

Análisis Canónico Flores

Melissa Ortega Galarza

2022-05-26

Instalar paqueterias

```
install.packages("tidyverse",dependencies = T)
library(tidyverse)
```

Preparación de la matriz

Importar la matriz de datos.

```
G<-data.frame(datos::flores)
```

Exploración de la matriz

```
dim(G)
```

```
## [1] 150  5
```

```
colnames(G)
```

```
## [1] "Largo.Sepalo" "Ancho.Sepalo" "Largo.Petalo" "Ancho.Petalo" "Especie"
```

```
str(G)
```

```
## 'data.frame':  150 obs. of  5 variables:
## $ Largo.Sepalo: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
## $ Ancho.Sepalo: num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
## $ Largo.Petalo: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
## $ Ancho.Petalo: num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
## $ Especie      : Factor w/ 3 levels "setosa","versicolor",...: 1 1 1 1 1 1 1 1 1 1 ...
```

```
anyNA(G)
```

```
## [1] FALSE
```

Escalamiento de la matriz

```
# Generacion de variables X
X <- G %>%
  select(Largo.Sepalo,Ancho.Sepalo ) %>%
  scale()
head(X)
```

```
##   Largo.Sepalo Ancho.Sepalo
## 1   -0.8976739   1.01560199
## 2   -1.1392005  -0.13153881
## 3   -1.3807271   0.32731751
## 4   -1.5014904   0.09788935
## 5   -1.0184372   1.24503015
## 6   -0.5353840   1.93331463
```

```
G$Largo.Sepalo
```

```
##   [1] 5.1 4.9 4.7 4.6 5.0 5.4 4.6 5.0 4.4 4.9 5.4 4.8 4.8 4.3 5.8 5.7 5.4 5.1
##  [19] 5.7 5.1 5.4 5.1 4.6 5.1 4.8 5.0 5.0 5.2 5.2 4.7 4.8 5.4 5.2 5.5 4.9 5.0
##  [37] 5.5 4.9 4.4 5.1 5.0 4.5 4.4 5.0 5.1 4.8 5.1 4.6 5.3 5.0 7.0 6.4 6.9 5.5
##  [55] 6.5 5.7 6.3 4.9 6.6 5.2 5.0 5.9 6.0 6.1 5.6 6.7 5.6 5.8 6.2 5.6 5.9 6.1
##  [73] 6.3 6.1 6.4 6.6 6.8 6.7 6.0 5.7 5.5 5.5 5.8 6.0 5.4 6.0 6.7 6.3 5.6 5.5
##  [91] 5.5 6.1 5.8 5.0 5.6 5.7 5.7 6.2 5.1 5.7 6.3 5.8 7.1 6.3 6.5 7.6 4.9 7.3
## [109] 6.7 7.2 6.5 6.4 6.8 5.7 5.8 6.4 6.5 7.7 7.7 6.0 6.9 5.6 7.7 6.3 6.7 7.2
## [127] 6.2 6.1 6.4 7.2 7.4 7.9 6.4 6.3 6.1 7.7 6.3 6.4 6.0 6.9 6.7 6.9 5.8 6.8
## [145] 6.7 6.7 6.3 6.5 6.2 5.9
```

Generación de variables Y

```
Y <- G %>%
  select(Largo.Petalo,Ancho.Petalo) %>%
  scale()
head(Y)
```

```
##   Largo.Petalo Ancho.Petalo
## 1   -1.335752   -1.311052
## 2   -1.335752   -1.311052
## 3   -1.392399   -1.311052
## 4   -1.279104   -1.311052
## 5   -1.335752   -1.311052
## 6   -1.165809   -1.048667
```

Análisis canonico con un par de variables

Libreria

```
install.packages("CCA")  
library("CCA")
```

Análisis

```
ac<-cancor(X,Y)
```

Visualización de la matriz X

```
ac$xcoef  
  
##                [,1]      [,2]  
## Largo.Sepalo -0.07251736 0.03932826  
## Ancho.Sepalo  0.03052965 0.07663824
```

Visualización de la matriz Y

```
ac$ycoef  
  
##                [,1]      [,2]  
## Largo.Petalo -0.12279948 -0.2774814  
## Ancho.Petalo  0.04332444 0.3003309
```

Visualización de la correlacion canonica

```
ac$cor  
  
## [1] 0.9409690 0.1239369
```

Obtención de la matriz de variables canonicas se obtiene multiplicando los coeficientes por cada una de las variables (X1 y Y1)

```
ac1_X <- as.matrix(X) %*% ac$xcoef[, 1]  
ac1_Y <- as.matrix(Y) %*% ac$ycoef[, 1]
```

Visualización de los primeros 20 datos

```
ac1_X[1:20,]
```

```
##           1           2           3           4           5           6           7
## 0.09610292 0.07859597 0.11011957 0.11187264 0.11186471 0.09784806 0.13288573
##           8           9          10          11          12          13          14
## 0.09785599 0.11537879 0.08560034 0.08383933 0.11537086 0.08735341 0.13114059
##          15          16          17          18          19          20
## 0.06982268 0.10659757 0.09784806 0.09610292 0.06457139 0.11711600
```

```
ac1_Y[1:20,]
```

```
##           1           2           3           4           5           6           7
## 0.10722900 0.10722900 0.11418530 0.10027270 0.10722900 0.09772779 0.11291285
##           8           9          10          11          12          13          14
## 0.10027270 0.10722900 0.09458885 0.10027270 0.09331640 0.10154515 0.12241406
##          15          16          17          18          19          20
## 0.12114160 0.11164040 0.12555300 0.11291285 0.09204394 0.10595655
```

Correlación canónica entre variable X1 y Y1

```
cor(ac1_X,ac1_Y)
```

```
##           [,1]
## [1,] 0.940969
```

Verificación de la correlacion canonica

```
assertthat::are_equal(ac$cor[1],
                       cor(ac1_X,ac1_Y)[1])
```

```
## [1] TRUE
```

Análisis canonico con dos pares de variables

Calculo de las variables X2 y Y2

```
ac2_X <- as.matrix(X) %*% ac$xcoef[, 2]
ac2_Y <- as.matrix(Y) %*% ac$ycoef[, 2]
```

Agregamos las variables generadas a la matriz

```
ac_df <- G %>%
  mutate(ac1_X=ac1_X,
         ac1_Y=ac1_Y,
         ac2_X=ac2_X,
         ac2_Y=ac2_Y)
```

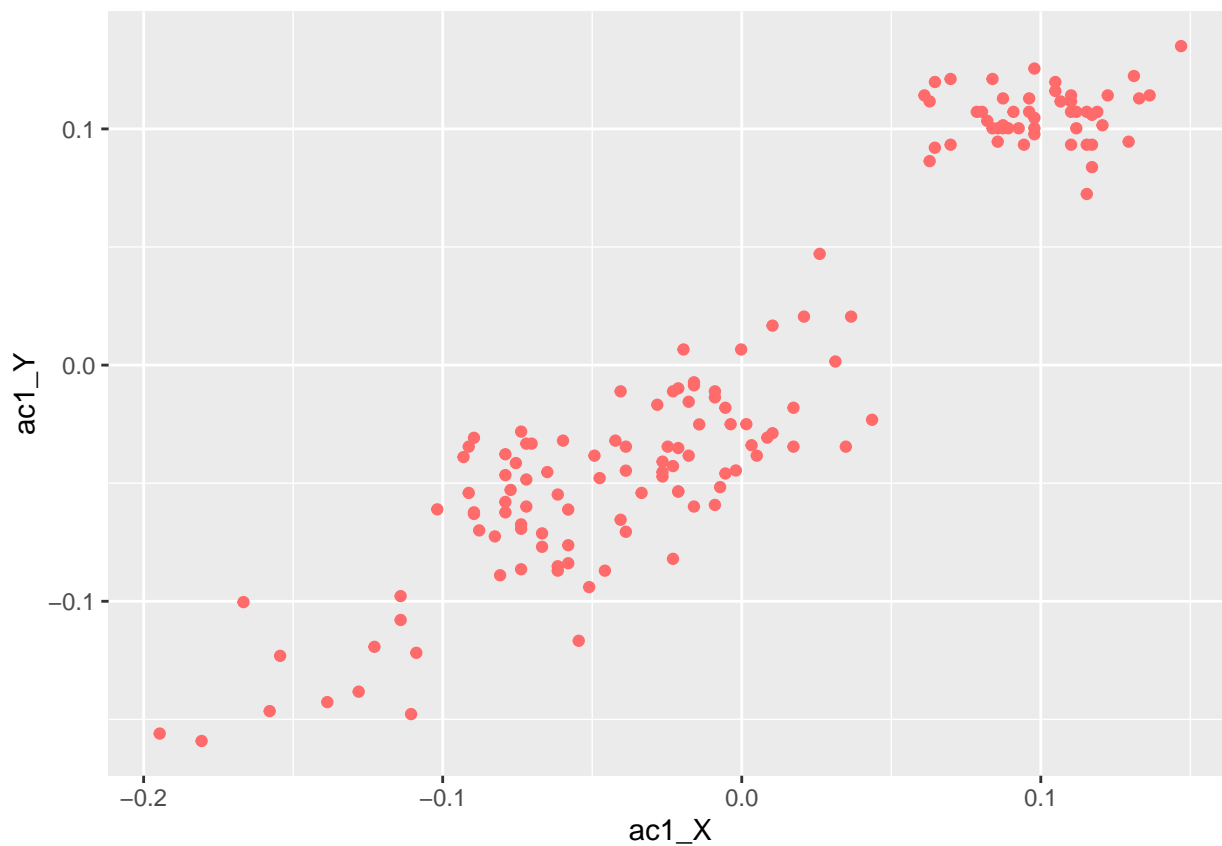
Visualización de los nombres de las variables

```
colnames(ac_df)
```

```
## [1] "Largo.Sepalo" "Ancho.Sepalo" "Largo.Petalo" "Ancho.Petalo" "Especie"  
## [6] "ac1_X"        "ac1_Y"        "ac2_X"        "ac2_Y"
```

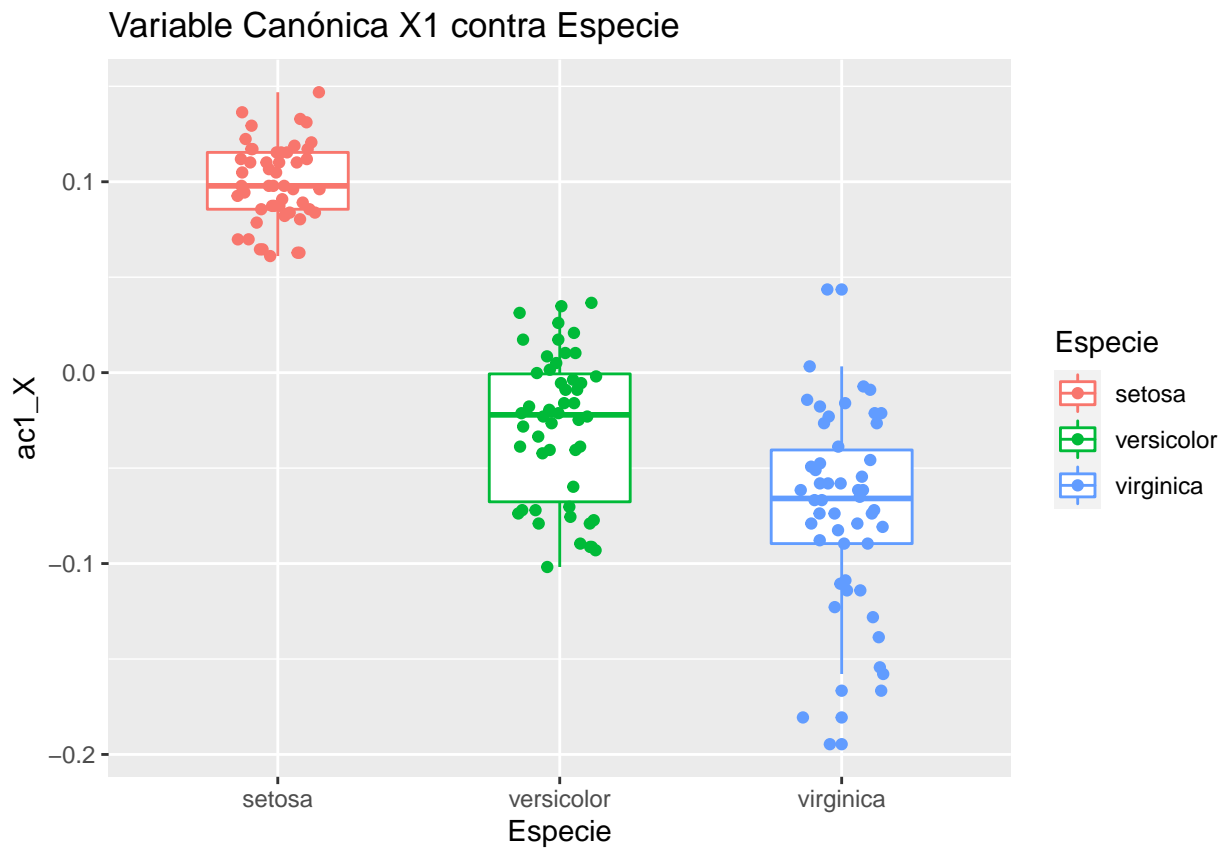
Generación del grafico scatter plot para la visualización de X1 y Y1

```
ac_df %>%  
  ggplot(aes(x=ac1_X,y=ac1_Y))+  
  geom_point(color="indianred1")
```



Generación de un boxplot

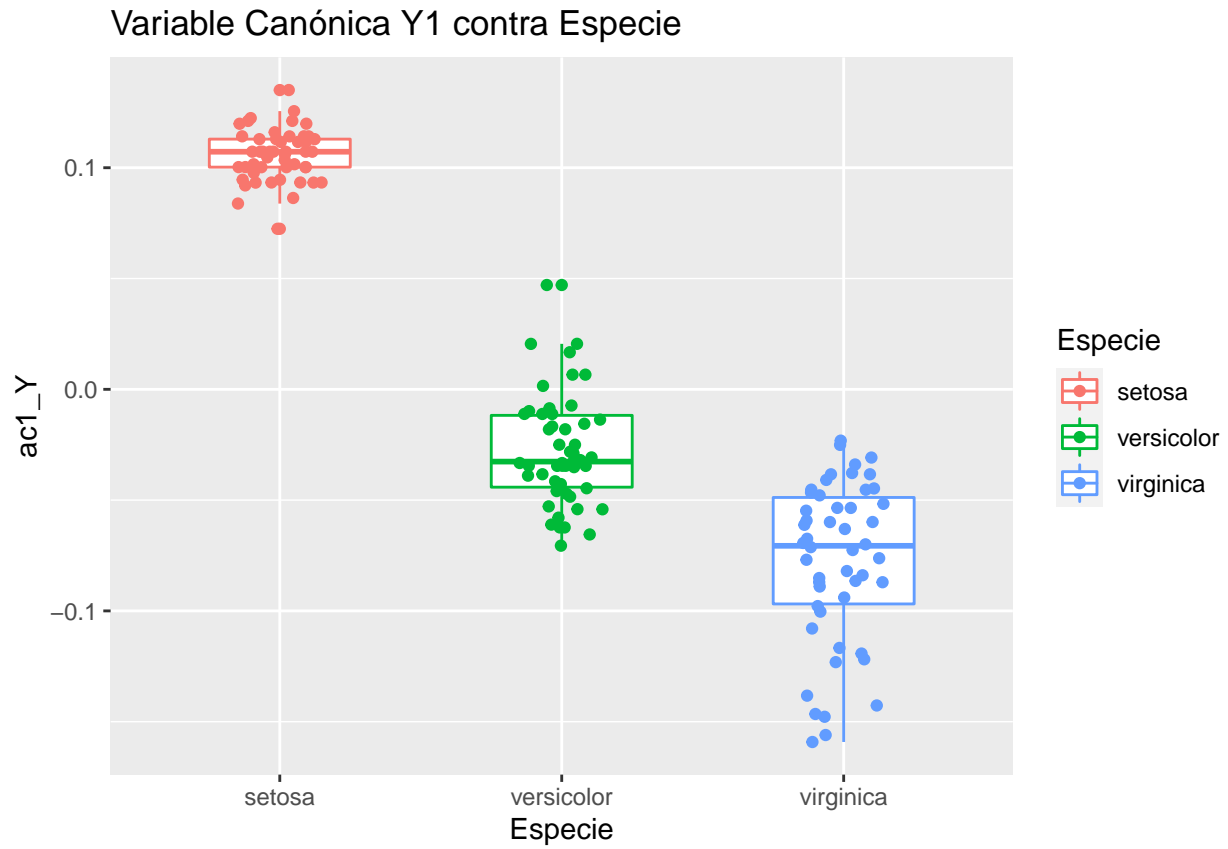
```
ac_df %>%  
  ggplot(aes(x=Especie,y=ac1_X, color=Especie))+  
  geom_boxplot(width=0.5)+  
  geom_jitter(width=0.15)+  
  ggtitle("Variable Canónica X1 contra Especie")
```



Interpretación:

Se observa una correlación entre la variable canónica X1 y la variable latente Especie.

```
ac_df %>%
  ggplot(aes(x=Especie,y=ac1_Y, color=Especie))+
  geom_boxplot(width=0.5)+
  geom_jitter(width=0.15)+
  ggtitle("Variable Canónica Y1 contra Especie")
```



```
ac_df %>%
  ggplot(aes(x=ac1_X,y=ac1_Y, color=Especie))+
  geom_point()+
  ggtitle("Variable Canónica X1 contra Y1")
```

