

Politecnico di Milano
Facoltà di Ingegneria dei Sistemi
APPELLO DI STATISTICA APPLICATA
24 Settembre 2008

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

Nome e cognome:

Numero di matricola:

Problema 1

L'Autorità dei Trasporti ha misurato il numero di passeggeri in partenza dai tre aeroporti gestiti dalla SEA (Linate, Malpensa ed Orio al Serio) in 30 giorni scelti caso dell'anno 2007 (dataset `flight.txt`). Assumendo che i dati siano un campione casuale da una normale trivariata:

- a) fornite 2 intervalli di confidenza globale 90% per la media e la varianza del “numero totale” di passeggeri quotidianamente in partenza dai tre aeroporti gestiti dalla SEA;
- b) verificate l'ipotesi secondo la quale ogni giorno dai tre aeroporti partono in media lo stesso numero di passeggeri;
- c) verificate l'ipotesi della SEA secondo la quale ogni giorno in media partono da Malpensa il doppio dei passeggeri sia di Linate che di Orio al Serio.

Problema 2

La Xair ha installato sui suoi aerei dei rilevatori ottici in grado di misurare la lunghezza e/o la larghezza dei bagagli da stiva imbarcati. Per ridurre il consumo di carburante è opportuno posizionare i bagagli di categoria H (peso maggiore di 15 kg) in prossimità delle ali mentre quelli di categoria L (peso inferiore a 15 kg) in testa o coda. Un campione casuale di 200 bagagli (per i quali sono riportate lunghezza, larghezza e categoria) è riportato nel file `luggage.txt`. Assumendo la lunghezza e la larghezza sia dei bagagli di categoria H che di quelli di categoria L variabili congiuntamente normali con stessa matrice di covarianza:

- a) secondo la regola di Bayes, si individuino tre classificatori dei bagagli che utilizzino come predittori rispettivamente:
 - solo la lunghezza,
 - solo la larghezza,
 - sia la lunghezza che la larghezza;
- b) mediante crossvalidazione si stimino gli AER dei tre classificatori individuati al punto (a).

Assumendo che i costi delle misurazioni di lunghezza e larghezza siano entrambi pari a 10€ per bagaglio ed il costo di un errato posizionamento di un bagaglio in stiva sia pari a 10€ per bagaglio per ogni 100 km di volo:

- c) scegliete quale classificatore è economicamente più vantaggioso utilizzarsi in un volo di 1000 km con 200 bagagli in stiva.

Problema 3

Tisséo, Atac e Atm hanno raccolto quotidianamente nel mese di maggio i dati relativi al numero di controlli ed al numero di multe effettuate sulle linee di superficie rispettivamente delle città di Tolosa, Roma e Milano (file `tolosa.txt`, `roma.txt`, `milano.txt`). I tre committenti vi chiedono di analizzare i dati raccolti; rispondete ai seguenti quesiti assumendo che i dati relativi a giorni differenti siano *iid* mentre i dati relativi allo stesso giorno ma a città differenti NON siano INDIPENDENTI:

- a) Vi è evidenza statistica di almeno una differenza tra le medie (variabili numero di controlli ed numero di multe) di Tolosa, Roma e Milano?
- b) Ricorrendo alla disuguaglianza di Bonferroni, fornite sei intervalli di confidenza globale 90% per le differenze (Tolosa *vs* Roma, Roma *vs* Milano, Milano *vs* Tolosa) delle medie (variabili numero di controlli e numero di multe).
- c) Redigete una relazione sintetica (massimo 10 righe) da fornire ai committenti mettendo in luce le assunzioni ed i risultati relativi all'analisi effettuata ai punti (a) e (b).

Problema 4

Nel file `time.txt` sono raccolti i tempi [min] di percorrenza di alcune tratte urbane di lunghezza [km] variabile, per diversi mezzi dell'Atm (autobus, filobus, tram) in diversi giorni (feriali, festivi).

- a) Ammettendo una dipendenza al più lineare tra la media del tempo di percorrenza e la lunghezza della tratta ed ignorando possibili interazioni tra il tipo di mezzo ed il tipo di giorno, introducete un modello (il più generale possibile) atto a descrivere i dati raccolti.
- b) Sulla base del modello (a) vi evidenza di un effetto del tipo di mezzo sulla media del tempo di percorrenza?
- c) Sulla base del modello (a) vi evidenza di un effetto del tipo di giorno sulla media del tempo di percorrenza?
- d) Sulla base del modello (a) vi evidenza di un effetto della lunghezza della tratta sulla media del tempo di percorrenza?
- e) Tra tutti i sottomodelli del modello (a) selezionate e commentate quello più adatto a descrivere i dati.
- f) Sulla base del modello (e), che tempo di percorrenza medio stimate per un filobus che percorre una tratta di 2 km in un giorno festivo?