

MESAS REDONDAS ACADEMICAS

ASPECTOS ESENCIALES DE LA METODOLOGIA EN LA INVESTIGACION CLINICA¹

I

INTRODUCCION

BERNARDO SEPÚLVEDA²

LA PRESENTE mesa redonda, organizada por indicación de la Mesa Directiva de la Academia Nacional de Medicina, tiene dos propósitos principales. El primero es insistir en la conveniencia de que los trabajos de investigación clínica se ajusten a las normas generalmente aceptadas del método científico. El método fue establecido en sus principios fundamentales por Claudio Bernard en su célebre Introducción al Estudio de la Medicina Experimental; y comprende, como es bien sabido, cuatro etapas esenciales: observación, hipótesis, experimentación y razonamiento lógico. La hipótesis y el razonamiento se relacionan directamente con la lógica; mientras que la experimentación pertenece a la metodología propia-

mente dicha. En las últimas décadas el método de la investigación científica se ha perfeccionado con la utilización de normas para la planeación sistémica de los experimentos y con la aplicación del análisis estadístico a los resultados.

Nadie discute en la actualidad la necesidad de seguir las reglas del método científico en la indagación clínica; y si bien se reconoce que la estricta aplicación de la metodología científica a la investigación en sujetos humanos, tiene limitaciones de orden ético y práctico, también se acepta que es el sistema más adecuado para resolver de manera eficaz, rápida y completa numerosos problemas de la medicina clínica. Sin embargo, y aunque mucho se ha avanzado en este sentido, todavía no se aplican con la suficiente extensión los principios que deben dirigir estas indagaciones, de lo cual resulta que mucho tiempo y mucho esfuerzo se pierden en la elabora-

¹ Mesa redonda presentada en la sesión ordinaria del 3 de junio de 1970.

² Académico numerario. Hospital General, Centro Médico Nacional, Instituto Mexicano del Seguro Social.

ción de trabajos cuyos resultados son de poco valor científico. Y no se crea que el defecto existe sólo en medios como el nuestro; para mencionar un ejemplo, puede citarse lo dicho por Chalmers, en su mensaje presidencial a la Asociación Americana de Gastroenterología, en mayo del año pasado, que titula "Un reto a los investigadores clínicos".¹ Chalmers alude a la escasez de trabajos de tipo clínico que llenen los requisitos exigidos en la pesquisa científica, situación que no parece haber mejorado en los últimos 20 años.

Por añadidura, en una revisión de los artículos aparecidos en las 10 revistas médicas más leídas en los Estados Unidos de Norteamérica, se encontró que en menos del 28 por ciento de los trabajos de tipo analítico, las conclusiones podían considerarse válidas;² y en otro estudio sobre los artículos relativos a ensayos con drogas, publicados en el "Canadian Medical Association Journal" durante cinco años, se halló que de 203 trabajos, sólo 11 se ajustaban a los criterios establecidos.³

Seguramente, si se hicieran revisiones críticas semejantes de la literatura publicada en nuestras revistas médicas, los resultados serían comparables.

El segundo propósito de la Mesa Redonda es subrayar la estrecha relación entre los aspectos de orden lógico y los estrictamente metodológicos en la investigación clínica. Probablemente, se ha puesto demasiado énfasis en los aspectos relacionados con la planeación del experimento y con el tratamiento estadístico; y parece haber cierta tendencia a considerar que estas etapas del proceso metódico son suficientes por sí mismas. Es indudable que ambas representan las innovaciones más importantes en el método de la indagación científica; pero es necesario recalcar que la mayor utilidad se obtiene cuando la planeación del experimento es precedida por la formulación adecuada de la hipótesis; y cuando además, el análisis estadístico de los datos, es seguido por la interpretación lógica de los resultados.

REFERENCIAS

1. Chalmers, T. C.: *A challenge to clinical investigators. Presidential address. Gastroenterology* 57: 631, 1969.
2. Schor, S. y Karten, I.: *Statistical evaluation of medical journals manuscripts. JAMA* 195: 145, 1966.
3. Mahon, W. A. y Daniel, E. E.: *A method for the assessment of reports of drug trials. Canad. Med. Ass. J.* 90: 565, 1964.

II

LA HIPOTESIS

BERNARDO SEPÚLVEDA¹

LA FORMULACIÓN de la hipótesis es una de las etapas esenciales en el proceso de la investigación científica; de hecho, constituye la base del método experimental, y tanto es así que la mayor parte de las investigaciones se llevan a cabo con el deliberado propósito de comprobar el grado de veracidad de una hipótesis. En esto consiste precisamente la importancia de la hipótesis, cuya función principal es servir como punto de partida a nuevos proyectos de experimentación, encaminados a resolver los problemas presentes. En consecuencia, la hipótesis es el móvil primordial en la pesquisa científica; y por ello desempeña papel relevante en la adquisición de nuevos conocimientos.

Las hipótesis generalmente proceden de la observación, y pueden originarse por diversos procesos mentales, tales como imaginación, intuición y razonamiento, que dependen de la aptitud intelectual del investigador y del conocimiento que tenga sobre las cuestiones que se analizan. Por consiguiente, la capacidad para enunciar hipótesis valiosas, es en buena parte una cualidad personal del investigador y no puede quedar sujeta a reglas fijas. Indudablemente, una propiedad característica

del genio es la formulación de hipótesis fecundas en sus implicaciones, y por ello los grandes avances de la ciencia se han debido a teorías surgidas de cerebros geniales. Sin embargo, es también indudable que hipótesis útiles pueden surgir de inteligencias comunes, cuando hay sólida preparación en los temas que se estudian e interés en resolver las incógnitas que presentan; y que a igualdad de circunstancias, es posible obtener mayor ventaja en la aplicación de la hipótesis, si se toman en cuenta algunas normas derivadas de la experiencia.

En primer término, conviene señalar algunas particularidades propias de la hipótesis. La primera es que ninguna hipótesis puede ser probada como absolutamente cierta, y que ello no obstante, puede ser aceptada en la práctica si los resultados del experimento le dan suficiente apoyo. Esto se debe a que todo el proceso de confirmación de la hipótesis es de naturaleza inductiva. Como es sabido, en la inducción, siempre es posible que las premisas sean verdaderas y la conclusión falsa, y viceversa. Cuando los resultados de un experimento realizado para probar una hipótesis, están de acuerdo con la misma, lo más que puede afirmarse es que se ha obtenido evidencia confirmatoria;

¹ Académico numerario. Hospital General, Centro Médico Nacional, Instituto Mexicano del Seguro Social.

pero a sabiendas de que hay siempre la posibilidad que la misma hipótesis sea rechazada o modificada posteriormente por nuevos descubrimientos; o bien, de que surjan nuevas hipótesis sobre el propio fenómeno, que asimismo sean confirmadas posteriormente.¹ Por añadidura, el grado de confirmación de la hipótesis puede ser alto, mediano o bajo,² lo cual depende de la cantidad y la variedad de la evidencia disponible.³ A pesar de estas limitaciones, y como antes se dijo, si la hipótesis obtiene suficiente evidencia experimental en su favor, puede aceptarse como válida en la práctica.

En la historia de la ciencia se encuentran numerosos casos de teorías que fueron útiles temporalmente para la explicación de los hechos, por tener cierta base experimental, y que han sido reemplazadas más tarde al adquirirse nuevos conocimientos. Un buen ejemplo de esto es la hipótesis formulada por Semmelweis en 1847, sobre la etiología de la fiebre puerperal, según la cual, el padecimiento era debido a la "materia cadavérica" o a la "materia pútrida de los organismos vivos", transmitida a las mujeres en trabajo de parto por las manos de los médicos y los estudiantes de medicina. Semmelweis obtuvo evidencia en apoyo de su hipótesis al comprobar que la mortalidad por fiebre puerperal descendió del 10 por ciento al uno por ciento, una vez que se ordenó al personal médico lavarse las manos en solución de cloruro de cal, antes de examinar a las parturientas.⁴ La hipótesis del genial cuanto desventurado Semmelweis poseía cierto fundamento teórico y, sobre todo,

tenía validez práctica. De haber sido universalmente aceptada, hubiera salvado la vida de muchos miles de mujeres; pero nunca pudo haber tenido comprobación absoluta; y fue sustituida treinta años más tarde por la teoría microbiana de Pasteur.

Otra particularidad de la hipótesis, quizá aún más interesante, se relaciona con la anterior. Es el hecho de que hipótesis incorrectas pueden ser de gran utilidad para el progreso científico. Según la original frase de Trotter, "En la ciencia, el deber primario de las ideas consiste en ser útiles e interesantes, más aún que ser *ciertas*".⁵ Como antes se dijo, la función principal de la hipótesis es servir como punto de partida a investigaciones, que de manera a veces inesperada conducen a valiosos descubrimientos.

Dos ejemplos ilustran lo que acaba de mencionarse. Ehrlich, en sus investigaciones sobre quimioterapia, partió de la hipótesis siguiente: puesto que algunos colorantes tiñen selectivamente las bacterias, puede haber sustancias que sean absorbidas también selectivamente por los microorganismos y que los destruyan sin dañar al huésped. Guiado por esta idea, y por la convicción de que las drogas sólo podían actuar cuando eran fijadas en forma específica por los gérmenes, ensayó numerosos compuestos, hasta que descubrió en 1907 la eficacia del "Salvarsan" contra el treponema de la sífilis. La hipótesis era errónea, ya que los agentes quimioterápicos pueden ser efectivos sin necesidad de fijación selectiva por los microorganismos; y sin embargo, fue fructífera en sus resultados.

El segundo ejemplo se refiere al descubrimiento de la sulfanilamida. La hipótesis original de Dogmak era *que* los colorantes azoados tenían propiedades antibacterianas, y, sobre esta base, demostró en 1932 la eficacia del "prontosil" en infecciones estreptocócicas. Por algún tiempo, se creyó que la eficacia terapéutica se debía al azocolorante; sin embargo, en 1935, Tréfoüels y colaboradores comprobaron que la fracción no colorante de la sustancia, o sea la sulfonamida, era la única dotada de acción quimioterápica.⁶ Nuevamente, una hipótesis errónea conducía a un descubrimiento de valor extraordinario.

Una vez señaladas estas particularidades de la hipótesis, procede indicar algunas normas para su aplicación en la investigación clínica.

La primera condición que debe exigirse a la hipótesis, es que sea formulada en términos precisos. Una hipótesis enunciada en forma general e indefinida, es inútil como punto de partida de cualquier estudio experimental; y sin embargo, éste es uno de los defectos más comunes en los proyectos de investigación y se comete con mayor frecuencia probablemente en los ensayos terapéuticos. Para ejemplificar, suponemos que se pretende probar la eficacia de una nueva droga en el tratamiento de la úlcera péptica. Sería un error expresar la hipótesis de manera tan general, que el estudio comprendiera toda clase de úlceras pépticas. Es necesario dividir la hipótesis global en una serie de hipótesis concretas, que sirvan de base a investigaciones especí-

ficamente planeadas. Esto es lo que constituye la formulación de las hipótesis unitarias, que es un proceso esencial en la iniciación del estudio.⁷ En el ejemplo indicado, sería indispensable enunciar hipótesis distintas para ensayar la droga separadamente en la úlcera gástrica y en la duodenal; pero todavía faltaría precisión a tal hipótesis, deberían especificarse otros aspectos relacionados con la definición de la enfermedad. ¿Se incluyen tanto úlceras pépticas antiguas como recientes? ¿Se incluyen tanto úlceras complicadas como sin complicaciones? Sólo de este primer análisis, pueden resultar varias hipótesis unitarias.

Pero, enseguida, habría que precisar otros tantos aspectos dependientes de la droga, tales como la dosis, el modo de administración y la duración del tratamiento. Al mismo tiempo, debería considerarse en la hipótesis si la droga va a ser administrada como tratamiento exclusivo o si va a ser usada en combinación con otros medicamentos o bien con dieta o con reposo en cama. Además, sería necesario definir la interpretación de los resultados; es decir, sobre qué bases se juzgará la eficacia terapéutica de la nueva droga. En el caso presente, sería totalmente distinto señalar cualquiera de estos dos criterios de curación: la simple desaparición del dolor o la cicatrización de la úlcera comprobada por rayos X. De acuerdo con cada uno de estos puntos concretos, deberían formularse las correspondientes hipótesis unitarias; en tal forma, que el planteamiento de las cuestiones individuales se hiciera de la manera más clara y precisa posible, lo

cual contribuirá a que las soluciones que se encuentren sean asimismo bien definidas. A este respecto, convendría repetir el acertado dicho de Bradford Hill: "La precisión en la pregunta, probablemente lleve a buscar la precisión en la respuesta".⁸

En segundo lugar, debe tomarse en cuenta que cada hipótesis unitaria sirve de base a un experimento, ensayo o estudio clínico, que está destinado específicamente a la comprobación de tal hipótesis y que podría llamarse asimismo experimento o estudio unitario. Por consiguiente, las conclusiones derivadas de esa investigación, se refieren en particular a la hipótesis unitaria y no son necesariamente aplicables a circunstancias distintas. Aun cuando a veces puede ser legítimo extender las conclusiones, siempre debe considerarse el riesgo de extrapolar los resultados más allá de lo que permiten los datos obtenidos.

A continuación, conviene mencionar dos precauciones necesarias en el uso de la hipótesis. Una, es la recomendación de abandonarla, tan pronto como se demuestre su inutilidad. Esto no siempre es fácil, puesto que el investigador se encariña paternalmente con su teoría y le cuesta trabajo rechazarla. Decía Claudio Bernard que aun cuando las hipótesis son esenciales en la planeación de un experimento, "no se debe estar nunca absorto en el pensamiento que se persigue, ni ilusionarse con el valor de las ideas o de las teorías científicas";⁹ y él mismo demostró en varias ocasiones su flexibilidad intelectual para desechar hipótesis que no eran confirmadas por los hechos. Como ilustración, podría citarse el abandono de

sus hipótesis iniciales sobre el órgano "destructor" del azúcar de los alimentos y sobre la función de los nervios simpáticos en la calorificación, después de que tales hipótesis eran refutadas por sus experimentos.¹⁰

La segunda precaución es la de estar alerta para evitar que el prejuicio influya de manera inconsciente en la interpretación de los resultados. Es una falla humana ver sólo lo que deseamos ver; y es por ello natural la tendencia a tratar de que los hechos se ajusten a las ideas. Para evitar este riesgo, es necesario tener presente siempre, que la hipótesis, por muy plausible que parezca, es sólo una suposición; y que la plausibilidad no sustituye a la evidencia.

Para terminar, quiero referirme a dos puntos importantes: el primero, es la necesidad de considerar en el planteamiento de la hipótesis, la llamada "hipótesis de nulidad"; el segundo, es la conveniencia de conocer el papel de las hipótesis auxiliares.

Todo experimento requiere tomar en cuenta la posibilidad de que los resultados sean debidos a la casualidad. Por supuesto, la investigación bien planeada, pretende que la casualidad intervenga al mínimo; y, en todo caso, pretende también medir la probabilidad de la participación de la casualidad. Esta es la función de la hipótesis de nulidad, que sirve de base para determinar la frecuencia probable con que interviene el azar en los resultados del experimento; y para cuya determinación, se utilizan los grupos testigos y el análisis estadístico.

En términos generales, la "hipótesis de nulidad" sostiene que los resulta-

dos de cualquier investigación son debidos al azar; o bien que las diferencias entre dos hechos no son significativas, sino expresión de meras coincidencias. Por extraño que parezca, en todo proyecto de experimento debe incluirse la "hipótesis de nulidad", aceptando la posibilidad teórica de que los resultados sean producto del azar; pero, en realidad, con el propósito de que el experimento refute la "hipótesis de nulidad". Esto parece un poco artificial, ya que se supone que la investigación tiene como primer objetivo, evidenciar la relación de causa a efecto en los resultados, más bien que demostrar la probabilidad de la intervención del azar. Tampoco parece lógico a primera vista, aceptar desde el principio una hipótesis contraria a la finalidad del experimento, para luego tratar de rechazarla con las conclusiones del propio experimento.¹¹

Sin embargo, debe considerarse por una parte la conveniencia de que el investigador adopte constantemente una actitud de cierto escepticismo, que le hará tener presente siempre la contingencia de que participe la casualidad en los resultados, y lo llevará en consecuencia a tomar todas las precauciones para evitarla. Por otra parte, debe también tomarse en cuenta que la "hipótesis de nulidad" representa un valioso método de comprobación, cuya utilidad ha sido confirmada con gran variedad de problemas de investigación.

En vista de tales razones, es necesario considerar la "hipótesis de nulidad" desde la iniciación del experimento. Para destacar su importancia, basta mencionar que Fisher, una de las figu-

ras más distinguidas en el campo de la bioestadística, ha dicho que, verdaderamente, "cada experimento existe para refutar la hipótesis de nulidad".²

Las hipótesis auxiliares son teorías que han sido confirmadas por investigaciones previas, por lo cual pueden servir de apoyo a la hipótesis principal; sin embargo, al igual que todas las hipótesis, pueden ser rechazadas o modificadas por descubrimientos posteriores. En el ejemplo antes mencionado de Semmelweis, la hipótesis principal sobre el modo de transmisión de la fiebre puerperal, se apoyaba en la hipótesis auxiliar de que la solución de cloruro de cal tenía poder suficiente para eliminar de las manos las materias contaminantes. En este caso, la hipótesis auxiliar sirvió de apoyo a la principal; pero aún suponiendo que no hubiera sido así, tal resultado no habría probado que la teoría de Semmelweis era necesariamente falsa, pues podría haberse debido a que la solución de cloruro de cal no tenía el poder que se le atribuía.

Este punto de las hipótesis auxiliares será tratado con mayor amplitud por el doctor Jinich, ya que se relaciona con el tema de la interpretación lógica.

REFERENCIAS

1. Cohen, M. R. y Nagel, E.: *An Introduction to Logic and Scientific Method*. New York, Harcourt, Brace and World, Inc. 1934, p. 393.
2. Salmon, W. C.: *Logic*. Englewood Cliffs, Prentice Hall, Inc. 1963, p. 77.
3. Russell, B.: *La Perspectiva Científica*. Barcelona, Ariel. 1969, p. 57.
4. Hempel, C. G.: *Philosophy of Natural Science*. Englewood Cliffs, Prentice Hall, Inc. 1966, p. 5.
5. Trotter, W.: Citado por Beveridge, W. I. B. en: *The Art of Scientific In-*

- investigation. New York, W. W. Norton and Company Inc. 1957, p. 4.
6. Goodman, L. S. y Gilman, A.: *The Pharmacological Basis of Therapeutics*. 3a. ed. New York, The Macmillan Company. 1967, p. 1145.
 7. Hamilton, M.: *Lectures on the Methodology of Clinical Research*. Edinburgh, E. and S. Livingstone Ltd. 1961, p. 15.
 8. Hill, A. B.: *Statistical Methods in Clinical and Preventive Medicine*. Edinburgh, E. and S. Livingstone Ltd. 1962, p. 9.
 9. Bernard, C.: *Introducción al Estudio de la Medicina Experimental*. Versión castellana por J. J. Izquierdo. México, Imprenta Universitaria. 1942, p. 262.
 10. Bernard, C.: Referencia anterior, pp. 257-266.
 11. Sepúlveda, B.: *Metodología de la investigación clínica en Gastroenterología*. Rev. Mex. Gastroent. 33: 505, 1968.
 12. Fisher, A.: *The Design of Experiments*. 6a. ed. Edinburgh, Oliver and Boyd. 1935, pp. 7-18.

III

LA PLANEACION DEL EXPERIMENTO

JESÚS KUMATE¹

EN LA APLICACIÓN del método experimental, el investigador trata de mantener constantes todos los factores involucrados en el determinismo de un fenómeno y observa los efectos producidos por la variación de uno de los factores participantes; de esta manera efectúa un análisis que evalúa la aportación de los diversos parámetros. El *desideratum* consiste en poder generalizar las observaciones y conclusiones para formular un principio general. Antes de que las conclusiones se acepten como verdaderas, deben ser confirmadas por la repetición del experimento, de preferencia por otros investigadores; por tanto, la reproducibilidad es condición *sine qua non* del método experimental.

Desgraciadamente, los resultados ob-

tenidos por la experimentación exhiben una dispersión mayor de la deseable y en el campo de la biología, la variación es aún mayor; se ha llegado a decir que... "*la característica más constante de los fenómenos biológicos es su variabilidad*"...; en la Medicina es frecuente oír el aforismo... "*no hay enfermedades, sino enfermos*"... o la expresión, mezcla de pesimismo y desesperación: "*en medicina todo es posible*".

Lo anterior ha hecho imperativo el empleo de grupos testigos; la utilidad de estas unidades comparativas, ha establecido firmemente el criterio de que los resultados valen lo que sus testigos o que la observación experimental más cuidadosamente planeada y ejecutada, es inútil si no existe un grupo comparativo.

En la investigación clínica son vá-

¹ Académico numerario. Hospital Infantil de México.

lidos los principios enunciados por Helmholtz y Dale de que "...*toda ciencia implica medición*"... y que... "*toda verdadera medición es, en esencia, comparativa*"...

Corresponde ahora concretar, después de haber formulado la hipótesis de trabajo, en base a la información disponible y a las posibilidades de asociación entre diversos factores, cómo se hará la prueba experimental de la hipótesis bajo condiciones en las que se ejerza control de todos los factores de variación involucrados y después, la definición de la población a la que se pretende generalizar los resultados de la experimentación. De hecho en la práctica de la investigación clínica, gran parte de la literatura relacionada con la planeación experimental es la descripción de los diferentes métodos de muestreo.^{1, 2}

La hipótesis ofrece una o más soluciones lógicas a la pregunta o incógnita planteada por la observación clínica. En esta etapa la intervención del ingrediente personal del investigador, su imaginación, genio, intuición, suerte, recursos materiales y equipo disponible, participarán para dar el toque que dará como resultado un trabajo clásico, elegante, ilustrativo, insuficiente, mediocre o intrascendente como la gran mayoría de las 600,000 comunicaciones registradas anualmente en el campo de la biología. Todo esto, independiente de la planeación experimental más cuidadosa. Hasta hace 20 años, los temas de la investigación clínica tenían como objeto, casi siempre, el descubrimiento de agentes etiológicos, la elucidación de un mecanismo patogénico, la identi-

cación de una nueva entidad nosológica y como era de esperarse en los médicos, la búsqueda y prueba de nuevos agentes terapéuticos o profilácticos. Vivimos actualmente un momento, en donde el equilibrio entre las metas perseguidas otrora, ha sido desplazado por el deseo de cultivar los ensayos terapéuticos y la explotación de la experiencia acumulada en los archivos de las grandes instituciones hospitalarias.

Es cierto que la complicación en la metodología del gabinete y la casi irresistible fascinación que la tecnología moderna, a través de las computadoras, radioisótopos, electrónica, etc. ejercen sobre la imaginación, han hecho que la búsqueda de agentes etiológicos y la curiosidad por aclarar la fisiopatología de una enfermedad, hayan sido relegados al gabinete o en el mejor de los casos, al estudio en equipo, asequible sólo a grandes y ricas instituciones. Conviene recordar que aún en nuestros días, muchos problemas fundamentales han sido resueltos por la utilización del mejor instrumento disponible en el laboratorio equipado con la mayor largueza, i.e.: el cerebro humano. Así, el descubrimiento del fenómeno de la tolerancia inmunológica³ (Premio Nobel de Medicina en 1959), requirió sólo de: ratones, jeringas, bisturíes, material de hemostasia y sutura, suero fisiológico, un hemocitómetro y un microscopio óptico común. La demostración de la vía placentaria en la transmisión de la inmunidad materna al feto,⁴ necesitó la comparación de los niveles de antitoxina diftérica en recién nacidos con atresia esofágica y sus madres; el material y aparatos usados fueron: co-

nejos, una muestra de toxina diftérica, jeringas y material para corte de pelo.

MUESTREO

Los grupos en estudio en cualquier investigación clínica se espera sean muestras representativas del universo estadístico (población ideal que comprende a todos los individuos que poseen características comunes al grupo sujeto al estudio comparativo); sin embargo, el peligro de que la muestra esté infiltrada de prejuicio en su elección es el fantasma omnipresente en todas las fases y clases de muestreo.

Muestreo aleatorio: es aquel en el que todos los miembros del universo estadístico, tienen la misma oportunidad de ser seleccionados para integrar la muestra en estudio. En sentido estricto, la muestra aleatoria es un concepto teórico, sin equivalente en la realidad ya que todas las maniobras para lograrla, incluyendo las tablas de números aleatorios, son simples aproximaciones. Ya que el término aleatorio se utiliza frecuentemente con otra connotación, Cochran⁵ aconseja un calificativo adicional y prefiere la frase: muestreo aleatorio sin restricción.

Muestreo sistemático: Se practica cuando el grupo es muy grande y se lleva a cabo por la elección de los miembros de la muestra, en base a un módulo constante. A partir de la lista completa del grupo, es elegido un miembro por cada 10, 50 ó 100 u otra cifra que reduzca la muestra a los recursos disponibles para el muestreo y los requerimientos de precisión y exactitud, en base al error de la muestra.

Aunque no es una muestra aleatoria, si no hay un agrupamiento tendencioso en la lista, puede ser equivalente a la ideal. Una falla del muestreo sistemático es que no se puede conocer la validez del error de muestreo de la estimación obtenida.⁶

Muestreo estratificado: consiste en subdividir el grupo en subgrupos, de acuerdo a las intenciones del estudio v.gr. sexo, edad, estado nutricional, condición socioeconómica; en lo demás se procede como en una muestra aleatoria.

Muestreo apareado: con este sistema, por cada uno de los miembros del grupo en estudio, se busca un testigo que posea todas las características del primero. De esta manera se obtienen dos muestras de idéntica magnitud y similares en los caracteres variables que pueden confundir la interpretación. La maniobra crítica es la selección de los individuos que aparecen a los problemas; se elige a la siguiente persona o paciente con las características buscadas, pero no se deja libertad para escoger a cualquier sujeto que tenga las credenciales clínicas solicitadas.

Muestreos tendenciosos: son todos aquellos que no se engloban bajo el rubro de muestreo probabilístico y que pueden ser útiles en algunos casos en donde se puede calcular la aportación de los factores tendenciosos y a la manera del "blanco" en la fotometría, se puede restar el componente espurio y obtener un valor libre de prejuicio de una muestra tendenciosa. La importancia de no introducir elemento de parcialidad en las muestras, radica en que los métodos estadísticos no son aplica-

bles en grupos obtenidos por un muestreo tendencioso.

Las fuentes más frecuentes de parcialidad en la muestra, son las estadísticas de los estudios *postmortem* ya que su material deriva casi siempre de muestras obtenidas por métodos que invalidan las pruebas empleadas en la estadística. El ejemplo más famoso de esta falla es el de Pearl,⁷ quien en 1929 comunicó la aparente incompatibilidad entre la tuberculosis y el cáncer; la falacia tuvo su origen en que la frecuencia de la tuberculosis pulmonar, encontrada en autopsias de cancerosos, fue mucho menor que la obtenida en *post-mortem* de enfermos sin cáncer. El haber realizado el estudio en el mismo hospital, en la misma época y con apareamiento de sexo, edad y raza, pareció reforzar la veracidad de la diferencia pero estudios posteriores señalaron que el hallazgo fue producto de la selección tendenciosa del material de estudio.

Errores de muestreo: resultan del hecho que se trabaja sólo con una parte del universo y comprenden las inherentes a la muestra *per se* y otros debidos a la ausencia de respuesta o de medición.

En la práctica del muestreo deben especificarse el máximo error permisible y la probabilidad de que pueda ser excedido. Lo anterior implica la estimación del error estándar (σ); en distribuciones normales, el valor del universo es 2σ en 95.5% de los casos. La especificación de σ debe ser hecha por quien va a utilizar los datos y dependerá de la precisión deseada.

El caso de falta de respuesta en una

encuesta o de no-medición en un grupo, constituye una deficiencia imposible de remediar, ya que si se trata de incidencia, ésta puede ser desde 0 a 100%; el decir que la cifra en la población no estudiada, no debe diferir mucho del grupo donde sí hubo respuestas, no pasa de ser un buen deseo, sin ningún valor estadístico real.

Tamaño de la muestra: Se obtiene por la fórmula.

$$n = \frac{t^2 a p q}{E^2}, \text{ en donde}$$

n = tamaño de la muestra,

t_a = desviación normal correspondiente a la probabilidad " α " de exceder el error permisible máximo,

p = tasa de prevalencia "desconocida"

q = $1-p$

E = error permisible máximo.

A primera vista resulta absurdo que deba conocerse " p " para calcular " n ", pero es inescapable que ambos parámetros están íntimamente relacionados. Es obvio que si " p " es muy elevada, v.gr. parasitosis intestinal en una área tropical preindustrial, " n " puede ser pequeño y dar una gran precisión; por el contrario si " p " es muy pequeño, v.gr. un defecto congénito del metabolismo, " n " debe alcanzar decenas de miles para lograr una aproximación de la verdadera incidencia.

En este capítulo, cualquier información previa: un estudio piloto previo, o una investigación similar en otra área, puede dar un dato que oriente acerca

del valor de "p" y pueda despejarse el valor de "n" en la fórmula.⁸

Con mucho el mayor volumen de los problemas afrontados en la investigación clínica son: a) los ensayos terapéuticos y b) las encuestas retrospectivas y prospectivas en busca de asociaciones etiológicas.

ENSAYOS TERAPÉUTICOS

Ocupan un primerísimo lugar en la actualidad, ofrecen una oportunidad terapéutica en muy diversas enfermedades. La modalidad clásica es la llamada "ciego doble"; consiste de los siguientes componentes:

- a) Definición del o de los grupos con tratamiento y el o los grupos testigos bajo tratamiento con medicamentos conocidos.
- b) Integración de una lista con números aleatorios, en cantidad igual a la necesaria para formar los grupos de estudio y testigos.
- c) Distribución a los números en la lista según un orden uniforme para formar los miembros de cada grupo y su reacomodo, en orden progresivo, una vez que cada número ha sido adscrito a un grupo determinado.
- d) Acuerdo de los criterios clínicos y gabinete para aceptar como adecuados, los pacientes presuntos candidatos al estudio y su distribución en los grupos, según la llegada al hospital o institución donde se realiza el proyecto. El clínico encargado de esta fase, desconoce la clase de medicamento que reciben las pacientes. (primer ciego).

- e) Evaluación de los resultados, en base a criterios objetivos determinados previamente, sobre los casos individuales y
- f) Tabulación de los resultados por un clínico o equipo, diferente del participante en "d" y quien desconoce qué pacientes recibieron los medicamentos y quiénes sirvieron de testigos (segundo ciego).

Una variante en la presentación del ciego doble⁹ es la que expresa que los pacientes no deben conocer qué clase de medicamento reciben (primer ciego), y los médicos deben ignorar a cuáles enfermos se les administraron los placebos o los medicamentos en prueba (2o. ciego). Los grupos testigos deben recibir el tratamiento convencional o el placebo (si está justificado); en forma tal que imite en la medida de lo posible, a la apariencia y forma de administración del medicamento de prueba.

Las limitaciones del ensayo terapéutico son de dos índoles:

1. No informa respecto al o los factores que condicionan la respuesta al tratamiento; v.gr. si un medicamento antileucémico induce un 80% de remisiones en tanto que otro ofrece sólo un 50%; no es seguro que este 50% también hubiera remitido con el quimioterápico que logró el 80% de remisiones.
2. La elección del tamaño de la muestra es completamente arbitraria y no hay manera de calcular, aunque sea aproximadamente, el error de la muestra, como no sea por comparación con el grupo control co-

nocido y se infiera que la variación con el nuevo medicamento será similar.

Los criterios de evaluación deben ser lo suficientemente sensibles para descubrir el o los efectos explorados, amén de su capacidad para exhibir el efecto de dosis/respuesta. Si no se atiende a estos dos requisitos, los resultados negativos no tendrán validez y las respuestas positivas no podrán cuantificarse.

La posibilidad de trabajar con el número adecuado de pacientes es la aplicación del análisis secuencial diseñado por Wald¹⁰ mediante el cual, previo conocimiento de la población normal y su variación, es posible suspender el estudio con el nivel de probabilidad deseado, cuando el efecto acumulado de los casos, conduce a la localización de la observación en aceptación o rechazo de la hipótesis de nulidad planteada.

ENCUESTAS RETROSPECTIVAS

En esta clase de estudios se trata de conocer, a partir de un efecto observado, la causa probablemente asociada mediante el interrogatorio de sus antecedentes. El ejemplo más ilustrativo de las ventajas e inconvenientes es la observación de Gregg¹¹ en 1941, de que en muchos de los niños con catarata congénita había el antecedente de rubéola en la madre durante el 1er. trimestre del embarazo. La observación fue muy valiosa porque sirvió para demostrar una asociación patogénica, i.e.: que la rubéola durante el período embrionario tiene efectos teratogénicos. Sin embargo, la cuantificación inicial

del efecto fue falsa ya que, como era de esperarse, el interrogatorio fue intencionado hacia las madres de recién nacidos con catarata y no se tomó en cuenta a las mujeres embarazadas que padecieron rubéola en las primeras 12 semanas del embarazo y dieron a luz neonatos sin ninguna malformación. *A posteriori* se trató el corregir el prejuicio de la muestra a base de interrogar muchas mujeres con embarazos recientes, independientemente del estado del producto, pero la rubéola es una enfermedad muy benigna y la anamnesis tuvo un valor muy dudoso.¹²

ESTUDIOS PROSPECTIVOS

Consisten en iniciar el estudio con un grupo en el que ha operado el presunto factor causal, v.gr., mujeres embarazadas que padecen rubéola en diferentes etapas de la gestación, y en las que se registra al término del embarazo, la incidencia de malformaciones congénitas en los recién nacidos.¹³ En contraste a los estudios retrospectivos que parten del efecto para buscar la causa, en los estudios prospectivos se escoge un agente etiológico y se observa el tiempo necesario para registrar la ocurrencia o ausencia del efecto buscado. En el caso de la rubéola, la observación longitudinal reveló que la incidencia del efecto causal había sido sobrevalorada y permitió en un tiempo razonable, aclarar el problema de la intensidad del efecto y confirmar la necesidad de la asociación causal entre rubéola y teratogénesis. Sin embargo, en otros casos una observación longitudinal puede requerir de muchos años, como el caso de la asociación del tabaquismo con el

cáncer broncogénico; en cambio una encuesta retrospectiva pudo señalar una asociación sugestiva de causalidad, cuando se analizaron los hábitos de tabaquismo en pacientes con cáncer broncogénico y en grupos comparativos de otras neoplasias o individuos normales, de edad y sexo similares al grupo de cáncer pulmonar. Los resultados pudieron ofrecer un efecto de dosis ya que se observó una gradación muy clara entre los no fumadores, los moderados y aquellos con tabaquismo intenso.^{14, 15}

ERRORES TIPO I Y II

Antes de efectuar la computación y tabulación de los resultados se establecen los niveles de significación estadística con que se va a juzgar la hipótesis de nulidad (H_0) planeada al iniciar el estudio.

Si la hipótesis de nulidad es cierta y es erróneamente rechazada, se comete un error tipo I; así cuando al analizar estadísticamente una diferencia se la declara significativa, se tiene el error tipo I. Por el contrario; si la hipótesis de nulidad es falsa y se acepta, se cae en un error tipo II que sería la situación conducente a declarar no significativa la diferencia encontrada durante la comparación estadística.¹⁶ En el caso de error tipo I, se reduce la posibilidad de cometerlo si en vez del nivel convencional de $p < 0.05$ se reduce a 0.1 o aún 0.001 y se conoce exactamente la probabilidad de caer en él, i.e.: 1/20 1/100 y 1/1000 respectivamente.

En el error tipo II, independiente del nivel de probabilidad contraria fijado, no es posible conocer qué tan frecuentes serán los errores tipo II; lo

único seguro es que habrá menos errores cuando se usa el nivel de 0.05 que en el caso de 0.01.

Si el tamaño de la muestra ha sido fijado de antemano (como ocurre en muchos casos de investigación clínica), no es posible disminuir ambos errores simultáneamente.

Resulta evidente que el investigador se encuentra ante el dilema de Scila y Caribdis y los elementos de juicio para tener un criterio correcto respecto al nivel más adecuado, son los derivados de la gravedad o importancia comparativa en caso de cometer un error o el otro; así en la evaluación comparada de varios medicamentos, si uno de ellos exhibe una toxicidad peligrosa, es aconsejable el plantear una diferencia grande en actividad terapéutica a favor del más tóxico para aceptar el riesgo de esos efectos desfavorables; en este caso p debe ser 0.01 ó 0.001. En el análisis químico de un lote de ese medicamento con relación a su actividad, el cometer un error tipo II equivaldría al uso de un fármaco peligroso, en tanto que un error tipo I llevaría al rechazo de los medicamentos. Las consecuencias desagradables de este último error serían económicas para el productor, pero en caso de cometer el error tipo II, si la toxicidad es elevada, conduciría a muertes o invalideces serias en los pacientes que los emplearan. Si comparamos la experimentación o investigación clínica con el viaje de un explorador, la herramienta a él asequible por el método estadístico y en particular por los lineamientos de la planeación experimental, le pueden informar sobre:

- a) Las reglas de tránsito en el camino elegido en la búsqueda de las incógnitas por develar (pruebas y distribuciones de muestras);
- b) La magnitud y duración del viaje, si hay una idea previa acerca del mismo problema (tamaño de la muestra);
- c) Cuándo se puede dar por terminada la exploración (análisis secuencial).

La planeación no puede prever si:

- a') El camino elegido es el adecuado;
- b') Una vez iniciada la exploración, ante una alternativa se está frente a un atajo, un callejón sin salida o ante el camino real;
- c') Una asociación estadísticamente significativa, implica causalidad o se trata de una conexión fortuita y sin importancia.

No debe extrañar que en un campo como la biología y la medicina en particular, las cautelas y precauciones que impone la planeación experimental sean más complicadas que en las ciencias exactas, pero aun en éstas, tiene validez lo asentado por Robert Boyle,¹⁷ físico-químico y filósofo, quien en 1673 escribió en un ensayo titulado "concerniente a los fracasos experimentales" el siguiente párrafo: "Siento Pirófilo, que a las muchas dificultades con las que te puedes encontrar y que debes vencer, en la prosecución seria y eficaz de la filosofía experimental, debo agregar una más que probablemente te sorprenda y desaliente: aprenderás que en muchos de los experimentos publicados

o comunicados a ti por autores dignos de confianza o tal vez recomendados por tu propia experiencia, puede ocurrir que en una prueba posterior no confirmen tu predicción, ya sea porque no ocurren de manera constante, o por lo menos varían mucho más de lo esperado por ti..."

REFERENCIAS

1. White, C.: *Sampling in medical research*. Brit. Med. J. 2: 1284, 1953.
2. Cochran, W. G. y Cox, G. M.: *Experimental Designs*. 2a. ed. Nueva York. John Wiley & Sons, Inc. 1957.
3. Billingham, R. E.; Brent, L. y Medawar, P. B.: *Actively acquired tolerance of foreign cells*. Nature 172: 603, 1953.
4. Wasz-Höckert, O.; Wager, O.; Hautala, T. y Widholm, O.: *Transmission of antibodies from mother to fetus. A study of the diphtheria antitoxin level in the newborn with esophageal atresia*. Ann. Med. Exp. Fenn. 34: 444, 1956.
5. Cochran, W. G.: *Sampling Techniques*. Nueva York. John Wiley & Sons, Inc., 1953.
6. Cochran, W. G.: *Modern methods in the sampling of human populations. General principles in the selection of sample*. Am. J. Pub. Health. 41: 647, 1951.
7. Pearl, R.: *Cancer and tuberculosis*. Am. J. Hyg. 9: 97, 1929.
8. Cornfield, J.: *Modern methods in the sampling of human populations. The determination of sample size*. Am. J. Pub. Health. 41: 634, 1951.
9. Modell, W. y Houde, R. W.: *Factors influencing clinical evaluation of drugs. With special reference to the double-blind technique*. J.A.M.A. 167: 2190, 1958.
10. Moroney, M. J.: *Facts from Figures*. 3a. edición. London, Penguin Books, 1956.
11. Gregg, N. M.: *Congenital cataract following german measles in the mother*. Trans. ophthal. Soc. Aust. 3: 35, 1941.
12. Brown, C. M. y Nathan, B. J.: *Maternal rubella and congenital defects*. Lancet 1: 975, 1954.
13. McDonald, A. D.: *Maternal health and congenital defect: a prospective investigation*. New Eng. J. Med. 258: 767, 1958.

14. Doll, R. y Hill, A. B.: *Smoking and carcinoma of the lung*. Brit. med. J. 2: 739, 1950.
15. Doll, R. y Hill, A. B.: *Lung cancer and other causes of death in relation to smoking*. Brit. med. J. 2: 1071, 1956.
16. Ostle, B.: *Statistics in Research. Basic concepts and Techniques for Research Workers*. Ames, The Iowa State College Press, 1954.
17. Boyle, R.: *Concerning the unsuccessfulness of experiments*. En: Fisher, R. A.: *The Design of Experiments*. 6a. edición. Edinburgh. Oliver and Boyd, 1951.

IV

ALGUNAS CONSIDERACIONES SOBRE EL ANALISIS ESTADISTICO EN LA INVESTIGACION CLINICA

JOAQUÍN CRAVIOTO¹

UNA INVESTIGACIÓN bien planificada debe incluir siempre en su diseño referencias claras y precisas acerca de las técnicas estadísticas que se van a utilizar en el análisis de los datos. Con demasiada frecuencia se recolecta información de manera tan errónea, que no existe operación estadística adecuada para manejar esos datos y poder ayudar a encontrar una respuesta valedera al problema planteado en la investigación. El análisis estadístico no es otra cosa que el procedimiento objetivo por medio del cual se puede llegar a la decisión de aceptar o rechazar un conjunto de datos como confirmatorios de una hipótesis, conociendo el riesgo que se corre al tomar tal decisión.

En los últimos 50 años el desarrollo de pruebas estadísticas se ha incrementado a tal grado que en la actualidad se cuenta con varias pruebas alternativas que pueden usarse para casi cada

uno de los diseños experimentales y el investigador se encuentra ante el problema de cómo seleccionar cuál de todas es la más apropiada y más económica para las preguntas que a través de la investigación desea contestar. Ante esta situación es necesario tener una base racional sobre la cual seleccionar la prueba más apropiada. Esta selección constituye el punto crítico del análisis estadístico.

En la selección de una prueba estadística deben emplearse los siguientes criterios: 1) El tipo de escala en que se expresan los datos de la investigación; 2) La potencia de la prueba; 3) La aplicabilidad del modelo estadístico en que se basa la prueba, a los datos particulares de la investigación; y 4) La potencia-eficiencia de la prueba.

I. TIPO DE MEDICIONES

Es importante tener siempre presente que las operaciones de cálculo aritmé-

¹ Académico numerario. Hospital Infantil de México.

tico inherentes a cada prueba estadística pueden aplicarse a cualquier conjunto de números independientemente de la manera como estos hayan sido obtenidos, pero la validez del método depende primariamente del modo de recolección de las observaciones; al seleccionar una prueba estadística es por lo tanto indispensable preguntarse en primer lugar, con qué tipo de mediciones se va a contar. En este sentido medir puede definirse en el terreno de la investigación, como la asignación de números a objetos y eventos de acuerdo a reglas aceptables de lógica. El sistema numérico es una creación altamente lógica ofreciendo múltiples posibilidades para manipulaciones también de carácter lógico. Si podemos de manera legítima asignar números al describir características, objetos y eventos podemos operar con esos números en todos sus modos permisibles y derivar de esas operaciones conclusiones aplicables a los fenómenos observados y medidos. Estamos justificados a describir cosas reales por medio de números siempre que exista un grado suficiente de isomorfismo (semejanza de propiedades) entre las cosas reales y el sistema numérico. Es decir, que ciertas propiedades de los números deben tener paralelismo con los fenómenos observados para que podamos confiadamente asignarles números.

Tres propiedades fundamentales de los números permiten su aplicación en el campo de la investigación científica: identidad, ordinalidad y aditividad.

Cada número sólo es igual a sí mismo, ningún otro es exactamente igual a él, es decir, posee identidad y por lo

tanto, a cualquier objeto o evento que pueda diferenciarse de los demás, que tenga identidad, podrá aplicársele un número. Además de identidad, los números poseen la propiedad de orden o rango, uno siempre mayor que otro que le precede en un continuum ascendente. Los objetos o eventos que son susceptibles de ordenamiento a lo largo de un continuum pueden recibir por lo tanto números. Finalmente, en el terreno de la investigación, es importante recordar que los números tienen la propiedad de ser aditivos, lo que quiere decir que la suma de un número con otro número debe dar de manera invariable otro número único. Es aparente de todo lo anterior que los fenómenos u objetos a los cuales deseamos aplicar números no necesitan poseer todas las propiedades de identidad, orden y aditividad para que legítimamente los podamos medir; dependiendo de cuáles propiedades estén presentes tendremos varios niveles de medida y por consiguiente también varios tipos de escalas.

TIPOS DE ESCALAS

Considerando la presencia únicamente del principio de identidad los objetos o eventos se pueden clasificar simplemente en categorías. A cada categoría se le asigna un número como etiqueta. Si dos observaciones son iguales en términos de lo que se está clasificando, ambas pertenecerán a la misma categoría y por consiguiente tendrán el mismo número. La asignación de números es absolutamente arbitraria pudiendo ser cualquiera del sistema numérico. Lo fundamental en este tipo de escalas llamadas nominales es que cada

una de las observaciones debe caer dentro de una categoría y sólo una, que las categorías deben ser mutuamente exclusivas y colectivamente exhaustivas. Esto quiere decir que una observación debe pertenecer siempre a la misma categoría, ninguna observación puede pertenecer a dos categorías y tampoco ninguna observación puede quedar fuera de la clasificación.

Se comprueba que siendo los números asignados a las categorías simplemente etiquetas para identificación, las clases a que corresponden no tienen orden, y no pueden efectuarse operaciones aritméticas con los números. Por ejemplo, si clasificamos a nuestros enfermos atendiendo a su lugar de origen, en urbanos y rurales y asignamos el número 1 a los urbanos y el 2 a los rurales, para así identificarlos, esto nunca querrá implicar que dos enfermos urbanos hacen uno rural.

Cuando además de la propiedad de identidad las observaciones que hacemos tienen orden a lo largo de un continuum podemos arreglarlas en clases o categorías y asignar a estas números utilizando su característica de orden; sin embargo, en estos casos no existe la libertad de hacer operaciones con esos números pues sólo indican situación a lo largo del continuum sin poder definir cuantitativamente qué tanto en más o menos una observación o grupo de observaciones está por arriba o por abajo de las incluidas en la categoría que le precede o sigue.

Conviene antes de seguir adelante hacer notar que el término escala está tomando en su sentido de "aquello que discrimina" y por esto podemos hablar

de escalas nominales donde no existe continuum de referencia. La discriminación o clasificación se aplica también a los otros tipos de escalas, como la ordinal que se acaba brevemente de describir, y a otras más refinadas en las cuales a la propiedad de identidad de las observaciones se añaden otras.

Las escalas ordinales son muy frecuentemente empleadas en la investigación clínica, donde el refinamiento cuantitativo muchas ocasiones no es posible. Por ejemplo cuando clasificamos una respuesta en pacientes en términos de: sin cambio, mejorados y curados, donde obviamente la suma de varios sin cambio no haría un mejorado, y donde 3 mejorados no darían un curado. Asimismo podemos hablar de clases sociales baja, media y alta sin que podamos sumar varias de la media para obtener uno de la alta. En conclusión podemos establecer que mientras que las escalas nominales sólo clasifican, las escalas ordinales clasifican y ordenan dando por resultado una serie de clases o categorías mutuamente exclusivas llamadas rangos.

Si se aplican escalas realmente métricas a las observaciones se obtienen categorías con diferencias absolutas en que los números que se asignan tienen las tres propiedades básicas: identidad, ordinalidad y aditividad. En términos operacionales las escalas a que dan lugar este tipo de mediciones clasifican, ordenan a lo largo de un continuum y producen intervalos iguales entre clases adyacentes. Los datos contenidos en estas escalas pueden legítimamente sujetarse a todas las operaciones adecuadas para números y las conclusiones

obtenidas de estas operaciones son valideras para las observaciones. Cuando la escala en que se arreglan estas mediciones carece de cero real se dice que se tiene una escala de intervalos.

Si las mediciones son métricas y además de todas las características de una escala de intervalo existe un cero verdadero como punto de origen se habla de escalas de tasa. En una escala de tasa la relación entre dos puntos de la escala es independiente de la unidad de medida. Por ejemplo si medimos masa en una escala de tasa, la escala de gramos tiene un cero verdadero y la relación entre dos pesos expresados en kilogramos es idéntica a la relación de los pesos tomados en gramos, o libras u onzas, o cualquier otra unidad de medida.

La escala de tasa es isomórfica a la estructura de la aritmética; por lo tanto las operaciones de aritmética son permitidas en los valores numéricos que se han asignado a los objetos mismos, así como a los intervalos entre los números. Estas escalas son más frecuentemente encontradas en las ciencias físicas y tienen como características particulares que la relación entre dos de sus valores no se altera cuando la escala se multiplica por una constante positiva. Todas las pruebas estadísticas son aplicables, en principio a estas escalas, pudiendo además emplear otras que como el promedio geométrico y el coeficiente de variación requieren el conocimiento del cero verdadero.

II. POTENCIA DE LA PRUEBA

La validez del análisis estadístico depende en mucho de la fuerza o poten-

cia de la prueba estadística empleada. Se acepta que una prueba estadística es potente cuando tiene una probabilidad muy pequeña de rechazar una hipótesis verdadera y una alta probabilidad de rechazar la hipótesis cuando ésta es falsa. En presencia de dos pruebas cuya probabilidad de rechazar hipótesis falsas sea igual, la selección en principio debe inclinarse hacia la prueba que tenga la probabilidad mayor de aceptar la hipótesis cuando sea verdadera.

III. MODELO ESTADÍSTICO

Aun cuando en principio es verdad el dicho de que todo análisis estadístico puede iluminar los datos recogidos en la investigación, es indispensable recordar que cada procedimiento tiene sus limitaciones. Toda prueba estadística es desarrollada como una idea matemática pura; por consiguiente toda prueba descansa en cierto número de premisas referentes a la población de donde proviene la muestra bajo estudio. Si esas premisas son verdaderas para los datos particulares bajo análisis el procedimiento será adecuado. Es decir, que al determinar la naturaleza de la población y por ende la manera propia de muestrearla se establece un modelo y una medición específicos. La prueba será aplicable sólo cuando satisfice ciertas condiciones establecidas como requisitos del modelo y la medida.

Algunas veces se está en capacidad de verificar si las condiciones de un modelo estadístico determinado se satisfacen, pero muy a menudo tenemos que suponer que sí se llenan sin poder demostrarlo. Por esta razón las con-

diciones del modelo estadístico de una prueba a menudo se denominan "presunciones o premisas". Todas las decisiones a las que se llega con el uso de una prueba estadística llevan implícita o explícitamente un condicional que establece: que si el modelo escogido fue el correcto y se satisfizo el requerimiento de medición entonces la conclusión es válida. Es obvio que mientras menor sea el número de presunciones que entren en el modelo, se necesitarán menos condicionales añadidos a la decisión. Esto indica que a menor número de premisas mayor generalidad de las conclusiones.

Las pruebas estadísticas pueden clasificarse en dos grandes grupos: pruebas paramétricas y pruebas no paramétricas. Una prueba paramétrica es aquella cuyo modelo especifica ciertas condiciones que debe tener la población de la cual se ha derivado la muestra bajo análisis, requiriéndose además que las observaciones puedan ser expresadas en escalas de tasa o de intervalo.

Las pruebas paramétricas más usadas en la investigación clínica son las pruebas de comparación de promedios y de análisis de varianzas. Cuando se llenan las premisas de los modelos base de estas pruebas puede afirmarse que no existen otras más potentes para efectuar un análisis adecuado. Sin embargo, estas pruebas tienen muchas y muy fuertes presunciones. Por ejemplo en la prueba "t", que compara promedios, se deben satisfacer las siguientes premisas:

1. Las observaciones deben ser inde-

pendientes. La selección de un caso, para ser incluido en la muestra, no debe prejuciar las probabilidades de selección de ningún otro caso de la población; asimismo el puntaje que se le dé a una observación no debe prejuciar el que se asignará a cualquier otra.

2. Las observaciones deben provenir de poblaciones cuya distribución sigue una curva normal.
3. Las poblaciones deben tener la misma varianza, aunque en casos especiales es suficiente con saber la tasa de sus varianzas.
4. Las variables consideradas en el estudio deben ser medidas por lo menos en escalas de intervalo, para que sea posible hacer operaciones aritméticas.

Las condiciones enumeradas, con excepción de la premisa de varianza semejante en las poblaciones, no se sujetan, por lo general, a prueba empírica en el curso del análisis estadístico. Por esta razón al ser aceptadas como supuestos, su falsedad o veracidad determina el significado real de las conclusiones a las que se llega mediante su empleo.

Cuando se tengan razones para creer que los supuestos señalados se han satisfecho en los datos bajo análisis, se debe escoger una prueba paramétrica. La selección de pruebas de "t" ó "F" es la óptima en estos casos porque constituyen el más poderoso instrumento que existe para rechazar la influencia del azar cuando éste no juega un papel realmente importante.

Aun cuando se tiene en la actualidad

cierta evidencia que demuestra que pequeñas desviaciones en la exactitud de las premisas, que forman la base de las pruebas paramétricas, pueden no tener efectos radicales sobre las conclusiones, no existe acuerdo general acerca de lo que constituye una desviación pequeña y a partir de qué magnitud la desviación ya no puede aceptarse como pequeña. Cuando la población no se distribuye en forma normal o cuando las mediciones no pueden expresarse en escalas de intervalo o tasa, o cuando las poblaciones no son iguales en varianza, es dudoso cualquier valor que se dé a las conclusiones derivadas de la aplicación de una prueba que exige precisamente esos requisitos en su modelo.

Cuando por cualquier razón no se pueden llenar los requisitos de las pruebas paramétricas, el investigador puede recurrir a las llamadas pruebas no paramétricas, que son aquellas cuyo modelo matemático no exige el conocimiento de los parámetros que tiene la población de la que proviene la muestra bajo estudio. Las premisas más importantes asociadas con estas pruebas son la independencia de las observaciones y la continuidad de las variables por analizar. Las pruebas no paramétricas por otra parte, no requieren que los datos sean estrictamente cuantitativos; escalas ordinales y aún escalas nominales son apropiadamente analizadas por estas técnicas.

En la investigación clínica, muy a menudo el tamaño de la muestra es muy pequeño, no siendo raro tener menos de 6 pacientes u observaciones. En estas circunstancias si la naturaleza de la población total (de casos u observa-

ciones) no es conocida de manera exacta no hay alternativa posible y debe emplearse una prueba no paramétrica. Tampoco existe alternativa en el caso de observaciones que sólo pueden clasificarse ya que no hay pruebas paramétricas aplicables a datos cualitativos.

Existen muchas ocasiones en que la investigación no puede obtener datos estrictamente cuantitativos, registrándose solamente gradaciones en las variables bajo estudio. Por ejemplo, al estudiar la gravedad de una enfermedad no mortal se puede medir si un enfermo está más grave que otro, sin poder precisar exactamente cuanto más grave. Estas observaciones se pueden catalogar solamente en rangos y su análisis realizarse por pruebas no paramétricas, ya que si se emplearan técnicas paramétricas las presunciones acerca de la verdadera distribución de las características en la población total serían irreales y llevarían por lo tanto a decisiones y conclusiones carentes de valor.

Es muy importante no olvidar que una ventaja de las pruebas no paramétricas es que en la mayoría de ellas se obtienen declaraciones de probabilidad exactas que son independientes de la forma en que se distribuyen las características en la población de la cual se derivan las muestras. Existen sin embargo, algunas pruebas estadísticas no paramétricas que suponen identidad en la forma de distribución de las poblaciones, y otras que demandan simetría.

IV. POTENCIA-EFICIENCIA

Como ya se mencionó con anterioridad, mientras menores en número son las presunciones en que se basa un mo-

delo estadístico, más generales son las conclusiones que se derivan de la aplicación de una prueba estadística asociada a ese modelo, pero también menos potente es la prueba para rechazar la influencia del azar cuando éste no juega un papel importante.

Esa aseveración es generalmente verdadera para muestras de un tamaño definido, pero puede carecer de veracidad al comparar dos pruebas estadísticas aplicadas a muestras de tamaños desiguales. Es decir, que si el tamaño de la muestra es de 30 por ejemplo en cada muestra, la prueba A podría ser más potente que la prueba B; pero la misma prueba B puede ser más potente con 30 casos que la prueba A con sólo 20. En otras palabras, podemos evitar el dilema de tener que escoger entre potencia y generalización seleccionando una prueba estadística que tenga amplia generalización y luego aumentar su potencia hasta que alcance la de la prueba más potente incrementando el tamaño de la muestra. El poder-eficiencia de una prueba define el aumento que debe hacerse en el tamaño de la muestra para hacer que una prueba menos potente alcance el poder de la más potente. Supongamos por ejemplo, que una prueba B requiere una muestra de 25 casos para tener la misma potencia que la prueba A con sólo 20 casos; el poder-eficiencia de B expresado como porcentaje de A será igual a 20 por 100 entre 25, lo que dará 80%. Esto significa que para igualar la potencia de las pruebas A y B se necesitan 10 casos para la prueba B por cada 8 que se hubieran requerido para la prueba A.

De este modo, se pueden escoger pruebas estadísticas con menor número de presunciones en su modelo y por consiguiente con mayor generalización que "t" y "F", y aumentando el tamaño de la muestra evitar el tener que satisfacer las condiciones de distribución normal, igualdad de varianza y efectos aditivos.

Cuando se satisfacen todas las presunciones contenidas en el modelo estadístico paramétrico y se tienen observaciones en escalas de intervalo o de tasa, se desperdician datos si se emplea una prueba no paramétrica. El grado de desperdicio se puede calcular del valor de poder-eficiencia que tiene la prueba no paramétrica. Por ejemplo, si una prueba no paramétrica tiene un poder-eficiencia de 80%, esto significaría que la prueba paramétrica correspondiente tendría el mismo grado de efectividad en una muestra 20% menor que la usada con la prueba no-paramétrica. Se comprende fácilmente que cuando se están colectando casos poco frecuentes, o se está en presencia de características con cambios estacionales el acortar el tamaño de la muestra puede ser muy importante pues ahorra tiempo y esfuerzo. Por otra parte, debe insistirse en que para emplear la prueba paramétrica y disminuir el tamaño de la muestra es necesario satisfacer todas las premisas del modelo base de la prueba.

En el momento actual un problema de análisis estadístico no totalmente resuelto por medio de métodos no paramétricos es la prueba de las interacciones de alto orden en el modelo de análisis de varianza, debido a las pre-

sunciones de propiedades aditivas que forzosamente tienen que hacerse en este análisis.

En resumen, la selección de una prueba estadística se basa en el tipo de datos con que se cuenta, el tamaño de la muestra por analizar, la satisfacción de los requisitos del modelo estadístico base de la prueba y la potencia de la misma.

Cuando se tienen escalas de intervalo o tasa y se pueden llenar las exigencias de modelos estadísticos paramétricos, las pruebas más potentes son sin duda alguna las paramétricas tipo "t" y "F". Sin embargo, la potencia de una prueba con menos requerimientos se puede incrementar mediante el aumento en el tamaño de la muestra.

Considerando las características de los problemas más frecuentes en la investigación clínica y la poca familiaridad del investigador clínico con pruebas no paramétricas, a juzgar por la revisión del tipo de pruebas que emplea al informar sobre sus hallazgos en las publicaciones médicas, a continuación se hace mención de algunas pruebas estadísticas útiles sobre todo cuando no se conocen los parámetros de las poblaciones bajo estudio, las muestras son pequeñas y no existen facilidades para la computación electrónica.

Las pruebas estadísticas, se han agrupado de acuerdo al diseño experimental para el que son apropiadas, en tres grandes grupos:

1) Pruebas que pueden emplearse cuando se desea conocer si una muestra pertenece a una población específica.

2) Pruebas que se usan cuando se quiere comparar los valores obtenidos en dos muestras relacionadas o independientes.

3) Pruebas que pueden emplearse cuando se desea comparar los valores obtenidos en 3 ó más muestras.

En la investigación científica en general, y en la investigación clínica en particular uno de los deseos del investigador es llegar a conclusiones aplicables a toda una población mediante el uso de la información contenida en una muestra estudiada. Por ejemplo, a través del estudio de un grupo de pacientes con hepatitis infecciosa llegar a describir el cuadro clínico de la hepatitis infecciosa, o mediante la respuesta terapéutica obtenida en un grupo de enfermos con absceso hepático amibiano encontrar la terapéutica aplicable a todos los casos de la misma afección, o midiendo el desarrollo mental de un grupo de niños con síndrome de Down, describir el desarrollo mental en las trisomías, etc., etc. Para poder llegar a conclusiones valederas para una población a partir del conocimiento de una muestra, es imperativo que dicha muestra sea una muestra al azar. Este problema es fundamental y a él hay que enfrentarse continuamente en el curso de toda investigación clínica. En la actualidad existen diversas técnicas que permiten llegar a saber la probabilidad de que la muestra en cuestión sea una muestra al azar. Estas técnicas se basan en el orden o secuencia en que fueron originalmente obtenidas las observaciones o datos individuales.

Una técnica relativamente sencilla propuesta por Swed, consiste en determinar el número de carreras que presenta la muestra. Una carrera es definida para este propósito, como una sucesión de símbolos idénticos que son seguidos y precedidos por símbolos diferentes o por ausencia de símbolo. En vista de que se conoce la distribución del número de carreras que es de esperarse al tomar repetidas muestras al azar de una misma población, se está en capacidad de decidir si una muestra específica tiene mayor o menor número de carreras de las que se presentarían si la muestra fuera al azar. Cuando se tienen puntajes o otros valores numéricos se pueden convertir éstos a signos de más y menos, por ejemplo de acuerdo a su situación con respecto a la mediana de la muestra, y con los signos establecer las carreras que se presentaron de acuerdo a la secuencia estricta en que fueron obteniéndose los valores individuales durante el proceso de recolección de la muestra.

En la investigación clínica no es raro que existan poblaciones formadas por sólo dos clases, por ejemplo, masculino y femenino, curado y no curado, hepatomegalia presente o ausente, febril y afebril, con efectos indeseables y sin efectos indeseables. En estos casos las observaciones caen de necesidad en alguna de las dos categorías. Si en estas condiciones conocemos la proporción de casos que caen en una clase, automáticamente conoceremos la proporción que hay en la otra clase restando de la unidad la proporción presente en la clase uno. Al obtener los datos empíricos en una investigación, simplemente por obra

del azar, las proporciones pueden numéricamente ser diferentes del valor teórico y la tarea del análisis estadístico es encontrar la probabilidad que tiene de ser verdad esta aseveración. La distribución de las proporciones que se presentan en muestras tomadas al azar de una población formada por dos clases es una distribución binominal que nos permite conocer los diversos valores que pudieron ocurrir debidos al azar, y de esta manera decidir si aceptamos que las proporciones o frecuencias observadas en la muestra investigada son diferentes de las que proporciona el azar.

Cuando los datos de la investigación son de naturaleza cualitativa y se reparten en más de dos clases dentro de una escala nominal, como por ejemplo al estudiar la frecuencia de una enfermedad en función de clase socioeconómica, o la frecuencia de un síntoma o una respuesta en función del lugar geográfico, etc., la prueba X^2 es adecuada para efectuar el análisis, pudiendo también emplearse en el caso de existir únicamente dos categorías en la escala. Esta prueba también pertenece al grupo de las de bondad del ajuste ya que estima la existencia o ausencia de una diferencia significativa entre la frecuencia observada en las diversas clases que entran en la escala y la frecuencia teórica que debería encontrarse si la repartición fuera producto del azar. Es necesario tener presente que esta prueba no debe emplearse cuando habiendo sólo dos clases en la escala la frecuencia teórica sea menor que cinco. En el caso de existir más de dos clases o categorías tampoco debe emplearse si cual-

quiera de las clases tiene una frecuencia teórica menor de uno, o si más del 20% de las frecuencias teóricas tienen valores inferiores a 5. En ocasiones se puede tratar de obviar esta dificultad combinando clases adyacentes, debiendo ser muy cuidadosos para efectuar sólo combinaciones que tengan significado; si a pesar de esta manipulación se llega a tener dos clases y una frecuencia teórica menor de 5, la prueba binomial deberá ser usada en lugar de la X^2 .

La prueba de Kolmogorov-Smirnov para una sola muestra, es una prueba de bondad de ajuste, es decir que mide el grado de acuerdo, existente entre la distribución de un conjunto de datos y una distribución teórica específica. Su objetivo es señalar si los datos provienen de una población que tiene precisamente la distribución teórica especificada. La prueba compara la distribución acumulada de las frecuencias teóricas con la distribución acumulada de las frecuencias observadas, encuentra el punto de divergencia máxima y determina qué probabilidades existen de que una divergencia de esa magnitud sea debida al azar. Cuando se trata de una variable continua y si la muestra es pequeña puede considerarse que la prueba de Kolmogorov-Smirnov es el instrumento de elección como prueba de bondad en el acuerdo.

Una segunda serie de situaciones en investigación clínica se plantean cuando el fenómeno a estudiar consiste en establecer si dos "tratamientos" son diferentes o si uno es mejor que otro. El concepto tratamiento dentro de este contexto es definido como la exposición natural o experimental de un grupo a

una situación, pudiendo ser ésta de una gran diversidad, por ejemplo, uso de un medicamento, presencia de patología, adiestramiento, aculturación, cambio climático, cambio topográfico, separación familiar, cambio económico, etc., etc. En todas estas situaciones el estudio comprende la comparación del grupo que sufrió el tratamiento contra el grupo que o bien no tuvo tratamiento o tuvo un tratamiento diferente del sufrido por el primer grupo. La idea base es tratar de eliminar circunstancias que no siendo parte del tratamiento influyen en realidad o en apariencia el resultado dando falsas positivas o falsas negativas. En estos casos, que estadísticamente corresponden al análisis de dos muestras relacionadas, el diseño de la investigación reclama el uso de controles apareados, que pueden ser dos grupos de individuos, o el mismo individuo antes y después del tratamiento.

En el primer caso un miembro de cada par de sujetos, idénticos en todas las características que pudieran influir en el resultado, se asigna al azar al grupo de tratamiento, y el otro miembro del par al otro tratamiento o al grupo control. En el segundo caso, se realizan las observaciones en los mismos sujetos estando constituido cada par por el individuo antes y después del tratamiento.

La técnica estadística más potente para analizar los datos provenientes de dos muestras relacionadas es la prueba de "t" aplicada a las diferencias obtenidas en cada par bajo estudio. La prueba "t" presupone que los valores de las diferencias se distribuyen de manera normal e independiente en la po-

blación a la que pertenece la muestra, requiriéndose además para su aplicación adecuada que la escala en que se expresan los datos sea de intervalo o de tasa.

Cuando los datos de la investigación dan solamente escalas nominales u ordinales, o las presunciones de distribución normal no se llenan, o las conclusiones se desea que puedan generalizarse a otro tipo de distribución, se pueden emplear pruebas de tipo no paramétrico de fácil ejecución. Entre las principales de éstas se encuentran la prueba de McNemar para estimar el significado de los cambios; la prueba de signos, la prueba de Wilcoxon de los pares y rangos con signo, y la prueba de aleatoriedad para casos apareados.

La prueba de McNemar para medir el significado del cambio es aplicable particularmente a aquellos diseños de pre y post-tratamiento en que cada individuo ha servido como su propio control y cuando los datos bajo estudio se expresan en escalas nominales u ordinales. Sabiendo que la distribución de las muestras empleando la fórmula de McNemar para ausencia real de cambio sigue muy de cerca la distribución de X^2 para un grado de libertad, se pueden conocer las frecuencias teóricas que son de esperarse por azar, especialmente si se emplea una corrección por continuidad. En estas circunstancias se puede estimar el grado de significación del efecto provocado por el tratamiento. Cuando la frecuencia teórica esperada es menor de 5, la prueba de McNemar no es adecuada, pudiendo utilizarse en su lugar la prueba binomial, donde el número de casos será

igual a la suma de aquellos que presenten cambio, en uno u otro sentido después del tratamiento, y la probabilidad de obtener x casos con cambio, es la menor de las frecuencias observadas.

En casos en que ha sido posible efectuar tanto prueba de "t", como prueba de McNemar, se ha podido demostrar que para muestras muy pequeñas (suma de frecuencias con cambio = 6), el poder-eficiencia de esta prueba es equivalente al 95% de "t"; para muestras mayores, el poder-eficiencia declina a medida que la frecuencia con que se observa cambio, aumenta hasta llegar a una asíntota de eficiencia de alrededor del 63%.

Cuando los datos obtenidos en cada par pueden ser clasificados en términos, de uno mayor que otro y la variable bajo estudio tiene continuidad, se puede emplear la prueba de signos. Esta prueba tiene un poder-eficiencia del 95% de la prueba "t", para muestras con un número de casos igual a 6, disminuyendo su eficiencia, a medida que el tamaño de la muestra aumenta hasta alcanzar un valor asíntótico de 63%.

En aquellas investigaciones en que los valores obtenidos en cada par de eventos pueden expresarse en términos de uno mayor que otro y además, las diferencias entre los diversos pares pueden también ser arreglados en orden de magnitud absoluta, la prueba de Wilcoxon de pares y rangos con signo es adecuada para el análisis. Esta prueba encuentra su mejor aplicación cuando los datos se obtienen en forma de puntajes, sin que exista seguridad de que los valores numéricos sean tan precisos como para asegurar que 2, sea el doble

de 1, pero sí para tener confianza en que una diferencia numérica mayor encontrada en un par con respecto a otro, indica una mayor diferencia; es decir que sin poder asegurar que las diferencias sean numéricamente exactas puede afirmarse que es totalmente adecuado arreglar las diferencias entre pares de acuerdo al rango dado por su magnitud absoluta. La distribución de la suma de los rangos afectados con el signo menos frecuente aún en muestras pequeñas es aproximadamente normal con promedio de cero, varianza la unidad. La prueba de Wilcoxon tiene para muestras pequeñas un poder-eficiencia del 95% de la prueba "t".

Cuando el investigador no tiene bases con qué afirmar que la población de donde sacó su muestra se distribuye de manera normal o que la varianza es homogénea y quiere obtener la probabilidad exacta de asociación de sus datos empíricos con el azar, la prueba de aleatoriedad para casos apareados es su mejor selección, siempre y cuando sus valores sean estrictamente numéricos.

Cuando el número de pares contenidos en el estudio es mayor de 12 el cálculo de esta prueba sólo puede lograrse de manera práctica por medio de un calculador electrónico. En ausencia de esta facilidad instrumental se puede substituir la prueba de aleatoriedad por la prueba de Wilcoxon de pares y rangos con signo, asignando a los rangos los valores numéricos. Esta alternativa es muy eficiente en virtud de que la prueba de Wilcoxon puede apropiadamente considerarse como una prueba de aleatoriedad de rangos.

Debido a que la prueba de aleatorie-

dad para casos apareados utiliza toda la información contenida en la muestra, se considera que su poder-eficiencia es del 100%, lo cual la clasifica como la más poderosa de las pruebas no paramétricas que pueden ser empleadas cuando los datos son expresados en escalas de intervalo.

El diseño de controles apareados frecuentemente no se puede llevar a la práctica por razones diversas y la investigación tiene que efectuarse en dos grupos independientes que pueden provenir de una o dos poblaciones estadísticas. En cualquiera de estas situaciones el tamaño de las muestras puede ser igual o diferente.

Para el análisis de datos provenientes de dos muestras independientes, la prueba de "t" aplicada a los promedios aritméticos constituye la técnica de elección. Sin embargo, como ya se dijo esta prueba demanda datos en escalas de intervalo o tasa y presupone el conocimiento de que los valores se distribuyen de manera normal en ambas poblaciones cuyas varianzas son iguales.

Cuando las premisas de "t" no se satisfacen, o los datos sólo pueden expresarse en escalas nominales u ordinales, o bien el investigador no desea sujetarse a las restricciones que le imponen las premisas contenidas en la prueba paramétrica, pues desea que sus conclusiones tengan mayor generalidad, se pueden usar pruebas no paramétricas, entre las cuales destacan la prueba de Fisher de probabilidades exactas, la X^2 para muestras independientes, la prueba de la mediana, la prueba U de Mann y Whitney, y la prueba de Kolmogorov-Smirnov.

Todas estas pruebas determinan la probabilidad de que dos muestras independientes pertenezcan o no a una misma población. Cada una sin embargo, tiene mayor o menor sensibilidad a cierta clase de diferencias entre las muestras.

La prueba de probabilidades exactas de Fisher en una prueba extremadamente útil para analizar muestras con un pequeño número de casos, conteniendo datos discretos arreglados en escalas nominales u ordinales de dos categorías mutuamente exclusivas. En investigaciones en que cada individuo no puede tener sino uno de los dos valores posibles, las frecuencias encontradas empíricamente son entradas en una tabla de contingencia de 2×2 en donde las filas representan los grupos independientes y las columnas el resultado del tratamiento. La probabilidad exacta de observar un conjunto de frecuencias en estas condiciones cuando los totales marginales son fijos está dada por la distribución hipergeométrica, o sea por la razón entre el producto de los factoriales de los cuatro totales marginales y el producto de los factoriales de las frecuencias multiplicado por el factorial del número total de casos. Con la misma fórmula es posible calcular la probabilidad exacta de aparición de las frecuencias más extremas y de este modo aumentar la potencia de la prueba.

Cuando la variable bajo estudio es discreta y los resultados de la investigación se obtienen en frecuencias, si se quiere conocer el valor de las diferencias entre dos grupos independientes se puede emplear la prueba de X^2 . La hipótesis que en estos casos se plantea

generalmente establece que los dos grupos difieren respecto a alguna característica y por ende en la frecuencia relativa con que cada grupo contribuye a cada categoría. La prueba de la hipótesis se hace contando el número de casos que cada grupo tiene en las diversas categorías y comparando después las proporciones de un grupo con otro.

La prueba de X^2 requiere que las frecuencias esperadas no sean muy pequeñas. Si se trata de una tabla de cuatro casillas (2×2) se recomienda que si el total de casos independientes es mayor de 40 se debe usar una corrección para continuidad. Si el número de casos está entre 20 y 40 sólo puede emplearse X^2 si todas las frecuencias esperadas son iguales o mayores de 5. Si alguna de ellas es menor de 5 se puede emplear la prueba de Fisher. Cuando el número de casos es menor de 20 debe emplearse la prueba de Fisher ya que X^2 no es aplicable.

Cuando los valores de dos grupos independientes se expresan en escalas ordinales y se quiere saber si difieren en sus tendencias centrales se puede usar la prueba de la mediana, cuya técnica consiste en obtener la mediana de ambos grupos combinados y usando este valor como eje, distribuir los valores de los grupos a comparar de acuerdo a su posición relativa a la mediana total. Como se sabe que en el caso de que ambos grupos fueran muestras provenientes de poblaciones con la misma mediana, la distribución de los valores que estarían por arriba de la mediana combinada, está dada por la distribución hipergeométrica, el problema se transforma en el manejo de una tabla de 2×2 y las

consideraciones hechas a propósito de X^2 son enteramente aplicables a la prueba de las medianas.

Se ha podido demostrar que la prueba de la mediana aplicada a escalas ordinales tiene una potencia de cerca del 95% para muestras muy pequeñas (6 o menos). Esta eficiencia disminuye con el aumento en el tamaño de la muestra alcanzando un valor asintótico del 63%.

La prueba no paramétrica más poderosa que se tiene para saber si dos grupos independientes son o no muestras de una misma población, es la prueba U de Mann-Whitney. Se considera que esta prueba es la mejor alternativa de la prueba de "t" y alcanza su máxima utilización cuando los datos de la investigación pertenecen a escalas ordinales. La técnica para muestras pequeñas es extremadamente sencilla, existiendo tablas preparadas por Mann y Whitney que permiten encontrar la probabilidad de que la diferencia en las muestras sea debida al azar.

Se ha podido demostrar que al aumentar el tamaño de las muestras la distribución de los valores de U rápidamente se aproxima a la distribución normal. El cálculo de valores Z, que para todo propósito práctico se distribuye normalmente con promedio cero y varianza unidad, nos permite conocer la significación de los valores de U encontrados. Cuando al calcular U existan muchos empates en los rangos asignados, es conveniente emplear una corrección, puesto que la prueba supone una distribución continua de las variables analizadas. Si esta continuidad fuera medida con una precisión absoluta la

probabilidad de encontrar dos valores iguales sería cero, sin embargo con nuestros sistemas de medida la discriminación no es óptima y se presentan valores iguales. Existiendo técnicas sencillas para hacer esta corrección, la probabilidad asociada con los nuevos valores es ligeramente mayor en favor del azar.

Se ha podido demostrar que si se aplica la prueba de U a datos que puedan adecuadamente sujetarse a prueba "t", la eficiencia se acerca al 95% para pruebas de tamaño moderado. Al aumentar el número de casos la eficiencia aumenta también hasta un valor límite de 95.5%.

En el caso de muestras muy pequeñas una alternativa a la prueba U es la prueba de Kolmogorov-Smirnov, en la cual se compara la distribución de la frecuencia acumulada usando los mismos intervalos en ambas muestras. La máxima diferencia encontrada se compara con aquella que podría ser dada por el azar, variando un tanto la fórmula específica de acuerdo al número de casos incluidos en el estudio. La prueba permite conocer si el grupo experimental difiere del control y si está, ésta diferencia lo coloca por encima del control. Comparada con la prueba "t" el poder-eficiencia de la prueba de Kolmogorov-Smirnov es de alrededor del 96% para muestras pequeñas (40 casos o menos entre los dos grupos bajo análisis).

Existen ocasiones en la investigación clínica en las cuales se desea saber si tres o más grupos estudiados de manera simultánea tienen diferencias significativas; es decir, si las muestras pertenecen a una misma población o pro-

vienen de poblaciones idénticas. En estas condiciones es necesario emplear una prueba que indique si existen o no diferencias globales entre los grupos, antes de seleccionar un par de ellos en quienes analizar el significado que pueda tener la diferencia encontrada.

La técnica de comparar pares sucesivos de muestras hasta cubrir todas las posibilidades de combinación es inadecuada, por aumentar muy considerablemente el riesgo de rechazar erróneamente la hipótesis de ausencia de diferencias. Se ha documentado que si se tomaran, por ejemplo 5 muestras y el nivel de significación se estableciera al 5% para propósitos de rechazo de la hipótesis de nulidad al comparar dos muestras, el nivel real de significación que debería establecerse para no cambiar las probabilidades al calcular las 10 combinaciones que dan las 5 muestras subiría al 0.40. De manera que sólo se justifica comparar pares de muestras buscando significación cuando una prueba global ha rechazado previamente la hipótesis de nulidad.

La prueba paramétrica con que investigar si varias muestras pertenecen a la misma población es el análisis de varianza o prueba "F". Las premisas asociadas con el modelo estadístico en que se basa la prueba incluyen: la independencia de las observaciones, la distribución normal de las poblaciones, la igualdad de las varianzas y el carácter aditivo del efecto de los promedios. Además, como en el caso de toda prueba paramétrica, las observaciones deben ser expresadas adecuadamente en escalas de intervalo o tasa.

Cuando no se pueden llenar los requi-

sitos de "F", o las variables sólo pueden obtenerse en escalas ordinales o nominales, se pueden emplear ciertas pruebas no paramétricas de acuerdo a que el diseño experimental con que se trabaje tenga muestras de igual tamaño, apareadas de acuerdo a uno o varios criterios que puedan afectar los valores que adquieran las observaciones, o bien que el diseño comprenda tres o más muestras independientes, de tamaño diferente, tomadas al azar cada una de una población diferente.

Para el primer tipo de diseño se cuenta con dos pruebas no paramétricas, la prueba "Q" de Cochran y la prueba de Friedman de análisis de varianza por rangos.

La prueba "Q" de Cochran se utiliza cuando las mediciones forman una escala nominal, o una escala ordinal dicotomizada. Esta prueba es en realidad una extensión de la prueba de McNemer para dos muestras relacionadas; su objetivo consiste en analizar si tres o más conjuntos apareados de frecuencias o proporciones difieren significativamente entre sí. El apareamiento puede estar basado en ciertas características importantes de los diferentes sujetos, o en el hecho de que el mismo individuo haya sido utilizado bajo diferentes condiciones. Cochran ha demostrado que bajo la hipótesis de nulidad, "Q" se distribuye aproximadamente igual a X^2 con grados de libertad igual al número de grupos menos 1.

Como la prueba "Q" de Cochran, se usa en escalas nominales y no habiendo prueba paramétrica utilizable no es posible conocer su eficiencia comparativa. Su uso con datos de intervalo sería una

aplicación errónea que desperdiciaría la calidad de los datos.

Si los datos de la investigación se obtienen en una escala ordinal y se desea saber si las muestras no difieren entre sí por pertenecer todas a la misma población estadística, el análisis de varianza puede hacerse utilizando los rangos en una prueba desarrollada por Friedman para tres o más muestras apareadas.

La técnica requiere transformar en rangos los valores obtenidos para cada grupo de sujetos sometidos a las diferentes condiciones, e investigar si los rangos obtenidos en cada condición por los distintos grupos aparecen con frecuencias casi iguales. Si las frecuencias son diferentes en las distintas condiciones se puede establecer una dependencia entre puntajes y condiciones. La prueba específicamente determina si las sumas de rangos contenidos en las distintas condiciones difieren significativamente. Para muestras muy pequeñas existen tablas de probabilidades exactas, para muestras mayores la probabilidad se determina por referencia a la distribución de X^2 con grados de libertad igual a número de condiciones menos 1.

Aun cuando no se conoce mucho acerca de la potencia de esta prueba, Friedman en un estudio empírico comparó los resultados de 56 análisis independientes de datos adecuados para la prueba "F" y el análisis de varianza por rangos, encontrando un acuerdo muy grande entre las dos pruebas.

Cuando no se pueden usar en la investigación muestras apareadas y se desea analizar datos procedentes de tres o más

muestras independientes para saber si existen diferencias entre ellas que puedan ser atribuidas a efectos de los tratamientos instituidos, se pueden usar técnicas no paramétricas cuando la prueba "F" no puede legítimamente aplicarse por el tipo de datos con que se cuenta o porque no se satisfacen las premisas del modelo estadístico base de la prueba.

Si se dispone de frecuencias en categorías discretas ya sean estas de tipo ordinal o nominal se puede determinar el significado de la diferencia empleando X^2 para más de 2 muestras independientes. Esta prueba es una simple extensión de la X^2 para dos muestras independientes, teniendo su empleo las mismas restricciones.

Si la variable bajo estudio ha sido medida cuando menos en una escala ordinal, la prueba a utilizar es la extensión de la prueba de la mediana, con la cual se puede determinar si tres o más grupos (no necesariamente del mismo tamaño) forman parte de la misma población estadística o provienen de poblaciones que tienen medianas iguales.

La prueba toma como punto de partida el cálculo de una mediana común a todos los grupos, después categoriza con signos de más y menos los datos en relación a esta mediana, las frecuencias resultantes las arregla en una tabla de 2 filas x K columnas (donde las filas representan los grupos por arriba y abajo de la mediana y las columnas las diversas condiciones) y calcula con estos datos una X^2 de grados de libertad iguales a $K-1$.

Otra prueba no paramétrica con la

cual se hace análisis de varianza para tres o más muestras independientes expresadas en escala ordinal es la prueba de Kruskal-Wallis, cuyo modelo estadístico requiere únicamente que la variable bajo estudio tenga una distribución continua. Esta prueba convierte los valores asignados a la variable en rangos, suma los rangos en cada muestra y determina por medio de tablas específicas la probabilidad de que esas sumas vengan o no de una misma población.

La prueba de Kruskal y Wallis comparada con la prueba "F" considerada como la más poderosa prueba paramétrica para efectuar análisis de varianza ha mostrado una potencia-eficiencia del 95.5%.

Parecería conveniente, para finalizar estas notas sobre el análisis estadístico en la investigación clínica, señalar un modelo fácil de seguir al tratar de establecer la confiabilidad que debe darse a los datos recogidos en una muestra como confirmatorios o no de una hipótesis bajo estudio. Este modelo, recomendado por muchos estadísticos e investigadores, permite fácilmente darse cuenta del lugar que ocupa dentro del diseño de una investigación el análisis estadístico.

1. *Declaración de la hipótesis que se desea analizar (H_1).*

El punto clave es asegurarse que la declaración es precisa, tan completa como sea posible y especificando la dirección que se espera o la ausencia de esta. Este último punto es fundamental para decidir si la prueba es de una o dos colas.

2. *Declaración de la hipótesis de nulidad (H_0).*

Esta hipótesis simplemente establece la ausencia de diferencia y se declara para percibir claramente que esta es precisamente la hipótesis que se sujeta a la prueba estadística.

3. *Selección de una prueba estadística.*

Esta selección debe hacerse en base a: 1) el tipo de escala en que estén las mediciones; 2) la potencia de la prueba; 3) la aplicabilidad del modelo estadístico a la muestra estudiada, y 4) el poder-eficiencia de pruebas alternas.

4. *Declaración acerca del tamaño de la muestra (N) y del nivel de significancia (α).*

α = probabilidad de rechazar H_0 erróneamente.

β = probabilidad de aceptar H_0 erróneamente.

5. *Encontrar la distribución teórica de la estadística a comparar.*

Para la mayoría de las estadísticas que se comparan existen tablas.

6. *Definir la región de rechazo.*

Conjunto de valores posibles tan extremos que si H_0 es verdadera la probabilidad de observar uno de ellos es muy pequeña (α).

Si H_1 predice dirección usar una cola.

7. *Hacer el cálculo y la decisión.*

H_0 se acepta o se rechaza.

Probablemente de todos estos pasos el menos importante es el referente a la ejecución de los cálculos aritméticos en la fórmula matemática de la prueba seleccionada. El investigador científico no debe olvidar que el valor de una prueba estadística no está dado por la magnitud del valor de significación encontrado al despejar una fórmula sino por la manera correcta de haberla se-

leccionado, ya que el valor numérico puede ser alto sin que la prueba sea apropiada, pues los cálculos aritméticos se aplican a cualquier serie de números siendo el resultado independiente de lo adecuado de la prueba empleada.

Para terminar debe decirse que el análisis estadístico no es un procedimiento que genere datos y que lo único que sí hace es precisar hasta donde las series numéricas que se le entregan se rechazan o aceptan como confirmatorias de una hipótesis, no siendo capaz de discriminar si los datos son o no adecuados al problema bajo estudio.

REFERENCIAS

1. Arkin, H. y Colton, R. R.: *Statistical Methods*, 4a. ed. College Outline Series, New York, Barnes y Noble, Inc. 1950.
2. Bowker, A. H.: *A test for symmetry in contingency tables*. J. Am. Statist. Ass. 43: 572, 1948.
3. Cochran, W. G.: *Some methods for strengthening the common X^2 tests*. Biometrics 10: 417, 1954.
4. Dixon, W. J. y Massey, F. G.: *Introduction to Statistical Analysis*. (Capítulo 13: *Enumeration Statistics*, y Capítulo 17: *Nonparametric Statistics*). New York, McGraw-Hill, 1951.
5. Domínguez Toledano, A.: *Bioestadística*. Serie Científica. México, Museo Nacional de Antropología, 1953.
6. Fisher, R. A.: *The Design of Experiments*. Edinburgh, Oliver and Boyd, 1953.
7. Freeman, L. C.: *Elementary Applied Statistics*. New York. John Wiley and Sons, 1965.
8. Freund, J. E.: *Modern Elementary Statistics*. New York, Prentice-Hall, 1952, Capítulo 11.
9. Garret, H. E. y Woodworth, R. S.: *Statistics in Psychology and Education*, 5a. ed. New York, David McKay Co. 1962.
10. Goodman, L. A.: *Kolmogorov-Smirnov tests for psychological research*. Psychol. Bull. 51: 160, 1954.
11. Guilford, J. P.: *Fundamental Statistics in Psychology and Education*. New York, McGraw-Hill. 1965.
12. Hays, W. L.: *Statistics*. New York, Holt, Rinehart and Winston. 1963.
13. Lewis, D. y Burke, C. J.: *The use and misuse of the chi-square test*. Psychol. Bull. 46: 433, 1953.
14. Mainland, D.: *Elementary Medical Statistics: The Principles of Quantitative Medicine*. Philadelphia, W. B. Saunders. 1952.
15. Mann, H. B. y Whitney, D. : *On a test of whether one of two random variables is stochastically larger than the other*. Ann. Math. Statist. 18: 52, 1947.
16. Massey, F. J. Jr.: *The Kolmogorov-Smirnov test for goodness of fit*. J. Am. Statist. Ass. 46: 68, 1951.
17. Mood, M. A.: *Introduction to the Theory of Statistics*. New York, McGraw-Hill. 1950.
18. Moore, G. H. y Wallis, W. A.: *Time series significance tests based on signs of differences*. J. Am. Statist. Ass. 38: 153, 1943.
19. McNemar, Q.: *Note on the sampling error of the difference between correlated proportions or percentages*. Psychometrika 12: 153, 1947.
20. McNemar, Q.: *Psychological Statistics*, 2a. ed. New York, Wiley. 1955.
21. Moses, L. E.: *Nonparametric statistics for psychological research*. Psychol. Bull. 49: 122, 1952.
22. Swed, F. S. y Eisenhart, C.: *Tables for testing randomness of grouping in a sequence of alternatives*. Ann. Math. Statist. 14: 66, 1943.
23. Wilcoxon, F.: *Probability tables for individual comparisons by ranking methods*. Biometrics 3: 119, 1947.

V

LA INTERPRETACION LOGICA

HORACIO JINICH¹

I. CONSIDERACIONES GENERALES

HACE POCOS años una prestigiada asociación médica norteamericana celebró su convención anual en la ciudad de Chicago. En la primera sesión, varios miembros distinguidos, profesores universitarios e investigadores de reconocido prestigio todos ellos, presentaron sus trabajos más recientes de investigación clínica. Algunos eran estudios experimentales de nuevos agentes terapéuticos, otros eran estudios genéticos, epidemiológicos y estadísticos. Los comentarios del auditorio fueron, en su gran mayoría, elogiosos y nadie puso en tela de juicio las conclusiones de los autores. A continuación hicieron uso de la palabra, uno tras otro, una serie de invitados a la sesión: uno experto en bioestadística, otro experto en diseño experimental, geneticista el tercero y otros más, profesores de lógica y filosofía de la ciencia. Todos ellos, sin excepción, hicieron trizas los trabajos científicos recién presentados, demostrando, sin piedad alguna, que adolecían de graves defectos, ora en la planeación

del experimento, ora en el cálculo estadístico, y casi todos en la interpretación lógica a tal grado, que no era difícil, a partir de estos mismos trabajos, llegar a veces a conclusiones exactamente contrarias a las alcanzadas por los autores. Después se supo que todo había sido una confabulación, que los autores habían inventado sus presentaciones, con objeto, precisamente, de ilustrar de manera dramática, la importancia de la metodología científica rigurosa en la investigación clínica.

Muchos otros estudios¹⁻³ confirman este hecho: que los clínicos realizan constantemente experimentos que carecen de las cualidades científicas esenciales: evidencia valedera, análisis lógico, demostración. Como dice Fainstein,⁴ "en una era de calculadoras y computadoras, muchos pronósticos se establecen en base a generalidades estadísticas vagas, a menudo inaplicables individualmente y erróneas. En una época de drogas potentes y procedimientos quirúrgicos formidables, los efectos exactos de muchos procedimientos terapéuticos son dudosos y sujetos a controversia. Su fundamentación consiste a menudo en datos de 'experiencia',

¹ Académico numerario. Hospital de Enfermedades de la Nutrición.

desprovistos de cuantificación, o en estadísticas grandiosas cuyas formulaciones matemáticas son de tal ingenuidad clínica, que su significado es puramente numérico, no biológico”.

Este estado de cosas no tiene justificación alguna. Es cierto que durante mucho tiempo sorprendió la complejidad del fenómeno vital y que, desconcertados por él, muchos investigadores desesperaban de la posibilidad de hallar leyes en biología y acostumbraban, por eso, abrir un abismo entre las ciencias de la materia inerte, que se plegaban con facilidad a la medición y a la sistematización, y la ciencia de la materia viva. Por fortuna ya no hay necesidad de defender el carácter científico de la biología. Sus logros, pasados y futuros, se obtienen gracias a la aplicación rigurosa de los métodos de la ciencia: planteada la hipótesis se planea con cuidado el experimento, se realiza, se observan y se miden con precisión los resultados, se tiene especial cuidado en la traducción de los fenómenos observados a palabras y juicios, se define con exactitud, se valoran los resultados de acuerdo con las leyes estrictas del cálculo estadístico y, finalmente, se obedecen escrupulosamente las leyes de la lógica en la interpretación de los hechos.

Con justicia se ha insistido mucho en la metodología del diseño experimental y en las reglas del análisis estadístico de los datos. No se ha insistido, sin embargo, con igual énfasis, en la importancia de los demás elementos de la investigación científica, especialmente la interpretación lógica de los resultados, sin la cual ningún experi-

mento, por impresionante que sea su técnica, y ningún cómputo, por elaborado que parezca, otorgan a las conclusiones legitimidad alguna. Con razón ha dicho Bradford Hill⁵ después de enumerar una serie de faltas de este tipo: “Los errores... (mencionados)... no se deben a ausencia de conocimiento de métodos estadísticos especializados o de adiestramiento matemático y sí, habitualmente, a la tendencia de los investigadores a aceptar las cifras sin discusión, sin preguntarse en cada oportunidad: ¿Qué hay detrás de estas cifras? ¿Qué factores pueden ser responsables de este valor? ¿De qué manera pueden haber surgido estas diferencias? Allí está siempre el núcleo de la cuestión”.

II. EL RAZONAMIENTO

La interpretación lógica de toda investigación científica consiste, en último análisis, en la aplicación de las leyes del razonamiento, que son del dominio de la lógica.

El conocimiento científico excepcionalmente es dado por actos de aprehensión inmediata. Casi siempre es el resultado de procesos más o menos complejos entre los que destacan los razonamientos deductivos, los analógicos y los inductivos. Los primeros reciben su validez de los grandes principios lógicos y en ellos se va de lo general a lo particular mediante leyes lógicas perfectamente establecidas desde la remota antigüedad. Entre las variaciones del proceso demostrativo científico, basado en la deducción, la lógica tradicional distingue la demostración

directa y la indirecta, la progresiva y la regresiva.⁶ Es *directa* cuando la verdad de lo que se demuestra en una serie lógica ininterrumpida, se deduce de la verdad de sus fundamentos. Es *indirecta* cuando la verdad de la tesis se deduce de la falsedad suficientemente fundada de su opuesto contradictorio, método tradicionalmente conocido como "deductio ad absurdum". La demostración es *progresiva* cuando, como en la prueba directa, de la verdad de los principios se deduce la verdad de la conclusión y es *regresiva* cuando de la falsedad de la consecuencia se deduce la falsedad de los principios. Esta última forma de demostración es muy usada en la refutación científica. Como todo razonamiento deductivo parte de lo universal, su comienzo nunca puede ser un juicio inmediato de experiencia, pues no hay experiencia inmediata sino de lo singular. Si el comienzo del razonamiento deductivo es una afirmación sobre los hechos, tiene que ser un juicio universal sobre los hechos, y por lo tanto el resultado de una inducción.

En el *razonamiento analógico* o por *analogía*, de que un objeto A' coincide con otro objeto A'' en ciertas notas, a, b, c, que son comunes a ambos, se concluye que A'' poseerá también la nota P que sabemos posee A'. Es un razonamiento de lo particular a lo particular análogo o de lo singular a lo singular análogo. En investigación clínica usamos a menudo este tipo de razonamiento cuando transferimos nuestras observaciones hechas en un hombre a otro hombre, o de un animal o máquina al hombre y viceversa. Por su

forma, el razonamiento analógico revisa la apariencia de un silogismo, pero en realidad no posee su fuerza y difiere fundamentalmente de él. El silogismo debe su rigor a que se atiene a las leyes ideales evidentes que rigen los pensamientos; en cambio en el razonamiento analógico se admite que a ciertas correspondencias entre dos objetos deben seguir otras, lo cual, si bien goza de cierta verosimilitud, carece en absoluto de seguridad; por este motivo el razonamiento analógico nunca termina en una resuelta afirmación. En el uso científico el razonamiento por analogía tiene dos papeles: o se aplica por sí cuando otro razonamiento no es posible, o se toman sus conclusiones como hipótesis, como datos verosímiles que hay que comprobar.

El procedimiento lógico por excelencia para establecer las leyes naturales es la *inducción*. No nos detendremos a analizar su fundamento, que según la opinión generalmente admitida consiste en "la regularidad del curso de la Naturaleza", regularidad que para unos es mera hipótesis (por muy elevado que sea su grado de probabilidad) mientras que para otros es una verdad segura. En general la inducción consiste en la averiguación mediante el análisis de los hechos, de las leyes que los rigen. Como estas leyes no se contentan con expresar relaciones entre los hechos observados sino que aspiran a una significación general, a valer para todos los hechos de una misma clase, hay en la inducción una comprobación y una extensión: comprobación de la relación en los casos examinados y extensión a todos los casos del mismo orden. Clá-

sicamente, la inducción se basa en varios métodos cuyo propósito consiste en aislar, entre las circunstancias que rodean a un fenómeno las que tienen con él una relación constante: *el método de concordancia*, *el método de diferencia* y el de *variaciones concomitantes*. En el método de concordancia, dado el fenómeno que se quiere investigar, se reúne el mayor número posible de casos en que aparezca, procurando que estos casos sean de la índole más diversa. Si se busca, por ejemplo, el antecedente de un fenómeno, se reúnen muchos casos en que ocurra ese fenómeno, muy diferentes en sí pero coincidentes en presentar el fenómeno en cuestión; el antecedente debe estar también presente en todos los casos, y si no lo está no sería el antecedente, pues el fenómeno se daría en este caso sin él. El método de concordancia supone un análisis eliminatorio: se eliminan las circunstancias que aparecen en las cosas examinadas sin darse en todas ellas. Ninguna de ellas puede ser el antecedente (o consecuente) del fenómeno que estudiamos, puesto que falta en algunos casos en que el fenómeno aparece. A medida que es mayor el número de casos examinados, la eliminación de circunstancias accidentales es mayor y crece por tanto la probabilidad de dar con el antecedente o consecuente verdadero. Si el número de casos analizados es pequeño, el término buscado puede ir por casualidad acompañado en todos los casos de circunstancias accidentales, que erróneamente se pueden considerar ligadas constantemente al fenómeno, y formando parte del término que se busca. Re-

cuérdese por ejemplo cómo los primeros casos de enfermedad de Bürger correspondieron a judíos, por lo que erróneamente se consideró que la enfermedad era privativa de ese grupo étnico.

En el método de diferencia se reúnen casos lo más parecidos posibles, pero en los cuales en unos aparezca el fenómeno estudiado y en otros falte. En los primeros se dará también lo que esté ligado constantemente al fenómeno mientras que estará ausente en el otro caso. Los ensayos terapéuticos con grupos testigos se basan, precisamente, en este método de la inducción.

Finalmente en el método de variaciones concomitantes se introduce la magnitud en la relación; se comparan casos bastante numerosos en los cuales el hecho en indagación se presenta en grados distintos y se busca la circunstancia que varía correlativamente con el hecho. Este método es frecuentemente usado en estudios epidemiológicos; por ejemplo para valorar el papel patógeno de los diversos contaminantes atmosféricos se intentan establecer relaciones entre las variaciones en su concentración, y la magnitud de diversos estados patológicos.

Establecidos, a través de estos métodos, juicios particulares, la inducción plantea sus conclusiones en forma de juicios universales. Aunque enunciados en sentido *asertórico*, los juicios universales de la inducción son todos *problemáticos*, aunque de una probabilidad que llega a ser a veces elevadísima. Se les llama, por ello "hipótesis probabilísticas" y en relación con ellos⁷ es necesario determinar qué grado de dife-

rencia existente entre las frecuencias observadas en el experimento y la probabilidad enunciada por la hipótesis debe servir de base para rechazar a esta última. Igualmente es preciso determinar qué tan estrecho debe ser el acuerdo entre las frecuencias observadas y la probabilidad hipotética, para que la hipótesis sea aceptada. Los requisitos pueden hacerse más o menos estrictos, de acuerdo con los objetivos de la investigación, y existen siempre dos tipos de errores en los que es posible caer: el de rechazar la hipótesis, aunque sea cierta, y el de aceptarla, siendo falsa.

III. LA OBSERVACIÓN

Sepúlveda se ha referido con gran profundidad, al punto de partida de la investigación científica: la hipótesis. Planteada ésta, y diseñado y ejecutado el estudio, el investigador observa los fenómenos. Consiste la *observación* en la captación del dato, en la recepción consciente de los elementos del saber. El hombre de ciencia se pone ante la realidad que suscita su interés y comprueba su modo de ser y su comportamiento, examina sus aspectos, los registra, reitera su examen. El resultado tiene alcance meramente individual y descriptivo: la observación se refiere a objetos singulares y termina con la enunciación de lo percibido en ellos. Tal enunciación no puede consistir sino en juicios de experiencia, singulares y, de ordinario, *asertóricos*. Cuando la observación está dificultada por impedimentos que nos hagan dudar del resultado, el juicio será *problemático*. La primera garantía de la fidelidad de

una observación es la ausencia de prejuicios; si existe alguno en la mente del observador, bajo la forma de una hipótesis, en el momento de observar hay que restarle cualquier influjo sobre la observación, evitando el conato de que lo prejuzgado cubra o reemplace lo verdaderamente observado. De aquí la importancia, en investigación clínica, de las llamadas pruebas ciegas y doble ciegas.

El segundo requisito de la observación es aprehender de la manera más exacta y completa lo que la observación nos muestra. Esta condición, que Kelvin hizo más estricta aún, al exigir la medición como prerequisite de la ciencia, ha obligado a los biólogos a intentar la medición dimensional de sus observaciones, y ha convencido a no pocos investigadores y filósofos de la ciencia de que, por este motivo, la labor del clínico nunca podrá adquirir rango científico. Esta actitud pesimista carece de fundamento; la investigación clínica puede tener carácter científico estricto; el clínico, limitado ciertamente en sus posibilidades de llevar a cabo mediciones dimensionales, puede, sin embargo, *contar* donde el número representa la *enumeración* de un grupo de entidades que han sido categorizadas en unidades; puede usar las matemáticas de la teoría de las series y el álgebra de Boole; puede, en fin, auxiliarse con los diagramas de Venn⁸ y satisfacer así las exigencias expresadas por Kelvin.

IV. EL LENGUAJE

Realizada la observación, captados los datos, surge uno de los problemas más graves de la ciencia: la conver-

sión de las vivencias a palabras, de los objetos a símbolos, de los sucesos a juicios. Torpe y limitado es el lenguaje humano, ante la infinita gama de imágenes que surgen como resultado del choque de los objetos y fenómenos con nuestros órganos de los sentidos. Y no siempre sabe el investigador escoger con tino las palabras más adecuadas ni definir con máxima precisión sus conceptos. El tiempo no nos permite discutir los apasionantes problemas planteados por las relaciones entre la semántica y la clínica. Basta señalar que a menudo se peca en este terreno fundamental. El lenguaje clínico es particularmente deficiente en estos aspectos. La nomenclatura y la taxonomía de las enfermedades son caóticas, no se hace la debida distinción entre enfermedad y enfermo ni se separan como es debido las lesiones anatómicas de las perturbaciones funcionales.

La definición de los términos empleados es fundamental para el investigador científico, quien no debe omitirla nunca. En ocasiones podrá utilizar definiciones "descriptivas" o "analíticas", o empleará definiciones "estipulativas", las cuales sirven para introducir una expresión que se va a utilizar en un sentido específico, en el contexto de una discusión o teoría. Así, por ejemplo, podrá el investigador definir ácido diciendo: "por ácido vamos a entender todo electrolito que da iones de hidrógeno" o "por bilirrubina directa vamos a entender bilirrubina combinada con dos moléculas de ácido glucurónico". Mejor aún es que el científico utilice las "definiciones operacionales", pro-

puestas por P. W. Bridgman.⁹ Consiste este tipo de definición en puntualizar, para cada término científico, la prueba, el procedimiento o la operación que deba practicarse para proporcionar un criterio para su aplicación. Así, por ejemplo, la definición operacional de ácido podría ser: "si se introduce una tira de papel tornasol azul, el líquido es ácido, sí, y sólo sí, el papel cambia de color a rojo". Y el término bilirrubina directa quedaría definido "operacionalmente" mediante la descripción de la prueba de laboratorio empleada y propuesta por el investigador para la medición de este pigmento. Resulta obvio que la "operación" mencionada en la definición deberá escogerse y describirse de tal manera que pueda ser ejecutada en forma inequívoca por cualquier observador competente y que el resultado pueda ser constatado de manera objetiva y no dependa esencialmente de quien ejecute la prueba. No escapa a nadie que mediante la adopción de criterios operacionales inequívocos aplicados a todos los términos científicos se asegura la posibilidad de repetir, de reproducir de manera objetiva todas las afirmaciones científicas.

No conviene terminar el análisis de la definición científica sin recordar las reglas prácticas clásicas¹⁰ para evitar las fallas más graves. Así, la definición no debe ser ni demasiado amplia ni demasiado restringida; no debe contener elementos superfluos; no debe ser tautológica, o sea, una mera repetición de lo definido; no debe ser negativa si puede ser positiva; y lo definido no debe estar supuesto o incluido en la definición.

V. VALORACIÓN ESTADÍSTICA

No escapa a nadie la importancia del análisis estadístico en la investigación biológica humana. Sólo por medio de él puede el científico, frente a la complejidad de los fenómenos que ocurren en los seres vivos, distinguir lo casual y azaroso de lo concatenado por relaciones causales. El Dr. Cravioto se ha referido con suficiente profundidad a este aspecto de la metodología científica. Pero conviene subrayar, una vez más, que el análisis estadístico, como tantos otros instrumentos poderosos, se usa a menudo en forma inapropiada y errónea. El análisis estadístico de los datos biológicos es de poco valor, y puede incluso ser un obstáculo, si el investigador, preocupado por los refinamientos matemáticos, ignora el origen biológico y el sentido de las cifras que maneja, o sí, obsesionado por su especulación, ignora o distorsiona los datos numéricos. En Medicina, un solo caso clínico, bien estudiado, puede ser suficiente para establecer o refutar, un concepto clínico importante, aunque desde el punto de vista estadístico carezca de valor. Bastó por ejemplo, un solo caso clínico, para probar que la insulina podía salvar la vida de un paciente en coma diabético. Por lo contrario, muchas investigaciones repletas de distinciones numéricas pueden tener significación estadística y ser al mismo tiempo insensatas desde el punto de vista clínico y terapéutico. A menudo se olvida que el análisis estadístico, a pesar de su potencialidad, tiene muchas limitaciones. Permite demostrar la precisión de los datos, pero no puede veri-

ficar su fidelidad; permite indicar las variaciones en una muestra, pero no puede determinar con qué seguridad dicha muestra es representativa del "universo" sujeto a estudio; permite definir que cierta relación probablemente tiene una causa específica, pero no nos señala la causa ni presta apoyo a lo que se decida designar como causa.

El empleo del análisis estadístico en la investigación biológica exige que el número de casos estudiados sea suficientemente grande para permitir la aplicación de las fórmulas apropiadas. Con este objeto, el investigador de laboratorio puede repetir su experimento tantas veces como sea necesario. Su material de experimentación es reproducible y sus aparatos de observación y medición están calibrados. De este modo le es fácil decidir cuando las relaciones estadísticas significativas se deben a sus manipulaciones experimentales y no a variaciones en el material o en el equipo de observación. Pero en la clínica humana las cosas ocurren de manera diferente: la limitación de casos clínicos obliga a menudo a combinar las observaciones obtenidas de muchos pacientes, estudiadas en diferentes tiempos y lugares por observadores diversos. De esta manera, es fácil entender como pueden resultar relaciones estadísticamente significativas pero que no se deben a la terapia misma sino a variaciones no identificadas o reconocidas en el material clínico o en el examinador clínico.

VI. LA CAUSALIDAD

El objetivo fundamental del conocimiento científico es entender los suce-

sos, los acontecimientos, es decir, conocer sus relaciones causales. Saber la causa es esencial, no sólo porque ello satisface una necesidad de la razón humana, sino, por encima de todo, porque ello es indispensable para la conservación del género humano. Entender significa prever, preocuparse, prevenir y controlar. No es este el lugar de describir las enconadas discusiones lógicas y epistemológicas acerca de la validez y la fundamentación del "principio de causalidad". Aceptemos sin atribuirle significación ontológica, que se trata de una exigencia de nuestra razón, de la cual no podemos prescindir, y la ciencia jamás prescinde. A menudo se tiene una idea demasiado simplista del concepto causa y en el discurso científico aparecen aún frecuentemente confundidos tres acepciones distintas, analizadas por Bergson. En efecto, una causa puede obrar por impulso, por descarga o por desarrollo. Una bola de billar que choca con otra determina su movimiento por impulso. La chispa que enciende la pólvora obra por descarga. El motor que hace dar vueltas al tocadiscos desarrolla la melodía grabada en el disco. Obsérvese que los tres casos son diferentes y que lo que los distingue entre sí es la mayor o menor solidaridad entre las causas y el efecto. En el primero, la cantidad y cualidad del efecto varían con la cantidad y cualidad de la causa. En el segundo ni la cantidad ni la cualidad del efecto varían con la cantidad y cualidad de la causa: el efecto es invariable. En el tercero la cantidad del efecto depende de la cantidad de la causa, pero la causa no influye en la ca-

lidad del efecto. Solamente en el primer caso la causa explica realmente el efecto; en los otros el efecto se prevé, más o menos anticipadamente, y el antecedente invocado es, en mayor o menor grado, desde luego, la ocasión más bien que la causa. Son raras, en Biología y Medicina, las ocasiones en que se puede aislar una causa, que obra "por impulso". La realidad es que los hechos antecedentes que condicionan los acontecimientos son una compleja pluralidad de factores recíprocamente dependientes e interactuantes. Aceptemos, pues, que cuando tratamos de averiguar la etiología de un determinado proceso mórbido, o de su curación, estamos buscando la totalidad de condiciones antecedentes que son a la vez necesarias para que ocurra el hecho y suficientes para llevarlo a efecto, y en ausencia de una de las cuales, no se produce.

A menudo se comete en medicina el error de buscar "la causa" de una enfermedad determinada, y no es raro que se formen bandos y escuelas que atribuyen a un determinado factor la etiología, olvidando o ignorando los conceptos sobre causalidad que hemos discutido.

Existe otro error de interpretación lógica, que los clínicos cometemos con gran frecuencia. Una vez que el investigador ha determinado que el resultado de su experimento es significativo y no se debe a meras variaciones en el material de estudio o en el equipo de observación y medición, trata de determinar cómo y por qué la maniobra experimental causó dicho resultado. Si el material de experimentación es

examinado, no puede cambiar espontáneamente, de modo que el investigador necesita considerar solamente a la maníobra experimental como la causa del cambio. Pero la materia viviente, en cambio, tiene capacidades biológicas intrínsecas de crecer, cicatrizar, sanar y en fin, de modificarse a sí misma. La materia viviente tiene, además, propiedades psicológicas, ausentes o poco notables en los animales inferiores, que pueden afectar muchas reacciones y sensaciones en los enfermos. Despréndese de todo ello que la mejoría del paciente, sometido a un tratamiento determinado, puede deberse no al agente terapéutico, sino a sus poderes naturales de regeneración o incluso a factores psíquicos diversos. A menos que el investigador reconozca que estos factores biológicos y psíquicos pueden cambiar el curso de la enfermedad humana, puede cometer la falacia clínica tradicional, tan frecuente y extendida, del razonamiento *post hoc ergo propter hoc*. El clínico administra un agente terapéutico, el paciente mejora y se concluye que el agente debe ser responsable del éxito obtenido. Gracias a este error de razonamiento, más de una droga, dieta o procedimiento de la medicina antigua y moderna ha recibido falso crédito por éxitos debidos en realidad a otros factores.¹¹

Error lógico parecido, que los investigadores clínicos de todos los tiempos han cometido tradicionalmente, es de intentar establecer conexiones causales mediante estudios retrospectivos. Por ejemplo, para demostrar la hipótesis de que la colitis ulcerosa inespecífica se debe a causas emocionales se han estu-

diado retrospectivamente grupos de enfermos con dicho diagnóstico y se ha encontrado que, en efecto, es muy frecuente el hallazgo de inmadurez emocional, nerviosismo, dependencia, hostilidad reprimida y culpa. Como estas alteraciones psíquicas no se han encontrado con igual frecuencia en grupos testigos de personas sanas o víctimas de otros padecimientos, se ha pretendido que la hipótesis psicogénica ha quedado así plenamente confirmada. Trátase de una falacia evidente, de un error de interpretación lógica; las alteraciones psíquicas pueden ser independientes de la enfermedad del colon e incluso pueden ser una consecuencia y no la causa.

Si se quiere estudiar la hipótesis de que A es causa de B debe procederse de otra manera: no debe buscarse la frecuencia con que aparece A como antecedente de B sino la frecuencia con que, en una serie de eventos A tiene lugar el evento B en comparación con la frecuencia con que tiene lugar el evento B en ausencia de A. Volviendo al ejemplo anterior, la investigación que pudiera establecer una relación entre colitis ulcerosa inespecífica y factores psíquicos debería ser *prospectiva*: estudiar un grupo de personas con las alteraciones psicológicas antes señaladas y un grupo testigo adecuado sin dichas alteraciones, y comparar la frecuencia respectiva con que aparezca la colitis ulcerosa inespecífica en un período adecuado de observación.

Este problema de interpretación lógica se parece a otro, clásico también, relacionado con los estudios de sensi-

bilidad y especificidad de datos clínicos o de laboratorio.

Supongamos que con objeto de establecer la importancia del dolor de cabeza en el diagnóstico de los tumores cerebrales o de la reacción de Wassermann en el diagnóstico de la sífilis determinamos la frecuencia de dichos datos y la encontramos sumamente elevada. El error lógico de conferirles elevado valor diagnóstico es evidente y, en el caso de la reacción de Wassermann, fue causa de innumerables equivocaciones, a menudo trágicas. En efecto hemos determinado sensibilidad de aquellos datos, pero para conocer su valor diagnóstico es indispensable también conocer su especificidad, o sea, determinar la relación recíproca: frecuencia con que, en poblaciones con dolor de cabeza, o con reacción de Wassermann positiva, éstos son causados por tumores cerebrales, o por sífilis, respectivamente.

Volvamos a los estudios prospectivos y veamos que, aunque definitivamente superiores a los retrospectivos para el establecimiento de relaciones causales, pueden dar lugar de todos modos a toda otra serie de errores de interpretación lógica.

En efecto, si tras cuidadosa investigación prospectiva hemos demostrado que B ocurre más frecuentemente después de A que en ausencia de A y que la diferencia es estadísticamente significativa, aún no hemos probado que A es la causa de B. De hecho, ningún cálculo estadístico permite afirmar la existencia de una relación causal. Lo único que hace es dar apoyo lógico a la

hipótesis la cual hubiera tenido que ser descartada en caso contrario.

En efecto¹² en el caso de que el análisis estadístico o el experimento encaminados a valorar una hipótesis, resultan negativos, se está desarrollando un argumento lógico que se ha denominado, clásicamente, *modus tollens*, de tipo deductivo, en el cual la hipótesis es la premisa mayor, el acontecimiento experimentalmente provocado o el análisis estadístico son la premisa menor y la conclusión es infaliblemente válida. Véase que la investigación puede esquematizarse de la siguiente manera:

Si H es verdad, también lo es I
 Pero I no es verdad.

Por tanto H no es verdad.

Ahora bien, si la observación o el experimento están acordes con la hipótesis, o sea, volviendo al esquema:

Si H es verdad, también lo es I
 I es verdad

Luego H es verdad

La afirmación de que la hipótesis es verdadera es una falacia, clásicamente conocida como la "falacia de afirmar el consecuente". En efecto, la conclusión puede ser falsa aunque el resultado del experimento o del estudio estadístico haya sido positivo.

Supongamos, por ejemplo, que sustentamos la hipótesis de que la úlcera péptica es causada por el uso crónico e indiscriminado de aspirina. Hacemos un estudio prospectivo bien ejecutado, reuniendo un grupo grande de personas que abusan de la aspirina y un grupo comparable, testigo, de personas que no utilizan dicho fármaco. Supon-

gamos que, en efecto, encontramos una frecuencia de úlcera péptica significativamente mayor en el primer grupo que en el segundo. La hipótesis parece haber quedado confirmada, pero en realidad no es así. Puede ser, por ejemplo, que la cefalea es muy frecuente en los pacientes con úlcera; y que por esa razón la proporción de los que consumen aspirina es muy elevada. El ejemplo es particularmente útil porque ilustra claramente la diferencia entre significación estadística y verdad lógica.

Al interpretar lógicamente los resultados de un análisis estadístico o de un estudio experimental es necesario tomar en cuenta que a menudo la hipótesis que se está investigando se apoya en "hipótesis secundarias". Por tanto es importante tomarlas en cuenta para decidir si un resultado experimental desfavorable permite descartar la hipótesis que se está investigando.

El esquema sería el siguiente:

Si tanto H como A son verdaderos,
también lo es I,

Pero I no es verdadero

H y A no son, ambos, verdaderos.

Supóngase, por ejemplo, que en un intento de investigar el origen psicogénico de la úlcera péptica estudiamos un grupo de pacientes ulcerosos sometidos a psicoterapia y encontramos con que el número de casos beneficiados fue muy reducido. Este resultado negativo de la pesquisa no necesariamente descarta la hipótesis de la etiología psicológica de la úlcera péptica, ya que se encuentra asociada una hipótesis secundaria: la de que los trastornos psicológicos siempre curan mediante psicoterapia; una

de las dos hipótesis es falsa, pero no necesariamente las dos.

VII. CONCLUSIONES

Hemos dicho, al principio de este trabajo, que la interpretación lógica constituye la parte más difícil y delicada de la metodología de la investigación científica. Después de todo la inteligencia humana tiene frecuentes fallas cuando trabaja sin vigilancia crítica. No siempre se somete al modo de ser de las cosas, no logra espontáneamente una adecuación a ellas, tiene sus exigencias propias, que deben ser denunciadas y corregidas. Quiere explicaciones a toda costa y es ya bien sabido que hasta en el pensamiento de mayor dignidad científica el afán explicativo ha perjudicado al seguro conocimiento, pasando por encima de una minuciosa aprehensión de los hechos, de una descripción completa y veráz, que es imprescindible supuesto para cualquier explicación válida y de otro lado, la urgencia explicativa nos presenta como plausibles explicaciones que de ninguna manera lo son. Una enérgica tendencia de nuestra mente a la unificación, a la simplificación, que nos proporciona una falaz claridad intelectual y una posibilidad de manejar con comodidad hechos y nociones, se traduce en dos resultados opuestos pero movidos por el mismo resorte simplificador: la falsa identificación y la falsa oposición. Nuestra inteligencia identifica de continuo, construye paradigmas con que substituye la diversidad riquísima de la experiencia. El presente ha sido un esfuerzo modesto de presentar,

de manera incompleta, los principales peligros y la manera de esquivarlos, las posibilidades y las limitaciones, de la interpretación lógica.

REFERENCIAS

1. Chalmers, T. C.: *Mortality rate versus funeral rate in clinical medicine*. *Gastroenterology* 46: 788, 1964.
2. Ingelfinger, F. J.; Relman, A. S. y Finland, M.: *Controversy in Internal Medicine*. Philadelphia, W. B. Saunders Co. 1966.
3. Sepúlveda, B.: *Metodología de la investigación clínica en gastroenterología*. *Rev. Gastroenterol. Méx.* 33: 505, 1968.
4. Feinstein, A. R.: *Clinical Judgment*. Baltimore, The Williams & Wilkins Co. 1967, p. 23.
5. Hill, A. B.: *Principios de Estadística Médica*, 3a. ed. Traducida de la 7a. inglesa. Coll, J. A. (Ed.) Editorial El Ateneo, Buenos Aires. p. 294.
6. Romero, F. y Pucciarelli, E.: *Lógica*. Espasa-Calpe Argentina. Buenos Aires, 1938. p. 85.
7. Hempel, C. G.: *Philosophy of Natural Science*. Englewood Cliffs, Prentice-Hall Inc. 1966, p. 65.
8. Feinstein, A. R.: *Op. cit.* Pág. 56.
9. Bridgman, P. W.: *The Logic of Modern Physics*. New York, The MacMillan Co. 1927.
10. Romero, F. y Pucciarelli, E.: *Op. cit.* Pág. 74.
11. Feinstein, A. R.: *Op. cit.* Pág. 51.
12. Hempel, C. G.: *Op. cit.* Pág. 7.

VI

CONCLUSIONES

BERNARDO SEPÚLVEDA¹

1. Hipótesis

POR LO QUE se refiere a la hipótesis, podría decirse en resumen lo siguiente: la hipótesis es elemento esencial de la investigación y sirve como punto de partida al método experimental para la adquisición de nuevos conocimientos.

A pesar de que la hipótesis nunca puede ser demostrada de manera absoluta, es aceptable como válida en la práctica, si hay suficiente evidencia ex-

perimental que la apoye; y a pesar de que sea incorrecta en sí misma, puede llevar a la realización de experimentos que produzcan valiosos resultados. Es condición fundamental para la utilización de la hipótesis, que sea formulada en términos definidos y precisos, a fin de integrar hipótesis unitarias; éstas a su vez son la base para la planeación de experimentos unitarios, cuyas conclusiones deben referirse esencialmente a la propia hipótesis.

Dos precauciones necesarias en el uso de la hipótesis son abandonarla cuando se demuestra su inutilidad y evitar que la idea preconcebida influya en la inter-

¹ Académico numerario. Hospital General, Centro Médico Nacional, Instituto Mexicano del Seguro Social.

pretación de los hechos recogidos. Por último, con objeto de estimar la frecuencia probable con que interviene el azar en los resultados experimentales, es necesario incluir en el planteamiento la llamada "hipótesis de nulidad"; y con el fin de interpretar debidamente la validez de los resultados experimentales, es necesario asimismo tomar en cuenta las hipótesis auxiliares.

2. *En lo relativo a la planeación del experimento*, sería posible resumir el trabajo del doctor Kumate en los términos siguientes:

Una de las tareas esenciales en la planeación experimental, es la definición de la muestra del universo estadístico, en la cual se va a probar la hipótesis de trabajo, para tratar después de generalizar las conclusiones. La muestra debe ser representativa del universo, por lo cual la selección de la misma debe sujetarse a las normas establecidas para el muestreo. Otros aspectos que deben tomarse en cuenta en la planeación del experimento son:

- a) el tamaño de la muestra;
- b) el error máximo permisible y la posibilidad de excederlo;
- c) la magnitud de los errores tipo I y II; y
- d) la posibilidad de practicar un análisis secuencial.

En las investigaciones planeadas para determinar relaciones etiológicas, se utilizan tanto los estudios retrospectivos como los prospectivos; en los primeros, se parte del efecto para buscar la causa, mientras que en los segundos, se observa si con un probable agente etiológico, ocurre o no el efecto.

En los ensayos terapéuticos, la técnica llamada "doble ciega" es la menos expuesta a la introducción de prejuicio que pudiera invalidar las conclusiones. Los elementos principales en esta técnica son: la integración aleatoria de la muestra; el desconocimiento de las sustancias usadas, por parte de quienes administran el medicamento y el placebo; y el desconocimiento, por parte de quienes están encargados de juzgar los resultados, de los individuos que recibieron medicamento o placebo.

3. En lo referente al trabajo del doctor Cravioto sobre análisis estadístico, los puntos principales podrían sintetizarse en la siguiente forma:

El análisis estadístico permite decidir si los resultados de una investigación confirman o rechazan la hipótesis, conociendo al mismo tiempo el riesgo inherente a tal decisión.

Dentro de la serie de pruebas estadísticas utilizadas en la actualidad, el investigador debe seleccionar la más adecuada al problema en estudio. Tales pruebas se dividen en dos grupos: paramétricas y no paramétricas. Las primeras son las más precisas, pero únicamente pueden aplicarse en aquellos casos en que las mediciones son estrictamente cuantitativas y el investigador conoce los valores de los parámetros de la población de la cual deriva su muestra. Las pruebas no paramétricas se emplean cuando las mediciones sólo pueden hacerse en forma de categorías o de escalas ordinales; asimismo, se utilizan cuando el tamaño de la muestra es muy pequeño, o cuando el investigador no conoce los parámetros de la

población bajo estudio o bien cuando a pesar de conocerlos desea que sus resultados tengan una generalización más amplia de la que permite un modelo estadístico base de una prueba paramétrica equivalente.

4. Finalmente, las conclusiones del trabajo del doctor Jinich sobre la lógica, podrían formularse en los términos que siguen:

La interpretación lógica de toda investigación científica, consiste esencialmente en la aplicación de las leyes del razonamiento. Los razonamientos más importantes son los deductivos, los analógicos y los inductivos. Los razonamientos analógicos o por analogía no permiten llegar a conclusiones seguras sino a hipótesis que es preciso comprobar. La inducción es el procedimiento lógico fundamental de la investigación científica; utiliza métodos cuyo propósito es aislar, entre las circunstancias que rodean a un fenómeno, las que tienen con él una relación constante. A pesar de que el razonamiento inductivo presenta problemas teóricos, debe aceptarse como procedimiento de gran valor práctico.

La fidelidad de la observación exige por una parte la ausencia de prejuicios y por la otra, la máxima precisión, que

debe incluir, cuando sea posible, la medición cuantitativa. Asimismo, debe procurarse la mayor precisión en el lenguaje, particularmente en la definición de los conceptos, con objeto de asegurar la posibilidad de verificar por medio de la repetición, las afirmaciones hechas. Por añadidura, es necesario advertir que el análisis estadístico, como tantos otros instrumentos poderosos, puede usarse de manera inapropiada; en consecuencia, conviene recordar sus limitaciones. Aun cuando el conocimiento de las causas es objetivo principal de la investigación, el concepto de causalidad es motivo de controversia. El investigador debe tomar en cuenta las diversas acepciones del término y las distintas formas como puede manifestarse la acción de la causa.

El investigador debe estar también prevenido contra los errores en la interpretación lógica; entre los principales figuran las tres falacias bien conocidas: la falacia "después de esto, luego por esto"; la falacia de "afirmar el consecuente"; y la de "negar el antecedente".

La hipótesis principal se apoya frecuentemente en hipótesis auxiliares, las cuales conviene tomar en cuenta para la correcta interpretación de los resultados.