

**Politecnico di Milano**  
**Facoltà di Ingegneria dei Sistemi**  
APPELLO DI STATISTICA APPLICATA  
18 Febbraio 2009

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

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

### Problema 1

La nave Polar Endeavour, sta trasmettendo a ciclo continuo un segnale di SOS utilizzando contemporaneamente tre diverse frequenze: 3 MHz, 30 MHz e 300 MHz. A causa della variabilità delle situazioni ambientali, le distanze massime alle quali i tre segnali possono essere captati possono essere modellizzati come quantità aleatorie congiuntamente normali di media [km] e matrice di covarianza [km<sup>2</sup>] rispettivamente:

$$\begin{pmatrix} 40 \\ 30 \\ 20 \end{pmatrix} \text{ e } \begin{pmatrix} 4 & 2 & 2 \\ 2 & 4 & 2 \\ 2 & 2 & 4 \end{pmatrix} .$$

Sapendo che in questo momento il segnale trasmesso a 3 MHz può essere captato fino a 50 km di distanza e che la nave Discovery si trova a 20 km dalla Polar Endeavour:

- a) definite una regione ellissoidale di probabilità 90% per la distanza percorsa dai segnali a 30 e 300 MHz;
- b) potete affermare che con probabilità almeno 90% tutti e tre i segnali sono captati dalla Discovery?

## Problema 2

Venerdì 17 ottobre 2008, nei cieli di Black Fortune è avvenuto uno scontro tra due satelliti artificiali. Circa un centinaio di detriti sono stati rivenuti al suolo (file `satellite.txt`); a causa dell'attrito con l'atmosfera non è stato però possibile definire la provenienza di nessuno dei detriti. Per definire la dinamica dell'incidente, l'aeronautica militare statunitense vi chiede di stimare la posizione relativa dei due punti di impatto al suolo.

Supponendo che i detriti relativi allo stesso satellite si siano dispersi al suolo secondo una legge normale di media e matrice di covarianza incognite:

- a) si attribuiscono i detriti ai due satelliti tramite un algoritmo di clustering gerarchico agglomerativo che utilizzi la distanza euclidea e l'average linkage;
- b) si riportino le numerosità dei due cluster ed il valore del coefficiente cofenetico;
- c) introducendo eventuali ipotesi, identificando i punti di impatto al suolo con le medie delle due normali ed assumendo corretta l'attribuzione dei detriti ai due satelliti, si fornisca una regione ellittica di confidenza globale 99% per la posizione relativa dei due punti di impatto al suolo.

### Problema 3

Durante l'estate australe nidificano nella penisola antartica tre specie di pinguini: Chinstrap, Adelia e Gentoo. Alcuni biologi della base Artowski hanno misurato il peso [kg] di 90 esemplari adulti: 15 maschi e 15 femmine per ciascuna delle 3 specie (file `penguins.txt`).

- a) Utilizzando un modello ANOVA a due fattori, si affermi se sesso e/o specie di appartenenza influenzano significativamente il peso.
- b) Utilizzando un'opportuno modello (eventualmente ridotto), si forniscano le stime (intervalli di confidenza globale 90%) delle medie e delle varianze dei gruppi individuati al punto (a).

## Problema 4

Nel file `PIL.txt` è riportato l'incremento percentuale annuo del prodotto interno lordo per Germania, Italia e Polonia relativo al periodo 2004-2009. Si assumano le diverse percentuali come realizzazioni di variabili aleatorie normali ed indipendenti di pari varianza e media invece dipendente dallo stato  $k \in \{\text{Germania, Italia, Polonia}\}$  e dagli anni  $\Delta t$  decorsi dal 2005 (0 per il 2005, 1 per il 2006 e così via) come segue:  $E(PIL) = \beta_{0k} + \beta_{1k}\Delta t + \beta_{2k}(\Delta t)^2$ .

- a) Fornite la stima ai minimi quadrati dei parametri che definiscono media e varianza del modello;
- b) utilizzando il modello (a) eseguite un test al 5% per la significatività del termine quadratico  $(\Delta t)^2$ ;
- c) utilizzando il modello (a) eseguite un test al 5% per la significatività del fattore  $k$ ;
- d) utilizzando il modello (a) eseguite un test al 5% per verificare l'ipotesi secondo cui il valore atteso del PIL della Germania coincide con quello dell'Italia;
- e) sulla base dei risultati dei tre test eseguiti ai punti (b), (c) e (d) individuate un modello ridotto per descrivere i dati e sulla base di questo fornite le nuove stime dei parametri;
- f) si scriva un report (di massimo 5 righe) che riassume i risultati più importanti dell'analisi.