Motor Trend Analysis

Iryna Garbuz

26 october 2014

Executive Summary

The regression analysis provided below shows changes of Miles per Gallon (MPG) to Transmission Type with different number of variables. In the simplest model, mpg over transmission types, shows that the manual transmission is 7.25 mpg better than automatic transmission. Taking cyliner, weight and horsepower into account, the multivariate regression model indicates that the manual transmission is near 1.81 mpg better than the automatic transmission while the goodness of fit has been reached 84%.

Data Processing

We are using mtcat dataset to proceed our regression analysis. You can see Figure 2 in Apendix showing the matrix of scatterplots for this dataset. Also, figure 1 shows the density distributuion of MPG by Transmittion type. First, convert numeric values to factor values.

Regression Model

Firstly, we built a simple model to directly measure the apparent influence of transmission on MPG. The first non-intercept coefficient of the model, 7.245, suggested an increase of 7.245 miles per gallon when switching from automatic to manual transmission (with p-value < 0.0001). However, the low R- squared value (0.3385) of the model indicated that transmission alone is not a very good predictor of the variation in MPG.

```
fit <- lm(mpg~am, mtcars)
summary(fit)$coeff
```

```
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept)   17.147      1.125  15.247 1.134e-15
## amManual       7.245      1.764   4.106 2.850e-04
```

We build several more models with different number of parameters to find the best one. I added the model with all parameters to compare it with others:

```
fitall <- lm(mpg~.,mtcars)
fitbest <- step(fitall, direction ="both")
```

```
anova(fitbest, fit, fitall)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ cyl + hp + wt + am
## Model 2: mpg ~ am
## Model 3: mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb
##   Res.Df RSS Df Sum of Sq     F  Pr(>F)
```



```
## 1     26 151
## 2     30 721 -4      -570 17.75 1.5e-05 ***
## 3     15 120 15       600  4.99  0.0018 **
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Results shows that the best moel includes such paramteres-transmission, cyliner, weight and horsepower into account. Looking at the above results, the p-value obtained is highly significant and we reject the null hypothesis that the confounder variables cyl, hp and wt don't contribute to the accuracy of the model. Let's see what influence of transmission on mpg is shown:

```
summary(fitbest)
```

```
##
## Call:
## lm(formula = mpg ~ cyl + hp + wt + am, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.939 -1.256 -0.401  1.125  5.051
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  33.7083     2.6049   12.94  7.7e-13 ***
## cyl6         -3.0313     1.4073   -2.15   0.0407 *
## cyl8         -2.1637     2.2843   -0.95   0.3523
## hp           -0.0321     0.0137   -2.35   0.0269 *
## wt           -2.4968     0.8856   -2.82   0.0091 **
## amManual      1.8092     1.3963    1.30   0.2065
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.41 on 26 degrees of freedom
## Multiple R-squared:  0.866,  Adjusted R-squared:  0.84
## F-statistic: 33.6 on 5 and 26 DF,  p-value: 1.51e-10
```

The multivariate regression model indicates that the manual transmission is near 1.81 mpg better than the automatic transmission while the goodness of fit has been reached 84%.

Model Diagnostics

Figure 3 in Apendix shows residuals variation and different aspects of model fit. This helped us in examining the residuals and finding leverage points to find any potential problems with the model.

Statistical Inference

In this section, we perform a t-test on the two subsets of mpg data: manual and automatic transmission assuming that the transmission data has a normal distribution and tests the null hypothesis that they come from the same distribution. Based on the t-test results, we reject the null hypothesis that the mpg distributions for manual and automatic transmissions are the same.


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

```
##
##  Welch Two Sample t-test
##
## data:  mpg by am
## t = -3.767, df = 18.33, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.28  -3.21
## sample estimates:
```

```
## sample estimates:
## mean in group Automatic    mean in group Manual
##                   17.15                   24.39
```

Conclusions

Based on the analysis done in this project, we can conclude that:

- Cars with Manual transmission get 1.8 more miles per gallon compared to cars with Automatic transmission. (1.8 adjusted for hp, cyl, and wt).
- mpg will decrease by 2.5 for every 1000 lb increase in wt.
- mpg decreases negligibly (only 0.32) with every increase of 10 in hp.

Apendix

```
## Package 'sm', version 2.2-5.4: type help(sm) for summary information
```

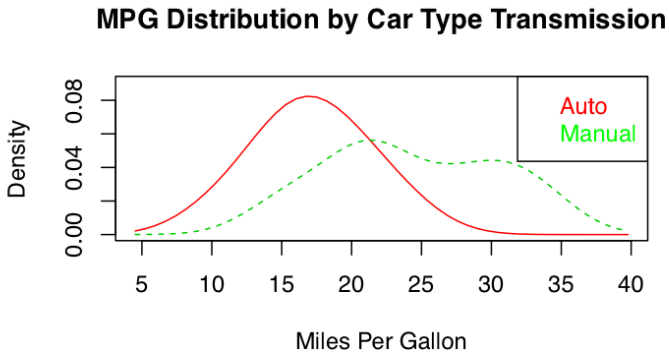
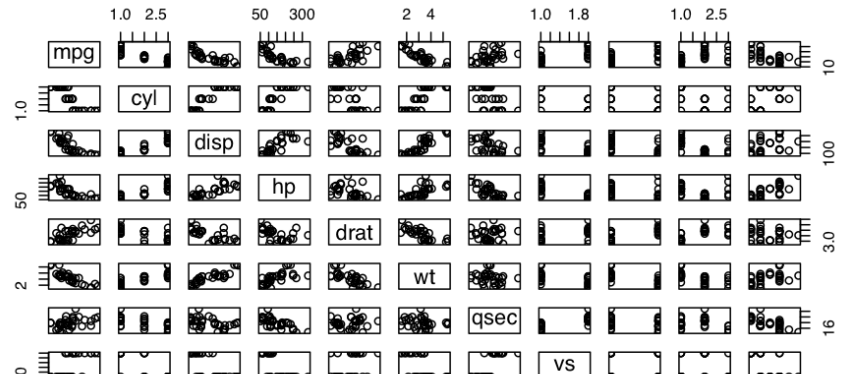


Figure 1: A Chart showing MPG distribution by Transmission Type


```
pairs(mpg ~ ., data = mtcars)
```



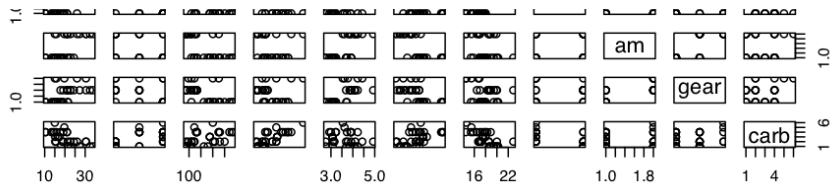


Figure 1: A matrix of scatterplots for mtcars dataset

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

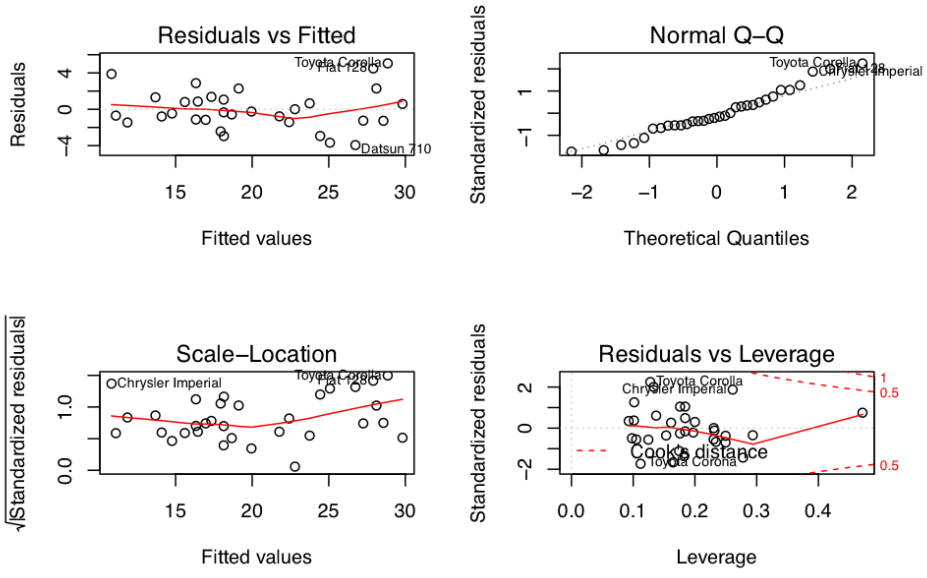



Figure 2: Residuals variation for Best Regression Model

```r
boxplot(mpg ~ am, data = mtcars, col = (c("red","blue")), ylab = "Miles Per Gallon", xlab = "Transmissi
```

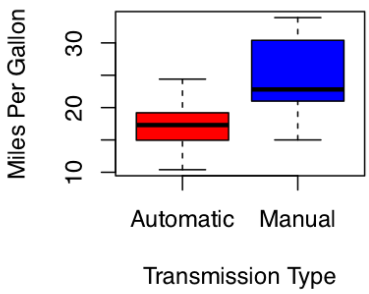


Figure 4: Boxplot of MPG by Transmission Type