

MÉTODOS BIOESTADÍSTICOS EN INVESTIGACIONES CIENTÍFICAS: MODELOS LOGÍSTICOS PROSPECTIVOS EN ESTUDIOS RETROSPECTIVOS*

Lic. Suzanne Loo de Lao

De la División de Epidemiología del Laboratorio Conmemorativo Gorgas

Regresiones logísticas han sido utilizadas frecuentemente para analizar data retrospectiva. Existen dos métodos diferentes: el modelo logístico "prospectivo" considera la enfermedad como la variable dependiente, mientras que el modelo logístico "retrospectivo" considera el factor de riesgo como la variable dependiente. Teóricamente, el modelo retrospectivo es el correcto al considerar el esquema de muestreo utilizado en estudios caso-control. El modelo retrospectivo se complica, sin embargo, cuando el factor de riesgo tiene más de dos valores o cuando hay que tomar en consideración más de un factor de riesgo. Matemáticamente, ambos modelos resultan en la misma desigualdad relativa con respecto a los parámetros β y α : $D.R. = e^{\beta} + \alpha' X$. Por consiguiente, el modelo prospectivo puede ser usado para estimar la desigualdad relativa en datos caso-control siempre que haya suficiente ajuste de covarianza.

Científicos, médicos y estudiantes están reconociendo el hecho de que en Panamá hay muchas posibilidades para la investigación científica en el campo biológico, clínico, epidemiológico, biomédico, social y otros. Al aumentar las investigaciones científicas se necesitarán métodos bioestadísticos para analizar los datos recogidos, ya sea en forma descriptiva o comprobando hipótesis, usando el modelo estadístico adecuado. La rápida evolución de las computadoras en los últimos años, en especial de las microcomputadoras, ha puesto al alcance de la investigación programas de computadoras o "software" para el manejo y el análisis estadístico de los datos. El investigador tiene a su alcance programas bastante flexibles y relativamente fáciles de usar, que le permitirán manejar y analizar sus propios datos. Existe entre otros, el programa "DBase III Plus" para el manejo de grandes cantidades de datos.

* Presentado para publicación en junio de 1987.

Para los análisis bioestadísticos existen programas como el "Systat," que es relativamente fácil de usar por cualquiera persona que tenga algún conocimiento de bioestadística. El Laboratorio Conmemorativo Gorgas tiene el programa "BMDP Statistical Software" para análisis más avanzados el cual consiste en programas de regresión logística y en tablas de sobrevivencia. Este "software" fue diseñado originalmente en la Universidad de California de los Estados Unidos, para computadoras grandes ("mainframes"). Debido al auge reciente de las microcomputadoras, fue adaptado a las computadoras personales.

La bioestadística es un campo que debemos desarrollar más en Panamá debido a las facilidades que tenemos a nuestro alcance. A continuación presentamos un tema relativamente nuevo en el campo de la epidemiología para el análisis del estudio caso-control.

A. Estudios retrospectivos y estudios prospectivos

Al estudiar las causas de una enfermedad, es preciso obtener la información descriptiva para identificar los grupos con altas o con bajas tasas de la enfermedad. El próximo paso es el de tratar de explicar por qué la tasa es alta o es baja en un grupo particular. Esto conduce a la formulación de hipótesis que pue-

den ser probadas a través de más estudios epidemiológicos, los cuales pueden ser de dos tipos: estudio de cohorte o estudio de caso-control.

En un estudio prospectivo, llamado también de cohorte, se le hace un seguimiento a una muestra de personas (llamado cohorte) que inicialmente no tenía la enfermedad en estudio, pero que varía en la exposición al factor de riesgo. Después de un tiempo predeterminado, se estudian las diferencias en la tasa de desarrollo de la enfermedad en relación con la exposición al factor de riesgo.

En un estudio retrospectivo, llamado más comúnmente caso-control, las personas que tienen una enfermedad (casos) son comparadas con aquéllas que no tienen la enfermedad (controles). Se determina retrospectivamente si los dos grupos difieren en el número de personas expuestas al factor de riesgo.

En el estudio prospectivo de una enfermedad rara o poco común debe seguirse a una gran cantidad de personas durante un largo período, para obtener un número limitado de casos, y por esa causa muchos estudios dependen de la investigación retrospectiva, porque es más económica. Las muestras pueden ser más pequeñas, debido a que los casos son identificados al comienzo del estudio retrospectivo;

y muchas veces son comparadas con una cantidad igual de controles. El estudio de caso-control ha sido frecuentemente utilizado como el primer paso en la búsqueda de los posibles factores, y cuando se sabe muy poco de las causas de la enfermedad (1).

En un estudio prospectivo se calcula el riesgo relativo de la enfermedad. El riesgo relativo es la incidencia de la enfermedad en los individuos expuestos al factor de riesgo en comparación con la incidencia en los individuos no expuestos. En un estudio retrospectivo el riesgo relativo no se puede calcular; pero sí puede estimarse la desigualdad relativa. En el caso de una enfermedad poco común la desigualdad relativa es un buen aproximado al riesgo relativo.

La desigualdad relativa (D.R.) se calcula de la siguiente manera:

para describir cómo un conjunto de k variables, por ejemplo, X_1, X_2, \dots, X_k , se relaciona en forma logística a una variable dicótoma Y . Hay dos métodos diferentes y a la vez controversiales que han surgido en el modelo de regresión logística para el análisis de estudios retrospectivos (2). El modelo logístico prospectivo considera la enfermedad como la variable dependiente. En el modelo retrospectivo el factor de riesgo es la variable resultante y la enfermedad es la variable regresiva. Teóricamente el segundo método es el correcto, al considerar el esquema utilizado en estudios retrospectivos. Aquí, el status de enfermedad es una variable fija y los factores de riesgo son considerados como variables aleatorias condicionadas al status de la enfermedad.

Ciertas complicaciones pueden surgir con el método retrospec-

ENFERMEDAD EN ESTUDIO

		PRESENTE	AUSENTE	TOTAL
FACTOR DE RIESGO	EXPUESTO	a	b	a+b
	NO EXPUESTO	c	d	c+d
TOTAL		a+c	b+d	

$$D.R. = \frac{ad}{bc}$$

B. El modelo logístico en estudios retrospectivos

El modelo logístico ha sido utilizado muchas veces en el análisis de una data retrospectiva

y sería preferible en esos casos usar el método prospectivo. Considérese la situación en donde el factor de riesgo tiene más de dos niveles. Si los facto-

res de exposición tienen muchos valores, estos valores deben ser agrupados en un número pequeño de categorías discretas (preferiblemente dos). Otra ventaja del modelo prospectivo se presenta cuando se considera varios factores de riesgo. ¿Cuál de éstos sería la variable resultante en el modelo retrospectivo? Si todos los factores de riesgo son considerados equitativos, podrían ser agrupados en valores discretos para representar combinaciones de exposición a varios factores. De otra manera, también podría utilizarse un modelo retrospectivo para cada factor de riesgo. Ambas alternativas son más complicadas que el modelo prospectivo.

Los modelos logísticos retrospectivos y prospectivos se interpretan de diferentes maneras, contienen diferentes parámetros y pueden dar diferentes resultados. El objetivo de este informe es el de comparar estos dos modelos e indicar dónde y cuándo pueden dar los mismos resultados con los mismos datos, justificando de esta manera la utilización del método más sencillo: el prospectivo.

LA DESIGUALDAD RELATIVA EN LOS MODELOS LOGISTICOS RETROSPECTIVOS Y PROSPECTIVOS

El parámetro clave en estudios retrospectivos es la desigualdad relativa en el cuadro de casos

o controles y expuestos o no al factor de riesgo. Si los casos escogidos fueran representativos de todos los casos y los controles seleccionados fueran representativos de la población estudiada, Cornfield demostró en 1951 que la desigualdad relativa en las enfermedades de baja incidencia es un buen aproximado al riesgo relativo (3). El riesgo relativo es la razón de la tasa de incidencia de aquellos expuestos al factor sobre la tasa de incidencia de aquellos no expuestos.

Asumamos que F es una variable dicótoma con $F=1$ (expuesta al factor) y $F=0$ (no expuesta), y E una variable indicadora de enfermedad con $E=1$ (caso) y $E=0$ (control). A continuación consideraremos un estudio con covarianza.

En el método retrospectivo la variable de exposición es la variable resultante del modelo logístico y la enfermedad es la variable regresiva. Como mencionamos anteriormente, ésta es la manera correcta para analizar datos de caso-control. El método retrospectivo para el modelo logístico regresivo sin interacción es descrita por la siguiente ecuación:

$$P(F=1 : E, z) = \frac{e^{\alpha + \beta E + \gamma z}}{1 + e^{\alpha + \beta E + \gamma z}}$$

donde α es un parámetro escalar desconocido, β es un coeficiente de E y z representa el efecto de la covarianza en F , con γ como su coeficiente.

El método prospectivo es ligeramente diferente ya que se invierten las funciones de la enfermedad y del factor de riesgo. La variable binaria de enfermedad es la variable resultante y el factor de exposición es la variable regresiva. Este modelo es descrito por la función:

$$P(E=1 : F, z) = \frac{e^{\alpha^* + \beta^* F + \gamma^* z^*}}{1 + e^{\alpha^* + \beta^* F + \gamma^* z^*}}$$

donde α^* es un parámetro desconocido, β^* es un coeficiente desconocido de la variable regresiva F y z^* representa el efecto de la covarianza en E , con γ^* como su coeficiente.

El método a seguir para calcular la desigualdad relativa en los modelos retrospectivos y prospectivos sería el siguiente. Asumamos que z y z^* tengan el mismo valor z_0 . Los cuadros siguientes ilustran el procedimiento a seguir para cada método.

Nótese que los parámetros β y β^* tienen los mismos significados en ambos modelos. La desigualdad relativa de las probabilidades de exposición resulta ser idéntica a la desigualdad relativa de las probabilidades de la enfermedad.

Ahora, si se incluyeran los términos de interacción, entonces el modelo logístico retrospectivo sería descrito como:

$$P(F=1 : E, z) = \frac{e^{\alpha + \beta E + \gamma z + \sigma(Ez)}}{1 + e^{\alpha + \beta E + \gamma z + \sigma(Ez)}}$$

donde σ es un coeficiente desconocido para el término de interacción entre el efecto de la covarianza y la enfermedad. De la misma manera el modelo logístico regresivo prospectivo seguiría la siguiente función:

$$P(E=1 : F, z) = \frac{e^{\alpha^* + \beta^* F + \gamma^* z^* + \sigma^*(Fz^*)}}{1 + e^{\alpha^* + \beta^* F + \gamma^* z^* + \sigma^*(Fz^*)}}$$

donde σ^* es un coeficiente desconocido para el término de in-

METODO RETROSPECTIVO			METODO PROSPECTIVO		
		F			F
		1			0
E	O	$e^{\alpha + \gamma z_0}$	1	1	1
		$1 + e^{\alpha + \gamma z_0}$	$1 + e^{\alpha + \gamma z_0}$		$1 + e^{\alpha^* + \gamma^* z_0}$
	E	$e^{\alpha + \beta + \gamma z_0}$	1		$e^{\alpha^* + \beta^* + \gamma^* z_0}$
	I	$1 + e^{\alpha + \beta + \gamma z_0}$	$1 + e^{\alpha + \beta + \gamma z_0}$		$1 + e^{\alpha^* + \beta^* + \gamma^* z_0}$

La desigualdad relativa (D.R.) se define en las siguientes ecuaciones:

METODO RETROSPECTIVO	METODO PROSPECTIVO
$D.R. = \frac{P(F=1 : E=1) / P(F=0 : E=1)}{P(F=1 : E=0) / P(F=0 : E=0)}$	$D.R. = \frac{P(E=1 : F=1) / P(E=0 : F=1)}{P(E=1 : F=0) / P(E=0 : F=0)}$
$D.R. = \frac{\left(\frac{e^{\alpha + \beta + \gamma z_0}}{1 + e^{\alpha + \beta + \gamma z_0}} \right) \left(\frac{1}{1 + e^{\alpha + \gamma z_0}} \right)}{\left(\frac{e^{\alpha + \gamma z_0}}{1 + e^{\alpha + \gamma z_0}} \right) \left(\frac{1}{1 + e^{\alpha + \gamma z_0}} \right)}$	$D.R. = \frac{\left(\frac{e^{\alpha^* + \beta^* + \gamma^* z_0}}{1 + e^{\alpha^* + \beta^* + \gamma^* z_0}} \right) \left(\frac{1}{1 + e^{\alpha^* + \gamma^* z_0}} \right)}{\left(\frac{e^{\alpha^* + \gamma^* z_0}}{1 + e^{\alpha^* + \gamma^* z_0}} \right) \left(\frac{1}{1 + e^{\alpha^* + \gamma^* z_0}} \right)}$
$\beta = \ln D.R.$	$\beta^* = \ln D.R.$
$\beta - \beta^* = \ln D.R.$	

teracción entre el efecto de la covarianza y el factor. Nuevamente, al comparar los dos métodos tenemos lo siguiente:

Los coeficientes β y σ se refieren a lo mismo en ambos modelos. Esto demuestra que la desigualdad relativa es la misma

METODO RETROSPECTIVO		METODO PROSPECTIVO	
Asumir $z=O$ (ausencia de efecto):		Asumir $z^*=O$ (ausencia de efecto):	
	F		F
	1 0		1 0
O	e^α	1	1
E	$1 + e^\alpha$	$1 + e^\alpha$	$1 + e^{\alpha^*}$
	$e^\alpha + \beta$	1	$e^{\alpha^*} + \beta^*$
1	$1 + e^\alpha + \beta$	$1 + e^\alpha + \beta$	$1 + e^{\alpha^*} + \beta^*$

Para una covarianza dada, las desigualdad relativa se define como:

en términos de los parámetros β (β^*) y σ (σ^*). Nótese que, estén o no estén fijos los totales

METODO RETROSPECTIVO	METODO PROSPECTIVO
$\left(\frac{e^{\alpha+\beta}}{1+e^{\alpha+\beta}} \right) / \left(\frac{1}{1+e^{\alpha+\beta}} \right)$	$\left(\frac{e^{\alpha^*+\beta^*}}{1+e^{\alpha^*+\beta^*}} \right) / \left(\frac{1}{1+e^{\alpha^*+\beta^*}} \right)$
D.R. = $\frac{\left(\frac{e^{\alpha+\beta}}{1+e^{\alpha+\beta}} \right) / \left(\frac{1}{1+e^{\alpha+\beta}} \right)}{\left(\frac{e^\alpha}{1+e^\alpha} \right) / \left(\frac{1}{1+e^\alpha} \right)} = e^\beta$	D.R. = $\frac{\left(\frac{e^{\alpha^*+\beta^*}}{1+e^{\alpha^*+\beta^*}} \right) / \left(\frac{1}{1+e^{\alpha^*+\beta^*}} \right)}{\left(\frac{e^{\alpha^*}}{1+e^{\alpha^*}} \right) / \left(\frac{1}{1+e^{\alpha^*}} \right)} = e^{\beta^*}$
$\left(\frac{e^\alpha}{1+e^\alpha} \right) / \left(\frac{1}{1+e^\alpha} \right)$	$\left(\frac{e^{\alpha^*}}{1+e^{\alpha^*}} \right) / \left(\frac{1}{1+e^{\alpha^*}} \right)$
$\beta = \beta^* = \text{logaritmo natural de D.R. donde } z \text{ y } z^* = O.$	

Nótese que β y β^* representan los mismos en ambos modelos para la desigualdad relativa.

Para el caso de z y z^* igual a 1 (el efecto está presente) la desigualdad relativa se define (véase la página siguiente).

Por lo general, para cualquier $z = z_0$ y $z^* = z_0$, el logaritmo natural de D.R. sería: $\ln \text{ D.R.} = \beta + \sigma z_0 = \beta^* + \sigma^* z_0$.

marginales del factor de riesgo o de la variable de enfermedad, la desigualdad relativa no cambia.

Los resultados presentados anteriormente se aplican de igual manera a un caso más general en donde existen más de una covarianza. En este caso \underline{z} y \underline{z}^* son p-vectores que representan los efectos de p covarianzas y $\underline{\gamma}$, $\underline{\gamma}^*$ y $\underline{\sigma}$ son vectores de coeficientes. El modelo logístico

	1	0
O	$\frac{e^{\alpha+\gamma}}{1+e^{\alpha+\gamma}}$	$\frac{1}{1+e^{\alpha+\gamma}}$
E	$\frac{e^{\alpha+\beta+\gamma+\sigma}}{1+e^{\alpha+\beta+\gamma+\sigma}}$	$\frac{1}{1+e^{\alpha+\beta+\gamma+\sigma}}$

	1	0
O	$\frac{1}{1+e^{\alpha^*+\beta^*+\gamma^*+\sigma^*}}$	$\frac{1}{1+e^{\alpha^*+\gamma^*}}$
E	$\frac{e^{\alpha^*+\beta^*+\gamma^*+\sigma^*}}{1+e^{\alpha^*+\beta^*+\gamma^*+\sigma^*}}$	$\frac{e^{\alpha^*+\gamma^*}}{1+e^{\alpha^*+\gamma^*}}$

$$\text{D.R.} = \frac{\left(\frac{e^{\alpha+\beta+\gamma+\sigma}}{1+e^{\alpha+\beta+\gamma+\sigma}}\right) / \left(\frac{1}{1+e^{\alpha+\beta+\gamma+\sigma}}\right)}{\left(\frac{e^{\alpha+\gamma}}{1+e^{\alpha+\gamma}}\right) / \left(\frac{1}{1+e^{\alpha+\gamma}}\right)}$$

$$\text{D.R.} = \frac{\left(\frac{e^{\alpha^*+\beta^*+\gamma^*+\sigma^*}}{1+e^{\alpha^*+\beta^*+\gamma^*+\sigma^*}}\right) / \left(\frac{1}{1+e^{\alpha^*+\beta^*+\gamma^*+\sigma^*}}\right)}{\left(\frac{e^{\alpha^*+\gamma^*}}{1+e^{\alpha^*+\gamma^*}}\right) / \left(\frac{1}{1+e^{\alpha^*+\gamma^*}}\right)}$$

$$e^{\beta+\sigma} = \text{D.R.} = e^{\beta^*+\sigma^*} \\ \beta + \sigma = \ln \text{D.R.} = \beta^* + \sigma^* \quad \text{cuando } Z \text{ y } Z^* = 1.$$

retrospectivo regresivo puede ser escrito como:

$$P(F=1; E, \underline{z}) = \frac{e^{\alpha+\beta E+\gamma \underline{z}+E \underline{\sigma} \underline{y}}}{1+e^{\alpha+\beta E+\gamma \underline{z}+E \underline{\sigma} \underline{y}}}$$

que equivale a

$$\ln \frac{P(E=1; E, \underline{z})}{1-P(F=1; E, \underline{z})} = \alpha + \beta E + \gamma \underline{z} + E \underline{\sigma} \underline{y}$$

El modelo logístico prospectivo se escribe como:

$$\ln \frac{P(E=1; F, \underline{z})}{1-P(E=1; F, \underline{z})} = \alpha + \beta F + \gamma \underline{z} + F \underline{\sigma} \underline{y}$$

En ambas ecuaciones \underline{y} es un vector q de las covarianzas que interaccionan con E o F respectivamente, donde q puede ser menor o igual que p (no todos los términos de covarianza interaccionan con la enfermedad o con el factor).

La desigualdad relativa para el modelo retrospectivo resulta ser:

$$\ln \text{D.R.} = \beta + \underline{\sigma} \underline{y} \quad \text{o} \quad \text{D.R.} = e^{\beta + \underline{\sigma} \underline{y}}$$

De igual manera, la desigualdad relativa del modelo prospectivo es:

$$\text{D.R.} = e^{\beta + \underline{\sigma} \underline{y}}$$

Los coeficientes β y $\underline{\sigma} \underline{y}$ tienen el mismo significado en ambos modelos, justificando así la forma en que fueron definidos en las ecuaciones anteriores.

Concluyendo, la desigualdad relativa no depende del método utilizado, ya sea retrospectivo o prospectivo. Ambos modelos proveen igual desigualdad relativa en términos de los parámetros β y $\underline{\sigma}$.

ESTIMACION DE LOS PARAMETROS β Y $\underline{\sigma}$ PARA LA DESIGUALDAD RELATIVA

Aunque β y $\underline{\sigma}$ tengan el mismo

significado para ambos modelos, sus estimados no son necesariamente los mismos.

$$\ln \frac{P(E=1; F, Z)}{1-P(E=1; F, Z)} = \alpha^* + \beta F + \gamma^* Z^*$$

Si las covarianzas son discretas, la data se puede presentar en una serie de tablas 2 x 2 en donde cada una corresponde a un resultado de covarianza. Breslow presentó dos métodos para estimar β y α (4). El primer método utiliza ecuaciones condicionales de máxima probabilidad obtenidas cuando los totales marginales son fijos y cuyas soluciones pueden requerir la técnica de iteración de Newton-Raphson. El segundo método asume una distribución normal asintótica y utiliza un modelo incondicional logarítmico-lineal para obtener aproximaciones a las ecuaciones de máxima probabilidad.

Breslow y Powers demostraron que los estimados de máxima probabilidad obtenidos del modelo logístico regresivo prospectivo son los mismos que los estimados utilizando la distribución asintótica, con la condición de que los efectos de la covarianza en enfermedad estén saturados en el modelo prospectivo (5). Los efectos de covarianza están saturados cuando son completamente dependientes del vector \underline{z} , cuando no hay grados de libertad.

Como ilustración, considérese el siguiente modelo logístico regresivo prospectivo:

sin término de interacción y con dos covarianzas que serían, por ejemplo, raza y ocupación. Supongamos que cada una tiene tres valores discretos: raza, los valores 1 = blanco, 2 = negro, 3 = mestizo; y ocupación, los valores 1 = profesional, 2 = trabajo manual, 3 = oficios domésticos. En este caso, habría nueve posibles combinaciones de covarianzas. Los efectos de las covarianzas en enfermedad son saturados cuando el vector \underline{z}^* incluye ocho términos. Supongamos que variables indicativas fueran utilizadas. Dos variables indicativas, R_1 y R_2 , serían utilizadas para representar las tres categorías de raza; y, de igual manera, dos variables indicativas, O_1 y O_2 , serían utilizadas para representar las tres categorías de ocupación. El vector \underline{z}^* puede tener los siguientes términos: $R_1, R_2, O_1, O_2, R_1O_1, R_2O_2, R_1O_2, R_2O_1$. Nótese que el número de parámetros α^* y γ^* equivale al número de resultados de covarianza ($1+8=9$).

Los efectos de covarianza en enfermedad están saturados cuando el número de parámetros α y γ es igual al número total de todos los posibles resultados de covarianza. En este caso, los estimados de γ y α obtenidos del modelo prospectivo son idénticos

a aquellos obtenidos con el modelo retrospectivo.

Si la covarianza tiene más valores (si es continuo, por ejemplo), el número de resultados de covarianza incrementa y muchos más parámetros tendrán que ser incluidos en el modelo para obtener saturación de los efectos de covarianza. Lógicamente que esto es impráctico. Se ha demostrado, sin embargo, que los estimados de β y σ en ambos modelos se asemejan más y más cuando más términos de covarianza son agregados a los modelos (5). Se puede llegar a un punto en que los estimados sean lo suficientemente similares sin tener que llegar a un número grande de parámetros.

LOS MODELOS LOGISTICOS RETROSPECTIVO Y PROSPECTIVO

Como se ha demostrado en las tablas 2 x 2, los modelos retrospectivos y prospectivos dan resultados equivalentes, ya que contienen la misma desigualdad relativa en términos de los parámetros de los modelos. Cuando hay suficiente ajuste de covarianza, el modelo prospectivo puede ser usado para estimar la desigualdad relativa en datos de caso-control, aun cuando las variables de exposición tienen más de dos niveles o cuando hay dos o más factores de riesgo, todos considerados por igual en importancia.

¿Cuánto ajuste de covarianza debe haber? La estrategia ideal sería incrementar el número de términos de covarianza en ambos modelos retrospectivo y prospectivo hasta que los estimados de β y σ sean similares y a la vez el número de parámetros sea razonable. Ya que esto no es siempre posible, la experiencia y el sentido común deben utilizarse para decidir la cantidad de términos de covarianza que deben incluirse en el modelo.

Se considera que cuando la distribución de covarianza es igual para los casos y controles, en caso de factor dicotómico, el modelo prospectivo requeriría menos ajuste de covarianza y sería el modelo preferido (5). El modelo prospectivo debe ser utilizado, por lo tanto, cuando los casos son iguales a los controles en las covarianzas. Cuando las covarianzas son fuertemente asociadas con la enfermedad y no tanto con el factor de exposición, se escoge el modelo retrospectivo. En resumen, el modelo que deba ser utilizado dependerá de cual requiere menos términos de covarianza para obtener un estimado de la desigualdad relativa.

Summary

Logistic regression has often been used to analyze retrospective data. There are two approaches to this analysis. One approach, the "prospective" logistic model, considers the disease

variable to be the dependent variable, while the "retrospective" model takes exposure to the risk factor to be the dependent variable. Theoretically, the correct model is the retrospective one, considering the sampling scheme of case-control studies. Complications may arise, however, with the "retrospective" approach: the exposure variable may have more than two levels

or there may be more than one risk factor to consider. It is shown, mathematically, that both models yield equivalent odds ratios in terms of the parameters β and α : $O.R. = e^{\beta + \alpha'y}$. The prospective model can therefore be used for estimating the odds ratio in case-control data, provided there is enough covariance adjustment.

AGRADECIMIENTO

El más sincero agradecimiento al Dr. Potter Chang, Profesor de la Universidad de California en Los Angeles, por su asesoramiento durante el transcurso de este trabajo y al Dr. William C. Reeves, Director del Laboratorio Conmemorativo Gorgas de Panamá, por su apoyo y sus valiosos comentarios.

BIBLIOGRAFIA

1. Mantel N, Haenszel W: Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 22 (4):719-748, 1959
2. Cox DR: *The analysis of binary data*, Methuen, London, 1970
3. Cornfield J: A method of estimating comparative rates from clinical data. Application to cancer of the lung, breast, and cervix. *JNCI* 11:1269-75, 1951
4. Breslow N: Regression analysis of the log odds ratio: A method for retrospective studies. *Biometrics* 32:409-16 1976
5. Breslow N, Powers W: Are there two logistic regressions for retrospective studies?, *Biometrics* 34:100-105, 1978
6. Kneale GW: Problems arising in estimating from retrospective survey data the latent period of juvenile cancers initiated by obstetric radiography. *Biometrics* 27:563-90, 1971
7. *DBase III Plus*. Ashton-Tate, Torrance, CA, 1985
8. Wilkinson L: *SYSTAT: The system for Statistics*. Systat, Inc. Evanston, IL, 1986
9. *BMDP Statistical Software*, BMDP Statistical Software, Inc. Los Angeles, 1985