Compito di Biostatistica

a.a.2009/2010

27/09/2010

Cognome.....**Nome**.....
matricola.....

1. Graficare il diagramma *ramo-foglia* del campione $\mathcal{C}_{50}(X)$ di misure della quantità X rappresentato dalla tabella di dati riportata qui di seguito

62	33	46	81	76	53	25	40	64	33
40	75	34	37	65	69	52	59	19	79
85	47	57	53	54	43	36	54	70	62
56	45	73	71	67	48	37	72	53	42
32	59	49	60	29	77	61	50	71	31

   e calcolare:

   (a) mediana campionaria;

   (b) media campionaria;

   (c) varianza e deviazione standard campionarie.

      i. Testare l'ipotesi di gaussianità delle misure della quantità X mediante il test del χ^2 (si consiglia di suddividere $\mathbb{R}$ in 4 intervalli).

2. Considerare il campione bivariato $\mathcal{C}_{10}(X, Y)$ relativo alle misure delle quantità X e Y rappresentato dalla tabella qui di seguito riportata

$(8, 19)$	$(13, 23)$
$(9, 20)$	$(14, 23)$
$(10, 20)$	$(15, 23)$
$(11, 21)$	$(16, 24)$
$(12, 22)$	$(17, 25)$

   e calcolare:

   (a) il coefficiente di correlazione campionaria;

   (b) nel caso si ritenga plausibile l'ipostesi che le misure delle quantità X e Y siano legate da una relazione lineare, giustificare questa affermazione e calcolare i parametri della retta di regressione.

ES. 1

1	9
2	5 9
3	3 3 4 7 6 7 2 1
4	6 0 0 7 3 5 8 2 9
5	3 2 9 7 3 4 4 6 3 9 0
6	2 4 5 9 2 7 0 1
7	6 5 9 0 3 1 2 7 1
8	1 5

Riordinando

1	9
2	5 9
3	1 2 3 3 4 6 7 7
4	0 0 2 3 5 6 7 8 9
5	0 2 3 3 3 4 7 2 1 0 0
6	0 1 2 2 4 5 7 9
7	0 1 1 2 3 5 6 7 9
8	1 5

$$\tilde{x} = \frac{\hat{x}_{\frac{N}{2}} + \hat{x}_{\frac{N}{2}+1}}{2} = \hat{x}_{25} + \hat{x}_{26}$$

$$= \frac{53 + 54}{2} = 53,50$$

$$\bar{x} = \frac{1}{50} \sum_{i=1}^{50} x_i = 53,74$$

$$s_x^2 = \frac{1}{49} \sum_{i=1}^{50} (x_i - \bar{x})^2 = \frac{1}{49} \left(\sum_{i=1}^{50} x_i^2 - 49 \bar{x}^2 \right)$$

$$= 266,25$$

$$s_x = \sqrt{s_x^2} = 16,32$$

$$A_{-2} = (-\infty ; \bar{x} - s_x] = (-\infty ; 37,42]$$

$$A_{-1} = (\bar{x} - s_x ; \bar{x}] = (37,42 ; 53,74]$$

$$A_1 = (\bar{x} ; \bar{x} + s_x] = (53,74 ; 70,06]$$

$$A_2 = (\bar{x} + s_x ; +\infty] = (70,06 ; +\infty)$$

$$O_{-2} = 11 ; O_{-1} = 4 ; O_1 = 5 ; O_2 = 0$$

$$E_{-2} = E_2 = 50 \cdot 0,16 = 8$$

$$E_{-1} = E_1 = 50 \cdot 0,34 = 17$$

Poiché $\frac{(O_2 - t_2)^2}{t_2} = \frac{9}{8} > 1$,

$\bar{\chi}^2 > 1$ in quanto somma di termini positivi. Dunque, l'ipotesi che i dati del campione siano in linea con una particolare famiglia di parametri $\mu = \bar{x}$ e $\sigma = S_x$ è da rigettarsi.

Es. 2

X	Y
8	19
9	20
10	20
11	21
12	22
13	23
14	23
15	23
16	24
17	25

$$\bar{x} = \frac{1}{10} \sum_{i=1}^{10} x_i = 12,50$$

$$S_x^2 = \frac{1}{9} \sum_{i=1}^{10} (x_i - \bar{x})^2 = 9,17$$

$$S_x = \sqrt{S_x^2} = 3,03$$

$$\bar{y} = 22$$

$$S_y^2 = 3,78$$

$$S_y = 1,94$$

$$S_{xy} = \frac{1}{9} \sum_{i=1}^{10} (x_i - \bar{x})(y_i - \bar{y}) = 5,46$$

$$r_{xy} = \frac{S_{xy}}{S_x S_y} = 0,92$$

Poiché $r_{xy} \cong 1$ è lecito supporre che esista una relazione lineare che lega X e Y cioè

$$Y = AX + B.$$

I valori ottimali stimati di A e B risultano

$$\check{A} = \frac{S_{xy}}{S_x^2} = 0,63$$

$$\check{B} = \bar{y} - \check{A}\bar{x} = 14,12$$

Pertanto la retta di regressione risulta

$$Y = 0,63X + 14,12$$