Car fuel consumption depending on transmission type

Evgeny Gorelov

3/22/2019

Executive summary

Studied the **mtcars** dataset using the most appropriate model, it was found that in the **mtcars** data set cars with manual transmition consume less fuel compared to cars with automatic transmittion. However this difference from **-11.280194** to **-3.209684** mpg is mostly explained by such confounders as power **hp** and weight **wt**, illustrating the fact, that heavier and more powerful cars are more often equipped with automatic transmission (see plots in the appendix). Taking into account power **hp** and weight **wt** confounder variables, the **mpg** difference between cars with manual and automatic transmission is **2.08371013** mpg, however the corresponding p-value is **0.1412682**, that does not allow us to draw a conclusion about fuel consumption difference between cars with manual and automatic transmission basing on the present **mtcars** dataset.

Exploratory data analyses

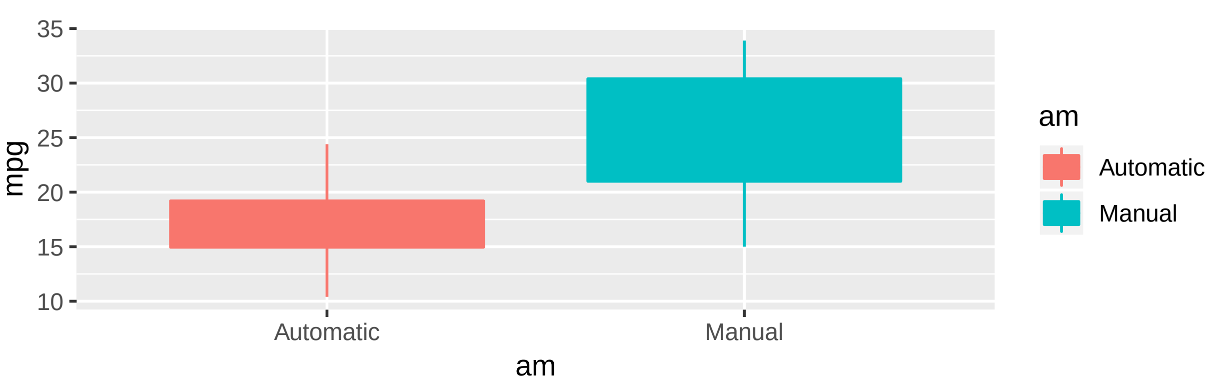
The **mtcars** data set is extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models). The data set includes data for **32** cars with **11** variables:

```
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
```

At the first glance, taking into account only **mpg** and **am** variables, it looks that cars with manual transmission have higher **mpg** with p-value = **0.001374**:

```
mtcars$am<-as.factor(as.character(mtcars$am))
levels(mtcars$am)[levels(mtcars$am)=="0"] <- "Automatic"
levels(mtcars$am)[levels(mtcars$am)=="1"] <- "Manual"
g<-ggplot(mtcars, aes(am, mpg, shape=am, colour=am, fill=am))
g+geom_boxplot()
```



```r
t.test(mpg~am, paired=FALSE, var.equal=FALSE, data=mtcars)
```

```
##
##  Welch Two Sample t-test
##
## data:  mpg by am
## t = -3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.280194  -3.209684
## sample estimates:
## mean in group Automatic    mean in group Manual
##                17.14737                24.39231
```

The difference in **mpg** is within 95 percent confidence interval **(-11.280194; -3.209684)**, showing that the difference of means is strictly negative, so cars with manual transmition should have larger **mpg** than ones with automatic transmission, if the sampled data set is representative. However, these preliminary results have to be checked for existance of confounder variables.

Best model search

First fit1 **mpg** vs **am** is created, and than additional predictors (**hp** and **wt**) are added to the model, and the significance level of new predictors is tested and found to be acceptable on any reasonable level:

```r
fit1<-lm(mpg~am, data=mtcars); fit3<-lm(mpg~am+hp+wt, data=mtcars)
Significant=c("hp", "wt")
sapply(Significant, function(x) anova(fit1,lm(paste("mpg~am+",x), data=mtcars))$Pr)[2,]
```

```
##           hp           wt
## 2.920375e-08 1.867415e-07
```

For the rest of variables (**cyl, disp, drat, qsec, vs, gear, carb**) it has been found, that adding them to the model is not necessary, since their p-values are well above 0.05:

```r
NonSignificant=c("cyl", "disp", "drat", "qsec", "vs", "gear", "carb")
sapply(NonSignificant, function(x) anova(fit3,lm(paste("mpg~am+hp+wt+",x), data=mtcars))$Pr)[2,]
```

```
##       cyl      disp      drat      qsec        vs      gear      carb
## 0.2119166 0.8122229 0.4823413 0.0757312 0.1896852 0.7081449 0.3708401
```

For the optimal model found the coefficient analysis leading to quantitative results is done:

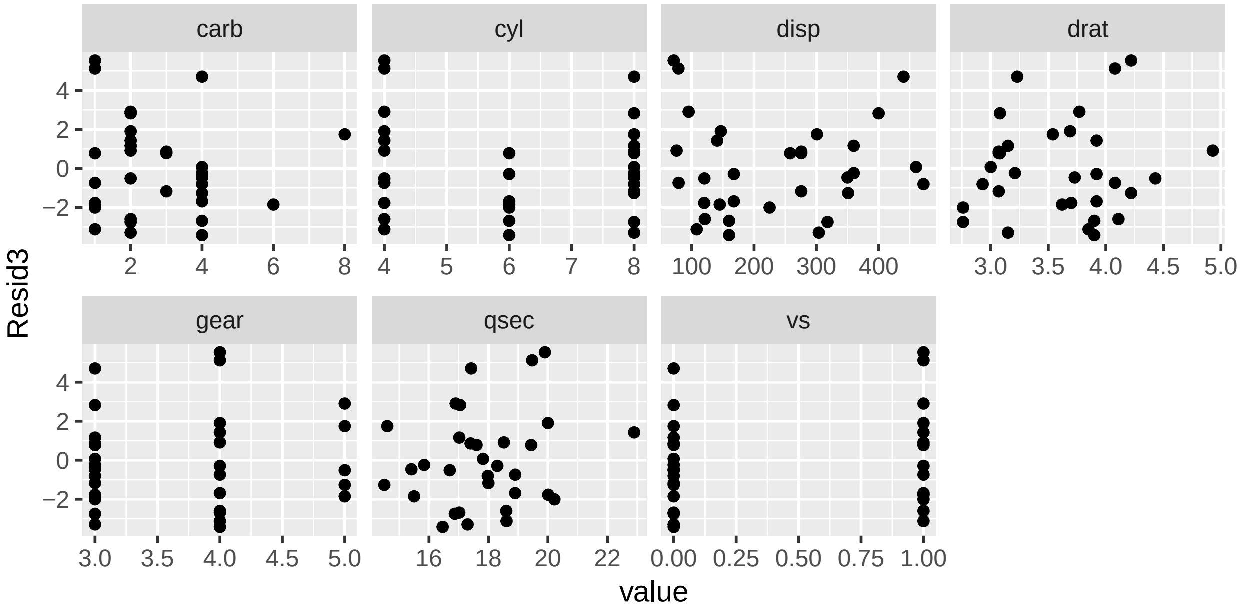
```r
coef(summary(fit3))
```

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

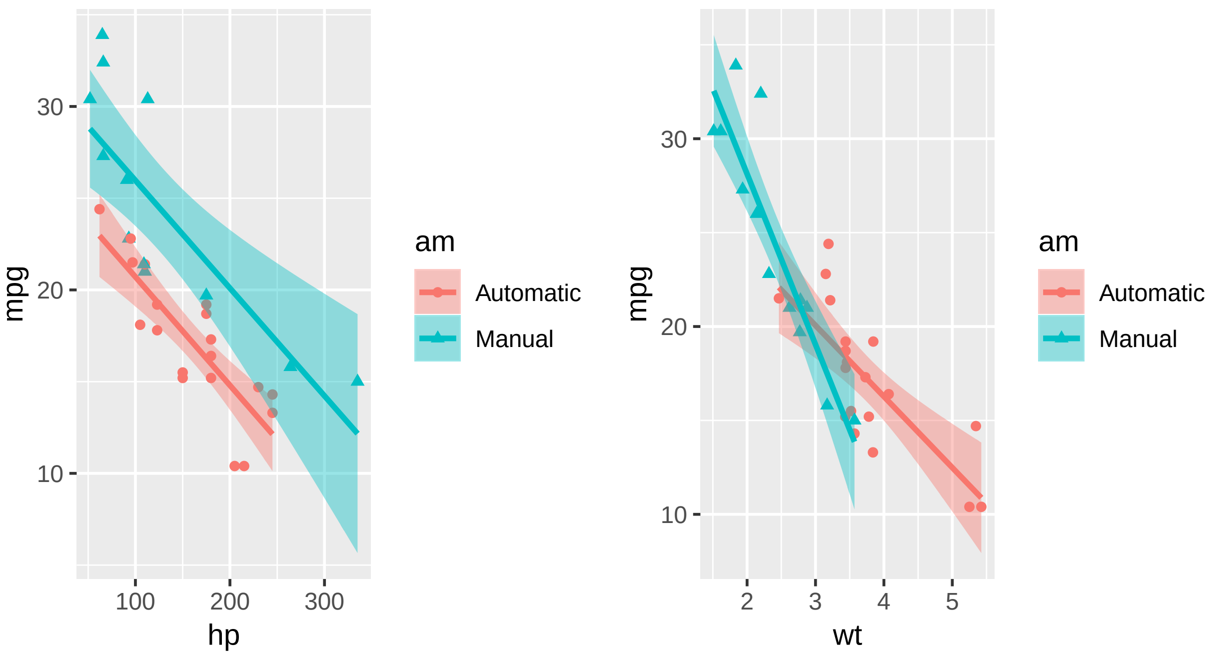
According to the summary above, mean value for **mpg** for cars with automatic transmission is **34.00287512**, and with manual **34.00287512+2.08371013=36.08659** correspondingly, however with quite large **0.1412682** p-value, definitelly outside 95% interval. With a unit increase of weight **wt** the **mpg** decreases by **2.87857541**, and a unit increase of power **hp** leads to **mpg** decrease by **0.03747873**.

Appendix

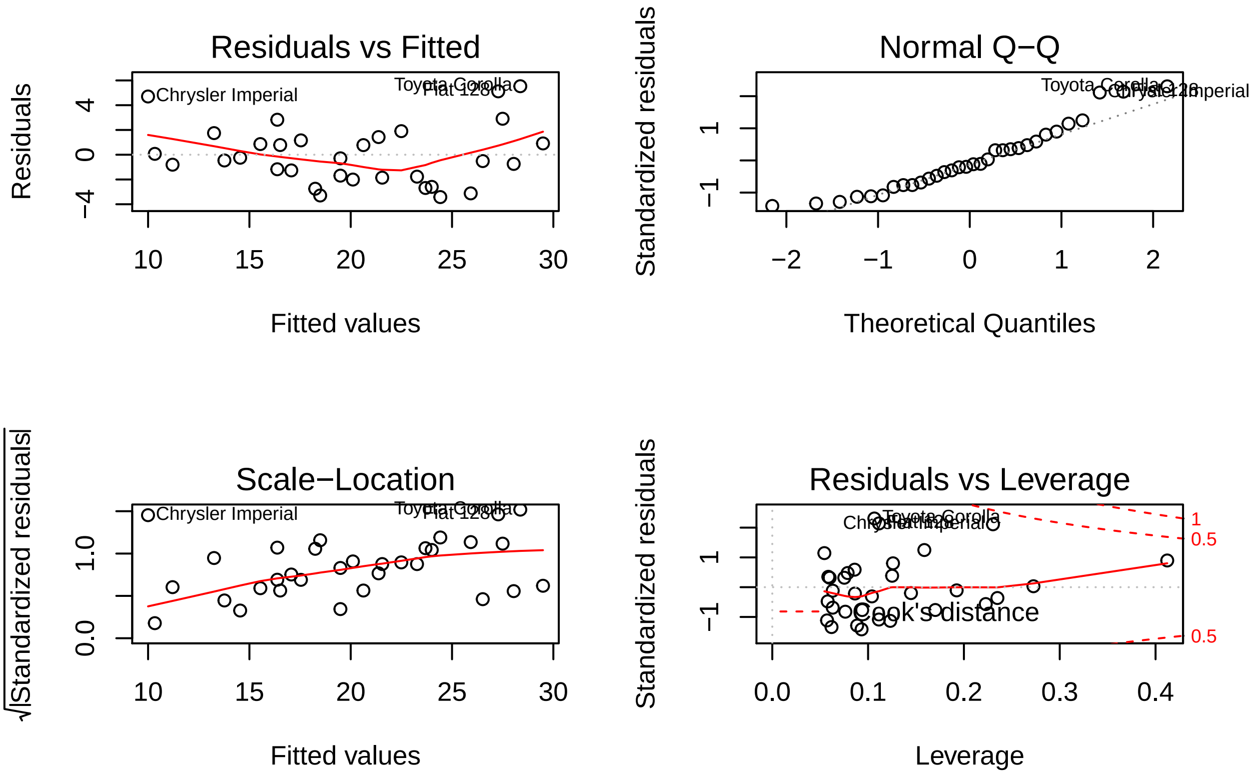
```
mtcars$Resid3<-fit3$residuals
mtcarsLong<-gather(mtcars, key="NSvar", value="value", NonSignificant)
ggplot(mtcarsLong, aes(value,Resid3))+geom_point()+labs(ylab="Residuals of fit3")+facet_wrap(~NSvar, sca
```




```
ghp<-ggplot(mtcars, aes(hp, mpg, shape=am, colour=am, fill=am))+geom_point()
gwt<-ggplot(mtcars, aes(wt, mpg, shape=am, colour=am, fill=am))+geom_point()
ggarrange(ghp+geom_smooth(method="lm"), gwt+geom_smooth(method="lm"),ncol=2,nrow=1)
```




```
par(mfrow=c(2,2))
plot(fit3)
```


Residuals vs Fitted
Residuals
Fitted values
Chrysler Imperial
Toyota Corolla
Fiat 128
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Chrysler Imperial
Toyota Corolla
Fiat 128
Scale−Location
√|Standardized residuals|
Fitted values
Chrysler Imperial
Toyota Corolla
Fiat 128
Residuals vs Leverage
Standardized residuals
Leverage
Cook's distance
Chrysler Imperial
Toyota Corolla
Fiat 128
1
0.5
0.5