Motor Trend: Manual transmission better than automatic for MPG

This report analyses the relationship between the car features and the fuel consumption expressed in MPG. By fitting various models, we conclude that the most significant features for the consumption are the transmission system, the cylinders, the horsepower and the weight of the car.

The cars with manual transmission system have better consumption efficiency (less mpg) than the ones with automatic. This is measured using the miles per gallon feature of our dataset, and is quantified as 1.8 mpg more for the cars with manual compared to the cars with automatic transmission system.

Getting and cleaning data

By checking the `mtcars` dataset, we identify that cyl, vs, am, gear and carb represent factors, so the datatype of these factors should be transformed as such in the data frame:

```
mtcars$cyl <- factor(mtcars$cyl); mtcars$vs <- factor(mtcars$vs); mtcars$am <-
 factor(mtcars$am); mtcars$gear <- factor(mtcars$gear); mtcars$carb <- factor(mtcars$carb)
```

Exploratory data analysis

Checking the relationship between the variables of mtcars using `pairs` in Figure 1. We observe that there is a good correlation between `mpg` (which represents the miles per galon) and each of the following variables: `disp`, `hp`, `wt`, `vs` and `am`.

To identify the relationship between `mpg` and the automatic/manual transmission system, we can plot it against the `am` variable only. In Figure 2 we observe that the manual transmission system (am=1) gives more miles per galon of fuel for a car. This answers the question of interest, whether automatic or manual transmission is better for MPG.

Fit multiple models and selection

To reach the conclusion of the exploratory data analysis, we build multiple regression models and select the best based on the adjusted R-squared value of each model. Initially we fit a linear model with all variables of `mtcars` in `fitall`. Then we select the best model using the `step()` function for the `fitall` model.

```
fitall<-lm(mpg ~ ., data = mtcars) ; fitbest<-step(fitall,direction="both")
```

```
summary(fitbest)$adj.r.squared
```

```
## [1] 0.8401
```

We observe that the `Adjusted R-squared` shows that the difference betwen automatic and manual transmission system is `0.84` or `84%`.

Checking how significantly different are the model using `am` variable only, against the best model (`cyl+hp+wt+am`), by running the anova() function on those two models. From the coefficient table shown

below, we see that the p-value almost 0, hense the hypothesis that `cyl`, `hq` and `wt` do not add value to the accuracy of the model is rejected.

```
fitam<-lm(mpg ~ am, data=mtcars); anova(fitam,fitbest)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ cyl + hp + wt + am
##   Res.Df RSS Df Sum of Sq    F  Pr(>F)
## 1     30 721
## 2     26 151  4       570 24.5 1.7e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual plot and diagnosis

By ploting the residuals (Figure 3), we see that some outliers are the Toyota Corolla, Fiat 128, Crysler Imperial and Toyota Corona. We also see that the distrubution (Normal Q-Q) of the residuals is normal.

To validate the observation, we're running the diagnostics with `hatvalues()` and `dfbetas()` and observe that indeed those cars have quite larger values than the rest of the cars:

```
tail(sort(round(hatvalues(fitbest),3)),3)
```

```
##       Toyota Corona Lincoln Continental       Maserati Bora
##               0.278               0.294               0.471
```

```
tail(sort(round(dfbetas(fitbest)[,6],3)),3)
```

```
## Chrysler Imperial          Fiat 128     Toyota Corona
##             0.351             0.429             0.731
```

Inference and Conclusion

Based on the fit model with the `am` variable only, and the observation of the normal distribution on the residual plots above, we're performing a `t.test` to see the significant difference between automatic and manual transmission systems (p-value < 0.05)

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

```
## [1] 0.001374
```

This also answers the question about the quantification of how different is the MPG between automatic and manual transmissions.

By reading the summary of the `fitbest` model, we conclude to the following:

- Increasing the cylinders of a car to 6 and 8 decreases the miles per galon by 3 and 2.2 respectively.
- Changing the horsepower of a car does not impact the fuel consumption (slight change).
- The heavier the car is, the more fuel it consumes (2.5 mpg, per 1000lb less).
- Cars with manual transmission system consume more fuel than the ones with automatic. Automatic cars have 1.8 more mpg compared to manual transmission.

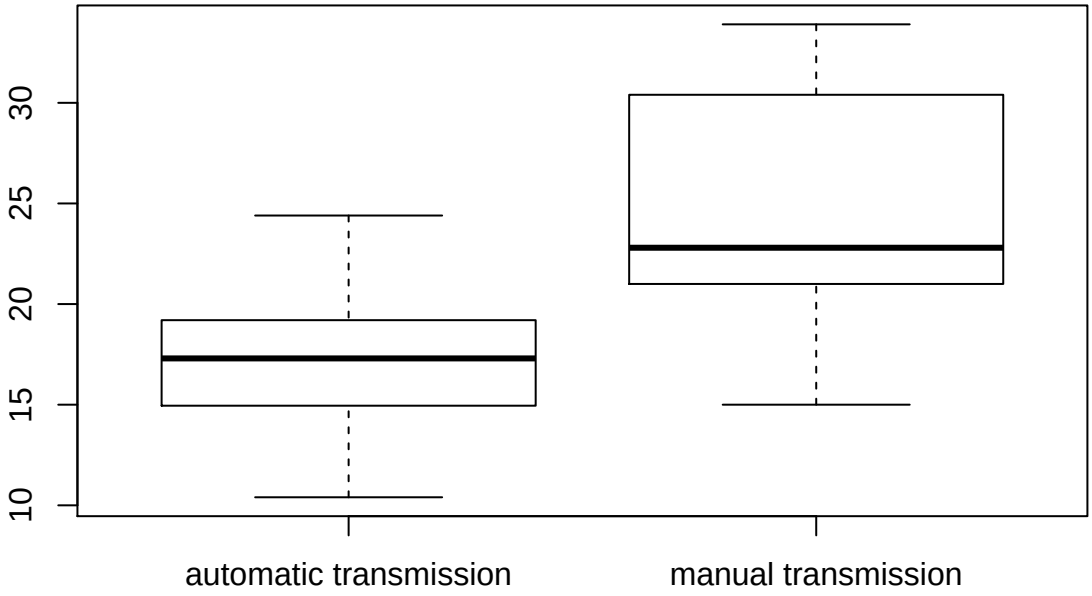
Appendix

```
pairs(mtcars,panel=panel.smooth,main="Figure 1. Relationship between mtcars variables")
```

Figure 1. Relationship between mtcars variables


```
boxplot(mpg~am,data=mtcars,main=("Figure 2. Miles per galon
vs. transmission system"),names=c("automatic transmission","manual transmission"))
```

Figure 2. Miles per galon vs. transmission system




```
par(mfrow = c(2, 2))
plot(fitbest)
mtext("Figure 3. Residual plot",side=3,line=-1,outer=TRUE)
```

Figure 3. Residual plot