

11. ORDENACIÓN

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

EL PROBLEMA DE ORDENACIÓN

Ordenación y **escalado** describen el proceso de producir un **menor número de variables** que pueden utilizarse para describir la relación entre un grupo de objetos, **partiendo de una matriz de distancias o similitudes entre los objetos, o a partir de los valores de algunas variables medidas para cada objeto.**

Muchos de los métodos que se han descrito anteriormente se pueden usar para la ordenación, y algunos de los ejemplos se han ocupado de este proceso (**ACP** – pájaros s y no-s, empleo en Europa, **EMD** – mapa de la isla sur de Nueva Zelanda, son ejemplos de ordenación).

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

EL PROBLEMA DE ORDENACIÓN

Además, el **AFD** puede considerarse como un tipo de ordenación diseñada para enfatizar las diferencias entre objetos en diferentes grupos, mientras que el **ACC** refleja un tipo de ordenación diseñada para enfatizar las relación entre dos grupos de variables medidas en los mismos objetos.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

EL PROBLEMA DE ORDENACIÓN

Aunque la ordenación cubre una amplia gama de situaciones, en **biología** se utiliza para ***sintetizar las relaciones entre diferentes especies a partir de su abundancia en diferentes sitios*** o, alternativamente, para **sintetizar las relaciones entre diferentes sitios en base a la abundancia de diferentes especies en esos sitios.**

Esto es válido para **calidad de aguas**, sustituyendo especies por elemento físico-químico y abundancia por concentraciones. O para **aguas subterráneas**, sustituyendo especies por parámetros hidrogeológicos y abundancia por valores.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

EL PROBLEMA DE ORDENACIÓN

Objetivo: Profundizar en este tipo de aplicaciones.

Más ejemplos del uso de **ACP** y **EMD** en este contexto, y

describir los métodos de **Análisis de Coordenadas**

Principales (ACoP) y Análisis de Correspondencias (AC).

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

RECORDATORIO ACP

ACP: Los valores de las variables X_1, X_2, \dots, X_p , medidos en cada uno de los n objetos, se utilizan para construir los C.P. Z_1, Z_2, \dots, Z_p , que son combinaciones lineales de las variables X y son tales que Z_1 tiene la máxima varianza posible, Z_2 tiene la mayor varianza posible con la condición de no estar correlacionada con Z_1 , Z_3 tiene la máxima varianza posible con la condición que no esté correlacionada con Z_1 y Z_2 , y así sucesivamente. **Objetivo:** Reemplazar las variables X por un número menor de C.P., con poca pérdida de información.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

RECORDATORIO ACP

ACP en términos de ordenación: Se puede esperar que los dos primeros C.P. sean suficientes para describir las diferencias entre los objetos, puesto que graficar Z_2 contra Z_1 , indicando los valores de Z_3 puede ser aceptable.

Si se requieren 4 o más C.P., luego, obviamente, no se obtiene una buena ordenación, por lo menos en lo que concierne a la representación gráfica.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

EJEMPLO 11.1 Reserva Natural Steneryd, Suecia

Ejemplo 11.1: La Tabla a continuación muestra la abundancia de 25 especies de plantas en 17 parcelas de un prado empastado en la Reserva Natural Steneryd de Suecia.

Consideremos una **ordenación de las parcelas**. En este caso,

Variabes para el ACP: abundancia de especies

Objetos: parcelas

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.


EJEMPLO 11.1 Reserva Natural Steneryd, Suecia

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Festuca ovina	38	43	43	30	10	11	20	0	0	5	4	1	1	0	0	0	0
Anemone nemorosa	0	0	0	4	10	7	21	14	13	19	20	19	6	10	12	14	21
Stallaria holostea	0	0	0	0	0	6	8	21	39	31	7	12	0	16	11	6	9
Agrostis tenuis	10	12	19	15	16	9	0	9	28	8	0	4	0	0	0	0	0
Ranunculus ficaria	0	0	0	0	0	0	0	0	0	0	13	0	0	21	20	21	37
Mercurialis perennis	0	0	0	0	0	0	0	0	0	0	1	0	0	0	11	45	45
Proa pratensis	1	0	5	6	2	8	10	15	12	15	4	5	6	7	0	0	0
Rumex acetosa	0	7	0	10	9	9	3	9	8	9	2	5	5	1	7	0	0
Veronica chamaedrys	0	0	1	4	6	9	9	9	11	11	6	5	4	1	7	0	0
Dactylis glomerata	0	0	0	0	0	8	0	14	2	14	3	9	8	7	7	2	1
Fraxinus excelsior (juv.)	0	0	0	0	0	8	0	0	6	5	4	7	9	8	8	7	6
Saxifraga granulata	0	5	3	9	12	9	0	1	7	4	5	1	1	1	3	0	0
Deschampsia flexuosa	0	0	0	0	0	0	30	0	14	3	8	0	3	3	0	0	0
Luzula campestris	4	10	10	9	7	6	9	0	0	2	1	0	2	0	1	0	0
Plantago lanceolata	2	9	7	15	13	8	0	0	0	0	0	0	0	0	0	0	0
Festuca rubra	0	0	0	0	15	6	0	18	1	9	0	0	2	0	0	0	0
Hieracium pilosella	12	7	16	8	1	6	0	0	0	0	0	0	0	0	0	0	0
Geum urbanum	0	0	0	0	0	7	0	2	2	1	0	7	9	2	3	8	7
Lathyrus montanus	0	0	0	0	0	7	9	2	12	6	3	8	0	0	0	0	0
Campanula persicifolia	0	0	0	0	2	6	3	0	6	5	3	9	3	2	7	0	0
Viola riviniana	0	0	0	0	0	4	1	4	2	9	6	8	4	1	6	0	0
Hepatica nobilis	0	0	0	0	0	8	0	4	0	6	2	10	6	0	2	7	0
Achillea millefolium	1	9	16	9	5	2	0	0	0	0	0	0	0	0	0	0	0
Allium sp.	0	0	0	0	2	7	0	1	0	3	1	6	8	2	0	7	4
Trifolium repens	0	0	6	14	19	2	0	0	0	0	0	0	0	0	0	0	0

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

EJEMPLO 11.1 Reserva Natural Steneryd, Suecia

Valores propios

Puesto que el nro. de variables > nro. parcelas, el nro. de valores propios en la matriz de correlación ( datos estandarizados) lo determina el nro. de parcelas.

En realidad, son **16 valores propios no nulos**.

Los primeros **3 C.P.** representan el **69% de la varianza**.

Valores propios:	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16
Valor propio	8,79	5,59	2,95	1,93	1,58	1,13	0,99	0,54	0,40	0,35	0,20	0,18	0,13	0,12	0,07	0,05
Variabilidad	35,17	22,34	11,82	7,72	6,32	4,52	3,97	2,18	1,61	1,39	0,78	0,70	0,51	0,46	0,30	0,21
% acumulada	35,17	57,51	69,33	77,04	83,37	87,89	91,86	94,04	95,64	97,04	97,82	98,53	99,03	99,50	99,79	100,00

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

EJEMPLO 11.1 Reserva Natural Steneryd, Suecia

Valores propios

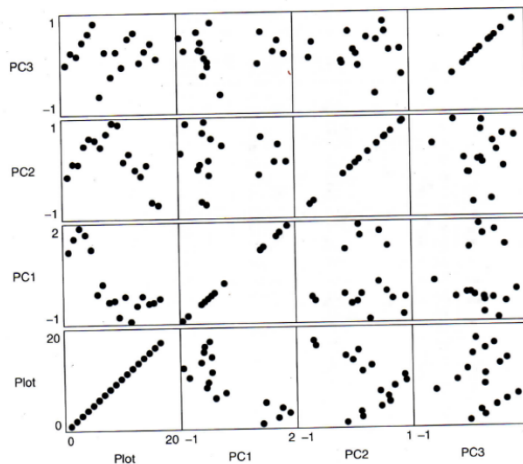
La tabla muestra los coeficientes de los 3 primeros C.P. Muestran contrastes entre la abundancia de diferentes especies que pueden ser significativos para un botánico, pero que no discutiremos en este contexto.

Vectores propios:			
	F1	F2	F3
Festuca ovina	0,30	0,00	0,07
Anemone nemorosa	-0,25	0,02	0,18
Stallaria holostea	-0,20	0,20	0,19
Agrostis tenuis	0,17	0,24	-0,01
Ranunculus ficaria	-0,11	-0,32	0,07
Mercurialis perennis	-0,08	-0,31	-0,02
Proa pratensis	-0,11	0,32	0,11
Rumex acetosa	-0,01	0,34	-0,23
Veronica chamaedrys	-0,15	0,36	0,06
Dactylis glomerata	-0,23	0,15	-0,18
Fraxinus excelsior (juv.)	-0,25	-0,11	-0,17
Saxifraga granulata	0,13	0,24	-0,23
Deschampsia flexuosa	-0,05	0,12	0,44
Luzula campestris	0,28	0,09	0,00
Plantago lanceolata	0,27	0,11	-0,26
Festuca rubra	-0,03	0,23	-0,19
Hieracium pilosella	0,27	-0,02	-0,05
Geum urbanum	-0,20	-0,18	-0,29
Lathyrus montanus	-0,15	0,26	0,19
Campanula persicifolia	-0,21	0,18	-0,07
Viola riviniana	-0,23	0,16	-0,11
Hepatica nobilis	-0,21	0,03	-0,34
Achillea millefolium	0,29	0,03	-0,10
Allium sp.	-0,18	-0,12	-0,36
Trifolium repens	0,21	0,11	-0,22

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

EJEMPLO 11.1 Reserva Natural Steneryd, Suecia

La figura muestra "la trama del dibujante" del nro de parcelas (1-17) y los 3 C.P. El primer C.P. (PC1) está estrechamente relacionado con el nro de la parcela. Esto refleja que las parcelas están ordenadas según la abundancia de especies con una alta respuesta a la luz y una baja respuesta a la humedad, la reacción del suelo y el nitrógeno. Por lo tanto, el análisis es al menos capaz de detectar esta tendencia.



AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

ANÁLISIS DE COORDENADAS PRINCIPALES

El análisis de coordenadas principales (ACoP) es **similar** al escalado multidimensional. Ambos métodos parten de una matriz de similitudes o distancias entre varios objetos y se esfuerzan por encontrar ejes de ordenación. Sin embargo, **difieren en el enfoque numérico** que se usa. El **análisis de coordenadas principales** utiliza un enfoque de **valor propio** que puede considerarse como una generalización de ACP. Sin embargo, el **escalado multidimensional**, al menos como se define en este curso, intenta en cambio **minimizar el estrés**, donde éste es una medida de si las posiciones de los objetos en una configuración t-D coinciden con las distancias o similitudes originales, después de un escalado apropiado.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

ANÁLISIS DE COORDENADAS PRINCIPALES

Conexión entre análisis de coordenadas principales (ACoP) y análisis de componentes principales (ACP)

Recordemos la teoría del ACP e introduzcamos conceptos adicionales:

1. El *i*-ésimo C.P. es una combinación lineal

$$Z_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p$$

de las variables X_1, X_2, \dots, X_p que se miden en cada uno de los objetos que se consideran. Hay **p** de estos componentes, y los coeficientes **a_{ij}** están dados por el vector propio **a_i** correspondiente al *i*-ésimo valor propio **λ_i** de la matriz de covarianza **C** de la muestra de las variables **X**. Es decir, la ecuación

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

ANÁLISIS DE COORDENADAS PRINCIPALES

$$\mathbf{C}\mathbf{a}_i = \lambda_i \mathbf{a}_i \quad \text{o} \quad (\mathbf{C} - \lambda_i \mathbf{I}) \mathbf{a}_i = \mathbf{0} \quad [1]$$

Donde $\mathbf{a}_i' = (a_{i1}, a_{i2}, \dots, a_{ip})$. Además, la varianza de Z_i es $\text{Var}(Z_i) = \lambda_i$ donde esto es cero si Z_i corresponde a una combinación lineal de las variables X que es constante.

2. Si las variables originales X están codificadas para tener media cero, entonces la matriz de covarianza \mathbf{C} ($p \times p$) tiene la forma:

$$\mathbf{C} = \frac{1}{(n-1)} \begin{bmatrix} \sum x_{i1}^2 & \sum x_{i1}x_{i2} & \dots & \sum x_{i1}x_{ip} \\ \sum x_{i2}x_{i1} & \sum x_{i2}^2 & \dots & \sum x_{i2}x_{ip} \\ \vdots & \vdots & \ddots & \vdots \\ \sum x_{ip}x_{i1} & \sum x_{ip}x_{i2} & \dots & \sum x_{ip}^2 \end{bmatrix}$$

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

ANÁLISIS DE COORDENADAS PRINCIPALES

donde n son los objetos, x_{ij} es el valor de X_j para el i -ésimo objeto, y las sumatorias para i son de 1 a n . De aquí:

$$\mathbf{C} = \frac{\mathbf{X}'\mathbf{X}}{n-1} \quad [2]$$

donde

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}$$

es la matriz de datos originales.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.