

Estudo 1

Jonatan Almeida and Helbert Paulino

2023-09-27

```
## Registered S3 method overwritten by 'GGally':  
##   method from  
##   +.gg      ggplot2
```

Resumo

Este estudo de caso é uma comparação do IMC médio de alunos do PPGE ao longo de dois semestres. Foram coletados dados de 2016/2 e 2017/2. Os dados são compostos de:

- altura;
- idade;
- sexo;
- peso;
- curso (PPGE ou PPGSEN) - aplica-se apenas para 2016/2.

A pergunta de interesse é a seguinte:

Existe alteração no estilo de vida entre os alunos do PPGE de um semestre para outro?

A análise será feita por subgrupos, masculino e feminino.

```
install.packages("devtools")           # you only have to install it once  
library(devtools)  
install_github("rstudio/rmarkdown")    # you only have to install it once  
library(rmarkdown)  
render("Estudo_1.Rmd", "pdf_document") # this renders the pdf
```

Design experimental

A pergunta de interesse nos leva a definir os seguintes teste de hipóteses:

$$\begin{cases} H_0 : \mu_{2016} = \mu_{2017} \\ H_1 : \mu_{2016} \neq \mu_{2017} \end{cases}$$

Onde o parâmetro μ significa o IMC médio de cada um dos anos. A hipótese H_0 significa que não houve alteração no estilo de vida entre os alunos e a hipótese H_1 significa que houve alteração, ou seja, as médias de IMC são diferentes entre os alunos.

Se tratando de IMC, existe as seguintes classificações:

- **IMC** < 18,5kg/m² - baixo peso
- **IMC** > 18,5 até 24,9kg/m² - eutrofia (peso adequado)
- **IMC** ≥ 25 até 29,9kg/m² - sobrepeso
- **IMC** > 30,0kg/m² até 34,9kg/m² - obesidade grau 1

- **IMC** > 35kg/m² até 39,9kg/m² - obesidade grau 2
- **IMC** > 40kg/m² - obesidade extrema

Nota-se que a alteração é sempre de 5 em 5 kg/m². Logo um valor interessante para o efeito mínimo relevante (δ^*) é uma alteração de 5 entre as médias ou uma alteração na classificação da média do IMC da turma.

Para o teste estatístico será dividido em duas análises, uma para o sexo masculino e uma para o sexo feminino. Então serão dois testes de hipóteses distintos, um para cada sexo.

Como a variância da população não é conhecida, utilizaremos o teste t com um $\alpha = 0,5$.

Description of the data collection

Whenever needed, you can also include an (optional) subsection describing the actual data collection, how it was performed, any adaptations or unexpected events that may have occurred, etc. Subsections like this can also be used for the sample size calculations or any other aspect that requires a longer discussion.

Exploratory Data Analysis

The first step is to load and preprocess the data. For instance,

```
data(mtcars)
fc<-c(2,8:11)
for (i in 1:length(fc)){mtcars[,fc[i]]<-as.factor(mtcars[,fc[i]])}
levels(mtcars$am) <- c("Automatic","Manual")
```

To get an initial feel for the relationships between the relevant variables of your experiment it is frequently interesting to perform some preliminary (exploratory) analysis. This is frequently referred to as *getting a feel* of your data, and can suggest procedures (such as outlier investigation or data transformations) to experienced experimenters.

```
library(GGally,quietly = T, warn.conflicts = F) # This is just me getting fancy.
                                                # There are much simpler ways ;-)
ggpairs(data=mtcars,columns=c(1,9),title="MPG by transmission type",
        upper=list(combo="box"),lower=list(combo="facethist"),
        diag=list(continuous="densityDiag",discrete="barDiag"))
```

Your preliminary analysis should be described together with the plots. In this example, two facts are immediately clear from the plots: first, **mpg** tends to correlate well with many of the other variables, most intensely with **drat** (positively) and **wt** (negatively). It is also clear that many of the variables are highly correlated (e.g., **wt** and **disp**). Second, it seems like manual transmission models present larger values of **mpg** than the automatic ones. In the next section a linear model will be fit to the data in order to investigate the significance and magnitude of this possible effect.

Statistical Analysis

Your statistical analysis should come here. This is the place where you should fit your statistical model, get the results of your significance test, your effect size estimates and confidence intervals.

```
model<-aov(mpg~am*disp,data=mtcars)
summary(model)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## am          1  405.2   405.2   47.948 1.58e-07 ***
## disp        1  420.6   420.6   49.778 1.13e-07 ***
## am:disp      1   63.7    63.7    7.537  0.0104 *
## Residuals   28  236.6     8.4
## ---
```

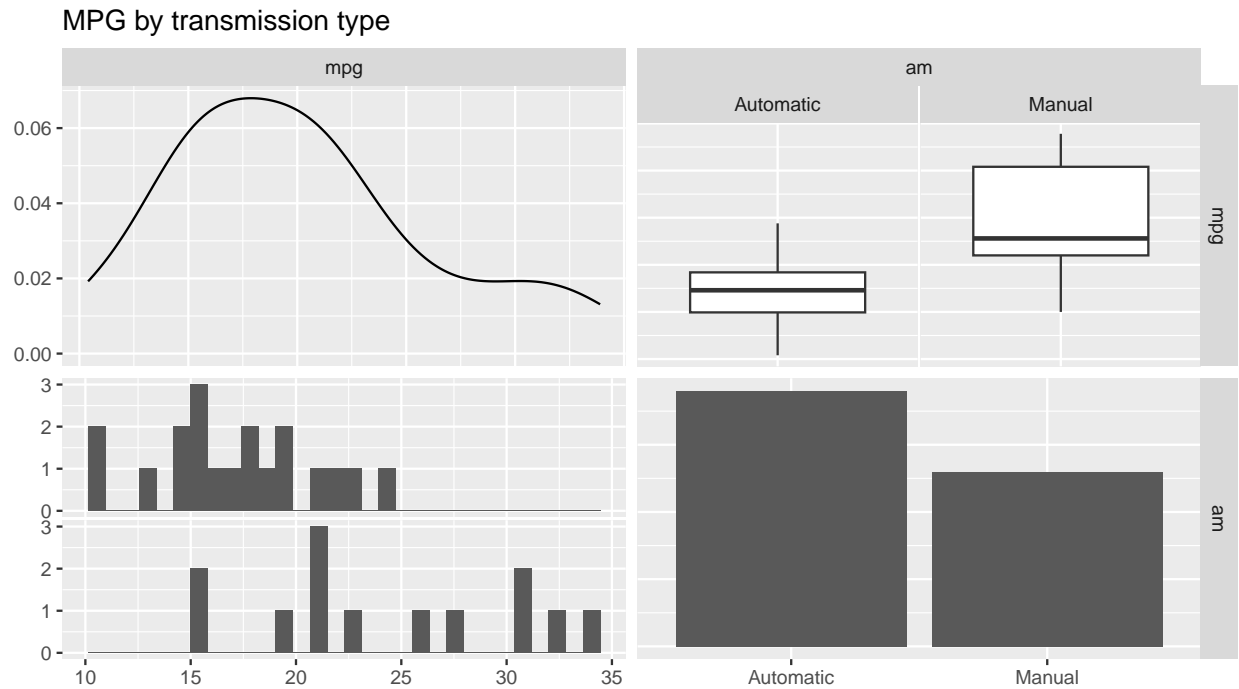


Figure 1: Exploring the effect of car transmission on mpg values

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Checking Model Assumptions

The assumptions of your test should also be validated, and possible effects of violations should also be explored.

```
par(mfrow=c(2,2), mai=.3*c(1,1,1,1))
plot(model,pch=16,lty=1,lwd=2)
```

Conclusions and Recommendations

The discussion of your results, and the scientific/technical meaning of the effects detected, should be placed here. Always be sure to tie your results back to the original question of interest!

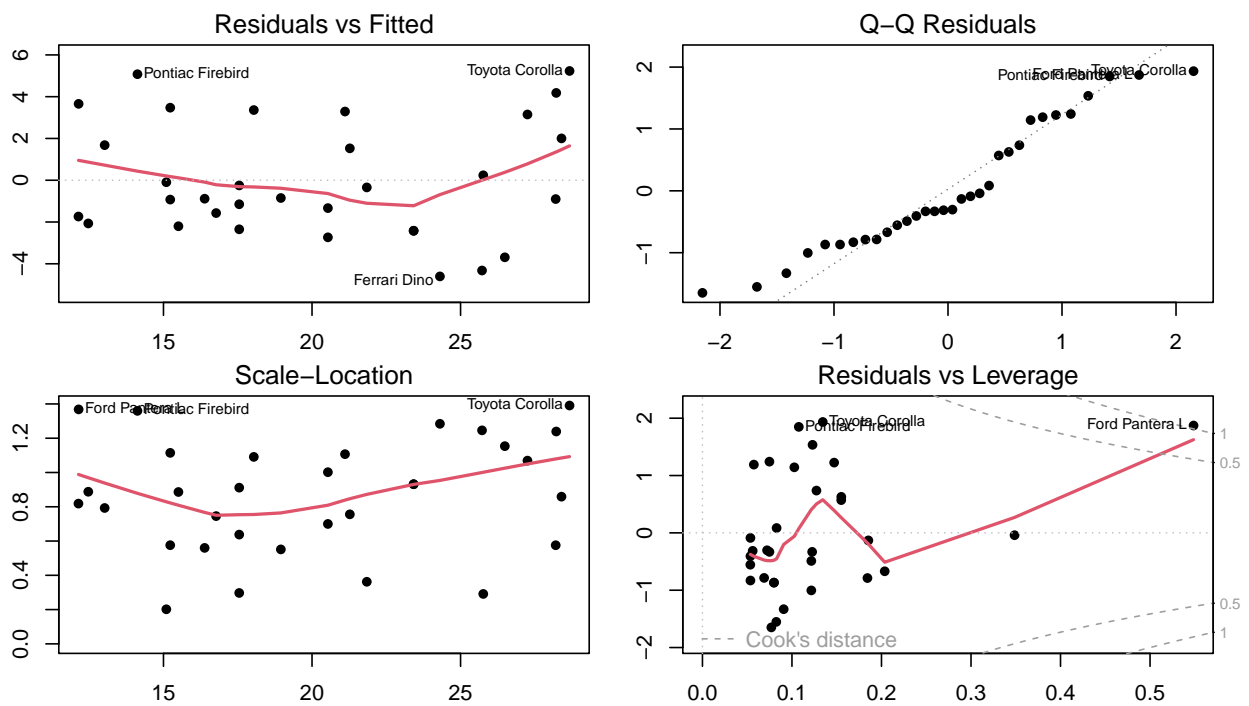


Figure 2: Residual plots for the anova model