FACOLTÀ DI SOCIOLOGIA - A.A. 2005-2006
ESAME DI ANALISI MULTIVARIATA
PRIMO RECUPERO
18.05.2006

Avvertenza: Fornire le formule utilizzate e tutti i passaggi dei calcoli eseguiti.
Utilizzare almeno 2 cifre decimali.

Esercizio 1:
a) Secondo i dati storici forniti da un'azienda di trasporto pubblico (metropolitana e autobus di superficie) su 100 persone controllate, mediamente 20 non risultano essere in possesso del titolo di viaggio. Calcolare la probabilità che un controllore trovi meno di 2 persone senza biglietto su un totale di 10 controllate su un autobus.
b) Sempre in tema di trasporti pubblici, uno studio ha stabilito che il tempo X (espresso in minuti) di attesa alla fermata di una linea urbana è ben interpretato da una variabile casuale normale con media (in minuti) $\mu=12$ e varianza $\sigma^2=144$. Si stabilisca la probabilità che il tempo di attesa ad una fermata sia inferiore a 5 minuti.
c) Da una popolazione casuale nella quale il fenomeno X ha media μ si estrae un campione casuale semplice di numerosità n. Al fine di stimare μ si utilizzano i seguenti stimatori:

(i) $T_1 = \frac{n+1}{n^2-1}\sum_{i=1}^{n} X_i$.

(ii) $T_2 = \frac{n-1}{n^2-1}\sum_{i=1}^{n} X_i$.

Stabilire se i due stimatori sono non distorti ed in caso non lo siano fornire un'adeguata procedura di correzione.

Esercizio 2:
a) Da un'indagine d'opinione sulla privatizzazione di un'importante azienda statale di telecomunicazioni condotta su un campione casuale di 50 individui, risulta che il 20% di essi giudica complessivamente in modo negativo il processo di privatizzazione. Costruire un intervallo di confidenza al 99% per la percentuale degli individui che giudicano negativa la privatizzazione.
b) Quale deve essere la numerosità campionaria minima per avere un intervallo di confidenza a livello 95% e con ampiezza non superiore a 0,3?
c) Esporre la metodologia per la costruzione di un intervallo di confidenza per la media, definendo e discutendo il livello di confidenza.

Esercizio 3:
a) Dal sito della provincia di Milano risulta che per legge il valore limite massimo delle sorgenti sonore è fissato a 50 decibel nelle zone residenziali durante l'orario diurno (dalle 6:00 alle 22:00). Si supponga che il fenomeno "rumore durante il giorno" possa essere ben interpretato da una v.c. Normale con media e varianza ignote. Si supponga che in una particolare zona siano stati effettuate 40 misurazioni che hanno prodotto una media pari a 54 decibel ed una varianza pari a 1,2. Impostare un adeguato test, fissato $\alpha = 0{,}10$, verificando l'ipotesi nulla per la quale la zona abbia un valore pari al limite massimo consentito dalla legge.
b) 40 bambini sono stati sottoposti a test di alimentazione ed in particolare è stata registrata l'età X (in giorni) allo svezzamento ed il peso alla nascita Y. I risultati sono riassunti nella seguente tabella a doppia entrata:

Y / X	Minore di 350 grammi	Tra 350 e 450 grammi	Oltre 450 grammi	
Più 180 giorni	16	4	0	20
Meno di 180 giorni	5	6	9	20
	21	10	9	40

Verificare a livello $\alpha = 0{,}05$, se esiste indipendenza tra X e Y utilizzando un opportuno test statistico e commentando il risultato.
c) Dare una definizione degli errori di prima e di seconda specie e discuterne il significato.

FACOLTÀ DI SOCIOLOGIA - A.A. 2005-2006
ESAME DI ANALISI MULTIVARIATA
Soluzioni appello del 18.05.2006[1]

Esercizio 1.

a) Nella singola prova che consiste nel controllo casuale di una persona a bordo di un autobus, si possono verificare i due eventi "persona munita di biglietto" oppure "persona non munita di biglietto". La legge binomiale la cui funzione di probabilità è

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

ben si presta ad interpretare il fenomeno. Si è interessati all'evento "meno di 2 persone non hanno il biglietto", cioè "o nessuna persona non non ha il biglietto, oppure 1 persona non ha". I parametri sono:

$p = \frac{20}{100} = 0{,}20$ $(1 - p = 0{,}80)$, $n = 10$ e si deve valutare la legge binomiale per X=0, X=1.

Quindi:

$$P(X = 0) + P(X = 1) = \binom{10}{0} 0{,}80^{10} + \binom{10}{1} 0{,}20 \bullet 0{,}80^9 = .$$
$$= 0{,}1073741824 + 0{,}268435456 = 0{,}376$$

b)
Evidentemente il quesito si risolve esplicitando la seguente probabilità per una variabile casuale normale X:

$$P(X < 5) = P\left(Z < \frac{5 - 12}{\sqrt{144}}\right) = P\left(Z < -\frac{7}{12}\right) = 1 - P\left(Z < \frac{7}{12}\right) = 1 - P(Z < 0{,}58) = 1 - 0{,}719 = 0{,}281$$

c)
(i) In generale, lo stimatore T si dice non distorto per il parametro θ se $E[T] = \theta$. In questo caso si ha:

$$E[T_1] = E\left[\frac{n+1}{n^2 - 1} \sum_{i=1}^{n} X_i\right] = \frac{n+1}{n^2 - 1} \sum_{i=1}^{n} E[X_i] = [i.i.d] = \frac{n+1}{n^2 - 1} \sum_{i=1}^{n} E[X] = \frac{n+1}{n^2 - 1} n\mu = \frac{n}{n-1} \mu \neq \mu .$$
$$E[T_2] = E\left[\frac{n-1}{n^2 - 1} \sum_{i=1}^{n} X_i\right] = \frac{n-1}{n^2 - 1} \sum_{i=1}^{n} E[X_i] = [i.i.d] = \frac{n-1}{n^2 - 1} \sum_{i=1}^{n} E[X] = \frac{n-1}{n^2 - 1} n\mu = \frac{n}{n+1} \mu \neq \mu .$$

Gli stimatori $\tilde{T}_1 = \frac{n-1}{n} T_1$ e $\tilde{T}_2 = \frac{n+1}{n} T_2$ sono non distorti.

Esercizio 2.

a) La proporzione campionaria è pari a:
$p = 0{,}2$.
Il livello di confidenza è:
$\alpha = 0{,}01$.
La numerosità campionaria è sufficientemente elevata; possiamo allora costruire l'intervallo di confidenza asintotico all'$(1 - \alpha) \times 100$ per la percentuale di individui contrari alla privatizzazione:

$P\left(p - z_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}} \leq P \leq p + z_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}\right) = 0{,}99$. L'intervallo di confidenza sarà dunque:

[1] A cura di Giancarlo Manzi

$\left(0,2 - z_{0,995}\sqrt{\frac{0,2\times 0,8}{50}};0,2 + z_{0,995}\sqrt{\frac{0,2\times 0,8}{50}}\right)$, e quindi:

$\left(0,2 - 2,575\sqrt{\frac{0,2\times 0,8}{50}};0,2 + 2,575\sqrt{\frac{0,2\times 0,8}{50}}\right)$.

Si ottiene, infine:

$(0,0543;0,346)$.

b) Deve essere:

$2z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}} \leqslant 0,3$, cioè:

$2\times 1,96\sqrt{\frac{0,2(0,8)}{n}} \leqslant 0,3$, ed elevando al quadrato ambo i membri:

$4\times 3,8416\frac{0,16}{n} \leqslant 0,09$, ed infine esplicitando n:

$n \geqslant \frac{15,3664\times 0,16}{0,09} = 27,318 \Rightarrow 28$.

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

Esercizio 3.

a) Si ha:

n=40;

$\overline{x} = 54$;

$\sigma = 1,0954$;

$\mu_0 = 50$;

$\alpha = 0,10$.

Si deve utilizzare il test Z normale, visto che la varianza è nota, con la statistica test seguente:

$\frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}$, che assume nel campione il valore:

$\frac{54 - 50}{1,0954 / \sqrt{40}} = \frac{4}{0,173} = 23,095$.

Poiché

$z_{0,95} = 1,645 < 23,095$ si rifiuta l'ipotesi nulla: $H_0 : \mu = 50$.

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:

$\chi^2 = 40\left(\frac{16^2}{20\times 21} + \frac{4^2}{20\times 10} + \frac{0^2}{20\times 9} + \frac{5^2}{21\times 20} + \frac{6^2}{10\times 20} + \frac{9^2}{9\times 20} - 1\right) = 40(0,6095 + 0,08 + 0 + 0,0595 + 0,18 + 0,45 - 1) = 15,16$

.

Poiché $\chi^2_{0,95;2} = 5.99146$ che è minore di 15,16, si rifiuta l'ipotesi nulla di indipendenza tra le due variabili.

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