FIRMA DELLO STUDENTE

SECONDA PROVA INTERMEDIA DI STATISTICA
CLEA - CLEFIN - CLAPI (cod. 5047/4038) CLELI (cod. 4038) 371-377
14 Gennaio 2004

Cognome *Nome*

Numero di matricola *Codice corso*

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. (2 punti) Sia X una variabile aleatoria Uniforme continua nell'intervallo $[0, 6k]$, $k > 0$.

a) Calcolare il valore atteso di X. **(1 punto)**

b) Calcolare la probabilità che X assuma valori nell'intervallo $\left(\frac{k}{3}, \frac{k}{2}\right]$. **(1 punto)**

a)

b)

2. (3 punti) Date le variabili aleatorie indipendenti X ~ N(3, 16) e Y ~ N(2, 25):

a) Calcolare il valore atteso e la varianza della v.a. W ottenuta dalla seguente combinazione lineare W = 42 + 7X − 3Y. **(2 punti).**

b) Come si distribuisce la v.a. W? **(1 punto)**

a)

b)

3. **(3 punti))** La probabilità di riuscire a salire sulla linea gialla della metropolitana nelle ore di punta è pari a 4/5.

a) Come si distribuisce la v.a. X_i che descrive se la *i*-esima persona riesce a salire sulla metropolitana oppure no? **(1 punto)**

b) Se si trovano sulla banchina di attesa 150 persone, che cercano di salire indipendentemente una dall'altra, come si distribuisce *approssimativamente* la v.a. $\overline{X}$ "frequenza relativa delle persone che riescono a salire all'arrivo del prossimo convoglio"? **(1 punto)**

c) Qual è la probabilità che almeno 114 di queste riescano a salire? **(1 punto)**

a)
b)
c)

4. **(6 punti)** Sia $(X_1, \ldots, X_6)$ un campione bernoulliano estratto da una popolazione statistica di media $E(X) = 3\theta$ e $Var(X) = 1$.

Siano $T_1 = \left(\frac{1}{5}X_1 + \frac{2}{15}X_2\right)$ e $T_2 = \frac{X_4 + X_6}{6}$ due stimatori per θ.

a) Dimostrare che sono entrambi corretti e calcolarne i rispettivi errori quadratici medi. **(4 punti)**

b) Qual è il più efficiente? E perché? **(2 punti)**

a)
b)

5. (6 punti) Sia ($X_1, \dots, X_6$) un campione bernoulliano di osservazioni estratto da una popolazione Normale (μ, σ^2) di media e varianza entrambe incognite.

a) Si definisca con l'opportuna espressione analitica un intervallo di confidenza al livello ($1-\alpha$) per il parametro μ. **(2 punti).**

b) Se il campione osservato risulta (11, 4, 17, 13, 3, 9), qual è l'intervallo di confidenza per μ a livello $1-\alpha = 0.90$? **(2 punti)** [Sugg. $\sum_{i=1}^{n} x_i = 57$, $\sum_{i=1}^{n} x_i^2 = 685$]

c) Se fosse noto che $\sigma = 7$, quale sarebbe la dimensione campionaria minima necessaria per ottenere, allo stesso livello di confidenza, un intervallo di lunghezza pari al massimo a 4? **(2 punti)**

a)
b)
c)

6. (6 punti) Sia ($X_1, \dots, X_5$) un campione da una popolazione $X \sim N(\mu, 144)$. Si vuole testare l'ipotesi nulla H_o: $\mu = 11$ contro l'ipotesi alternativa H_1: $\mu < 11$.

a) Proponete l'espressione analitica dell'opportuna regione di rifiuto. **(2 punti)**

b) Per $\alpha = 0.05$, se (7, 4, 12, 18, 5) è il campione osservato, accetto o rifiuto l'ipotesi nulla? **(1 punto)**

c) Se il P-value osservato è 0.02, accetto o rifiuto l'ipotesi nulla? (Giustificare la risposta) **(1 punto)**

d) Se il test fosse caratterizzato dalla regione di rifiuto $R = \{(x_1, \dots, x_n) : \bar{x}_n < 9.5\}$ quale risulterebbe la probabilità dell'errore di prima specie α? **(2 punti)**

a)
b)

c)

d)

7. **(4 punti)** Dato un campione di osservazioni (Y_1, x_1), (Y_2, x_2), …, (Y_n, x_n), assumere che il modello lineare semplice spieghi la dipendenza della variabile aleatoria Y dalla variabile deterministica x significa ipotizzare che $Y_i=\beta_0+\beta_1 x_i+\varepsilon_i$, per i=1, …, n.
E' noto che come stime puntuali dei parametri incogniti β_0 e β_1 si utilizzano le stime dei minimi quadrati $\hat{\beta}_0$ e $\hat{\beta}_1$.

a) Quale ipotesi aggiuntiva sulla distribuzione degli errori ε_i, rende *forti* le ipotesi *deboli* per il modello lineare semplice e consente, ad esempio, di calcolare stime di intervallo per β_0 e β_1? **(2 punti)**
b) La formula sotto riportata è nota sotto il nome di **scomposizione della devianza totale.**

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2+\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$$

Quale delle quantità a destra del segno di uguale è la **devianza del modello**? **(1 punto)**

c) Quale rapporto tra le quantità coinvolte nella scomposizione fornisce il coefficiente di determinazione, che misura la bontà di adattamento del modello lineare ? **(1 punto)**

a)

b)

c)