

Tema 2

Construcción de curvas ROC

Contenidos

- Elección del punto de corte conocidos la Sensibilidad y la Especificad. Curva ROC
- Caso 1.- Construcción de la curva ROC en el caso de distribuciones normales.
- Caso 2.- Construcción de la curva ROC con SPSS conocidos los resultados en dos pruebas diagnósticas.

Elección del punto M, basándose en la Sensibilidad y la Especificidad: curvas ROC

El valor de M clasifica a los individuos como sano o enfermos.

Valor máximo de colesterol, por encima del cual deberemos establecer una dieta o un tratamiento

Ejemplo

En un estudio se incluyó a 2.641 pacientes con sospecha de cáncer prostático que acudieron a una consulta de Urología durante un periodo de tiempo determinado. Durante su exploración, se recogió el resultado del tacto rectal realizado a cada uno de estos pacientes, según fuese éste normal o anormal, y se contrastó con el posterior diagnóstico obtenido de la biopsia prostática.

		Resultado de biopsia prostática		
		Cáncer	Patología benigna	Total
Resultado tacto rectal	Anormal	634	269	903
	Normal	487	1251	1738
	Total	1121	1520	2641

Especificidad=0.82

Sensibilidad=0.56

¿Está bien escogido
el valor de M?

Caso 1. Construcción de la curva ROC en el caso de distribuciones normales. (Modelo teórico)

- Es una distribución continua.
- Es la más frecuente.
- Su función de densidad tiene forma de campana (campana de Gauss).
- Es simétrica.
- Viene identificada por el valor de su esperanza, μ y de su varianza σ^2 .



Peso

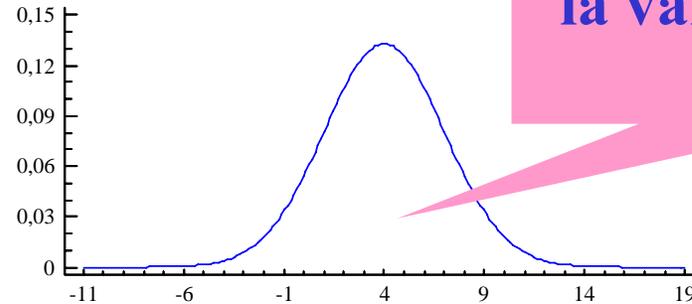


Alturas

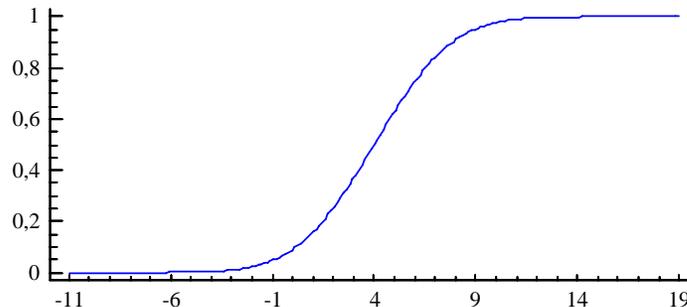
Distribución Normal $N(\mu, \sigma)$

Más concentración de individuos cerca de μ ; individuos con valores de la variable lejos de μ son los raros.

Función de densidad

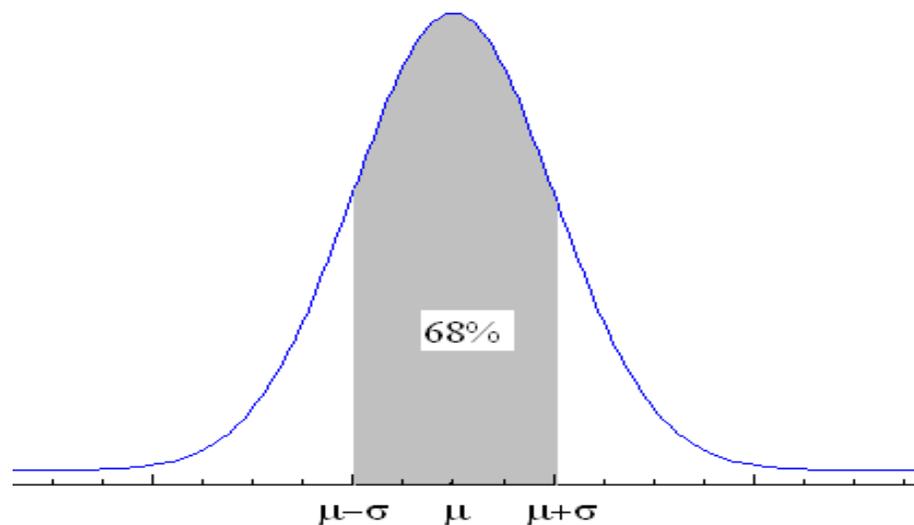


Función de distribución



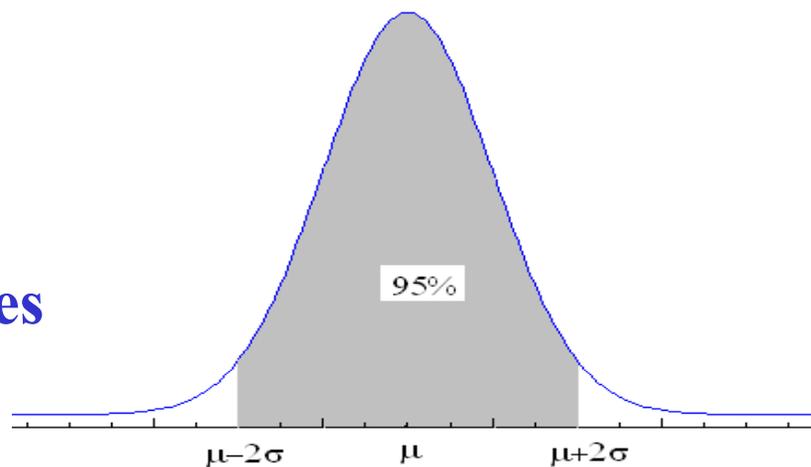
Distribución Normal $N(\mu, \sigma)$

- Entre la media y una desviación típica tenemos siempre la misma probabilidad: aprox. **68%**



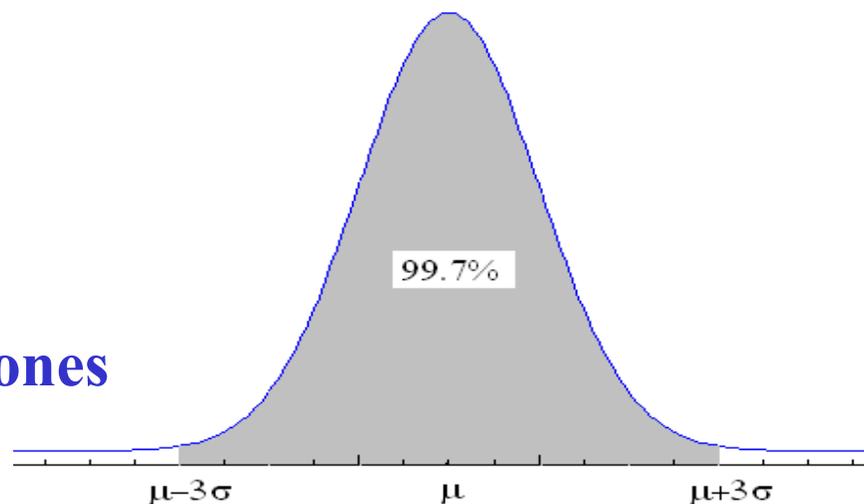
Distribución Normal $N(\mu, \sigma)$

- Entre la media y una desviación típica tenemos siempre la misma probabilidad: **aprox. 68%**
- Entre la media y dos desviaciones típicas **aprox. 95%**



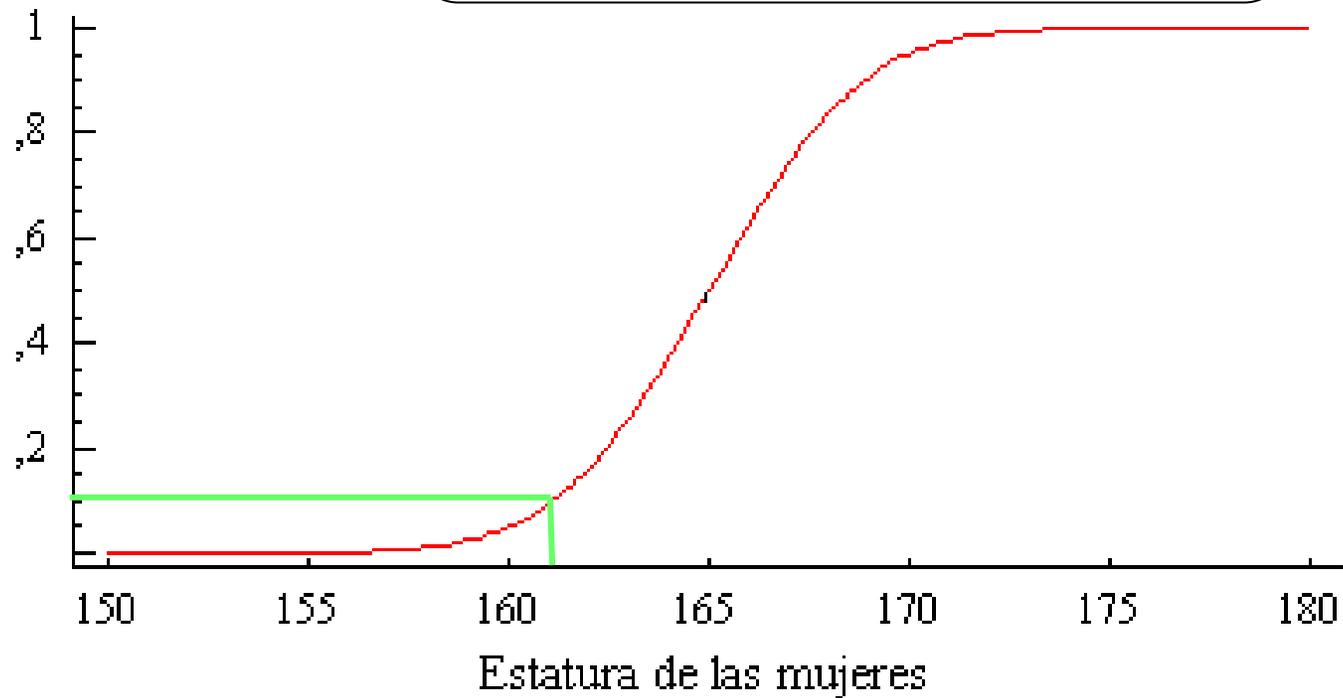
Distribución Normal $N(\mu, \sigma)$

- Entre la media y una desviación típica tenemos siempre la misma probabilidad: **aprox. 68%**
- Entre la media y dos desviaciones típicas **aprox. 95%**
- Entre la media y tres desviaciones típicas **aprox. 99%**



Función de distribución $F(x)$

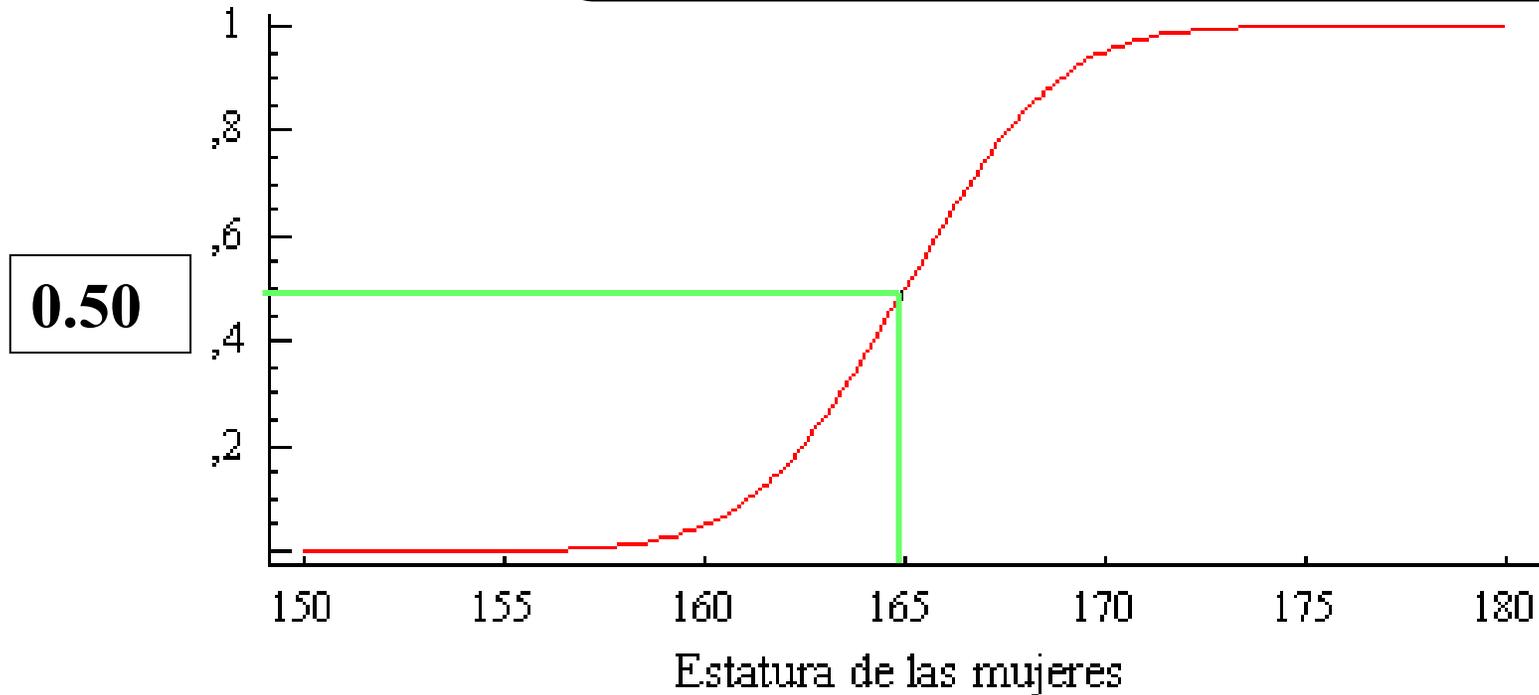
Una estatura de 161
¿A qué percentil corresponde?



0.1

Inversa de la Función de distribución

Teniendo en cuenta la altura de las mujeres
¿Cuál es el percentil 50?



Caso 1. Construcción de la curva ROC en el caso de distribuciones normales.

Ejemplo

El diámetro, en micras, de los hematíes de individuos normales sigue una $N(7.5, 0.2)$ y el de los individuos cirróticos sigue una distribución $N(8.5, 0.6)$ y se conviene en clasificar sólo erróneamente a un 5% de los normales como cirróticos.

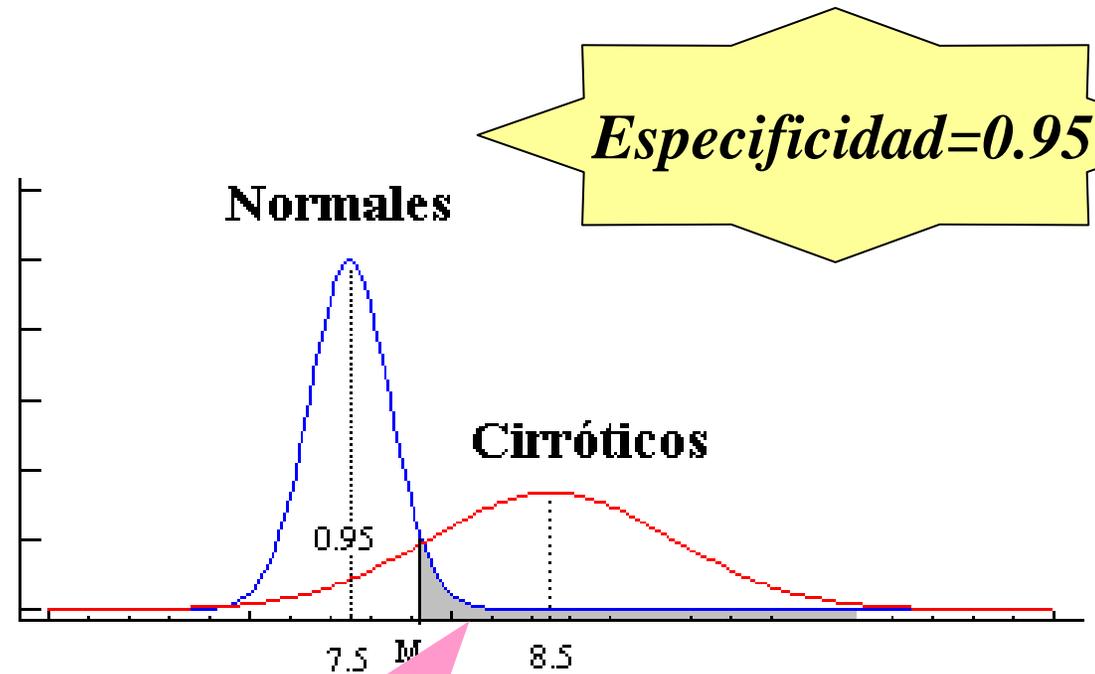
Caso 1. Construcción de la curva ROC en el caso de distribuciones normales.

Ejemplo

Normales $\sim N(7.5, 0.2)$

Cirróticos $\sim N(8.5, 0.6)$

se conviene en clasificar erróneamente a un 5% de los normales.



Probabilidad de considerar enfermo a un sano

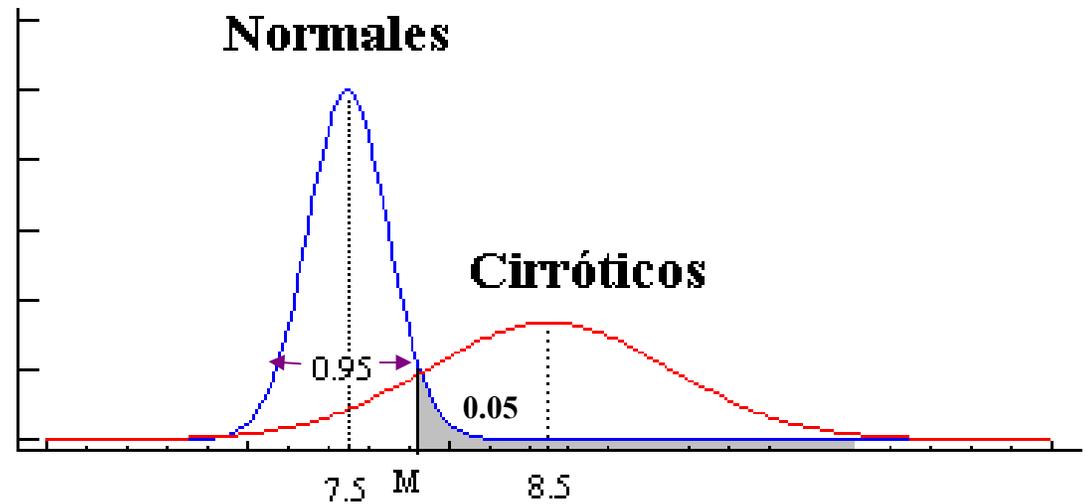
Caso 1. Construcción de la curva ROC en el caso de distribuciones normales.

Ejemplo

Normales $\sim N(7.5, 0.2)$

Cirróticos $\sim N(8.5, 0.6)$

se conviene en clasificar erróneamente a un 5% de los normales.



¿qué valor determina que un individuo pase a ser considerado cirrótico?

Distribuciones con SPSS

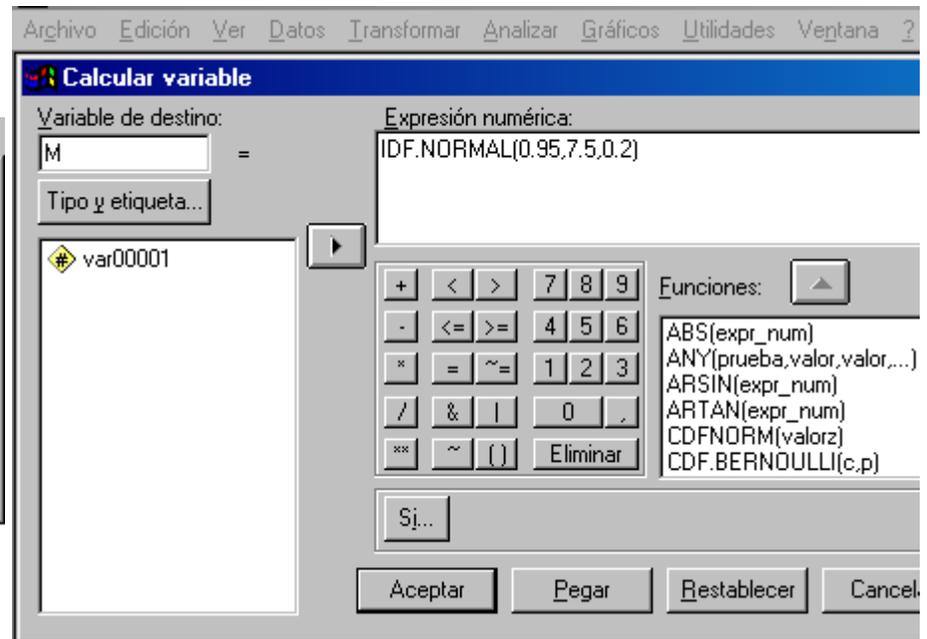
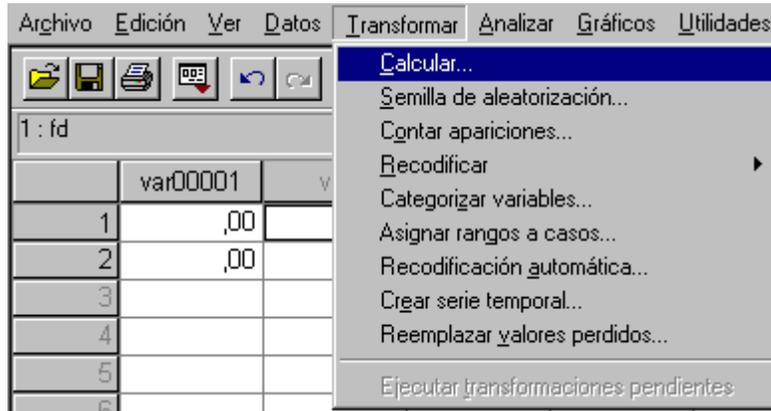
Cálculo del punto de la distribución $N(7.5,0.2)$ que acumula una probabilidad 0.95

Qué buscamos:

- ¿El punto que acumula una probabilidad fija? → Inversa de la FD**
- ¿La probabilidad acumulada en un punto fijo? → FD**

Distribuciones con SPSS

Cálculo del punto de la distribución $N(7.5,0.2)$ que acumula una probabilidad 0.95



$M=7.82$

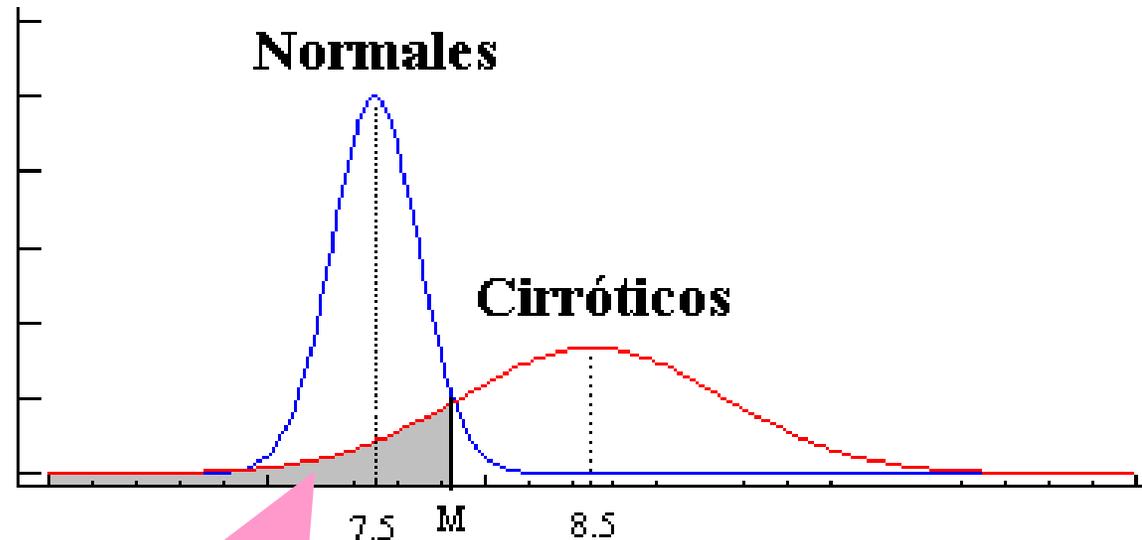
Caso 1. Construcción de la curva ROC en el caso de distribuciones normales.

Ejemplo

Normales $\sim N(7.5, 0.2)$

Cirróticos $\sim N(8.5, 0.6)$

se conviene en clasificar erróneamente a un 5% de los normales.



Probabilidad de considerar sano a un enfermo

Distribuciones con SPSS

Cálculo de la probabilidad obtener valores menores o iguales que 7.83 en una distribución $N(8.5,0.6)$

Qué buscamos:

¿El punto que acumula una probabilidad fija? → Inversa de la FD

¿La probabilidad acumulada en un punto fijo? → FD

Distribuciones con SPSS

Cálculo de la probabilidad obtener valores menores o iguales que 7.83 en una distribución $N(8.5,0.6)$

The screenshot shows the 'Calcular variable' (Calculate Variable) dialog box in SPSS. The 'Variable de destino' (Destination Variable) is 'FD'. The 'Expresión numérica' (Numerical Expression) is 'CDF.NORMAL(7.83,8.5,0.6)'. The 'Funciones' (Functions) list includes 'CDF.NORMAL'. A yellow starburst graphic highlights the value '0.13'.

Variable de destino: FD = Expresión numérica: CDF.NORMAL(7.83,8.5,0.6)

Funciones: CDF.NORMAL

0.13

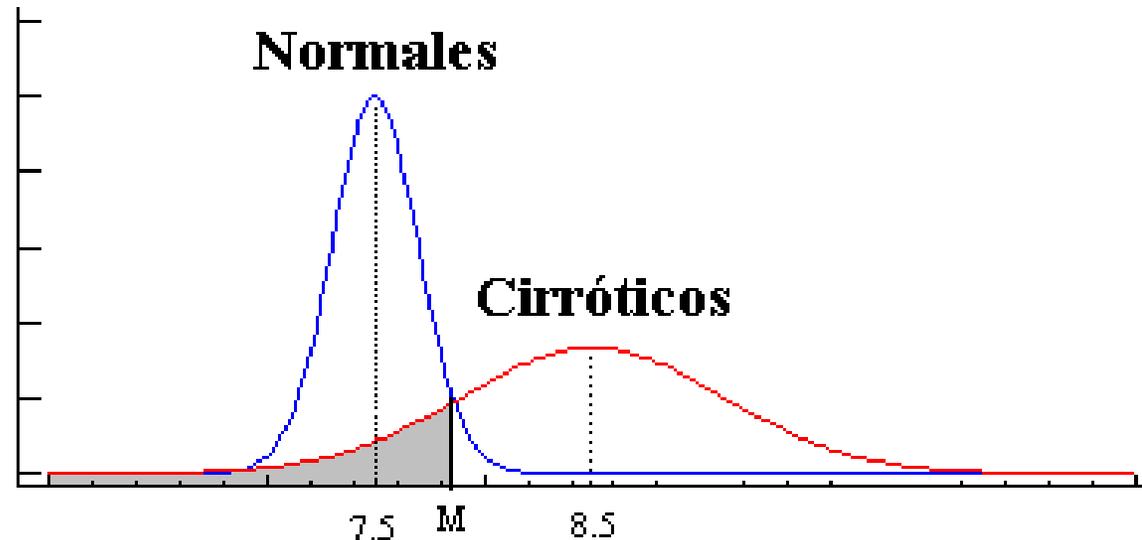
Caso 1. Construcción de la curva ROC en el caso de distribuciones normales.

Ejemplo

Normales $\sim N(7.5, 0.2)$

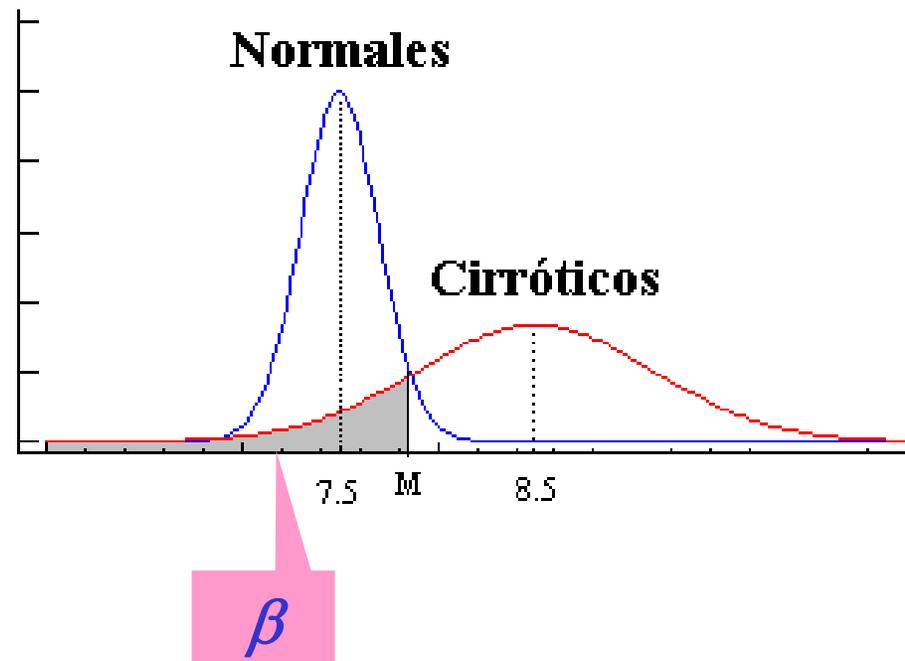
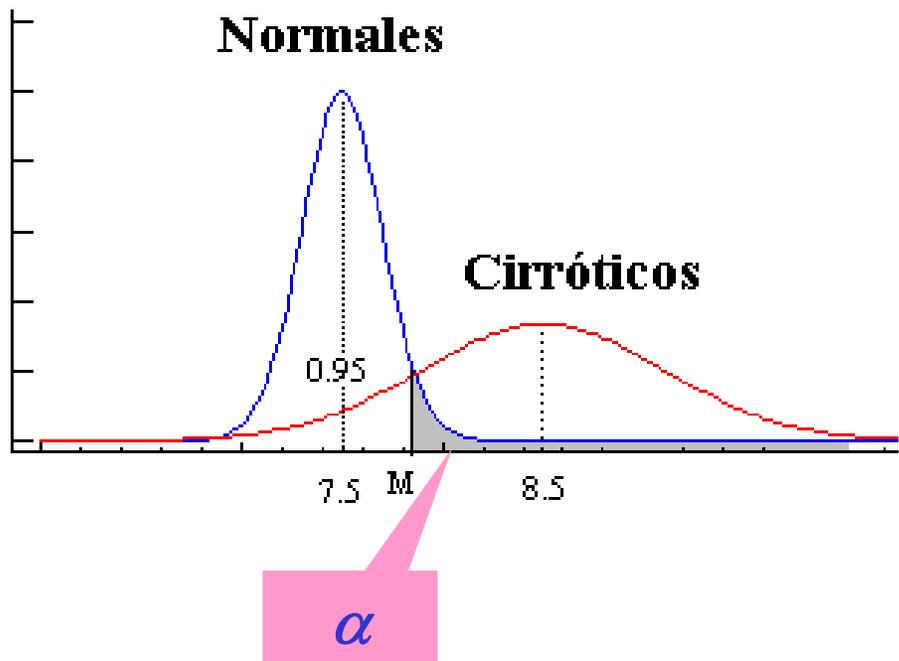
Cirróticos $\sim N(8.5, 0.6)$

se conviene en clasificar erróneamente a un 5% de los normales.



¿cuántos individuos cirróticos son declarados normales?

Caso 1. Construcción de la curva ROC en el caso de distribuciones normales.



100 α % el porcentaje de falsos positivos (en ej. 5%)

100 β % el porcentaje de falsos negativos (el 13.1%)

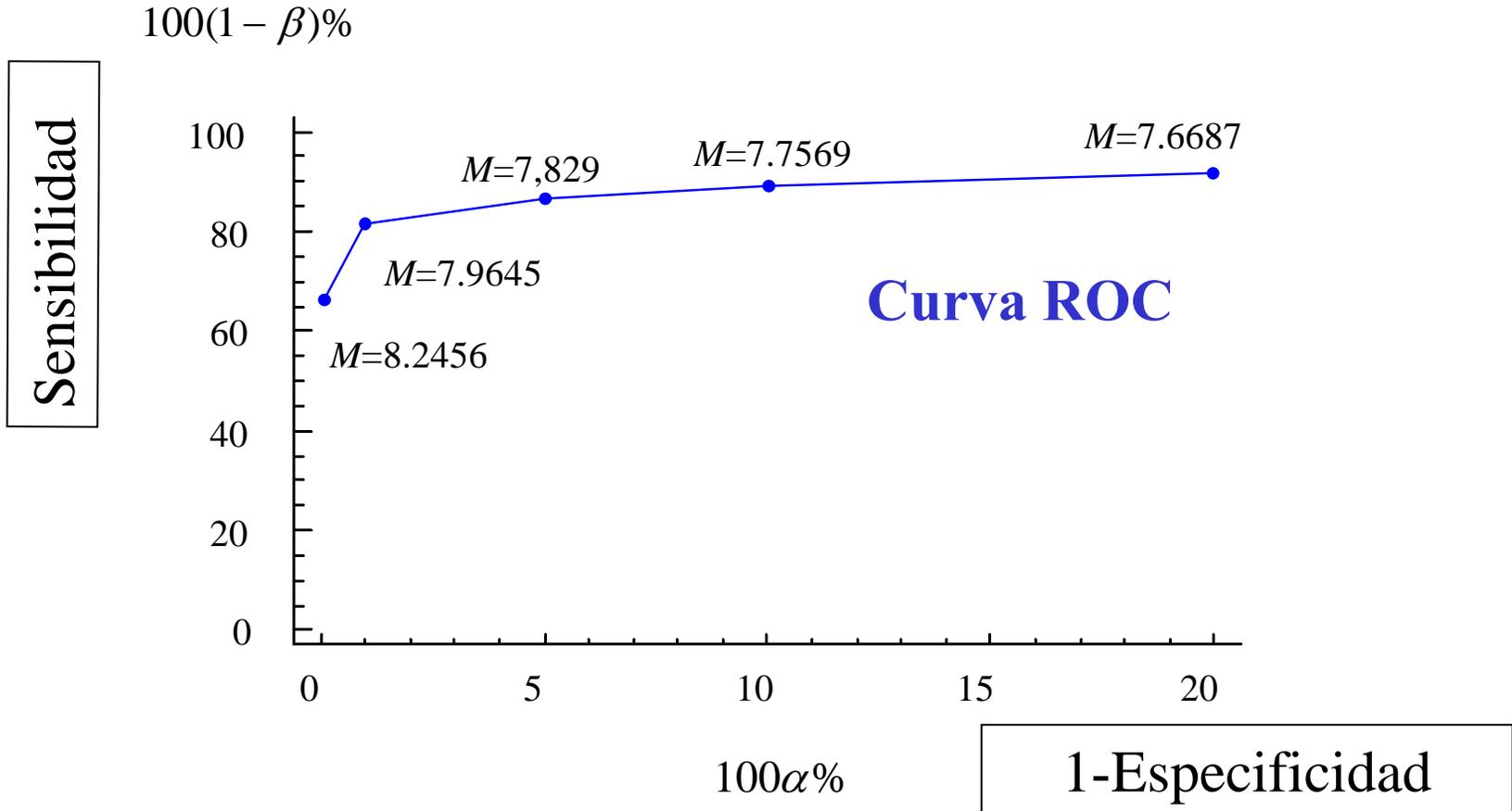
Caso 1. Construcción de la curva ROC en el caso de distribuciones normales.

Ejemplo

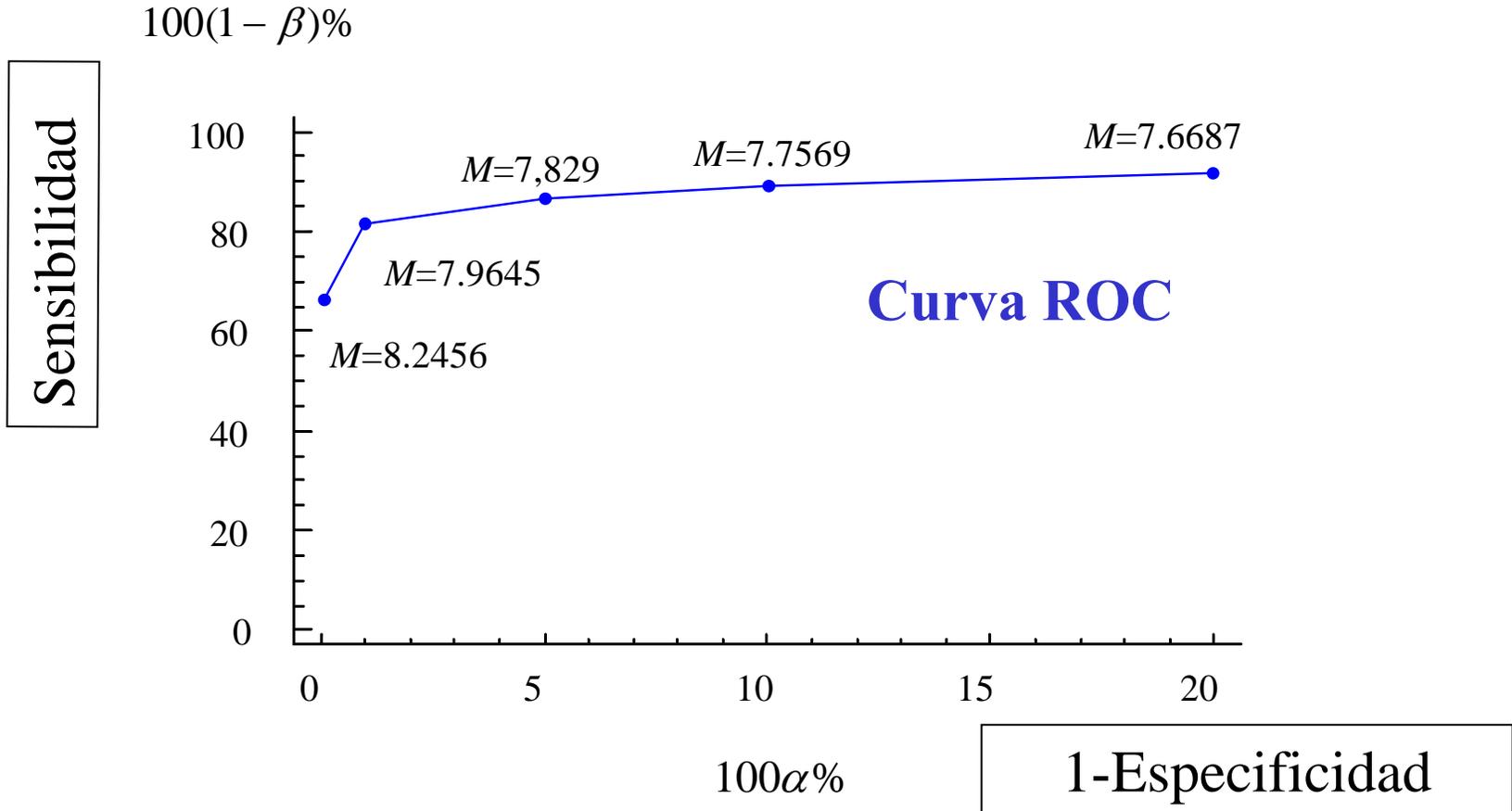
El diámetro, en micras, de los hematíes de individuos normales sigue una $N(7.5,0.2)$ y el de los individuos cirróticos sigue una distribución $N(8.5, 0.6)$ y se conviene en clasificar sólo erróneamente a un 5% de los normales como cirróticos.

% de individuos sanos clasificados como cirróticos	$100\alpha\%$	1	5	10	20
Punto de corte	M		7.82		
% de individuos cirróticos clasificados correctamente	$100(1-\beta)\%$		86.83		

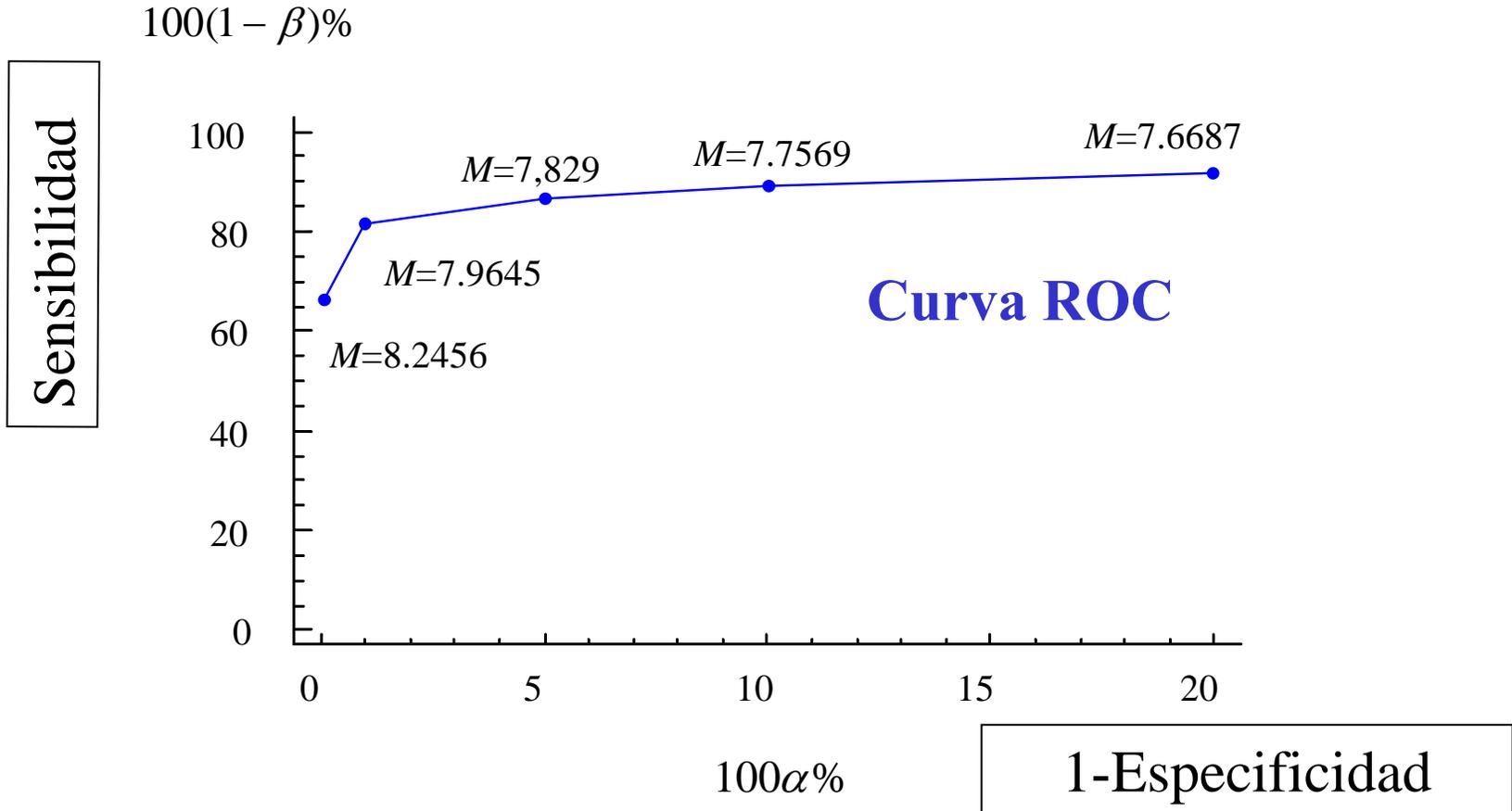
¿El valor $M=7.6687$ qué porcentaje de individuos sanos clasificaría como cirróticos?



*¿y qué porcentaje de individuos
cirróticos clasificaría como sanos?*



¿Cuál es el valor óptimo para M ?



Sensibilidad de la prueba

Interesa que sea alto cuando la enfermedad es grave y es especialmente importante detectarla, bien porque admita tratamiento, bien porque exista la posibilidad de que se transmita a otros individuos, como puede ser el VIH.

¿cuándo se aplicará una prueba con alta sensibilidad?

Si es peor clasificar a un enfermo como sano que a un sano como enfermo.

Especificidad de la prueba (*E*):

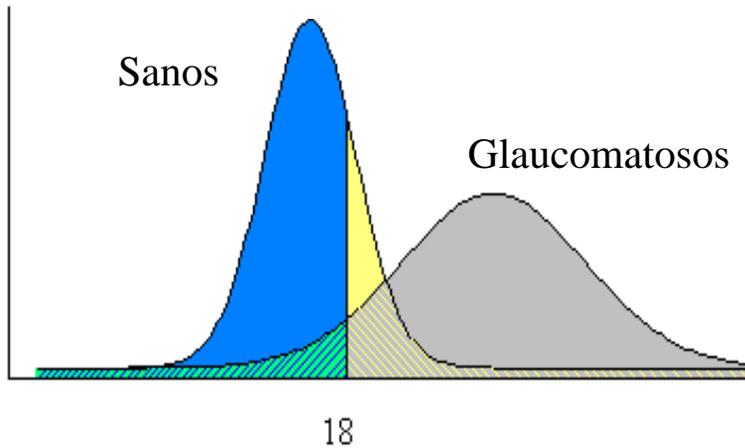
Tiene por objeto confirmar que alguien no presenta la enfermedad. Este valor interesa que sea alto, bien cuando la enfermedad es seria y difícilmente tratable, o bien cuando el hecho de saber que no se tiene la enfermedad posee una gran importancia sanitaria y psicológica.

No es posible maximizar Especificidad y Sensibilidad al mismo tiempo.

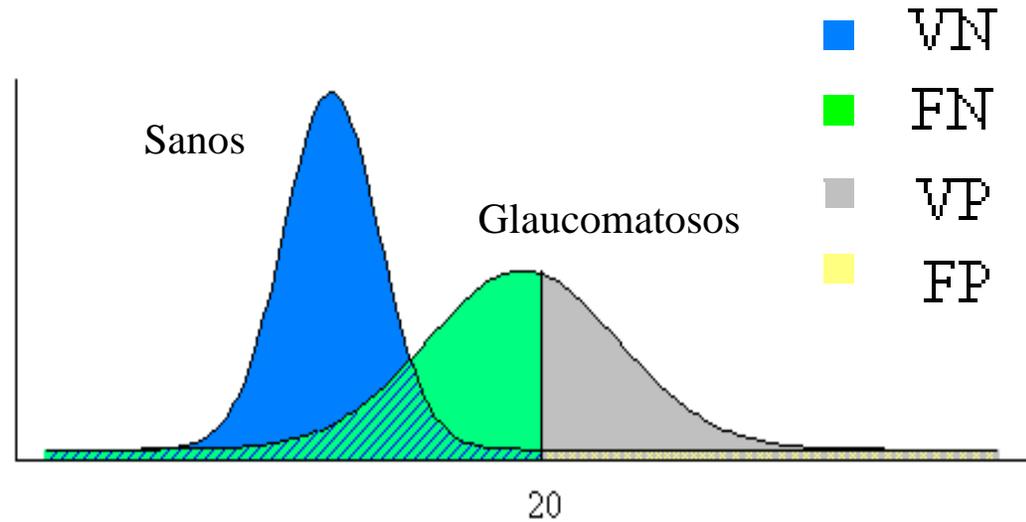
Ejemplo

Se quiere estudiar la capacidad diagnóstica de la tonometría ocular en el diagnóstico del glaucoma. Representamos las dos curvas de tensiones oculares, la de los no glaucomatosos y la de los glaucomatosos. Se observa un solapamiento de las curvas porque existen individuos con glaucoma normotensional y, por otra parte, también hay sujetos que nunca tendrán glaucoma a pesar de presentar unos valores elevados de tensión ocular.

Inicialmente $M=18$



Se desplaza el corte a $M=20$



¿Cuándo es mayor la sensibilidad?

¿Cuándo es mayor la especificidad?

Caso 2.- Construcción de la curva ROC con SPSS

conocidos los resultados en dos pruebas diagnósticas
(Caso práctico)

Ejemplo

El grado de afección abdominal en enfermos de Hodking se puede predecir a través de un modelo de regresión logística sin necesidad de recurrir a la cirugía. Disponemos de los siguientes datos:

Probabilidad de padecer afección abdominal (M)	Cirugía	
	La padecen	No la padecen
Menor que 0.1	15	155
Menor que 0.2	20	342
Menor que 0.3	29	454
Menor que 0.4	51	602
Menor que 0.5	72	609
Menor que 0.6	81	630
Menor que 0.7	123	685
Menor que 0.8	124	693
Menor que 0.9	153	699
Menor que 1	300	700

Es decir fijando como criterio:

Si se obtiene $p < 0.1$, considero que no la padece

Si se obtiene $p \geq 0.1$, considero que sí la padece

Probabilidad de padecer afección abdominal (M)	Cirugía	
	La padecen	No la padecen
Menor que 0.1	15	155
Mayor o igual que 0.1	285	545
Total	300	700

Sensibilidad=0.95

Especificidad=0.22

Curvas ROC con SPSS

Primer caso: tabla de frecuencia, construimos clases disjuntas

Probabilidad de padecer afección abdominal (M)	Cirugía	
	La padecen	No la padecen
Menor que 0.1	15	155
Entre 0.1 y 0.2	5	187
Entre 0.2 y 0.3	9	112
Entre 0.3 y 0.4	22	148
Entre 0.4 y 0.5	21	7
Entre 0.5 y 0.6	9	21
Entre 0.6 y 0.7	42	55
Entre 0.7 y 0.8	1	8
Entre 0.8 y 0.9	29	6
Entre 0.9 y 1	147	1

Curvas ROC con SPSS

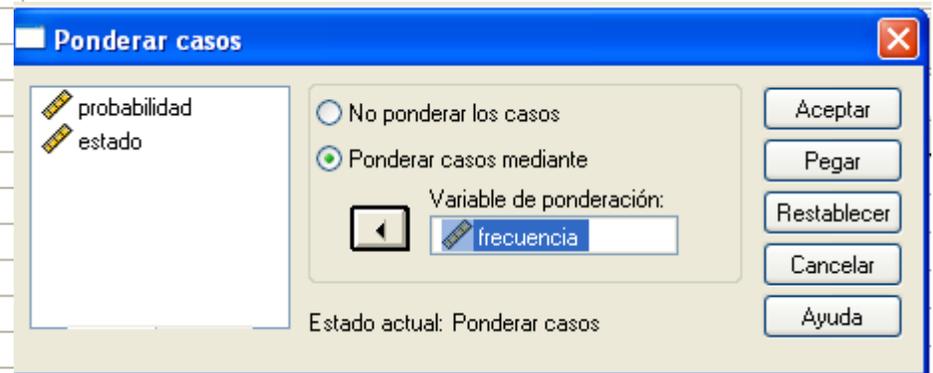
*frecuencias.sav [Conjunto_de_datos1] - Editor de datos SPSS

Archivo Edición Ver Datos Transformar Analizar Gráficos Utilidades Ventana ?

7 :

	probabilidad	frecuencia	estado	var
1	Menor de 0.1	15,00	Enfermo	
2	Entre 0.1 y 0.2	5,00	Enfermo	
3	Entre 0.2 y 0.3	9,00	Enfermo	
4	Entre 0.3 y 0.4	22,00	Enfermo	
5	Entre 0.4 y 0.5	21,00	Enfermo	
6	Entre 0.5 y 0.6	9,00	Enfermo	
7	Entre 0.6 y 0.7	42,00	Enfermo	
8	Entre 0.7 y 0.8	1,00	Enfermo	
9	Entre 0.8 y 0.9	29,00	Enfermo	
10	Entre 0.9 y 1	147,00	Enfermo	
11	Menor de 0.1	155,00	Sano	
12	Entre 0.1 y 0.2	187,00	Sano	
13	Entre 0.2 y 0.3	112,00	Sano	
14	Entre 0.3 y 0.4	148,00	Sano	
15	Entre 0.4 y 0.5	7,00	Sano	
16	Entre 0.5 y 0.6	21,00	Sano	
17	Entre 0.6 y 0.7	55,00	Sano	
18	Entre 0.7 y 0.8	8,00	Sano	
19	Entre 0.8 y 0.9	6,00	Sano	
20	Entre 0.9 y 1	1,00	Sano	

Datos → Ponderar casos



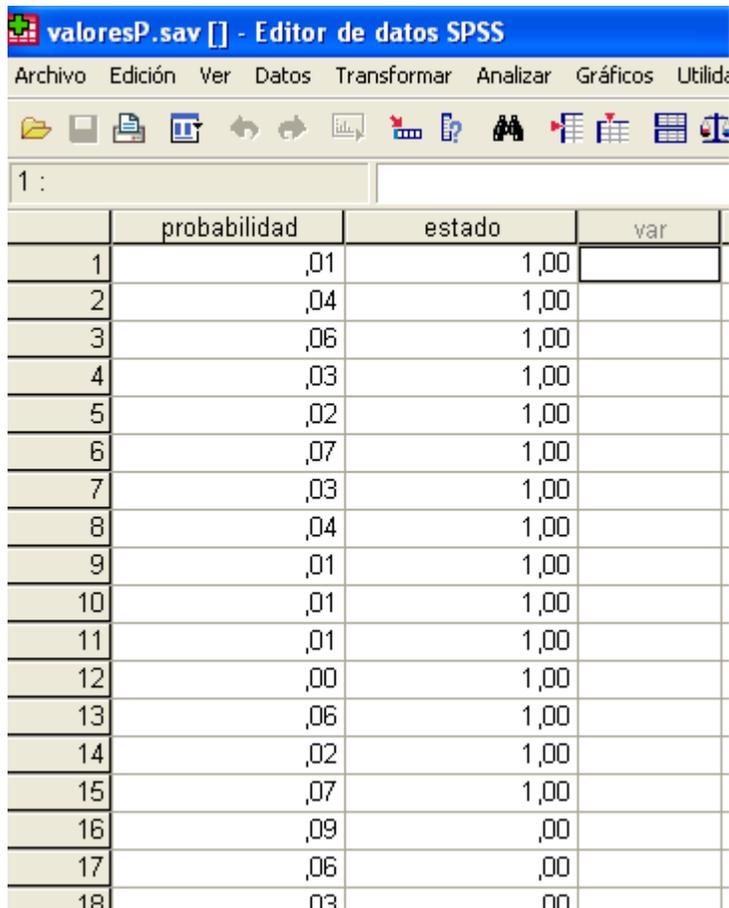
Curvas ROC con SPSS

Analizar → Curva COR



Curvas ROC con SPSS

Segundo caso: valores de los pacientes



valoresP.sav [] - Editor de datos SPSS

Archivo Edición Ver Datos Transformar Analizar Gráficos Utilida

1 :

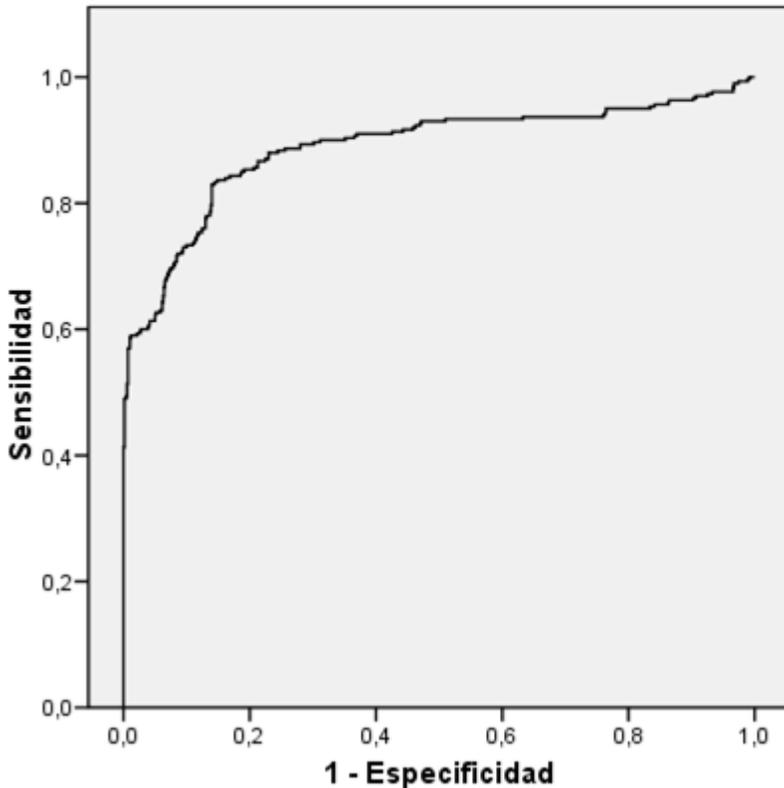
	probabilidad	estado	var
1	,01	1,00	
2	,04	1,00	
3	,06	1,00	
4	,03	1,00	
5	,02	1,00	
6	,07	1,00	
7	,03	1,00	
8	,04	1,00	
9	,01	1,00	
10	,01	1,00	
11	,01	1,00	
12	,00	1,00	
13	,06	1,00	
14	,02	1,00	
15	,07	1,00	
16	,09	,00	
17	,06	,00	
18	,03	,00	

Analizar → Curva COR



Curvas ROC con SPSS

Curva COR



Área bajo la curva

Variables resultado de contraste: probabilidad

Área	Error típ. ^a	Sig. asintótica ^b	Intervalo de confianza asintótico al 95%	
			Límite inferior	Límite superior
,889	,014	,000	,862	,916

a. Bajo el supuesto no paramétrico
b. Hipótesis nula: área verdadera = 0,5

Área bajo la curva:
Seleccionamos al azar dos individuos, uno sano y uno enfermo, probabilidad de que la prueba de un valor más alto en el enfermo que en el sano

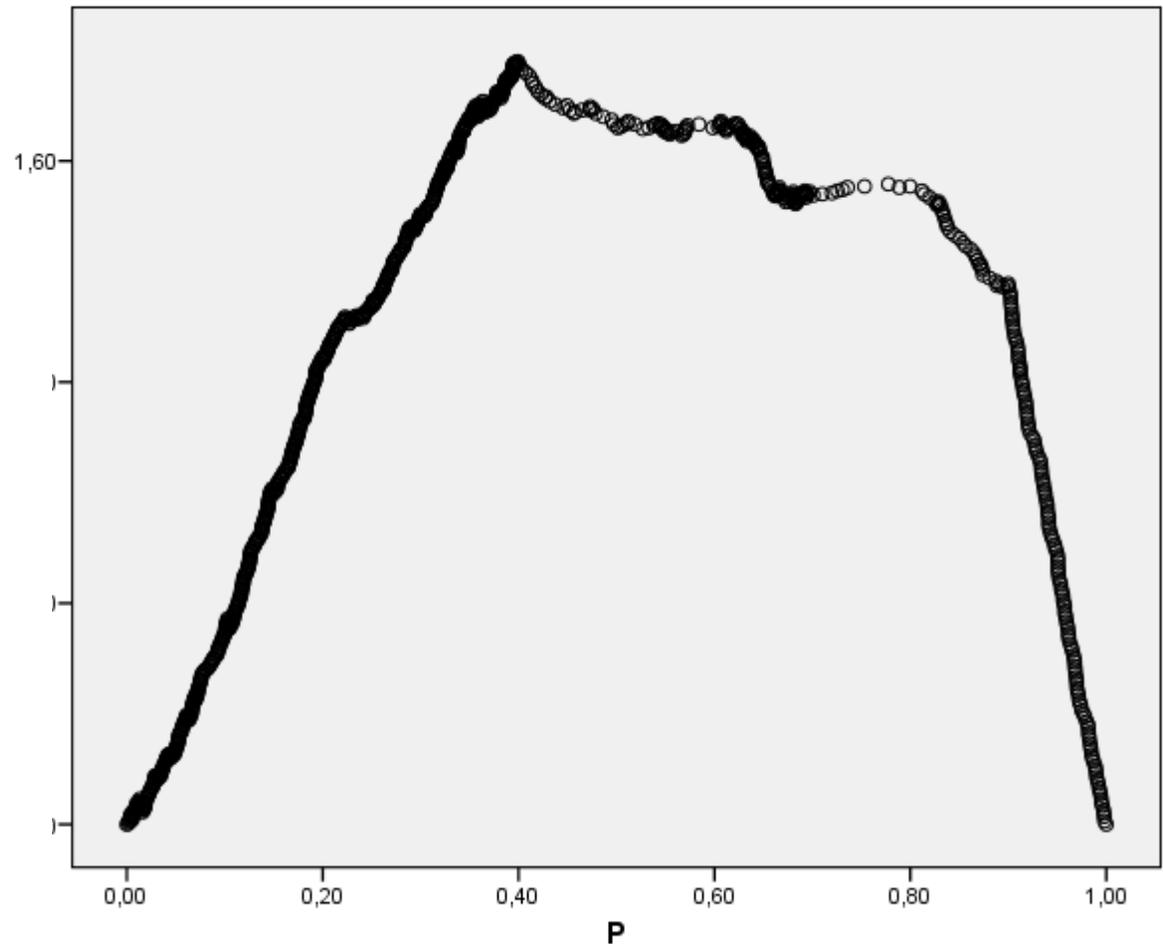
Es mejor hacerle la prueba que tirar una moneda

Curvas ROC con SPSS

Coordenadas de la curva

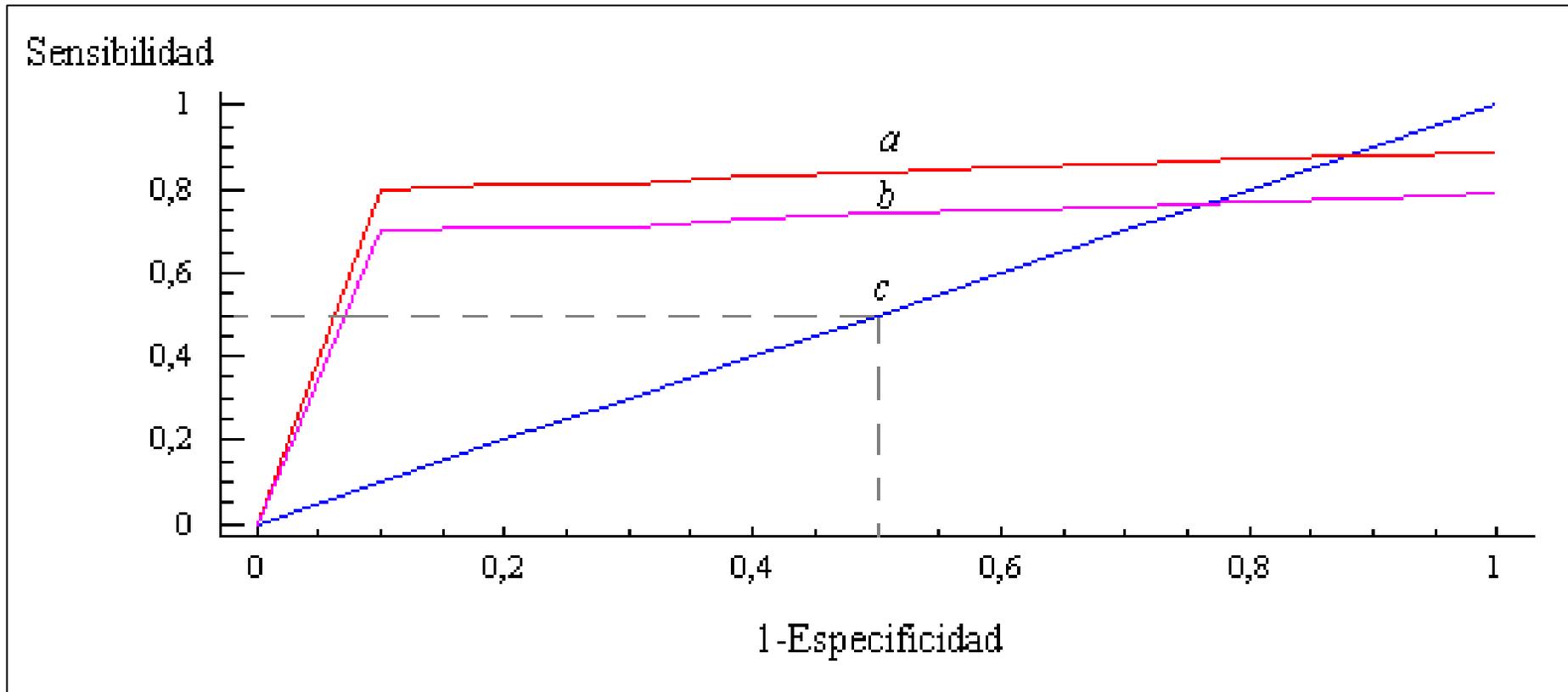
Variables resultado de contraste: probabilidad

Positivo si es mayor o igual que ^a	Sensibilidad	1 - Especificidad
,0000	1,000	1,000
,0014	1,000	,999
,0020	1,000	,997
,0031	1,000	,996
,0037	1,000	,994
,0038	1,000	,993
,0041	1,000	,991
,0045	,997	,991
,0048	,997	,990
,0050	,997	,989
,0053	,993	,989
,0056	,993	,987



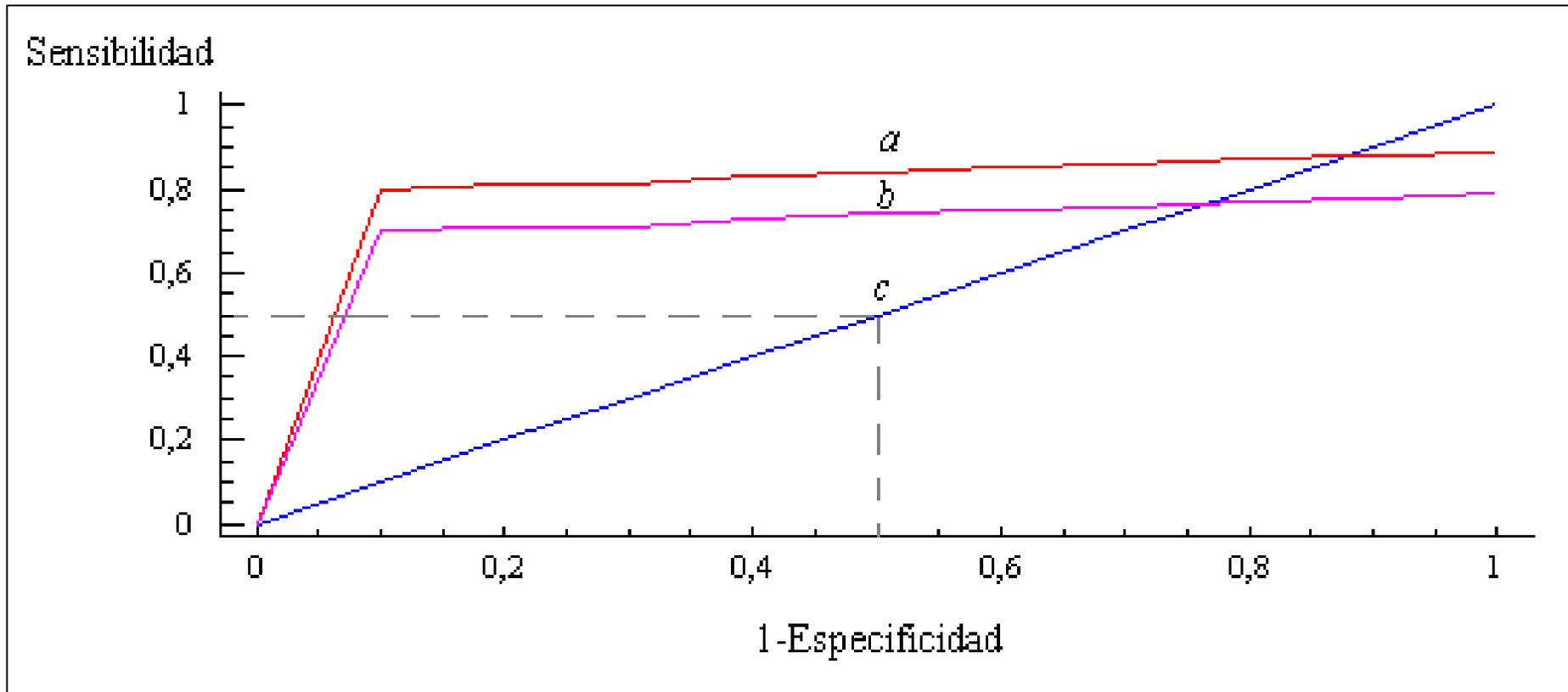
Para determinar el mejor punto de corte necesitamos conocer los costes de los FP y FN.

Comparación de curvas



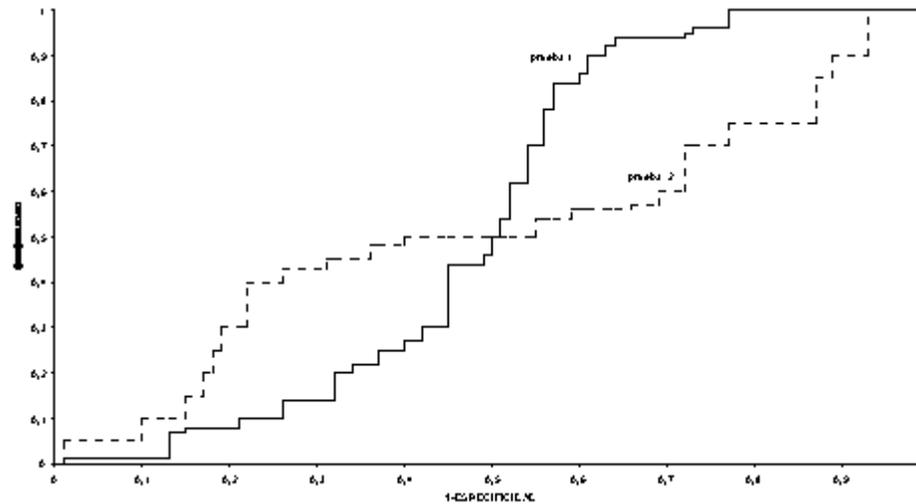
¿Cuándo es la mejor prueba de las tres?

Comparación de curvas



¿Cuándo es la peor prueba de las tres?

Comparación de curvas



¿Cuándo es la mejor prueba?

Curvas ROC frente a Curvas DET

Martín et al.

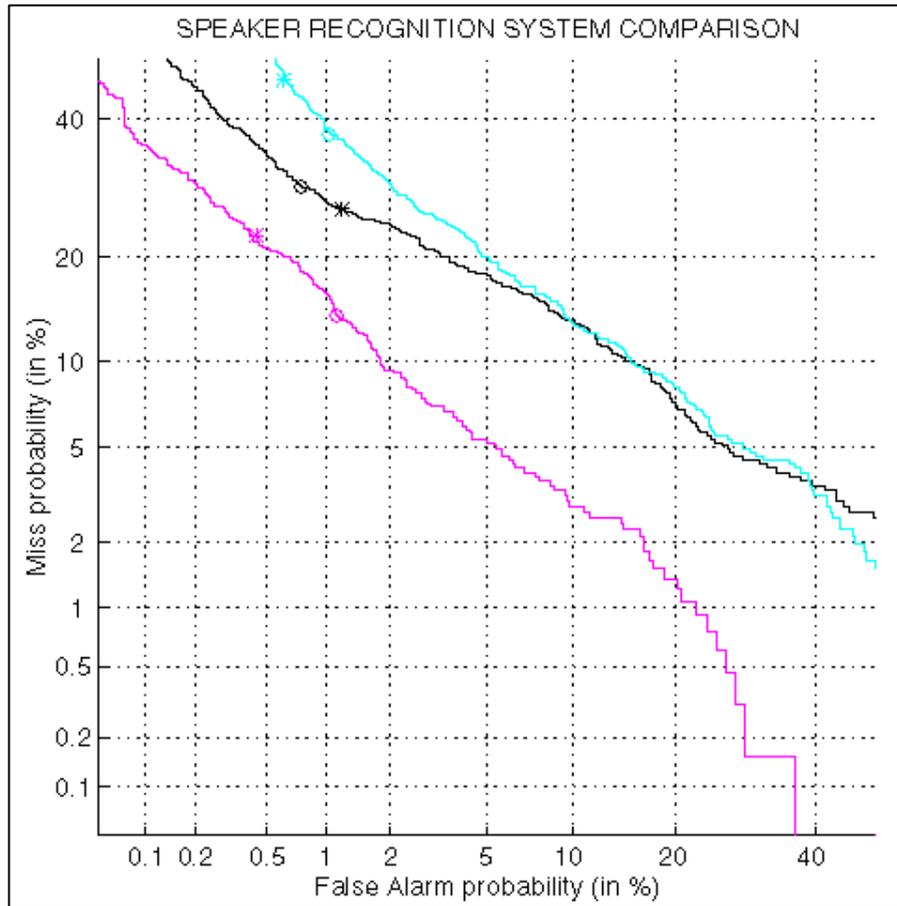


Figure 1: Plot of DET Curves for a speaker recognition evaluation.

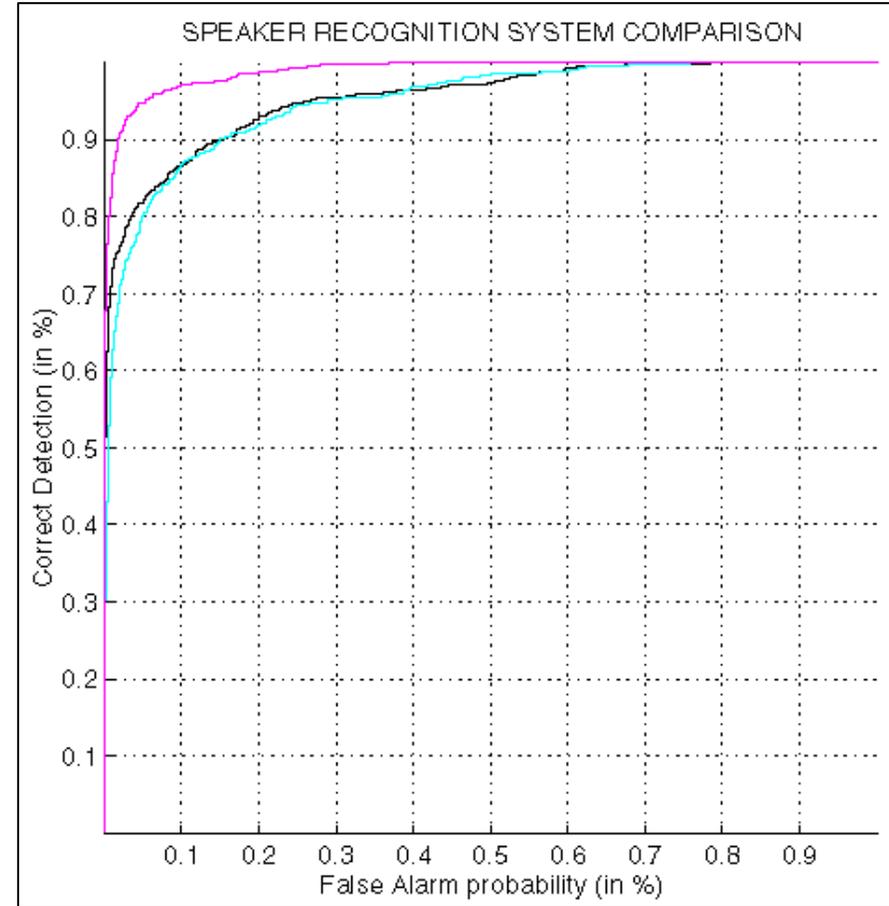
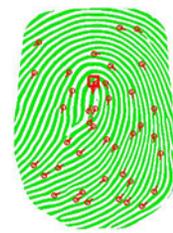


Figure 2: Plot of ROC Curves

Motor de reconocimiento de huellas dactilares

Este software asegura alta confiabilidad en el reconocimiento de huellas dactilares y una velocidad de comparación de hasta 20000 huellas por segundo. La comparación directa entre la imagen de la huella a ser identificada y las numerosas imágenes almacenadas, no servirían para una comparación fiable, debido a la alta sensibilidad a los errores (Ej. Ruidos de la imagen, áreas de huellas dañadas, posturas diferentes del dedo...). Una solución avanzada a este problema es extraer características de los puntos de minutia (puntos donde las huellas dactilares tienen bifurcaciones o finalizaciones) a partir de la imagen de la huella y comparar entre estos conjuntos de características.

Imagen original



**Imagen filtrada y
dirección de la minutia**

Motor de reconocimiento de huellas dactilares. Confiabilidad

Las curvas ROC obtenidas a partir de cuatro bases de datos estándar, fueron usadas para estudiar la confiabilidad de este sistema, obteniéndose para las diferentes bases de datos, resultados mejores o iguales que los alcanzados por otros algoritmos.

