Is an Automatic or Manual Transmission Better for Fuel Consumption (a blast from the past - 1974 dataset analysis)?

Rajesh Sinha

Executive Summary

This analysis is based on 1974 Motor Tend US dataset which comprises of fuel consumption of 31 automobiles of (1973-74 models). Based on regression models, we can confirm the following

- Fuel consumption is best explained by a regression model which has weight, cyclinder, horsepower as confounding variables and transmission type as independent predictor
- Automatic Transmission is less fuel efficient than Manual transmission
- Based on our model, we can estimate that the manual transmission cars are 1.80 miles per gallon more fuel efficient than automatic cars.

Exploratory Analysis

```
data(mtcars)
df <- mtcars
df$am <- as.factor(df$am)
df$cyl <- factor(df$cyl)
df$vs <- factor(df$vs)
df$gear <- factor(df$gear)
df$carb <- factor(df$carb)
levels(df$am) <- c("Automatic", "Manual")
x <- df %>% select(mpg, am) %>% group_by(am) %>% summarize(mn = mean(mpg))
```

Within exploratory analysis, we see that

- The boxplot in appendix (Figure-1) clearly shows that there is a difference betwen automatic and manual fuel consumption. The average fuel consumption of automatic car is 17.1473684 and for manual is 24.3923077 miles per gallon.
- This difference is statistically significant. If we perform t test for manual and automatic transmission data using t-test, we can see that we can reject the null hypothesis that the there is no difference between the averages of the mpg between the transmission types.

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

```
##
##  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
```

Building Regression Models

First we see how well does a simple regression model of mpg as a predicted value with only transmission type as predictor is

Initial Model based on Transmission Type only

```r
fit.am <- lm(mpg ~ am, data = df)
summary(fit.am)
```

```
##
## Call:
## lm(formula = mpg ~ am, data = df)
##
## 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
```

Note the following

- the p-value is smaller than 0.05 implying that there exists a statistically significant difference between means of manual versus automatic transmission cars
- from intercepts, we can see that automatic transmission achieve 17.147 mpg whereas the manual ones do 24.39 mpg
- The adjusted R^2 value is 0.34 implying that only 36% is explained by transmission type as predictor

Based on the above, we **cannot** confirm that the difference between the manual and automatic transmission is actually **7.254** as the model is not good enough.

Finding the Best Fit Model

Given the poor explanation of variability of mpg only by transmission type, we now find what other predictors are important.

```
base.model <- lm(mpg ~ ., data = df)
best.model <- step(base.model, direction = "both")
```

```
summary(best.model)
```

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

Note the following

- The adjusted R^2 is about 84% which explains sizeable variability of mpg
- Beyond transmission type as an independent variable, it also includes weight, cyclinder and horsepower as confounding variables.
- The difference between the automatic and manual transmission now stands at **1.80 mpg**.

To confirm the importance of additional predictors, we can run anova of best.model with fit.am

```
anova(best.model, fit.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 low p value is indicative of statistically significant difference between the two models.

Figure 2 in the appendix shows the checking of residuals for signs of non-normality & residuals vs fitted values plot for signs of heteroskedasticity. residual diagnostics show normality & no evidence of heteroskedasticity. Over here we can find that the points in the Residuals vs. Fitted plot are randomly scattered on the plot that verifies the independence condition.The Normal Q-Q plot consists of the points which mostly fall on the line indicating that the residuals are normally distributed. The Scale-Location plot consists of points scattered in a constant band pattern, indicating constant variance. There are some distinct points of interest (outliers or leverage points) in the top right of the plots that may indicate values of increased leverage of outliers.

Appendix

Figure 1. Boxplot of automated vs manual cars

```
means <- aggregate(mpg ~ am, df, mean)
ggplot(df, aes(x = am, y = mpg)) + geom_boxplot(aes(fill = am)) + facet_wrap(~am) +
    ylab("MPG") + xlab("Trasmission") + ggtitle("Figure-1: MPG by transmission type") +
    stat_summary(fun.y = mean, colour = "darkred", geom = "point", shape = 18,
        size = 3, show_guide = FALSE) + geom_text(data = means, aes(label = round(mpg,
    2), y = mpg + 0.08), size = 3)
```

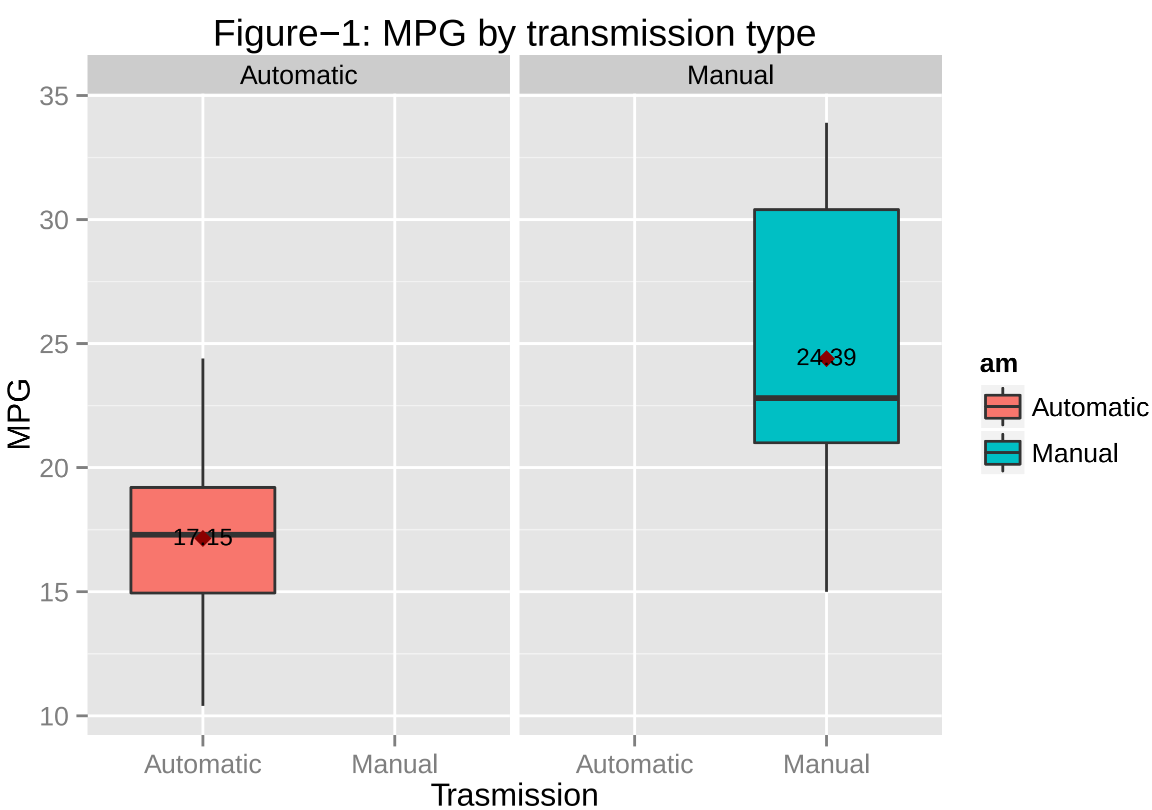

Figure 2. Residuals Plotting

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