

FACOLTÀ DI SOCIOLOGIA - A.A. 2005-2006
SIMULAZIONE D'ESAME DI ANALISI MULTIVARIATA
22.03.2006

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

Esercizio 1:

- a) Un proprietario di una piccola azienda decide di inserire 12 annunci pubblicitari su 12 numeri di una rivista, uno per ogni numero della rivista. Per esperienza crede che la probabilità che un lettore sia esposto ad un annuncio pubblicitario (cioè lo veda e ne rimanga colpito) sia pari a $p=0,50$, ed inoltre crede che le "esposizioni" all'annuncio pubblicitario siano indipendenti tra loro, cioè il fatto che un lettore veda l'annuncio pubblicitario nel primo numero della rivista in cui compare non influenza il fatto che lo veda nel secondo, nel terzo, ecc. Fornire la probabilità che, fatte queste assunzioni, un lettore qualsiasi veda almeno quattro volte l'annuncio pubblicitario.
- b) La durata X degli spot pubblicitari di marche automobilistiche su una televisione commerciale segue la legge normale con media $\mu=45$ secondi e scarto quadratico medio pari a $\sigma=10$. La durata Y degli spot pubblicitari di marche di telefoni cellulari su una televisione commerciale segue la legge normale con media $\mu=35$ secondi e scarto quadratico medio pari a $\sigma=15$. Si stabilisca la probabilità che
- (i) uno spot di una marca automobilistica sia compreso tra 25 e 30 secondi;
- (ii) uno spot di una marca di telefoni cellulari sia compreso tra 30 e 40 secondi.
- c) Sia X una v.c. cosiddetta *rettangolare* con funzione di probabilità:

$$P(X = x) = \frac{1}{N} \text{ per } x = 1, 2, \dots, N.$$

(i) Se ne calcoli il valore atteso (Suggerimento: utilizzare la relazione $\sum_{i=1}^N i = \frac{N(N+1)}{2}$).

(ii) Se ne tracci il grafico per $N=6$.

Esercizio 2:

È stata misurata l'altezza di 5 piantine di mais a 30 giorni dalla semina, ottenendo le seguenti misurazioni (in cm):
 22, 23, 30, 31, 22.

- a) Supponendo che l'altezza segua una legge normale, determinare un intervallo di confidenza a livello 95% per l'altezza media.
- b) Quale deve essere la numerosità campionaria minima perché la probabilità di commettere un errore assoluto di stima non superiore ad 1 cm sia pari a $(1-\alpha)=0.9$?
- c) Dopo aver esposto il concetto di *precisione* di una stima intervallare, discutere e interpretare le relazioni fra livello di confidenza, ampiezza dell'intervallo e numerosità campionaria.

Esercizio 3:

a) A proposito del faccia a faccia televisivo di due candidati ad un'elezione politica, un quotidiano riporta la notizia che gli ascoltatori che hanno visionato il programma sono stati per il 70% uomini. Dopo aver selezionato un campione di 400 ascoltatori che hanno visto il faccia a faccia, si vuole verificare la notizia riportata sul quotidiano. I risultati campionari hanno evidenziato che 220 ascoltatori sono stati di sesso maschile. Che test si può usare per la verifica della notizia data? Questo test scelto risulta significativo a livello di significatività del 5%?

b) In un sondaggio sulle preferenze elettorali in base al sesso, viene chiesto a 2800 individui quali fossero le loro preferenze politiche, ottenendo la seguente distribuzione doppia:

X	Y			
	Primo schieramento	Secondo schieramento	Altri	
M	650	750	100	1500
F	500	700	100	1300
	1150	1450	200	2800

Si può formulare l'ipotesi, scegliendo $\alpha=0,05$, che la preferenza politica è indipendente dal sesso?

c) A che cosa serve il p -value? Discuterne i valori per la valutazione di un test a due code.

FACOLTÀ DI SOCIOLOGIA - A.A. 2005-2006
ESAME DI ANALISI MULTIVARIATA
Soluzioni simulazione del 22.03.2006

Esercizio 1.

a)

Passo 1. *Individuazione del problema.*

Le "estrazioni" sono indipendenti ed inoltre le prove sono bernoulliane. Possiamo utilizzare il modello binomiale espresso dalla seguente f.p.:

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

Passo 2. *Individuazione dei dati a disposizione.*

$$p = 0,50$$

$$1 - p = 0,50$$

$$n = 12$$

Passo 3. *Probabilità da calcolare.*

$$\begin{aligned} P(X \geq 4) &= 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)] = \\ &= 1 - \left[\binom{12}{0} 0,5^0 (0,5)^{12} + \binom{12}{1} 0,5^1 (0,5)^{11} + \binom{12}{2} 0,5^2 (0,5)^{10} + \binom{12}{3} 0,5^3 (0,5)^9 \right] = \\ &= 1 - [0,000244 + 0,002930 + 0,016113 + 0,053711] = 0,927 \end{aligned}$$

b)

Passo 2. *Individuazione dei dati a disposizione.*

Marche automobilistiche:

$$\mu = 45$$

$$\sigma = 10$$

Telefoni cellulari:

$$\mu = 35$$

$$\sigma = 15$$

Passo 3. *Probabilità da calcolare.*

Marche automobilistiche:

$$\begin{aligned} P(25 < X < 30) &= [\text{standardizzando}] = P \frac{25 - 45}{10} < \frac{X - 45}{10} < \frac{30 - 45}{10} = P(-2 < Z < -1,5) = \\ &= P(Z < -1,5) - P(Z < -2) = P(Z < 2) - P(Z < 1,5) = 0,9772 - 0,9332 = 0,044 \end{aligned}$$

Telefoni cellulari:

$$\begin{aligned} P(30 < X < 40) &= [\text{standardizzando}] = P \frac{30 - 35}{15} < \frac{X - 35}{15} < \frac{40 - 35}{15} = P(-0,333 < Z < 0,333) = \\ &= 2[P(Z < 0,333) - P(Z < 0)] = 2 \times 0,1293 = 0,2586 \end{aligned}$$

c)

(i)

Passo 1. *Individuazione del problema.*

Si tratta di una v.c. discreta.

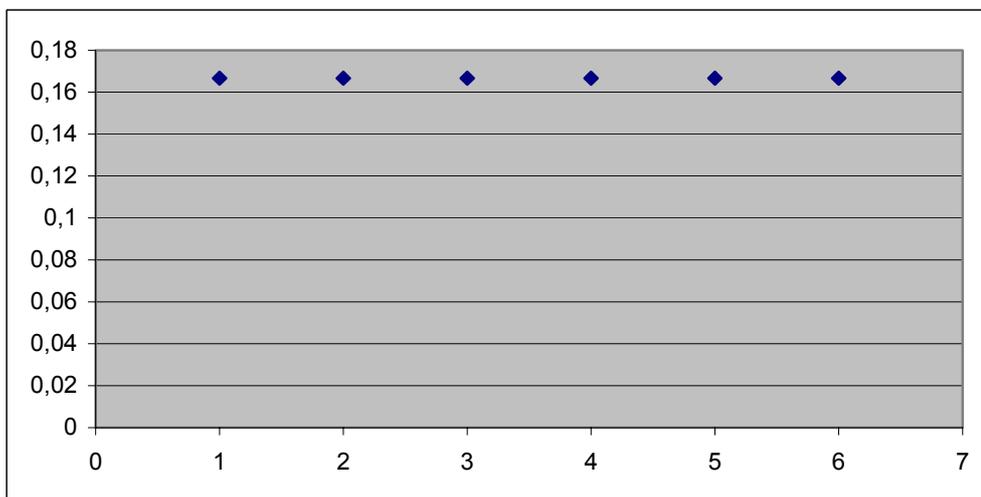
Passo 2. *Formula del valore atteso per una v.c. discreta.*

$$E(X) = \sum_{i=1}^n x_i P(X = x_i)$$

Passo 3. *Calcolo del valore atteso utilizzando il suggerimento dato.*

$$E(X) = \frac{1}{N} (\sum_{i=1}^n x_i) = \frac{1}{N} \frac{N(N+1)}{2} = \frac{N+1}{2}$$

(ii)



Esercizio 2.

a) Dato che il campione proviene da una popolazione normale con varianza ignota, l'IC risulta basato su una v.c. T di Student con $n-1=4$ gdl.

Per prima cosa calcolo la media campionaria e la varianza campionaria corretta.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{22 + 23 + 30 + 31 + 22}{5} = 25.6$$

$$\begin{aligned} s^2 &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \\ &= \frac{1}{4} [(22 - 25.6)^2 + (23 - 25.6)^2 + (30 - 25.6)^2 + (31 - 25.6)^2 + (22 - 25.6)^2] = 20.3 \end{aligned}$$

Un teorema di Teoria delle Probabilità ci assicura che:

$$\frac{\bar{X} - \mu}{\sqrt{S^2/n}} \sim T_{n-1}.$$

Sfruttiamo questo risultato per costruire l'intervallo di confidenza a livello $(1-\alpha)=0.95$.

$$P\left(t_{n-1, \alpha/2} \leq \frac{\bar{X} - \mu}{\sqrt{S^2/n}} \leq +t_{n-1, \alpha/2}\right) = 1 - \alpha$$

$$P\left(\bar{X} - t_{n-1, \alpha/2} \sqrt{\frac{S^2}{n}} \leq \mu \leq \bar{X} + t_{n-1, \alpha/2} \sqrt{\frac{S^2}{n}}\right) = 1 - \alpha$$

L'intervallo di confidenza risulta perciò:

$$\left[\bar{x} - t_{4, 0.025} \sqrt{\frac{s^2}{n}}, \bar{x} + t_{4, 0.025} \sqrt{\frac{s^2}{n}}\right] = \left[25.6 - 2.78 \sqrt{\frac{20.3}{5}}, 25.6 + 2.78 \sqrt{\frac{20.3}{5}}\right]$$

$$= [20, 31.2]$$

b) Si richiede che con probabilità 0.9 valga la seguente disuguaglianza:

$$1 - \alpha = P(|\bar{X} - \mu| \leq Err)$$

$$= P(\mu - Err \leq \bar{X} \leq \mu + Err)$$

$$= P\left(\frac{\mu - Err - \mu}{\sqrt{S^2/n}} \leq \frac{\bar{X} - \mu}{\sqrt{S^2/n}} \leq \frac{\mu + Err - \mu}{\sqrt{S^2/n}}\right)$$

Vale allora l'uguaglianza: $\frac{Err}{\sqrt{s^2/n}} = t_{n-1, \alpha/2} \Rightarrow n = \frac{t_{n-1, \alpha/2}^2 \cdot s^2}{Err^2} = \frac{2.13^2 \cdot 20.3}{1} = 92.1 \approx 93$

c)

Intuitivamente, un IC è tanto *più preciso* (e utile per fare inferenza sul parametro ignoto) quanto più è *stretto*, quanto meno è *ampio*. La precisione di un IC è in relazione *inversa* col livello di confidenza e in relazione *diretta* con la numerosità campionaria. A parità di ampiezza campionaria n un aumento del livello di confidenza $(1 - \alpha)$ provoca una diminuzione di precisione (un aumento di ampiezza dell'IC). E viceversa. A parità di livello di confidenza $(1 - \alpha)$ un aumento della numerosità campionaria n provoca un aumento della precisione (una diminuzione dell'ampiezza dell'IC). E viceversa.

Esercizio 3.

a)

Passo 1. *Impostazione del sistema d'ipotesi.*

$$H_0 : p = 0,7$$

$$H_1 : p \neq 0,7$$

Passo 2. *Individuazione dei dati per l'impostazione del test.*

$$p_0 = 0,7;$$

$$n=400;$$

$$\hat{p} = \frac{220}{400} = 0,55$$

Passo 3. *Individuazione della statistica test.*

Si può fare riferimento al test normale per una frequenza secondo cui:

$$\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0,1)$$

Passo 4. *Calcolo della statistica test.*

$$Z = \frac{\hat{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0,55 - 0,7}{\sqrt{\frac{0,7 \times 0,3}{400}}} = \frac{-0,15}{0,023} = -6,522 .$$

Passo 5. *Confronto con il valore teorico che individua la regione di rifiuto.*

Il valore teorico per una significatività $\alpha = 0,05$ è dato da -1,96 che è maggiore del valore sperimentale trovato.

Passo 6. *Decisione.*

Si rifiuta l'ipotesi nulla.

b)

Passo 1. *Individuazione dei valori teorici d'indipendenza stocastica.*

X \ Y	CDL	Unione	Altri	
M	616,071	776,786	107,143	1500
F	533,929	673,214	92,857	1300
	1150	1450	200	2800

Passo 2. *Calcolo del χ^2 empirico.*

$$\chi^2 = 7,04 .$$

Passo 3. *Individuazione dei df.*

Numero delle righe: 2

Numero delle colonne: 3

$$df = 2 .$$

Passo 4. *Individuazione del χ^2 teorico.*

$$\chi_{0,95;2}^2 = 5,991 \text{ che è minore del valore trovato.}$$

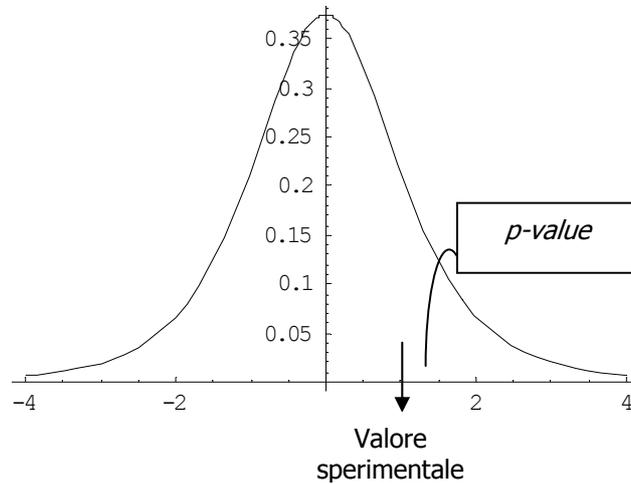
Passo 6. *Decisione.*

Si rifiuta l'ipotesi nulla.

c)

Il p -value è la probabilità *a destra* del valore sperimentale sotto la curva della statistica test.

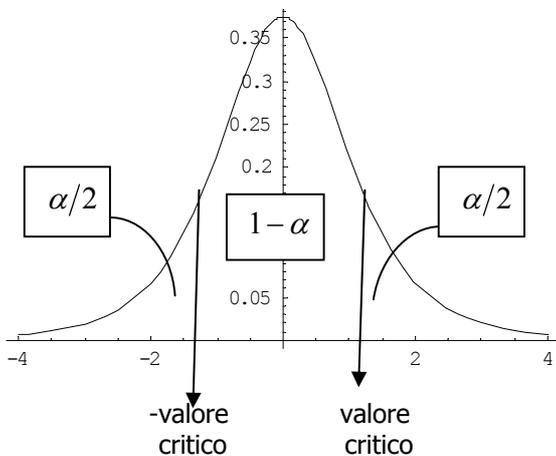
$$p\text{-value} \rightarrow P(\text{statistica test} \geq \text{valore sperimentale})$$



Ipotesi bilaterale/Test a due code \rightarrow si rifiuta H_0 a livello di significatività $(1 - \alpha)$ se

$$p\text{-value} \leq \alpha/2 \text{ oppure se } p\text{-value} \geq 1 - \alpha/2$$

Infatti:



- Se: valore sperimentale \geq valore critico allora cade nella coda di destra e si ha:

$$p\text{-value} = P(\text{a destra del valore sperimentale}) \leq \alpha/2$$

- Se: valore sperimentale \leq - valore critico allora cade nella coda di sinistra e si ha:

$$p\text{-value} = P(\text{a destra del valore sperimentale}) \geq 1 - \alpha/2$$

Zona di accettazione

Regione critica