

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

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

Nome e cognome:

Numero di matricola:

Problema 1

Nei file `risoimport.txt` e `risoexport.txt` sono riportati i dati giornalieri relativi alle importazioni e alle esportazioni cinesi di cinque qualità di riso (in migliaia di tonnellate) dal 1 gennaio al 31 dicembre 2012. Assumendo indipendenti i dati relativi a giorni diversi, si risponda ai seguenti quesiti.

- a) Vi è evidenza statistica per affermare che le importazioni medie di riso differiscano dalle esportazioni medie?
- b) Si forniscano 5 intervalli di confidenza T^2 simultanei (livello globale 95%) a supporto del test al punto (a) e si commentino i risultati.
- c) Mantenendo il livello globale al 95%, si fornisca un ulteriore intervallo di confidenza per la differenza media tra le importazioni totali e le esportazioni totali di riso e si commenti il risultato.

Problema 2

L'Istituto Cinese di Genomica ha ricostruito la mappa genetica di 200 individui di nazionalità cinese, 100 dei quali allergici alle mandorle e 100 non allergici alle mandorle. Nei file `hatingalmonds.txt` e `lovingalmonds.txt` sono registrate, rispettivamente per gli individui appartenenti ai due gruppi, la presenza (1) o l'assenza (0) di 520 mutazioni note. I genetisti sospettano che alcune di queste mutazioni possano essere positivamente associate all'allergia alle mandorle e per questo decidono di confrontare l'incidenza delle 520 mutazioni nelle due popolazioni.

- a) Si riportino le mutazioni significativamente influenti imponendo una probabilità al massimo dell'1% che la singola mutazione, se non influente, venga erroneamente giudicata influente.
- b) Si riportino le mutazioni significativamente influenti imponendo una probabilità al massimo dell'1% che almeno una tra le mutazioni non influenti venga erroneamente giudicata influente.
- c) Si riportino le mutazioni significativamente influenti imponendo che la proporzione attesa di mutazioni giudicate erroneamente influenti tra quelle giudicate influenti sia minore o uguale al 1%.

Problema 3

L'archeologo cinese Mai Trovai ha recentemente trovato nelle acque del Lago Dongting i resti di un'imbarcazione di lunghezza 12.73 metri. Volendo datare l'imbarcazione, vi chiede di attribuire l'imbarcazione alla dinastia Yuan (1271-1368, 97 anni) o alla dinastia Ming (1368-1644, 276 anni) partendo dalle lunghezze di 103 imbarcazioni di indubbia attribuzione (file `dongting.txt`).

- a) Assumendo che la lunghezza delle imbarcazioni segua una distribuzione uniforme e che le probabilità a priori possano assumersi proporzionali alla durata delle dinastie si costruisca un opportuno classificatore. Si riportino in particolare su di un grafico le probabilità a posteriori e le corrispondenti regioni di classificazione.
[Si ricordi che gli stimatori di massima verosimiglianza per il valor minimo e il valor massimo di una distribuzione uniforme sono rispettivamente il minimo e il massimo campionario]
- b) Si calcoli e commenti l'APER del classificatore.
- c) Si attribuisca l'imbarcazione ad una delle due dinastie fornendo anche la corrispondente probabilità di errore.

Problema 4

L'azienda artigianale Maog Net produce magneti per frigoriferi dal 2003. L'azienda dispone di tre catene di montaggio rispettivamente per la produzione di magneti di tipo *Mao*, di tipo *Panda* e di tipo *Pagoda*. Nei file `mao.txt`, `panda.txt` e `pagoda.txt` sono riportati (rispettivamente per i tre tipi di magneti) il numero di unità prodotte e il corrispondente costo totale di produzione relativi agli anni dal 2003 al 2012.

- a) Assumendo indipendenza e uguale variabilità tra i 30 costi di produzione e un modello economico per la produzione annuale del tipo $C_{totali} = C_{fissi} + C_{unitari} \cdot Quantita + \epsilon$, si stimino i costi fissi e unitari per le tre linee di produzione.
- b) Vi è evidenza statistica per affermare che i costi fissi delle tre linee di produzione siano diversi?
- c) Vi è evidenza statistica per affermare che i costi unitari delle tre linee di produzione siano diversi?
- d) Sulla base delle risposte alle domande (b) e (c) si fornisca un eventuale modello ridotto e si ristimino coerentemente i costi fissi e unitari per le tre linee di produzione.
- e) La Maog Net ha deciso di produrre nell'anno in corso 2500 magneti *Mao*, 4000 magneti *Panda* e 1000 magneti *Pagoda*. Si forniscano tre intervalli di previsione di livello globale 90% per i costi di produzione delle tre linee relativi all'anno in corso.