Análisis estadístico de la curva dosis-respuesta en farmacología: conceptos generales

M. Campillo y M.L. Martín-Mateo

Laboratorio de Bioestadística y Epidemiología. Facultad de Medicina. Universidad Autónoma de Barcelona. Bellaterra, Barcelona.

La proliferación y vulgarización de los microordenadores ha supuesto una mayor utilización de modelos de ajuste en todos aquellos supuestos experimentales englobables en el diseño clásico de experiencias; tal es el caso de la curva dosis-respuesta. No obstante, este fenómeno no está exento de graves problemas, entre los que destacan la falta de control de las condiciones de aplicabilidad de los modelos así como el olvido frecuente de su naturaleza y objetivos.

Este hecho conduce a una situación preocupante, ya detectada por nuestro laboratorio [1] a principios de esta década. Dicha situación muestra cómo el aumento del uso de técnicas estadísticas se acompaña de un crecimiento de la utilización inadecuada de las mismas. Por ello planteamos la presente ponencia como un recordatorio de las bases estadísticas en las que debe fundamentarse el análisis de los modelos.

Intencionadamente recordaremos los aspectos más básicos, pues son los que se vulneran con más frecuencia, con el fin de suscitar en lo posible el debate posterior.

La ecuación dosis-respuesta como diseño experimental clásico

Hipótesis de trabajo

Dentro del diseño experimental clásico debemos suponer que el objeto observable de nuestro interés, la variable y, es una variable aleatoria continua. Ello quiere decir que, a pesar de que todos los posibles factores que pueden modificar su valor promedio se encuentren fijos y controlados, la observación repetida de y conduce a $\{y_i\}$, conjunto disperso de valores alrededor de otro central y.

En primer lugar debe recordarse que, en una situación experimental sin errores sistemáticos, dicha dispersión es superior a la precisión de la técnica de medida y en ningún caso debe asimilarse a un error. Así, las distintas respuestas obtenidas para una dosis dada no son erróneas frente a un valor medio determinado, sino que constituyen la fluctuación natural de la variable. A riesgo de incurrir en una reiteración, si y presenta dispersión es porque es una variable. Lo opuesto sería considerar y como una constante medida con imprecisión.

En segundo lugar, el diseño clásico experimental modifica una de las condiciones experimentales, x, manteniendo el resto constante.

Así, para distintos valores x o niveles de x, dosis en este caso, se efectúan diversas observaciones de la respuesta, de la variable y.

Por lo tanto, en el dominio prefijado de x, $\{x_1 \ldots x_k\}$ se obtienen N valores de y_{ij} con el objetivo final de observar la posible relación entre ambos conjuntos.

En general, dentro de este contexto experimental la hipótesis de trabajo es la relación existente entre el efecto medio observado y la variación no aleatoria de la dosis.

Modelo

El modelo más básico, punto de partida de elaboraciones posteriores, consiste en analizar la variación de los y_{ij} en diversos componentes:

$$y_{ij} = \overline{Y} + F(x_i) + \epsilon_{ij}$$

es decir, la variabilidad natural del observable medida respecto a su correspondiente media. El efecto medio producido al modificar x respecto al valor promedio obtenido y este último $\overline{Y}$ como elemento de referencia.

Hipótesis estadística

Una vez establecida la hipótesis de trabajo, así como el modelo que se pretende ajustar, debe

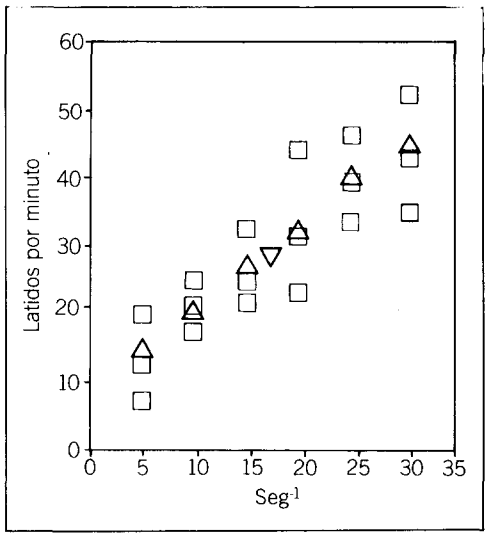

Fig. 1. Ritmo cardiaco versus *inotropía (+).*

seleccionarse el conjunto de hipótesis estadísticas adecuadas que nos permitan evaluar el riesgo de que dicho modelo no sea debido al azar o diferente a otro posible.

En el caso también más simple de modelo lineal, los tests estadísticos se engloban dentro del contexto del análisis de la varianza y de la regresión lineal a la media, característica esta última que se olvida muy frecuentemente.

Descomposición de la dispersión global observada

En la figura 1 se representa gráficamente un ejemplo de lo expuesto hasta el momento utilizando los datos de la tabla I. En la misma se observa cómo el observable y posee una varianza clara para cada situación experimental, definida por el valor adjudicado a x. En ningún caso, por la propia naturaleza del diseño, puede considerarse que x posea varianza, ya que no se trata de una variable aleatoria sino de un factor fijo predeterminado por el experimentador.

En la figura 2 se muestra la descomposición de los términos de dispersión. Así, en la figura 2a se describe la variabilidad natural del observable. Dicha dispersión, posee N—K grados de libertad, ya que se encuentra referida a los k valores y_i obtenidos en cada situación del rango $\{x_1 \ldots x_k\}$.

En la figura 2b se describe la variación de los R efectos medios respecto al valor promedio global Y : $\Sigma_{ni} (\bar{y}_i - \bar{y})^2$. Por esta razón esta dispersión posee k-1 grados de libertad.

La hipótesis estadística adecuada en este estadio puede plantearse como: ¿Hasta qué punto la varianza de los efectos medios $\bar{y}_i$ es superior

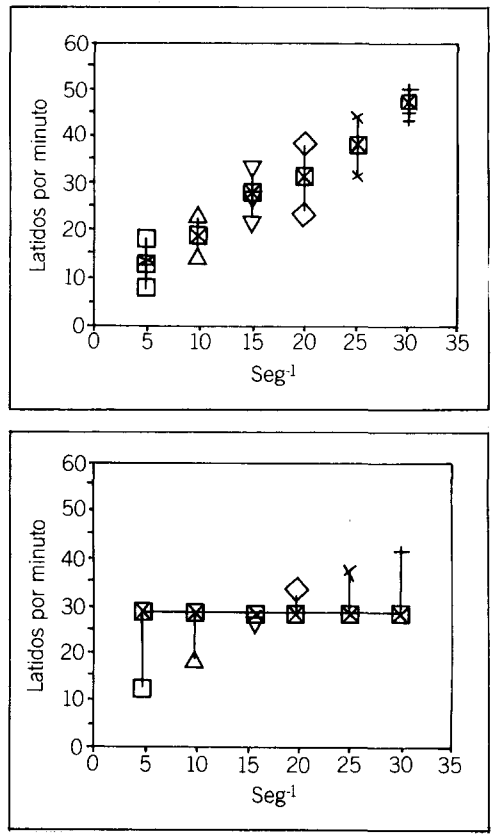

Fig. 2. a) *Dispersión intragrupos;* b) *dispersión entre las medias.*

TABLA I
RITMO CARDÍACO* EN FUNCIÓN DE INOTROPÍA POSITIVA MEDIDO EN GATOS ANESTESIADOS

Inotropía (+) (seg^{-1})					
5	10	15	20	25	30
12	20	26	31	37	42
18	22	31	38	44	52
9	17	24	25	32	33

*Latidos/minuto.

a la varianza propia de la variable y_{ij} en cada condición experimental x_i?

Este análisis de comparación de varianzas nos permite determinar, con un índice de riesgo preestablecido, si en el rango de dosis estudiado, $\{x_1 ... x_k\}$, se producen significativamente respuestas medias, $\bar{y}_i$ diferentes.

En el ejemplo descrito en la tabla I la varianza entre los efectos medios es 10,4 veces superior a la varianza promedio de cada situación experimental.

Una vez pasado satisfactoriamente este test deberían analizarse cuáles son los efectos medios observados diferentes realmente mediante el test *a posteriori* más adecuado: Tukey, Scheffe, etc. Después, debe recordarse el objetivo del modelo: explicitar F (x_i), es decir, analizar los efectos medios y su posible relación con x.

Reiteramos esta idea dada la frecuencia con que se considera toda la dispersión en la consecución de un modelo. Es evidente que los paquetes matemáticos y estadísticos no poseen criterio y ajustan en lo posible la dispersión que se les introduce. No obstante, conceptualmente es como mínimo sorprendente que se intente evaluar un modelo ajustando una dispersión aleatoria que se produce en condiciones experimentales constantes.

Explicitación del modelo F (x_i): regresión lineal a la media

Existen dos maneras de abordar el problema de la explicitación del modelo; ambas poseen diferentes ventajas e inconvenientes.

Presunción a priori

La respuesta previa del modelo que se pretende ajustar implica un conocimiento fisiológico y farmacológico del problema que se estudia. Por esta razón, puede prefijarse el riesgo α que se asume en su elaboración así como el del conjunto de hipótesis estadísticas apropiadas para juzgar el cumplimiento de los objetivos que persigue el modelo: explicación de la varianza entre medias de y, y su utilidad en predicciones posteriores.

Dichas condiciones no deben transgredirse *a priori* aunque no dejan de ser un esquema de referencia a la hora de establecer conclusiones.

Normalidad de la variable estudiada. Esta condición se traduce en que, en condiciones constantes, la distribución de valores se corresponde a una función gausiana.

Teniendo en cuenta este hecho, la valoración de los posibles *outliers* cambia de perspectiva. Así, existe un 5 % de probabilidades de que en el transcurso de una experiencia se presente un valor alejado de valor medio en más de 2 desviaciones estándar. ¿Es un valor anómalo? No, sencillamente el 5 % de los datos cumplen la condición de *outliers* si presuponemos la normalidad.

Además, su peso en la media está muy ponderado por el resto de las repeticiones. Evidentemente si introducimos los datos individuales en el modelo poseerá una influencia no despreciable. Sin embargo, sólo aporta información a la dispersión natural y parcialmente al promedio.

¿Por qué no repetir experiencias en dicho nivel en vez de eliminar un dato que debe aparecer en una de cada 20 experiencias?

Homocedasticidad. El modelo de regresión lineal, basado en el esquema ANOVA, se basa en que dentro de cada dosis, x_i, en cada situación experimental, la varianza del observable se mantiene constante.

Esta condición se rebate cotidianamente por la experiencia. En general, la varianza es función del valor de x fijado por el experimentador.

Afortunadamente, este problema sólo repercute en dos aspectos:

1. Cuando no puede rechazarse la hipótesis estadística nula porque puede deberse a la heterocedasticidad. (La diferente dispersión puede enmascarar las verdaderas diferencias, sobre todo porque el número de repeticiones no supera en general a $n = 4$.)

2. Cuando mediante una transformación de la variable se logra dicha homocedasticidad.

Es frecuente buscar la transformación que elimine este problema. Así, comúnmente se utiliza el logaritmo natural o decimal de la variable o el arcsen cuando el observable constituye un tanto por ciento del máximo de respuesta disponible.

Cualquier tipo de transformación se justifica en sí misma si con ella se logra entrar dentro de las condiciones de aplicabilidad. No obstante, no debe olvidarse que la regresión, en este caso, se efectúa sobre la variable transformada y no sobre el original.

Por ejemplo, una transformación logarítmica de y incide sobre el hecho que lo que realmente se analiza son medias geométricas y no aritméticas.

Otra cuestión es linealizar la función, proceso que en la actualidad no tiene demasiado sentido debido a la disponibilidad de una gran variedad de algoritmos de ajuste no lineal.

Contraste de hipótesis nulas

Sea cual sea el modelo analizado debe efectuarse un contraste de hipótesis estadístico adecuado que nos permita discernir cada uno de los tres objetivos parciales definidos a la hora de generar dicho modelo:

Test de ajuste. En resumidas cuentas, debe analizarse si la varianza objeto de estudio, que podemos descomponer en la que explica el modelo y la residual, se encuentra suficientemente descrita.

En otras palabras, si la varianza explicada supera con creces a la residual.

Test de significación. Si dicho cociente no supera a lo previsible por el azar.

Ambos tests se encuentran muy interrelacionados pero no constituyen, por separado, condición necesaria y suficiente de la bondad de un modelo.

Test de utilidad. La elaboración de un modelo no constituye únicamente un esquema descriptivo de la experiencia. En la mayoría de los casos lo que se pretende es su posterior utilización con el fin de precedir efectos medios para dosis no consideradas.

Debemos destacar que en este contexto, los modelos sólo pueden predecir efectos medios y no casos individuales. (Hemos renunciado a explicar la variabilidad natural.)

Adicionalmente, y no de manera gratuita, debe recalcarse que el modelo sólo es útil dentro del dominio de valores de x explorado. Es decir, la interpolación conlleva un error valorable; la extrapolación no tiene sentido.

Comparación de modelos

La comparación de modelos debe efectuarse a partir de dos aspectos que en principio no son contrapuestos.

En primer lugar debe analizarse la bondad de ajuste, y en segundo lugar su utilidad.

En la mayoría de programas de ajuste se valora exclusivamente la bondad del mismo en términos de P, probabilidad *a posteriori* que no considera en ningún caso el número de modelos contrastados simultáneamente.

Presunción a posteriori

Las facilidades que supone la búsqueda automática de un modelo presenta la ventaja de obtener el mejor ajuste numérico a la dispersión administrada. Sin embargo, debe imperar la cautela con el fin de evitar dos grandes problemas: *a)* la obtención de un modelo sin significado farmacológico claro, y *b)* el aumento del riesgo de obtener un modelo fruto del azar.

En el caso más simple, es decir en aquél en que los modelos analizados mediante una técnica automática fuesen probabilísticamente independientes, el riesgo total, (α'), de rechazar la hipótesis nula: no sigue un modelo sino que se adapta a la función $\alpha' = (1 - [1 - \alpha]^n)$ siendo α el valor habitual de p = 0,05 y n el número de modelos analizados.

Por ejemplo, el ajuste de diez modelos independientes cada uno de los cuales se contrasta estadísticamente con una hipótesis rechazable al 5 %, posee un riesgo total del 40,3 % de que al azar pueda rechazarse la hipótesis nula al mentos para un modelo.

Modelo lineal: $\mu y_i = \beta x_i + \alpha$

Dentro de la innumerable gama de posibles modelos, el modelo lineal es el de concepción más simple debido fundamentalmente a la facilidad de interpretación.

Generalmente, se abusa más de un modelo de este tipo de lo que se usa. La razón es sencilla y puede resumirse en el hecho de que, a menudo, sus condiciones de aplicabilidad se olvidan o se obvian sin análisis previo.

Globalmente el conjunto de condiciones que permiten contrastar la hipótesis nula de carencia de lineabilidad es el mismo que configuran cualquier análisis de tipo ANOVA.

Así mismo, no se dispone de ningún programa que efectúe el análisis de utilidad adecuado, si bien existe un gran número de textos en los que se detallan los algoritmos y tests adecuados para efectuar la comparación entre modelos [3-6]. Por este motivo expondremos a continuación los fundamentos conceptuales de los mismos.

Paralelismo

El estudio del paralelismo se traduce en la práctica en la comparación de las pendientes de los modelos lineales ajustados. El test adecuado se basa en la hipótesis nula según la cual ambas pendientes son valores muestrales estimadores de una misma poblacional.

Por lo tanto el test posee una estructura formal similar a la comparación de dos medias muestrales de varianzas desconocidas. Es decir, la diferencia de pendientes se distribuye se-

gún una función t con una varianza igual al promedio de las varianzas estimadas de cada pendiente.

En el caso de no poder rechazar la igualdad pueden substituirse ambos valores b_1 y b_2 por su valor promedio, ponderando cada pendiente por el rango de valores x para el que se ha obtenido.

Intercepción

El mismo esquema de análisis debe aplicarse a la comparación de pendientes así como al contraste de que la pendiente obtenida no se diferencia del valor cero.

También en este caso la hipótesis nula supone que las pendientes obtenidas no son más que valores muestrales de una población que posee media poblacional igual a la y, cuya desviación típica se estima mediante el promedio de las desviaciones de las ordenadas en el origen ajustadas.

Igualmente, si no puede rechazarse la hipótesis nula de igualdad entre pendientes es preferible utilizar el promedio ponderado de las mismas como mejor estimador de la pendiente poblacional.

Unicidad

Una situación que se presenta con cierta frecuencia consiste en la obtención de modelos lineales en condiciones experimentales ligeramente distintas, por ejemplo, como caso simple, en dos rangos de x diferentes. La hipótesis nula a contrastar es en este caso si los modelos obtenidos no constituyen en realidad más que aspectos parciales de un único modelo.

De manera simple, la comparación debe efectuarse entre el ajuste obtenido para el rango total de x y el promedio del que se obtiene para los modelos parciales. En los textos referenciados se encuentran descritos los test así como los algoritmos precisos.

También debe apuntarse la idea de que en el caso de no poder rechazar la hipótesis de unicidad es preferible la utilización del modelo único. Entre otras ventajas que supone este esquema está el aumentar considerablemente el número de grados de libertad.

Conclusiones

Como conclusiones generales sólo queremos hacer notar la necesidad que tiene todo experimentalista de prefijar con claridad cuál es el fenómeno que quiere explicar con un modelo, observar si se cumplen las condiciones de aplicabilidad del mismo y evitar en lo posible la determinación del modelo basándose como único criterio en el mejor ajuste numérico y en el valor de P *a posteriori*.

BIBLIOGRAFÍA

1. Martín M, Sanz F. Efecto de la introducción de la Estadística en el *curriculum* de Medicina: análisis de una década, 1970-1980. Medicina Clínica 1982; 79 (6): 273-276.
2. Kenakin TP, Johnson SF. Eur J Pharmacol 1985; 3: 347-354.
3. Kenakin TP. Pharmacologic analysis of drug-receptor interaction. Nueva York, Raven Press, 1987.
4. Kleinbaum D, Kupper L. Applied regression analysis and other multivariate methods. Boston, Duxbury, 1978.
5. Berenson ML, Levine DM, Goldstein M. Intermediate statistical methods ans applications. New Jersey, Prentice-Hall Inc, 1983.
6. Afflifi AA, Clark V. Computer aided multivariate analysis. California Lifetime Learning Pub, 1984.

DISCUSIÓN

R. OBACH: Quisiera continuar con el tema de los *outliers*. Cuanto se demuestra de forma repetitiva que efectivamente un punto se separa sistemáticamente de un presupuesto modelo y ello no es atribuible a problemas metodológicos o al menos explicables, ¿cuál es la solución más factible? Particularmente, la supresión de un punto no me parece razonable en cuanto que hay una constatación experimental de que se va repitiendo. En todo caso deberían probarse otros modelos que justificasen o anulasen esta separación. Otra cuestión es la postura a adoptar y hasta qué punto un solo *outlier* ante una nube de puntos numerosa puede o no desvirtuar un modelo.

M.L. MARTÍN-MATEO: La importancia del punto que se separa de la media, que no pretendo explicar, es posiblemente muy pequeña. Lo que ocurre, en general, y éste es el riesgo de la utilización de ordenadores, es que uno in-

troduce todos los datos uno a uno y las máquinas no interpretan en absoluto cuál es la voluntad de análisis por lo que analizan lo que se les introduzca cuando en realidad lo que el investigador pretende explicar generalmente son efectos medios. La solución que a veces se propone consiste en hacer más réplicas de la dosis en cuestión para que no pese tanto el *outlier*, pero no eliminarlo.

J.M. PEINADO: En este contexto yo comparto la opinión del Dr. García Sevilla en el sentido de que muchas veces lo que importa es el sentido común ante un dato que se aparta y que además muchas veces no es posible repetir, por ejemplo porque no sea factible obtener nuevas muestras para la analítica. Cuando un dato de este tipo no tiene ninguna explicación experimental creo que el sentido común obliga muchas veces a prescindir del mismo sin que por ello se alteren los resultados. Quisiera además hacer un comentario sobre el número de datos de un experimento. Yo entiendo que cuanto mayor es este número menor es el efecto del azar.

M.L. MARTÍN-MATEO: La *n* a la que he hecho mención en mi presentación no se refería al número de datos sino al número posible de modelos que se pueden probar. No hay que olvidar que estadístisticamente cualquier modelo comporta un riesgo de obtener un resultado significativo simplemente por azar y cuantos más modelos aplicamos mayor es este riesgo. Esta es la situación cuando no se define *a priori* el número de parámetros.

E. MORCILLO: Quisiera hacer una pregunta, ¿cuál es el argumento definitivo para utilizar un análisis matemático en lugar de un análisis estadístico de los datos para la construcción de una curva dosis-respuesta? Si se hace una revisión de las comunicaciones en una revista científica de farmacología de buena calidad, se puede advertir que la mayoría de los investigadores optan en general por presentar sus datos de curvas concentración-respuesta simplemente en unos clásicos ejes de coordenadas y yo diría que en un 90 % de los casos se resisten a aplicar un modelo «matemático» al cual se ajusten esos datos.

M.L. MARTÍN-MATEO: No tengo dudas sobre la diferencia entre análisis estadístico y análisis matemático, el problema es que en general ambos términos se confunden. La valoración de un modelo matemático, aunque quizá sería mejor hablar de modelo numérico, está en la función de que no sea obtenido por azar. En sentido estricto, lo que se establece en un estudio experimental es un contraste de hipótesis que permite valorar únicamente el riesgo de la afirmación que se hace. Normalmente se confunde la determinación de esa probabilidad con la determinación de un modelo por la técnica matemática o numérica. El mal uso de la estadística a que yo me refería se produce cuando se aplican indiscriminadamente multitud de pruebas hasta obtener la significación deseada o la confusión existente entre correlación y regresión.

J.L. GONZÁLEZ: Estoy completamente de acuerdo con el Dr. Martín Mateo sobre el riesgo que entraña trabajar con paquetes estadísticos que aplican multitud de pruebas sin que muchas veces exista una justificación racional para ello. Quisiera además plantear la cuestión de cuál sería la actitud a adoptar cuando dos modelos proporcionan parámetros estadísticos coincidentes, ¿cuál debería ser la actitud del estadístico ante una situación de este tipo?

M.L. MARTÍN-MATEO: En realidad, el que elabora la parte estadística de un problema no toma posteriormente decisiones sobre el resultado obtenido. Es el investigador que ha planteado la experiencia quien debe interpretarlo. Lo que es realmente importante es que el investigador y el estadístico se pongan de acuerdo *a priori* sobre lo que pretende realmente el experimento.

A. GARCÍA: Hay muchos tipos de respuestas fisiológicas que los farmacólogos valoramos pero fundamentalmente podrían agruparse en dos: unas que son todo o nada y que denominamos cuantales y otras que son graduales y que podemos monitorizar más finamente. Me pregunto si el análisis estadístico de ambos tipos de respuesta en curvas dosis-respuesta es extrapolable o por el contrario presenta diferencias apreciables.

M.L. MARTÍN-MATEO: Efectivamente existen diferencias. En el caso de respuestas cuantitativas cuantales se recurre a una transformación consistente en la valoración en tanto por ciento de las veces que se ha presentado. Por tanto, trabajando no con la variable sino con el porcentaje de veces que ésta ocurre dentro de una muestra. El problema que existe es que al repetir una variable dicotómica dentro de una muestra pequeña, jamás se obtiene una distribución normal en sentido purista sino que la distribución es binomial. Lo que ocurre es que afortunadamente se puede pasar de una a otra con unos riesgos de error pequeños utilizando transformaciones matemáticas.