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

CLEA, CLEFIN *19 gennaio 2005*

Firma dello studente

Cognome Nome N. di matricola

COMPITO A1

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 $X \sim N(0, 16)$ e $Y \sim N(0, 25)$.

a) Si determini la legge della variabile aleatoria $W = 3X + 2Y$. **(2 punti)**

> W è sicuramente gaussiana. D'altro canto,
> $\mathbb{E}(W) = 3\mathbb{E}(X) + 2\mathbb{E}(Y) = 0$ e $\mathbb{V}ar(W) = 9\mathbb{V}ar(X) + 4\mathbb{V}ar(Y) = 244$.
> Allora, $W \sim N(0, 244)$

b) Si calcoli $\mathbb{P}(\{Y \leq -1\} \cup \{0 \leq Y < 1\})$. **(2 punti)**

> Osserviamo, anzitutto, che $\{Y \leq -1\}$ e $\{0 \leq Y < 1\}$ sono eventi disgiunti. Inoltre, si vede subito che la legge di Y è simmetrica. Segue che
> $\mathbb{P}(\{Y \leq -1\} \cup \{0 \leq Y < 1\}) = \mathbb{P}(\{Y \leq -1\}) + \mathbb{P}(\{0 \leq Y < 1\})$
> $= \mathbb{P}(\{Y > 1\}) + \mathbb{P}(\{0 \leq Y < 1\}) = \mathbb{P}(Y \geq 0) = 1/2$.

Esercizio II (5 punti)

Viene estratto un campione bernoulliano $(X_1, ..., X_n)$ da una popolazione X con media μ e varianza σ^2.

a) Si proponga uno stimatore non distorto per $2\mu + 10$, lo si indichi con T_n e se ne calcoli la varianza. **(2 punti)**

Stimatore	Varianza dello stimatore
$T_n = 2\bar{X} + 10$	$\mathbb{V}ar(T_n) = 4\mathbb{V}ar(\bar{X}) = \frac{4\sigma^2}{n}$

b) Dato $n = 3$, si calcoli la stima ottenuta tramite T_n in corrispondenza della realizzazione campionaria $(x_1, x_2, x_3) = (3, 6, 9)$. **(1 punto)**

> $t_3(3, 6, 9) = 2 \times \frac{3+6+9}{3} + 10 = 22$.

c) Usando uno stimatore non distorto, si fornisca una stima della varianza della popolazione (si utilizzino i dati del punto precedente) **(2 punti)**

> Utilizzando lo stimatore lo stimatore S_c^2 si ottiene
> $s_c^2(3, 6, 9) = \frac{3}{2}s^2(3, 6, 9) = \frac{3}{2} \times 6 = 9$

Esercizio III (4 punti)

Sia (X, Y) un vettore aleatorio la cui distribuzione è riassunta nella tabella seguente:

$X \backslash Y$	3	7	
4	0.48	0.12	0.6
6	0.32	0.08	0.4
	0.8	0.2	

a) Si scriva la distribuzione di $X \mid Y = 7$ e se ne calcoli il valore atteso. **(3 punti)**

Si vede subito dalla distribuzione congiunta di X e Y che trattasi di variabili aleatorie indipendenti.
Segue che la legge di $X \mid Y = 7$ e la legge di X sono identiche. Più precisamente,
$\mathbb{P}(X = 4 \mid Y = 7) = 0.6$ e $\mathbb{P}(X = 6 \mid Y = 7) = 0.4$. Inoltre,
$\mathbb{E}(X \mid Y = 7) = (4 \times 0.6) + (6 \times 0.4) = 4.8$.

b) Si determini $Cov(X, Y)$. **(1 punto)**

Per l'indipendenza di X e Y, si ha immediatamente che $Cov(X, Y) = 0$.
Volendo effettivamente calcolare la covarianza,
si avrebbe $Cov(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = (12 \times 0.48) + (28 \times 0.12) + (18 \times 0.32) + (42 \times 0.08)$
$- [(4 \times 0.6) + (6 \times 0.4)] [(3 \times 0.8) + (7 \times 0.2)] = 0$

Esercizio IV (3 punti)

Si consideri un modello lineare semplice

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

e si supponga di aver effettuato n osservazioni $(x_1, Y_1), ..., (x_n, Y_n)$. Si forniscano le ipotesi forti del modello lineare. **(3 punti)**

1) $\varepsilon_i \sim N(0, \sigma^2)$, per ogni $i = 1, ..., n$,
2) $Cov(\varepsilon_i, \varepsilon_j) = 0$, per ogni $1 \leq i \neq j \leq n$.

Esercizio V (5 punti)

Sia $(X_1, ..., X_9)$ un campione estratto da una popolazione $X \sim N(\mu, \sigma^2)$. Nessuna informazione su μ e σ^2 è disponibile. Si voglia verificare le ipotesi

$$H_0 : \mu = 0 \quad vs. \quad H_1 : \mu < 0.$$

a) Si scriva l'espressione generale della regione critica di livello α per verificare le ipotesi H_0 e H_1. **(2 punti)**

$$R = \left\{ (x_1, ..., x_n) : \bar{x} < \mu_0 - t_{1-\alpha}^{(n-1)} \frac{s_c}{\sqrt{n}} \right\}.$$

b) Supponiamo di aver osservato la realizzazione campionaria $x = (x_1, ..., x_9)$ tale che $\sum_{i=1}^{9} x_i = -27$ e $\sum_{i=1}^{9} x_i^2 = 135$. Si determini la regione critica associata al test relativo a H_0 e H_1 al livello $\alpha = 0.01$ e si decida a favore o contro H_0. **(3 punti)**

Regione critica
$\left\{ (x_1, ..., x_9) : \bar{x} < -\frac{2.896 s_c}{3} \right\}$

Decisione
Si rifiuta H_0 in quanto
$\bar{x}_9 = -3 < -\frac{2.896 \times 2.5981}{3} = -2.5080$

Esercizio VI (2 punti)

Siano $(X_1, ..., X_9)$ e $(Y_1, ..., Y_{16})$ due campioni indipendenti estratti rispettivamente dalle popolazioni $X \sim N\left(\mu_X, \sigma^2\right)$ e $Y \sim N\left(\mu_Y, \sigma^2\right)$. Nessuna informazione su μ_X, μ_Y e σ^2 è disponibile. Si vogliano verificare le ipotesi

$$H_0 : \mu_X = \mu_Y \quad vs. \quad H_1 : \mu_X \neq \mu_Y.$$

Si scriva l'espressione analitica dello stimatore della varianza comune σ^2. **(2 punti)**

$$\boxed{S_p^2\left(X_1, ..., X_9, Y_1, ..., Y_{16}\right) = \frac{1}{23}\left(\sum_{i=1}^{9}\left(X_i - \bar{X}\right)^2 + \sum_{i=1}^{16}\left(Y_i - \bar{Y}\right)^2\right)}$$

Esercizio VII (2 punti)

Sia $(X_1, ..., X_n)$, $n > 5$, un campione bernoulliano estratto da una popolazione X con $\mathbb{E}(X) = \mu$ e $\mathbb{V}ar(X) = \sigma^2$. Si dica, fornendo una giustificazione, se lo stimatore

$$T_n = \frac{n-5}{n}X_1 + \frac{5}{n}X_n$$

è distorto per μ. **(2 punti)**

Lo stimatore è non distorto. Infatti
$\mathbb{E}(T_n) = \left(\frac{n-5+5}{n}\right)\mu = \mu$

Esercizio VIII (5 punti)

Sia θ la proporzione di teleutenti che hanno seguito un programma d'intrattenimento nella serata del 3 novembre 2004. Sia $(X_1, \ldots, X_n)$ un campione estratto dalla suddetta popolazione.

a) Supponendo di aver estratto un campione di $n = 100$ teleutenti e di aver rilevato che 75 di essi hanno seguito il programma in questione, si costruisca un intervallo di confidenza per θ al livello $1 - \alpha = 0.9$. **(3 punti)**

$$\boxed{\left(0.75 - 1.645\sqrt{\frac{0.75 \times 0.25}{100}}, 0.75 + 1.645\sqrt{\frac{0.75 \times 0.25}{100}}\right) = (0.6788, 0.8212)}$$

b) Supponendo $n = 3$ e $\theta = 0.4$, si calcoli la probabilità $\mathbb{P}(W > 2)$, essendo $W = X_1 + X_2 + X_3$. **(2 punti)**

Sappiamo che $W \sim Bin(3, 0.4)$. Inoltre
$\mathbb{P}(W > 2) = \mathbb{P}(W = 3) = (0.4)^3 = 0.064$