Motor Trend Course Project Analysis

John Letteboer

11/25/2018

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.

Exploratory analysis and visualizations are located in the Appendix to this document.

Data prep

Load the data and do some checks.

```
data(mtcars)
dim(mtcars)
```

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

```
head(mtcars,3)
```

```
##                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
```

```
str(mtcars)
```

```
## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

Exploratory Analysis

Let's check the summary of the variable "mpg".

```
summary(mtcars$mpg)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.40   15.43   19.20   20.09   22.80   33.90
```

Let's check automatic and manual seperate

```
summary(mtcars$mpg[mtcars$am==0])
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.40   14.95   17.30   17.15   19.20   24.40
```

```
summary(mtcars$mpg[mtcars$am==1])
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   15.00   21.00   22.80   24.39   30.40   33.90
```

To see a significant difference in the mean of "mpg" for automatic and manual transmission let's do a T-Test and show the estimate. See Appendix A.1 for the T-Test and A.2 for the boxplot.

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

```
## mean in group 0 mean in group 1
##        17.14737        24.39231
```

```
t$p.value
```

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

The p-value is 0.0013736, so we reject the null hypothesis.

Let's check the correlations.

```
cor(mtcars[1], mtcars[2:11] )
```

```
##           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
```

According to the correlation table, there are at least four variables with a high correlation to our outcome variable "mpg". The highest value comes from the weight variable "wt". Let's have a look to this variable separately for automatic (0) and manual (1) transmission in Appendix A.3.

Linear Models

The linear dependencies suggest to analyse linear models as follows:

```
fit1 <- lm(mpg ~ am , data = mtcars)
fit2 <- lm(mpg ~ am + wt, data = mtcars)
fit3 <- lm(mpg ~ am + wt + hp , data = mtcars)
fit4 <- lm(mpg ~ am + wt + hp+ disp, data = mtcars)
fit5 <- lm(mpg ~ ., data = mtcars)
```

We start with the variable "mpg" as a function of the vairable "am" add one variable after another and do the ANOVA routine (see Appendix A.4) to find the simplest modell that explains significantly the change in "mpg". I didn't take the variable "cyl" for its high corellation with the variable "disp". We see that adding the variables "wt" and "hp" significantly improve the model, so it's the modell "fit3" which we use further. In Appendix A.5 you find the correlations of the four variables "used". In Appendix A.6 you find the summay of the modell "fit3" that explans about 84% of the variability of the variable "mpg".

Let's turn to the residuals of modell "fit3". In Appendix A.7 you find the plot of the residuals. It seems that some "outliers" should be analized more carefully but over all the fit of modell "fit3" and its residuals seem to satisfy basic requirement for a linear modell to explain the variation of the variable "mpg".

Conclusion

Is an automatic or manual transmission better for MPG?

It appears that manual transmission cars are better for MPG compared to automatic cars. However when modeled with confounding variables like displacement, HP and weight, the difference is not as significant as it seems in the beginning: a big part of the difference is explained by other variables.

Quantify the MPG difference between automatic and manual transmissions

Analysis shows that when only transmission was used in the model manual cars have an mpg increase of 7.245. However, when variables wt and hp are included, the manual car advantage drops to 2.084 with other variables contributing, sometimes more (e.g. weight) to the effect.

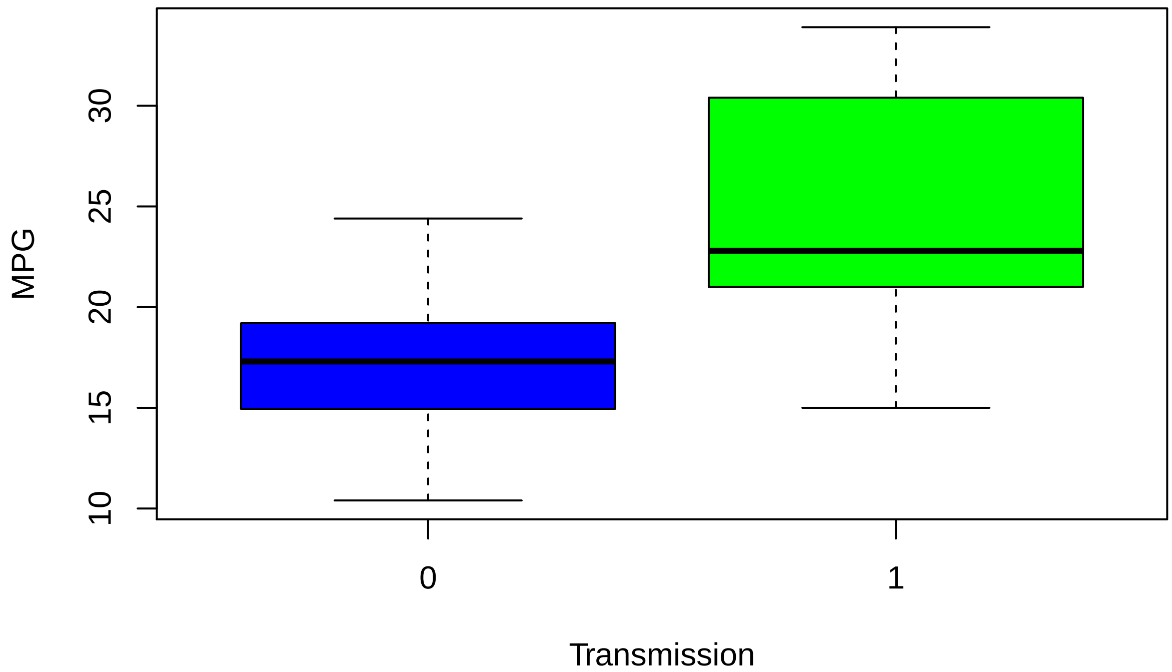
Appendix

A.1 t-Test for the variable "mpg" for Automatic and Manual Transmission

```
##
##  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 0 mean in group 1
##        17.14737        24.39231
```

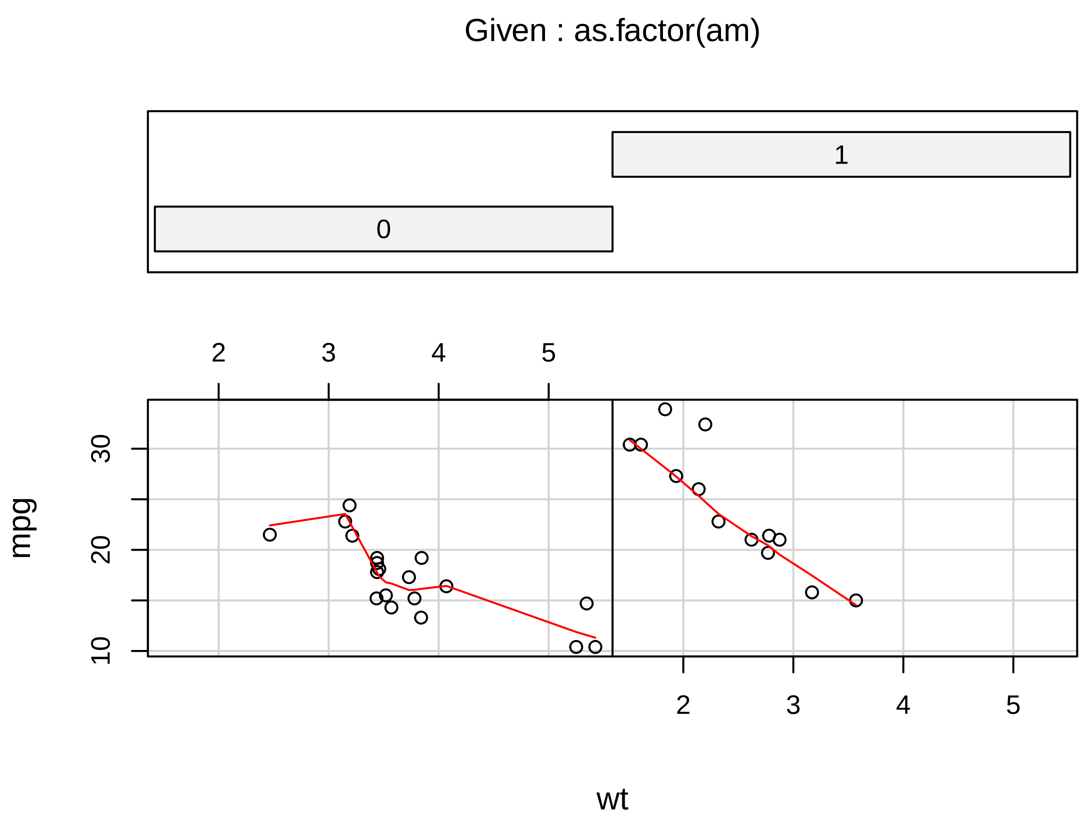
A.2 Boxplot: Mean of the variable "mpg" for Automatic and Manual Transmission

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



A.3 Coplot: Dependencies of weight for Automatic and Manual Transmission

```
coplot(mpg ~ wt | as.factor(am), data = mtcars, panel = panel.smooth, rows = 1)
```



There seems to be a quite linear dependency that differs in function of the variable “am”.

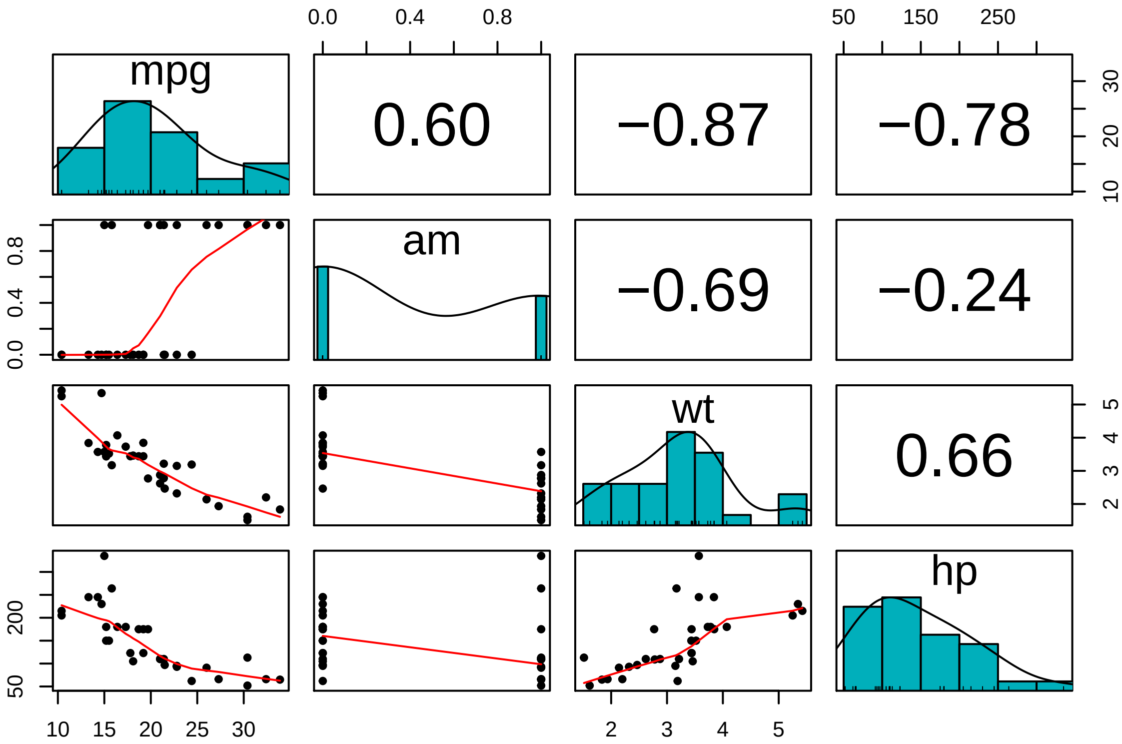
A.4 ANOVA

```
anova(fit1, fit2, fit3, fit4, fit5)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + wt
## Model 3: mpg ~ am + wt + hp
## Model 4: mpg ~ am + wt + hp + disp
## Model 5: mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 720.90
## 2     29 278.32  1    442.58 63.0133 9.325e-08 ***
## 3     28 180.29  1     98.03 13.9571  0.001219 **
## 4     27 179.91  1      0.38  0.0546  0.817510
## 5     21 147.49  6     32.41  0.7692  0.602559
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

A.5 Correlation of the variables of the Model "fit3"

```
library(psych)
mtcars_vars <- mtcars[, c(1, 9, 6, 4)]
pairs.panels(mtcars_vars, method = "pearson", hist.col = "#00AFBB", ellipses = FALSE)
```



A.6 Summary of the Model "fit3"

```
##
## Call:
## lm(formula = mpg ~ am + wt + hp, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4221 -1.7924 -0.3788  1.2249  5.5317
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 34.002875   2.642659  12.867 2.82e-13 ***
## am           2.083710   1.376420   1.514 0.141268
## wt          -2.878575   0.904971  -3.181 0.003574 **
## hp          -0.037479   0.009605  -3.902 0.000546 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.538 on 28 degrees of freedom
## Multiple R-squared:  0.8399, Adjusted R-squared:  0.8227
## F-statistic: 48.96 on 3 and 28 DF,  p-value: 2.908e-11
```

A.7 Plot of the Residuals of the Model "fit3"

```r
par(mfrow = c(2,2))
plot(fit3)
```