Regression Models Final Project

Lucas Falcao Monteiro

26/08/2021

Analyzing fuel efficiency for automatic and manual transmission car models.

Executive summary

In this study we analyzed the impact of a manual transmission over fuel efficiency in cars. We used our database with 32 car models from 73/74 to answer a question that many of our readers ask us. Fitting a model with the transmission type, weight and 1/4 mile type as predictors led us to find a statistically significant relationship between transmission type and fuel efficiency as measured in miles per gallon. However, the strength of this relationship seemed to rely on a few outliers, and our study could be repeated with a better database that includes more vehicles.

Introduction

Our work in the *Motor Trend* magazine includes analyzing technical details about automobiles. One question of interest to our readers is the efficiency of cars, as measures in miles per gallon, since fuel is a considerable cost of maintaining a vehicle. In this report we aim to discuss whether automatic cars are as efficient as those with manual transmission.

Data collection

The data was extracted from our own collection, including 32 different 1973 and 1974 car models. Each observation contains 10 aspects, including the number of cylinders, horsepower, weight etc.

```
data <- mtcars
data$am <- as.factor(data$am)
data$cyl <- as.factor(data$cyl)
names(data)
```

```
##  [1] "mpg"  "cyl"  "disp" "hp"   "drat" "wt"   "qsec" "vs"   "am"   "gear"
## [11] "carb"
```

Results

We tested various models and how additional predictors offered more information. We decided to adjust for the weight and 1/4 mile time. We find the following estimate and confidence interval for the influence of transmission type over fuel efficiency:

```
summary(lm(mpg ~ am + wt + qsec, data))$coef[2, ]
```

```
##   Estimate Std. Error    t value   Pr(>|t|)
## 2.93583719 1.41090451 2.08081919 0.04671551
```

```
confint(lm(mpg ~ am + wt + qsec, data))[2, ]
```

```
##      2.5 %     97.5 %
## 0.04573031 5.82594408
```

Adjusting for the other parameters, there seems to be about a 3 mile per gallon increase in manual cars, and this is a significant change, since our 95% confidence interval does not include 0. However, excluding a few data points takes away the significance of our model (Appendix).

Conclusion

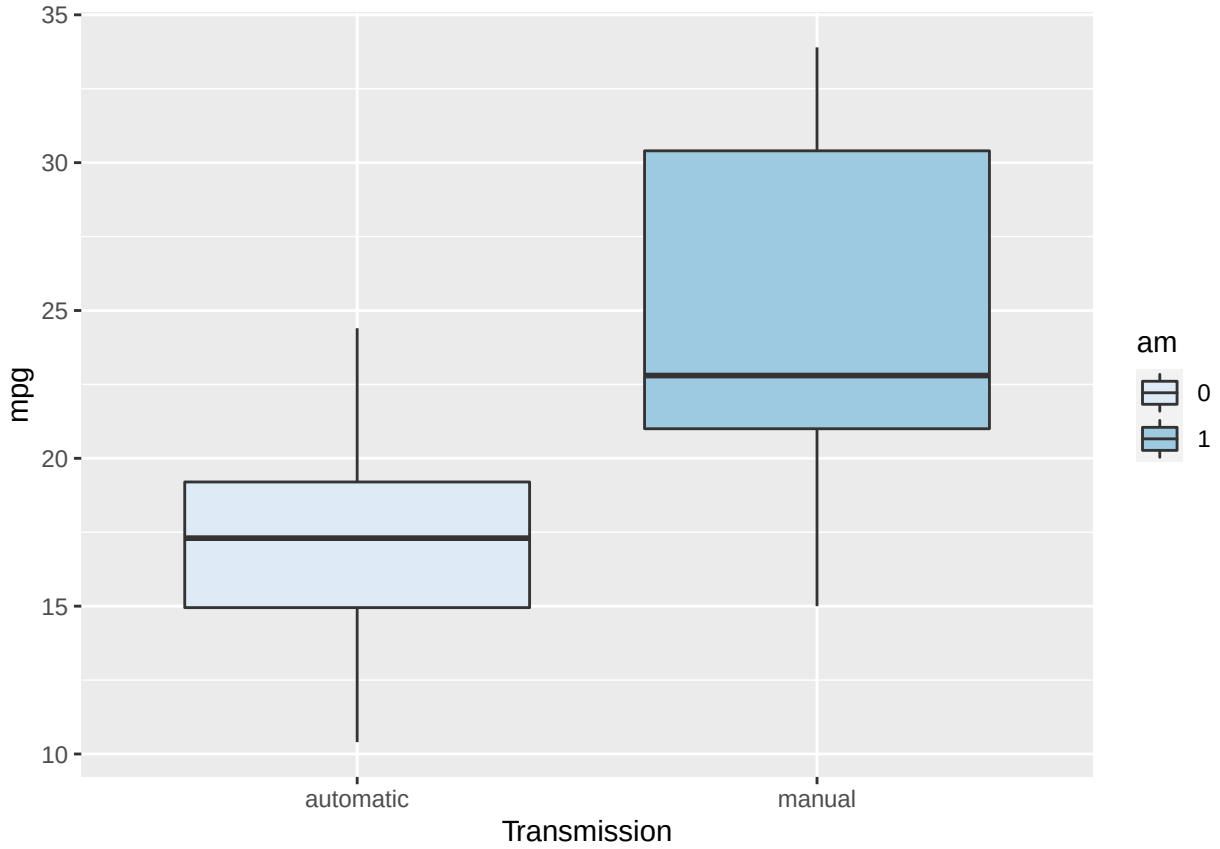
We can conclude from this data that transmission type has a significant effect over fuel efficiency for vehicles when controlling for weight and 1/4 mile time. However, taking away a few data points takes away the significance of this relationship, suggesting our result is not very robust. Also, our dataset is pretty small (32 observations), so a better analysis would be performed over a larger dataset, perhaps with hundreds of vehicles.

Appendix

The entries indicate whether the transmission is automatic (0) or manual (1) in the 'am' column. We can plot a histogram of mpg for these two groups.

```
library(ggplot2)


g <- ggplot(data = data, aes(y = mpg, x = am, fill = am))
g + geom_boxplot() + scale_fill_brewer(palette = "Blues") + scale_x_discrete(name = "Transmission",
    labels = c(`0` = "automatic", `1` = "manual"))
```



It seems that manual transmission cars are more efficient than automatic ones. We can fit a linear model using transmission as a predictor for the outcome 'mpg' and check its significance.

```
mdltrans <- lm(mpg ~ am, data)
summary(mdltrans)$coef
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## am1          7.244939   1.764422  4.106127 2.850207e-04
```

This preliminary model indicates there is a significant effect of transmission over fuel efficiency. However, this model might be biased because we are not correcting for other parameters that clearly affect efficiency, such as number of cylinders, horsepower and weight. Adding those into a separate model:

```
mdl <- lm(mpg ~ am + cyl + hp + wt, data)
summary(mdl)$coef
```

```
##                Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 33.70832390 2.60488618 12.940421 7.733392e-13
## am1          1.80921138 1.39630450  1.295714 2.064597e-01
## cyl6        -3.03134449 1.40728351 -2.154040 4.068272e-02
## cyl8        -2.16367532 2.28425172 -0.947214 3.522509e-01
## hp          -0.03210943 0.01369257 -2.345025 2.693461e-02
## wt          -2.49682942 0.88558779 -2.819404 9.081408e-03
```

Now we see that for a fixed number of cylinders, horsepower and weight, the change from automatic to manual transmission might increase the fuel efficiency by 1.8 miles per gallon. We can use the best function to identify whether there is a better combination.

```
mdl0 <- lm(mpg ~ am + cyl + hp + wt + qsec + drat + disp + vs +
    gear + carb, data)
mdlbest <- step(mdl0, direction = "both", trace = FALSE)
summary(mdlbest)$coef
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)  9.617781  6.9595930  1.381946 1.779152e-01
## am1          2.935837  1.4109045  2.080819 4.671551e-02
## wt          -3.916504  0.7112016 -5.506882 6.952711e-06
## qsec         1.225886  0.2886696  4.246676 2.161737e-04
```

The best model that was found included the transmission type, weight and the 1/4 mile time, which is how fast you can transverse 1/4 of a mile. We can perform an anova test to see if adding these predictors improves the model over just using the transmission mode.

```
mdl1 <- lm(mpg ~ am + wt, data)

anova(mdltrans, mdl1, mdlbest)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + wt
## Model 3: mpg ~ am + wt + qsec
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     29 278.32  1    442.58 73.203 2.673e-09 ***
## 3     28 169.29  1    109.03 18.034 0.0002162 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can also obtain the confidence interval for the slope of the transmission type predictor.

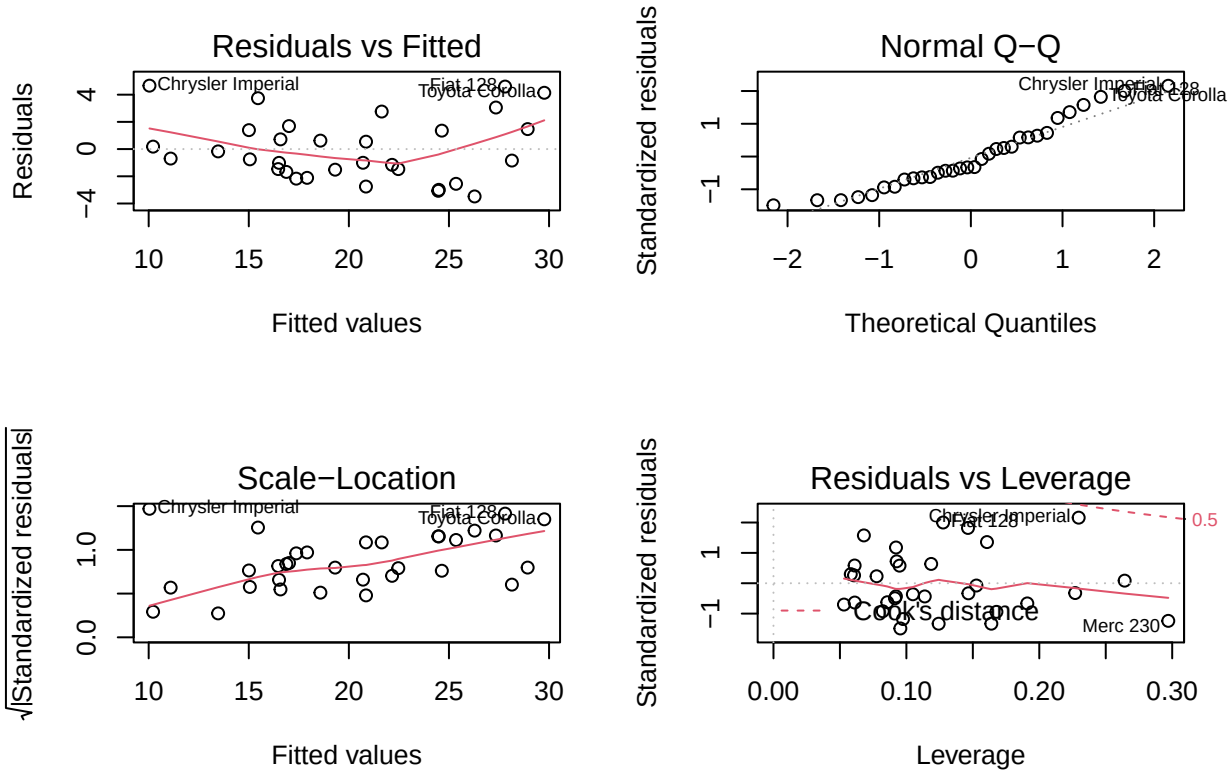
```
confint(mdlbest)[2, ]
```

```
##      2.5 %     97.5 %
## 0.04573031 5.82594408
```

This confidence interval does not contains 0, so we can attest that transmission type affects fuel efficiency.

Checking the residuals and other diagnostic plots:

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



We see that there are some models influencing the model, such as *Toyota Corolla*, *Toyota Corona*, *Chrysler Imperial* and *Fiat 128*. We can try to remove those points and evaluate the effect over the model.

```
data$car <- rownames(data)
data1 <- data[!(data$car %in% c("Toyota Corolla", "Toyota Corona",
    "Chrysler Imperial", "Fiat 128")), ]
mdlnew <- lm(mpg ~ am + wt + qsec, data1)
confint(mdlnew)[2, ]
```

```
##     2.5 %    97.5 %
## -1.990623  2.817522
```

In this case, the confidence interval includes 0, so excluding 4 data points takes away the significance of our model and we can no longer state that transmission type influences fuel efficiency.