**Facoltà di Sociologia - A.A. 2005-2006**
**Esame di Analisi Multivariata – Prof.ssa Mecatti**
**Terzo recupero**
**25.09.2006**

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

**Esercizio 1**

a) Si supponga che la probabilità di *non* ricevere una multa se si parcheggia in una zona centrale con divieto di sosta tra le ore 10.00 e le ore 12.00 di un giorno feriale è 0.9 a Roma, 0.8 a Milano e 0.1 a Lugano in Svizzera. Stabilire in quale delle tre città si ha la probabilità maggiore di non ricevere multe, se si parcheggia in sosta vietata 44 volte a Roma, 20 volte a Milano e 2 volte a Lugano. E in quale città si ha la probabilità minore? (*Suggerimento*: assumere l'indipendenza dei tre eventi di parcheggio nelle tre città ed usare tre binomiali).

b) Il tempo medio di ritardo in partenza dei voli in 10 aeroporti inglesi è risultato 13,83 minuti nel 2004 e 14,89 minuti nel 2005. I rispettivi s.q.m. sono risultati 13,2 e 12,1. Supponendo che il ritardo $X$ (espresso in minuti) dei voli in partenza nei 10 scali possa essere interpretato da una v.c. normale, si stabilisca in quale dei due anni è stata maggiore la probabilità di partire con un ritardo superiore a 45 minuti.

c) Si considerino i seguenti tre stimatori della media ignota $\mu$ di una popolazione con varianza pari a $\sigma^2$, basati su un campionamento bernoulliano di numerosità 3:

$$T_1 = \frac{1}{3}X_1 + \frac{2}{3}X_3;$$

$$T_2 = \frac{1}{4}X_2 + \frac{3}{4}X_3;$$

$$T_3 = \frac{1}{5}X_1 + \frac{4}{5}X_2.$$

Stabilire se i tre stimatori sono non distorti per $\mu$ e, chiarendo il criterio, stabilire quale dei tre è preferibile.

**Esercizio 2**

a) Da un'indagine condotta su un campione bernoulliano di 105 italiani, risulta che 35 di essi risultano molto interessati alla politica. Costruire un intervallo di confidenza al 95% per la proporzione di italiani molto interessati alla politica.

b) Quale deve essere la numerosità campionaria minima per costruire un intervallo di confidenza a livello 95% di ampiezza non superiore a 0,2?

c) Esporre il problema della stima puntuale della % di fenomeni categoriali, definire lo stimatore *proporzione campionaria* e discuterne le proprietà.

**Esercizio 3**

a) "Supererai tre esami al mese!" Così recita lo spot di una società privata di preparazione agli esami universitari. Da un'indagine statistica su un collettivo di 100 studenti di sociologia risulta invece che il tempo medio di preparazione ad un modulo d'esame è di circa 13 giorni con una varianza di 4,5 giorni. Esso inoltre risulta distribuito normalmente. La varianza nella popolazione è ignota. Costruire un adeguato test bilaterale (considerando un mese di 30 giorni) sull'affermazione della società privata e sulla base dei risultati campionari, fissando $\alpha = 0{,}10$.

b) 40 consumatori di un supermercato hanno risposto a due domande sul prezzo dei generi alimentari (*X*) e di quello degli elettrodomestici (*Y*) dopo l'entrata in vigore dell'euro. I risultati sono riassunti nella seguente tabella a doppia entrata:

| ***Y*** / ***X*** | **Sono aumentati molto** | **Sono aumentati di poco** | **Sono rimasti stabili** | **Sono diminuiti di poco** | **Sono diminuiti molto** | **Totale** |
|---|---|---|---|---|---|---|
| **Sono aumentati molto** | 20 | 10 | 0 | 2 | 0 | **32** |
| **Sono aumentati di poco** | 2 | 0 | 4 | 0 | 0 | **6** |
| **Sono rimasti stabili** | 0 | 2 | 0 | 0 | 0 | **2** |
| **Sono diminuiti di poco** | 0 | 0 | 0 | 0 | 0 | **0** |
| **Sono diminuiti molto** | 0 | 0 | 0 | 0 | 0 | **0** |
| **Totale** | **22** | **12** | **4** | **2** | **0** | **40** |

Verificare a livello $\alpha = 0{,}05$, se esiste indipendenza tra *X* e *Y* utilizzando un opportuno test statistico e commentando il risultato.

c) Esporre le caratteristiche e le proprietà della variabile casuale Normale e discuterne l'utilità nella costruzione di un test statistico.

**Facoltà di Sociologia - A.A. 2005-2006**
**Esame di Analisi Multivariata – Prof.ssa Mecatti**
**Soluzioni appello del 25.09.2006**[1]

**Esercizio 1.**

a) La prova singola consiste in un parcheggio in sosta vietata, ciascuno in una delle tre città citate. La probabilità di "farla franca" parcheggiando una volta è data da $p_R = 0.90$ a Roma, da $p_M = 0.80$ a Milano e da $p_L = 0.10$ a Lugano. La legge binomiale, la cui funzione di probabilità è

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

si presta ad interpretare il fenomeno considerato. L'evento d'interesse è *su 44, 20 e 2 parcheggi, rispettivamente a Roma, Milano e Lugano*. Si deve valutare la legge binomiale con i parametri *n*=44, *X*=44 e $p_R = 0.90$ per Roma, con i parametri *n*=20, *X*=20 e $p_M = 0{,}80$ per Milano e con i parametri *n*=2, *X*=2 e $p_R = 0.10$ per Lugano. Si ottiene:

$$P_{Roma}(X = 44) = \binom{44}{44} 0{,}90^{44} \bullet 0{,}10^{0} = 0{,}90^{44} = 0{,}009698\,;$$

$$P_{Milano}(X = 20) = \binom{20}{20} 0{,}80^{20} \bullet 0{,}20^{0} = 0{,}80^{20} = 0{,}011529\,;$$

$$P_{Lugano}(X = 2) = \binom{2}{2} 0{,}10^{2} \bullet 0{,}90^{0} = 0{,}10^{2} = 0{,}01\,.$$

Per cui, sia pur di poco, la probabilità maggiore si riscontra a Milano. La probabilità minore a Roma.

b)

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

$$P_{2004}(X > 45) = P_{2004}\left(Z > \frac{45 - 13{,}83}{13{,}2}\right) = P_{2004}(Z > 2{,}36) = 0{,}0091$$

$$P_{2005}(X > 45) = P_{2005}\left(Z > \frac{45 - 14{,}89}{12{,}1}\right) = P_{2005}(Z > 2{,}49) = 0{,}0064\,,$$

per cui, nonostante il tempo medio di ritardo sia stato maggiore nel 2004 che nel 2005, la probabilità di un ritardo maggiore di 45 minuti è stato maggiore nel 2004.

c)

In generale, lo stimatore *T* si dice non distorto per il parametro $\mu$ se $E[T] = \mu$. In questo caso si ha:

$$E[T_1] = E\left[\frac{X_1 + 2X_3}{3}\right] = \frac{1}{3}E[X_1 + 2X_3] = \frac{1}{3}\{E[X_1] + 2E[X_3]\} = \mu$$

---

[1] A cura di Giancarlo Manzi

$$E[T_2] = E[\frac{X_2 + 3X_3}{4}] = \frac{1}{4}E[X_2 + 3X_3] = \frac{1}{4}\{E[X_2] + 3E[X_3]\} = \mu$$
$$E[T_3] = E[\frac{X_1 + 4X_2}{5}] = \frac{1}{5}E[X_1 + 4X_2] = \frac{1}{5}\{E[X_1] + 4E[X_2]\} = \mu$$

I tre stimatori sono dunque non distorti. Poiché, però:

$$Var[T_1] = Var[\frac{X_1 + 2X_3}{3}] = \frac{1}{9}Var[X_1 + 2X_3] = \frac{1}{9}\{Var[X_1] + Var[2X_3]\} =$$
$$= \frac{1}{9}\{Var[X] + 4Var[X]\} = \frac{5}{9}\sigma^2$$
$$Var[T_2] = Var[\frac{X_2 + 3X_3}{4}] = \frac{1}{16}Var[X_2 + 3X_3] = \frac{1}{16}\{Var[X_2] + Var[3X_3]\} =$$
$$= \frac{1}{16}\{Var[X] + 9Var[X]\} = \frac{5}{8}\sigma^2$$
$$Var[T_3] = Var[\frac{X_1 + 4X_2}{5}] = \frac{1}{25}Var[X_1 + 4X_2] = \frac{1}{25}\{Var[X_1] + Var[4X_2]\} =$$
$$= \frac{1}{25}\{Var[X] + 16Var[X]\} = \frac{17}{25}\sigma^2$$

è da preferire il primo perché ha varianza inferiore rispetto agli altri due.

**Esercizio 2.**

a) La proporzione campionaria è pari a:

$p = \frac{35}{105} = \frac{1}{3} \approx 0{,}33$.

Il livello di confidenza è:

$\alpha = 0{,}05$.

La numerosità campionaria ($n$=105) è sufficientemente elevata; possiamo allora costruire l'intervallo di confidenza asintotico all' $(1 - \alpha) \times 100$ per la percentuale di individui molto interessati alla vita politica:

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

$(0{,}33 - z_{0,975}\sqrt{\frac{0{,}33 \times 0{,}66}{105}}; 0{,}33 + z_{0,975}\sqrt{\frac{0{,}33 \times 0{,}66}{105}})$, e quindi:

$(0{,}33 - 1{,}96\sqrt{\frac{0{,}33 \times 0{,}66}{105}}; 0{,}33 + 1{,}96\sqrt{\frac{0{,}33 \times 0{,}66}{105}})$.

Si ottiene, infine:

$(0{,}33 - 0.089; 0{,}33 + 0.089)$, cioè $(0{,}244; 0{,}422)$.

b) Deve essere:

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

$2\times 1{,}96\sqrt{\frac{0{,}33(0{,}666)}{n}} \leqslant 0{,}2$, ed elevando al quadrato ambo i membri:

$4\times 3{,}8416\frac{0{,}22}{n} \leqslant 0{,}04$, ed infine esplicitando $n$:

$n \geqslant \frac{15{,}3664\times 0{,}22}{0{,}04} = 84{,}52 \Rightarrow 85$.

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

**Esercizio 3.**

a) Si ha:

$n$=100;

$\bar{x} = 13$;

$s = 2{,}12$;

$\mu_0 = 10$;

$\alpha = 0{,}10$.

Si può utilizzare il test $Z$ normale perché, pur essendo la varianza della popolazione ignota, la numerosità campionaria è sufficientementemente elevata. Allora, la statistica-test sarà la seguente:

$\frac{\bar{X} - \mu_0}{s/\sqrt{n}}$, che assume nel campione il valore:

$\frac{13-10}{2{,}12/\sqrt{100}} = \frac{3}{0{,}212} = 14{,}15$.

Poiché

$z_{0,95} = 1{,}645 < 14{,}15$ si rifiuta l'ipotesi nulla: $H_0 : \mu = 10$ e l'affermazione della società privata di preparazione agli esami universitari può venire contestata.

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:

$\chi^2 = 30{,}189$.

Poiché $\chi^2_{0,95;6} = 12{,}5916$ che è minore di 30,189, si rifiuta dunque l'ipotesi nulla di indipendenza tra le due variabili.

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