
6. EL CALIBRATGE DELS INSTRUMENTS DE MESURA ANALÍTICA:¹ CALIBRATGE UNIVARIANT

Romà Tauler i Anna Izquierdo-Ridorsa*

6.1. INTRODUCCIÓ

El concepte de calibratge és bàsic en qualsevol procés de mesura. Mesurar és essencialment un procés de comparació pel qual uns valors desconeguts que s'han de determinar es comparen amb uns estàndards o patrons coneguts. Aquesta comparació pot ésser directa, com en el cas d'una balança amb dos braços iguals, o indirecta, quan s'utilitza un instrument o una escala prèviament calibrada. El propòsit del calibratge és eliminar o minimitzar el biaix (la manca d'exactitud) en el procés de mesura. En canvi, la precisió d'un sistema ja calibrat o la d'un sistema no calibrat pot ser la mateixa.

En les mesures químic-analítiques, el calibratge es refereix al procés mitjançant el qual la resposta d'un sistema de mesura es relaciona amb la concentració o quantitat d'analit d'interès. Els diferents tipus de calibratge es poden classificar en: 1) *directe*, quan es coneix el model que relaciona les variables (concentracions) amb les respostes instrumentals (on s'inclouen els mètodes absoluts, és a dir, aquells en què es coneix de manera absoluta la llei que els relaciona, per exemple, la relació estequiomètrica en una reacció química); 2) *indirecte*, quan el model no es coneix i s'ha

1. Al llarg de la present discussió, els exemples i les notacions emprades es refereixen al calibratge d'instruments de mesura en la química analítica. Aquesta és una branca de la química on l'aplicació de les eines estadístiques i numèriques per a millorar les tècniques de mesura i de calibratge té una gran importància.

* Departament de Química Analítica, Universitat de Barcelona, Diagonal 647, 08028 Barcelona.

d'estimar estadísticament; 3) *de patró intern*, quan no es poden controlar totes les variables del procés de mesura instrumental, i 4) *d'addició estàndard*, quan hi són presents els efectes matriu.

La dependència entre dues variables (per exemple, les concentracions i les respostes analítiques) pot ser de diferents tipus: 1) *funcional*, quan no hi ha errors de tipus aleatori; 2) *de regressió*, quan una de les dues variables és de tipus aleatori i l'altra es considera exacta (valor fix), i 3) *de correlació*, quan ambdues variables són aleatòries.

En el mètode de calibratge tradicional, es considera que la dependència entre les variables del calibratge és del tipus de regressió, on la concentració se suposa que és la variable independent (lliure d'error aleatori) i la resposta analítica és la variable dependent. Generalment, s'assumeix que la distribució de les respostes analítiques té un comportament normal o gaussià. Això no és així en tots els casos, com, per exemple en les mesures radioquímiques de comptatge, on la distribució de les mesures no és normal (gaussiana) sinó del tipus de Poisson. Normalment, quan ens acostem al límit de detecció, les mesures s'aproximen millor a una distribució del tipus *log-normal* que a una distribució normal. Es pot determinar si les mesures instrumentals segueixen o no una distribució gaussiana aplicant el test de David [1].

Suposant que hem escollit el tipus de calibratge per regressió, hi ha una sèrie d'etapes a considerar en el procés de calibratge:

- 1) Adopció del model.
- 2) Establiment del pla experimental.
- 3) Realització de les mesures.
- 4) Càlcul de l'estimació dels paràmetres desconeguts.
- 5) Càlcul de la variància de l'ajust i de la dels paràmetres.
- 6) Verificació de la validesa de les hipòtesis assumides en el model proposat.

Generalment, es considera que el model és de tipus lineal, ja que així ho és en la majoria dels casos, almenys en un interval petit (les corbes de calibratge analítiques no solen ser lineals a concentracions baixes i altes). El primer requeriment per al calibratge és el de disposar i utilitzar patrons o estàndards apropiats i exactes. Com a mínim, aquests patrons han de ser similars, idealment iguals, a l'objecte que es compara. En les mesures físiques, això és relativament fàcil, però en les mesures químiques això és moltes

vegades molt difícil i àdhuc impossible. Els problemes relacionats amb la manca de coincidència (*matching*) de la matriu i amb el nivell de concentració de l'analit s'han de considerar i avaluar a partir de consideracions teòriques i/o experimentals. És també important considerar el requeriment d'exactitud dels patrons emprats per al calibratge i la del període màxim entre calibratges successius. L'interval inicial de temps s'ha d'escollir sobre la base del coneixement previ o la intuïció. Basant-se en l'experiència adquirida durant la seva utilització, l'interval podrà estendre's si la metodologia es manté dins de la tolerància de cada calibratge, o s'haurà de disminuir si s'observa que surt de la tolerància. Els gràfics estadístics de control poden ajudar en el seguiment del canvi del valor mesurat d'una mostra test amb el temps. Molts tipus de mesures químiques no es basen en la comparació directa amb patrons sinó en comparacions indirectes i intermitents, en les quals l'ús dels patrons només es fa a fi d'establir una funció de resposta analítica que s'utilitzarà després en les mesures analítiques subsegüents.

El pla experimental inclou l'elecció dels patrons que cobreixin l'interval d'interès. Podem classificar els patrons en quatre tipus:

a) Patrons certificats (SMR). Són materials de referència certificats per laboratoris autoritzats per a fer-ho: NBS als EUA, BCR a la UE, i IAEA a Viena. Per a la certificació, s'utilitzen mètodes de la màxima exactitud, diferents mètodes independents i les anàlisis interlaboratoris. Generalment són cars.

b) Patrons de matriu sintètica. Són patrons que només reproduiran els elements majoritaris de la matriu de la mostra a analitzar.

c) Patrons model, en els quals s'escull una substància model.

d) L'analit o el compost a analitzar en el seu estat pur. És el cas de simplificació màxima, on es preparen solucions de l'analit pur en el dissolvent de la mostra.

En l'anomenat *calibratge univariant* s'estudia la relació que hi ha entre dues variables x i y . En el mètode tradicional de calibratge la resposta instrumental R és la variable dependent y , i la concentració o quantitat d'analit en els patrons, c , és la variable independent o variable de control x . Se suposa que la variable x no està subjecta a errors (la qual cosa és justificable quan la concentració dels analits en els patrons es coneix amb força exactitud).

De manera general, els models que descriuen la relació que hi ha entre x i y poden representar-se amb la funció:

$$y = f(x, \beta_0, \beta_1, \dots, \beta_m),$$

on les β són els paràmetres de la funció. Es distingeix entre models lineals i no lineals; el terme *lineal* no es refereix a la forma geomètrica que pren la funció ajustada sinó a la forma matemàtica del model. En els models lineals Y és una funció lineal dels paràmetres; per exemple, $Y = \beta_0 + \beta_1x$, $Y = \beta_0 + \beta_1x + \beta_2x^2$ (model lineal de segon ordre), o $Y = \beta_0 + \beta_1x + \beta_2x^2 + \beta_3x^3$ (model lineal de tercer ordre). Moltes vegades, per a evitar la confusió, s'utilitza el terme curvilineal per a expressar aquells models lineals la representació geomètrica dels quals és en forma de corba. Exemples de models no lineals són la funció exponencial $Y = \beta_1 \exp(\beta_2x)$ o la funció de Gauss, $Y = \beta_1 \exp(-(\beta_2(x - \beta_3))^2)$.

6.1.1. Covariància i correlació

Aquests dos conceptes es refereixen a la interrelació entre dues variables, x i y (en els mètodes de calibratge analític serà entre les concentracions i les respostes o mesures instrumentals). La covariància entre x i y es troba, en la pràctica, a partir de l'expressió següent:

$$\text{Cov}(x, y) = \sum (x_i - x_m)(y_i - y_m)/(N - 1),$$

on x_m i y_m són, respectivament, les mitjanes dels valors de les x i de les y , i N és el nombre de punts. D'aquesta definició, se'n dedueix que, quan diferències grans en les x es corresponen amb diferències grans en les y , el valor de la covariància resultarà gran; quan no hi ha cap relació entre les dues variables (la variació a x no proporciona informació sobre com variarà y), el valor de la covariància serà nul o molt petit. De totes maneres, aquest valor de la covariància depèn de l'escala de les unitats de mesura de les x i de les y . Precisament per a evitar aquesta ambigüitat, es defineix així la correlació entre x i y :

$$r_{x,y} = \text{Cov}(x, y)/(s_x s_y),$$

El calibratge dels instruments de mesura analítica: calibratge univariant

on s_x i s_y són les desviacions estàndard de x i de y , respectivament. Aquest valor de r estarà comprès entre 1 i -1 (figura 1).

El concepte de matriu de covariàncies per a una sèrie de variables x_1, x_2, \dots, x_k es defineix a partir de la matriu que té per elements diagonals les variàncies i com a elements de fora de la diagonal les covariàncies.

$$Cov(x) = \begin{pmatrix} s^2(x_1) & cov(x_1, x_2) & cov(x_1, x_3) & \dots & \dots \\ cov(x_2, x_1) & s^2(x_2) & cov(x_2, x_3) & \dots & \dots \\ cov(x_3, x_1) & cov(x_3, x_2) & s^2(x_3) & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix}$$

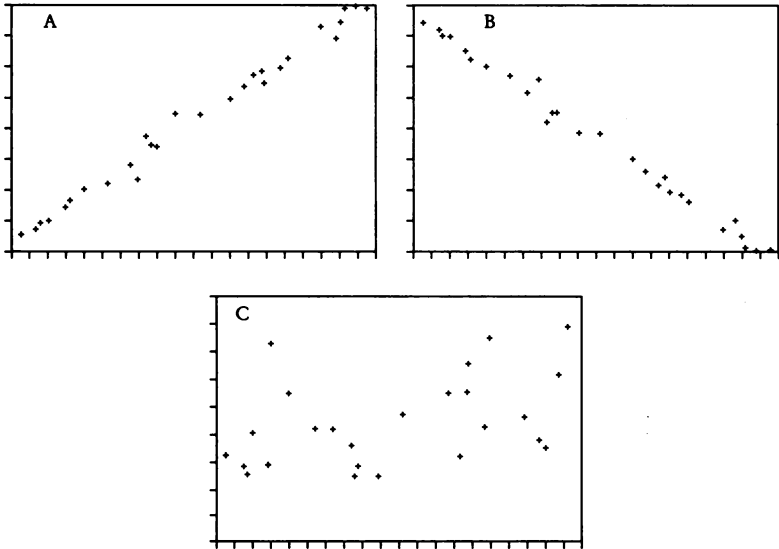


FIGURA 1. Diferents correlacions entre dues variables: A) $r \cong 1$; B) $r \cong -1$; C) $r \cong 0$

6.2. CALIBRATGE LINEAL PER MÍNIMS QUADRATS

La tècnica d'ajust de corbes per mínims quadrats permet construir la corba de calibratge (de manera no esbiaxada) amb una fa-

Del plaer dels sentits al plaer de les xifres

mília de punts $(c_1, R_1), (c_2, R_2), (c_3, R_3), \dots, (c_n, R_n)$, on les c_n es refereixen a les concentracions d'analit en les n mostres, i les R_n es refereixen a les n respostes o mesures analítiques (instrumentals) corresponents. S'han de complir els requeriments següents per tal que l'aplicació del mètode de mínims quadrats no ponderats sigui adequada:

1) Els únics errors de les mesures s'han de trobar a la variable dependent (la resposta del sistema analític, R). Si la variable independent (la concentració de l'analit, c) està subjecta a errors, aquests últims han d'ésser més petits que els corresponents a la variable dependent. És a dir, que les variàncies associades a les concentracions i a les respostes analítiques compleixen que $\sigma_c^2 = 0$, o bé que $\sigma_c^2 \ll \sigma_R^2$. Això vol dir que els errors fets en la preparació dels patrons han d'ésser negligibles en comparació amb els errors comesos en les mesures instrumentals.

2) Les variàncies en els valors de les respostes instrumentals, R , han d'ésser estadísticament iguals: $\sigma_{R1}^2 = \sigma_{R2}^2 = \dots = \sigma_{RN}^2 = \sigma_R^2$.

3) Els errors o variàncies en els valors de R són independents.

4) Les variàncies associades a les respostes instrumentals segueixen una distribució normal, $\sigma_R^2 \{N(0, \sigma^2)\}$.

Quan es compleixen aquestes condicions, es pot utilitzar el mètode de mínims quadrats no ponderats per a calcular els paràmetres de la corba de calibratge:

$$R = \beta_0 + \beta_1 c + e,$$

on R és la resposta instrumental, c és la concentració d'analit, β_0 i β_1 són els paràmetres del model lineal simple i e és l'error aleatori.

El model lineal proposat descriu que la resposta analítica és funció de la quantitat d'analit present en la mostra, a més del senyal de fons β_0 . Els paràmetres β_0 i β_1 del model no són coneguts, però pot utilitzar-se la informació proporcionada per les mesures en l'etapa de calibratge per a calcular unes estimacions b_0 i b_1 de β_0 i β_1 . Aquestes estimacions es fan de manera que la línia estimada s'ajusti tan bé com sigui possible als punts experimentals. Aquesta línia serà aquella que faci mínima la suma de quadrats dels residuals, definida de la manera següent:

El calibratge dels instruments de mesura analítica: calibratge univariant

$$SS = \sum e_i^2 = \sum (R_i - b_0 - b_1 x_i)^2.$$

La condició de mínim comporta la resolució del sistema d'equacions que es deriva de:

$$\partial SS / \partial b_0 = \partial SS / \partial b_1 = 0,$$

que té per solució:

$$b_0 = [(\sum c_i^2)(\sum R_i) - (\sum c_i)(\sum c_i R_i)] / [N(\sum c_i^2) - (\sum c_i)^2]$$

$$b_1 = [N(\sum c_i R_i) - (\sum c_i)(\sum R_i)] / [N(\sum c_i^2) - (\sum c_i)^2],$$

on N representa el nombre total d'observacions o mesures.

L'efecte que la variància en les respostes instrumentals R , σ_R^2 , té sobre les estimacions b_0 i b_1 es pot trobar per aplicació directa dels principis de la propagació d'errors; les fórmules per a la seva estimació són:

$$s_{b_0}^2 = s_R^2 \sum c_i^2 / N \sum (c_i - c_m)^2$$

$$s_{b_1}^2 = s_R^2 / (\sum c_i - c_m)^2,$$

on s_R^2 és un estimador no esbiaixat de σ_R^2 (variància dels residuals) que és donat per:

$$s_R^2 = \{\sum (R_i - R_m)^2 - b_1^2 [\sum (c_i - c_m)^2]\} / (N - 2),$$

on c_m i R_m són, respectivament, els valors mitjans de les N concentracions i les N respostes.

Es pot comprovar que l'estimació de β_0 (senyal de fons) millora en augmentar N i per a concentracions baixes dels patrons; l'estimació de β_1 millora quan es cobreix un ventall ampli de concentracions dels patrons.

Els intervals de confiança per a β_0 i β_1 associats amb una probabilitat determinada $P = 1 - \alpha$ (α , nivell de significació) són donats per:

$$b_0 - [t(\alpha/2, n - 2)]S_{b_0} < \beta_0 < b_0 + [t(\alpha/2, n - 2)]S_{b_0}$$

Del plaer dels sentits al plaer de les xifres

$$b_1 - [t(\alpha/2, n - 2)]S_{b_1} < \beta_1 < b_1 + [t(\alpha/2, n - 2)]S_{b_1}.$$

Les incerteses associades en l'estimació de β_0 i β_1 impliquen que la recta de regressió no és única. Al contrari, s'obté una banda de regressió. Per a un valor determinat de c , c^* , es pot calcular un interval de confiança per al valor de la mitjana que correspon a una resposta mitjana de rèpliques R^* ; per a un nivell de confiança $(1 - \alpha)$:

$$b_0 + b_1c^* - t_{\alpha/2}S_R [1/N + (c^* - c_m)^2/\Sigma(c_i - c_m)^2]^{1/2} < R^* < \\ < b_0 + b_1c^* + t_{\alpha/2}S_R [1/N + (c^* - c_m)^2/\Sigma(c_i - c_m)^2]^{1/2},$$

on $t_{\alpha/2}$ és el valor de la distribució t amb $N-2$ graus de llibertat i un nivell de significació α . A la figura 2 es mostra de manera esquemàtica l'anomenada *banda de regressió*. L'interval de confiança de R^* és mínim quan $c^* = c_m$.

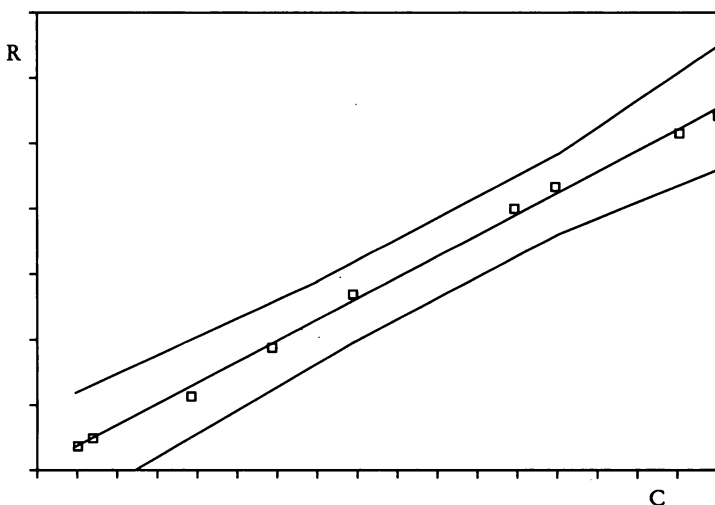


FIGURA 2. Banda de regressió.

Idealment, la corba ha de passar per l'origen; és a dir, que quan $c = 0$ no hi hauria d'haver resposta analítica significativa. Això implicaria unes condicions difícils de trobar en la pràctica, en les quals el soroll de fons experimental i instrumental, així com qualsevol ti-

El calibratge dels instruments de mesura analítica: calibratge univariant

pus d'interferència no tindrien cap efecte sobre el senyal analític, és a dir, quan $E(\beta_0) = 0$. Si això es compleix, s'ha d'emprar el model $R = \beta_1 c$ en lloc de l'equació que inclou β_0 . De totes maneres, generalment és millor incloure β_0 en el model, i en tot cas avaluar la seva significació estadística (per exemple, amb un test estadístic t de Student, per tal de veure si b_0 és estadísticament diferent de zero).

6.2.1. Generalització matricial del model lineal (anàlisi multicomponent)

Suposem que la resposta analítica no sols depèn d'un dels components de la mostra sinó també dels altres; en aquest cas, hom escriurà:

$$R_i = \beta_1 c_{i1} + \beta_2 c_{i2} + \dots + \beta_k c_{ik},$$

on ara les c_{ik} són les concentracions dels k analits en els $i = 1, 2, \dots, n$ patrons; el conjunt de respostes R_i es poden expressar en forma matricial en termes d'una matriu de coeficients β i d'una matriu de variables C :

$$\begin{pmatrix} R_1 \\ R_2 \\ R_3 \\ \vdots \\ \vdots \\ R_n \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & c_{13} & \cdot & \cdot & \cdot & c_{1k} \\ c_{21} & c_{22} & c_{23} & \cdot & \cdot & \cdot & c_{2k} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ c_{n1} & c_{n2} & c_{n3} & \cdot & \cdot & \cdot & c_{nk} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \cdot \\ \cdot \\ \beta_k \end{pmatrix}$$

o, en forma matricial reduïda:

$$R = C \beta.$$

La matriu de coeficients β pot estimar-se a partir de l'expressió:

$$b = (C'C)^{-1}C'R,$$

on C' és la matriu transposada de C i $(C'C)^{-1}C'$ és la matriu pseudo-inversa de C .

Aquest model lineal generalitzat es pot estendre a respostes multicanal o multivariants, i es tractarà amb més detall en la discussió dels mètodes de calibratge multivariant.

6.3. VALIDACIÓ DEL MODEL LINEAL

6.3.1. Coeficient de correlació r

$$r = \frac{\sum(c_i - c_m)(R_i - R_m)}{[\sum(c_i - c_m)^2 \sum(R_i - R_m)^2]^{1/2}}$$

Encara que està molt estès el càlcul del coeficient de correlació, r , no és convenient utilitzar-lo en la validació del model lineal, per diferents raons: *a*) aquests valors de r no es poden comparar quan s'han obtingut a partir de corbes amb patrons diferents, ja que r depèn de l'interval de calibratge; *b*) la forma de la recta ajustada pot canviar sense afectar r , i *c*) r no fa possible de realitzar comparacions quantitatives de la qualitat de l'ajust, ni tampoc indica si s'ajusta adequadament.

6.3.2. Test F sobre rèpliques

Si es fan rèpliques de les mesures R_i , es pot aplicar el test F següent:

$$F = s^2(\text{total})/s^2(\text{intra}) \longrightarrow F_{1-\alpha}(n-2, N-n),$$

on n és el nombre de patrons de concentració diferent i N el nombre de mesures totals; $s^2(\text{intra})$ es refereix a la variància entre rèpliques i $s^2(\text{total})$ a la variància total associada als residuals obtinguts en l'ajust minimoquadràtic. Quan el valor de F calculat és menor o igual que el tabulat per a un nivell de confiança $1-\alpha$ i un nombre determinat de graus de llibertat, no hi ha cap raó per suposar que el model no és adequat (vegeu ANOVA més avall).

6.3.3. Test t sobre b_1

Es pot aplicar un test t per tal de veure si b_1 (pendent) difereix significativament de zero, ja que en cas que b_1 no difereixi significativament de zero (hipòtesi nula) no hi ha cap raó per a suposar que hi ha dependència lineal.

6.3.4. ANOVA per a models lineals

L'aplicació de l'anàlisi de la variància, ANOVA, permet detectar la manca d'ajust en una regressió, i, per tant, conèixer si el model escollit és el correcte. La millor utilització de l'anàlisi de la variància es fa quan es tenen determinacions replicades de les mesures o respostes analítiques.

La variació total observada en les respostes instrumentals R o variables y respecte a la seva mitjana és donada per l'expressió:

$$SS_T = \sum \sum (y_{ij} - y_m)^2,$$

on y_{ij} es refereix a una de les j mesures replicades corresponents a un valor particular de x_i , i on y_m és la mitjana de totes les mesures. Es pot realitzar la descomposició

$$(y_{ij} - y_m) = (y_{ij} - y_{mi}) + (y_{mi} - y_{ci}) + (y_{ci} - y_m)$$

← residual →

on y_{mi} és la mitjana de les rèpliques y_{ij} corresponents a un valor particular de x_i , i y_{ci} és el valor de y en aquest mateix valor de x_i calculat a partir de la funció de regressió trobada. El significat de tot això queda reflectit en la representació gràfica de la figura 3.

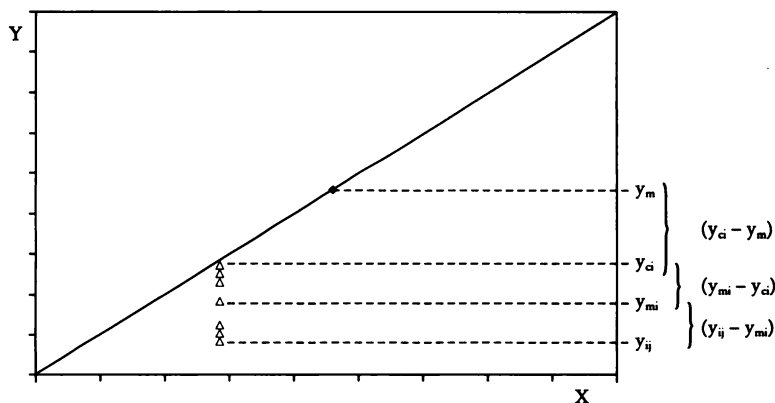


FIGURA 3. Representació de la descomposició que es porta a terme per tal de fer l'anàlisi de la variància.

Del plaer dels sentits al plaer de les xifres

En elevar al quadrat l'expressió anterior i sumar respecte a i i a j , queda:

$$SS_T = \sum \sum (y_{ij} - y_{mi})^2 + \sum n_i (y_{mi} - y_{ci})^2 + \sum n_i (y_{ci} - y_m)^2$$

$$\begin{array}{c} \longleftarrow SS_R \longrightarrow \\ \longleftarrow SS_{PE} \longrightarrow \quad \longleftarrow SS_{LOF} \longrightarrow, \quad \longleftarrow SS_{REG} \longrightarrow \end{array}$$

on SS_R representa la variació residual, SS_{REG} la variació explicada per la regressió, SS_{PE} la variació associada a l'error aleatori (rèpliques), SS_{LOF} la variació associada a la manca d'ajust (*lack of fit*), és a dir, a la variació no explicada ni pel model ni per l'error aleatori; $j = 1, 2, \dots, n_i$; n_i és el nombre d'observacions replicades sobre x_i .

D'aquesta manera, la variació total a y se separa en dos components: SS_{REG} , que és la variació que pot atribuir-se a la línia de regressió, i SS_R , que és la variació residual que mesura la variació que no pot explicar-se per la línia de regressió.

Quan es disposa de mesures replicades, la variació residual pot separar-se en dos termes més:

- 1) Un component que mesura la variació deguda a la incertesa experimental pura, que és la suma d'errors purs, SS_{PE} .
- 2) Un component que mesura la variació de les mitjanes de cada grup de rèpliques y_{mi} , respecte a la línia de regressió i que s'anomena SS_{LOF} , suma de quadrats per manca d'ajust (*lack of fit*).

Es pot construir la taula ANOVA per als models lineals de la manera següent:

Anàlisi de la variància

Causa de variació	SS suma de quadrats	g. ll. graus de llibertat	MS mitjana de quadrats	valor de F
Regressió	SS_{REG}	1	MS_{REG}	MS_{REG}/MS_R
Residual	SS_R	$n - 2$	MS_R	
Manca d'ajust	SS_{LOF}	$k - 2$	MS_{LOF}	MS_{LOF}/MS_{PE}
Error pur	SS_{PE}	$n - k$	MS_{PE}	
Total	SS_T	$n - 1$		

En aquesta taula, les mitjanes de quadrats, MS , s'obtenen per divisió de la suma de quadrats, SS , pels graus de llibertat corresponents. Aquestes mitjanes de quadrats són estimacions de les variàncies. MS_{PE} és una estimació de l'error pur σ^2 . MS_{LOF} serà una estimació de σ^2 , només en el cas d'haver escollit el model correcte.

El procediment usual per a examinar la manca d'ajust és comparar el valor de $F = MS_{LOF}/MS_{PE}$ amb la distribució F amb $k - 2$ i $n - k$ graus de llibertat. Quan el quocient és gran (quan MS_{LOF} és significativament més gran que MS_{PE}), es conclou que el model no és adequat, ja que la variació observada entre les mitjanes del grup de rèpliques i la línia ajustada no pot explicar-se en termes d'incertesa experimental pura.

Quan les dues estimacions independents de σ^2 , MS_{LOF} i MS_{PE} són del mateix ordre, el model es compleix i la variància combinada MS_R pot utilitzar-se com a estimació de σ^2 en els càlculs posteriors.

6.3.5. Estudi dels residuals

Els residuals es defineixen com la diferència entre la resposta observada, R_i , i la resposta ajustada o calculada pel model, R_C :

$$e_i = R_i - R_C; i = 1, 2, \dots, N.$$

L'estudi dels residuals pot mostrar si les assumpcions inicials es compleixen, o si el model és inadequat. En cas d'un ajust correcte, la representació gràfica dels residuals en forma d'histograma s'assemblarà a la distribució d'observacions aleatòria de mitjana zero. Moltes vegades els residuals es representen en funció de la variable independent c . Si les assumpcions inicials es compleixen, la representació dels residuals e en funció de les concentracions respectives c , resulta en una banda horitzontal (figura 4a).

Les anormalitats en el model o en les dades donen bandes de residuals diferents. Tres casos freqüents, que no donen banda horitzontal, són:

— Cas I (figura 4b): suggereix que les variàncies a R_i no són constants, per la qual cosa es necessita un procediment d'ajust de corbes per mínims quadrats ponderats (vegeu més avall).

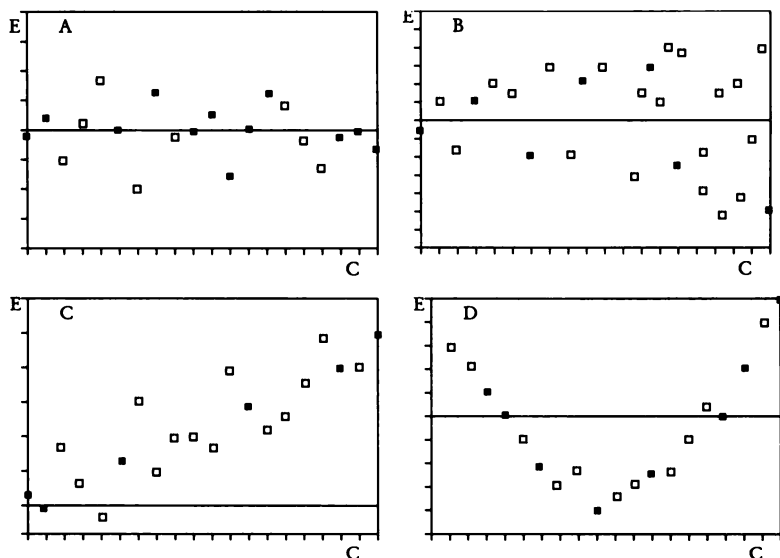


FIGURA 4. Representació dels errors respecte a la concentració (variable independent): *a*) distribució aleatòria amb homocedasticitat; *b*) distribució aleatòria amb heterocedasticitat; *c*) efecte lineal de la variable independent sobre els residuals; *d*) residuals quan el model postulat no és l'adequat.

— Cas II (figura 4c): indica que la variable independent (la concentració) té un efecte lineal sobre els residuals.

— Cas III (figura 4d): mostra que el model necessita termes addicionals de grau superior, per exemple un terme quadràtic c^2 .

Els residuals també es poden representar respecte als valors modelats de les respostes, R_c . El compliment del model i les assumpcions respecte a R_c seran certes en cas d'obtenir una banda horitzontal. Si es troben bandes com les mostrades anteriorment (residuals e respecte a les respostes calculades R_c), se'n dedueix: cas I, el mateix que quan considerem els residuals e respecte a les concentracions c ; cas II, la representació mostra desviacions del model proposat, desviacions que són negatives amb els valors baixos de R i positives amb els valors alts de R ; això pot ésser causat per errors experimentals o per falta d'algun terme en el model. El cas III suggereix la necessitat de més termes en el

El calibratge dels instruments de mesura analítica: calibratge univariant

model, o d'una transformació de la resposta R . Quan els residuals han estat obtinguts a partir d'unes dades adquirides durant un temps llarg, és recomanable fer la representació en funció del temps.

Tot el que s'ha dit, no només és vàlid respecte a l'anàlisi de les corbes de calibratge analítiques, sinó també per a qualsevol altre procediment d'examen dels residuals en proposar un determinat model matemàtic en un experiment. La construcció de l'anomenada *corba de calibratge* és un cas particular dels procediments de modelat matemàtic general i de l'estimació de paràmetres presents en moltes àrees de la ciència i de l'enginyeria.

6.4. REGRESSIÓ INVERSA. CALIBRATGE TRADICIONAL I CALIBRATGE INVERS. CALIBRATGE DIRECTE I CALIBRATGE INDIRECTE

Un cop s'ha construït la corba analítica correctament, la quantitat o concentració d'un analit en una mostra desconeguda es determina a partir del valor de la resposta obtinguda en l'anàlisi, R_s , per interpolació. Aquest procés s'anomena també *regressió inversa*.

$$R = b_0 + b_1c \longrightarrow c = (R - b_0)/b_1$$

El valor de la concentració de l'analit pot estimar-se a partir de la seva resposta R_s i a partir del model, que donarà c_s^m . El valor vertader c_s estarà acotat per dos valors, c_s^l i c_s^u . Aquests valors es troben per intersecció dels dos límits de la banda de regressió i de la línia recta $R = R_s$ (figura 5).

La precisió de l'estimació depèn de la qualitat de la línia ajustada i també de la precisió associada a la mesura R_s . El càlcul precís de l'error comès en la predicció és complex, i generalment es fa aproximadament a partir d'equacions simplificades. Quan s'utilitzen n observacions en la construcció de la corba de calibratge, i s'obtenen m respostes replicades d'una mostra desconeguda amb un valor mitjà de R_s , la quantitat d'analit a la mostra és donada per l'equació:

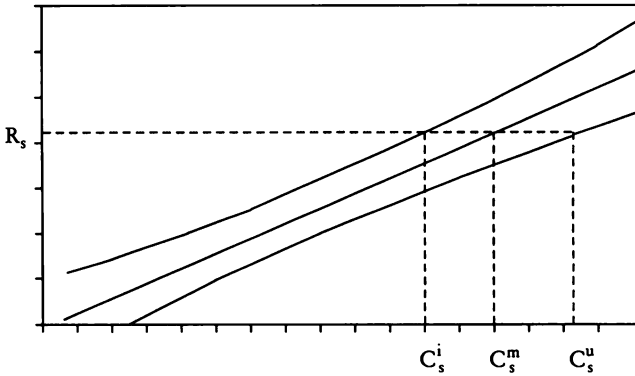


FIGURA 5. Efecte de la banda de regressió en l'estimació de c_s .

$c_s = c_s^m \pm t s_c$ (t de Student per a $n + m - 3$ graus de llibertat i un nivell de confiança $1 - \alpha$)

$$c_s^m = (R_s - b_0)/b_1$$

$$s_c = s_R/b_1 [1/m + 1/n + (R_s - R_m)/(b_1^2 \sum (c_i - c_m)^2)]^{1/2},$$

on R_m és la mitjana de les respostes de les n observacions, c_m és la mitjana de les concentracions de les n observacions, b_1 és l'estimació de β_1 i s_R^2 és la variància associada a les mesures o respostes instrumentals.

L'interval més petit de c es troba quan $R = R_m$. L'equació donada és en realitat una simplificació de l'equació rigorosa, ja que implica suposar que $s_b^2 = 0$, però pot emprar-se de manera aproximada.

6.4.1. Calibratge tradicional o clàssic versus calibratge invers

Fins ara només s'ha fet referència al procediment de calibratge clàssic o tradicional, en el qual es considera que les respostes analítiques R són les variables dependents y i que les concentracions analítiques c són les variables independents x . Aquesta distinció és una mica artificial, ja que es podria fer justament al revés. De fet, des d'un punt de vista analític, estem interessats en la

predicció de les concentracions desconegudes c a partir de les respostes mesurades R , i no a l'inrevés, com es pressuposa en el mètode clàssic o tradicional. Pel que fa als errors, hom no pot considerar sempre certa l'assumpció que aquests són més grans per a les mesures de la resposta analítica que per a les concentracions dels patrons. Hi ha casos on les concentracions dels patrons emprats també porten associades un cert error. Les diferències entre les dues aproximacions, la del calibratge clàssic i la del calibratge invers, està relacionada amb les dues maneres diferents d'utilitzar el model de calibratge, de manera causal en el procés de calibratge i de manera predictiva quan s'estimen les concentracions desconegudes: per a modelar o bé per a predir. Això equival a dir que la regressió es fa de les concentracions respecte a les respostes, o bé que es fa de les respostes respecte a les concentracions.

Per al calibratge univariant es pot escriure la predicció de la concentració de dues maneres diferents, segons que s'utilitzi el mètode clàssic o el calibratge invers:

$$c \text{ (clàssic)} = c_m + (s_c^2 / \text{cov}(R, c)) (R - R_m)$$

$$c \text{ (invers)} = c_m + (\text{cov}(R, c) / s_R^2) (R - R_m).$$

A la figura 6 es mostra un exemple de les diferents representacions gràfiques obtingudes quan s'apliquen els mètodes de regressió tradicional i de regressió inversa sobre un mateix grup de dades. Observant els residuals en ambdós casos, es pot veure que el calibratge invers dóna, en general, millors resultats, millor capacitat predictiva, menors residuals i menor error quadràtic mitjà, MSE, que el calibratge pel mètode clàssic. Les prediccions «inverses» estan estabilitzades contra el soroll de fons en les dades de calibratge, ja que totes les prediccions s'encongeixen cap a l'estimació del centre de la població. Per això, per a les mostres que tenen valors de y més grans que la mitjana, les prediccions són sistemàticament massa petites, i al contrari, per a mostres amb valors baixos de y , les prediccions són sistemàticament massa elevades. De totes maneres, les diferències entre els dos procediments de calibratge són petites i no s'han d'exagerar. Quan les dades s'ajusten bé al model, les dues aproximacions donen pràcticament els mateixos resultats.

Del plaer dels sentits al plaer de les xifres

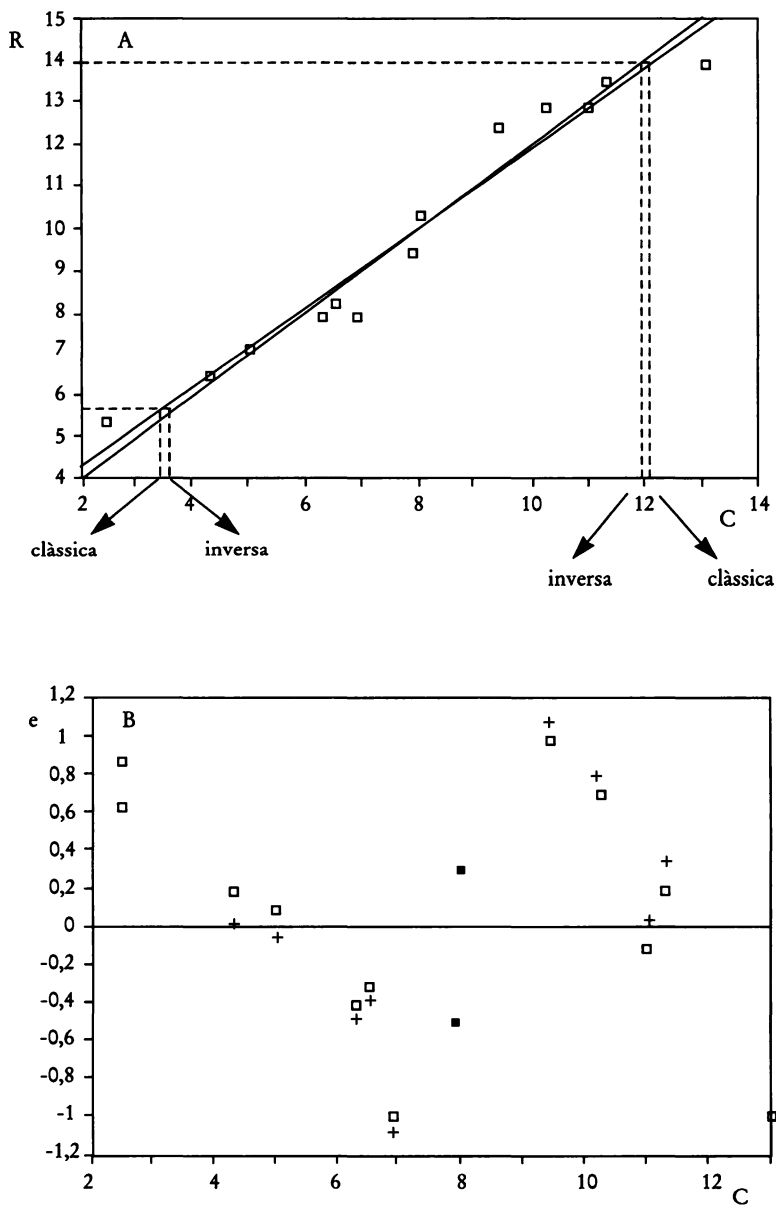


FIGURA 6. A) Representació gràfica de les regressions clàssica i inversa. B) Representació gràfica dels residuals (□ clàssica, + inversa).

6.4.2. Calibratge directe versus calibratge indirecte

La utilització de l'un o de l'altre tipus de calibratge depèn de la informació disponible prèviament. El calibratge directe significa que els paràmetres del model es coneixen *a priori* i poden utilitzar-se directament en la predicció. Per exemple, aquest és el cas quan el que es pretén és la resolució directa d'espectres coneguts que es troben superposats. Quan es coneixen les respostes analítiques de tots els constituents rellevants i les mesures es comporten de manera lineal, aleshores el model de barreja lineal pot resoldre's directament; això és el que ocorre en situacions simples d'anàlisi de solucions amb pocs constituents a concentracions baixes per espectroscòpia d'absorció; però en moltes situacions analítiques no es poden utilitzar aquests models de calibratge totalment causal, ja que hi ha interferències no identificades, respostes no lineals, o bé els diversos constituents presents interaccionen entre ells i canvien les seves característiques.

Per altra banda, el calibratge indirecte es refereix al cas quan els paràmetres del model s'han d'estimar en el mateix procés de calibratge. Aquest tipus de calibratge pot necessitar moltes dades experimentals per a permetre l'estimació estadística dels paràmetres desconeguts en el model de calibratge. Hi ha situacions intermèdies entre aquestes dues, on alguns dels paràmetres són coneguts *a priori*, mentre que d'altres s'estimen a partir de les dades de calibratge.

6.5. LÍMIT DE DETECCIÓ

A la figura 7, la intersecció de la banda de regressió i de l'eix vertical defineix un camp de respostes obtingudes a partir d'una mostra sense analit, $c = 0$, i per a un cert nivell de confiança. Les respostes menors que el punt d'intersecció donen com a resultats, per tant, concentracions negatives.

Una resposta d'un blanc R_b estarà associada a un interval $[0, c_b^u]$, on c_b^u és el límit superior del valor veritable c_b , corresponent a R_b .

$$c_b^u = c_b + t_{s_c} = 0 + t(s_R/b_1) [1/m + 1/n + (R_b - R_m)/(b_1^2 \sum (c_i - c_m)^2)]^{1/2}$$

Amb un nivell de confiança $(1 - \alpha)$, la probabilitat que un blanc produeixi un senyal més gran o igual que R_b és $\alpha/2$. Si això és un risc acceptable, aleshores la mínima quantitat detectable d'analit és

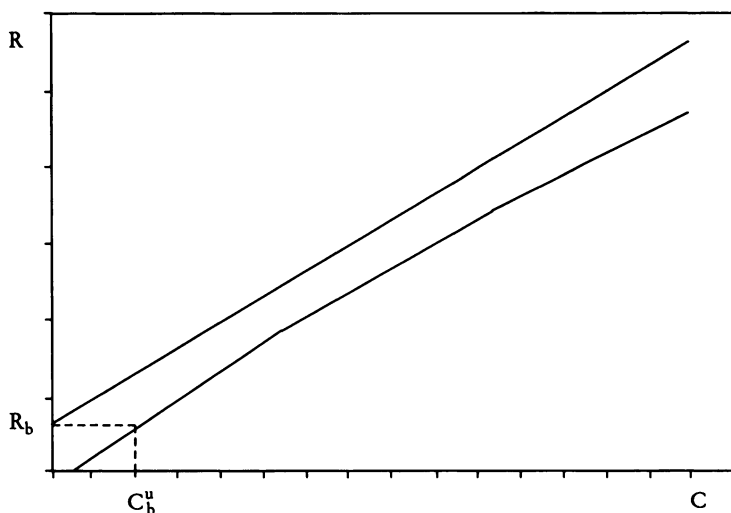


FIGURA 7. Límit de detecció.

c_b^u , que és el límit de detecció. La presència d'una quantitat d'anàlit més petita que aquesta no es pot verificar. L'interval de confiança per a c_b depèn del valor de R_b . Així, doncs, el límit de detecció pot trobar-se a partir de la quantitat c_b obtinguda per utilització de l'equació del càlcul de c , en la regressió inversa, associada a una resposta R_b , per un determinat nivell de confiança.

6.6. HETEROCEDASTICITAT I MÍNIMS QUADRATS PONDERATS (RESPOSTES AMB VARIÀNCIES DIFERENTS)

Quan ens trobem en el cas *b* de la figura 4 (anàlisi de residuals), on les variàncies a R_i no són iguals en tot el camp de concentracions, c , estudiat, es diu que les dades són heterocedàstiques. Hi ha uns valors de R que són més precisos que els altres. La utilització dels mètodes de mínims quadrats ordinaris no és vàlida. La suma de quadrats a minimitzar s'ha de ponderar.

$$SS = \sum (R_i - \beta_0 - \beta_1 c)^2 / \sigma_{R_i}^2$$

El calibratge dels instruments de mesura analítica: calibratge univariant

Els valors de β_0 i β_1 que minimitzen la funció anterior s'estimen a partir dels b_0 i b_1 , trobats amb les equacions:

$$b_1 = [(\sum w_i)(\sum w_i c_i R_i) - (\sum w_i c_i)(\sum w_i R_i)] / [(\sum w_i)(\sum w_i c_i^2) - (\sum w_i c_i)^2]$$

$$b_0 = [(\sum w_i R_i) - b_1(\sum w_i c_i)] / \sum w_i$$

on els factors de ponderació $w_i = 1/s_{Ri}^2$ són les aproximacions a $1/\sigma_{Ri}^2$. La banda de regressió obtinguda quan la variància a R augmenta en fer-ho també R es mostra a la figura 8:

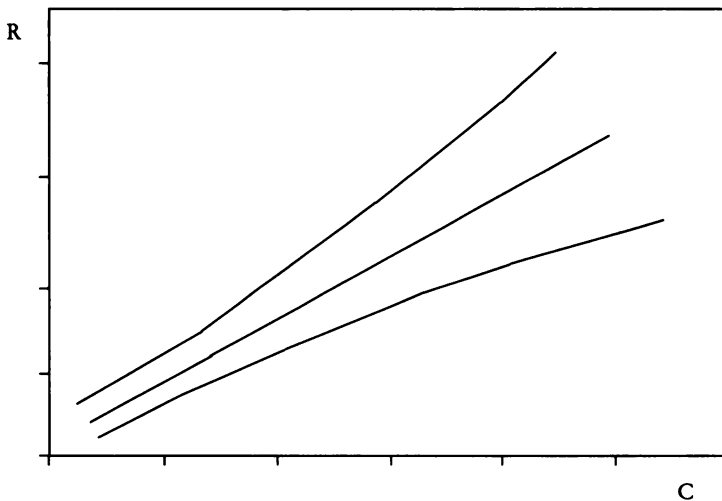


FIGURA 8. Banda de regressió per a dades heterocedàstiques.

Els intervals de confiança s'obtenen per inclusió adequada dels termes de ponderació en les equacions obtingudes prèviament pel mètode de mínims quadrats no ponderats; els pesos es poden normalitzar de manera que la seva mitjana sigui igual a una constant de normalització, per exemple, igual a 1. Això s'aconsegueix transformant cada pes antic w_i en un de nou w'_i :

$$w'_i = N (w_i / \sum w_i),$$

Del plaer dels sentits al plaer de les xifres

Una altra manera d'enfocar el problema de la manca d'uniformitat de la variància és mitjançant una transformació adequada de les dades. Aquesta transformació depèn del tipus de funció de variància, és a dir, de com varia s_R^2 en funció de la resposta. Per exemple, en el cas que la variància sigui proporcional a la resposta, una transformació arrel quadrada de les dades dona una variància constant

$$\sqrt{y} = b_0 + b_1 \sqrt{x},$$

o en el cas en què la variància és proporcional als quadrats de les respostes, és millor emprar la transformació logarítmica

$$\log y = b_0 + b_1 \log x.$$

En aquests casos, ambdues variables y i x són transformades per evitar que la línia recta es converteixi en una representació no lineal.

6.7. MODEL LINEAL QUAN LES DUES VARIABLES ESTAN SUBJECTES A ERROR. MÍNIMS QUADRATS RIGOROSOS

Els models lineals proposats fins ara no són aplicables en el cas que les dues variables estiguin sotmeses a errors. Aquesta situació ocorre, per exemple, en el cas que els errors de les dues variables estiguin correlacionats. En el model:

$$y_i = \beta_0 + \beta_1 x_i$$

trobarem els valors del b_0 i b_1 , estimacions dels paràmetres β_0 i β_1 , a partir de la minimització de la suma de quadrats SS , definida ara com:

$$SS = \sum (y_i - b_0 - b_1 x_i)^2 / (s^2_{y_i} + b_1^2 s^2_{x_i}).$$

Precisament, quan $s^2_{x_i} = 0$ i $s^2_{y_i} = s^2$ (dades homocedàstiques), ens trobem en el cas anterior dels mínims quadrats ordinaris. També, quan $s^2_{x_i} = 0$ o $s^2_{x_i} \ll s^2_{y_i}$ i les variàncies a y_i i $s^2_{y_i}$ no són iguals, ens trobem en el cas dels mínims quadrats ponderats (dades heterocedàstiques). La resolució de l'equació anterior referida al cas en què les dues variables estan sotmeses a error és més complicada que en els casos anteriors, ja que la funció mínimoquadràtica no és lineal res-

El calibratge dels instruments de mesura analítica: calibratge univariant

pecte als paràmetres b_0 i b_1 . Això complica relativament les coses, i bé s'ha de recórrer a sistemes iteratius de minimització (vegeu-ho més avall per al cas de calibratge no lineal), o bé s'han d'utilitzar certes tècniques, com la proposada per Wald, d'agrupament de conjunts de dades [2].

Hi ha casos en què l'anàlisi de regressió es fa amb dades que contenen errors, tant a les x com a les y . També en les tècniques de calibratge analític aquest pot ésser el cas, ja que els estàndards es preparen a partir de pesada, dilució, etc., i les seves concentracions poques vegades es coneixen de manera exacta. Generalment, s'ignoren aquests errors, ja que se suposa que són més petits que els associats a la mesura instrumental. De totes maneres, aquest procediment no es pot considerar totalment correcte. En el futur, probablement es farà més ús dels mètodes alternatius esmentats, que, com el seu nom indica, són més rigorosos.

6.8. CALIBRATGE NO LINEAL

Com ja s'ha esmentat, es troben amb relativa freqüència casos de calibratge no lineal en anàlisi instrumental, per exemple, en espectroscòpia d'emissió i en fluorescència de raigs X. L'aproximació més simple, en aquests casos, és la que assumeix la linealitat en intervals curts de la corba de calibratge, que queda així dividida en petites seccions [3]. Altres vegades, el que es fa és tractar d'ajustar polinomis d'ordre superior, per exemple:

$$R = \beta_0 + \beta_1c + \beta_2c^2.$$

Aquest cas, però, no es pot considerar no lineal en els paràmetres β , i és un cas de model lineal d'ordre superior a 1, o model curvilini de segon ordre. Per tal d'evitar la confusió en les denominacions, cal distingir clarament què és allò que es considera en realitat un model no lineal en els paràmetres.

6.8.1. Tipus i exemples de models lineals en els paràmetres

Anomenarem *models lineals en els paràmetres* els que responen a formulacions generals del tipus:

Del plaer dels sentits al plaer de les xifres

$$Y = \beta_0 Z_0 + \beta_1 Z_1 + \beta_2 Z_2 + \dots + \beta_p Z_p + \text{error}$$

$$\text{on } Z_0 = 1 \text{ i } Z_j = Z_j(X_1, X_2, \dots, X_k),$$

i els *models no lineals* seran els que no responen a aquesta formulació general.

Exemples:

1) Quan $p = 1$, $Z_1 = X_1$, tenim model lineal d'ordre 1 amb una variable independent:

$$Y = \beta_0 + \beta_1 X_1 + \text{error}.$$

2) Quan $p = k$, i $Z_j = X_j$, tenim un model lineal d'ordre 1 amb k variables independents o predictors:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \text{error}.$$

3) Quan $p = 3$ i $Z_1 = X$, $Z_2 = X^2$ i $Z_3 = X^3$, el model és d'ordre 3 i tenim una sola variable independent:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \text{error}.$$

4) Models lineals per transformació:

$$Z_1 = 1/X_1, Z_2 = 1/X_2, p = 2,$$

$$Y = \beta_0 + \beta_1(1/X_1) + \beta_2(1/X_2) + \text{error},$$

$$Z_1 = \ln X_1, Z_2 = \ln X_2, p = 2,$$

$$Y = \beta_0 + \beta_1 \ln X_1 + \beta_2 \ln X_2 + \text{error},$$

$$Z_1 = (X_1)^{1/2}, Z_2 = (X_2)^{1/2},$$

$$Y = \beta_0 + \beta_1(X_1)^{1/2} + \beta_2(X_2)^{1/2} + \text{error}.$$

6.8.2. Models no lineals que són intrínsecament lineals

Exemples:

1) El model multiplicatiu $Y = \alpha X_1^\beta X_2^\Sigma X_3^\mu$.

$$\ln Y = \ln \alpha + \beta \ln X_1 + \Sigma \ln X_2 + \mu \ln X_3 + \ln (\text{error})$$

2) El model exponencial $Y = e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2}$.

$$\ln Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ln (\text{error})$$

3) El model recíproc $Y = 1/(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \text{error})$.

$$1/Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \text{error}$$

6.8.3. Models que són intrínsecament no linealitzables en els paràmetres

Exemples:

$$Y = \alpha + (0,49 - \alpha) \exp(-\beta(x - 8)) + \text{error}$$

$$W = \alpha(1 - \beta e^{-kx}) + \text{error}$$

$$Y = \Phi_1 \log(\Phi_2 X + \Phi_3) + \text{error}$$

$$Y = \exp[-\Phi_1 X_1 \exp[-\Phi_2(1/X_2 - 1/620)]] + \text{error}$$

$$Y = \phi_1 X^{\phi_2} + \text{error}$$

$$Y = \phi_1(1 - \exp(-\phi_2 t)) + \text{error}$$

Per als models no linealitzables en els paràmetres, s'escriu l'expressió general:

$$Y = f(\Sigma_1, \Sigma_2, \Sigma_3, \dots, \Sigma_k; \phi_1, \phi_2, \dots, \phi_p) + \text{error}$$

$$Y = f(\Sigma, \phi) + \text{error},$$

Del plaer dels sentits al plaer de les xifres

on les Σ són les variables independents i les ϕ són els paràmetres del model no lineal. La funció quadràtica a minimitzar es defineix com sempre:

$$SS(\phi) = \sum (Y_i - f_i(\Sigma, \phi))^2,$$

on Y_i és la resposta analítica en el punt experimental i .

El valor de la funció f_i en un punt qualsevol i es pot aproximar fent servir la fórmula de Taylor (primer ordre) al voltant d'uns valors inicials dels paràmetres ϕ_0 .

$$Y_i = f_i(\Sigma, \phi_0) + \sum (\delta(f(\Sigma, \phi)) / \delta(\phi_i))_{\phi} (\phi_i - \phi_0)$$

Això és precisament un sistema d'equacions lineal que es pot resoldre per $(\phi_i - \phi_0)$. A partir d'aquesta solució es torna a realitzar una nova aproximació i a trobar un nou valor dels paràmetres ϕ , fins a arribar a una solució satisfactòria (convergència). La solució, doncs, és iterativa. Els mètodes més emprats en la resolució del problema de la regressió no lineal són els de cerca directa, el simplex seqüencial i sobretot el de Newton Raphson amb l'algorisme modificat de Marquardt [4].

6.9. DIAGNÒSTICS D'ABERRANTS I REGRESSIÓ ROBUSTA

El mètode de mínims quadrats descrit fins ara és el procediment emprat generalment en els processos de calibratge. No obstant això, la possible presència de punts «aberrants» (*outliers*) té una influència notable en les estimacions fetes per mínims quadrats i, per tant, es poden obtenir estimacions errònies del pendent i de l'ordenada a l'origen de la recta ajustada. A la figura 9 es mostra un exemple on la presència d'un punt aberrant modifica considerablement la posició de la millor recta ajustada per mínims quadrats (aquest punt es diu que produeix un efecte d'alçaprem, *leverage*, de la recta).

Pot argüir-se que els aberrants poden descobrir-se a partir de l'examen dels residuals, però això, malauradament, no sempre és cert; per exemple, a la figura anterior el punt aberrant no té un residual que sigui especialment més gran que el dels altres punts experimentals. Una solució possible a aquest problema és realitzar un

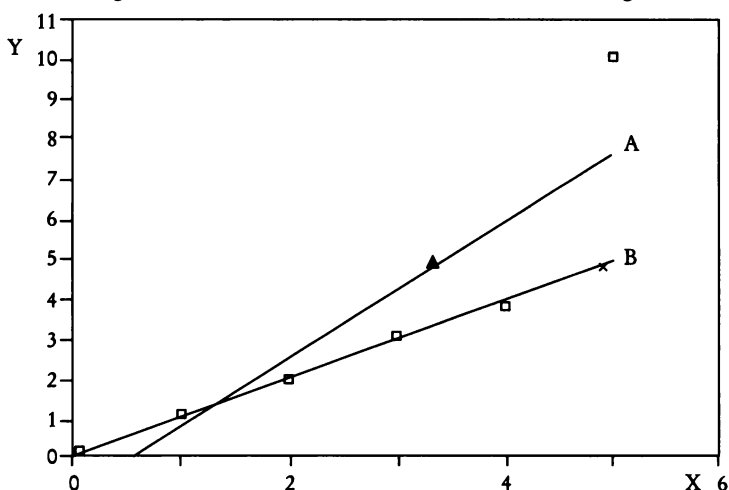


FIGURA 9. Rectes de calibratge obtingudes a partir de les dades de la taula 1 (conjunt de dades 1): a) tenint en compte el resultat ($y = 0,08 + 0,96x$), b) sense tenir-lo en compte ($y = -0,90 + 1,69x$).

diagnòstic d'aberrants. Això es pot fer a partir d'estadístics que centren la seva atenció en les observacions que tenen una influència gran sobre els estimadors minimoquadràtics. Un exemple d'aquests diagnòstics és el que es basa en la distància al quadrat de Cook, el qual mesura la variació dels coeficients de regressió que es produeix si s'omet la observació i :

$$CD^2(i) = [\sum(y_{ij} - y_{ij}(i))^2] / 2s_y^2 \text{ per a tot } j \neq i,$$

on (i) a CD^2 indica que és l'observació i la que s'exclou en el càlcul de CD ; y_{ij} i $y_{ij}(i)$ són, respectivament, els valors calculats de y_j , quan es consideren tots els punts de la recta i quan s'exclou el punt i . Així es calculen tantes CD^2 com observacions i es tenen. La distància de Cook (CD) considera les variacions de tots els residuals quan s'omet l'observació i . Un valor alt de $CD^2(i)$ indica que l'observació i té una influència gran en els estimadors minimoquadràtics. Se sol considerar que un $CD^2(i) \geq 1$ és un valor gran. Els valors de CD^2 per als punts de la figura 9 es donen a la taula 1 [5]. Per a les dades de la taula, el valor de CD^2 es més gran que 1 només per al punt 6, que és el punt aberrant; aquest punt empeny la

Del plaer dels sentits al plaer de les xifres

recta ajustada cap a la seva nova posició, és a dir, cap amunt (*leverage point*).

Com es pot veure, la distància de Cook és molt sensible als aberrants en els extrems del camp obert per les dades, però és insensible envers els aberrants de la meitat de la línia. Això pot observar-se en el segon grup de dades de la taula següent, on s'ha introduït el punt aberrant a la meitat de la línia (punt 4). A més, el procediment basat en les distàncies de Cook no funciona quan hi són presents més d'una observació aberrant, ja que la influència d'un cert punt pot quedar emmascarada per la d'un altre punt.

TAULA 1

Punt núm.	Dades 1			Dades 2		
	<i>x</i>	<i>y</i>	$CD^2(i)$	<i>x</i>	<i>y</i>	$CD^2(i)$
1	0,0	0,0	0,296	0,0	0,0	0,055
2	1,0	1,1	0,009	1,0	1,1	0,018
3	2,0	2,0	0,010	2,0	2,0	0,015
4	3,0	3,1	0,050	3,0	10,0	0,441
5	4,0	3,8	0,407	4,0	3,8	0,080
6	5,0	10,0	2,192	5,0	5,1	0,271

6.9.1. Mètodes de regressió robusta

Els mètodes de regressió robusta es troben menys afectats per les observacions atípiques o aberrants que els mètodes de regressió tradicionals. Els mètodes de regressió robusta més utilitzats són els que es basen en l'ús de la mediana. Aquests mètodes tenen un punt crític (*breakdown*) més gran que els mètodes tradicionals LS (*least-squares* = mínims quadrats). El *punt crític* es defineix com aquell percentatge d'aberrants que es pot tolerar sense afectar la línia ni els seus paràmetres estimats. En el cas del mètode tradicional, el punt crític és del 0 %, ja que àdhuc un sol punt aberrant (*outlier*) té una influència gran en els estimadors.

El calibratge dels instruments de mesura analítica: calibratge univariant

6.9.1.1. Mètode de la mediana única (*single median method, SM*)

En aquest mètode, el paràmetre b_1 (el pendent de la recta) es determina de la manera següent: es calculen tots els pendents entre tots els parells, $(C_n^2)_{i,j}$, de dades; la mediana d'aquests pendents és el nou estimador del pendent:

$$b_1 = \text{med } (y_j - y_i) / (x_j - x_i) \text{ per a } 1 \leq j < i \leq n.$$

Amb aquesta estimació robusta del pendent es calcula l'ordenada a l'origen b_0 :

$$b_0 = \text{med } (y_i - b_1 x_i).$$

Pot demostrar-se que aquest mètode té un punt crític del 29 %. Quan s'aplica a les dades de la figura, la recta ajustada és:

$$y = 0,0 + 1,03x.$$

El punt aberrant té ara un residual molt més gran que abans i, per tant, és fàcilment detectable. De totes maneres, quan hi ha dos aberrants (taula 2, conjunt de dades 3), la contaminació introduïda per aquests aberrants és massa gran per a poder tenir bons estimadors.

6.9.1.2. Mètode de la mediana repetida (*repeated median method, RM*)

Aquest mètode és una millora de l'anterior, ja que el punt crític s'eleva fins a un 50 %. Els estimadors del pendent i de l'ordenada són:

$$b_1 = \text{med } \text{med } (y_j - y_i) / (x_j - x_i) \\ i, j \neq i$$

$$b_0 = \text{med } (y_j - b_1 x_i) \\ i$$

Aquest procediment s'explica a la figura 10 amb les dades de la taula 2: a cada punt i es calcula la mediana dels pendents entre aquest punt i tots els altres punts j ($j \neq i$), és a dir, es troba la mediana $b_1(i)$. S'obtenen d'aquesta manera n medianes (una per cada punt i) i la mediana d'aquestes n medianes, és a dir, $\text{med}(\text{med } b_1(i))$ és l'es-

Del plaer dels sentits al plaer de les xifres

TAULA 2

*Comparació dels mètodes de regressió robusta**

Punt núm.	Dades 1		Residuals			
	x	y	LS	SM	RM	LMS
1	0,0	0,0	0,90	0,00	-0,03	0,00
2	1,0	1,1	0,30	0,07	0,06	0,07
3	2,0	2,0	-0,49	-0,07	-0,06	-0,07
4	3,0	3,1	-1,08	0,00	0,03	0,00
5	4,0	3,8	-2,07	-0,33	-0,29	-0,33
6	5,0	10,0	2,44	4,83	4,89	4,83

Paràmetres de calibratge

b_0	-0,90	0,00	0,03	0,00
b_1	1,69	1,03	1,02	1,03

Punt núm.	Dades 3		Residuals			
	x	y	LS	SM	RM	LMS
1	0,0	0,0	1,19	0,45	0,00	0,00
2	1,0	1,1	0,07	-0,45	0,00	0,07
3	2,0	2,0	-1,26	-1,55	-0,20	-0,07
4	3,0	3,1	-2,38	-2,45	-0,20	0,00
5	4,0	10,0	2,30	2,45	5,60	5,87
6	5,0	10,0	0,08	0,45	4,50	4,83

Paràmetres de calibratge

b_0	-1,19	-0,45	0,00	0,00
b_1	2,22	2,00	1,10	1,03

* De l'anglès, LS = *least-squares*, SM = *single median*, RM = *repeated median*, LMS = *least median squares*. (Vegeu la referència 5.)

timador del pendent pel mètode de la mediana repetida. A partir d'aquest pendent robust b_1 , l'estimador de l'ordenada b_0 s'obté tal com s'indica, a partir de la mediana de les i diferències entre y_i i $b_1 x_i$. Quan s'aplica a les dades de la taula 2, la recta ajustada resulta ser:

$$y = 0,03 + 1,02x.$$

Àdhuc en el cas de tenir dos *outliers* en les dades, el mètode RM funciona bé.

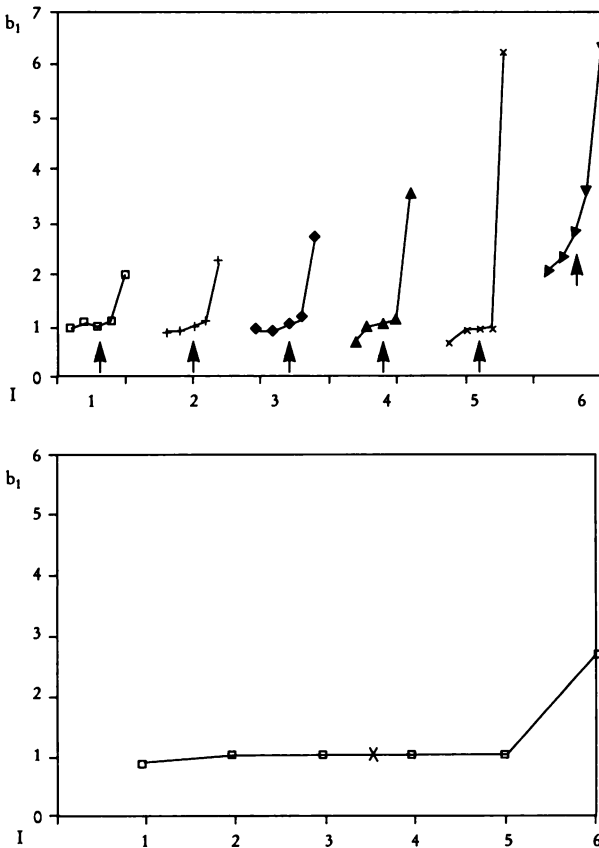


FIGURA 10. Aplicació del mètode de la mediana repetida al conjunt de dades 1.

Del plaer dels sentits al plaer de les xifres

6.9.1.3. Mètode de la mediana mínima de quadrats (*least median squares, LMS*)

Aquest és un altre mètode important de regressió robusta proposat per Rousseeuw. Es basa en la minimització del valor de la mediana dels residuals al quadrat. L'estimador LMS és donat per la minimització de:

$$\underset{i}{\text{med}} (y_j - b_1 x_j - b_0)^2,$$

on la mediana dels residuals és el $[n/2] + 1$ residual quan s'ordenen tots els valors residuals de menor a major, i $[n/2]$ és refereix a la part entera de $n/2$.

Recordem que en el mètode LS tradicional es minimitza la suma de quadrats:

$$\Sigma (y_j - b_1 x_j - b_0)^2.$$

En canvi, en la versió més simple del mètode LMS, l'estimació de l'ordenada i del pendent es fa de la manera següent: es calculen les línies entre tots els possibles parells de punts; això dóna C_n^2 parelles d'estimacions de b_0 i b_1 ; es calcula per a cada línia els residuals al quadrat de totes les n dades; finalment, es pren com a bona la línia que dóna la mediana mínima dels residuals al quadrat.

Aquest mètode té un punt crític del 50 %; quan s'aplica a les dades anteriors, la recta ajustada té per equació:

$$y = 0,00 + 1,03x.$$

Al conjunt de dades 3 de la taula 2 es pot veure que, àdhuc en el cas de tenir diversos punts aberrants en el mateix grup de dades, els paràmetres estimats es veuen poc afectats per la presència d'aberrants, a diferència del que passa en el mètode de mínims quadrats tradicional (LS) i en el mètode de la mediana única (SM), i menys afectats que en el cas d'emprar el mètode de la mediana repetida (RM). Una discussió dels mètodes de regressió robusta es pot trobar al llibre de Rousseeuw [5].

6.10. EFECTE DE LA MATRIU DE LA MOSTRA

6.10.1. *Coincidència de matriu* (matrix matching)

Els conceptes de *selectivitat* i d'*especificitat* es refereixen a la capacitat que té un mètode per a determinar l'analit sense interferència de les altres substàncies presents a la matriu. Encara que és difícil trobar un mètode totalment específic per a un analit determinat, es poden evitar moltes de les interferències químiques mitjançant l'elecció correcta dels patrons i de les condicions de l'anàlisi (*matrix matching standards*). En aquest mètode, es preparen els patrons de manera que s'assemblin tant com sigui possible a la matriu de la mostra on l'analit s'ha de determinar. Això vol dir, per exemple, emprar la mateixa matriu d'anions en la determinació d'un ió metàl·lic. Quan la matriu és molt complexa i, per tant, difícil de preparar, el mètode millor és el de l'addició estàndard.

6.10.2. *Mètode de l'addició d'un estàndard* (SAM)

En aquest mètode, els estàndards s'afegeixen sobre la matriu de la mostra, i la resposta de l'analit més l'estàndard se segueix en funció de la quantitat afegida d'estàndard. Suposem que la resposta obtinguda abans d'afegir l'estàndard és R_0 ; suposem que la relació entre la concentració d'analit i la resposta és donada per l'equació lineal (d'on s'han corregit les contribucions constants i, per tant, no conté el terme constant):

$$R_0 = kC_0 = kn_0/V_0,$$

on n_0 és el nombre inicial de mols d'analit i V_0 és el volum inicial de la mostra (com en el model lineal, en cas que $R = \beta_1 c$).

Si el model es compleix també després de l'addició de n_i mols d'analit, escriurem:

$$R_i = k[(n_0 + n_i)/(V_0 + V_i)],$$

on R_i és la nova resposta després de l'addició i , n_i és el nombre de mols afegits fins l'addició i i V_i és el volum de l'addició i .

Del plaer dels sentits al plaer de les xifres

En considerar l'efecte de la dilució, es pot escriure:

$$Q_i = R_i(V_0 + V_i) = kn_0 + kn_i,$$

que és la resposta corregida per al nou volum, $Q_0 = R_i(V_0 + V_i)$. La representació de Q_i enfront de n_i dóna una línia recta que té un pendent igual a k i una ordenada a l'origen igual a kn_0 . Així, el nombre inicial de mols es pot trobar per extrapolació:

$$\text{quan } Q_i = 0, kn_0 = -kn_i \text{ o } n_0 = -n_i.$$

Un desavantatge del mètode és que generalment l'estimació del nombre de mols n_0 es realitza lluny del valor mitjà de Q , precisament on l'interval de confiança per a n_0 és més gran (figura 11).

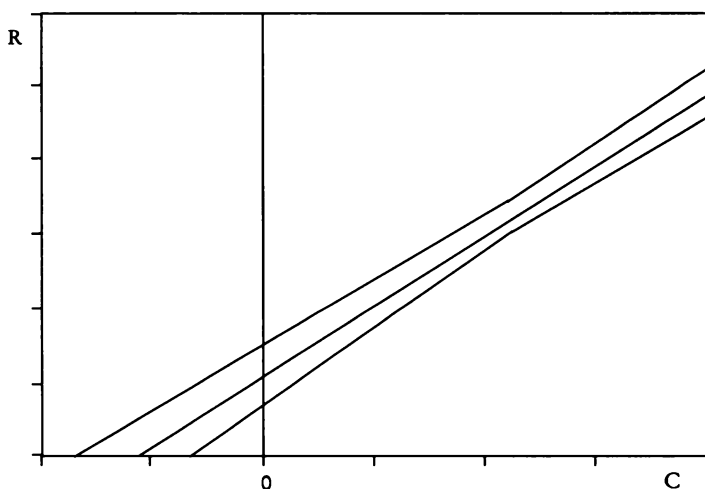


FIGURA 11. Representació de l'interval de confiança per a la recta de l'addició estàndard.

En el mètode d'addició múltiple és costum de fer cada addició igual a l'anterior. El procediment requereix conèixer aproximadament quina és la concentració inicial de l'analit en la mostra a analitzar. Pel que fa al nombre de patrons a utilitzar, s'ha de tenir en compte el següent: a) el camp de linealitat del mètode analític, i

El calibratge dels instruments de mesura analítica: calibratge univariant

b) la precisió requerida de l'anàlisi. En la pràctica, se solen emprar uns cinc patrons.

6.10.3. *Mètode generalitzat de l'addició estàndard (GSAM)*

Quan se sospita de la presència d'efectes matriu i no es disposa de sensors específics, els mètodes d'anàlisi tradicionals són inadequats. Fins fa poc, l'únic recurs de l'analista era fer patrons amb matrius iguals als problemes i realitzar l'anàlisi multicomponent amb les tècniques de calibratge multivariant (vegeu el capítol que segueix).

Actualment, hom disposa del mètode de l'addició estàndard generalitzat amb la inclusió de l'anàlisi multicomponent GSAM. L'anàlisi es realitza sobre la mateixa matriu, però ara s'afegeixen estàndards diferents (de diferents components) sobre la mateixa mostra i s'estimen les concentracions inicials de cada component C_0 a partir dels canvis observats de les respostes:

$$R_0 = C_0K,$$

on R_0 és la matriu que conté les respostes dels p sensors per a les n mostres inicials, C_0 és la matriu que conté les concentracions inicials dels r analits a cada mostra i K és la matriu dels coeficients de sensibilitat (constants de resposta lineal).

Quan es té en compte el canvi de volum,

$$Q_0 = N_0K,$$

Q_0 serà la matriu de respostes R_0 multiplicada pel volum inicial de la solució; N_0 és la matriu de les quantitats inicials de cada analit.

En el cas desitjable que el nombre de sensors sigui més gran que el d'analits $p \geq r$ i que s'afegeixin a cada addició n estàndards ($n = r$),

$$Q = (\Delta N + N_0)K.$$

Q ($n \times p$) és la matriu de respostes després de les addicions corregides tenint en compte el canvi de volum; ΔN ($n \times r$) és la matriu de quantitats afegides a cada pas; N_0 és la matriu de les quantitats ini-

Del plaer dels sentits al plaer de les xifres

cials de cada analit, i K ($r \times p$) és la matriu de respostes lineals de cada analit per cada sensor.

$$\Delta Q = Q - Q_0 = \Delta N K$$

$$K = (\Delta N^T \Delta N)^{-1} \Delta N^T \Delta Q$$

Les quantitats presents inicialment objecte de l'anàlisi, n_0 , s'obtin-
dran a partir de l'equació:

$$n_0 = (K K^T)^{-1} K Q_0.$$

A la figura 12 es mostra un cas on hi ha dos analits que donen resposta (un és l'analit d'interès i l'altre es considera una interferència). Si s'aplica el mètode d'addició estàndard, SAM, s'obté una predicció errònia, ja que SAM no elimina l'efecte de la interferència. En canvi, el mètode generalitzat de l'addició estàndard (GSAM) pot corregir aquest efecte.

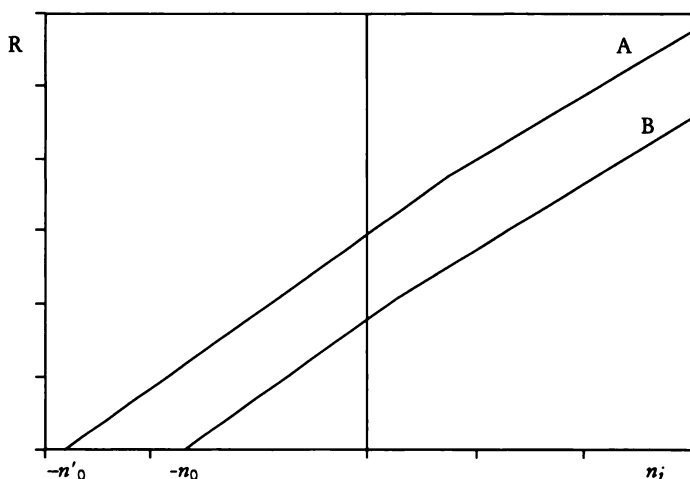


FIGURA 12. Efecte d'una espècie interferent sobre la predicció:

A) Aplicació de SAM; corba no corregida respecte a la interferència; el valor estimat de la quantitat inicial és n'_0 .

B) Aplicació de GSAM; corba real corregida respecte a la interferència; el valor estimat de la quantitat inicial és n_0 (correcta).

6.10.4. Propagació d'errors en el mètode d'addició estàndard

L'efecte que els procediments de calibratge i d'anàlisi per mínims quadrats tradicionals produeixen sobre l'estimació de la concentració desconeguda de l'analit a la mostra pot avaluar-se a partir dels principis de la propagació dels errors. Quan es té

$$R = \beta_0 + \beta_1 c,$$

la variància sobre c causada per les variacions a R és donada per:

$$c = (R - \beta_0)/\beta_1 \quad \sigma_c^2 = \sigma_R^2/\beta_1^2.$$

Com ja s'ha dit, la sensibilitat d'un mètode analític és donada per la relació que hi ha entre la variació de la resposta i la de la concentració ($\delta R/\delta c$); així, β_1 és en realitat un factor de sensibilitat.

A partir de les equacions anteriors, es pot veure fàcilment que, com més sensible sigui el mètode (β_1 gran), menys subjecte estarà a la propagació d'errors (σ_c petita).

En el cas del mètode d'addició estàndard, SAM, i quan el canvi de volums és negligible:

$$n_0 = n_1 R_0 / (R_1 - R_0)$$

$$\sigma_{n_0}^2 = n_0^2 [(\sigma_{R_0}^2/R_0^2 + \sigma_{R_1}^2/R_1^2)] [1 + n_0/n_1]^2.$$

Aquesta equació es pot simplificar si se suposa que el coeficient de variació σ/μ és constant per a tots els R , per exemple, $\sigma/\mu = \theta$

$$\sigma_{n_0}^2 = n_0^2 [1 + n_0/n_1]^2 [1/N_0 + 1/N_1] \theta^2,$$

on N_0 representa el nombre d'anàlisis replicades realitzades per tal d'estimar R_0 , i on N_1 és el nombre d'anàlisis replicades realitzades per tal d'estimar R_1 . A partir d'aquesta equació es veu que, per tenir valors de $\sigma_{n_0}^2$ petits s'ha d'afegir la màxima quantitat possible d'estàndard n_1 (sempre que ens mantinguem en l'interval lineal). Per a un nombre fix d'anàlisis $N_1 + N_0 = \text{constant}$, i n_1 fix, la variància serà mínima quan $N_1 = N_0$. Aquest resultat és un exemple dels avantatges que un analista pot obtenir a partir de l'estimació

Del plaer dels sentits al plaer de les xifres

acurada dels errors experimentals, pel que fa al millor disseny del seu experiment.

6.11. EFECTES MULTIPLICATIUS. MÈTODE DE CALIBRATGE DEL PATRÓ INTERN

En tots els mètodes emprats fins ara, s'han plantejat les dues hipòtesis següents:

a) La quantitat de mostra presa, tant per al calibratge com per a l'anàlisi, es coneix de manera acurada.

b) Tots els passos del procediment són els mateixos i els paràmetres que poden afectar la mesura física es mantenen constants.

Hi ha, però, mètodes analítics importants on no és possible garantir la reproductibilitat del procediment o de les quantitats preses per a l'anàlisi. Això ocorre, per exemple, en les tècniques analítiques següents: cromatografia gas/líquid (GLC), espectrografia i espectrometria d'emissió d'arc o de guspira, i fluorescència de raigs X.

El mètode de calibratge adequat per a aquest tipus de problema s'anomena *mètode del patró intern*. En aquest mètode, la resposta del detector donada per l'analit es compara amb la que dona un altre element o compost de concentració coneguda (el patró intern), que també és present en la mostra quan l'anàlisi es duu a terme. Encara que el patró intern ja pot formar part dels constituents de la mostra, en la majoria dels casos el patró intern s'afegeix a la mostra abans de l'anàlisi.

Considerem un mètode analític en el qual el paràmetre que es mesura depèn de tres termes:

$$R = mkc,$$

on R és la resposta instrumental que es mesura, m és un terme constant que depèn de la resposta de detector per a l'analit, o sensibilitat, k és un paràmetre variable, el valor exacte del qual no es pot conèixer per a cada experiment, i c és la concentració d'analit. L'equació donada es diferencia de les equacions lineals anteriors perquè té un tercer terme k ; aquest terme, en realitat, s'exclou en aquells mètodes, ja que es considera constant; p. e., en el cas de la

El calibratge dels instruments de mesura analítica: calibratge univariant

lleï de Beer $A = \epsilon cl$, l'absorbància es pot relacionar directament amb la concentració només quan l (camí òptic) i ϵ (absortivitat unitària) són constants.

En el mètode del patró intern, per als dos components considerats, analit (1) i patró intern (2), es pot escriure

$$R_1 = m_1 k_1 c_1 \text{ i } R_2 = m_2 k_2 c_2$$

si es divideix una igualtat per l'altra:

$$R_1/R_2 = m_1 k_1 c_1 / m_2 k_2 c_2.$$

Quan les dues mesures R_1 i R_2 s'obtenen simultàniament de la mateixa part de la mostra que s'analitza, pot suposar-se que els paràmetres variables k_1 i k_2 són iguals i, per tant, s'anul·len per al mateix experiment:

$$R_1/R_2 = m_1 c_1 / m_2 c_2.$$

En aquesta equació m_1/m_2 s'anomena *terme de relació de resposta* i pot expressar-se amb M . Com que en unes condicions determinades els valors de m_1 i m_2 no varien, M també serà constant, i, per tant, l'equació anterior es pot reescriure com:

$$R_1/R_2 = M c_1/c_2.$$

6.11.1. Mètode del patró intern únic

Primer es necessita conèixer el valor de M . Això es fa preparant una solució que conté quantitats conegudes de l'analit i del patró intern. D'aquesta solució s'obtenen valors de R_1 i R_2 . Com que c_1 i c_2 són conegudes, es pot calcular M . Si ara afegim una quantitat coneguda de patró intern a la mostra que conté l'analit (desconeguda), la concentració de l'analit a la mostra pot deduir-se a partir de:

$$c_1 = R_1 c_2 / R_2 M.$$

6.11.2. *Mètode del patró intern múltiple*

En aquesta variant, es preparen una sèrie de solucions de calibratge de manera que continguin concentracions diferents de l'analit pur c_i , amb una quantitat constant coneguda del patró intern c_2 . Es pot escriure:

$$R_1/R_{c_2} = M c_i/c_2 = M' c_i, \text{ on } M' = M/c_2.$$

Es pot fer el gràfic dels valors de R_1/R_2 enfront dels de c_i , i aquest ha de donar una línia recta que passa per l'origen i té un pendent M' . Si la mostra desconeguda es tracta de la mateixa manera que les dels patrons interns utilitzats per a preparar el gràfic de calibratge, la concentració de l'analit es pot deduir per extrapolació.

El principal avantatge del mètode de patró intern sobre els altres mètodes de calibratge és que pot donar millor exactitud i precisió; sota condicions òptimes l'exactitud pot arribar a ésser del 0,5 % en una tècnica instrumental (en els altres mètodes era de l'1-2 %). Això és degut, fonamentalment, al fet que les concentracions de l'analit i del patró es determinen sobre la mateixa solució i al mateix temps. Aquest mètode pot emprar-se no solament per a les tècniques esmentades, sinó que es pot estendre a d'altres quan es requereixi més exactitud i més precisió.

REFERÈNCIES

1. H. HUBERT i R. KLOCKENKÄMPER (1983). *Fresenius Z. Anal. Chem.*, núm. 316, p. 186.
2. A. WALD (1940). *Ann. Math. Statist.*, 11, p. 284-300.
3. L. M. SCHWARTZ (1977). *Anal. Chem.*, 49, p. 2062-2068.
4. N. R. DRAPER i H. SMITH (1981). *Applied Regression Analysis*. 2a ed. Nova York: John Wiley and Sons.
5. P. J. ROUSSEUW i A. Leroy (1987). *Robust Regression and Outlier Detection*. Nova York: John Wiley and Sons.