

PROVA SCRITTA DI STATISTICA (COD 4038-5047)
CLEA-CLEFIN-CLAPI-CLEMIT
10 luglio 2003
MODALITÀ A

APPROSSIMARE TUTTI I CALCOLI ALLA QUARTA CIFRA DECIMALE

ESERCIZIO 1 (10 punti)

L'agenzia di viaggio "Volare Volare Single" preoccupata di un'improvvisa diminuzione di clienti, decide di intervistare un collettivo di 15 clienti circa la soddisfazione della loro recente vacanza.

Nella seguente tabella, per ciascun cliente è indicato:

DESTINAZIONE Destinazione del viaggio (Italia, Europa, America, Africa)

SESSO (f=femmina, m=maschio)

GIORNI Durata vacanza in giorni

TEMA vacanza (i clienti sono solo quelli che hanno scelto una vacanza per motivi CULTURALI, o per SCIARE o per SUB, immersioni subacquee)

SODD Soddisfazione: 1=poca, 2=media= 3=alta

Destinazione	Sesso	Giorni	Tema	Sodd	Giorni²
Europa	m	15	Sci	3	225
Italia	f	7	Sci	2	49
Africa	f	7	Sub	1	49
Europa	m	7	Sci	3	49
Europa	f	14	Culturale	2	196
Italia	m	3	Sub	2	9
America	m	14	Culturale	3	196
America	m	8	Sci	2	64
Italia	m	3	Culturale	2	9
Africa	f	13	Culturale	3	169
America	m	7	Culturale	1	49
Italia	f	7	Culturale	1	49
Africa	f	15	Sub	3	225
Europa	f	7	Culturale	3	49
Italia	m	3	Culturale	2	9
somma		130			1396

- 1) Dopo aver rappresentato tramite una tabella la distribuzione del carattere GIORNI, calcolare un'opportuna misura di posizione (1+1 punti)

DISTRIBUZIONE DI FREQUENZA

Giorni	Frequenze assolute	Frequenze relative	Funzione di ripartizione
3	3	20.00%	20.00%
7	6	40.00%	60.00%
8	1	6.67%	66.67%
13	1	6.67%	73.33%
14	2	13.33%	86.67%
15	2	13.33%	100.00%
Totale complessivo	15	100.00%	

Misure di posizione: **MEDIANA: 7 MEDIA:8.66**
(MODA:7)

2) I caratteri SODD e TEMA sono statisticamente indipendenti? Giustificare la risposta (2 punti)

Frequenze assolute	Sodd				
Tema	1	2	3	Totale complessivo	
Culturale	2	3	3		8
Sci		2	2		4
Sub	1	1	1		3
Totale complessivo	3	6	6		15

Frequenze relative	Sodd				
Tema	1	2	3	Totale complessivo	
Culturale	13.33%	20.00%	20.00%		53.33%
Sci		13.33%	13.33%		26.67%
Sub	6.67%	6.67%	6.67%		20.00%
Totale complessivo	20.00%	40.00%	40.00%		100.00%

Nella tabella 2x2 è presente una cella con frequenza assoluta (relativa) nulla, quindi i due caratteri non sono indipendenti

3) Qual è la moda del carattere TEMA tra clienti i più soddisfatti? (1 punto)

MODA (tema|SODD=3)=CULTURALE

Distribuzione condizionata

Tema	Frequenze assolute	Frequenze relative
Culturale	3	50.00%
Sci	2	33.33%
Sub	1	16.67%
Totale complessivo	6	100.00%

4) Possiamo affermare che in media le donne hanno scelto viaggi più lunghi degli uomini? Giustificare la risposta (2 punti)

$$M(\text{Giorni} \mid \text{Maschio}) = \frac{15 + 7 + 3 + 14 + 8 + 3 + 7 + 3}{8} = \frac{60}{8} = 7.5$$

$$M(\text{Giorni} \mid \text{Femmina}) = \frac{7 + 7 + 14 + 13 + 7 + 15 + 7}{7} = \frac{70}{7} = 10$$

la media del carattere giorni condizionato a Sesso=F è maggiore della media condizionata a Sesso=M

5) La lunghezza del viaggio è più variabile tra gli uomini o tra le donne? Motivare la risposta (3 punti)

$$M(\text{Giorni}^2 \mid \text{Maschio}) = \frac{15^2 + 7^2 + 3^2 + 14^2 + 8^2 + 3^2 + 7^2 + 3^2}{8} = \frac{610}{8} = 76.25$$

$$M(\text{Giorni}^2 \mid \text{Femmina}) = \frac{7^2 + 7^2 + 14^2 + 13^2 + 7^2 + 15^2 + 7^2}{7} = \frac{786}{7} = 112.2857$$

$$\text{Var}(\text{Giorni} \mid \text{Maschio}) = M(\text{Giorni}^2 \mid \text{Maschio}) - M(\text{Giorni} \mid \text{Maschio})^2 = 76.25 - 7.5^2 = 20$$

$$\text{Var}(\text{Giorni} \mid \text{Femmine}) = M(\text{Giorni}^2 \mid \text{Femmine}) - M(\text{Giorni} \mid \text{Femmine})^2 = 112.2857 - 10^2 = 12.2857$$

$$CV(\text{Giorni} \mid \text{Maschi}) = \frac{\sqrt{20}}{7.5} = 0.5963$$

$$CV(\text{Giorni} \mid \text{Femmine}) = \frac{\sqrt{12.2857}}{10} = 0.3505$$

La lunghezza del viaggio è più variabile tra gli uomini che tra le donne

ESERCIZIO 2 (6 punti)

Per prevenire la diffusione dell'epidemia della Sars (polmonite atipica), si è deciso di controllare tutti i passeggeri provenienti da paesi infetti. Particolare attenzione si presta ai voli provenienti da Toronto (supponiamo che la settimana precedente ci siano stati alcuni casi di pazienti affetti da Sars proprio all'aeroporto internazionale di Toronto).

Si hanno a disposizione le seguenti informazioni:

- una persona affetta da Sars ha febbre alta (superiore a 38°) con probabilità 0.8,
- una persona sana non contagiata da Sars, ha probabilità di avere febbre alta uguale a 0.1,
- la probabilità che un passeggero proveniente da Toronto sia affetto da Sars è 0.0001

- Quale è la probabilità che un viaggiatore proveniente da Toronto abbia febbre alta al suo arrivo?
- Nell'ipotesi che si sia trovato nel volo proveniente da Toronto un viaggiatore con la febbre alta (superiore a 38°), quale è la probabilità che tale viaggiatore sia affetto da SARS?

$$P(\text{Febbre} \mid \text{SARS}) = 0.8$$

$$P(\text{Febbre} \mid \overline{\text{SARS}}) = 0.1$$

$$P(\text{SARS}) = 0.0001$$

$$P(\text{Febbre}) = P(\text{Febbre} \mid \text{SARS}) \times P(\text{SARS}) + P(\text{Febbre} \mid \overline{\text{SARS}}) \times P(\overline{\text{SARS}}) = 0.8 \times 0.0001 + 0.1 \times 0.9999 = 0.$$

$$P(\text{SARS} \mid \text{Febbre}) = \frac{P(\text{Febbre} \mid \text{SARS}) \times P(\text{SARS})}{P(\text{Febbre})} = 0.000799$$

ESERCIZIO 3 (3 punti)

Enunciare la definizione di stimatore consistente (in senso forte o in media quadratica) per un parametro θ .

Uno stimatore $T_n=T(X_1,\dots,X_n)$ si dice consistente in media quadratica (o in senso forte) per un parametro $\theta \in \Theta$ se

$$\lim_{n \rightarrow \infty} E\left[(T_n - \theta)^2\right] = 0 \quad \forall \theta$$

ESERCIZIO 4 (8 punti)

Uno psichiatra sospetta che anche piccole dosi di caffeina abbiano un effetto stimolante sul sistema nervoso tale da modificarne il tempo di reazione. Decide di verificare tale sospetto, tramite il test ABC, un test di reazione del sistema nervoso.

Sia Y la variabile aleatoria "tempo di reazione al test ABC senza assunzione di farmaci", distribuita in accordo ad una Gaussiana con $E(Y)=\mu_Y$, parametro non noto e $\text{Var}(Y)=\sigma_Y^2=0.08$.

Sia, invece, X la variabile aleatoria "tempo di reazione al test ABC sotto effetto della caffeina", distribuita sempre in accordo ad una Gaussiana con $E(X)=\mu_X$, parametro non noto e $\text{Var}(X)=\sigma_X^2=0.05$.

Un campione di 64 soggetti estratti casualmente è stato sottoposto al test ABC senza aver assunto alcun farmaco nelle ultime 24 ore, ottenendo il seguente risultato:

$$\sum_{i=1}^{64} y_i = 9.6$$

Un altro campione di 81 soggetti estratti casualmente è stato sottoposto al test ABC dopo aver assunto una piccola dose di caffeina ottenendo il seguente risultato:

$$\sum_{i=1}^{81} x_i = 9.72$$

- 1) Proporre un opportuno stimatore non distorto per $\mu_X - \mu_Y$, indicarne la stima e specificare la distribuzione. (3 punti)

stimatore non distorto per $\mu_X - \mu_Y$: $\bar{X} - \bar{Y}$

stima = $0.12 - 0.15 = -0.03$

$$\bar{X} - \bar{Y} \sim N\left(\mu_X - \mu_Y, \frac{\sigma_X^2}{N_X} + \frac{\sigma_Y^2}{N_Y}\right)$$

Si intende verificare al livello di significatività $\alpha=0.05$ l'ipotesi nulla $H_0: \mu_X = \mu_Y$ contro $H_1: \mu_X \neq \mu_Y$.

- 2) Sotto quali ipotesi riuscite a scrivere una regione di rifiuto per questo problema? (2 punti)

Le due popolazioni devono essere indipendenti

- 3) Scrivere la regione di rifiuto specificando le quantità coinvolte. (2 punti)

$$R = \left\{ |\bar{x} - \bar{y}| \geq z_{1-\alpha/2} \sqrt{\frac{\sigma_X^2}{N_X} + \frac{\sigma_Y^2}{N_Y}} \right\}$$

4) Decidere se accettare o rifiutare l'ipotesi nulla (**1 punto**)

$$R = \left\{ |\bar{x} - \bar{y}| \geq z_{1-\alpha/2} \sqrt{\frac{s_X^2}{N_X} + \frac{s_Y^2}{N_Y}} \right\}$$

$$|\bar{x} - \bar{y}| = 0.03$$

$$z_{1-\alpha/2} = 1.96$$

$$\frac{s_X^2}{N_X} + \frac{s_Y^2}{N_Y} = \frac{0.05}{81} + \frac{0.08}{64} = 0.000617 + 0.00125 = 0.001867$$

$$\sqrt{\frac{s_X^2}{N_X} + \frac{s_Y^2}{N_Y}} = 0.0432$$

$$0.03 < 0.08467$$

Accetto l'ipotesi nulla: la caffeina non modifica il tempo di reazione al test ABC