



HAL
open science

Tutoriel sur les courbes ROC et leur création grâce au site Internet easyROC

Loic Desquilbet

► **To cite this version:**

Loic Desquilbet. Tutoriel sur les courbes ROC et leur création grâce au site Internet easyROC. 2020.
hal-02870055v1

HAL Id: hal-02870055

<https://hal.science/hal-02870055v1>

Preprint submitted on 16 Jun 2020 (v1), last revised 22 Mar 2023 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Tutoriel sur les courbes ROC et leur création grâce au site Internet easyROC

Loïc Desquilbet, PhD en Santé Publique

Professeur en Biostatistique et en Epidémiologie Clinique
Département des Sciences Biologiques et Pharmaceutiques
Ecole nationale vétérinaire d'Alfort

Contrat de diffusion



Cette œuvre est mise à disposition selon les termes de la [Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International](https://creativecommons.org/licenses/by-nc-nd/4.0/) (BY NC ND 4.0). Le résumé de la licence se trouve ici : <https://creativecommons.org/licenses/by-nc-nd/3.0/fr/>.

Attribution — Vous devez créditer l'Œuvre, intégrer un lien vers la licence et indiquer si des modifications ont été effectuées à l'Œuvre. Vous devez indiquer ces informations par tous les moyens raisonnables, sans toutefois suggérer que l'Offrant vous soutient ou soutient la façon dont vous avez utilisé son Œuvre.

Pas d'Utilisation Commerciale — Vous n'êtes pas autorisé à faire un usage commercial de cette Œuvre, tout ou partie du matériel la composant.

Pas de modifications — Dans le cas où vous effectuez un remix, que vous transformez, ou créez à partir du matériel composant l'Œuvre originale, vous n'êtes pas autorisé à distribuer ou mettre à disposition l'Œuvre modifiée.

Table des matières

Contrat de diffusion	2
I. Objectifs du tutoriel.....	4
II. Un peu de théorie sur les courbes ROC.....	4
A. La courbe ROC	4
1. Contexte et objectif de réalisation d'une courbe ROC.....	4
2. La construction de la courbe ROC	5
3. L'aire sous la courbe ROC.....	6
B. Les méthodes d'identification du seuil optimal	7
1. Introduction.....	7
2. L'indice de Youden	7
3. Le point le plus proche du coin supérieur gauche de la courbe ROC.....	9
4. L'indice de Youden généralisé.....	9
C. Indicateurs statistiques de la qualité d'un test diagnostique une fois le seuil optimal identifié	11
1. Valeurs prédictives positive et négative.....	11
2. Rapports de vraisemblance positif et négatif	12
III. Utiliser le site Internet easyROC pour dresser une ou plusieurs courbes ROC	14
A. Fichier de données	14
1. Préparation du fichier de données avant importation dans le site Internet easyROC	14
2. Présentation du fichier de données fictives utilisé	15
B. Importation des données dans le site Internet easyROC.....	15
C. Courbe ROC avec une seule variable quantitative et indicateurs statistiques	16
1. Obtention de la courbe ROC.....	16
2. Indicateurs statistiques de la courbe ROC.....	17
3. Liste des valeurs de la sensibilité et de la spécificité	18
D. Seuil optimal, indicateurs statistiques et représentations graphiques de la qualité du test diagnostique.....	19
1. Identification du seuil optimal.....	19
2. Indicateurs statistiques de la qualité du test de diagnostic	21
3. Représentations graphiques de la qualité du test diagnostique.....	23
E. Comparaison de plusieurs courbes ROC	24
1. Comparaison de deux courbes ROC	24
2. Comparaison de trois courbes ROC ou plus	25
IV. Bibliographie.....	26

I. Objectifs du tutoriel

Ce tutoriel a pour objectifs (1) de vous fournir un minimum de théorie concernant les courbes ROC¹, (2) de vous permettre de créer une ou plusieurs courbes ROC grâce au site Internet easyROC (<http://www.biosoft.hacettepe.edu.tr/easyROC/>)², et (3) d'interpréter les résultats fournis par le site Internet easyROC. Ce tutoriel met aussi l'accent sur les méthodes d'identification de la valeur seuil du test diagnostique qui permet discriminer de façon *optimale* les individus malades des individus sains. Par ailleurs, l'article de Goksuluk et coll. décrit en détail tous les calculs mathématiques et statistiques du site Internet easyROC (Goksuluk et al., 2016).

Même si ce tutoriel fournit le minimum à savoir sur les courbes ROC, je vous recommande néanmoins la lecture des articles suivants (Drobatz, 2009; Greiner et al., 2000; Soreide, 2009; Sweig and Campbell, 1993), ainsi que celui, en français, de Delacour et coll. (Delacour et al., 2005). En revanche, les notions de sensibilité et de spécificité sont un pré-requis, et doivent donc être acquises avant de poursuivre ce tutoriel. (La lecture de l'article de Grimes et Schulz peut être une excellente opportunité de revoir tout cela (Grimes and Schulz, 2002) !)

Dans tout ce qui suit, je ferai référence aux éléments présents sur une figure en utilisant la convention suivante : « [Figure X.Y] », où « X » est le numéro de la Figure, et « Y » le numéro de l'élément pointé dans cette Figure X.

II. Un peu de théorie sur les courbes ROC

A. La courbe ROC

1. Contexte et objectif de réalisation d'une courbe ROC

Nous souhaitons réaliser une courbe ROC lorsque nous souhaitons montrer que la valeur d'un paramètre quantitatif obtenue à l'aide d'un instrument de mesure est capable de bien discriminer les individus présentant un caractère des individus ne le présentant pas. Dans ce contexte, il faut s'assurer avec certitude, grâce à une méthode de référence (méthode dite « Gold Standard »), que les individus présentent bien ce caractère tandis que les autres ne le présentent effectivement pas. Comme je vais me placer dans un contexte médical dans toute la suite de ce tutoriel, je vais considérer que cet instrument de mesure est un « test diagnostique » (au sens le plus large du terme), et je vais considérer que le caractère binaire (présent/absent) est la « maladie » étudiée. Dans ce contexte médical, nous souhaitons réaliser une courbe ROC lorsque nous souhaitons savoir si un test diagnostique permet de correctement discriminer les individus malades des individus sains (Lusted, 1971). Dans ce contexte, le test diagnostique est une nouvelle méthode d'évaluation de la présence de la maladie, et l'existence d'une méthode Gold Standard (ou considérée comme telle) pour évaluer correctement la présence de la maladie est indispensable pour réaliser une courbe ROC.

Ainsi, dans la majorité des situations, l'objectif de l'utilisation de courbes ROC est de savoir si des valeurs « élevées » d'un test diagnostique correspondent à des individus malades et si de « faibles » valeurs de ce test diagnostique correspondent à des individus sains. Ces notions de « élevées » et « faibles » font référence au fait que les valeurs fournies par le test diagnostique seront supérieures ou inférieures à un seuil, qui sera identifié par une méthode dont nous reparlerons dans ce tutoriel. (Dans certaines situations cliniques, ce sont plutôt des valeurs faibles d'un test diagnostique qui

¹ Pour « Receiver Operating Characteristic ».

² Il arrive parfois que ce site ne soit pas accessible. Si tel est le cas, veuillez essayer celui-ci : <http://opensoft.turcosa.com.tr/easyROC/>

pourraient indiquer la présence de la maladie, et des valeurs élevées qui pourraient indiquer l'absence de la maladie. Mais la démarche globale reste identique.)

2. La construction de la courbe ROC

Supposons le fichier de données de 15 individus présenté dans le Tableau 1 ci-dessous, dont 4 sont malades et 11 sont sains (soit un pourcentage de prévalence de la maladie dans l'échantillon de 27%), avec chacun leur valeur fournie par un test diagnostique dont on souhaite quantifier sa capacité à discriminer les individus malades des individus sains.

Pour chacune des différentes valeurs V du test diagnostique, nous pouvons calculer la proportion d'individus dont la valeur du test diagnostique est supérieure ou égale à V parmi les individus malades de l'échantillon, et nous pouvons calculer la proportion d'individus dont la valeur du test diagnostique est inférieure à V parmi les individus sains de l'échantillon. Ces deux valeurs sont respectivement la sensibilité (Se) et la spécificité (Sp) du test diagnostique en prenant comme valeur seuil la valeur V , toutes deux estimées dans l'échantillon (cf. les deux colonnes « Se » et « Sp » du Tableau 1 en face de chaque valeur V). Par exemple, si l'on prend la valeur $V=16$ (valeur de l'individu n°8), parmi les 4 individus malades, 3 ont une valeur ≥ 16 (soit une Se de $3/4=0,75$), et parmi les 11 individus sains, 5 ont une valeur < 16 (soit une Sp de $6/11=0,55$). C'est la raison pour laquelle, à la ligne de l'individu n°8, la colonne « Se » indique la valeur de 0,75, et la colonne « Sp » indique la valeur de 0,55.

La courbe ROC est un graphique dont les points ont pour abscisse la colonne « $1-Sp$ » et pour ordonnées la colonne « Se » du Tableau 1. Il y aura donc autant de points sur la courbe ROC que de valeurs différentes du test diagnostique.

Tableau 1. Données fictives de 15 individus pour les calculs des coordonnées d'une courbe ROC.

Id	Valeur du test diagnostique	Malade	Se	Sp	1-Sp	Se+Sp-1	$(1-Se)^2 + (1-Sp)^2$
1	2	0	1,00	0,00	1,00	0,00	1,00
2	4	0	1,00	0,09	0,91	0,09	0,83
3	6	1	1,00	0,18	0,82	0,18	0,67
4	8	0	0,75	0,18	0,82	-0,07	0,73
5	10	0	0,75	0,27	0,73	0,02	0,59
6	12	0	0,75	0,36	0,64	0,11	0,47
7	14	0	0,75	0,45	0,55	0,20	0,36
8	16	0	0,75	0,55	0,45	0,30	0,27
9	18	1	0,75	0,64	0,36	0,39	0,19
10	20	0	0,50	0,64	0,36	0,14	0,38
11	22	1	0,50	0,73	0,27	0,23	0,32
12	24	0	0,25	0,73	0,27	-0,02	0,64
13	26	0	0,25	0,82	0,18	0,07	0,60
14	28	1	0,25	0,91	0,09	0,16	0,57
15	30	0	0,00	0,91	0,09	-0,09	1,01

Une courbe ROC doit forcément commencer par le point de coordonnées (0,0) et terminer par le point de coordonnées (1,1). Les points sont ensuite reliés par un segment (segment qui peut être vertical, oblique, ou horizontal). La courbe ROC issue du Tableau 1 est présentée dans la Figure 1. Dans cette Figure 1, chaque point [Figure 1.i] correspond à la valeur du test diagnostique de l'individu n° i ($i \in \{1, \dots, 15\}$), dont l'abscisse et l'ordonnée sont respectivement les valeurs de Se et $1-Sp$ correspondant à cette valeur du test diagnostique. Par exemple, le point [Figure 1.6] correspond à la valeur 12 du test diagnostique de l'individu n°6 (cf. Tableau 1). A cette valeur de 12 correspondent

(cf. Tableau 1) une Sp de 0,36, donc une valeur de $1-Sp$ de 0,64 (l'abscisse du point ; Figure 1.16), et une Se de 0,75 (l'ordonnée du point ; Figure 1.17).

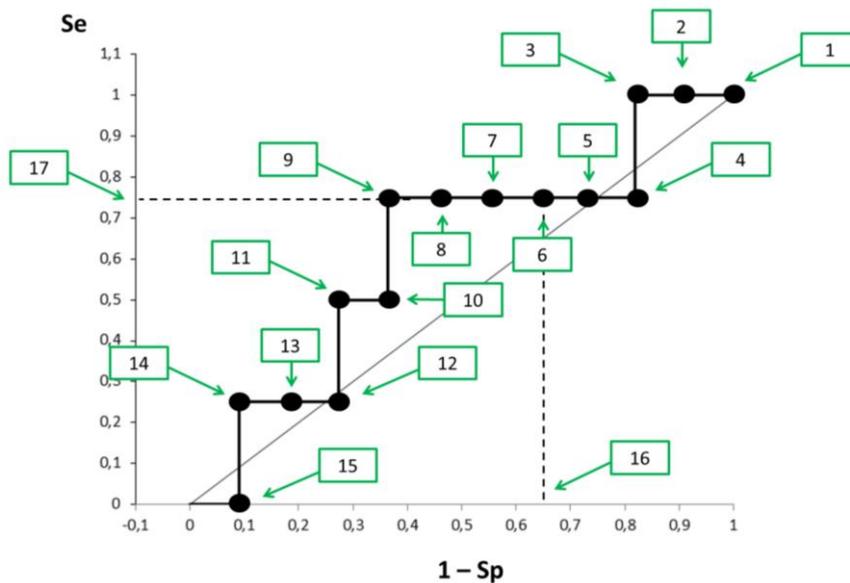


Figure 1. Courbe ROC à partir des données du Tableau 1.

Plus la courbe ROC s'éloigne de la droite à 45° (représentée par un trait fin sur la Figure 1) tout en restant au-dessus, plus le test diagnostique est capable de bien discriminer les individus malades des individus sains.

3. L'aire sous la courbe ROC

L'aire sous la courbe ROC est un indicateur statistique quantifiant la capacité du test diagnostique à discriminer les individus malades des individus sains. Si la courbe ROC est confondue avec la droite à 45°, l'aire sous cette courbe vaut 0,5. Par conséquent, un test diagnostique sera d'autant plus capable de discriminer les individus malades des individus sains que cette aire sous la courbe ROC est éloignée, par valeur supérieure, de 0,5. Cette aire a une valeur maximale de 1. Dans la Figure 1, l'aire sous la courbe ROC vaut 0,61 (somme des aires rectangulaires sous la courbe). Cette valeur de 0,61 signifie que, dans 61% des cas, un individu tiré au sort de la population des individus malades a une valeur du test diagnostique supérieure à celle d'un individu tiré au sort de la population des individus sains (Sweig and Campbell, 1993). Swets a proposé des seuils d'aire sous la courbe ROC facilitant l'interprétation de la valeur de cette aire (Swets, 1988) : « Values of [the area under the ROC curve] between 0.50 and 0.70 or so represent a rather low accuracy - the true-positive proportion is not much greater than the false-positive proportion anywhere along the curve. Values of [this area] between about 0.70 and 0.90 represent accuracies that are useful for some purposes, and higher values represent a rather high accuracy ».

B. Les méthodes d'identification du seuil optimal

1. Introduction

Si un test diagnostique possède une bonne capacité à discriminer les individus malades des individus sains (c'est-à-dire, une valeur d'aire sous la courbe ROC proche de 1), il est alors quasi indispensable d'identifier la valeur seuil V_{seuil} qui permette justement de bien discriminer ces deux groupes d'individus. De nombreuses méthodes existent pour identifier V_{seuil} , en fonction des qualités du test diagnostique recherchées (Lopez-Raton et al., 2014). Les trois méthodes d'identification du seuil optimal principalement utilisées en recherche clinique sont l'indice de Youden, le point le plus proche du coin supérieur gauche de la courbe ROC, et l'indice de Youden généralisé (Perkins and Schisterman, 2006).

2. L'indice de Youden

a) Calcul de l'indice de Youden

Pour calculer l'indice de Youden, il est au préalable nécessaire de calculer, pour chaque valeur différente du test diagnostique, la valeur de $Se+Sp-1$ (cf. Tableau 1). L'indice de Youden a pour valeur $\max(Se+Sp-1)$ parmi toutes les valeurs de $Se+Sp-1$ calculées (Schisterman et al., 2005; Youden, 1950). A partir des données du Tableau 1, nous pouvons voir que l'indice de Youden vaut 0,39. Il correspond à la valeur de 18 pour le test diagnostique. Par conséquent, le seuil du test diagnostique permettant de discriminer les individus malades des individus sains de façon optimale selon l'indice de Youden est 18. Avec un tel seuil, le test diagnostique a une Se de 0,75 et une Sp de 0,64 (cf. Tableau 1).

b) Hypothèses de calcul et limites dans l'interprétation de l'indice de Youden

La simplicité du calcul de l'indice de Youden ne doit pas faire oublier les hypothèses de calcul qui sont indispensables à connaître et à comprendre pour correctement interpréter la valeur du seuil optimal identifié par l'indice de Youden. Ces hypothèses d'interprétation sont souvent peu connues par les utilisateurs de l'indice de Youden.

L'indice de Youden identifie le seuil du test diagnostique qui maximise la somme $Se+Sp$. Autrement dit, l'indice de Youden identifie le seuil du test diagnostique qui minimise la somme $(1-Se)+(1-Sp)$. Or, $(1-Se)$ représente la proportion d'individus faux négatifs et $(1-Sp)$ représente la proportion d'individus faux positifs. Donc, l'indice de Youden identifie le seuil du test diagnostique qui minimise la somme de ces deux proportions, en donnant le même poids à chacune de ces deux proportions : les termes « Se » et « Sp » sont en effet complètement interchangeables dans la formule de l'indice de Youden.

Nous pouvons tout d'abord faire un premier commentaire. D'un point de vue clinique, ce n'est pas tant la somme des *proportions* d'individus faux négatifs et d'individus faux positifs que l'on aimerait minimiser, mais la somme des *nombre*s d'individus faux négatifs et d'individus faux positifs. Cela dit, quand, dans la population, il y a autant d'individus malades que d'individus non malades (\Leftrightarrow pourcentage de prévalence de la maladie dans la population égale à 50%), minimiser la somme de ces proportions revient à minimiser la somme des nombres d'individus correspondants. Mais dès que le pourcentage de prévalence de la maladie dans la population s'éloigne de 50%, ce n'est plus le cas ! Nous allons illustrer cela ci-dessous.

Reprenons l'exemple du Tableau 1. L'indice de Youden a identifié le seuil de 18, conduisant à une Se de 0,75 et une Sp de 0,64. Supposons que l'on applique ce test diagnostique et ce seuil de 18 au sein d'une population de 2000 individus, au sein de laquelle le pourcentage de prévalence de la maladie est de 50% (soit « π » le pourcentage de prévalence de la maladie dans la population). Supposons aussi que ces valeurs de Se et de Sp dans l'échantillon sont exactement celles, intrinsèques, du test diagnostique (nous ferons cette hypothèse dans toute la suite de ce tutoriel). Dans cette population,

parmi les 1000 individus malades, il y aura 250 individus faux négatifs, et parmi les 1000 individus sains, il y aura 360 individus faux positifs, soit au total, $250+360=610$ individus mal classés. Si l'on avait choisi tout autre seuil que 18, avec $\pi=50\%$, le nombre total d'individus mal classés dans cette population aurait été supérieur à celui avec le seuil de 18. Par exemple, avec un seuil de 20, le Tableau 1 indique que la Se vaut 0,50 et la Sp vaut 0,64, ce qui aurait conduit à 500 individus faux négatifs et 360 individus faux positifs parmi les 2000 individus de la population, soit au total 860 individus mal classés (donc 250 individus mal classés *de plus* qu'avec l'utilisation du seuil de 18). Autre exemple, avec un seuil de 16 cette fois-ci, le Tableau 1 indique que la Se vaut 0,75 et la Sp vaut 0,55. Dans cette même population, l'utilisation du seuil de 16 aurait conduit à 250 individus faux négatifs et 450 individus faux positifs, soit au total 700 individus mal classés (donc 90 individus mal classés *de plus* qu'avec l'utilisation du seuil de 18). Si maintenant $\pi=10\%$ dans cette population de 2000 individus (comprenant donc maintenant 200 individus malades et 1800 individus sains), l'utilisation du seuil optimal selon l'indice de Youden de valeur 18 (Se de 0,75 et Sp de 0,64) aurait conduit à 50 individus faux négatifs et 648 individus faux positifs, soit au total 698 individus mal classés. Dans cette même population, si on avait utilisé le seuil de 26 par exemple, dont les Se et Sp correspondantes sont 0,25 et 0,82 (cf. Tableau 1), le nombre d'individus faux négatifs aurait été de 150 et le nombre d'individus faux positifs aurait été de 324, soit un total de 474 individus mal classés, et donc 224 individus mal classés *de moins* par rapport au nombre d'individus mal classés avec le seuil de 18 identifié par l'indice de Youden. Nous venons donc de voir que, même si dans les deux cas ($\pi=50\%$ et $\pi=10\%$) l'indice de Youden identifie systématiquement le seuil qui minimise la somme des proportions d'individus mal classés, il n'identifie pas *systématiquement* le seuil qui minimise la somme du nombre d'individus mal classés (il le fait uniquement lorsque $\pi=50\%$).

En résumé, et de façon très importante, utiliser dans la population le test diagnostique et prendre comme seuil optimal celui identifié par l'indice de Youden part du principe que les quatre hypothèses de travail suivantes sont acceptées par l'utilisateur de l'indice de Youden (il faut donc en avoir conscience) :

- (1) on cherche à minimiser les proportions d'individus faux négatifs et d'individus faux positifs, et non pas à minimiser les nombres d'individus faux négatifs et d'individus faux positifs ;
- (2) l'indice de Youden identifie le seuil du test diagnostique qui minimise ces nombres d'individus faux négatifs et d'individus faux positifs dans le même temps si et seulement si le pourcentage de prévalence de la maladie dans la population (π) vaut 50% ;
- (3) identifier un seuil qui fait diminuer de X% la sensibilité du test diagnostique a le même « coût » (clinique, sanitaire, social, ...) qu'identifier un seuil qui fait diminuer de la même quantité de X% la spécificité (car les termes « Se » et « Sp » sont interchangeable dans la formule de l'indice de Youden) ;
- (4) dans la situation où $\pi=50\%$, un individu faux négatif a le même « coût » (clinique, sanitaire, sociétal, ...) qu'un individu faux positif.

Par conséquent, si la valeur de π est éloignée de 50% et/ou si un individu faux négatif n'a pas le même « coût » (clinique, sanitaire, social, ...) qu'un individu faux positif, alors l'indice de Youden ne sélectionne plus de façon optimale le seuil du test diagnostique. Dans ce cas-là, je vous recommande d'utiliser l'indice de Youden généralisé, car il prend justement en compte dans son calcul la valeur de π (estimée par l'utilisateur de la méthode à partir de ses connaissances du terrain) ainsi que le « coût » d'un individu faux négatif par rapport à celui d'un individu faux positif (estimé, là encore, par l'utilisateur de la méthode à partir de ses connaissances du terrain).

3. Le point le plus proche du coin supérieur gauche de la courbe ROC

Une autre méthode d'identification du seuil optimal consiste à identifier la valeur du test diagnostique correspondant au point sur la courbe ROC le plus proche du coin supérieur gauche (de coordonnées (0,1)). Pour identifier ce seuil, il est nécessaire de calculer au préalable, pour chaque valeur différente du test diagnostique, la valeur de $(1-Se)^2+(1-Sp)^2$. Le seuil correspondant au point le plus proche du coin supérieur gauche est la valeur du test diagnostique minimisant cette valeur de $(1-Se)^2+(1-Sp)^2$ (Perkins and Schisterman, 2006). Dans le Tableau 1, on peut voir que la méthode d'identification du seuil optimal selon le point le plus proche du coin supérieur gauche identifie le même seuil que celui identifié par l'indice de Youden : 18. Cela dit, cette méthode d'identification ne minimise pas forcément ni la somme des proportions d'individus faux négatifs et faux positifs, ni les nombres d'individus correspondants, ni même les coûts des individus faux négatifs et faux positifs. Par conséquent, les limites dans l'utilisation de l'indice de Youden énoncées ci-dessus se retrouvent au moins tout autant dans l'utilisation de la méthode d'identification du seuil optimal selon le point le plus proche du coin supérieur gauche.

4. L'indice de Youden généralisé

Pour calculer l'indice de Youden généralisé, il est au préalable nécessaire de calculer, pour chaque valeur différente du test diagnostique, la valeur suivante : $Se + \frac{1-\pi}{CFN/CFP \times \pi} \times Sp - 1$, avec π le pourcentage de prévalence de la maladie dans la population, et CFN/CFP le rapport du coût d'un individu faux négatif (CFN) par celui d'un individu faux positif (CFP) (Perkins and Schisterman, 2006). L'indice de Youden généralisé a pour valeur $\max\left(Se + \frac{1-\pi}{CFN/CFP \times \pi} \times Sp - 1\right)$, parmi toutes les valeurs calculées (une par valeur différente du test diagnostique). L'utilisateur de l'indice de Youden généralisé doit utiliser ses connaissances du terrain pour estimer la valeur de π ainsi que le rapport CFN/CFP . Par exemple, choisir $CFN/CFP=2$ signifie que l'on pense que le « coût » (clinique, sanitaire, social, ...) d'un individu faux négatif est deux fois plus élevé que celui d'un individu faux positif. La situation où $CFN/CFP \gg 1$ est celle où ne pas identifier un individu vraiment malade serait beaucoup plus problématique par la suite (décompensation, traitements beaucoup plus lourds, décès rapide sans ce diagnostic, ...) que de prendre en charge un individu sain comme s'il était malade (initiation d'un traitement à tort, suivi clinique inutile, stress inutile suite à l'annonce erronée de la présence de la maladie, ...). Nous allons voir ci-dessous que l'indice de Youden généralisé minimise la somme des *nombres* d'individus faux négatifs et d'individus faux positifs, en prenant en compte le pourcentage de prévalence de la maladie dans la population (π) et le rapport du coût d'un individu faux négatif par celui d'un individu faux positif.

Le Tableau 2 reprend les données du Tableau 1, selon plusieurs scénarios : $\pi=50\%$ ou $\pi=30\%$, et $CFN/CFP=1$ ou $CFN/CFP=3^3$. Le Tableau 2 présente les valeurs permettant d'identifier le seuil optimal selon l'indice de Youden (les valeurs de $Se+Sp-1$), et celui selon l'indice de Youden généralisé (les valeurs de $Se + \frac{1-\pi}{CFN/CFP \times \pi} \times Sp - 1$).

³ Ce qui veut dire que l'on considère que le « coût » d'un individu faux négatif est trois fois plus élevé que celui d'un individu faux positif.

Tableau 2. Données du Tableau 1 permettant de calculer l'indice de Youden généralisé selon plusieurs scénarios en fonction du pourcentage de prévalence de la maladie (π) et du rapport du coût d'un individu faux négatif (CFN) par celui d'un individu faux positif (CFP).

Valeur du test diagnostique	Se	Sp	Se+Sp-1	$Se + \frac{1 - \pi}{CFN/CFP} \times Sp - 1$			
				CFN/CFP=1		CFN/CFP=3	
				$\pi=50\%$	$\pi=30\%$	$\pi=50\%$	$\pi=30\%$
2	1,00	0,00	0,00	0,00	0,00	0,00	0,00
4	1,00	0,09	0,09	0,09	0,21	0,03	0,07
6	1,00	0,18	0,18	0,18	0,42	0,06*	0,14
8	0,75	0,18	-0,07	-0,07	0,17	-0,19	-0,11
10	0,75	0,27	0,02	0,02	0,39	-0,16	-0,04
12	0,75	0,36	0,11	0,11	0,60	-0,13	0,03
14	0,75	0,45	0,20	0,20	0,81	-0,10	0,10
16	0,75	0,55	0,30	0,30	1,02	-0,07	0,17
18	0,75	0,64	0,39*	0,39*	1,23	-0,04	0,24*
20	0,50	0,64	0,14	0,14	0,98	-0,29	-0,01
22	0,50	0,73	0,23	0,23	1,20	-0,26	0,07
24	0,25	0,73	-0,02	-0,02	0,95	-0,51	-0,18
26	0,25	0,82	0,07	0,07	1,16	-0,48	-0,11
28	0,25	0,91	0,16	0,16	1,37*	-0,45	-0,04
30	0,00	0,91	-0,09	-0,09	1,12	-0,70	-0,29

* : valeur maximale conduisant à sélectionner le seuil optimal pour le test diagnostique.

Nous pouvons commenter chacune des quatre dernières colonnes du Tableau 2.

(1^{ère} colonne) Pour une valeur de $\pi=50\%$ et pour un « coût » des individus faux négatifs identique à celui des individus faux positifs (c'est-à-dire, CFN/CFP=1), la valeur de $Se + \frac{1 - \pi}{CFN/CFP} \times Sp - 1$ est identique à celle de Se+Sp-1, ce qui est cohérent avec le fait que l'indice de Youden sélectionne le seuil optimal dans un contexte où $\pi=50\%$ et où le « coût » d'un individu faux négatif est identique à celui d'un individu faux positif (indépendamment du taux de prévalence de la maladie dans l'échantillon, qui est ici de 27%).

(2^{ème} colonne) Pour une valeur de $\pi=30\%$, et pour un rapport CFN/CFP de 1, l'indice de Youden généralisé conduit à identifier le seuil de 28, conduisant à une Se de 0,25 et une Sp de 0,91. Reprenons l'exemple de la population de 2000 individus au sein de laquelle $\pi=30\%$. Avec l'utilisation d'un tel seuil de 28, il y aurait 450 individus faux négatifs, et 126 individus faux positifs, soit un total de 576 individus mal classés. L'indice de Youden, lui, identifie le seuil de 18, avec une Se de 0,75 et une Sp 0,64. Ainsi, dans cette même population, l'utilisation du seuil de 18 conduirait à 150 individus faux négatifs et 504 individus faux positifs, soit un total de 654 individus mal classés (donc 78 de plus qu'avec l'indice de Youden généralisé). Nous venons donc de voir que l'indice de Youden généralisé minimise la somme du nombre d'individus mal classés en prenant en compte la valeur de π , tandis que l'indice de Youden ne minimise cette somme du nombre d'individus mal classés que si $\pi=50\%$. Ensuite, nous pouvons remarquer que pour minimiser la somme du nombre d'individus mal classés, plus π diminue, plus l'indice de Youden généralisé privilégie une valeur élevée de Sp plutôt qu'une valeur élevée de Se (tout en garantissant un nombre d'individus mal classés le plus faible possible, quelle que soit la valeur de π).

(3^{ème} colonne) Pour une valeur de $\pi=50\%$, et pour un rapport CFN/CFP de 3, l'indice de Youden généralisé conduit à sélectionner le seuil de 6, conduisant à une Se de 1,00 et une Sp de 0,18 (cf. Tableau 2). Si l'on illustre en € le rapport CFN/CFP=3 de façon totalement théorique, cela signifierait qu'un individu faux négatif coûterait 3€ tandis qu'un individu faux positif coûterait 1€. Reprenons l'exemple de la population de 2000 individus au sein de laquelle $\pi=50\%$. Avec l'utilisation d'un tel seuil de 6, il n'y aurait aucun individu faux négatif et 820 individus faux positifs, soit un coût total de 820€. L'indice de Youden, lui, sélectionne toujours le seuil de 18, avec une Se de 0,75 et une Sp 0,64. Dans cette même population, utiliser le seuil de 18 conduirait à 250 individus faux négatifs et 360 individus faux positifs, soit un total de 610 individus mal classés, donc 210 individus de moins qu'avec l'indice de Youden généralisé. *Mais* le « coût » d'un tel seuil de 18 est le suivant : $250 \times 3€ + 360 \times 1€ = 1110€$, donc 290€ de plus qu'en sélectionnant le seuil de 6 selon l'indice de Youden généralisé !

(4^{ème} colonne) Nous retrouvons bien la fin du commentaire de la 2^{ème} colonne, à savoir que pour un même rapport de coût (ici, CFN/CFP=3), le fait de passer de $\pi=50\%$ à $\pi=30\%$ conduit l'indice de Youden généralisé à sélectionner un seuil (en l'occurrence, 18 ici) qui privilégie une valeur élevée de Sp plutôt qu'une valeur élevée de Se (entre la 3^{ème} et la 4^{ème} colonne, on passe d'une Se de 1,00 à une Se de 0,75, et une Sp de 0,18 à une Sp de 0,64). Ensuite, nous pouvons remarquer, en comparant les valeurs de Se et Sp correspondant aux quatre seuils optimaux identifiés par l'indice de Youden généralisé (repérés par le signe « * » dans chacune des quatre dernières colonnes du Tableau 2), que pour minimiser le « coût » (clinique, sanitaire, social, ...) des individus mal classés, plus le ratio CFN/CFP augmente, plus l'indice de Youden généralisé privilégie une valeur élevée de Se plutôt qu'une valeur élevée de Sp , et ce, quelle que soit la valeur de π .

C. Indicateurs statistiques de la qualité d'un test diagnostique une fois le seuil optimal identifié

Nous avons déjà longuement parlé des Se et Sp estimées et correspondantes au seuil optimal identifié. Deux autres familles d'indicateurs statistiques sont au moins tout autant pertinentes que Se et Sp : les valeurs prédictives (positive et négative ; respectivement VPP et VPN) (Grimes and Schulz, 2002) et les rapports de vraisemblance (positif et négatif ; respectivement RV^+ et RV^-) (Timsit et al., 2018).

1. Valeurs prédictives positive et négative

Brièvement, la valeur de la VPP s'interprète de la façon suivante : parmi les individus dont le test diagnostique est positif, $VPP\%$ sont malades. (Et celle de la VPN s'interprète de la façon suivante : parmi les individus dont le test diagnostique est négatif, $VPN\%$ sont sains.) Contrairement aux valeurs de Se et Sp , celles de la VPP et de la VPN dépendent du pourcentage de prévalence de la maladie dans la population cible (π) (Grimes and Schulz, 2002). On peut ainsi déduire des valeurs de Se , de Sp , et de π la valeur de la VPP et de la VPN (Altman and Bland, 1994) : $VPP = \frac{\pi \times Se}{(1-\pi) \times (1-Sp) + (\pi \times Se)}$ et $VPN = \frac{(1-\pi) \times Sp}{(1-\pi) \times Sp + \pi \times (1-Se)}$. Par exemple, en reprenant les valeurs du Tableau 2, pour un seuil optimal de 18 (identifié par l'indice de Youden), conduisant à une Se de 0,75 et une Sp de 0,64, et en fixant π à 50%, les VPP et VPN du test diagnostique valent respectivement 0,68 et 0,72. Ainsi, en utilisant le seuil de 18 dans une population où $\pi=50\%$, 68% des individus dont la valeur du test diagnostique est supérieure ou égale à 18 seront malades, et 72% des individus dont la valeur du test diagnostique est inférieure à 18 seront sains.

2. Rapports de vraisemblance positif et négatif

La valeur du rapport de vraisemblance (RV) quantifie la modification de la probabilité qu'un individu soit malade entre celle *avant* l'annonce du résultat du test diagnostique (probabilité pré-test) et celle *après* le résultat du test diagnostique (probabilité post-test) (Hayden and Brown, 1999). La probabilité pré-test est une probabilité estimée par le clinicien à partir du tableau clinique de l'individu. La formule liant la probabilité pré-test ($P_{\text{pré}}$) à la probabilité post-test (P_{post}), par l'intermédiaire de RV, est la suivante : $P_{\text{post}} = P_{\text{pré}} \times \frac{RV}{(1-P_{\text{pré}}) + P_{\text{pré}} \times RV}$. L'utilisation d'un graphique appelé « nomogramme » (Fagan, 1975; Grimes and Schulz, 2005) (cf. Figure 2) permet d'en déduire la valeur de P_{post} à partir de $P_{\text{pré}}$ et de la valeur de RV. Deux RV doivent être calculés : RV^+ et RV^- . RV^+ quantifie la modification de la probabilité qu'un individu soit malade entre celle avant l'annonce du résultat *positif* du test diagnostique et celle après l'annonce de ce résultat positif. RV^- quantifie la modification de la probabilité qu'un individu soit malade entre celle avant l'annonce du résultat *négatif* du test diagnostique et celle après l'annonce de ce résultat négatif. Les valeurs de RV^+ et de RV^- dépendent uniquement des valeurs intrinsèques de Se et Sp du test diagnostique. Ce sont donc des indicateurs statistiques de la qualité du test diagnostique qui, comme Se et Sp , ne dépendent par conséquent pas de π . Les formules sont les suivantes : $RV^+ = \frac{Se}{1-Sp}$ et $RV^- = \frac{1-Se}{Sp}$.

La Figure 2 représente un nomogramme vierge (Figure 2.A), ainsi que trois exemples d'utilisation du nomogramme. Le nomogramme est constitué de trois axes verticaux, le premier représentant la probabilité $P_{\text{pré}}$, l'axe du milieu représentant la valeur de RV (RV^+ ou RV^- selon le résultat du test diagnostique), et l'axe de droite représentant la probabilité P_{post} . En traçant une droite passant par $P_{\text{pré}}$ et par RV, on en déduit la valeur de P_{post} . Par exemple, en reprenant les valeurs du Tableau 2, pour un seuil optimal de 18 (identifié par l'indice de Youden), conduisant à une Se de 0,75 et une Sp de 0,64, les RV^+ et RV^- du test diagnostique valent respectivement 2,4 et 0,4. Supposons qu'un individu se présente en clinique avec un tableau clinique qui fait penser à une probabilité de 20% qu'il soit malade ($P_{\text{pré}}=20\%$). Maintenant, de deux choses l'une. Supposons que le test diagnostique fournisse une valeur supérieure ou égale à 18, le test diagnostique est ainsi positif. Nous utilisons donc la valeur de RV^+ dans le nomogramme, et nous faisons passer une droite par la graduation de gauche de « 20 » (car $P_{\text{pré}}=20\%$) et par la graduation du milieu de « 2,4 » (car $RV^+=2,4$) (cf. Figure 2.B). La lecture de la valeur de la graduation de droite de la Figure 2.B nous donne environ « 40 » (la valeur de P_{post}). Ainsi, l'annonce du résultat positif du test diagnostique (valeur supérieure ou égale à 18) augmente la probabilité que cet individu soit malade à 40% (elle était jugée à 20% avant l'annonce du résultat, positif, du test diagnostique). Supposons maintenant que le test diagnostique fournisse une valeur inférieure à 18 pour ce même individu, le test est ainsi négatif. Nous utilisons donc la valeur de RV^- dans le nomogramme, et nous faisons passer une droite par la graduation de gauche de « 20 » (car $P_{\text{pré}}=20\%$) et par la graduation du milieu de « 0,4 » (car $RV^-=0,4$) (cf. Figure 2.C). La lecture de la valeur de la graduation de droite de la Figure 2.C nous donne environ « 7 » (P_{post}). Ainsi, l'annonce du résultat négatif du test diagnostique (valeur inférieure à 18) diminue la probabilité que cet individu soit malade à 7% (elle était jugée à 20% avant l'annonce du résultat, négatif, du test diagnostique).

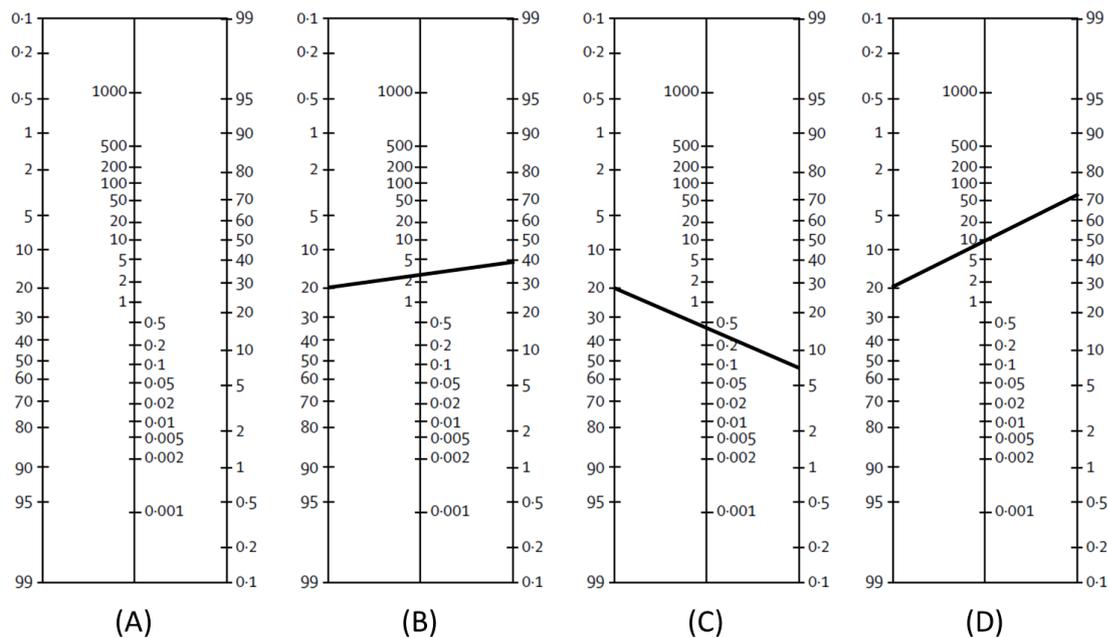


Figure 2. Nomogramme (issu de (Grimes and Schulz, 2005)) permettant de quantifier la probabilité post-test de maladie (P_{post}) à partir de la probabilité pré-test de maladie (P_{pre}) et du rapport de vraisemblance (RV). L'axe vertical de gauche représente la valeur de P_{pre} , celui du milieu représente la valeur de RV, et celui de droite représente la valeur de P_{post} . (A) représente un nomogramme vierge. (B), (C), et (D) représentent trois exemples détaillés dans le texte.

Il est reconnu qu'une valeur de RV^+ au moins égale à 10 permet de dire que le résultat positif d'un test diagnostique augmente fortement la présomption qu'un individu soit malade (Davidson, 2002). (De même, une valeur de RV^- inférieure ou égale à 0,1 permet de dire que le résultat négatif d'un test diagnostique augmente fortement la présomption qu'un individu soit sain.) En effet, si l'on reprend le nomogramme avec l'individu dont on estime la probabilité qu'il soit malade à 20%, et si la RV^+ d'un test diagnostique est égale à 10, alors l'annonce d'un résultat positif à ce test diagnostique nous permet de penser que cet individu est malade avec cette fois-ci une probabilité de plus de 70% (Figure 2.D). Ensuite, une valeur de RV^+ comprise entre 5 et 10 (ou un valeur de RV^- comprise entre 0,1 et 0,2) indique un effet « modéré » du résultat du test diagnostique sur la probabilité post-test, et une valeur de RV^+ comprise entre 2 et 5 (ou un valeur de RV^- comprise entre 0,2 et 0,5) indique un effet « faible » du résultat du test diagnostique sur la probabilité post-test (Hayden and Brown, 1999).

Une autre façon d'interpréter les valeurs de RV^+ et RV^- est la suivante (Timsit et al., 2018). Il y a RV^+ fois plus de chances que le test diagnostique soit positif lorsque l'individu est malade que lorsqu'il est sain ; il y a $1/RV^-$ fois plus de chances que le test diagnostique soit négatif lorsque l'individu est sain que lorsqu'il est malade. En reprenant les valeurs précédentes de RV^+ (2,4) et RV^- (0,4) pour le seuil identifié de 18, cela signifie qu'il y a 2,4 fois plus de chances que la valeur du test diagnostique soit supérieure ou égale à 18 lorsqu'un individu est malade que lorsqu'il est sain, et cela signifie qu'il y a $1/0,4=2,5$ fois plus de chances que la valeur du test diagnostique soit inférieure à 18 lorsque l'individu est sain que lorsqu'il est malade.

III. Utiliser le site Internet easyROC pour dresser une ou plusieurs courbes ROC

Cette partie présente, à partir de données fictives qui sont décrites ci-dessous, les résultats fournis par le site Internet easyROC, ainsi que l'interprétation de ces résultats.

A. Fichier de données

1. Préparation du fichier de données avant importation dans le site Internet easyROC

Si votre fichier de données est un fichier Excel, avant de l'importer dans le site Internet easyROC, vous devez vérifier que tous les points ci-dessous sont respectés dans votre fichier de données. Je vais appeler « VAR_QUANT_i » la variable quantitative correspondant au test diagnostique n°i dont vous souhaitez quantifier la capacité à discriminer les individus malades des individus sains ($i \in \{1, \dots, n\}$, avec n le nombre total de tests diagnostiques étudiés ; en général, $n=1$).

- La première ligne du fichier de données doit contenir le nom des intitulés de colonnes (variables). Vous choisissez le nom que vous souhaitez pour chacune de vos variables, tant qu'il respecte les trois conditions suivantes : pas d'espace (mettez un « _ » pour matérialiser visuellement un espace si besoin), pas d'accent (sinon, le site Internet easyROC met un « . » à la place de l'accent), et un nombre de caractères limité (de préférence ≤ 20).
- Le fichier de données doit contenir la variable relative à la maladie, que je vous recommande de coder de la façon suivante : « 0 » pour les individus sains et « 1 » pour les individus malades.
- Il ne doit y avoir aucune donnée manquante sur VAR_QUANT_i. S'il y en a, vous devez supprimer les individus pour lesquelles la donnée manque sur VAR_QUANT_i.
- Si vous souhaitez comparer statistiquement la capacité discriminante sur la présence de la maladie de plusieurs tests diagnostiques entre eux, toutes ces variables VAR_QUANT_i doivent figurer dans le même fichier de données, et les individus du fichier de données doivent être tels qu'ils n'aient aucune donnée manquante sur *toutes* ces variables VAR_QUANT_i. Si vous ne souhaitez pas comparer statistiquement leur capacité discriminante, mais simplement calculer la courbe ROC pour chacun de ces tests diagnostiques, et s'il y a des données manquantes sur ces variables, alors je vous conseille de créer autant de fichiers de données que de variables VAR_QUANT_i.
- Chaque variable VAR_QUANT_i ne doit contenir que des chiffres (sauf la première ligne qui contient le nom de la variable, ou sauf le signe « - » si votre test diagnostique peut prendre des valeurs négatives).

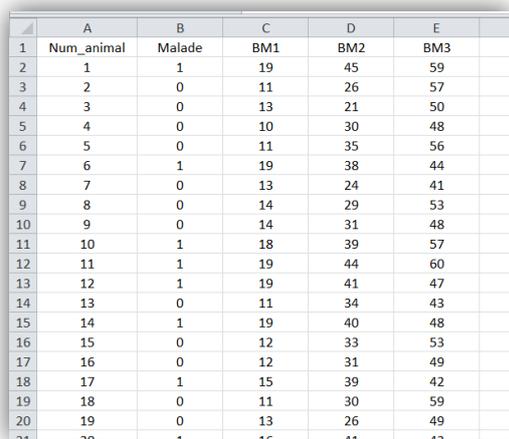
Notez le séparateur de décimal dans votre fichier de données. Si votre ordinateur est paramétré dans la langue française, il y a de très bonnes chances pour que ce séparateur soit la virgule (le séparateur de décimal est le point dans les systèmes anglo-saxons).

Une fois que vous vous êtes assuré que tous les points ci-dessus sont vérifiés, enregistrez-le (sous le format Excel), puis enregistrez-le sous le format .csv (séparateur : point-virgule). Au moment de le faire, Excel vous demande « Voulez-vous conserver le format du classeur ? ». Cliquez sur « Oui », puis fermez le fichier. Excel vous demande alors si vous souhaitez enregistrer les modifications, cliquez sur « Non »⁴.

⁴ Je sais, c'est contre-intuitif, mais comme vous avez déjà tout enregistré avant, sous le format Excel, puis sous le format .csv, vous pouvez fermer le fichier sans avoir besoin de l'enregistrer à nouveau !

2. Présentation du fichier de données fictives utilisé

Un nouveau fichier de données va être utilisé pour vous présenter le site Internet easyROC. Ce fichier de données fictives comprend 60 animaux, dont 28 animaux malades et 32 animaux sains. Ce fichier de données contient les valeurs de trois tests diagnostiques, qui correspondent à trois biomarqueurs BM1, BM2, et BM3, et dont on souhaite quantifier leur capacité à discriminer les animaux malades des animaux sains, à l'aide de courbes ROC. Comme nous allons être amenés à comparer statistiquement leur capacité de discrimination, les trois variables BM1, BM2, et BM3 doivent figurer dans le même fichier de données. La Figure 3 présente les 19 premières lignes du fichier de données (pour information, il n'y a aucune donnée manquante sur BM1, BM2, et BM3 pour les 60 animaux du fichier de données).



	A	B	C	D	E
1	Num_animal	Malade	BM1	BM2	BM3
2	1	1	19	45	59
3	2	0	11	26	57
4	3	0	13	21	50
5	4	0	10	30	48
6	5	0	11	35	56
7	6	1	19	38	44
8	7	0	13	24	41
9	8	0	14	29	53
10	9	0	14	31	48
11	10	1	18	39	57
12	11	1	19	44	60
13	12	1	19	41	47
14	13	0	11	34	43
15	14	1	19	40	48
16	15	0	12	33	53
17	16	0	12	31	49
18	17	1	15	39	42
19	18	0	11	30	59
20	19	0	13	26	49
21	20	1	16	41	43

Figure 3. Présentation des 19 premières lignes du fichier de données fictives utilisé pour présenter le site Internet easyROC.

B. Importation des données dans le site Internet easyROC

Une fois sur la page d'accueil du site Internet easyROC, nous cliquons sur « Data upload » [Figure 4.1], puis sur « Upload a file » [Figure 4.2], puis nous allons chercher notre fichier de données enregistré sur notre ordinateur, en cliquant sur « Browse » [Figure 4.3]. Nous sélectionnons le séparateur de colonne du fichier de données (le point-virgule⁵ si le fichier de données a été enregistré sous le format .csv) [Figure 4.4]. Nous devons spécifier que le séparateur de décimal est la virgule (« comma » ; la case n'est pas cochée par défaut⁶) [Figure 4.5]. Ensuite, nous sélectionnons la variable relative à la maladie [Figure 4.6], et la valeur de cette variable pour les animaux malades (« 1 », comme recommandé ci-dessus) [Figure 4.7]. Le site Internet présente les dix premières lignes du fichier de données [Figure 4.8] afin de vérifier que les données ont bien été importées.

⁵ « Semicolon » dans le site Internet.

⁶ Cela dit, si vous n'avez que des variables quantitatives *entières*, il n'est pas nécessaire de cocher cette case.

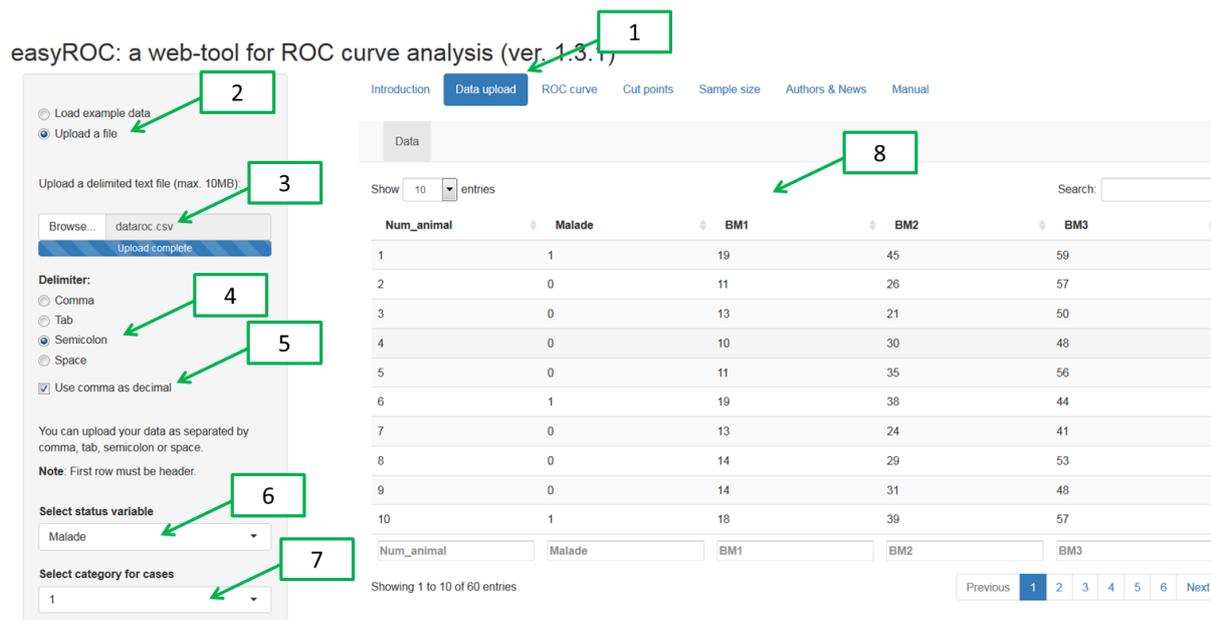


Figure 4. Copie d'écran de la page « Data upload » du site internet easyROC.

C. Courbe ROC avec une seule variable quantitative et indicateurs statistiques

1. Obtention de la courbe ROC

Nous allons maintenant créer une première courbe ROC, pour le test diagnostique BM2. Pour cela, nous cliquons sur « ROC curve » en haut de la page [Figure 5.1], et nous sélectionnons la variable quantitative BM2 correspondant au test diagnostique dont nous souhaitons obtenir la première courbe ROC [Figure 5.2].

Nous devons ensuite nous poser la question de savoir si des valeurs élevées de la variable quantitative indiquent *a priori* un risque accru de présenter la maladie, ou bien au contraire un risque moindre. Dans notre exemple, nous allons supposer qu'une augmentation des trois biomarqueurs indique *a priori* un risque accru de présenter la maladie. Donc, nous laissons cochée la case « Higher values indicate risks » (qui l'est effectivement par défaut) [Figure 5.3]. La situation où il faudrait la décocher serait celle où une augmentation de la variable quantitative indique *a priori* un risque moindre de présenter la maladie (comme par exemple la valeur de la protéinémie et la présence d'une glomérulopathie chez les chiens : une valeur élevée de la protéinémie indique *a priori* un risque moindre de présenter une glomérulopathie).

La courbe ROC est dressée en bas de la page [Figure 5.4], sous les résultats statistiques. On peut remarquer que cette courbe ROC est relativement éloignée de la droite à 45° tout en étant au-dessus, ce qui indique que le test diagnostique BM2 semble bien discriminer les animaux malades des animaux sains. Cela dit, il est nécessaire de quantifier cette capacité de discrimination à l'aide d'indicateurs statistiques, tels que l'aire sous la courbe ROC.

2. Indicateurs statistiques de la courbe ROC

Dans la mesure où nous sommes dans l'onglet « Statistics » de la page ([Figure 5.5] ; c'est l'onglet qui apparaît dès que l'on clique sur « ROC curve » [Figure 5.1]), ce sont les indicateurs statistiques de la courbe ROC qui sont présentés [Figure 5.6]. Ces indicateurs sont présentés dans l'ordre suivant : l'aire sous la courbe ROC (« AUC »), la *standard error* de cette aire (« SE.AUC »), les bornes inférieure et supérieure de l'intervalle de confiance à 95% de l'aire sous la courbe ROC (« LowerLimit » et « UpperLimit »), la statistique du test (« z »), et le degré de signification du test statistique (« p-value »). Ce test statistique teste l'hypothèse nulle suivante : la valeur réelle de l'aire sous la courbe ROC, dans la population cible qu'est censé représenter l'échantillon de l'étude, vaut 0,5 (en effet, une valeur de 0,5 pour l'aire sous la courbe ROC indique que la variable quantitative est absolument indépendante de la présence de la maladie). Ici, l'aire sous la courbe ROC vaut 0,908 [Figure 5.7], et cette valeur est significativement différente de 0,5 ($p < 0,05$; [Figure 5.8]). Nous retrouvons par ailleurs que cette aire sous la courbe ROC est significativement différente de 0,5 en voyant que son intervalle de confiance à 95% (0,831-0,985 ; [Figure 5.9]) ne comprend pas la valeur « 0,5 ». Attention, ce n'est pas parce qu'une aire sous la courbe ROC est significativement différente de 0,5 que le test diagnostique est performant ! En effet, c'est bien la valeur de l'aire sous la courbe ROC qui quantifie la capacité à discriminer les individus malades des individus sains. Un degré de signification inférieur à 0,05 permet uniquement d'inférer à la population qu'il y a de grandes chances pour que la valeur fournie par le test diagnostique ne soit pas indépendante de la présence de la maladie. Avec une taille d'échantillon importante, il est tout à fait possible d'obtenir une valeur d'aire sous la courbe ROC égale, par exemple, à 0,67 (traduisant le fait que le test diagnostique n'est pas très utile pour discriminer les individus malades des individus sains (Swets, 1988)), et qui soit pourtant significativement différente de 0,5.

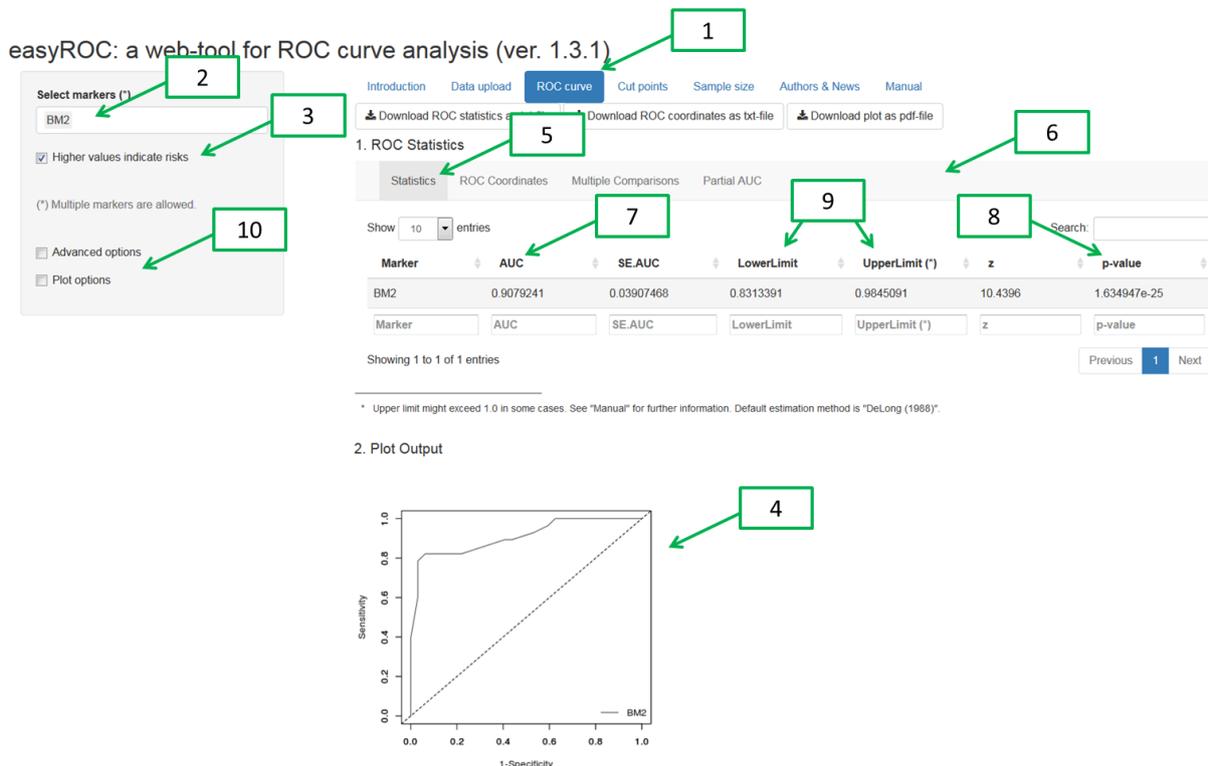


Figure 5. Copie d'écran de la page « ROC curve », onglet « Statistics ».

En cliquant sur « Advanced options » [Figure 5.10], nous pouvons spécifier la méthode de construction de la courbe ROC (paramétrique *versus* non paramétrique (Detilleux et al., 1999)), la méthode d'estimation de la *standard error* de l'aire sous la courbe ROC, et la méthode de calcul de

l'intervalle de confiance à 95%. En cliquant sur « Plot options » [Figure 5.10], nous pouvons modifier les caractéristiques graphiques de la courbe ROC. Les valeurs par défaut de toutes ces options sont suffisamment satisfaisantes pour que l'on n'ait pas besoin d'y toucher. Notamment, la construction par défaut de la courbe ROC est la méthode non paramétrique, qui la méthode classiquement utilisée dans les articles.

3. Liste des valeurs de la sensibilité et de la spécificité

Toujours en restant sur la page « ROC curve », mais cette fois-ci en cliquant sur « ROC Coordinates » [Figure 6.1], nous obtenons les trois informations suivantes : les valeurs que prend la variable quantitative dans le fichier de données et qui sont triées par ordre croissant (colonne « Cutpoint », [Figure 6.2]), la valeur correspondante de $(1-Sp)^7$ (colonne « FPR », [Figure 6.3]), et celle de la valeur de Se (colonne « TPR », [Figure 6.4]). Attention, pour correctement interpréter ces valeurs de $(1-Sp)$ et Se , il faut prendre en compte l'information de la case à cocher [Figure 5.3]. Ici, elle est cochée. Donc, le logiciel calcule la sensibilité Se et la spécificité Sp du test diagnostique qui dit « quand la variable quantitative est supérieure ou égale à V , l'animal est considéré comme malade, et quand elle est inférieure stricte à V , l'animal considéré comme sain ». (Si la case [Figure 5.3] n'avait pas été cochée, la phrase eût été : « quand la variable quantitative est inférieure ou égale à V , l'animal est considéré comme malade, et quand elle est supérieure stricte à V , l'animal est considéré comme sain ».) Le site Internet permet aussi de télécharger sous forme de fichier texte (ou pdf) les résultats dont nous venons de parler [Figure 6.5], sous forme de quatre colonnes, exactement comme celles présentées dans la Figure 6.

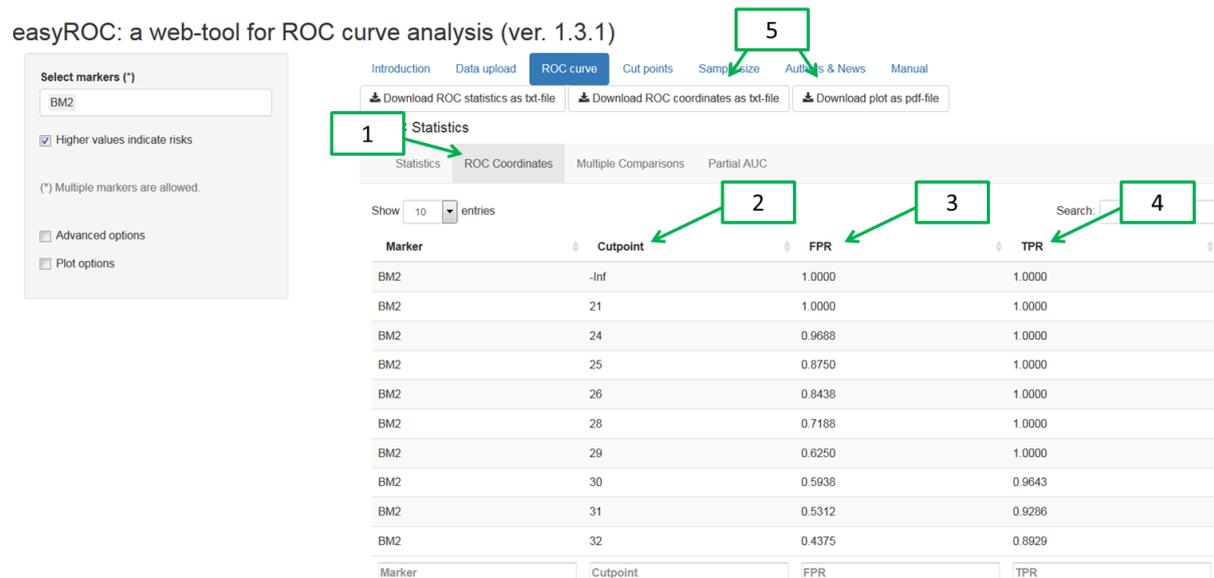


Figure 6. Copie d'écran de la page « ROC curve », onglet « ROC Coordinates ».

⁷ « FPR » = false-positive rate

⁸ « TPR » = true-positive rate

D. Seuil optimal, indicateurs statistiques et représentations graphiques de la qualité du test diagnostique

1. Identification du seuil optimal

a) Identification du seuil optimal selon l'indice de Youden

Nous allons maintenant cliquer sur « Cut points », tout en haut de la page du site Internet [Figure 7.1]. Cette page présente l'identification du seuil optimal de la variable quantitative, selon la méthode d'identification que l'on décide. Par défaut, la méthode d'identification est celle de l'indice de Youden [Figure 7.2]. Sur la Figure 7, on peut lire la valeur du seuil optimal pour le test diagnostique BM2, qui est de 36 (cf. Table 1 « Cut-off Results », [Figure 7.3]), avec un indice de Youden correspondant de valeur 0,7589 (valeur en face de « Optimal criterion », [Figure 7.3]).

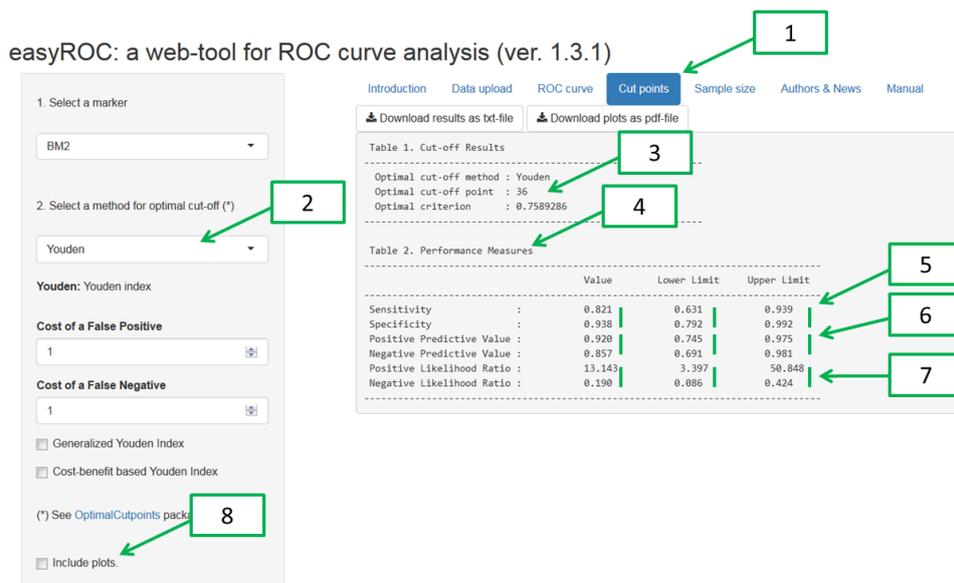


Figure 7. Copie d'écran de la page « Cut points », identification du seuil optimal selon l'indice de Youden.

b) Identification du seuil optimal selon le point le plus proche du coin supérieur gauche

Pour identifier le seuil optimal selon la méthode du point le plus proche du coin supérieur gauche, il suffit de cliquer sur « ROC01 » dans la liste déroulante [Figure 8.1]. Il se trouve que dans ces données fictives de 60 animaux, le seuil optimal identifié selon cette méthode est le même que celui identifié par l'indice de Youden (valeur de 36 dans la Table 1 « Cut-off Results », [Figure 8.2]).

easyROC: a web-tool for ROC curve analysis (ver. 1.3.1)

The screenshot shows the 'Cut points' page of the easyROC web-tool. On the left, there is a sidebar with the following settings:

- 1. Select a marker: BM2
- 2. Select a method for optimal cut-off (*): ROC01
- ROC01: minimizes distance between ROC plot and point (0,1)
- (*) See OptimalCulpoints package from R
- Include plots.

The main content area has a navigation bar with 'Introduction', 'Data upload', 'ROC curve', 'Cut points', 'Sample size', 'Authors & News', and 'Manual'. Below the navigation bar are two download links: 'Download results as txt-file' and 'Download plots as pdf-file'. The main content displays two tables:

Table 1. Cut-off Results

Optimal cut-off method :	ROC01
Optimal cut-off point :	36
Optimal criterion :	0.03579401

Table 2. Performance Measures

	Value	Lower Limit	Upper Limit
Sensitivity :	0.821	0.631	0.939
Specificity :	0.938	0.792	0.992
Positive Predictive Value :	0.920	0.745	0.975
Negative Predictive Value :	0.857	0.691	0.981
Positive Likelihood Ratio :	13.143	3.397	50.848
Negative Likelihood Ratio :	0.190	0.086	0.424

Figure 8. Copie d'écran de la page « Cut points », identification du seuil optimal selon le point le plus proche du coin supérieur gauche.

c) Identification du seuil optimal selon l'indice de Youden généralisé

Pour identifier le seuil optimal selon l'indice de Youden généralisé, vous devez sélectionner « Youden » pour la méthode d'identification [Figure 9.1], puis cliquer sur « Generalized Youden Index » [Figure 9.2]. Par défaut, les coûts des individus faux positifs (CFP) et celui des individus faux négatifs (CFN) sont fixés à « 1 » [Figure 9.3] & [Figure 9.4]. Le seuil optimal est présenté dans la Table 1 « Cut-off Results » [Figure 9.5], à partir de la valeur de l'indice de Youden généralisé, de valeur 0,8929. Ici, deux seuils optimaux sont proposés : 36 et 37, car la valeur de 0,8929 est atteinte pour ces deux seuils (ce sont en quelque sorte des ex-aequo). Nous avons vu dans la partie sur l'indice de Youden généralisé que la valeur de cet indice dépendait du pourcentage de prévalence de la maladie dans la population (π) ainsi que du ratio CFN/CFP. Comme vous pouvez le remarquer sur la Figure 9, le site Internet ne demande pas de fournir la valeur de π . En fait, le site Internet fixe pour π la valeur du pourcentage de prévalence dans l'échantillon étudié. Autrement dit, le site Internet fait l'hypothèse que le pourcentage de prévalence de la maladie dans l'échantillon est égal à π . Notamment, dans l'échantillon, le pourcentage de prévalence de la maladie vaut 28/60, soit 47%. Par conséquent, le site Internet fournit la valeur du seuil optimal en considérant que $\pi=47\%$. Cela a une conséquence importante : si le pourcentage de prévalence de la maladie dans l'échantillon n'est pas voisin de celui de la population, le site Internet easyROC ne permet pas d'identifier le seuil optimal selon l'indice de Youden généralisé (en tout cas, ce seuil ne peut pas être considéré comme « optimal » pour la population visée). Pour remédier à ce problème, j'ai créé un fichier Excel qui permet, à partir des résultats fournis par le site Internet easyROC, de fixer vous-même la valeur de π , et le fichier Excel identifiera le seuil optimal pour cette valeur de π (je peux vous envoyer ce fichier Excel sur demande à l'adresse e-mail loic.desquilbet(at)vet-alfort.fr).

easyROC: a web-tool for ROC curve analysis (ver. 1.3.1)

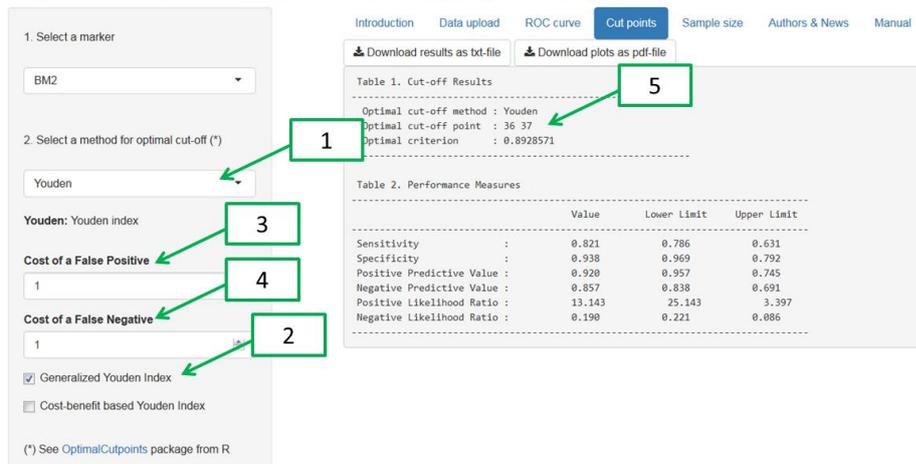


Figure 9. Copie d'écran de la page « Cut points », identification du seuil optimal selon l'indice de Youden généralisé.

Le site Internet permet néanmoins de modifier les valeurs de CFP et de CFN pour sélectionner le seuil optimal selon l'indice de Youden généralisé [Figure 9.3 et Figure 9.4]. Par exemple, en sélectionnant la valeur « 1 » pour « Cost of a false positive » (CFP) et « 3 » pour « Cost of false negative » (CFN), le site Internet identifiera le seuil optimal selon l'indice de Youden généralisé pour une valeur de $\pi=47\%$, et pour un ratio CFN/CFP=3.

d) Identification du seuil optimal selon d'autres méthodes

D'autres méthodes d'identification du seuil optimal sont proposées par le site Internet, mais ne seront pas traitées dans ce tutoriel. Ces méthodes d'identification sont en effet beaucoup moins présentes dans la littérature. Elles sont cependant détaillées dans l'article de Lopez-Raton et coll. (Lopez-Raton et al., 2014).

2. Indicateurs statistiques de la qualité du test de diagnostic

Revenons à la situation où l'on identifie le seuil optimal selon l'indice de Youden (cf. Figure 7). Dans la Table 2 « Performance Measures » [Figure 7.4], on peut y lire différents indicateurs statistiques de la qualité du test diagnostique correspondant au seuil optimal identifié (selon la méthode au choix de l'utilisateur).

Tout d'abord, la valeur de *Se* pour le seuil de 36 [Figure 7.3] est d'environ 0,82 [Figure 7.5]. Cela signifie que, dans l'échantillon, parmi les animaux malades, 82% d'entre eux ont la valeur de BM2 ≥ 36 . L'intervalle de confiance à 95% de *Se* est de [0,63 ; 0,94]. Ainsi, si l'estimation de *Se* n'est pas biaisée, il y a 95% de chances pour que la vraie valeur de *Se* dans la population cible soit comprise entre 0,63 et 0,94. La valeur de *Sp* ainsi que son intervalle de confiance sont indiqués juste en dessous : environ 0,94 [0,72 ; 0,99]. Ainsi, dans l'échantillon, parmi les animaux sains, 94% d'entre eux ont la valeur de BM2 < 36 .

Le site Internet fournit aussi les *VPP* et *VPN*, avec leur intervalle de confiance à 95% [Figure 7.6]. Ainsi, avec un seuil de 36 pour BM2, la *VPP* est de 0,92. Cela signifie que, dans l'échantillon, parmi les animaux dont la valeur de BM2 est ≥ 36 , 92% d'entre eux sont malades. La *VPN* est d'environ 0,86. Cela signifie que, dans l'échantillon, parmi les animaux dont la valeur de BM2 est < 36 , 86% d'entre eux sont sains. Pour que ces valeurs de *VPP* et *VPN* puissent être utilisées dans la population, il faut faire l'hypothèse que le pourcentage de prévalence de la maladie dans la population (π) soit égal (ou très voisin) de celui dans l'échantillon (ici, 47%). Si ce n'est pas le cas, alors il faudra calculer les valeurs de *VPP* et *VPN* à partir des valeurs de *Se* et *Sp* correspondant au seuil optimal identifié, et à

partir de la valeur estimée de π (à partir de ses propres connaissances du terrain), en utilisant les formules fournies dans la partie II.C.1 (ainsi que dans (Altman and Bland, 1994)).

Le site Internet fournit enfin les valeurs des rapports de vraisemblance positif (RV^+) et négatif (RV^-), avec leur intervalle de confiance à 95%. Ici, la valeur de RV^+ est d'environ 13,1 (« Positive likelihood ratio », [Figure 7.7]). Cela signifie qu'il y a 13,1 fois plus de chances qu'un animal ait une valeur de $BM2 \geq 36$ lorsque l'animal est malade que lorsqu'il est sain. La valeur de RV^+ est supérieure à 10, donc l'utilisation du test diagnostique $BM2$ avec une valeur seuil de 36 permet de dire que l'annonce d'un résultat positif de ce test diagnostique (obtenir une valeur supérieure ou égale à 36) augmente fortement la présomption qu'un individu soit malade. Utilisons le nomogramme pour interpréter cette valeur de RV^+ de 13,1. Supposons qu'un animal se présente en clinique vétérinaire avec un tableau clinique qui laisse penser qu'il y a 20% de risques qu'il soit malade. Si la valeur de son test diagnostique $BM2$ est supérieure ou égale à 36 (le test diagnostique est donc positif), alors l'annonce d'un tel résultat (positif) conduit désormais à une probabilité que cet animal soit malade de près de 80% [Figure 10.A].

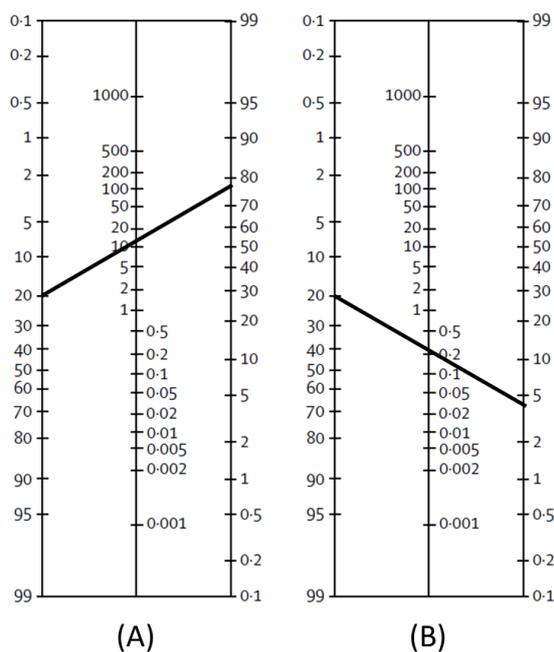


Figure 10. Nomogramme permettant l'interprétation des valeurs de RV^+ (13,4 ; (A)) et RV^- (0,19 ; (B)) fournies par le site Internet easyROC, à partir d'une valeur de probabilité pré-test de 20%.

La valeur de RV^- est de 0,19 (« Negative likelihood ratio », [Figure 7.7]). Cela signifie qu'il y a $1/0,19=5,3$ fois plus de chances qu'un animal ait une valeur de $BM2 < 36$ lorsque l'animal est sain que lorsqu'il est malade. Cette valeur de RV^- est comprise entre 0,1 et 0,2, indiquant un effet « modéré » de l'annonce du résultat du test diagnostique négatif (obtenir une valeur inférieure à 36) sur la probabilité post-test. Utilisons là encore le nomogramme pour interpréter cette valeur de RV^- de 0,19. Supposons le même animal que ci-dessus dont le tableau clinique laisse penser qu'il y a 20% de risques qu'il soit malade. Si la valeur de son test diagnostique $BM2$ est cette fois-ci inférieure à 36 (le test diagnostique est donc négatif), alors l'annonce d'un tel résultat (négatif) conduit désormais à une probabilité que cet animal soit malade de moins de 5% [Figure 10.B].

3. Représentations graphiques de la qualité du test diagnostique

En cliquant sur « Include plots » [Figure 7.8], nous obtenons les résultats de la Figure 11. Là encore, des informations importantes sont présentes sur ces graphiques. Tout d'abord, rappelons que le seuil optimal identifié par l'indice de Youden est de 36 [Figure 11.1]. Le graphique en haut à gauche (intitulé « ROC curve ») représente la courbe ROC que nous avons déjà obtenue précédemment [Figure 5.4]. Mais trois nouvelles informations sont désormais présentes. Tout d'abord, nous voyons où le seuil optimal se situe sur la courbe [Figure 11.2]. Le point représentant ce seuil optimal a pour abscisse $1-Sp$, avec $Sp=0,938$ [Figure 11.3], soit une abscisse de 0,062 [Figure 11.4]. Ce point a pour ordonnée Se , soit 0,821 [Figure 11.3 & Figure 11.5]. Le graphique [Figure 11.6] représente les valeurs de Se (en rouge) et de Sp (en bleu) en fonction de la valeur de BM2 (« Marker Name » en abscisse). Le trait vertical en pointillés correspond au seuil optimal de 36. Le graphique [Figure 11.7] représente la distribution (lissée) de BM2 dans chacun des sous-échantillons des animaux malades (en rouge) et sains (en bleu). Moins ces deux distributions se recoupent, plus le test diagnostique est capable de bien discriminer les animaux malades des animaux sains. Le graphique [Figure 11.8] représente d'une autre façon la distribution de BM2 dans les deux sous-échantillons d'animaux malades et sains. Dans ces deux graphiques [Figure 11.7 & Figure 11.8], le seuil optimal est représenté par un trait en pointillés. Sur le graphique [Figure 11.8], avec un seuil de 36 (seuil optimal selon l'indice de Youden, cf. [Figure 11.1]), on se rend compte que la valeur de Sp (0,938) est plus élevée que celle de Se (0,821) : il y a en effet beaucoup moins de points *au-dessus* du seuil de 36 parmi les animaux sains ($n=1$ individu faux positif) qu'il y en a *en dessous* de la valeur seuil de 36 parmi les animaux malades ($n=5$ individus faux négatifs).

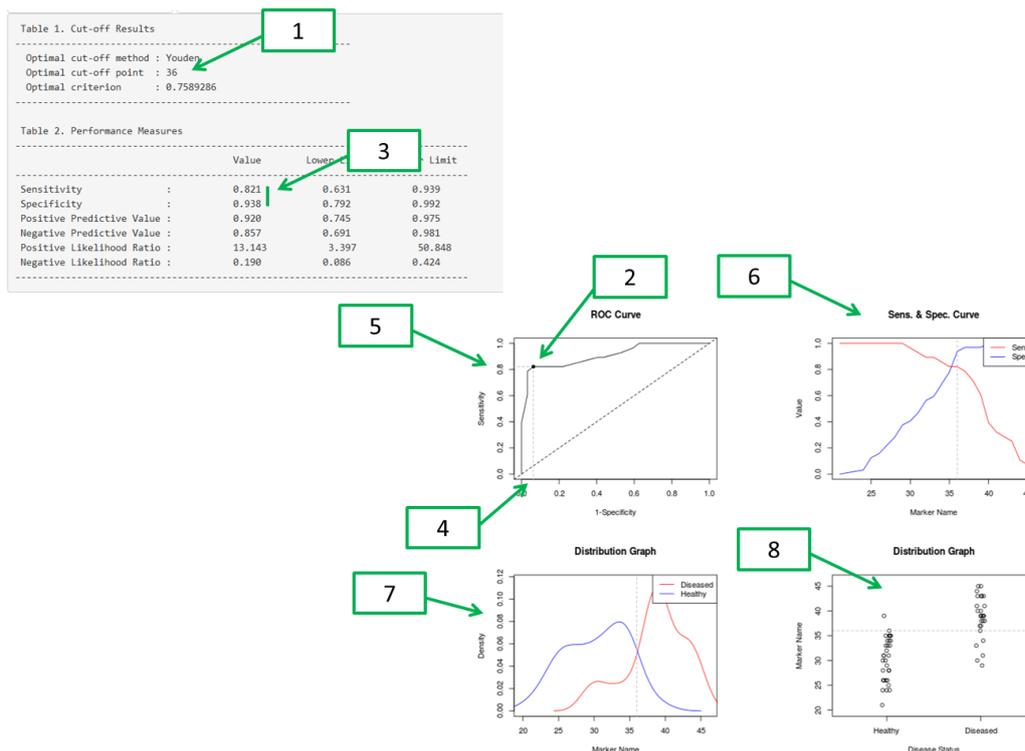


Figure 11. Copie d'écran de la page « Cut points » après avoir cliqué sur « Include plots ».

E. Comparaison de plusieurs courbes ROC

1. Comparaison de deux courbes ROC

Nous allons voir maintenant que le site Internet easyROC permet de comparer statistiquement plusieurs courbes ROC. Sur le site Internet, nous cliquons à nouveau sur « ROC curve » [Figure 12.1], puis dans le champ « Select markers (*) », nous ajoutons « BM1 » en plus de « BM2 » [Figure 12.2]. Dans l'onglet « Statistics » [Figure 12.3], nous voyons les valeurs des deux aires sous la courbe ROC, une pour chaque test diagnostique (colonne « AUC » dans [Figure 12.4]). Juste en dessous, nous avons la juxtaposition des deux courbes ROC (courbe noire pour BM1 et courbe rouge pour BM2) [Figure 12.5]. Nous pouvons voir sur le graphique [Figure 12.5] que la courbe noire (BM1) est plus éloignée de la droite à 45° que la courbe rouge (BM2), ce que traduit le fait que l'aire sous la courbe ROC de BM1 (0,994) est supérieure à celle de BM2 (0,908). Ainsi, dans l'échantillon, le test diagnostique BM1 était plus capable de discriminer les animaux malades des animaux sains que le test diagnostique BM2.



Figure 12. Copie d'écran de la page « ROC curve », onglet « Statistics », avec deux variables quantitatives sélectionnées.

En cliquant sur l'onglet « Multiple Comparisons » [Figure 13.1], la colonne |I-J| correspond à la différence entre les deux aires sous la courbe ROC [Figure 13.2]. Le site Internet teste statistiquement cette différence entre les aires sous ces courbes ROC, et fournit le degré de signification de ce test ($p=0,03 < 0,05$; [Figure 13.3]). Ainsi, l'aire sous la courbe ROC du test diagnostique BM1 est significativement différente de l'aire sous la courbe ROC du test diagnostique BM2. Par conséquent, il y a des chances pour que, dans la population, le test diagnostique BM1 soit plus capable de discriminer les animaux malades des animaux sains que le test diagnostique BM2.

easyROC: a web-tool for ROC curve analysis (ver. 1.3.1)

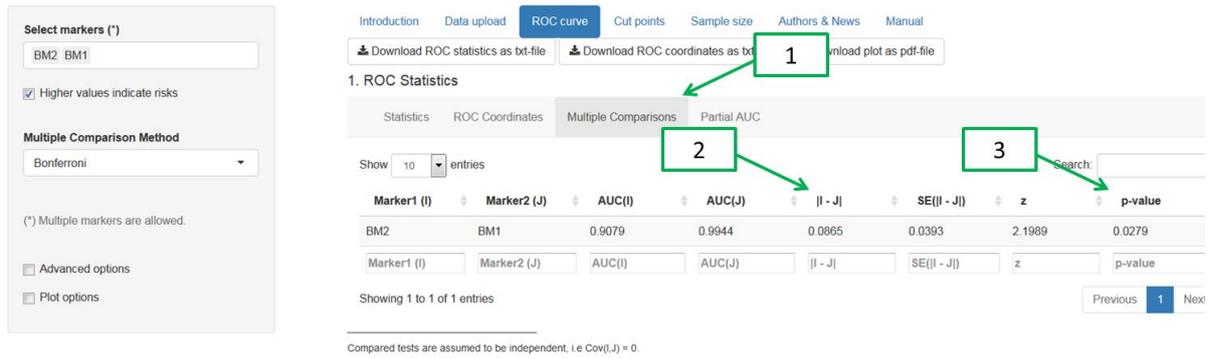
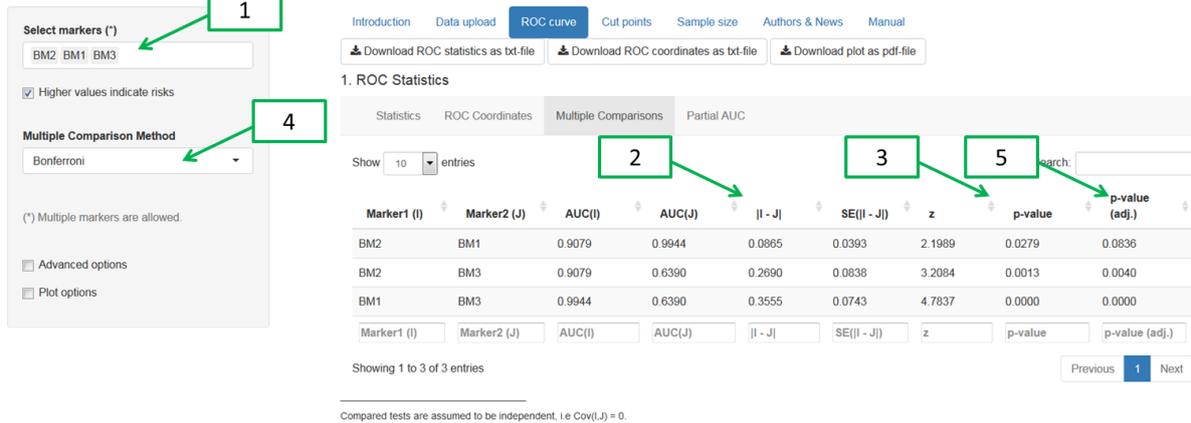


Figure 13. Copie d'écran de la page « ROC curve », onglet « Multiple comparisons », avec deux variables quantitatives sélectionnées.

2. Comparaison de trois courbes ROC ou plus

Nous décidons maintenant de comparer trois aires sous la courbe ROC, une par test diagnostique. Nous ajoutons donc « BM3 » à « BM1 » et « BM2 » dans la liste des variables quantitatives [Figure 14.1].

easyROC: a web-tool for ROC curve analysis (ver. 1.3.1)



2. Plot Output

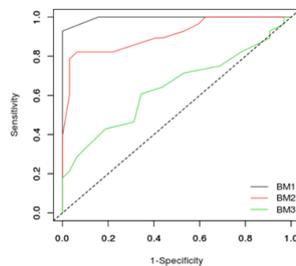


Figure 14. Copie d'écran de la page « ROC curve », onglet « Multiple comparisons », avec trois variables quantitatives sélectionnées.

Le site Internet propose des comparaisons deux à deux d'aires sous la courbe ROC [Figure 14.2]. Le degré de signification associé au test statistique testant la différence entre deux aires sous la courbe ROC est présenté sous la colonne « p-value » [Figure 14.3]. Cependant, ce degré de signification n'est pas correctement estimé, puisqu'il ne prend pas en compte la situation de tests statistiques multiples dans laquelle nous sommes ici (Bender and Lange, 2001). Brielèvement, nous sommes en effet ici dans

cette situation de tests statistiques multiples car trois tests statistiques ont été utilisés : aire « BM2 » *versus* aire « BM1 », aire « BM2 » *versus* aire « BM3 », et aire « BM1 » *versus* aire « BM3 », sans savoir quelle aire allait être significativement différente d'une autre (il s'agit ici d'une sorte de « pêche au degré de signification significatif »). Si, en vrai, les trois tests diagnostiques sont purement équivalents dans leur capacité à discriminer les animaux malades des animaux sains, il y a de bonnes chances pour que, dans l'échantillon, au moins un des trois tests statistiques comparant les tests diagnostiques deux à deux soit malgré tout significatif⁹. (Lors d'une situation de tests statistiques multiples, on dit que le risque d'erreur de 1^{ère} espèce $\alpha=0,05$ est sur-évalué (Dmitrienko and D'Agostino, 2013).) Une des méthodes pour réaliser ces tests statistiques multiples sans que le risque d'erreur de 1^{ère} espèce α ne soit sur-évalué est celle de Bonferroni (Perneger, 1998). Si n est le nombre de tests statistiques réalisés (ici, $n=3$), alors la correction de Bonferroni consiste à multiplier par n le degré de signification, puis de comparer ce degré de signification corrigé à la valeur de 0,05 (comme on le fait classiquement) pour savoir si le test statistique est significatif. Le site Internet propose par défaut cette méthode de correction du risque d'erreur de 1^{ère} espèce [Figure 14.4]. Ici, si nous multiplions les degrés de signification de la colonne « p-value » [Figure 14.3] par 3 (car $n=3$), nous obtenons les degrés de signification de la colonne « p-value (adj.) » qui sont corrigés par la correction de Bonferroni [Figure 14.5]. Après correction, nous nous apercevons que l'aire sous la courbe ROC du test diagnostique BM2 n'est plus significativement différente de l'aire sous la courbe ROC du test diagnostique BM1. Ce résultat n'est pas contradictoire avec ce que nous avons trouvé avant d'ajouter « BM3 » dans la liste des variables quantitatives [Figure 14.1] : avant d'ajouter « BM3 », nous voulions répondre à la question suivante (et nous devons avoir une bonne raison de nous poser cette question) : « est-ce que les tests diagnostiques BM1 et BM2 sont réellement différents dans leur capacité à discriminer les animaux malades des animaux sains ? » En ajoutant « BM3 », là, nous sommes beaucoup plus dans une démarche exploratoire, avec la question sous-jacente suivante : « quel est, des trois tests diagnostiques, celui qui se distingue des deux autres dans sa capacité à discriminer les animaux malades des animaux sains ? » Cette démarche exploratoire a un coût : la sur-évaluation (qui doit être corrigée) du risque d'erreur de 1^{ère} espèce α .

IV. Bibliographie

- Altman, D.G. and Bland, J.M., 1994. Diagnostic tests 2: Predictive values. *BMJ*. 309, 102.
- Bender, R. and Lange, S., 2001. Adjusting for multiple testing--when and how? *J Clin Epidemiol*. 54, 343-9.
- Davidson, M., 2002. The interpretation of diagnostic test: a primer for physiotherapists. *Aust J Physiother*. 48, 227-32.
- Delacour, H., Servonnet, A., Perrot, A., Vigezzi, J.F. and Ramirez, J.M., 2005. [ROC (receiver operating characteristics) curve: principles and application in biology]. *Ann Biol Clin (Paris)*. 63, 145-54.
- Detilleux, J., Arendt, J., Lomba, F. and Leroy, P., 1999. Methods for estimating areas under receiver-operating characteristic curves: illustration with somatic-cell scores in subclinical intramammary infections. *Prev Vet Med*. 41, 75-88.
- Dmitrienko, A. and D'Agostino, R., Sr., 2013. Traditional multiplicity adjustment methods in clinical trials. *Stat Med*. 32, 5172-218.
- Drobatz, K.J., 2009. Measures of accuracy and performance of diagnostic tests. *J Vet Cardiol*. 11 Suppl 1, S33-40.
- Fagan, T.J., 1975. Letter: Nomogram for Bayes theorem. *N Engl J Med*. 293, 257.

⁹ Cette probabilité vaut $(1 - (1 - 0,05)^3) = 14\%$ de chances, si le risque d'erreur de 1^{ère} espèce α est fixé à 0,05.

- Goksuluk, D., Korkmaz, S., Zararsiz, G. and Karaagaoglu, A.E., 2016. easyROC: An Interactive Web-tool for ROC Curve Analysis Using R Language Environment. *The R Journal*. 8, 213-230.
- Greiner, M., Pfeiffer, D. and Smith, R.D., 2000. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Prev Vet Med*. 45, 23-41.
- Grimes, D.A. and Schulz, K.F., 2002. Uses and abuses of screening tests. *Lancet*. 359, 881-4.
- Grimes, D.A. and Schulz, K.F., 2005. Refining clinical diagnosis with likelihood ratios. *Lancet*. 365, 1500-5.
- Hayden, S.R. and Brown, M.D., 1999. Likelihood ratio: A powerful tool for incorporating the results of a diagnostic test into clinical decisionmaking. *Ann Emerg Med*. 33, 575-80.
- Lopez-Raton, M., Rodriguez-Alvarez, M.X., Cadarso-Suarez, C. and Gude-Sampedro, F., 2014. OptimalCutpoints: An R Package for Selecting Optimal Cutpoints in Diagnostic Tests. *J Stat Software*. 61.
- Lusted, L.B., 1971. Decision-making studies in patient management. *N Engl J Med*. 284, 416-24.
- Perkins, N.J. and Schisterman, E.F., 2006. The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. *Am J Epidemiol*. 163, 670-5.
- Perneger, T.V., 1998. What's wrong with Bonferroni adjustments. *BMJ*. 316, 1236-8.
- Schisterman, E.F., Perkins, N.J., Liu, A. and Bondell, H., 2005. Optimal cut-point and its corresponding Youden Index to discriminate individuals using pooled blood samples. *Epidemiology*. 16, 73-81.
- Soreide, K., 2009. Receiver-operating characteristic curve analysis in diagnostic, prognostic and predictive biomarker research. *J Clin Pathol*. 62, 1-5.
- Sweig, M.H. and Campbell, G., 1993. Receiver-Operating Characteristic (ROC) Plots: A Fundamental Evaluation Tool in Clinical Medicine. *Clin Chem*. 39, 561-77.
- Swets, J.A., 1988. Measuring the accuracy of diagnostic systems. *Science*. 240, 1285-93.
- Timsit, E., Leguillette, R., White, B.J., Larson, R.L. and Buczinski, S., 2018. Likelihood ratios: an intuitive tool for incorporating diagnostic test results into decision-making. *J Am Vet Med Assoc*. 252, 1362-1366.
- Youden, W.J., 1950. Index for rating diagnostic tests. *Cancer*. 3, 32-5.