

Politecnico di Milano
Facoltà di Ingegneria dei Sistemi
V APPELLO DI STATISTICA APPLICATA
27 Settembre 2007

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

Nome e cognome:

Numero di matricola:

Problema 1

L'Assessorato al Turismo di Barcellona vuole lanciare una campagna pubblicitaria per convincere i milanesi a spendere le loro ferie a Barcellona. Le condizioni climatiche (medie annuali) relative alle due città descritte tramite temperatura ($^{\circ}C$), umidità relativa (valore percentuale) e condizioni ventose (*nodi*) degli ultimi 50 anni sono riportate nei file `barcellona.txt` e `milano.txt`.

- a) Assumendo indipendenti le condizioni climatiche delle due città, si esegua un test per provare che le medie sono diverse per le due città. Si accompagni il risultato del test con tre intervalli T^2 -simultanei (livello di confidenza globale 90%) per la differenza delle medie delle tre variabili.
- b) Si ripeta quanto richiesto al punto (a) ammettendo una possibile dipendenza tra le condizioni climatiche delle due città relative allo stesso anno.
- c) Si confrontino le assunzioni necessarie allo sviluppo dei punti (a) e (b).

Problema 2

Il preside di un piccolo asilo svizzero vorrebbe comprare banchi di dimensione diversa per gli alunni di prima, seconda e terza e/o per maschi e femmine. Per far ciò viene misurata l'altezza di tutti gli alunni dell'asilo (file `asilo.txt`).

- a) Introducendo un opportuno modello statistico, si aiuti il preside a capire quanti tipi di banchi comprare.
- b) Si forniscano al preside degli intervalli di Bonferroni di confidenza globale 90% per la differenza delle medie delle altezze dei gruppi individuati al punto (a).

Problema 3

Nel file `galaxy.txt` sono riportate le coordinate celesti (latitudine e longitudine) di 116 astri presenti nel settore A12 della volta celeste. Il noto astronomo O'Repek vi chiede aiuto per individuare e caratterizzare una galassia che ci si aspetta essere presente in tale settore.

- a) Noto che le galassie compaiono nella volta celeste come ammassi ellissoidali di stelle molto ravvicinate, si utilizzi un opportuno algoritmo di clustering agglomerativo per individuare la galassia. (Si utilizzi la distanza euclidea e il Ward linkage)
- b) Si individui una regione ellissoidale di confidenza 99% per il centro della galassia.
- c) Assumendo che le stelle di tale galassia si distribuiscano sulla volta celeste secondo una legge normale bivariata, si utilizzi la stima di massima verosimiglianza della distribuzione per individuare una regione ellissoidale dove si stima cadano il 99% degli astri della galassia.

Problema 4

Presso il centro di ricerca SmokySmoke si vuole stimare il danno ambientale ($\text{€}/\text{km}$) prodotto dalle automobili circolanti nel comune di Milano. Scopo dello studio è proporre al comune l'istituzione di una pollution charge basata sulla cilindrata (cm^3) e sul tipo di carburante (diesel o benzina). 30 automobili diesel (file **diesel.txt**) e 30 automobili benzina (**benzina.txt**) di diversa cilindrata e circolanti nella sola area urbana del comune di Milano, sono state monitorate per un anno.

- a) Sulla base di un opportuno modello statistico si costruisca uno stimatore puntuale del danno prodotto da un autoveicolo condizionatamente alla cilindrata ed al tipo di carburante dell'autoveicolo. (Si giustifichi ovviamente la scelta del modello introducendo se necessario opportune ipotesi)
- b) Vi è evidenza statistica del fatto che il modello individuato spieghi il danno ambientale prodotto dagli autoveicoli?
- c) Si stimino media e varianza del danno prodotto da un diesel di 1200 cm^3 e del danno prodotto da un benzina di 1000 cm^3 .
- d) Si costruisca un intervallo di confidenza 90% per la differenza delle medie stimate al punto (c).