Regression Models Course Project

YE

may, 15th 2016

Executive Summary

The goal of this project is to explore the relationship between a set of variables and Miles per Galon (MPG) in the MTCARS dataset.

The dataset

Format

A data frame with 32 observations on 11 variables.

Source

Henderson and Velleman (1981), Building multiple regression models interactively. Biometrics, 37, 391–411.

```
head(mtcars)
```

```
##                   mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

ANALYSIS

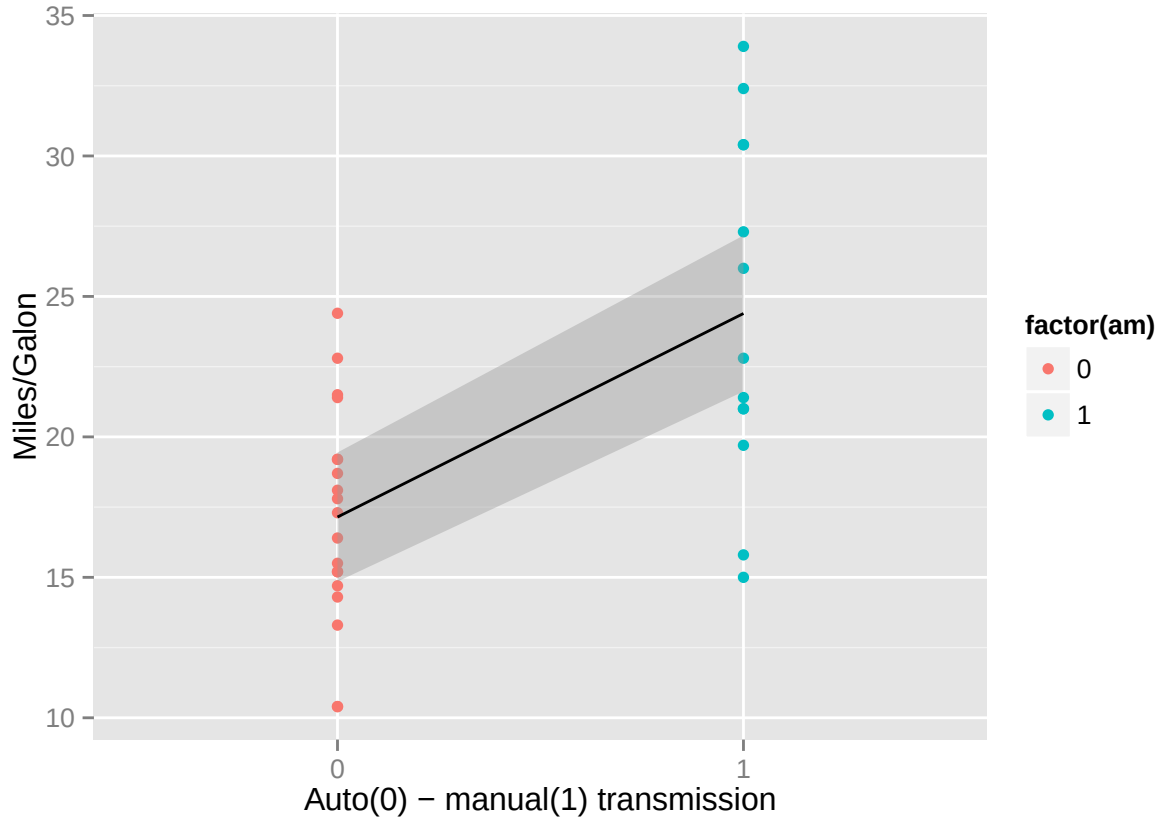
1. Is an automatic or manual transmission better for MPG ?

EXPLORATORY DATA ANALYSIS

```
    data(mtcars)
    require(ggplot2)
```

```
## Loading required package: ggplot2
```

```
    g = ggplot(data=mtcars,aes(y=mpg,x=factor(am), colour=factor(am) ) )
    g = g+ geom_point()
    g = g + geom_smooth(method='lm', colour="black", aes(group=0))
    g = g + xlab("Auto(0) - manual(1) transmission") + ylab("Miles/Galon")
    g
```




```
    aggregate(mpg~am, data = mtcars, mean)
```

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

The Manual MPG is higher than the Auto MPG, we are going to investigate if this difference is important and find if the others variables have an impact.

By fitting a linear regression, we have the following:

```
    data(mtcars)
    fit <- lm(formula= mpg~as.factor(am),data=mtcars)
    summary(fit)
```

```
##
## Call:
## lm(formula = mpg ~ as.factor(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 ***
## as.factor(am)1    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
```

Conclusion:

- P-value = 0.000285, we reject the null hypothesis that is, we declare a relationship to exist between MPG and AM (transmission).
- The manual transmission (AM=1) spends 17.147+7.245 = 24.392 Galon each mile. The automatic Transmission spends 17.147 Galons every mile. **We can say that Manual transmission is better for MPG**.

2. Quantify the MPG difference between automatic and manual transmissions

A look to the R-Squared above, 0.3598 (= 36%), shows that, this regression has explained only 36% of the variability of the MPG. We need to fit a multiple linear model to explain the difference (7.245) between automatic and manual transmission.

Multiple linear regression

```
data(mtcars)
summary(lm(formula= mpg~.,data=mtcars))
```

```
##
## Call:
## lm(formula = mpg ~ ., data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4506 -1.6044 -0.1196  1.2193  4.6271
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.30337   18.71788   0.657   0.5181
## cyl         -0.11144    1.04502  -0.107   0.9161
## disp         0.01334    0.01786   0.747   0.4635
## hp          -0.02148    0.02177  -0.987   0.3350
## drat         0.78711    1.63537   0.481   0.6353
## wt          -3.71530    1.89441  -1.961   0.0633 .
## qsec         0.82104    0.73084   1.123   0.2739
## vs           0.31776    2.10451   0.151   0.8814
## am           2.52023    2.05665   1.225   0.2340
## gear         0.65541    1.49326   0.439   0.6652
## carb        -0.19942    0.82875  -0.241   0.8122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.65 on 21 degrees of freedom
## Multiple R-squared:  0.869,  Adjusted R-squared:  0.8066
## F-statistic: 13.93 on 10 and 21 DF,  p-value: 3.793e-07
```

Through this regression we can seen that: * wt (p-value=0.0633), qsec (p-value=0.2739) and am(p-value=0.2340) have the most fewest p-values, so we can fit another model with that only 3 variables:

– ### Final Multiple linear regression

```
    data(mtcars)
    summary(lm(formula= mpg~as.factor(am)+qsec+wt,data=mtcars))
```

```
##
## Call:
## lm(formula = mpg ~ as.factor(am) + qsec + wt, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4811 -1.5555 -0.7257  1.4110  4.6610
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)      9.6178     6.9596   1.382 0.177915
## as.factor(am)1   2.9358     1.4109   2.081 0.046716 *
## qsec             1.2259     0.2887   4.247 0.000216 ***
## wt              -3.9165     0.7112  -5.507 6.95e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.459 on 28 degrees of freedom
## Multiple R-squared:  0.8497, Adjusted R-squared:  0.8336
## F-statistic: 52.75 on 3 and 28 DF,  p-value: 1.21e-11
```

Conclusion:

*According to this final model, we have R-Squared = 0.8497, so 85% of the variability of MPG is explained through this regression.

***The difference between Automatic and Manual transmission is 2.9358 when we include qsec and wt varaiables**.