Motor Trend Car Road Tests - Analysis of Fuel Efficiency

Candice MH

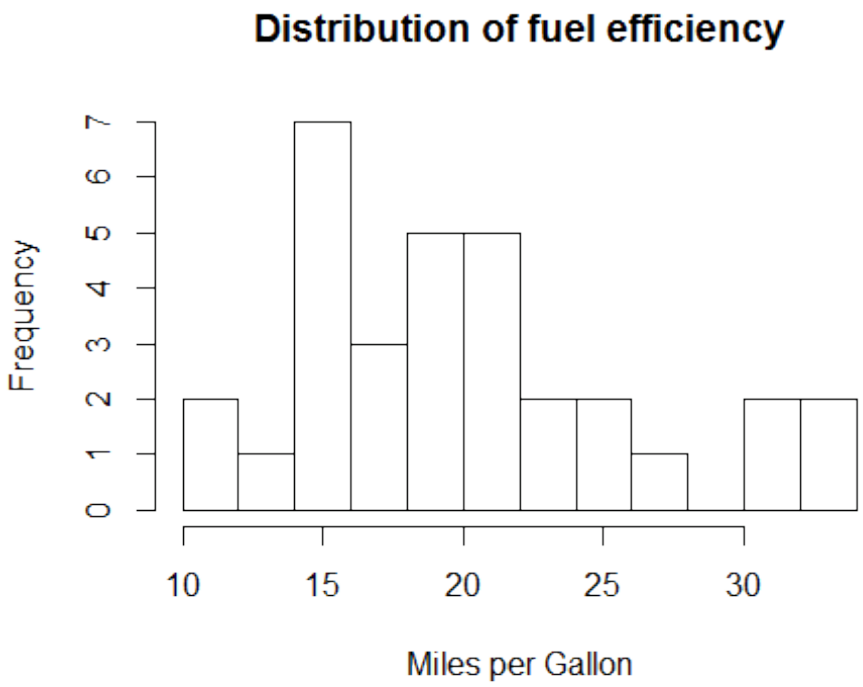
Sunday, November 09, 2014

Executive Summary

This report describes a linear regression model which predicts fuel efficiency (miles per US gallon) based on a number of other predictors. In particular, this report describes the impact of transmission on fuel economy and whether or not automatic or manual transmissions have better fuel economy. The report finds that, for vehicles of equal weight and speed (quarter mile time), automatic vehicles should be chosen over manual vehicles if fuel efficiency is a consideration. The author acknowledges the source of the data from https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/mtcars.html.

Exploratory Data Analysis

Miles per US gallon (*mpg*) varies between 10.4 and 33.9 with a mean of 20.0906 and standard deviation equal to 6.0269. *mpg* is displayed in a histogram below, and reveals a distribution that is roughly normal apart from a slight right skew.



The explanatory variables available to perform this analysis are *cyl*, *disp*, *hp*, *drat*, *wt*, *qsec*, *vs*, *am*, *gear* and *carb*, where *cyl* is the number of cylinders, *disp* is the displacement, *hp* is the gross horsepower, *drat* is the rear axle ratio, *wt* is measured in thousand pounds, *qsec* is the quarter mile time, *vs* is V/S, *am* is a categorical variable that takes the value 0 for automatic vehicles and 1 for manual vehicles, *gear* is the number of forward gears, and *carb* is the number of carburetors.

Without accounting for relationships between explanatory variables, the correlation matrix shows us that *cyl*, *disp*, *hp*, *wt* and *carb* are all negatively related to *mpg*, and *drat*, *qsec*, *vs*, *am* and *gear* are all negatively related to *mpg*. The variable with strongest correlation to fuel efficiency is weight at -0.8677. The correlation between transmission and fuel efficiency is 0.5998. There is a large problem with multicolinearity evident in this data set, illustrated by the cross correlation of 0.902 between cylinders and displacement.

Transformation is used to mitigate the numerical issues caused by multicolinearity present between explanatory variables. This data set contains a number of categorical variables (*cyl*, *vs*, *am* and *gear*) that are not be transformed. The remaining continuous variables (*carb*, *disp*, *drat*, *hp*, *qsec* and *wt*) are standardised. These variables are entered into the model with a postscript of 2.

Model Selection and Results

Stepwise model assessment is used to automate covariate selection, thereby fitting multiple models but ultimately selecting the model with the minimal AIC. The resultant model is calculated but suppressed from output. See the appendix for details.

The R-squared of the model is 0.8497 indicating a relatively strong model fit. Only weight, quarter mile time and transmission are significant in predicting fuel efficiency, with p-values<0.05. Analysis of the residuals from this model demonstrates that assumptions of normally distributed errors with constant error variance appears to be met (see appendix). Data are presumed to be independent.

Conclusion

Co-efficients of weight and quarter mile time have been adjusted to account for the standardisation so that intuitive interpretation of co-efficients can be achieved. For every additional thousand pounds of weight, fuel efficiency declines by 3.9165 miles per US gallon, after accounting for the impact of quarter mile time and transmission. For every one second improvement in quarter mile time, fuel efficiency increases by 1.2259 miles per US gallon, after accounting for the impact of weight and transmission.

For manual transmission vehicles, after accounting for the impact of weight and quarter mile time, the fuel efficiency in miles per US gallon is 2.1906 higher than automatic transmission vehicles. For vehicles of equal weight and quarter mile time, automatic transmission should be chosen if fuel efficiency is a consideration.

Appendix

Data Summary

```
summary(mtcars)
```

```
##       mpg             cyl             disp             hp       
##  Min.   :10.4   Min.   :4.00   Min.   : 71.1   Min.   : 52.0  
##  1st Qu.:15.4   1st Qu.:4.00   1st Qu.:120.8   1st Qu.: 96.5  
##  Median :19.2   Median :6.00   Median :196.3   Median :123.0  
##  Mean   :20.1   Mean   :6.19   Mean   :230.7   Mean   :146.7  
##  3rd Qu.:22.8   3rd Qu.:8.00   3rd Qu.:326.0   3rd Qu.:180.0  
##  Max.   :33.9   Max.   :8.00   Max.   :472.0   Max.   :335.0  
##       drat             wt             qsec             vs        
##  Min.   :2.76   Min.   :1.51   Min.   :14.5   Min.   :0.000  
##  1st Qu.:3.08   1st Qu.:2.58   1st Qu.:16.9   1st Qu.:0.000  
##  Median :3.69   Median :3.33   Median :17.7   Median :0.000  
##  Mean   :3.60   Mean   :3.22   Mean   :17.8   Mean   :0.438  
##  3rd Qu.:3.92   3rd Qu.:3.61   3rd Qu.:18.9   3rd Qu.:1.000  
##  Max.   :4.93   Max.   :5.42   Max.   :22.9   Max.   :1.000  
##        am              gear            carb      
##  Min.   :0.000   Min.   :3.00   Min.   :1.00  
##  1st Qu.:0.000   1st Qu.:3.00   1st Qu.:2.00  
##  Median :0.000   Median :4.00   Median :2.00  
##  Mean   :0.406   Mean   :3.69   Mean   :2.81  
##  3rd Qu.:1.000   3rd Qu.:4.00   3rd Qu.:4.00  
##  Max.   :1.000   Max.   :5.00   Max.   :8.00
```

Correlation Analysis

```
cor(mtcars)
```

```
##          mpg     cyl    disp      hp     drat      wt    qsec
vs
## mpg   1.0000 -0.8522 -0.8476 -0.7762  0.68117 -0.8677  0.4187
0.6640
## cyl  -0.8522  1.0000  0.9020  0.8324 -0.69994  0.7825 -0.5912
-0.8108
## disp -0.8476  0.9020  1.0000  0.7909 -0.71021  0.8880 -0.4337
-0.7104
## hp   -0.7762  0.8324  0.7909  1.0000 -0.44876  0.6587 -0.7082
-0.7231
## drat  0.6812 -0.6999 -0.7102 -0.4488  1.00000 -0.7124  0.0912
0.4403
## wt   -0.8677  0.7825  0.8880  0.6587 -0.71244  1.0000 -0.1747
-0.5549
## qsec  0.4187 -0.5912 -0.4337 -0.7082  0.09120 -0.1747  1.0000
```

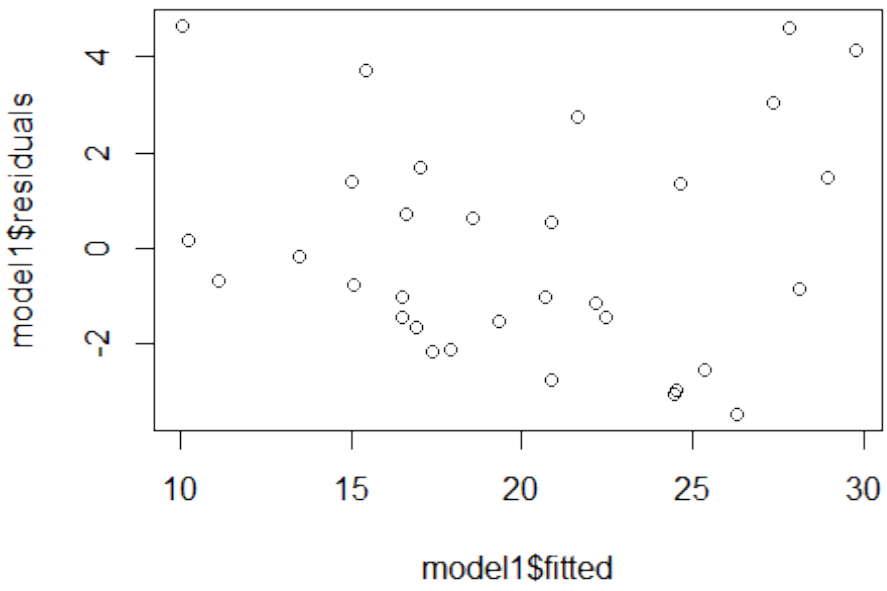
```
0.7445
## vs    0.6640 -0.8108 -0.7104 -0.7231  0.44028 -0.5549  0.7445
1.0000
## am    0.5998 -0.5226 -0.5912 -0.2432  0.71271 -0.6925 -0.2299
0.1683
## gear  0.4803 -0.4927 -0.5556 -0.1257  0.69961 -0.5833 -0.2127
0.2060
## carb -0.5509  0.5270  0.3950  0.7498 -0.09079  0.4276 -0.6562
-0.5696
##            am    gear     carb
## mpg   0.59983  0.4803 -0.55093
## cyl  -0.52261 -0.4927  0.52699
## disp -0.59123 -0.5556  0.39498
## hp   -0.24320 -0.1257  0.74981
## drat  0.71271  0.6996 -0.09079
## wt   -0.69250 -0.5833  0.42761
## qsec -0.22986 -0.2127 -0.65625
## vs    0.16835  0.2060 -0.56961
## am    1.00000  0.7941  0.05753
## gear  0.79406  1.0000  0.27407
## carb  0.05753  0.2741  1.00000
```

Final model - residual checking

```
summary(model1)

##
## Call:
## lm(formula = mpg ~ wt2 + qsec2 + as.factor(am), data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.481 -1.556 -0.726  1.411  4.661
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)      18.898      0.719   26.27  < 2e-16 ***
## wt2              -3.832      0.696   -5.51    7e-06 ***
## qsec2             2.191      0.516    4.25  0.00022 ***
## as.factor(am)1    2.936      1.411    2.08  0.04672 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.46 on 28 degrees of freedom
## Multiple R-squared:  0.85,   Adjusted R-squared:  0.834
## F-statistic: 52.7 on 3 and 28 DF,  p-value: 1.21e-11

plot(model1$fitted,model1$residuals)
```




```
qqnorm(model1$residuals)
```