

IKERKETA EPIDEMIOLOGIKORAKO
ESTADISTIKA-AZTERPIDEAK



METODOS ESTADISTICOS
PARA LA INVESTIGACION EPIDEMIOLOGICA

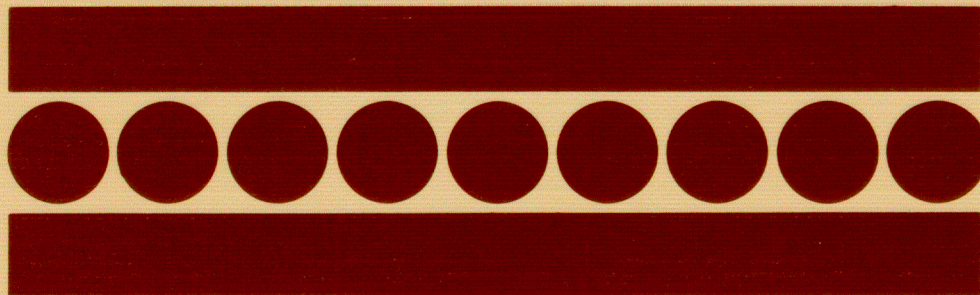


METHODES STATISTIQUES
POUR L'INVESTIGATION EPIDEMIOLOGIQUE



STATISTICAL METHODS
FOR EPIDEMIOLOGY RESEARCH

L.C. SILVA



Lanketa / Elaboración:

Euskal Estatistika-Erakundea /
Instituto Vasco de Estadística

Argitalpena / Edición:

Euskal Estatistika-Erakundea /
Instituto Vasco de Estadística
C/ Dato, 14-16 - 01005 Vitoria-Gasteiz

Botaldia / Tirada

1.000 ejemplares
IV-1988

© **Euskadiko K.A.ko Administrazioa**
Administración de la C.A. de Euskadi

Fotokonposaketarako tratamendu informatikoa:
Tratamiento informático de fotocomposición:

Fotocomposición Didot, S.A.
Nervión, 3-5.º BILBAO

Inprimaketa eta koadernaketa /
Impresión y encuadernación:

Itxaropena, S.A.
Araba kalea, 45 - Zarautz (Gipuzkoa)

Lege-gordailua / Depósito legal: S.S. 605/92

ISBN: 84-7749-019-8
ISBN: 84-7542-127-10

NAZIOARTEKO ESTADISTIKA
·MINTEGIA EUSKADIN

1987

SEMINARIO INTERNACIONAL
DE ESTADISTICA EN EUSKADI

**IKERKETA EPIDEMIOLOGIKORAKO
ESTADISTIKA-AZTERPIDEAK**



**METODOS ESTADISTICOS
PARA LA INVESTIGACION EPIDEMIOLOGICA**



**METHODES STATISTIQUES
POUR L'INVESTIGATION EPIDEMIOLOGIQUE**



**STATISTICAL METHODS
FOR EPIDEMIOLOGY RESEARCH**

L.C. SILVA

KOADERNOA **14** CUADERNO



AURKEZPENA

Estatistikako Mintegi Internazionalak sustatzean, hainbat xedekin bete nahi luke Euskal Estatistika Erakundeak hala nola:

- Unibertsitatearekiko eta, Estatistika Sailarekiko lankidetzaz bultzatu.
- Funtzionario, irakasle, ikasle eta estatistikaren alorrean interesaturik leudekeen guztien birziklapen profesionala erraztu.
- Estatistikako alorrean eta mundu-mailan irakasle prestu eta abangoardiako ikerlari diren pertsonaiak Euskadira ekarri, guzti horrek zuzeneko harremanei eta esperientzien eza-gupenei dagokienez suposatzen duen ondorio positiboarekin.

Iharduketa osagarri bezala eta interesaturik leudekeen ahalik eta pertsona eta Erakunde gehienetara iristearren, ikastaro hauetako txostenak argitaratzea erabaki da, horrela gure Herrian gai honi buruzko ezagutza zabaltzen laguntzeko asmoarekin.

Vitoria-Gasteiz, 1988ko martxoa
JOSE LUIS NARVAIZA SOLIS
Zuzendari Orokorra

PRESENTACION

Al promover los Seminarios Internacionales de Estadística, el Instituto Vasco de Estadística pretende cubrir varios objetivos:

- Fomentar la colaboración con la Universidad y en especial con los Departamentos de Estadística.
- Facilitar el reciclaje profesional de funcionarios, profesores, alumnos y cuantos puedan estar interesados en el campo estadístico.
- Traer a Euskadi a ilustres profesores e investigadores de vanguardia en materia estadística, a nivel mundial, con el consiguiente efecto positivo en cuanto a la relación directa y conocimiento de experiencias.

Como actuación complementaria y para llegar al mayor número posible de personas e Instituciones interesadas, se ha decidido publicar las ponencias de estos Cursos, para contribuir así a acrecentar el conocimiento sobre esta materia en nuestro País.

Vitoria-Gasteiz, marzo, 1988
JOSE LUIS NARVAIZA SOLIS
Director General

PRESENTATION

L'Institut Basque de la Statistique se propose d'atteindre plusieurs objectifs par la promotion des Séminaires Internationaux de la Statistique:

- Encourager la collaboration avec l'Université et spécialement avec les départements de la statistique.
- Faciliter le recyclage professionnel des fonctionnaires, professeurs, élèves, et tous ceux qui pourraient être intéressés par la statistique.
- Inviter en Euskadi des professeurs mondialement renommés et des chercheurs de premier ordre en matière de Statistique avec tout ce que cela pourrait entraîner comme avantage dans les rapports et l'échange d'expériences.

En outre, il a été décidé de publier les exposés de ces rencontres afin d'atteindre le plus grand nombre de personnes et d'institutions intéressées, et pour contribuer ainsi à développer dans notre pays les connaissances sur cette matière.

Vitoria-Gasteiz, mars, 1988

JOSE LUIS NARVAIZA SOLIS
Directeur Général

PRESENTATION

In promoting the International Seminars on Statistics, the Basque Statistical Office is attempting to achieve a number of objectives:

- Encourage joint working with the Basque University and, in particular, with its Department of Statistics.
- Facilitate the in-training of civil servants, teachers and students and of all those interested in the field on statistics.
- Bring to Euskadi distinguished academics and researchers in the front line of statistics work, at a world-wide level, with all the benefits that this will bring through direct contacts and the interchange of experiences and ideas.

As an additional step it has been decided to publish in advance the papers to be presented at these courses. This is in order that as many interested people and institutions as possible are made aware. In this way we hope to contribute to the growth and awareness concerning this topic in our country.

Vitoria-Gasteiz, march, 1988

JOSE LUIS NARVAIZA SOLIS
Director General

SARRERA

Euskal Estatistika-Erakundeak eta Euskal Herriko Unibertsitateak antolatuta, eta Euskadiko V. Nazioarteko Estatistika-Miantegiaren barnean, L.C. Silva-k «Ikerketa Epidemiologikorako Metodo Estatistikoak» gaiaren inguruan emandako ikastaroa laburbiltzen du liburu honek. V. Mintegi horretan irakasle/ponente izan dira baita honako beste jaun-andere hauek ere: E. Sanromá, Barcelonako Unibertsitateko Irakasle eta Ikerlaria, «Ekonomia Erregionala. Azterketa ekonomikoa eta metodo kuantitatiboak» gaiarekin; M. Gettler Summa, Dauphine Unibertsitateko Ikerlaria eta L. Hauesler, Parisko CREDOCeke Ikerlaria, «Sailkapen Automatikoko Metodoak» gaiari buruzko hitzaldi batekin, eta F. Saltor, Kataluniako Unibertsitate Politeknikoko Lengoaia eta Sistema Informatikoen Katedraduna, «Datu-Biltegiak» gaiarekin.

INTRODUCCION

Este libro resume el curso que sobre «Métodos estadísticos para la investigación epidemiológica», ha impartido L.C. Silva, dentro del V Seminario Internacional de Estadística en Euskadi, organizado por el Instituto Vasco de Estadística y la Universidad del País Vasco. Este V Seminario cuenta además con la participación de E. Sanromá, Profesor e Investigador de la Universidad de Barcelona, con el tema «Economía Regional. Análisis económico y métodos cuantitativos», M. Gettler Summa, Investigadora de la Universidad Dauphine y L. Hauesler, Investigadora del CREDOC, de París, con un curso sobre «Métodos de Clasificación Automática», y F. Saltor, Catedrático de Lenguajes y Sistemas Informáticos, de la Universidad Politécnica de Cataluña, con el tema «Bases de Datos».

INTRODUCTION

Ce livre resume le cours «Méthodes Statistiques pour l'Investigation épidémiologique» fait par L.C. Silva dans le V Séminaire International de la Statistique au Pays Basque, organisé par l'Institut Basque de la Statistique et l'Université du Pays Basque. Ce V Séminaire compte aussi avec la participation de E. Sanromá, Professeur et Investigateur de l'Université de Barcelone, avec le sujet «Economie Régionale, Analyse Economique et méthodes quantitatives», M. Gettler Summa, Investigatrice de l'Université Dauphine et L. Hauesler, Investigatrice du CREDOC à Paris, avec le sujet «Méthodes de Classification automatique» et F. Saltor, Professeur de Langages et Systèmes Informatiques de l'Université Polytechnique de Catalunya, avec le sujet «Bases de Données».

INTRODUCTION

This book summarizes the course on «Statistical methods for epidemiology research», given by L.C. Silva, during the 5th International Seminar on Statistics, organized by the Basque Statistical Institute and the University of the Basque Country. This 5th Seminar also counted on the participation of E. Sanromá, Professor and Researcher at the University of Barcelona, with the theme «Regional Economics: Economic analysis and quantitative methods», M. Gettler Summa, Researcher at the Dauphine University and L. Hauesler, Researcher at the CREDOC, from Paris, with the course «Automatic classification methods», and F. Saltor, Professor of Computer Languages and Systems of the Politecnia University of Catalonia, with the theme «Data Bases».

BIOGRAFIA

Luis Carlos Silva, Habanako Medikuzientzi Institutu Goreneko titulodun ikerlaria da eta Osasun Publiko Fakultateko irakaslea. Langiketa, Saio Biologiko eta Ikerketaren Metodologiako goi mailako kurtsoak eman ditu Osasun Publiko Ministerioaren Institutuetan. Ann Arbor Michiganeko Unibertsitateko Sampling Program for Foreign Statisticians-eko egresatua da (1979). Pragan Carlos-eko Zientzi Fisiko Matematiko Unibertsitatean bere doktoregoa lortu zuen (1982). Cubatar Matematika Elkartearen bazkide fundatzailea da, baita ere Cubatar Osasun Administrazio eta Inkesta Estatistikalarien Nazioarteko Elkarte (I.A.S.S.)-ko bazkidea. Anitz herrietako aldizkari espezializatuetan artikulak eta «Langinketa-teknika eta bere aplikazioak ikerketa-sozialean» liburua idatzi ditu.

Luis Carlos Silva es investigador titular del Instituto Superior de Ciencias Médicas de la Habana y docente de su Facultad de Salud Pública. Ha impartido Muestreo, Ensayo Biológico y Metodología de la Investigación a nivel postgradual en Institutos de Investigación del Ministerio de Salud Pública de Cuba y en España. Es egresado del Sampling Program for Foreign Statisticians de la Universidad de Michigan, Ann Arbor (1979) y alcanzó su doctorado en Ciencias Físico Matemáticas en la Universidad de Carlos (Praga, 1982). Es miembro fundador de la Sociedad Cubana de Matemáticas, Sociedad Cubana de Administración de Salud y miembro de la Asociación Internacional de Estadísticos de Encuestas (I.A.S.S.). Ha publicado numerosos artículos en revistas especializadas de varios países y el libro «Las técnicas muestrales y su aplicación en la investigación higiénico-social».

Luis Carlos Silva est investigateur titulaire de l'Institut Supérieur de Sciences Médicales de l'Havane et enseignant de sa Faculté de Santé Publique. Il a enseigné Echantillonnage, Essai Biologique et Méthodologie dans des Instituts d'Investigation du Ministère de Santé Publique de Cuba et d'Espagne. Il est gradué du Sampling Program for Foreign Statisticians à l'Université de Michigan, Ann Arbor (1979) et atteint son doctorat en Sciences Physiques et Mathématiques à l'Université de Carlos (Prague, 1982). Il est membre fondateur de la Société Cubaine de Mathématique, Société Cubaine de l'Administration de la Santé et membre de l'Association Internationale de Statisticiens d'Enquêtes (I.A.S.S.). Il a publié de nombreux articles dans des revues spécialisées dans de nombreux pays et le livre «Les Techniques d'échantillonnage et leur application dans l'investigation hygiénico-sociale».

Luis Carlos Silva, is Official Researcher at the Higher Institute of Medical Sciences of Havana and Professor at its Public Health Faculty. He has lectured on Sampling, Biological Testing and Research Methodology at postgraduate level in Research Institutes of the Ministry of Public Health in Cuba and Spain. He graduated in the Sampling Programme for Foreign Statisticians at Michigan University, Ann Arbor (1979) and obtained his doctorate in Physic-Mathematics Sciences in Carlos University (Prague, 1982). He is a founder member of the Cuban Mathematics Society, the Cuban Society for Health Administration and a member of the International Association of Survey Statisticians (I.A.S.S.). He has published numerous articles in specialist magazines in various countries and the book «Sampling techniques and their application in Social-Hygenic Research».



AURKIBIDEA

1. Proba diagnostikoen ebaluazio estatistikoari buruzko oharrrak	15
1.1. Sarrera	15
1.2. Oinarrizko kontzeptu eta definizio batzuk	16
1.3. Estimazio-arazoak	19
1.4. Diseinu alternatiboen konparazioa	22
1.5. Azken oharrrak	27
Bibliografia	29
I Eranskina	31
II Eranskina	33
2. Inkesta Seroepidemiologikoen ebaluazioa	35
2.1. Sarrera	35
2.2. Antigorputzen kopuruaren neurketa	36
2.3. Estimazio-arazoak	36
2.4. Irtenbide alternatibo bat	37
2.5. Prozeduraren ebaluazioa	38
2.6. Simulazioaren emaitzak	39
Bibliografia	41
Eranskina	43
3. Metodo estatistikoak epidemiologian ...	45
3.1. Sarrera	45
3.2. Ikerketa epidemiologikoen sailkapena	46
3.3. Neurketa ikerketa epidemiologikoan	48
3.4. Efektuen neurketa	49
3.5. Estimazio-arazoak	51
3.6. Konfidantza-tarteak	53
3.7. Hautespren-arazoei buruzko oharrrak	53
3.8. Kausalitate-erlazioei buruzko iruzkin bat	54
3.9. Faktore nahastaitleen kontrola	55
3.10. Erregresio logistikoa	57
Bibliografia	61

INDICE

1. Apuntes sobre la evaluación estadística de las pruebas diagnósticas	15
1.1. Introducción	15
1.2. Algunos conceptos y definiciones básicas	16
1.3. Problemas de estimación	19
1.4. Comparación de diseños alternativos	22
1.5. Apuntes finales	27
Bibliografía	29
Anexo I	31
Anexo II	33
2. Evaluación de Encuestas Seroepidemiológicas	35
2.1. Introducción	35
2.2. Medición de títulos de anticuerpos	36
2.3. Problema de estimación	36
2.4. Una solución alternativa	37
2.5. Evaluación del procedimiento	38
2.6. Resultados de la simulación	39
Bibliografía	41
Anexo	43
3. Sobre los métodos estadísticos en epidemiología	45
3.1. Introducción	45
3.2. Clasificación de estudios epidemiológicos	46
3.3. La medición en la investigación epidemiológica	48
3.4. Medición de efectos	49
3.5. Problemas de estimación	51
3.6. Intervalos de confianza	53
3.7. Notas sobre problemas de selección	53
3.8. Un comentario sobre relaciones de causalidad	54
3.9. Control de factores confusores ...	55
3.10. La regresión logística	57
Bibliografía	61

**METODOS ESTADISTICOS
PARA LA INVESTIGACION EPIDEMIOLOGICA**

1. APUNTES SOBRE LA EVALUACION ESTADISTICA DE PRUEBAS DIAGNOSTICAS

1.1. Introducción

El número de pruebas de laboratorio que se utilizan en la práctica médica contemporánea es extremadamente voluminoso y tiende a incrementarse exponencialmente. Tales pruebas, unidas a un vasto arsenal de procedimientos médicos, constituyen importantes guías para el diagnóstico de las enfermedades.

La manera de clasificar las enfermedades es en sí misma polémica y en cierto sentido carente de uniformidad (ver Susser, 1973) ya que algunas se definen a partir de sus determinantes o causas —por ejemplo la sífilis o la influenza— mientras que otras se definen a partir de sus manifestaciones, tales como el retraso mental, la enfermedad coronaria o ciertos desórdenes metabólicos. No obstante, la nosología —ciencia que se ocupa de la clasificación, descripción y diferenciación de las enfermedades— ha alcanzado un notable desarrollo, de suerte que la inmensa mayoría de las entidades poseen perfiles nítidos y vías relativamente seguras para su identificación.

Mucho se ha escrito en torno a las pruebas diagnósticas y a su eficacia real como elementos contribuyentes a la correcta clasificación diagnóstica de un paciente bajo análisis. Una revisión exhaustiva de lo realizado hasta mediados de la década del 70 puede hallarse en el libro de Galen y Gambino (1975).

En principio, la calidad del proceso diagnóstico resulta importante tanto a nivel individual

como poblacional. Se está en el primer caso cuando se trata de identificar si un sujeto está afectado o no por cierta dolencia, de lo cual puede derivarse la definición de una u otra conducta terapéutica. El problema se presenta a nivel poblacional cuando se intenta establecer la prevalencia de una enfermedad en un conglomerado humano; su importancia estriba en que tal conocimiento contribuye a la planificación de recursos por parte del Sistema de Salud Pública, al conocimiento del estado de salud de la población y al diseño eficiente de estudios epidemiológicos.

La distinción básica entre ambos casos reposa en que si bien a nivel individual el interés principal concierne a la precisión del diagnóstico de cada sujeto, en el segundo lo que se pretende es evitar el sesgo con que pudiera hacerse la estimación del parámetro poblacional.

La sensibilidad y la especificidad de las pruebas diagnósticas que se definen más adelante constituyen sin duda los dos indicadores más difundidos para evaluar su eficiencia. Aun siendo conceptos relativamente sencillos, no es difícil detectar confusiones tanto con su interpretación y alcance como con la relación que guardan con otros parámetros de interés. Tales problemas son particularmente acusados entre el personal llamado a usufructuarlas en tanto herramientas de su trabajo. En efecto, Schechter y Sheps (1985) señalan: «Desafortunadamente algunos, a partir de su falta de comprensión, tienden a concluir que esos conceptos son esotéricos y clínicamente irrelevantes. Nada podría estar más alejado de la verdad.»

Paralelamente, existen no pocas líneas en las que se han producido interesantes desarrollos y en cuyo contexto pueden identificarse problemas abiertos, susceptibles de promisorias continuaciones.

Los apuntes que conforman la presente conferencia procuran dar una visión general del tratamiento estadístico inherente a la evaluación y al uso de las pruebas diagnósticas.

1.2. Algunos conceptos y definiciones básicas

Es claro que algunas entidades nosológicas pueden ser padecidas de acuerdo a una gradación; por ejemplo, un hipertenso puede padecer de esa enfermedad con una u otra intensidad, dependiendo de cuán altas sean sus tensiones diastólica y sistólica y de las diversas combinaciones en que éstas pueden presentarse. Sin embargo, en la inmensa mayoría de los casos el concepto de interés primario es delimitar si el sujeto padece de una enfermedad especificada o está libre de ella. Para muchas entidades, existen delimitaciones inequívocas que permiten ubicar al sujeto en una de esas dos categorías.

Por ejemplo, el significado de que un sujeto tenga fractura del cráneo —por su propia naturaleza— no se presta a interpretaciones equívocas, independientemente de que un medio concreto para establecerlo pueda dar lugar a errores; en un ejemplo como el de la tensión arterial, también se puede postular una frontera formal que separe al hipertenso del que no lo es (por ejemplo, declarar que se está en el primer caso si la tensión diastólica excede los 90 mm Hg), aunque la vía para medir esa tensión pueda ser tal que en algunos casos declare hipertensos a quienes no lo son y viceversa.

En síntesis, debe distinguirse la condición por conducto de la cual se define la posesión del estado patológico que se analiza del procedimiento usado para evaluar si ella se cumple.

En lo sucesivo partiremos del supuesto de que tal condición definitoria está perfectamente definida. Un sujeto específico, por tanto, cumple con esa condición (circunstancia que se denotará por E) o no cumple con ella (lo cual se denotará por \bar{E}).

Genéricamente, supongamos que hay un test o prueba que, aplicado sobre un sujeto concreto,

puede dar lugar a sólo uno de dos resultados posibles: T+ y T-, símbolos con que se denotan los resultados positivos y negativos a la prueba.

En principio, el resultado T+ constituye un *indicio* de que el sujeto tiene la condición E (es decir, de que está enfermo) y el resultado T- induciría a pensar que el sujeto en cuestión tiene la condición complementaria \bar{E} (o sea, que no tiene la enfermedad).

Definiremos como *prevalencia* de la enfermedad en cuestión a la probabilidad P(E) de que un sujeto cualquiera tenga la enfermedad en estudio; P(E) es la expresión probabilística de lo que, en sentido epidemiológico, Hill (1967) llama «Tasa de prevalencia» y define como el resultado de dividir el número de enfermos existentes en un instante dado entre el número de individuos en riesgo de padecer tal enfermedad en ese momento.

Con frecuencia, este importante parámetro se estima a través de una prueba como la descrita previamente. La proporción de sujetos de una muestra para los que se obtiene un resultado positivo (T+) suele tomarse directamente como estimador de P(E) en los programas de pesquiasaje (screening programs) y exámenes masivos. (Ver Zelen y Feinleb, 1969).

Tal estimador está afectado, naturalmente, por el error muestral que dimana del hecho de que no toda la población es objeto de examen.

Sin embargo, de tanta o mayor importancia son las consecuencias de que raramente se cuenta con pruebas diagnósticas perfectas. En efecto, no todos los enfermos son detectados como tales por el test ni todos los sanos resultan clasificados como negativos.

El grado de eficiencia inherente a una prueba diagnóstica se resume en dos parámetros conocidos como *sensibilidad* y *especificidad*.

El primero mide la capacidad del test para detectar a un sujeto enfermo (expresa cuán «sensible» es la presencia de la enfermedad) y viene definido por la probabilidad condicional.

$$\alpha = P(T+/E)$$

O sea, la sensibilidad es la probabilidad de que la prueba identifique como enfermo a aquel que efectivamente lo es.

El otro parámetro mide la capacidad que tiene la prueba de identificar como sanos a los que efectivamente lo son.

tivamente lo son. Se define como la probabilidad condicional:

$$\beta = P(T-\bar{E})$$

Es decir, mide cuán «específico» es en el sentido siguiente: cuanto mayor sea β , menor es su complemento $P(T+/\bar{E})$; o sea, menor es la probabilidad de que declare como enfermos a sujetos que no lo son.

Los parámetros y sugeridos por Yerushalmy (1947) hace ya cuatro décadas, sintetizan la eficiencia intrínseca de una prueba y constituyen una vía, por ejemplo, para comparar dos o más pruebas que compiten entre sí.

Una primera medida conjunta de esa eficiencia fue propuesta varias décadas atrás. Se trata del Índice de Youden (1950):

$$I_1 = \alpha + \beta - 1$$

El caso $\alpha + \beta = 2$ ($I_1 = 1$) se alcanza sólo cuando la prueba diagnóstica es óptima.

Ahora bien, para evaluar el interés práctico de una prueba hay que centrar la atención en la medida en que sus resultados modifican realmente el grado de conocimiento que se tenía sobre el estado del paciente (ver Sox, 1986).

Más concretamente, interesa evaluar la probabilidad de que un sujeto para el que se haya obtenido T+ sea efectivamente un enfermo, y compararla con la probabilidad incondicional.

En ese contexto surgen otros dos indicadores de interés: el valor predictivo de un resultado positivo y el de un resultado negativo. En términos probabilísticos, se trata de los parámetros:

$$P(E/T+) \text{ y } P(\bar{E}/T-)$$

El primero, por ejemplo, cuantifica la posibilidad de que un sujeto esté enfermo dado que el test lo declaró positivo. Su complemento:

$$1 - P(E/T+) = P(\bar{E}/T+)$$

es la expresión probabilística del concepto conocido por los epidemiólogos como «tasa de falsos positivos».

Obviamente, entonces, una prueba diagnóstica debe cumplir las siguientes condiciones mínimas:

$$P(E/T+) \geq P(E) \quad (1)$$

$$P(\bar{E}/T-) \geq P(\bar{E}) \quad (2)$$

Si llamamos $p = P(E)$ y $q = 1 - p$, aplicando el teorema de Bayes se obtienen de inmediato:

$$P(E/T+) = \frac{p\alpha}{p\alpha + q(1 - \beta)} \quad (3)$$

$$P(\bar{E}/T-) = \frac{q\beta}{q\beta + p(1 - \alpha)} \quad (4)$$

Al sustituir (3) y (4) en (1) y (2) respectivamente se obtiene que estas últimas son ecuaciones equivalentes entre sí y equivalentes a la condición:

$$\alpha + \beta \geq 1 \quad (5)$$

Cabe subrayar que (5) es, ciertamente, una condición elemental puesto que su incumplimiento exige que al menos uno de los sumandos sea inferior a 0.5 lo cual viene a significar que, al menos en uno de los sentidos, el test no es mejor que tomar la decisión mediante el lanzamiento de una moneda.

Consecuentemente, algunos autores como Quade y col (1980) parten incluso del supuesto más fuerte de que tanto α como β superan a 0.5.

Sin embargo, no faltan ejemplos reales donde incluso la condición mínima (5) se incumple (ver, por ejemplo, las ilustraciones que se dan en Rogan y Gladen, 1978). Por otra parte, puede ocurrir que la condición sea cierta pero que el aporte hecho por el test al conocimiento que ya se tenía sea virtualmente despreciable. Por ejemplo Weiner y col (1979) concluyen que cierto test predictor de la enfermedad coronaria es de poco interés pues, si bien el valor $P(E/T+)$ asciende a 0.96 cuando se aplica a pacientes anginosos, la prevalencia de la enfermedad es de 0.89 entre dichos pacientes.

Es fácil demostrar que la condición

$$\alpha + \beta = 1 \quad (6)$$

es equivalente a

$$P(E/T+) + P(\bar{E}/T-) = 1$$

La diferencia esencial que existe entre la pareja α, β y los valores predictivos es la siguiente.

Los primeros caracterizan al test propiamente dicho, independientemente de la prevalencia vigente en la población a la que se aplique. Naturalmente, como señala Hilden (1980), estos pa-

rámetros alcanzan un sentido claro sólo en el contexto de un problema clínico bien delimitado.

El valor predictivo de un test (positivo o negativo) sin embargo, depende además de la magnitud de p , la prevalencia de la enfermedad en la población.

Nótese, por otra parte, que bajo el supuesto de que se cumple (5), el índice de Youden satisface la relación:

$$0 \leq I_1 \leq 1$$

con lo cual se revela cabalmente su expresividad. Sin embargo, éste ha sido criticado por Hilden (1981) sobre la base de que la suma de probabilidades condicionales, donde el evento condición no es el mismo, carece de interpretación.

Por otra parte, curiosamente y en fecha mucho más reciente, Connell y Koespell (1985) luego de consideraciones teóricas y desarrollos algebraicos relativamente complicados arriban a un indicador que mide «la ganancia porcentual en certeza a que da lugar un resultado positivo y un resultado negativo» y luego crean un indicador combinado de ambos. De ese elaborado proceso resulta un indicador que no es otra cosa que el susodicho índice de Youden. Aunque ello aporta otra interpretación cualitativa para I_1 , sorprende la capacidad de algunos resultados para ser «redescubiertos» (los autores no mencionan que este indicador fue propuesto tres décadas antes).

A diferencia de este índice, que no depende de la frecuencia en que se presenta la dolencia bajo estudio, se ha concebido otro indicador que caracteriza la eficiencia de una prueba diagnóstica en una situación práctica y concreta. Se trata del llamado *Índice de validez* (ver Feinstein, 1977).

$$I_2 = \alpha p + \beta q \quad (6)$$

Es fácil comprender que I_2 no es otra cosa que la probabilidad de que el diagnóstico sea correcto. Como puede apreciarse, I_2 depende linealmente de la prevalencia:

$$I_2 = A + Bp$$

$$\text{donde } A = \beta \text{ y } B = (\alpha - \beta).$$

Schechter y Sheps (1985) han intentado establecer algunos principios para la toma de decisiones en función de los resultados de las pruebas.

Allí se insiste en que la verdadera utilidad clínica de una prueba depende de la probabilidad

intrínseca de cada sujeto y del grado de evidencia que ella agrega (o elimina) de que éste tenga en efecto la dolencia. Por ejemplo, un individuo específico, por sus características personales, puede tener una probabilidad dada de padecerla (dependiente de su edad, sexo, síntomas, etc.); en función de ésta, se calcula la probabilidad a posteriori que dimane del resultado del test. La magnitud alcanzada entonces puede o no determinar un cambio en la conducta a seguir (mantenerlo en observación, hacer otras pruebas, indicar un medicamento, etc.). Puesto que los valores $P(E/T+)$ y $P(E/T-)$ pueden calcularse antes de aplicarse el test, a partir de ellos podría el clínico abstenerse incluso de ordenar una prueba que no debería modificar su conducta posterior. De tal suerte, se evitaría una práctica ya impugnada por Wong y Lincoln (1983) según la cual algunos clínicos «disparan antes de apuntar».

Las consideraciones de Schechter y Sheps sugieren la idea siguiente. Supongamos que se ha realizado un estudio en que a cada sujeto de una muestra además de clasificarlo según posea o no el rasgo de interés (E o \bar{E}) se le ha preguntado un conjunto de datos X_1, X_2, \dots, X_k supuestamente relevantes en tanto potenciales modificadores de la probabilidad de estar en una u otra clase.

En tales circunstancias puede estimarse una función del tipo siguiente:

$$P(E/X_1, X_2, \dots, X_k) = f(X_1, X_2, \dots, X_k)$$

O sea, una función que exprese la probabilidad de tener el rasgo en función del perfil del individuo dado por los valores de X_i .

Por ejemplo pudiera tratarse de lo siguiente:

E = el sujeto tiene enfermedad de la arteria coronaria

\bar{E} = el sujeto no tiene la enfermedad

$$X_1 = \begin{cases} 1 & \text{si es varón} \\ 0 & \text{si es mujer} \end{cases}$$

X_2 = edad del sujeto en años

$$X_3 = \begin{cases} 1 & \text{el suj. pres. hist. de dol. angin.} \\ 0 & \text{en otro caso} \end{cases}$$

$$X_4 = \begin{cases} 1 & \text{el individuo es fumador} \\ 0 & \text{el individuo no fuma} \end{cases}$$

Con los datos X_1, X_2, X_3 y X_4 para cada paciente de cada uno de los dos grupos se pueden, por ejemplo, estimar los parámetros A y B .

($i = 1, 2, 3, 4$) que definen la función logística, de uso creciente en la investigación epidemiológica:

$$P(E/X_1, X_2, X_3, X_4) = (1 + \exp(-A - \sum_{i=1}^k B_i X_i))^{-1}$$

El proceso de ajuste y evaluación de la pertinencia de tal modelo puede realizarse de manera relativamente fácil con el software disponible actualmente.

Ahora para evaluar la conveniencia de la realización, pongamos por caso, de una angiografía radionuclídea (prueba diagnóstica T) sería necesario calcular $P(E/T+)$ y $P(E/T-)$. Para ello puede utilizarse el programa SCH.BAS que se anexa a este material (tomado del propio trabajo de Schechter y Sheps).

El resultado de evaluar la función puede constituir el dato de entrada («pretest probability») que junto con la especificidad y sensibilidad de la angiografía (0.87 y 0.54 respectivamente según Austin y col., 1982) constituyen los datos de entrada al programa.

1.3. Problemas de estimación

Supongamos que α y β son conocidos para cierto test T y cierta dolencia E, y que se desea estimar la prevalencia $P(E)$ de esta última en una población dada.

Puesto que T no es perfecto ya que declara enfermos a una parte de los sanos y viceversa, es evidente que \hat{p} es un estimador sesgado de p .

En efecto, el valor esperado de p es:

$$P(T+) = P(T+/E) P(E) + P(T+/\bar{E}) P(\bar{E}) = \alpha P + (1 - \beta)(1 - P)$$

Consecuentemente, despejando P de esta relación se obtiene una corrección que elimina el sesgo:

$$\hat{p}_1 = \frac{\hat{p} + \beta - 1}{\alpha + \beta - 1}$$

Los efectos de tal corrección pueden ser considerables, aún en caso de pruebas con altas sensibilidad y especificidad.

La Tabla 1 refleja la modificación que ella supone para una prueba donde $\alpha = \beta = 0.9$ en términos del sesgo relativo $|p - p_1| / \hat{p}$.

Naturalmente, si α y β no se conocen, deben ser a su vez estimadas y luego calcularse:

$$\hat{p}_2 = \frac{\hat{p} + \hat{\beta} - 1}{\hat{\alpha} + \hat{\beta} - 1}$$

Rogan y Gladen (1978) hacen un estudio comparativo de \hat{p} y \hat{p}_2 en términos de sus respec-

Tabla 1. Valores de p y sesgo relativo (en %) para una prueba con $\alpha = \beta = 0.9$ según valores de \hat{p} .

\hat{p}	0.100	0.200	0.300	0.400	0.500	0.600	0.700	0.800	0.900
\hat{p}_1	0.000	0.125	0.250	0.375	0.500	0.625	0.750	0.875	1.000
SR	100	38	17	6	0	4	7	9	11

El método regular sería seleccionar una muestra simple aleatoria de tamaño n y aplicarle a todos los sujetos la prueba diagnóstica en cuestión.

Una primera estimación de $P(E)$ sería simplemente la siguiente:

$$\hat{p} = t/n$$

donde t es el número de sujetos para los cuales la prueba arrojó resultado positivo (T+).

tivos errores cuadráticos medios en el que este último resulta claramente favorecido.

El problema de la estimación de α y β ha merecido múltiples consideraciones. La vía canónica consiste en seleccionar una muestra de n_1 enfermos y otra de n_2 sujetos sanos y aplicar el test a los $n = n_1 + n_2$ individuos de suerte que pueda conformarse una clasificación cruzada tal y como la que se expone en la Tabla 2.

Tabla 2. Tabla 2×2 para los resultados de una prueba diagnóstica.

Resultados de la prueba	Enfermedad		Total
	Presente (E)	Ausente (E)	
Positivo (T+)	a	b	a+b
Negativo (T-)	c	d	c+d
Total	n_1	n_2	n

Los estimadores obvios de α y β vienen dados por las razones siguientes:

$$\hat{\alpha} = a / n_1 \quad \text{y} \quad \hat{\beta} = d / n_2 \quad (7)$$

Varias observaciones cabe hacer aquí. En primer lugar, hay que enfatizar que el criterio de clasificación de los sujetos como enfermos o no debe ser independiente de la prueba diagnóstica en estudio; de ese criterio—tomado como «criterio de verdad»— se supone que tiene sensibilidad y especificidad máximas. Por otra parte, debe notarse que el diseño planteado no consiente la estimación de $P(E)$ a través de la razón n_1/n puesto que n_1 y n_2 son números elegidos por el evaluador (con frecuencia iguales entre sí). Consecuentemente, tampoco es posible en este caso estimar los valores predictivos definidos en (3) y (4).

Para lograr esto último hay que contar con una estimación *independiente* del parámetro $P(E)$.

Otra variante es obtener una muestra simple aleatoria de n individuos y aplicar a cada uno de ellos tanto «el criterio de verdad» con que se cuenta como la prueba diagnóstica que se evalúa. En tal caso, todos los marginales ($a+b$, $c+d$, n , y n) son aleatorios y además de (7) pueden usarse los estimadores:

$$P(\hat{E} / T+) = a / (a + b)$$

$$P(\hat{E} / T-) = d / (c + d)$$

$$\widehat{P(E)} = n_1 / n$$

Tal estrategia, sin embargo, es raramente factible. Efectivamente, téngase en cuenta que lo que ocurre usualmente es que el criterio estándar de comparación es un método sofisticado o econó-

micamente prohibitivo en tanto que el test que se evalúa se propone precisamente como un suceso económicamente ventajoso de ser aplicado bajo condiciones de apremio de tiempo o de recursos.

Un ejemplo elocuente pudiera ser el caso en que el criterio de verdad viene dado por un examen histopatológico de una pieza anatómica (ocasionalmente sólo evaluable tras la muerte del paciente) y la prueba diagnóstica pudiese ser resultado de aplicar una regla en función de síntomas y signos clínicos.

Con frecuencia la prueba tiene un sentido predictivo y por ende es imprescindible optar por esta segunda vía. Por ejemplo, un sujeto presenta síntomas de la llamada «depresión del anciano»; la determinación inequívoca de si se trata sólo de un estado depresivo (reversible) o del debut de una demencia senil (dolencia que supone conductas terapéuticas diferentes), no es posible hasta que haya transcurrido un período de evolución relativamente prolongado.

Si se está evaluando un criterio diagnóstico precoz de la demencia en base al perfil sintomático del anciano, no se puede hacer otra cosa que aplicar dicha prueba a una muestra de n sujetos que cumplan un conjunto básico de síntomas (común a ambos desórdenes) y esperar el lapso debido para conocer los restantes componentes de la Tabla 2.

La falta de sensibilidad y especificidad de una prueba puede tener su origen en dos fuentes básicas: inconsistencia interna por una parte y falta de sensibilidad o especificidad intrínsecas.

En el primer caso se encuentran aquellas pruebas en que influyen la apreciación personal del evaluador o del paciente, aspectos volitivos relacionados con uno u otro, variabilidad inherente al proceso de medición, carácter lábil de las manifestaciones, etc. Retomando un ejemplo anterior, conceptualmente puede entenderse que un sujeto es hipertenso cuando su tensión arterial mínima supere los 90 mm Hg. Ahora bien, es menester precisar qué se entiende por ello: se está postulando la existencia de una tensión diastólica inherente al sujeto, pero es bien conocido que un registro concreto puede depender de factores tales como el equipo con que se realiza la prueba, el proceso de lectura y, sobre todo, de las condiciones sico-físicas del paciente en el momento de la prueba. Tales coyunturas pueden explicar entonces la falta de eficiencia de la prueba.

Por otra parte, puede ocurrir que la prueba sea, por su naturaleza, internamente consistente (siempre arroja el mismo resultado para cada individuo) y poseer una imperfección que le es inherente.

Un ejemplo absurdo en la práctica pero ilustrativo de este tipo de test es el caso en que la regla consiste en declarar enfermo a todos los sujetos (en tal caso $\alpha = 1$ y $\beta = 0$).

Para el caso en que la fuente de imperfección sea esencialmente del primer tipo, la estimación simultánea de α , β y $P(E)$ puede realizarse sin necesidad de un criterio de verdad.

Dempster y col (1977) y luego Dawid y Skene (1979) desarrollaron un algoritmo basado en el llamado «principio de pérdida de información».

El procedimiento consiste en aplicar en k ocasiones independientes la prueba T a cada sujeto de una muestra simple aleatoria de tamaño n . Supongamos que $k = 3$.

Se define como n_i a la cantidad de individuos para la que el número de resultados positivos fue i ($i=0, 1, 2, 3$).

Basándose en esos $K+1$ valores (4 en este caso) se procede a un proceso iterativo que discurre del modo siguiente:

1. Se propone o sugiere a un número r_i de sujetos que, siendo enfermos verdaderos, obtuvieron un resultado positivo (T+) en i de las 3 pruebas. Se define entonces $s_i = n_i - r_i$.

2. Se calculan $s = \sum_{i=0}^3 s_i$ y $r = \sum_{i=0}^3 r_i$

3. Se computan $\alpha_1 = (\sum_{i=1}^3 i r_i) / 3r$, $\beta_1 = [\sum_{i=0}^2 (3 - i) s_i] / 3s$

y $p_1 = r / n$.

4. Se calculan r'_i y s'_i :

$$r'_0 = nr_1(1-\alpha)^3 \quad s'_0 = n(1-p_1)\beta_1^3$$

$$r'_1 = 3nr_1\alpha_1(1-\alpha)^2 \quad s'_1 = 3n(1-p_1)\beta_1^2(1-\beta_1)$$

$$r'_2 = 3nr_1\alpha_1^2(1-\alpha) \quad s'_2 = 3n(1-p_1)\beta_1(1-\beta_1)^2$$

$$r'_3 = nr_1\alpha_1^3 \quad s'_3 = n(1-p_1)(1-\beta_1)^3$$

5. Se corrigen r'_i y s'_i de suerte que sumen n_i :

$$r_i = \frac{n_i r'_i}{r'_i + s'_i} \quad s_i = n_i - r_i$$

Se vuelve al paso 2., se calculan α_2, β_2 y p_2 según las fórmulas del tercer paso, y así sucesivamente hasta que las sucesiones $(\alpha_i), (\beta_i), (p_i)$ converjan.

Se anexa el programa EM.BAS en lenguaje BASIC que realiza este proceso iterativo.

A modo de ejemplo supongamos que habiéndose realizado 3 veces cierta prueba a cada uno de 700 individuos, se obtienen los resultados que se aprecian en la segunda columna de la Tabla 3.

Tabla 3. Valores de n_i ($i=0, 1, 2, 3$) obtenidos para un ejemplo hipotético y valores supuestos de r_i (número de enfermos entre los n_i).

i	Número n_i de sujetos con i resultados positivos	Primera aproximación de los r_i
0	189	12
1	180	42
2	136	85
3	202	190
Total	700	329

En la tercera columna se han puesto las cantidades «adivinadas» de verdaderos positivos (E) entre los tienen i resultados positivos de la prueba.

Con estos datos se obtienen las primeras estimaciones:

$$\alpha_1 = .79 \quad \beta_1 = .75 \quad p_1 = .47$$

Aplicando el algoritmo iterativo, se obtienen las sucesiones que se recogen en la Tabla 4.

Tabla 4. Sucesivos valores de α_i, β_i, p_i al aplicar el algoritmo.

Iteración (j)	α_i	β_i	p_i
1	.79	.75	.47
2	.82	.78	.47
3	.84	.79	.46
4	.85	.79	.46

Los límites de las sucesiones constituyen estimaciones máximo verosímiles de la sensibilidad, especificidad y prevalencia respectivamente.

Volviendo al método convencional explicado arriba, basado en la contrastación de los resultados

de T con un «criterio de verdad» (llamado a veces en la literatura sajona como «Golden Standard») cabe advertir que muchas veces el supuesto $\alpha = \beta = 1$ no se cumple.

Cabe detenerse sobre los efectos que sobre las estimaciones de la sensibilidad y la especificidad de una prueba pudiese tener el hecho de que el criterio de comparación sea deficiente. Como plantean Robertson y col (1983), los criterios de rutina suelen ser totalmente inadecuados para la evaluación de un nuevo test ya que sus propias imperfecciones pueden distorsionar sensiblemente la verdadera capacidad de la mejor prueba diagnóstica que se analice.

Supongamos que se tienen dos tests. Denotemos por A y \bar{A} a los resultados T+ y T- respectivamente para el primero de ellos y mediante B y \bar{B} a tales resultados para el segundo.

Llamemos $\alpha' = P(A/E)$ y $\beta' = P(\bar{A}/\bar{E})$.

Supongamos que la primera de esas pruebas se toma como referencia para evaluar la segunda; se eligen n_1 individuos para los que se obtuvo A y n_2 para los que se obtuvo \bar{A} .

Al realizar el cruzamiento según los resultados (B o \bar{B}) del test bajo estudio se obtiene una tabla similar a la que, a modo de ejemplo, se resume en la Tabla 5.

Tabla 5. Resultado de la aplicación de una prueba diagnóstica bajo estudio a 200 sujetos positivos y 200 negativos según una prueba de referencia.

	A	\bar{A}	
B	a = 120	b = 70	190
\bar{B}	c = 80	d = 130	210
	n = 200	n = 200	n = 400

Según los resultados de la Tabla 5 se harían las estimaciones regulares para la sensibilidad y especificidad del test en cuestión:

$$\hat{\alpha} = \frac{a}{n} = 0.60 \quad \hat{\beta} = \frac{d}{n} = 0.65$$

Sin embargo, puede probarse que los parámetros verdaderos se obtendrían según las fórmulas siguientes:

$$\alpha = \frac{(a+b)\beta' - b}{n_1 - (1-\beta')n} \quad (8)$$

$$\beta = \frac{(c+b)\alpha' - c}{n\alpha' - n_1} \quad (9)$$

Aplicadas a nuestro ejemplo para, digamos $\alpha' = 0.85$ y $\beta' = 0.80$, se obtendría $\alpha = 0.68$ y $\beta = 0.70$, lo cual constituye una sensible corrección a las estimaciones $\hat{\alpha}$ y $\hat{\beta}$.

Si por ejemplo se hubiese tenido $\alpha' = \beta' = 0.95$, entonces los parámetros reales serían $\alpha = 0.61$ y $\beta = 0.66$ (virtualmente iguales a las estimaciones). Quiere esto decir que el problema dista de ser grave si la imperfección del test de referencia no es acusada.

Las fórmulas (8) y (9), por tanto, permiten que cualquier prueba cuyos valores de α y β sean conocidos puedan servir como criterio de verdad. (Sin que sea necesario el conocimiento de P(E) que plantea Staquet y col, 1981).

El caso en que tampoco los parámetros de la prueba usada como referencia son conocidos ofrece sin duda complicaciones adicionales. El tema desborda el alcance del presente material por lo que sólo se mencionarán los trabajos relevantes que pueden ser consultados.

Greenberg y Jekel (1969) han sugerido que se trabaje con dos poblaciones, una con baja prevalencia de la enfermedad y otra con prevalencia elevada. La primera se usaría para estimar α y la otra para estimar β .

Sucesivas contribuciones alrededor de ese punto son debidas a Hui y Walter (1980), Thibodeau (1981) y Vacek (1985).

1.4. Comparación de diseños alternativos

A fin de mejorar la calidad de una prueba se ha sugerido, Nissen-Mayer (1964), la idea de aplicar la misma prueba r veces a cada sujeto y declarar positivo a aquel para el que se hayan obtenido resultados positivos en k o más ocasiones.

Sandifer y col (1968), por ejemplo, sugieren tomar $r = 3$ y $k = 2$, de modo que con frecuencia resulta innecesario realizar la tercera prueba (es decir, siempre que los dos primeros resultados coincidan, el tercero es superfluo a los efectos de clasificar al individuo).

Otra variante puede ser la siguiente. Realizar una vez la prueba; si es negativa, declarar negativo al sujeto; si es positiva confirmarlo con una segunda prueba; si esta segunda es negativa, decidir entonces a través de una tercera tomando como válido el resultado predominante.

Naturalmente, la evaluación de la eficiencia relativa de estos (y otros) diseños debe incluir alguna consideración sobre los «los costos» de tener un falso positivo y un falso negativo. Como es obvio, no se trata de un costo en unidades monetarias sino de una valoración (que necesariamente habrá de transitar por cierta subjetividad del investigador) sobre los efectos en términos de ansiedad, gastos, etc., que produciría declarar probablemente enfermo a quien no lo es (falso positivo) y en términos del riesgo para la salud que implica sugerir que es sano un individuo enfermo (falso negativo). Por ejemplo, Dobie (1985) fundamenta que el «costo» de declarar libre de neuroma acústico a un sujeto que posee esa dolencia es 5 veces mayor que el de atribuírsela indebidamente.

Para fijar ideas en cuanto a cómo se podría realizar el análisis, consideremos una situación simple en que se comparan dos pruebas T_A y T_B :

- A: realizar cierto test T que tiene sensibilidad α y especificidad β .
- B: realizar 2 veces el test: declarar positivo sólo al sujeto para el que ambas veces se haya obtenido T_{A+} .

Comencemos calculando la sensibilidad y especificidad de T_B .

$$\alpha^* = P(T_{B+} / E) = P(X = 2 / E) = \alpha^2$$

Donde X = número de veces que se obtiene T_{A+} cuando se aplica en dos ocasiones.

$$\begin{aligned} \beta^* &= P(T_{B-} / \bar{E}) = P(X < 2 / \bar{E}) \\ &= \beta^2 + 2\beta(1 - \beta) = \beta(2 - \beta) \end{aligned}$$

el procedimiento B, como es de esperar, aumenta la especificidad (es el resultado de multiplicar β por $(2 - \beta)$, número mayor que 1) pero disminuye la sensibilidad.

Para comparar el comportamiento de un método respecto de otro cuando se aplican en una población de n sujetos cuya prevalencia de la enfermedad es $P(E)$ es menester definir un criterio racional. El procedimiento usual es el de es-

tablecer una función de costo: se suma el costo concerniente a los falsos negativos (FN) al que producen los falsos positivos (FP). Sean esos costos CFN y CFP respectivamente.

El número esperado de sujetos enfermos es $n P(E)$. Una parte de ellos (FN) serán declarados como negativos por la prueba:

$$FN = P(T- / E) P(E) n$$

Análogamente:

$$FP = P(T+ / \bar{E}) P(\bar{E}) n$$

Se define entonces:

$$C' = (CFN) (FN) + (CFP) (FP)$$

Llamando $C = CFN / CFP$ (que expresa cuanto más costoso es un falso negativo que un falso positivo), y $p = P(E)$, se tiene:

$$C' = n (CFP) \{ [1 - P(T+ / E)] P(E) C + [1 - P(T- / \bar{E})] [1 - P(E)] \}$$

Ahora, calcular el mínimo de C' es equivalente a calcularlo para:

$$C = (1 - \alpha) p C_R + (1 - \beta) (1 - p) \quad (10)$$

Consideremos $\alpha = .80$, $\beta = .70$ y $C_R = 5$. Los resultados de C para ambas pruebas y para diferentes prevalencias se resumen en la Tabla 6.

Tabla 6. Valores del costo total resultante de aplicar T_A y T_B para diferentes valores de la prevalencia (supuestos $\alpha = .80$, $\beta = .70$ y $C_R = 5$).

P(E)	C_A	C_B
0.10	0.37	0.26
0.20	0.44	0.43
0.30	0.51	0.60
0.40	0.58	0.77
0.50	0.65	0.95
0.60	0.72	1.11
0.70	0.79	1.29
0.80	0.86	1.46
0.90	0.93	1.62

Como se ve, en este caso la prueba T_B sería más recomendable para poblaciones en que la prevalencia sea inferior al 20 %.

Otra vía para comparar ambos sistemas pudiera ser la de confrontar sus respectivos índices de validez I_2 , definido en (6).

Expresando la relación entre I_2 y p a través de la recta que los vincula, se tendría para A y B respectivamente.

$$I_{2A} = 0.10 p + 0.70$$

$$I_{2B} = -0.27 p + 0.91$$

Tal y como sugiere Taube (1986), ellas pueden representarse en un sistema cartesiano como el del Gráfico 1.

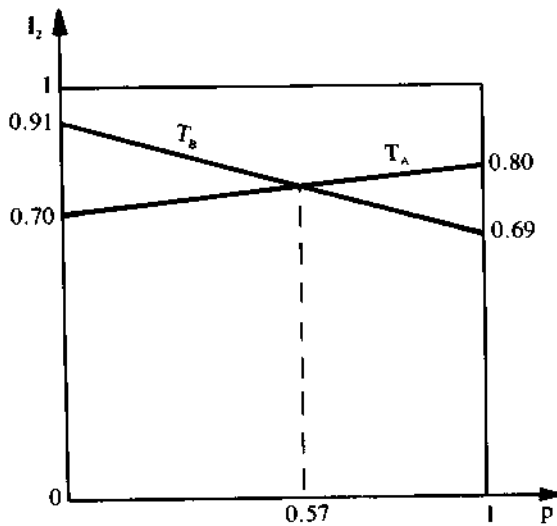


Gráfico 1. Relación entre el índice de validez y la prevalencia para los tests A y B ($\alpha = 0.8$ y $\beta = 0.7$).

De aquí se desprende que para $P(E) > 0.57$ el test A es mejor (tiene superior índice de validez) en tanto que para prevalencias menores, el B sería preferible.

Aparentemente hay falta de consistencia entre los resultados que arrojan uno y otros criterios de comparación. Por ejemplo, según el primer análisis había que elegir T_A si, digamos, $P(E) = 0.4$, en tanto que para esa prevalencia el segundo procedimiento aconseja hacer uso de T_B .

La contradicción es, en efecto sólo aparente. En rigor el segundo procedimiento asume tácitamente que $CFN = CFP$. Más concretamente, el costo C definido en (10), cuando $C_R = 1$, no es otra cosa que $1 - I_2$. De modo que la diferencia sólo puede atribuirse a que se ha partido de supuestos totalmente diferentes en cuanto a la magnitud de C_R .

Hasta el momento todo el análisis se ha circunscrito a pruebas diagnósticas que arrojan resultados dicotómicos ($T+$ o $T-$).

Sin embargo, los resultados de una gran cantidad de pruebas usuales de valor diagnóstico se mueven en un «continuum» y corresponden a la decisión del clínico al establecer por debajo o por encima de qué valor va a considerar que el resultado indica anormalidad.

Supongamos que se considera la glucosa en sangre de un individuo como factor discriminante entre diabéticos y no diabéticos. Tomando un ejemplo de Lilienfeld y Lilienfeld (1983), consideremos que se ha realizado una prueba postprandial de dos horas para dosificación de la glicemia a 70 diabéticos confirmados y 510 sujetos no diabéticos. Al considerar positivo al sujeto cuyo nivel de glucosa en sangre excede los 110 mg/100 cc y negativo al que exhiba un valor inferior, se obtienen los resultados de la Tabla 7.I. Si el nivel elegido es 140 mg/100 cc, entonces los resultados son los de la Tabla 7.II.

Tabla 7. Clasificación en positivos y negativos de un grupo de 70 diabéticos y 510 no diabéticos según una prueba post-pandrial de dos horas según se tome el nivel 110 mg/100 cc o 140 mg/100 cc.

I	Diabéticos	No diabéticos
más de 110 mg/100 cc	65	253
110 mg/100 cc o menos	5	247
Totales	70	510

II	Diabéticos	No diabéticos
más de 140 mg/100 cc	52	45
menos de 140 mg/100 cc	18	465
Totales	70	510

Las estimaciones de sensibilidad y especificidad para ambas pruebas son, respectivamente:

$$\alpha_I = .93 \quad \beta_I = .48$$

$$\alpha_{II} = .74 \quad \beta_{II} = .91$$

Es decir, se observa lo que típicamente ocurre en tales situaciones: una modificación que aumente la especificidad, disminuye la sensibilidad y viceversa. Por otra parte, es obvio que para cada nivel escogido para discriminar se tendría una tabla 2×2 como la de la Tabla 7.

La Tabla 8 reproduce los datos de α y β que dan Lilienfeld y Lilienfeld (1983) para 13 niveles (puntos de corte) escogidos.

Tabla 8. Sensibilidad y especificidad para varios niveles de glucosa en sangre aplicados a 70 diabéticos y 510 no diabéticos.

Nivel de glucosa (mg/100 cc)	α	β
80	1.00	0.01
90	0.99	0.07
100	0.97	0.25
110	0.93	0.48
120	0.89	0.68
130	0.81	0.82
140	0.74	0.91
150	0.64	0.96
160	0.56	0.99
170	0.53	1.00
180	0.50	1.00
190	0.44	1.00
200	0.37	1.00

De hecho, se está ante una «familia de pruebas» y resulta natural plantearse la tarea de identificar la mejor de ellas; es decir, establecer el nivel óptimo.

Un instrumento que ha despertado creciente interés (ver Metz, 1978) ha sido el de la llamada curva característica de operación, conocida por la sigla ROC («Receiver Operating Characteristic Curve»).

La capacidad discriminativa de una prueba a lo largo de todos los posibles «puntos de corte» puede ser representada a través de una curva. Situada en un sistema de ejes cartesianos, en el eje de ordenadas se ubican los distintos valores de la sensibilidad y en el eje de abscisas las correspondientes tasas de falsos positivos ($1 - \beta$).

El Gráfico 2 muestra la forma típica de una curva ROC. Como se ve, el valor de α crece en la medida que crece $1 - \beta$ (β disminuye).

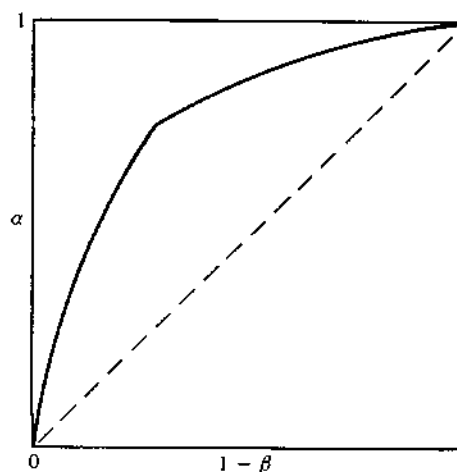


Gráfico 2. Forma genérica de una curva ROC

Un caso extremo se produce cuando la curva coincide con la diagonal; este es el caso en que siempre se tiene $\alpha + \beta = 1$, cuando la prueba nunca ofrece la mínima información necesaria como para ser útil. El otro caso es aquel en que la curva coincide con la paralela al eje de abscisas que pasa por el punto $\alpha = 1$ (es decir, cuando existe un punto de corte para el cual $\alpha = 1$ y $1 - \beta = 0$).

En el Gráfico 3 se refleja la curva ROC empírica correspondiente a los datos de la Tabla 8.

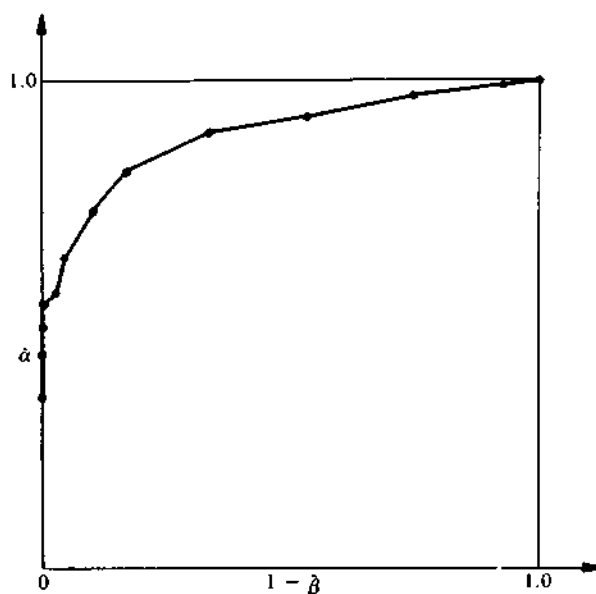


Gráfico 3. Curva ROC empírica correspondiente a los datos de la Tabla 8.

Para evaluar el mejor nivel de decisión («cut-off») se puede acudir nuevamente a la función definida en (10). Si se va a usar la prueba postpandreal en un pesquisaje masivo («Screening») de la población, hay que establecer el nivel $P(E)$ de la prevalencia vigente en esa población (aunque sea aproximadamente) y fijar un costo relativo C de declarar libre de diabetes a un enfermo respecto de declararlo diabético sin que lo sea. Si por ejemplo se fija $C_R = 4$ entonces usando los valores de la Tabla 8 se puede calcular C para cada uno de los 13 niveles y para la prevalencia que tenga la población en cuestión.

La Tabla 9 recoge los resultados de tal situación para dos poblaciones, una con $p(E) = 0.1$ y otra con $p(E) = 0.3$.

Como se ve en la Tabla 9 para el caso $P(E) = 0.1$, el mínimo valor de C se alcanza cuando el punto de corte escogido es 150 mg/100 cc. Si la prevalencia fuese $p(E) = 0.3$, entonces lo óptimo sería tomar 130 mg/100 cc como punto de separación.

Una aplicación directa de esta idea se relaciona con la regresión logística mencionada en la Sección III. Supongamos que se ha ajustado una curva de regresión logística múltiple: en la situación siguiente.

Tabla 9. Valores de C para los puntos de corte considerados en la Tabla 8, considerando $C_R = 4$ y dos valores de $P(E)$.

Nivel de decisión mg/100 cc	Prevalencia 10 %	Prevalencia 30 %
80	.89	.69
90	.84	.66
100	.68	.56
110	.50	.45
120	.33	.36
130	.24	.35
140	.19	.38
150	.18	.46
160	.19	.54
170	.19	.56
180	.20	.60
190	.22	.67
200	.25	.76

Se quiere vaticinar si un nódulo de mama es o no maligno a partir de algunos datos de la mujer.

Trabajando con una muestra de casos confirmados y otra de casos con tumores benignos, se registran para unos y otros los resultados de 4 variables:

- X_1 = edad de la primera menstruación
- X_2 = número de episodios tumorales previos
- X_3 = edad de la mujer
- X_4 = número de hijos

Con estos datos se estima la función:

$$p = \text{Prob (tumor maligno)} = \frac{1}{1 + \exp(-A - \sum_{i=1}^4 B_i X_i)}$$

Para cada caso nuevo del que se conozca el vector (X_1, X_2, X_3, X_4) se puede obtener entonces la probabilidad p de que sea portadora de un tumor maligno.

Es natural procurar la determinación de una frontera F tal que se pueda aplicar la regla: si $p < F$, declarar el nódulo como benigno y declararlo maligno en caso contrario, con la consiguiente toma de medidas.

Si el cálculo de p se verifica ahora sobre otras n portadoras de nódulos (m malignos y $n-m$ benignos), se puede fijar un punto de corte —por ejemplo 0.7— y aplicar la regla de diagnosticar como maligno aquel nódulo para el cual $p > 0.7$ y como benigno al que no cumpla esa condición. Naturalmente sólo a de los m tumores malignos y sólo d de los $n-m$ benignos habrán sido correctamente diagnosticados. Las razones a/m y $d/(n-m)$ son precisamente las estimaciones de α y β para la frontera $F = 0.7$.

Definiendo, para ese juego de datos, diferentes fronteras (por ejemplo 19 entre 0.05 y 0.95 aumentando 0.05 cada vez) se puede obtener una curva ROC y también realizar un análisis como el que se resumió en la Tabla 9.

Supongamos finalmente que otro investigador afirma que las variables predictoras X_1, X_2, X_3 y X_4 no son todas realmente relevantes. A su juicio, la edad de la menarquía (X_1) debe suprimirse dada la dificultad de obtener el dato verdadero; opina además que deben incluirse otras dos variables: X_5 = consumo previo de estrógenos (sí o no) y X_6 = antecedentes de tumor maligno de la madre (sí o no).

Siguiendo el mismo método podría entonces ajustarse el modelo:

$$P(\text{tumor maligno}) = \frac{1}{1 + \exp(-A' - \sum_{i=2}^6 B'_i X_i)}$$

Aunque la calidad del ajuste de uno y otro modelo a los datos originales puede obtenerse mediante procedimientos ad hoc (ver Hosmer y Lemeshow, 1980), resultará de interés comparar ambos modelos en cuanto a sus respectivas eficiencias diagnósticas. Una vía es la de comparar las correspondientes curvas ROC. Si una de ellas está por encima de la otra a lo largo de todo el intervalo (0,1), no quedan dudas de que el primer modelo es más eficiente que el segundo. Si ambas curvas se cruzan, entonces habría que definir un indicador que permita decidir.

Supongamos que la curva ROC empírica se construyó a partir de k parejas (α_i, β_i) correspondientes a k niveles o puntos de corte.

Un posible indicador global de la eficiencia de la prueba es:

$$A = \frac{1}{2} \sum_{i=0}^{k-1} (\beta_i - \beta_{i-1}) (\alpha_i + \alpha_{i+1})$$

donde se entiende que $\alpha_0 = 0$ y $\beta_0 = 1$. (Se supone también que tal y como cabe esperar, se cumple $\alpha_i \leq \alpha_{i+1}$ y $\beta_i \geq \beta_{i+1}$ para todo i).

El indicador A expresa el área bajo la curva ROC empírica y por tanto se mueve necesariamente entre las cotas 0 y 1. Sus atributos en función de los fines con los que aquí se propone son susceptibles de posterior estudio teórico y práctico.

1.5. Apuntes finales

Múltiples aspectos de interés —no desarrollados en esta conferencia— abren la posibilidad de futura indagación. Tal es el caso de los problemas de muestreo y sesgos de selección asociados a la estimación de parámetros, aspectos parcialmente abordados en trabajos como Pearson y col. (1982), Begg y Greenes (1983) y Diamond y col. (1986).

De particular interés es el problema que surge cuando sólo una parte de los sujetos elegidos inicialmente reciben a la postre una evaluación diagnóstica definitiva.

Otro punto importante que aún espera por desarrollos rigurosos es el que se desprende de que en la práctica clínica regular las pruebas diagnósticas se verifican secuencialmente: lo usual es que se comience con una prueba relativamente imprecisa y barata; se realiza una más sofisticada en dependencia del resultado de la precedente y así sucesivamente. El trazado de una estrategia óptima que contemple simultáneamente la eficiencia desde el punto de vista médico y de los costos, demanda sin dudas atención especializada.

Otra área que consiente nuevas contribuciones se relaciona con la posibilidad de generalizar la teoría para el caso en que los resultados de la prueba no se reducen a dos. Por ejemplo, pudiera existir, además de T+ y T-, una categoría T₀ de sujetos para los que no se puede hacer clasificación nítida. Un esfuerzo inicial en esta dirección es el debido a White y Landis (1982).

BIBLIOGRAFIA

- AUSTIN, E.H. y col (1982). *Prospective evaluation of radionuclide angiocardiology for the diagnosis of coronary heart disease*. American Journal of Cardiology 50:1212-1216.
- BEGG, C.B. and GREENES, R.A. (1983). *Assessment of diagnostics tests when disease verification is subject to selection bias*. Biometrics 39:207-215.
- CONNELL, F.A. and KOESPELL, T.D. (1985). *Measures of gain in certainty from a diagnostic test*. American Journal of Epidemiology 121:744-763.
- DAWID, A.P. and SKENE, A.M. (1979). *Maximum likelihood estimation of observer error-rates using the EM algorithm*. Applied Statistics 28:20-28.
- DEMPSTER, P.A. y col. (1977). *Maximum likelihood from incomplete data via the EM algorithm*. Journal of the Royal Statistics Society, Series B 39:1-38.
- DIAMOND, G.A. y col. (1986). *A model for assessing the sensitivity and specificity of tests subject to selection bias*. Journal of Chronic Diseases 39:343-355.
- DOBIE, R.A. (1985). *The use of relative cost ratios in choosing a diagnostic test*. Ear and Hearing 6(2):113-116.
- FEINSTEIN, A.R. (1977). *Clinical biostatistics*. C.V. Mosby Company, St. Louis.
- GALEN, R.S. and GAMBINO, S.R. (1975). *Beyond normality: the predictive value and efficiency of medical diagnosis*. Wiley and Sons. New York.
- GREENBERG, R.A. and JEKEL, J.F. (1969). *Some problems in the determination of the false positive and false negative rates of tuberculin tests*. American Review of Respiratory Disease 100:645-650.
- HILDEN, J. (1980). *The concept of medical usefulness in clinical chemistry and the difficulty of applying it to questions of research policy*. Scandinavian Journal of Clinical Laboratory Investigation 155:31-46.
- HILDEN, J. (1981). *Sensible or specious? Another critical look at sensitivity and specificity*. Institute of Medical Genetics, University of Copenhagen (mimeo).
- HILL, A.B. (1975). *Principios de estadística médica*. Editorial R, La Habana.
- HOSMER, D.W. and LEMESHOW, S. (1980). *Goodness of fit for the multiple logistic regression model*. Communications in Statistics A9:1043-1069.
- HUI, S.L. and WALTER, S.D. (1980). *Estimating the error rates of diagnostic tests*. Biometrics 36:167-171.
- LILIENFELD, A.M. y LILIENFELD, D.E. (1983). *Fundamentos de epidemiología*. Fondo Educativo Interamericano. México DF.
- McCLISH, D. and QUADRE, D. (1985). *Improving estimates of prevalence by repeated testing*. Biometrics 41:81-89.
- METZ, C.E. (1978). *Basic principles of ROC analysis*. Seminars of Nuclear Medicine 8:283-298.

- NISSEN-MAYER, S. (1964). *Evaluation of screening tests in medical diagnosis*. *Biometrics* 20:730-750.
- PEARSON, T. y col. (1982). *Selection bias in persons undergoing coronary arteriography*. *American Journal of Epidemiology* 116:568 (resumen).
- QUADE, D. y col. (1980). *Effects of misclassifications on statistical inferences in epidemiology*. *American Journal of Epidemiology* 111:503-515.
- ROBERTSON, E.A. y col. (1983). *Evaluating the clinical efficacy of laboratory test*. *American Journal of Clinical Pathology* 79(1):78-86.
- ROGAN, W.J. and GLADEN, B. (1979). *Estimating prevalence from the results of a screening test*. *American Journal of Epidemiology* 107:71-76.
- SANDIFER, M.G. y col. (1968). *Sampling selection by diagnosis in clinical drug evaluation*. *Psychopharmacology* 13:118-128.
- SCHECHTER, M.T. and SHEPS, S.B. (1985). *Diagnostic testing revisited: pathways through uncertainty*. *The Canadian Medical Association Journal* 132:755-760.
- SOX, H.C. (1986). *Probability theory in the use of diagnostic tests*. *Annals of Internal Medicine* 104(1):60-66.
- STAQUET, M. y col. (1981). *Methodology for the assessment of new dichotomous diagnostic tests*. *Journal of Chronic Diseases* 34:599-610.
- SUSSER, M. (1973). *Causal thinking in health sciences*. Oxford University Press. New York.
- TAUBE, A. (1986). *Sensitivity, specificity and predictive values: a graphical approach*. *Statistics in Medicine* 5:585-591.
- VACEK, P.M. (1985). *The effect of conditional dependence on the evaluation of diagnostic tests*. *Biometrics* 41:959-96.
- THIBODEAU, L.A. (1981). *Evaluating diagnostic tests*. *Biometrics* 37:801-804.
- WEINER, D.A. y col. (1979). *Exercise stress testing: correlations among history of angina, ST-segment response and prevalence of coronary artery disease in the CASS*. *New England Journal of medical Science* 301:230-235.
- WHITE, A.A. and LANDIS, J.R. (1982). *A general categorical data methodology for evaluating medical diagnostic tests*. *Communications in Statistics* 11:567-605.
- WONG, E.T. and LINCOLN, T.L. (1983). *Ready' Eire' Aim': an inquiry into laboratory test ordering*. *Journal of the American Medical Association* 250:2510-2513.
- YERUSHALMY, J. (1947). *Statistical problems in assessing methods of medical diagnosis, with special reference to X-ray techniques*. *Public Health Reports* 62:1432-1449.
- YOU DEN, W.J. (1950). *Index for rating diagnostic tests*. *Cancer* 3:32-35.
- ZELEN, M. and FEINLAB, M. (1969). *On the theory of screening for chronic diseases*. *Biometrika* 56:601-614.

ANEXO 1. Programa SCH.BAS para el cálculo de probabilidades a posteriori.

```
10 DEFINT A-J
20 DEFSNG L-S
30 DIM B(19), BP(19), BN(19)
40 CLS:PRINT 'ENTRE SENSIBILIDAD(%), ESPECIFICIDAD(%), PROBABILIDAD
PRE-TEST%)'
50 PRINT:PRINT 'Por ej.: 90, 60, 15'
60 PRINT:INPUT SE, SP, P
70 LP=SE/(100!-SP)
80 LN=(100!-SE)/SP
90 DEF FNPOST(Q, L) = (100!*Q*L) / (100!-Q+Q*L)
100 PTLP=FNPOST(P, LP)
110 PTLN=FNPOST(P, LN)
120 CLS
130 PRINT 'SENSIBILIDAD', 'ESPECIFICIDAD', SPC(2) 'P'
140 PRINT SE, SP, P
150 PRINT:PRINT TAB(15) 'Diagrama de árbol'
160 PRINT TAB(15) '-----'
170 PRINT TAB(50) PTLP
180 FOR I=44 TO 20 STEP -6
190 PRINT TAB(I) '*'
200 NEXT I
210 PRINT TAB(10) P
220 FOR I=20 TO 44 STEP 6
230 PRINT TAB(I) '*'
240 NEXT I
250 PRINT TAB(50) PTLN
260 PRINT 'Teclee cualquier tecla para continuar'
270 X$=INKEY$: IF LEN(X$)=0 THEN 270
280 PRINT
290 PRINT TAB(15) 'P(-)', TAB(33) 'P', TAB(45) 'P(+)'
300 PRINT TAB(15) '-----', TAB(33) '--', TAB(45) '-----'
310 FOR J=1 TO 19
320 B(J)=5*J
330 BP(J)=FNPOST(B(J), LP)
340 BN(J)=FNPOST(B(J), LN)
350 PRINT TAB(17) BN(J), TAB(32) B(J), TAB(47) BP(J)
360 NEXT J
370 PRINT 'Para recomenzar: SI; para terminar: NO'
380 INPUT Y$
390 IF Y$='SI' THEN 40
400 IF Y$='NO' THEN 410 ELSE 370
410 END
```

ANEXO 2. Programa EM.BAS para estimación simultánea de la sensibilidad, especificidad y prevalencia.

```
10 CLEAR; C(0)=1:C(1)=3:C(2)=3:C(3)=1:LC=.01
20 DIM P(100), A(100), B(100)
30 FOR I=0 TO 3
40 PRINT 'N(' ; I ; ')=' ; : INPUT N(I) : SN=SN+N(I)
50 NEXT I
60 PRINT : PRINT
70 PRINT 'Dar número supuesto de enfermos: ' : PRINT
80 FOR I=0 TO 3
90 PRINT 'Para los que tuvieron' ; I ; 'resultados negativos' ; : INPUT
    R(I) : S(I)=N(I)-R(I)
100 NEXT I
110 PRINT : PRINT 'SENS', 'ESPEC', 'PREV' : PRINT
120 T=T+1
130 FOR I=0 TO 3
140 SR=SR+R(I) : SS=SS+S(I)
150 NEXT I
160 FOR I=1 TO 3
170 A=A+I*R(I) : B=B+(4-I)*S(I-1)
180 NEXT I
190 A(T)=A/3/SR : B(T)=B/3/SS : P(T)=SR/SN : SR=0 : SS=0
200 FOR I=0 TO 3
210 R(I)=SN*P(T)*(1-A(T))^(3-I)*A(T)^I*C(I)
220 S(I)=SN*(1-P(T))*(1-B(T))^(3-I)*B(T)^I*C(I)
230 R(I)=N(I)*R(I)/(R(I)+S(I)) : S(I)=N(I)-R(I)
240 NEXT I
250 PRINT A(T), B(T), P(T) : A=0 : B=0
260 IF T=1 THEN 120
270 E1=ABS(A(T)-A(T-1))
280 E2=ABS(B(T)-B(T-1))
290 E3=ABS(P(T)-P(T-1))
300 IF E1>LC OR E2>LC OR E3>LC THEN 120 ELSE PRINT : PRINT 'Convergencia' :
    END
```

2. EVALUACION DE ENCUESTAS SEROEPIDEMIOLOGICAS

2.1. Introducción

La vigilancia epidemiológica constituye una tarea permanente de la Salud Pública de cualquier país cuyos principios han venido delineándose desde la primera mitad del siglo XIX.

En su concepción más general esta actividad transita por la recolección sistematizada de datos sobre el acaecimiento de enfermedades, su análisis e interpretación, y la difusión de la información ya procesada (Raska, 1966). Una forma de tal vigilancia se concreta a través de las encuestas seroepidemiológicas, particularmente importantes en relación con ciertas enfermedades virales (hepatitis B, dengue, etc.).

Con el propósito de identificar la situación presente y pasada en relación con enfermedades infecciosas mediante pruebas de anticuerpos y antígenos se recurre periódicamente a la recolección y evaluación de muestras de sangre obtenidas de la población de interés, usualmente mediante una muestra representativa de ella.

Ante una muestra serológica concreta puede evaluarse la presencia o ausencia de anticuerpos, lo cual indica si el sujeto ha estado o no en contacto con la enfermedad respectivamente. Esta información cualitativa permite la estimación de la «tasa de prevalencia de anticuerpos» (TPA) en la población mediante la razón del número de sujetos «seropositivos» entre el número de examinados. Puesto que muchos anticuerpos permanecen en el organismo durante decenas de años, ocasionalmente toda la vida, esa tasa refleja la experiencia

acumulada —pasada y presente— de la población con el agente infeccioso (Evans, 1976).

Por otra parte, no todos los sujetos tienen igual densidad de anticuerpos. En efecto, además de la posibilidad de detectar la presencia de anticuerpos en la sangre, existen técnicas para cuantificar el grado en que esa presencia se verifica. A la variable que refleja esa densidad y cuya forma de medición se examina más adelante, se le denomina «Título de anticuerpos». Consecuentemente, otro parámetro importante puede estimarse en la población: la llamada «línea de base» (LB) que no es otra cosa que la mediana de la distribución de títulos. Ella constituye una medida de resumen, complementaria a la TPA y que sintetiza la intensidad del contacto de los seropositivos con el agente.

Además de constituir descriptores de interés intrínseco a los efectos de caracterizar el nivel inmunológico de la población, el TPA y la LB tienen una virtualidad detectora en relación con la aparición de un nuevo foco. Un incremento apreciable en estos parámetros respecto de un corte evaluativo previo es indicio de la irrupción de un brote epidémico, cuya virulencia se refleja, especialmente, en el incremento de la LB.

Ahora bien, el proceso de la medición de los títulos y la determinación de positividad a él asociado, exhiben características sui-géneris. El tratamiento estadístico que a partir de ello se ha dado a la información proveniente de las encuestas, consiente un análisis detenido y resulta susceptible de

refinamiento. Ese es el propósito central del presente trabajo.

2.2. Medición de título de anticuerpos

El procedimiento habitual para determinar la concentración de anticuerpos a cierto virus en la sangre —llamémosle C — puede describirse del modo siguiente.

La presencia de anticuerpos se evalúa poniendo en contacto el suero con un antígeno adecuado; si el suero contiene anticuerpos en cantidad suficiente, se produce una reacción observable en el mismo.

El método consiste en determinar la máxima dilución del suero a la cual aún se produce reacción. Para ello se crea una sucesión de preparados de suero diluido en la que el grado de concentración del suero va disminuyendo. Usualmente el grado de dilución se hace crecer en progresión geométrica; por ejemplo, puede ocurrir que el primer preparado sea de 1 en 4 (de cada 4 partes, una corresponde al suero, lo que se denota 1:4), la del segundo de 1 en 8, la del tercero de 1 en 16, etc. Si la máxima dilución a la que aún se mantiene la reacción fuese —digamos— 1 en 32, esto significa que el suero debe contener 32 veces la concentración de anticuerpos requerida para la reacción.

En general, supongamos que la sucesión es:

$$1:C_0 \quad 1:C_1 \quad \dots \quad 1:C_k$$

Los recíprocos C_0, C_1, \dots, C_k se definen a partir de C_0 , de suerte que $C_i = \alpha C_{i-1}$ para $i=1, \dots, k$, donde α es la razón de la progresión geométrica ($\alpha > 1$).

Regularmente se trabaja con $\alpha = 2$ en tanto que k se define de manera que al nivel de concentración $1:C_k$ ya no haya suero alguno para el que se registre la presencia de anticuerpos.

Como se verá más adelante, un punto de singular importancia a los efectos del análisis del problema es que siempre se trabaja con $C_0 > 1$. Las elecciones más comunes para C_0 son 2, 4, 8, 10, 16 y 64 dependiendo del problema en cuestión. El caso $C_0 = 1$ (no se diluye en absoluto el suero), se evita por razones técnicas bien conocidas de los inmunólogos. En tal caso, hay un gran riesgo de reacciones inespecíficas; vale decir: las reacciones que se verifican no necesariamente

pueden atribuirse al tipo de anticuerpo que se estudia.

Resumiendo con un ejemplo, supongamos que $C_0 = 10$, $\alpha = 2$ y $k = 6$; se trabaja entonces con 7 concentraciones: 1:10, 1:20, 1:40, 1:80, 1:160, 1:320, 1:640. Supongamos que un suero dado reacciona ante el antígeno para las 3 primeras concentraciones, pero ya deja de reaccionar a partir de la cuarta inclusive. Se la atribuye entonces el título 1:40.

Ahora bien, debe notarse que —de hecho— lo que se sabe es que la concentración a la que reacciona es superior a 1:40 pero inferior a 1:80. O sea, el valor de C —que en general es un número real no negativo— no puede determinarse unívocamente; de él sólo se conoce que cumple la condición $40 \leq C < 80$.

2.3. Problema de estimación

Supongamos provisionalmente que la concentración de anticuerpos de cada sujeto puede conocerse exactamente.

Bajo ese supuesto, si se incluyen n' sujetos en el estudio, se tendrían respectivos valores de la variable: c_1, c_2, \dots, c_n .

La estimación de T —suponiendo, como es usual, que la muestra es simple aleatoria— sería simplemente:

$$t^* = \frac{m}{n'}$$

donde m es el número de sueros para las cuales $C > 0$.

Por otra parte, en la literatura especializada (White, 1973) se admite el supuesto (empíricamente corroborado) de que la distribución de C para los sujetos seropositivos ($C > 0$) es la logarítmico-normal. O sea; $\ln C \sim N(\mu, \sigma^2)$.

Si llamamos M a la mediana de esta distribución; por definición se tiene que

$$P\{C < M\} = 1/2.$$

$$\text{Ahora } P\{C < M\} = P\left\{Z < \frac{\ln M - \mu}{\sigma}\right\}$$

$$\text{donde } Z = \frac{\ln C - \mu}{\sigma} \sim N(0,1).$$

y de aquí se deduce que $\ln(M) = \mu$; o sea $M = \exp(\mu)$.

Si c_1, \dots, c_n representa la submuestra de valores de C mayores que 0, entonces $(\sum_{i=1}^n \ln c_i) / n$ es el estimador máximo verosímil de μ y de aquí se desprende que:

$$\hat{M} = \exp\left(\frac{1}{n} \sum_{i=1}^n \ln c_i\right)$$

sería un estimador de M .

De las propiedades del logaritmo y el exponencial, se deduce que \hat{M} no es otra cosa que la media geométrica (MG) de las observaciones:

$$\hat{M} = \sqrt[n]{\prod_{i=1}^n c_i}$$

Este es el origen de que se sugiera (aunque sin justificación explícita en la literatura consultada) el uso de la MG en la estimación de la LB.

Sin embargo el problema presenta ciertas singularidades en la práctica como consecuencia de la forma en que se realiza la determinación de las concentraciones y que se explicó en la sección precedente.

Asumamos que n' personas han sido incluidas en una encuesta serológica. Llamemos f_i al número de sueros para los que se produjo reacción en la $(i-1)$ -ésima concentración pero no para la i -ésima ($i:1, \dots, k$).

Los resultados pueden resumirse del modo siguiente:

Título	$0 \leq C < C_0$	$C_0 \leq C < C_1$	\dots	$C_{k-1} \leq C < C_k$
Frecuencia	f_0	f_1	\dots	f_k

El procedimiento usual entre los inmunólogos para estimar T es calcular miméticamente:

$$t = n / n'$$

donde $n = \sum_{i=1}^k f_i$

Sin embargo, al hacerlo se está asumiendo que C es 0 en todos los sujetos para los que $C < C_0$.

En realidad existe una parte de los f_i individuos que tienen efectivamente anticuerpos, sólo que en concentración tan baja que éstos no se detectan ni siquiera para la primera dilución considerada (convirtiéndose así en falsos negativos).

Esto tiene importancia obvia ya que es cualitativamente muy diferente el caso $C = 0$ (no estuvo en contacto con el agente) al caso $0 < C < C_0$ (estuvo en contacto, pero el rastro dejado por el agente fue débil). Consecuentemente, t produce una subestimación de la TPA.

En cuanto al tratamiento usual para la estimación de la LB, la situación es mucho más caótica. Se observa una notable falta de concordancia en los procedimientos seguidos por distintos autores.

Entre las diversas maneras en que se estima M , la más usada es la que sugiere White (1973):

$$m = \exp\left\{\left[\left(f_0 \ln(C_0/\alpha) + \sum_{i=1}^k f_i \ln C_{i-1}\right) / \left(\sum_{i=0}^k f_i\right)\right]\right\}$$

Ejemplos del uso de ese estimador en situaciones reales pueden verse en Blackwelder (1982), Bond y Lewis (1970), Crossman (1973), Kendall (1978), Tong Ming (1977).

Esta forma de estimación es susceptible de varias observaciones críticas. Por una parte, a los f_i sujetos cuyas concentraciones están en el intervalo $[C_{i-1}, C_i]$ se les atribuye el extremo izquierdo. Por otra parte, se procura tener en cuenta el hecho de que no todas las concentraciones inferiores a C_0 son nulas: se le atribuye la concentración C_0/α a los f_0 sujetos que están en ese caso; sin embargo, esa estrategia resulta arbitraria y, en última instancia, insuficiente.

2.4. Una solución alternativa

Supongamos nuevamente que se tienen n' determinaciones de la variable C pero que sólo se tiene acceso a los n valores mayores que cierto C_0 . Estamos en presencia entonces de una muestra de una distribución lognormal truncada a la izquierda en el punto C_0 . Equivalentemente, lo que se tiene es una muestra de la distribución normal truncada a la izquierda en $X_0 = \ln C_0$, que corresponde a la variable $X = \ln C$.

La función de distribución de X es la siguiente (Cramér, 1946):

$$F_{x_0}(x) = P(X/X \geq X_0) = \frac{\Phi\left(\frac{x-\mu}{\sigma}\right) - \Phi\left(\frac{X_0-\mu}{\sigma}\right)}{1 - \Phi\left(\frac{X_0-\mu}{\sigma}\right)}$$

donde Φ es la función de distribución normal estándar:

$$\Phi(t) = \int_{-\infty}^t f(h) dh \quad \text{y} \quad f(t) = [\exp(-t^2/2)] / \sqrt{2\pi}$$

El propósito es estimar μ y σ^2 a partir de la muestra disponible. Se usará para ello el método de los momentos.

Llamando $\beta_0 = \frac{X_0 - \mu}{\sigma}$ y derivando se obtiene que la función de densidad es:

$$f(x) = \frac{\lambda}{\sigma} \quad \text{donde} \quad \lambda = \frac{f(\beta_0)}{1 - \Phi(\beta_0)} \quad (1)$$

Es fácil deducir ahora el valor esperado y la varianza de X:

$$E_T = E_{X_0}(X) = E(X/X > X_0) = X_0 + \sigma(\lambda - \beta_0) \quad (2)$$

$$V_T = V_{X_0}(X) = E(X^2/X > X_0) - E_{X_0}^2(X) = \sigma^2(1 + \beta_0\lambda - \lambda^2) \quad (3)$$

La idea ahora es estimar los términos de la izquierda mediante la muestra a la que se tiene acceso (la distribución empírica truncada) y resolver el sistema de ecuaciones de suerte que puedan estimarse μ y σ^2 , parámetros que definen la verdadera distribución de C.

Trabajando sólo con f_1, f_2, \dots, f_k definidos en la página 16, es inmediato que:

$$\hat{E}_T = (\sum_{i=1}^k f_i H_i) / n$$

$$\hat{V}_T = (\sum_{i=1}^k f_i H_i^2 - n \hat{E}_T^2) / n$$

donde $H = \ln(\sqrt{C_i C_{i-1}})$.

Sustituyendo en (2) y (3) y dividiendo la segunda entre la primera se tiene:

$$\frac{\hat{V}_T}{(\hat{E}_T - X_0)^2} = \frac{1 - \lambda(\lambda - \beta_0)}{(\lambda - \beta_0)^2}$$

Por la monotonía de la función

$$\psi(\beta) = \frac{1 - \lambda(\lambda - \beta)}{(\lambda - \beta)^2}$$

en un entorno de cualquier valor de β_0 que razonablemente pueda esperarse (por ejemplo, en el intervalo $[-3, 3]$), se puede estimar mediante:

$$\hat{\beta}_0 = \psi^{-1}\left(\frac{\hat{V}_T}{(\hat{E}_T - X_0)^2}\right)$$

Sustituyendo en (1), se tiene

$$\hat{\lambda} = \frac{f(\hat{\beta}_0)}{1 - \Phi(\hat{\beta}_0)} \quad \text{y de (2) se obtiene:}$$

$$\hat{\sigma} = \frac{\hat{E}_T - X_0}{\hat{\lambda} - \hat{\beta}_0}$$

Finalmente, como $\beta_0 = \frac{X_0 - \mu}{\sigma}$, se tiene:

$$\hat{\mu} = X_0 - \hat{\sigma}\hat{\beta}_0$$

Cabe comentar que mediante el método de máxima verosimilitud se arriba a los mismos resultados. (Ver Cohen, 1957).

Una vez en posesión de $\hat{\mu}$ y $\hat{\beta}_0$, las estimaciones de T y M son inmediatas:

$$\hat{T} = \frac{n}{n' (1 - \Phi(\hat{\beta}_0))}$$

$$\hat{M} = \exp(\hat{\mu})$$

A modo de ejemplo, consideremos la siguiente muestra: $k = 5$, $C_0 = 10$, $\alpha = 2$, $f_0 = 235$, $f_1 = 145$, $f_2 = 112$, $f_3 = 56$, $f_4 = 21$ y $f_5 = 5$. El cálculo de \hat{M} y \hat{T} para estos datos da:

$$\hat{M} = 15.48 \quad \text{y} \quad \hat{T} = 0.88$$

(Para este cálculo puede usarse el programa VIRUS.BAS, que se anexa). La estimación regular sería $m = 10.91$ y $t = 0.59$.

2.5. Evaluación del procedimiento

Una vía para evaluar el método aquí sugerido es por conducto de la simulación.

El procedimiento sería el siguiente:

1. a) fijar μ , σ^2 , C_0 y α
- b) fijar un valor para TPA, digamos, T_0
- c) fijar un tamaño de muestra general n

A partir de esas delimitaciones se tendría que $n(T_0)$ sujetos serían los seropositivos en la muestra (aquellos con $C > 0$).

2. En este punto puede calcularse la distribución exacta de individuos que —de acuerdo con la distribución logarítmico-normal definida por μ y σ^2 — caerían en cada clase; o sea, pueden hallarse $f_0^*, f_1, f_2, \dots, f_k$ donde f_i es el número de sujetos cuya concentración está entre C_{i-1} y C_i y f_0^* es el número de individuos para los que $0 < C < C_0$. El número k de clases es aquél necesario para cubrir la muestra completa (o sea, se van calculando f_i hasta llegar a $f_{k+1} = 0$).

Más concretamente, se calculan:

$$f_i = n' T_0 \left[\Phi \left(\frac{\ln C_i - \mu}{\sigma} \right) - \Phi \left(\frac{\ln C_{i-1} - \mu}{\sigma} \right) \right]$$

para $i: 1, 2, \dots, k$ y

$$f_0^* = n' T_0 \left[1 - \Phi \left(\frac{\ln C_0 - \mu}{\sigma} \right) \right]$$

donde k es tal que $\Phi \left(\frac{\ln C_k - \mu}{\sigma} \right) = 1$

Poniendo ahora $f_0 = n' (1 - T_0) + f_0^*$ se ha completado la simulación de una muestra real como la descrita en la Sección III.

2.6. Resultados de la simulación

Para llevar adelante la comparación se escogieron diferentes juegos de parámetros. La Tabla 1 refleja algunos de los resultados obtenidos con los estimadores usuales (t y m) y los modificados (\hat{T} y \hat{M}). En este caso, T y M se prefijan como parte del modelo y son por tanto conocidos. Ello permite comparar la eficiencia entre los dos métodos en juego.

En todos los casos se simularon muestras de tamaño $n' = 1000$ y se prefijó $a = 2$ que es la razón escogida para las progresiones geométricas en la inmensa mayoría de los estudios.

La Tabla recoge lo que a nuestro juicio constituye una «muestra representativa» de las decenas de casos simulados. Como se ve, se han elegido valores de μ que van de 1 a 8 (es decir M entre 2.7 y 2981.0) ya que valores fuera de ese rango son ajenos a los que prevalecen en la práctica independientemente del virus en estudio. Para la elección de σ^2 , C_0 y T_0 se eligieron diversos valores procurando recorrer una amplia gama de posibilidades.

El primer resultado que se observa claramente es que \hat{T} siempre superior a t . O sea, formalmente siempre se tiene:

$$|\hat{T} - T_0| \leq |t - T_0|$$

En cuanto a la estimación de M , los esti-

Tabla 2. Estimaciones de T_0 y M por los dos métodos (usual y modificado) para muestras de tamaño 1000 y progresiones de razón 2 simuladas según distintos parámetros prefijados (μ, σ, C_0, t_0).

Parámetros de partida			TPA y estimadores			LB y estimadores		
μ	σ^2	C_0	T_0	t	\hat{T}	M	m	\hat{M}
1	2	2	.40	.23	.36	2.7	1.5	3.4
2	.5	8	.30	.14	.18	7.4	4.5	11.7
3	2	10	.20	.14	.19	20.1	6.4	22.5
4	1	20	.30	.25	.29	54.6	15.2	55.5
5	.5	20	.70	.44	.44	148.4	34.5	112.5
5	13	10	.50	.39	.50	148.4	27.9	154.8
6	2	10	.50	.50	.50	403.4	37.7	402.0
8	4	20	.50	.50	.50	2981.0	144.5	2958.7

3. Ahora se usa la muestra para estimar T_0 y $M = \exp(\mu)$ mediante los procedimientos convencionales: t y m , mediante los estimados modificados: \hat{T} y \hat{M} . Puesto que T_0 y M son —en este caso— conocidos, la comparación resulta altamente elocuente.

madores m y \hat{M} son esencialmente similares en cuanto a precisión para valores pequeños de μ ($\mu \leq 2$) pero \hat{M} es *muchísimo* más preciso que m para valores mayores de μ . Cabe destacar que en la práctica real, lo más común es, precisamente, tener valores de μ mayores que 3.

BIBLIOGRAFIA

- BLACKWELDER, W.C. et al (1982). *A population study of herpesvirus infections and HLA antigens*. American Journal of Epidemiology 115:569-576.
- BOND, J.O. and LEWIS, A.L. (1970). *Persistence of antibody to St. Louis encephalitis virus in survivors of Florida epidemic*. American Journal of Epidemiology 91:593-600.
- COHEN, A.C. (1957). *On the solution of estimating equations for truncated and censored samples from normal population*. Biometrika 44:225-236.
- CRAMER, H. (1946). *Mathematical methods on statistics*. Princeton University Press, Princeton.
- CROSSMAN, R.A. et al (1973). *Study of Japanese encephalitis virus in Chiang Mai Valley, Thailand*. American Journal of Epidemiology 98:133-149.
- EVANS, A.S. (1976). *Surveillance and seroepidemiology*. In «Viral infections of humans» (Alfred S. Evans, Ed.), Ch. 2, Plenum Publishing Corporation, New York, 1976.
- KENDALL, A.P. et al (1978). *C-circulation of two influenza A (H3N2) antigenic variants detected by virus surveillance in individual communities*. American Journal of Epidemiology 108:303-311.
- RASKA, A.S. (1966). *National and international surveillance of communicable diseases*, WHO Chron. 20:313-321.
- TONG MING LIN et al (1977). *Antibody to E-B virus capsid antigen in nasopharyngeal carcinoma comparison groups*. American Journal of Epidemiology 106:336-339.
- WHITE, C. (1973). *Statistical methods in serum surveys* en «Serological Epidemiology», J.R. Paul and C. White Eds., Academic Press, N.Y.

ANEXO: Programa VIRUS:BAS para el cómputo de \hat{T} y \hat{M} a partir de una muestra de títulos organizados en clases y frecuencias.

```

10 CLEAR
20 DIM C(30), F(30), H(30)
30 INPUT 'N.º de sujetos que no reaccionaron (negativos): '; F(0):PRINT
40 INPUT 'N.º de clases con sujetos positivos: '; K:PRINT
50 INPUT 'Razón de la progresión geométrica: '; A:PRINT
60 INPUT 'Primera concentración: '; C(0):X0=LOG(C(0)):PRINT
70 PRINT 'N.º de sujetos que reaccionan a la concentración: '
80 FOR I=1 TO K
90 PRINT I; : INPUT F(I):N=N+F(I)
100 C(I)=C(0)*A^I
110 CH(I)=LOG(C(I-1)):H(I)=LOG(SQR(C(I-1)*C(I)))
120 EET=EET+F(I)*CH(I)
130 VT=VT+F(I)*H(I)^2:ET=ET+F(I)*H(I)
140 NEXT I
150 PRINT:TL=N/(N+F(0))
160 ET=ET/N
170 VT=(VT-N*ET^2)/N
180 W=VT/(ET-X)^2
190 DEF FNF(X)=1/SQR(2*3.14159)*EXP(-X^2/2)
200 LI=-3:LS=3:GOSUB 2000
210 SIGNUE=(ET-X0)/(R-LM):MUNUE=X0-SIGNUE*LM
220 TNUE=TL/Z:MGNUE=EXP(MUNUE)
230 MGL=EXP((EET+F(0)*LOG(C(0)/A)/(N+F(0)))
240 PRINT 0; C(0), F(0)
250 FOR I=1 TO K
260 PRINT C(0)*A^(I-1); C(0)*A^I, F(I)
270 NEXT I
280 PRINT
290 PRINT 'Método', 'Tasa', 'Línea de base'
300 PRINT
310 PRINT 'Usual', TL, MGL
320 PRINT 'Modificado', TNUE, MGNUE
330 PRINT:PRINT
340 END
1000 IF U>=0 THEN B=U ELSE B=-U
1010 T=1/(1+.231642*B)
1020 YY=EXP(1)
1030 Q=1/SQR(2*3.14159)*YY^(-B*B/2)
1040 AA=.31938153#:BB=-.3565637882#:CC=1.78147937#:
DD=-1.821255978#:EE=1.330274429#
1050 Z=1-Q*(AA*T+BB*T^2+CC*T^3+DD*T^4+EE*T^5)
1060 IF U<0 THEN Z=1-Z
1070 RETURN
2000 IF ABS(LI-LS)>=.01 THEN 2020
2010 RETURN
2020 LM=(LI+LS)/2:GOSUB 2060
2030 IF FF>=0 THEN 2050
2040 LI=LM:GOTO 2000
2050 LS=LM:GOTO 2000
2060 UU=FNF(LM)
2070 U=-LM:GOSUB 1000:R=UU/Z
2080 FF=(1-R*(R-LM))/(R-LM)^2-W
2090 RETURN

```

3. SOBRE LOS METODOS ESTADISTICOS EN EPIDEMIOLOGIA

3.1. Introducción

Si bien el pensamiento epidemiológico contemporáneo reposa sobre antecedentes que se ubican en el siglo pasado y en la primera mitad del presente, no es hasta las décadas del 50 y el 60 que, con los famosos trabajos que arrojan luz sobre la relación causal existente entre el hábito de fumar y el cáncer pulmonar y sobre las circunstancias que condicionan el desarrollo de enfermedades coronarias, se comienza a delinear lo que ha dado en llamarse «epidemiología moderna». Como ciencia con objeto de estudio delimitado, esta rama se ha venido consolidando en el último vicenio, con la conformación de un cuerpo teórico sistematizado.

Una definición simple pero comúnmente aceptada es la que Cole (1979) atribuye a G. Anderson: «el estudio de la ocurrencia de las enfermedades». Otras definiciones, por ejemplo la de Lilienfeld y Lilienfeld (1980), se basan en los métodos y propósitos que se le atribuyen a esta rama de la ciencia.

Ellos pueden resumirse en lo fundamental del modo siguiente:

- descripción del estado de salud de la población mediante indicadores e identificación de tendencias.
- predicción de magnitud y forma que puede asumir ese estado.
- explicación de los mecanismos de transmisión de los procesos morbosos e identificación de sus causas.

- control de las enfermedades en el sentido de prevenir casos nuevos, erradicar los existentes y disminuir sus efectos con particular orientación hacia la prolongación de la vida.

Debe enfatizarse que, pese al impetuoso desarrollo alcanzado por esta disciplina en función de tan amplio espectro de propósitos —o, quizás, precisamente debido a él— existen numerosas zonas matizadas por la controversia. En efecto, no es difícil advertir la falta de consenso en materia de nomenclatura, enfoques y definiciones.

El nuevo pensamiento epidemiológico ha desafiado incluso la validez de la división clásica entre enfermedades «crónicas» y enfermedades «agudas» o «infecciosas». Históricamente se desarrolló la *epidemiología de las enfermedades infecciosas* con métodos específicos para el estudio y control de las enfermedades transmisibles. Durante mucho tiempo incluso, el concepto de epidemiología se identificó exclusivamente con las enfermedades que —como el tifus o el cólera— afectan a muchas personas simultáneamente y como resultado de un brote explosivo. Luego se amplió considerablemente el concepto para abarcar a las dolencias como el cáncer cuyo lapso de inducción (período entre la aparición de la causa y el establecimiento de la dolencia o de sus síntomas) era prolongado, y surgió lo que dio en llamarse *epidemiología de las enfermedades crónicas*.

Sin embargo, la existencia de fenómenos morbosos tales como los derivados de un politrau-

matismo de origen accidental (que no se inscriben ni en una ni en otra definición) han llevado a algunos a declarar que tal división se resiente de un carácter fosilizado. (Rothman, 1986).

A mi juicio, sin embargo, la diferenciación genérica de esas ramas de la epidemiología es esencialmente correcta y, sobre todo, fecunda ya que —en efecto— los métodos de análisis y el pensamiento subyacente se distinguen nítidamente entre sí.

Por otra parte, el empuje alcanzado mundialmente por la política de poner el énfasis en la atención primaria y la orientación preventiva de la Salud Pública ha venido a subrayar el interés por las enfermedades crónico-degenerativas y sus causas y, por tanto, por los métodos epidemiológicos a ellas asociados. Finalmente, otros fenómenos con impacto en la salud de la comunidad, tales como el suicidio o los accidentes, si bien nada tienen de «crónicos» en el sentido clásico, son susceptibles de ser estudiados, prevenidos y controlados con ayuda de procedimientos análogos a los que se usan para encarar otros males que sí lo son, tales como las cardiopatías o la diabetes.

Para su desempeño, la epidemiología se vale de múltiples disciplinas pertenecientes a diversas áreas de ciencias. Por ejemplo, la fisiología, la virología y la patología —del universo biomédico— se combinan con ciencias sociales tales como la sicología, la sociología y la economía. La matemática, la demografía, investigación de operaciones y particularmente, la estadística se insertan de manera singular en un entramado cuya armonía va consolidándose rápidamente. Esta última disciplina ha tenido una creciente incidencia en el pensamiento epidemiológico. Múltiples son los procedimientos concebidos al calor de nuevos problemas y sucesivos los refinamientos para enfrentarlos.

Paralelamente a la inestimable contribución de notables estadísticos (Mantel, Fliess, Cornfield, Cox, Miettinen, etc.) las aplicaciones de las herramientas estadísticas a la epidemiología no han estado exentas de cierto mimetismo que en ocasiones ha conducido al arraigo de concepciones erróneas, sobre todo en materia de interpretación.

En años recientes han surgido varios libros orientados a la metodología de los estudios epidemiológicos tales como Kleinbaum y col. (1982), Schlesselman (1982), Breslow and Day (1980) y Rothman (1986).

En las secciones siguientes se intenta dar una visión global sobre algunos conceptos y procedimientos básicos utilizados para el desarrollo de las tareas que en la epidemiología de las enfermedades crónicas se resuelven con la participación singular de la estadística.

Sin embargo el énfasis esencial —y en cierto sentido el «leit motiv» de todo el texto— está en la evaluación de la asociación entre variables. Como es obvio, la existencia de correlación entre una variable y otra es condición inexcusable aunque, desde luego, no suficiente para declarar una relación causal entre ellas.

Siendo el descubrimiento y la descripción de los nexos causales entre los fenómenos uno de los propósitos más importantes de la ciencia y, muy particularmente, de la que ahora nos ocupa, el asunto adquiere especial relevancia. Adviértase que con excepción de la actividad meramente descriptiva, ninguna de las tareas que se identificaron como características de la epidemiología puede concebirse al margen de las relaciones de causalidad. Por otra parte la descripción nunca se erige como finalidad en sí misma sino que cobra su máximo valor en la medida que sea capaz de estimular el planteamiento de hipótesis explicativas de los mecanismos según los que se encadenan los fenómenos estudiados.

Así, en este caso, la asociación que más interesa es la que pueda exhibir un factor (F) de presunta importancia patogénica con una enfermedad (E).

Las formas de medir tal asociación, el efecto que sobre ella puedan ejercer otros factores y la validez que pueda conferírsele a la luz de los procedimientos seguidos para cuantificarla son elementos de interés primordial.

3.2. Clasificación de estudios epidemiológicos

Una primera clasificación de los estudios epidemiológicos según el diseño seguido es la que los divide en experimentales y observacionales.

El primer tipo se caracteriza por el hecho de que el investigador manipula previa y deliberadamente la realidad, modifica el curso natural de los acontecimientos a fin de observar lo que acaece bajo un régimen de condiciones u otro. Típicamente, ese es el caso de los ensayos clínicos controlados en que el investigador somete a un grupo

de sujetos a cierto tratamiento y al resto a un tratamiento diferente (ocasionalmente ausencia de toda acción o suministración de un placebo). Bajo esta rúbrica se inscribe también el llamado «ensayo en la comunidad».

Los estudios experimentales han merecido profunda atención por parte de los estadísticos y alcanzaron gran desarrollo teórico y práctico en el ámbito de la agronomía y la farmacología primero, y en la actividad clínica y otras luego. Sobre su uso en epidemiología puede consultarse, por ejemplo, Weddall (1979).

En el estudio de las enfermedades crónicas, sin embargo, raramente se dan condiciones propicias para el experimento; particularmente en aquellos llamados a identificar factores etiológicos, el investigador ha de ingeniárselas para sacar partido de lo que puede observar prospectiva o retrospectivamente de la realidad. De ahí que los estudios obligatoriamente reducidos a ese marco se hayan convertido en la vía fundamental de la investigación epidemiológica y hayan alcanzado impetuoso protagonismo en las últimas décadas.

Las formas que pueden asumir los estudios de observación son ciertamente numerosas.

Un primer eje de clasificación concierne al tipo de unidad de análisis (entidad sobre la que se realizan las mediciones).

Lo más usual es que la unidad de análisis sea directamente el individuo pero ocasionalmente se trabaja con grupos o conglomerados de sujetos (fábricas, ciudades, países, etc.) en lo que ha dado en llamarse estudio ecológico. (Dogan y Rokkan, 1969). Este procedimiento es usado con cierta frecuencia en los estudios de mortalidad y en menor medida en indagaciones sobre morbilidad. A modo de ilustración puede recurrirse por ejemplo a Silva y col. (1986), Stavrazy (1976).

Usualmente se trata de establecer correlaciones entre tasas de mortalidad o de incidencia de enfermedades y factores predisponentes cuya presencia se mide al nivel de las comunidades incluidas en el estudio. Sin embargo, en la transferencia de tales correlaciones al nivel de los individuos se corre el riesgo de caer en lo que se conoce como «falacia ecológica» y que ha sido motivo de no poca atención. (Ver Kalimo y Bice, 1973; Firebaugh, 1987; Morgenstern, 1982).

Por ejemplo Susser (1973) menciona varios trabajos en que se consigna una clara asociación

negativa entre la dureza del agua y las tasas de mortalidad por enfermedades cardíacas: tal asociación no se ha observado sin embargo en investigaciones realizadas a nivel individual.

Los estudios más elocuentes son sin duda los que se realizan tomando a los sujetos directamente como unidades de análisis. En términos generales los diseños posibles son tres (ver Friedman, 1980):

- estudios de cohorte
- estudios transversales
- estudios caso-control

El primero (también llamado prospectivo) consiste en identificar dos grupos de sujetos. En uno, los sujetos tienen determinado factor *F* en tanto que los del segundo están libres de él. Ambas muestras se observan durante un período dado a través de indagaciones periódicas o vigilancia sistemática a fin de observar la aparición de casos nuevos de la enfermedad *E* (o muertes por concepto de ella).

El análisis discurre típicamente a través de la estimación —y posterior comparación— de las probabilidades de aparición de enfermedad o muerte para ambos grupos; si tales magnitudes difieren, se declara la existencia de correlación entre *F* y *E*.

Esta es la forma de observación más parecida a un experimento y, por ende, la que más eficiente resulta a los efectos de sacar conclusiones de causalidad.

La dificultad para realizar el seguimiento y los lapsos de espera que muchas veces se necesitan constituyen, sin embargo, defectos de magnitud tal que suelen imposibilitar su aplicación. A ello se adiciona que —si la enfermedad *E* es por su naturaleza infrecuente— el número de sujetos necesarios para que la comparación sea útil puede ser prohibitivamente grande.

Los estudios transversales (o de prevalencia) se basan en la selección de una muestra de la población. Se indaga entonces la situación de salud y la presencia de factores de interés para cada sujeto elegido. Estos estudios reposan vitalmente sobre la eficiencia del diseño muestral y son especialmente útiles para caracterizar el estado de salud de la población estudiada y como fuente de datos para la planificación. Su utilidad, sin embargo, para la detección de nexos causales es reducida ya que no se puede controlar otro de los rasgos de que debe acompañarse la asociación de

dos variables para que las modificaciones de una sean consecuencia de los que experimenta la otra: en efecto, no puede establecerse si la precedencia temporal entre F y E en cada sujeto es la que concuerda con la hipótesis de causalidad. Por otra parte, al igual que en los estudios prospectivos si la frecuencia de E en la población es muy baja o su duración es muy breve, entonces el tamaño de la muestra necesaria puede ser enorme.

Finalmente, en su forma básica, los estudios caso-control (o retrospectivos) proceden al revés que los de cohorte. En estos se transita del efecto hacia sus presuntas causas. Vale decir: se elige un grupo de «casos» (individuos con la enfermedad E) y otro de «controles». Para uno y otros se indaga la concurrencia presente o pasada del factor o los factores de interés.

El diseño de casos y controles probablemente es, al decir de Kleinbaum y col (1982), la mayor contribución de los epidemiólogos a la metodología de la investigación. Por ser independiente de la frecuencia con que se presenten las dolencias (y por ende adecuados para el estudio de enfermedades relativamente raras) y por no demandar de largos períodos de observación, los estudios retrospectivos se han convertido en la forma dominante de la investigación analítica en epidemiología.

La metodología asociada a este tipo de estudios no se reduce a la identificación de factores etiológicos sino que puede extenderse a otras áreas tales como la evaluación de servicios de salud.

Consecuentemente con todo ello, el presente material concede preferencial atención a los estudios retrospectivos.

Una discusión detallada de las ventajas y desventajas asociadas a los estudios prospectivos y retrospectivos puede hallarse en Schlesselman (1982).

3.3. La medición en la investigación epidemiológica

Uno de los aspectos más equívocos y complejos, pero también uno de los puntos medulares de la investigación epidemiológica es el que se relaciona con la medición de las variables y procesos en juego. Cabe alertar por otra parte acerca del complejo entramado de definiciones y, sobre todo, de diferentes denominaciones que la literatura recoge para los mismos conceptos.

Definamos por *riesgo* a la probabilidad que tiene un individuo de desarrollar una enfermedad dada a lo largo de un período preestablecido.

Para la evaluación del riesgo de enfermar se han utilizado básicamente dos indicadores; a ambos se les ha llamado *tasa de incidencia* (con lo cual se ha arrojado cierta confusión adicional sobre el tema).

Para describirlos, aquí se usará la nomenclatura que probablemente está universalizándose.

La primera medida es la *densidad de incidencia* DI, definida como la razón entre el número de casos nuevos durante un período dividida por la suma de los períodos de observación correspondientes a cada sujeto desde que entran al estudio hasta la aparición de la enfermedad (si se produce) o hasta que sale por cualquier razón.

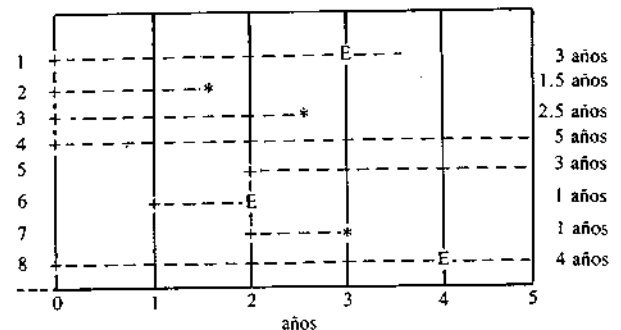
Esto significa que la DI se estima mediante:

$$p^* = \frac{e}{\sum_{i=1}^n t_i} \quad (1)$$

donde e es el número de casos nuevos en el período, n el total de individuos observados y t_i es el tiempo que transcurrió para el i-ésimo sujeto entre el momento que dejó de observarse (por haber adquirido la enfermedad, por haberse perdido del estudio, o por haber concluido éste).

Para fijar las ideas imaginemos que se realiza un estudio longitudinal que dura 5 años y que a lo largo del quinquenio se observa la situación sintetizada en el Cuadro 1.

Cuadro 1. Ejemplo gráfico de un seguimiento de 8 individuos a lo largo de 5 años.



+ comienzo de la observación.

* fin de la observación (salió del estudio o murió por una causa ajena a la enfermedad).

E aparición de la enfermedad.

Aplicando (1) a los datos del ejemplo se tendría que el tiempo total acumulado de observación de los 8 sujetos es 21 años (3 + 1.5 + 2.5 + 5 + 3 + 1 + 1 + 4) y por tanto:

$$p^* = \frac{3}{21} = 0.143$$

O sea, la densidad de incidencia es tal que de cada 1000 personas, 143 desarrollan la enfermedad cada año.

Usualmente la DI se computa por grupos de edades. El valor numérico de la DI en sí mismo no tiene una interpretación cualitativa clara y su utilidad estriba esencialmente en tanto herramienta para la comparación. Una discusión esclarecedora en este sentido puede hallarse en Elandt-Johnson (1975) y sobre todo en Rothman (1986).

La otra medida de incidencia es la llamada *tasa de incidencia acumulada* TIA, que se define simplemente como la razón del número de casos nuevos que aparecen en un lapso dado entre el número de sujetos susceptibles de desarrollar la enfermedad al comienzo del mismo.

Aunque depende del período establecido, la TIA es un parámetro sin dimensiones que se mueve entre 0 y 1 y puede interpretarse como la probabilidad de enfermar.

Formalmente si de n sujetos observados durante el período, e se enferman, la TIA se estima mediante:

$$p = \frac{e}{n} \quad (2)$$

pero si existiera cierto número de sujetos, digamos r , para los que no se puede completar el período de observación, la estimación (2) debe corregirse (ver Chiang, 1968) mediante:

$$P = \frac{e}{(n - r/2)}$$

En cierto sentido los parámetros DI y TIA procuran aprehender cuantitativamente el mismo concepto a pesar de ser estructuralmente muy diferentes.

Si DI permanece constante durante todo el período Δt de observación, puede demostrarse (ver Chiang, 1968) la siguiente relación entre la tasa

de incidencia acumulada en el período y la densidad de incidencia por persona:

$$p = 1 - \exp(-p^* \Delta t)$$

Una tercera medida básica en el estudio de la morbilidad es la *tasa de prevalencia*. Esto es la razón P del número de personas que padecen la enfermedad en un instante dado entre el número de individuos en riesgo de padecerla en ese mismo momento.

La diferencia fundamental entre la tasa de prevalencia y la de incidencia radica en que mientras la segunda concierne al número de casos nuevos en un período dado, la primera se refiere a los casos en que la enfermedad está presente, con independencia del momento en que se inició.

Freeman y Hutchinson (1980) han demostrado que:

$$P = \frac{(DI) (DM)}{1 + (DI) (DM)} \quad (3)$$

donde DM es la duración que en promedio tiene la enfermedad en cuestión.

Si la prevalencia es pequeña (menor que 0.1) entonces (3) puede simplificarse a:

$$P = (DI) (DM) \quad (4)$$

La fórmula (4) ayuda a comprender porqué en los estudios de causalidad no suele usarse la prevalencia.

Una tasa de prevalencia puede ser alta como consecuencia de que la enfermedad es en promedio muy prolongada, aun cuando la incidencia sea relativamente baja. Consecuentemente, bajo la presencia de un factor puede haber una prevalencia mucho más alta que cuando él está ausente, debiéndose ello más a las medidas aplicadas para prolongar la vida del enfermo que a la presencia del factor.

3.4. Medición de efectos

Consideremos ahora la situación en que se tiene cierto factor que puede estar o no presente en un sujeto dado (F o \bar{F} respectivamente) y que se quiere evaluar su posible efecto sobre el desarrollo de una enfermedad que a su vez puede ser o no padecida (E o \bar{E}) por el sujeto.

Lo primero que debe confirmarse en el camino de esa evaluación es la existencia de una asociación o correlación real entre ambas condiciones.

La forma más generalmente utilizada para medir el grado de asociación es el llamado *riesgo relativo* (R).

Por ejemplo, para confirmar la sospecha de que estar aquejado de hipertensión (F) puede influir en el desarrollo de una demencia senil (E) es menester corroborar previamente que, aunque un sujeto puede desarrollar esa dolencia sin ser hipertenso (y viceversa, bien puede un normotenso derivar hacia una demencia senil), el riesgo de que E se produzca bajo la condición F es mayor que la de que se produzca bajo \bar{F} . R pretende precisamente medir cuánto más probable es lo primero que lo segundo.

Su definición formal es:

$$R = \frac{P(E / F)}{P(E / \bar{F})} \quad (5)$$

donde $P(A / B)$ denota la probabilidad condicional de que se produzca el evento A dado que se produjo B.

A su vez (5) puede expresarse en términos de las tasas de incidencia respectivas. Considerando un lapso determinado, puede escribirse:

$$R = \frac{TIA_1}{TIA_0} \quad (6)$$

ó

$$R = \frac{DI_1}{DI_0} \quad (7)$$

donde los subíndices 1 y 0 indican la presencia y ausencia del factor respectivamente.

Las expresiones (6) o (7) valen indistintamente en la mayoría de las situaciones prácticas. En Rothman (1986) puede hallarse una discusión al respecto.

El riesgo relativo tiene gran utilidad como herramienta en el proceso de indagación etiológica y sus atributos en tal sentido serán considerados más adelante. Sin embargo, tiene poco interés si lo que se quiere medir es el impacto real que tiene la presencia de cierto factor en la población.

Por ejemplo supongamos que para cierta enfermedad E se tiene que las tasas de incidencia entre sujetos con y sin el factor se han estimado respectivamente así:

$$p_1 = 0.00005 \quad \text{y} \quad p_0 = 0.00001$$

El riesgo relativo estimado sería $R = 5$; quiere decir que sería 5 veces más probable padecer E si se está bajo el influjo de F que en caso contrario.

Ahora, si para otro factor F' se tuviese:

$$p'_1 = 0.5 \quad \text{y} \quad p'_0 = 0.1$$

se estimaría también que el riesgo relativo de estar bajo sus efectos es 5.

Sin embargo las complicaciones para un sujeto concreto serían muy diferentes, ya que en el segundo caso el riesgo a que está sometido un sujeto expuesto a F' es 10 000 veces mayor que si está expuesto a F. Por otra parte, desde el punto de vista poblacional un componente de primordial importancia es la proporción PF de sujetos afectados por el factor.

A fin de medir este efecto social se ha sugerido (Levin, 1953) la llamada *fracción etiológica* (FE) definida como la proporción de casos que se presentan en la población cuya existencia puede atribuirse a los efectos del factor.

Por esa vía se redimensiona el concepto de riesgo ahora en el contexto de las políticas de salud puesto que refleja la reducción que experimentaría la aparición de la enfermedad si el factor fuese eliminado (ver, por ejemplo, Kleinbaum y col., 1982).

Puede probarse que la fracción etiológica se conecta con el riesgo relativo y la proporción de afectados por F mediante la siguiente relación:

$$FE = \frac{(PF)(R-1)}{1 + (PF)(R-1)} \quad (8)$$

La Tabla I recoge el valor de la fracción etiológica para diferentes valores de PF y R.

Tabla I. Fracción etiológica en función del riesgo relativo (R) y la proporción de sujetos expuestos (PF).

Porcentaje con el factor (PF)	Riesgo Relativo (R)					
	1.5	2	3	5	10	15
.05	0.02	0.04	0.09	0.17	0.31	0.41
.10	0.05	0.09	0.17	0.29	0.47	0.58
.25	0.11	0.20	0.33	0.50	0.69	0.78
.50	0.20	0.33	0.50	0.67	0.82	0.88
.75	0.27	0.43	0.60	0.75	0.87	0.91
.90	0.31	0.47	0.64	0.78	0.89	0.93
.95	0.32	0.49	0.66	0.79	0.90	0.93

Como puede observarse, por ejemplo, si uno de cada 4 sujetos está expuesto a un factor ($PF = 0.25$) cuya presencia multiplica por 10 el riesgo de enfermar ($R = 10$) entonces, si se pudiese bloquear la aparición del factor, dejarían de producirse 69 de cada 100 enfermos.

3.5. Problemas de estimación

La manera más eficiente para estimar R y FE es a través de un estudio prospectivo.

Supongamos que se ha logrado establecer un grupo de N sujetos sanos (una muestra «representativa» de la población de sanos) y que se ha delimitado que N_1 de ellos están expuestos al factor F en tanto que los restantes N_2 no lo están.

Se observan ambos grupos durante cierto lapso Δt , conluido el cual se identifica el número de individuos que enfermaron dentro de cada subgrupo. Así, se puede confeccionar una clasificación como la de la Tabla 2.

Tabla 2. Distribución de enfermos (casos nuevos) entre expuestos y no expuestos a cierto factor.

		Enferman		
		SI (E)	NO (\bar{E})	Total
Factor	Presente (F)	A	B	N_1
	Ausente (\bar{F})	C	D	N_2
	Total	A+C	B+D	N

A partir de estos datos pueden estimarse las tasas de incidencia en cada grupo:

$$p_1 = A / N_1 \quad , \quad p_0 = D / N_2$$

y a partir de estas probabilidades condicionales, el riesgo relativo:

$$\hat{R} = p_1 / p_0$$

Con una estimación independiente de PF puede estimarse entonces FE sustituyendo ambas estimaciones en (8).

El esquema de muestreo en un estudio retrospectivo (caso-control) es totalmente diferente. Se toman dos grupos: n enfermos y n controles y se examina en cada caso el número de sujetos que estuvieron expuestos al factor de interés, de suerte que los resultados pueden resumirse como en la Tabla 3.

Tabla 3. Distribución de casos y controles según hayan estado o no expuestos a un factor.

	Casos (E)	Controles (\bar{E})	
F	a	b	a+b
\bar{F}	c	d	c+d
Total	n_1	n_2	n_1+n_2

El rasgo distintivo básico de este tipo de estudios es que aquí no se pueden estimar las tasas de incidencia separadamente.

En efecto, $a/(a+b)$ no constituye una estimación admisible de $P(E/F)$ ni $c/(c+d)$ puede aceptarse como estimación de $P(E/\bar{F})$. En este caso el número absoluto a , por ejemplo sería aproximadamente igual a $n_1 P(F/E)$ de modo que bastaría modificar n_1 y mantener igual n_2 para modificar la magnitud de a y por tanto la de $a/(a+b)$ como resultado exclusivo de la decisión del investigador.

Supongamos que $P(F/E) = 0.05$

y $P(F/\bar{E}) = 0.01$.

Si se toma $n_1 = n_2 = 1000$, se tendría aproximadamente $a = 50$, $b = 10$, $c = 950$ y $d = 990$; de modo que $a/(a+b) = 0.83$ y $c/(c+d) = 0.49$.

Si se tomara $n_1 = 3000$ y $n_2 = 1000$ la Tabla 3 se conformaría del modo siguiente: $a = 150$, $b = 10$, $c = 2850$ y $d = 990$, de modo que la presunta tasa de incidencia entre expuestos se elevaría hasta el valor $a/(a+b) = 0.94$ en tanto que dicha tasa entre no expuestos sería 0.74 .

La gran ventaja de los estudios caso-control, además de las señaladas en la sección II es que, aunque no puedan estimarse ni el numerador ni el denominador de (5), sí puede obtenerse una estimación de la razón como tal (o sea, del riesgo relativo R).

Ello unido a las consideraciones hechas en la sección II conducen a que en lo sucesivo se preste la máxima atención a dicho diseño.

Puede demostrarse que cuando la enfermedad en la población es relativamente rara (digamos $P(E) < 0.1$, como ocurre casi siempre) R puede estimarse mediante la llamada *razón de productos cruzados* («odds ratio»), que se denota por:

$$\psi = \frac{a d}{b c} \quad (9)$$

Una prueba de ello se obtiene mediante el teorema de Bayes, que establece que:

$$P(A/B) = \frac{P(A) P(B/A)}{P(A) P(B/A) + P(\bar{A}) P(B/\bar{A})} \quad (10)$$

Aplicando (10) a $P(E/F)$ y a $P(E/\bar{F})$ se tiene:

$$R = \frac{P(E/F)}{P(E/\bar{F})} = \frac{P(F/E)}{P(\bar{F}/E)} \quad (11)$$

$$\frac{P(E) P(\bar{F}/E) + P(\bar{E}) P(\bar{F}/\bar{E})}{P(E) P(F/E) + P(\bar{E}) P(F/\bar{E})}$$

Considerando $P(\bar{E}) \approx 0$ y, por ende, $P(E) \approx 1$ se tiene:

$$R \approx R' = \frac{P(F/E) / P(\bar{F}/E)}{P(\bar{F}/\bar{E}) / P(F/\bar{E})} \quad (12)$$

Lo cual obviamente se estima mediante el estadígrafo definido en (9).

Tabla 4. Valores de R y R' para valores de P(F/E) y P(F/\bar{E}) entre 0.1 y 0.5.

		P(F/E)					
		0.1	0.2	0.3	0.4	0.5	
P(F/\bar{E})	0.1	R	1.00	2.25	3.85	6.00	9.00
		R'	1.00	2.02	3.14	4.46	6.14
	0.2	R	0.44	1.00	1.17	2.67	4.00
		R'	0.47	1.00	1.61	2.36	3.35
	0.3	R	0.26	0.58	1.00	1.56	2.33
		R'	0.29	0.61	1.00	1.48	2.13
	0.4	R	0.17	0.38	0.64	1.00	1.50
		R'	0.19	0.41	0.67	1.00	1.44
	0.5	R	0.11	0.25	0.43	0.67	1.00
		R'	0.13	0.28	0.46	0.69	1.00

Para tener una idea de la calidad de la aproximación de R mediante R', en la Tabla 4 se resumen ambos valores cuando P(F/E) y P(F/\bar{E}) varían entre 0.1 y 0.5 bajo el supuesto más desfavorable dentro de la restricción para P(E), o sea, P(E) = 0.1.

Actualmente se debate acerca de diseños mediante los cuales R puede estimarse en estudios caso-control aun sin que se cumpla el requerimiento P(E) ≈ 0. (Ver Hogue y col., 1983; Greenland y col., 1986 y Hogue y col., 1986).

Ahora la fracción etiológica se estima sustituyendo $\hat{P}\bar{F}$ y PF en (B) donde PF es una estimación independiente de PF:

$$\varphi = \frac{(\hat{P}\bar{F}) (\Psi - 1)}{1 + (\hat{P}\bar{F}) (\Psi - 1)} \quad (13)$$

Naturalmente, a tal estimación de PF no siempre se tiene acceso. Una variante para resolver tal carencia fue dada por Levin y Bertell (1978) quienes sugieren estimar FE mediante:

$$\varphi' = \frac{(a d - b c)}{d n} \quad (14)$$

En rigor φ' no es otra cosa que φ cuando en lugar de $\hat{P}\bar{F}$ se pone $b/(b+d)$, o sea cuando se asume que la frecuencia del factor es igual en la población que entre los controles, lo cual es intuitivamente razonable siempre que la frecuencia de la enfermedad sea bastante pequeña.

En términos formales se tiene $P(F) \approx P(F/E)$ siempre que P(E) sea pequeño ya que $P(F) = P(F/E) P(E) + P(F/\bar{E}) P(\bar{E})$.

A modo de ilustración, supongamos que se ha hecho un estudio de 700 pacientes que tuvieron un infarto al miocardio y 1400 controles. Retrospectivamente se indaga para cada uno de los 2100 individuos si era o no fumador y se obtiene la Tabla 5 (similar a la Tabla 3).

Tabla 5. Distribución de infartados y controles respecto del hábito de fumar.

	Infartados	Controles	Total
Fumadores	586	745	1331
No fumadores	114	655	769
Total	700	1400	2100

De acuerdo a estos datos y según (9), se estimaría R como:

$$\varphi = \frac{(585) (655)}{(114) (745)} = 4.25$$

A través de (14) se tendría que la fracción etiológica asciende a:

$$\varphi' = \frac{(585) (655) - (114) (745)}{(655) (700)} = 0.65$$

En principio estos datos se interpretarían por una parte como que el hábito de fumar hace 4.5 veces más probable la aparición del infarto y, por otra, como que la supresión del tabaquismo reduciría el número de infartos a un 35 % de los que se presentan con la «ayuda» del tabaco.

3.6. Intervalos de confianza

Una vez estimado R lo natural es preguntarse si ese número es efectivamente diferente de 1. Ello equivale a plantearse una prueba para docimar la hipótesis de independencia entre la presencia de E y la de F.

Basándose en la notación de la Tabla 2, tal prueba se realiza mediante el conocido estadígrafo:

$$\chi^2 = \frac{(n_1 + n_2)(ad - bc)^2}{n_1 n_2 (a + b)(c + d)} \quad (15)$$

que se distribuye Ji-cuadrado con 1 grado de libertad.

Si χ^2 es suficientemente grande (por ejemplo mayor que 3.84, percentil correspondiente al «sacralizado» $p = 0.05$) se declara que es suficientemente diferente de 1 como para afirmar que esa diferencia no es atribuible al azar sino a que, en realidad, $R = 1$.

Por ejemplo, aplicado (15) a los datos de la Tabla 5 se tiene:

$$\chi^2 = \frac{(2100) [(586)(655) - (114)(7451)]^2}{(700)(1400)(1331)(769)} = 187.04$$

Según este resultado habría amplia evidencia muestral de que existe una asociación entre el hábito de fumar y la aparición de infarto cardíaco.

Haldane (1956) propuso el siguiente procedimiento para la estimación de un intervalo de confianza para R:

$$[\exp(L - Z_\alpha S), \exp(L + Z_\alpha S)] \quad (16)$$

donde $L = \ln \Psi$, $S = \left(\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d} \right)^{1/2}$ y Z_α es el percentil $(1 - \alpha/2)$ de la distribución normal estándar.

En el ejemplo $L = 1.51$ y $S = 0.12$. tomando $Z = 1.96$ (correspondiente a $\alpha = 0.05$), se tiene, aplicando (16), que R se halla entre 3.60 y 5.67 con confiabilidad del 95 %. O sea, el riesgo de infarto sería por lo menos 3.6 y a lo sumo 5.7 veces mayor para un fumador que para quien no lo sea.

El tratamiento de este problema para estudios prospectivos puede hallarse en Katz y col. (1978).

Para el caso de FE, Walter (1975) demostró que el estimador definido en (14) es tal que $\ln(1 - \varphi')$ se distribuye aproximadamente normal con media $\ln(1 - FE)$ y varianza igual a:

$$V = \frac{a}{cn} + \frac{b}{dn}$$

Consecuentemente, un intervalo de confianza viene dado por:

$$[1 - (1 - \varphi') \exp(z_\alpha \sqrt{V}), 1 - (1 - \varphi') \exp(-Z_\alpha \sqrt{V})] \quad (17)$$

Walter (1978) dio más tarde un método esencialmente equivalente que no involucra la función exponencial:

$$[\varphi' - Z_\alpha S', \varphi' + Z_\alpha S'] \quad (18)$$

$$\text{donde } S' = \frac{cn_2}{dn_1} \sqrt{V}$$

aplicando (17), el intervalo al 95 % para FE es (0.582, 0.707)

y mediante (18) se obtiene virtualmente lo mismo: (0.588, 0.712)

3.7. Notas sobre problemas de selección

La muestra elegida para el estudio constituye un aspecto básico para la calidad del mismo. Amplia es la gama de sesgos que pueden producirse a partir de una muestra incorrecta. Sackett (1979) se ha detenido en su clasificación y análisis. Aquí se bosquejarán algunas ideas básicas sobre el tema, parte de las cuales se han expuesto en Silva (1986).

Un *sesgo de selección* alude a una afectación de las estimaciones producto del modo en que se eligen los individuos, ya sea por deficiencias del marco del que se produce la selección como de

los grupos que se comparan. Otro ejemplo importante es la pérdida de sujetos a lo largo de un estudio prospectivo.

Los problemas más sutiles se producen en los estudios caso-control, particularmente en virtud de que a veces el modo usado para identificar a los enfermos (casos) varía en dependencia de que tengan o no el factor cuyo efecto se investiga.

Las fuentes para este problema pueden ser muy diversas (Kleinbaum y col., 1982). Como una ilustración cabe referirse al ejemplo clásico descrito cuatro décadas atrás por Berkson (1946), quien observó que un paciente que posee dos condiciones diferentes que demandan atención tiene más probabilidad de ser hospitalizado que aquel que sólo tiene una de ellas; de tal suerte, si una de dichas condiciones es evaluada en calidad de factor predisponente para la adquisición de la otra y los grupos a comparar (caso-control) se eligen de entre los pacientes hospitalizados, puede producirse una distorsión en la evaluación del riesgo relativo debido a la mencionada asociación previa que sólo se daría entre hospitalizados y no en la población general.

Robertson y col., (1978) y Conn y col., (1979) han dado ejemplos empíricos de esta falacia.

Otro punto que ha suscitado atención es el tamaño muestral. Tal y como ocurre en muestreo y en diseño experimental, para encarar este asunto deben establecerse ciertas premisas.

En este caso, supongamos que se quiere establecer el número mínimo de casos y controles (se asume que ambos tamaños son iguales) de suerte que la prueba para evaluar la hipótesis $R = 1$ conduzca a un rechazo. Es menester prefijar 4 parámetros. (Schlesselman, 1974): el nivel de significación α deseado, la potencia $1 - \beta$ que se aspira conferirle al test, la proporción P_0 de sujetos expuestos al factor entre los controles (que es esencialmente igual a P_F como ya se explicó) y un valor R^* mínimo con suficiente relevancia biológica.

Bajo tales definiciones se tiene:

$$n = 2 S(1 - S) (Z_\alpha + Z_\beta)^2 / (p_1 - p_0)$$

donde Z_γ es el percentil $1 - \gamma$ de la distribución normal.

$$p_1 = p_0 R^* / [1 + p_0 (R^* - 1)] \text{ y } S = \\ = (p_0 + p_1) / 2$$

McKeown y col. (1985) profundiza en este tema y trata el problema para el caso en que el objeto del estudio se centra en la FE en lugar de R.

Schlesselman (1982) y O'Neill (1984) ofrecen Tablas para el cálculo del tamaño de los grupos.

3.8. Un comentario sobre las relaciones de causalidad

Los procesos que discurren de la observación empírica a la generalización que se consagra en principios o leyes, transitan necesariamente por el método de inducción científica (Burlatski, 1981) y éste —si pretende ser correcto y fecundo— ha de someterse tanto a reglas generales de este tipo de razonamiento como a las que dimanen de la disciplina específica que se investiga.

Por causa de una enfermedad se entiende un evento, condición o característica que juega un papel esencial en que ésta ocurra.

Si se supera una visión estrechamente determinística, hay que apreciar la causalidad como un concepto relativo. Las leyes que establecen los patrones de causalidad no se expresan de manera siempre nítida, sino que se abren paso a través de manifestaciones particulares regidas por el azar y por la confluencia de múltiples factores que inciden simultáneamente (Silva, 1977).

La reflexión y la experiencia nos persuaden de que la causa de cualquier efecto consiste en una constelación de componentes que actúan de consumo (Rothman, 1986). Más que determinar *la* o *las* causas de un efecto dado —en este caso una enfermedad— corresponde evaluar en qué medida determinados factores pueden ser contribuyentes para que se produzca el desenlace bajo estudio.

El enfoque probabilístico para el estudio de las relaciones causa-efecto es acorde a esta manera de entender los nexos causales. El razonamiento lógico, las reglas empírica y teóricamente elaboradas y los conocimientos de la disciplina en cuestión se integran bajo ese enfoque para encarar el problema.

En el afán de sintetizar reglas generales de conducta, pueden estipularse algunas condiciones mínimas necesarias para dejar establecida una relación de causalidad entre un factor F y un efecto E.

- En primer lugar, desde luego, debe existir una forma de concomitancia o correlación entre ellas. Múltiples son los ejemplos que ilustran las falacias en que puede incurrirse al confundir la mera correlación con una asociación causa-efecto. En cualquier caso, cuanto más intensa sea tal correlación, más verosímil se torna la hipótesis causal.
- En segundo lugar debe verificarse una relación de precedencia temporal entre F y E acorde con la hipótesis en juego: o sea, F ha de ocurrir necesariamente antes que E.
- La hipótesis que establece el vínculo causal debe tener plausibilidad biológica y poseer fundamento teórico.
- La asociación observada no puede deberse por entero a la influencia de factores ajenos a la relación entre las variables analizadas.

Acerca de estos criterios, pueden consultarse muy diversas fuentes; particularmente lúcidas son las reflexiones y ejemplos expuestos por Susser (1973). Cabe detenerse aquí, sin embargo, en la última de las condiciones enumeradas.

Los grupos que se comparan no deben diferir entre sí salvo en cuanto a la condición que supuestamente sería responsable del efecto que se estudia.

El investigador procura adelantarse a identificar qué otras variables, diferentes al factor, pudieran dar lugar a diferencias en los resultados del efecto. Consecuentemente, actúa de suerte que la influencia de tales variables sea neutralizada o mantenida bajo control.

En una situación experimental, el método clásico consiste en crear grupos similares «en todos los sentidos» (usualmente mediante la asignación aleatoria) y luego someter los grupos así confeccionados a la influencia de distintos niveles del factor en estudio.

Como ya se ha dicho, en los estudios epidemiológicos tales manipulaciones no son por lo general factibles. No se puede, por ejemplo, formar al azar 2 grupos de 100 embarazadas cada uno e indicar a cada mujer del primero que fume 20 cigarrillos diarios y a las del segundo que no fume nada, a fin de evaluar los respectivos porcentajes de malformaciones congénitas en los recién nacidos.

En los estudios observacionales los grupos no se *crean* sino que sólo pueden *elegirse*. Así, pueden formarse dos grupos, uno de fumadoras y otro de no fumadoras, y observar el resultado (estudio prospectivo); análogamente, podría crearse un grupo de madres que tuvieron hijos malformados y otro con hijos sanos a fin de examinar retrospectivamente el hábito de fumar en cada grupo; sin embargo, bajo tales circunstancias no pueden controlarse otros factores que pudieran oscurecer luego el análisis.

Por ejemplo, si en el grupo de fumadoras del primer diseño la frecuencia de mujeres con edades extremas (adolescentes o mayores de 40 años) es mayor que entre las no fumadoras, un porcentaje menor de malformaciones en estas últimas podría deberse más a la composición de edades que al consumo de tabaco y, en principio, no puede evaluarse separadamente la influencia del factor de la que tiene el que no pudo controlarse.

En los estudios no experimentales, usualmente el efecto de los factores confusores o extraños se controla mediante técnicas adecuadas *en la fase de análisis*.

(NOTA: Los métodos de estratificación en los estudios prospectivos y de pareamiento en los retrospectivos son alternativas para ejercer el control en la fase de diseño. Su eficiencia es, sin embargo limitada y su implementación engorrosa. Por ello se pasan por alto en el presente material.)

3.9 Control de factores confusores

Los factores que pueden condicionar la asociación entre F y E pueden ser de muy diversa naturaleza y han sido objeto de diferentes clasificaciones. Susser (1973) propone e ilustra un amplio espectro de variantes que en ocasiones se orienta a distinguir sutiles matices.

La presente discusión se concentra en lo que se conoce en la literatura como *factores confusores* (Miettinen, 1974; Greenland y Neutra, 1980; Miettinen y Cook, 1981).

En esencia, cuando la estimación del efecto de un factor (F) se distorsiona a raíz de la mezcla con el efecto de otro factor F', se dice que este último es un factor confusor.

Esta situación ha sido identificada (Rothman, 1976) con la «paradoja» sobre la que Simpson

(1951) llamó la atención por primera vez y que pudiera ilustrarse con el siguiente ejemplo:

Una persona quiere evaluar si existe alguna asociación entre la aparición de sus dolores reumáticos y el hecho de que llueva.

En el mes de junio registra 10 días lluviosos y advierte dolores en 9 de ellos (90 %); por otra parte, sintió dolores en sólo 15 de los 20 días no lluviosos (75 %).

Repite la experiencia en julio y observa que el dolor sobrevino en 4 de 20 días lluviosos (25 %) y sólo en 1 de los 11 no lluviosos (9 %).

En cada uno de los meses se produce entonces que el dolor aparece más frecuentemente en los días lluviosos que en días que no lo son. Al considerar el bimestre completo, sin embargo, del número total de días de lluvia —treinta— en sólo 13 (43 %) se produjo el dolor en tanto que éste apareció en 17 (55 %) de los 31 días secos.

Volviendo a los factores confusores, en el contexto de los estudios de causalidad de las enfermedades, un factor F' se considera confusor de la relación entre F y E si cumple tres condiciones:

- F' es en sí mismo un factor de riesgo
- F' está correlacionado simultáneamente con F y con E .
- F' no es un paso intermedio en el tránsito causal entre F y E .

Un ejemplo sencillo es el siguiente: si se estudia la enfermedad coronaria (E) y el efecto del sedentarismo (F) sobre ella, entonces el hábito de fumar (F') sería un factor de confusión. Si en un estudio caso-control, en el grupo de casos hay una mayor frecuencia de fumadores que entre los controles (cosa verosímil ya que F' es en efecto un factor de riesgo para E), la asociación entre F y E se vería incrementada artificialmente.

El control de factores extraños es por lo tanto imprescindible para la justa evaluación de una asociación.

Consideremos el riesgo relativo estimado en los siguientes ejemplos (el primero tomado de Kleinbaum y col., 1982), donde se está usando la estructura de la Tabla 2 o de la Tabla 3 según sea el caso:

Ejemplo 1. Estudio prospectivo en que se han seleccionado 1000 sujetos expuestos a un factor y 1000 libres de él.

	E	\bar{E}		
F	200	800	1 000	$p = 200/1\ 000 = 0.20$
\bar{F}	50	950	1 000	$p = 50/1\ 000 = 0.05$
$\hat{R} = p_1 / p_0 = 4.00$				

Supongamos que la tabla se desdobra (post-estratificación) en dos partes, según se trate de sujetos con F' y sin F'

	Sujetos con F'			Sujetos con \bar{F}'		
	E	\bar{E}		E	\bar{E}	
F	194	706	900	F	6	94
\bar{F}	21	79	100	\bar{F}	29	871
$\hat{R}_1 = 1.02$			$\hat{R}_2 = 1.86$			

La drástica disminución de \hat{R} en cada uno de los estratos indica que, efectivamente, F' estaba confundiendo la verdadera relación. Por otra parte, se observa que mientras F no se relaciona con E entre los que tienen F' , factor y enfermedad sí están asociados entre los que no están expuestos a F' . Cuando ello ocurre se dice que el factor es, además, un *modificador de efecto*.

Ejemplo 2. Estudio retrospectivo en que se han seleccionado 1000 casos y 1000 controles.

	casos	controles	
F	600	250	$\Psi = ad / bc = 4.5$
\bar{F}	400	750	
	1000	1000	

Ahora, supongamos que mediante la post-estratificación se tiene:

	Sujetos con F'		Sujetos con \bar{F}'	
	E	\bar{E}	E	\bar{E}
F	400	20	F	200
\bar{F}	350	170	\bar{F}	50
	750	190		250
$\Psi_1 = 9.71$		$\Psi_2 = 10.08$		

F' es un factor de confusión que debilitaba la verdadera asociación (no modificador de efectos).

Con los mismos datos, supongamos que se controla otro factor F'' y que se obtiene:

	Sujetos con F''		Sujetos con \bar{F}''	
	E	\bar{E}	E	\bar{E}
F	500	100	F	100
\bar{F}	100	200	\bar{F}	300
	600	300		400
$\Psi_1 = 10.00$		$\Psi_2 = 1.22$		

En este ejemplo, por último, se observa que F'' posee un fuerte carácter modificador de efecto ya que la diferencia entre Ψ_1 y Ψ_2 es notable pero no juega un papel confusor ya que no estaba ni atenuando ni magnificado el riesgo relativo cuando no estaba controlado.

Cabe preguntarse ahora: cómo tener un estimador único de R luego de haberse efectuado la post-estratificación. Supongamos que se han creado k estratos y se tienen respectivas tablas 2×2 .

Para los estudios casos-control, Mantel y Haenszel (1959) propusieron una media ponderada de los Ψ_i ($i:1, \dots, k$) que se puede expresar del modo siguiente:

$$\Psi_{MH} = \frac{\sum_{i=1}^k \frac{a_i d_i}{n_i}}{\sum_{i=1}^k \frac{b_i c_i}{n_i}} \quad (18)$$

donde $n_i = a_i + b_i + c_i + d_i$ y donde $a = \sum_{i=1}^k a_i$, $b = \sum_{i=1}^k b_i$, etc.

Hauck (1979) ha dado una fórmula para calcular intervalos de confianza para R en tal caso:

$$[\exp(-Z_\alpha S_{MH}), \exp(Z_\alpha S_{MH})] \quad (19)$$

donde $S_{MH} = \frac{\sum_{i=1}^k w_i r_i}{(\sum_{i=1}^k w_i)^2}$

$$w_i = b_i c_i / n_i, r_i = \frac{a_i + c_i}{a_i c_i} + \frac{b_i + d_i}{b_i d_i}$$

Para los estudios prospectivos Tarone (1981) propuso la siguiente estimación puntual:

$$\hat{R}_T = \frac{\sum_{i=1}^k W_i R_i}{\sum_{i=1}^k W_i} \quad (20)$$

donde $W_i = \frac{N_{ii} C_i}{B_i + D_i}$.

y el intervalo:

$$[\hat{R}_T - Z_\alpha S, \hat{R}_T + Z_\alpha S] \quad (21)$$

donde

$$S_T^2 = \frac{\sum_{i=1}^k W_i^2 \hat{R}_i^2 \left(\frac{B_i}{A_i N_{ii}} + \frac{D_i}{C_i N_{2i}} \right)}{\left(\sum_{i=1}^k W_i \right)^2}$$

La evaluación de la fracción etiológica (FE) teniendo en cuenta el efecto confusor de una tercera variable ha sido abordado en Ejigou (1979).

3.10. La regresión logística

Consideremos la función f definida en el eje real:

$$f(Z) = [1 + \exp(-Z)]^{-1}$$

De su definición se desprende que f , conocida como la función logística, es creciente y cumple la condición $0 < f(Z) < 1$.

Esta propiedad ha sugerido a los epidemiólogos considerar el riesgo de desarrollar una enfermedad (o, eventualmente de morir) durante cierto período como una función de variables independientes X_1, X_2, \dots, X_k supuestamente influyentes en ese resultado.

En este sentido, se ha planteado la posibilidad de evaluar la probabilidad condicional de que se produzca la enfermedad dado que un sujeto sano exhibe el perfil $X_1 = x_1, \dots, X_k = x_k$ a través del siguiente modelo:

$$\begin{aligned} P(x) &= P\{Y = 1 / x_1, \dots, x_k\} = \\ &= \{1 + \exp(-A - \sum_{i=1}^k B_i x_i)\}^{-1} \end{aligned} \quad (22)$$

donde Y es la variable dependiente o de respuesta definida como $Y = 1$ si se produce E y como $Y = 0$ en otro caso.

La peculiaridad básica de la función logística proviene del hecho de que cumple la propiedad siguiente:

$$\ln \left(\frac{f(Z)}{1 - f(Z)} \right) = Z$$

Aplicado a nuestro modelo, ello significa que:

$$\ln \frac{P(x)}{Q(x)} = A + \sum_{i=1}^k B_i X_i$$

donde $Q(x) = 1 - P(x)$.

O sea, aplicada la transformación logit a la probabilidad, se obtiene una función lineal de las variables predictoras.

La aplicación más fecunda de la regresión logística ha sido en el análisis de estudios retros-

pectivos. En buena medida, ello se puede atribuir a la siguiente propiedad. Si los valores x para un sujeto dado son $x = (x_1, x_2, \dots, x_k)$ mientras que para otros son $x' = (x'_1, \dots, x'_k)$ se puede definir:

$$L(x, x') = \frac{P(x') Q(x)}{Q(x') P(x)} \quad (23)$$

que no es otra cosa que la razón de productos cruzados que puede interpretarse como cuántas veces más probable resulta enfermar bajo las condiciones x' que bajo x .

Si se desarrolla (21) se comprueba que:

$$L(x, x') = \exp \left[\sum_{i=1}^k B_i (x'_i - x_i) \right] \quad (24)$$

lo cual abre interesantes interpretaciones de los coeficientes de regresión logística.

En efecto, si suponemos que $x'_i = x_i$ para cierto i , el sumando $B_i(x'_i - x_i)$ se anula, de modo que la razón $L(x, x')$ sólo depende de las variables predictoras en que el sujeto difiere.

Por lo tanto, si dos sujetos son iguales en todos los sentidos salvo en lo que respecta a cierta variable X_j , se tendría:

$$L(x, x') = \exp [B_j (x'_j - x_j)]$$

Si X_j representa a un factor de riesgo que vale $X_j = 1$ si ese factor está presente y $X_j = 0$ en caso contrario, entonces la razón de productos cruzados se reduciría a $\exp (B_j)$ cuando se compara un sujeto con el factor con uno que no lo tiene.

En otras palabras, de este notable resultado se desprende que a través del modelo logístico *puede estimarse el riesgo relativo de poseer cierto factor supuesto que todas las demás variables en el modelo se han controlado*; además, esa estimación sería en extremo simple: $\exp (\hat{B}_j)$ donde \hat{B}_j sería una estimación del coeficiente B_j .

El problema de estimar los parámetros A, B_1, \dots, B_k fue resuelto por Walker y Duncan (1967). Para ello, naturalmente, hace falta contar con una matriz de datos:

Y_1	X_{11}	X_{1k}
Y_2	X_{21}	X_{2k}
.	.		.
.	.		.
Y_n	X_{n1}	X_{nk}

donde Y es 1 ó 0, dependiendo de que el sujeto i -ésimo tenga o no respectivamente la dolencia (caso o control, si se tratase de un estudio de ese tipo) en tanto que X_{ij} es el valor de X_j para dicho sujeto, la cual puede ser una variable de cualquier tipo (dicotómica, discreta o continua).

En la actualidad puede contarse con software adecuado para la rápida solución de este problema. Por ejemplo, Mac Gee (1986) ha ofrecido recientemente el programa LOGRESS que, además de resolver el problema de estimación calcula intervalos de confianza para los coeficientes así como las estimaciones de los riesgos relativos asociados al caso ($x'_j - x_j$) = 1 y sus intervalos para cada j ($j:1, \dots, k$).

Otro programa aplicable es el de Hosmer y col. (1978) o el contenido en el paquete BMDP. Finalmente, el autor del presente material posee un programa confeccionado en BASIC por Richard Jones en 1986 que puede obtenerse en discos flexibles para su explotación en prácticamente cualquier PC.

Cabe advertir que cuando se ha hecho un estudio caso control, el resultado de la aplicación de (20) a un caso cualquiera de la población no estima la probabilidad de que dicho sujeto desarrolle la enfermedad sintetizada en la expresión $Y = 1$. Ciertamente, hay que reparar en que la proporción de casos y controles en la muestra normalmente no se corresponde con la de la población general.

Si P_1 es la fracción de muestreo n_1 / N_1 y P_2 la fracción n_2 / N_2 donde N_1 y N_2 son los tamaños poblacionales de enfermos y sanos respectivamente, entonces para estimar la probabilidad arriba mencionada, habría que calcular (Mantel, 1973):

$$[1 + \exp \{b - A - Bx + \ln \hat{\theta}\}]^{-1}$$

donde $\hat{\theta}$ es una estimación de P_1 / P_2 . Desafortunadamente, lo natural es no contar con tal estimación.

Un ejemplo de aplicación de la regresión logística en un estudio caso-control (tomado de Mac Gee, 1986) es el siguiente. Se considera *caso* a un sujeto que murió ($y = 1$) por concepto de una enfermedad cardíaca en el marco de cierto período y *control* a un sobreviviente ($y = 0$). Las variables predictoras consideradas son:

- X_1 = edad
- X_2 = hábito de fumar (0 si no fuma, 1 si fuma)
- X_3 = tensión arterial sistólica

Trabajando con n igual a 191 casos y n igual a 226 controles, se obtienen las siguientes estimaciones:

$$\begin{aligned}\hat{A} &= -6.614 & \hat{B}_2 &= 0.312 \\ \hat{B}_1 &= 0.075 & \hat{B}_3 &= 0.018\end{aligned}$$

Entonces se puede estimar el riesgo relativo de morir dado que se trataba de un fumador respecto de un no fumador, eliminando el efecto de la edad y de la tensión arterial (es decir, controlando estas dos fuentes de confusión a la vez) mediante:

$$\exp(0.312) = 1.37$$

Otras preguntas pueden ser similarmente respondidas. Por ejemplo, ¿cuánto más riesgo tiene un sujeto fumador de 60 años que uno no fumador de 35? Aplicando (22) se tendría:

$$\begin{aligned}\exp[0.075(60-35) + 0.312(1-0)] &= \\ = \exp(2.187) &= 8.908\end{aligned}$$

Es decir, el riesgo de muerte del primero es 9 veces mayor que el del segundo.

Un punto que despierta natural interés es el concerniente a la selección de modelos. Si inicialmente se incluyeron k variables, es posible que no todas ellas sean relevantes al problema. Existen varios procedimientos para eliminar las

variables superfluas. El más usado es el de la «regresión paso a paso», que a grandes rasgos consiste en ir eliminando variables del modelo y comparando los resultados de las sucesivas versiones. La comparación se realiza a través de la prueba conocida como «razón de verosimilitud». En Mac Gee y col. (1984) se explica e ilustra detalladamente tanto este procedimiento como otras alternativas. Se enfatiza además la falta de concordancia que puede resultar al aplicar diferentes algoritmos de selección a los mismos datos.

La regresión logística se ha sugerido además como herramienta con finalidad diagnóstica en el sentido predictivo asociado a todo modelo de regresión. Lee (1975) ha propuesto este método para contribuir a la evaluación de ensayos clínicos controlados, con lo cual ha incorporado una ingeniosa aplicación al creciente uso de dicho modelo.

Bruzzi y col. (1985) han planteado una solución al problema de la estimación de la FE atribuible a varios factores simultáneamente, enfatizando los beneficios que se desprenden de usar la regresión logística de suerte que no sea menester calcular los riesgos separadamente en muchos estratos, procedimiento a menudo poco factible por razón de tamaño muestral insuficiente.

BIBLIOGRAFIA

- BERKSON, J. (1946). *Limitations of the application of fourfold table analysis to hospital data*. Biometrics Bulletin 2:47-53.
- BRESLOW, N.E. and DAY, N.E. (1980). *Statistical methods in cancer research*. IARC Scientific Publications N.º 32, Lyon.
- BRUZZI, P.S.B. and col. (1985). *Estimating the population attributable risk for multiple risk factors using case-control data*. American Journal of Epidemiology 122:904-914.
- BURLASKI, F. (1981). *Materialismo dialéctico*. Editorial Progreso. Moscú.
- CHIAN, C.L. (1980). *Introduction to stochastic processes in biostatistics*. Wiley, N.Y.
- COLE, P. and MAC. MAHON, B. (1971). *Attributable risk percent in case-control studies*. British Journal of Preventive and Social Medicine 25:242-244.
- COLE, P. (1979). *The evolving case-control study*. Journal of Chronic Diseases 32:15-27.
- CONN, H.O. y col. (1979). *The Berkson bias in action*. Yale Journal of Biological Medicine 2:241-147.
- DOGAN, M. and ROKKAN, S. (1969). *social ecology*. MIT Press, Massachusstes.
- EJIGOU, A. (1979). *Estimation of attributable risk in the presence of confounding*. Biomedical Journal 21:155-165.
- ELANDT-JOHSON, R.C. (1975). *Definition of rates: some remarks on their use and misuse*. American Journal of epidemiology 102:267-271.
- FIREBAUGH, B. (1978). *A rule for inferring individual-level relationships from aggregated data*. American Sociology Review 43:557-572.
- FREEMAN, J. and HUTCHINSON, G.B. (1980). *Prevalence, incidence and duration*. American Journal of Epidemiology 112:707-723.
- FRIEDMAN, G.D. (1981). *Primer of epidemiology*. 2nd. Ed. Mac Graw-Hill, N.Y.
- GREENLAND, S. and NEUTRA, R. (1980). *Control of confounding in the assessment of medical technology*. International Journal of Epidemiology 9:361-37.
- GREENLAND, S. y col. (1986). *The rare-disease assumption revisited*. American Journal of Epidemiology 124:869-876.
- HALDANE, J.B.S. (1956). *The estimation and significance of the logarithm of a ratio frequencies*. Annals of Human Genetics 20:309-311.
- HAUCK, W.W. (1979). *The large sample variance of the Mantel-Haenszel estimator of a common odds ratio*. Biometrics 35:817-819.
- HOGUE, C.J.R. y col. (1983). *Estimators of relative risk for case-control studies*. American Journal of Epidemiology 118:396-407.
- HOGUE, C.J.R. y col. (1986). *The case-exposure study. A further explication and response to a critique*. American Journal of Epidemiology 124:877-883.
- HOSMER, D.W. and col. (1978). *A computer program for step-wise logistic regression using maximum likelihood estimation*. Computer Programs in Biomedicine 8:121:134.

- KALIMO, E. and BICE, T.W. (1973). *Causal analysis and ecological fallacy in cross-national epidemiological research*. Scandinavian Journal of social Medicine 1:17-24.
- KATZ, D. (1978). *Obtaining confidence intervals for the risk ratio in cohort studies*. Biometrics 34:469-474.
- KLEINBAUM, D.G. y col. (1982). *Epidemiologic research*. Lifetime Learning Pub., Belmont.
- LEE, J. (1981). *Covariance adjustment of rates bases on the multiple logistic regression model*. Journal of Chronic Diseases 34:415-426.
- LEVIN, M.L. (1953). *The occurrence of lung cancer*. Acta Union International Contra Cancrum 9:531-541.
- LEVIN, M.L. and BERTELL, R. (1978). *Re: simple estimation of population attributable risk from case-control studies*. American Journal of Epidemiology 108:78-79.
- LILIENFELD, A.M. and LILIENFELD, D.E. (1983). *Fundamentos de epidemiología*. Fondo Educativo Interamericano. México DF.
- MAC GEE, D.L. and col. (1984). *The results of logistic analyses when the variables are highly correlated*. Journal of Chronic Diseases 37:713-719.
- MAC GEE, D.L. (1986). *A program for logistic regression on the IBM PC*. American Journal of Epidemiology 124:702-705.
- MANTEL, N. and HAENSZEL, W. (1959). *Statistical aspects of the analysis of data from retrospective studies disease*. Journal of the National Cancer Institute 22:719-784.
- MIETTINEN, O.S. (1974). *Confounding and effect-modification*. American Journal of Epidemiology 100:350-353.
- MIETTINEN, O.S. and COOK, E.F. (1981). *Confounding: essence and detection*. American Journal of Epidemiology 114:609-620.
- MORGENSTERN, H. (1982). *Uses of ecologic analysis in epidemiologic research*. American Journal of Public Health 72:1336-1344.
- O'NEIL, R.T. (1984). *Sample sizes for estimation of the odds ratio in unmatched case-control studies*. American Journal of Epidemiology 120:145-153.
- ROBERTS, R.S. y col. (1978). *An empirical demonstration of Berkson's bias*. Journal of Chronical Diseases 31:119-128.
- ROTHMAN, K.J. (1986). *Modern epidemiology*. Little Brown and Co., Boston.
- SACKET, D.L. (1979). *Bias in analytic research*. Journal of Chronic Diseases 32:51-68.
- SCHLESSELMAN, J.J. (1974). *Sample size requirements in cohort and case-control studies of diseses*. American Journal of Epidemiology 99:381-384.
- SCHLESSELMAN, J.J. (1982). *Case-control studies*. Oxford University Press, N.Y.
- SILVA, L.C. (1977). *El razonamiento en las aplicaciones de la estadística*. Revista Cubana de Administración de Salud 3:275-279.
- SILVA, L.C. y col. (1986). *Evaluación de la mortalidad infantil según condiciones higiénico-sociales del municipio*. Revista Cubana de Administración de Salud 11:200-210.
- SILVA, L.C. (1986). *Sobre el concepto de representatividad y los métodos de selección muestral*. Revista Cubana de Investigaciones Biomédicas 5:363-369.
- SIMPSON, E.H. (1951). *The interpretation of interaction in contingency tables*. Journal of the Royal Statistical Society, Series B 13:238-241.
- STAVRAKY, K.M. (1976). *The role of ecologic analysis in studies of the etiology of disease: a discussion with reference to large bowel cancer*. Journal of Chronic Diseases 29:435-444.
- SUSSER, M. (1973). *Causal thinking inthe health science*. Oxford University Press, N.Y.
- TARONE, R.E. (1981). *On sumary estimators of relative risks*. Journal of Chronic Diseases 34:463-468.
- WALKER, S. and DUNCAN, D. (1967). *Estimation of the probability of an event as a function of several variables*. Biometrika 54:167-179.
- WALTER, S.D. (1975). *The distribution of Levin's measure of attributable risk*. Biometrika 62:371-374.
- WALTER, S.D. (1978). *Calculation of attributable risk from epidemiological data*. International Journal of Epidemiology 7:175-182.
- WEDDELL, J. (1979). *Experimental studies*. En «Health care and epidemiology», editado por W. Holland y L. Karhansen, Cap. 14. Hall, Boston.