▶ Test per la verifica dell'indipendenza di due variabili

Per verificare se due variabili possono essere considerate correlate oppure indipendenti fra loro si può usare uno dei seguenti test.

Test di verifica di *r*. Con il coefficiente di correlazione, si può effettuare un test di ipotesi per verificare l'indipendenza di due variabili distribuite normalmente.

Ipotesi

1. *test a una coda*
   H_0: I valori delle due coppie di variabili sono indipendenti
   H_1: I valori sono legati da una relazione lineare diretta
2. *test a una coda*
   H_0: I valori delle due coppie di variabili sono indipendenti
   H_1: I valori sono legati da una relazione lineare inversa
3. *test a due code*
   H_0: I valori delle due coppie di variabili sono indipendenti
   H_1: I valori sono legati da una relazione lineare

Funzione discriminante

Si calcola il valore di r in base a una delle formule (31.16 o 17).

Decisione

Si confronta il valore di r calcolato (r_c) con quello riportato nelle tabelle 31.11 o 31.12 (r_t), in corrispondenza dell'α prescelto e per $\nu = n - 2$ e, a seconda che il tipo di test sia a una o due code. L'ipotesi nulla è accettata se:

$$r_c < r_t$$

Test di verifica di *r* con il t di Student. Si può verificare la significatività del coefficiente di correlazione (r) usando la seguente funzione discriminante:

$$t = r \cdot \sqrt{\frac{n-2}{1-r^2}}$$

Si confronta poi il valore così calcolato (t_c) con il valore (t_t) ricavato dalla tabella 31.6 o 31.7, in funzione dell'α prescelto e $v = n - 2$; se:

$$t_c \geq t_t$$

si può rifiutare l'ipotesi nulla e affermare che «esiste una relazione lineare significativa al livello α» fra le due variabili.

Test di verifica di r con l'F di Fisher. Si può verificare la significatività del coefficiente di correlazione usando un test basato sull'analisi della varianza, calcolando la seguente funzione discriminante:

$$F = r^2 \cdot \frac{n-2}{1-r^2}$$

Tabella 31.11 Valori di *r* per un test a una coda

ν	α	
	0,05	0,01
1	0,9877	0,999507
2	0,9000	0,9800
3	0,805	0,9343
4	0,729	0,882
5	0,669	0,833
6	0,621	0,789
7	0,582	0,750
8	0,549	0,715
9	0,521	0,685
10	0,497	0,658
11	0,476	0,634
12	0,457	0,612
13	0,441	0,592
14	0,426	0,574
15	0,412	0,558
16	0,400	0,543
17	0,389	0,529
18	0,378	0,516
19	0,369	0,503
20	0,360	0,492
25	0,323	0,445
30	0,296	0,409
35	0,275	0,381
40	0,257	0,358
45	0,243	0,338
50	0,231	0,322
60	0,211	0,295
70	0,195	0,274
80	0,183	0,257
90	0,173	0,242
100	0,164	0,230

Tabella 31.12 Valori di *r* per un test a due code

ν	α	
	0,05	0,01
1	0,99692	0,999877
2	0,9500	0,99000
3	0,878	0,9587
4	0,811	0,9172
5	0,754	0,875
6	0,707	0,834
7	0,666	0,798
8	0,632	0,765
9	0,602	0,735
10	0,576	0,708
11	0,553	0,684
12	0,532	0,661
13	0,514	0,641
14	0,497	0,623
15	0,482	0,606
16	0,468	0,590
17	0,456	0,575
18	0,444	0,561
19	0,433	0,549
20	0,423	0,537
25	0,381	0,487
30	0,349	0,449
35	0,325	0,418
40	0,304	0,393
45	0,288	0,372
50	0273	0,354
60	0250	0,325
70	0232	0,302
80	0,217	0,283
90	0,205	0,267
100	0,195	0,254

Si confronta poi il valore così calcolato (F_c) con il valore (F_t) ricavato dalle tabelle 31.8 o 31,9, in funzione di α e $v_1 = 1$ e di $v_2 = n - 2$. Se:

$$F_c \geq F_t$$

Si può rifiutare l'ipotesi nulla e affermare che «esiste una relazione lineare significativa al livello α» fra le due variabili.

ESEMPIO 18 **L'analisi di 20 campioni di acqua minerale provenienti da zone diverse (►fig. 31.11) ha fornito le concentrazioni di calcio e bicarbonati riportate nella seguente tabella.**

n	Ca (mg/L)	HCO_3^- (mg/L)	n	Ca (mg/L)	HCO_3^- (mg/L)
1	92,0	405,6	11	12,8	24,1
2	11,0	40,3	12	94,4	384,3
3	3,4	13,7	13	57,2	280,6
4	26,2	77,8	14	101,0	409,0
5	50,0	101,4	15	338,7	1134,6
6	52,9	103,1	16	142,5	451,5
7	28,4	122,0	17	108,0	338,7
8	4,7	18,3	18	50,0	311,1
9	39,0	190,6	19	40,1	134,2
10	22,5	74,4	20	72,0	478,3

Determinare se le concentrazioni dei due ioni sono correlabili per $p = 0{,}95$.

$\Sigma x_i = 1346{,}8$ $\Sigma y_i = 5093{,}6$ $\Sigma x_i^2 = 195965{,}30$
$\Sigma y_i^2 = 2\,594\,105{,}66$ $\Sigma x_i \cdot y_i = 697\,552{,}75$

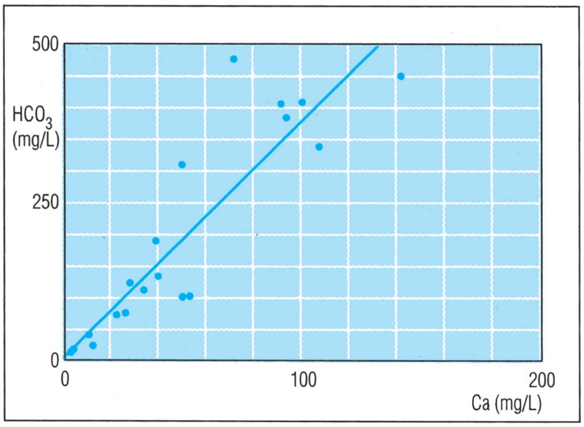

Figura 31.11
Analisi di 20 campioni di acqua minerale: diagramma della concentrazione di HCO_3^- in funzione di quella di Ca^{2+}.

Ipotesi (test a due code)

H_0: i valori delle due coppie di variabili sono indipendenti
H_1: i valori sono legati da una relazione lineare positiva o negativa

Il test è a due code, per $\alpha = 0{,}05$. In base alla (18.29*c*), il coefficiente di correlazione vale:

$$r = \frac{\left[697\,552{,}75 - \dfrac{1346{,}8 \cdot 5093{,}6}{20}\right]}{\sqrt{\left[195\,965{,}3 - \dfrac{(1346{,}8)^2}{20}\right] \cdot \left[2\,594\,105{,}66 - \dfrac{(5093{,}6)^2}{20}\right]}} = 0{,}9596$$

- Test di verifica di r

Si confronta il valore calcolato (r_c) con quello tabulato ($r_t = 0{,}444$) in corrispondenza di $\alpha = 0{,}05$, per $v = 18$ (*v.* tabella 31.12).

Dato che $r_c > r_t$, l'ipotesi nulla è respinta e si può affermare che «esiste una stretta correlazione fra la concentrazione di ioni calcio e bicarbonato nelle acque minerali prese in considerazione». Per approfondire il discorso si potrebbe ipotizzare che le acque in questione solubilizzino il bicarbonato di calcio durante il loro percorso sotterraneo.

- Test di verifica con t

Si calcola il valore della funzione discriminante:

$$t_c = 0{,}9596 \cdot \sqrt{\frac{20-2}{1-(0{,}9596)^2}} = 14{,}47$$

Si confronta il valore calcolato con quello tabulato ($t_t = 2{,}10$) in corrispondenza di $\alpha = 0{,}05$ e $v = 18$ (*v.* tabella 31.6).

Dato che $r_c > t_t$, si può rifiutare l'ipotesi nulla e affermare che «esiste una relazione lineare significativa al livello α» fra le due variabili.

- Test di verifica con F

Si calcola il valore della funzione discriminante:

$$F_c = (0{,}9596)^2 \cdot \frac{20-2}{1-(0{,}9596)^2} = 209{,}36$$

Si confronta poi il valore calcolato con quello tabulato ($F_t = 5{,}98$) in corrispondenza di $\alpha = 0{,}05$, $v_1 = 1$ e $v_2 = 18$ (*v.* tabella 31.8).

Dato che $F_c > F_t$, si può rifiutare l'ipotesi nulla e affermare che «esiste una relazione lineare significativa al livello α» fra le due variabili.