Multivariate analyze

Contents

- Introduction
- Principal outcomes
- Discussion
- Conclusions
- R code

Introduction

After cleaning and sorting the COVID-19 data, it is possible to merge and analyze it to reveal patterns and trends in the spread of COVID-19. the next step is to perform a multivariate analysis, which will help to generate hypotheses about the spread of COVID-19 in the future.

Principal outcomes

I merged 20 variables, including date, molecular positivity, number of samples for molecular and antigen tests, serological positivity, number of samples for serological tests, excess of SINADEF's deaths, MINSA's deaths, DIRESA+DIRIS's deaths, DIRESA+DIRIS's COVID-19 patients, non-free beds in ICUs, and vaccinated people. These variables were then clustered into four groups based on their affinity: infected people, deaths, available ICU beds (%), and vaccinated people.

Factorial Principal components (FPCs)

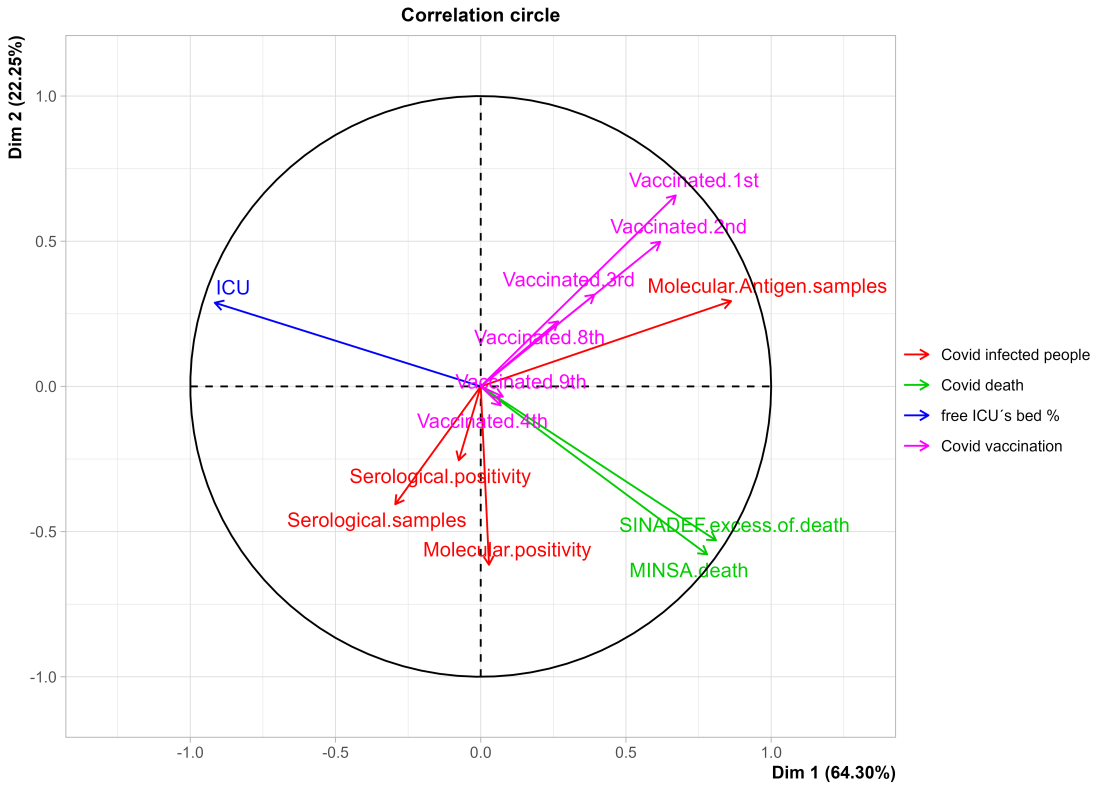

Figure 1. Factorial analysis of COVID-19:infected people, deaths, free bed(%) and vaccinated people

The first principal component, DIM1, which explains approximately 64% of the variance in the data, is interpreted as a signal of the effect of COVID-19 mortality on people in intensive care units (ICUs). DIM1 has the highest loading for variables such as the percentage of free ICU beds, the number of COVID-19 deaths, and the number of COVID-19 infected people. The second principal component, DIM2, which explains approximately 22% of the variance in the data, is interpreted as a signal of the decreasing effect of COVID-19 mortality due to vaccination. DIM2 has the highest loading for variables such as the number of free vaccinated people and the number of COVID-19 deaths.

Discussion

In fig.1, it is highlighted that there is an inverse relationship between the number of COVID-19 deaths and the number of free ICU beds, as well as between COVID-19 positivity and the number of vaccinated people. This suggests that the number of COVID-19 deaths and positivity is modulated by the number of free ICU beds and vaccination, respectively. In other words, the more free ICU beds there are, the lower the number of COVID-19 deaths, and the higher the number of vaccinated people, the lower the COVID-19 positivity.

Conclusions

The four principal components (FPCs) account for 87% of the data variance and help to reduce the number of variables, improving the interpretation of the data.

R code

Bellow I attached a R-script. Contact Us here, if you consider to give opinions, suggestions and questions.

```
################################################################################
#to start

library("magrittr")
library("plyr")
library("dplyr")
library("janitor")#paquetes que se deben instalar
library("ggplot2")
library("viridis")
library("scales")
library("stringr")
library("gganimate")
library("transformr")
library("gifski")
library("data.table")
library("pracma")
library("tidyr")
library("bestNormalize")
library("FactoMineR")
library("vegan")
library("vars")
library("forecast")
library("mgcv")
library("mgcViz")
library("itsadug")
library("visreg")
library("gridExtra")
library("magrittr")
library("devtools")
library("gamm4")
```

```
library("tidymv")

setwd("~/COVID19/")#directory

d=read.csv("d.csv",sep=",",dec=".",header=TRUE)

colnames(d)=c("X","dates",
              "Molecular positivity","Molecular&Antigen samples",
              "Serological positivity","Serological samples",
              "SINADEF excess of death","MINSA death",
              "DIRESA & DIRIS death","DIRESA & DIRIS infected",
              "free ICU´s bed %",
              "Vaccinated-1st","Vaccinated-2nd",
              "Vaccinated-3rd","Vaccinated-4th",
              "Vaccinated-5th","Vaccinated-6th",
              "Vaccinated-7th","Vaccinated-8th",
              "Vaccinated-9th","Vaccinated-10th",
              "Vaccinated-11th")

#d$`Molecular positivity`[c(which(d$dates=="2022-07-26")+1):length(d$`Molecular positivity`)]=NA
View(d)
dd=d[1:1125,] #03-04-2023

Positivity=dd[,c(3,4,5,6)]
fallecidos=dd[,7:8]
ICU=dd[,c(11)]*100
vaccination=dd[,c(12:22)]

tab<- data.frame(Positivity,fallecidos,ICU,vaccination)
tab=na.omit(tab)
gr<- c(ncol(Positivity),ncol(fallecidos),1,ncol(vaccination))


# Compute the MFA without multiple plots
t.mfa <- MFA(tab,
             group = gr,
             type = c("c","s","s","c"),
             ncp =,
             name.group = c("Covid infected people","Covid death","free ICU´s bed %","Covid vaccination
             graph =FALSE)

# Plot the results
MFA1=plot(t.mfa,
          choix = "axes",
          habillage = "group",
          shadowtext = TRUE)

ggsave("AMV.dimesiones.png", dpi = 600,   width = 250,
       height = 159,unit="mm",plot =MFA1)

#x11();plot(
#t.mfa,
#choix = "ind",
```

```
#partial = "all",
#habillage = "group")

MFA2=plot(t.mfa,
          choix = "var",
          habillage = "group",
          shadowtext =TRUE)

ggsave("AMV.biplot.png", dpi = 600,   width = 250,
       height = 159,unit="mm",plot =MFA2)

MFA3=plot(t.mfa, choix = "group")
ggsave("AMV.grupos.png", dpi = 600,   width = 250,
       height = 159,unit="mm",plot =MFA3)
# Eigenvalues, scree plot and broken stick model
#ev<- t.mfa$eig[, 1]
#names(ev) <- paste("MFA", 1 : length(ev))
#x11();screestick(ev, las = 3)

# RV coefficients with tests (p-values above the diagonal of
# the matrix)
pvalue <- t.mfa$group$RV
pvalue

##############################################################################################
```