Regression Models Course Project

Huascar Duarte

September 20th, 2014

Executive Summary

Famous automobile magazine Motor Trend has comissioned us to explore the relationship between a set of variables and fuel consumption mesured in miles per gallon. They are particularly interested in the following two questions:

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

Our study confirms that cars with manual transmission are more economical than cars with automatic transmission, but economy varies with the car's power.

Data

The `mtcars` dataset provided in R's `datasets` library will be used to explore the relaitonship berween fuel consumption and other variables.

The data in this dataset was extracted from the 1974 *Motor Trend* US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models), providing the information below, some o which we will not use in our analysis, since they are not directly related to fuel consumption:

1. **mpg** - Miles/(US) gallon
2. cyl - Number of cylinders *(not used)*
3. **disp** - Displacement (cu.in.), i.e., total cubic capacity of cylinders
4. **hp** - Gross horsepower
5. drat - Rear axle ratio *(not used)*
6. **wt** - Weight (lb/1000)
7. qsec - 1/4 mile time *(not used)*
8. vs - Disposition of cylinders (V or Straight) *(not used)*
9. **am** - Transmission (0 = automatic, 1 = manual)
10. gear - Number of forward gears *(not used)*
11. carb - Number of carburetors *(not used)*

We will use a subset of the data with the variables we are interested in the data frame `cars`.

Data Exploration

As we can see in figure 1, there appears to exit a linear relationship between `mpg` and the remaining variables.

Models

Simple model

As shown in figure 2, adopting a simple model considering only the transmission type, we can say, with 95% confidence, that cars with automatic transmissions have a fuel consumption between **14.9** and **19.4 mpg**, with a mean of **17.1 mpg**, while manual transmission cars have a fuel consumption between **21.6** and **27.2 mpg**, with a mean of **24.4 mpg**, i.e.:

> We are 95% confident that cars with manual transmission are between **39.7%** and **45.6%** more economical than cars with automatic transmission, with a difference between **6.8 mpg** and **7.7 mpg** in economy.

Multivariable Models

We will now attempt to analyse the influence of the other variables in the estimation of mpg. Figure 3 analyses how a liner model with all variables adjusts to the data. We can see that the `disp` variable has a high p value, meaning that it does not correlate well with our data, while the other two variables (`hp` and `wt`) present p values below 0.05 indicating a good correlation. We can also see that the coefficient for manual cars (am1) is higher than the coefficient for automatic cars (am0), meaning that manual cars will have higher mpg than automatic cars.

Figure 4 analyses a model without the `disp` variable, which presents a nice fit, with low p values and high R2.

Figure 5 shows a third model without the `wt` variable, keeping only the `hp` variable. This model presents the best p value of all with a R2 very close to the previous model.

The result of the `anova` function comparing the 3 models shown in figure 7 present P values that indicate that the third model (f3) is the best, which is nice because it has only one regressor (the `hp` variable) which makes the analysis much easier.

In figure 6 we plot the residuals diagnostics, which indicates a good fit for our model `f3`. Maybe we should consider the Maseratti Bora an outlier and leave it out from the model, but we chose to keep it.

Conclusion with selected model

As we can see in figure 8, our fitted model indicates that there is an almost constant variation of 5.3 mpg between cars with manual and automatic transmission, and that at 63 and 222 hp's our 95% confidence intervals overlap. So we can state that:

> We are 95% confident that cars with manual transmission, with power between 63 and 222 HP's, are more economical than cars with automatic transmission. On average manual cars are **5.3 mpg** more economical than automatic ones. Fuel economy between manual and automatic cars varyies between **22.9%** and **39.3%** for cars from 63 to 222 hp's.

Appendix

Fig. 1: Pairs with red dots for manual and green dots for automatic

```
pairs(cars, col = 2 + (cars$am == 0))
```

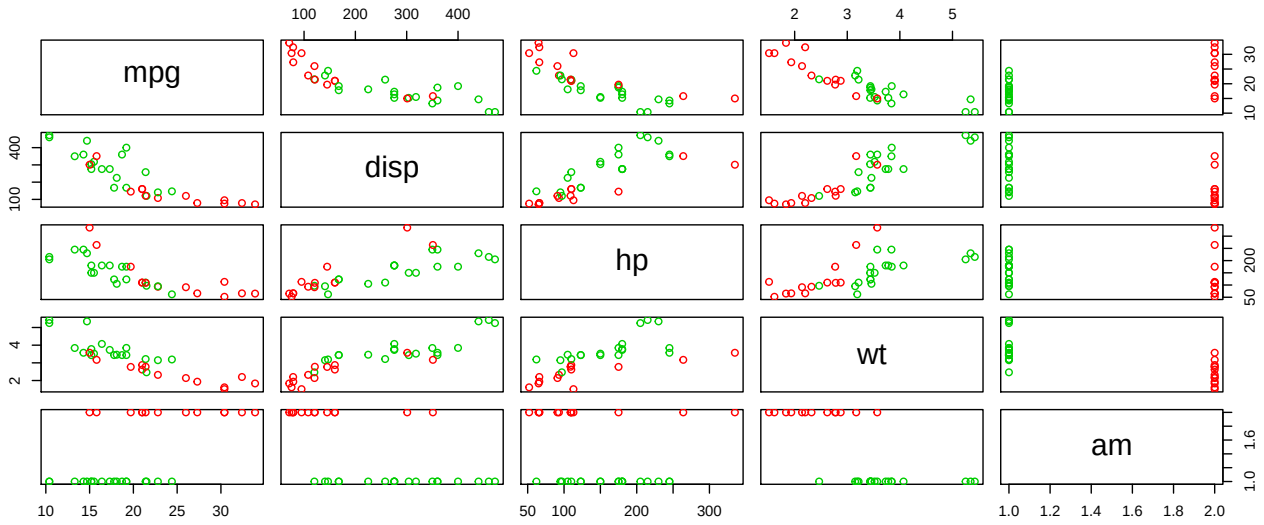

Fig. 2: Boxplot with 95% confidence intervals

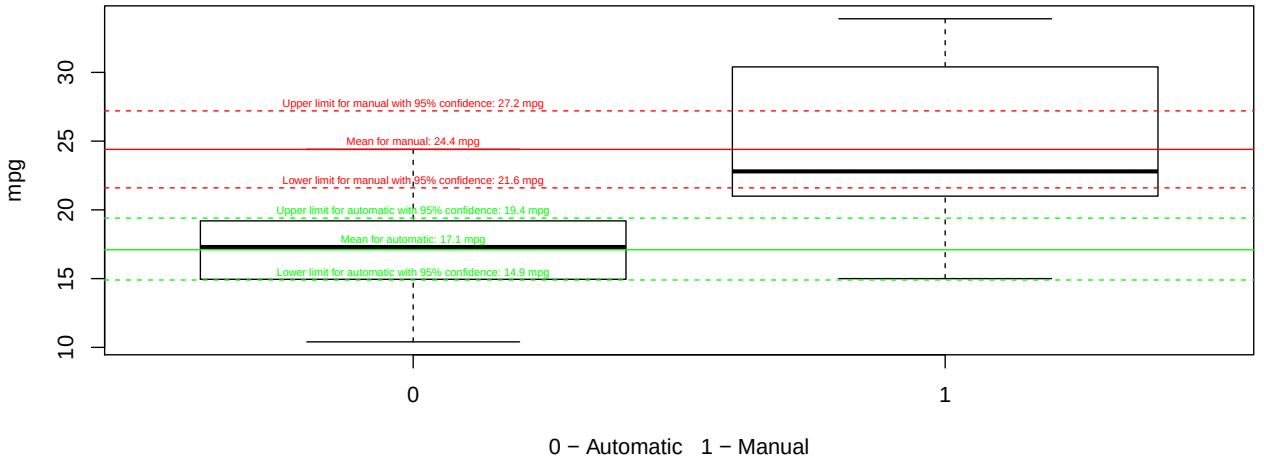

Fig. 3: Linear model f1 with all variables

```
f1<-lm(mpg ~ . - 1, cars)
summary(f1)$coefficients
```

```
##       Estimate Std. Error t value  Pr(>|t|)
## disp  0.002489    0.01038  0.2399 8.122e-01
## hp   -0.039323    0.01243 -3.1627 3.842e-03
## wt   -3.046747    1.15712 -2.6330 1.383e-02
## am0  34.209443    2.82283 12.1189 1.980e-12
## am1  36.368714    2.12202 17.1387 4.902e-16
```

```
summary(f1)$r.squared
```

```
## [1] 0.9872
```

Fig. 4: Linear model f2 without the `disp` variable

```r
f2<-lm(mpg ~ hp + wt + am - 1, cars)
summary(f2)$coefficients
```

```
##     Estimate Std. Error t value  Pr(>|t|)
## hp  -0.03748   0.009605  -3.902 5.464e-04
## wt  -2.87858   0.904971  -3.181 3.574e-03
## am0 34.00288   2.642659  12.867 2.824e-13
## am1 36.08659   1.736338  20.783 1.478e-18
```

```r
summary(f2)$r.squared
```

```
## [1] 0.9872
```

Fig. 5: Model f3 with only the hp variable

```r
f3<-lm(mpg ~ hp + am - 1, cars)
summary(f3)$coefficients
```

```
##     Estimate Std. Error t value  Pr(>|t|)
## hp  -0.05889   0.007857  -7.495 2.920e-08
## am0 26.58491   1.425094  18.655 1.074e-17
## am1 31.86200   1.282279  24.848 4.252e-21
```

```r
summary(f3)$r.squared
```

```
## [1] 0.9825
```

Fig. 6: Residuals for the selected model

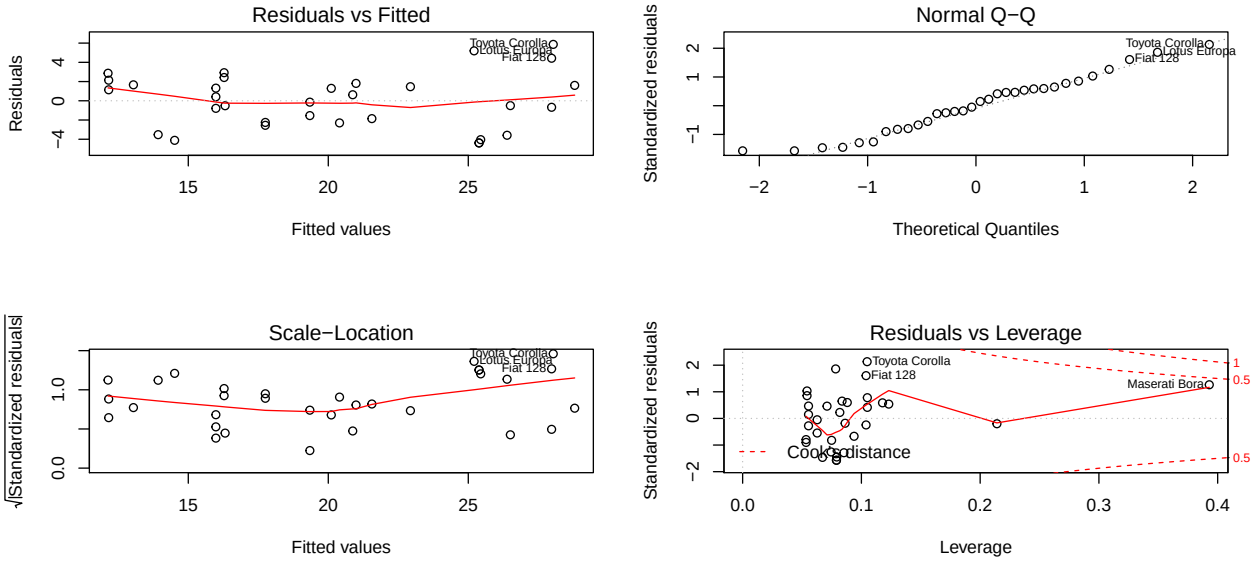

Fig. 7: Anova function results for the models

```r
anova(f1, f2, f3)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ (disp + hp + wt + am) - 1
## Model 2: mpg ~ hp + wt + am - 1
## Model 3: mpg ~ hp + am - 1
##   Res.Df RSS Df Sum of Sq    F Pr(>F)
## 1     27 180
## 2     28 180 -1      -0.4 0.06 0.8122
## 3     29 245 -1     -65.1 9.78 0.0042 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Fig. 8: Mpg x hp plot for manual (red) and automatic (green) cars with 95% confidence interval