

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

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

*Nome e cognome:*  
*Numero di matricola:*

### **Problema 1**

La compagnia aerea Speed Airways, vola quotidianamente sulla tratta aerea Madrid/Boston. Nel file `MB.txt` sono contenuti dati (durata effettiva del volo, ritardo all'arrivo e ritardo alla partenza espressi in ore) relativi ai 366 voli dell'anno 2008. L'associazione Happy Fliers vi chiede di valutare la qualità del servizio mediante una semplice analisi statistica.

- a) La Speed Airways dichiara che la durata media del volo sia di 6.5 ore, il ritardo medio all'arrivo nullo e ritardo medio alla partenza nullo. Vi è evidenza statistica per smentire tale affermazione?
- b) Fornite tre intervalli  $T^2$ -simultanei di confidenza globale 99% per le medie delle tre variabili riportate nel dataset e commentate il risultato.
- c) Per verificare la correttezza della procedura di acquisizione dei dati si esegua un test per la varianza di un'opportuna combinazione lineare delle variabili. Vi è evidenza statistica per affermare che la deviazione standard sia maggiore di 0.11 ore, ovvero che i dati non siano stati acquisiti correttamente?

## Problema 2

Nei distributori automatici di carburante della Exxon sono inseriti dei rivelatori ottici in grado di misurare le dimensioni delle banconote inserite. Sapendo che lo 0.1% delle banconote da 10 dollari in circolazione risultano contraffatte, la Exxon vorrebbe implementare un software in grado di identificare le banconote da 10 dollari false in modo da ridurre al minimo le perdite economiche. Assumendo che:

- sia la popolazione delle banconote vere che quella delle banconote false seguano una distribuzione normale (medie e matrici di covarianza diverse);
- accettare una banconota falsa porti un danno economico di 10 dollari;
- rifiutare una banconota vera porti un danno economico quantificabile in 5 centesimi di dollaro;

soddisfate le seguenti richieste della Exxon:

- a) costruite un opportuno classificatore, stimando i parametri sconosciuti a partire dai due dataset `true.txt` e `false.txt`, contenenti rispettivamente i dati relativi a 100 banconote vere ed a 100 banconote false (dati in mm); e riportate qualitativamente le due regioni di classificazione su di un grafico;
- b) calcolate l'APER del classificatore e partendo dall'APER stimate il danno economico atteso del classificatore.
- c) con che probabilità stimate venga rifiutata la prima banconota da 10 dollari che inserite nel distributore?

### Problema 3

L'editore Ted Singapore, sta monitorando il numero di errori di battitura (ogni 18000 battute) sui 30 giornali di sua proprietà. I dati relativi allo scorso anno sono riportati nel file `singapore.txt`. I giornali sono categorizzate per frequenza d'uscita (quotidiana, settimanale, mensile) e per target di riferimento (specializzato, generalista).

- a) Si implementi un modello ANOVA a due fattori senza interazione per stimare i parametri del modello.
- b) Sulla base del modello (a), vi evidenza statistica di una dipendenza della media del numero di errori dalla frequenza d'uscita?
- c) Sulla base del modello (a), vi evidenza statistica di una dipendenza della media del numero di errori dal target di riferimento?
- d) Sulla base dei test precedenti si costruisca un modello ridotto e se ne stimino i parametri.
- e) Soerentemente col modello ridotto (d), si costruiscano degli intervalli di confidenza globale 95% per le differenze delle medie dei gruppi individuati.
- f) Si riassuma in massimo tre righe il risultato delle analisi.

## Problema 4

Uno zoologo sta studiando l'evoluzione temporale dell'altezza al garrese di una nuova razza di capre scoperta sull'altipiano andino (dataset `goat.txt`). Utilizzando per l'individuo  $i$ -esimo un modello di crescita esponenziale del seguente tipo:  $h_i = A + B(1 - e^{-t_i}) + C\epsilon_i$ , con  $t_i$  l'età dell'individuo [anni],  $h_i$  l'altezza dell'individuo [cm] e  $\epsilon_i$  un termine casuale distribuito secondo una normale standard,  $A$  e  $B$  parametri dipendenti esclusivamente dal sesso dell'individuo,  $C$  un parametro uguale per tutta la popolazione.

- a) Si stimino col metodo dei minimi quadrati i 5 parametri del modello;
- b) sulla base del modello (a), vi è evidenza statistica che l'altezza media alla nascita ( $t = 0$ ) sia diversa per maschi e femmine?
- c) sulla base del modello (a), vi è evidenza statistica che l'altezza media in età adulta ( $t = +\infty$ ) sia diversa per maschi e femmine?
- d) sulla base dei test (b) e (c), si implementi un opportuno modello ridotto e se ne stimino i parametri;
- e) sulla base del modello ridotto (d), si forniscano intervalli di confidenza globale 90% per l'altezza media alla nascita e in età adulta degli individui maschi e femmine;
- f) sulla base del modello ridotto (d), vi è evidenza statistica per affermare che in età adulta gli individui maschi siano mediamente più alti delle femmine?