Statistical Exploration of Tranmission Type on Miles Per Gallon

S. van der Westhuizen

April 10, 2017

Executive Summary

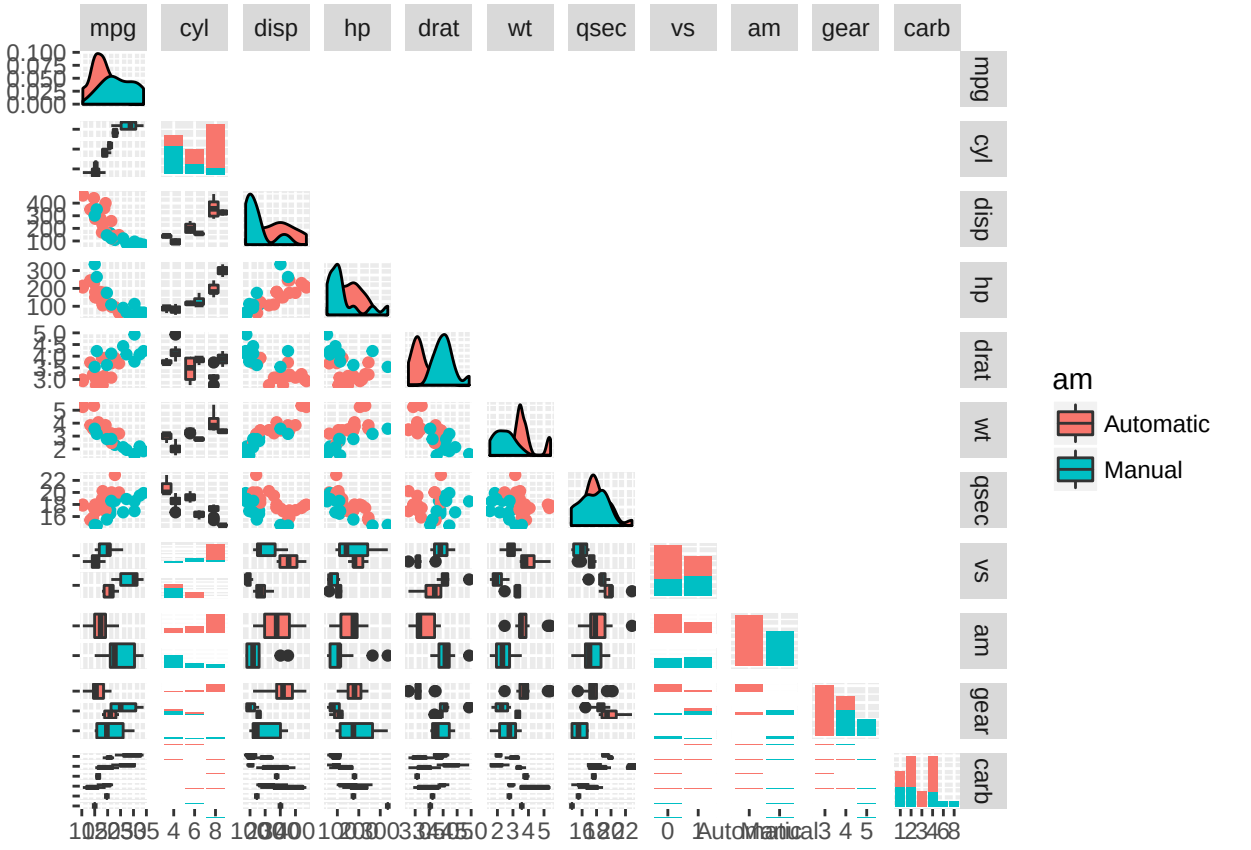
This analysis was conducted on the mtcars dataset. The dataset was extracted from the 1974 Motor Trend US magazine and comprises of 32 automobiles between 1973 and 1974. The purpose of this analysis is to determine:
* "Is an automatic or manual transmission better for MPG"
* "Quantify the MPG difference between automatic and manual transmissions"

The analysis below indicates that transmission type does not have a significant impact on MPG. Manual transmissions account for only a 1.8 MPG increase when holding everything else the same.

Data Exploration

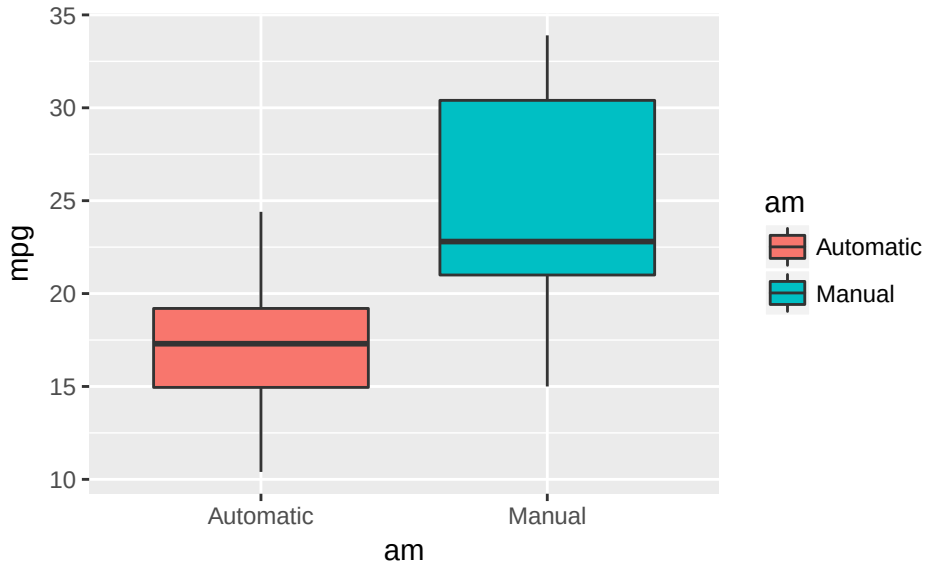
The dataset mtcars consists of a data frame with 32 observations on 11 variables.

```
# Loading the required libraries
suppressMessages(library(datasets))
suppressMessages(library(ggplot2))
suppressMessages(library(GGally))

# Loading the 'mtcars' dataset from the datasets library
data("mtcars")

# Convert the following columns to type factor
mtcars$cyl <- as.factor(mtcars$cyl)
mtcars$vs <- as.factor(mtcars$vs)
mtcars$gear <- as.factor(mtcars$gear)
mtcars$carb <- as.factor(mtcars$carb)
mtcars$am <- factor(mtcars$am,labels=c('Automatic','Manual'))

ggpairs(mtcars, ggplot2::aes(colour = am), legend = c(2,1),
  lower = list(
    continous = "smooth",
    combo =  "box",
    discrete = "facetbar"
  ),
  diag = list(
    continous = "densityDiag",
    discrete = "barDiag"
  ),
  upper = "blank"
)
```



From the pairs plot above, we can observe a few trends which will be investigated further:
1. The boxplot of MPG ~ am seems to indicate a significant difference
2. Manual transmission tend to cluster towards the higher end of MPG
3. There appears to be relationships between the independant variables which may imply that there may be covariance

```
ggplot(mtcars, aes(am, mpg))+
  geom_boxplot(aes(fill=am))
```



```r
ttest <- t.test(mpg ~ am, mtcars)
print(ttest)
```

```
##
##  Welch Two Sample t-test
##
## data:  mpg by am
## t = -3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.280194  -3.209684
## sample estimates:
## mean in group Automatic    mean in group Manual
##                17.14737                24.39231
```

From the t test above, P = 0.0013736 < 0.05 which implies that their mean difference is not equal to 0 so we can reject the null hypothesis. There is a significant difference between manual and automatic transmissions.

However, we need to investigate it further to understand how much of the variance in MPG is accounted for by the transmission type.

Summary by transmission type

Manual

```r
man<- mtcars[mtcars$am == "Manual",]
summary(man)
```

```
##       mpg        cyl        disp             hp             drat
##  Min.   :15.00   4:8   Min.   : 71.1   Min.   : 52.0   Min.   :3.54
##  1st Qu.:21.00   6:3   1st Qu.: 79.0   1st Qu.: 66.0   1st Qu.:3.85
##  Median :22.80   8:2   Median :120.3   Median :109.0   Median :4.08
##  Mean   :24.39         Mean   :143.5   Mean   :126.8   Mean   :4.05
##  3rd Qu.:30.40         3rd Qu.:160.0   3rd Qu.:113.0   3rd Qu.:4.22
##  Max.   :33.90         Max.   :351.0   Max.   :335.0   Max.   :4.93
##        wt             qsec       vs           am     gear  carb
##  Min.   :1.513   Min.   :14.50   0:6   Automatic: 0   3:0   1:4
##  1st Qu.:1.935   1st Qu.:16.46   1:7   Manual   :13   4:8   2:4
##  Median :2.320   Median :17.02                       5:5   3:0
##  Mean   :2.411   Mean   :17.36                             4:3
##  3rd Qu.:2.780   3rd Qu.:18.61                             6:1
##  Max.   :3.570   Max.   :19.90                             8:1
```

Automatic

```r
auto<- mtcars[mtcars$am == "Automatic",]
summary(auto)
```

```
##       mpg        cyl         disp             hp             drat
##  Min.   :10.40   4: 3   Min.   :120.1   Min.   : 62.0   Min.   :2.760
##  1st Qu.:14.95   6: 4   1st Qu.:196.3   1st Qu.:116.5   1st Qu.:3.070
##  Median :17.30   8:12   Median :275.8   Median :175.0   Median :3.150
##  Mean   :17.15          Mean   :290.4   Mean   :160.3   Mean   :3.286
```

```
##  3rd Qu.:19.20          3rd Qu.:360.0   3rd Qu.:192.5   3rd Qu.:3.695
##  Max.   :24.40          Max.   :472.0   Max.   :245.0   Max.   :3.920
##        wt             qsec       vs              am     gear   carb
##  Min.   :2.465   Min.   :15.41   0:12   Automatic:19   3:15   1:3
##  1st Qu.:3.438   1st Qu.:17.18   1: 7   Manual   : 0   4: 4   2:6
##  Median :3.520   Median :17.82                          5: 0   3:3
##  Mean   :3.769   Mean   :18.18                                 4:7
##  3rd Qu.:3.842   3rd Qu.:19.17                                 6:0
##  Max.   :5.424   Max.   :22.90                                 8:0
```

From the statistical summarise above, it appears that not only is the MPG for manual transmissions higher, their weight (wt), displacement (disp) and horsepower (hp) all appear to be on average lower than the automatic transmissions. It is also worth to note that the number of vehicles per cylinder class is very different as well. These factor may be confounding the results of transmission type on MPG.

Models

```
model1 <- lm(mpg~ am, mtcars)
model2 <- lm(mpg~ ., mtcars)
```

Final model selection was performed using stepwise regression using the AIC metric. Cross validation was not considered due to the limited data size.

```
suppressMessages(library(MASS))
modelfinal <- stepAIC(model2, trace = 0)
anova(model1, modelfinal, model2)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ cyl + hp + wt + am
## Model 3: mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 720.90
## 2     26 151.03  4    569.87 17.7489 1.476e-05 ***
## 3     15 120.40 11     30.62  0.3468    0.9588
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Conclusion

The anova summary above establishes that the model as selected through strepwise regression is highly significant and accounts for the greatest reduction in residual sum of squares. The best model fit comprises of the following coefficients:

```
modelfinal$coefficients
```

```
## (Intercept)        cyl6        cyl8          hp          wt    amManual
## 33.70832390 -3.03134449 -2.16367532 -0.03210943 -2.49682942  1.80921138
```

This indicates that manual transmissions only account for a 1.8 MPG change above automatic transmissions. This is significantly less than other variables influence on the overall MPG and the mean difference between transmissions types according to the t test above. This is a result of the confounding effects of the other

variables. The image in the appendix below highlights the residual plots of the model and confirm the best model fit.

Appendix

Residual plots for the final model

```
par(mfrow =c(2,2))
plot(modelfinal)
```