

Politecnico di Milano
II Scuola - Ingegneria dei Sistemi (MI)
APPELLO DI STATISTICA APPLICATA
5 settembre 2013

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

Nome e cognome:

Numero di matricola:

Problema 1

Nel file `tennis.txt` sono riportate le velocità in battuta [km/h] relative all'ultima partita del tennista S. Peed. Nel file è anche indicato se al momento della battuta S. Peed fosse in vantaggio/parità/svantaggio e se stesse servendo controsole o col sole alle spalle.

- a) Si costruisca un opportuno modello ANOVA a due fattori senza interazione per la velocità in battuta e si riportino i valori dei parametri.
- b) Sulla base del modello precedentemente, si eseguano due test per l'influenza dei due fattori, si proponga un eventuale modello ridotto e si riportino i valori dei parametri.
- c) Si fornisca un intervallo di confidenza 95% per la velocità media in vantaggio e controsole sulla base del modello (b).

Problema 2

La Fifa sta monitorando da alcuni mesi la durata delle partite dell'arbitro inglese Tim Consumer. Nel file `match.txt` sono riportate le durate [min] del primo e del secondo tempo delle ultime 20 partite da lui arbitrate. Gli standard Fifa richiedono una durata media del primo tempo di 46 minuti e una durata media del secondo tempo di 48 minuti.

- a) Con un livello di significatività del 5% si può provare che Tim Consumer si stia discostando dagli standard Fifa?
- b) Si forniscano e commentino tre intervalli T^2 -simultanei di confidenza globale 95% la durata media dei primi tempi, dei secondi tempi e delle partite intere arbitrate da Tim Consumer.

Problema 3

Nel file `hotel.txt` sono riportati gli incassi per camera [€] relativi ai 130 alberghi della catena Holisleep Inn nelle quattro stagioni relative all'ultimo anno.

- a) Si esegua un'analisi delle componenti principali; in particolare si individuino le componenti principali più importanti e se ne riporti la frazione di varianza totale spiegata;
- b) si interpretino le componenti principali selezionate.

Problema 4

La Polocoli ha postato il primo Agosto sulla propria pagina facebook due foto relative ad un nuovo modello di occhiali. Le due foto sono identiche a meno del colore della montatura: rossa in un caso e nera nell'altro. Nei file `red.txt` e `black.txt` sono riportati il numero di *like* giornalieri ricevuti dalle due foto nel mese di Agosto. Assumendo un modello per il numero medio giornaliero di *like* $\mu(t) = A + B/t$, con A e B due costanti che dipendono dal colore degli occhiali e t il tempo [giorni] trascorso dal primo agosto ($t = 1$), ed un errore additivo a media nulla e a varianza costante:

- a) si stimino col metodo dei minimi quadrati i 5 parametri del modello;
- b) si esegua un test per confermare o smentire l'ipotesi secondo la quale appena messe online ($t = 1$) le due foto abbiano ricevuto lo stesso apprezzamento medio;
- c) si esegua un test per confermare o smentire l'ipotesi secondo la quale finito il periodo di transizione ($t = +\infty$) le due foto riceveranno lo stesso apprezzamento medio;
- d) sulla base dei due precedenti test si proponga un modello ridotto e se ne stimino i parametri;
- e) sulla base del modello (d) si forniscano due intervalli di previsione di livello globale 90% per il numero di *like* che verranno ricevuti dalle due foto il 5 settembre ($t=36$).