

Revisión Practica Segundo Parcial AID

Ejercicio 1

Se analiza la leche de las hembras de 20 mamíferos, obteniéndose los porcentajes en agua, proteínas, grasa y lactosa siguientes:

Mamífero	Agua	Proteínas	Grasa	Lactosa
CABALLO	90.1	2.6	1	6.9
BURRO	90.3	1.7	1.4	6.2
CEBRA	86.2	3	4.8	5.3
HAMSTER	81.9	7.4	7.2	2.7
RATA	72.5	9.2	12.6	3.3
OVEJA	82	5.6	6.4	4.7
RENO	64.8	10.7	20.3	2.5
MULA	90	2	1.8	5.5
CERDO	82.8	7.1	5.1	3.7
CAMELLO	87.7	3.5	3.4	4.8
BÚFALO	82.1	5.9	7.9	4.7
ZORRO	81.6	6.6	5.9	4.9
CONEJO	71.3	12.3	13.1	1.9
LLAMA	86.5	3.9	3.2	5.6
CIERVO	65.9	10.4	19.7	2.6
BISONTE	86.9	4.8	1.7	5.7
GATO	81.6	10.1	6.3	4.4
PERRO	76.3	9.3	9.5	3
FOCA	46.4	9.7	42	0
DELFIN	44.9	10.6	34.9	0.9

Clasificar estos mamíferos de forma jerárquica atendiendo a dichas variables realizando el estudio de acuerdo a las siguientes opciones de medidas de asociación y métodos de amalgamamiento:

- a) Distancia euclídea y linkage simple
- b) Distancia Manhattan y linkage simple
- c) Distancia euclídea y método del centroide
- d) Distancia Manhattan y método del centroide

Para cada uno de las anteriores opciones se pide la matriz de distancias, el dendrograma y la clasificación resultante si se toman 2, 3 o 4 grupos diferentes.

Elegir una de todas estas clasificaciones que considere adecuada y explicar por qué la ha elegido.

Ejercicio 2

Un fabricante sospecha que los lotes de materia prima recibidos de un proveedor difieren significativamente de su contenido en calcio. Elige al azar 5 lotes diferentes y un químico hace cinco determinaciones del contenido en calcio de cada lote. Los resultados obtenidos han sido

lote1	lote2	lote3	lote4	lote5
23.46	23.59	23.51	23.28	23.29
23.48	23.46	23.64	23.4	23.46
23.56	23.42	23.46	23.37	23.37
23.39	23.49	23.52	23.46	23.32
23.4	23.5	23.49	23.29	23.38

- a- Establecer las hipótesis de interés
- b- Señalar los supuestos del modelo a testear.
- c- Realizar un test con un nivel de significación del 0.05
- d- Analizar el cumplimiento de los supuestos del modelo, detallando en cada caso las hipótesis testeadas.
- e- Si el fabricante tiene dos líneas de productos, cuál de los lotes seleccionaría para la línea con alto contenido de calcio y cuál para la línea de bajo contenido de calcio. Fundamente su respuesta.

Ejercicio 3

En este ejercicio utilizaremos la base de datos Wisconsin cáncer de mama.

La base consiste en 683 casos de potenciales tumores cancerígenos en Wisconsin, de los cuales 283 resultaron malignos. La determinación de la malignidad del tumor es generalmente posterior a un procedimiento quirúrgico. El objetivo de este estudio es determinar si la extracción con una pequeña aguja de tejido.

Se han registrado para cada caso las siguientes 10 variables:

Class: 0 si es maligno, 1 si es benigno

Adhes: adhesión marginal

BNucl: núcleos desnudos

Chrom: cromatina suave

Epith: tamaño de las células epiteliales

Mitos: mitosis

NNucl: núcleos normales

Thick: grosor del borde

UShap: uniformidad celular

USize: tamaño celular

- a- Hallar el vector medio total y el vector medio por grupos.
- b- Comparar las variables de a una entre los grupos mediante análisis univariado. Cual/es le parece que pueden aportar a discriminar entre los grupos?
- c- Compare en forma multivariada los dos grupos.
- d- Utilice toda la información disponible para construir una función discriminante lineal o cuadrática según corresponda, entre los grupos.
- e- Estime en forma ingenua la capacidad discriminante de esta función.
- f- Seleccione aleatoriamente el 70% de la base. Construya una regla discriminante lineal y utilice el 30% restante de la base para cuantificar el poder discriminante de la regla construida.
- g- Compare las dos reglas y las dos calificaciones.