FACOLTÀ DI SOCIOLOGIA - A.A. 2005-2006
ESAME DI ANALISI MULTIVARIATA
Appello del 15.02.2006
ULTIMO RECUPERO
<u>Avvertenza</u>: Fornire le formule utilizzate e tutti i passaggi dei calcoli eseguiti.
Utilizzare almeno 2 cifre decimali.

Esercizio 1:
a) Un'indagine sul livello di preparazione degli studenti lavoratori iscritti nelle facoltà scientifiche di un'università italiana ha riscontrato che l'85% di essi supera un esame in materie statistiche al primo tentativo. Fornire la probabilità che, estratti casualmente 10 studenti lavoratori, fra questi almeno 8 superano un esame in materie statistiche al primo tentativo.
b) La stessa indagine ha stabilito che il tempo X (espresso in giorni) necessario agli studenti lavoratori per la preparazione dell'esame di statistica è ben interpretato da una variabile casuale Normale con media μ =30 e scarto quadratico medio σ =6. Scelto casualmente uno studente lavoratore che ha preparato l'esame di statistica, si stabilisca la probabilità che abbia dedicato alla preparazione dell'esame tra i 33 ed i 36 giorni.
c) Dalla popolazione di studenti di lavoratori di un'altra università si estrae un campione casuale di ampiezza 4 al fine di stimare il tempo medio μ di preparazione dell'esame di statistica. A tale scopo si considerano i seguenti stimatori:

$$T_1 = \frac{X_1 + X_2}{2} \qquad \text{e} \qquad T_2 = \frac{X_1 + X_2 + X_3}{3}.$$

(i) Verificare se i due stimatori sono non distorti per μ ed eventualmente quantificarne la distorsione.
(ii) Stabilire, specificando il criterio e motivando la risposta, quale fra i due stimatori è preferibile.

Esercizio 2:
a) Da un'indagine sulla conflittualità condotta su un campione casuale di 50 lavoratori dipendenti nell'industria, risulta che, nell'ultimo anno, 30 di essi hanno effettuato almeno un'ora di sciopero. Costruire un intervallo di confidenza al 95% per la percentuale dei lavoratori industriali che hanno scioperato nell'ultimo anno.
b) Quale deve essere la numerosità campionaria minima per avere un intervallo di confidenza a livello 90% e con ampiezza non superiore a 0,3?
c) Dare il concetto di intervallo di confidenza approssimato (per grandi campioni) e discuterne l'applicazione ai casi reali.

Esercizio 3:
a) Si supponga che il tempo X di attesa in coda (regolata attraverso l'"elimina-code" elettronico) agli sportelli per il pagamento di bollettini di c/c presso gli uffici postali italiani, sia interpretabile da una variabile casuale Normale con media e varianza ignote. Si è chiesto a 30 utenti, scelti casualmente in un ufficio postale nel centro di Milano, quanti minuti hanno atteso in coda prima di effettuare un pagamento di bollettini c/c. Ne è risultata un'attesa media di 30 minuti e una varianza campionaria corretta $\bar{s}^2 = 25$. Fissato $\alpha = 0{,}10$, verificare l'ipotesi secondo cui il numero medio di minuti di attesa allo sportello sia pari a 27 e commentare il risultato.
b) Agli stessi 30 utenti viene chiesta una valutazione Y degli uffici postali in base al sistema di regolazione delle code ed in base alla loro percezione X del tempo medio di attesa superiore od inferiore a 30 minuti. I risultati sono riassunti nella seguente tabella a doppia entrata:

X \ Y	Una fila per ciascuno sportello	Fila unica per tutti gli sportelli	"Elimina-code" elettronico	
Più di 30 minuti	6	6	8	20
Meno di 30 minuti	4	4	2	10
	10	10	10	30

Verificare a livello α = 0,05, se esiste indipendenza tra X e Y utilizzando un opportuno test statistico e commentando il risultato.
c) Esporre il concetto di p-value e discuterne i valori per la valutazione di un test a una o a due code.

FACOLTÀ DI SOCIOLOGIA - A.A. 2005-2006
ESAME DI ANALISI MULTIVARIATA
Soluzioni appello del 15.02.2006[1]

ULTIMO RECUPERO

Esercizio 1.
a) Per trovare la probabilità richiesta bisogna utilizzare la formula

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x},$$

con

$p = 0{,}85$,
$1 - p = 0{,}15$ e
$n = 10$,

e calcolare la seguente probabilità

$$P(X \geqslant 8) = P(X = 8) + P(X = 9) + P(X = 10) = 0{,}28 + 0{,}35 + 0{,}20 = 0{,}83$$

b)
Si deve trovare la seguente probabilità per una variabile casuale normale *X*:

$$P(33 < X < 36) = [\,\mathit{standardizzando}\,] = P \quad \frac{33 - 30}{6} < \frac{X - 30}{6} < \frac{36 - 30}{6} \quad = P(0{,}5 < Z < 1) =$$
$$= P(Z < 1) - P(Z < 0{,}5) = 0{,}8413 - 0{,}6915 = 0{,}1498$$

c)
(i) La media dello stimatore T_1 è data da:

$E(T_1) = E\left(\frac{X_1 + X_2}{2}\right) = \frac{1}{2}\left(E(X_1) + E(X_2)\right) = \frac{1}{2}(2\mu) = \mu$. Lo stimatore è corretto.

La media dello stimatore T_2 è data da:

$E(T_2) = E\left(\frac{X_1 + X_2 + X_2}{3}\right) = \frac{1}{3}\left(E(X_1) + E(X_2) + E(X_3)\right) = \frac{1}{3}(3\mu) = \mu$. Lo stimatore è corretto

(ii) Le varianze sono date da:

$$Var(T_1) = Var\left(\frac{X_1 + X_2}{2}\right) = \frac{1}{4}\left(Var(X_1) + Var(X_2)\right) = \frac{1}{4}\left(2\sigma^2\right) = \frac{\sigma^2}{2}$$
$$Var(T_2) = Var\left(\frac{X_1 + X_2 + X_3}{3}\right) = \frac{1}{9}\left(Var(X_1) + Var(X_2) + Var(X_3)\right) = \frac{1}{9}\left(3\sigma^2\right) = \frac{\sigma^2}{3},$$

e si sceglie il secondo stimatore perché è più efficiente.

[1] A cura di Giancarlo Manzi

Esercizio 2.
a) La proporzione campionaria è:
$p = \frac{30}{50} = 0,6$.
Il livello di confidenza è:
$\alpha = 0,05$.
Quindi, poiché la numerosità campionaria è sufficientemente elevata, possiamo costruire l'intervallo di confidenza asintotico all'$(1 - \alpha)\times 100$ per la percentuale di lavoratori che hanno scioperato nella popolazione:
$P\left(p - z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}} \le P \le p + z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}} \right) = 0,95$. L'intervallo di confidenza sarà:
$(0,6 - z_{0,975}\sqrt{\frac{0,6\times 0,4}{50}};0,6 + z_{0,975}\sqrt{\frac{0,6\times 0,4}{50}})$, e quindi:
$(0,6 - 1,96\sqrt{\frac{0,6\times 0,4}{50}};0,6 + 1,96\sqrt{\frac{0,6\times 0,4}{50}})$.
Svolgendo i calcoli l'intervallo risulta:
$(0,464207;0,735792)$.

b) Deve essere:
$2z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}} \leqslant 0,30$, cioè:
$2\times 1,645\sqrt{\frac{0,6(0,4)}{n}} \leqslant 0,30$, ed elevando al quadrato ambo i membri:
$4\times 2,706\frac{0,24}{n} \leqslant 0,09$, ed infine esplicitando *n*:
$n \geqslant \frac{10,8241\times 0,24}{0,09} = 28,86 \approx 29$.
c) Si vedano appunti e/o libri di testo.

Esercizio 3.
a) Si ha:
n=30;
$\bar{x} = 30$;
$\bar{s}^2 = 25$;
$\mu_0 = 27$;
$\alpha = 0,10$.
Si deve utilizzare la *statistica test*, visto che la varianza è ignota:
$\frac{\bar{X} - \mu_0}{s/\sqrt{n}} \approx T_{29}$, che assume nel campione il valore:
$t = \frac{30 - 27}{5/\sqrt{30}} = \frac{3}{0,9129} = 3,2863$.
Poiché
$t_{0,95;29} = 1,699 < 3,2863$ si rifiuta l'ipotesi nulla: $H_0 : \mu = 27$.

b) Si deve utilizzare il test d'indipendenza secondo cui si rifiuta l'ipotesi nulla H_0 di indipendenza se $\chi^2 > \chi^2_{1-\alpha;(h-1)(k-1)}$

Il valore empirico del chi quadro è dato da:

$$\chi^2 = 30\left(\frac{6^2}{20\times 10}+\frac{6^2}{20\times 10}+\frac{8^2}{20\times 10}+\frac{4^2}{10\times 10}+\frac{4^2}{10\times 10}+\frac{2^2}{10\times 10}-1\right) = 30(0{,}36+0{,}32+0{,}32+0{,}04-1) = 1{,}2\ .$$

Poiché $\chi^2_{0,95;2} = \mathit{5.99146}$ che è maggiore di 1,2, si accetta l'ipotesi nulla di indipendenza tra le due variabili.

c) Si vedano appunti e/o libri di testo.