

# Regression Models Course Project: Effects of manual and automatic transmission on MPG

Amber Beasock

31 January 2016

---

## Executive Summary

In this project, we analyze the `mtcars` (Motor Trend Car Road Tests) data set and explore the relationship between a set of variables and miles per gallon (MPG). 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 main focus of the project is to answer the following questions:

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

In our analysis, we use exploratory data analysis and regression models to explore the impact of manual transmissions (`am=1`) and automatic transmissions (`am=0`) on MPG (Miles/(US) gallon).

Our analysis concludes that manual transmission is better for MPG. The mean for manual is about 7 mpg more than the mean for automatic transmission. Linear regression analysis shows that manual transmission cars get an average of 2.94 mpg more than automatics when weight and 1/4 mile time are held constant.

## Data Processing

```
library(corrplot)
```

Load the data set `mtcars`. Use `?mtcars` to view the R documentation on the data set to find out more information. From this we learn the mapping of the variable `am`, 0=automatic, 1=manual. Transform some variables from numeric to factor.

```
data(mtcars)
head(mtcars,3) #Preview of the dataset
```

```
##           mpg cyl  disp  hp  drat    wt  qsec vs am gear carb
## Mazda RX4    21.0   6  160 110  3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag 21.0   6  160 110  3.90 2.875 17.02  0  1    4    4
## Datsun 710    22.8   4  108  93  3.85 2.320 18.61  1  1    4    1
```

```
dim(mtcars)
```

```
## [1] 32 11
```

```
mtcars_num <- mtcars #Need the variables as numeric to check correlation
mtcars$cyl <- as.factor(mtcars$cyl)
mtcars$vs <- as.factor(mtcars$vs)
mtcars$am <- factor(mtcars$am, labels=c("automatic", "manual"))
mtcars$gear <- as.factor(mtcars$gear)
mtcars$carb <- as.factor(mtcars$carb)
```

## Exploratory Data Analysis

The plots created in the exploratory data analysis can be found in the **Appendix: Figures** section at the end. According to the box plot **Figure 1**, manual transmission yields a higher MPG. In **Figure 2**, a pair graph is used to generate scatterplots to show the relationship between variables. A correlation matrix is also used to further examine the correlation between variables. **Figure 3** shows the strength of correlation between all variables.

```
x <- mtcars_num[1]
y <- mtcars_num[2:11]
cor(x, y)
```

```
##           cyl      disp      hp      drat      wt      qsec
## mpg -0.852162 -0.8475514 -0.7761684 0.6811719 -0.8676594 0.418684
##           vs      am      gear      carb
## mpg 0.6640389 0.5998324 0.4802848 -0.5509251
```

There are stronger correlations between variables: “cyl”, “disp”, “hp”, and “wt”.

## Statistical Inference

A two sample T-test is performed on the two types of transmissions (manual and automatic) to test the null hypothesis, which is the MPG of the automatic and manual transmissions are from the same distribution.

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

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

```
t_test$estimate
```

```
## mean in group automatic    mean in group manual
##           17.14737           24.39231
```

The p-value is 0.0014, so we reject the null hypothesis. The mean of manual transmission is about 7 MPG more than that of an automatic transmission.

## Regression Analysis

In the first model, all variables are included as predictors of MPG.

```
initModel <- lm(mpg ~ ., data=mtcars)
summary(initModel) # Results hidden
```

The residual standard error is 2.833 on 15 degrees of freedom. The Adjusted R-Squared value is 0.779, which is interpreted to mean this model can explain about 77.9% of the variance of the MPG. All of the coefficients are above the 0.05 significant level, so we can conclude none of the coefficients are significant.

Backward elimination can be used to find the most statistically significant predictor variables. The `step()` function starts with all the predictors in the model and then drops the one with the largest p-value (least significant). Then the model is re-fitted and this is repeated until only variables that are statistically significant remain.

```
bestModel <- step(initModel, k=log(nrow(mtcars)))
summary(bestModel) # Results hidden
```

The best model shows that “wt”, “qsec”, and “am” are statistically significant variables. The Adjusted R-Squared value is 0.8336, which is interpreted to mean this model can explain about 83.36% of the variance of the MPG.

Since the focus of this project is on the relationship between transmission and MPG, the next model will fit transmission as a predictor of MPG.

```
mpg_amModel <- lm(mpg ~ am, data=mtcars)
summary(mpg_amModel) # Results hidden
```

This model shows that a car with an automatic transmission has an average of 17.147 mpg, and manual transmission increases mpg by 7.245. However, this model has an Adjusted R-Squared of 0.3385, which means the model can only explain about 33.85% of the variance of the MPG. Other variables should be added in to get a higher Adjusted R-Squared value.

Lastly, let’s compare the model with only “am” as the predictor variable with the best model obtained through backwards elimination.

```
anova <- anova(mpg_amModel, bestModel) # Results hidden
```

This results in a p-value of  $1.5504951 \times 10^{-9}$ , which is highly significant. Therefore, we reject the null hypothesis that variables “wt”, “qsec”, and “am” do not contribute to the prediction of mpg.

The best fitted model is `lm(formula = mpg ~ wt + qsec + am, data = mtcars)`

```
confint(bestModel)
```

```
##              2.5 %    97.5 %
## (Intercept) -4.63829946 23.873860
## wt          -5.37333423 -2.459673
## qsec         0.63457320  1.817199
## ammanual     0.04573031  5.825944
```

```
summary(bestModel)$coef
```

```
##              Estimate Std. Error  t value    Pr(>|t|)
## (Intercept)  9.617781   6.9595930  1.381946 1.779152e-01
## wt          -3.916504   0.7112016 -5.506882 6.952711e-06
## qsec         1.225886   0.2886696  4.246676 2.161737e-04
## ammanual     2.935837   1.4109045  2.080819 4.671551e-02
```

This model shows that when “wt” and “qsec” remain constant, manual transmission cars get an average of 2.94 more MPG than those with automatic transmission.

## Residual Plots & Diagnostics

Refer to Figure 4 in the **Appendix: Figures** section for the residual plots. Interpretation of the residual plots:

- The Residuals vs Fitted plot - the points are randomly scattered on the plot confirming the independence condition.
- The Normal Q-Q plot - the majority of points fall on the line indicating normal distribution of the residuals.
- The Scale-Location plot - there is a random band around the line with no clear pattern indicating constant variance.
- The Residual vs Leverage plot - there are some outliers (leverage points) that may indicate values of increased leverage of outliers.

Next, regression diagnostics can be used to further investigate our model.

The `hatvalues()` function is used to find values far from the average (usually 2-3x the average) because these may have substantial influence on the regression parameters.

```
hv_mean<- mean(hatvalues(bestModel))
hv <- hatvalues(bestModel)
tail(sort(hv),3)
```

```
##   Chrysler Imperial Lincoln Continental      Merc 230
##           0.2296338           0.2642151           0.2970422
```

The average hatvalue is 0.125, and from the above step we get the cars that are about 2 times the average or greater. These are the same cars that we see mentioned on the residual plots.

## Conclusion

Cars with manual transmission get better miles per gallon compared to those with automatic transmission.

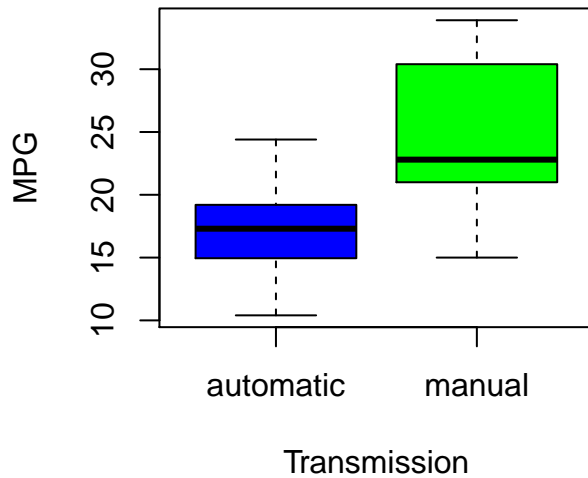
- The t-test shows that manual transmission gets an average of 7 MPG more than cars with automatic transmission.
- Several linear regression models were fitted to evaluate different aspects that could impact MPG. The best fitted model `lm(formula = mpg ~ wt + qsec + am, data = mtcars)` showed that when “wt” (Weight (lb/1000)) and “qsec” (1/4 mile time) remain constant, manual transmission cars get an average of 2.94 more MPG than those with automatic transmission.

## Appendix: Figures

Figure 1 - Boxplot of MPG vs Transmission

```
boxplot(mtcars$mpg~mtcars$am, xlab="Transmission", ylab="MPG", main="Figure 1: Boxplot", col=c("blue", "red"))
```

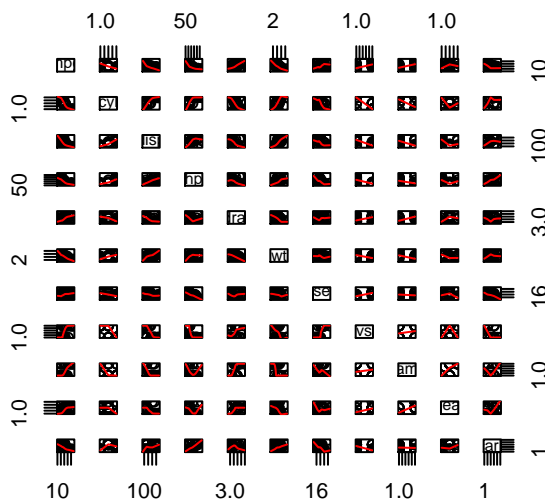
**Figure 1: Boxplot**



**Figure 2 - Pair Graph of Motor Trend Car Road Tests**

```
pairs(mpg ~ ., mtcars, panel=panel.smooth, main="Figure 2: Pair Graph")
```

**Figure 2: Pair Graph**



**Figure 3 - Correlation Plot**

```
cor2 <- cor(mtcars_num)
corrplot(cor2, method="color")
```

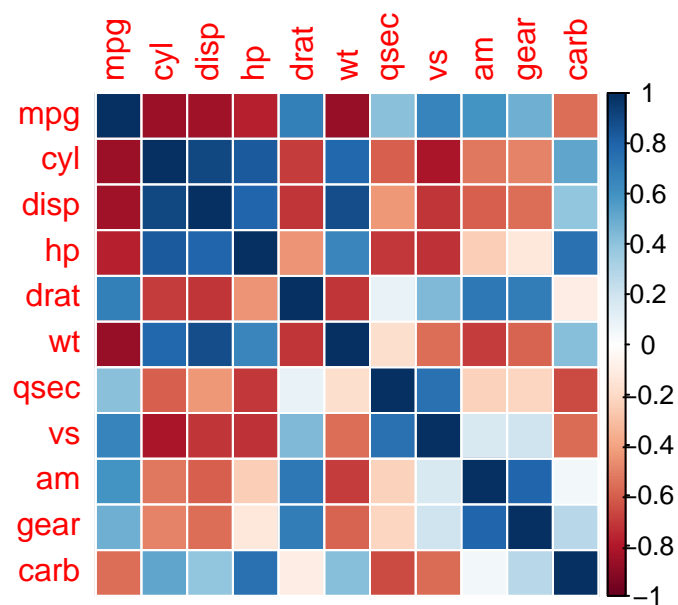


Figure 4 - Residual Plots

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

