RegressionModels

xiao

August 16, 2015

executive summary

You work for Motor Trend, a magazine about the automobile industry. Looking at a data set of a collection of cars, they are interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome). They are particularly interested in the following two questions:

Is an automatic or manual transmission better for MPG

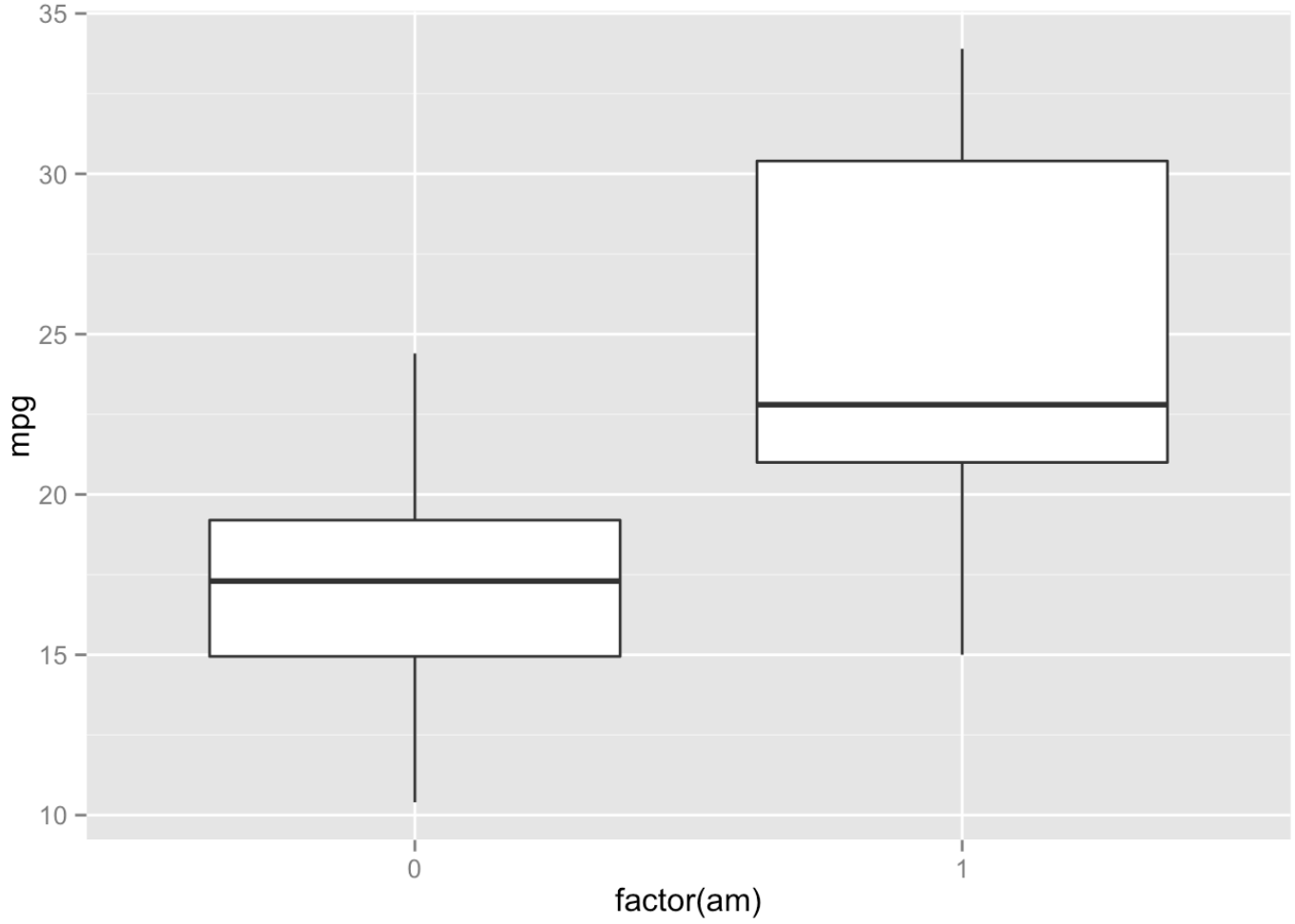
Quantify the MPG difference between automatic and manual transmissions

From the following analysis, we can see manual (1) is better for MPG, from the exploratory analysis, the boxplot and the t-stats both shows that the difference is more than one standard deviation.

Some Exploratory Analysis

1. Is an automatic or manual transmission better for MPG

```
library(ggplot2)
qplot(factor(am), mpg, data = mtcars, geom = "boxplot")
```



2. Quantify the MPG difference between automatic and manual transmissions

```
myMean = tapply(mtcars$mpg, mtcars$am, mean)
myStd = tapply(mtcars$mpg, mtcars$am, sd)
(myMean[1]-myMean[2])/myStd[1]
```

```
##         0
## -1.889672
```

```
(myMean[1]-myMean[2])/myStd[2]
```

```
##         0
## -1.174886
```

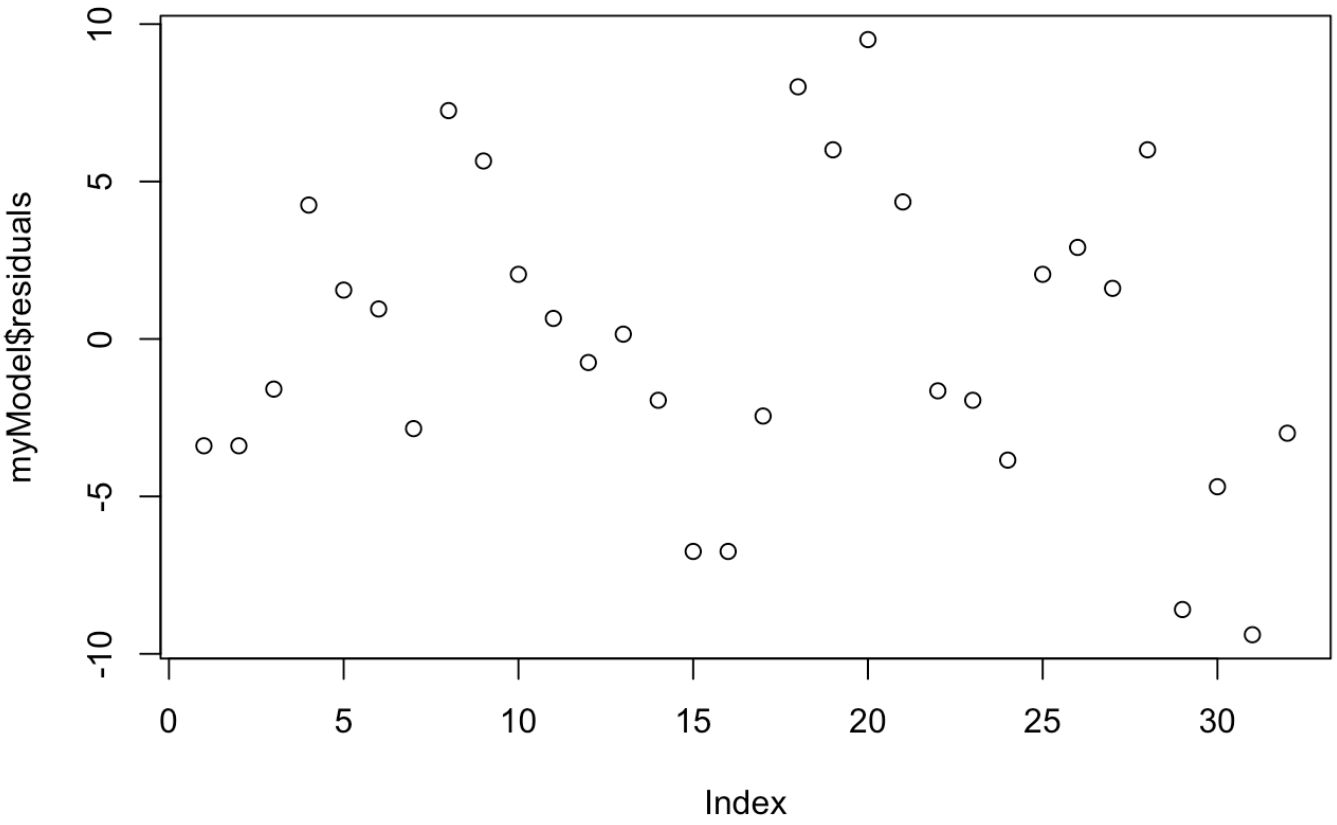
Regression Analysis

From the following result, we can see that the regression shows am (whether the car is automatic or manual) is a significant predictor of mpg. The P-Value for factor(am)1 is smaller than 5%. The residual of this result seems random, there isn't an obvious trend that am fails to predict.

```
myModel = lm(formula = mpg~factor(am), data=mtcars)
summary(myModel)
```

```
##
## Call:
## lm(formula = mpg ~ 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 ***
## 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
```

```
plot(myModel$residuals)
```


myModel$residuals
10
5
0
-5
-10
0
5
10
15
20
25
30
Index