# Is an automatic or manual transmission better for MPG?

*K. Divis*

*August 16, 2015*

### Executive summary

The mtcars data was used to answer the question of whether an automatic or manual transmission is better for MPG, quantifying the difference. The model that best fit the data included transmission type, weight of the car, and quarter mile time to predict gas milage. It revealed that manual transmission vehicles get better gas milage than automatic transmission vehicles (a difference of about 2.94 mpg). The report below outlines how this conclusion was reached, including exploratory analyses, model fitting/selection procedures, explanation of the best model, and model diagnostics and residuals.

*(This report was completed as part of (and under the guidelines of) the Johns Hopkins Data Science specialization Regression Models class on Coursera.)*

---

### Exploratory analysis

The mtcars dataset includes data from 32 vehicles (from 1973 or 1974), with values for mpg (miles/US gallon), cyl (number of cylinders), disp (displacement in cu. in.), hp (gross horsepower), drat (rear axle ratio), wt (weight in lb/1000), qsec (quarter mile time), vs (V-engine or straight engine), transmission (automatic or manual), gear (number of forward gears), and carb (number of carburetors).

Since the questions are about transmission and mpg, Figure 1 (see the appendix) shows the mpg for automatic and manual transmissions, not factoring in the other variables. At first glance, automatic transmissions appear to get fewer mpg than manual transmissions.

It seems reasonable that many of these variables will be related to one another (and thus may pose collinearity problems in model building). In an effort to further explore those relationships, Figure 2 shows the correlations between the variables. As suspected, many of the variables are highly correlated.

### Model fitting/selection

*Strategy for model selection:* Model fitting and selection was completed using a stepwise approach. The initial model included all the variables (am, cyl, disp, hp, drat, wt, qsec, vs, gear, and carb) as predictors of mpg. Each subsequent model included one less variable and was compared to the previous model using a log likelihood test. If removing that variable did not significantly change the fit of the model, then the subsequent model removed another variable. Which variable was removed was determined by the coefficient beta weights for that variable. The variable that appeared to contribute the least (highest p-value) was removed.

*Journey through model selection:* The initial model included all the variables (am, cyl, disp, hp, drat, wt, qsec, vs, gear, and carb). The second model removed cyl, which did not compromise the model ($p = 0.916$). The third mdoel additionally removed vs, which did not compromise the model ($p = 0.843$). The forth model additionally removed carb, which did not compromise the model ($p = 0.747$). The fifth model additionally removed gear, which did not compromise the model ($p = 0.62$). The sixth model additionally removed drat, which did not compromise the model ($p = 0.462$). The seventh model additionally removed disp, which did not compromise the model ($p = 0.299$). The eighth model additionally removed hp, which did not compromise the model ($p = 0.223$). This model looked like it would not need to be modified further. As a check, the nineth model removed qsec, but it did compromise the model ($p = 0$). Therefore the best model, based on this selection strategy was the model using am, wt, and qsec to predict mpg (Adjusted R Squared = 0.83; AIC = 154.12).

**Best model and results**

The best model used am, wt, and qsec to predict mpg. Mathematically, it can be written as mpg ~ B0 + B1(am) + B2(wt) + B3(qsec) + error. The beta coefficients for the model are: B0 = 9.62 (p = 0.178, CI = -4.64 to 23.87), B1 = 2.94 (p = 0.047, CI = 0.05 to 5.83), B2 = -3.92 (p = 0, CI = -5.37 to -2.46), and B3 = 1.23 (p = 0, CI = 0.63 to 1.82). Based on their respective p-values and confidence intervals, all three of the variables are important. Importantly, the effect of manual vs. automatic transmission is significant (p = 0.047).

The coefficients can be interpretted as follows: holding wt and qsec constant (i.e., equal to 0), an automatic transmission is expected to get 9.62 mpg (B0) and a manual transmission is expected to get 12.55 mpg (B0 + B1). The average wt was 3.22 half tons and the average qsec was 17.85 seconds. Using those values (rather than zeros), an automatic transimission is expected to get 18.9 mpg (B0 + B2(mean wt) + B3(mean qsec)) and a manual transimssion is expected to get 21.83 mpg (B0 + B1 + B2(mean wt) + B3(mean qsec).

**Residual plot and some diagnostics**

Figure 3 shows residual plots and diagnostics. They do show some deviations from what we would expect from a "perfect" model (e.g., the dip in the regression line on the Residuals vs. Fitted plot), but many aspects look promising (e.g., most of the points fall close to the diagonal in the Normal Q-Q plot, indicating our assumptions of normality may be justified). See the results from the gvlma() function below for further analysis.

One early question was whether collinearity would be an issue with this dataset. Variance inflation factors (VIF) were calculated for all 3 predictors: am = 2.54, wt = 2.48, and qsec = 1.36. A standard test to see whether collinearity is an issue is whether the sqrt(VIF) is greater than 2. This was not the case for any of the predictors in the model, so collinearity does not appear to be a problem in this final model.

I also assessed the assumptions of the linear model using the gvlma() function in the gvlma library. While the skewness (of error distributions), kurtosis (normality of error distibutions), and heteroscedasticity (of errors) assumptions were met, the link function direction test statistic was not (p = .003). It appears this model does not use the best link function for this data, so it should be interpretted with some caution. However, the simplicity of the model and ease of intuitively understanding it provide much benefit.

**Answers to questions**

The answers to the questions are not black and white–their exact interpretation depends on what model (and which predictors) are ultimately chosen. However, based on my analyses and the decisions I made when building the model, I came to the following conclusions. In general, manual transmissions get better gas milage than automatic transmissions (a difference of 2.94 mpg, p = 0.047, CI = 0.05 to 5.83). I found that it was important to adjust for the influence of weight and quarter mile time when evaluating the effect of transmission on gas milage. (Please note that exact values and examples can be found in the "Best Model and Results" section above).

---

## Appendix

**Figure 1: MPG by transmission type**

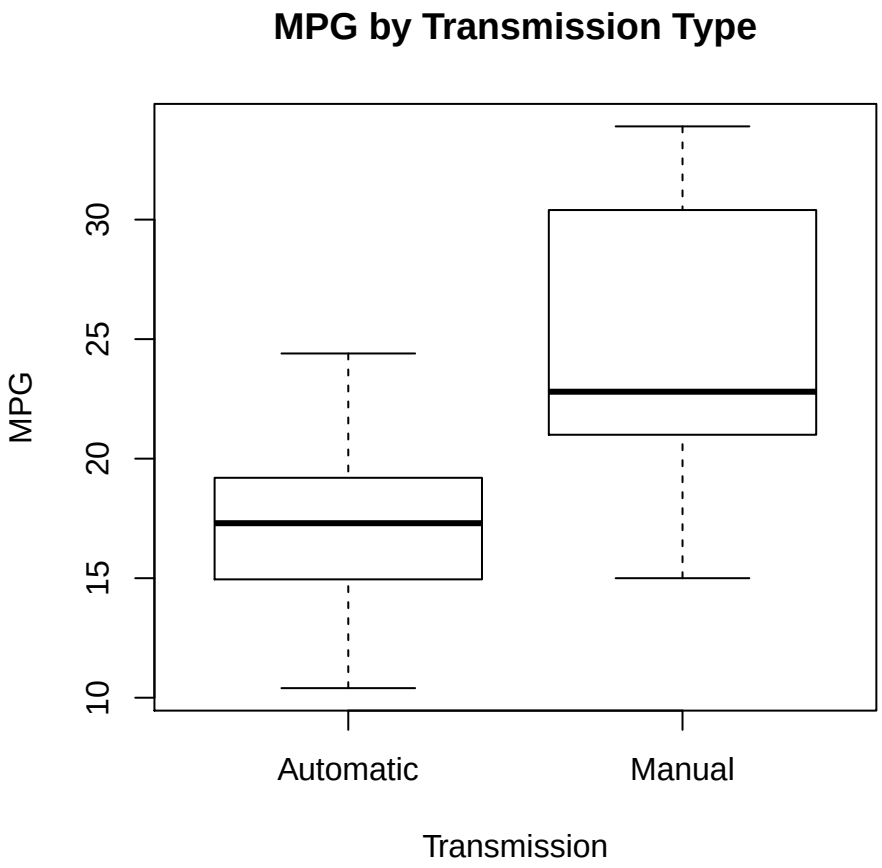

**Figure 2: Correlation Plot**

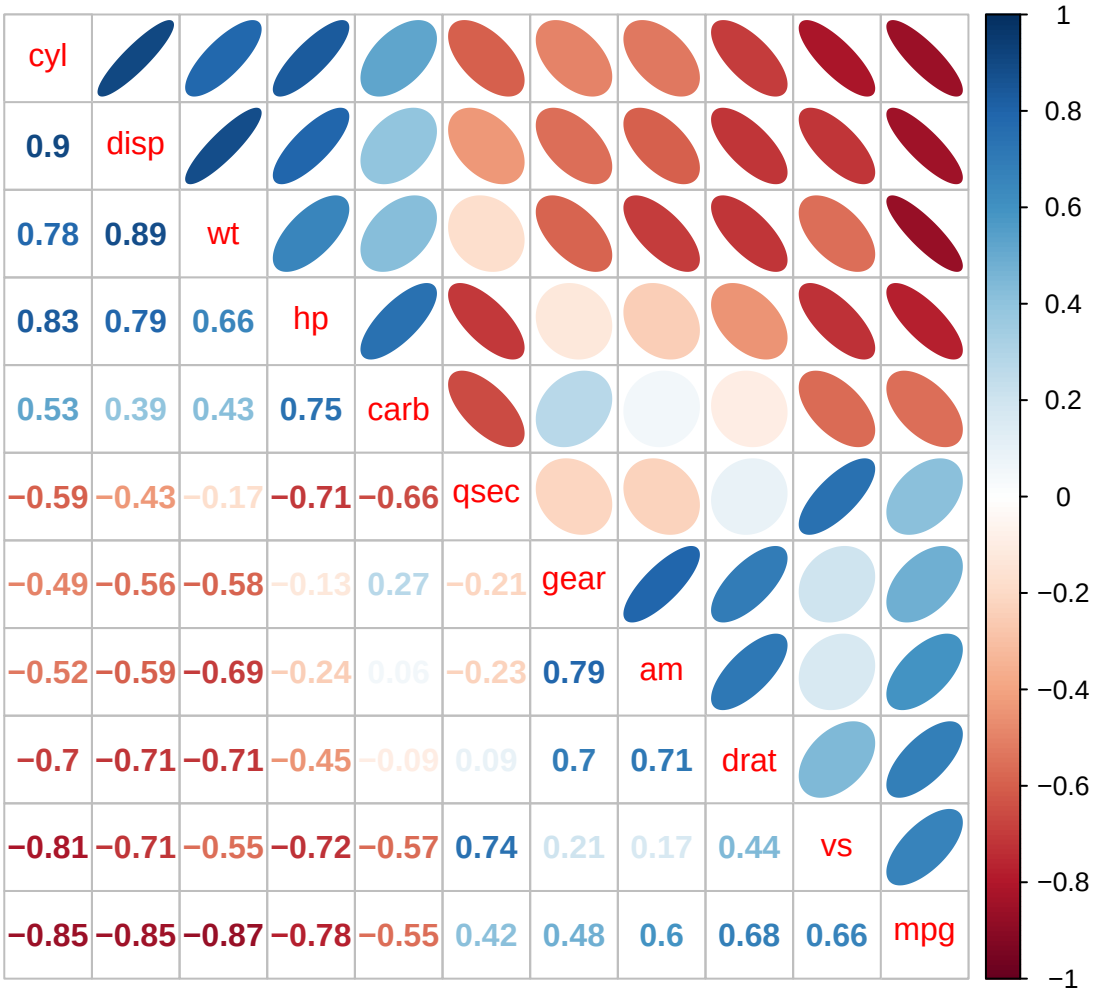

**Figure 3: Diagnostic Plots**