

Sensibilité, spécificité, courbe ROC etc.

Gilbert Saporta
Conservatoire National des Arts et Métiers

Gilbert.saporta@cnam.fr
<http://cedric.cnam.fr/~saporta>

octobre 2012

1. Dépistage, diagnostic et tests

Example 6.16. Remein and Wilkerson [1961] considered a number of screening tests for diabetes. They had a group of consultants establish criteria, their gold standard, for diabetes. On each of a number of days, they recruited patients being seen in the outpatient department of the Boston City Hospital for reasons other than suspected diabetes. The table below presents results on the Folin–Wu blood test used 1 hour after a test meal and using a blood sugar level of 150 mg per 100 mL of blood sugar as a positive test.

Test	Diabetic	Nondiabetic	Total
+	56	49	105
–	14	461	475
Total	70	510	580

Test	Diabetic	Nondiabetic	Total
+	56	49	105
-	14	461	475
Total	70	510	580

- Faux positifs : 49
- Faux négatifs: 14
- Vrais positifs: 56
- Vrais négatifs: 461

Test	Diabetic	Nondiabetic	Total
+	56	49	105
-	14	461	475
Total	70	510	580

- **Sensibilité (Se)** : proportion de vrais positifs parmi les personnes à dépister:
 $56/70=80\%$
- **Spécificité (Sp)**: proportion de vrais négatifs chez les non-malades:
 $461/510=90\%$

- Facile d'avoir une bonne sensibilité: déclarer tout le monde positif... mais peu spécifique
- Un bon test doit être sensible et spécifique
- Sachant que le test est positif, quelle est la probabilité d'être vraiment malade?

valeur prédictive positive ou VPP

On définit également **la valeur prédictive négative**

Calcul de la VPP

- Nécessite de connaître la prévalence p
- VPP doit être $> p$

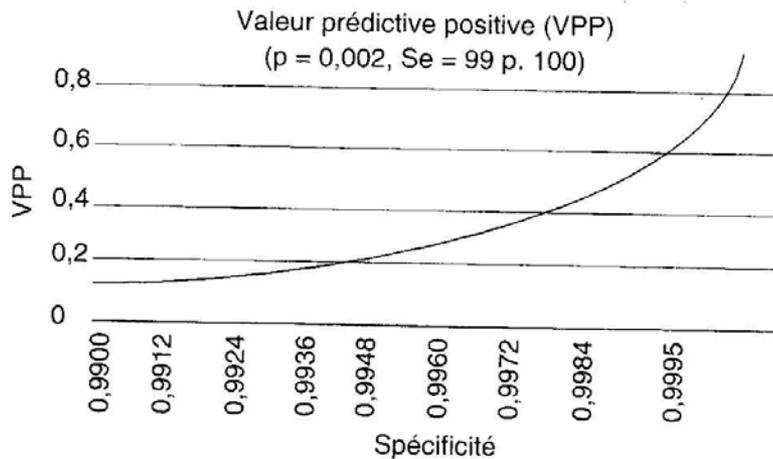
Application de la formule de Bayes

$$VPP = P(M / +) = \frac{P(+ / M)P(M)}{P(+ / M)P(M) + P(+ / \bar{M})P(\bar{M})}$$

$$VPP = \frac{pSe}{pSe + (1 - Sp)(1 - p)}$$

- Exercice 1 (Valleron, 2007, p.20):
Un test a une sensibilité de 0.99, une spécificité de 0.9. La maladie touche un sujet sur 100. Quelles sont les proportions de faux négatifs et faux positifs dans la population? Calculer la VPP et la VPN
- Exercice 2 : calculer l'IC à 95% de la VPN dans l'exemple introductif

- Dépistage des maladies rares
 - Si p est très faible la VPP est faible si la spécificité n'est pas TRES élevée



© Valleron, 2007

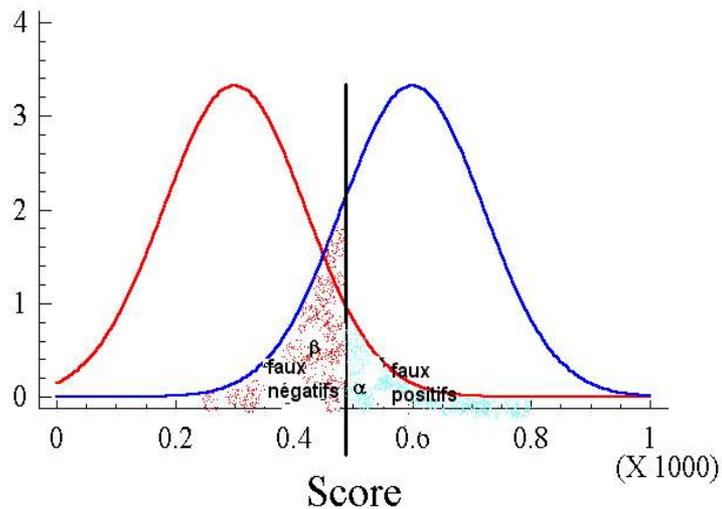
Figure 3-2 Variation de la valeur prédictive positive (fraction parmi les sujets trouvés positifs par le test de ceux qui sont réellement séropositifs) en fonction de la spécificité. On a supposé une sensibilité de 99 p. 100 et une prévalence du VIH de 2 p. 1 000.

- dépistage systématique ou dépistage orienté?

2. La courbe ROC

- Receiver Operating Characteristic (transmission de signal)
- Un test s'appuie souvent sur une mesure quantitative continue S
 - Au-delà d'un seuil s , on est déclaré positif
 - Comment choisir le seuil s ?
 - Comment varient la sensibilité et la spécificité en fonction du seuil s ?

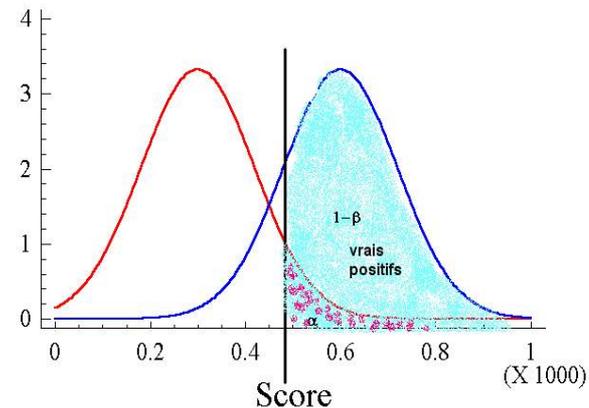
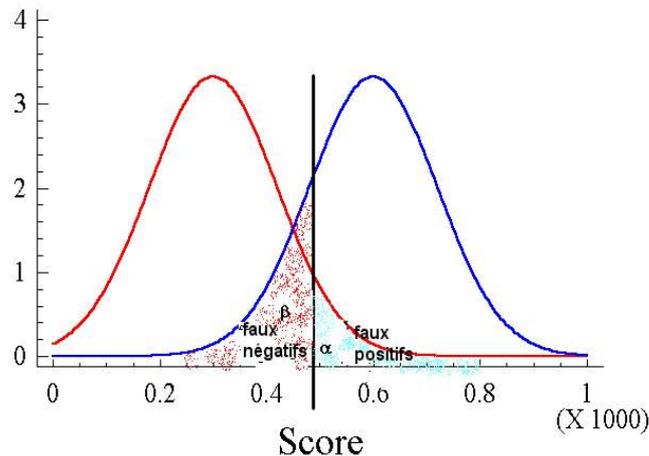
- **Retour sur la théorie des tests:** groupe à détecter pour des valeurs élevées de S



Risque de première espèce = α
taux de faux positifs

Risque de deuxième espèce = β
taux de faux négatifs

- Sensibilité $1-\beta = P(S > s/M)$: % de vrais positifs = puissance du test
- Spécificité $1-\alpha = P(S < s/NM)$: % de vrais négatifs



Courbe ROC

- Evolution de $1-\beta$ puissance du test en fonction de α , risque de première espèce lorsque le seuil varie
- Proportion de vrais positifs en fonction de la proportion de faux positifs

- Exemple : répartition du pic de fièvre pour les vraies et fausses gripes

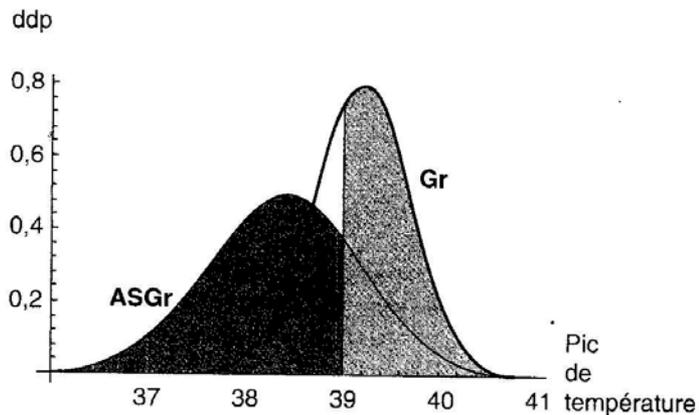


Figure 3-4 Distribution de probabilité du pic de fièvre chez des sujets ayant une « vraie » grippe (Gr) et une « fausse » grippe (autres syndromes grippaux, ASGr).

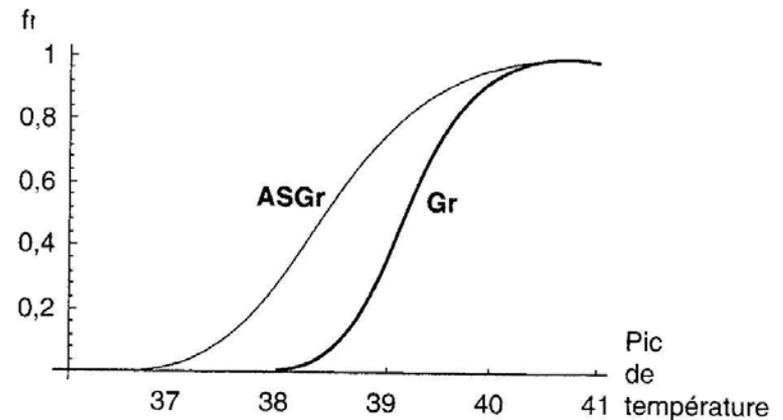


Figure 3-5 Fonction de répartition du pic de fièvre chez des sujets ayant une « vraie » grippe (Gr) et une « fausse » grippe (autres syndromes grippaux, ASGr).

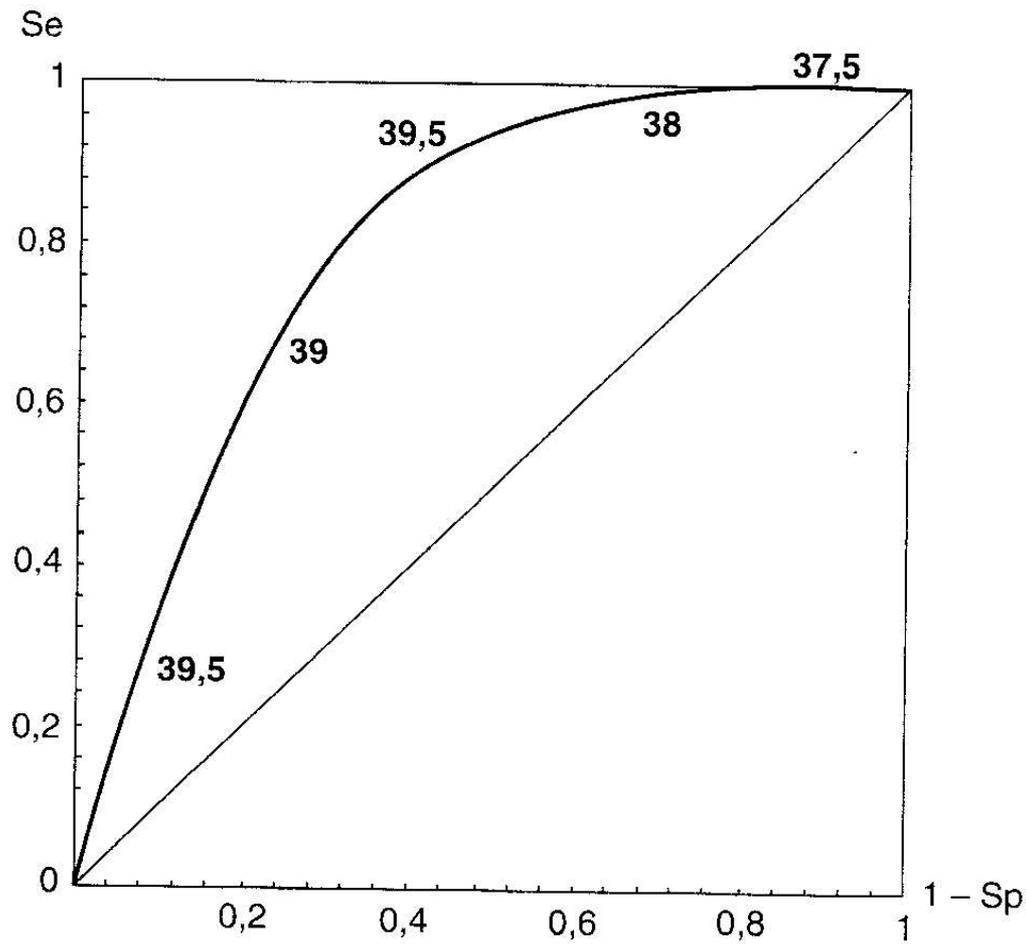
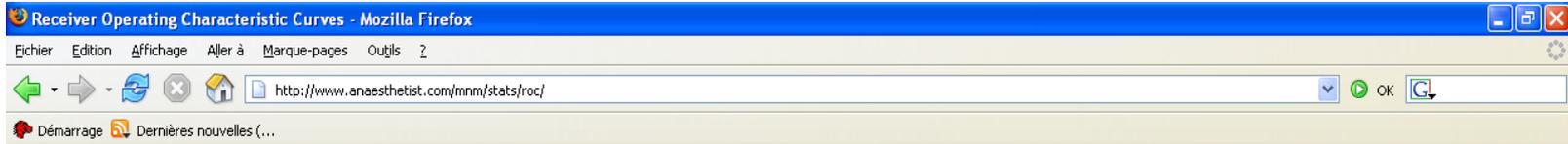


Figure 3-6 Exemple du diagnostic de la « vraie » grippe versus celui de la « fausse grippe ». Courbe ROC correspondant aux données des figures 3-4 et 3-5.

■ Un site: <http://www.anaesthetist.com/mnm/stats/roc/>

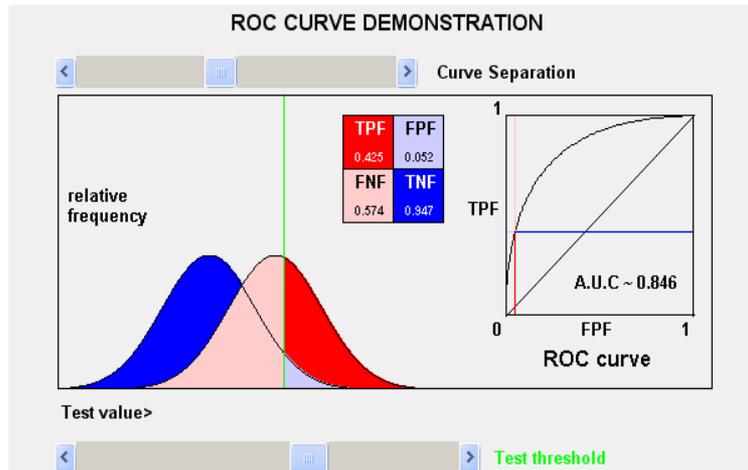


Playing with ROCs

In this section we will fool around with ROCs. We will:

1. [Create](#) ROC curves;
2. Find out why the area under the ROC curve is [non-parametric](#), and why this is important;
3. Learn to calculate required [sample sizes](#);
4. Compare the areas under [two ROC curves](#);
5. Examine the effects of [noise](#), a [bad 'gold standard'](#), and [other sources of error](#).

Let's play some more. In the following example, see how closely the two curves are superimposed, and how flat the corresponding ROC curve is! This demonstrates an important property of ROC curves - the greater the overlap of the two curves, the smaller the area under the ROC curve.



Vary the curve separation using the upper "slider" control, and see how the ROC curve changes. When the curves overlap almost totally the ROC curve turns into a diagonal line from the bottom left corner to the upper right corner. What does this mean?

Comparer des courbes ROC correspondant à plusieurs tests

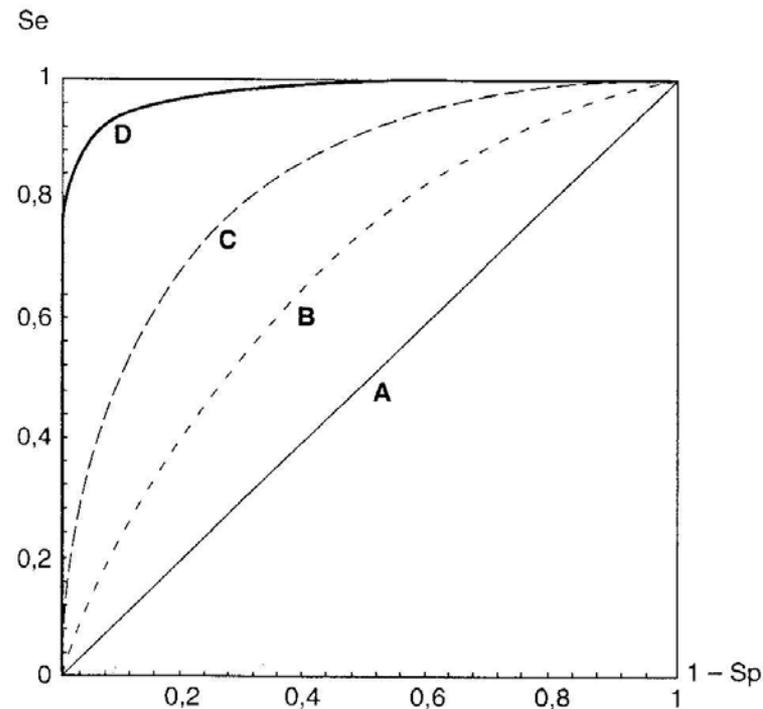
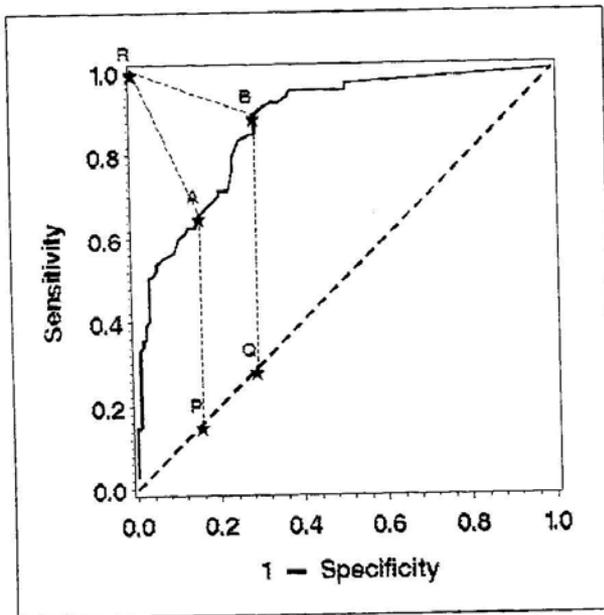


Figure 3-7 Différentes courbes ROC. A : test sans intérêt diagnostique. B : mauvais test diagnostique. C : test meilleur que B. D : bon test diagnostique.

Choix du seuil optimal à l'aide de la courbe ROC

Figure 3.5 Choosing the Optimal Threshold Using the Empirical ROC Curve



Le seuil correspondant

- au point le plus proche de l'idéal (1;1)
- au point le plus loin de la diagonale

Choix du seuil optimal selon la minimisation du coût d'erreur

C_M coût de déclarer malade à tort

C_{NM} coût de déclarer non malade à tort

Espérance du coût global pour un seuil s :

$$p(1-Se(s)) C_{NM} + (1-p)(1-Sp(s)) C_M$$

$Se(s) = F_M(s)$ fonction de répartition des malades

$1-Sp(s)=F_{NM}(s)$ fonction de répartition des non malades

Coût minimum si la dérivée par rapport à s est nulle

$$pf_M(s)C_{NM} - (1-p)f_{NM}(s)C_M = 0$$

$$\frac{f_M(s)}{f_{NM}(s)} = \frac{C_M}{C_{NM}} \frac{1-p}{p}$$

Rapport des densités = pente de la tangente à la courbe ROC

Exercice Tracer la courbe ROC correspondant aux données suivantes:

		Serum Calcium Levels mg/dL					Total
		8 mg/dL	9 mg/dL	10 mg/dL	11 mg/dL	12 mg/dL	
Disease Status	Negative	40	7	4	2	1	54
	Positive	2	3	5	8	17	35
	Total	42	10	9	10	18	89

3. L'AUC ou surface sous la courbe ROC (Area Under Curve)

- Plus l'AUC est grand, meilleur est le test.
- Fournit un ordre partiel sur les tests
- Problème si les courbes ROC se croisent
- Courbe ROC et surface sont des mesures intrinsèques de séparabilité, invariantes pour toute transformation monotone croissante de la mesure S

- Surface théorique sous la courbe ROC:
 $P(X_1 > X_2)$ si on tire au hasard et
indépendamment une observation M et
une observation nonM

$$AUC = \int_{s=+\infty}^{s=-\infty} (1 - \beta(s)) d\alpha(s)$$

- Comment estimer l'AUC?
 - mesurer la surface
 - estimer $P(X_1 > X_2)$

Cas de données individuelles: on mesure
S pour n_1 individus M et n_2 individus
nonM

Pour un test parfait, les mesures des n_1
individus M sont supérieures aux mesures
des n_2 individus nonM

Mesures de concordance

- Paires formées d'une obs M et d'une NM .
 - Nombre de paires $t = n_1 n_2$ $n = n_1 + n_2$
- Si l'observation M a une valeur de S plus grande que celle de l'observation NM la paire est dite concordante.
- nc = nombre de paires concordantes; nd = nombre de paires discordantes; $t - nc - nd$ = nombre d'ex-aequo

- La surface sous la courbe est estimée par le pourcentage de paires concordantes
- Lien avec les tests U de Mann-Whitney et W de Wilcoxon $n_c = U$

$$U+W = n_1n_2 + 0.5n_1(n_1+1)$$

$$\text{AUC} = U/n_1n_2$$

Autres mesures

- D de Somers = $(nc - nd) / t$
- Gamma = $(nc - nd) / (nc + nd)$
- Tau-a = $2 (nc - nd) / n(n-1)$

Indice de Gini

- Double de la surface entre la courbe ROC et la diagonale

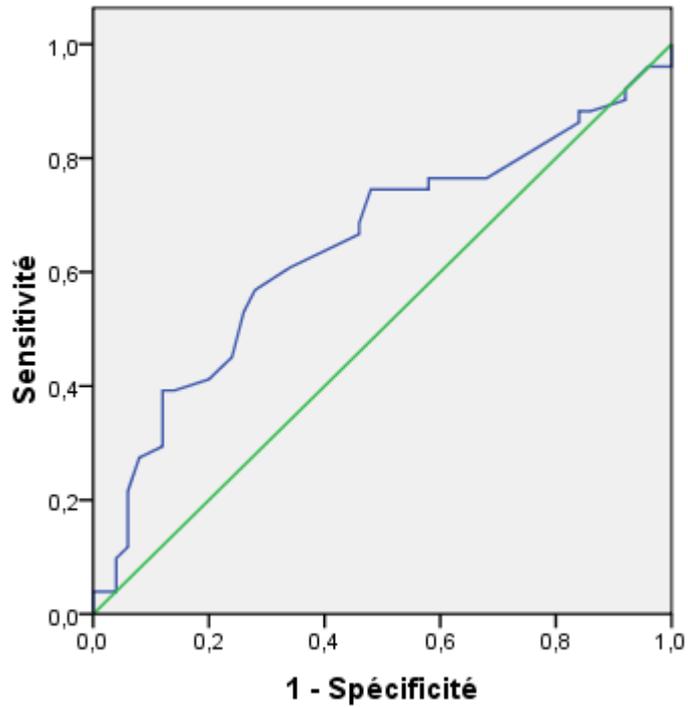
$$G=2AUC-1$$

- En l'absence d'ex-aequo: G identique au D de Somers

Exemple « Infarctus » $n_1=51, n_2=50$

Obs	C	PRONO	FRCAR	INCAR	INSYS	PRDIA	PAPUL	PVENT	REPUL
1	2	SURVIE	90	1.71	19.0	16	19.5	16.0	912
2	1	DECES	90	1.68	18.7	24	31.0	14.0	1476
3	1	DECES	120	1.40	11.7	23	29.0	8.0	1657
4	2	SURVIE	82	1.79	21.8	14	17.5	10.0	782
5	1	DECES	80	1.58	19.7	21	28.0	18.5	1418
6	1	DECES	80	1.13	14.1	18	23.5	9.0	1664
7	2	SURVIE	94	2.04	21.7	23	27.0	10.0	1059
8	2	SURVIE	80	1.19	14.9	16	21.0	16.5	1412
9	2	SURVIE	78	2.16	27.7	15	20.5	11.5	759
10	2	SURVIE	100	2.28	22.8	16	23.0	4.0	807
11	2	SURVIE	90	2.79	31.0	16	25.0	8.0	717
12	2	SURVIE	86	2.70	31.4	15	23.0	9.5	681
13	2	SURVIE	80	2.61	32.6	8	15.0	1.0	460
14	2	SURVIE	61	2.84	47.3	11	17.0	12.0	479
15	2	SURVIE	99	3.12	31.8	15	20.0	11.0	513
16	2	SURVIE	92	2.47	26.8	12	19.0	11.0	615
17	2	SURVIE	96	1.88	19.6	12	19.0	3.0	809
18	2	SURVIE	86	1.70	19.8	10	14.0	10.5	659
19	2	SURVIE	125	3.37	26.9	18	28.0	6.0	665
20	2	SURVIE	80	2.01	25.0	15	20.0	6.0	796

Courbe ROC

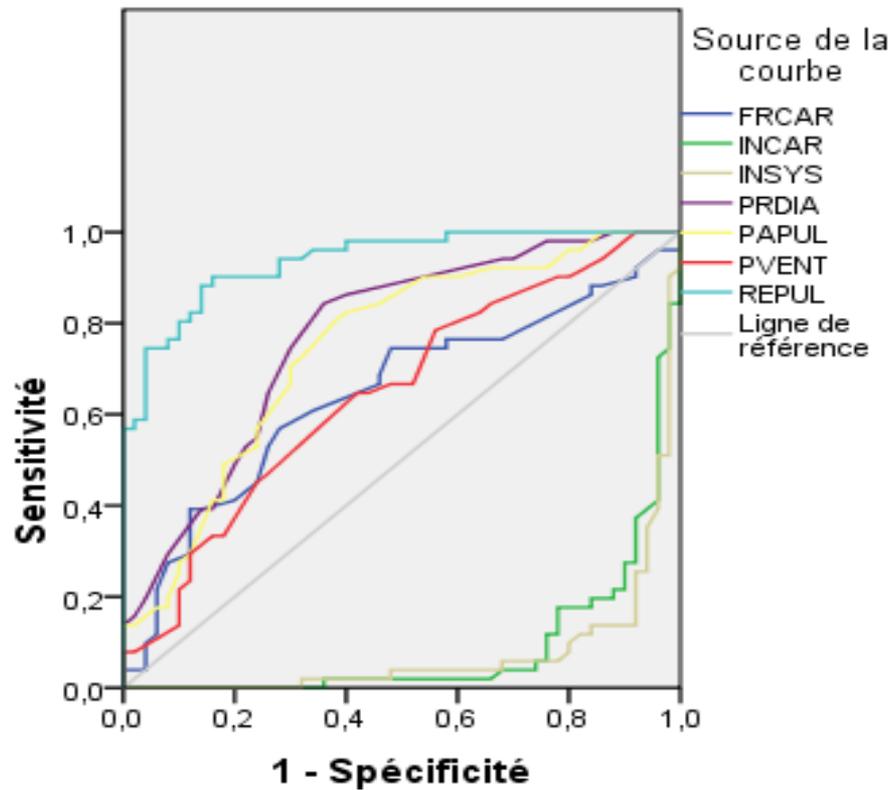


Zone sous la courbe

Variable(s) de
résultats tests:
FRCAR

Zone
,645

Courbe ROC

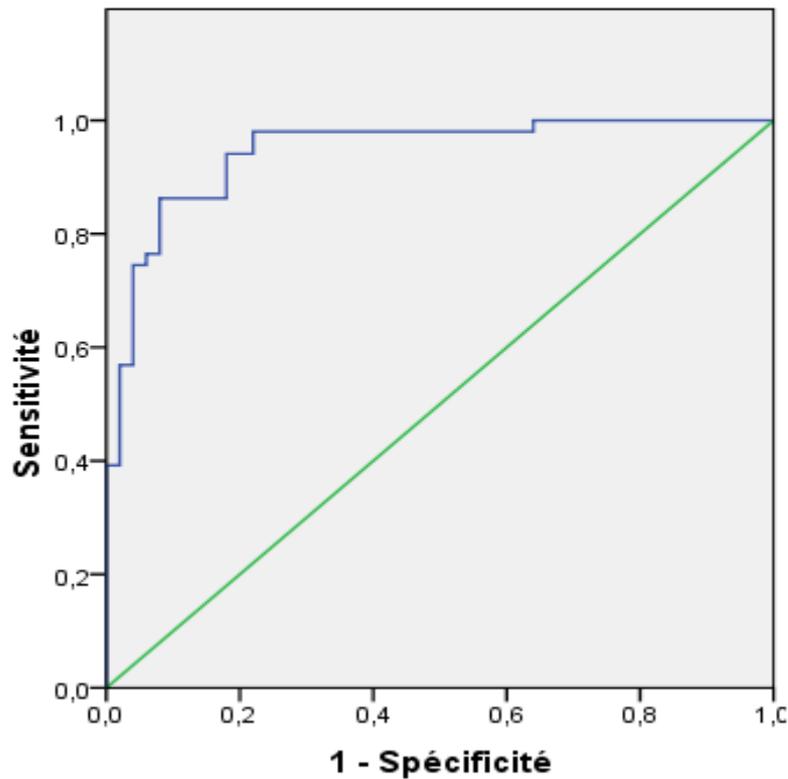


Zone sous la courbe

Variable(s) de résultats tests	Zone
FRCAR	,645
INCAR	,088
INSYS	,075
PRDIA	,772
PAPUL	,746
PVENT	,648
REPUL	,937

Score optimal

Courbe ROC



Zone sous la courbe

Variable(s) de résultats tests:scorfish

Zone	Erreur Std. ^a	Signif. asymptotique ^b	Intervalle de confiance 95% asymptotique	
			Borne inférieure	Borne supérieure
,945	,022	,000	,903	,988

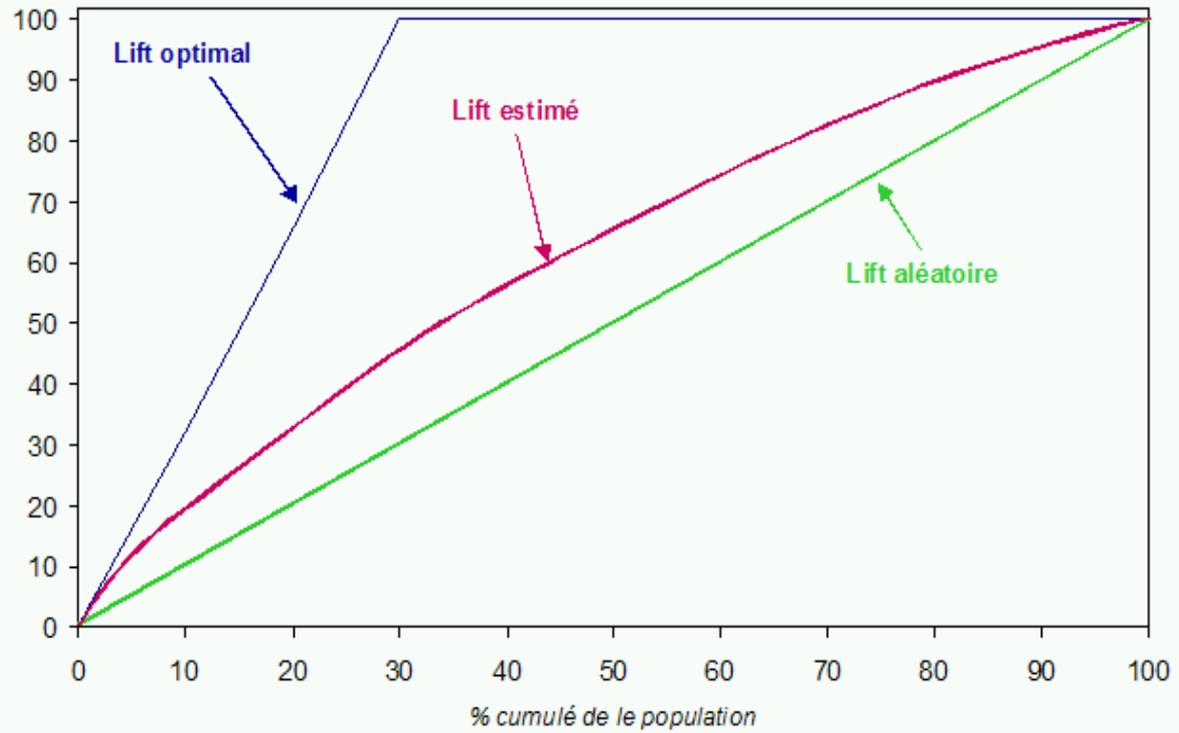
a. Dans l'hypothèse non-paramétrique

b. Hypothèse nulle f : zone vraie = 0.5

4. Ciblage et courbe de lift

- Dépistage systématique à l'aide d'un marqueur numérique S
- Quelle fraction de la population à risque (de prévalence p) détecte t-on en ciblant les $x\%$ de la population ayant le marqueur supérieur à un seuil s ?

% de la cible



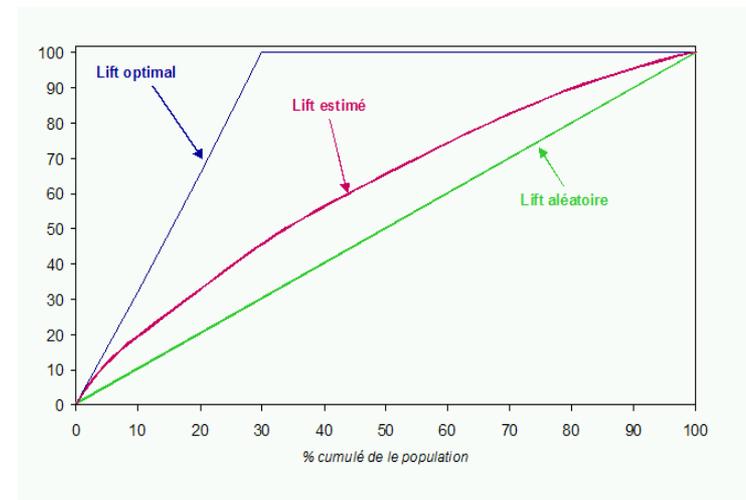
Surface sous la courbe lift

- Pourcentage d'individus ayant un score $> s$

- Surface $p(1-\beta) + (1-p)\alpha$
$$L = \int (1-\beta) d\{p(1-\beta) + (1-p)\alpha\} =$$
$$\left[p \int (1-\beta) d(1-\beta) \right] + \left[(1-p) \int (1-\beta) d\alpha \right]$$
$$= \frac{p}{2} + (1-p)AUC$$

Coefficient Ki (Kxen)

- $K_i = (\text{surface entre lift estimé et aléatoire}) / (\text{surface entre lift idéal et aléatoire})$
- $K_i = 2(\text{surface ROC}) - 1$



$$K_i = \frac{L - \frac{1}{2}}{\frac{1-p}{2}} = \frac{p + 2(1-p)AUC - 1}{1-p} = 2AUC - 1 = Gini$$