

Università degli Studi di Milano-Bicocca – Facoltà di Economia
Esame di Analisi dei Dati (modulo B)
 8 settembre 2004
N.B.: Commentare sempre i risultati ottenuti

1) Si illustri la situazione nella quale la matrice di varianze e covarianze, valutata su p variabili quantitative, può essere scomposta in somma di due matrici e se ne fornisca la dimostrazione analitica.

2) Si è interessati a valutare la relazione esistente fra le previsioni del tempo effettuate da una società di previsioni meteo e il tempo effettivamente osservato. Nella seguente tabella è riportata la distribuzione congiunta di frequenze delle variabili $X =$ “Tempo previsto” e $Y =$ “Tempo osservato” su un campione di 100 giorni scelti a caso nell’arco di un anno.

Y X	Pioggia	Variabile	Sereno	TOTALE
Pioggia	19	11	2	32
Variabile	15	5	14	34
Sereno	3	7	24	34
TOTALE	37	23	40	100

Sono di seguito riportati gli autovalori e gli autovettori di norma chi-quadrato unitaria della matrice $S = P_r^T P_c$, essendo P_r e P_c rispettivamente la matrice dei profili riga e la matrice dei profili colonna nel problema considerato (l’autovalore unitario e il corrispondente autovettore sono stati omessi):

$$\lambda_1 = 0.29832 \qquad \lambda_2 = 0.02239$$

$$v_1 = \begin{bmatrix} -0.37843 \\ -0.09914 \\ 0.47758 \end{bmatrix} \qquad v_2 = \begin{bmatrix} 0.29983 \\ -0.40898 \\ 0.10919 \end{bmatrix}$$

- a) Si determinino e si rappresentino graficamente le proiezioni dei profili riga e le proiezioni dei profili colonna sugli assi principali estraibili.
- b) Si calcolino i contributi assoluti e quelli relativi ritenuti utili per il tipo di analisi e si commentino i risultati ottenuti.

3) Si descrivano gli obiettivi dell’analisi discriminante e si illustri la metodologia qualora il numero di gruppi nei quali è suddivisa la popolazione sia pari a due.

4) Nella seguente tabella 100 contratti assicurativi sono stati classificati in base al tipo di rischio coperto e al tipo di canale distributivo.

Canale distributivo	Agenzia assicurativa	Banca	Intermediario assicurativo	Totale
Rischio				
Responsabilità civile auto	30	4	14	48
Infortuni	15	2	5	22
Malattia	3	4	3	10
Furto	14	2	4	20
Totale	62	12	26	100

Si descriva il modello log-lineare di indipendenza distributiva per la tabella sopra riportata e si forniscano le stime di massima verosimiglianza dei parametri di tale modello.

5) La distanza chi-quadrato: definizione, proprietà, impieghi e relazione con la misura di connessione X^2 .