SOLUZIONI PROVA SCRITTA DI STATISTICA (cod. 4038, 5047, 371, 377)

8 settembre 2005

MODALITÀ B

APPROSSIMARE TUTTI I CALCOLI ALLA QUARTA CIFRA DECIMALE

Esercizio 1. (7 punti)

Su un collettivo di 13 nuclei familiari sono stati rilevati i seguenti caratteri

X : Stato civile del capofamiglia (C=CONIUGATO, CE=CELIBE/NUBILE, D=DIVORZIATO)
Y : Numero di bambini

I dati sono contenuti nella seguente tabella

Nucleo	1	2	3	4	5	6	7	8	9	10	11	12	13
Y	3	3	4	2	1	4	1	0	5	3	2	3	0
X	C	C	C	C	CE	C	D	CE	C	C	C	C	D

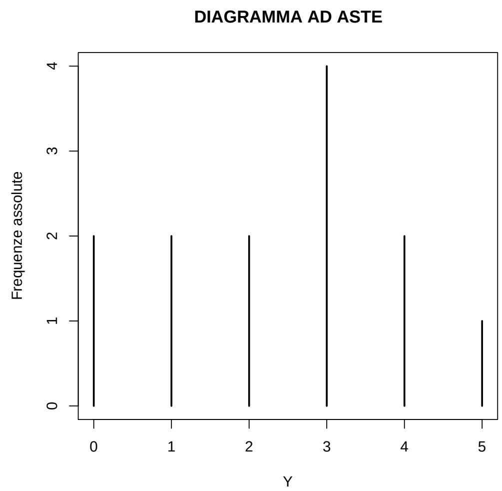
a) Si costruisca la distribuzione di frequenze assolute del carattere Y. (**1 punto**)

Il carattere Y è quantitativo discreto. La distribuzione di frequenze assolute è la seguente

Y	n_i
0	2
1	2
2	2
3	4
4	2
5	1
Totale	13

b) Si fornisca un'opportuna rappresentazione grafica della distribuzione del carattere Y. (**1 punto**)

La distribuzione di Y può essere rappresentata mediante un diagramma ad aste.



c) Si determinino la mediana e la media di Y. Si utilizzino i valori trovati per trarre informazioni sulla forma della distribuzione del carattere in esame. (**2 punti**)

La modalità di Y che occupa la posizione mediana $(N+1)/2 = 7$ nell'ordinamento crescente dei dati è 3. Dunque, $Me(Y) = 3$. La media è

$$\mu = \frac{0 \times 2 + 1 \times 2 + 2 \times 2 + 3 \times 4 + 4 \times 2 + 5 \times 1}{13} = 2,3846.$$

La distribuzione di Y è asimmetrica, obliqua a sinistra. Lo si può dedurre dall'osservazione del grafico precedente ed è confermato dal fatto che $\mu < Me(X)$.

d) Indicato con $Me(Y)$ il valore della mediana di Y, si completi con i valori numerici delle frequenze assolute congiunte la seguente tabella a doppia entrata riportandola sul foglio protocollo. (**2 punti**)

Y / X	$0 \leq y \leq Me(Y)$	$y > Me(Y)$	
C	.	.	.
CE	.	.	.
D	.	.	.
	.	.	13

La tabella richiesta è

Y / X	$0 \leq y \leq 3$	$y > 3$	
C	6	3	9
CE	2	0	2
D	2	0	2
	10	3	13

e) Si consideri la tabella a doppia entrata ottenuta al punto d). Lasciando fissa una delle due distribuzioni marginali, si costruisca una tabella di massima connessione unilaterale di Y con X. (**1 punto**)

Una tabella di massima connessione unilaterale di Y con X si ottiene, ad esempio, lasciando fissa la marginale di X

Y / X	$0 \leq y \leq 3$	$y > 3$	
C	0	9	9
CE	2	0	2
D	2	0	2
	4	9	13

Esercizio 2. (4 punti)

Lo stipendio lordo base annuo dei dipendenti di un certo livello professionale è una variabile casuale X con valore atteso pari a 40.000 Euro e varianza pari a 1.000 Euro^2.

a) Tutti i dipendenti devono pagare una quota fissa di 200 Euro per un fondo previdenziale e a quello che rimane viene tolto il 20% per le tasse. Sia Y la v.c. "stipendio al netto del fondo previdenziale e delle tasse". Risulta $Y = \frac{4}{5}X - 160$. Si determinino il valore atteso e la varianza di Y. (**2 punti**)

Risultano
$$\mathbb{E}[Y] = \mathbb{E}\left[\frac{4}{5}X - 160\right] = \frac{4}{5}\mathbb{E}[X] - 160 = \frac{4}{5} \times 40.000 - 160 = 31.840$$
e
$$\mathbb{V}\text{ar}[Y] = \mathbb{V}\text{ar}\left[\frac{4}{5}X - 160\right] = \frac{16}{25}\mathbb{V}\text{ar}[X] = \frac{16}{25} \times 1.000 = 640.$$

b) Fornire una limitazione superiore per la probabilità di osservare un valore di Y inferiore a 31.640 Euro o superiore a 32.040 Euro. (**2 punti**)

Utilizzando la diseguaglianza di Chebychev,
$$\begin{aligned}
\mathbb{P}\left((Y < 31.640) \cup (Y > 32.040)\right) &= 1 - \mathbb{P}\left(31.640 \leq Y \leq 32.040\right) \\
&= 1 - \mathbb{P}\left(31.840 - 200 \leq Y \leq 31.840 + 200\right) \\
&= 1 - \mathbb{P}(|Y - \mathbb{E}[Y]| \leq 200) \\
&= \mathbb{P}(|Y - \mathbb{E}[Y]| > 200) \\
&< \frac{1}{200^2}\mathbb{V}\text{ar}[Y] = \frac{640}{200^2} = 0,016.
\end{aligned}$$

Esercizio 3. (2 punti)

Siano A e B due eventi definiti su uno spazio Ω.

a) Sapendo che $\mathbb{P}(B|A) = 0,3$ e $\mathbb{P}(A \cap B) = 0,21$, si determini la probabilità $\mathbb{P}(A)$. (**1 punto**)

Dalla definizione di probabilità condizionata di B dato A, ossia $\mathbb{P}(B|A) = \mathbb{P}(A \cap B)/\mathbb{P}(A)$, si ricava
$$\mathbb{P}(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B|A)} = \frac{0,21}{0,3} = 0,7.$$

b) Dimostrare che se A e B sono due eventi indipendenti allora anche $\overline{A}$ e B sono indipendenti. (**1 punto**)

Per ipotesi A e B sono eventi indipendenti, ossia $\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B)$. È sufficiente mostrare che anche $\mathbb{P}(\overline{A} \cap B) = \mathbb{P}(\overline{A}) \times \mathbb{P}(B)$. Dalla relazione $\mathbb{P}(B) = \mathbb{P}(B \cap A) + \mathbb{P}(B \cap \overline{A})$ si ricava
$$\begin{aligned}
\mathbb{P}(\overline{A} \cap B) &= \mathbb{P}(B \cap \overline{A}) \\
&= \mathbb{P}(B) - \mathbb{P}(B \cap A) \\
&= \mathbb{P}(B) - \mathbb{P}(B) \times \mathbb{P}(A) \qquad \text{[Per l'indipendenza tra } A \text{ e } B] \\
&= \mathbb{P}(B)\,[1 - \mathbb{P}(A)] \\
&= \mathbb{P}(B) \times \mathbb{P}(\overline{A}) \\
&= \mathbb{P}(\overline{A}) \times \mathbb{P}(B).
\end{aligned}$$

Esercizio 4. (3 punti)

Siano X ed Y due variabili casuali discrete con la seguente distribuzione congiunta di probabilità

X \ Y	0	2
1	0,3	0,1
2	0,2	0,4

a) Stabilire se X ed Y sono probabilisticamente indipendenti. (**1 punto**)

Le variabili casuali X ed Y non sono indipendenti in quanto la condizione

$$p_{ij} = p_{i\cdot} \times p_{\cdot j}, \quad \forall\,(i,\, j)$$

non è verificata.

b) Determinare la distribuzione di probabilità della variabile casuale quoziente $Z = Y \div X$. (**1 punto**)

La variabile casuale Z ha la seguente distribuzione di probabilità

z	0	1	2
$\mathbb{P}(Z = z)$	0,5	0,4	0,1

c) Calcolare il valore atteso di Z. (**1 punto**)

Si ha $\mathbb{E}[Z] = 0 \times 0,5 + 1 \times 0,4 + 2 \times 0,1 = 0,6$.

Esercizio 5. (4 punti)

Sia X una variabile casuale con valore atteso $\mu > 0$ non noto e varianza σ^2 nota. Si vuole stimare il parametro μ sulla base di un campione bernoulliano $(X_1,\, \ldots,\, X_n)$ di numerosità $n = 5$. Vengono proposti i due stimatori

$$T_1 = \frac{2X_1 + 3X_3 + 2X_5}{5}, \qquad T_2 = \frac{X_1 + X_3 + X_5}{3}.$$

a) Si stabilisca se ciascuno dei due stimatori T_1 e T_2 è non distorto per μ. (**2 punti**)

Lo stimatore T_1 è distorto per μ, mentre T_2 è non distorto. Si ha infatti

$$\begin{aligned} \mathbb{E}[T_1] &= \frac{1}{5}\left(2\mathbb{E}[X_1] + 3\mathbb{E}[X_3] + 2\mathbb{E}[X_5]\right) = \frac{7}{5}\mu \neq \mu, \qquad \forall\, \mu > 0, \\ \mathbb{E}[T_2] &= \frac{1}{3}\left(\mathbb{E}[X_1] + \mathbb{E}[X_3] + \mathbb{E}[X_5]\right) = \mu, \qquad \forall\, \mu > 0. \end{aligned}$$

b) Si stabilisca quale tra T_1 e T_2 è preferibile utilizzare adottando il criterio dell'errore quadratico medio. (**2 punti**)

Lo stimatore T_2 è preferibile a T_1 in quanto

$$EQM[T_2] < EQM[T_1], \qquad \forall\, \mu > 0,$$

l'errore quadratico medio associato a T_2 è cioè minore di quello associato a T_1 quale che sia $\mu > 0$. Indicata con $D[T_i] = \mathbb{E}[T_i] - \mu$ la distorsione di T_i, per $i = 1,\, 2$, risulta infatti

$$\begin{aligned} EQM[T_1] &= \mathbb{V}\mathrm{ar}[T_1] + (D[T_1])^2 = \frac{17}{25}\sigma^2 + \left(\frac{7}{5}\mu - \mu\right)^2 = \frac{17}{25}\sigma^2 + \frac{4}{25}\mu^2, \\ EQM[T_2] &= \mathbb{V}\mathrm{ar}[T_2] = \frac{\sigma^2}{3}. \end{aligned}$$

Esercizio 6. (7 punti)

Si ipotizza che il salario X di una categoria di metalmeccanici si distribuisca secondo una v.c. gaussiana, $X \sim N(\mu, \sigma^2)$, con σ^2 noto. L'intervallo di confidenza per μ di livello $1-\alpha = 0,95$ basato sull'osservazione di un campione bernoulliano $(x_1, \ldots, x_n)$ di $n = 9$ dipendenti è $ic_{0,95}(\mu) = (1.478, 24,\ \ 1.501, 76)$.

a) Si determinino i valori di $\overline{x}_9$ e σ. (**3 punti**)

Poiché l'intervallo di confidenza è centrato in $\overline{x}_9$,

$$\overline{x}_9 = \frac{1.478, 24 + 1.501, 76}{2} = 1.490.$$

Dalla relazione $L[ic_{1-\alpha}(\mu)] = 2z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$, si ricava

$$\sigma = \frac{\sqrt{n}L[ic_{1-\alpha}(\mu)]}{2z_{1-\alpha/2}}.$$

Essendo nel caso specifico $z_{1-\alpha/2} = z_{0,975} = 1,96$, risulta

$$\sigma = \frac{\sqrt{9}\,(1.501, 76 - 1.478, 24)}{2 \times 1,96} = 18.$$

b) Sulla base di un campione bernoulliano $(X_1, \ldots, X_n)$ di numerosità n tratto da X si vuole sottoporre a verifica l'ipotesi che il salario medio sia superiore ad un certo valore μ_0. Si scriva l'espressione generale della regione critica di ampiezza α per il problema di verifica d'ipotesi $H_0 : \mu > \mu_0$ *vs* $H_1 : \mu \leq \mu_0$. (**1 punto**)

L'espressione generale della regione di rifiuto è

$$R = \left\{(x_1, \ldots, x_n) : \overline{x}_n \leq \mu_0 - \frac{\sigma}{\sqrt{n}} z_{1-\alpha}\right\}.$$

c) Sulla base della realizzazione campionaria osservata con $n = 9$ di cui sopra, si dica se si accetta o si rifiuta H_0 per $\mu_0 = 1.500$ ed $\alpha = 0,05$. (Nel caso non si fosse risolto il punto a), si utilizzino i valori $\overline{x}_9 = 1.495$ e $\sigma = 15$). (**2 punti**)

Nel caso specifico, $z_{1-\alpha} = z_{0,95} = 1,645$ e

$$R = \left\{(x_1, \ldots, x_9) : \overline{x}_9 \leq 1.500 - \frac{18}{\sqrt{9}}\,1,645\right\} = \{(x_1, \ldots, x_9) : \overline{x}_9 \leq 1.490, 13\}.$$

Poiché $\overline{x}_9 = 1.490 < 1.490, 13$, la realizzazione campionaria osservata appartiene alla regione di rifiuto, dunque, non si accetta l'ipotesi nulla.
(Se si sono utilizzati i valori $\overline{x}_9 = 1.495$ e $\sigma = 15$, allora $R = \{(x_1, \ldots, x_9) : \overline{x}_9 \leq 1.491, 775\}$. Essendo $\overline{x}_9 = 1.495 > 1.491, 775$, non si rifiuta l'ipotesi nulla).

d) Si calcoli la probabilità di commettere un errore di seconda specie per $\mu_1 = 1.498$. (Nel caso non si fosse risolto il punto a), si utilizzi il valore $\sigma = 15$). (**1 punto**)

La probabilità di commettere un errore di seconda specie è, per definizione, la probabilità di accettare H_0 quando H_1 è vera, diciamo $\mu = \mu_1$. In generale si ha

$$\begin{aligned}
\mathbb{P}_{\mu_1}((X_1, \ldots, X_n) \in A) &= \mathbb{P}_{\mu_1}\left(\overline{X}_n > \mu_0 - \frac{\sigma}{\sqrt{n}} z_{1-\alpha}\right) \\
&= \mathbb{P}_{\mu_1}\left(\frac{\overline{X}_n - \mu_1}{\sigma/\sqrt{n}} > \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} - z_{1-\alpha}\right) \\
&= \mathbb{P}\left(Z > \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} - z_{1-\alpha}\right).
\end{aligned}$$

Per $\mu_1 = 1.498$,

$$\mathbb{P}\left(Z > \frac{1.500 - 1.498}{18/\sqrt{9}} - 1,645\right) = \mathbb{P}\left(Z > -1,31\right) = 0,9049.$$

(Se si è utilizzato il valore $\sigma = 15$, allora $\mathbb{P}\left(Z > \frac{1.500-1.498}{15/\sqrt{9}} - 1,645\right) = \mathbb{P}\left(Z > -1,24\right) = 0,8925$).