

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

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

*Nome e cognome:*

*Numero di matricola:*

**Problema 1**

Famosi dantisti ritengono che il numero di caratteri ( $N_C$ ) ed il numero di parole ( $N_P$ ) contenuti in un generico sonetto dantesco seguano approssimativamente una distribuzione congiuntamente normale di media  $\mu = \begin{pmatrix} 400 \\ 90 \end{pmatrix}$  e matrice di covarianza  $\Sigma = \begin{pmatrix} 100 & 20 \\ 20 & 10 \end{pmatrix}$ . Recentemente sono stati scoperti due nuovi sonetti attribuiti a Dante, provvisoriamente identificati coi codici 2008A e 2008B. Assumendo il numero di caratteri e di parole del sonetto 2008A indipendenti dal numero di caratteri e di parole del sonetto 2008B:

- a) identificate nel piano  $N_C \times N_P$  una regione ellissoidale nella quale il sonetto 2008A caschi con probabilità 0.9.
- b) Con che probabilità solo uno dei due sonetti casca nella regione individuata al punto (a)?
- c) Sapendo che il sonetto 2008A è costituito da 105 parole ed il sonetto 2008B da 95 parole, con che probabilità il sonetto 2008A avrà almeno 20 caratteri in più del sonetto 2008B?

## Problema 2

Nel dataset `eighteen.txt` sono riportati, per 100 comuni italiani, le frazioni percentuali di ragazzi minorenni e ragazze minorenni relative all'intera popolazione residente nel comune. Supponendo che questo sia un campione casuale bivariato estratto da una distribuzione normale:

- a) si esegua un test per verificare che in Italia risiedono lo stesso numero di ragazzi minorenni e di ragazze minorenni;
- b) sapendo che in Italia risiedono 60 milioni di abitanti, si riportino tre intervalli  $T^2$ -simultanei (di confidenza globale 90%) per:
  - il numero assoluto di ragazzi minorenni che risiedono in Italia,
  - il numero assoluto di ragazze minorenni che risiedono in Italia,
  - il numero assoluto di minorenni che risiedono in Italia.

### Problema 3

La West Sussex Bread Association ha selezionato casualmente 60 attività commerciali nelle quali vengono abitualmente vendute ciambelle. 30 attività hanno sede nella città di Brighton e 30 nella città di Worthing. Per ognuna delle due città, in 10 attività è stato registrato il prezzo di una ciambella liscia, in 10 il prezzo di una ciambella ripiena di crema ed in 10 il prezzo di una ciambella ripiena di marmellata. I dati raccolti sono riportati nel dataset `doughnut.txt`.

- a) Descrivete il modello ANOVA che ritenete opportuno per l'analisi di questi dati.
- b) Identificando i fattori che influenzano significativamente la distribuzione del prezzo delle ciambelle, individuate un eventuale modello ridotto.
- c) Utilizzando la disuguaglianza di Bonferroni stimate tramite intervalli confidenza bilateri (con confidenza globale 95%) medie e varianze delle sottopopolazioni associate al modello ridotto individuato al punto (b).

## Problema 4

A causa dell'aumento del costo del petrolio, la compagnia aerea FlyDown è interessata ad individuare un modello che permetta di stimare il peso di un passeggero (dato tipicamente non disponibile) a partire dalla sua età e dal suo sesso (dati sempre disponibili). Per questo motivo, la FlyDown ha raccolto (dataset `flydown.txt`) tramite un questionario anonimo dati riguardanti il peso di 126 clienti (63 uomini e 63 donne) di età compresa tra i 18 e gli 80 anni.

- a) Introducete un modello di regressione quadratico nell'età in grado di descrivere un'eventuale dipendenza del peso atteso di un passeggero dall'età e dal sesso.
- b) Vi è evidenza statistica di una dipendenza del peso atteso dal sesso?
- c) Vi è evidenza statistica di una dipendenza (lineare o quadratica) del peso atteso dall'età?
- d) Individuate un modello ridotto del modello (a) adatto a descrivere i dati raccolti e stimatene i parametri.
- e) Sulla base del modello individuato al punto (d), vi è evidenza statistica per smentire l'ipotesi secondo la quale il massimo del peso atteso (per uomini e/o donne) viene raggiunto in corrispondenza dei 50 anni?
- f) Individuate un modello ridotto del modello (d) che tenga in considerazione quanto affermato al punto (e) e stimatene i parametri.