# Cars' MPG Analysis for Motor Trend Magazine

by Diana H

### Executive Summary

The mtcars dataset is loaded and examined using explanatory data analysis and regression models were fit in R to give Motor Trend magazine data-based answers for:
"Is an automatic or manual transmission better for MPG" *Answer :yes*
"Quantify the MPG difference between automatic and manual transmissions" *Answer :1.8 miles/gallon*

### Loading the Data

**Description of the Dataset** As quoted from R help file: "The data was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models)." The names of the variables and their description can be seen in this table.

| names | mpg | cyl | disp | hp | drat | wt | qsec | vs | am | gear | carb |
|---|---|---|---|---|---|---|---|---|---|---|---|
| desc | Miles/(US) gallon | Number of cylinders | Displacement (cu.in.) | Gross horsepower | Rear axle ratio | Weight (1000 lbs) | 1/4 mile time | Engine (0 = V-shaped, 1 = straight) | Transmission (0 = automatic, 1 = manual) | Number of forward gears | Number of carburetors |

```
data <- mtcars
```

**Loading the Data** The types of all the variables were checked using the *class* function and they are all numeric. The variables that refer to categorical CYL, VS, AM, GEAR, CARB, were changed to factors.

```
# changing categorical variables to factors
for(i in c(2,8:11)){
data[,i]<- factor(data[,i])}
```

### Setting Glowbal Options

```
options(warn = -1)
knitr::opts_chunk$set(comment = NA)
knitr::opts_chunk$set(message = FALSE)
knitr::opts_chunk$set(opts.label="kill_prefix")
knitr::opts_template$set("kill_prefix"=list(comment=NA, null_prefix=TRUE))
```

### Exploratory Data Analysis

Plotting MPG as a function of AM (transmission type) using the code below. The result, Figure 1, can be seen in the appendix. From Figure 1, it can be seen that the MPG for manual cars is higher than that for automatic cars. So, transmission type (AM) affects MPG and manual transmission is better for MPG.

```
library(tidyverse)
data %>% ggplot(aes(am,mpg)) + geom_boxplot() +
  labs(title="Variation of MPG in Automatic and Manual Cars", x="Transmission Type",
  caption="0 automatic cars\n1 manual cars    ",y="MPG")+ theme_bw()
```

### Regression Model

First, a simple regression model was fit, then a multivariable regression model. The two were then examined then compared to each other and finally the better fit was selected.

**Fitting a Simple Linear Model** As a first step, a linear regression model with AM as the only predictor for MPG will be made, and the value of the resultant adjusted R squared will be checked.

```
fit_linear1 <- lm(mpg~am,data=data)
summary(fit_linear1)$coef
```

```
             Estimate Std. Error   t value     Pr(>|t|)
(Intercept) 17.147368   1.124603 15.247492 1.133983e-15
am1          7.244939   1.764422  4.106127 2.850207e-04
```

```
summary(fit_linear1)$r.squared
```

```
[1] 0.3597989
```

**Interpretation of the Coefficients (Simple Linear Model)** The adjusted R squared value for this simple linear regression is 0.3385. This means that the model used explains only 33.8% of the regression variance. Since the dataset has other variables, then a multivariable regression model will be tried. From the summary coefficients, it can be seen that the average MPG for automatic transmission is 17.147 and (7.2 + 17.147) for the manual.

**Fitting a Multivariable regression Model and Comparison** To make a good choice for the multivariable model, a pair plot, Figure 2, showing the correlation of each variable with MPG is made. From this figure, the variables that have the strongest correlation with MPG are: cyl,disp,hp,and wt.

```
pairs(mpg ~ .,data)
```

The variables cyl,disp,hp,and wt are used for the multivariable regression model. And the latter will be compared to the previous model.

```
fit_multi <- lm(mpg~ am +cyl + disp+ hp + wt,data)
summary(fit_multi)$coef
```

```
               Estimate Std. Error    t value     Pr(>|t|)
(Intercept) 33.864276061 2.69541569 12.5636562 2.668321e-12
am1          1.806099494 1.42107933  1.2709350 2.154510e-01
cyl6        -3.136066556 1.46909031 -2.1346996 4.277253e-02
cyl8        -2.717781289 2.89814941 -0.9377644 3.573375e-01
disp         0.004087893 0.01276729  0.3201848 7.514890e-01
hp          -0.032480178 0.01398322 -2.3227963 2.862128e-02
wt          -2.738694608 1.17597755 -2.3288664 2.824553e-02
```

```
summary(fit_multi)$r.squared
```

```
[1] 0.8664276
```

Anova test is used to compare the two models.

```
anova(fit_linear1,fit_multi)
```

```
Analysis of Variance Table

Model 1: mpg ~ am
Model 2: mpg ~ am + cyl + disp + hp + wt
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1     30 720.90
2     25 150.41  5    570.49 18.965 8.637e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

**Interpretation of the Coefficients (Multivariable regression Model) and of Anova Test** The p-value from the anova test is very small:8.637e-08, and hence the new multivariable regression model is better than the initial simple model. In the model's summary, the adjusted R squared value now suggests that the multivariable model used with those predictors explains 83.44% of the regression variance. The quantified difference is now 1.8 miles/gallon while in the previous model it was 7.2 miles/gallon.

### Residual Plot and Diagnostics

Looking at the residual vs fitted graph in figure below we can see that our residuals do not follow a known pattern neither are they heteroskedastic. This means that the model is well fitted and no overly used predictors. Also the quantile plot shows that the data are approximately normally distributed.

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

### Conclusions and Uncertainties

So, the multivariable model with predictors (model followed) AM, CYL, DISP, HP, WT predictors is a good fit and better than the simple model with only the AM predictor. This analysis showed that manual transmission is better than automatic transmission for MPG and the quantified difference according to the final model followed is 1.8 miles more for manual transmission per US gallon. This difference is less than in the first model, 7.2 miles/gallon and less than the difference on the boxplot, Figure 1, which is around 7 miles/gallon. So, even though it is evident/certain that manual transmission is better than automatic transmission for MPG yet it is not very certain that the difference is 1.8 as per the model followed. This is maybe due to the fact that when a regression model improves variance explanation, the mean is to some extent compensated.

**Appendix**

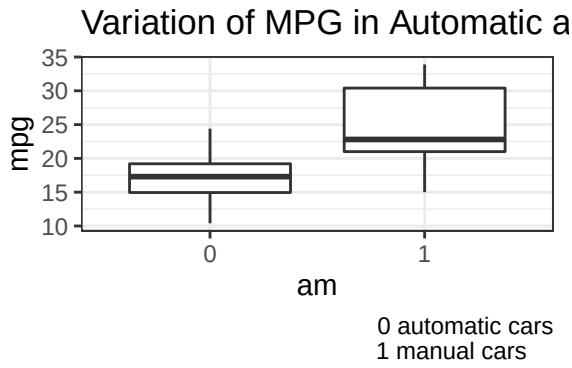


Figure 1: Boxplot of MPG for the different transmission types, automatic and manual.

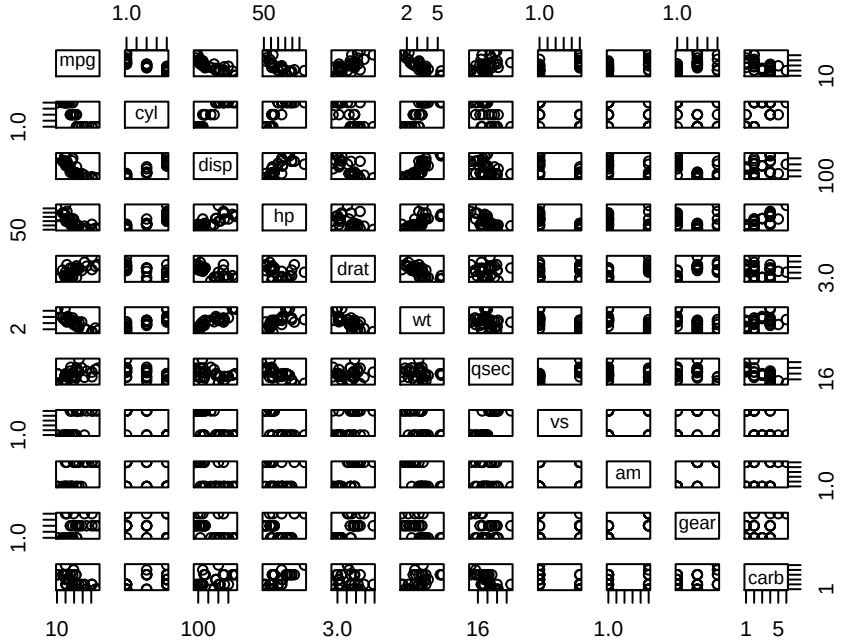


Figure 2: Correlation of MPG with the other variables

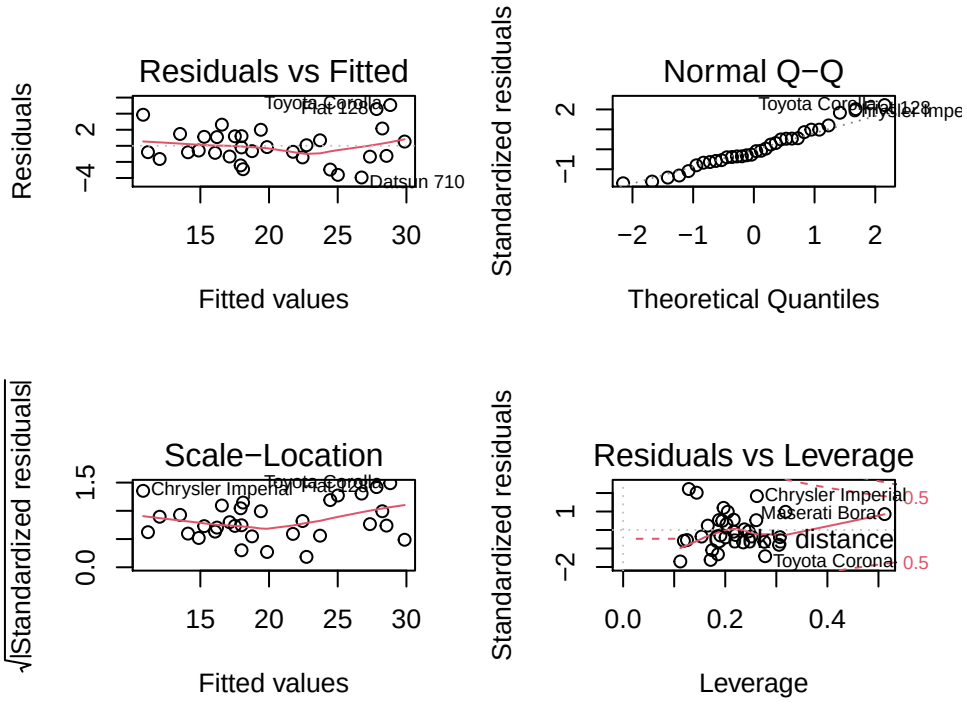


Figure 3: Residual Plots of the Multivariable regression Model