FIRMA DELLO STUDENTE

SECONDA PROVA INTERMEDIA DI STATISTICA

CLEA, CLEFIN, CLAPI (cod. 5047/4038)
CLELI 4038
371/377

14 Gennaio 2004

Cognome ***Nome***

Numero di matricola

COMPITO D1

Ai fini della valutazione si terrà conto solo ed esclusivamente di quanto riportato negli appositi spazi. Al termine della prova, è OBBLIGATORIO consegnare il presente foglio ed il foglio di brutta (DI CUI NON SI TERRÀ CONTO AI FINI DELLA VALUTAZIONE).

APPROSSIMARE TUTTI I CALCOLI ALLA QUARTA CIFRA DECIMALE

1. (2 punti) Sia X una v.a. con funzione di densità

$$f(x) = 4x^3 \qquad 0 \le x \le 1.$$

a) (1 punto) Si calcoli il valore atteso di X.
b) (1 punto) Si scriva l'espressione analitica della funzione di ripartizione di X.

a) $E(X) = \int_0^1 x \cdot 4x^3 dx = 4\left[\frac{x^5}{5}\right]_0^1 = \frac{4}{5}$

b) $F(x) = \begin{cases} 0 & x < 0 \\ x^4 & 0 \le x \le 1 \\ 1 & x > 1 \end{cases}$

2. (3 punti) Si consideri il seguente esperimento. Si lancia due volte una moneta, "testa" vale 1 e "croce" vale 2. Siano X=somma dei punteggi nei due lanci e Y=massimo punteggio sui due lanci.
a) (1 punto) Si scriva la distribuzione di frequenza congiunta di (X,Y).
b) (2 punti) Si calcoli $P(X=4|Y=2)$.

a)

X \ Y	1	2	$p_X(x_i)$
2	1/4		1/4
3		2/4	2/4
4		1/4	1/4
$p_Y(y_j)$	1/4	3/4	1

b) P(X=4|Y=2) = (1/4):(3/4) = 1/3

3. (3 punti)

a) (1 punto) Siano X_i, i=1,…,10, v.a. i.i.d. con $E(X)$=20 e $Var(X)$=5. Si dica quanto valgono il valore atteso e la varianza della media campionaria.

b) (2 punti) Sia (X_1,X_2) un campione bernoulliano da una popolazione con valore atteso q. Sia $T_1=2X_1+X_2/2$ uno stimatore per q. Si corregga la distorsione di T_1 determinando un nuovo stimatore T_2 non distorto.

a) $E(\overline{X}) = 20$; $\mathrm{Var}(\overline{X}) = 5/10 = 0.5$

b) $E(T_1) = 2\vartheta + \dfrac{\vartheta}{2} = \dfrac{5}{2}\vartheta$; $D(T_1) = \dfrac{5}{2}\vartheta - \vartheta = \dfrac{3}{2}\vartheta$

Si ponga allora $T_2 = \dfrac{2}{5}(T_1) = \dfrac{2}{5}(2X_1 + \dfrac{X_2}{2})$

si avrà $E(T_2) = \dfrac{4}{5}\vartheta + \dfrac{2}{10}\vartheta = \vartheta$ (Lo stimatore T_2 è quindi non distorto).

4. (6 punti) Siano X_i, i=1,…,10, i.i.d. $N(\mu_X,\sigma^2_X)$ e Y_j, j=1,…,10, i.i.d. $N(m_Y, s^2_Y)$.

a) (2 punti) Si determini uno stimatore non distorto per $q=m_X+m_Y$.

b) (3 punti) Assumendo di avere un campione di ampiezza 25 da una $N(m,s^2)$ con m=9 e s^2=100, si determini uno stimatore non distorto per m e si calcoli la probabilità che tale stimatore sia minore o uguale a 10.

c) (1punto) Si confronti lo stimatore ottenuto al punto (a) con $T=(X_1+Y_1)/2$ e si dica qual è il più efficiente (si assume X_i indipendente da Y_j per ogni i,j).

a) Uno stimatore non distorto per $q=m_X+m_Y$ è $T = \overline{X}+\overline{Y} = \dfrac{1}{10}\sum_{i=1}^{10} X_i + \dfrac{1}{10}\sum_{j=1}^{10} Y_j$

b) Uno stimatore non distorto per m è $\overline{X} = \dfrac{1}{25}\sum_{i=1}^{25} X_i$.

Poiché $\overline{X} \sim N(9, \dfrac{100}{25})$ si avrà $P(\overline{X} \le 10) = P(Z \le \dfrac{10-9}{2}) = P(Z \le \dfrac{1}{2}) = 0.6915$

c) $E(\overline{X}+\overline{Y}) = \mu_X + \mu_Y$; $\mathrm{Var}(\overline{X}+\overline{Y}) = \dfrac{\sigma^2_X}{10} + \dfrac{\sigma^2_Y}{10}$;

$E(T) = \dfrac{\mu_X+\mu_Y}{2}$; $D(T) = \dfrac{\mu_X+\mu_Y}{2} - (\mu_X+\mu_Y) = -\dfrac{1}{2}(\mu_X+\mu_Y)$; $\mathrm{Var}(T) = \dfrac{1}{4}(\sigma^2_X+\sigma^2_Y)$;

$$EQM(\overline{X}+\overline{Y}) = \mathrm{Var}(\overline{X}+\overline{Y}) = \frac{1}{10}(\sigma^2_X+\sigma^2_Y)$$

$$EQM(T) = \mathrm{Var}(T) + D^2(T) = \frac{1}{4}(\sigma^2_X+\sigma^2_Y) + \frac{1}{4}(\mu_X+\mu_Y)^2$$

Poiché $EQM(\overline{X}+\overline{Y}) < EQM(T)$ lo stimatore $\overline{X}+\overline{Y}$ è più efficiente di T.

5. (6 punti) Sia $X_1,\dots,X_{100}$ un campione da una popolazione con media m e varianza s^2.

a) (2 punti) Si scriva l'espressione analitica dell' intervallo di confidenza asintotico al 95% per m.

b) (2 punti) Assumendo che la popolazione sia $N(m,25)$ e che $\sum_{i=1}^{100} X_i = 500$, si costruisca un intervallo di confidenza al 95% per m.

c) (2 punti) Se l'ampiezza del campione fosse 225, quale dovrebbe essere il valore di α affinché l'intervallo di livello $(1-\alpha)$ per m sia di lunghezza inferiore a 1.5?

a) $\text{I.C.}_{0.95}(\mu) = (\overline{X} - z_{0.975}\frac{S_c}{\sqrt{n}}; \overline{X} + z_{0.975}\frac{S_c}{\sqrt{n}})$ (n = 100)

b) $\text{i.c.}_{0.95}(\mu) = (5 - 1.96\frac{5}{10}; 5 + 1.96\frac{5}{10}) = (4.02;\ 5.98)$

c) $2 \cdot z_{1-\frac{\alpha}{2}} \cdot \frac{5}{15} < 1.5 \quad \Rightarrow \quad z_{1-\frac{\alpha}{2}} < \frac{22.5}{10} \quad \Rightarrow \quad 1 - \frac{\alpha}{2} < 0.9878 \quad \Rightarrow \quad \alpha > 0.0244$

6. (6 punti) Sia $X_1,\dots,X_n$ un campione da $N(m,100)$.

a) (2 punti) Si costruisca un test per $H_0: \mu=5$ contro $H_1: m=7$ con livello di significatività 0.05.

b) (2 punti) Si calcoli la probabilità dell'errore di secondo tipo, sapendo che n=16.

c) (2 punti) Supponendo di aver osservato una media campionaria pari a 10, per quale valore di a si rifiuta l'ipotesi nulla (n = 16)?

a) $\begin{cases} H_0: & \mu = 5 \\ H_1: & \mu = 7 \end{cases};$ $\qquad R = \left\{(x_1, x_2, \dots, x_n): \overline{x} > 5 + 1{,}645 \cdot \frac{10}{\sqrt{n}}\right\};$

b) Poiché $\overline{X} \sim N(\mu, \frac{100}{16})$,

$$\beta = P(\overline{X} \le 5 + 1{,}645 \cdot \frac{10}{4} \mid \mu = 7) = P(\overline{X} \le 9.1125 \mid \mu = 7) =$$

$$= P(Z \le \frac{2.1125}{10/4}) = P(Z \le 0.85) = \Phi(0.85) = 0.8023$$

c) $10 \in R$ se $10 > 5 + z_{1-\alpha} \cdot \frac{10}{4} \quad \Rightarrow \quad z_{1-\alpha} < 2 \quad \Rightarrow \quad 1-\alpha < 0.9772 \quad \Rightarrow \quad \alpha > 0.0228$

7. (4 punti) Si consideri un modello di regressione lineare con Y variabile dipendente e X variabile esplicativa.

a) (2 punti) Si scriva l'equazione del modello lineare semplice e si enuncino le ipotesi deboli.

b) (2 punti) Supponendo di aver osservato

Y	*X*
2	10
3	15
3	20
4	10
2	20

si forniscano le stime dei minimi quadrati dei coefficienti.

a) Secondo il modello di regressione lineare semplice

$$Y_i = \beta_0 + \beta_1 \cdot x_i + \varepsilon_i .$$

Le ipotesi deboli di tale modello sono date da:

i) $E(\varepsilon_i) = 0, \forall i = 1, \ldots, n$

ii) $Var(\varepsilon_i) = \sigma^2, \forall i = 1, \ldots, n$

iii) $Cov(\varepsilon_i, \varepsilon_j) = 0, \forall i \neq j$

b) Poiché

$M(X) = 15;$ $M(Y) = 2.8;$ $M(XY) = 41;$ $M(X^2) = 245;$

$cov(X, Y) = -1;$ $var(X) = 20$

le stime dei minimi quadrati dei coefficienti sono:

$$\widehat{\beta}_0 = 3.55 ; \quad \widehat{\beta}_1 = -0.05$$