Guión de Scripts

Giovany Babativa-Márquez, Ph.D.

Enero 2023

Tabla de Contenido

# 1 PROTEINAS

######========================================================  
###### Script por Giovany Babativa-Márquez, PhD.   
######========================================================  
#== Use el conjunto de datos sobre consumo de proteínas en varios   
#== países para realizar un análisis HJ-Biplot,  
#== Utilice los paquetes  
#== 1. MultBiplotR  
#== 2. BiplotGUI  
#============================================================  
  
rm(list = ls())  
  
#.... Paquetes.  
  
library(pacman)  
  
p\_load(readxl, tidyverse, MultBiplotR, BiplotML, janitor, here, skimr, foreign,  
 factoextra, FactoMineR, Factoshiny, FactoInvestigate,  
 explor, BiplotGUI,  
 corrplot, patchwork,  
 RSpectra)  
  
  
######-------------------- Ejemplo Proteinas  
  
proteinas <- read\_excel(here::here("data/Consumo Proteinas.xlsx")) |>   
 clean\_names()  
  
glimpse(proteinas)  
  
datos <- proteinas |>   
 column\_to\_rownames(var = "pais") |>   
 select(where(is.numeric))  
  
  
skim(datos)  
  
#-- Análisis HJ-Biplot usando el paquete MultBiplotR  
  
  
hjProteinas <- MultBiplotR::HJ.Biplot(datos)  
  
hjProteinas$Inertia  
  
  
data.frame(value=hjProteinas$Inertia) |>   
 mutate(Eje = 1:n()) |>   
 ggplot(aes(x = Eje, y = value)) +  
 geom\_col(fill = "lightblue") +  
 labs(y = "% Inercia") +  
 geom\_text(aes(label=round(value, 1)), position=position\_dodge(width=0.9), vjust=-0.25) +  
 theme\_bw()   
  
  
hjProteinas$ColContributions  
  
corrplot(hjProteinas$ColContributions, is.corr = F) # method = "pie"  
  
hjProteinas$RowContributions  
  
  
corrplot(hjProteinas$RowContributions, is.corr = F)   
  
  
summary(hjProteinas)  
  
plot(hjProteinas)  
CorrelationCircle(hjProteinas)  
  
#####---- Biplot Interactivo usando el paquete BiplotGUI  
  
BiplotGUI::Biplots(datos)

# 2 JUGADORES DE FÚTBOL

######========================================================  
###### Script por Giovany Babativa-Márquez, PhD.   
######========================================================  
#== Use el conjunto de datos sobre el rendimiento de los jugadores   
#== profesionales de fútbol para realizar un análisis multivariante   
#== y un análisis clúster. Utilice los paquetes  
#==  
#== Utilice los paquetes  
#== 1. FactoMineR y factoextra  
#== 2. explor  
#== 3. Factoshiny  
#============================================================  
  
rm(list = ls())  
  
  
library(pacman)  
  
p\_load(readxl, tidyverse, MultBiplotR, BiplotML, janitor, here, skimr, foreign,  
 factoextra, FactoMineR, Factoshiny, FactoInvestigate,  
 explor, BiplotGUI,  
 corrplot, patchwork,  
 RSpectra)  
  
#----- Ejemplo jugadores.  
  
jugadores <- read\_excel(here::here("data/jugadores.xlsx")) |>   
 clean\_names()  
  
glimpse(jugadores)  
  
skim(jugadores)  
  
datosJ <- jugadores |>   
 column\_to\_rownames(var = "nombre") |>   
 select(where(is.numeric), -id)  
  
############################ Análisis de componentes principales  
  
comprin <- PCA(datosJ , scale.unit=TRUE, ncp=5, graph = FALSE)  
  
fviz\_screeplot(comprin, ncp=10, title="Valores propios", ylab="Porcentaje de varianza", xlab="Eje")  
  
comprin$eig  
  
g3 <- fviz\_pca\_ind(comprin, col.ind="cos2", title="Proyección de los jugadores") +  
 scale\_color\_gradient2(low="white", mid="blue",   
 high="red", midpoint=0.50)  
g3  
  
  
g4 <- fviz\_pca\_var(comprin, col.var="contrib", axes=c(1,2), title="Proyección de las habilidades")+  
 scale\_color\_gradient2(low="white", mid="blue",   
 high="red", midpoint=2)+theme\_bw()  
  
g4  
  
jugadores$posicion <- as.factor(jugadores$posicion)  
  
dev.new()  
fviz\_pca\_ind(comprin, label="none", habillage=jugadores$posicion, title="Jugadores por posición")  
  
  
fviz\_pca\_biplot(comprin)  
  
View(comprin$ind$coord)  
View(comprin$var$coord)  
  
library(explor)  
explor(comprin)  
  
  
######### Factoshiny  
  
res <- Factoshiny(datosJ)

# 3 POSICIONAMIENTO CARROS

######========================================================  
###### Script por Giovany Babativa-Márquez, PhD.   
######========================================================  
#== El conjunto de datos de marcas y atributos de los carros contiene   
#== la percepción de 1000 personas mayores de 25 años propietarias de   
#== vehículos. Realice un análisis multivariante que permita identificar   
#== el posicionamiento de las marcas.  
#==  
#== Utilice los paquetes  
#== 1. FactoMineR  
#== 2. explor  
#== 3. Factoshiny  
#============================================================  
  
rm(list = ls())  
  
  
library(pacman)  
  
p\_load(readxl, tidyverse, MultBiplotR, BiplotML, janitor, here, skimr, foreign,  
 factoextra, FactoMineR, Factoshiny, FactoInvestigate,  
 explor, BiplotGUI,  
 corrplot, patchwork,  
 RSpectra)  
  
#----- Ejemplo marcas de carros.  
#------ CA Simple  
  
datos2 <- read.spss("data/CarrosREES.sav", to.data.frame = T)  
  
glimpse(datos2)  
  
datos <- datos2 |>   
 select(MARCAR, Atributo, Frecuencia) |>   
 pivot\_wider(names\_from = Atributo, values\_from = Frecuencia) |>   
 column\_to\_rownames(var = "MARCAR")  
  
dev.off()  
res <- CA(datos, graph = F)  
  
res$eig  
corrplot(res$row$cos2)  
corrplot(res$row$coord, is.corr = FALSE)  
  
corrplot(res$col$cos2)  
corrplot(res$col$coord, is.corr = FALSE)  
  
  
out <- HCPC(res, nb.clust = 3)  
  
out$desc.var  
  
  
sout <- Factoshiny(datos)  
  
library(explor)  
explor(res)

# 4 MANOVA BIPLOT

#-------------------------------------------  
  
rm(list = ls())  
  
  
library(pacman)  
  
p\_load(readxl, tidyverse, MultBiplotR, BiplotML, janitor, here, skimr, foreign,  
 factoextra, FactoMineR, Factoshiny, FactoInvestigate,  
 explor, BiplotGUI,  
 corrplot, patchwork,  
 RSpectra)  
  
  
#----- MANOVA Biplot  
  
vinos <- read.spss("data/vinos.sav", to.data.frame = T) |>   
 janitor::clean\_names()  
  
table(vinos$denomina, vinos$a\_o)  
  
glimpse(vinos)  
  
X <- as.matrix(vinos[,4:21])  
  
#---- Primero haremos una análisis multivariante de la varianza  
?manova  
  
manvin <- manova(X ~ vinos$grupo)  
summary(manvin) # Se rechaza Ho  
  
summary(manvin, test = "Wilks")  
summary(manvin, test = "Hotelling")  
summary(manvin, test = "Roy")  
  
#---- Representación multivariante -- Manova Biplot  
?CanonicalBiplot  
canbip <- CanonicalBiplot(X, group=vinos$grupo)  
summary(canbip)  
  
plot(canbip)  
plot(canbip, mode="s")

# 5 BIPLOT LOGÍSTICO - SPIDERS

######========================================================  
###### Script por Giovany Babativa-Márquez, PhD.   
######========================================================  
#--- El paquete MultBiplotR contiene el conjunto de datos “spiders” el cual   
#--- contiene 28 sitios de muestreo donde se identifica la presencia o ausencia   
#--- de algunas especies de arañas. Realice un biplot logístico para identificar   
#--- las especies que son más probables en los mismos sitios y concluya sobre las   
#--- diferencias que se observen.  
#============================================================  
  
rm(list = ls())  
  
  
library(pacman)  
  
p\_load(readxl, tidyverse, MultBiplotR, BiplotML, janitor, here, skimr, foreign,  
 factoextra, FactoMineR, Factoshiny, FactoInvestigate,  
 explor, BiplotGUI,  
 corrplot, patchwork,  
 RSpectra)  
  
  
#-------.  
data("spiders")  
?spiders  
  
datos <- spiders |>   
 mutate(across(everything(), ~ifelse(. == "Absent", 0, 1)))  
  
  
glimpse(datos)  
  
outLB <- LogBip(x = datos, method = "MM")  
  
View(outLB$Ahat)

# 6 BIPLOT LOGÍSTICO - DATOS EPIGENETICOS

#============================================================  
  
rm(list = ls())  
  
  
library(pacman)  
  
p\_load(readxl, tidyverse, MultBiplotR, BiplotML, janitor, here, skimr, foreign,  
 factoextra, FactoMineR, Factoshiny, FactoInvestigate,  
 explor, BiplotGUI,  
 corrplot, patchwork,  
 RSpectra)  
  
  
#-------.  
   
load("data/xMethy.rda")  
  
  
cvMet\_MM <- cv\_LogBip(data = xMethy |>   
 select(-`Cancer Type`), k=0:5, method = "MM")  
  
bipMethy\_MM <- LogBip(x = xMethy |>   
 select(-`Cancer Type`),  
 k = 2,  
 method = "MM",  
 col.ind = xMethy$`Cancer Type`)  
  
PlotMet\_MM <- plotBLB(bipMethy\_MM, xylim = c(-100,100),  
 col.ind = xMethy$`Cancer Type`, repel = T )+  
 labs(title = "Biplot Logístico",  
 subtitle = "Estimación con el algoritmo MM-BCD",  
 caption = "Gráfico elaborado por Giovany Babativa") +  
 theme\_void()   
  
  
PlotMet\_MM  
  
  
Pi <- fitted\_LB(bipMethy\_MM, type = "response")  
  
View(Pi)

# 7 ANÁLISIS FACTORIAL MÚLTIPLE

rm(list = ls())  
  
library(FactoMineR)  
library(factoextra)  
library(tidyverse)  
  
data(wine)  
colnames(wine)  
  
glimpse(wine)  
  
res.mfa <- MFA(wine, group = c(2, 5, 3, 10, 9, 2),   
 type = c("n", "s", "s", "s", "s", "s"),  
 name.group = c("origen","olor.pre.agi","asp.visual",  
 "olor.pos.agi", "sabor","desemp.gral"),  
 num.group.sup = c(1, 6),  
 graph = FALSE)  
  
print(res.mfa)  
eig.val = get\_eigenvalue(res.mfa)  
head(eig.val)  
  
fviz\_screeplot(res.mfa)  
  
grupos <- get\_mfa\_var(res.mfa, "group")  
grupos  
  
# Coordenadas de los grupos  
head(grupos$coord)  
  
# Cos2: calidad de la representación en el mapa factorial  
head(grupos$cos2)  
  
# Contribución a las dimensiones  
head(grupos$contrib)  
  
# Grafico para el grupo de variables  
fviz\_mfa\_var(res.mfa, "group")  
  
quanti.var = get\_mfa\_var(res.mfa, "quanti.var")  
quanti.var  
  
# Coordinates  
head(quanti.var$coord)  
# Cos2: quality on the factore map  
head(quanti.var$cos2)  
# Contributions to the dimensions  
head(quanti.var$contrib)  
  
fviz\_mfa\_var(res.mfa, "quanti.var", palette = "jco",   
 col.var.sup = "violet", repel = TRUE)  
  
fviz\_mfa\_var(res.mfa, "quanti.var", palette = "jco",   
 col.var.sup = "violet", repel = TRUE,  
 geom = c("point", "text"), legend = "bottom")  
  
# Contributions to dimension 1  
fviz\_contrib(res.mfa, choice = "quanti.var", axes = 1, top = 20,  
 palette = "jco")  
  
# Contributions to dimension 2  
fviz\_contrib(res.mfa, choice = "quanti.var", axes = 2, top = 20,  
 palette = "jco")  
  
  
g1 =fviz\_mfa\_var(res.mfa, "quanti.var", col.var = "cos2",   
 gradient.cols = c("white", "blue", "red"),   
 col.var.sup = "violet", repel = TRUE,  
 geom = c("point", "text"))  
  
g1  
  
#### Grafica de los individuos  
  
ind = get\_mfa\_ind(res.mfa)  
  
g2 = fviz\_mfa\_ind(res.mfa, col.ind = "cos2",   
 gradient.cols = c("white", "blue", "red"),  
 repel = TRUE)  
  
g2  
  
library(gridExtra)  
  
dev.new()  
grid.arrange(g1, g2, nrow=1)  
  
  
###La primera dimensión representa a la intensidad...   
###1 DAM es el más intenso y armonioso, contrario a 1VAU y 2ING  
  
#### El eje 2 esta asociado a los vinos T2 y T1 caracterizados  
#### fuerte valor en Spice antes de agitar  
  
### la mayoría de variables cualitativas están sobre el origen  
### mostrando que no hay asociación con la intensidad y armonia,o  
### con el segundo eje  
  
### Aunque la categoria Soil=REFERENCE se relaciona con vinos  
### de calidad, frutas, armoniosos como 1ING o 1POY  
  
  
table(wine$Label)  
  
fviz\_mfa\_ind(res.mfa,   
 habillage = "Label", # color by groups   
 palette = c("#00AFBB", "#E7B800", "#FC4E07"),  
 addEllipses = TRUE, ellipse.type = "confidence",   
 repel = TRUE # Avoid text overlapping  
)   
  
table(wine$Soil)  
  
#### Grafico para multiples variables categoricas  
fviz\_ellipses(res.mfa, c("Label", "Soil"), repel = TRUE)  
  
  
##### Grafica parcial de individuos  
fviz\_mfa\_ind(res.mfa, partial = "all")   
  
### Solo algunos  
  
fviz\_mfa\_ind(res.mfa, partial = c("1DAM", "1VAU", "2ING"))  
  
###El vino 1DAM ha sido descrito   
###como particularmente "intenso" y "armonioso",   
### note el grupo de olores pre: tiene una alta coordenada   
#### en el primer eje desde el punto de vista   
  
  
### Desde el grupo de olores, 2ING fue más "intenso" y "armonioso"   
## que 1VAU, pero desde el punto de vista del grupo de sabor  
### 1VAU fue más "intenso" y "armonioso" que 2ING.  
  
##### Grafic de ejes parciales  
  
fviz\_mfa\_axes(res.mfa)  
  
  
#### Ahora hagamos el ejercicio con Factoshiny  
  
library(Factoshiny)  
out <- Factoshiny::Factoshiny(wine)