

Nombre:

DNI:

## Cuestionario

*Cada respuesta válida se contabiliza con 1 punto. Por cada fallo se resta 0,5 puntos*

*Marcar la opción correcta*

1) La distancia de Mahalanobis se calcula con la expresión:

a)  $d_i = [(x_i - \bar{x})' S^{-1} (x_i - \bar{x})]^{1/2}$

b)  $d_{ij} = \left[ \sum_{s=1}^p (x_{is} - x_{js})^2 \right]^{1/2}$

c)  $H_j = \frac{1}{n} \frac{\sum (d_{ij} - s_j^2)^2}{s_j^4}$

d) Ninguna de las anteriores

2) La medida de la varianza adimensional que tiene en cuenta la dependencia entre las variables y no depende de las unidades es:

a)  $V_{mp} = \frac{1}{np} \sum_{i=1}^n (x_i - \bar{x})'(x_i - \bar{x}) = \bar{s}^2$

b)  $V_m = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})'(x_i - \bar{x}) = \text{Tr}(\mathbf{S})$

c)  $VE = |\mathbf{S}|^{1/p}$

d)  $VG = |\mathbf{S}|$

## CLASE 1

Cada respuesta válida se contabiliza con 0,4 puntos. Por cada fallo se resta 0,2 puntos

Las siguientes afirmaciones ¿son verdaderas o falsas?

Nº	AFIRMACIONES	V/F
1	Un objetivo del análisis multivariante es resumir los datos sin perder la información	V
2	Un objetivo del análisis multivariante es encontrar grupos de individuos y asignar individuos nuevos a éstos grupos	V
3	Un objetivo del análisis multivariante es aumentar la cantidad de variables e información	F
4	Las medidas empleadas para describir los datos multivariantes son la media y la varianza de las variables	F
5	Las medidas empleadas para describir los datos multivariantes son la media de las medias y la media de las varianzas	F
6	Las medidas empleadas para describir los datos multivariantes son las matrices de medias, matriz de varianzas-covarianzas y medidas de dependencia lineal	V
7	Los datos multivariantes se representan con una matriz <b>X</b> de dimensiones $n \times p$ , llamada matriz de datos, de elementos $x_{ij}$ , con $i=1, \dots, n$ que representan los individuos y $j=1, \dots, p$ que representan las variables.	V
8	Los datos multivariantes se representan con una matriz <b>X</b> de dimensiones $n \times n$ , llamada matriz de datos, de elementos $x_{ij}$ , con $i=1, \dots, n$ que representan las medias para los individuos y $j=1, \dots, p$ que representan las variables.	F
9	Los datos multivariantes se representan con una matriz <b>X</b> de dimensiones $p \times n$ , llamada matriz de datos, de elementos $x_{ij}$ , con $i=1, \dots, p$ que representan las variables y $j=1, \dots, n$ que representan el individuo	F
10	La transformación de datos se realiza para facilitar el uso de las variables y obtener homogeneidad en sus propiedades.	V
11	La transformación más usada en datos multivariantes es la Box -Cox	F
12	La transformación más usada en datos multivariantes es la estandarización	V
13	Las observaciones atípicas son equivalentes a datos no observados y deben calcularse con el resto	F
14	Las observaciones atípicas afectan a los resultados globales modificando la matriz de medias y la de varianzas-covarianzas	V
15	Las observaciones atípicas son diferentes del resto y se deben eliminar del análisis conjunto y estudiar con detalle	V
16	Las observaciones atípicas son diferentes del resto y no se deben eliminar del análisis conjunto	F
17	Los gráficos de dispersión y los gráficos de figuras planas se utilizan para reducir las variables.	F
18	Los gráficos de dispersión y los gráficos de figuras planas se utilizan para detectar datos atípicos.	V
19	El análisis de componentes principales permite representar óptimamente en un espacio de dimensión pequeña, observaciones multivariantes	V
20	El análisis de componentes principales permite transformar las variables originales, en general correladas, en nuevas variables incorreladas	V