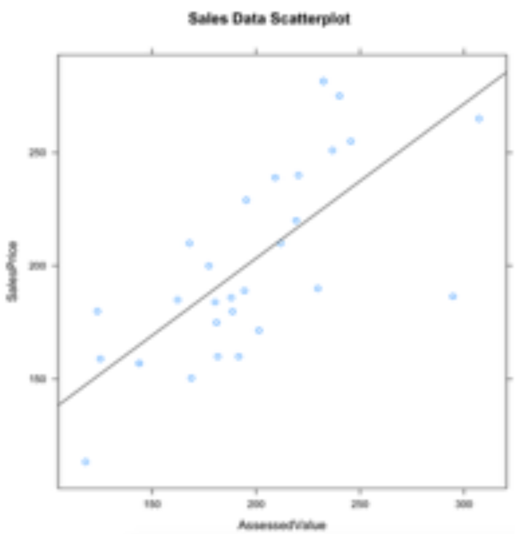
Harris Christiansen
christih@purdue.edu
STAT350 - Lab 8
2015-04-22

A. House Prices
   1. Scatterplot



   2. The scatterplot appears to have fairly constant standard deviation throughout, is primarily positive, and shows a trend of higher sales prices for higher assessed value homes. However, especially on the high AssessedValue side, the strength of the sales price linear regression line is not very strong.

   3. The correlation between SalesPrice and AssessedValue is 0.7061533, which is closer to 1 than 0 and indicates there is correlation between the two variables. This answer agrees with that from part 2, indicating that while there is correlation it is not incredibly strong.

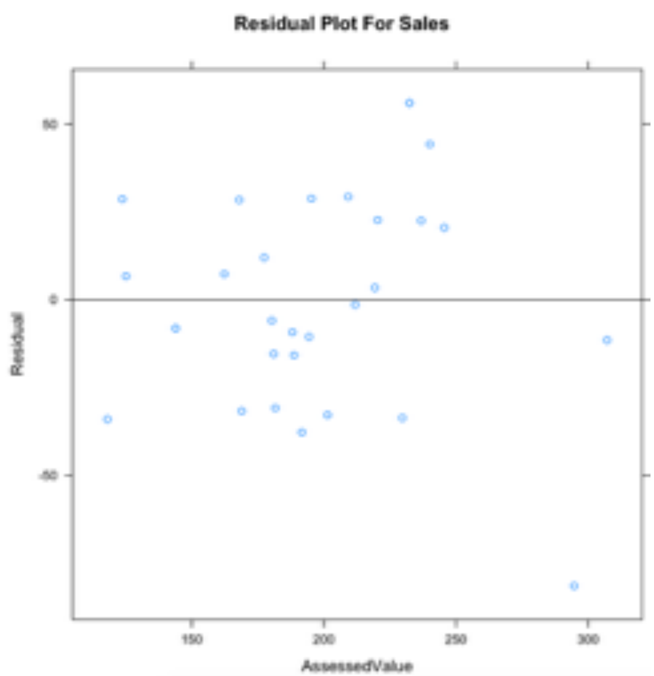
   4. Yes, the correlation value is a good summary of the appearance of the data.

   5. SalesPrice = 66.9460 + 0.6819*AssessedValue
      R^2 = 0.4987

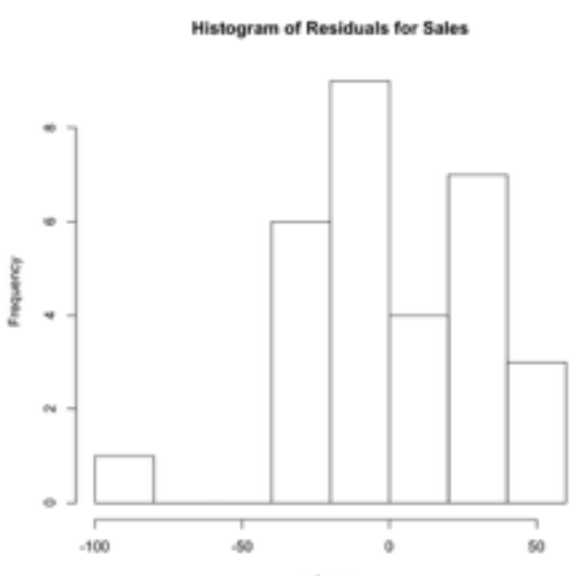
```
> summary(sales.lm)

Call:
lm(formula = SalesPrice ~ AssessedValue)

Residuals:
    Min      1Q  Median      3Q     Max 
-81.411 -15.638  -3.638  22.711  56.074 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)    66.9460    26.1086   2.564    0.016 *  
AssessedValue   0.6819     0.1292   5.277  1.3e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 31.41 on 28 degrees of freedom
Multiple R-squared:  0.4987,	Adjusted R-squared:  0.4807 
F-statistic: 27.85 on 1 and 28 DF,  p-value: 1.3e-05
```

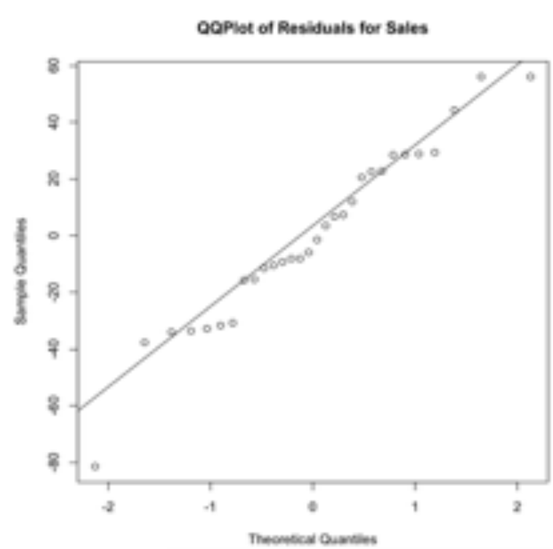
6. For Property=1:
   Predicted SalesPrice = 66.9460 + 0.6819*188.7 = 195.6205
   Residual = -15.726038

7. The residuals plot shows that the majority of the residuals fall evenly between -50 and 50, however there is one outlier with a high assessed value and extremely low residual. Because the residuals are evenly spread out, the conclusion is the same as in parts 2 and 4.



8. Based on the histogram and qqplot, the residuals appear to be normal.





9. Based on parts 1, 7, and 8, a linear regression analysis is reasonable because the data shows a trend with a significant amount of correlation and the residuals are evenly distributed.

10.

```
> confint(sales.lm, level=0.99)
                   0.5 %     99.5 %
(Intercept)   -5.1988794 139.090975
AssessedValue  0.3248595   1.038998
```

Slope: 99% CI = (0.3248595, 1.038998)
We are 99% sure that the population slope is between 0.3248595 and 1.038998.

Intercept: 99% CI = (-5.1988794, 139.090975)
We are 99% sure that the population slope is between -5.1988794 and 139.090975.

The significance of this result is that we are fairly confident the population data yields a correlation between AssessedValue and SalesPrice. The inference on the intercept is of interest in this problem because it covers such a large range.

11. Part 0: B is the population slope.

Part 1:
H_0: B = 0
H_a: B ≠ 0

Part 2:
t_t = 5.277
DF = 28

Part 3:
P-Value = .000013

Part 4:
a = 0.01
Since 0.000013 < 0.01, we should reject H_a and conclude that there is an association between AssessedValue and SalesPrice.

12. Parts 10 and 11 came to the same conclusion that there is significant correlation between AssessedValue and SalesPrice.

13. Using a Linear Regression Analysis, we have shown that there is a strong correlation between AssessedValue and SalesPrice. The model is appropriate for predicting the SalesPrice based on AssessedValue. The effects of switching x and y would result in predicting AssessedValue using SalesPrices, and should also be appropriate with this model. This model should only be used in the same demographic as that from which it was gathered.

Appendix:

```
################ Problem A ###############

sales <- read.table(file="sales.txt", header=TRUE)

################ Part 1 ###############

attach(sales)
library(lattice)
xyplot(SalesPrice ~ AssessedValue, data = sales, panel=function(x,y) {
 panel.xyplot(x, y)
 panel.lmline(x, y)
}, main="Sales Data Scatterplot")

################ Part 3 ###############
cor(sales$SalesPrice, sales$AssessedValue)

################ Part 5 ###############

sales.lm = lm(SalesPrice ~ AssessedValue)
sales.res = sales.lm$res
summary(sales.lm)
# Answer: SalesPrice = 66.9460 + 0.6819AssesedValue
# R^2: 0.4987

################ Part 6 ###############

# Answer (Property=1): 66.9460 + 0.6819*188.7 = 195.6205
# TO-DO: Residual

################ Part 7 ###############

xyplot(sales.res ~ AssessedValue, data=sales, main="Residual Plot For
 Sales", ylab="Residual", panel=function(x,y){
 panel.xyplot(x,y)
 panel.abline(h=0)
})

################ Part 8 ###############

hist(sales.res, main="Histogram of Residuals for Sales")
qqnorm(sales.res, main="QQPlot of Residuals for Sales")
qqline(sales.res)
```

```
################ Part 10 ###############

confint(sales.lm, level=0.99)
```