Coursera - Regression Models Assignment

Arkadiusz Oliwa

20 march 2018

Executive summary

In this assignment I compared manual transmission and automatic transmission for MPG (Miles/US galon). There was used simple exploratory analysis as well as hypothesis testing and linear regression. As you can see in this paper having a manual car cause increase in MPG in comparison to automatic transmission.

Dataset

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

Simple exploratory data analysis

Table below and box plot in appendix shows car transmission types by MPG. As you can see there is increase in MGP for manual transmission in comparison to automatic transmission We can also see in appendix that there is linear correlation between mpg and the variables disp, hp, wt (strong negative) and drat (strong positive)

```
library(dplyr)

mtcars %>%
  mutate(Transmission = ifelse(am == 0, no = "manual", yes = "automatic")) %>%
  group_by(Transmission) %>%
  summarise(median = median(mpg), mean = mean(mpg), sd = sd(mpg),
            min = min(mpg), max = max(mpg))

## # A tibble: 2 x 6
##   Transmission median  mean    sd   min   max
##   <chr>         <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 automatic      17.3  17.1  3.83  10.4  24.4
## 2 manual         22.8  24.4  6.17  15.0  33.9
```

Hypothesis testing

I'm interested in if the average value of MPG differs significantly from a manual and automatic transmission within a defined confidence level 0.05

```
t.test(formula = mpg ~ am, 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 0 mean in group 1
##        17.14737        24.39231
```

The p-value is 0.001374 so we reject the null hypothesis and we inference that there is a significant statistical difference in the mean MPG between manual transmission cars and that of automatic transmission cars

Linear regression

I'm going to use function step() to automatically choose the best model by AIC criterion in a stepwise algorithm

```
fit <- step(object = lm(formula = mpg ~ ., data = mtcars), direction = "both")
```

Based on diagnostic plots (appendix) we can say that residual are normally distributed and homoskedastic.

```
summary(fit)
```

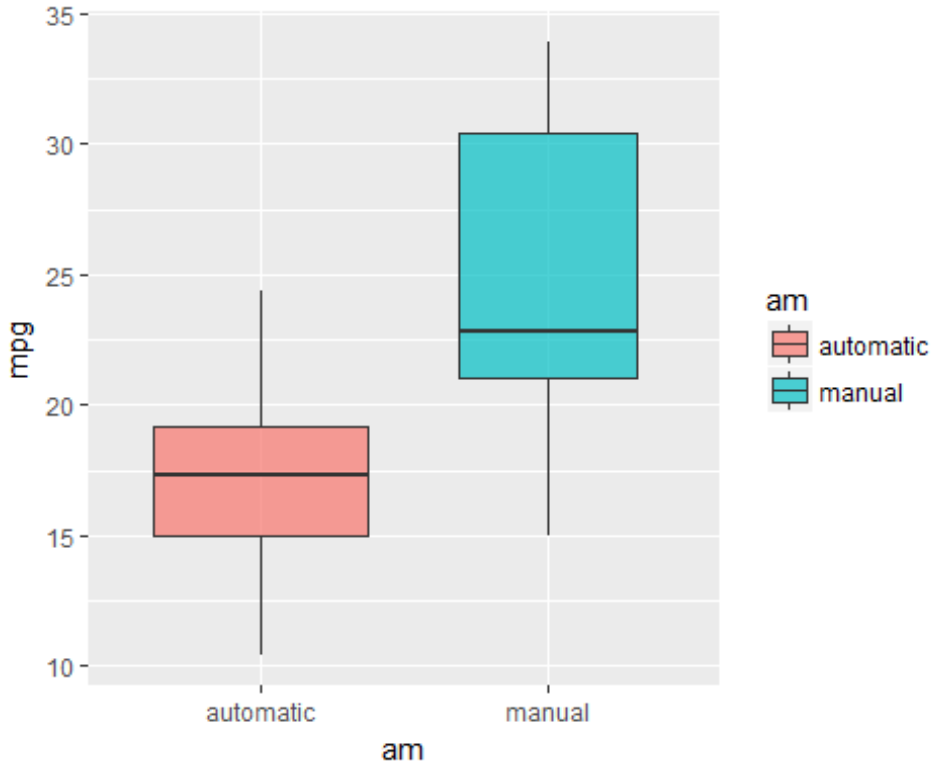
```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4811 -1.5555 -0.7257  1.4110  4.6610
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.6178     6.9596   1.382 0.177915
## wt           -3.9165     0.7112  -5.507 6.95e-06 ***
## qsec          1.2259     0.2887   4.247 0.000216 ***
## am1           2.9358     1.4109   2.081 0.046716 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.459 on 28 degrees of freedom
## Multiple R-squared:  0.8497, Adjusted R-squared:  0.8336
## F-statistic: 52.75 on 3 and 28 DF,  p-value: 1.21e-11
```

Looking at result we can say that: The adjusted R-squared for the model is 0.8336, and three variables coefficients are all significant at the 1% confidence level. Based on coefficients we can say that having a manual car we gain 2.9358 MPG above that of an automatic

Appendix

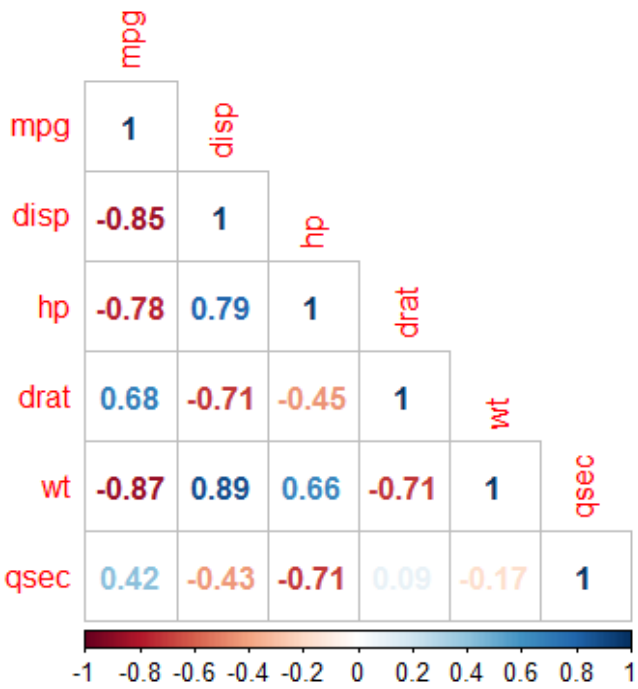
Box Plot of MPG vs. Transmission

```
library(ggplot2)

mtcars %>%
  mutate(am = ifelse(am == 0, no = "manual", yes = "automatic")) %>%
  ggplot(aes(y = mpg, x = am, fill = am)) +
  geom_boxplot(alpha = .7, varwidth = TRUE)
```



Correlation between viariables

```
library(corrplot)

corrplot(corr = cor(mtcars[, c("mpg", "disp", "hp", "drat", "wt", "qsec")],
                    method = "pearson"),
         method = "number", type = "lower")
```



Diagnostic plots

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