

Análisis de correspondencias de tablas de contingencia estructuradas

CAMPO ELÍAS PARDO
Universidad Nacional de Colombia - Bogotá
Departamento de Estadística.
e-mail: cepardot@unal.edu.co

1. Introducción

Una tabla de contingencia (TC) se obtiene al clasificar dos variables cualitativas. El conjunto de las filas representan a las categorías de una variable y el conjunto de las columnas, a las categorías de la otra variable. El valor de una celda representa al número de *unidades estadísticas* = *individuos*, que poseen simultáneamente las categorías de la fila y de la columna que se cruzan. Al dividir la tabla sobre su total se obtiene la tabla de frecuencias relativas asociada a la TC, esta representa la distribución conjunta de probabilidad bivariada de la población en estudio, según las dos variables cualitativas que dan origen a la tabla. La suma de las filas (marginal fila), representa la distribución de probabilidad según la variable fila y la suma de las columnas (marginal columna), la distribución de probabilidad asociada a la variable columna. La división de cada una de las filas (respectivamente, columnas) por su marginal origina las distribuciones condicionales o perfiles fila (columna). El análisis de correspondencias simples (ACS) permite la descripción de una TC, cumpliendo con tres objetivos: 1) obtener y comparar una tipología de los perfiles fila; 2) obtener y comparar una tipología de los perfiles columnas y 3) estudiar las asociaciones entre las dos variables.

En aplicaciones más complejas las filas y las columnas pueden estar estructuradas mediante sendas particiones, definidas por otras variables cualitativas. En este documento se utiliza como ejemplo una tabla faunística, que cruza especies, en filas, con sitios \times fechas en columnas. La estructura de las filas esta dada por los grupos taxonómicos de pertenencia de cada una de las especies y la de las columnas por las fechas en que se hacen los conteos de las especies. Sobre este tipo de tabla son posibles varios análisis de correspondencias (AC), teniendo en cuenta o no las estructuras de partición. La presentación algunos de tales métodos es el objetivo de este cursillo.

En §2 se hace un repaso del análisis en componentes principales (ACP), método sobre el cual se construye el análisis de correspondencias (AC). En §3 se hace una descripción del ejemplo *Ardèche*, en §4 se presenta un resumen del ACS. Luego de resumir en §5 el AC con respecto a un modelo, se presenta en §6 el AC intra-tablas, que tienen en cuenta una de las dos particiones y finalmente en §7 el análisis de correspondencias interno (ACI), que tiene en cuenta simultáneamente las dos particiones. La sección 8 contiene una serie de talleres para ejecutar los métodos utilizando el lenguaje estadístico R (R Development Core Team 2004).

La simbología que se adopta en estas notas, es usual en muchos textos y es la siguiente: las letras mayúsculas en negrilla hacen referencia a matrices (\mathbf{A}), la minúsculas en negrilla a vectores (\mathbf{a}), las letras mayúsculas y minúsculas en itálica a variables (escalares) (A, a). En el caso de conjuntos

se utiliza la misma letra mayúscula para indicar al conjunto y a su cardinalidad.

Para ejecutar los métodos se utiliza el modulo ADE-4: análisis de datos ecológicos y ambientales (en francés, “environnementales”) con procedimientos exploratorios euclidianos (Thioulouse, Dufour & Chessel 2004). Para complementar este cursillo se recomienda consultar los documentos: Ortiz & Pardo (2004), Cabarcas & Pardo (2001) y Montenegro & Pardo (1996).

2. Análisis en componentes principales (ACP)

En esta sección se hace una introducción sencilla del ACP, recurriendo a un ejemplo muy pequeño. Para un estudio más detallado, se pueden consultar, entre otros, Lebart, Morineau & Piron (1995), Lebart, Morineau & Warwick (1984) y Escofier & Pagès (1992).

2.1. Ejemplo café

El café tostado comercial puede estar contaminado con otros granos de menor costo, lo que altera las propiedades de las tasas preparadas. Se realiza un experimento, haciendo preparaciones de café excelso (sin contaminar) y contaminado con maíz y cebada, al 20 % y al 40 %; además se considera el efecto del tiempo de tostación, 8 minutos (tostación clara) y 10 minutos (tostación oscura). El experimento consta de 10 tratamientos y sobre las tasas de café, preparadas con cada una de las mezclas, se miden propiedades químicas, físicas y sensoriales (Duarte, Suarez, Moreno & Ortiz 1996). En este ejemplo se utilizan solamente las variables físicas: *Color*, *DA*: Densidad aparente, *EA*: extracto acuoso (contenido de sólidos solubles). Los valores obtenidos para los 10 tratamientos se muestran en la tabla 1.

Tabla 1: Características físicas de los cafés

Cafés	Color	DA	EA
ExcelsoClaro	298.0	385.1	25.0
Claro40Maiz	361.0	481.3	41.0
Claro40Cebad	321.0	422.6	40.0
Claro20Maiz	335.0	444.3	33.0
Claro20Cebad	314.0	368.7	32.0
ExcelsoOscur	186.0	346.6	28.0
Oscuro40Maiz	278.0	422.6	43.0
Oscuro40Ceba	238.0	403.0	42.0
Oscuro20Maiz	226.0	368.7	36.0
Oscuro20Ceba	210.0	368.7	35.0

2.2. El ACP paso a paso

El ACP, desde el punto de vista geométrico, primero hace una transformación adecuada de la tabla de datos, para representarla como dos nubes de puntos en dos espacios: el de los individuos y el de las variables. El parecido estadístico entre individuos se traduce en una distancia geométrica y la correlación entre variables en el ángulo entre los vectores que la representan. Las nubes de puntos son abstractas y para visualizarlas se proyectan sobre las rectas y planos que conserven la mayor parte posible de la variabilidad. En esta sección se ilustra tal procedimiento.

2.2.1. Transformaciones de la tabla de datos

La matriz \mathbf{Y} , de dimensión $I \times K$, que ingresa a un procedimiento de ACP, representa los datos de un conjunto de I individuos, descrito por K variables continuas (ver figura 1). La escala de las mediciones influye en el análisis, por lo que generalmente, los datos se suelen estandarizar antes de realizar un ACP, al que se denomina ACP normado. Para repasar el procedimiento del ACP se presenta el caso más sencillo, donde los individuos tienen el mismo peso en el análisis y las variables la misma influencia.

Centrado en la nube de individuos N_I . En la nube de los I individuos en \mathbb{R}^K los ejes son las variables y las coordenadas de cada punto-individuo son los valores de las variables que asume. En la figura 1 se muestra en 3D la nube de individuos del ejemplo café.

El centro de gravedad de la nube de individuos es: $\mathbf{g} = \frac{1}{I} \mathbf{X}' \mathbf{1}_I$, donde $\mathbf{1}_I$ es un vector de I unos.

Al restar a cada vector individuo el centro de gravedad \mathbf{g} se obtiene la matriz centrada \mathbf{Y}_C : $\mathbf{Y}_C = \mathbf{Y} - \mathbf{1}_I \mathbf{g}'$. Al representar \mathbf{Y}_C en \mathbb{R}^K el origen se traslada al centro de gravedad de N_I (figura 1).

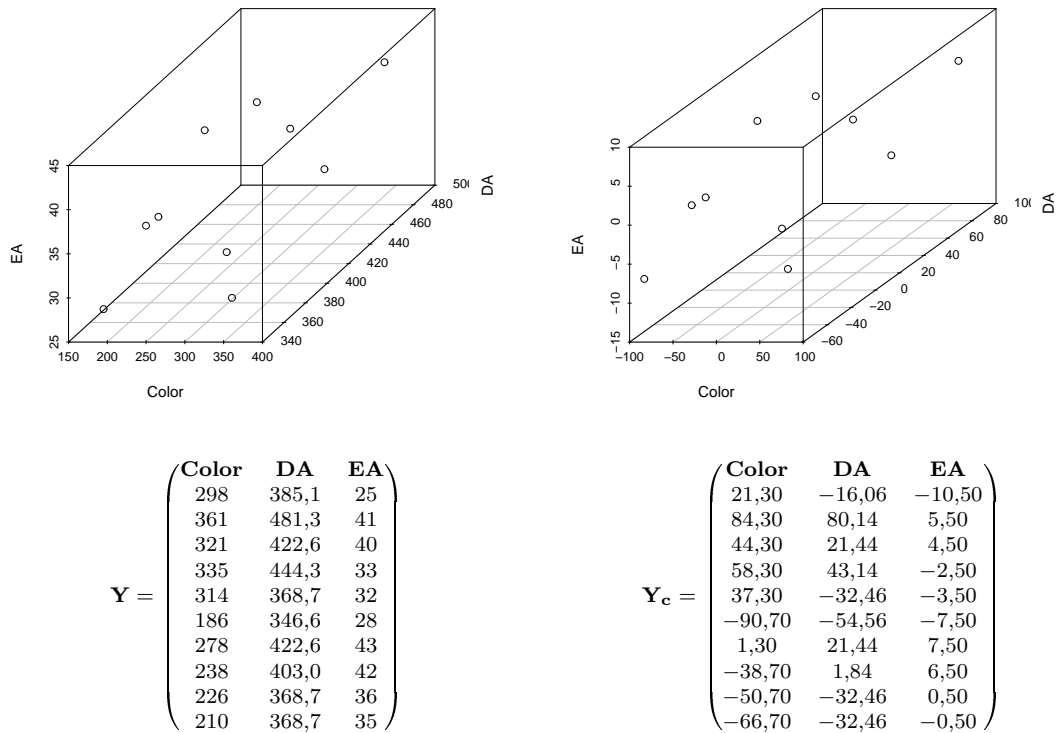


Figura 1: Representación de la tabla de datos de café en 3D

Reducción de la nube de puntos (cambio de escala). La matriz de varianzas y covarianzas, \mathbf{V} asociada a la tabla \mathbf{Y} es: $\mathbf{V} = \frac{1}{I} \mathbf{Y}'_C \mathbf{Y}_C$. En la diagonal de \mathbf{V} se tienen las varianzas, de modo que la suma de varianzas es igual a $\text{traza}(\mathbf{V})$.

La matriz normalizada \mathbf{X} (matriz de datos \mathbf{Y} centrada y reducida) tiene término general: $x_{ik} = \frac{y_{ik} - \bar{y}_k}{\sigma_k}$, donde \bar{y}_k y σ_k , son la media y la desviación estándar de la variable k . En términos matriciales \mathbf{X} se obtiene mediante: $\mathbf{X} = \mathbf{Y}_C \mathbf{D}_{1/\sigma}$, donde $\mathbf{D}_{1/\sigma} = \text{diag}(\frac{1}{\sigma_k})$. En la figura 2 se muestra \mathbf{X} para el ejemplo café. El valor que un individuo asume para una variable es la diferencia con respecto al promedio, pero ahora medida en el número de desviaciones estándar.

2.2.2. Nube de individuos estandarizados

Para el ACP normado del ejemplo las coordenadas de la nube de individuos son las 10 filas de la matriz \mathbf{X} (figura 2).

La matriz de correlaciones de las variables iniciales registradas en la tabla \mathbf{Y} , es la matriz de varianzas y covarianzas de \mathbf{X} : $\mathbf{V} = \frac{1}{I} \mathbf{X}' \mathbf{X}$. Para el ejemplo la matriz de correlaciones se puede ver en la figura 4.

Inercia de la nube de individuos N_I . La noción física de momento de inercia alrededor de un punto se utiliza como medida de dispersión de la nube de puntos, alrededor de su centro de gravedad y se denomina *inercia*.

Si cada individuo i se dota del peso p_i , la inercia de la nube es: $In(N_I) = \sum_{i=1}^I p_i d^2(\mathbf{i}, \mathbf{g})$

En el caso del ACP normado, que se está desarrollando, el peso de cada individuo es el mismo $p_i = 1/I$ y el centro de gravedad coincide con el origen de la representación, entonces:

$$In(N_I) = \frac{1}{I} \sum_{i=1}^I \sum_{k=1}^K x_{ik}^2 = \frac{1}{I} \sum_{k=1}^K \sum_{i=1}^I x_{ik}^2 = \frac{1}{I} \sum_{k=1}^K \sigma_k^2 = \frac{1}{I} \text{traza}(\mathbf{X}' \mathbf{X}) = K$$

En el ejemplo la inercia total de nube es 3 (el número de variables) y está repartida equitativamente entre los tres ejes, es decir entre las tres variables.

Objetivo y solución del ACP en la nube de individuos. El objetivo geométrico del ACP es buscar un nuevo sistema de ejes de tal manera que la mayoría de la inercia se concentre en los primeros ejes. Es decir se trata de descomponer la inercia de la nube de puntos en ejes ortogonales ordenados, de tal manera que en el primer eje este la mayor inercia posible, en el segundo la mayor inercia residual posible, etc. Este es un problema de optimización que se encuentra resuelto en los textos de análisis multivariado de datos. La solución se tiene con los vectores y valores propios de la matriz de varianzas y covarianzas, que en el caso del ACP normado coincide con la matriz de correlaciones $\frac{1}{I} \mathbf{X}' \mathbf{X}$. El primer vector propio normado \mathbf{u}_1 , asociado al valor propio más grande λ_1 , da la dirección del primer eje. El valor inercia recogida sobre este primer eje es el valor propio λ_1 .

El significado de graficar a la nube de puntos N_I sobre el nuevo sistema de ejes, es decir sobre la base de los vectores propios normados: $\{\mathbf{u}_1, \dots, \mathbf{u}_s, \dots, \mathbf{u}_K\}$, es hacer una rotación del sistema de ejes. Como los ejes están ordenados por la cantidad de inercia que recogen, las proyecciones sobre los primeros ejes son las “mejores”, en el sentido de conservar la mayor inercia posible. El primer plano factorial, conformado por los ejes 1 y 2, generados por u_1 y u_2 , respectivamente, retiene una inercia igual a $\lambda_1 + \lambda_2$.

Una primera decisión que debe tomar el analista, usuario del ACP, es determinar el número de ejes S que va a retener para la interpretación, es decir la dimensión S del subespacio de proyección

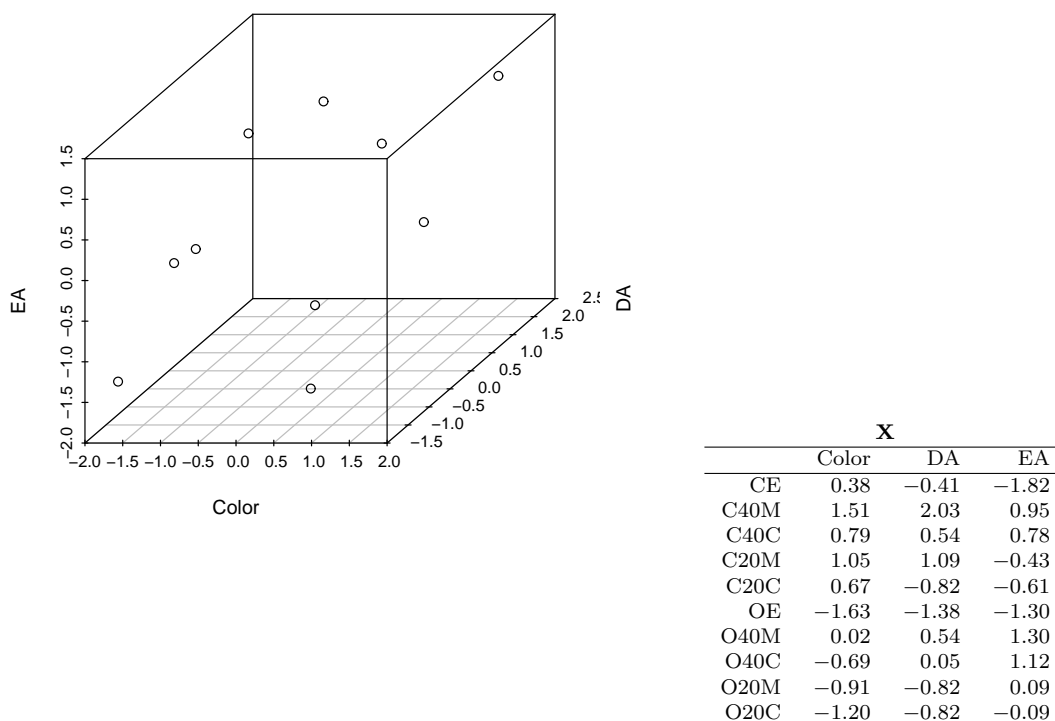


Figura 2: Nube de individuos asociada a los datos estandarizados de café

de la nube de puntos. El gráfico de valores propios y la fracción de inercia retenida en el subespacio son una buena guía para tomar esa decisión.

En el ejemplo café, los valores propios son: $\lambda_1 = 2,067$, $\lambda_2 = 0,822$ y $\lambda_3 = 0,111$ y los vectores propios: $\mathbf{u}_1 = (-0,58 \ -0,67 \ -0,46)'$, $\mathbf{u}_2 = (-0,57 \ -0,07 \ 0,82)'$ y $\mathbf{u}_3 = (0,58 \ -0,74 \ 0,35)'$ o sus opuestos, ya que hay dos soluciones \mathbf{u}_1 y $-\mathbf{u}_1$, etc.

El primer plano factorial recoge $2,067+0,822 = 2,889$ de inercia que es el $2,889*100/3 = 96,3\%$, es decir que casi nada se pierde al leer el primer plano factorial, en lugar de la representación en 3D, pero en cambio la lectura se hace mucho más fácil.

El sentido de los ejes no tiene significado. Cada eje factorial puede ser generado por uno de los dos vectores propios normados que definen su dirección \mathbf{u}_s o $-\mathbf{u}_s$. Escoger uno u otro significa cambiar el signo de las coordenadas, es decir un rotación de 180 grados. Esto implica que para un mismo análisis se pueden tener planos rotados, según el paquete y el procedimiento utilizado. Para comparar planos de diferentes análisis, que aparecen rotados, se pueden cambiar los signos de las coordenadas de manera conveniente, para graficarlos de nuevo.

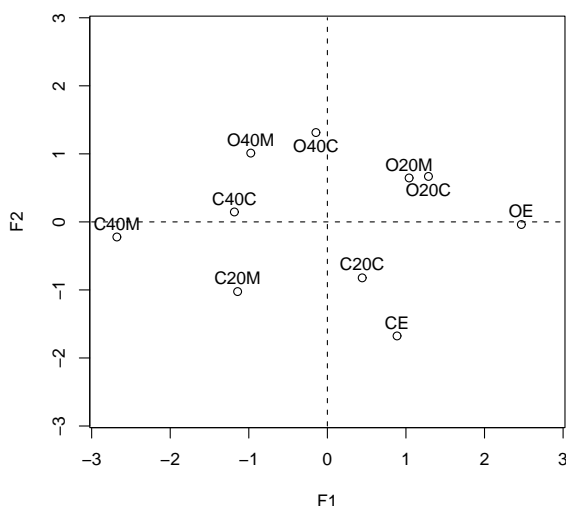
Gráficas y ayudas a la interpretación El primer plano factorial se construye buscando las coordenadas de los individuos sobre los ejes 1 y 2. El vector de todas las coordenadas sobre el eje 1 se nota F_1 y es: $F_1 = \mathbf{X}u_1$, si se arreglan los vectores propios como columnas en una matriz \mathbf{U} , la tabla de las coordenadas sobre los nuevos ejes es $\mathbf{F} = \mathbf{X}\mathbf{U}$. El primer plano factorial del

ACP normado del ejemplo café se muestran en la figura 3, donde se incluyen los valores de las coordenadas y ayudas a la interpretación.

La interpretación se hace teniendo en cuenta los vectores propios: Al lado negativo del primer eje se sitúan los cafés con mayores valores en las tres variables, mientras que al lado positivo del segundo eje los de mayor valor en *EA* y al lado negativo los de mayor valor en *Color*.

Un plano factorial es una aproximación de la nube de puntos y como tal tendrá puntos bien representados, pero podrá contener puntos con mala calidad de proyección. Se utiliza el coseno al cuadrado que es, para un punto, el cuadrado de la relación entre la norma de la proyección sobre la norma en el espacio completo. La suma de los cosenos cuadrados sobre todos los ejes factoriales es 1. La distancia de un punto al origen, en el espacio completo, es un buen complemento en la lectura de los ejes factoriales, esta dada por la norma del vector-individuo.

Para conocer los individuos que más influyen sobre la dirección de un eje factorial se utiliza el cociente de la contribución a la inercia del individuo sobre la inercia total del eje (valor propio). En la figura 3 se encuentran los valores de las ayudas a la interpretación para el ejemplo café.



Coordenadas y ayudas a la interpretación								
	d^2	Coordenadas		Contribución		Cosenos ²		Plano
		F1	F2	F1	F2	F1	F2	
CE	3.61	0.89	-1.68	3.80	34.19	21.80	77.89	99.69
C40M	7.31	-2.68	-0.22	34.72	0.60	98.20	0.67	98.87
C40C	1.53	-1.18	0.15	6.78	0.26	91.45	1.39	92.84
C20M	2.47	-1.14	-1.02	6.31	12.77	52.75	42.41	95.16
C20C	1.49	0.44	-0.82	0.95	8.23	13.20	45.42	58.62
OE	6.24	2.47	-0.04	29.49	0.02	97.70	0.02	97.73
O40M	1.98	-0.98	1.01	4.60	12.44	48.09	51.72	99.81
O40C	1.75	-0.15	1.31	0.10	20.98	1.21	98.65	99.86
O20M	1.51	1.04	0.65	5.24	5.07	71.67	27.58	99.25
O20C	2.12	1.29	0.67	8.00	5.43	78.22	21.10	99.32

Figura 3: Primer plano factorial del ACP normado del ejemplo café

2.2.3. La nube de variables N_K

La nube de variables está constituida por K puntos en \mathbb{R}^I , las coordenadas de cada punto son las columnas de la matriz \mathbf{X} . El procedimiento es análogo al de la nube de individuos pero su interpretación es diferente.

Significado del centrado. Sea \mathbf{Y}_k el vector columna asociado de la variable k , es decir la columna k de \mathbf{Y} , \mathbf{Y}_{C_k} y \mathbf{X}_k las columnas k de las matrices \mathbf{Y}_C y \mathbf{X} , respectivamente. El centrado de un vector \mathbf{Y}_k se logra mediante: $\mathbf{Y}_{C_k} = \mathbf{Y}_k - \bar{Y}_k \mathbf{1}_I$, donde \bar{Y}_k es la media de la variable k . El vector $\mathbf{1}_I$ es la primera bisectriz de \mathbb{R}^I y el vector $\bar{Y}_k \mathbf{1}_I$ es la homotecia (amplificación o reducción) de la primera bisectriz por la media de la variable k . Entonces todas las variables están representadas como puntos sobre la primera bisectriz (recta generada por el vector $\mathbf{1}_I$, que es un subespacio de dimensión 1). Finalmente, el vector diferencia, $\mathbf{Y}_k - \bar{Y}_k \mathbf{1}_I$, es una proyección de \mathbf{Y}_k sobre el subespacio ortogonal a la primera bisectriz. Esto significa que las variables centradas están todas en el subespacio ortogonal a la primera bisectriz y que en el proceso de centrado se pierde una dimensión.

Significado del reducido. Una columna k de \mathbf{X} se obtiene mediante $\mathbf{X}_k = \frac{1}{\sigma_k} \mathbf{Y}_{C_k}$. Esto hace que la norma al cuadrado del vector \mathbf{X}_k sea:

$$\mathbf{X}'_k \mathbf{X}_k = \frac{1}{\sigma_k^2} \mathbf{Y}'_{C_k} \mathbf{Y}_{C_k} = \frac{1}{\sigma_k^2} \frac{1}{I} (\mathbf{Y}_k - \bar{Y}_k \mathbf{1}_I)' (\mathbf{Y}_k - \bar{Y}_k \mathbf{1}_I) = \frac{1}{\sigma_k^2} \sum_{i=1}^I (y_{ik} - \bar{y}_k)^2 = I$$

Cambiando el producto escalar canónico, asociado a la matriz \mathbf{I}_{d_I} , matriz identidad de dimensión I , por el productor escalar asociado a $\frac{1}{I} \mathbf{I}_{d_I}$, se logra que la norma de cada una de las variables sea 1:

$$\langle \mathbf{X}'_k, \mathbf{X}_k \rangle_{\frac{1}{I} \mathbf{I}_{d_I}} = \mathbf{X}'_k \frac{1}{I} \mathbf{I}_{d_I} \mathbf{X}_k = \frac{1}{I} \mathbf{X}'_k \mathbf{X}_k = 1$$

y entonces, la representación de las variables estandarizadas se puede ver como flechas que terminan en el cascarón hiperesférico de radio 1 y centro en el origen.

Significado del ángulo entre dos flechas variables. El coseno del ángulo entre dos vectores variables en el ACP normado es: $\langle \mathbf{X}_k, \mathbf{X}_{k'} \rangle_{\frac{1}{I} \mathbf{I}_{d_I}}$ y es exactamente igual al coeficiente de correlación entre las dos variables k y k' , ya que $\|\mathbf{X}_k\|_{\frac{1}{I} \mathbf{I}_{d_I}} = 1$ (norma=desviación estándar = 1).

Entonces el espacio de las variables de un ACP normado es una representación de la matriz de correlaciones.

Objetivo del ACP y solución. Para poder leer la nube de variables se buscan los ejes de proyección que conservan lo más posible las normas de las variables originales. La solución se consigue con los valores y vectores propios de la matriz $\frac{1}{I} \mathbf{X} \mathbf{X}'$, esta matriz tiene K valores propios que son iguales a los valores propios de $\frac{1}{I} \mathbf{X}' \mathbf{X}$ y los restantes $I - K$ valores propios son cero. Cada valor propio λ_s tiene asociado el vector propio \mathbf{v}_s

Círculo de correlaciones y ayudas a la interpretación. Un plano factorial de las variables se denomina círculo de correlaciones (figura 4), ya que es la proyección de la *hiperesfera* de correlaciones. La longitud de la proyección, sin error, de un vector-variable es 1. La calidad de la representación en el plano se observa visualmente al dibujar un círculo de radio uno en el plano factorial.

El vector de coordenadas sobre un eje s , \mathbf{G}_s , se obtiene mediante $\frac{1}{I} \mathbf{X}'_k \mathbf{v}_k$ y coincide con la correlación entre la variable k y el eje s . Los valores de la variable sintética representada por el eje s están en el vector F_s , que contiene las coordenadas de los individuos sobre el eje factorial s .

La contribución de cada variable a un eje s sirve para seleccionar las variables que dan más significado al eje.

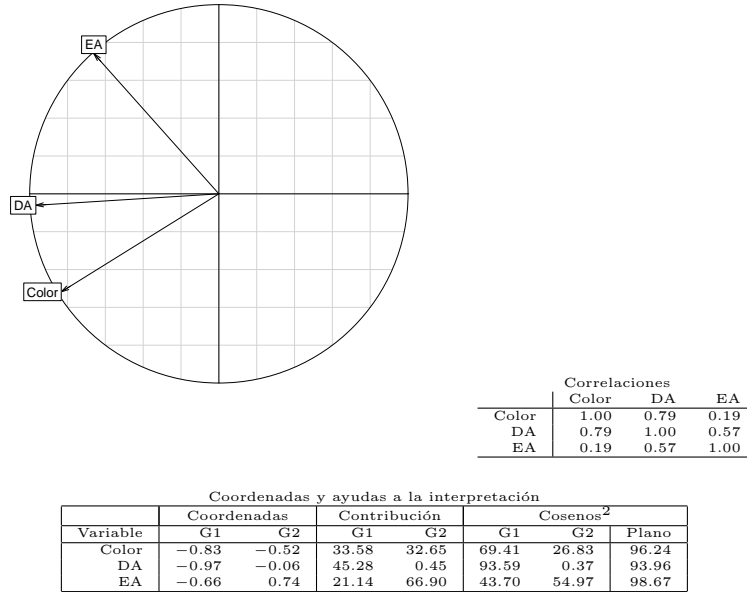


Figura 4: Círculo de correlaciones del ejemplo café

2.3. ACP generalizado $ACP(\mathbf{X}, \mathbf{M}, \mathbf{D})$

Los métodos en ejes principales (ACP, ACS, AC múltiples, AC intra, ACI, etc) se pueden ver como un ACP de una matriz \mathbf{X} , que contiene los datos a analizar; definiendo las matrices de métrica y pesos en los dos espacios: de individuos y de variables. Se nota \mathbf{M} a la matriz de métrica en el espacio de los individuos y de pesos en el de las variables y \mathbf{D} a la matriz de pesos de los individuos y de métrica en el espacio de las variables. Las matrices \mathbf{M} y \mathbf{D} generalmente son diagonales. La matriz \mathbf{X} esta centrada con los pesos dados en \mathbf{D} , es decir $\mathbf{g} = \mathbf{X}'\mathbf{D}\mathbf{1}_I = \mathbf{0}$, siendo $\mathbf{1}_I$ un vector columna de I unos. Con la definición de la tripleta $(\mathbf{X}, \mathbf{M}, \mathbf{D})$, el análisis queda completamente determinado y se nota $ACP(\mathbf{X}, \mathbf{M}, \mathbf{D})$.

El ACP normado de la sección anterior se puede ver como $ACP(\mathbf{Y}_c, \text{diag}(\frac{1}{\sigma_k^2}), \frac{1}{I}\mathbf{I}_{d_I})$, la matriz a analizar es la matriz de datos centrados, la métrica $\text{diag}(\frac{1}{\sigma_k^2})$ es la matriz diagonal de las inversas de las varianzas y $\frac{1}{I}\mathbf{I}_{d_I}$ es la matriz de pesos, donde \mathbf{I}_{d_I} es la matriz identidad de dimensión I . El ACP normado también es el $ACP(\mathbf{X}, \mathbf{I}_{d_K}, \frac{1}{I}\mathbf{I}_{d_I})$ o el $ACP(\mathbf{Z}, \mathbf{I}_{d_K}, \mathbf{I}_{d_I})$, donde \mathbf{Z} es la matriz de datos normalizados dividida por \sqrt{I} , es decir que se diferencia de \mathbf{X} por la homotecia (multiplicación por) $\frac{1}{\sqrt{I}}$.

2.3.1. Análisis en \mathbb{R}^K

Con una métrica diagonal \mathbf{M} , la distancia entre dos filas i y l es:

$$d_{il}^2 = \sum_{k=1}^K m_k (x_{ik} - x_{lk})^2 \quad (1)$$

La matriz diagonal de pesos \mathbf{D} está involucrada en el cálculo del centro de gravedad y de la inercia. La inercia total es la suma ponderada de las distancias al cuadrado de los puntos-fila al centro de gravedad de la nube:

$$I = \sum_{i=1}^I p_i d^2(\mathbf{i}, \mathbf{0}) \quad (2)$$

Cada punto-fila contribuye a la inercia con el producto de su peso por el cuadrado de la distancia al centro de gravedad, el cual coincide con el origen de la representación. La contribución de un punto-fila a la inercia de la nube proyectada sobre un eje cualquiera es el peso de la fila por la coordenada al cuadrado sobre ese eje.

Lo que busca el ACP es encontrar un sistema de ejes \mathbf{u} , \mathbf{M} -ortonormales (normado $\mathbf{u}'\mathbf{M}\mathbf{u} = 1$ y ortogonal $\mathbf{u}'_s\mathbf{M}\mathbf{u}_t = 0, s \neq t$) de \mathbf{M} -inercia máxima.

Sea $F = \mathbf{X}\mathbf{M}\mathbf{u}$ el vector de las coordenadas sobre el eje definido por \mathbf{u} , entonces la inercia proyectada sobre el eje es $\sum p_i F^2 = \mathbf{u}'\mathbf{M}\mathbf{X}'\mathbf{D}\mathbf{X}\mathbf{M}\mathbf{u}$. Maximizar esta cantidad es encontrar la dirección de mayor inercia proyectada, es decir la mejor proyección condicionada a los pesos de los puntos-fila. La solución es el vector propio \mathbf{u}_1 correspondiente al mayor valor propio λ_1 de la matriz $\mathbf{X}'\mathbf{D}\mathbf{X}\mathbf{M}$.

Se procede luego a encontrar una segunda dirección \mathbf{M} ortogonal a la primera que maximiza la inercia proyectada sobre ese eje, luego una tercera, \mathbf{M} ortogonal a las dos primeras. Encontrándose un nuevo sistema de ejes \mathbf{u}_s , con coordenadas F_s . Los \mathbf{u}_s son los vectores propios asociados a los valores propios de la matriz $\mathbf{X}'\mathbf{D}\mathbf{X}\mathbf{M}$, ordenados de mayor a menor.

La inercia proyectada sobre un subespacio de dimensión S es la suma de las inercias proyectadas sobre los ejes ortogonales que lo conforman. La aproximación de la proyección de la nube sobre ese subespacio se suele medir con el cociente de inercias:

$$\sum_{s=1}^S \lambda_s / \sum_{s=1}^K \lambda_s \quad (3)$$

2.3.2. Análisis en \mathbb{R}^I y dualidad

En la nube de los K puntos columna en \mathbb{R}^I , las distancias están definidas por la matriz de métrica \mathbf{D} y los pesos están en la matriz \mathbf{M} . Se busca la dirección \mathbf{v} sobre la cual la $\sum_{k=1}^K m_k G_k^2$ sea máxima, donde \mathbf{G}_k es la proyección de la variable \mathbf{k} sobre la dirección \mathbf{v} . El vector de todas las proyecciones sobre \mathbf{v} es $\mathbf{X}'\mathbf{D}\mathbf{v}$ y la cantidad a maximizar es: $\mathbf{v}'\mathbf{D}\mathbf{X}\mathbf{M}\mathbf{X}'\mathbf{D}\mathbf{v}$ sujeta a la restricción $\mathbf{v}'\mathbf{D}\mathbf{v} = 1$. Sin embargo no es necesario realizar la diagonalización de la matriz $\mathbf{X}\mathbf{M}\mathbf{X}'\mathbf{D}$ pues los ejes factoriales de los dos espacios están relacionados. Sea μ_s un valor propio de la matriz $\mathbf{X}\mathbf{M}\mathbf{X}'\mathbf{D}$. Las relaciones entre los dos espacios son:

- los valores propios diferentes de cero son iguales en los dos espacios: $\mu_s = \lambda_s$
- el vector de coordenadas F_s sobre \mathbf{u}_s es un vector propio asociado a μ_s

- la \mathbf{M} -norma al cuadrado del vector F_s es λ_s
- el vector \mathbf{v}_s es igual a $\frac{1}{\sqrt{\lambda_s}}F_s$
- el vector de coordenadas G_s es igual a $\sqrt{\lambda_s}\mathbf{u}_s$

2.3.3. Ayudas a la interpretación de las gráficas.

Calidad de la representación. Un indicador de esa calidad es el coseno que es una relación entre las magnitudes de la proyección y del vector original. Sin embargo se utiliza coseno cuadrado, ya que para un punto, la suma de los cosenos cuadrados sobre todos los ejes factoriales es 1 y su coseno cuadrado en un subespacio se obtiene sumando sus cosenos cuadrados sobre los ejes factoriales que generan al subespacio. Sobre un eje s el coseno cuadrado de un punto-fila i es:

$$Cos_s^2(i) = \frac{F_s^2(i)}{\|\mathbf{i}\|^2} \quad (4)$$

Las coordenadas de un vector fila \mathbf{i} están en la fila i de la matriz \mathbf{X} . El valor del coseno al cuadrado coincide con la relación de contribuciones del individuo i a la inercia: *contribución a la inercia proyectada sobre el eje s /contribución a la inercia total* y se llama también contribución relativa.

Contribución absoluta. Otro aspecto que ayuda a la interpretación es identificar las filas que más estén contribuyendo a la inercia de una eje s . Este indicador se obtiene dividiendo la inercia proyectada del punto-fila i sobre la inercia total del eje, que es igual al valor propio. Se suele expresar en porcentaje y recibe el nombre de contribución absoluta de la fila i al eje s :

$$Con_s(i) = \frac{p_i F_s^2(i)}{\lambda_s} \quad (5)$$

Para las columnas se definen los mismos indicadores.

2.3.4. Elementos suplementarios o ilustrativos

Sobre los subespacios factoriales se pueden proyectar elementos que no participaron en el análisis, ya sean filas o columnas. También es posible proyectar filas *artificiales*, por ejemplo las filas promedio de grupos construidos a partir de variables nominales. La fórmulas de proyección son las mismas de los elementos activos. Es válido calcular la calidad de la representación de los elementos suplementarios. La contribución a la formación de los ejes es obviamente nula.

2.3.5. Fórmulas del ACP($\mathbf{X}, \mathbf{M}, \mathbf{D}$)

Un método específico, en ejes principales, queda completamente determinado definiendo las matrices: \mathbf{X} , que se obtiene de los datos mediante la transformación adecuada; \mathbf{M} , métrica en el espacio de las filas y pesos en el espacio de las columnas y \mathbf{D} , pesos en el espacio de las filas y métrica en el espacio de las columnas. Entonces las fórmulas del método específico se obtienen reemplazando en las fórmulas ACP($\mathbf{X}, \mathbf{M}, \mathbf{D}$) (tabla 2).

Tabla 2: Fórmulas del ACP($\mathbf{X}, \mathbf{M}, \mathbf{D}$)

Espacio	\mathbb{R}^K	\mathbb{R}^I
Nube	N_I	N_K
Coordenadas	filas de \mathbf{X}	columnas de \mathbf{X}
Pesos	diagonal de \mathbf{D}	diagonal de \mathbf{M}
Métrica	\mathbf{M}	\mathbf{D}
Inercia	$traza(\mathbf{X}'\mathbf{D}\mathbf{X}\mathbf{M})$	$traza(\mathbf{X}\mathbf{M}\mathbf{X}'\mathbf{D})$
Valor propio	λ_s	λ_s
Vector propio	\mathbf{u}_s	\mathbf{v}_s
Coordenadas factoriales	$F_s = \mathbf{X}\mathbf{M}\mathbf{u}_s = \sqrt{\lambda_s}\mathbf{v}_s$	$G_s = \mathbf{X}'\mathbf{D}\mathbf{v}_s = \sqrt{\lambda_s}\mathbf{u}_s$
Fórmulas de transición	$F_s = \frac{1}{\sqrt{\lambda_s}}\mathbf{X}\mathbf{M}\mathbf{G}_s$ $F_s(i) = \frac{1}{\sqrt{\lambda_s}} \sum_{k=1}^K x_{ik}m_k G_s(k)$	$G_s = \frac{1}{\sqrt{\lambda_s}}\mathbf{X}'\mathbf{D}\mathbf{F}_s$ $G_s(k) = \frac{1}{\sqrt{\lambda_s}} \sum_{i=1}^I x_{ik}d_i F_s(i)$

3. Ejemplo *Ardèche*

El ejemplo *Ardèche* (Cazes, Chessel & Dolédec 1988) consiste en una tabla faunística (tabla 3), que cruza 43 especies (filas) y fechas \times sitios (35 columnas, cada una correspondiendo a un muestreo en un sitio y una fecha). Estos datos están disponibles en el paquete ADE-4 de R. Las 43 especies están distribuidas en 4 grupos taxonómicos (*Ephemeroptera*, *Plecoptera*, *Coleoptera*, *Trichoptera*), que inducen la partición de las filas (4 grupos). Los seis sitios (*A*, *B*, *C*, *D*, *E*, *F*), se observan en seis fechas (*jul82*, *ago82*, *nov82*, *feb83*, *abr83*, *jul83*), seleccionadas de estaciones diferentes, sin embargo, la observación del sitio *F* en *jul82*, es un dato faltante. En las columnas se considera la partición inducida por las diferentes fechas (6 grupos).

El análisis del ejemplo *Ardèche* está orientado por las siguientes preguntas:

1. En qué momento y para cual grupo taxonómico se obtiene una capacidad óptima para separar los sitios según su contenido faunístico?.
2. Es la tipología estable con las fechas?

4. Análisis de correspondencias simples (ACS)

4.1. Pequeño ejemplo y notación

4.1.1. La tabla de contingencia

La tabla 3 vista globalmente es una TC que cruza especies y sitios \times fechas. Cada subtabla (l, j) es, por si sola, una tabla de contingencia que cruza las especies de un grupo taxonómico con los sitios, solamente en una fecha. Veamos por ejemplo la subtabla (2, 5), que corresponde a las tres especies del grupo *Plecoptera*, encontradas en las seis estaciones en abril de 1983 (ver tabla 4).

4.1.2. Tabla de frecuencias relativas

La tabla de frecuencias relativas se nota \mathbf{F} de término general f_{ik} , el término general de su marginal fila se nota f_i . y el de su marginal columna f_k . Se definen $\mathbf{D}_I = \text{diag}(f_i)$ y $\mathbf{D}_K = \text{diag}(f_k)$ (ver tabla 4).

Tabla 3: Tabla faunística del ejemplo *ardeche*

	1: jul82					2: aug82					3: nov82					4: feb83					5: apr83					6: jul83																
	A	B	C	D	E	A	B	C	D	E	A	B	C	D	E	A	B	C	D	E	A	B	C	D	E	A	B	C	D	E												
Eph1	9	9	6	6	9	7	3	6	6	8	5	11	8	9	10	10	8	6	8	8	11	7	4	9	8	8	8	7	7	1	5	4	0	4	6							
Eph2	7	4	6	6	8	8	4	7	7	8	1	4	1	2	3	1	1	9	8	9	5	8	3	6	5	7	6	5	5	9	9	9	10	9								
Eph3	6	2	5	4	7	5	0	1	5	7	0	4	0	0	0	2	1	4	3	3	2	6	3	5	4	2	2	2	5	7	5	6	6	5	2							
Eph4	0	0	1	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	2	1	1	10	10	9	9	8	8	6	9	9	7	2								
Eph5	5	0	5	5	4	0	0	0	3	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0						
Eph6	0	2	0	3	4	4	0	0	5	5	0	1	2	4	2	1	6	6	6	5	6	2	3	1	6	2	4	3	3	3	0	6	1	0	8							
Eph7	6	7	0	4	5	1	1	0	6	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0						
Eph8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	6	7	1	6	6	8	0	7	4	3	0							
Eph9	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0						
Eph10	0	3	0	0	4	0	3	0	0	5	1	4	0	0	0	0	0	0	0	1	0	0	0	2	2	1	0	0	0	0	0	0	0	0	0	0	0					
Eph11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	6	0	0	0	4	3	2	6	5	3	5	7	0						
Ple1	3	9	7	3	6	1	6	5	5	8	1	1	0	0	0	0	0	1	0	1	0	0	0	7	7	5	0	4	4	7	1	0	7	6	4	8						
Ple2	3	7	5	4	5	0	2	1	5	5	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0					
Ple3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	3	3	0	3	3	4	4	0	0	3	0	0	0	0	0	0	0	0	0	0				
Col1	4	5	3	0	6	7	5	4	0	6	0	3	0	2	0	0	0	4	4	5	3	0	7	5	5	0	3	4	6	5	7	4	6	0	7	7	4	6				
Col2	8	9	8	5	8	9	8	6	9	5	6	0	1	4	5	4	7	5	8	4	2	7	8	6	7	5	6	5	5	6	6	7	7	5	6	6	7	7				
Col3	9	9	5	9	8	9	9	6	10	9	8	7	1	5	7	5	4	5	9	6	9	5	5	7	8	5	7	5	5	6	8	5	8	4	6	5	8	4	6			
Col4	9	7	4	6	5	9	7	4	4	6	7	4	0	4	4	4	0	4	6	8	6	1	0	6	8	5	0	2	6	5	6	7	4	5	6	7	4	5				
Col5	5	3	0	0	4	4	1	0	0	3	0	1	0	0	0	1	0	0	3	2	1	1	0	2	3	1	1	2	0	3	3	4	2	4	1	1	1					
Col6	8	4	2	9	6	8	5	4	9	7	7	8	3	4	7	4	6	5	4	6	8	4	3	6	4	6	7	5	4	0	0	2	4	0	0	2	4	0				
Col7	7	7	5	4	6	8	5	4	8	7	0	4	1	1	4	1	1	2	0	3	6	0	0	3	3	3	6	1	1	7	7	7	7	5	2	7	7	5				
Col8	7	0	0	6	3	8	2	1	5	6	1	1	0	1	3	0	0	0	3	7	0	0	0	3	0	0	6	0	2	5	1	7	7	0	2	5	1	7	0			
Col9	1	0	0	4	2	1	1	0	0	1	0	0	0	0	0	0	0	0	1	3	0	0	0	0	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0			
Col10	2	2	0	0	1	1	5	1	1	0	0	3	0	5	0	0	0	0	0	1	5	0	0	1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
Col11	5	0	0	4	0	5	0	3	1	0	1	0	1	0	1	0	0	0	0	1	5	0	0	1	0	1	3	1	1	2	0	0	4	0	0	0	0	0				
Col12	3	1	0	4	0	5	0	0	1	0	1	0	3	0	0	3	0	0	1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
Col13	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	3	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0			
Tri1	8	3	6	7	8	10	1	3	8	8	2	8	4	5	7	8	6	2	2	5	6	5	2	5	1	4	5	4	4	7	4	8	8	5	6	7	4	8	8	5	6	
Tri2	9	5	8	7	8	8	0	2	6	6	1	8	0	3	5	5	4	4	0	2	7	3	1	6	4	5	7	5	5	4	4	6	6	5	3	0	7	1	3	7	4	0
Tri3	8	1	5	3	4	9	1	1	5	0	0	6	0	2	5	3	0	2	0	2	5	0	0	5	3	4	6	3	0	7	1	3	7	4	0	0	0	2	0	0	0	
Tri4	0	2	1	0	0	3	1	0	0	2	0	3	1	2	2	0	0	0	3	7	0	0	0	5	0	4	7	1	1	0	0	0	2	0	0	0	0	0	0	0	0	
Tri5	6	3	4	4	6	6	0	5	4	0	0	6	0	4	2	0	2	2	4	5	1	0	5	3	3	4	1	0	5	4	6	6	2	2	4	6	6	2	2			
Tri6	4	7	1	6	5	2	8	1	0	4	0	2	1	3	3	3	0	0	5	0	4	0	0	1	4	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Tri7	4	5	2	3	4	1	1	1	2	5	1	0	0	0	0	0	0	0	0	0	0	0	0	1	4	1	0	0	1	0	3	3	0	0	1	0	0	0	0	0		
Tri8	0	6	0	2	0	1	1	0	0	1	0	0	0	1	0	0	0	1	1	3	5	1	0	0	3	0	5	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
Tri9	2	0	3	4	4	3	1	4	0	3	2	2	0	0	2	4	0	0	0	7	0	0	0	1	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Tri10	0	2	0	0	8	3	0	0	5	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	1	0	0	1	0	1	0	1	0	1	0	1	0	
Tri11	1	0	0	0	3	0	0	0	4	0	4	2	0	0	2	0	0	0	0	6	0	0	0	1	1	1	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Tri12	2	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	3	4	3	5	0	1	2	4	1	3	1	1	2	4	1	3	1	1	0	0	0	0	0	0	
Tri13	0	0	0	0	4	0	0	0	1	2	0	3	2	0	0	0	0	3	0	0	2	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Tri14	0	0	0	1	4	0	0	0	0	4	0	0	0	0	3	0	0	0	5	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Tri15	1	0	1	2	5	6	0	0	0	5	0	2	0	1	2	1	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Tri16	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	3	0	3	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0

En el ejemplo la marginal fila representa a la distribución de frecuencias de todos los insectos del grupo *Plecoptera* según su especie (3 especies) y la marginal columna es la distribución de los insectos recolectados en las seis estaciones.

4.1.3. Tabla de perfiles fila

Para cada especie: *Ple1*, *Ple2* y *Ple3* se tiene una distribución de frecuencias entre los 6 sitios, que se denomina distribución condicional o perfil fila. Se obtiene al dividir cada celda de la respectiva fila por la suma de la fila, en la TC o en la tabla de frecuencias **F**. La marginal columna de la tabla **F** se constituye en la distribución promedio de los perfiles fila y es la distribución de todos los 53 insectos en los 6 sitios, sin importar su especie.

Un perfil fila *i* se nota: $\{\frac{f_{ik}}{f_i}; k = 1, \dots, K\}$. El conjunto de los perfiles fila se notan y calculan mediante $\mathbf{D}_I^{-1}\mathbf{F}$ (ver tabla 4 y figura 5).

4.1.4. Tabla de perfiles columna

Cada sitio tiene su distribución de insectos en las 3 especies (condicionales o perfiles columna). La distribución marginal fila es la distribución de todos los 53 insectos, en las 3 especies, sin importar el sitio.

Un perfil columna *k* se nota: $\{\frac{f_{ik}}{f_{.k}}; i = 1, \dots, I\}$. El conjunto de los perfiles columna se calcula mediante $\mathbf{F}'\mathbf{D}_K^{-1}$ (ver tabla 4 y figura 5).

4.1.5. El modelo de independencia

Si se supone que no hay asociación, es decir que hay independencia estadística entre las variables fila y columna, el modelo es $a_{ik} = f_i \cdot f_k$, término general de la tabla de independencia **A** (ver tabla 4). En esta tabla las distribuciones condicionales fila (respectivamente, columna) son todas iguales a la marginal de las columnas (filas) de tabla **F**. Las desviaciones al modelo de independencia son **F-A**(ver tabla 4 y figura 5).

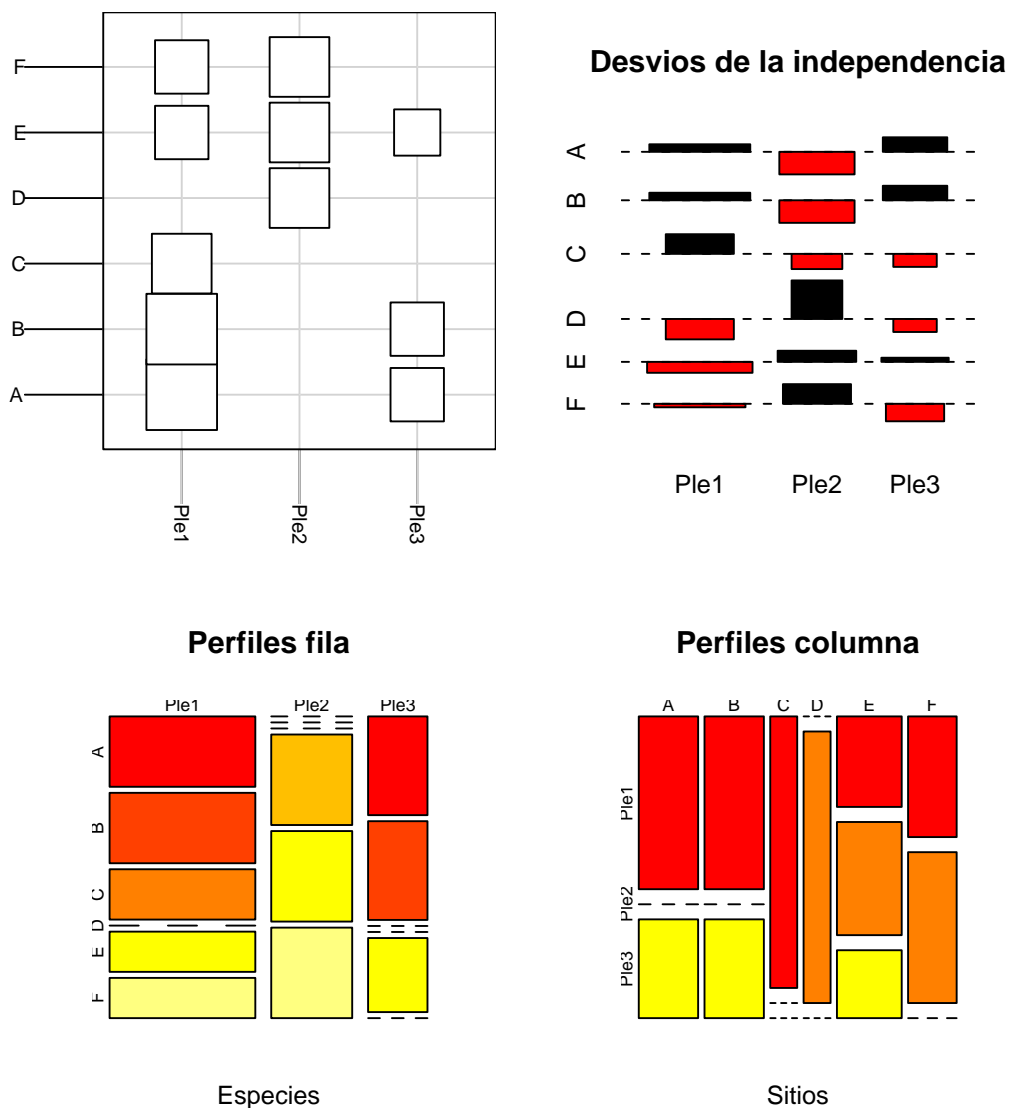


Figura 5: Gráficas asociadas a la tabla (2, 5)

Tabla 4: Subtabla de contingencia (2, 5) y tablas derivadas

Tabla de contingencia

	A	B	C	D	E	F	Mar.
Ple1	7	7	5	0	4	4	27
Ple2	0	0	0	5	5	5	15
Ple3	4	4	0	0	3	0	11
Mar.	11	11	5	5	12	9	53

Frecuencias relativas

	A	B	C	D	E	F	Mar.
Ple1	13.2	13.2	9.4	0.00	7.6	7.6	50.9
Ple2	0.0	0.0	0.0	9.43	9.4	9.4	28.3
Ple3	7.6	7.6	0.0	0.00	5.7	0.0	20.8
Mar.	20.8	20.8	9.4	9.4	22.6	17.0	100.0

Notación

$$\mathbf{F} = \begin{pmatrix} f_{11} & f_{12} & f_{13} & f_{14} & f_{15} & f_{16} \\ f_{21} & f_{22} & f_{23} & f_{24} & f_{25} & f_{26} \\ f_{31} & f_{32} & f_{33} & f_{34} & f_{35} & f_{36} \end{pmatrix}$$

Marginales fila

$$\mathbf{D_I} = \begin{pmatrix} f_{1\cdot} & 0 & 0 \\ 0 & f_{2\cdot} & 0 \\ 0 & 0 & f_{3\cdot} \end{pmatrix} \quad \mathbf{D_I} = \begin{pmatrix} 50,9 & 0 & 0 \\ 0 & 28,3 & 0 \\ 0 & 0 & 20,8 \end{pmatrix}$$

Marginales columna

$$\mathbf{D_K} = \begin{pmatrix} f_{\cdot 1} & 0 & 0 & 0 & 0 & 0 \\ 0 & f_{\cdot 2} & 0 & 0 & 0 & 0 \\ 0 & 0 & f_{\cdot 3} & 0 & 0 & 0 \\ 0 & 0 & 0 & f_{\cdot 4} & 0 & 0 \\ 0 & 0 & 0 & 0 & f_{\cdot 5} & 0 \\ 0 & 0 & 0 & 0 & 0 & f_{\cdot 6} \end{pmatrix} \quad \mathbf{D_K} = \begin{pmatrix} 20,8 & 0 & 0 & 0 & 0 & 0 \\ 0 & 20,8 & 0 & 0 & 0 & 0 \\ 0 & 0 & 9,4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 9,4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 22,6 & 0 \\ 0 & 0 & 0 & 0 & 0 & 17,0 \end{pmatrix}$$

Perfiles fila

$$\mathbf{D^{-1}F} = \begin{pmatrix} \frac{f_{11}}{f_{1\cdot}} & \frac{f_{12}}{f_{1\cdot}} & \frac{f_{13}}{f_{1\cdot}} & \frac{f_{14}}{f_{1\cdot}} & \frac{f_{15}}{f_{1\cdot}} & \frac{f_{16}}{f_{1\cdot}} \\ \frac{f_{21}}{f_{2\cdot}} & \frac{f_{22}}{f_{2\cdot}} & \frac{f_{23}}{f_{2\cdot}} & \frac{f_{24}}{f_{2\cdot}} & \frac{f_{25}}{f_{2\cdot}} & \frac{f_{26}}{f_{2\cdot}} \\ \frac{f_{31}}{f_{3\cdot}} & \frac{f_{32}}{f_{3\cdot}} & \frac{f_{33}}{f_{3\cdot}} & \frac{f_{34}}{f_{3\cdot}} & \frac{f_{35}}{f_{3\cdot}} & \frac{f_{36}}{f_{3\cdot}} \end{pmatrix}$$

	A	B	C	D	E	F	Suma
Ple1	25.9	25.9	18.5	0.0	14.8	14.8	100
Ple2	0.0	0.0	0.0	33.3	33.3	33.3	100
Ple3	36.4	36.4	0.0	0.0	27.3	0.0	100

Perfiles columna

$$\mathbf{F'D_K^{-1}} = \begin{pmatrix} \frac{f_{11}}{f_{\cdot 1}} & \frac{f_{12}}{f_{\cdot 2}} & \frac{f_{13}}{f_{\cdot 3}} & \frac{f_{14}}{f_{\cdot 4}} & \frac{f_{15}}{f_{\cdot 5}} & \frac{f_{16}}{f_{\cdot 6}} \\ \frac{f_{21}}{f_{\cdot 1}} & \frac{f_{22}}{f_{\cdot 2}} & \frac{f_{23}}{f_{\cdot 3}} & \frac{f_{24}}{f_{\cdot 4}} & \frac{f_{25}}{f_{\cdot 5}} & \frac{f_{26}}{f_{\cdot 6}} \\ \frac{f_{31}}{f_{\cdot 1}} & \frac{f_{32}}{f_{\cdot 2}} & \frac{f_{33}}{f_{\cdot 3}} & \frac{f_{34}}{f_{\cdot 4}} & \frac{f_{35}}{f_{\cdot 5}} & \frac{f_{36}}{f_{\cdot 6}} \end{pmatrix}$$

	A	B	C	D	E	F
Ple1	63.6	63.6	100.0	0.0	33.3	44.4
Ple2	0.0	0.0	0.0	100.0	41.7	55.7
Ple3	36.4	36.4	0.0	0.0	25.0	0.0
Suma	100.0	100.0	100.0	100.0	100.0	100.0

Tabla de independencia **A**

	A	B	C	D	E	F
Ple1	10.57	10.57	4.81	4.81	11.53	8.65
Ple2	5.87	5.87	2.67	2.67	6.41	4.81
Ple3	4.31	4.31	1.96	1.96	4.70	3.52

F-A

	A	B	C	D	E	F
Ple1	2.63	2.63	4.63	-4.81	-3.99	-1.10
Ple2	-5.87	-5.87	-2.67	6.76	3.03	4.63
Ple3	3.24	3.24	-1.96	-1.96	0.96	-3.52

4.2. El ACS como dos ACPs

Con el ACS se obtienen imágenes sintéticas que permiten las mismas lecturas de las gráficas de la figura 5. Sin embargo, para lograr una interpretación correcta, se requiere del conocimiento de la lógica del método y de sus propiedades. La presentación más común de los textos de *análisis*

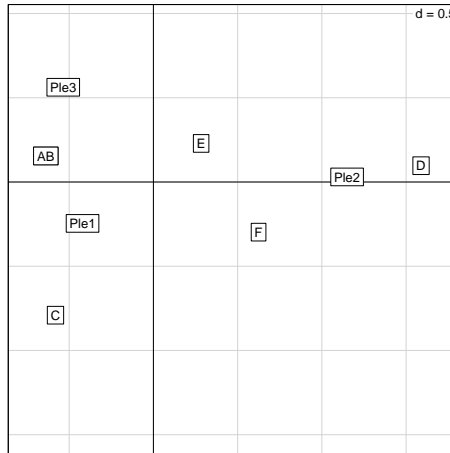


Figura 6: Plano factorial del ACS del ejemplo

multivariado de datos, es la de dos ACP inter-relacionados, la cual se resume a continuación.

4.2.1. ACP de los perfiles-fila

La tabla que se analiza es la de perfiles fila, es decir que el histograma que representa al perfil se ve como un punto en \mathbb{R}^K . La diferencia entre dos histogramas se traduce en una distancia entre los puntos que los representan. Los pesos de los puntos fila son la distribución marginal y se ordenan en la matriz diagonal \mathbf{D}_I , las distancias entre distribuciones condicionales se definen a partir del producto punto dado por la matriz \mathbf{D}_K^{-1} . Las imágenes para los perfiles fila son los planos factoriales derivados del ACP($\mathbf{D}_I^{-1}\mathbf{F}$, \mathbf{D}_K^{-1} , \mathbf{D}_I). La matriz $\mathbf{D}_I^{-1}\mathbf{F}$ no está centrada, pero su valor propio más grande de la matriz a diagonalizar ($\mathbf{F}'\mathbf{D}_I^{-1}\mathbf{F}\mathbf{D}_K^{-1}$, ver tabla 2) es 1 y el vector propio asociado es el centro de gravedad de la nube. De modo que lo que se hace, en lugar de centrar, es eliminar este valor propio y su vector propio. Partir del segundo vector propio de esta matriz es equivalente a centrar y se logra una simplificación de las formulas del ACS.

4.2.2. ACP de los perfiles-columna

Los histogramas de las distribuciones condicionales columna se representan como puntos en \mathbb{R}^I , a cada punto k se le asigna el peso $f_{.k}$. El análisis de los perfiles columna es el ACP($\mathbf{F}'\mathbf{D}_K^{-1}$, \mathbf{D}_I^{-1} , \mathbf{D}_K).

4.2.3. Representación simultánea

Los dos ACP están relacionados, debido a que los perfiles fila y columna se derivan de la misma matriz \mathbf{F} y la inversa de la matriz de pesos en un espacio es la métrica en el otro. Las relaciones de transición entre los dos espacios (§4.3) permiten la representación simultánea de los mapas factoriales.

En cada ACP del ACS, los mapas factoriales y sus ayudas a la interpretación son análogos a los de los individuos en el ACP clásico.

4.3. El ACS como un ACP(X,M,D)

El AC de la tabla \mathbf{F} también se obtiene mediante el ACP de la tabla \mathbf{X} cuyo término general está dado por (6), usando $\mathbf{D} = \mathbf{D}_I = \text{diag}(f_{i\cdot})$, como pesos de las filas y matriz de métrica en el espacio de las columnas, y $\mathbf{M} = \mathbf{D}_K = \text{diag}(f_{\cdot k})$, como pesos de las columnas y matriz de métrica en el espacio de las filas.

$$x_{ik} = \frac{f_{ik} - f_{i\cdot}f_{\cdot k}}{f_{i\cdot}f_{\cdot k}} \quad (6)$$

Todas las fórmulas del ACS se pueden derivar de las fórmulas correspondientes al ACP generalizado (ver tabla 2). La \mathbf{M} -distancia al cuadrado entre dos filas i y l y la \mathbf{D} -distancia al cuadrado entre las columnas k y q de \mathbf{X} son:

$$d^2(i, l) = \sum_{k=1}^K \frac{1}{f_{\cdot k}} \left(\frac{f_{ik}}{f_{i\cdot}} - \frac{f_{lk}}{f_{l\cdot}} \right)^2; \quad d^2(k, q) = \sum_{i=1}^I \frac{1}{f_{i\cdot}} \left(\frac{f_{ik}}{f_{\cdot j}} - \frac{f_{iq}}{f_{\cdot q}} \right)^2 \quad (7)$$

Las expresiones de (7) son las distancias ji-cuadrado entre los perfiles fila y columna, respectivamente, derivados de la tabla \mathbf{F} (ver tabla 4). Esta distancia tiene dos propiedades muy importantes para la interpretación de las salidas del ACS:

Equivalencia distribucional. El ACS no se modifica si se unen dos puntos que tienen el mismo perfil, por ejemplo los sitios A y B del ejemplo (ver figuras 5 y 6), el peso del punto colapsado es la suma de los pesos de los puntos que se unen. Esto permite unir filas o columnas con perfiles parecidos, para simplificar las tablas originales. Esta propiedad hace que el ACS sea robusto ante la “arbitrariedad” en la conformación de las categorías de una variable en un estudio.

Representación simultánea. En el ACS las relaciones de transición son:

$$F_s(i) = \frac{1}{\sqrt{\lambda_s}} \sum_{k=1}^K \frac{f_{ik}}{f_{i\cdot}} G_s(k); \quad G_s(k) = \frac{1}{\sqrt{\lambda_s}} \sum_{i=1}^I \frac{f_{ik}}{f_{\cdot k}} F_s(i) \quad (8)$$

Las relaciones de la expresión (8) además de hacer posible técnicamente la representación simultánea permiten su interpretación. Un punto fila se ubica en el promedio, ponderado por los valores de su perfil, de las coordenadas de todos los puntos columna, dilatado por el inverso de la norma del vector propio. Entonces el punto fila se ubica más cerca de los puntos de las columnas que más contribuyen a su perfil. La dilatación hace que la asociación más destacada sea también la más alejada. Por ejemplo la asociación entre la especie *Ple2* y el sitio D , se manifiesta con mayor intensidad en el sitio. De hecho en el sitio D solo se encontraron individuos de la especie *Ple2*, en cambio la especie *Ple2* se encontró en la misma proporción en los sitios D , E y F (ver figuras 5 y 6). Las filas o columnas correspondientes a las categorías de menor frecuencia son las más alejadas del origen de la representación.

4.4. ACS del ejemplo Ardèche

Sobre la tabla 3, de 43 filas por 35 columnas, se realiza el ACS clásico, el cual no tiene en cuenta la doble estructura de partición de la tabla. Sin embargo tal estructura emerge en los resultados pues está contenida en los datos. Los resultados que se muestran se obtienen directamente con las funciones *dudi.coa* y *inertia.dudi* del ADE-4.

4.4.1. Valores propios

Las dos nubes de puntos están en un subespacio de dimensión 34 ($\min(35, 43) - 1$), la lectura de la tabla y de la gráfica (figura 7) permite tomar la decisión de cuántos ejes analizar. En este ejemplo los dos primeros valores propios se destacan sobre los demás y retienen 34 % de la inercia, sin embargo también sobresalen los ejes 3 y 4. En los 4 primeros ejes se conserva el 52 % de la inercia. La forma más o menos homogénea del gráfico de valores propios a partir del 5 eje hace pensar que no será necesario analizar más ejes.

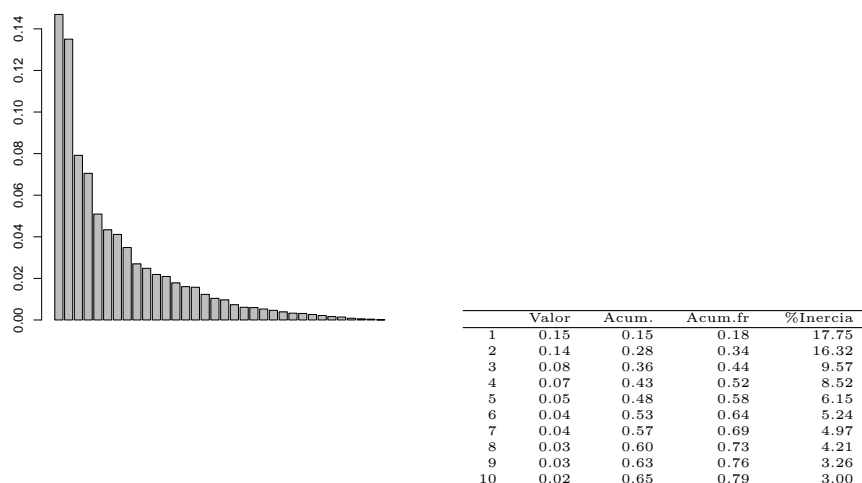


Figura 7: Valores propios de ACS de Ardèche

4.4.2. Planos factoriales

En la figura 8 se presenta el plano factorial 1-2 tres veces, para destacar diferentes aspectos, proyectando las particiones de fechas y estaciones como variables nominales ilustrativas. La inercia del plano 1-2 se debe principalmente a las fechas y la del plano 3-4 a las estaciones. La dispersión de las especies se da al interior de los grupos taxonómicos ya que los cuatro se ubican muy cerca al centro.

4.4.3. Ayudas a la interpretación

Las ayudas presentadas en la tabla 5 permiten controlar y complementar las lecturas de los planos factoriales. La contribución a la inercia del eje ayuda a encontrar un significado del eje. Por ejemplo, en el eje 3, $D5$, $D4$, $B1$ y $B4$ son las que más contribuyen (acumulan el 47.5 % de la inercia del eje), entonces el eje 3 se interpreta principalmente como la contraposición de $D4$ y $D5$ (lado negativo) y $B1$ y $B4$ (lado positivo). La calidad de la representación ayuda a evitar lecturas erróneas de puntos mal representados, por ejemplo $B2$, $C2$ y $D2$ en el plano 1-2.

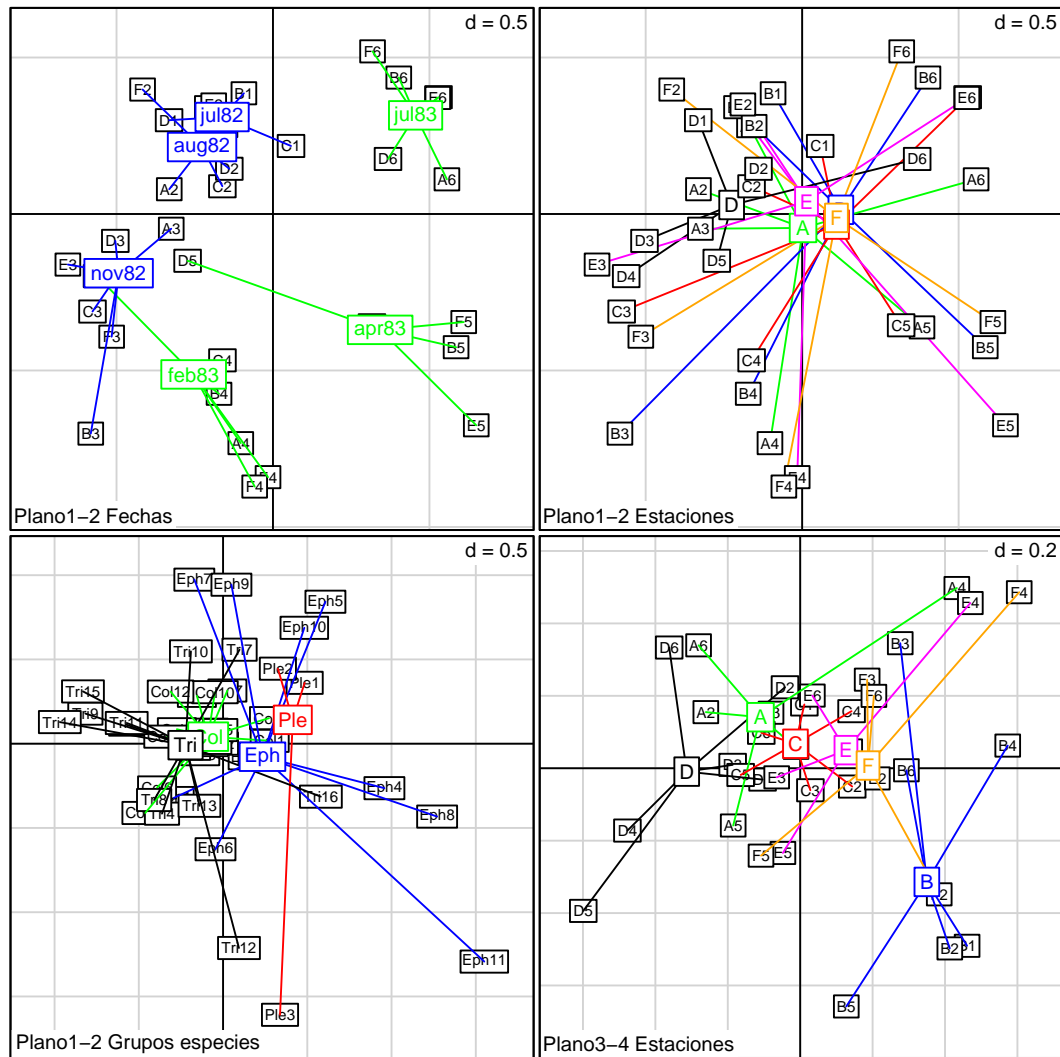


Figura 8: Planos 1-2 y 3-4 del ACS de Ardèche

5. AC con respecto a un modelo

Escofier (1984) generaliza el AC a un modelo cualquiera, donde los principios y las propiedades del AC también son aplicables. La tabla de frecuencias \mathbf{F} y el modelo \mathbf{A} tienen la misma dimensión $I \times K$. Las matrices de pesos y métricas son: $\mathbf{D} = \text{diag}(d_i)$ y $\mathbf{M} = \text{diag}(m_k)$. A continuación se resumen las propiedades de este AC, denominado también AC generalizado.

5.1. Las nubes de puntos

Al conjunto de filas I , se asocia una nube de puntos N_I en \mathbb{R}^K . Las coordenadas son las diferencias entre los perfiles fila de \mathbf{F} y los perfiles filas de \mathbf{A} . El espacio \mathbb{R}^K está dotado de la métrica χ^2 , la distancia al cuadrado entre dos puntos fila i y l es:

Tabla 5: Coordenadas y ayudas a la interpretación para las columnas del ACS de Ardèche

Col.	Coordenadas				con.tra	Contribución a la inercia del eje				
	Comp1	Comp2	Comp3	Comp4		Comp1	Comp2	Comp3	Comp4	remain
A1	-0.17	0.28	-0.01	0.06	2.00	A1	0.83	2.52	0.01	0.20
B1	-0.09	0.39	0.46	-0.49	3.76	B1	0.20	3.83	9.33	11.85
C1	0.06	0.22	0.01	0.18	1.54	C1	0.06	0.94	0.00	1.18
D1	-0.33	0.30	-0.10	-0.03	2.66	D1	2.61	2.31	0.47	0.05
E1	-0.21	0.34	0.14	0.05	3.11	E1	1.39	4.00	1.09	0.16
A2	-0.33	0.08	-0.26	0.15	2.38	A2	3.23	0.20	3.71	1.47
B2	-0.15	0.28	0.41	-0.50	5.20	B2	0.37	1.34	4.94	8.03
C2	-0.16	0.09	0.14	-0.05	1.53	C2	0.34	0.11	0.49	0.07
D2	-0.14	0.14	-0.04	0.22	2.21	D2	0.45	0.52	0.08	2.40
E2	-0.19	0.35	0.21	-0.03	3.11	E2	1.10	4.07	2.57	0.05
F2	-0.42	0.40	0.38	-0.35	4.12	F2	1.87	1.85	2.95	2.72
A3	-0.33	-0.05	-0.08	0.15	1.98	A3	2.19	0.05	0.25	1.02
B3	-0.58	-0.70	0.28	0.35	2.47	B3	1.56	2.47	0.65	1.15
C3	-0.58	-0.31	0.03	-0.06	1.63	C3	3.01	0.96	0.01	0.07
D3	-0.50	-0.09	-0.19	0.01	2.41	D3	4.39	0.14	1.12	0.00
E3	-0.66	-0.16	-0.06	-0.03	2.61	E3	5.20	0.34	0.09	0.02
F3	-0.52	-0.39	0.18	0.24	2.81	F3	2.09	1.32	0.50	0.98
A4	-0.10	-0.73	0.43	0.50	3.29	A4	0.14	7.51	4.47	6.65
B4	-0.17	-0.57	0.57	0.06	2.81	B4	0.43	5.18	8.79	0.12
C4	-0.16	-0.47	0.14	0.16	2.41	C4	0.48	4.39	0.69	0.96
D4	-0.56	-0.20	-0.48	-0.17	4.45	D4	9.12	1.24	12.19	1.78
E4	-0.02	-0.84	0.47	0.46	3.22	E4	0.00	9.61	5.12	5.42
F4	-0.06	-0.87	0.60	0.49	2.17	F4	0.02	4.76	3.92	2.84
A5	0.38	-0.36	-0.19	-0.16	2.38	A5	3.72	3.63	1.61	1.32
B5	0.59	-0.43	0.13	-0.66	5.20	B5	8.10	4.66	0.71	21.33
C5	0.32	-0.36	-0.16	-0.02	2.59	C5	2.08	2.86	1.04	0.01
D5	-0.27	-0.15	-0.60	-0.39	4.23	D5	1.92	0.63	17.16	8.33
E5	0.65	-0.68	-0.05	-0.23	3.80	E5	7.30	8.58	0.07	1.97
F5	0.61	-0.35	-0.11	-0.24	2.81	F5	6.00	2.09	0.35	1.96
A6	0.56	0.11	-0.28	0.34	2.91	A6	7.49	0.32	3.54	5.82
B6	0.40	0.44	0.29	-0.01	2.59	B6	3.51	4.49	3.47	0.00
C6	0.53	0.37	-0.11	0.10	2.72	C6	7.58	4.11	0.58	0.54
D6	0.37	0.18	-0.36	0.33	2.55	D6	3.72	0.92	6.68	6.37
E6	0.53	0.37	0.03	0.20	3.37	E6	5.81	3.13	0.04	1.76
F6	0.32	0.52	0.20	0.20	3.10	F6	1.69	4.91	1.29	1.39

$$d^2(i, l) = \sum_k \left\{ \frac{f_{ik} - a_{ik}}{d_i} - \frac{f_{lk} - a_{lk}}{d_l} \right\}^2 \frac{1}{m_k} \tag{9}$$

Dos puntos i y l están próximos si las desviaciones entre la fila de la tabla y la del modelo, ponderadas por \mathbf{D} se parecen para todo k . Un punto i está alejado del origen si la fila i se desvía bastante del modelo. Cada punto i esta dotado del peso d_i . La nube N_K se define simétricamente.

5.2. Los centros de gravedad

Las coordenadas de los centros de gravedad de N_I y de N_K notadas $g_I(k)$ y $g_K(i)$, valen respectivamente:

$$g_I(k) = \sum_i d_i \frac{(f_{ij} - a_{ij})}{d_i} = f_{.k} - a_{.k} ; \quad g_K(i) = \sum_k m_k \frac{(f_{ik} - a_{ik})}{m_k} = f_{i.} - a_{i.} \quad (10)$$

Si las márgenes sobre $K(I)$ de la tabla y del modelo son iguales entre ellas, la nube de las filas (columnas) es centrada.

Las distancias entre los puntos de la nube solo estarán bien representadas por sus proyecciones sobre los primeros ejes de inercia si el centro de gravedad no está muy alejado del origen. Se requiere entonces que las márgenes del modelo no sean muy diferentes a las de la tabla (Escofier 1984, p.26).

5.3. Relaciones de transición

Las relaciones de transición se escriben:

$$F_s(i) = \frac{1}{\sqrt{\lambda_s}} \sum_{k \in K} \left(\frac{f_{ik} - a_{ik}}{d_i} \right) G_s(j) ; \quad G_s(k) = \frac{1}{\sqrt{\lambda_s}} \sum_{i \in I} \left(\frac{f_{ij} - a_{ij}}{m_k} \right) F_s(i) \quad (11)$$

5.4. AC con respecto a un modelo como un AC clásico

El AC clásico (ACS) se puede ver también como el AC de \mathbf{F} con respecto al modelo de independencia, dado por el producto de sus marginales $(f_{i.} f_{.k})$. El AC de \mathbf{F} con respecto a otro modelo \mathbf{A} con métricas y pesos $\mathbf{D}_K = \text{diag}(f_{.k})$ y $\mathbf{D}_I = \text{diag}(f_{i.})$ se puede obtener mediante el AC clásico de $f_{ik} - a_{ik} + f_{i.} f_{.k}$.

6. AC intra-tablas

Se considera el caso en que una tabla de contingencia está conformada por columnas que están agrupadas en subtablas, que se puede considerar también como una yuxtaposición de tablas de contingencia. La notación que se utiliza se muestra en la figura 9.

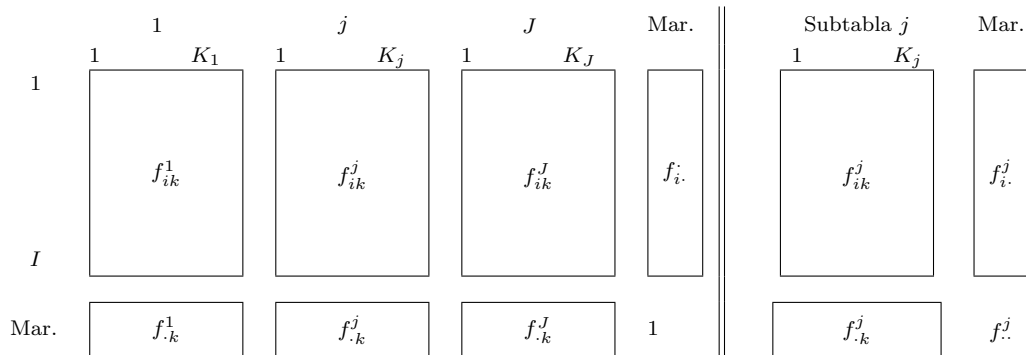


Figura 9: Tabla \mathbf{F} estructurada en J subtablas columna

El AC intra-tablas (Benzécri 1983, Escofier 1984, Escofier & Pagès 1998), toma el modelo de independencia intra-tablas, para estudiar globalmente las desviaciones de cada subnube de columnas con respecto a su propio centroide. Por ejemplo, si se toma en cuenta la partición sobre las columnas de \mathbf{F} , considerando esta tabla como una yuxtaposición de J tablas, el término general del modelo de independencia intra-tablas es (ver figura 9):

$$a_{ik}^j = \frac{f_{i\cdot}^j f_{\cdot k}^j}{f_{\cdot\cdot}^j} \quad (12)$$

Este modelo tiene las mismas márgenes que \mathbf{F} . El análisis intra-tablas es el ACP($\mathbf{X}, \mathbf{M}, \mathbf{D}$) con:

$$x_{ik}^j = \frac{f_{ik}^j}{f_{i\cdot}^j f_{\cdot k}^j} - \frac{f_{i\cdot}^j}{f_{i\cdot}^j f_{\cdot\cdot}^j} \quad ; \quad \mathbf{M} = \text{diag}(f_{\cdot k}^j) \quad ; \quad \mathbf{D} = \text{diag}(f_{i\cdot}^j) \quad (13)$$

El AC intra-tablas recentra las subnubes de puntos-columna correspondientes a cada subtabla o matriz-banda j sobre sus propios centros de gravedad $\frac{f_{i\cdot}^j}{f_{\cdot\cdot}^j}$. Es decir que las subnubes correspondientes a cada subtabla j están centradas, en el origen de la representación.

En AC intratablas es la yuxtaposición de AC de las tablas separadas, utilizando en cada una de ellas los pesos y la métrica \mathbf{D}_I de la tabla global \mathbf{F} , en lugar de la métrica y pesos derivados de sus propias marginales fila, notadas \mathbf{D}_I^j . La marginal fila de \mathbf{F} es el promedio ponderado de las marginales de cada una de las tablas.

Las propiedades del AC intra-tablas se derivan del AC con respecto a un modelo.

6.1. Ayudas a la interpretación asociadas a las subnubes del AC intra-tablas

Las ayudas de un AC clásico (contribuciones, calidades de representación y distancias al origen), son aplicables en el AC intra-tablas. Sin embargo son posibles ayudas adicionales asociadas a las subnubes.

La nube N_K de los K puntos columna en \mathbb{R}^I está particionada en J subnubes N_{K_j} , cada una conformada por los K_j puntos correspondientes a las columnas de la subtabla j .

6.1.1. Peso de las subnubes

El peso de una subnube columna N_{K_j} , es la suma de los pesos de las columnas que la conforman:

$$f_{\cdot\cdot}^j = \sum_{k \in K_j} f_{\cdot k}^j.$$

6.1.2. Inercia de cada subnube

La inercia de la subnube N_{K_j} , $In(K_j)$, es la suma de la contribución a la inercia de sus K_j puntos (recordar que la métrica en \mathbb{R}^I es $\mathbf{D}_I = \text{diag}(f_{i\cdot}^j)$):

$$In(K_j) = \sum_{k \in K_j} f_{\cdot k}^j \sum_{i=1}^I f_{i\cdot}^j (x_{ik}^j)^2 \quad (14)$$

La inercia de la subnube N_{K_j} proyectada sobre el eje s es:

$$In_s(K_j) = \sum_{k \in K_j} f_{.k}^j F_s(j, k)^2 \quad (15)$$

Las inercias calculadas son intra-nubes o intra-tablas, debido a que el AC intra-tablas elimina la inercia entre-tablas, al llevar cada subnube al centro de la representación, se cumple entonces: $In_s(N_K) = \sum_{j=1}^J f_{.j} In_s(K_j)$. Cada elemento de esta suma es la contribución de la subnube j a la inercia del eje s . Para la inercia total en \mathbb{R}^I también se puede calcular esta contribución.

En resumen para las subnubes proyectadas se definen las siguientes ayudas a la interpretación:

Calidad de representación: $In_s(K_j)/In(K_j)$

Contribución a la inercia del eje: $f_{.j} In_s(K_j)/\lambda_s$

6.2. Análisis intra-fechas de Ardèche

En el ACS de Ardèche (sección 4.4) la variabilidad intra-fechas es la que más influye. Aunque el efecto de los sitios se puede ver en el plano factorial 3-4, es posible eliminar el efecto de las fechas haciendo un análisis intra-tablas columna, cada una correspondiendo a una fecha. En este análisis las nubes de cada fecha están centradas en el origen de la representación (figura 11).

La gráfica de valores propios (figura 10) es similar a la del ACS (figura 7) eliminando los dos primeros valores propios. De hecho los valores propios 3 y 4 del ACS son 0.08 y 0.07 y los valores propios 1 y 2 del ACintra-fechas son 0.09 y 0.07. Es decir que el plano 1-2 del ACintra-fechas es similar al plano 3-4 ACS (ver figuras 8 y 11).

La tabla 6 muestra la repartición de la inercias intra-fecha: total y en proyección sobre los primeros ejes factoriales. Las nubes menos dispersas en el espacio completo son *nov82* y *jul83*, sin embargo en el primer plano factorial *nov82* tiene poca inercia comparada con las demás nubes, como consecuencia su calidad de representación es también baja. Al primer eje contribuyen sobre todo *feb83* y *abr83*, de modo que la dispersión de las especies a lo largo de este eje conduce a una tipología de los sitios en invierno, en cambio la dispersión en el segundo eje se da sobretodo en verano.

Tabla 6: Ayudas a la interpretación asociadas a las subnubes

Fechas	Total		Comp1			Comp2			Plano	
	Cont.Inercia x1000	%	Cont.Inercia x1000	%	Calidad %	Cont.Inercia x1000	%	Calidad %	Calidad %	Pesos %
jul82	73.20	14.71	4.15	4.66	5.67	12.49	18.63	17.07	22.74	18.59
aug82	111.86	22.48	3.45	3.87	3.08	24.68	36.80	22.06	25.14	17.97
nov82	58.97	11.85	2.66	2.98	4.51	0.85	1.27	1.44	5.95	10.54
feb83	88.08	17.70	39.61	44.43	44.97	4.49	6.70	5.10	50.07	13.64
apr83	100.13	20.12	36.53	40.98	36.48	8.32	12.40	8.31	44.79	18.95
jul83	65.28	13.12	2.74	3.07	4.20	16.22	24.19	24.85	29.05	20.31
Total	497.53	100.00	89.13	100.00		67.06	100.00			100.00

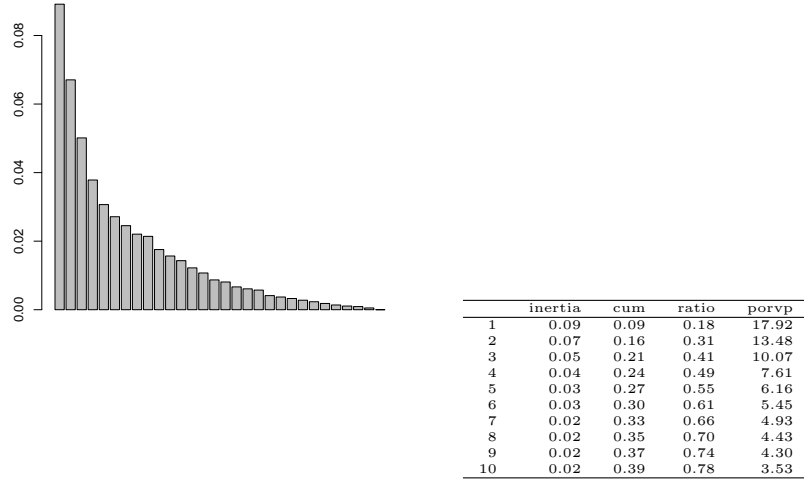


Figura 10: Valores propios del AC intra-fechas de Ardèche

7. AC interno, ACI

El ACI fue propuesto por Cazes et al. (1988), para el análisis de tablas faunísticas en ecología. Este análisis tiene en cuenta las dos particiones de la tabla \mathbf{F} : la partición de las filas y la partición de las columnas. La notación utilizada aparece en la figura 12.

El ACI es una generalización del AC intra-tablas, para tener en cuenta la partición tanto en filas como en columnas y es el AC con respecto al modelo:

$$a_{ik}^{lj} = \frac{f_{\cdot k}^{lj} f_{i \cdot}^{l \cdot}}{f_{\cdot \cdot}^{l \cdot}} + \frac{f_{i \cdot}^{lj} f_{\cdot k}^{j \cdot}}{f_{\cdot \cdot}^{j \cdot}} - \frac{f_{i \cdot}^{l \cdot} f_{\cdot k}^{j \cdot} f_{\cdot \cdot}^{lj}}{f_{\cdot \cdot}^{l \cdot} f_{\cdot \cdot}^{j \cdot}} \quad (16)$$

Este modelo tiene las mismas márgenes que la tabla \mathbf{F} . El modelo (16) se obtiene al aplicar dos veces el AC intra-tablas: AC intra-tablas fila de $f_{ik}^{lj} - \frac{f_{i \cdot}^{lj} f_{\cdot k}^{j \cdot}}{f_{\cdot \cdot}^{j \cdot}} + f_{i \cdot}^{l \cdot} f_{\cdot k}^{j \cdot}$ o el AC intra-tablas columna

de $f_{ik}^{lj} - \frac{f_{\cdot k}^{lj} f_{i \cdot}^{l \cdot}}{f_{\cdot \cdot}^{l \cdot}} + f_{i \cdot}^{l \cdot} f_{\cdot k}^{j \cdot}$. En ambos casos el resultado es el mismo y el ACI se denomina, también, doble AC intra (Cazes & Moreau 1991), entonces (16) se puede llamar modelo de independencia intra-intra. El ACI es el ACP($\mathbf{X}, \mathbf{M}, \mathbf{D}$) con:

$$x_{ik}^{lj} = \frac{f_{ik}^{lj}}{f_{i \cdot}^{l \cdot} f_{\cdot k}^{j \cdot}} - \frac{f_{\cdot k}^{lj}}{f_{\cdot \cdot}^{l \cdot} f_{\cdot k}^{j \cdot}} - \frac{f_{i \cdot}^{lj}}{f_{\cdot \cdot}^{j \cdot} f_{i \cdot}^{l \cdot}} + \frac{f_{\cdot \cdot}^{lj}}{f_{\cdot \cdot}^{l \cdot} f_{\cdot \cdot}^{j \cdot}} \quad ; \quad \mathbf{M} = \text{diag}(f_{\cdot k}^{j \cdot}) \quad ; \quad \mathbf{D} = \text{diag}(f_{i \cdot}^{l \cdot}) \quad (17)$$

7.1. Propiedades del ACI

1. Las subnubes asociadas a las J bandas-columna están centradas.
2. Las subnubes asociadas a las L bandas filas están centradas.
3. Las coordenadas factoriales están centradas y centradas por banda-fila y banda-columna.

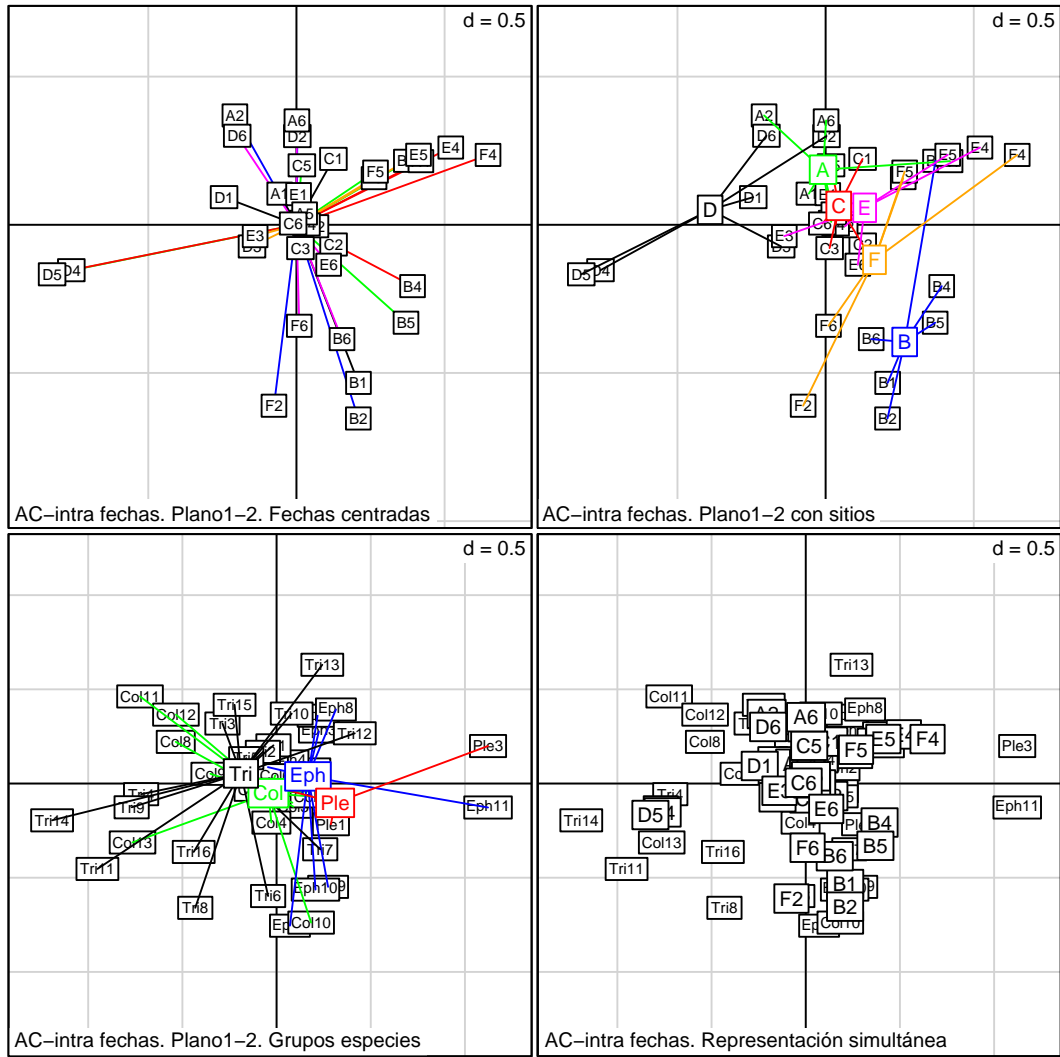


Figura 11: Primer plano factorial del AC intra-fechas de Ardèche

4. Las formulas de transición son:

$$F_s(l, i) = \frac{1}{\sqrt{\lambda_s}} \sum_{j,k} \left(\frac{f_{jk}^{lj}}{f_{i.}^{lj}} - \frac{f_{.k}^{lj}}{f_{..}^{lj}} \right) G_s(j, k) ; \quad G_s(j, k) = \frac{1}{\sqrt{\lambda_s}} \sum_{l,i} \left(\frac{f_{ik}^{lj}}{f_{i.}^{lj}} - \frac{f_{.k}^{lj}}{f_{..}^{lj}} \right) F_s(l, i) \quad (18)$$

5. El análisis intra-tablas es un ACI, cuando solo se tiene en cuenta una partición.

7.1.1. Ayudas a la interpretación asociadas a las subnubes fila (columna)

De la misma manera que en el AC-intra columnas, la nube N_K en \mathbb{R}^I esta dividida en J subnubes N_{K_j} , de K_j puntos cada una y las ayudas allí descritas se aplican en el ACI (§6.1). Adicionalmente, la nube N_I en \mathbb{R}^K está dividida en L subnubes N_{I_l} de I_l puntos cada una. La fórmulas de las ayudas asociadas a éstas subnubes se derivan de las de la sección 6.1.

	1	j	J	Mar.		Banda j	Mar.
	1	K_1	K_j	K_J		1	K_j
1	f_{ik}^{11}	f_{ik}^{1j}	f_{ik}^{1J}	$f_{i\cdot}^{1\cdot}$		f_{ik}^{1j}	$f_{i\cdot}^{1j}$
I_1							
l	f_{ik}^{l1}	f_{ik}^{lj}	f_{ik}^{lJ}	$f_{i\cdot}^{l\cdot}$		f_{ik}^{lj}	$f_{i\cdot}^{lj}$
I_l							
L	f_{ik}^{L1}	f_{ik}^{Lj}	f_{ik}^{LJ}	$f_{i\cdot}^{L\cdot}$		f_{ik}^{Lj}	$f_{i\cdot}^{Lj}$
I_L							
Mar.	$f_{\cdot k}^{1\cdot}$	$f_{\cdot k}^{j\cdot}$	$f_{\cdot k}^{J\cdot}$	1		$f_{\cdot k}^{j\cdot}$	$f_{\cdot\cdot}^{j\cdot}$

	Subtabla banda fila l			Mar.		Subtabla (l, j)	Mar.
1	f_{ik}^{l1}	f_{ik}^{lj}	f_{ik}^{lJ}	$f_{i\cdot}^{l\cdot}$		f_{ik}^{lj}	$f_{i\cdot}^{lj}$
I_l							
Mar.	$f_{\cdot k}^{l1}$	$f_{\cdot k}^{lj}$	$f_{\cdot k}^{lJ}$	$f_{\cdot\cdot}^{l\cdot}$		$f_{\cdot k}^{lj}$	$f_{\cdot\cdot}^{lj}$

Figura 12: Tabla F estructurada en $L \times J$ subtablas

7.2. ACI del ejemplo Ardèche

En el ACS (§4.4) y en el AC intra-fechas (§6.2) se ha visto que las proyecciones de grupos taxonómicos como ilustrativos, están muy cerca al centro, lo que significa que la inercia entre ellos es baja. De todas maneras se puede realizar un análisis sin tener en cuenta esta dispersión. En el ACI de este ejemplo se elimina la inercia entre-fechas y la inercia entre-grupos taxonómicos, lo que permite interpretar la tipología espacial y evaluar el potencial de los grupos taxonómicos como descriptores biológicos. El análisis se realizó con ADE4, exportando las coordenadas factoriales a SPAD (<http://www.cisia.com/>) para producir y editar los planos. Los nombres científicos de las especies se encuentran en el artículo citado de Cazes et al. (1988).

7.2.1. Valores propios

Los valores propios (figura 13) son similares a los del AC intra-fechas (figura 10), debido a que la inercia entre-grupos taxonómicos es baja. Se destacan los dos primeros valores propios, pero también puede ser interesante interpretar los ejes 3 y 4. Estos 4 primeros ejes retienen el 47% de la inercia. El plano 1-2 del ACI (figuras 14 y 15) es muy parecido al plano 3-4 de ACS (figura 8) y al plano 1-2 del AC intra-fechas (figura 11). Hay que tener en cuenta que los planos del ACI están rotados en ambos ejes con respecto al ACS y AC intra-fechas (§2.2.2).

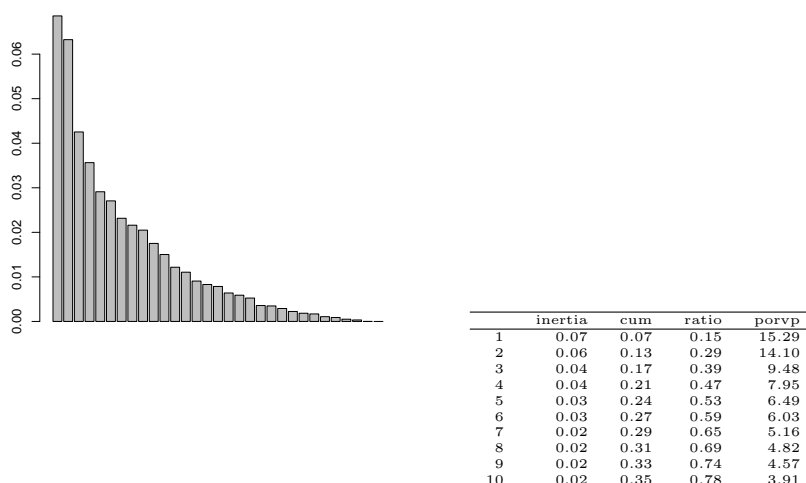


Figura 13: Valores propios del ACI de Ardèche

7.2.2. Análisis de las columnas

La figura 14 muestra los puntos $fechas \times sitios$ en el primer plano factorial del ACI. Como se presenta en el artículo de Cazes et al. (1988), el ACI destaca la originalidad del sitio *B*, en contraste con *A* y *D*, principalmente. El sitio *D* tiene una composición faunística muy específica en invierno (*D-feb83* y *D-abr83*). *F* y, un poco menos, *D* presentan diferencias marcadas entre invierno (al lado izquierdo del primer eje: *F-feb83*, *E-feb83*, *F-abr83* y *E-abr83*) y verano (al lado derecho del primer eje: *F-ago82* y *F-jul83*). Finalmente la crecida del río, en noviembre, homogeniza la distribución faunística y, entonces la subnube $nov82 \times sitios$ se ubica cerca al centroide. Ver también la tabla 7.

Tabla 7: Descomposición de la inercia del ACI. Bandas columna (fechas)

Fechas	Total		Comp1			Comp2			Plano	
	Cont.Inercia x1000	%	Cont.Inercia x1000	%	Calidad %	Cont.Inercia x1000	%	Calidad %	Calidad %	Pesos %
jul82	66.18	14.76	2.02	2.95	3.06	13.12	20.75	19.83	22.89	18.59
aug82	101.67	22.67	7.43	10.84	7.31	19.10	30.20	18.79	26.10	17.97
nov82	55.79	12.44	1.79	2.62	3.22	0.07	0.11	0.13	3.35	10.54
feb83	74.72	16.66	26.13	38.10	34.97	4.38	6.92	5.86	40.83	13.64
apr83	88.25	19.68	26.24	38.27	29.73	11.63	18.39	13.18	42.92	18.95
jul83	61.82	13.79	4.95	7.22	8.01	14.94	23.63	24.17	32.18	20.31
Total	448.43	100.00	68.58	100.00		63.25	100.00			100.00

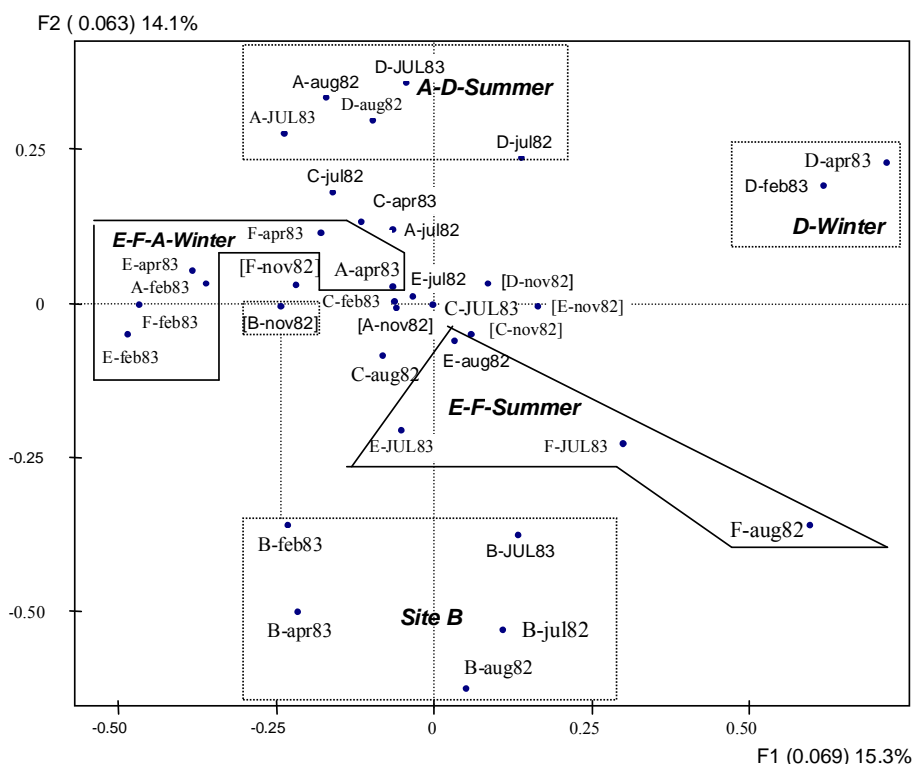


Figura 14: Puntos columna sobre el primer plano factorial del ACI

7.2.3. Análisis de las filas

En cuanto a las especies, la inercia del primer eje se debe principalmente a la gran dispersión de los trichópteros (ver tabla 8): en este grupo las especies con vaina son atraídas por los sitios arenosos y rocosos, mientras que los trichópteros libres prefieren el sustrato duro del suelo (ver figura 15). La dispersión del grupo *Coleoptera* contribuye principalmente a la inercia del segundo eje (tabla 8), contrastando a las especies según su preferencia o no a las corrientes fuertes (figura 15). Cazes et al. (1988) concluyen que hay una tipología estival, definida principalmente por el grupo *Coleoptera* y otra tipología invernal, debida al grupo *Trichoptera*, destacando la originalidad del sitio *D* y la homogenización de la fauna en noviembre.

Tabla 8: Descomposición de la inercia del ACI. Bandas fila (grupos taxonómicos)

Fechas	Total		Comp1			Comp2			Plano	
	Cont.Inercia x1000	%	Cont.Inercia x1000	%	Calidad %	Cont.Inercia x1000	%	Calidad %	Calidad %	Pesos %
Eph	106.39	23.72	14.71	21.45	13.83	16.20	25.62	15.23	29.06	28.79
Ple	26.64	5.94	7.53	10.98	28.26	1.72	2.71	6.45	34.71	6.53
Col	116.52	25.98	7.54	10.99	6.47	26.25	41.51	22.53	29.00	37.03
Tri	198.88	44.35	38.80	56.57	19.51	19.08	30.16	9.59	29.10	27.66
Total	448.43	100.00	68.58	100.00		63.25	100.00			100.00

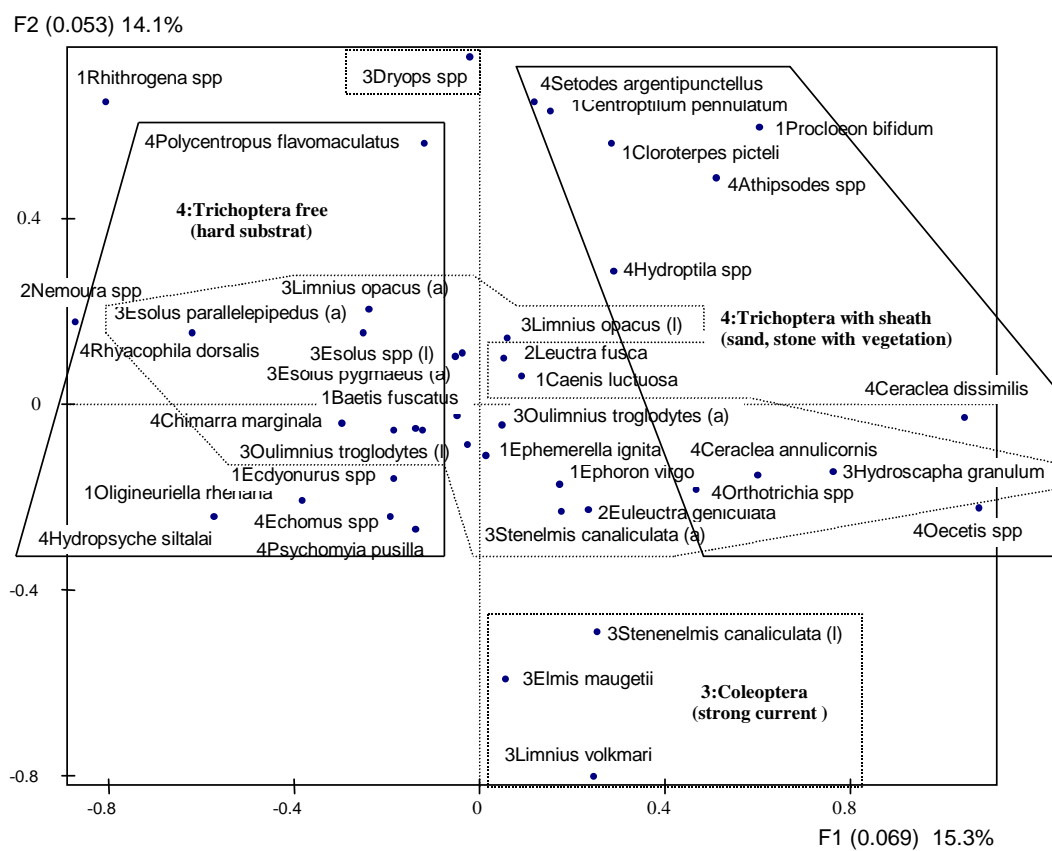


Figura 15: Especies sobre el primer plano factorial del ACI

8. Talleres

8.1. Instalación del R, del ADE-4 y otros

1. Abrir la página <http://www.R-project.org>
2. Bajar el programa *rw2001.exe*
3. Instalar RGui (R bajo Windows), haciendo doble click en *rw2001.exe*
4. Iniciar R, haciendo click en el acceso directo **R 2.0.1** o en *Inicio* → *Programas* → *R* → *R 2.0.1*
5. Bajar ADE-4 de <http://pbil.univ-lyon1.fr/R/rplus/ade4dsR.html> o una página de R.
6. Instalar ADE-4: en el menú de R hacer click en *Packages*, luego en *Install package(s) from local zip files ...*, buscar el archivo *ade4_1.3-3.zip* y hacer click en *Abrir*
7. Bajar *scatterplot3d* de <http://cran.au.r-project.org/>:
Packages → *scatterplot3d* → *scatterplot3d_0.3-19.zip* → *Guardar*.
8. Instalar *scatterplot3d*: El mismo procedimiento del numeral 6.

Los interesados en utilizar el editor WinEdt o editar resultados de R en Latex pueden consultar Ortiz & Pardo (2004).

Ejercicio. Instalar el paquete *xtable*.

8.2. ACP paso a paso

1. Definición de la tabla de datos **Y** y gráfica en 3D

```
#-----
# creación de la matriz Y con las variables físicas de café
#-----
color <-c(298,361,321,335,314,186,278,238,226,210)
DA <- c(385.1,481.3,422.6,444.3,368.7,346.6,422.6,403,368.7,368.7)
EA <- c(25,41,40,33,32,28,43,42,36,35)
Y <- matrix(0,10,3) # se define Y como matriz 0 de 10x3
Y[,1] <- color # se le asigna a la primer columna el vector color
Y[,2] <- DA # segunda columna de Y es DA
Y[,3] <- EA # tercera columna de Y es EA
Y # se ve la matriz en R
#-----

# Gráfica de Y
library(scatterplot3d)
scatterplot3d(Y)
#-----
```

2. Cálculo del centro de gravedad **g**

```
#-----
I<- 10 # número de individuos
K <- 3 # número de variables
```

```

UnoI <- matrix(1,I,1) # vector columna de I unos
g <- t(Y)%*%UnoI/I # calculo del centro de gravedad
g # para ver los valores
#-----

```

3. Centrado y reducido de **Y**. Gráfica. Correlaciones.

```

#-----
Yc <- Y - UnoI)%*%t(g)
Yc
scatterplot3d(Yc) # grafica 3D
#-----
# cálculo de Des
V <- t(Yc)%*%Yc/I # matriz de covarianzas
Des <- matrix(0,K,K) # define matrix Des (KxK) y la inicializa con ceros
for (k in 1:K) Des[k,k] <- sqrt(1/V[k,k]) # asigna diagonales de Des
Des # para ver Des
#-----
# cálculo de X y gráfica 3D
X <- Yc %*% Des
colnames(X) <- colnames(Y)
X
scatterplot3d(X) # grafica 3D
#-----
# matriz de correlaciones
V <- t(X) %*% X/I
rownames(V) <- colnames(Y); colnames(V) <- colnames(Y)
V
#-----

```

4. Valores y vectores propios. Coordenadas filas. Gráfica filas primer plano.

```

#-----
eigen(V)
U <- eigen(V)$vectors # cada columna de U es un vector propio
F <- X %*% U # obtención de las coordenadas
colnames(F) <- c("F1","F2","F3")
rownames(U) <- colnames(Y)
colnames(U) <- colnames(F)
# gráfica
par(mfrow=c(1,1)) # una sola gráfica
F1 <- F[,1]; F2 <- F[,2]
E1 <- F1; E2 <- F2+0.2; E2[8] <-F2[8]-0.2; E2[10]<-F2[10]-0.2 # coordenadas etiquetas
plot(F1,F2,xlim=c(-2.8,2.8),ylim=c(-2.8,2.8)) # grafica
text(E1,E2,labels=rownames(F)) # pone etiquetas de cafés
lines(c(-3,3),c(0,0),lty=2) # eje horizontal en 0
lines(c(0,0),c(-3,3),lty=2) # eje vertical en cero
# fin gráfica
#-----

```

5. Círculo de correlaciones, coordenadas y ayudas

```

#-----
G <- t(X) %*% eigen(X %*% t(X))$vectors[,1:2]/sqrt(I)
# verificación del centrado

```

```

D <- diag(1/I,I)
t(X) %*% D %*% UnoI
# ayudas a la interpretación
cont <- G*G/c(1,1,1) %*% t(eigen(V)$values[1:2])*100
cos2 <- G*G*100
cos2plano <- rowSums(cos2)
ayu <- cbind(G,cont,cos2,cos2plano)
ayu
# círculo de correlaciones
library(ade4)
s.corcircle(G)
#-----

```

Ejercicios

1. Compare los resultados obtenidos paso a paso con los del programa princomp:

```

#-----
# 1.ACP con princomp
acp1 <- princomp(Y,cor=TRUE) # ACP normado de los datos
class(acp1) # clase de objeto que produce princomp
names(acp1) # contenido del objeto
# [1] "sdev" "loadings" "center" "scale" "n.obs" "scores" "call"
acp1$sdev^2 # valores propios
unclass(acp1) # ver todo el contenido de acp1
summary(acp1) # proporción de varianza de cada eje
plot(acp1) # gráfica de valores propios
biplot(acp1) # gráfica de individuos con antiguos ejes unitarios
#-----

```

2. Compare los resultados anteriores con los de ADE4:

```

#-----
#2. ACP con ade4
library(ade4)
acp2 <- dudi.pca(Y)
# aparece el gráfico de valores propios en la ventana gráfica
# Select the number of axes: teclee 2 y return
class(acp2) # clase de objeto que produce dudi.pca
names(acp2) # contenido del objeto que sale de dudi.pca
unclass(acp2) # ver el contenido de acp2
s.corcircle(acp2$co) # círculo de correlaciones
s.label(acp2$li) # primer plano factorial de individuos
scatter(acp2) # gráfica de individuos con antiguos ejes unitarios
#-----

```

8.3. AC de una tabla de contingencia pequeña

1. Entrada de datos. Marginales. Frecuencias relativas

```

#-----

```

```

#1-Entrada de datos
N <- matrix(0,3,6)
N[1,] <- c(7,7,5,0,4,4)
N[2,] <- c(0,0,0,5,5,5)
N[3,] <- c(4,4,0,0,3,0)
rownames(N) <- c("Ple1","Ple2","Ple3")
colnames(N) <- c("A","B","C","D","E","F")
N
#2-Marignales fila y columna y total
rowSums(N)
colSums(N)
sum(N)
#3-Tabla de frecuencias relativas y sus marginales
F <- prop.table(N)
rowSums(F)*100
colSums(F)*100
sum(F)
#-----

```

2. Tablas derivadas de TC y F

```

#-----
PF <- prop.table(N,1)
PF*100
rowSums(PF)
PC <- prop.table(N,2)
PC*100
rowSums(PC)
# tabla independencia
A <- rowSums(F) %*% t(colSums(F))
rownames(A) <- rownames(F); colnames(A) <- colnames(F)
FI*100
(F - A)*100
#-----

```

3. Gráficas derivadas de la TC

```

#-----
par(mfrow=c(2,2)) # para montar 4 gráficas
library(ade4)
table.cont(t(N),csize=4,col.labels=rownames(N))
assocplot(N,main="Desvios de la independencia")
mosaicplot(F,color=TRUE,main="Perfiles fila")
mosaicplot(t(F),color=TRUE,main="Perfiles columna")
#-----

```

4. Primer plano factorial

```

#-----
eje1 <- dudi.coa(as.data.frame(N))
par(mfrow=c(1,1))
s.label(eje1$co,label=c("AB","AB",colnames(N)[3:6]))
s.label(eje1$li,add.plot=TRUE)
#-----

```

Ejercicio. Analice las salidas del ACS de la TC pequeña.

8.4. ACS del ejemplo Ardeche

Desarrolle el ejemplo respondiendo a las siguientes preguntas:

1. Explique el contenido y los tipos de información del ejemplo *Ardèche*:

```
#-----
#Ejemplo Ardeche
#-----
library(ade4) # activa libreria
data(ardeche) # lee datos de ejemplo
class(ardeche)
summary(ardeche)
unclass(ardeche)
#-----
```

2. Explique cómo se hace el cambio de nombre de las columnas de la tabla faunística de Ardèche, que aparece a continuación:

```
#-----
# cambio de nombres de las columnas
names(ardeche$tab)
names(ardeche$tab) <-
  paste(substr(names(ardeche$tab),2,2),substr(names(ardeche$tab),1,1),sep="")
names(ardeche$tab)
#-----
```

3. Comprenda las instrucciones que originan la figura 8:

```
#-----
# cuatro gráficas del ACS de Ardeche
par(mfrow=c(2,2)) # 4 gráficas
# AC global (sin tener en cuenta la particion)
ACglobal <- dudi.coa(ardeche$tab, scan = FALSE,nf=4) # realiza AC
# plano 1-2 de columnas mostrando fechas
s.label(ACglobal$co, clab = 0.75,sub="Plano1-2 Fechas") # pone etiquetas sobre primer plano
s.class(ACglobal$co, ardeche$dat.fac,wt=ACglobal$cw,cellipse = 0,cpoint = 0,
  add.p = TRUE,col=rep(c("green","blue"),3))
# plano 1-2 de columnas mostrando estaciones
s.class(ACglobal$co,ardeche$sta.fac,wt=ACglobal$cw,cpoint = 0,cellipse = 0,
  col=c("green","blue","red","black","magenta","orange"),sub="Plano1-2 Estaciones")
s.label(ACglobal$co,clab = 0.75, add.p = TRUE)
# plano 3-4 de columnas mostrando estaciones
s.label(ACglobal$co,clab = 0.75,xax=3,yax=4,sub="Plano3-4 Estaciones")
s.class(ACglobal$co,ardeche$sta.fac,wt=ACglobal$cw,cpoint = 0,cellipse = 0,
  col=c("green","blue","red","black","magenta","orange"),xax=3,yax=4, add.p = TRUE)
# plano 1-2 de filas mostrando grupos taxonómicos
s.label(ACglobal$li,clab = 0.75)
s.class(ACglobal$li,ardeche$esp.fac,wt=ACglobal$lw,cpoint = 0,cellipse = 0,
  col=c("green","blue","red","black"),sub="Plano1-2 Grupos especies", add.p = TRUE)
par(mfrow=c(1,1)) # una gráfica
#-----
```

4. Utilice las ayudas a la interpretación de la función *inertia.dudi* para complementar la interpretación de los ejes y planos factoriales.

a) Identifique los objetos incluidos en la salida de *inertia.dudi*

```
? inertia.dudi          # para conocer la función
ayu <- inertia.dudi(ACglobal,TRUE,TRUE)
```

b) Identifique las filas y columnas que más contribuyen a cada uno de los cuatro primeros ejes.

```
# filas más contributivas a los cuatro primeros ejes
sort(ayu$row.abs[,1],decreasing=TRUE)
sort(ayu$row.abs[,2],decreasing=TRUE)
sort(ayu$row.abs[,3],decreasing=TRUE)
sort(ayu$row.abs[,4],decreasing=TRUE)
# columnas más contributivas
sort(ayu$col.abs[,1],decreasing=TRUE)
sort(ayu$col.abs[,2],decreasing=TRUE)
sort(ayu$col.abs[,3],decreasing=TRUE)
sort(ayu$col.abs[,4],decreasing=TRUE)
```

c) Identifique las filas y columnas mejor y peor representadas en cada uno de los 4 primeros ejes

```
# calidad de representación de las filas
sort(abs(ayu$row.rel[,1]),decreasing=TRUE)
sort(abs(ayu$row.rel[,2]),decreasing=TRUE)
sort(abs(ayu$row.rel[,3]),decreasing=TRUE)
sort(abs(ayu$row.rel[,4]),decreasing=TRUE)
# calidad de representación de las columnas
sort(abs(ayu$col.rel[,1]),decreasing=TRUE)
sort(abs(ayu$col.rel[,2]),decreasing=TRUE)
sort(abs(ayu$col.rel[,3]),decreasing=TRUE)
sort(abs(ayu$col.rel[,4]),decreasing=TRUE)
```

d) Identifique las filas y columnas mejor y peor representadas en los planos 1-2 y 3-4, respectivamente

```
# calidad en el plano1-2
sort(abs(ayu$row.rel[,1])+abs(ayu$row.rel[,2]),decreasing=TRUE)
sort(abs(ayu$col.rel[,1])+abs(ayu$col.rel[,2]),decreasing=TRUE)
# calidad en el plano3-4
sort(abs(ayu$row.rel[,3])+abs(ayu$row.rel[,4]),decreasing=TRUE)
sort(abs(ayu$col.rel[,3])+abs(ayu$col.rel[,4]),decreasing=TRUE)
```

Ejercicio. Realice el ACS de los coleópteros en *feb83*.

8.5. AC intra-fechas de Ardèche

Observe los resultados del siguiente código:

```
#-----
# Análisis intra fechas
#-----
# Gráfica de valores propios
par(mfrow=c(1,1))
witwit.coa(ACglobal,43,ardeche$col.blocks)
#-----
ACwwcol <- witwit.coa(ACglobal,43,ardeche$col.blocks,
                    scannf=FALSE,nf=4)
```

```

# GRAFICAS
par(mfrow=c(2,2))
# plano 1-2 de columnas mostrando fechas
s.class(ACwwcol$co, ardeche$dat.fac,wt=ACwwcol$cw,cellipse = 0,cpoint = 0,
  col=c("green","blue","red","black","magenta","orange"),
  clabel=0,sub="AC-intra fechas. Plano1-2. Fechas centradas")
s.label(ACwwcol$co, add.p = TRUE,clab = 0.75) # pone etiquetas sobre primer plano
# plano 1-2 de columnas mostrando sitios
s.label(ACwwcol$co,clab = 0.75)
s.class(ACwwcol$co,ardeche$sta.fac,wt=ACwwcol$cw,cpoint = 0,cellipse = 0,
sub="AC-intra fechas. Plano1-2 con sitios",
  col=c("green","blue","red","black","magenta","orange"), add.p = TRUE)
# plano 1-2 de filas mostrando grupos taxonómicos
s.label(ACwwcol$li,clab = 0.75)
s.class(ACwwcol$li,ardeche$esp.fac,wt=ACglobal$lw,cpoint = 0,cellipse = 0,
  col=c("green","blue","red","black"),
  sub="AC-intra fechas. Plano1-2. Grupos especies",add.p = TRUE)
# representacion simultánea
scatter.coa(ACwwcol,clab.row = 0.75,
  clab.col = 1.0, posieig = "none",
  sub = "AC-intra fechas. Representación simultánea", csub = 1.0)
#-----
# Ayudas a la interpretación
inertia.dudi(ACwwcol,TRUE,TRUE)
#-----

```

Ejercicio. Realice el AC intra-grupos taxonómicos de Ardèche.

8.6. ACI de Ardèche

Compruebe las conclusiones del ACI de Ardeche del texto (§7.2), utilizando directamente las gráficas y salidas del ADE4 y R.

1. Ejecución del ACI de Ardèche. Identifique cada uno de los objetos de la salida de *witwit.coa* y de *summary.witwit*.

```

#-----
# ACI
#-----
library(ade4) # activa libreria
# cambio de nombres de las columnas
names(ardeche$tab)
names(ardeche$tab) <-
  paste(substr(names(ardeche$tab),2,2),substr(names(ardeche$tab),1,1),sep="")
names(ardeche$tab)
ACglobal <- dudi.coa(ardeche$tab, scan = FALSE,nf=4) # realiza AC
ACww <- witwit.coa(ACglobal,ardeche$row.blocks,ardeche$col.blocks,
  scannf=FALSE,nf=4) # realiza ACI
names(ACww)
unclass(ACww)
summary(ACww)
#-----

```

2. Verifique que las fechas y los grupos taxonómicos están centrados.

```

#-----
# plano 1-2 de columnas mostrando fechas centradas
par(mfrow=c(1,1))
s.class(ACww$co, ardeche$dat.fac,wt=ACww$cw,cellipse = 0,cpoint = 0,
  col=c("green","blue","red","black","magenta","orange"),clabel=0)
s.label(ACww$co,clab = 0.75, add.p = TRUE)
# plano 1-2 de filas mostrando grupos centrados
s.class(ACww$li, ardeche$esp.fac,wt=ACww$lw,cellipse = 0,cpoint = 0,
  col=c("green","blue","red","black"),clabel=0)
s.label(ACww$li,clab = 0.75, add.p = TRUE)
#-----

```

3. Gráficas separadas para cada fecha

```

#-----
# gráficas de ACI por fechas
par(mfrow=c(3,2))
for (j in 1:6) {
  cla <- unique(ardeche$dat.fac)[j]
  x <- ACww$co[ardeche$dat.fac==cla,]
  s.label(ACww$co,clab=0,cpoi=1,sub=as.character(cla),csub=2.5)
  s.label(x,label=c('A','B','C','D','E','F'),add.p=T,clab=1.5)
}
#-----

```

4. Gráficas separadas por grupos taxonómicos

```

#-----
# gráficas de ACI por grupos taxonómicos
par(mfrow=c(2,2))
for (j in 1:4) {
  cla <- unique(ardeche$esp.fac)[j]
  x <- ACww$li[ardeche$esp.fac==cla,]
  s.label(ACww$li,clab=0,cpoi=1,sub=as.character(cla),csub=1.5)
  s.label(x,add.p=T,clab=1)
}
#-----

```

5. Primer plano factorial mostrando sitios

```

#-----
# plano 1-2 de columnas mostrando sitios
par(mfrow=c(1,1))
s.label(ACww$co,clab = 0.75)
s.class(ACww$co,ardeche$sta.fac,wt=ACww$cw,cpoint = 0,cellipse = 0,
  col=c("green","blue","red","black","magenta","orange"), add.p = TRUE)
#-----

```

6. Ayudas asociadas a las subnubes (tablas 7 y 8)

```

#-----
# Contribuciones de las subnubes a la Inercia en ACI y AC intra
#-----
# Función inerti.ica.

```

```

# Ayudas ACI (no parciales) modificada Feb-23-05
# organiza las ayudas dadas en witwit.coa
# ENTRA
# ACww salida de witwit.coa; cbl=columnas por bloque columna; rbl = idem. fila
# default es número de filas y número de columnas
# xax=eje horizontal, yax = eje vertical. Default es 1 y 2
# SALE
# porcentaje de inercia de la subnube fila(columna) al eje
# calidad de representación de la subnube,
# contribución a la inercia de los ejes y total
#tabla con dos ejes
# LAS INERCIAS ESTAN MULTIPLICADAS POR 1000 EN LA TABLA
#-----
inertia.ica <- function(ACww,cbl=ncol(ACww$tab),rbl=nrow(ACww$tab),xax=1,yax=2)
{
  J <- nrow(ACww$cbvar); L <- nrow(ACww$lvar)
  nf <- ACww$nf
  # inercia total subnubes (no parciales)
  X <- as.matrix(ACww$tab)
  normK <- apply(X*X*ACww$lw,2,sum);normI <- apply(t(X*X)*ACww$cw,2,sum)
  cbl.fac <- rep(1:J,cbl);rbl.fac <- rep(1:L,rbl)
  # columnas
  inJ <- tapply(normK*ACww$cw,cbl.fac,sum);names(inJ) <- names(cbl)
  cinJ <- inJ/sum(inJ)*100
  #filas
  inL <- tapply(normI*ACww$lw,rbl.fac,sum);names(inL) <- names(rbl)
  cinL <- inL/sum(inL)*100
  # calidad o contribucion relativa
  Unonf <- t(rep(1,nf))
  # tabla para dos ejes
  icol <- NULL
  icol <- cbind(icol,inJ*1000,cinJ)
  icol <- cbind(icol,((ACww$cbvar * (ACww$cbw %>% t(rep(1,nf))))[xax])*1000)

  inertia <- NULL
  inertia$cloud.col <- cinJ
  inertia$cloud.row <- cinL
  inertia$rel.col <- ((ACww$cbvar * (ACww$cbw %>% t(rep(1,nf))))/
    (inJ %>% Unonf))*100
  inertia$rel.row <- ((ACww$lvar * (ACww$lw %>% t(rep(1,nf))))/
    (inL %>% Unonf))*100
  inertia$abs.col <- ((ACww$cbvar * (ACww$cbw %>% t(rep(1,nf))))/
    (rep(1,J) %>% t(ACww$eig[1:nf])))*100
  inertia$abs.row <- ((ACww$lvar * (ACww$lw %>% t(rep(1,nf))))/
    (rep(1,L) %>% t(ACww$eig[1:nf])))*100

  # tabla para dos ejes COLUMNAS
  icol <- NULL
  # I total
  icol <- cbind(icol,inJ*1000,cinJ); colnames(icol) <- c("Inubej","Cnubej")
  # comp 1
  icol <- cbind(icol,((ACww$cbvar * (ACww$cbw %>% t(rep(1,nf))))[xax])*1000)
  colnames(icol)[3] <- paste("IjComp",xax,sep="")
  icol <- cbind(icol,inertia$abs.col[,xax])
  colnames(icol)[4] <- paste("CIjComp",xax,sep="")
  icol <- cbind(icol,inertia$rel.col[,xax])
  colnames(icol)[5] <- paste("CaljComp",xax,sep="")
  # comp 2
  icol <- cbind(icol,((ACww$cbvar * (ACww$cbw %>% t(rep(1,nf))))[yax])*1000)
  colnames(icol)[6] <- paste("IjComp",yax,sep="")
  icol <- cbind(icol,inertia$abs.col[,yax])
  colnames(icol)[7] <- paste("CIjComp",yax,sep="")
  icol <- cbind(icol,inertia$rel.col[,yax])
  colnames(icol)[8] <- paste("CaljComp",yax,sep="")
  # calidad plano
  icol <- cbind(icol,inertia$rel.col[,xax]+inertia$rel.col[,yax])

```

```

colnames(icol)[9] <- paste("CaljPlano",xax,"-",yax,sep="")
# peso
icol <- cbind(icol,ACww$cbw*100)
colnames(icol)[10] <- "PesoJ"
# sumas
suma <- colSums(icol);suma[5] <- NA;suma[8:9] <- NA
icol <- rbind(icol,suma)
rownames(icol)[(J+1)] <- "Total"
# tabla para dos ejes FILAS
ifil <- NULL
# I total
ifil <- cbind(ifil,inL*1000,cinL); colnames(ifil) <- c("Inubel","Cnubel")
# comp 1
ifil <- cbind(ifil,((ACww$lbvar * (ACww$lbw %>% t(rep(1,nf))))[xax])*1000)
colnames(ifil)[3] <- paste("I1Comp",xax,sep="")
ifil <- cbind(ifil,inertia$abs.row[,xax])
colnames(ifil)[4] <- paste("CI1Comp",xax,sep="")
ifil <- cbind(ifil,inertia$rel.row[,xax])
colnames(ifil)[5] <- paste("CallComp",xax,sep="")
# comp 2
ifil <- cbind(ifil,((ACww$lbvar * (ACww$lbw %>% t(rep(1,nf))))[yax])*1000)
colnames(ifil)[6] <- paste("I1Comp",yax,sep="")
ifil <- cbind(ifil,inertia$abs.row[,yax])
colnames(ifil)[7] <- paste("CI1Comp",yax,sep="")
ifil <- cbind(ifil,inertia$rel.row[,yax])
colnames(ifil)[8] <- paste("CallComp",yax,sep="")
# calidad plano
ifil <- cbind(ifil,inertia$rel.row[,xax]+inertia$rel.row[,yax])
colnames(ifil)[9] <- paste("CallPlano",xax,"-",yax,sep="")
# peso
ifil <- cbind(ifil,ACww$lbw*100)
colnames(ifil)[10] <- "PesoI"
# sumas de filas
suma <- colSums(ifil);suma[5] <- NA;suma[8:9] <- NA
ifil <- rbind(ifil,suma)
rownames(ifil)[(L+1)] <- "Total"
inertia$col <- icol
inertia$row <- ifil
return(inertia)
}
#-----
ayu <- inertia.ica(ACww,ardeche$col.blocks,ardeche$row.blocks)
ayu$col
ayu$row
#-----

```

7. Ayudas para las subnubes del AC intra-fechas (tabla 6)

```

#-----
# Ayudas para subnubes de AC intra-fechas
ayu <- inertia.ica(ACwwcol,ardeche$col.blocks)
ayu$col
#-----

```

Lectura. Métodos lineales de ordenación para ciencias ambientales: Dolédec & Chessel (1991).

Referencias

- Benzécri, J. (1983), 'Analyse de l'inertie intraclasse par l'analyse d'un tableau de correspondances', *Les Cahiers de l'Analyse des Données* **8**(3), 351–358.
- Cabarcas, G. & Pardo, C.-E. (2001), 'Métodos estadísticos multivariados en investigación social', Cursillo, Simposio de Estadística - Santa Marta. Universidad Nacional. Departamento de Estadística.
*<http://es.geocities.com/socccad.colombia/documentos/cepardot/MetEstMulInvSoc.html>
- Cazes, P., Chessel, D. & Dolédec, S. (1988), 'L'analyse des correspondances internes d'un tableau partitionné. Son usage en hydrobiologie', *Revue de Statistique Appliquée* **36**(1), 39–54.
*<http://pbil.univ-lyon1.fr/R/articles/arti054.pdf>
- Cazes, P. & Moreau, J. (1991), 'Analysis of a contingency table in which the rows and columns have a graph structure', *Symbolic-numeric data analysis and learning* pp. 271–280. Eds. Diday E. y Lechevailler Y.
- Dolédec, S. & Chessel, D. (1991), 'Recent developments in linear ordination methods for environmental sciences', *Advances in Ecology* **1**, 133–155.
*<http://pbil.univ-lyon1.fr/R/articles/arti066.pdf>
- Duarte, R., Suarez, M., Moreno, E. & Ortiz, P. (1996), 'Análisis multivariado por componentes principales, de cafés tostados y molidos adulterados con cereales', *Cenicafé* **47**(2), 65–76.
- Escofier, B. (1984), 'Analyse factorielle en reference a un modele. Application a l'analyse de tableaux d'échanges', *Revue de Statistique Appliquée* **32**(4), 25–36.
- Escofier, B. & Pagès, J. (1992), *Análisis factoriales simples y múltiples. Objetivos, métodos e interpretación*, Universidad del País Vasco, Bilbao.
- Escofier, B. & Pagès, J. (1998), *Analyses factorielles simples et multiples. Objectifs, méthodes et interprétation*, 3 edn, Dunod, Paris.
- Lebart, L., Morineau, A. & Piron, M. (1995), *Statistique exploratoire multidimensionnelle*, Dunod, Paris.
- Lebart, L., Morineau, A. & Warwick (1984), *Multivariate Descriptive Statistical Analysis*, Wiley, New York.
- Montenegro, A. & Pardo, C. (1996), 'Introducción al análisis de datos textuales', Cursillo, Universidad Nacional de Colombia. Departamento de Matemáticas y Estadística. Bogotá.
*<http://es.geocities.com/socccad.colombia/documentos/cepardot/CursoTex.html>
- Ortiz, J. & Pardo, C. (2004), 'Análisis de datos multivariados en R', Cursillo, Simposio de Estadística - Cartagena. Universidad Nacional. Departamento de Estadística.
*<http://es.geocities.com/socccad.colombia/documentos/cepardot/AnalMultiR.html>
- R Development Core Team (2004), *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria.
*<http://www.R-project.org>
- Thioulouse, J., Dufour, A. & Chessel, D. (2004), *ADE4: Analysis of Environmental Data : Exploratory and Euclidean method Multivariate data analysis and graphical display*, Lyon, France.
*<http://cran.univ-lyon1.fr/src/contrib/Descriptions/ade4.html>