Fuel Economy Report

Mikerah Quintyne-Collins

Abstract

Introduction

Fuel economy of a vehicle is way to determine how much fuel a vehicle uses per unit distance, e.g., miles or kilometers. It is a measure that enables us to improve how much consumers spend on fuel and determine how it affects the environment.

Previous studies have shown that various design components of vehicles affect fuel economy such as design of the body of the car. This present study aims to assess the impact of tranmission types and wheel-drive-type by class of vehicle on fuel economy.

The report is presented by given a description of the data, the methods and tools used during the analysis, the results of the analysis and finally recommendations for future work.

Data

Overview

The dataset used for this study consists of a subset of the data provided by the EPA. It only consists of the car models that had a new release every year between 1999 and 2008. The dataset contains 38 car models, 11 variables and 234 observations. The dataset is available in the R package *ggplot2*.

Exploratory Data Analysis

In order to get a better understanding of the dataset provided and our research questions, we will analyze boxplots and tables. Figure 1 shows a side by side boxplot of fuel economy in miles per gallon split by type of transmission. Table 1 shows a frequency table for every combination of type of tranmission and the number of cylinders. Figure 2 shows an interaction plot between the wheel-drive-type and class of vehicle on fuel economy.

In summary, it seems like the transmission of the vehicle doesn't affect the fuel economy of the vehicle. The medians for most of the transmission types are between 20 and 30 mpg and 2 are less than 20 mpg. There are 2 manual and automatic transmissions with a outliers but the manual(m5) has the most. Moreover, by observing table 1, we see that the number of cylinders in a car may be a counfounding variable for the transmission type because usually the number of cylinders is dependent on what type of transmission a car has. Furthermore, if we look at the table below, we see that is it unbalanced. In other words, there is an unequal number of observations for every combination of the number of cylinders and transmission types.

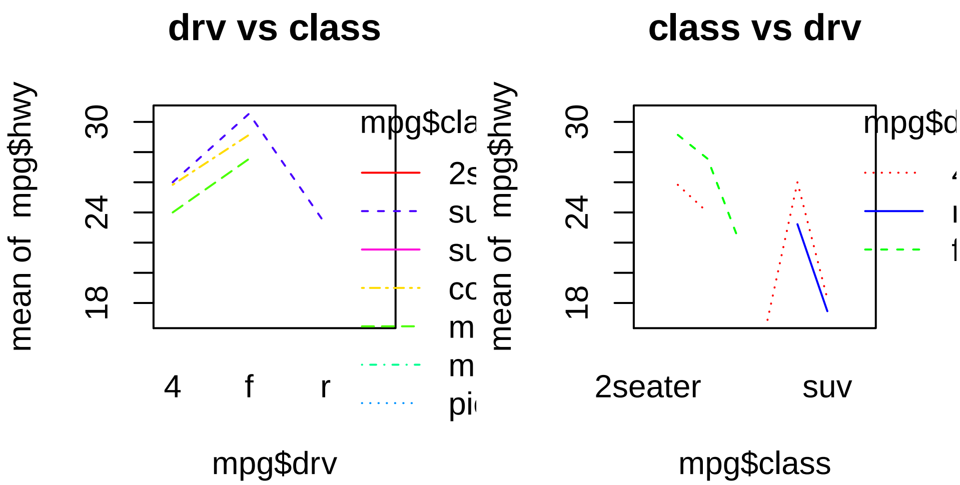

Figure 1: Figure 3: Interaction plots between wheel drive type and class of vehicle

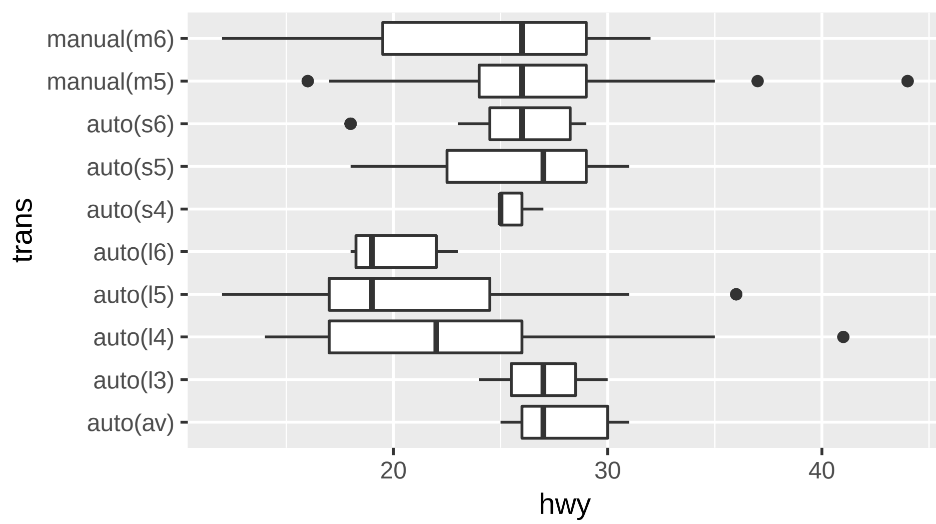

Figure 1: boxplot of fuel economy by type of transmission

Methods

Table 1: Table for the number of cylinders and transmission type

	4	5	6	8
auto(av)	2	0	3	0
auto(l3)	2	0	0	0
auto(l4)	24	0	29	30
auto(l5)	6	0	16	17
auto(l6)	0	0	2	4
auto(s4)	2	0	0	1
auto(s5)	1	0	1	1
auto(s6)	4	2	5	5
manual(m5)	33	2	18	5
manual(m6)	7	0	5	7

An exploratory data analysis was carried out for all the variables of interest in order to develop appropriate hypothesis with respect to our research questions.

Linear models where used to determine the impact of type of transmission and wheel-drive-type on the fuel economy. One-way ANOVA and ANCOVA was used to analyze whether the type of transmission in a vehicle impacts the fuel economy of the vehicle. Two-way ANOVA was use to study the impact of wheel-drive-type on fuel economy by the class of the vehicle.

The statistical programming language R using the functions *lm* and *aov* was used to fit the linear models and perform one-way ANOVA. The package *car* was used to perform ANCOVA and two-way ANOVA using the *anova* function. The package *ggplot2* was used for plotting. All statistical tests were performed at a significance level of $\alpha = 0.05$, unless otherwise stated.

Analysis

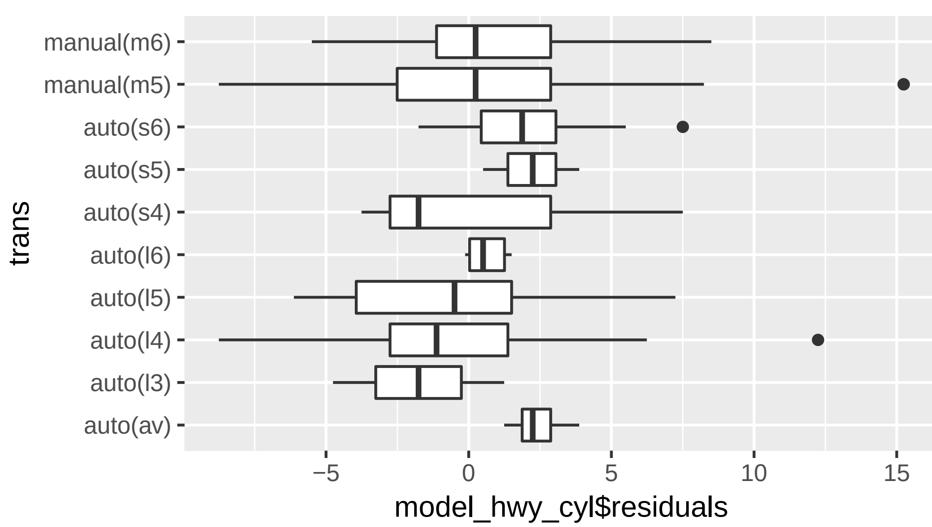

Figure 4: Boxplot of residuals split by transmission type

First, we use a linear model without taking into account the possible confounder to analyze the impact of transmission type on fuel economy. The model we used is

$Y = \beta_0 + \beta_{L3}X_{L3,i} + \beta_{L4,i}X_{L4,i} + \beta_{L5,i}X_{L5,i} + \beta_{L6,i}X_{L6,i} + \beta_{S4,i}X_{S4,i} + \beta_{S5,i}X_{S5,i} + \beta_{S6,i}X_{S6,i} + \beta_{M5,i}X_{M5,i} + \beta_{M6,i}X_{M6,i} + \epsilon_i \quad for \quad i = 1, ..., 236$ and $\epsilon_i \sim N(0, \sigma^2)$

Before analyzing the impact of transmission type on fuel economy, we further analyze the relationship between our confounder variable, the number of cylinders, and fuel economy. The model we used is $Y = \beta_{cyl}X_{cyl} + \epsilon \quad \epsilon \sim N(0, \sigma^2)$. The corresponding coefficient in this model was found to be significant and conclude that we should take this into account when analyzing the effects of transmission type on fuel economy. Figure 4 shows a boxplot of the residuals of this model split by transmission type.

Afterwards, we use a linear model taking into the possible confounder to analyze the impact of tranmission type on fuel economy. The model used is $Y_{ik} = \mu + \alpha_i + \beta_{cyl}X_{ik}^{(cyl)} + \epsilon_{ik(cyl)} \quad k = 1, .., 236$ where each i is for the different transmission types and $\epsilon_{ik(cyl)} \sim N(0, \sigma^2)$.

Results and Discussion

Table 2: Analysis of Variance Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
cyl	1	4796	4796	330.4	6.632e-46
trans	9	229.1	25.45	1.754	0.07841
Residuals	223	3237	14.51	NA	NA

Table 3: Analysis of Variance Table We hypothesize that there is the type of transmission does not have a significant impact on the fuel economy of a vehicle and that the impact of wheel-drive-type on fuel economy does not vary significantly across class of vehicle. As explained, in Analysis section, we use two models to analyze the first hypothesis and a single model to answer the second one.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
drv	2	4385	2192	234.4	1.88e-55
class	6	1773	295.5	31.6	2.381e-27
drv:class	3	28.03	9.344	0.9992	0.3941
Residuals	222	2076	9.352	NA	NA

For determining the impact of type of transmission on fuel economy, we decided to use the model including the number of cylinders, as it is the more conservative of the two, in order to answer if the type of transmission impacts fuel economy. Using this model, we fail to reject the null hypothesis (p-value=0.07841). We conclude that we have little evidence that the type of transmission impacts the fuel economy of a vehicle. This conclusion confirms our hypothesis.

For the impact of wheel-drive-type on fuel economy by class of vehicle, we fail to reject the null hypothesis. Thus, we conclude that there is little evidence that the impact of wheel-drive-type on fuel economy varies across vechicle class. This conclusion confirms our hypothesis.

These resutls add an important discussion to the debate on how governments try to regulate the emissions and how consumers decided on vechicle purchases. Car manufacturers will need to find ways in which the respect the regulations regarding fuel economy and balance this with consumer wants.

Conclusion

We summarize the findings as follows:

1. There is little evidence that the type of transmission impacts the fuel economy of a vehicle.
2. There is little evidence that the wheel-drive-type impacts fuel economy by vehicle class.

Appendix

```
library(ggplot2)
data(mpg)

# Data Wrangling
```

```r
# Change all the chr variables to factor variables
for ( i in names(Filter(is.character,mpg))){
  mpg[[i]] <- as.factor(mpg[[i]])}
mpg <- as.data.frame(mpg)
# Check data types
str(mpg)

# Boxplot for fuel economy split by transmission type
hwy_and_trans_boxplot <- ggplot(mpg, aes(trans, hwy)) +
  geom_boxplot(varwidth = TRUE) + coord_flip()
hwy_and_trans_boxplot

# Initial ANOVA model without confounder
model_without_confounder <- lm(hwy~trans, data=mpg)
summary(aov(model_without_confounder))

# table to clearly see why the number of cylinders
# may be a counfounding variable for transmission type
cyl_and_trans <- table(mpg$trans, mpg$cyl)

# model for analyzing confounder
model_confounder <- lm(hwy~cyl, data=mpg)
summary(model_confounder)
# Boxplot for above model
ggplot(mpg, aes(trans, model2$residuals)) + geom_boxplot() + coord_flip()

# ANCOVA model with confounder
model_with_confounder <- lm(hwy~cyl+trans, data=mpg)
anova(model_with_confounder)

# Interaction plot for analyzing impact of wheel-drive-type by vehicle class
par(mfrow=c(2,1))
interaction.plot(mpg$drv, mpg$class,mpg$hwy, col=rainbow(7), main="drv vs class")
interaction.plot(mpg$class, mpg$drv,mpg$hwy, col=rainbow(3), main="class vs drv")

# two-way ANOVA model
model_drv_class <- lm(hwy~drv*class, data=mpg)
anova(model_drv_class)
```