

Analisis Canonico

SERGIO ADRIAN ORTIZ ORTEGA

2022-05-14

```
#Analisis Canonico
##Instalar paqueterias
rm(list=ls())
library(tidyverse)
library(readxl)
```

cargamos los datos

```
tension<- read_excel("tenso.xlsx")
```

Exploracion de matriz

```
dim(tension)
```

```
## [1] 50 13
```

```
str(tension)
```

```
## tibble [50 x 13] (S3: tbl_df/tbl/data.frame)
## $ edad : num [1:50] 75 55 65 50 65 90 75 60 65 80 ...
## $ anemia : num [1:50] 0 0 0 1 1 1 1 0 1 ...
## $ diabetes : num [1:50] 0 0 0 0 1 0 0 1 0 0 ...
## $ Alta_presión_sanguínea : num [1:50] 1 0 0 0 0 1 0 0 0 1 ...
## $ sexo : num [1:50] 1 1 1 1 0 1 1 1 0 1 ...
## $ fuma : num [1:50] 0 0 1 0 0 1 0 1 0 1 ...
## $ MUERTE_EVENTO : num [1:50] 1 1 1 1 1 1 1 1 1 1 ...
## $ creatinina_fosfoquinasa: num [1:50] 582 7861 146 111 160 ...
## $ fracción_de_eyección : num [1:50] 20 38 20 20 20 40 15 60 65 35 ...
## $ plaquetas : num [1:50] 265000 263358 162000 210000 327000 ...
## $ suero_creatinina : num [1:50] 1.9 1.1 1.3 1.9 2.7 2.1 1.2 1.1 1.5 9.4 ...
## $ suero_sodio : num [1:50] 130 136 129 137 116 132 137 131 138 133 ...
## $ tiempo : num [1:50] 4 6 7 7 8 8 10 10 10 10 ...
```

```
colnames(tension)
```

```
## [1] "edad" "anemia"
## [3] "diabetes" "Alta_presión_sanguínea"
## [5] "sexo" "fuma"
## [7] "MUERTE_EVENTO" "creatinina_fosfoquinasa"
## [9] "fracción_de_eyección" "plaquetas"
## [11] "suero_creatinina" "suero_sodio"
## [13] "tiempo"
```

##transformar las variables para que funcionen con el codigo

```
fuma<-factor(tension$fuma,
             levels= c("1","0"))
anemia<-factor(tension$anemia,
              levels= c("1","0"))
sexo<-factor(tension$sexo,
            levels= c("1","0"))
diabetes<-factor(tension$diabetes,
               levels= c("1","0"))
Alta_presión_sanguínea<-factor(tension$Alta_presión_sanguínea,
                              levels= c("1","0"))
MUERTE_EVENTO<-factor(tension$MUERTE_EVENTO,
                     levels= c("1","0"))
edad<-as.numeric(tension$edad,strict = TRUE)
creatinina_fosfoquinasa<-as.numeric(tension$creatinina_fosfoquinasa,strict = TRUE)
fracción_de_eyección<-as.numeric(tension$fracción_de_eyección,strict = TRUE)
plaquetas<-as.numeric(tension$plaquetas,strict = FALSE)
suero_creatinina<-as.numeric(tension$suero_creatinina,strict = TRUE)
suero_sodio<-as.numeric(tension$suero_sodio,strict = TRUE)
tiempo<-as.numeric(tension$tiempo,strict = TRUE)
```

ya con los datos convertidos como requerimos los tranformaresmos de nuevo a una matrix

```
tension<-data.frame(edad, anemia, diabetes,Alta_presión_sanguínea,sexo,fuma,
                   MUERTE_EVENTO,creatinina_fosfoquinasa,fracción_de_eyección,
                   plaquetas,suero_creatinina,suero_sodio,tiempo)
```

Escalamiento de la matriz

##Generacion de variables X

```
X <-tension %>%
  select(suero_creatinina, suero_sodio) %>%
  scale()
head(X)
```

```
##      suero_creatinina suero_sodio
## [1,]      0.06949762  -0.9799705
## [2,]     -0.46407335   0.1639639
## [3,]     -0.33068061  -1.1706262
## [4,]      0.06949762   0.3546197
## [5,]      0.60306859  -3.6491508
## [6,]      0.20289036  -0.5986590
```

Generacion de variables Y

```
Y <- tension %>%
  select(creatinina_fosfoquinasa,fracción_de_eyección) %>%
  scale()
head(Y)
```

```
##      creatinina_fosfoquinasa fracción_de_eyección
## [1,]          0.06493473         -1.2999320
## [2,]          6.41106639          0.1552158
```

```
## [3,]          -0.31518801          -1.2999320
## [4,]          -0.34570245          -1.2999320
## [5,]          -0.30298223          -1.2999320
## [6,]          -0.40150028           0.3168988
```

Analisis canonico con un par de variables

```
##Libreria
```

```
library(CCA)
```

```
## Loading required package: fda
## Loading required package: splines
## Loading required package: fds
## Loading required package: rainbow
## Loading required package: MASS
##
## Attaching package: 'MASS'
## The following object is masked from 'package:dplyr':
##
##      select
## Loading required package: pcaPP
## Loading required package: RCurl
##
## Attaching package: 'RCurl'
## The following object is masked from 'package:tidyr':
##
##      complete
## Loading required package: deSolve
##
## Attaching package: 'fda'
## The following object is masked from 'package:graphics':
##
##      matplot
## Loading required package: fields
## Loading required package: spam
## Spam version 2.8-0 (2022-01-05) is loaded.
## Type 'help( Spam)' or 'demo( spam)' for a short introduction
## and overview of this package.
## Help for individual functions is also obtained by adding the
## suffix '.spam' to the function name, e.g. 'help( chol.spam)'.
##
## Attaching package: 'spam'
## The following objects are masked from 'package:base':
##
##      backsolve, forwardsolve
```

```
## Loading required package: viridis
## Loading required package: viridisLite
##
## Try help(fields) to get started.
```

Analisis

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

Visualizacion de la matriz X y matriz Y

```
ac$xcoef
```

```
##                [,1]      [,2]
## suero_creatinina 0.04270985 -0.13965934
## suero_sodio      -0.12773810 -0.07079434
```

```
ac$ycoef
```

```
##                [,1]      [,2]
## creatinina_fosfoquinasa -0.03071297  0.1395576
## fracción_de_eyección   -0.14024559 -0.0274004
```

Visualizacion de la correlacion canonica

```
ac$cor
```

```
## [1] 0.20691613 0.07046846
```

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] 0.12814780 -0.04076494 0.13541025 -0.04233021 0.49189256 0.08513698
## [7] -0.06227035 0.08100507 -0.07807858 0.26873017 0.16361423 -0.14387813
## [13] -0.06511894 -0.06511894 -0.09232154 -0.03506776 -0.14387813 0.16987530
## [19] -0.14102954 0.34733382
```

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

```
## [1] 0.18031540 -0.21867120 0.19199009 0.19292728 0.19161522 -0.03211240
## [7] 0.24600076 -0.26604199 -0.31849960 0.02254091 -0.01034747 0.13302571
## [13] 0.05625474 -0.01267706 0.08038066 -0.15437901 -0.01216830 0.24834142
## [19] 0.13586406 -0.21650305
```

Correlacion canonica entre variable X1 y Y1

```
cor(ac1_X,ac1_Y)

##           [,1]
## [1,] 0.2069161

## Verificación de la correlación canónica
assertthat::are_equal(ac$cor[1],
                      cor(ac1_X,ac1_Y)[1])

## [1] TRUE
```

Analisis 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]

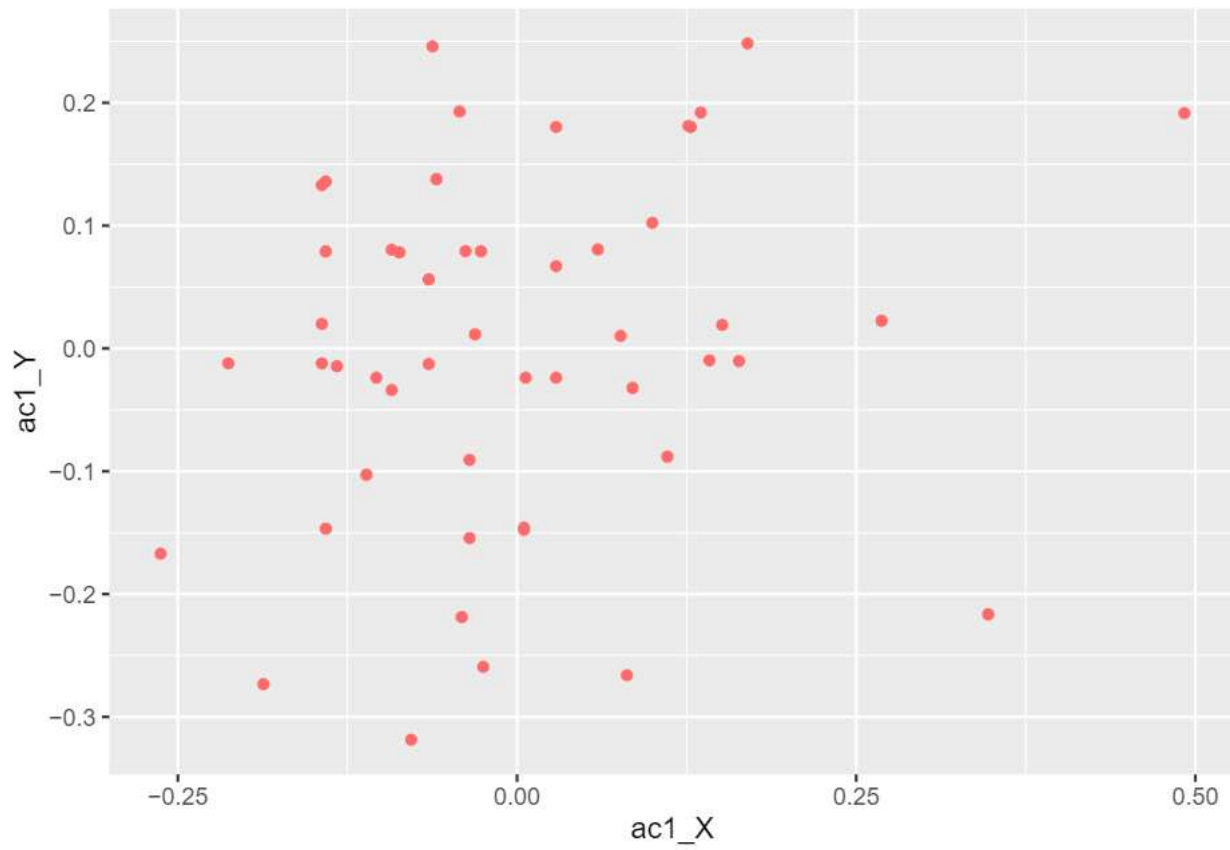
## Agregamos las variables generadas a la matriz original de tension
ac_df <- tension %>%
  mutate(ac1_X=ac1_X,
         ac1_Y=ac1_Y,
         ac2_X=ac2_X,
         ac2_Y=ac2_Y)

## Visualización de los nombres de las variables
colnames(ac_df)

## [1] "edad" "anemia"
## [3] "diabetes" "Alta_presión_sanguínea"
## [5] "sexo" "fuma"
## [7] "MUERTE_EVENTO" "creatinina_fosfoquinasa"
## [9] "fracción_de_eyeción" "plaquetas"
## [11] "suero_creatinina" "suero_sodio"
## [13] "tiempo" "ac1_X"
## [15] "ac1_Y" "ac2_X"
## [17] "ac2_Y"
```

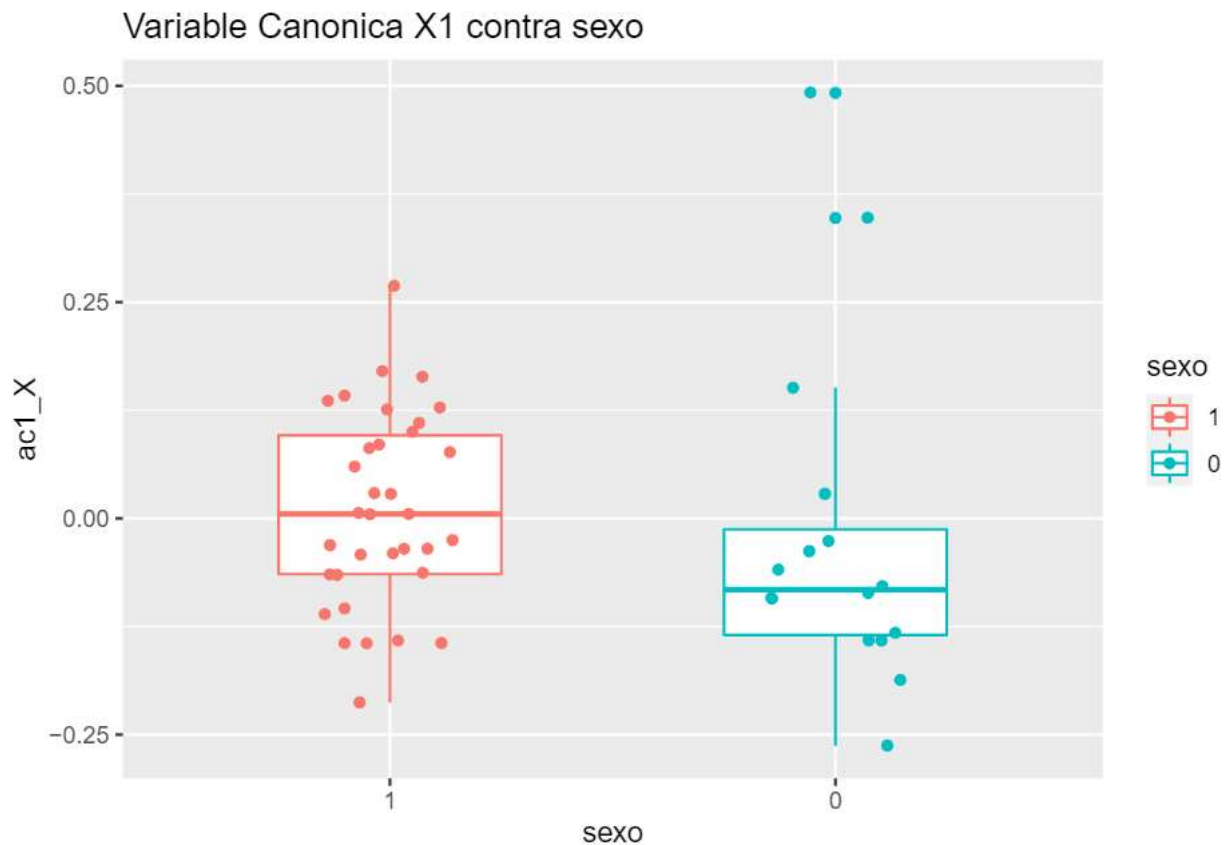
Generacion del grafico scatter plot para la visualizacion de X1 y Y1

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



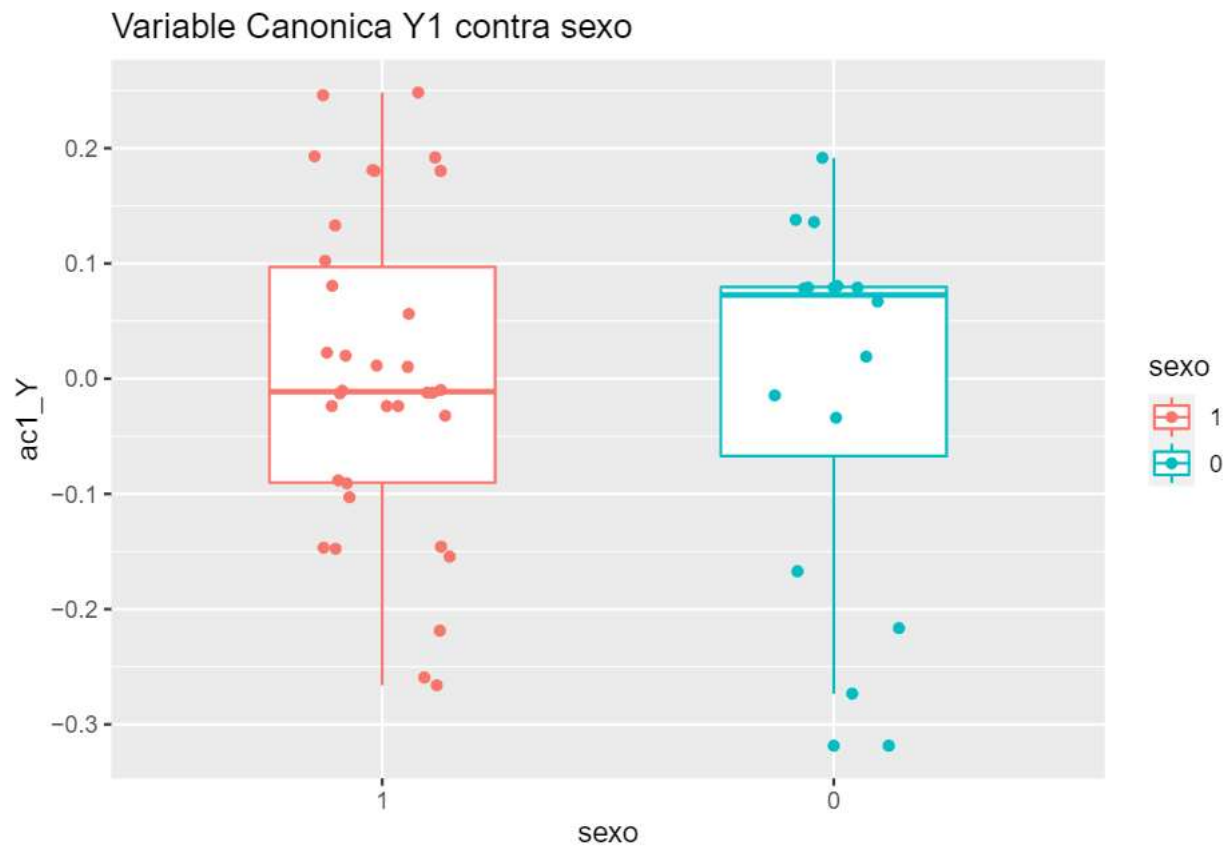
Generacion de un boxplot

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

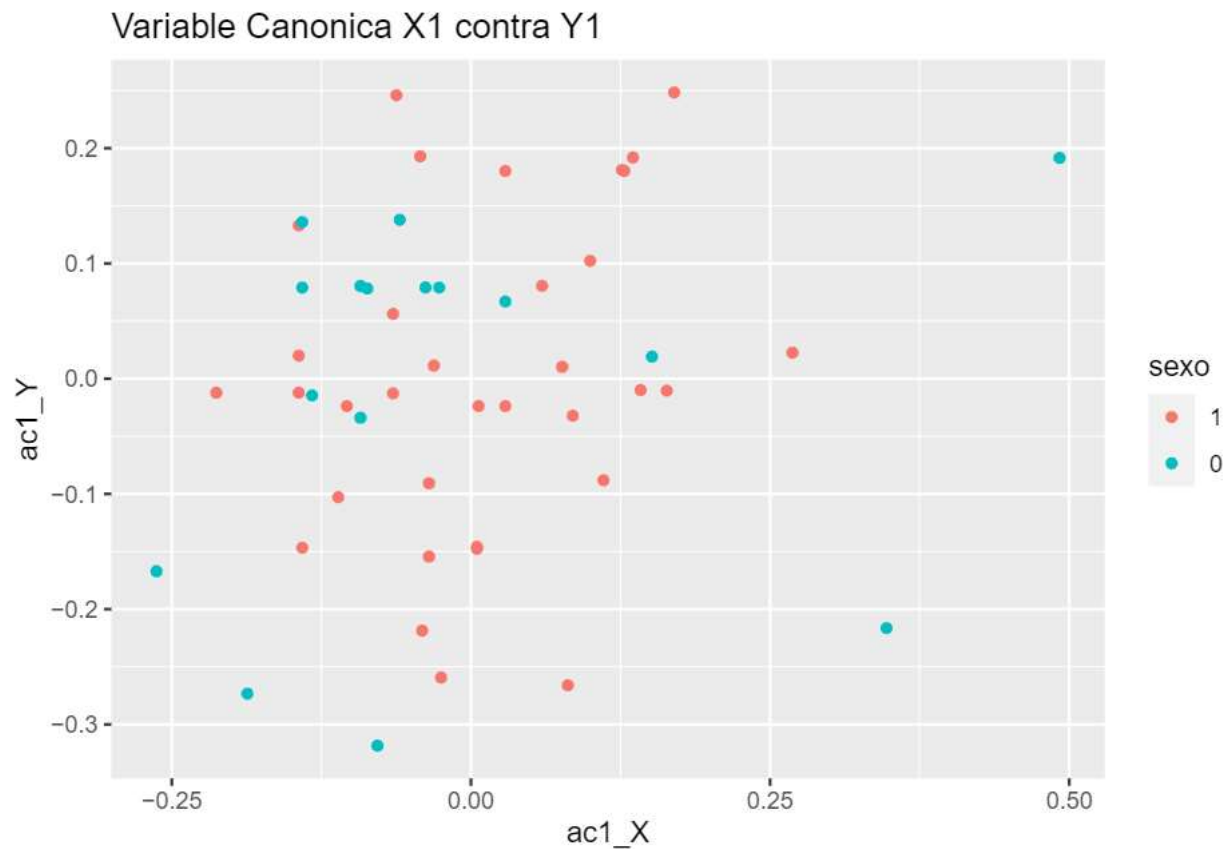



Interpretacion: se observa una correlacion entre la variable canonica X1 y la variable latente Especie

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



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

Scarter plot con las variables canonicas X2 y Y2 separadas por fuma.

```
ac_df %>%
  ggplot(aes(x=ac2_X,y=ac2_Y, color=fuma))+
  geom_point()+
  ggtitle("Variable Canonica X2 contra Y2")
```

