Motor Trend: The Effect of Transmission Types on MPG

James Brink

October 24th, 2014

Executive Summary In this analysis we explore the data we collected from our 1974 road tests. We tested 32 different cars and recorded their fuel consumption as well as 10 different aspects of their design. Our goal is to explore the relationship between these 10 aspects of automobile design and fuel consumption (miles per gallon or mpg). On a more granular level we wanted to show the effects, if any, that an automatic transmission had on mpg when compared to a manual transmission. Futhermore we aimed to quantify any mpg differences between the transmission types.

After performing our analysis we found that after adjusting for weight, number of cylinders, and horsepower, that cars with manual transmission tend to have higher mpg.

Loading and processing the data First we load in the mtcars dataset from the datasets package. We assign the dataset to the variable 'data'. Next we convert am (transmission type), cyl (number of cylinders), vs (V or straight cylinder arrangment), gear (number of gears), and carb (number of carburetors) to factor variables. After that we rename the levels on am so they are easier to work with.

Exploratory Data Analysis We created a scatterplot matrix which plotted the relatioships between all the variables in our dataset. You can view this figure (Figure 1) in the appendix. There appears to be relationships between mpg and cyl, disp, hp, drat, wt, and qsec. This shouldn't be unexpected. Logically we would expect these variables to effect mpg since they are measures of engine size, engine power, and vehicle weight. The remaining varables, vs, am, gear, and carb may have a relationship with mpg but it is tough to see because they are all factor variables with only a few levels. We will have to do further testing to determine which variables we want to include in our model.

Model Fitting To help us figure out which variables to include in our model we used a stepwise alogrithm to do reduce the number of variables down to those that provide us with an optimal model. This function starts by including all the variables and then removes one at a time to determine what regressors we should include in an optimal model.

```r
fit <- lm(mpg ~ ., data = data)
optFit <- step(fit, direction = "backward")
```

```
## lm(formula = mpg ~ cyl + hp + wt + am, data = data)
```

```
## (Intercept)        cyl6        cyl8          hp          wt       amMan
##    33.70832    -3.03134    -2.16368    -0.03211    -2.49683     1.80921
```

The model returned includes cyl, hp, wt, and am. The adjusted R-squared is 0.84 meaning it explains 84% of the variability in mpg. We can interpret the coefficients in the model as an increase from 4 cylinders to 6 causes a 3.03 drop in mpg and increasing from 6 cylnders to 8 drops causes a further 2.16 drop in mpg. We also see that an increase in horsepower causes a drop of 3.2 mpg for every 100hp and that weight causes a 2.50 drop in mpg for every additional 1000 pounds. Finally we see that a change from an automatic transmission to a manual transmission increased mpg by 1.81.

Since we are ultimately trying to show the relationship between am and mpg we should compare this model to a single variable model with just am to see if the inclusion of these regressors do in-fact create a better model. Also, intuitively one would think that disp and qsec might be useful to our model since they represent engine size and engine power, which we typcially associate with lower mpg so we will aslo compare our optimal model to a model including all the regressors in the optimal model as well as qsec and disp.

```
singleFit <- lm(mpg ~ am, data=data)
optPlusFit <- lm(mpg ~ cyl + hp + wt + am + qsec + disp, data=data)
anova(singleFit, optFit, optPlusFit)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ cyl + hp + wt + am
## Model 3: mpg ~ cyl + hp + wt + am + qsec + disp
##   Res.Df RSS Df Sum of Sq     F  Pr(>F)
## 1     30 721
## 2     26 151  4       570 24.02 4.3e-08 ***
## 3     24 142  2         9  0.73    0.49
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We observe a very small P value for the optimal model vs. the single variable model. However, for the optimal model plus disp and qsec vs the optimal model we see a relatively large P value meaning that the added regressors do not provide any improvement to the optimal model.

Residuals and Diagnostics Residuals were plotted for the optimal model and can be found in Figure 2 of the appendix. The first plot, Residuals vs. Fitted, seems to show no relationship in the residuals meaning that we can verify independence. The second plot, Normal Q-Q shows us that our residuals are normally distributed. The third plot, Scale-Location shows no pattern in the variance of our residuals, meaning that variance is relatively constant. Finally, the fourth plot, Residuals vs. Leverage, shows us the leverage of the different data points, we see that the all but one of the datapoints is within two standard deviations.

Uncertainty Before we draw conclusions about transmission types on mpg we must test to make sure that the difference in mpg for automatic and manual transmissions are actually different. We accomplished this with a simple t test which returned a P value of 0.0014 which is small enough to reject the null hypothesis that the mean mpg for manual and automatic transmissions is the same.

Conclusion In conclusion we see that our model is valid and we can use our coefficients to correctly identify the relationships between the variables in our model and mpg. As we set out to do initially we can not only show that cars with manual transmissions have a higher mpg than those with automatic, but we can also quantify that as 1.81 mpg.

Figure 1 Scatterplot Matrix

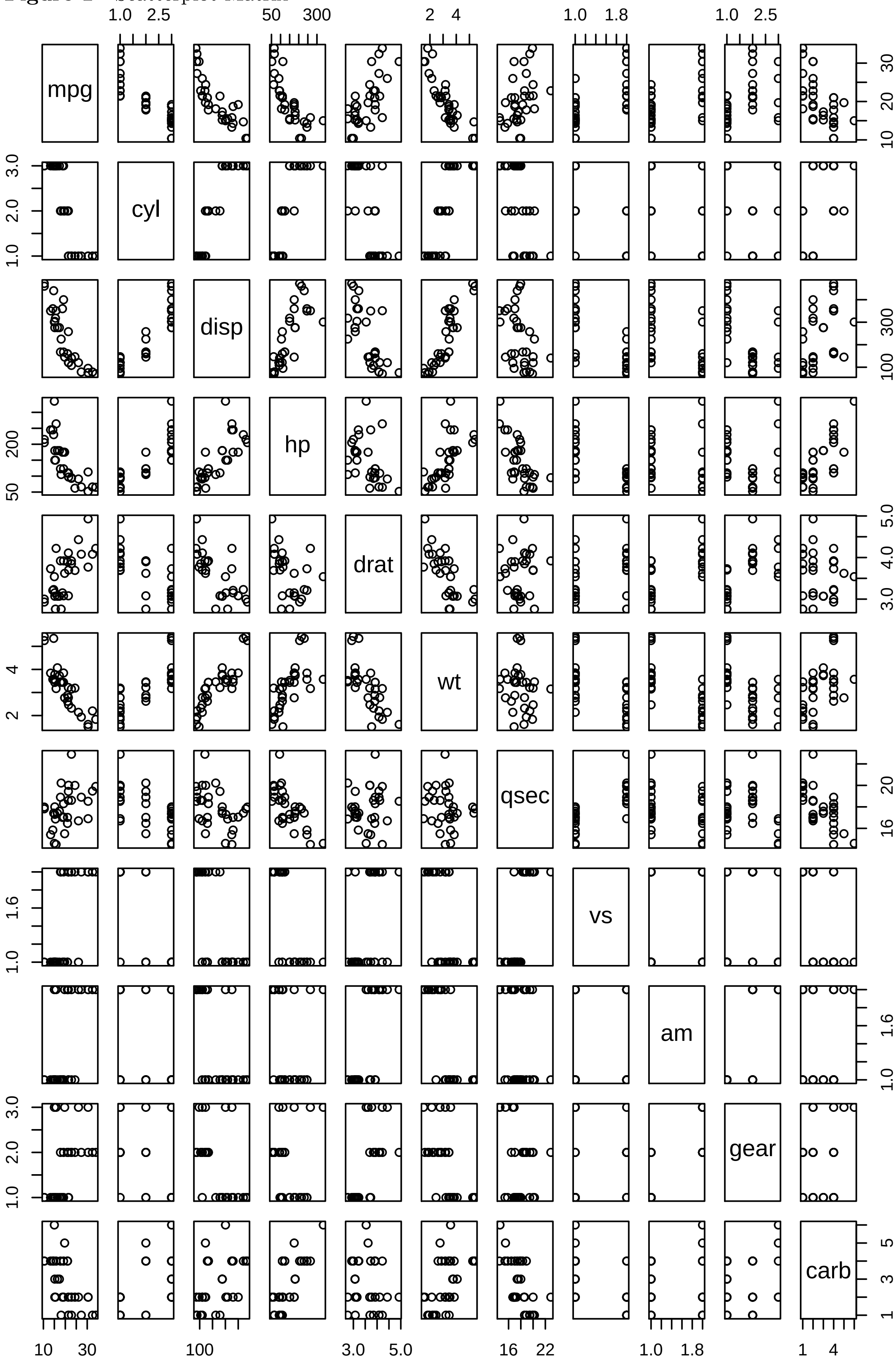

Figure 2 Residuals and Diagnostics