

# Empleo del análisis multivariante en la investigación clínica

F. Hervás Maldonado<sup>1</sup>

*Sanid. mil.* 2007; 63 (2): 111-117

## RESUMEN

Se revisan las técnicas de análisis multivariante. INTRODUCCIÓN Y OBJETIVOS: estudiar simultáneamente numerosas variables requiere una metódica especial con dos objetivos concretos, que son el análisis de todos los datos de manera conjunta y la asistencia lógica en la toma de decisiones. TIPOS DE ANÁLISIS MULTIVARIANTE: se relacionan las técnicas descriptivas y explicativas más comunes. SISTEMÁTICA: se indican los siete pasos más comunes que lo desarrollan. CONCLUSIÓN: se propone la creación de asesorías metodológicas para la vigilancia y racionalización de los procesos de investigación. También se destaca la necesidad de más inventores, así como de investigadores asistidos con mejores presupuestos.

PALABRAS CLAVE: Análisis multivariante, Estadística médica.

## INTRODUCCIÓN Y OBJETIVOS

Con frecuencia vemos artículos de magnífica factura en los que se alude al análisis multivariante como modelo justificativo de un hecho presuntamente probado.<sup>1-26</sup> Sin embargo, no siempre es realmente así, por muchas razones, pese a la gran pureza del método estadístico seguido.

En primer lugar, hemos de ser humildes y reconocer que los modelos matemáticos estandarizados suelen cojear por algún lado en los casos concretos.<sup>27-30</sup> Esto no es válido para pequeñas consideraciones de datos, por muchas variables que les afecten, pero sí lo es cuando el número de datos crece de manera espectacular<sup>31-34</sup>. Por ejemplo, trabajar con dos mil individuos no es lo mismo que trabajar con cincuenta, como es lógico, porque siempre existen desvíos significativos de la ley normal, por mucho que nos empeñemos en meter todos los datos en un mismo «saco» estadístico. Pero es que desde un punto de vista militar no nos interesan tanto los cincuenta individuos como los dos mil, como es lógico, pues nuestra labor es de un espectro amplio y no reducido. Tal vez, si nos concentrásemos en la pura asistencia de un área sanitaria, esto no sería así, pero la movilidad de nuestra gente, la variedad de nuestras misiones y el interés prospectivo de nuestra asistencia (llamado también inteligencia sanitaria) hacen que la variabilidad sea inmensa. Por ejemplo, en nuestro laboratorio de Microbiología no podemos considerar una procedencia limitada de la cepa aislada, perteneciente a cada especie microbiana, sino que más bien hemos de considerarla heredera de la diáspora bacteriana o vírica a través de la historia, por los diversos continentes. Por tanto, intentar valorar nuestros datos con un método estadístico convencional suele ser la

antesala del fracaso. No obstante, si tratamos previamente los datos en la forma adecuada, entonces sí podemos recurrir a la metódica indicada.<sup>35-41</sup> Esto, lamentablemente, no suele hacerse.

En segundo lugar, los modelos utilizados suelen referirse a programas estadísticos concretos, orientados – por lo general – a tipos de muestra muy determinados, salvo cuando los epidemiólogos utilizan los procedimientos descriptivos recogidos en programas especiales, como el EPI-INFO u otros parecidos<sup>42-62</sup>. Aunque, como ya hemos dicho, son descriptivos y algo analíticos, de hecho les falta cierta profundidad y, sobre todo, robustez en la prospectiva, que es más bien limitada. Por otra parte, las asociaciones interdependientes suelen ser deficitarias, no tanto las dependientes y sí mucho las estructurales, como luego veremos.

Por último, insistir en la necesidad de un soporte consecuente<sup>31, 32, 63-96</sup> de esta clase en cualquier posible instituto de investigación – sobre todo de estirpe militar – con personal y medios adecuados, bien formados y entrenados, soportados por una estructura matemática de suficiente envergadura, como la que proporcionaría el adecuado departamento de ingeniería matemática, matemática aplicada o incluso diseño matemático de nuestras universidades. Sin ese apoyo, nuestro proyecto está claramente lastrado y es muy poco prometedor, porque la voluntad – en este caso – no puede suplir con su esfuerzo la falta de medios y de conocimiento. Los medios son claros: ordenadores especiales, con muchas gigas de memoria RAM, para poder aunar gráficos y datos múltiples, con posibilidad de adaptar sensores y lectores de todo tipo y con software muy abierto, modificable para cada caso. Porque investigar es caro, no cabe duda.

El *análisis multivariante* viene a ser un conjunto de metódicas estadísticas que analizan muchos datos de muchas variables (más de dos, en sentido purista, varios cientos de forma habitual). Estas variables<sup>36, 59, 97-104</sup> pueden ser dependientes e independientes, o bien solo estas últimas. Por otra parte, los datos suelen ser miles para cada variable. A partir de ahora, y por motivos de estirpe informática, llamaremos *inputs* a estos datos y *outputs* a las conclusiones (o conclusión única) que generen. La secuencia de trabajo se definirá

<sup>1</sup> TCol. Médico. Microbiología  
Hospital Central de la Defensa «Gómez Ulla»

Dirección para correspondencia: fhermal@oc.mde.es, fhervas@terra.es

Recibido: 5 de marzo de 2007  
Aceptado: 28 de junio de 2007

mediante un *algoritmo*, como luego veremos. Pensemos en una ecuación cuando digamos algoritmo,<sup>29,105-107</sup> una ecuación que representamos gráficamente mediante un sistema convencional (y no libre) de grafos. Muchas veces vemos algoritmos representados de cualquier manera. Eso no son algoritmos, son otra cosa,<sup>108-111</sup> generalmente árboles de la decisión ampliados.

Los dos principales objetivos del análisis multivariante son: el análisis *conjunto*<sup>103</sup> de los datos multivariantes y la asistencia *lógica*<sup>107</sup> en nuestra toma de decisiones. Dicho de otro modo: el estudio de una muestra grande y su inferenciación poblacional. Al fin y al cabo, como en cualquier otro problema estadístico, pero con una metódica especial.

## TIPOS DE ANÁLISIS MULTIVARIANTE

Podríamos dividir los análisis multivariantes, según su finalidad, en dos grandes grupos:

- Las técnicas multivariantes descriptivas, en las que no se consideran más que variables independientes. Pueden ser pruebas sincrónicas o diacrónicas. Dentro de ellas, siguiendo la pauta del programa SPSS u otro software, el tratamiento de grupos puede hacerse de manera diversa
- Técnicas explicativas, según posean una o más variables dependientes.

La clasificación, según su metódica, podemos sistematizarla así:

### I. *Métodos de dependencia*, según la *variable* (o variables) dependiente sea:

- a) Con *una sola variable dependiente (VD)*:
  - *VD cuantitativa o métrica*:
    - A. *Variables independientes (VI) cuantitativas*:
      - *Una sola VI*:
        - Coeficiente de correlación de *Pearson* <sup>36, 101, 102, 104</sup>
      - *Varias VI*:
        - Regresión múltiple <sup>36, 102</sup>,
        - Análisis de supervivencia <sup>61, 103</sup>
        - MANOVA <sup>36, 61, 102</sup>,
        - Correlación canónica <sup>36, 100, 102</sup>
    - B. *VI cualitativas*:
      - *Una sola VI*:
        - Dos grupos: t de *Student* <sup>36</sup>
        - Más de dos grupos: ANOVA <sup>36, 61, 102</sup>
      - *Varias VI*:
        - ANOVA factorial <sup>36, 61, 102</sup>
        - Análisis conjunto <sup>61, 103</sup>
    - *VD cualitativa o no métrica*:
      - *VI cuantitativas*:
        - Análisis discriminante <sup>61, 103</sup>
      - *VI cualitativas*:
        - $\chi^2$  <sup>36, 101</sup>
        - Modelos log-lineales <sup>61, 103</sup>
        - Análisis conjunto <sup>103</sup>
        - Regresión logística: <sup>103</sup>
  - Con VD de *outputs dicotómicos* (binaria).
  - Con VD de *outputs politómicos* (categorizada): Modelos LOGIT y PROBIT.

- Con VD de *outputs censurados* (o truncada).
  - b) Con *varias VD*:
    - *VI cuantitativas*:
      - Correlación canónica <sup>36, 100, 102</sup>
      - Modelos causales <sup>61</sup>
    - *VI cualitativas*:
      - *Una sola VI*:
        - Dos grupos: coeficiente  $r^2$ , Hotelling <sup>36, 62, 102</sup>
        - Más de dos grupos: MANOVA <sup>36, 61, 102</sup>
      - *Varias VI*:
        - MANOVA <sup>36, 61, 102</sup>,
- II. *Métodos de interdependencia*, o modelos descriptivos, donde no hay variable dependiente conocida:
- 1) *Modelos con descripción sincrónica*:
    - a) Variables con *inputs cuantitativos o métricos*:
      - Determinación de componentes principales <sup>100</sup>
      - Cluster análisis <sup>29, 111</sup>
      - Análisis Factorial <sup>36, 100</sup>
      - Escalas multidimensionales <sup>36, 100</sup>
    - b) Variables de *inputs cualitativos o no métricos*:
      - Escalas multidimensionales <sup>36, 100</sup>
      - Análisis de correspondencias <sup>61, 62</sup>
      - Modelos log-lineales <sup>61, 103</sup>
  - 2) *Modelos con descripción diacrónica*:
    - Series temporales <sup>61</sup>.
    - Análisis de supervivencia <sup>61, 103</sup>

### III. *Métodos estructurales*, que se diseñan de manera personalizada para cada caso <sup>61, 103</sup>.

Algunos de los procedimientos indicados tienen las siguientes aplicaciones en la Sanidad Militar:

**Regresión múltiple:** se utiliza cuando diversas variables cuantitativas dependen de otras varias también cuantitativas e independientes. *Por ejemplo, calcular la previsión de gasto en material, las necesidades de personal y el consumo de fármacos por zonas en un proceso epidémico a partir de la incidencia de casos, la prevalencia zonal, las edades de la población, la previsión de temperatura y lluvia, la densidad de población por zonas, las tasas migratorias y la evolución del IPC interanual por meses (índice de precios al consumo).*

**Análisis de supervivencia:** Es similar al anterior, pero con la diferencia de que la única variable dependiente es el tiempo de supervivencia de un individuo u objeto a partir de diversas variables independientes. *Un ejemplo: cuanto tiempo puede permanecer un EMAT sin suministro antibiótico a partir de la tasa de incidencia de las diversas patologías infecciosas, ajustando consumo por patología, teniendo en cuenta las caducidades medias de los fármacos, las condiciones de conservación, el personal para su administración y la demanda de asistencia.*

**ANOVA o MANOVA:** el análisis de la varianza (ANOVA) o análisis multivariante de la varianza (MANOVA) se realiza cuando confrontamos grupos de una o más variables independientes cualitativas con diversas variables dependientes cuantitativas. Tiene por objeto el determinar las diferencias significativas entre los grupos con respecto a las variables cuantitativas. *Como ejemplo, averiguar si hay diferencias en la tasa de colesterol por sexos en grupos de*

edad, o si afecta también el tipo de actividad o la vestimenta, así como la influencia de grupos de alimentos.

**Correlación canónica:** se trata de relacionar entre sí, simultáneamente, diversas variables cuantitativas dependientes e independientes, mediante la determinación de las diferentes correlaciones lineales de cada conjunto de variables, las cuales logren maximizar la correlación existente entre los dos conjuntos de variables, dependientes e independientes. *El ejemplo, podría versar sobre ver la relación existente entre minutos diarios de práctica deportiva, calorías de ingesta diarias, capacidad pulmonar y peso, con días de afectación por enfermedad común al año, respuesta refleja inducida, potencia muscular del brazo dominante, tensión arterial mínima y pulsaciones por minuto.*

**Análisis discriminante:** proporciona reglas óptimas de clasificación de inputs, desconociendo su grupo de procedencia y basándose en los valores de las variables independientes, que son todas cuantitativas. *Para ejemplo puede servir el establecer puntos de corte en un grupo de parámetros bioquímicos, que sirvan para discriminar enfermos de la población en estudio para el diagnóstico de una determinada enfermedad. Nosotros solo sabemos la tasa de prevalencia o incidencia de la enfermedad, «acoplándola» con la distribución de inputs encontrada.*

**Modelos de regresión logística:** la variable dependiente es cualitativa. Es una alternativa al análisis discriminante, cuando la distribución no sigue una ley normal. *Por ejemplo, en dos grupos de sanos y enfermos, ver como se comportan pruebas positivas y negativas de diagnóstico. O bien, en un grupo de enfermos con varias categorías (muy graves, graves y leves) vemos como se comportan diversas pruebas diagnósticas positivas y negativas. E incluso en un grupo de enfermos de una edad, frente a otro de enfermos de otra edad, ver cómo se comportan dichas pruebas positivas y negativas, condicionándolas según el día de la semana.*

**Análisis conjunto:** esta técnica estudia el efecto de inputs de variables independientes cualitativas sobre variables dependientes cuantitativas o cualitativas. La diferencia con el análisis de la varianza (ANOVA) se basa en que las variables dependientes pueden ser cualitativas y en que los inputs de las variables independientes no son aleatorios, sino fijados por el investigador. Es lo que se conoce como diseño de experimentos. *Por ejemplo, un investigador desea saber la respuesta de distintos modelos dietéticos en un grupo de animales de experimentación. Utiliza cuatro grupos de ratas whistar y administra combinaciones de piensos, valorando su acción hipo o hiperlipemiente, así como la respuesta intelectual en un recorrido de laberinto, en tiempo. Para la primera variable dependiente, los outputs los somete a un punto de corte, por encima o por debajo del valor x de colesterol. Para la segunda variable dependiente (output cuantitativo) mide el tiempo tardado en el recorrido en segundos, transcurridos 24 horas de la ingesta del pienso sólido o 6 del suplemento líquido. Las variables independientes son dos. En la primera, los inputs poseen cuatro categorías: pienso estándar, pienso estándar con suplemento graso, pienso estándar con suplemento graso y producto hipolipemiente y pienso estándar solo con producto hipolipemiente. La segunda tiene dos categorías de inputs, agua con y sin azúcar.*

**Análisis factorial y análisis de componentes principales:** Se interrelacionan un número elevado de variables cuantitativas, todas ellas independientes. Sus interrelaciones vienen dadas a través de

un número menor de variables que llamamos factores (si no son observables) o componentes principales (si son observables). *Como ejemplo de Componentes Principales podemos considerar la estandarización de valores biométricos para elaborar una tabla de incapacidades para el servicio, en la que estudiamos a una población amplia y medimos urea, glucosa, colesterol, GOT, GPT, GGT, bilirrubina, presión sistólica, presión diastólica, densidad de la orina, pH urinario, etc. Habrá que buscar la interrelación entre todas las variables observadas para llegar al número y tipo de las mismas a considerar. Otro ejemplo es cuando un psicólogo desea definir los factores que determinan la inteligencia de unos individuos, a partir de las respuestas a un test de inteligencia, utilizando un Análisis Factorial para resolverlo.*

**Escalas multidimensionales:** tienen por objeto el poder representar en un espacio multidimensional las preferencias o semejanzas. Construiremos un mapa, dibujando las posiciones de los objetos, comparándolos de forma que aquellos que percibimos como similares estén más próximos y los que percibimos como distintos, estén más lejanos. *Un ejemplo es analizar entre un grupo de soldados los alimentos percibidos como más apetitosos y los que no les gusta comer; según su consumo en raciones por día, haciendo un seguimiento de una dieta controlada durante un tiempo. Nuestro objetivo es elaborar una dieta más a su gusto. La prueba puede efectuarse igualmente con variables cualitativas, sobre todo si poseen más de dos categorías.*

**Cluster análisis o análisis de conglomerados:** Tiene como finalidad la clasificación de variables cuantitativas en grupos cada vez menores, basándonos en la similaridad y disimilaridad de algunos de sus estadísticos, recogida mediante coeficientes de comparación y representada gráficamente mediante un grafo denominado dendrograma (por parecerse a las dendritas de las neuronas). Difiere del análisis discriminante en que se desconoce el número y composición de los distintos grupos. *Por ejemplo, determinar la probabilidad de que una determinada secuencia genética obtenida defina un microsatélite familiar de ADN (comúnmente llamada la prueba del ADN de identificación de cadáveres). Esta prueba (cluster) puede utilizarse también con variables cualitativas o mezcla. Las métodos de análisis de conglomerados son muchas (tablas I y II), como en otras técnicas multivariantes, por lo que conviene elegir la más adecuada y, ante la duda, pedir asesoría a personas expertas.*

**Análisis de correspondencias:** se puede utilizar con inputs de variables independientes cualitativas, pero no cuantitativas. Se utiliza en las tablas de contingencia multidimensionales. Se diferencia de las escalas multidimensionales en que se representan simultáneamente las filas y columnas de las tablas de contingencia, pero se busca un objetivo similar en ambos procedimientos. *Por ejemplo, estudiar la gripe en los soldados, teniendo en cuenta tamaño de la unidad (pequeña, mediana, grande), zona geográfica de ubicación (norte, centro o sur; montaña o costa, ciudad o rural), si ha habido desplazamientos o no de la unidad en los últimos seis meses, sexo analizado, etc. Es decir, relacionar todo con la presencia o ausencia de gripe en cada caso y sumar.*

**Tabla I.** Procedimientos de clasificación de grupos de SPSS.

Clasificación de Grupos	Cluster Analysis Discriminantes	Quick Cluster Cluster
		Discriminante

**Tabla II.** Clasificación de grupos de inputs mediante conglomerados.

Criterio de Selección
Agrupación por distancias (conglomerados o clusters) con criterios de similitud/disimilitud
1. Métodos Jerárquicos
1.1. Algoritmos Jerárquicos Aglomerativos
1.1.1. Algoritmos de Contractura Espacial * Método de la Ligadura Simple
1.1.2. Algoritmos de Dilatación Espacial * Método de la ligadura completa
1.1.3. Algoritmos de Conservación Espacial * Método de la Ligadura Media * Método de la Ligadura Media Ponderada * Método del Centroide * Método de la Mediana * Método de la Varianza o de Ward
1.2. Algoritmos Jerárquicos Disociativos
2. Métodos Particionales
2.1. Algoritmo Convergente de K - Medias o de Hartigan
2.2. Otros Algoritmos
2.2.1. Análisis de Densidad y Moda de Wishart
2.2.2. Algoritmo Líder

**Modelos log-lineales:** Se aplican en las tablas de contingencia multidimensionales, buscando establecer un modelo de relación dependiente de tipo multidimensional entre los inputs observados de las variables, con objeto de poder explicar las frecuencias calculadas. Es un método logarítmico similar al anterior, pero más orientado a problemas complejos que requieren una mayor precisión. *Por ejemplo, analizar la gripe aviar en la tropa, según las mismas variables del caso anterior y con modelos simulados de progresión de la enfermedad sobre la base de los acuíferos.*

**Métodos estructurales:** estudian las relaciones existentes entre grupos de variables expresadas mediante sistemas de ecuaciones que se valoran a un tiempo, considerándose que cierto tipo de ellas (llamadas constructos) son medidas erróneamente en cierto grado, a partir de otras variables observables que se denominan indicadores. Estos modelos constan de dos partes: una parte estructural (que indica las relaciones de dependencia existentes entre las ecuaciones constructos latentes) y otro modelo de medida (que define cómo los indicadores se relacionan con sus constructos correspondientes). *Por ejemplo, analizar cómo se relacionan las imágenes obtenidas mediante un estudio de PET con las percepciones de enfermedad del médico (signos) y del paciente (síntomas), así como con otras pruebas diagnósticas sugestivas.*

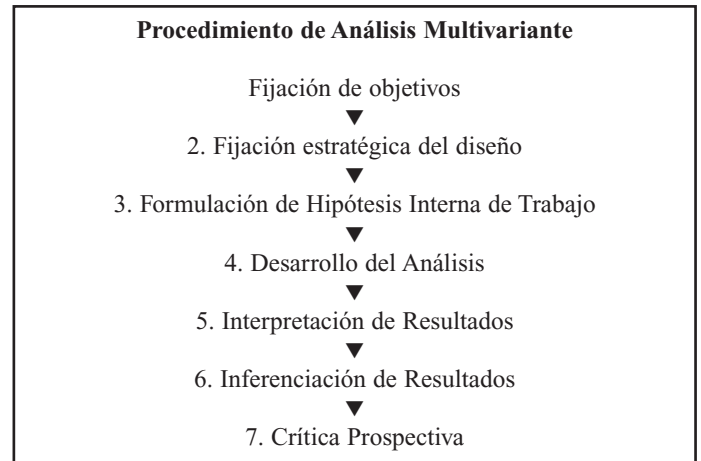
**SISTEMÁTICA**

Existen muchos enfoques posibles a la hora de efectuar un análisis multivariante,<sup>29, 30, 61, 103</sup> de manera que cualquier metódica ge-

neral de trabajo es simplemente orientativa, aunque si se basa – como es natural – en la experiencia aplicada a la sanidad (o a cualquier otro campo más concreto del conocimiento), el rango de variabilidad metódica se estrecha. Por tanto, nuestra opción está un poco más clara.

Siete fases (tabla III) presiden un estudio de este tipo:

**Tabla III.** Fases de un Análisis Multivariante.



1ª) Fijación de objetivos del análisis. Junto con ellos, habremos de fijar las técnicas de análisis multivariante que los posibiliten.<sup>29, 30</sup> Por tanto, el investigador habrá de establecer la relación conceptual de lo que quiere investigar. Definirá los conceptos en liza y cómo se afectan o pueden afectar mutuamente. Es decir, que habrá de establecer las relaciones de dependencia o de interdependencia de las variables a considerar, para saber si son dependientes o independientes y si son cualitativas o cuantitativas.

2ª) Fijará el tamaño muestral,<sup>36, 59, 60, 100</sup> de acuerdo con su presupuesto y casuística, en primer lugar, la estimación de ecuaciones previsible (si es procedente), así como las distancias a calcular (igualmente, si procede) y las técnicas empleables para la estimación. A partir de ahí se recogen los inputs o datos.

3ª) Elaborar las hipótesis internas de trabajo.<sup>36, 59, 60, 61, 100, 102</sup> Evaluaremos las posibles hipótesis de normalidad, linealidad, independencia, homocedasticidad (cuando la varianza de los errores estocásticos – de azar – de la regresión no se ve modificada por el incremento de las observaciones)... Igualmente decidiremos lo que hemos de hacer con el ajuste imputable a los datos perdidos (missing inputs).

4ª) Efectuar el análisis. Estimamos el modelo y ajustamos los datos al mismo <sup>36, 59, 60, 61, 100, 102, 107</sup>. Si nos apareciesen observaciones atípicas (conocidas técnicamente como outliers) o que influyan anormalmente sobre la estimación y su ajuste, habremos de considerar la utilidad real de su inclusión. Esto no quiere decir que podamos «maquillar» nuestra prueba, sino todo lo contrario. Lo que deberemos hacer es impedir que nos la «maquillen» desde fuera terceras razones no conocidas que hayan modificado previamente los datos observados. Pero ha de comprobarse que tales terceras razones existen.

5ª) Interpretar los resultados <sup>60, 61</sup>. Esto no es nada sencillo. A veces, los resultados nos obligan a volver a la 3ª fase y reconsiderar nuestras hipótesis, e incluso a retomar la 4ª fase, repitiendo los análisis con la misma u otra metódica.

6ª) Habremos de validar nuestros análisis, mediante su expresión sobre la población a que se refiere la muestra estudiada. Es decir, nuestra inferenciación ha de tener un claro fundamento de razón para que la utilidad de nuestro estudio sea patente.<sup>39, 60, 61, 62</sup>

7ª) Por último, fijaremos la línea o líneas de posible continuidad<sup>32, 107</sup> en nuestro trabajo, mediante la sugerencia de enfoques futuros, siempre a la luz de una autocrítica.

*Pondremos un ejemplo, queremos ver como se asocian dos determinados genes de resistencia a una familia de antimicrobianos con cuatro antimicrobianos de esa misma familia en dosis crecientes, en seis especies bacterianas a las que se adicionan cantidades ínfimas de antimicrobiano en medio líquido y se comprueba su crecimiento, secuenciando los aislados con la ayuda de microarrays. Ese es nuestro objetivo. Para eso se propone utilizar un método asociativo de análisis de conglomerados y seleccionamos – por razones que no hacen al caso – el método de la mediana, de estimación mediante coeficientes de medianas relacionadas<sup>29, 30</sup>. Por tanto, tenemos una variable cualitativa independiente (gen de resistencia, con dos categorías), otra variable cualitativa independiente (especie bacteriana, con seis categorías), una variable cuantitativa de antimicrobiano independiente (dosis por ml de antimicrobiano, repetidas tras un número variable de ciclos de crecimiento), una más independiente cualitativa (tipo de antimicrobiano añadido), una dependiente cuantitativa (UFC o, si queremos, densidad por ml, tras el último de los ciclos de crecimiento expuesto) y otra dependiente cuantitativa (porcentaje de homología genética con cada gen de resistencia). Es la primera fase.*

*En segundo lugar, fijamos el tamaño de la muestra (básicamente el presupuesto nos lo va a fijar, en este caso, pero podríamos recurrir a otros procedimientos, aunque al ser inmensa la población bacteriana, van a ser poco útiles) y la forma de medir: microarrays, e-test, densitometría... Recogemos los datos y los tabulamos a continuación.*

*La tercera fase conlleva revisar nuestro trabajo y comprobar si todo lo hemos hecho de la misma manera, así como si existen factores externos de discordancia que no han sido considerados en alguna de las observaciones. Reajustamos la muestra, desechando lo que probamos que no procede.*

*En cuarto lugar, efectuamos la prueba de análisis de conglomerados (método de la mediana), de manera que vemos cómo se agrupan los elementos y las variables, estableciéndose las distancias entre unos y otros o entre unas y otras, mediante valoración de similitudes y disimilitudes.*

*Interpretamos, en quinto lugar, los resultados, viendo cómo se asocia la presencia de un gen de resistencia con la exposición repetida «in vitro» a un antimicrobiano o no.*

*A continuación (sexta fase) inferimos sobre población en virtud de la frecuencia e intensidad encontradas en esa asociación.*

*Por último, proponemos un modelo sencillo de seguimiento futuro de resistencias y una posible causa de esta asociación que ha de ser investigada más profundamente en una determinada línea de trabajo.*

### CONCLUSIÓN

No debemos conformarnos con el estudio somero de los problemas, sino que el estado del conocimiento matemático actual, re-

quiere profundizar más en los diseños<sup>29, 39, 61, 102, 103</sup>. Es preferible que las investigaciones sean menos y de más calidad, porque así se rentabiliza mucho mejor el presupuesto para ello. Además, la posibilidad de abordaje cada vez más complejo, hace recomendable plantearse asociaciones de trabajo un poco más estables, con unas líneas suficientemente definidas y respaldadas. Es muy conveniente que en estos grupos se establezcan asesorías metodológicas profesionales, lideradas por expertos en el diseño, producción y difusión de la investigación, así como en el tratamiento adecuado de los datos obtenidos y difundidos. En el caso concreto de los institutos de investigación, la necesidad de asesoría es mayor, aunque solo sea para racionalizar esfuerzos y unificar criterios. Este es nuestro caso militar, donde la investigación tiene un propósito muy concreto: promoción de la defensa nacional y protección del personal.

Hoy en día, el análisis multivariante no es algo inalcanzable, pues disponemos de magnífico software para poder desarrollarlo. Por eso, conviene ser cautos y desconfiar de softwares que no presenten una clara dimensión de calidad internacionalmente reconocida, como es el caso de EPI-INFO, SPSS, Statgraphics o Mathematics, por citar los cuatro más utilizados por lo general. Hay otros también muy buenos, dependiendo del problema en que nos ocupemos. Pero tampoco seamos excesivamente críticos, porque ni estos ni ningún software sirven para todos los casos. En situaciones muy concretas (29), nos lo habrán de desarrollar «ex – profeso».

Por último, pensemos un poco en términos de rentabilidad. ¿Necesitamos investigadores o inventores? Porque no es la misma cosa ni genera los mismos beneficios. Los inventores generan grandes beneficios a corto plazo y con bajo coste. Los investigadores pueden generar también grandes beneficios – o tal vez no, porque su trabajo se desarrolla en situaciones de incertidumbre, al menos relativa – pero siempre con un coste muy elevado y a un plazo más largo. Debe haber investigadores, sin duda, pero con buenos presupuestos. Faltan cantidades ingentes de inventores, que son los que realmente proporcionan los fondos necesarios para costear el gasto en la investigación<sup>112</sup>. Su trabajo desarrolla el de los investigadores y lo rentabiliza, con inversiones muy ajustadas a demanda. Al menos así es como se costea la investigación en los países de PIB más ventajoso.

### BIBLIOGRAFÍA

1. Lang ES, Boivin JF et al. Do electronic linkages between the Emergency Department and Primary Care Physicians reduce resource utilization in the Emergency Department? Results of a randomized controlled trial. *Acad Emerg Med*, 2003; 10(5): 547-8.
2. Saitz R, Horton NJ et al. Addressing alcohol problems in primary care: a cluster randomized, controlled trial of a systems intervention. The screening and intervention in primary care (SIP) study. *Ann Intern Med*, 2003; 138(5): 372-82.
3. Croudace T, Evans J et al. Impact of the ICD-10 primary health care (PHC) diagnostic and management guidelines for mental disorders on detection and outcome in primary care. Cluster randomised controlled trial. *Br J Psychiatry*, 2003; 182: 20-30.
4. Sheikh K, Smeeth L, Ashcroft R. Randomised controlled trials in primary care: scope and application. *Br J Gen Pract*, 2002; 52(482): 746-51.
5. Bierma-Zeinstra SM, Bohnen AM et al. Hip problems in older adults: classification by cluster analysis. *J Clin Epidemiol*, 2001; 54(11): 1139-45.
6. Campbell MJ. Cluster randomized trials in general (family) practice research. *Stat Methods Med Res*, 2000; 9(2): 81-94.
7. Kendrick D, Marsh P et al. Preventing injuries in children: cluster randomised controlled trial in primary care. *BMJ*, 1999; 318(7189): 980-3.

8. Cotter B, Bremer V et al. Assessment of neonatal tetanus elimination in African setting by lot quality assurance cluster sampling (LQA-CS). *Epidemiol Infect*, 2003; 130(2): 221-6.
9. Heegaard WG, Plummer D et al. Ultrasound in Helicopter EMS. *Acad Emerg Med*, 2003; 10(5): 449.
10. McLachlan RI, Baker HW et al. Semen analysis: its place in modern reproductive medical practice. *Pathology*, 2003; 35(1): 25-33.
11. Dhatt GS, Pum JK et al. Evaluation of HbA1C results in an external quality assessment scheme in South Africa. *Clin Chem Acta*, 2003; 331(1-2): 147-51.
12. Graf J. Do you know the frequency of errors in your intensive care unit?. *Crit Care Med*, 2003; 31(4): 1277-8, 1006-11.
13. Mason I, Brady C. Benchmarking and nutrition. *Nurse Times*, 2003; 99(3): 49-50.
14. Murray M, Bodenheimer T et al. Improving timely access to primary care: case studies of the advanced access model. *JAMA*, 2003; 289(8): 1042-6.
15. Whitten KL, Rein MF et al. The emotional experience of intercourse and sexually transmitted diseases: a decision-tree analysis. *Sex Transm Dis*, 2003; 30(4): 348-56.
16. Delaney G, Barton M et al. A model for decision making for the use of radio therapy in lung cancer. *Lancet Oncol*, 2003; 4(2): 120-8.
17. Birkmeyer JD, Liu JY. Decision analysis models: opening the black box. *Surgery*, 2003; 133(1): 1-4.
18. Hunsberger B, Bagwell CB, Herbert D, Bray C, Langweiler M. Effects of resolution reduction on data analysis. *Cytometry*, 2003; 53A(2): 103-11.
19. Tan CK, Law NM, et al. Simple clinical prognostic model for hepatocellular carcinoma in developing countries and its validation. *J Clin Oncol*, 2003; 21(12): 2294-8.
20. Redleaf MI, Bauer CA. Topical antiseptic mouthwash in oncological surgery of the oral cavity and oropharynx. *J Laringol Otol*, 1994. 108(11): 973-9.
21. Hawrylyshyn PA, Bernstein P, Papsin FR. Risk factors associated with infection following cesarean section. *Am J Obstet Gynecol*, 1981. 139(3): 294-8.
22. Raz R, Okev N, Kennes Y, Gilboa A, Lavi I, Bisharat N. Demographic characteristics of patients with community-acquired bacteriuria and susceptibility of urinary pathogens to antimicrobials in northern Israel. *Isr Med Assoc J*, 2000. 2(6): 426-9.
23. Ruza F, Alvarado F, Herruzo R, Delgado MA, García S, Dorao P, Goded F. Prevention of nosocomial infection in a pediatric intensive care unit (PICU) through the use of selective digestive decontamination. *Eur J Epidemiol*, 1998. 14(7): 719-27.
24. Herruzo-Cabrera R, García JI, García-Magán P, Del Rey-Calero J. Nosocomial infection in a neonatal intensive care unit and its prevention with selective intestinal decolonization. A multivariate evaluation of infection reduction. *Eur J Epidemiol*, 1994. 10(5): 573-8.
25. El tratamiento del dolor de garganta con antibióticos en atención primaria. *Rev Panam Salud Pública*, 2001. 10(4): 258-9.
26. Rubio LA, Sedeño C, Fernández SA, Arbesú MA, Puig L. Grado esperado de uso de los antimicrobianos aminoglucósidos en pacientes oncológicos. *Rev Cubana Farm*, 2002. 36(1): 42-9.
27. Varo J. Gestión estratégica de la calidad en los Servicios sanitarios. Díaz de Santos. Madrid, 1994.
28. Pineault R, Daveluy C. La planificación sanitaria: conceptos, métodos, estrategias, 2ª ed. Masson. Barcelona, 1992.
29. Hervás F. Procedimientos de Inteligencia Artificial en el Estudio de las Enfermedades Infecciosas. Madrid: Ed Díaz de Santos, 1999.
30. Vogt W, Nagel D. Cluster analysis in diagnosis. *Clin Chem*, 1992; 38(2): 182-98.
31. Carrasco JL. El método estadístico en la investigación médica. 5ª ed. Madrid. Editorial Ciencia.
32. Hulley SB, Cummings SR. Diseño de la investigación clínica. Ed. Doyma. Barcelona 1.993
33. Milton, JS., Toscos, JO. Estadística para biología y ciencias de la salud. Madrid: Interamericana M<sub>C</sub> Graw Hill; 2001
34. Eppinger SD. Innovation at the speed of formation. *Har Bus rev*, 2001; 79(1):149-58, 178.
35. Colton, T., Estadística en medicina. Ed. Barcelona. Salvat . 1979.
36. Martín Andrés, A., Luna del Castillo, JD. Bioestadística para las ciencias de la salud. 4ª ed. Madrid: NORMA; 1993
37. Sackett, D.L., Haynes, R.B., Guyatt, G.H., Tugwell, P. Epidemiología clínica . Ciencia básica para la medicina clínica. 2ª ed. Madrid: Médica Panamericana; 1994.
38. Lasky FD, Boser RB. Designing in quality through design control: a manufacturer's perspective. *Clin Chem*, 1997; 43(5):866-72.
39. Kelsey IL, Thompson WD, Evans A. Methods in observational epidemiology. New York. Oxford University Press. 1.986
40. Featherstone D, Brown D, Sanders R. Development of the global measles laboratory network. *J Infect Dis*, 2003; 187(Suppl 1): S264-9.
41. Larson E. Status of practice guidelines in the United States: CDC guidelines as an example. *Prev Med*, 2003; 36(5): 519-24.
42. Shih SC, Bost JE, Pawlson LG. Standardized health plan reporting in four areas of preventive health care. *Am J Prev Med*, 2003; 24(4): 293-300.
43. Williams ID, Bird A. Public perceptions of air quality and quality of life in urban and suburban areas of London. *J Environ Monit*, 2003; 5(2): 253-9.
44. Sweeney J, Brooks AM, Leahy A. Development of the Irish national perception of quality of care survey. *Int J Qual Health Care*, 2003; 15(2): 163-8.
45. Cardew PT. A method for assessing the effect of water quality changes on plumbosolvency using random daytime sampling. *Water Res*, 2003; 37(12): 2821-32.
46. Peterson AT, Vieglais DA, Andreasen JK. Migratory birds modeled as critical transport agents for west Nile virus in north America. *Vector Borne Zoonotic Dis*, 2003; 3(1): 27-37.
47. Philippe P. Sartwell's incubation period model revisited in the light of dynamic modeling. *J Clin Epidemiol* 1994; 47(4):419-433.
48. Piccardi C, Lazzaris S. Vaccination policies for chaos reduction in childhood epidemics. *IEEE Trans Biomed Eng* 1998; 45(5):591-595.
49. Cavalieri LF, Kocak H. Chaos: a potential problem in the biological control of insect pests. *Math Biosci* 1995; 127(1):1-17.
50. Olsen LF, Truty GL, Schaffer WM. Oscillations and chaos in epidemics: a nonlinear dynamic study of six childhood diseases in Copenhagen, Denmark. *Theor Popul Biol* 1988; 33(3):344-370.
51. Koella JC, Doebeli M. Population dynamics and the evolution of virulence in epidemiological models with discrete host generations. *J Theor Biol* 1999; 198(3):461-475.
52. Colimon KH. Fundamentos de Epidemiología. Ed Díaz de Santos. Madrid, 1990.
53. Greenberg RS. Medical Epidemiology. 3ª ed. Lange. New York, 2001.
54. Jenicek Cleroux R. Epidemiología. La lógica de la medicina moderna. Masson. Barcelona, 1996.
55. McMahon B, Trichopoulos D. Epidemiología. 2ª ed. Marban. Madrid, 2001
56. Piédrola Gil y cols. Medicina Preventiva y Salud Pública. 10ª edición. Masson-Salvat. Barcelona 2001.
57. Rey Calero J, Herruzo Cabrera R, Rodríguez Artalejo F. Fundamentos de Epidemiología clínica. Síntesis. Madrid. 1996.
58. Rothman KJ, Greenland S. Modern epidemiology. 2ª ed. Lippincott-Raven. Philadelphia. 1998.
59. Lwanga SK, Lemeshow S. Determinación del tamaño de las muestras en los estudios sanitarios. OMS. Ginebra, 1991.
60. Goldberg M. L'Epidémiologie sans peine. Editions Frison-Roche, 2ème éd. Paris, 1990.
61. Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic research. Principles and quantitative methods. Van Nostrand Reinhold. New York, 1982.
62. Szklo M, Nieto J. Epidemiología intermedia. Conceptos y aplicaciones. Ed Díaz de Santos. Madrid, 2003.
63. Overtveit J, Gustafson D. Using research to inform quality programs. *BMJ*, 2003; 326(7392): 759-61.
64. Mills AE, Rorty MV. Total quality management and the silent patient. *Bus Ethics Q*, 2002; 12(4): 481-504.
65. Murray M, Berwick DM. Advanced access: reducing waiting and delays in primary care. *JAMA*, 2003; 289(8): 1035-40.
66. Korting HC. Total quality management in dermatomycology in Germany. *Mycoses*, 2003; 46(1-2): 2-4.
67. Holl RW, Grabert M. The quality circle: how to improve the outcome of paediatric diabetes care. *Horm Res*, 2002; 57(Suppl 1): 105-9.
68. Tausch BD, Harter MC. Perceived effectiveness of diagnostic and therapeutic guidelines in primary care quality circles. *Int J Qual Health Care*, 2001; 13(3): 239-46.
69. Forster DH, Krause G et al. Can quality circles improve hospital-acquired infection control?. *J Hosp Infect*, 2000; 45(4): 302-10.
70. Henderson AR. Assessing test accuracy and its clinical consequences: a primer for receiver operating characteristic curve analysis. *Ann Clin Biochem*, 1993; 30:521-39.
71. The eighteenth annual Arnold O. Beckman conference in clinical chemistry. Pursuing medical relevance: a crucial role for the clinical laboratory scientist - from a focus on analytical performance to one on medical outcome. *Clin Chem*, 1995; Special issue, 8B.

## Empleo del análisis multivariante en la investigación clínica

72. Wie S, Shanklin CW, Lee KE. A decision tree for selecting the most cost-effective waste disposal strategy in foodservice operations. *J Am Diet Assoc*, 2003; 103(4): 475-82.
73. Mandelbrot B.B. *Fractal Geometry of Nature*. Nueva York: W H Freeman, 1983.
74. Skinner JE, Molnar M, Vybiral T, Mitra M. Application of chaos theory to biology and medicine. *Integr Physiol Behav Sci* 1992; 27(1):39-53.
75. Tsonis PA, Tsonis AA. Chaos: principles and implications in biology. *Comput Appl Biosci* 1989; 5(1):27-32.
76. Philippe P. Chaos and public health: implications for an epidemic. *Can J Public Health* 1992; 83(2):165-166.
77. Olsen LF, Schaffer WM. Chaos versus noisy periodicity: alternative hypotheses for childhood epidemics. *Science* 1990; 249(4968):499-504.
78. Cross S. Fractals in pathology. *J Path* 1997; 182:1-8.
79. Kaitala V, Ylikarjula J, Heino M. Dynamic complexities in host-parasitoid interaction. *J Theor Biol* 1999; 197(3):331-341.
80. Davey HM, Davey CL, Woodward AM, Edmonds AN, Lee AW, Kell DB. Oscillatory, stochastic and chaotic growth rate fluctuations in permissively controlled yeast cultures. *Biosystems* 1996; 39(1):43-61.
81. Lehning M, Shonnard DR, Chang DP, Bell RL. An inversion algorithm for determining area-source emissions from downwind concentration measurements. *Air Waste* 1994; 44(10):1204-1213.
82. Renshaw E. Chaos in biometry. *IMA J Math Appl Med Biol* 1994; 11(1):17-44.
83. Skinner JE. Low-dimensional chaos in biological systems. *Biotechnology (NY)* 1994; 12(6):596-600.
84. Nieminen H, Takala EP. Evidence of deterministic chaos in the myoelectrical signal. *Electromyogr Clin Neurophysiol* 1996; 36(1):49-58.
85. Markham FW. A method for introducing the concepts of chaos theory to medical students. *Theor Med Bioeth* 1998; 19(1):1-4.
86. Griffiths F, Byrne D. General practice and the new science emerging from the theories of chaos and complexity. *Br J Gen Pract* 1998; 48(435):1697-1699.
87. Sedivy R. The potential role of apoptosis (programmed cell death) in a chaotic determined carcinogenesis. *Med Hypotheses* 1996; 46(5):455-457.
88. Lindberg C, Herzog A, Merry M, Goldstein J. Health care applications of complexity science. *Life at the edge of chaos. Physician Exec* 1998; 24(1):6-20.
89. Kirkwood TB, Bangham CR. Cycles, chaos, and evolution in virus cultures: a model of defective interfering particles. *Proc Natl Acad Sci USA* 1994; 91(18):8685-8689.
90. Matioli GT. On the unreliability of the axiom of randomness: the case of random mutations in biology. *Med Hypotheses* 1997; 49(2):181-182.
91. Grenfell BT, Kleczkowski A, Gilligan CA, Bolker BM. Spatial heterogeneity, nonlinear dynamics and chaos in infectious diseases. *Stat Methods Med Res* 1995; 4(2):160-183.
92. Heindl B, Schmidt R, Schmid G, Haller M, Pfaller P, Gierl L et al. A case-based consiliarius for therapy recommendation (ICONS): computer-based advice for calculated antibiotic therapy in intensive care medicine. *Comp Methods Prog Biomed* 1997; 52:117-127.
93. Dharan M. *Control de calidad en los laboratorios clínicos*. Ed Reverté. Barcelona, 1980.
94. *Manual Basic de Control de Qualitat dels Laboratoris Clínic*. Generalitat de Catalunya. Departament de Sanitat y Seguretat Social. Barcelona, 1ª ed, 1986.
95. Trillas E, Alsina C, Terricabras JM. *Introducción a la lógica borrosa*. Ed Ariel. Barcelona, 1995.
96. Varo J. *Gestión estratégica de la calidad en los servicios sanitarios. Un modelo de gestión hospitalaria*. Ed Díaz de Santos. Madrid, 1994.
97. Leguina J. *Fundamentos en Demografía. Siglo XXI*. Madrid, 1989.
98. Martínez-González MA, De Irala J, Faulin FJ. *Bioestadística amigable*. Díaz de Santos. Madrid, 2001.
99. Doménech JM. *Bioestadística. Métodos estadísticos para investigadores*. Herder. Barcelona, 1980.
100. Solanas A, Salafranca LI, Fauquet J, Núñez MI. *Estadística descriptiva en ciencias del comportamiento*. Thomson editores. Madrid, 2004.
101. Pascua M. *Metodología bioestadística para médicos y oficiales sanitarios*. Ed Paz Montalvo, 2ª ed. Madrid, 1974.
102. Sánchez M, Frutos G, Cuesta PL. *Estadística y matemática aplicadas*. Síntesis. Madrid, 1996.
103. Juez P, Díez FJ. *Probabilidad y estadística en medicina. Aplicaciones en la práctica clínica y en la gestión sanitaria*. Ed Díaz de Santos. Madrid, 1997.
104. Marín-Pliego J, Riz-Maya L. *Estadística I. Probabilidad*. Thomson editores. Madrid, 2004.
105. Grossman Z. Mathematical modeling of thymopoiesis in HIV infection: real data, virtual data, and data interpretation small star, filled. *Clin Immunol*, 2003; 107(3): 137-39.
106. Lorenz E.N. Deterministic nonperiodic flow. *J Atmos Sci* 1963; 20:130-141.
107. Aranda J, Fernández JL, Jiménez J, Morilla F. *Fundamentos de lógica matemática*. Ed Sanz y Torres, 2ª ed. Madrid, 2001.
108. Moldofsky H. Sleep and the immune system. *Int J Immunopharmacol*, 1995; 17(8): 649-54.
109. Whitmire D, Cornelius L, Whitmire P. Effects of food on ethanol metabolism. *Drug Metabol Drug Interact*, 2002; 19(2): 83-96.
110. Rubinstein RY. *Simulation and the Monte Carlo method*. John Wiley & Sons, NY, 1981.
111. Kerry SM, Bland JM. The intraclass correlation coefficient in cluster randomisation. *BMJ*, 1998; 316(7142): 1455.
112. Hervás F. *Hospital empresa y sanidad asistencial. Calidad y rentabilidad*. Ed Díaz de Santos. Madrid, 2007.