Analisi delle Correlazioni Canoniche (ACC)

Analisi Statistica dei Dati Multidimensionali[1]

[1]Corso di Laurea in Scienze Statistiche e Attuariali

Facoltà di Scienze Economiche e Aziendali
Università degli Studi del Sannio

Fonte: *Stawinoga. Lucidi del Corso. Università di Napoli "Federico II".*
Gherghi-Lauro. Appunti di Analisi dei Dati Multidimensionali, RCEedizioni.
Manly. Multivariate Statistical Methods: A Primer. Chapman & Hall, CRC.

Fonte: Stawinoga

- L' Analisi delle Correlazioni Canoniche è stata proposta da H. Hottelling nel 1936.
- Nel 1968 J. D.Carroll ha proposto una generalizzazione dell' ACC per tre o più gruppi di variabili.
- L'ACC è molto importante dal punto di vista teorico e metodologico. Essa può essere considerata come il caso generale delle:
  - Regressione Multipla
  - Analisi delle Corrispondenze
  - Analisi Discriminante

Fonte: Stawinoga

- L' obiettivo dell' ACC è identificare le relazioni lineari esistenti tra due gruppi di variabili quantitative osservate su uno stesso insieme di individui.

- Lo scopo è trovare una combinazione lineare delle variabili del primo gruppo e una combinazione lineare delle variabili del secondo gruppo che siano le più correlate possibile.

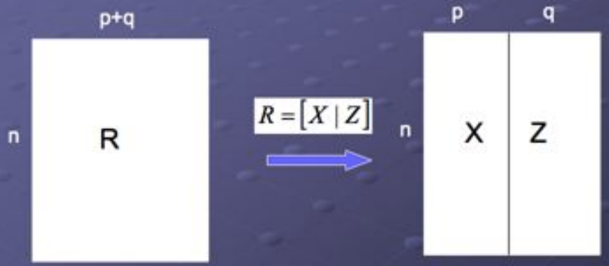
- L' ACC opera su una matrice R ad n righe e $p+q$ colonne partizionabile in due sottomatrici $X\ (n,p)$ e $Z(n,q)$.

  Le colonne della matrice X sono costituite dalle variabili del primo insieme.

  Le variabili del secondo insieme costituiscono le colonne della matrice Z.

Fonte: Stawinoga

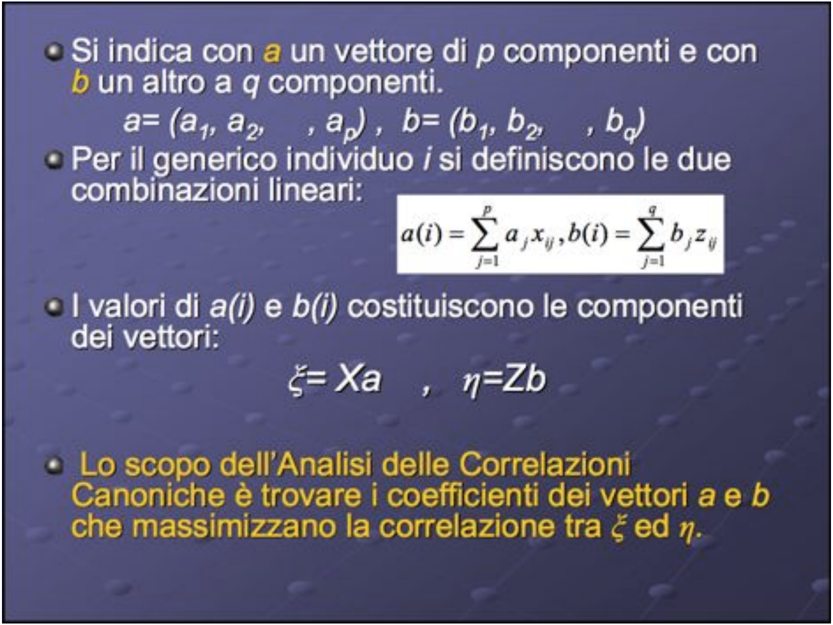
- Si supponga che tutte le variabili siano centrate quindi per ogni colonna della matrice R la somma degli elementi è uguale a 0 .
- Dal vettore $(x_{i1}, x_{i2}, \quad , x_{ip}, z_{i1}, \quad , z_{iq})$ si identifica il generico individuo *i* della matrice *R*.



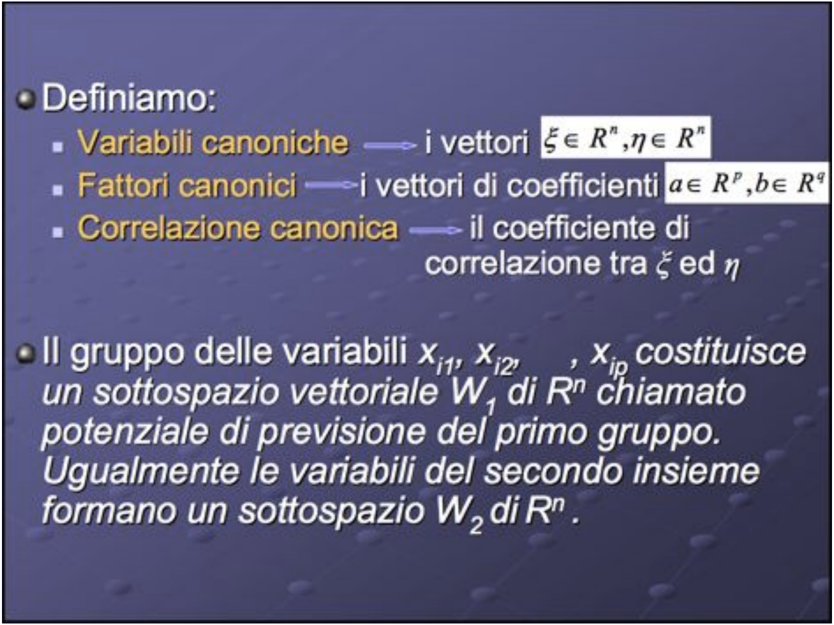
- La matrice di varianza-covarianza della matrice R può essere ottenuta da:

$$V(R) = \frac{1}{n} R'R = \frac{1}{n} \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z \end{bmatrix}$$

Fonte: Stawinoga

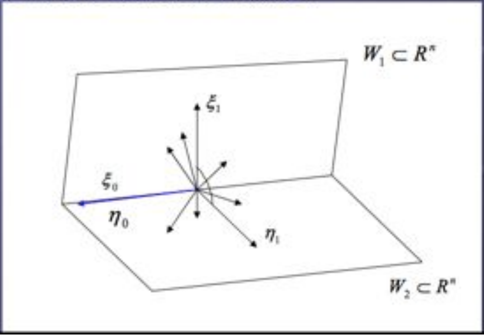


Fonte: Stawinoga



Fonte: Stawinoga

- Indichiamo con $(\xi_1,\xi_2, \quad ,\xi_k)$, $(\eta_1,\eta_2, \quad ,\eta_k)$ due basi ortonormali rispettivamente di W_1 e W_2 tali che le coppie (ξ_i, η_i) $i=1, \quad ,k$ siano le più correlate possibile.
- Poiché le variabili sono centrate l' obiettivo è cercare le coppie (ξ_i, η_i) $i=1, \quad ,k$ tali che massimizzano il coseno dell' angolo formato da loro.
- Il vettore comune di due sottospazi (la prima bisettrice) rappresenta la soluzione banale.



Fonte: Stawinoga

- Il valore del coseno dell' angolo formato tra due variabili canoniche ξ ed η *è uguale al valore del loro coefficiente di correlazione.*

$$\cos(\xi,\eta) = \frac{a'X'Zb}{\sqrt{(a'X'Xa)(b'Z'Zb)}}$$

- *L' angolo tra due vettori non dipende dalla loro norma quindi possiamo porre:*

$$\|\xi\| = a'X'Xa = 1, \|\eta\| = b'Z'Zb = 1$$

- *Usando la funzione di Lagrange si possono ottenere sotto le condizioni di normalizzazione i vettori a e b che massimizzano la quantità a'X'Zb.*

$$L = a'X'Zb - \lambda(a'X'Xa - 1) - \mu(b'Z'Zb - 1)$$

Fonte: Stawinoga

- Derivando l'equazione di Lagrange rispetto ad *a* e *b* e ponendo i risultati uguali a 0 otteniamo:

$$
(1)\quad \begin{aligned} X'Zb - 2\lambda X'Xa &= 0 \\ Z'Xa - 2\mu Z'Zb &= 0 \end{aligned}
$$

- Sotto le condizioni di normalizzazione moltiplichiamo le equazioni riportate sopra rispettivamente per *a'* e *b'*:

$$
\begin{aligned} a'X'Zb &= 2\lambda \\ b'Z'Xa &= 2\mu \end{aligned}
$$

- Ricordando che il trasposto di uno scalare è lo scalare stesso otteniamo la quantità:

$$
\boxed{\beta = 2\lambda = a'X'Zb}
$$

Questa quantità è il coefficiente di correlazione massimo che abbiamo trovato.

Fonte: Stawinoga

- Il nostro sistema da risolvere lo possiamo scrivere nella forma:

$$(2)\quad \begin{cases} X'Zb = \beta X'Xa \\ Z'Xa = \beta Z'Zb \end{cases}$$

- Per poter risolvere questo sistema **le matrici *X'X* e *Z'Z* devono essere non singolari** (det(X'X) ≠0, det(Z'Z) ≠0)

- Ricavando *a* dalla prima equazione del sistema (2)e *b* dalla seconda otteniamo:

$$(3)\quad a = \frac{1}{\beta}(X'X)^{-1}X'Zb \qquad (4)\quad b = \frac{1}{\beta}(Z'Z)^{-1}Z'Xa$$

- Adesso sostituendo *a* nella equazione (4) si ottiene:

$$Z'X(X'X)^{-1}X'Zb = \beta^2 Z'Zb$$

Fonte: Stawinoga

- Il vettore *b* soluzione del sistema è quindi l'autovettore della matrice:

$$(Z'Z)^{-1}Z'X(X'X)^{-1}X'Z$$

associato al più grande autovalore β^2 .

- β^2 rappresenta il quadrato del coefficiente di correlazione tra ξ ed η.
- Il vettore *a* può essere calcolato dalla equazione (3) o come autovettore associato al più grande autovalore della matrice:

$$(X'X)^{-1}X'Z(Z'Z)^{-1}Z'X$$

- Per poter ottenere le variabili canoniche ξ *ed* η moltiplichiamo le equazioni (3) e (4) rispettivamente per X e Z.

Fonte: Stawinoga

- Si ottiene:

(5) $\xi = Xa = \frac{1}{\beta} \underbrace{X(X'X)^{-1}X'}_{P_1} Zb$

(6) $\eta = Zb = \frac{1}{\beta} \underbrace{Z(Z'Z)^{-1}Z'}_{P_2} Xa$

- Dalle equazioni (5) e (6) deriva un risultato importante. Le matrici:

$$P_1 = X(X'X)^{-1}X' \qquad P_2 = Z(Z'Z)^{-1}Z'$$

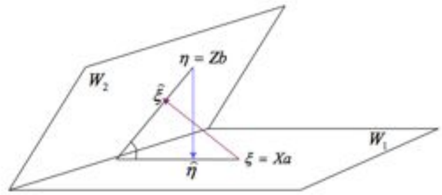
- Le matrici P_1 e P_2 sono simmetriche ed idempotenti. Possiamo considerarle come operatori di proiezione di R^n sui sottospazi W_1 W_2 generati dalle colonne delle matrici X e Z.

Fonte: Stawinoga

- Ciascun vettore η (oppure ξ) è collineare con la proiezione ortogonale dell' altro.

Quindi se η_1 è un vettore unitario di W_2 , il vettore di W_1 che forma l' angolo minimo con η_1 è il vettore $\boxed{\hat{\eta}_1}$ proiezione ortogonale di η_1 su W_1 .

Collinearità dei vettori $\hat{\eta}$ e ξ



Fonte: Stawinoga

Il legame dell' ACC con la regressione multipla

In questo caso la matrice Z è formata da una sola colonna($q = 1$).
La matrice Z è costituita dalla variabile da spiegare z e X è costituita dalle p variabili esplicative x_1, x_2, , x_p. Il vettore b ha una sola componente ed è quindi uno scalare, lo stesso il prodotto $Z'Z$.
Possiamo scrivere:

$$\beta^2 = \frac{Z'X(X'X)^{-1}X'Z}{Z'Z}$$

La quantità β^2 costituisce il coefficiente di correlazione multipla tra la variabile da spiegare e le variabili esplicative. Dalla (3) otteniamo:

$$a = \frac{b}{\beta}(X'X)^{-1}X'Z$$

Il vettore a è proporzionale al vettore dei coefficienti della regressione multipla con x_1, x_2, , x_p variabili esplicative e z la variabile dipendente.
Dalla condizione di normalizzazione si ricava:

$$b = \frac{1}{\sqrt{Z'Z}}$$

Fonte: Stawinoga

Il legame con l' Analisi delle Corrispondenze

In questo caso si considerano due matrici Z_1 e Z_2 di dimensioni rispettivamente (n, q_1) e (n, q_2).
L' Analisi delle Corrispondenze di una tabella di contingenza può essere vista come l' Analisi delle Correlazioni Canoniche della matrice $[Z_1 Z_2]$.
Nella matrice Z_1 vengono riportati i valori in codifica disgiuntiva completa delle q_1 modalità di una variabile X.
La matrice Z_2 è costituita dalle q_2 modalità di una variabile Y.
Tutte due osservate su n unità. Dal prodotto $Z_1'Z_2$ otteniamo la tabella di contingenza F di dimensioni (q_1, q_2).
Con D_1 e D_2 indichiamo le matrici diagonali rispettivamente dei marginali di riga e di colonna.
Otteniamo :

$$\varphi_1 = \frac{1}{\sqrt{\lambda}} D_1^{-1} Z_1' Z_2 \varphi_2 \qquad \varphi_2 = \frac{1}{\sqrt{\lambda}} D_2^{-1} Z_2' Z_1 \varphi_1$$

Fonte: Stawinoga

Se moltiplichiamo per Z_1 e Z_2 otteniamo:

$$Z_1\varphi_1 = \frac{1}{\sqrt{\lambda}} Z_1 D_1^{-1} Z_1' Z_2 \varphi_2 \qquad Z_2\varphi_2 = \frac{1}{\sqrt{\lambda}} Z_2 D_2^{-1} Z_2' Z_1 \varphi_1$$

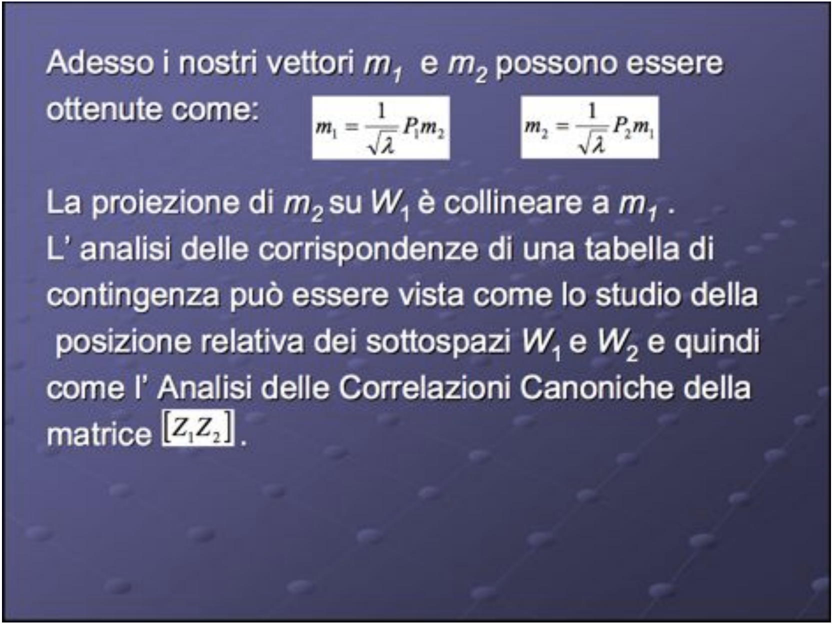
La matrice $Z = [Z_1 Z_2]$ ha $q_1 + q_2$ colonne con corrispondenti $q_1 + q_2$ punti nello spazio R^n. Ogni sottomatrice Z_i (i = 1,2) genera in R^n un sottospazio lineare W_i a q_i dimensioni. Le componenti del vettore φ_1 $(q_1, 1)$ costituiscono le coordinate di un punto m_1 nel sottospazio W_1.

Le coordinate di m_1 in R^n: $m_1 = Z_1\varphi_1$

Con P_1 e P_2 abbiamo indicato gli operatori di proiezione su W_1 e W_2.

$$P_1 = Z_1 (Z_1' Z_1)^{-1} Z_1' \qquad P_2 = Z_2 (Z_2' Z_2)^{-1} Z_2'$$

Fonte: Stawinoga



Adesso i nostri vettori m_1 e m_2 possono essere ottenute come: $m_1 = \frac{1}{\sqrt{\lambda}} P_1 m_2$ $m_2 = \frac{1}{\sqrt{\lambda}} P_2 m_1$

La proiezione di m_2 su W_1 è collineare a m_1 .
L' analisi delle corrispondenze di una tabella di contingenza può essere vista come lo studio della posizione relativa dei sottospazi W_1 e W_2 e quindi come l' Analisi delle Correlazioni Canoniche della matrice $[Z_1 Z_2]$.

Statistical Study $(\mathbf{Y}^T\mathbf{DX}, \mathbf{Q_X}, \mathbf{Q_Y})$

$$\mathbf{Q_X} = (\mathbf{X}^T\mathbf{DX})^{-1} \qquad \mathbf{Q_Y} = (\mathbf{Y}^T\mathbf{DY})^{-1}$$

$$\begin{tikzcd}
\Re^p \arrow[r, "(\mathbf{X}^T\mathbf{DX})^{-1}"] & \Re^{p*} \arrow[d, "\mathbf{Y}^T\mathbf{DX}"] \\
\Re^{q*} \arrow[u, "\mathbf{X}^T\mathbf{DY}"] & \Re^q \arrow[l, "(\mathbf{Y}^T\mathbf{DY})^{-1}"]
\end{tikzcd}$$

$$\begin{tikzcd}
\Re^p \arrow[r, "(\mathbf{X}^T\mathbf{DX})^{-1}"] & \Re^{p*} \arrow[d, "\mathbf{X}"] \\
\Re^{n*} \arrow[u, "\mathbf{X}^T"] & \Re^n \arrow[d, "\mathbf{D}"] \\
\Re^n \arrow[u, "\mathbf{D}"] & \Re^{n*} \arrow[d, "\mathbf{Y}^T"] \\
\Re^{q*} \arrow[u, "\mathbf{Y}"] & \Re^q \arrow[l, "(\mathbf{Y}^T\mathbf{DY})^{-1}"]
\end{tikzcd}$$

Statistical Study

$$(\hat{\mathbf{Y}}^T\mathbf{D}\hat{\mathbf{X}}, \mathbf{Q}_{\mathbf{X}}, \mathbf{Q}_{\mathbf{Y}})$$

$$\hat{\mathbf{X}} = \mathbf{X}(\mathbf{X}^T\mathbf{D}\mathbf{X})^{-1}$$

$$\hat{\mathbf{Y}} = \mathbf{Y}(\mathbf{Y}^T\mathbf{D}\mathbf{Y})^{-1}$$

$$\mathbf{Q}_{\mathbf{X}} = (\mathbf{X}^T\mathbf{D}\mathbf{X}),\ \mathbf{Q}_{\mathbf{Y}} = (\mathbf{Y}^T\mathbf{D}\mathbf{Y})$$

$$\begin{tikzcd}
\Re^{p} \arrow[r, "(\mathbf{X}^T\mathbf{D}\mathbf{X})"] & \Re^{p*} \arrow[d, "\hat{\mathbf{Y}}^T\mathbf{D}\hat{\mathbf{X}}"] \\
\Re^{q*} \arrow[u, "\hat{\mathbf{X}}^T\mathbf{D}\hat{\mathbf{Y}}"] & \Re^{q} \arrow[l, "(\mathbf{Y}^T\mathbf{D}\mathbf{Y})"]
\end{tikzcd}$$

$$\begin{tikzcd}
\Re^{p} \arrow[r, "(\mathbf{X}^T\mathbf{D}\mathbf{X})"] & \Re^{p*} \arrow[d, "(\mathbf{X}^T\mathbf{D}\mathbf{X})^{-1}"] \\
\Re^{p} \arrow[u, "(\mathbf{X}^T\mathbf{D}\mathbf{X})^{-1}"] & \Re^{p*} \arrow[d, "\mathbf{X}"] \\
\Re^{n*} \arrow[u, "\mathbf{X}^T"] & \Re^{n} \arrow[d, "\mathbf{D}"] \\
\Re^{n} \arrow[u, "\mathbf{D}"] & \Re^{n*} \arrow[d, "\mathbf{Y}^T"] \\
\Re^{q*} \arrow[u, "\mathbf{Y}"] & \Re^{q} \arrow[d, "(\mathbf{Y}^T\mathbf{D}\mathbf{Y})^{-1}"] \\
\Re^{q*} \arrow[u, "(\mathbf{Y}^T\mathbf{D}\mathbf{Y})^{-1}"] & \Re^{q} \arrow[l, "(\mathbf{Y}^T\mathbf{D}\mathbf{Y})"]
\end{tikzcd}$$

Fonte: Gherghi-Lauro

5-Un esempio

Si consideri la matrice seguente sulle cui righe sono rappresentate n=22 cantine friulane produttrici del vino bianco "Tocai friulano" e sulle cui colonne sono invece riportate le diverse votazioni (con punteggi da zero a 3) di un gruppo di esperti e di un gruppo di consumatori relativamente a p=q=4 variabili riguardanti il vino in questione: l'*aspetto*, l'*aroma*, il *gusto* e il *retrogusto*.

Matrice dei dati:

	ESPERTI				CONSUMATORI			
	ASP1	ARO1	GUS1	RET1	ASP2	ARO2	GUS2	RET2
Cantina 1	3,0	2,0	1,5	1,0	2,0	2,0	2,0	2,0
Cantina 2	3,0	2,0	2,0	2,0	2,0	2,0	2,0	3,0
Cantina 3	2,5	1,5	2,0	1,5	2,0	2,0	2,0	1,0
Cantina 4	3,0	2,0	2,0	2,0	2,0	2,0	2,0	1,0
:	:	:	:	:	:	:	:	:
Cantina 21	3,0	1,5	2,0	1,0	2,0	1,0	2,0	2,0
Cantina 22	3,0	2,0	1,5	2,0	2,0	2,0	3,0	1,0

Fonte: Gherghi-Lauro

Caratteristiche delle variabili

		Media	Sqm	Max	Min
Esperti	Asp1	2,932	0,228	3,0	2,0
	Aro1	1,795	0,359	2,0	1,0
	Gus1	1,727	0,419	2,0	1,0
	Ret1	1,568	0,484	2,0	1,0
Consumatori	Asp2	2,091	0,287	3,0	2,0
	Aro2	1,864	0,547	3,0	1,0
	Gus2	1,909	0,668	3,0	0,0
	Ret2	1,818	0,490	3,0	1,0

Fonte: Gherghi-Lauro

Matrici di correlazione nei due gruppi

Esperti

	Asp1	Aro1	Gus1	Ret1
Asp1	1,000			
Aro1	0,523	1,000		
Gus1	0,281	0,536	1,000	
Ret1	0,248	0,473	0,652	1,000

Consumatori

	Asp2	Asp2	Gus2	Ret2
Asp2	1,000			
Aro2	0,368	1,000		
Gus2	0,043	0,339	1,000	
Ret2	0,117	0,077	0,088	1,000

Matrice di correlazione tra i due gruppi

		Consumatori			
		Asp2	Asp2	Gus2	Ret2
Esperti	Asp1	0,094	-0,074	-0,041	0,092
	Aro1	-0,261	-0,489	-0,078	-0,082
	Gus1	-0,171	-0,360	-0,007	0,091
	Ret1	-0,045	-0,051	0,301	-0,044

Fonte: Gherghi-Lauro

Fattori canonici

		F1	F2	F3	F4
Esperti	Asp1	1,843	2,378	-2,620	-3,237
	Aro1	-2,713	-1,379	1,815	-1,308
	Gus1	-1,477	1,452	-1,930	1,788
	Ret1	1,566	-2,212	-0,547	-0,291
Consumatori	Asp2	0,900	0,598	-1,489	-3,299
	Aro2	1,493	0,538	0,976	0,959
	Gus2	0,242	-1,357	-0,818	0,005
	Re2	-0,127	1,035	-1,441	1,047

Autovalori, correlazioni canoniche e percentuali di variabilità spiegata

N.	Valore	Corr. Can.	%	%cum.	Istogramma
1	0,428	0,654	65,2	65,2	•••••••••••••••••••••••••••••••••
2	0,169	0,410	25,7	90,9	•••••••••••••
3	0,050	0,224	7,6	98,5	••••
4	0,010	0,100	1,5	100,0	•

Correlazioni tra variabili iniziali e variabili canoniche (ξ_i e η_i, $i=1,\dots,4$)

		ξ_1	ξ_2	ξ_3	ξ_4
Esperti	Asp1	-0,07	0,19	-0,55	-0,81
	Aro1	-0,73	-0,39	-0,22	-0,52
	Gus1	-0,53	-0,20	-0,80	0,20
	Ret1	0,00	-0,77	-0,63	-0,06
Consumatori	Asp2	0,37	0,12	-0,08	-0,07
	Aro2	0,63	0,37	0,03	0,02
	Gus2	0,29	-0,31	-0,10	0,02
	Ret2	0,03	0,19	-0,17	0,04

		η_1	η_2	η_3	η_4
Consumatori	Asp2	0,56	0,30	-0,34	-0,70
	Aro2	0,96	0,09	0,14	0,22
	Gus2	0,44	-0,75	-0,45	0,19
	Ret2	0,05	0,47	-0,76	0,44
Esperti	Asp1	-0,05	0,08	-0,12	-0,08
	Aro1	-0,48	-0,16	-0,05	-0,05
	Gus1	-0,35	-0,08	-0,18	0,02
	Ret1	0,00	-0,32	-0,14	-0,01

Fonte: Gherghi-Lauro

Coseni degli angoli tra punti-unità e sottospazio canonico

	ξ_1	ξ_2	ξ_3	ξ_4	η_1	η_2	η_3	η_4
Ca1	0,56	0,73	0,91	1,00	0,19	0,23	0,25	1,00
Ca2	0,17	0,77	0,99	1,00	0,00	0,45	0,75	1,00
Ca3	0,21	0,22	0,22	1,00	0,14	0,57	0,97	1,00
Ca4	0,17	0,77	0,99	1,00	0,14	0,57	0,97	1,00
Ca5	0,61	0,61	0,96	1,00	0,19	0,23	0,25	1,00
Ca6	0,17	0,77	0,99	1,00	0,19	0,23	0,25	1,00
Ca7	0,17	0,77	0,99	1,00	0,85	0,98	0,98	1,00
Ca8	0,17	0,77	0,99	1,00	0,19	0,23	0,25	1,00
Ca9	0,75	1,00	1,00	1,00	0,07	0,83	0,94	1,00
Ca10	0,17	0,77	0,99	1,00	0,40	0,75	0,99	1,00
Ca11	0,17	0,77	0,99	1,00	0,19	0,23	0,25	1,00
Ca12	0,17	0,77	0,99	1,00	0,14	0,57	0,97	1,00
Ca13	0,17	0,77	0,99	1,00	0,19	0,23	0,25	1,00
Ca14	0,10	0,10	0,76	1,00	0,21	0,77	0,93	1,00
Ca15	0,17	0,77	0,99	1,00	0,31	0,37	0,60	1,00
Ca16	0,15	0,27	0,71	1,00	0,19	0,23	0,25	1,00
Ca17	0,86	0,99	0,99	1,00	0,64	0,69	0,78	1,00
Ca18	0,75	1,00	1,00	1,00	0,85	0,98	0,98	1,00
Ca19	0,10	0,10	0,76	1,00	0,53	0,65	0,90	1,00
Ca20	0,86	0,99	0,99	1,00	0,76	0,84	0,86	1,00
Ca21	0,14	0,88	0,95	1,00	0,76	0,80	0,98	1,00
Ca22	0,31	0,81	0,84	1,00	0,20	0,96	0,98	1,00

Fonte: Gherghi-Lauro

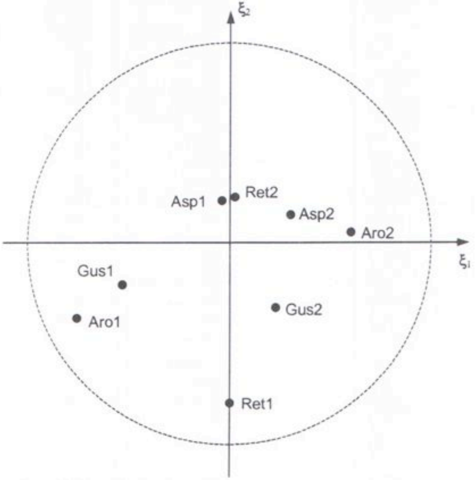


Figura 4 – *Cerchio delle correlazioni*

Fonte: Gherghi-Lauro

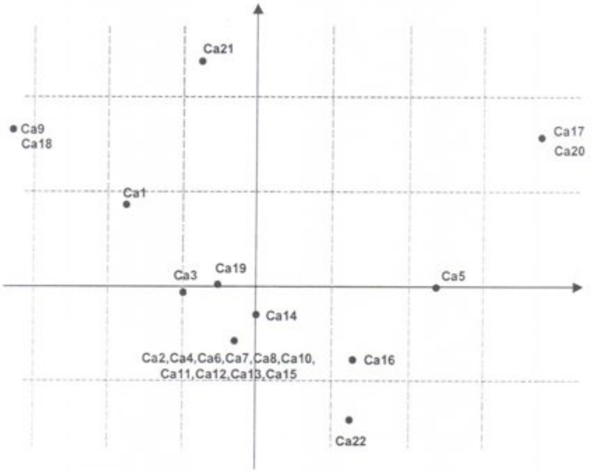


Figura 5 – *Rappresentazione dei punti-unità*

Distribution of butterfly

The data concern **four environmental variables**

(altitude, annual precipitation, and the minimum and maximum temperatures)

and **four genetic variables**

(percentage frequencies for different phosphoglucose-isomerase [Pgi] genes as determined by the technique of electrophoresis)

regarding **16 colonies of the butterfly** *Euphydryas editha*
in California and Oregon (McKechnie et al., 1975; Manly, 2005) .

The study of the relationships between the environmental and genetic variables could indicate the adaptation of the *Euphydryas editha* to the local environments.

The environmental variables have been treated as **X** variables and the gene frequencies as the **Y** variables.

For a deeper analysis of this data set see also Manly (2005).

Fonte: Manly.

	Altitude	Rainfall	MaxTp	MinTp	G(-)	G(0.80)	G(1.00)	G(1.16)
SS	500	43	98	17	3	22	57	17
SB	808	20	92	32	16	20	38	13
WSB	570	28	98	26	6	28	46	17
JRC	550	28	98	26	4	19	47	27
JRH	550	28	98	26	1	8	50	35
SJ	380	15	99	28	2	19	44	32
CR	930	21	99	28	0	15	50	27
UO	650	10	101	27	31	40	25	4
LO	600	10	101	27	40	32	28	0
DP	1500	19	99	23	1	6	80	12
PZ	1750	22	101	27	5	34	33	22
MC	2000	58	100	18	7	14	66	13
IF	2500	34	102	16	9	15	47	21
AF	2000	21	105	20	10	17	32	27
GH	7850	42	84	5	5	7	84	4
GL	10500	50	81	-12	3	1	92	4

Fonte: Manly.

Table 10.1 Correlation Matrix for Variables Measured on Colonies of *Euphydryas editha*, Partitioned into A, B, C, and C′ Submatrices

	X_1	X_2	X_3	X_4	Y_1	Y_2	Y_3	Y_4
X_1	1.000	0.568	−0.828	−0.936	−0.201	−0.573	0.727	−0.458
X_2	0.568	1.000	−0.479	−0.705	−0.468	−0.550	0.699	−0.138
X_3	−0.828	−0.479	1.000	0.719	0.224	0.536	−0.717	0.438
X_4	−0.936	0.705	0.719	1.000	0.246	0.593	−0.759	0.412
				A	C			
				C′	B			
Y_1	−0.201	−0.468	0.224	0.246	1.000	0.638	−0.561	−0.584
Y_2	−0.573	−0.550	0.536	0.593	0.638	1.000	−0.824	−0.127
Y_3	0.727	0.699	−0.717	−0.759	−0.561	−0.824	1.000	−0.264
Y_4	−0.458	−0.138	0.438	0.412	−0.584	−0.127	−0.264	1.000

Fonte: Manly.

The eigenvalues obtained from Equation 10.1 are 0.7425, 0.2049, 0.1425, and 0.0069. Calculation of square roots gives the corresponding canonical correlations of 0.8617, 0.4527, 0.3775, and 0.0833, respectively, and the canonical variables are found to be as follows:

$$
\begin{aligned}
U_1 &= -0.09X_1 - 0.29X_2 + 0.48X_3 + 0.29X_4 \\
V_1 &= +0.54Y_1 + 0.42Y_2 - 0.10Y_3 + 0.82Y_4 \\
U_2 &= +2.31X_1 - 0.73X_2 + 0.45X_3 + 1.27X_4 \\
V_2 &= -1.66Y_1 - 2.20Y_2 - 3.71Y_3 + 2.77Y_4 \\
U_3 &= +3.02X_1 + 1.33X_2 + 0.57X_3 + 3.58X_4 \\
V_3 &= -3.56Y_1 - 1.35Y_2 - 3.86Y_3 - 2.86Y_4 \\
U_4 &= +1.43X_1 + 0.26X_2 + 1.72X_3 - 0.03X_4 \\
V_4 &= +0.60Y_1 - 1.44Y_2 - 0.58Y_3 + 0.58Y_4
\end{aligned}
$$

Fonte: Manly.

The correlations between the environmental variables and U_1 are:

	Altitude	Precipitation	Maximum temperature	Minimum temperature
U_1	−0.92	−0.77	0.90	0.92

This suggests that U_1 is best interpreted as a measure of high temperatures and low altitude and precipitation. The correlations between V_1 and the gene frequencies are:

	Mobility 0.40/0.60	Mobility 0.80	Mobility 1.00	Mobility 1.16
V_1	0.38	0.74	−0.96	0.49

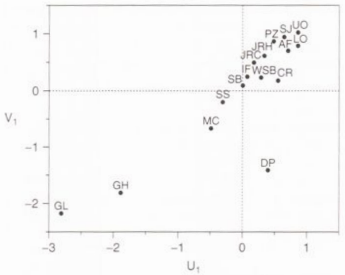
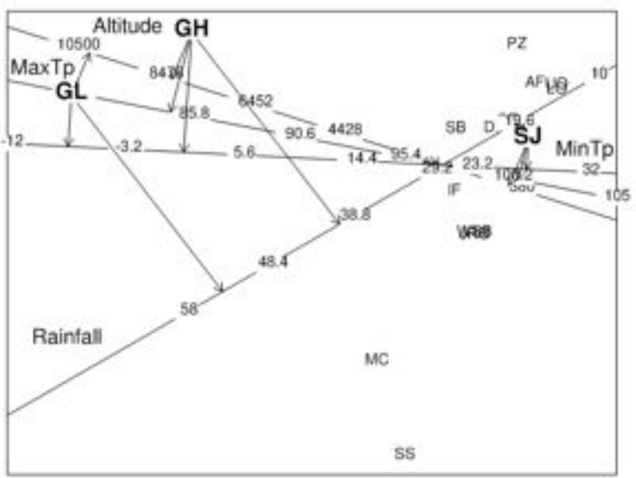


Figure 10.1 Plot of V_1 against U_1 for 16 colonies of *Euphydryas editha*.

Fonte: Manly.


Altitude
GH
10500
MaxTp
GL
6452
85.8
90.6
4428
95.4
-12
-3.2
5.6
14.4
23.2
SB
D
SJ
MinTp
PZ
10
32
105
IF
38.8
48.4
58
Rainfall
MC
SS