PROVA SCRITTA DI STATISTICA (cod. 4038,5047,371,377)

1 Luglio 2004

Modalità A

SOLUZIONI

APPROSSIMARE I CALCOLI ALLA QUARTA CIFRA DECIMALE

ESERCIZIO N. 1 (6 punti)

In una università vi sono 5 corsi di laurea: A, B, C, D ed E. Per ciascuno dei 5 corsi di laurea, sono riportati nella tabella che segue i due caratteri X:"numero studenti iscritti" ed Y:"numero docenti che insegnano":

CORSO DI LAUREA	NUMERO STUDENTI (X)	NUMERO DOCENTI (Y)
A	320	21
B	280	14
C	415	43
D	80	12
E	105	10

a) (2 punti) Si dica, calcolando un opportuno indicatore statistico, se il livello della concentrazione del carattere X è da considerarsi basso, medio oppure alto.

b) (2 punti) Si calcolino media e mediana di Y e, sulla base di questi valori, si dica se la distribuzione di Y può considerarsi simmetrica oppure no.

c) (2 punti) Si dica se il valore assoluto del coefficiente di correlazione lineare tra X e Y è oppure no strettamente minore di 1, giustificando la risposta.

a) Calcoliamo il rapporto di concentrazione di Gini.

$$R = \frac{\sum(F_i - Q_i)}{\sum F_i} = \frac{(0.2-\frac{80}{1200})+(0.4-\frac{185}{1200})+(0.6-\frac{465}{1200})+(0.8-\frac{785}{1200})}{0.2+0.4+0.6+0.8} = \frac{0.7375}{2} = 0.3687.$$

Essendo R compreso tra 0 ed 1, il valore 0.3687 indica un livello di concentrazione medio, medio-basso.

b) La mediana di Y è 14, mentre la media aritmetica di Y è 20. Questi valori indicano una asimmetria verso destra della distribuzione di Y.

c) Il coefficiente di correlazione lineare tra X e Y è certamente inferiore a 1 in valore assoluto, in quanto non c'è perfetta dipendenza lineare tra X e Y.

ESERCIZIO N.2 (4 punti)

a) (2 punti) Si dia la definizione di stimatore asintoticamente non distorto per un parametro q e si fornisca un esempio di stimatore distorto ma asintoticamente non distorto per la media di una popolazione.

b) (2 punti) Uno stimatore asintoticamente distorto può essere consistente? Perché?

a) Sia T_n uno stimatore del parametro incognito q associato al campione $X_1,...,X_n$. T_n si dice asintoticamente non distorto per q se risulta E(T_n) $\to q$ per $n \to \infty$.

Un esempio di stimatore distorto ma asintoticamente non distorto per la media è $T_n = \overline{X}_n + \frac{1}{n}$.

b) Uno stimatore asintoticamente distorto non può essere consistente in media quadratica; infatti uno stimatore T_n è consistente in media quadratica se e solo se Var(T_n) $\rightarrow 0$ e E(T_n) $\rightarrow q$.

ESERCIZIO N.3 (4 punti)

La sede provinciale di una società di fornitura di gas naturale ha rinnovato la propria struttura di intervento tecnico e dichiara di aver raggiunto un livello di efficienza descrivibile da un tempo medio di intervento m inferiore a 4 giorni. Allo scopo di verificare questa affermazione vengono rilevati i tempi di attesa di un campione sufficientemente grande di clienti e si ottengono i risultati riassunti dalle seguenti quantità (espresse in giorni): $\sum_{i=1}^{50} x_i = 160$ e $\sum_{i=1}^{50} x_i^2 = 700$.

a) (2 punti) Si determinino stime puntuali del tempo medio di intervento e della varianza del tempo di intervento.

b) (2 punti) Utilizzando un test di livello 0.1, si decida, sulla base dei dati forniti, se rifiutare oppure no l'ipotesi nulla $H_0 : m \geq 4$ contro l'alternativa $H_1 : m < 4$.

a) La stima del tempo medio di intervento è 160/50=3.2. La stima della varianza (utilizzando lo stimatore non distorto "varianza campionaria corretta") è $s_c^2 = \frac{1}{n-1}\sum(x_i - \overline{x})^2 = \frac{50}{49}\left(\sum x_i^2 / 50 - \overline{x}^2\right) = \frac{50}{49}(14 - 10.24) = 3.8367.$

b) Il test da utilizzare ha regione di rifiuto $R = \left\{ \frac{\overline{x} - 4}{s_c / \sqrt{50}} \leq -z_{0.9} = -1.282 \right\}$; poiché

$$\frac{\overline{x} - 4}{s_c / \sqrt{50}} = \frac{3.2 - 4}{\sqrt{3.8367} / \sqrt{50}} = -2.888 < -1.282,$$ si rifiuta l'ipotesi nulla.

ESERCIZIO N.4 (3 punti) Si consideri una popolazione normale, con media incognita e varianza pari a 9, ed un campione estratto da questa, di ampiezza n=36.

Supponendo di ottenere, in base alla realizzazione del campione, l'intervallo di confidenza (2.8,5.2), si dica quale è la stima puntuale della media incognita e quale è il coefficiente di confidenza dell'intervallo considerato.

L'intervallo di confidenza per la media di una popolazione normale con varianza nota è $(\overline{x} - z_{1-a/2} s / \sqrt{n}, \overline{x} + z_{1-a/2} s / \sqrt{n})$ e la sua lunghezza è $L = 2 z_{1-a/2} s / \sqrt{n}$ .Poiché la lunghezza dell'intervallo ottenuto è 2.4, si ha $L = 2 z_{1-a/2} 3/6 = z_{1-a/2} = 2.4$, da cui $1 - a = 0.9836$; quindi l'intervallo ha coefficiente di confidenza 0.9836. Inoltre, la stima puntuale della media è il punto centrale dell'intervallo, quindi 4.

ESERCIZIO N.5 (7 punti)

L'assemblaggio automatico di un certo manufatto si svolge in 2 fasi indipendenti; indichiamo con X_1 e X_2 i tempi (in minuti) necessari per completare la I e la II fase rispettivamente e supponiamo che X_1 e X_2 abbiano distribuzioni normali di medie 4 e 6 e varianze 0.09 e 0.16 rispettivamente. Si determini:

a) (2 punti) la probabilità che il tempo di completamento della II fase sia superiore a 5 minuti.

b) (1 punto) il tempo in media necessario per completare l'assemblaggio.
c) (1 punto) la probabilità che il tempo necessario per l'assemblaggio completo del manufatto sia compreso tra 9 e 10.5 minuti.
d) (2 punti) Si considerano ora 60 manufatti del tipo in esame e si è interessati al tempo complessivo di assemblaggio degli stessi (ovvero, alla somma dei tempi di assemblaggio dei 60 manufatti). Si calcoli la probabilità che il tempo complessivo sia inferiore a 595 minuti.
e) (1 punto) Si potrebbe rispondere alla domanda d) togliendo l'ipotesi che le distribuzioni dei tempi di completamento delle due fasi di assemblaggio siano normali? Si giustifichi la risposta.

a) $P(X_2 > 5) = P(\frac{X_2 - 6}{0.4} > \frac{5-6}{0.4}) = P(Z > -2.5) = 0.9938.$
b) Il tempo medio richiesto è $E(X)=E(X_1) + E(X_2) = 4 + 6 = 10.$
c) Indicato con $X = X_1 + X_2$ il tempo necessario per l'assemblaggio completo, si ha che X ha distribuzione normale con media 10 e varianza 0.25, per cui $P(9 < X < 10.5) = P(-2 < Z < 1) = 0.8185.$
d) I tempi di assemblaggio dei 60 manufatti sono variabili aleatorie indipendenti e tutte distribuite in accordo ad una normale con media 10 e varianza 0.25; quindi, il tempo complessivo è una variabile aleatoria Y con distribuzione normale con media 600 e varianza 15. Dunque

$$P(Y < 595) = P(\frac{Y-600}{3.873} < \frac{595-600}{3.873}) = P(Z < -1.29) = 1 - \Phi(1.29) = 1 - 0.9015 = 0.0985.$$

e) Si potrebbe ugualmente rispondere alla domanda, utilizzando il teorema centrale del limite, applicabile in questo caso perché n=60 è sufficientemente grande; la probabilità cercata sarebbe la stessa, da intendersi però questa volta come approssimata.

ESERCIZIO N.6 (3 punti)
Un carattere X ha media aritmetica pari a 10 e scarto quadratico medio pari a 0.5.
a) (2 punti) Si determini l'intervallo più piccolo, centrato in 10, nel quale abbiamo la certezza, sulla base delle sole informazioni in nostro possesso (media e scarto quadratico medio), cadano almeno il 95% delle osservazioni su X.
b) (1 punto) Si consideri l'intervallo (9,12). E' possibile che risulti Fr(9<X<12)=0.6? Si giustifichi la risposta.

a) In base alla disuguaglianza di Cebicev, si ha che $Fr(10 - k < X < 10 + k) \geq 1 - s^2 / k^2 = 1 - 0.25 / k^2$; ponendo quindi quest'ultimo valore pari a 0.95, si ottiene k=2.236, per cui il più piccolo intervallo in cui certamente cade almeno il 95% delle osservazione è (7.764,12.236).
b) Sempre utilizzando la disuguaglianza di Cebicev, si osserva che $Fr(9 < X < 12) \geq Fr(9 < X < 11) \geq 1 - 0.25/1 = 0.75$, per cui la Fr(9<X<12) non può essere 0.6.