Transmission Type and MPG: Is Automatic or Manual Better?

Executive Summary

This article discusses how the relationship between miles per gallon that a car model has is predicted by the interaction between transmission type and weight. The model explains 83% of the variance in MPG, and predicts that at lighter weights, manual cars have better MPG, but at heavier weights, automatic cars have better MPG. However, the conclusions that can be drawn from this model are limited by a small sample, a different weight range in the data for manual and automatic cars, and some violations of model assumptions. The prediction intervals demonstrated high uncertainty for new variables, indicating this model may be unsuitable for making conclusions about car models not in the data.

Is an automatic or manual transmission better for MPG?

Let's consider a dataset of 32 car models from 1973-1974 which records the miles per gallon (MPG) of each of the models. In this dataset, we have 19 cars with automatic transmission and 13 with manual transmission. We also have information about the number of cylinders, displacement, gross horsepower, rear axle ratio, weight, 1/4 mile time, type of engine (V- or straight), number of forward gears and number of carburetors. After screening the MPG and transmission type variables (see Supporting Appendix), we could simply regress transmission type onto MPG to see if there is a significant difference in their mean MPG.

We can see from Figure 1 that manual cars (24.39) have a significantly higher MPG than automatic cars (17.15). However, this model does not explain much of the variance in MPG, having an R^2 of 0.338. Exploring the data further, we can see that a number of variables have very high correlations with MPG (see Figure 2 in the Supporting Appendix). Furthermore, some of these variables have a high relationship with transmission type, suggesting they may influence how transmission relates to MPG (Figure 2). In order to build a model predicting MPG including both transmission type and any additional covariates, variables were added into the model in the order of their strength of correlation with MPG (Figure 2). However, disp, hp and vs were not included as predictors as they were collinear (had a correlation at or above 0.8) with other possible predictors.

Figure 3 in the Supporting Appendix demonstrates the model building strategy. The predictors were entered one-by-one into a model containing transmission type. Only those that significantly improved model fit were retained. After all significant variables were entered into the model, the improvement to model fit by including interaction terms between these variables and transmission type was assessed. The two final models was one containing transmission type (am.f), weight (wt), and their interaction term (am.f * wt), and another containing the above as well as number of cylinders (cyl). The variance inflation factors (VIFs) of the two models are shown below:

```
## Loading required package: car
```

```
##       am.f          wt    am.f:wt
## 20.901259   2.728248  15.366853
```

```
##       am.f          wt       cyl.f    am.f:wt
## 24.302147   3.983261   3.060691  18.189413
```

The difference in R^2 between the two models is small (0.833 compared to 0.877), but the inclusion of number of cylinders in the model both increases the variance inflation, especially of weight, and decreases the

interpretability of the model. Moreover, number of cylinders is highly correlated with car weight (0.782), meaning it is likely explaining a lot of the same variance as weight. As such, the final model included transmission type (am.f), weight (wt), and their interaction term (am.f * wt).

Quantifying the MPG difference between automatic and manual transmissions.

To increase the ease of interpretation of the final model, the weight term was centred, so that the intercept represents the mean MPG for an automatic car with a mean weight. The final model indicates that the way that the transmission type affects a car's MPG depends on the weight of the car, due to the significant interaction between weight and transmission types.

```
##                           Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)              19.235844  0.7356848 26.146856 3.243563e-21
## am.fManual               -2.167728  1.4188862 -1.527767 1.377893e-01
## I(wt - mean(wt))         -3.785908  0.7856478 -4.818836 4.551182e-05
## am.fManual:I(wt - mean(wt)) -5.298360  1.4446993 -3.667449 1.017148e-03
```

At the mean weight for the sample (3217 lbs), both transmission types have a similar mean miles/gallon, with automatic cars having a mean of 19.24 MPG (95% CI: [17.72, 20.75]) and and manual cars having a mean of 17.07 MPG (95% CI: [14.3, 19.84]). However, as demonstrated in Figure 3, the mileage performance of each transmission type depends on how heavy the car is. The model predicts that automatic cars perform worse at lower car weights. For example, a 2000 lb automatic car would have a mean of 23.85 MPG (95% prediction interval: [17.67, 30.02]), while a manual car of the same weight would have a mean of 28.13 (95% prediction interval: [21.89, 34.36]). However, the model also predicts that conversely, automatic cars perform better at higher car weights. For example, a 4000 lb automatic car would have a mean of 16.27 MPG (95% prediction interval: [10.79, 21.76]), while a manual car of the same weight would have a mean of 9.95 MPG (95% prediction interval: [2.41, 17.5]).

Model assumptions and conclusions

Normality (Figure 2) and linearity (Figure 2) appear to be satisfied. Homoscedasticity appears to be satisfied for weight (Figure 2), but not transmission type (Figure 1). The plots of the residual versus fitted values indicate a pattern, where higher fitted values tend to have higher negative residuals (Figure 5). In addition, there are three car models with unusually high residuals (Merc 240DD, Fiat 128 and Toyota Corolla) indicating this model is a poor fit for both cars with high MPG (past about 28 MPG) and these three models. Moreover, the normal Q-Q plot of residuals indicates that errors are not normally distributed. The top hatvalues (leverage) and dfbetas (influence) demonstrate that while some values (especially the Maserati Bora) have potential to be outliers, the trend lines don't seem to be distorting around any given point (Figure 3).

```
##       Maserati Bora Lincoln Continental   Chrysler Imperial
##               0.371               0.304               0.281
## Cadillac Fleetwood        Lotus Europa         Honda Civic
##               0.254               0.253               0.216


## Maserati Bora     Merc 240D Ford Pantera L     Merc 230      Fiat 128
##         0.354         0.340         0.314         0.218         0.207
##    Duster 360
##         0.172
```

However, this plot does highlight two final issues with these data which limit the conclusions that can be drawn. The first is the overall small sample size, which means that each regression line does not have a large

amount of evidence to support it. In addition, weight has a different distribution for each transmission type (automatic: 2.465 and 5.424 lbs; manual: 1.513 and 3.57 lbs). As such, there is no support for the predictions this model makes about MPG for low-weight automatic cars, or high-weight manual cars. Predictions outside the range of values should therefore be interpreted with caution.

Finally, the fact that the model explained 83% of the variance in MPG leaves 17% variance in MPG unexplained. Given the model fitting undertaken in this article, we cannot explain this variance using something in our existing dataset. Instead, other unmeasured factors may explain this leftover variance.

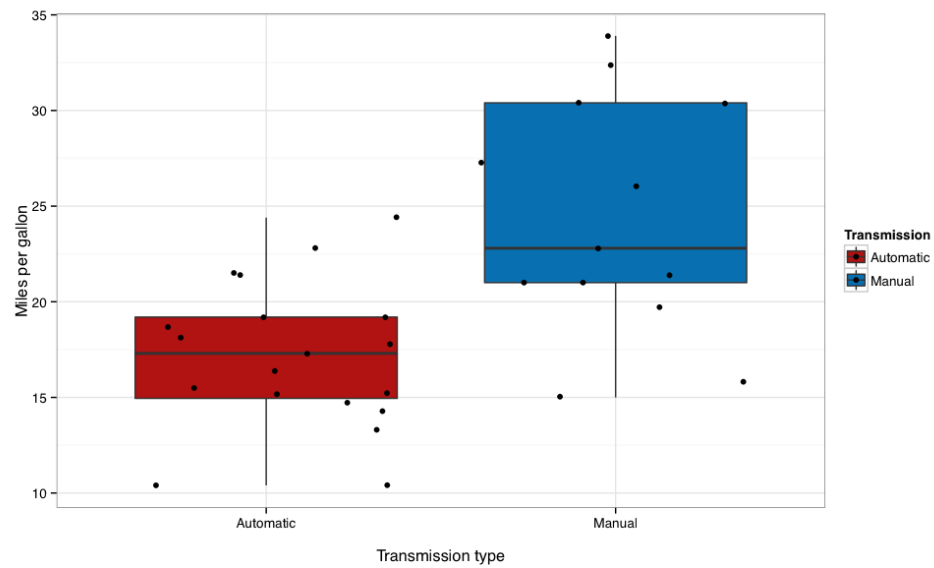## Supporting Appendix

Exploratory data analyses

Figure 1. Difference between mean MPG for automatic and manual transmissions.
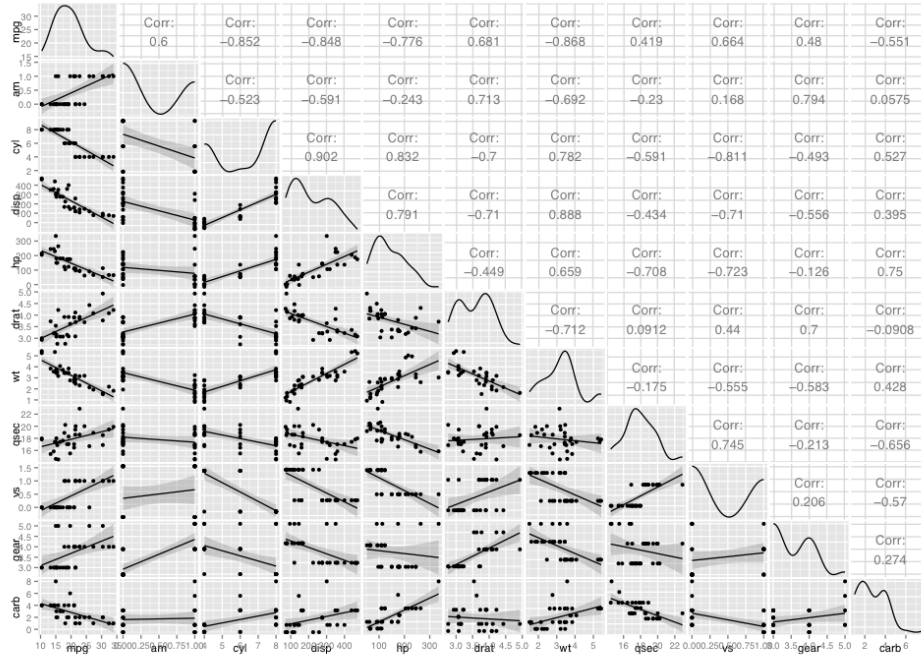
Figure 2. Correlation and scatterplot matrix between all predictors in the dataset and the outcome, including linear fit lines with 95% confidence intervals.

Building the model

Model 1	Model 2	p-value of model improvement
mpg ~ am.f	mpg ~ am.f + wt	0
mpg ~ am.f + wt	mpg ~ am.f + wt + cyl.f	0.003
mpg ~ am.f + wt + cyl.f	mpg ~ am.f + wt + cyl.f + drat	0.861
mpg ~ am.f + wt + cyl.f	mpg ~ am.f + wt + cyl.f + carb.f	0.764
mpg ~ am.f + wt + cyl.f	mpg ~ am.f + wt + cyl.f + gear.f	0.564
mpg ~ am.f + wt + cyl.f	mpg ~ am.f + wt + cyl.f + qsec	0.061
mpg ~ am.f + wt + cyl.f	mpg ~ am.f + wt + cyl.f + am.f * wt	0.007
mpg ~ am.f + wt + cyl.f + am.f * wt	mpg ~ am.f + wt + cyl.f + am.f * wt + am.f * cyl.f	0.802
mpg ~ am.f + wt	mpg ~ am.f + wt + am.f * wt	0.001

Figure 3. Comparison of model fit between nested models.

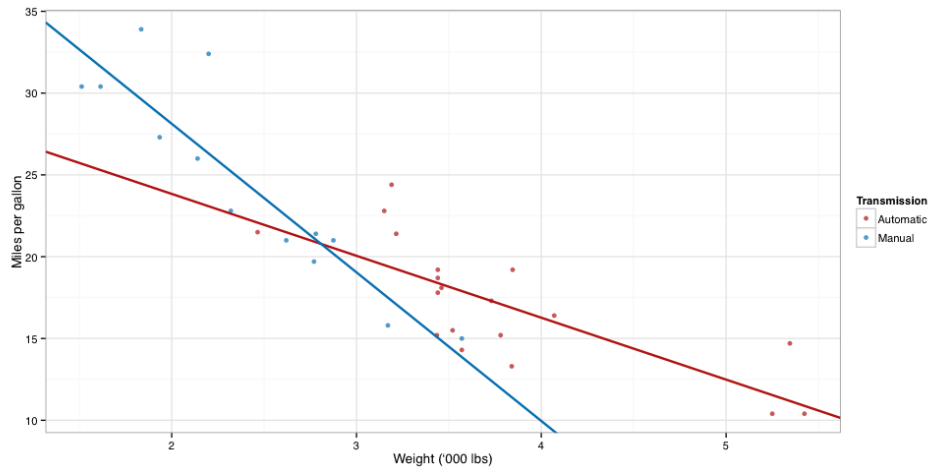4

Plotting the final model

Figure 4. Plot demonstrating the final model, with the regression lines predicting MPG by weight separated by transmission type.
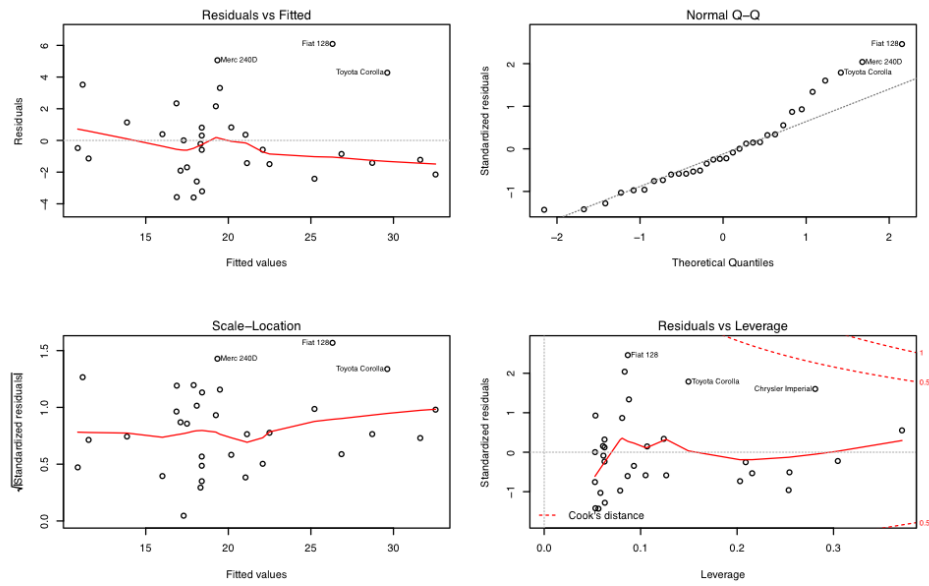
Diagnostic plots

Figure 5. Diagnostic plots demonstrating the behaviour of the residuals.