

# Bioestadística

## Práctica 4

Vamos a insistir en esta práctica en los conceptos de *variable modificadora de efecto* y *variable de confusión*. Veremos también la *regresión logística*. Aplicaremos a unos mismos datos tanto el análisis propio de las tablas de contingencia como la regresión logística; esto nos permitirá resaltar algunas propiedades de la regresión logística, a la vez que relacionar los resultados de ambos métodos.

### Ejercicio 1

En uno de los últimos capítulos de teoría de la asignatura se han introducido los conceptos de *modificación de efecto* y *confusión*, cruciales en Epidemiología a la hora de interpretar los resultados de un estudio. Hay ocasiones en las que alguna variable distinta a las que hemos incluido como causa y efecto es más o menos «responsable» de esos resultados. Podemos pensar entonces que se trata de una variable que modifica el efecto o de una variable que confunde. Veamos qué diferencias hay entre unas y otras (de «¿Qué es una variable modificadora de efecto?»; Jokin de Irala, Miguel Ángel Martínez- -González y Francisco Guillén Grima; Medicina Clínica; 117, 297-302, 2001; y de «¿Qué es una variable de confusión?»; —; Medicina Clínica; 117, 377-385, 2001):

Variable de confusión: Siempre que se estima una medida de asociación, el resultado es suma de dos factores: el efecto real y la «confusión». Esta última se debe a que es imposible, en general, estudiar exactamente la misma población en las situaciones de expuesta y no expuesta a una determinada causa. Una variable (o factor) de confusión es una variable que distorsiona la medida de la asociación entre otras dos variables. Las condiciones que debe verificar una variable para ser de confusión son: 1) Estar asociada con el desenlace, tanto en expuestos como en no expuestos; 2) Estar asociada con la exposición pero no ser un resultado de la misma; 3) No ser un eslabón causal intermedio entre la exposición y el desenlace. La edad y el sexo son las principales candidatas a ser variables de confusión. Se habla de confusión cuando existen diferencias importantes entre los resultados brutos y los ajustados por los posibles factores de confusión. Entre los métodos para evitarla están: aleatorización de la muestra, restricción en la admisión de sujetos para la muestra y el emparejamiento de los datos. Entre los métodos para identificarla están: la estratificación y el análisis multivariante. Dentro del primer método, el estimador de Mantel-Haenszel combina las medidas de los distintos estratos mediante un promedio ponderado. Se suele considerar que una variable es confusora si el sesgo entre las razones cruda y ponderada es superior al 10%.

Variable modificadora de efecto: Es un concepto complejo que debe distinguirse claramente de la confusión, ya que su identificación determinará una actitud radicalmente opuesta por parte del investigador: así como en presencia de confusión el objetivo es eliminar una distorsión de la medida de asociación objeto de la investigación, ante la presencia de interacción el objetivo es describir mejor un fenómeno, una riqueza existente en los datos. Se puede concluir que hay modificación de efecto o lo contrario según la escala —aditiva o multiplicativa— escogida para evaluar el fenómeno. Es preciso, por tanto, especificar la escala en que se está midiendo. Además, para hablar del efecto de

una exposición es necesario tener en cuenta el valor de la otra variable o el estrato de la otra variable en el que se evalúa dicho efecto. Cuando no hay interacción se puede explicar el efecto de cada variable independientemente de las otras. En cuanto a la identificación: Se evalúa la relación causa-efecto, se excluyen las variables de confusión, se hace un análisis por estratos.

El siguiente cuadro, tomado de los mismos trabajos, aclara bastante la diferencia entre ambos tipos de variables.

**Características de las variables de confusión y de los modificadores del efecto**

Características	Confusión	Interacción
Significado biológico	No corresponde a un fenómeno biológico Es una distorsión de la asociación entre una exposición y un desenlace debido a una tercera variable que es el factor de confusión	Puede corresponder a un fenómeno biológico, sobre todo cuando existe una modificación de la aditividad de las medidas de efecto Corresponde al cambio del verdadero valor de la asociación entre una exposición y un desenlace, en los diversos niveles de una tercera variable que es la «modificadora del efecto»
Consecuencia de su presencia	Introduce un error o distorsión en la estimación de la medida de asociación	Enriquece la información que se puede dar de la medida del efecto
Reproducibilidad	No se reproduce necesariamente en el tiempo ni en otros estudios	En el caso de representar un fenómeno biológico, debería reproducirse en el tiempo o en otros estudios
Identificación	Elaboración de gráficas causales (DAG) Comparar asociaciones brutas con un promedio de las ajustadas (Mantel-Haenszel) Análisis estratificado Análisis multivariable+	Análisis del efecto de una variable en subgrupos de la otra Utilizar escalas aditivas y multiplicativas Análisis por subgrupos Análisis multivariable (términos de producto)
Actuación del investigador	Eliminar el efecto de confusión Prevenirlo en el diseño, controlarlo en el análisis, ajustando por cada factor de confusión	Describir en detalle este fenómeno Tablas con medidas del efecto de la exposición separados para cada subgrupo del modificador del efecto Su existencia no depende del diseño del estudio
Metodología analítica	Análisis estratificado, análisis multivariable, ajuste de tasas, otros procedimientos	Estimación manual o automática de medidas de asociación en subgrupos a partir de los datos de un modelo
Presentación científica	Estimación del efecto de la exposición ajustando por los factores de confusión  (En todo caso, comparación entre medidas del efecto brutas y ajustadas)	No se puede presentar un solo valor de la medida de efecto. Valores de medidas de asociación en cada subgrupo de interés Ayuda de gráficos para una mejor comprensión

En los dos ejemplos desarrollados en los apuntes de la asignatura (con *Epidat*, no los repetimos aquí), pueden verse los siguientes pasos para la identificación y actuación de una variable que sospechamos que puede ser modificadora de efecto o confusora:

Ejemplo 1: Se estratifica por la variable sospechosa (sexo) -> Para estudiar al capacidad modificadora de efecto de la variable, por un lado se comparan los valores de las razones de puntos de ambos estratos y por otro lado se observa el resultado del contraste de homogeneidad -> Como las razones son cercanas y no hay evidencia para rechazar la hipótesis de homogeneidad, se considera que la variable no es modificadora de efecto -> Se evalúa si la variable es confusora: para ello se unen los datos en una única muestra (son homogéneos) y se comparan las razones de puntos cruda y combinada (Mantel-Haenszel) -> Como son bastante distintas y el sesgo debido a la variable es del 155%, se concluye que es una variable de confusión -> El resultado que se considera correcto es el de la ponderación de Mantel-Haenszel, que ajusta por la variable.

Ejemplo 2: Se estratifica por la variable sospechosa (obesidad) -> Para estudiar al capacidad modificadora de efecto de la variable, por un lado se comparan los valores de las razones de puntos de ambos estratos y por otro lado se observa el resultado del contraste de homogeneidad -> Como las razones son bastante distintas en los estratos y hay evidencia para rechazar la hipótesis nula, se considera que se trata de una variable modificadora de efecto -> No es necesario pasar a estudiar la confusión, y los resultado que se proporcionan son los del análisis en cada estrato.

Vamos a hacer ahora uno de los ejemplos incluido en la ayuda de Epidat.

*Se estudia la infección hospitalaria posquirúrgica en pacientes operados de la cadera. El resultado se mide a través de la variable INFEC (INFEC=1 cuando el paciente se infecta a lo largo de la primera semana, INFEC=0 si no se infecta). Se desea evaluar un nuevo régimen técnico-organizativo de la atención de enfermería que se dispensa a estos pacientes. Se define la variable REGIMEN, de naturaleza dicotómica, que vale 1 si el sujeto estuvo ingresado bajo el nuevo régimen y 0 en caso de que haya estado atendido bajo el régimen convencional. Imagínese que se han estudiado 80 pacientes de diferentes edades, 36 de los cuales se han ubicado en el régimen convencional y 44 en el régimen en estudio, y que los resultados son los que se recogen en la siguiente tabla.*

Régimen	Infección	
	Sí (1)	No (0)
Nuevo (1)	7	37
Convencional (0)	14	22

OR=0,30

Considérese, además, que se quiere evaluar si la edad del paciente (se nombrará EDAD a esta variable) constituye una variable de confusión en la relación que pudiera existir entre el régimen organizativo y el hecho de desarrollar una infección.

Está claro que la variable EDAD cumple con los tres criterios convencionalmente admitidos para ser considerada como variable de confusión. Primero, el riesgo de infección aumenta con la edad. Segundo la proporción de pacientes mayores de 40 años es mayor en el grupo que recibió el régimen de atención convencional. Por último, es inverosímil creer que el efecto protector del régimen de intervención sobre el hecho de desarrollar una infección se produzca a través de la edad.

Para valorarlo, los datos se dividen en dos categorías de edad (menores e iguales o mayores de 40 años). En este caso, se codifica la variable del modo siguiente: EDAD=1 si el sujeto es menor de 40 años y EDAD=2 si no lo es, lo que produce la configuración que recoge la siguiente tabla.

		Infección	
		Sí (1)	No (0)
Edad<40 (1)	Régimen nuevo (1)	2	22
	Régimen convencional (0)	2	9
Edad≥40 (2)	Régimen nuevo (1)	5	15
	Régimen convencional (0)	12	13

OR<sub>1</sub>=0,41  
OR<sub>2</sub>=0,36

Hacemos el análisis estratificado siguiendo los menús:

*Métodos --> Tablas de contingencia --> Tablas 2x2 --> Estratificadas*

e introducimos los datos como se indica en la siguiente figura

Los resultados de este análisis son

```

Tipo de estudio : Transversal
Número de estratos: 2
Nivel de confianza: 95,0%

Tabla global
-----
                Enfermos      Sanos      Total
-----
Expuestos          7          37          44
No expuestos      14          22          36
-----
Total             21          59          80

RAZÓN DE PREVALENCIAS DE ENFERMEDAD (RP)

Estrato          RP          IC (95,0%)
-----
1                0,458333  0,073860  2,844155  (Katz)
2                0,520833  0,220003  1,233018  (Katz)
-----
Cruda            0,409091  0,185072  0,904273  (Katz)
Combinada (M-H) 0,508049  0,233023  1,107675
Ponderada        0,508839  0,233416  1,109250
    
```

Prueba de homogeneidad			
	Ji-cuadrado	gl	Valor p
Combinada (M-H)	0,0154	1	0,9012
Ponderada	0,0154	1	0,9012

RAZÓN DE PREVALENCIAS DE EXPOSICIÓN (RP)			
Estrato	RP	IC (95,0%)	
1	0,704545	0,257763	1,925743 (Katz)
2	0,549020	0,243466	1,238049 (Katz)
Cruda	0,531532	0,281380	1,004071 (Katz)
Combinada (M-H)	0,596818	0,315939	1,127406
Ponderada	0,605924	0,321976	1,140284

Prueba de homogeneidad			
	Ji-cuadrado	gl	Valor p
Combinada (M-H)	0,1451	1	0,7033
Ponderada	0,1429	1	0,7054

ODDS RATIO (OR)			
Estrato	OR	IC (95,0%)	
1	0,409091	0,049706	3,366928 (Woolf)
2	0,361111	0,100339	1,299602 (Woolf)
Cruda	0,297297	0,104080	0,849210 (Woolf)
Combinada (M-H)	0,372585	0,124791	1,112414
Ponderada	0,373463	0,125006	1,115749

Prueba de homogeneidad			
	Ji-cuadrado	gl	Valor p
Combinada (M-H)	0,0098	1	0,9210
Ponderada	0,0098	1	0,9210

PRUEBA DE ASOCIACIÓN DE MANTEL-HAENSZEL			
Ji-cuadrado	gl	Valor p	
3,1470	1	0,0761	

Como vemos, las razones de puntos por estratos son bastante cercanas y el contraste de homogeneidad refleja que los datos dan un apoyo grande a la hipótesis nula, por lo que no podemos considerar que la variable sea modificadora de efecto y pasamos a evaluar su sesgo de confusión, según la fórmula (de los apuntes de la asignatura):

$$\left(\frac{OR_{MH}}{OR_{crudo}} - 1\right) \cdot 100 = \left(\frac{0,372585}{0,297297} - 1\right) \cdot 100 = 25.32417$$

Como el sesgo de confusión es superior al 10%, podemos afirmar que es una variable de confusión.

Obs: Por lo que significan los datos, podrían haberse analizado como si se tratase de un estudio de cohortes, en vez de un estudio transversal; entonces habría que considerar la razón de tasas de incidencia en lugar de la razón de puntos.

## Ejercicio 2

En este ejercicio vamos a practicar los ejemplos de regresión logística que se incluyen en la ayuda de *Epidat*, e insistiremos en algunas características de esta regresión comparando los resultados de este ejercicio con los del anterior.

Entre los propósitos de muchas investigaciones epidemiológicas se halla el establecimiento de las leyes que rigen los fenómenos que se examinan. El examen se realiza típicamente en un marco complejo, donde la coexistencia de factores mutuamente relacionados determina el comportamiento de otros. Para sondear o incluso desentrañar la naturaleza de tales relaciones, el investigador puede auxiliarse, entre otras alternativas, del análisis de regresión. La regresión logística (RL) forma parte del conjunto de métodos estadísticos que caen bajo tal denominación y es la variante que corresponde al caso en que se valora la contribución de diferentes factores en la ocurrencia de un evento simple.

En general, la RL es adecuada cuando la variable de respuesta Y es politómica (admite varias categorías de respuesta, tales como MEJORA MUCHO, MEJORA, SE MANTIENE IGUAL, EMPEORA, EMPEORA MUCHO), pero es especialmente útil en particular cuando sólo hay dos posibles respuestas (cuando la variable de respuesta es dicotómica), que es el caso más común. Tal es el caso, por ejemplo, de las siguientes situaciones: el paciente muere o sobrevive en las primeras 48 horas de su ingreso, el organismo acepta o no un trasplante, se produjo o no un intento suicida antes de los 60 años, etc.) y lo que se quiere es construir un modelo que exprese la probabilidad de ocurrencia del evento de que se trate en función de un conjunto de variables independientes. Y se codifica como 1 (si se produce cierto desenlace) y como 0 en caso opuesto, de modo que la RL expresa  $P(Y=1)$  en función de ciertas variables relevantes a los efectos del problema que se haya planteado. La finalidad con que se construye ese modelo no es única: básicamente, puede tratarse de un mero esfuerzo descriptivo de cierto proceso, puede hacerse en el contexto de la búsqueda de explicaciones causales o puede desearse la construcción de un modelo para la predicción.

La RL es una de las técnicas estadístico-inferenciales más empleadas en la producción científica contemporánea.

**El modelo logístico.** El problema que resuelve la regresión logística es el de expresar la probabilidad de cierto desenlace ( $Y=1$ ) en función de r variables  $X_1, X_2, \dots, X_r$ . Concretamente, lo que hace el programa es hallar los coeficientes  $\beta_i$  que mejor se ajustan a la siguiente representación funcional:

$$P(Y = 1) = \frac{1}{1 + \exp(-\beta_0 - \beta_1 X_1 - \dots - \beta_r X_r)}$$

donde  $\exp(\cdot)$  representa la función exponencial.

Una idea que merece la pena resaltar sobre la regresión logística es la utilización de las variables ficticias (*dummy*) donde se tengan variables cualitativas nominales:

**Variables ficticias (*dummy*).** Las variables explicativas de tipo nominal con más de dos categorías deben ser incluidas en el modelo definiendo variables *dummy*. Epidat 3.1 permite indicar que una variable independiente sea tratada de este modo y, en tal caso, construye automáticamente las *dummy* correspondientes.

Brevemente dicho, el sentido de las variables *dummy* es el siguiente: supóngase que cierta variable es nominal (raza, religión profesada, grupo sanguíneo, etc.) y consta de k categorías; deben crearse entonces k-1 variables dicotómicas que son las llamadas variables *dummy* asociadas a esta variable nominal. Se denotarán por  $Z_1, Z_2, \dots, Z_{k-1}$ . A cada categoría o clase de la variable nominal le corresponde un conjunto de valores de los  $Z_i$  con el cual se identifica dicha clase.

La manera más usual de definir estas k-1 variables es la siguiente: si el sujeto pertenece a la primera categoría, entonces las k-1 variables *dummy* valen 0: se tiene  $Z_1 = Z_2 = \dots = Z_{k-1} = 0$ ; si el sujeto se halla en la segunda categoría, entonces  $Z_1 = 1$  y las restantes valen 0;  $Z_2$  vale 1 solo para aquellos individuos que están en la tercera categoría, en cuyo caso las otras variables asumen el valor 0, y así sucesivamente hasta llegar a última categoría, para la cual  $Z_{k-1}$  es la única que vale 1.

Por ejemplo, si la variable nominal de interés es el grupo sanguíneo, la cual tiene k=4 categorías (sangre tipo O, tipo A, tipo B y tipo AB); entonces se tendrían los siguientes valores de las variables *dummy* para cada grupo sanguíneo:

Variable nominal (grupo sanguíneo)	$Z_1$	$Z_2$	$Z_3$
O	0	0	0
A	1	0	0
B	0	1	0
AB	0	0	1

Repitamos el análisis de los datos del ejercicio anterior pero ahora aplicando regresión logística.

Al usar este submódulo hay que teclear los datos de una tabla de contingencia de 3 entradas con 8 celdas, o prepararla en EXCEL, Dbase o ACCESS para que el programa la lea automáticamente según la siguiente estructura:

INFECCION	REGIMEN	EDAD	FREQ
0	0	1	9
0	0	2	13
0	1	1	22
0	1	2	15
1	0	1	2
1	0	2	12
1	1	1	2
1	1	2	5

El archivo CADERA.xls que se incluye en Epidat 3.1 contiene la tabla arriba expuesta. Al emplear el programa, el usuario puede elegir cuántas y cuáles variables independientes incorporar al modelo. A continuación se exponen los resultados que se obtienen cuando se pone una sola variable (REGIMEN), y luego los que se producen cuando se añade la variable EDAD.



Parece que algunas versiones de *Epidat* dan un error al intentar cargar ese archivo. Una solución es crear otro archivo de un formato del que también puedan importarse los datos; abriendo el archivo con *Excel* y guardándolo con el formato de *Dbase (4)* se soluciona el problema, puesto que *Epidat* sí lo carga.

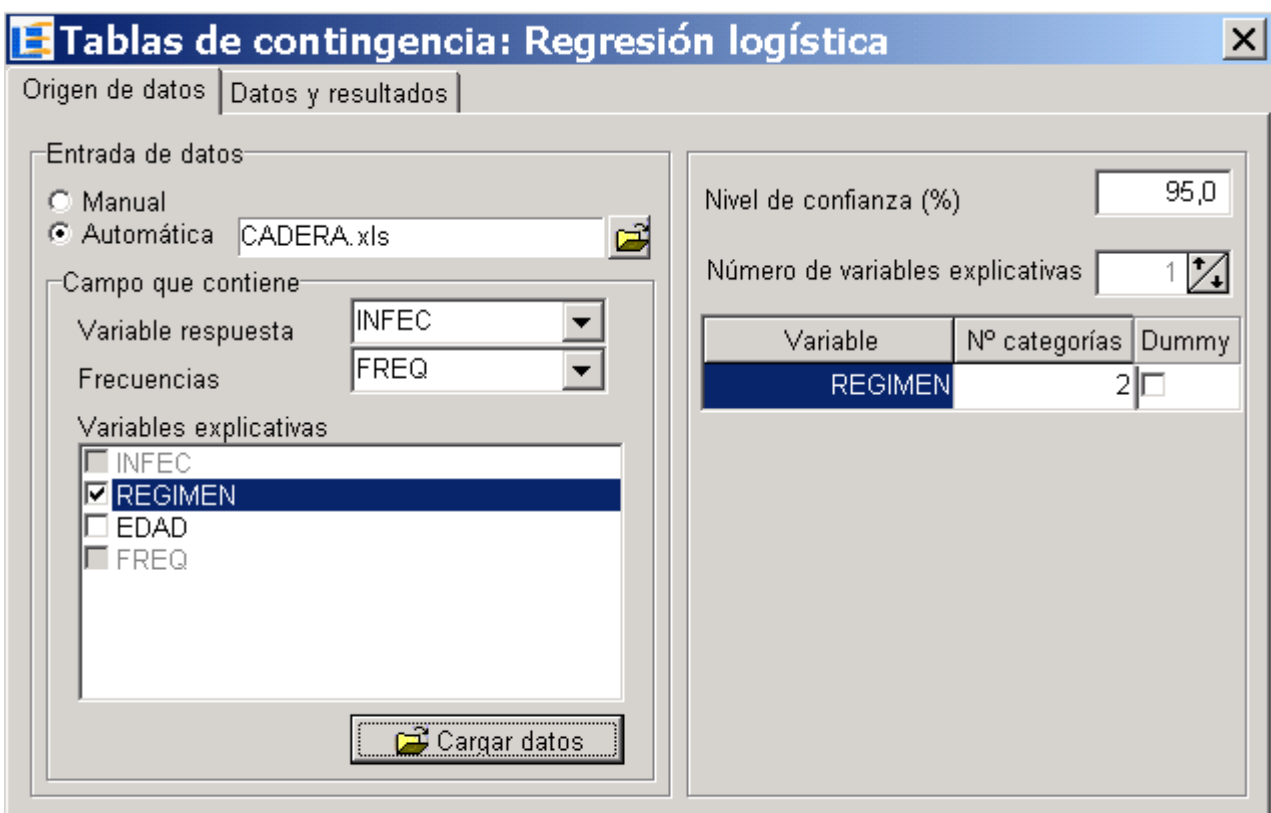
Hacemos el análisis siguiendo los menús:

*Métodos --> Tablas de contingencia --> Regresión logística*

e introducimos los datos de forma automática (desde el archivo que los contiene), desde

*C:\Archivos de programa\Epidat 3.1\Ejemplos\Tablas de contingencia*

Marcamos primero sólo REGIMEN como variable independiente



Al pulsar el botón *Cargar datos* y el de hacer lo cálculos, obtenemos:

```
Archivo de trabajo: C:\Archivos de programa\Epidat 3.1\Ejemplos\Tablas de
contingencia\CADERA.xls
Campo que contiene:
Variable respuesta: INFEC
Frecuencias: FREQ
Variables explicativas: REGIMEN

Nivel de confianza: 95,0%
```



Variable respuesta:

Valor N° sujetos

0	59
1	21
Total	80

La sucesión de estimadores ha convergido

N° iteraciones necesarias 3

-2 ln Verosimilitud inicial: 92,104901

-2 ln Verosimilitud final : 86,671970

Cociente de verosimilitud

Estadístico	gl	Valor p
5,4329	1	0,0198

Coefficiente de determinación: 0,0675

Variable	Coefficiente	EE	Valor de Z	Valor p
Constante	-0,451473			
REGIMEN	-1,210425	0,535158	-2,261807	0,0237

Variable	Odds ratio	IC (95,0%)	
REGIMEN	0,298071	0,104422	0,850838

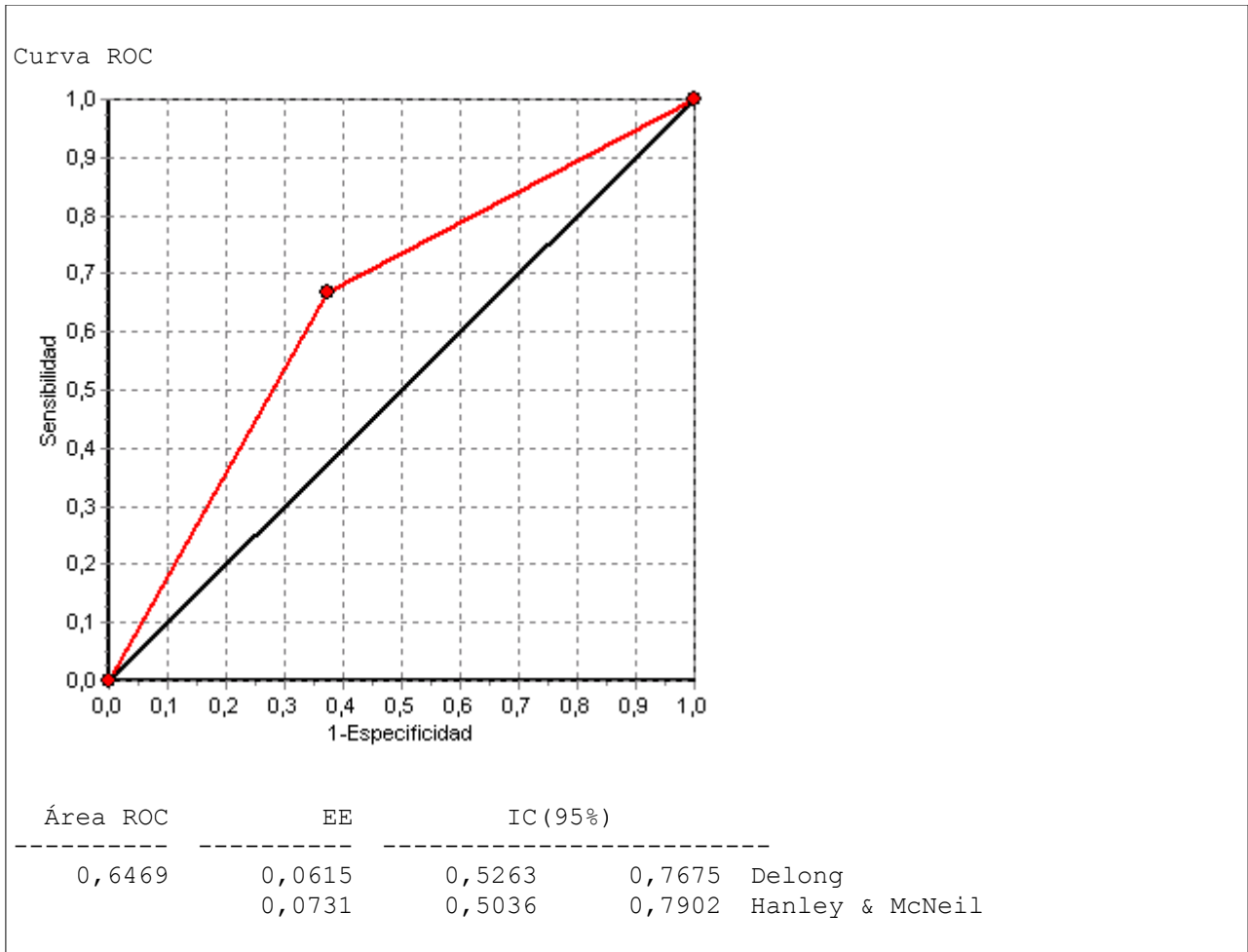
PRUEBA DE BONDAD DE AJUSTE DE HOSMER Y LEMESHOW

Grupos basados en los deciles

Grupo de Probabilidad	Respuesta = 0 Valor observado	Valor esperado	Respuesta = 1 Valor observado	Valor esperado
1	37	36,98	7	7,02
2	22	22,00	14	14,00

Ji-cuadrado	gl	Valor p
0,0001	0	NAN

Compara los resultados encerrados en el círculo con los de la razón de puntos cruda de los resultados de la página 5.



Si ahora se repite la regresión incluyendo además la variable EDAD como independiente (además de REGIMEN)

**Tablas de contingencia: Regresión logística**

Origen de datos | Datos y resultados

Entrada de datos

Manual  
 Automática CADERA.xls

Campo que contiene

Variable respuesta: INFECCION  
 Frecuencias: FREQ

Variables explicativas

INFECCION  
 REGIMEN  
 EDAD  
 FREQ

Nivel de confianza (%): 95,0  
 Número de variables explicativas: 2

Variable	Nº categorías	Dummy
REGIMEN	2	<input type="checkbox"/>
EDAD	2	<input type="checkbox"/>

Cargar datos

Se tiene:

Archivo de trabajo: C:\Archivos de programa\Epidat 3.1\Ejemplos\Tablas de contingencia\CADERA.xls

Campo que contiene:

Variable respuesta: INFEC

Frecuencias: FREQ

Variabes explicativas: REGIMEN EDAD

Nivel de confianza: 95,0%

Variable respuesta:

Valor N° sujetos

0	59
1	21
-----	
Total	80

La sucesión de estimadores ha convergido

N° iteraciones necesarias 3

-2 ln Verosimilitud inicial: 92,104901

-2 ln Verosimilitud final : 81,324308

Cociente de verosimilitud

Estadístico	gl	Valor p
10,7806	2	0,0046

Coefficiente de determinación: 0,1308

Variable	Coefficiente	EE	Valor de Z	Valor p
Constante	-2,759493			
REGIMEN	-0,974758	0,554901	-1,756635	0,0790
EDAD	1,332184	0,622533	2,139941	0,0324

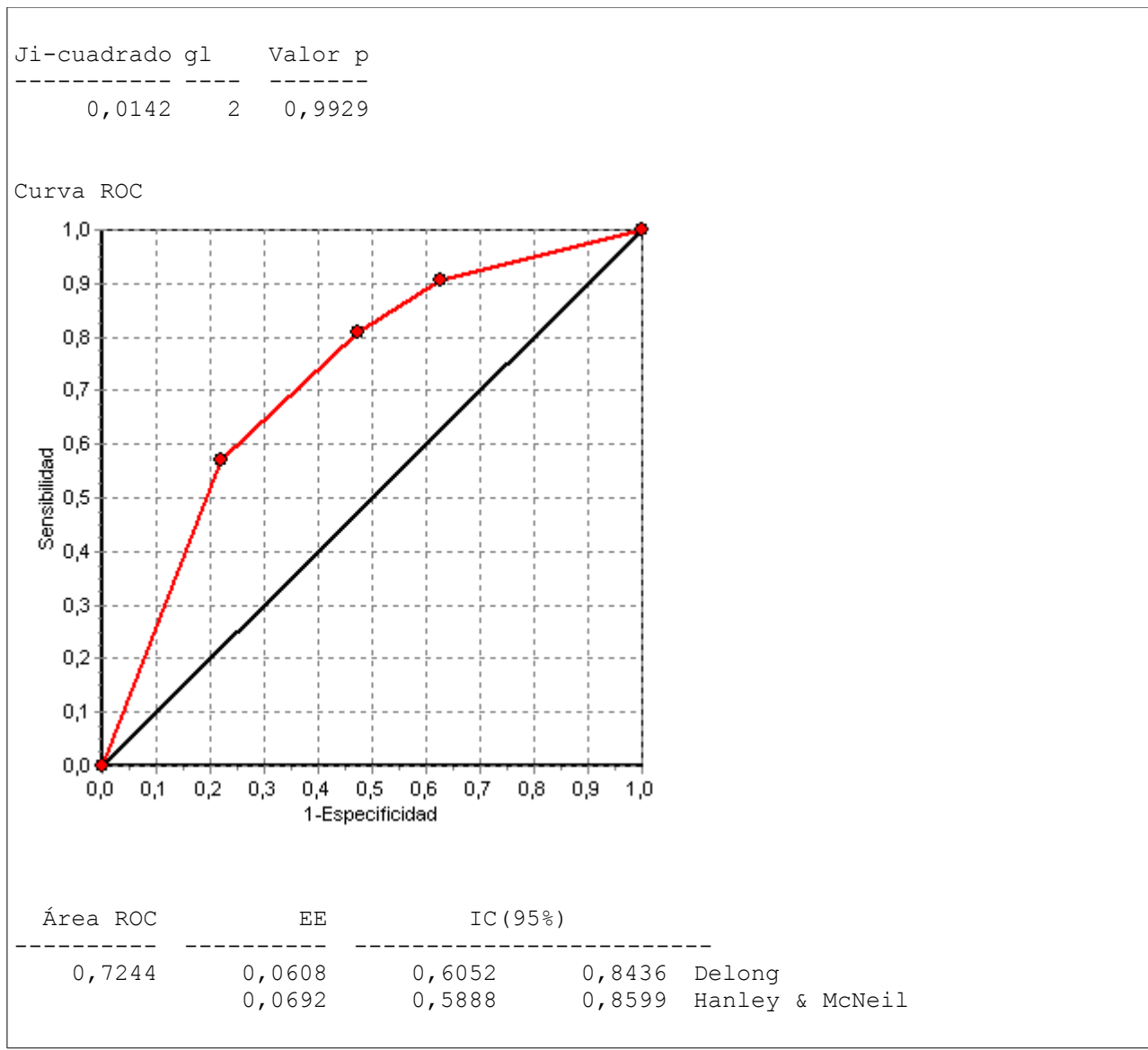
Variable	Odds ratio	IC (95,0%)	
REGIMEN	0,377284	0,127156	1,119438
EDAD	3,789310	1,118560	12,836931

PRUEBA DE BONDAD DE AJUSTE DE HOSMER Y LEMESHOW

Grupos basados en los deciles

Grupo de Probabilidad	Respuesta = 0		Respuesta = 1	
	Valor observado	Valor esperado	Valor observado	Valor esperado
1	22	22,01	2	1,99
2	9	8,87	2	2,13
3	15	14,89	5	5,11
4	13	13,09	12	11,91

Compara los resultados encerrados en el círculo con los de la razón de puntos cruda de los resultados de la página 5.



Obsérvese que ahora el área encerrada bajo la curva ROC es mayor que en el análisis sin la variable EDAD, lo que demuestra que ha mejorado el poder de predicción de la regresión.

Varias cosas que merece la pena resaltar sobre la regresión logística son:

**a) El análisis de la RL suple al análisis estratificado.** Nótese que, en el caso de los pacientes operados de la cadera, el odds ratio (0,298) coincide con la razón de productos cruzados correspondiente a la tabla. El intervalo de confianza que produce la RL [0,10 ; 0,85] es también coincidente con el que se obtiene mediante el análisis no paramétrico que arroja el análisis hecho a través de tablas de 2x2 incluido en otro submódulo del presente módulo. Por otra parte, el OR=0,377 que se obtiene a través del exponencial del coeficiente que corresponde a REGIMEN en el modelo que incluye las dos variables independientes, no es otra cosa que la estimación de Mantel Haenszel (lo mismo ocurre con el intervalo de confianza).

**b) La valoración sobre el posible papel confusor de un factor se desarrolla de manera ágil.** Basta correr el modelo con y sin el factor y comparar los coeficientes de la variable independiente. En el ejemplo de los operados de la cadera, se compara (puede utilizarse la fórmula que se incluye al principio de la página 6) 0,298 con 0,377 lo cual permite pensar que sí hay efecto confusor. El OR correspondiente a REGIMEN tiene, en el primer caso, un intervalo de confianza que no contiene al 1 (significativo al nivel 0,05) mientras que el que se obtiene cuando se controla la edad sí lo contiene (pierde la significación).

c) **El ajuste suele ser bueno.** El resultado que se ha obtenido en estos ejemplos, donde los valores esperados y observados son muy parecidos, es típico.

d) **Si el contexto del problema es predictivo, la probabilidad del suceso para un perfil de entrada dado ha de computarse independientemente empleando los coeficientes estimados.**

Y algunas recomendaciones que también se dan en la ayuda de *Epidat* son:

- Las variables explicativas deben tener una relación monótona con la probabilidad del evento que se estudia.
- Las variables independientes involucradas en el modelo no deben estar correlacionadas entre sí. Si la correlación entre dos variables es alta, entonces los resultados de la RL son poco confiables. Concretamente, los errores estándares se incrementan apreciablemente y suele ocurrir que los coeficientes no son significativamente diferentes de cero, aunque la aportación global de las variables sí lo sea.
- Debe recordarse que el conjunto de variables *dummy* constituye un todo indisoluble con el cual se sule a una variable nominal. Cualquier decisión que se adopte o valoración que se haga concierne al conjunto íntegro.
- Es muy importante distinguir entre un contexto explicativo y un contexto predictivo. En el primer caso, el modelo para cada posible factor de riesgo o protector se ajusta con los factores que pueden ser confusores para él. Solo en los estudios predictivos se ajusta el mejor modelo. Debe tenerse en cuenta, en este caso, que una variable puede tener valor predictivo aunque no sea parte del mecanismo causal que produce el fenómeno en estudio.

05/06/07

David Casado de Lucas: <http://www.est.uc3m.es/dcasado/>  
Departamento de Estadística  
Universidad Carlos III de Madrid

