Does transmission type affect fuel efficiency?

Rashmitha Rallapalli, Regression Models,Coursera Peer Assignment 2016

Executive Summary

This week Motor Trend is going to look at the affect of automatic transmissions on fuel efficiency. To do this we will use a data set that examines the fuel efficiency and 10 aspects of automobile design and performance for 32 automobiles (all 1973 - 1974 models). Out of the 32 cars, 13 have manual transmissions and 19 have automatic transmissions. In this data set on average there is a difference in fuel efficiency depending on transmission type such that on average manual vehicles achieve a fuel efficiency of 7.2 miles per gallon more than automatic vehicles. However, transmission type is not a particularly good predictor of fuel efficiency. By applying analysis of variance (ANOVA) to the dataset, calculating the correlations between the variables, and building a number of models, we were able to identify that the number of cylinders and the weight of the automobile are good predictors of fuel efficiency, achieving an adjusted R squared of 0.82. If we add transmission type to this model, then the difference in fuel efficiency for a manual transmission is much smaller, just 0.18 miles per gallon for a vehicle with the same weight and number of cylinders.

Therefore we conclude that number of cylinders and weight are good predictors of fuel efficiency, but transmission type is not.

Method :

The data set

The data set was extracted from the 1974 edition of Motor Trend US Magazine and it deals with 1973 - 1974 models. It consists of 32 observations on 11 variables:

mpg: Miles per US gallon

cyl: Number of cylinders

 disp: Displacement (cubic inches)

hp: Gross horsepower

drat: Rear axle ratio

wt: Weight (lb / 1000)

qsec: 1 / 4 mile time

vs: V/S

am: Transmission (0 = automatic, 1 = manual)

gear: Number of forward gears

carb: Number of carburetors Results

The exploratory analysis of the data is described in Appendix. Based on the exploratory analysis, we selected three models to explore the question posed by this report:

Model **3** relates the number of cylinders and the weight to fuel efficency and achieves adjusted R squared of **0.8185**. Model **7** relates the type of type transmission to fuel efficiency and achieves adjusted R squared of **0.3385**. Model **8** relates the number of cylinders, the weight and the type of transmission to fuel effiency and achieves adjusted R squared of **0.8122**.

First we will examine the coefficients for model 3:

Estimate Std. Error t value Pr (>|t|)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	39.6863	1.7150	23.14	0.0000
cyl	-1.5078	0.4147	-3.64	0.0011
wt	-3.1910	0.7569	-4.22	0.0002

Here we see as the number of cylinders increase, assuming weight stays the same, for one additional cylinder fuel efficiency decreases by 1.5 miles per gallon. We also see that as the weight increases by 1000 lb, assuming the number of cylinders stays the same, the fuel efficiency decreases by 3.2 miles per gallon. The intercept for fuel efficiency is 39.7 miles per gallon, for a non-existant car with 0 weight and no cylinders - the highest actual efficiency car in the dataset is the Toyota Corolla which achieves 33.9 mpg.

Second we will consider the relationship between the type of transmission and fuel efficiency in Model 7:

Estimate Std. Error t value Pr(>|t|)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	17.1474	1.1246	15.25	0.0000
am	7.2449	1.7644	4.11	0.0003

The coefficient for transmission indicates that cars with a manual transmission achieve a fuel efficiency of **7.24** miles per gallon higher than cars with an automatic transmission. However this model using transmission achieves an adjusted R squared value **of 0.3385**, which is much worse than model **3**, so transmission is much poorer predictor of fuel efficiency than the number of cylinders and weight.

Third we will consider the relationship between type of transmission and fuel efficiency in Model 8: **Estimate Std. Error t value Pr(>|t|)**

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	39.4179	2.6415	14.92	0.0000
cyl	-1.5102	0.4223	-3.58	0.0013
wt	-3.1251	0.9109	-3.43	0.0019
am	0.1765	1.3045	0.14	0.8933

If we look at the coefficients, we see that compared to model **7** the coefficient for transmission type has decreased, indicating that for the same number of cylinders and car weight, this model predicts that a car with a manual transmission would achieve an improvement in fuel efficiency of just **0.18** miles per gallon over an automatic transmission. This model achieved an adjusted R squared value of **0.8122**. This is slightly worse than model **3**.

Finally we will compare predictions and their confidence intervals from model **3** and model 8. We will predict the fuel efficiency in mpg for a vehicle with the mean weight and mean number of cylinders in the data set using model **3**:

fit   lwr   upr

1 20.09 19.16 21.02

Next we will predict the fuel efficiency for the same vehicle, only first automatic transmission, and second with manual transmission, using model **8**:

fit lwr upr

1 20.02 18.58 21.46

fit lwr upr

1 20.2 18.35 22.04

We see there is a small difference between the predictions of model **3 (20.09)** and model **8 (20.2),** which is much smaller than the confidence intervals. Therefore we propose the transmission type makes a negligible difference to the fuel efficiency.

Conclusion

Although in this data set on average manual vehicles achieve a fuel efficiency of **7.2** miles per gallon more than automatic vehicles, transmission type is not a particularly good predictor of fuel efficiency. We were able to identify that the number of cylinders and the weight of the automobile are good predictors of fuel efficiency, achieving an adjusted R squared of **0.82**. If we add transmission type to this model, then the difference in fuel effiency for a manual transmission is much smaller, just **0.18** miles per gallon for a vehicle with the same weight and number of cylinders. Therefore we conclude that number of cylinders and weight are good predictors of fuel efficiency, but transmission type is not.

Appendix - Exploratory Data Analysis

First we will use analysis of variance to investigate the relationship between fuel efficiency (mpg) and the other variables:

```
data(mtcars)
options(contrasts = c("contr.sum", "contr.poly"))
aov.1 <- aov(mpg ~ ., data = mtcars)
library(xtable)
 print(xtable(aov.1), type = "html")
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
cyl	1	817.71	817.71	116.42	0.0000
disp	1	37.59	37.59	5.35	0.0309
hp	1	9.37	9.37	1.33	0.2610
drat	1	16.47	16.47	2.34	0.1406
wt	1	77.48	77.48	11.03	0.0032
qsec	1	3.95	3.95	0.56	0.4617
vs	1	0.13	0.13	0.02	0.8932
am	1	14.47	14.47	2.06	0.1659
gear	1	0.97	0.97	0.14	0.7137
carb	1	0.41	0.41	0.06	0.8122
Residuals	21	147.49	7.02		

We see that the number of cylinders (cyl), the weight (wt) and the displacement (disp) are all significant at the 0.05 level. Next we will attempt to identify possible confounders by looking at the correlations between these variables:

```
c <- cor(mtcars)
c[upper.tri(c)] <- NA
print(xtable(c), type = "html")
```

mpg cyl disp hp drat wt qsec vs am gear carb

mpg   1.00
cyl   -0.85   1.00
disp -0.85   0.90   1.00
hp   -0.78   0.83   0.79   1.00
drat 0.68 -0.70 -0.71 -0.45 1.00
wt   -0.87   0.78   0.89   0.66 -0.71 1.00
qsec 0.42 -0.59 -0.43 -0.71 0.09 -0.17   1.00
vs   0.66 -0.81   -0.71 -0.72 0.44 -0.55   0.74   1.00
am   0.60 -0.52   -0.59 -0.24   0.71 -0.69 -0.23   0.17 1.00
gear 0.48 -0.49   -0.56 -0.13   0.70 -0.58 -0.21   0.21 0.79   1.00
carb -0.55   0.53   0.39 0.75 -0.09 0.43 -0.66 -0.57 0.06   0.27 1.00

Here we see that there is a strong relationship between displacement and cylinders (0.9), displacement and weight (0.89) and a slightly less strong relationship between weight and cylinders (0.78). The next step is build a number of different regression models to investigate whether these variables can be used to predict fuel efficency.

Fitting multiple regression models

Next we build a number of models, using combinations of the variables we identified in the previous section:

```
fit1 <- lm(mpg ~ cyl, data = mtcars)
fit2 <- lm(mpg ~ wt, data = mtcars)
fit3 <- lm(mpg ~ cyl + wt, data = mtcars)
fit4 <- lm(mpg ~ disp, data = mtcars)
fit5 <- lm(mpg ~ disp + cyl, data = mtcars)
fit6 <- lm(mpg ~ disp + cyl + wt, data = mtcars)
```

The models achieve the following adjusted R squared values:

Model 1: Cylinders to fuel efficiency **0.7171**.

Model 2: Weight to fuel efficiency **0.7446**.

Model 3: Cylinders and weight to fuel efficiency **0.8185**.

Model 4: Displacement to fuel efficiency **0.709**.

Model 5: Displacement and cylinders to fuel efficiency **0.743**.

Model 6: Displacement, cylinders and weight to fuel efficiency **0.8147**.

Clearly model 3 has the best adjusted R squared value, and even though model 6 has an additional term, displacement, it achieves a slightly lower adjusted R squared value. This agrees with our previous analysis that although cylinders, weight and displacement all have a significant relationship to fuel efficiency, displacement is strongly related to cylinders and weight so is a confounder and does not impart any additional information.

Next we check the diagnostic plots of the residuals for model 3.

```
par(mfrow = c(2, 2))
 plot(fit3)
```
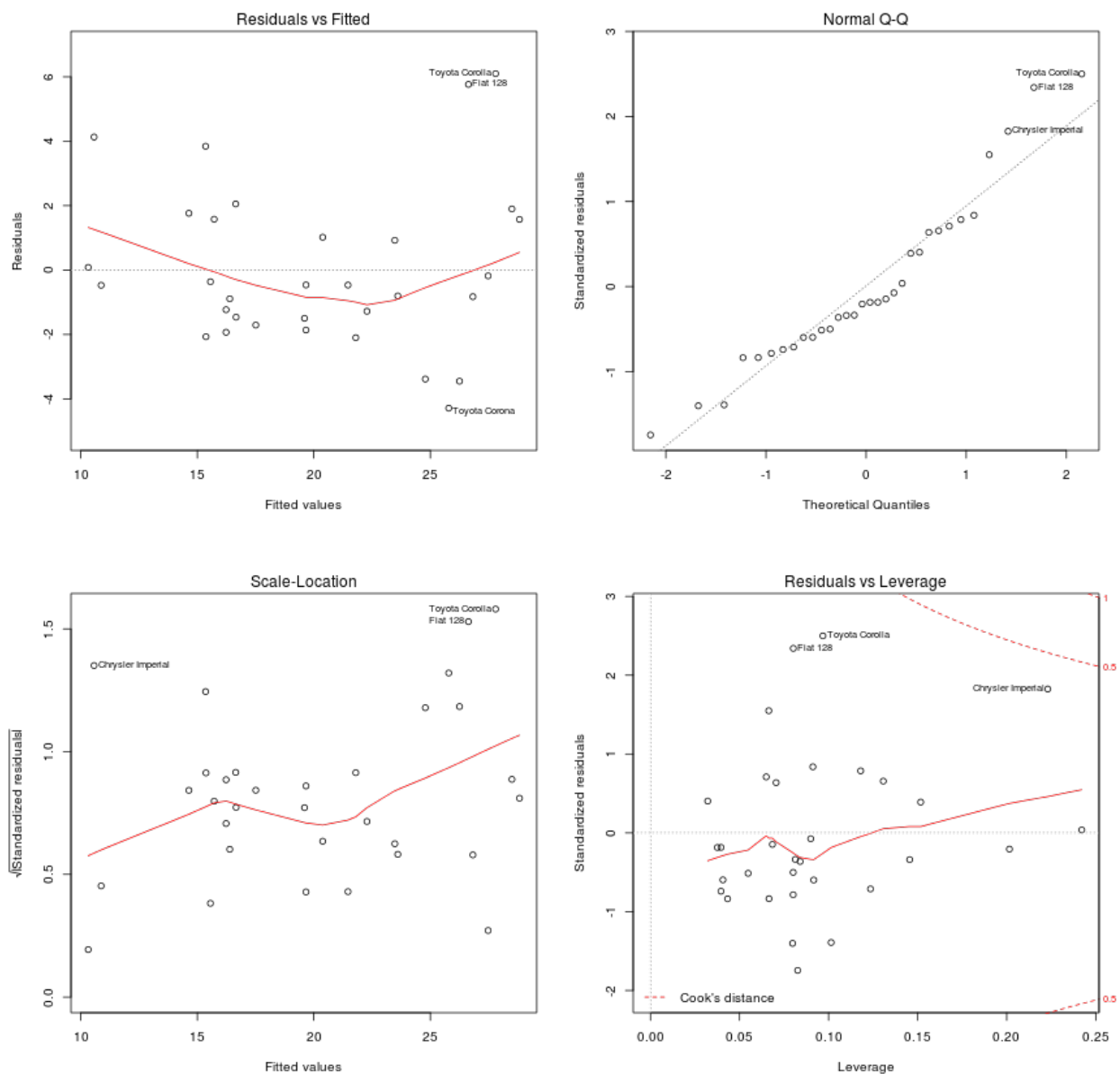
Figure 1: Diagnostic plots of residuals for model 3

The plots show that there are a see a number of outliers in the dataset, specifically the Toyota Corolla, Toyota Corona, Fiat 128 and Chrysler Imperial. The Toyota Corolla and Fiat 128 achieve a very high fuel efficiency (33.9 and 32.4 mpg respectively), whereas the Imperial has low fuel efficiency (14.7 mpg). The Toyota Corona achieves medium fuel efficiency (24.9 mpg).

```
boxplot(mpg ~ am, data = mtcars, xlab = "Transmission", ylab = "MPG", col = terrain.colors(2))
title(main = "Figure 2: Boxplot of transmission against MPG")
legend("topleft", inset = 0.05, title = "Transmission type", c("automatic", "manual"), fill =
terrain.colors(2), horiz = TRUE)
```

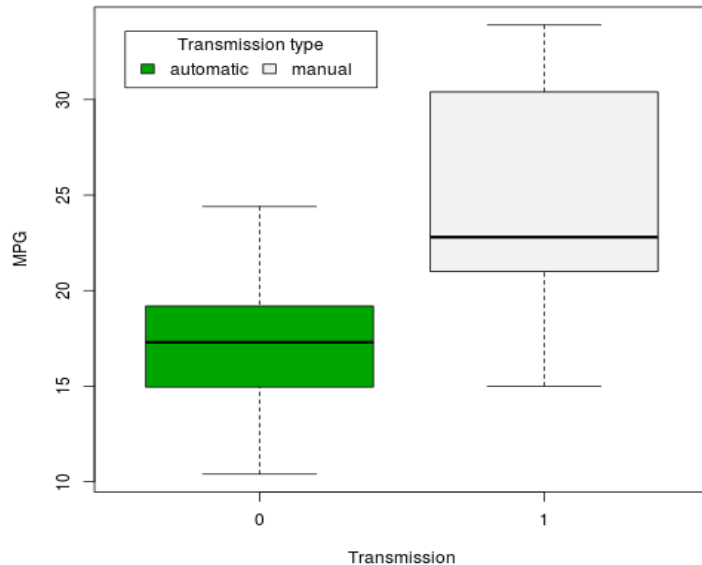**Figure 2: Boxplot of transmission against MPG**

Figure 2 shows that on average there is a difference between the fuel efficiency depending on transmission type.

```
fit7 <- lm(mpg ~ am, data = mtcars)
print(xtable(fit7), type = "html")
```

Estimate Std. Error t value Pr(>|t|)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	17.1474	1.1246	15.25	0.0000
am	7.2449	1.7644	4.11	0.0003

If we build a linear regression model that predicts the fuel efficiency solely based on transmission, then the coefficient indicates that cars with a manual transmission achieve a fuel efficiency of **7.24** miles per gallon higher than cars with an automatic transmission. However this model using transmission achieves an adjusted R squared value of **0.3385**, which is much worse than model 3, so transmission is much poorer predictor of fuel efficiency than the number of cylinders and weight.

Next we build a model that predicts the fuel efficiency using the number of cylinders, the weight and the transmission type:

```
fit8 <- lm(mpg ~ cyl + wt + am, data = mtcars)
print(xtable(fit8), type = "html")
```

Estimate Std. Error t value Pr(>|t|)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	39.4179	2.6415	14.92	0.0000
cyl	-1.5102	0.4223	-3.58	0.0013
wt	-3.1251	0.9109	-3.43	0.0019
am	0.1765	1.3045	0.14	0.8933

The model using cylinders, weight and transmission achieved an adjusted R squared value of **0.8122**. This is slightly worse than model 3, that did not use transmission. If we look at the coefficients, we see that the coefficient for transmission type has decreased, indicating that for the same number of cylinders and car weight, this model predicts that a car with a manual transmission would achieve an improvement in fuel efficiency of just **0.1765** miles per gallon.