

Gonzalo Moreno Course Project

Gonzalo Moreno

Saturday, October 25, 2014

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”

EXECUTIVE SUMMARY

In the next project, is proved that manual transmissions are better for mpg with a mpg difference of 2.935 and a 95% confidence interval from 0.0457 to 5.826.

variables of MTCARS [, 1] mpg Miles/(US) gallon [, 2] cyl Number of cylinders [, 3] disp Displacement (cu.in.) [, 4] hp Gross horsepower [, 5] drat Rear axle ratio [, 6] wt Weight (lb/1000) [, 7] qsec 1/4 mile time [, 8] vs V/S [, 9] am Transmission (0 = automatic, 1 = manual) [,10] gear Number of forward gears [,11] carb Number of carburetors

QUESTION 1

Question 1: is better manual than automatic transmission? Manual is better than automatic transmission.

```
data.automatic<- subset(mtcars, mtcars[,9]==0)
data.manual<- subset(mtcars, mtcars[,9]==1)
```

The boxplot suggests strong evidence of manual is better than automatic transmission. (see appendix fig 1) Furthermore, it seems that manual has a bigger variance compared to automatic transmission. We do a t.test and calculate a 95% confidence interval.

```
t.test(data.manual[1], data.automatic[1], alternative= "greater", paired=FALSE, var.equal=FALSE)

##
## Welch Two Sample t-test
##
## data: data.manual[1] and data.automatic[1]
## t = 3.7671, df = 18.332, p-value = 0.0006868
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  3.913256      Inf
## sample estimates:
## mean of x mean of y
##  24.39231  17.14737
```

The t.test shows that mpg is greater with manual than automatic transmission. The 95% confidence interval shows that the difference can be at least of 3.9 or more.

QUESTION 2

Quantify the MPG difference between automatic and manual transmissions.

So we calculate a lineal regression for $mpg = B_0 + B_1 \cdot \text{factor}(am)$

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

As we can see the MPG difference between manual and automatic is 7.245. This could be answered using the t.test of the previous question and calculating the difference between manual = 24.39231 - automatic = 17.14737 The 95% confidence interval for the mpg difference is between 3.64151 and 10.84837 However, when we plot mpg residuals we found troubles in the model chosen, so we decide first to find a better model using stepwise starting with all possible variables. (see appendix fig 2)

```
library(MASS)
mpg2<-lm(mpg ~ ., data = mtcars)
step <- stepAIC(mpg2, direction="both")
```

Now we find the best model

```
step$coeff
```

```
## (Intercept)      wt      qsec      am
##    9.617781   -3.916504   1.225886   2.935837
```

```
bestmodel.mpg<-lm(mpg ~ wt + qsec + factor(am), data=mtcars)
confint(bestmodel.mpg)
```

```
##              2.5 %    97.5 %
## (Intercept) -4.63829946 23.873860
## wt          -5.37333423 -2.459673
## qsec         0.63457320  1.817199
## factor(am)1  0.04573031  5.825944
```

When we use bestmodel.mpg we find that mpg difference is of 2.9358 with an 95% confidence interval between 0.04573031 and 5.825944 When we plot bestmodel.mpg, we found a better performance of residuals (see Appendix fig 3)

APPENDIX

fig 1. Boxplot of manual vs automatic

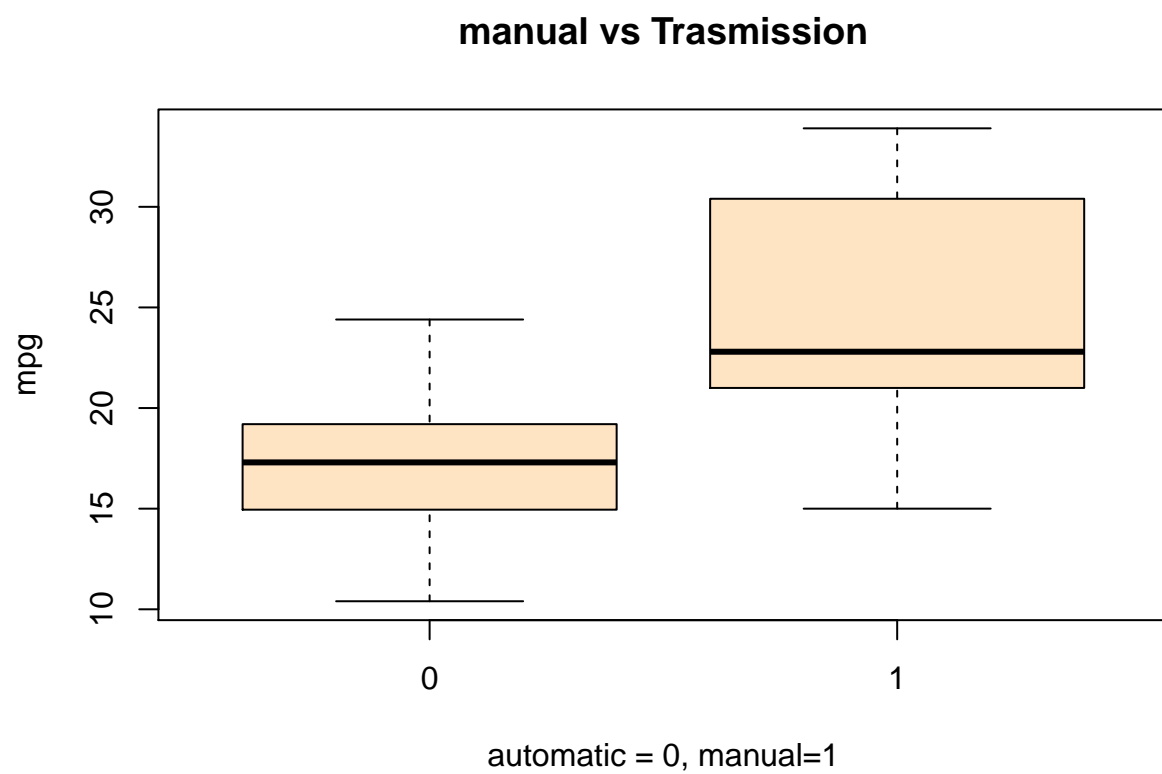


fig 2. plot of residuals of mpg

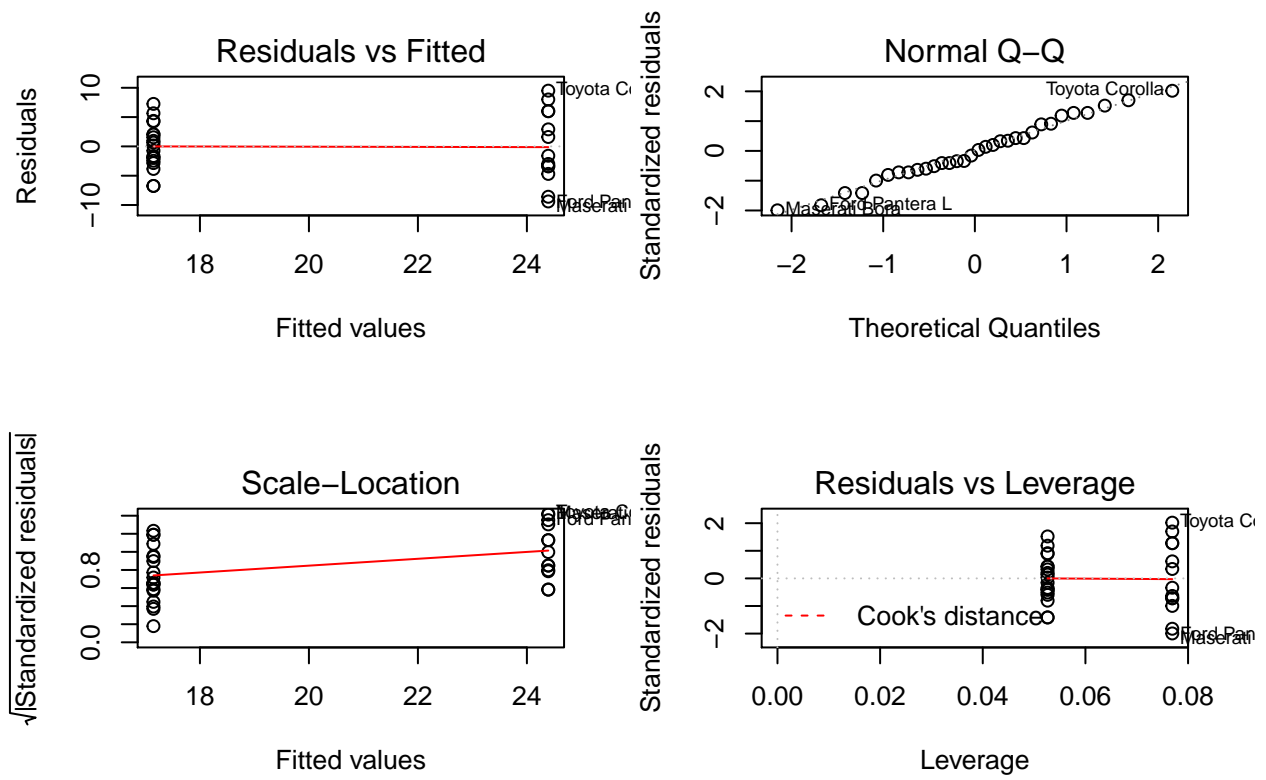
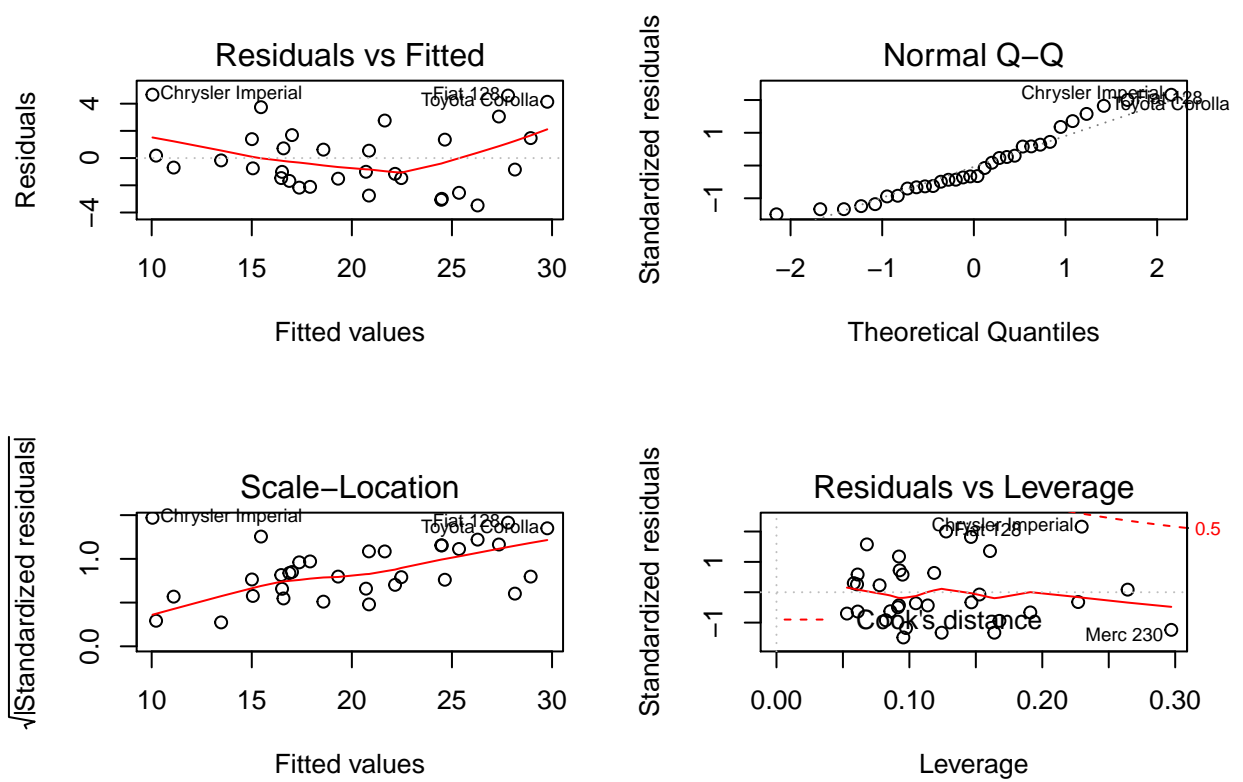


fig 3 plot of residuals of bestmodel.mpg



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.