mtcars; mpg-transmission

Ishikawa

2019/8/24

```
library(tidyverse)
```

Introduction

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

We use "mtcars" dataset built in R. First, read the dataset and show outline of it. Second, do some exploratory data analyses for mpg and transmission that is auto or manual. In last, compare some regression models, consider whether the transmission influence to mpg or not.

Dataset

We use “mtcars” dataset built in R. The data was 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). Here are the outline.

```
mtcars %>% dim()
```

```
## [1] 32 11
```

```
mtcars %>% head(4)
```

```
##                    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
```

Analysis

From the format descriptions in the help, we focus following two variables.

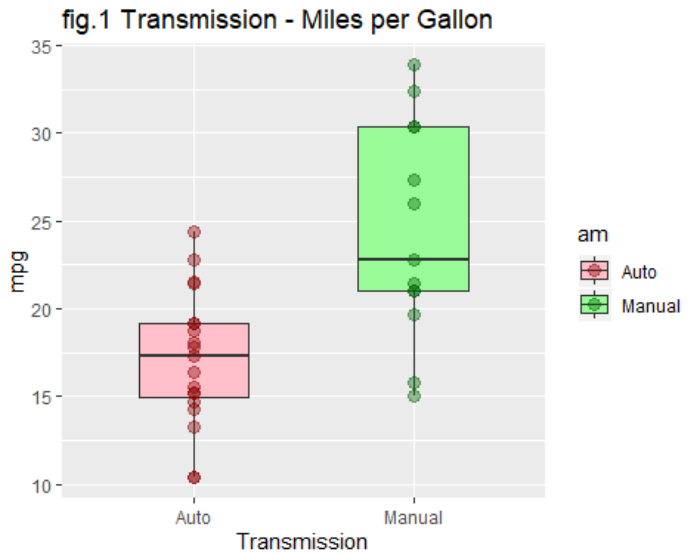
[, 1] mpg Miles/(US) gallon
[, 9] am Transmission (0 = automatic, 1 = manual)

Check the relation between mpg and am.

```
mtcars2 <- mtcars %>% mutate(am=factor(am, levels=(c(0,1)), labels=(c('Auto','Manual'))))
table(mtcars2$am)

##   Auto Manual
##     19     13

ggplot(mtcars2) + ggtitle('fig.1 Transmission - Miles per Gallon') + xlab('Transmission') +
    geom_boxplot(aes(x=am, y=mpg, fill=am ), width=0.5) +
    geom_point(  aes(x=am, y=mpg, color=am), size=3, alpha=0.4) +
    scale_color_manual(values=c('darkred','darkgreen')) +
    scale_fill_manual( values=c('pink','palegreen'))
```



It looks like Manual Transmission is better for mpg (longer distance runs at same fuel). We do T test to confirm the statistically significant.

```
g1 <- mtcars2 %>% filter(am=='Auto'   ) %>% pull(mpg)
g2 <- mtcars2 %>% filter(am=='Manual') %>% pull(mpg)
p  <- t.test(g2,g1, paired=FALSE, var.equal=TRUE, alternative='greater')$p.value %>% format(digit
s=3)
cat('T test: p-value=', p, ' < 0.05 is reject the null hypothesis (There is no deffrence.).\n')

## T test: p-value= 0.000143  < 0.05 is reject the null hypothesis (There is no deffrence.).
```

By T test, we found out “Manual Transmission is better for mpg” is statistically significant.

Regression

Check the am’s strength as regressor.

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

## Call:
## lm(formula = mpg ~ am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
## Coefficients:
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.147      1.125  15.247 1.13e-15 ***
## am             7.245      1.764   4.106 0.000285 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.902 on 30 degrees of freedom
## Multiple R-squared:  0.3598, Adjusted R-squared:  0.3385
## F-statistic: 16.86 on 1 and 30 DF,  p-value: 0.000285
```

By the regression from just “am”, Adjusted R-squared is 0.3385. There is not a strong reration between mpg and am.

Compare Regressin Models

We set following regression models, compare them.

Model 1) from just “am”
Model 2) all regressors exclude “am”
Model 3) all regressors

```
fit_excl <- lm(mpg~., data=(mtcars%>%select(-am)))
fit_all  <- lm(mpg~., data= mtcars)

anova(fit_am, fit_excl, fit_all)

## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ cyl + disp + hp + drat + wt + qsec + vs + gear + carb
## Model 3: mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb
##   Res.Df    RSS Df Sum of Sq       F   Pr(>F)
## 1     30 720.90
## 2     22 158.04  8    562.86 10.0173 1.17e-05 ***
## 3     21 147.49  1     10.55  1.5016    0.234
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Focus to the RSS (residual sum of suquare) of each model. Model 1 is 721, Model 2 is 158 and Model 3 is 147. Model 2 is mutch better from Model 1. But Model 3 is not so better from Model 2. The P-statistic of Model 3 is 0.23. It can reject that Model 3 is better.

Conclusion

In focus to factors just mpg (miles per gallon) and am (auto or manual transmission), the manual transmission is better for mpg. Otherwise include the other factors, in comparation of regression models, mpg is almost predicted by other factors. So, am is not important for mpg.

As outcome of the product line of the cars, the choice the manual transmission is better for mpg. But whether auto or manual, transmission do not makes influence for mpg better essentially.