

**Politecnico di Milano**  
**Facoltà di Ingegneria dei Sistemi**  
APPELLO DI STATISTICA APPLICATA  
1 Luglio 2009

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

***Nome e cognome:***

***Numero di matricola:***

### **Problema 1**

All'interno dell'acceleratore di particelle del CERN di Ginevra sono stati fatti collidere due fasci di positroni; un rivelatore di particelle ha registrato la posizione (misurata in picometri nel sistema di riferimento cartesiano  $(X,Y,Z)$  del rivelatore) dei positroni circa 10 femptosecondi dopo l'urto (file `particles.txt`). Assumendo che le posizioni della singole particelle costituiscano variabili aleatorie indipendenti ed identicamente distribuite secondo la medesima legge normale trivariata:

- a) si effettui un test di livello 99% per verificare che l'urto sia avvenuto in corrispondenza dell'origine del sistema di riferimento del rivelatore (si identifichi il punto d'impatto con la media della normale trivariata);
- b) si identifichi la direzione di massima dispersione delle particelle;
- c) si quantifichi (in termini di distanza euclidea) la dispersione lungo la direzione individuata al punto (b) e la dispersione residua.

## Problema 2

Uno storico dell'arte richiede il vostro aiuto per individuare un criterio di classificazione per discriminare le sculture create da Gian Lorenzo Bernini da quelle di altri scultori suoi contemporanei, partendo dal peso [quintali] e dall'altezza [metri] di 100 sculture di attribuzione indubbia (file `sculptures.txt`). Tenendo conto che le sculture del Bernini costituiscono circa il 25% delle sculture che dovranno essere classificate e che lo scopo dello storico è quello di minimizzare il numero di misclassificazioni attese:

- a) si costruiscano due classificatori C1 e C2, assumendo rispettivamente i dati provenienti da due popolazioni normali di matrice di covarianza uguale per il classificatore C1, ed invece provenienti da due popolazioni normali di matrice di covarianza diversa per il classificatore C2;
- b) tramite crossvalidazione si stimino gli AER dei due classificatori e se ne commentino i valori;
- c) come verrà classificata dai due classificatori una statua alta 2 metri e pesante 4 quintali?

### Problema 3

Il Diagramma Veritatis è una misteriosa opera attribuita a Galileo. Alcuni semiologi ritengono che in alcune pagine del libro sia nascosto un messaggio in codice; inoltre ritengono che tali pagine siano caratterizzate da un'anomala numerosità di alcune lettere dell'alfabeto. Nel file `veritatis.txt` sono riportate, per le 132 pagine del libro, le frequenze assolute delle cinque vocali dell'alfabeto latino.

- a) Utilizzando un algoritmo di clustering agglomerativo (distanza di Manhattan, average linkage), si individuino due cluster e si riportino le pagine sospette;
- b) assumendo che, per ciascuno dei due cluster, le cinque frequenze assolute siano (approssimativamente) distribuite normalmente con stessa matrice di covarianza, si esegua un test per provare l'esistenza di una differenza nei valori medi delle due distribuzioni;
- c) utilizzando cinque intervalli di Bonferroni di livello globale 90%, si commenti il risultato del test implementato al punto (b).

## Problema 4

Nell'Index Librorum Prohibitorum (edizione del 1948), vengono elencate circa 10000 opere considerate eretiche dalla chiesa cattolica. Nel file '`index.txt`' sono riportati, per gli anni che vanno dal 1300 al 1899, il numero di opere aggiunte annualmente all'Index. La maggior parte degli storici ritiene che il numero medio di opere aggiunte annualmente sia decresciuto linearmente nel tempo nell'arco di questo periodo (modello A). Recentemente, il prof. Langdon ha proposto una teoria secondo la quale il trend lineare è "momentaneamente" cambiato (in modo discontinuo) durante il periodo di egemonia francese (dal 1768, Trattato di Versailles, al 1815, Battaglia di Waterloo, inclusi), durante il quale si è assistito ad un crollo delle opere aggiunte annualmente all'Index (modello B). Definita come  $\mu(t)$  il numero medio delle opere aggiunte all'Index in corrispondenza dell'anno  $t$  e formalizzati i due modelli come segue:

**Modello A:**  $\mu(t) = \alpha + \beta t$ ;

**Modello B:**  $\mu(t) = \alpha_1 + \beta_1 t$  per  $1768 \leq t \leq 1815$  e  $\mu(t) = \alpha_2 + \beta_2 t$  per  $t \leq 1767$  o  $t \geq 1816$ ;

si risponda ai seguenti quesiti:

- a) si stimino i parametri di entrambi i modelli utilizzando il metodo dei minimi quadrati; che assunzioni è necessario introdurre sulla componente di errore affinché le stime fornite dal metodo dei minimi quadrati non siano distorte?
- b) vi è evidenza statistica di un diverso trend lineare nel periodo di egemonia francese?
- c) utilizzando il modello B e la disuguaglianza di Bonferroni, si forniscano 2 intervalli di confidenza globale 90% per la media e la varianza del numero di opere inserite nell'indice nell'anno 1800;
- d) utilizzando il modello B, si fornisca un intervallo di confidenza 90% per la differenza tra numero medio di opere aggiunte all'Index nel 1816 ed numero medio di opere aggiunte nel 1815.