PROVA SCRITTA DI STATISTICA
CLEA-CLEFIN-CLELI (COD. 4038)
27 gennaio 2003
SOLUZIONI

<u>ESERCIZIO 1</u> (2 punti)
Si calcoli la mediana del carattere X la cui distribuzione di frequenze è la seguente:

x_i	n_i
-1	20
3	40
4	10
6	40
12	40

La mediana è 6, ovvero la prima modalità cui corrisponde una frequenza relativa cumulata maggiore o uguale a 0.5 (in questo caso, alla modalità 6 corrisponde una frequenza relativa cumulata pari a 11/15=0.733.

<u>ESERCIZIO 2</u> (4 punti)
In un'indagine volta a rilevare la puntualità dei treni su un determinato percorso, si analizza un campione di 200 treni e si osserva che 140 di essi sono giunti a destinazione in ritardo.

a) Si determini una stima puntuale, ottenuta da uno stimatore non distorto, per la frequenza incognita q di treni puntuali su quel percorso.
b) Si determini un intervallo di confidenza per q con coefficiente di confidenza 0.9.

a) Una stima puntuale per la frequenza incognita di treni puntuali è la frequenza campionaria, uguale, nel caso in esame, a 60/200=0.3.
b) Un intervallo di confidenza per q di coefficiente 0.9 è:

$$(\overline{x} - z_{1-a/2}\frac{\sqrt{\overline{x}(1-\overline{x})}}{\sqrt{n}}, \overline{x} + z_{1-a/2}\frac{\sqrt{\overline{x}(1-\overline{x})}}{\sqrt{n}}) =$$

$$(0.3-1.645\frac{\sqrt{0.3\cdot 0.7}}{\sqrt{200}}, 0.3+1.645\frac{\sqrt{0.3\cdot 0.7}}{\sqrt{200}}) = (0.2467, 0.3533).$$

<u>ESERCIZIO 3</u> (5 punti)
Si vuole verificare se le votazioni medie di laurea in due università differiscono tra loro. A questo scopo si considerano campioni (indipendenti) di $n_1 = 5$ laureati della prima

università e di $n_2 = 4$ laureati della seconda università e dalla rilevazione si ottengono i seguenti dati:

votazioni I università	103	102	90	88	107

votazioni II università	99	99	106	96

Supponiamo che le distribuzioni dei voti di laurea in entrambe le università siano normali (o gaussiane) con varianze (incognite) uguali.

a) Si fornisca una stima, basata su uno stimatore non distorto, per la varianza incognita comune delle votazioni di laurea nella due università.
b) Si decida se rifiutare oppure no l'ipotesi che le votazioni medie nelle due università siano uguali, utilizzando un test di dimensione 0.05.

Si tratta di due popolazioni normali indipendenti con varianze incognite ed uguali.

a) Una stima per la varianza incognita comune alle due popolazioni è:

$$s_p^2 = \frac{\sum (x_i - \bar{x})^2 + \sum (y_i - \bar{y})^2}{n_1 + n_2 - 2} = \frac{286 + 54}{7} = 48.5714$$

b) Per decidere se rifiutare oppure no l'ipotesi di uguaglianza delle medie delle due popolazioni, utilizziamo la regione di rifiuto

$$R = \left\{ (x_1, \ldots, x_5), (y_1, \ldots, y_4) : \frac{|\bar{x} - \bar{y}|}{s_p \sqrt{1/5 + 1/4}} > t_{0.975}^{7} \right\}.$$

Poiché $\dfrac{|\bar{x} - \bar{y}|}{s_p \sqrt{1/5 + 1/4}} = 0.4278$ e $t_{0.975}^{7} = 2.365$, l'ipotesi di uguaglianza delle medie delle due popolazioni non deve essere rifiutata.

<u>ESERCIZIO 4</u> (4 punti)

Si consideri una popolazione con distribuzione normale con media incognita m e varianza $s^2 = 16$ ed un campione bernoulliano di ampiezza n estratto da questa.

a) Si scriva l'espressione, in funzione di *n*, della lunghezza dell'intervallo di confidenza per m, di coefficiente di confidenza 0.99.
b) Si dica quale deve essere il numero minimo di elementi del campione affinchè l'intervallo di cui al punto a) abbia lunghezza non superiore a 2.

a) La lunghezza dell'intervallo è $L(n) = 2 z_{1-a/2} \dfrac{s}{\sqrt{n}} = 2 \cdot 2.576 \dfrac{4}{\sqrt{n}} = \dfrac{20.608}{\sqrt{n}}$.
b) Imponendo la condizione $L(n) \leq 2$, si ottiene $\sqrt{n} \geq 10.304$, ovvero $n \geq 106.1724$; quindi, l'ampiezza minima richiesta del campione è *n*=107.

ESERCIZIO 5 (4 punti)

a) Si scriva la funzione di probabilità di una variabile aleatoria X con distribuzione binomiale di parametri $n=7$ e $q=0.4$.

b) Con riferimento alla variabile aleatoria X di cui al punto a), si calcoli $P(X<2)$.

c) Si calcoli E(X), sempre con riferimento alla variabile aleatoria in a).

a) $p(x)=\begin{cases}\binom{7}{x}(0.4)^x(0.6)^{7-x} & x=0,1,...,7\\ 0 & altrove\end{cases}$.

b) $P(X<2)=P(X=0)+P(X=1)=0.0279+0.1306=0.1585$.

c) $E(X)=nq=2.8$.

ESERCIZIO 6 (8 punti)

Si considerino due caratteri X e Y tali che:

$m(X)=30, m(Y)=6, \text{cov}(X,Y)=-3, \text{var}(X)=9, \text{var}(Y)=1$.

a) Si dica, giustificando la risposta, se è da ritenersi più disperso X oppure Y.

b) Si scriva l'equazione della retta dei minimi quadrati di Y su X.

c) Si dica, giustificando la risposta, se X e Y sono oppure no statisticamente indipendenti.

d) Si dica se è possibile, con i dati a disposizione, calcolare $\tilde{\chi}^2(X,Y)$ e, in caso affermativo, se ne indichi il valore.

a) Si ha: $cv(X)=\frac{\sqrt{Var(X)}}{|m(X)|}=\frac{3}{30}=0.1$ e $cv(Y)=\frac{\sqrt{Var(Y)}}{|m(Y)|}=\frac{1}{6}=0.1667$, per cui X è meno disperso di Y.

b) L'equazione è $\hat{y}=\frac{\text{cov}(X,Y)}{Var(X)}x+m(Y)-m(X)\frac{\text{cov}(X,Y)}{Var(X)}=-\frac{1}{3}x+16.$

c) X e Y non sono statisticamente indipendenti; infatti, se lo fossero, la loro covarianza sarebbe uguale a 0.

d) E' possibile affermare che $\tilde{\chi}^2(X,Y)=1$ in quanto, essendo $r^2(X,Y)=1$, X e Y hanno una perfetta associazione lineare e quindi, di conseguenza, il loro livello di connessione è massimo.