

Análisis Canónico

Oscar Elí Bonilla Morales

2022-05-24

#Análisis Canonico

Preparacion de la matriz

Nota:

Se utilizará la matriz “Flores”, dataset ubicado dentro del paquete “datos”

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

install.packages("tidyverse")
library(tidyverse)

datos <- flores
head(datos)
```

```
##   Largo.Sepalo Ancho.Sepalo Largo.Petalo Ancho.Petalo Especie
## 1          5.1          3.5          1.4          0.2  setosa
## 2          4.9          3.0          1.4          0.2  setosa
## 3          4.7          3.2          1.3          0.2  setosa
## 4          4.6          3.1          1.5          0.2  setosa
## 5          5.0          3.6          1.4          0.2  setosa
## 6          5.4          3.9          1.7          0.4  setosa
```

Exploracion de la matriz

```
dim(datos)
```

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

```
colnames(datos)
```

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

```
str(datos)
```

```
## '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(datos)
```

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

Quitar los espacios de los nombres

```
colnames(datos)[1]="LargoSepalo"  
colnames(datos)[2]= "AnchoSepalo"  
colnames(datos)[3]="LargoPetalos"  
colnames(datos)[4]= "AnchoPetalos"  
head(datos)
```

```
##   LargoSepalo AnchoSepalo LargoPetalos AnchoPetalos Especie  
## 1         5.1         3.5         1.4         0.2   setosa  
## 2         4.9         3.0         1.4         0.2   setosa  
## 3         4.7         3.2         1.3         0.2   setosa  
## 4         4.6         3.1         1.5         0.2   setosa  
## 5         5.0         3.6         1.4         0.2   setosa  
## 6         5.4         3.9         1.7         0.4   setosa
```

Escalamiento de la matriz

Generacion de variables X

```
X <- datos %>%  
  select(LargoSepalo, AnchoSepalo) %>%  
  scale()  
head(X)
```

```
##   LargoSepalo AnchoSepalo  
## 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
```

Generacion de variables Y

```
Y <- datos %>%  
  select(LargoPetalos,AnchoPetalos) %>%  
  scale()  
head(Y)
```

```
##   LargoPetalos AnchoPetalos  
## 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
```

Analisis canonico con un par de variables

Libreria

```
library(CCA)
```

Analisis

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

Visualizacion de la matriz X

```
ac$xcoef
```

```
##           [,1]      [,2]
## LargoSepalo -0.07251736 0.03932826
## AnchoSepalo  0.03052965 0.07663824
```

Visualizacion de la matriz Y

```
ac$ycoef
```

```
##           [,1]      [,2]
## LargoPetaló -0.12279948 -0.2774814
## AnchoPetaló  0.04332444  0.3003309
```

Visualizacion de la correlacion canonica

```
ac$cor
```

```
## [1] 0.9409690 0.1239369
```

Los valores obtenidos en este análisis demuestran que el primer paquete (x1,y1) es de 94%

Obtencion 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]
```

```
#Visualizacion 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
```

Correlacion canonica entre variable X1 y Y1

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

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

Verificacion de la correlacion canonica

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

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

Analisis canonico con dos pares de variables

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

Agregamos las variables generadas a la matriz original de penguins

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

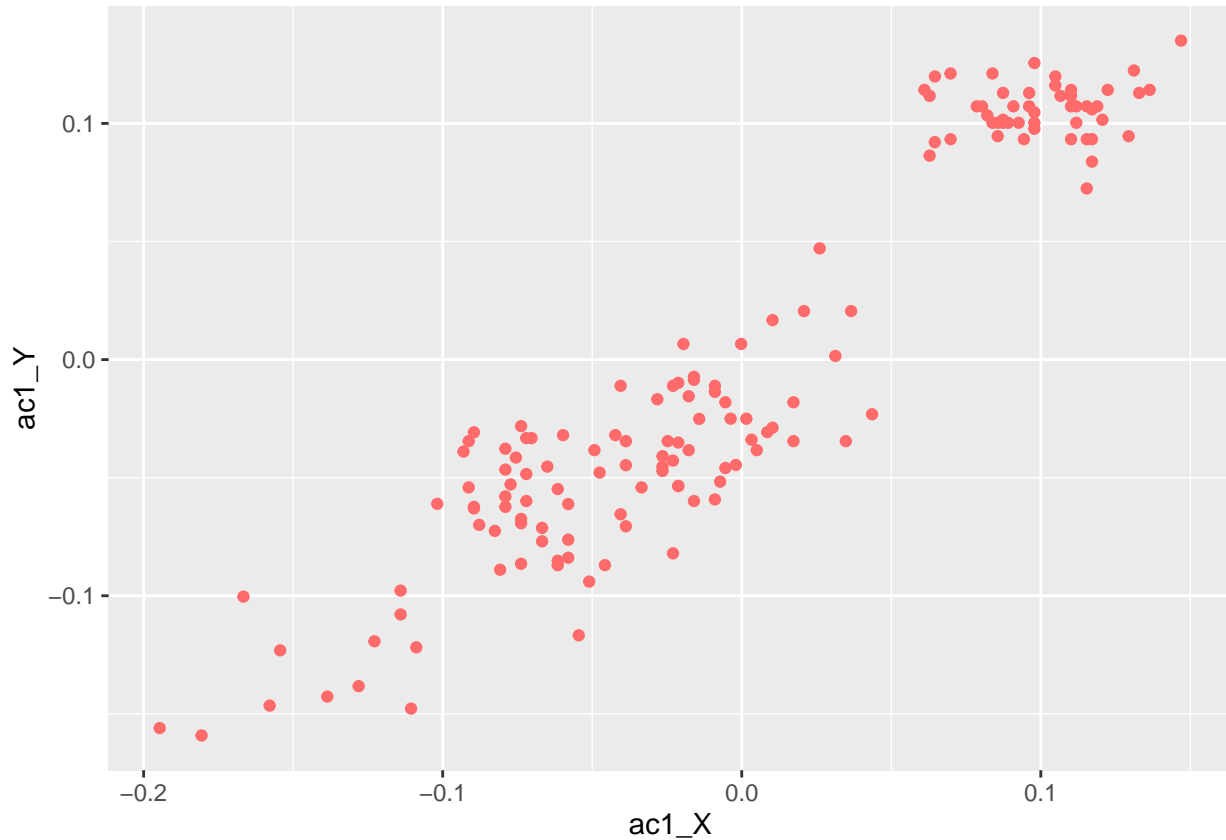
Visualizacion de los nombres de las variables

```
colnames(ac_df)
```

```
## [1] "LargoSepalo" "AnchoSepalo" "LargoPetal" "AnchoPetal" "Especie"
## [6] "ac1_X"       "ac1_Y"       "ac2_X"       "ac2_Y"
```

Generacion del grafico scater plot para la visualizacion de X1 y Y1

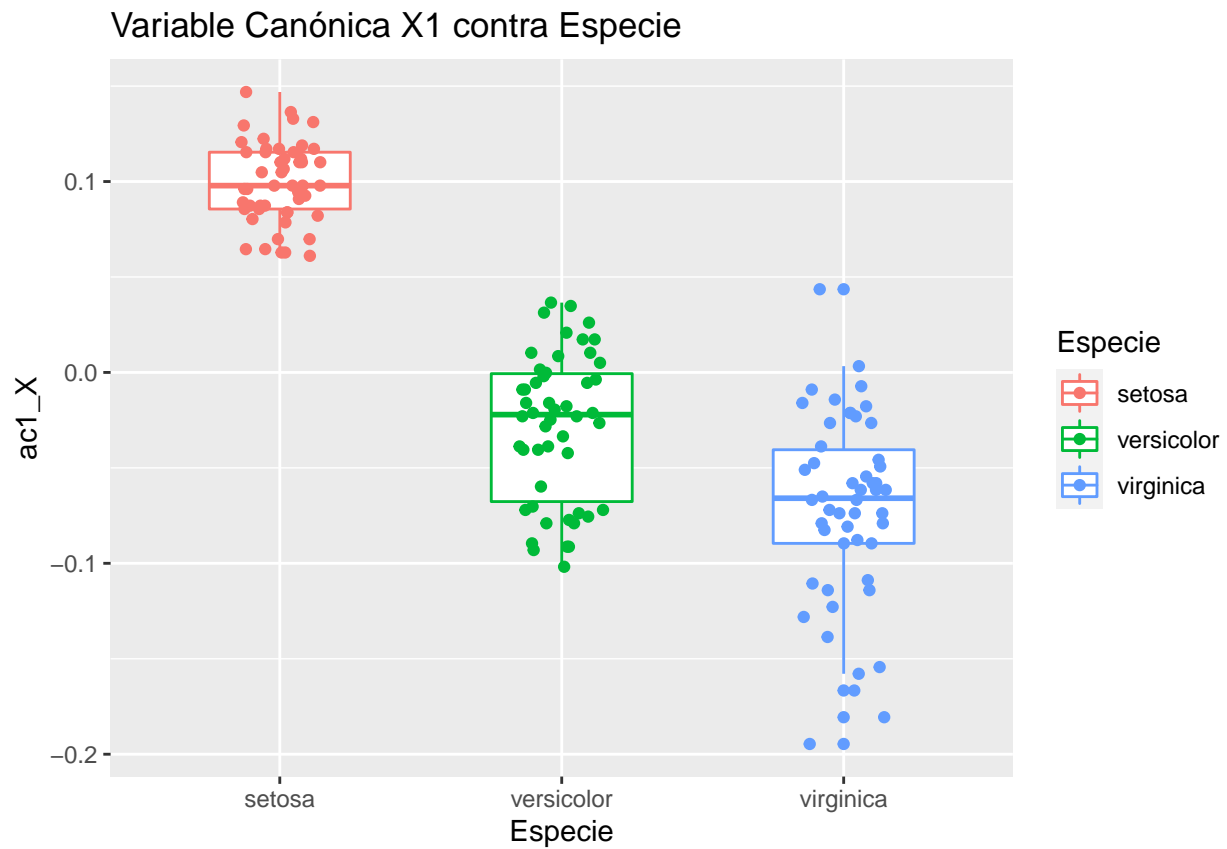
```
ac_df %>%  
  ggplot(aes(x=ac1_X,y=ac1_Y))+  
  geom_point(color="indianred1")
```



En este gráfico es posible observar que existe una correlación alta en la distribución de nuestros datos, en el eje x se encuentra nuestra variable canónica x1 y la variable obtenida de nuestro paquete 1

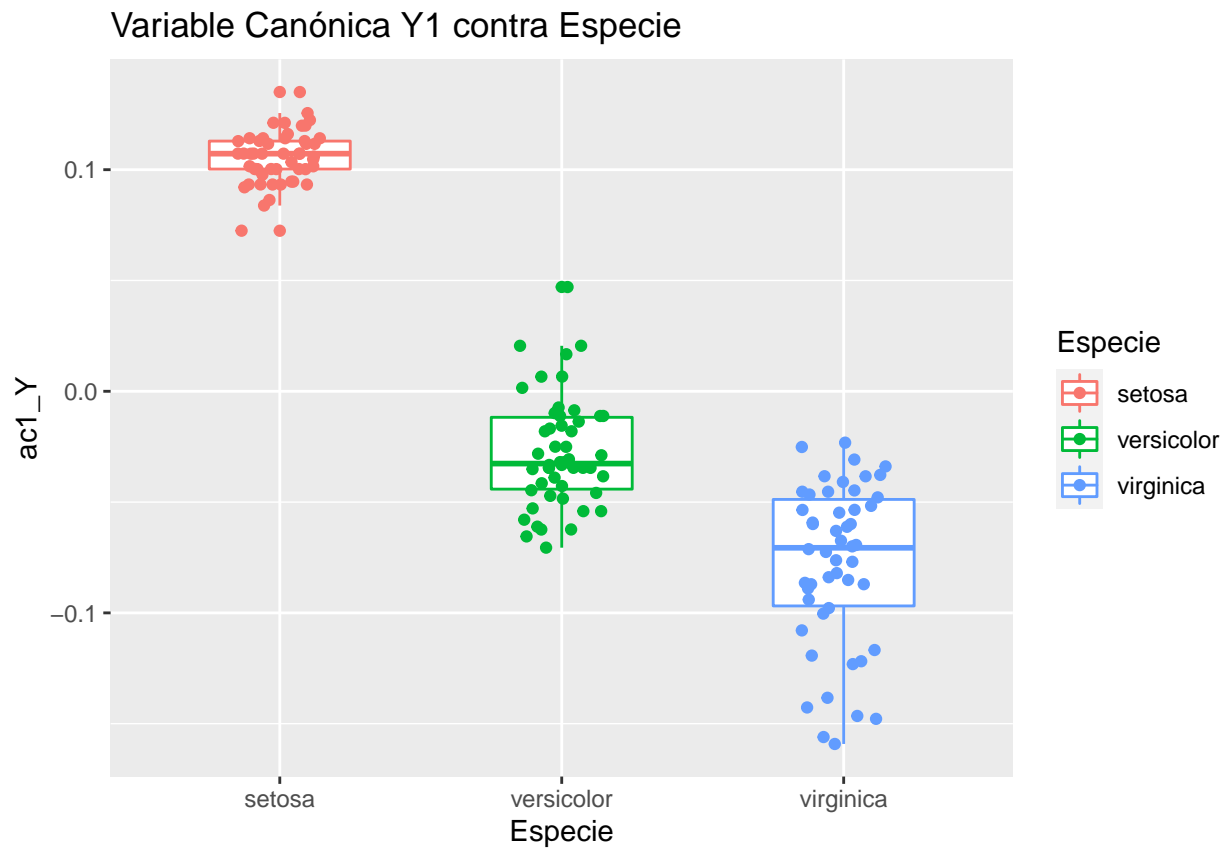
Generacion de un boxplot

```
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")
```



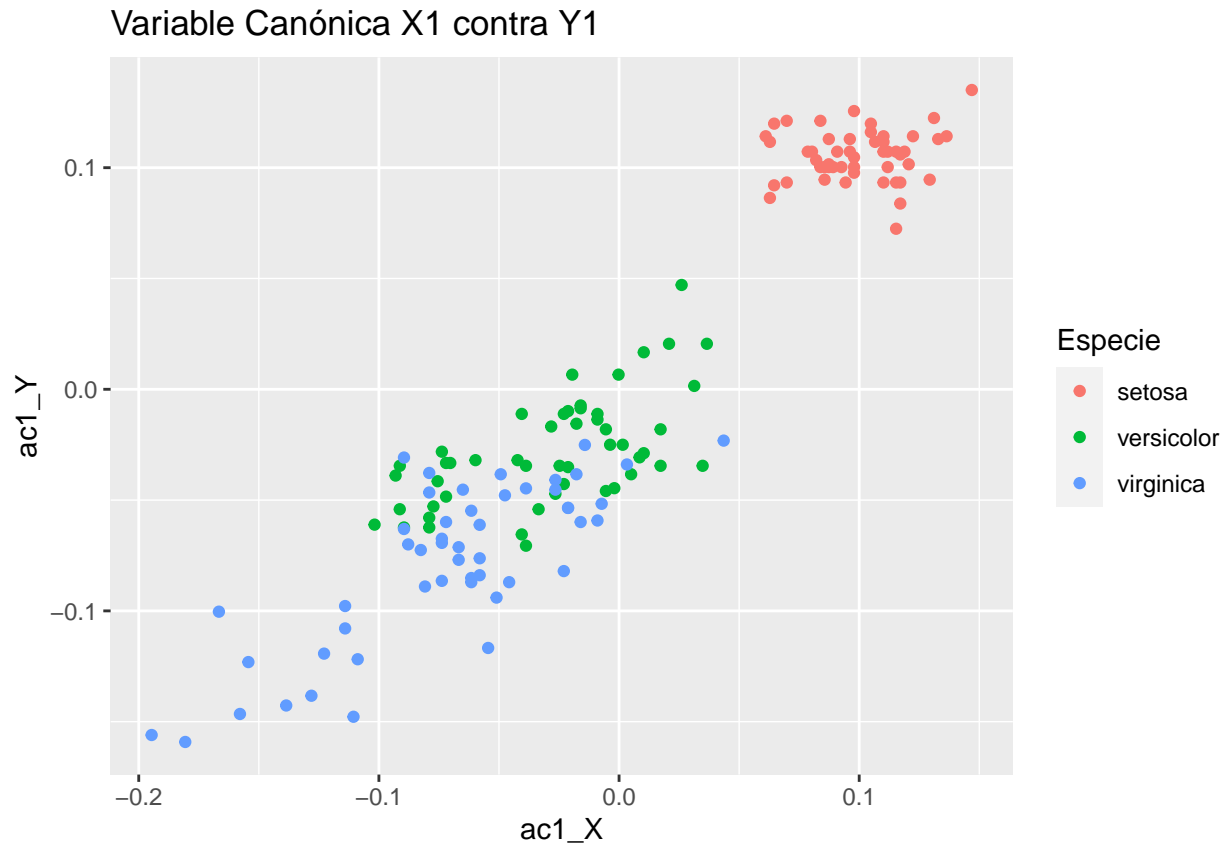
En este gráfico de igual manera es posible apreciar la correlación entre nuestra variable `acx1` y la variable `especie`.

```
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")
```



En este gráfico, igual que el anterior se presenta una buena correlación entre nuestra variable ac1_Y1

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



Por ultimo, dentro de este gráfico se encuentran representadas las correlaciones de la variable canonica x1, contra la variable y1, podemos observar que en ambas existe una buena correlación.

Generar la ecuación canonica

```
ac$xcoef
```

```
##           [,1]      [,2]
## LargoSepalo -0.07251736 0.03932826
## AnchoSepalo  0.03052965 0.07663824
```

```
ac$ycoef
```

```
##           [,1]      [,2]
## LargoPetalto -0.12279948 -0.2774814
## AnchoPetalto  0.04332444  0.3003309
```

```
head(ac_df)
```

```
##   LargoSepalo AnchoSepalo LargoPetalto AnchoPetalto Especie   ac1_X   ac1_Y
## 1         5.1         3.5          1.4          0.2   setosa 0.09610292 0.10722900
## 2         4.9         3.0          1.4          0.2   setosa 0.07859597 0.10722900
## 3         4.7         3.2          1.3          0.2   setosa 0.11011957 0.11418530
## 4         4.6         3.1          1.5          0.2   setosa 0.11187264 0.10027270
## 5         5.0         3.6          1.4          0.2   setosa 0.11186471 0.10722900
## 6         5.4         3.9          1.7          0.4   setosa 0.09784806 0.09772779
##           ac2_X      ac2_Y
## 1  0.04252999 -0.023103195
```



```
## 2 -0.05488368 -0.023103195
## 3 -0.02921656 -0.007384524
## 4 -0.05154894 -0.038821867
## 5  0.05536355 -0.023103195
## 6  0.12711010  0.008543219
```

- $u1 = -0.0725 \text{ GrosorSepalo} + 0.0393 \text{ (LargoSepalo)}$
- $v2 = -0.1227 \text{ largoPetal} + 0.3003 \text{ (AnchoPetal)}$