Reg Analysis Course Project

NDR

September 9, 2015

Executive Summary: I was tasked with evaluating whether an automatic or manual transmission car would have better miles per gallon (mpg) and see if we could qunatify the improvement between one and the other. At first glance it appeared that an manual transmission incraesed the mpg of a car by 7.25 mpg. However, after building other models it was determined that weight and horsepower (hp) were the only two variables that predicted the mpg of a car. After accounting for these variables, there was no statitically significant relationship (95%) between transmission types. Thus, there is no differenc ein mpg between automatic and manual transmission cars.

Analysis

To start the analysis, first the data was loaded in:

```
data("mtcars")
```

To examine what variables might have an impact on mpg, a pairs plot was constructed (Appendix Figure 1). Based on a review of this pairs plot, there appears to be a relationship between mpg and all the other variables, except the number of forward gears and the number of carbureators.

To evaluate if there is a significant relationship between transmission type and mpg a simple linear model was generated

```
fit1 <- lm(mpg~factor(am), data = mtcars)
summary(fit1)$coef
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## factor(am)1  7.244939   1.764422  4.106127 2.850207e-04
```

From the summary of the linear fit, it appears that a manual transmission gives you 7.25 more mpg than an automatic transmission and that the realtionship is significant well below 95% confidence. However, because the pairs plot suggested relationships between mpg and other variables, a residual plot, along with other diagnostic plots, were examined (Appendix Figure 2). The residual plot doesn't show any noticable trend, but there does seem to be wide variation in the residuals that might be explaiend by other variables.

To figure out what other variables are really important a number of linear models were constructed by adding in additional variables based on a review of the pairs plot (Figure 1).

```
##                Estimate Std. Error     t value     Pr(>|t|)
## (Intercept) 37.32155131  3.0546385 12.21799285 5.843477e-13
## factor(am)1 -0.02361522  1.5456453 -0.01527855 9.879146e-01
## wt          -5.35281145  0.7882438 -6.79080719 1.867415e-07
```

When weight is added into the linear model, the significance of transmission type coefficient disapears as can be seen from the summary of fit2 coefficients. When we add additional variables weight continues to be the only variable whose coefficient is significant, except in model 3 where horsepower is also significant.

To evaluate these models, we used analysis of variance.

```r
anova(fit1, fit2, fit3, fit4, fit5, fit6, fit7, fit8)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ factor(am)
## Model 2: mpg ~ factor(am) + wt
## Model 3: mpg ~ factor(am) + wt + hp
## Model 4: mpg ~ factor(am) + wt + hp + factor(cyl)
## Model 5: mpg ~ factor(am) + wt + hp + factor(cyl) + drat
## Model 6: mpg ~ factor(am) + wt + hp + factor(cyl) + drat + factor(vs)
## Model 7: mpg ~ factor(am) + wt + hp + factor(cyl) + drat + factor(vs) +
##     qsec
## Model 8: mpg ~ factor(am) + wt + hp + factor(cyl) + drat + factor(vs) +
##     qsec + disp
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 720.90
## 2     29 278.32  1    442.58 70.0365 2.771e-08 ***
## 3     28 180.29  1     98.03 15.5128 0.0007004 ***
## 4     26 151.03  2     29.27  2.3156 0.1222799
## 5     25 150.81  1      0.22  0.0348 0.8536386
## 6     24 143.35  1      7.46  1.1799 0.2891341
## 7     23 140.57  1      2.78  0.4396 0.5142023
## 8     22 139.02  1      1.55  0.2450 0.6255116
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The analysis of the different models shows that model 2 (transmission and weight) and 3 (transmission, weight, and hp) are significantly different from model 1. This suggests that weight and horsepower are important variables impacting mpg. To evaluate if weight is even important, a similar analysis was done with 3 new models.

```
## Analysis of Variance Table
##
## Model 1: mpg ~ wt
## Model 2: mpg ~ wt + hp
## Model 3: mpg ~ wt + hp + am
##   Res.Df    RSS Df Sum of Sq       F   Pr(>F)
## 1     30 278.32
## 2     29 195.05  1    83.274 12.9328 0.001226 **
## 3     28 180.29  1    14.757  2.2918 0.141268
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

This analysis shows that the model with horsepower and weight is significantly different from the model with just weight. However the model with transmission type is not significantly different from the model of just weight and horsepower. Thus the only two variables that appear to have an effect on mpg for cars is weight and horsepower. The answer to our question of which type of transmission is better for mpg is that neither autmatic nor manual transmission is better in regards to mpg after accounting for the weight and horsepower of the car. Thus, there is no difference between automatic and manual transmission cars.

Appendix

Figure 1 Pairs plot of mtcars

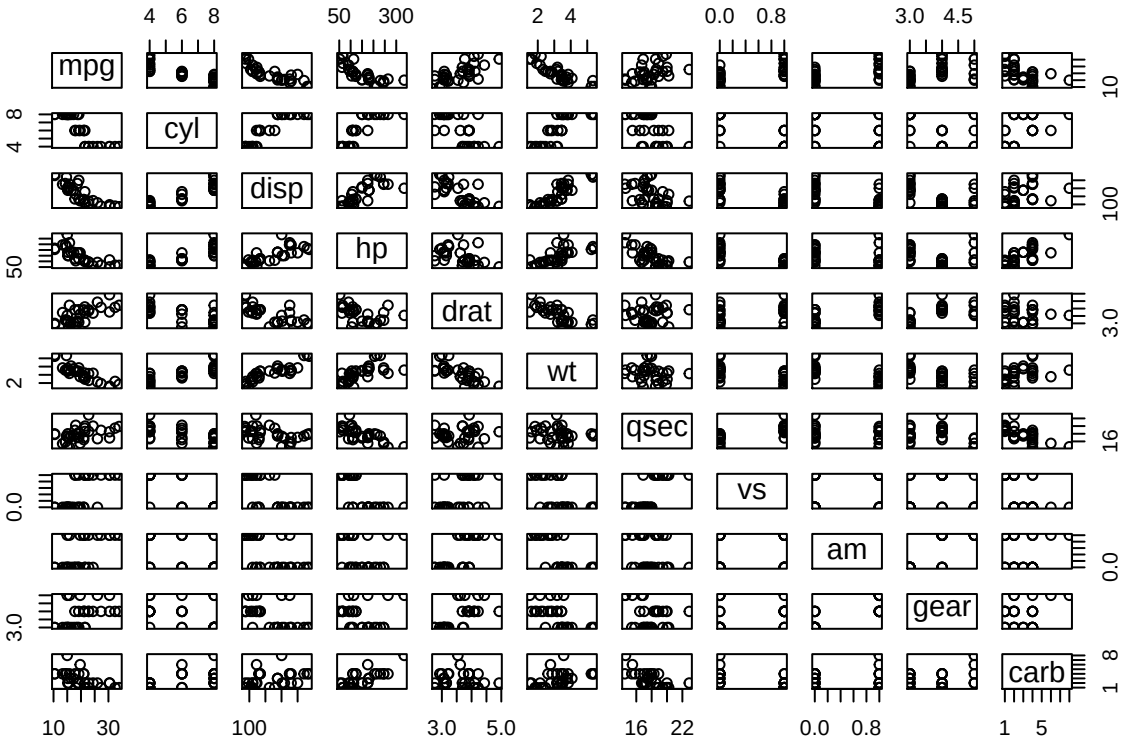

Figure 2 Residual Plot of Simple Linear Model