Politecnico di Milano
Facoltà di Ingegneria dei Sistemi
II Appello di Statistica Applicata
28 Febbraio 2007

Nome e cognome:
Numero di matricola:

Problema 1

Siano $X \sim N_p(\mu, \Sigma)$, $p \geq 2$ e $rango(\Sigma) = s$.
Si dimostri che se $1 \leq s \leq p-1$ e Σ^- è l'inversa generalizzata di Σ allora:

$$(X-\mu)'\Sigma^-(X-\mu) \sim \chi(s)$$

Problema 2

Nel dataset `Pb2.txt` sono riportate le temperature medie mensili (^oC) registrate nel 2006 in tre località canadesi: Edmonton, Resolute e Montreal.
È frequente in meteorologia assumere che le temperature medie mensili oscillino sinusoidalmente attorno ad un valor medio annuale:

$$Temp_g(t) = \beta_{0g} + \beta_{1g}\ \sin\left(\frac{2\pi}{12}t\right) + \beta_{2g}\ \cos\left(\frac{2\pi}{12}t\right) + \epsilon$$

con $\epsilon \sim N(0, \sigma^2)$, $t = 1, 2, 3, \ldots, 12$ (mese) e $g = Edmonton, Resolute, Montreal$ (località).

(a) Utilizzando il metodo dei minimi quadrati si stimino i 10 parametri del modello.

(b) Si verifichino le assunzioni del modello.

(c) Sfruttando la nota relazione trigonometrica $\sin(\alpha - \beta) = \sin(\alpha)\cos(\beta) - \cos(\alpha)\sin(\beta)$ e reinterpretando il modello nella forma:

$$Temp_g(t) = \mu_g + A_g\ \sin\left(\frac{2\pi}{12}(t - \varphi_g)\right) + \epsilon$$

si riportino le relazioni analitiche che legano i nuovi parametri (μ_g, A_g, φ_g) ai vecchi parametri $(\beta_{0g}, \beta_{1g}, \beta_{2g})$.

(d) Si stimino i parametri della nuova formulazione, ovvero:

- I valori medi annuali (μ_g).
- Le ampiezze delle oscillazioni (A_g).
- Le fasi delle oscillazioni (φ_g).

(e) Mediante l'utilizzo di un opportuno test statistico (del quale si riporti il p-value) si giustifichi la possibilità di utilizzare un modello ridotto che assuma oscillazioni di ampiezza e fase uguali, ma valori medi annuali diversi, per le due centraline di Edmonton e Montreal.

Problema 3

Nel dataset `Pb3.txt` sono riportate Lunghezza (cm) e Larghezza (cm) del torace di 50 passeri. Il biologo che ha raccolto le misure vuol dimostrare che i passeri si possono suddividere in due gruppi dalle caratteristiche ben distinte in termini di Lunghezza e Larghezza del torace. Aiutatelo a provare la sua teoria implementando e commentando le seguenti analisi:

(a) Mediante un algoritmo di clustering gerarchico agglomerativo che utilizzi la distanza di Manhattan ed il Single linkage si affermi se è ragionevole assumere una clusterizzazione in due gruppi.

(b) Si implementi un test per provare la differenza delle medie dei due gruppi.

(c) Si identifichino e si commentino i quattro intervalli di Bonferroni di confidenza globale 90% (estremo inferiore, valore centrale, estremo superiore) per:

- La differenza delle medie della variabile Lunghezza.
- La differenza delle medie della variabile Larghezza.
- La differenza delle medie della somma delle variabili Lunghezza e Larghezza.
- La differenza delle medie della differenza delle variabili Lunghezza e Larghezza.

(d) Si verifichino le assunzioni necessarie all'implementazione del test.

Problema 4

Si vuole individuare un criterio di classificazione per due popolazioni bivariate A e B tali che:

$$
\begin{aligned}
\pi_A &= 1/2 \\
\pi_B &= 1/2 \\
f_A(x) &\sim N_2(0, \Sigma) \\
f_B(x) &\sim N_2(0, \lambda\Sigma)
\end{aligned}
$$

Con Σ definita positiva e $\lambda \in (0, +\infty)$.

(a) Al variare di λ in $(0, +\infty)$ si riportino le espressioni analitiche di $R_A(\lambda)$ e di $R_B(\lambda)$.

(b) Al variare di λ in $(0, +\infty)$ si riporti l'espressione analitica di $AER(\lambda)$ e si riporti un grafico qualitativo $AER = AER(\lambda)$. (Potete farvi aiutare da R ...)

(c) Qual è il valore di $AER(\lambda)$ per $\lambda = 10^{-1}, 10^{0}, 10^{+1}$.