Regression Models - Analysis of MPG in mtcars Data Set

M. Nelson

November 18, 2015

Executive Summary

The purpose of this project is to explore the relationship between miles per gallon (MPG) and the set of variables available in the mtcars data set. The goal is specifically to address these two questions:

1. “Is an automatic or manual transmission better for MPG”
2. “Quantify the MPG difference between automatic and manual transmissions”

Exploratory Data Analysis

The first exploratory check is a T-test of a null hypothesis that there is no difference between automatic or manual transmissions:

```
## [1]  3.209684 11.280194
## attr(,"conf.level")
## [1] 0.95
```

Looking at transmission type alone would reject that null hypothesis and suggest that manual transmission provides 3-11 better MPG than automatic.

However, intuition suggests that other variable such as weight of the car may be strongly correlated with MPG. A second exploratory check in Figure 1 shows at some weights (see the indicated range), that automatics appear to have better MPG than manuals. So it is important to build a more complete model considering additional variables of the data set. Then check again under that improved model whether manual transmission still indicates higher MPG than automatic.

Coefficient Interpretation

Fitting all variable in mtcars to the output of mpg provides the set of coefficients shown in Figure 2. Scanning the p-values shows that wt, am, and qsec have the lowest values and are thus most statistically significant when considering all variables The corresponding slope coefficients are -3.7(wt), 2.5(am), and 0.8(qsec). These indicate that mpg will decrease 3.7 for each 1,000 lb increase in weight. Also that moving from automatic to manual will increase mpg by 2.5. And finally that mpg will increase by 0.8 for every additional second it takes the car to run a quarter mile. This also seems intuitively correct as the faster a car goes, the more quickly it may burn gas resulting in lower mpg.

Model Selection

Model selection will use a nested likelihood ratio test to select which variables to include. This will compare four fitted models:

1. am (only the transmission variable.)
2. am + wt + qsec (The suspected best set based on the p-value analysis above.)
3. am + wt + qsec + hp (Adding in hp which had the next lowest p-value.)
4. . (All available variables.)

Figure 3 indicates that using only the am+wt+qsec variables gives a very low p-value (8.025e-08). But adding the hp variable has a much higher p-value (.265) and all variables even higher (0.9308). So the model chosen is the one that includes only transmission, weight, and quarter mile seconds.

Residual Diagnostics

Figure 4 is a series of plots to diagnose residuals. In the standardized residual plot, there seems to be a slight upward sloping trend in the size of the residual value as the fitted values of mpg increases. This could indicate that some other variable not available in this data set is influencing this trend. However, the residuals versus leverage chart shows that these data points are well-behaved within this model. None exert unusually high influence to the model as measured against the Cook's Distance line plotted at 0.5.

Conclusions and Uncertainty

With the selected model, the fitted coefficients are shown here:

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)  9.617781  6.9595930  1.381946 1.779152e-01
## am           2.935837  1.4109045  2.080819 4.671551e-02
## wt          -3.916504  0.7112016 -5.506882 6.952711e-06
## qsec         1.225886  0.2886696  4.246676 2.161737e-04
```

This indicates that moving from automatic to manual transmission gives an increase of 2.9 MPG. The p-value for the transmission variable is 0.047. This is outside the 95% confidence interval of the null hypothesis that there is no difference in MPG due to transmission type. The null hypothesis is thus rejected and the analysis concludes that manual transmissions provide better MPG than automatic.

Responses to the specific questions of this project:

1. "Is an automatic or manual transmission better for MPG"

   Manual transmission generally provides a higher MPG than automatic transmission.

2. "Quantify the MPG difference between automatic and manual transmissions"

   Within a model that also includes wt and qsec, moving from automatic to manual transmission is expected to increase MPG by 2.9.

Appendix

Figure 1:

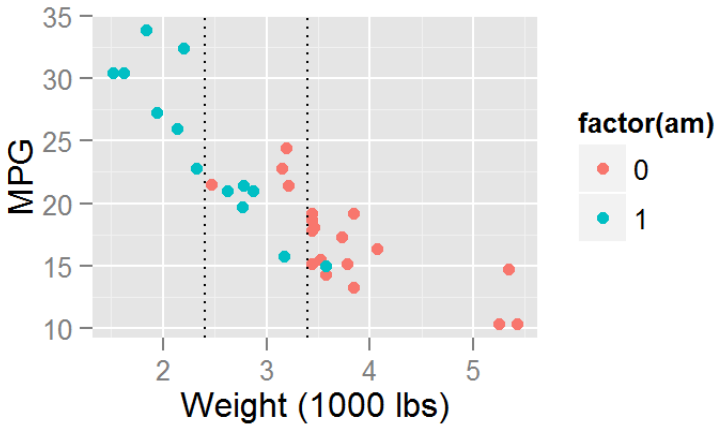


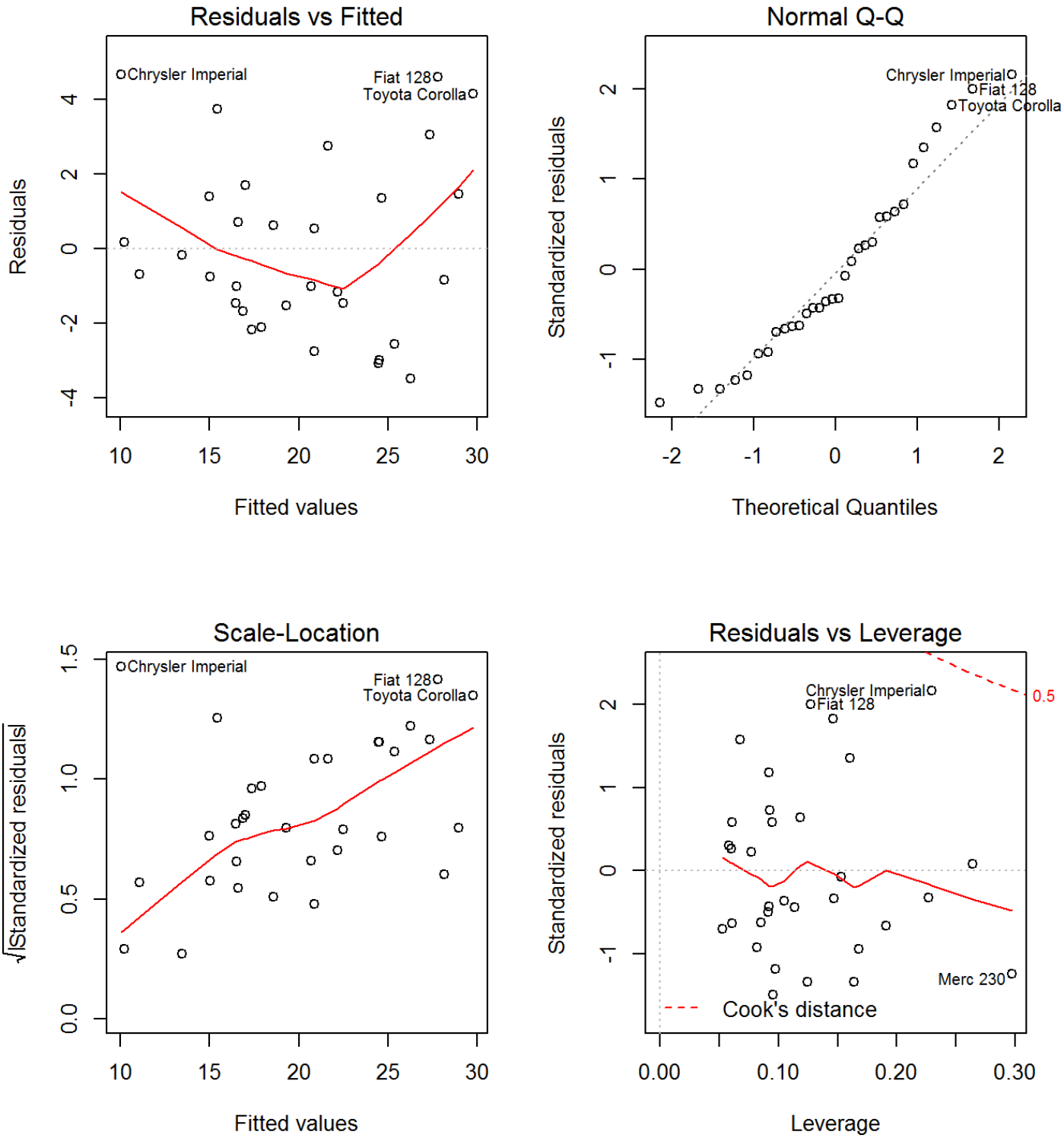
Figure 2:

```
##                Estimate  Std. Error    t value   Pr(>|t|)
## (Intercept) 12.30337416 18.71788443  0.6573058 0.51812440
## cyl         -0.11144048  1.04502336 -0.1066392 0.91608738
## disp         0.01333524  0.01785750  0.7467585 0.46348865
## hp          -0.02148212  0.02176858 -0.9868407 0.33495531
## drat         0.78711097  1.63537307  0.4813036 0.63527790
## wt          -3.71530393  1.89441430 -1.9611887 0.06325215
## qsec         0.82104075  0.73084480  1.1234133 0.27394127
## vs           0.31776281  2.10450861  0.1509915 0.88142347
## am           2.52022689  2.05665055  1.2254035 0.23398971
## gear         0.65541302  1.49325996  0.4389142 0.66520643
## carb        -0.19941925  0.82875250 -0.2406258 0.81217871
```

Figure 3:

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + wt + qsec
## Model 3: mpg ~ am + wt + qsec + hp
## Model 4: mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 720.90
## 2     28 169.29  2    551.61 39.2687 8.025e-08 ***
## 3     27 160.07  1      9.22  1.3127    0.2648
## 4     21 147.49  6     12.57  0.2983    0.9308
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 4:


Residuals vs Fitted
Chrysler Imperial
Fiat 128
Toyota Corolla
Residuals
Fitted values
Normal Q-Q
Chrysler Imperial
Fiat 128
Toyota Corolla
Standardized residuals
Theoretical Quantiles
Scale-Location
Chrysler Imperial
Fiat 128
Toyota Corolla
√|Standardized residuals|
Fitted values
Residuals vs Leverage
Chrysler Imperial
Fiat 128
Merc 230
0.5
Cook's distance
Standardized residuals
Leverage