



OptimalCutpoints: An R Package for Selecting Optimal Cutpoints in Diagnostic Tests

Mónica López-Ratón

Universidad de Santiago de Compostela

María Xosé Rodríguez-Álvarez

Universidade de Vigo

Carmen Cadarso-Suárez

Universidad de Santiago de Compostela

Francisco Gude-Sampedro

Complejo Hospitalario Universitario
de Santiago de Compostela (CHUS)

Abstract

Continuous diagnostic tests are often used for discriminating between healthy and diseased populations. For the clinical application of such tests, it is useful to select a cutpoint or discrimination value c that defines positive and negative test results. In general, individuals with a diagnostic test value of c or higher are classified as diseased. Several search strategies have been proposed for choosing optimal cutpoints in diagnostic tests, depending on the underlying reason for this choice. This paper introduces an R package, known as **OptimalCutpoints**, for selecting optimal cutpoints in diagnostic tests. It incorporates criteria that take the costs of the different diagnostic decisions into account, as well as the prevalence of the target disease and several methods based on measures of diagnostic test accuracy. Moreover, it enables optimal levels to be calculated according to levels of given (categorical) covariates. While the numerical output includes the optimal cutpoint values and associated accuracy measures with their confidence intervals, the graphical output includes the receiver operating characteristic (ROC) and predictive ROC curves. An illustration of the use of **OptimalCutpoints** is provided, using a real biomedical dataset.

Keywords: optimal cutpoint, diagnostic tests, accuracy measures, ROC curve, R.

1. Introduction

Continuous diagnostic tests or biomarkers are often used for discriminating between healthy and diseased populations ($D = 0$ and $D = 1$, respectively). For their application in clinical

practice, it is useful to select a cutpoint or discrimination value c to define positive and negative test results. In general, assuming that higher marker values are associated with disease, individuals with a diagnostic test value T equal to or higher than c are classified as diseased (positive test T_+), whereas patients with a lower value are classified as healthy (negative test T_-). It should be noted, however, that this classification is not error-free. The test may err or fail in its task of detecting disease in two different ways, namely, by classifying a healthy patient incorrectly (false positive, FP) or, alternatively, by declaring a patient to be healthy when he/she is in fact diseased (false negative, FN). Conversely, the test may correctly classify a healthy patient (true negative, TN) or a diseased patient (true positive, TP). Accordingly, before routine application of a diagnostic test in practice, any errors of classification must be quantified. With respect to any given cutpoint, different measures of a test's diagnostic accuracy can be considered. The most popular accuracy measures are sensitivity (Se) and specificity (Sp). Furthermore, on the basis of these two measures, other measures can be also defined, such as positive and negative predictive values (PPV and NPV) and diagnostic likelihood ratios (DLR_+ and DLR_-).

The selection of the appropriate cutpoint is crucial to avoid erroneous conclusions being drawn in clinical practice. While the choice of the number and values of the cutpoints may be made in accordance with criteria already established by earlier studies or, for theoretical reasons, be based on clinical or physiological information (this is the desirable approach), at other times it is the researcher him/herself who has to decide on the cutpoints that are to be set on the basis of certain criteria. In medical decision-making theory and epidemiologic research, determining a cutpoint for a quantitative variable is a common problem, and has indeed been an active area of study (Miller and Siegmund 1982; Altman, Lausen, Sauerbrei, and Schumacher 1994; Lausen and Schumacher 1996; Mazumdar and Glassman 2000). Depending on the ultimate goal, several strategies for selecting optimal cutpoints in diagnostic tests have been proposed in the literature, yet all of these have, in great measure, been based on optimizing the preceding accuracy measures (see Youden 1950; Feinstein 1975; Metz 1978; Albert and Harris 1987; England 1988; Schäfer 1989; Vermont, Bosson, François, Robert, Rueff, and Demongeot 1991; Greiner 1995; Riddle and Stratford 1999, among others).

To facilitate the task of selecting optimal values in clinical practice, it is essential to have software for implementing the different optimal-cutpoint selection criteria in an environment which biomedical researchers will find user-friendly and easily understandable. Important contributions to this issue have been made by the R (R Core Team 2014) packages **DiagnosisMed** (Brasil 2010), **pROC** (Robin, Turck, Hainard, Tiberti, Lisacek, Sanchez, and Müller 2011) and **Epi** (Carstensen, Plummer, Laara, and Hills 2013). The **DiagnosisMed** package includes the estimation of optimal cutpoints by means of 10 different methods, such as the method that selects the cutoff at which sensitivity is equal to specificity (Amaro, Gude, Gonzalez-Juanatey, Iglesias, Fernandez-Vazquez, Garcia-Acuna, and Gil 1995; Greiner 1995; Hosmer and Lemeshow 2000), or that based on maximizing the diagnostic odds ratio (Kraemer 1992; Böhning, Holling, and Patilea 2011). Even though their main objective is not the selection of an optimal cutpoint, the **pROC** and **Epi** packages also include some specific function for selecting the optimal value using only one or two criteria. Specifically, the **pROC** package provides the method based on the Youden index (Youden 1950) and the criterion for the optimal point on the receiver operating characteristic (ROC) curve (Metz 1978; Swets and Swets 1979; Swets and Pickett 1982) closest to the point $(0, 1)$ (Metz 1978; Vermont *et al.* 1991), allowing the costs of the different diagnostic decisions to be incorporated into both

criteria. The **Epi** package enables the value that maximizes the sum of sensitivity plus specificity measures to be selected as the optimal cutpoint. There are other packages which provide methods for obtaining classification rules for specific models/contexts other than a medical setting. These include the R packages such as, **PresenceAbsence** which includes 12 different criteria (Freeman and Moisen 2008), and **SDMTools** with 8 methods (VanDerWal, Falconi, Januchowski, Shoo, and Storlie 2012).

However, all these packages have some limitations. Firstly, as pointed out above, some of them include very few selection criteria. Secondly, none of the packages include criteria based on predictive values and/or likelihoods ratios, i.e., they only consider optimal-cutpoint selection criteria based on sensitivity and specificity measures. This entails an important limitation from the standpoint of clinical applicability: although sensitivity and specificity are considered the fundamental operating characteristics of a diagnostic test, there are nevertheless times when such measures may not be so useful in clinical practice, since clinical staff do not have prior information about the patient's true disease status. Indeed, the problem tends to be just the opposite, and involves the need to ascertain the probability of the patient being healthy (or diseased) in a case where the test result is negative (or positive). Hence, strategies for selecting the optimal cutpoint based on predictive values can be more useful in certain situations (see Vermont *et al.* 1991; Itoh, Takahashi, Nishida, Sakagami, and Okubo 1996; Gallop, Crits-Christoph, Muenz, and Tu 2003, among others).

To address some of the remaining gaps in or limitations of the previous packages, we have implemented an R package known as **OptimalCutpoints**, specifically designed for selecting optimal cutpoints in continuous diagnostic tests. It is freely available from the Comprehensive R Archive Network (CRAN) at <http://CRAN.R-project.org/package=OptimalCutpoints>. This package enables end-users to choose from among a considerable number of strategies (34) commonly used in clinical practice for optimal-cutpoint selection (see Table 2 in Section 3). **OptimalCutpoints** includes all the methods considered in the abovementioned packages plus others that have been proposed in the literature for selection of optimal values in diagnostic tests, such as criteria based on predictive values or likelihood ratios. Moreover, it incorporates several criteria that take into account the costs of the different diagnostic decisions as well as the prevalence of the disease under study.

To illustrate the different optimal selection criteria implemented in this package, this paper goes on to consider a study conducted on 141 consecutive patients admitted to the Cardiology Department of a Teaching Hospital in Galicia (northwest Spain) for evaluation of chest pain or cardiovascular disease. The study sought to investigate the clinical usefulness of leukocyte elastase determination in the diagnosis of coronary artery disease (CAD). All patients underwent coronary angiography during investigation: 96 had coronary lesions (diseased patients) and 45 had non-stenotic coronaries (non-diseased patients). Fuller details of this data set can be found in Amaro *et al.* (1995).

The remainder of the paper is structured as follows: Section 2 briefly reviews methods for selecting optimal cutpoints in clinical practice; Section 3 explains the use of the main functions and methods of **OptimalCutpoints**; Section 4 gives an illustration of the practical application of the package using the CAD dataset; and lastly, Section 5 concludes with a discussion and possible future extensions of the package.

2. Optimal-cutpoint selection methods

The need to determine a cutpoint in continuous diagnostic tests is widely acknowledged, mainly in the medical sciences, and, as mentioned in the preceding section, diverse criteria have been proposed for selecting such cutpoints. Some authors (Gönen and Sima 2013) talk in general of two main statistical approaches to the problem of selecting an optimal cutpoint, one of which uses the ROC curve (Metz 1978; Swets and Swets 1979; Swets and Pickett 1982), and the other seeks to maximize an appropriately chosen statistical test (Mazumdar and Glassman 2000). The ROC curve is a global measure of diagnostic accuracy of a continuous test that reflects the degree of overlapping of test results in healthy and diseased populations. Moreover, it is independent of the disease prevalence. The ROC curve is obtained by plotting the coordinates $(1 - Sp(c); Se(c))$ for all possible cutpoints c , where $Se(c)$ and $Sp(c)$ are defined as

$$\begin{aligned} Se(c) &= P(T_+ | D = 1) = P(T \geq c | D = 1), \\ Sp(c) &= P(T_- | D = 0) = P(T < c | D = 0), \end{aligned}$$

under the conventional assumption that larger T values are more indicative of disease. It should be noted however that, when high marker values are linked to health, a positive test would be one where $T \leq c$, and the definition of Se and Sp should be changed accordingly. It is common for the information of a ROC curve to be summarized in a single value or index. Several such indices are proposed in the literature, with the most widely used being the area under the ROC curve (AUC) (Bamber 1975; Swets 1979). The AUC takes values ranging from 0.5 (uninformative test) to 1 (perfect test).

ROC analysis furnishes several optimal-cutpoint selection criteria based on Se and Sp measures (Green and Swets 1966; Zweig and Campbell 1993; Coffin and Sukhatme 1997; Pepe 2003), by imposing specifications of one kind or another in respect of such measures, assuming certain values, or defining a linear combination or function of both. Furthermore, ROC curve criteria allow for the choice of optimal cutpoints, based on the prevalence of the disease of interest and the relative cost ratio (risks and benefits) of the possible medical decisions (correct and incorrect) flowing from the diagnostic test result. The other widely used approach is maximization of an appropriate statistical test, often one that is based on two samples and compares the groups resulting from dichotomization. This method first appeared in the context of a binary result using the Pearson χ^2 test (Miller and Siegmund 1982) and is frequently known as the maximum χ^2 or minimum p value method (Mazumdar and Glassman 2000).

Despite this initial general grouping, the large number of optimal-cutpoint selection criteria in diagnostic tests has persuaded us to summarize the following outline by splitting the criteria into a series of subgroups. A detailed description of all the criteria incorporated in the **OptimalCutpoints** package is provided in Section 3. For the sake of clarity, however, in each of the subgroups considered in this Section, the criteria names used in the **OptimalCutpoints** package are indicated.

2.1. Criteria based on sensitivity and specificity measures

In some diagnostic situations, it is desirable to have a higher probability of detecting a TN or TP result, or for both to exceed certain values, and in such a case the optimal cutpoint should therefore be chosen with this aim in mind: hence, some minimum value is selected

which one (Se or Sp) or both of the two measures are required to exceed, and subject to this condition, the other measure is as high as possible (Schäfer 1989; Vermont *et al.* 1991; Gallop *et al.* 2003). In the **OptimalCutpoints** package, these criteria have been included under the names "MinValueSe", "MinValueSp" and "MinValueSpSe".

In a manner similar to the above strategies, analogous criteria can be defined, in which the end-user, rather than setting a single minimum value for either or both measures, sets a single target value for either the sensitivity or the specificity measures (Rutter and Miglioretti 2003) ("ValueSe" and "ValueSp" methods in **OptimalCutpoints**).

Another criterion is based on the maximization of one of the two measures, i.e., "MaxSe" and "MaxSp" methods (Bortheyry, Malerbi, and Franco 1994; Filella, Alcover, Molina, Giménez, Rodríguez, Jo, Carretero, and Ballesta 1995; Álvarez García, Collantes-Fernández, Costas, Rebordosa, and Ortega-Mora 2003), though this procedure proves more extreme, since the choice of an optimal cutpoint should generally imply an equilibrium between Se and Sp . Consequently, the equality or simultaneous maximization of these two quantities ("SpEqualSe" and "MaxSpSe" methods in **OptimalCutpoints**, Riddle and Stratford 1999; Peng and So 2002; Gallop *et al.* 2003) or the maximization/minimization of a given combination tends to be more appropriate. Some examples of criteria in this setting are: the *Youden index* (Youden 1950) or incorporating the costs of incorrect classifications of the diagnosis, the *generalized Youden index* (*GYI*, Geisser 1998; Greiner, Pfeiffer, and Smith 2000; Schisterman, Perkins, Liu, and Bondell 2005, both possibilities are included in the "Youden" method of the package); *efficiency or proportion of cases correctly classified* (Feinstein 1975, "MaxEfficiency" method); criterion for the *optimal point on the ROC curve closest to the point (0, 1)* (Metz 1978, "ROC01" method); or criteria based on predictive values and likelihood ratios (see sections below).

As pointed out above, in clinical practice the selection of the criterion to be used should depend on the ultimate goal of the diagnostic test. For instance, in the case of CAD, if one wished to use leukocyte elastase determination as a screening test prior to performing a coronary angiography for detecting CAD, one would seek high sensitivity so as to be able to identify all the diseased patients. Thus, using the "MinValueSe" criterion with an $Se \geq 0.95$, the cutpoint obtained would be $22 \mu\text{gl}^{-1}$, and so coronary angiography would be performed on any patient having a leukocyte elastase level $\geq 22 \mu\text{gl}^{-1}$. Using this optimal cutpoint, 96% of CAD patients would be correctly classified, whereas only 38% of patients without CAD would be correctly identified (28 false positive classifications). If, however, one were seeking an equilibrium between sensitivity and specificity, by, e.g., the "SpEqualSe" method, an optimal cutpoint of $38 \mu\text{gl}^{-1}$ would be obtained, with which 68% of patients with CAD and 67% of patients without CAD would be correctly detected. This value is very close to that obtained using the Youden index (the "Youden" method), which would afford a value of 37, similarly seeking an equilibrium between the two measures of sensitivity and specificity ($Se = 0.69$ and $Sp = 0.67$). It should be noted that, in some diseases which are incurable or rapidly lethal, interest may lie in selecting the greatest specificity possible, e.g., $Sp \geq 0.95$ ("MinValueSp" method). Applied to our CAD example, a value of $54 \mu\text{gl}^{-1}$ would be obtained, a value that, as will be readily appreciated, is very much higher than the previous ones. In our example, however, this approach would not be appropriate, in view of the fact that angioplasty, with or without a stent, is usually a successful treatment in CAD.

Criteria based on predictive values

Despite the fact that Se and Sp are considered the fundamental operating characteristics of a diagnostic test, in practice their capacity for quantifying medical uncertainty is limited. A clinician (or observer of a diagnostic test) is sometimes interested in knowing what the probability is of an individual who has tested positive actually proving to be diseased (i.e., the positive predictive value, PPV), and vice-versa, i.e., the probability of an individual who has tested negative being disease-free (the negative predictive value, NPV). These two quantities can be expressed, in terms of the Se and Sp , as:

$$PPV(c) = P(D = 1 | T_+) = \frac{pSe(c)}{pSe(c) + (1-p)(1-Sp(c))},$$

$$NPV(c) = P(D = 0 | T_-) = \frac{(1-p)Sp(c)}{(1-p)Sp(c) + p(1-Se(c))},$$

where p denotes the disease prevalence, i.e., $p = P(D = 1)$. As with Se and Sp , in the case of PPV and NPV measures, there are a number of strategies for selecting an optimal cutpoint (Vermont *et al.* 1991), such as setting minimal values selected previously for a given predictive value or for both ("MinValuePPV", "MinValueNPV" and "MinValueNPVPPV" methods in the package), setting a single target value for one of the predictive values ("ValueNPV" and "ValuePPV" in **OptimalCutpoints**), selecting the point at which the predictive values are practically the same, "NPVEqualPPV" method (Vermont *et al.* 1991; Gallop *et al.* 2003) or using the criterion of the point on the predictive ROC (PROC) curve closest to the point (0,1) ("PROC01" method). Similar to the ROC curve, the PROC is defined as the plot of $(1 - NPV(c); PPV(c))$ for all possible cutpoints c (Vermont *et al.* 1991; Gallop *et al.* 2003).

From an applied point of view, it is usual to seek elevated positive predictive values in any case where treating false positives may have serious consequences, be these psychological, physical or economic (e.g., chemotherapy in cancer or AIDS). Taking the CAD example, in view of the fact that 1) coronary disease is potentially curable (there is a treatment), 2) a false positive does not produce serious disorders for the patient, and 3) coronary angiography enjoys good results with low risk, one would seek elevated negative predictive values. This is related to the ability to rule out the disease with a greater degree of certainty. For the purpose, one could, for instance, use the "MinValueNPV" criterion. Thus, for a $NPV \geq 0.95$, the cutpoint obtained would be $13 \mu\text{gl}^{-1}$, with a $PPV = 0.72$ and an $NPV = 1$ (the maximum value). This means that all patients with elastase below $13 \mu\text{gl}^{-1}$ are identified as patients without CAD and can be correctly classified as healthy (i.e., there are no false negative results).

Criteria based on diagnostic likelihood ratios

Where the aim of the diagnostic test is predictive, cutpoints based on the DLR may be more useful (Boyko 1994). $DLRs$ provide a summary of how many times patients with a disease are more (or less) likely to have a particular result than patients without the disease. Specifically, the positive and negative diagnostic likelihood ratios (DLR_+ and DLR_- respectively) are defined as:

$$DLR_+(c) = \frac{P(T_+ | D = 1)}{P(T_+ | D = 0)} = \frac{Se(c)}{1 - Sp(c)},$$

$$DLR_-(c) = \frac{P(T_- | D = 1)}{P(T_- | D = 0)} = \frac{1 - Se(c)}{Sp(c)}.$$

Optimal-cutpoint selection criteria based on these measures have also been proposed, with pre-established values being defined in the same way as described for the Se , Sp and predictive values measures (Rutter and Miglioretti 2003). These criteria have been denoted as "ValueDLR.Positive" and "ValueDLR.Negative" in the **OptimalCutpoints** package.

In our CAD data, for instance, if one wished to seek a cutpoint above which a positive result doubled the probability of having the disease as opposed to not having it, one would seek a cutpoint that yielded a DLR_+ equal to 2. To this end, one would use the "ValueDLR.Positive" criterion implemented in **OptimalCutpoints**, and would obtain a cutpoint equal to $41 \mu\text{gl}^{-1}$. This means that any patient having a leukocyte elastase value higher than or equal to $41 \mu\text{gl}^{-1}$ would be twice as likely to have than not to have coronary stenosis.

2.2. Methodology based on cost-benefit analysis of the diagnosis

When undertaking a diagnostic procedure, a price is paid (in terms of money and/or risk of possible complications) to gain information that may be beneficial for the subsequent treatment and care of the patient. According to ROC methodology, information so gained on patients' current health or disease status can be measured and described, in a statistical sense, by attempting to respond to the following questions: (1) How can the benefits obtained from correct diagnostic decisions be balanced (offset) against the costs of incorrect decisions?; and (2) How can one judge if the information 'bought' is worth the price 'paid'?

In this case, the benefits and costs of each type of decision are combined with the prevalence of the disease of interest to find the Se and $1 - Sp$ values on the ROC curve which will yield the minimum mean cost (maximum mean benefit) in a given diagnosis (McNeill, Keeler, and Adelstein 1975; Metz, Starr, Lusted, and Rossmann 1975; Metz 1978; Swets and Swets 1979), where the term 'cost' can be construed as a combination of various aspects and not exclusively as a monetary term (Edwards, Guttentag, and Snapper 1975). The mean cost of the consequences of conducting a diagnostic test should include the price that must be paid for performing the test ('overhead cost' C_0), and the costs of the medical consequences of each type of diagnostic decision, weighted by the probability of these occurring. Hence, for a situation in which there are two possible alternative decisions (though it may easily be extended to situations with a larger number of decisions), the expected cost C of the use of the diagnostic test can be expressed as:

$$C(c) = C_0 + C_{TP} p Se(c) + C_{TN} (1 - p) Sp(c) + C_{FP} (1 - p) (1 - Sp(c)) + C_{FN} p (1 - Se(c)),$$

where C_{TP} , C_{TN} , C_{FP} , C_{FN} represent the mean costs of medical consequences flowing from each type of diagnostic decision. On minimizing the previous expression, the optimal point is that where the slope of the ROC curve ('slope of iso-utility') is given by (Lusted 1968; England 1988; Halpern, Albert, Krieger, Metz, and Maidment 1996) $S = \frac{1 - p}{p} \frac{C_{FP} - C_{TN}}{C_{FN} - C_{TP}}$.

This criterion is included in **OptimalCutpoints** under the name "CB".

The problem of this approach is that it requires the consequences of each possible test result to be quantified, and as a rule, allocating costs to different classifications is complex. In many situations, the value of the cost ratio is determined directly, without knowing the individual values of the four costs that appear in the expression. Without better information, one generally tends to assume that $p = 0.5$, $C_{FP} = C_{FN}$, and $C_{TN} = C_{TP}$, and a cutpoint would

thus be chosen such that $S = 1$. This implies that the prevalence in the study population is around 50%, and that the costs of the FP and FN test results are the same. It is therefore important to stress that this cutpoint might not be optimal for other prevalences and cost ratios.

Some authors (McNeill *et al.* 1975; Zweig and Campbell 1993; Burgueño, García-Bastos, and González-Buitrago 1995) talk only of the cost ratio of an FP to an FN result because the costs of true decisions are assumed to be null. This is why another criterion for optimal-cutpoint selection has been proposed, based on minimization of a term that measures the cost of incorrect classifications ("MCT", *Misclassification-cost term*, in the package), (Smith 1991; Greiner 1995, 1996):

$$MCT(c) = \frac{C_{FN}}{C_{FP}}p(1 - Se(c)) + (1 - p)(1 - Sp(c)).$$

Returning to the CAD example, if one assumed that, despite being a severe disease, coronary stenosis is usually treated successfully with minimal risk for the patient, clinically speaking it would make more sense to consider that the false negative results have a higher 'cost' than do the false positive results. If one noted, say, that an FN had triple the cost of an FP , the "MCT" method with a ratio of $C_{FN}/C_{FP} = 3$ could be used. One would thus obtain an optimal value of $21 \mu\text{gl}^{-1}$, so that patients with elastase higher than or equal to $21 \mu\text{gl}^{-1}$ would be classified as patients who present with CAD, so minimizing false negative classifications.

Finally, note that some criteria, such as the Youden index, the generalized Youden index and efficiency (Section 2.1), can also be viewed from a cost-benefit standpoint. The slope of the ROC curve at the optimal cutpoint obtained by means of the Youden index is equal to 1 (Perkins and Schisterman 2006); at the optimal cutpoint calculated on the basis of the generalized Youden index, the slope is calculated by solely considering the costs deriving from false positive and false negative decisions, $S = \frac{1-p}{p} \frac{C_{FP}}{C_{FN}}$; and lastly, at the cutpoint that maximizes the proportion of cases correctly classified, the slope is calculated solely by reference to the prevalence, $S = \frac{1-p}{p}$.

2.3. Maximum χ^2 or minimum p value criterion

Another approach for selecting the optimal cutpoint consists of maximizing a statistical test which represents the association between this marker and the binary result obtained on using the cut value (Mazumdar and Glassman 2000). The pertinent χ^2 test is calculated for each of the observed diagnostic marker values (candidates for the optimal cutpoint) -except for a proportion of the extreme values- with the point for which the maximum χ^2 or, equivalently, the corresponding minimum p value is obtained, being selected as the optimal value (Miller and Siegmund 1982; Mazumdar and Glassman 2000) ("MinPvalue" method in **OptimalCutpoints**). A number of correction methods have, moreover, been proposed for adjusting for the increase in the type-I error which is associated with the minimum p value approach, such as the maximally selected rank statistical method (Schulgen, Lausen, Olsen, and Schumacher 1994; Lausen and Schumacher 1996) or the use of a permutation test approach (Hilsenbeck, Clark, and McGuire 1992). The former is an easily applicable method but has the drawback of being too conservative in cases where there are few cutpoints.

2.4. Prevalence-based methods

Finally, strategies have also been proposed for the selection of optimal prevalence-based cut-points, designed mainly for situations in which the marker assumes values from 0 to 1 (prevalence values), e.g., the probabilities obtained on the basis of a statistical model. Here the following can be considered: (1) *the observed prevalence criterion* ("**ObservedPrev**" method in **OptimalCutpoints**), which simply consists of selecting as optimal the value closest to the observed prevalence; (2) *the mean predicted probability criterion* ("**MeanPrev**" method), in which the closest value to the mean of the diagnostic test values is chosen as optimal, for instance, the mean probability of occurrence based on the results of the model; (3) selection of a cut value for which prevalence predicted on the basis of the model is practically equal to the observed prevalence ("**PrevMatching**" method). Criteria (1) and (3) are useful strategies in cases where preserving prevalence is of crucial interest (Manel, Williams, and Ormerod 2001; Kelly, Dunstan, Lloyd, and Fone 2008).

3. Description of **OptimalCutpoints** package

The previous section outlines several methods proposed in the literature for selecting optimal cutpoints in diagnostic tests. This section introduces **OptimalCutpoints**, an R package in which all these methods have been incorporated in a way designed to be clear and user-friendly for the end-user. **OptimalCutpoints** provides numerical and graphical results. The numerical results include the optimal cutpoint according to the selected criterion, and the associated accuracy measures with their confidence intervals. The program's graphical output shows the ROC and PROC curves of the test analyzed and, where possible, the plot of the pertinent criterion according to the different test values (candidates for the optimal cutpoint). In addition, **OptimalCutpoints** enables optimal levels to be calculated automatically according to levels of certain (categorical) covariates and this will be illustrated in the biomedical example in the next section. This is of great interest because a diagnostic marker's discriminatory capacity can often depend on specific characteristics, such as a patient's age or gender, or the severity of the disease (Pepe 2003). Moreover, no restriction has been imposed with respect to the range of values of the diagnostic test, i.e., it can take some values in a continuous range or a risk score obtained from a predictive diagnostic model (values from 0 to 1). Finally, insofar as computation is concerned, for all methods in this package, the optimal cut value obtained is always one of the observed diagnostic marker values, and the ROC and PROC curves and accuracy measures are empirically estimated.

In R language, programming is based on objects, and computations are basically functions that are specialized in performing specific calculations. Table 1 provides a summary of the main functions in the package.

The main function of the package is the `optimal.cutpoints` function, which uses the selected method(s) to compute the optimal cutpoint with its accuracy measures, and creates an object of class `optimal.cutpoints`. Usage is as follows:

```
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, ...)
```

Function	Description
<code>optimal.cutpoints</code>	Computes the optimal cutpoint with its accuracy measures and, optionally, the pertinent confidence intervals for such measures.
<code>control.cutpoints</code>	Function used to set several parameters that control the optimal-cutpoint computing process.
<code>print</code>	Print method for objects fitted with <code>optimal.cutpoints</code> .
<code>summary</code>	Produces a summary of an <code>optimal.cutpoints</code> object.
<code>plot</code>	Plot method for objects fitted with <code>optimal.cutpoints</code> . Includes the plots of the ROC and PROC curves, indicating the optimal cutpoint on these plots.

Table 1: Summary of functions in the **OptimalCutpoints** package.

The `X` argument is either a character string with the name of the diagnostic test variable or a formula. When `X` is a formula, it must be an object of class `formula`. The right side of `~` must contain the name of the variable that distinguishes healthy from diseased individuals, and the left side of `~` must contain the name of the diagnostic test variable. The `status` argument only applies when the `X` argument contains the name of the diagnostic test variable, and is a character string with the name of the variable that distinguishes healthy from diseased individuals. The `tag.healthy` argument is the value codifying healthy individuals in the `status` variable.

The `methods` argument is a character vector specifying which method/s is/are used for selecting optimal cutpoints. A total of 34 methods have been implemented in **OptimalCutpoints** (see Table 2). Various optimal-cutpoint selection methods can be selected simultaneously.

The `data` argument is a data frame which must, at minimum, contain the following variables: diagnostic marker; disease status (diseased/healthy); and whether adjustment is to be made for any (categorical) covariate of interest, a variable that indicates the levels of this covariate.

The `direction` argument is a character string specifying the direction in which the ROC curve must be computed. By default, individuals with a test value lower than the cutoff are classified as healthy (negative test), whereas patients with a test value greater than (or equal to) the cutoff are classified as diseased (positive test). If this is not the case, however, and the high values are related to health, this argument should be established at `>`.

The `categorical.cov` argument is an optional argument, and is a character string with the name of the categorical covariate according to which optimal cutpoints are to be calculated. By default it is `NULL`, i.e., no categorical covariate is considered in the analysis.

The `pop.prev` argument is the value of the disease's prevalence. By default it is `NULL`, and in such a case, prevalence is estimated on the basis of sample prevalence, taking into account the number of patients in the sample (cross-sectional study). However, the end-user can also specify a given value for prevalence, as, say, in other types of studies (case-control study) where it cannot be estimated on the basis of the sample. Where the `categorical.cov` is not `NULL`, the prevalence value can be specified by a single value if the same prevalence is assumed for the different levels of the covariate, or by a vector having as many components as levels if different values are assumed.

Criterion name	Description
<i>Criteria based on sensitivity and specificity measures</i>	
ValueSe	A value set for sensitivity (Rutter and Miglioretti 2003): the cutpoint c fulfilling the condition $Se(c) = \text{valueSe}$.
ValueSp	A value set for specificity (Rutter and Miglioretti 2003): the cutpoint c fulfilling the condition $Sp(c) = \text{valueSp}$.
SpEqualSe	Sensitivity = specificity: the cutpoint c minimizing $\{ Sp(c) - Se(c) \}$ (Amaro <i>et al.</i> 1995; Greiner 1995; Hosmer and Lemeshow 2000).
MaxSe	Maximizes sensitivity: the cutpoint c maximizing $Se(c)$ (Filella <i>et al.</i> 1995; Hoffman, Clanon, Littenberg, Frank, and Peirce 2000; Álvarez García <i>et al.</i> 2003).
MaxSp	Maximizes specificity: the cutpoint c maximizing $Sp(c)$ (Bortheyri <i>et al.</i> 1994; Hoffman <i>et al.</i> 2000).
MaxSpSe	Maximizes sensitivity and specificity simultaneously: the cutpoint c maximizing $\{\min\{Sp(c), Se(c)\}\}$ (Riddle and Stratford 1999; Gallop <i>et al.</i> 2003).
Youden	Youden index: the cutpoint c maximizing $YI(c) = Se(c) + Sp(c) - 1$ (Youden 1950; Aoki, Misumi, Kimura, Zhao, and Xie 1997; Greiner <i>et al.</i> 2000) or generalized Youden index: the cutpoint c maximizing $GYI(c) = Se(c) + \frac{1-p}{p} \frac{C_{FN}}{C_{FP}} Sp(c) - 1$ (Geisser 1998; Greiner <i>et al.</i> 2000).
MaxProdSpSe	Maximizes the product of sensitivity and specificity: the cutpoint c maximizing $\{Sp(c)Se(c)\}$ (Lewis, Chuai, Nessel, Lichtenstein, Aberra, and Ellenberg 2008).
Minimax	Minimizes the most frequent error (Hand 1987): the cutpoint c minimizing $\{\max\{p(1 - Se(c)), (1 - p)(1 - Sp(c))\}\}$.
MinValueSe	A minimum value set for sensitivity: the cutpoint c fulfilling the condition $Se(c) \geq \text{minValueSe}$ (Schäfer 1989; Vermont <i>et al.</i> 1991; Gallop <i>et al.</i> 2003).
MinValueSp	A minimum value set for specificity: the cutpoint c fulfilling the condition $Sp(c) \geq \text{minValueSp}$ (Schäfer 1989; Vermont <i>et al.</i> 1991; Gallop <i>et al.</i> 2003).
MinValueSpSe	A minimum value set for specificity and sensitivity (Schäfer 1989): the cutpoint c fulfilling the condition $Sp(c) \geq \text{minValueSp}$ and $Se(c) \geq \text{minValueSe}$.
ROC01	Minimizes distance between ROC plot and point (0,1): the cutpoint c minimizing $\{(Sp(c) - 1)^2 + (Se(c) - 1)^2\}$ (Metz 1978; Vermont <i>et al.</i> 1991).
MaxDOR	Maximizes Diagnostic Odds Ratio: the cutpoint c maximizing $DOR(c) = \frac{Se(c)}{1 - Se(c)} \frac{Sp(c)}{1 - Sp(c)}$ (Kraemer 1992; Böhning <i>et al.</i> 2011).

Continued on next page

Table 2 – continued from previous page

Criterion name	Description
MaxEfficiency	Maximizes efficiency or accuracy: the cutpoint c maximizing $Ef(c) = pSe(c) + (1 - p)Sp(c)$ (Feinstein 1975; Galen 1986; Greiner 1995, 1996).
MaxKappa	Maximizes Kappa index (Cohen 1960; Greiner <i>et al.</i> 2000) or Weighted Kappa index (Kraemer 1992; Kraemer, Periyakoil, and Noda 2002).
<i>Criteria based on predictive values</i>	
ValueNPV	A value set for negative predictive value: the cutpoint c fulfilling the condition $NPV(c) = \text{valueNPV}$.
ValuePPV	A value set for positive predictive value: the cutpoint c fulfilling the condition $PPV(c) = \text{valuePPV}$.
NPVEqualPPV	negative predictive value = positive predictive value (Vermont <i>et al.</i> 1991): the cutpoint c minimizing $ NPV(c) - PPV(c) $.
MaxNPVPPV	Maximizes negative predictive value and positive predictive value simultaneously: the cutpoint c maximizing $\{\min\{NPV(c), PPV(c)\}\}$.
MaxSumNPVPPV	Maximizes the sum of negative predictive value and positive predictive value: the cutpoint c maximizing $\{NPV(c) + PPV(c)\}$.
MaxProdNPVPPV	Maximizes the product of negative predictive value and positive predictive value: the cutpoint c maximizing $\{NPV(c)PPV(c)\}$.
MinValueNPV	A minimum value set for negative predictive value (Vermont <i>et al.</i> 1991): the cutpoint c fulfilling the condition $NPV(c) \geq \text{minValueNPV}$.
MinValuePPV	A minimum value set for positive predictive value (Vermont <i>et al.</i> 1991): the cutpoint c fulfilling the condition $PPV(c) \geq \text{minValuePPV}$.
MinValueNPVPPV	A minimum value set for predictive values (Vermont <i>et al.</i> 1991): the cutpoint c fulfilling the condition $NPV(c) \geq \text{minValueNPV}$ and $PPV(c) \geq \text{minValuePPV}$.
PROC01	Minimizes distance between PROC plot and point $(0, 1)$ (Vermont <i>et al.</i> 1991; Gallop <i>et al.</i> 2003): the cutpoint c minimizing $\{(NPV(c) - 1)^2 + (PPV(c) - 1)^2\}$.
<i>Criteria based on diagnostic likelihood ratios</i>	
ValueDLR.Negative	A value set for negative diagnostic likelihood ratio: the cutpoint c fulfilling the condition $DLR_-(c) = \text{valueDLR-}$ (Boyko 1994; Rutter and Miglioretti 2003).
ValueDLR.Positive	A value set for positive diagnostic likelihood ratio: the cutpoint c fulfilling the condition $DLR_+(c) = \text{valueDLR+}$ (Boyko 1994; Rutter and Miglioretti 2003).

Continued on next page

Table 2 – continued from previous page

Criterion name	Description
<i>Criteria based on cost-benefit analysis of the diagnosis</i>	
CB	Cost-benefit method, computing slope of ROC curve at optimal cutpoint, as $S = \frac{1-p}{p} \frac{C_{FP} - C_{TN}}{C_{FN} - C_{TP}}$ (McNeill <i>et al.</i> 1975; Metz <i>et al.</i> 1975; Metz 1978).
MCT	Minimizes Misclassification Cost Term: the cutpoint c minimizing $MCT(c) = \frac{C_{FN}}{C_{FP}} p(1 - Se(c)) + (1-p)(1 - Sp(c))$ (Smith 1991; Greiner 1995, 1996).
<i>Maximum χ^2 or minimum p value criterion</i>	
MinPvalue	Minimizes p value associated with the statistical χ^2 test which measures the association between the marker and the binary result obtained on using the cutpoint (Miller and Siegmund 1982; Altman <i>et al.</i> 1994).
<i>Prevalence-based methods</i>	
MeanPrev	The closest value to the mean of the diagnostic test values. This criterion is usually used in cases where the diagnostic test takes values in the interval $(0, 1)$, i.e., the mean probability of occurrence, e.g., based on the results of a statistical model (Manel <i>et al.</i> 2001; Kelly <i>et al.</i> 2008).
ObservedPrev	The closest value to the observed prevalence: the cutpoint c minimizing $ c - p $, with p being prevalence estimated from the sample. This criterion is thus indicated/valid in cases where the diagnostic test takes values in the interval $(0, 1)$ (Manel <i>et al.</i> 2001).
PrevalenceMatching	The value for which predicted prevalence is practically equal to observed prevalence: the cutpoint c minimizing $\{ p(1 - Se(c)) - (1-p)(1 - Sp(c)) \}$. This criterion is usually used in cases where the diagnostic test takes values in the interval $(0, 1)$, i.e., the predicted probability, e.g., based on a statistical model (Manel <i>et al.</i> 2001; Kelly <i>et al.</i> 2008).

Table 2: Available methods in the **OptimalCutpoints** package.

The `control` argument indicates the output of the `control.cutpoints` function, which controls the whole optimal-cutpoint calculation process. This function will be explained in detail in the following subsection.

The `ci.fit` argument is a logical value, and if it is `TRUE` then inference is performed on the accuracy measures at the optimal cutpoint (by default it is `FALSE`). Finally, `conf.level` is the value of the confidence level $(1 - \alpha)$, and by default is equal to 0.95.

Summarizing, the `X`, `status`, `tag.healthy`, `methods` and `data` arguments of the `optimal.cutpoints` function are essential arguments, and if one or more is not introduced,

this will lead to error. The remaining arguments (`direction`, `categorical.cov`, `pop.prev`, `control`, `ci.fit`, `conf.level` and `trace`) are optional and, in the event of having no value, will operate with the values established by default.

3.1. Controlling the optimal-cutpoint computation/selection process

It should be noted that there are arguments that are specific to each method. We decided to include all of these in the `control` argument; `control` is a list of control values for the selection process designed to replace the default values yielded by the `control.cutpoints` function. The arguments of the `control.cutpoints` function, as well as the methods for which they apply, are shown in Table 3.

The values of the costs in general (necessary in criteria which make use of cost/benefit-based methodology, i.e., "CB", "MCT", "Youden" and "MaxEfficiency"), and the costs ratio and costs of incorrect classifications in particular, must be indicated in the `costs.ratio`, `CFP` and `CFN` arguments, respectively. By default, the value 1 is established for all, a situation equivalent to classification costs not being considered. The values established by default for the accuracy measures (necessary in the "MinValueSp", "ValueSp", "MinValueSe", "ValueSe", "MinValueSpSe", "MinValueNPV", "ValueNPV", "MinValuePPV", "ValuePPV", "MinValueNPVPPV", "ValueDLR.Positive" and "ValueDLR.Negative" methods) are indicated in the `valueSp`, `valueSe`, `valueNPV`, `valuePPV`, `valueDLR.Positive` and `valueDLR.Negative` arguments, respectively. By default, a value of 0.85 appears for sensitivity, specificity and predictive values measures, a value of 2 for the positive likelihood ratio, and a value of 0.5 for the negative likelihood ratio. These values were set on the basis of values usually indicated in the literature but end-users will have to set them in line with their own goals.

The `adjusted.pvalue` argument of the `control.cutpoints` function should be used in the "MinPvalue" method to indicate whether the Miller and Siegmund method ("PADJMS" option) or Altman method ("PALT5" and "PALT10" options) is selected for adjusting the p value (Miller and Siegmund 1982; Altman *et al.* 1994). The default is "PADJMS". The first method uses the minimum p value (p_{min}) observed and the proportion (ϵ) of sample data which is below the lowest (ϵ_{low}) (or above the highest ϵ_{high}) cutpoint considered:

$$p_{acor} = \phi(z) \left(z - \frac{1}{z} \right) \log \left(\frac{\epsilon_{high}(1 - \epsilon_{low})}{(1 - \epsilon_{high})\epsilon_{low}} \right) + 4 \frac{\phi(z)}{z},$$

where z is the $(1 - p_{min}/2)$ quantile of the standard normal distribution, and ϕ denotes the density function of the standard Normal. Altman *et al.* (1994) furnished the following simplifications of the above formula that work well for low minimum p values ($0.0001 < p_{min} < 0.1$) and are easily applicable: For $\epsilon = \epsilon_{low} = \epsilon_{high} = 5\%$: $p_{alt5} = -3.13p_{min}(1 + 1.65 \ln(p_{min}))$ and for $\epsilon = \epsilon_{low} = \epsilon_{high} = 10\%$: $p_{alt10} = -1.63p_{min}(1 + 2.35 \ln(p_{min}))$

Various approaches are considered in **OptimalCutpoints** for calculating the confidence intervals of the accuracy measures. The `ci.SeSp`, `ci.PV` and `ci.DLR` arguments in the `control.cutpoints` function indicate the methods selected for computing confidence intervals for Se/Sp , PPV/NPV and DLR_+/DLR_- , respectively. They are meaningful only when the argument `ci.fit` is TRUE.

Argument	Method	Description
CFP	MCT Youden MaxKappa	A numerical value specifying the cost of a false positive decision C_{FP} . The default value is 1.
CFN	MCT Youden MaxKappa	A numerical value specifying the cost of a false negative decision C_{FN} . The default value is 1.
costs.ratio	CB	A numerical value specifying the costs ratio $CR = \frac{C_{FP} - C_{TN}}{C_{FN} - C_{TP}}$. The default value is 1.
costs.benefits. Youden	Youden	A logical value. If TRUE, the optimal cutpoint based on cost-benefit methodology is computed. The default is FALSE.
costs.benefits. Efficiency	MaxEfficiency	A logical value. If TRUE, the optimal cutpoint based on cost-benefit methodology is computed. The default is FALSE.
generalized.Youden	Youden	A logical value. If TRUE, the generalized Youden index is computed. The default is FALSE.
weighted.Kappa	MaxKappa	A logical value. If TRUE, the weighted kappa index is computed. The default is FALSE.
valueSe	MinValueSe ValueSe MinValueSpSe	A numerical value specifying the (minimum or specific) value set for sensitivity. The default value is 0.85.
valueSp	MinValueSp ValueSp MinValueSpSe	A numerical value specifying the (minimum or specific) value set for specificity. The default value is 0.85.
valueNPV	MinValueNPV ValueNPV MinValueNPVPPV	A numerical value specifying the minimum value set for negative predictive value. The default value is 0.85.
valuePPV	MinValuePPV ValuePPV MinValueNPVPPV	A numerical value specifying the minimum value set for positive predictive value. The default value is 0.85.
valueDLR.Positive	ValueDLR.Positive	A numerical value specifying the value set for the positive diagnostic likelihood ratio. The default value is 2.
valueDLR.Negative	ValueDLR.Negative	A numerical value specifying the value set for the negative diagnostic likelihood ratio. The default value is 0.5.

Continued on next page

Table 3 – continued from previous page

Argument	Method	Description
<code>maxSp</code>	<code>MinValueSpSe</code>	A logical value meaningful only in a case where there is more than one cutpoint fulfilling the conditions. If <code>TRUE</code> , those of the cutpoints which yield maximum specificity are computed. Otherwise, the cutoff that yields maximum sensitivity is computed. The default is <code>TRUE</code> .
<code>maxNPV</code>	<code>MinValueNPVPPV</code>	A logical value meaningful only in the case where there is more than one cutpoint fulfilling the conditions. If <code>TRUE</code> , those of the cutpoints which yield the maximum negative predictive value are computed. Otherwise, the cutoff that yields the maximum positive predictive value is computed. The default is <code>TRUE</code> .
<code>ci.SeSp</code>	All methods	A character string meaningful only when the argument <code>ci.fit</code> of the <code>optimal.cutpoints</code> function is <code>TRUE</code> . It indicates the method for estimating the confidence interval for sensitivity and specificity measures. Options are "Exact" (Clopper and Pearson 1934), "Quadratic" (Fleiss 1981), "Wald" (Wald and Wolfowitz 1939), "AgrestiCoull" (Agresti and Coull 1998) and "RubinSchenker" (Rubin and Schenker 1987). The default is "Exact".
<code>ci.PV</code>	All methods	A character string meaningful only when the argument <code>ci.fit</code> of the <code>optimal.cutpoints</code> function is <code>TRUE</code> . It indicates the method for estimating the confidence interval for predictive values. Options are "Exact" (Clopper and Pearson 1934), "Quadratic" (Fleiss 1981), "Wald" (Wald and Wolfowitz 1939), "AgrestiCoull" (Agresti and Coull 1998), "RubinSchenker" (Rubin and Schenker 1987), "Transformed" (Simel, Samsa, and Matchar 1991), "NotTransformed" (Koopman 1984) and "GartNam" (Gart and Nam 1998). The default is "Exact".

Continued on next page

Table 3 – continued from previous page

Argument	Method	Description
<code>ci.DLR</code>	All methods	A character string meaningful only when the argument <code>ci.fit</code> of the <code>optimal.cutpoints</code> function is <code>TRUE</code> . It indicates the method for estimating the confidence interval for diagnostic likelihood ratios. Options are "Transformed" (Simel <i>et al.</i> 1991), "NotTransformed" (Koopman 1984) and "GartNam" (Gart and Nam 1998). The default is "Transformed".
<code>adjusted.pvalue</code>	<code>MinPvalue</code>	A character string specifying the method for adjusting the p value, i.e., "PADJMS" for the Miller and Siegmund method (Miller and Siegmund 1982), and "PALT5", "PALT10" for the Altman method (Altman <i>et al.</i> 1994). The default is "PADJMS".
<code>standard.deviation.accuracy</code>	<code>MaxEfficiency</code>	A logical value. If <code>TRUE</code> , standard deviation associated with accuracy at the optimal cut-point is computed. The default is <code>FALSE</code> .

Table 3: Summary of arguments of the `control.cutpoints` function.

In the `ci.SeSp` argument, the options are "Exact", "Quadratic", "Wald", "AgrestiCoull" and "RubinSchenker". "Exact" is the exact confidence interval based on the exact distribution of a proportion (Clopper and Pearson 1934). It should be noted that this method cannot be applied for proportions where the numerator or the difference between the denominator and the numerator is equal to zero. If this occurs for any value of the corresponding accuracy measure, i.e., the sensitivity or the specificity, the program shows a warning message and returns a `NaN` for the limit of the confidence interval that could not be computed. It is worth noting, however, that this problem only happens for values of sensitivity/specificity equal to zero or one, i.e., on values which are not of interest in clinical practice. "Quadratic" refers to Fleiss' quadratic confidence interval (Fleiss 1981), based on the asymptotic normality of the estimator of a proportion but adding a continuity correction, and this approach is valid in a situation where both the numerator and the difference between the denominator and the numerator of the proportion are greater than 5. "Wald" indicates Wald's confidence interval (Wald and Wolfowitz 1939) with continuity correction, based on maximum-likelihood estimation of a proportion, and adding a continuity correction; it is valid where the numerator and the difference between the denominator and numerator are greater than 20. Similarly to the "Exact" method, when "Quadratic" or "Wald" approaches are not valid for any value of the corresponding accuracy measure, the program shows a warning message. However, in these cases the confidence intervals are computed. We therefore recommend the user to check the conditions under which these methods are valid at the optimal cutpoint. The "AgrestiCoull" option computes the confidence interval proposed by Agresti and Coull (1998), and is a score confidence interval that does not use the standard calculation for the binomial proportion.

Finally, "RubinSchenker" means Rubin and Schenker's logit confidence interval (1987), and uses logit transformation and bayesian arguments with an a priori Jeffreys distribution. The default is "Exact".

In the `ci.DLR` argument, "Transformed", "NotTransformed" and "GartNam" are the available options. "Transformed" indicates the confidence interval based on the logarithmic transformation of the diagnostic likelihood ratios (Simel *et al.* 1991), "NotTransformed" is the confidence interval without transformation (Koopman 1984), and "GartNam" is the confidence interval based on the calculation of the interval for the ratio of two independent proportions (Gart and Nam 1998). The default is "Transformed". Inference of the predictive values depends on the type of study, i.e., whether cross-sectional (prevalence can be estimated on the basis of the sample) or case-control. In the former case, the approaches for calculating the confidence intervals of the predictive values are the same as for the sensitivity and specificity measures. Accordingly, in such a case, the possible options for the `ci.PV` argument are "Exact", "Quadratic", "Wald", "AgrestiCoull" and "RubinSchenker". In a case-control study, however, the confidence intervals of the predictive values should be based on the intervals of the likelihood ratios, so that the available options are "Transformed", "NotTransformed" and "GartNam". The default is "Exact".

For greater detail, the help manual of the `optimal.cutpoints` and `control.cutpoints` functions can be consulted. In addition, the following section gives an illustration of the use of these two functions, based on the real biomedical CAD example.

3.2. Summaries: numerical and graphical output

Numerical and graphical summaries of the created object can be obtained by using the `summary`, `print` and `plot` methods.

Numerical results are printed on the screen, and the output yielded by the `summary` method always includes: the matched call to the main function `optimal.cutpoints`; the AUC value with its confidence interval (DeLong, DeLong, and Clarke-Pearson 1988); the method used for selecting the optimal value together with the number of optimal cutpoints (in some cases there may be more than one value); the optimal cutoff(s) and its/their accuracy-measure estimates (*Se*, *Sp*, etc.); the number of false positive and false negative classifications; and, where possible, the value of the optimal criterion. Furthermore, accuracy measures will be accompanied by their confidence levels, if the `se.fit` argument is TRUE. All this information will be shown for each level of categorical covariate, if specified. Graphical output shows the empirical ROC and PROC curves and, where possible, the plot of the chosen criterion versus all the different test values (candidates for the optimal cutpoint).

3.3. Technical features

In this subsection, certain specific characteristics of some methods and the behavior of the package in such cases are briefly explained. The methods in which a minimum value is set for sensitivity, specificity or the predictive values (the "MinValueSe", "MinValueSp", "MinValuePPV" and "MinValueNPV" methods, respectively), can take several or even zero values. In the latter case, an error message is shown and the user can enter a new minimum value, if desired. In a case where there is more than one cutpoint fulfilling the condition, that which maximizes the other measure is chosen as the optimal cutpoint(s). For example, in the "MinValueSp" method, if there is more than one cutpoint with $Sp \geq \text{minValueSp}$, that

which yields the maximum sensitivity is chosen. So, the cutpoint(s) that achieves the highest sensitivity and specificity under the condition $Sp \geq \text{minValueSp}$ are finally chosen.

The same behavior has been used for the methods that set minimum values for both the sensitivity and specificity measures ("**MinValueSpSe**" method) or for both predictive values ("**MinValueNPVPPV**" method). The only difference is that if there is more than one cutpoint fulfilling these conditions, those which yield maximum sensitivity or maximum specificity (in "**MinValueSpSe**") or maximum predictive positive value or negative predictive value (in "**MinValueNPVPPV**") are chosen. The user can select one of the two options by means of the **maxSp** and **maxNPV** arguments in the **control.cutpoints** function, respectively (see Table 3). If **TRUE** (the default value), the cutpoint/s yielding maximum specificity or maximum negative predictive value is/are computed as the optimal cutpoint(s).

Finally, it should be noted that there are several criteria proposed in the literature that provide the same optimal value. For instance, the "**Youden**" method is identical (from an optimization point of view) to the method that maximizes the sum of sensitivity and specificity (Albert and Harris 1987; Zweig and Campbell 1993) and to the criterion that maximizes concordance, which is a function of the AUC defined as $Se + Sp - 0.5$ (Begg, Cramer, Venkatraman, and Rosai 2000; Gönen and Sima 2013). Similarly, "**MaxProdSpSe**" is the same as the method which maximizes the accuracy area just defined as the product of sensitivity and specificity (Lewis *et al.* 2008). Moreover, the method that maximizes efficiency or accuracy ("**MaxEfficiency**" method in **OptimalCutpoints**) provides the same optimal cutpoint as the method that minimizes the classification error rate (Metz 1978).

4. Practical application of the **OptimalCutpoints** package

This section describes the application of the **OptimalCutpoints** R package. As mentioned in the Introduction, to illustrate the use of this package, we shall consider the study that sought to investigate the clinical usefulness of leukocyte elastase for diagnosis of CAD (Amaro *et al.* 1995). Usefulness refers to the practical value of information when it comes to managing patients. The main research question here is to select optimal cutpoints for elastase concentrations at the date of diagnosing patients with CAD. Depending on a predetermined elastase concentration cutpoint, subjects are chosen for coronary angiography. Since it is well established that elastase concentrations behave differently according to gender, the analyses were performed separately for males and females.

The first step consists of loading the **OptimalCutpoints** package and the data set (included in the package) in R:

```
R> library("OptimalCutpoints")
R> data("elas")
```

To view summary statistics of the variables included in the data set:

```
R> summary(elas)
```

	elas	status	gender
Min.	: 5.0	Min. :0.000	Female: 37
1st Qu.:	27.0	1st Qu.:0.000	Male :104

```

Median : 39.0   Median :1.000
Mean   : 43.3   Mean    :0.681
3rd Qu.: 51.0   3rd Qu.:1.000
Max.   :163.0   Max.    :1.000

```

To compute the optimal cutpoint using the *elas* data set, simply use the syntax shown below:

```

R> cutpoint1 <- optimal.cutpoints(X = "elas", status = "status",
+   tag.healthy = 0, methods = c("Youden", "SpEqualSe"), data = elas,
+   categorical.cov = "gender", pop.prev = NULL,
+   control = control.cutpoints(), ci.fit = TRUE)

```

In this case, by way of example, two methods are considered for calculating the optimal cutpoint, namely, the method based on the Youden index, and the sensitivity-specificity equality criterion, since these are two of the best-known and most widely used methods in practice (Youden 1950; Greiner 1995; Aoki *et al.* 1997; Shapiro 1999; Greiner *et al.* 2000). If no adjustment is made for any categorical covariate, one is left with the default value of the `categorical.cov` argument: `categorical.cov = NULL`, and so there is no need for it to be indicated. As the intention here, however, is to perform separate analyses for males and females, the `categorical.cov = "gender"` must be indicated. As this example involves a cross-sectional study, disease prevalence is estimated on the basis of sample prevalence, and so the default value, `pop.prev = NULL`. In another type of study, end-users could indicate a given value for the population prevalence. Finally, as the argument `ci.fit` is `TRUE`, the confidence intervals for the accuracy measures are computed. By default, the "Exact" method is used for the sensitivity, the specificity and the predictive values. As pointed out before, this method cannot be applied in those situations where the *TPs*, the *FPs*, the *TNs* or the *FNs* are equal to zero. In these cases, the program produces a warning, and returns a `NaN` for the limit of the confidence interval that could not be computed. Specifically, in this example eight warning messages appear: four warnings for females (for the sensitivity, the specificity, and the positive and negative predictive values), and four similar warnings for males.

The `cutpoint1` object is a list that consists of the following components:

```

R> names(cutpoint1)

[1] "Youden"      "SpEqualSe"   "methods"     "levels.cat"  "call"
[6] "data"

```

The component "methods" is a character vector with the value of the argument `methods` used in the call; "levels.cat" is a character vector indicating the levels of the categorical covariate; "call" is the matched call; and finally, "data" is the data frame used in the analysis. The first two components ("Youden" and "SpEqualSe") contain the results associated with each of the methods selected for computing the optimal cutpoint. In this case, each of these components is itself a two-component list (for "Male" and "Female") containing:

```

R> names(cutpoint1$Youden$Male)

[1] "measures.acc"      "optimal.cutoff"   "criterion"
[4] "optimal.criterion"

```


Each of the previous components contains the following information:

1. "measures.acc": a list with all cutoffs, their accuracy measures (Se , Sp , PPV , NPV , DLR_+ and DLR_-), the AUC, and prevalence and sample size in healthy and diseased populations:

```
R> names(cutpoint1$Youden$Male$measures.acc)
```

```
[1] "cutoffs"      "Se"           "Sp"           "PPV"         "NPV"
[6] "DLR.Positive" "DLR.Negative" "AUC"         "pop.prev"    "n"
```

2. "optimal.cutoff": a list with the optimal cutoff(s), its/their accuracy measures (Se , Sp , PPV , NPV , DLR_+ and DLR_-), and the number of false positive and false negative decisions:

```
R> names(cutpoint1$Youden$Male$optimal.cutoff)
```

```
[1] "cutoff"      "Se"           "Sp"           "PPV"         "NPV"
[6] "DLR.Positive" "DLR.Negative" "FP"          "FN"
```

3. "criterion": the numerical value of the method considered for selecting the optimal cutpoint for each cutoff:

```
R> cutpoint1$Youden$Male$criterion
```

```
      1      2      3      4      5      6      7
0.000000 0.0434783 0.0869565 0.1057434 0.0933977 0.1368760 0.1803543
      8      9     10     11     12     13     14
0.2238325 0.2549651 0.2302738 0.2179281 0.1932367 0.1438540 0.2249061
```

```
[...]
```

```
     50     51     52     53     54     55     56
0.0864198 0.0740741 0.0617284 0.0493827 0.0370370 0.0246914 0.0123457
```

4. "optimal.criterion": the numerical value of the criterion at the optimal cutpoint:

```
R> cutpoint1$Youden$Male$optimal.criterion
```

```
[1] 0.393451
```

Accordingly, end-users can easily access each of these components, e.g., to see the sensitivity values for all cut values in males:

```
R> cutpoint1$Youden$Male$measures.acc$Se
```

or

```
R> cutpoint1$SpEqualSe$Male$measures.acc$Se
```

4.1. Numerical results

A numerical summary of the results can be obtained by calling up the `print` or `summary` methods:

```
R> summary(cutpoint1)
```

Call:

```
optimal.cutpoints.default(X = "elas", status = "status", tag.healthy = 0,
  methods = c("Youden", "SpEqualSe"), data = elas,
  categorical.cov = "gender", pop.preval = NULL,
  control = control.cutpoints(), ci.fit = TRUE)
```

```
*****
Female
*****
```

Area under the ROC curve (AUC): 0.818 (0.684, 0.952)

CRITERION: Youden

Number of optimal cutoffs: 1

	Estimate	95% CI lower limit	95% CI upper limit
cutoff	46.0000000	-	-
Se	0.6666667	0.3838037	0.8817589
Sp	0.8181818	0.5971542	0.9481327
PPV	0.7142857	0.4516107	0.9031160
NPV	0.7826087	0.5285570	0.9359958
DLR.Positive	3.6666667	1.4096667	9.5373214
DLR.Negative	0.4074074	0.1939348	0.8558589
FP	4.0000000	-	-
FN	5.0000000	-	-
Optimal criterion	0.4848485	-	-

CRITERION: SpEqualSe

Number of optimal cutoffs: 1

	Estimate	95% CI lower limit	95% CI upper limit
cutoff	41.0000000	-	-
Se	0.7333333	0.4489968	0.9221285
Sp	0.6818181	0.4512756	0.8613535
PPV	0.6111111	0.3762084	0.8712446
NPV	0.7894736	0.5263331	0.9157684
DLR.Positive	2.3047619	1.1634481	4.5656764
DLR.Negative	0.3911111	0.1611869	0.9490094
FP	7.0000000	-	-

```

FN          4.00000000          -          -
Optimal criterion 0.05151515          -          -
    
```

```

*****
Male
*****
    
```

Area under the ROC curve (AUC): 0.722 (0.612, 0.831)

CRITERION: Youden

Number of optimal cutoffs: 1

	Estimate	95% CI lower limit	95% CI upper limit
cutoff	38.0000000	-	-
Se	0.6543210	0.5404147	0.7565737
Sp	0.7391304	0.5159480	0.8977139
PPV	0.8983051	0.7686835	0.9355009
NPV	0.3777778	0.2738718	0.6528595
DLR.Positive	2.5082305	1.2382379	5.0807845
DLR.Negative	0.4676834	0.3180320	0.6877538
FP	6.0000000	-	-
FN	28.0000000	-	-
Optimal criterion	0.3934514	-	-

CRITERION: SpEqualSe

Number of optimal cutoffs: 1

	Estimate	95% CI lower limit	95% CI upper limit
cutoff	36.0000000	-	-
Se	0.6666667	0.5531734	0.7675667
Sp	0.6086956	0.3854190	0.8029236
PPV	0.8571429	0.7075091	0.9083152
NPV	0.3414634	0.2429772	0.5759222
DLR.Positive	1.7037037	1.0003372	2.9016279
DLR.Negative	0.5476190	0.3492857	0.8585711
FP	9.0000000	-	-
FN	27.0000000	-	-
Optimal criterion	0.0579710	-	-

In this case, the `summary` method displays the information relating to the optimal cutpoint, i.e., the methods used for selecting the optimal value here were "Youden" and "SpEqualSe", and the optimal cutpoints as well as their accuracy measures are shown for both methods and for males and females. Optimal criteria refer to the Youden index (i.e., the value $Se + Sp - 1$ at optimal cutpoint) and the $Se - Sp$ difference at optimal cutoff. It must be borne in mind that, as the choice of the cutpoint is made on the basis of the empirical ROC curve, it is not

always possible to obtain the value c of the sample for which $Se(c) = Sp(c)$; here the cutpoint $c / \min_c\{|\widehat{Se}(c) - \widehat{Sp}(c)|\}$ is obtained.

As in the call to the function, `ci.fit = TRUE`, the accuracy measures -as pointed out above- appear accompanied by their corresponding confidence levels. By default, confidence intervals for the AUC and accuracy measures are calculated for a confidence level α of 0.95, though this value can be changed in the `conf.level` argument of the main `optimal.cutpoints` function. Moreover, the exact confidence interval (Clopper and Pearson 1934) is being calculated by default for the sensitivity, specificity and predictive values measures (in this case the inference for these values is correct, since it is a cross-sectional study; if it were a case-control study, however, the method of calculating the confidence interval for the predictive values would have to be changed), and for the interval based on the transformation logarithm for the likelihood ratios (Simel *et al.* 1991). These methods for calculating confidence intervals can be changed by means of the `ci.SeSp`, `ci.PV` and `ci.DLR` arguments, which appear in the `control.cutpoints` function. Hence, if one wanted to change the method of calculating the confidence interval for the sensitivity and specificity measures, one would only have to indicate this using the following syntax:

```
R> cutpoint2 <- optimal.cutpoints(X = "elas", status = "status",
+   tag.healthy = 0, methods = c("Youden", "SpEqualSe"), data = elas,
+   pop.prev = NULL, categorical.cov = "gender",
+   control = control.cutpoints(ci.SeSp = "AgrestiCoull"), ci.fit = TRUE)
R> summary(cutpoint2)
```

In this case, the results would be as follows:

Call:

```
optimal.cutpoints.default(X = "elas", status = "status", tag.healthy = 0,
  methods = c("Youden", "SpEqualSe"), data = elas,
  categorical.cov = "gender", pop.prev = NULL,
  control = control.cutpoints(ci.SeSp = "AgrestiCoull"), ci.fit = TRUE)
```

```
*****
```

```
Female
```

```
*****
```

```
Area under the ROC curve (AUC): 0.818 (0.684, 0.952)
```

```
CRITERION: Youden
```

```
Number of optimal cutoffs: 1
```

	Estimate	95% CI lower limit	95% CI upper limit
cuttoff	46.0000000	-	-
Se	0.6666667	0.4171355	0.8482368
Sp	0.8181818	0.6148339	0.9269312
[...]			

```
CRITERION: SpEqualSe
```

Number of optimal cutoffs: 1

	Estimate	95% CI lower limit	95% CI upper limit
cutoff	41.00000000	-	-
Se	0.73333333	0.4804957	0.8910255
Sp	0.68181818	0.4731860	0.8363941
[...]			

Male

Area under the ROC curve (AUC): 0.722 (0.612, 0.831)

CRITERION: Youden

Number of optimal cutoffs: 1

	Estimate	95% CI lower limit	95% CI upper limit
cutoff	38.00000000	-	-
Se	0.6543210	0.5458938	0.7487735
Sp	0.7391304	0.5353000	0.8745138
[...]			

CRITERION: SpEqualSe

Number of optimal cutoffs: 1

	Estimate	95% CI lower limit	95% CI upper limit
cutoff	36.00000000	-	-
Se	0.66666667	0.5585284	0.7597123
Sp	0.60869565	0.4078552	0.7784238
[...]			

Note that the exploration of the usefulness of medical information involves many factors which, rather than being properties of the test system, are instead properties of the circumstances of the clinical application. With respect to clinical interpretation, the following result was obtained: leukocyte elastase determination was shown to be a test that displays good 'diagnostic accuracy' or ability to discriminate between patients with and without CAD, particularly in women (AUC = 0.82 in women versus AUC = 0.72 in men). The cutpoint obtained using the criterion based on the Youden index for women was $46 \mu\text{g l}^{-1}$. Accordingly, women with an elastase value higher than or equal to 46 were classified as patients with CAD. Using this cutpoint, 82% of women without CAD and 67% of women with CAD were correctly classified (4 false positive and 5 false negative classifications). Furthermore, 78% of women who registered a negative test result (i.e., an elastase value lower than 46) did not really present with CAD, while 71% of women with a positive result did in fact present with the disease. The likelihood of a female having CAD increased by 3.67 in the case of a positive test result and, conversely, decreased by 0.41 in the case of a negative test result.

In the case of men, the cutpoint obtained using the criterion based on the Youden index

was $38 \mu\text{gl}^{-1}$, a value lower than that obtained for women. This means that men with elastase values higher than or equal to $38 \mu\text{gl}^{-1}$ were classified as CAD patients. On the basis of this value, 65% of the men who presented with the disease were correctly classified by determination of elastase (positive value) and 74% of those who did not present with CAD were likewise correctly classified (negative value, elastase lower than $38 \mu\text{gl}^{-1}$). Of the men who registered a positive elastase value, almost 90% really had CAD but among those who registered a negative elastase value, only 37% did not really have the disease. Moreover, the likelihood of a male having CAD was 2.5-fold if the test result was positive (elastase $\geq 38 \mu\text{gl}^{-1}$) and 0.47-fold if the test result was negative.

If, instead of the Youden criterion, one applies the method that selects the cutoff at which sensitivity is equal to specificity, one obtains optimal cutpoints lower than those previously obtained (equal to $41 \mu\text{gl}^{-1}$ in women and $36 \mu\text{gl}^{-1}$ in men), and the same conclusions can be drawn as in the previous case. The only difference is that, when the optimal value falls, the measures of sensitivity (and thus the false positive decisions) and negative predictive value increase, while the measures of specificity (and false negatives), positive predictive value and likelihood ratios all decrease.

The Youden index can also be interpreted from a cost-benefit analysis perspective. The slope of the ROC curve at the optimal cutpoint obtained using this index is equal to 1 (Perkins and Schisterman 2006), which is equivalent to having a prevalence equal to 0.5 and a costs ratio equal to 1. If one wished to calculate the optimal value taking this into account, one would have to specify that the `costs.benefits.Youden` argument of the `control.cutpoints` function was set as `TRUE`:

```
R> cutpoint3 <- optimal.cutpoints(X = "elas", status = "status",
+   tag.healthy = 0, methods = c("Youden", "SpEqualSe"), data = elas,
+   pop.prev = NULL, categorical.cov = "gender",
+   control = control.cutpoints(costs.benefits.Youden = TRUE),
+   ci.fit = TRUE)
```

The Youden index gives equal weight to sensitivity and specificity. Sometimes, however, different weights are suitable (based on the cost of the different types of error and the prevalence of the disease), and in such a case the generalized Youden index can be used. This possibility was also implemented in the **OptimalCutpoints** package, within the "Youden" method. For the purpose, this only has to be indicated in the `generalized.Youden` argument of the `control.cutpoints` function. If `generalized.Youden = TRUE`, the generalized Youden index is computed. In the absence of a value being indicated for costs of incorrect diagnostic decisions or prevalence, $C_{FP} = C_{FN} = 1$ is considered by default (a situation equivalent to having no costs), and prevalence is estimated on the basis of the sample, in this case $p = 0.77885$ in males, and $p = 0.40541$ in females. If the end-user wishes to specify any given costs, these must be indicated in the `CFP` and `CFN` arguments of the `control.cutpoints` function, within the `control` argument in the main function, e.g.:

```
R> cutpoint4 <- optimal.cutpoints(X = "elas", status = "status",
+   tag.healthy = 0, methods = c("Youden", "SpEqualSe"), data = elas,
+   pop.prev = NULL, categorical.cov = "gender", control =
+     control.cutpoints(generalized.Youden = TRUE, CFP = 1, CFN = 3),
+   ci.fit = TRUE)
```


So, in this case, assuming that an *FN* result has triple the cost of an *FP* result, the Youden index would yield some cutpoints that were lower than those obtained without considering misclassification costs, ($25 \mu\text{g l}^{-1}$ in women and $13 \mu\text{g}^{-1}$ in men). Hence, with these optimal values, the presence of false negatives (zero false negative decisions) is avoided, and some maximum values (equal to 1) are attained for sensitivity and also the negative predictive value.

To change the value of the population prevalence, this only has to be directly indicated in the `pop.prev` argument of the `optimal.cutpoints` function. For instance, with a prevalence equal to 0.5 and costs equal to 1, the generalized Youden index is equivalent to the Youden index (the results would be the same):

```
R> cutpoint5 <- optimal.cutpoints(X = "elas", status = "status",
+   tag.healthy = 0, methods = c("Youden", "SpEqualSe"), data = elas,
+   pop.prev = 0.5, categorical.cov = "gender",
+   control = control.cutpoints(generalized.Youden = TRUE),
+   ci.fit = TRUE)
```

It could also be indicated as follows, by using a two-component vector (number of levels of the covariate) to specify the prevalence:

```
R> cutpoint5 <- optimal.cutpoints(X = "elas", status = "status",
+   tag.healthy = 0, methods = c("Youden", "SpEqualSe"), data = elas,
+   pop.prev = c(0.5, 0.5), categorical.cov = "gender",
+   control = control.cutpoints(generalized.Youden = TRUE),
+   ci.fit = TRUE)
```

4.2. Graphical results

The graphical output of the results can be obtained by calling up the `plot` method:

```
R> plot(cutpoint1)
```

By default, the `plot` method depicts the plots of the ROC and PROC curves (`which = c(1, 2)`). However, the plot of the values of the optimal criterion as a function of the cutoffs can, where applicable, be obtained by specifying the argument `which = 3`:

```
R> plot(cutpoint1, which = 3, ylim = c(0, 1))
```

Figures 1 and 2 show the figures that appear as a result of the above calls in females and males, respectively. This is the default output but the end-user can add specific graphic parameters, such as color, legend, etc.

Furthermore, other figures are also possible, e.g., in the method in which sensitivity and specificity are equal, the plot of these measures (jointly) according to the cutpoint may be of interest. This graph can be created on the basis of the pertinent components yielded with the `optimal.cutpoints` function. For instance, for males, the code is as follows:

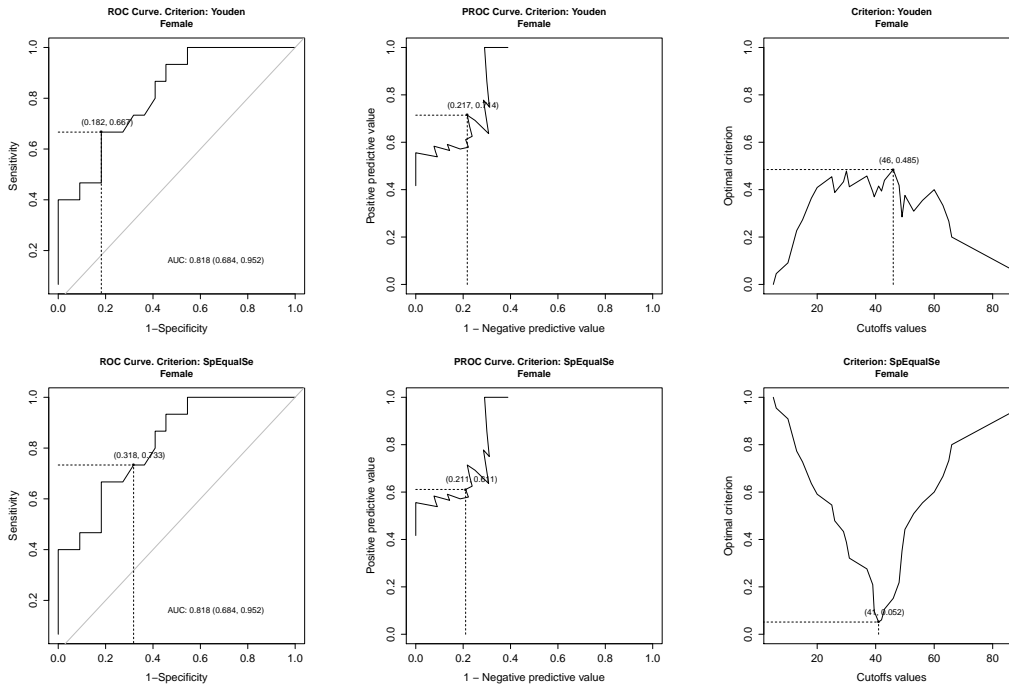


Figure 1: Graphical output in females. From left to right, ROC curve, PROC curve and optimal criterion (according to cutpoints). Top to bottom, the Youden index and sensitivity-specificity equality criteria.

```
R> plot(cutpoint1$SpEqualSe$Male$measures.acc$cutoffs,
+       cutpoint1$SpEqualSe$Male$measures.acc$Se[, 1],
+       xlab = "Cutpoint", ylab = "Sensitivity and Specificity", type = "l",
+       lty = 2, main = "Sensitivity and Specificity \n Males")
R> lines(cutpoint1$SpEqualSe$Male$measures.acc$cutoffs,
+        cutpoint1$SpEqualSe$Male$measures.acc$Sp[, 1], xlab = "Cutpoint",
+        ylab = "Sensitivity and Specificity", type = "l")
R> legend("topright", legend = c("Se", "Sp"), lty = c(2, 1), bty = "n")
```

And similarly for females (see Figure 3).

5. Discussion

The selection of a cutpoint or optimal threshold is useful in continuous diagnostic tests. This develops **OptimalCutpoints**, a user-friendly R package that allows users to choose from among several popular methods in clinical practice. Unlike other packages (Freeman and Moisen 2008; Brasil 2010), **OptimalCutpoints** enables optimal levels to be calculated directly according to levels of given (categorical) covariates. This is of great interest, since discrimination of a biodiagnostic marker may often be different depending on certain characteristics, such as a particular patient's age group or sex (Pepe 2003), and so when it comes to selecting the optimal cutpoint, this must be borne in mind in order to avoid drawing erroneous conclusions. Moreover, some packages only allow for the diagnostic test to take values from

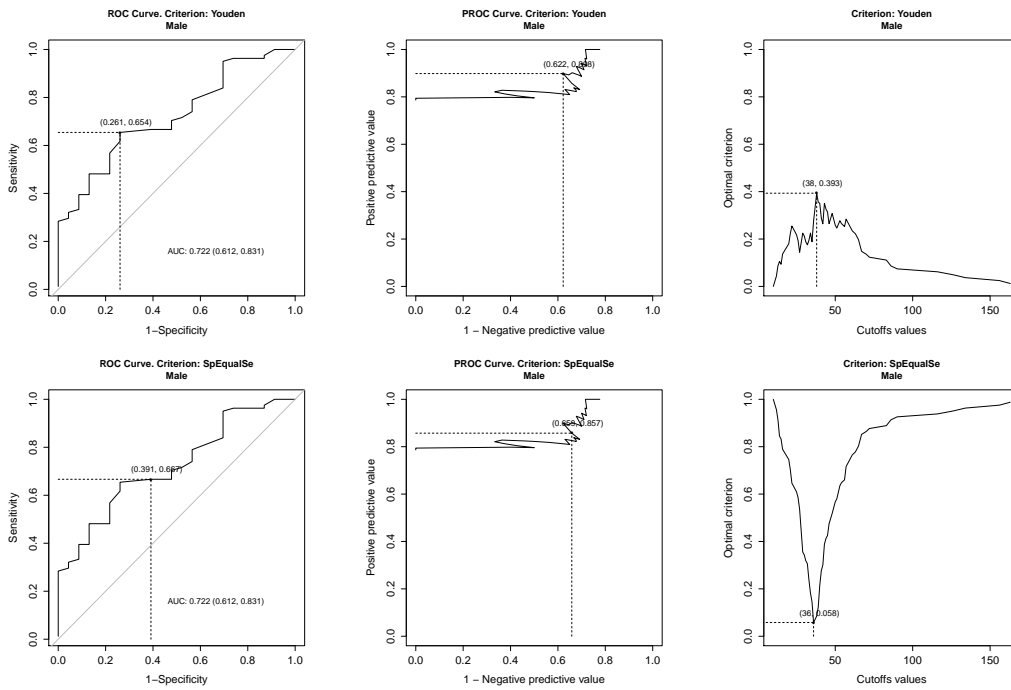


Figure 2: Graphical output in males. From left to right, ROC curve, PROC curve and optimal criterion (according to cutpoints). Top to bottom, the Youden index and sensitivity-specificity equality criteria.

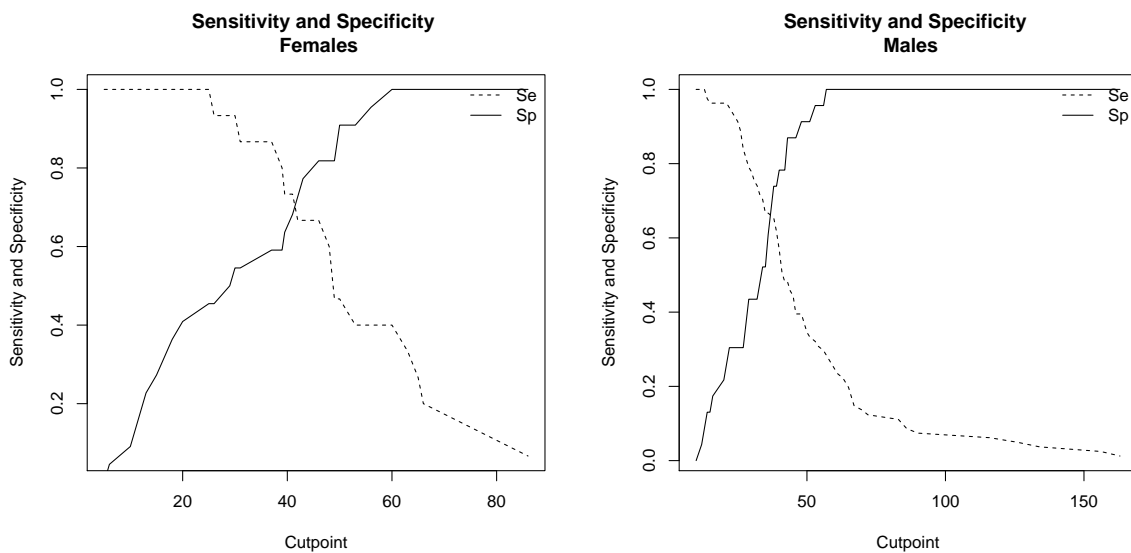


Figure 3: Sensitivity (dotted line) and specificity (solid line) plots according to cutpoints for females (top) and males (bottom).

0 to 1, since they are specifically designed for predictive diagnostic models, but in **Optimal-Cutpoints**, no restriction has been imposed with respect to the range of the values of the

diagnostic test. Thus, it can take some values in a continuous range or a risk score obtained from a predictive diagnostic model (values from 0 to 1). In addition, more optimal-cutpoint selection methods proposed in the literature were included, such as criteria based on predictive values or likelihood ratios, and the incorporation of costs and/or prevalence in some of these criteria. In general, the choice of method to be applied in practice ought to be based on the researcher's specific goal and the diagnostic properties (sensitivity, specificity, predictive values, etc.) sought, depending mainly on the disease under study.

Moreover, with our package, users can easily obtain numerical (point and confidence interval estimates) and graphical output for all methods with just one input command, and make decisions accordingly. We thus trust that a program displaying these features will prove useful to the biomedical community; a program which we are thinking of continually improving, by enabling, say, the incorporation of covariates of a continuous nature and implementing new, more efficient methods for estimating optimal cutpoints under each of the criteria outlined, e.g., under certain parametric assumptions (only empirical estimators were considered in this first version of the package). It would also be useful to extend the **OptimalCutpoints** package to the situation of partial disease verification (Begg and Greenes 1983; Zhou 1993, 1994, 1998), i.e., where the true disease status of all the patients in the sample is not known, or the costs of incorrect classifications of diagnoses and/or prevalence of the disease must be taken into account in other criteria. This is where the focus of our future research should be concentrated.

This study centered on the field of diagnostic tests, but **OptimalCutpoints** may also be applied in any field where signal-to-noise analysis is performed, such as screening, radio-diagnostic techniques or biology, among others.

Acknowledgments

M. López-Ratón, M.X. Rodríguez-Álvarez and C. Cadarso-Suárez acknowledge the financial support received in the form of grants MTM2008-01603, MTM2010-09213-E and MTM2011-28285-C02-00 from Spain's Ministry of Science & Innovation. The work of M.X. Rodríguez-Álvarez was supported by Carlos III Institute of Health grant CA09/0053.

References

- Agresti A, Coull BA (1998). "Approximate Is Better than "Exact" for Interval Estimation of Binomial Proportions." *The American Statistician*, **52**, 119–126.
- Albert A, Harris EK (1987). *Multivariate Interpretation of Clinical Laboratory Data*. Marcel Dekker, New York.
- Altman DG, Lausen B, Sauerbrei W, Schumacher M (1994). "Dangers of Using "Optimal" Cutpoints in the Evaluation of Prognostic Factors." *Journal of the National Cancer Institute*, **86**(11), 829–835.
- Álvarez García G, Collantes-Fernández E, Costas E, Rebordosa X, Ortega-Mora LM (2003). "Influence of Age and Purpose for Testing on the Cut-Off Selection of Serological Methods in Bovine Neosporosis." *Veterinary Research*, **34**, 341–352.

- Amaro A, Gude F, Gonzalez-Juanatey R, Iglesias F, Fernandez-Vazquez F, Garcia-Acuna J, Gil M (1995). "Plasma Leukocyte Elastase Concentration in Angiographically Diagnosed Coronary Artery Disease." *European Heart Journal*, **16**, 615–622.
- Aoki K, Misumi J, Kimura T, Zhao W, Xie T (1997). "Evaluation of Cutoff Levels for Screening of Gastric Cancer Using Serum Pepsinogens and Distributions of Levels of Serum Pepsinogens I, II and of PG I/PG II Ratios in a Gastric Cancer Case-Control Study." *Journal of Epidemiology*, **7**, 143–151.
- Bamber D (1975). "The Area above the Ordinal Dominance Graph and the Area below the Receiver Operating Graph." *Journal of Mathematical Psychology*, **12**, 387–415.
- Begg C, Greenes R (1983). "Assessment of Diagnostic Tests when Disease Verification Is Subject to Selection Bias." *Biometrics*, **39**, 207–215.
- Begg CB, Cramer LD, Venkatraman ES, Rosai J (2000). "Comparing Tumour Staging and Grading Systems: A Case Study and a Review of the Issues, Using Thymoma as a Model." *Statistics in Medicine*, **19**, 1997–2014.
- Böhning D, Holling H, Patilea V (2011). "A Limitation of the Diagnostic-Odds Ratio in Determining an Optimal Cut-Off Value for a Continuous Diagnostic Test." *Statistical Methods in Medical Research*, **20**, 541–550.
- Bortheyry AL, Malerbi DA, Franco LJ (1994). "The ROC Curve in the Evaluation of Fasting Capillary Blood Glucose As a Screening Test for Diabetes and IGT." *Diabetes Care*, **17**, 1269–1272.
- Boyko EJ (1994). "Ruling out or Ruling in Disease with the Most Sensitive or Specific Diagnostic Test: Short Cut or Wrong Turn?" *Medical Decision Making*, **14**, 175–179.
- Brasil P (2010). *DiagnosisMed: Diagnostic Test Accuracy Evaluation for Medical Professionals*. R package version 0.2.3, URL <http://R-Forge.R-project.org/projects/diagnosimed>.
- Burgueño MJ, García-Bastos JL, González-Buitrago JM (1995). "Las Curvas ROC en la Evaluación de Pruebas Diagnósticas." *Medicina Clínica*, **104**, 661–670.
- Carstensen B, Plummer M, Laara E, Hills M (2013). *Epi: A Package for Statistical Analysis in Epidemiology*. R package version 1.1.49, URL <http://CRAN.R-project.org/package=Epi>.
- Clopper C, Pearson ES (1934). "The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial." *Biometrika*, **26**, 404–413.
- Coffin M, Sukhatme S (1997). "Receiver Operating Characteristic Studies and Measurement Errors." *Biometrics*, **53**, 823–837.
- Cohen J (1960). "A Coefficient of Agreement for Nominal Scales." *Educational and Psychological Measurement*, **20**, 37–46.
- DeLong ER, DeLong DM, Clarke-Pearson DL (1988). "Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach." *Biometrics*, **44**, 837–845.

- Edwards W, Guttentag M, Snapper K (1975). “A Decision Theoretic Approach to Evaluation Research.” In EL Struening, M Guttentag (eds.), *Handbook of Evaluation Research*, volume I, pp. 139–181. Sage Publications, Beverly Hills, California.
- England WL (1988). “An Exponential Model Used for Optimal Threshold Selection on ROC Curves.” *Medical Decision Making*, **8**, 120–131.
- Feinstein SH (1975). “The Accuracy of Diver Sound Localization by Pointing.” *Understanding Biomedical Research*, **2**(3), 173–184.
- Filella X, Alcover J, Molina R, Giménez N, Rodríguez A, Jo J, Carretero P, Ballesta AM (1995). “Clinical Usefulness of Free PSA Fraction As an Indicator of Prostate Cancer.” *International Journal of Cancer*, **63**, 780–784.
- Fleiss JL (1981). *Statistical Methods for Rates and Proportions*. John Wiley & Sons, New York.
- Freeman EA, Moisen G (2008). “**PresenceAbsence**: An R Package for Presence-Absence Analysis.” *Journal of Statistical Software*, **23**(1), 1–31. URL <http://www.jstatsoft.org/v23/i11/>.
- Galen RS (1986). “Use of Predictive Value Theory in Clinical Immunology.” In NR Rose, H Friedmann, JL Fahey (eds.), *Manual of Clinical Laboratory Immunology*, pp. 966–970. American Society of Microbiology, Washington, DC.
- Gallop RJ, Crits-Christoph P, Muenz LR, Tu XM (2003). “Determination and Interpretation of the Optimal Operating Point for ROC Curves Derived through Generalized Linear Models.” *Understanding Statistics*, **2**(4), 219–242.
- Gart JJ, Nam J (1998). “Approximate Interval Estimation of the Ratio of Binomial Parameters: A Review and Corrections for Skewness.” *Biometrics*, **44**, 323–338.
- Geisser S (1998). “Comparing Two Tests Used for Diagnostic or Screening Processes.” *Statistics Probability Letters*, **40**, 113–119.
- Gönen M, Sima C (2013). “Optimal Cutpoint Estimation with Censored Data.” *Journal of Statistical Theory and Practice*, **7**, 345–359.
- Green DM, Swets JA (1966). *Signal Detection Theory and Psychophysics*. John Wiley & Sons, New York.
- Greiner M (1995). “Two-Graph Receiver Operating Characteristic (TG-ROC): A Microsoft **Excel** Template for the Selection of Cut-Off Values in Diagnostic Tests.” *Journal of Immunological Methods*, **185**(1), 145–146.
- Greiner M (1996). “Two-Graph Receiver Operating Characteristic (TG-ROC): Update Version Supports Optimisation of Cut-Off Values That Minimise Overall Misclassification Costs.” *Journal of Immunological Methods*, **191**, 93–94.
- Greiner M, Pfeiffer D, Smith RD (2000). “Principals and Practical Application of the Receiver Operating Characteristic Analysis for Diagnostic Tests.” *Preventive Veterinary Medicine*, **45**, 23–41.

- Halpern EJ, Albert M, Krieger AM, Metz CE, Maidment AD (1996). “Comparison of Receiver Operating Characteristic Curves on the Basis of Optimal Operating Point.” *Academic Radiology*, **3**, 245–253.
- Hand D (1987). “Screening versus Prevalence Estimation.” *Applied Statistics*, **36**, 1–7.
- Hilsenbeck SG, Clark GM, McGuire WL (1992). “Why Do so Many Prognostic Factors Fail to Pan Out?” *Breast Cancer Research and Treatment*, **22**, 197–206.
- Hoffman RM, Clanon DL, Littenberg B, Frank JJ, Peirce JC (2000). “Using the Free-To-Total Prostate-Specific Antigen Ratio to Detect Prostate Cancer in Men with Nonspecific Elevations of Prostate-Specific Antigen Levels.” *Journal of General Internal Medicine*, **15**, 739–748.
- Hosmer DW, Lemeshow S (2000). *Applied Logistic Regression*. John Wiley & Sons, New York.
- Itoh M, Takahashi K, Nishida H, Sakagami K, Okubo T (1996). “Estimation of the Optimal Cut Off Point in a New Immunological Faecal Blood Test in a Corporate Colorectal Cancer Screening Programme.” *Journal of Medical Screening*, **3**(2), 66–71.
- Kelly MJ, Dunstan FD, Lloyd K, Fone DL (2008). “Evaluating Cutpoints for the MHI-5 and MCS Using the GHQ-12: A Comparison of Five Different Methods.” *BMC Psychiatry*, **8**, 10.
- Koopman PAR (1984). “Confidence Limits for the Ratio of Two Binomial Proportions.” *Biometrics*, **40**, 513–517.
- Kraemer HC (1992). “Risk Ratios, Odds Ratio, and the Test QROC.” In *Evaluating Medical Tests*, pp. 103–113. Sage Publications, Beverly Hills, California.
- Kraemer HC, Periyakoil VS, Noda A (2002). “Kappa Coefficients in Medical Research.” *Statistics in Medicine*, **21**, 2109–2129.
- Lausen B, Schumacher M (1996). “Evaluating the Effect of Optimized Cutoff Values in the Assessment of Prognostic Factors.” *Computational Statistics & Data Analysis*, **21**, 307–326.
- Lewis JD, Chuai S, Nessel L, Lichtenstein GR, Aberra FN, Ellenberg JH (2008). “Use of the Noninvasive Components of the Mayo Score to Assess Clinical Response in Ulcerative Colitis.” *Inflammatory Bowel Diseases*, **14**(12), 1660–1666.
- Lusted LB (1968). *Introduction to Medical Decision Making*. 2nd edition. Springfield, Illinois, CC Thomas.
- Manel S, Williams H, Ormerod S (2001). “Evaluating Presence-Absence Models in Ecology: The Need to Account for Prevalence.” *Journal of Applied Ecology*, **38**, 921–931.
- Mazumdar M, Glassman JR (2000). “Categorizing a Prognostic Variable: Review of Methods, Code for Easy Implementation and Applications to Decision-Making about Cancer Treatments.” *Statistics in Medicine*, **19**, 113–132.

- McNeill BJ, Keeler E, Adelstein SJ (1975). “Primer on Certain Elements of Medical Decision Making, with Comments on Analysis ROC.” *The New England Journal of Medicine*, **293**, 211–215.
- Metz CE (1978). “Basic Principles of ROC Analysis.” *Seminars in Nuclear Medicine*, **8**, 183–298.
- Metz CE, Starr SJ, Lusted LB, Rossmann K (1975). “Progress in Evaluation of Human Observer Visual Detection Performance Using the ROC Curve Approach.” In C Raynaud, AE Todd-Pokropek (eds.), *Information processing in scintigraphy*, pp. 420–436. Commissariat à l’Energie Atomique, Département de Biologie, Service Hospitalier Frédéric Joliot, Orsay.
- Miller R, Siegmund D (1982). “Maximally Selected Chi Square Statistics.” *Biometrics*, **38**, 1011–1016.
- Peng CYJ, So TSH (2002). “Logistic Regression Analysis and Reporting: A Primer.” *Understanding Statistics*, **1**(1), 31–70.
- Pepe M (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford University Press, New York.
- Perkins NJ, Schisterman EF (2006). “The Inconsistency of “Optimal” Cutpoints Obtained Using Two Criteria Based on the Receiver Operating Characteristics Curve.” *American Journal of Epidemiology*, **163**, 670–675.
- R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Riddle DL, Stratford PW (1999). “Interpreting Validity Indexes for Diagnostic Tests: An Illustration Using the Berg Balance Test.” *Physical Therapy*, **79**, 939–950.
- Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Müller M (2011). “**pROC**: An Open-Source Package for R and S+ to Analyze and Compare ROC Curves.” *BMC Bioinformatics*, **12**(77).
- Rubin DB, Schenker N (1987). “Logit-Based Interval Estimation for Binomial Data Using the Jeffreys Prior.” *Sociological Methodology*, **17**, 131–144.
- Rutter CM, Miglioretti DL (2003). “Estimating the Accuracy of Psychological Scales Using Longitudinal Data.” *Biostatistics*, **4**(1), 97–107.
- Schäfer H (1989). “Constructing a Cut-Off Point for a Quantitative Diagnostic Test.” *Statistics in Medicine*, **8**, 1381–1391.
- Schisterman EF, Perkins NJ, Liu A, Bondell H (2005). “Optimal Cutpoint and Its Corresponding Youden Index to Discriminate Individuals Using Pooled Blood Samples.” *Epidemiology*, **16**, 73–81.
- Schulgen G, Lausen B, Olsen JH, Schumacher M (1994). “Outcome-Oriented Cutpoints in Analysis of Quantitative Exposures.” *American Journal of Epidemiology*, **140**, 172–184.

- Shapiro DE (1999). “The Interpretation of Diagnostic Tests.” *Statistical Methods in Medical Research*, **8**, 113–134.
- Simel DL, Samsa GP, Matchar DB (1991). “Likelihood Ratios with Confidence: Sample Size Estimation for Diagnostic Test Studies.” *Journal of Clinical Epidemiology*, **44**(8), 763–770.
- Smith RD (1991). “Evaluation of Diagnostic Tests.” In RD Smith (ed.), *Veterinary Clinical Epidemiology*, 3rd edition, pp. 29–43. Butterworth-Heinemann, Stoneham.
- Swets JA (1979). “ROC Analysis Applied to the Evaluation of Medical Imaging Techniques.” *Investigative Radiology*, **14**, 109–121.
- Swets JA, Pickett RM (1982). *Evaluation of Diagnostic Systems: Methods from Signal Detection Theory*. Academic Press, New York.
- Swets JA, Swets JB (1979). “ROC Approach to Cost/Benefit Analysis.” In *Proceedings of the Sixth IEEE Conference on Computer Applications in Radiology*, pp. 203 – 206. Reprinted in Ripley KL, Murray A (eds.).
- VanDerWal J, Falconi L, Januchowski S, Shoo L, Storlie C (2012). *SDMTools: Species Distribution Modelling Tools: Tools for Processing Data Associated with Species Distribution Modelling Exercises*. R package version 1.1-20, URL <http://CRAN.R-project.org/package=SDMTools>.
- Vermont J, Bosson JL, François P, Robert C, Rueff A, Demongeot J (1991). “Strategies for Graphical Threshold Determination.” *Computer Methods and Programs in Biomedicine*, **35**, 141–150.
- Wald A, Wolfowitz J (1939). “Confidence Limits for Continuous Distribution Functions.” *The Annals of Mathematical Statistics*, **10**, 105–118.
- Youden WJ (1950). “Index for Rating Diagnostic Tests.” *Cancer*, **3**, 32–35.
- Zhou XH (1993). “Maximum Likelihood Estimators of Sensitivity and Specificity Corrected for Verification Bias.” *Communication in Statistics-Theory and Methods*, **22**, 3177–3198.
- Zhou XH (1994). “Effect of Verification Bias on Positive and Negative Predictive Values.” *Statistics in Medicine*, **13**, 1737–1745.
- Zhou XH (1998). “Correcting for Verification Bias in Studies of a Diagnostic Test’s Accuracy.” *Statistical Methods in Medical Research*, **7**, 337–353.
- Zweig MH, Campbell G (1993). “Receiver-Operating Characteristics (ROC) Plots: A Fundamental Evaluation Tool in Clinical Medicine.” *Clinical Chemistry*, **39**, 561–577.

Affiliation:

Mónica López-Ratón

Unidad de Bioestadística. Departamento de Estadística e Investigación Operativa

Facultad de Medicina. San Francisco s/n

Santiago de Compostela, Spain

E-mail: monica.lopez.raton@gmail.com