Final project: Regression

Bobby den Bezemer

19 augustus 2015

Abstract

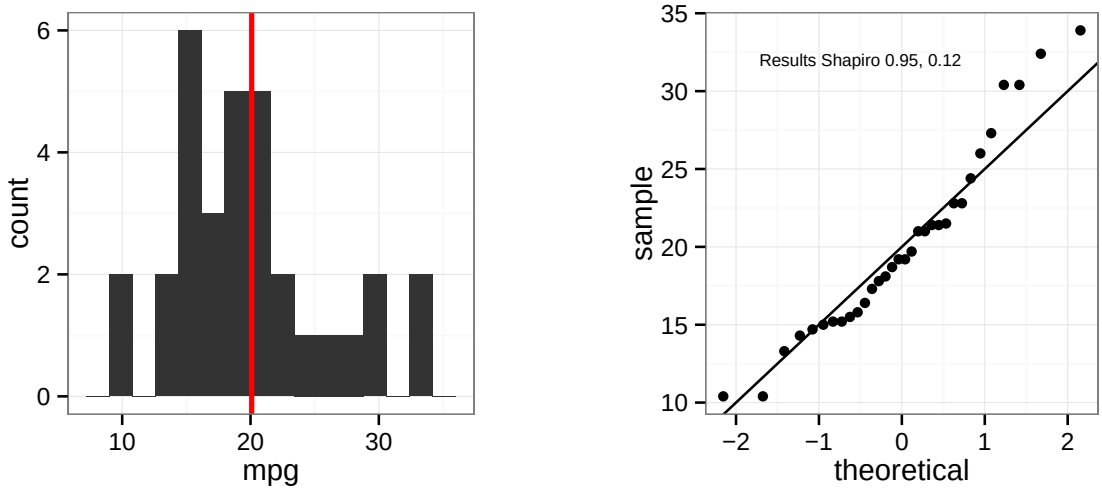
The current article proposes two models on the relationship between fuel efficiency (miles per gallon), car type and car weight. It concludes that a model consisting of car type, car weight and an interaction between car weight and type best explains the fuel efficiency of a car. Future research may expand the current model by including additional variables such as number of gears and cylinders.

Introduction

Motor Trend is the leading magazine on the automobile industry. In this article, we aim to investigate the relationship between different car types and fuel efficiency (miles per gallon). Second, it probes a second model expanding the earlier relationship by including the weight of cars as well as an interaction term of weight and car type. The report itself is written in Rmarkdown and expanded with customized latex. All graphs are included in the text to clarify the points being made to the reader. The reader is therefore suggested to evaluate the size of the paper as though it contains both text and the appendix (so a limit of 5 pages).

Exploratory analyses

To explore the data, several graphs have been made. Figure 1 shows the distribution of miles per gallon and the average miles per gallon. As can be seen that, the variable miles per gallon looks roughly normal (although there seems to be a slight deviation).



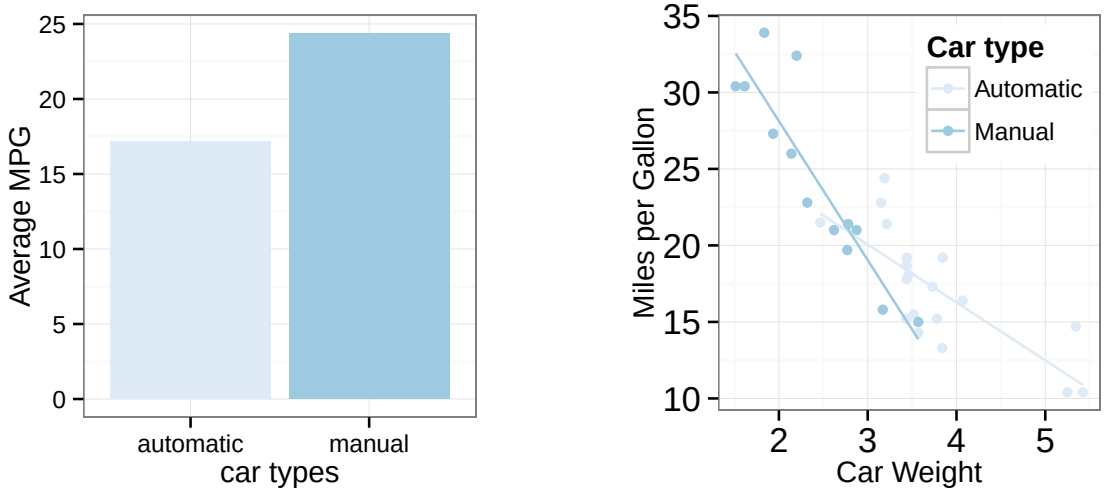
(a) Histogram of Distribution

(b) Normality plot

Figure 1: Figure 1

Figure 2 visualizes the effects of the predictors and how they relate to miles per gallon. As can be seen on the left part of figure 2, automatic cars have a higher average miles per gallon than manual cars. The right most

part of figure 2 shows the relationships between weight and miles per gallon for both automatic and manual cars. The next section will statistically analyze manual and automatic cars in terms of their miles per gallon.



(a) Average of Mpg per car type

(b) Car type, weight and MPG

Figure 2: Figure 1

Statistical analyses

The previous section explored automatic and manual cars in terms of their miles per gallon. This section will statistically analyze these car types.

Model 1

Below is displayed the output of regressing automatic and manual types on mpg.

```
##              Estimate Std. Error   t value     Pr(>|t|)     Mean Std. Dev
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15 3.833966 17.14737
## factor(am)1  7.244939   1.764422  4.106127 2.850207e-04 6.166504 24.39231
```

First of all, what follows from the output is that car type is a significant predictor of miles per gallon $\beta = 7.24, t(30) = 4.11, p < .001$. On average, automatic cars have a lower miles per gallon than manual cars ($M = 17.14$, $SD = 3.83$ vs $M = 24.39$, $SD = 6.17$). In order to see how much variance is explained by the model, we could extract the Rsquared. The model has the following $R^2 = 0.36$ meaning that approximately 36 percent of the variation in miles per gallon is explained by the model. The next section will add the weight of the cars as a predictor, as well as the interaction between car type and weight.

Model 2

Below is displayed the output from the regressing miles per gallon on manual/automatic car type and the weight of a car.

```
##                  Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)    31.416055  3.0201093 10.402291 4.001043e-11
## factor(am)1    14.878423  4.2640422  3.489277 1.621034e-03
## wt             -3.785908  0.7856478 -4.818836 4.551182e-05
## factor(am)1:wt -5.298360  1.4446993 -3.667449 1.017148e-03
```

First of all, there is a significant main effect of car type on miles per gallon, β = 14.88, $t(28)$ = 3.49, $p <$ .01. The intercept for automatic cars is 31.41, while the coefficient for manual cars is 14.88 turning the intercept of manual cars to 46.29. This means that, in the case of a car with zero weight, it would have a fuel efficiency of 31.41 miles per gallon for automatic cars and 46.29 for manual cars. Second, there is a significant main effect of weight on miles per gallon, β = -3.79, $t(28)$ = -4.82, $p <$ .001. The miles per gallon decreases by 3.79 per pound increase in weight for automatic cars as this is the reference category. Third, there is a significant interaction effect of weight and car type on miles per gallon, β = -5.30, $t(28)$ = -3.67,$p <$ .01. This is demonstrated by the difference in slope of automatic and manual cars for the relationship between weight and fuel efficiency. The slope for automatic cars is -3.79 and for manual cars it is -9.08. Finally, in order to see how much variance is explained by the model (including car type, weight and the interaction between car type and weight to predict the miles per gallon of a car), one could take a glance at the R^2. The model has the following R^2 = 0.83. This means that approximately 83% of the variance is explained by the model.

Model selection

Although all the predictors included in the last model are significant, one could ask whether this more complicated model is preferred to the simpler model with only car type. The widely endorsed Ockham's razor, for instance, attaches much importance to parsimony. In order to test this, we could use nested model testing. The following output displays the results of comparing the first and the second model.

```
## Analysis of Variance Table
##
## Model 1: mpg ~ factor(am)
## Model 2: mpg ~ factor(am) * wt
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     28 188.01  2    532.89 39.682 6.733e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The F test is highly significant, meaning that the second model is preferred to the first model, $F(2, 28)$ = 39.68, $p <$ .001. The second model will therefore be used to predict miles per gallon of all sorts of different cars in the future.

Diagnotics

The following section will display the residual plots and its interpretation.

Above is displayed the residual plot for the second model with the 2 main effects and the interaction effect. In this plot the predicted values are to be found on the x axis while the residuals are plotted on the y axis. Most values are evenly and randomly distributed above and below the red abline. This firstly suggests that a linear trend fits the data reasonably well. In additon, the residuals don't seem toh ave a funnel shape, suggesting that there is no indication of heteroscedasticity. The observation of the Fiat128 in the top right looks slightly like an outlier. In order to verify this, the studentized residuals have been calculated. Studentized residuals with a value over 3 are possible outliers. The observation however has a studentized residual of 2.7, suggesting this is no outlier. All in all, the residual plot does not suggest that there are any major problems with the model.

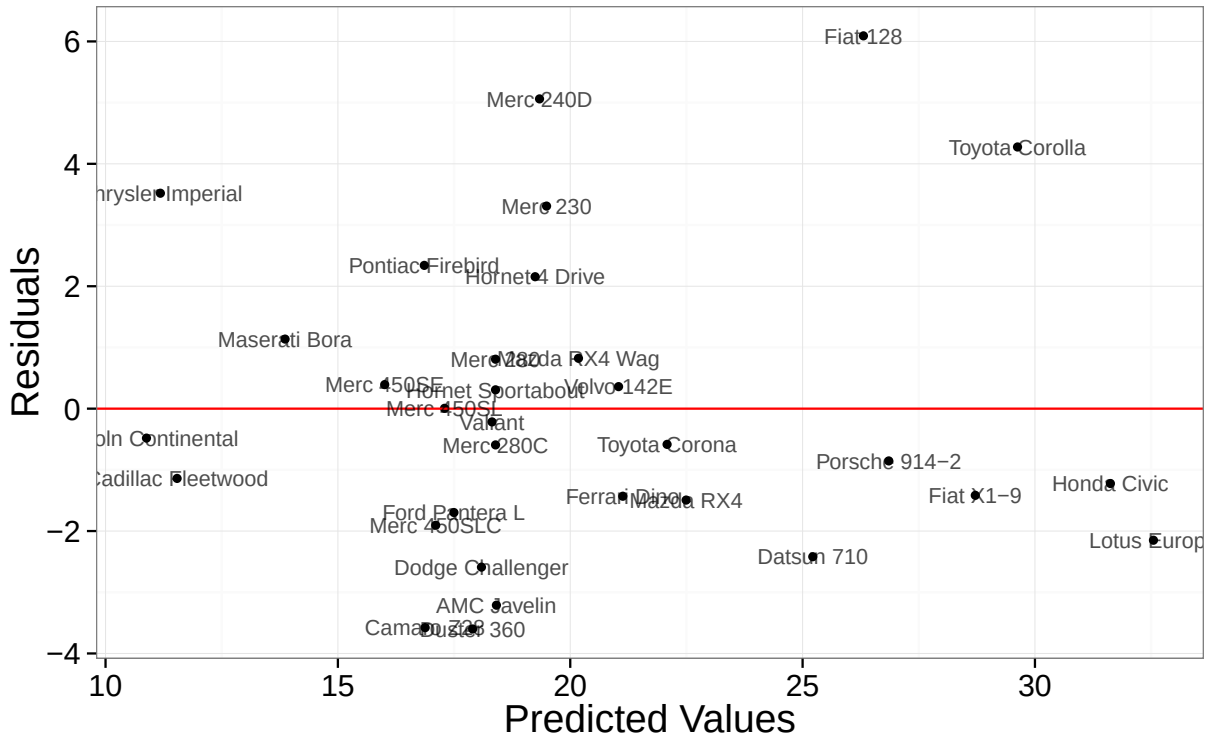


Figure 3

Conclusion

All in all, my final model for predicting the miles per gallon of cars includes the predictors of car type, weight and the interaction of these two variables. This model explains approximately 83% of the variation in the miles per gallon of all the different cars. By all means, the limitations of this model should be noted. Although 83% of the variation explained is high, the researchers could obviously have tested more or different models. However, a model with only 3 predictors that explains that much variance in the data is considered both powerful and parsimonious. I am therefore of the opinion that this model will effectively predict the miles per gallon of future cars as well as providing potential researchers who are interested in the predictions of miles per gallon of cars with an easily interpretable model.