

TRABAJO PRÁCTICO Nº 6 Análisis Discriminante

Ejercicio 1

La base de datos **Gorriones.xls** consiste en 49 registros de gorriones sobre los que se han medido las variables “Largo total”, “extensión alar”, “largo del pico y la cabeza”, “largo del humero”, “largo de la quilla del esternón”. Cada uno de estos gorriones ha sobrevivido o no. Hay también registrada una variable categórica respecto de la clasificación como sobreviviente o no del gorrión.

- i Compare las medias de cada una de las variables entre los grupos. Realice una exploración grafica.
- ii Compare los vectores medios de ambos grupos. Tiene sentido realizar un análisis discriminante?
- iii Realice el análisis discriminante a partir de las variables que considere adecuada. incluir.
- iv Satisface los supuestos del modelo?. Resulta una buena clasificación?.

Ejercicio 2

El archivo de datos **hemofilia.xls** contiene 74 registros de pacientes sobre los cuales se han medido dos variables continuas Actividad AHF y Antígeno AHF. Con estas variables se pretende predecir el grupo de pertenencia respecto de la portación de Hemofilia.

- a. Considera que ambas variables pueden ser de ayuda para esta clasificación?
- b. Realice un gráfico bivariado de ambos grupos conjuntamente.
- c. Le parece a partir de este grafico que una función discriminante lineal es adecuada.
- d. Realice un análisis discriminante con 50 registros elegidos al azar.
- e. Utilice los restantes registros para estimar la calidad de la regla discriminante.

Ejercicio 3

Los datos del archivo **pulso.idb** contienen el pulso antes de realizar un ejercicio y el pulso después de realizarlo de un conjunto de 40 individuos, hombres y mujeres entre los cuales hay fumadores y no fumadores.

- a- Interesa saber si la información del pulso antes y después de correr permite discriminar el sexo.
- b- Ídem con la categoría de fumador
- c- En cuál de los dos casos discrimina mejor?.

Ejercicio 4

La base de datos **IRIS.TXT** contiene 150 registros correspondientes a 3 especies de la flor Iris. Los datos pertenecen a un clásico ejemplo debido a Fisher (1936). El objetivo consiste en clasificar estas subespecies a partir de las 4 variables que incluyen medidas del sépalos y del pétalo de cada flor.

- i. analice cuales valores medios son diferentes en las especies.
- ii. aplique alguna regla de clasificación para discriminar los 3 grupos.
- iii. ¿Cuál es el porcentaje de bien clasificados? ¿Y los porcentajes de bien clasificados para cada especie?
- iv. ¿Qué registros no clasificó correctamente? ¿Puede dar una opinión sobre por qué pasó esto?

Ejercicio 5

La base de datos **BAROTRO2.DAT** consiste en 37 registros geoposicionales que informan sobre las coordenadas en las cuales 37 tormentas se transformaron en huracanes, para 2 clasificaciones de huracanes: Baro y Trop.

Estos datos son ficticios y pertenecen a Elsner, Lehmiller, and Kimberlain (1996).

i Realice un análisis discriminante teniendo como objetivo la clasificación de los huracanes.

ii Encuentre la expresión de la función discriminante.

iii ¿Qué cantidad de huracanes han sido bien clasificados? ¿Qué opina sobre este resultado?

iv Apóyese en el uso de gráficos para entender el resultado del ítem anterior.