Motor Trend Data Analysis (Regression Model)

unkame

Saturday, May 23, 2015

Executive Summary

The aims of this report was to analyze the relationship between miles per gallon (MPG) and a set of variables, to understand two main questions: 1. Whether an automatic or manual transmission is better for MPG 2. Quantify the MPG difference between automatic and manual transmissions

Using exploratory data analysis and basic linear model, we could conclude that manual transmission had better MPG usage than that of automatic one, by a average of 7.25 MPG higher. While using multivariate linear model, we understood that other confounding variables of weight (lb/1000) and 1/4 mile time contributed significant impacts on the difference on MPG apart from types of transmission. In that case the manual transmission only had 2.94 MPG better than automatic transmission if the two variables were constant.

DataSet used

Data used in this report are the data set "mtcars", which has 32 observations of 11 variables. Preliminary codes required would be:

```
data(mtcars)       ## load data "mtcars"
```

Notes that the attribute am (transmission), value 1 represents manual transmission while 0 is automatic transmssion.

Exploratory Data Analysis

This study focused on the relationship between MPG and Types of Transmission (AM). First of all the data MPG were checked if it was normally distributed, which is the fundamental assumption of linear regression. The result was in Appendix Figure 1, which MPG was approximately normal without apparent outliers in data.

Then a boxplox analysis was used to illustrate the MPG verus AM and the result was in Appendix Figure 2. Statistically, the graph shown that manual transmission had higher MPG than automatic one.

Inference of mileages of automatic vs manual transmission

Finding the mean of both types of transmissions:

```
##   am      mpg
## 1  0 17.14737
## 2  1 24.39231
```

The mean of MPG of manual transmission cars is 7.25 MPGs higher than that of automatic transmission cars. And the t-test was:

```
##
##  Welch Two Sample t-test
##
## data:  autoData$mpg and manualData$mpg
## t = -3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.280194  -3.209684
## sample estimates:
## mean of x mean of y
##  17.14737  24.39231
```

The p-value was 0.001374, which was low enough (< 0.05) to reject the null hupothesis. They were from different populations, and so that the maual transmission had higher MPG than that of automatic one. But multiple linear regression analysis was needed as the above result was assumed that on all other characteristics of auto and manual cars were the same.

Basic Linear Regression

Here was the basic linear model using mpg and am :

```
fit <- lm(mpg~am, data = mtcars)
summary(fit)          ## Lengthly Result hidden, see explanation below
```

The estimate of am was 7.25, which meant the manual transmission had 7.25 MPG more than automatic transmission. While multiple R squared is 0.3598, meaning that this model had just 35.98% of the variability which was low. So multivariate regression analysis would be used to quantify the mpg differences between automatic and manual transmission in better way.

Multivariate Regression Analysis

To achieve multivariate analysis, here function Step() was used to find out the best variable combinations describing MPG using AIC algorithm:

```
stepModel = step(lm(data = mtcars, mpg ~ .), trace=0, steps=10000)
summary(stepModel)     ## Lengthly Result hidden, see explanation below
```

This showed that in adddition to transmission, wt (weight lb/1000) as well as qsec (1/4 mile time) had high relation to explaining the variation in mpg. The adjusted R-squared was ~83% which means that the model explaint ~83% of the variation in mpg indicating it was a robust and highly predictive model.

Final model to quantify the mpg difference between transmissions

Using variance analysis "anova" to compare basic model and final model obtained in :

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + wt + qsec
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
```

```
## 1     30 720.90
## 2     28 169.29  2    551.61 45.618 1.55e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

This model captured 83% of the overall variation in mpg. With a p-value of 1.55e-09, null hypothesis were rejected and this multivariate model was significantly different basic linear regression model.

- Residual Analysis & Diagnostic - Before reporting the details of the model, the residuals diagnostocs (Appendix Figure 3) were examined to check non-normality and heteroskedasticity. The plots showed the normality and there was no evidence of heteroskedasticity.

Here was the final model that best fits the relationship of MPG in data mtcars:

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)  9.617781  6.9595930  1.381946 1.779152e-01
## am           2.935837  1.4109045  2.080819 4.671551e-02
## wt          -3.916504  0.7112016 -5.506882 6.952711e-06
## qsec         1.225886  0.2886696  4.246676 2.161737e-04
```

The estimate of am showed manual transmission had 2.94 MPG more than automatic transmission on average, without measuring the other 2 variables wt and qsec. The final model would be (ignore gaussian noise):

```
## MPG = 9.62 + 2.94*am - 3.92*wt + 1.23*qsec
```

Conclusion

From the study of data mtcars, through exploratory data analysis manual transmission has better MPG, with 7.25 MPGs higher than that of automatic transmission, based on rejecting null hupothesis by P-value of 0.001374.

A multivariate linear regression model [MPG = 9.62 + 2.94am - 3.92wt + 1.23*qsec] including variables am (transmission), wt (weight lb/1000) and qsec (1/4 mile time) was most suitable combination to quantify the MPG difference between automatic and manual transmissions. This model explaint 83% of the variance in MPG. And typically on average the manual transmission cars have 2.94 MPGs more than automatic transmission cars if wt and qsec conditions of both transmission types of cars were the same.

Appendix

Figure 1 - Histogram & Density Plot of MPG

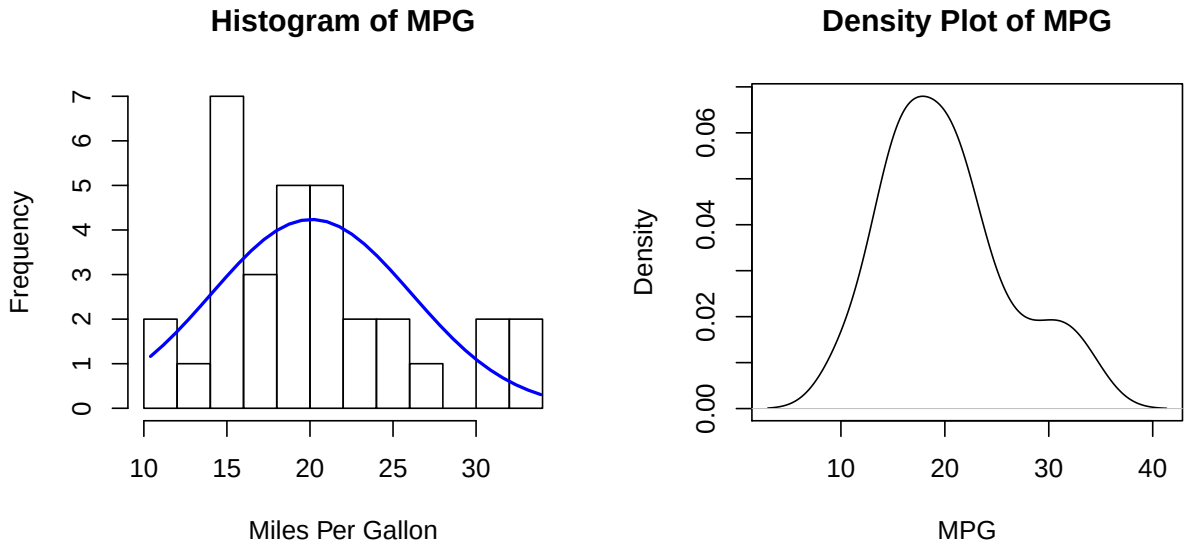

Figure 2 - Boxplot of MPG vs Types of Transmission

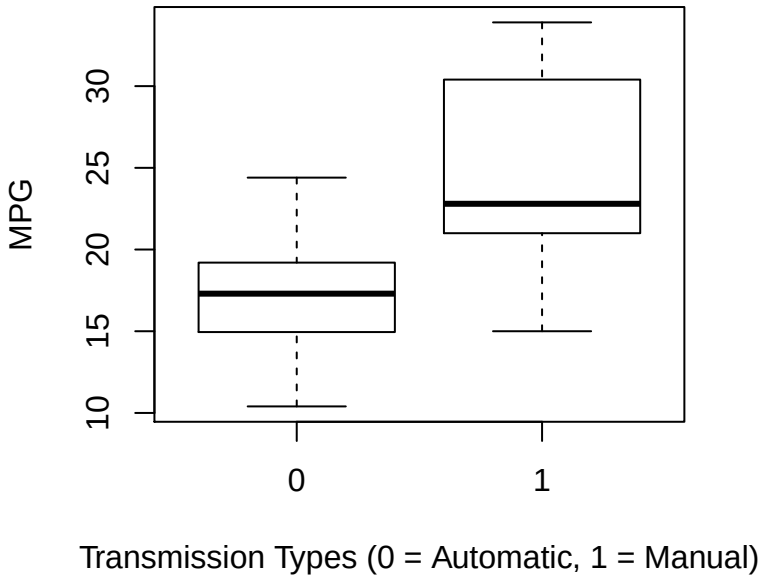

Figure 3 - Residual Diagnostic