

```

# Assignment: Housing Data
# Name: Smitshoek, Stephen
# Date: 2022-05-06

library(readxl)
library(ggplot2)
setwd("C:\\Users\\sksmi\\PeytoAccess\\Personal\\Bellevue\\DSC520\\dsc520")
housing_data <- read_xlsx("data\\week-7-housing.xlsx")
colnames(housing_data)[1:2] <- c("sale_date", "sale_price")

housing_data$sale_year <- as.numeric(format(housing_data$sale_date, format =
"%Y"))
housing_data$sale_date <- as.Date(housing_data$sale_date)
housing_data$bath_total <- with(housing_data,
                                bath_full_count +
                                bath_half_count * .5 +
                                bath_3qtr_count * .75)

sale_price_lm1 <- lm(sale_price ~ sq_ft_lot,
                    data=housing_data)

sale_price_lm2 <- lm(sale_price ~ sq_ft_lot + year_built + sale_year +
                    bath_total + bedrooms,
                    data=housing_data)

summary(sale_price_lm1)
summary(sale_price_lm2)

library(lm.beta)
lm.beta(sale_price_lm2)

confint(sale_price_lm2, level=.95)

anova(sale_price_lm1, sale_price_lm2)

housing_data$residuals <- resid(sale_price_lm2)

housing_data$std.resid <- rstandard(sale_price_lm2) > 2 |
rstandard(sale_price_lm2) < -2

sum(housing_data$std.resid)

housing_data[housing_data$std.resid,]

housing_data$leverage <- hatvalues(sale_price_lm2)
housing_data$cooks <- cooks.distance(sale_price_lm2)
housing_data$cov.ratio <- covratio(sale_price_lm2)

avg.lvg <- (5 + 1) / 12865
cvr.pos <- 1 + (3 * (5+1) / 12865)
cvr.neg <- 1 - (3 * (5+1) / 12865)

subset(housing_data, std.resid & (cooks > 1 | leverage > avg.lvg * 3 |
cov.ratio > cvr.pos | cov.ratio < cvr.neg))[,c("sale_date", "std.resid",
"cooks", "leverage", "cov.ratio")]

```

```
housing_data[housing_data$sale_date == "2006-03-28",]  
  
library(car)  
durbinWatsonTest(sale_price_lm2)  
  
vif(sale_price_lm2)  
mean(vif(sale_price_lm2))  
1/vif(sale_price_lm2)  
  
plot(sale_price_lm2)  
hist(rstudent(sale_price_lm2))
```