

SOLUZIONI
PROVA SCRITTA DI STATISTICA (COD 4038)
27 Marzo 2003
MODALITÀ A

APPROSSIMARE TUTTI I CALCOLI ALLA QUARTA CIFRA DECIMALE

ESERCIZIO 1 (8 punti)

Il quotidiano "Repubblica" ha pubblicato i 10 libri più venduti nella settimana (dal 3 al 9 Febbraio).
 Nella seguente tabella, per ciascun libro è indicato:

TITOLO

AUTORE

EDITORE

PREZZO

PAGINE

numero pagine

GEN

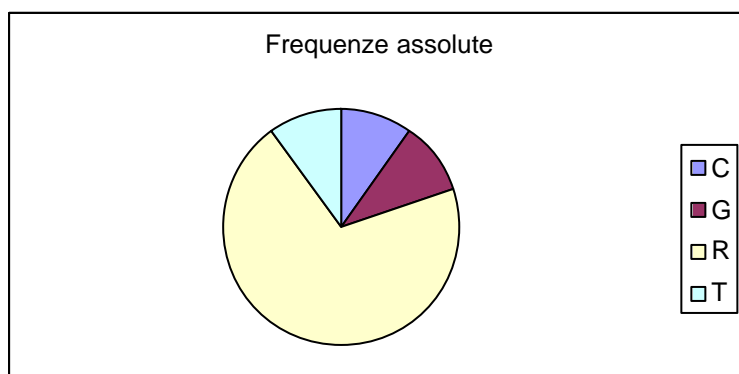
Genere del libro (C=comico, G=giallo, R=romanzo, T=thriller)

TITOLO	Autore	Editore	Prezzo	Pagine	Gen	Prezzo ²	Pagine ²
Il volo del calabrone	Ken Follett	Mondadori	19.00	444	G	361	197136
Io uccido	G. Faletti	Baldini	17.20	681	T	295.84	463761
Famiglia	Mario Puzo	Sonzogno	17.50	393	R	306.25	154449
La principessa sul pisello	L. Littizzetto	Mondadori	13.00	240	C	169	57600
Buick 8	Stephen King	Sperling & K.	18.00	476	R	324	226576
La città delle bestie	I. Allende	Feltrinelli	14.00	224	R	196	50176
L'amore dura tre anni	F. Beigbeder	Feltrinelli	8.00	160	R	64	25600
Non ti muovere	M. Mazzantini	Mondadori	16.50	304	R	272.25	92416
Harry Potter e il prigioniero di Azkaban	J.K. Rowling	Salani	15.00	376	R	225	141376
Il sogno più dolce	Doris Lesing	Feltrinelli	18.00	456	R	324	207936
SOMMA			156.2	3754		2537.34	1617026

1) Fornire un'opportuna rappresentazione grafica del carattere GEN (1 punti)

DISTRIBUZIONE DI
FREQUENZE

Gen	Frequenze assolute	Frequenze relative
C	1	10.00%
G	1	10.00%
R	7	70.00%
T	1	10.00%
Totale complessivo	10	100.00%



2) Calcolare un'opportuna misura di posizione per descrivere il carattere GEN (1 punti)

MODA del carattere GEN=R

Romanzo è la modalità del carattere GEN con frequenza relativa maggiore

3) Considerando solo i libri editi da Feltrinelli e Mondadori, possiamo affermare che il PREZZO è regressivamente indipendente dall'EDITORE? Giustificare la risposta (2 punti)

$$m(\text{Prezzo} | \text{Editore} = \text{Mondadori}) = \frac{19 + 13 + 16.50}{3} = 16.1\bar{6}$$

$$m(\text{Prezzo} | \text{Editore} = \text{Feltrinelli}) = \frac{14 + 8 + 18}{3} = 13.\bar{3}$$

PREZZO non è regressivamente indipendente da EDITORE perché le medie condizionate sono diverse.

4) Quale tra i due caratteri PREZZO e PAGINE presenta maggiore variabilità? (2 punti)

$$m_{\text{PAGINE}} = 375.4 \quad S_{\text{PAGINE}}^2 = 161702.6 - (375.4)^2 = 20777.44$$

$$m_{\text{PREZZO}} = 15.62 \quad S_{\text{PREZZO}}^2 = 253.734 - (15.62)^2 = 9.7496$$

$$CV_{\text{PREZZO}} = \frac{\sqrt{S_{\text{PREZZO}}^2}}{m_{\text{PREZZO}}} = 0.1999 \quad CV_{\text{PAGINE}} = \frac{\sqrt{S_{\text{PAGINE}}^2}}{m_{\text{PAGINE}}} = 0.3840$$

Il carattere PAGINE è più variabile del carattere PREZZO

5) Prevedere il prezzo di un nuovo libro di 500 pagine che uscirà tra breve, sulla base della relazione lineare tra i caratteri PREZZO e PAGINE presenti nella precedente tabella. (la covarianza tra PREZZO e PAGINE risulta 335.722) (2 punti)

$$\text{COV}(\text{Prezzo}, \text{Pagine}) = 335.722$$

$$S_{\text{Pagine}}^2 = 20777.44$$

$$b = \text{COV}(\text{Prezzo}, \text{Pagine}) / \text{VAR}(\text{Pagine}) = 0.0162$$

$$a = m_{\text{PREZZO}} - m_{\text{Pagine}} * b = 15.62 - 0.0162 * 375.4 = 15.62 - 6.0815 = 9.5385$$

$$\text{PREZZO} = 9.5385 + 0.0162 * \text{PAGINE}$$

$$\text{Prezzo}(\text{di un libro di 500 pagine}) = 17.64 \text{ euro}$$

ESERCIZIO 2 (3 punti)

Siano X e Y due variabili statistiche, rispondere con VERO o FALSO alle seguenti affermazioni:

a) Se X è regressivamente indipendente da Y, allora Y è regressivamente indipendente da X.

FALSO

b) $\rho(X, Y) = 0$ implica $f^*(X, Y) < 1$

FALSO

c) $f^*(X, Y) = 0$ implica $\rho(X, Y) = 0$

VERO

ESERCIZIO 3 (4 punti)

Sia X una variabile aleatoria con funzione di ripartizione $F(x)$ data dalla seguente espressione:

$$F(x) = \begin{cases} 0 & x < 2 \\ \frac{x-2}{3} & 2 \leq x < 5 \\ 1 & x \geq 5 \end{cases}$$

a) Calcolare la funzione di densità di probabilità della variabile aleatoria X . **(2 punti)**

$$f(x) = \begin{cases} 0 & x < 2 \\ \frac{1}{3} & 2 \leq x < 5 \\ 0 & x \geq 5 \end{cases}$$

b) Calcolare il valore atteso, $E(X)$ **(2 punti)**

$$E(X)=3.5$$

ESERCIZIO 4 (4 punti)

Sia X una variabile aleatoria con $E(X)=\mu+3$ e $V(X)=1$.

Dato un campione Bernoulliano di ampiezza n :

a) proporre uno stimatore non distorto per μ ; **(2 punti)**

$$T(X_1, \dots, X_n) = \bar{X} - 3$$

b) calcolare la varianza dello stimatore trovato al punto precedente. **(2 punti)**

$$\text{Var}(T(X_1, \dots, X_n)) = \frac{1}{n}$$

ESERCIZIO 5 (8 punti)

Il direttore dell'Ospedale XYZ, situato in un quartiere molto povero nei sobborghi di New York, sospetta che i neonati che nascono lì abbiano un peso inferiore rispetto alla media nazionale (pari a 3.2 Kg), tale da dover richiedere un intervento di prevenzione sulla malnutrizione delle donne del quartiere.

Possiamo supporre che la variabile aleatoria X ="peso alla nascita" si distribuisca (nel quartiere) come una normale di valore atteso non noto μ e varianza non nota σ^2 . Il direttore decide allora di verificare l'ipotesi $H_0: \mu=3.2$ in alternativa all'ipotesi $H_1: \mu < 3.2$, misura quindi il peso $(x_1, x_2, \dots, x_{81})$ di 81 bambini scelti casualmente tra i neonati nell'ultimo anno e riscontra che

$$\sum_{i=1}^{81} x_i = 234.9 \quad \sum_{i=1}^{81} x_i^2 = 1001.21$$

- 1) Proporre un opportuno stimatore non distorto per il parametro σ^2 e fornirne una stima. **(2 punti)**
- 2) Calcolare l'intervallo di confidenza a livello $1-\alpha=0.9$ per il parametro μ . **(2 punti)**

3) Scrivere l'espressione analitica del p-value corrispondente alla realizzazione campionaria osservata e calcolarne il valore. Decidere inoltre se il direttore riterrà necessario un intervento di sanità pubblica contro la malnutrizione, avendo fissato α , il livello di significatività del test uguale a 0.06 (motivare la risposta). **(4 punti)**

1)

$$s_c^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{80} \sum_{i=1}^{81} (X_i - \bar{X})^2$$

$$\bar{x} = 2.9$$

$$s^2 = \frac{1}{81} \sum_{i=1}^{81} x_i^2 - \left(\frac{1}{81} \sum_{i=1}^{81} x_i \right)^2 = \frac{1001.21}{81} - \left(\frac{234.9}{81} \right)^2 = 12.3606 - 2.9^2 = 12.3606 - 8.41 = 3.9506$$

$$s_c^2 = \frac{n}{n-1} s^2 = \frac{81}{80} s^2 = 4$$

2) Poiché n è sufficientemente grande la distribuzione T di student tende alla distribuzione Normale

$$0.90\% CI = \left[\bar{x} - z_{1-\alpha/2} \frac{s_c}{\sqrt{n}}; \bar{x} + z_{1-\alpha/2} \frac{s_c}{\sqrt{n}} \right] = \left[2.9 - 1.645 \frac{2}{9}; 2.9 + 1.645 \frac{2}{9} \right] = [2.5344, 3.2656]$$

3)

$$p\text{-value} = P(\bar{X} < \bar{x} \mid \mathbf{m} = 3.2) = P(\bar{X} < 2.9 \mid \mathbf{m} = 3.2) = P\left(\frac{\bar{X} - \mathbf{m}}{\frac{s_c}{\sqrt{n}}} < \frac{2.9 - 3.2}{\frac{s_c}{\sqrt{n}}} \right) = P\left(T^{n-1} < \frac{2.9 - 3.2}{\frac{s_c}{\sqrt{n}}} \right) =$$

$$= P\left(Z < \frac{2.9 - 3.2}{\frac{s_c}{\sqrt{n}}} \right) = P\left(Z < \frac{2.9 - 3.2}{\frac{2}{9}} \right) = P(Z < -1.35) = 1 - 0.9115 = 0.0885$$

Decisione

Accettiamo l'ipotesi nulla, poiché il p-value è maggiore di α . Non si ritiene necessario un intervento di prevenzione sulla malnutrizione delle donne del quartiere.