**FIRMA DELLO STUDENTE**

# SECONDA PROVA INTERMEDIA DI STATISTICA
## CLEA - CLEFIN - CLELI (COD. 5047/4038)
## 15 gennaio 2002

***Cognome*** ***Nome***

***Numero di matricola***

# COMPITO A1

**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 popolazione distribuita secondo la legge Uniforme sull'intervallo $(\theta-2,\theta+2)$; sia $X_1$, …., $X_4$ un campione bernoulliano estratto da $X$. Dimostrare (*esplicitando il procedimento seguito*) che il seguente stimatore:

$$T(X_1, X_2, X_3, X_4) = \frac{2X_1 - X_2}{2} + \frac{2X_3 + X_4}{3}$$

è distorto per $\theta$ e *valutare* la distorsione di $T$.

$$E(X) = q$$

$$E(T) = E\left(\frac{2X_1 - X_2}{2} + \frac{2X_3 + X_4}{3}\right) = \frac{2E(X_1) - E(X_2)}{2} + \frac{2E(X_3) + E(X_4)}{3} = \frac{3}{2}q$$

$$D(T) = E(T) - q = \frac{1}{2}q$$

2. **(2 punti)** Sia $X_1,\dots,X_n$ un campione bernoulliano estratto da una popolazione con funzione di densità caratterizzata da un parametro $\theta$ non noto. Siano inoltre $H_0$: $\theta = \theta_0$ e $H_1$: $\theta = \theta_1$ due ipotesi *semplici*, e sia $R = \{(x_1,\dots,x_n) : \sum_{i=1}^{n} x_i^2 > 10\}$ la regione critica di un test T* per verificare $H_0$ contro $H_1$. Scrivere l'espressione analitica di $\alpha$ (probabilità di commettere errore di primo tipo) e $\beta$ (probabilità di commettere errore di secondo tipo) in relazione al test T*.

$$a = P\left((x_1,\dots,x_n) : \sum_{i=1}^{n} x_i^2 > 10 \,|\, q = q_0\right)$$

$$b = P\left((x_1,\dots,x_n) : \sum_{i=1}^{n} x_i^2 \le 10 \,|\, q = q_1\right)$$

3. **(4 punti)** Il direttore della scuola elementare XYZ presume che i suoi studenti abbiano un quoziente intellettivo (QI) superiore alla media degli studenti di scuola elementare (pari a 120), così da poter considerare la sua scuola una "Scuola di Eccellenza". Possiamo supporre che la variabile aleatoria "quoziente intellettivo " (QI) si distribuisca (nella scuola) come una normale di valore atteso non noto $\mu$ e varianza 225. Il direttore decide allora di verificare l'ipotesi $H_0$:$\mu$=120 contro $H_1$: $\mu$>120: misura quindi il quoziente di intelligenza di 49 bambini scelti casualmente tra gli studenti e riscontra un valore medio di 124. Scrivere l'espressione analitica del p-value corrispondente alla realizzazione campionaria osservata e calcolarne il valore. Decidere inoltre se il direttore può considerare la sua scuola una "Scuola di Eccellenza", avendo fissato il livello di significatività $\alpha$=0.04 (motivare la risposta).

| |
|---|
| $p-value = P(\overline{X} > \overline{x} \mid m=120) = P(\overline{X} > 124 \mid m=120) = P\left(\frac{\overline{X}-120}{\sqrt{\frac{225}{49}}} > \frac{124-120}{\sqrt{\frac{225}{49}}}\right)$ <br> $= P(Z > 1.867) = 1 - P(Z \le 1.87) = 1 - 0.9693 = 0.0307$ |
| Decisione <br> Rifiutiamo l'ipotesi nulla, la scuola XYZ può essere considerata una Scuola di Eccellenza. |

**4. (3 punti)** Sia $X_1, X_2, \ldots, X_n$ un campione casuale semplice (bernoulliano) estratto da una popolazione X distribuita secondo una legge di distribuzione non nota, tale che E(X)=μ (incognito) e Var(X)=$\sigma^2$(noto). Enunciare le proprietà di cui gode la media campionaria, $\overline{X}$, come stimatore di μ. Specificare inoltre la sua distribuzione asintotica.

| |
|---|
| 1) la media campionaria è stimatore non distorto della media della popolazione $E(\overline{X}) = m \quad \forall m$ <br><br> 2) la media campionaria è stimatore consistente (in senso forte o in media quadratica) della media della popolazione $Var(\overline{X}) = \frac{s^2}{n}$ <br><br> 3) la media campionaria si distribuisce asintoticamente in accordo a una Gaussiana <br> $\overline{X} \sim N\left(m, \frac{s^2}{n}\right)$ |

**5. (4 punti)** Si consideri la popolazione X, con funzione di probabilità

$$p_X(x) = \begin{cases} 0.4 & x=-4, x=4 \\ 0.2 & x=0 \\ 0 & altrove \end{cases}$$

Estratto da X il campione bernoulliano $(X_1, X_2, ..., X_{100})$, e definita la media campionaria $\overline{X} = \frac{\sum_{i=1}^{100} X_i}{100}$, si calcoli il valore atteso di $\overline{X}$, la varianza di $\overline{X}$ e la probabilità approssimata con cui $\overline{X}$ assume valori maggiori di 0.4.

| |
|---|
| $E(\overline{X}) = 0$ $\qquad$ $Var(\overline{X}) = \frac{Var(X)}{100} = \frac{12.8}{100} = 0.128$ <br><br> $P(\overline{X} > 0.4) = P\left(\frac{\overline{X}-0}{\sqrt{0.128}} > \frac{0.4}{0.3578}\right) = P(Z > 1.118) = 1 - P(Z \le 1.12) = 1 - 0.8686 = 0.1314$ |

**6.(2 punti)** Un'urna contiene 40 palline rosse e 60 blu. Vengono estratte con reimmissione due palline. Definiti i due numeri aleatori:

X = {numero di palline rosse estratte}
Y = { numero di palline blu estratte }

si determini la distribuzione del vettore aleatorio (*X, Y*).

| X\Y | 0 | 1 | 2 |
|---|---|---|---|
| **0** | | | 0.36 |
| **1** | | 0.48 | |
| **2** | 0.16 | | |

**7 (13 punti)** Per poter decidere se è necessario un intervento a livello nazionale di propaganda contro il furto, si selezionano a caso 300 persone e si chiede se hanno mai commesso un furto nella loro vita. Se la proporzione di "chi ha commesso almeno un furto" risultasse superiore a 0.55 allora si riterrebbe necessario un intervento. Tra i 300 soggetti intervistati in totale riservatezza, 180 hanno ammesso di aver rubato almeno una volta nella loro vita.

a) Si fornisca uno stimatore non distorto e una stima della proporzione effettiva *p* di coloro che hanno commesso un furto almeno una volta nella loro vita. Si calcoli inoltre la varianza dello stimatore utilizzato **(1+1 + 2 punti)**.

| Stimatore | Stima |
|---|---|
| $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad o \quad \hat{P} = \frac{\# successi}{n}$ proporzione campionaria | **0.6** |

| Varianza |
|---|
| $\frac{\hat{p}(1-\hat{p})}{n} = \frac{.6 * .4}{300} = 0.0008$ |

b) Si vuole determinare l'intervallo di confidenza per *p* al livello $(1-\alpha) = 0.99$. Indicare

| | |
|---|---|
| l'espressione analitica (formula generale in simboli) dell'opportuno intervallo di confidenza **(1 punto)** | $99\% CI = \left[ \hat{p} - z_{1-a/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}};\ \hat{p} + z_{1-a/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]$ |
| il valore del percentile ottenuto sulle tavole a tale scopo **(1 punto)** | 2.576 |
| l'intervallo numerico ottenuto (usare quattro cifre decimali). **(1 punto)** | [0.5271, 0.6729] |

c) Si decide allora di verificare l'ipotesi nulla $H_0: p$ £ $0.55$ contro $H_1: p > 0.55$ al livello $\alpha = 0.05$. Si indichi:

l'espressione analitica (formula generale in simboli) dell'opportuna regione critica
**(2 punti)**

$$R = \left\{ (x_1,..x_{300}) : \hat{p} > p_0 + z_{1-a} \sqrt{\frac{p_0(1-p_0)}{n}} \right\}$$

il valore del percentile letto sulle tavole
**(2 punti)**

1.645

La decisione presa.
**(2 punti)**

Poiché $R = \{\hat{p} > 0.5972\}$
$0.6 > 0.5972$
Rifiuto l'ipotesi nulla