

Regression Models Course Project: Effect of Transmission Type on MPG

Patrick Close

December 27, 2015

Executive Summary

In this project we will analyze the **mtcars** dataset to determine the effect of several variables on the response (mpg). In particular, we are interested in two questions:

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

Data Transformations

This analysis requires the following transformations:

```
mtcars$cyl <- factor(mtcars$cyl)
mtcars$vs <- factor(mtcars$vs)
mtcars$am <- factor(mtcars$am, labels = c("Auto", "Man"))
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
```

Exploratory Data Analysis

To better understand the data set we plot the relationships between all variables of the dataset. The resulting scatterplot matrix can be seen as Figure 2 in the appendix. It appears several variables have at least a moderate correlation with mpg, something we will try to confirm in subsequent analysis.

As can be seen in the boxplot in Figure 1 in the appendix, cars with a manual transmission appear to have a higher mpg than cars with an automatic transmission. Additionally, as confirmed by the p-value < 0.05 from the below t-test, there is a significant difference in mpg by transmission type.

```
t_test <- t.test(mpg ~ am, data = mtcars)
print(t_test$p.value)
```

```
## [1] 0.001373638
```

Regression analysis

First we will build a simple linear model with the response mpg with a single factor, transmission type (am)

```
#Regression Analysis
#simple linear model of mpg by am
lmodel <- lm(mpg ~ am, data = mtcars)
summary(lmodel)
```

Second, we will build a multifactor model and compare the resulting adjusted R-squared values to determine if the multifactor model offers better explanatory power

```
#multiple regression model
lmodels_all <- lm(mpg ~., data = mtcars)
bestmodel <- step(lmodels_all, direction = "both")
sum_lmodel <- summary(lmodel)
sum_best <- summary(bestmodel)
```

```
##                r.squared adj.r.squared
## single factor model 0.3597989      0.3384589
## multi-factor model  0.8658799      0.8400875
```

From the superior adjusted R-squared of the multi-factor model, we can conclude that the inclusion of additional factors adds explanatory power to the model. Additionally, using the **anova** function, we can see the p-value < 0.05.

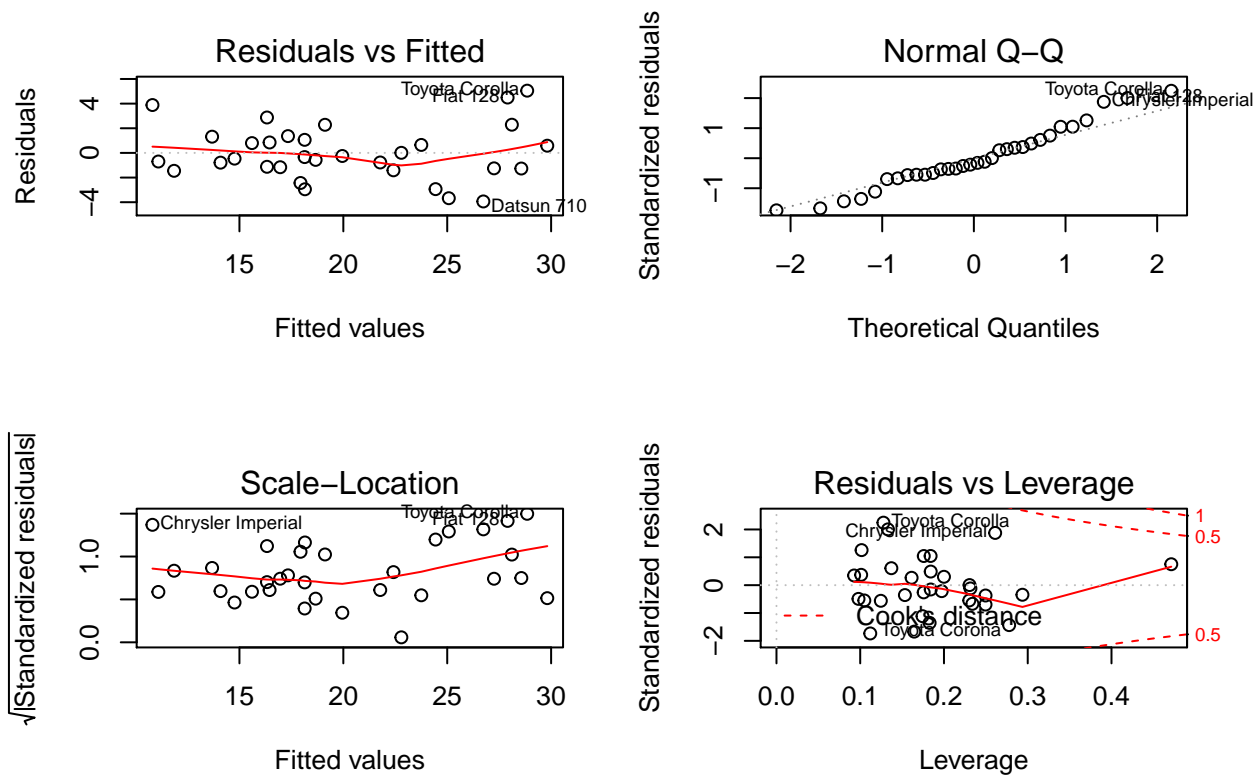
```
anova(lmodel, lmodels_all)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb
##   Res.Df  RSS Df Sum of Sq    F   Pr(>F)
## 1      30 720.9
## 2      15 120.4 15    600.49 4.9874 0.001759 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residuals

bestmodel produces the following residual plots:

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



From the above plots, we can make the following observations:

- * The points in the Residuals vs. Fitted plot seem to be randomly scattered on the plot and verify 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.

Conclusions

- Cars with Manual transmission get more miles per gallon mpg compared to cars with Automatic transmission. (1.8 adjusted by hp, cyl, and wt).
- mpg will decrease by 2.5 (adjusted by hp, cyl, and am) for every 1000 lb increase in wt.
- mpg decreases negligibly with increase of hp.
- If number of cylinders, cyl increases from 4 to 6 and 8, mpg will decrease by a factor of 3 and 2.2 respectively (adjusted by hp, wt, and am).

Appendix

Figure 1: Boxplot of mpg by am

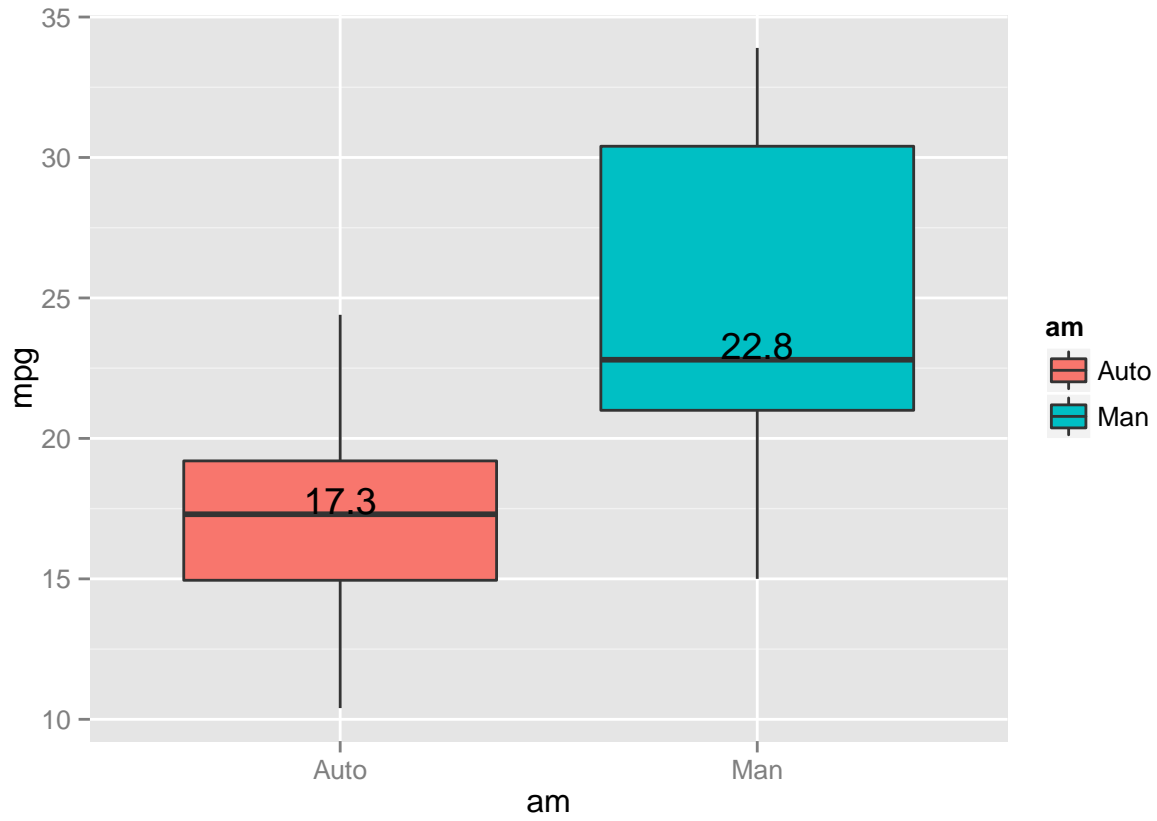


Figure 2: Scatterplot matrix of factors of mtcars

