Car Transmission vs. Fuel Consumption – Regression Models

fjelltronen

May 13, 2018

Executive Summary

This report explores the relationship between a set of car-related variable and the miles-per-gallon attribute. First, we introduce the `mtcars` dataset. Then we fit a linear model and use its coeficients to determine that manual transmissions lead to a better fuel consumption when compared with automatic ones, by an estimated 7.245 miles-per-gallon.

MTCARS Dataset

The Motor Trend Car Road Tests data consists of fuel consumption and 10 aspects of automobile design/performance for 32 cars. A brief analysis of the data is shown in the Appendix below.

Convert some of the numeric variables into factors, including `am` (transmission information, where `0` is automatic and `1` is manual):

```
mtcars$am <- factor(mtcars$am)
```

A quick exploration of the relationship between the `am` and `mpg` variables is shown graphically in the Appendix.

Is an automatic or manual transmission better for `mpg`? / Quantify the `mpg` difference between automatic and manual transmissions

Let us fit a linear model with `mpg` as the outcome and `am` as the regressor. The coeficients of this least square regression model can help answer these questions.

```
fit <- lm(mpg~am, data = mtcars)
coefficients(fit)
```

```
## (Intercept)         am1
##   17.147368    7.244939
```

The adjusted estimate for the expected change in `mpg` when comparing automatic transmission (`(Intercept)` or `am0`) and manual transmission (`am1`) is `7.245`. This positive number indicates that the automatic transmission is worse for `mpg` (less miles-per-gallon imply worse fuel consumption). More specifically, the estimated difference in `mpg` between automatic and manual transmissions is 7.245 (the mean of `am1` is `24.392`, `7.245` higher than the mean of `am0`, which is `17.147`).

```
summary(fit)
```

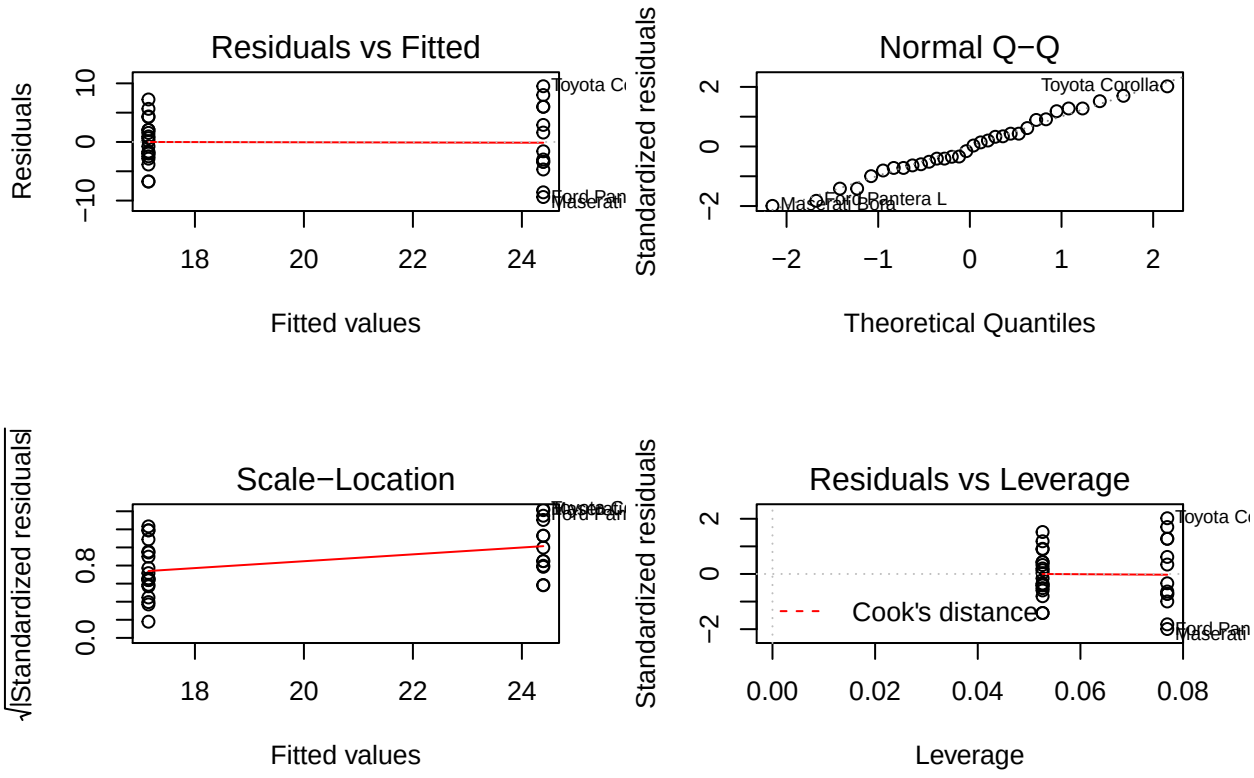
```
##
## Call:
## lm(formula = mpg ~ am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
```

```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.147      1.125  15.247 1.13e-15 ***
## am1            7.245      1.764   4.106 0.000285 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.902 on 30 degrees of freedom
## Multiple R-squared:  0.3598, Adjusted R-squared:  0.3385
## F-statistic: 16.86 on 1 and 30 DF,  p-value: 0.000285
```

Since the p-value of this coefficient is `0.000285` (< 0.05), the t-test for H_0: no difference between `am0` and `am1` versus H_a: difference between `am0` and `am1` is *significant*.

Furthermore, in order to evaluate the fit and residuals of the generated model, we use diagnostic plots.

```
par(mfrow = c(2, 2)); plot(fit)
```



The model does not capture the variance in `mpg` for each type of transmission. Its residuals are normally distributed and spred equally along the range of predictors. No cases are influential with respect to Cook's distance.

Appendix

mtcars dataset

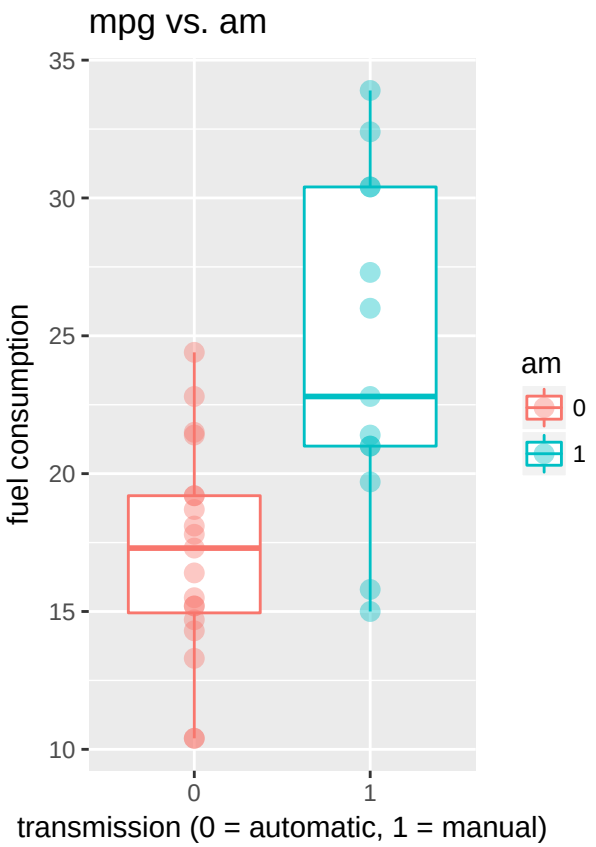
```
# structure of the data set
str(mtcars)
```

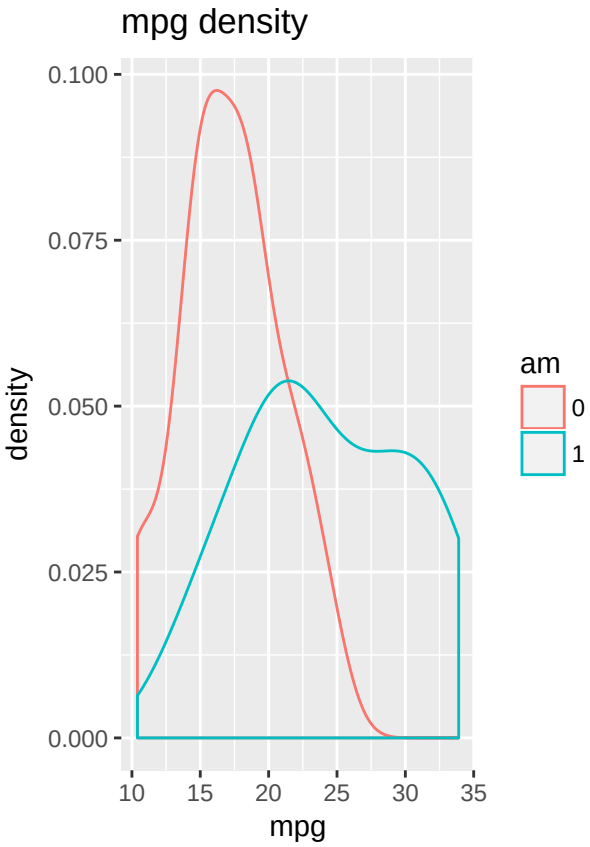
```
## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : Factor w/ 2 levels "0","1": 2 2 2 1 1 1 1 1 1 1 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

```
# value counts for `am`
table(mtcars$am)
```

```
##
##  0  1
## 19 13
```

```
# mpg for automatic/manual transmission
library(ggplot2)
library(gridExtra)
g1 <- ggplot(data = mtcars, aes(x = am, y = mpg, color = am)) +
  geom_boxplot() + geom_point(size = 3, alpha = 0.4) +
  labs(title = "mpg vs. am",
       x = "transmission (0 = automatic, 1 = manual)", y = "fuel consumption")
g2 <- ggplot(data = mtcars, aes(x = mpg, color = am)) + geom_density() +
  labs(title = "mpg density")
grid.arrange(g1, g2, ncol = 2)
```


mpg vs. am
fuel consumption
35
30
25
20
15
10
0
1
transmission (0 = automatic, 1 = manual)
am
0
1


mpg density
density
0.100
0.075
0.050
0.025
0.000
10
15
20
25
30
35
mpg
am
0
1