Impact of the type of type of transmission type on MPG

An analysis

Olivia Urbanowski

6/8/2020

Executive summary

Based on data provided by R, we conducted a comparative analysis of the fuel economy (mpg) for automatic and manual cars, for cars on the market in 1973-1974.
Considering the limited amount of data available, we decided to infer the fuel economy by using the best possible linear model including transmission type.
It appears that manual transmissions are better for fuel economy, ie. mpg is higher for manual cars.

Exploratory data analysis

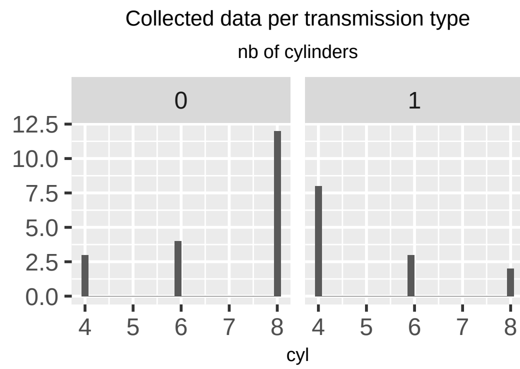
The data used to explore the question of the impact of the type of transmission on the MPG is available in R - `mtcars`.

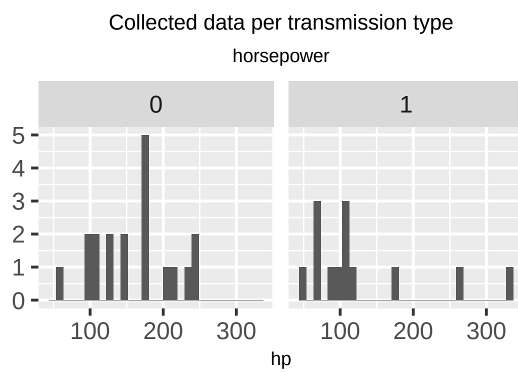
Table 1: Extract of mtcars

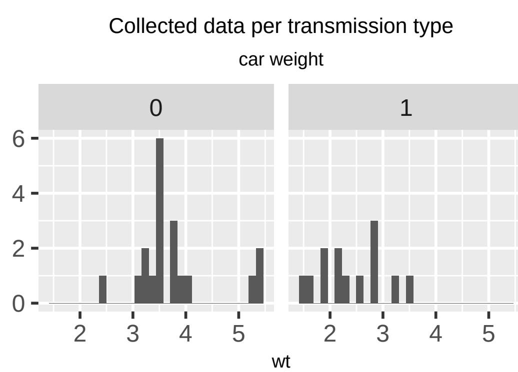
	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2

This dataset collects for 32 car models the following information:
mpg - Miles/(US) gallon ; cyl - Number of cylinders ; disp - Displacement (cu.in.) ; hp - Gross horsepower ; drat - Rear axle ratio ; wt - Weight (1000 lbs) ; qsec - 1/4 mile time ; vs - Engine (0 = V-shaped, 1 = straight) ; am - Transmission (0 = automatic, 1 = manual) ; gear - Number of forward gears ; carb - Number of carburetors

Focusing on the type of transmission - `am = 0` for automatic and `am = 1` for manual - it turns out it is difficult to compare directly the data we have, since the cars the data was collected from present different characteristics:







Nonetheless, we will try to fit a linear model to predict the fuel economy, with the type of transmission as a predictor (amongst others).

1. Model selection

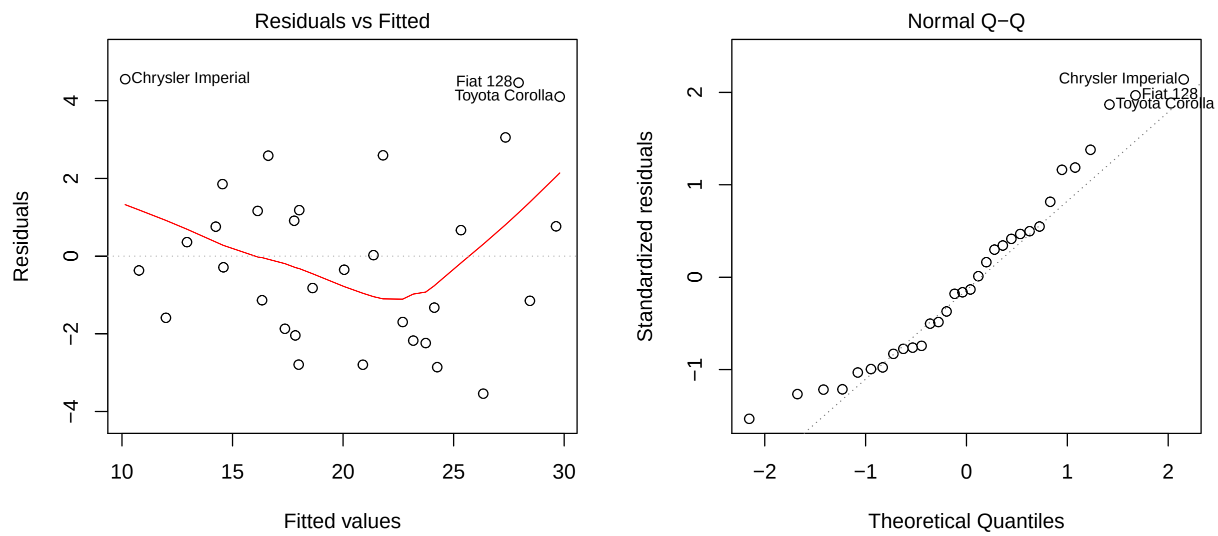
To analyze the impact of the type of transmission on fuel economy (mpg), we chose to select the best possible model - including the type of transmission `am` as a predictor - using adjusted R^2.
The code for this part can be checked in the appendix.
A comparison of all possible models - thanks to the olsrr package - reveals the best predictors are:

```
## [1] "disp hp wt qsec am"
```

All following analysis will thus be performed keeping only these as predictors.

2. Validity of the linear regression model

Plotting the residuals vs. the fitted model seems to confirm linearity of the model, and no heteroskedasticity (i.e. contant variance) of the residuals:



Performing a normality test on the residuals also returns a rather conslusive results: a p-value of 0.186, which means we fail to reject the hypothesis of a normal distribution for the residuals.
Our linear model `mpg ~ disp + hp + wt + qsec + am` is thus deemed as valid.

3. Impact of the type of transmission on the fuel economy

Running the selected linear regression model and comparing the coefficients reveals that for all other predictors kept constant, cars with a manual transmission are predicted to have an increased fuel economy of 3.47 mpg.

```
##              Estimate Std. Error  t value   Pr(>|t|)
## (Intercept) 14.361904   9.740795 1.474408 0.15237837
## am1          3.470453   1.485780 2.335779 0.02748781
```

The 95% confidence interval associated with this inscrease is:

```
##     2.5 %    97.5 %
## 0.4163887 6.5245181
```

Conclusion

From the collected data, we infered a linear model which, based on the performed analysis, selects displacement, gross horsepower, weight, 1/4 mile time and transmission as the best predictors for fuel economy.
Assuming our sample is representative of the cars on the market, the analysis we performed reveal that **a manual transmission is better for mpg. All other car characteristics being equal, a manual car would travel between 0.42 and 6.52 more miles per US gallon** (confidence interval : 95%).

Code

```r
knitr::opts_chunk$set(
  echo = FALSE,          # don't show code
  warning = FALSE,       # don't show warnings
  message = FALSE,       # don't show messages (less serious warnings)
  cache = TRUE,
  fig.align = "center",  # center figures
  fig.height=4,          # set default figure size
  fig.width=6
)
library(dplyr)           # load my libraries
library(ggplot2)
library(olsrr)           # useful library for selecting the best model
mydata <- mtcars
## converting 2 columns into factor variables
cols <- c("vs", "am")
mydata[,cols] <- lapply(mydata[,cols], as.factor)

knitr::kable(mydata[1:5,],  caption = "Extract of mtcars")
p <- ggplot(data = mydata)
p + geom_histogram(aes(x=cyl)) + facet_grid(.~am) + labs(x=
    "cyl", y="", title =
    "Collected data per transmission type", subtitle =
      "nb of cylinders")+theme(plot.title=
     element_text(hjust = 0.5, size = 8), plot.subtitle=
     element_text(hjust = 0.5, size = 7), axis.title.x = element_text(size = 7))
p + geom_histogram(aes(x=hp)) + facet_grid(.~am) + labs(x="hp", title =
  "Collected data per transmission type", subtitle ="horsepower", y=
    "") +theme(plot.title= element_text(hjust = 0.5, size = 8), plot.subtitle=
     element_text(hjust = 0.5, size = 7), axis.title.x = element_text(size = 7))
p + geom_histogram(aes(x=wt)) + facet_grid(.~am) + labs(x="wt", y="", title=
  "Collected data per transmission type",subtitle="car weight")+theme(plot.title=
     element_text(hjust = 0.5, size = 8), plot.subtitle=
     element_text(hjust = 0.5, size = 7), axis.title.x = element_text(size = 7))
fit <- lm(mpg ~ .,mydata)
model <- ols_step_all_possible(fit)
## this handy function calculates the R squared
## and adjusted R squared for all possible models
model <- arrange(model, desc(adjr))
model[1,"predictors"]
fit1 <- lm(mpg ~ disp + hp + wt + qsec + am, mydata)
par(mfrow=c(1,2))
plot(fit1, which =1)
plot(fit1, which =2)
pval <- shapiro.test(fit1$residuals)[2] ## normality test
summary(fit1)$coef[c(1,6),]
confint(fit1)[6,]
```