Exploring the Relationship Between Auto Transmission Type and MPG

Executive Summary

In this study, we showed that changing from an automatic transmission to a manual transmission, all other parameters fixed, produces an increase in *mpg*. We performed this analysis by using backward stepwise regression from a full linear model to a reduced model, and detailed the goodness-of-fit using a model comparison ANOVA. We substantiated the assumption of normality of the reduced model by showing that the regression residual plot had no discernable trends. Finally, we created a 95% confidence interval for each of the regression coefficients showing that there is an increase in *mpg* by using a manual transmission over an automatic transmission.

Exploratory Data Analysis and Data Summary

This data set has 0 incomplete cases, and each variable's summary statistics are

```
##       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
```

In the Appendix, a scatterplot matrix is shown for the full dataset. In the plot, we see many variables that have a linear relationship with *mpg* along with *disp* and *hp* potentially having a quadratic relationship with *mpg*. However, we will only fit a model with linear terms since we are only interested in the affect of *am* on *mpg*, which appears to be linear.

Model Fit and Data Analysis

To study the affect of transmission type on *mpg*, we will fit the model as mentioned above as the full model, followed by fitting a reduced model using backward stepwise regression. The summary of the full model fit is in the Appendix, while the summary of the reduced model is

```
##
## 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 ***
## am             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
```

From this summary, we see that the model is statistically significant with such a low p-value, and if, holding all other regressors fixed, changing from an automatic transmission to a manual transmission adds 2.9358 *mpg* on the average. Additionally, by viewing the residual plot in the Appendix, we see that there is no discernable skweness from the model, thereby justifying an assumption of model normality.

Comparing the full model to the reduced model gives the following ANOVA table:

```
## Analysis of Variance Table
##
## Model 1: mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb
## Model 2: mpg ~ wt + qsec + am
##   Res.Df RSS Df Sum of Sq    F Pr(>F)
## 1     21 148
## 2     28 169 -7     -21.8 0.44   0.86
```

The $\Pr(>F)$ value shows that the reduced model is not significantly different from the full model, but it has fewer regressors (-7). Additionally, the 95% confidence interval for *am* shows that we cna expect an increase in *mpg* by using a manual transmission instead of an automatic transmission.

```
##               2.5 % 97.5 %
## (Intercept) -4.63830 23.874
## wt          -5.37333 -2.460
## qsec         0.63457  1.817
## am           0.04573  5.826
```

Appendix

Scatterplot of mtcars Data
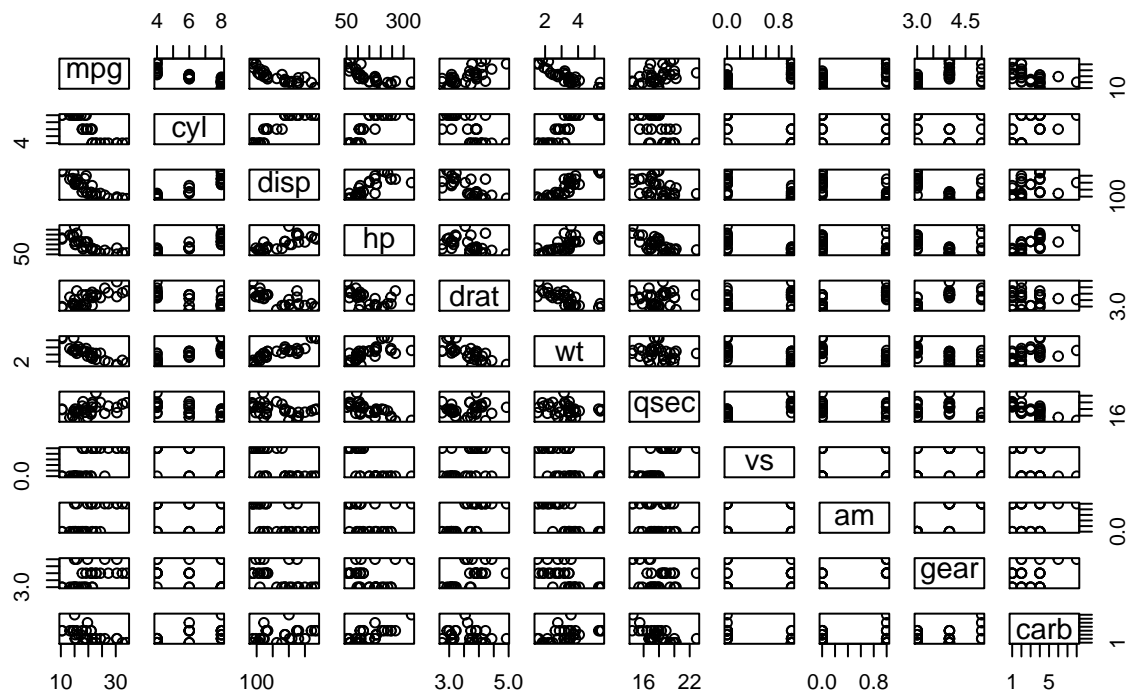
Summary of full model fit:

```
##
## Call:
## lm(formula = mpg ~ ., data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -3.45  -1.60  -0.12   1.22   4.63
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  12.3034    18.7179    0.66    0.518
## cyl          -0.1114     1.0450   -0.11    0.916
## disp          0.0133     0.0179    0.75    0.463
## hp           -0.0215     0.0218   -0.99    0.335
## drat          0.7871     1.6354    0.48    0.635
## wt           -3.7153     1.8944   -1.96    0.063 .
## qsec          0.8210     0.7308    1.12    0.274
## vs            0.3178     2.1045    0.15    0.881
## am            2.5202     2.0567    1.23    0.234
## gear          0.6554     1.4933    0.44    0.665
## carb         -0.1994     0.8288   -0.24    0.812
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 2.65 on 21 degrees of freedom
## Multiple R-squared:  0.869,  Adjusted R-squared:  0.807
## F-statistic: 13.9 on 10 and 21 DF,  p-value: 3.79e-07
```

Regression residuals for the reduced model: