

Verosimilitud empírica aplicada a las curvas ROC



Emilio Letón
Dpto. IA
UNED



Elisa M. Molanes
Dpto. Estadística
UC3M



elisamaria.molanes@uc3m.es

emilio.leton@dia.uned.es

16-Oct-2009

TESTS DE DIAGNÓSTICO DICOTÓMICOS

Índices de diagnóstico
Tests múltiples

TESTS DE DIAGNÓSTICO CONTINUOS

INFERENCIA DEL ÍNDICE DE YOUDEN

TESTS DE DIAGNÓSTICO DICOTÓMICOS

	T+	T-	
Enf	a	b	r_1
Sano	c	d	r_2
	s_1	s_2	n

	T+	T-	
Enf	36	4	40
Sano	12	93	105
	48	97	145

OR=69.75 (p<0.0001)

	T+	T-	
Enf	10	30	40
Sano	12	93	105
	22	123	145

OR=2.58 (p=0.04)

Índices de diagnóstico

- Sensibilidad y Especificidad.
- Valores predictivos.
- Razones de verosimilitud.

Sensibilidad y especificidad

$$\text{Sensibilidad} = \pi'_1 = \Pr(T+ | \text{Enf})$$

$$\text{Especificidad} = 1 - \pi'_2 = \Pr(T- | \text{Sano})$$

$$\text{Sens} = p'_1 = \frac{a}{r_1}$$

$$\text{Espe} = 1 - p'_2 = 1 - \frac{c}{r_1} = \frac{d}{r_2}$$

Dependencia de la prevalencia

	T+	T-		T+	T-		T+	T-	
Enf	36	4	40	36	4	40	360	40	400
Sano	120	930	1050	12	93	105	12	93	105

Prev 3.7% 27.6% 79.2%

Sens 90.0% 90.0% 90.0%

Espe 88.6% 88.6% 88.6%

Índices relacionados

$$\text{Eficacia} = \text{Pr ev} \cdot \text{Sens} + (1 - \text{Pr ev}) \cdot \text{Espe}$$

$$\text{Youden} = \text{Sens} + \text{Espe} - 1$$

$$\text{Efic} = \frac{a + d}{n}$$

$$J = \frac{a}{r_1} + \frac{d}{r_2} - 1$$

Valores predictivos

$$\text{VP}_{\text{positivo}} = \pi_1 = \Pr(\text{Enf} \mid T +)$$

$$\text{VP}_{\text{negativo}} = 1 - \pi_2 = \Pr(\text{Sano} \mid T -)$$

$$\text{VP}(+) = p_1 = \frac{a}{s_1}$$

$$\text{VP}(-) = 1 - p_2 = 1 - \frac{b}{s_2} = \frac{d}{s_2}$$

Dependencia de la prevalencia

	T+	T-		T+	T-		T+	T-	
Enf	36	4	40	36	4	40	360	40	400
Sano	120	930	1050	12	93	105	12	93	105

Prev 3.7% 27.6% 79.2%

VP(+) 23.1% 75.0% 96.8%

VP(-) 99.6% 95.9% 69.9%

Incorporación de la prevalencia

$$VP(+)=\frac{\text{Pr ev} \cdot \text{Sens}}{\text{Pr ev} \cdot \text{Sens} + (1 - \text{Pr ev}) \cdot (1 - \text{Espe})}$$

$$VP(-)=\frac{(1 - \text{Pr ev}) \cdot \text{Espe}}{(1 - \text{Pr ev}) \cdot \text{Espe} + \text{Pr ev} \cdot (1 - \text{Sens})}$$

$$\text{Pr}(\text{Enf} \mid T +) = VP(+)$$

$$\text{Pr}(\text{Enf} \mid T -) = 1 - VP(-)$$

Razones de verosimilitud

$$\text{LR}_{\text{positivo}} = \frac{\pi'_1}{\pi'_2} = \frac{\Pr(T+ | \text{Enf})}{\Pr(T+ | \text{Sano})}$$

$$\text{LR}_{\text{negativo}} = \frac{1 - \pi'_1}{1 - \pi'_2} = \frac{\Pr(T- | \text{Enf})}{\Pr(T- | \text{Sano})}$$

$$\text{LR}(+) = \frac{p'_1}{p'_2} = \frac{\text{Sens}}{1 - \text{Espe}}$$

$$\text{LR}(-) = \frac{1 - p'_1}{1 - p'_2} = \frac{1 - \text{Sens}}{\text{Espe}}$$

Dependencia de la prevalencia

	T+	T-		T+	T-		T+	T-	
Enf	36	4	40	36	4	40	360	40	400
Sano	120	930	1050	12	93	105	12	93	105

Prev 3.7%

27.6%

79.2%

LR(+) 7.875

7.875

7.875

LR(-) 0.113

0.113

0.113

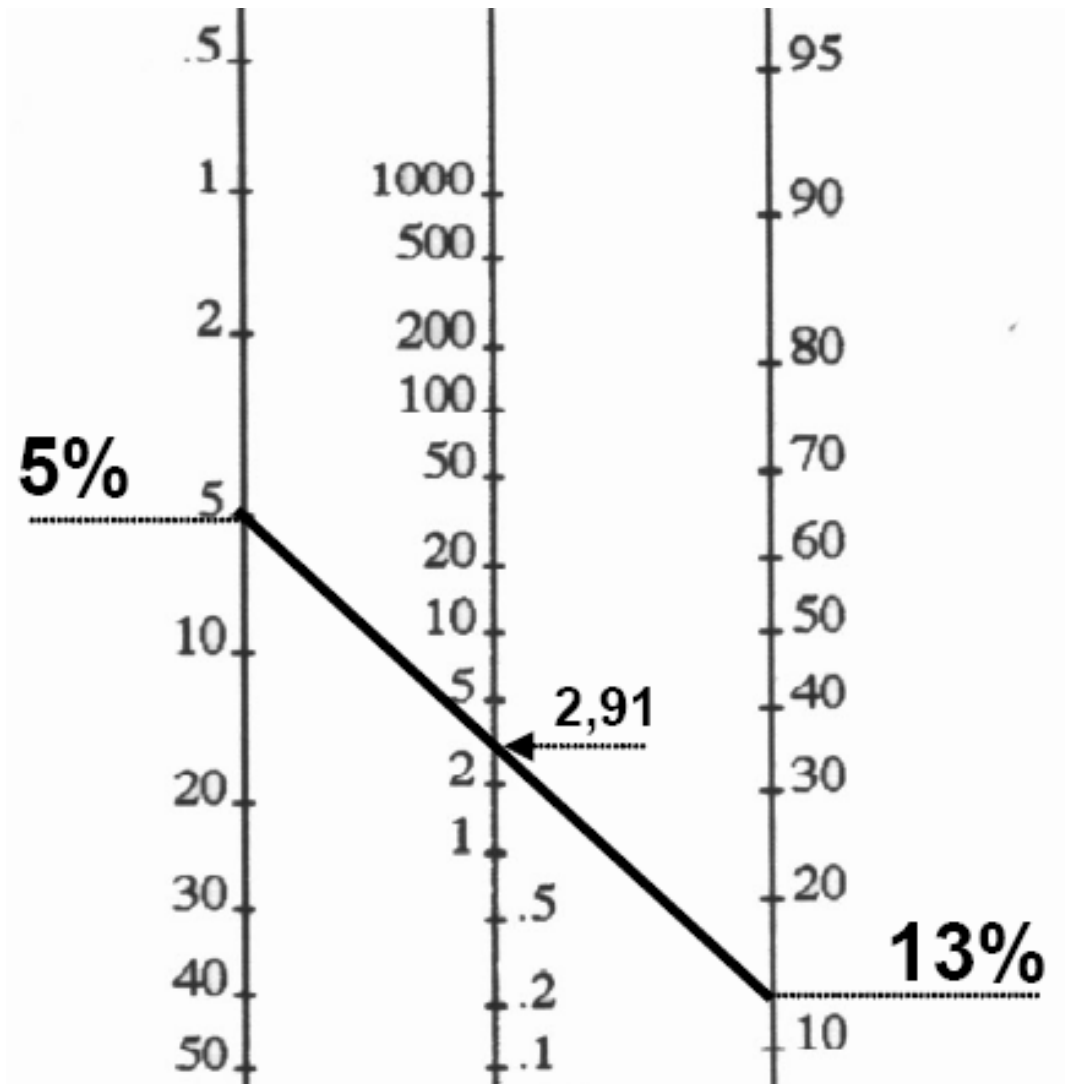
Incorporación de la prevalencia

$$\Pr(\text{Enf} \mid T +) = \frac{\text{Pr ev} \cdot \text{LR}(+)}{\text{Pr ev} \cdot \text{LR}(+) + (1 - \text{Pr ev})}$$

$$\Pr(\text{Enf} \mid T -) = \frac{\text{Pr ev} \cdot \text{LR}(-)}{\text{Pr ev} \cdot \text{LR}(-) + (1 - \text{Pr ev})}$$

$$\Pr(\text{Enf} \mid \text{res}) = \frac{\text{Pr ev} \cdot \text{LR}(\text{res})}{\text{Pr ev} \cdot \text{LR}(\text{res}) + (1 - \text{Pr ev})}$$

Nomograma de Fagan



Ganancia de la información

$$\text{Odds priori} = \frac{\text{Pr}(\text{Enf})}{1 - \text{Pr}(\text{Enf})} = \frac{\text{Pr ev}}{1 - \text{Pr ev}}$$

$$\text{Odds pos}(+) = \frac{\text{Pr}(\text{Enf} \mid T +)}{1 - \text{Pr}(\text{Enf} \mid T +)} = \text{Odds priori} \cdot \text{LR}(+)$$

$$\text{Odds pos}(-) = \frac{\text{Pr}(\text{Enf} \mid T -)}{1 - \text{Pr}(\text{Enf} \mid T -)} = \text{Odds priori} \cdot \text{LR}(-)$$

$$\text{OR} = \frac{\text{LR}(+)}{\text{LR}(-)}$$

Tests múltiples

	Test1	Test2	Test3
Sens	0.90	0.70	0.75
Espe	0.80	0.75	0.90
LR(+)	4.50	2.80	7.50
LR(-)	0.12	0.40	0.27

Prev=0.15

Combinación de razones de verosimilitud

$$\begin{aligned} \text{LR}_{\text{global}}(+, -, +) &= \text{LR}(+) \cdot \text{LR}(-) \cdot \text{LR}(+) \\ &= 4.5 \cdot 0.4 \cdot 7.5 = 13.5 \end{aligned}$$

$$\begin{aligned} \text{Pr}(\text{Enf} \mid +, -, +) &= \frac{\text{Pr ev} \cdot \text{LR}_{\text{global}}(+, -, +)}{\text{Pr ev} \cdot \text{LR}_{\text{global}}(+, -, +) + (1 - \text{Pr ev})} \\ &= \frac{0.15 \cdot 13.5}{0.15 \cdot 13.5 + (1 - 0.15)} = 0.70 \end{aligned}$$

TESTS DE DIAGNÓSTICO DICOTÓMICOS

TESTS DE DIAGNÓSTICO CONTINUOS

Curvas ROC

INFERENCIA DEL ÍNDICE DE YAUDEN

TESTS DE DIAGNÓSTICO CONTINUOS

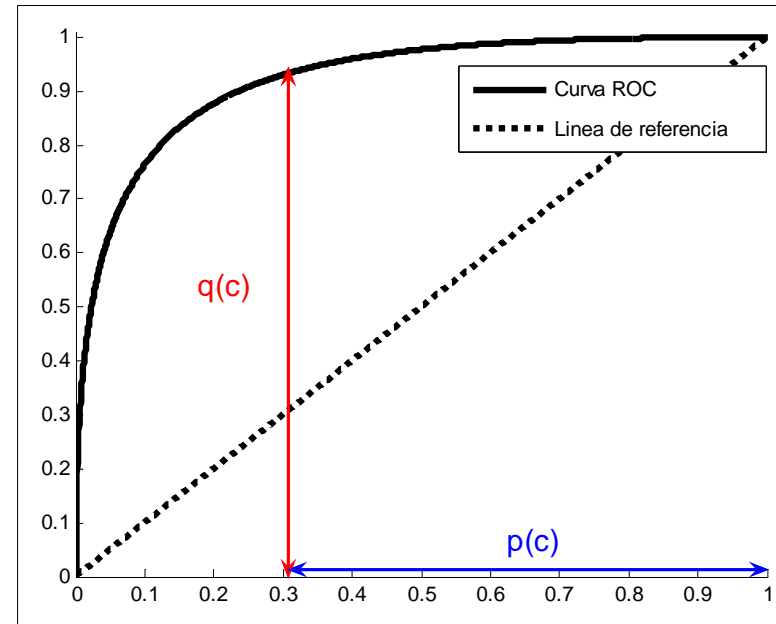
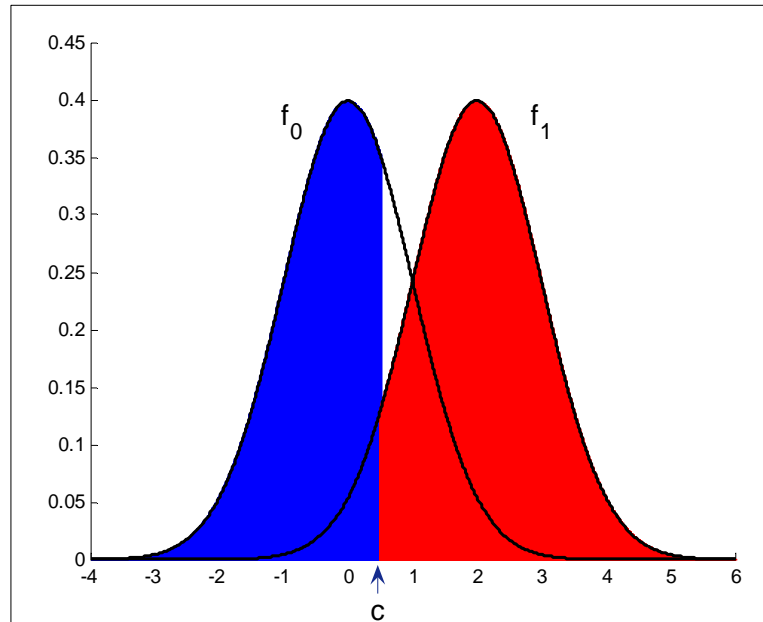
- PSA: antígeno específico de la próstata.
- TSH: hormona específica de la tiroides.

Umbral c

$$\text{Sens}(c) = q(c) = \Pr(X > c \mid D) = \Pr(X_1 > c)$$

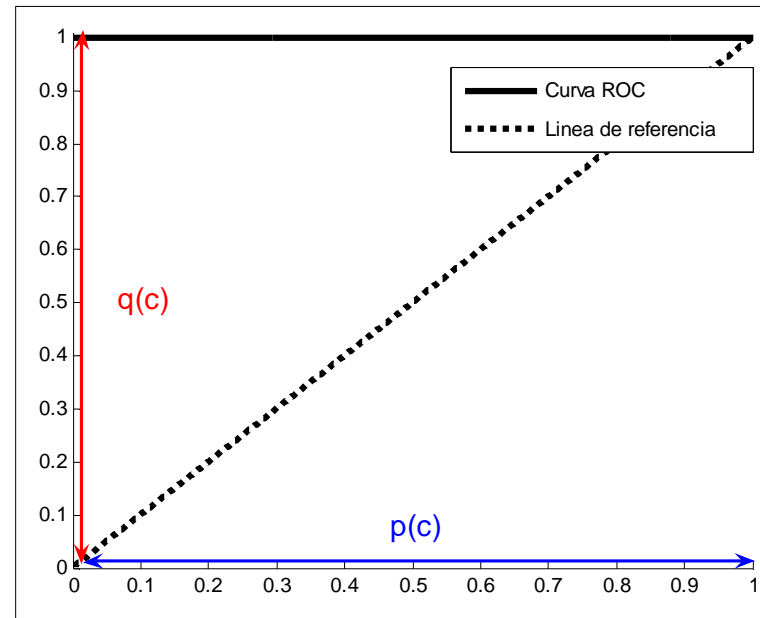
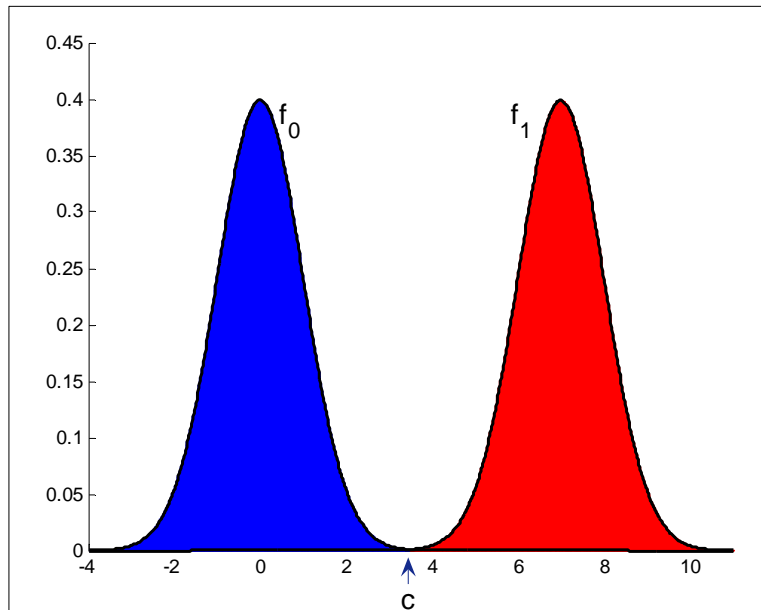
$$\text{Espe}(c) = p(c) = \Pr(X \leq c \mid H) = \Pr(X_0 < c)$$

Curvas “Receiver Operating Characteristic”



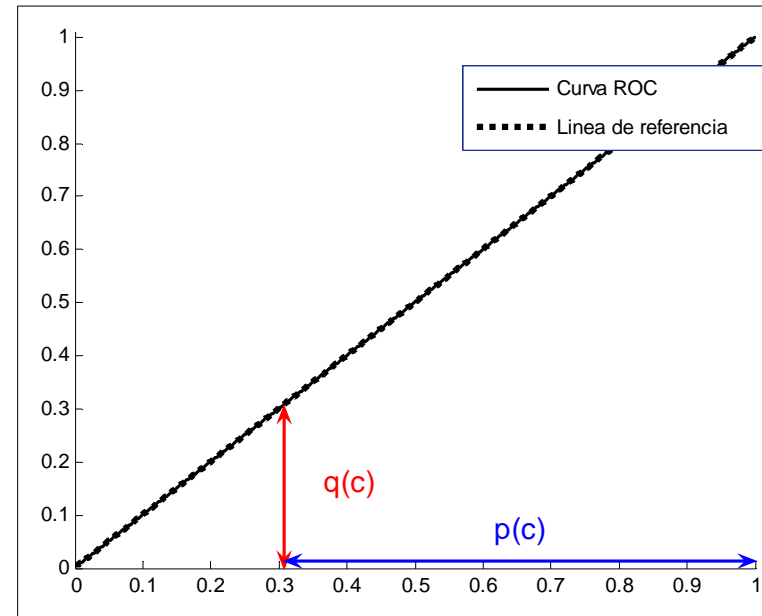
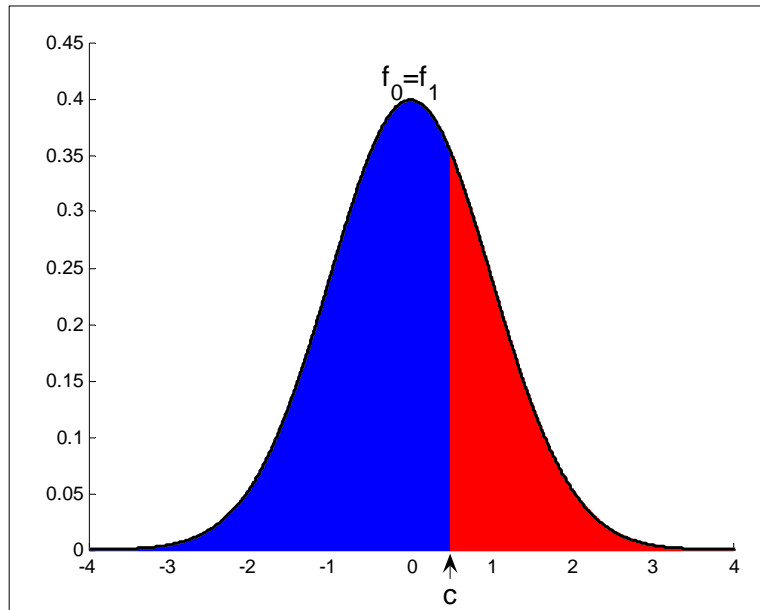
$$(1 - p(c), q(c)) = (1 - \text{Espe}, \text{Sens}) \quad \text{ROC}(t) = \bar{F}_1(\bar{F}_0^{-1}(t))$$

Poblaciones no solapadas



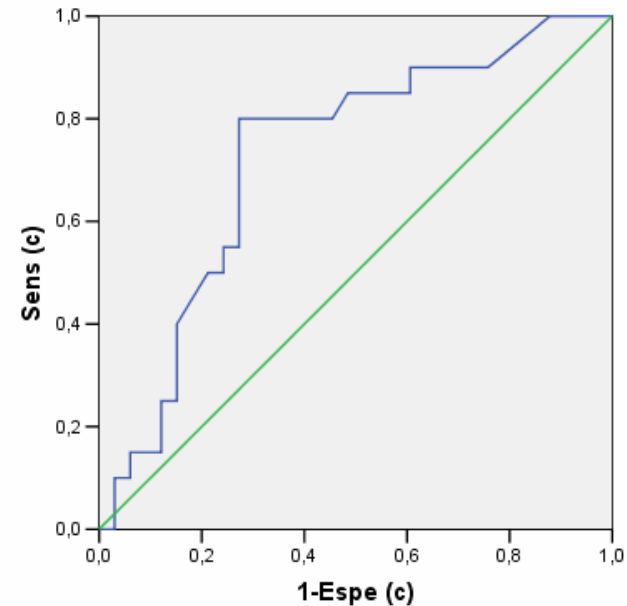
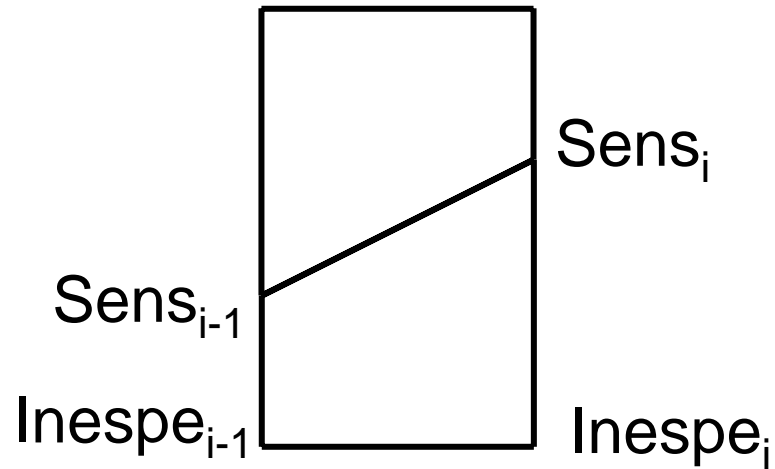
$$(1 - p(c), q(c)) = (1 - 1, 1) = (0, 1)$$

Poblaciones completamente solapadas



$$(1 - p(c), q(c)) = (1 - p(c), 1 - p(c))$$

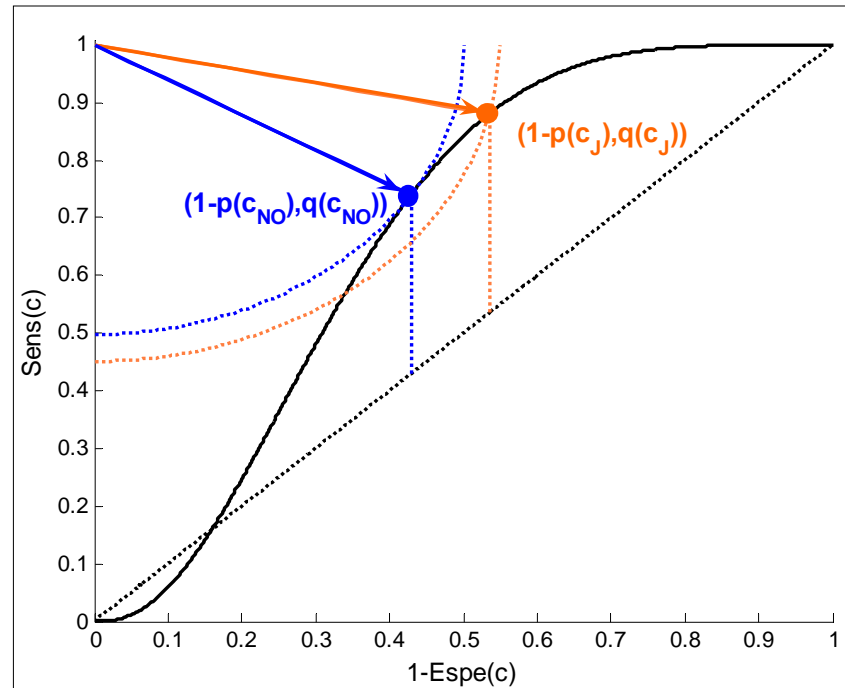
AUC



$$\sum_{i=1}^k (Inespe_i - Inespe_{i-1}) \cdot \left(\frac{Sens_i + Sens_{i-1}}{2} \right)$$

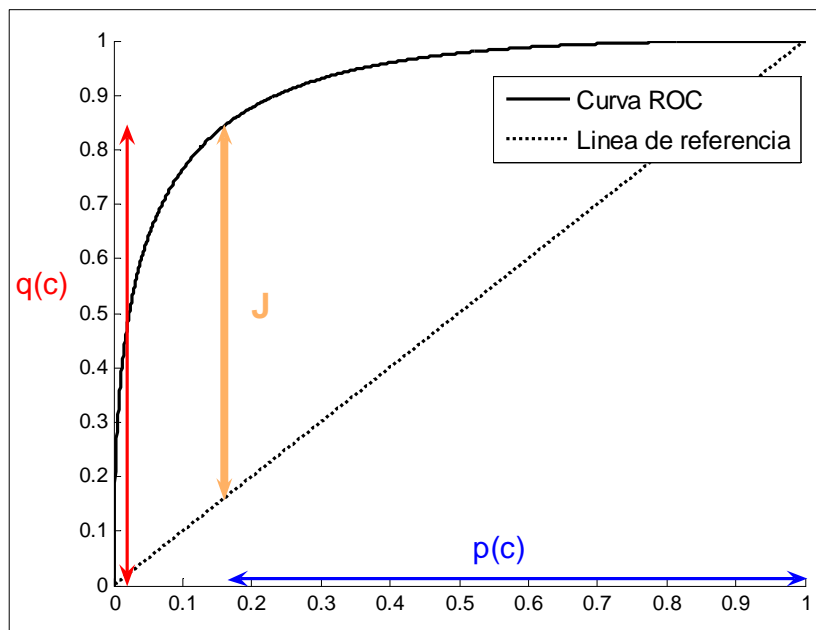
Punto de corte óptimo

- Esquina Noroeste.
- Índice de Youden.



Youden (1950), Fluss et al. (2005), Schisterman et al. (2005), Le (2006), Schisterman and Perkins (2007), ...

Índice de Youden (1/3)



$$J = \max \{J(c); c \in \mathfrak{R}\}$$

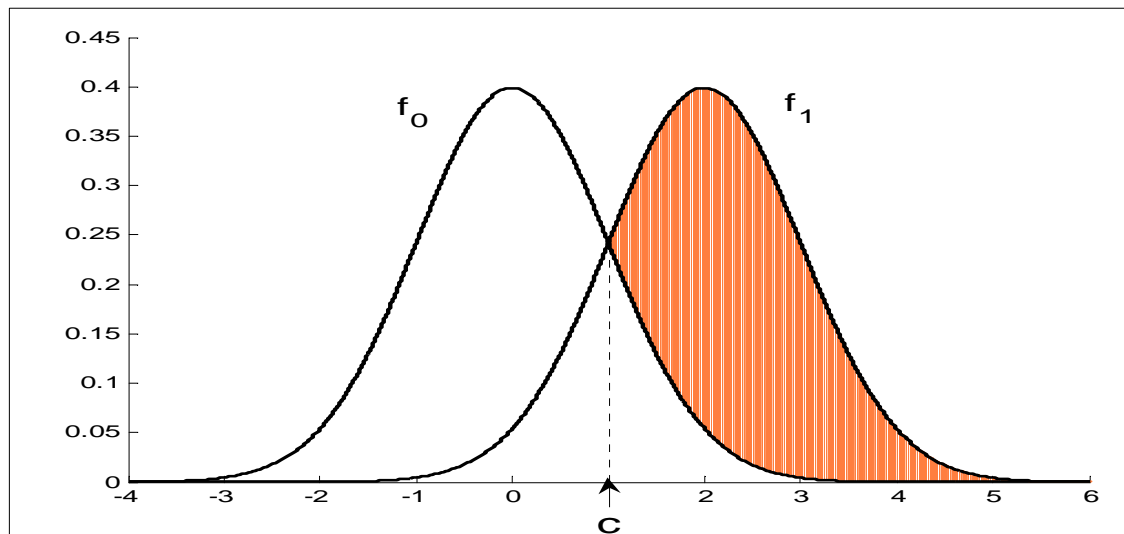
$$J(c) = q(c) + p(c) - 1$$

$$(1 - p(c), q(c))$$

$$(1 - p(c), 1 - p(c))$$

$$\sqrt{((1 - p(c)) - (1 - p(c)))^2 + (q(c) - (1 - p(c)))^2} = J(c)$$

Índice de Youden (2/3)



$$J(c) = q(c) + p(c) - 1 = \Pr(X_1 > c) + \Pr(X_0 < c) - 1$$

$$J(c) = \bar{F}_1(c) - \bar{F}_0(c)$$

$$0 = J'(c) = -f_1(c) + f_0(c)$$

TESTS DE DIAGNÓSTICO DICOTÓMICOS

TESTS DE DIAGNÓSTICO CONTINUOS

INFERENCIA DEL ÍNDICE DE YOUDEN

Método delta
Verosimilitud empírica ajustada
Estudio de simulación
Ejemplo
Conclusiones

Método delta

- Basado en el teorema de Taylor.
- Útil para aproximar momentos de transformadas de v.a.
- Schisterman y Perkins (2007) para modelo binormal y modelo bigamma.

$$v[h(X, Y)] \approx \left[\frac{\partial h}{\partial x}(\mu_x, \mu_y) \right]^2 v[X] + \left[\frac{\partial h}{\partial y}(\mu_x, \mu_y) \right]^2 v[Y] \\ + 2 \left[\frac{\partial h}{\partial x}(\mu_x, \mu_y) \right] \left[\frac{\partial h}{\partial y}(\mu_x, \mu_y) \right] \text{Cov}(X, Y)$$

Modelo binormal

$$f_i(\mu_i, \sigma_i, x) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma_i^2} (x - \mu_i)^2\right], \quad x \in \mathfrak{R}, \sigma_i > 0$$

$$J(c) = \int_c^\infty f_1(x) dx - \int_c^\infty f_0(x) dx$$

- Hay una expresión cerrada para c_{opt} .
- No es necesario tener dicha expresión.
- Se puede usar MV.

Varianza de J

$$\begin{aligned} \text{Var}[\hat{J}] \approx & \left[\frac{\partial J}{\partial \mu_0} \right]^2 \text{Var}[\hat{\mu}_0] + \left[\frac{\partial J}{\partial \mu_1} \right]^2 \text{Var}[\hat{\mu}_1] \\ & + \left[\frac{\partial J}{\partial \sigma_0} \right]^2 \text{Var}[\hat{\sigma}_0] + \left[\frac{\partial J}{\partial \sigma_1} \right]^2 \text{Var}[\hat{\sigma}_1] \end{aligned}$$

- Las parciales de J utilizan las parciales de c.
- De forma análoga se deduce una expresión para la varianza de \hat{C} .

Parciales de J

$$\frac{\partial J}{\partial \mu_i}(c) = (-1)^i \bar{F}_i(c) \frac{\mu_i}{\sigma_i^2} + (-1)^{i+1} \frac{1}{\sigma_i^2} \int_c^\infty f_i(x) x \, dx$$
$$+ \frac{\partial c}{\partial \mu_i} (f_0(c) - f_1(c))$$

$$\frac{\partial J}{\partial \sigma_i}(c) = (-1)^i \bar{F}_i(c) \left(\frac{1}{\sigma_i} - \frac{\mu_i^2}{\sigma_i^3} \right) + (-1)^{i+1} \frac{1}{\sigma_i^3} \int_c^\infty f_i(x) x^2 \, dx$$
$$+ (-1)^i \frac{2\mu_i}{\sigma_i^3} \int_c^\infty f_i(x) x \, dx + \frac{\partial c}{\partial \sigma_i} (f_0(c) - f_1(c))$$

Formas cerradas para las parciales de c

$$\frac{\partial J}{\partial c} (c_{\text{opt}}) = 0$$

$$\frac{\partial^2 J}{\partial \mu_i \partial c} (c_{\text{opt}}) + \frac{\partial^2 J}{\partial c^2} (c_{\text{opt}}) \frac{\partial c}{\partial \mu_i} (c_{\text{opt}}) = 0$$

$$\frac{\partial c}{\partial \mu_i} (c_{\text{opt}}) = \frac{-\frac{\partial^2 J}{\partial \mu_i \partial c} (c_{\text{opt}})}{\frac{\partial^2 J}{\partial c^2} (c_{\text{opt}})} = \frac{-\frac{\partial(f_0 - f_1)}{\partial \mu_i} (c_{\text{opt}})}{\frac{\partial(f_0 - f_1)}{\partial c} (c_{\text{opt}})}$$

Caso especial ($\sigma_0 = \sigma_1 = \sigma$)

$$\text{Var}[\hat{J}] \approx \left[\frac{\partial J}{\partial \mu_0} \right]^2 \text{Var}[\hat{\mu}_0] + \left[\frac{\partial J}{\partial \mu_1} \right]^2 \text{Var}[\hat{\mu}_1] \\ + \left[\frac{\partial J}{\partial \sigma} \right]^2 \text{Var}[\hat{\sigma}]$$

- Ajustar las parciales de J.
- Ajustar las parciales de c.

Modelo bigamma

$$f_i(\alpha_i, \beta_i, x) = \frac{e^{-x/\beta_i} x^{\alpha_i-1}}{\beta_i^{\alpha_i} \Gamma(\alpha_i)}, \quad x > 0, \alpha_i > 0, \beta_i > 0$$

$$\Gamma(p) = \int_0^{\infty} e^{-x} x^{p-1} dx, \quad p > 0$$

$$J(c) = \int_c^{\infty} f_1(x) dx - \int_c^{\infty} f_0(x) dx$$

- No hay una expresión cerrada para c_{opt} .
- No es necesario tener dicha expresión.
- Se puede usar MV.

Varianza de J

$$\begin{aligned}\text{Var}[\hat{J}] \approx & \left[\frac{\partial J}{\partial \alpha_0} \right]^2 \text{Var}[\hat{\alpha}_0] + \left[\frac{\partial J}{\partial \beta_0} \right]^2 \text{Var}[\hat{\beta}_0] \\ & + \left[\frac{\partial J}{\partial \alpha_1} \right]^2 \text{Var}[\hat{\alpha}_1] + \left[\frac{\partial J}{\partial \beta_1} \right]^2 \text{Var}[\hat{\beta}_1] \\ & + 2 \left[\frac{\partial J}{\partial \alpha_0} \right] \left[\frac{\partial J}{\partial \beta_0} \right] \text{Cov}(\hat{\alpha}_0, \hat{\beta}_0) \\ & + 2 \left[\frac{\partial J}{\partial \alpha_1} \right] \left[\frac{\partial J}{\partial \beta_1} \right] \text{Cov}(\hat{\alpha}_1, \hat{\beta}_1)\end{aligned}$$

Caso especial 1 ($\alpha_0 = \alpha_1 = \alpha$)

$$\begin{aligned}\text{Var}[\hat{J}] \approx & \left[\frac{\partial J}{\partial \alpha} \right]^2 \text{Var}[\hat{\alpha}] + \left[\frac{\partial J}{\partial \beta_0} \right]^2 \text{Var}[\hat{\beta}_0] + \left[\frac{\partial J}{\partial \beta_1} \right]^2 \text{Var}[\hat{\beta}_1] \\ & + 2 \left[\frac{\partial J}{\partial \alpha} \right] \left[\frac{\partial J}{\partial \beta_0} \right] \text{Cov}(\hat{\alpha}, \hat{\beta}_0) \\ & + 2 \left[\frac{\partial J}{\partial \alpha} \right] \left[\frac{\partial J}{\partial \beta_1} \right] \text{Cov}(\hat{\alpha}, \hat{\beta}_1) \\ & + 2 \left[\frac{\partial J}{\partial \beta_0} \right] \left[\frac{\partial J}{\partial \beta_1} \right] \text{Cov}(\hat{\beta}_0, \hat{\beta}_1)\end{aligned}$$

Caso especial 2 ($\beta_0 = \beta_1 = \beta$)

$$\begin{aligned}\text{Var}[\hat{J}] \approx & \left[\frac{\partial J}{\partial \alpha_0} \right]^2 \text{Var}[\hat{\alpha}_0] + \left[\frac{\partial J}{\partial \alpha_1} \right]^2 \text{Var}[\hat{\alpha}_1] + \left[\frac{\partial J}{\partial \beta} \right]^2 \text{Var}[\hat{\beta}] \\ & + 2 \left[\frac{\partial J}{\partial \alpha_0} \right] \left[\frac{\partial J}{\partial \beta} \right] \text{Cov}(\hat{\alpha}_0, \hat{\beta}) \\ & + 2 \left[\frac{\partial J}{\partial \alpha_1} \right] \left[\frac{\partial J}{\partial \beta} \right] \text{Cov}(\hat{\alpha}_1, \hat{\beta}) \\ & + 2 \left[\frac{\partial J}{\partial \alpha_0} \right] \left[\frac{\partial J}{\partial \alpha_1} \right] \text{Cov}(\hat{\alpha}_0, \hat{\alpha}_1)\end{aligned}$$

Verosimilitud empírica ajustada

- Los métodos basados en la verosimilitud son más potentes.
- Los métodos no paramétricos evitan los problemas derivados de la incorrecta especificación de un modelo paramétrico.
- Verosimilitud empírica (EL): combina ambas metodologías.
- Propuesto por Thomas y Grunkemeier (1975) para calcular mejores intervalos de confianza para el estimador de KM.
- Las principales ventajas de los IC calculados por EL es que respetan el rango del espacio de parámetros, que son invariantes frente a transformaciones y que su forma es "data-driven" (los datos hablan por sí mismos).
- EL ajustada: método propuesto para estimar cuantiles por Zhou y Jing (2003).

Distribución relativa

- Handcock y Morris (1999).
- Muy relacionada con el concepto de curva ROC.

$$R_{01}(t) = \Pr(Z \leq t) = \Pr(F_0(X_1) \leq t) = F_1(F_0^{-1}(t))$$

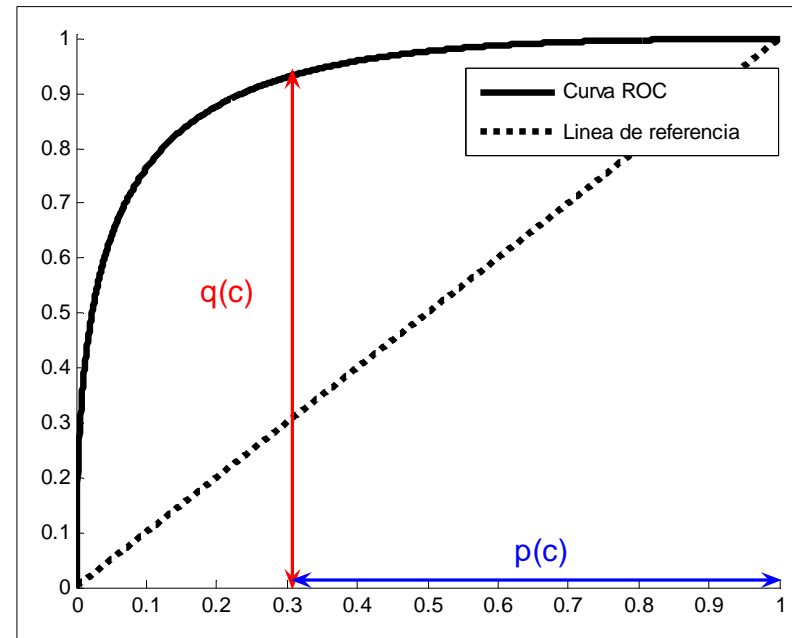
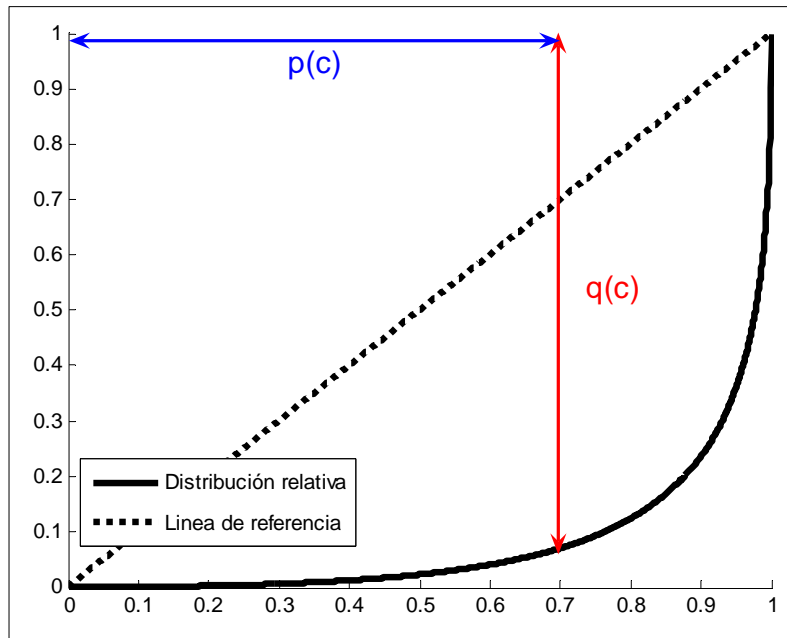
$$F_1(c) = s, F_0(c) = t \Rightarrow c \text{ es } s\text{-cuantil de } X_1$$

$c \text{ es } t\text{-cuantil de } X_0$

$$R_{01}(t) = 1 - \text{ROC}(1 - t)$$

Distribución relativa vs. curva ROC

$$R_{01}(t) = 1 - \text{ROC}(1 - t)$$



Método nuevo

- Basado en EL y bootstrapping.
- Estimación del punto de corte óptimo y del índice de Youden asociado.
- IC para ambos.
- En primer lugar se calcula el umbral, y luego como valor añadido la estimación de J.
- Cuatro pasos.

Paso 1

$$\hat{R}_{01}(t) = \frac{1}{n_1} \sum_{k=1}^{n_1} G\left(\frac{t - F_{0n_0}(X_{1k})}{h_1}\right)$$

$$G(x) = \int_{-\infty}^x K(y) dy$$

$$t_0 = \operatorname{argmax} |\hat{R}_{01}(t) - t|$$

$$\hat{R}_{10}(t), t_1$$

Paso 2

- Aplicamos el método de EL ajustado para estimar cuantiles (Zhou y Jing (2003)).
- Estimación del t_0 -ésimo cuantil de la población sana c_0 .
- Estimación del t_1 -ésimo cuantil de la población enferma c_1 .

$$\hat{L}(c) = 2n_i \left(\hat{F}_i(c) \log \frac{\hat{F}_i(c)}{t_i} + (1 - \hat{F}_i(c)) \log \frac{1 - \hat{F}_i(c)}{1 - t_i} \right)$$

$$\hat{F}_i(x) = \frac{1}{n_i} \sum_{k=1}^{n_i} G \left(\frac{x - X_{ik}}{g_i} \right)$$

Pasos 3 y 4

$$\hat{c} = \hat{c}_0 \quad \text{si} \quad \hat{J}_0 > \hat{J}_1$$

$$\hat{c} = \hat{c}_1 \quad \text{si} \quad \hat{J}_0 \leq \hat{J}_1$$

$$\hat{J}_0 = \hat{F}_0(\hat{c}_0) - \hat{F}_1(\hat{c}_0) \qquad \hat{J}_1 = \hat{F}_0(\hat{c}_1) - \hat{F}_1(\hat{c}_1)$$

- Remuestras bootstrap.
- Método del percentil para los IC.

Estudio de simulación

- MATLAB.

- 1000 ensayos considerados: $n_0=50$, $n_1=50$.

- Paso 1: kernel gaussiano, parámetros de ventana:

$$h_0 = \frac{0.9}{1.34} \min \left\{ \text{std}(F_{1n_{-1}}(X_{0k})), \text{iqr}(F_{1n_{-1}}(X_{0k})), k = 1, \dots, n_0 \right\} n_0^{-1/3}$$

$$h_1 = \frac{0.9}{1.34} \min \left\{ \text{std}(F_{0n_{-0}}(X_{1k})), \text{iqr}(F_{0n_{-0}}(X_{1k})), k = 1, \dots, n_1 \right\} n_1^{-1/3}$$

-Paso 2: kernel específico, parámetros de ventana

$$K(x) = \left\{ \frac{21 - 9\sqrt{21}}{8} x^2 + \frac{-3 + 3\sqrt{21}}{8} \right\} 1_{\{|x| \leq 1\}}, \quad g_i = n_i^{-1/2}$$

-B=299 remuestras bootstrap: kernel gaussiano, ventana:

$$\frac{0.9}{1.34} \min \left\{ \text{std}(X_{0k}), \text{iqr}(X_{0k}), k = 1, \dots, n_0 \right\} n_0^{-1/5}$$

Escenarios

	Índice de Youden J			
Forma α_1	$J=0.4$	$J=0.6$	$J=0.8$	$J=0.9$
$\alpha_1=1.5$	2.4828	4.3565	9.7847	19.8020
$\alpha_2=2.0$	1.6622	2.7650	5.6517	10.3842

$$\alpha_0 = 1.5, \beta_0 = 1$$

Resultados: c

		ELM		Método delta	
α_1	J	Cobertura (%)	Anchura	Cobertura (%)	Anchura
1.5	0.4	89.60	1.3335	94.30	1.1748
1.5	0.6	96.50	1.7301	93.30	0.9998
1.5	0.8	98.30	2.4271	93.70	1.3767
1.5	0.9	94.20	2.9930	94.00	1.9729
2.0	0.4	91.70	1.1244	95.30	1.7671
2.0	0.6	96.20	1.4935	97.50	1.1625
2.0	0.8	96.70	2.1673	95.50	1.4250
2.0	0.9	95.40	2.7118	95.70	1.9728

Resultados: J

		ELM		Método delta	
α_1	J	Cobertura (%)	Anchura	Cobertura (%)	Anchura
1.5	0.4	94.50	0.2988	94.20	0.2773
1.5	0.6	94.00	0.2519	93.90	0.2445
1.5	0.8	92.90	0.1785	92.50	0.1793
1.5	0.9	90.40	0.1222	90.80	0.1218
2.0	0.4	94.40	0.3045	93.10	0.2801
2.0	0.6	93.80	0.2573	93.10	0.2485
2.0	0.8	91.50	0.1837	91.90	0.1800
2.0	0.9	89.90	0.1277	90.50	0.1215

Ejemplo

- Le (2006).
- 53 pacientes con cáncer de próstata: 20 con nódulos infectados y 33 sin nódulos infectados.
- Biomarcador: nivel de “acid phosphatase” en sangre (x100).

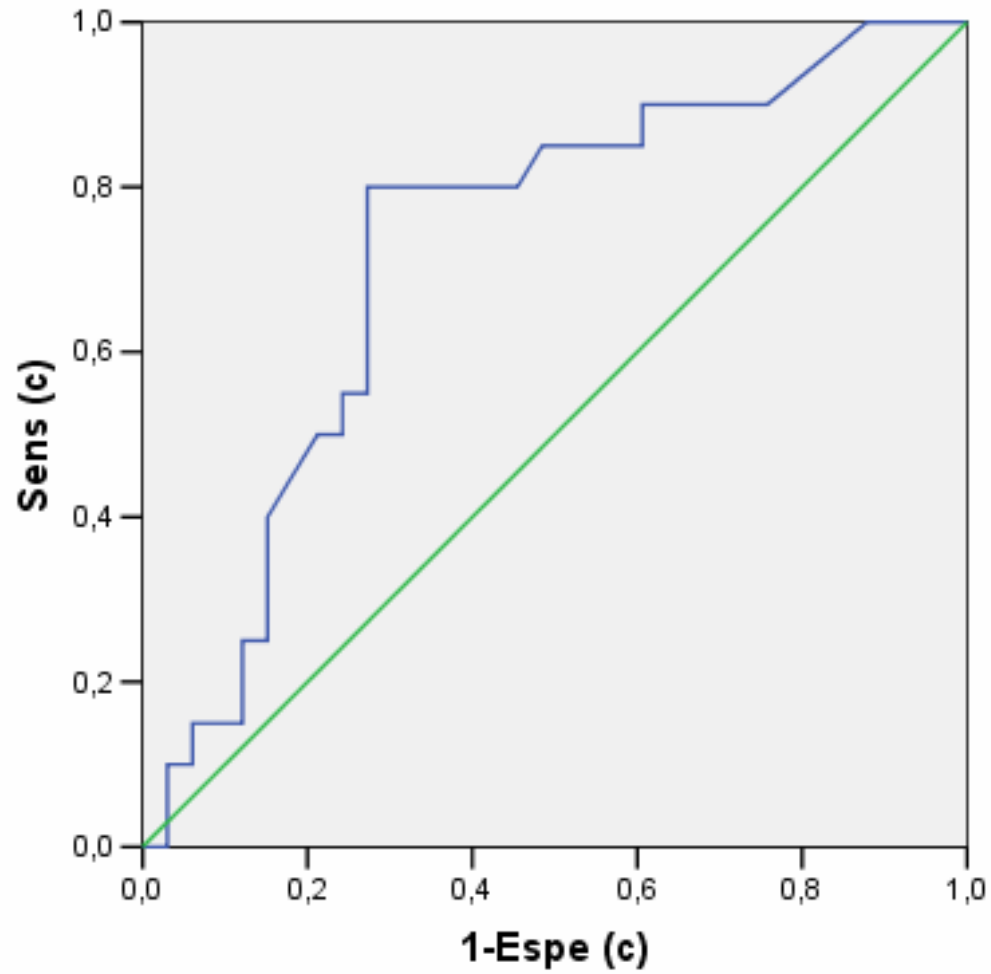
Con nódulos infectados:

48, 49, 51, 56, 67, 67, 67, 70, 70, 72, 76, 78, 81, 82, 82, 84, 89, 99, 126, 136.

Sin nodos infectados:

40, 40, 46, 47, 48, 48, 49, 49, 50, 50, 50, 50, 50, 52, 52, 55, 55, 56, 59, 62, 62, 63, 65, 66, 71, 75, 76, 78, 83, 95, 98, 102, 187.

Resultados (1/2)



Resultados (2/2)

- Nuestro método (sin asunciones paramétricas):

Umbral=60.67, IC95%=(51.40, 67.50).

- Método delta:

No aplicable.

- Le (Alternativas de Lehmann):

Umbral=75.00.

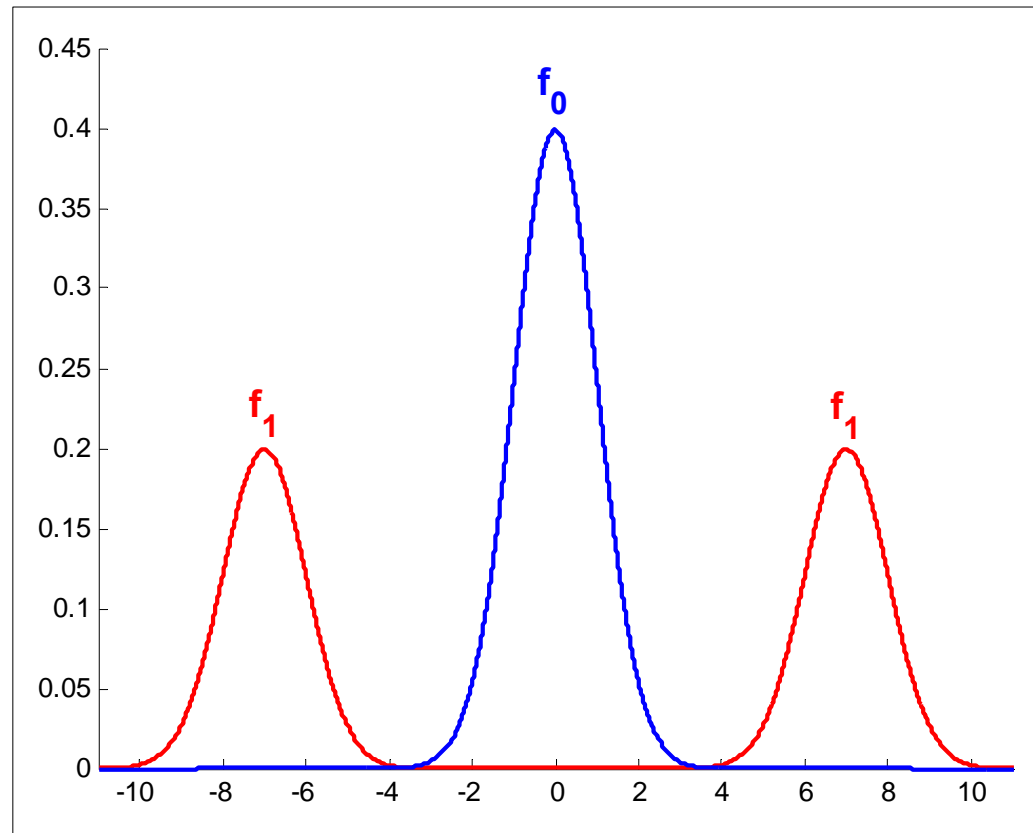
Conclusiones

- El nuevo IC tiene buen comportamiento en términos de la cobertura y de la anchura.
- Competitivo con el método delta.
- No requiere de asunciones paramétricas.
- Mayor coste computacional en términos de tiempo que el método delta.

Trabajo futuro (1/2)

- LOD: límite de detección.
- Datos concentrados en torno a un valor (“spiked data”).
- Muestras combinadas.
- Covariables.

Trabajo futuro (2/2)



Referencias

- Le, C.T. A solution for the most basic optimization problem associated with an ROC curve. *Statistical Methods in Medical Research* **2006**, *15*, 571-584.
- Letón, E.; Molanes, E.M. *Statistics and Econometrics Series* **2009**, *07*, Working Paper 09-19.
- Pepe, M.S. *The statistical evaluation of medical tests for classification and prediction*. Oxford University Press, New York, **2003**.
- Schisterman, E.F.; Perkins, N.J. Confidence intervals for the Youden index and corresponding optimal cut-point. *Communications in Statistics – Simulation and Computation* **2007**, *36*, 549-563.