

Análisis Canónico - Iris

Karina Itzel Rodríguez Conde

3/6/2022

ANÁLISIS CANÓNICO

Introducción

El Análisis Canónico es una técnica multivariante que cuantifica la validez de la relación entre dos conjuntos de variables, sea dependiente o independiente y tiene como objetivo principal el determinar si estos dos conjuntos de variables son independientes uno de otro.

Matriz de datos

Para esta práctica, se trabajó con la matriz **iris**, la cual fue extraída del paquete *MASS* que se encuentra precargada en R.

Importar la matriz de datos

```
library(MASS)
iris<-as.data.frame(iris)
```

Exploración de la matriz

1.- Dimensión de la matriz:

```
dim(iris)
```

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

La base de datos cuenta con 150 observaciones y 5 variables.

2.- Nombres de las columnas:

```
colnames(iris)
```

```
## [1] "Sepal.Length" "Sepal.Width" "Petal.Length" "Petal.Width" "Species"
```

3.- Tipo de variables:

```
str(iris)
```

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

4.- Saber si existen datos nulos:

```
anyNA(iris)
```

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

Esta base de datos no contiene datos nulos.

Escalamiento de la matriz

Paqueterías a utilizar

```
library(tidyverse)
```

1.- Generación de variables X

```
X <- iris %>%
  select(Sepal.Length, Sepal.Width) %>%
  scale()
head(X)
```

```
##      Sepal.Length Sepal.Width
## [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
```

2.- Generación de variables Y

```
Y <- iris %>%
  select(Petal.Length,Petal.Width) %>%
  scale()
head(Y)
```

```
##      Petal.Length Petal.Width
## [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 canónico con un par de variables

Librería y paquetería a utilizar

```
library(CCA)
```

1.- Análisis

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

```
## $cor
## [1] 0.9409690 0.1239369
##
## $xcoef
##           [,1]      [,2]
## Sepal.Length -0.07251736 0.03932826
## Sepal.Width   0.03052965 0.07663824
##
## $ycoef
##           [,1]      [,2]
## Petal.Length -0.12279948 -0.2774814
## Petal.Width   0.04332444 0.3003309
##
## $xcenter
## Sepal.Length Sepal.Width
## -4.480675e-16 2.035409e-16
##
## $ycenter
## Petal.Length Petal.Width
## -2.844947e-17 -3.714621e-17
```

2.- Visualización de la matriz X

```
ac$xcoef
```

```
##           [,1]      [,2]
## Sepal.Length -0.07251736 0.03932826
## Sepal.Width   0.03052965 0.07663824
```

3.- Visualización de la matriz Y

```
ac$ycoef
```

```
##           [,1]      [,2]
## Petal.Length -0.12279948 -0.2774814
## Petal.Width   0.04332444 0.3003309
```

4.- Visualización de la correlación canónica

```
ac$cor
```

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

5.- Obtención de la matriz de variables canónicas

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]
```

6.- Visualización de los primeros 20 datos

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

```
## [1] 0.09610292 0.07859597 0.11011957 0.11187264 0.11186471 0.09784806  
## [7] 0.13288573 0.09785599 0.11537879 0.08560034 0.08383933 0.11537086  
## [13] 0.08735341 0.13114059 0.06982268 0.10659757 0.09784806 0.09610292  
## [19] 0.06457139 0.11711600
```

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

```
## [1] 0.10722900 0.10722900 0.11418530 0.10027270 0.10722900 0.09772779  
## [7] 0.11291285 0.10027270 0.10722900 0.09458885 0.10027270 0.09331640  
## [13] 0.10154515 0.12241406 0.12114160 0.11164040 0.12555300 0.11291285  
## [19] 0.09204394 0.10595655
```

7.- Correlación canónica entre variable X1 y Y1

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

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

8.- Verificación de la correlación canónica

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

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

Análisis canónico con dos pares de variables

1.- Cálculo de las variables X2 y Y2

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

2.- Agregamos las variables generadas a la matriz original de penguins

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

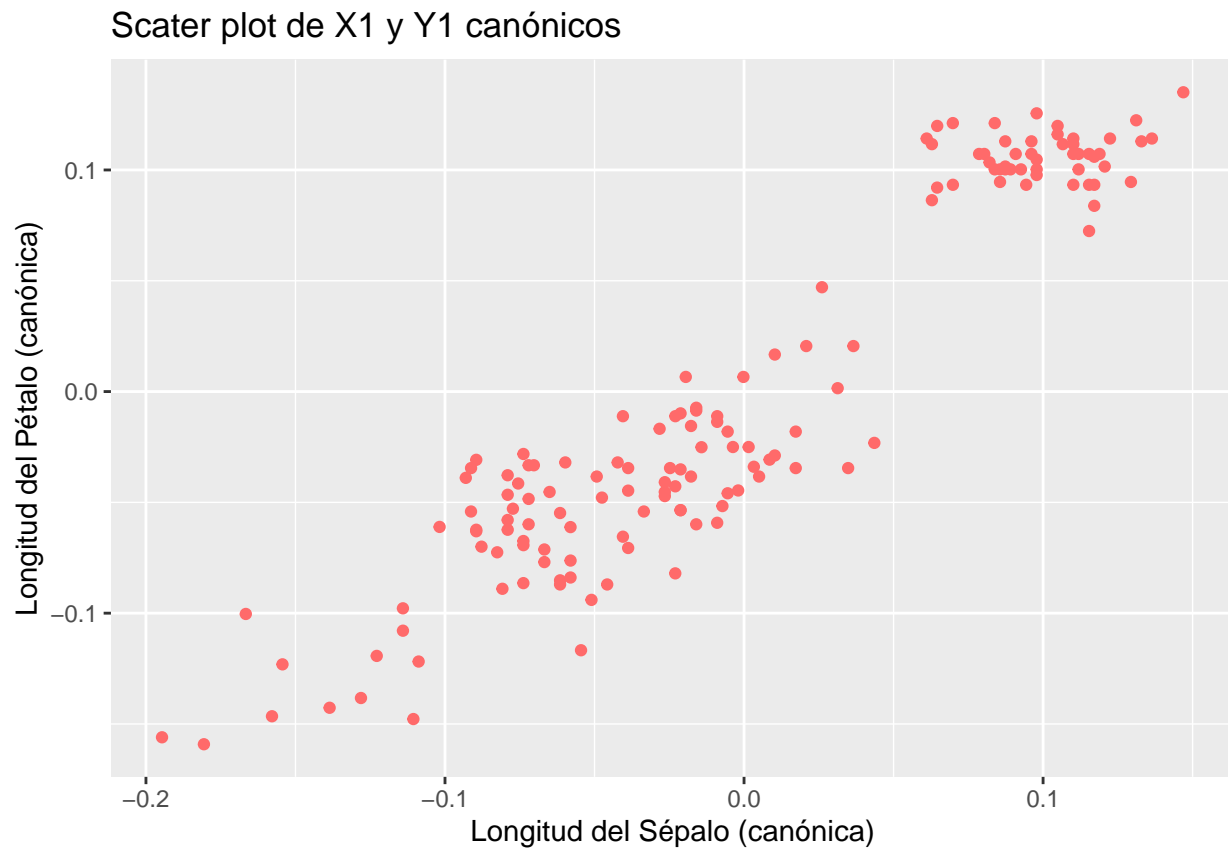
3.- Visualización de los nombres de las variables

```
colnames(ac_df)
```

```
## [1] "Sepal.Length" "Sepal.Width" "Petal.Length" "Petal.Width" "Species"
## [6] "ac1_X"        "ac1_Y"        "ac2_X"        "ac2_Y"
```

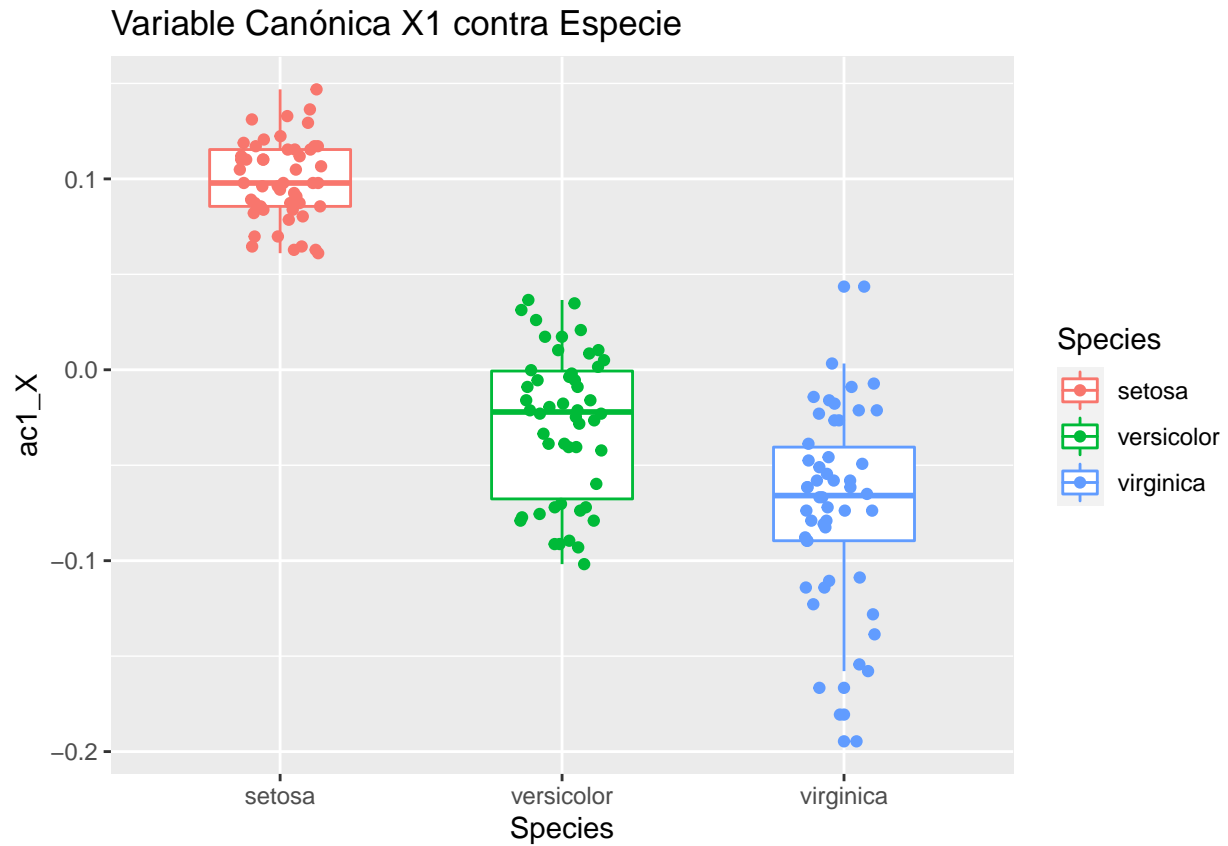
4.- Generación del grafico scatter plot para la visualización de X1 y Y1

```
ac_df %>%
  ggplot(aes(x=ac1_X,y=ac1_Y))+
  geom_point(color="indianred1") +
  xlab("Longitud del Sépalo (canónica)") +
  ylab("Longitud del Pétalo (canónica)") +
  ggtitle("Scater plot de X1 y Y1 canónicos")
```



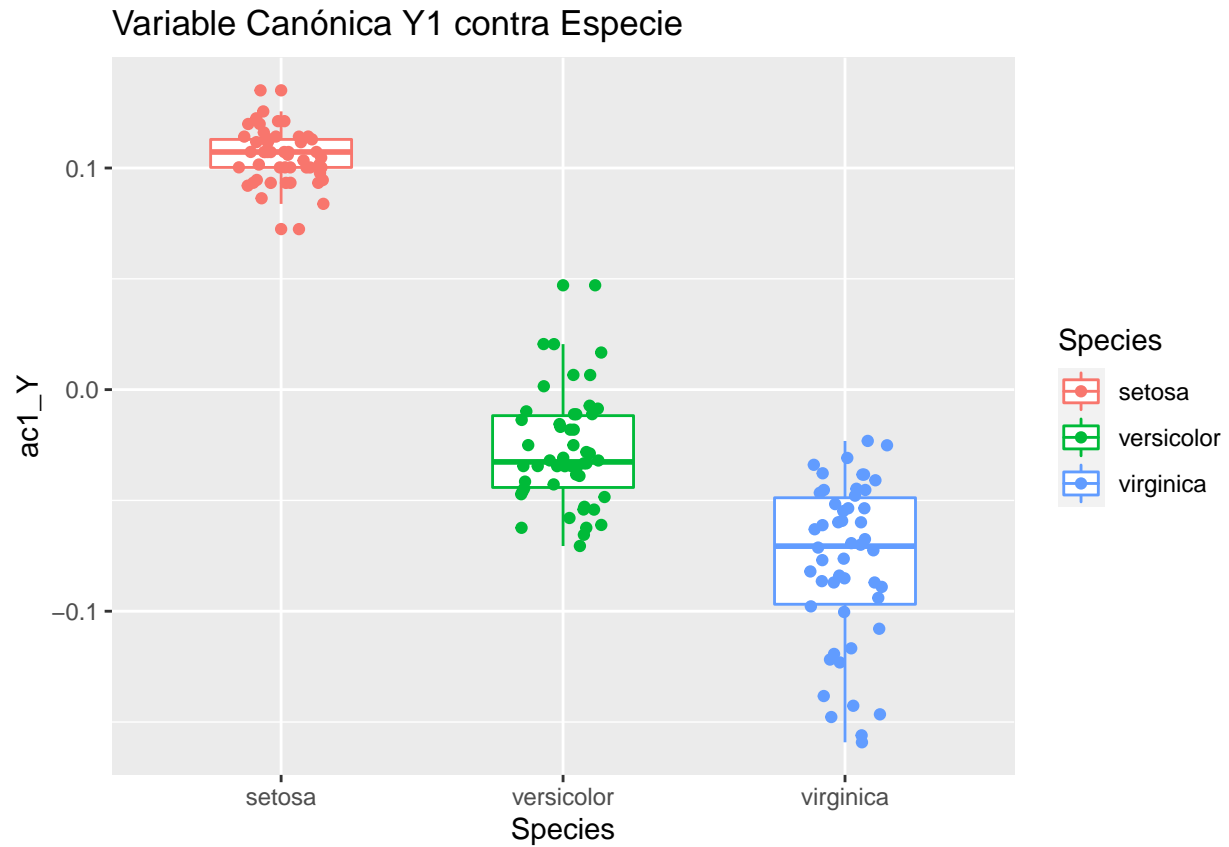
5.- Generación de un boxplot

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



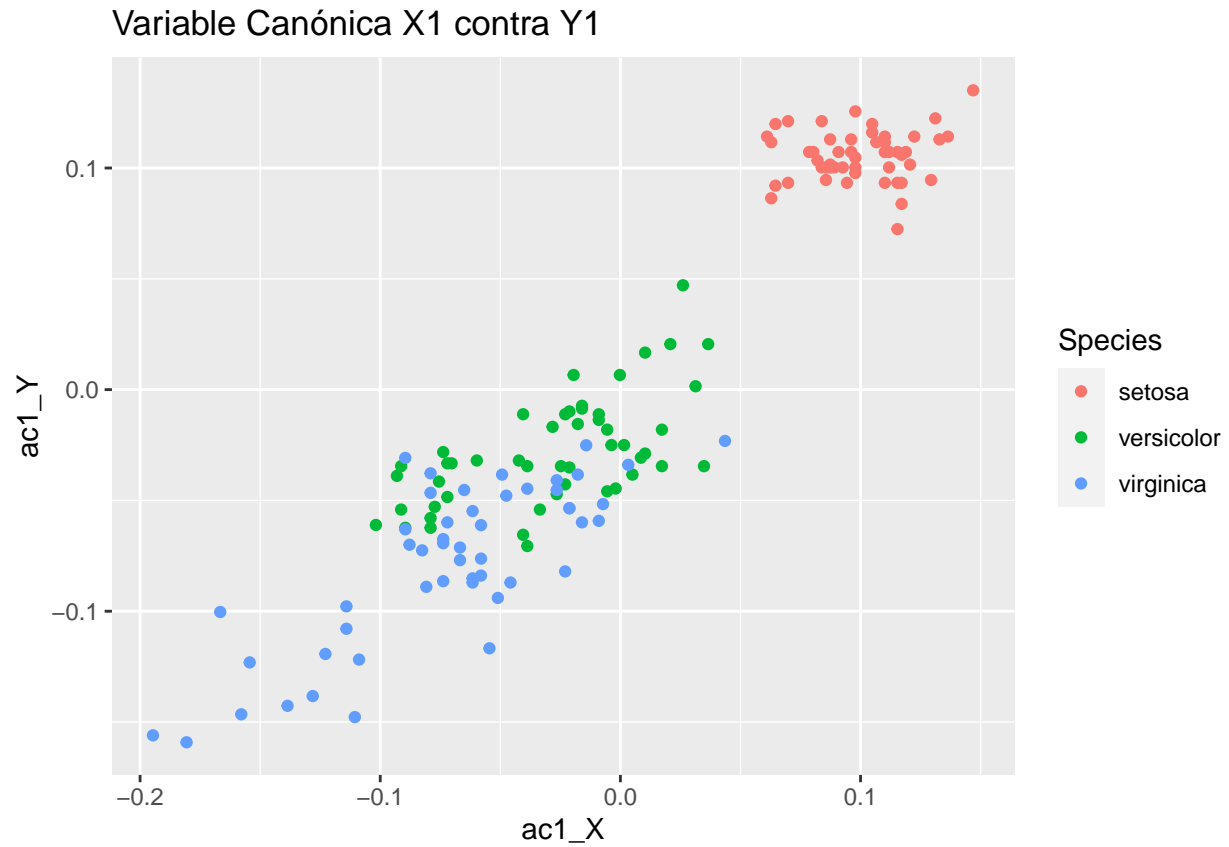
Se observa una correlacion entre la variable canónica X1 y la variable latente Especie.

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



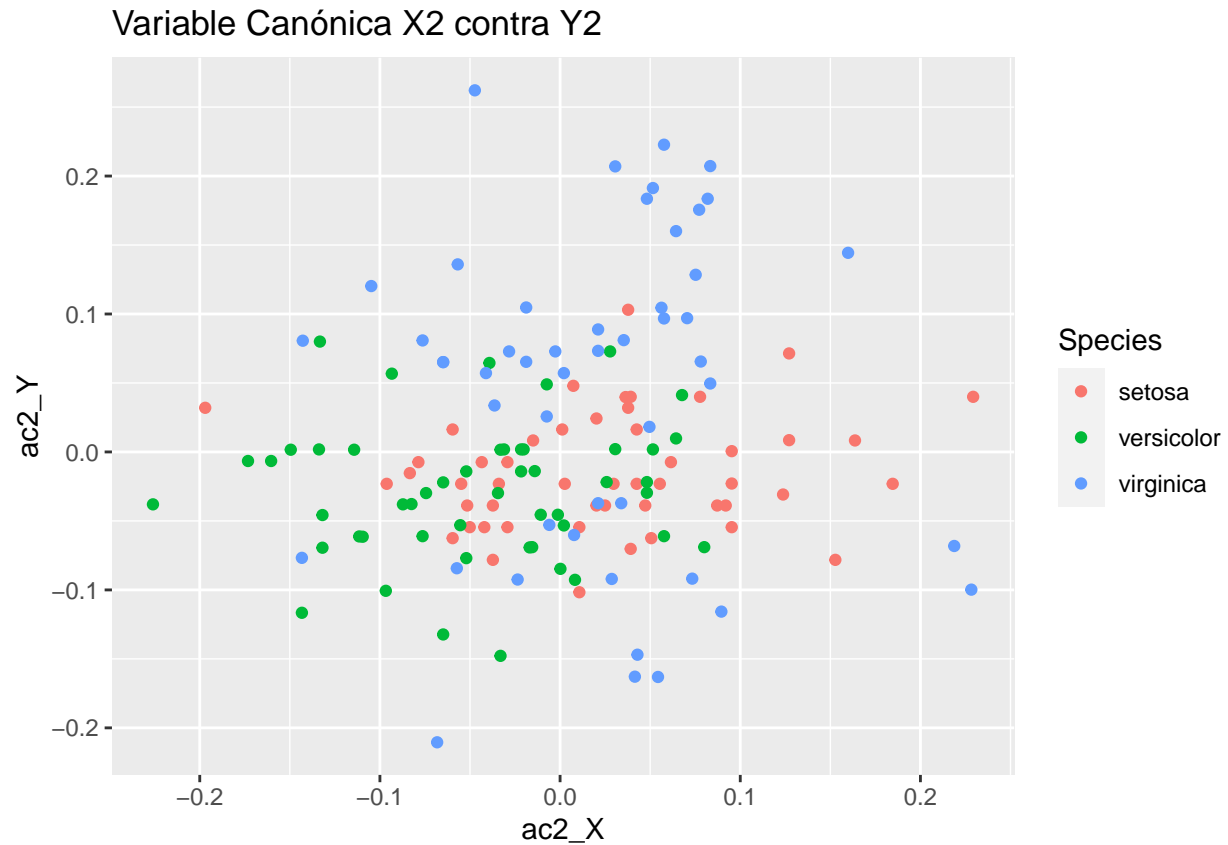
6.- Construcción de un scatter plot con las variables X1 y Y1, separados por especie

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



7.- Scatter plot con las variables canónicas X2 y Y2 separadas por género.

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

No se identifica correlación entre el conjunto de variables X2 y Y2 separadas por género.

8.- Generación de la ecuación canónica

```
ac$xcoef
```

```
##           [,1]      [,2]
## Sepal.Length -0.07251736 0.03932826
## Sepal.Width  0.03052965 0.07663824
```

```
ac$ycoef
```

```
##           [,1]      [,2]
## Petal.Length -0.12279948 -0.2774814
## Petal.Width  0.04332444 0.3003309
```

9.- Sustitución en la ecuación canónica general $U1 = -0.072(\text{Sepal.Length}) + 0.030 (\text{Sepal.Width})$

$V1 = -0.122(\text{Petal.Length}) + 0.043(\text{Petal.Width})$