Case Study 01: This is a template

Hernane Braga and Lucas and Vinicius

January 19, 2021

Summary

This document provides a template for the Case Study reports. Reports should always start a short executive summary (call it an *Abstract* if you want) to give the reader a general idea of the topic under investigation, the kind of analysis performed, the results obtained, and the general recommendations of the authors.

There are (at least) two ways to render this document to **pdf** in R: one easy and one... well, less easy.

Experimental design

First Experiment

A section detailing the experimental setup. This is the place where you will define your test hypotheses, e.g.:

$$\begin{cases} H_0: \mu_1 \leq \mu_2\\ H_1: \mu_1 > \mu_2 \end{cases}$$

Second Experiment

A section detailing the experimental setup. This is the place where you will define your test hypotheses, e.g.:

$$\begin{cases} H_0: \mu_1 \leq \mu_2\\ H_1: \mu_1 > \mu_2 \end{cases}$$

Third Experiment

A section detailing the experimental setup. This is the place where you will define your test hypotheses, e.g.:

$$\begin{cases} H_0: \mu_1 \leq \mu_2\\ H_1: \mu_1 > \mu_2 \end{cases}$$

including the reasons behind your choices of the value for H_0 and the directionality (or not) of H_1.

This is also the place where you should discuss (whenever necessary) your definitions of minimally relevant effects (δ^*), sample size calculations, choice of power and significance levels, and any other relevant information about specificities in your data collection procedures.

Description of the data collection

The data was split in two datasets, the first containing data from students from 2016 class and the other from 2017.It was necessary clean data for getting the variables of weigth and heigth from PPGE students.

First experiment dataset

- Only man

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

```
# Loading dataset
dataset_2016 <- read.csv('data/imc_20162.csv', sep=',')
dataset_2017 <- read.csv('data/CS01_20172.csv', sep=';')


# Renaming columns
names(dataset_2017)[names(dataset_2017) == "Sex"] <- "Gender"
names(dataset_2017)[names(dataset_2017) == "height.m"] <- "Height.m"


# Selecting variables
model_var <- c("Height.m", "Weight.kg", "IMC")


# Calculating IMC
dataset_2016[,'IMC'] = dataset_2016[,'Weight.kg'] / (dataset_2016[,'Height.m'])^2
dataset_2017[,'IMC'] = dataset_2017[,'Weight.kg'] / (dataset_2017[,'Height.m'])^2


# Selecting PPGEE students
PPGEE_2016students <- dataset_2016[which(dataset_2016$Course == 'PPGEE'), ]


# Male students
male_students2016 <- PPGEE_2016students[which(PPGEE_2016students$Gender == 'M'), ]
male_students2017 <- dataset_2017[which(dataset_2017$Gender == 'M'), ]


# Female students
female_students2016 <- PPGEE_2016students[which(PPGEE_2016students$Gender == 'F'), ]
female_students2017 <- dataset_2017[which(dataset_2017$Gender == 'F'), ]


# Experiment dataset - Male x Female IMC
male_students <- rbind(male_students2016[model_var], male_students2017[model_var])
female_students <- rbind(female_students2016[model_var], female_students2017[model_var])
```

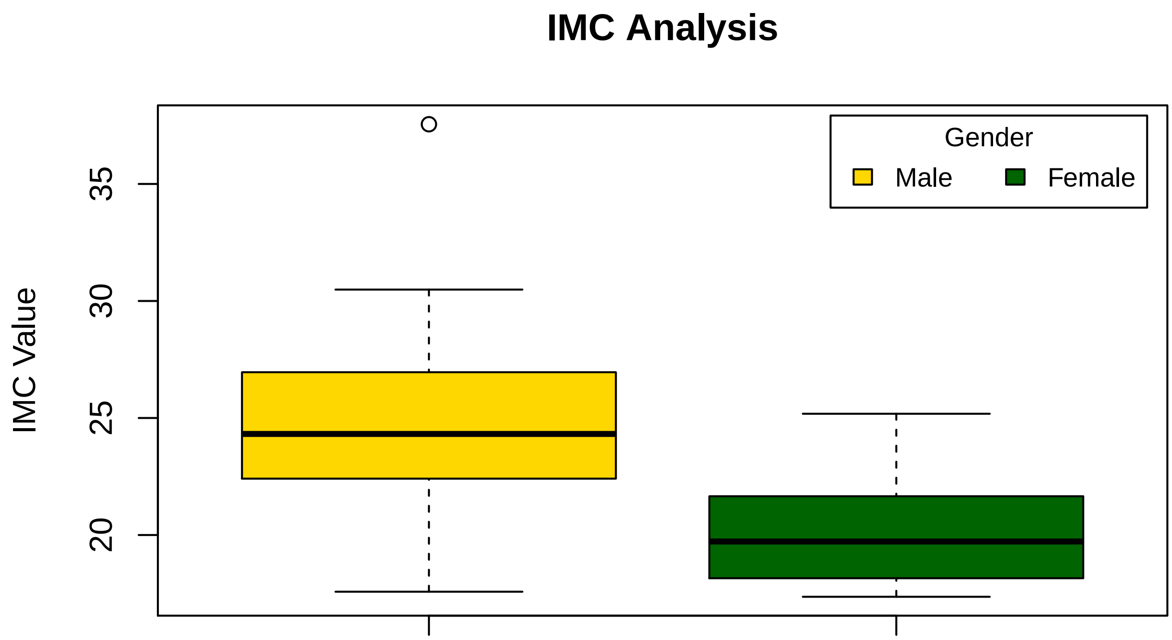
Exploratory Data Analysis

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

```
# Boxplot analysis

boxplot(male_students$IMC, female_students$IMC, main="IMC Analysis", col=(c("gold","darkgreen")),
        ylab="IMC Value")


legend("topright", inset=.02, title="Gender",
       c("Male","Female"), fill=(c("gold","darkgreen")), horiz=TRUE, cex=0.8)
```



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.

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)
```

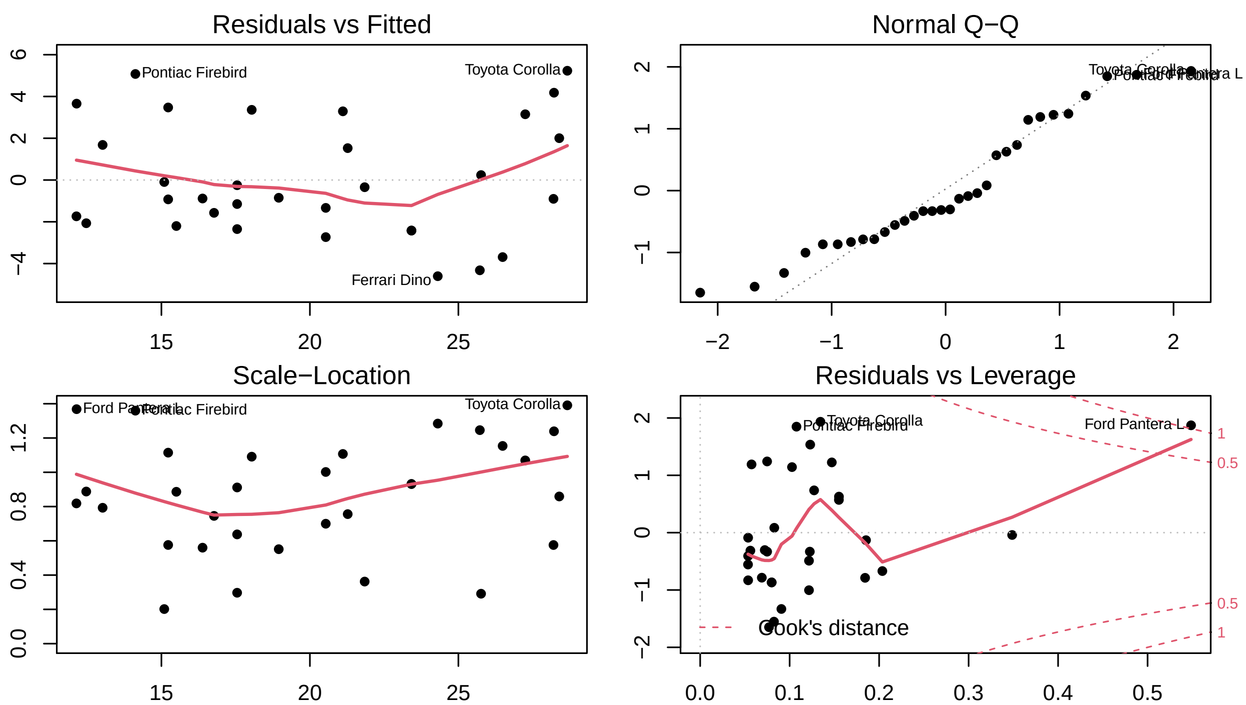


Figure 1: Residual plots for the anova model

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!