Regression Models Final Project

Lucas Falcao Monteiro

26/08/2021

Analyzing fuel efficiency for automatic and manual transmission car models.

Executive summary

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
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 number of cylinders, horsepower and weight, and we find the following estimate and confidence interval for the influence of transmission type over fuel efficiency:

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

```
##   Estimate Std. Error    t value   Pr(>|t|)
##  1.4780477  1.4411493  1.0256035  0.3141799
```

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

```
##      2.5 %    97.5 %
## -1.478946  4.435042
```

While there seems to be a small increase in fuel efficiency (1.48 miles per gallon) for manual transmission cars, the confidence interval includes 0. The p-value is 0.31, meaning there is a 31% chance of getting this result if we consider that there is no influence of the predictor on the outcome.

Conclusion

We cannot conclude from this data that transmission type significantly improves fuel efficiency for vehicles when controlling for weight, number of cylinders and horsepower. However, 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.

```r
library(ggplot2)

data$am <- as.factor(data$am)

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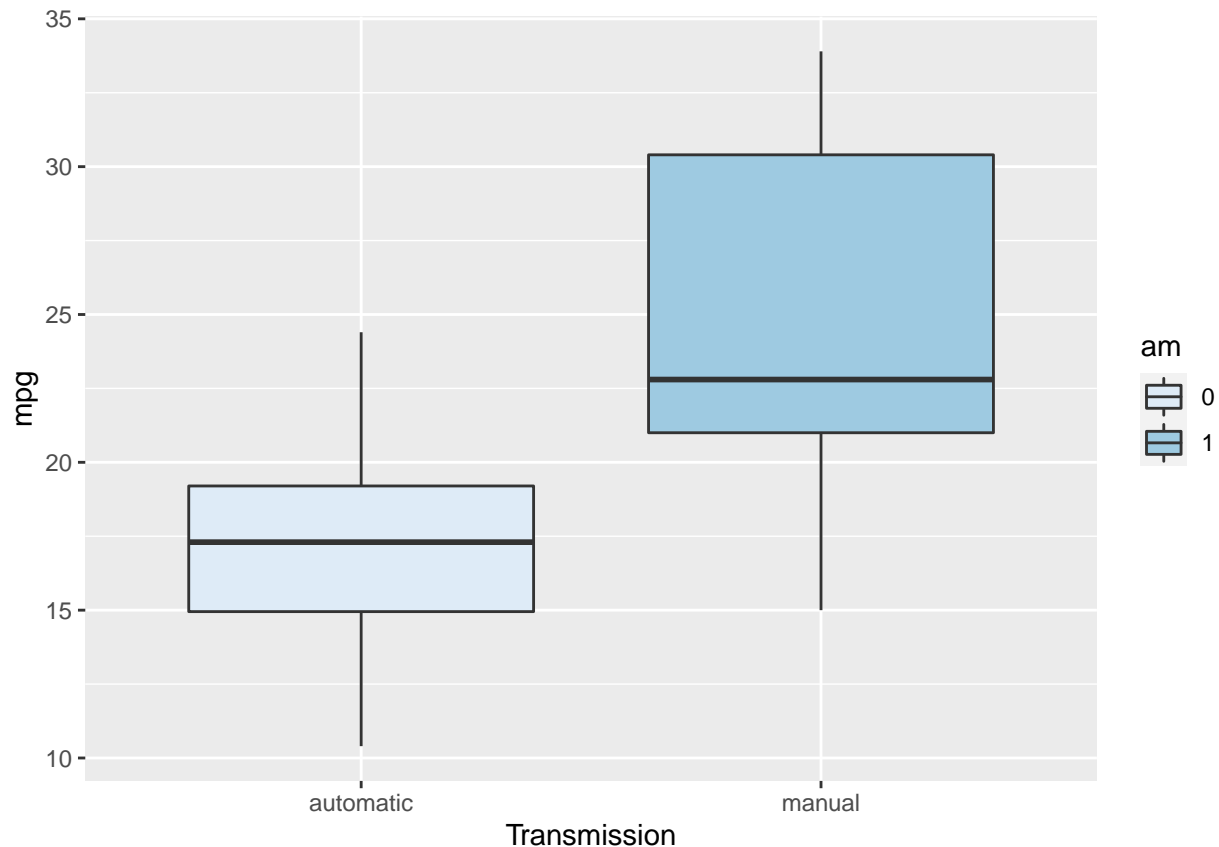
It seems that manual transmission cars are more efficient than automatic ones. We can fit a linear model using transmission as a predictor for the outcome 'mpg' and check its significance.

```
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) 36.14653575 3.10478079 11.642218 4.944804e-12
## am1          1.47804771 1.44114927  1.025603 3.141799e-01
## cyl         -0.74515702 0.58278741 -1.278609 2.119166e-01
## hp          -0.02495106 0.01364614 -1.828433 7.855337e-02
## wt          -2.60648071 0.91983749 -2.833632 8.603218e-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 1.5 miles per gallon. We can create several models and test the additional information each predictor adds to the model.

```
mdl0 <- lm(mpg ~ am + cyl, data)
mdl1 <- lm(mpg ~ am + cyl + hp, data)
mdl2 <- lm(mpg ~ am + cyl + hp + wt, data)
mdl3 <- lm(mpg ~ am + cyl + hp + wt + qsec, data)
mdl4 <- lm(mpg ~ am + cyl + hp + wt + qsec + drat, data)
mdl5 <- lm(mpg ~ am + cyl + hp + wt + qsec + drat + disp, data)
mdl6 <- lm(mpg ~ am + cyl + hp + wt + qsec + drat + disp + vs,
    data)
mdl7 <- lm(mpg ~ am + cyl + hp + wt + qsec + drat + disp + vs +
    gear, data)
mdl8 <- lm(mpg ~ am + cyl + hp + wt + qsec + drat + disp + vs +
    gear + carb, data)

anova(mdltrans, mdl0, mdl1, mdl2, mdl3, mdl4, mdl5, mdl6, mdl7,
    mdl8)
```

```
## Analysis of Variance Table
##
## Model  1: mpg ~ am
## Model  2: mpg ~ am + cyl
## Model  3: mpg ~ am + cyl + hp
## Model  4: mpg ~ am + cyl + hp + wt
## Model  5: mpg ~ am + cyl + hp + wt + qsec
## Model  6: mpg ~ am + cyl + hp + wt + qsec + drat
## Model  7: mpg ~ am + cyl + hp + wt + qsec + drat + disp
## Model  8: mpg ~ am + cyl + hp + wt + qsec + drat + disp + vs
## Model  9: mpg ~ am + cyl + hp + wt + qsec + drat + disp + vs + gear
## Model 10: mpg ~ am + cyl + hp + wt + qsec + drat + disp + vs + gear + carb
##     Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1       30 720.90
## 2       29 271.36  1    449.53 64.0039 8.231e-08 ***
## 3       28 220.55  1     50.81  7.2341   0.01372 *
## 4       27 170.00  1     50.56  7.1980   0.01393 *
## 5       26 159.82  1     10.18  1.4495   0.24201
## 6       25 158.64  1      1.18  0.1680   0.68605
## 7       24 149.09  1      9.55  1.3594   0.25672
## 8       23 148.87  1      0.22  0.0309   0.86214
## 9       22 147.90  1      0.97  0.1384   0.71365
## 10      21 147.49  1      0.41  0.0579   0.81218
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Therefore, we confirm that adding the number of cylinders, the horsepower and the weight of the vehicles offers more information about the fuel efficiency. That would be model 2 in our code. Finding the confidence interval for the slope of the transmission in this model:

```
confint(mdl2)
```
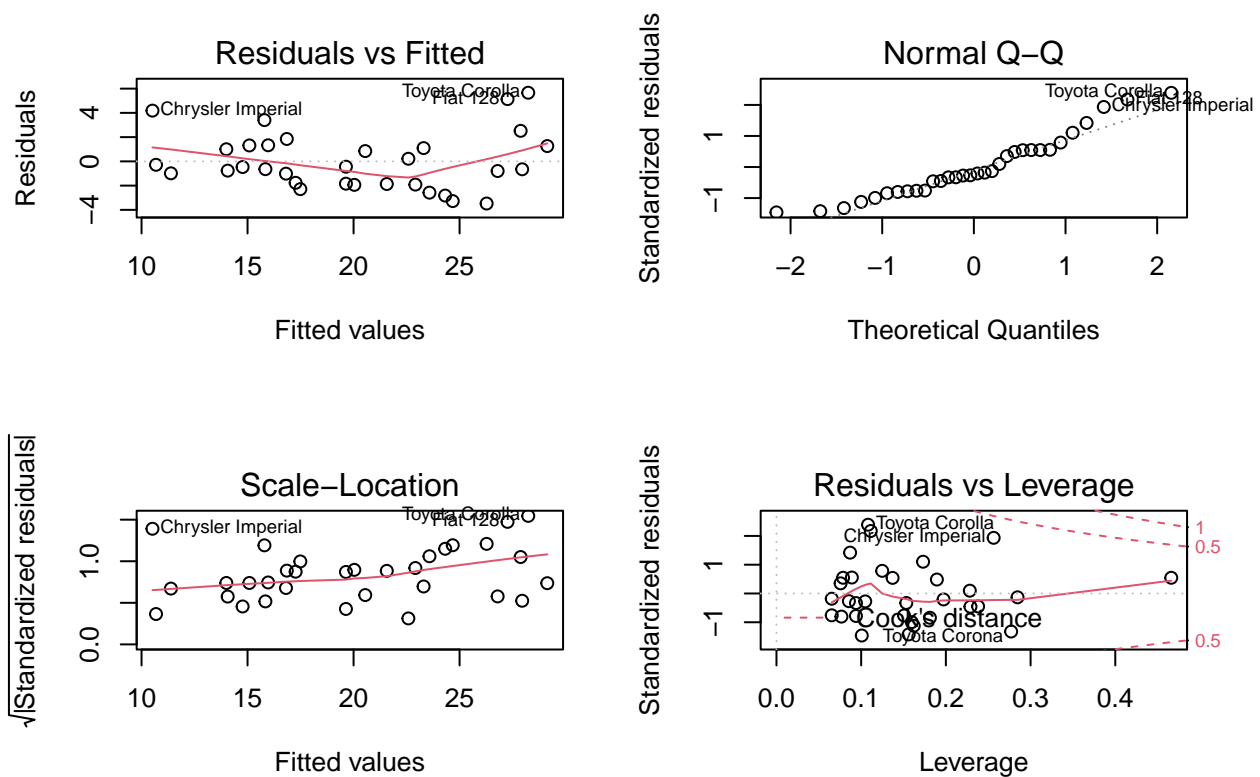
```
##                      2.5 %       97.5 %
```

```
## (Intercept) 29.77605177 42.517019733
## am1           -1.47894635  4.435041763
## cyl           -1.94093802  0.450623969
## hp            -0.05295064  0.003048517
## wt            -4.49383134 -0.719130075
```

This confidence interval contains 0, so we cannot safely attest that transmission type affects fuel efficiency.

Checking the residuals and other diagnostic plots:

```
par(mfrow = c(2, 2))
plot(mdl2)
```


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 + cyl + hp + wt, data1)
confint(mdlnew)
```

```
##                    2.5 %       97.5 %
## (Intercept) 34.36084008 44.701408206
## am1          -3.23948625  1.385417757
## cyl          -1.79104388 -0.034170646
```

```
## hp         -0.03187145  0.008764671
## wt         -5.24847338 -2.322621204
```

In this case, the center of the distribution is closer to 0, indicating an even less pronounced effect of transmission over fuel efficiency.