Automatic and manual transmissions: comparative study

Elena Fedorova

25 August 2018

## Executive Summary

The report explores relationship between transmission type (manual or automatic) and miles per gallon (MPG). The analysis is based on the mtcars dataset. The following questions were addressed in the report: define which type of transmission is better for MPG, and quantify the difference in MPG. The simple linear regression and the multiple regression models with hypothesis testing will be used in the analysis. Both models ultimately confirmed that the cars in this study with manual transmissions had on average significantly higher MPG’s than the cars with automatic transmissions. Data visualisation is presented in the Appendix section.

## Loading, processing and exloring the data

data(mtcars)  
head(mtcars, n = 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

summary(mtcars)

## mpg cyl disp hp   
## Min. :10.40 Min. :4.000 Min. : 71.1 Min. : 52.0   
## 1st Qu.:15.43 1st Qu.:4.000 1st Qu.:120.8 1st Qu.: 96.5   
## Median :19.20 Median :6.000 Median :196.3 Median :123.0   
## Mean :20.09 Mean :6.188 Mean :230.7 Mean :146.7   
## 3rd Qu.:22.80 3rd Qu.:8.000 3rd Qu.:326.0 3rd Qu.:180.0   
## Max. :33.90 Max. :8.000 Max. :472.0 Max. :335.0   
## drat wt qsec vs   
## Min. :2.760 Min. :1.513 Min. :14.50 Min. :0.0000   
## 1st Qu.:3.080 1st Qu.:2.581 1st Qu.:16.89 1st Qu.:0.0000   
## Median :3.695 Median :3.325 Median :17.71 Median :0.0000   
## Mean :3.597 Mean :3.217 Mean :17.85 Mean :0.4375   
## 3rd Qu.:3.920 3rd Qu.:3.610 3rd Qu.:18.90 3rd Qu.:1.0000   
## Max. :4.930 Max. :5.424 Max. :22.90 Max. :1.0000   
## am gear carb   
## Min. :0.0000 Min. :3.000 Min. :1.000   
## 1st Qu.:0.0000 1st Qu.:3.000 1st Qu.:2.000   
## Median :0.0000 Median :4.000 Median :2.000   
## Mean :0.4062 Mean :3.688 Mean :2.812   
## 3rd Qu.:1.0000 3rd Qu.:4.000 3rd Qu.:4.000   
## Max. :1.0000 Max. :5.000 Max. :8.000

## Analysis

# Simple linear regression

ModelFit <- lm(mpg ~ am, data = mtcars)  
summary(ModelFit)

##   
## Call:  
## lm(formula = mpg ~ am, data = mtcars)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -9.3923 -3.0923 -0.2974 3.2439 9.5077   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 17.147 1.125 15.247 1.13e-15 \*\*\*  
## am 7.245 1.764 4.106 0.000285 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.902 on 30 degrees of freedom  
## Multiple R-squared: 0.3598, Adjusted R-squared: 0.3385   
## F-statistic: 16.86 on 1 and 30 DF, p-value: 0.000285

summary(ModelFit)$coeff

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

The Beta0/intercept coefficient is mean MPG for cars with automatic transmissions; the Beta1/am coefficient is the mean increase in MPG for cars with manual transmissions (am = 1). (Beta0 + beta1) is the mean MPG for cars with manual transmissions. So, the mean difference in MPG is 7.244939.

Thus, the 95% confidence interval for beta1 (mean MPG difference) is following:

alpha <- 0.05  
n <- length(mtcars$mpg)  
pe <- coef(summary(ModelFit))["am", "Estimate"]  
se <- coef(summary(ModelFit))["am", "Std. Error"]  
t <- qt(1 - alpha/2, n - 2)  
pe + c(-1, 1) \* (se \* t)

## [1] 3.64151 10.84837

Based on the results, we can reject the null hyposthesis in favor of the alternative one: that there is a significant difference in MPG between the two groups at alpha = 0.05.

# Multiple regression

The following predictor variables will be included into analysis: wt (weight), qsec (1/4 mile time) and am (transmission type).The following step-by-step approach will be used in the modelling: 1) Start with the predictor whose correlation with mpg is highest (wt) 2) The variables that are highly correlated with wt are to be removed 3) Add the remaining predictor, qsec 4) Finally add am, to see if it is a significant predictor.

MultiFit <- lm(mpg ~ wt + qsec + am, data=mtcars)  
summary(MultiFit)

##   
## 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 \*\*\*  
## am 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

coef(summary(MultiFit))

## 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  
## am 2.935837 1.4109045 2.080819 4.671551e-02

So, the mean difference in MPG is 2.935837. Thus, the 95% confidence interval for beta1 (mean MPG difference) is following:

alpha <- 0.05  
n <- length(mtcars$mpg)  
pe <- coef(summary(MultiFit))["am", "Estimate"]  
se <- coef(summary(MultiFit))["am", "Std. Error"]  
t <- qt(1 - alpha/2, n - 2)  
pe + c(-1, 1) \* (se \* t)

## [1] 0.05438576 5.81728862

Based on the results, we can also reject null hypothesis in favor of the alternative one: that there is a significant difference in MPG between the two groups at alpha = 0.05.

## Conclusion

The analysis performed confirmed that there is difference in MPG associated with transmission type.In the simple model, the mean MPG difference is 7.25 MPG, while the multiple regression model delivers the difference of 2.93 MPG.

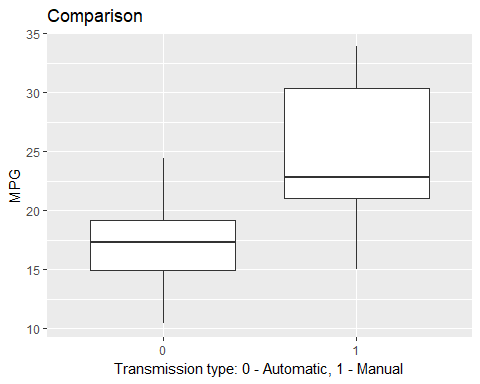
## Appendix

This section contains basic exploratory data analysis and all the required visualisations supporting the final conclusion.

# 1. Exploratory comparison of Automatic and Manual transmission MPG

The presented boxplots based on the observations of our data set demonstrate that on average the cars with manual transmission generally have higher MPG.

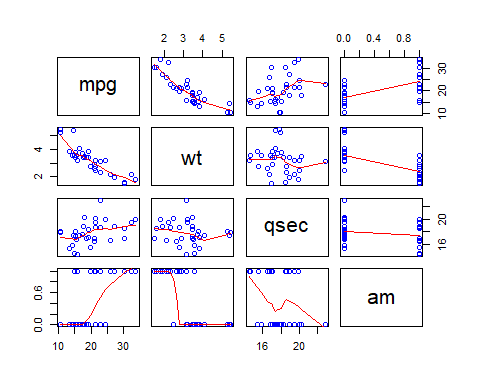
library(ggplot2)  
ggplot(data = mtcars, aes(x = as.factor(mtcars$am), y = mtcars$mpg)) + geom\_boxplot() + labs(x = "Transmission type: 0 - Automatic, 1 - Manual", y = "MPG") + ggtitle("Comparison")



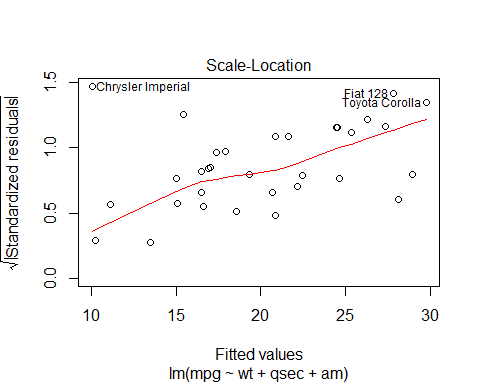
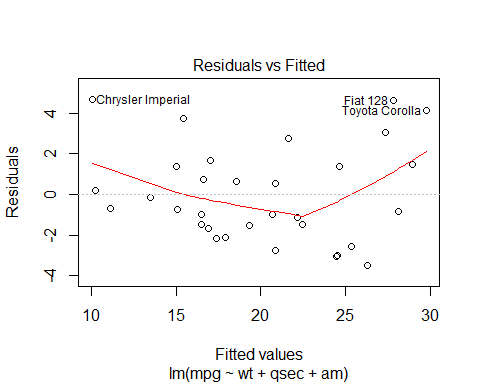
# 2. Scatterplots

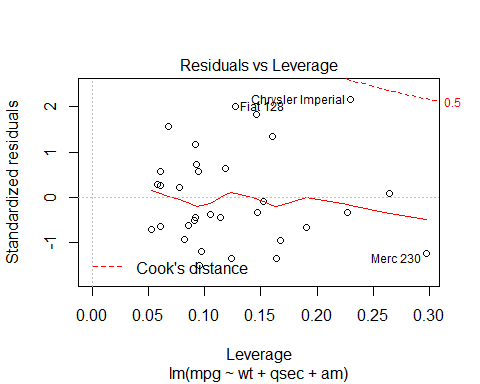
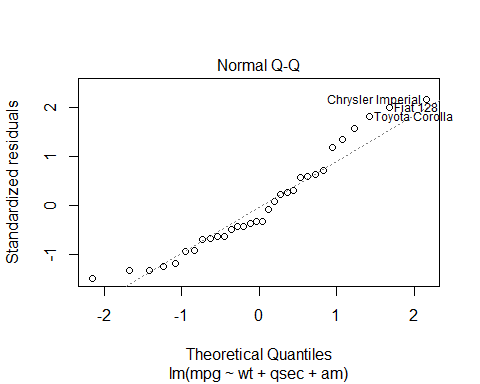
The presented scatterplots visually demonstrate correlations: moderate association can be noticed

mtcarsv <- mtcars[, c(1, 6, 7, 9)]  
pairs(mtcarsv, panel = panel.smooth, col = "blue")



# 3. Residual diagnostics

plot(MultiFit)



The following plots lead us to the following conclusion that the residuals and fitted values are independent. The points of the Normal Q-Q plot following closely to the line conclude that the distribution of residuals is normal.The Scale-Location plot random distribution confirms the constant variance assumption. As all the points are within the 0.05, the Residuals vs. Leverage plot demostrates that there are no outliers.