Course Project - Outcome difference between automatic and manual transmissions reasearch

Tatiana Lezhneva

Overveiw

There is a dataset **mtcars** which was extracted from the *1974 Motor trend US* magazine, comprises fuel con- sumption and 10 aspects for 32 automobiles. Questions of our research are: Is an automatic or manual transmission better for MPG; Quantify the MPG difference between automatic and manual transmissions.

1. Exploratory data analysis

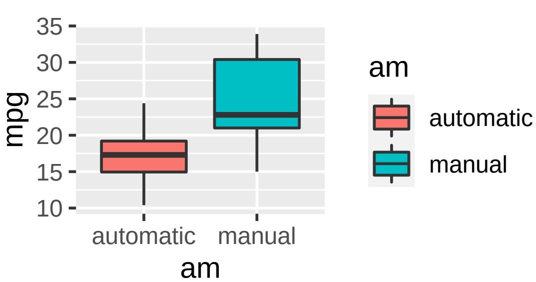
```
library(dplyr); library(ggplot2); library(car); library(cowplot);   Cars <- mtcars
```

Some variables in dataset (**vs**, **am**, **cyl**, **gear**, **carb**) need to be treated like factors:

```
Cars$vs<- factor(Cars$vs, levels = c(0, 1), labels = c("V-shaped", "straight"))
Cars$am<- factor(Cars$am, levels = c(0, 1), labels = c("automatic", "manual"))
Cars$cyl <- factor(Cars$cyl); Cars$gear <- factor(Cars$gear); Cars$carb <- factor(Cars$carb)
```

The first question of our research is about better transmission type for MPG. Let's take a look on the boxplot

```
ggplot(Cars, aes(x = am, y = mpg, fill = am)) + geom_boxplot()
```



On the first sight cars with automatic transmission appear to be more effective with outcome then cars with manual one, that is they have a significant difference. Plots of other variables from the set are in Appendix 1.

Statiatical Inference

So to figure out is there a sighnificant difference between the impact of transmission type on mpg, we use t-test. Before the test will have done lets assume that all observations are independant and cars were chosen randomly.

```
t.test(subset(Cars,Cars$am == "automatic")$mpg, subset(Cars,Cars$am == "manual")$mpg)
```

```
##
##  Welch Two Sample t-test
##
## data:  subset(Cars, Cars$am == "automatic")$mpg and subset(Cars, Cars$am == "manual")$mpg
## 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
```

We've got the p-value = 0.001374, which means the null hypothesis is true and there is a sighnificant difference between auto and manual transmissions. Another question to answer is Quantity the MPG difference between automatic and manual transmissions. Looking at the relation between MPG and variables we can see a kind of linear associations so we are going to build multyvariant regression model. Let's start with full model that include all the variables (exept **gear** because its unbalanced)

```
Cars_model0 <- lm(mpg ~ .-gear, Cars)
```

Next step is checking variance inflation factor (VIF) for all the parameters and leaving in model only the uncorrelated ones with VIF < 2 (Applendix 2) . The model that we've got:

```
Cars_model6
```

```
##
## Call:
## lm(formula = mpg ~ cyl + am + carb, data = Cars)
##
## Coefficients:
## (Intercept)         cyl6         cyl8     ammanual        carb2        carb3
##     23.6804      -2.7286      -6.9025       4.2735      -0.2288      -0.4779
##       carb4        carb6        carb8
##     -3.9396      -5.5254      -6.0515
```

After using F-test(Appendix 3), which shows that **carb** is unnesessery parameter(p=0.195), the final model

```
Cars_model_final <- update(Cars_model6, .~.-carb); summary(Cars_model_final)
```

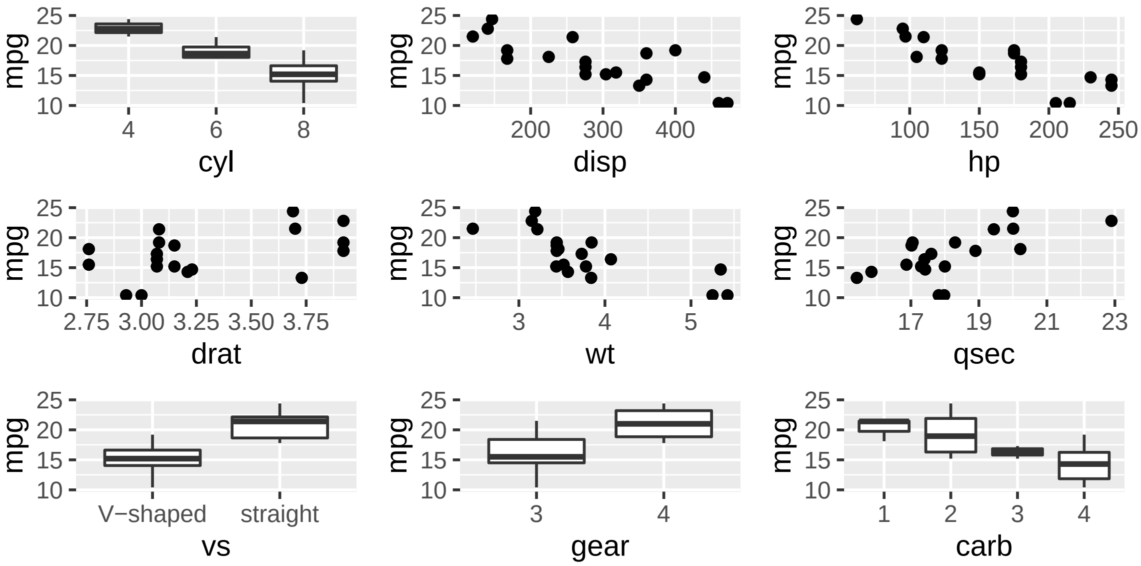
```
##
## Call:
## lm(formula = mpg ~ cyl + am, data = Cars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.9618 -1.4971 -0.2057  1.8907  6.5382
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   24.802      1.323  18.752  < 2e-16 ***
## cyl6          -6.156      1.536  -4.009 0.000411 ***
## cyl8         -10.068      1.452  -6.933 1.55e-07 ***
## ammanual       2.560      1.298   1.973 0.058457 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.073 on 28 degrees of freedom
## Multiple R-squared:  0.7651, Adjusted R-squared:  0.7399
## F-statistic:  30.4 on 3 and 28 DF,  p-value: 5.959e-09
```

The model describes about 98% of the variance, p-value=2.2e-16. Evaluation of residuals is valid(Appendix4).

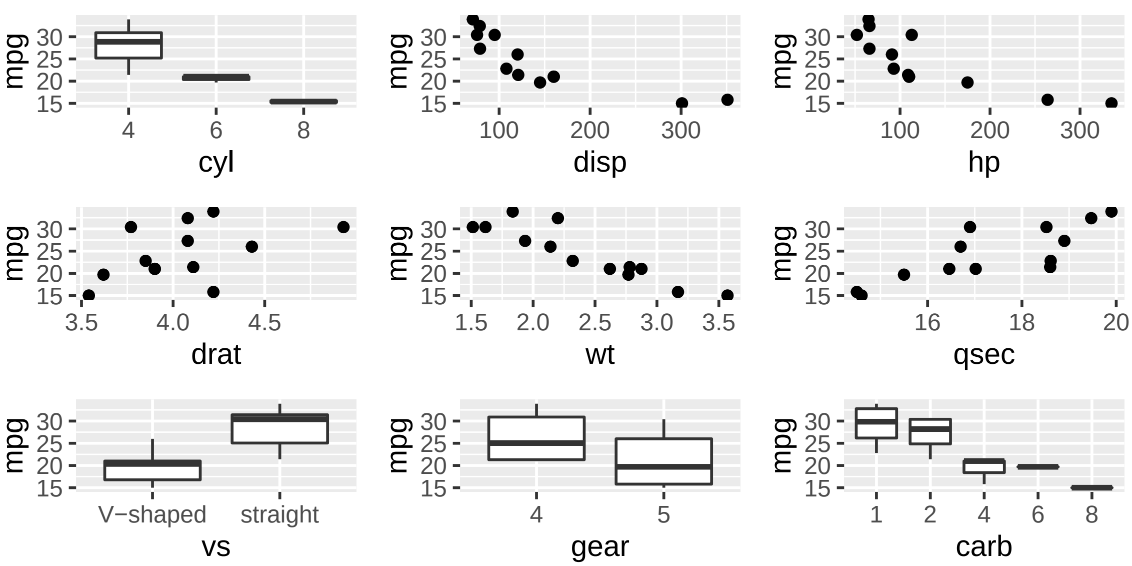
Conclusions

The transmission type does impact on MPG, auto transmission is better for mpg than manual one (p-value = 0.001374). In the final model there are two parameters: transmission and number of cylinders, which describe about 98% of the variance. According to this model manual transmission get 2.56 more mpg then manual one (holding number of cylenders cobstant).

Appendix 1



Plot 1. MPG for Automatic Transmission



Plot 2. MPG for Manual Transmission

Appendix 2

```
Cars_model1 <- update(Cars_model0, .~.-disp); vif(Cars_model1)
```

```
##            GVIF Df GVIF^(1/(2*Df))
## cyl   53.926155  2        2.709879
```

```
## hp   16.770720  1        4.095207
## drat  6.409877  1        2.531774
## wt    9.907262  1        3.147580
## qsec  9.068990  1        3.011476
## vs    6.663306  1        2.581338
## am    5.434991  1        2.331307
## carb 54.378009  5        1.491222
```

```r
Cars_model2 <- update(Cars_model1, .~.-hp); vif(Cars_model2)
```

```
##           GVIF Df GVIF^(1/(2*Df))
## cyl  28.847347  2        2.317536
## drat  6.390472  1        2.527938
## wt    9.735362  1        3.120154
## qsec  7.773235  1        2.788052
## vs    6.113767  1        2.472603
## am    5.354759  1        2.314035
## carb 22.048458  5        1.362504
```

```r
Cars_model3 <- update(Cars_model2, .~.-wt); vif(Cars_model3)
```

```
##           GVIF Df GVIF^(1/(2*Df))
## cyl  20.013591  2        2.115102
## drat  4.888260  1        2.210941
## qsec  5.253062  1        2.291956
## vs    6.020006  1        2.453570
## am    5.333782  1        2.309498
## carb  9.258349  5        1.249261
```

```r
Cars_model4 <- update(Cars_model3, .~.-vs); vif(Cars_model4)
```

```
##           GVIF Df GVIF^(1/(2*Df))
## cyl  14.783614  2        1.960854
## drat  4.689795  1        2.165593
## qsec  5.058412  1        2.249091
## am    4.776215  1        2.185455
## carb  6.996224  5        1.214748
```

```r
Cars_model5 <- update(Cars_model4, .~.-qsec); vif(Cars_model5)
```

```
##          GVIF Df GVIF^(1/(2*Df))
## cyl  7.525753  2        1.656294
## drat 4.648333  1        2.155999
## am   2.762317  1        1.662022
## carb 5.799609  5        1.192175
```

```r
Cars_model6 <- update(Cars_model5, .~.-drat); vif(Cars_model6)
```

```
##          GVIF Df GVIF^(1/(2*Df))
## cyl  4.038267  2        1.417584
## am   1.878638  1        1.370634
## carb 3.645833  5        1.138098
```

Appendix 3

```r
drop1(Cars_model6, test = "F")
```

```
## Single term deletions
```

```
##
## Model:
## mpg ~ cyl + am + carb
##        Df Sum of Sq    RSS    AIC F value   Pr(>F)
## <none>              195.61 75.933
## cyl     2   117.505 313.12 86.987  6.9081 0.004471 **
## am      1    75.038 270.65 84.323  8.8230 0.006850 **
## carb    5    68.885 264.50 75.587  1.6199 0.194581
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Appendix 4