Effect of Transmission Type on Automobile Efficiency (MPG)

Fenton Taylor

September 8, 2016

Executive Summary

Using the car data collected by *Motor Trend*, we will determine two things:

- Are manual or automatic transmissions better for mpg?
- What is the mpg difference, if any, for transmission types?

First, we will calculate the average mpg for each transmission type to see if there is a potential difference. Then we will fit a basic logistic regression to see if mpg alone is a good predictor of transmission type. If it isn't we will fit other models to determine the best set of predictors. Finally we will quantify the difference between average MPG for each transmission type and determine if that difference is significant.

After analysis, it was determined that manual transmissions have a higher mpg than automatic when all other variables are held constant. Interstingly, the MPG ratings converge as car weight increases and acceleration decreases. Specifically, lighter/faster manual cars get about 2.8 better mpg, average manual cars get about 1.5 better mpg, and heavier/slower manual cars get about 1.0 better mpg than automatic cars.

The Data

The **mtcars** data set contains 11 observations about 32 different cars. The variables are: mpg, cyl, disp, hp, drat, wt, qsec, vs, am, gear, carb. (Full variable documentation **HERE**).

The column **mpg** contains numeric measurements of miles/(US) gallon. The column **am** contains binary values of transmission type (0 = automatic, 1 = manual). However, all the variables are numeric in the data set, so **am** should be converted to a factor along with other appropriate variables.

```
model <- row.names(mtcars)
mtcars <- mutate(mtcars,
                 am = factor(am, labels = c("Automatic", "Manual")),
                 cyl = factor(cyl),
                 vs = factor(vs),
                 gear = factor(gear),
                 carb = factor(carb)
                 )
row.names(mtcars) <- model
```

Exploratory Analysis

Now that the data are in the correct form, let's take a look at the mean mpg for each transmission type and see if they are significantly different by using R's *t.test* function.

```
## mean in group Automatic     mean in group Manual
##                     17.1                    24.4
```

```
## p-value
## 0.00137
```

With a difference in the means of 7.245 and a p-value of $0.001 < .05$ we can reject the null hypothesis that the means for the two groups are the same. However, let's first take a look at how **mpg** might be correlated with the other variables. (See *Figure 1* in the Appendix).

- **mpg** is positively correlated with **drat**, **qsec**, **vs**, **am**, and **gear**.
- **mpg** is negatively correlated with **cyl**, **disp**, **hp**, **wt**, and **carb**.

Next, we will need to look at some regression models to see which of these variables, or combination of them, are the best predictors of **mpg** and to quantify the difference.

Regression Analysis

First, we will fit a basic linear model with just **am** as the predictor for **mpg**.

```
## lm(formula = mpg ~ am, data = mtcars)
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)     17.15       1.12   15.25 1.13e-15
## amManual         7.24       1.76    4.11 2.85e-04
```

```
## Adjusted R^2
##        0.338
```

Even though the transmission type is a significant predictor of **mpg** (p-value $= .000285 < .05$), the adjusted R^2 is only 0.338, meaning only about 33.8% of the variability in the data is explained by the model. Surely, we can find a better model to predict **mpg**.

Finding the Best Model

We will use R's *step* function to determine the best model to predict **mpg**. First, we must generate a model that includes all variables as predictors and let *step* test all possible combinations. The result is the following:

```
## lm(formula = mpg ~ wt + qsec + am, data = mtcars)
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      9.62      6.960    1.38 1.78e-01
## wt              -3.92      0.711   -5.51 6.95e-06
## qsec             1.23      0.289    4.25 2.16e-04
## amManual         2.94      1.411    2.08 4.67e-02
```

```
## Adjusted R^2
##        0.834
```

The best-fitting model according to the step function uses **am** (transmission type), **wt** (weight in 1000s of lbs), and **qsec** (quarter mile time in seconds–a measure of acceleration). Since we are really interested in transmisison type, let's fit one more model with an interaction between **am:wt** and **am:qsec**.

```
## lm(formula = mpg ~ am:wt + am:qsec, data = mtcars)


##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)       13.969      5.776    2.42 2.26e-02
## amAutomatic:wt    -3.176      0.636   -4.99 3.11e-05
## amManual:wt       -6.099      0.969   -6.30 9.70e-07
## amAutomatic:qsec   0.834      0.260    3.20 3.46e-03
## amManual:qsec      1.446      0.269    5.37 1.12e-05


## Adjusted R^2
##       0.879
```

This model with variable interaction seems to be a solid fit. All of the coefficients are significant with p-values $< .01$, except for the intercept, which is still significant at 0.023. Furthermore, this model has an adjusted R^2 term of 0.879, which is greater than the adjusted R^2 (0.834) from the model produced by the *step* function. As a final check, we will use R's *anova* function to see if the interaction model is an improvement over the previous ones.

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ wt + qsec + am
## Model 3: mpg ~ am:wt + am:qsec
##   Res.Df RSS Df Sum of Sq    F  Pr(>F)
## 1     30 721
## 2     28 169  2       552 62.7 7.1e-11 ***
## 3     27 119  1        51 11.5  0.0022 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

With a p-value of $0.002 < .05$, we can say with 95% confidence that the interaction model is an improvement over the best model returned by the *step* function. Therefore we will choose the interaction model.

Residual Diagnostics

According to the residual diagnostic plots (see *Figure 2* in Appendix), we can verify the following assumptions.

1. The *Residuals vs Fitted* plot shows random distribution and no non-linear trends, therefore the independence assumption is satisfied.
2. The *Normal Q-Q* plot is approximately linear, therefore the residuals are approximately normally distributed.
3. The *Scale Location* plot shows random distribution, which confirms constant variance.
4. The *Residuals vs Leverage* confirms that there are no outliers.

Conclusion

Interpreting the Coefficients

The interaction model has different coefficients for **am:wt** and **am:qsec**. That means automatic and manual transmissions will have different predicted values if **wt** and **qsec** are held constant. The expected value for mpg can be predicted by the following equation, where $am_{0i} = 1$ or 0, if the transmission is Automatic or Manual, respectively, and $am_{1i} = 1$ or 0, if the transmission is Manual or Automatic.

$$E[MPG_i] = 13.969 - 3.176(wt_i)(am_{0i}) - 6.099(wt_i)(am_{1i}) + 0.834(qsec_i)(am_{0i}) + 1.446(qsec_i)(am_{1i})$$

So, if all other variables are held constant, the difference in expected mpg is

$$E[MPG_i|am_{1i} = 1] - E[MPG_i|am_{0i} = 1] = -2.93wt_i + 0.612qsec_i$$

That means for every 1,000 lb increase in car weight, the expected difference in mpg decreases by 2.93 mpg. And for every 1 sec increase in quarter mile time, the expected difference in mpg increases by 0.612 mpg.

Results

The following table shows the differences in expected mpg for three cars in range of weight and quarter mile times. The **wt** and **qsec** parameters were taken from the 1st quartile, mean, and 3rd quartile of those variables in the original data set.

Table 1: Predicted values for MPG by transmission type.

	Automatic MPG	Manual MPG	Difference
1st Qu. (wt = 2.581, qsec = 16.89)	19.9	22.7	2.81
Mean (wt = 3.217, qsec = 17.85)	18.6	20.1	1.49
3rd Qu. (wt = 3.610, qsec = 18.90)	18.3	19.3	1.02

After analysis, it was determined that manual transmissions have a higher mpg than automatic when all other variables are held constant. Interstingly, the MPG ratings converge as car weight increases and acceleration decreases. Specifically, lighter/faster manual cars get about 2.811 better mpg, average manual cars get about 1.491 better mpg, and heavier/slower manual cars get about 1.025 better mpg.

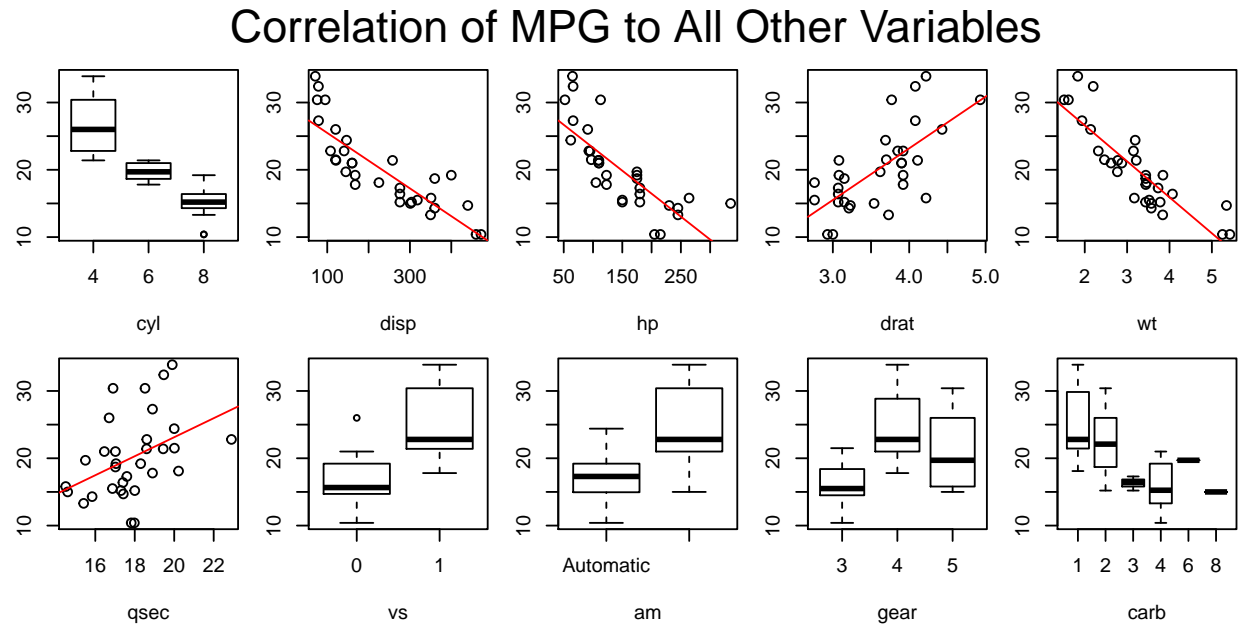---

Appendix

Figure 1 - Exploratory Analysis

Correlation of MPG to All Other Variables

Figure 2 - Residual Diagnostics

lm(mpg ~ am:wt + am:qsec)