

Mauro Scimone, Paola Ganis e Enrico Feoli

PROGRAMMI BASIC PER IL CALCOLO DI MISURE DI DIVERSITA' IN
COMUNITA' ECOLOGICHE

Quaderni del
Gruppo Elaborazione Automatica Dati
Ecologia Quantitativa

Dipartimento di Biologia
Universita' di Trieste

GEAD-EQ n.5

Trieste, 1987



INDICE

Introduzione	pag. 3
I dati	pag. 3
Programmi	pag. 4
1. PERDIV	pag. 4
2. DIV	pag. 10
3. ALFADIV	pag. 14
4. BETADIV	pag. 18
5. PRODIV	pag. 22
6. INFORMA	pag. 28
Bibliografia	pag. 34

Ringraziamenti

Il lavoro è stato svolto nell'ambito di ricerche finanziate dal CNR (Progetto bilaterale Italia-Svezia - E. Feoli; Elaborazione Dati Flora e Vegetazione - Gruppo Biologia Naturalistica - D. Lausi), MPI 40 % e MPI 60% (E. Feoli e L. Poldini). Ringraziamo il prof. M. Lagonegro per la sua costante consulenza.

Introduzione

La misura della diversità delle comunità ecologiche può essere effettuata secondo diversi criteri e modelli (Hurlbert, 1971; Whittaker, 1972; Pielou, 1975; Grassle et Al., 1979). In questi ultimi anni sono stati prodotti molti lavori sulle applicazioni dei vari indici di diversità e sul significato della diversità stessa (vedi Giller, 1984). In questo quaderno presentiamo un insieme di programmi per l'analisi della diversità di comunità animali o vegetali basati principalmente sulla teoria dell'informazione. Viene presa in considerazione la funzione di Reyni (1961) che generalizza l'entropia di Shannon e la funzione di Patil e Taillie (1976) che generalizza l'indice di Gini-Simpson. Inoltre vengono considerate diverse funzioni informazionali per il calcolo della somiglianza o eterogeneità di un insieme di stati di comunità, che possono essere utili per il calcolo di ciò che impropriamente viene chiamata beta diversità da Whittaker (1972). Infatti la terminologia di Whittaker a proposito di questa misura, non solo esula dallo stesso concetto di diversità (Feoli, 1984), ma può creare confusione in quanto con 'beta' si indica il parametro per il calcolo dei profili di diversità di Patil & Taillie (1976).

Queste funzioni che sono state descritte da Feoli, Lagonegro e Orloci (1984) e di cui già esiste una versione di programmi FORTRAN, possono essere usate anche per produrre ordinamenti e classificazioni automatiche di comunità qualora vengano scelte le opzioni che producono matrici di distanza o somiglianza e vengano usati i programmi opportuni (es. Lagonegro e Feoli, 1985).

Il quaderno presenta 6 programmi (PERDIV, DIV, PRODIV, ALFADIV, BETADIV, INFORMA) per ciascuno dei quali vengono descritti i parametri di input e di output.

I dati

I programmi elaborano i dati organizzati in tabelle (matrici) nelle cui righe sono riportate le variabili (specie o categorie superiori di diversa natura, Feoli, 1984) e le cui colonne rappresentano gli stati delle comunità. Successivamente nella descrizione dei programmi parleremo sempre di righe e di colonne della matrice. I dati nelle matrici dovrebbero rappresentare delle frequenze o conte (numero di individui), in tal caso, alcuni programmi sarebbero utili per testare ipotesi mediante la statistica del chi-quadrato (Kullback, 1959; Feoli, Lagonegro e Orloci, 1984), ad ogni modo sono ammissibili dati numerici di diversa natura. La responsabilità del test è tutta del ricercatore.

Sono qui di seguito riportati alcuni simboli che ricorrono nelle formule delle funzioni informazionali riportate nel testo.

Σ = simbolo di sommatoria

m = numero di righe della tabella

n = numero di colonne della tabella

c = numero di righe o di colonne minore o uguale a m o n

i = indice di riga

j e k = indici di colonna

x = vettore riga
 y = vettore colonna
 ln = logaritmo naturale
 ** = simbolo di operazione esponenziale

Per un esempio di applicazione dei singoli programmi viene usata una matrice di 15 righe e 5 colonne (qui sotto riportata) che rappresentano 5 tipi di vegetazione dei boschi ad *Ostrya carpinifolia* della regione Friuli Venezia-Giulia già descritti in Feoli, Ganis e Poldini (1987, in stampa).

	A	B	C	D	E
1 Fanerofite cespitose	35.3	19.9	19.8	17.2	18.8
2 Fanerofite scapose	6.9	8.2	8.9	9.4	8.2
3 Fanerofite lianose	7.2	4.4	2.9	3.0	1.4
4 Camefite reptanti	0	0	0.7	0.9	0.2
5 Camefite suffruticose	2.1	4.4	1.8	0.4	7.0
6 Camefite fruticose	0	0.6	0.2	0.5	2.1
7 Terofite scapose	1.2	1.5	1.0	0	0.6
8 Geofite radicegemma	0	0.7	1.4	0.5	0.7
9 Geofite bulbose	3.0	0.8	3.5	3.7	4.0
10 Geofite rizomatose	8.8	13.3	16.3	21.0	9.3
11 Emicriptofite cespitose	7.8	10.2	10.0	6.7	12.6
12 Emicriptofite reptanti	0	0.7	1.4	1.6	0.5
13 Emicriptofite scapose	11.5	23.9	23.2	26.1	25.5
14 Emicriptofite rosulate	1.8	5.7	3.9	4.4	1.9
15 Nanofanerofite	14.3	5.8	5.0	4.6	7.1

Programmi

1. PERDIV

Questo programma elabora una matrice $m \times n$ calcolando i seguenti valori:

- totali di riga e colonna
- totale generale
- entropia delle righe (1.1) e delle colonne (1.2) secondo la formula di Shannon:

$$h(x) = - \sum_{j=1, \dots, n} p(j) \ln p(j) \quad (1.1)$$

$$h(y) = - \sum_{i=1, \dots, m} p(i) \ln p(i) \quad (1.2)$$

dove $p(j)$ e' il rapporto tra un valore di riga e il corrispondente totale di riga e $p(i)$ rappresenta il rapporto tra un valore di colonna e il suo totale corrispondente ;

- entropia massima di righe e colonne data dal logaritmo naturale del numero c di valori di riga o colonna maggiori di zero ;

$$H_{\max} = \ln c \quad (2)$$

e) entropia relativa o "evenness" data dal rapporto tra (1) e (2)

f) entropia dei totali marginali di riga (H_x) e colonna (H_y) usando la formula di Shannon :

$$H_x = -\sum p(i.) \ln p(i.) \quad i=1, \dots, m \quad (3.1)$$

$$H_y = -\sum p(.j) \ln p(.j) \quad j=1, \dots, n \quad (3.2)$$

dove $p(i.)$ e' il rapporto tra un totale di riga ($F_{i.}$) e il totale generale ($F_{..}$) e $p(.j)$ e' il rapporto tra un totale di colonna ($F_{.j}$) e il totale generale .

g) entropia congiunta:

$$H_{xy} = -\sum p(i,j) \ln p(i,j) \quad \begin{matrix} i=1, \dots, m \\ j=1, \dots, n \end{matrix} \quad (4)$$

dove $p(i,j)$ e' il rapporto tra il valore che c' e' nella riga i-esima e colonna j-esima ed il totale generale

h) mutua entropia :

$$H(x;y) = H_x + H_y - H_{xy} \quad (5)$$

i) il chi quadrato sulla base della precedente:

$$2I = 2 F_{..} H(x;y) \quad (6)$$

l) i corrispondenti gradi di liberta' $(m - 1) (n - 1)$.

PERDIV tratta i dati considerando opzionalmente il loro effettivo valore o trasformandoli in dati binari; e' inoltre possibile su opzione salvare su files le matrici dei valori percentuali rispetto alle righe, alle colonne e al totale generale secondo richiesta.

Esempio di applicazione

Il programma PERDIV richiede in input:

- 1) nome del file della tabella dati
- 2) numero righe, numero colonne della tabella dati, opzione dati binari (0=non binari; 1=binari)
- 3) nomi dei files per le tabelle delle percentuali rispetto ai totali di riga, ai totali di colonna, al totale generale. Se non si desidera calcolare ne' salvare una o piu' di queste tabelle, invece del nome, si scrive il carattere '.' punto. Per esempio se a questa richiesta si risponde scrivendo: '.,TABCOL,.' il programma calcola e salva solo la tabella delle percentuali di colonna con il nome TABCOL; se si risponde 'TABRIG,TABCOL,TABGEN' il programma calcola e salva tutte e tre le matrici ; infine rispondendo '.,.,.' non viene stampata ne' salvata alcuna tabella di valori percentuali.

```

RUN
Nome tabella? %bio5
N.righe,n.colonne,opz.dat binari (0 o 1)? 15,5,0
Nomi tabelle % riga, % colonna, % totale generale
(se non servono scrivere invece del nome il punto )? prova1,.,prova3

```

Programma Entropia e Percentuali

elaborazione sul file %bio5 - dati quantitativi
dimensioni tabella : 15 , 5

	tot.col.	h	Hmax	h/Hmax
1	99.90001	1.993906	2.397895	.8315234
2	100.1	2.173515	2.639058	.8235954
3	100	2.182226	2.70805	.8058291
4	100	2.074332	2.639058	.7860125
5	99.89999	2.174063	2.70805	.8028148

	tot.rig.	h	Hmax	h/Hmax
1	111	1.569658	1.609438	.9752832
2	41.6	1.60421	1.609438	.9967519
3	18.9	1.479532	1.609438	.9192851
4	1.8	.9580003	1.098612	.8720094
5	15.7	1.327546	1.609438	.8248508
6	3.4	1.052272	1.386294	.7590538
7	4.3	1.337571	1.386294	.9648538
8	3.3	1.307516	1.386294	.9431732
9	15	1.515517	1.609438	.9416436
10	68.70001	1.555438	1.609438	.9664481
11	47.30001	1.58579	1.609438	.9853068
12	4.2	1.285842	1.386294	.9275386
13	110.2	1.575166	1.609438	.9787057
14	17.7	1.516216	1.609438	.9420783
15	36.8	1.507101	1.609438	.9364144

tabella percentuali su tot. di riga (su file prova1)

31.8	17.9	17.8	15.5	16.9
16.6	19.7	21.4	22.6	19.7
38.1	23.3	15.3	15.9	7.4
0.0	0.0	38.9	50.0	11.1
13.4	28.0	11.5	2.5	44.6
0.0	17.6	5.9	14.7	61.8
27.9	34.9	23.3	0.0	14.0
0.0	21.2	42.4	15.2	21.2
20.0	5.3	23.3	24.7	26.7
12.8	19.4	23.7	30.6	13.5
16.5	21.6	21.1	14.2	26.6
0.0	16.7	33.3	38.1	11.9
10.4	21.7	21.1	23.7	23.1
10.2	32.2	22.0	24.9	10.7
38.9	15.8	13.6	12.5	19.3

tot.generale = 499.9
 entropia righe (Hx) = 2.186717
 entropia colonne (Hy) = 1.609438
 entropia tabella (Hxy) = 3.729072
 mutua entropia (H(x;y)=Hx+Hy-Hxy) = .0670836

 2I = 67.07018
 gradi di liberta' = 56

tabella percentuali sul totale generale (su file prova3)

```

-----
 7.1  4.0  4.0  3.4  3.8
 1.4  1.6  1.8  1.9  1.6
 1.4  0.9  0.6  0.6  0.3
 0.0  0.0  0.1  0.2  0.0
 0.4  0.9  0.4  0.1  1.4
 0.0  0.1  0.0  0.1  0.4
 0.2  0.3  0.2  0.0  0.1
 0.0  0.1  0.3  0.1  0.1
 0.6  0.2  0.7  0.7  0.8
 1.8  2.7  3.3  4.2  1.9
 1.6  2.0  2.0  1.3  2.5
 0.0  0.1  0.3  0.3  0.1
 2.3  4.8  4.6  5.2  5.1
 0.4  1.1  0.8  0.9  0.4
 2.9  1.2  1.0  0.9  1.4
  
```

```

10 REN Programma Entropia e Percentuali. Calcola l'entropia delle righe
20 REN delle colonne e la mutua entropia di una matrice. Calcola i
30 REN totali di riga, di colonna e il totale generale e trasforma i valori
40 REN della tabella in valori percentuali sui totali suddetti. (P.Ganis,1986)
50 INPUT "Nome tabella";F$
60 INPUT "N.righe,n.colonne,opz.dat binari (0 o 1)" ;NR,NC,B
70 DIM TB(NR,NC),SR(NR),SC(NC),TC(NC),TR(NR),EC(NC),ER(NR)
80 OPEN "I",#1,F$
90 FOR I=1 TO NR:FOR J=1 TO NC:INPUT #1,TB(I,J)
100 IF B THEN IF TB(I,J) THEN TB(I,J)=1
110 NEXT J,I
120 CLOSE #1
130 PRINT "Nomi tabelle X riga, X colonna, X totale generale"
140 INPUT (se non servono scrivere invece del nome il punto)";FR$,F$,FT$
150 CLS
160 PRINT "-----"
170 PRINT "Programma Entropia e Percentuali":PRINT
180 IF B THEN PRINT "elaborazione sul file ";F$;" - presenza/assenza":ELSE

PRINT "elaborazione sul file ";F$;" - dati quantitativi"
190 PRINT "dimensioni tabella : ";NR;" ";NC
200 PRINT "-----":PRINT
210 ST=0
220 IF FC$(1)="" THEN OPEN "o",#2,FC$
230 IF FR$(1)="" THEN OPEN "o",#3,FR$
240 IF FT$(1)="" THEN OPEN "o",#4,FT$
250 FOR I=1 TO NC:SC(I)=0!:TC(I)=0!:EC(I)=0!:NEXT I: FOR I=1 TO NR:FOR J=1 TO NC: SC(J)=SC(J)+TB(I,J):IF TB(I,J) THEN TC(J)=TC(J)+1
260 NEXT J,I
270 FOR I= 1 TO NR: FOR J=1 TO NC
280 IF TB(I,J) THEN 290 ELSE 300
290 EC(J)=EC(J)+TB(I,J)/SC(J)*LOG(TB(I,J)/SC(J))
300 NEXT J,I
310 FOR I=1 TO NC:EC(I)=-EC(I):TC(I)=LOG(TC(I)):NEXT I
320 PRINT " ", "tot.col.", "h", "Hmax", "h/Hmax":PRINT
330 FOR I=1 TO NC:IF TC(I) THEN HDM=EC(I)/TC(I) ELSE HDM=0
340 PRINT I,SC(I),EC(I),TC(I),HDM:NEXT I
350 IF FC$="" GOTO 430
360 ZZ4=INPUT$(1)
370 CLS
380 PRINT:PRINT "tabella percentuali su tot. di colonna (su file ";F$;)"
390 PRINT "-----":PRINT
400 FOR I=1 TO NR: FOR J=1 TO NC: TBPC=TB(I,J)/SC(J)*100
410 PRINT#2,USING"#####.##"; TBPC;:PRINT USING"#####.##";TBPC;
420 NEXT J: PRINT #2,: PRINT: NEXT I
430 FOR I=1 TO NR: SR(I)=0!:TR(I)=0!:ER(I)=0: NEXT I
440 FOR I=1 TO NC:FOR J=1 TO NR: SR(J)=SR(J)+TB(J,I):IF TB(J,I) THEN TR(J)=TR(J)+1
450 ST=ST+TB(J,I):NEXT J,I
460 FOR I=1 TO NC: FOR J=1 TO NR:
470 IF TB(J,I) THEN 480 ELSE 490
480 ER(J)=ER(J)+TB(J,I)/SR(J)*LOG(TB(J,I)/SR(J))
490 NEXT J,I
500 ZZ4=INPUT$(1)

```

```

500 ZZ4=INPUT$(1)
510 CLS
520 FOR I=1 TO NR:ER(I)=-ER(I):TR(I)=LOG(TR(I)):NEXT I
530 PRINT:PRINT " ", "tot.rig.", "h", "Hmax", "h/Hmax":PRINT
540 FOR I=1 TO NR: IF TR(I) THEN HDM=ER(I)/TR(I) ELSE HDM=0
550 PRINT I,SR(I),ER(I),TR(I),HDM:NEXT I
560 ZZ4=INPUT$(1)
570 CLS
580 IF FR$="" GOTO 650
590 PRINT:PRINT"tabella percentuali su tot. di riga (su file ";FR$;)"
600 PRINT"-----":PRINT
610 FOR I=1 TO NR: FOR J=1 TO NC:TBPC=TB(I,J)/SR(I)*100:PRINT #3,USING"#####",TBPC;: PRINT USING"#####"; TBPC;: NEXT J
620 PRINT #3,:PRINT : NEXT I
630 ZZ4=INPUT$(1)
640 CLS
650 PRINT:PRINT"tot.generale = ";ST
660 HTB=0
670 FOR I=1 TO NR: FOR J=1 TO NC
680 IF TB(I,J) THEN 690 ELSE 700
690 HTB=HTB+TB(I,J)/ST*LOG(TB(I,J)/ST)
700 NEXT J,I
710 HRIG=0:HCOL=0
720 FOR I=1 TO NR: HRIG=HRIG+SR(I)/ST*LOG(SR(I)/ST): NEXT I
730 FOR J=1 TO NC: HCOL=HCOL+SC(J)/ST*LOG(SC(J)/ST): NEXT J
740 PRINT"entropia righe (Hx) = ";-HRIG
750 PRINT"entropia colonne (Hy) = ";-HCOL
760 PRINT"entropia tabella (Hxy) = ";-HTB
770 MUTUAINF=-HRIG-HCOL+HTB
780 PRINT"mutua entropia (H(x,y)=Hx+Hy-Hxy) = ";MUTUAINF:PRINT
790 PRINT"2I = ";2*ST*MUTUAINF
800 PRINT"gradi di liberta' = ";(NR-1)*(NC-1)
810 IF FT$="" GOTO 890
820 ZZ4=INPUT$(1)
830 CLS
840 PRINT"tabella percentuali sul totale generale (su file ";FT$;)"
850 PRINT"-----":PRINT
860 FOR I=1 TO NR: FOR J=1 TO NC: TBPC=TB(I,J)/ST*100
870 PRINT#4,USING"#####",TBPC;:PRINT USING"#####",TBPC;:NEXT J
880 PRINT#4,:PRINT: NEXT I
890 IF FC{<}"." THEN CLOSE#2
900 IF FR{<}"." THEN CLOSE#3
910 IF FT{<}"." THEN CLOSE#4
920 END

```

2. DIV

Questo programma elabora una matrice $m \times n$ producendo una tabella dei seguenti parametri di diversita':

- a) numero di variabili per colonna e rapporto tra questo numero e il numero di righe della tabella (frequenza assoluta e relativa degli elementi di colonna)
- b) entropia massima per ogni colonna
- c) rapporto tra l'entropia massima della colonna e l'entropia massima della tabella
- d) indice di Gini-Simpson:

$$D = \sum_{i=1, \dots, m} p(i)^2 \quad (7)$$

dove $p(i)$ e' il rapporto tra ogni singolo valore e il corrispondente totale di colonna; l'indice tende a zero con l'aumentare della diversita', e' uguale ad 1 quando c' e' solo una entita' (diversita' = 0).

- e) entropia secondo la formula di Shannon (1.2)

f) entropia relativa (o indice di uniformita') data dal rapporto tra l'entropia (1.2) e l'entropia massima di colonna (2)

- g) indice di ricchezza di Margalef dato da:

$$M = (c - 1) / \ln \sum y(i) \quad (8)$$

dove $y(i)$ e' un valore di colonna.

Questo programma puo' servire per determinare i profili di diversita' gerarchica come suggerito da Feoli & Scimone (1984). Per ottenerli basta effettuare una differenza tra le matrici corrispondenti ai livelli gerarchici inferiori e quelle dei livelli gerarchici superiori. Per far cio' si puo' usare il programma OPSUMAT (Lagonegro e Feoli, 1985).

Esempio di applicazione

Il programma DIV richiede in input:

- 1) nome del file di lettura della tabella dati e il nome del file di scrittura della matrice dei valori di diversita' calcolati
- 2) numero di righe e di colonne della tabella in input

RUN
Programma DIV

Calcola i parametri di diversita' di una tabella specie x rilievi

NOME FILE DATI
? %bio5
NOME FILE SCRITTURA
? prova2
ORDINE DELLA MATRICE DEI DATI
? 15,5

FILE: %bio5

N. TOTALE SPECIE = 15
N. RILIEVI = 5
HMAX TOTALE = 2.70805

RILIEVO	S/R	R/T	HMAX	HMAXT	SP	H'	EVEN	RICH
1	11	0.733	2.398	0.885	0.184	1.994	0.833	2.172
2	14	0.933	2.639	0.975	0.142	2.174	0.824	2.822
3	15	1.000	2.708	1.000	0.144	2.182	0.806	3.040
4	14	0.933	2.639	0.975	0.162	2.074	0.786	2.823
5	15	1.000	2.708	1.000	0.145	2.174	0.803	3.041

S/R= N.SPECIE/RILIEVO
R/T= N.SPECIE PER RILIEVO/N.TOTALE SPECIE
HMAX= DIVERSITA' MASSIMA PER RILIEVO (LOG N. SPECIE)
HMAXT= HMAX PER RILIEVO/HMAX TOTALE
SP=INDICE DI DOMINANZA SIMPSON
H'=INDICE DI DIVERSITA' DI SHANNON
EVEN=INDICE DI UNIFORMITA' (ENTROPIA RELATIVA)
RICH=INDICE DI RICCHEZZA DI MARGALEF

```

10 REM DIV
20 LPRINT CHR$(15):WIDTH LPRINT 134
30 CLS
40 PRINT "Programma DIV":PRINT
50 PRINT "Calcola i parametri di diversita' di una tabella specie x rilievi"
60 PRINT
70 C$="      "
80 F$="      "
90 PRINT "NOME FILE DATI "
100 INPUT V$
110 PRINT "NOME FILE SCRITTURA ":INPUT VV$
120 OPEN "o",2,VV$
130 OPEN "i",1,V$
140 R$=""
150 PRINT "ORDINE DELLA MATRICE DEI DATI"
160 INPUT S,R
170 DIM X(S,R)
180 FOR I=1 TO S
190 FOR J=1 TO R
200 INPUT#1,X(I,J)
210 NEXT J
220 NEXT I
230 DIM C(R ),D(R ),E(R ),L(R ),C2(R ),C1(R )
240 DIM H1(R ),H2(R ),H3(R ),PIE(R ),MOR(R)
250 DIM T1(S ),T2(S )
260 DIM S1(S ),S2(S ),S3(S )
270 DIM A9(R )
280 DIM P1(S ),P2(S ),P3(S )
290 DIM N(S ),M2(S )
300 B=LOG(S)
310 FOR J=1 TO R
320 C(J)=0
330 C1(J)=0
340 FOR I=1 TO S
350 IF X(I,J)=0 GOTO 380
360 C(J)=C(J)+1
370 C1(J)=C1(J)+X(I,J)
380 NEXT I
390 D(J)=C(J)/S
400 E(J)=LOG(C(J))
410 L(J)=E(J)/B
420 C2(J)=0
430 FOR I= 1 TO S
440 K1= (X(I,J)/C1(J))^2
450 C2(J)=C2(J)+K1
460 NEXT I
470 NEXT J
480 GOSUB 840
490 LPRINT ""
500 LPRINT

```

```

510 LPRINT " FILE: ",Y%,LPRINT
520 LPRINT " N. TOTALE SPECIE = ",S
530 LPRINT " N. RILIEVI = ",R
540 LPRINT " HMAX TOTALE = ",B
550 LPRINT
560 LPRINT
570 LPRINT " RILIEVO  S/R  R/T  HMAX "
580 LPRINT " HMAXT  SP  H' "
590 LPRINT " EVEN  RICH "
600 A$=""
610 FOR I=1 TO 120
620 LPRINT A$;
630 NEXT I
640 LPRINT
650 LPRINT
660 FOR J=1 TO R
670 PIE(J)=H1(J)/LOG(C(J)):MAR(J)=(C(J)-1)/LOG(C1(J))
680 LPRINT USING G$;J;C(J):LPRINT USING F$;D(J);E(J);
690 LPRINT USING F$;L(J);C2(J);H1(J);PIE(J);MAR(J)
700 PRINT W2,USING F$;J;C(J);D(J);E(J);L(J);C2(J);H1(J);PIE(J);MAR(J)
710 NEXT J
720 FOR I=1 TO 100
730 LPRINT A$;
740 NEXT I
750 LPRINT
760 LPRINT"S/R= N.SPECIE/RILIEVO"
770 LPRINT"R/T= N.SPECIE PER RILIEVO/N.TOTALE SPECIE"
780 LPRINT"HMAX= DIVERSITA' MASSIMA PER RILIEVO (LOG N. SPECIE)"
790 LPRINT"HMAXT= HMAX PER RILIEVO/HMAX TOTALE"
800 LPRINT"SP=INDICE DI DOMINANZA SIMPSON"
810 LPRINT"H'=INDICE DI DIVERSITA' DI SHANNON"
820 LPRINT"EVEN=INDICE DI UNIFORMITA' (ENTROPIA RELATIVA)": LPRINT "RICH=INDICE DI RICCHEZZA DI MARGALEF"
830 END
840 FOR J=1 TO R:H1(J)=0
850 FOR I=1 TO S
860 X=X(I,J)/C1(J)
870 IF X=0 THEN GOTO 900
880 X1=X*LOG(X)
890 H1(J)=H1(J)-X1
900 NEXT I
910 A9(J)=H1(J)/E(J)
920 NEXT J
930 RETURN

```

3. ALFADIV

Il programma calcola 10 misure di diversità dei singoli vettori colonna di una tabella secondo l'entropia di Renyi (9) facendo variare il parametro alfa.

$$H(\alpha) = (-\ln_5 p(i)^{\alpha}) / (1-\alpha) \quad i=1, \dots, m \quad (9)$$

dove alfa varia da 0 a 2 ed e' incrementata del valore 0,2, p(i) e' il rapporto tra un valore e il suo totale di colonna. Questa rappresenta la formula generalizzata dell'entropia di Shannon (1.2) e diventa uguale a quest'ultima per alfa tendente a 1. Per alfa = 0 la funzione diventa uguale alla (2), mentre per alfa = 2 diventa uguale al logaritmo negativo dell'indice di Gini-Simpson.

Il programma calcola anche l'entropia massima di colonna secondo la formula (2).

Esempio di applicazione

Il programma ALFADIV richiede in input:

- 1) nome del file della tabella
- 2) numero di righe e di colonne della tabella
- 3) indice di colonna per la quale si richiedono le 10 misure di diversità

Il programma calcola per ogni colonna richiesta i 10 indici di diversità corrispondenti ai valori di alfa e da' la possibilità di stampare il profilo di diversità mettendo in ascissa il parametro alfa ed in ordinata il valore di diversità corrispondente. I profili di diversità di più colonne sono tra loro confrontabili (sovrapponibili) essendo la scala dell'asse y rapportata sempre all'entropia massima.

RUN

ALFA-DIV

Calcola la diversità α dei vettori colonna a dieci livelli α per α da 0 a 2

nome del file di dati: Zbio5

n. righe e n. colonne: 15,5

dare l'indice del vettore desiderato (0 per uscire): 1

HMAX (Y) = 2.70805

vettore: 1

H (0.0) = 2.397895

H (0.2) = 2.310629

H (0.4) = 2.226038

H (0.6) = 2.144791

H (0.8) = 2.067323

H (1.0) = 1.993906

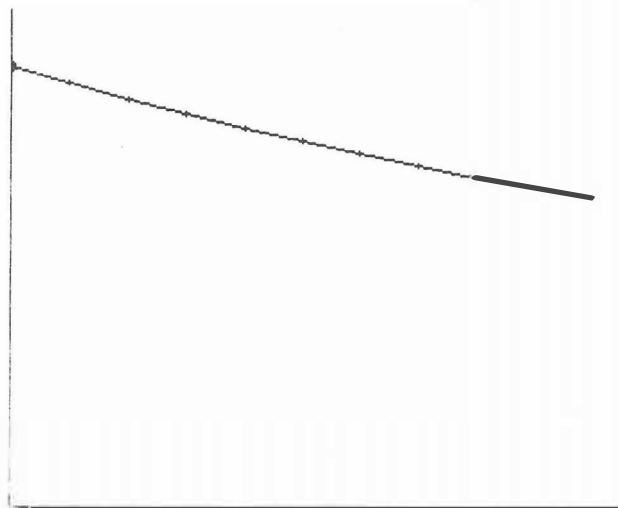
H (1.2) = 1.924708

H (1.4) = 1.859836

H (1.6) = 1.799351

H (1.8) = 1.74327

H (2.0) = 1.691553



H (0.0) = 2.397895
H (0.2) = 2.310629
H (0.4) = 2.226038
H (0.6) = 2.144791
H (0.8) = 2.067323
H (1.0) = 1.993906
H (1.2) = 1.924708
H (1.4) = 1.859836
H (1.6) = 1.799351
H (1.8) = 1.74327
H (2.0) = 1.691553

HMAX (Ψ) = 2.70805

vettore: 1

```

10 REM ALFADIV
20 SCREEN 2
30 CLS:PRINT "ALFA-DIV":PRINT "Calcola la diversita'",CHR$(224);" dei vettori colonna a dieci livelli ",CHR$(224)
40 PRINT "per ";CHR$(224);" da 0 a 2":PRINT
50 INPUT "nome del file di dati: ", A$
60 INPUT "n. righe e n. colonne: ",R,C
70 OPT=1
80 DIM X(R,C),P(R,C),T(C),H(50),N(C)
90 OPEN "i",1,A$
100 FOR I=1 TO R:FOR J=1 TO C:INPUT #1,X(I,J):NEXT J:NEXT I
110 REM --- calcola i totali
120 TOT=0
130 FOR J=1 TO C:T(J)=0
140 FOR I=1 TO R
150 T(J)=T(J)+X(I,J)
160 NEXT I
170 TOT=TOT+T(J)
180 NEXT J
190 REM --- calcola le probabilita'
200 FOR I=1 TO R:FOR J=1 TO C
210 IF OPT=1 THEN GOTO 220 ELSE 240
220 P(I,J)=X(I,J)/T(J)
230 GOTO 250
240 P(I,J)=X(I,J)/TOT
250 NEXT J:NEXT I
260 REM --- calcola H'
270 IF OPT <> 1 THEN GOTO 450
280 CLS:INPUT "dare l'indice del vettore desiderato (0 per uscire): ",K
290 IF K=0 THEN GOTO 590
300 FOR ALFA=0 TO 20 STEP 2
310 H(ALFA)=0
320 FOR I=1 TO R
330 DELTA=1-(ALFA/10)
340 IF P(I,K)=0 THEN GOTO 380
350 IF DELTA=0 THEN H(ALFA)=H(ALFA)-P(I,K)*LOG(P(I,K))
360 IF DELTA=0 THEN GOTO 380
370 H(ALFA)=H(ALFA)+P(I,K)^(ALFA/10)
380 NEXT I
390 IF DELTA =0 THEN GOTO 410
400 H(ALFA)=LOG(H(ALFA))/DELTA
410 NEXT ALFA
420 HMAX=LOG(R)
430 GOSUB 600
440 GOTO 280
450 FOR ALFA=0 TO 2 STEP .2 : H(ALFA)=0
460 PRINT ALFA;
470 FOR I=1 TO R:FOR J=1 TO C
480 DELTA=1-ALFA
490 IF DELTA=0 THEN H(ALFA)=H(ALFA)-P(I,J)*LOG(P(I,J))
500 IF DELTA=0 THEN GOTO 520

```

```

510 H(ALFA)=H(ALFA)+P(I,J)^ALFA
520 NEXT J
530 NEXT I
540 IF DELTA=0 THEN GOTO 560
550 H(ALFA)=LOG(H(ALFA))/DELTA
560 NEXT ALFA
570 HMAX=LOG(NTOT)
580 GOSUB 600
590 END
600 REM --- traccia la curva
610 ZZ=0
620 STP=.2
630 CLS :SCREEN 3
640 VIEW (10,10)-(350,250)
650 WINDOW (0,HMAX)-(21,0)
660 LINE (0,0)-(0,HMAX):LINE(0,0)-(21,0)
670 CIRCLE (0,H(0)),.1
680 FOR ALFA=2 TO 20 STEP 2
690 LINE (ALFA-2 ,H(ALFA-2 ))-(ALFA,H(ALFA))
700 CIRCLE (ALFA,H(ALFA)),.05
710 NEXT ALFA
720 LOCATE 20,1:PRINT "HMAX (V) = ";HMAX
730 LOCATE 20,55:PRINT "vettore: ";k
740 FOR ALFA=0 TO 20 STEP 2
750 LOCATE ZZ+3,55:PRINT "H (";:PRINT USING "##.##";ALFA/10;:PRINT ") = ";H(ALFA)
760 ZZ=ZZ+1
770 NEXT ALFA
780 LOCATE 22,1:PRINT "(per stampare premere spazio+RETURN)"
790 LOCATE 23,1:LINE INPUT X$
800 IF X$(1) = " " THEN GOTO 830
810 LOCATE 22,1:PRINT "
820 LCOPY
830 VIEW
840 SCREEN 2
850 RETURN

```

4. BETADIV

Il programma calcola 10 misure di diversita' dei singoli vettori colonna di una tabella secondo la funzione di Patil e Taillie (10) facendo variare il parametro beta:

$$H(\beta) = (1 - \sum p(i)^{\beta+1}) / \beta \quad i=1, \dots, m \quad (10)$$

dove beta varia da -1 a +1 ed e' incrementato del valore 0.2, p(i) e' il rapporto tra un valore ed il suo totale colonna. Questa rappresenta la formula generalizzata di (1- indice di Gini-Simpson) e diventa uguale a quest'ultima per beta=+1.

Esempio di applicazione

Il programma BETADIV richiede in input:

- 1) nome del file della tabella
- 2) numero di righe e di colonne della tabella
- 3) indice della colonna per la quale si richiedono le 10 misure di diversita'

Il programma stampa i 10 indici di diversita' per ogni colonna richiesta e da' la possibilita' di stampare il profilo di diversita' ponendo nelle ascisse il parametro beta e nelle ordinate il valore di diversita' corrispondente. I profili di piu' colonne sono tra loro confrontabili (sovrapponibili) essendo la scala dell'asse y sempre rapportata al numero di righe della tabella.

RUN

BETA-DIV

Calcola la diversita' β dei vettori colonna a dieci livelli β per β da -1 a +1

nome del file di dati: %bio5

n. righe e n. colonne: 15,5

dare l'indice del vettore desiderato (0 per uscire): 1

Y MAX = 15

vettore: 1

H (-1.0) = 10

H (-0.8) = 6.687883

H (-0.6) = 4.670606

H (-0.4) = 3.395604

H (-0.2) = 2.560237

H (0.0) = 1.994371

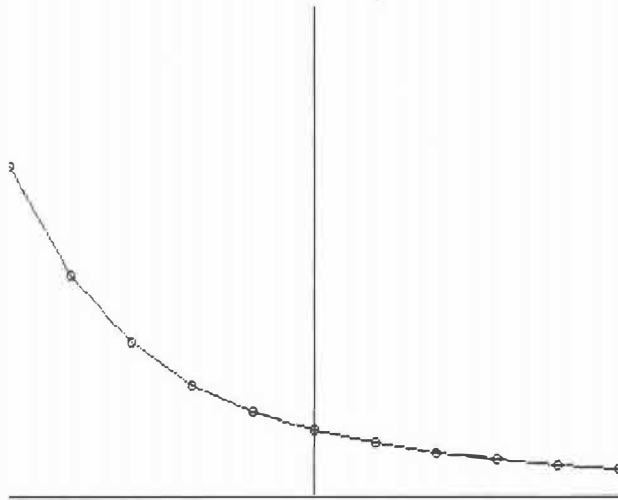
H (0.2) = 1.597548

H (0.4) = 1.311899

H (0.6) = 1.100454

H (0.8) = .9400896

H (1.0) = .8157668



H (-1.0)	=	10
H (-0.8)	=	6.687883
H (-0.6)	=	4.670606
H (-0.4)	=	3.395604
H (-0.2)	=	2.560237
H (0.0)	=	1.994371
H (0.2)	=	1.597548
H (0.4)	=	1.311899
H (0.6)	=	1.100454
H (0.8)	=	.9400896
H (1.0)	=	.8157668

Y MAX = 15

vettore: 1

```

10 REM BETADIV
20 SCREEN 2
30 CLS:PRINT "BETA-DIV":PRINT "Calcola la diversita' ";CHR$(225);" dei vettori colonna a dieci livelli ";CHR$(225)
40 PRINT "per ";CHR$(225);" da -1 a +1":PRINT
50 INPUT "nome del file di dati: ", A$
60 INPUT "n. righe e n. colonne: ",R,C
70 OPT=1
80 DIM X(R,C),P(R,C),T(C),H(250),N(C)
90 OPEN "i",1,A$
100 FOR I=1 TO R:FOR J=1 TO C:INPUT #1,X(I,J):NEXT J:NEXT I
110 REM --- calcola i totali
120 TOT=0
130 FOR J=1 TO C:T(J)=0
140 FOR I=1 TO R
150 T(J)=T(J)+X(I,J)
160 NEXT I
170 TOT=TOT+T(J)
180 NEXT J
190 REM --- calcola le probabilita'
200 FOR I=1 TO R:FOR J=1 TO C
210 IF OPT=1 THEN GOTO 220 ELSE 240
220 P(I,J)=X(I,J)/T(J)
230 GOTO 250
240 P(I,J)=X(I,J)/TOT
250 NEXT J:NEXT I
260 REM --- calcola H'
270 CLS
280 BETA1=-1:BETA2=1
290 PASSO=(BETA2-BETA1)/20
300 IF OPT <> 1 THEN GOTO 450
310 CLS:INPUT "dare l'indice del vettore desiderato (0 per uscire): ", K
320 IF K=0 THEN GOTO 590
330 FOR ALFA= 0 TO 20:BETA=ALFA/10-1
340 IF BETA=0 THEN BETA=.0001
350 H(ALFA)=0
360 FOR I=1 TO R
370 IF P(I,K)=0 THEN GOTO 390
380 H(ALFA)=H(ALFA)+P(I,K)^(BETA+1)
390 NEXT I
400 H(ALFA)=(1-H(ALFA))/BETA
410 NEXT ALFA
420 HMAX=LOG(R)
430 GOSUB 600
440 GOTO 310
450 FOR ALFA=BETA1 TO BETA2 STEP PASSO:H(ALFA)=0
460 PRINT ALFA;
470 FOR I=1 TO R:FOR J=1 TO C
480 DELTA= ALFA
490 IF ALFA=0 THEN H(ALFA)=H(ALFA)-P(I,J)*LOG(P(I,J))
500 IF ALFA=0 THEN GOTO 520

```

```

510 H(ALFA)=H(ALFA)+P(I,J)^ALFA
520 NEXT J
530 NEXT I
540 IF ALFA=0 THEN GOTO 560
550 H(ALFA)=LOG(H(ALFA))/DELTA
560 NEXT ALFA
570 HMAX=LOG(NTOT)
580 GOSUB 600
590 END
600 REM --- traccia la curva
610 CLS :SCREEN 3
620 VIEW (10,10)-(350,250)
630 WINDOW (BETA1,R)-(BETA2,0)
640 LINE (0,0)-(0,R):LINE(BETA1,0)-(BETA2,0)
650 CIRCLE (BETA1,H(0)),.01
660 FOR ALFA=2 TO 20 STEP 2
670 P= ALFA/10-1
680 LINE (P-.2,H (ALFA-2))-(P,H(ALFA))
690 CIRCLE (P,H(ALFA)),.01
700 NEXT ALFA
710 LOCATE 20,1:PRINT "Y MAX = ";R
720 LOCATE 20,55:PRINT "vettore: ";K
730 ZX=4
740 FOR ALFA=0 TO 20 STEP 2
750 P=ALFA/10-1
760 LOCATE ZX,55:PRINT "H (";:PRINT USING "##.##";P;:PRINT ") = ";H(ALFA)
770 ZX=ZX+1
780 NEXT ALFA
790 LOCATE 22,1:PRINT "(per stampare premere spazio+RETURN)"
800 LOCATE 23,1:LINE INPUT X$
810 IF X$(0)=" " THEN GOTO 840
820 LOCATE 22,1:PRINT "
830 LCOPY
840 VIEW
850 SCREEN 2
860 RETURN

```

5. PRODIV

Il programma traccia i grafici relativi ai profili di diversita' di Whittaker (1972) dei vettori colonna di una tabella $m \times n$. Questi vengono costruiti riordinando la sequenza delle righe secondo valori decrescenti del vettore colonna considerato oppure secondo i logaritmi dei valori percentuali di colonna. Opzionalmente e' possibile il riordinamento anche sulla base del contributo all' entropia di Shannon (1) e all'indice di Gini-Simpson (7) di ciascun elemento del vettore colonna. Viene calcolato e disegnato a richiesta anche il profilo di diversita' relativo alla frequenza assoluta dei valori nelle m righe: e' bene sottolineare che in questo caso viene calcolato un unico profilo per tutta la matrice.

Esempio di applicazione

Il programma PRODIV richiede in input:

- 1) nome del file della tabella
- 2) numero di righe e di colonne della tabella
- 3) numero che seleziona una delle seguenti opzioni:
 - 1 - ordinamento decrescente della frequenza assoluta sulle righe
 - 2 - ordinamento decrescente dei valori dei singoli vettori colonna
 - 3 - ordinamento decrescente dei valori di entropia di Shannon calcolata sui valori dei singoli vettori colonna
 - 4 - ordinamento decrescente degli indici di Gini-Simpson calcolati sui valori dei singoli vettori colonna
 - 5 - ordinamento decrescente dei logaritmi dei valori percentuali di colonna

Il programma stampa la sequenza delle m righe ordinate in senso decrescente rispetto ai valori della funzione richiesta e disegna il grafico corrispondente.

RUN
PRODIV - Traccia profili di diversita'

DARE NOME FILE: %bio5
DIMENSIONI DEL FILE: 15,5
OPZIONI:

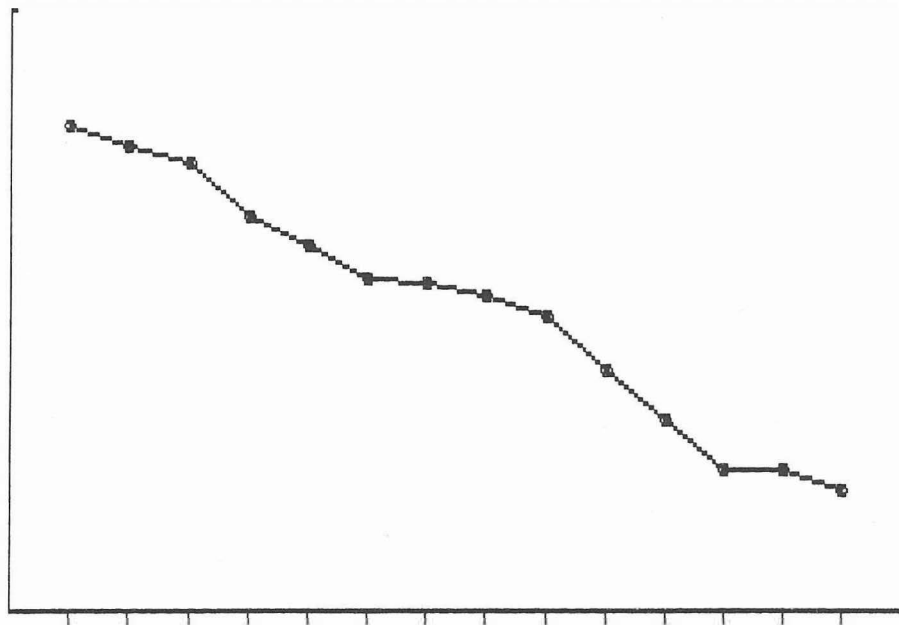
1 .FREQUENZA ASSOLUTA SULLA MATRICE
2 .COPERTURA PER RILIEVO
3 .ENTROPIA DI SHANNON PER RILIEVO
4 .INDICE DI SIMPSON PER RILIEVO
5 .LOG DELLA COPERTURA PERCENTUALE
6 .QUIT
? 5
DARE INDICE DEL VETTORE CONSIDERATO: (0 PER USCIRE) 4
STAMPA DELLA SEQUENZA (SI/NO) ? SI
SPECIE LOG DELLA COPERTURA PERCENTUALE

13	5.564521
10	5.347108
1	5.147495
2	4.543295
11	4.204693
15	3.828642
14	3.78419
9	3.610918
3	3.401197
12	2.772589
4	2.197225
8	1.609438
6	1.609438
5	1.386294
7	0

6.907755
N.specie: 14

6.907755

LOG DELLA COPERTURA PERCENTUALE - vettore: 4



N.specie: 14

```

10 REM PRODIV
20 SCREEN 2
30 CLS:PRINT "PRODIV - Traccia profili di diversita':":PRINT
40 DIM Z$(10)
50 Z$(1)="FREQUENZA ASSOLUTA SULLA MATRICE"
60 Z$(2)="COPERTURA PER RILIEVO"
70 Z$(3)="ENTROPIA DI SHANNON PER RILIEVO"
80 Z$(4)="INDICE DI SIMPSON PER RILIEVO"
90 Z$(6)="QUIT"
100 Z$(5)="LOG DELLA COPERTURA PERCENTUALE"
110 INPUT "DARE NOME FILE: ",A$
120 OPEN "I",1,A$
130 INPUT "DIMENSIONI DEL FILE: ",S,R
140 DIM X(S,R),A(S),B(S),F(S),T(R)
150 FOR I=1 TO S:FOR J=1 TO R:INPUT #1,X(I,J):NEXT J:NEXT I
160 CLS
170 PRINT "OPZIONI:":PRINT
180 FOR I=1 TO 6:PRINT I;" ";Z$(I):NEXT I
190 INPUT 0
200 IF 0=1 THEN GOTO 250
210 IF 0=6 THEN END
220 INPUT "DARE INDICE DEL VETTORE CONSIDERATO: (0 PER USCIRE) ",P
230 Q=P-U:0
240 IF P=0 THEN GOTO 160
250 ON 0 GOTO 260,370,540,540,540
260 REM ----- OPZIONE 1 -----
270 FOR I=1 TO S:F(I)=0
280 FOR J=1 TO R
290 IF X(I,J)=0 THEN GOTO 310
300 F(I)=F(I)+1
310 NEXT J
320 NEXT I
330 U=R
340 GOSUB 840
350 GOSUB 1000
360 CLS:GOTO 160
370 REM ----- OPZIONE 2 -----
380 FOR I=1 TO S:F(I)=0
390 FOR J=P TO Q
400 F(I)=F(I)+X(I,J):NEXT J
410 NEXT I
420 U=100
430 GOSUB 840
440 GOSUB 1000
450 CLS:GOTO 220
460 REM ----- OPZIONE 3 -----
470 FOR I=1 TO S
480 IF F(I)=0 THEN GOTO 500
490 F(I)=LOG(F(I))
500 NEXT I

```

```

510 GOSUB 840
520 GOSUB 1000
530 GOTO 160
540 REM ----- OPZIONI 4,5 -----
550 FOR J=P TO Q: T(J)=0
560 FOR I=1 TO 5
570 T(J)=T(J)+X(I,J)
580 F(I)=0
590 NEXT I:NEXT J
600 IF 0=4 THEN GOTO 740
610 FOR J=P TO Q
620 FOR I=1 TO 5
630 IF X(I,J)=0 THEN GOTO 660
640 F(I)=-((X(I,J)/T(J))*LOG(X(I,J)/T(J)))
650 IF 0=5 THEN F(I)=LOG(1000*X(I,J)/T(J))
660 NEXT I
670 NEXT J
680 U=1
690 IF 0=5 THEN U=LOG(1000)
700 GOSUB 840
710 GOSUB 1000
720 IF 0=3 THEN GOTO 220
730 IF 0=5 THEN GOTO 220
740 FOR J=P TO Q
750 FOR I=1 TO 5
760 F(I)=(X(I,J)/T(J))^2
770 IF F(I)=0 THEN GOTO 780
780 NEXT I
790 NEXT J
800 U=.3
810 GOSUB 840
820 GOSUB 1000
830 CLS:GOTO 220
840 REM ----- RANKING -----
850 FOR J=P TO Q
860 FOR I=1 TO 5:A(I)=I:NEXT I
870 FOR K=1 TO 5-1:FOR H=K+1 TO 5
880 IF F(K)>F(H) THEN GOTO 910
890 V=F(K):F(K)=F(H):F(H)=V
900 Z=A(K):A(K)=A(H):A(H)=Z
910 NEXT H:NEXT K:NEXT J
920 CLS
930 PRINT "STAMPA DELLA SEQUENZA (SI/NO) ";:INPUT ST$
940 CLS
950 IF ST$="SI" THEN LPRINT "SPECIE",Z$(0):LPRINT
960 PRINT "SPECIE",Z$(0):PRINT
970 IF ST$="SI" THEN FOR I=1 TO 5:LPRINT A(I),F(I):NEXT I
980 FOR I=1 TO 5:PRINT A(I),F(I):NEXT I
990 RETURN
1000 REM ----- PLOT -----

```

```

1010 CLS
1020 SCREEN 2
1030 VIEW (0,0)-(500,175)
1040 WINDOW (0,-.1*U)-(5+.005*S,U+.1*U)
1050 LINE (0,0)-(0,U):LINE (0,0)-(5*1,0)
1060 CIRCLE (1,F(1)),.007*S:PSET (1,0):LINE -(1,-.02*U)
1070 PSET (1,F(1))
1080 FOR I=2 TO 5
1090 IF F(I)=0 THEN GOTO 1140
1100 NUM=I
1110 LINE -(I,F(I)):CIRCLE (I,F(I)),.007*S
1120 PSET (I,0):LINE -(I,-.02 *U):PSET (I,F(I))
1130 NEXT I
1140 LINE(0,U)-(.007*S,U):LOCATE 1,1:PRINT U
1150 LOCATE 21,65:PRINT "N.specie:";NUM
1160 LOCATE 1,20:PRINT "PER STANPARE, PRENERE SPAZIO+(CR)";
1170 LINE INPUT X$
1180 IF X$("<")" " THEN GOTO 1230
1190 LOCATE 1,20:PRINT "
1200 IF P=0 THEN LOCATE 1,20:PRINT Z$(0)
1210 IF P(<)>0 THEN LOCATE 1,20:PRINT Z$(0);" - vettore:";P
1220 LCOPY
1230 VIEW:CLS
1240 RETURN

```

6. INFORMA

Il programma calcola tra i vettori colonna di una matrice $m \times n$ indici di somiglianza o di distanza basati sulla teoria dell'informazione e crea matrici simmetriche $n \times n$. E' possibile scegliere tra i seguenti 5 indici:

1) indice di distanza basato sulla mutua informazione (11) calcolata sulla sottomatrice $m \times 2$ costituita dai due vettori colonna j e k messi a confronto:

$$d(j,k) = I(x,y) = I_x + I_y - I_{xy} \quad (11)$$

la mutua informazione (11) e' un F_{ij} multiplo della mutua entropia (5) cosi' come l'informazione di riga I_x e' un F_{ij} multiplo dell' entropia delle righe (3.1) e l'informazione di colonna I_y e' un F_{ij} multiplo dell' entropia delle colonne (3.2).

2) indice di somiglianza che varia tra 0 (somiglianza nulla) ed 1 (massima somiglianza) calcolato sulla sottomatrice $m \times 2$ costituita dai due vettori colonna j e k messi a confronto e dato dal seguente rapporto:

$$s(j,k) = S I(x(i))/S I(x_{\max}(i)) \quad i=1, \dots, m \quad (12)$$

dove l'informazione $I(x)$ e' un F_{ij} multiplo della entropia (1.1); la sommatoria delle informazioni di riga ($I(x)$) viene rapportata alla sommatoria delle informazioni di riga ($I(x_{\max})$) di una ipotetica matrice $m \times 2$ i cui valori di riga sono entrambi uguali al valore maggiore dei due della sottomatrice originaria.

3) indice di divergenza dalla media calcolato sulla matrice $m \times 2$ costituita dai vettori colonna j e k messi a confronto; in questo caso l' indice di distanza e' l' informazione calcolata nella seguente maniera:

$$d(j,k) = I(x, x_{\text{med}}) = S S x(i,j) \ln(x(i,j)/x_{\text{med}}) \quad (13)$$
$$i=1, \dots, m; \quad j=1, \dots, n$$

dove x_{med} e' il valore medio di ogni riga. Quanto piu' questa divergenza e' alta, tanto piu' i due vettori sono diversi. Si consiglia l' uso di questo indice quando il numero di variabili per vettore e' piu' o meno costante.

4) indice di somiglianza dato dalla differenza tra l'informazione congiunta (I_{xy}) e la mutua informazione ($I(x;y)$), chiamato anche equivocazione, calcolato anche in questo caso sulla matrice $m \times 2$ dei due vettori j e k da confrontare:

$$s(j,k) = I_{xy} - I(x;y) \quad (14)$$

l'indice varia da 0 (somiglianza nulla) al valore della informazione congiunta (somiglianza massima).

5) indice di somiglianza di Raiski dato dal rapporto tra l'equivocazione (13) e l'informazione congiunta (Ixy):

$$r(j,k) = (I_{xy} - I(x;y))/I_{xy} \quad (15)$$

che varia da 0 (somiglianza nulla) ad 1 (somiglianza massima).

6) indice di coerenza:

$$d(j,k) = (1 - r(j,j)**2)**1/2 \quad (16)$$

che varia da 0 (somiglianza massima) ad 1 (somiglianza nulla).

Esempio di applicazione

Il programma richiede in input:

1) numero che seleziona uno degli indici descritti secondo l'ordine dato con le seguenti denominazioni:

- 1) MUTUA INFORMAZIONE
- 2) INFORMAZIONE ALL'INTERNO DELLE RIGHE
- 3) DIVERGENZA DALLA MEDIA
- 4) EQUIVOCAZIONE
- 5) INDICE DI RAISKI
- 6) COERENZA

2) nome del file della tabella di input
3) numero di righe e di colonne della tabella
4) opzione di stampa degli indici calcolati
5) nome del file della tabella degli indici: questa e' la matrice triangolare alta della matrice simmetrica degli indici, compresi gli elementi diagonali.

Il programma calcola anche alcune funzioni informazionali su tutta tabella di input quali l'informazione di riga e di colonna, l'informazione congiunta, la mutua informazione, l'equivocazione, l'indice di Raiski e la coerenza.

RUN
PROGRAMMA INFORMA

Calcola indici di somiglianza o distanza fra le colonne di una tabella
specie/rilievi, basati sulla teoria dell' informazione

OPZIONI :

1. MUTUA INFORMAZIONE
 2. INFORMAZIONE ALL'INTERNO DELLE RIGHE
 3. DIVERGENZA DALLA MEDIA
 4. EQUIVOCAZIONE
 5. INDICE DI RAJSKI
 6. COEFFICIENTE DI COERENZA
- ? 5

NOME DEL FILE DELLA TABELLA ? %bio5
N.RIGHE, N.COLONNE ? 15,5

FILE: %bio5

COMPONENTI DELL' ENTROPIA GENERALE:

INFORMAZIONE DI RIGA	1093.139
INFORMAZIONE DI COLONNA	804.5571
INFORMAZIONE CONGIUNTA	1864.162
MUTUA INFORMAZIONE	33.53455
EQUIVOCAZIONE	1830.627
INDICE DI RAJSKI9820109
COERENZA1888243

MATRICE DI SIMMETRICA FRA I VETTORI COLONNA PER:
INDICE DI SOMIGLIANZA DI RAJSKI
STAMPA DELLA MATRICE (Y/N)

? y

1.000	0.980	0.979	0.969	0.976
1.000	0.995	0.988	0.992	
1.000	0.996	0.991		
1.000	0.983			
1.000				

NOME FILE DI SCRITTURA
? prova2

```

10 REM INFORMA
20 REM ---- menu -----
30 CLS:PRINT "PROGRAMMA INFORMA ":PRINT:PRINT "Calcola indici di somiglianza o distanza fra le colonne di una tabella"
40 PRINT "specie/rilievi, basati sulla teoria dell' informazione"
50 PRINT:PRINT:PRINT "OPZIONI ":PRINT
60 PRINT "1. MUTUA INFORMAZIONE"
70 PRINT "2. INFORMAZIONE ALL' INTERNO DELLE RIGHE"
80 PRINT "3. DIVERGENZA DALLA MEDIA"
90 PRINT "4. EQUIVOCAZIONE"
100 PRINT "5. INDICE DI RAJSKI"
110 PRINT "6. COEFFICIENTE DI COERENZA"
120 INPUT OPT
130 PRINT :INPUT "NOME DEL FILE DELLA TABELLA ",M$
140 INPUT "N.RIGHE, N.COLONNE ",R,C
150 OPEN "I",1,M$
160 DIM X(R,C),A(C,C)
170 FOR I=1 TO R:FOR J=1 TO C:INPUT #1,X(I,J):NEXT J:NEXT I
180 REM ---- TOTALI DI RIGA E COLONNA
190 DIM TR(R),TC(C)
200 FOR I=1 TO R:FOR J=1 TO C
210 TR(I)=TR(I)+X(I,J):TC(J)=TC(J)+X(I,J):GT=GT+X(I,J)
220 NEXT J:NEXT I
230 GOSUB 280
240 IF OPT=2 THEN GOSUB 740
250 IF OPT=3 THEN GOSUB 860
260 GOSUB 970
270 END
280 REM ---- OPZIONI 1,4,5,6
290 DIM DR(R),DC(C)
300 FOR I=1 TO R:FOR J=1 TO C
310 A=1:IF X(I,J) <> 0 THEN A=X(I,J)/TR(I):B=A
320 DR(I)=DR(I)+B*LOG(A):NEXT J:NEXT I
330 FOR J=1 TO C:FOR I=1 TO R
340 A=1:IF X(I,J) <> 0 THEN A=X(I,J)/TC(J):B=A
350 DC(J)=DC(J)+B*LOG(A):NEXT I:NEXT J
360 BR=GT*LOG(GT):BC=BR:JT=BR
370 FOR I=1 TO R:B=TR(I)*LOG(TR(I)):BR=BR-B:NEXT I
380 FOR J=1 TO C:B=TC(J)*LOG(TC(J)):BC=BC-B:NEXT J
390 FOR I=1 TO R:FOR J=1 TO C
400 O=1:IF X(I,J) <> 0 THEN O=X(I,J)
410 JT=JT-X(I,J)*LOG(O):NEXT J:NEXT I
420 M=BR+BC-JT:E=JT-M:RJ=E/JT:CH=SQR(1-RJ^2)
430 REM ---- OPZIONE 1: MUTUA INFORMAZIONE
440 DIM MU(C,C)
450 FOR J=1 TO C:FOR K=J TO C
460 FOR I=1 TO R:AA=X(I,J)+X(I,K):TT=TC(J)+TC(K)
470 B=1:IF X(I,J) <> 0 THEN B=X(I,J)*TT/(TC(J)*AA)
480 O=1:IF X(I,K) <> 0 THEN O=X(I,K)*TT/(TC(K)*AA)
490 BB=X(I,J)*LOG(B):CC=X(I,K)*LOG(O)
500 MU(J,K)=MU(J,K)+BB+CC:A(J,K)=MU(J,K)

```

```

510 NEXT I:NEXT K:NEXT J
520 IF OPT=1 THEN RETURN
530 REM ---- OPZIONE 4: EQUIVOCAZIONE
540 DIM JO(C,C),EQ(C,C)
550 FOR J=1 TO C:FOR K=J TO C:TT=TC(J)+TC(K):FOR I=1 TO R
560 A=1:IF X(I,J)(<)D THEN A=X(I,J)/TT
570 B=1:IF X(I,K)(<)D THEN B=X(I,K)/TT
580 A=X(I,J)*LOG(A):B=X(I,K)*LOG(B)
590 JO(J,K)=JO(J,K)-A-B:NEXT I
600 EQ(J,K)=JO(J,K)-MU(J,K):A(J,K)=EQ(J,K):NEXT K:NEXT J
610 IF OPT=4 THEN RETURN
620 REM ---- OPZIONE 5: INDICE DI RAJSKI
630 DIM RA(C,C)
640 FOR J=1 TO C:FOR K=J TO C
650 RA(J,K)=EQ(J,K)/JO(J,K):A(J,K)=RA(J,K)
660 NEXT K:NEXT J
670 IF OPT=5 THEN RETURN
680 REM --- OPZIONE 6: INDICE DI COERENZA
690 DIM CO(C,C)
700 FOR J=1 TO C:FOR K=J TO C
710 IF RA(J,K)>1 THEN RA(J,K)=1
720 CO(J,K)=SQRT(1-RA(J,K)^2):A(J,K)=CO(J,K):NEXT K:NEXT J
730 RETURN
740 REM ---- OPZIONE 2: MULTIPLIO DELLA W-ENTROPIA
750 DIM MEN(C,C)
760 FOR J=1 TO C:FOR K=J TO C:MX=D:FOR I=1 TO R
770 IF X(I,J)=X(I,K) THEN MAX=X(I,J)
780 IF X(I,J)X(I,K) THEN MAX=X(I,K)
790 T=X(I,J)+X(I,K)
800 A=1:IF X(I,J)(<)D THEN A=X(I,J)/T
810 B=1:IF X(I,K)(<)D THEN B=X(I,K)/T
820 A=X(I,J)*LOG(A):B=X(I,K)*LOG(B)
830 MEN(J,K)=MEN(J,K)-A-B:MX=MX+2*MAX*LOG(2):NEXT I
840 MEN(J,K)=MEN(J,K)/MX:A(J,K)=MEN(J,K):NEXT K:NEXT J
850 RETURN
860 REM ---- OPZIONE 3: DIVERGENZA DALLA MEDIA
870 DIM DIV(C,C),MS(C,C)
880 FOR J=1 TO C:FOR K=J TO C
890 FOR I=1 TO R:T=X(I,J)+X(I,K)
900 IF X(I,J)(<)D THEN A=X(I,J)/T ELSE A=.5
910 IF X(I,K)(<)D THEN B=X(I,K)/T ELSE B=.5
920 A=X(I,J)*LOG(A/.5):B=X(I,K)*LOG(B/.5)
930 DIV(J,K)=DIV(J,K)+A+B:MS(J,K)=MS(J,K)+T*LOG(2)
940 NEXT I
950 A(J,K)=DIV(J,K):NEXT K:NEXT J
960 RETURN
970 REM ---- STAMPA E SALVA
980 DIM IT$(6)
990 IT$(1)="DIVERGENZA PER MUTUA INFORMAZIONE"
1000 IT$(2)="DIVERGENZA PER INFORMAZIONE ALL'INTERNO DELLE RIGHE"

```

```

1010 IT$(3)="DIVERGENZA DALLA MEDIA DI RIGA"
1020 IT$(4)="EQUIVOCAZIONE"
1030 IT$(5)="INDICE DI SOMIGLIANZA DI RAJSKI"
1040 IT$(6)="COEFFICIENTE DI COERENZA"
1050 LPRINT :LPRINT "FILE: ";M$:LPRINT :LPRINT "COMPONENTI DELL' ENTROPIA GENERALE:"
1060 LPRINT "INFORMAZIONE DI RIGA ..... ";BR
1070 LPRINT "INFORMAZIONE DI COLONNA ..... ";BC
1080 LPRINT "INFORMAZIONE CONGIUNTA ..... ";JT
1090 LPRINT "MUTUA INFORMAZIONE ..... ";M
1100 LPRINT "EQUIVOCAZIONE ..... ";E
1110 LPRINT "INDICE DI RAJSKI ..... ";RJ
1120 LPRINT "COERENZA ..... ";CH
1130 LPRINT :LPRINT "MATRICE DI SIMMETRICA FRA I VETTORI COLONNA PER: ";LPRINT IT$(OPT)
1140 PRINT "STAMPA DELLA MATRICE (Y/N) ":INPUT Y$
1150 F$="#####.###"
1160 IF Y$="N" THEN GOTO 1200
1170 FOR J= 1 TO C:FOR K= J TO C
1180 LPRINT USING F$;A(J,K);
1190 NEXT K:LPRINT :NEXT J
1200 PRINT "NONE FILE DI SCRITTURA ":INPUT UT$
1210 OPEN "0",2, UT$
1220 FOR J= 1 TO C:FOR K= J TO C
1230 PRINT #2,USING F$;A(J,K);
1240 NEXT K:PRINT #2,:NEXT J
1250 CLOSE
1260 RETURN

```

Bibliografia

- Feoli E. 1984 - Some aspects of classification and ordination of vegetation data in perspective. *Studia Geobotanica* 4: 7-21.
- Feoli E. Ganis P. e Poldini L. 1987 - Relazioni tra corologia e descrizioni tassonomiche e morfologiche della vegetazione dei boschi ad *Ostrya carpinifolia* Scop. del Friuli Venezia-Giulia. (in stampa)
- Feoli E. e Scimone M. 1984 - Hierarchical diversity: an application to broad-leaved woods of the Apennines. *Giornale Botanico Italiano*. Vol. 118, n1-2: 1-15.
- Feoli E., Lagonegro M. e Orloci L. 1984 - Information analysis of vegetation data. Junk, The Hague.
- Giller P.S. 1984 - Community Structure and the Niche. Chapman and Hall, London.
- Grassle J.F., Patil G.P., Smith W. e Taillie C. (eds.). 1979 - Ecological diversity in theory and practice. Statistical Ecological series, Vol.6. International Co-operative Publishing House, Fairland, Maryland, U.S.A.
- Hulbert S. H. 1971 - The non-concept of species diversity: a critique and alternative parameters. *Ecology* 52: 577-586.
- Kullback S. 1959 - Information Theory and Statistics. Wiley, New York.
- Lagonegro M. e Feoli E. 1985 - Analisi multivariata di dati. Manuale d'uso di programmi BASIC per personal computers. Libreria Goliardica . Trieste.
- Patil G.P. e Taillie C. 1976 - Ecological diversity: concepts, indices and applications. Proceedings of the 9th Int. Biometric Conference. Vol. II. The Biometric Society, 383-411.
- Pielou E.C. 1975 - Ecological Diversity. Wiley. New York.
- Renyi A. 1961 - On measure of entropy and information. In: J. Neyman (ed.), Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability, pp 547-561. Univ. of Calif. Press, Berkeley.
- Whittaker, R.H. 1972 - Evolution and measurement of species diversity. *Taxon* 21: 213-251.

Tipo/lito Astra