Regression Models Course Project

Fervallen

August 28, 2018

1. Overview

In this assignment we will look at a data set of a collection of cars to explore the relationship between a set of variables and miles per gallon (MPG) (outcome). We 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

Let's check what data we have for analysis:

```
dim(mtcars)
```

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

```
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
```

```
mtcars$mpg
```

```
##  [1] 21.0 21.0 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 17.8 16.4 17.3 15.2
## [15] 10.4 10.4 14.7 32.4 30.4 33.9 21.5 15.5 15.2 13.3 19.2 27.3 26.0 30.4
## [29] 15.8 19.7 15.0 21.4
```

```
unique(mtcars$am)
```

```
## [1] 1 0
```

We are interested in "mpg" (miles per galon) and "am" (automatic - 0 or manual - 1 transmission) fields.

2. Exploratory data analysis and answer to question "Is an automatic or manual transmission better for MPG?"

At first - let's check the MPG means for the different transmission types:

```
autoMpg <- mtcars[mtcars$am == 0, ]$mpg
mean(autoMpg)
```

```
## [1] 17.14737
```

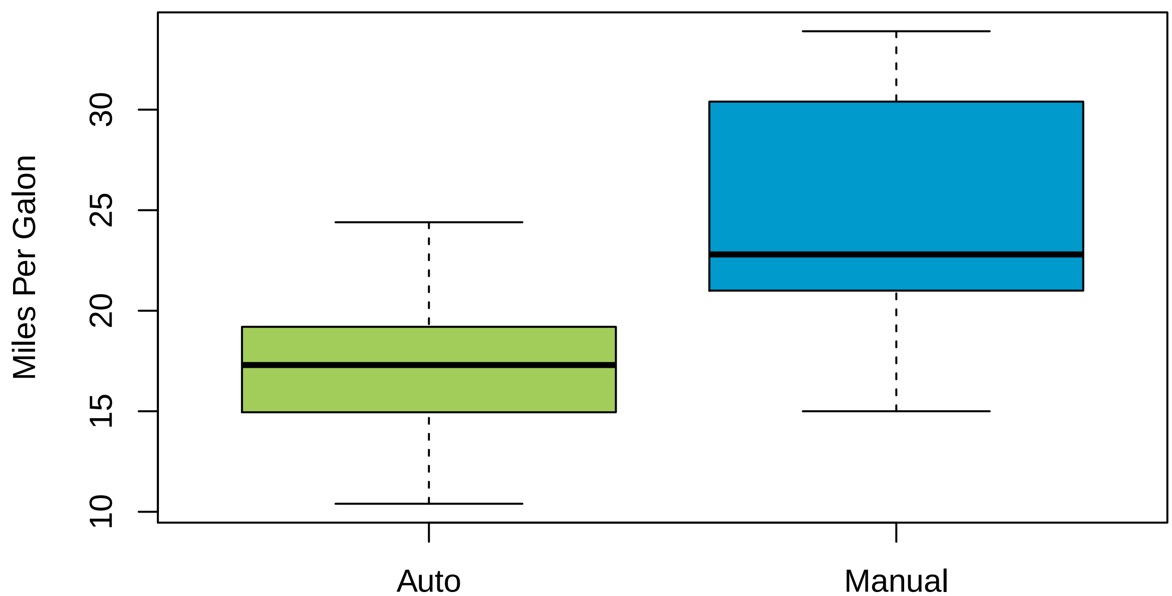
```
manualMpg <- mtcars[mtcars$am == 1, ]$mpg
mean(manualMpg)
```

```
## [1] 24.39231
```

So, at the first glance it looks like cars with automatic transmission have lower MPG.

Let's draw a boxplot so visualize this data:

```r
boxplot(mpg ~ am,
  data = mtcars,
  names = c('Auto', 'Manual'),
  ylab = 'Miles Per Galon',
  col = c('darkolivegreen3', 'deepskyblue3')
)
```



Now we are even more sure, that automatic transmission cars have bigger fuel consumption.

Let's test a hypotesis, that the transmission type does not affect MPG:

```r
t.test(autoMpg, manualMpg)
```

```
##
##  Welch Two Sample t-test
##
## data:  autoMpg and manualMpg
## 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 of x mean of y
##  17.14737  24.39231
```

Resulting P-value is really small, so hypotesis needs to be rejected.

3. Building linear models to "Quantify the MPG difference between automatic and manual transmissions"

For the start let's build a full multivariable model to check how all the variables affect MPG.
As many people have suggested we'll use a stepwise approach to find the best model solution:

```
stepModel <- step(lm(data = mtcars, mpg ~ .), trace = 0)
summary(stepModel)
```

```
##
## 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 ***
## am            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
```

As we see transmission type(am) is quite important, but weight(wt) makes more difference and 1/4 mile time(qsec) has it's impact too.

In average, choosing automatic transmission drops 2.94 MPG, but the std. error are very big.

Let's also check how weight and qsec impact MPG with different types of transmission:

```
mpgByAmModel <- lm(mpg ~ factor(am):wt + factor(am):qsec, data = mtcars)
summary(mpgByAmModel)
```

```
##
## Call:
## lm(formula = mpg ~ factor(am):wt + factor(am):qsec, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9361 -1.4017 -0.1551  1.2695  3.8862
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       13.9692     5.7756   2.419  0.02259 *
## factor(am)0:wt    -3.1759     0.6362  -4.992 3.11e-05 ***
## factor(am)1:wt    -6.0992     0.9685  -6.297 9.70e-07 ***
## factor(am)0:qsec   0.8338     0.2602   3.205  0.00346 **
## factor(am)1:qsec   1.4464     0.2692   5.373 1.12e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 2.097 on 27 degrees of freedom
## Multiple R-squared:  0.8946, Adjusted R-squared:  0.879
## F-statistic: 57.28 on 4 and 27 DF,  p-value: 8.424e-13
```

So when the weight is increased by 1000 lbs fuel consumption is decreased by 3.18 miles per galon for automatic transmission and by 6.1 miles per galon for cars with manual transmission.
When 1/4 mile time is increased by 1 sec fuel consumption is decreased by 0.834 miles per galon for automatic transmission and by 1.446 miles per galon for cars with manual transmission.