

ANÁLISIS DE CORRESPONDENCIAS

A continuación se enumeran algunas **denominaciones alternativas** con que el método de análisis de correspondencias (ACorr) aparece citado en la literatura:

- Análisis de tablas de contingencias (*Contingency table analysis*)
- Metodología R-Q (*RQ-technique*)
- Análisis de homogeneidad (*Homogeneity analysis*)
- Ponderación recíproca (*reciprocal averaging*)
- Ordenamiento recíproco (*reciprocal ordering*)
- Escalado dual (*Dual scaling*)

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

ANÁLISIS DE CORRESPONDENCIAS

El método puede explicarse en el contexto de la ordenación de p sitios en función de la abundancia de n especies, aunque puede utilizarse igualmente para datos que pueden presentarse como una tabla bidireccional, con las filas correspondientes a un tipo de clasificación y las columnas a un segundo tipo de clasificación.

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

ANÁLISIS DE CORRESPONDENCIAS

Con sitios y especies, la situación es la que se muestra a continuación:

Species	Site				Row Sum	Species Value
	1	2	...	p		
1	x_{11}	x_{12}	...	x_{1p}	R_1	a_1
2	x_{21}	x_{22}	...	x_{2p}	R_2	a_2
.
.
.
n	x_{n1}	x_{n2}	...	x_{np}	R_n	a_n
Column sum	C_1	C_2	...	C_p		
Site value	b_1	b_2	...	b_p		

Se pueden entender como una masa o perfil medio asociado a la fila o columna

a_1, a_2, \dots, a_n valores de especies asociados con las filas

b_1, b_2, \dots, b_p valores de sitios asociados con las columnas

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

ANÁLISIS DE CORRESPONDENCIAS

Una interpretación del análisis de correspondencia es que **se trata de elegir las especies y los valores del sitio para que estén lo más correlacionados posible con la distribución bivariada que se representa por las abundancias en el cuerpo de la tabla.**

Es decir, **los valores de sitio y especie se eligen para maximizar su correlación para la distribución donde la cantidad de veces que la especie i ocurre en el sitio j es proporcional a la abundancia observada x_{ij} .**

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

ANÁLISIS DE CORRESPONDENCIAS

Resulta que la solución a este problema de maximización está dada por el conjunto de ecuaciones:

$$a_1 = \left\{ (x_{11}/R_1)b_1 + (x_{12}/R_1)b_2 + \dots + (x_{1p}/R_1)b_p \right\} / r$$

$$a_2 = \left\{ (x_{21}/R_2)b_1 + (x_{22}/R_2)b_2 + \dots + (x_{2p}/R_2)b_p \right\} / r$$

$$\vdots$$

$$a_n = \left\{ (x_{n1}/R_n)b_1 + (x_{n2}/R_n)b_2 + \dots + (x_{np}/R_n)b_p \right\} / r$$

$$b_1 = \left\{ (x_{11}/C_1)a_1 + (x_{21}/C_1)a_2 + \dots + (x_{n1}/C_1)a_n \right\} / r$$

$$b_2 = \left\{ (x_{12}/C_2)a_1 + (x_{22}/C_2)a_2 + \dots + (x_{n2}/C_2)a_n \right\} / r$$

$$\vdots$$

$$b_p = \left\{ (x_{1p}/C_p)a_1 + (x_{2p}/C_p)a_2 + \dots + (x_{np}/C_p)a_n \right\} / r$$

R_i abundancia total de especies i ; C_j abundancia total en el sitio j
 r es la correlación máxima que se busca.

Así, el **valor de la i -ésima especie a_i** es un promedio ponderado de los valores de los distintos sitios, con el **sitio j** teniendo un peso que es proporcional a x_{ij}/R_i , y el valor del **j -ésimo sitio b_j** es un promedio ponderado de los valores de la especie, con **especies i** que tienen un peso que es proporcional a x_{ji}/C_j .

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

ANÁLISIS DE CORRESPONDENCIAS

"**Promedio recíproco**" es un término que se usa para describir estas ecuaciones, puesto que los valores de la especie son promedios (ponderados) de los valores del sitio, y los valores del sitio son promedios (ponderados) de los valores de la especie.

Estas ecuaciones se utilizan a menudo como *punto de partida para justificar el análisis de correspondencia*, como medio para producir valores de especies en función de los valores del sitio, y viceversa.

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

ANÁLISIS DE CORRESPONDENCIAS

Estas ecuaciones se pueden resolver de forma iterativa después de haber sido modificadas para eliminar la **solución trivial** con **$a_i = 1$ para todo i , $b_j = 1$ para todo j , y $r = 1$.**

Sin embargo, es más conveniente expresar las ecuaciones en forma matricial, puesto que muestra que puede haber varias soluciones posibles a las ecuaciones y que éstas se pueden encontrar a partir de un análisis de valores propios.

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

ANÁLISIS DE CORRESPONDENCIAS

$$\mathbf{a} = \mathbf{R}^{-1}\mathbf{X}\mathbf{b}/r \quad [8]$$

$$\mathbf{b} = \mathbf{C}^{-1}\mathbf{X}'\mathbf{a}/r \quad [9]$$

$$\mathbf{a}' = (a_1, a_2, \dots, a_n)$$

$$\mathbf{b}' = (b_1, b_2, \dots, b_p)$$

R matriz diagonal $n \times n$, con R_i en la i -ésima fila e i -ésima columna

C matriz diagonal $p \times p$, con C_j en la j -ésima fila e j -ésima columna

X matriz $n \times p$ con x_{ij} en el i -ésima fila y j -ésima columna

Sustituyendo [9] en [8] se llega a

$$r^2(\mathbf{R}^{1/2}\mathbf{a}) = (\mathbf{R}^{-1/2}\mathbf{X}\mathbf{C}^{-1/2}) (\mathbf{R}^{-1/2}\mathbf{X}\mathbf{C}^{-1/2})'(\mathbf{R}^{1/2}\mathbf{a})$$

$\mathbf{R}^{1/2}$ matriz diagonal, con $\sqrt{R_i}$ en la i -ésima fila e i -ésima columna

$\mathbf{C}^{1/2}$ matriz diagonal, con $\sqrt{C_j}$ en la j -ésima fila y j -ésima columna

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

ANÁLISIS DE CORRESPONDENCIAS

Las soluciones al problema de maximizar la correlación están dadas por los valores propios de la matriz $n \times n$

$$(\mathbf{R}^{-1/2}\mathbf{X}\mathbf{C}^{-1/2}) (\mathbf{R}^{-1/2}\mathbf{X}\mathbf{C}^{-1/2})'$$

Para cualquier **valor propio** λ_k , **la correlación entre las especies y los sitios será** $r_k = \sqrt{\lambda_k}$ **y el vector propio** para esta correlación será

$$\mathbf{R}^{1/2}\mathbf{a}_k = (\sqrt{R_1}a_{1k}, \sqrt{R_2}a_{2k}, \dots, \sqrt{R_n}a_{nk})'$$

a_{ik} son los valores de las especies y los correspondientes valores de los sitios: $\mathbf{b}_k = \mathbf{C}^{-1}\mathbf{X}'\mathbf{a}_k/r_k$

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

ANÁLISIS DE CORRESPONDENCIAS

El valor propio más grande siempre será $r^2 = 1$, dando la solución trivial $a_i = 1$ para todos los i y $b_j = 1$ para todos los j .

Los valores propios restantes serán positivos o nulos y **reflejarán diferentes dimensiones posibles para representar las relaciones entre las especies y los sitios.**

Se puede demostrar que **estas dimensiones son ortogonales**, puesto que las especies y los valores del sitio para una dimensión no estarán correlacionados con las especies y los valores del sitio en otras dimensiones para la distribución de datos de las abundancias x_{ij} .

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

ANÁLISIS DE CORRESPONDENCIAS

Ordenación por medio de análisis de correspondencias implica utilizar las especies y los valores del sitio para los primeros mayores valores propios menores que 1, porque estas son las soluciones para las cuales las correlaciones entre los valores de las especies y los valores del sitio son más fuertes.

Se suele representar las especies y los sitios en los mismos ejes puesto que los valores de las especies son un promedio de los valores del sitio y viceversa. En otras palabras, **el análisis de correspondencia proporciona una ordenación de las especies y los sitios al mismo tiempo.**

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

ANÁLISIS DE CORRESPONDENCIAS

Ej. de uso del ACorr - Especies de plantas en la Reserva Natural Steneryd:

Se obtuvo 16 valores propios menores que 1. Sus raíces cuadradas representan las correlaciones existentes entre los valores de las especies y los valores de la parcela. Solo los valores de especie y parcela para los dos primeros valores propios se usarán para la ordenación.

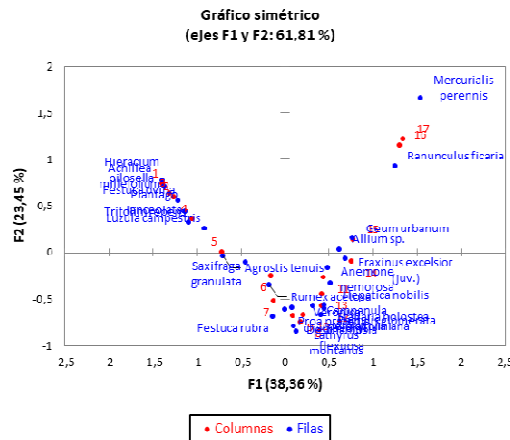
XLSSTAT 2015.2.01.17502 - Análisis Factorial de Correspondencias (AFC) - el 13/06/2018 a las 8:48:09
Tabla de contingencia: Libro = Ejemplo 1.12.xlsm / Hoja = Datos / Rango = Datos!\$A\$1:\$R\$26 / 25 filas y 17 columnas

Valores propios y porcentajes de inercia:																
	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16
Valor propio	0,665	0,406	0,199	0,136	0,094	0,074	0,057	0,028	0,020	0,019	0,010	0,008	0,007	0,005	0,003	0,001
Inercia (%)	38,362	23,449	11,468	7,873	5,439	4,293	3,269	1,638	1,143	1,086	0,551	0,475	0,394	0,295	0,198	0,068
%acumulad	38,362	61,810	73,278	81,152	86,591	90,884	94,152	95,790	96,933	98,019	98,570	99,045	99,439	99,734	99,932	100,000

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

ANÁLISIS DE CORRESPONDENCIAS

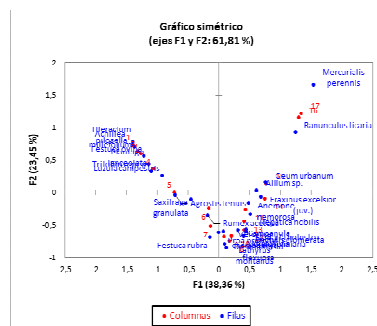
Ej. de uso del
ACorr - Especies
de plantas en la
Reserva Natural
Steneryd:



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

ANÁLISIS DE CORRESPONDENCIAS

La ordenación de los sitios es bastante clara, con una secuencia casi perfecta desde la parcela 1 a la izquierda a la parcela 17 a la derecha, moviéndose alrededor del arco. Las especies se intercalan entre los sitios de las parcelas a lo largo del mismo arco, desde *Mercurialis perennis* a la derecha hasta *Hieracium pilosella* a la izquierda. Por ejemplo, *Mercurialis perennis* es abundante solo en las parcelas con el número más alto y *Hieracium pilosella* es abundante solo en las parcelas con número más bajo.



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

Características de los diagramas de AC

- Los **agrupamientos de observaciones** deben interpretarse como resultantes de un mismo proceso o pertenecientes a una misma familia.
- La **cercanía de las variables** debe interpretarse como una indicación de la correlación entre las mismas.
- Si se distinguen diferentes agrupamientos de observaciones, éstos se caracterizan por la(s) variable(s) cercana(s) a tales grupos.

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

Características de los diagramas de AC

La síntesis bidimensional de puntos que originalmente se encuentran en dos espacios multidimensionales puede resultar engañosa. **¿Quién puede garantizar que dos muestras o variables que en 2-D se encuentran próximas también lo estén en n-D o p-D?**

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

Enunciado de las contribuciones (Benzécri, 1970)

Las contribuciones son coeficientes que ayudan a determinar qué tan fidedignamente una muestra o una variable ha sido restituida en un espacio factorial limitado. Además, es posible determinar cuál es el aporte de las muestras y de las variables a un determinado factor.

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

Enunciado de las contribuciones (Benzécri, 1970)

Contribución absoluta (CA): Expresa la contribución de la observación i o de la variable j a la dispersión a lo largo del factor k . Permite detectar anomalías ("outliers") en la matriz de datos.

Contribución relativa (CR): Indica la contribución del factor k a la separación de la observación i o de la variable j con respecto a su situación promedio (el centro de gravedad de ambas nubes).

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

Ej.: Aplicación a datos de precipitación anual

Matriz de datos: **P**anual **n(filas) xp(columnas)**

n = observaciones = nro. de años con registros

p = variables = número de pluviómetros

Interpretación del gráfico 2-D:

Observaciones (precipitación anual):

Cada punto representa una observación (año) conteniendo toda la información (precipitación anual) registrada en p pluviómetros. Esto significa condensar p registros en un único punto.

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

Ej.: Aplicación a datos de precipitación anual

La distancia entre puntos que representan distintas observaciones (años) es una medida de similitud con respecto a la precipitación anual en una cuenca o región.

Variables (pluviómetros):

Cada punto representa una variable (pluviómetro), conteniendo toda la información (precipitación anual) registrada en n observaciones (años). Esto significa condensar n registros en un único punto.

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

Ej.: Aplicación a datos de precipitación anual

Relación entre las observaciones (precipitación anual) y las variables (pluviómetros):

En general, cada punto que representa una observación estará situado en una posición cercana al pluviómetro donde la precipitación registrada fue mayor, mientras que aquellos puntos ubicados en una posición opuesta indican que en dicho año el pluviómetro registró poca precipitación.

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

Ej.: Aplicación a datos de precipitación anual

Obs! La interpretación anterior de la birrepresentación en un espacio factorial limitado es válida solamente si se logra la restitución del espacio n- y p-dimensional. El grado de precisión de está restitución debe analizarse controlando las contribuciones (Benzecri, 1973 & 1977)

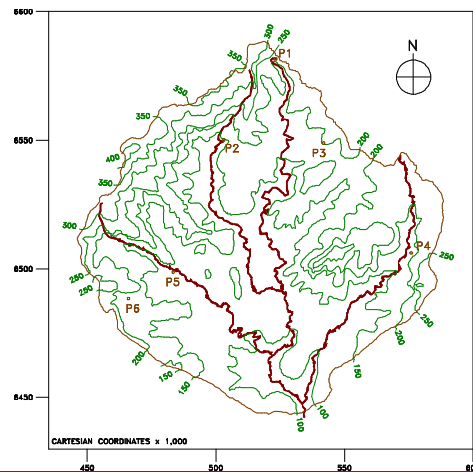
AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

Ej.: Detección de errores en datos de precipitación anual

Pluviómetro	Ubicación	Lectura	X	Y
P1 1147	Rivera	DNM	524.0	6582.5
P2 1220	Tranqueras	MI	503.0	6550.0
P3 1224	Ataques	MI	542.0	6549.0
P4 1379	Moirones	MI	577.6	6504.0
P5 1405	Tacuarembó	DNM	482.5	6492.0
P6 1440	Valle Edén	MI	464.5	6480.0

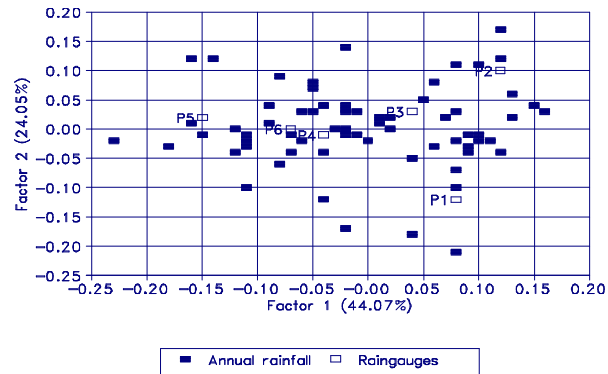
AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

Ej.: Detección de errores en datos de precipitación anual



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

Ej.: Detección de errores en datos de precipitación anual



Varianza explicada por los factores 1 y 2: 68.12 %

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

Ej.: Detección de errores en datos de precipitación anual

Pluviómetros	FACTOR 1		FACTOR 2	
	CA(1)	CR(1)	CA(2)	CR(2)
P1	13	28	61	72
P2	30	62	33	38
P3	2.7	59	3.5	41
P4	3.2	92	0.5	8.1
P5	43	98	1.5	1.9
P6	7.8	100	0.0	0.0

La restitución de las variables en el plano de los dos primeros factores es satisfactoria puesto que la suma de las contribuciones relativas de cada variable es aproximadamente 100 %.

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

Ej.: Detección de errores en datos de precipitación anual

Las **contribuciones absolutas** indican el peso de cada variable sobre cada factor. Valores altos pueden representar pluviómetros donde la precipitación medida es mayor que la precipitación medida en otros pluviómetros o datos anómalos ("outliers") por exceso.

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

Ej.: Detección de errores en datos de precipitación anual

Interpretación de la birrepresentación:

- ❑ **P3 (Ataques), P4 (Moirones) y P6 (Valle Edén)** están situados cerca del centro de gravedad (0,0). Por lo tanto, representan el comportamiento medio de la cuenca.
- ❑ **P1 (Rivera), P2 (Tranqueras) y P5 (Tacuarembó)** se ubican lejos del centro de gravedad (0,0).

P1 (Rivera) - Cuchilla Negra

P2 (Tranqueras) - Cuchilla de Haedo

Por tanto, mayor precipitación puede explicarse por la influencia topográfica (altitud 300 a 400 m snm).

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

Ej.: Detección de errores en datos de precipitación anual

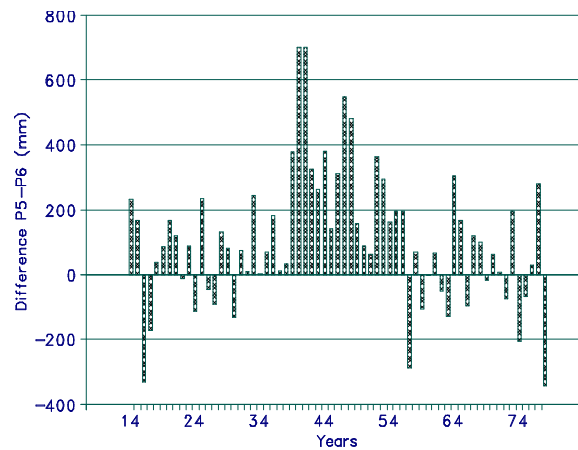
- **P5 (Tacuarembó) y P6 (Valle Edén)** están ubicados a una distancia de 21.6 km. Es decir, son pluviómetros relativamente cercanos.

No obstante, el **Análisis de Correspondencias** muestra un comportamiento diferente.

¿Cómo explicarlo?

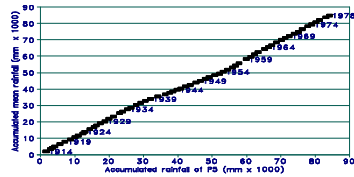
AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

Ej.: Detección de errores en datos de precipitación anual



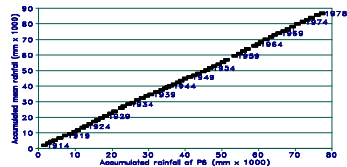
AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

Ej.: Detección de errores en datos de precipitación anual



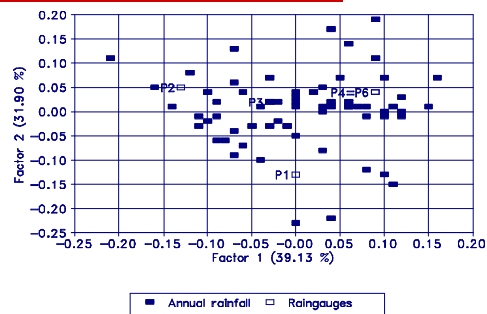
Análisis de doble masa:

- 1) P5 vs Pmedia (P1, P2, P3, P4, P6)
Cambio de pendiente en 1937 y 1952.
- 2) P6 vs Pmedia (P1, P2, P3, P4).



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

Ej.: Detección de errores en datos de precipitación anual



- **Análisis de correspondencias:** P1, P2, P3, P4, P6
- **Varianza explicada** por los factores 1 y 2: 71.03 %

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

Ej.: Detección de errores en datos de precipitación anual

- La representación gráfica sugiere una **variabilidad espacial** en la distribución de la precipitación media anual.

Pmedia anual (mm):

P1=1453, P2=1327, P3=1326, P4=1235, P5=1180

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

ANÁLISIS DE CORRESPONDENCIAS

Comparación de métodos de ordenación

ACP se utiliza solo cuando los valores de las p variables son conocidos para cada uno de los objetos que se estudian y las variables tienen una distribución aproximadamente normal. Por lo tanto, este método de análisis no puede utilizarse cuando solo se dispone de la matriz de distancia o similitud.

ACoP y **EMD** se utilizan cuando se requiere una ordenación a partir de una matriz de distancias o similitudes entre los objetos bajo estudio. ACoP y EMD métrica deberían dar resultados similares.

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

ANÁLISIS DE CORRESPONDENCIAS

Comparación de métodos de ordenación

ACorr fue inicialmente desarrollado para situaciones donde los objetos de interés se dividen por medidas de la abundancia de diferentes características. Cuando este es el caso, este método parece dar ordenaciones que son relativamente fáciles de interpretar. Actualmente sus aplicaciones son mucho más amplias.

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

ORDENACIÓN - RESUMEN

- ✓ **Ordenación** es el proceso de generar un menor número de variables para representar las relaciones entre una cantidad de objetos, generalmente gráficamente. También se utiliza el término **escalado** en lugar de ordenación.
- ✓ Muchos de los métodos descritos en el curso pueden usarse para la ordenación. Se analizaron ACP, ACoP, EMD y ACorr.

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

ORDENACIÓN - RESUMEN

- ✓ El análisis de coordenadas principales (ACoP) es un método de ordenación, que parte de una matriz de similitudes entre n objetos, de forma similar que el escalado multidimensional (EMD).

Se revisó la teoría del ACP y se mostro la conexión entre ACP y ACoP, y se utilizó ACoP con los datos de la Reserva Natural Steneryd.

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

ORDENACIÓN - RESUMEN

- ✓ Se aplicó el EMD con los datos de la Reserva Natural Steneryd.
- ✓ El análisis de correspondencia es el 4º método de ordenación analizado.
- ✓ Se hicieron recomendaciones sobre cuándo usar cada uno de los cuatro métodos de ordenación.

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