Motor Trend Weekly - MTCars Regression Analysis

Kie Gouveia

Executive Summary

The purpose of this report is to concisely explore the relationship between a set of variables and miles per gallon(mpg) using the mtcars dataset. In particular, the report will attempt to address the following:
- Is an automatic or manual transmission better for MPG?
- Quantify the mpg difference between automatic and manual transmissions.

Inital exploratory analysis (see Appendix: Exploratory Analysis) revealed that the mean mpg among automatic cars is 17.15 with a standard deviation of 3.83, and 24.39 among manual cars, with a standard deviation of 6.17.

Univariate Analysis

It will be useful to begin this analysis by creating a linear univariate model using automatic/manual (am) as a predictor and mpg as the outcome. This will provide insight into the relationship between these variables and act as a baseline for further multivariate analysis.

```
fit <- lm(mpg ~ am, data = mtcars)
summary(fit)$coeff; summary(fit)$r.squared
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## am1          7.244939   1.764422  4.106127 2.850207e-04
```

```
## [1] 0.3597989
```

Here, the estimate of am0 (the mean mpg for automatic vehicles) is our intercept and the slope associated with am1(manual) 7.25, is simply the difference in mean mpg between automatic and manual cars. This suggests that on average, the mean mpg among manual cars is 7.25 higher than that of automatic cars. This result is statistically significant using a confidence level of 95%. However, the returned R^2 indicates that this model is only capable of predicting 36% of the variance in mpg, suggesting that the model could be improved by including other variables.

Multivariate Analysis

To revise the model to include other variables which are significant predictors of mpg, we will do some simple nested model testing using the anova function.

```
fitAll <- lm(mpg ~ ., data = mtcars)
anova(fitAll) # not shown.
```

According to this analysis, cyl(number of cylinders), disp (displacement(cu.in.)) and wt(weight (lb/1000)) all appear to contain groups with significant levels of variation. Therefore, they could serve as good predictors for mpg. Given that there is a known interaction between automatic and manual transmission and weight, we will also include this in the model.

After seeing that certain numbers of cylinders did not show statistically significant variance (at 95% confidence), we settle on this model which shows that all figures are statistically significant and returns a much stronger .833 R^2.

```
newfit <- lm(mpg ~ am:wt + am + wt, data = mtcars)
summary(fit)$coeff;
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## am1          7.244939   1.764422  4.106127 2.850207e-04
```

```
summary(fit)$r.squared
```

```
## [1] 0.3597989
```

```
anova(fit, newfit) # not shown.
```

An anova analysis comparing our multivariate model to our univariate model shows statistically significant p-values, meaning we can reject the hypothesis that our newly added variables do not contribute information around variance in mpg.

Diagnostics

```
par(mfrow = c(2,2)); plot(newfit)
plot(resid(newfit)) # not shown.
```

In our residual plot analysis the plots seem fairly healthy. There may be a slight negative pattern in the Residuals vs Fitted Values plot that warrants further exploration. The Q-Q plot appears to show that residuals are relatively normalalthough here, too, there may be a slight trend.

Inference

```
t.test(mpg ~ am, data = mtcars)$conf.int
```

```
## [1] -11.280194  -3.209684
## attr(,"conf.level")
## [1] 0.95
```

Finally, we conduct a t-test with the null hypothesis that there is not significant variation in mpg among automatic and manual cars. The results of this test allow us to say that the likelihood of variation we see in mpg between automatic and manual vehicles would happen in less 5% of cases. Thus, with 95% confidence there is variance in mpg between these transmission types.

Findings and Considerations

According to our most accurate model: - Holding weight constant and disregarding the interaction between weight and am, manual transmission cars will get 14.9 more miles per gallon than automatic cars. - In order to ensure that the coefficients for transmission type were statistically significant, other strong predictors were omitted. In another analysis, it would be worth exploring models including other strong predictor candidates. - The data itself may be skewed by driver skill, which is not accounted for by the dataset.

Appendix

Preparation

Load data and necessary packages:

```
library(ggplot2)
library(car)
library(graphics)
library(gclus)
data(mtcars)
mtcars$am <- factor(mtcars$am)
```

Exploratory Analysis

Data Summary:

```
##       mpg         cyl         disp             hp             drat      
##  Min.   :10.40   4:11   Min.   : 71.1   Min.   : 52.0   Min.   :2.760  
##  1st Qu.:15.43   6: 7   1st Qu.:120.8   1st Qu.: 96.5   1st Qu.:3.080  
##  Median :19.20   8:14   Median :196.3   Median :123.0   Median :3.695  
##  Mean   :20.09          Mean   :230.7   Mean   :146.7   Mean   :3.597  
##  3rd Qu.:22.80          3rd Qu.:326.0   3rd Qu.:180.0   3rd Qu.:3.920  
##  Max.   :33.90          Max.   :472.0   Max.   :335.0   Max.   :4.930  
##        wt             qsec        vs     am     gear   carb  
##  Min.   :1.513   Min.   :14.50   0:18   0:19   3:15   1: 7  
##  1st Qu.:2.581   1st Qu.:16.89   1:14   1:13   4:12   2:10  
##  Median :3.325   Median :17.71                 5: 5   3: 3  
##  Mean   :3.217   Mean   :17.85                        4:10  
##  3rd Qu.:3.610   3rd Qu.:18.90                        6: 1  
##  Max.   :5.424   Max.   :22.90                        8: 1  
```

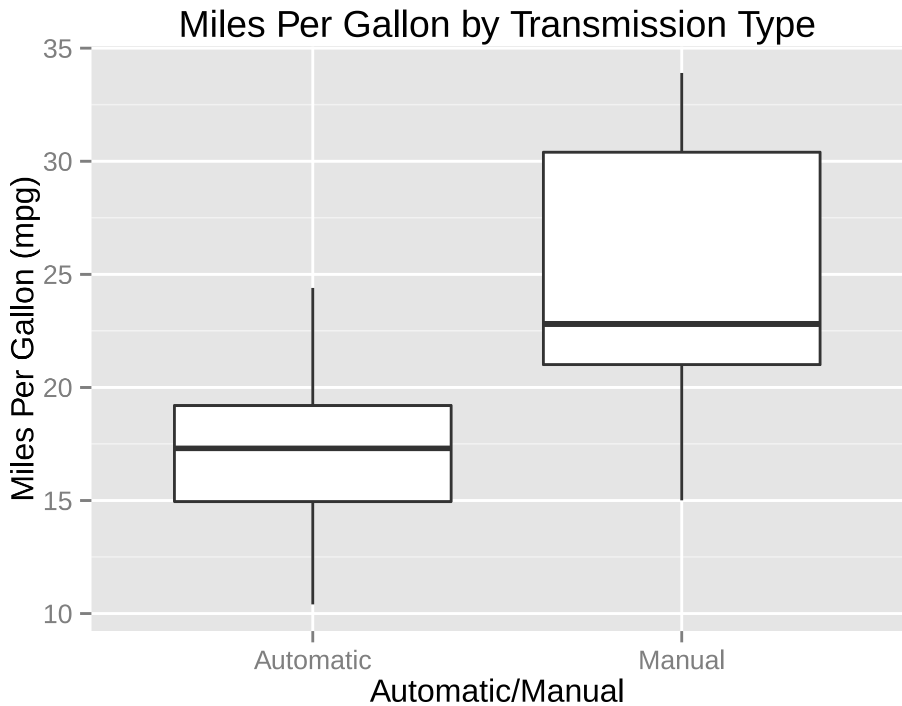
```
## [1] 17.15
```

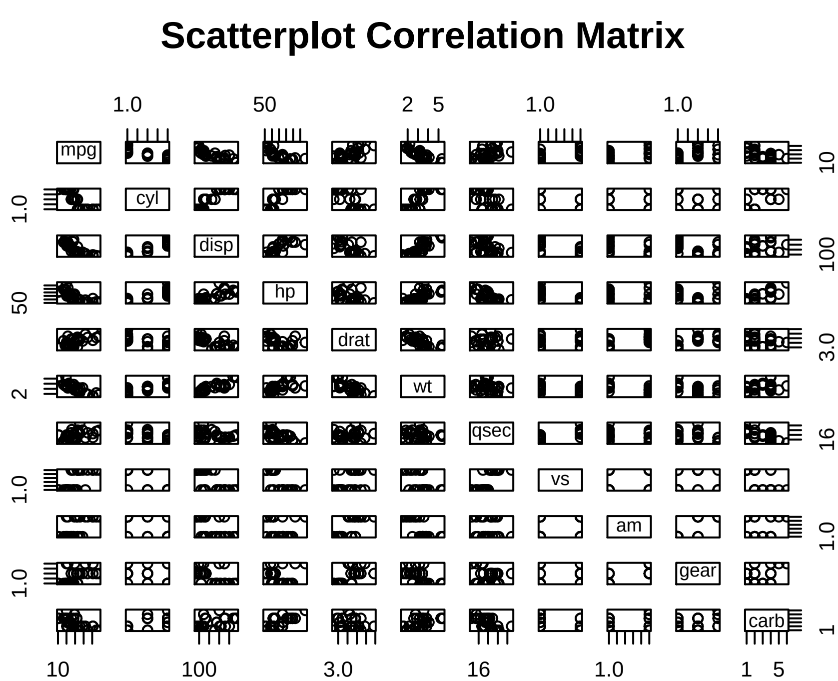
```
## [1] 24.39
```

```
## [1] 3.83
```

```
## [1] 6.17
```

Exploratory Graphs:


Miles Per Gallon by Transmission Type
Miles Per Gallon (mpg)
35
30
25
20
15
10
Automatic
Manual
Automatic/Manual


Scatterplot Correlation Matrix
mpg
cyl
disp
hp
drat
wt
qsec
vs
am
gear
carb

Diagnostics

```
par(mfrow = c(2,2)); plot(newfit)
```