CdS in Scienze Ambientali / L. Specialistica TeCoRe
Fondamenti di Statistica / Metodi Statistici
Prova Scritta del 3/09/2010

1. (8 punti) I seguenti dati

H+	11	11	22	21	23	40	21	41	27	29
SO4-2	11	16	25	26	30	53	31	58	40	32

si riferiscono alla quantità di ioni H+ e SO4-2 riscontrate in pioggia acida in 10 siti osservati. Riportare le coppie di dati su un grafico. Dopo aver determinato il coefficiente di correlazione, valutare con un test di ipotesi (liv.sign. 5%) se la correlazione tra i due parametri è statisticamente significativa.

2. (14 punti) In uno studio sull'acidità (pH) della pioggia, un campione di misurazioni in 15 stazioni di una zona industriale è costituito dai seguenti valori

3.3 5.2 4.8 3.7 3.9 4.3 4.9 4.3 4.0 4.8 4.2 4.9 4.1 5.0 4.3

a) Supponendo che il campione provenga da una popolazione distribuita normalmente, determinare intervalli di confidenza al 95% per la media e la varianza.

b) Dopo alcuni mesi, le misurazioni sono ripetute, e la media campionaria ottenuta è 4.53. Possiamo concludere che ci sia stata una variazione nel pH dell'acidità della pioggia? (liv.sign. 5%)

c) Un'azienda privata riporta i seguenti valori di acidità, ottenuti con proprie centraline negli stessi siti:

3.4 5.3 5.1 3.7 3.2 4.7 4.2 4.3 4.4 5.0 4.5 4.5 4.5 4.7 4.7

Mediante un confronto di coppie rispetto ai valori nel punto a), possiamo affermare che le misurazioni delle centraline sono equivalenti (liv.sign. 5%)?

3. (10 punti) In uno studio sulla riproduzione di due specie di animali, A e B, vengono considerati campioni casuali di 80 animali della specie A e 65 della specie B. I valori dei pesi medi e delle relative deviazioni standard sono riportati di seguito

$$A: \ \bar{x}_1 = 5.13, s_1 = 0.4 \qquad B: \ \bar{x}_2 = 4.86, s_2 = 0.7.$$

Determinare intervalli di confidenza al 95% per la differenza tra i due pesi medi. Fare un test dell'ipotesi che i pesi medi della popolazione siano uguali, con $\alpha = 0.05$ e $\alpha = 0.01$. Calcolare inoltre i p-valori ed interpretare il risultato.

Soluzione

1. Il diagramma di dispersione è riportato in Figura 1. Il coefficiente di correlazione è:

$$r = \frac{1}{n-1} \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{s_x s_y} = 0.97158,$$

dove $n = 10$, $\bar{x} = 24.6$, $\bar{y} = 32.2$, $s_x = 10.2$ e $s_y = 14.786$. Mediante la statistica $t = r\sqrt{(n-2)/(1-r^2)}$, confrontiamo t con la t di Student t_{n-2} con $\alpha = 0.05$. Si ha $t = 11.61$ e $t_8(\alpha) = 2.306$ (test a due code). Dato che $t \gg t_8(\alpha)$, rifiuto l'ipotesi di zero correlazione, a favore dell'ipotesi di correlazione.

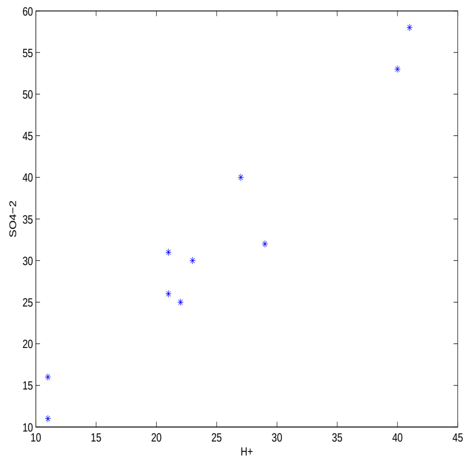

Figure 1: Diagramma di dispersione. Es.1

2. (a). Gli intervalli si ottengono come $[\bar{x} - z_0 s/\sqrt{n}, \bar{x} + z_0 s/\sqrt{n}]$. Si ha $n = 15$, $\bar{x} = 4.38$ ed $s = 0.54011$. Per livello di confidenza 95% si ha $z_0 = 1.96$, mentre per il 99% si ha $z_0 = 2.5758$, quindi

$$95\%: \ [4.1067, 4.6533], \qquad 99\%: \ [4.0208, 4.7392].$$

Per la varianza, si ha $\sigma^2 \in [(n-1)s^2/\chi^2_U, (n-1)s^2/\chi^2_L]$. Per livello di confidenza 95% si ha $\chi^2_U = 26.119$ e $\chi^2_L = 5.6287$, mentre per il 99%, $\chi^2_U = 31.319$ e $\chi^2_L = 4.0747$. Sostituendo, si ottiene

$$95\%: \ [0.15636, 0.7255], \qquad 99\%: \ [0.13040, 1.0023].$$

(b). Considerando come valori storici la media e la varianza del campione iniziale, cioè $s = \sigma$ e $\bar{x} = \mu$, impostiamo il test con H_0: $\mu = 4.38$ e H_1: $\mu \neq 4.38$ (test a due code). Se il nuovo valore è considerato come campione di un solo valore, allora si ha $z = (x - \mu)/\sigma = (4.53 - 4.38)/0.54011 = 0.27772$. Confrontato con $z_0 = 1.96$ (liv.sign. 5%), si ottiene che $z \ll z_0$, e quindi non ci sono motivi per rifiutare l'ipotesi nulla, cioè non risulta esserci stata una variazione nel pH.

Se il nuovo valore è invece considerato come media campionaria $\bar{x}$ di un campione di dimensione $n = 15$, allora si usa $z = (\bar{x} - \mu)/(\sigma/\sqrt{n}) = (4.53 - 4.38)/(0.54011/\sqrt{15}) = 1.07$, e la conclusione rimane invariata.

(c). Siano y_i, $i = 1, n$, i valori del nuovo campione. Definiamo $d_i = x_i - y_i$ ed impostiamo il test H_0: $\delta = 0$ e H_1: $\delta \neq 0$ (test a due code). Si ha $t = (\bar{d} - \delta)/(s/\sqrt{n}) = (-0.0333 - 0)/(0.38668/\sqrt{15}) = -0.3335$ e $t_{n-1} = -2.1448$ (n piccolo), si osserva che t non ricade nella zona critica, quindi non abbiamo motivo di rifiutare l'ipotesi nulla. Concludiamo che non ci sono differenze tra i valori rilevati dalle due centraline.

3. ....