SECONDA PROVA INTERMEDIA DI STATISTICA CLEA 5047 – 371- 377-4038 19 gennaio 2005
Cognome ***Nome***
Numero di matricola
COMPITO C1 **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 Siano X e Y due variabili aleatorie indipendenti entrambe distribuite secondo la distribuzione Normale con valore atteso pari a 4 e varianza pari a 4.

a) Calcolare P(2<X<5) **(2 punti)**
b) Calcolare P(X+2Y>15) **(2 punti)**

a)	b)
$P(2 < X < 5) = P\left(\frac{2-4}{2} < X < \frac{5-4}{2}\right)$ $= P(-1 < x < 0.5) = \Phi(0.5) - 1 + \Phi(1) =$ $= 0.6915 - 1 + 0.8413 = 0.5328$	$W = X + 2Y \sim N(12{,}20)$ $P(W > 15) = P\left(\frac{W-12}{\sqrt{20}} > \frac{15-12}{\sqrt{20}}\right) = 1 - \Phi(0.67) = 1 - 0.7486 =$ $= 0.2514$

2. Si consideri il seguente vettore aleatorio

X\Y	**1**	**2**	**3**
1	0	0.1	0.6
2	0	0.15	0
3	0.15	0	0

Calcolare la distribuzione di Z=X+2Y **(2 punti)**

$$Z = \begin{cases} 5 & 6 & 7 \\ 0.25 & 0.15 & 0.6 \end{cases}$$

3. Fornire la definizione di campione bernoulliano. **(1 punto)**

Un campione costituito da *n* variabili aleatorie indipendenti ed identicamente distribuite come la popolazione di partenza.

4. Dato $X_1, X_2, \ldots X_n$ un campione bernoulliano estratto da una popolazione normale con media μ e varianza pari a 1. La seguente funzione campionaria

$$\sum_{1}^{n} X_i^{\,2}$$

è una statistica? Motivare la risposta. (**2 punti**)

Sì perché è una funzione delle n variabili aleatorie che costituiscono il campione e non dipende da quantità non note.

5. Sia X una variabile aleatoria con distribuzione normale di media θ e varianza 4 e sia (X_1, X_2, X_3) un campione bernoulliano estratto da X.

a) Dimostrare, esplicitando il procedimento seguito, che

$$T = \frac{X_1 - X_2}{2} + X_3 + 3$$

è uno stimatore distorto per θ. **(2 punti)**

$$E(T) = E\left(\frac{X_1 - X_2}{2} + X_3 + 3\right) = q + 3 \neq q$$

b) Calcolare l'errore quadratico medio di tale stimatore. **(2 punti)**

$$EQM(T) = V(T) + D^2(T)$$
$$V(T) = V\left(\frac{X_1 - X_2}{2} + X_3 + 3\right) = V\left(\frac{X_1 - X_2}{2}\right) + V(X_3) = \frac{8}{4} + 4 = 6$$
$$D^2(T) = (q + 3 - q)^2 = 9$$
$$EQM(T) = 15$$

c) A partire da T costruire uno stimatore non distorto per θ. **(2 punti)**

$T_1 = (T - 3)$ è uno stimatore non distorto per q

6. Fornire la definizione di intervallo di confidenza per un generico parametro θ. **(2 punti)**

Sia $(X_1, \ldots, X_n)$ un campione casuale estratto da una popolazione X caratterizzata da una funzione di ripartizione, F_θ, indicizzata da un parametro $\theta \in \Theta \subseteq \Re$. Si considerino due statistiche: $T_1 = T_1(X_1, \ldots, X_n)$ e $T_2 = T_2(X_1, \ldots, X_n)$ tali che:

1) $T_1 < T_2$;
2) $P_q(T_1 < q < T_2) = 1 - a \qquad \forall q \in \Theta$

dove $0 < \alpha < 1$ è una costante nota. Allora l'intervallo aleatorio (T_1,T_2) viene detto intervallo di confidenza aleatorio per θ al livello (1- α).

7. Da una popolazione X normale con media μ e varianza 9 viene estratto il seguente campione (1, 4, 5, 10, 30).

a) Scrivere l'espressione dell'intervallo di confidenza generico per μ al 95%. **(1 punto)**

$$\left(\bar{x} - z_{1-\frac{a}{2}} \frac{s}{\sqrt{n}} ; \bar{x} + z_{1-\frac{a}{2}} \frac{s}{\sqrt{n}} \right) = \left(\bar{x} - 1.96 \frac{3}{\sqrt{5}} ; \bar{x} + 1.96 \frac{3}{\sqrt{5}} \right)$$

b) Calcolare gli estremi dell'intervallo di confidenza per μ al 95% in corrispondenza del campione osservato. **(1 punto)**

$$\left(10 - 1.96 \frac{3}{\sqrt{5}} ; 10 + 1.96 \frac{3}{\sqrt{5}} \right) = (7.3704 ; 12.6296)$$

c) Sulla base dei risultati precedenti, calcolare la lunghezza dell'intervallo se il campione osservato fosse stato di ampiezza 20. Considerate nuovamente un livello di confidenza del 95%(**2 punti**).

$l = 2z_{1-\frac{a}{2}} \frac{s}{\sqrt{n}}$ poiché il nuovo campione ha dimensione n_1 quadrupla rispetto a quello precedente

$$l_1 = 2z_{1-\frac{a}{2}} \frac{s}{\sqrt{n_1}} = 2z_{1-\frac{a}{2}} \frac{s}{\sqrt{4n}} = \frac{l}{\sqrt{4}} = \frac{l}{2} = 2.6296$$

8. Sia X~N(μ,σ^2) con entrambi i parametri non noti.

a) Avendo osservato il seguente campione (10, 8, 5, 3) accettereste o rifiutereste ad un livello α=0.1 l'ipotesi H_0: μ=6 contro l'ipotesi alternativa H_1: μ ? 6? (**2 punti**) Scrivere l'espressione della regione di rifiuto e motivare la risposta

$$R = \left\{ \left| \bar{x} - m_0 \right| > t^{n-1}_{1-\frac{a}{2}} \frac{s_C}{\sqrt{n}} \right\}$$

$$\bar{x} = 6.5 \quad \sum x_i^2 = 198 \quad s_C^2 = \frac{198 - 4 \times 6.5^2}{3} = 9.6666 \quad t^3_{0.95} = 2.353 .$$

$$R = \left\{ \overline{x} < 6 - 2.353 \frac{\sqrt{s_C}}{2}; \overline{x} > 6 + 2.353 \frac{\sqrt{s_C}}{2} \right\}$$

in questo caso

$$R = \left\{ \overline{x} < 6 - 2.353 \frac{\sqrt{9.6666}}{2}; \overline{x} > 6 + 2.353 \frac{\sqrt{9.6666}}{2} \right\}$$

$$R = \{ \overline{x} < 2.3421; \overline{x} > 9.6578 \}$$

Poiché $\overline{x} \notin R$ accetto

9. Da una popolazione Bernoulliana di parametro θ viene estratto un campione di ampiezza 4. Al fine di verificare l'ipotesi θ=0.5 contro l'ipotesi alternativa θ=0.8 si considera un test caratterizzato dalla seguente regione di rifiuto

$$R = \left\{ (x_1, x_2, x_3, x_4) : \sum_{i=1}^{4} x_i \geq 3 \right\}$$

a) Calcolare la probabilità di errore di seconda specie. (**2 punti**)

$$b = P(A \mid H_1) = P\left(\sum X_i < 3 \mid q = 0.8\right) = 1 - P\left(\sum X_i = 3 \mid q = 0.8\right) - P\left(\sum X_i = 4 \mid q = 0.8\right) =$$
$$= 1 - 4 \times 0.8^3 \times 0.2 - 0.8^4 = 0.1808$$

b) Se il campione osservato fosse (1, 1, 0, 1) quale sarebbe la vostra decisione? Motivare la risposta. **(1 punto)**

Rifiuto poiché $\sum_{1}^{4} x_i = 3$

10 Siano X e Y due variabili aleatorie continue la cui relazione può essere spiegata attraverso un modello lineare del tipo Y= $\beta_0 + \beta_1 x + \varepsilon$

Scrivere l'espressione degli stimatori degli stimatori dei minimi quadrati di β_0 e β_1. (**3 punti**)

$$\hat{b}_1 = \frac{\sum (x_i - \overline{x})(Y_i - \overline{Y})}{\sum (x_i - \overline{x})^2} = \frac{\sum (x_i - \overline{x}) Y_i}{\sum (x_i - \overline{x})^2}$$

$$\hat{b}_0 = \overline{Y} - \hat{b}_1 \overline{x}$$

SECONDA PROVA INTERMEDIA DI STATISTICA
CLEA 5047 – 371- 377-4038
19 gennaio 2005

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).

1. Sia data la seguente funzione:

$$f(x) = \begin{cases} \frac{k}{3}(x-1)^2 & 0 < \mathrm{x} < 2 \\ 0 & \text{altrove} \end{cases}$$

Si determini il valore di k in modo tale che $f(x)$ sia una funzione di densità di probabilità (**2punti**).

1) $f_X \geq 0$

$k \geq 0$

2) $\int_{-\infty}^{+\infty} \mathrm{f}_X(x)dx = 1 \qquad \int_0^2 \frac{k}{3}(x-1)^2 dx = \frac{k}{3}\left|\frac{x^3}{3} + x - x^2\right|_0^2 = \frac{k}{3}\left(\frac{2}{3}\right)$

$k = \frac{9}{2}$

2. Una moneta regolare viene lanciata due volte. Ad ogni testa viene associato un punteggio pari a 1 mentre ad ogni croce un punteggio pari a 0. Indicata con X la variabile aleatoria somma dei punteggi e con Y la variabile aleatoria prodotto dei punteggi calcolare

a) la distribuzione marginale di X. **(2 punti)**

$$X = \begin{cases} 0 & 1 & 2 \\ \frac{1}{4} & \frac{2}{4} & \frac{1}{4} \end{cases}$$

b) la distribuzione congiunta di X e Y. **(2 punti)**

X\Y	**0**	**1**
0	1/4	0
1	2/4	0
2	0	1/4

3. Sia X una Normale di media 3 e varianza 1. Si consideri un campione bernoulliano di ampiezza *10* estratto da questa popolazione. Determinare la distribuzione di probabilità di $S_{10} = \sum_{1}^{10} X_i$.**(1 punto)**

S_{10}~N(30,10)

4. Dato un campione bernoulliano di ampiezza *n* estratto da una popolazione X, definire il momento campionario di ordine 3. (**1 punto**).

$$M_{2,n} = \frac{1}{n}\sum_{i=1}^{n} X_i^3$$

5. Da una popolazione X ~ Uniforme (θ-2,θ+5) viene estratto un campione bernoulliano di ampiezza *n*.

a) Proporre uno stimatore non distorto per E(X). (**2 punti**)

Lo stimatore cercato è $T = \overline{X}$

b) Proporre uno stimatore non distorto per θ. (**2 punti**)

$E(T) = E(\overline{X}) = \frac{2q+3}{2}$ di conseguenza lo stimatore non distorto per θ è $T^1 = \frac{2T-3}{2} = \frac{2\overline{X}-3}{2}$

6. Fornire la definizione di stimatore consistente in media quadratica. (**2 punti**)

Sia $T_1, T_2, \ldots, T_n, \ldots$ una successione di stimatori per q, dove $T_n = T_n(X_1,\ldots,X_n)$.

Si dice che la successione, o più brevemente T_n, è **consistente in media quadratica** per θ se:

$$\lim_{n\to\infty} E_q\,(T_n - q)^2 = 0 \qquad \forall q \in \Theta$$

7. Si consideri il campione (2, 5, 8, 5) estratto da una popolazione X distribuita secondo una legge normale di media μ e varianza σ^2 entrambe non note.

a) Costruire un intervallo di confidenza al 99% per la media μ. (**2 punti**)

$$\bar{x} = 5 \quad \sum x_i^2 = 118 \quad s_C^2 = \frac{118 - 4 \times 5^2}{3} = 6$$

$$\left(\bar{x} - t^3{}_{0.995} \frac{s_C}{\sqrt{n}} ; \bar{x} + t^3{}_{0.995} \frac{s_C}{\sqrt{n}} \right) = \left(5 - 5.841 \frac{\sqrt{6}}{2} ; 5 + 5.841 \frac{\sqrt{6}}{2} \right) = (-2.1537;\ 12.1537)$$

b) Sulla base del campione osservato accettereste o rifiutereste l'ipotesi $H_0:\mu=8$ contro l'ipotesi $H_1:\mu$? 8 al livello $\alpha=0.01$? Motivare la risposta. (**2 punti**).

Accetto poiché il valore 8 appartiene all'intervallo di confidenza

c) Come cambierebbe la risposta la punto precedente se il livello significatività del test (α) fosse pari a 0.003? Motivare la risposta. (**2 punti**)

Continuerei ad accettare, l'intervallo di confidenza considerato diventerebbe più ampio e conterebbe al suo interno l'intervallo precedente.

8. Si considerino due popolazioni indipendenti, $X\sim N(\mu_1,1)$ e $Y\sim N(\mu_2,8)$. Viene estratto un campione di ampiezza n_1 dalla prima popolazione ed un campione di ampiezza n_2 dalla seconda popolazione. Scrivere l'espressione generica della regione di rifiuto, ad un livello α, per testare l'ipotesi che le due medie siano uguali contro l'ipotesi alternativa che le due medie siano diverse. (**2 punti**)

$$R := \left\{ (x_1, \cdots, x_{n1}), (y_1, \cdots, y_{n2}) : \left| (\bar{x} - \bar{y}) \right| > z_{1-a/2} \sqrt{\frac{1}{n_1} + \frac{8}{n_2}} \right\}$$

9. Una società produttrice di articoli sportivi vuole lanciare un nuovo tipo di sci sul mercato. Al fine di valutare se l'investimento sarà proficuo regala a un campione di 100 sportivi il nuovo tipo di sci. Di questi solo 35 esprimono un parere negativo sul prodotto.

a) Si consideri la proporzione p di sportivi, dell'intera popolazione, che esprimeranno parere negativo sul prodotto. Costruire l'intervallo di confidenza al 95 % per p. (**2 punti**)

$$\left(\bar{x} - z_{1-a/2} \sqrt{\frac{\bar{x}(1-\bar{x})}{n}} ; \quad \bar{x} + z_{1-a/2} \sqrt{\frac{\bar{x}(1-\bar{x})}{n}} \right)$$

$$(0.35 - 1.96 \times 0.0477 ;\ 0.35 + 1.96 \times 0.0477)$$

(0.2565;0.4435)

b) Quale valore dovrebbe assumere α affinché la lunghezza dell'intervallo considerato sia inferiore a 0.1? Scrivere il procedimento seguito (**2 punti**)

$$l = 2z_{1-a/2}\sqrt{\frac{\overline{x}(1-\overline{x})}{n}} < 0.1$$

$$z_{1-a/2} < \frac{0.1}{2\sqrt{\frac{\overline{x}(1-\overline{x})}{n}}} \approx 1.05$$

$$1-\frac{a}{2} < 0.8531$$

$$\frac{a}{2} > 1-0.8532$$

$$a > 0.2938$$

10. Utilizzando il software EXCEL si è studiata la dipendenza lineare tra le variabili X e Y ottenendo i seguenti risultati:

Statistica della regressione	
R multiplo	0,980200427
R al quadrato	0,960792877
R al quadrato corretto	0,955891986
Errore standard	0,782381756
Osservazioni	10

	Coefficienti	*Errore standard*	*Stat t*	*Valore di significatività*
Intercetta	3,466666667	0,534468489	6,486195	0,000190863
Variabile X 1	1,206060606	0,086137396	14,00159	6,56516E-07

a) Valutare sulla base di un opportuno indice la bontà del modello lineare proposto. Motivare la risposta. **(2 punti)**

Il modello presenta un buon adattamento. Poiché R^2=0.9608, il modello di regressione lineare spiega circa il 96% della variabilità totale.

b) Sulla base delle informazioni fornite in tabella è possibile accettare ad un livello α=0.05 l'ipotesi che il coefficiente angolare della retta sia uguale a zero? Motivare la risposta. **(2 punti)**

p-value=6,56516E -07

Poiché p-value<α l'ipotesi viene rifiutata.