FIRMA DELLO STUDENTE

SOLUZIONI SECONDA PROVA INTERMEDIA DI STATISTICA

CLEA, CLEFIN, CLAPI (COD. 5047/4038)
CLELI 4038
371/377

14 Gennaio 2004

Cognome ***Nome***

Numero di matricola ***Codice corso***

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) La Befana ha preparato dei sacchetti con dei dolci, che possono essere di due dimensioni:
- sacchetti piccoli, il cui peso in grammi è distribuito come una normale, di *media* 100 e *scarto quadratico medio* pari a 5;
- sacchetti grandi, il cui peso in grammi è distribuito come una normale, di *media* 200 e *varianza* 75.

a) Calcolare la probabilità che scelto un sacchetto piccolo a caso tra quelli preparati dalla Befana, questo pesi più di 115 grammi. **[1 punto]**

$$P(X > 115) = P\left(Z > \frac{115-100}{5}\right) = P(Z > 3) = 1 - \Phi(3) = 1 - 0.9987 = 0.0013$$

b) Scrivere la distribuzione del peso di un pacco regalo, costituito come insieme di un sacchetto piccolo e un sacchetto grande scelti indipendentemente l'uno dall'altro. **[1 punto]**

W è *Normale con valore atteso* 300 *e* var*ianza* 100

c) Calcolare la probabilità che un pacco regalo pesi meno di 280 grammi. **[1 punto]**

$$P(W < 280) = P\left(Z < \frac{280-300}{10}\right) = P(Z < -2) = P(Z > 2) = 1 - \Phi(2) = 1 - 0.9772 = 0.0228$$

2) Il vettore aleatorio (X,Y) ha una distribuzione di probabilità rappresentata da

X \ Y	0	2	
1	0,1	0,3	0,4
3	0	0,6	0,6
	0,1	0,9	1

a) Dare la distribuzione di W=X-Y **[1 punto]**
b) Calcolare Cov(W,X) **[1 punto]**

a)

$$W=\begin{cases} -1 & 1 \\ 0.3 & 0.7 \end{cases}$$

b)

$$m(X\cdot W)=1\cdot 0.1-1\cdot 0.3+3\cdot 0.6=1.6$$
$$m(X)=1\cdot 0.4+3\cdot 0.6=2.2$$
$$m(W)=-1\cdot 0.3+1\cdot 0.7=0.4$$

$$Cov(X,W)=m(X\cdot W)-m(X)\cdot m(W)=0.72$$

3) Sia X una popolazione distribuita secondo la legge Bernoulliana di parametro p. Sia inoltre (X_1, X_2, X_3, X_4, X_5) un campione casuale semplice estratto da X.
a) Dimostrare (esplicitando il procedimento seguito) che

$$T_5=\frac{1}{5}(4X_1-3X_2+X_3+X_4+X_5)$$

è distorto per p e valutare la distorsione **[2 punti]**

$$E(T_5)=\frac{1}{5}[4E(X_1)-3E(X_2)+E(X_3)+E(X_4)+E(X_5)]\,ed\ essendo\ le\ X_i\ v.\ a.\ id$$
$$=\frac{1}{5}[4E(X)-3E(X)+E(X)+E(X)+E(X)]=\frac{1}{5}[4p-3p+p+p+p]=\frac{4}{5}p\neq p$$
$$D(T_5)=E(T_5)-p=\frac{4}{5}p-p=-\frac{1}{5}p$$

b) Calcolare la varianza di tale stimatore **[2 punti]**

$Essendo\ le\ X_i\ v.a.\ indipendenti$

$$V(T_5)=\frac{1}{25}[16V(X_1)+9V(X_2)+V(X_3)+V(X_4)+V(X_5)]\ ed\ essendo\ le\ X_i\ v.\ a.\ id$$
$$=\frac{1}{25}[16V(X)+9V(X)+V(X)+V(X)+V(X)]=\frac{28}{25}p(1-p)$$

c)Partendo da T_5 trovare uno stimatore non distorto per p **[1 punto]**

$$T''=\frac{5}{4}T_5$$

d) Supponendo che la popolazione sia di 1000 elementi, trovare uno stimatore non distorto per 1000p (totale di X per la popolazione) a partire dallo stimatore calcolato in c) **[1 punto]**

$$T^*=1000T''$$

4) In un concorso di Miss Italia viene rilevata, tra le altre informazioni, l'altezza (in cm) delle partecipanti (X). Si ipotizza che X ~ $N\left(m, s^{2}\right)$, con m e σ^2 *ignote*. Estratto un campione Bernoulliano di 5 Miss, la realizzazione campionaria ottenuta è la seguente: (176, 184, 169, 190, 181).
Determinare l'intervallo di confidenza per μ a livello 1-α =0,95 indicando:

a) l'espressione analitica in simboli dell'intervallo di confidenza **[1 punto]**

$$\left(\bar{x} - t^{(n-1)}_{(1-\frac{a}{2})}\frac{s_c}{\sqrt{n}}\ ;\ \bar{x} + t^{(n-1)}_{(1-\frac{a}{2})}\frac{s_c}{\sqrt{n}}\right) \qquad \left(\bar{x} - t^{(4)}_{(0.975)}\frac{s_c}{\sqrt{5}}\ ;\ \bar{x} - t^{(4)}_{(0.975)}\frac{s_c}{\sqrt{5}}\right)$$

b) valore del percentile letto sulle tavole **[1 punto]**

$$t^{(4)}_{(0.975)} = 2.776$$

c) intervallo di confidenza (basato sulla realizzazione campionaria sopra riportata) specificando il valore di tutte le quantità coinvolte **[2 punti]**

$$\left(180 - 2.776\sqrt{\frac{63.5}{5}}\ ;\ 180 + 2.776\sqrt{\frac{63.5}{5}}\right) \qquad (170.1072;1\ 89.8928\ )$$

d) La lunghezza dell'intervallo cambierebbe se avessimo osservato un altro campione (sempre di ampiezza 5)? Perché? **[2 punti]**

Sì, cambierebbe perché dipende dalla stima della varianza della popolazione s_c^2 *che varia al variare della realizzazione campionaria.*

5) In un sacchetto di mandarini ci sono dei mandarini guasti e dei mandarini buoni. Basandoci su un campione casuale semplice di 4 estrazioni (in cui ad ogni estrazione si associa 1 se si è pescato un mandarino guasto e 0 altrimenti) si vuole valutare l'ipotesi che la probabilità di trovare un mandarino guasto sia ¼ contro l'ipotesi che la probabilità di trovare un mandarino guasto sia ¾. (H_0: p = ¼; H_1: p = ¾)
Allo scopo è stata proposta la seguente regione di rifiuto: R = $\{(x_1, x_2, x_3, x_4): (x_1 + x_2 + x_3 + x_4) \geq 3\}$.
a)Se osserviamo x_1=1, x_2=0, x_3=1, x_4=1 accettiamo o rifiutiamo l'ipotesi nulla? Perché? **[2 punti]**

Rifiutiamo l'ipotesi nulla perché $\sum_{i=1}^{4} x_i = 3 \in R$

b)Definire, basandosi sulla regione di rifiuto sopra scritta, l'errore di primo tipo **[2 punti]** e l'errore di secondo tipo. **[1 punto]**

$a = P\left(\sum_{i=1}^{4} X_i \geq 3 \mid p = \frac{1}{4}\right)$	$b = P\left(\sum_{i=1}^{4} X_i < 3 \mid p = \frac{3}{4}\right)$

c)Calcolare α. **[2 punti]**

$$\alpha = P\left(\sum_{i=1}^{4} X_i \geq 3 \mid p = \frac{1}{4}\right) = P\left(\sum_{i=1}^{4} X_i = 3 \mid p = \frac{1}{4}\right) + P\left(\sum_{i=1}^{4} X_i = 4 \mid p = \frac{1}{4}\right) = \binom{4}{3}\left(\frac{1}{4}\right)^3\left(\frac{3}{4}\right) + \left(\frac{1}{4}\right)^4 = 0.0508$$

d) Se sulla base di una realizzazione campionaria accetto H_0, posso affermare che H_0 è vera? Perché? **[2 punti]**

No, perché non saprò mai se H_0 è vera o no perché non conoscerò mai il vero valore del parametro incognito.

6) Osservate il seguente output di excel, ottenuto dalla regressione lineare di Y su X1:

Statistica della regressione	
R multiplo	0,8448
R al quadrato	0,7137
R al quadrato corretto	0,7112
Errore standard	27,2427
Osservazioni	117

	Coefficienti	*Errore standard*	*Stat t*
Intercetta	348,45	12,49	27,89
Variabile X1	-5,87	0,35	-16,93

a) Scrivete l'equazione della retta di regressione lineare di Y su X1 **[1 punto]**

$\hat{y}_i = 348.45 - 5.87 x_i$

b) Sulla base della tabella accettereste l'ipotesi che β_1=0? Scrivere il ragionamento seguito e il/i valore/i da cui si capisce. **[2 punti]**

La regione di rifiuto in questo caso è: $R = \left\{ (y_1, x_1), \ldots (y_{117}, x_{117}) : \frac{|\hat{b}_1|}{\hat{s}_{\hat{b}_1}} > t^{(115)}_{(1-\frac{a}{2})} \right\}$

$\frac{|\hat{b}_1|}{\hat{s}_{\hat{b}_1}}$ =16.93 *è talmente elevato che sarà maggiore di qualunque* $t^{(115)}_{(1-\frac{a}{2})}$ *e cadrà quindi nella regione di rifiuto.*

$|Stat\ t|$ =16.93 *è talmente elevato che il p-value sarà 0 e quindi rifiuterò con qualunque* α

b) Sulla base della tabella, ritenete che la relazione lineare tra Y e X1 sia forte o debole? Da quale valore si capisce? **[1 punto]**

La relazione lineare tra Y e X1 è abbastanza forte e lo si vede da R al quadrato=0.7137

$$0 \leq R^2 = \frac{SQM}{SQT} \leq 1$$