

# ANÁLISIS DE REGRESIÓN PARA UN ESTUDIO CASO-CONTROL CON DATOS APAREADOS

M.T. Alves-Pérez<sup>(1)</sup>, I. Martínez-Silva<sup>(1)</sup>, J. Cubiella-Fernández<sup>(2)</sup>, M.J. García-García<sup>(3)</sup>

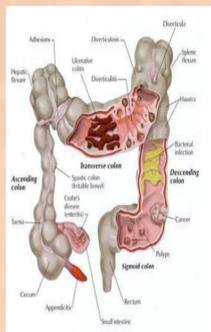
(1) Unidad de Bioestadística - Departamento de Estadística e I.O., Universidad de Santiago de Compostela; (2) Servicio de Aparato Digestivo, Complejo Hospitalario de Ourense; (3) Unidad de Epidemiología, Complejo Hospitalario de Ourense

## 1.- Introducción

La Colitis Isquémica (C.I.) es una enfermedad inducida por la reducción del flujo sanguíneo colónico<sup>(1)</sup>. Aunque se ha asociado a diversas enfermedades cardiovasculares, no se han definido claramente los factores de riesgo.

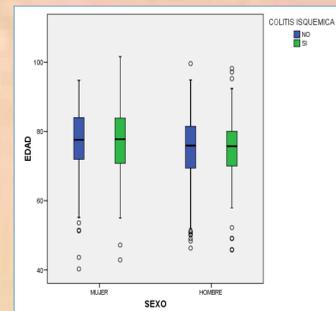
El objetivo de este estudio es evaluar qué factores de riesgo se relacionan con el desarrollo de la colitis isquémica<sup>(2)</sup> así como valorar el efecto sobre el desarrollo de CI que tienen los fármacos hipotensores y los antiinflamatorios no esteroideos<sup>(3)</sup>.

Para ello se ha diseñado un estudio retrospectivo caso-control con datos apareados.



## 2.- Selección de casos y apareamiento de datos

Se han incluido en este estudio como casos, una cohorte de 169 pacientes que presentaron un episodio de CI. Se han seleccionado los controles realizando un apareamiento de razón 1:2 en base a las variables: edad, sexo y año de diagnóstico.



En este gráfico se puede observar cómo las variables se encuentran apareadas de modo correcto; fruto de ello es la no existencia de grandes diferencias entre la media de edad de hombres y mujeres ni tampoco entre los casos y los controles.

## 4.- Modelos de Regresión<sup>(4)</sup>

**Regresión Lineal:** dadas  $X_1, \dots, X_n$  variables explicativas siendo y la variable respuesta, obtenemos el modelo

$$y = \beta_0 + \sum_{j=1}^{p-1} \beta_j x_j + \varepsilon \quad \text{donde } \beta_i$$

son los coeficientes del modelo y  $\varepsilon$  es el término de error.

**Regresión logística:** en este caso y es una variable dicotómica en  $\{0,1\}$ , y la  $E(y/x) \in [0,1]$  por lo que se define la transformación logit como :

$$g(x) = \ln \left[ \frac{\pi(x)}{1 - \pi(x)} \right] = \beta_0 + \sum_{j=1}^{p-1} \beta_j x_j, \quad \pi(x) = \frac{e^{\beta_0 + \sum_{j=1}^{p-1} \beta_j x_j}}{1 + e^{\beta_0 + \sum_{j=1}^{p-1} \beta_j x_j}}$$

y así logramos una función  $g(x)$  que ajusta a  $y = \pi(x) + \varepsilon$ .

**Regresión logística condicionada:** es necesaria la utilización de k funciones  $\pi_k(x)$ , con  $k=1, \dots, K$  para K valores de la variable identificadora, quedando definida:

$$\pi_k(x) = \frac{e^{\beta_{0,k} + \sum_{j=1}^{p-1} \beta_j x_j}}{1 + e^{\beta_{0,k} + \sum_{j=1}^{p-1} \beta_j x_j}}$$

donde  $\alpha_k$  indica la contribución al logit de todos los términos constantes dentro del k-ésimo extracto.

## 3.- Limitaciones y problemas



La gran ventaja de este tipo de estudios, retrospectivo y caso-control, es el bajo coste que poseen. En nuestro caso, como la incidencia de CI es del 0,9%; en una muestra aleatoria de 500 pacientes, sólo tendríamos 4'5 pacientes con dicha enfermedad.

### LIMITACIONES:

- Se trata de un estudio retrospectivo
- No se podrá analizar la relación con otras patologías
- No se podrá valorar el efecto de las variables por las que se ha realizado el apareamiento

### PROBLEMAS:

¿Cómo debemos analizar estos datos?

¿Es correcto una regresión logística?

¿O debemos realizar una regresión logística condicional?

FALSO

VERDADERO

## 5.- Recogida y análisis de los datos

SEXO	EDAD	PAR	CASO	HTA	DIAB.
0	40	1	0	0	0
0	44	1	0	0	0
0	43	1	1	0	0
1	50	169	0	0	0
1	51	169	0	0	1
1	49	169	1	0	0

Código Stata (5):

```
clogit variable-respuesta var1 var2 var3, group(PAR) # Modelo de regresión
clogit ,or # OR e intervalo de confianza del OR del modelo anterior
```

Código R(6):

```
library(survival)
fit.CLR <- clogit(variable-respuesta ~ var1+var2+var3+strata(par),method = "exact",data
=datos) # Modelo de regresión
summary(fit.CLR) # Resumen del modelo
```

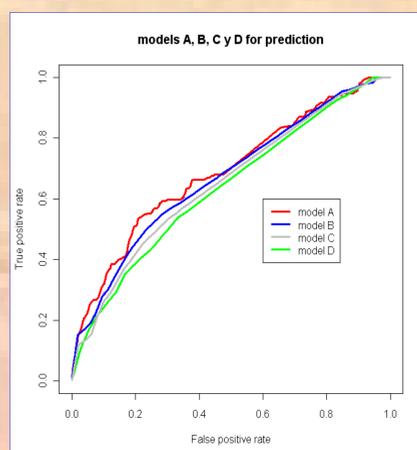


En la tabla superior izquierda, tenemos un ejemplo de cómo deben ser almacenados los datos, en un estudio caso-control para datos apareados. Y en el cuadro de la derecha, el código de R y Stata para su análisis, mediante regresión logística condicionada.

## 6.- Conclusiones

La tabla inferior recoge la comparativa de modelos y en la imagen de la derecha pueden observarse las curvas ROC para los mismos.

Modelo	Nº vbles.	Nº signif.	Df	R2	ANOVA	AUC
A	28	7	25	0.122	-----	0.6873
B	8	7	8	0.099	0.9144	0.6689
C	7	6	7	0.077	0.2591	0.6354
D	6	5	6	0.066	0.1053	0.6510



### Modelo B:

Variables	p-valor	Coefficientes	I.C.	OR	I.C.
Diabetes	0.054	0.541	-0.010 1.092	1.718	0.990 2.980
Dislipemia	0.006	0.709	-0.203 1.215	2.033	1.225 5.372
Insuficiencia cardíaca	0.033	0.882	0.070 1.694	2.416	1.073 5.441
Arteriopatía periférica	0.009	1.474	0.361 2.586	4.365	1.435 13.272
Antecedente de CI	0.006	3.032	0.875 5.189	20.74	2.399 179.329
Tratamiento con digital	0.017	-1.356	-2.467 -0.245	0.258	0.085 0.783
Tratamiento con AAS	0.028	0.546	0.058 1.035	1.727	1.059 2.816

Es evidente la elección del modelo B, dada la poca pérdida de explicación y la significativa reducción del número de variables utilizadas en él. Cabe resaltar, que estos modelos son explicativos y no predictivos, debido al bajo valor del R<sup>2</sup>-ajustado. De hecho, en todo el estudio han estado siempre por debajo del 20%, confirmando que una predicción a este nivel no es adecuada.

### REFERENCIAS:

- Gandhi SK, Hanson MM, Vernava AM, Kaminski DL, Longo WE. Ischemic colitis. Dis Colon Rectum 1996; 39:88-100.
- Montoro MA, Santolaria S.B, Sánchez-Puértolas B, Vera F and GTECIE Group. Clinical characteristics and outcome of ischemic colitis in Spain. A multicentre and prospective study (ICS study). Gut 2006; 55 (suppl V) A208.
- Higgins PDR, Davis KJ, Laine L. Systematic review: the epidemiology of ischaemic colitis. Aliment Pharmacol Ther 2004; 19: 729-38.
- David W. Hosmer, Stanley Lemeshow.(2000) Applied Logistic Regression Ed. Willey-Interscience Publication
- 14-user Stata for Windows (network) perpetual license. Serial number: 1910540659. Licensed to: Consellería de Sanidade - Servizo Galego de Saúde.
- R Development Core Team, 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.

### AGRADECIMIENTOS:

Carmen Cadarso Suárez  
Francisco Gude  
Sampedro

### MÁS INFORMACIÓN:

[malvesperez@gmail.com](mailto:malvesperez@gmail.com)  
[isabelmaria.martinez@usc.es](mailto:isabelmaria.martinez@usc.es)