

**Politecnico di Milano**  
**II Scuola - Ingegneria dei Sistemi (MI)**  
APPELLO DI STATISTICA APPLICATA  
29 giugno 2012

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

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

### Problema 1

Il sindaco delle Isole Lofoten utilizza, per la previsione delle produzioni **trimestrali** di aringhe [migliaia di tonnellate], un modello normale multivariato  $(X_I, X_{II}, X_{III}, X_{IV})$  con media e matrice di covarianza rispettivamente:

$$\begin{pmatrix} 10 \\ 30 \\ 50 \\ 10 \end{pmatrix} \text{ e } \begin{pmatrix} 1 & 0.5 & 0.25 & 0.125 \\ 0.5 & 1 & 0.5 & 0.25 \\ 0.25 & 0.5 & 1 & 0.5 \\ 0.125 & 0.25 & 0.5 & 1 \end{pmatrix}.$$

- a) Si costruisca sulla base del modello un intervallo di previsione (probabilità 90%) per la produzione **annuale** del 2012;
- b) si costruisca sulla base del modello un intervallo di previsione (probabilità 90%) per la produzione relativa al secondo **semestre** del 2012.

Pochi giorni fa è stata resa nota la produzione relativa al primo **semestre** del 2012, pari a 50 migliaia di tonnellate. Sulla base di questa nuova informazione:

- c) si riaggiorni l'intervallo di previsione per la produzione **annuale** del 2012;
- d) si riaggiorni l'intervallo di previsione per la produzione relativa al secondo **semestre** del 2012.

## Problema 2

Nel 1962 sono stati scoperti nel Mare del Nord degli enormi giacimenti di petrolio. Nei file `fifties.txt` e `seventies.txt` sono riportati rispettivamente per il decennio 1950-1959 e il decennio 1970-79 i redditi mensili medi [NOK2012] nelle 307 municipalità norvegesi. Assumendo l'indipendenza tra i dati relativi ad anni diversi e introducendo opportune ipotesi di normalità e omoschedasticità, per ciascuna municipalità si esegua un test unilatero per provare l'esistenza di un incremento dei salari medi.

- a) Si riportino a titolo indicativo le municipalità associate al p-value più basso e al p-value più alto.
- b) Quali municipalità hanno un p-value minore o uguale all'1%?
- c) Imponendo col metodo di Bonferroni un Familywise Error Rate minore o uguale all'1%, si individuino le municipalità dove vi è evidenza statistica di un incremento dei salari medi.
- c) Imponendo col metodo di Benjamini-Hochberg un False Discovery Rate minore o uguale all'1%, si individuino le municipalità dove vi è evidenza statistica di un incremento dei salari medi.

### Problema 3

Un giovane statistico di Tromsø vuole costruire un classificatore in grado di prevedere la presenza o assenza di precipitazioni nevose nel giorno  $G+1$  sulla base della temperatura [ $^{\circ}\text{C}$ ] e dell'umidità medie del giorno  $G$ . Utilizzando i dati relativi agli ultimi 30 giorni (file `snow.txt`) e sapendo che negli ultimi 40 anni a Tromsø è nevicato in media 207 giorni all'anno:

- a) Si costruisca un classificatore [Si riportino in particolare le assunzioni modellistiche e si fornisca un grafico qualitativo delle regioni di classificazione].
- b) Si stimi l'APER del classificatore (a) e lo si confronti con quello del classificatore banale.
- c) La temperatura e l'umidità misurate ieri a Tromsø sono rispettivamente pari a  $-10^{\circ}\text{C}$  e 0.75. Si stimi la probabilità di neve nella giornata odierna utilizzando sia il classificatore (a) che il classificatore banale.

## Problema 4

In Norvegia volano due compagnie aeree low-cost: la Widerøe e la Norwegian Airlines. Un'azienda norvegese vuole minimizzare il tempo speso in volo dai suoi dipendenti scegliendo di volta in volta a quale compagnia appoggiarsi sulla base della lunghezza della tratta. Nei file `wideroe.txt` e `norwegian.txt` sono riportate per le 17 tratte operate da entrambe le compagnie i tempi di volo relativi ai voli degli ultimi 5 giorni. Assumendo indipendenti le 170 misure raccolte e assumendo per ciascuna compagnia una dipendenza lineare tra la durata media del volo [min] e la lunghezza della tratta [km]:

- a) Si stimino le intercette e i coefficienti angolari delle due rette di regressione.
- b) Vi è evidenza statistica di una diversa dipendenza lineare nelle due compagnie?
- c) Sulla base del precedente test si proponga un eventuale modello ridotto.
- d) Sulla base del modello (c) si dica per quali voli risulta più conveniente appoggiarsi alla Widerøe e per quali alla Norwegian.
- e) Sulla base del modello (c) si dica che compagnia utilizzare per un eventuale tratta 950 km e si fornisca congiuntamente una stima intervallare di confidenza 95% per il risparmio medio di tempo rispetto all'altra compagnia.