successione ecologica

Share your science. Make it better

INTRODUZIONE ALL'ANALISI MULTIVARIATA CON R

#iorestoacasa #acasaconse

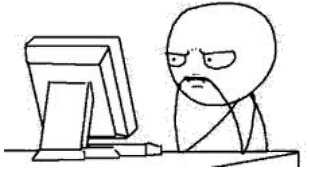
Marco Sigovini

CNR-ISMAR Venezia

marco.sigovini@ve.ismar.cnr.it

successione ecologica
Share your science. Make it better

- *R is a free software environment for statistical computing and graphics*
- Progetto *open source* GNU multipiattaforma (Windows, Linux, ...)
- Offre strumenti per la gestione, analisi e visualizzazione di dati
- Aggiornamento continuo tramite pacchetti sviluppati da una vasta comunità
- Consente flessibilità, controllo, riproducibilità, uso algoritmi iterativi e ricorsivi
- Supportato da documentazione e supporti didattici (manuali, tutorial, etc.)
- https://www.r-project.org/



(inizialmente un po' ostico...)

- R si basa su **oggetti** (definiti da **classi**) su cui si applicano **funzioni**
- **Classi**: specificano la struttura degli oggetti (numeric, factor, matrix, data.frame, list, ...)
- **Tipo di dati**: logical (T/F), integer, double (= numeric: numeri reali), character
- **Sintassi** segue (con alcune eccezioni) il modello:

```
>   a <- mean( c(1,3) )
```

(NB: *case sensitive*!)

- Gli elementi di un oggetto si estraggono con specifiche sintassi (uso di '[ ]', '$')
- Il simbolo '#' impedisce l'esecuzione della riga
- '?' seguito da una funzione apre una finestra di help

Analisi multivariata in Ecologia

Si rivolge all'osservazione ed all'analisi simultanea di più **variabili di interesse**

In ambito ecologico è considerata una parte della più vasta *Numerical Ecology* (Legendre & Legendre, 1979) e trova applicazione nell'Ecologia di comunità, nelle Scienze ambientali ed in altri settori dell'ecologia

Ha fondamentalmente gli stessi obiettivi della statistica classica:

- la descrizione di pattern
- l'inferenza sulla "popolazione" a partire da un "campione" (in senso statistico)
- la predizione di un fenomeno

Un obiettivo accessorio è la riduzione della dimensionalità dei dati originali per facilitare la comprensione dei fenomeni

Alcuni principali metodi multivariati

- Metodi di classificazione: ***hierarchical cluster analysis***, *K-means*, ...
- Metodi di ordinamento (→ *Gradient analysis)*:

	eigenvectors methods			*nonmetric*
unconstrained	**PCA**	**CA**	PCOA	**NMDS**
constrained	**RDA**	CCA	CAP/dbRDA	

- Test statistici multivariati: **ANOSIM**, **PERMANOVA**, …
- Identificazione variabili [ad es. specie] “rappresentative”, etc.

Tipologie di dati

per componente ecologica:

- dati di comunità (abbondanza, biomassa, copertura, presenza/assenza)
- variabili ambientali
- variabili spaziali (coordinate)

per forma matematica:

- variabili qualitative nominali, o v. categoriche (ad es. tipo di habitat)
- v. qualitative ordinali (ad es. classi di qualità)
- v. binarie (ad es. M/F, presenza/assenza, v. dummy)
- v. quantitative continue (ad es. misure di proprietà continue, biomasse)
- v. quantitative discrete, conteggi

per ruolo nell'analisi (a seconda del tipo di analisi):

- v. di risposta (*response*) VS predittive o esplicative (*predictive*, *explanatory*)
- v. dipendenti VS indipendenti
- fattori

successione ecologica
Share your science. Make it better

Tipologie di dati

Matrice di comunità (abbondanze)

	Brachy	PHTH	HPAV	RARD	SSTR	Protopl	MEGR	MPRO	TVIE	HMIN
1	17	5	5	3	2	1	4	2	2	1
2	2	7	16	0	6	0	4	2	0	0
3	4	3	1	1	2	0	3	0	0	0
4	23	7	10	2	2	0	4	0	1	2
5	5	8	13	9	0	13	0	0	0	3
6	19	7	5	9	3	2	3	0	0	20
7	17	3	8	2	3	0	3	0	0	19
8	5	4	8	2	1	2	3	0	0	1
9	3	3	2	2	1	1	12	0	0	0
10	22	4	5	3	0	0	0	0	0	11

Matrice delle variabili ambientali

	SubsDens	WatrCont	Substrate	Shrub	Topo
1	39.18	350.15	Sphagnl	Few	Hummock
2	54.99	434.81	Litter	Few	Hummock
3	46.07	371.72	Interface	Few	Hummock
4	48.19	360.50	Sphagnl	Few	Hummock
5	23.55	204.13	Sphagnl	Few	Hummock
6	57.32	311.55	Sphagnl	Few	Hummock
7	36.95	378.93	Sphagnl	Few	Hummock
8	80.59	266.78	Interface	Many	Blanket
9	61.43	310.70	Litter	Many	Blanket
10	32.14	220.73	Sphagnl	Many	Hummock

Procedure preliminari sui dataset

Analisi iniziale dei dati:

- analisi multivariate (generalmente) non parametriche non richiedono assunzioni sulla distribuzione; può essere comunque opportuno trattare eventuali outlayers e forti deviazioni dalla normalità ai fini dell'interpretazione

- analisi di collinearità (correlazione tra covariate)

Standardizzazione (→ variabili ambientali): stessa scala di variabilità, eliminate udm

Trasformazione (→ dati di comunità):

- in base al peso dato a specie rare e comuni (ad es. radq(x), log(x+1), eliminazione specie rare, ...)

- per poter applicare certe tecniche ai dati di comunità (ad es. trasf. di Hellinger)

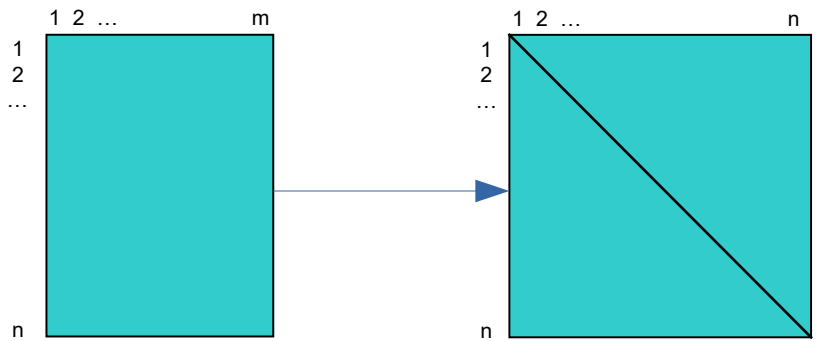
Coefficienti di dissimilarità

- Distanza* Euclidea d[jk] = sqrt(sum(x[ij]-x[ik])^2) [0,+inf[

  → variabili ambientali

- Bray-Curtis (*% difference*) d[jk] = (sum abs(x[ij]-x[ik]))/(sum (x[ij]+x[ik])) [0,1]

  → dati di comunità (quantitativi): l'indice evita che l'assenza congiunta di specie contribuisca alla similarità tra campioni (*double-zero problem*)



("*Q-mode*")

* distanze = dissimilarità che presentano proprietà metriche, inclusa diseguaglianza triangolare

Percorso di analisi