Exam 3 - Question 1

Adam McQuistan

Saturday, April 30, 2016

Problem 1 - 8.24

The dataset

Variable	Description
Price	Selling price of house
Assessed	Quantitative variable. Assessed value in thousands USD
Lot	Qualitative variable. 1 for corner lot, 0 for non corner

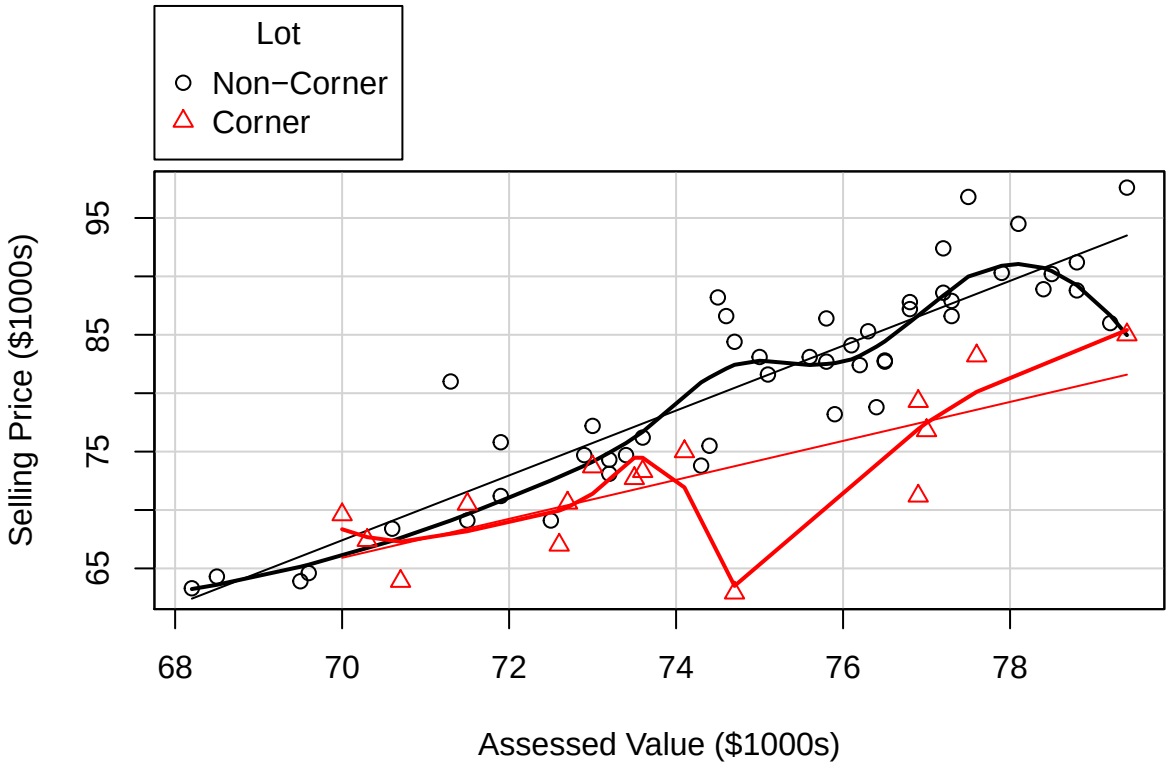
```
df <- read.csv(file="data/8.24.csv")
df$Case <- NULL
df$Lot <- factor(df$Lot, levels=c(0,1), labels=c("Non-Corner", "Corner"))
summary(df)
```

```
##      Price          Assessed             Lot
##  Min.   :62.90   Min.   :68.20   Non-Corner:48
##  1st Qu.:72.33   1st Qu.:72.85   Corner    :16
##  Median :79.05   Median :74.85
##  Mean   :79.02   Mean   :74.72
##  3rd Qu.:86.45   3rd Qu.:76.92
##  Max.   :97.60   Max.   :79.40
```

Part A

Plot the sample data for the two populations as a symbolic scatter plot. Does the regression appear to be the same for the two populations?

```
library(car)
scatterplot(Price ~ Assessed | Lot, data=df,
            ylab="Selling Price ($1000s)",
            xlab="Assessed Value ($1000s)")
```



Yes they are appear to be different. The slope for non-corner lots is greater and, on average the non-corner lots tend to sell above the corner lots for the same assessed values.

Note that the zero symbol is for non-corner lots and triangles are for corner lots.

Part B

Test for identity of the regression functuons for dwellings on corner and non-corner lots. Control for risk of type 1 error at 0.05.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \varepsilon$$

```
result1 <- lm(Price ~ Assessed + Lot + I(Assessed * as.numeric(Lot)),
              data=df)
result1_smry <- summary(result1)
df$residuals <- result1_smry$residuals
print(result1_smry)
```

```
##
## Call:
## lm(formula = Price ~ Assessed + Lot + I(Assessed * as.numeric(Lot)),
##     data = df)
##
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -10.8470  -2.1639   0.0913   1.9348   9.9836
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  -126.9052    14.7225  -8.620 4.33e-12 ***
## Assessed                        3.8834     0.5292   7.338 6.61e-10 ***
## LotCorner                      76.0215    30.1314   2.523  0.01430 *
## I(Assessed * as.numeric(Lot))  -1.1075     0.4055  -2.731  0.00828 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.893 on 60 degrees of freedom
## Multiple R-squared:  0.8233, Adjusted R-squared:  0.8145
## F-statistic: 93.21 on 3 and 60 DF,  p-value: < 2.2e-16
```

```
result1_aov <- fullRegressionAnova(anova(result1))
```

```
##                 VariationSource DF        SS         MS   F_stats
## 1                    Regression  3 4237.0502 1412.35007  93.21370
## 2                      Assessed  1 3670.9042 3670.90425 242.27602
## 3                           Lot  1  453.1474  453.14744  29.90728
## 4 I(Assessed * as.numeric(Lot))  1  112.9985  112.99852   7.45779
## 5                     Residuals 60  909.1046   15.15174        NA
## 6                         Total 63 5146.1548         NA        NA
```

Inferences about Two Regression Lines

$H_o : \beta_2 = \beta_3 = 0$

H_a : Not both $\beta_2 = 0$ and $\beta_3 = 0$

```
term1 <- (result1_aov$SS[3] + result1_aov$SS[4]) / (result1_aov$DF[3] + result1_aov$DF[4])
term2 <- result1_aov$SS[5] / result1_aov$DF[5]
F_stat <- term1 / term2
F_crit <- qf(0.95, 2,60)

msg <- paste("F stat: ", F_stat, "\nF crit: ", F_crit, sep="")
result <- ifelse(F_stat > F_crit, "Conclude Ha, not identical", "Conclude Ho, they are identical")
cat(msg, "\n", result, sep="")
```

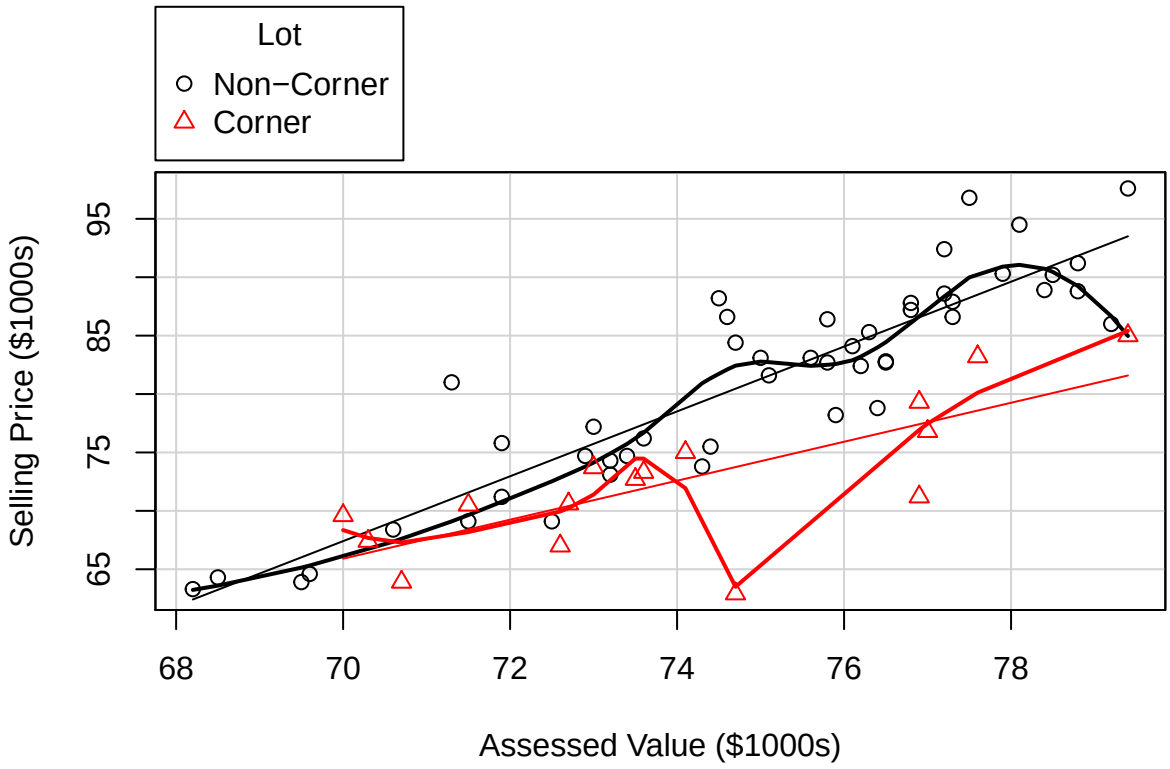
```
## F stat: 18.6825349040819
## F crit: 3.15041131058273
## Conclude Ha, not identical
```

Since F stat (18.7) > F crit (F(0.95, 2, 60) = 3.2) we conclude Ha that the regression functions are not identical.

Part C

Plot the estimated regression function for the two populations and describe the nature of the differences between them.

```
scatterplot(Price ~ Assessed | Lot, data=df,
            ylab="Selling Price ($1000s)",
            xlab="Assessed Value ($1000s)")
```



The slope for non-corner lots is greater and, on average the non-corner lots tend to sell above the corner lots for the same assessed values.