

Politecnico di Milano
Facoltà di Ingegneria dei Sistemi
APPELLO DI STATISTICA APPLICATA
21 Settembre 2009

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

Nome e cognome:

Numero di matricola:

Problema 1

Nel file `sco2009.txt` sono riportati i tempi [minuti] dei diversi intervenanti della conferenza S.Co.2009 (18 sessioni di 4 interventi ciascuna). Assumendo un modello normale quadrivariato per i tempi relativi al primo, secondo, terzo e quarto intervento della stessa sessione:

- a) Vi è evidenza statistica per affermare che le durate medie dei quattro interventi siano diverse?
- b) Utilizzando la disuguaglianza di Bonferroni, si forniscano otto intervalli di confidenza globale 90% per la media e la varianza del primo, del secondo, del terzo e del quarto intervento.

Sulla base del modello introdotto, utilizzando le stime campionarie della media e della matrice di covarianza e sapendo che il tempo massimo a disposizione di ogni sessione è di 2 ore:

- c) Stimate con che probabilità il tempo a disposizione per le domande conclusive sia inferiore a 15 minuti.
- b) Stimate con che probabilità la somma dei tempi dei quattro interventi ecceda il tempo massimo di 2 ore.

Problema 2

Nel file 100m sono contenuti i tempi [secondi] impiegati per correre i 100m piani dai 20 studenti iscritti alla squadra di atletica del Politecnico. Per ogni studente sono riportati 4 tempi: al rientro dalle vacanze estive e una, due e tre settimane dal rientro. Inquadrando il problema nell'ottica di un problema di misure ripetute:

- a) Vi è evidenza statistica per affermare che i tempi medi cambino nel tempo?
- b) Utilizzando opportuni intervalli di confidenza descrivete la dinamica temporale della media dei tempi.
- c) Dopo quante settimane si può pensare che la media dei tempi si sia stabilizzata?

Problema 3

Nel periodo estivo centinaia di esemplari di balene si trasferiscono nel Golfo del Maine per nutrirsi in vista dell'Inverno. In questo periodo le balene si organizzano in colonie ben separate e ben localizzate geograficamente. Nel file `whales.txt` sono riportati le coordinate geografiche degli avvistamenti relativi alla zona di mare di fronte a Boston nella quale è nota la presenza di 2 colonie.

- a) Utilizzando un algoritmo di clustering gerarchico agglomerativo (metrica euclidea e Ward linkage), si individuino gli avvistamenti relativi alle due colonie.
- b) Si discuta la bontà dell'algoritmo.
- c) Assumendo corretta la suddivisione ottenuta al punto (a) ed introducendo opportune ipotesi, si fornisca una stima, sia puntuale che mediante una regione di confidenza 90%, del dislocamento relativo delle due colonie.

Problema 4

Nel file `BGMI.txt` sono riportati i tempi di percorrenza [ore] della tratta autostradale Bergamo-Milano, l'orario di partenza da Bergamo [ore] e il giorno della settimana di 1127 autoveicoli partiti tra le 6:00 e le 9:00. Assumendo una dipendenza al più quadratica della media del tempo di percorrenza dal tempo di partenza e diversa tra giorni feriali e festivi:

$$T_{\text{percorrenza}} = \alpha_g + \beta_g(t_{\text{partenza}} - 7.5) + \gamma_g(t_{\text{partenza}} - 7.5)^2 + \epsilon$$

con $\epsilon \sim N(0, \sigma^2)$ e $g \in \{\text{feriale}, \text{festivo}\}$.

- a) Si forniscano le stime ai minimi quadrati dei parametri del modello.
- b) Sulla base di un opportuno test, vi è evidenza statistica di un diverso andamento della media nei giorni feriali e nei giorni festivi?
- c) Si costruisca un opportuno modello ridotto a 4 gradi di libertà e se ne stimino i parametri.
- d) Sulla base del modello (c), si forniscano delle stime intervallari (confidenza globale 90%) del valore della curva di regressione e della sua derivata in corrispondenza di una partenza feriale e/o festiva alle 7:30.