

Case Study 00: This is a template

Author 1 and Author 2

March 00, 2015

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

Descrição do Problema

O IMC é uma medida utilizada na área de saúde como um indicador simples capaz de encontrar correlações entre o peso e altura do paciente com doenças decorrentes de obesidade, sendo inclusive capaz de classificar em diferentes grupos para obter maior precisão na análise em questão [ADD. REF].

Apesar dos problemas decorrentes da simplicidade do teste [ADD REF.] e métodos mais robustos tais como a bioimpedância [ADD REF.] estarem presentes, continua sendo uma medida interessante pela facilidade e baixo custo de encontrá-la. A partir de tal importância médica, muitas questões naturalmente surgem relacionadas a testes entre diferentes populações.

Neste estudo, serão propostos três testes de forma a compreender possíveis relações e diferenças presentes na amostragem realizada nos anos de 2016 e 2017 no Departamento de Engenharia da UFMG. Os testes propostos são:

- Diferença entre IMC de Homens e Mulheres;
- Diferença de IMC entre os dois anos em estudo;
- Diferença entre alunos de Graduação e Pós-Graduação em 2016.

Por meio destes testes, podemos obter informações relevantes capazes de compreender diferenças baseadas em idade, sexo e ano de estudo.

Design dos Experimentos

1. The easy way

Use RStudio as your editor, open the **.Rmd** file and click the *Knit PDF* button at the top of the editor.

2. The slightly-less-easy way

If you're using any other R editor (such as the basic R editor), you have to use the *render()* function from the **rmarkdown** package:

```
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("report_template.Rmd", "pdf_document") # this renders the pdf
```

Design dos Experimentos

Nos experimentos aqui realizados, procuramos extrair o máximo de informações possíveis de dados que já foram extraídos. A partir disto, problemas estatísticos naturalmente irão surgir, sendo alguns deles:

- Diferentes tamanhos de amostra;
- Poucos dados;
- Amostra não representativa de todo o departamento.

Apesar de tais problemas existirem, será aproveitada ao máximo a robustez existente nos experimentos.

Diferença entre IMC de Homens e Mulheres

Dada a suposta natureza do IMC das mulheres ser inferior aos homens [ADD REF.], torna-se relevante realizar a mesma análise dentro do universo de estudo. Para isso, podemos testar a alternativa de o imc masculino (μ_{imcm}) ser maior que o feminino (μ_{imcf}).

$$\begin{cases} H_0 : \mu_{imcf} = \mu_{imcm} \\ H_1 : \mu_{imcf} < \mu_{imcm} \end{cases}$$

Estudos referentes à normalidade de ambas distribuições foram realizados, assim como um teste de comparação das variâncias de Fligner-Killen.

O tamanho de (δ^*) utilizado neste teste baseou-se na relevância das faixas pré-definidas de IMC, sendo que para que pessoa no meio de uma faixa precisa ter uma alteração de 2.5 em seu IMC para alternar de faixa, sendo ($\delta^* = 2.5$).

O valor de $\alpha = 0.05$ buscou manter o padrão encontrado na literatura médica, assim como esperava-se encontrar um valor de $\beta = 0.8$ para corroborar com não encontrarmos falsos negativos.

Diferença de IMC entre os dois anos em estudo;

Realizamos para este teste dois experimentos, um para o sexo masculino e outro para o feminino. Esta estratificação baseou-se no mesmo critério da suposta diferença dos IMCs entre os sexos.

Para cada um dos testes, iremos avaliar se (μ_{2016}) é igual a (μ_{2017}).

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

Comentários referentes ao δ^* , α e β são análogos aos do primeiro experimento.

Diferença entre alunos de Graduação e Pós-Graduação em 2016;

Realizamos para este experimento um teste entre a diferença da média do IMC entre alunos de graduação (μ_{grad}) e pós graduação (μ_{pos}) de forma a testar se a idade é um fator que influencia a saúde dos alunos do departamento.

$$\begin{cases} H_0 : \mu_{grad} = \mu_{pos} \\ H_1 : \mu_{grad} \neq \mu_{pos} \end{cases}$$

Comentários referentes ao δ^* , α e β são análogos aos do primeiro experimento.

Experimentos

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"))
```

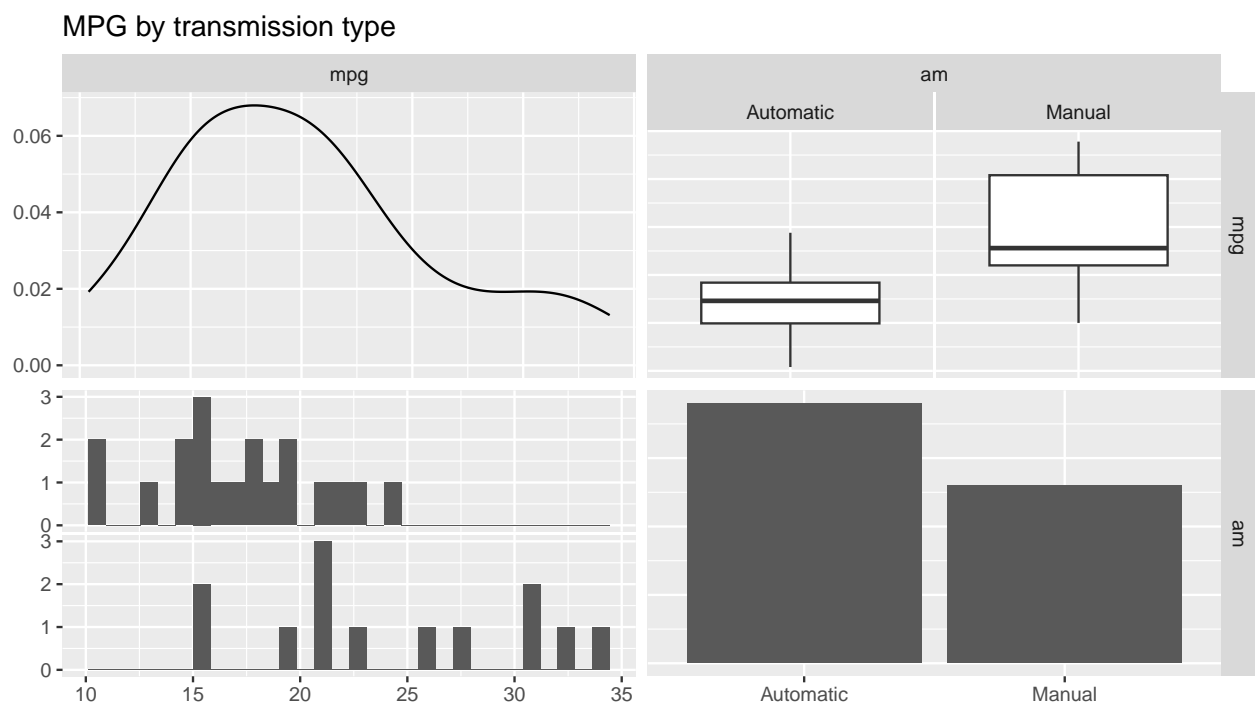


Figure 1: Exploring the effect of car transmission on mpg values

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
## ---
## 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)
```

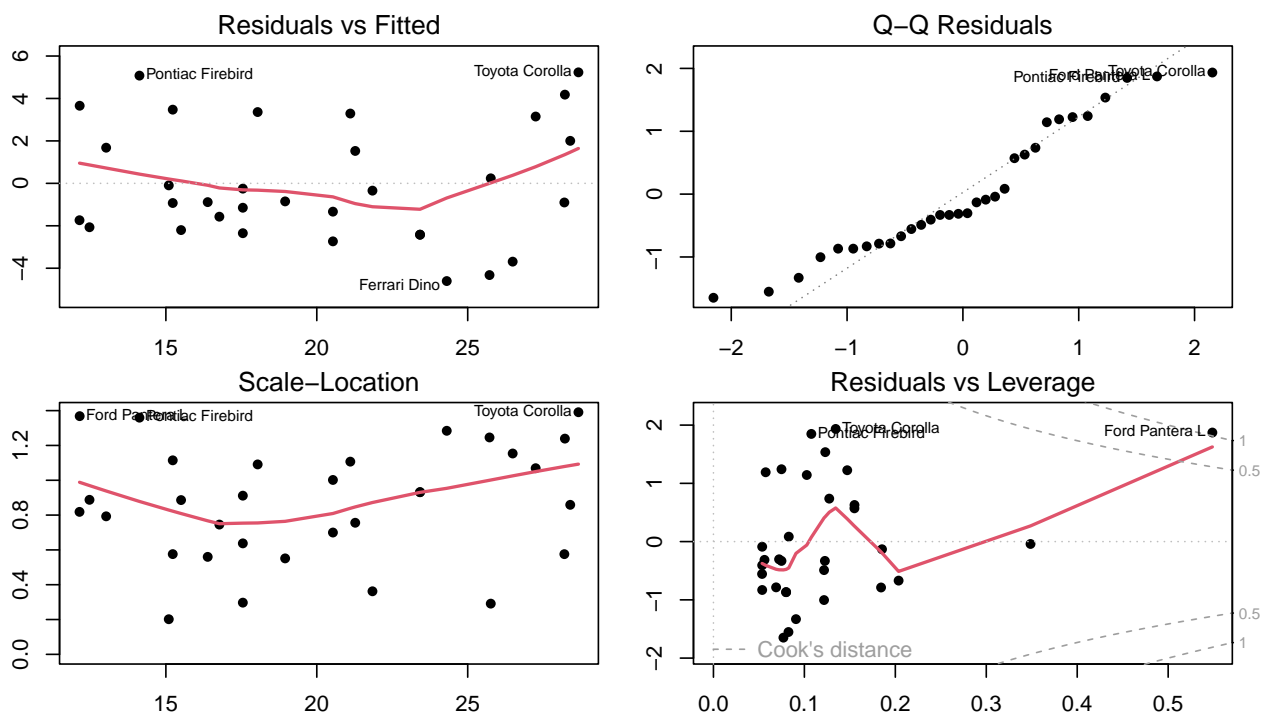


Figure 2: Residual plots for the anova model

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!