

# Extra - Análisis Canónico

Itzel Teodocio Olivares

2022-06-02

## Analisis Canonico

### Instalar paqueterias

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

### Preparacion de la matriz

```
millas<-data.frame(datos::millas)
```

### Exploracion de la matriz

```
dim(millas)
```

```
## [1] 234 11
```

```
colnames(millas)
```

```
## [1] "fabricante" "modelo" "cilindrada" "anio" "cilindros"  
## [6] "transmision" "traccion" "ciudad" "autopista" "combustible"  
## [11] "clase"
```

```
str(millas)
```

```
## 'data.frame': 234 obs. of 11 variables:  
## $ fabricante : chr "audi" "audi" "audi" "audi" ...  
## $ modelo : chr "a4" "a4" "a4" "a4" ...  
## $ cilindrada : num 1.8 1.8 2 2 2.8 2.8 3.1 1.8 1.8 2 ...  
## $ anio : int 1999 1999 2008 2008 1999 1999 2008 1999 1999 2008 ...  
## $ cilindros : int 4 4 4 4 6 6 6 4 4 4 ...  
## $ transmision: chr "auto(l5)" "manual(m5)" "manual(m6)" "auto(av)" ...  
## $ traccion : chr "d" "d" "d" "d" ...  
## $ ciudad : int 18 21 20 21 16 18 18 18 16 20 ...  
## $ autopista : int 29 29 31 30 26 26 27 26 25 28 ...  
## $ combustible: chr "p" "p" "p" "p" ...  
## $ clase : chr "compacto" "compacto" "compacto" "compacto" ...
```

```
anyNA(millas)
```

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

## Escalamiento de la matriz

### Generacion de variables X

```
X <- millas %>%
  select(cilindrada, anio) %>%
  scale()
head(X)
```

```
##      cilindrada      anio
## [1,] -1.2939999 -0.997861
## [2,] -1.2939999 -0.997861
## [3,] -1.1391962  0.997861
## [4,] -1.1391962  0.997861
## [5,] -0.5199816 -0.997861
## [6,] -0.5199816 -0.997861
```

### Generacion de variables Y

```
Y <- millas %>%
  select(cilindros, ciudad) %>%
  scale()
head(Y)
```

```
##      cilindros      ciudad
## [1,] -1.1721058  0.2681016
## [2,] -1.1721058  0.9729978
## [3,] -1.1721058  0.7380324
## [4,] -1.1721058  0.9729978
## [5,]  0.0689474 -0.2018293
## [6,]  0.0689474  0.2681016
```

## Analisis canonico con un par de variables

### Libreria

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

### Analisis

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

### Visualizacion de la matriz X

```
ac$xcoef
```

```
##              [,1]      [,2]
## cilindrada -0.065760581 -0.007955922
```

```
## anio          0.001853736  0.066214156
```

## Visualizacion de la matriz Y

```
ac$ycoef
```

```
##           [,1]      [,2]
## cilindros -0.05703566 0.0947826
## ciudad    0.01017508 0.1101511
```

## Visualizacion de la correlacion canonica

```
ac$cor
```

```
## [1] 0.9342558 0.1150952
```

## 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] 0.08324442 0.08324442 0.07676398 0.07676398 0.03234452 0.03234452
## [7] 0.02077409 0.08324442 0.08324442 0.07676398 0.07676398 0.03234452
## [13] 0.03234452 0.02077409 0.02077409 0.03234452 0.02077409 -0.03521579
## [19] -0.09120568 -0.09120568
```

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

```
## [1] 0.069579782 0.076752158 0.074361366 0.076752158 -0.005986089
## [6] -0.001204505 -0.001204505 0.069579782 0.064798198 0.074361366
## [11] 0.071970574 -0.008376882 -0.003595297 -0.003595297 -0.008376882
## [16] -0.008376882 -0.003595297 -0.076770377 -0.081551961 -0.088724337
```

## Correlacion canonica entre variable X1 y Y1

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

```
##           [,1]
## [1,] 0.9342558
```

## 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

### 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 millas

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

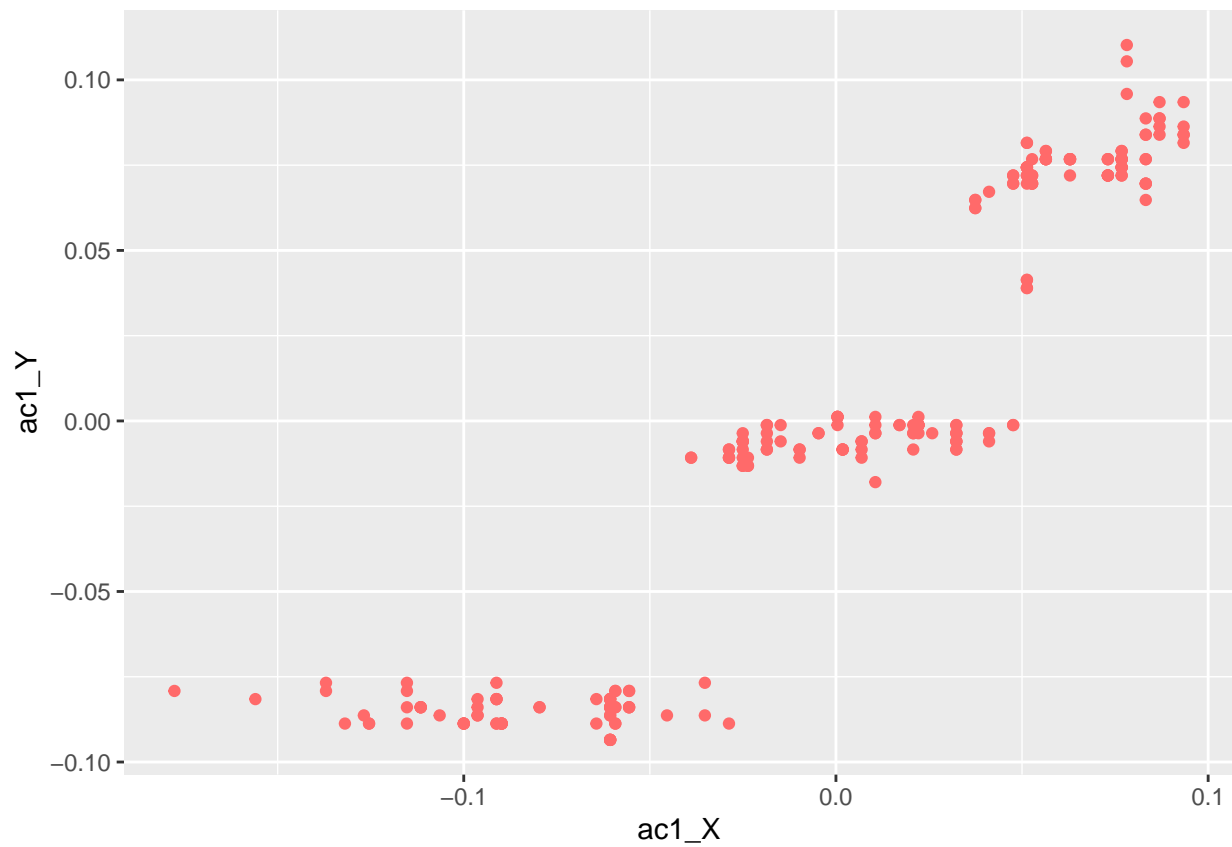
## Visualizacion de los nombres de las variables

```
colnames(ac_df)
```

```
## [1] "fabricante" "modelo" "cilindrada" "anio" "cilindros"  
## [6] "transmision" "traccion" "ciudad" "autopista" "combustible"  
## [11] "clase" "ac1_X" "ac1_Y" "ac2_X" "ac2_Y"
```

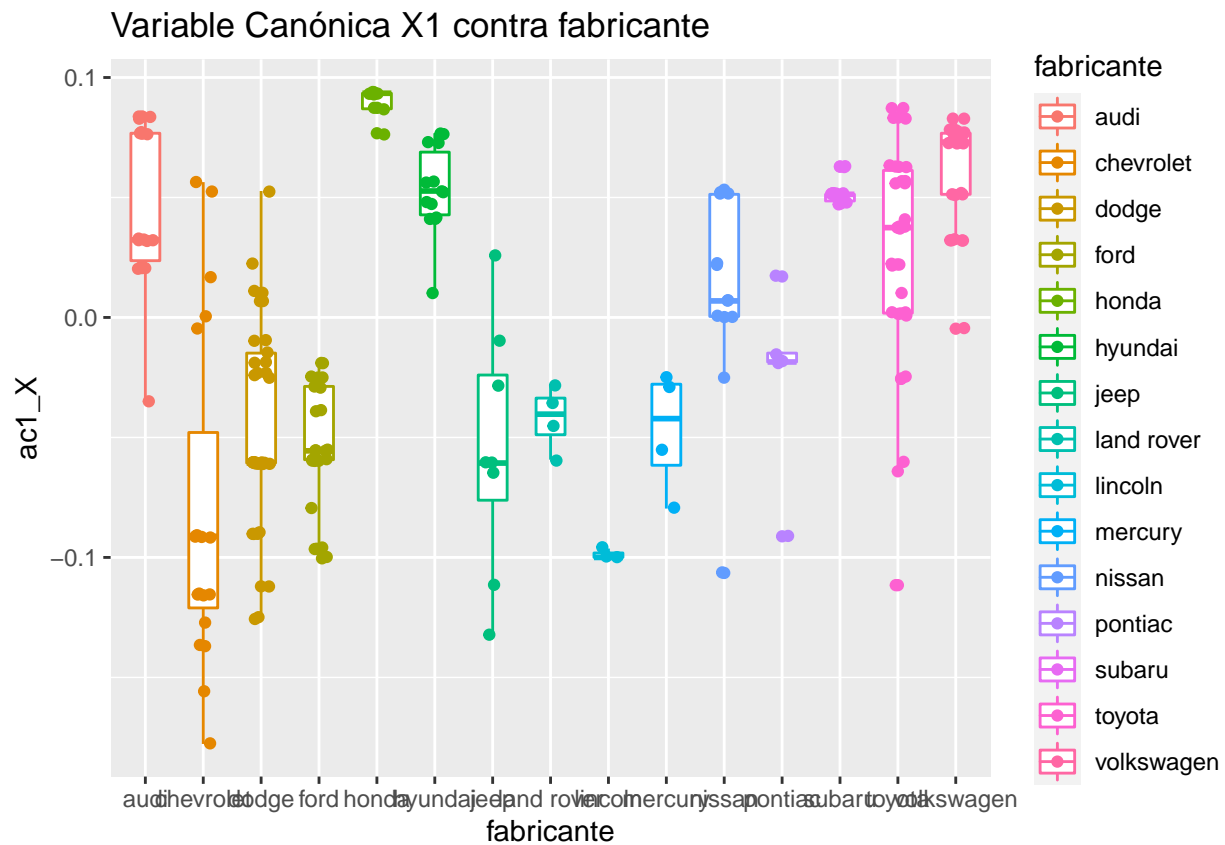
## Generacion del grafico scater plot para la visualizacion de X1 y Y1

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



## Generacion de un boxplot

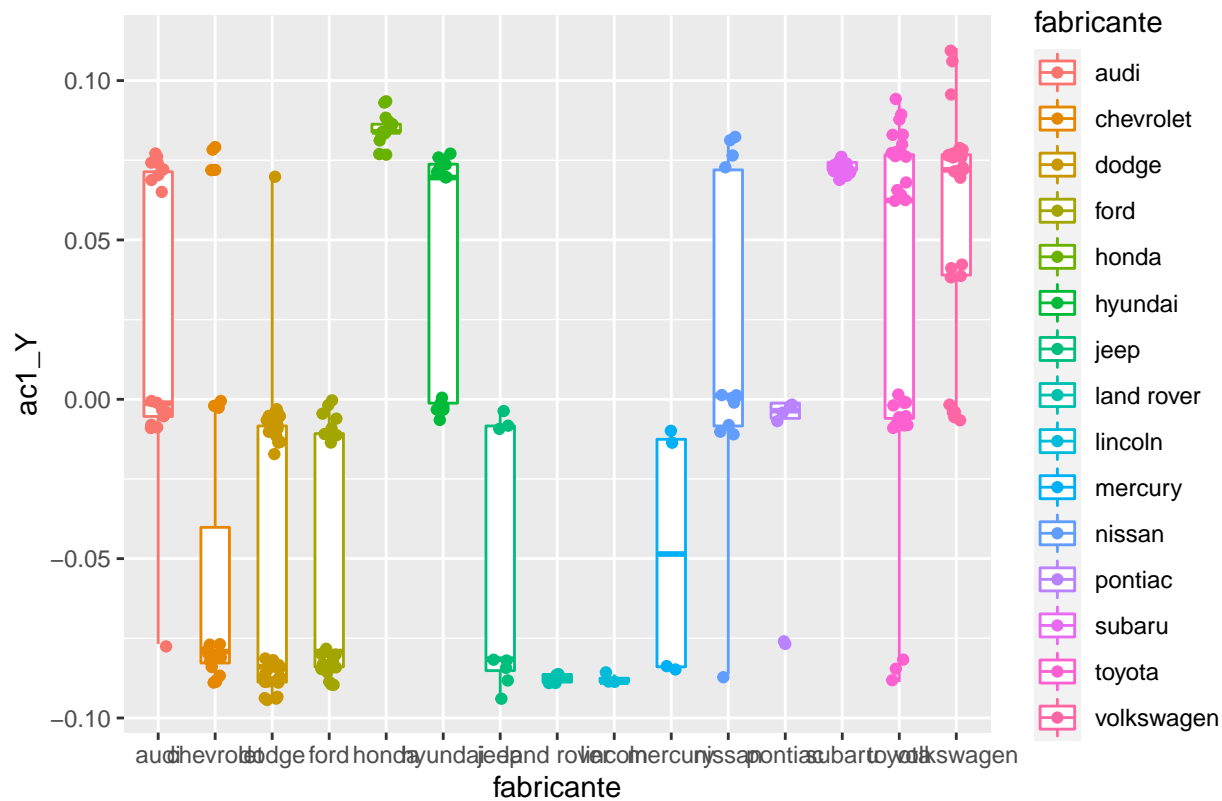
```
ac_df %>%
  ggplot(aes(x=fabricante,y=ac1_X, color=fabricante))+
  geom_boxplot(width=0.5)+
  geom_jitter(width=0.15)+
  ggtitle("Variable Canónica X1 contra fabricante")
```



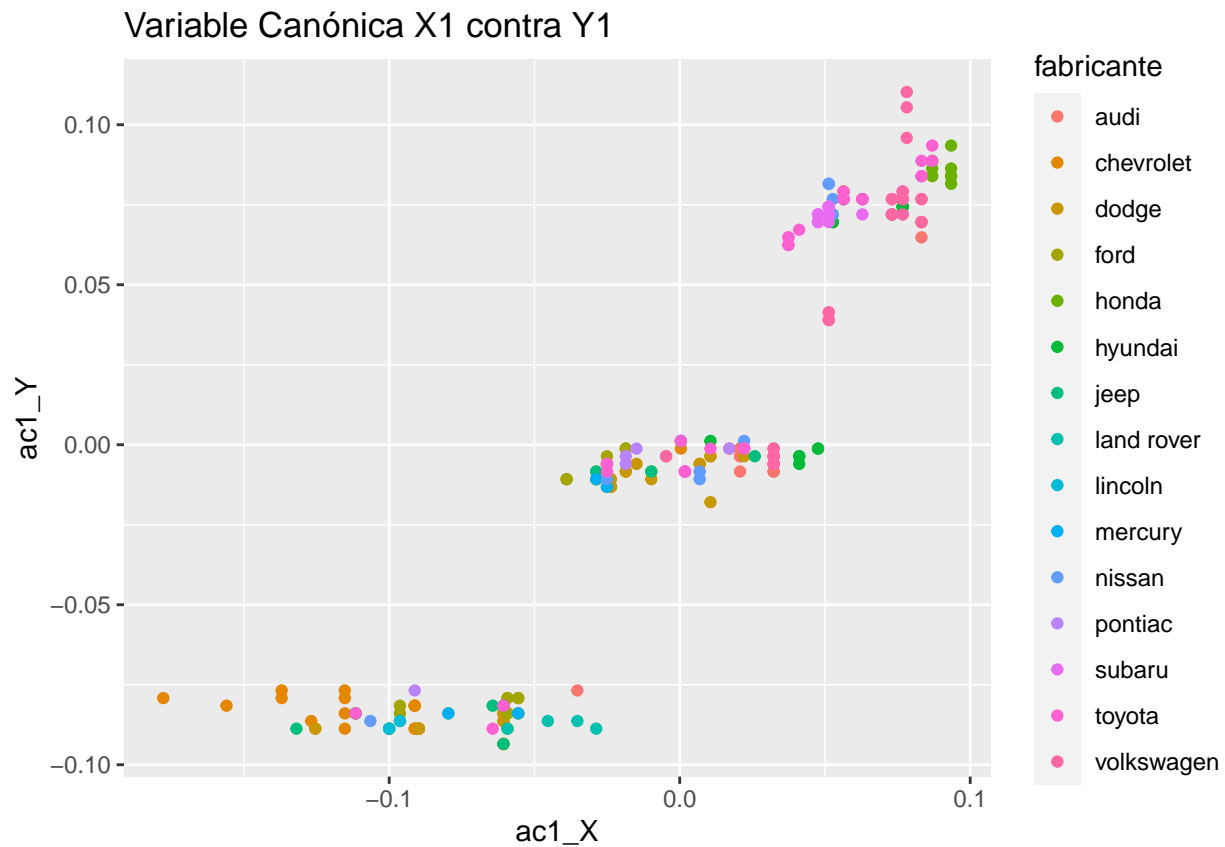
**Interpretación:** se observa una correlación entre la variable canónica X1 y la variable latente Fabricante

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

Variable Canónica Y1 contra Fabricante



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



Scarter plot con las variables canonicas

X2 y Y2 separadas por modelo.

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



