

Politecnico di Milano
Facoltà di Ingegneria dei Sistemi
APPELLO DI STATISTICA APPLICATA
29 Giugno 2010

©I diritti d'autore sono riservati. Ogni sfruttamento commerciale non autorizzato sarà perseguito.

Nome e cognome:
Numero di matricola:

Problema 1

Il noto etologo James Viper sostiene che la tana del famoso mostro di Lochness, si trovi in fondo all'omonimo lago in una posizione X [espressa in km a Est e a Nord dalla città di Inverness] modellizzabile tramite una legge normale di media $\mu = (30, 50)$ e matrice di covarianza $\Sigma = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$. Assumendo corretta la teoria di James Viper:

- a) si identifichi una regione ellissoidale nella quale con probabilità 99% si trovi la tana (si riportino in particolare il centro, la direzione e la lunghezza dei semiassi);
- b) si calcoli la distanza quadratica media della tana dalla cittadina di Snakeburgh, situata a 31 km a Est e 52 km Nord di Inverness.

Problema 2

Nel file `england.txt` sono riportati il numero medio giornaliero di rapine, di furti in casa e di borseggi nelle 82 contee dell'Inghilterra nell'anno 2008.

- a) Mediante un algoritmo di clustering gerarchico agglomerativo basato sulla distanza di Manhattan e il *single linkage* si identifichino eventuali cluster presenti nei dati.
- b) Una volta identificati i due cluster di numerosità maggiore, se ne quantifichino le differenze mediante un test pooled per il confronto tra le rispettive medie.
- c) Mediante tre intervalli T^2 -simultanei si mettano in luce e si interpretino le differenze tra i due cluster.

Problema 3

Nel file `scotland.txt` sono riportati il numero di abitanti residenti in Scozia secondo l'ultimo censimento del 2001, suddivisi per fascia di età e per contea. Assumendo indipendenti e identicamente distribuiti i dati relativi alle diverse contee ed invece dipendenti i dati relativi a diverse fasce di età all'interno della singola contea:

- a) si esegua un test per verificare l'ipotesi secondo la quale il numero medio di residenti non cambia al variare della fascia d'età;
- b) si fornisca un intervallo di confidenza 95% per il numero medio di abitanti per contea;
- c) si effettui una riduzione dimensionale del data set mediante un'analisi delle componenti principali e se ne interpretino le componenti.

Problema 4

Nel file `energy.txt` sono riportati i consumi elettrici istantanei [MW] allo scoccare di ogni ora nella città di Belfast, per il mese di Giugno. Assumendo le misurazioni indipendenti e normalmente distribuite con pari varianza e media $\mu = A_g + B_g [1 - \cos(\frac{2\pi}{24}t)]$ dipendente dall'ora di misurazione t ($t = 0, 1, 2, \dots, 23$) e dalla natura infrasettimanale o meno del giorno di misurazione ($g \in \{\text{"Lun-Ven"}, \text{"Sab-Dom"}\}$).

- a) Si stimino i parametri del modello.
- b) Si fornisca, se possibile, un opportuno modello ridotto, se ne giustifichi la scelta e se ne interpretino i parametri.
- c) Sulla base del modello ridotto (b), si forniscano degli intervalli di confidenza globale 90% per la media e la varianza dei consumi raggiunti in data odierna (martedì 29 giugno 2010) alle ore 12:00 ed alle ore 18:00.
- d) Sulla base del modello ridotto (b), vi è evidenza che di venerdì il consumo medio massimo sia più del doppio del consumo medio minimo?
- e) Sulla base del modello ridotto (b), vi è evidenza che di sabato il consumo medio massimo sia più del doppio del consumo medio minimo?