Effects on MPG, Automatic vs. Manual Transmission

Tyler Richardett

Executive Summary

Using the `mtcars` data set, we answer the following two questions:

1. Is an automatic or manual transmission better for MPG?
2. Quantify the MPG difference between automatic and manual transmissions.

A two-sided t-test concludes that manual transmission is better for MPG. And a multivariate linear regression concludes that manual transmission accounts for a roughly 3 MPG difference over automatic transmission.

Loading and Cleaning the Data

First, load in the data from the {datasets} library.

```
library(datasets)
data(mtcars)
```

Reclassify the "vs" and "am" variables as factors.

```
mtcars$vs <- factor(mtcars$vs, labels = c("V", "Straight"))
mtcars$am <- factor(mtcars$am, labels = c("Automatic", "Manual"))
```

Exploratory Data Analysis

Use the `summary` function to get a glimpse of the clean data. In the interest of saving space, we'll just look at the "mpg" and "am" variables.

```
summary(mtcars[,c(1,9)])
```

```
##       mpg                am
##  Min.   :10.40   Automatic:19
##  1st Qu.:15.43   Manual   :13
##  Median :19.20
##  Mean   :20.09
##  3rd Qu.:22.80
##  Max.   :33.90
```

See **Appendix A** for a boxplot that compares range of average MPG for vehicles with automatic and manual transmissions.

Comparing Automatic and Manual Transmissions

Using the `summarize` function from the {dplyr} library, calculate the mean MPG for vehicles with automatic and manual transmissions.

```
library(dplyr)
am.means <- mtcars %>%
        group_by(am) %>%
        summarize(mean = mean(mpg))
```

am	mean
Automatic	17.15
Manual	24.39

Because the sample size for each group is low (< 30), use a two-sample t-test to compare the two groups. The null hypothesis we are testing is that there is no difference in MPG between vehicles with automatic transmissions and those with manual transmissions.

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

```
##
##  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
```

The p-value (0.001374) is less than 0.05, so we reject the null hypothesis. This answers our first question: **Manual transmission is better than automatic transmission for MPG**.

Quantifying the Difference Between Automatic and Manual Transmissions

Use the `step` function to choose the best model.

```
step.model <- step(lm(mpg ~ ., data = mtcars))
```

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4811 -1.5555 -0.7257  1.4110  4.6610
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.6178     6.9596   1.382 0.177915
## wt           -3.9165     0.7112  -5.507 6.95e-06 ***
## qsec          1.2259     0.2887   4.247 0.000216 ***
## amManual      2.9358     1.4109   2.081 0.046716 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.459 on 28 degrees of freedom
## Multiple R-squared:  0.8497, Adjusted R-squared:  0.8336
## F-statistic: 52.75 on 3 and 28 DF,  p-value: 1.21e-11
```

As you can see, the best fit model includes the "am" variable, which is fortunate for the purposes of this analysis, and it explains about 85 percent of the variance. See **Appendix B** for the residual plots, which

show no clear pattern.

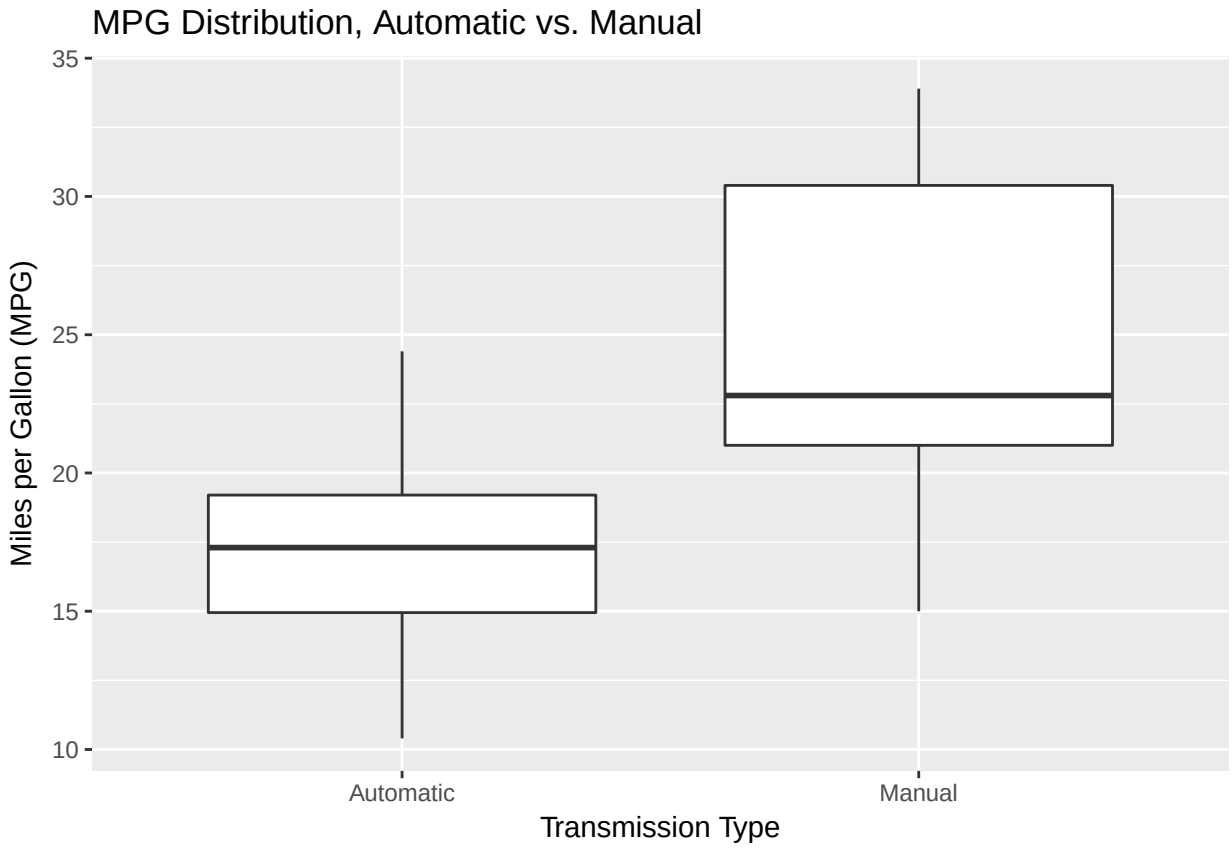
This answers our second question: **Manual transmission improves MPG by 2.94 units, over automatic transmission**. The p-value (0.046716) is less than 0.05.

Appendices

Appendix A

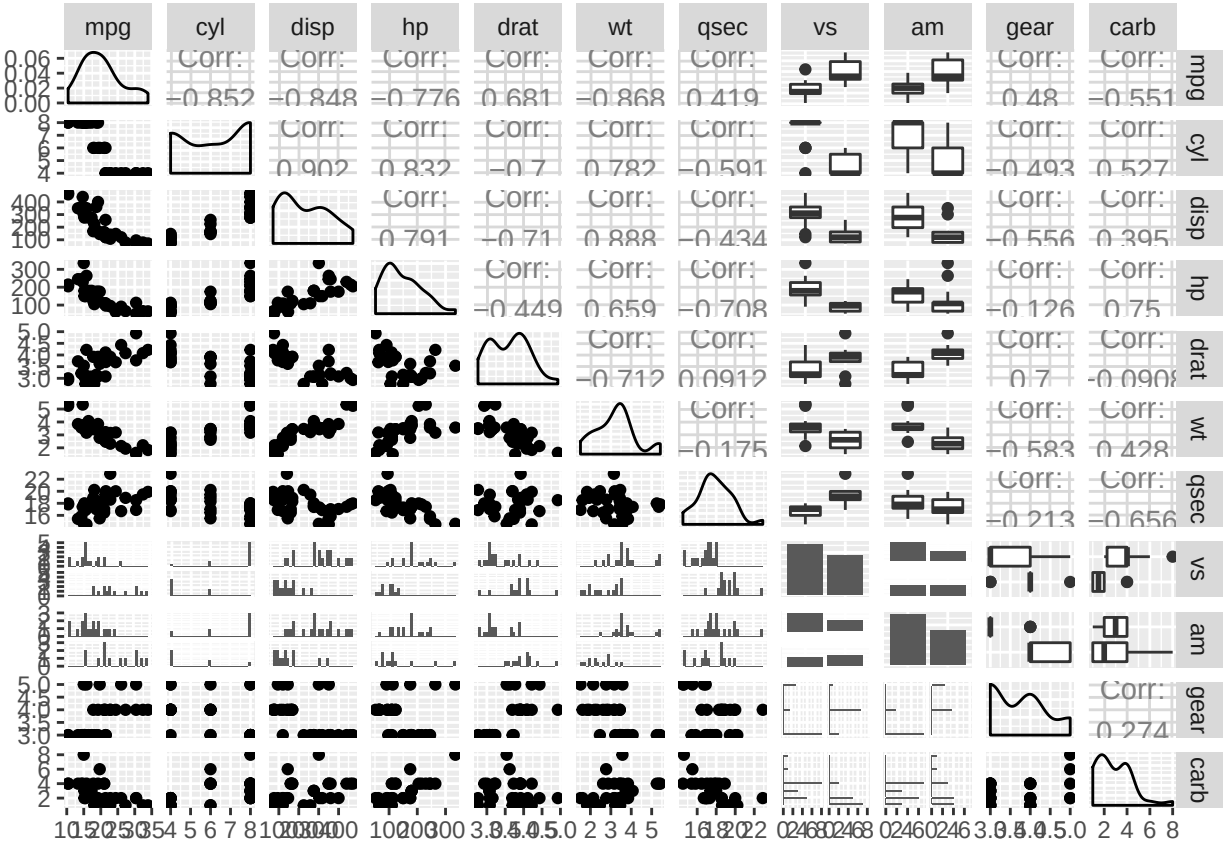
Use `ggplot` to create a boxplot of MPG by automatic vs. manual transmissions.

```
library(ggplot2)
ggplot(mtcars, aes(x = am, y = mpg)) +
        geom_boxplot() +
        labs(y = "Miles per Gallon (MPG)", x = "Transmission Type") +
        ggtitle("MPG Distribution, Automatic vs. Manual")
```



And use `ggpairs` to create a matrix that shows the intersection of all variables.

```
library(GGally)
ggpairs(mtcars)
```



Appendix B

Below, find the residual plots for the model used in the analysis.

```
par(mfrow = c(2,2))
plot(step.model)
```