Hw3 Part 1 Review

CS100

Linear Regression

Assumed Linear Model: $Y = A + Bx$

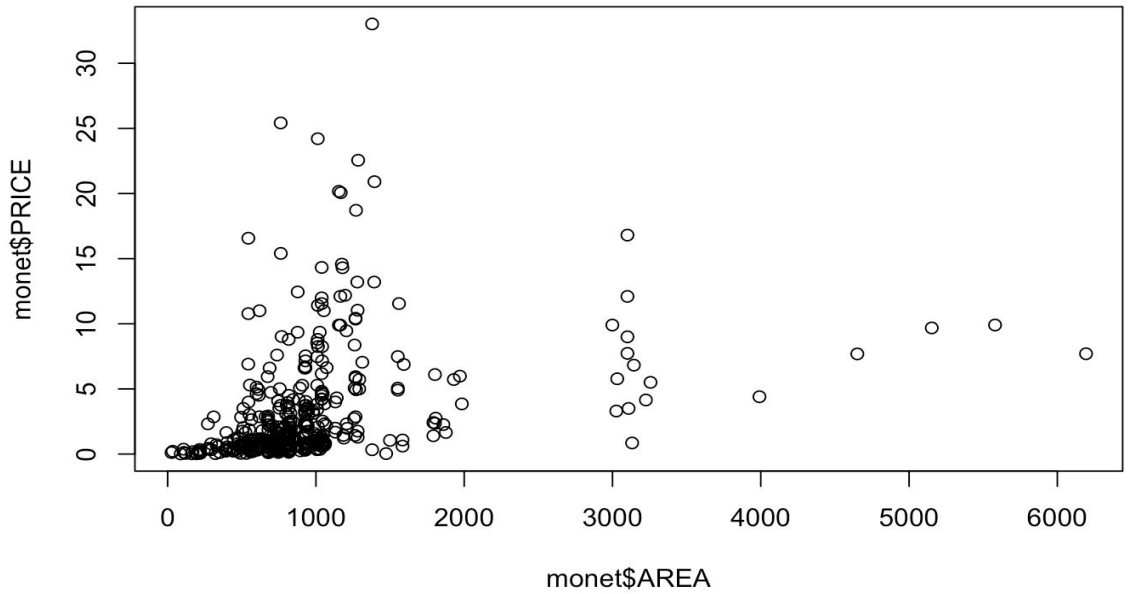
abline(a = A, b = B)

Log/Sqrt Transformation

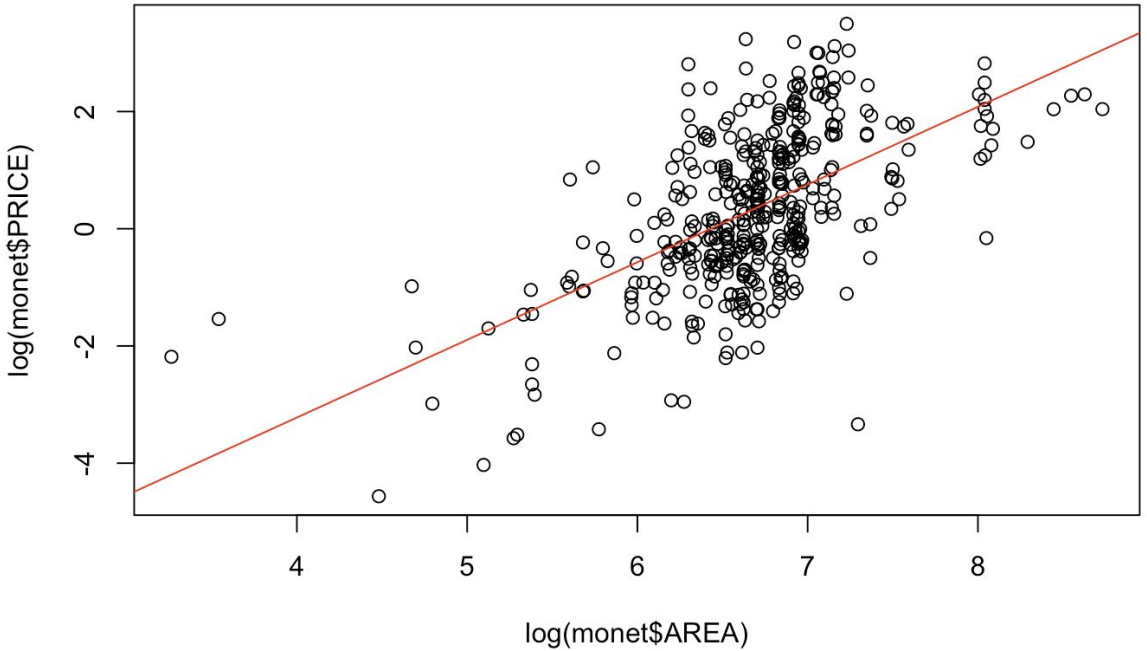
When/Why do we do Log or Sqrt Transformation?

- Not all relationships are Linear (EX) monet.PRICE VS monet.AREA
- Some of the relationships can be **MADE** linear with **APPROPRIATE** transformation.
- If residual plot is random, it is good fit for Linear Model.
- If residual plot is not random (U curve, inverted-U curve), not a good fit for Linear Model → We transform the data to use fit linear model.

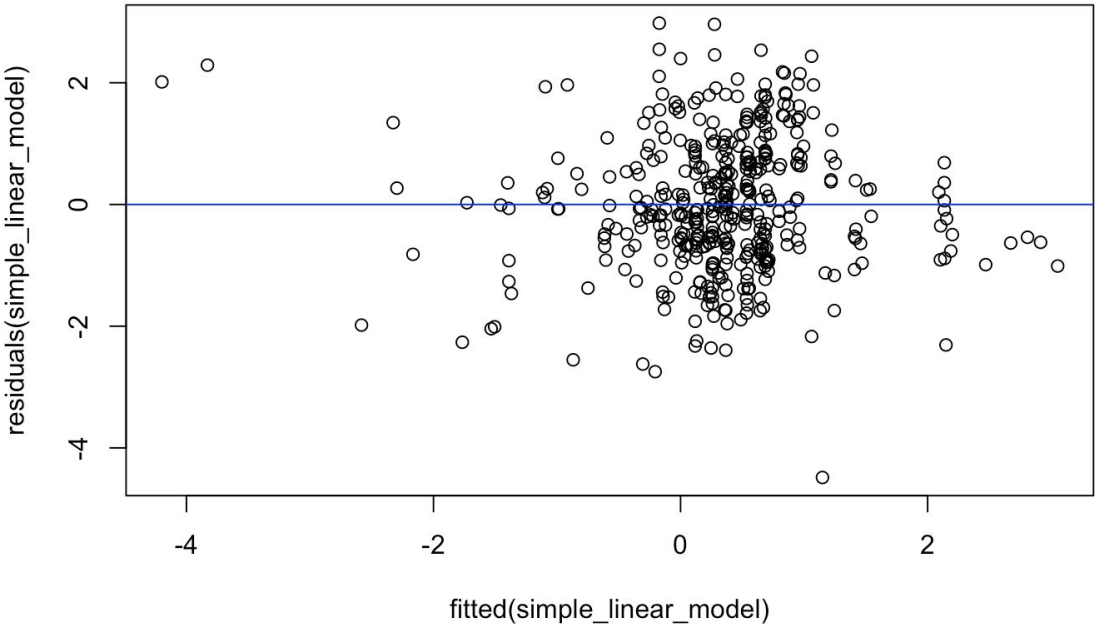
PRICE vs AREA (LINEAR RELATIONSHIP?)



Log(PRICE) vs Log(AREA) (BETTER?)



RESIDUAL (remove the first 2...??)



CORRELATION COEFFICIENT

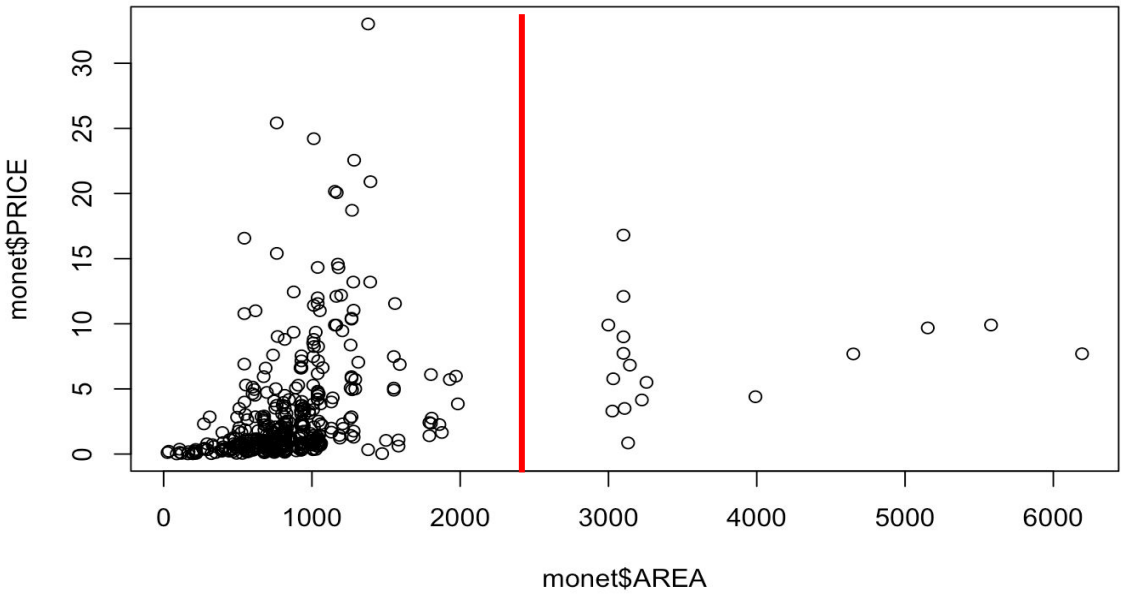
- measures the **STRENGTH** of linear relationship between 2 variables.

REGRESSION SLOPE

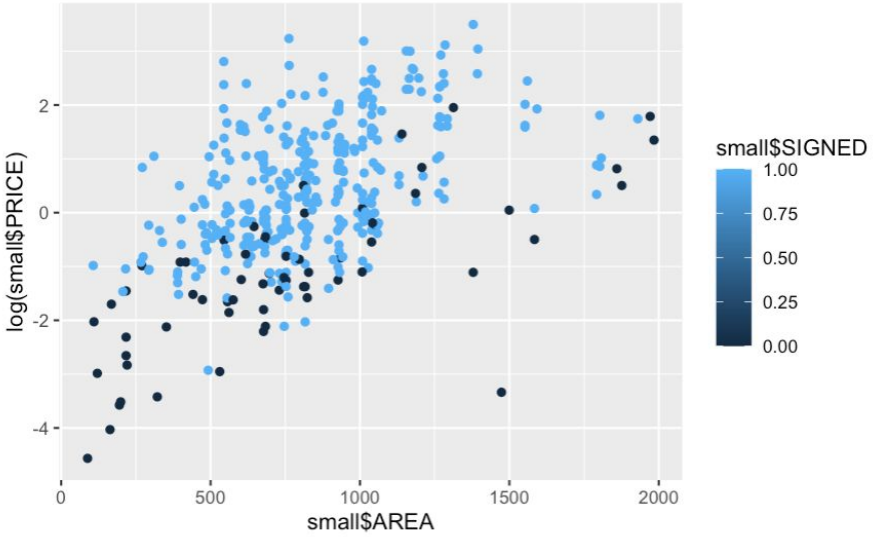
- measures the **STEEPNESS** of the linear relationship between 2 variables.

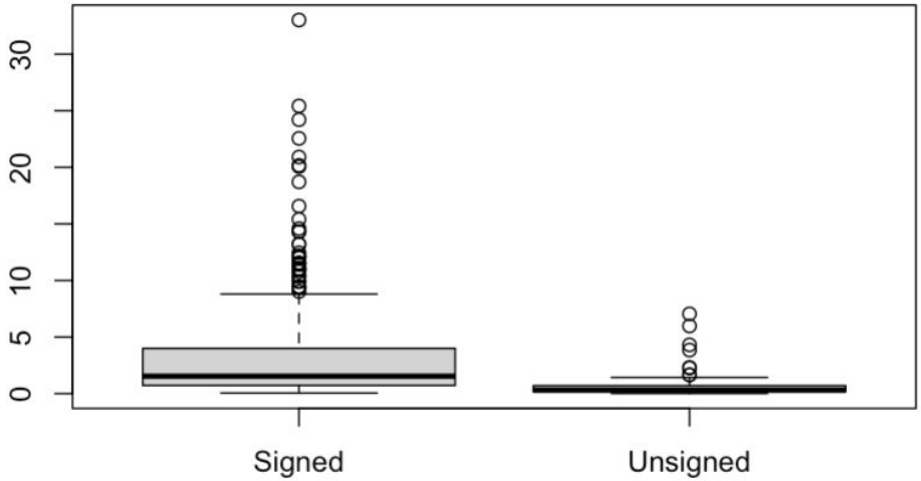
"Intuitively, the more correlated 2 variables are, the easier it is to draw a line of best linear fit. Hence, we computed the correlation of each transformation."

PRICE vs AREA



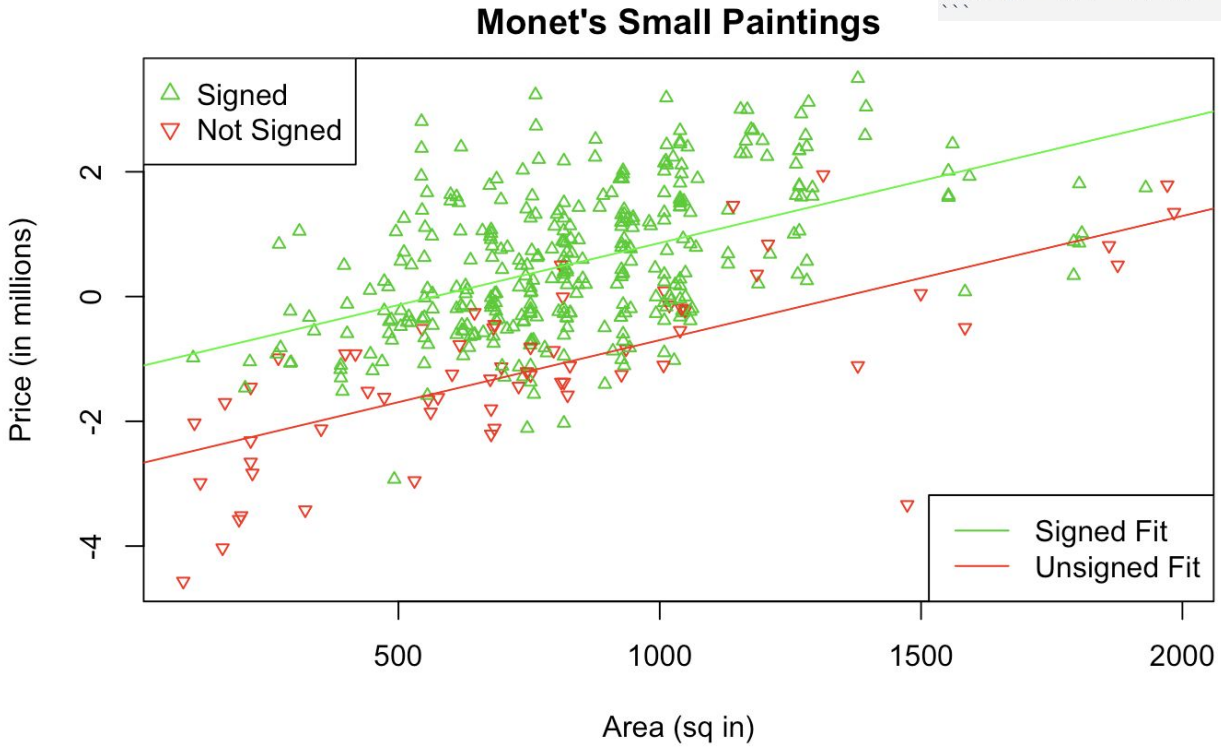
SIGNED AND UNSIGNED PLOT OF SMALL PICTURES





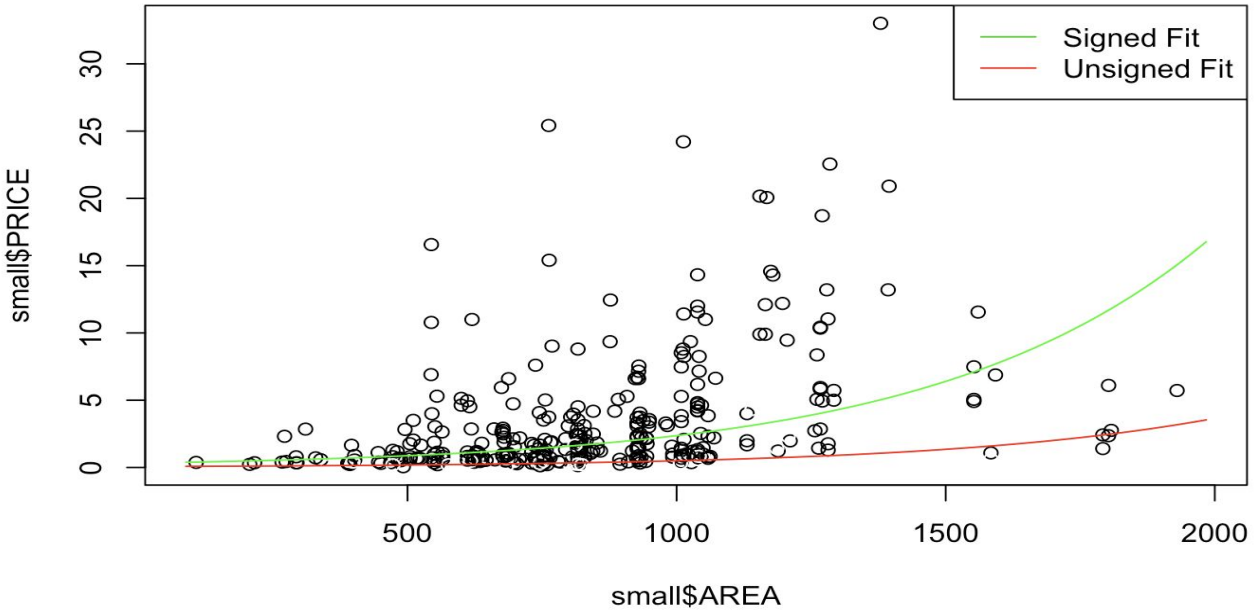
log(PRICE) vs AREA fit curve

```r
signed_linear_model <- lm(log(small$PRICE) ~ small$AREA + small$SIGNED)
a <- signed_linear_model$coefficients[1]
b <- signed_linear_model$coefficients[2]
dummy <- signed_linear_model$coefficients[3]
plot(log(small$PRICE) ~ small$AREA, col = as.factor(small$SIGNED))
abline(a = a + dummy, b = b, col = "red")
abline(a = a, b = b, col = "black")
```

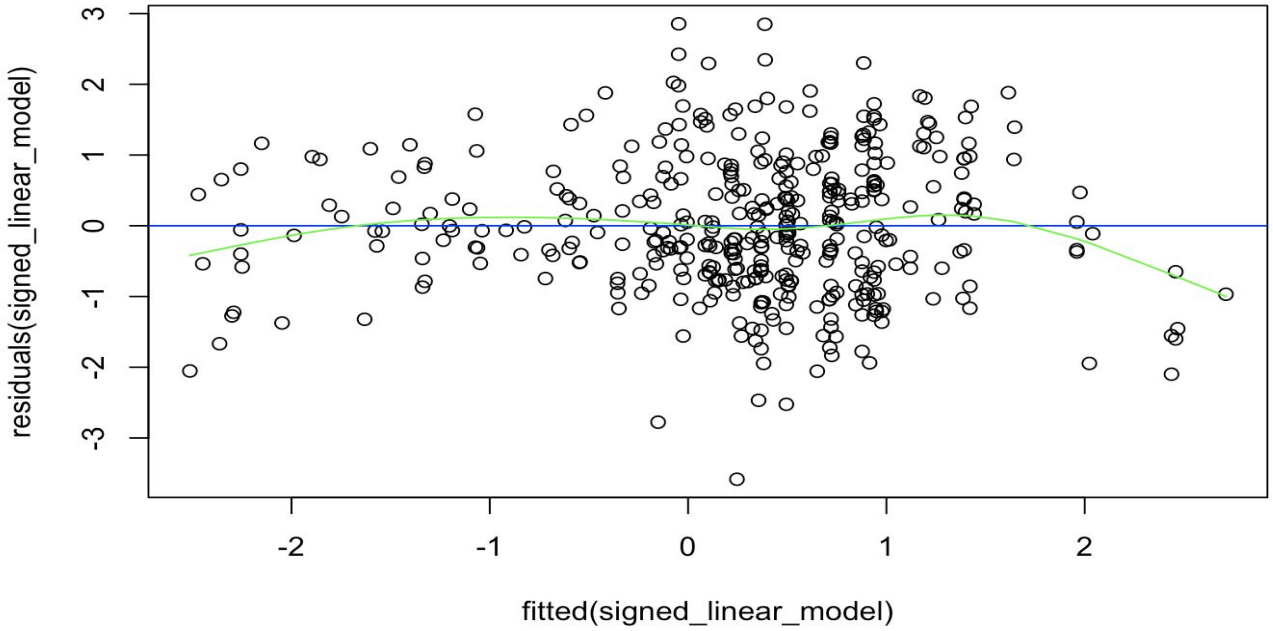


PRICE vs AREA fit curve

```r
```{r}
#Plot model in non - log space.
a <- signed_linear_model$coefficients[1]
b <- signed_linear_model$coefficients[2]
dummy <- signed_linear_model$coefficients[3]
plot(small$PRICE~small$AREA)
curve(exp(a + b * x + dummy), col = "green", add = TRUE)
curve(exp(a + b * x), col = "red", add = TRUE)
```
```



RESIDUAL


residuals(signed_linear_model)
3
2
1
0
-1
-2
-3
-2
-1
0
1
2
fitted(signed_linear_model)