

# **Máster en Estadística Aplicada**

Departamento de Estadística e Investigación  
Operativa



**Universidad de Granada**

## **El índice de Youden y su aplicación a la determinación del punto de corte en un test cuantitativo**

Autor: Jorge Mario Estrada Álvarez  
Tutor: Juan de Dios Luna Del castillo PhD.

Curso: 2015/2016

# Índice

0	Introducción y objetivos del presente trabajo .....	4
Las pruebas diagnósticas en medicina y cualquier área de la salud, son un instrumento que junto con la pericia clínica y examen físico contribuyen a la identificación de situaciones anómalas en el individuo. ... 4		
1	Definición del Índice de Youden, propiedades y usos.....	6
1.1	Definición de los parámetros básicos de un test diagnóstico: Sensibilidad, Especificidad, Razones de verosimilitudes, del positivo y del negativo. ....	6
1.1.1	Sensibilidad y Especificidad.....	7
1.1.2	Razones de verosimilitud ( <b>LR</b> ).....	8
1.2	Definición del Índice de Youden. Valores del índice de Youden y sus interpretaciones. Índice de Youden Ponderado: Valores e Interpretación. ....	9
1.2.1	Índice de Youden Ponderado.....	10
1.3	Relación entre el Índice de Youden y la incertidumbre y los errores del test diagnóstico. ....	13
1.4	La curva ROC y la determinación del punto de corte óptimo empleando el índice de Youden. ....	15
1.4.1	La curva ROC.....	16
1.4.2	Propiedades de la curva ROC.....	18
1.4.3	Estadísticos resumen de la curva ROC .....	19
2	Problemas de inferencia con el índice de Youden.....	21
2.1	Estimación puntual del Índice de Youden.....	21
2.1.1	Estimación puntual del índice de Youden por máxima verosimilitud .....	22
2.2	Determinación del punto de corte optimo en una curva ROC con el índice de Youden .....	23
2.2.1	Métodos paramétricos.....	24
2.2.2	Métodos No paramétricos .....	27
2.3	Intervalos de confianza para el índice de Youden .....	29
2.4	Contraste de hipótesis para un índice de Youden .....	40
2.5	Comparación de dos índices de Youden con muestras independientes.....	41
2.6	Comparación de dos índices de Youden con muestras apareadas .....	42
3	La aplicación de índice de Youden a la determinación del punto de corte óptimo en dos ejemplos reales. ....	45
3.1	Software en R de determinación del punto de corte óptimo. ....	45

3.2	Determinación del valor diagnóstico de la edad, el sexo y el índice de masa corporal para la hipertensión a través de los datos de la Encuesta Nacional de Salud: comparación de la calidad diagnóstica y de los puntos de cortes óptimos a través del índice de Youden.....	47
3.2.1	Puntos de Corte para Hipertensión por Índice de Youden .....	49
3.2.2	Puntos de Corte para Diabetes por Índice de Youden.....	51
3.2.3	Puntos de Corte para Hipercolesterolemia por Índice de Youden .....	53
3.2.4	Puntos de corte para la edad con regresión logística en patología crónicas .....	55
3.3	Valor pronóstico de una regla predictiva de la depresión en pacientes de Atención Primaria: determinación del punto de corte óptimo a partir del índice de Youden para la decisión de instaurar o no un tratamiento preventivo. ....	62
4	Conclusiones .....	65
5	Índice de Gráficos .....	67
6	Índice de Tablas .....	68
7	Bibliografía .....	69

# 0 Introducción y objetivos del presente trabajo

Las pruebas diagnósticas en medicina y cualquier área de la salud, son un instrumento que junto con la pericia clínica y examen físico contribuyen a la identificación de situaciones anómalas en el individuo.

Estas pruebas diagnósticas corresponden a resultados de test en ocasiones con respuestas en forma de medición nominal, con 2 opciones de resultado (Binario) o con múltiples opciones de repuestas ordenadas (ordinales) e incluso resultados que asumen una escala continua.

Es de interés para la comunidad científica el estudiar la capacidad diagnostica de los diferentes test existentes o producidos de novo, lo cual implica la aplicación de diversos diseños de investigación, así como de métodos estadísticos que permitan estimar cuantitativamente la exactitud diagnostica de la prueba utilizada en la identificación de condiciones anómalas (enfermedad).

En la teoría general de métodos estadísticos para diagnostico en medicina, medidas de exactitud de un test diagnostico han sido muy bien descritas, siendo clasificadas según su dependencia o no de la prevalencia de la condición a identificar; este es el caso por ejemplo de la Sensibilidad y la Especificidad, que corresponde a la capacidad de la prueba de detectar enfermos y sanos respectivamente, estas dos, siendo no dependientes de la prevalencia. Otras medidas dependientes de la prevalencia, como los valores predictivos positivo y negativo, que dan cuenta de la cercanía a un diagnostico cuando el test es positivo y negativo respectivamente, y que tienen un alto uso en la práctica clínica por parte de los profesionales de salud.

Combinaciones de los anteriores han hecho surgir diferentes medidas que propenden por ventajas a los ya planteados. Este es el caso del índice de Youden definido como la suma entre la sensibilidad y especificidad menos la unidad, el cual será estudiado a profundidad en este trabajo y especialmente su uso como un método para determinar el punto de corte óptimo para clasificación de enfermos y sanos cuando el test asume un resultado en escala continua. Por tal motivo este trabajo tiene como objetivos los siguientes:

Definir y dar a conocer las propiedades y usos del índice de Youden como medida estadística de exactitud de una prueba diagnóstica.

Describir el índice de Youden originalmente propuesto por W. Youden y algunas variaciones intentando integrar ponderaciones para la sensibilidad y especificidad basado en la razón de costo.

Identificar la relación entre la Curva ROC de un test diagnóstico cuantitativo y el índice de Youden.

Plantear el índice de Youden como medida para identificar el punto de corte óptimo en un test diagnóstico cuantitativo y describir los diferentes métodos inferenciales para su estimación puntual y por intervalo.

Aplicar en dos ejemplos reales el uso del índice de Youden como método para determinar el punto de corte óptimo para un test cuantitativo.

# 1 Definición del Índice de Youden, propiedades y usos.

## 1.1 Definición de los parámetros básicos de un test diagnóstico: Sensibilidad, Especificidad, Razones de verosimilitudes, del positivo y del negativo.

El uso de pruebas diagnósticas es una herramienta de uso común en muchas áreas de la salud, con el fin de establecer la presencia de una condición, que en la mayoría de las veces corresponde con la presencia o ausencia de la enfermedad, de esta manera es de interés conocer el desempeño de los instrumentos diagnósticos, este desempeño está directamente evaluado desde diferentes puntos de vista, específicamente en la capacidad de discriminar los sujetos enfermos de los sanos. Es claro que asociado a este desempeño también se encuentra la clasificación errónea de los sujetos, es decir, los errores cometidos por la prueba.

Existen diferentes indicadores estadísticos o parámetros que permiten evaluar una prueba diagnóstica en los aspectos anteriormente mencionados, algunos de estos como la sensibilidad, la especificidad, las razones de verosimilitud positiva y negativa, evalúan características intrínsecas de los test sin depender de la prevalencia de la condición a detectar (probabilidad de ocurrencia de la enfermedad), mientras que parámetros como el valor predictivo positivo y negativo, que dependen de la sensibilidad y especificidad del test, también dependen fuertemente de la prevalencia de la enfermedad.

A continuación se estudiara cada uno de estos parámetros teniendo como base una tabla de contingencia de  $2 \times 2$ , donde la primera fila, que se denomina  $T^+$ , representa el resultado positivo en una prueba, la segunda fila, se denomina  $T^-$ , representa un resultado negativo; en las columnas, la primera representa un grupo de sujetos con presencia segura de la enfermedad o condición de interés en diagnosticar notada como  $D^+$ , la segunda representa un grupo de sujetos sin la condición de interés o sanos, con notación  $D^-$ , la condición representada en la columnas de la tabla, se asumen que fueron establecidas por una prueba gold estándar o de referencia que mejor clasifica correctamente la enfermedad.

La tabla resume los resultados del test diagnóstico que se pueden clasificar como verdaderos positivos ( $a$ ), verdaderos negativos ( $d$ ), falsos positivos ( $c$ ) o falsos negativos ( $b$ ). Un verdadero positivo ocurre cuando un sujeto enfermo es correctamente clasificado como positivo en el resultado del test, un verdadero negativo ocurre cuando un sujeto no afectado por la enfermedad tiene un resultado negativo en el test diagnóstico, un falso positivo ocurre cuando un sujeto no afectado por la enfermedad tiene un

resultado positivo en el test diagnóstico y un falso negativo ocurre cuando un sujeto afectado por la enfermedad obtiene un resultado negativo en el test diagnóstico.

**Tabla 1** Tabla de contingencia 2×2 para los resultados de evaluación de una prueba diagnóstica con resultado binario

		Resultado del test		
		$T^+$	$T^-$	
Estado	$D^+$	$a$	$b$	$F_1 = a + b$
	$D^-$	$c$	$d$	$F_2 = c + d$
		$C_1$	$C_2$	$n$

### 1.1.1 Sensibilidad y Especificidad

La sensibilidad y Especificidad, permiten dar cuenta de dos aspectos muy importantes del test, sobre la capacidad que tiene el test para discriminar. La sensibilidad acoge el aspecto de la capacidad de detectar la enfermedad o condición de interés cuando la misma está presente y está dada por la probabilidad de tener un resultado positivo en presencia de la enfermedad  $Se = P(T^+/D^+)$ , también es conocida como la fracción de verdaderos positivos, vista bajo la notación de la tabla de  $2 \times 2$  :

$$Se = P(T^+/D^+) = \frac{a}{F_1} = \frac{a}{a + b}$$

( 1 )

La especificidad, por el contrario, da cuenta la capacidad que tiene un test para descartar la enfermedad cuando esta es ausente, en forma de probabilidad condicionada se da como  $Sp = P(T^-/D^-)$ , también es conocida como la fracción de verdaderos negativos, también bajo la notación de la tabla  $2 \times 2$ , se puede escribir como:

$$Sp = P(T^-/D^-) = \frac{d}{F_2} = \frac{d}{c + d}$$

( 2 )

Estos dos indicadores toman valores entre  $[0 - 1]$ , cuando los dos toman valores la unidad, el test es perfecto, sin embargo, esto no es tan común en la práctica y por lo general estos son importantes cuando son mayores a  $\frac{1}{2}$ . La sensibilidad y la especificidad son indicadores de la exactitud diagnóstica de un test

que no dependen de la prevalencia de la enfermedad, es caso que decir que el valor de estos es independiente de la población en que se apliquen, y estas evalúan características intrínsecas del test, es decir, más su desempeño diagnóstico, por tal motivo son también llamadas medidas de exactitud.

### 1.1.2 Razones de verosimilitud ( $LR$ )

Estos indicadores, son más informativos, consiste en el cociente de dos probabilidades, entre la probabilidad de tener un resultado positivo estando enfermo y la probabilidad de tener un resultado positivo si se estuviera sano, la anterior define la razón de probabilidades positiva, mientras que si el cociente es entre la probabilidad de tener un resultado negativo mientras se está enfermo y la probabilidad de tener un resultado negativo cuando se está sano, se denomina razón de probabilidades negativa. De esta manera se pueden representar como:

$$LR^+ = \frac{P(T^+/D^+)}{P(T^+/D^-)} = \frac{Se}{1 - Sp}$$

( 3 )

$$LR^- = \frac{P(T^-/D^+)}{P(T^-/D^-)} = \frac{1 - Se}{Sp}$$

( 4 )

La razón de verosimilitudes, pueden tomar valores entre 0 e  $\infty$ . Las  $LR$ s cuantifican el aumento en el conocimiento sobre la presencia de la enfermedad que es adquirido a través de la prueba diagnóstica. Por lo tanto, los parámetros ( $LR^+$ ,  $LR^-$ ) cuantifican el cambio en las probabilidades de enfermedad obtenido por el conocimiento de los resultados de la prueba diagnóstica. Se esperaría que, en condiciones adecuadas de un test, la probabilidad de tener un resultado positivo cuando se está enfermo debe ser mayor que cuando no se está enfermo, por tal motivo el  $LR^+$  siempre tendera a ser mayor que 1, siendo más informativo entre más se aleje de 1. En el mismo sentido, la probabilidad de tener un resultado negativo cuando se está enfermo debe ser menor que cuando se está sano, por tal motivo el  $LR^-$  siempre tendera a ser menor a 1 y entre más informativo será más por debajo de 1.



## 1.2 Definición del Índice de Youden. Valores del índice de Youden y sus interpretaciones. Índice de Youden Ponderado: Valores e Interpretación.

El índice  $J$  de Youden fue introducido por W. J. Youden en 1950 (Youden, 1950), y se define como:

$$J = 1 - [P(T^+/D^-) + P(T^-/D^+)] = 1 - [1 - Sp + 1 - Se] = Se + Sp - 1$$

el índice se expresa como la suma de sensibilidad y especificidad del test menos la unidad, de esta manera una de las interpretaciones dadas a los resultados del índice es la ganancia promedio de certidumbre neta en la clasificación de un individuo sano o enfermo. Dado que se asume igualmente graves los falsos positivos y falsos negativos, sensibilidades y especificidades por debajo de 0.5, harían que el índice tome valores menores a 1, lo que indicaría que el test tiene un comportamiento negativo con la enfermedad, lo cual no es común en la práctica. Mientras que si se toman valores de 0.5 para sensibilidad y especificidad (asumido ser un test inútil en la práctica) el índice  $J$  toma valor de cero. Por el contrario, si el valor de sensibilidad y especificidad tienden a valores de 1 como máximo, entonces el valor del índice de Youden tomaría como valor máximo 1, indicando un desempeño excelente del test.

Bajo el razonamiento anterior se puede concluir que el índice toma valores entre  $[0,1]$  siendo 1 un test perfecto y cero un test inútil para detectar o descartar la enfermedad.

Este índice tiene unas características importantes que fueron descritas por W.J. Youden (Youden, 1950) y que se resumen a continuación:

- El índice tiene valor cero cuando tiene la misma proporción de positivos en el grupo de enfermos y sanos (controles).
- El índice se vuelve 1 únicamente cuando los falsos positivos y falsos negativos no están presentes. Si solo se presenta un tipo de error, el índice es controlado por ese error.
- El índice es independiente de los tamaños relativos y absolutos de los grupos en que se evalúa la prueba (Enfermos vs Controles).
- Todas las pruebas con el mismo índice, hacen el mismo número total de malas clasificaciones por cada 100 pacientes.
- Es posible calcular un error estándar para el índice, siendo el tamaño de la muestra un factor que influye naturalmente la confiabilidad en la estimación.

Las anteriores condiciones hacen del índice de Youden un indicador con ventajas en su uso, siempre y cuando se de la misma importancia a la sensibilidad y la especificidad, sin embargo, en la práctica clínica

es necesario de acuerdo a la situación de aplicación del test asignarle una mayor importancia a una de las dos, esto llevo a que Li y cols. desarrollaran un índice de Youden que permitiera dicha ponderación y lo denominaron índice de Youden ponderado.

### 1.2.1 Índice de Youden Ponderado

Li y cols. (Li, Shen, Yin, Peng, & Chen, 2013) subsanando en parte la desventaja del índice  $J$  de Youden, frente a que asume igualmente importantes la sensibilidad y especificidad y colocando el hecho, que en la realidad estos parámetros tienen una relación inversa, es decir, en la medida en que un test mejora su sensibilidad, su especificidad disminuye y viceversa, plantearon una modificación al índice, agregando una ponderación ( $w$ ) que daría importancia relativa y diferencial a la sensibilidad y especificidad, dado que esto en la práctica clínica es de importancia cuando se requieren test preferencialmente con alta sensibilidad a expensas de su especificidad o por el contrario con un alta especificidad. El nuevo índice de Youden ponderado ( $J_w$ ) es definido como:

$$J_w = 2(w \times Se + (1 - w) \times Sp) - 1 \quad \text{con } 0 \leq w \leq 1$$

( 5 )

La formulación para el índice  $J_w$  ponderado trae consigo el cumplimiento de tres principios: 1) la suma de las ponderaciones asignadas a sensibilidad y especificidad es 1; por tanto, si  $w$  es asignado a la sensibilidad,  $(1 - w)$  será la ponderación para especificidad. 2) el  $J_w$  toma los mismos valores que  $J$ , es decir entre  $[0,1]$ . 3) Cuando los mismos pesos son asignados a sensibilidad y especificidad, es decir 0.5, el índice de Youden ponderado es igual al índice de Youden.

Los aspectos relacionados a la asignación de los pesos o ponderaciones fueron abordados por Smits (Smits, 2010), quien discute frente al costo relacionado con clasificación correcta o incorrecta en un individuo sano o enfermo, y este costo puede utilizarse como ponderación dentro de la estimación del índice de Youden. Por lo anterior Smits definió los siguientes resultados en una tabla de contingencia de  $2 \times 2$  y organizó los resultados así:

**Tabla 2. Tabla de contingencia de  $2 \times 2$  para un test cuantitativo para un punto de corte  $c$**

		Resultado del test		
		$T^+$	$T^-$	
$D^+$		VP	FN	$P$

<b>Estado</b>	$D^-$	<table> <tr> <td><math>FP</math></td> <td><math>VN</math></td> </tr> </table>	$FP$	$VN$	$P' = 1 - P$
$FP$	$VN$				
		$Q_c$ $Q'_c = 1 - Q_c$			

1

Siendo:

$P$ : Prevalencia

$Q_c$ : Proporción de personas que tienen un resultado positivo

$TP, TN, FP, FN$ : proporciones de la población general de clasificaciones correctas e incorrectas

Las siguientes relaciones se pueden deducir de la tabla 2:

$$P = VP + FN$$

$$Q_c = FP + VP$$

$$Se = \frac{VP}{P}$$

$$Sp = \frac{VN}{P'}$$

Smits define el costo esperado (CE) asociado a un resultado (punto de corte o nivel o umbral de un test) como la suma entre los productos de costos de decisiones incorrectas y la proporción de clasificados incorrectos (falsos positivos y falsos negativos) y beneficios de decisiones correctas con la proporción de clasificados correctamente (verdaderos positivos y verdaderos negativos), la ecuación de costos esperados sería:

$$CE = [(C_{FP} \times FP) + (C_{FN} \times FN)] - [(B_{VP} + VP) + (B_{VN} + VN)]$$

( 6 )

siendo:

$C_{FP}$ : Costos asociados a falsos positivos

$C_{FN}$ : Costos asociados a falsos negativos

$B_{VP}$ : Beneficios asociados a verdaderos positivos

$B_{VN}$ : Beneficios asociados a verdaderos negativos

El umbral  $c$  asociado con el costo esperado más bajo determinara el punto de corte óptimo para la clasificación de un test. Sin embargo, es más conveniente considerar la razón de costos ( $r$ ) como sigue:

$$r = \frac{B_{VP} - C_{FN}}{B_{VP} - C_{FN} + B_{VN} - C_{FP}} \quad (7)$$

$B_{VP} - C_{FN}$  en el numerador de la razón de costos ( $r$ ) refleja que tanta diferencia en costos se da si las personas enfermas son clasificadas correcta o incorrectamente, mientras que  $B_{VN} - C_{FP}$  en el denominador refleja que tanta diferencia en costos se muestra si las personas sanas son clasificadas correcta o incorrectamente. Así  $r$  tomara valores cercanos a 1 cuando la diferencia en costos asociados a clasificaciones correctas o incorrectas en personas enfermas son mayores y tomara valores cercanos a cero cuando la diferencia en costos es mucho mayor para las clasificaciones correctas e incorrectas de las personas sanas. Únicamente los costos serán iguales entre ambos grupos (enfermos y sanos) cuando  $r \approx 0,5$ .

De esta manera el costo esperado puede ser expresado en términos de  $r$  así:

$$CE' = P \times Se \times r + (1 - P) \times Sp \times (1 - r)$$

utilizando las relaciones deducidas de la tabla 2:

$$CE' = VP \times r + VN \times (1 - r) \quad (8)$$

De nuevo tomando el índice de Youden para un nivel de un test  $c$  y sustrayendo una constante que no afecta la elección para un  $c$ , se puede expresar como:

$$J' = \frac{VP}{P} + \frac{VN}{1 - P} = \frac{VP(1 - P) + VN(P)}{P(1 - P)}$$

Multiplicando por una constante no cambia la elección del  $c$ :

$$J'' = VP(1 - P) + VN(P) \quad (9)$$

Como se aprecia la ecuación (8) y (9) son iguales, verificando que  $r = 1 - P$ , significando que la razón de costo esta en función de la prevalencia de la enfermedad de la población diana/objetivo donde se aplicaría la prueba, realizando una analogía con el índice de Youden ponderado propuesto por Li y cols. las ponderaciones a utilizar podrían ser de hecho la prevalencia de la enfermedad o condición a detectar o

diagnosticar, teniendo las siguientes interpretaciones: Si la población objetivo se presumen una alta prevalencia entonces  $r < 0,5$  de esta manera los errores de clasificación sobre las personas sanas tendrían un mayor costo, por el contrario, si la prevalencia es baja  $r > 0,5$  lo que implicaría que los errores de clasificación sobre los enfermos tendría un mayor costo. Finalmente, si  $r = 0,5$  los errores de clasificación serían valorados de manera igualitaria frente a los costos.

### 1.3 Relación entre el Índice de Youden y la incertidumbre y los errores del test diagnóstico.

Hughes (Hughes, 2015) mostró la relación entre el índice de Youden y la razón de verosimilitud positiva utilizando el análisis descrito por Shannon sobre el contenido de información de un mensaje, para este caso el resultado de un test. De esta manera, Si se espera que el resultado positivo del test sea poco probable y este sale positivo, el contenido de información o aporte del resultado obtenido es significativo o importante, contrario a lo que sucede si un resultado positivo es altamente probable, el contenido de información de este resultado es más bien pequeño. Shannon definió el contenido de información de un mensaje como  $h(p)$  en función de la probabilidad ( $p$ ) de ese mensaje:

$$h(p) = -\log(p)$$

( 10 )

El índice  $J$ -Youden puede ser expresado teniendo que  $Sp = 1 - FFP$  y  $Se = FVP$ , siendo  $FFP$  la proporción de falsos positivos y un estimador de la probabilidad condicionada  $P(T^+ / D^+)$  y  $FVP$  la proporción de verdaderos positivos correspondiendo al estimador de la probabilidad  $P(T^+ / D^-)$ :

$$J = Se + Sp - 1$$

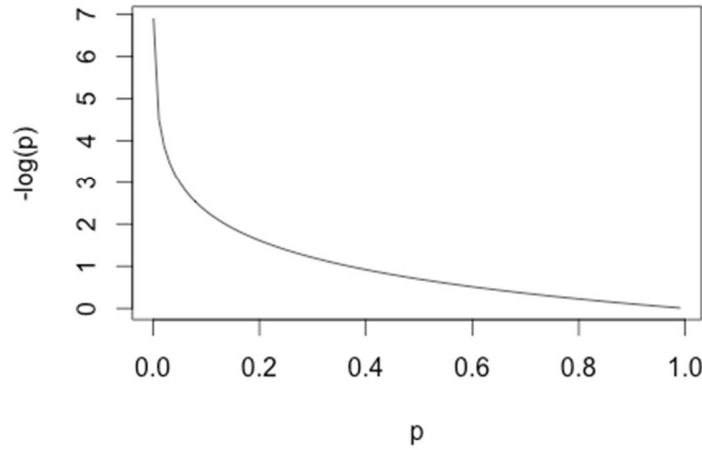
$$J = FVP + (1 - FFP) - 1$$

$$J = FVP - FFP = P(T^+ / D^+) - P(T^+ / D^-)$$

( 11 )

Si en una representación como el gráfico 1 se ubicarán en la coordenada "x" el intervalo producido por la diferencia entre las proporciones de verdaderos y falsos positivos, lo cual correspondería al índice  $J$ -Youden y en la coordenada "y" la diferencia entre  $-\log(p)$  para cada probabilidad se obtendría:

**Grafico 1 Grafico de Shannon para el contenido de información de un mensaje**



$$-\log[P(T^+ / D^+)] = P(T^+ / D^+)$$

$$-\log[P(T^+ / D^-)] = P(T^+ / D^-)$$

Para la coordenada "x"

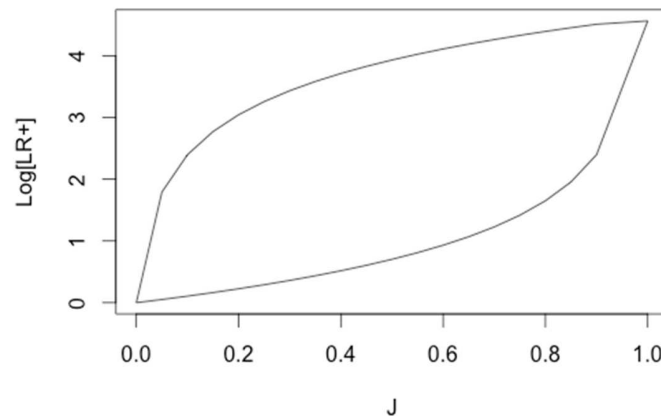
$$P(T^+ / D^+) - P(T^+ / D^-) = PVP - PFP = J$$

Para la coordenada "y"

$$-\log[P(T^+ / D^+)] - (-\log[P(T^+ / D^-)]) = \log \left[ \frac{P(T^+ / D^+)}{P(T^+ / D^-)} \right] = \log[LR^+]$$

La correspondencia entre  $J$  – Youden y el  $\log[LR^+]$  graficamente es tal que incrementa monotonicamente el  $\log[LR^+]$  desde  $J = 0$  hasta  $J = 1$ . como se muestra en la gráfica 2. La curva inferior fue simulada para valores grandes de sensibilidad ( $Se \approx 0.99$ ) y variando valores de especificidad, mientras que la curva superior se obtuvo para valores altos de especificidad ( $Sp \approx 0.99$ ) y variando valores de sensibilidad. como se nota la relación en ambos casos es monótonamente creciente. Sin embargo, en los extremos esta relación no es uno a uno, debido a que para valores similares de  $J$ -Youden se tiene necesariamente valores similares en el  $\log[LR^+]$  (D Böhning, 2015; Dankmar Böhning, Böhning, & Holling, 2008).

**Grafico 2 Relación entre el índice  $J$  – Youden y el logaritmo de la razon de verosimilitud positiva  $\log[LR^+]$**

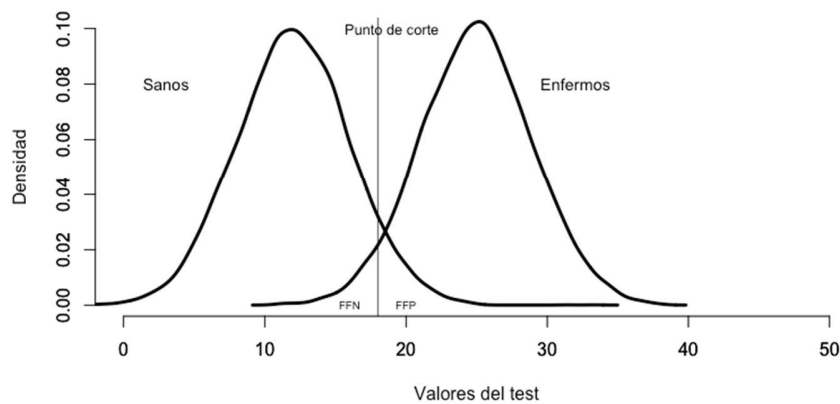


#### 1.4 La curva ROC y la determinación del punto de corte óptimo empleando el índice de Youden.

Hasta al momento se ha definido los indicadores de exactitud diagnóstica para un test donde el resultado es binario (Positivo – Negativo; Presente – Ausente). Sin embargo, es muy frecuente que el resultado del test diagnóstico sea un valor numérico (variable continua) de manera que conforme sea mayor (o menor) sea un mayor indicador de la enfermedad; en este caso estamos en la presencia de un test cuantitativo. como ejemplos de este tipo de test son frecuentemente vistos los biomarcadores, donde su presencia se cuantifica e indica la presencia temprana de la enfermedad. En cardiología resultados como la fracción de eyección, el strain global longitudinal, son indicadores de la afectación del miocardio por disminución de la irrigación sanguínea (presencia de Infarto agudo al miocardio). De esta manera se pueden considerar la distribución de los resultados del test tanto en población sana como enferma, dando a lugar a la siguiente gráfica. Es importante que se puede seleccionar un punto de corte por encima (por debajo) que pueda clasificar como enfermos (o como sanos). Sin embargo, este mismo punto puede generar dos tipos de errores.

Cuando para valores mayores a un punto de corte elegido se declaran los enfermos y normales o sanos en el caso contrario, da lugar a que un porcentaje de los normales sean clasificados como enfermos (falsos positivos), este error es denominado alfa ( $\alpha$ ), y al mismo tiempo este punto de corte hace que un porcentaje de los enfermos sean clasificados como sanos (falsos negativos), este error es denominado beta ( $\beta$ ).

**Grafico 3 Distribución de los valores de un test cuantitativo en población sana y enferma**



Si se quisiera disminuir el error asociado a la clasificación de falsos positivos, un corrimiento del punto de corte hacia la izquierda convendría, sin embargo, esta decisión repercutiría inmediatamente con un incremento en el error asociado a falsos negativos. Para el caso contrario ocurriría lo mismo, de esta manera la selección del punto de corte podría estar basado en la fijación de antemano de cualquiera de los dos errores  $\alpha$  o  $\beta$ . Es clara la relación entre el punto de corte y los errores de clasificación, de modo que fijado uno de ellos quedan determinados los demás.

Dado que  $\alpha$  y  $\beta$  están relacionados entre sí (a través del punto de corte), suele ser costumbre reflejar esta relación, representando en el plano una curva, llamada curva ROC (siglas en inglés Receiver Operating Characteristic curve), que relaciona el porcentaje de falsos positivos del test  $\alpha$  (en el eje horizontal) con la sensibilidad  $1 - \beta$  del mismo (en el eje vertical).

La curva ROC permite evaluar cómo se comporta el test diagnóstico para cada punto de corte seleccionado y, a partir de ella, es factible deducir cuál es el punto de corte óptimo: aquel valor en el que un pequeño aumento de la sensibilidad ocasiona un incremento excesivo (en términos relativos) de la proporción de falsos positivos.

### 1.4.1 La curva ROC

La curva ROC es un gráfico bidimensional en los que la sensibilidad ( $Se$ ) y la fracción de falsos positivos ( $1 - Sp$ ) son representados. para su mejor comprensión se dan las siguientes condiciones:

Considérese una variable aleatoria  $Y$ , continua, que es el resultado de un test diagnóstico continuo, y supongamos, sin pérdida de generalidad, que valores altos de esa variable aleatoria suponen más evidencia de enfermedad que valores pequeños. Supongamos un punto de corte  $c$  de manera que si  $Y \geq c$  consideraremos que el tests es positivo y si  $Y < c$  el test es negativo. Ligados a ese punto de corte  $c$  tenemos dos proporciones, una de acierto y otra de fallo:



$$FVP(c) = P(Y \geq c/D^+)$$

$$FFP = P(Y \geq c/D^-)$$

La primera es la fracción de verdaderos positivos, ligada a  $c$ , pues es la probabilidad de que el test diagnóstico desborde a  $c$  en la población de enfermos, mientras que la segunda es la fracción de falsos positivos puesto que es la probabilidad de que el test diagnóstico esté por encima de  $c$  en la población de personas no enfermas.

A partir de aquí, para cada valor diferente de  $c$  podemos tener una pareja de valores  $\{(FVP(c), FFP(c))\}$ , y a ese conjunto de parejas de valores es a lo que se denomina curva ROC:

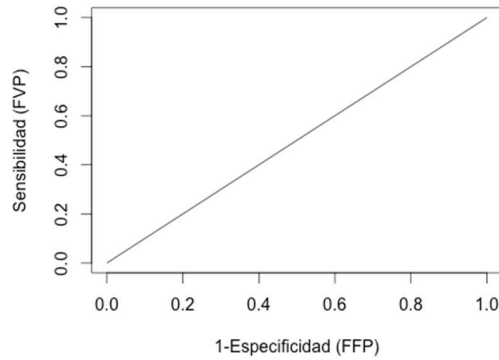
$$ROC(c) = \{(FVP(c), FFP(c)) / c \in (-\infty, +\infty)\}$$

( 12 )

La curva está representada, que para cada valor  $c$  existe un valor de sensibilidad (fracción de verdaderos positivos) contra la fracción de falsos positivos, y como estas dos fracciones toman valores entre 0 y 1, la curva se representa en el cuadrante (0,0) ; (0,1); (1,1); (1,0) que define el espacio ROC (Noguera Moreno, 2010), con los siguientes puntos importantes a ser destacados:

- El punto (0,0) que representa la no clasificación nunca de un positivo, y no comete errores falsos positivos.
- El punto (1,1) que representa la opuesta al anterior de emitir incondicionalmente clasificaciones positivas.
- El punto (0,1) que representa la clasificación perfecta. totalidad de verdaderos positivos, y ningún falso positivo.
- La diagonal  $x = y$  representa la clasificación al azar.

#### Grafico 4. Representación en un plano del espacio ROC



Por tanto, los puntos de corte que caen en el triángulo inferior, se dice que tiene una clasificación peor que el azar, el test sería inútil, ya que no muestra capacidad de detección de enfermos y tampoco de sanos de manera correcta. Por el contrario, un test que forma una curva en el triángulo superior del área, significaría un test con propiedades interesantes en el diagnóstico de la enfermedad, siendo los puntos más cercanos al nordeste (punto (0,1)) con las mejores condiciones en terminos de sensibilidad y especificidad.

#### 1.4.2 Propiedades de la curva ROC

- La curva ROC es una función monótona en  $(0,1)$  con  $\lim_{t \rightarrow 0} ROC(t) = 0$  y con  $\lim_{t \rightarrow 1} ROC(t) = 1$  con  $t \in (0,1)$ .
- Si se denotan por  $F_{D+}$  y  $F_{D-}$  las distribuciones de resultados o puntuaciones del test diagnóstico con variable aleatoria  $Y$  (Grafica 3), en las poblaciones enferma y sana respectivamente, la función ROC se puede escribir como:

$$ROC(t) = 1 - F_{D+}(F_{D-}^{-1}(1-t)), \quad t \in (0,1)$$

( 13 )

- La curva ROC es invariante si los resultados o puntuaciones del marcador son sometidas a transformaciones monótonas estrictamente crecientes.
- La curva ROC es la función en  $(0,1)$  con  $ROC(0) = 0$  y  $ROC(1) = 1$  con pendiente:

$$\frac{\partial ROC(t)}{\partial t} = \frac{f_{D+}(F_{D-}^{-1}(t))}{f_{D-}(F_{D-}^{-1}(t))}$$

( 14 )

donde  $f_{D+}$  y  $f_{D-}$  son las funciones de densidad de los resultados del test diagnóstico.

### 1.4.3 Estadísticos resumen de la curva ROC

La curva ROC describe el comportamiento de un test diagnóstico cuantitativo en todo su rango de posibles valores, sin embargo es necesario en algunas ocasiones comparar varios test diagnósticos en cuanto a su desempeño diagnóstico, por tal razón se han construido indicadores resumen de la curva, en forma de escalares únicos que permitan fácilmente comparar y resumir las características y atributos de la curva ROC, entre los más comúnmente mencionados en la literatura se encuentra el área bajo la curva ROC y el índice de Youden, los cuales a continuación se revisaran (Noguera Moreno, 2010).

- **Área bajo la Curva ROC (AUC):** dado que AUC es una porción del área de un cuadrado de lado 1, su valor siempre estará entre 0 y 1, cualquier test que produzca un  $AUC < 0.5$ , estaría con una capacidad diagnóstica peor que el azar. se define la AUC como:

$$AUC = \int_0^1 ROC(t) dt$$

( 15 )

Una de las interpretaciones para el AUC, es la capacidad discriminativa entre sanos y enfermos, sin embargo, otras interpretaciones pueden ser dadas como la siguiente: si un test tiene una AUC de 0.8, significa que un individuo seleccionado al azar del grupo de enfermos tiene un valor mayor en el test que el de un individuo elegido al azar del grupo de sanos un 80% de las veces.

- **Índice de J-Youden:** Otro de los estadísticos utilizados para resumir resultados de una curva ROC es el ya definido previamente índice de Youden, el cual fue definido como la suma de sensibilidad y especificidad menos la unidad. De esta manera se puede determinar como punto de corte óptimo en la curva ROC como aquel donde el índice de Youden sea máximo, este mismo índice también puede ser expresado como se vio en la ecuación (11), siendo la diferencia entre la fracción de verdaderos positivos y fracción de falsos positivos, por tanto, el punto de corte óptimo puede ser seleccionado como aquel que maximice la diferencia entre estas fracciones. Perkins y cols. (Perkins & Schisterman, 2006) justificaron la elección del punto de corte óptimo en la curva ROC bajo el índice de Youden como el más apropiado, dado a los desacuerdos que se presentan con la metodología de la elección basado en el valor de punto de corte más cercano al punto (0,1) en la curva ROC. Dado lo anterior se define el punto de corte  $c_j$  basado en el índice de Youden:

$$J = \max\{Se(c) - (1 - Sp(c))\}$$

( 16 )

Según lo anterior en una curva ROC, el índice  $J$ -Youden es la distancia vertical máxima desde la curva ROC a la diagonal principal o de clasificación por azar.

## 2 Problemas de inferencia con el índice de Youden

### 2.1 Estimación puntual del Índice de Youden

Para la estimación puntual del índice  $J$ -Youden, se debe considerar las formas de muestreo que se utilizan, para obtener los datos con objeto de calcular las estimaciones necesarias y a partir de los mismos obtener el índice de Youden. Se han planteado dos tipos de muestreo que el investigador puede utilizar, el primero de ellos denominado muestreo tipo I (usado frecuentemente en diseños de investigación de corte transversal), el cual consiste en la selección de una muestra aleatoria de individuos de tamaño  $n$ , a los cuales se les aplica al tiempo tanto la prueba diagnóstica en evaluación como el gold estándar que define el estado real del paciente. El resultado de la recolección de datos de este esquema se puede resumir en la siguiente tabla, no perdiendo generalidad se asume que el test es positivo cuando logra un valor mayor o igual a un umbral o punto de corte  $c$ :

		Resultado del test		
		$T^+ \geq c$	$T^- < c$	
Gold Estándar	$D^+$	$a$	$b$	$F_1$
	$D^-$	$c$	$d$	$F_2$
		$C_1$	$C_2$	$n$

El segundo esquema es llamado Muestreo tipo II, este consiste en la selección de dos muestras independientes  $n_1$  y  $n_2$  correspondientes a individuos sanos y enfermos respectivamente. bajo este esquema los marginales correspondientes a total de sano y total de enfermos son fijos, las estimaciones de sensibilidad y especificidad pueden ser calculados y por consecuencia el índice de Youden. El muestreo tipo II es frecuentemente asociado al diseño de investigación denominado de casos y controles.

		Resultado del test	
		$T^+ \geq c$	$T^- < c$
$D^+$	$a$	$b$	$n_1$
$D^-$	$c$	$d$	$n_2$

Youden (Youden, 1950) definió un nuevo índice que permitiera expresar el desempeño de un test y que además favorecería la comparación entre nuevos o mejorados test.

Asumiendo la notación presentada en la tabla 1, y que es igualmente grave un resultado falso positivo como un resultado falso negativo, se puede definir la ganancia en certidumbre entre los individuos enfermos cuando tienen un resultado en un test cuantitativo mayor o igual a un punto de corte  $c$  ( $GN_c^+$ ) como:

$$GN_c^+ = P(T^+/D^+) - P(T^-/D^+) = \frac{a}{a+b} - \frac{b}{a+b}$$

$$GN_c^+ = \frac{a-b}{a+b}$$

Bajo las mismas condiciones se puede definir la ganancia neta de certidumbre entre el grupo de sanos o individuos sin la enfermedad cuando el test cuantitativo tiene un valor menor que un punto de corte o umbral  $c$  ( $GN_c^-$ ):

$$GN_c^- = P(T^-/D^-) - P(T^+/D^-) = \frac{d}{c+d} - \frac{c}{c+d}$$

$$GN_c^- = \frac{d-c}{c+d}$$

Youden definió su índice como el promedio de las ganancias netas en certidumbre y lo simbolizo como el índice  $J$ , sin perder generalidad para un punto de corte cualquiera  $c$ :

$$J_c = \frac{1}{2} \left[ \frac{a-b}{a+b} + \frac{d-c}{c+d} \right] = \frac{a}{a+b} + \frac{d}{c+d} - 1$$

### 2.1.1 Estimación puntual del índice de Youden por máxima verosimilitud

Teniendo en cuenta que el índice de Youden es dependiente de la sensibilidad y especificidad, se puede asumir la estimación de estas por sus estimadores máximos verosímiles, de esta manera con datos provenientes de un muestreo tipo I (transversal), según la tabla de contingencia tomando totales marginales para las filas  $F_1$  y  $F_2$  se puede mostrar que para la sensibilidad tiene una distribución de una proporción binominal  $a \rightarrow B(F_1 = n, Se = p)$ , donde  $a = \sum x_i$  es el número de enfermos y  $x_i \begin{cases} 1 \text{ si el test es positivo} \\ 0 \text{ si el test es negativo} \end{cases}$

$$p(x_1, x_2, x_3, \dots, x_n) = P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$$

$$L(p) = P(X_1 = x_1)P(X_2 = x_2) \dots P(X_n = x_n)$$

$$L(p) = p^{\sum x_i} (1-p)^{n-\sum x_i}$$

$$L(p) = p^a (1-p)^{n-a}$$

$$\log[L(p)] = a \log p + (n-a) \log(1-p)$$

$$\frac{d \log[L(p)]}{dp} = \frac{a}{\hat{p}} - \frac{n-a}{1-\hat{p}} = 0$$

$$\hat{p} = \widehat{Se} = \frac{a}{n} = \frac{a}{F_1}$$

Siguiendo el mismo procedimiento se obtiene el estimador de máxima verosimilitud que corresponde a la especificidad ( $Sp$ ), encontrándose:

$$\widehat{Sp} = \frac{d}{F_2}$$

Si se toma un muestreo tipo II, es fácil mostrar los mismos estimadores de máxima verosimilitud únicamente reemplazando los marginales de las filas por los correspondientes tamaños de muestra para los dos grupos independientes muestreado  $n_1$  y  $n_2$ .

La estimación puntual del índice de Youden se puede expresar como la combinación de estos dos estimadores de máxima verosimilitud:

$$\hat{J} = \widehat{Se} + \widehat{Sp} - 1$$

$$\hat{J} = \frac{a}{F_1} + \frac{d}{F_2} - 1 = \frac{ad - bc}{(a+b)(c+d)}$$

$$\hat{J}_c = \frac{ad - bc}{(a+b)(c+d)}$$

## 2.2 Determinación del punto de corte optimo en una curva ROC con el índice de Youden

Para estimar el punto de corte óptimo para un test cuantitativo usando el índice de Youden existen diferentes métodos, principalmente estos se clasifican en métodos paramétricos y no-paramétricos.

Sea los valores de un test cuantitativo  $x_1, x_2, x_3, \dots, x_n$  e  $y_1, y_2, y_3, \dots, y_m$  para dos muestras aleatorias sobre una población de enfermos y sanos respectivamente, igualmente sean sus respectivas funciones de distribución acumulada  $G_{D+}$  y  $F_{D-}$ . Para cualquier valor o punto de corte  $c$  del test la  $Se(c) = 1 - G_{D+}$  y  $Sp(c) = F_{D-}$ . El índice de Youden es:

$$J = \max_c \{Se(c) + Sp(c) - 1\} = \max_c \{1 - G_{D+} + F_{D-} - 1\}$$

$$J = \max_c \{F_{D^-} - G_{D^+}\} \quad (17)$$

El valor  $c^*$  que logra este maximo será el punto de corte óptimo. El valor del índice será dado por la estimación de  $G_{D^+}$  y  $F_{D^-}$  ( $\hat{G}_{D^+}, \hat{F}_{D^-}$ ), sustituyendo estos estimadores en la ecuación anterior se obtendría:

$$\hat{J} = \max_{c^*} \{\hat{F}_{D^-} - \hat{G}_{D^+}\} \quad (18)$$

Como se mencionó, dos enfoques se pueden utilizar para la estimación de  $G_{D^+}$  y  $F_{D^-}$  sin embargo como fue discutido por Fluss y cols. (Fluss, Faraggi, & Reiser, 2005) estos diferentes métodos pueden llevar a diferentes estimaciones de  $J$  y  $c^*$ .

### 2.2.1 Métodos paramétricos

Un enfoque simple bajo el método paramétrico para la estimación de  $J$  consiste en asumir que los valores del test en población sana y enferma asumen distribuciones normales con varianzas distintas. De esta manera sea  $X$  e  $Y$  los valores del test en población sana y enferma respectivamente, y que asumen una distribución normal e independiente con varianzas diferentes así:  $X \rightarrow N(\mu_{D^+}, \sigma_{D^+}^2)$  y  $Y \rightarrow (\mu_{D^-}, \sigma_{D^-}^2)$ . Se asume que  $\mu_{D^+} > \mu_{D^-}$  entonces:

$$\begin{aligned} Se(c) &= 1 - \Phi\left(\frac{c - \mu_{D^+}}{\sigma_{D^+}}\right), & Sp(c) &= \Phi\left(\frac{c - \mu_{D^-}}{\sigma_{D^-}}\right) \\ J &= \max_c \left\{ 1 - \Phi\left(\frac{c - \mu_{D^+}}{\sigma_{D^+}}\right) + \Phi\left(\frac{c - \mu_{D^-}}{\sigma_{D^-}}\right) - 1 \right\} \\ J &= \max_c \left\{ \Phi\left(\frac{c - \mu_{D^-}}{\sigma_{D^-}}\right) - \Phi\left(\frac{c - \mu_{D^+}}{\sigma_{D^+}}\right) \right\} \end{aligned} \quad (19)$$

Donde  $\Phi$  hace referencia a la función de distribución normal estándar acumulada.

Para obtener el máximo de la función en la ecuación 19, se calcula la primera derivada e igualando a cero se obtiene:



$$c^* = \frac{(\mu_{D+}\sigma_{D-}^2 - \mu_{D-}\sigma_{D+}^2) - \sigma_{D-}\sigma_{D+} \sqrt{(\mu_{D-} - \mu_{D+})^2 + (\sigma_{D-}^2 - \sigma_{D+}^2) \log\left(\frac{\sigma_{D-}^2}{\sigma_{D+}^2}\right)}}{(\sigma_{D-}^2 - \sigma_{D+}^2)}$$

Si se asumen varianzas iguales para las distribuciones de sanos y enfermos,  $c^* = \frac{\mu_{D-} - \mu_{D+}}{2}$ , sustituyendo  $c^*$  en la ecuación 19, se pueden obtener una solución para  $J$ :

$$J = \Phi\left(\frac{c^* - \mu_{D-}}{\sigma_{D-}}\right) - \Phi\left(\frac{c^* - \mu_{D+}}{\sigma_{D+}}\right)$$

Las correspondientes estimaciones de los parámetros desconocidos ( $\mu_{D+}$   $\sigma_{D+}$   $\mu_{D-}$   $\sigma_{D-}$ ) pueden ser sustituidos para estimar  $J$  y  $c^*$  en las ecuaciones anteriores. El asumir las distribuciones normales será denominado bajo el enfoque paramétrico como el **Método Normal**. Es claro que asumir esta normalidad para valores del test puede ser cuestionable en muchas ocasiones.

Otro método menos restrictivo es asumir que hay una transformación monotonica  $t(\cdot)$ . Tal que  $t(X)$  y  $t(Y)$  son distribuidas normalmente. Dado que la curva ROC es invariante ante cualquier transformación, después de aplicar la transformación en los datos de la muestra se puede utilizar el método normal para estimar  $J_t$  y  $c_t^*$  con dichos datos transformados. Para obtener las estimaciones de los datos originales  $\hat{J}_t = \hat{J}$  y  $\hat{c}^* = t^{-1}(c_t^*)$ . La transformación  $t(\cdot)$  según diferentes autores puede ser definida mediante una transformación de potencias Box-Cox del tipo:

$$t(y) = y^{(\lambda)} = \begin{cases} \frac{(y^\lambda - 1)}{\lambda} & \lambda \neq 0 \\ \log(y) & \lambda = 0 \end{cases}$$

$\lambda$  puede ser estimado bajo máxima verosimilitud, teniendo el resultante  $\hat{\lambda}$  y usándolo en la transformación junto con el método normal quedaría entonces descrito como el **Método de Transformada Normal**.

Ruopp et al. (Ruopp, Perkins, Whitcomb, & Schisterman, 2008) plantearon también la estimación de  $J$  y  $c^*$  mediante el **Método de Máxima Verosimilitud**. Perkins et al. en un trabajo publicado en el 2007 (Perkins, Schisterman, & Vexler, 2007) abordan el problema de estimación por máxima verosimilitud teniendo en cuenta que el test tiene un límite de detección (LOD), esto hace referencia a valores del marcador que el test por debajo de este límite no pudo detectarlos o son faltantes, este LOD lo simbolizaremos como  $d$ . El enfoque de máxima verosimilitud logra parametrizar una distribución observada para estimar  $J$ . Utilizando la misma notación y asumiendo que un test cuantitativo toma valores  $x_1, x_2, x_3, \dots, x_n$  de una

variable aleatoria  $X$  e  $y_1, y_2, y_3, \dots, y_m$  de una variable aleatoria  $Y$  con distribuciones continuas  $G_{D^+}(x, \theta_x)$  e  $F_{D^-}(y, \theta_y)$  y funciones de densidad de probabilidad  $g_{D^+}(x)$  y  $f_{D^-}(y)$  para poblaciones enferma y sana respectivamente. suponemos que para los valores del test tanto en enfermos como sanos están sujetos a un LOD y son cuantificables únicamente por encima de este limite  $d$  tal que:

$$Z_{X_i} = \begin{cases} X_i & ; X_i \geq d \\ N/A & ; X_i < d; \end{cases} \quad , \quad Z_{Y_j} = \begin{cases} Y_j & ; Y_j \geq d \\ N/A & ; Y_j < d; \end{cases} \quad (20)$$

$N/A$  es observado para valores del test en niveles donde no es cuantificable. Una solución a estas observaciones no disponibles es crear estos valores con un valor específico " $a$ ", esto crea una masa de observaciones hacia ese valor específico que no son intrínsecas a la naturaleza continua de los valores del test, pero permitirían estas observaciones faltantes ser incluidas en los procedimientos de estimación.

Considerando una muestra aleatoria, (por ejemplo enfermos) y se obtienen los resultados de un test cuantitativo, correspondientes a una variable  $X \{x_i; i = 1, 2, \dots, n\}$  con distribución  $G_{D^+}(x, \theta_x)$  y parametro desconocido  $\theta_x$  y sea  $z_{X_j} j = 1, 2, 3, \dots, n$  denotando los valores mediante la transformación con la ecuación 20 y bajo la existencia de un límite de detección fijo  $d$ , donde  $k$  valores son cuantificados y  $(n - k)$  observaciones son censuradas por debajo de  $d$ . La funcion de verosimilitud para el parametro  $\theta_x$  dado la muestra de todas las  $z_{X_j}$  pueden ser vistas como una Bernoulli, donde cada valor de la observación está presente o no. Si es faltante, no podría obtenerse información más allá de su valor estando por debajo del LOD ( $d$ ). Si está presente, podemos valorar la verosimilitud de  $\theta_x$  dado el valor  $z_{X_j}$  por encima de  $d$ . colocando los  $z_{X_j}$  en orden ascendente iniciando con los  $N/A$  se puede escribir la función de verosimilitud:

$$L(\theta_X; z) \propto \prod_{j=1}^n \left[ (1 - G(d; \theta_X)) \frac{g(z_{X_j}; \theta_X)}{(1 - G(d; \theta_X))} I(z_{X_j} \neq N/A) + G(d; \theta_X) I(z_{X_j} = N/A) \right]$$

$$L(\theta_X; z) = [G(d; \theta_X)]^{n-k} \prod_{j=n-k+1}^n [g(z_{X_j}; \theta_X)] \quad (21)$$

donde  $g(x; \theta_X)$  es la funcion de densidad de probabilidad de  $X$  e  $I$  es una funcion indicadora tomando valores de 1 si es verdadero y 0 de otra manera. Para calcular el estimador máximo verosímil (MLE)  $\hat{\theta}_X$  de  $\theta_X$ , se maximiza  $L$  con respecto al parámetro o en el caso en que sea un vector de parámetros se maximiza  $L$  con respecto a cada parametro mientras se fija los demas.

Aplicando el anterior procedimiento como se usó para población enferma, de la misma forma se obtiene para la estimación máximo verosímil del parámetro de la distribución de población sana.

Teniendo los MLEs de ambas distribuciones y siendo MLE equivariante, la sustitución en las funciones correspondientes permite calcular  $\hat{f} = J(\hat{\theta}_X, \hat{\theta}_Y)$  y  $\hat{c}^* = c(\hat{\theta}_X, \hat{\theta}_Y)$ .

### 2.2.2 Métodos No paramétricos

Uno de los métodos no paramétricos más simples es usar las funciones de distribución acumuladas empíricas, como estimaciones de las funciones de distribución acumuladas de  $X$  e  $Y$ . Las funciones de distribución acumuladas empíricas pueden ser escritas como:

$$G_{D^+}(c) = \frac{1}{m} \sum_{i=1}^m I(x_i \leq c), \quad F_{D^-}(c) = \frac{1}{n} \sum_{i=1}^n I(y_i \leq c)$$

Donde  $I(u \leq c) = \begin{cases} 1 & u \leq c \\ 0 & u > c \end{cases}$

Siendo el índice de Youden estimado así:

$$\hat{J} = \max_c \{ \hat{F}_{D^-}(c) - \hat{G}_{D^+}(c) \} \quad c \in \{x_1, \dots, x_m, y_1, \dots, y_n\}$$

( 22 )

Para estimar  $c^*$  como punto de corte óptimo de un test, según Fluss (Fluss et al., 2005) se puede optar por dos enfoques:

- La observación donde el máximo fue encontrado.
- Suponer que las dos muestras son combinadas y organizadas en orden ascendente denotadas por  $d_1, \dots, d_{m+n}$  y que el máximo fue logrado en  $d_j$ . Luego el valor de  $\hat{J}$  es constante en el intervalo  $[d_j, d_{j+1})$  es razonable tomar  $\frac{d_j + d_{j+1}}{2}$  como un estimado de  $c^*$ .

Estudios de simulación han encontrado pequeñas diferencias entre los dos enfoques, con una pequeña diferencia en el último. Bajo el enfoque no paramétrico, el método anteriormente explicado corresponde al denominado **Método Empírico**.

Otro método bajo el enfoque no paramétrico es el **Método Kernel**, el cual usa una función kernel de suavización sobre las funciones de distribución acumuladas empíricas. Utilizando la función de kernel gaussiana:

$$\hat{F}_{D^-}(t) = \frac{1}{n} \sum_{i=1}^n \Phi\left(\frac{t - y_i}{h_y}\right), \quad \hat{G}_{D^+}(t) = \frac{1}{m} \sum_{i=1}^m \Phi\left(\frac{t - x_i}{h_x}\right)$$

Para controlar la cantidad de suavización se usa un ancho de la banda dada por:

$$h_y = 0.9 \min \left\{ \frac{s_y, iqr_y}{1.34} \right\} n^{-0.2}, \quad h_x = 0.9 \min \left\{ \frac{s_x, iqr_x}{1.34} \right\} m^{-0.2}$$

Donde  $s_y = \left( \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 \right)^{1/2}$  es la desviación estándar e  $iqr_y$  es el rango intercuartilico para la muestra de sujetos saludables. Se puede igualmente definir para la muestra de sujetos enfermos. De esta manera la estimación del índice de Youden por el método kernel seria:

$$\hat{J} = \max_c \{ \hat{F}_D^-(c) - \hat{G}_D^+(c) \}$$

Métodos numéricos iterativos pueden ser usados para encontrar el máximo, así como también un punto de corte optimo  $c^*$ .

Es importante aclarar que los diferentes métodos forman similares curvas ROC con máximos muy cercanos entre sí, únicamente el método normal configura una curva un poco diferente y por ende una estimación del máximo de  $J$  alejados de los demás producidos por los otros métodos.

Con relación al punto de corte  $c^*$  donde se obtiene el máximo, al obtenerse curvas distintas, este  $c^*$  es distinto. En lo concluido por Fluss et al. ningún método es superior al otro en términos de sesgo y error cuadrático medio para diferentes escenarios de distribución de los datos seleccionados en estudios de simulación. Un buen desempeño de algunos de los métodos en la estimación de  $J$  no necesariamente significa tener un buen desempeño en la estimación correspondiente de  $c^*$ .

El método empírico siendo el más usado, tuvo el peor desempeño en el estudio de simulación y no es recomendado a menos que se tenga un tamaño de muestra grande.

Cuando los datos son unimodales y con distribución normal, los métodos normal, transformado normal y kernel tienen un buen comportamiento. Para datos que no poseen una distribución normal, el método normal tiene mayor sesgo y error cuadrático medio que el transformado normal y el kernel.

El método kernel mostro ser mejor cuando el tamaño de la muestra es pequeño o cuando existe bimodalidad en los datos. En situaciones donde las poblaciones están bien separadas, detalles de la forma de las distribuciones pierden importancia y el método transformado normal provee una adecuada solución, esto a pesar del hecho que la transformación a normalidad no es efectiva para distribuciones bimodales.

En cuanto a la estimación de  $c^*$ , el método empírico y el normal mostraron ser peores que el transformado normal y el método kernel. Cuando la distribución del test es unimodal el método transformado normal tiene mejor desempeño; para distribuciones bimodales con una separación alta el kernel es mejor método y para

nivel de separación moderada el transformado normal es el procedimiento óptimo. Para niveles de separación pequeña no hubo un método preferido.

Según Fluss et al. los métodos transformado normal y kernel pueden ser preferidos como métodos de estimación puntual de  $J$  y  $c^*$ , según escenario de forma de distribución; por tanto, la construcción de intervalos de confianza para el índice de Youden y el punto de corte óptimo bajo estos dos métodos serían necesarios.

### 2.3 Intervalos de confianza para el índice de Youden

En la sección anterior se discutió y se presentaron diferentes métodos de estimación puntual del índice de Youden, de esta manera la inferencia se realizaría con un solo valor puntual sobre el parámetro. Para subsanar esta desventaja revisaremos diferentes intervalos de confianza que han sido propuestos por diferentes autores bajo diferentes metodologías y considerando cada vez más matices en el problema que lo hacen un tópico interesante.

#### a) Intervalo de confianza de W. Youden

Youden (Youden, 1950) propuso un intervalo de confianza basado en que su índice  $J$ , asumiendo que sensibilidad y especificidad son proporción binomiales de grupos independientes, para este caso tomando los resultados en la forma de la tabla de contingencia de  $2 \times 2$  de la tabla 1 y siendo  $a \rightarrow B(F_1 = a + b, Se)$  y  $d \rightarrow B(F_2 = c + d, Sp)$ :

$$\widehat{Se} = \frac{a}{a + b}$$

Siendo para la varianza de una proporción binomial  $p(1 - p)/n$ , en el caso de la sensibilidad se tiene:

$$Var(\widehat{Se}) = \frac{\frac{a}{a+b} \left(1 - \frac{a}{a+b}\right)}{a+b} = \frac{\frac{ab}{(a+b)^2}}{(a+b)}$$

$$Var(\widehat{Se}) = \frac{ab}{(a+b)^3}$$

( 23 )

Para la especificidad se aplica el mismo procedimiento obteniendo:

$$Var(\widehat{Sp}) = \frac{cd}{(c+d)^3}$$

( 24 )

Por consiguiente, la varianza para  $\hat{J}$  sería la suma de varianzas:

$$Var(\hat{J}) = \frac{ab}{(a+b)^3} + \frac{cd}{(c+d)^3}$$

( 25 )

y el error estándar de  $\hat{J}$ :

$$\hat{\sigma}_J = \sqrt{Var(\hat{J})} = \sqrt{\frac{ab}{(a+b)^3} + \frac{cd}{(c+d)^3}}$$

( 26 )

El intervalo de confianza al  $(1 - \alpha)100\%$  para  $J$ :

$$\hat{J} \pm z_{1-\alpha/2} \sqrt{\frac{ab}{(a+b)^3} + \frac{cd}{(c+d)^3}}$$

( 27 )

#### b) Intervalos de confianza paramétricos de Shcisterman y Perkins asumiendo distribución normal y gamma

Uno de los primeros en plantear la construcción de intervalos de confianza para el índice de Youden fue Schisterman y Perkins (Schisterman & Perkins, 2007), quienes basados en la distribución normal y gamma dedujeron intervalos de confianza.

Los mencionados autores asumieron que los resultados del test en casos  $X$  y controles  $Y$  son distribuidos de manera **Normal** tal que  $X_i \sim N(\mu_X, \sigma_X^2); i = 1, \dots, n_X$  y  $Y_j \sim N(\mu_Y, \sigma_Y^2); j = 1, \dots, n_Y$  con  $\mu_X < \mu_Y$  siendo:

$$J = \Phi\left(\frac{\mu_X - c}{\sigma_X}\right) - \Phi\left(\frac{c - \mu_Y}{\sigma_Y}\right) - 1$$

( 28 )

Donde  $\Phi$  denota la función de distribución normal estándar acumulada. Cuando los parámetros son desconocidos los estimados  $\hat{J}$  y  $\hat{c}$  son encontrados con la sustitución de las medias y desviaciones estándar muestrales y luego la ecuación **22** puede ser utilizada para el caso de varianzas iguales. Luego  $\hat{J}$  y  $\hat{c}$  son funciones de estimadores máximos verosímiles, ellos por si mismos son estimadores máximos verosímiles de  $J$  y  $c$ , por tanto, los resultantes estimadores son asintóticamente insesgados y normalmente distribuidos. Un Intervalo de confianza al  $(1 - \alpha)100\%$  para  $J$  tendría la forma:

$$\hat{J} \pm z_{\alpha/2} \sqrt{\text{Var}(\hat{J})}$$

( 29 )

En la ecuación **29** la varianza de  $\hat{J}$  es desconocida. De esta manera se puede tener una aproximación como la realizada por Schisterman utilizando el método delta y construyendo el intervalo de confianza aproximado al  $(1 - \alpha)100\%$  (Schisterman & Perkins, 2007):

$$\hat{J} \pm z_{\alpha/2} \sqrt{\widehat{\text{Var}}(\hat{J})}$$

( 30 )

$z_{\alpha/2}$  es el percentil de  $\alpha/2$  percentil de la distribución normal

$$\text{Var}(\hat{J}) \approx \left( \frac{\partial J}{\partial \mu_X} \right)^2 \text{Var}(\hat{\mu}_X) + \left( \frac{\partial J}{\partial \mu_X} \right)^2 \text{Var}(\hat{\sigma}_X) + \left( \frac{\partial J}{\partial \mu_Y} \right)^2 \text{Var}(\hat{\mu}_Y) + \left( \frac{\partial J}{\partial \mu_Y} \right)^2 \text{Var}(\hat{\sigma}_Y)$$

( 31 )

donde las varianzas de  $\hat{\mu}_X$  y  $\hat{\mu}_Y$  son  $\sigma_X^2/n_X$  y  $\sigma_Y^2/n_Y$  respectivamente. Y las varianzas de  $\hat{\sigma}_X$  y  $\hat{\sigma}_Y$ , obtenidas mediante el método delta son:

$$\text{Var}(\hat{\sigma}_X) = \frac{\sigma_X^2}{2(n_X - 1)} ; \text{Var}(\hat{\sigma}_Y) = \frac{\sigma_Y^2}{2(n_Y - 1)}$$

( 32 )

Obteniendo las derivadas parciales de  $J$  con respecto a los parámetros, reemplazando junto con las varianzas de  $\hat{\mu}_X$ ,  $\hat{\mu}_Y$ ,  $\hat{\sigma}_X$ ,  $\hat{\sigma}_Y$  en la ecuación **31** se tiene la varianza aproximada con la cual se puede calcular el intervalo de confianza en la ecuación **30**.

$$\begin{aligned}
\widehat{Var}(\hat{f}) \approx & \frac{S_X^2}{n_X} \left[ \phi(\hat{z}_X) + (\phi(\hat{z}_Y) - \phi(\hat{z}_X)) \left( \frac{-1 \pm \hat{a}\hat{b}(rad)^{-1/2}}{\hat{b}^2 - 1} \right) \right]^2 \\
& + (-1) \frac{S_Y^2}{n_Y} \left[ \phi(\hat{z}_Y) + (\phi(\hat{z}_Y) - \phi(\hat{z}_X)) \left( \frac{\hat{b}^2 \pm (-1)\hat{a}\hat{b}(rad)^{-1/2}}{\hat{b}^2 - 1} \right) \right]^2 \\
& + \frac{1}{2(n_X - 1)} \left[ \hat{z}_X \phi(\hat{z}_X) + \frac{\hat{b}\phi(\hat{z}_Y) - \phi(\hat{z}_X)}{(\hat{b}^2 - 1)^2 (S_Y^2)^{1/2}} \right. \\
& * \left. \left( 2\hat{a}\hat{b}^2 \pm \left( (-\hat{b}^2 - 1)(rad)^{1/2} + S_X^2(\hat{b}^2 - 1)(rad)^{-1/2} (\ln(\hat{b}^2) + 1 - \hat{b}^{-2}) \right) \right) \right]^2 \\
& + \frac{-1}{2(n_Y - 1)} \left[ \hat{z}_Y \phi(\hat{z}_Y) + \frac{(\hat{b}\phi(\hat{z}_Y) - \phi(\hat{z}_X))}{(\hat{b}^2 - 1)^2 (S_Y^2)^{1/2}} \right. \\
& * \left. \left( 2\hat{a}\hat{b}^2 \pm \left( (-\hat{b}^2 - 1)(rad)^{1/2} + S_Y^2(\hat{b}^2 - 1)(rad)^{-1/2} (\ln(\hat{b}^2) + \hat{b}^2 - 1) \right) \right) \right]^2
\end{aligned}
\tag{33}$$

Donde:

$$a = \mu_X - \mu_Y ; \quad b = \frac{\sigma_X}{\sigma_Y}$$

$$rad = a^2 + (b^2 - 1)\sigma_Y^2 \ln(b^2)$$

$$z_X = \frac{\mu_X - c}{\sigma_X} ; \quad z_Y = \frac{\mu_Y - c}{\sigma_Y}$$

La ecuación 33 de aproximación a la varianza de  $\hat{f}$  aplica cuando las varianzas son desiguales. Cuando estas se asumen ser iguales,  $\hat{b} = 1$ , la formula de varianza aproximada de  $\hat{f}$  seria:

$$\begin{aligned}
\widehat{Var}(\hat{f}) \approx & \left( \frac{S_X^2}{n_X} \right) \left[ \phi(\hat{z}_X) + \frac{1}{2}(\phi(\hat{z}_Y) - \phi(\hat{z}_X)) \right]^2 - \left( \frac{S_Y^2}{n_Y} \right) \left[ \phi(\hat{z}_Y) + \frac{1}{2}(\phi(\hat{z}_Y) - \phi(\hat{z}_X)) \right]^2 \\
& + \frac{1}{2(n_X - 1)} [\hat{z}_X \phi(\hat{z}_X)]^2 - \frac{1}{2(n_Y - 1)} [\hat{z}_Y \phi(\hat{z}_Y)]^2
\end{aligned}
\tag{34}$$

Sustituyendo los respectivos valores necesarios tanto en la ecuación 33 y 34 según el caso de varianzas iguales o desiguales respectivamente, se puede obtener un estimado aproximado de la varianza de  $\hat{f}$  y usarla en la ecuación 30 para obtener un intervalo de confianza aproximado para  $J$ .



Schisterman et al. también plantearon intervalos mediante el método delta asumiendo distribuciones **Gamma**, tal que  $X_j \sim \Gamma(\alpha_X, \beta_Y) j = 1, \dots, n_X$  y  $Y_i \sim \Gamma(\alpha_Y, \beta_Y) i = 1, \dots, n_Y$  tanto para enfermos como en sanos respectivamente.

$$\Gamma(x, \alpha, \beta) = \frac{e^{-x/\beta} x^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} \quad (35)$$

Donde  $\alpha > 0$  y  $\beta > 0$  corresponden a los parámetros de pendiente y escala, respectivamente, y  $\Gamma(\alpha)$  es la función gamma.  $J$  estaría estimado por la intersección de estas dos funciones de densidad de probabilidad. Así:

$$J = \int_c^\infty \frac{e^{-x/\beta_X} x^{\alpha_X-1}}{\beta_X^{\alpha_X} \Gamma(\alpha_X)} dx + \int_c^\infty \frac{e^{-y/\beta_Y} y^{\alpha_Y-1}}{\beta_Y^{\alpha_Y} \Gamma(\alpha_Y)} dy - 1 \quad (36)$$

Sustituyendo por sus estimadores máximos verosímiles  $\hat{\alpha}_X, \hat{\beta}_X, \hat{\alpha}_Y, \hat{\beta}_Y$  se puede estimar  $\hat{f}$ . Un intervalo de confianza al  $(1 - \alpha)100\%$  asumiendo la distribución gamma se puede obtener con la ecuación 30. De igual manera que para el caso normal la varianza de  $\hat{f}$  es desconocida, por tal motivo se hace una aproximación mediante el método delta sustituyendo los estimadores máximos verosímiles de:  $\hat{\alpha}_X, \hat{\beta}_X, \hat{\alpha}_Y, \hat{\beta}_Y$ :

$$\begin{aligned} Var(\hat{f}) \approx & \left( \frac{\partial J}{\partial \alpha_X} \right)^2 Var(\hat{\alpha}_X) + \left( \frac{\partial J}{\partial \beta_X} \right)^2 Var(\hat{\beta}_X) + \left( \frac{\partial J}{\partial \alpha_Y} \right)^2 Var(\hat{\alpha}_Y) + \left( \frac{\partial J}{\partial \beta_Y} \right)^2 Var(\hat{\beta}_Y) \\ & + \left( \frac{\partial J}{\partial \alpha_X} \right) \left( \frac{\partial J}{\partial \beta_X} \right) Cov(\hat{\alpha}_X, \hat{\beta}_X) + \left( \frac{\partial J}{\partial \alpha_Y} \right) \left( \frac{\partial J}{\partial \beta_Y} \right) Cov(\hat{\alpha}_Y, \hat{\beta}_Y) \end{aligned} \quad (37)$$

Las varianzas y covarianza que involucran estimadores son:

$$\begin{aligned} Var(\hat{\alpha}_X) &= \frac{\alpha_X}{\beta_X^2 * \det(\Sigma_X^{-1}) * n_X} \\ Var(\hat{\beta}_X) &= \frac{d^2 \ln \Gamma(\alpha_X)}{d\alpha_X^2} * \frac{1}{\det(\Sigma_X^{-1}) * n_X} \\ Cov(\hat{\alpha}_X, \hat{\beta}_X) &= \frac{-1}{\beta_X^2 * \det(\Sigma_X^{-1}) * n_X} \end{aligned}$$

Donde  $\det(\Sigma_X^{-1}) = \left( \frac{d^2 \ln \Gamma(\alpha_X)}{d\alpha_X^2} * \alpha_X - 1 \right) * \beta_X^{-2}$ .

Reemplazando en la ecuación 37 las derivadas parciales de  $J$  además de las varianzas y covarianzas se obtiene la formula explicita de aproximación de la varianza de  $\hat{f}$ . Para más detalle en la derivadas parciales se puede revisar el apéndice 3 del artículo original de Schisterman y Perkins (Schisterman & Perkins, 2007).

En estudios de simulación Schisterman evaluó los intervalos de confianza aproximados mediante el método delta, comparándolos extensivamente con intervalos calculados mediante métodos de bootstrap. Sus resultados mostraron que la cobertura de probabilidades es más cercana al valor nominal; para el mismo porcentaje de cobertura los intervalos de confianza bootstrap son aproximadamente dos a tres veces más amplios que los encontrados por el método delta. Sin embargo, los intervalos derivados por este método son dependientes sobre los supuestos de distribución, supuestos que no son de ninguna manera no razonables, pero que siempre deben ser revisados porque violaciones a los mismos pueden llevar a un sustancial sesgo en la estimación. Cuando hay alta sospecha del no cumplimiento de los supuestos en la distribución los intervalos de confianza bootstrap son más robustos que los calculados por el método delta y proveen una mejor cobertura de probabilidades.

**c) Intervalos de confianza no paramétricos de Zhou y Qin.**

Siguiendo con el razonamiento anterior sobre la violación de los supuestos de distribución, Zho et al. (Zhou & Qin, 2012) publicaron nuevos intervalos de confianza **No paramétricos (NB)**, justificándose en la desventaja que tienen los intervalos paramétricos al ser dependientes del supuesto de distribución, aunado a la complejidad de las expresiones para su cálculo, lo cual los harían poco aplicables en la práctica. Adicionalmente, teniendo en cuenta que Schisterman y Perkins utilizaron intervalos no paramétricos (Bootstrap) y los compararon con los paramétricos, sin tener en cuenta que los métodos bootstrap utilizados (BP y BPca) producen intervalos más amplios y con una menor cobertura de probabilidades.

Zhou et al. reescribieron el índice de Youden como la diferencia de dos proporciones binomiales:

$$Se = P(X \geq c) ; Sp = P(Y \leq c)$$

$$J = P(X \geq c) + P(Y \leq c) - 1 = P(Y \leq c) - [1 - P(X \geq c)]$$

$$J = P(Y \leq c) - P(X < c)$$

Su estimador empírico considerando un punto de corte optimo  $c^*$  sería:

$$\hat{J} = \frac{\sum_{j=1}^m I(Y_j \leq c^*)}{m} - \frac{\sum_{i=1}^n I(X_i < c^*)}{n}$$

( 38 )

De esta manera el estimado empírico de  $J$  sería la diferencia entre dos proporciones binomiales y las funciones indicadoras  $I(Y_j \leq c^*), j = 1, \dots, m$  y  $I(X_i < c^*), i = 1, \dots, n$  son dependientes. Por lo tanto, dejarían de ser una simple diferencia de dos proporciones binomiales y los métodos usuales para construir intervalos de confianza no son aplicables.

Basados en un método denominado AC intervalo, el cual ajusta la diferencia de dos proporciones binomiales Zho et al. plantearon un nuevo procedimiento para calcular los intervalos de confianza para  $J$ . El nuevo estimado ajustado del índice de Youden consistiría en sumar dos éxitos y 2 fallas a las observaciones Bernoulli:

$$\hat{J}_{AC} = \frac{\sum_{j=1}^m I(Y_j \leq c^*) + z_{1-\alpha/2}^2 / 2}{m + z_{1-\alpha/2}^2} - \frac{\sum_{i=1}^n I(X_i < c^*) + z_{1-\alpha/2}^2 / 2}{n + z_{1-\alpha/2}^2} \quad (39)$$

Luego  $z_{1-\alpha/2}$  es aproximadamente 2 cuando  $\alpha=0.05$ .

Dado que  $\hat{J}_{AC}$  es también la diferencia entre dos proporciones correlacionadas, la estimación de su varianza es difícil, por tanto Zhou (Zhou & Qin, 2012) plantearon el siguiente procedimiento para estimar la varianza de  $\hat{J}_{AC}$  mediante bootstrap:

1. Se extrae una remuestra de tamaño  $n$   $X_i^{*'}s$  con reemplazo de una muestra de enfermos  $X_i's$  y una remuestra de tamaño  $m$ ,  $Y_j^{*'}s$  con reemplazo de una muestra de sanos  $Y_j's$ .
2. Se calcula la versión bootstrap de  $\hat{J}_{AC}$  donde  $c^{**}$  es la versión bootstrap de  $c^*$ :

$$\hat{J}_{AC}^* = \frac{\sum_{j=1}^m I(Y_j^* \leq c^{**}) + z_{1-\alpha/2}^2 / 2}{m + z_{1-\alpha/2}^2} - \frac{\sum_{i=1}^n I(X_i^* < c^{**}) + z_{1-\alpha/2}^2 / 2}{n + z_{1-\alpha/2}^2} \quad (40)$$

3. Se repite el paso 1 y 2  $B$  veces hasta obtener replicas bootstrap  $\{\hat{J}_{AC}^{*b} \mid b = 1, 2, \dots, B\}$  se sugiere que  $B \geq 200$ , luego se calcula la varianza del estimador  $V^*(\hat{J}_{AC})$ :

$$V^*(\hat{J}_{AC}) = \frac{1}{B-1} \sum_{b=1}^B (\hat{J}_{AC}^{*b} - \bar{J}_{AC}^*)^2 \quad (41)$$

Donde  $\bar{J}_{AC}^* = \frac{1}{B} \sum_{b=1}^B \hat{J}_{AC}^{*b}$ .

El intervalo de confianza al nivel  $1 - \alpha$ )100% seria:

$$\left( \bar{J}_{AC}^* - z_{1-\alpha/2} \sqrt{V^* \hat{J}_{AC}} , \quad \bar{J}_{AC}^* + z_{1-\alpha/2} \sqrt{V^* \hat{J}_{AC}} \right) \quad (42)$$

Como subproducto del anterior intervalo se puede obtener mediante el bootstrap percentiles (BP) un intervalo para el estimado ajustado  $\hat{J}_{AC}$ , este es denominado el intervalo **BPAC** y se construye como sigue:

$$\left( \hat{J}_{AC}^{*([B\alpha/2])} , \quad \hat{J}_{AC}^{*[B(1-\frac{\alpha}{2})]} \right) \quad (43)$$

Donde  $\hat{J}_{AC}^{*([B\alpha/2])}$  y  $\hat{J}_{AC}^{*[B(1-\frac{\alpha}{2})]}$  son el  $\alpha/2$ -iesimo y  $(1-\alpha/2)$ -iesimo cuantil de  $\hat{J}_{AC}^{*1}, \hat{J}_{AC}^{*2}, \hat{J}_{AC}^{*3}, \dots, \hat{J}_{AC}^{*B}$  respectivamente.

Según los estudios de simulación realizados por Zho, el intervalo NB tiene comportamiento competitivo contra los intervalos delta paramétricos, sin embargo, cuando el índice de Youden está en su frontera ( $J \geq 0.9$ ) y el tamaño de muestra es pequeño, los intervalos paramétricos tienen mejor desempeño, esto es lógico debido a que los intervalos bootstrap están basados en un enfoque fuertemente no paramétrico (sin supuestos en la distribución de los resultados del test en sanos y enfermos). De otra manera la validez de los intervalos paramétricos está sujetos fuertemente al cumplimiento de los supuestos distribucionales. También cuando el índice de Youden está en un valor de moderado a alto ( $J \geq 0.8$ ) el intervalo BPAC es competitivo contra el NB y mucho mejor que el no paramétrico BP propuesto por Shisterman y Perkins (Schisterman & Perkins, 2007).

#### d) Intervalos de confianza de Shan (The Wilson score method – Square-and-add method)

Shan (Shan, 2015) basado en los intervalos de confianza propuestos por Zho y Qin (Zhou & Qin, 2012), planteo una mejora a la estimación de la diferencia de dos proporciones binomiales ajustada (denominada AC) de  $J$ , usando el método Wilson Score para la construcción del intervalo de confianza para cada proporción y el método square-and-add para construir el intervalo de confianza para el índice de Youden.

Al igual que Zho, se asumió que el índice  $\hat{f}$  puede ser reescrito como la diferencia entre dos proporciones binomiales (ecuación 38) o bajo el método de estimado ajustado de proporción  $\hat{f}_{AC}$  (suma de dos éxitos y dos fallos, ecuación 39), de esta manera los límites del intervalo de confianza Wilson  $(i_1, s_1)$  para la proporción  $p_1 = P(Y \leq c^*)$  corresponden a las raíces de la siguiente ecuación:

$$(p_1 - \hat{p}_1)^2 = z_{1-\alpha/2}^2 \frac{p_1(1-p_1)}{m} \quad (44)$$

sus raíces son:

$$\begin{aligned} i_1 &= \frac{1}{1 + z_{1-\alpha/2}^2/m} \left[ \hat{p}_1 + \frac{z_{1-\alpha/2}^2}{2m} - z_{1-\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{m} + \frac{z_{1-\alpha/2}^2}{4m^2}} \right] \\ s_1 &= \frac{1}{1 + z_{1-\alpha/2}^2/m} \left[ \hat{p}_1 + \frac{z_{1-\alpha/2}^2}{2m} + z_{1-\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{m} + \frac{z_{1-\alpha/2}^2}{4m^2}} \right] \end{aligned} \quad (45)$$

De igual manera los límites del intervalo de confianza Wilson  $(i_2, s_2)$  para la proporción  $p_2 = P(X < c^*)$  son:

$$(p_2 - \hat{p}_2)^2 = z_{1-\alpha/2}^2 \frac{p_2(1-p_2)}{n} \quad (46)$$

$$\begin{aligned} i_2 &= \frac{1}{1 + z_{1-\alpha/2}^2/n} \left[ \hat{p}_2 + \frac{z_{1-\alpha/2}^2}{2n} - z_{1-\alpha/2} \sqrt{\frac{\hat{p}_2(1-\hat{p}_2)}{n} + \frac{z_{1-\alpha/2}^2}{4n^2}} \right] \\ s_2 &= \frac{1}{1 + z_{1-\alpha/2}^2/n} \left[ \hat{p}_2 + \frac{z_{1-\alpha/2}^2}{2n} + z_{1-\alpha/2} \sqrt{\frac{\hat{p}_2(1-\hat{p}_2)}{n} + \frac{z_{1-\alpha/2}^2}{4n^2}} \right] \end{aligned} \quad (47)$$

El intervalo de confianza para  $\hat{f}$ :

$$\left( \hat{f} - z_{1-\alpha/2} \sqrt{\frac{i_1(1-i_1)}{m} + \frac{s_2(1-s_2)}{n}}, \hat{f} + z_{1-\alpha/2} \sqrt{\frac{s_1(1-s_1)}{m} + \frac{i_2(1-i_2)}{n}} \right)$$

las estimaciones de  $\hat{p}_1$  y  $\hat{p}_2$  cuando se estiman por la ecuación 38 Shan los denomino **método NP** y estima  $\hat{f}$  y si se utiliza la ecuación 39 los denomino intervalos método **NPAC** y estima  $\hat{f}_{AC}$ , esa misma notacion sera utilizada en este trabajo.

En los estudios de simulación realizados por Shan, comparo el desempeño de los intervalos NP , NPAC y el Bootstrap AC (BAC) propuesto por Zho y Qin. La comparación fue en términos de cobertura de probabilidades y amplitud del intervalo; 5000 muestras en población enferma y sana con distribuciones normal y gamma fueron simuladas,  $B = 500$  muestras bootstrap fueron generadas para el metodo BAC. Para los métodos NP y NPAC no se generaron muestras bootstrap, ya que computacionalmente es simple de aplicar dichas expresiones, siendo esta una de las ventajas en términos de aplicabilidad a la práctica real por parte de este tipo de estimación por intervalo.

Los resultados tanto con distribución normal como gamma fueron similares mostrando que las coberturas de probabilidad, definidas como la proporción de veces que los intervalos generados contienen el valor del índice de Youden predefinido, son frecuentemente más bajas que el nivel nominal para el método BAC y esto ocurre más cuando el tamaño de muestra es pequeño y el índice  $J$  es alto. Los métodos mejorados por Shan, NP y NPAC muestran mejor cobertura de probabilidades que el BAC. NPAC no se desempeña muy bien como el NP para índices  $J$  altos, hasta de 0.9, en tales casos la cobertura de probabilidades del NPAC puede ser hasta de 57.3%. NPAC podría tener mejor cobertura de probabilidad que NP cuando el índice  $J$  es pequeño, pero con una amplitud del intervalo mayor. En general el NP tuvo mejor desempeño que los otros dos métodos.

Shan también realizo comparación entre métodos asumiendo que no necesariamente las distribuciones de enfermos y sano fueran las mismas. el combino distribución t-student - normal y normal-gamma, sin embargo, el método NP mostro mejor desempeño comparado con los demás métodos. En conclusión frente a los dos nuevos intervalos propuestos por Shan, se encontró que los métodos NP y NPAC presentan mejor cobertura de probabilidad que el BAC, el desempeño del NPAC depende sobre el valor de  $J$ , mientras que el NP mostro ser robusto, adicinalmente el NP es mas facil de calcular que el BAC, ya que no requiere de muestras bootstrap, por tanto el NP es recomendado para su aplicación en la práctica.

#### e) Intervalo de confianza de Chen

Chen et al. (Chen, Xue, Tan, & Chen, 2015), plantearon un intervalo de confianza desde la perspectiva de que la sensibilidad y la especificidad no son independientes, si bien se obtienen de muestras independientes (sano y enfermos), su dependencia está ligada al error de mala clasificación, cuando este no

existe (test perfecto) su dependencia desaparece. De hecho, esto puede ser evidenciado cuando se grafica la curva ROC, la sensibilidad y la especificidad están negativamente asociadas.

Asumiendo de nuevo una muestra aleatoria con la aplicación de gold estandar se obtiene los resultados como en la tabla 1 (tabla de contingencia  $2 \times 2$ ) y  $J$  puede ser estimado asumiendo independencia entre sensibilidad y especificidad como lo propuso W. Youden, y puede verse en la expresión 25

$$Var(\hat{J}) = Var(\widehat{Se} + \widehat{Sp} - 1) = \frac{ab}{(a+b)^3} + \frac{cd}{(c+d)^3}$$

dado que la sensibilidad y especificidad no son independientes su verdadera varianza estaría dada por:

$$Var(\hat{J}) = Var(\widehat{Se} + \widehat{Sp} - 1) = Var(\widehat{Se}) + Var(\widehat{Sp}) + 2Cov(\widehat{Se}, \widehat{Sp}) \quad (49)$$

utilizando el método delta se puede estimar la varianza de la suma de dos proporciones relacionadas:

$$Var(X + Y) = Y^2 VarX + X^2 Var(Y)$$

$$Var(\hat{J}) = (\widehat{Sp})^2 Var(\widehat{Se}) + (\widehat{Se})^2 Var(\widehat{Sp})$$

$$Var(\hat{J}) = \left(\frac{d}{c+d}\right)^2 \left(\frac{ab}{(a+b)^3}\right) + \left(\frac{a}{a+b}\right)^2 \left(\frac{cd}{(c+d)^3}\right) \quad (50)$$

Siendo su error estándar:

$$\hat{\sigma}_J = \sqrt{Var(\hat{J})} = \sqrt{\left(\frac{d}{c+d}\right)^2 \left(\frac{ab}{(a+b)^3}\right) + \left(\frac{a}{a+b}\right)^2 \left(\frac{cd}{(c+d)^3}\right)} \quad (51)$$

Teniendo la varianza de  $\hat{J}$  se construye un intervalo de confianza al  $(1 - \alpha)100\%$  para el índice  $J$  de Youden:

$$\hat{J} \pm z_{1-\alpha/2} \sqrt{\left(\frac{d}{c+d}\right)^2 \left(\frac{ab}{(a+b)^3}\right) + \left(\frac{a}{a+b}\right)^2 \left(\frac{cd}{(c+d)^3}\right)}$$

( 52 )

donde  $\alpha$  es el nivel de significancia y  $z_{1-\alpha/2}$  es 100(1 -  $\alpha/2$ )- iesimo percentil de la distribución normal estándar.

Según la simulaciones realizadas por Chen, este método de calcular intervalos de confianza para el índice de Youden asumiendo sensibilidad y especificidad dependientes, genera varianzas más pequeñas que el método descrito por Youden (Youden, 1950), por tanto muestra ser un método más potente que el original. Comparándolo también con intervalos bootstrap, la diferencias en la varianza estimada esta alrededor de 0.146%, significando que el método asumiendo sensibilidad y especificidad dependientes tiene intervalos muy parecidos a los obtenidos por bootstrap.

Chen también planteo contrastes de hipótesis para diferentes situaciones basados en la dependencia entre sensibilidad y especificidad, los cuales son revisados a continuación.

## 2.4 Contraste de hipótesis para un índice de Youden

Dentro del trabajo de Chen, basado en asumir la dependencia entre sensibilidad y especificidad, propone el contraste de hipótesis para un índice de Youden obtenido en una muestra aleatoria en la cual a partir de un gold estandar se obtienen enfermos y sanos y a la misma muestra se obtienen los resultados del nuevo test al cual se pretende evaluar su exactitud diagnostica.

Sea:

$$H_0 \equiv J = J_0$$

$$H_1 \equiv J \neq J_0$$

Donde  $J$  es el parametro o valor verdadero de indice en cuestion y  $J_0$  corresponde a un valor umbral dado. Acorde al teorema del limite central, el estadístico de prueba para comparar el índice de Youden muestral con el parámetro conocido  $J_0$  seleccionado por los investigadores o por estudios previos sera:

$$Z = \frac{\hat{J} - J_0}{\hat{\sigma}_J} = \frac{\hat{J} - J_0}{\sqrt{\left(\frac{d}{c+d}\right)^2 \left(\frac{ab}{(a+b)^3}\right) + \left(\frac{a}{a+b}\right)^2 \left(\frac{cd}{(c+d)^3}\right)}}$$

( 53 )

Es de notar que el error estándar utilizado en el estadístico de prueba es teniendo en cuenta la dependencia entre sensibilidad y especificidad, el mismo utilizado en la construcción de intervalo de confianza para un Índice de Youden descrito en la sección anterior.



Los términos de  $\widehat{Se}^2, \widehat{Sp}^2$  tiene como valores máximos 1, la varianza es siempre más pequeña que la calculada cuando sensibilidad y especificidad se toman como independientes, por tanto, el test propuesto es más potente que el planteado originalmente por Youden.

## 2.5 Comparación de dos índices de Youden con muestras independientes

El siguiente test aplica para comparar dos índices de Youden calculados en muestras independientes, con el fin de comparar el desempeño del mismo test en dos poblaciones diferentes (Chen et al., 2015).

Sea:

$$H_0 \equiv J_1 = J_2$$

$$H_1 \equiv J_1 \neq J_2$$

Donde  $J_1$  y  $J_2$  son los parametros, los valores verdaderos de los índices de Youden de dos muestras independientes. Luego la varianza para la diferencias entre estos dos índices independientes es (Chen et al., 2015):

$$Var(\hat{J}_1 - \hat{J}_2) = Var(\hat{J}_1) + Var(\hat{J}_2) \quad (54)$$

aplicando la ecuación 51 para cada índice de Youden que conforman la diferencia, quedaría:

$$Var(\hat{J}_1 - \hat{J}_2) = \sqrt{\left(\frac{d_1}{c_1+d_1}\right)^2 \left(\frac{a_1 b_1}{(a_1+b_1)^3}\right) + \left(\frac{a_1}{a_1+b_1}\right)^2 \left(\frac{c_1 d_1}{(c_1+d_1)^3}\right)} + \sqrt{\left(\frac{d_2}{c_2+d}\right)^2 \left(\frac{a b}{(a+b)^3}\right) + \left(\frac{a}{a+b}\right)^2 \left(\frac{c d}{(c+d)^3}\right)} \quad (55)$$

De nuevo aplicando el teorema del límite central el estadístico de prueba siendo  $d = J_1 - J_2$ :

$$Z = \frac{\hat{J}_1 - \hat{J}_2}{\hat{\sigma}_{D_J}} = \frac{\hat{J}_1 - \hat{J}_2}{\sqrt{Var(\hat{J}_1 - \hat{J}_2)}} \quad (56)$$

De igual manera con la ecuación 55 se puede construir un intervalo de confianza  $(1 - \alpha/2)100\%$  para la diferencia de dos índices de Youden:

$$d \pm Z_{\alpha/2} \sqrt{Var(\hat{J}_1 - \hat{J}_2)} \quad (57)$$

## 2.6 Comparación de dos índices de Youden con muestras apareadas

En el diseño pareado, el desarrollo de la aplicación de la muestra se puede realizar de dos maneras: 1) cuando el mismo test es aplicado al enfermo y al control con el cual esta pareado y 2) dos test diferentes son aplicados a la misma muestra de pacientes. Para estos dos casos se puede plantear un contraste de hipótesis (Chen et al., 2015).

Sea:

$$H_0 \equiv J_1 = J_2$$

$$H_1 \equiv J_1 \neq J_2$$

donde  $J_1$  y  $J_2$  son los valores verdaderos del índice de Youden de los dos grupos relacionados. Si se toma en cuenta la asociación entre sensibilidad y especificidad como también la de los test pareados, la varianza del índice de Youden pareado es:

$$\begin{aligned} Var(\hat{J}_1 - \hat{J}_2)^* &= Var(\hat{J}_1) - Var(\hat{J}_2) - 2Cov(\hat{J}_1, \hat{J}_2) \\ (58) \end{aligned}$$

Las varianzas  $Var(\hat{J}_1)$  y  $Var(\hat{J}_2)$  pueden ser obtenidas con la ecuación 51 y la covarianza entre dos índices de Youden así:

$$\begin{aligned} Cov(\hat{J}_1, \hat{J}_2) &= Cov(\widehat{Se}_1 + \widehat{Sp}_1 - 1, \widehat{Se}_2 + \widehat{Sp}_2 - 1) \\ &= Cov(\widehat{Se}_1, \widehat{Se}_2) + Cov(\widehat{Sp}_1, \widehat{Sp}_2) + Cov(\widehat{Se}_1, \widehat{Sp}_2) + Cov(\widehat{Se}_2, \widehat{Sp}_1) \\ (59) \end{aligned}$$

Entonces la varianza entre  $\hat{J}_1 - \hat{J}_2$  es:

$$\begin{aligned} Var(\hat{J}_1 - \hat{J}_2)^* &= Var(\widehat{Se}_1) + Var(\widehat{Sp}_1) + Var(\widehat{Se}_2) + Var(\widehat{Sp}_2) \\ &\quad - 2[Cov(\widehat{Se}_1, \widehat{Se}_2) + Cov(\widehat{Sp}_1, \widehat{Sp}_2) + Cov(\widehat{Se}_1, \widehat{Sp}_2) + Cov(\widehat{Se}_2, \widehat{Sp}_1) \\ &\quad - Cov(\widehat{Se}_1, \widehat{Sp}_1) - Cov(\widehat{Se}_2, \widehat{Sp}_2)] \\ (60) \end{aligned}$$

Luego la ecuación 60 puede reescribirse así:

$$\begin{aligned} Var(\hat{J}_1 - \hat{J}_2)^* &= V - 2c - 2[Cov(\widehat{Se}_1, \widehat{Se}_2) + Cov(\widehat{Sp}_1, \widehat{Sp}_2)] \\ (61) \end{aligned}$$

siendo:

$$V = Var(\widehat{Se}_1) + Var(\widehat{Sp}_1) + Var(\widehat{Se}_2) + Var(\widehat{Sp}_2)$$

( 62 )

$$c = Cov(\widehat{Se}_1, \widehat{Sp}_2) + Cov(\widehat{Se}_2, \widehat{Sp}_1) - Cov(\widehat{Se}_1, \widehat{Sp}_1) - Cov(\widehat{Se}_2, \widehat{Sp}_2)$$

( 63 )

Dado lo anterior,  $V$  puede ser calculado por la ecuación de varianza para sensibilidad y especificidad propuesta originalmente por Youden, mientras que  $c$  que corresponde a la covarianza de dos sensibilidades y 2 especificidades pueden ser estimadas a través de un coeficiente de contingencia para dos proporciones relacionadas:

$$Cov(\widehat{Se}_1, \widehat{Se}_2) = \hat{r}_{Se} \sqrt{Var(\widehat{Se}_1)} \sqrt{Var(\widehat{Se}_2)}$$

$$Cov(\widehat{Sp}_1, \widehat{Sp}_2) = \hat{r}_{Sp} \sqrt{Var(\widehat{Sp}_1)} \sqrt{Var(\widehat{Sp}_2)}$$

Donde  $\hat{r}_{Se}$  representa el coeficiente de asociación entre las sensibilidades en el grupo de enfermos y  $\hat{r}_{Sp}$  el coeficiente entre las especificidades del grupo de sanos, desde muestras pareadas. En efecto el amplio uso que tiene el coeficiente kappa ( $\kappa$ ), Chen reemplazo el coeficiente de contingencia por los  $\hat{\kappa}_{Se}$  y  $\hat{\kappa}_{Sp}$  para facilitar el uso de software existente.

En relación a la estimación de  $c$  en la ecuacion 63, este puede ser obtenido por el método de multiplicadores de LaGrange, en el estudio de esta expresión, Chen demostró que  $c$  puede ser ignorado, debido a que en varios procedimientos de derivación matemática constato que los valores a los que se acerca como máximo es cero y como valor mínimo y medio son altamente cercanos a cero, del orden de  $-1.089 \times 10^{-9}$  y  $-2.675 \times 10^{-9}$  para el minimo y la media respectivamente. Estos valores fueron estimados para diferentes tamaños de muestra, mostrando que entre mayor sea el tamaño de muestra mas se acercara al valor cero, incluso en tamaños de muestra de 20 por grupo.

Teniendo en cuenta lo anterior la varianza del índice  $J$  de Youden pareado seria:

$$Var(\hat{J}_1 - \hat{J}_2)^* = \sum_{i=1}^2 [Var(\widehat{Se}_i) + Var(\widehat{Sp}_i)]$$

$$- 2 \left[ \hat{\kappa}_{Se} \sqrt{Var(\widehat{Se}_1)} \sqrt{Var(\widehat{Se}_2)} + \hat{\kappa}_{Sp} \sqrt{Var(\widehat{Sp}_1)} \sqrt{Var(\widehat{Sp}_2)} \right]$$

$$\begin{aligned}
Var(\hat{f}_1 - \hat{f}_2)^* &= \sum_{i=1}^2 \left[ \frac{a_i b_i}{(a_i + b_i)^3} + \frac{c_i d_i}{(c_i + d_i)^3} \right] \\
&\quad - 2 \left[ \hat{\kappa}_{se} \sqrt{\frac{a_1 b_1}{(a_1 + b_1)^3}} \sqrt{\frac{a_2 b_2}{(a_2 + b_2)^3}} + \hat{\kappa}_{sp} \sqrt{\frac{c_1 d_1}{(c_1 + d_1)^3}} \sqrt{\frac{c_2 d_2}{(c_2 + d_2)^3}} \right] \quad (i = 1, 2)
\end{aligned}$$

( 64 )

Con ecuación **64** de la varianza estimada de  $J$  se puede plantear el estadístico de prueba para el contraste de comparación de dos índices de Youden con datos pareados:

$$Z = \frac{\hat{f}_1 - \hat{f}_2}{\hat{\sigma}_{D_J}} = \frac{\hat{f}_1 - \hat{f}_2}{\sqrt{Var(\hat{f}_1 - \hat{f}_2)^*}}$$

Igualmente, un intervalo de confianza al  $(1 - \alpha)100\%$  para la diferencia de dos índices de Youden con datos pareados:

$$d \pm Z_{1-\alpha/2} \sqrt{Var(\hat{f}_1 - \hat{f}_2)^*}$$

## 3 La aplicación de índice de Youden a la determinación del punto de corte óptimo en dos ejemplos reales.

### 3.1 Software en R de determinación del punto de corte óptimo.

En actualidad existen diversos softwares para el análisis de datos de evaluación de pruebas diagnósticas. Stata, SPSS son algunos de los más comerciales e incluyen rutinas para obtener resultados sobre medidas de exactitud diagnóstica como sensibilidad, especificidad, valores predictivos y razones de verosimilitud positiva y negativa.

Todos incluyen análisis de curva ROC, pero no incluyen criterios de selección del punto de corte óptimo. El software estadístico R, es un software libre en el cual permitido la implementación de diversos paquetes desarrollados por sus usuarios y distribuidos a otros para su uso de forma libre. Ese es el caso del paquete publicado en el 2014 denominado `OptimalCutpoint` (Lopez-Raton, Monica; Rodriguez-Alvarez, Maria Xose; Cadarso-Suarez, Carmen; Gude-Sampedro, 2014) el cual comprende funciones básicas para el análisis de selección de puntos de corte óptimos en un test cuantitativo, incluye más de 34 métodos diferentes para selección de puntos de corte, incluye inferencia para las medidas de exactitud basado en intervalos de confianza con diferentes abordajes como lo son intervalos exactos, intervalos tipo wald, cuadrático de Fleiss, entre otros.

Las principales funciones de este paquete con su proposito se describen a continuación:

`optimal.cutpoints`: Calcula el punto de corte optimo con sus correspondientes medidas de exactitud, intervalos de confianza para cada medida.

`control.cutpoints`: Función para programar diferentes parámetros que controlan los procesos de cálculo de la función `optimal.cutpoints`

`print`: Imprime resultados de objetos creados mediante la función `optimal.cutpoints`

`summary`: produce un resumen del objeto tipo `optimal.cutpoints`

`plot`: grafica para objetos de tipo `optimal.cutpoints`, incluye curvas ROC y PROC indicando el punto de corte óptimo sobre los gráficos.

La función principal del paquete contiene los siguientes argumentos y son descritos a continuación:

```
optimal.cutpoints(X, status, tag.healthy, methods, data, direction=
c("<", ">"), categorical.cov = NULL, pop.prev = NULL, control=
control.cutpoints(), ci.fit = FALSE, conf.level = 0.95, trace = FALSE,
...)
```

`X`: es una cadena de caracteres con el nombre de la variable que contiene el resultado de la prueba diagnóstica en los individuos. Puede `X` también puede ser una fórmula. Cuando es así, debe ser un objeto de la clase fórmula. Al lado derecho del símbolo `~` debe colocarse el nombre de la variable que distingue los sanos de las personas enfermas, y al lado izquierdo del símbolo `~` debe contener el nombre de la variable que contiene la prueba diagnóstica.

`Status`: este argumento es una cadena de caracteres indicando el nombre de la variable que contiene la información para distinguir enfermos de sanos. Este argumento sólo aplica cuando `X` contiene el nombre de la variable de la prueba diagnóstica.

`tag.healthy`: Es un argumento que identifica el valor con el cual esta codificado los individuos sanos en la variable `status` del dataset.

`methods`: argumento vector carácter especificando el o los métodos que serán usados para seleccionar el punto de corte óptimo. Este paquete contiene 34 métodos diferentes para el cálculo de un punto de corte óptimo. Se pueden seleccionar varios métodos al mismo tiempo. Para examinar todos los métodos incluidos en el paquete se puede revisar la publicación de López et al (Lopez-Raton, Monica; Rodriguez-Alvarez, Maria Xose; Cadarso-Suarez, Carmen; Gude-Sampedro, 2014).

`Data`: con este argumento se declara el data frame el cual debe contener al menos las siguientes variables: el valor del test diagnóstico, la variable que indica el estado de sano y enfermo, y alguna covariable de ajuste para los resultados si se desea. Esta última debe ser categórica.

`Direction`: este argumento permite declarar mediante un carácter la dirección en la cual la curva ROC tiene que ser calculada. Por defecto los individuos con menor valor son clasificados como sanos. Mientras que pacientes con un valor en el test mayor indica los individuos enfermos. Si el caso es al contrario, el argumento `direction`, debe ser especificado con `>`

`categorical.cov`: es un argumento opcional y es un carácter cadena, con el nombre de la variable categórica, que indica los niveles por los cuales los puntos de corte deben ser calculados. Por defecto el valor de este argumento es `NULL`.

`pop.prev`: este argumento permite declarar un valor de prevalencia. Por defecto es `NULL`, y en tal caso la prevalencia estimada se realiza sobre la base de una prevalencia muestral, tomando en cuenta el número de pacientes en la muestra, esto aplica cuando el diseño de muestreo proviene de un estudio transversal (cross-sectional).

`Control`: indica la salida de la función `control.cutpoints`, la cual permite establecer el procedimiento de cálculo de aspectos importantes, como el cambio de método de construcción de intervalos de confianza, entre otros aspectos importantes.

`ci.fit`: este argumento es un valor lógico, el cual si es `TRUE` luego se calculan los intervalos de confianza para medidas de exactitud al punto de corte identificado por algunos de los métodos declarados en el argumento `methods` (por defecto es `FALSE`).

`conf.level`: bajo este argumento se declara un valor de nivel de confianza en las estimaciones, por defecto es 0.95.

### **3.2 Determinación del valor diagnóstico de la edad, el sexo y el índice de masa corporal para la hipertensión a través de los datos de la Encuesta Nacional de Salud: comparación de la calidad diagnóstica y de los puntos de cortes óptimos a través del índice de Youden.**

La idea que subyace debajo de este trabajo es la de determinar de manera rápida y sencilla cuando una persona elegida de la población Española tiene alguna de las tres enfermedades siguientes: hipertensión, diabetes o hipercolesterolemia o disfunción en el control lipídico. Como esa determinación se hace a partir del Encuesta Nacional de Salud Española en la que no se hace un diagnóstico riguroso de esas enfermedades hemos decidido que emplearemos el consumo de antihipertensivos, antidiabéticos y anticolésterolémicos, como indicador cierto de la enfermedad. En el caso del consumo de esos fármacos podemos asegurar que si una persona los consume lo hace porque presenta la enfermedad ya que son fármacos que no se consumen sin receta médica. Sin embargo, el objetivo del trabajo es determinar si una variable tan fácil de recoger como la edad de una persona nos podría permitir afirmar el consumo o no del medicamento y por tanto la presencia de la enfermedad. Para ello enfrentaremos la edad con el consumo de cada uno de los tres tipos de fármacos buscando el punto de corte idóneo para la edad (por el procedimiento que hemos estado viendo

con el índice de Youden) de forma que pudiéramos usar ese punto de corte (uno para cada enfermedad) ya que en él conoceremos la sensibilidad y la especificidad para el consume de cada uno de esos tres fármacos. Con ese punto de corte conocidos dichos parámetros podremos, en otros estudios, determinar el efecto de la hipertensión sobre otras variables si bien empleando modelos que manejen errores en las covariables. Todo ello nos ahorraría el trabajo del diagnóstico de la hipertensión en encuestas poblacionales en las que no intervenga personal sanitario, como es lo más común. Lo mismo se podría decir para la diabetes y la disfunción lipídica. La base de datos con la que hemos trabajado está conformada por 5 variables extraídas de la encuesta nacional de salud de España. Se tuvo en cuenta el sexo, la edad, y el consumo reportado por el encuestado en medicamentos para la hipertensión, colesterol y diabetes. Estas últimas son consideradas un proxy a la presencia de cada una de las patologías mencionadas en la muestra encuestada. Esta aplicación, como se ha dicho, consiste en encontrar los puntos de corte óptimos en la edad para el diagnóstico de las tres patologías: Hipertensión, Hipercolesterolemia y Diabetes. Caracterizados por el consumo de medicamentos para las patologías mencionadas. Las variables descritas son:

```
library(OptimalCutpoints)
load("~/Documents/MAESTRIA ESTADISTICA/TFM/datos_encuesta_nacional_salud/datos_salud.RData")
head(ensalud)
```

	sexo	EDADa	tension_c	colesterol_c	diabetes_c
1	1	25	0	0	0
2	0	75	0	0	0
3	1	46	0	0	0
4	0	19	0	0	0
5	1	22	0	0	0
6	1	22	0	0	0

Para la variable sexo representada por los codigos 1=Masculino 0=Femenino, EDADa= variable cuantitativa representada en años cumplidos al momento de la encuesta, tension\_c 1=consumo de medicamento para la tension arterial 0= no consumo, colesterol\_c 1=consumo de medicamento para el colesterol 0= no consumo, diabetes\_c 1= consumo de medicamento para la diabetes 0= no consumo. Estadísticos descriptivos de cada uno de ellos se muestran a continuación:

```
summary(ensalud)
```

sexo		EDADa		tension_c		colesterol_c	
Min.	:0.0000	Min.	: 15.0	Min.	:0.0000	Min.	:0.0000
1st Qu.:	:0.0000	1st Qu.:	: 37.0	1st Qu.:	:0.0000	1st Qu.:	:0.0000
Median	:0.0000	Median	: 50.0	Median	:0.0000	Median	:0.0000
Mean	:0.4593	Mean	: 51.6	Mean	:0.2096	Mean	:0.1336
3rd Qu.:	:1.0000	3rd Qu.:	: 67.0	3rd Qu.:	:0.0000	3rd Qu.:	:0.0000



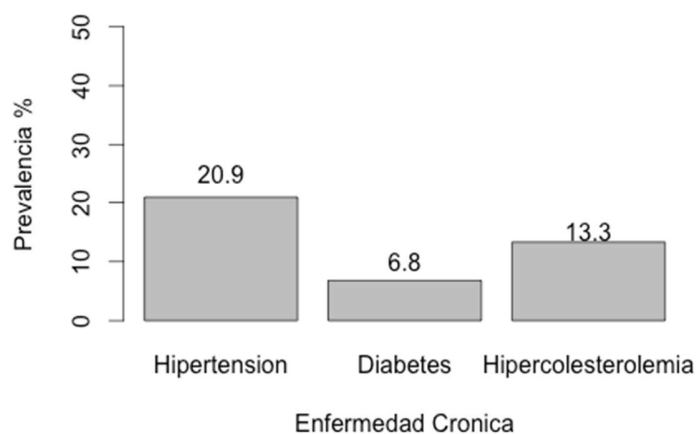
```

Max.    :1.0000   Max.    :103.0   Max.    :1.0000   Max.    :1.0000
diabetes_c
Min.    :0.00000
1st Qu.:0.00000
Median :0.00000
Mean    :0.06831
3rd Qu.:0.00000
Max.    :1.00000

```

Predomina el sexo femenino con el 54.07%, el promedio de edad del encuestado esta en 51.6 años (DE 19.09), las prevalencias son:

**Grafico 5. Prevalencias reportadas de hipertensión, Diabetes e Hipercolesterolemia, Encuesta Nacional de Salud**



### 3.2.1 Puntos de Corte para Hipertensión por Índice de Youden

Usando el paquete Optimalcutpoint de R, se genera el punto de corte de la edad para la identificación de la hipertensión sobre los datos de la encuesta nacional de salud. Se obtiene un punto de corte general bajo el siguiente comando:

```

cortehta<-optimal.cutpoints("EDADa", status = "tension_c", tag.healthy = 0, m
ethods = "Youden", data=ensalud, pop.prev = NULL, ci.fit = TRUE, conf.level =
0.95)

```

explorando los resultados obtenemos:

```
summary(cortehta)
```

Area under the ROC curve (AUC): 0.843 (0.837, 0.848)

CRITERION: Youden

Number of optimal cutoffs: 1

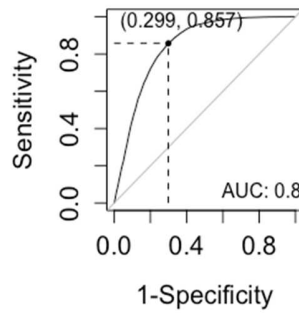
	Estimate	95% CI lower limit	95% CI upper limit
cutoff	56.0000000	-	-
Se	0.8573700	0.8466904	0.8675731
Sp	0.7009757	0.6939470	0.7079339
PPV	0.4319222	0.4237681	0.4531480
NPV	0.9488057	0.9445296	0.9504050
DLR.Positive	2.8672248	2.7930230	2.9433978
DLR.Negative	0.2034736	0.1891308	0.2189040
FP	4965.0000000	-	-
FN	628.0000000	-	-
Optimal criterion	0.5583456	-	-

La variable edad tiene una buena exactitud diagnostica presenta un área bajo la curva ROC (AUC) de 0.84 (IC95% 0.837 - 0.848) Se encuentra un punto de corte optimo en la edad de 56 años basado en el índice J-Youden (J=0.558), dicho punto de corte presenta una sensibilidad de 85.7% IC95% (84.6 - 86.75), especificidad de 70.0% IC95% (69.4 - 70.7).

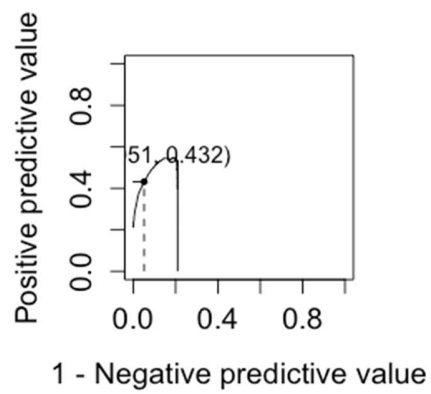
El valor tan alto para la sensibilidad encontrado con el punto de corte de 56 años nos permite pensar en el uso de la edad como test de screening para el diagnóstico rápido de la hipertensión en el sentido de que si no se tiene edad y por tanto se decide que no tiene hipertensión se tiene un valor predictivo negativo en una población como la que tenemos de 0.9488 que es ya una probabilidad muy alta. Ese punto de corte tiene una especificidad sensiblemente inferior, un 70%, lo que nos lleva a un valor predictivo positivo para una población como la anterior de un 43% que muestra claramente dónde radicarían los problemas de ese punto de corte. En cualquier caso, conociendo esas dos medidas en futuros estudios se puede emplear el punto de corte de la edad como indicador de hipertensión como covariables sujeta a error con la sensibilidad y especificidad que acabamos de estimar.

`plot(cortehta)`

**ROC Curve. Criterion: Youden**

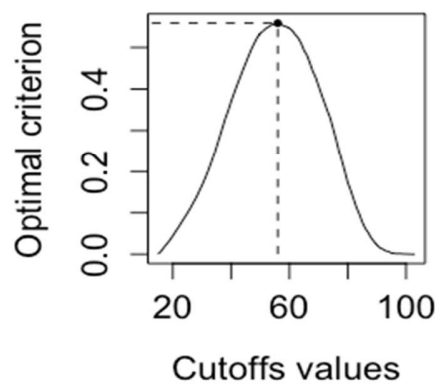


**PROC Curve. Criterion: Youden**



```
plot(cortehta, which=3)
```

**Criterion: Youden**



### 3.2.2 Puntos de Corte para Diabetes por Índice de Youden

Usando el paquete Optimalcutpoint de R, se genera el punto de corte de la edad para la identificación de la diabetes sobre los datos de la encuesta nacional de salud. Se obtiene un punto de corte general bajo el siguiente comando:

```
cortediabetes<-optimal.cutpoints("EDADa", status = "diabetes_c", tag.healthy
= 0, methods = "Youden", data=ensalud, pop.prev = NULL, ci.fit = TRUE, conf.l
evel = 0.95)
```

explorando los resultados obtenemos:

```
summary(cortediabetes)
```

Area under the ROC curve (AUC): 0.778 (0.768, 0.787)

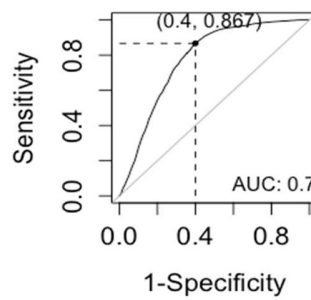
CRITERION: Youden

Number of optimal cutoffs: 1

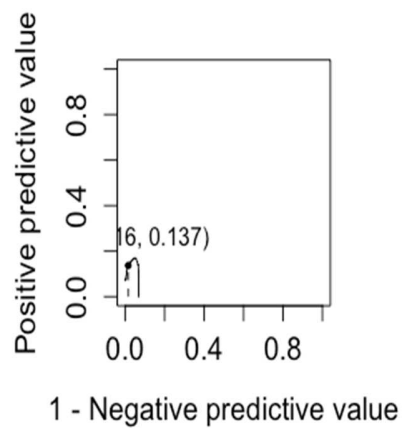
	Estimate	95% CI lower limit	95% CI upper limit
cutoff	55.0000000	-	-
Se	0.8668990	0.8482232	0.8840577
Sp	0.6001942	0.5932913	0.6070672
PPV	0.1371706	0.1338107	0.1569134
NPV	0.9840007	0.9814033	0.9844467
DLR.Positive	2.1682999	2.1114588	2.2266711
DLR.Negative	0.2217633	0.1942359	0.2531919
FP	7825.0000000	-	-
FN	191.0000000	-	-
Optimal criterion	0.4670931	-	-

La variable edad tiene una buena exactitud diagnostica, aunque no tan buena como la de la hipertensión, presenta una área bajo la curva ROC (AUC) de 0.778 (IC95% 0.768 - 0.787) Se encuentra un punto de corte optimo en la edad de 55 años basado en el índice J-Youden (J=0.0.48), dicho punto de corte presenta una sensibilidad de 86.6% IC95% (84.8 - 88.4), especificidad de 60.0% IC95% (59.3 - 60.7). El valor alto para la sensibilidad encontrado con el punto de corte de 55 años nos permite pensar en el uso de la edad como test de screening para el diagnóstico rápido de la diabetes en el sentido de que si no se tiene edad y por tanto se decide que no tiene hipertensión, se tiene un valor predictivo negativo en una población como la que tenemos de 98.4% que es ya una probabilidad muy alta. Ese punto de corte tiene una especificidad muy baja, 60% lo que nos lleva a un valor predictivo positivo para una población como la anterior de tan solo un 13.7% que muestra claramente dónde radicarían los problemas de ese punto de corte. En cualquier caso, conociendo esas dos medidas en futuros estudios se puede emplear el punto de corte de la edad como indicador de diabetes como covariables sujeta a error con la sensibilidad y especificidad que acabamos de estimar.

**ROC Curve. Criterion: Youden**

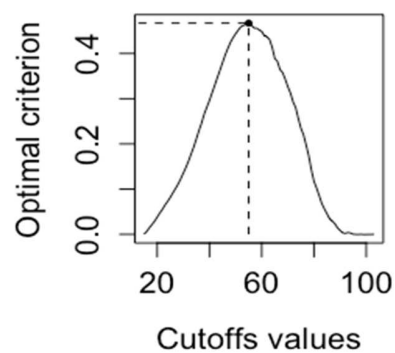


**PROC Curve. Criterion: Youden**



```
plot(cortediabetes, which=3)
```

**Criterion: Youden**



### 3.2.3 Puntos de Corte para Hipercolesterolemia por Índice de Youden

Usando el paquete Optimalcutpoint de R, se genera el punto de corte de la edad para la identificación de hipercolesterolemia sobre los datos de la encuesta nacional de salud. Se obtiene un punto de corte general bajo el siguiente comando:

```
cortecolest<-optimal.cutpoints("EDADa", status = "cholesterol_c", tag.healthy = 0, methods = "Youden", data=ensalud, pop.prev = NULL, ci.fit = TRUE, conf.level = 0.95)
```

explorando los resultados obtenemos:

```
summary(cortecolest)
```

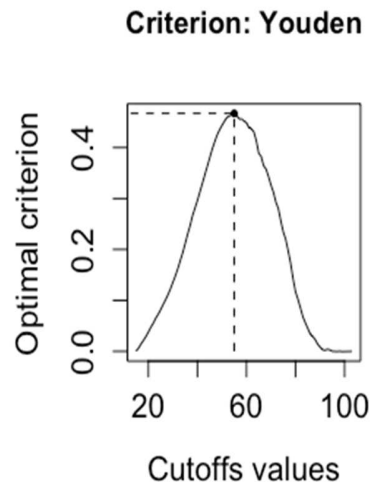
Area under the ROC curve (AUC): 0.782 (0.775, 0.789)

CRITERION: Youden

Number of optimal cutoffs: 1

	Estimate	95% CI lower limit	95% CI upper limit
cutoff	53.0000000	-	-
Se	0.8842180	0.8718008	0.8958276
Sp	0.6009341	0.5937764	0.6080595
PPV	0.2546947	0.2490869	0.2778759
NPV	0.9711419	0.9677065	0.9719655
DLR.Positive	2.2157191	2.1668677	2.2656719
DLR.Negative	0.1926700	0.1738276	0.2135549
FP	7263.0000000	-	-
FN	325.0000000	-	-
Optimal criterion	0.4851521	-	-

La variable edad tiene una buena exactitud diagnostica presenta una área bajo la curva ROC (AUC) de 0.782 (IC95% 0.775 - 0.789) Se encuentra un punto de corte optimo en la edad de 53 años basado en el índice J-Youden (J=0.48), dicho punto de corte presenta una sensibilidad de 88.4% IC95% (87.1 - 89.5), especificidad de 60.0% IC95% (59.3 - 60.7). El valor alto para la sensibilidad encontrado con el punto de corte de 53 años nos permite pensar en el uso de la edad como test de screening para el diagnóstico rápido de la disfunción lipídica, en el sentido de que si no se tiene edad y por tanto se decide que no tiene hipertensión, se tiene un valor predictivo negativo en una población como la que tenemos de 97.1.4% que es ya una probabilidad muy alta. Ese punto de corte tiene una especificidad muy baja, 60% lo que nos lleva a un valor predictivo positivo para una población como la anterior de 25.4%. Al igual que en hipertensión y diabetes en el colesterol conociendo esas dos medidas en futuros estudios se puede emplear el punto de corte de la edad como indicador de estas enfermedades como covariables sujeta a error con la sensibilidad y especificidad.



### 3.2.4 Puntos de corte para la edad con regresión logística en patología crónicas

#### 3.2.4.1 Punto de corte para hipertensión por regresión logística

Se asumen un modelo de regresión logística incluyendo las variable edad y sexo como covariable.

```
modelo.rlm<-glm(tension_c~EDADa+sexo, family=binomial, data= ensalud)
summary(modelo.rlm)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.1447	-0.5806	-0.3260	-0.1567	3.0022

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-5.981616	0.090671	-65.971	<2e-16 ***
EDADa	0.079377	0.001326	59.884	<2e-16 ***
sexo	-0.022149	0.040220	-0.551	0.582

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance:	21571	on 21006	degrees of freedom
Residual deviance:	16130	on 21004	degrees of freedom
AIC:	16136		

El coeficiente de la variable edad, es estadísticamente significativo, mostrando un cambio en el logaritmo del odds de sufrir de hipertensión. para facilitar su interpretación clínica se examina el coeficiente

exponenciado para obtener en forma de Odds Ratio (OR). Sin embargo, el coeficiente del sexo no es significativamente diferente de 1 lo que indica un efecto pequeño de esta variable sobre el consume de fármacos antihipertensivos.

```
exp(modelo.rlm$coefficients)

(Intercept)      EDADa      sexo
0.002524742 1.082612429 0.978094150

          2.5 %      97.5 %
(Intercept) 0.002113672 0.003015757
EDADa      1.079803490 1.085428675
sexo       0.903952079 1.058317348
```

En términos de OR, por cada incremento en un año de edad el odds de padecer hipertensión incrementa aproximadamente 1.08 veces IC95%(1.079 - 1.058). Las mujeres tienen 1.02 veces mayor frecuencia de consume de antihipertensivos comparado con los hombres, aunque como se mencionó no es estadísticamente significativo.

Si realizamos un cálculo de la probabilidad predicha de consumo de antihipertensivos (Hipertensos), bajo el modelo de regresión logística que incluye edad y sexo del paciente, se construye una variable llamada **tension.modelo**

```
tension.modelo<-modelo.rlm$fitted.values
ensalud2<-data.frame(ensalud,tension.modelo)
```

Usando de nuevo el paquete OptimalCutpoints se puede obtener el punto de corte para definir la probabilidad ajustada de consumo de hipertensión como variable proxy al diagnóstico de hipertensión, bajo el modelo de regresión logística que se ajustó y que incluía información adicional sobre el sexo del paciente tenemos:

```
prueba<-optimal.cutpoints("tension.modelo", status = "tension_c", tag.healthy
= 0, methods = "Youden", data=ensalud2, pop.prev = NULL, ci.fit = TRUE, conf
.level = 0.95)

summary(prueba)

Area under the ROC curve (AUC): 0.843 (0.837, 0.848)
CRITERION: Youden
Number of optimal cutoffs: 1

          Estimate 95% CI lower limit 95% CI upper limit
## cutoff      0.1738402          -          -
```

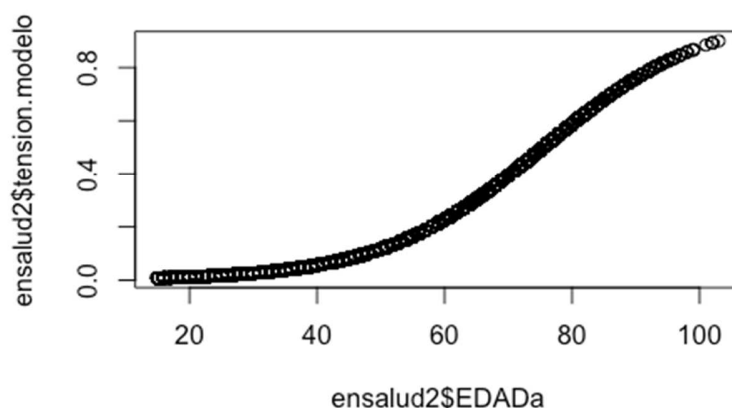


## Se	0.8573700	0.8466904	0.8675731
## Sp	0.7009757	0.6939470	0.7079339
## PPV	0.4319222	0.4237681	0.4531480
## NPV	0.9488057	0.9445296	0.9504050
## DLR.Positive	2.8672248	2.7930230	2.9433978
## DLR.Negative	0.2034736	0.1891308	0.2189040
## FP	4965.0000000	-	-
## FN	628.0000000	-	-
## Optimal criterion	0.5583456	-	-

De acuerdo al resultado anterior, se identifica como punto de corte para probabilidad predicha de consumo de antihipertensivos de 0.173 ( $Se = 85.7\%$ ,  $Sp = 70\%$ ) y se identifica como punto de corte para la edad aproximadamente de 56 años como punto de screening de probabilidad de consumo de antihipertensivos bajo un modelo de regresión logística ajustado por sexo.

Como se puede evidenciar el punto de corte determinado por regresión logística es igual al obtenido por el criterio del índice de Youden, esto debido probablemente a que el sexo no tiene un efecto significativo sobre la probabilidad de consumo de antihipertensivos

```
plot(ensalud2$EDADa, ensalud2$tension.modelo)
```



#### 3.2.4.2 Punto de corte para Diabetes por Regresión Logística

Se asumen un modelo de regresión logística incluyendo las variable edad y sexo como variable de control.

```
modelo.rlmdiab<-glm(diabetes_c~EDADa+sexo, family=binomial, data= ensalud)
```

```
summary(modelo.rlmdiab)
```

```
Deviance Residuals:
```

Min	1Q	Median	3Q	Max
-1.0714	-0.4153	-0.2600	-0.1772	3.1452

```
Coefficients:
```

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-6.128832	0.127850	-47.938	< 2e-16 ***
EDADa	0.056653	0.001755	32.284	< 2e-16 ***
sexo	0.265641	0.057652	4.608	4.07e-06 ***

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Dispersion parameter for binomial family taken to be 1)
```

```
Null deviance: 10471.9 on 21006 degrees of freedom  
Residual deviance: 9183.7 on 21004 degrees of freedom  
AIC: 9189.7
```

```
Number of Fisher Scoring iterations: 6
```

El coeficiente de la variable edad y sexo, son estadísticamente significativos y están relacionados con el cambio en el logaritmo del odds de sufrir de diabetes. para facilitar su interpretación clínica se examina el coeficiente exponenciado para obtener en forma de Odds Ratio (OR).

```
exp(modelo.rlmdiab$coefficients)
```

(Intercept)	EDADa	sexo
0.002179124	1.058288012	1.304266584

	2.5 %	97.5 %
(Intercept)	0.001696118	0.002799677
EDADa	1.054654399	1.061934144
sexo	1.164912519	1.460291048

En término de OR, por cada incremento en un año de edad el odds de padecer diabetes incrementa aproximadamente 1.06 veces IC95%(1.05 - 1.061) independiente del sexo. Por el contrario, la variable sexo muestra un OR 1.3 (IC95% 1.16 - 1.46), siendo 1.3 veces más frecuente la diabetes entre hombres comparado con las mujeres, independiente de la edad.

Si realizamos un cálculo de la probabilidad predicha de consumo de antidiabéticos (Diabéticos), bajo el modelo de regresión logística que incluye edad y sexo del paciente, se construye una variable llamada `diabetes.modelo`

```
diabetes.modelo<-modelo.rlmdiab$fitted.values
ensalud2<-data.frame(ensalud,tension.modelo, diabetes.modelo)
```

Usando de nuevo el paquete `OptimalCutpoints` se puede obtener el punto de corte para definir la probabilidad ajustada de consumo de antidiabéticos como variable proxy al diagnóstico de diabetes, bajo el modelo de regresión logística que se ajustó y que incluía información adicional sobre el sexo del paciente, tenemos:

```
prueba1<-optimal.cutpoints("diabetes.modelo", status = "diabetes_c", tag.heal
thy = 0, methods = "Youden", data=ensalud2, pop.prev = NULL, ci.fit = TRUE, c
onf.level = 0.95)
summary(prueba1)
Area under the ROC curve (AUC): 0.78 (0.77, 0.79)
```

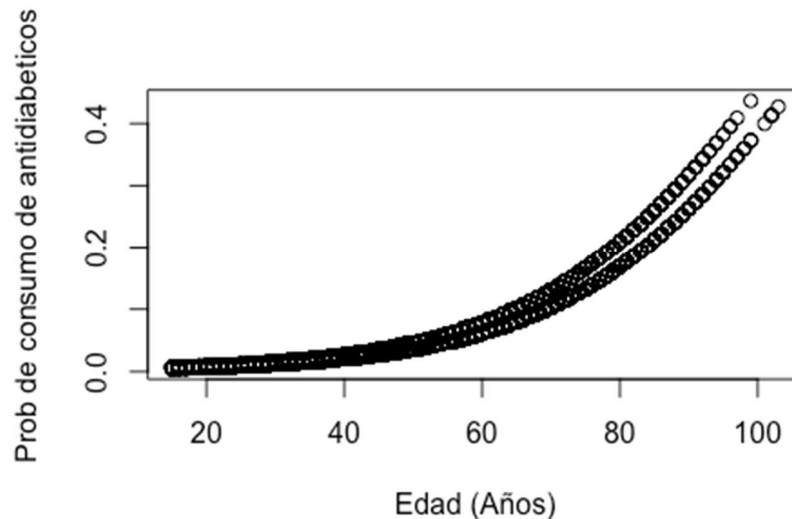
CRITERION: Youden

Number of optimal cutoffs: 1

	Estimate	95% CI lower limit	95% CI upper limit
cutoff	5.130689e-02	-	-
Se	8.801394e-01	0.8622037	0.8965015
Sp	5.927856e-01	0.5858635	0.5996802
PPV	1.367919e-01	0.1334496	0.1574928
NPV	9.853915e-01	0.9828996	0.9857981
DLR.Positive	2.161366e+00	2.1069469	2.2171908
DLR.Negative	2.021989e-01	0.1756632	0.2327432
FP	7.970000e+03	-	-
FN	1.720000e+02	-	-
Optimal criterion	4.729250e-01	-	-

De acuerdo al resultado anterior, se identifica como punto de corte para probabilidad predicha de consumo de antidiabéticos de 0.05 (Se= 88.01%, Sp= 59.2%). y se identifica como punto de corte para la edad aproximadamente de 52 años como punto de screening de probabilidad de consumo de antidiabéticos bajo un modelo de regresión logística ajustado por sexo. Este punto de corte es un poco menor al encontrado con el criterio de índice de Youden (55 años), esto probablemente debido al efecto significativo que tiene el sexo sobre la probabilidad de consumo, haciendo que al tener a la variable sexo como variable de ajuste el punto de corte disminuya.

```
plot(ensalud2$EDADa, ensalud2$diabetes.modelo, xlab = "Edad (Años)", ylab = "
Prob de consumo de antidiabeticos")
```



### 3.2.4.3 Puntos de Corte para Hipercolesterolemia por regresión logística

se asumen un modelo de regresión logística incluyendo las variable edad y sexo como variable de control.

```
modelo.rlmcolest<-glm(colesterol_c~EDADa+sexo, family=binomial, data= ensalud
)
```

```
summary(modelo.rlmcolest)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.4444	-0.5496	-0.3576	-0.2365	2.8774

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-5.193992	0.091900	-56.518	<2e-16 ***
EDADa	0.056338	0.001313	42.902	<2e-16 ***
sexo	0.053709	0.044053	1.219	0.223

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 16521 on 21006 degrees of freedom  
Residual deviance: 14206 on 21004 degrees of freedom  
AIC: 14212

Number of Fisher Scoring iterations: 5

El coeficiente de la variable edad al igual que los modelos anteriores es estadísticamente significativa y esta relacionada con el cambio en el logaritmo del odds de sufrir de disfunción lipídica. Para facilitar su interpretación clínica se examina el coeficiente exponenciado para obtener en forma de Odds Ratio (OR). Sin embargo el coeficiente del sexo no es significativamente diferente de 1 (valor  $p = 0.223$ ) lo que indica el poco efecto de esta variable sobre el consume de fármacos para el manejo de colesterol.

```
exp(modelo.rlmcolest$coefficients)

(Intercept)      EDADa      sexo
0.005549809 1.057955041 1.055177734

      2.5 %      97.5 %
(Intercept) 0.004635031 0.006645129
EDADa      1.055235576 1.060681515
sexo       0.967893958 1.150332679
```

En término de OR, por cada incremento en un año de edad el odds de padecer una disfunción lipídica incrementa aproximadamente 1.06 veces IC95%(1.05 - 1.061) independiente del sexo. Por el contrario, la variables sexo muestra un OR 1.05 (IC95% 0.96 - 1.15), este intervalo de confianza incluye el valor 1, mostrando igualmente que el efecto del sexo no es significativo.

Si realizamos un cálculo de la probabilidad predicha de consumo de medicamento para el colesterol (Hipercolesterolemico), bajo el modelo de regresión logística que incluye edad y sexo del paciente, se construye una variable llamada `colesterol.modelo`

```
colesterol.modelo<-modelo.rlmcolest$fitted.values
ensalud2<-data.frame(ensalud,tension.modelo, diabetes.modelo, colesterol.modelo)
```

Usando de nuevo el paquete `OptimalCutpoints` se puede obtener el punto de corte para definir la probabilidad ajustada de consumo de medicamentos para el colesterol como variable proxy al diagnóstico de disfunción lipídica, bajo el modelo de regresión logística que se ajustó y que incluía información adicional sobre el sexo del paciente, tenemos:

```
prueba2<-optimal.cutpoints("colesterol.modelo", status = "colesterol_c", tag.
healthy = 0, methods = "Youden", data=ensalud2, pop.prev = NULL, ci.fit = TRUE,
conf.level = 0.95)
summary(prueba2)
```

```
Area under the ROC curve (AUC): 0.782 (0.775, 0.789)
```

CRITERION: Youden

Number of optimal cutoffs: 1

	Estimate	95% CI lower limit	95% CI upper limit
cutoff	0.1039229	-	-
Se	0.8763805	0.8636324	0.8883378
Sp	0.6096154	0.6024835	0.6167123
PPV	0.2571877	0.2515223	0.2798199
NPV	0.9696731	0.9661744	0.9705408
DLR.Positive	2.2449155	2.1941759	2.2968284
DLR.Negative	0.2027828	0.1836356	0.2239265
FP	7105.0000000	-	-
FN	347.0000000	-	-
Optimal criterion	0.4859959	-	-

De acuerdo al resultado anterior, se identifica como punto de corte para probabilidad predicha de consumo de medicamentos para el colesterol de 0.103 (Se= 87.63%, Sp= 60.96%). y se identifica como punto de corte para la edad aproximadamente de 53 años como punto de screening de probabilidad de consumo de antihipertensivos bajo un modelo de regresión logística ajustado por sexo. Punto de corte exactamente igual al obtenido bajo el método del índice de Youden.

```
plot(ensalud2$EDADa, ensalud2$colesterol.modelo, xlab = "Edad (Años)", ylab =  
"Prob de consumo de med Colesterol")
```

### **3.3 Valor pronóstico de una regla predictiva de la depresión en pacientes de Atención Primaria: determinación del punto de corte óptimo a partir del índice de Youden para la decisión de instaurar o no un tratamiento preventivo.**

El estudio PREDICT-Spain, es un estudio de cohorte prospectiva que fue desarrollado para crear un algoritmo de predicción de depresión mayor a 12 meses. Conto con la participación de hombres y mujeres de centros de salud de 7 provincias de España. Se aplicaron regresiones logísticas multinivel y diferentes estadísticos para evaluar su bondad de ajuste.

El archivo con el que se trabaja dispone de una variable que fue la presencia o no de depresión al cabo de 12 meses de seguimiento del paciente y la otra variable que fue la probabilidad de presentar esa depresión predicha para cada paciente según el modelo ajustado.

El objetivo de este estudio era determinar, en un futuro ensayo clínico los pacientes con una alta probabilidad de depresión para establecer la efectividad de un tratamiento que previniera dicha depresión. Para identificar un punto de corte en el score del modelo predictivo, es decir la probabilidad de deprimirse al año, obtenido al aplicar el cuestionario se utilizará el criterio del índice de Youden, generando un punto de corte optimo y sus respectivas medidas de exactitud diagnostica.

```
library(OptimalCutpoints)
load("~/Documents/MAESTRIA ESTADISTICA/TFM/archivos_predict_tfm/predict.RData")

corte.predict<-optimal.cutpoints("prob", status = "dep", tag.healthy = 0, methods = "Youden", data=predict, pop.prev = NULL, ci.fit = TRUE, conf.level = 0.95)
summary(corte.predict)
```

Area under the ROC curve (AUC): 0.823 (0.798, 0.847)

CRITERION: Youden

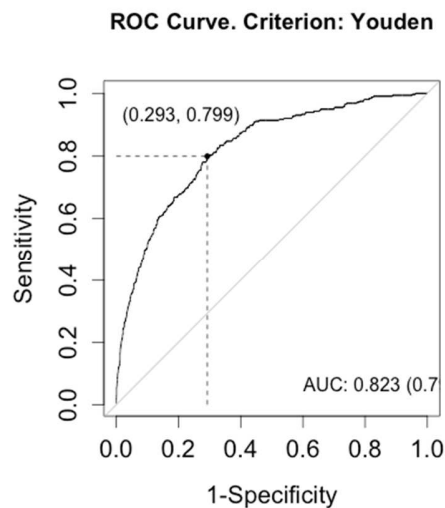
Number of optimal cutoffs: 1

	Estimate	95% CI lower limit	95% CI upper limit
cutoff	0.09197832	-	-
Se	0.79872204	0.7499628	0.8417257
Sp	0.70741646	0.6889727	0.7253640
PPV	0.25826446	0.2418512	0.3181665
NPV	0.96498054	0.9541870	0.9678468
DLR.Positive	2.72989401	2.5126331	2.9659409
DLR.Negative	0.28452540	0.2278469	0.3553030
FP	718.00000000	-	-
FN	63.00000000	-	-
Optimal criterion	0.50613851	-	-

El área bajo la curva ROC para el score obtenido en el cuestionario de Riesgo de depresión mayor a 12 meses muestra un valor de 0.823 (IC95% 0.798 – 0.847), con una buena capacidad discriminante, de hecho fue la más alta obtenida en problemas similares. Se identifica como punto de corte optimo, basado en el criterio del índice de Youden, un score (probabilidad de desarrollar depresión a 12 meses) de 0.09. La sensibilidad asociada a este punto de corte es de 79.8% (IC95% 74.9 - 84.1), especificidad 70.7% (IC95% 68.8 - 72.5) y el índice de J-Youden para este punto de corte es 0.506. Como se puede ver en la tabla el valor predictivo negativo en una población como ésta fue de un 96,5% valor bastante elevado. Con ese valor a la hora del ensayo clínico estamos bastante seguros de que si un paciente candidato a entrar en el ensayo clínico tiene una probabilidad de deprimirse inferior al 9% efectivamente no se va a deprimir y por tanto sería descartable para solicitar su participación en el ensayo. Con ese punto de corte se tiene un valor

predictivo positivo mucho menor, 25,8%, debido a que la prevalencia del problema es muy pequeña; este dato lo que nos dice es que, si nosotros consideramos como candidato a participar en el ensayo clínico a aquellas personas con probabilidad de deprimirse al año superior al 9%, solo el 25% de ellas realmente se deprimirían de forma que el tratamiento tendría que ser muy efectivo para que realmente pudiéramos poner de manifiesto su bondad. Esta forma tan conservadora de comportarnos al tomar el 9% obtenido del índice de Youden, es la idónea en este caso porque por un lado cumple claramente las normas éticas, se dejará sin tratar a aquellos es muy seguro que no desarrollarán el problema, y solo se tratarán aquellos que tienen un riesgo alto, aunque sepamos que tres cuartos de ellos no desarrollarían la enfermedad lo que obliga a que el tratamiento probado deba ser muy potente. El ensayo clínico se llevó a cabo en estas condiciones y demostró la efectividad del tratamiento.

```
plot(corte.predict)
```





## 4 Conclusiones

Con este trabajo de fin de master se ha presentado el índice de Youden como un estadístico resumen sobre la exactitud de un test diagnóstico que no depende de la prevalencia, es versátil en la medida que el investigador puede usarla asumiendo igual costo en la clasificación errónea de sanos y enfermos, sin embargo también versión ponderada del índice existe y permite que dicha ponderación sean por el costo de una calcificación errónea.

Con este trabajo hemos logrado evidenciar diferentes enfoques (paramétricos, no-paramétricos, semi-paramétricos y bootstrap), reportados en la literatura científica, para la construcción de intervalos de confianza para el índice de Youden. Aunque entre ellos se presentan ventajas y desventajas, la mayoría de ellos tienen la posibilidad de una implementación en el análisis de datos en investigación biomédica. En referencia a contrastes de hipótesis, se identificó una publicación reportando dicho método para el caso del contraste de prueba sobre un índice de Youden, comparación de dos índices en muestras independiente y muestras apareadas.

En referencia a la obtención del punto de corte óptimo por el índice de Youden en una curva ROC, con este trabajo se encontraron diferentes enfoques para encontrar el valor  $c^*$  del test donde se hace máxima diferencia entre la distribución de los valores del test para población sana y enferma. Los enfoques de estimación de estas distribuciones pueden ser paramétricas y no paramétricas y se han reportado trabajos sobre las distribuciones paramétricas como la Normal y Gamma, sin embargo, la construcción de distribuciones empíricas también han sido consideradas.

En cuanto a la implementación del índice de Youden en los problemas prácticos propuestos, en el primer ejemplo mediante el paquete de R denominado `OptimalcutPoints` se pretendió determinar la presencia de hipertensión usando como indicador la edad. Tal punto de corte resultó muy sensible y poco específico en las tres patologías estudiadas, dejando como resultado que la edad puede ser usado como test de screening para el diagnóstico rápido de estas patologías y dejando la puerta abierta a ser usado en estudios posteriores, sujeto a los errores de clasificación reportados en este análisis. En el segundo ejemplo, con los datos del estudio PREDCIT-Spain que tuvo como objetivo principal validar un modelo predictivo de depresión a 12 meses; bajo la filosofía de Youden se encontró un punto de corte en la probabilidad de desarrollar depresión al año con una buena sensibilidad y especificidad, permitiendo en futuros ensayos clínicos tamizar participantes con alta probabilidad de desarrollar depresión y ser susceptibles de beneficiarse de la

intervención, y por el contrario quienes no tienen esta alta probabilidad no ser incluidos y no someterlos a tratamientos innecesarios, lo anterior en consonancia con el desarrollo ético de investigación en la que participan seres humanos.

## 5 Índice de Gráficos

Grafico 1 Grafico de Shannon para el contenido de información de un mensaje.....	14
Grafico 2 Relación entre el índice J-Youden y el logaritmo de la razón de verosimilitud positiva $\log[LR+]$ .....	14
Grafico 3 Distribución de los valores de un test cuantitativo en población sana y enferma .....	15
Grafico 4. Representación en un plano del espacio ROC .....	17
Grafico 5. Prevalencias reportadas de hipertensión, Diabetes e Hipercolesterolemia, Encuesta Nacional de Salud.....	49

# 6 Índice de Tablas

Tabla 1 Tabla de contingencia  $2 \times 2$  para los resultados de evaluación de una prueba diagnostica con  
resultado binario..... 7

Tabla 2. Tabla de contingencia de  $2 \times 2$  para un test cuantitativo para un punto de corte c ..... 10

## 7 Bibliografía

- Böhning, D. (2015). Youden ' s Index and the Likelihood Ratio Positive in Diagnostic Testing, 1–2.
- Böhning, D., Böhning, W., & Holling, H. (2008). Revisiting Youden's index as a useful measure of the misclassification error in meta-analysis of diagnostic studies. *Statistical Methods in Medical Research*, 17(6), 543–54. <http://doi.org/10.1177/0962280207081867>
- Chen, F., Xue, Y., Tan, M. T., & Chen, P. (2015). Efficient statistical tests to compare Youden index: Accounting for contingency correlation. *Statistics in Medicine*, 34(9), 1560–1576. <http://doi.org/10.1002/sim.6432>
- Fluss, R., Faraggi, D., & Reiser, B. (2005). Estimation of the Youden Index and its associated cutoff point. *Biometrical Journal*, 47(4), 458–472. <http://doi.org/10.1002/bimj.200410135>
- Hughes, G. (2015). Youden ' s Index and the Weight of Evidence. *Methods of Information in Medicine*, 198–199. <http://doi.org/10.1186/s12911>
- Li, D. L., Shen, F., Yin, Y., Peng, J. X., & Chen, P. Y. (2013). Weighted youden index and its two-independent-sample comparison based on weighted sensitivity and specificity. *Chinese Medical Journal*, 126(6), 1150–1154. <http://doi.org/10.3760/cma.j.issn.0366-6999.20123102>
- Lopez-Raton, Monica; Rodriguez-Alvarez, Maria Xose; Cadarso-Suarez, Carmen; Gude-Sampedro, F. (2014). OptimalCutpoints : An R Package for Selecting Optimal Cutpoints in Diagnostic Tests. *Journal of Statistical Software*, 61(8). <http://doi.org/10.18637/jss.v061.i08>
- Noguera Moreno, T. L. (2010). Metodología ROC en la Evaluación de Medidas Antropométricas como Marcadores de la Hipertensión Arterial. Aplicación a Población Gallega Adulta.
- Perkins, N. J., & Schisterman, E. F. (2006). The Inconsistency of “ Optimal ” Cut-points Using Two ROC Based Criteria. *Am J Epidemiol*, 163(7), 670–675. <http://doi.org/10.1093/aje/kwj063>
- Perkins, N. J., Schisterman, E. F., & Vexler, A. (2007). Receiver operating characteristic curve inference

from a sample with a limit of detection. *American Journal of Epidemiology*, 165(3), 325–333.  
<http://doi.org/10.1093/aje/kwk011>

Ruopp, M. D., Perkins, N. J., Whitcomb, B. W., & Schisterman, E. F. (2008). Youden Index and optimal cut-point estimated from observations affected by a lower limit of detection. *Biometrical Journal*, 50(3), 419–430. <http://doi.org/10.1002/bimj.200710415>

Schisterman, E. F., & Perkins, N. (2007). Confidence intervals for the Youden index and corresponding optimal cut-point. *Communications in Statistics - Simulation and Computation*, 36(3), 549–563.  
<http://doi.org/10.1080/03610910701212181>

Shan, G. (2015). Improved confidence intervals for the Youden Index. *PLoS ONE*, 10(7), 1–19.  
<http://doi.org/10.1371/journal.pone.0127272>

Smits, N. (2010). A note on Youden's J and its cost ratio. *BMC Medical Research Methodology*, 10(1), 89.  
<http://doi.org/10.1186/1471-2288-10-89>

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, 3(1), 32–35. [http://doi.org/10.1002/1097-0142\(1950\)3:1<32::AID-CNCR2820030106>3.0.CO;2-3](http://doi.org/10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3)

Zhou, H., & Qin, G. (2012). New Nonparametric Confidence Intervals for the Youden Index. *Journal of Biopharmaceutical Statistics*, 22(February 2015), 1244–1257.  
<http://doi.org/10.1080/10543406.2011.592234>