Motor Trend Car Road Tests - Effect of transmission on MPG

Dhruv Mishra

Executive Summary

In this project, we will analyze the `mtcars` data set and explore the relationship between a set of variables and miles per gallon (MPG) (outcome). Two particular questions answered are:

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

Data loading and preprocessing

```
data(mtcars)
library(ggplot2)
mtcars$cyl <- factor(mtcars$cyl)
mtcars$vs <- factor(mtcars$vs)
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
mtcars$am <- factor(mtcars$am,labels=c('Automatic','Manual'))
```

Exploratory data analysis

Please refer to appendix for the plots. As can be seen, MPG is generally more for manual transmission as compared to automatic transmission

Which is better: Automatic or Manual?

To answer this question, let's fit a basic model with am as the only predictor.

```
fit.with.am <- lm(mpg~am, data = mtcars)
summary(fit.with.am)
```

```
##
## Call:
## lm(formula = mpg ~ am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.147      1.125  15.247 1.13e-15 ***
```

```
## amManual       7.245      1.764   4.106 0.000285 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.902 on 30 degrees of freedom
## Multiple R-squared:  0.3598, Adjusted R-squared:  0.3385
## F-statistic: 16.86 on 1 and 30 DF,  p-value: 0.000285
```

Transmission accounts for 36% of the variance with a p-value<0.0003 implying that the model is statistically significant and the difference between the two transmission modes exits. The intercept and slope coefficients show that automatic transmissions achieve 17.147 mpg and manual transmission cars achieve 17.147 + 7.245 = 24.39 mpg. The 95% Confidence Intervals for manual and automatic transmission can be seen as below:

```
am.automatic <- t.test(mtcars$mpg[mtcars$am == "Automatic"], mu = 0)
am.manual <- t.test(mtcars$mpg[mtcars$am == "Manual"], mu = 0)
am.automatic$conf.int
```

```
## [1] 15.29946 18.99528
## attr(,"conf.level")
## [1] 0.95
```

```
am.manual$conf.int
```

```
## [1] 20.66593 28.11869
## attr(,"conf.level")
## [1] 0.95
```

We can see that the 95% CI for automatic transmission is (15.3, 19) is strictly below the CI for manual transmission which is (20.67, 28.12). This is evidence that manual transmission yields better fuel economy.

Quantifying the difference

To be able to quantify the difference, we will have to create a model which explains the maximum variance. We build an initial model with all the variables as predictors and use both forward selection and backward elimination to arrive at the best model.

```
fit.all <- lm(mpg ~ ., data = mtcars)
fit.best <- step(fit.all, direction = "both")
```

```
summary(fit.best)
```

```
##
## Call:
## lm(formula = mpg ~ cyl + hp + wt + am, data = mtcars)
##
## 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 **
## amManual     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
```

From the above details, we observe that the model consists of the variables `cyl`, `wt` and `hp` as confounders and `am` as the independent variable. The adjusted R2 value is 0.84 which is the maximum obtained considering all combinations of variables. Thus, we can conclude that more than 84% of the variability is explained by this model.

Lets compare `fit.best` with `fit.with.am`.

```
anova(fit.best,fit.with.am)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ cyl + hp + wt + am
## Model 2: mpg ~ am
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     26 151.03
## 2     30 720.90 -4   -569.87 24.527 1.688e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value is highly significant implying that the additional variables actually contribute to the accuracy of the model.

Conclusion

- Manual transmission is better for MPG by a factor of 1.8 compared to automatic transmission.
- Means and medians for automatic and manual transmission cars are significantly different.

Appendix

Figure-1

```
ggplot(data=mtcars, aes(y=mpg, x = am)) + geom_boxplot() + facet_wrap(~ am) + ylab("MPG") + xlab("Trasmission") + ggtitle("MPG by transmission type")
```