

Estadística en la investigación en medicina

V. Barrio y R. Madero *

Servicio de Nefrología. Hospital Regional Reina Sofía. Córdoba.

* Sección de Bioestadística. Hospital La Paz. Madrid.

Introducción

Los métodos estadísticos, en su sentido más amplio, constituyen un instrumento de trabajo cada vez más importante en diversos campos experimentales. Un conocimiento de la metodología estadística proporciona al investigador una herramienta que le permite analizar adecuadamente los datos sobre los que trabaja y adoptar de forma racional las decisiones oportunas¹. No existe una metodología estadística particular de la economía, de la educación o de la medicina, sino una teoría general aplicable a cualquier campo de estudio en el que se realicen observaciones aunque hay procedimientos más utilizados en ciertas áreas.

De forma general, el primer paso en cualquier estudio, después del diseño del mismo, consiste en la recogida y descripción de los datos. Este análisis preliminar y experiencias previas permiten enunciar una hipótesis que contrastaremos para confirmarla o rechazarla. En el caso de que se confirme podremos enunciar una ley que nos servirá para predecir resultados futuros en circunstancias similares.

La estadística interviene en cada uno de los pasos descritos:

- En la búsqueda de las observaciones (teoría de muestreo).
- En la presentación y resumen de los datos (estadística descriptiva).
- En la estimación de parámetros (teoría de estimación).
- En la formulación y verificación de hipótesis (test de hipótesis).

Siguiendo este esquema vamos a desarrollar brevemente los conceptos fundamentales de la metodología estadística, resaltando a continuación los test que se deben aplicar en diferentes circunstancias y los programas estadísticos disponibles, que a nuestro juicio son los más completos actualmente.

Teoría de muestreo

Una muestra es una parte representativa de la población objeto de estudio. Gracias al método estadístico, del estudio de una muestra, podemos inferir las características de la población considerada.

El método utilizado para escoger una muestra es un factor determinante del uso que puede hacerse de dicha muestra². La condición más importante que debe cumplirse es que sea elegida al azar, es decir, que cualquier individuo de la población tenga la misma probabilidad de formar parte de la misma. Estas muestras se denominan aleatorias y están libres de un error llamado sesgo que se podría cometer con un muestreo no probabilístico o con un muestreo sistemático si alguna de las características de interés siguiera un patrón cíclico. En medicina este tipo de muestras es francamente difícil de conseguir, por lo que si un investigador intenta hacer inferencias desde su muestra debe tomar precauciones para no incurrir en errores y falacias. Por ejemplo, se puede construir artificialmente una asociación espúrea, o bien se puede enmascarar alguna asociación que realmente exista si las probabilidades de admisión a un hospital son diferentes para aquellos que padecen o no una determinada enfermedad y presentan la característica de interés³. La mejor manera de evitar el sesgo en el proceso de muestreo es comparar la muestra con datos que procedan de fuentes externas. Existen otros tipos de muestreo, como el muestreo sistemático o el muestreo estratificado, más complejos y utilizados, sobre todo en medicina comunitaria y en estudios sociológicos a gran escala, que requieren ajustes ponderados a la hora de realizar el análisis⁴. Otro aspecto importante en cuanto al proceso de muestreo es el tamaño que como mínimo debe tener la muestra para que ésta sirva para demostrar la hipótesis considerada. Dependiendo del diseño del trabajo y de las características de las observaciones, es necesario dicho tamaño.

Estadística descriptiva

Se puede considerar la estadística descriptiva como el conjunto de técnicas que facilitan la organización, resumen y comunicación de los datos. Utiliza

Correspondencia: Dr. V. Barrio.
Servicio de Nefrología.
Hospital Regional Reina Sofía.
Avda. Menéndez Pidal, s/n.
14004 Córdoba.

el número como soporte objetivo. Desde el punto de vista estadístico, una población es un conjunto de individuos u objetos con una característica común observable. Cada individuo de la población puede describirse según una o varias características, como, por ejemplo, edad, sexo, peso, etc. Estas variables se pueden clasificar de acuerdo con su distribución y sus niveles de medida, que, como se describe más adelante, son determinantes a la hora de elegir el método de análisis de los datos. El primer criterio de clasificación trata de determinar si existen intervalos entre las posibles medidas de una variable discreta, y en caso contrario, es decir, si cualquier resultado es observable, ante una variable continua. El segundo criterio de clasificación se refiere al grado de precisión matemática de la medida de la variable. El nivel más bajo es el nominal, utilizado simplemente para asignar diferentes categorías. La variable sexo, por ejemplo, es nominal y dicotómica, ya que sólo puede tomar dos posibles valores. Un nivel de medida algo más alto es aquel que permite no sólo agrupar en diferentes categorías, sino asignar un orden a las mismas, por lo que se denomina ordinal. Por último, el nivel más alto, llamado intervalo, se otorgará a aquellas variables ordenadas cuya distancia entre las diferentes categorías sea medible. Para ser considerada intervalo, una variable debe tener unas unidades físicas de medida o algún estándar bien establecido. El peso, la estatura, la tensión arterial, o el número de fallecimientos reúnen estas condiciones y se consideran variables intervalo. La figura 1 muestra las posibles situaciones resultantes al combinar estas clasificaciones. Así, observamos que toda variable nominal es discreta, aunque una variable discreta puede ser nominal, ordinal o intervalo. Igualmente, una variable continua puede ser ordinal o intervalo; sin embargo, hay variables ordinales o intervalo que no

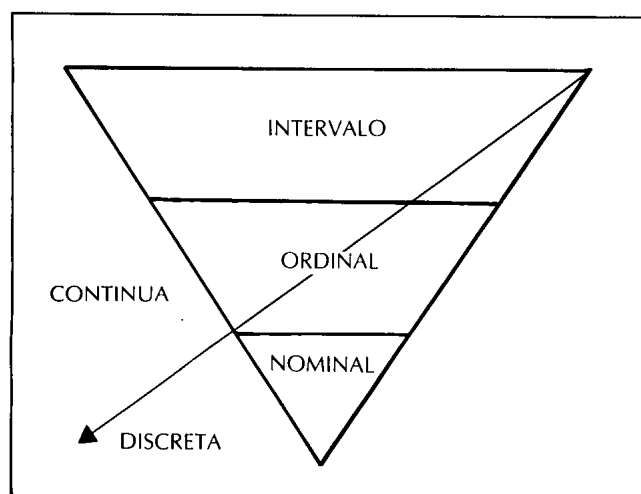


Fig. 1.—Clasificación de variables según su distribución y su nivel de medida.

son continuas. Por ejemplo, el sexo es una variable nominal y discreta; la edad puede ser considerada intervalo y continua, aunque si se agrupa en intervalos puede ser nominal y discreta; la clase social, dependiendo de cómo se mida y del interés del investigador, puede ser considerada como ordinal y continua, ordinal y discreta o normal y discreta.

Las variables de tipo cualitativo o variables discretas se representan a través de las proporciones o porcentajes de individuos en cada categoría. Se utilizan dos tipos de parámetros para representar una variable cuantitativa o continua, parámetros de centralización referidos a la localización del centro de la distribución y parámetros de dispersión que dan una idea de cómo se aproximan o se alejan los datos del centro. Los parámetros de centralización que se utilizan con más frecuencia son la media aritmética, la mediana y la moda, mientras que para la dispersión los más usados son la desviación típica y el error estándar de la media, aunque existen otros que dan una idea de qué posición ocupa un determinado dato en la distribución, como son los cuartiles, deciles y percentiles. Hay que resaltar que la desviación típica, también llamada desviación estándar, y el error estándar están relacionados, de forma que para una muestra dada el error estándar se calcula directamente dividiendo el valor de la desviación típica por la raíz cuadrada del tamaño de la muestra. Esto hace que el error estándar aparezca como una medida de dispersión más pequeña y sea utilizada con frecuencia en las publicaciones, a pesar del error conceptual que supone, ya que no cuantifica la variabilidad que existe entre las diferentes observaciones como la desviación estándar, sino que mide la precisión con la que la media de la muestra estima la media real de la población⁵.

La mayor parte de los fenómenos biológicos en los que la variable de interés es cuantitativa quedan perfectamente representados a través de la media y la desviación estándar, aunque cada vez se observa con mayor frecuencia en las publicaciones la utilización de la mediana y del rango intercuartil en el caso de que las variables objeto de estudio no sigan una distribución normal.

Para que la estadística sea eficaz tiene necesariamente que simplificar, y las unidades sobre las que trabaja deben pertenecer a un conjunto homogéneo y bien delimitado. La estadística no se ocupa de forma general de los casos raros.

Teoría de la estimación

Inferencia estadística es el nombre con el que se conocen una serie de técnicas que se proponen extender conclusiones a una población a partir de los resultados obtenidos en una muestra. La estimación es el proceso por el cual se calcula de forma aproxima-

mada el valor del parámetro a estudiar por el valor obtenido en la muestra. A este valor del parámetro en la muestra se le llama estadístico. Los parámetros más usados son la media (μ) y la desviación típica (σ), que se representan en la muestra como \bar{x} y s , respectivamente.

Otra forma de estimación que cada vez se utiliza con mayor frecuencia^{6, 7} son los llamados intervalos de confianza. Un estadístico calculado en una muestra nunca coincidirá exactamente con el valor real del parámetro que se pretende estimar, ya que el propio muestreo introduce cierto grado de error. Es importante, por tanto, que un estimador vaya acompañado de la precisión de la estimación. Los intervalos de confianza proporcionan un método para establecer cuánto se acerca el valor del estadístico al parámetro y la casualidad de este acercamiento. Los resultados de un estudio también pueden, y en muchas ocasiones deben expresarse como intervalos de confianza, ya que no sólo tiene interés el valor de significación del test de hipótesis, sino que muchas veces el dato de mayor relevancia de un trabajo es la magnitud de la diferencia entre los grupos. Esta manera de presentar los resultados es aplicable a la mayoría de los test utilizados habitualmente, como son comparación de medias, proporciones o de sus diferencias respectivas⁶, análisis de supervivencia^{7, 8}, regresión y correlación simple⁹ e incluso análisis no paramétricos¹⁰.

Test de hipótesis

El objeto de un test de significación estadística consiste en determinar el grado de consistencia entre una hipótesis que describe algún aspecto de una población particular y una serie de datos obtenidos de una muestra. Aunque la muestra sea representativa de la población estudiada, es evidente que una muestra diferente proporcionará una serie de datos diferentes. Por lo tanto, siempre existirán dos explicaciones obvias para los resultados obtenidos. En la primera de ellas, la hipótesis es cierta y la incertidumbre del muestreo o el azar son las razones del resultado obtenido. En el segundo caso, la hipótesis es falsa y la diferencia entre la situación real y la hipótesis considerada es la causa del resultado obtenido.

La medida objetiva que se utiliza para evaluar el grado de consistencia entre los datos recogidos y la hipótesis es un cálculo de probabilidad de la incertidumbre de muestreo. Suponiendo que la hipótesis sea cierta, se calcula la probabilidad de que el resultado obtenido con la muestra sea más o menos probable. Un test de significación asume que una hipótesis nula (H_0) sobre la población objeto de estudio; obtener datos sobre la población por medio de una

muestra aleatoria; presentar la serie de posibles resultados, de los cuales el observado en la muestra representa una única posibilidad; representar la distribución de probabilidad del test estadístico basado en las asunciones de que la hipótesis nula es verdadera y la incertidumbre de la muestra aleatoria; ordenar los posibles resultados de acuerdo con su consistencia con la hipótesis nula y, por último, calcular la probabilidad de que la incertidumbre del muestreo pudiera producir un resultado no más de acuerdo con la hipótesis nula que el resultado realmente observado. Esta última probabilidad se denomina nivel de significación de los datos con respecto a la hipótesis nula (H_0) o valor de p . Por lo tanto, el valor de p representa la probabilidad de obtener un valor del test estadístico tan extremo o aún mayor que el obtenido con los datos si la diferencia entre los grupos se debiera únicamente a la variabilidad de la medida o de la respuesta de los sujetos¹¹.

Sin embargo, como la situación real con respecto a la hipótesis nula en la población objeto de estudio nunca se podrá conocer con certeza, siempre existen posibilidades de error en las conclusiones que alcanza un investigador. Estos errores, denominados tipo I y II, o falsos positivos y falsos negativos, respectivamente, y sus probabilidades asociadas se expresan en la tabla I. Si no existe realmente diferencia alguna entre los grupos, entonces el valor de α , el nivel de significación de los datos con respecto a la hipótesis nula fijado de antemano habitualmente en 0,05, es simplemente la probabilidad de obtener un resultado improbable en esta situación y concluir, por tanto, que los datos contradicen la hipótesis nula. Si, al contrario, existe realmente una diferencia entre los grupos, la probabilidad de que el investigador alcance una conclusión correcta es $1-\beta$. Esta probabilidad se conoce con el nombre de potencia del test y depende de la magnitud de la diferencia entre los grupos, pero, sobre todo, del tamaño total de la muestra, ya que cuanto mayor esta muestra, mayor información proporcionará sobre la población, y la estimación de la situación real será más precisa. Consecuentemente, al aumentar el tamaño de la muestra disminuye la probabilidad de obtener falsos negativos (β) si existe una diferencia real entre los grupos. Como α y β representan las probabilidades de obtener un resultado erróneo, en el mejor de los casos nos gustaría que fueran lo más pequeñas posibles. Sin embargo, si disminuimos α sin modificar el tamaño de la muestra, el valor de β aumenta. Lo mismo ocurrirá con el valor de α si intentamos disminuir la probabilidad de falsos negativos sin modificar el tamaño de la muestra.

La discusión precedente sienta las bases del cálculo del tamaño de la muestra (n) necesario para poder detectar con una alta probabilidad ($1-\beta$) una diferencia de cierta magnitud (δ) entre dos grupos si la pro-

Tabla I. Posibles resultados de un test de significación estadística. Errores de tipo I y II y probabilidades asociadas

Situación real	Conclusión del investigador	
	H ₀ verdadero	H ₀ falso
H ₀ verdadero	Correcto (1- α)	Falso positivo (α) Error tipo I
H ₀ falso	Falso negativo (β) Error tipo II	Correcto (1- β)

(*) Probabilidades asociadas a cada resultado.

abilidad de obtener un resultado falso positivo está fijada de antemano (α). No existe una respuesta única a este planteamiento, siendo habitual el presentar una tabla o gráfico con los posibles valores de n en función de α y de la potencia del test (1- β). Otras consideraciones aparte de las reseñadas, como el tamaño relativo de cada grupo, los posibles abandonos del tratamiento o la falta de «compliance» del mismo, son también determinantes y, por tanto, han de ser tenidas en cuenta a la hora de realizar los cálculos^{12, 13}.

Esquema de test a utilizar según el tipo de variables objeto de estudio

Como la mayor dificultad para cualquier investigador parece residir en la elección del test que debe utilizar, vamos a indicar brevemente las condiciones que se deben cumplir para la correcta aplicación de los análisis más empleados. En realidad, la mayor dificultad del desarrollo de un proyecto de investigación reside en la fase de diseño, ya que en este momento es cuando se debe decidir qué hipótesis se va a investigar, qué variables se van a considerar y cuántos sujetos se van a estudiar, lo que determinará el tipo de análisis a utilizar y la potencia del estudio.

De acuerdo con el tipo de variable que utilicemos (categórica, ordinal o continua) y su condición de variable dependiente o independiente, la tabla II muestra los tests a utilizar. Se deben resaltar algunas puntualizaciones con respecto a esta tabla, aunque recomendamos la consulta de algunos textos clásicos de bioestadística¹⁴⁻¹⁶ y de una serie de artículos¹⁷⁻²² recién publicada para mayores detalles.

La inclusión de datos pareados a la hora de realizar el estudio obliga a la utilización de test específicos en el momento del análisis. Así, para tablas de contingencia con datos pareados será preciso recurrir al test de McNemar en lugar del test de chi cuadrado; para variables continuas, la T de Student se reemplazará por la T de Student para datos pareados,

Tabla II. Relación entre el tipo de variable y el test de significación que se debe utilizar en el análisis univariado

Primera variable (*)	Segunda variable (*)	Test de significación
Dicotómica:		
Pareado	Dicotómica	Test de McNemar
No pareado	Dicotómica	Chi cuadrado, test de Fisher
Dicotómica	Nominal	Chi cuadrado
Dicotómica:		
Pareado	Ordinal	Test de Wilcoxon
No pareado	Ordinal	U de Mann-Whitney
Dicotómica:		
Pareado	Continua	T de Student para datos pareados
No pareado	Continua	T de Student
Nominal	Nominal	Chi cuadrado
Nominal:		
Pareado	Ordinal	Test de Friedman
No pareado	Ordinal	Test de Kruskal-Wallis
Nominal	Continua	ANOVA de un factor
Ordinal	Ordinal	Rho de Spearman Tau de Kendall
Ordinal	Continua	Categorizar variable continua y utilizar rho, tau o chi cuadrado
Continua	Continua	Correlación (r) Regresión lineal

(*) Por simplicidad, y para no duplicar las entradas, no se especifican las variables dependientes e independientes en esta tabla. La mayoría de los test reseñados son correctos, considerando como variable dependiente tanto la primera como la segunda variable. Sin embargo, para la regresión lineal es importante especificar las variables dependientes e independientes.

y en el caso de variables ordinales o de utilizar tests no paramétricos, la U de Mann-Whitney será sustituida por el test de Wilcoxon.

El test exacto de Fisher se utilizará en lugar de chi cuadrado cuando la frecuencia esperada en cualquier celda sea inferior a 5 o el número total de sujetos sea inferior a 50¹⁵. La utilización de la T de Student para la comparación de dos medias asume una distribución normal de las variables e igualdad de varianzas. En el caso de no cumplirse estas condiciones se pueden transformar los datos —habitualmente logaritmos si la distribución está sesgada a la derecha o si la desviación estándar es proporcional a la media— o recurrir a test no paramétricos que utilizan rangos y no asumen ninguna distribución particular de las variables^{23, 24}. Estos tests son también los recomendados para el análisis de variables ordinales^{25, 26}.

El análisis de la varianza (ANOVA) constituye la extensión de la T de Student cuando se comparan tres o más grupos. El utilizar este último test cuando se comparan más de dos medias es un error, ya que no se aprovecha toda la información sobre la variabilidad entre los grupos y la probabilidad de cometer un error de tipo I aumenta de forma importante. Para evitar estos problemas se debe realizar un test global

basado en el análisis de varianza para detectar si existen diferencias significativas entre los grupos y a continuación, y únicamente si el test global previo (F) es significativo, utilizar tests de Student modificados para identificar dónde residen estas diferencias. El análisis de la varianza, al igual que la T de Student, asume que los datos son independientes, que siguen una distribución normal y que la varianza de cada grupo es similar. De nuevo podremos transformar los datos para que el análisis sea válido o recurrir a tests no paramétricos que obvian estas dificultades si estas condiciones no se cumplen. Para las comparaciones múltiples se emplearán uno de los procedimientos descritos que corrigen el aumento de la probabilidad de falsos positivos, como son los métodos de Bonferroni, Duncan, Tukey o Scheffé, entre otros²⁷.

Es frecuente encontrar en estudios biológicos diseños en los cuales se siguen grupos experimentales en diferentes momentos evolutivos, siendo interesante contrastar diferencias a lo largo del tiempo. Este tipo de diseño ha de tener en cuenta el tiempo en el análisis, y el correcto tratamiento es el análisis de varianza con medidas repetidas¹⁶.

Un tipo de análisis que está cobrando especial relevancia en los últimos tiempos, y que no está incluido en la tabla II, es el análisis de supervivencia^{28, 29}. En este caso la variable dependiente es el tiempo transcurrido hasta el suceso de interés, y la variable independiente es dicotómica, utilizando para el análisis un test no paramétrico (log-rank test). Si existe más de una variable independiente, o ésta es continua, se requieren test paramétricos más complejos, como el modelo de Cox³⁰.

Análisis multivariante

El análisis multivariante es un conjunto de métodos estadísticos y matemáticos para describir, analizar e interpretar las observaciones multidimensionales, es decir, el material estadístico que proviene de la observación de más de una variable³¹⁻³⁴. Dadas las múltiples aplicaciones que tiene en todas las ciencias experimentales y el desarrollo que ha alcanzado en los últimos años, su utilización se ha hecho casi indispensable. Las espectaculares posibilidades que ofrece la informática en la actualidad han influido de forma decisiva en este desarrollo. Los ordenadores de alta capacidad y velocidad permiten la utilización de métodos complejos y de modelos con un elevado número de variables.

La estadística univariante consiste en describir una sola variable mediante características muestrales o realizar inferencias sobre la información de una muestra obtenida bajo una o varias condiciones experimentales. El análisis multivariante cumple también con los objetivos descriptivo o inductivo, pero

trabajando simultáneamente con más de dos variables. Es una estadística más compleja, pero también más potente, que utiliza métodos de álgebra lineal, cálculo numérico y geometría lineal, entre otros. En líneas generales, los métodos de análisis multivariante si diferencian según el área de aplicación (una o más poblaciones) y según los tipos de variables considerados en el análisis (tablas III y IV). El análisis factorial y el análisis de las correspondencias se utilizan en el caso de existir una única población y un grupo de variables de naturaleza razonablemente homogénea para intentar descubrir y reducir la estructura de la dependencia entre las variables que componen el grupo. El análisis discriminante y la regresión logística tratan de clasificar las observaciones en dos o más poblaciones utilizando para ello características propias definidas por variables que pueden ser de igual o diferente naturaleza. El análisis log-lineal representa la extensión multidimensional del análisis de tablas de contingencia, mientras que el análisis de regresión múltiple trata de definir o predecir una variable dependiente a través de un grupo de variables denominadas independientes o explicativas. Por último, el análisis multivariante de la varianza o de la covarianza —similar al análisis de la varianza, pero ajustando por diferentes factores— se refiere a la relación entre dos grupos de variables de naturaleza diferente entre ellos.

Paquetes de programas

En la actualidad, cualquier trabajo de investigación requiere de una forma u otra la utilización de la estadística. El investigador puede sentirse intimidado por dicha empresa, pero los ordenadores pueden serle de gran utilidad gracias a la facilidad para repartir tareas, almacenar y recuperar la información, procesar los datos y realizar el análisis deseado.

Existen calculadoras de bolsillo programables que para un análisis sencillo, con pocas variables y pocos casos, son suficientes para realizar los cálculos. El problema se complica cuando alguna de las tres premisas anteriores no se cumple: aumenta el número de casos, el número de variables o bien se necesita hacer un análisis más sofisticado. Por ejemplo, un estudio con 100 casos o con más de 30 variables, o que necesita realizar un análisis multivariante que normalmente arrastra muchos datos en forma de vectores y matrices de grandes proporciones, puede desbordar la capacidad de la calculadora e incluso de algún microordenador. En esta situación se hace imprescindible por la complejidad de los cálculos necesarios, la utilización de programas «empaquetados» bien enfocados para un microordenador o para un sistema mayor.

Un «paquete» es un conjunto de programas princi-

Tabla III. Relación entre el tipo de variable y los procedimientos que se deben utilizar en el análisis multivariante

Variable dependiente	Variabes independientes	Análisis apropiado
Dicotómica	Todas continuas.	Regresión logística, función discriminante.
Dicotómica	Dicotómicas y/o nominales.	Análisis log-lineal.
Dicotómica	Nominales y continuas.	Regresión logística.
Nominal	Todas continuas.	Categorizar variables continuas para hacer análisis log-lineal, función discriminante.
Ordinal	No existe análisis formal multivariante para este tipo de variable dependiente. Si se puede asumir una distribución continua, utilizar el análisis descrito debajo. En caso contrario, hacer análisis log-lineal.	
Continua	Todas continuas.	Regresión múltiple.
Continua	Nominales o dicotómicas.	ANOVA con <i>n</i> factores.
Continua	Nominales y continuas.	Análisis de covarianza (ANCOVA).

Tabla IV. Otros análisis multivariantes**Tipos de variables**

Dicotómicas y/o nominales.	Análisis factorial.
Continuas.	Análisis de las correspondencias.

pales que tienen una tarea común, un lenguaje de programación más o menos complejo y que puede ser utilizado por diferentes ordenadores. Existe en el mercado una gran variedad de ofertas de programas estadísticos. Uno de los más populares y completos es el BMDP (Biomedical Computer Program's), desarrollado bajo la dirección de W. J. Dixon en el centro médico de la Universidad de California, Los Angeles³⁵. Los programas se agrupan en varias categorías estadística descriptiva, tablas de contingencia, análisis multivariante, regresión paso a paso, análisis de varianza y covarianza, etc. Sus principales ventajas residen en la gran versatilidad en el manejo de ficheros, transformación de variables y edición de datos, además de ser capaz de realizar prácticamente cualquier tipo de análisis, incluyendo análisis de supervivencia y modelo de Cox. Otro conocido paquete es el SPSS (Statistical Package for the Social Sciences), desarrollado por N. H. Nie y colaboradores en la Universidad de Chicago³⁶. Como su nombre indica, los programas están diseñados para el análisis de datos de ciencias sociales. Proporciona también gran flexibilidad en cuanto al formato de los datos; transformación de éstos y manejo de ficheros. Los procedimientos en el SPSS incluyen análisis descriptivo, correlación simple y parcial, múltiples formas del análisis de la varianza y análisis de clasificación multitas. Otro apreciado paquete con posibilidades similares a las anteriormente descritas, tanto en el manejo de datos como en los tipos de análisis asequibles, es el SAS (Statistical Analysis System), desarro-

llado por la Universidad Estatal de Carolina del Norte³⁷. Hay un largo etcétera de paquetes, pero los tres anteriormente citados son los más utilizados y, a nuestro juicio, los más completos. Debemos recordar que el paquete estadístico no es capaz de decidir qué tipo de análisis es el correcto, dependiendo del diseño del estudio y del tipo de variables consideradas, ni qué condiciones deben cumplirse a la hora de realizar un análisis determinado. Es el investigador quien debe elegir y escribir el programa, valorar si es adecuado e interpretar los resultados.

Hay ciertos criterios que se deben tener en cuenta al evaluar la calidad de un paquete de programas estadísticos. La efectividad estadística se refiere al hecho de que el paquete permita al usuario no sólo obtener unos resultados que le hagan aceptar o rechazar un modelo, sino autoevaluar su análisis, comprobar que todas las condiciones requeridas para el modelo se cumplen y saber qué hipotéticos errores puede cometer. Además es importante que la variedad de técnicas estadísticas que se puedan realizar sea elevada, ya que en muchos casos las necesidades también lo son. Los ficheros de entrada de datos y de salida de resultados deben ser de fácil manejo para el usuario. También es importante que exista la posibilidad de utilizar algún fichero de salida, como, por ejemplo, residuos de una regresión o matriz de covarianza, como datos de entrada para agilizar el trabajo y ahorrar tiempo. Las fórmulas utilizadas deben ser correctas, estables numéricamente e indicar cuándo un procedimiento no es aplicable en el caso de que exista alguna operación que no se pueda realizar.

Hay algunas desventajas en cuanto a la utilización de los paquetes, como son la necesidad de ajustarse a los requerimientos de notación y lenguaje del programa, las limitaciones impuestas por los procedimientos numéricos que éstos realizan y por la salida de los resultados, que en ocasiones proporcionan una información que no nos interesa, faltando otra

que sí es necesaria (por ejemplo, estimación puntual de un estadístico, pero no el intervalo de confianza), aunque la ayuda que proporcionan para el cálculo hacen prácticamente imprescindible hoy día su utilización.

Conclusión

Quizá muchos clínicos se sientan desbordados en ocasiones por el crecientemente complejo tratamiento estadístico que deben dar a sus estudios de investigación con el fin de aprovechar toda la información recogida en los datos. Aunque esperamos que los conceptos someramente expuestos en este artículo ayuden en cierta medida a despejar alguna duda, el mensaje que queremos transmitir es que siempre que sea factible se debe consultar a un bioestadístico³⁸ desde las fases iniciales de diseño del proyecto de investigación y de recogida de datos, así como para el análisis e interpretación de resultados. Las ventajas en términos de tiempo, eficacia y, sobre todo, validez científica del estudio nos parecen incuestionables.

Bibliografía

- Moses LE: Statistical concepts fundamental to investigations. En *Medical uses of statistics*. Waltham, Massachussets, NEJM Books, 3-26, 1986.
- Cochran WG: *Sampling techniques*. New York, John Wiley & Sons, 1977.
- Berkson J: Limitations of the application of fourfold table analysis to hospital data. *Biomet Bull* 2:47-53, 1946.
- Abramson JH: Sampling methods. En *Survey methods in community medicine*. Edinburgh, Churchill Livingstone, 70-80, 1984.
- Glantz SA: Biostatistics: how to detect, correct and prevent errors in the medical literature. *Circulation* 61:1-7, 1980.
- Gardner MJ y Altman DG: Confidence intervals rather than p values: estimation rather than hypothesis testing. *Br Med J* 292:746-750, 1986.
- Simon R: Confidence intervals for reporting results of clinical trials. *Ann Intern Med* 105:429-435, 1986.
- Machín D y Gardner MJ: Calculating confidence intervals for survival time analyses. *Br Med J* 296:1369-1371, 1988.
- Altman DG y Gardner MJ: Calculating confidence intervals for regression and correlation. *Br Med J* 296:1238-1242, 1988.
- Campbell MJ y Gardner MJ: Calculating intervals for some non-parametric analyses. *Br Med J* 296:1454-1456, 1988.
- Ingelfinger JA, Mosteller F, Thibodeau LA y Ware JH: What are P values? En *Biostatistics in Clinical Medicine*. New York. Macmillan Publishing Co., Inc., 160-176, 1983.
- Freiman JA, Chalmers TC, Smith H Jr y Kuelber RR: The importance of Beta, the type II error, and sample size in the design and interpretation of the randomized clinical trial. En *Medical uses of statistics*. Waltham, Massachussets NEJM Books, 289-304, 1986.
- Friedman LM, Furberg CD y DeMets DL: Sample size. En *Fundamentals of Clinical Trials*. Littleton, Massachussets. PSG Publishing Company, Inc, 83-107, 1985.
- Carrasco JL: *El método estadístico en la investigación médica*. Ciencia 3, Madrid, 1983.
- Fleiss JL: *Statistical methods for rates and proportions*. New York. John Wiley & Sons, 1981.
- Armitage P y Berry G: *Statistical methods in medical research*. Blackwell. Oxford, 1987.
- Appleton DR: Analysis of data in nephrology I. Choosing the correct statistical test: dichotomous variables. *Nephrol Dial Transplant* 3:91-94, 1988.
- Appleton DR: Analysis of data in nephrology II. When to use the t-Test. *Nephrol Dial Transplant* 3:225-227, 1988.
- Appleton DR: Analysis of data in nephrology III. Survival analysis and analysis of variance. *Nephrol Dial Transplant* 3:340-344, 1988.
- Appleton DR: Analysis of data nephrology IV. Relationships between continuous variables. *Nephrol Dial Transplant* 3:478-481, 1988.
- Appleton DR: Analysis of data in nephrology B. Contingency tables. *Nephrol Dial Transplant* 3:694-696, 1988.
- Appleton DR: Analysis of data in nephrology VI. The generalised linear model. *Nephrol Dial Transplant* 3:830-832, 1988.
- Hollander M y Wolfe D: *Nonparametric statistical methods*. New York. John Wiley & Sons, 1973.
- Conover WJ: *Practical non-parametric statistics*. New York. John Wiley & Sons, 1980.
- Forrest M y Andersen B: Ordinal scales and statistics in medical research. *Br Med J* 292:537-538, 1986.
- Moses LE, Emerson JD y Hosdseini H: Analyzing data from ordered categories. *N Engl J Med* 311:442-448, 1984.
- Wallenstein S, Zucker CL y Fleiss JL: Some statistical methods useful in circulation research. *Circ Res* 47:1-9, 1980.
- Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, Mantel N, McPherson K, Peto J y Smith PG: Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I: Introduction and design. *Br J Cancer* 34:585-612, 1976.
- Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howards SV, Mantel N, McPherson K, Peto J y Smith PG: Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II: Analysis and examples. *Br J Cancer* 35:1-39, 1977.
- Cox DR: Regresion models and life tables (with discussion). *J Roy Stat Soc (B)* 34:187-220, 1972.
- Escudero LF: *Reconocimiento de patrones*. Ed. Paraninfo. Madrid, 1977.
- Afifi AA y Azen SP: *Statistical analysis. A computer oriented approach*. New York, Academic Press Inc., 1979.
- Cuadras CM: *Métodos de análisis multivariante*. Barcelona, Editorial Universitaria de Barcelona, 1981.
- Srivastava MS y Carter EM: *An introduction to applied multivariate statistics*. New York, North-Holland, 1983.
- BMDP Statistical Software: Berkeley, California, University Press, 1985.
- SPSS/PC for the IBM PC/XT/AT. Chicago. SPSS Inc., 1986.
- Statistical Analysis System (SAS). User's guide. Cary, North Carolina: SAS Institute Inc., 1985.
- Moses LE y Louis TA: Statistical consultation in clinical research: a two way street. En *Medical uses of statistics*. Waltham, Massachussets. NEJM Books, 338-345, 1986.