B27
Hector G. FLores Rodriguez

Stats 110 - HW5
November 13, 2017

1. 3.23

```r
# Load data
diamonds = read.csv("../data/Diamonds.csv")
```

1-(a). Quadratic model using Depth

```r
model_a = lm(TotalPrice ~ Depth + I(Depth^2), data=diamonds)
summary(model_a)
```

```
##
## Call:
## lm(formula = TotalPrice ~ Depth + I(Depth^2), data = diamonds)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -9323  -4251  -2676   2134  45513
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -28406.783 112211.790  -0.253    0.800
## Depth          766.369   3353.222   0.229    0.819
## I(Depth^2)      -3.233     24.869  -0.130    0.897
##
## Residual standard error: 7616 on 348 degrees of freedom
## Multiple R-squared:  0.04748,    Adjusted R-squared:  0.042
## F-statistic: 8.673 on 2 and 348 DF,  p-value: 0.0002111
```

1-(b). Two-predictor model using Carat and Depth

```r
model_b = lm(TotalPrice ~ Carat + Depth, data=diamonds)
summary(model_b)
```

```
##
## Call:
## lm(formula = TotalPrice ~ Carat + Depth, data = diamonds)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9234.7 -1223.7  -274.3  1161.0 16368.6
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1059.24    1918.36   0.552    0.581
## Carat       15087.01     320.96  47.006  < 2e-16 ***
## Depth        -134.94      30.92  -4.364 1.68e-05 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2809 on 348 degrees of freedom
## Multiple R-squared:  0.8704, Adjusted R-squared:  0.8696
## F-statistic:  1168 on 2 and 348 DF,  p-value: < 2.2e-16
```

1-(c). A three-predictor model that adds interaction for Carat and Depth

```r
model_c = lm(TotalPrice ~ Carat + Depth + Carat*Depth, data=diamonds)
summary(model_c)
```

```
##
## Call:
## lm(formula = TotalPrice ~ Carat + Depth + Carat * Depth, data = diamonds)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8254.4 -1311.5  -157.2  1131.8 14513.9
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  31171.41    4219.58   7.387 1.13e-12 ***
## Carat       -11827.73    3436.47  -3.442 0.000648 ***
## Depth         -598.18      65.47  -9.137  < 2e-16 ***
## Carat:Depth    408.45      51.96   7.861 4.84e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2592 on 347 degrees of freedom
## Multiple R-squared:   0.89,  Adjusted R-squared:  0.889
## F-statistic: 935.7 on 3 and 347 DF,  p-value: < 2.2e-16
```

1-(d). Complete second-order model using Carat and Depth.

```r
model_d = lm(TotalPrice ~ Carat + Depth + I(Carat^2) + I(Depth^2) + Carat*Depth,
             data=diamonds)
summary(model_d)
```

```
##
## Call:
## lm(formula = TotalPrice ~ Carat + Depth + I(Carat^2) + I(Depth^2) +
##     Carat * Depth, data = diamonds)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12196.1   -652.7    -38.5    485.7  10582.2
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 24338.820  30297.912   0.803   0.4223
## Carat        7573.620   3040.787   2.491   0.0132 *
## Depth        -728.700    904.439  -0.806   0.4210
## I(Carat^2)   4761.592    330.246  14.418   <2e-16 ***
## I(Depth^2)      5.276      6.727   0.784   0.4333
```

```
## Carat:Depth   -83.891     53.530  -1.567   0.1180
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2053 on 345 degrees of freedom
## Multiple R-squared:  0.9313, Adjusted R-squared:  0.9304
## F-statistic: 936.1 on 5 and 345 DF,  p-value: < 2.2e-16
```

The quadratic model where CaratSq is the quadratic term provides the better model. The model captures most of the variance yet it is a more parsimonious model than the second order model and the three predictor model with the interaction term.

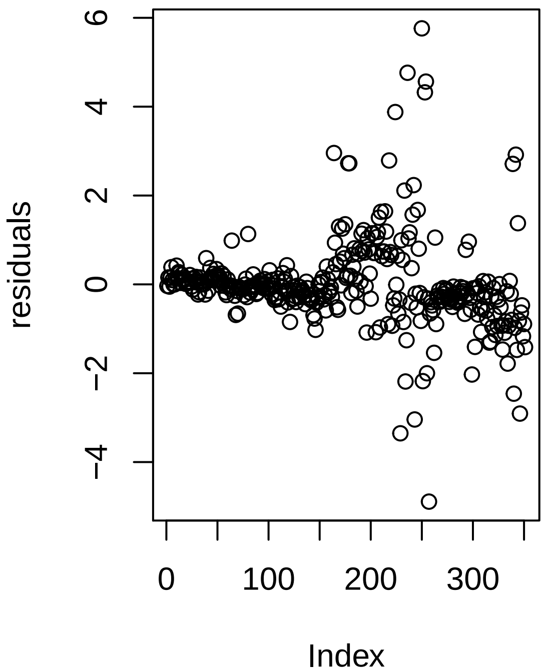
2. Exercise 3.24

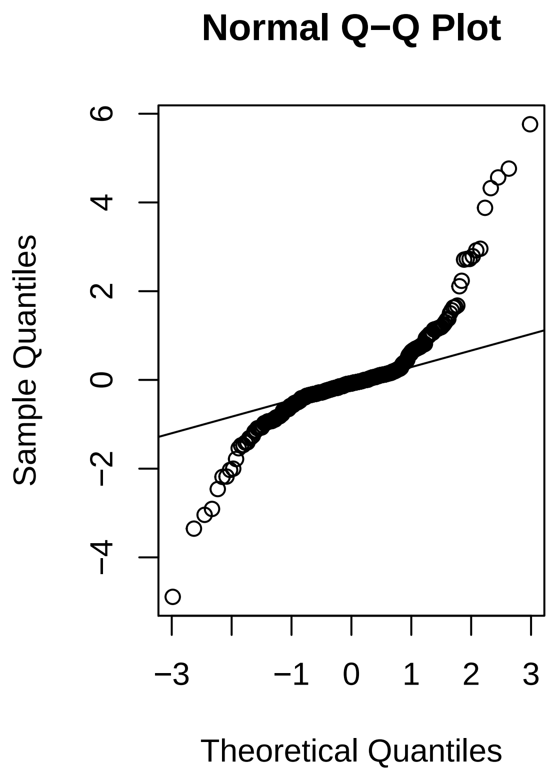
```
my_model = lm(TotalPrice ~ Carat + I(Carat^2), data=diamonds)
summary(my_model)
```

```
##
## Call:
## lm(formula = TotalPrice ~ Carat + I(Carat^2), data = diamonds)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10207.4   -711.6   -167.9    355.0  12147.3
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -522.7      466.3  -1.121  0.26307
## Carat         2386.0      752.5   3.171  0.00166 **
## I(Carat^2)    4498.2      263.0  17.101  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2127 on 348 degrees of freedom
## Multiple R-squared:  0.9257, Adjusted R-squared:  0.9253
## F-statistic:  2168 on 2 and 348 DF,  p-value: < 2.2e-16
```

2-(a). Check for homoscedacity and normality

```
par(mfrow=c(1,2))
residuals = rstandard(my_model)
plot(residuals)
qqnorm(residuals)
qqline(residuals)
```





There is heteroskedacity and a lack of normality in my model.

2-(b). Transform response variable to natural log

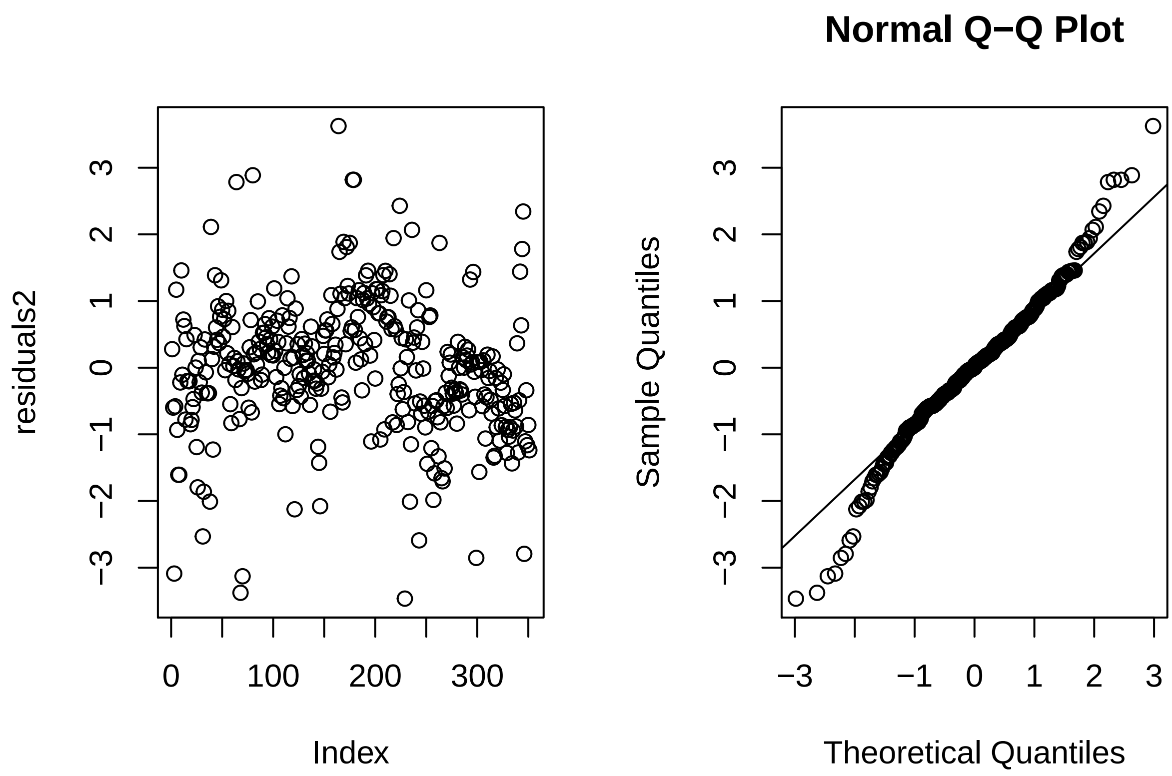
```
log_model = lm(log(TotalPrice) ~ Carat + I(Carat^2), data=diamonds)
summary(log_model)
```

```
##
## Call:
## lm(formula = log(TotalPrice) ~ Carat + I(Carat^2), data = diamonds)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.8215 -0.1313  0.0003  0.1391  0.8615
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.13042    0.05218  117.48   <2e-16 ***
## Carat        3.05963    0.08422   36.33   <2e-16 ***
## I(Carat^2)  -0.52730    0.02944  -17.91   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.238 on 348 degrees of freedom
## Multiple R-squared:  0.925,  Adjusted R-squared:  0.9246
## F-statistic:  2146 on 2 and 348 DF,  p-value: < 2.2e-16
```

My “best” choice of predictors is still a reasonable choice for predicting logPrice.

2-(c). Check graphs for log model

```
par(mfrow=c(1,2))
residuals2 = rstandard(log_model)
plot(residuals2)
qqnorm(residuals2)
qqline(residuals2)
```



The log transformation helped with heteroskedacity and normality. The assumptions are now fairly reasonable.

4. Exercise 3.25

Use nested F-test for the complete second-order model to determine if all the terms in the model that involve Depth could be removed withought signifcantly impairing its effectiveness.

```
full_model = model_d
reduced_model = my_model
anova(reduced_model, full_model)
```

```
## Analysis of Variance Table
##
## Model 1: TotalPrice ~ Carat + I(Carat^2)
## Model 2: TotalPrice ~ Carat + Depth + I(Carat^2) + I(Depth^2) + Carat *
##     Depth
```

```
##   Res.Df        RSS Df Sum of Sq      F   Pr(>F)
## 1    348 1574044410
## 2    345 1454702094  3 119342316 9.4345 5.24e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

There exist some term that involves information on Depth that is significant.

4. Exercise 3.26

Couple found a 0.5-carat diamond with depth of 62%.

4-(a). Determine average total price for quadratic model

```r
new_diamond = data.frame(Carat=0.5, Depth=62.0)
predict(my_model, new_diamond)
```

```
##        1
## 1794.843
```

The quadratic model predicts $1794.84 as the average total price.

4-(b). 95% confidence interval for the mean total price

```r
predict(my_model, new_diamond, interval="c")
```

```
##        fit      lwr      upr
## 1 1794.843 1424.296 2165.389
```

With 95% confidence, the mean total price for a 0.5-carat diamond with a depth of 62% somewhere between $1424.30 and $2165.40

4-(c). 95% prediction interval

```r
predict(my_model, new_diamond, interval="predict")
```

```
##        fit       lwr      upr
## 1 1794.843 -2404.462 5994.147
```

With 95% confidence, we predict that the total price for a 0.5-carat diamond with depth of 62% is somewhere between $0 and $5994.15

Note: Negative dollars or recieving money for purchasing a diamond doesn't make sense.

4-(d). Repeat two intervals for logPrice

```r
exp(predict(log_model, new_diamond, interval="c"))
```

```
##        fit      lwr      upr
## 1 1860.149 1784.588 1938.908
```

With 95% confidence, the mean total price for a 0.5-carat diamond with a depth of 62% somewhere between $1784.59 and $1938.91

```r
exp(predict(log_model, new_diamond, interval="p"))
```

```
##        fit      lwr     upr
## 1 1860.149 1162.651 2976.09
```

With 95% confidence, we predict that the total price for a 0.5-carat diamond with depth of 62% is somewhere between $1162.65 and $2976.09