

CLASE 08 - INTRODUCCIÓN ANÁLISIS MULTIVARIANTE

OCE 313 - Técnicas de análisis no paramétricos.

Dr. José Gallardo Matus

Pontificia Universidad Católica de Valparaíso

22 May 2022

PLAN DE LA CLASE

1.- Introducción

- ▶ ¿Qué son los análisis mulvariantes?.
- ▶ Estudios de caso: Análisis de cluster, análisis de componentes principales.
- ▶ Matriz de distancia: cálculo manual

2). Práctica con R y Rstudio cloud.

- ▶ Matriz de distancia: cálculo con R

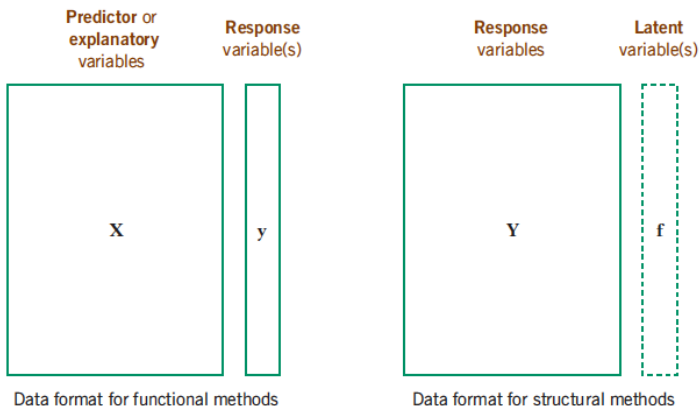
INTRODUCCIÓN ANÁLISIS MULTIVARIANTE

¿Qué son los análisis multivariantes?

Son un conjunto diverso de métodos que estudian y examinan el efecto simultáneo de múltiples variables.

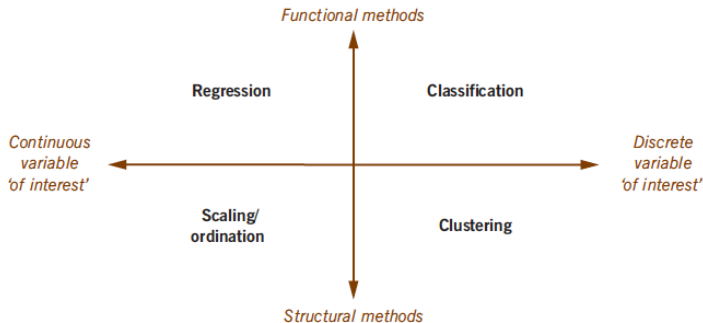
Sitio	a	b	c	d	e	Depth	Pollution	Temp.	Sediment
s1	0	2	9	14	2	72	4.8	3.5	S
s2	26	4	13	11	0	75	2.8	2.5	C
s3	0	10	9	8	0	59	5.4	2.7	C
s4	0	0	15	3	0	64	8.2	2.9	S
s5	13	5	3	10	7	61	3.9	3.1	C
s6	31	21	13	16	5	94	2.6	3.5	G

TIPOS DE MÉTODOS MULTIVARIANTES (MM)



Fuente: Multivariate Statistic, 2014

MÉTODOS MULTIVARIANTES SEGÚN TIPO DE VARIABLE



Fuente: Multivariate Statistic, 2014

EJEMPLO: ZOOGEOGRAFÍA DE CRUSTACEOS

Agrupamiento jerárquico.

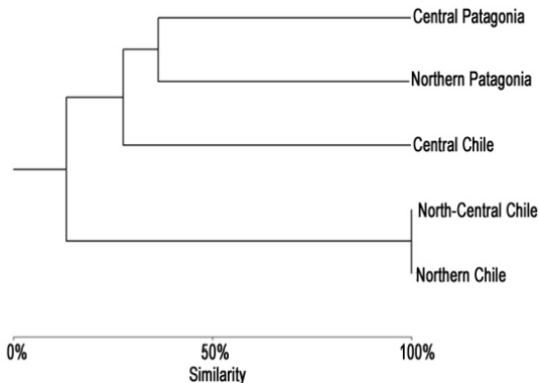
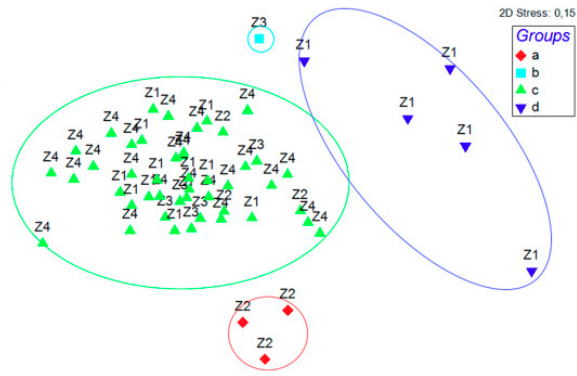


Figure 8. Dendrogram of inland water Copepoda considered in the present study.

Fuente: De los Ríos-Escalante et al, 2013

EJEMPLO: POLIQUETOS

Escalamiento multidimensional no métrico (NMDS).



Fuente: Sanchis, Soto y Quiroga, 2021

MATRIZ DE DISTANCIA O SIMILARIDAD

¿Qué es y para que sirve?

- Las matrices de distancia o similaridad están en la base de todos los análisis multivariados de estructura.

Algunas consideraciones

- Las matrices de distancia se pueden elaborar tanto para variables cuantitativas continuas, como discretas.
 - ▶ Debido a que las variables pueden tener diferente escala o magnitud es necesario muchas veces transformar o estandarizar las variables antes de calcular las matrices de distancia.
 - ▶ Cuando una variable tiene muchos ceros también es conveniente transformarla.

TIPOS DE MATRICES DE DISTANCIA

- ▶ **Euclidean** Para variables cuantitativas continuas.

Con base en el teorema de pitágoras

$$c^2 = a^2 + b^2$$

$$a = \sqrt{c^2 - b^2}$$

$$b = \sqrt{c^2 - a^2}$$

$$c = \sqrt{a^2 + b^2}$$

- ▶ **No euclidean**: Para variables cuantitativas discretas (conteos).

Diferentes alternativas: Bray-Curtis.

DISTANCIA EUCLIDEANA

- Evaluemos efecto de escala de las variables.

Sitio	Depth	Pollution	Temp.
s29	51	6.0	3.0
s30	99	1.9	2.9

$$s29 - s30 = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2}.$$

$$s29 - s30 = \sqrt{(51 - 99)^2 + (6.0 - 1.9)^2 + (3.0 - 2.9)^2}.$$

$$s29 - s30 = \sqrt{(2304) + (18.81) + (0.01)} = 48.17$$

ESTANDARIZACIÓN

	Depth	Pollution	Temperature
Mean	74,43	4,52	3,06
sd	15,61	2,14	0,28

Valor estandarizado : (valor original – mean) / sd

Valor estandarizado s29 : $(51 - 74,43) / 16,61 = -1,501$

Sitio	Depth	Pollution	Temperature
s29	-1,501	0,693	-0,201
s30	1,573	-1,222	-0,557

DISTANCIA EUCLIDEANA ESTANDARIZADA

Sitio	Depth	Pollution	Temperature
s29	-1,501	0,693	-0,201
s30	1,573	-1,222	-0,557

$$s_{29} - s_{30} = \sqrt{(-1,50 - 1,57)^2 + (0,69 - 1,22)^2 + (0,20 - 0,55)^2}.$$

Distancia estandarizada.

$$s_{29} - s_{30} = \sqrt{(9,499) + (3,667) + (0,127)} = 3,639.$$

Distancia no estandarizada.

$$s_{29} - s_{30} = \sqrt{(2304) + (18.81) + (0.01)} = 48.17$$

DISTANCIA NO EUCLIDEANA

Sitio	a	b	c	d	e
s29	11	0	7	8	0
s30	24	37	5	18	1

$$s_{29} - s_{30} = \frac{\sum |n_i - n_j|}{\sum n_i + \sum n_j}.$$

$$s_{29} - s_{30} = \frac{|11-24|+|0-37|+|7-5|+|8-18|+|0-1|}{26+85} = \frac{63}{111} = 0,568$$

RESUMEN DE LA CLASE

- ▶ Revisión e importancia de análisis multivariantes.
- ▶ Estudios de caso: analisis de cluster jerarquico (dendograma) y análisis de .
- ▶ Calculo de matriz de distancia manual
- ▶ ▶ Calculo de matriz de distancia con R