Summer 2022

Investigating Cut-point Methods in ROC Analysis: A Critique of Alternative Approaches

Gregory W. McGuire

SHAWNEE STATE UNIVERSITY

Investigating Cut-point Methods in ROC Analysis: A Critique of Alternative Approaches

A Thesis

By

Gregory W. McGuire

Department of Mathematical Sciences

Submitted in partial fulfillment of the requirements

for the degree of

Master of Science, Mathematics

July 2022

Accepted by the Graduate Department

7/27/2022

Graduate Director, Date

The thesis entitled '**Investigating Cut-point Methods in ROC Analysis: A Critique of Alternative Approaches**' presented by **Gregory W. McGuire**, a candidate for the degree of **Master of Science in Mathematics**, has been approved and is worthy of acceptance.

7/27/2022
Date

Graduate Director

7/27/2022
Date

Student

ABSTRACT

In 2017, Ilker Unal presented the Index of Union method for obtaining optimal cut-points in ROC analysis and claimed that it outperformed other methods, including the historied Youden Index. This is an investigation into that claim using generated data. It specifically pits the Youden Index method against the Index of Union (IU) method under various circumstances. The data sets have different ratios of diseased and non-diseased data points along with different ratios of true and false results based on a theoretical true cut-point. The data was analyzed to see if any patterns emerged as to when the Youden Index obtain a cut-point closer to the theoretical true cut-point and when the IU method does. Although the IU method performed better in the majority of data sets, the Youden Index method did outperform the IU method at times. The IU method had a clear advantage in the case that the specificity and sensitivity were equal, while the Youden Index had an advantage when the area under the ROC curve was between 0.27 and 0.47. The results imply that there is good reason for the uncertainty in the landscape of methods for obtaining optimal cut-points, but there may be a good argument as to when to use one over another. More research should be done into the relationship between the area under the curve and these methods. In the meantime, medical researchers should not rely on a single method but rather take into the range of cut-points obtained by various methods.

ACKNOWLEDGMENTS

I would like to thank my wife, Lauren, without whom this thesis would not have been possible. (I love you with all my heart!) If not for her encouragement, I would have neither begun nor finished my degree which saw the births of our now 17-month-old son and 5-month-old daughter. I know that it has been as challenging for her as it has for me, if not more so, and I appreciate every sacrifice made along the way!

Next, although she will tell me that she needs no such acknowledgement, I would like to thank my mom for spending countless hours of quality time helping with her grandbabies as I completed my thesis and coursework.

Lastly, I would like to thank Dr. Doug Darbro for being understanding as I struggled to meet deadlines near the birth of our daughter, and I would like to thank Dr. John Whitaker for publishing assignments in advance to allow me the opportunity to get ahead in Analysis II as her birth and this thesis loomed large in my mind.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

CHAPTER I: INTRODUCTION

Chapter 1 will provide an introduction to ROC analysis and methods for determining optimal cut-points. Chapter 1 will also address the research problem being investigated, the purpose of the study, the research hypotheses, and the significance of the present study. The chapter will then conclude with a preview of the remainder of the thesis.

ROC Analysis and Optimal Cut-points

A receiver operating characteristic (ROC) curve is a graphical way of dichotomizing data into those items that carry a trait and those that do not. It was originally developed during World War II to help the U.S. military differentiate between signals and noise in radar detection, but it has since proven to be beneficial for medical research in particular. In order to categorize data into signal and noise, carriers and non-carriers, a cut-point must be calculated and used to divide the two groups (Zou, O'Malley, & Mauri, 2007, Zweig & Campbell, 1993). Over the years, various methods for determining the cut-point have been proposed, but given the stakes involved with the data being analyzed, it is paramount that researchers use the method that gives the optimal cut-point, i.e., the one that leads to the most accurate classification of the data.

Originally published in 1950, Youden's index, also known as Youden's *J* statistic, has been the industry standard for calculating a cut-point for over 50 years (Youden, 1950, Perkins & Schisterman, 2005). However, in 2017, Ilker Unal conducted a study to compare Youden's index and three other contenders (Minimum P Value, Closest to (0,1), and the Concordance Probability methods) with a new method he proposed, the Index of Union (IU) Method, and his simulations led to the study's conclusion that the IU method should be preferred to even Youden's index (Unal, 2017). A number of follow-up studies as well as additional methods have been published in the last four years, but Youden's index has not been displaced in the field at large. Whether it ought to be remains debatable, but one thing is certain – with so much riding on accuracy in ROC analysis, the IU method should be given due consideration (Linden & Yarnold, 2018, Mira, et al, 2019, Feng, Griffin, Kethireddy, & Mei, 2021, Hong, Choi, & Lim, 2020).

Statement of the Problem

The problem that will be addressed in this study is that there is a lack of consensus on what is the best method for defining an optimal cut-point value in ROC analysis. This lack of consensus may lead to misclassification of biomarkers for diseased and non-diseased patients. It also may cause researchers to present findings using multiple methods unnecessarily, which burdens the reader and obscures the literature.

Purpose of the Study

The purpose of the study is to extend the research done by Ilker Unal, in an effort to provide data analysts with the best means by which to classify biomarker data correctly into diseased or not. In particular, the study will compare Unal's Index of Union Method and Youden's Index under various situations. Consistently determining the optimal cut-point is the main concern of the study, but it would be helpful to pinpoint the circumstances under which a given method outperforms the other.

Significance of the Study

Given the great importance of preventing errors in diagnoses and the value of being able to accurately identify disease with a single test, medical research could be greatly enhanced by having a clear, mathematically optimal, optimal cut-point method. Time will tell whether the IU method can fulfill that role and surpass the Youden Index method, and perhaps, this study can fit one more piece into that puzzle.

Research Questions

Primary Question: Does the IU Method for calculating the optimal cut-point consistently outperform the Youden Index?

Secondary Question 1: Does the IU Method for calculating the optimal cut-point outperform the Youden Index when the Sensitivity and Specificity are equal?

Secondary Question 2:[1] Does the IU Method for calculating the optimal cut-point outperform the Youden Index when the sample size is lower (n = 50) or higher (n = 1000)?

Secondary Question 3: Does whether the sensitivity or specificity is higher determine whether the IU Method for calculating the optimal cut-point outperforms the Youden Index?

Hypotheses

Primary Hypothesis: The hypothesis is that the IU Method will fail to outperform the Youden Index under certain conditions within the simulations.

Secondary Hypothesis 1: The hypothesis is that the IU Method will outperform the Youden Index when the Sensitivity and Specificity are equal.

Secondary Hypothesis 2: The hypothesis is that sample size will have no bearing on which method performs better.

Secondary Hypothesis 3: The hypothesis is that it does not matter whether the sensitivity or specificity is higher.

Research Design

This study is an article review that will involve a simulation of generated data with comparisons between methods. It is hypothesis testing in a non-traditional sense.

---

[1] Secondary Question 2 was later removed.

Data will be generated for the simulation. Those data sets will include various sample sizes (n = 50, 100, 1000)[2] with various values of true positives (a), false positives (c), true negatives (d), and false negatives (b) to have been found by a supposed test. Those a, b, c, and d values will be manufactured at (a+b) is 10%, 50%, and 90% of the total sum, and for each of those, the ratio of a:b and c:d will also be varied at 1:9, 1:1, and 9:1. Particular numerical values will be obtained using an online random number generator.

Data will be recorded in Excel, and R will be used for the statistical analyses and computations. A number of packages in R include commands to perform the Youden method, e.g., cutpointr, pROC, and OptimalCutpoints, but there is none available for the IU method. In Unal's study, he mentioned that he would make his R code available upon request. (Unal, 2017) That request has been made without response. Alternatively, computations will be done using R and employ commands for sensitivity and specificity to help simplify the process.

Definition of Terms

1. Optimal Cut-point: Although specific methods define the optimal cut-point in slightly different terms, a generalized definition would be when the point classifies the largest percent of the data items correctly. (Linden & Yarnold, 2018)
2. ROC curve: A receiver operating characteristic (ROC) curve is a plot of sensitivity on the y-axis from 0 to 1 against (1-specificity) on the x-axis from 0 to 1. The following three terms are important summary values for an ROC curve:

   (a) AUC: AUC stands for "Area under the Curve". This is a summary of diagnostic accuracy. An AUC = 0.5 would be a diagonal line from (0,0) to (1,1),

[2] The research design was later modified to only include a sample size of 1000.

and it represents random chance. The closer the area gets to 1, the more accurate it is. (Zou, O'Malley, & Mauri, 2007)

(b) Sensitivity: Sensitivity is the true positive rate, Se(c). This would be the probability that a test correctly classifies a diseased subject as positive, a/(a+b) where a = true positives and b = false negatives.

(c) Specificity: Specificity is the true negative rate, Sp(c). This would be the probability that a test correctly classifies a non-diseased subject as negative, d/(c+d) where c = false positives and d = true negatives.

3. Unal's IU Method: The Index of Union method attempts to maximize the sensitivity and specificity values at the same time. Given the function IU(c) = (|Se(c) - AUC| + (Sp(c) – AUC|), the optimal cut-point value $\hat{c}_{IU}$ minimizes the IU(c) function and the |Se(c)-Sp(c)| difference. (Unal, 2017)

4. Youden's *J* statistic: *J*(c) = Se(c) + Sp(c) – 1 over all cut-points c. $\hat{c}_j$ is the optimal cut-point value when J is maximum. Equivalently, Youden originally gave his statistic as J = (ad – bc)/((a+b)(c+d)) where there are "a" correctly diagnosed and "b" falsely negative diseased patients and there are "c" false positives and "d" correctly reported as non-diseased. (Youden, 1950, Unal, 2017, Fluss, Faraggi, & Reiser, 2005, Perkins & Schisterman, 2005)

Summary

Chapter 1 introduced the problem statement to be investigated and provided the purpose and significance of the current study. Chapter 2 will delve deeper into the current research literature relevant to the study. In Chapter 3, the methodology used for the study will be laid out, while Chapter 4 will present the results of the analysis. Finally, Chapter 5 will tie everything together with conclusions and recommendations.

CHAPTER II: BACKGROUND AND LITERATURE REVIEW

Chapter 2 will provide a review of the literature designed to highlight the need for the study at hand. Chapter 2 takes the reader through 5 sub-chapters that are grouped according to key ideas that begin broadly and narrow in on the focus of the study, conveniently following chronological order all the while. The chapter begins with ROC Analysis in general and then optimal cut-points within that, and the chapter concludes by considering Youden's Index, the Index of Union, and, most importantly, comparisons of the methods.

ROC Analysis and Optimal Cut-points

Receiver operating characteristic (ROC) analysis began as a U.S. military project to decipher Japanese radar signals from noise. After the war, its ability to dichotomize data was coopted by medical researchers who wanted to make sure that the diagnostic tests being used to classify people into groups who have a certain condition and those who do not, i.e., diseased and non-diseased, were doing that accurately with reliable cut-off marks, referred to as "cut-points". ROC analysis has been extremely useful in laboratory testing, epidemiology, radiology, bioinformatics, oncology, and cardiology, to name just a few areas (Zou, O'Malley, & Mauri, 2007, Zweig & Campbell, 1993).

A major concern in ROC analysis is that there are two obvious problems with these medical diagnostic tests: telling a healthy person he is diseased (false positive) and telling a diseased person he is healthy (false negative). In order to avoid those errors, researchers need a way to determine the value that leads to the fewest errors occurring on these medical tests, i.e., the optimal cut-point. To analyze the situation, ROC curves give us a visual display of a test's performance. They plot the true positive rates (sensitivity) along the y-axis and the false positive rates (1-specificity) along the x-axis that would result for various cut-points. Since these rates are probabilistic, the values range from 0 to 1. Ideally, the sensitivity and specificity would both be 1, which would occur at the point (0,1) on the ROC curve. Demonstrably, in the real world, our tests are not so perfect, and identifying an optimal cut-point and calculating its value is not so obvious.

Thus, the last 75 years have seen a number of methods put forth as the mathematically best way to calculate the optimal cut-point. These methods include the Concordance Probability method, the Minimum P Value method, and the ODA method, but they are not particularly intuitive or geometrically interpretable methods, despite having merits of their own (Perkins & Schisterman, 2006, Linden & Yarnold, 2018, Unal, 2017). There is another method that is clearly more intuitive and popular: Closest to (0, 1). Given an ROC curve, as described above, the closest to (0, 1) method simply calculates the Euclidean distances from the points on the curve to the ideal point to see which one is closest to it. It makes perfect geometric sense, although it is deceptively difficult to give an interpretation of the quadratic terms that result in the distance formula. Simple and logical, but does it perform consistently better than other methods?

Youden's Index

The other method that holds intuitively, albeit less geometrically apparent, is the Youden Index method. This method was originally published in 1950, and it has been in standard use for ROC analysis ever since. Like the Closest to (0, 1) method, this Youden

Index method is intuitive with a natural interpretation. The goal is still to maximize the true results and minimize the false ones. It has been suggested that the longevity of Youden's index stems from that simplicity and clarity, but that does not mean it succeeds in determining what it purports to determine or whether it does so in the best way possible.

Youden's *J* statistic is calculated by subtracting 1 from the sum of the sensitivity and specificity of a cut-point, and the maximum value of J is the optimal cut-point. Equivalently, Youden gave his statistic as $J = (ad - bc)/((a+b)(c+d))$ where there are "a" true positives,"b" false negatives,"c" false positives, and "d" true negatives. (Youden, 1950, Unal, 2017, Fluss, Faraggi, & Reiser, 2005, Perkins & Schisterman, 2005)

This can be given a more palatable geometric meaning in connection to the area under the curve (Hilden & Glasziou, 1996), and the Youden *J* statistic turns out to be the point on the curve farthest from the y = x line, which would be a 50/50 chance (Perkins & Schisterman, 2006). Thus, undergirded by a simple notion, calculated as a statistic that employs basic mathematical operations, and connected to geometric reality, Youden's Index is a prime candidate for researchers who are using ROC analysis and finding optimal cut-points.

Index of Union Method

In 2017, Ilker Unal conducted a study to compare Youden's method and three of the other competing optimal cut-point methods (Minimum P Value, Closest to (0,1), and the Concordance Probability methods) with his newly proposed Index of Union (IU) Method. The main objective of the IU method is to minimize the sum of the absolute difference between the sensitivity and the area under the ROC curve (AUC) and the absolute difference between the specificity and the AUC:

$$\min(IU(c)) = \min(|Se(c) - AUC| + |Sp(c) - AUC|).$$

The secondary objective is to minimize the absolute difference between the specificity and sensitivity, but this second assumption only needs to be addressed when there would otherwise be a tie for the optimal cut-point. Geometrically, the method is minimizing the half perimeter of the rectangle with the following vertices: (1 – AUC, AUC), (1 – Sp(c), AUC), (1 – AUC, Se(c)), and (1 – Sp(c), Se(c)). In a way that is not true for the Closest to (0,1) and Youden methods, this appears to beg the question, why? Nonetheless, after conducting simulations, Unal was led to conclude that the IU method is preferable to all other methods (Unal, 2017).

Since the conclusion of Unal's study, researchers have begun to give it due consideration. Studies quickly began to report the optimal cut-point using the IU method alongside others (Linden & Yarnold, 2018, Mira, et al, 2019, Feng, Griffin, Kethireddy, & Mei, 2021, Hong, Choi, & Lim, 2020). Sadly, the literature lacks a comprehensive study on the Index of Union method itself, but that has not deterred researchers from calculating and reporting it.

Comparison Studies

Much to a researcher's dismay, the conclusion of an article only opens the door to the discussion rather than closing it. Of course, for the IU method to become the industry standard, it must overcome an "ever-increasing body of supporting literature" for the Youden Index, not to mention others (Perkins & Schisterman, 2006).

Since the publication of the Youden Index in 1950, other methods for determining optimal cut-points have made the in-roads that the IU method has begun in such a short amount of time. Unfortunately, although the Index of Union has gained some level of popularity in the last five years, there is not much available by way of comparison studies between the Index of Union Method and the Youden Index Method to date.

Besides Unal's own study in which he concluded that his IU method outperforms other methods, including Youden's, the literature includes a single full-length comparison of the IU method with Youden's Index. In 2020, Hong, Choi, and Lim from Sungkyunkwan University in Seoul, South Korea ran simulations using various methods, including both the IU and Youden's Index. The major problem with this study is that it was published in Korean, and only a poor translation to English was obtainable.

Although the English translation was extremely difficult to follow, there were intelligible parts worth discussion, especially the results that were originally presented in tables with English headings. Hong, Choi, and Lim found that the IU method and Youden's Index had fewer type 1 and type 2 errors than the others, and interestingly, they reported that the errors for these methods converge to the same value. However, they ultimately concluded that IU Method is the most efficient method, because it converged much more quickly (Hong, Choi, & Lim, 2020).

Despite the findings of Unal and Hong, Choi, and Lim as well as a plethora of studies comparing other methods with the Youden Index Method, researchers have been left defending a choice or presenting multiple methods for the reader to wade through and make their own choices (Perkins & Schisterman, 2006, Zou, O'Malley, & Mauri, 2007).

One tactic suggested in the literature is for different methods to be presented for different reasons, depending on what the researcher wants to highlight. The Youden index method could be chosen for situations in which the researcher is interested in interpreting the net gain of the true positive accounting for the false positive, for example (Rota, Antolini, & Valsecchi, 2015). Most articles seem to opt for presenting as much as possible to the reader. Thus, the diagnosticians never get to the bottom of the situation and report ranges, and clinicians must have secondary tools for disease identification (Perkins & Schisterman, 2006, Zou, O'Malley, & Mauri, 2007).

Given the great importance of preventing errors in diagnoses and the value of being able to accurately identify disease with a single test, medical research could be greatly enhanced by having a clear, mathematically optimal, optimal cut-point method. Time will tell whether the IU method can fulfill that role and surpass the Youden Index method, and perhaps, this study can fit one more piece into that puzzle.

Summary

Chapter 2 reviewed literature relevant to ROC Analysis, Optimal Cut-points, Youden Index Method, Index of Union Method, and Comparison Studies. The lack of comprehensive studies on the IU Method shows a need for future research, and there is a further need of comparison studies between the IU and Youden Index Methods that the present study attempts to fill in part. In the end, the literature showed that the failure of comparison studies to put forth a clear optimal cut-point creates a real burden on research that makes use of ROC Analysis, which is the real justification of this study.

Now that the reader has become acquainted with the study and reviewed the literature, the methodology will be laid out in Chapter 3, the results in Chapter 4, and the conclusions and recommendations in Chapter 5.

CHAPTER III: METHODOLOGY

Chapter 3 will provide an overview of the research design for this study. Chapter 3 lays out how data was generated and according to what parameters. Finally, the method of statistical analysis is described, which ties to the statistical results discussed in the next chapter.

Research Design

This study is a simulation of generated data to compare the performances of the Index of Union and Youden Index methods. It is a direct extension of the work done by Ilker Unal, in which he similarly conducted a simulation with generated data. In the end, he presented the Index of Union as the best choice for obtaining Optimal Cut-points in ROC analysis.

Data Generation

Data was generated for the simulation using R. The data sets included various sample sizes (n = 50, 100, 1000 with $\mu = 0$ and $\sigma = 1$)[3] with various numbers of true positives (a), false positives (c), true negatives (d), and false negatives (b) to have been found by a supposed test with a true theoretical cut-point. The simulation was also conducted once with the true theoretical cut-point was set at 0 and again set arbitrarily at 1.5.[4] The a, b, c, and d values were manufactured at (a+b) is 10%, 50%, and 90% of the total sum, and for each of those, the ratio of a:b and c:d were also varied at 1:9, 1:1, and 9:1.

---

[3] This was later modified to include only a sample size of 1000.
[4] This was later modified to include only a true theoretical cut-point of 0.5.

To be more specific, the numerical values were obtained using the function rnorm in R. Values for a, b, c, and d were accepted in alphabetical order until the desired number of data points for each category was reached. The rnorm function was chosen to delimit the data sets to values originating from a normal distribution.

Data Analysis and Interpretation

Data was recorded in Excel, and R was used for the statistical analyses and computations. A number of packages in R include commands to perform the Youden method, e.g., cutpointr, pROC, and OptimalCutpoints, but there is none available to calculate the Index of Union directly. In Unal's study, he mentioned that he would make his R code available upon request. (Unal, 2017) That request was made via email with no response to date, so the computations were done by employing commands for sensitivity and specificity to help simplify the process and create a formula to find the Index of Union, i.e., $\min(IU(c)) = \min(|Se(c) - AUC| + |Sp(c) - AUC|)$.[5]

Summary

Chapter 3 described how data was generated and analyzed for this simulation study. Next, chapter 4 will present the results of that analysis. Then, finally, chapter 5 will interpret and draw conclusions from those results.

[5] This was later modified to only consider the second objective of the IU method, i.e., minimizing the absolute difference between specificity and sensitivity.

CHAPTER IV: RESULTS

Chapter 4 will provide the results of statistical analyses. The purpose of this study was to compare methods for determining optimal cut-points. Data was generated according to certain specifications, in order to compare the cut-points found by the IU Method presented by Unal with those found using the Youden Index in particular. After the data is described, there will be a discussion of the research questions and hypotheses.

Generated Data Sets

The specifications in the methodology led to 27 distinct data sets. There were 9 data sets with equal sensitivity and specificity (3, 5, 7, 12, 14, 16, 21, 23, 25), nine with sensitivity higher (1, 2, 4, 10, 11, 13, 19, 20, 22), and nine with specificity higher (6, 8, 9, 15, 17, 18, 24, 26, 27). It should be noted that, although the theoretical true cut-point to form those ratios was set at 0.5, the random number generator did not produce any values between 0.48933642 and 0.5023473. For certain data sets, that range was wider in one direction or the other. Table 1 below is a guide to the data set number and the specified ratios of true/false positive and true/false negative used for the data generation. It also gives the cut-points determined by both methods, which was closer to the theoretical true cut-point, and whether the sensitivity was higher, lower, or equal to the specificity.

Table 1: Optimal Cut-points

Data Set #	Ratio of TP:FN:FP:TN	Youden Cutpoint	Unal's Cutpoint	Closer to 0.5	Sens or Spec Higher?
1	a:b:c:d = 10:90:90:810	-1.2009	-0.324	Unal	Specificity
2	a:b:c:d = 10:90:450:450	0.451	-0.0105	Youden	Specificity
3	a:b:c:d = 10:90:810:90	0.4599	0.5023	Unal	Tie
4	a:b:c:d = 50:50:90:810	2.7981	-0.0394	Unal	Specificity
5	a:b:c:d = 50:50:450:450	-0.2164	0.5023	Unal	Tie
6	a:b:c:d = 50:50:810:90	0.523	0.7491	Youden	Sensitivity
7	a:b:c:d = 90:10:90:810	2.726	0.5038	Unal	Tie
8	a:b:c:d = 90:10:450:450	2.7177	0.7235	Unal	Sensitivity
9	a:b:c:d = 90:10:810:90	0.8381	0.915	Youden	Sensitivity
10	a:b:c:d = 50:450:50:450	-0.6227	-0.2954	Unal	Specificity
11	a:b:c:d = 50:450:250:250	0.4935	-0.0202	Youden	Specificity
12	a:b:c:d = 50:450:450:50	0.5023	0.5023	TIE	Tie
13	a:b:c:d = 250:250:50:450	Infinity	-0.0202	Unal	Specificity
14	a:b:c:d = 250:250:250:250	-0.7049	0.5023	Unal	Tie
15	a:b:c:d = 250:250:450:50	0.5023	0.7582	Youden	Sensitivity
16	a:b:c:d = 450:50:50:450	3.3334	0.5023	Unal	Tie
17	a:b:c:d = 450:50:250:250	3.3334	0.7582	Unal	Sensitivity
18	a:b:c:d = 450:50:450:50	0.9411	0.9307	Unal	Sensitivity
19	a:b:c:d = 90:810:10:90	-0.3689	-0.2927	Unal	Specificity
20	a:b:c:d = 90:810:50:50	0.5038	-0.0626	Youden	Specificity
21	a:b:c:d = 90:810:90:10	0.5038	0.5038	TIE	Tie
22	a:b:c:d = 450:450:10:90	-1.5469	-0.0249	Unal	Specificity
23	a:b:c:d = 450:450:50:50	0.6772	0.5023	Unal	Tie
24	a:b:c:d = 450:450:90:10	0.5437	0.7851	Youden	Sensitivity
25	a:b:c:d = 810:90:10:90	Infinity	0.5023	Unal	Tie
26	a:b:c:d = 810:90:50:50	3.1512	0.7373	Unal	Sensitivity
27	a:b:c:d = 810:90:90:10	1.7451	0.915	Unal	Sensitivity

Research Questions and Relevant Results

Below is a reminder of the driving questions for this study along with the hypotheses from Chapter One and a discussion of the relevant results.

Primary Question: Does the IU Method for calculating the optimal cut-point consistently outperform the Youden Index?

Primary Hypothesis: The hypothesis that the IU Method will fail to outperform the Youden Index under certain conditions was confirmed. The Youden Index obtained a cut-point closer to the theoretical true value in 7 out of 27 instances (data sets 2, 6, 9, 11, 15, 20, 24) with 2 ties (data sets 12 and 21). No pattern emerged as to why those specifications led the Youden Index to an optimal cut-point that was closer to the theoretical true cut-point.[6]

Secondary Question 1: Does the IU Method for calculating the optimal cut-point outperform the Youden Index when the Sensitivity and Specificity are equal?

Secondary Hypothesis 1: The hypothesis that the IU Method will outperform the Youden Index when the Sensitivity and Specificity are equal was confirmed. In fact, all nine of the data sets that generated equal sensitivity and specificity resulted in the IU Method obtaining a cut-point that would delineate the data into exactly the same categories as had been initially generated, while two of those nine also saw the Youden Index obtain the same delineation.

Secondary Question 3:[7] Does whether the sensitivity or specificity is higher determine whether the IU Method for calculating the optimal cut-point outperforms the Youden Index?

Secondary Hypothesis 3: The hypothesis that it does not matter whether the sensitivity or specificity is higher was also confirmed. Out of the seven cases that the Youden Index outperformed the IU Method, the specificity was higher three times and the sensitivity higher four times. In the two cases that the methods tied, there was also a tie between the sensitivity and specificity.

[6] Detailed results for both methods can be viewed for every data set in Appendix.
[7] Secondary Question 2 was removed when the methodology was modified.

Additional Statistical Considerations

First, although all 27 data sets were distinct in terms of scores with their corresponding diseased or non-diseased status, the methodology used led to only 11 different pairs of mean and standard deviation. This curious statistical fact can be seen in Table 2 below, which displays the means and standard deviations for all the data sets along with which method obtained the closer cut-point to the theoretical true value.

Table 2: Means & SD

			Data Sets/Better Method				
Pair 1	Mean = -0.3527431	SD = 0.837857	1 Unal	10 Unal	19 Unal		
Pair2	Mean = 0.2312934	SD = 1.049908	2 Youden	22 Unal			
Pair 3	Mean = 0.8254608	SD= 0.8433006	3 Unal	25 Unal			
Pair 4	Mean = -0.28105	SD = 0.8805964	4 Unal	20 Youden			
Pair 5	Mean = 0.2914549	SD = 1.056868	5 Unal	12 Tied	14 Unal	16 Unal	23 Unal
Pair 6	Mean = 0.8958743	SD = 0.7913209	6 Youden	26 Unal			
Pair 7	Mean = -0.2010222	SD = 0.9382004	7 Unal	21 Tied			
Pair 8	Mean = 0.3543923	SD = 1.050842	8 Unal	24 Youden			
Pair 9	Mean = 0.9693953	SD = 0.7321414	9 Youden	18 Unal	27 Unal		
Pair 10	Mean = -0.01390235	SD = 0.9990552	11 Youden	13 Unal			
Pair 11	Mean = 0.6293439	SD = 0.9532248	15 Youden	17 Unal			

There was no noticeable connection between means and standard deviations and which method obtained an optimal cut-point closer to the theoretical true, though. The seven cases for which the Youden Index outperformed had a mean and standard deviation shared by at least one other data set for which the IU Method outperformed the Youden. However, mean/standard deviation Pair 1 led to Unal's method each time and Pair 5 led to the IU Method four out of five times with the 5th being one of the ties. Those means and standard deviations are neither higher nor lower than other pairs.

The next statistical fact to consider is that, despite the methodology presented (or rather because of it), only two of the data sets obeyed normality based on Shapiro-Wilks tests conducted on all 27 data sets[8]. The methodology laid out a way to pull data values from a normal distribution at random, which it did using rnorm in R, but it failed to safeguard that distribution in the final data sets. Since the initial set pulled was subsequently selected at various ratios and sums according to whether it was above or below the theoretical true cut-point, the resulting data sets were not themselves normally distributed. Instead of being a simple random sample from the normal distribution, it turned out to be more of a stratified random sample upon reflection. The two data sets that obeyed normality were data sets 11 and 13. Those two data sets are the only ones that have 700 of the 1000 scores below the theoretical cut-point. Since the cut-point was above the mean of the normally distributed set from which the values were pulled, it makes sense that a negative heavy set would have the best chance at swinging back to normality, as long as it does not become too negative heavy like those with 900 below.

[8] Shapiro-Wilks Tests can be viewed at the end of the Appendix

All that being said, data set 11 was Youden's victory, while data set 13 was Unal's. Thus, in that small sample, failing to reject normality did not correlate to one method over the other.

Lastly, an aspect of ROC analysis that may provide insight is the ROC curve[9] that plots the Sensitivity vs. 1 – Specificity. In particular, the areas under the curves shown in Table 3 below may be useful for figuring out why certain specifications led to the Youden Index performing better and not others.

Table 3: AUC's

Data Set	AUC	Better Method
1	0.493	Unal
2	0.3039	Youden
3	0.1053	Unal
4	0.6857	Unal
5	0.4991	Unal
6	0.3038	Youden
7	0.8763	Unal
8	0.6824	Unal
9	0.4806	Youden
10	0.4681	Unal
11	0.2748	Youden
12	0.0986	TIE
13	0.6908	Unal
14	0.5035	Unal
15	0.3012	Youden
16	0.894	Unal
17	0.6951	Unal
18	0.4895	Unal
19	0.5141	Unal
20	0.3081	Youden
21	0.0929	TIE
22	0.6729	Unal
23	0.4804	Unal
24	0.2889	Youden
25	0.9042	Unal
26	0.708	Unal
27	0.4937	Unal

[9] ROC curves are provided in the Appendix for all data sets

The data sets where the Youden Index performed better had AUC's between 0.2748 and 0.4806, while those for Unal were below 0.11 or above 0.4681. The only data set on Youden's list above that 0.4681 is one where the cut-points were very close but did not tie (0.8381 for Youden vs. 0.915 for Unal). Even more interestingly, the two times that the methods tied were the only times that the AUC was below 0.10.

Summary

Chapter Four presented the results of the statistical analyses conducted in the study. The discussion was primarily connected to the research questions that motivated the study, and additional considerations were made in attempt to shed light on what may be at play. Next, in Chapter Five, this information will be interpreted. Conclusions will be drawn, and recommendations for future studies will be made.

CHAPTER V: SUMMARY

This final chapter will provide a summary of the study considering the literature review and motivation for the study. The chapter then turns to a discussion of the limitations of the study and threats to generalizability. Finally, it presents recommendations for future studies.

Motivation

Considering the use of ROC analysis for medical diagnostic tests and the importance of finding the optimal cut-point for minimizing false positives and false negatives when it comes to matters of life-and-death, there is strong motivation to settle the debate over which method determines the optimal cut-point. The motivation is so strong, in fact, that it would be tempting to cautiously conclude that the IU Method does outperform the Youden Index and should supplant it as a gold standard. However, that is simply not warranted by this study.

Revisiting the Literature

The literature review painted a picture of uncertainty as to which method for determining a cut-point is the best. In fact, it was suggested that many diagnosticians report several methods with no sense of preference. That seems to be maintained in this study. Although minimizing the absolute difference between the sensitivity and specificity indicated a cut-point that was closer to the theoretical true cut-point more often, it was not absolute. Even worse, there was no clear reason as to why, based on the parameters considered here, except when the specificity and sensitivity are the same. In that last case, the absolute difference between them will be zero, so this will lead to the true theoretical cut-point.

This study confirmed the fact that the Youden Index method is easier to implement as the literature suggested. It also made more sense in terms of fundamental logic as well as geometric interpretation, which made its use more attractive. Thus, it is no wonder that researchers find themselves in the situation they do regarding the established Youden Index, even if it may fail to consistently outperform other methods under certain circumstances.

Limitations and Threats to Generalizability

Most of the limitations and threats to generalizability amount to flaws in the original research design. Firstly, what was deemed to be Unal's IU Method in the results of Chapter Four was restricted to only the secondary objective of that method, for which R code could be written. The R code for the primary objective of the IU Method was not received from Unal in time for this study, and it proved too difficult to reproduce. Secondly, although the data generated came from a normal distribution, the selection process betrayed that normality. Thus, this study is only reasonably reproducible with the specific data sets used here. Lastly, this study was limited by time constraints. The original methodology was not achievable, as it would have resulted in 162 data sets each requiring the same analyses applied to the 27 data sets here. As it is, this study could have also benefited from a narrower, more in-depth focus on a particular consideration.

Recommendations for Future Studies

Besides addressing the issues discussed above, there are a couple other recommendations that might aid a future researcher in a comparison of cut-point methods. First, the closest to (0, 1) method presents itself as an alternative worth inclusion in any serious conversation about optimal cut-points. Secondly, when the ROC curves were analyzed in this study, it was striking that there seemed to be a connection between the handful of data sets that led Youden to outperform the IU Method and a particular range of AUC's. This should not have been altogether surprising, given the geometric interpretations of the methods. Still yet, what seemed promising is that a delineation in this way could potentially provide medical researchers a guideline for when to use one method for determining a cut-point rather than another, which was the real motivation for the study in the first place.

REFERENCES

Bohning, D., Bohning, W., & Holling, H. (2008). Revisiting Youden's index as a useful measure of the misclassification error in meta-analysis of diagnostic studies. Statistical Methods in Medical Research, 3(3), 1-12.

Feng, C., Griffin, P., Kethireddy, S., & Mei Y. (2021). A boosting inspired personalized threshold method for sepsis screening. Journal of Applied Statistics, 48(1), 154-175. doi: 10.1080/02664763.2020.1716695

Fluss, R., Faraggi, D., & Reiser, B. (2005). Estimation of the Youden index and its associated cutoff point. Biometrical Journal, 47(4), 458-472. doi: 10.1002/bimj.200410135

Hilden J, Glasziou P. (1996). Regret graphs, diagnostic uncertainty and Youden's index. Statistics in Medicine, 15, 969–86.

Hong, C.S., Choi, S.Y., & Lim, D.H. (2020). Index of union and other accuracy measures. The Korean Journal of Applied Statistics, 33(4), 395-407. doi:10.5351/KJAS.2020.33.4.395

Linden, A. & Yarnold, P.R. (2018). Identifying maximum-accuracy cutpoints for diagnostic indexes. Optimal Data Analysis, 7, 59-65. Retrieved from https://www.researchgate.net/publication/327405683_Identifying_Maximum-Accuracy_Cut-Points_for_Diagnostic_Indexes_via_ODA

Mira, A., et al. (2019). Capturing the severity and impairment associated with depression: The overall depression severity and impairment scale (ODSIS) validation in a Spanish clinical sample. Frontiers in Psychiatry, 10(180), 1-9. doi: 10.3389/fpsyt.2019.00180

Perkins, N.J. & Schisterman, E.F. (2005). The Youden index and the optimal cut-point corrected for measurement error. Biometrical Journal, 47(4), 428-441. doi: 10.1002/bimj.200410133

Perkins, N.J. & Schisterman, E.F. (2006). The inconsistency of optimal cut-points using two ROC based criteria. American Journal of Epidemiology, 163(7), 670-675. doi: 10.1093/aje/kwj063

Rota, M., Antolini, L., & Valsecchi, M.G. (2015). Optimal cut-point definition in biomarkers: the case of censored failure time outcome. BMC Medical Research Methodology, 15(24). doi: 10.1186/s12874-015-0009-y

Unal, I. (2017). Defining an optimal cut-point value in ROC analysis: An alternative approach. Computational and Mathematical Methods in Medicine, 2017(3762651), 1-14. doi: 10.1155/2017/3762651

Youden, W.J. (1950). Index for rating diagnostic tests. Cancer, 3(1), 32-35. doi: 10.1002/1097-0142(1950)3:1<32

Zou, K. H., O'Malley, A.J., & Mauri, L. (2007). Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. Circulation, 115(5), 654-657. doi: 10.1161/CIRCULATIONAHA.105.594929

Zweig, M.H. & Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. Clinical Chemistry, 39(4), 561-577. doi: 10.1093/CLINCHEM/39.4.561

APPENDIX

<u>Data Set #1</u> (a:b:c:d = 10:90:90:810)
Sensitivity = 10%, Specificity = 90%

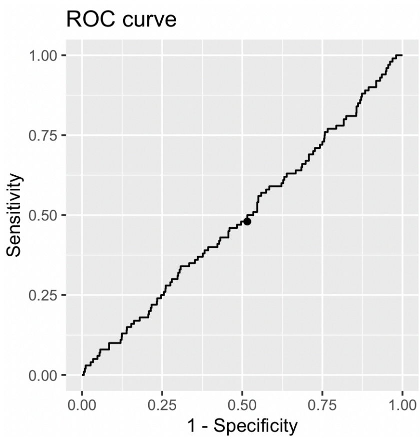
mean(data1$Score_1)
[1] -0.3527431

> sd(data1$Score_1)
[1] 0.8378757

<u>Unal</u>

AUC	n	n_pos	n_neg
0.493	1000	100	900

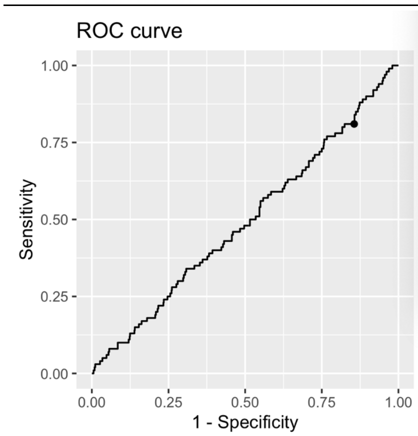
optimal_cutpoint	abs_d_sens_spec	acc	sensitivity	specificity	tp	fn	fp	tn
-0.324	0.0044	0.484	0.48	0.4844	48	52	464	436



<u>Youden</u>

AUC	n	n_pos	n_neg
0.493	1000	100	900

optimal_cutpoint	youden	acc	sensitivity	specificity	tp	fn	fp	tn
-1.2009	-0.0456	0.211	0.81	0.1444	81	19	770	130



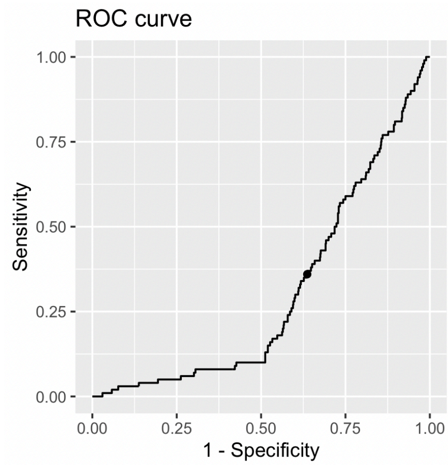
Data Set #2 (a:b:c:d = 10:90:450:450)
Sensitivity = 10%, Specificity = 50%

```
> mean(data1$Score_2)
[1] 0.2312934
```

```
> sd(data1$Score_2)
[1] 1.049908
```

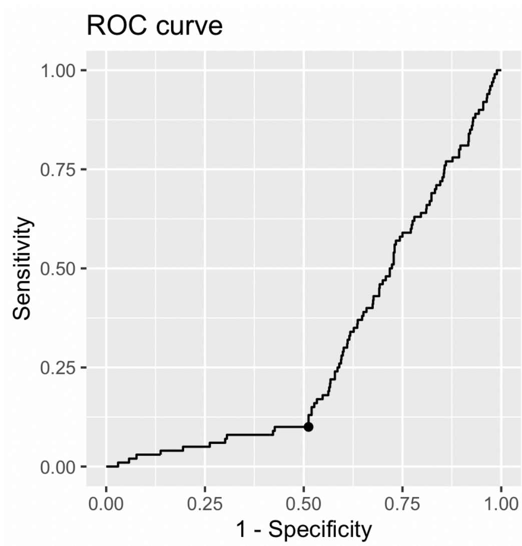
Unal

```
   AUC    n n_pos n_neg
0.3039 1000   100   900

optimal_cutpoint abs_d_sens_spec   acc sensitivity specificity tp fn  fp  tn
         -0.0105          0.0033 0.363        0.36      0.3633 36 64 573 327
```



Youden

```
   AUC    n n_pos n_neg
0.3039 1000   100   900

optimal_cutpoint  youden   acc sensitivity specificity tp fn  fp  tn
           0.451 -0.4122 0.449         0.1      0.4878 10 90 461 439
```



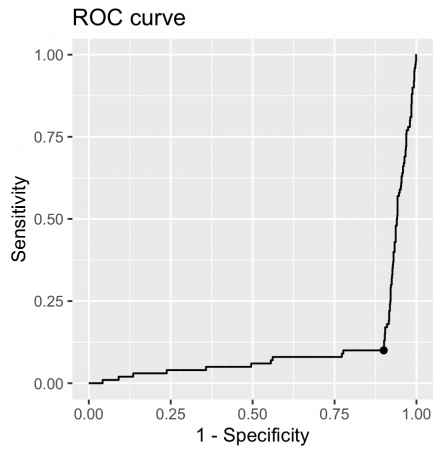
Data Set #3 (a:b:c:d = 10:90:810:90)
Sensitivity = 10%, Specificity = 10%

```
> mean(data1$Score_3)
[1] 0.8254608
```

```
> sd(data1$Score_3)
[1] 0.8433006
```

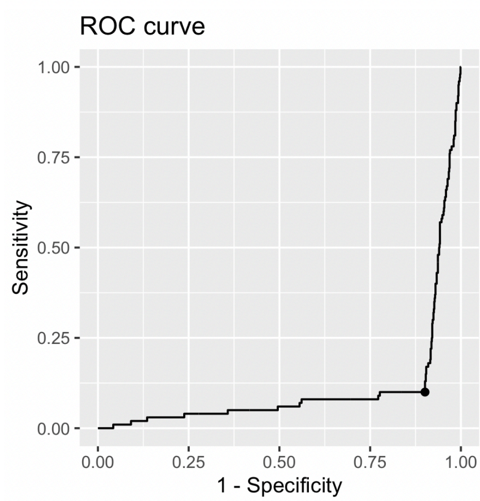
Unal

```
   AUC    n n_pos n_neg
0.1053 1000   100   900

optimal_cutpoint abs_d_sens_spec acc sensitivity specificity tp fn  fp tn
          0.5023               0 0.1         0.1         0.1 10 90 810 90
```



Youden

```
   AUC    n n_pos n_neg
0.1053 1000   100   900

optimal_cutpoint  youden   acc sensitivity specificity tp fn  fp tn
          0.4599 -0.8011 0.099         0.1      0.0989 10 90 811 89
```



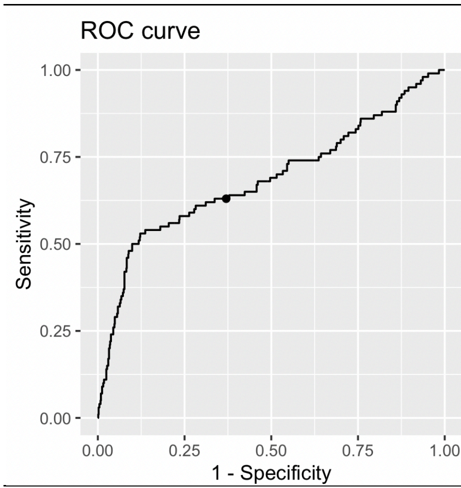
Data Set #4 (a:b:c:d = 50:50:90:810)
Sensitivity = 50%, Specificity = 90%

```
> mean(data1$Score_4)
[1] -0.28105
```

```
> sd(data1$Score_4)
[1] 0.8805964
```

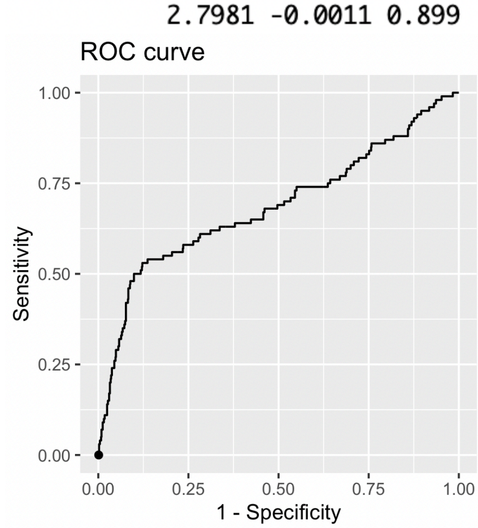
Unal

```
   AUC    n n_pos n_neg
0.6857 1000   100   900

optimal_cutpoint abs_d_sens_spec  acc sensitivity specificity tp fn  fp  tn
         -0.0394               0 0.63        0.63        0.63 63 37 333 567
```



Youden

```
   AUC    n n_pos n_neg
0.6857 1000   100   900

optimal_cutpoint  youden   acc sensitivity specificity tp  fn fp  tn
          2.7981 -0.0011 0.899           0      0.9989  0 100  1 899
```



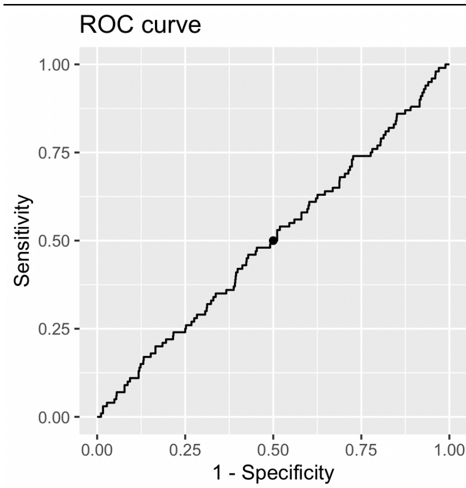
Data Set #5 (a:b:c:d = 50:50:450:450)
Sensitivity = 50%, Specificity = 50%

```
> mean(data1$Score_5)
[1] 0.2914549
```

```
> sd(data1$Score_5)
[1] 1.056868
```

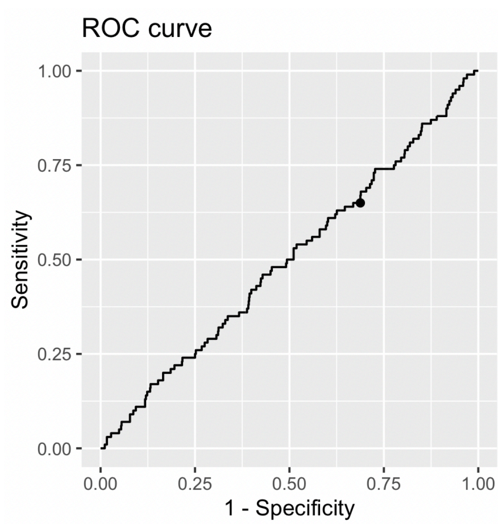
Unal

```
   AUC    n n_pos n_neg
0.4991 1000   100   900

optimal_cutpoint abs_d_sens_spec acc sensitivity specificity tp fn  fp  tn
          0.5023               0 0.5         0.5         0.5 50 50 450 450
```



Youden

```
   AUC    n n_pos n_neg
0.4991 1000   100   900

optimal_cutpoint  youden   acc sensitivity specificity tp fn  fp  tn
         -0.2164 -0.0378 0.346        0.65      0.3122 65 35 619 281
```



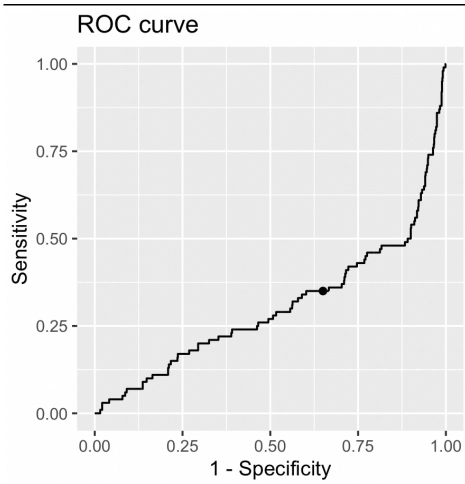
<u>Data Set #6</u> (a:b:c:d = 50:50:810:90)
Sensitivity = 50%, Specificity = 10%

```
> mean(data1$Score_6)
[1] 0.8958743
```

```
> sd(data1$Score_6)
[1] 0.7913209
```

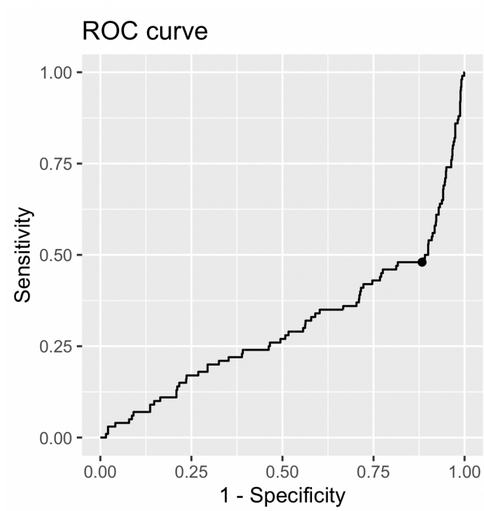
<u>Unal</u>

```
   AUC    n n_pos n_neg
0.3038 1000   100   900

optimal_cutpoint abs_d_sens_spec  acc sensitivity specificity tp fn  fp  tn
          0.7491               0 0.35        0.35        0.35 35 65 585 315
```



<u>Youden</u>

```
   AUC    n n_pos n_neg
0.3038 1000   100   900

optimal_cutpoint  youden   acc sensitivity specificity tp fn  fp  tn
           0.523 -0.4033 0.153        0.48      0.1167 48 52 795 105
```



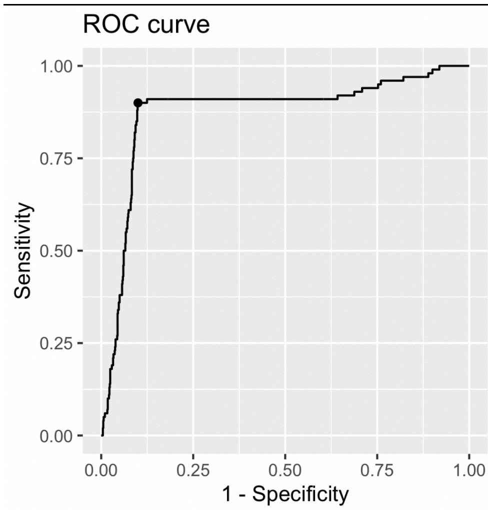
<u>Data Set #7</u> (a:b:c:d = 90:10:90:810)
Sensitivity = 90%, Specificity = 90%

> mean(data1$Score_7)
[1] -0.2010222

> sd(data1$Score_7)
[1] 0.9382004

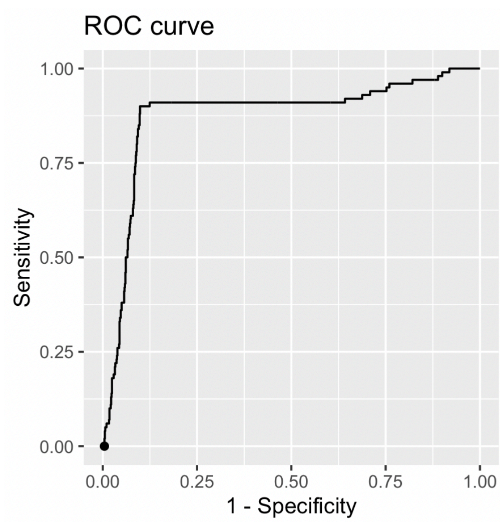
<u>Unal</u>

```
   AUC    n n_pos n_neg
0.8763 1000   100   900

optimal_cutpoint abs_d_sens_spec acc sensitivity specificity tp fn fp  tn
          0.5038               0 0.9         0.9         0.9 90 10 90 810
```



<u>Youden</u>

```
   AUC    n n_pos n_neg
0.8763 1000   100   900

optimal_cutpoint  youden   acc sensitivity specificity tp  fn fp  tn
           2.726 -0.0044 0.896           0      0.9956  0 100  4 896
```



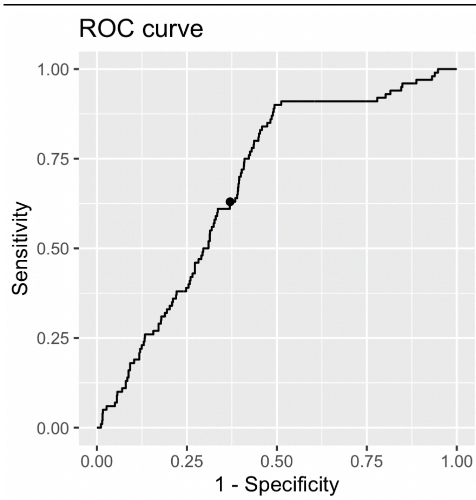
Data Set #8 (a:b:c:d = 90:10:450:450)
Sensitivity = 90%, Specificity = 50%

> mean(data1$Score_8)
[1] 0.3543923

> sd(data1$Score_8)
[1] 1.050842

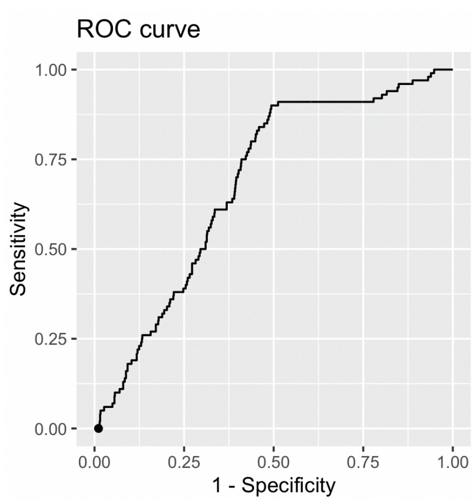
Unal

```
   AUC    n n_pos n_neg
0.6824 1000   100   900

optimal_cutpoint abs_d_sens_spec  acc sensitivity specificity tp fn  fp  tn
          0.7235               0 0.63        0.63        0.63 63 37 333 567
```



Youden

```
   AUC    n n_pos n_neg
0.6824 1000   100   900

optimal_cutpoint  youden  acc sensitivity specificity tp  fn fp  tn
          2.7177 -0.0111 0.89           0      0.9889  0 100 10 890
```



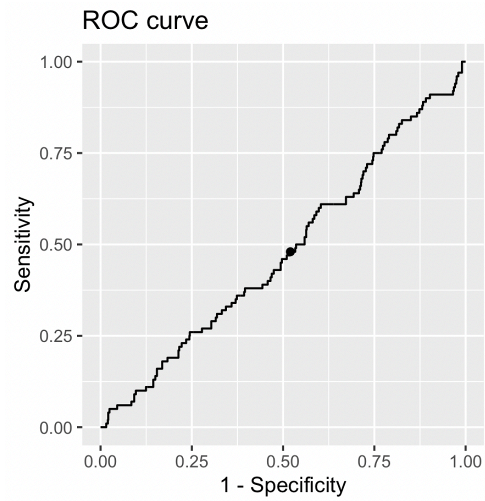
Data Set #9 (a:b:c:d = 90:10:810:90)
Sensitivity = 90%, Specificity = 10%

```
> mean(data1$Score_9)
[1] 0.9693953
```

```
> sd(data1$Score_9)
[1] 0.7321414
```

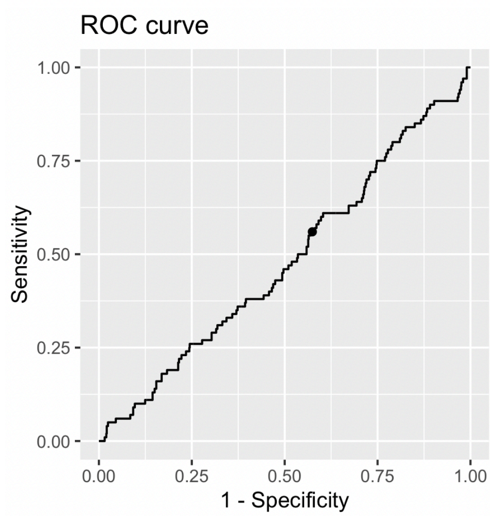
Unal

```
   AUC    n n_pos n_neg
0.4806 1000   100   900

optimal_cutpoint abs_d_sens_spec  acc sensitivity specificity tp fn  fp  tn
           0.915               0 0.48        0.48        0.48 48 52 468 432
```



Youden

```
   AUC    n n_pos n_neg
0.4806 1000   100   900

optimal_cutpoint  youden   acc sensitivity specificity tp fn  fp  tn
          0.8381 -0.0144 0.439        0.56      0.4256 56 44 517 383
```



Data Set #10 (a:b:c:d = 50:450:50:450)
Sensitivity = 10%, Specificity = 90%

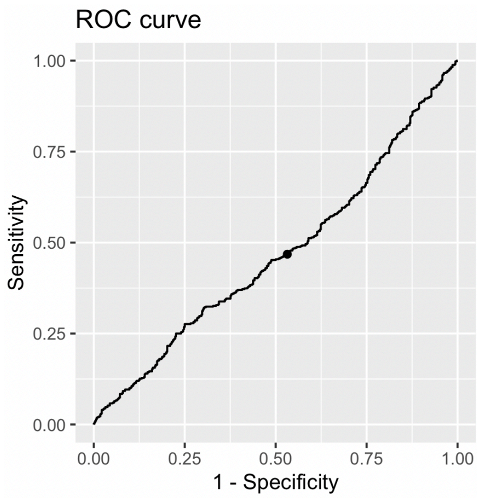
> mean(data1$Score_10)
[1] -0.3527431

> sd(data1$Score_10)
[1] 0.8378757

Unal

AUC	n	n_pos	n_neg
0.4681	1000	500	500

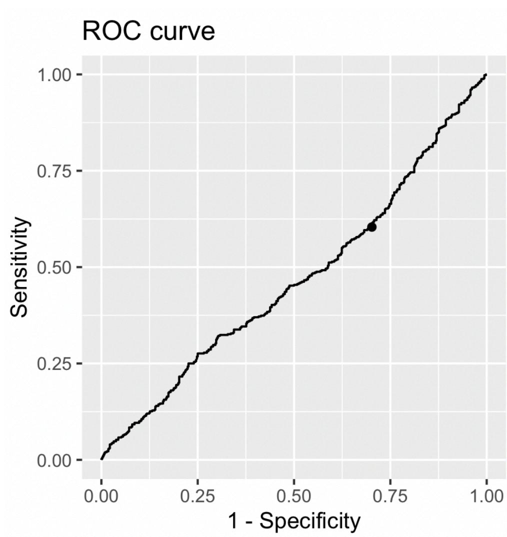
optimal_cutpoint	abs_d_sens_spec	acc	sensitivity	specificity	tp	fn	fp
-0.2954	0	0.468	0.468	0.468	234	266	266



Youden

AUC	n	n_pos	n_neg
0.4681	1000	500	500

optimal_cutpoint	youden	acc	sensitivity	specificity	tp	fn	fp	tn
-0.6227	-0.098	0.451	0.604	0.298	302	198	351	149



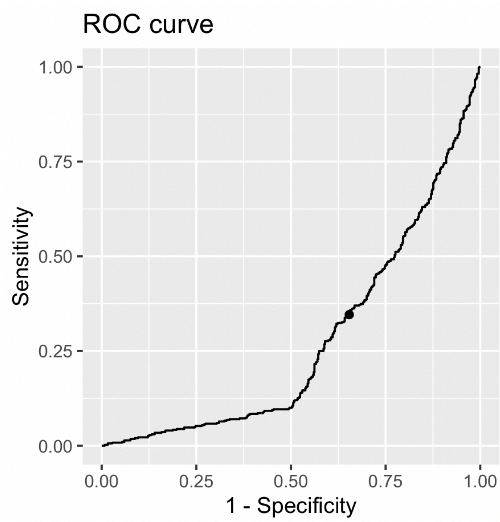
<u>Data Set #11</u> (a:b:c:d = 50:450:250:250)
Sensitivity = 10%, Specificity = 50%

```
> mean(data1$Score_11)
[1] -0.01390235
```

```
> sd(data1$Score_11)
[1] 0.9990552
```

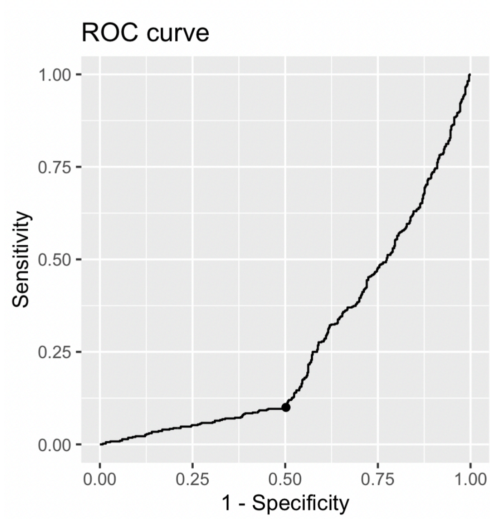
<u>Unal</u>

```
   AUC    n n_pos n_neg
0.2748 1000   500   500

optimal_cutpoint abs_d_sens_spec   acc sensitivity specificity  tp  fn  fp
         -0.0202               0 0.346       0.346       0.346 173 327 327
```



<u>Youden</u>

```
   AUC    n n_pos n_neg
0.2748 1000   500   500

optimal_cutpoint youden   acc sensitivity specificity tp  fn  fp  tn
          0.4935 -0.402 0.299         0.1       0.498 50 450 251 249
```



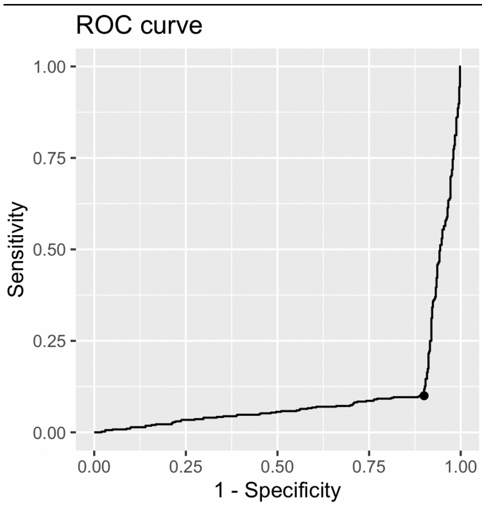
Data Set #12 (a:b:c:d = 50:450:450:50)
Sensitivity = 10%, Specificity = 10%

```
> mean(data1$Score_12)
[1] 0.2914549
```

```
> sd(data1$Score_12)
[1] 1.056868
```

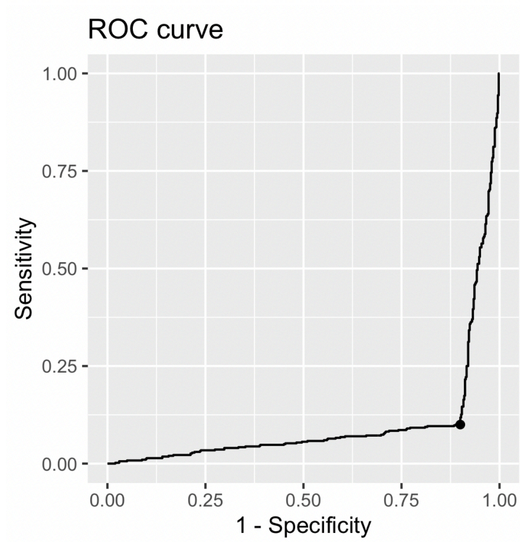
Unal

```
   AUC    n n_pos n_neg
0.0986 1000   500   500

optimal_cutpoint abs_d_sens_spec acc sensitivity specificity tp  fn  fp tn
          0.5023               0 0.1         0.1         0.1 50 450 450 50
```



Youden

```
   AUC    n n_pos n_neg
0.0986 1000   500   500

optimal_cutpoint youden acc sensitivity specificity tp  fn  fp tn
          0.5023   -0.8 0.1         0.1         0.1 50 450 450 50
```



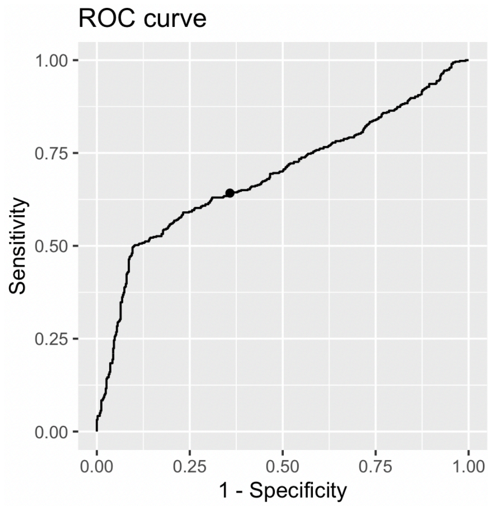
<u>Data Set #13</u> (a:b:c:d = 250:250:50:450)
Sensitivity = 50%, Specificity = 90%

```
> mean(data1$Score_13)
[1] -0.01390235
```

```
> sd(data1$Score_13
[1] 0.9990552
```

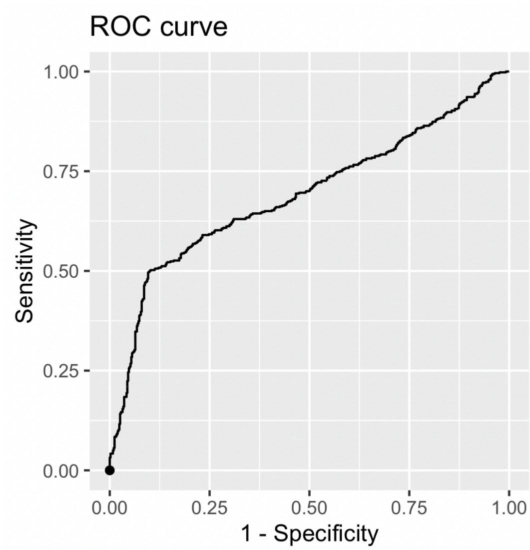
<u>Unal</u>

```
   AUC    n n_pos n_neg
0.6908 1000   500   500

optimal_cutpoint abs_d_sens_spec   acc sensitivity specificity  tp  fn  fp
         -0.0202               0 0.642       0.642       0.642 321 179 179
  tn
```



<u>Youden</u>

```
   AUC    n n_pos n_neg
0.6908 1000   500   500

optimal_cutpoint youden acc sensitivity specificity tp  fn fp  tn
             Inf      0 0.5           0           1  0 500  0 500
```



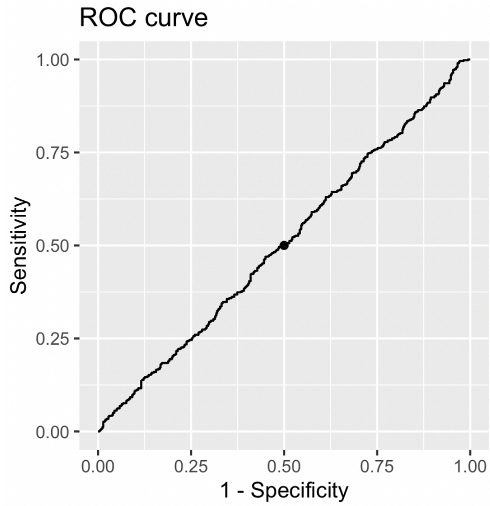
Data Set #14 (a:b:c:d = 250:250:250:250)
Sensitivity = 50%, Specificity = 50%

```
> mean(data1$Score_14)
[1] 0.2914549
```

```
> sd(data1$Score_14)
[1] 1.056868
```

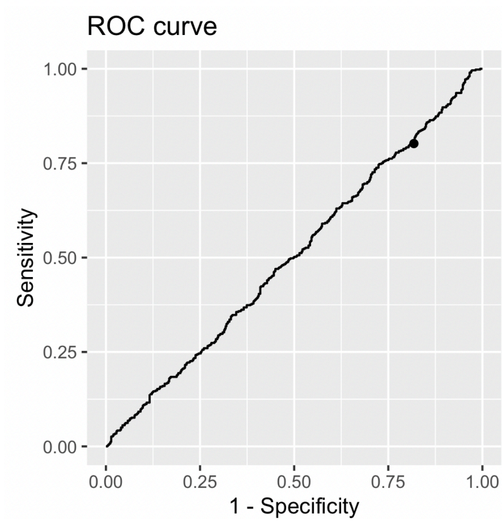
Unal

```
   AUC    n n_pos n_neg
0.5035 1000   500   500

optimal_cutpoint abs_d_sens_spec acc sensitivity specificity  tp  fn  fp  tn
          0.5023               0 0.5         0.5         0.5 250 250 250 250
```



Youden

```
   AUC    n n_pos n_neg
0.5035 1000   500   500

optimal_cutpoint youden   acc sensitivity specificity  tp fn  fp tn
         -0.7049 -0.016 0.492       0.802       0.182 401 99 409 91
```



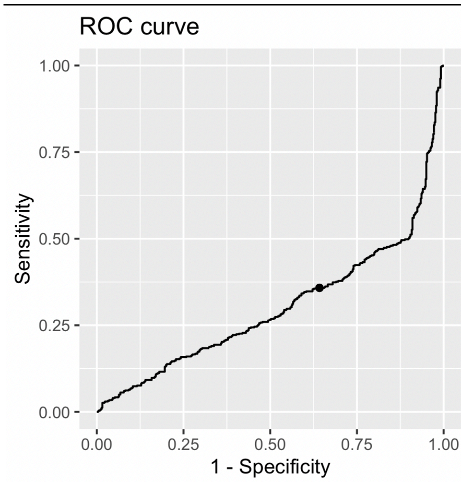
Data Set #15 (a:b:c:d = 250:250:450:50)
Sensitivity = 50%, Specificity = 10%

> mean(data1$Score_15)
[1] 0.6293439

> sd(data1$Score_15)
[1] 0.9532248

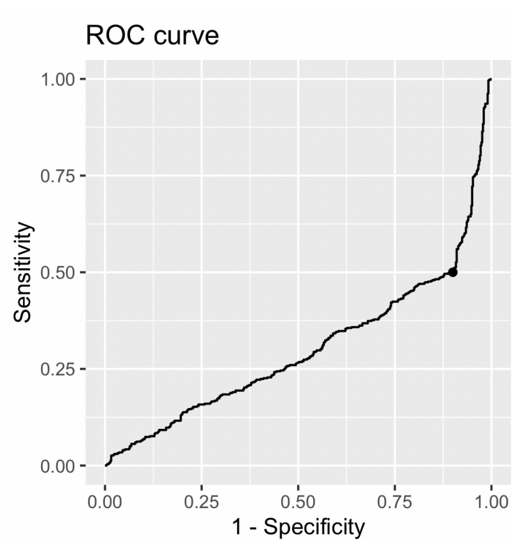
Unal

```
   AUC    n n_pos n_neg
0.3012 1000   500   500

optimal_cutpoint abs_d_sens_spec   acc sensitivity specificity  tp  fn  fp
          0.7582               0 0.358       0.358       0.358 179 321 321
```



Youden

```
   AUC    n n_pos n_neg
0.3012 1000   500   500

optimal_cutpoint youden acc sensitivity specificity  tp  fn  fp tn
          0.5023   -0.4 0.3         0.5         0.1 250 250 450 50
```



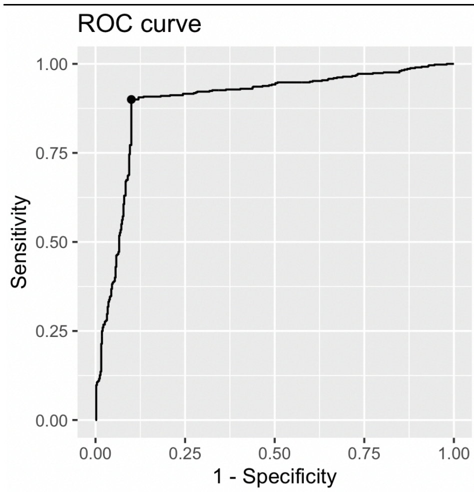
Data Set #16 (a:b:c:d = 450:50:50:450)
Sensitivity = 90%, Specificity = 90%

> mean(data1$Score_16)
[1] 0.2914549

> sd(data1$Score_16)
[1] 1.056868

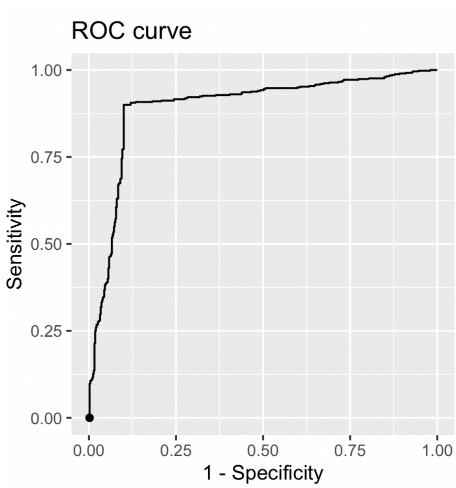
Unal

```
   AUC    n n_pos n_neg
 0.894 1000   500   500

 optimal_cutpoint abs_d_sens_spec acc sensitivity specificity  tp fn fp  tn
           0.5023               0 0.9         0.9         0.9 450 50 50 450
```



Youden

```
   AUC    n n_pos n_neg
 0.894 1000   500   500

 optimal_cutpoint youden   acc sensitivity specificity tp  fn fp  tn
           3.3334 -0.002 0.499           0       0.998  0 500  1 499
```



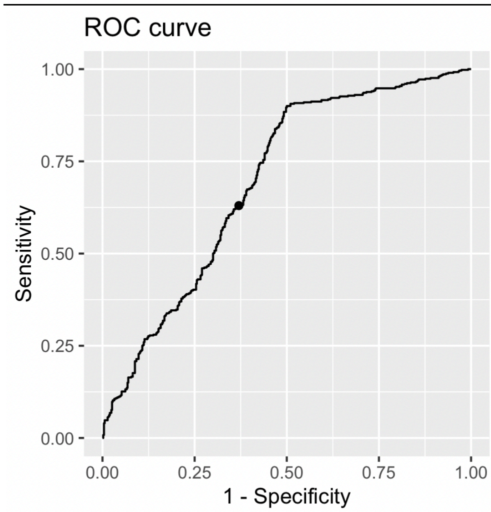
Data Set #17 (a:b:c:d = 450:50:250:250)
Sensitivity = 90%, Specificity = 50%

> mean(data1$Score_17)
[1] 0.6293439

> sd(data1$Score_17)
[1] 0.9532248

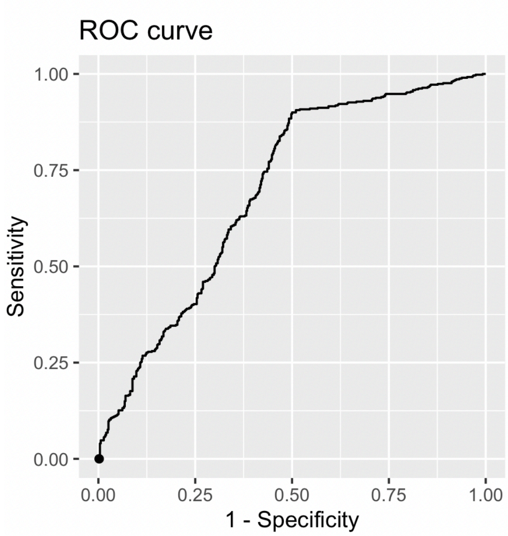
Unal

```
   AUC    n n_pos n_neg
 0.6951 1000   500   500

 optimal_cutpoint abs_d_sens_spec  acc sensitivity specificity  tp  fn  fp  tn
           0.7582               0 0.63        0.63        0.63 315 185 185 315
```



Youden

```
   AUC    n n_pos n_neg
 0.6951 1000   500   500

 optimal_cutpoint youden   acc sensitivity specificity tp  fn fp  tn
           3.3334 -0.002 0.499           0       0.998  0 500  1 499
```



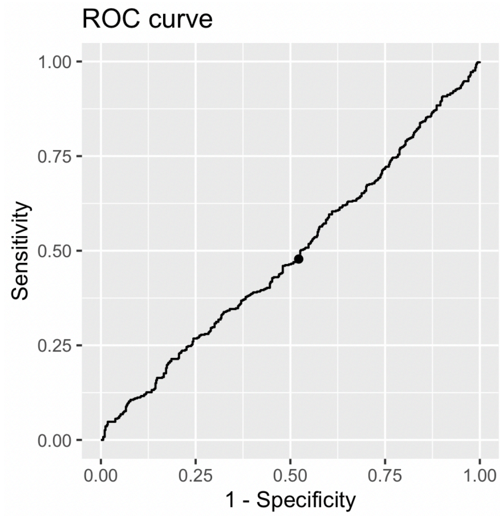
<u>Data Set #18</u> (a:b:c:d = 450:50:450:50)
Sensitivity = 90%, Specificity = 10%

```
> mean(data1$Score_18)
[1] 0.9693953
```

```
> sd(data1$Score_18)
[1] 0.7321414
```

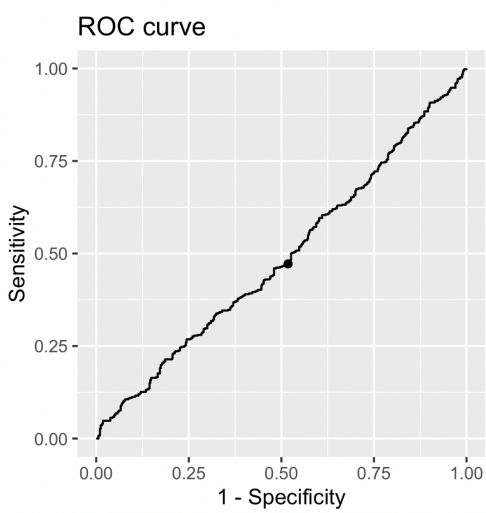
<u>Unal</u>

```
   AUC    n n_pos n_neg
0.4895 1000   500   500

optimal_cutpoint abs_d_sens_spec   acc sensitivity specificity  tp  fn  fp
          0.9307               0 0.478       0.478       0.478 239 261 261
```



<u>Youden</u>

```
   AUC    n n_pos n_neg
0.4895 1000   500   500

optimal_cutpoint youden   acc sensitivity specificity  tp  fn  fp  tn
          0.9411 -0.046 0.477       0.472       0.482 236 264 259 241
```



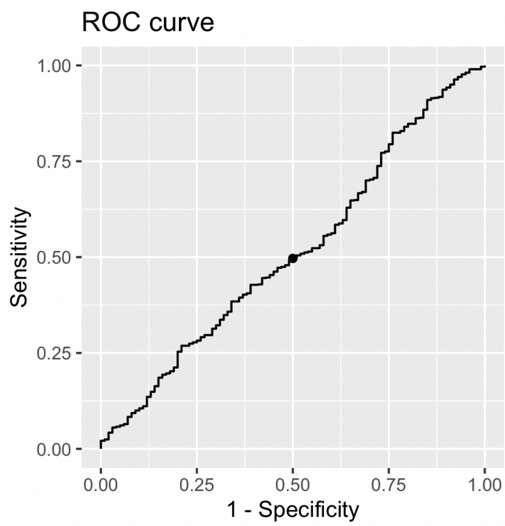
Data Set #19 (a:b:c:d = 90:810:10:90)
Sensitivity = 10%, Specificity = 90%

```
> mean(data1$Score_19)
[1] -0.3527431
```

```
> sd(data1$Score_19)
[1] 0.8378757
```

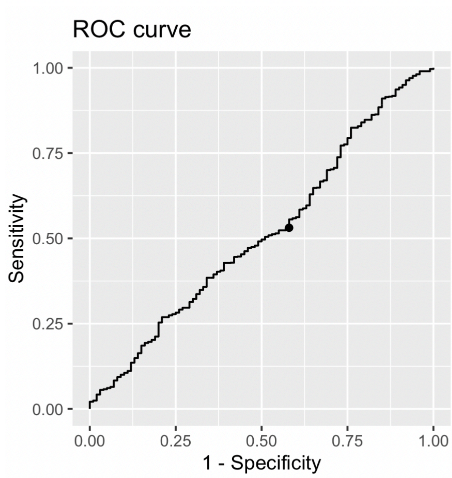
Unal

```
   AUC    n n_pos n_neg
0.5141 1000   900   100

optimal_cutpoint abs_d_sens_spec   acc sensitivity specificity  tp  fn fp tn
         -0.2927          0.0033 0.497      0.4967         0.5 447 453 50 50
```



Youden

```
   AUC    n n_pos n_neg
0.5141 1000   900   100

optimal_cutpoint  youden  acc sensitivity specificity  tp  fn fp tn
         -0.3689 -0.0489 0.52      0.5311        0.42 478 422 58 42
```



Data Set #20 (a:b:c:d = 90:810:50:50)
Sensitivity = 10%, Specificity = 50%

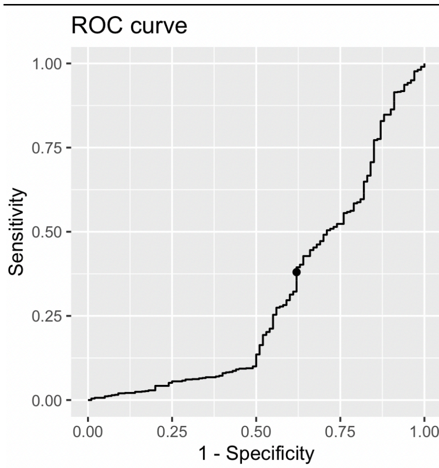
> mean(data1$Score_20)
[1] -0.28105

> sd(data1$Score_20)
[1] 0.8805964

Unal

AUC	n	n_pos	n_neg
0.3081	1000	900	100

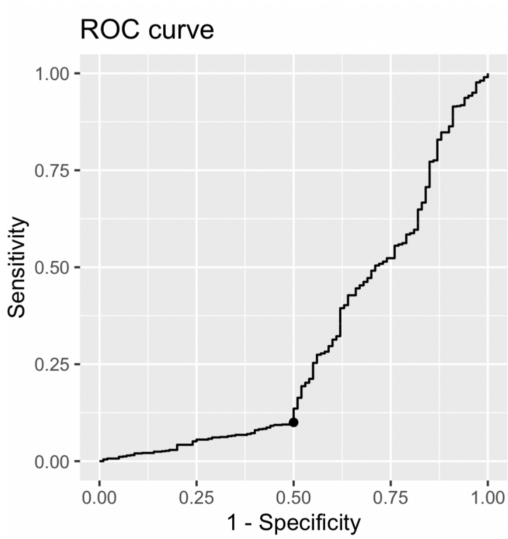
optimal_cutpoint	abs_d_sens_spec	acc	sensitivity	specificity	tp	fn	fp	tn
-0.0626	0	0.38	0.38	0.38	342	558	62	38



Youden

AUC	n	n_pos	n_neg
0.3081	1000	900	100

optimal_cutpoint	youden	acc	sensitivity	specificity	tp	fn	fp	tn
0.5038	-0.4	0.14	0.1	0.5	90	810	50	50



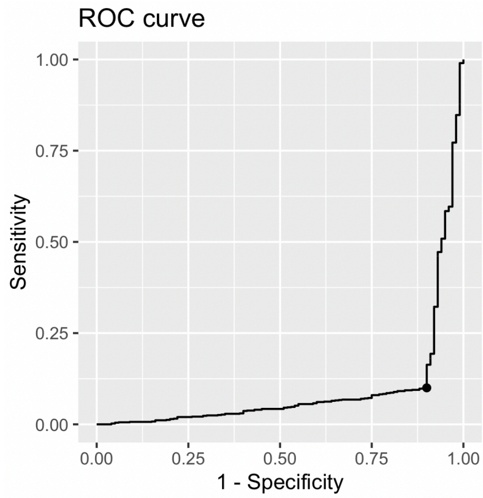
Data Set #21 (a:b:c:d = 90:810:90:10)
Sensitivity = 10%, Specificity = 10%

```
> mean(data1$Score_21)
[1] -0.2010222
```

```
> sd(data1$Score_21)
[1] 0.9382004
```

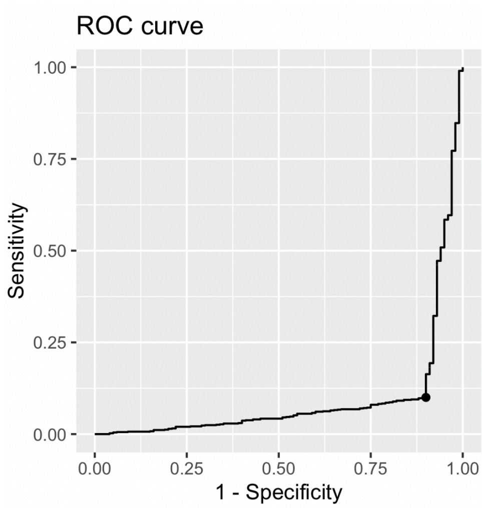
Unal

```
   AUC    n n_pos n_neg
0.0929 1000   900   100

optimal_cutpoint abs_d_sens_spec acc sensitivity specificity tp  fn fp tn
          0.5038               0 0.1         0.1         0.1 90 810 90 10
```



Youden

```
   AUC    n n_pos n_neg
0.0929 1000   900   100

optimal_cutpoint youden acc sensitivity specificity tp  fn fp tn
          0.5038   -0.8 0.1         0.1         0.1 90 810 90 10
```



<u>Data Set #22</u> (a:b:c:d = 450:450:10:90)
Sensitivity = 50%, Specificity = 90%

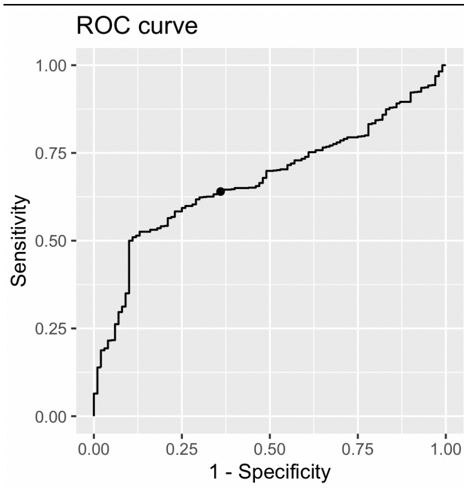
```
> mean(data1$Score_22)
[1] 0.2312934
```

```
> sd(data1$Score_22)
[1] 1.049908
```

<u>Unal</u>

```
   AUC    n n_pos n_neg
0.6729 1000   900   100
```

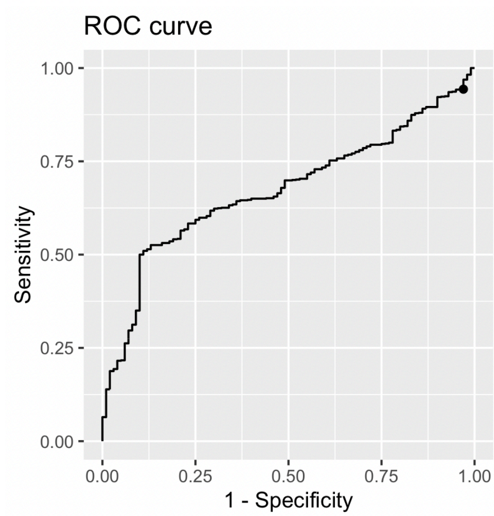
```
optimal_cutpoint abs_d_sens_spec  acc sensitivity specificity  tp  fn fp tn
         -0.0249               0 0.64        0.64        0.64 576 324 36 64
```



<u>Youden</u>

```
   AUC    n n_pos n_neg
0.6729 1000   900   100
```

```
optimal_cutpoint  youden   acc sensitivity specificity  tp fn fp tn
         -1.5469 -0.0267 0.852      0.9433        0.03 849 51 97  3
```



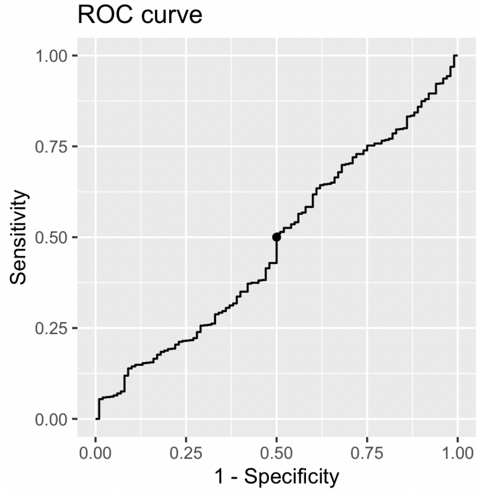
Data Set #23 (a:b:c:d = 450:450:50:50)
Sensitivity = 50%, Specificity = 50%

```
> mean(data1$Score_23)
[1] 0.2914549
```

```
> sd(data1$Score_23)
[1] 1.056868
```

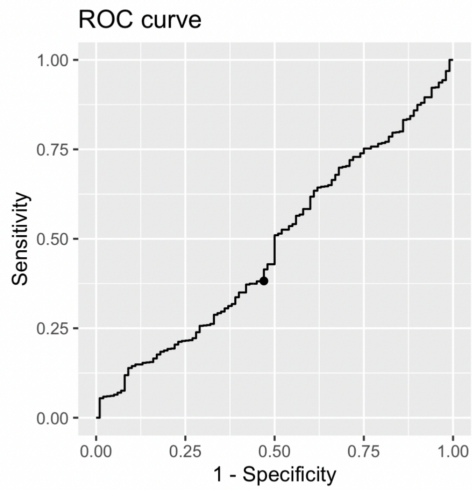
Unal

```
    AUC    n n_pos n_neg
 0.4804 1000   900   100

 optimal_cutpoint abs_d_sens_spec acc sensitivity specificity  tp  fn fp tn
           0.5023               0 0.5         0.5         0.5 450 450 50 50
```



Youden

```
    AUC    n n_pos n_neg
 0.4804 1000   900   100

 optimal_cutpoint  youden   acc sensitivity specificity  tp  fn fp tn
           0.6772 -0.0878 0.397      0.3822        0.53 344 556 47 53
```




```
> mean(data1$Score_24)
[1] 0.3543923
```

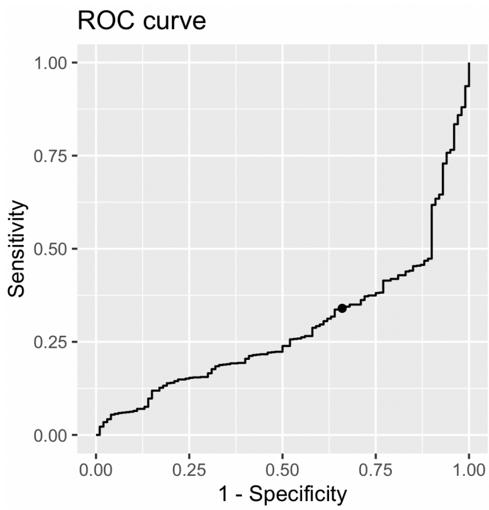
```
> sd(data1$Score_24)
[1] 1.050842
```

Data Set #24 (a:b:c:d = 450:450:90:10)
Sensitivity = 50%, Specificity = 10%

Unal

AUC	n	n_pos	n_neg
0.2889	1000	900	100

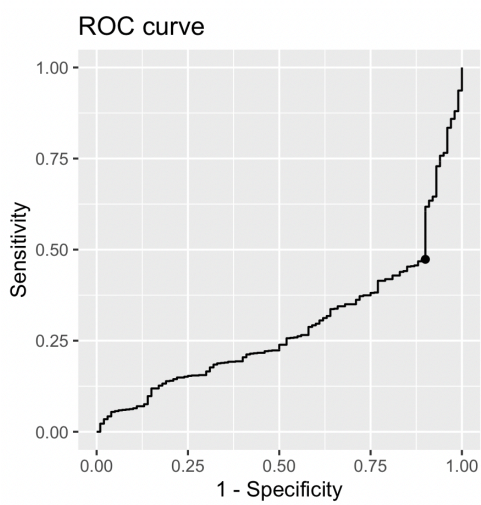
optimal_cutpoint	abs_d_sens_spec	acc	sensitivity	specificity	tp	fn	fp	tn
0.7851	0	0.34	0.34	0.34	306	594	66	34



Youden

AUC	n	n_pos	n_neg
0.2889	1000	900	100

optimal_cutpoint	youden	acc	sensitivity	specificity	tp	fn	fp	tn
0.5437	-0.4267	0.436	0.4733	0.1	426	474	90	10



Data Set #25 (a:b:c:d = 810:90:10:90)
Sensitivity = 90%, Specificity = 90%

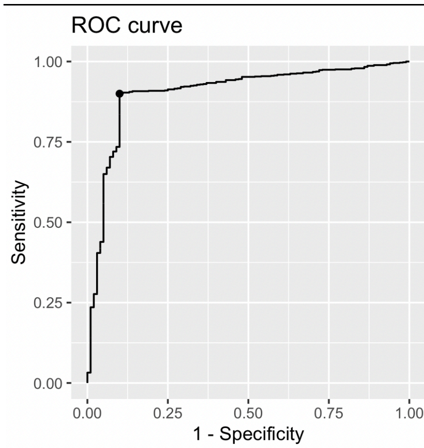
> mean(data1$Score_25)
[1] 0.8254608

> sd(data1$Score_25)
[1] 0.8433006

Unal

AUC	n	n_pos	n_neg
0.9042	1000	900	100

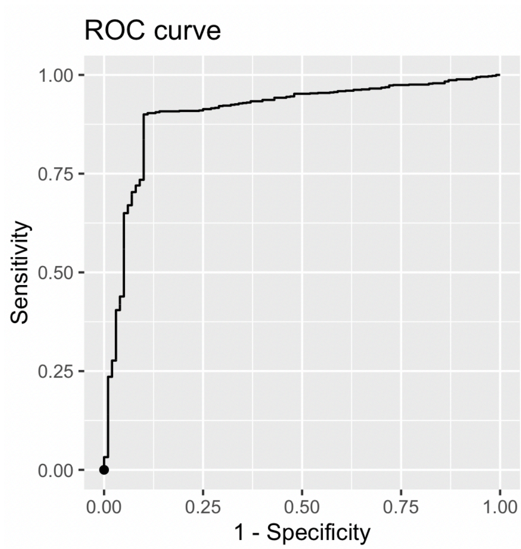
optimal_cutpoint	abs_d_sens_spec	acc	sensitivity	specificity	tp	fn	fp	tn
0.5023	0	0.9	0.9	0.9	810	90	10	90



Youden

AUC	n	n_pos	n_neg
0.9042	1000	900	100

optimal_cutpoint	youden	acc	sensitivity	specificity	tp	fn	fp	tn
Inf	0	0.1	0	1	0	900	0	100



Data Set #26 (a:b:c:d = 810:90:50:50)
Sensitivity = 90%, Specificity = 50%

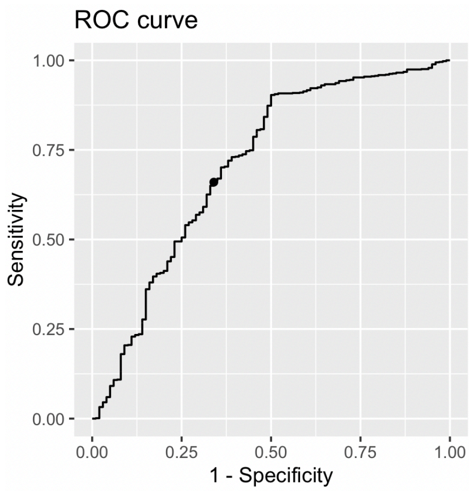
> mean(data1$Score_26)
[1] 0.8958743

> sd(data1$Score_26
[1] 0.7913209

Unal

AUC	n	n_pos	n_neg
0.708	1000	900	100

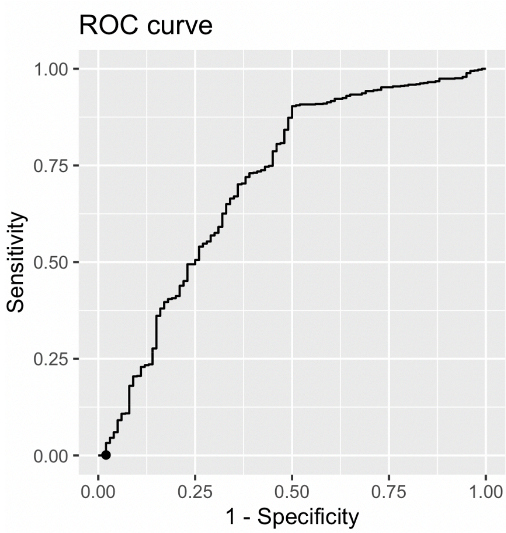
optimal_cutpoint	abs_d_sens_spec	acc	sensitivity	specificity	tp	fn	fp	tn
0.7373	0	0.66	0.66	0.66	594	306	34	66



Youden

AUC	n	n_pos	n_neg
0.708	1000	900	100

optimal_cutpoint	youden	acc	sensitivity	specificity	tp	fn	fp	tn
3.1512	-0.0189	0.099	0.0011	0.98	1	899	2	98



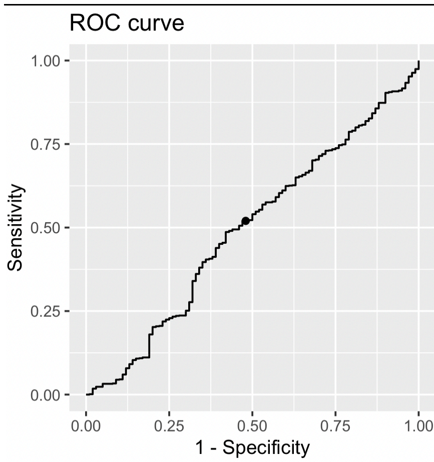
<u>Data Set #27</u> (a:b:c:d = 810:90:90:10)
Sensitivity = 90%, Specificity = 10%

```
> mean(data1$Score_27)
[1] 0.9693953
```

```
> sd(data1$Score_27)
[1] 0.7321414
```

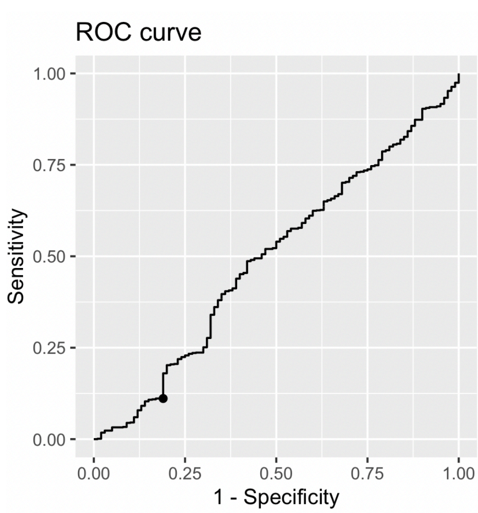
Unal

```
   AUC    n n_pos n_neg
0.4937 1000   900   100

optimal_cutpoint abs_d_sens_spec  acc sensitivity specificity  tp  fn fp tn
           0.915               0 0.52        0.52        0.52 468 432 48 52
```



Youden

```
   AUC    n n_pos n_neg
0.4937 1000   900   100

optimal_cutpoint  youden   acc sensitivity specificity  tp  fn fp tn
          1.7451 -0.0789 0.181      0.1111        0.81 100 800 19 81
```



<u>Normality Tests</u>

```
> shapiro.test(data1$Score_1)

	Shapiro-Wilk normality test

data:  data1$Score_1
W = 0.98799, p-value = 2.657e-07

> shapiro.test(data1$Score_2)

	Shapiro-Wilk normality test

data:  data1$Score_2
W = 0.99391, p-value = 0.000432

> shapiro.test(data1$Score_3)

	Shapiro-Wilk normality test

data:  data1$Score_3
W = 0.93815, p-value < 2.2e-16

> shapiro.test(data1$Score_4)

	Shapiro-Wilk normality test

data:  data1$Score_4
W = 0.99166, p-value = 1.972e-05

> shapiro.test(data1$Score_13)

	Shapiro-Wilk normality test

data:  data1$Score_13
W = 0.99827, p-value = 0.4144

> shapiro.test(data1$Score_14)

	Shapiro-Wilk normality test

data:  data1$Score_14
W = 0.99032, p-value = 3.716e-06

> shapiro.test(data1$Score_15)

	Shapiro-Wilk normality test

data:  data1$Score_15
W = 0.96032, p-value = 7.691e-16

> shapiro.test(data1$Score_16)

	Shapiro-Wilk normality test

data:  data1$Score_16
W = 0.99032, p-value = 3.716e-06

> shapiro.test(data1$Score_25)

	Shapiro-Wilk normality test

data:  data1$Score_25
W = 0.93815, p-value < 2.2e-16

> shapiro.test(data1$Score_5)

	Shapiro-Wilk normality test

data:  data1$Score_5
W = 0.99032, p-value = 3.716e-06

> shapiro.test(data1$Score_6)

	Shapiro-Wilk normality test

data:  data1$Score_6
W = 0.93291, p-value < 2.2e-16

> shapiro.test(data1$Score_7)

	Shapiro-Wilk normality test

data:  data1$Score_7
W = 0.99137, p-value = 1.367e-05

> shapiro.test(data1$Score_8)

	Shapiro-Wilk normality test

data:  data1$Score_8
W = 0.98593, p-value = 3.207e-08

> shapiro.test(data1$Score_17)

	Shapiro-Wilk normality test

data:  data1$Score_17
W = 0.96032, p-value = 7.691e-16

> shapiro.test(data1$Score_18)

	Shapiro-Wilk normality test

data:  data1$Score_18
W = 0.92805, p-value < 2.2e-16

> shapiro.test(data1$Score_19)

	Shapiro-Wilk normality test

data:  data1$Score_19
W = 0.98799, p-value = 2.657e-07

> shapiro.test(data1$Score_20)

	Shapiro-Wilk normality test

data:  data1$Score_20
W = 0.99166, p-value = 1.972e-05

> shapiro.test(data1$Score_26)

	Shapiro-Wilk normality test

data:  data1$Score_26
W = 0.93291, p-value < 2.2e-16

> shapiro.test(data1$Score_9)

	Shapiro-Wilk normality test

data:  data1$Score_9
W = 0.92805, p-value < 2.2e-16

> shapiro.test(data1$Score_10)

	Shapiro-Wilk normality test

data:  data1$Score_10
W = 0.98799, p-value = 2.657e-07

> shapiro.test(data1$Score_11)

	Shapiro-Wilk normality test

data:  data1$Score_11
W = 0.99827, p-value = 0.4144

> shapiro.test(data1$Score_12)

	Shapiro-Wilk normality test

data:  data1$Score_12
W = 0.99032, p-value = 3.716e-06

> shapiro.test(data1$Score_21)

	Shapiro-Wilk normality test

data:  data1$Score_21
W = 0.99137, p-value = 1.367e-05

> shapiro.test(data1$Score_22)

	Shapiro-Wilk normality test

data:  data1$Score_22
W = 0.99391, p-value = 0.000432

> shapiro.test(data1$Score_23)

	Shapiro-Wilk normality test

data:  data1$Score_23
W = 0.99032, p-value = 3.716e-06

> shapiro.test(data1$Score_24)

	Shapiro-Wilk normality test

data:  data1$Score_24
W = 0.98593, p-value = 3.207e-08

> shapiro.test(data1$Score_27)

	Shapiro-Wilk normality test

data:  data1$Score_27
W = 0.92805, p-value < 2.2e-16
```

BIBLIOGRAPHY

Gregory W. McGuire

Candidate for the Degree of

Master of Science Mathematics

Thesis: INVESTIGATING CUT-POINT METHODS IN ROC ANALYSIS: A CRITIQUE OF ALTERNATIVE APPROACHES

Major Field: Mathematics

Biographical: Born on Mar. 11, 1985. Married to Lauren (Hennig) McGuire on Dec. 21, 2018. Son, Lucas McGuire, born on Feb. 14, 2021. Daughter, Sophia McGuire, born on Feb. 11, 2022.

Education: M.A.T., Secondary Mathematics; M.A., Philosophy; B.A., Psychology

Completed the requirements for the Master of Science in Mathematics, Portsmouth, Ohio in July 2022.

7/27/2022

ADVISER'S APPROVAL: Doug Darbro