

## PCorteSoft - Sistema para la determinación del punto de corte óptimo en escalas predictivas y la categorización de variables continuas

PCorteSoft – System for determination of optimal cut-off point in predictive scales and categorization of continuous variables

Larisa Zamora Matamoros<sup>1\*</sup>

[0000-0003-2210-0806](tel:0000-0003-2210-0806)

Jorge Rey Díaz Silvera<sup>1</sup>

[0000-0001-7243-0809](tel:0000-0001-7243-0809)

<sup>1</sup> Universidad de Oriente, Santiago de Cuba, Cuba.

\* Autor para la correspondencia: [larisa@uo.edu.cu](mailto:larisa@uo.edu.cu)

### RESUMEN

Al desarrollar modelos de predicción para su aplicación en la práctica clínica, los profesionales de la salud suelen categorizar las variables clínicas que son de naturaleza continua. En muchas ocasiones estos modelos constituyen la base para la confección de escalas predictivas, a partir de las cuales se estratifica a los pacientes en varias categorías atendiendo al fenómeno estudiado. En estos casos se requiere la determinación de uno o varios puntos de cortes que permitan dividir el recorrido de la variable, variables continuas o puntuaciones de una escala, en dos o más categorías.

El presente trabajo tiene como objetivo la automatización de diferentes métodos para dicotomizar variables continuas en modelos de predicción clínica, donde la variable respuesta es dicotómica, y determinar el punto de corte óptimo en la estratificación de pacientes en dos categorías, a partir de escalas de predicción. Para ello se elaboró un software en el lenguaje de programación R, que implementa diferentes métodos para la determinación del punto de corte óptimo, lo cual agiliza el trabajo investigativo de los especialistas de salud en el proceso de elaboración de modelos predictivos y/o escalas de predicción.

**Palabras clave:** categorización de variables continuas; punto de corte; escalas de predicción; software en R.



## ABSTRACT

When developing predictive models for application in clinical practice, health professionals often categorize clinical variables that are continuous in nature. In many cases, these models are the basis for the development of predictive scales from which patients are stratified into various categories according to the phenomenon under study. In both cases, it is necessary to determine one or more cut-off points that allow dividing the path of the variable, continuous variables, or scores of a scale into two or more categories.

The aim of the present work is to automate different existing methods for dichotomizing continuous variables in clinical prediction models where the response variable is dichotomous, as well as to determine the optimal cut-off point for stratifying patients into two categories, based on prediction scales. For this purpose, a software was developed in the R programming language, which implements different existing methods for the determination of the optimal cut-off point, speeding up the research work of health specialists in the process of developing predictive models and/or prediction scales.

**Keywords:** categorization of continuous variables; cut-off point; prediction scales; R software.

**Recibido:** 20/03/2023

**Aprobado:** 01/09/2023

## Introducción

El pronóstico y los modelos de predicción son esenciales en la atención médica integral al paciente. La función de pronosticar es inherente a la ciencia contemporánea, por lo que un elemento básico que debe demostrar la madurez de la clínica como ciencia es precisamente la certeza de su pronóstico. El desarrollo tecnológico bien empleado permite emitir predicciones más precisas; sin embargo, tienen la influencia de incertidumbre inherente a la clínica. El método clínico y la revolución tecnológica van en un proceso de desarrollo continuo donde ninguno sustituye al otro, sino que ambos se complementan. <sup>(1)</sup>

El pronóstico es un proceso probabilístico en el cual se estima alguna situación que genera incertidumbre. En la práctica clínica, entre los propósitos principales de los estudios de pronóstico se incluyen describir la historia natural o el curso clínico de las enfermedades, investigar las variables o factores asociados a los distintos desenlaces de interés, construir escalas para estimar las probabilidades individuales para desarrollar dichos desenlaces y establecer las consecuencias de intervenciones posteriores a su implementación. <sup>(2)</sup> En estos estudios se deben construir modelos de



predicción, los cuales se definen como un grupo de técnicas que, mediante el aprendizaje automático, la recolección de datos históricos, el Big Data y el reconocimiento de patrones, pretenden predecir eventos o resultados futuros mediante el análisis de patrones y así ayudar en la toma de decisiones en cualquier dominio de aplicación. <sup>(3)</sup>

Los modelos de predicción son actualmente relevantes en varios campos, incluida la medicina. Decisiones como el tratamiento más adecuado para una enfermedad, si se debe o no dar de alta a un determinado paciente, etc., se basan en el riesgo individual del paciente de sufrir algún evento desfavorable, y dicho riesgo se suele medir en función de variables clínicas que son de naturaleza continua.<sup>(3)</sup> Estos modelos son construidos a partir de variables medidas en sujetos sanos y enfermos, y se busca encontrar el modelo con el menor número de variables (modelo más parsimonioso) que mejor pronostique la probabilidad del desenlace. Para lograr esto, las variables pueden ser modificables en su escala de medición. <sup>(2)</sup>

Al desarrollar modelos de predicción para su aplicación en la práctica clínica, los profesionales de la salud suelen categorizar las variables clínicas que son de naturaleza continua, aunque esto no resulte aconsejable desde el punto de vista clínico y estadístico, ya que implica una pérdida importante de información. <sup>(3-5)</sup> Además, si los puntos de corte se eligen en base a la información proporcionada por los propios datos del estudio, las conclusiones que se obtienen pueden ser menos extrapolables a otras situaciones. <sup>(6)</sup>

Para determinar los puntos de corte en el proceso de categorización, existen al menos dos aproximaciones: una interna, orientada a los datos, y otra externa orientada al instrumento (orientada a los resultados). <sup>(6, 7)</sup> La interna se basa a menudo en el cálculo de percentiles, aunque tiene sus inconvenientes, incluso cuando la categorización se base en criterios clínicos se ha demostrado que puede variar enormemente de un médico, hospital, o país a otro.<sup>(3, 4)</sup> La externa exige contar con un criterio dicotómico, a manera de estándar de oro, con el cual contrastar los resultados obtenidos en el instrumento. <sup>(7)</sup>

Para la obtención de puntos de corte candidatos se proponen varias metodologías: la determinación biológica, la representación gráfica de los datos para un efecto de umbral, el cálculo de percentiles, la exploración de todos los valores observados de la variable continua para determinar el que mejor separe los grupos de riesgo según una prueba Chi-cuadrado y a partir de los coeficientes del modelo de la regresión logística que se obtiene al considerar la variable continua como única variable independiente. La primera mencionada es la más deseable, pero a menudo no disponible. <sup>(4, 6, 8)</sup>

Una vez obtenido el modelo predictivo, en dependencia del objetivo de la investigación, se elaboran escalas predictivas. <sup>(9, 10)</sup> Una escala debe contar con criterios para evaluar la gravedad del fenómeno estudiado y realizar una clasificación diagnóstica de las personas investigadas, lo cual pertenece al dominio de la clasificación diagnóstica. Esta clasificación requiere también determinar un punto de



corte en las puntuaciones brutas de la escala, para diferenciar entre quienes sufren el fenómeno investigado y quienes no lo hacen. <sup>(7)</sup>

René GF y colaboradores plantean que un método estadístico apropiado para evaluar la eficacia de una prueba diagnóstica y localizar puntos de corte que logren un balance óptimo entre la sensibilidad y la especificidad del diagnóstico, es el basado en las curvas ROC (acrónimo de Receiver Operating Characteristic, o Característica Operativa del Receptor). <sup>(7)</sup>

La categorización de variables pronósticas es fundamental para su uso en la toma de decisiones clínicas. A menudo se busca un único punto de corte en las puntuaciones que estratifique a los pacientes en dos categorías (alto-bajo riesgo, fallecido-vivo, etc.). Estas categorías se pueden usar para hacer recomendaciones de tratamiento, determinar la elegibilidad del estudio o controlar los diferentes pronósticos de los pacientes en el diseño de un ensayo clínico. <sup>(4, 8)</sup>

En los últimos años la autora de este artículo ha observado el incremento de tesis doctorales en medicina que tienen como objetivo construir y validar escalas predictivas para la estratificación del riesgo de ocurrencia de algún evento de interés, en las cuales el proceso de determinación de los puntos de corte se realiza de forma manual, por lo cual surge la necesidad de elaborar un software que automatice todo este proceso.

De aquí que el objetivo del presente trabajo es la elaboración de un software desarrollado en el lenguaje de programación R que implemente los diferentes métodos para dicotomizar variables continuas en modelos de predicción clínica, donde la variable respuesta es dicotómica, y determine el punto de corte óptimo en la estratificación de pacientes en dos categorías, a partir de escalas de predicción.

## Métodos

Para la realización de este trabajo se hizo una revisión sistemática de la literatura relacionada con el tema en el sector de la salud. Se investigaron los métodos existentes para dicotomizar variables y cómo estos son aplicados por los especialistas de la salud.

De los métodos encontrados en la literatura reportada, para dicotomizar variables continuas, en el presente trabajo se implementan: la representación gráfica, el cálculo de percentiles, la determinación del estadígrafo Chi-Cuadrado en tablas de contingencia 2x2 para cada valor de la variable a dicotomizar, y el que emplea los coeficientes de la regresión logística; los cuales se describen brevemente a continuación, junto con las funciones implementadas para su ejecución. Fuentes presenta una descripción detallada de cada uno de los métodos antes mencionados. <sup>(6)</sup>



Además, se implementa el método basado en las curvas ROC para evaluar la eficacia de una prueba diagnóstica y localizar su punto de corte óptimo, en el sentido que logre un balance óptimo entre la sensibilidad y la especificidad de la prueba diagnóstica.

Luego de determinar los puntos de corte candidatos por cada método, para la selección del punto de corte óptimo se calculan la sensibilidad, especificidad, el índice de Youden y el área bajo la curva (AUC), los cuales son empleados como índices de exactitud. <sup>(3, 4, 6)</sup>

### La representación gráfica

Verificada la existencia de una relación entre la variable cuantitativa y la variable dependiente que caracteriza el fenómeno estudiado, se debe examinar de forma gráfica dicha relación. Si la variable dependiente es una variable cualitativa (por ejemplo, dicotómica, como en los modelos de regresión logística binaria), la gráfica será en general poco informativa, salvo que haya una separación bastante definida de los casos investigados. <sup>(6)</sup> Fuentes Smith plantea que una gráfica mucho más ilustrativa se obtiene dividiendo la variable continua en intervalos de igual longitud y calculando la proporción de eventos de interés para cada uno de ellos, representando entonces dicha proporción frente al valor correspondiente al centro de cada intervalo. <sup>(6)</sup> En el software PCorteSoft se implementan ambas formas de representar gráficamente la relación entre la variable cuantitativa y la dependiente.

### El cálculo de percentiles

En algunas tesis doctorales, una vez obtenida la escala para pronosticar determinado evento, se procede a clasificar el riesgo de presentar el evento de interés en dos o más categorías, para lo cual se divide la distribución de la escala en dos o más zonas delimitadas por percentiles. Por tal razón, en el software elaborado se implementa el cálculo de percentiles, para lo cual se ordenan los valores de la escala obtenidos en cada paciente y se calcula, mediante la función **quantile** del sistema base de R, los percentiles solicitados por el usuario.

### Área bajo la curva ROC

Para la categorización de escalas pronósticas ( $Y$ ), el investigador debe previamente haber aplicado la regresión logística y determinado los pesos de las variables que conformarán la escala ( $x$ ), siempre que, el investigador haya comprobado el cumplimiento de los supuestos de la regresión logística, con énfasis en el referente a la existencia de una relación lineal entre las variables independientes y el logit de la variable respuesta. Para asignar estos pesos ( $w$ ), algunos investigadores utilizan la variante de redondear, al número entero superior, el resultado de multiplicar por 10 los coeficientes  $\beta_i$  correspondientes a los predictores del modelo de predicción obtenido, y otros emplean el valor redondeado de la exponencial de este coeficiente ( $e^{\beta_i}$ ). <sup>(9,10, 11)</sup>



Hanley y McNeil plantean que Bamber reconoció que la probabilidad de clasificar correctamente a un par (normal, anormal) está íntimamente relacionada con el valor del estadígrafo  $W$  en la prueba de Wilcoxon-Mann-Whitney, y demuestran empíricamente, que si se aplica esta prueba a las valoraciones dadas a imágenes de tomografía computarizada de sujetos normales y enfermos, se obtiene la misma cantidad que si se calcula el área bajo la curva ROC correspondiente, utilizando la regla trapezoidal.<sup>(12)</sup> Además, muestran como las propiedades estadísticas de  $W$  pueden utilizarse para predecir las propiedades estadísticas del área bajo la curva ROC.

Para la implementación del método del área bajo la curva ROC se emplean las expresiones dadas por Hanley y McNeil para el cálculo del área bajo la curva y su intervalo de confianza.<sup>(12)</sup> Estas expresiones son válidas solo en el caso de tener una variable categórica ordenada, la cual surge, en el caso de escalas, cuando las variables que intervienen en la definición de la misma son discretas. De no cumplirse esta condición, se emplea la regla trapezoidal para calcular el área bajo la curva ROC y se da la posibilidad de elegir el método a emplear para obtener el intervalo de confianza, el cual puede ser el de Wald o el del *score*, también conocido como de Wilson.<sup>(13)</sup> Además, tiene implementada la prueba de hipótesis para la significación estadística del AUC, mostrada en la expresión (1):

$$H_0: AUC = 0,5; H_a: AUC \neq 0,5(1)$$

Para la implementación de PCorteSoft se empleó el lenguaje de programación R (versión 4.0.2) usando las ventajas que este presenta para el trabajo con vectores y RStudio (versión 1.3.1093), entorno de desarrollo integrado libre y de código abierto para desarrollar software escrito en R. Todas las salidas se muestran en forma de tablas y gráficos.

### La determinación del estadígrafo Chi-Cuadrado

En el proceso para dicotomizar una variable continua ( $x$ ), el número máximo de puntos de corte candidatos es  $k - 1$ , donde  $k$  es el número de valores únicos (valores no repetidos) de la variable en el rango de valores que toma en la muestra. Como se plantea Brent, algunos investigadores han sugerido excluir el 5 o 10 % de los valores más extremos a ambos lados de la distribución de la variable para evitar tener valores pequeños en uno de los grupos tras la dicotomización, evitando así pérdidas sustanciales de potencia estadística. El 90 u 80 % de los datos internos de la distribución se denomina intervalo de selección (IS) y es el que se emplea para elegir el punto de corte.<sup>(4)</sup>

Para cada punto de corte candidato dentro del IS ( $x_c$ ), se construye una tabla de contingencia de 2x2, como la que se muestra en la tabla 1, donde Estado representa la variable dependiente dicotómica, y se aplica la prueba Chi cuadrado de asociación. Un modelo de punto de corte puede ser apropiado si la prueba Chi cuadrado resulta significativa (p-valor menor que el nivel de significación elegido por el investigador). El punto de corte óptimo se define a menudo como el punto de corte candidato con el p-valor más pequeño. Este método para estimar un punto de corte se denomina método o enfoque del p-valor mínimo.<sup>(4)</sup>



El enfoque del p-valor mínimo implica múltiples pruebas que ponderan el tamaño del error de tipo I.<sup>(4, 8)</sup> El método más sencillo y quizás el más popular para contrarrestar este problema es el de la corrección de Bonferroni, el cual divide el nivel de significación deseado (usualmente 0,05) por el número de puntos de corte candidatos examinados. La corrección de Bonferroni asume que todas las pruebas estadísticas que se realizan son independientes entre sí y, por lo tanto, se considera una técnica conservadora cuando las pruebas no son independientes. Brent y colaboradores plantean que el conservadurismo de la técnica puede ser una característica atractiva a la hora de determinar la significación estadística de un punto de corte óptimo, ya que sólo las relaciones fuertes entre la variable dicotomizada y la variable dependiente se considerarían estadísticamente significativas.<sup>(4, 14)</sup>

**Tabla 1** - Tabla de contingencia 2x2.

	$x \leq x_c$	$x > x_c$
Estado = 0	a	b
Estado = 1	c	d

Otros métodos para contrarrestar la ponderación de las tasas de error de tipo I se basan en el empleo de funciones matemáticas de los p-valores observados para estimar los p-valores ajustados, entre los cuales se encuentran las propuestas por Altman et al. para el caso de que se excluya el 5 % o el 10 % de los valores más extremos de la variable a ambos lados de su distribución, y se muestran en las expresiones (2) y (3), respectivamente.

$$p_{ajustado} = -3,13 * p_{min} * (1 + 1,65 \ln(p_{min})) \quad (2) \quad p_{ajustado} =$$

$$-1,63 * p_{min} * (1 + 2,35 \ln(p_{min})) \quad (3)$$

donde  $p_{min}$  es el p-valor más pequeño obtenido al aplicar k veces la prueba Chi cuadrado, y  $p_{ajustado}$  es el valor corregido.<sup>(6)</sup>

La función de PCorteSoft en la que se implementa este método determina la cantidad de valores únicos o no repetidos, de la variable a dicotomizar (puntos de corte candidatos) y para cada uno de ellos calcula el estadígrafo Chi cuadrado. Además, obtiene los p-valores asociados a la prueba de asociación y los p-valores ajustados según el método de Bonferroni o el de Altman et al. para cada uno de los puntos de cortes candidatos. Devuelve como punto de corte óptimo aquel para el cual el Chi cuadrado alcanza el máximo.



El área bajo la curva ROC se calcula a partir de los valores de la variable dependiente (Estado) y su estimación (Estado.E) obtenida a partir del punto de corte óptimo, esto es:

$$Estado.E(i) = \begin{cases} 1 & \text{si } x(i) \geq x_c \\ 0 & \text{si } x(i) < x_c \end{cases} \quad (4)$$

donde  $x(i)$  representa el valor de la variable continua en el  $i$ -ésimo paciente analizado.

Para el cálculo del AUC se generan varios puntos de corte, el valor más pequeño es el valor mínimo de la variable *Estado.E* menos 1 y el más grande, su valor máximo más 1. Todos los demás puntos de corte son los promedios de dos valores de *Estado.E* consecutivos.

PCorteSoft brinda la posibilidad de elegir el método a emplear para obtener el intervalo de confianza del área bajo la curva ROC, el cual puede ser el de Wald o el de Wilson, como se explicó anteriormente. <sup>(13)</sup>

### Empleando los coeficientes de la regresión logística binaria

En este caso se toma como valor de punto de corte ( $x_c$ ) para la clasificación, el valor de la variable continua  $x$  para el cual la probabilidad de ocurrencia del evento de interés es 0,5. Esto permite

obtener  $x_c = \frac{-\beta_0}{\beta_1}$  como punto de corte candidato, donde  $\beta_0$  y  $\beta_1$  son los coeficientes del modelo de regresión logística binario:

$$\ln \left( \frac{p}{1-p} \right) = \beta_0 + \beta_1 * x \quad (5)$$

para el cual se debe verificar que:  $\beta_1$  sea significativamente diferente de cero, el coeficiente de determinación  $R^2$  sea un valor cercano a uno, y la prueba de bondad de ajuste de Hosmer y Lemeshow resulte no significativa.

La función en la que se implementa el método, realiza el ajuste del modelo de regresión, verifica si el coeficiente  $\beta_1$  es significativamente diferente de cero, y solo en este caso devuelve el punto de corte  $x_c$ . Para calcular el área bajo la curva ROC y su intervalo de confianza se emplea el mismo procedimiento descrito en el método anterior.

## Resultados

Se desarrollaron cinco métodos para dicotomizar variables cuantitativas, para los que se implementaron sendas funciones en el lenguaje R, las cuales se describen a continuación:



**Punto.Corte.Grafico(Estado, x, Particiones=0)**

Esta función permite hacer la representación gráfica de los datos particionando o no la variable continua a dicotomizar. Sus parámetros son:

- **Estado:** valores de la variable dependiente o de respuesta en el modelo predictivo; toma valor uno (1) si el resultado investigado está presente y cero (0) en caso contrario.
- **x:** variable continua que se desea dicotomizar para emplear en el modelo predictivo.
- **Particiones:** determina el método empleado para hacer la representación gráfica. Si tiene valor cero (valor que toma por defecto), la variable  $x$  no es particionada; en otro caso, el recorrido de esta variable es particionado en tantos intervalos como valor se le haya asignado. Se calcula la proporción del evento de interés para cada uno de esos intervalos y se representa gráficamente la relación existente entre dicha proporción y el valor correspondiente al centro de cada intervalo.

**Punto.Corte.Percentil(Y, p)**

- **Y:** vector con los valores de la escala obtenidos en cada paciente.
- **p:** el o los valores del o de los percentiles que se desean calcular.

La función devuelve el o los percentiles solicitados.

**Punto.Corte.Escala(Estado, x, w = NULL, conf = 0.95, Tipo = c("D", "C"),**

**IC = c("Wald", "Wilson", "Bamber"), g.ROC = TRUE, g.IY = TRUE)**

Esta función calcula el punto de corte óptimo aplicando el método del área bajo la curva ROC. Sus parámetros se describen a continuación:

- **Estado:** valores de la variable dependiente o de respuesta en el modelo predictivo; toma valor uno (1) si el resultado investigado está presente y cero (0) en caso contrario.
- **x:** matriz de orden  $n \times p$  con los valores de las  $p$  variables que definen la escala en la muestra de tamaño  $n$  seleccionada, o un vector de longitud  $n$  con los valores de la escala obtenidos en cada paciente.
- **w:** los pesos asociados a cada una de las variables que definen la escala (solo en el caso de que  $x$  sea una matriz).
- **conf:** nivel de confianza deseado para construir el intervalo de confianza del área bajo la curva ROC. Por defecto asume el valor 0,95.
- **Tipo:** para especificar si  $x$  toma valores discretos ("D") o continuos ("C").



- **IC**: especifica el método a emplear para construir el intervalo de confianza del área bajo la curva ROC, el cual puede ser el de Wald, el de Wilson o el de Bamber.
- **g.ROC**: variable booleana indicando si el gráfico del área bajo la curva ROC debe mostrarse.
- **g.IY**: variable booleana indicando si el gráfico de los valores del índice de Youden debe mostrarse.

Como salida devuelve los valores de los puntos de corte analizados; los valores de la sensibilidad, la especificidad y el índice de Youden para cada uno de estos puntos de corte. Además, para el o los valores del punto de corte óptimo, calculado por el software como el punto de corte donde el índice de Youden es mayor, devuelve el valor estimado del área bajo la curva ROC; el límite inferior y superior del intervalo de confianza de dicha área; calculado con el nivel de confiabilidad especificado; y el p-valor asociado a la prueba de hipótesis mostrada antes en la expresión (1). Muestra, además, el gráfico de la curva ROC y de los valores del índice de Youden, en caso de ser solicitados.

**Punto.Corte.Chi(Estado, x, Excluir=c(0, 5, 10), Metodo=c("Bonf", "Altman"), conf=0.95,**

**IC = c("Wald", "Wilson", "Bamber"), Tipo=c("D", "C"))**

Esta función implementa la determinación del estadígrafo Chi-Cuadrado, dándole al usuario la posibilidad de elegir el método a emplear para corregir la ponderación del error del tipo I (Bonferroni o Altman et al.). Devuelve el valor del punto de corte óptimo ( $x_c$ ), una estimación del área bajo la curva ROC (AUC), su desviación estándar, intervalo de confianza, el p-valor asociado a la prueba de hipótesis (1) y una tabla con los valores de los puntos de corte candidatos, el valor del estadígrafo Chi cuadrado asociado a cada uno de ellos, el p-valor de la prueba Chi cuadrado de asociación y el p-valor ajustado según el método seleccionado. Además, muestra el gráfico de los valores de los puntos de corte candidatos contra su correspondiente valor Chi cuadrado. Sus parámetros son:

- **Estado, conf, IC y Tipo**: tienen igual significado que en la función anterior.
- **x**: variable continua que se desea dicotomizar para emplear en el modelo predictivo.
- **Excluir**: porcentaje de valores extremos que serán excluidos a ambos lados de la distribución de la variable continua. Puede tomar los valores 5 o 10, que representan el 5 % y 10 % de los datos a excluir, respectivamente, o el valor 0 si no se desea excluir valores extremos.
- **Método**: parámetro para especificar el método a emplear para corregir la ponderación del error del tipo I. Sus valores posibles son "Bonf", si se desea emplear el método de Bonferroni, o "Altman", si se desea emplear el método propuesto por Altman et al.

**Punto.Corte.RL(Estado, x, alfa=0.05, IC=c("Wald", "Wilson", "Bamber"),**



### Tipo=c("D","C"))

Esta función implementa el cálculo del punto de corte a partir de los coeficientes del modelo de regresión logística binaria. <sup>(5)</sup> Realiza el ajuste de este modelo y si la prueba de Wald para el coeficiente  $\beta_1$  resulta significativa ( $p - \text{valor} < \text{alfa}$ ) devuelve el valor del punto de corte calculado, una estimación del área bajo la curva ROC, su desviación estándar, intervalo de confianza y el p-valor asociado a la prueba de hipótesis. <sup>(1)</sup> En caso contrario, la variable independiente ( $x$ ) no influye en la dependiente (Estado) y no tiene sentido calcular el punto de corte, mostrándose el comentario: "El coeficiente B1 no es significativamente diferente de cero". Los parámetros de la función son:

- **Estado,  $x$ , IC y Tipo:** tienen igual significado que en las funciones antes definidas.
- **alfa:** nivel de significación para realizar la prueba de Wald ( $H_0: \beta_1 = 0$  contra  $H_a: \beta_1 \neq 0$ ). Por defecto toma el valor 0,05.

Además de las funciones descritas anteriormente, fueron implementadas otras para el cálculo del área bajo la curva ROC y sus intervalos confidenciales, las cuales son llamadas desde las funciones **Punto.Corte.Chi**, **Punto.Corte.Escala** y **Punto.Corte.RL**, y se describen a continuación.

Para calcular área bajo la curva ROC a partir de diferentes datos de entrada:

### **AUC.IC.Bamber( $k$ , $XE$ , $XS$ , $NE$ , $NS$ , $conf = 0.95$ )**

Devuelve el valor estimado del área bajo la curva ROC; el límite inferior y superior del intervalo de confianza de dicha área y el p-valor asociado a la prueba de hipótesis, según las expresiones dadas por Bamber. <sup>(11)</sup> Solo es empleada para el caso de variable categórica ordenada. Sus parámetros son:

- **$k$ :** niveles de la variable categórica ordenada que se desea dicotomizar.
- **$XE$  y  $XS$ :** vectores de dimensión  $k$ , con la cantidad de enfermos y sanos que toman los  $k$  niveles de la variable a dicotomizar.
- **$NE$  y  $NS$ :** cantidad de enfermos y sanos en la muestra analizada.
- **conf:** nivel de confiabilidad deseado, por defecto 0,95.

### **AUC(Estado, Estado.E)**

Devuelve el área bajo la curva ROC obtenida a partir de los valores de la variable dependiente (Estado) y su estimación a partir de un punto de corte (Estado.E). Sus parámetros son:

- **Estado:** valores de la variable dependiente o de respuesta en el modelo predictivo.



- **Estado.E:** valores estimados de la variable dependiente obtenidos a partir del punto de corte óptimo (ver expresión (4))

Para el cálculo del intervalo de confianza a partir de los métodos de Wald y de Wilson:

#### **IC.Wilson(Area.est, conf, n)**

Calcula el intervalo de confianza según el método de Wilson. Sus parámetros son:

- **Area.est:** valor estimado del área bajo la curva ROC.
- **conf:** nivel de confianza con el cual se desea construir el intervalo de confianza.
- **n:** tamaño de la muestra

#### **IC.Wald(Area.est, conf, n)**

Calcula el intervalo de confianza según el método clásico o de Wald. Sus parámetros son los mismos definidos para la función anterior.

Por último, y con el objetivo de poder definir el mejor punto de corte luego de haber aplicado todos o varios de los métodos descritos en este artículo, se implementó la función **Indices.Exactitud**, la cual, partiendo de los valores de la variable dependiente (**Estado**), los de la variable a categorizar (**Y**) y el vector con los valores de los puntos de corte obtenidos de la aplicación de los métodos empleados (**PC**), calcula la sensibilidad, especificidad, el índice de Youden, el área bajo la curva ROC, su desviación estándar e intervalo de confianza y el p-valor asociado a la prueba de hipótesis referente al área bajo la curva ROC; para que estos sean empleados por el usuario en su elección del punto de corte óptimo. Su encabezado es **Indices.Exactitud(Estado, Y, PC)**.

Estos resultados solo ayudan al investigador en la elección del punto de corte óptimo, la decisión final dependerá también de su experticia.

## Discusión

Para la ejemplificación del software desarrollado se tomó como muestra los datos asociados al cáncer de pulmón de una investigación desarrollada en el 2013, en la cual se realizó un estudio analítico, transversal y retrospectivo de casos y controles, que abordó como evento de interés la presencia de factores de riesgo en pacientes tratados por cáncer de pulmón, próstata o mama, en la consulta externa de los Hospitales Oncológico “Conrado Benítez” y General “Dr. Juan Bruno Zayas” de la provincia Santiago de Cuba, desde el 1º de julio del 2012 hasta el 31 de junio del 2013. En el caso de los controles se estudiaron residentes de esa provincia que no tenían ningún tipo de cáncer.

(15)



En la tabla 2 se presentan las estadísticas asociadas al modelo de regresión lineal simple obtenido por Mederos. <sup>(15)</sup> En el anexo se muestra la codificación de estas variables.

**Tabla 2** - Resultados de la regresión logística binaria.

Variables	B	Error estandar	Wald	gl	p-valor	Exp(B)	IC 95 % para	
							Inferior	Superior
EDAD	2,816	0,724	15,144	1	0,000	16,705	4,045	68,979
PAG	2,367	0,828	8,173	1	0,004	10,662	2,105	54,011
POR	1,742	0,728	5,726	1	0,017	5,711	1,371	23,801
PCA	2,147	0,707	9,217	1	0,002	8,557	2,140	34,216
RESID	-3,808	0,992	14,726	1	0,000	0,022	0,003	0,155
Constante	-0,815	0,792	1,060	1	0,303	0,442	-----	----

Tomada de Yohani García Mederos, 2013.

B: coeficientes del modelo; gl: grados de libertad; Exp(B): exponencial de los coeficientes del modelo; IC: intervalo de confianza; EDAD: edad mayor de 50 años, PAG: Antecedente genético; POR: Ocupación de riesgo; PCA: Contaminación ambiental; RESID: residencia urbana o rural.

Como plantea Mederos, y se puede observar en la tabla 2, la residencia (RESID) tiene un coeficiente negativo y un intervalo de confianza con sus dos límites por debajo de uno, por lo que constituye un factor protector tener residencia urbana (RESID=1).<sup>(15)</sup> Por tal razón, se decidió, para el presente estudio, invertir la codificación de esta variable, definiendo la nueva variable:

$$RESID2 = \begin{cases} 1 & \text{si el paciente vive en zona rural} \\ 0 & \text{si el paciente vive en zona urbana} \end{cases}$$

lo cual equivale a invertir el signo del coeficiente de la variable residencia en el modelo obtenido por Mederos. A partir del modelo predictivo, se obtuvo la escala predictiva de cáncer de pulmón, considerando como pesos los coeficientes  $\beta$  del modelo de regresión logística binaria redondeados al número entero superior, la cual viene dada por:

$$Y = 30 * EDAD + 20 * PAG + 20 * POR + 20 * PCA + 38 * RESID2$$

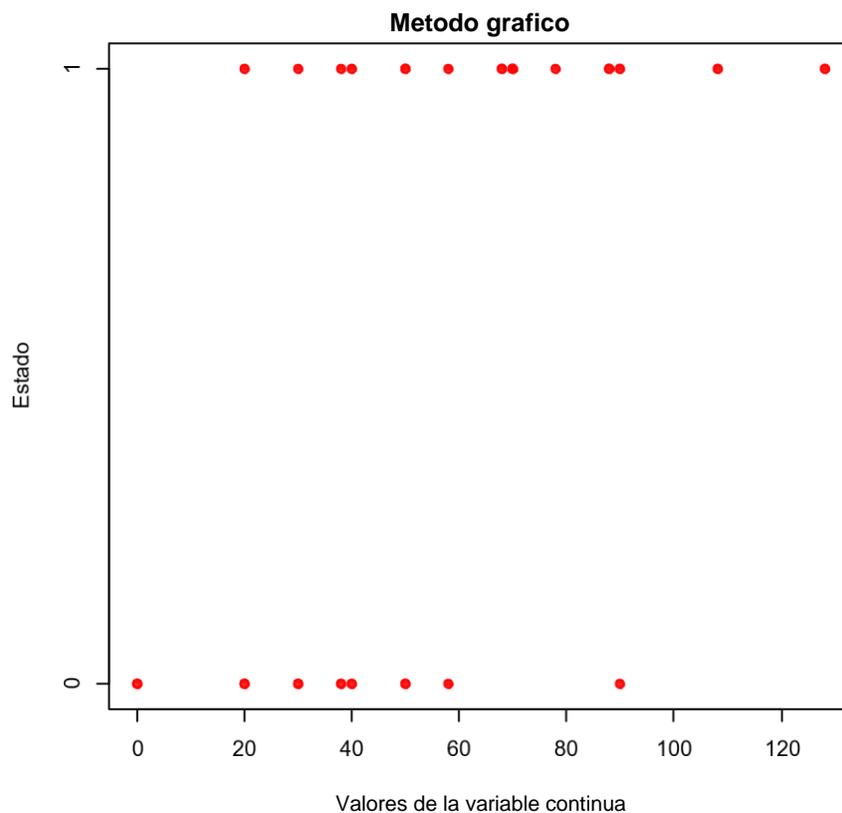


En la ejemplificación se consideró aplicar los métodos presentados para determinar el punto de corte óptimo en la estratificación de pacientes en dos categorías, a partir de la escala predictiva de cáncer de pulmón. De forma similar se procede si lo que se desea es dicotomizar una variable continua.

En las figuras 1a y 1b se presentan dos gráficos que muestran la relación entre la variable cuantitativa a dicotomizar (la escala en este caso) y la variable dependiente (cáncer), las cuales son el resultado de las siguientes llamadas:

> Punto.Corte.Grafico(Estado = Cancer, x = Y, Particiones = 0)

> Punto.Corte.Grafico(Estado = Cancer, x = Y, Particiones = 8)

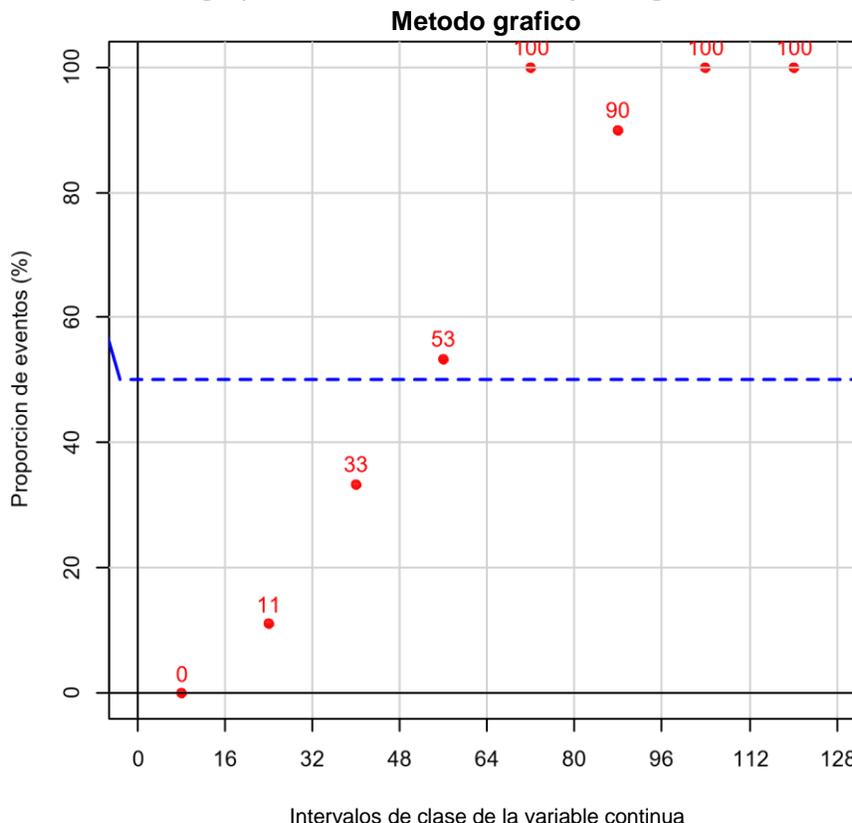


**Fig.1a** - Representación gráfica de la relación entre la variable de interés y la variable continua sin particionar.

Al aplicar el método gráfico sin particionar la variable cuantitativa, la gráfica que se obtiene (figura 1a) no es informativa, lo cual coincide con lo planteado por Fuentes. <sup>(6)</sup> Si se divide el rango de la variable cuantitativa en ocho intervalos de igual longitud (figura 1b) se estima como punto de corte



óptimo un valor entre 64 y 80. Tomando el punto medio de este intervalo, el punto de corte vendría dado por el valor 72, es decir, pacientes con valores de la escala superiores o iguales a 72 puntos clasificarían como de alto riesgo y en caso contrario, de bajo riesgo.



**Fig.1b** - Representación gráfica de la relación entre la variable de interés y la variable continua particionando en ocho intervalos.

Para ejemplificar el cálculo del punto de corte por el método del percentil, se solicitó arbitrariamente el percentil 75, obteniéndose como punto de corte el valor 70, para el cual las zonas de riesgo se definirían como:

Zona de riesgo bajo: si  $Y \leq 70$  Zona

de riesgo alto: si  $Y > 70$ .

Estos resultados se obtuvieron con el llamado de la función Punto.Corte.Percentil, cuya salida se muestra a continuación:

> Punto.Corte.Percentil(Y, p = 0.75)

75%



70

Al invocar la función Punto.Corte.Escala para calcular el punto de corte mediante el método del área bajo la curva y empleando el método de Bamber para hallar el intervalo de confianza, se obtiene la siguiente salida:

```
> Punto.Corte.Escala(Estado = Cancer, x = Y, conf = 0.95, Tipo = "D", IC = "Bamber",
                    g.ROC = TRUE, g.IY = TRUE)
```

Punto(s) de corte óptimo: 45

Área bajo la curva ROC: 0.86

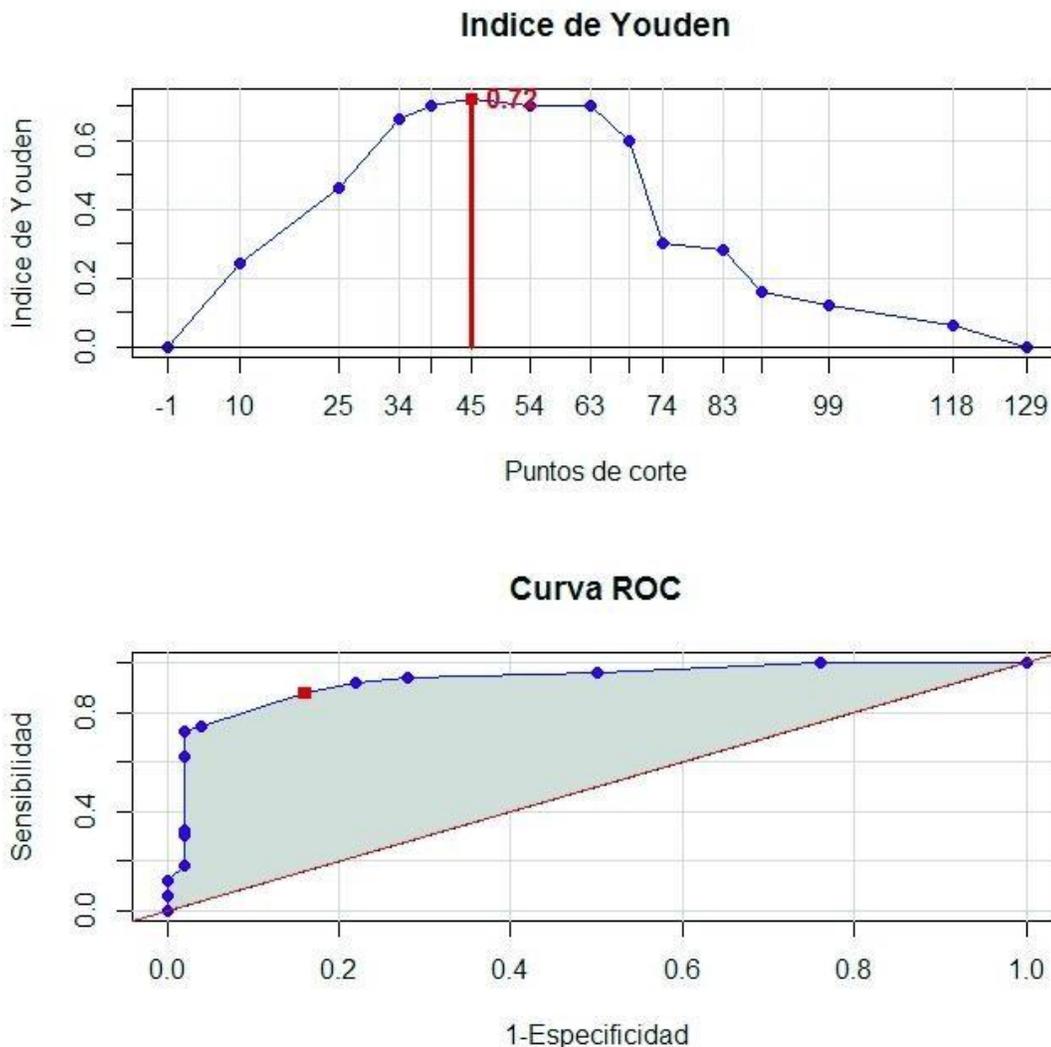
Desviación Estándar: 0.0402664

Intervalo de confianza según Bamber: (0.781079, 0.938921) p-valor:  
3.87602e-19

	Ptos. de Corte	Sensibilidad	1-Especificidad	Indice de Youden
1	-1	1.00	1.00	0.00
2	10	1.00	0.76	0.24
3	25	0.96	0.50	0.46
4	34	0.94	0.28	0.66
5	39	0.92	0.22	0.70
<b>6</b>	<b>45</b>	<b>0.88</b>	<b>0.16</b>	<b>0.72</b>
7	54	0.74	0.04	0.70
8	63	0.72	0.02	0.70
9	69	0.62	0.02	0.60
10	74	0.32	0.02	0.30
11	83	0.30	0.02	0.28
12	89	0.18	0.02	0.16
13	99	0.12	0.00	0.12
14	118	0.06	0.00	0.06
15	129	0.00	0.00	0.00



Este método define como punto de corte óptimo el valor de la escala correspondiente al valor 45. Para este punto de corte, se obtiene una estimación del área bajo la curva ROC igual a 0,86, con una desviación estándar de 0,0402664. Se puede afirmar, con un nivel de confiabilidad del 95 %, que el área bajo la curva ROC se encuentra en el intervalo (0,7811; 0,9389) y que la misma es significativamente diferente de 0,5 con un nivel de significación del 5 % ( $p$ -valor= 3,87602e-19). En la figura 2 se muestra la salida gráfica que brinda la función Punto.Corte.Escala.



**Fig. 2** - Representación gráfica del índice de Youden según punto de corte (figura superior) y área bajo la curva ROC (figura inferior).



En el método de la determinación del estadígrafo Chi-Cuadrado se ha elegido el método propuesto por Altman para contrarrestar la ponderación de las tasas de error de tipo I, y el de Wilson para construir el intervalo de confianza del área bajo la curva ROC. Como estamos trabajando con los valores de una escala de predicción, se decidió no eliminar valores extremos (Excluir = 0), lo cual si es aconsejable cuando se trabaja con variables continuas que se desean dicotomizar.

Al llamar a la función con los parámetros correspondientes, se obtiene la siguiente salida:

```
>Punto.Corte.Chi(Estado = Cancer, x = Y, Excluir = 0, Metodo = "Altman", IC = "Wilson",
                conf = 0.95, IC = "Wilson", Tipo = "D")
```

Punto(s) de corte óptimo: 68

Área bajo la curva ROC: 0.85

Desviación Estándar: 0.0356577

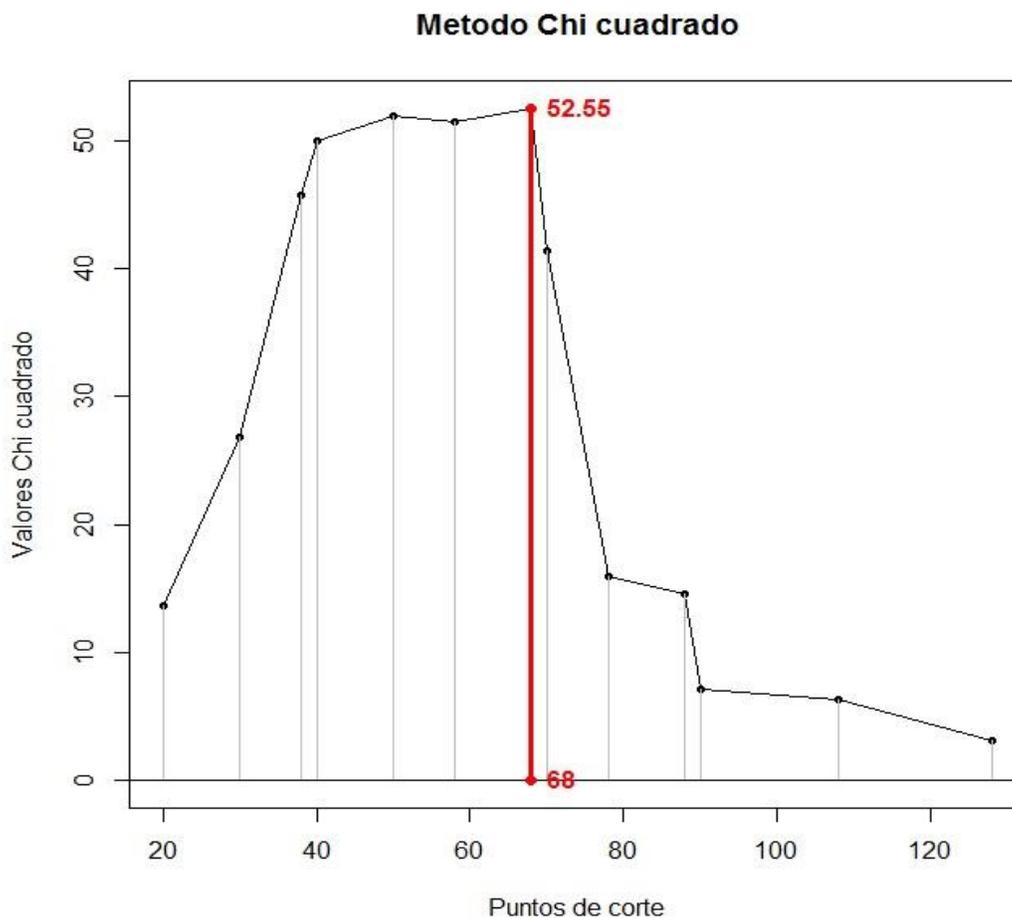
Intervalo de confianza según Wilson: (0.767164, 0.906940) p-valor:  
4.82586-23

	PCorte	Chi.cuadrado	p.valor	p.ajustado
1	20	13.640	2.218e-04	6.788066e-03
2	30	26.840	2.211e-07	1.261896e-05
3	38	45.780	1.326e-11	1.250632e-09
4	40	49.980	1.554e-12	1.592726e-10
5	50	51.920	5.772e-13	6.136346e-11
6	58	51.490	7.188e-13	7.581696e-11
7	<b>68</b>	<b>52.550</b>	<b>4.189e-13</b>	<b>4.504878e-11</b>
8	70	41.360	1.266e-10	1.084569e-08
9	78	15.950	6.517e-05	2.299890e-03
10	88	14.580	1.341e-04	4.361640e-03
11	90	7.111	7.661e-03	1.304691e-01
12	108	6.383	1.152e-02	1.782156e-01
13	128	3.093	7.864e-02	6.378091e-01



De la salida y la figura 3, que se muestra a continuación, se obtiene que este método define como punto de corte óptimo el valor de la escala correspondiente al 68, lo cual no entra en contradicción con lo obtenido por el método gráfico, pero dista del obtenido con el método del área bajo la curva ROC.

Para este punto de corte se obtiene una estimación del área bajo la curva ROC igual a 0,85; con una desviación estándar de 0,0356577. Se puede afirmar, con un nivel de confiabilidad del 95 %, que el área bajo la curva ROC se encuentra en el intervalo (0,767164; 0,906940) y que la misma es significativamente diferente de 0,5 con un nivel de significación del 5 % (p-valor= 4,82586-23).



**Fig. 3** - Representación gráfica de los valores de Chi cuadrado obtenidos a partir de todas las tablas de contingencia posibles, según valores de la variable continua.



Para calcular el intervalo de confianza del área bajo la curva ROC usando el método de la regresión logística binaria se empleó el método de Wilson. Al llamado de la función correspondiente, el software muestra la siguiente salida:

>Punto.Corte.RL(Estado = Cancer, x = Y, alfa = 0.05, IC = "Wilson")

Punto de corte óptimo: 46.6044 Área

bajo la curva ROC: 0.86

Desviación Estándar: 0.0347222

Intervalo de confianza según Wilson: (0.778628,0.914737) p-valor:

1.733298e-25

Este método define como punto de corte óptimo el valor de la escala correspondiente al valor 46,6; el cual se aproxima al valor obtenido por el método del área bajo la curva ROC. Para este punto de corte, se obtiene una estimación del área bajo la curva ROC igual a 0,86; con una desviación estándar de 0,03472. Se puede afirmar, con un nivel de confiabilidad del 95 %, que el área bajo la curva ROC se encuentra en el intervalo (0,77863; 0,9147) según Wilson, y que la misma es significativamente diferente de 0,5 con un nivel de significación del 5 % (p-valor = 1,733298e-25).

Obtenidos los puntos de corte a partir de la aplicación de los cinco métodos implementados, se procede a calcular la sensibilidad, especificidad y el índice de Youden, para emplearlos como índices de exactitud, los cuales se obtienen al invocar la función `Indice.Exactitud`. Esta función también devuelve el área bajo la curva ROC, la desviación estándar e intervalo de confianza del área bajo la curva ROC y el p-valor asociado a la prueba de hipótesis referente al área bajo la curva ROC. Estos resultados se muestran en las tablas 3 y 4, para los tres tipos de intervalos de confianza que admite el software, el de Wald, de Wilson y el de Bamber.

De esta tabla se puede concluir que los métodos que mejores resultados muestran son el de la Regresión Logística Binaria y el del área bajo la curva ROC, cuyos resultados coinciden (para los datos suministrados), excepto en el valor reportado como punto de corte. Sin embargo, ambos puntos de corte se encuentran entre dos valores consecutivos de la escala, esto es,  $40 < 45 < 50$  y  $40 < 46,60 < 50$ .

**Tabla 3** - Punto de corte, sensibilidad, especificidad e índice de Youden para cada punto de corte.

Método	Punto de Corte	Sen	Esp	IY
Gráfico	72	0,32	0,98	0,30
Percentil 75	70	0,62	0,98	0,60



Chi cuadrado	68	0,72	0,98	0,70
Regresión Logística Binaria	46,60	0,88	0,84	0,72
Área bajo la curva ROC	45	0,88	0,84	0,72

**Tabla 4** - Punto de corte, área bajo la curva ROC, desviación estándar e intervalo de confianza del área bajo la curva ROC y p-valor.

Método	Punto de Corte	Área estimada	DS	Intervalo de confianza	p-valor
<b>Gráfico</b>					
IC="Wald"	72	0,65	0,0477	(0,5565; 0,7435)	8,30847e-04
IC="Wilson"	72	0,65	0,0469	(0,5525; 0,7364)	6,89901e-04
IC="Bamber"	72	0,65	0,0553	(0,5415; 0,7585)	6,72817e-03
<b>Percentil 75</b>					
IC="Wald"	70	0,80	0,0400	(0,7216; 0,8784)	3,19089e-14
IC="Wilson"	70	0,80	0,0397	(0,7112; 0,8666)	1,94831e-14
IC="Bamber"	70	0,80	0,0464	(0,7090; 0,8910)	1,02677e-10
<b>Chi cuadrado<sup>1</sup></b>					
IC="Wald"	68	0,85	0,0357	(0,7800; 0,9199)	5,52106e-23
IC="Wilson"	68	0,85	0,0357	(0,7672; 0,9069)	4,82586e-23
IC="Bamber"	68	0,85	0,0414	(0,7688; 0,9312)	2,99861e-17
<b>Regresión Logística Binaria</b>					
IC="Wald"	46,6044	0,86	0,0347	(0,7920; 0,9280)	1,61060e-25
IC="Wilson"	46,6044	0,86	0,0347	(0,7786; 0,9147)	1,73330e-25
IC="Bamber"	46,6044	0,86	0,0403	(0,7811; 0,9389)	3,87602e-19
<b>Área bajo la curva ROC</b>					
IC="Wald"	45	0,86	0,0347	(0,7919; 0,9280)	1,61060e-25
IC="Wilson"	45	0,86	0,0347	(0,7786; 0,9147)	1,73330e-25
IC="Bamber"	45	0,86	0,0403	(0,7811; 0,9389)	3,87602e-19

<sup>1</sup> No se especifica método, porque el método elegido no afecta estos valores.



## Conclusiones

En este artículo se ha presentado la funcionalidad del software PCorteSoft, elaborado en el lenguaje de programación R, en el cual se han implementado diferentes métodos para dicotomizar variables continuas en modelos de predicción clínica, donde la variable respuesta es dicotómica, y determinar el punto de corte óptimo en la estratificación de pacientes en dos categorías, a partir de escalas de predicción.

Con este paquete, el especialista de la salud puede disponer de una herramienta que contiene los métodos más empleados para dicotomizar variables y en dependencia de los resultados y su experticia, decidir qué punto de corte emplear.

En el empleo del software debe tenerse en cuenta el cumplimiento de ciertos supuestos y desventajas propias de algunas técnicas estadísticas utilizadas para su implementación, que el investigador analizará al seleccionar el método a emplear atendiendo a la calidad de los datos que posea.

## Referencias

1. Mendieta Pedroso M, Sotolongo López JC. El pronóstico y su importancia en la práctica clínica. Rev cubana med [Internet]. 2019 Sep [Citado 14/02/2023];58(3):e1322. Disponible en: [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S0034-75232019000300006&lng=es](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0034-75232019000300006&lng=es)
2. Rendón Macías ME, Castillo Ivón AS. Metodología para la elaboración de los estudios sobre pronóstico. RAM [Internet]. 2023 Feb [Citado 14/02/2023];69(1):48-55. Disponible en: <https://revistaalergia.mx/ojs/index.php/ram/article/view/1100>
3. Barrio I, Arostegui I, Rodríguez Álvarez MX, Quintana JM. A new approach to categorising continuous variables in prediction models: Proposal and validation. Statistical Methods in Medical Research [Internet]. 2017 Dec;26(6):2586-602. Disponible en: <http://doi.org/10.1177/0962280215601873>
4. Brent AW, Jayawant NM, Sumithra JM, Stephen SC, Alfred FF. Finding Optimal Cutpoints for Continuous Covariates with Binary and Time-to-Event Outcomes. Technical Report Series #79. 2006 June [Internet]. Minnesota: Mayo Clinic Rochester; 2006 [Citado 14/02/2023]. Disponible en: <https://www.mayo.edu/research/documents/biostat-79pdf/doc-10027230>
5. Cumsille F, Bangdiwala SI. Categorización de variables en el análisis estadístico de datos: consecuencias sobre la interpretación de resultados. Rev Panam Salud Publica [Internet]. 2000 Nov



[Citado 14/02/2023];8(5):348-54. Disponible en: [http://www.scielo.org/scielo.php?script=sci\\_arttext&pid=S102049892000001000005&lng=pt&nr\\_m=iso](http://www.scielo.org/scielo.php?script=sci_arttext&pid=S102049892000001000005&lng=pt&nr_m=iso)

6. Fuentes SLE. Metodología para la elección de punto de corte óptimo para dicotomizar covariables continuas. Rev Cub Gen. 2013;7(3):36-42.
7. Gemp Fuentealba R, Avendaño Bravo C, Muñoz Urrutia C. Normas y punto de corte para la Escala de Depresión del Centro para Estudios Epidemiológicos (CES-D) en población juvenil chilena. Terapia Psicológica [Internet]. 2004 [Citado 14/02/2023];22(2):145-56. Disponible en: <https://www.redalyc.org/articulo.oa?id=78522205>
8. Mazumdar M, Glassman JR. Categorizing a prognostic variable: review of methods, code for easy implementation and applications to decision-making about cancer treatments. Statistics in Medicine [Internet]. 2000 Jan;19(1):113-2. Disponible en: [http://doi.org/10.1002/\(sici\)1097-0258\(20000115\)19:1<113::aid-sim245>3.0.co;2-o](http://doi.org/10.1002/(sici)1097-0258(20000115)19:1<113::aid-sim245>3.0.co;2-o)
9. Suárez Quesada A. Índice para predecir el desarrollo de la neumonía asociada al ictus isquémico [Tesis Doctoral]. Bayamo: Universidad de Ciencias Médicas de Santiago de Cuba; 2021.
10. Santos Medina M. Escala predictiva de muerte hospitalaria por infarto agudo de miocardio [Tesis Doctoral]. Santiago de Cuba: Universidad de Ciencias Médicas de Santiago de Cuba; 2022.
11. Nieto Cabrera M. Diseño y validación de un modelo predictivo de mediastinitis en cirugía cardiaca [Tesis Doctoral]. Madrid: Universidad Complutense de Madrid; 2014.
12. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982 Apr;143(1):29-36.
13. Cepeda Cuervo E, Aguilar W, Cervantes V, Corrales M, Díaz I, Rodríguez D. Intervalos de confianza e intervalos de credibilidad para una proporción. Revista Colombiana de Estadística [Internet]. 2008 Dic [Citado 14/02/2023];31(2):211-28. Disponible en: <https://www.redalyc.org/articulo.oa?id=89912222006>
14. Liqueur B, Riou J. CPMGLM: an R package for p-value adjustment when looking for an optimal transformation of a single explanatory variable in generalized linear models. BMC Med Res Methodol [Internet]. 2019 Apr 16;19(1):79. Disponible en: <https://doi.org/10.1186/s12874-019-0711-2>
15. García Mederos Y. El Análisis Estadístico Implicativo en la identificación de factores de riesgo de cáncer de pulmón, próstata y mama en la provincia de Santiago de Cuba [Tesis Especialidad en Bioestadística Médica]. Santiago de Cuba: Universidad de Ciencias Médicas de Santiago de Cuba; 2013.



### **Conflicto de interés**

Los autores declaran que no existen conflictos de intereses.

### **Contribución de los autores**

Dra. C. Larisa Zamora Matamoros: Formulación inicial del problema a resolver, conceptualización, análisis formal, investigación, metodología, administración del proyecto, curación de datos, desarrollo del software, validación, visualización, redacción del borrador original, revisión y edición.

Dr. C. Jorge Rey Díaz Silvera: Investigación, metodología, administración del proyecto, supervisión técnica, validación, visualización, redacción del borrador original, revisión y edición.



**Anexos**

Variable	Categorías	Código	Descripción de las categorías
Variable de respuesta:			
Cáncer de pulmón (Cancer)	Si	1	Según la presencia o no del cáncer de pulmón.
	No	0	
Variables explicativas:			
1. Edad	Si	1	Paciente mayor de 50 años.
	No	0	Paciente menor o igual de 50 años.
2. Antecedente genético (PAG)	Si	1	Según la presencia del antecedente de cáncer de pulmón en familiares de primera línea (padres y/o hermanos)
	No	0	
3. Ocupación de riesgo (POR)	Si	1	Según realice o no una labor que conlleve a la exposición con cualquiera de las siguientes sustancias: asbesto, radón, arsénico, azufre, cloruro de vinilo, hematita, materiales cromados, cromatos de níquel, productos de carbón, gas mostaza, éteres de clorometilo, gasolina, derivados del hierro o berilio
	No	0	
4. Contaminación ambiental (PCA)	Si	1	Según tenga contacto mantenido o no con humos de motores de explosión (coches y vehículos de motor en general) y sistemas de calefacción, cocinas de keroseno, partículas de brea del pavimento de las calles, partículas radiactivas.
	No	0	
5. Residencia (RESID)	Urbana	1	Según viva el paciente en zona urbana o no.
	Rural	0	

