▶ Intervallo di fiducia del coefficiente angolare e dell'intercetta

L'intervallo di fiducia del coefficiente angolare (b_1) è dato da:

$$b_1 \pm t_{\alpha;\,\nu} \frac{\sqrt{\dfrac{n-1}{n-2}\left(s_y^2 - b_1^2 s_x^2\right)}}{\sqrt{\sum x_i^2 - \dfrac{\left(\sum x_i\right)^2}{n}}} \tag{22.24}$$

dove • s^2y è la varianza dei valori di y • b_1 il coefficiente angolare dell'equazione di regressione • s^2x la varianza dei valori di x • n il numero delle coppie di valori (x; y) • $t_{\alpha;\,\nu}$ il t di Student per una probabilità α pari a 0,05 (test a due code) e ν (= $n - 2$) gradi di libertà.

L'intervallo di fiducia dell'intercetta (b_0) è dato da:

$$b_0 \pm t_{\alpha;\,\nu} \cdot \sqrt{\frac{n-1}{n-2}\left(s_y^2 - b_1^2 s_x^2\right) \cdot \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum x_i^2 - \dfrac{\left(\sum x_i\right)^2}{n}}\right]}$$

L'intervallo di fiducia del coefficiente angolare consente di verificare quanto esso sia significativamente diverso da zero (nessuna correlazione fra x e y); l'intervallo di fiducia dell'intercetta, invece, consente di verificare se la retta passa per lo zero, così come impongono alcune leggi (come per esempio la legge di Beer).

ESEMPIO 21

In base ai dati dell'Esempio 19, calcolare l'intervallo di fiducia del coefficiente angolare e dell'intercetta per α = 0,05 (ovvero p = 0,95) e ν = 5 (cui corrisponde, in base alla tabella 22.6, t = 2,571).

In base alla (31.21) l'intervallo di fiducia del coefficiente angolare vale:

$$0{,}6816 \pm 2{,}571 \cdot \frac{\sqrt{\dfrac{6}{5}\left[0{,}0366899 - 0{,}6816^2 \cdot 0{,}07886667\right]}}{\sqrt{1{,}5379 - \dfrac{2{,}73^2}{7}}} = 0{,}6816 \pm 0{,}02919$$

In base alla (31.22) l'intervallo di fiducia dell'intercetta vale:

$$0{,}007464 \pm 2{,}571 \cdot \sqrt{\frac{6}{5}\left[0{,}0366899 - 0{,}6816^2 \cdot 0{,}07886667\right] \cdot \left[\frac{1}{7} + \frac{0{,}39^2}{1{,}5379 - \dfrac{2{,}73^2}{7}}\right]} = : 0{,}007464 \pm 0{,}01368$$

Il coefficiente angolare b_1 risulta dunque significativamente diverso da zero e quindi si può ribadire (dal momento che R^2 = 0,9986), che esiste una relazione lineare fra le due variabili.

Inoltre, siccome nell'intervallo di fiducia dell'intercetta (che va da −0,0062 a 0,0211) **è compreso lo zero** e dato che la legge di Beer prevede che la retta di taratura passi per tale punto, si può procedere al calcolo del coefficiente angolare della *retta forzata a passare per lo zero*.

Test per verificare se l'intercetta è diversa da zero

Nella legge di Beer non compare l'intercetta (b_0) e quindi, in teoria, tutte le rette sperimentali dovrebbero passare per lo zero. In pratica ciò non si verifica quasi mai, ma è interessante controllare quanto l'intercetta sia significativamente diversa da zero, anche per verificare eventualmente l'opportunità di adottare un modello di regressione che «forzi» la retta di regressione a passare per tale punto.

Ipotesi

$H_0 : b_0 = 0$ *test a due code*
$H_1 : b_0 \neq 0$

Funzione discriminante

$$t = \frac{|0 - b_0|}{\sqrt{\frac{n-1}{n-2}\left(s_y^2 - b_1^2 s_x^2\right) \cdot \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}}\right]}}$$

Decisione

Si confronta il valore calcolato, t_c con quello, t_t riportato nella tabella 22.6 per $\alpha = 0{,}05$ e $v = n - 2$, se:

$$t_c \leq t_t$$

si può accettare l'ipotesi nulla e considerare l'intercetta uguale a zero.

Il test è utile per verificare, per esempio, se un retta passa effettivamente per lo zero, ma anche per controllare se la retta di regressione, applicata al metodo dell'aggiunta multipla, porta a un risultato significativamente diverso da zero (perché solo in tal caso si può considerare che l'analita sia presente nel campione).

ESEMPIO 22

In base ai dati dell'Esempio 19, verificare se il termine noto (b_0) è significativamente diverso da 0, per $\alpha = 0{,}05$ e $v = 5$.

In base alla (22.28) la funzione discriminante vale:

$$t_c = \frac{0 - 0{,}0074643}{\sqrt{\frac{6}{5}\left[0{,}00366899 - 0{,}6816^2 \cdot 0{,}07886667\right] \cdot \left[\frac{1}{7} + \frac{0{,}39^2}{1{,}5379 - \frac{2{,}73^2}{7}}\right]}} = 1{,}41$$

Dalla tabella 22.6 si ricava, per $\alpha = 0{,}05$ e $\nu = 5$ il valore $t = 2{,}571$. Siccome $t_c < t_t$, si può accettare l'ipotesi nulla e considerare che la retta passi per lo zero.

Test per controllare se un dato è aberrante

Se si suppone che una coppia di dati di una retta di taratura sia aberrante, si può effettuare un semplice controllo usando il test F di Fischer-Snedecor sui valori della varianza residua $s^2_{y/x}$ calcolata prima di eliminare il dato sospetto e dopo averlo eliminato:

$$s^2_{y/x} = \frac{\sum[y_i - (b_1 \cdot x_i + b_0)]^2}{n-2} \tag{22.27}$$

Ipotesi

$H_0: s^2_{\mathrm{y/x,\,dopo}} = s^2_{\mathrm{y/x,\,prima}}$ *test a 2 code*
$H_1: s^2_{\mathrm{y/x,\,dopo}} \neq s^2_{\mathrm{y/x,\,prima}}$

Funzione discriminante

$$F = \frac{\left(\nu \cdot s^2_{y/x}\right)_{prima} - \left(\nu \cdot s^2_{y/x}\right)_{dopo}}{\left(\nu \cdot s^2_{y/x}\right)_{dopo}} \tag{22.28}$$

Decisione

Nella tabella 22.8, si cerca il valore F_t secondo il livello di probabilità prescelto. Tale valore si colloca all'incrocio fra la colonna verticale che corrisponde a ν=1 e la riga orizzontale che corrisponde ai gradi di libertà della serie di dei dati senza il dato sospetto (dopo). Se:

$$F_c \le F_t$$

allora l'ipotesi nulla è da ritenersi valida, ovvero le due varianze possono essere considerate simili e il dato non può essere eliminato. Se invece:

$$F_c > F_t$$

le due varianze sono da ritenersi diverse e il dato si può eliminare.

ESEMPIO 23

Supponiamo che un dato della serie considerata nell'esempio 19 sia diverso e sospetto perché potrebbe giacere al di fuori della retta: ad esempio il 4° dato (evidenziato in neretto):

x (mg/L Cr)	0	0,13	0,26	0,39	**0,52**	0,65	0,78
y (A_{450})	0	0,095	0,194	0,283	**0,313**	0,444	0,540

Per stabilire se il dato sospetto possa fare parte o meno della serie si calcolano b_0 e b_1 considerando dapprima tutti i valori (prima) e poi togliendo il dato sospetto (dopo).

In base alle formule già viste per l'esempio 19:

$b_{0\mathrm{prima}} = 0{,}005893$ $b_{1\mathrm{prima}} = 0{,}6695$ n = 7 $\nu = 5$
$b_{0\mathrm{dopo}} = 0{,}007677$ $b_{1\mathrm{dopo}} = 0{,}6832$ n = 6 $\nu = 4$

Poi si calcola la deviazione standard residua per ogni retta:

$s^2_{\mathrm{y/x,}\,prima} = 4{,}6519 \cdot 10^{-4}$
$s^2_{\mathrm{y/x,}\,dopo} = 0{,}5519 \cdot 10^{-4}$

E infine F:

$$F = \frac{\left(5 \cdot 4{,}6519 \cdot 10^{-4}\right) - \left(4 \cdot 0{,}5519 \cdot 10^{-4}\right)}{\left(4 \cdot 0{,}5519 \cdot 10^{-4}\right)} = 9{,}54$$

Siccome F tabulato, per $\alpha = 0{,}05$ e $\nu_{\text{dopo}} = 4$ è pari a 12,2, il dato non può essere eliminato*.

* Se la retta avesse avuto più di 8 punti invece che 7, il dato sarebbe stato considerato aberrante…

Intervallo di fiducia di un valore di y

Nella pratica di laboratorio accade spesso che una retta di taratura venga preparata ogni volta che si preparano i reagenti di fresco e che nelle sessioni analitiche successive non venga rifatta tutta la retta di taratura, ma si prepari uno *standard di controllo* (meglio se replicato 3 o 5 volte) verificando che esso fornisca un segnale che confermi la validità della retta determinata in precedenza. In questo caso è utile verificare se tale misura (o la media delle misure) cade, o meno, nell'intervallo di fiducia della retta.

In altri termini, si prepara uno standard di concentrazione nota e si controlla che il segnale misurato (y_c) sia compreso nell'intervallo di fiducia, calcolato come segue:

$$b_1 x + b_0 \pm t_{\alpha;\,\nu} \cdot \sqrt{\frac{n-1}{n-2}\left(s_y^2 - b_1^2 s_x^2\right)} \cdot \sqrt{\frac{1}{m} + \frac{1}{n} + \frac{(x-\bar{x})^2}{\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}}} \qquad (22.29)$$

dove • m è il numero di repliche di y_c (che corrisponde al numero di soluzioni standard a concentrazione x preparate per il controllo) • $\alpha = 0{,}05$ (test a due code).

In base alla (18.40) si può notare come l'intervallo di fiducia tende a diventare più piccolo:

- se n, ovvero il numero di soluzioni standard a diversa concentrazione usate per preparare la retta, è elevato;
- se m, ovvero il numero di repliche della misura dello standard di controllo, è elevato (soprattutto maggiore di l);
- se il valore di $(x - \bar{x})$ è basso (ovvero se la concentrazione dello standard di controllo cade vicino al centro della retta);
- se la pendenza della retta (b_1) è elevata.

ESEMPIO 24

Dopo avere preparato la retta di taratura dell'Esempio 19, e avere effettuato una serie di analisi, si vuole procedere a una nuova analisi a distanza di due settimane dalla precedente, usando gli stessi reagenti. Per verificare la possibilità di riutilizzare la stessa retta, si preparano 3 soluzioni standard con una concentrazione di 0,45 mg/L e si misura un'assorbanza media di 0,331.

Per controllare se la retta è ancora valida si determina l'intervallo di fiducia del valore di assorbanza ricavato dalla retta per $x = 0{,}45$, $\alpha = 0{,}05$ e $v = 5$ ($t = 2{,}571$, ricavato dalla tabella 22.6):

$$0{,}6816\cdot 0{,}45+0{,}0074643\pm 2{,}571\cdot\sqrt{\frac{6}{5}\left(0{,}0366899-0{,}6816^{2}\cdot 0{,}007886667\right)}\cdot$$

$$\cdot\sqrt{\frac{1}{3}+\frac{1}{7}+\frac{0{,}45-0{,}39}{1{,}5379-\frac{2{,}73^{2}}{7}}}=0{,}314\pm 0{,}015$$

Dato che il valore di assorbanza medio misurato (0,331) è maggiore del limite superiore dell'intervallo di fiducia (0,329), si deve ricontrollare la retta di taratura perché è probabile che i reagenti siano scaduti o degradati.

▶ Intervallo di fiducia di un valore di x ricavato da un valore misurato di y

Da una retta di taratura generalmente si ricava la concentrazione incognita di un campione (sull'asse x), a partire da una lettura sperimentale (riportata lungo l'asse y) che *può* essere singola oppure la media di letture replicate più volte. L'intervallo di fiducia del valore di x ricavato da una misura di y_m è dato da:

$$\frac{y_m-b_0}{b_1}\pm\frac{t_{\alpha;\nu}\cdot\sqrt{\frac{n-1}{n-2}\left(s_y^2-b_1^2s_x^2\right)}}{b_1}\cdot\sqrt{\frac{1}{m}+\frac{1}{n}+\frac{\left(\frac{y_m-b_0}{b_1}-\bar{x}\right)^2}{\sum x_i^2-\frac{\left(\sum x_i\right)^2}{n}}}\qquad(22.30)$$

dove • m è il numero di repliche delle misure di y • y la media di tali misure e α =0,05 (test a due code).

Anche in questo caso valgono le considerazioni che abbiamo espresso alla fine del paragrafo precedente.

ESEMPIO 25

Un campione di acqua di discarica inquinata sottoposta all'analisi del Cr(VI), ha fornito un'assorbanza di 0,054 a 540 nm (lettura singola). Determinare la concentrazione incognita.

Usando la retta ottenuta con il metodo dei minimi quadrati (*vedi* es. 19) si ottiene il valore:

$$C\left(mg/\mathrm{L_{Cr}}\right)=\frac{0{,}054-0{,}0075}{0{,}6816}=0{,}068$$

In base alla (31.30) l'intervallo di fiducia del risultato per $\alpha = 0{,}05$ vale:

$$\frac{2{,}571\cdot\sqrt{\frac{6}{5}\cdot\left(0{,}0366899-0{,}6816^{2}\cdot 0{,}07886667\right)}}{0{,}6816}\cdot\sqrt{1+\frac{1}{7}+\frac{\left(\frac{0{,}054-0{,}07464}{0{,}6816}-0{,}39\right)^{2}}{1{,}5379-\frac{2{,}73^{2}}{7}}}=$$

$$=0{,}0378$$

L'intervallo di fiducia del risultato è quindi:

$$C\ (\text{mg/L Cr}) = 0{,}068 \pm 0{,}038$$

Supponiamo ora di avere misurato l'assorbanza di tre diverse soluzioni del campione, invece di una sola, e che 0,054 sia la media delle tre misure effettuate. L'intervallo di fiducia, in questa ipotesi, risulta pari a:

$$\frac{2{,}571 \cdot \sqrt{\frac{6}{5} \cdot \left(0{,}0366899 - 0{,}6816^2 \cdot 0{,}07886667\right)}}{0{,}6816} \cdot \sqrt{\frac{1}{3} + \frac{1}{7} + \frac{\left(\frac{0{,}054 - 0{,}07464}{0{,}6816} - 0{,}39\right)^2}{1{,}5379 - \frac{2{,}73^2}{7}}} =$$

$$= 0{,}030$$

Questo intervallo, come del resto ci si poteva aspettare, è meno ampio di quello calcolato nel caso di una misura singola.

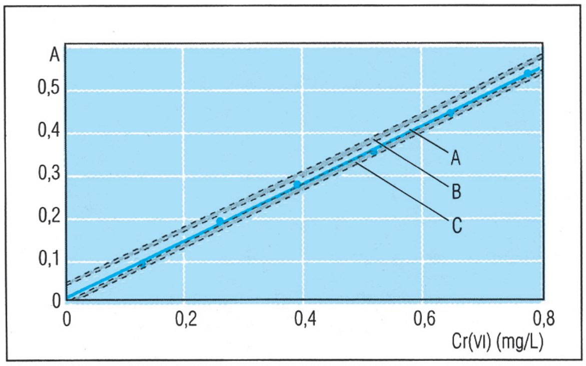

Figura 22.13
Retta di taratura di Cr(VI), ottenuta in base ai dati riportati nell'Esempio 19.
(A) Retta di regressione.
(B) Intervallo di fiducia della retta (per $\alpha = 0{,}05$), che include il valore vero di y.
(C) Intervallo di fiducia della retta (per $\alpha = 0{,}05$) dei valori di x ricavati da singole misure di y.

Il valore anomalo è il primo della serie. Nella prima colonna sono riportate le formule per il calcolo di r a seconda del diverso numero (n) di valori che costituiscono la serie da controllare.

Tabella 22.11

Valori critici di r per $\alpha = 0{,}05$ e $\alpha = 0{,}01$

Formula	n	r ($\alpha = 0{,}05$)	r ($\alpha = 0{,}01$)
$r = \dfrac{x_2 - x_1}{x_n - x_1}$	3	0,941	0,988
	4	0,765	0,899
	5	0,642	0,780
	6	0,560	0,698
	7	0,507	0,673
$r = \dfrac{x_2 - x_1}{x_{n-1} - x_1}$	8	0,554	0,683
	9	0,512	0,635
	10	0,477	0,597
$r = \dfrac{x_3 - x_1}{x_{n-1} - x_1}$	11	0,576	0,679
	12	0,546	0,642
	13	0,521	0,615
$r = \dfrac{x_3 - x_1}{x_{n-2} - x_1}$	14	0,546	0,641
	15	0,525	0,616
	16	0,507	0,595
	17	0,490	0,577
	18	0,475	0,561
	19	0,462	0,547
	20	0,450	0,535
	21	0,440	0,524
	22	0,430	0,514
	23	0,421	0,505
	24	0,413	0,497
	25	0,406	0,489
	26	0,399	0,486
	27	0,393	0,475
	28	0,387	0,469
	29	0,381	0,463
	30	0,376	0,457

I valori anomali sono il primo e l'ultimo della serie. Nella prima colonna sono riportate le formule per il calcolo di r a seconda del diverso numero (n) di valori che costituiscono la serie da controllare.

Tabella 22.12

Valori critici di r per $\alpha = 0{,}05$ e $\alpha = 0{,}01$

Formula	n	r	
		$\alpha = 0{,}05$	$\alpha = 0{,}01$
per il primo valore: $r = \dfrac{x_2 - x_1}{x_{n-1} - x_1}$ per l'ultimo valore: $r = \dfrac{x_n - x_{n-1}}{x_n - x_2}$	4	0,955	0,991
	5	0,807	0,916
	6	0,689	0,805
	7	0,610	0,740
	8	0,554	0,683
	9	0,512	0,635
	10	0,477	0,597
per il primo valore: $r = \dfrac{x_2 - x_1}{x_{n-1} - x_1}$ per l'ultimo valore: $r = \dfrac{x_n - x_{n-1}}{x_n - x_2}$	11	0,450	0,566
	12	0,428	0,541
	13	0,410	0,520
	14	0,395	0,502
	15	0,381	0,486
	16	0,369	0,472
	17	0,359	0,460
	18	0,349	0,449
	19	0,341	0,439
	20	0,334	0,430
	21	0,327	0,421
	22	0,320	0,414
	23	0,314	0,407
	24	0,309	0,400
	25	0,304	0,394
	26	0,299	0,389
	27	0,295	0,383
	28	0,291	0,378
	29	0,287	0,374
	30	0,283	0,369

La serie presenta una coppia di valori anomali all'inizio o alla fine. Nella prima colonna sono riportate le formule per il calcolo di r a seconda del diverso numero (n) di valori che costituiscono la serie da controllare.

Tabella 22.13

Valori critici di r per $\alpha = 0,05$ e $\alpha = 0,01$

Formula	n	r	
		$\alpha = 0,05$	$\alpha = 0,01$
$r = \dfrac{x_3 - x_1}{x_n - x_1}$	4	0,967	0,992
	5	0,845	0,929
	6	0,736	0,836
	7	0,661	0,778
	8	0,607	0,710
	9	0,565	0,667
	10	0,531	0,632
	11	0,504	0,603
	12	0,481	0,579
	13	0,461	0,557
	14	0,445	0,538
	15	0,430	0,522
	16	0,418	0,508
	17	0,406	0,495
	18	0,397	0,484
	19	0,379	0,473
$r = \dfrac{x_3 - x_1}{x_n - x_1}$	20	0,372	0,464
	21	0,365	0,455
	22	0,358	0,447
	23	0,352	0,440
	24	0,347	0,434
	25	0,343	0,438
	26	0,338	0,422
	27	0,334	0,417
	28	0,330	0,412
	29	0,326	0,407
	30	0,322	0,402

Tabella 22.14

Valori critici di r per $\alpha = 0{,}05$ e $\alpha = 0{,}01$

Formula	n	r ($\alpha = 0{,}05$)	r ($\alpha = 0{,}01$)
per il primo valore: $r = \frac{x_3 - x_1}{x_{n-1} - x_1}$ per l'ultimo valore: $r = \frac{x_n - x_{n-2}}{x_n - x_3}$	5	0,976	0,995
	6	0,872	0,951
	7	0,780	0,885
	8	0,710	0,829
	9	0,657	0,776
	10	0,612	0,726
	11	0,576	0,679
	12	0,546	0,642
	13	0,521	0,615
	14	0,501	0,593
	15	0,483	0,574
	16	0,467	0,557
	17	0,453	0,542
	18	0,440	0,529
	19	0,428	0,517
	20	0,419	0,506
	21	0,410	0,496
	22	0,402	0,487
	23	0,395	0,479
	24	0,388	0,471
	25	0,382	0,464
	26	0,376	0,457
	27	0,370	0,450
	28	0,365	0,444
	29	0,360	0,438
	30	0,355	0,433

La serie presenta una coppia di valori anomali a un estremo e un singolo valore all'altro. Nella prima colonna sono riportate le formule per il calcolo di *r* a seconda del diverso numero (*n*) di valori che costituiscono la serie da controllare.