Transmission type and fuel efficiency

Joris Schut

Friday, March 13, 2015

Executive summary

Using simple and multivariate regression, the effect of transmission type on fuel efficiency was examined. A positive relation between a manual transmission and fuel efficiency was found. A simple linear model including only transmission type only took 35% of the variation of the data into account. Including weight in lb / 1000 (wt) and 1/4 mile time (qsec) in the model along the transmission type increased quality of the model so 85% of the variance could be explained. In this model, cars with a manual transmission had on average a better fuel efficiency of 2.936 miles per gallon than those with an automatic transmission.

Introduction

In cars, the type of transmission (automatic or manual) used might affect the fuel efficiency (measured in miles per gallon). This study aims to identify if such difference exist and how big such difference is. It does so by looking at the mtcars data set present in R. Using simple and multivariate linear regression the difference will be determined.

Research questions

The research questions for this study are as follows:

- Is an automatic or manual transmission better for MPG?
- What is the MPG difference between automatic and manual transmissions?

Exploratory analisis

To see if a difference exist in fuel efficiency between the different transmission types, exploratory data analysis was carried out. Plotting the fuel efficiency by transmission type resulted in the belief fuel efficiency was affected by the type of transmission (see figure 1). To confirm this, a two sided t-test was carried out with H0: Xbar(automatic) equals Xbar(manual) and Ha: Xbar(automatic) does not equals Xbar(manual). Here, H0 was rejected with 95% certainty (p=0.001374).

A single linear regression model

After establishing there is a difference between transmission type and fuel efficiency, a simple linear model was generated. This resulted in a model with the folowwing parameters:

Intercept	Transmission type
17.147	7.245

R^2: 0.3598

Although this model explains some of the variation in fuel efficiency, it is of limited predictive value. With a R^2 value of 0.359, only 36% of the variation in fuel efficient can be explained by the type of transmission.

A multivariate linear regression model

To create a better predictive model additional factors from the data set are added to the model. Factors to be included were found by using a step algorithm to determine the key factors that predict fuel efficiency.

Coefficient	Estimate	Std. Error	t-value	Pr(>
Intercept	9.6178	6.9596	1.382	0.177915
Weight (in lb)/1000	-3.9165	0.7112	-5.507	6.95e-06
1/4 mile time	1.2259	0.2887	4.247	0.000216
Transmission type	2.9358	1.4109	2.081	0.046716

R^2: 0.8497

In the constructed model weight in lb / 1000 (wt), 1/4 mile time (qsec) and transmission type (am) were found to be explaining 85% of the variation of the variation in fuel efficiency. These factors were used in a multivariate regression model. This resulted in the following parameters in the linear model.

(Intercept)	Transmission type	Weight (in lb)/1000	1/4 mile time
9.618	2.936	-3.917	1.226

Comparing the simple and multivariate model

To compare the outcomes of the two models, an analysis of variance was carried out with H0: both models yield similar results and Ha: both models do not yield similar results. Here, H0 was rejected with 95% certainty (p=1.55e-09).

Analysis of Variance Table

- Model 1: mpg ~ am
- Model 2: mpg ~ am + wt + qsec

#	Res.	Df	RSS	Df	Sum of Sq	F	Pr(>F)
1		30	720.90				
2	28	169.29	2	551.61	45.618		1.55e-09

Before reporting on the results, the normality of the data is checked for. The quartile-quartile (Q-Q) plot (figure 2) shows the data is approximately normal.

Conclusion and discussion

Using simple and multivariate regression, the effect of transmission type on fuel efficiency was examined. A positive relation between a manual transmission and fuel efficiency was found. When only looking at transmission types, cars with a manual transmission had on average a better fuel efficiency of 7.245 miles per gallon. However, this model only took 35% of the variation of the data into account. Including weight in lb / 1000 (wt) and 1/4 mile time (qsec) in the model along the transmission type increased quality of the model so 85% of the variance could be explained. In this model, cars with a manual transmission had on average a better fuel efficiency of 2.936 miles per gallon compared to an automatic transmission. The data did not show signs of no-normality.

Annexes

Histogram

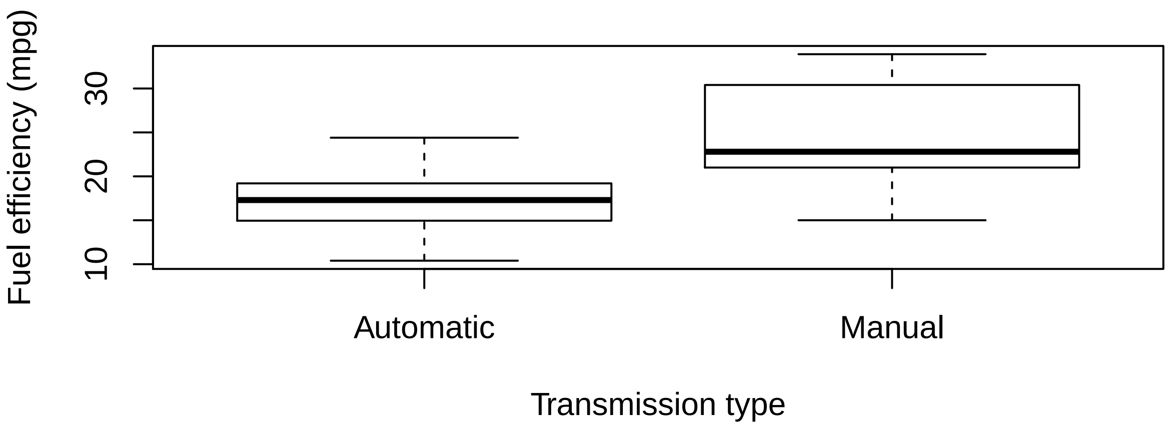


Figure 1: fuel efficiency by transmission type

Quartile-quartile plot

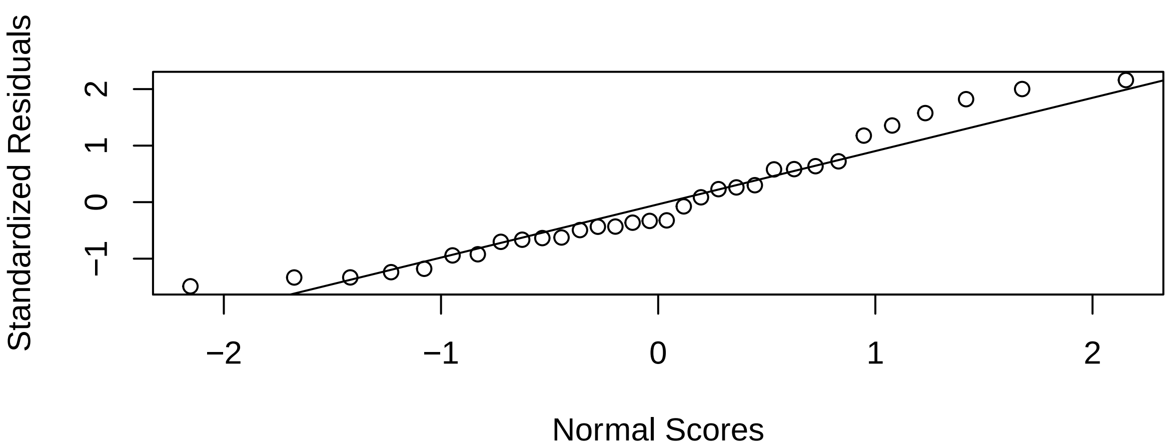
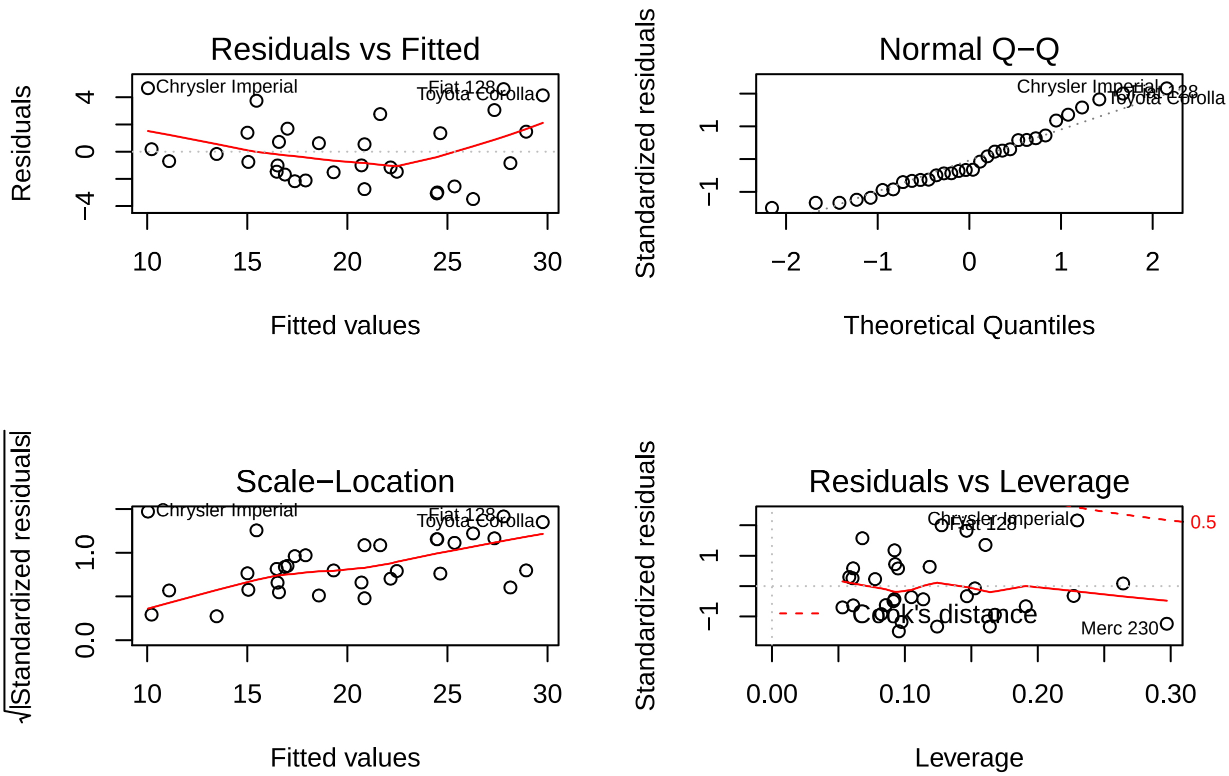


Figure 2: Q−Q plot of the residuals

Residual plot



Code used

```r
#Loads the data
data(mtcars)

#Sets the transmission as factors and names them
mtcars$am <- as.factor(mtcars$am)
levels(mtcars$am) <- c("Automatic", "Manual")

#Plots a boxplot
boxplot(mpg~am, data=mtcars, sub="Figure 1: fuel efficiency by transmission type",
        xlab="Transmission type", ylab="Fuel efficiency (mpg)")

#T-test of mpg by transmission type
automatic <- mtcars[mtcars$am == "Automatic",]
manual <- mtcars[mtcars$am == "Manual",]
t.test(automatic$mpg, manual$mpg)

#Simple linear regression
simplelm <- lm(mpg~am, data=mtcars)
print(simplelm)
summary(simplelm)

#Multivatiate regression
#Step algorithm
data(mtcars)
stepmodel = step(lm(data = mtcars, mpg ~ .),trace=0,steps=10000)
summary(stepmodel)

#Best approximation
bestfit <- lm(mpg~am + wt + qsec, data = mtcars)
print(bestfit)

#Comparing the 2 models
anova(simplelm, bestfit)

#checking for signs of non-normality
lm.stdres = rstandard(bestfit)

qqnorm(lm.stdres, main="", sub="Figure 2: Q-Q plot of the residuals",
       ylab="Standardized Residuals", xlab="Normal Scores")
qqline(lm.stdres)

#Residual plots
par(mfrow = c(2,2))
plot(bestfit)
```