Miles per gallon in different transmission types

Radomir Nowacki

December 25, 2015

Executive summary of effect of transmission type on miles per gallon of road vehicles

This study will analyze `mtcars` dataset in an attempt to provide information of automatic and manual transmission types on petrol efficiency measured in miles per gallon. The findings of this study are following:

- According to simple t-test analysis, manual transmission is more efficient by 7.24 miles per galon
- Improved multivariate model lowers this prediction to 2.45 miles per galon due to inclusion of confounding variables

In both cases the manual transmission is better for the miles per galon.

Exploratory data analysis

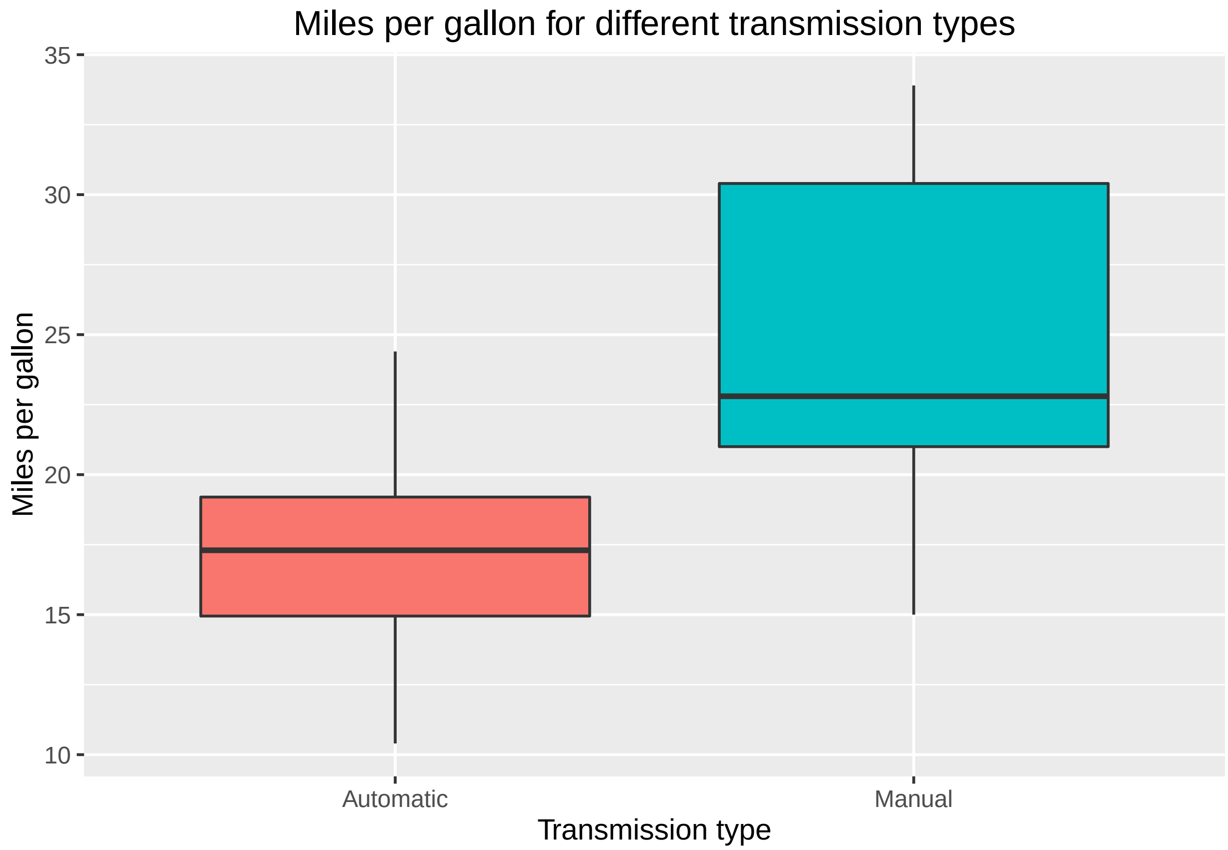
A boxplot of both transmission types will be produced to visually inspect if there is a difference between the two transmission types.

```
library(ggplot2)
library(knitr)
library(broom)
library(dplyr)

data("mtcars")

flist <- c('am','gear','carb','vs')
mtcars[flist] <- lapply(mtcars[flist],as.factor)

ggplot(data=mtcars,aes(y=mpg,x=am,fill=am))+
      geom_boxplot(aes(fill=am))+
      xlab('Transmission type')+
      ylab('Miles per gallon')+
      labs(title='Miles per gallon for different transmission types')+
      scale_fill_discrete(guide=FALSE)+
      scale_x_discrete(labels=c('Automatic','Manual'))
```




```
options(scipen = 999)
```

Above graph strongly hints to a better fuel efficiency of manual transmission. T-test will be performed to check if this is the case.

```
mpgtest <- t.test(mpg~am,data=mtcars)
mpgtest %>% tidy %>% select (estimate1,estimate2,statistic,p.value,conf.low,conf.high) %>%
 kable(digits=4, col.names=c('Automatic','Manual','t-statistic','p','CI low','CI high'),
            caption='t-test of mpg between manual and automatic transmission')
```

Table 1: t-test of mpg between manual and automatic transmission

Automatic	Manual	t-statistic	p	CI low	CI high
17.1474	24.3923	-3.7671	0.0014	-11.2802	-3.2097

P-value of 0.0014 lets us reject the null hypothesis of no difference between transmission types and allows for the claim of difference in the mean MPG between manual and automatic transmission.

Regression analysis

Firstly a simple linear regression of miles per gallon by transmission type is performed

```
model_1 <- lm(mpg~am,data=mtcars)
model_1_ar2 <- summary(model_1)$adj.r.squared
model_1 %>% tidy %>% mutate(term=c('Intercept','Manual')) %>%
 kable(digits=4,caption='Linear model between mpg and transmission')
```

Table 2: Linear model between mpg and transmission

term	estimate	std.error	statistic	p.value
Intercept	17.1474	1.1246	15.2475	0.0000
Manual	7.2449	1.7644	4.1061	0.0003

This simple model is essentialy the same as the one in t-test. However with adjusted R^2 of only 0.3385 we explain only 33.85% of variability in mpg by transmission type.

To improve on this model a stepwise regression is run to find a better model, 1/4 mile times and displacement variables are removed as the first one is nto a raw predictive variable and second one isi highly correlated with weight

```
mtcarslight <- mtcars %>% select(-disp,-qsec)
model_2 <- step(lm(mpg~.,data=mtcarslight),direction='both',trace=0)
model_2_ar2 <- summary(model_2)$adj.r.squared
model_2 %>% tidy %>% mutate(term=c('Intercept','Horsepower','Weight','Engine type','Manual')) %>%
 kable(digits=4,caption='Multivariate linear model')
```

Table 3: Multivariate linear model

term	estimate	std.error	statistic	p.value
Intercept	31.0788	3.3928	9.1603	0.0000
Horsepower	-0.0301	0.0109	-2.7508	0.0105
Weight	-2.5910	0.9174	-2.8243	0.0088
Engine type	1.7855	1.3271	1.3454	0.1897
Manual	2.4171	1.3794	1.7523	0.0911

```
anova(model_1,model_2) %>% tidy %>% cbind(model=c('mpg~am','mpg~am+hp+wt+vs'),.) %>% kable(digits=2,cap
```

Table 4: Analysis of variance between models

model	res.df	rss	df	sumsq	statistic	p.value
mpg~am	30	720.90	NA	NA	NA	NA
mpg~am+hp+wt+vs	27	168.96	3	551.93	29.4	0

Second model is siginificantly different from the simple one. It improves adjusted R^2 to 0.8277, increasing explanatory power of the model to 82.77% of variability in mpg. According to this improved model manual transmission type results in 2.42 increase in miles per gallon of the car.

Visual inspection of residual plots in the appendix provides no signs of violation of heteroskedasticity.

It is to be noted that there is significant corelation between horsepower and weight, as well as engine type and also transmission and weight. Those correlations may invalidate the findings of this study, but the author lacks the skills to further analyze the data.

Appendix

```
par(mfrow=c(2,2))
plot(model_2)
```