

Educación médica

Bioestadística para todos. Tercera parte: el análisis estadístico de la investigación clínica

Dr. ARMANDO GARSD*

ARCH ARG PEDIATR / 1998 / VOL. 96:
258

1. INTRODUCCION

En la primera de tres entregas hemos pasado revista al material necesario para hacer estadísticas.¹ En la segunda entrega nos hemos referido al rol de la estadística en la investigación clínica.² Finalmente, en esta última parte, presentamos ciertos conceptos básicos para el análisis estadístico de esas investigaciones.

Una vez concluida la fase experimental del estudio clínico se llega a la quintaesencia de la investigación, esto es, a los datos resultantes. Rara vez ocurre que estos resultados sean tan contundentes que hablen por sí mismos y no requieran confirmación estadística. La mayoría de las investigaciones clínicas tienen un clímax más modesto y precisan de un análisis que discierna efectos válidos de resultados espurios o atribuibles al azar.

Decíamos en la primera entrega que en algunos países la estadística del ensayo clínico está en parte regulada por ley. Esta regulación es parcial porque no es posible normatizar íntegramente el arte y ciencia del análisis de datos. La cantidad y variedad de problemas que se presentan en la práctica es de tal magnitud que sería imposible prescribir una guía para resolverlos. A su vez, la disciplina estadística avanza permanentemente y el advenimiento de computadoras de alta capacidad permite efectuar más y más pruebas estadísticas exactas que de otro modo deberían resolverse por aproximación.

Por lo tanto, en esta entrega nos limitamos a formular algunas sugerencias básicas, de ningún modo exhaustivas, de cómo orientar correctamente el análisis. Más allá, el lector deberá buscar literatura especializada o consejo profesional. Por

brevidad supondremos que los datos ya han sido correctamente ingresados en una planilla electrónica de cálculo estadístico.

2. CONTABILIZACION DE PACIENTES

Contrariamente a lo que podría suponerse a primera vista, el análisis de los resultados de un estudio no empieza con la aplicación espectacular de un modelo o prueba estadística, sino con un simple ejercicio contable, en el cual se da debida cuenta de qué pasó con cada paciente que participó del estudio. De este desglose (*disposition of patients*) surgen los subgrupos, o líneas de abordaje, del análisis estadístico ulterior. La contabilización resultante se suele presentar en forma de tabla o de diagrama de flujo.

Considérese el caso de un estudio en consultorio externo hospitalario, en paralelo, doble ciego, doble placebo, con verificación a 15 días, para investigar la equivalencia terapéutica de lindano (aplicación externa) e ivermectina (vía oral), ambos en una única dosis, para el tratamiento de pacientes con escabiosis.

La *Tabla 1* describe la contabilización de pacientes a los 15 días de estudio. La tabla es de naturaleza descriptiva y aunque no parezca de gran profundidad conceptual tampoco es trivial. La misma revela deserción diferencial en el grupo en ivermectina (24% vs. 8% en lindano, $p=0,002$). (Estos datos son hipotéticos).

TABLA 1
Contabilización de pacientes a 15 días
en un estudio de
ivermectina vs. lindano en escabiosis

Tratamiento	Ingresaron	Desertaron	Completaron	Curaron	Fracasaron
Ivermectina	100	24	76	56	20
Lindano	100	8	92	46	46

* Profesor de Estadística. Universidad de Buenos Aires.
Correspondencia: Ramsay 2498, piso 3. (1428) Buenos Aires.

téticos. Los valores de p se refieren a la prueba de la significatividad estadística de la diferencia entre dos proporciones binomiales, vía prueba de χ^2). La deserción aparentemente fue seguida de mayor eficacia de ivermectina entre los que completaron el estudio (por deserción [*dropout*] se entiende el caso del paciente que a partir de un cierto momento abandona el estudio y del que no se tiene ninguna información ulterior). En estos casos ningún análisis único de eficacia resulta satisfactorio. El primer paso estadigráfico consiste en plantear por lo menos dos niveles de análisis de eficacia, a saber, considerando todos los pacientes aleatorizados (*all randomized patients*) y aquellos que completaron el estudio (*completers*).

Así, si se toman en cuenta todos los pacientes aleatorizados, la diferencia en eficacia es de 56% (56/100) en ivermectina vs. 46% (46/100) en lindano. Nótese que dentro de este lineamiento los pacientes que desertan del estudio se contabilizan efectivamente como fracasos terapéuticos. Así planteadas las cosas, la diferencia no es estadísticamente significativa ($p=0,154$) y lleva a efectuar pruebas formales de equivalencia de eficacia.³

Por otro lado, si se considera que las deserciones no son informativas, sino completamente al azar y se procede al análisis de los que completaron los 15 días de estudio, las diferencias son 73,7% (56/76) vs. 50% (46/92), cuya significatividad a favor de ivermectina es de $p=0,002$. (informalmente, la deserción es completamente al azar si la probabilidad de que un paciente deserte no depende de la condición de su escabiosis previa o posterior a la deserción).

Incluso existe un planteo más avezado que consiste en considerar a las deserciones como éxitos terapéuticos, o sea, presuponer que los pacientes que desertaron son pacientes que se curaron pero que no volvieron al hospital a reportarlo.

El consenso de la comunidad estadigráfica es que, en general, cuanto más se parecen entre sí los resultados de los análisis de esos subgrupos, más sólida es la confianza en las conclusiones del estudio.⁴

Como vemos, la contabilización de pacientes es útil porque revela aspectos inusitados de la evolución del estudio. El tema de las deserciones diferenciales es complejo pero vital para el buen análisis.⁵

3. TEORIA Y PRACTICA ESTADISTICA

Toda ciencia experimental procura reducir los fenómenos en estudio a un conjunto básico de principios que expliquen todo lo observado. La

física y la química son ejemplos de esa reducción. La moderna disciplina estadística procura, dentro de su ámbito, una reducción similar. La síntesis resultante se manifiesta, en primer término, en los modelos estadísticos y sus pruebas de asociados. Pero son los conceptos de reducción, más que las pruebas en sí mismas, los que guían y simplifican el análisis práctico de los datos.

Existen varios puntos de unificación estadística, algunos bien complejos.⁶ No obstante, aquí nos limitamos a ejemplificar uno, interesante por su belleza y simplicidad.

Considérese el caso de un estudio comparativo, aleatorizado, para evaluar la seguridad y tolerancia de stavudina (d4T) vs. zidovudina (ZDV, AZT) en niños HIV⁺. Uno de los eventos de interés fue la aparición en estos pacientes de toxicidad hematológica de grado 3 o mayor, en algún momento del estudio. Los resultados se presentan en la *Tabla 2*, con la proporción de pacientes con toxicidad y sin ella indicada entre paréntesis.⁷

Ahora bien, si a cada paciente en el grupo en stavudina se le asigna un número 1 si el paciente hizo el evento de interés o el número 0 si no lo hizo, el resultado de esta binarización es una sucesión de ceros y unos, con un total de 93 unos y 12 ceros. Se puede demostrar fácilmente que el promedio aritmético de esa sucesión es 0,886. O sea, la proporción de eventos en el grupo en stavudina. Vemos entonces que lo que habitualmente llamamos proporción tiene una interpretación como promedio aritmético de variables binarias. Esto nos permite, en principio, usar todo el valioso arsenal de conocimientos sobre estadísticas de promedios a casos de proporciones y porcentajes. Si procedemos de igual manera con el grupo en zidovudina, la diferencia de proporciones de eventos entre uno y otro tratamiento, 0,886 vs. 0,785, se transforma en una diferencia de promedios, susceptible de análisis estadístico de promedios, por ejemplo vía prueba de Student según el enfoque de D'Agostino.⁸ Por extensión, los análisis que indirectamente involucran proporciones, por ejemplo los estudios

TABLA 2
Presencia de evento hematológico
en un estudio de
stavudina vs. zidovudina en pacientes HIV⁺

	Evento		Total
	Sí	No	
Stavudina	93 (0,886)	12 (0,114)	105 (1)
Zidovudina	84 (0,785)	23 (0,215)	107 (1)

de sobrevivencia,⁹ y los que modelizan probabilidades, por ejemplo los logísticos,¹⁰ también admiten, en principio, una representación estadística como promedios.

4. TIPOS DE ANALISIS

En teoría, la estadística es la ciencia de las distribuciones probabilísticas de los datos así como de los parámetros que las caracterizan. La contraparte práctica es la exploración visual de los datos, el estudio de los histogramas, los promedios y otros valores muestrales. Los puentes entre la teoría y la práctica se establecen mediante las pruebas de hipótesis, los valores *p* de significatividad, los intervalos de confianza y otros procedimientos más especializados.

El caso más conocido es el de la comparación de dos promedios poblacionales normales e independientes. En rigor uno compara, a partir de los promedios de las muestras, dos distribuciones normales, que bajo ciertas suposiciones añadidas se les permite diferir sólo en sus promedios poblacionales. La prueba *t* de Student es la expresión práctica de este enfoque. Los llamados procedimientos no paramétricos no precisan del supuesto de normalidad. En el caso considerado, se comparan dos distribuciones a través de la prueba de Mann-Whitney.¹¹

Estos conceptos pueden extenderse a la comparación de varias poblaciones, bajo un comporta-

miento de promedios arbitrario, por ejemplo en el análisis de variancia, o suponiendo que esas poblaciones se alinean en sus promedios poblacionales, o sea que obedecen a una regresión lineal, simple o múltiple.¹²

Cuando se consideran varias distribuciones simultáneas se entra en el mundo del análisis estadístico multivariado. El coeficiente simple de correlación es el más sencillo y conocido de los parámetros provenientes de distribuciones multivariadas, en este caso particular llamada bivariada. En otros problemas la situación es más intrincada.

Considérese el caso de una investigación neonatológica de la presión de succión en función del tiempo en 6 lactantes, medida cada 30 segundos (en unidades estandarizadas de energía) en 12 ocasiones, comunes a todos los lactantes. Los datos se grafican en el *Gráfico 1*.¹³ El detalle por lactante complica la interpretación del fenómeno al revelar la presencia de algunos individuos con escaso decrecimiento o incluso reinicio de la succión. Se ajusta una regresión lineal para cada unidad experimental, el lactante, y se advierte que los resultados son mayormente significativos pero con cierta variación entre las rectas ajustadas. Se desea conocer la significatividad del descenso durante esos primeros 6 minutos de succión, cuál es en definitiva la recta correcta, cómo manejar varias regresiones simultáneas y cuánto de la variabilidad total corresponde a

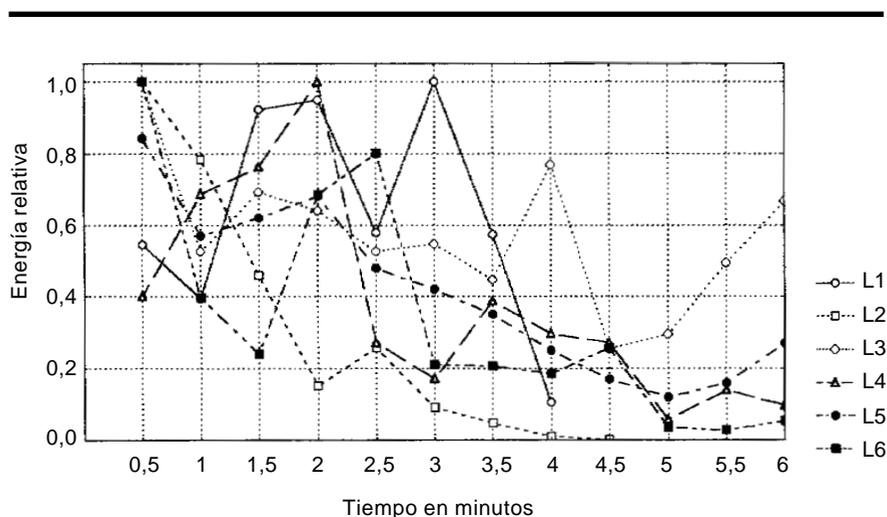


GRÁFICO 1
Energía de succión en función del tiempo
individualizada por lactante (L)

diferencias en las regresiones individuales.

Para abordar el problema deben reconocerse sus fuentes de variación aleatoria, a saber, una variación aleatoria en el comportamiento individual, otra de error experimental y, finalmente, una suerte de memoria, o proceso estocástico, conectando cada error con los anteriores. El tratamiento estadístico más satisfactorio del problema planteado implica el ajuste de un modelo lineal general, de naturaleza multivariada.¹⁴ No obstante, este enfoque multidimensional tiene varias limitaciones prácticas que deben respetarse a la hora de aconsejar el tratamiento estadístico. El método multivariado es complejo de seguir e involucra un tratamiento matricial elaborado. A su vez, los programas estadísticos de computación (*software*) no siempre contienen métodos multivariados, o requieren adaptaciones especiales. Algunos métodos multivariados son menos potentes que sus contrapartes univariadas o requieren ciertas condiciones que el problema en consideración no reúne, por ejemplo, que el número de medidas repetidas sea menor que el número de unidades experimentales. Sin ir más lejos, el ejemplo planteado no cumple este requisito. La pericia estadística consiste en reducir la complejidad del problema a un modelo simple, aproximadamente correcto, que se pueda entender intuitivamente. Felizmente, tales modelos existen. Pero hay que encontrarlos.

5. LA VARIABILIDAD COMO OBJETO DE ANALISIS

No obstante, las distribuciones pueden diferir en formas que trascienden los promedios, por ejemplo porque representan fenómenos de diferente variación. Considérese un ensayo efectuado en 100 sujetos, en los cuales el colesterol sérico fue medido en cada uno 10 veces con un analizador estándar (Ektachem 700) y 10 veces con un método de referencia (Cobas Bio) calibrado por el *Centers for Disease Control and Epidemiology*. El objetivo del ensayo fue determinar si la variación individual bajo ambos métodos era la misma, independientemente de variaciones de nivel. Un aspecto curioso de este estudio fue que las observaciones originales no estaban disponibles por razones de confidencialidad. No obstante, se disponía de dos variancias muestrales por individuo, una por método, provenientes de esas observaciones. O sea, que se disponía de 100 pares de variancias, dos por individuo. Sobre esta base los estadígrafos idearon recientemente un método para probar, y en efecto probaron que la variación intraindividual

del método estándar era significativamente mayor que la del método de referencia.¹⁵

Siempre es útil entonces, enfocar, tempranamente y correctamente el tipo de problema que se va a analizar en forma estadística.

6. COMPARACIONES MULTIPLES

El denominado problema de las comparaciones múltiples es una de las causas más frecuentes de cuestionamiento en la literatura de análisis estadísticos de otro modo perfectamente válidos.¹⁶

En su forma más simple, el problema se origina porque cuando se comparan entre sí varias poblaciones, el valor de significatividad estadística de cada comparación no es una medida válida del error total (la expresión "error o significatividad de una comparación" aquí se refiere al error estadístico de tipo 1, en este caso, la probabilidad de rechazar la igualdad de placebo y tratamiento cuando en realidad para esa comparación específica [o *endpoint*] son iguales. Por "error total", nos referimos a la probabilidad de cometer al menos un error de comparación entre todas las comparaciones planeadas). El problema de los puntos múltiples de análisis (*multiple endpoints*) es otra variante de la misma dificultad.

El siguiente ejemplo surge de un ensayo clínico, multicéntrico, doble-ciego y en paralelo efectuado para evaluar la eficacia y seguridad de una droga experimental en pacientes asmáticos.¹⁷ Se planearon cuatro puntos de análisis (*endpoints*) a saber, FEV₁, PEF_R, puntaje de síntomas y cantidad de medicación adicional requerida por el paciente. Se puede demostrar que si el nivel de significatividad de cada una de las cuatro comparaciones entre tratamiento experimental y placebo se toma como $\alpha=0,05$, el error total es mayor de 0,05. Por ejemplo, la diferencia entre placebo y medicación para el puntaje sintomático resultó en un p de 0,027.¹⁸ Sin embargo, si se ajusta la significatividad globalmente, por ejemplo mediante el ajuste de Bonferroni, ese resultado no es estadísticamente significativo. O sea, con ajuste de significatividad, el puntaje sintomático de los pacientes tratados con la medicación experimental es similar al de aquellos que recibieron placebo.

Vemos entonces que el análisis estadístico procura ajustar la significación hacia un marco de sobriedad, de manera de evitar resultados significativos espurios, surgidos por chance, a fuerza de tomar muchos puntos de análisis.

7. COMENTARIO FINAL

Hemos titulado esta serie de entregas "Bioes-

estadística para todos". Cabe preguntarse hasta qué punto hemos justificado ese título. Ofrecemos la siguiente respuesta.

De todo lo expuesto, resulta evidente que las buenas estadísticas en la investigación clínica no son obras de estadísticos geniales sino el producto de un trabajo interdisciplinario, interactivo y eficaz entre todos los investigadores participantes. Como consecuencia, los resultados de esas buenas estadísticas son entendidos, especialmente, por aquellos que no entienden de metodología estadística.

Por eso, las buenas estadísticas están al alcance de todos. ■

BIBLIOGRAFIA

1. Garsd A. Bioestadística para todos. Primera parte: Las herramientas del oficio estadigráfico. Arch Arg Pediatr 1997; 95: 205-207.
2. Garsd A. Bioestadística para todos. Segunda parte: El rol de la estadística en la investigación clínica. Arch Arg Pediatr 1997; 95: 448-451.
3. Blackwelder WC. Proving the null hypothesis in clinical trials. Control Clin Trials 1982; 3: 345-353.
4. International Conference on Harmonisation. 1997. E9: Statistical principles for clinical trials. Documento disponible en Internet (<http://www.pharmweb.net/pwmirror/pw9/ifpma/ich5.html/>) o impreso en The Federal Register 1997, May 9.
5. Little RJ, Rubin DB. Statistical analysis with missing data. New York: John Wiley, 1987.
6. Cox DR, Hinkley DV. 1974. Theoretical statistics. London: Chapman and Hall 1974; 36-48.
7. Kline MW. The AIDS Clinical Trials Group 240 Team. A randomized comparative trial of stavudine (d4T) versus zidovudine (ZDV, AZT) in children with human immunodeficiency virus infection. Pediatrics 1998; 101: 214-220.
8. D'Agostino RB, Chase W, Belanger A. The appropriateness of some common procedures for testing the equality of two independent binomial proportions. Am Statistician 1988; 42: 198-202.
9. Lee ET. Statistical methods for survival data analysis. 2nd ed. John Wiley. New York: John Wiley, 1992.
10. Hosmer Jr., Hosmer DW, Lemeshow S. Applied logistic regression. New York: John Wiley, 1989.
11. Conover WJ. Practical nonparametric statistics. New York: John Wiley, 1980.
12. Neter J, Wasserman W, Kutner MH. Applied linear statistical models. Second edition. Irwin, Homewood, Illinois, 1985.
13. Voloschin LM, Athabe O, Olivé H et al. A new tool for measuring the suckling stimulus during breast-feeding in humans: the orokinogram and the Fourier series. J Reproduct Fertil. En prensa, 1998.
14. Crowder MJ, Hand DJ. Analysis of repeated measures. New York: Chapman and Hall, 1993.
15. Padmanabhan AR, Chinchilli VM, Babu GJ. Robust analysis of within-unit variances in repeated measurement experiments. Biometrics 1997; 53: 1520-1526.
16. Pocock SJ, Hughes DM, Lee RJ. Statistical problems in the reporting of clinical trials. N Engl J Med 1987; 317: 426-432.
17. National Asthma Education Program Expert Panel Report 1991. Guidelines for the diagnosis and management of asthma. Bethesda, MD. National Institutes of Health Publication N° 91-3042.
18. Zhang J, Quan Q et al. Some statistical methods for multiple endpoints in clinical trials. Control Clin Trials 1997; 18: 204-221.