Regression Models Project

Beth Skierski

4/21/2021

Executive Summary

This analysis looks at a data set for a collection of cars (mtcars) and explores the relationship between transmission type (automatic or manual) and car fuel efficiency (MPG). The approach to analyze this question is to perform exploratory analysis, create a linear model with the terms of significance only, and to ultimately quantify the effect of transmission type on mpg.

The analysis concludes that transmission type does indeed have an effect on mpg, and that when transmission type is manual vs. automatic, fuel efficiency goes up about 2.94 miles per gallon.

Exploratory Analysis

First, I took a look at the data and plotted mpg as a function of transmission type. The first figure shown in the appendix suggests that transmission type does indeed have an effect on fuel efficiency.

Creation of Linear Model

Next, I created a linear model with the "lm" function, then performed a stepwise regression to determine which factors should be included in the final linear model.

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4811 -1.5555 -0.7257  1.4110  4.6610
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.6178     6.9596   1.382 0.177915
## wt           -3.9165     0.7112  -5.507 6.95e-06 ***
## qsec          1.2259     0.2887   4.247 0.000216 ***
## am            2.9358     1.4109   2.081 0.046716 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.459 on 28 degrees of freedom
## Multiple R-squared:  0.8497, Adjusted R-squared:  0.8336
## F-statistic: 52.75 on 3 and 28 DF,  p-value: 1.21e-11
```

The summary of the stepwise regression shows that the p-values are significant for wt(weight, 1000lbs), qsec(1/4 mile time), and am(transmission type, 0=automatic, 1=manual). Therefore, these three factors should be included in the linear model. The summary shows the formula with weights for each of these factors. As shown, the recommended formula for mpg is:

mpg= 9.6178 - 3.9165*wt* + *1.2259*qsec + 2.9358*am

Residual Plots and Diagnostics

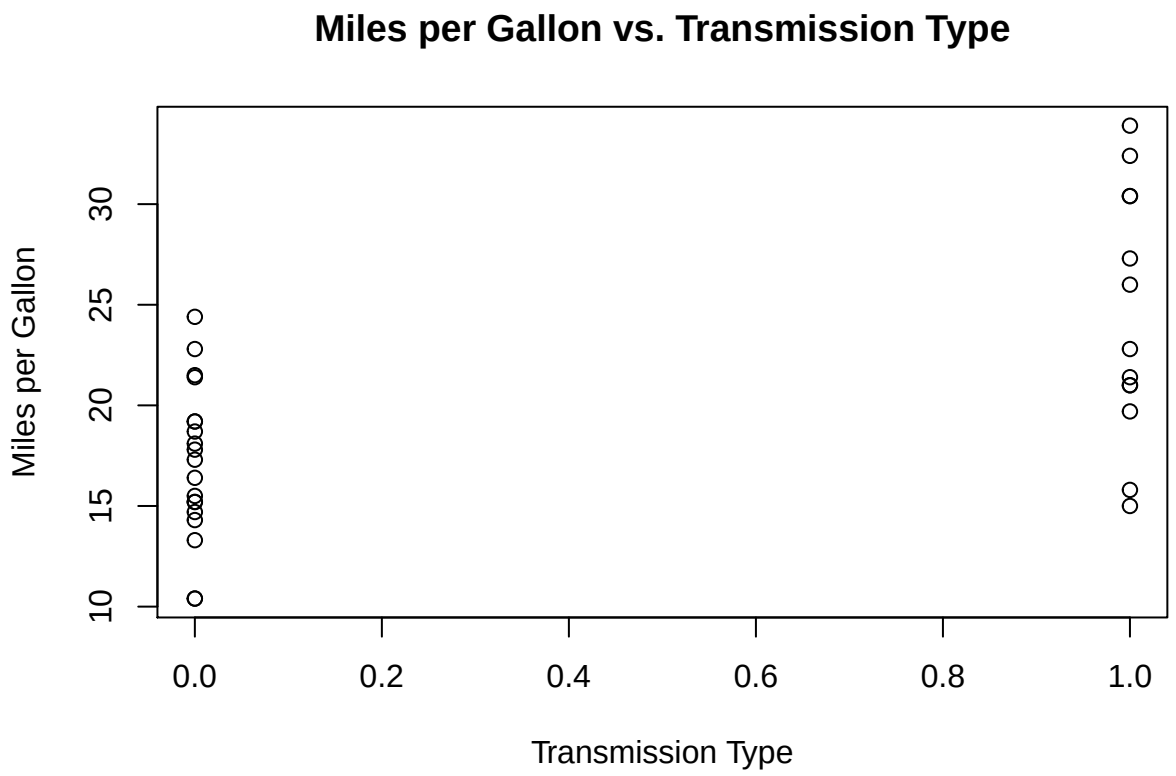
As shown above, the R-squared value of the linear model is ~0.85, which suggests that the model explains 85% of the total variability of the data. This is rather good, but not great. Furthermore, the residual plots shown in the appendix demonstrate that the model is a pretty good fit. Specifically, the "Residuals vs. Fitted" plot shows that the distance between the fitted values (the red line) and the residuals (the points) are about equal in both directions. There do not seem to be any extreme outliers; the Chrysler Imperial seems to have strong influence on the model, but the fit is generally good, even when it is included.

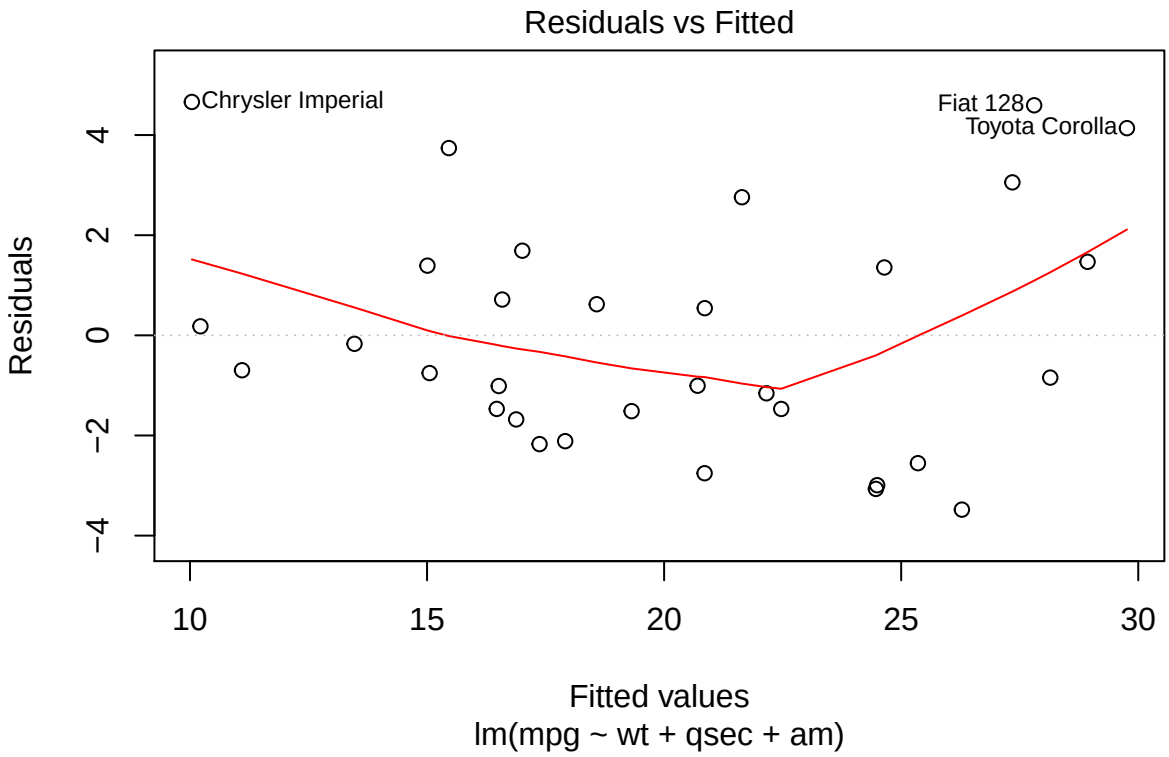
Conclusions

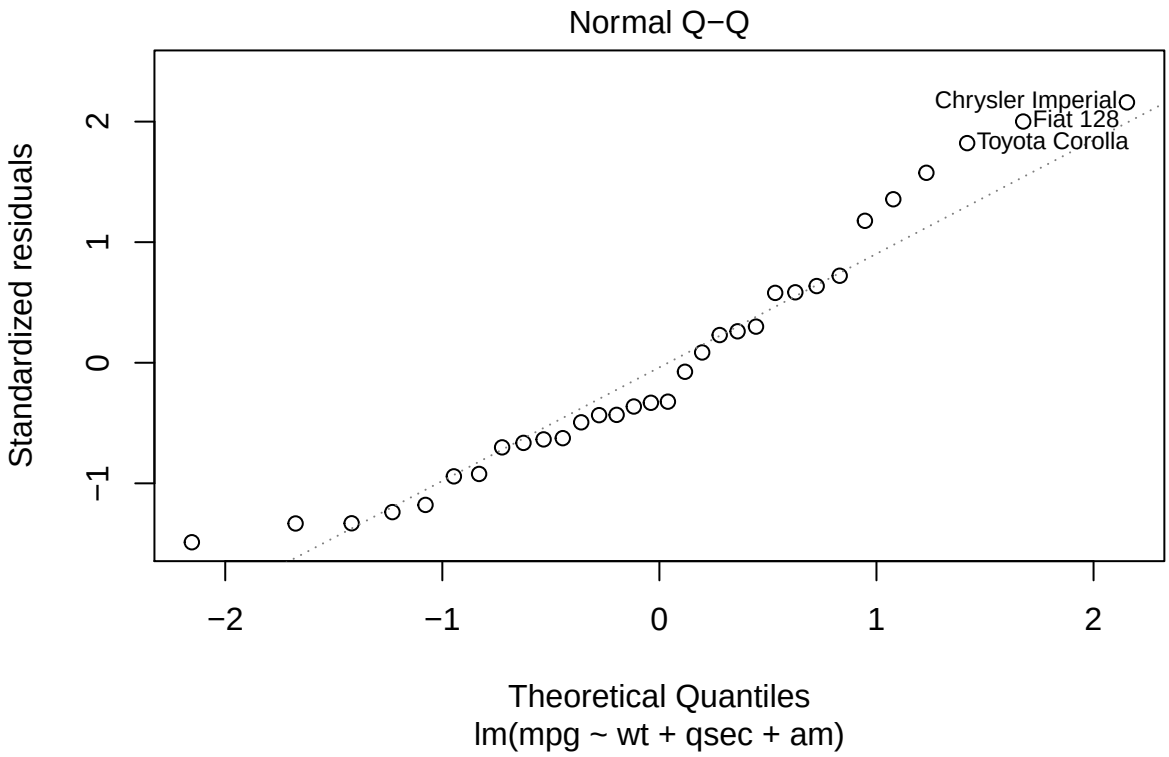
Given the good fit of the model, it can be concluded that transmission type does indeed have an effect on fuel efficiency. When transmission type is manual vs. automatic, it is expected that the fuel efficiency will go up by about 2.94 miles per gallon.

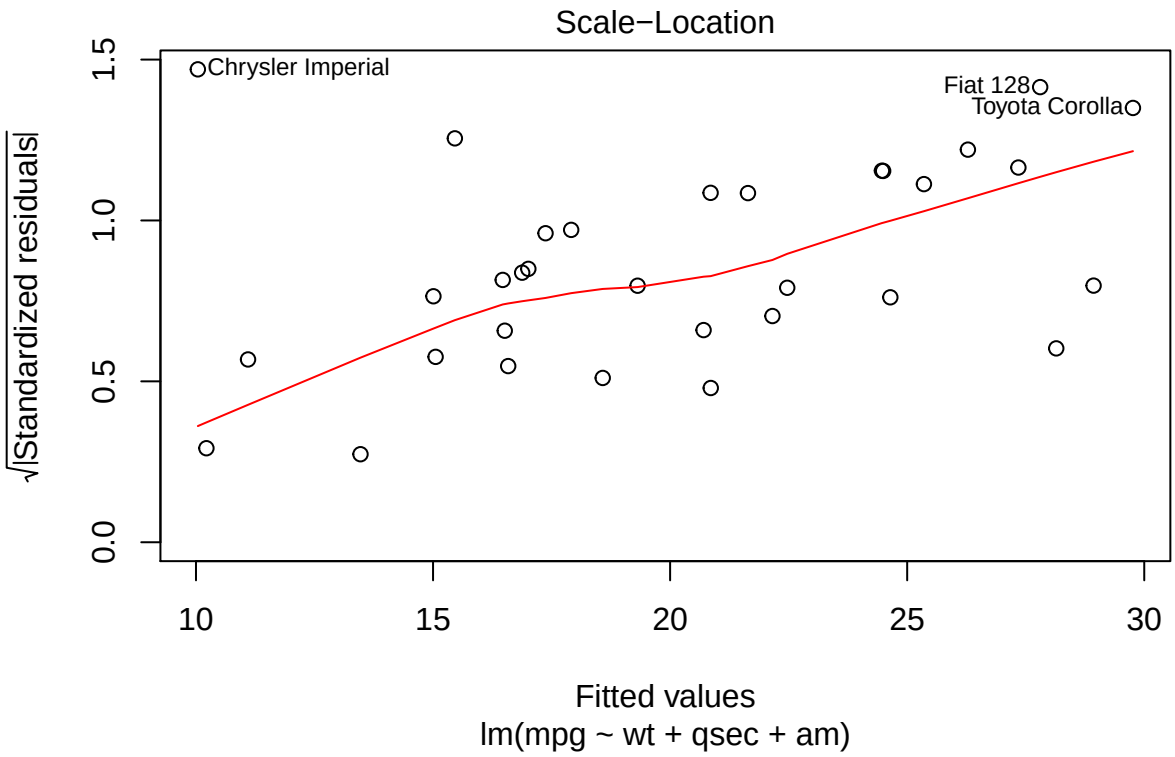
Appendix: Code and Figures

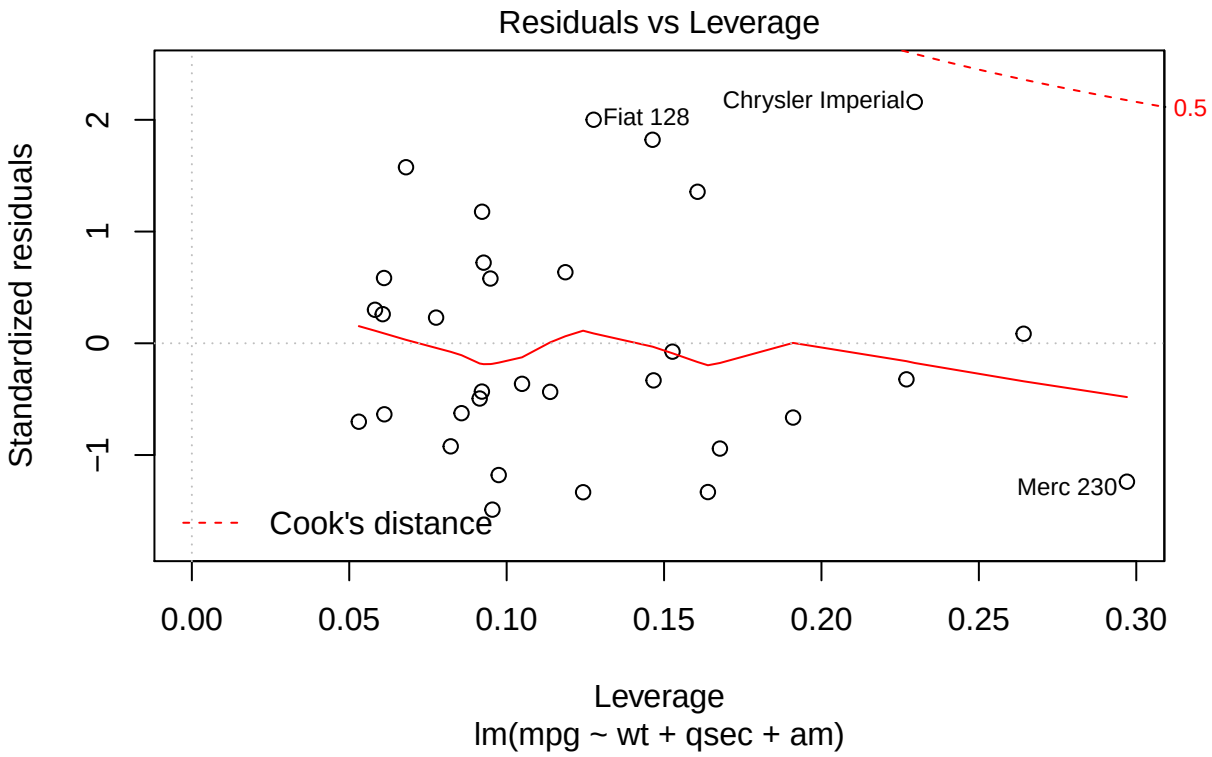
```
knitr::opts_chunk$set(echo = FALSE)
#create linear model with all factors
fit <- lm(mpg ~., data=mtcars)
#run stepwise algorithm to determine which factors to include
step1 <- step(fit, direction = "both")
summary(step1)
plot(mpg~am, mtcars, xlab="Transmission Type", ylab="Miles per Gallon", main="Miles per Gallon vs. Trans
plot(step1)
```




Residuals vs Fitted
Chrysler Imperial
Fiat 128
Toyota Corolla
Residuals
4
2
0
−2
−4
10
15
20
25
30
Fitted values
lm(mpg ~ wt + qsec + am)


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(mpg ~ wt + qsec + am)
Chrysler Imperial
Fiat 128
Toyota Corolla
−2
−1
0
1
2


Scale−Location
Chrysler Imperial
Fiat 128
Toyota Corolla
√|Standardized residuals|
0.0
0.5
1.0
1.5
10
15
20
25
30
Fitted values
lm(mpg ~ wt + qsec + am)


Residuals vs Leverage
Standardized residuals
Leverage
lm(mpg ~ wt + qsec + am)
Fiat 128
Chrysler Imperial
Merc 230
0.5
Cook's distance
2
1
0
−1
0.00
0.05
0.10
0.15
0.20
0.25
0.30