MPG difference between automatic and manual transmission

X. Zeng

August 19, 2015

Introduction

Looking at a data set of a collection of cars, people in Motor Trend, a magazine about the automobile industry, 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"

In this report, I'll use regression models to answer the questions, and we will find that manual has higher mpg than automatic transmission.

Exploratory data analysis

First, let's load the data:

```
library(ggplot2)
data(mtcars)
head(mtcars,3)
```

```
##                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
```

Then, let's use a boxplot to see the mpg difference between automatic (`am=0`) and manual (`am=1`) transmission. For the figure (in Appendix), we can see there are some difference between to two mpgs.

Regression

Now, let's use two linear regression models to see if there's some difference between different transimission systems.

```
df1 <- mtcars[mtcars$am==1,]
df2 <- mtcars[mtcars$am==0,]
dim(df1)
```

```
## [1] 13 11
```

```r
dim(df2)
```

```
## [1] 19 11
```

```r
ind <- sample(1:nrow(df2), nrow(df1))
df2 <- df2[ind,]
fit1 <- lm(mpg~.-am, data=df1)
fit2 <- lm(mpg~.-am, data=df2)
summary(fit1)
```

```
##
## Call:
## lm(formula = mpg ~ . - am, data = df1)
##
## Residuals:
##      Mazda RX4  Mazda RX4 Wag     Datsun 710       Fiat 128    Honda
Civic
##        1.07088       -0.77411       -0.75528        2.69590
0.03134
## Toyota Corolla      Fiat X1-9  Porsche 914-2   Lotus Europa Ford
Pantera L
##       -1.23282       -0.63755        0.27164        0.29677       -
0.30177
##   Ferrari Dino  Maserati Bora     Volvo 142E
##       -0.84005        0.57341       -0.39836
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -137.9074    69.1493  -1.994    0.140
## cyl           -1.2813     4.5374  -0.282    0.796
## disp           0.1799     0.1756   1.024    0.381
## hp            -0.1605     0.1434  -1.119    0.345
## drat          -4.9498     5.4684  -0.905    0.432
## wt           -10.5419     4.9958  -2.110    0.125
## qsec           8.0950     3.4515   2.345    0.101
## vs             0.9431     5.0886   0.185    0.865
## gear          12.3285     6.6603   1.851    0.161
## carb           4.6885     4.0645   1.154    0.332
##
## Residual standard error: 2.078 on 3 degrees of freedom
## Multiple R-squared:  0.9716, Adjusted R-squared:  0.8864
## F-statistic: 11.41 on 9 and 3 DF,  p-value: 0.03498
```

```r
summary(fit2)
```

```
##
## Call:
## lm(formula = mpg ~ . - am, data = df2)
##
## Residuals:
##   Hornet Sportabout             Valiant            Merc 280
```

```
##           6.421e-01          -2.048e+00           8.806e-01
##          Merc 450SE          Camaro Z28 Lincoln Continental
##           3.943e-01          -1.184e+00          -7.786e-01
##   Chrysler Imperial           Merc 280C         Merc 450SLC
##           9.514e-01          -8.806e-01          -4.211e-01
##      Hornet 4 Drive         AMC Javelin           Merc 240D
##           2.048e+00          -1.653e+00          -6.106e-16
##          Merc 450SL
##           2.048e+00
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -48.98789  132.39058  -0.370    0.736
## cyl           0.40500    3.67161   0.110    0.919
## disp         -0.03139    0.04345  -0.723    0.522
## hp            0.12475    0.16283   0.766    0.499
## drat          2.23852   14.57005   0.154    0.888
## wt            2.57193    6.39299   0.402    0.714
## qsec          0.60207    3.72899   0.161    0.882
## vs           -0.69996    5.76734  -0.121    0.911
## gear         13.70281   17.08460   0.802    0.481
## carb         -6.98933    6.81109  -1.026    0.380
##
## Residual standard error: 2.616 on 3 degrees of freedom
## Multiple R-squared:  0.8659, Adjusted R-squared:  0.4637
## F-statistic: 2.153 on 9 and 3 DF,  p-value: 0.2859
```

As we can see from the summary, the model generally represent the variation of mpg for both cases. Now, let's use t-test to see if there's any difference between the two fitted models.

```
t.test(fit1$fitted.values,fit2$fitted.values)
```

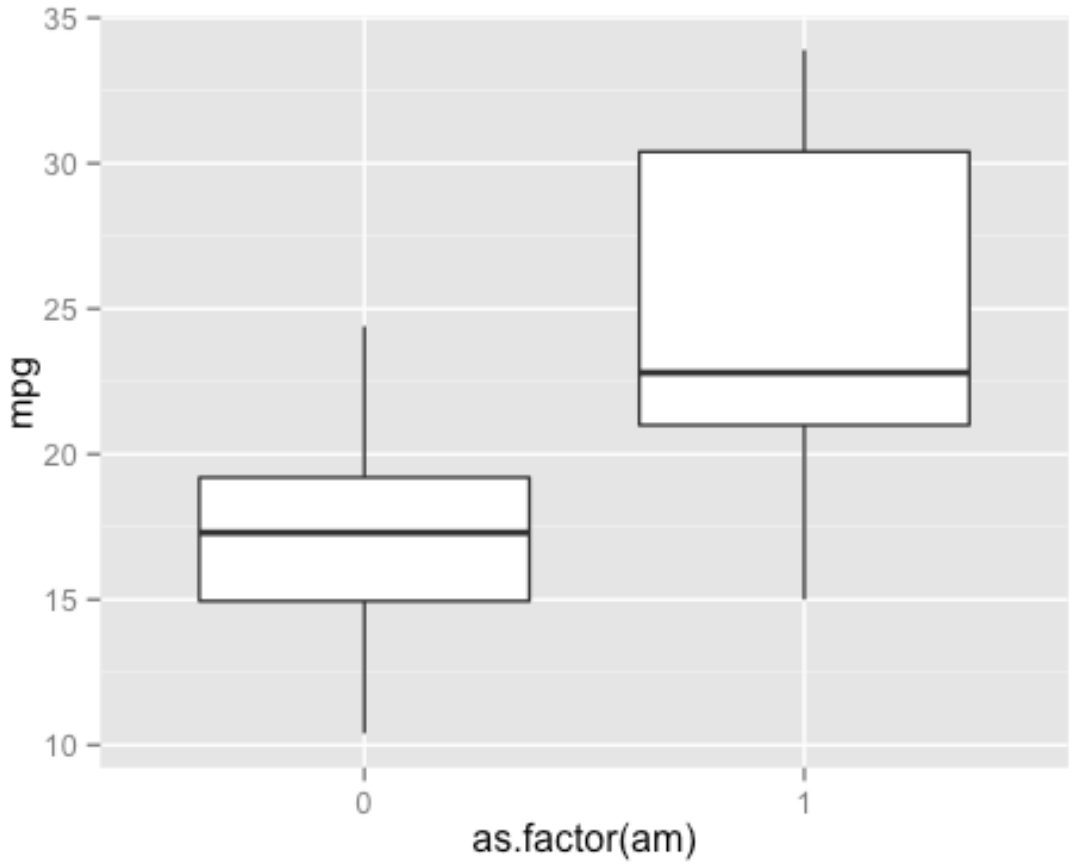
```
##
##  Welch Two Sample t-test
##
## data:  fit1$fitted.values and fit2$fitted.values
## t = 3.8034, df = 18.587, p-value = 0.00124
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   3.280138 11.335246
## sample estimates:
## mean of x mean of y
##  24.39231  17.08462
```

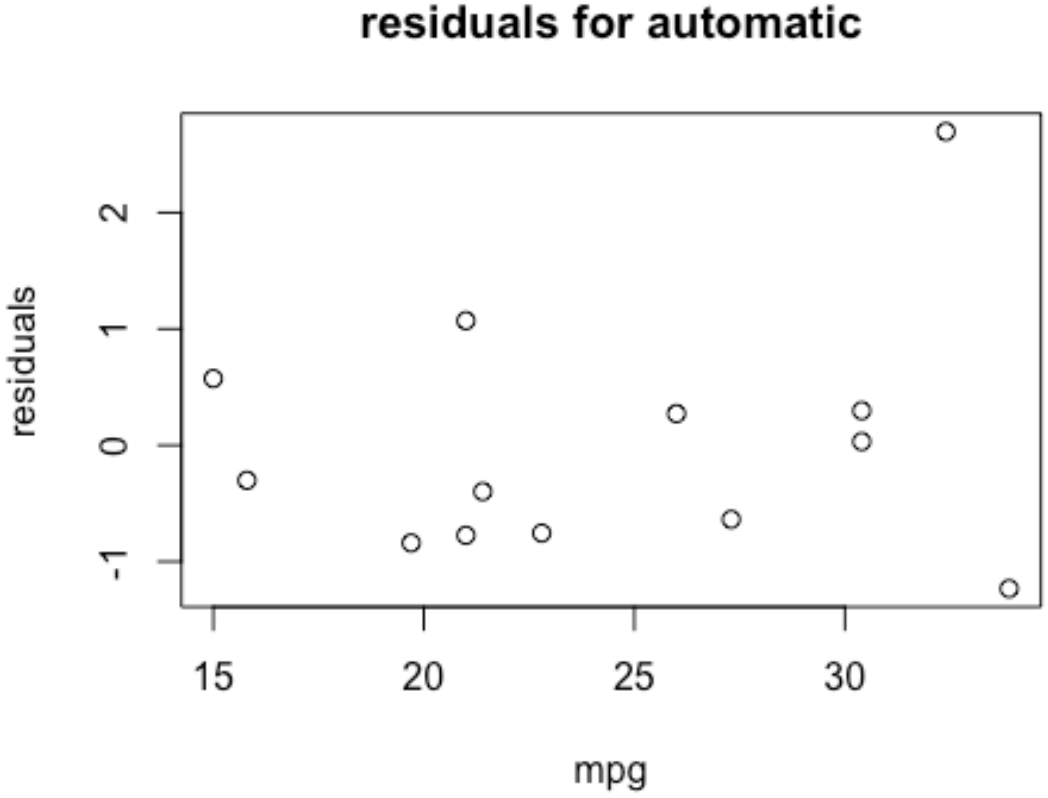
Apparently, the p-value is much less than 0.05. So we can reject the null hyposis that there's no difference between the two means related to automatic and manual transmission.

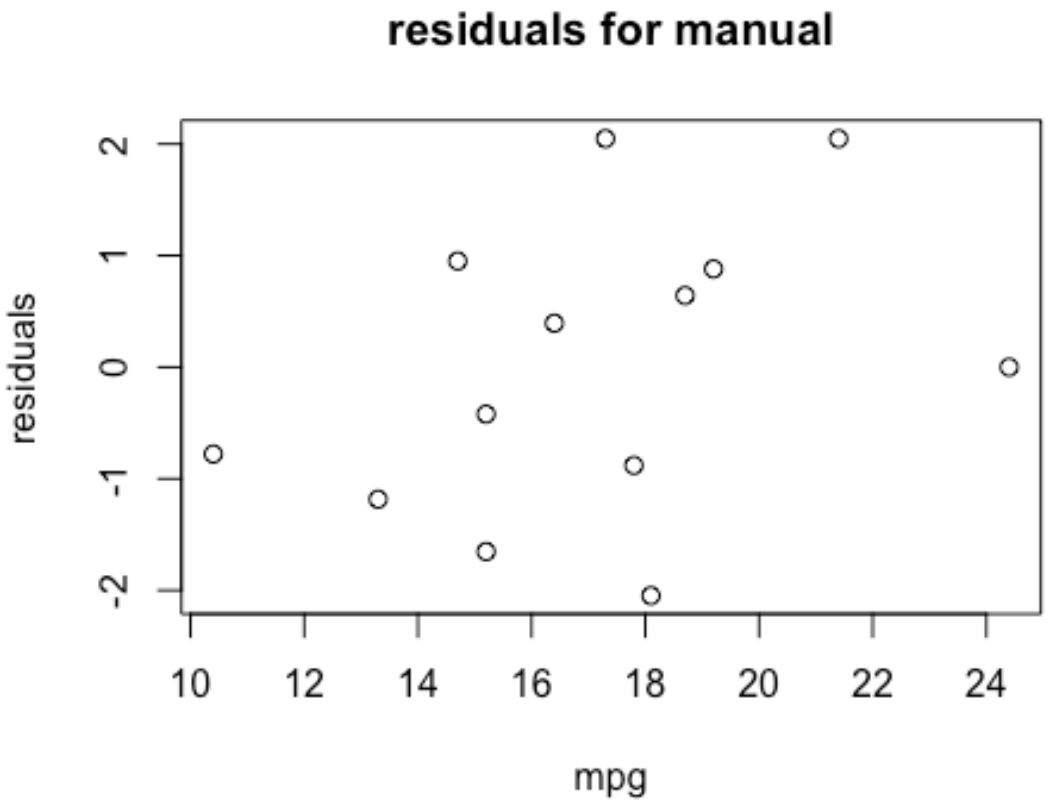
Conclusion

The MPG difference between manual and automatic transmissions is 7.3076923. That is, manual has higher mpg than automatic transmission.

Appendix




residuals for automatic
residuals
-1
0
1
2
15
20
25
30
mpg


residuals for manual
residuals
2
1
0
-1
-2
10
12
14
16
18
20
22
24
mpg