Automobile designs on fuel consumption

LG

1 1 2022

Executive summary

In current times as fuel prices are rising, let's explore how different types of transmission effect fuel consumption. We collected data on 32 automobiles from brands like Porsche, Mercedes, Fiat, Mazda and many more, all designed in 1973/ 1974. To answer if and if yes, by how much, type of transmission, automatic or manual, saves more fuel.

Exploring data

First of all, let's explore what types of cars we have: 18 out of 32 models use automatic transmission. On average, our tested models consume 20.1 miles/ gallon. However, looking at Figure 1, models with automatic transmission consume on average far less fuel compared to those with manual transmission.

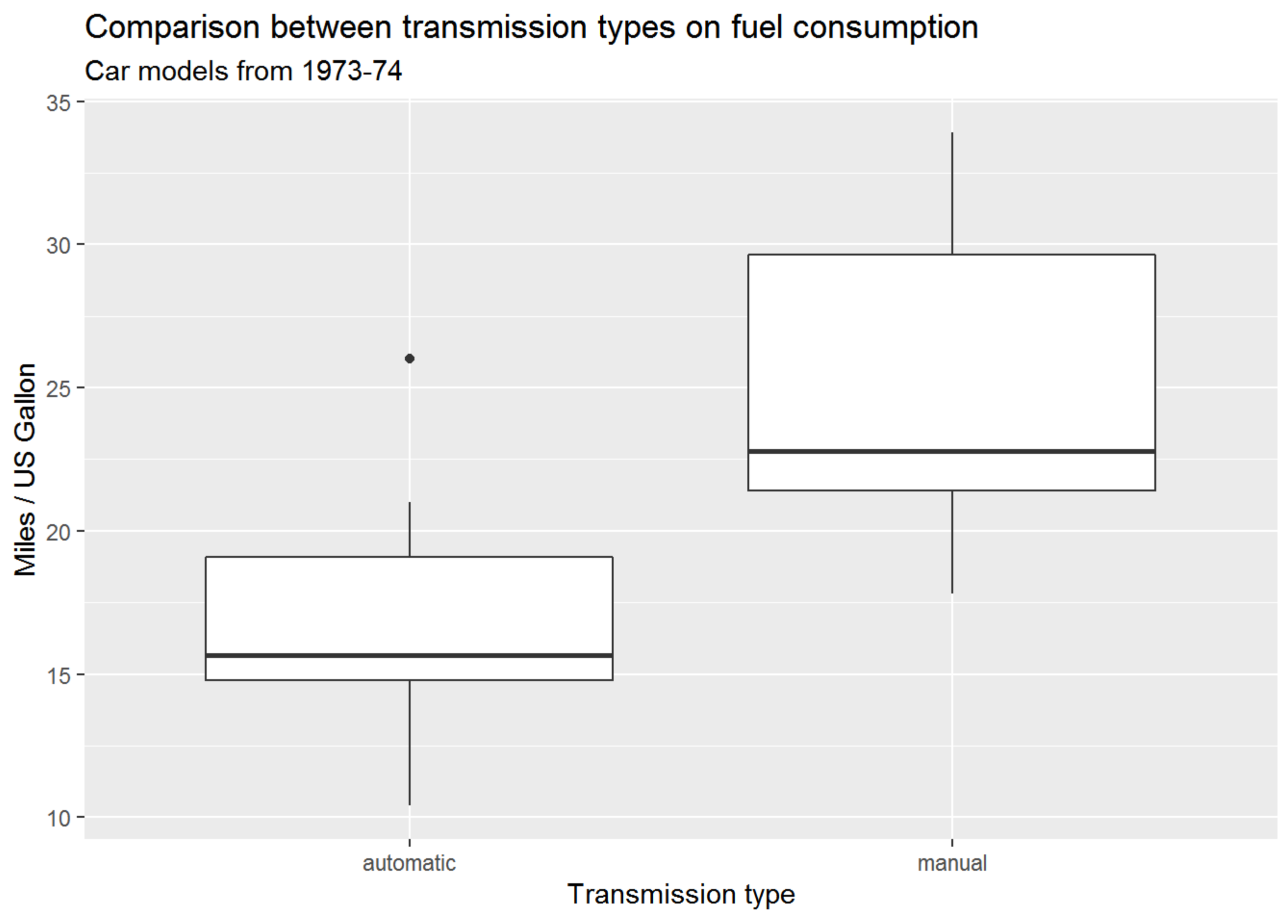


Figure 1: Differences in fuel consumption due to transmission type

Relationship: transmission on fuel consumption

To verify, if transmission type in fact effects fuel consumption, we drive a statistical analysis by creating a linear regression model. Precisely, the question of interest is: **can we detect a significant difference between automatic and manual transmission on fuel consumption measured in miles/ gallon, and if so, how much does it differ?**

First of all, the model of linear regression `lm(formula = mpg ~ Transmission - 1, data = mtcars)` confirms that either transmission type has a statistical significant effect on fuel consumption. Automatic transmission has a mean of 16.62 miles/ gallon and manual transmission has a mean of 24.56 miles/ gallon, both at p-value < 0.001. This results in a very strong statistical significant difference of 7.94 miles/ gallon between the two types (p-value < 0.001) thus automatic transmission turns out to be more efficient. However, the goodness of fit reveals that only 35% of variability are explained (R^2 = 0.44). Plotting the residuals shows a clear pattern assuming two groups (see figures in Appendix).

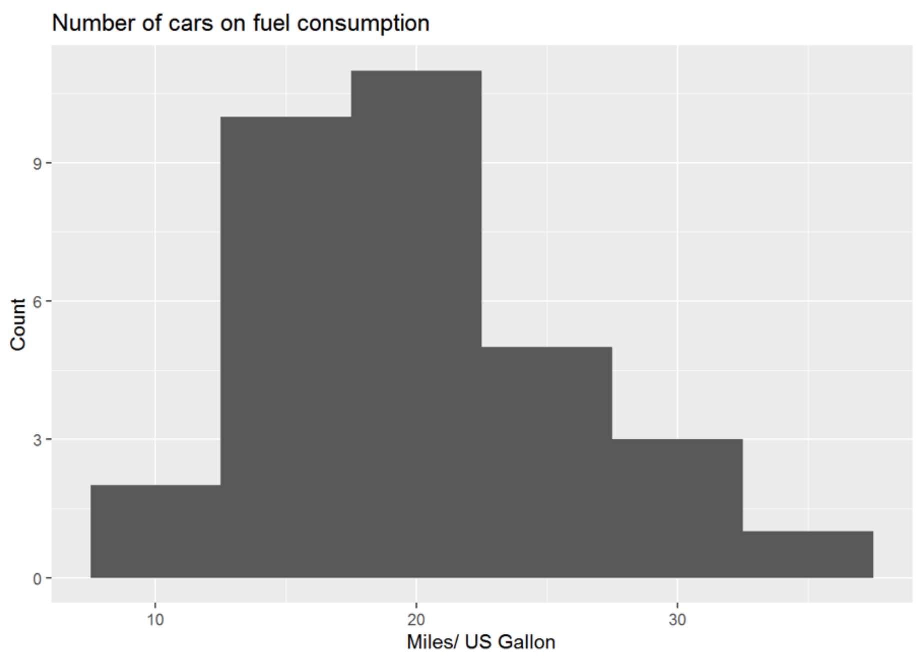
Therefore, we cannot explain fuel consumption solely on transmission type. Luckily, more variables have been collected on our tested models. Subsequently, we created a new linear model with all variables collected. Unsurprisingly, the model explained 99% of variation (R^2 = 0.99. However, no variable has been found as significantly contributing to explain fuel consumption (all p-values > 0.05). To avoid overfitting, we can exclude one after another variable from the model until the best model is evaluated. It turns out, that a model with gross horsepower, weight, transmission type and number of cylinders results in the best fitted model. The goodness of fit is very high (explained variability: R^2 = 0.989) along with a standard residual error of 2.48. (see residual plots in Appedix below)

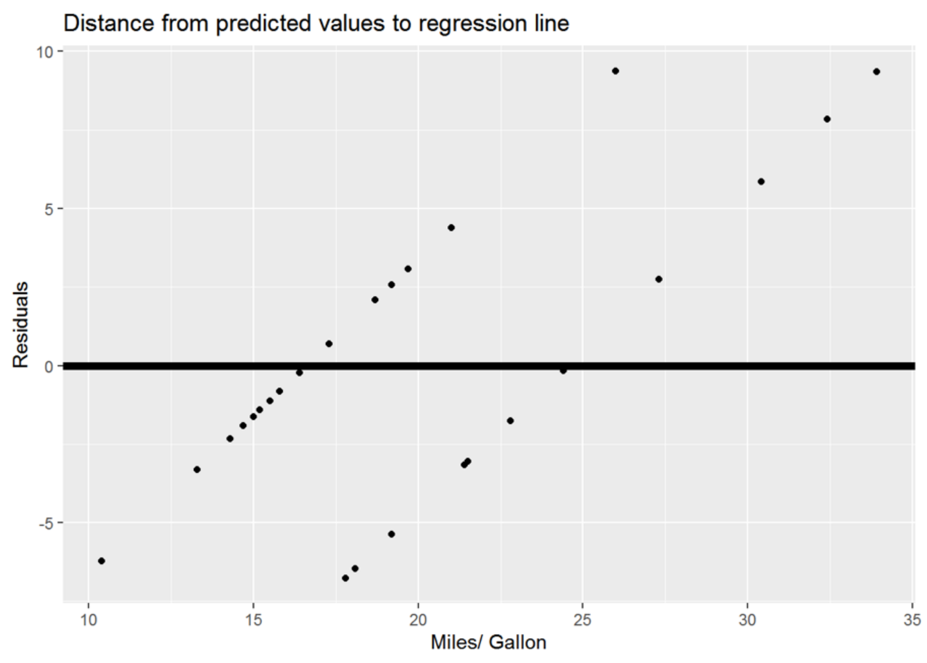
Conclusion

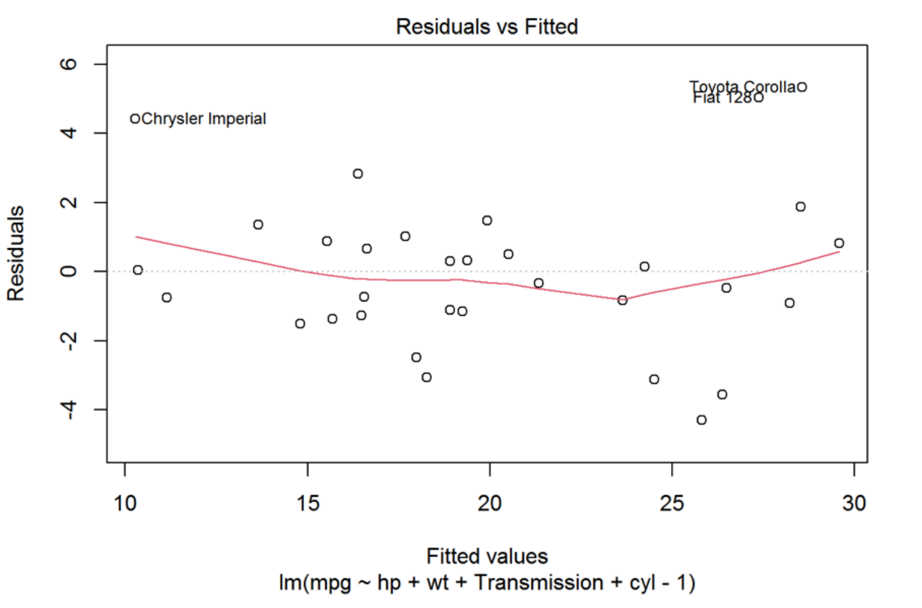
Explaining fuel consumption is depending on multiple automotive car designs. Mostly associated with significant effects are horse power, weight, transmission type and number of cylinders.

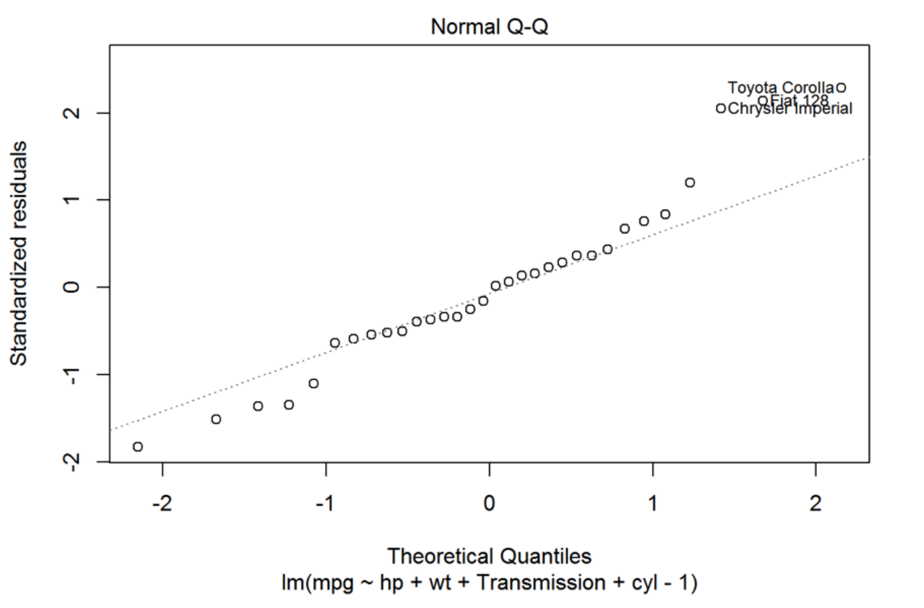
Appendix

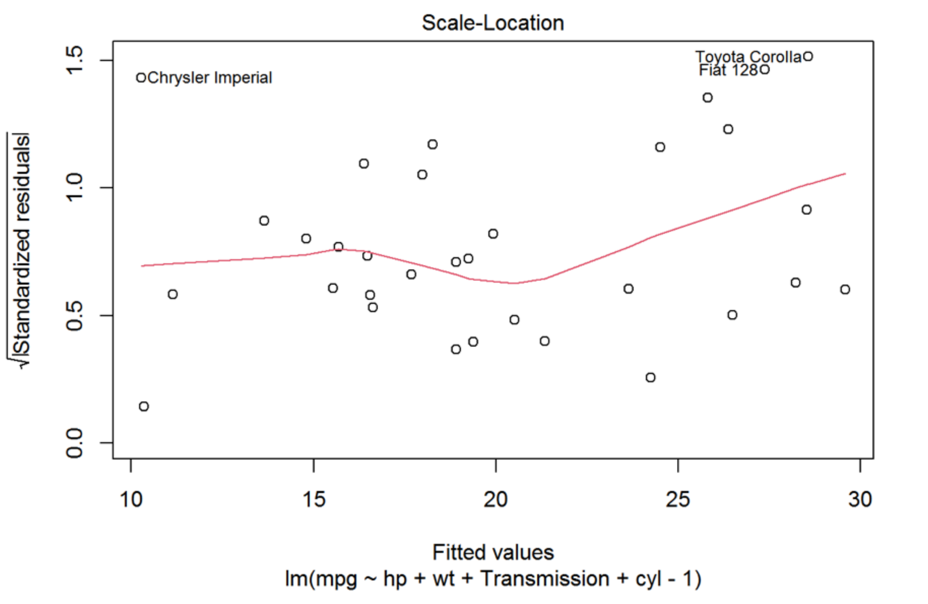
Figures:

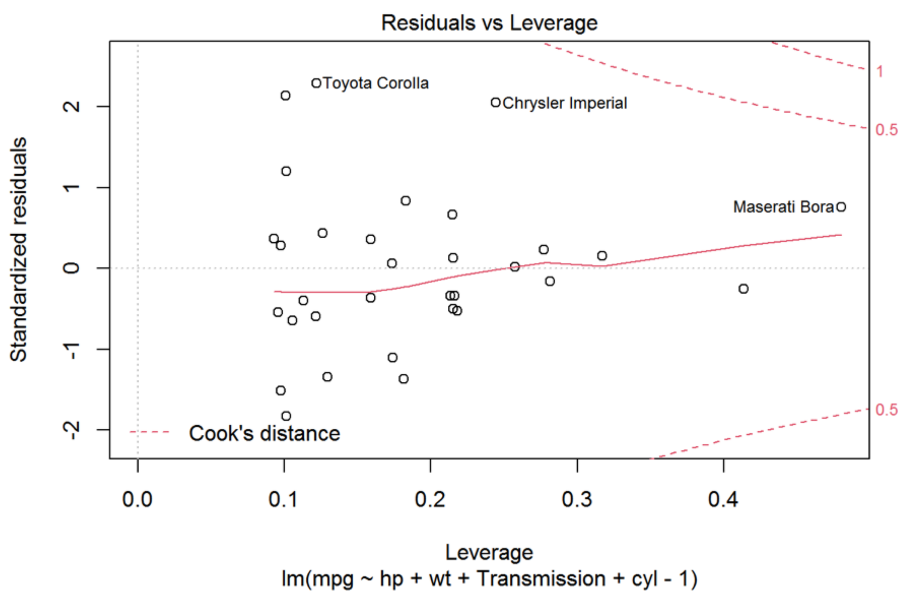













```
knitr::opts_chunk$set(echo = FALSE, cache = TRUE)
#loading data and packages
data("mtcars")
library(ggplot2)
library(dplyr)
library(knitr)

#cleaning data, variable vs represents transmission, where 0 = automatic, 1= manual
head(mtcars)
mtcars_o <- mtcars # saving original dataset
mtcars <- mtcars_o %>% mutate(am = factor(case_when(  # transform am to factor variable and
  vs == 0 ~ "automatic",
  vs == 1 ~ "manual"))) %>% rename(Transmission = am)
#steps to transform several columns to factor variables
mtcars1 <- mtcars %>%  select(-c(2,8,10:11))
mtcars2 <- mtcars %>% select(c(2,8,10:11)) %>% mutate(across(where(is.numeric), as.factor))
mtcars <- cbind(mtcars1,mtcars2)
n <- nrow(mtcars)
count_vs <- count(mtcars, vars=Transmission) #number of models per transmission type
overall_mpg_m <- mean(mtcars$mpg)
group_mpg_m <- mtcars %>% group_by(Transmission) %>% summarize(mean = mean(mpg))
ggplot(data = mtcars, aes(x=Transmission, y = mpg)) +
  geom_boxplot() +
  labs(x= "Transmission type",
       y="Miles / US Gallon",
       title = "Comparison between transmission types on fuel consumption",
       subtitle= "Car models from 1973-74")
model1 <- lm(mpg ~ Transmission -1, data= mtcars)   #checking whether types have any effect on outcome variable
model2 <- lm(mpg ~ Transmission, data= mtcars)  #checking the change in the mean between the two types
summary(model2)
sigma <- summary(model2)$sigma
residuals <- resid(model2)
all_model <- lm(mpg ~ .-1 , data = mtcars)
summary(all_model)
selected <- step(all_model, trace=0)
summary(selected)
# exploring distribution of mpg
ggplot(data= mtcars, aes(x=mpg))+
  geom_histogram(binwidth = 5)+
  labs(x= "Miles/ US Gallon",
       y= "Count",
       title = "Number of cars on fuel consumption")

#plotting residual of model: mpg as outcome, transmission as predictor
ggplot(data.frame(mpg = mtcars$mpg, residuals = residuals), aes(x= mpg, y = residuals))+
    geom_point()+
  geom_hline(yintercept= 0, size =2)+
  labs(x= "Miles/ Gallon",
       y= "Residuals",
       title = "Distance from predicted values to regression line")

# plotting residuals and diagnostics
plot(selected)
```