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

8 settembre 2005

MODALITÀ A

APPROSSIMARE TUTTI I CALCOLI ALLA QUARTA CIFRA DECIMALE

Esercizio 1. (7 punti)

Su un collettivo di 15 nuclei familiari sono stati rilevati i seguenti caratteri

X : Numero di bambini

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

I dati sono contenuti nella seguente tabella

Nucleo	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X	1	3	4	2	1	4	0	1	5	3	2	2	0	3	2
Y	C	C	C	C	CE	C	D	CE	C	C	C	C	D	C	CE

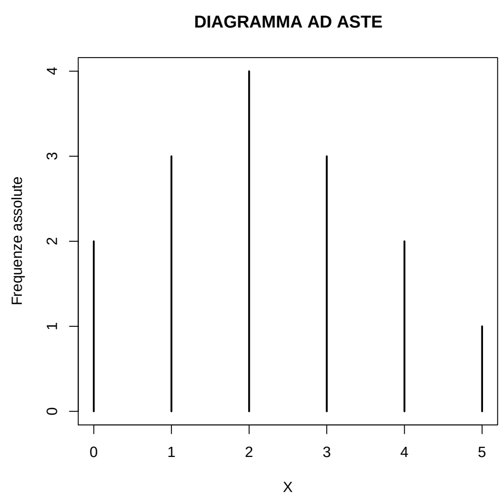
a) Si costruisca la distribuzione di frequenze assolute del carattere X. (**1 punto**)

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

X	n_i
0	2
1	3
2	4
3	3
4	2
5	1
Totale	15

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

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



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

La modalità di X che occupa la posizione mediana $(N+1)/2 = 8$ nell'ordinamento crescente dei dati è 2. Dunque, $Me(X) = 2$. La media è

$$\mu = \frac{0 \times 2 + 1 \times 3 + 2 \times 4 + 3 \times 3 + 4 \times 2 + 5 \times 1}{15} = 2,2.$$

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

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

X \ Y	C	CE	D	
$0 \leq x \leq Me(X)$	·	·	·	·
$x > Me(X)$	·	·	·	·
	·	·	·	15

La tabella richiesta è

X \ Y	C	CE	D	
$0 \leq x \leq 2$	4	3	2	9
$x > 2$	6	0	0	6
	10	3	2	15

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 X con Y. (**1 punto**)

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

X \ Y	C	CE	D	
$0 \leq x \leq 2$	0	3	2	5
$x > 2$	10	0	0	10
	10	3	2	15

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 20.000 Euro e varianza pari a 500 Euro^2.

a) Tutti i dipendenti devono pagare una quota fissa di 100 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 $Y = \frac{4}{5}X - 80$. Si determinino il valore atteso e la varianza di Y. (**2 punti**)

Risultano

$$\mathbb{E}[Y] = \mathbb{E}\left[\frac{4}{5}X - 80\right] = \frac{4}{5}\mathbb{E}[X] - 80 = \frac{4}{5} \times 20.000 - 80 = 15.920$$

e

$$\mathbb{V}\mathrm{ar}[Y] = \mathbb{V}\mathrm{ar}\left[\frac{4}{5}X - 80\right] = \frac{16}{25}\mathbb{V}\mathrm{ar}[X] = \frac{16}{25} \times 500 = 320.$$

b) Fornire una limitazione inferiore per la probabilità di osservare un valore di Y compreso tra 15.820 e 16.020 Euro. (**2 punti**)

Si noti che l'intervallo (15.820, 16.020) è centrato in $15.920 = \mathbb{E}[Y]$. Per la diseguaglianza di Chebychev,

$$\begin{aligned}\mathbb{P}(15.820 < Y < 16.020) &= \mathbb{P}(15.920 - 100 < Y < 15.920 + 100) \\ &= \mathbb{P}(|Y - \mathbb{E}[Y]| < 100) \\ &> 1 - \frac{1}{100^2}\mathbb{V}\mathrm{ar}[Y] = 1 - \frac{320}{100^2} = 0.968.\end{aligned}$$

Esercizio 3. (2 punti)

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

a) È noto che $\mathbb{P}(A) = 0,2$ e $\mathbb{P}(A \cap B) = 0,1$. Si determini la probabilità $\mathbb{P}(B)$ nel caso in cui A e B siano eventi indipendenti. (**1 punto**)

Poiché per ipotesi A e B sono eventi indipendenti,

$$\mathbb{P}(B) = \mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{0,1}{0,2} = 0,5.$$

b) Dimostrare che se A e B sono due eventi indipendenti allora anche A e $\overline{B}$ 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 $\mathbb{P}(A \cap \overline{B}) = \mathbb{P}(A) \times \mathbb{P}(\overline{B})$. Dalla relazione $\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap \overline{B})$ si ricava

$$\begin{aligned}\mathbb{P}(A \cap \overline{B}) &= \mathbb{P}(A) - \mathbb{P}(A \cap B) \\ &= \mathbb{P}(A) - \mathbb{P}(A) \times \mathbb{P}(B) \qquad \text{[Per l'indipendenza tra } A \text{ e } B\text{]} \\ &= \mathbb{P}(A)\,[1 - \mathbb{P}(B)] \\ &= \mathbb{P}(A) \times \mathbb{P}(\overline{B}).\end{aligned}$$

Esercizio 4. (3 punti)

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

X \ Y	0	1	2
0	0,2	0,16	0,04
1	0,3	0,24	0,06

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

Le variabili casuali X ed Y sono indipendenti in quanto per ogni coppia $(i,\ j)$, risulta $p_{ij} = p_{i\cdot} \times p_{\cdot j}$.

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

La variabile casuale Z ha la seguente distribuzione di probabilità

z	0	1	2
$\mathbb{P}(Z=z)$	0,7	0,24	0,06

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

Si ha $\mathbb{E}[Z] = 0 \times 0,7 + 1 \times 0,24 + 2 \times 0,06 = 0,36$.

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 $n = 4$. Vengono proposti i due stimatori

$$T_1 = \frac{X_1 + X_2}{2}, \qquad T_2 = \frac{X_1 + 2X_2 + 3X_3 + X_4}{4}.$$

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

$$\begin{aligned}
\mathbb{E}[T_1] &= \frac{1}{2}\left(\mathbb{E}[X_1] + \mathbb{E}[X_2]\right) = \mu, \qquad \forall\, \mu > 0, \\
\mathbb{E}[T_2] &= \frac{1}{4}\left(\mathbb{E}[X_1] + 2\mathbb{E}[X_2] + 3\mathbb{E}[X_3] + \mathbb{E}[X_4]\right) = \frac{7}{4}\mu \neq \mu, \qquad \forall\, \mu > 0.
\end{aligned}$$

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_1 è preferibile a T_2 in quanto

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

l'errore quadratico medio associato a T_1 è cioè minore di quello associato a T_2 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

$$\begin{aligned}
EQM[T_1] &= \mathbb{V}\text{ar}[T_1] = \frac{\sigma^2}{2}, \\
EQM[T_2] &= \mathbb{V}\text{ar}[T_2] + (D[T_2])^2 = \frac{15}{16}\sigma^2 + \left(\frac{7}{4}\mu - \mu\right)^2 = \frac{15}{16}\sigma^2 + \frac{9}{16}\mu^2.
\end{aligned}$$

Esercizio 6. (7 punti)

Sull'etichetta di un fustino di detersivo si dichiara che il peso in Kg del contenuto è pari a μ_0. Si ipotizza che il peso effettivo X del prodotto 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 $1-\alpha = 0,975$ basato sull'osservazione del contenuto di un campione bernoulliano $(x_1, \ldots, x_n)$ di $n = 50$ fustini è $ic_{0,975}(\mu) = (4,6416, \ \ 4,9584)$.

a) Si determinino i valori di $\overline{x}_{50}$ e σ. (**3 punti**)

Poiché l'intervallo di confidenza è centrato in $\overline{x}_{50}$,

$$\overline{x}_{50} = \frac{4,6416 + 4,9584}{2} = 4,8.$$

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 $z_{1-\alpha/2} = z_{0,9875} = 2,24$, risulta

$$\sigma = \frac{\sqrt{50}\,(4,9584 - 4,6416)}{2 \times 2,24} = 0,5.$$

b) Un'associazione di consumatori ritiene che il contenuto medio effettivo dei fustini sia inferiore al peso dichiarato μ_0. Sia $(X_1, \ldots, X_n)$ un campione bernoulliano di numerosità n tratto da X. Si scriva l'espressione generale della regione critica di ampiezza α per il problema di verifica d'ipotesi $H_0 : \mu < \mu_0$ *vs* $H_1 : \mu \geq \mu_0$. (**1 punto**)

L'espressione generale della regione di rifiuto è

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

c) Sulla base della realizzazione campionaria osservata con $n = 50$ di cui sopra, si dica se si accetta o si rifiuta H_0 per $\mu_0 = 5$ ed $\alpha = 0,025$. (Nel caso non si fosse risolto il punto a), si utilizzino i valori $\overline{x}_{50} = 4,9$ e $\sigma = 0,4$). (**2 punti**)

Nel caso specifico, $z_{1-\alpha} = z_{0,975} = 1,96$ e

$$R = \left\{(x_1, \ldots, x_{50}) : \overline{x}_{50} \geq 5 + \frac{0,5}{\sqrt{50}}\ 1,96\right\} = \{(x_1, \ldots, x_{50}) : \overline{x}_{50} \geq 5,1386\}.$$

Poiché $\overline{x}_{50} = 4,8 < 5,1386$, la realizzazione campionaria osservata non appartiene alla regione di rifiuto, dunque, non si rifiuta l'ipotesi nulla.
(Se si sono utilizzati i valori $\overline{x}_{50} = 4,9$ e $\sigma = 0,4$, allora $R = \{(x_1, \ldots, x_{50}) : \overline{x}_{50} \geq 5,1109\}$. Essendo $\overline{x}_{50} = 4,9 < 5,1109$, non si rifiuta l'ipotesi nulla).

d) Si calcoli la probabilità di commettere un errore di seconda specie per $\mu_1 = 5,2$. (Nel caso non si fosse risolto il punto a), si utilizzi il valore $\sigma = 0,4$). (**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

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

Per $\mu_1 = 5,2$,

$$\Phi\left(\frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} + z_{1-\alpha}\right) = \Phi\left(\frac{5 - 5,2}{0,5/\sqrt{50}} + 1,96\right) = \Phi(-0,87) = 1 - 0,8078 = 0,1922.$$

(Se si è utilizzato il valore $\sigma = 0,4$, allora $\Phi\left(\frac{5-5,2}{0,4/\sqrt{50}} + 1,96\right) = \Phi(-1,57) = 1 - 0,9418 = 0,0582$).