Regression models on Motor Trend cars dataset

Molina Rafidison

11 Apr 2016

Report

Introduction

Looking at a data set of a collection of cars, we are interested in exploring the relationship between a set of variables and miles per gallon (MPG). We are particularly interested in the following two questions:

- "Is an automatic or manual transmission better for MPG?"
- "Quantifying the MPG difference between automatic and manual transmissions"

Exploratory data analysis

First load the needed packages which are datasets, ggplot2 and gridExtra.

Load the dataset from the `datasets` package and look at the basic information.

```
##                mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4     21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag 21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710    22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
```

Look at the visual relationships of weight and acceleration on mileage per gallon according to transmission on Fig. 0 *(cf. Appendix Fig. 0)*.

We convert `am` into a factor; it is the variable we are interested in.

```
mtcars$am <- factor(mtcars$am, labels = c("Auto", "Manual"))
```

And we plot transmission versus miles per gallon *(cf. Appendix Fig. 1)* to highlight any suspect behavior.

Statistical inference

From Fig. 1, the manual transmission seems to allow more mileage per gallon than the automatic transmission. We use a t-test.

```
tTest <- t.test(mpg ~ am, data = mtcars)
```

```
## [1] "P-value: 0.00137363833307103"
```

```
## [1] "-11.2801943550402" "-3.20968418746996"
```

The p value is statistically significant (< 0.05). Plus, 0 is not in the confidence interval. We reject the Null hypothesis and confirm that automatic transmission allows less mileage per gallon than manual one **without considering any other variable**.

Regression model

We use the linear regression to fit a model, starting with a backward removal method for building the regression model.

```r
stepModel <- summary(step(lm(mpg ~ ., mtcars), trace = 0))
```

```
## [1] "Model: mpg ~ wt + qsec + am"
```

```
## [1] "Adjusted R-squared: 0.833556080257604"
```

This method tells us to keep the weight, the quarter mile time and thankfully the transmission. The model is not bad with an adjusted R-squared of 83%.

Test the model to optimize it by adding interaction from the transmission.

```r
fineModel <- lm(mpg ~ (am*wt) + qsec, mtcars)
```

```
##               Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)   9.723053  5.8990407  1.648243 0.1108925394
## amManual     14.079428  3.4352512  4.098515 0.0003408693
## wt           -2.936531  0.6660253 -4.409038 0.0001488947
## qsec          1.016974  0.2520152  4.035366 0.0004030165
## amManual:wt  -4.141376  1.1968119 -3.460340 0.0018085763
```

```
## [1] "Adjusted R-squared: 0.880421944614729"
```

We reached a 88% adjusted R-squared that we will check to avoid bias thanks to the residual plots *(cf. Appendix Fig. 2)*.

What we can say about the plots is that: the Residual vs. Fitted plot shows homoscedasticity that justifies the independance assumption; the residuals are approximately normally distributed according to the Normal Q-Q plot; the line representing the variance is rather constant on the Scale-Location plot; and everything is fine in the Residuals vs. Leverage plot.

Results

Conclusions

To answer the two questions initally asked:

Is an automatic or manual transmission better for MPG? We can conclude that choosing the best transmission for mileage per gallon depends on the weight and acceleration of a car.

Quantify the MPG difference between automatic and manual transmissions Choosing a manual transmission increases the mpg (+10.64 to +17.51) and so does the automatic transmission (+3.82 to +15.62). But the heavier the car, the more mpg decreases - -3.61 to -2.27 every 1,000 lbs - and even more with a manual transmission with -5.34 to -2.94 every 1,000 lbs. And the longer the quarter mile time, the slightly more mpg - +0.76 to +1.27 per second whatever the transmission.

That means that the heavier the cars with a quick acceleration the better thinking about an automatic transmission for better fuel efficiency. Small light cars would have a little bit more efficiency with manual transmission.

Appendix

Find figures related to the Report part.

Fig. 0

```
wtPlot <- ggplot(mtcars, aes(x = wt, y = mpg)) +
    geom_point(aes(color = am)) +
    labs(title = "Miles/gallon versus weight \n according to transmission",
         x = "Weight (in K lbs)",
         y = "Miles/gallon") +
    scale_fill_discrete(guide = FALSE) +
    theme(text = element_text(size = 7))

qsecPlot <- ggplot(mtcars, aes(x = qsec, y = mpg)) +
    geom_point(aes(color = am)) +
    labs(title = "Miles/gallon versus 1/4 mile time \n according to transmission",
         x = "Time to reach 1/4 mile (in sec)",
         y = "Miles/gallon") +
    scale_fill_discrete(guide = FALSE) +
    theme(text = element_text(size = 7))

grid.arrange(wtPlot, qsecPlot, ncol = 2)
```

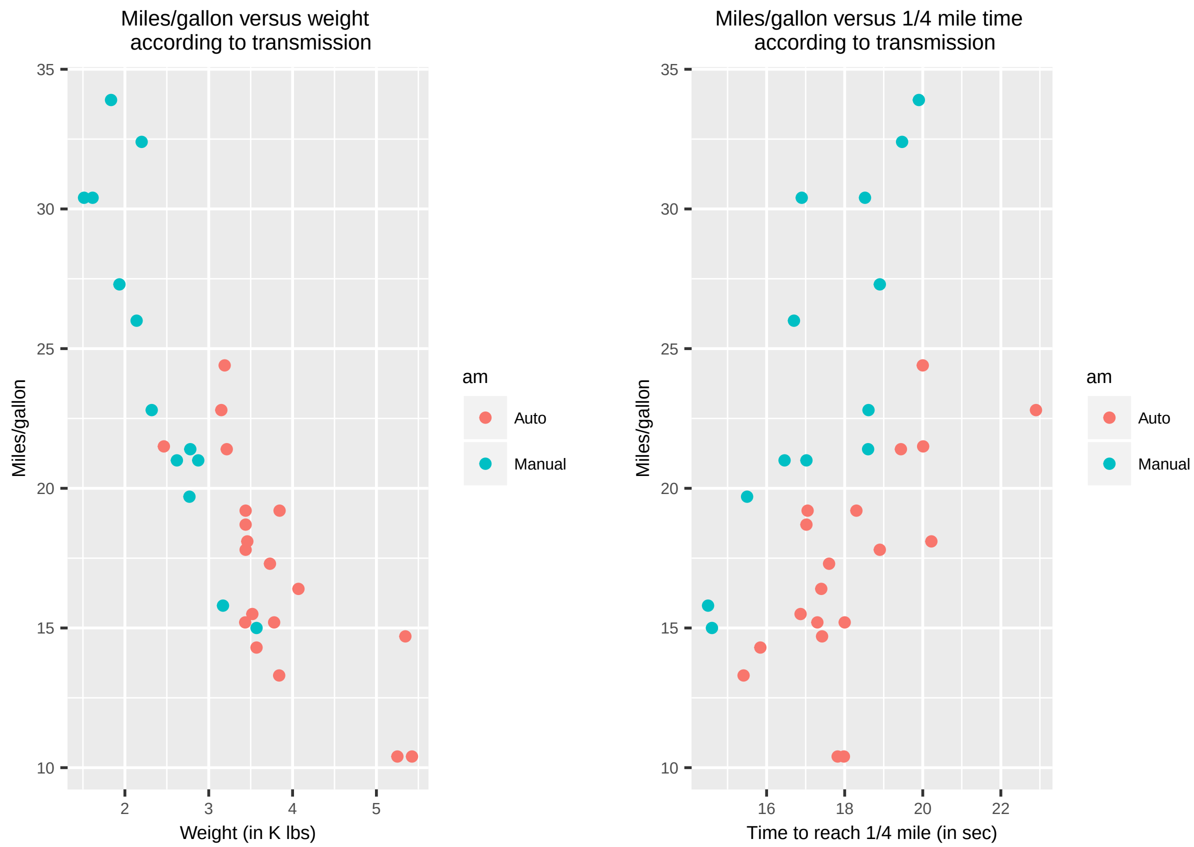


Fig. 1

```
ggplot(mtcars, aes(x = am, y = mpg, group = am)) +
    geom_boxplot() +
    labs(title = "Fig. 1: Miles per gallon versus type of transmission",
         x = "Transmission",
         y = "Miles/gallon") +
    scale_fill_discrete(guide = FALSE) +
    theme(text = element_text(size = 7))
```

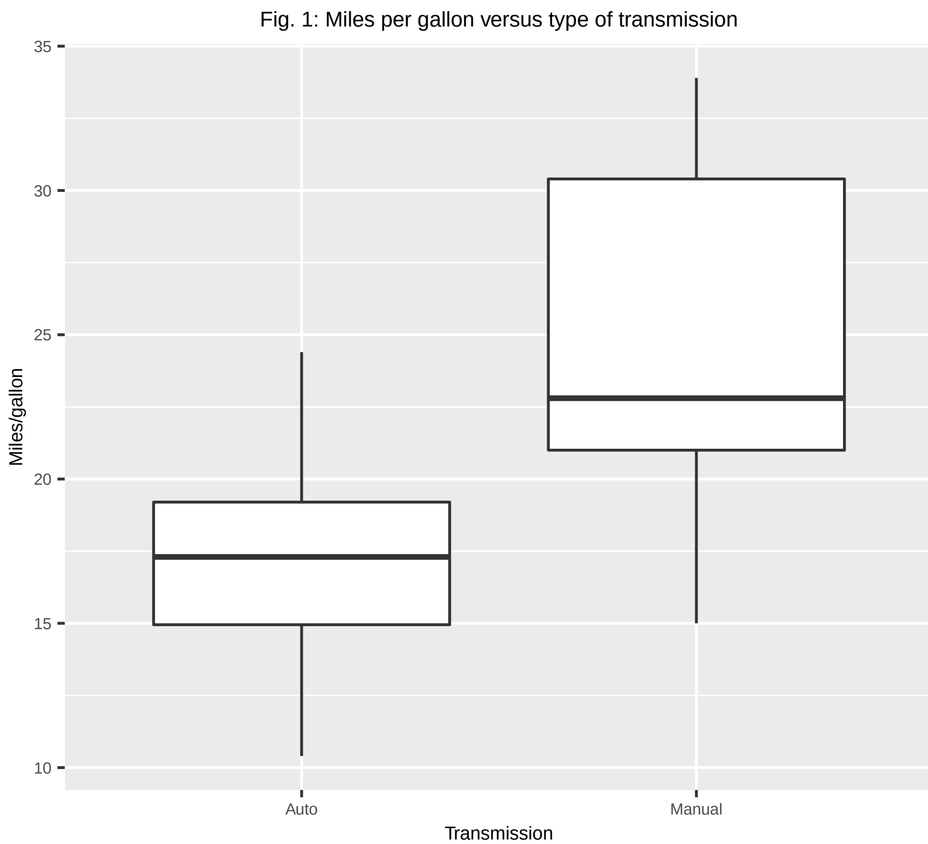
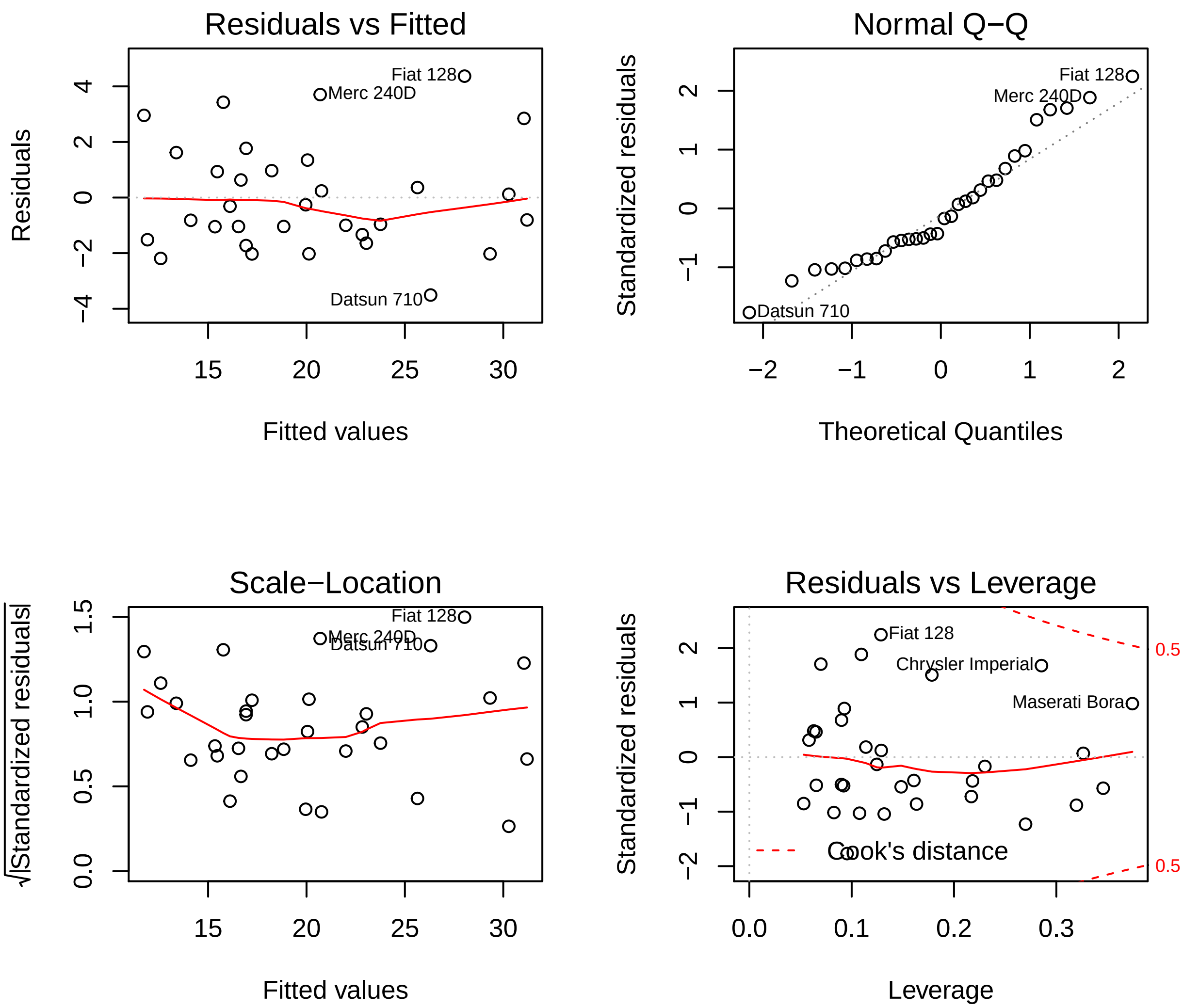


Fig. 2

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


Residuals vs Fitted
Residuals
Fitted values
Fiat 128
Merc 240D
Datsun 710
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Fiat 128
Merc 240D
Datsun 710
Scale−Location
√|Standardized residuals|
Fitted values
Fiat 128
Merc 240D
Datsun 710
Residuals vs Leverage
Standardized residuals
Leverage
Fiat 128
Chrysler Imperial
Maserati Bora
Cook's distance
0.5
0.5