# HW6Q2

Zachary Lazerick

11/4/2021

Load the Lawstat Library to Access the Brown-Forscythe-Levene Test

```
library(lawstat)
```

Read in the Data

```
Advertising = read.csv("Advertising.csv")
attach(Advertising)
```

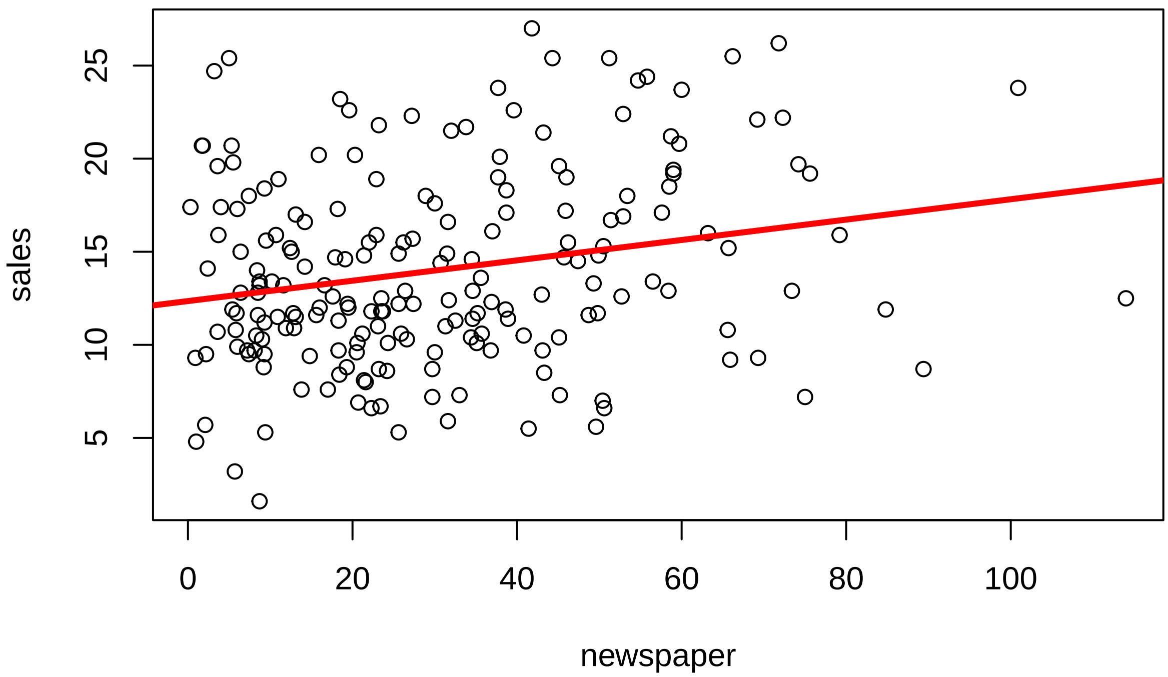
**Question 2: Building an SLR Model Sales ~ Newspaper** (Initial Results)

Construct a Simple Linear Model with Predictor: Newspaper and Response: Sales

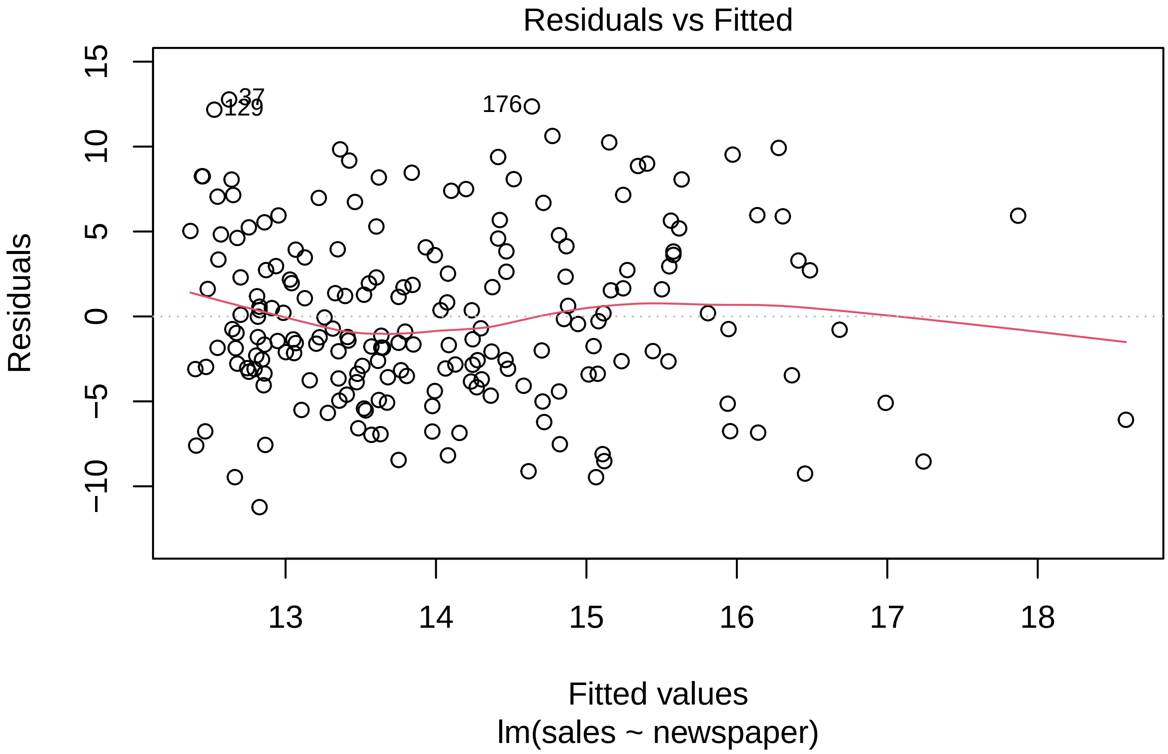
```
lm.fit.news = lm(sales~newspaper, data = Advertising)
```

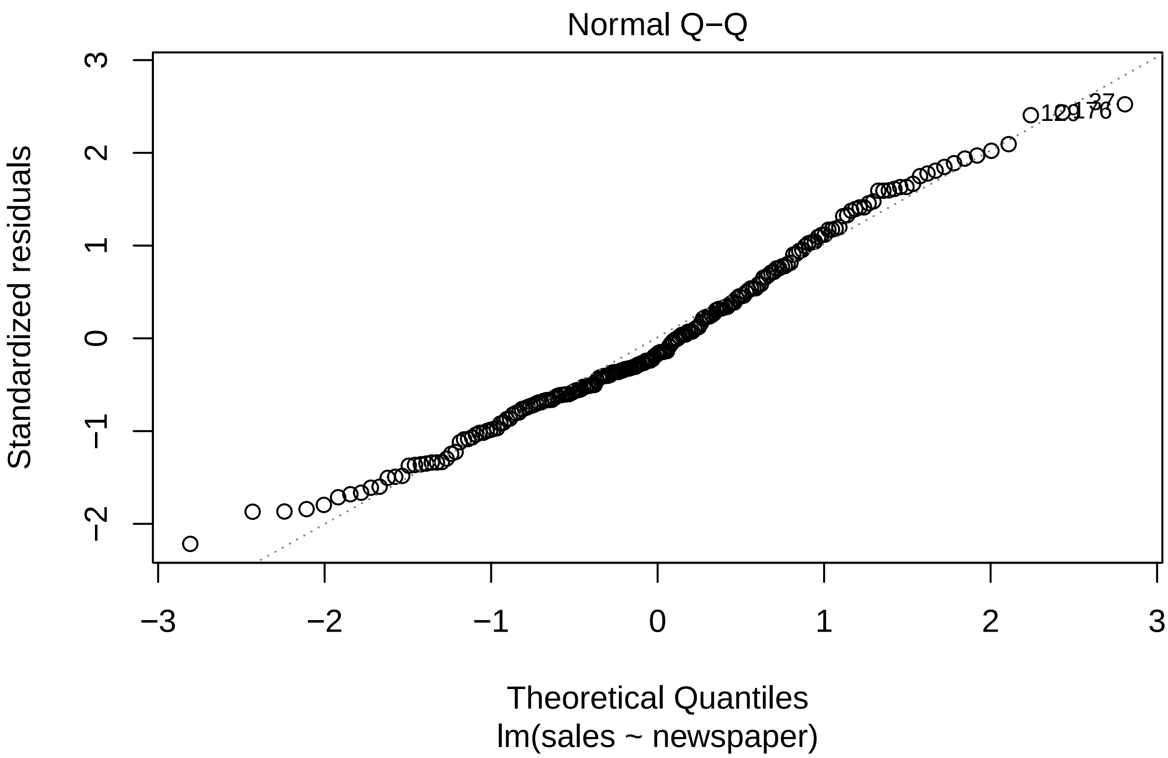
```
##
## Call:
## lm(formula = sales ~ newspaper, data = Advertising)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.2272  -3.3873  -0.8392   3.5059  12.7751
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.35141    0.62142   19.88  < 2e-16 ***
## newspaper    0.05469    0.01658    3.30  0.00115 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.092 on 198 degrees of freedom
## Multiple R-squared:  0.05212,    Adjusted R-squared:  0.04733
## F-statistic: 10.89 on 1 and 198 DF,  p-value: 0.001148
```

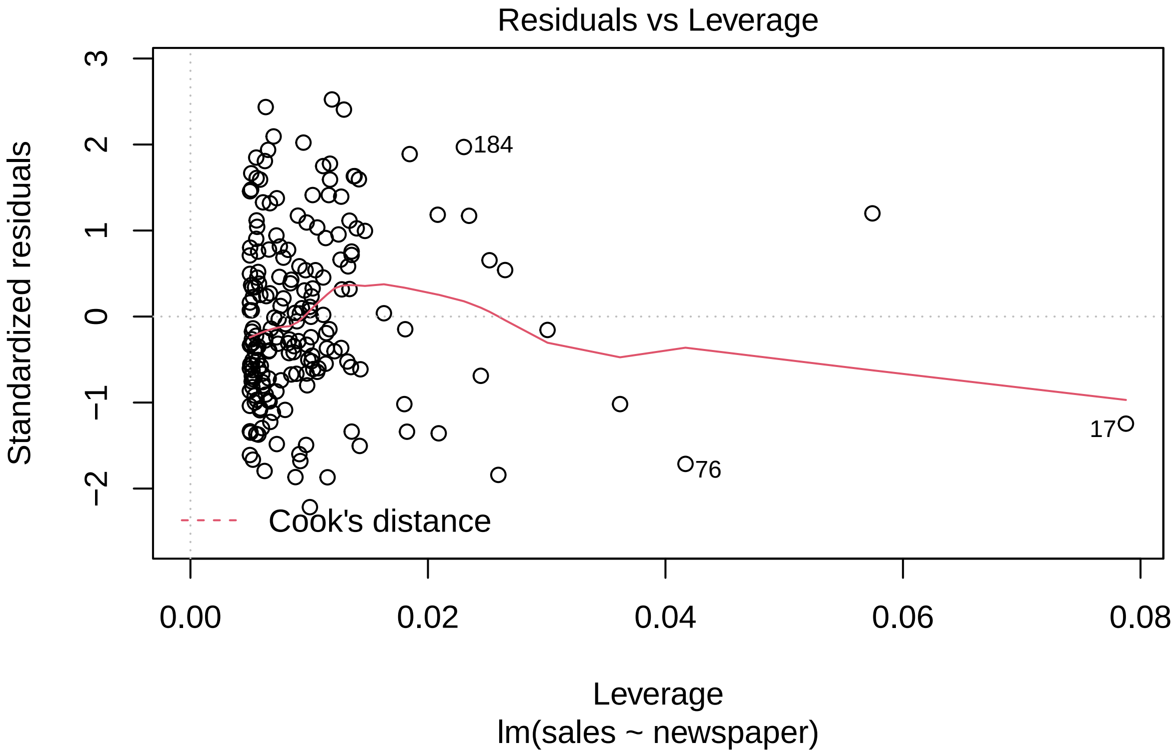
Construct a Scatter Plot with the Calculated Linear Model "Eye-Test" for Abnormalities (Non-linearity, etc.)



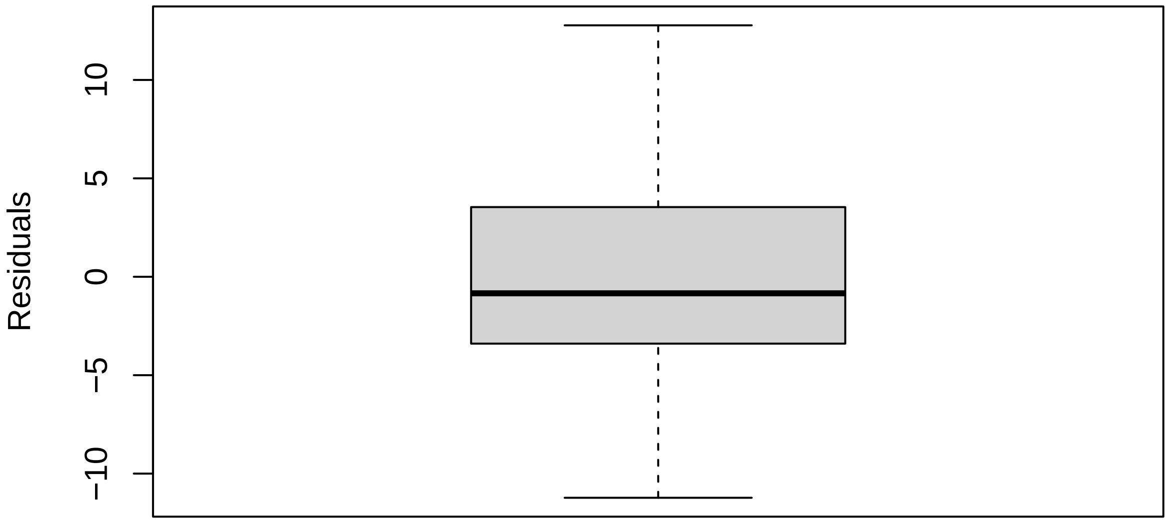

Construct Plots to Check Diagnostics

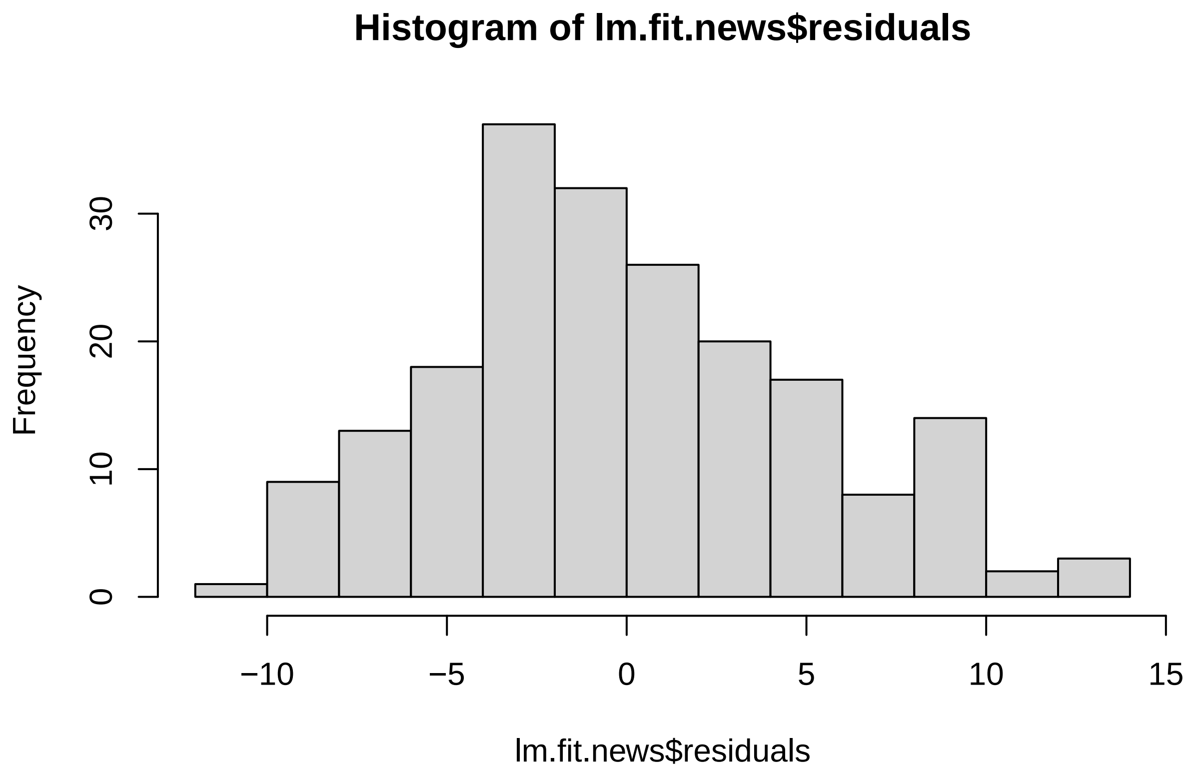

Residuals vs Fitted
Residuals
Fitted values
lm(sales ~ newspaper)
37
129
176
15
10
5
0
−5
−10
13
14
15
16
17
18


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(sales ~ newspaper)
129
176
37
−3
−2
−1
0
1
2
3




Check Distribution of Residuals for Obvious Deviations


Residuals
10
5
0
−5
−10




Run Shapiro-Wilk and Brown-Forscythe-Levene Tests

```
shapiro.test(lm.fit.news$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  lm.fit.news$residuals
## W = 0.98197, p-value = 0.0114
```

```
levene.test(lm.fit.news$residuals, Advertising$Group, location = c("median"))
```

```
##
##  Modified robust Brown-Forsythe Levene-type test based on the absolute
##  deviations from the median
##
## data:  lm.fit.news$residuals
## Test Statistic = 2.7715, p-value = 0.09754
```

**Conclusion** From the Shapiro-Wilks Test, we conclude that the data is not from a normal distribution. However, from the Levene Test, we conclude that the data is homoscedastic at the .05 significance level, although not very convincingly. Therefore, we must transform the data in hopes of obtaining approximately normal and more convincing homoscedastic data to draw statistically significant conclusions from this data.

**Data Transformation Test Results**

| Response ~ Predictor | Shapiro-Wilks P-Value | Levene P-Value |
|---|---|---|
| Sales ~ Newspaper | .0114 | .09754 |
| sqrt(Sales) ~ Newspaper | .308 | .3703 |
| log10(Sales) ~ Newspaper | 5.41e-06 | .9783 |
| (1/Sales) ~ Newspaper | 2.2e-16 | .2057 |
| log10(Sales) ~ sqrt(Newspaper) | 8.661e-06 | .9341 |
| sqrt(Sales) ~ sqrt(Newspaper) | .2702 | .4016 |
| (1/Sales) ~ sqrt(Newspaper) | 2.2e-16 | .185 |
| log10(Sales) ~ log10(Newspaper) | 8.478e-06 | .9743 |
| sqrt(Sales) ~ log10(Newspaper) | .1401 | .3659 |
| (1/Sales) ~ log10(Newspaper) | 2.2e-16 | .1831 |

**Conclusion** From transforming our data, we obtain more convincing homoscedastic data in all cases, some of which also yielding approximate normally distributed data. The three transformations that show this best are 1. sqrt(Sales) ~ sqrt(Newspaper), 2. sqrt(Sales) ~ Newspaper, and 3. sqrt(Sales) ~ log10(Newspaper). Ultimately we will use the sqrt(Sales) ~ sqrt(Newspaper) transformation to analyze our data because it yielded the greatest p-value for the Levene-Test of these three transformations

**Analyzing the sqrt(Sales) ~ sqrt(Newspaper) Model**

```
summary(lm.fit.news.6)
```

```
##
## Call:
## lm(formula = sqrtSales ~ sqrtNews, data = Advertising)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.25846 -0.43085 -0.06373  0.51934  1.56655
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.31633    0.13396  24.756   <2e-16 ***
## sqrtNews     0.07019    0.02424   2.896   0.0042 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.697 on 198 degrees of freedom
## Multiple R-squared:  0.04065,    Adjusted R-squared:  0.0358
## F-statistic: 8.389 on 1 and 198 DF,  p-value: 0.0042
```

**Interpretation of the Model** From the Summary of our Simple Linear Model, we obtain a regression equation of $Y = 3.31633 + .07019X$, with both coefficients being statistically significant at the .05 significance level. This means that there is an association between the level of Newspaper Advertising and Number of Unit Sales. Therefore the company should increase its Newspaper Advertising budget to indirectly increase Sales. In fact, for every increment of 1 in level of sqrt(Newspaper), (in thousands of dollars), we expect that the response, sqrt(Sales), (in thousands of dollars), increases by .07019. However, the interpretation of our intercept does not make sense. This is because at sqrt(Newspaper) = 0, it would be expected that sqrt(Sales) = 0 by the fact that it would be impossible to have sales with no produced product.