Automatic and Manual Tranmissions on Miles per Gallon

Charin Polpanumas

February 10, 2559 BE

Executive Summary

We examine the effect of transmissions, either being automatic or manual, of 32 automobiles controlling for other design and performance aspects, based on the *mtcars* dataset. When controlling for confounders selected using stepwise AIC algorithm (number of cylinders, horsepower, and weight), transmissions have a slight effect on miles per gallon but it is not statistically significant.

Exploratory Data Analysis

The dataset was extracted from the 1947 *Motor Trend* magazine. The columns include 10 aspects of automobile design and performance. The rows represent 32 automobiles.

cyl,*vs*, *am*, *gear*, and *carb* are more suitably treated as factors, whereas the rest are numeric. Create a new dataset m and apply this treatment

Plot correlation matrix in order to see if there is a chance of collinearity. We notice from the plot that pairs of variables with correlation higher than 0.8 or lower than -0,8 are (mpg,disp), (mpg,wt) (cyl,disp), (cyl,hp), (cyl,vs), (disp,wt) and should be careful when including them in the model. (See Figure 1)

The boxplot suggests that automatic vehicles have smaller miles per gallon than manual vehicles. However, we need to investigate further, controlling for effects of confounders, by using a linear regression model. (See Figure 2)

Regression Analysis

Model Selection

The dependent variable is mpg, and the indepedent variable is am. Thus we can utilize multiple regression analysis using other variables in the dataset as possible confounders.

We perform an AIC stepwise selection, relying on similar principles as the likelihood ratio tests using anova() discussed in class, in order to determine which variables to include. See Table 1 for all the trial models.

```
base_fit <- lm(mpg~., data=m)
step <- step(base_fit, direction="both")
```

Results

According to the result, we include cyl, hp, wt as confounders. The model we chose explains 84.01% of the variations in mpg (adjusted R-square). The model specification is statistically significant (p-value of F-stat being 1.506e-10).

On average a manual vehicle has 1.809 miles better miles per gallon than an automatic vehicle but the difference is not statistically significant at 5%. We cannot conclude with the current dataset that

```
##
## Call:
## lm(formula = mpg ~ cyl + hp + wt + am, data = m)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9387 -1.2560 -0.4013  1.1253  5.0513
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.70832    2.60489  12.940 7.73e-13 ***
## cyl6        -3.03134    1.40728  -2.154  0.04068 *
## cyl8        -2.16368    2.28425  -0.947  0.35225
## hp          -0.03211    0.01369  -2.345  0.02693 *
## wt          -2.49683    0.88559  -2.819  0.00908 **
## am1          1.80921    1.39630   1.296  0.20646
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.41 on 26 degrees of freedom
## Multiple R-squared:  0.8659, Adjusted R-squared:  0.8401
## F-statistic: 33.57 on 5 and 26 DF,  p-value: 1.506e-10
```

Residual Analysis

According to Figure 3, Residuals vs Fitted and Q-Q plots demonstrate that the residuals can be said to be normally distributed. The band pattern in Scale-Location plot demonstrates constant variance of the residuals. The Residuals vs Leverage plot illustrates that all data points fall within 0.5 Cook's distance, thus no outlier has an extreme effect on overall changes in regression coefficients when removed.

Conclusion

Although on average a manual vehicle seems to have 1.809 miles better miles per gallon than an automatic vehicle 79.4% of the time (p=0.206), from the given dataset, we cannot conclude if transmissions have a statistically significant influence on miles per gallon when controlled for number of cylinders, horsepower, and weight. This may be because the influence has already been captured by the confounders.

Appendix

Table 1 All Models Performed by AIC Stepwise Selection

Model 1 includes all predictors. At each step a predictor is removed. This iterates until the AIC is minimized.

```
##     Step Df   Deviance Resid. Df Resid. Dev      AIC
## 1        NA         NA        15   120.4027 76.40339
## 2 - carb  5 13.5988573        20   134.0015 69.82769
```

```
## 3 - gear  2  5.0215145        22   139.0230 67.00492
## 4 - drat  1  0.9672159        23   139.9903 65.22678
## 5 - disp  1  1.2473996        24   141.2377 63.51066
## 6 - qsec  1  2.4420033        25   143.6797 62.05921
## 7   - vs  1  7.3459298        26   151.0256 61.65483
```

Figure 1 Correlation Matrix of Variables in mtcars

```
## Loading required package: corrplot
```

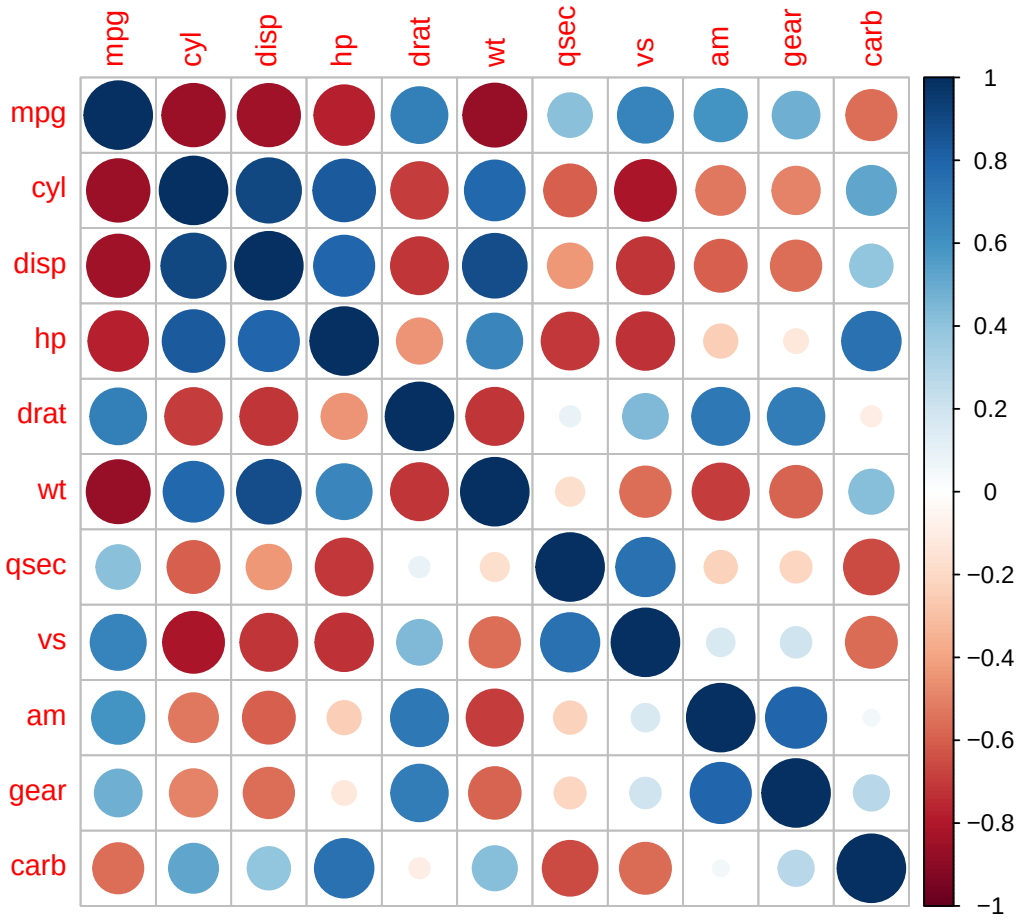

Figure 2 Miles per Gallon of Automatic vs Manual Vehicles

```
## Loading required package: ggplot2
```

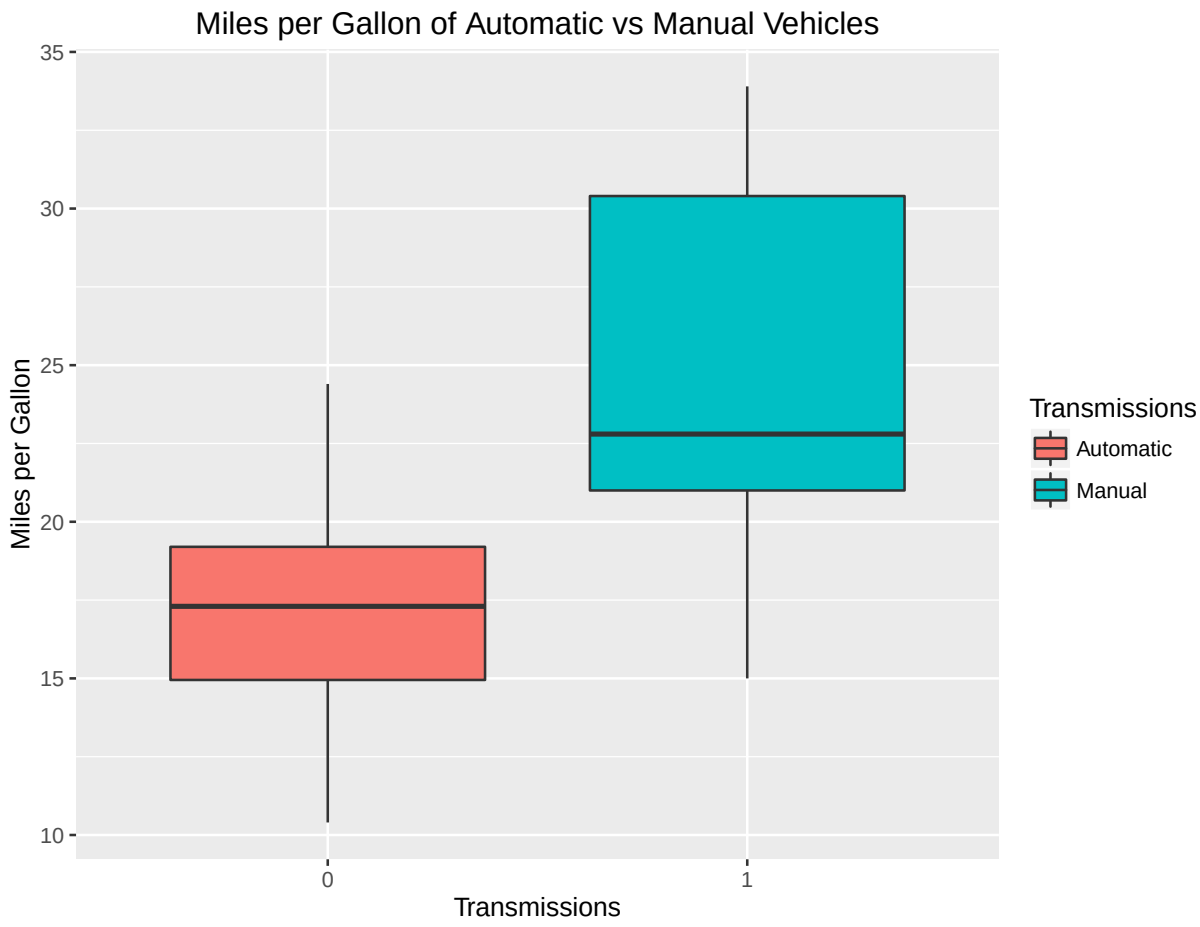
Miles per Gallon of Automatic vs Manual Vehicles
Miles per Gallon
35
30
25
20
15
10
0
1
Transmissions
Transmissions
Automatic
Manual

Figure 3 Residual Plots