STAT 608 HW 3

Jack Cunningham (jgavc@tamu.edu)

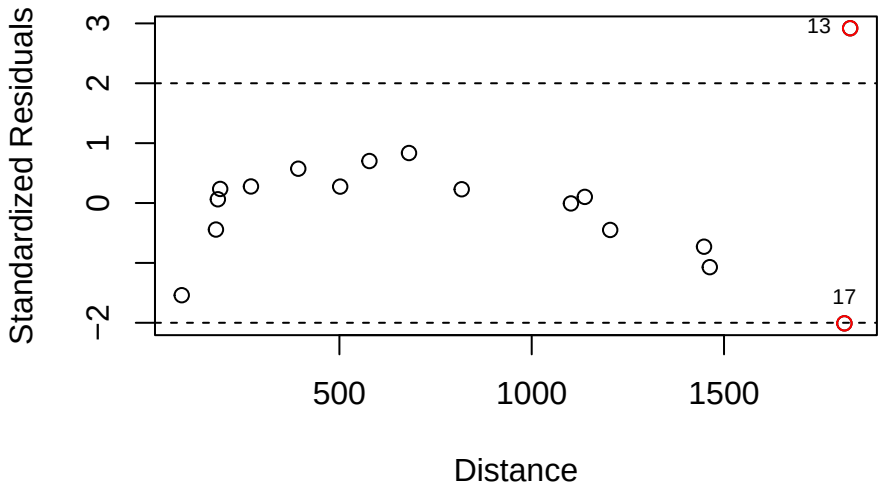
9/23/24

1.

a)

The business analyst claims that this model is highly effective for understanding the effects of Distance on Fare and predicting future values of Fare, there a few issues with this conclusion. First when we look at the standard residual plot we see that there appears to be a discernible quadratic pattern. With $n = 17$ it is very important that residuals be distributed i.i.d $N(0, \sigma^2)$ in order to make inference on coefficient estimates. Additionally with **any** sized sample residuals must be distributed i.i.d $N(0, \sigma^2)$ in order to create prediction intervals about the response variable.

Wwith our current model there are also two outlier points I have labeled in red below. These need to be further analyzed.



b)

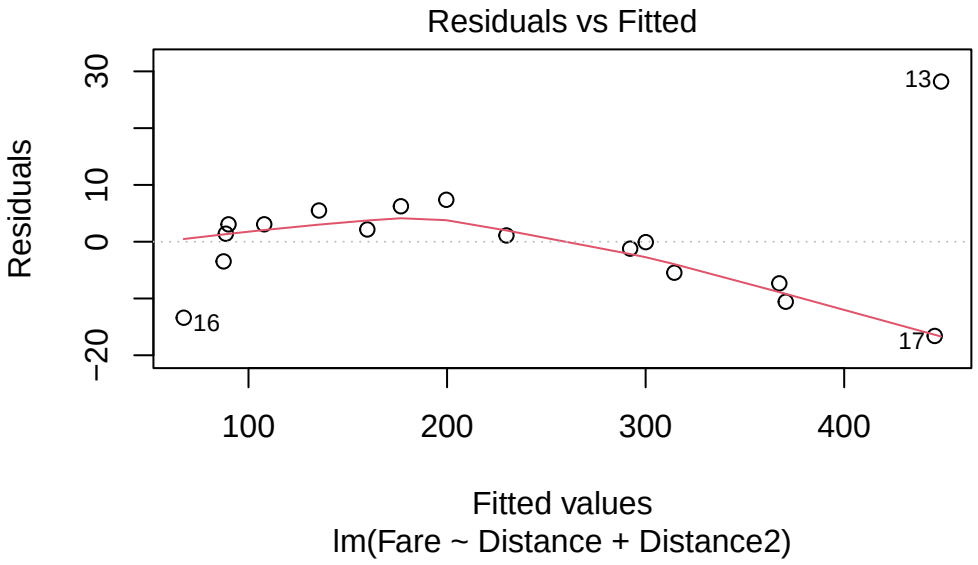
The ordinary straight line regression does not appear to fit the data well. We can see a clear quadratic pattern in the standardized residuals. I would recommend introducing the term $Distance^2$ and analyzing the standardized residuals again. I would also examine the properties of point 13 as this appears to be a bad leverage point.

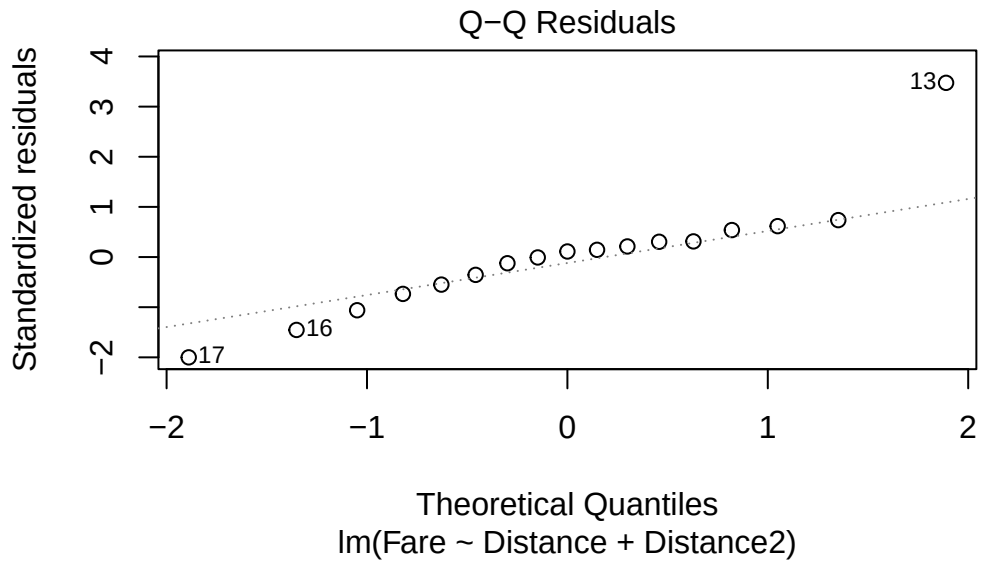
```
airfares$Distance2 <- airfares$Distance^2
airline_quad <- lm(Fare~Distance + Distance2, data = airfares)
summary(airline_quad)
```

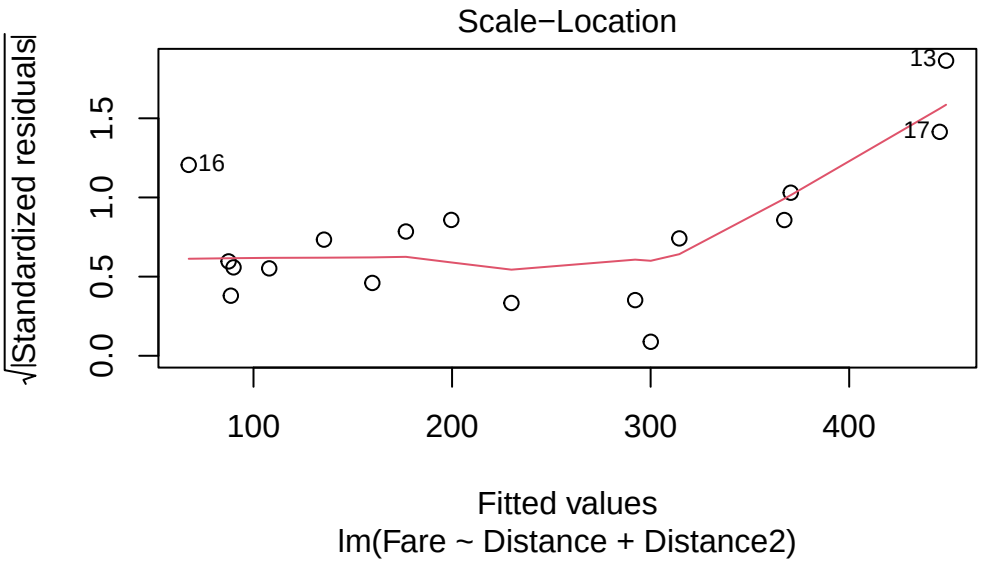
```
Call:
lm(formula = Fare ~ Distance + Distance2, data = airfares)

Residuals:
    Min      1Q  Median      3Q     Max
-16.599  -5.457   1.102   3.064  28.206

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.702e+01  6.808e+00   6.907 7.25e-06 ***
Distance     2.266e-01  1.861e-02  12.177 7.75e-09 ***
```

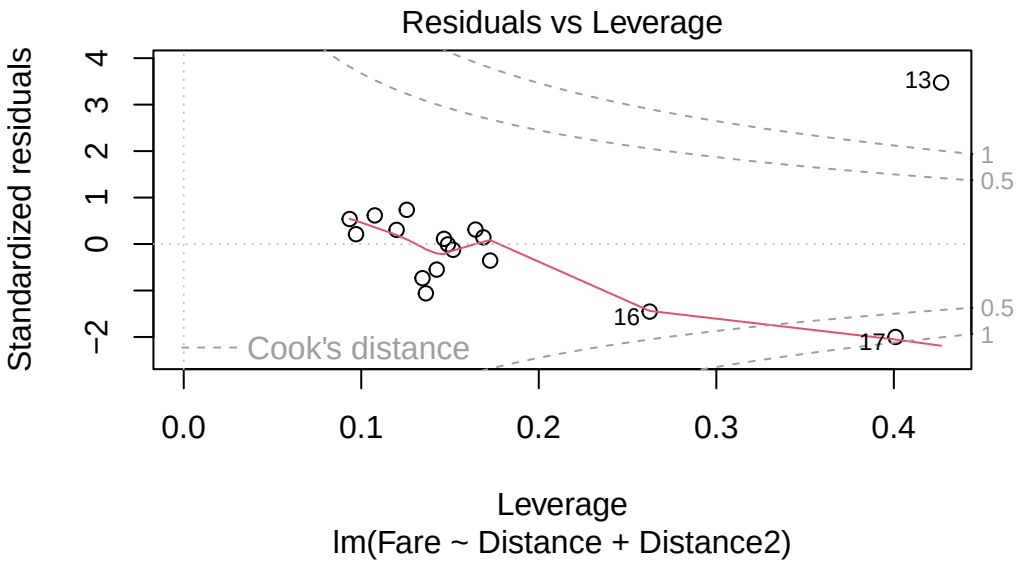
```
Distance2   -3.742e-06  9.726e-06  -0.385    0.706
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.72 on 14 degrees of freedom
Multiple R-squared:  0.994, Adjusted R-squared:  0.9932
F-statistic:  1165 on 2 and 14 DF,  p-value: 2.718e-16
```

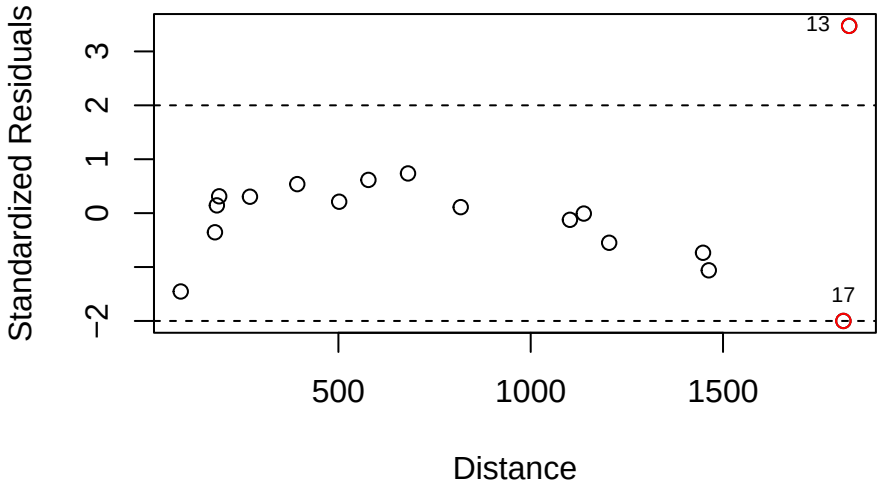
```r
plot(airline_quad)
```




Q−Q Residuals
Standardized residuals
−2
0
1
2
3
4
13
16
17
−2
−1
0
1
2
Theoretical Quantiles
lm(Fare ~ Distance + Distance2)


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
13
17
16
100
200
300
400
Fitted values
lm(Fare ~ Distance + Distance2)


Residuals vs Leverage
Standardized residuals
Cook's distance
13
16
17
1
0.5
0.5
1
0.0
0.1
0.2
0.3
0.4
Leverage
lm(Fare ~ Distance + Distance2)


Standardized Residuals
13
17
500
1000
1500
Distance