regmodel

Jack

26 December 2015

Context

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"

Question

Take the mtcars data set and write up an analysis to answer their question using regression models and exploratory data analyses.

Your report must be:

Written as a PDF printout of a compiled (using knitr) R markdown document. Brief. Roughly the equivalent of 2 pages or less for the main text. Supporting figures in an appendix can be included up to 5 total pages including the 2 for the main report. The appendix can only include figures. Include a first paragraph executive summary.

Executive Summary: Using the dataset mtcars, I explored the correlation between various variables and the mpg.Wt, cyl, disp and hp were found to be most correlated with mpg.

Based on the boxplot, I found that automatic transmission has a better mpg than manual transmission.

Exploratory Analysis

```
data(mtcars)
names(mtcars)
```

```
##  [1] "mpg"  "cyl"  "disp" "hp"   "drat" "wt"   "qsec" "vs"   "am"   "gear"
## [11] "carb"
```

```
cor(mtcars, mtcars$mpg)[-1,]
```

```
##        cyl       disp         hp       drat         wt       qsec
## -0.8521620 -0.8475514 -0.7761684  0.6811719 -0.8676594  0.4186840
##         vs         am       gear       carb
##  0.6640389  0.5998324  0.4802848 -0.5509251
```

The above helps us to see the correlation between all variables and mpg. We can see that wt, cyl, disp and hp are most correlated with mpg, in decreasing orders of correlation.

Question 1: "Is an automatic or manual transmission better for MPG" Question 2: "Quantify the MPG difference between automatic and manual transmissions"

```
mtcars$am <- as.factor(mtcars$am)
levels(mtcars$am) <- c("Automatic", "Manual")

stepmodel = step(lm(data = mtcars, mpg ~ .),trace=0,steps=10000)
summary(stepmodel)
```

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am, 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
## wt           -3.9165     0.7112  -5.507 6.95e-06 ***
## qsec          1.2259     0.2887   4.247 0.000216 ***
## amManual      2.9358     1.4109   2.081 0.046716 *
## ---
## 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
```

```
model <- lm(mpg~ factor(am):wt + factor(am):qsec,data=mtcars)
summary(model)
```

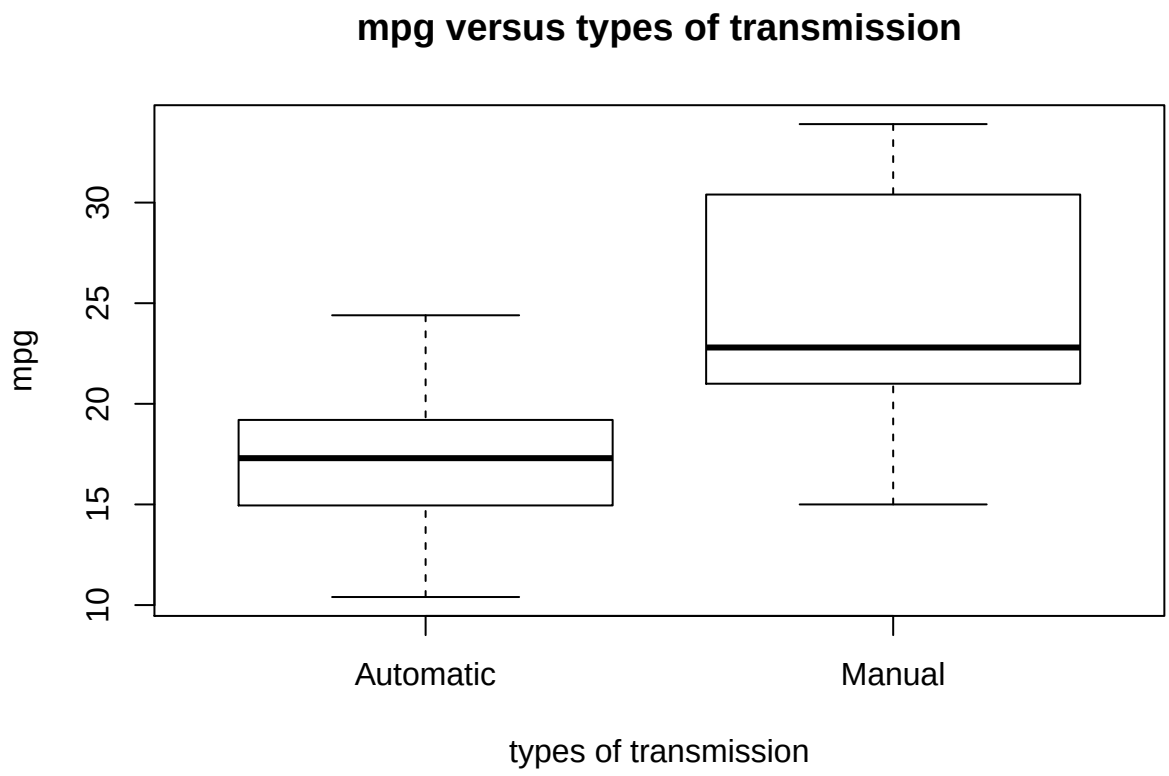
```
##
## Call:
## lm(formula = mpg ~ factor(am):wt + factor(am):qsec, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9361 -1.4017 -0.1551  1.2695  3.8862
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)                13.9692     5.7756   2.419  0.02259 *
## factor(am)Automatic:wt     -3.1759     0.6362  -4.992 3.11e-05 ***
## factor(am)Manual:wt        -6.0992     0.9685  -6.297 9.70e-07 ***
## factor(am)Automatic:qsec    0.8338     0.2602   3.205  0.00346 **
## factor(am)Manual:qsec       1.4464     0.2692   5.373 1.12e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.097 on 27 degrees of freedom
## Multiple R-squared:  0.8946, Adjusted R-squared:  0.879
## F-statistic: 57.28 on 4 and 27 DF,  p-value: 8.424e-13
```

In Appendix A, a boxplot was created showing mpg versus types of transmission. From the boxplot, it does appear that automatic transmission has a better mpg, while manual transmission has a poorer mpg.

From the coeefficients, it can be seen that as the weight of the car increased, the advantage of the automatic transmission over manual transmission decreases, until it may be better to select manual transmission for heavy cars above a certain weight.

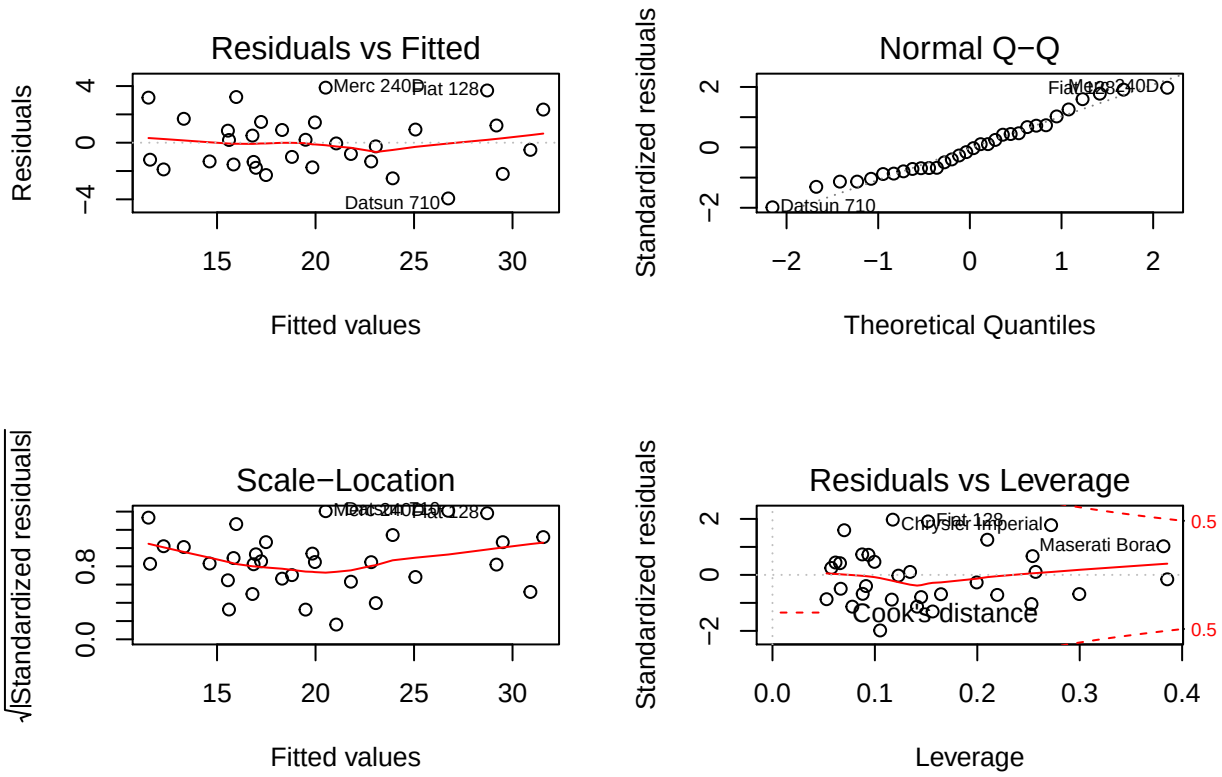
Appendix A:

```r
boxplot(mtcars$mpg ~ mtcars$am, data = mtcars, outpch = 21, ylab="mpg",xlab="types of transmission",mai
```



Appendix B:

```r
par(mfrow=c(2,2))
plot(model)
```


Residuals vs Fitted
Residuals
Fitted values
Merc 240D
Fiat 128
Datsun 710
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Datsun 710
Scale−Location
√|Standardized residuals|
Fitted values
Residuals vs Leverage
Standardized residuals
Leverage
Chrysler Imperial
Maserati Bora
Cook's distance
0.5