   
## Residual Std. Error 79.427 (df = 480) 78.588 (df = 477) 75.696 (df = 474)   
## F Statistic 12.758\*\*\* (df = 1; 480) 6.585\*\*\* (df = 4; 477) 9.790\*\*\* (df = 7; 474)   
## ============================================================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

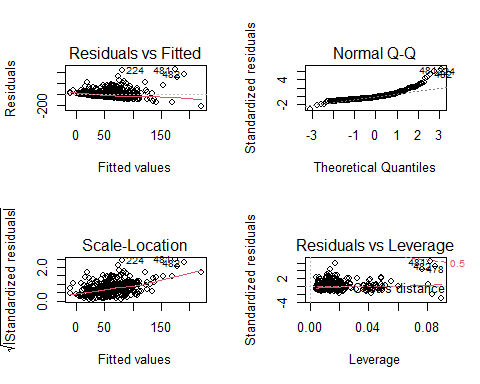
stargazer(a1, a2, a3, type='text', ci=TRUE, ci.level=0.95, single.row = TRUE)

##   
## =================================================================================================================================  
## Dependent variable:   
## -------------------------------------------------------------------------------------------------------------  
## adom   
## (1) (2) (3)   
## ---------------------------------------------------------------------------------------------------------------------------------  
## yrblt 4.938\*\*\* (2.105, 7.771) 2.751\* (-0.125, 5.627)   
## pendingdate -9.417\*\*\* (-14.584, -4.250) -9.234\*\*\* (-14.723, -3.744) -8.396\*\*\* (-13.772, -3.020)   
## lppersqft 0.050 (-0.324, 0.425) -0.343\* (-0.738, 0.053)   
## spl\_sale -5.870 (-37.469, 25.730) -8.855 (-39.530, 21.819)   
## baths\_total 35.547\*\*\* (23.968, 47.126)   
## private\_pool -8.169 (-25.581, 9.243)   
## community\_pool -4.274 (-19.653, 11.106)   
## Constant 19,054.870\*\*\* (8,632.630, 29,477.110) 8,832.522 (-4,282.157, 21,947.200) 11,461.770\* (-1,550.770, 24,474.300)  
## ---------------------------------------------------------------------------------------------------------------------------------  
## Observations 482 482 482   
## R2 0.026 0.052 0.126   
## Adjusted R2 0.024 0.044 0.113   
## Residual Std. Error 79.427 (df = 480) 78.588 (df = 477) 75.696 (df = 474)   
## F Statistic 12.758\*\*\* (df = 1; 480) 6.585\*\*\* (df = 4; 477) 9.790\*\*\* (df = 7; 474)   
## =================================================================================================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

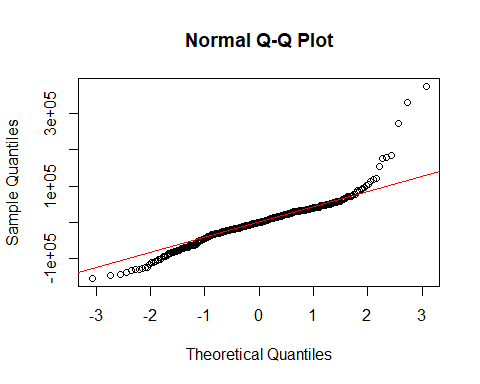
#Assumptions tests  
#Plots  
par(mfrow = c(2, 2))  
plot(p3)



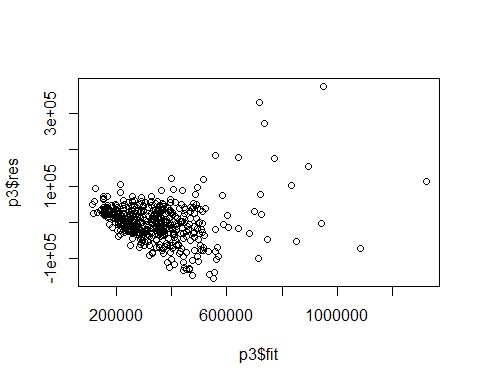
par(mfrow=c(1,1))  
  
par(mfrow = c(2, 2))  
plot(a3)



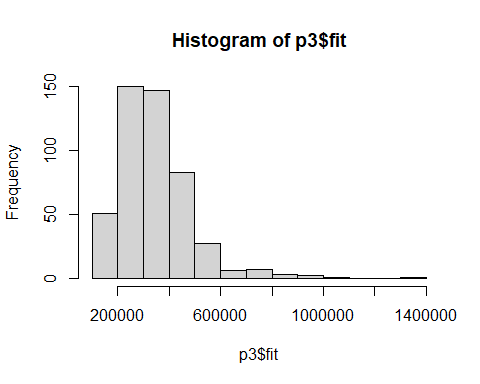
par(mfrow=c(1,1))  
  
qqnorm(p3$res)  
qqline(p3$res, col = 'red')



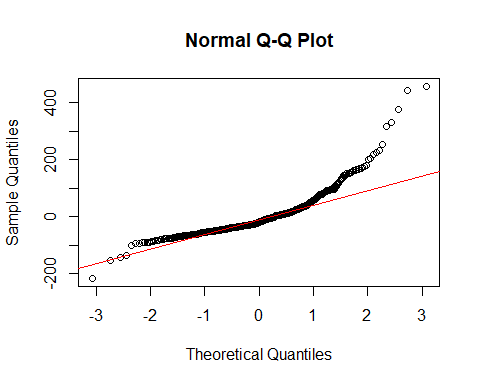
plot(p3$res ~ p3$fit)



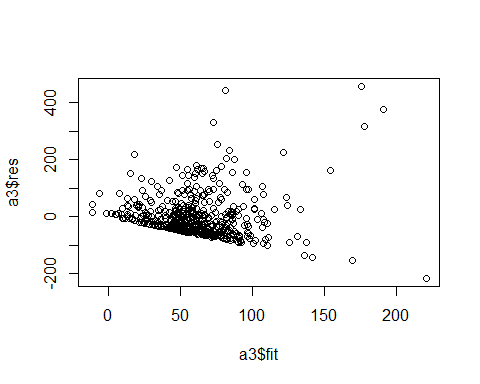
hist(p3$fit)



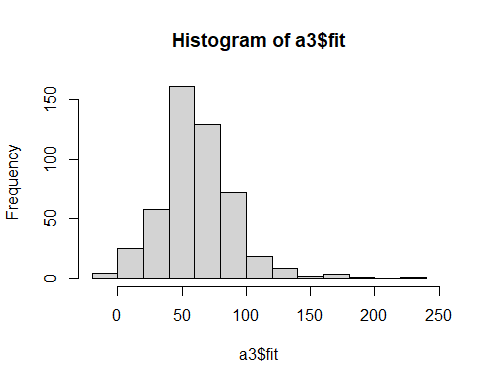
qqnorm(a3$res)  
qqline(a3$res, col = 'red')



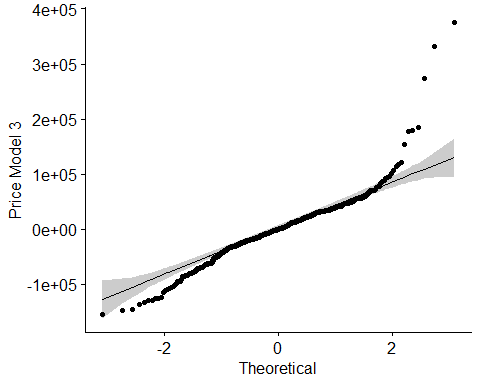
plot(a3$res ~ a3$fit)



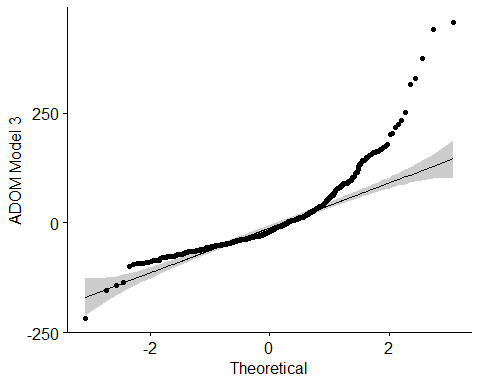
hist(a3$fit)



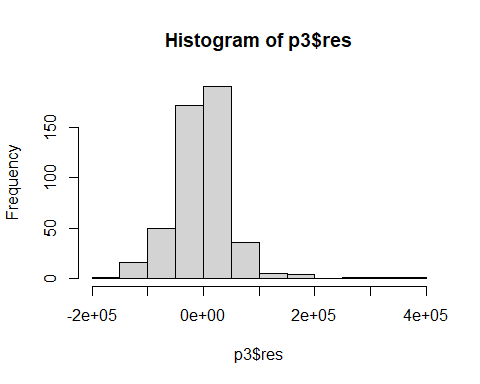
ggqqplot(p3$res, ylab = "Price Model 3")



ggqqplot(a3$res, ylab = "ADOM Model 3")



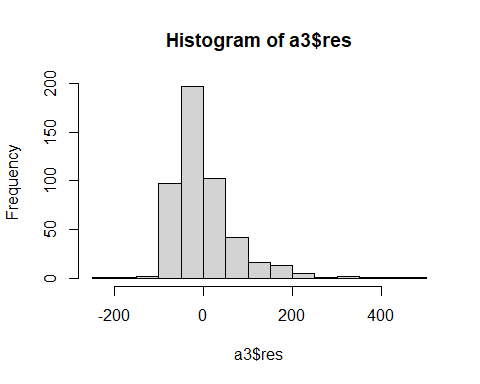
#Normality  
#Shapiro-Wilk's Test  
hist(p3$res)



shapiro.test(p3$res)

##   
## Shapiro-Wilk normality test  
##   
## data: p3$res  
## W = 0.90232, p-value < 2.2e-16

hist(a3$res)



shapiro.test(a3$res)

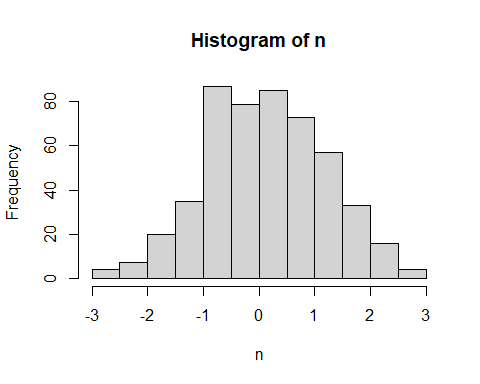
##   
## Shapiro-Wilk normality test  
##   
## data: a3$res  
## W = 0.83243, p-value < 2.2e-16

#Kolmogorov-Smirnov Test  
n <- rnorm(500)  
hist(n)  
ks.test(n, p3$res)

## Warning in ks.test(n, p3$res): p-value will be approximate in the presence of  
## ties

##   
## Two-sample Kolmogorov-Smirnov test  
##   
## data: n and p3$res  
## D = 0.5, p-value < 2.2e-16  
## alternative hypothesis: two-sided

hist(n)



ks.test(n, a3$res)

##   
## Two-sample Kolmogorov-Smirnov test  
##   
## data: n and a3$res  
## D = 0.60166, p-value < 2.2e-16  
## alternative hypothesis: two-sided

#Homoscedasticity  
#Bartlett's Test  
bartlett.test(list(p3$res, p3$fit))

##   
## Bartlett test of homogeneity of variances  
##   
## data: list(p3$res, p3$fit)  
## Bartlett's K-squared = 379.74, df = 1, p-value < 2.2e-16

bartlett.test(list(a3$res, a3$fit))

##   
## Bartlett test of homogeneity of variances  
##   
## data: list(a3$res, a3$fit)  
## Bartlett's K-squared = 392.91, df = 1, p-value < 2.2e-16

#Levene's Test  
leveneTest(p3$res, p3$fit, center=mean)

## Warning in leveneTest.default(p3$res, p3$fit, center = mean): p3$fit coerced to  
## factor.

## Warning in anova.lm(lm(resp ~ group)): ANOVA F-tests on an essentially perfect  
## fit are unreliable

## Levene's Test for Homogeneity of Variance (center = mean)  
## Df F value Pr(>F)   
## group 470 4.1755e+26 < 2.2e-16 \*\*\*  
## 7   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

leveneTest(a3$res, a3$fit, center=mean)

## Warning in leveneTest.default(a3$res, a3$fit, center = mean): a3$fit coerced to  
## factor.

## Levene's Test for Homogeneity of Variance (center = mean)  
## Df F value Pr(>F)  
## group 481   
## 0

#Breusch-Pagan Test  
bptest(p3)

##   
## studentized Breusch-Pagan test  
##   
## data: p3  
## BP = 99.524, df = 10, p-value < 2.2e-16

bptest(a3)

##   
## studentized Breusch-Pagan test  
##   
## data: a3  
## BP = 72.683, df = 7, p-value = 4.232e-13

#Multicollinearity  
vif(p3)

## beds sqft garages yrblt spl\_sale   
## 2.075706 4.173145 1.947956 1.228472 1.028684   
## baths\_total tile\_roof private\_pool community\_pool pendingdate   
## 4.338559 1.358516 1.536042 1.083330 1.048908

vif(a3)

## yrblt pendingdate lppersqft spl\_sale baths\_total   
## 1.203863 1.191815 1.640892 1.165093 1.515354   
## private\_pool community\_pool   
## 1.373933 1.080759

#Autocorrelation (Independence)  
#Durbin-Watson Test  
dwtest(p3)

##   
## Durbin-Watson test  
##   
## data: p3  
## DW = 1.556, p-value = 3.798e-07  
## alternative hypothesis: true autocorrelation is greater than 0

dwtest(a3)

##   
## Durbin-Watson test  
##   
## data: a3  
## DW = 1.7583, p-value = 0.00321  
## alternative hypothesis: true autocorrelation is greater than 0