

Politecnico di Milano
II Scuola - Ingegneria dei Sistemi (MI)
APPELLO DI STATISTICA APPLICATA
16 luglio 2013

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

Nome e cognome:
Numero di matricola:

Problema 1

Il DUNESCO ha intenzione di inserire anche i trulli di Alberobritto tra i patrimoni dell'umanità. Pertanto, vi chiede di verificare la correttezza di alcune informazioni relative alle caratteristiche medie dei trulli contenute nel documento descrittivo dell'opera. Il file `trulli.txt` contiene le misurazioni relative a 80 trulli, per ognuno dei quali sono riportate le misure di diametro esterno [m], altezza del tronco [m], altezza del tetto [m].

- a) Si esegua un test statistico per verificare se le misure medie dei trulli di Alberobritto siano rispettivamente di 5m, 2m e 2.5m.
- b) Si forniscano 3 intervalli T^2 -simultanei di livello globale 95% a supporto del test al punto a) e si commentino i risultati. Ritenete necessario che il DUNESCO modifichi la descrizione dell'opera?
- c) Mantenendo lo stesso livello di confidenza, si individui un intervallo di confidenza T^2 -simultaneo per la lunghezza media della muratura a secco a cinzione dei trulli (si approssimi il diametro della cinzione al diametro esterno del trullo).

Problema 2

L'Unione Pescatori dello Stretto (UPS) ritiene che le quantità di pescato nelle acque dello Stretto di Messina dipendano dalle caratteristiche fisico-chimiche delle acque stesse, che alternativamente provengono dal bacino ionico e dal bacino tirrenico. L'UPS ha pertanto deciso di monitorare l'effetto delle correnti marine sul pescato; il file `fishing.txt` raccoglie i dati relativi alla quantità [kg] di orate, di mormore e di cernie pescate al largo di Scilla (RC) nei 90 giorni della scorsa stagione estiva.

- a) Utilizzando un algoritmo di clustering gerarchico agglomerativo (distanza Euclidea e Ward linkage) si individuino due gruppi di giorni di pesca (se ne riportino le numerosità e medie stimate).
- b) Si verifichi attraverso un test di livello 5% se la quantità media di pescato differisca tra i due gruppi di giorni.
- c) Attraverso tre intervalli di Bonferroni (livello globale 95%) per le differenze medie delle quantità di pescato tra i gruppi individuati al punto a), si segnalino all'UPS le specie il cui pescato risulta significativamente influenzato dal fenomeno delle correnti marine.

Problema 3

Ogni anno a Palmi (RC), il 16 agosto, si svolge la processione di San Rocco. Il file `sanrocco.txt` raccoglie la durata [ore] della processione negli ultimi 46 anni, i livelli di calore registrati davanti alla chiesa di San Rocco (caldo ($<35^{\circ}\text{C}$)/torrido ($\geq 35^{\circ}\text{C}$)) e la presenza o assenza di vento (ventilato/fermo).

- a) Si adatti un modello ANOVA additivo a due fattori (caldo/torrido; ventilato/fermo) e si forniscano delle stime puntuali delle medie e delle varianze della durata della processione nelle quattro condizioni climatiche individuate.
- b) Sulla base del modello (a) si esegua un test al 95% per provare la significatività del fattore temperatura.
- c) Sulla base del modello (a) si esegua un test al 95% per provare la significatività del fattore vento.
- d) Coerentemente con i risultati ai punti (b) e (c) si individui un eventuale modello ridotto e si ristimino di conseguenza le medie e le varianze della durata della processione per i gruppi omogenei di condizioni climatiche individuate.

Problema 4

La catena di fastfood *Non solo orecchiette e cime di rapa* vende esclusivamente orecchiette con le cime di rapa e conta 57 punti vendita in Puglia e 4 in Lombardia. Nei file `puglia.txt` e `lombardia.txt` sono riportate le spese sostenute dai 61 punti vendita nell'anno 2012 [euro] e le corrispondenti quantità di orecchiette [kg] e cime di rapa [kg] utilizzate.

- a) Assumendo indipendenza e uguale variabilità tra le spese sostenute nei 61 punti vendita e un modello economico (diverso per le due regioni) per la spesa totale annuale del tipo

$$Spesa_{totale} = C_{fissi} + C_{orecchiette} \cdot Quantita_{orecchiette} + C_{cime} \cdot Quantita_{cime} + \epsilon ,$$

si stimino i costi fissi per punto vendita, i costi unitari delle orecchiette e i costi unitari delle cime di rapa per le due regioni.

- b) Sulla base del modello (a) vi è evidenza statistica per affermare che sia necessario utilizzare due modelli economici diversi in Puglia e in Lombardia?
- c) Sulla base del modello (a) vi è evidenza statistica per affermare che i costi fissi per punto vendita, i costi unitari delle orecchiette e i costi unitari delle cime di rapa siano diversi in Puglia e in Lombardia? Si eseguano tre distinti test a riguardo e si commentino i p -value.
- d) Sulla base delle risposte alle domande precedenti si fornisca un eventuale modello ridotto e si ristimino coerentemente i parametri del modello.
- e) Sulla base del modello (f) si fornisca un intervallo di confidenza 95% per la deviazione standard della spesa totale annuale per punto vendita.