

Politecnico di Milano
II Scuola - Ingegneria dei Sistemi (MI)
APPELLO DI STATISTICA APPLICATA
29 giugno 2011

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

Nome e cognome:

Numero di matricola:

Problema 1

L'Autority spagnola per la paella ha misurato il contenuto di vongole [g] e gamberetti [g] di alcune confezioni da 200g di "Paella del Cantabro" (file `cantabro.txt`) al fine di verificare che la media del contenuto di vongole e gamberetti non sia significativamente diversa da quanto dichiarato sull'etichetta: 30g di vongole e 50g di gamberetti.

- a) Si esegua un opportuno test statistico per verificare quanto detto sopra.
- b) Si commenti il test precedente utilizzando tre intervalli T^2 -simultanei (di confidenza globale 99%) per la media del contenuto di vongole, del contenuto di gamberetti e della loro somma.
- c) Ritenete necessario un intervento dell'Autority? Motivate la risposta.

Problema 2

Nella baia di Santander sfociano due fiumi le cui sabbie trasportate vanno ad alimentare le spiagge della costa limitrofa. Un giovane geochimico ha misurato le concentrazioni di silicio, calcio e magnesio in 1000 granelli di sabbia raccolti dal suo asciugamano (file `sand.txt`). Vi chiede una consulenza affinché lo aiutate a dimostrare la presenza nella baia di due sabbie ben distinte.

- a) Utilizzando un algoritmo di clustering gerarchico agglomerativo (distanza Manhattan e Ward linkage) raggruppare i granelli in due gruppi (se ne riportino le numerosità e le concentrazioni medie).
- b) Si esegua un test al 90% per provare la diversa composizione chimica media delle due sabbie.
- c) Utilizzando tre intervalli di Bonferroni di confidenza globale 90% si identificano gli elementi con concentrazioni significativamente diverse nei due gruppi.

Problema 3

Juan de los Euros, noto tassista di Santander, ha cronometrato la durata degli ultimi suoi viaggi da/per l'aeroporto (file `time.txt`).

- a) Dopo aver fittato un modello anova additivo a due fattori (centro-aero/aero-centro, feriale/festivo) si stimino puntualmente le medie e le varianze dei quattro possibili tipi di viaggio.
- b) Sulla base del modello (a) si esegua un test al 90% per provare la significatività del fattore centro-aero/aero-centro.
- c) Sulla base del modello (a) si esegua un test al 90% per provare la significatività del fattore feriale/festivo.
- d) Sulla base dei test (b) e (c) si proponga un eventuale modello ridotto e si ristimino puntualmente – coerentemente col modello ridotto – le medie e le varianze dei quattro possibili tipi di viaggio.

Problema 4

Nel file `people.txt` sono riportati i quintali di rifiuti raccolti mensilmente nella città di Santander a partire dal gennaio 2009 ($t = 1$) fino al maggio 2011 ($t = 29$). Assumendo un modello del tipo:

$$\text{Rifiuti} = A + B \cdot t + C \cdot \left(1 - \cos\left(\frac{2\pi}{12}t\right)\right) + \epsilon \quad \text{con} \quad \epsilon \sim N(0, \sigma^2)$$

e attribuendo ai residenti i primi due addendi mentre ai turisti il terzo addendo, si risponda ai seguenti quesiti.

- a) Si stimino i parametri del modello
- b) Sulla base del modello (a), vi è evidenza statistica di una crescita imputabile ai residenti?
- c) Sulla base del modello (a), vi è evidenza statistica di un contributo significativo da parte dei turisti?
- d) L'Università della Cantabria ritiene che la crescita imputabile ai residenti sia quantificabile in un incremento di 10 quintali al mese. Possiamo smentire questa affermazione?
- e) Sulla base dei test (b), (c) e (d) si proponga un eventuale modello ridotto e/o vincolato e se ne stimino i parametri.
- f) Sulla base del modello (e), si forniscano tre previsioni puntuali per i rifiuti che verranno raccolti nel giugno 2011, per i rifiuti che verranno raccolti nel giugno 2011 imputabili ai residenti e per quelli che verranno raccolti nel giugno 2011 imputabili ai turisti.