MPG and transmission

Shuai Wang

Sunday, May 24, 2015

Executive Summary

To attract more readers of our magazine, we want to explore the relationship between a set of variables and miles per gallon (MPG), especially that between automatic/manual transmission and MPG. After investigating it, we find that manual transmission do have larger MPG than automatic one with high statistical significance if unadjusted for all other variables. However, when looking at all the variables available, we discover that the most two significant variables related to MPG are car weight and gross horsepower, and after adjusting for them, automatic/manual transmission is no longer significantly related to MPG. In fact, all other variables available become insignificant after adjusting for them.

Introduction

Nowadays the interest in fuel efficiency is still hot for both individuals and companies, as the overall economy has been weak for years. Articles about the relationship between a set of automobile related variables and miles per gallon (MPG) will be very attractive in our *Motor Trend* magazine, especially about the relationship and its magnitude between automatic/manual transmission and MPG, which has a lot of rumors.

Exploratory Data Analysis

The data available to use is the `mtcars` dataset in the `R` software, which is in fact collected by us in 1974. This data is attached to `R`, so there is no need to read it in.

First let's look at the relationship between MPG and automatic/manual transmission and see what we can get (Appendix A). Note that the x axis value is 1 for manual transmission, and 0 for automatic.

From this plot, one may see that manual transmission will have larger MPG than automatic one if unadjusted for any other variable. We will confirm it later in the inference section.

Now we want to explore what will happen if adjusted. First take a look at the full model with all predictors included.

```r
summary(lm(mpg~.-cyl-gear+as.factor(cyl)+as.factor(gear),data=mtcars))$coef
```

```
##                     Estimate  Std. Error    t value   Pr(>|t|)
## (Intercept)      15.09261548 17.13627433  0.8807408 0.38946336
## disp              0.01256810  0.01774024  0.7084518 0.48726645
## hp               -0.05711722  0.03174603 -1.7991927 0.08789210
## drat              0.73576811  1.98461241  0.3707364 0.71493502
## wt               -3.54511861  1.90895437 -1.8570997 0.07886857
## qsec              0.76801287  0.75221895  1.0209964 0.32008122
## vs                2.48849171  2.54014636  0.9796647 0.33956206
## am                3.34735713  2.28948094  1.4620594 0.16006890
## carb              0.78702815  1.03599487  0.7596834 0.45676696
## as.factor(cyl)6  -1.19939698  2.38736481 -0.5023937 0.62116357
## as.factor(cyl)8   3.05491692  4.82986776  0.6325053 0.53459525
## as.factor(gear)4 -0.99921782  2.94657533 -0.3391116 0.73824498
## as.factor(gear)5  1.06454635  3.02729599  0.3516492 0.72897110
```

From the result, We can see car weight has the smallest p value. So we will choose it as our initial model and add other variables one by one, and see which one is more significant by looking at their p values.

```
fit<-lm(mpg~wt,data=mtcars)
sapply(c("hp","vs","carb","drat","am","qsec"), function(var){
    fit2<-update(fit,paste0("mpg~wt+",var))
    anova(fit,fit2)[2,6]})
```

```
##          hp          vs        carb        drat          am        qsec
## 0.001451229 0.012925801 0.025645772 0.330854410 0.987914586 0.001499883
```

So you can see horsepower has the least p-value, which is less than 0.05. Add it into the model, repeat the above process, and this time no other variables can be added, because their p-values are all bigger than 0.05 (output is omitted). Hence we will only choose car weight and horsepower as our predictors.

Statistical Inference

First let us look at the unadjusted regression with only the automatic/manual transmission as the predictor.

```
options(scipen=5)
fit<-lm(mpg~am,data=mtcars)
coef<-summary(fit)$coef
coef
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## am           7.244939   1.764422  4.106127 2.850207e-04
```

The regression coefficient for automatic/manual transmission (variable `am`) is 7.2449393. It means that, the expected MPG for cars with manual transmission is 7.2449393 miles per US gallon more than that for cars with automatic transmission. Its p value is 0.000285, which is the probability of obtaining evidence as extreme or more extreme than what we have obtained, under the null hypothesis that the regression coefficient is zero, i.e., there is no difference on the expected MPG between manual transmission and automatic transmission. This value is now very small, so we reject the null hypothesis and conclude that manual transmission has a higher expected MPG than automatic transmission with significance level 0.000285.

The residual plot looks good, scattering around 0 with roughly equal variability (Appendex B).

Because we already get a final model with only car weight and horsepower as predictors, let us look at the relation between automatic/manual transmission and MPG adjusted for them.

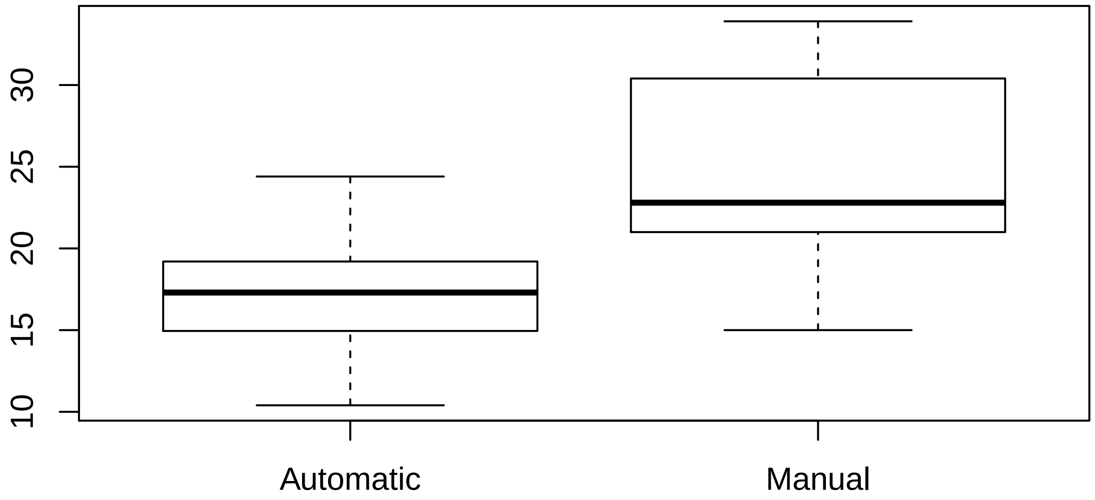
```
fit<-lm(mpg~wt+hp+am,data=mtcars)
coef<-summary(fit)$coef
coef
```

```
##                Estimate  Std. Error   t value     Pr(>|t|)
## (Intercept) 34.00287512 2.642659337 12.866916 2.824030e-13
## wt          -2.87857541 0.904970538 -3.180850 3.574031e-03
## hp          -0.03747873 0.009605422 -3.901830 5.464023e-04
## am           2.08371013 1.376420152  1.513862 1.412682e-01
```

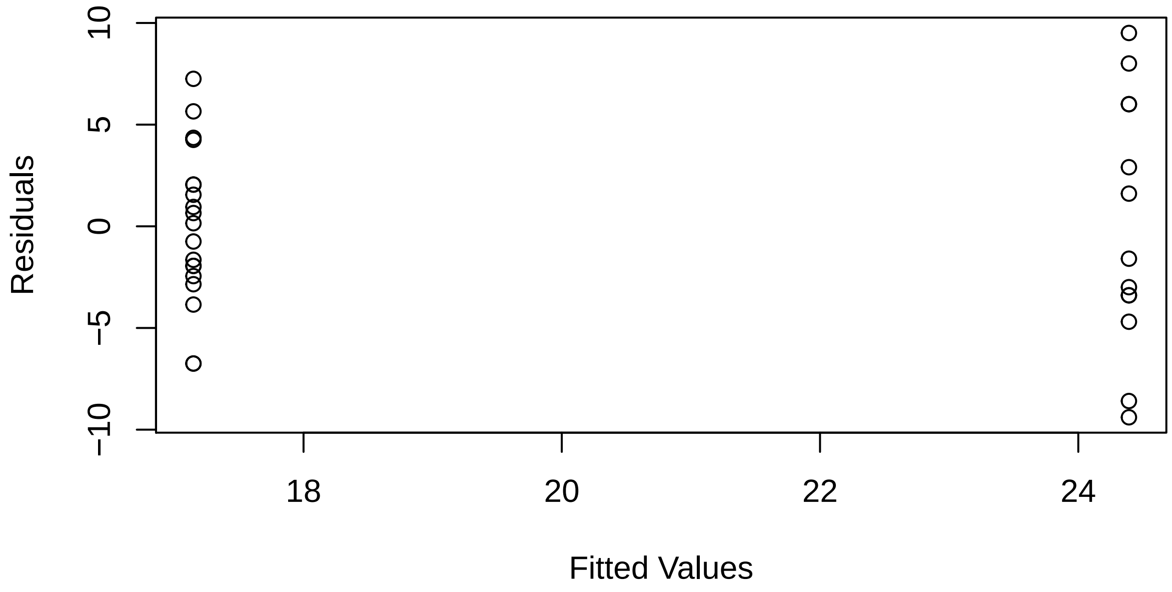
The regression coefficient for automatic/manual transmission (variable `am`) is 2.0837101 with p value 0.1412682, which is the probability of obtaining evidence as extreme or more extreme than what we have obtained, under the null hypothesis that the regression coefficient is zero after adjusting for reciprocals of car weight and horsepower. This value is now very big, so we fail to reject the null hypothesis and conclude that there is no difference on the expected MPG between manual transmission and automatic transmission for the cars with the same car weight and horsepower.

The residual plot is also good in this case (Appendix C).

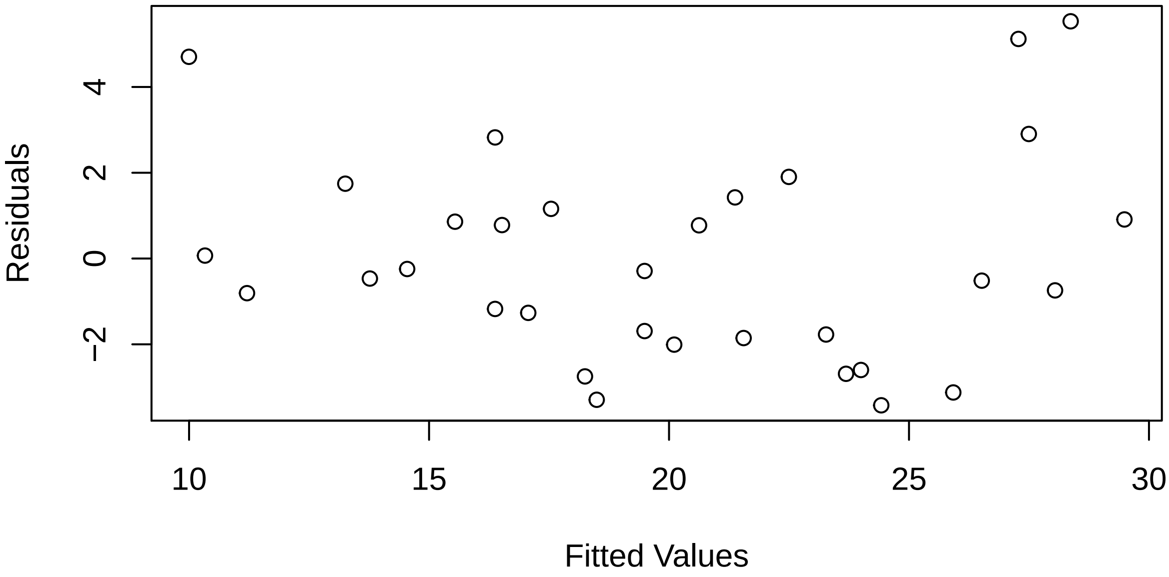
Appendix A. Boxplot of MPG by Transmission Type



Appendix B. Residual Plot for the Unadjusted Model



ppendix C. Residual Plot for the Model Adjusted for Weight and Horsep


Residuals
Fitted Values