SECONDA PROVA INTERMEDIA DI STATISTICA (Cod. 5047/4038/371/377)

CLEA, CLEFIN *19 gennaio 2005*

Firma dello studente

Cognome Nome N. di matricola

COMPITO B1

Istruzioni. 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

Esercizio I (4 punti)

Siano X e Y due variabili aleatorie indipendenti tali che $X \sim Uniforme\,(0,1)$ e $Y \sim Uniforme\,(0,2)$.

a) Si calcoli $\mathbb{P}\left(\left\{X<\frac{1}{3}\right\}\cap\{Y>1\}\right)$. **(2 punti)**

Per l'indipendenza, si ha che
$\mathbb{P}\left(\left\{X<\frac{1}{3}\right\}\cap\{Y>1\}\right)=\mathbb{P}\left(\left\{X<\frac{1}{3}\right\}\right)\mathbb{P}\left(\{Y>1\}\right)=\frac{1}{6}$

b) Si calcolino il valore atteso e la varianza della variabile aleatoria $W=3Y+8$. **(2 punti)**

$\mathbb{E}\,(W)=3\mathbb{E}\,(Y)+8=3+8=11$
$\mathbb{V}ar\,(W)=9\mathbb{V}ar\,(Y)=9\times\frac{4}{12}=3.$

Esercizio II (4 punti)

Viene estratto un campione casuale $(X_1,...,X_n)$ da una popolazione X con valore atteso $\mathbb{E}\,(X)=\theta+4$ e varianza σ^2. Al fine di stimare θ viene proposto lo stimatore

$$T_n=\frac{1}{n}\sum_{i=1}^{n}X_i-2$$

a) Si mostri che T_n è uno stimatore distorto per θ. Si verifichi, inoltre, se esso è consistente in media quadratica per il parametro θ. **(2 punti)**

$\mathbb{E}\,(T_n)=\frac{n(\theta+4)}{n}-2=\theta+2$ da cui la distorsione.
Inoltre, si vede immediatamente che T_n è asintoticamente distorto. Pertanto, non è consistente in media quadratica

b) Si proponga uno stimatore T_n' non distorto per θ e se ne scriva la varianza. **(2 punti)**

Stimatore: Per il punto a, $T_n'=T_n-2$ è non distorto per θ. Si noti, inoltre che $T_n'=\bar{X}-4$.

Varianza dello stimatore: Per quanto visto, $\mathbb{V}ar\,(T_n')=\mathbb{V}ar\,(T_n-2)=\mathbb{V}ar\,(T_n)=\mathbb{V}ar\left(\bar{X}\right)$ da cui $\mathbb{V}ar\,(T_n')=\frac{\sigma^2}{n}$.

Esercizio III (2 punti)

Sia dato il vettore aleatorio (X, Y),

$X \backslash Y$	0	1
0	0.42	0.18
1	0.28	0.12

Si determini la distribuzione di $Z = X - Y$.

Il supporto di Z è $\{-1, 0, 1\}$ e
$\mathbb{P}(Z = -1) = 0.18$, $\mathbb{P}(Z = 0) = 0.42 + 0.12 = 0.54$ e $\mathbb{P}(Z = 1) = 0.28$

Esercizio IV (3 punti)

Si consideri il modello lineare semplice

$$Y = \beta_0 + \beta_1 x + \varepsilon,$$

e si supponga di aver effettuato nell'ambito di questo le osservazioni $(x_1, Y_1), ..., (x_n, Y_n)$. Si forniscano le ipotesi deboli del modello.

1) $\mathbb{E}(\varepsilon_i) = 0$, $i = 1, ..., n$.
2) $\mathbb{V}ar(\varepsilon_i) = \sigma^2 > 0$, $i = 1, ..., n$.
3) $Cov(\varepsilon_i, \varepsilon_j) = 0$, per ogni $1 \leq i \neq j \leq n$.

Esercizio V (8 punti)

Un campione casuale $(X_1, ..., X_n)$ viene estratto da una popolazione X bernoulliana di parametro θ. Si vogliono verificare le ipotesi

$$H_0 : \theta = 0.4 \quad vs. \quad H_1 : \theta = 0.5$$

a) Si scriva l'espressione generale della regione critica di livello α per verificare le ipotesi **(2 punti)**

$R \equiv \left\{(x_1, \ldots, x_n) : \bar{x} > \theta_0 + z_{1-\alpha} \sqrt{\frac{\theta_0(1-\theta_0)}{n}}\right\}$ supponendo $n \geq 30$

b) Per $n = 100$, supponiamo si sia osservata la realizzazione campionaria $(x_1, ..., x_{100})$ tale che $\sum_{i=1}^{100} x_i = 45$. Si determini la regione di rifiuto di livello $\alpha = 0.05$ e si decida a favore o contro H_0. Infine, si calcoli la probabilità di commettere un errore di seconda specie utilizzando il test. **(5 punti)**

Regione critica: $\left\{\bar{x} > 0.4 + 1.645\sqrt{0.24/100}\right\} = \{\bar{x} > 0.4806\}$

Decisione:
Si accetta H_0 in quanto $\bar{x} = 0.45 < 0.4806$

Errore di seconda specie
$\mathbb{P}\left(Z \leq \frac{0.4806-0.5}{\sqrt{0.25/100}}\right) = \mathbb{P}(Z \leq -0.388) \simeq \mathbb{P}(Z \leq -0.39) = 1 - \mathbb{P}(Z \leq 0.39) = 1 - 0.6517 = 0.3483$

c) Si dica se la decisione alla quale si è giunti nel punto precedente cambierebbe se si fossero volute verificare le ipotesi

$$H_0' : \theta = 0.4 \quad vs. \quad H_1' : \theta > 0.4. \qquad \textbf{(1 punto)}$$

La decisione non cambierebbe. Le regioni di rifiuto associate ai test asintotici relativi a H_0' e H_1' e H_0 e H_1, rispettivamente, coincidono.

Esercizio VI (3 punti)

In una cittadina il 30% degli abitanti é favorevole alla costruzione di un nuovo parcheggio interrato. Sia $(X_1, ..., X_n)$ un campione casuale estratto dalla popolazione.

a) Si fornisca la distribuzione di $W = \sum_{i=1}^{n} X_i$. **(2 punto)**

$W \sim Bin\,(n, 0.3)$

b) Supponendo $n = 5$, si calcoli $\mathbb{P}\,(W > 4)$, essendo $W = \sum_{i=1}^{5} X_i$. **(1 punto)**

$\mathbb{P}\,(W > 4) = \mathbb{P}\,(W = 5) = \mathbb{P}\,(X_i = 1,\ i = 1, ..., 5) = (0.3)^5 = 0.0024$

Esercizio VII (6 punti)

Sia $(X_1, ..., X_n)$ un campione casuale estratto da una popolazione $X \sim N\left(\mu, \sigma^2 = 625\right)$.

a) Si determini l'ampiezza minima del campione affinché si possa costruire un intervallo di confidenza al 95% per μ di lunghezza non superiore a 8. (Si indichi tale ampiezza con n_0) **(2 punti)**

$L = 2z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq 8 \Rightarrow 8\sqrt{n} \geq 2 \times 1.96 \times 25 \Rightarrow n \geq 150.0625 \Rightarrow n_0 = 151.$

b) Utilizzando l'ampiezza campionaria n_0 precedentemente determinata, si costruisca l'intervallo di confidenza al 95% per μ, supponendo di aver ottenuto una realizzazione campionaria $x = (x_1, ..., x_{n_0})$ tale che $\sum_{i=1}^{n_0} x_i = 3020$. (Nel caso non si sia risolto il punto precedente, si utilizzi il valore $n_0 = 313$) **(2 punti)**

$\left(20 - \frac{1.96\times25}{12.2882}, 20 + \frac{1.96\times25}{12.2882}\right) = (16.0124, 23.9876)$
Si fosse utilizzato $n_0 = 313$, l'intervallo di confidenza ottenuto sarebbe stato
$\left(\frac{3020}{313} - \frac{1.96\times25}{17.6918}, \frac{3020}{313} + \frac{1.96\times25}{17.6918}\right) = (6.879, 12.4182)$

c) Come cambierebbe la risposta al punto precedente se si abbandonasse l'ipotesi di normalità della popolazione? Commentare. **(2 punti)**

Se si abbandonasse l'ipotesi di normalità della popolazione, in corrispondenza della realizzazione campionaria x di cui al punto b, si avrebbe lo stesso intervallo ottenuto sopra come intervallo di confidenza asintotico.