

Análisis Discriminante ejercicio 1

Problema 1

Datos malaria

Description
The malaria data frame has 100 rows and 4 columns.

Type
malaria

Format
This data frame contains the following columns:
 -surper: subject code
 -age: age in years.
 -ab: antibody level
 -mal: a numeric vector code; Malaria if no. 1-yes.

Details
A random sample of 100 children aged 4-15 years from a village in Ghana. The children were followed for a period of 6 months. At the beginning of the study, values of a particular antibody were assessed. Based on observation during the study period, the children were categorized into two groups: individuals with and without symptoms of malaria.

(1) Determine las funciones discriminantes que serían necesarias para representar las poblaciones consideradas

(2) Analice el ajuste del modelo discriminante que construyó comparando la predicción de pertenencia a los grupos con la realidad

(3) Utilice las funciones discriminantes para representar gráficamente los grupos y caracterizarlos

(4) Si una persona tiene

| | |
|-----|-----|
| age | ab |
| 20 | 200 |

Diga si clasificaría o no en el grupo con malaria al aplicar el análisis discriminante.

Solución:

```
> #PAQUETE
> library(MASS)
> #CONSTRUCCION MODELO FUNCION DISCRIMINANTE
> m=lda(mal~age+ab)
> #visualizar la salida
> m
call:
lda(mal ~ age + ab)
```

Prior probabilities of groups:

| | |
|------|------|
| 0 | 1 |
| 0.73 | 0.27 |

$$S = \min(q-1, p)$$

Group means:

| | |
|------------|----------|
| age | ab |
| 0 9.273973 | 380.0411 |
| 1 7.740741 | 126.1111 |

Coefficients of linear discriminants:

| | |
|-----|--------------|
| LD1 | |
| age | -0.121733727 |
| ab | 0.001781491 |

$$LD1 = -0.12age - 0.001ab$$

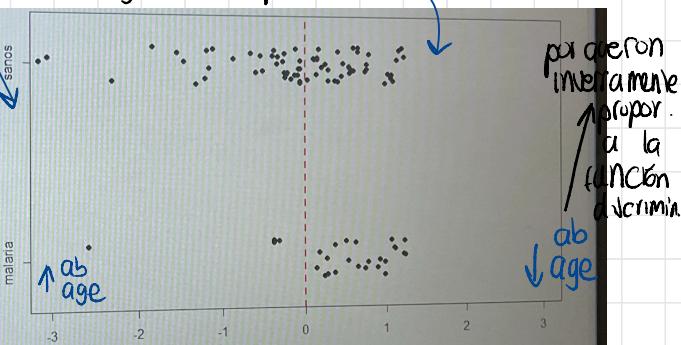
titulos de grupos Ade variables

2-1

→ 1 función discriminante (LD)

Ya que solo tenemos una función discriminante (solo un eje entonces) faremos un diagrama de puntos

```
9 #visualizar la salida
10 m
11
12 #coordenadas de los datos(proyecciones)
13 g= predict(m,dimen=1)$x
14
15 #grafico de los datos
16 #es un eje entonces solo se hace undiagrama de puntos
17 z=list("malaria"=g[mal==1],"sanos"=g[mal==0])
18 stripchart(z,method="jitter",xlim=c(-3,3),pch=16)
19 abline(v=0,col="red",lty=2,lwd=2)
20
21 #primero hay que determinar las coordenadas de los dat
22 #primeras funciones discriminantes
23 prediccion=predict(m)$class
24
25 *
```



```
> #primeras funciones discriminantes  
> prediccion=predict(m)$class  
> #tabla cruzada  
> table(mal,prediccion)  
prediccion  
mal  0  1  
 0 73  0  
 1 27  0
```

Ejercicio 2

BASE DE DATOS: Iris (esta base de datos forma parte de R)

Base de datos: 150 observaciones de flores de la planta iris.

Existen tres tipos de clases de flores iris:

- virginica
- setosa
- versicolor

Hay 50 observaciones de cada una.

Las variables o atributos que se miden de cada flor son:

- 1 El tipo de flor (grupo) como variable categórica.
- 2 El largo y el ancho del pétalo en cm como variables numéricas.
- 3 El largo y el ancho del sépalo en cm como variables numéricas.

(4) Si usted camina por el bosque y se encuentra 3 flores, las mide y obtiene los resultados siguientes:

| Sepal.Length | Sepal.Width | Petal.Length | Petal.Width |
|--------------|-------------|--------------|-------------|
| 5.1 | 2.9 | 1.8 | 0.1 |
| 6 | 2.5 | 4.2 | 1 |
| 7 | 3 | 6 | 1.9 |

Utilice el modelo para la discriminación que acaba de construir para clasificar cada una de las flores que se encontró



Solución:

```
> #PAQUETE
> library(MASS)
> #CONSTRUCCION MODELO FUNCION DISCRIMINANTE
> m=lda(Species~Sepal.Length+Sepal.Width+Petal.Length+Petal.Width)
> #visualizar la salida
> m
Call:
lda(Species ~ Sepal.Length + Sepal.Width + Petal.Length + Petal.Width)

Prior probabilities of groups:
  setosa versicolor virginica
0.3333333 0.3333333 0.3333333

Group means:
  Sepal.Length Sepal.Width Petal.Length Petal.Width
setosa      5.006     3.428     1.462     0.246
versicolor   5.936     2.770     4.260     1.326
virginica    6.588     2.974     5.552     2.026

Coefficients of linear discriminants:
LD1 LD2
Sepal.Length 0.8293776 -0.02410215
Sepal.Width  1.5344731 -2.16452123
Petal.Length -2.2021171  0.93192121
Petal.Width  -2.8104603 -2.83918785

Proportion of trace:
LD1 LD2
0.9912 0.0088
```

2 ← función discriminante
 3-1 4 ↑ q 1, p ↓ variables

$$LD_1 = 0.81 \text{Sepal.length} + 1.53 \text{Sepal.width} - 2.20 \text{Petal.length} - 2.81 \text{Petal.width}$$

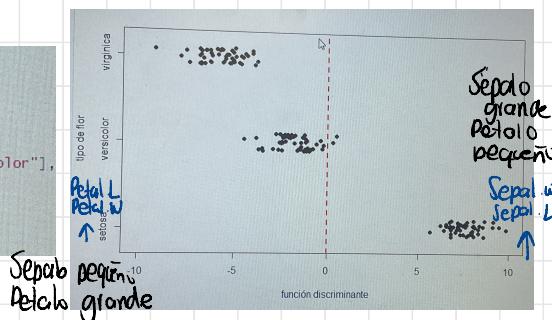
Como la segunda función discriminante es tan chiquita solo trabajaremos con una (LD_1) que va a determinar el eje de discriminación

Logra una separación de grupos del 99%

Diagrama de puntos

```
#coordenadas de los datos(proyecciones)
g=predict(m,dimen=1)$x
```

```
#grafico de los datos
#es un eje etonos solo se hace un diagrama de puntos
z=list("setosa"=g[Species=="setosa"], "versicolor"=g[Species=="versicolor"],
"virginica"=g[Species=="virginica"])
stripchart(z,method="jitter",pch=16,xlim=c(-10,10),
,xlab="función discriminante", ylab="tipo de flor")
abline(v=0,col="red",lty=2,lwd=2)
```



```

> #primeras funciones discriminantes
> prediccion=predict(m)$class
> #tabla cruzada
> table(species,prediccion)

```

| Species | prediccion | | |
|------------|------------|------------|-----------|
| | setosa | versicolor | virginica |
| setosa | 50 | 0 | 0 |
| versicolor | 0 | 48 | 2 |
| virginica | 0 | 1 | 49 |

- Había 50 setosa y clasifico las 50 como setosa
- Había 50 versicolor y clasifico las 48 como versicolor y 2 como virginica
- Clasificados mal
- Había 50 virginica y clasifico las 49 como virginica y 1 como versicolor

- Pronóstico

→ observaciones

```

#pronostico
PP=iris[-c(1:150),-5]
data.entry(PP)
#solo tomo la probabilidad
predict(m,PP)$posterior

```

| | setosa | versicolor | virginica |
|---|--------------|--------------|--------------|
| 1 | 1.000000e+00 | 4.156756e-15 | 1.009250e-33 |
| 2 | 5.080435e-18 | 9.999980e-01 | 2.001384e-06 |
| 3 | 4.812765e-41 | 1.645849e-04 | 9.998354e-01 |

→ quitar columna a predecir
"Species"

→ entrar los datos que quiero predecir

→ la primera flor tiene mayor probabilidad de ser setosa, la segunda de ser versicolor y la tercera de ser virginica .

3er ejercicio

| Base de datos vino | |
|--|--------------------------------|
| Type The type of wine, into one of three classes, 1 (59 obs), 2(71 obs), and 3 (48 obs). | |
| Alcohol | Alcohol |
| Malic | Malic acid |
| Ash | Ash |
| Alcalinity | Alkalinity of ash |
| Magnesium | Magnesium |
| Phenols | Total phenols |
| Flavanoids | Flavanoids |
| Nonflavanoids | Nonflavanoid phenols |
| Pronanthocyanins | Pronanthocyanins |
| Color | Color intensity. |
| Hue | Hue |
| Dilution | D280/OD315 of diluted wines. |
| Proline | Proline |
| (4) Un vino que tenga | |
| Alcohol | Alcohol 15 |
| Malic | Malic acid 2 |
| Ash | Ash 2.5 |
| Alcalinity | Alkalinity of ash 15 |
| Magnesium | Magnesium 112 |
| Phenols | Total phenols 4 |
| Flavanoids | Flavanoids 38 |
| Nonflavanoids | Nonflavanoid phenols 0.5 |
| Pronanthocyanins | Pronanthocyanins 2 |
| Color | Color intensity. 7 |
| Hue | Hue 1.5 |
| Dilution | D280/OD315 of diluted wines. 3 |
| Proline | Proline 1500 |

En cuál de los tipos clasificaria

Solución:

```

attach(wine)
head(wine)
#PAQUETE
library(MASS)

#CONSTRUCCION MODELO FUNCION DISCRIMINANTE
m lda(Type~Alcohol+Malic+Ash+Alcalinity+Magnesium+Phenols+Flavanoids+Nonflavanoids+Pronanthocyanins+color+Hue+Dilution+Proline)
#visualizar la salida
m

Prior probabilities of groups:
      1      2      3 
0.3314607 0.3988764 0.2696629 

Group means:
  Alcohol   Malic     Ash Alcalinity Magnesium Phenols Flavanoids Nonflavanoids Pronanthocyanins   Color 
1 13.74475 2.010678 2.455593 17.03729 106.3390 2.840169 2.9823729 0.290000 1.899322 5.528305 
2 12.27873 1.932676 2.244789 20.23803 94.5493 2.258873 2.0808451 0.363662 1.630282 3.086620 
3 13.15375 3.333750 2.437083 21.41667 99.3125 1.678750 0.7814583 0.447500 1.153542 7.396250 
      Hue Dilution Proline 
1  1.0620339 3.157797 1115.7119 
2  1.0562817 2.785352 519.5070 
3  0.6827083 1.683542 629.8958 

Coefficients of linear discriminants:
          LD1        LD2
Alcohol -0.403399781  0.8717930699 
Malic   0.165254596  0.3053797325 
Ash    -0.369075256  2.3458497486 
Alcalinity 0.154797889 -0.1463807654 
Magnesium -0.002163496 -0.0004627565 
Phenols  0.618052068 -0.0322128171 
Flavanoids -1.661191235 -0.4919980543 
Nonflavanoids -1.495818440 -1.6309537953 
Pronanthocyanins 0.134092628 -0.3070875776 
Color    0.355055710  0.2532306865 
Hue     -0.818036073 -1.5156344987 
Dilution -1.157559376  0.0511839665 
Proline  -0.002691206  0.0028529846 

Proportion of trace:
         LD1       LD2
  0.6875  0.3125

```