ANÁLISIS DISCRIMINANTE Y CORRELACIONES CANÓNICAS

Prof. ESPERANZA AYUGA TÉLLEZ

ANÁLISIS DISCRIMINANTE

Propuesto por primera vez por Fisher, físico que trabajó en una estación experimental agrícola y célebre por sus estudios genéticos.

Tiene por objeto clasificar un nuevo elemento observado, usando valores de las variables conocidas, en alguna de las poblaciones que originan éstas. Conocido como reconocimiento de patrones o clasificación supervisada.

p.e. Clasificar los restos de un cráneo descubierto en una excavación como humano, partiendo de medidas físicas de cráneos humanos y de antropoides.

- **Clasificación entre dos poblaciones**:

Tenemos dos poblaciones, P_1 y P_2, con un vector aleatorio **X** continuo, p-dimensional, definido en ambas y con funciones de densidad multivariantes conocidas ($f_1(\mathbf{X})$ y $f_2(\mathbf{X})$).

Queremos clasificar x_0 en una de las dos poblaciones.

Si conocemos π_i, probabilidad "a priori" de que el elemento proceda de P_i, con $\pi_1+\pi_2=1$, entonces (por Bayes)

$$x_0 \in P_2 \Leftrightarrow \pi_2 f_2(x_0) > \pi_1 f_1(x_0)$$

- **Consecuencias de la clasificación errónea**:

Clasificamos en P_2 si:

a) π_2 ("a priori") es más alta (a igualdad del resto)

b) f_2 (verosimilitud) es más alta (a igualdad del resto)

c) El coste de equivocarnos es más bajo (a igualdad del resto)

- **Caso de dos poblaciones normales**:

Tenemos dos poblaciones de $f_1(\mathbf{X})$ y $f_2(\mathbf{X})$ normales con $\mathbf{V}_1=\mathbf{V}_2=\mathbf{V}$ entonces la regla general anterior se reduce a clasificar en P_2 si $D_1^2>D_2^2$,

con $D_i^2 = (\mathbf{X}-\mu_i)'\mathbf{V}^{-1}(\mathbf{X}-\mu_i)$= distancia de Mahalanobis

O bien, construir la variable indicador z= $\mathbf{w}'\mathbf{x}$ con $\mathbf{w}= \mathbf{V}^{-1} (\mu_2 -\mu_1)$ y clasificar z en P_2 si $|z-m_1| > |z-m_2|$ con $m_i= \mathbf{w}'\mu_i$.

Esto equivale a buscar la dirección óptima de proyección para discriminar.

- **Probabilidad de error**:

$$P(2/1) = P(1/2) = \Phi\,(-D/2)$$

con Φ función de distribución de la Normal estandar y $D^2 = (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)'\mathbf{V}^{-1}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)$= distancia de Mahalanobis.

- **Probabilidad de acertar**:

$$P(1/x) = \frac{1}{1 + \frac{\pi_1}{\pi_2}\exp\left\{-\frac{1}{2}\left(D_2^2 - D_1^2\right)\right\}}$$

que nos indica la confianza en la clasificación.

- **Generalización a varias poblaciones**:

Si tenemos G poblaciones, se divide el espacio en G regiones A_g tales que si $x \in A_g$ se clasifica el punto en la población P_g.

La regla de decisión de máxima verosimilitud es:

$$A_g = \{x \in A_g / \pi_g f_g(x) > \pi_i f_i(x); \forall i \neq g\}$$

Esto equivale a calcular las D^2 de x al centro de cada población y clasificarla en la P_g que haga esta distancia mínima (si todas las π_i son iguales y $f_i(x)$ normales con la misma matriz de varianzas).

Para G poblaciones se necesitan r=min(G-1,p)

- **Poblaciones desconocidas**:

Si sólo disponemos de la muestra:

clasificamos x_0 en la población P_g si $\min_g (x_0 - \bar{x}_g)' \hat{S}_w^{-1} (x_0 - \bar{x}_g$

o construimos $z_{g,g+1} = \hat{w}'_{g,g+1} x_0$, con $\hat{w}_{g,g+1} = \hat{w}_g - \hat{w}_{g+1}$

y clasificamos en g frente a g+1 si

$$\left| z_{g,g+1} - \hat{m}_g \right| < \left| z_{g,g+1} - \hat{m}_{g+1} \right|$$

El error de clasificación es

ε= total mal clasificados/total bien clasificados

También podemos construir n funciones discriminantes con n-1 observaciones y clasificamos el dato con la regla construida sin él (validación cruzada)

V. CANÓNICAS DISCRIMINANTES

- Se construyen las variables canónicas que tengan máximo poder discriminante (proyecciones en las direcciones de máxima distancia) mediante los autovalores y que son incorreladas.

- Cuando p y G son grandes es frecuente que la mayor discriminación se consiga con pocas variables canónicas.

OTRAS FORMAS DE DISCRIMINACIÓN

- Cuadrática: Si las V son distintas se clasifica la observación en el grupo con más probabilidad "a posteriori" ⇒ f. Discrim. Cuadrática (con regiones no disjuntas y nº de parámetros a estimar mayor).

- Bayesiana: Para v.a. con cualquier distribución. Con la probabilidad "a posteriori" ⇒ f. Discrim. Cuadrática.

USO DEL A. DISCRIMINANTE

En muchos casos en que se necesita clasificar elementos con informaciones incompletas

Fases de su aplicación

1º

PREPARACIÓN DE LOS DATOS PARA EL ANÁLISIS

Comprobación de los supuestos básicos

Elección de las variables predictoras

Selección de los casos a analizar

Análisis de las distribuciones univariantes de los grupos

2º

LAS FUNCIONES DISCRIMINANTES

Estimación de las funciones discriminantes

Análisis discriminante simultáneo

Análisis discriminante no simultáneo

Derivación de los coeficientes de las funciones discriminantes

Significatividad de las funciones

Obtención de las puntuaciones discriminantes

ANÁLISIS DISCRIMINANTE

ESTADÍSTICOS:

•Determinar la importancia relativa de cada v. indep. En la diferenciación de los grupos:

- opesos discriminantes estandarizados
- oCorrelaciones de estructuras discriminantes
- oValores F parciales

•Examen de las medias grupales en relación con cada función discriminante

•Cálculo de los índices de potencialidad

GRÁFICOS:

•Centroides grupales

•Correlaciones discriminantes

•Mapas territoriales

•Histogramas de puntuaciones discriminantes

•Diagramas de dispersión para todas las funciones

ES

EV
De
Ob
Ap

IN

R
E
P
L

GRÁFICOS

N DEL MODELO

5º

SITIVA

Conclusión del análisis

ación de nuevos objetos a los grupos

DECISIONES INICIALES

1. Elección de variables predictoras:

Se basa en técnicas previas de clasificación grupal.

2. Selección de casos a analizar:

La eliminación de dichos casos se basará en el estudio detallado de éstos. Si son demasiadas, si son relevantes, etc.

Si se emplea la validación cruzada para contrastar las funciones discriminantes hay determinar qué parte de la muestra se elimina de la estimación y se emplea en la validación.

DECISIONES INICIALES

3. Modalidad de análisis:

Dependiendo de si sólo se quiere discriminar o también se quiere emplear un número reducido de ***variables predictoras***:

•Análisis discriminante simultáneo: se emplean todas.

•Análisis discriminante secuencial: serie reducida en consonancia con su poder discriminatorio. La incorporación es secuencial, se introduce una nueva variable en consonancia con su poder discriminante y se analiza la colinealidad.

4. Descriptiva univariante:

Se analizan las diferencias entre grupos de los estadísticos de cada variable.

ESTIMACIÓN DE LAS FUNCIONES

Función discriminante canónica: combinación lineal de "p" variables predictoras que más discriminan entre los grupos definidos "a priori"

$$f_{km} = u_0 + u_1 X_{1km} + \ldots + u_p X_{pkm}$$

f_{km} = puntuación para el caso m en el grupo k

X_{ikm} = valor de la v. X_i para el caso m en el grupo k

La puntuación discriminante representa la proyección de ese caso a lo largo del eje discriminante definido por la función.

Los coeficientes se calculan para maximizar diferencias entre centroides y los valores u_i incorrelados entre las diferentes funciones.

ESTIMACIÓN DE LAS FUNCIONES

Nº de funciones y significatividad:

$$N\ max = \min(p, g-1) \quad \begin{cases} p = \text{nº de variables usadas} \\ g = \text{nº de grupos} \end{cases}$$

Relevancia de las funciones: se comprueba con la conjunción

Autovalores: λ_i= SCEG/SCIG, cuanto mayor es λ_i más discriminación

Porcentaje de varianza: % de V relativo que representa cada función.

Correlación canónica: $r_i = [\lambda_i/(1+ \lambda_i)]^{1/2}$, mide el grado de asociación entre el grupo y la función, cuanto más próximo a 1 mejor asociación.

ESTIMACIÓN DE LAS FUNCIONES

Nº de funciones y significatividad:

Se comprueba con los estadísticos:

Lambda de Wilks: se tienen valores pequeños cuando hay mucha variabilidad entre grupos y poca dentro de ellos, valores cercanos a 1 indican que la función no logra diferenciar entre grupos.

Chi-cuadrado: mide la discriminación residual. Si $p>0{,}05$ no procede seguir estimando funciones discriminantes.

Estandarización de coeficientes:

Las puntuaciones se obtienen con los coef. sin estandarizar.

Los coef. estandarizados se emplean como referentes de la contribución de la variable a la función discriminante

ESTIMACIÓN DE LAS FUNCIONES

Procedimiento secuencial para elegir v. predictoras:

Hacia delante (forward): se van incluyendo variables por su poder discriminante.

Hacia atrás (backward): se van eliminando variables.

Criterios de selección:

Lambda de Wilks: se selecciona la variable con lambda menor

La razón F parcial: razón de variabilidad entre y var. intra. Interesan valores elevados. Hay que elegir los valores:

F mínimo (F-to enter) para entrar >2, entre 2,5 y 5
F máximo (F-to remove) para salir >2, entre 2,5 y 5

$$F_{min} > F_{max}$$

Para incorporar, F debe ser elevado y α=0,05

CAPACIDAD PREDICTIVA

Procedimiento para VALORAR la capacidad predictiva:

Tabla de clasificación: Se incluyen los casos bien y mal clasificados y en qué grupo.

Se determinará una probabilidad "a priori" de asignación a cada uno de los grupos:

- Para todos igual
- Proporcional al número de casos en cada grupo.
- Otra asignación

El caso se clasifica en el grupo con mayor probabilidad "a posteriori"

El éxito se mide con el porcentaje de casos correctamente clasificados.

INTERPRETACIÓN

1. DESCRIPCIÓN de las funciones : destacando el poder discriminatorio de las variables que la forman.

- Los coeficientes estandarizados
- Las correlaciones
- Los valores de F parciales

2. EXAMEN de los centroides : su finalidad es obtener una visión global de las diferencias grupales respecto a las funciones obtenidas.

ULTIMAS APLICACIONES

Patrones de comportamiento (2005):

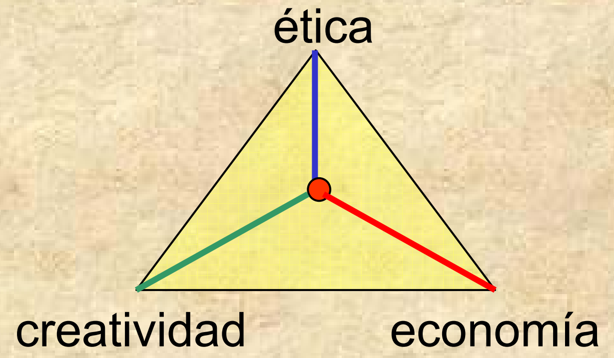
Estudio sobre diferencias entre los patrones de asimilación de CO_2, eficiencia fotosintética y crecimiento del *Schinus* frente a cinco especies nativas de Florida en diferentes condiciones de salinidad: neutra, baja y alta.

ULTIMAS APLICACIONES

ULTIMAS APLICACIONES

Patrones de comportamiento (2005):

Estudio sobre modelos de gestión de organizaciones culturales mediante medidas de los valores del individuo sobre tres ejes: economía-práctica; creatividad-emocional y ética-social.



Control de procesos (2005):

Programa informático para el comprobar si el proceso está o no bajo control, entrando múltiples variables físicas y tratamiento estadístico de éstas, incluido el AD.

ULTIMAS APLICACIONES

Tipología estructural (2005):

Asignación de parcelas forestales a diferentes tipologías estructurales de los hayedos burgaleses, atendiendo a variables dasométricas y medidas de la biodiversidad de las parcelas, con un 95% de eficiencia en la clasificación.

Ejemplo 1:

Una máquina que admite monedas realiza 3 mediciones de cada moneda para determinar su valor: peso(X1), espesor(X2) y densidad de estrías en su canto(X3). Los instrumentos de medición de estas variables no son muy precisos y se ha comprobado en una amplia experimentación con 3 tipos de monedas M1, M2 y M3, que las medidas son $N(\mu, \mathbf{V})$

Ejemplo 1:

$$\underset{\mu_1}{\begin{bmatrix} 20 \\ 8 \\ 8 \end{bmatrix}} \qquad \underset{\mu_2}{\begin{bmatrix} 19{,}5 \\ 7{,}8 \\ 10 \end{bmatrix}} \qquad \underset{\mu_3}{\begin{bmatrix} 20{,}5 \\ 8{,}3 \\ 5 \end{bmatrix}}$$

$$\mathbf{V} = \begin{bmatrix} 4 & 0{,}8 & -5 \\ 0{,}8 & 0{,}25 & -0{,}9 \\ -5 & -0{,}9 & 9 \end{bmatrix}$$

Clasificar la moneda de medidas

(22; 8,5;7)´

Aparentemente está más próxima a M3

Pero podría ser M1

$$\mathbf{V}^{-1} = \begin{bmatrix} 1.5 & -2.8 & 0.55 \\ -2.8 & 11.5 & -0.42 \\ 0.55 & -0.42 & 0.37 \end{bmatrix}$$

Ejemplo 1:

Funciones discriminantes:

$$Z1 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_3)' \mathbf{V}^{-1}\mathbf{x} = 1{,}77x_1 - 3{,}31x_2 + 0{,}98x_3$$

$$Z2 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \mathbf{V}^{-1}\mathbf{x} = -0{,}93x_1 + 1{,}74x_2 - 0{,}56x_3$$

$$Z3 = (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_3)' \mathbf{V}^{-1}\mathbf{x} = Z1 - Z2$$

$$\left.\begin{array}{l} Z1(\boldsymbol{\mu}_1) = 1{,}77\times 20 - 3{,}31\times 8 + 0{,}98\times 8 = 16{,}71 \\ Z1(\boldsymbol{\mu}_2) = 1{,}77\times 20{,}5 - 3{,}31\times 8{,}3 + 0{,}98\times 5 = 13{,}65 \end{array}\right\}$$

La media o punto de corte es 15,17

Z1=1,77x22-3,31x8,5+0,98x7=17,61 > 15,17⇒M1

Ejemplo 1:

Equivale a calcular D²:

$$D_1^2 = (\mathbf{x} - \boldsymbol{\mu}_1)' \mathbf{V}^{-1} (\mathbf{x} - \boldsymbol{\mu}_1) = 1{,}84$$

$$D_2^2 = (\mathbf{x} - \boldsymbol{\mu}_2)' \mathbf{V}^{-1} (\mathbf{x} - \boldsymbol{\mu}_2) = 2{,}01$$

$$D_3^2 = (\mathbf{x} - \boldsymbol{\mu}_3)' \mathbf{V}^{-1} (\mathbf{x} - \boldsymbol{\mu}_3) = 6{,}69$$

D_1^2 es la menor $\Rightarrow$ clasificamos en M1

La moneda que queremos clasificar tiene mucho peso y espesor ($\in$M3) entonces la densidad de las estrías debía ser bajo

Ejemplo 1:

Hemos clasificado la moneda en M1 y no en M3 como pensamos al principio. Para explicarlo estudiamos la matriz de correlaciones entre coeficientes estandarizados.

$$\mathbf{R} = \begin{bmatrix} 1 & 0{,}8 & -0{,}83 \\ 0{,}8 & 1 & -0{,}6 \\ -0{,}83 & -0{,}6 & 1 \end{bmatrix}$$

La moneda que queremos clasificar tiene mucho peso y espesor ($\in$M3) entonces la densidad de estrías debía ser bajo (correlación negativa). Sin embargo es alto, valor compatible con una moneda M1 sucia (la suciedad aumenta peso y espesor)

Ejemplo 2:

Se tienen 360 observaciones de distintas zonas de la costa, que se han clasificado, atendiendo al grado de salinidad y contaminantes, en aguas no degradadas (1), algo degradadas (2) y muy degradadas (3).

Vamos determinar si la presencia de determinados organismos nos permite asignar el ecosistema marino a alguno de estos grados.

Ejemplo 2:

Las variables que se consideran en la obtención de las funciones discriminantes son:

bivalvos

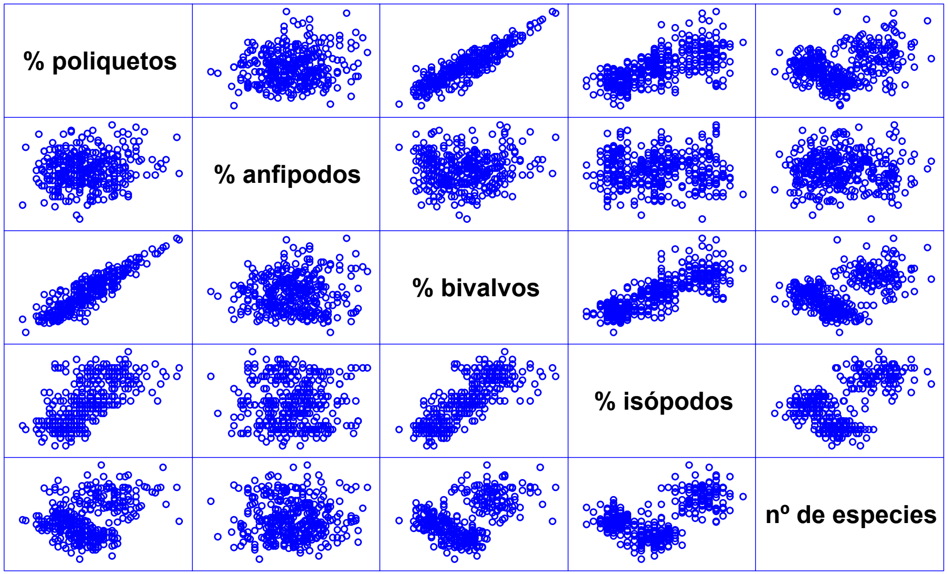
poliquetos

Nº de sp distintas

isópodos

anfípodos

Ejemplo 2:



Resumen estadístico

	% poliq	% anfip	% bivs	% isóp	nº de sp
Nº	360	360	360	360	360
$\bar{X}$	5,95833	2,9525	4,37028	1,24028	52,5889
S	1,00765	1,20146	1,1505	0,53954	25,8341
CV	16,9116%	40,6929%	26,3256%	43,5015%	49,1247%

Las variables isópodos/bivalvos y poliquetos/bivalvos están correlacionadas linealmente.

El % de isópodos y nº de especies distintas son los más diferentes en cuanto a medias y desviaciones típicas.

Ejemplo 2:

Number of complete cases: 360
Number of groups: 3

Función	Valor propio	% Relativo	Correlación Canonica
1	9,36485	83,00	0,95054
2	1,91833	17,00	0,81076

Las dos funciones son discriminantes (f1> f2)

Funciones Derivadas	Wilks Lambda	Chi	g.d.l.	P-Valor
1	0,0330599	1213,7581	6	0,0000
2	0,342661	381,2804	2	0,0000

Las dos funciones obtienen grupos con medias diferentes

Ejemplo 2:

Stepwise regression: **Method: forward selection**

F-to-enter: 4,0 / F-to-remove: 4,0

Step 0: 0 variables in the model.

Step 1: Adding variable % isópodos with F-to-enter = 940,328

---

1 variables in the model.
Wilk's lambda = 0,159542 Approximate F = 940,328 with P-value = 0,0000

Step 2: Adding variable nº de especies with F-to-enter = 430,292

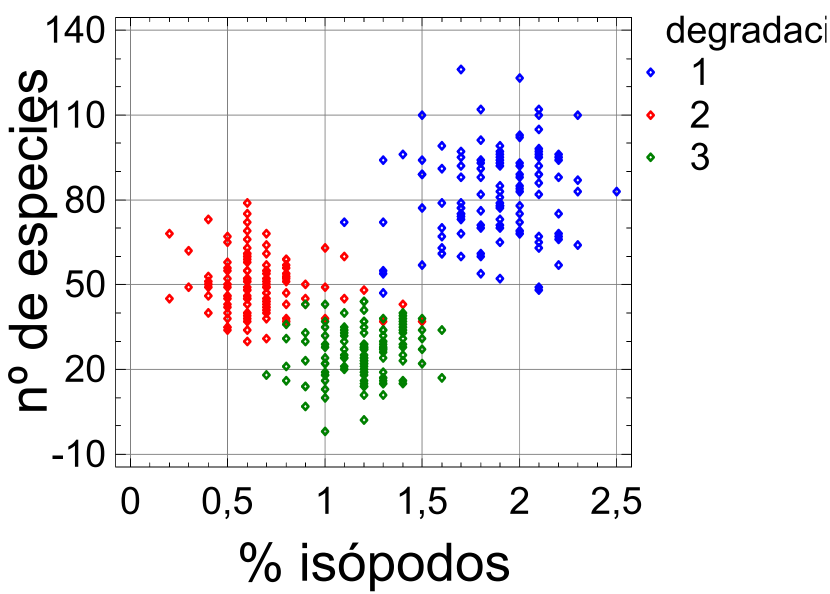
---

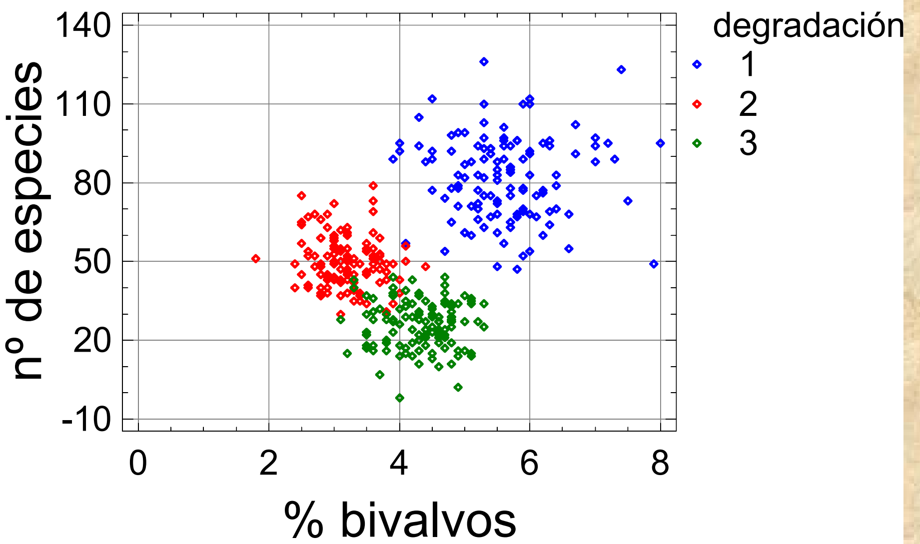
2 variables in the model.
Wilk's lambda = 0,0466856 Approximate F = 645,813 with P-value = 0,0000

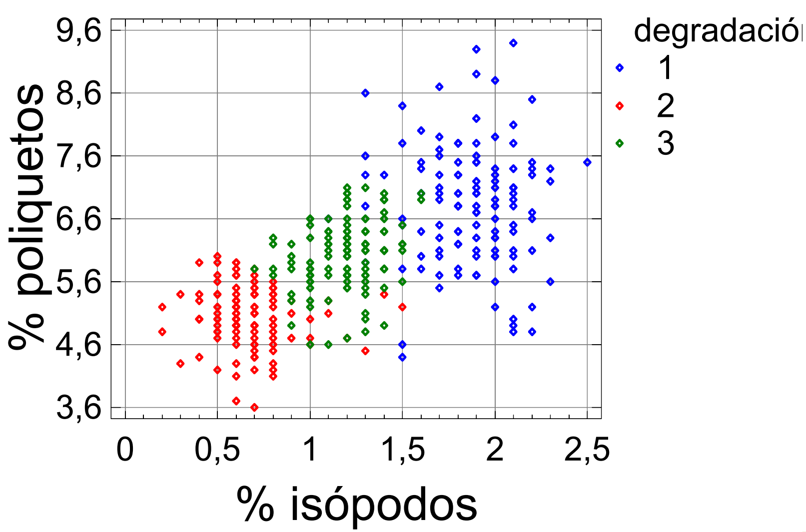
Step 3: Adding variable % bivalvos with F-to-enter = 73,1565

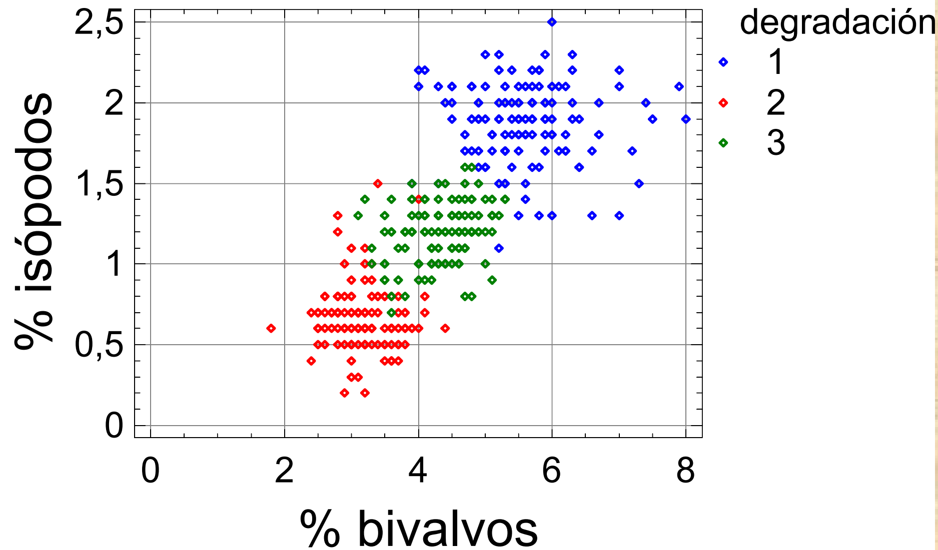
---

3 variables in the model.
Wilk's lambda = 0,0330599 Approximate F = 532,479 with P-value = 0,0000


degradaci
1
2
3
nº de especies
140
110
80
50
20
-10
0
0,5
1
1,5
2
2,5
% isópodos


degradación
1
2
3
nº de especies
140
110
80
50
20
-10
0
2
4
6
8
% bivalvos


degradació
1
2
3
% poliquetos
9,6
8,6
7,6
6,6
5,6
4,6
3,6
0
0,5
1
1,5
2
2,5
% isópodos


degradación
1
2
3
% isópodos
2,5
2
1,5
1
0,5
0
0
2
4
6
8
% bivalvos

Ejemplo 2:

Grupo actual	tamaño grupo	Degradación asignada 1	2	3
1	120	119	0	1
		(99,17%)	(0,00%)	(0,83%)
2	120	0	117	3
		(0,00%)	(97,50%)	(2,50%)
3	120	0	2	118
		(0,00%)	(1,67%)	(98,33%)

% de casos correctamente clasificados: 98,33%

La clasificación más acertada es la de los ecosistemas no degradados

G	Prob. *a priori*
1	0,3333
2	0,3333
3	0,3333

La probabilidad "a priori" es igual para todos los grupos y proporcional al tamaño del grupo

Ejemplo 2:

Classification Function Coefficients for degradación

	1	2	3
% bivalvos	16,018	9,24273	12,244
% isópodos	36,2747	11,8938	23,6155
nº de especies	0,553299	0,350885	0,173621
CONSTANT	-102,461	-28,5171	-43,9113

Z1= 16,081%biv+36,2747%isop+0,553299nºsp-102,461

Z2= 9,24273%biv+11,8938%isop+0,50885nºsp-28,5171

Z3= 12,244%biv+23,6155%isop+0,173621nºsp-43,9113

Clasificamos en el grupo 2 si Z2>Z1 y Z2>Z3

Coeficientes de la Función Discriminante para degradación

En la f1 todas las variables contribuyen por igual y en f2 la variable de mayor importancia es la diversidad

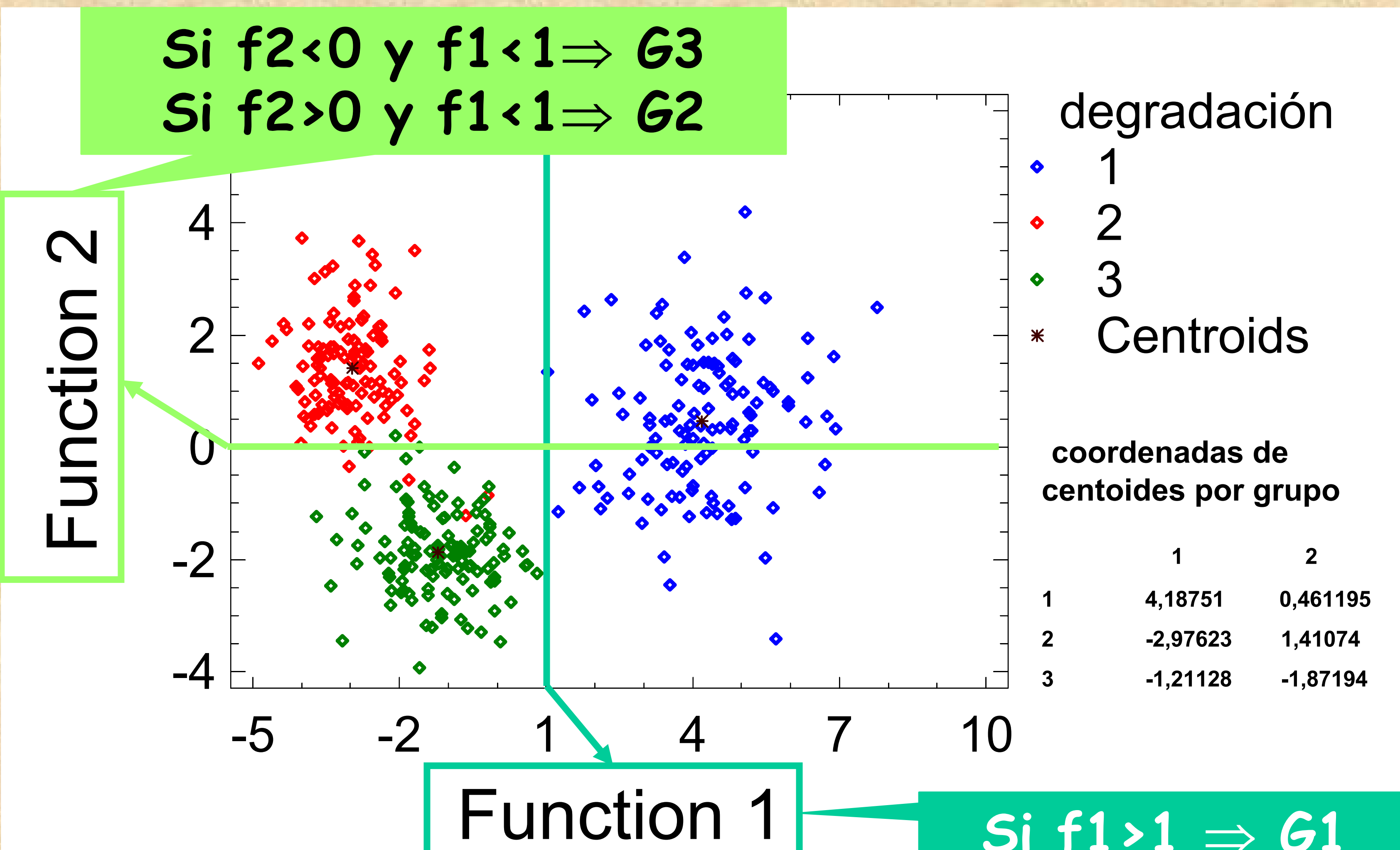
Standardized Coefficients

	1	2
% bivalvos	0,524842	-0,258276
% isópodos	0,681804	-0,405098
nº de especies	0,456872	0,892658

Unstandardized Coefficients

	1	2
% bivalvos	0,887861	-0,436919
% isópodos	3,1549	-1,8745
nº de especies	0,0381303	0,0745006
CONSTANT	-9,79838	0,316459

Ejemplo 2:



coordenadas de centoides por grupo

	1	2
1	4,18751	0,461195
2	-2,97623	1,41074
3	-1,21128	-1,87194

Ejemplo 2:

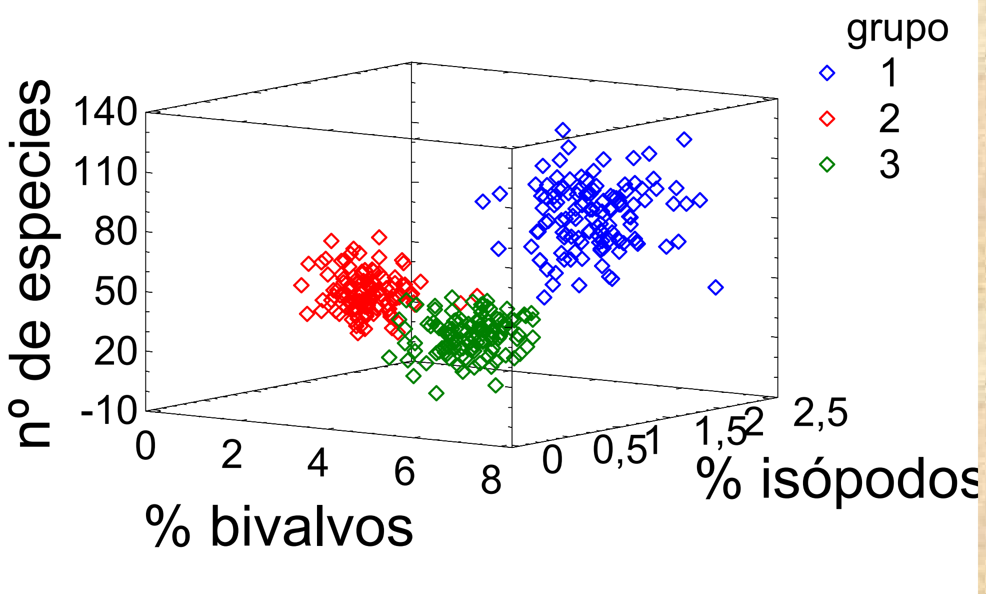
degradación	1	2	3	TOTAL
COUNTS	120	120	120	360
MEANS				
% bivalvos	5,5975	3,18167	4,33167	4,37028
% isópodos	1,86917	0,6625	1,18917	1,24028
nº de sp	81,8	50,0167	25,95	52,5889
STD. DEVIATIONS				
% bivalvos	2,3659	1,78372	2,08127	2,09052
% isópodos	1,36717	0,813941	1,09049	1,11368
nº de sp	9,04434	7,07225	5,09411	7,25182

Within-Group Covariance Matrix

	% bivalvos	% isópodos	nº de especies
% bivalvos	0,349436	0,0043401	-0,284127
% isópodos	0,0043401	0,0467033	0,190672
nº de especies	-0,284127	0,190672	143,565

Within-Group Correlation Matrix

	% bivalvos	% isópodos	nº de especies
% bivalvos	1,0	0,0339737	-0,0401147
% isópodos	0,0339737	1,0	0,0736357
nº de especies	-0,0401147	0,0736357	1,0

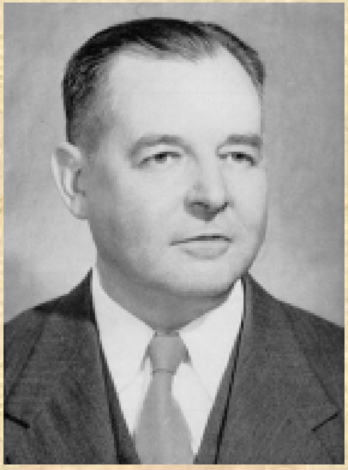


Ejemplo 2:

- La degradación de las costas se puede determinar por la biodiversidad de los ecosistemas marinos.
- El A. discriminante es una técnica estadística muy eficaz (menos del 20% de error) para predecir la degradación de las costas por medio de los organismos encontrados en ellas.

DEPENDENCIA ENTRE CONJUNTOS DE VARIABLES

Hotelling propone en 1936 las CORRELACIONES CANÓNICAS como una extensión de las componentes principales



Tiene por objeto relacionar las variables en dos grupos.

p.e. Para relacionar un conjunto de variables que midan el rendimiento escolar y otro grupo que mida el uso de los tiempos de ocio, o cuando tratamos de relacionar las variables que miden el rendimiento en Secundaria con las notas de la Universidad.

La relación se puede buscar con dos enfoques:

- Simétrico: cuando no existe un conjunto que sea la causa del otro (p.e. las variables que miden características fisiológicas y morfológicas de las plantas, están relacionadas sin causalidad).
- Asimétrico: cuando unas variables explican las otras pero no al revés (p.e. el tipo de suelo y el crecimiento de las plantas)

VARIABLES CANÓNICAS

El problema es encontrar 2 v. resumen, una de cada conjunto, que tengan correlación máxima:

$$x^* = \mathbf{X}\boldsymbol{\alpha} = \sum_{i=1}^{p} \alpha_i x_i \quad \text{y} \quad y^* = \mathbf{Y}\boldsymbol{\beta} = \sum_{j=1}^{q} \beta_j y_j$$

Si las variables son normales de media 0, la solución consiste en construir las 2 matrices:

$$\mathbf{A}_{pxp} = \mathbf{V}_{11}^{-1}\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21} \quad \text{y} \quad \mathbf{B}_{qxq} = \mathbf{V}_{22}^{-1}\mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{V}_{12}$$

Y calcular el vector asociado a su máximo valor propio, vector que proporciona las v. canónicas.

VARIABLES CANÓNICAS

El objetivo es relacionar de forma global un grupo de variables $x_1, x_2,\ldots,x_m$ con otro grupo de variables, $y_1, y_2,\ldots,y_n$ todas ellas medidas en la misma población a través de nuevas variables no medibles, $U=(u_{1,\ldots} u_m)$ y $V=(v_1,\ldots,v_n)$:

$U_1=u_{11}\, x_1+\ldots+u_{1m}\, x_m$

$V_1=v_{11}\, y_1+\ldots+v_{1n}\, y_n$

Con la propiedad de que la correlación entre U y V es máxima.

VARIABLES CANÓNICAS

- **Metodología**

Para relacionar (x_1, x_2,...,x_m) con,(y_1, y_2,...,y_n) habrá que encontrar 2 vectores ($u_{1,}$ ...u_m) y (v_1 ,...,v_n) con las propiedades:

1. $u_{1,}$ ...u_m son mutuamente incorreladas.
2. v_1 ,...,v_n son mutuamente incorreladas.
3. Las correlaciones cuadráticas o correlaciones canónicas entre (u_1,v_1), (u_2, v_2), ..., (u_n, v_n) son máximas: $\rho_1^2 >= \rho_2^2 >= ... >= \rho_n^2$
4. Las variables canónicas se obtienen a partir de:

VARIABLES CANÓNICAS

$$\mathrm{Det}(C_{12}\, C_{22}^{-1}\, C_{21} - \rho_i^2\, C_{11}) = 0$$

Matriz de Covarianzas de y con x

Matriz de Covarianzas de $y_1,\ldots,y_m$

Matriz de Covarianzas de $x_1,\ldots,x_m$

Correlaciones canónicas-autovalores

$C_{12}\, C_{22}^{-1}\, C_{21}\, u_i = \rho_i^2\, C_{11} u_i$

Donde $(u_{1i}, \ldots u_{mi})$ es el vector canónico.

$C_{12}\, C_{11}^{-1}\, C_{21}\, v_i = \rho_i^2\, C_{22} v_i$

Donde $(v_{1i}, \ldots v_{ni})$ es el vector canónico.

CONSTRUCCIÓN DE V. C.

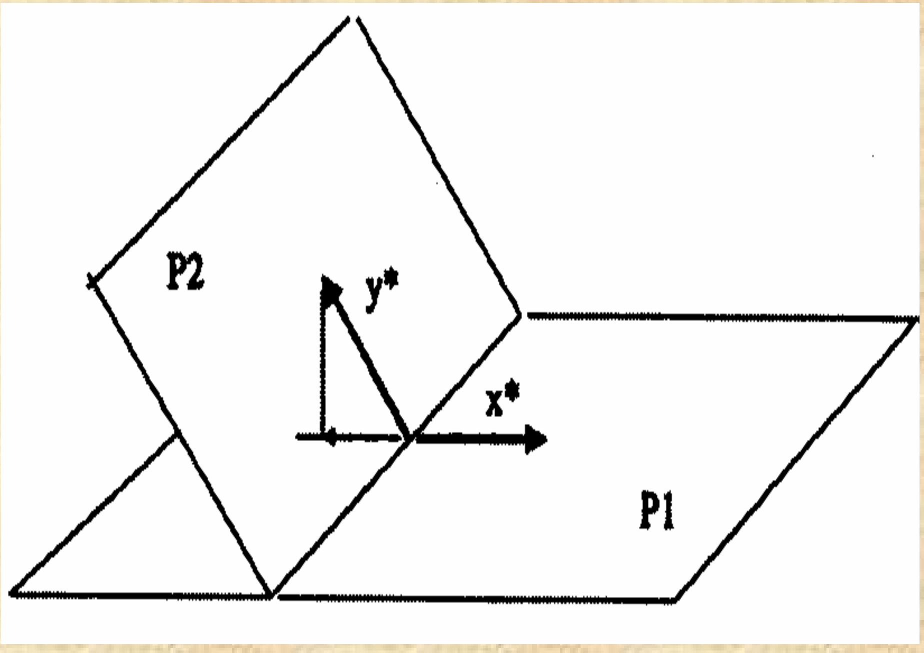
Una vez obtenidas las dos variables, es posible que esta primera relación entre las 2 v. indicadores explique completamente los dos conjuntos y no exista más relación entre ambas.

Si no es así, se puede buscar una 2ª v. Indicadora del primer conjunto incorrelada con la 1ª y que tenga correlación máxima con otra v. Indicadora del segundo conjunto.

CORRELACIÓN CANÓNICA

Las correlaciones canónicas representan relaciones de dependencia entre los subespacios generados por los dos conjuntos de variables.

Los vectores **x*** e **y*** están lo más cerca posible (es decir, **x*** es colineal con la proyección de **y*** sobre P_1 y viceversa.



PROPIEDADES DE LAS V. C.

- Son indicadores de los dos conjuntos de variables que tienen máxima correlación.
- Los coeficientes de la v. c. son los vectores propios ligados al mismo valor propio de **A** y **B**.
- Si $\alpha_{\mathbf{i}}'\mathbf{x}$ es una v.c. también - $\alpha_{\mathbf{i}}'\mathbf{x}$ lo es.
- Las correlaciones canónicas son el cuadrado del coeficiente de correlación entre las dos v. c.

PROPIEDADES DE LAS V. C.

- Las correlaciones canónicas λ_i^2 son invariantes ante transformaciones lineales de las variables.

- La primera correlación canónica λ_1^2 nunca es menor que el mayor coeficiente de correlación simple al cuadrado, entre una variable de cada conjunto.

PROPIEDADES DE LAS V. C.

- La correlación canónica λ_i^2 es el coeficiente de determinación en una regresión múltiple con respecto a la variable **y*** y variables explicativas las **x** (idem para la regresión de **x*** con las **y**).

- Las v. c. son los predictores óptimos en el sentido de minimizar $E(\|\mathbf{x}^*-\mathbf{y}^*\|^2)$

CONTRASTES

- Contrastamos que los dos conjuntos de variables están incorrelados que equivale a decir que todas las correlaciones canónicas son nulas .

Bajo las hipótesis de que **X** e **Y** siguen distribuciones normales de media **0**:

H_0: $V_{12}=0$

H_1: $V_{12}\neq 0$

$$\lambda = -m\sum_{j=1}^{r}\log(1-\lambda_j^2) \approx \chi^2_{pq},$$

$$m = n - \frac{1}{2(p+q+3)} \text{ y } r = \min(p,q)$$

CONTRASTES

- Podemos contrastar por otra parte que los *s* 1os coeficientes de la correlación canónica son ≠0 y los restantes nulos.

H_0: $\lambda_i>0$ i=1,...,s; y $\lambda_{s+1}=...=\lambda_r=0$

H_1: $\lambda_i>0$ i=1,...,s; y al menos un $\lambda_j>0$ con j=s+1,...,r

$$\lambda = -m\sum_{j=s+1}^{r}\log(1-\lambda_j^2) \approx \chi^2_{(p-s)(q-s)}, \text{ con } m = n - \frac{1}{2(p+q+3)}$$

Ejemplo:

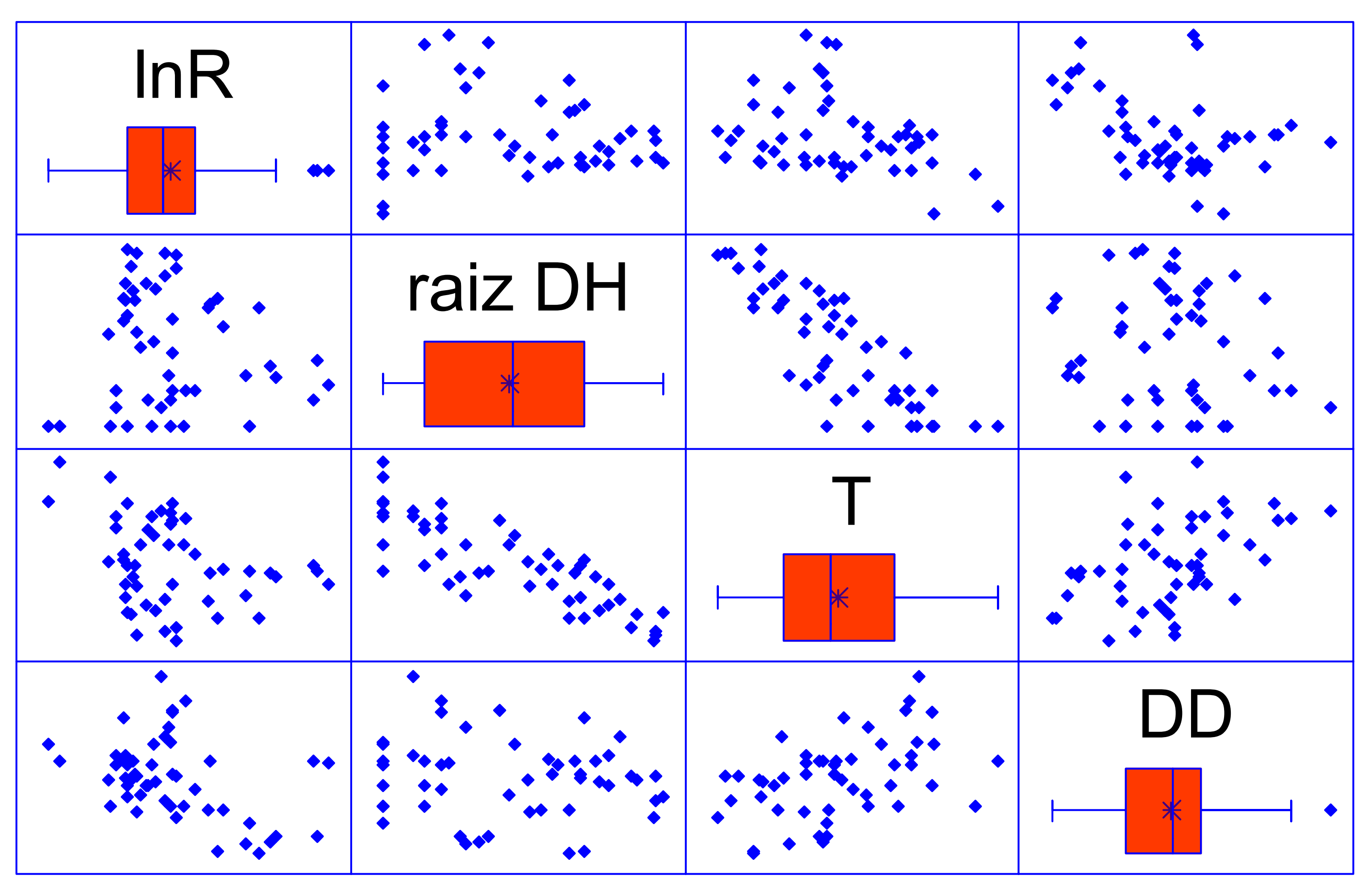
Se han medido en 50 poblaciones españolas variables climáticas que se agrupan en variables relacionadas con la pluviometría (conjunto 1) y las relacionadas con las temperaturas (conjunto 2):

Conjunto 1: Precipitaciones anuales y número de días de niebla.

Conjunto 2: Temperatura media anual y número de días despejados al año.

Se comprobó que las variables de cada conjunto estaban incorreladas y se transformaron para obtener normalidad.

Ejemplo:



Ejemplo:

Correlaciones Canónicas

Nº	Valor propio	Correlación Canónica	Lambda Wilks	Chi	g.d.l.	P-Valor
1	0,863625	0,929315	0,119216	98,8971	4	0,0000
2	0,12582	0,354711	0,87418	6,25279	1	0,0124

Correlación alta

Correlaciones significativas

Ejemplo:

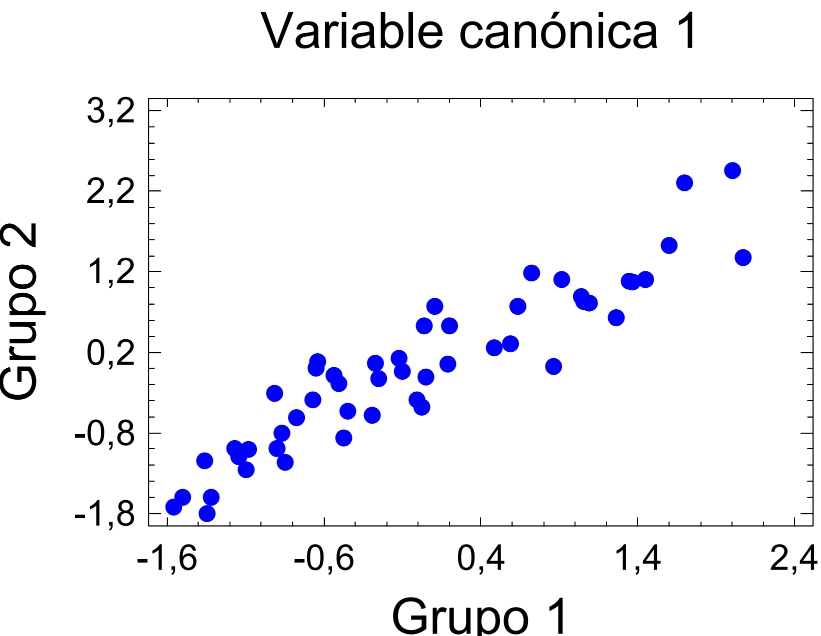
Coeficientes de las variables del primer grupo para las dos vvcc

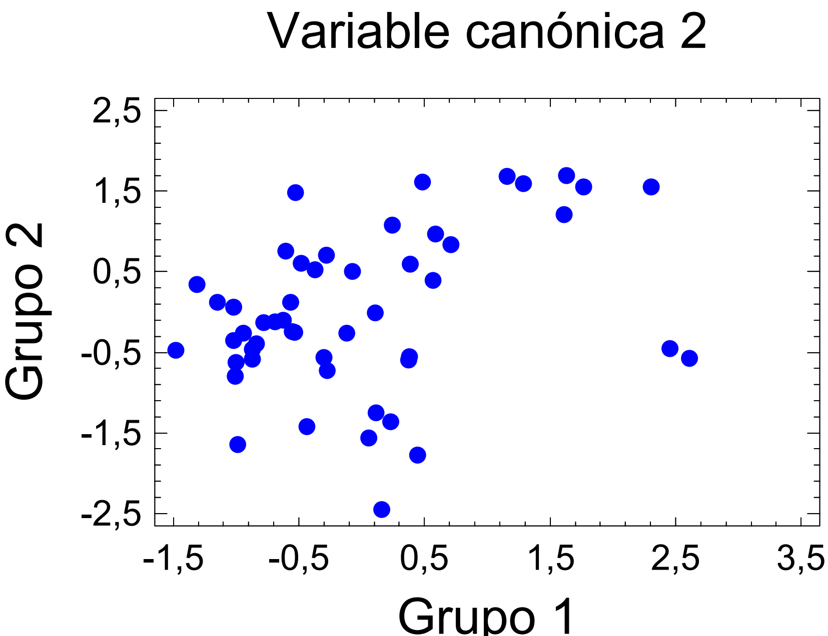
lnR	-0,365702	0,93433
raiz DH	-0,961068	-0,288198

Coeficientes de las variables del segundo grupo para las vvcc

T	1,06964	0,333641
DD	-0,184722	-1,10514

Ejemplo:


Variable canónica 1
Grupo 2
Grupo 1
3,2
2,2
1,2
0,2
-0,8
-1,8
-1,6
-0,6
0,4
1,4
2,4


Variable canónica 2
Grupo 2
Grupo 1
2,5
1,5
0,5
-0,5
-1,5
-2,5
-1,5
-0,5
0,5
1,5
2,5
3,5

Ejemplo:

Las nuevas variables que representan temperatura y pluviometría están muy correlacionadas linealmente según la VC1:

$$\underbrace{-0{,}365702*\ln R - 0{,}961068*\text{raiz } DH}_{\boxed{u_1}}=\underbrace{1{,}06964*T - 0{,}184722*DD}_{\boxed{v_1}}$$

Otra combinación de estas variables correlacionadas más ligeramente es la VC2:

$$\underbrace{0{,}93433*\ln R - 0{,}288198*\text{raiz } DH}_{\boxed{u_2}}=\underbrace{0{,}333641*T - 1{,}10514*DD}_{\boxed{v_2}}$$

RELACIÓN CON OTRAS TÉCNICAS

- La regresión es un caso particular de las correlaciones canónicas: *si cada uno de los conjuntos tiene una sola variable* ($r^2=\lambda^2$)

- La correlación canónica entre X (v. explicativas) y las G variables y_i conduce a los mismos resultados que el análisis discriminante si:

$$y_i = \begin{cases} 1 & \text{si } i \in \text{grupo } i \\ 0 & \text{en el resto de casos} \end{cases}$$

ANÁLISIS CANÓNICO ASIMÉTRICO

El objetivo del estudio es prever cada uno de los componentes de Y mediante las variables X.

La correlación canónica no resuelve el problema ya que puede existir alta correlación entre **x*** e **y*** y baja entre cada **y** con las **x***.

➡Construyendo q ecuaciones distintas de regresión.

➡Buscando una única α'X que tenga buenas propiedades para predecir las Y$\Rightarrow$A. C. Asimétrico.

ANÁLISIS CANÓNICO ASIMÉTRICO

Si las variables originales están estandarizadas el coeficiente de redundancia se define como:

$$\mathrm{CR}(y/\mathbf{x}'\boldsymbol{\alpha}) = \frac{1}{q}\boldsymbol{\alpha}'\mathbf{R}_{\mathbf{yx}}\mathbf{R}_{\mathbf{xy}}\boldsymbol{\alpha}$$

La medida de la correlación del conjunto de las r combinaciones lineales $\mathbf{x}'\alpha_{\mathbf{1}},\ldots,\mathbf{x}'\alpha_{\mathbf{r}}$ es la redundancia total: $R(y/x) = \sum_{i=1}^{r} \mathrm{CR}(y/\mathbf{x}'\boldsymbol{\alpha}_{\mathbf{i}}) \neq R(x/y)$

A. C. ASIMÉTRICO

Para encontrar la combinación lineal $\mathbf{x}'\alpha$ con máxima correlación con cada variable y_i individualmente de manera que la suma de correlaciones al cuadrado entre $\mathbf{x}'\alpha$ y las y hay que maximizar $\alpha'\mathbf{R}_{\mathbf{xy}}\mathbf{R}_{\mathbf{yx}}\alpha$ con la restricción: $\alpha'\mathbf{R}_{\mathbf{xx}}\alpha=1$

Por tanto, α es el vector propio de la matriz:

$$\mathbf{H} = \mathbf{R}_{\mathbf{xx}}^{-1}\mathbf{R}_{\mathbf{xy}}\mathbf{R}_{\mathbf{yx}}$$

A. C. ASIMÉTRICO

- Como en A. C. Simétrico podemos buscar una 2ª variable canónica asimétrica, ortogonal a la primera y con máxima correlación con la v. endógena.

- Este mismo análisis puede hacerse para explicar las **X** con las **Y**, pero el problema no es simétrico.

- R(**y**/**x**) no tiene en cuenta las correlaciones entre las variables **y** $\Rightarrow$ no es una medida multivariante de la dependencia entre los conjuntos.

ULTIMAS APLICACIONES

Ictiología (2004):

Estudio sobre asociaciones entre variables de tipo morfológico de las sp. de rayas de Bahía Almejas y la composición de sus dietas.

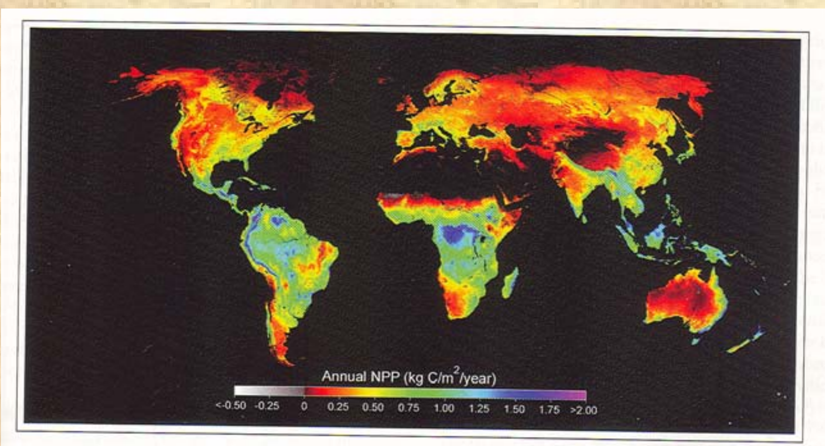
Ecosistemas (2004):

Estudio de relaciones entre el número de Zebrasoma flavescens y las características de los arrecifes de coral de Hawai.

ULTIMAS APLICACIONES

Edafología (2005):

Estudio sobre asociaciones entre características físicas del suelo de Gana y el uso de éste: agrícola, cultivo forestal, vegetación natural...



Teledetección (2005):

Estudio de relaciones entre v. espectrales y nodos y entre v. temporales y la longitudes de onda fijas.