Facoltà di Sociologia - A.A. 2005-2006

Esame di Analisi Multivariata – Prof.ssa Mecatti

Secondo recupero

04.07.2006

Avvertenza: Fornire le formule utilizzate e tutti i passaggi dei calcoli eseguiti.

Utilizzare almeno 2 cifre decimali.

Esercizio 1

a) Il numero medio di divorzi ogni 100 matrimoni in una comunità circoscritta del nord Europa è pari a 26. Calcolare la probabilità che in questa comunità, scelte a caso 11 persone che hanno contratto matrimonio, ce ne siano da 4 a 6 divorziate.

b) La durata X (espressa in mesi) del matrimonio prima del divorzio nella stessa comunità è interpretabile da una v.c. normale con media μ=40 e varianza σ^2=16. Si stabilisca la probabilità che la durata di un matrimonio nella comunità sia inferiore a 30 mesi.

c) In una data popolazione il fenomeno X è interpretato da una v.c. con media μ ignota; si estraggono due campioni bernoulliani in due momenti successivi, rispettivamente di numerosità 3 e 2. Le medie campionarie dei due campioni rappresentano due diversi stimatori per μ. Stabilire se i due stimatori sono non distorti e, spiegando perchè, stabilire quale dei due è preferibile.

Esercizio 2

a) Da un'indagine condotta su 60 immigrati in Italia, risulta che 24 di essi non hanno una conoscenza sufficiente della lingua italiana. Costruire un intervallo di confidenza al 95% per la proporzione di immigrati che non conosce la lingua italiana.

b) Quale deve essere la numerosità campionaria minima per avere un intervallo di confidenza a livello 95% ma di ampiezza non superiore a 0,4?

c) Esporre i concetti di errore campionario e di errore medio di stima e discutere la metodologia per la stima di quest'ultimo.

Esercizio 3

a) L'età di un collettivo di donne tra 20 e 60 anni che si sono separate legalmente entro tre anni dal matrimonio si distribuisce normalmente con s.q.m. pari a 2. In una recente indagine, è risultato che 60 donne tra 20 e 60 anni, separate entro tre anni dal matrimonio, hanno un'età media pari a 42 anni. Impostare un adeguato test sulla base di questi risultati campionari, verificando l'ipotesi nulla per cui l'età media è pari a 41 anni, fissato $\alpha = 0,10$.

b) 40 studenti di terza media sono stati sottoposti a un test di matematica (X) e uno di lingua inglese (Y). I risultati possibili sono *del tutto insufficiente*, *insufficiente*, *quasi sufficiente*, *sufficiente*, *buono* e *ottimo*. I risultati sono riassunti nella seguente tabella a doppia entrata:

Y / X	del tutto insuff.	insuff.	quasi suff.	suff.	buono	ottimo	totale
del tutto insuff.	6	2	2	1	1	0	12
insuff.	3	5	3	2	0	0	13
quasi suff.	3	2	3	0	0	0	8
suff.	0	0	0	2	1	0	3
buono	0	0	0	0	2	2	4
totale	12	9	8	5	4	2	40

Verificare a livello α= 0,05, se esiste indipendenza tra X e Y utilizzando un opportuno test statistico e commentando il risultato.

c) Definire e classificare le ipotesi statistiche, discutere i concetti di accettazione/rifiuto dell'ipotesi nulla e definire l'errore di I specie ed il livello di significatività di un test statistico.

Facoltà di Sociologia - A.A. 2005-2006

Esame di Analisi Multivariata – Prof.ssa Mecatti

Soluzioni appello del 04.07.2006[1]

Esercizio 1. a) La prova singola consiste in una "'estrazione" da un insieme di persone coniugate. La probabilità di "estrarre" una persona divorziata è data da $p = 0.26$. La legge binomiale, la cui funzione di probabilità è

$$P\left(X = x\right) = \left(\begin{array}{c} n \\ x \end{array}\right) p^x \left(1 - p\right)^{n-x},$$

si presta ad interpretare il fenomeno considerato. L'evento d'interesse è *su 11 persone ce ne sono da 4 a 6 divorziate*. Si deve valutare la legge binomiale per n=11, X=4,5,6, e poi sommare i valori delle probabilità ottenute. Quindi:

$$\begin{array}{l} P\left(X = 4\right) + P\left(X = 5\right) + P\left(X = 6\right) = \left(\begin{array}{c} 11 \\ 4 \end{array}\right) 0,26^4 \cdot 0,74^7 + \left(\begin{array}{c} 11 \\ 5 \end{array}\right) 0,26^5 \cdot 0,74^6 + \left(\begin{array}{c} 11 \\ 6 \end{array}\right) 0,26^6 \cdot 0,74^5 = \\ 330 \cdot 0{,}0046 \cdot 0{,}122 + 462 \cdot 0{,}0012 \cdot 0{,}164 + 462 \cdot 0{,}0003 \cdot 0{,}222 = 0{,}0185 + 0{,}091 + 0{,}0307 = 0{,}14 \end{array}.$$

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

$$P\left(X < 30\right) = P(Z < \frac{30 - 40}{\sqrt{16}}) = P(Z < -\frac{10}{4}) = 1 - P(Z < \frac{10}{4}) = 1 - P(Z < 2,50) = 1 - 0,9938 = 0,0062$$

c) (i) e (ii). In generale, lo stimatore T si dice non distorto per il parametro μ se $E\left[T\right] = \mu$. In questo caso si ha:

$$\begin{array}{l} E\left[T_1\right] = E[\frac{X_1+X_2+X_3}{3}] = \frac{1}{3}E[X_1 + X_2 + X_3] = \frac{1}{3}\left\{E[X_1] + E[X_2] + E[X_3]\right\} = [i.i.d] = \\ = \frac{1}{3}\left\{E[X] + E[X] + E[X]\right\} = \frac{3}{3}\mu = \mu \end{array}$$

$$\begin{array}{l} E\left[T_2\right] = E[\frac{X_1+X_2}{2}] = \frac{1}{2}E[X_1 + X_2] = \frac{1}{2}\left\{E[X_1] + E[X_2]\right\} = [i.i.d] = \\ = \frac{1}{2}\left\{E[X] + E[X]\right\} = \frac{2}{2}\mu = \mu \end{array}$$

I due stimatori sono non distorti. Poiché, però:

$$\begin{array}{l} Var\left[T_1\right] = Var[\frac{X_1+X_2+X_3}{3}] = \frac{1}{9}Var[X_1 + X_2 + X_3] = \frac{1}{9}\left\{Var[X_1] + Var[X_2] + Var[X_3]\right\} = [i.i.d] = \\ = \frac{1}{9}\left\{Var[X] + Var[X] + Var[X]\right\} = \frac{1}{3}\sigma^2 \end{array}$$

$$\begin{array}{l} Var\left[T_2\right] = Var[\frac{X_1+X_2}{2}] = \frac{1}{4}Var[X_1 + X_2] = \frac{1}{4}\left\{Var[X_1] + Var[X_2]\right\} = [i.i.d] = \\ = \frac{1}{4}\left\{Var[X] + Var[X]\right\} = \frac{1}{2}\sigma^2 \end{array}$$

è da preferire il primo perché $Var[T_1] < Var[T_2]$

Esercizio 2. a) La proporzione campionaria è pari a:

$$p = \frac{24}{60} = \frac{2}{5} = 0,4.$$

Il livello di confidenza è:

$$\alpha = 0,05.$$

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:

$P(p - z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}} \leq P \leq p + z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}) = 0,95$. L'intervallo di confidenza sarà dunque:

$(0,4 - z_{0,975}\sqrt{\frac{0,4\times0,6}{60}}; 0,4 + z_{0,975}\sqrt{\frac{0,4\times0,6}{60}})$, e quindi:

$$(0,4 - 1,96\sqrt{\frac{0,4 \times 0,6}{60}}; 0,4 + 1,96\sqrt{\frac{0,4 \times 0,6}{60}}).$$

Si ottiene, infine:

$$\left(0,276; 0,524\right).$$

b) Deve essere:

$2z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}} \leq 0,4$, cioè:

[1] A cura di Giancarlo Manzi

$2 \times 1,96\sqrt{\frac{0,4(0,6)}{n}} \leq 0,4$, ed elevando al quadrato ambo i membri:
$4 \times 3,8416\frac{0,24}{n} \leq 0,16$, ed infine esplicitando n:

$$n \geq \frac{15,3664 \times 0,24}{0,16} = 23,0496 \Rightarrow 24.$$

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

Esercizio 3. a) Si ha:

$$n{=}60;$$

$$\overline{x} = 42;$$

$$\sigma = 2;$$

$$\mu_0 = 41;$$

$$\alpha = 0,10.$$

Si può utilizzare il test Z normale con la statistica-test seguente:
$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$, che assume nel campione il valore:

$$\frac{42-41}{2/\sqrt{60}} = \frac{1}{0,258} = 3,873.$$

Poiché
$z_{0,95} = 1,645 < 3,873$ si rifiuta l'ipotesi nulla: $H_0 : \mu = 41$.
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:

$$\chi^2 = 47,023.$$

Poiché $\chi^2_{0,95;2} = 5,99146$ che è minore di 47,023, si rifiuta dunque l'ipotesi nulla di indipendenza tra le due variabili.
c) Si vedano appunti e/o libri di testo.