

# La problemática del ajuste en la etapa de la calibración

A.G. GONZÁLEZ, ANA SAYAGO y AGUSTÍN G. ASUERO

---

*Departamento de Química Analítica, Facultad de Farmacia,  
Universidad de Sevilla*

## Resumen

En este trabajo se pasa revista a los factores que influyen en el proceso de calibrado en el análisis químico. Los postulados en los que se basa la aplicación del método de los mínimos cuadrados, la bondad del ajuste del modelo empleado y el análisis de residuales son objeto de consideración. Este último tópico constituye una poderosa herramienta que permite comprobar las suposiciones en las que se basa el modelo y adoptar decisiones con respecto a la independencia u homogeneidad (homocedasticidad) de los datos. En casos de varianza no uniforme (heterocedasticidad) se procede a la ponderación, o a la transformación de los datos, en su caso. Se analizan también estrategias a aplicar en situaciones más complejas tales como curvas de calibración no lineales con varianza no uniforme. Por último, se considera el uso de funciones racionales y ajuste por funciones splines.

## Abstract

The factors, which exert an influence over the calibration process in chemical analysis, are subject of consideration. The postulates in which the least squares method are based; the goodness-of-fit of the model and the residual analysis are subject of study. This last topic is a powerful tool, which allows checking the assumptions assumed, as well as taking decisions concerning data independence or homogeneity (homocedasticity). When the variance is non-uniform (heterocedasticity), data are either weighted or a transformation is carried out. The strategies to be applied in more complex cases such as non-linear calibration curves with non-uniform variance are also analysed. Finally, the use of rational or spline functions are considered.

**Palabras clave:** calibración, análisis químico, varianzas uniforme y no uniforme, curvas de calibrado no lineales.

**Key words:** calibration, chemical analysis, uniform and non-uniform variances, non-linear calibration curves.

---

Correo electrónico: agonzale@us.es

## INTRODUCCIÓN

El calibrado, tal como lo define Sharaf y col. (1986), es esencialmente un proceso mediante el cual, la respuesta de un sistema de medida (señal analítica) puede expresarse en términos de una cualidad o cantidad de interés (concentración de analito). De manera matemática, podemos asimilarlo a una función  $y = f(x)$  donde  $y$  es la respuesta analítica correspondiente a la concentración  $x$ . El aspecto de mayor interés analítico consiste en encontrar la forma de esa función de respuesta. En el análisis tradicional por métodos químicos como la volumetría y gravimetría, la cantidad de analito está relacionada linealmente con el volumen del agente valorante o con el peso de la forma de pesada que constituyen un modelo lineal de respuesta

$$(1) \quad x = F \text{ (factor gravimétrico)} \cdot y \text{ (peso de la forma de pesada)}$$

$$(2) \quad x = N \text{ (normalidad)} \cdot \text{Peso equivalente} \cdot y \text{ (Volumen de valorante)}$$

Pero, es en análisis instrumental donde más se emplea la noción de función de respuesta, existiendo en algunos casos relaciones teóricas (Agtenderbos, 1979; Penninckx y col., 1996; Kleijburg y Pijpers, 1985; Wagenaar y col., 1974; Andrews y Jowett, 1982; O'Connell y col., 1990) que permiten pronosticar ésta (Tabla 1). Sin embargo, en otras muchas situaciones, la función de respuesta ha de establecerse empíricamente midiendo problemas ficticios, esto es, utilizando disoluciones patrones (diferentes concentraciones conocidas de analito presentes en la misma matriz). Bajo tales condiciones, la señal instrumental está definida por la cantidad de analito y el ruido de fondo.

**Tabla 1.** Funciones de respuesta empleadas en análisis instrumental

Función de respuesta	Nombre	Técnica (ejemplos)
$y = A + B x$	Ley de Beer	Espectroscopía de absorción
$y = A + B \log x$	Nernst	
$y = A x^B$	Scheibe-Lomakin	Electroquímica
$y = A + B x + C x^2$	Wagenaar y col.	Espectroscopía emisión
$y = A + B [1 - \exp(-C x)]$	Andrews y col.	Absorción atómica
$y = \frac{A - D}{1 + \left(\frac{x}{C}\right)^B} + D$	Rodbard	" Radioinmunoensayo

De los modelos matemáticos empleados para determinar la función de respuesta, el lineal es el más sencillo, más estudiado y fácil de tratar. Muchas funciones de respuesta presentan zonas lineales generalmente a bajas concentraciones de analito y otras donde aparece curvatura, e incluso, a veces, regiones donde la señal de respuesta es independiente (Mc Dowell, 1981) de la concentración de analito. El analista está interesado en la porción de la función de respuesta donde la variación de la señal instrumental con la

concentración de analito contiene información útil. El segmento de función de respuesta con interés analítico para establecer el calibrado y determinar el contenido de analito en una muestra desconocida recibe el nombre de *curva de calibrado*, estimándose, a partir de ella, la concentración desconocida por interpolación (método de patrones externos). Cualquier extrapolación puede conducir a resultados sometidos a grandes errores.

Como ya hemos indicado, si la porción de la curva de calibrado es lineal (recta de calibrado), tanto mejor. En casos de respuestas no lineales es común efectuar transformaciones para linealizarlas (Carroll y Rupert, 1988). El ajuste de los modelos matemáticos a los datos de la calibración (Cuadros-Rodríguez y col., 2001) para establecer la curva de calibrado es una tarea que puede, hoy día, llevar a cabo el analista de manera sistemática, sin olvidar la rigurosidad. En lo que sigue trataremos el ajuste del modelo de la calibración a los datos cuando la señal responde a un solo analito y el error se asocia a la variable  $x$ . El análisis multicomponente, aunque también es tratable por el método de patrones externos, quizás sea más lógico abordarlo desde las perspectivas del método de la adición estándar generalizada (Jochum y col., 1981; Herrero y col., 1994) y no será tratado aquí, al igual que la calibración en aquellos casos en que en ambas variables  $x$  e  $y$  están sujetas a error (Ripley and Thompson, 1987; González y Asuero, 1993).

## MÉTODO DE LOS MÍNIMOS CUADRADOS

La técnica más usual para establecer una curva de calibrado por ajuste de puntos es la de los mínimos cuadrados, pero para poder aplicarla justificadamente deben de cumplirse, entre otros (Agterdenbos, 1979; Sayago y col., 2004) tres requisitos fundamentales: i) Que la variable  $x$  se encuentre exenta de error. ii) Que el error  $\varepsilon_i$  asociado a la variable  $y_i$  se encuentre normalmente distribuido con media cero y varianza  $\sigma_i^2$ ;  $N(0, \sigma_i^2)$ . iii) Que la desviación observada en los valores de  $y$  para un valor fijo de  $x$  (replicaciones) permanezca constante en todo el intervalo de concentraciones. (Postulado de la uniformidad de la varianza u homocedasticidad).

En calibración, la concentración de analito se conoce, dentro de unos márgenes, con gran exactitud y precisión  $y$ , por lo tanto, podemos admitir que (i) se cumple. Muchos analistas dan por supuesto que la condición (ii) de normalidad se cumple y raras veces lo comprueban. Sin embargo, Clancey (1947), ya en la década de lo 40, examinó 250 distribuciones basadas en 50.000 análisis químicos y encontró que menos de un 15% de aquellos podían considerarse normales con el propósito de aplicar las técnicas estadísticas comunes. Esto nos recuerda la sentencia socarrona de Schmitt (1969) de que “La distribución normal es la menos normal de las distribuciones”.

Los analistas han prestado mucha más atención al postulado de *homocedasticidad* (iii). En situaciones de varianza no constante (*heterocedasticidad*), la técnica de los mínimos cuadrados aún puede emplearse, pero utilizando factores de ponderación (Gort y Hoogerbrugge, 1995; Desimoni, 1999). Garden y col. (1980) y Schwartz (1979) han considerado el problema de la varianza no uniforme en la construcción de curvas de calibrado.

Comenzaremos por estudiar el modelo de ajuste más sencillo –y aplicado– que el analista puede ensayar: El lineal. Supongamos que entre dos variables existe (Hunter, 1981) la relación lineal

$$(3) \quad \eta = \beta_0 + \beta_1 \xi$$

donde  $\beta_0$  y  $\beta_1$  son los valores de la ordenada en el origen y la pendiente de la recta, y  $\eta$  y  $\xi$ , los valores verdaderos desconocidos de la respuesta analítica y la concentración de analito calibrado, relacionados con los valores experimentales según

$$(4) \quad \eta = y + \varepsilon \quad \xi = x + \lambda$$

siendo  $\varepsilon$  y  $\lambda$  una medida del error aleatorio asociado a nuestro intento de medir  $\eta$  y  $\xi$ .

Una vez establecido el control de calidad en el proceso de medida, los errores tipo “blunder”, pérdidas no aleatorias de control y errores sistemáticos (Blanco y Cerdá, 1988) no se contemplan. Afortunadamente, como ya hemos indicado, la medida de la concentración del analito puede considerarse exenta de error (i) y, por tanto,  $\lambda \approx 0$  o mejor  $\lambda \varepsilon$ , con lo cual el modelo por ajustar sería

$$(5) \quad y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

Por otra parte, si aceptamos la distribución normal del error en la  $y_i$  (ii), la función de distribución de un valor cualquiera  $y_i$  alrededor del valor verdadero vendrá dada por la campana de Gauss

$$(6) \quad \phi(y_i) = \frac{1}{\sigma_1 \sqrt{2\pi}} \exp\left(\frac{-(y_i - \eta_i)^2}{2\sigma_1^2}\right)$$

La distribución conjunta independientemente de todos los valores de  $y$  será

$$(7) \quad \Phi(y_i) = \prod \phi(y_i) = \frac{1}{\prod \sigma_1 \sqrt{(2\pi)^n}} \exp\left(\frac{-(y_i - \eta_i)^2}{2\sigma_1^2}\right)$$

puesto que

$$(8) \quad \varepsilon_i^2 = (y_i - \eta_i)^2$$

y si existe homocedasticidad (iii)

$$(9) \quad \sigma_1^2 = \sigma_2^2 = \dots = \sigma^2$$

con lo cual

$$(10) \quad \phi(y) = \frac{1}{\sigma^n \sqrt{(2\pi)^n}} \exp\left(-\frac{1}{2} \sum \left(\frac{\varepsilon_i^2}{\sigma^2}\right)\right)$$

Nos interesa ajustar el modelo a los datos, de manera que la distribución conjunta tome un valor máximo, es decir, se desvíe lo mínimo de los valores verdaderos. Para ello basta

con que  $\sum \varepsilon_i^2$  sea mínimo. Desgraciadamente, trabajamos con una muestra de tamaño finito y desconocemos los valores de  $\beta_0, \beta_1, \eta_i$ , por lo que debemos contentarnos con sus estimaciones  $b_0, b_1, y_i$ . De este modo, en lugar de hacer mínima la suma  $\sum \varepsilon_i^2$ , debe hacerse mínima la cantidad  $\sum (y_i - \hat{y}_i)^2 = \sum r_i^2$ . Al número  $r_i$ , diferencia entre el dato experimental  $y_i$  y el valor estimado  $\hat{y}_i$ , se le llama residual y, para nuestros propósitos, deberá distribuirse tal como  $\varepsilon_i$ . Así pues, como

$$(11) \quad Q = \sum (y_i - b_0 - b_1 x_i)^2$$

de la condición necesaria de mínimo

$$(12) \quad \frac{\partial Q}{\partial b_0} = 0 \quad \frac{\partial Q}{\partial b_1} = 0$$

obtenemos las ecuaciones normales de la regresión lineal, de las que se calcula fácilmente los estimadores

$$(13) \quad b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$(14) \quad b_0 = \bar{y} - b_1 \bar{x}$$

### BONDAD DEL AJUSTE

Sin embargo, las cosas no siempre son tan sencillas. A veces, los datos se acomodan aparentemente a una línea recta cuando en la realidad adolecen de curvatura. El ojo humano reconoce con mucha más facilidad que cualquier técnica de procesamiento matemático la presencia de conductas diferentes entre grupos de puntos en una gran cantidad de casos. Por ello, la representación gráfica de la función de respuesta es un paso esencial para detectar curvaturas. De todos modos, la bondad del ajuste debe confrontarse con algún procedimiento matemático. Los analistas, para estimar la bondad del ajuste han utilizado demasiado a menudo el coeficiente de correlación  $r$  definido por

$$(15) \quad r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

estando los valores de  $r$  comprendidos (Koch, 1984) entre  $-1$  y  $+1$ .

Sin embargo, el empleo tan difundido del coeficiente de correlación ha sido nefasto, debido al poco valor real de este parámetro para detectar curvatura (AMC, 1988; Thompson, 1990; Asuero y col., 2006). Como indica Hunter (1981), en la teoría estadística, la correlación es la medida de la asociación entre dos variables aleatorias pero, en nuestro caso,  $\xi$  y  $\eta$  están funcionalmente relacionados. Correlación, en su sentido matemático-estadístico no existe, y los valores de  $r$  próximos a  $+1$  o  $-1$  proporcionan un ambiente de respetabilidad, pero no mucho más. Las curvas de calibrado analíticas dan a menudo un valor muy próximo a  $r = 1,000$  aunque haya diferencia significativa entre los puntos de la linealidad. Generalmente, se considera que  $100 r^2$  proporciona una estimación del % de variación

de los puntos experimentales a la linealidad (Kvålseth, 1985). Así, un valor de  $r = 0.980$  indica que el 96% de la variación está explicado por el modelo, pero debemos recordar, sin embargo, que ese 4% restante de la variación no explicada podría concentrarse en una pequeña porción de la curva. En consecuencia, el coeficiente de correlación es un medio poco sensible para estimar la calidad del ajuste de una recta de calibrado, y su uso discriminado puede conducir a conclusiones erróneas.

Algunos analistas utilizan ensayos estadísticos para saber si el coeficiente de correlación es significativo, como el  $t$  de Student (Miller and Miller, 2000)

$$(16) \quad t = \frac{|r| \sqrt{n-2}}{\sqrt{1-r^2}}$$

o la transformación de Fischer (Akhazarova y Kafarov, 1982)

$$(17) \quad z = \frac{1}{2} \ln \frac{1+r}{1-r}$$

pero no por ello se conjura el peligro, pues pueden indicar correlación cuando no la hay. Más interesante es el análisis de la varianza del ajuste (Sharaf y col., 1986; Akhazarova y Kafarov, 1982; AMC, 1994), calculando la varianza debida al modelo y la varianza residual. Se puede aplicar el criterio  $F$  para ensayar la bondad del ajuste. En el caso de una recta sin error  $F = \infty$  y en el caso de independencia entre  $y$  y  $x$ ,  $F = 0$ . Sin embargo, este ensayo por sí mismo no es evidencia (Sharaf, 1986) de que el modelo sea el adecuado.

## ANÁLISIS DE RESIDUALES

El analista necesita, pues, algunos medios, para comprobar si las suposiciones ii y iii se cumplen, así como para asegurarse de que el modelo escogido es el que mejor se ajusta a los datos experimentales. Para ello cuenta con una herramienta muy válida, el análisis de residuales (Draper y Smith, 1998; Meloun y col., 2002). Si estudiamos el comportamiento de los residuales después de haber obtenido la recta de regresión lineal y representamos tales residuales frente a los valores de  $x$ , pueden observarse cuatro situaciones (Belloto y Sokolowski, 1985) (figura 1).

El comportamiento I indica que el modelo utilizado era el adecuado y los residuales se han distribuido a lo largo del intervalo de medida con media cero y desviación constante. El comportamiento II aparece cuando la ecuación utilizada para el ajuste no es la adecuada y podría necesitar probablemente, términos de orden superior para el ajuste completo. La situación III es típica cuando no existe uniformidad en la varianza de  $y$ . El sistema es heterocedástico, y la dispersión disminuye con  $x$ , como indica la figura, aunque también podría aumentar o tomar la forma de una pajarita. El comportamiento IV se refiere a una situación en la que los errores dependen de alguna variable que aún no ha sido descubierta, vulnerando la suposición de independencia del error. Para corregir esta tendencia deben considerarse otras variables que tácitamente pudieran intervenir en el proceso. De todos modos, a veces, las situaciones reales no son tan claras como las que idealmente se

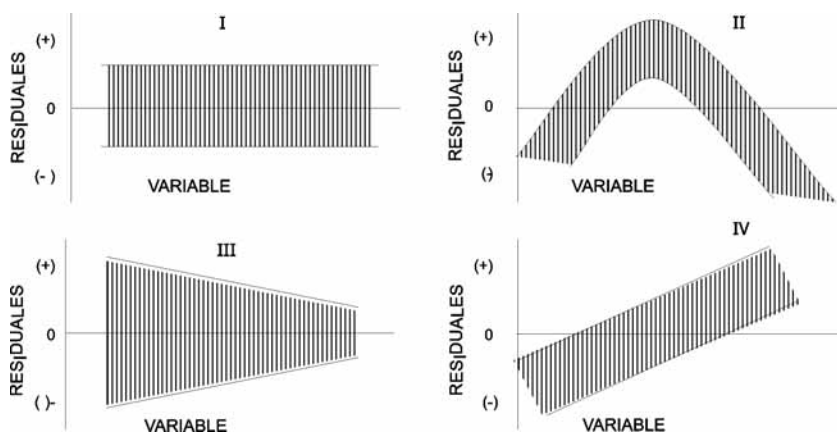


Figura 1. Representación de los residuales frente a valores de la variable

muestran en la figura. El comportamiento I ya hemos dicho que corresponde al caso de un ajuste perfecto en que los residuales se distribuyen según  $N(0, \sigma^2)$ .

Para examinar matemáticamente si los residuales se distribuyen como hemos supuesto existen numerosos ensayos. Los más sencillos consisten en representar la curva de frecuencia acumulativa (%) frente al valor del residual, que da la familiar curva en forma de S o una recta si se traza sobre papel de probabilidad normal (Miller y Miller, 2000; Shapiro, 1990). Métodos más específicos son el de Kolmogorof-Smirnov (Miller y Miller, 2000; Kateman y Pijpers, 1993), el de Bowman y Shenton de los parámetros de simetría y kurtosis (Bowman y Shenton, 1975) y el de Anderson-Darling (Pettit, 1977) de los percentiles. Un test no paramétrico muy útil para comprobar si los residuales se ajustan bien a las suposiciones de una distribución al azar, y que se usa mucho para detectar curvaturas en rectas de calibrados, es el de Wald-Wolfowitz (Miller y Miller, 2000). Se define "ristra" (run en la literatura anglosajona) como cada secuencia de signos distintos que aparecen separados cuando se observa el signo de los residuales a valores crecientes de  $x$  (Thompson, 1982). Así, en la siguiente serie de residuales: + + + + - - - - + + hay tres ristra. Evidentemente, si existe una secuencia no aleatoria de signos + y - aparecerá un número menor de ristra que en el caso de distribución aleatoria (o en algún caso, mucho mayor, como sucede en las variaciones sistemáticas periódicas).

El ensayo anterior nos indica si los residuales se distribuyen al azar. No obstante, puede ser que aunque su distribución sea aleatoria, no sea normal. Ya hemos indicado que muchos experimentos conducen a distribuciones no normales del error. Los estudios Monte-Carlo indican que el estimador de centralización en mínimos cuadrados, la media, es uno de los más pobres estimadores frente a la variedad de distribuciones de error diferentes a la normal (Staudte y Sheather, 1990; Rousseeuw, 1991). La puesta a punto de los estimadores de mínimos cuadrados cuando se viola la imposición de normalidad basados en estadísticas robustas es de una gran importancia (Rousseeuw, 1991; Walczak y Massart, 1998; Massart y col., 1986; Danzar, 1989; Thompson, 1989). Un estimador se llama robusto (Rousseeuw, 1991) cuando es insensible a desviaciones suaves de las suposiciones establecidas, y solo es ligeramente ineficiente relativo a los mínimos

cuadrados cuando las suposiciones se cumplen. Así, los métodos robustos proporcionan protección frente a las distribuciones no normales y son además resistentes a la presencia de resultados sospechosos (*outliers* en la literatura anglosajona) en los datos (Walczak y Massart, 1998).

Si después de haber ajustado nuestro modelo a los datos observamos una distribución de los residuales como en el caso III nos encontramos ante una violación del postulado de homocedasticidad. En tales casos, la aplicación directa de la técnica convencional de los mínimos cuadrados puede conducir a conclusiones erráticas (Draper and Smith, 1998).

Es por ello que se reconoce en la literatura química que usualmente la varianza no es constante en análisis químico (Garden y col., 1980). Como indica Schwarz (1979), aunque la suposición de homocedasticidad es válida para muchos procesos analíticos, hay algunas típicamente heterocedásticas, que incluyen todos los métodos basados en las medidas de cuentas, así como los métodos fotométricos y cromatográficos en ciertos casos. Efectivamente, en determinadas ocasiones, la heterocedasticidad puede aparecer en cualquier método analítico, como en espectrofotometría de absorción y voltametría de redisolución anódica, cromatografía, electroforesis capilar (Baumann y Wätzig, 1995), espectroscopía de plasma acoplado por inducción (Ketkar y Bzik, 2000), y otros.

Una vez hemos observado dicha tendencia en los residuales, se comprueba la heterocedasticidad (a la vez que se soslaya) mediante replicado de los valores de  $y$  para un valor dado de  $x$ . Los ensayos de Barlett o Cochran (Akhazarova y Kaarov, 1982; Sayago and Asuero, 2004; UNE 82009-2, 1999), son suficientemente rigurosos.

## INDEPENDENCIA

Analicemos ahora la situación correspondiente al caso IV en la que se observa una tendencia de los residuales cuya causa es la dependencia de una variable aún no analizada. Se trata de una situación que por su curiosidad e importancia vamos a aclarar con un ejemplo. Cuando se estudia la solubilidad del Diazepam en propilenglicol (Belloto y Sokolowski, 1985), la relación obtenida entre la concentración y la absorbancia (ley de Beer) fue:  $C = -0,034084 + 0,117424 A$  ( $r = 0,999$ ). Sin embargo, al representar los residuales respecto al orden numérico en que cada muestra fue medida, se observa una situación típica del caso IV. Si al representar los residuales frente a una variable considerada o no, se observa una tendencia como ésta, debe sospecharse la dependencia de una variable oculta o no, relacionada con las anteriores. En el caso del Diazepam, esta representación sugiere que los errores no son independientes, pudiendo estar relacionados con el tiempo de alguna manera desconocida, ya que los residuales parecen ser función del orden de las medidas. Si el tiempo se incluye en el modelo, se obtiene la ecuación  $C = -0,071217 + 0,118201 A + 0,0003625566 t$  ( $r = 1,00$ ). Los residuales así obtenidos representados frente a la variable tiempo siguen una representación típica de adecuación al modelo como el caso I.

Gracias al análisis de residuales se detecta la influencia de la variable tiempo que de otro modo hubiese pasado inadvertida. Un modelo residual dependiente del tiempo, como el que hemos visto, puede indicar que se está produciendo una seria



contaminación, o que la fuente de luz decae con el tiempo. El modelo residual indica que existe una variable dependiente del tiempo, pero no indica por qué existe esa dependencia.

Cuando los residuales obtenidos al ajustar el modelo lineal a nuestros datos muestra un comportamiento como II, esto nos indica, desafortunadamente, que el modelo sugerido no se adecua a la realidad y que son necesarios términos adicionales en  $x$  para asegurar el ajuste a los datos analíticos. En este caso, los residuales no se distribuyen al azar, sino que muestran una disposición cóncava o convexa. El ensayo de Wald-Wolfowitz (Miller and Miller, 2000) nos sugeriría que existe una curvatura, y que nuestra función de respuesta no es lineal en  $x$ .

### REGRESIÓN LINEAL PONDERADA

Si aplicamos la condición de mínimo a la distribución conjunta en el caso en que las varianzas no son constantes, la expresión a minimizar sería  $\sum (\varepsilon_i / \sigma_i)^2$ . En situaciones reales, tratando con residuales, es conveniente definir (Asuero and González, 1989) una cantidad denominada *peso o factor de ponderación*, que denotaremos por  $w_i$ , asociado con una observación  $y_i$ , y definido por

$$(18) \quad w_i = \frac{\sigma_0^2}{\sigma_i^2}$$

El peso de una observación es inversamente proporcional a la varianza. El término  $\sigma_0^2$  es simplemente un factor de proporcionalidad que coincide con la varianza de peso unidad, y que muchos analistas (Cuadro-Rodríguez y col., 2001) toman arbitrariamente como unidad. El objeto de introducir los pesos es estabilizar la varianza de los residuales. En el caso del modelo ponderado tendríamos

$$(19) \quad \sqrt{w_i} y_i = \sqrt{w_i} \beta_0 + \sqrt{w_i} \beta_1 + \sqrt{w_i} \varepsilon_i$$

y por tanto

$$(20) \quad \text{var}(\sqrt{w_i} \varepsilon_i) = w_i \text{var}(\varepsilon_i) = w_i \sigma_i^2 = \frac{\sigma_0^2}{\sigma_i^2} \sigma_i^2 = \sigma_0^2 = cte$$

Ahora, la condición de minimización es

$$(21) \quad Q = \sum w_i (y_i - \hat{y}_i)^2 = \min$$

que conduce a expresiones similares a las deducidas para la regresión lineal simple

$$(22) \quad b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$(23) \quad b_0 = \bar{y} - b_1 \bar{x}$$

donde

$$(24) \quad \bar{x} = \frac{\sum w_i x_i}{\sum w_i}$$

$$(25) \quad \bar{y} = \frac{\sum w_i y_i}{\sum w_i}$$

### NORMALIZACIÓN Y MODELIZACIÓN DE LOS PESOS

Es conveniente normalizar los pesos (Connors, 1987) pues, en medidas precisas,  $w_i$  así como  $\sum w_i$  pueden tomar valores elevados y provocar errores de redondeo o resultados anómalos cuando se usan pocas cifras significativas. La manera más adecuada de normalizar los pesos es aquella cuya expresión se reduzca, en el caso de homocedasticidad, a las ya conocidas. Esto requiere que  $\sum w_i = N$ , siendo  $N$  el número de puntos de la curva de calibración. Entonces,

$$(26) \quad \sum \left( \frac{\sigma_0^2}{\sigma_i^2} \right) = N$$

de donde

$$(27) \quad \sigma_0^2 = \frac{N}{\sum \sigma_i^{-2}}$$

$$(28) \quad w_i = \frac{N \sigma_i^{-2}}{\sum \sigma_i^{-2}}$$

Como resulta que  $\sigma_i$  es desconocida, la estimamos a partir de las replicaciones efectuadas

$$(29) \quad s_i = \sqrt{\frac{\sum (y_{ij} - \bar{y}_i)^2}{n_j - 1}}$$

Garden y col. (1980) indican una técnica experimental más sofisticada para el cálculo de los factores de peso. Para ello se hacen al menos replicaciones a cada concentración  $x$ , y se obtiene la desviación estándar  $s_i$ . Puesto que este parámetro tiene una distribución log-normal, se calculan los valores  $\log s_i$ , y se ajusta la curva  $\log s_i$  frente a  $x_i$  por el método de los mínimos cuadrados sin pesar, con lo que se obtiene una función suavizada

$$(30) \quad \log s = f(x)$$

Una vez hecho esto, los valores de los factores de ponderación para un valor determinado de  $x$ , pueden calcularse sencillamente a partir de la Eq. 18 con  $\sigma_0^2 = 1$ , de acuerdo con

$$(31) \quad \log w_i = -2 \log s_i = -2 f(x)$$

Éste es el procedimiento por seguir cuando la heterocedasticidad es errática (el residual ya aumenta con el valor de  $x$ , ya disminuye o sufre ambas tendencias a lo largo de una serie de valores).

Si esta tendencia fuera constante, habría que sospechar en cambio de la existencia de relaciones funcionales de la dispersión de los residuales con el valor de la abscisa, consecuencia inherente de una función de la varianza de la señal con respecto al valor que toma dicha señal. Estas relaciones pueden encontrarse bien empíricamente (cuando se observa una tendencia definida al representar  $s_i$  frente a  $y_i$  –mejor que frente a  $x_i$ – que puede representarse por una relación funcional) o bien, a partir de principios teóricos. Los casos más típicos que suministra la literatura química son dependencias potenciales o cuadráticas (Zorn y col., 1997; Prudnikov y Shapkina, 1984; Bubert y Klockenkämper, 1983)

$$(32) \quad \sigma_y^2 = A y^B$$

$$(33) \quad \sigma_y^2 = C_0 + C_1 y + C_2 y^2$$

Un ejemplo de relaciones teóricas a priori que indican una tendencia de la varianza son las medidas de cuentas (Schwartz, 1978) que, al seguir una distribución de Poisson, tiene como medida de la varianza el valor medio (Finney, 1979). Otro ejemplo lo constituye la cromatografía gaseosa cuando se mide la altura del pico correspondiente a diferentes cargas molares inyectadas de analito, donde (Bocek y Novak, 1970) la varianza es proporcional al cuadrado de la altura.

Para concluir el epígrafe, mencionaremos todavía un procedimiento para el ajuste en situaciones heterocedásticas, que consiste en tomar en principio los  $w_i = 1$  y calcular los parámetros de regresión convencionalmente. A continuación se estima la varianza de  $y_i$ ,  $s_{y_i}^2$ , y se hace

$$(34) \quad w_i = \frac{1}{s_{y_i}^2}$$

recalculando los parámetros. El procedimiento se reitera hasta (Schmidt, 1969; Das y Tydeman, 1982) que no se encuentren diferencias significativas entre dos varianzas sucesivas utilizando el ensayo  $F$ . La regresión lineal pesada proporciona valores más fiables (Koch, 1985; Tyson, 1988, Mac Taggart y Farwell, 1992) de la precisión y del límite de confianza de las concentraciones.

## TRANSFORMACIÓN DE COORDENADAS

En ciertas ocasiones, los autores proponen como una alternativa al empleo de los pesos, el uso de transformaciones de coordenadas que aseguren una estabilidad de la varianza (Asuero, 2005). Así pues, si efectivamente existe una función conocida de la varianza  $\sigma_y^2 = u(y)$ , podemos efectuar una transformación  $Y = \phi(y)$  cuya forma es desconocida a priori. Según la ley de propagación de errores

$$(35) \quad \sigma_Y^2 = \left( \frac{dY}{dy} \right)^2 \sigma_y^2$$

Impongamos ahora que la varianza de la nueva variable  $Y$  sea constante e igual a  $\sigma_0^2$ . Según esto

$$(36) \quad \left(\frac{dY}{dy}\right)^2 = \frac{\sigma_0^2}{\sigma_y^2} = \frac{\sigma_0^2}{u(y)}$$

y, por tanto,

$$(37) \quad \Phi(y) = Y = \sigma_0 \int \frac{dy}{\sqrt{u(y)}}$$

Es posible tomar un valor arbitrario para  $\sigma_0$ . Una vez hecho esto, la varianza está uniformizada y solo hay que minimizar la cantidad

$$(38) \quad Q = \sum (Y_i - \hat{Y}_i)^2$$

Solo que ahora, la función  $Y = G(x)$  es generalmente una expresión no lineal en  $x$  y a veces no lineal en los parámetros, con lo que habría que recurrir a métodos de regresión no lineal. Esta situación parece más embarazosa que la de una regresión lineal ponderada, por lo que este tipo de estrategia es útil cuando la función de respuesta  $y = f(x)$  de por sí no es lineal.

## PESOS DEPENDIENTES DE LAS TRANSFORMACIONES

Pero, el analista todavía puede jugar una carta antes de resignarse a calcular los parámetros de calibración con un modelo curvilíneo. A veces, es posible transformar los datos (Atkinson, 1985) de manera que se obtenga un comportamiento lineal. Un método muy común consiste en representar los datos en papel semilogarítmico. Si se observa un comportamiento lineal, es posible la existencia de una relación de tipo potencial o exponencial y se procede a efectuar un cambio de variables. Por ejemplo, si se sospecha que la función es  $y = A x^B$ , tomamos logaritmos para linealizar se tiene:  $\log y = \log A + B \log x$ . Los cambios son  $Y = \log y$  y  $X = \log x$ , pero estas nuevas variables no pueden emplearse para una regresión lineal convencional, pues si aplicamos la ley de propagación de errores a la  $Y$  (a la  $X$  no, porque la abscisa está exenta de error) tenemos

$$(39) \quad \sigma_Y^2 = \left(\frac{\partial Y}{\partial y}\right)^2 \sigma_y^2$$

Si  $\sigma_y^2$  es constante (homocedasticidad en  $y$ ), la linealización produce una no uniformidad en la varianza  $Y$ , pues generalmente  $(\partial Y/\partial y) = f(y)$ , con lo que se hace necesario introducir un peso dependiente de la transformación (Asuero y González, 1989; de Levie, 2000; de Levie, 1986; Paré y col., 1988; Asuero, 1993)

$$(40) \quad w_{Y_i} = \frac{\sigma_0^2}{\sigma_Y^2} = \left(\frac{\partial y}{\partial Y}\right)^2 \frac{\sigma_0^2}{\sigma_y^2}$$

si  $\sigma_y^2 = \sigma_0^2$

$$(41) \quad w_{Y_i} = \left( \frac{\partial y}{\partial Y} \right)^2 = w_{Y/y}$$

donde  $W_{Y/y}$  es el peso dependiente de la transformación. Si  $\sigma_y^2$  no es constante, entonces

$$(42) \quad w_{y_i} = \left( \frac{\partial y}{\partial Y} \right)^2 \frac{\sigma_0^2}{\sigma_y^2} = w_{Y/y} \cdot w_{y_i}$$

Los pesos atribuidos a las transformaciones más usuales en química analítica se muestran en la Tabla 2.

**Tabla 2.** Factores de peso dependientes de la transformación

Transformación	Factor de peso
$Y = 1/y$	$y^4 \sigma_0^2 / \sigma_y^2$
$Y = \ln y$	$y^2 \sigma_0^2 / \sigma_y^2$
$Y = y^2$	$\sigma_0^2 / (4y^2 \sigma_y^2)$
$Y = e^y$	$\sigma_0^2 / (e^{2y} \sigma_y^2)$
$Y = \text{logit } y$	$y^2(1-y)^2 \sigma_0^2 / \sigma_y^2$

### CURVAS NO LINEALES CON VARIANZA NO UNIFORME

En muchos casos no se conoce exactamente la forma de una ecuación no lineal en  $x$ . Si no es posible una transformación que convierta los datos experimentales en otros que se comporten linealmente, será necesario ajustar los datos a una curva. Schwartz (1977) ha estudiado tres aproximaciones prácticas para la construcción de una curva de calibrado  $y = f(x)$ .

El procedimiento más sencillo es el de los segmentos rectilíneos, que consiste en construir la curva conectando puntos  $(x_i, y_i)$  adyacentes mediante segmentos rectilíneos. Si la medida experimental cae entre dos puntos  $(x_a, y_a)$  y  $(x_b, y_b)$ , el segmento que nos interesa tiene la ecuación de una recta que pasa por dos puntos

$$(43) \quad y = y_a + \frac{y_b - y_a}{x_b - x_a} (x - x_a)$$

Este es el procedimiento más sencillo, pero también el que proporciona situaciones menos satisfactorias a la hora de calcular la concentración del analito y estudiar la precisión de la medida.

Otro método algo más complejo es el de los tres parámetros, donde se hace

$$(44) \quad y = c_0 + c_1 f_1(x) + c_2 f_2(x)$$

Con  $f_1(x)$  y  $f_2(x)$  lineales o no. El analista ha de ser lo suficientemente avisado para seleccionar las funciones según la forma de la curva. Una vez hecho, se utiliza un método de regresión multilínea para estimar los valores de  $c_0, c_1$  y  $c_2$ . Si las funciones no están bien escogidas, pueden cometerse graves errores debidos a falta de ajuste.

El tercer método consiste en construir la curva de calibrado con tantos parámetros ajustables como sean necesarios para que se adapte adecuadamente a los datos. Una estrategia común es ajustar la curva a un expresión polinómica en  $x$

$$(45) \quad y = a + bx + cx^2 + dx^3 + \dots$$

El mejor procedimiento es probar polinomiales de grado sucesivamente mayor, hasta que el ensayo  $F$  de las varianzas residuales indica que el error sistemático debido a la falta de ajuste ha sido eliminado. Si la representación tiene  $N$  puntos, el polinomial de mayor grado que puede usarse es de grado  $N-1$  pero, en general, polinomiales de orden más bajo proporcionan ajustes bastante adecuados. Este tipo de ajuste se resuelve por regresión multilínea (Draper and Smith, 1998; Pattengill y Sands, 1979; Heilbronner, 1979). Para un polinomial de grado  $m$ ,  $F(x) = a_0 + a_1x + a_2x^2 + \dots + a_mx^m$ , la cantidad para minimizar es

$$(46) \quad Q(a_0, a_1, \dots, a_m) = \sum w_i (y_i - F(x_i))^2$$

En este caso, el conjunto de ecuaciones normales se obtiene de las diversas

$$(47) \quad \frac{\partial Q}{\partial a_j} = 0 \quad j=0,1,\dots,m$$

Estas ecuaciones normales conducen a un sistema lineal que en forma matricial puede escribirse:  $C \cdot A = B$ , donde

$$(48) \quad C = \begin{bmatrix} \sum w_i & \sum w_i x_i & \dots & \sum w_i x_i^m \\ \sum w_i x_i & \sum w_i x_i^2 & \dots & \sum w_i x_i^{m+1} \\ \dots & \dots & \dots & \dots \\ \sum w_i x_i^m & \sum w_i x_i^{m+1} & \dots & \sum w_i x_i^{2m} \end{bmatrix}$$

$$(49) \quad A = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ \vdots \\ a_m \end{bmatrix}$$

$$(50) \quad B = \begin{bmatrix} \sum w_i y_i \\ \sum w_i x_i y_i \\ \vdots \\ \vdots \\ \sum w_i x_i^m y_i \end{bmatrix}$$

y por tanto:  $A = C^{-1} \cdot B$ . La varianza del ajuste viene dada por  $s^2 = Q/(N - m - 1)$ .

Los elementos de la matriz de covarianzas vienen dados por

$$(51) \quad V_{ij} = s^2 C_{ij}^{-1}$$

$$(52) \quad V_{ij} = s^2 C_{ij}^{-1}$$

donde  $C_{ij}^{-1}$  es un elemento de la inversa de la matriz  $C$ .

Este método falla a veces porque la matriz de coeficientes  $C$  se hace prácticamente singular. Una aproximación alternativa es utilizar polinomios ortogonales (Lee and Lee, 1982; Valkó y Vajda, 1989; Kragten, 1990). Sean dos polinomios ortogonales en  $x$ ,  $P_i(x)$  y  $P_j(x)$  entonces,

$$(53) \quad \sum P_i(x_k) P_j(x_k) = 0 \quad \forall i \neq j$$

La función de ajuste puede ponerse como

$$(54) \quad F(x) = \sum a_j P_j(x) = a_0 P_0(x) + \dots + a_m P_m(x)$$

o combinando los polinomios en serie de potencias

$$(55) \quad F(x) = c_0 + c_1 x + c_2 x^2 + \dots + c_m x^m$$

El uso de los polinomios ortogonales conduce a la anulación de los elementos no diagonales de la matriz de los coeficientes, evitando cuasisingularidades y simplificando los cálculos. Los más usados en ajustes de curvas son los de Chebichev (Akhnazarova y Kafarov, 1982) y Forsythe (Lee and Lee, 1982; Valkó y Vajda, 1989; Kragten, 1990). Los primeros se generan con la fórmula recurrente

$$(56) \quad \begin{aligned} P_0(x) &= 1 \\ P_1(x) &= x - \frac{n+1}{2} \\ P_{k-1}(x) &= P_1(x)P_k(x) - \frac{k^2(n^2 - k^2)}{4(4k^2 - 1)} P_{k-1}(x) \end{aligned}$$

los de Forsythe, en cambio, con

$$(57) \quad \begin{aligned} P_{-1}(x) &= 0 \\ P_0(x) &= 1 \\ &\vdots \\ P_k(x) &= (x - u_k) P_{k-1}(x) - V_{k-1} P_{k-2}(x) \end{aligned}$$

donde

$$(58) \quad u_k = \frac{\sum x_i P_{k-1}^2(x_i)}{\sum P_{k-1}^2(x_i)}$$

$$(59) \quad V_{k-1} = \frac{\sum x_i P_{k-1}(x_i) P_{k-2}(x_i)}{\sum P_{k-1}^2(x_i)}$$

Los valores de  $a_j$  se calculan fácilmente

$$(60) \quad a_j = \frac{\sum y_i P_j(x_i)}{\sum P_j^2(x_i)}$$

Para encontrar la ecuación de regresión por el método de aproximaciones sucesivas, se calculan las varianzas residuales para la regresión de orden  $k$  donde  $Q_k$  es la suma de los cuadrados de los residuales, que se calcula por la recurrencia:

$$(61) \quad Q_k = Q_{k-1} - a_k^2 \sum P_k^2(x_i)$$

El prerrequisito (Akhazarova y Kafarov, 1982) es que  $Q_0 = \sum (y_i - \bar{y})^2$ . Cuando al aumentar el grado del polinomial,  $s_{k+1}^2 > s_k^2$ , mediante el ensayo  $F$ , se da por terminado el proceso. Al ser ortogonales los polinomiales, la matriz de covarianza es diagonal y sus términos son

$$(62) \quad V_{jj} = s_{aj}^2 = s_k^2 C_{jj}^{-1}$$

Haciendo un resumen de lo visto hasta ahora, el método de los segmentos rectilíneos puede ser útil en los casos de datos altamente precisos y curvas de calibrado de poca curvatura. El método de los tres parámetros es siempre intrínsecamente inexacto para curvas no lineales de elevada curvatura, a menos que se sepa la forma correcta de las funciones. En general, el método de ajuste polinomial proporciona la mayor confianza porque utiliza el test  $F$  para ajustar apropiadamente los puntos, aunque se trata, sin lugar a dudas, del modelo más complicado.

## **FUNCIONES RACIONALES Y AJUSTE POR FUNCIONES DE SPLINES**

A pesar de las grandes ventajas y numerosísimas aplicaciones de los polinomiales ortogonales, no son en absoluto el arma definitiva. Rice (1964) ha propuesto funciones polinomiales del tipo

$$(63) \quad F(x) = \frac{\sum a_i x^i}{\sum b_i x^i}$$

que tienen un grado de flexibilidad mucho mayor que los polinomiales, pues una función racional de grado relativamente bajo puede ajustarse a una forma que nunca podría aproximarse de manera efectiva mediante el uso de polinomiales de grado bajo o medio. Este tipo de función es más difícil de evaluar que un polinomial, siendo necesario el uso de la regresión no lineal.

Otro método de ajuste poco usado pero muy recomendable (Jost y Rock, 1954) es el de los polinomiales ortogonales discontinuos de Vettin, análogos a los de Legendre, que se expresan siempre en función de un parámetro que es la longitud entre puntos



equidistantes. Se trata de un procedimiento muy similar al de la Serie de Fourier solo que utiliza un número finito de puntos. El inconveniente es la necesidad de conocer los puntos equidistantes, pero esto puede soslayarse mediante lectura a ojo de la curva diseñada.

Estos dos últimos tipos de ajuste sirven para curvas de calibrado, donde se superponen dos o más fenómenos. A título de ejemplo, en fluorimetría, las representaciones de la intensidad de emisión frente a la concentración son prácticamente lineales en disolución diluida, pero muestran curvatura negativa a altas concentraciones debido a la presencia de artefactos ópticos o a la no adecuación del modelo que predecía linealidad. Como ambos efectos son independientes, en la práctica pueden aparecer curvas de muy diferente aspecto, ya que se superponen dos o más procesos. Es lógico pensar que no habrá ninguna función algebraica que, por sí sola, sea capaz de ajustar completamente los datos.

La mejor solución sería tal vez ajustar los puntos a una curva que consistiera en varias secciones unidas de diferente forma matemática. Éste es exactamente el fundamento del ajuste por funciones Spline (Price y Simonnen, 1962; Ahlberg y col., 1967; Hwang y Wineformer, 1988). En la práctica, los Splines cúbicos (Frigieri y Rossi, 1979) son los más empleados, y aproximan los datos a una serie de ecuaciones cúbicas. Estos “eslabones” cúbicos se superponen en  $p$  puntos de interpolación (*Knots* en la literatura anglosajona) y es esencial que los Splines muestren una transición continua en tales puntos. Esta continuidad debe seguirla tanto la función como sus  $m-1$  primeras derivadas. Un Spline Cúbico completo tiene  $p-1$  eslabones, cada uno de ellos con coeficientes ( $S = a + bx + cx^2 + dx^3$ ). Así, es necesario calcular  $4x$  ( $p-1$ ) coeficientes. Este procedimiento ha sido utilizado exitosamente en radioinmunoensayo, cromatografía gas-líquido y espectrometría de absorción atómica (Miller and Miller, 2000).

## BIBLIOGRAFÍA

- Agterdenbos, J. (1979). Calibration in quantitative-analysis .I. General-considerations. *Analytica Chimica Acta*, 108, 315-323.
- Ahlberg, J.H., Nilson, N., Walsh, J.L. (1967), *The Theory of Splines and Their Application*, New York: Academic Press.
- AMC (1988), Uses (proper and improper) of correlation coefficients. *Analyst*, 113, 1469-1471.
- AMC (1994), ¿Is my calibration linear?. *Analyst*, 119, 2363-2366.
- Andrews, J.A.S., Jowett, A. (1982). A numerical aid for evaluation of atomic-absorption spectrometric results. *Analytica Chimica Acta*, 134, 383-388
- Akhnazarova, S., Kafarov, V. (1982), *Experiment Optimization in Chemistry and Chemical Engineering*, Moscú: Ed. Mir.
- A.G. Asuero (2005), Calibración, comparación de métodos y estimación de parámetros en análisis químico y farmacéutico. *Anales de la Real Academia Nacional de Farmacia*, 71, 153-173.


- Asuero, A.G., González, A.G. (1989), Some observations on fitting a straight line to data. *Microchemical Journal*, 40, 216-225.
- Asuero, A.G. (1993), Evaluation of acidity constants of dibasic acids and diacid bases from liquid-liquid distribution measurements. *International Journal of Pharmaceutics*, 89, 103-110.
- Asuero AG, Sayago A, Gonzalez AG. The correlation coefficient: an overview. *CRC Critical Reviews in Analytical Chemistry*, 36, 41-59.
- Atkinson, A.C. (1985), Plots, Transformations and Regression- an introduction to graphical methods of diagnostic regression analysis, Oxford: Oxford University Press.
- Baumann, K., Watzig, H. (1995), Appropriate calibration functions for capillary electrophoresis. 2. Heterocedasticity and its consequences. *Journal of Chromatography A*, 700, 9-20.
- Belloto, R.J., Sokoloski, T.D.(1985). Residual analysis in regression. *American Journal of Pharmaceutical Education*, 49, 295-303
- Blanco, M., Cerdá, M. (Eds.) (1988), *Quimiometría*, Barcelona: Servei de publicacions de la Universitat Autònoma de Barcelona.
- Bocek, P., Novak, J. (1970), Statistical processing of calibration data in quantitative analysis by gas chromatography. *Journal of Chromatography*, 51, 375-383.
- Bowman, K.O., Shenton, L.R. (1975). Omnibus test contours for departures from normality based on square root and . *Biometrika*, 62, 243-250.
- Bubert, H., Klockenkamper, R. (1983), Precision-dependent calibration in instrumental analysis. *Fresenius Zeitschrift fur Analytische Chemie*, 316, 186-193.
- Carroll, J., Ruppert, D. (1988), *Transformation and Weighing in Regression*, London: Chapman & Hall.
- Clancey, V.J. (1947), *Nature*, 159, 339-340.
- K.A. Connors (1987), *Binding Constants, The Measurement of Molecular Complex Stability*, New York: Wiley.
- Cuadros-Rodriguez, L., Gamiz-Cracia, L., Almansa-Lopez, E.M., Bosque-Sendra, J.M.(2001). Calibration in chemical measurement processes: I. A metrological approach. *Trends in Analytical Chemistry*, 20, 195-206
- Danzer, K. (1989), Robust statistics in analytical chemistry, *Fresenius Zeitschrift fur Analytische Chemie* 335, 869-875.
- Das, R.E.G., Tydeman, M.S. (1982), Iterative weighted regression analysis of logit responses –a computer- program for analysis of bioassays and immuno assays. *Computer Programs in Biomedicine*, 15, 13-21.
- De Levie, R. (2000), Curve fitting least squares. *Critical Reviews in Analytical Chemistry*, 30, 59-74.

- De Levie, R. (1986), When, why and how to use weighted least squares. *Journal of Chemical Education*, 63, 10-15.
- Desimoni, E. (1999). A program for the weighted linear least-squares regression of unbalanced response arrays. *Analyst*, 124, 1191-1196.
- Draper, N.R., Smith, H. (1998), *Applied Regression Analysis*, 3th ed., New York: Wiley.
- Finney, D.J., Bioassay and the practice of statistical-inference. *International Statistical Review*, 47, 1-12.
- Frigieri, P.F., Rossi, F.B., Calibration of photographic emulsions by cubic spline functions and application in spectrochemical analysis. *Analytical Chemistry*, 51, 54-57.
- Garden, J.S., Mitchell, D.G., Mills, W.N. (1980). Non-constant variance regression techniques for calibration-curve based analysis. *Analytical Chemistry*, 52, 2310-2315
- Gonzalez, A.G., Asuero, A.G., (1993), Computational program for validating analytical methods. *Fresenius Journal of Analytical Chemistry*, 346, 885-887
- Gort S.M., Hoogerbrugge, R. (1995). A user-friendly spreadsheet program for calibration using weighted regression. *Chemometrics and Intelligent Laboratory Systems*, 28, 193-199
- Heilbronner, E. (1979), Position and confidence limits of an extremum. Determination of the absorption maximum in wide bands. *Journal of Chemical Education*, 56, 240-243.
- Herrero, A., Ortiz, Mc., Arcos, J., López-Palacios, J., Sarabia, L. (1994), Multiple standard addition with latent variables (MSALV). Application to the determination of copper in wine by using differential-pulse anodic-stripping voltammetry. *Analytica Chimica Acta*, 293, 277-293
- Hunter, J.S. (1981), Calibration and the straight-line –current statistical practices. *Journal of the Association of Official Analytical Chemists*, 64, 574-583
- Jochum, C., Jochum, P., Kowalski, B.R. (1981), Error propagation and optimal performance in multicomponent analysis. *Analytical Chemistry*, 53, 85-92
- Hwang, J.D., Winefordner, J.D. (1988), Regression methods in analytical chemistry. *Progress in Analytical Spectroscopy*, 11, 209-249.
- Jost, W., Rock, H. (1954), A convenient method for the evaluation of vapour liquid equilibria of binary mixtures. *Chemical Engineering Science*, 3, 17-25.
- Kateman, G., Pijpers (1993), F.W., *Quality Control in Analytical Chemistry*, New York: Wiley.
- Ketkar SN, Bzik TJ, (2000), Calibration of analytical instruments. Impact of nonconstant variance in calibration data. *Analytical Chemistry*, 72, 4762-4765.
- Kleijburg, M.R., Pijpers, F.W. (1985), Calibration graphs in atomic-absorption spectrophotometry. *Analyst*, 110, 147-150.
- Koch, G.G. (1985), A basic demonstration of the [-1, 1] range for the correlation-coefficient. *American Statistician*, 39, 201-202.

- Kragten, J. (1990), Least-squares polynomial curve-fitting for calibration purposes (STATCALIBRA). *Analytica Chimica Acta*, 241, 1-13.
- Kvalseth, T.O. (1985), Cautionary note about  $r^2$ . *American Statistician*, 39, 279-285
- Lee, J.D., Lee, T.D. (1982), *Statistics and Computer Methods in Basic*, Amsterdam: van Nostrand Reinhold.
- MacTaggart, D.L., Farwell, S.O. (1992), Analytical use of linear regression 1. Regression procedures for calibration and quantitation. *Journal of the Association of Official Analytical Chemists International*, 75, 594-608.
- Massart DL, Kaufman L, Rousseeuw PJ, Leroy, A. (1986). Least median of squares – a robust method for outlier and model error-detection in regression and calibration. *Analytica Chimica Acta*, 187, 171-179.
- Mc Dowell, L.M. (1981), Effect of detector nonlinearity on the height, area, width, and moments of peaks in liquid chromatography with absorbance detectors. *Analytical Chemistry* 53, 1373-1376.
- Meloun, M., Militky, J., Hill, M., Brereton, R.G. (2002), Crucial problems in regression modelling and their solutions. *Analyst*, 127, 433-450
- Miller, J.C., Miller, J.N. (2000), *Statistics and Chemometrics for Analytical Chemistry*, 4<sup>th</sup> ed. Chichester, Ellis Horwood.
- O'Connell, M.A., Belanger, B.A., Haaland, P.D., Calibration and assay development using the four parameter logistic model. *Chemometrics and Intelligent Laboratory Systems*, 20, 97-114.
- Pare, A., Rius, J., Tomas, X. (1988), Applications of a spreadsheet of calculations – linear regression and variance analysis. *Afinidad*, 45, 325-328.
- Pattengill, M.D., Sands, D.E. (1979), Statistical significance of linear least squares parameters. *Journal of Chemical Education*, 56, 244-247.
- Penninckx, W., Hartmann, C., Massart, DL., Smeyers-Verbeke, J. (1996), Validation of the calibration procedure in atomic absorption spectrometric methods. *Journal of Analytical Atomic Spectrometry*, 11, 237-246
- Pettit, A.N. (1977), Testing the normality of several independent samples using the Anderson-Darling statistic. *Journal Royal Statistical Society C*, 26, 156.
- Price, J.F., Simonsen, R.M. (1962), *Various Methods and Computer Routines for Approximation, Curve Fitting and Interpolation*, Boeing Scientific Research Laboratories, Mathematical Note N° 249, 34.
- Prudnikov, E.D., Shapkina, Y.S. (1984), Random errors in analytical methods. *Analyst*, 109, 305-307.
- Rice, J.R. (1964), *The Approximation of Functions*, Reading: Addison-Wesley,
- Ripley, B.D., Thompson, M. (1987), Regression techniques for the detection of analytical bias. *Analyst*, 112, 377-383

- Rousseeuw, P.J. (1991), Tutorial to robust statistics. *Journal of Chemometrics*, 5, 1-20.
- Sayago A, Asuero A.G. (2004), Fitting straight lines with replicated observations by linear regression: Part II. Testing for homogeneity of variances. *CRC Critical Reviews in Analytical Chemistry*, 34, 133-146.
- Sayago A., Boccio M., Asuero A.G., Fitting straight lines with replicated observations by linear regression: the least squares postulates. *CRC Critical Reviews in Analytical Chemistry*, 34, 39-50.
- Schwartz, L.M. (1979), Calibration curves with nonuniform variance. *Analytical Chemistry*, 51, 723-727.
- Schwartz, L.M. (1977), Nonlinear calibration. *Analytical Chemistry*, 49, 2062-2068.
- Schwartz, L.M. (1978), Statistical uncertainties of analyses by calibration of counting measurements. *Analytical Chemistry*, 50, 980-985.
- Schmidtt, S.A. (1969), Measuring Uncertainty - An Elementary Introduction to Bayesian Statistics, Reading MA: Addison-Wesley,
- Sharaf, M.A., Illman, D.L., Kowalski, B.R. (1986), Chemometrics, New York: Wiley.
- Shapiro, S.S. (1990), How to test normality and other distributional assumptions, American Society for Quality Control, Wisc.: ASQC.
- Staudte, G.G., Sheather, S.J. (1990), Robust Estimation and Testing, New York: Wiley.
- Thompson, M. (1990), Abuse of statistics software package. *Analytical Proceedings*, 27, 142-144.
- Thompson, M. (1982), Regression methods in the comparison of accuracy. *Analyst*, 107, 1169-1180.
- Thompson, M. (1989), Robust statistics and functional relationship estimation for comparing bias of analytical procedures over extended concentration ranges. *Analytical Chemistry*, 61, 1942-1945.
- Tyson, J.F.(1988), Conventional calibration strategies for flame AAS and some conventional alternatives, pp 371-382, in C.S. Creaser and A.M.C. Davies (Eds.), "Analytical Applications of Spectroscopy", London: RSC.
- UNE 82009-2 (1999). Exactitud (veracidad y precision) de resultados y métodos de medición. Parte 2. Método básico para la determinación de la repetibilidad y la reproducibilidad de un método de medición normalizado (ISO 5725-2: 1994), Madrid: AENOR.
- Valkó, P., Vajda, S. (1989). Advanced Scientific Computing in BASIC with applications in Chemistry, Biology and Pharmacology, Amsterdam: Elsevier.
- Wagenaar, H.C., Novotny, I., Degalan, L. (1974), Influence of hollow-cathode lamp line-profiles upon analytical curves in atomic-absorption spectroscopy. *Spectrochimica Acta Part B-Atomic Spectroscopy B*, 29, 301-317

Walczak, B., Massart, D.L. (1998), Multiple outlier detection revisited. *Chemometrics and Intelligent Laboratory Systems* 41, 1-15.

Zorn. M.E., Gibbons, R.D., Sonzogni, W.C., Weighted least squares approach to calculating limits of detection and quantification by modelling variability as a function of concentration. *Analytical Chemistry*, 69, 3069-3075. 

Referencia	Fecha de recepción	Fecha de aprobación
González A. G., Sayago A., y García, A. (2006). La problemática del ajuste en la etapa de la calibración. <i>Revista Tumbaga</i> , 1, 107-128	Día/mes/año 29/01/2006	Día/mes/año 13/06/2006