

Analysis of MPG vs Transmission

Eric Scuccimarra

9 January 2018

Executive Summary

The objective is to analyze the data from the 1974 Motor Trend road tests of cars to determine if there is a relationship between type of transmission and miles per gallon. To accomplish this multiple linear models are fitted and examined. The best fitting model includes three regressors - weight, quarter mile time and transmission type. In this model manual transmission provides a 2.9 MPG advantage over automatic transmissions.

Loading and Preprocessing Data

The data is part of R's datasets library and is loaded as follows:

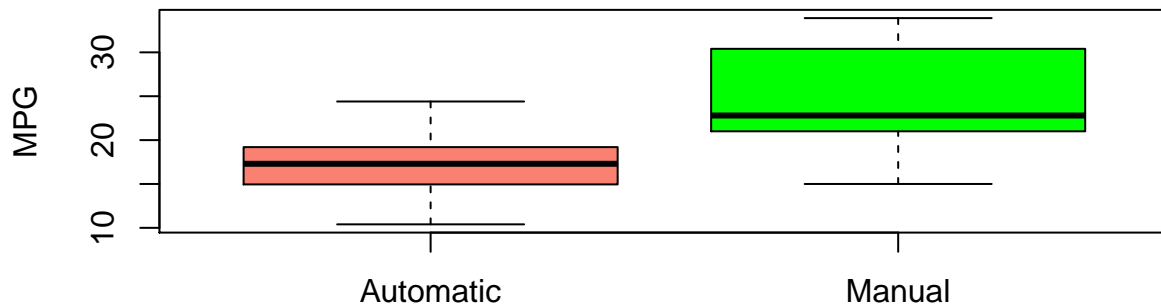
```
library(datasets)
mtcars$am <- as.factor(mtcars$am)
levels(mtcars$am) <- c("Automatic", "Manual")
mtcars$cyl <- as.factor(mtcars$cyl)
mtcars$gear <- NULL
```

The data was preprocessed by converting transmission type and number of cylinders to factors and removing the number of gears as that only is relevant to manual cars.

Exploratory Analysis

A boxplot of MPG per transmission type:

```
boxplot(mtcars$mpg ~ mtcars$am, col=c("salmon","green"), ylab="MPG")
```



This boxplot indicates that there does appear to be a substantial difference between the ranges for Automatic and Manual transmissions, with manual having not only a higher median, but also having a higher 1st quartile than the 3rd quartile of automatics.

However almost all of the variables in the data appear to be correlated with MPG, with the most significant correlation existing between MPG and weight, cylinders, displacement and horsepower, in order of descending correlation.

I will also fit a linear model comparing MPG to transmission:

```
fit0 <- lm(mpg ~ am, data = mtcars)
summary(fit0)$coef

##              Estimate Std. Error  t value    Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## amManual     7.244939   1.764422  4.106127 2.850207e-04

summary(fit0)$r.squared

## [1] 0.3597989
```

As we already knew, transmission type is correlated with MPG, but only explains 36% of the variability, and given the correlations it would seem we can fit a better model. There were also some concerns about leverage raised by the diagnostic plot of this fit which are in the appendix.

Regression Analysis

I will use the step function in the MASS library which will automate selection of a model using the AIC step-wise algorithm.

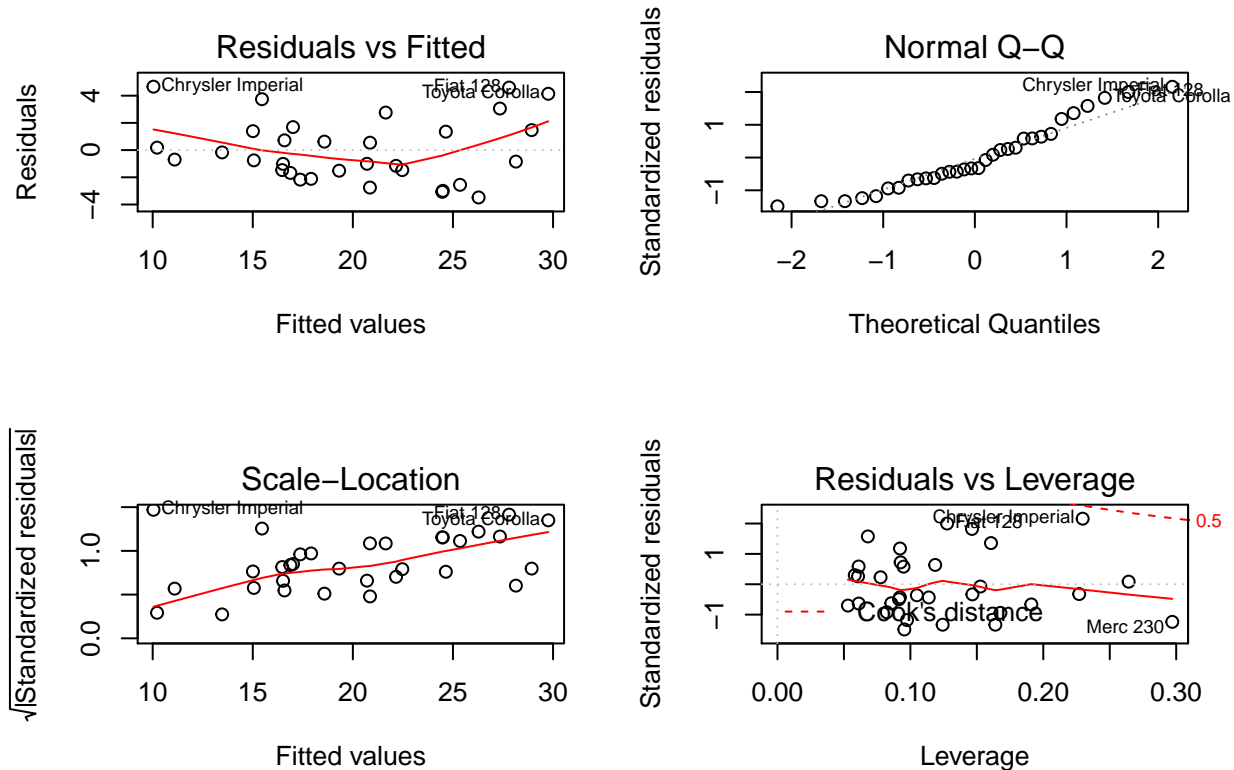
```
library(MASS)
fit1 <- lm(mpg ~ ., data=mtcars)
sfit1 <- step(fit1, direction="both")

summary(sfit1)

##
## 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
```

Step has chosen weight + quarter mile time plus transmission type as the best model, which explains 85% of the variability in MPG. The ANOVA output of the step command are in the appendix.

```
par(mfrow=c(2,2))
plot(sfit1)
```



There are also some concerns with this model, namely the Q-Q plot deviates from the diagonal, indicating some non-normal data, and the residuals vs fitted line appears to be skewed. To verify the selection of this model I attempted to fit some other models which intuitively seemed to make sense and while came close to being as accurate as the model picked by step, they all had higher standard error and lower R^2 . For the sake of brevity those calculations are not detailed here.

```
mmanual <- mean(subset(mtcars, am == "Manual")$wt)
mauto <- mean(subset(mtcars, am == "Automatic")$wt)
```

One notable caveat is that there is a substantial difference in the mean weights between manual and automatic cars. The mean weight of manual cars is 2.411 while the mean weight of automatic cars is 3.7688947. This makes it difficult to extrapolate how much of the difference in MPG is due to the transmission and how much is due to the extra weight of automatic autos.

This is accounted for in the summary of the model above, where significance of including the transmission type in the model is dangerously close to the 0.05 cutoff, with a $\Pr(>t)$ of 0.467.

Conclusion

Disregarding other variables manual transmission provides a 7.2 MPG advantage over automatic transmission. Our best fit model includes weight, quarter mile time and transmission type with manual transmission with

manual transmission providing an average 2.9 MPG gain over automatic transmission.

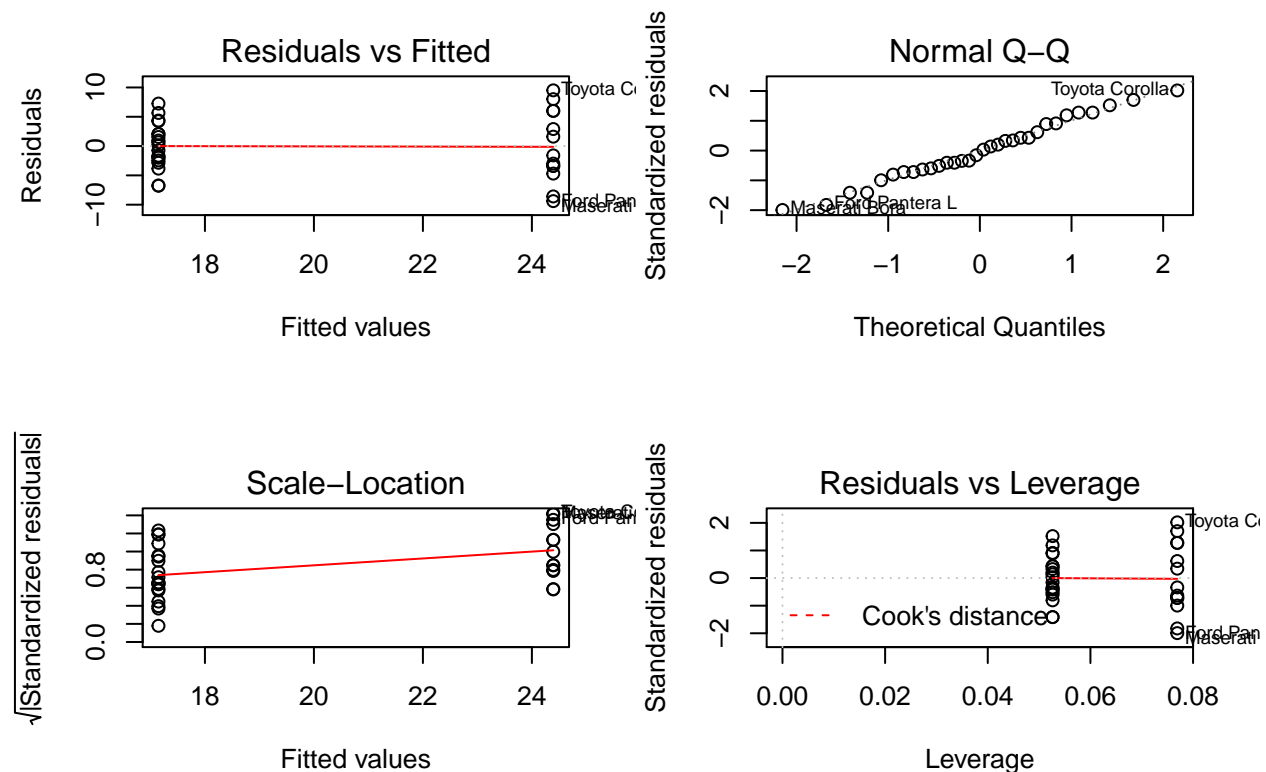
The feature most correlated to the MPG is the weight of the vehicle, which is also highly correlated to the transmission type. While the model include the transmission type does provide a better fit, the significance of adding the transmission type variable is close enough to the 5% threshold that I cannot confidently assert that the difference in MPG between automatic and manual transmissions is not due to the extra weight associated with automatic transmissions without additional data.

I should also stress that this data is from 1974 so may not be relevant to modern automobiles.

Appendix

Diagnostic plot of fit between MPG and transmission

```
par(mfrow=c(2,2))
plot(fit0)
```



ANOVA Results of Step Process

```
sfit1$anova
```

##	Step	Df	Deviance	Resid. Df	Resid. Dev	AIC
## 1	NA	NA	NA	21	135.1573	68.10251
## 2	- drat	1	0.02529014	22	135.1826	66.10850
## 3	- vs	1	4.25043766	23	139.4330	65.09916

## 4 - carb	1	2.89754287	24	142.3306	63.75733
## 5 - disp	1	1.65114072	25	143.9817	62.12642
## 6 - cyl	2	16.08472969	27	160.0665	61.51530
## 7 - hp	1	9.21946935	28	169.2859	61.30730