Regression Model for MPG

Andrew Hope

April 26, 2018

Executive Summary

This analysis aims to quantify the difference in miles per gallon (MPG) that will result from a car being equipped with an automatic transmission vs. a manual transmission. The analysis uses an ordinary linear regression, and concludes that the type of transmission does have a significant effect on MPG.

Load Data and Packages

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.4.3
```

```
df <- mtcars
```

Exploratory Analysis

```
dim(df)
```

```
## [1] 32 11
```

```
head(df)
```

```
##                   mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

The ‘am’ variable indicates the transmission (0 = automatic, 1 = manual).

```
table(df$am)
```

```
##
##  0  1
## 19 13
```

Model Generation

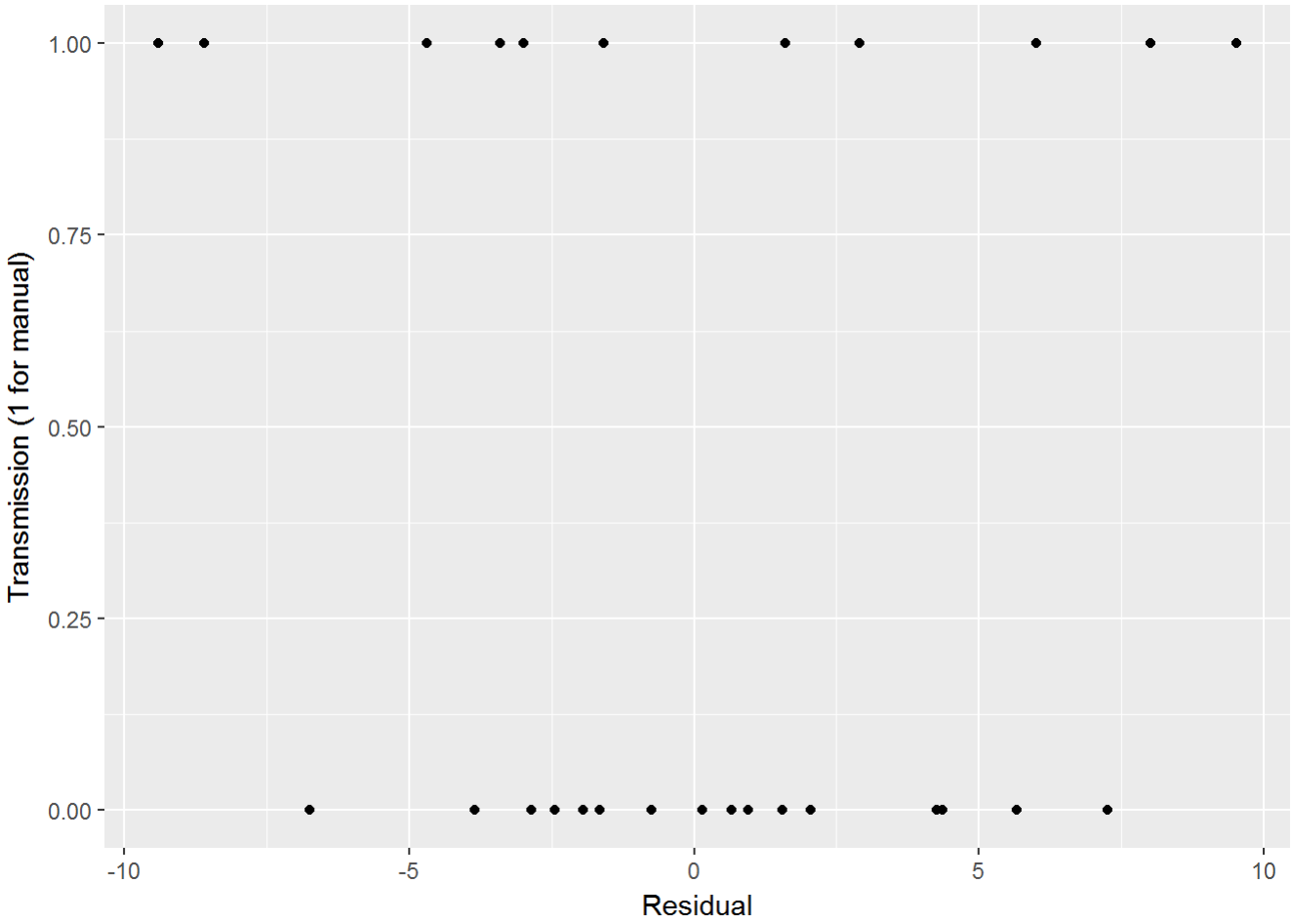
```
fit <- lm(mpg ~ am, data = df)
summary(fit)
```

```
##
## Call:
## lm(formula = mpg ~ am, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.147      1.125  15.247 1.13e-15 ***
## am             7.245      1.764   4.106 0.000285 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.902 on 30 degrees of freedom
## Multiple R-squared:  0.3598, Adjusted R-squared:  0.3385
## F-statistic: 16.86 on 1 and 30 DF,  p-value: 0.000285
```

This model tells us that the mean MPG for an automatic transmission is 17.147, and the mean MPG for a manual increases by 7.245 (to a total of 24.392).

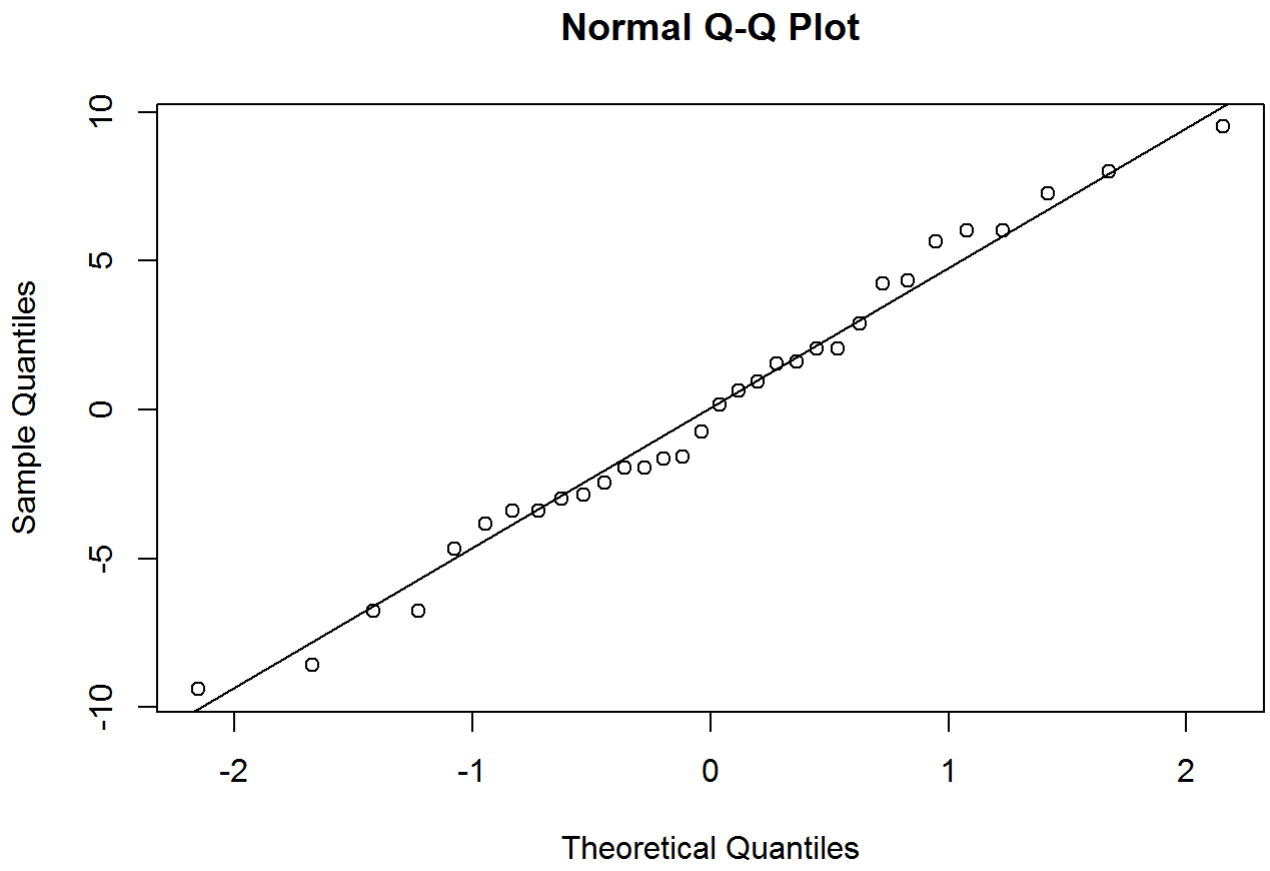
Residual Analysis

```
qplot(fit$residuals, df$am, ylab = "Transmission (1 for manual)", xlab = "Residual")
```



This residual plot shows values evenly distributed around 0. It appears that the variance of residuals is larger when am == 1. This difference is noticable but not dramatic.

```
qqnorm(fit$residuals)
qqline(fit$residuals)
```



A Q-Q plot of the residuals shows that residual values do not deviate from a normal distribution in a clear and systematic way. The residuals appear to be approximately normal.

Explanation of Uncertainty

The model's coefficient has a p-value of 0.000285, which is low enough to conclude that the transmission variable has a non-zero effect on MPG. However, the r-squared and residual standard error indicate that using the transmission variable as the only predictor will not provide an especially precise estimate of MPG.

The adjusted r-squared value is 0.3385, meaning 33% of the variability of MPG is explained by the transmission variable. There is significant variablility outside of the transmission variable that should be investigated.

The residual standard error is 4.9. This is low enough to confirm that the model outperforms a null hypothesis, but it would be reasonable to continue the analysis to look for a model with a lower residual standard error.