

Esame di MQAT - Statistica del 20-01-2010

Nome / eMail :

- 1) Calcolare la mediana e l'indice di Gini relativo nella seguente tabella di frequenze

$$\begin{pmatrix} j & x_j \text{ - } x_{j+1} & n_j \\ 1 & 0 \text{ - } 10 & 30 \\ 2 & 10 \text{ - } 25 & 5 \\ 3 & 25 \text{ - } 50 & 15 \end{pmatrix}$$

- 2) Costruire LA matrice di frequenze sotto condizioni di INdipendenza ed UNA matrice di frequenze sotto condizioni di perfetta dipendenza, della seguente tabella 3x3 di frequenze congiunte reali

$$\begin{pmatrix} 0 & 1 & 2 \\ 3 & 4 & 0 \\ 0 & 5 & 6 \end{pmatrix}$$

- 3) Si consideri una variabile X di tipo binomiale, con parametri $\pi = 0.4$ ed $n = 15$.
Si chiede di disegnare il grafico della funzione e di calcolare la probabilita' condizionata $P(x \leq 4 | x \leq 7)$
- 4) Date due variabili (x, y) indipendenti e standardizzate ($\mu = 0, \sigma = 1$) e la statistica $z = (1 + 2x - y/3)$, calcolare la media e la varianza di z .
- 5) Due campioni indipendenti di uguale numerosita' $n = 12$, hanno prodotto le seguenti statistiche: $\bar{x} = +1, \bar{y} = -2, S_x^2 = 6, S_y^2 = 9$. Costruire un intervallo di confidenza per la differenza delle medie e verificare l'ipotesi di omogeneita' in media. A parita' di condizioni, determinare il valore di n per cui la decisione e' in "bilico" (il valore ipotizzato "cade" sull'estremo dell'intervallo).
- 6) Dato il modello lineare $Y_i = 1 - 0.5 X_i + e_i$, con $n=10, \bar{X}=5, S_x=2$ e $S_e^2=3$.
Fare il grafico della retta e calcolare la previsione \hat{Y}_{n+1} per $X_{n+1}=11$. Con l'intervallo di previsione verificare se il valore osservato $Y_{n+1}=-3$ e' anomalo.

Soluzioni

1) bisogna costruire la matrice

$\dot{x}_j = (x_j + x_{j+1})/2$	f_j	F_j	$X_j = n_j \dot{x}_j$	$X_{(j)}$	A_j	Q_j	P_j
5	0.6	0.6	150	87.5	87.5	0.11	0.33
17.5	0.1	0.7	87.5	150	237.5	0.3	0.66
37.5	0.3	1	562.5	562.5	800	1	1

La classe mediana e' la prima essendo $F_1 \geq 0.5$. L'indice di Gini assoluto diventa $G=0.58$ e avendo $G_{\max}=(3-1)/2$ quello relativo $g=0.29$, e' medio basso

2) La matrice di frequenza (assoluta) di perfetta indipendenza e' data dal prodotto delle frequenze marginali come $n_{ij}^* = n_i n_j / n$, quindi $n_{11}^*=0.43$, $n_{21}^*=1$, ecc.. La matrice di perfetta dipendenza e' data da UNA matrice diagonale, con somma dei valori uguali ad n , ad esempio, quella con dipendenza negativa e' $n_{13}^{**}=3$, $n_{22}^{**}=7$, $n_{31}^{**}=11$.

3) La funzione binomiale $B(0.4,15)$ ha 16 valori $x=0,1 \dots 15$ e media $\mu=6$. La probabilita' cercata diventa $P(x \leq 4 | x \leq 7) = P(x \leq 4) / P(x \leq 7)$, ed approssimando alla normale $N(6, 3.6)$ si ha $P(x \leq 4) \approx P(z \leq -1.05) = 15\%$, ecc..

4) La media $E(z) = 1 + 2E(x) - E(y) / 3 = 1$, la varianza $V(z) = 4V(x) + V(y) / 9 = 4.111$

5) L'intervallo di confidenza approssimato al 95% per la differenza delle medie in campioni non normali e'

$$\hat{\delta}_{1,2} \approx (1 - (-2)) \pm 2\sqrt{6/12 + 9/12} = [0.76, 5.24]$$

da cui si Rifiuta l'ipotesi di medie uguali (c'e' disomogeneita'). Il valore "critico" di n e' quello per cui, a parita' di condizioni, l'estremo inferiore dell'intervallo eguaglia zero, quindi $3 = 2\sqrt{15/n}$ da cui $n \approx 7$

6) La retta e' inclinata negativamente a partire dalla intercetta $\alpha = 1$. La previsione e' $\hat{Y}_{11} = 1 - 0.5 * 11 = -4.5$ il suo intervallo di confidenza approssimato al 95% e' $\hat{Y}_{11} \pm 2 S_e = [-1, -8]$. L'intervallo esatto sotto ipotesi di normalita' e regressori deterministici e' $\hat{Y}_{11} \pm t_{0.025}(n-2) S_e \sqrt{1 + 1/n + (X_{11} - \bar{X})^2 / (n * S_x^2)}$ piu' largo. Il valore osservato -3 non e' quindi anomalo.