Ejercicios de Análisis Multivariante
Máster en Estadística. Marzo 2013.

1. Generar una matriz $\mathbf{X}$, de dimensión 4×3 y un vector $\mathbf{u}$, 4×1, ambos de números aleatorios y construir las matrices simétricas $\mathbf{A} = \mathbf{X}'\mathbf{X}$ y $\mathbf{B} = \mathbf{u}\,\mathbf{u}'$.
   (a) Calcular la traza y el determinante de $\mathbf{A}$ y $\mathbf{B}$.
   (b) Obtener los autovalores y autovectores de $\mathbf{A}$ y $\mathbf{B}$.
   (c) Comprobar que la traza y el determinante de $\mathbf{A}$ coinciden respectivamente con la suma y el producto de los autovalores de $\mathbf{A}$.
   (d) Obtener los rangos de $\mathbf{A}$ y $\mathbf{B}$ y comprobar que coinciden, respectivamente, con el número de autovalores no nulos de $\mathbf{A}$ y $\mathbf{B}$.

2. Dada la matriz
$$\mathbf{A} = \begin{pmatrix} 3 & 2 & 0 \\ 2 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$
   (a) Calcular sus autovalores, los de $\mathbf{A}^2$ y los de $\mathbf{A}^{-1}$.
   (b) Calcular una base ortogonal que la diagonalice.

3. Considerar la siguiente matriz
$$\mathbf{A} = \begin{pmatrix} 6 & 10 \\ 10 & 6 \\ 1 & 5 \end{pmatrix}.$$
   (a) Encontrar la inversa generalizada de Moore-Penrose, $\mathbf{A}^-$, de $\mathbf{A}$.
   (b) Comprobar que se cumple la propiedad
$$\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}. \tag{1}$$
   (c) Comprobar que se cumplen las propiedades
      i) $\mathbf{A}^-\mathbf{A}\mathbf{A}^- = \mathbf{A}^-$,
      ii) $\mathbf{A}^-\mathbf{A}$ es simétrica,
      iii) $\mathbf{A}\mathbf{A}^-$ es simétrica.

4. Calcula la matriz simétrica asociada a cada una de las siguientes formas cuadráticas y determina si es definida positiva.
   (a) $Q(x_1, x_2) = 2x_1^2 - 3x_1x_2 + 3x_2^2$,
   (b) $Q(x_1, x_2, x_3) = x_1^2 + x_1x_3 + 0.25x_3^2 + 1.6x_1x_2 + 0.6x_2^2 + 0.8x_2x_3$.

5. Se define la *matriz de centrado* de dimensión n como $\mathbf{H} = \mathbf{I} - \frac{1}{n}\mathbf{1}\,\mathbf{1}'$, donde $\mathbf{I}$ es la matriz identidad de dimensión $n \times n$ y $\mathbf{1}$ es un vector $n \times 1$ de unos. La utilidad de esta matriz $\mathbf{H}$ radica en que, como su nombre indica, se usa para centrar configuraciones de datos: si $\mathbf{X}$ es una matriz de datos de dimensión $n \times p$, entonces $\mathbf{H}\,\mathbf{X}$ es una matriz cuyas columnas tienen media cero.

   Utilizar Matlab para comprobar las dos siguientes propiedades de la matriz de centrado (tomando, por ejemplo, $n = 5$):

(a) $\mathbf{H}$ es idempotente,

(b) $\mathsf{rang}(\mathbf{H}) = \mathsf{tr}(\mathbf{H}) = n - 1$.

6. (Problema 2.2) Los datos de la Tabla 1 corresponden a chalets construidos por diez promotoras que operan a lo largo de la costa española.

Table 1: Diez promotoras de la costa española

Promotora	X_1 =Duración media hipoteca (años)	X_2 =Precio medio (millones euros)	X_3 =Superficie media (m^2) de cocina
1	8.7	0.3	3.1
2	14.3	0.9	7.4
3	18.9	1.8	9.0
4	19.0	0.8	9.4
5	20.5	0.9	8.3
6	14.7	1.1	7.6
7	18.8	2.5	12.6
8	37.3	2.7	18.1
9	12.6	1.3	5.9
10	25.7	3.4	15.9

(a) Dibujar el diagrama de dispersión múltiple y comentar el aspecto del gráfico.

(b) Para X_1 y X_2 calcular, respectivamente, las medias muestrales $\bar{x}_1$ y $\bar{x}_2$, las varianzas muestrales s_{11} y s_{22}, la covarianza entre X_1 y X_2, s_{12}, y la correlación entre ambas, r_{12}. Interpretar el valor obtenido de r_{12}.

(c) Utilizando la matriz de datos $\mathbf{X}$ y la de centrado $\mathbf{H}$ definida en el Ejercicio 5, calcular el vector de medias muestrales $\bar{\mathbf{x}}$ y la matriz de covarianzas muestrales $\mathbf{S}$. A partir de ésta obtener la matriz de correlaciones $\mathbf{R}$.

7. (Problema 2.3) La contaminación por mercurio de peces de aguadulce comestibles es una amenaza directa contra nuestra salud. Entre 1990 y 1991 se llevó a cabo un estudio en 53 lagos de Florida con el fin de examinar los factores que influían en el nivel de contaminación por mercurio. Las variables que se midieron fueron:

X_1 =número de identificación, X_2 =nombre del lago, X_3 =alcalinidad (mg/l de carbonato de calcio), X_4 =pH, X_5 =calcio (mg/l), X_6 =clorofila (mg/l), X_7 =concentración media de mercurio (partes por millón) en el tejido múscular del grupo de peces estudiados en cada lago, X_8 =número de peces estudiados por lago, X_9 =mínimo de la concentración de mercurio en cada grupo de peces, X_{10} =máximo de la concentración de mercurio en cada grupo de peces, X_{11} =estimación (mediante regresión) de la concentración de mercurio en un pez de 3 años (o promedio de mercurio cuando la edad no está disponible), X_{12} =indicador de la edad de los peces. La Tabla 2 contiene los datos de este estudio, disponible en la página web http://lib.stat.cmu.edu/DASL.

(a) Representar de forma conjunta las variables X_3, X_6, X_7 y ver cómo se modifica su dispersión cuando se producen transformaciones (lineales y no lineales) sobre las variables. Considerar como medidas de dispersión global la traza y el determinante de la matriz de covarianzas.

Table 2: Datos del Ejercicio 7 (http://lib.stat.cmu.edu/DASL/Datafiles/MercuryinBass.html)

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}
1	Alligator	5.9	6.1	3.0	0.7	1.23	5	0.85	1.43	1.53	1
2	Annie	3.5	5.1	1.9	3.2	1.33	7	0.92	1.90	1.33	0
3	Apopka	116.0	9.1	44.1	128.3	0.04	6	0.04	0.06	0.04	0
4	Blue Cypress	39.4	6.9	16.4	3.5	0.44	12	0.13	0.84	0.44	0
5	Brick	2.5	4.6	2.9	1.8	1.20	12	0.69	1.50	1.33	1
6	Bryant	19.6	7.3	4.5	44.1	0.27	14	0.04	0.48	0.25	1
7	Cherry	5.2	5.4	2.8	3.4	0.48	10	0.30	0.72	0.45	1
8	Crescent	71.4	8.1	55.2	33.7	0.19	12	0.08	0.38	0.16	1
9	Deer Point	26.4	5.8	9.2	1.6	0.83	24	0.26	1.40	0.72	1
10	Dias	4.8	6.4	4.6	22.5	0.81	12	0.41	1.47	0.81	1
11	Dorr	6.6	5.4	2.7	14.9	0.71	12	0.52	0.86	0.71	1
12	Down	16.5	7.2	13.8	4.0	0.50	12	0.10	0.73	0.51	1
13	Eaton	25.4	7.2	25.2	11.6	0.49	7	0.26	1.01	0.54	1
14	East Tohopekaliga	7.1	5.8	5.2	5.8	1.16	43	0.50	2.03	1.00	1
15	Farm-13	128.0	7.6	86.5	71.1	0.05	11	0.04	0.11	0.05	0
16	George	83.7	8.2	66.5	78.6	0.15	10	0.12	0.18	0.15	1
17	Griffin	108.5	8.7	35.6	80.1	0.19	40	0.07	0.43	0.19	1
18	Harney	61.3	7.8	57.4	13.9	0.77	6	0.32	1.50	0.49	1
19	Hart	6.4	5.8	4.0	4.6	1.08	10	0.64	1.33	1.02	1
20	Hatchineha	31.0	6.7	15.0	17.0	0.98	6	0.67	1.44	0.70	1
21	Iamonia	7.5	4.4	2.0	9.6	0.63	12	0.33	0.93	0.45	1
22	Istokpoga	17.3	6.7	10.7	9.5	0.56	12	0.37	0.94	0.59	1
23	Jackson	12.6	6.1	3.7	21.0	0.41	12	0.25	0.61	0.41	0
24	Josephine	7.0	6.9	6.3	32.1	0.73	12	0.33	2.04	0.81	1
25	Kingsley	10.5	5.5	6.3	1.6	0.34	10	0.25	0.62	0.42	1
26	Kissimmee	30.0	6.9	13.9	21.5	0.59	36	0.23	1.12	0.53	1
27	Lochloosa	55.4	7.3	15.9	24.7	0.34	10	0.17	0.52	0.31	1
28	Louisa	3.9	4.5	3.3	7.0	0.84	8	0.59	1.38	0.87	1
29	Miccasukee	5.5	4.8	1.7	14.8	0.50	11	0.31	0.84	0.50	0
30	Minneola	6.3	5.8	3.3	0.7	0.34	10	0.19	0.69	0.47	1
31	Monroe	67.0	7.8	58.6	43.8	0.28	10	0.16	0.59	0.25	1
32	Newmans	28.8	7.4	10.2	32.7	0.34	10	0.16	0.65	0.41	1
33	Ocean Pond	5.8	3.6	1.6	3.2	0.87	12	0.31	1.90	0.87	0
34	Ocheese Pond	4.5	4.4	1.1	3.2	0.56	13	0.25	1.02	0.56	0
35	Okeechobee	119.1	7.9	38.4	16.1	0.17	12	0.07	0.30	0.16	1
36	Orange	25.4	7.1	8.8	45.2	0.18	13	0.09	0.29	0.16	1
37	Panasoffkee	106.5	6.8	90.7	16.5	0.19	13	0.05	0.37	0.23	1
38	Parker	53.0	8.4	45.6	152.4	0.04	4	0.04	0.06	0.04	0
39	Placid	8.5	7.0	2.5	12.8	0.49	12	0.31	0.63	0.56	1
40	Puzzle	87.6	7.5	85.5	20.1	1.10	10	0.79	1.41	0.89	1
41	Rodman	114.0	7.0	72.6	6.4	0.16	14	0.04	0.26	0.18	1
42	Rousseau	97.5	6.8	45.5	6.2	0.10	12	0.05	0.26	0.19	1
43	Sampson	11.8	5.9	24.2	1.6	0.48	10	0.27	1.05	0.44	1
44	Shipp	66.5	8.3	26.0	68.2	0.21	12	0.05	0.48	0.16	1
45	Talquin	16.0	6.7	41.2	24.1	0.86	12	0.36	1.40	0.67	1
46	Tarpon	5.0	6.2	23.6	9.6	0.52	12	0.31	0.95	0.55	1
47	Trafford	81.5	8.9	20.5	9.6	0.27	6	0.04	0.40	0.27	0
48	Trout	1.2	4.3	2.1	6.4	0.94	10	0.59	1.24	0.98	1
49	Tsala Apopka	34.0	7.0	13.1	4.6	0.40	12	0.08	0.90	0.31	1
50	Weir	15.5	6.9	5.2	16.5	0.43	11	0.23	0.69	0.43	1
51	Tohopekaliga	25.6	6.2	12.6	27.7	0.65	44	0.30	1.10	0.58	1
52	Wildcat	17.3	5.2	3.0	2.6	0.25	12	0.15	0.40	0.28	1
53	Yale	71.8	7.9	20.5	8.8	0.27	12	0.15	0.51	0.25	1

(b) Dibujar el histograma tridimensional correspondiente a X_3 y X_7. Elegir sendas transformaciones no lineales para estas variables de entre las utilizadas en el apartado anterior y dibujar el histograma tridimensional de las variables transformadas.

8. Sea $\mathbf{X}$ un vector con distribución uniforme en el rectángulo $[0,2] \times [3,4]$.

   (a) Especificar la función de densidad de $\mathbf{X}$. Calcular $\mathsf{E}(\mathbf{X})$ y $\mathsf{Var}(\mathbf{X})$.

   (b) Sea $\mathbf{X}_1, \ldots, \mathbf{X}_{30}$ una muestra aleatoria simple de $\mathbf{X}$ y $\bar{\mathbf{X}} = \sum_{i=1}^{30} X_i/30$ la media muestral correspondiente. Calcular $\mathsf{E}(\bar{\mathbf{X}})$ y $\mathsf{Var}(\bar{\mathbf{X}})$.

   (c) Generar con Matlab una realización de la muestra del apartado anterior. Calcular la media $\bar{\mathbf{x}}$ y la matriz de covarianzas muestrales $\mathbf{S}$. Dibujar en un gráfico de dispersión la muestra y marcar los puntos $\mathsf{E}(\bar{\mathbf{X}})$ y $\bar{\mathbf{x}}$.

   (d) Generar con Matlab 40 muestras de tamaño 5, calcular sus correspondientes medias muestrales y dibujar éstas en un gráfico en el que se marque

también $\mathsf{E}(\bar{\mathbf{X}})$. Repetir este proceso en gráficos distintos para 40 muestras de tamaño 20 y otras 40 de tamaño 50. ¿Qué se observa?

9. Sea $\boldsymbol{\mu}$ un vector $p \times 1$ y $\boldsymbol{\Sigma}$ una matriz $p \times p$ simétrica y definida positiva. Fijar un valor de p y generar muestras de tamaño n de una normal $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ para distintos valores de n. Para cada muestra obtener el vector de medias muestrales, $\overline{\mathbf{x}}$, y la matriz de covarianzas muestrales, $\mathbf{S}$, y comprobar que a medida que aumenta n, los valores de $\overline{\mathbf{x}}$ y $\mathbf{S}$ se van acercando a $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$, respectivamente.

   Indicación: El vector $\mathbf{X} = (X_1, X_2, \ldots, X_p)'$ tiene ley normal p-variante si existen p variables aleatorias independientes con ley $N(0,1)$, $Y_1, Y_2, \ldots, Y_p$, tales que

$$\mathbf{X} = \boldsymbol{\mu} + \mathbf{A}\,\mathbf{Y}, \tag{2}$$

   donde $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_p)'$, $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_p)'$ y $\mathbf{A}$ es una matriz $p \times p$. Si las p columnas de $\mathbf{A}$ no son linealmente independientes, alguna de las X_i puede expresarse como combinación lineal de las otras; en caso contrario, se trata de una distribución p-variante no singular.

   Si el vector $\mathbf{X}$ verifica (2), entonces

$$\mathsf{E}(\mathbf{X}) = \boldsymbol{\mu}, \quad \mathsf{Var}(\mathbf{X}) = \mathbf{A}'\,\mathbf{A},$$

   y se dice que $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, donde $\boldsymbol{\Sigma} = \mathbf{A}'\,\mathbf{A}$ es definida positiva si $\mathbf{A}$ es regular. Por ejemplo, $\mathbf{A}$ puede ser la matriz de Cholesky de $\boldsymbol{\Sigma}$ (ver Peña 2002), que calculamos en Matlab con la orden `A = chol(Sigma)`.

10. Una distribución muy relacionada con la ley normal multivariante, y que es el análogo multivariante de la ley χ^2, es la distribución Wishart. Dados $\mathbf{X}_1, \ldots, \mathbf{X}_n$ vectores aleatorios i.i.d.$\sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$, la matriz $p \times p$

$$\mathbf{Q} = \sum_{i=1}^{n} \mathbf{X}_i\,\mathbf{X}_i' \sim W_p(\boldsymbol{\Sigma}, n)$$

   sigue una ley Wishart con parámetro de escala $\boldsymbol{\Sigma}$ y n grados de libertad.

   Dadas las variables aleatorias $\mathbf{Z} \sim N_p(\mathbf{0}, \mathbf{I})$ y $\mathbf{Q} \sim W_p(\mathbf{I}, n)$ estocásticamente independientes, la variable aleatoria

$$T^2 = n\,\mathbf{Z}'\,\mathbf{Q}^{-1}\,\mathbf{Z} \sim T^2(p, n)$$

   sigue una ley T^2 de Hotelling con p y n grados de libertad. Si $p = 1$, entonces $T^2(1, n)$ es el cuadrado de una variable aleatoria con ley t de Student y n grados de libertad. En general, $T^2(p, n)$ es proporcional a una F de Fisher

$$\frac{n - p + 1}{n\,p}\,T^2(p, n) = F(p, n - p + 1). \tag{3}$$

   La variable T^2 se utiliza de manera análoga a la ley t de Student, en contrastes sobre medias multivariantes.

   Para p y n fijos, generar una muestra de tamaño N de una ley $T^2(p, n)$ de Hotelling. Representar los resultados mediante un histograma.

11. Si $\mathbf{A} \sim W_p(\mathbf{\Sigma}, a)$ y $\mathbf{B} \sim W_p(\mathbf{\Sigma}, b)$ son independientes, $\mathbf{\Sigma}$ es regular y $a \geq p$, la variable aleatoria
$$\Lambda = \frac{|\mathbf{A}|}{|\mathbf{A}+\mathbf{B}|}$$
tiene una ley Lambda de Wilks, $\Lambda(p, a, b)$, con parámetros p, a y b.

La ley Λ no depende del parámetro $\mathbf{\Sigma}$ de $\mathbf{A}$ y $\mathbf{B}$, por lo que es suficiente considerarla para $\mathbf{\Sigma} = \mathbf{I}$. Tiene la misma distribución que un producto de b v.a. independientes con distribución Beta, es decir, si $L \sim \Lambda(p, a, b)$ entonces
$$L = \prod_{i=1}^{b} u_i, \quad \text{donde } u_i \sim \text{Beta}\left(\frac{a+i-p}{2}, \frac{p}{2}\right).$$
Generar una muestra de tamaño N de una ley Λ de Wilks. Representar los resultados mediante un histograma.

12. (Problema 3.19) En una fábrica de zumos se diseña el siguiente procedimiento de control de calidad. Se toma una muestra piloto (véase la Tabla 3) de $n = 50$ extracciones de zumo cuando el proceso de fabricación funciona correctamente y en ella se mide la concentración de $p = 11$ aminoácidos, $\mathbf{X} = (X_1, \ldots, X_{11})'$. Suponer que $\mathbf{X}$ sigue una distribución normal. A continuación cada día se observan estas mismas variables con objeto de detectar algún cambio significativo en la calidad del proceso (véase Tabla 4). Suponer que estas sucesivas observaciones, $\mathbf{y}_i$, $i = 1, \ldots, 10$, son independientes de la muestra piloto y entre sí.

Construir un gráfico de control para estos nuevos diez días como se indica a continuación. En primer lugar calcular la media $\bar{\mathbf{x}}$ y la matriz de covarianzas $\mathbf{S}$ para la muestra piloto. A continuación para la observación $\mathbf{y}_i$ construir el estadístico
$$T^2(i) = \frac{n}{n+1}(\mathbf{y}_i - \bar{\mathbf{x}})'\mathbf{S}^{-1}(\mathbf{y}_i - \bar{\mathbf{x}})$$
que debería seguir una $T^2(p, n-1)$ si la distribución de $\mathbf{y}_i$ es la misma que la de la muestra piloto.

Representar secuencialmente los valores de $T^2(i)$ en un gráfico y marcar en él un límite de control $\text{LC} = \frac{(n-1)\,p}{n-p} F^{\alpha}(p, n-p)$, siendo α el nivel de significación que deseemos fijar ($\alpha = 0.05$, por ejemplo). Parar el proceso de fabricación el primer día i que una observación $\mathbf{y}_i$ esté fuera de la región de control, es decir, $\mathbf{y}_i > \text{LC}$.

13. Los datos de la Tabla 2 contienen información sobre la contaminación por mercurio de peces de aguadulce comestibles en 53 lagos de Florida. Considerar solamente las variables X_3 =alcalinidad (mg/l de carbonato de calcio), X_4 =pH, X_5 =calcio (mg/l), X_6 =clorofila (mg/l), X_7 =concentración media de mercurio (partes por millón) en el tejido múscular del grupo de peces estudiados en cada lago, X_9 =mínimo de la concentración de mercurio en cada grupo de peces, X_{10} =máximo de la concentración de mercurio en cada grupo de peces, y realizar un análisis de componentes principales. Razonar a partir de qué matriz, $\mathbf{S}$ o $\mathbf{R}$, es más adecuado realizar dicho análisis. Interpretar las dos primeras componentes principales.

Table 3: Concentraciones de 11 aminoácidos en 50 zumos (Ejercicio 12)

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}
0.480	5.234	2.620	2.857	0.803	13.897	0.326	0.902	0.164	0.183	4.155
0.245	1.312	2.115	8.077	0.974	9.227	0.252	2.703	-0.006	-0.061	1.995
0.276	3.402	2.527	5.447	0.957	13.474	0.299	2.341	0.094	0.113	3.541
0.482	6.554	2.631	5.134	0.671	12.333	0.259	1.473	0.216	0.112	3.941
0.400	4.011	2.528	3.716	0.805	10.382	0.266	0.697	0.201	0.159	4.361
0.336	4.001	3.083	4.626	0.904	7.834	0.156	0.898	0.130	0.061	2.444
0.379	3.366	2.099	6.142	0.977	17.366	0.384	2.451	0.204	0.063	3.177
0.369	4.550	2.242	3.609	0.672	12.353	0.291	0.975	0.158	0.201	3.185
0.396	5.479	2.231	4.264	0.786	15.248	0.244	1.318	0.064	0.116	3.989
0.325	3.573	2.446	5.087	0.708	10.791	0.183	1.500	0.075	0.122	3.675
0.404	4.195	3.226	4.959	0.948	14.880	0.460	0.910	0.151	0.280	5.071
0.367	4.756	2.891	4.264	0.799	13.443	0.270	0.927	0.195	0.194	3.932
0.340	3.640	3.075	4.937	0.821	13.782	0.296	1.659	0.214	0.107	3.507
0.281	2.872	2.299	4.543	0.926	8.921	0.205	0.901	0.072	0.102	2.567
0.373	4.212	2.769	5.014	1.060	15.577	0.288	1.664	0.175	0.095	3.788
0.356	3.629	3.435	4.694	0.843	11.503	0.253	1.249	0.106	0.198	3.147
0.426	5.087	2.797	3.029	0.758	11.412	0.311	0.912	0.175	0.154	3.759
0.262	2.722	3.439	6.223	1.018	8.324	0.233	1.200	0.083	0.108	3.065
0.422	5.769	1.948	4.525	0.576	15.151	0.342	1.282	0.014	0.087	4.773
0.242	2.074	3.090	6.822	0.987	10.655	0.274	1.858	0.065	0.072	2.754
0.288	3.413	3.338	5.562	1.054	9.265	0.276	1.830	0.181	0.071	2.710
0.409	4.701	3.340	5.531	1.237	13.800	0.274	1.598	0.159	0.102	3.032
0.382	4.362	2.588	3.941	0.779	14.441	0.265	1.480	0.213	0.147	3.372
0.277	3.261	2.730	4.335	0.747	7.909	0.181	1.014	0.102	0.108	2.910
0.416	3.511	2.822	5.128	0.992	15.695	0.298	1.864	0.268	0.108	4.097
0.238	2.840	3.180	6.392	1.293	9.059	0.209	1.529	0.120	0.043	3.000
0.544	6.523	3.333	3.431	0.759	13.712	0.334	0.423	0.128	0.240	5.209
0.404	4.119	2.689	4.599	0.744	13.960	0.264	1.241	0.099	0.126	4.185
0.384	4.126	2.440	5.626	0.965	11.960	0.224	1.647	0.203	0.086	3.102
0.290	2.823	2.731	6.063	0.688	7.677	0.217	1.343	0.065	0.073	3.250
0.598	5.807	2.525	4.633	0.889	16.131	0.368	1.462	0.221	0.169	4.544
0.337	4.067	2.902	4.826	0.772	14.203	0.343	1.577	0.167	0.074	3.355
0.403	4.327	2.660	4.993	0.863	14.668	0.402	1.720	0.125	0.091	3.617
0.241	4.281	2.984	4.369	0.828	9.670	0.243	1.036	0.201	0.105	3.089
0.412	4.038	3.731	4.341	0.971	12.550	0.244	1.197	0.135	0.180	3.309
0.154	1.840	3.533	6.902	1.308	8.954	0.190	2.047	0.091	0.018	1.608
0.352	5.170	2.945	2.187	0.866	11.566	0.306	0.765	0.194	0.165	2.959
0.288	3.336	3.430	5.054	0.896	10.608	0.258	1.017	0.104	0.175	2.689
0.447	5.060	3.240	5.462	0.937	18.099	0.339	1.762	0.196	0.164	3.649
0.420	5.828	2.898	4.121	0.793	14.167	0.347	1.133	0.180	0.199	4.181
0.492	5.230	2.116	3.516	0.584	16.289	0.374	1.241	0.262	0.188	4.687
0.385	4.707	2.350	4.655	0.882	15.452	0.357	1.789	0.208	0.153	3.213
0.354	4.626	2.854	4.885	0.753	14.250	0.273	1.332	0.072	0.098	3.228
0.244	3.112	3.245	6.687	1.095	11.960	0.240	2.001	0.177	0.080	2.440
0.221	2.715	2.848	5.216	0.978	6.625	0.137	1.202	0.075	0.015	1.833
0.374	2.819	2.694	5.560	0.804	10.830	0.268	1.472	0.069	0.137	2.838
0.416	3.943	2.908	6.660	1.076	14.812	0.313	2.033	0.173	0.069	3.716
0.356	3.874	2.739	4.778	0.894	11.158	0.215	1.099	0.149	0.093	3.510
0.410	4.898	2.362	3.565	0.630	11.763	0.342	0.783	0.119	0.169	4.037
0.246	2.761	2.914	4.860	0.799	5.649	0.168	1.192	0.016	0.069	2.180

Table 4: Concentraciones de aminoácidos en 10 nuevos zumos (Ejercicio 12)

Día	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}
1	0.275	3.693	2.785	6.812	1.086	12.206	0.262	2.152	0.091	0.106	2.851
2	0.295	3.401	2.594	5.903	0.964	9.945	0.189	1.719	0.069	0.058	2.271
3	0.370	3.865	2.935	7.034	1.122	18.572	0.354	2.354	0.148	0.043	3.779
4	0.385	3.585	3.601	5.454	1.139	11.033	0.255	0.857	0.078	0.130	3.625
5	0.248	3.188	2.966	7.090	1.205	7.800	0.199	1.657	0.046	0.024	2.733
6	0.480	4.512	2.142	4.533	0.762	18.385	0.345	1.710	0.093	0.167	4.872
7	0.417	5.260	2.554	3.404	0.773	13.679	0.277	0.908	0.122	0.161	3.734
8	0.327	4.388	3.110	4.396	0.774	9.041	0.213	0.669	0.129	0.141	3.725
9	0.251	3.125	2.589	6.390	1.106	13.410	0.235	1.898	0.107	0.044	2.864
10	0.422	4.810	2.002	3.322	1.144	15.986	0.348	1.147	0.154	0.178	3.511

14. (Problema 4.5) En la Tabla 5 se recogen las siguientes variables medidas sobre 30 olmos hembra. Este conjunto de datos pertenece a un estudio realizado por el Departamento de Industria Primaria y Pesca de Tasmania (Australia) en 1994. Los datos completos están disponibles en Nash *et al.* (1994).

	nombre	unidades	breve descripción
X_1	Longitud	mm	mayor medida de la corteza
X_2	Diámetro	mm	perpendicular a la longitud
X_3	Altura	mm	con madera dentro de la corteza
X_4	Peso total	g	todo el olmo
X_5	Peso desvainado	g	peso de la madera
X_6	Peso de las vísceras	g	peso de la tripa (después de sangrar)
X_7	Peso de la corteza	g	después de ser secado

Realizar un análisis de componentes principales e interpretar las dos primeras componentes.

Table 5: Datos para el Ejercicio 14

X1	X2	X3	X4	X5	X6	X7
0.53	0.42	0.135	0.677	0.2565	0.1415	0.21
0.53	0.415	0.15	0.7775	0.237	0.1415	0.33
0.545	0.425	0.125	0.768	0.294	0.1495	0.26
0.55	0.44	0.15	0.8945	0.3145	0.151	0.32
0.525	0.38	0.14	0.6065	0.194	0.1475	0.21
0.535	0.405	0.145	0.6845	0.2725	0.171	0.205
0.47	0.355	0.1	0.4755	0.1675	0.0805	0.185
0.44	0.34	0.1	0.451	0.188	0.087	0.13
0.565	0.44	0.155	0.9395	0.4275	0.214	0.27
0.55	0.415	0.135	0.7635	0.318	0.21	0.2
0.615	0.48	0.165	1.1615	0.513	0.301	0.305
0.56	0.44	0.14	0.9285	0.3825	0.188	0.3
0.58	0.45	0.185	0.9955	0.3945	0.272	0.285
0.68	0.56	0.165	1.639	0.6055	0.2805	0.46
0.68	0.55	0.175	1.798	0.815	0.3925	0.455
0.705	0.55	0.2	1.7095	0.633	0.4115	0.49
0.54	0.475	0.155	1.217	0.5305	0.3075	0.34
0.45	0.355	0.105	0.5225	0.237	0.1165	0.145
0.575	0.445	0.135	0.883	0.381	0.2035	0.26
0.45	0.335	0.105	0.425	0.1865	0.091	0.115
0.55	0.425	0.135	0.8515	0.362	0.196	0.27
0.46	0.375	0.12	0.4605	0.1775	0.11	0.15
0.525	0.425	0.16	0.8355	0.3545	0.2135	0.245
0.47	0.36	0.12	0.4775	0.2105	0.1055	0.15
0.5	0.4	0.14	0.6615	0.2565	0.1755	0.22
0.505	0.4	0.125	0.583	0.246	0.13	0.175
0.53	0.41	0.13	0.6965	0.302	0.1935	0.2
0.565	0.44	0.16	0.915	0.354	0.1935	0.32
0.595	0.495	0.185	1.285	0.416	0.224	0.485
0.475	0.39	0.12	0.5305	0.2135	0.1155	0.17

15. Considerar dos variables aleatorias con media cero y matriz de covarianzas

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma^2 & 1 \\ 1 & \sigma^2 \end{pmatrix}$$

con $\sigma > 0$. Se pide:

(a) Calcular los autovalores de $\mathbf{\Sigma}$. ¿Para que valores de σ es la matriz $\mathbf{\Sigma}$ definida positiva?

(b) Encontrar las componentes principales a partir de $\mathbf{\Sigma}$.

(c) Calcular la proporción de variabilidad explicada por la primera componente principal.

16. (Problema 4.15) Determinar la edad de un árbol contando el número de anillos de una sección del tronco a través del microscopio es un trabajo muy laborioso.

Por ello se busca la forma de predecir la edad de un árbol utilizando otras medidas más sencillas de obtener. La Tabla 6 contiene ocho variables medidas sobre 151 olmos. Las variables $X_1, \ldots, X_7$ son las mismas que las descritas en el Ejercicio 14. La variable y es el número de anillos del olmo. Obtener un modelo de regresión que permita predecir la edad de un olmo en función del resto de variables.

Table 6: Datos del Ejercicio 16.

X_1	X_2	X_3	X_4	X_5	X_6	X_7	y	X_1	X_2	X_3	X_4	X_5	X_6	X_7	y
0.455	0.365	0.095	0.514	0.2245	0.101	0.15	15	0.595	0.475	0.14	0.944	0.3625	0.189	0.315	9
0.35	0.265	0.09	0.2255	0.0995	0.0485	0.07	7	0.6	0.47	0.15	0.922	0.363	0.194	0.305	10
0.53	0.42	0.135	0.677	0.2565	0.1415	0.21	9	0.555	0.425	0.14	0.788	0.282	0.1595	0.285	11
0.44	0.365	0.125	0.516	0.2155	0.114	0.155	10	0.615	0.475	0.17	1.1025	0.4695	0.2355	0.345	14
0.33	0.255	0.08	0.205	0.0895	0.0395	0.055	7	0.575	0.445	0.14	0.941	0.3845	0.252	0.285	9
0.425	0.3	0.095	0.3515	0.141	0.0775	0.12	8	0.62	0.51	0.175	1.615	0.5105	0.192	0.675	12
0.53	0.415	0.15	0.7775	0.237	0.1415	0.33	20	0.52	0.425	0.165	0.9885	0.396	0.225	0.32	16
0.545	0.425	0.125	0.768	0.294	0.1495	0.26	16	0.595	0.475	0.16	1.3175	0.408	0.234	0.58	21
0.475	0.37	0.125	0.5095	0.2165	0.1125	0.165	9	0.58	0.45	0.14	1.013	0.38	0.216	0.36	14
0.55	0.44	0.15	0.8945	0.3145	0.151	0.32	19	0.57	0.465	0.18	1.295	0.339	0.2225	0.44	12
0.525	0.38	0.14	0.6065	0.194	0.1475	0.21	14	0.625	0.465	0.14	1.195	0.4825	0.205	0.4 13	
0.43	0.35	0.11	0.406	0.1675	0.081	0.135	10	0.56	0.44	0.16	0.8645	0.3305	0.2075	0.26	10
0.49	0.38	0.135	0.5415	0.2175	0.095	0.19	11	0.46	0.355	0.13	0.517	0.2205	0.114	0.165	9
0.535	0.405	0.145	0.6845	0.2725	0.171	0.205	10	0.575	0.45	0.16	0.9775	0.3135	0.231	0.33	12
0.47	0.355	0.1	0.4755	0.1675	0.0805	0.185	10	0.565	0.425	0.135	0.8115	0.341	0.1675	0.255	15
0.5	0.4	0.13	0.6645	0.258	0.133	0.24	12	0.555	0.44	0.15	0.755	0.307	0.1525	0.26	12
0.355	0.28	0.085	0.2905	0.095	0.0395	0.115	7	0.595	0.465	0.175	1.115	0.4015	0.254	0.39	13
0.44	0.34	0.1	0.451	0.188	0.087	0.13	10	0.625	0.495	0.165	1.262	0.507	0.318	0.39	10
0.365	0.295	0.08	0.2555	0.097	0.043	0.1	7	0.695	0.56	0.19	1.494	0.588	0.3425	0.485	15
0.45	0.32	0.1	0.381	0.1705	0.075	0.115	9	0.665	0.535	0.195	1.606	0.5755	0.388	0.48	14
0.355	0.28	0.095	0.2455	0.0955	0.062	0.075	11	0.535	0.435	0.15	0.725	0.269	0.1385	0.25	9
0.38	0.275	0.1	0.2255	0.08	0.049	0.085	10	0.47	0.375	0.13	0.523	0.214	0.132	0.145	8
0.565	0.44	0.155	0.9395	0.4275	0.214	0.27	12	0.47	0.37	0.13	0.5225	0.201	0.133	0.165	7
0.55	0.415	0.135	0.7635	0.318	0.21	0.2	9	0.475	0.375	0.125	0.5785	0.2775	0.085	0.155	10
0.615	0.48	0.165	1.1615	0.513	0.301	0.305	10	0.36	0.265	0.095	0.2315	0.105	0.046	0.075	7
0.56	0.44	0.14	0.9285	0.3825	0.188	0.3	11	0.55	0.435	0.145	0.843	0.328	0.1915	0.255	15
0.58	0.45	0.185	0.9955	0.3945	0.272	0.285	11	0.53	0.435	0.16	0.883	0.316	0.164	0.335	15
0.59	0.445	0.14	0.931	0.356	0.234	0.28	12	0.53	0.415	0.14	0.724	0.3105	0.1675	0.205	10
0.605	0.475	0.18	0.9365	0.394	0.219	0.295	15	0.605	0.47	0.16	1.1735	0.4975	0.2405	0.345	12
0.575	0.425	0.14	0.8635	0.393	0.227	0.2	11	0.52	0.41	0.155	0.727	0.291	0.1835	0.235	12
0.58	0.47	0.165	0.9975	0.3935	0.242	0.33	10	0.545	0.43	0.165	0.802	0.2935	0.183	0.28	11
0.68	0.56	0.165	1.639	0.6055	0.2805	0.46	15	0.5	0.4	0.125	0.6675	0.261	0.1315	0.22	10
0.665	0.525	0.165	1.338	0.5515	0.3575	0.35	18	0.51	0.39	0.135	0.6335	0.231	0.179	0.2	9
0.68	0.55	0.175	1.798	0.815	0.3925	0.455	19	0.435	0.395	0.105	0.3635	0.136	0.098	0.13	9
0.705	0.55	0.2	1.7095	0.633	0.4115	0.49	13	0.495	0.395	0.125	0.5415	0.2375	0.1345	0.155	9
0.465	0.355	0.105	0.4795	0.227	0.124	0.125	8	0.465	0.36	0.105	0.431	0.172	0.107	0.175	9
0.54	0.475	0.155	1.217	0.5305	0.3075	0.34	16	0.435	0.32	0.08	0.3325	0.1485	0.0635	0.105	9
0.45	0.355	0.105	0.5225	0.237	0.1165	0.145	8	0.425	0.35	0.105	0.393	0.13	0.063	0.165	9
0.575	0.445	0.135	0.883	0.381	0.2035	0.26	11	0.545	0.41	0.125	0.6935	0.2975	0.146	0.21	11
0.355	0.29	0.09	0.3275	0.134	0.086	0.09	9	0.53	0.415	0.115	0.5915	0.233	0.1585	0.18	11
0.45	0.335	0.105	0.425	0.1865	0.091	0.115	9	0.49	0.375	0.135	0.6125	0.2555	0.102	0.22	11
0.55	0.425	0.135	0.8515	0.362	0.196	0.27	14	0.44	0.34	0.105	0.402	0.1305	0.0955	0.165	10
0.24	0.175	0.045	0.07	0.0315	0.0235	0.02	5	0.56	0.43	0.15	0.8825	0.3465	0.172	0.31	9
0.205	0.15	0.055	0.042	0.0255	0.015	0.012	5	0.405	0.305	0.085	0.2605	0.1145	0.0595	0.085	8
0.21	0.15	0.05	0.042	0.0175	0.0125	0.015	4	0.47	0.365	0.105	0.4205	0.163	0.1035	0.14	9
0.39	0.295	0.095	0.203	0.0875	0.045	0.075	7	0.385	0.295	0.085	0.2535	0.103	0.0575	0.085	7
0.47	0.37	0.12	0.5795	0.293	0.227	0.14	9	0.515	0.425	0.14	0.766	0.304	0.1725	0.255	14
0.46	0.375	0.12	0.4605	0.1775	0.11	0.15	7	0.37	0.265	0.075	0.214	0.09	0.051	0.07	6
0.325	0.245	0.07	0.161	0.0755	0.0255	0.045	6	0.36	0.28	0.08	0.1755	0.081	0.0505	0.07	6
0.525	0.425	0.16	0.8355	0.3545	0.2135	0.245	9	0.27	0.195	0.06	0.073	0.0285	0.0235	0.03	5
0.52	0.41	0.12	0.595	0.2385	0.111	0.19	8	0.375	0.275	0.09	0.238	0.1075	0.0545	0.07	6
0.4	0.32	0.095	0.303	0.1335	0.06	0.1	7	0.385	0.29	0.085	0.2505	0.112	0.061	0.08	8
0.485	0.36	0.13	0.5415	0.2595	0.096	0.16	10	0.7	0.535	0.16	1.7255	0.63	0.2635	0.54	19
0.47	0.36	0.12	0.4775	0.2105	0.1055	0.15	10	0.71	0.54	0.165	1.959	0.7665	0.261	0.78	18
0.405	0.31	0.1	0.385	0.173	0.0915	0.11	7	0.595	0.48	0.165	1.262	0.4835	0.283	0.41	17
0.5	0.4	0.14	0.6615	0.2565	0.1755	0.22	8	0.44	0.35	0.125	0.4035	0.175	0.063	0.129	9
0.445	0.35	0.12	0.4425	0.192	0.0955	0.135	8	0.325	0.26	0.09	0.1915	0.085	0.036	0.062	7
0.47	0.385	0.135	0.5895	0.2765	0.12	0.17	8	0.35	0.26	0.095	0.211	0.086	0.056	0.068	7
0.245	0.19	0.06	0.086	0.042	0.014	0.025	4	0.265	0.2	0.065	0.0975	0.04	0.0205	0.028	7
0.505	0.4	0.125	0.583	0.246	0.13	0.175	7	0.425	0.33	0.115	0.406	0.1635	0.081	0.1355	8
0.45	0.345	0.105	0.4115	0.18	0.1125	0.135	7	0.305	0.23	0.08	0.156	0.0675	0.0345	0.048	7
0.505	0.405	0.11	0.625	0.305	0.16	0.175	9	0.345	0.255	0.09	0.2005	0.094	0.0295	0.063	9
0.53	0.41	0.13	0.6965	0.302	0.1935	0.2	10	0.405	0.325	0.11	0.3555	0.151	0.063	0.117	9
0.425	0.325	0.095	0.3785	0.1705	0.08	0.1	7	0.375	0.285	0.095	0.253	0.096	0.0575	0.0925	9
0.52	0.4	0.12	0.58	0.234	0.1315	0.185	8	0.565	0.445	0.155	0.826	0.341	0.2055	0.2475	10
0.475	0.355	0.12	0.48	0.234	0.1015	0.135	8	0.55	0.45	0.145	0.741	0.295	0.1435	0.2665	10
0.565	0.44	0.16	0.915	0.354	0.1935	0.32	12	0.65	0.52	0.19	1.3445	0.519	0.306	0.4465	16
0.595	0.495	0.185	1.285	0.416	0.224	0.485	13	0.56	0.455	0.155	0.797	0.34	0.19	0.2425	11
0.475	0.39	0.12	0.5305	0.2135	0.1155	0.17	10	0.475	0.375	0.13	0.5175	0.2075	0.1165	0.17	10
0.31	0.235	0.07	0.151	0.063	0.0405	0.045	6	0.49	0.38	0.125	0.549	0.245	0.1075	0.174	10
0.555	0.425	0.13	0.7665	0.264	0.168	0.275	13	0.46	0.35	0.12	0.515	0.224	0.108	0.1565	10
0.4	0.32	0.11	0.353	0.1405	0.0985	0.1	8	0.28	0.205	0.08	0.127	0.052	0.039	0.042	9
0.595	0.475	0.17	1.247	0.48	0.225	0.425	20	0.175	0.13	0.055	0.0315	0.0105	0.0065	0.0125	5
0.57	0.48	0.175	1.185	0.474	0.261	0.38	11	0.17	0.13	0.095	0.03	0.013	0.008	0.01	4
0.605	0.45	0.195	1.098	0.481	0.2895	0.315	13	0.59	0.475	0.145	1.053	0.4415	0.262	0.325	15
0.6	0.475	0.15	1.0075	0.4425	0.221	0.28	15								

17. (Problema 8.4) Las Tablas 7, 8 y 9 contienen varias variables medidas sobre 250 olmos, divididos en 3 grupos, según su sexo (*Grupo 1*: 100 olmos femeninos, *Grupo 2*: 100 olmos masculinos, *Grupo 3*: 50 olmos juveniles o plántulas). Véase el Ejercicio 14 para una descripción completa de las variables.

(a) Realizar la representación canónica de los tres grupos, especificando los porcentajes de variabilidad explicados por cada eje canónico.

(b) Suponiendo normalidad multivariante, construir las regiones de confianza (al 95%) para los individuos medios de cada grupo.

(c) Interpretar los ejes canónicos.

Table 7: Datos para el Ejercicio 17. Grupo 1: olmos femeninos.

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_1	X_2	X_3	X_4	X_5	X_6	X_7
0.53	0.42	0.135	0.677	0.2565	0.1415	0.21	0.53	0.415	0.115	0.5915	0.233	0.1585	0.18
0.53	0.415	0.15	0.7775	0.237	0.1415	0.33	0.49	0.375	0.135	0.6125	0.2555	0.102	0.22
0.545	0.425	0.125	0.768	0.294	0.1495	0.26	0.56	0.43	0.15	0.8825	0.3465	0.172	0.31
0.55	0.44	0.15	0.8945	0.3145	0.151	0.32	0.47	0.365	0.105	0.4205	0.163	0.1035	0.14
0.525	0.38	0.14	0.6065	0.194	0.1475	0.21	0.515	0.425	0.14	0.766	0.304	0.1725	0.255
0.535	0.405	0.145	0.6845	0.2725	0.171	0.205	0.44	0.35	0.125	0.4035	0.175	0.063	0.129
0.47	0.355	0.1	0.4755	0.1675	0.0805	0.185	0.325	0.26	0.09	0.1915	0.085	0.036	0.062
0.44	0.34	0.1	0.451	0.188	0.087	0.13	0.425	0.33	0.115	0.406	0.1635	0.081	0.1355
0.565	0.44	0.155	0.9395	0.4275	0.214	0.27	0.305	0.23	0.08	0.156	0.0675	0.0345	0.048
0.55	0.415	0.135	0.7635	0.318	0.21	0.2	0.405	0.325	0.11	0.3555	0.151	0.063	0.117
0.615	0.48	0.165	1.1615	0.513	0.301	0.305	0.565	0.445	0.155	0.826	0.341	0.2055	0.2475
0.56	0.44	0.14	0.9285	0.3825	0.188	0.3	0.55	0.45	0.145	0.741	0.295	0.1435	0.2665
0.58	0.45	0.185	0.9955	0.3945	0.272	0.285	0.49	0.38	0.125	0.549	0.245	0.1075	0.174
0.68	0.56	0.165	1.639	0.6055	0.2805	0.46	0.605	0.5	0.185	1.1185	0.469	0.2585	0.335
0.68	0.55	0.175	1.798	0.815	0.3925	0.455	0.635	0.515	0.19	1.3715	0.5065	0.305	0.45
0.705	0.55	0.2	1.7095	0.633	0.4115	0.49	0.605	0.485	0.16	1.0565	0.37	0.2355	0.355
0.54	0.475	0.155	1.217	0.5305	0.3075	0.34	0.565	0.45	0.135	0.9885	0.387	0.1495	0.31
0.45	0.355	0.105	0.5225	0.237	0.1165	0.145	0.575	0.46	0.19	0.994	0.392	0.2425	0.34
0.575	0.445	0.135	0.883	0.381	0.2035	0.26	0.58	0.455	0.17	0.9075	0.374	0.2135	0.285
0.45	0.335	0.105	0.425	0.1865	0.091	0.115	0.575	0.46	0.165	1.124	0.2985	0.1785	0.44
0.55	0.425	0.135	0.8515	0.362	0.196	0.27	0.605	0.485	0.16	1.222	0.53	0.2575	0.28
0.46	0.375	0.12	0.4605	0.1775	0.11	0.15	0.725	0.56	0.21	2.141	0.65	0.398	1.005
0.525	0.425	0.16	0.8355	0.3545	0.2135	0.245	0.65	0.545	0.23	1.752	0.5605	0.2895	0.815
0.47	0.36	0.12	0.4775	0.2105	0.1055	0.15	0.725	0.575	0.175	2.124	0.765	0.4515	0.85
0.5	0.4	0.14	0.6615	0.2565	0.1755	0.22	0.68	0.57	0.205	1.842	0.625	0.408	0.65
0.505	0.4	0.125	0.583	0.246	0.13	0.175	0.68	0.515	0.175	1.6185	0.5125	0.409	0.62
0.53	0.41	0.13	0.6965	0.302	0.1935	0.2	0.53	0.395	0.145	0.775	0.308	0.169	0.255
0.565	0.44	0.16	0.915	0.354	0.1935	0.32	0.52	0.405	0.115	0.776	0.32	0.1845	0.22
0.595	0.495	0.185	1.285	0.416	0.224	0.485	0.56	0.45	0.16	1.0235	0.429	0.268	0.3
0.475	0.39	0.12	0.5305	0.2135	0.1155	0.17	0.62	0.475	0.175	1.0165	0.4355	0.214	0.325
0.4	0.32	0.11	0.353	0.1405	0.0985	0.1	0.645	0.51	0.2	1.5675	0.621	0.367	0.46
0.595	0.475	0.17	1.247	0.48	0.225	0.425	0.63	0.48	0.15	1.0525	0.392	0.336	0.285
0.605	0.45	0.195	1.098	0.481	0.2895	0.315	0.63	0.5	0.185	1.383	0.54	0.3315	0.38
0.6	0.475	0.15	1.0075	0.4425	0.221	0.28	0.63	0.48	0.16	1.199	0.5265	0.335	0.315
0.6	0.47	0.15	0.922	0.363	0.194	0.305	0.585	0.46	0.17	0.9325	0.365	0.271	0.29
0.555	0.425	0.14	0.788	0.282	0.1595	0.285	0.51	0.4	0.14	0.8145	0.459	0.1965	0.195
0.615	0.475	0.17	1.1025	0.4695	0.2355	0.345	0.505	0.41	0.15	0.644	0.285	0.145	0.21
0.575	0.445	0.14	0.941	0.3845	0.252	0.285	0.45	0.345	0.12	0.4165	0.1655	0.095	0.135
0.52	0.425	0.165	0.9885	0.396	0.225	0.32	0.5	0.4	0.145	0.63	0.234	0.1465	0.23
0.57	0.465	0.18	1.295	0.339	0.2225	0.44	0.53	0.435	0.17	0.8155	0.2985	0.155	0.275
0.46	0.355	0.13	0.517	0.2205	0.114	0.165	0.44	0.34	0.14	0.482	0.186	0.1085	0.16
0.575	0.45	0.16	0.9775	0.3135	0.231	0.33	0.525	0.415	0.17	0.8325	0.2755	0.1685	0.31
0.625	0.495	0.165	1.262	0.507	0.318	0.39	0.49	0.365	0.145	0.6345	0.1995	0.1625	0.22
0.475	0.375	0.125	0.5785	0.2775	0.085	0.155	0.415	0.325	0.105	0.38	0.1595	0.0785	0.12
0.52	0.41	0.155	0.727	0.291	0.1835	0.235	0.485	0.395	0.16	0.66	0.2475	0.128	0.235
0.545	0.43	0.165	0.802	0.2935	0.183	0.28	0.415	0.305	0.13	0.32	0.1305	0.0755	0.105
0.5	0.4	0.125	0.6675	0.261	0.1315	0.22	0.445	0.325	0.125	0.455	0.1785	0.1125	0.14
0.51	0.39	0.135	0.6335	0.231	0.179	0.2	0.47	0.35	0.145	0.5175	0.187	0.1235	0.18
0.435	0.395	0.105	0.3635	0.136	0.098	0.13	0.49	0.375	0.15	0.5755	0.22	0.144	0.19
0.545	0.41	0.125	0.6935	0.2975	0.146	0.21	0.445	0.355	0.15	0.485	0.181	0.125	0.155

18. (Problema 8.5) La Tabla 10 contiene once variables medidas sobre un total de 44 individuos pertenecientes a cuatro especies de cocodrilos: 1. *Alligator mississippiensis*, 2. *Crocodylus niloticus*, 3. *Crocodylus porosus*, 4. *Osteolaemus tetraspis*. La Figura 1 muestra las regiones geográficas donde se encuentran estas especies de cocodrilos. Las variables medidas sobre cada individuo son: X_1 =longitud del cráneo, X_2 =ancho del cráneo, X_3 =ancho del hocico, X_4 =longitud del hocico, X_5 =longitud dorsal del cráneo, X_6 =ancho máximo orbital, X_7 =ancho mínimo inter-orbital, X_8 =longitud máxima orbital, X_9 =longitud del paladar post-orbital, X_{10} =ancho posterior del paladar, X_{11} =ancho máximo entre orificios nasales (Fuente: Iordansky 1973).

    Realizar la representación canónica de las cuatro especies, especificando los porcentajes de variabilidad explicados por cada eje canónico. Suponiendo normalidad multivariante, construir las regiones de confianza (al 90%) para los individuos medios de cada grupo.

Table 8: Datos para el Ejercicio 17. Grupo 2: olmos masculinos

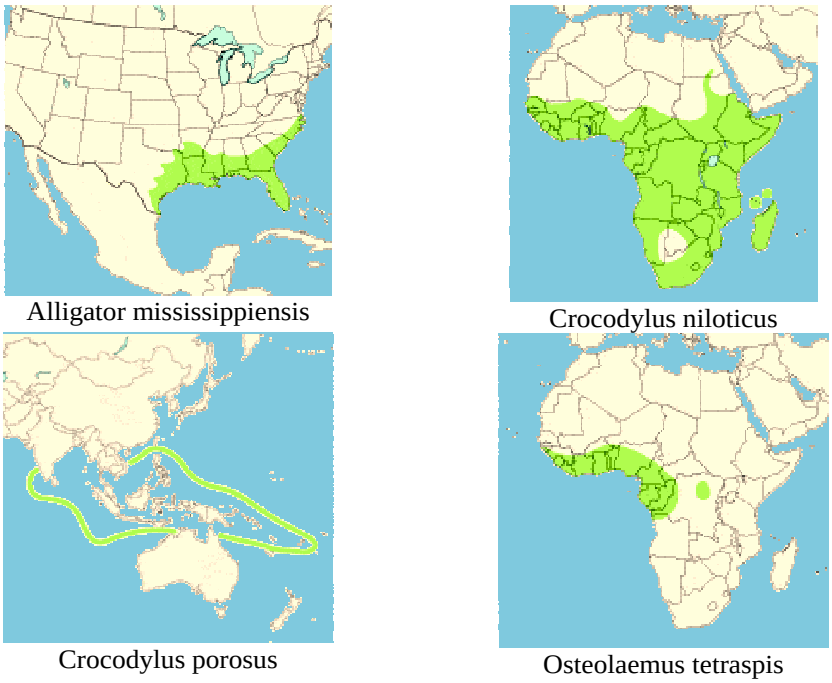
X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_1	X_2	X_3	X_4	X_5	X_6	X_7
0.665	0.525	0.165	1.338	0.5515	0.3575	0.35	0.515	0.405	0.13	0.722	0.32	0.131	0.21
0.465	0.355	0.105	0.4795	0.227	0.124	0.125	0.645	0.485	0.215	1.514	0.546	0.2615	0.635
0.355	0.29	0.09	0.3275	0.134	0.086	0.09	0.605	0.465	0.165	1.056	0.4215	0.2475	0.34
0.47	0.37	0.12	0.5795	0.293	0.227	0.14	0.61	0.485	0.175	1.2445	0.544	0.297	0.345
0.4	0.32	0.095	0.303	0.1335	0.06	0.1	0.725	0.57	0.19	2.55	1.0705	0.483	0.725
0.485	0.36	0.13	0.5415	0.2595	0.096	0.16	0.705	0.56	0.22	1.981	0.8175	0.3085	0.76
0.405	0.31	0.1	0.385	0.173	0.0915	0.11	0.695	0.55	0.215	1.9565	0.7125	0.541	0.59
0.445	0.35	0.12	0.4425	0.192	0.0955	0.135	0.525	0.435	0.155	1.065	0.486	0.233	0.285
0.47	0.385	0.135	0.5895	0.2765	0.12	0.17	0.58	0.475	0.15	0.97	0.385	0.2165	0.35
0.45	0.345	0.105	0.4115	0.18	0.1125	0.135	0.57	0.48	0.18	0.9395	0.399	0.2	0.295
0.505	0.405	0.11	0.625	0.305	0.16	0.175	0.64	0.51	0.175	1.368	0.515	0.266	0.57
0.425	0.325	0.095	0.3785	0.1705	0.08	0.1	0.62	0.49	0.19	1.218	0.5455	0.2965	0.355
0.52	0.4	0.12	0.58	0.234	0.1315	0.185	0.615	0.48	0.18	1.1595	0.4845	0.2165	0.325
0.475	0.355	0.12	0.48	0.234	0.1015	0.135	0.61	0.485	0.17	1.0225	0.419	0.2405	0.36
0.555	0.425	0.13	0.7665	0.264	0.168	0.275	0.58	0.45	0.15	0.927	0.276	0.1815	0.36
0.57	0.48	0.175	1.185	0.474	0.261	0.38	0.5	0.405	0.155	0.772	0.346	0.1535	0.245
0.595	0.475	0.14	0.944	0.3625	0.189	0.315	0.64	0.5	0.185	1.3035	0.4445	0.2635	0.465
0.62	0.51	0.175	1.615	0.5105	0.192	0.675	0.56	0.45	0.16	0.922	0.432	0.178	0.26
0.595	0.475	0.16	1.3175	0.408	0.234	0.58	0.585	0.46	0.185	0.922	0.3635	0.213	0.285
0.58	0.45	0.14	1.013	0.38	0.216	0.36	0.5	0.4	0.165	0.825	0.254	0.205	0.285
0.625	0.465	0.14	1.195	0.4825	0.205	0.4	0.42	0.335	0.115	0.369	0.171	0.071	0.12
0.56	0.44	0.16	0.8645	0.3305	0.2075	0.26	0.335	0.25	0.09	0.181	0.0755	0.0415	0.06
0.565	0.425	0.135	0.8115	0.341	0.1675	0.255	0.5	0.405	0.14	0.6155	0.241	0.1355	0.205
0.555	0.44	0.15	0.755	0.307	0.1525	0.26	0.55	0.405	0.14	0.8025	0.244	0.1635	0.255
0.595	0.465	0.175	1.115	0.4015	0.254	0.39	0.45	0.35	0.13	0.46	0.174	0.111	0.135
0.695	0.56	0.19	1.494	0.588	0.3425	0.485	0.47	0.36	0.135	0.501	0.1665	0.115	0.165
0.665	0.535	0.195	1.606	0.5755	0.388	0.48	0.555	0.445	0.135	0.836	0.336	0.1625	0.275
0.535	0.435	0.15	0.725	0.269	0.1385	0.25	0.565	0.44	0.175	0.9025	0.31	0.193	0.325
0.47	0.375	0.13	0.523	0.214	0.132	0.145	0.625	0.505	0.215	1.4455	0.496	0.287	0.435
0.47	0.37	0.13	0.5225	0.201	0.133	0.165	0.565	0.425	0.16	0.9425	0.3495	0.2185	0.275
0.55	0.435	0.145	0.843	0.328	0.1915	0.255	0.59	0.47	0.18	1.1235	0.4205	0.2805	0.36
0.53	0.435	0.16	0.883	0.316	0.164	0.335	0.6	0.495	0.165	1.2415	0.485	0.2775	0.34
0.53	0.415	0.14	0.724	0.3105	0.1675	0.205	0.56	0.45	0.175	1.011	0.3835	0.2065	0.37
0.605	0.47	0.16	1.1735	0.4975	0.2405	0.345	0.56	0.45	0.185	1.07	0.3805	0.175	0.41
0.495	0.395	0.125	0.5415	0.2375	0.1345	0.155	0.545	0.46	0.16	0.8975	0.341	0.1655	0.345
0.465	0.36	0.105	0.431	0.172	0.107	0.175	0.53	0.42	0.165	0.8945	0.319	0.239	0.245
0.425	0.35	0.105	0.393	0.13	0.063	0.165	0.27	0.2	0.08	0.1205	0.0465	0.028	0.04
0.44	0.34	0.105	0.402	0.1305	0.0955	0.165	0.52	0.45	0.15	0.895	0.3615	0.186	0.235
0.405	0.305	0.085	0.2605	0.1145	0.0595	0.085	0.35	0.275	0.11	0.2925	0.1225	0.0635	0.0905
0.37	0.265	0.075	0.214	0.09	0.051	0.07	0.47	0.39	0.15	0.6355	0.2185	0.0885	0.255
0.7	0.535	0.16	1.7255	0.63	0.2635	0.54	0.59	0.5	0.2	1.187	0.412	0.2705	0.37
0.71	0.54	0.165	1.959	0.7665	0.261	0.78	0.62	0.485	0.205	1.219	0.3875	0.2505	0.385
0.595	0.48	0.165	1.262	0.4835	0.283	0.41	0.63	0.505	0.225	1.525	0.56	0.3335	0.45
0.345	0.255	0.09	0.2005	0.094	0.0295	0.063	0.63	0.515	0.155	1.259	0.4105	0.197	0.41
0.375	0.285	0.095	0.253	0.096	0.0575	0.0925	0.655	0.54	0.215	1.844	0.7425	0.327	0.585
0.65	0.52	0.19	1.3445	0.519	0.306	0.4465	0.61	0.5	0.24	1.642	0.532	0.3345	0.69
0.56	0.455	0.155	0.797	0.34	0.19	0.2425	0.635	0.525	0.205	1.484	0.55	0.3115	0.43
0.475	0.375	0.13	0.5175	0.2075	0.1165	0.17	0.485	0.395	0.14	0.6295	0.2285	0.127	0.225
0.46	0.35	0.12	0.515	0.224	0.108	0.1565	0.515	0.38	0.175	0.9565	0.325	0.158	0.31
0.59	0.475	0.145	1.053	0.4415	0.262	0.325	0.53	0.435	0.155	0.699	0.288	0.1595	0.205

Table 9: Datos para el Ejercicio 17. Grupo 3: olmos juveniles o plántulas.

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_1	X_2	X_3	X_4	X_5	X_6	X_7
0.28	0.205	0.08	0.127	0.052	0.039	0.042	0.33	0.255	0.085	0.1655	0.063	0.039	0.06
0.175	0.13	0.055	0.0315	0.0105	0.0065	0.0125	0.35	0.26	0.085	0.174	0.0705	0.0345	0.06
0.17	0.13	0.095	0.03	0.013	0.008	0.01	0.32	0.245	0.08	0.1585	0.0635	0.0325	0.05
0.235	0.16	0.04	0.048	0.0185	0.018	0.015	0.36	0.275	0.085	0.1975	0.0745	0.0415	0.07
0.36	0.26	0.09	0.1785	0.0645	0.037	0.075	0.305	0.245	0.075	0.156	0.0675	0.038	0.045
0.315	0.21	0.06	0.125	0.06	0.0375	0.035	0.345	0.27	0.11	0.2135	0.082	0.0545	0.07
0.315	0.245	0.085	0.1435	0.053	0.0475	0.05	0.33	0.25	0.105	0.1715	0.0655	0.035	0.06
0.225	0.16	0.045	0.0465	0.025	0.015	0.015	0.245	0.195	0.06	0.095	0.0445	0.0245	0.026
0.355	0.275	0.085	0.22	0.092	0.06	0.15	0.36	0.285	0.105	0.2415	0.0915	0.057	0.075
0.4	0.3	0.11	0.315	0.109	0.067	0.12	0.295	0.215	0.085	0.128	0.049	0.034	0.04
0.435	0.34	0.11	0.3795	0.1495	0.085	0.12	0.275	0.205	0.075	0.1105	0.045	0.0285	0.035
0.37	0.28	0.095	0.2655	0.122	0.052	0.08	0.28	0.21	0.085	0.1065	0.039	0.0295	0.03
0.405	0.3	0.12	0.324	0.1265	0.07	0.11	0.2	0.145	0.06	0.037	0.0125	0.0095	0.011
0.425	0.38	0.105	0.3265	0.1285	0.0785	0.1	0.165	0.12	0.03	0.0215	0.007	0.005	0.005
0.365	0.27	0.085	0.205	0.078	0.0485	0.07	0.45	0.355	0.11	0.4585	0.194	0.067	0.14
0.275	0.215	0.075	0.1155	0.0485	0.029	0.035	0.33	0.255	0.095	0.172	0.066	0.0255	0.06
0.44	0.35	0.135	0.435	0.1815	0.083	0.125	0.265	0.21	0.06	0.0965	0.0425	0.022	0.03
0.295	0.225	0.08	0.124	0.0485	0.032	0.04	0.19	0.145	0.04	0.038	0.0165	0.0065	0.015
0.075	0.055	0.01	0.002	0.001	0.0005	0.0015	0.265	0.205	0.07	0.1055	0.039	0.041	0.035
0.13	0.1	0.03	0.013	0.0045	0.003	0.004	0.355	0.275	0.09	0.251	0.097	0.053	0.08
0.11	0.09	0.03	0.008	0.0025	0.002	0.003	0.32	0.255	0.1	0.1755	0.073	0.0415	0.065
0.16	0.12	0.035	0.021	0.0075	0.0045	0.005	0.36	0.28	0.09	0.2255	0.0885	0.04	0.09
0.27	0.2	0.07	0.1	0.034	0.0245	0.035	0.3	0.22	0.08	0.121	0.0475	0.042	0.035
0.23	0.175	0.065	0.0645	0.026	0.0105	0.02	0.235	0.175	0.04	0.0705	0.0335	0.015	0.02
0.3	0.23	0.08	0.1275	0.0435	0.0265	0.04	0.34	0.26	0.08	0.2	0.08	0.0555	0.055

19. (Datos del Problema 4.4) La Tabla 11 contiene 11 indicadores económicos y sociales de 96 países del mundo. Las variables observadas son: X_1 = Tasa anual de crecimiento de la población, X_2 = Tasa de mortalidad infantil por cada 1000 nacidos vivos, X_3 = Porcentaje de mujeres en la población activa, X_4 = PNB en 1995 (en millones de dólares), X_5 = Producción de electricidad (en millones kW/h), X_6 = Líneas telefónicas por cada 1000 habitantes, X_7 = Consumo de

Figure 1: Hábitat de las cuatro especies de cocodrilos. (Ejercicio 18)



agua per cápita, X_8 = Proporción de la superficie del país cubierta por bosques, X_9 = Proporción de deforestación anual, X_{10} = Consumo de energía per cápita, X_{11} = Emisión de CO2 per cápita.

Calcular la matriz de distancias de Mahalanobis entre los 20 primeros países.

20. En muchas situaciones las variables que se observan sobre un conjunto de individuos son de naturaleza binaria. En estos casos para poder disponer de una matriz de distancias entre individuos se utilizan los coeficientes de similaridad.

    El coeficiente de similaridad entre el individuo i y el individuo j, s_{ij}, se calcula a partir de las frecuencias:
    $a =$"número de variables con respuesta 1 en ambos individuos",
    $b =$"número de variables con respuesta 0 en el primer individuo y con respuesta 1 en el segundo individuo",
    $c =$"número de variables con respuesta 1 en el primer individuo y con respuesta 0 en el segundo individuo",
    $d =$"número de variables con respuesta 0 en ambos individuos".

    Existen muchísimos coeficientes de similaridad (véase Cuadras 2004), pero los de Sokal-Michener y de Jaccard son especialmente interesantes porque dan lugar a una configuación euclídea. Se definen como:

$$\textit{Sokal y Michener:} \quad s_{ij} = \frac{a+d}{p}, \qquad \textit{Jaccard:} \quad s_{ij} = \frac{a}{a+b+c},$$

Table 10: Datos para el Ejercicio 18.

especie	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
1	72	40	37	35	71	17	5	20	15	25	11
1	220	112	98	138	216	30	16	46	36	64	31
1	225	150	89	140	220	32	17	52	37	82	30
1	272	138	120	175	262	24	25	54	44	78	38
1	288	148	126	180	275	40	22	58	42	82	40
1	290	150	117	183	270	40	20	54	46	82	40
1	292	150	127	166	284	49	26	56	48	86	39
1	320	150	124	203	310	40	25	62	46	80	38
1	354	178	137	240	337	42	25	69	50	89	51
1	366	186	160	232	348	39	32	68	54	98	53
1	380	236	210	238	358	52	27	63	63	120	64
2	160	64	46	100	153	20	9	22	30	39	9
2	198	94	70	121	186	25	13	31	32	48	13
2	248	243	76	159	235	30	16	41	42	105	15
2	254	114	71	158	235	28	16	40	42	65	15
2	420	235	170	270	400	37	42	60	68	105	42
2	440	250	170	280	420	42	50	65	70	120	48
2	525	290	220	360	495	45	48	72	82	145	54
2	582	336	218	382	554	48	58	72	76	105	57
2	610	345	268	400	564	46	90	85	76	164	56
3	76	30	22	41	73	13	4	17	16	20	4
3	548	74	56	364	513	23	10	29	26	44	48
3	238	292	68	154	230	29	12	36	30	55	48
3	408	200	148	274	390	38	36	57	54	110	32
3	548	300	210	364	513	46	55	68	65	150	48
3	565	292	216	405	550	45	64	70	90	160	48
3	672	384	302	452	620	50	70	90	85	185	64
3	800	416	324	516	740	63	82	100	105	204	75
4	164	90	70	90	160	36	16	42	32	57	20
4	188	107	71	92	160	29	13	38	35	65	18
4	170	98	72	98	165	31	14	42	35	60	20
4	173	107	70	100	165	33	12	40	35	60	22
4	175	102	73	102	165	32	14	42	38	64	24
4	185	105	77	105	175	32	14	44	40	61	22
4	185	105	78	105	175	33	16	40	40	61	22
4	188	107	82	108	180	33	16	40	40	65	24
4	188	104	80	110	178	34	15	44	40	64	24
4	190	108	80	112	180	32	16	45	38	65	24
4	194	110	82	114	182	34	15	44	38	67	24
4	194	117	92	117	180	34	18	43	42	70	23
4	203	108	88	116	193	35	16	46	40	69	26
4	210	107	91	124	178	36	19	48	40	65	26
4	225	128	105	128	215	40	20	52	45	75	28
4	240	136	91	133	222	38	19	51	46	76	27

donde p es el número de variables observadas. Aplicando uno de estos coeficientes a un conjunto de n individuos se obtiene una matriz de similaridades $\mathcal{S} = (s_{ij})_{n\times n}$.

Una forma de obtener una distancia a partir de un coeficiente de similaridad es la siguiente:

$$\mathbf{D}^{(2)} = 2\,(\mathbf{1}_n\,\mathbf{1}_n' - \mathcal{S}).$$

Se considera el siguiente conjunto de seis individuos formado por cinco animales, *león, girafa, vaca, oveja, gato doméstico*, junto con el *hombre*. Se miden seis variables binarias sobre estos individuos: X_1 =tiene cola, X_2 =es salvaje, X_3 =tiene el cuello largo, X_4 =es animal de granja, X_5 =es carnívoro, X_6 =camina sobre cuatro patas.

(a) Obtener la matriz de datos.

(b) Calcular los coeficentes de similaridad de Sokal-Michener y de Jaccard para cada par de individuos y obtener las matrices de distancias asociadas.

Table 11: Indicadores económicos y sociales sobre países del mundo

País	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
1. Albania	1	30	41	2199	3903	12	94	53	0	341	1.2
2. Angola	3	124	46	4422	955	6	57	19	0,7	89	0.5
3. Arabia Saudi	4.3	21	13	133540	91019	96	497	1	0	4566	13.1
4. Argelia	2.3	34	24	44609	19883	42	180	2	0.8	906	3
5. Argentina	1.3	22	31	278431	65962	160	1043	22	0.1	1504	3.5
6. Australia	1.4	6	43	337909	167155	510	933	19	0	5341	15.3
7. Austria	0,6	6	41	216547	53259	465	304	47	-0.4	3301	7.2
8. Bangladesh	2	79	42	28599	9891	2	220	6	4.1	64	0.5
9. Belgica	0.3	8	40	250710	72236	457	917	20	-0.3	5120	10.1
10. Benin	3	95	48	2034	2	5	26	45	1.3	20	0.1
11. Bielorrusia	0.4	13	49	21356	31397	190	295	31	-0.4	2392	9.9
12. Bolivia	2.3	69	37	5905	2824	35	201	45	1.2	213	1
13. Brasil	1.6	44	25	579157	260882	13	246	66	0.6	718	1.4
14. Bulgaria	-0.6	15	38	11225	38133	335	1544	33	-0.2	2438	6.4
15. Camerun	2.9	56	38	8615	2740	4	38	44	0.6	103	0.2
16. Canadá	1.3	6	45	573695	554227	590	1602	49	-1.4	7854	14.4
17. Colombia	1.8	26	37	70263	43354	100	174	52	0.7	632	1.8
18. Congo	3.1	90	43	1734	435	8	20	58	0.2	331	1.6
19. Corea del Norte	1.8	26	45	12870	38000	47	687	74	0	1129	11.2
20. Corea del Sur	0.9	10	40	435137	164993	415	632	66	0,1	2982	6.6
21. Costa de Marfil	3.4	86	33	9248	2305	8	66	34	1	103	0.5
22. Costa Rica	2.5	13	30	8884	4772	164	780	28	3	558	1.2
23. Cuba	0.9	9	38	7159	10982	32	870	16	1	923	2.6
24. Chile	1.6	12	32	59151	25276	132	1626	12	-0.1	1012	2.6
25. China	1.3	34	45	744890	928083	34	461	13	0,7	664	2.3
26. Dinamarca	0,2	6	46	156027	40097	613	233	12	0	3977	10.4
27. R. Dominicana	2	37	29	11390	6182	79	446	23	2.9	237	1.4
28. Ecuador	2.3	36	26	15997	8256	61	581	43	1.8	565	1.8
29. Egipto	2.2	56	29	45507	51947	46	956	0	0	600	1.2
30. El Salvador	1.8	36	34	9057	3211	53	245	6	2.3	370	0.7
31. E. Arabes Unidos	5.8	16	13	42806	18870	283	884	0	0	10531	33.9
32. Eslovaquia	0.3	11	48	15848	24740	208	237	38	0,1	3243	7
33. España	0.3	7	36	532347	161654	385	781	51	0	3458	5.7
34. Etiopia	2.6	112	41	5722	1293	2	51	13	0.3	22	0.1
35. Filipinas	2.3	39	37	71865	27062	21	686	26	3.4	316	0.8
36. Finlandia	0.4	5	48	105174	65546	550	440	77	0	5997	8.2
37. Francia	0.5	6	44	1451051	476200	558	665	25	-0.1	4042	9.3
38. Gabon	2.9	89	44	3759	933	30	97	71	0.6	652	5.3
39. Ghana	3	73	31	6719	6115	4	35	42	1.4	93	0.2
40. Grecia	0.5	8	36	85885	40623	493	523	47	0	2260	7.2
41. Guatemala	2.9	44	26	14255	3161	27	139	39	1.8	210	0.6
42. Haití	2	72	43	1770	369	8	1	1	5.1	29	0.1
43. Países Bajos	0.6	6	40	371039	79647	525	518	10	-0.3	4580	9.2
44. Honduras	3	45	30	3566	2672	29	294	41	2.2	204	0.6
45. Hungría	-0.3	11	44	42129	33486	185	661	18	-0.5	2383	5.8
46. India	1.9	68	32	319660	386500	13	612	17	0.6	248	0.9
51. Indonesia	1.7	51	40	190105	32414	17	96	60	1.1	366	1
47. Irak	2.7	108	18	24600	27060	33	4575	4	0.1	1213	3.4
48. Irán	3.2	45	24	113400	79128	79	1362	11	0	1505	4
49. Irlanda	0.1	6	33	52765	17105	365	233	6	-1.2	3137	8.7
50. Islandia	1.1	4	44	6686	4780	555	636	1	0	7932	6.8
52. Israel	2.7	8	40	87875	32781	418	428	6	-0.3	7717	8.1
53. Jordania	3.7	31	21	6354	3076	73	173	1	-1	1067	3
54. Kenia	2.9	58	46	7783	3539	9	87	2	0.6	110	0.2
55. Kuwait	-0.3	11	28	28941	22798	230	525	0	0	8622	11.2
56. Líbano	2.3	32	28	16673	5184	89	241	8	0.6	964	2.9
57. Libia	3.9	61	21	23400	17800	59	880	0	-1.4	2499	3.1
58. Malasia	2.6	12	37	78321	39093	166	768	54	2.1	1699	3.8
59. Marruecos	2	55	35	29545	11100	43	437	20	-1.4	327	1.1
60. México	2.1	33	31	304596	147926	96	899	25	1.3	1561	3.8
61. Mozambique	1.8	113	48	1353	490	3	55	22	0.8	40	0.1
62. Birmania	1.8	83	43	35840	3500	3	101	44	1.3	49	0.1
63. Nepal	2.5	91	40	4391	927	4	150	37	1	28	0.1
64. Nicaragua	3.1	46	36	1659	1688	23	367	50	1.9	300	0.6
65. Nigeria	2.9	80	36	28411	15530	4	41	17	0.7	162	0.9
66. Noruega	0.5	5	46	136277	113488	526	488	31	-1.4	5318	14.1
67. Nueva Zelanda	1	7	44	51655	35139	429	589	28	0	4245	7.6
68. Omán	4.5	18	15	10578	6187	77	564	19	0	2392	5.3
69. Pakistán	3	90	26	59991	58529	16	2053	2	3.5	254	0.6
70. Panamá	1.9	23	34	7293	3380	114	754	42	1.9	618	1.7
71. Paraguay	2.7	41	29	8558	36415	31	109	32	2.8	298	0.6
72. Perú	2.1	47	29	53019	15963	47	300	53	0.4	367	1
73. Polonia	0.4	14	46	107829	135347	148	321	28	-0.1	2401	8.9
74. Portugal	-0.1	7	43	96829	31380	361	739	34	-0.5	1827	4.8
75. Reino Unido	0.3	6	42	1094734	325383	502	205	10	-1.1	3732	9.2
76. Rep. Checa	0	8	47	39900	58705	236	266	34	0	3863	13.1
77. Rumania	0	23	44	33488	55136	131	113	37	0	1733	5.4
78. Senegal	2.8	62	42	5070	1002	10	202	39	0.7	97	0.4
79. Singapur	1.8	4	38	79831	20046	478	84	7	2.3	8103	17.7
80. Siria	3.1	32	26	15760	15186	63	435	4	-4.3	997	3.3
81. Sri Lanka	1.3	19	35	12619	4387	11	503	27	1.4	91	0.3
82. Sudán	2.3	77	28	7910	1333	3	633	18	1.1	66	0.1
83. Suecia	0.6	4	48	209720	142895	681	341	68	0	5723	6.6
84. Suiza	0.8	6	40	286014	65724	613	173	30	-0.6	8629	9.4
85. Suráfrica	2.3	50	37	130918	189416	95	359	4	-0.8	2146	7.5
86. Tailandia	1.3	39	46	192630	71177	59	602	25	3.9	769	2
87. Tanzania	3.1	82	49	3703	1913	3	40	38	1.2	34	0.1
88. Túnez	2.1	39	30	16369	6714	58	381	4	-1.9	595	1.6
89. Turquía	1.9	48	35	169452	78322	212	283	26	0	957	2.5
90. Ucrania	0.1	15	49	84084	202995	157	673	16	-0.3	3180	11.7
91. Uruguay	0.6	18	40	16458	7617	196	241	4	-0.6	629	1.6
92. Venezuela	2.4	23	33	65382	73116	111	382	52	1.2	2186	5.7
93. Vietnam	2.2	41	49	17634	12270	11	414	26	1.5	101	0.3
94. Yemen	4.2	100	29	4044	2159	12	335	8	0	206	0.7
95. Zambia	2.6	109	45	3605	1785	8	186	43	1.1	149	0.3
96. Zimbabue	2.8	55	44	5933	7334	14	136	23	0.7	438	1.8

21. Una situación muy habitual en análisis multivariante es disponer de un conjunto de datos mixto, es decir, un conjunto de individuos sobre los que se han observado tanto variables cuantitativas como cualitativas (o categóricas). En estos casos es de gran utilidad la *distancia de Gower*, cuyo cuadrado se define como $d_{ij}^2 = 1 - s_{ij}$, donde

$$s_{ij} = \frac{\sum_{h=1}^{p_1} (1 - |x_{ih} - x_{jh}|/G_h) + a + \alpha}{p_1 + (p_2 - d) + p_3} \tag{4}$$

es el coeficiente de similaridad de Gower, p_1 es el número de variables cuantitativas continuas, p_2 es el número de variables binarias, p_3 es el número de variables cualitativas(no binarias), a es el número de coincidencias $(1,1)$ en las variables binarias, d es el número de coincidencias $(0,0)$ en las variables binarias, α es el número de coincidencias en las variables cualitativas(no binarias)

y G_h es el rango (o recorrido) de la h-ésima variable cuantitativa.

Si $p_1 = p_3 = 0$ entonces (4) coincide con el coeficiente de similaridad de Jaccard. Si se consideran las variables binarias como categóricas (es decir, $p_1 = p_2 = 0$) entonces (4) coincide con el coeficiente de similaridad de Sokal y Michener.

(Datos del Problema 5.5) La Tabla 12 contiene información sobre 50 jugadores de fútbol de la liga española (temporada 2006/07). Las variables observadas son:
X_1 =número de goles marcados, X_2 =edad (años), X_3 =altura (m), X_4 =peso (kg), X_5 =pierna buena del jugador (1 =derecha, 0 =izquierda), X_6 =nacionalidad (1 =Argentina, 2 =Brasil, 3 =Camerun, 4 =Italia, 5 =España, 6 =Francia, 7 =Uruguay, 8 =Portugal, 9 =Inglaterra), X_7 =tipo de estudios (1 =sin estudios, 2 =básicos, 3 =medios, 4 =superiores).
Obtener la matriz de distancias de Gower entre estos individuos.

Table 12: Variables observadas sobre jugadores de la liga española de fútbol 2006/07.

	Jugador	X_1	X_2	X_3	X_4	X_5	X_6	X_7
1.	Ronaldinho	15	26	1.78	71	1	2	2
2.	Etoo	21	25	1.8	75	0	3	2
3.	Xavi	6	26	1.7	68	0	5	4
4.	Messi	7	19	1.69	67	0	1	3
5.	Puyol	1	28	1.78	78	0	5	3
6.	Raúl	7	29	1.8	73.5	1	5	3
7.	Ronaldo	18	30	1.83	82	0	2	1
8.	Beckham	4	31	1.8	67	0	9	3
9.	Casillas	0	25	1.85	70	0	5	4
10.	Cannavaro	0	33	1.76	75.5	0	4	2
11.	Torres	24	22	1.83	70	0	5	4
12.	Agüero	14	18	1.72	68	0	1	3
13.	Maxi	10	25	1.8	79	0	1	3
14.	Pablo	3	25	1.92	80	0	5	4
15.	Maniche	3	29	1.73	69	0	8	2
16.	Morientes	13	30	1.86	79	0	5	3
17.	Joaquin	5	25	1.79	75	0	5	4
18.	Villa	22	24	1.75	69	0	5	3
19.	Ayala	1	33	1.77	75.5	0	1	1
20.	Cañizares	0	36	1.81	78	1	5	3
21.	Jesus Navas	2	20	1.7	60	0	5	3
22.	Puerta	6	21	1.83	74	1	5	3
23.	Javi Navarro	7	32	1.82	75	0	5	3
24.	Daniel Alves	2	23	1.71	64	0	2	2
25.	Kanouté	12	29	1.92	82	1	6	1
26.	Valerón	9	31	1.84	71	0	5	3
27.	Arizmendi	8	22	1.92	78	0	5	3
28.	Capdevila	3	28	1.81	79	1	5	4
29.	Riki	7	26	1.86	80	0	5	3
30.	Coloccini	2	24	1.82	78	1	1	2
31.	Riquelme	10	28	1.82	75	0	1	2
32.	Forlán	17	27	1.72	75	0	7	3
33.	Cani	4	25	1.75	69.5	0	5	3
34.	Javi Venta	0	30	1.8	73	1	5	3
35.	Tachinardi	4	31	1.87	80	1	4	4
36.	Pandiani	6	30	1.84	74	0	7	1
37.	Tamudo	10	28	1.77	74	0	5	3
38.	De la Peña	2	30	1.69	69	0	5	3
39.	Luís García	8	25	1.8	68	0	5	3
40.	Jonathan	4	21	1.8	72	1	5	3
41.	Aimar	6	26	1.68	60	1	1	2
42.	Diego Milito	9	27	1.81	78	0	1	2
43.	Savio	3	32	1.71	68	1	2	2
44.	Sergio García	7	23	1.76	69	0	5	3
45.	Zapater	5	21	1.73	70.5	0	5	3
46.	Edú	6	27	1.82	74	1	2	3
47.	Juanito	2	30	1.83	80	0	5	4
48.	Melli	5	22	1.81	78	0	5	3
49.	Capi	7	29	1.75	73	0	5	2
50.	Doblas	0	25	1.84	78	0	5	3

22. Utilizando la matriz de distancias del Ejercicio 21 obtener una representación de los jugadores en coordenadas principales. Determinar cuál es el porcentaje

de variabilidad explicado por las dos primeras coordenadas principales. ¿Qué se puede decir de las semejanzas entre jugadores?

23. Considerar los datos del Ejercicio 20. Sea $\mathbf{D}^{(2)}$ la matriz de cuadrados de distancias obtenida a partir del coficiente de similaridad de Sokal y Michener.

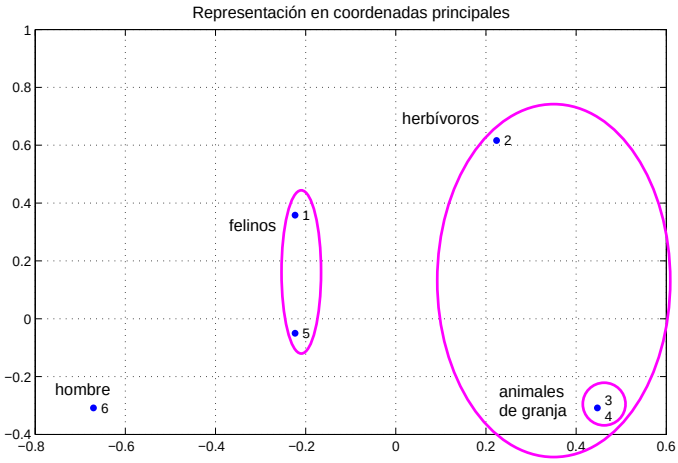
    (a) Verificar que $\mathbf{D}$ no es ultramétrica.

    (b) Realizar clasificaciones jerárquicas mediante los métodos del mínimo, del máximo y UPGMA. ¿Qué diferencias se observan?

    (c) Calcular la correlación cofenética en cada caso.

    (d) Comparar los dendrogramas con la representación en coordenadas principales que muestra la Figura 2.

Figure 2: Representación en coordenadas principales y agrupaciones (Ejercicio 20)



24. La Tabla 11 contiene una serie de indicadores económicos y sociales sobre 96 países del mundo. Sea $\mathbf{Y}$ la matriz que contiene las dos primeras componentes principales calculadas a partir de la matriz de correlaciones Obtener las distancias euclídeas entre países a partir de $\mathbf{Y}$ y realizar una clasificación jerárquica mediante el método UPGMA. Comentar los resultados obtenidos.

25. (Datos Problema 9.4) Un enólogo analiza dos componentes X_1 y X_2 en sendas muestras de dos tipos de vinos. Los resultados del análisis se pueden ver en la Tabla 13. Los datos se han extraído de Newman *et al.* (1998).

    (a) Denotemos $\mathbf{X} = (X_1, X_2)'$. Expresar la regla de clasificación lineal de Fisher para una nueva observación $\mathbf{x} = (x_1, x_2)'$. Programarla como una función de Matlab.

    (b) Aplicar la regla de clasificación obtenida en el apartado anterior al caso concreto en que $\mathbf{x} = (13.05, 515)'$. ¿A qué tipo de vino corresponde?

26. (Datos del Problema 8.2) La Tabla 14 contiene cuatro variables numéricas X_1 =longitud del sépalo, X_2 =anchura del sépalo, X_3 =longitud del pétalo,

Table 13: Muestras de dos vinos (Ejercicio 25). Fuente: Newman *et al.* (1998)

Vino 1		Vino 2	
X_1	X_2	x_1	x_2
14.23	1065	12.37	520
13.20	1050	12.33	680
13.16	1185	12.64	450
14.37	1480	13.67	630
13.24	735	12.37	420
14.20	1450	12.17	355
14.39	1290	12.37	678
14.06	1295	13.11	502
14.83	1045	12.37	510
13.86	1045	13.34	750
14.10	1510	12.21	718
14.12	1280	12.29	870
13.75	1320	13.86	410
14.75	1150	13.49	472
14.38	1547	12.99	985
13.63	1310		
14.30	1280		
13.83	1130		
14.19	1680		
13.64	845		

X_4 =anchura del pétalo medidas sobre tres especies de flores del género Iris: *Iris setosa*, *Iris versicolor* e *Iris virginica* (Fuente: Fisher 1936).

Supondremos que el vector $\mathbf{X}$ observado sigue una distribución normal. Dadas las tres nuevas flores (individuos)

ind.	X_1	X_2	X_3	X_4
$\mathbf{x}_1$	4.6	3.6	1.0	0.2
$\mathbf{x}_2$	6.8	2.8	4.8	1.4
$\mathbf{x}_3$	7.2	3.2	6.0	1.8

asignarlas a alguna de las tres especies (*I. setosa*, *I. virginica* o *I. versicolor*) mediante

(a) el discriminador lineal,

(b) el discriminador cuadrático.

Table 14: Datos para el Ejercicio 26

X_1	X_2	X_3	X_4	X_1	X_2	X_3	X_4	X_1	X_2	X_3	X_4
5.1	3.5	1.4	0.2	7.0	3.2	4.7	1.4	6.3	3.3	6.0	2.5
4.9	3.0	1.4	0.2	6.4	3.2	4.5	1.5	5.8	2.7	5.1	1.9
4.7	3.2	1.3	0.2	6.9	3.1	4.9	1.5	7.1	3.0	5.9	2.1
4.6	3.1	1.5	0.2	5.5	2.3	4.0	1.3	6.3	2.9	5.6	1.8
5.0	3.6	1.4	0.2	6.5	2.8	4.6	1.5	6.5	3.0	5.8	2.2
5.4	3.9	1.7	0.4	5.7	2.8	4.5	1.3	7.6	3.0	6.6	2.1
4.6	3.4	1.4	0.3	6.3	3.3	4.7	1.6	4.9	2.5	4.5	1.7
5.0	3.4	1.5	0.2	4.9	2.4	3.3	1.0	7.3	2.9	6.3	1.8
4.4	2.9	1.4	0.2	6.6	2.9	4.6	1.3	6.7	2.5	5.8	1.8
4.9	3.1	1.5	0.1	5.2	2.7	3.9	1.4	7.2	3.6	6.1	2.5
5.4	3.7	1.5	0.2	5.0	2.0	3.5	1.0	6.5	3.2	5.1	2.0
4.8	3.4	1.6	0.2	5.9	3.0	4.2	1.5	6.4	2.7	5.3	1.9
4.8	3.0	1.4	0.1	6.0	2.2	4.0	1.0	6.8	3.0	5.5	2.1
4.3	3.0	1.1	0.1	6.1	2.9	4.7	1.4	5.7	2.5	5.0	2.0
5.8	4.0	1.2	0.2	5.6	2.9	3.6	1.3	5.8	2.8	5.1	2.4
5.7	4.4	1.5	0.4	6.7	3.1	4.4	1.4	6.4	3.2	5.3	2.3
5.4	3.9	1.3	0.4	5.6	3.0	4.5	1.5	6.5	3.0	5.5	1.8
5.1	3.5	1.4	0.3	5.8	2.7	4.1	1.0	7.7	3.8	6.7	2.2
5.7	3.8	1.7	0.3	6.2	2.2	4.5	1.5	7.7	2.6	6.9	2.3
5.1	3.8	1.5	0.3	5.6	2.5	3.9	1.1	6.0	2.2	5.0	1.5
5.4	3.4	1.7	0.2	5.9	3.2	4.8	1.8	6.9	3.2	5.7	2.3
5.1	3.7	1.5	0.4	6.1	2.8	4.0	1.3	5.6	2.8	4.9	2.0
4.6	3.6	1.0	0.2	6.3	2.5	4.9	1.5	7.7	2.8	6.7	2.0
5.1	3.3	1.7	0.5	6.1	2.8	4.7	1.2	6.3	2.7	4.9	1.8
4.8	3.4	1.9	0.2	6.4	2.9	4.3	1.3	6.7	3.3	5.7	2.1
5.0	3.0	1.6	0.2	6.6	3.0	4.4	1.4	7.2	3.2	6.0	1.8
5.0	3.4	1.6	0.4	6.8	2.8	4.8	1.4	6.2	2.8	4.8	1.8
5.2	3.5	1.5	0.2	6.7	3.0	5.0	1.7	6.1	3.0	4.9	1.8
5.2	3.4	1.4	0.2	6.0	2.9	4.5	1.5	6.4	2.8	5.6	2.1
4.7	3.2	1.6	0.2	5.7	2.6	3.5	1.0	7.2	3.0	5.8	1.6
4.8	3.1	1.6	0.2	5.5	2.4	3.8	1.1	7.4	2.8	6.1	1.9
5.4	3.4	1.5	0.4	5.5	2.4	3.7	1.0	7.9	3.8	6.4	2.0
5.2	4.1	1.5	0.1	5.8	2.7	3.9	1.2	6.4	2.8	5.6	2.2
5.5	4.2	1.4	0.2	6.0	2.7	5.1	1.6	6.3	2.8	5.1	1.5
4.9	3.1	1.5	0.2	5.4	3.0	4.5	1.5	6.1	2.6	5.6	1.4
5.0	3.2	1.2	0.2	6.0	3.4	4.5	1.6	7.7	3.0	6.1	2.3
5.5	3.5	1.3	0.2	6.7	3.1	4.7	1.5	6.3	3.4	5.6	2.4
4.9	3.6	1.4	0.1	6.3	2.3	4.4	1.3	6.4	3.1	5.5	1.8
4.4	3.0	1.3	0.2	5.6	3.0	4.1	1.3	6.0	3.0	4.8	1.8
5.1	3.4	1.5	0.2	5.5	2.5	4.0	1.3	6.9	3.1	5.4	2.1
5.0	3.5	1.3	0.3	5.5	2.6	4.4	1.2	6.7	3.1	5.6	2.4
4.5	2.3	1.3	0.3	6.1	3.0	4.6	1.4	6.9	3.1	5.1	2.3
4.4	3.2	1.3	0.2	5.8	2.6	4.0	1.2	5.8	2.7	5.1	1.9
5.0	3.5	1.6	0.6	5.0	2.3	3.3	1.0	6.8	3.2	5.9	2.3
5.1	3.8	1.9	0.4	5.6	2.7	4.2	1.3	6.7	3.3	5.7	2.5
4.8	3.0	1.4	0.3	5.7	3.0	4.2	1.2	6.7	3.0	5.2	2.3
5.1	3.8	1.6	0.2	5.7	2.9	4.2	1.3	6.3	2.5	5.0	1.9
4.6	3.2	1.4	0.2	6.2	2.9	4.3	1.3	6.5	3.0	5.2	2.0
5.3	3.7	1.5	0.2	5.1	2.5	3.0	1.1	6.2	3.4	5.4	2.3
5.0	3.3	1.4	0.2	5.7	2.8	4.1	1.3	5.9	3.0	5.1	1.8