Fuel Economy and Analysis of Motor Trend Data Transmissions

Executive Summary

The contribution of automatic or manual transmissions (AM) to miles per gallon (MPG) was analyzed using the 1974 *Motor Trend Magazine* data of 11 variables on 32 automobiles. The resulting analysis of the *mtcars* dataset in the standard R datasets package indicate:

- Cars with manual transmissions had better MPGs than automatic transmissions
- Holding all other variables constant, cars with manual transmissions had 7.25 MPG better fuel economy
- There were several confounding variables including weight in lbs/1000 and quarter mile time in seconds which were included in the final model in addition to transmission type
- The final model explained nearly 0.85 of the variance as described by the R^2 value
- In the final model including these additional variables, manual transmissions had better mileage but only about 2.9 MPG different than automatic transmission cars holding all other variables constant

Data Transformation and Exploration

Several of the 11 variables in the dataset were initially numeric and had to be transformed into factor variables to be appropriately analyzed.

```
mtcars <- datasets::mtcars  # start with the baseline data from the datasets package
mtcars$am <- factor(mtcars$am, levels=c(0,1), labels=c("Automatic","Manual"))
mtcars$cyl <- factor(mtcars$cyl); mtcars$vs <- factor(mtcars$vs)
```

As can be seen in **Appendix Figure 1**, considering no other variables, there is a significant difference in MPG based on transmission type.

Linear Models and Selection

Base model: A baseline model was initially developed using just transmission type for an independent variable and MPG as the dependent.

```
##
## Call:
## lm(formula = mpg ~ am, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.392 -3.092 -0.297  3.244  9.508
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    17.15       1.12   15.25  1.1e-15 ***
## amManual        7.24       1.76    4.11  0.00029 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.9 on 30 degrees of freedom
## Multiple R-squared:  0.36,   Adjusted R-squared:  0.338
## F-statistic: 16.9 on 1 and 30 DF,  p-value: 0.000285
```

```
##              2.5 % 97.5 %
## (Intercept) 14.851  19.44
## amManual     3.642  10.85
```

As can be seen, with no other variables there is a statistically significant difference of 7.2449 increase in MPG for manual transmissions (p=0.000285, conf interval= (3.6415 to 10.8484)). However, the model only explains 0.3598 of the variance in mpg based on the R^2 value. (Residual tests of this model may be found in **Appendix Figure 3**.)

Alternate model: However, in reviewing other parameters in the dataset, it was determined that including only the transmission type created a bias model. We created several models using and then verified it with the *step()* function (see appendix for optional exploratory analysis) and looked at the significance of comparing models using ANOVA tests, and the variance inflation to find a more parsimonious, interpretable representation. We also looked at outliers using *hatvalues* function, *dfbetas* and *PRESS* valueson the most significant model to see if any of the model coefficients were significantly influence by specific outlier values.

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + 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)    9.618      6.960    1.38  0.17792
## wt            -3.917      0.711   -5.51    7e-06 ***
## qsec           1.226      0.289    4.25  0.00022 ***
## amManual       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
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ wt + qsec + am
##   Res.Df RSS Df Sum of Sq    F  Pr(>F)
## 1     30 721
## 2     28 169  2       552 45.6 1.6e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

It was determined that a model using transmission type, quarter mile time in seconds, and weight in lbs/1000 had a significantly better result (p-value for each coefficient < 0.05, R^2 = 0.8497) and was statistically significantly better than the base model based on anova tests (p = 1.6e-9). The confidence interval for manual transmissions was reduced to 2.9358 (conf interval = 0.0457 to 5.8259) with these additional parameters added to the model. The residual tests of this new model are available in **Appendix Figure 4**.

Conclusions

Taken on its own, manual transmission cars would save on average 7.2449 MPG if it was the only modeled independent variable. However, a significantly better model with quarter mile time, weight and transmission type would explain about 0.8497 of the variance in MPG. Holding quarter mile time and weight constant, this model results in manual transmissions having 2.9358 MPG better than automatic transmissions.

In reviewing the residuals and PRESS values, the Datsun 710, Merc 230, Toyota Corona & Corolla, Fiat 128 and Chrysler Imperial had impacts on the regression (see **Appendix Figure 5**).

The reproducible research used to produce this report can be found at https://github.com/svonkleeck/RegressionProject.

Appendix

The following information is available as an appendix for review or reference.

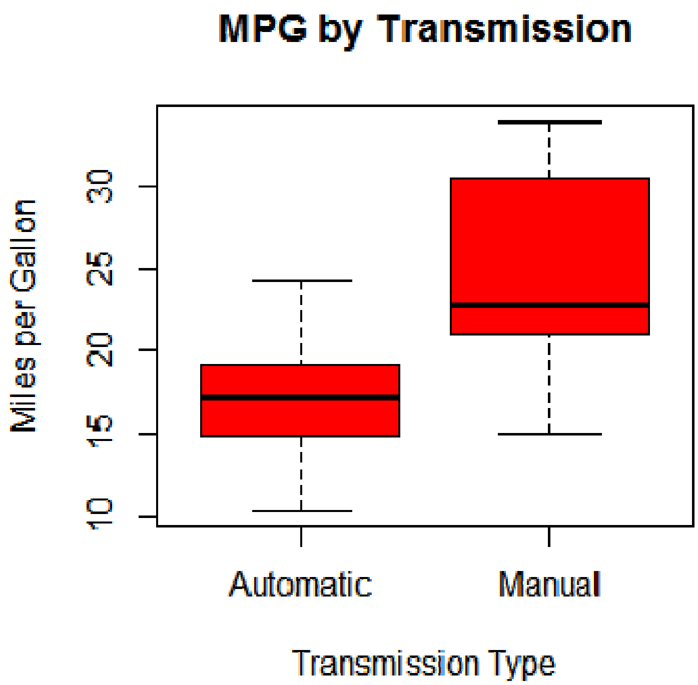

Figure 1: Automatic vs. Manual transmissions and Miles per Gallon with no other variables

```
options(width=90)
summary(mtcars)
```

```
##       mpg         cyl         disp             hp             drat             wt
##  Min.   :10.4   4:11   Min.   : 71.1   Min.   : 52.0   Min.   :2.76   Min.   :1.51
##  1st Qu.:15.4   6: 7   1st Qu.:120.8   1st Qu.: 96.5   1st Qu.:3.08   1st Qu.:2.58
##  Median :19.2   8:14   Median :196.3   Median :123.0   Median :3.69   Median :3.33
##  Mean   :20.1          Mean   :230.7   Mean   :146.7   Mean   :3.60   Mean   :3.22
##  3rd Qu.:22.8          3rd Qu.:326.0   3rd Qu.:180.0   3rd Qu.:3.92   3rd Qu.:3.61
##  Max.   :33.9          Max.   :472.0   Max.   :335.0   Max.   :4.93   Max.   :5.42
##       qsec        vs              am          gear             carb
##  Min.   :14.5   0:18   Automatic:19   Min.   :3.00   Min.   :1.00
##  1st Qu.:16.9   1:14   Manual   :13   1st Qu.:3.00   1st Qu.:2.00
##  Median :17.7                         Median :4.00   Median :2.00
##  Mean   :17.8                         Mean   :3.69   Mean   :2.81
##  3rd Qu.:18.9                         3rd Qu.:4.00   3rd Qu.:4.00
##  Max.   :22.9                         Max.   :5.00   Max.   :8.00
```

Figure 2: Summary of mtcars data used for analysis

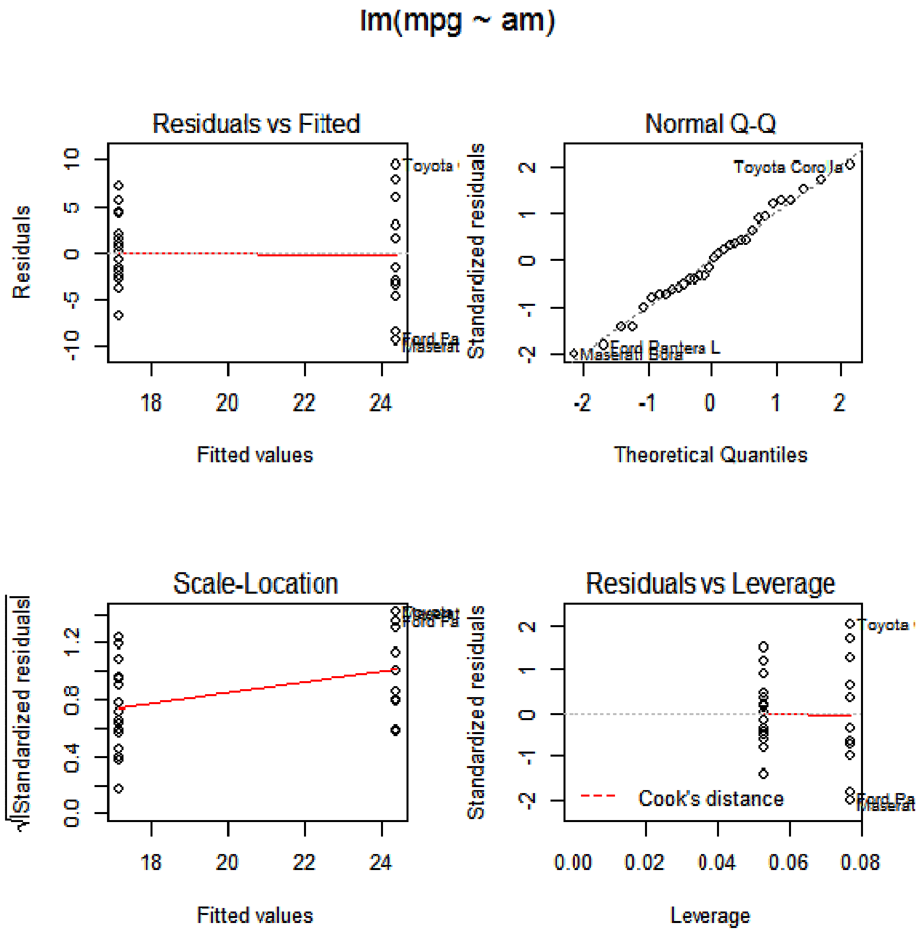

Figure 3: Residuals analysis of base model including just mpg and transmission type

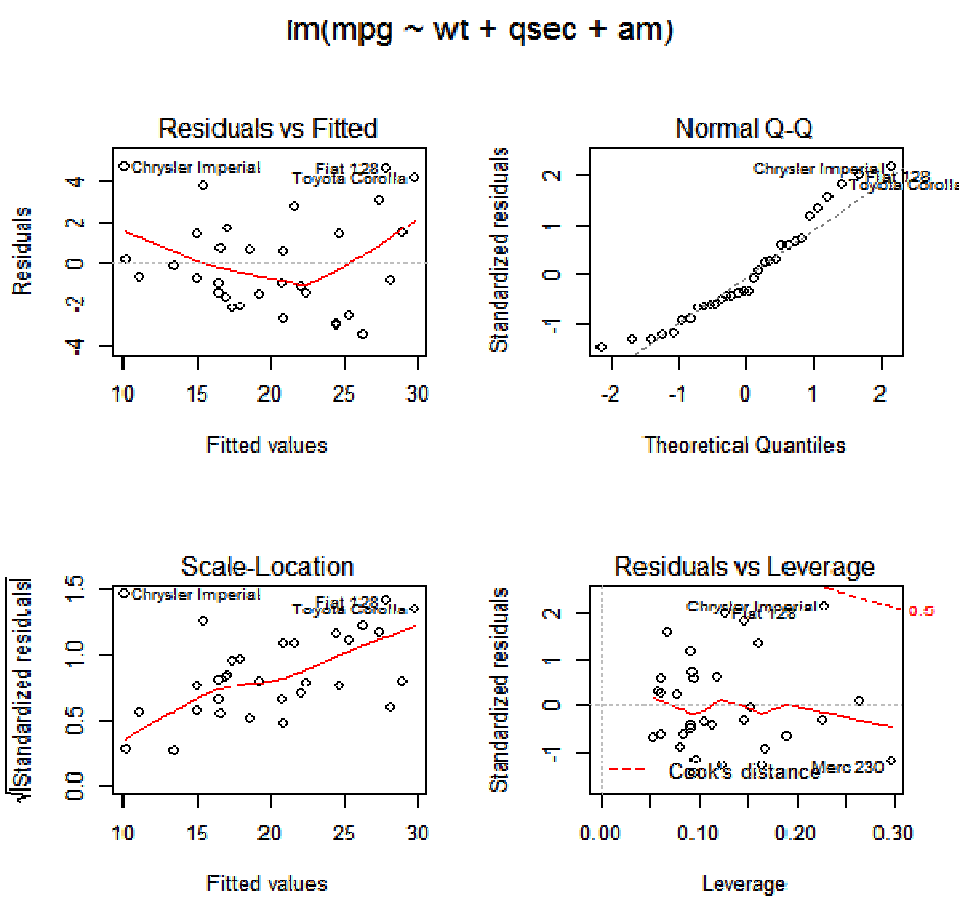

Figure 4: Residuals analysis of best model including mpg, weight, quarter mile time and transmission type

```
# influence potential
hat <- hatvalues(bestlm)
#hat[order(hat, decreasing = TRUE)]
PRESS<-resid(bestlm)/(1-hat)
PRESS[order(PRESS)]
```

```
##         Datsun 710            Merc 230       Toyota Corona          Volvo 142E
##            -3.8485             -3.6326             -3.5805             -3.5025
##            Valiant      Ford Pantera L         AMC Javelin       Maserati Bora
##            -3.0515             -2.5397             -2.3671             -1.8166
##         Merc 450SLC          Merc 280C           Mazda RX4       Mazda RX4 Wag
##            -1.7734             -1.6129             -1.6082             -1.2748
##        Ferrari Dino    Dodge Challenger           Fiat X1-9   Cadillac Fleetwood
##            -1.1354             -1.1151             -0.9434             -0.9024
##          Duster 360          Camaro Z28 Lincoln Continental      Hornet 4 Drive
##            -0.8835             -0.1998              0.2447              0.5882
##            Merc 280         Merc 450SL          Merc 450SE       Porsche 914-2
##             0.6614              0.7594              1.4825              1.4974
##         Honda Civic   Hornet Sportabout          Merc 240D        Lotus Europa
##             1.6658              1.8631              3.0383              3.6392
##    Pontiac Firebird      Toyota Corolla           Fiat 128    Chrysler Imperial
##             4.0128              4.8475              5.2669              6.0504
```

```
dfbetas(bestlm)
```

```
##                     (Intercept)         wt      qsec amManual
## Mazda RX4            -0.0349775 -6.600e-03  0.049747 -0.08285
## Mazda RX4 Wag         0.0319133 -5.892e-02 -0.012225 -0.11094
## Datsun 710            0.1889626 -6.962e-02 -0.214995 -0.29858
## Hornet 4 Drive       -0.0001294 -2.117e-02  0.016780 -0.03323
## Hornet Sportabout     0.1632501 -1.202e-01 -0.136542 -0.17418
## Valiant               0.1738851 -3.047e-02 -0.245818  0.05241
## Duster 360           -0.1128598  6.659e-02  0.106922  0.09601
## Merc 240D            -0.0706888 -8.215e-02  0.166362 -0.14237
## Merc 230              0.5481272 -1.258e-01 -0.699292 -0.06704
## Merc 280              0.0202295 -2.343e-02 -0.008336 -0.04170
## Merc 280C            -0.0067919  3.576e-02 -0.025251  0.07973
## Merc 450SE            0.0238269  2.583e-02 -0.030098 -0.04203
## Merc 450SL            0.0276831 -1.331e-02 -0.022421 -0.04075
## Merc 450SLC          -0.0259822  5.012e-03  0.014273  0.07138
## Cadillac Fleetwood    0.1002672 -1.506e-01 -0.064948 -0.08089
## Lincoln Continental  -0.0290793  4.481e-02  0.018084  0.02475
## Chrysler Imperial    -0.6259919  1.094e+00  0.336678  0.56264
## Fiat 128             -0.4244326  1.290e-01  0.496886  0.47657
## Honda Civic           0.0292163 -1.109e-01  0.017887  0.01695
## Toyota Corolla       -0.3208270 -5.112e-02  0.451493  0.31746
## Toyota Corona        -0.1449114  4.065e-01 -0.051887  0.40504
## Dodge Challenger     -0.0968076  6.493e-02  0.084185  0.09961
## AMC Javelin          -0.1801169  1.401e-01  0.143630  0.20811
## Camaro Z28           -0.0246187  1.072e-02  0.025203  0.01890
## Pontiac Firebird      0.2130648 -6.925e-02 -0.205751 -0.23603
## Fiat X1-9             0.0253263  1.952e-02 -0.043551 -0.04460
## Porsche 914-2         0.0787753 -6.933e-02 -0.069045  0.02157
## Lotus Europa          0.3666001 -4.292e-01 -0.266924 -0.13714
## Ford Pantera L       -0.1560943 -6.170e-02  0.238474 -0.14025
## Ferrari Dino         -0.0579853  3.523e-06  0.076881 -0.04804
## Maserati Bora        -0.0397653 -1.325e-01  0.120368 -0.16843
## Volvo 142E            0.3119833 -2.543e-01 -0.283781 -0.41565
```

Figure 5: Press values, dfbetas influence and hatvalues of influence on the best model

We can see the influence specific vehicles had on each coefficient.

Exploring aternate models

Optionally how to find the best fit using the step() function. This is not part of the official analysis but was used to verify the optimal model selection.

```
library(car)
(summaryBestlmStep <- summary(bestlmStep <- step(lm(mpg ~ ., data=mtcars))))
```

Other significant methods explored:

```
fitAll <- lm(mpg ~ ., data=mtcars)  # no good
summary(fitAll); vif(fitAll)
(summaryBest2lm <- summary(best2lm <- lm(mpg ~ am + hp, data=mtcars)));
anova (baselm, best2lm)
```

Note AM and HP leaves AM signifcant and explains 78% variance but not best model as confirmed by step analysis.