

Anal. Real Acad. Nac. Farm., 2005, 71: 153-173

Revisión

Calibración, comparación de métodos y estimación de parámetros en el análisis químico y farmacéutico (*)

AGUSTÍN GARCÍA ASUERO

Académico Correspondiente de la Real Academia Nacional de Farmacia

RESUMEN

En este trabajo se pasa revista a la problemática del ajuste de una línea recta cuando se realizan observaciones replicadas. Este tópico se encuentra estrechamente relacionado con cuestiones básicas, tales como curvas de calibrado relacionando el valor medido de una respuesta con la propiedad de un material, comparación de dos métodos analíticos aplicado a diferentes concentraciones de material, relaciones en las que el tiempo es la variable x , y estimación de parámetros. El tema reviste en adición un gran interés en el ámbito de las medidas físicas y químicas, de tanta aplicación en las Ciencias Farmacéuticas y afines.

Palabras clave: Mínimos cuadrados.—Replicación.—Líneas rectas.

SUMMARY

Fitting straight lines with replicated observations is considered in this paper. This topic is closely related to very basic operations, for example, calibration curves relating measured value of response to a property of materials, comparison of two analytical methods applied to a range of test materials, relationships in which time is the x -variate, and parameter estimation methods. The subject has a great

(*) Discurso de Toma de Posesión como Académico Correspondiente, leído en la Junta Pública de la Real Academia Nacional de Farmacia, celebrada el jueves 17 de junio de 2004.

interest in the field of chemical and physical measurements, which are widely applied in the pharmaceuticals and related sciences.

Key words: Least squares.—Replicated observations.—Straight lines.

EXTENSIVE ABSTRACT

Calibration, method comparison and parameter evaluation in chemical and pharmaceutical analysis

Several aspects of least squares, particularly in regard to the use of replication, error analysis and weighing and data transformations, appears to be poorly understood by a number of experimenters. Examples are given from older and more recent literature where experimental data were processed in an incorrect way from the point of view of statistics. As a matter of fact, however, the statistical methods most commonly misapplied by analytical chemists are correlation and regression. There is no doubt of the importance of these topics, which are closely related to very basic operations, e.g., calibration and the comparison of two analytical methods applied to a range of test material. In analytical chemistry as well as in other quantitative sciences it is often necessary to fit a mathematical equation or model to experimental data. Common situations that may be described by functional relationships include calibration curves relating measured value of response to a property of material, comparison of analytical procedures, relationships in which time is the x-variate and parameter estimation methods. Parameters of the approximating function, however, are frequently derived using the least-squares methodology.

To demonstrate that a least squares criterion is valid it is necessary to assume: i) that the errors, ε_i , are random rather than systematic, with mean zero and variances $\sigma_i^2 = \sigma^2/w_i$ (where σ is a constant and w_i is the weight of point) and follow a gaussian distribution; this distribution is so common that is also referred to as the normal one; ii) that the independent variable, i.e., x , the abscissa, is known exactly or can be set by the experimenter either; iii) the observations, y_i , are in an effective sense uncorrelated and statistically independent, i.e. for $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$, with means equal to their respective expectations or true values, $E\{y_i\} = \eta_i$; and (iv) that the correct weights, w_i , are known. The least squares criterion gives indeed poor results, however, if the observations are incorrectly weighted or if the data contain «outliers», i.e., very poor observations at higher frequency than allowed for by the normal distribution. When the conditions are met, the parameter estimates found by minimization of a least squares criterion are best unbiased linear estimates of the regression parameters.

Real data are often subject to problems that make the use of classical statistics based on the normal distribution, difficult. The main practical problem probably is the occurrence of outliers. Another difficulty can be that the distribution of the data is not normal. The normality assumption is, in fact, quite reasonable to expect the y_i to be independent in many situations if they are the results of separate

isolated non interfering measurements. Thus, assumption normality, nevertheless, is a plausible assumption as an error term is made up of the combination of a large number of small chance effects arising from several sources. Such a combination tends to produce a normal distribution, regardless of the distribution of the separate errors (the Central Limit Theorem) if its variance is finite.

In the context of most calibration problems the assumption relative to the abscissa variable is reasonable because the analyte concentrations (x values) are precise enough. Particular attention must be given to equation in which one variable is involved on both sides. Then an error in this quantity appears in both coordinates mutually correlated in both conditions, i.e., the independent variable x is not an exact quantity and the independence of errors is not fulfilled.

When the abscissa range, e.g. concentration, span several orders of magnitude, the precision of the y values vary greatly over the range of the x values. There two main solutions to the problem of non constant variance: Transform the data, or perform a weighted least squares regression analysis as several authors have pointed out is a better solution.

It is obvious that estimates of error variances independent both of the assumed model and the method of fitting can only be obtained from replicates at each point. It is important to understate that repeated runs must be genuine repeats and not just repetitions of the same reading. The question of how many replicate measurements to take must include consideration of the magnitude of variability, availability of the test material and reagent, the time required, the cost of each measurement, and the variability required in the final result. Even within the concepts that are based on the construction of a calibration curve, there is no consensus about the choice of calibration samples and the number of replicates. As the number of replicate increases, however, the central limit theorem states that the frequency distribution for the mean value approaches normality (very rapidly indeed, especially if the parent distribution is symmetric). This fortunate circumstance provides a very important, but little recognized basis for replication of analyses.

INTRODUCCIÓN

Las técnicas estadísticas se basan en suposiciones, y la validez de los resultados obtenidos en la práctica depende de que las condiciones supuestas se satisfagan, al menos con un grado suficiente de aproximación. Exner ha suministrado recientemente ejemplos variados en los que los datos experimentales se procesan de manera incorrecta desde el punto de vista estadístico. Vamos a tratar la problemática del ajuste de una línea recta a un conjunto de datos biva-riantes, cuando se llevan a cabo observaciones replicadas, sin perder de vista el ámbito de las medidas físicas y químicas, de tanta aplicación en Química y Farmacia. La «Conferencia Internacional sobre

Harmonización (ICH)», por ejemplo, establece el análisis por regresión para definir la vida media de los fármacos en su guía sobre «Ensayos de Estabilidad de Nuevos Fármacos y Productos».

Los modelos sencillos facilitan en gran número de ocasiones la interpretación de complejos fenómenos físico-químicos. Si existe un competidor al «test t» en popularidad, probablemente éste sea la regresión. En adición, quizá la solicitud más común de ayuda que reciben los estadísticos de sus colegas no estadísticos sea el ajuste de una relación lineal a un conjunto de datos. A pesar de esto, correlación y regresión son los métodos más mal aplicados por los analistas. No hay duda de la importancia de estos temas, relacionados con cuestiones básicas tales como la calibración, la comparación de dos métodos analíticos aplicados a materiales de ensayo, a diferentes concentraciones, o la estimación de parámetros. Por otra parte, la introducción de regulaciones para el control de la producción de alimentos, de productos farmacéuticos y del medio ambiente, ha originado un gran interés en la validación de los métodos analíticos.

POSTULADOS DE LOS MÍNIMOS CUADRADOS

En las Ciencias Farmacéuticas, como en otras ciencias cuantitativas, es a menudo necesario ajustar una ecuación matemática o modelo a un conjunto de datos experimentales. Situaciones que pueden describirse por medio de relaciones funcionales incluyen curvas de calibración que relacionan un valor medido o respuesta con la propiedad de un material, comparación de procedimientos analíticos y relaciones en las que la variable x es el tiempo. A este respecto, Deming enfatiza que algunos investigadores todavía «ajustan los datos a un modelo», lo que sugiere una falta de integridad científica. Lo que hacen, naturalmente, es ajustar un modelo a los datos.

Si son necesarias meras estimaciones de parámetros, puede emplearse cualquier criterio de acuerdo con las preferencias estéticas; la suma mínima de los módulos, el criterio de *minimax*, etc. El problema se presenta cuando es necesario deducir inferencias acerca de la fiabilidad de los parámetros y/o sobre la fiabilidad del modelo. Bajo condiciones ideales, el método de los mínimos cuadrados es el preferido para ajustar ecuaciones teóricas a un conjunto de

datos experimentales. La suma (ponderada) de los cuadrados de los residuales —desviaciones con respecto a la función de ajuste—, se hace tan pequeña como sea posible. Su principal ventaja es que suministra una estimación de la incertidumbre de los parámetros. El ajuste estadístico de una línea recta se conoce generalmente como regresión lineal, en donde la palabra regresión tiene tan sólo un significado histórico.

Para demostrar que un criterio de mínimos cuadrados (ponderados) es válido es necesario asumir: i) que los errores cometidos son aleatorios y no sistemáticos, con medio cero y varianza no uniforme; ii) que siguen una distribución gaussiana; esta distribución es tan común que nos referimos a ella como normal; iii) que la variable independiente, i.e., x , la abscisa, se conoce exactamente o puede ser establecida (fijada) por el experimentador; iv) que las observaciones y_i son estadísticamente independientes, esto es, no se encuentran correlacionadas, coincidiendo las medias con sus respectivas expectativas o verdaderos valores; y v) que los pesos correctos, números positivos, son conocidos; esto requiere conocer, a su vez, la forma funcional de la dependencia de x con la varianza de y . El criterio de los mínimos cuadrados da lugar, no obstante, a resultados cuestionables si las observaciones no se ponderan correctamente o si los datos contienen resultados atípicos, esto es, observaciones (pobres) a frecuencias mayores que las permitidas por la distribución normal.

Se asume el caso general de precisión no uniforme o heterocedasticidad; la cantidad medida es determinable no con varianza constante (homocedástica) sino dependiente de su magnitud. De hecho, en muchos fenómenos, conforme el nivel de la señal aumenta, se incrementa también el del ruido. De esta manera, se obtiene una aproximación más consistente, que permite, en adición, el uso de un modelo de varianza para definir los posibles pesos. Aunque la regresión lineal ponderada es bien conocida, las aplicaciones en análisis químico y farmacéutico no son muy amplias, presumiblemente porque este tipo de estadística y la forma en cómo se trata en el software estadístico, no son especialmente atractivas.

Muchos investigadores no son conscientes de que los datos que publican violan, claramente, al menos, alguna de las suposiciones inherentes al modelo de regresión. De hecho, las desviaciones de

dichas suposiciones son la regla en el análisis químico y físico. No sólo se dispone en contadas ocasiones de la información exacta concerniente a la relación funcional, sino que, en general, se presentan errores sistemáticos y existe evidencia de que las observaciones siguen distribuciones que tienen colas más alargadas que la normal. «*Muchos, si no la mayor parte de los análisis, pueden tener un sistema de error leptocúrtico*», indicaba Student en los veinte. De hecho, una característica de la mayor parte de los análisis químicos que actúa en contra de la eficacia de muchos métodos clásicos de contrastes de hipótesis es el pequeño número de muestras procesadas. El poder de los métodos estadísticos para detectar pequeñas discrepancias se encuentra severamente limitado por el número de observaciones.

Normalidad

Un aspecto del análisis de datos en el que la práctica habitual no es a veces soportada por la evidencia experimental es la suposición de que las observaciones están normalmente distribuidas. Esta suposición es crítica, ya que desviaciones de la misma pueden ser la causa de que el método de los mínimos cuadrados conduzca a resultados dudosos, incluso con una correcta ponderación. La suposición de que las respuestas están distribuidas normalmente se hace con frecuencia para calcular intervalos de confianza, ensayos de efectos significativos o comparaciones de datos adicionales. Existen ensayos para comprobar la normalidad, pero requieren, en general, más observaciones que las que se llevan a cabo, en general, en los experimentos usuales. Incluso cuando se dispone de datos abundantes, lo usual es no aplicar ensayo alguno.

La condición de normalidad se asume pues *a priori*, a menos que existan objeciones teóricas o indicaciones empíricas en su contra. Aunque Tukey y MacLaughlin sugieren que la distribución normal es tan rara que debería denominarse patológica, su asunción no ocasiona serios errores. La suposición de normalidad, por otra parte, es plausible, ya que un término de error se obtiene como una combinación de un gran número de pequeñas causas (errores aleatorios independientes). Tal combinación tiende a producir una distribución normal con independencia de la distribución de los errores separados

(Teorema Central del Límite) si su varianza es finita. Puesto que la mayor parte de los experimentos implican muchas operaciones para establecer los resultados de las medidas, es razonable suponer que las perturbaciones o errores estarán normalmente distribuidos. De ahí que la normalidad sea una suposición muy frecuente en el análisis por regresión. Los parámetros del modelo se estiman, en la mayor parte de los casos, por mínimos cuadrados, puesto que son eficientes si los errores están normalmente distribuidos.

Muchos resultados normales se satisfacen para poblaciones no normales, y respuestas no normales pueden ser transformadas en normales. Una transformación que estabiliza la varianza es capaz, a menudo, de transformar una variable no normal, sesgada, en una variable razonablemente simétrica y aproximadamente normal. La transformación de escala logra a menudo pues, un doble efecto beneficioso. Hay que tener en cuenta que si la variable original se distribuye normalmente, la variable transformada puede no hacerlo. Con frecuencia, sin embargo, la falta de constancia de la varianza (heterocedasticidad) se asocia simultáneamente con la ausencia de normalidad y la transformación que da lugar a una varianza constante (homocedasticidad) también origina simultáneamente una distribución próxima a la normal.

Robusteza

Una o más observaciones, que no sigan la misma pauta que el resto de los datos, pueden ejercer una gran influencia sobre el modelo de regresión. Los datos reales están sujetos a problemas que dificultan el uso de la estadística clásica basada en la distribución normal. El principal problema práctico es la existencia de datos atípicos. Otra dificultad radica en que, a veces, la distribución de los datos no es normal. Esto puede tratar de corregirse con ensayos para eliminar datos anómalos o con procedimientos que conviertan la distribución en normal. La detección y el rechazo de datos anómalos, en particular, no es a menudo evidente, como indican Walczak y Massart, y una posible alternativa entonces es usar procedimientos estadísticos que sean robustos frente a los datos atípicos o frente a las desviaciones de la normalidad.

Se ha despertado un interés considerable en la estadística «no paramétrica», describiéndose una variedad de métodos robustos de regresión, válidos frente a violaciones de las suposiciones clásicas. Tales herramientas requieren sólo simples suposiciones, tales como aleatoriedad, independencia y simetría. El cálculo de la regresión robusta conlleva mucho tiempo y está ideado para uso con varios errores de distribución que pueden presentarse en la práctica, así como con datos normales contaminados con observaciones indeseables. Los métodos robustos son menos exactos que la estimación por mínimos cuadrados cuando las suposiciones de éstos son ciertas, pero más exactos —a veces mucho más— cuando ellas son falsas, esto es, son insensibles a las violaciones de estas suposiciones. Métodos intensivos de computación de inferencias como «bootstrap» o «jackknife» son, por otra parte, herramientas poderosas con sólo unas cuantas suposiciones referentes a la distribución.

Independencia

La suposición de independencia implica que las perturbaciones en diferentes experimentos son independientes unas de otras, esto es, la perturbación de series separadas no se encuentran sistemáticamente relacionadas, una suposición que se satisface de manera apropiada mediante aleatoriedad. Debe también confirmarse que los errores son independientes o, en otras palabras, no correlacionados con alguna variable. En muchas situaciones es bastante razonable esperar que las y_i sean independientes si son resultado de medidas aisladas, separadas, no interferentes. Sin embargo, en el trabajo de calibración puede aparecer correlación entre los términos de error si la muestra se lleva a cabo de forma inapropiada o las condiciones analíticas varían en función del tiempo (temperatura, degradación, evaporación, dispositivo de deriva, etc.), tal como reconoce Baumann.

La causa más común de dependencia en las respuestas es que sean tomadas de forma secuencial en el tiempo o de una manera sistemática, lo que introduce generalmente una correlación positiva entre las observaciones. La correlación serial o autocorrelación en los datos representa, de esta manera, una violación en lo que respecta a la independencia de los errores de medida.

La correlación pasa a menudo desapercibida. El análisis de series temporales incorpora la estructura de correlación en el modelo usado para analizar los datos. Se aplican los criterios de Durbin y Watson o de Durbin, en orden a comprobar el postulado de independencia como muestran Draper y Smith. Si los y_i son obtenidos a través de alguna combinación funcional de un número de valores medidos, en general, no serán independientes y debe emplearse un formalismo completo. Las perturbaciones no independientes pueden tratarse por mínimos cuadrados generalizados, pero, como en el caso en el que la varianza no es constante, las modificaciones al modelo pueden hacerse bien a partir de la información adquirida por los datos o mediante suposiciones adicionales tales como la naturaleza de la interdependencia.

Abscisa libre de error

Se asume que todos los errores se presentan en los valores medidos de la ordenada y , y_i , y que los errores en los valores de x son despreciables en relación con los de y . Esta no es una restricción en la práctica. Si los errores en x son los mayores, x e y pueden intercambiarse. Es la estructura del experimento antes que la conveniencia o confort del programador lo que determina cuál es la variable independiente y cuál la dependiente.

La regresión lineal por mínimos cuadrados se aplica a menudo para determinar un modelo matemático de calibración que aproxima la relación entre la concentración y la respuesta. La preparación de estándares siempre conlleva un error. Sin embargo, los errores en la x no tienen consecuencias si son menores a una décima parte de los errores en la y . Si el error en la x es mayor, entonces el error total se incrementa significativamente. Además, los parámetros de regresión y los intervalos de confianza calculados a partir de una curva de calibrado son, en dicho caso sesgados, si se usan mínimos cuadrados (ordinarios) ponderados.

En el contexto de la mayor parte de los problemas de calibración, la suposición relativa a la variable abscisa es razonable porque las concentraciones analíticas (valores de x) son suficientemente precisos. La suposición de que los errores se presentan solamente en la

«dirección y», por otra parte, es válida en muchos experimentos; los errores en las señales instrumentales son a menudo de, al menos, un 2-3 por 100 RSD (desviación estándar relativa), mientras que los errores en la preparación de los estándares no deben superar la décima parte de esto.

Los errores en la preparación de los estándares deben reducirse por debajo del 0,05 por 100, con objeto de aprovechar los beneficios que se derivan del uso de un equipo instrumental preciso. Esto resulta siempre posible usando balanzas y equipos volumétricos precisos. La precisión de operaciones volumétricas, tales como la dilución, se mejora a menudo usando balanzas en lugar de pipetas y pesando frascos para el cálculo de volúmenes o utilizando estándares internos. Las modernas técnicas automáticas han mejorado considerablemente la precisión de muchos métodos instrumentales; cromatografía gas-líquido (GL), cromatografía líquida (LC) y electroforesis capilar (CE) suministran señales repetidas con una precisión del 0,5 al 1 por 100; las técnicas espectroscópicas se comportan igual o mejor y el análisis por inyección en flujo muestra muchos ejemplos de RSD del 0,5 por 100 o menores. En tales casos puede ser necesario, bien abandonar la suposición de que x se encuentra libre de error, o mantener la validez de la suposición preparando los estándares gravimétricamente, en lugar de volumétricamente, esto es, lograr una exactitud mayor que la usual.

En el caso de curvas de valoración lineales, el valor de abscisa es el volumen de valorante añadido, que puede considerarse exento de errores aleatorios. Las medidas de los volúmenes pueden efectuarse con una gran precisión, usando una jeringa adecuada o una micropipeta con un dispositivo automático.

En estudios cinéticos, « y » es alguna función que representa la concentración de los materiales de partida o de los productos de reacción, mientras que « x » representa una escala de tiempo. Una práctica común en la determinación de un mecanismo cinético es evaluar las constantes de velocidad a un elevado número de concentraciones. En general, se supone, que en todos los casos las concentraciones se conocen con exactitud, y por tanto que sólo las velocidades contienen términos de error. Las mismas consideraciones se aplican en cinética enzimática; la precisión con la que se conocen las concentraciones de sustrato depende de la exactitud del pipeteo en la preparación de

mezclas de reacción, y si se llevan a cabo diferentes diluciones a partir de una disolución patrón y se adoptan las precauciones oportunas, los errores aleatorios resultantes en las concentraciones de sustrato serán pequeños. Por tanto, es casi siempre razonable suponer que las concentraciones de sustrato se conocen mucho más exactamente que las velocidades de reacción, y en consecuencia, no es una suposición demasiado arriesgada tratar todos los errores achacándolos a la velocidad, lo que normalmente se hace.

La exactitud y precisión del dispositivo usado cuando se mide la absorbancia de una especie transitoria en función del tiempo, por ejemplo, es muy superior a la medida de la absorbancia, en cuyo caso se justifica considerar sólo los errores aleatorios en esta última. En algunos métodos analíticos, no obstante, tales como fluorescencia de rayos X, se usan a menudo como estándares de calibración materiales de referencia certificados, ya que las muestras reales (i.e., materiales geológicos) son demasiado complejas. Por esta razón las incertidumbres, como contemplan «Rius y col.» se encuentran asociadas a ambos valores de concentración de los materiales de referencia y a las respuestas instrumentales.

Aunque puede haber muchos experimentos donde es razonable suponer que una variable está ampliamente exenta de errores, existen otros en los que tal suposición es manifiestamente absurda, como aquellos casos en los que ambas variables se calculan a partir de la misma observación. Debe prestarse particular atención a la ecuación en la que una variable se encuentra situada en ambos miembros, de especial importancia en los «métodos de estimación de parámetros». En estos casos, un error en esta cantidad aparece en ambas coordenadas mutuamente correlacionadas; la variable independiente x no es una cantidad exacta y la independencia de los errores no se satisface.

En el estudio de comparación de métodos se examina un número de muestras por cada uno de los dos métodos a estudiar, y los dos conjuntos de resultados obtenidos se representa en los ejes x - y . Cada punto de esta gráfica representa, por tanto, una muestra examinada por los dos métodos. En este ejemplo, es obvio que los errores de medida deben producirse en ambas direcciones x e y , como se trata en los trabajos de Mac Taggart y Farwell, Martín, y Rius y col., entre otros autores.

En general, cuando los valores para x e y se obtienen mediante medidas: $y_i = \eta_i + \varepsilon_i$, $x_i = \xi_i + \delta_i$, donde x_i e y_i son los valores medidos de las variables. Cuando ambas variables contienen errores, la distinción entre variables dependiente e independiente es ambigua aunque en general se atiende a controlar una de ellas x , y observar la otra y . Sus verdaderos valores son η_i y ξ_i y sus respectivos errores ε_i y δ_i . Note, que se obtienen resultados sesgados cuando se usa el método de los mínimos cuadrados ponderados en la comparación de dos métodos, en base a la suposición de que los valores de x son conocidos sin error alguno. En este contexto pueden plantearse situaciones erróneas cuando se usan los mínimos cuadrados ponderados en el cálculo de los coeficientes de regresión.

Si por alguna razón la precisión con la que los valores de x son conocidos no es considerablemente mejor que la precisión de las medidas de los valores de y , el análisis estadístico basado en el método (ordinario) de los mínimos cuadrados ponderados no es válido y se requiere una aproximación más general. El problema fundamental que surge si las desviaciones se miden en cualquier dirección distinta a la paralela con uno u otro eje, es que tales desviaciones no poseen dimensiones adecuadamente definidas (excepto en el caso usual en el que x e y posean la misma dimensión y, naturalmente, las mismas escalas). La consecuencia práctica de esto es que el ajuste obtenido en este caso dependerá arbitrariamente de las escalas elegidas para la representación. La solución de este problema, que plantea la regresión ortogonal, se aborda mediante diversas estrategias.

Pesos apropiados

Cuando el rango de la abscisa, por ejemplo concentración, se extiende a varios órdenes de magnitud, como ocurre con la calibración, en aquellos casos en los que se investigan concentraciones de drogas en orina u otros fluidos biológicos, la precisión de los valores de y puede variar grandemente a lo largo del rango de valores de x . Esta condición contraviene el requerimiento de homocedasticidad de la regresión lineal (simple) no ponderada. Con la espectrometría de emisión de plasma acoplada a la espectrometría de masas, se

requiere el uso de mínimos cuadrados ponderados, incluso cuando la calibración se lleva a cabo sobre un rango de concentraciones relativamente estrecho.

Matsuda et col. sostienen que debe efectuarse un análisis de ruido y considerar siempre la regresión ponderada como el modo general y la regresión simple como el especial. Casi siempre se encuentra que la precisión absoluta de la determinación, esto es, la desviación estándar, incrementa con la concentración, mientras que la precisión relativa (la desviación estándar relativa) disminuye con la misma. Un tipo común de heterocedasticidad se presenta en la práctica cuando los errores poseen una magnitud relativa constante. Es posible establecer una relación entre la precisión y la concentración sobre el rango de concentración ensayado, cuando los métodos propuestos se aplican a diferentes concentraciones. Varios autores han sugerido relaciones y la norma ISO 5725 (1986) suministra guías para el establecimiento de la existencia de una relación dada. A los datos más fidedignos (menor variabilidad) hay que darles mayor énfasis o peso. De esta manera, no tiene que recalcularse la respuesta frente a la concentración, puesto que los datos originales permanecen inalterados.

El método de los mínimos cuadrados es una herramienta poderosa para el tratamiento de datos, pero sus ventajas pueden estar contrarrestadas si no se contemplan los pesos apropiados. El problema se agudiza debido al hecho de que el criterio de los mínimos cuadrados es altamente sensible a los datos anómalos y a menudo se produce una situación paradójica en la que la observación reconocida como la peor, es la que contribuye más fuertemente a la estimación de los parámetros. Aunque la replicación puede constituir una severa limitación (desde el punto de vista experimental), también posee la ventaja de suministrar una especie de regresión robusta. El método más común de llevar a cabo la regresión ponderada es simplemente usar para los pesos el recíproco de las varianzas. Esta relación asegura que si se emplea replicación, los valores anómalos de y_i tengan pesos más bajos.

Bondad de ajuste del modelo

Se sabe que si el modelo lineal es correcto, los residuales (diferencias entre los valores experimentales y los calculados), obtenidos por mínimos cuadrados, pueden usarse para estimar la varianza del error, y no son necesarias medidas repetidas. Sin embargo, si el modelo lineal no es aplicable, los residuales estiman la suma de un error de ajuste y el error experimental. Es obvio que las estimaciones de la varianza del error independiente de ambos modelos, asumido y método de ajuste, pueden sólo obtenerse mediante replicaciones en cada punto (Feinberg). Por tanto, una comparación de la suma de los cuadrados de los residuales con la suma de los cuadrados de los errores obtenidos de los replicados, suministra un test de la bondad de ajuste del modelo.

MÚLTIPLES MEDIDAS A UNO O MÁS PUNTOS

Con frecuencia es útil realizar experiencias en las que se preparan una o más muestras a los mismos valores de la variable de entrada x . El término conjunto se define como referente a un número de medidas repetidas, independientes, de la misma propiedad. Las medidas repetidas de tal manera que estén sujetas a todas las fuentes de errores aleatorios del experimento se llaman replicados. En aquellos casos en los que estas muestras repetidas se preparan de tal manera que están sujetas a todas las fuentes plausibles de error, se denominan replicados genuinos. La replicación se define así como la realización independiente de dos o más experimentos al mismo nivel, encontrándose todos los factores bajo control.

Para evitar ambigüedad la IUPAC (1994), sin embargo, indica que el término «replicado» o «replicación» debe utilizarse sólo en el contexto de la medida (análisis) y no en el sentido de «separación de múltiples unidades» o recolecta de replicados, a menos que su uso sea explícito. Cuando no todos los factores están presentes, tenemos pseudoreplicados, no válidos para estimar la varianza de la distribución a partir de las observaciones realizadas.

Supongamos que existen k especímenes (muestras) de una variable simple (población normal) a ser analizados. Para cada espécimen

se lleva a cabo un número diferente de replicados: $y_{j,nk}$, nk observaciones repetidas a y_j . Una o más medidas reales replicadas se considera una muestra aleatoria de esta población hipotética infinita. El tamaño (estadístico) de la muestra es el número de medidas que la constituyen. Así, un conjunto de tres medidas replicadas es de tamaño 3 (no tres muestras). Debe prestarse atención al desafortunado pero inevitable uso dual de la palabra «muestra» con dos significados distintos: la muestra química y la muestra estadística.

Los resultados replicados para cada muestra se distribuyen de forma dispersa alrededor de su valor medio, y_j , debido a los errores aleatorios de medida (AMC, 1994); siguiendo estas fluctuaciones la distribución normal. Estas fluctuaciones son impredecibles, debido a los factores experimentales que no se encuentran bajo control rígido y a las limitaciones mecánicas inherentes a los aparatos de medida experimental y a veces también, a la variabilidad inherente al fenómeno estudiado. Llamamos homogéneo a cualquier conjunto de medidas iguales, una vez descartado el error experimental. Para tales conjuntos existen reglas para el rechazo de datos anómalos basados en la teoría de la distribución normal.

La respuesta de interés depende en adición, de la propiedad medida y también de otros factores, tales como la temperatura que rodea la muestra y la humedad relativa, el nivel del instrumento, las impurezas de los reactivos químicos, la exactitud de la balanza o la destreza del operador. Se trata siempre de mantener el control sobre los factores ambientales, y los factores que puedan afectar al proceso de medida, aunque lograr dicho control de forma total es humanamente imposible.

Puede ocurrir incluso en experimentos bien diseñados, que algunos resultados se pierdan debido a un error grosero, rotura del equipo o alguna otra razón. La pérdida de los resultados no puede siempre recuperarse por medio de ensayos repetidos, aunque sea posible esto en experiencias sencillas. No puede excluirse tampoco el caso en que resulte un número desigual de replicados como consecuencia del diseño del experimento.

Replicados genuinos

Es importante comprender que las replicaciones, como insisten Draper y Smith, deben ser muestras repetidas genuinas y no simples repeticiones de la misma lectura. En experiencias químicas, una sucesión de lecturas hechas durante el estado de equilibrio no suministra puntos genuinos repetidos. Cuando los replicados genuinos se realizan bajo un conjunto dado de condiciones experimentales, la variación entre las observaciones asociadas puede emplearse para estimar la desviación estándar de los efectos.

Por replicados genuinos se entiende la variación entre muestras preparadas en las mismas condiciones experimentales, un reflejo de la variabilidad total que afecta a series preparadas a diferentes condiciones experimentales. Este punto requiere cuidadosa consideración. En particular, varios análisis químicos a partir de una muestra sencilla suministran sólo una estimación de la varianza analítica, generalmente solo una pequeña parte de la varianza «run-to-run». Este problema de determinar, de manera errónea, la varianza experimental, ha sido, en general, particularmente inoportuno.

Un beneficio obvio de la replicación es que mejora la fiabilidad de los resultados. Otro beneficio es la facilidad de ensayo de la bondad de ajuste del modelo. La incertidumbre puramente experimental puede obtenerse sólo controlando todos los factores a niveles fijados de antemano y replicando el experimento.

En la preparación y análisis de una muestra de control que comprende las operaciones preliminares de secado, pesada, dilución y posterior doble inyección de una muestra simple preparada siguiendo esta pauta en una columna de HPLC, los únicos factores que pueden producir diferencias entre los resultados medidos son los que operan a partir de la etapa de inyección hacia adelante, esto es, inyección, separación y detección (Mullins, 1994). De acuerdo con esto, dos replicados genuinos en este caso suponen no sólo diferencias debidas a estos últimos factores, sino también diferencias debidas a los errores de secado, pesada y dilución. Por tanto, los replicados parciales podrían sobreestimar seriamente el tamaño del error aleatorio adscrito a las mediciones.

Las repeticiones que comienzan en una etapa tardía (alícuotas de la misma porción de muestras de ensayo disueltas) no suministran (IUPAC, 1994) una estimación de la repetibilidad, puesto que la variabilidad introducida por la omisión de las restantes etapas no se incluye en la medida final. Las medidas repetidas realizadas sobre una disolución muestra, con un instrumento, suministran sólo una estimación de la precisión instrumental.

Número de replicados (normalidad)

Los replicados permiten estimar las magnitudes de las variaciones aleatorias, y es de esperar que las medias de los replicados, si están ausentes errores sistemáticos, se encuentren más próximas al valor verdadero que las lecturas individuales. La cuestión de cuántas medidas replicadas debemos hacer incluye consideraciones acerca de la magnitud de la variabilidad, disponibilidad de reactivos y material de ensayo, tiempo requerido, coste de cada medida y variabilidad requerida en el resultado final.

Esta es una cuestión para la que no hay una contestación clara y simple. Incluso dentro del contexto de la curva de calibrado (IUPAC, ISO y AOAC, 1989; IUPAC, 1990), no existe un consenso acerca de la elección de las muestras de calibración y del número de replicados. Cinco replicados en cada grupo han sido propuestos por Jacquez y col. (1968). Demasiados replicados suponen un esfuerzo adicional (Castillo y Castell, 2001), mientras que con pocos no se consigue la sensibilidad requerida. Tres o cuatro replicados no son suficientes para estimar adecuadamente una varianza; al menos se requieren generalmente de 8 a 10. El Comité de Mejora Ambiental de la ACS (1980) recomienda tres replicados (replicados experimentales) a cinco valores de concentración igualmente espaciadas. No obstante, la reducción en trabajo experimental (diseño de calibración y menores replicados experimentales e instrumentales), no implica necesariamente, según González Casado y col. (1998) una pérdida de información analítica.

Conforme el número de replicados aumenta, la estimación de la varianza total mejora, como es obvio. Sin embargo, superficialmente, uno podría estar tentado en contestar: cuanto más mejor,

basándonos en el teorema del error estándar de la media. Tal razonamiento, sin embargo, es una falacia, como muestra Mandel, dado que el error de replicación es sólo una porción, algunas veces pequeña, del error total. A este respecto deben considerarse además la significación estadística y la práctica, como nos recuerda Davies. Los errores sistemáticos no se reducen con el proceso de promedio.

Conforme el número de replicados aumenta, el teorema central del límite establece, que la distribución de frecuencia para el valor medio se aproxima a la normalidad (muy rápidamente si se parte de una distribución simétrica). Esta afortunada circunstancia suministra una muy importante pero poco reconocida base para la replicación de los análisis. En efecto, la replicación nos permite suponer normalidad (para los valores de la media, no para los valores individuales), una suposición que es bastante difícil de sostener de otra manera; la media de cuatro observaciones es ya muy próxima a la normal. La buena práctica analítica toma este efecto en cuenta cuando se establece el mínimo número deseable de observaciones. En particular, las distribuciones uniforme y binomial suministran un ejemplo ilustrativo de aproximación a la normalidad, siendo posible generar números aleatorios con distribución normal a partir de las mismas (Güell y Holcombe, 1990) con ayuda de lenguajes de alto nivel.

REGRESIÓN ORTOGONAL

En el ámbito del análisis clínico, en donde se aplica con asiduidad la comparación de métodos, a la regresión ortogonal con errores similares en ambas variables o con cociente de varianzas constante se la denomina regresión de Deming. La metodología de esta regresión ha sido estudiada con detalle en *Clinical Chemistry* e incluso Linnet (1997) y Philippe Marquis (1999) han elaborado sendos programas, CBstat y Method Validator, respectivamente, este último de acceso libre en la red. Este tipo de problemas admite solución exacta.

La solución satisfactoria en el caso general cuando hay errores estadísticos en ambas coordenadas ha sido objeto de una intensa investigación. En este último caso resulta necesaria la aplicación de

procedimientos numéricos iterativos. York propone la solución de una ecuación cúbica como punto de partida, aunque una ecuación lineal es más simple, como demuestra Williamson, en quien se basa recientemente Martin (Clinical Chemistry, 2000) para proponer el Programa de «Regresión general ponderada iterativa de Deming». Lisy y col. proponen asimismo un método elegante de resolución del problema basado en el uso de ecuaciones normales. Rius y col. han desarrollado ampliamente las aplicaciones analíticas.

CONCLUSIÓN

El tema de la regresión reviste una importancia vital, ya que se encuentra estrechamente relacionado con el de la calibración, comparación de dos métodos analíticos, validación de métodos analíticos y estimación de errores. Bajo condiciones ideales, el método de los mínimos cuadrados es el preferido, lo que implica una serie de suposiciones (por ejemplo, normalidad, independencia, abscisa libre de error, y ponderación apropiada). Sin embargo, algunos datos publicados violan de forma grosera al menos una de las suposiciones inherentes al modelo de regresión. Si las desviaciones experimentales pueden adscribirse a la variable dependiente y , la cuestión se simplifica considerablemente. Cuando las medidas se obtienen sobre un rango amplio de valores de x , la suposición de uniformidad en la varianza no resulta válida. Ocurre a veces que algunas de las observaciones utilizadas en el análisis por regresión son más dignas de confianza que otras, por lo que la aplicación directa de los mínimos cuadrados convencionales puede conducir a un serio error. Aunque la suposición de homocedasticidad es válida para algunos procedimientos analíticos, otros tales como medidas de cuentas y análisis fotométrico y cromatográfico bajo ciertas condiciones, no la soportan. Existen dos soluciones al problema de la varianza no constante: llevar a cabo una transformación de los datos, o bien un análisis de regresión lineal ponderada, la mejor solución para algunos investigadores. En general, se ha prestado menos atención a la suposición de normalidad subyacente al método de los mínimos cuadrados, que al tema de la varianza no constante. La replicación de las observaciones está asociada con la desviación estándar de los efectos. Los replicados deben ser genuinos a este respecto. Demasia-

da replicación supone un gran esfuerzo, mientras que poca no nos permite alcanzar la información requerida.

AGRADECIMIENTO

La generosidad de los Excmos. Señores Don Manuel Ortega Mata, Don Vicente Vilas Sánchez y Don Segundo Jiménez Gómez (q.e.p.d.), justifica mi presencia en la Real Academia Nacional de Farmacia. A ellos mi más profundo agradecimiento. Trataré como en la parábola de los talentos de corresponder prestando el mejor servicio a esta honorable Institución. Gracias también al Excmo Señor Presidente de esta Real Academia por sus buenos oficios; él sabe que me tiene como uno de sus declarados admiradores, así como al Excmo Señor Don Benito del Castillo, Decano de la Facultad de Farmacia de la Universidad Complutense de Madrid, espejo en el cual me he mirado y que me ha honrado a lo largo de los años con su amistad y confianza. Decía John Wayne, cuando se le otorgó el oscar cinematográfico: «de haberlo sabido me hubiera puesto en movimiento mucho antes de entrar en el vientre de mi madre». Muchas gracias también a la Doctora Doña Ana Sayago Gómez y al Doctor Don Antonio Gustavo González González por su colaboración y ayuda en la preparación de este discurso.

BIBLIOGRAFÍA

- (1) A. G. ASUERO, A. G. GONZÁLEZ (1989). Some observations of fitting a straight line to data. *Microchem. J.* 40: 216-225.
- (2) A. G. ASUERO; A. SAYAGO, M. BOCCIO. Calibration in chemical and pharmaceutical análisis, comunicación privada.
- (3) K. BAUMANN (1997). Regression and calibration for analytical separation techniques. Part II. Validation, weighted and robust regression. *Process Control Qual.* 10: 75-112.
- (4) A. G. GONZÁLEZ, A. G. ASUERO (1993). Computational program for validating analytical results. *Fresenius Z' Anal. Chem.* 346: 885-887.
- (5) A. G. GONZÁLEZ; M. A. HERRADOR, A. G. ASUERO (1999). Intralaboratory testing of method accuracy from recovery assays. *Talanta* 48: 729-726.
- (6) R. F. MARTIN (2000). General Deming regression for estimating systematic bias and its confidence interval in method-comparison studies. *Clin. Chem.* 46: 100-104.

- (7) J. RIU, F. X. RIUX (1995). Univariate regression models with errors in both axes. *J. Chemometrics* 9: 342-362.
- (8) SAYAGO, A. G. ASUERO (2004). Fitting straight lines with replicated observations by linear regression. II Testing for homogeneity of variances. *CRC Crit. Rev. Anal. Chem.*, 34: 133-146.
- (9) SAYAGO; M. BOCCIO, A. G. ASUERO (2004). Fitting straight lines with replicated observations by linear regression: the least squares postulates. *CRC Crit. Rev. Anal. Chem.* 34: 39-50.