Regression Model Peer Assessment

Lai Yiu Ming, Tom

Sunday, June 14, 2015

Overview

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

1. Is an automatic or manual transmission better for MPG?
2. Quantify the MPG difference between automatic and manual transmissions?

Basic summary of data

The dataset mtcars 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). Manual transmission/automatic transmission is defined in the column "AM" with 0 = automatic and 1 = manual. Mile per gallon data in recorded in column "MPG".

In Appendix A1 Pairwise scatter plot, there is a pairwise scatter plot for selected features for reader to explore the dataset.

Study and result

Relationship between MPG and automatic/manual transmission

In this section, we use hypothesis test to test if manual transmission is difference than automatic transmission in term of MPG.

```
t.test(mpg~am, data=mtcars, alternative="two.sided")$p.value
```

```
## [1] 0.001373638
```

From the result, as P-value = 0.001374 < 0.025, hence the null hypothesis is rejected. Manual transmission has mean of mpg (mile per gallons) of 24.39231 which is higher than that of automatic tranmission, which implies that manual transmission is better in term of MPG.

Quantify the MPG difference between automatic and manual transmissions

In this section, we fit a linear regression model which mpg is the outcome that include automatic/manual transmission as a factor variable. There are 2 difference models we can apply:

1. Without intercept: $MPG = AM\beta_0 + \epsilon$

2. With intercept: $MPG = \beta_0 + \mathsf{AM}\beta_1 + \epsilon$

There is no need for extra dummy variables because there are only 2 levels for factor variable "AM".

For the model comparison between first option and second option, refer to Appendix A2 Linear regression model comparison.

In this study, we decide to choose the second option because regression line fits better. A linear regression model is created.

```
model <- lm(mpg ~ am, data=mtcars)
summary(model)
```

```
##
## Call:
## lm(formula = mpg ~ am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.147      1.125  15.247 1.13e-15 ***
## am             7.245      1.764   4.106 0.000285 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.902 on 30 degrees of freedom
## Multiple R-squared:  0.3598, Adjusted R-squared:  0.3385
## F-statistic: 16.86 on 1 and 30 DF,  p-value: 0.000285
```

Residue plots are created, and by inspection, the residue follow normal distribution. Hence, the model should be a good model fit without missing any important feature. To see the residue plot, refer to Appendix A3 Residue plots.

From the model:

$\beta_0 = 17.147$ - MPG for automatic transmission.

$\beta_1 = 7.245$ - change in MPG forchanging from automatic transmission to manual transmission.

ϵ - error term

By inspecting the β_1, MPG for automatic tranmission is 7.245 miles/gallon less than that for manual transmission.

Conclusion

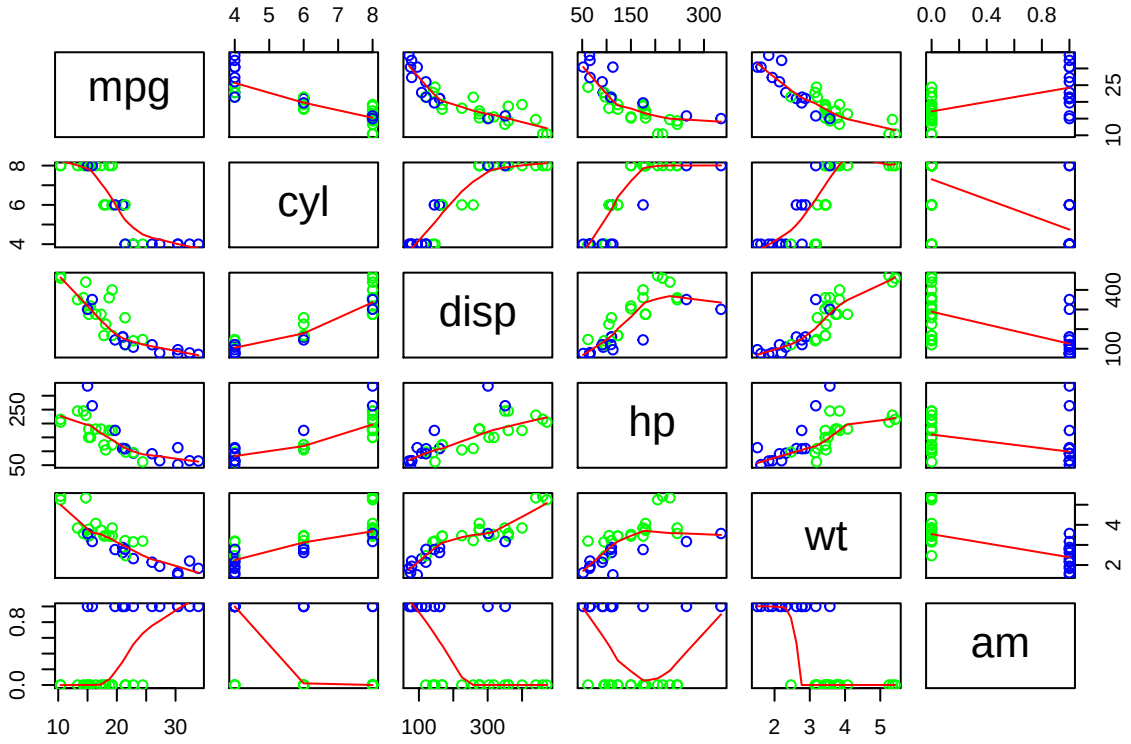
In this study, we have use hypothesis test to show that the manual transmission perform better than automatic transmission in term of MPG.

We also fit a linear regression model and it suggested that for 1 gallon, manual transmission run 7.245 miles more compare to automatic transmission.

Appendix

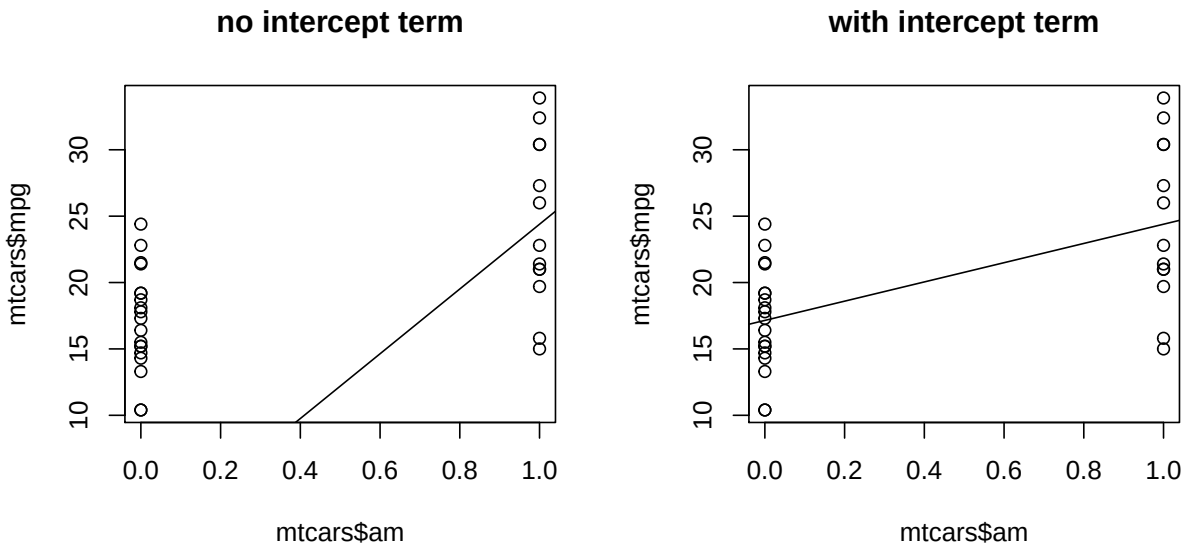
A1 Pairwise scatter plot

```
require(stats)
require(graphics)
pairs(mtcars[,c(1,2,3,4,6,9)], panel=panel.smooth, col=ifelse(mtcars$am==1, 'BLUE', 'GREEN'))
```



A2 Linear regression model comparison

```
par(mfrow=c(1,2))
model <- lm(mpg ~ am-1, data=mtcars)
plot(mtcars$am, mtcars$mpg, main="no intercept term")
abline(model)
model <- lm(mpg ~ am, data=mtcars)
plot(mtcars$am, mtcars$mpg, main="with intercept term")
abline(model)
```



A3 Residue plots

```
par(mfrow=c(2,2))
modelFit <- lm(mpg ~ am, data=mtcars)
plot(modelFit)
```