

Créer et Valider une signature diagnostique à partir de biopuces

Partie 1 : Evaluer les capacités prédictive d'un marqueur

Yohann.Foucher@univ-nantes.fr

Equipe d'Accueil 4275 "Biostatistique, recherche clinique et mesures subjectives en santé", Université de Nantes

MASTER 1 Bionformatique et Biostatistique - UE données omics

Plan

Plan

- Les nouveaux outils d'analyse à haut débit permettent la mesure simultanée de milliers caractéristiques d'un individu
ex : Puces ADN : elles peuvent mesurer plusieurs milliers de gènes.
- Objectif : Identifier un ensemble de marqueur qui permettent de discriminer deux types d'individus.
ex : Gènes associés à la rémission d'un cancer après traitement.

Exemple : Rémission d'une leucémie *

- 128 patients avec un diagnostic de leucémie aiguë lymphoblastique (ALL).
- Des puces contenant l'expression de 12625 gènes ont été collectés.
- Objectif de l'exemple : Identifier un ensemble de gènes prédisant la rémission après traitement.

*. S. Chiaretti et al. Gene expression profile of adult T-cell acute lymphocytic leukemia identifies distinct subsets of patients with different response to therapy and survival. Blood, 2004, Vol. 103, No. 7

```
> library(Biobase)
> library(multtest)
> library(ALL)
> library(genefilter)
> data(ALL)
> dim(exprs(ALL))

[1] 12625   128

> X<-exprs(ALL)
> pheno<-pData(ALL)
>
```

```
> # Etape de filtration
>
> ffun<-filterfun(pOverA(p=0.2, A=100), cv(a=0.7, b=4))
> filt<-genefilter(2^X, ffun)
> sum(filt)

[1] 430

> as.character(pheno$remission)[1:10]

[1] "CR" "CR" "CR" "CR" "CR" "CR" "CR" "CR" "CR" "CR"

> rem<-1*(as.character(pheno$remission)=="CR")
> table(rem)

rem
 0  1
15 99

> filtX<-X[filt,!is.na(rem)]
> dim(filtX)

[1] 430 114

> rem <- rem[!is.na(rem)]
> length(rem)

[1] 114
```

```
> # selection des gènes  
>  
> rem.boot<-MTP(X = filtX, Y=rem, test = "t.twosamp.equalvar",  
+ robust=TRUE, alpha=0.05, B = 20, get.cutoff = TRUE)  
  
running bootstrap...  
iteration = 20  
  
> sum(rem.boot@adjp<=0.05)  
  
[1] 10  
  
> sum(rem.boot@adjp<=0.05)/sum(filt)  
  
[1] 0.02325581  
  
>
```

Problèmes

- Ces deux gènes sont-ils prédictifs ?

$$\text{logit}(\pi) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \epsilon$$

où :

- π est une probabilité associée à la modalité d'intérêt d'une variable binaire : $\pi = Pr(Y = 1 | X_1, X_2, \dots, X_p)$
 - $Y = 1$ si rémission ou 0 sinon.
- X_1, X_2, \dots, X_p sont des biomarqueurs explicatifs
 - X_1 et X_2 pour les deux gènes.
- β_0 est l'intercept
- $\beta_1, \beta_2, \dots, \beta_p$ sont des coefficients de régression
 - β_1 pour l'effet du gène 1.
 - β_2 pour l'effet du gène 2.
- ϵ est l'erreur du modèle

- X_k est une variable quantitative, alors $\exp(\beta_k)$ correspond à l'odds ratio pour une augmentation de 1 unité du biomarqueur k .
 - Si $\beta_k > 0$ alors la probabilité d'appartenir à modalité d'intérêt augmente quand l'expression du gène augmente.
 - Si $\beta_k < 0$ alors la probabilité d'appartenir à modalité d'intérêt diminue quand l'expression du gène augmente.
- Il est donc naturel d'écrire le score prédictif suivant :

$$S = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

- Plus S est grand, plus le sujet a de chance d'appartenir à $Y = 1$.
- Plus S est grand, plus le sujet a de chance d'appartenir à $Y = 0$.

Exemple : Rémission d'une leucémie

```
> dim(filtX)
[1] 430 114
> X1 <- filtX[rem.boot@adjp<=0.05,][1,]
> X2 <- filtX[rem.boot@adjp<=0.05,][2,]
> summary(glm(rem ~ X1 , binomial))$coef
      Estimate Std. Error  z value  Pr(>|z|)
(Intercept)  9.0441043  2.3182710  3.901228 9.570588e-05
X1           -0.9616988  0.2978995 -3.228265 1.245433e-03
> summary(glm(rem ~ X2 , binomial))$coef
      Estimate Std. Error  z value  Pr(>|z|)
(Intercept) -5.539538  2.9867187 -1.854724 0.06363570
X2           1.245638  0.5136125  2.425248 0.01529794
>
```

Attention : ces p-values ne sont pas corrigées.

Exemple : Rémission d'une leucémie

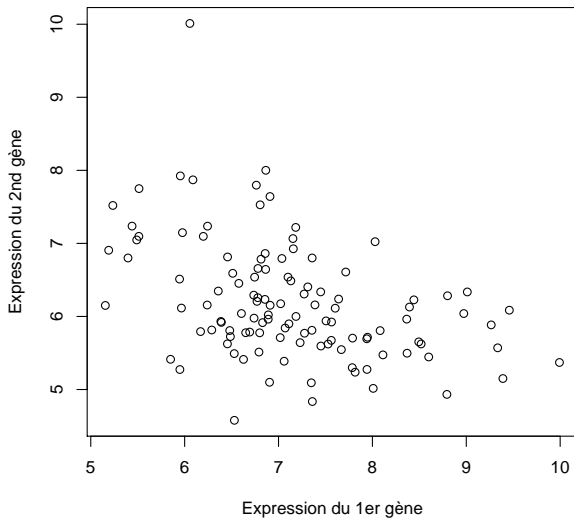
```
> summary(logit.model <- glm(rem ~ X1 + X2 , binomial))$coef
```

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.5000572	4.4915031	0.5566193	0.57778759
X1	-0.7980853	0.3149071	-2.5343514	0.01126557
X2	0.8972820	0.5694723	1.5756375	0.11510936

```
> plot(X1, X2, xlab="Expression du 1er gène", ylab="Expression du 2nd gène")
>
```

Pourquoi une telle différence dans le modèle multivarié : colinéarité ?

Exemple : Rémission d'une leucémie



Exemple : Rémission d'une leucémie

```
> logit.model$coef

(Intercept)          X1          X2
    2.5000572   -0.7980853   0.8972820

> S <- logit.model$coef[2] * X1 + logit.model$coef[3] * X2
> summary(S[rem==1])

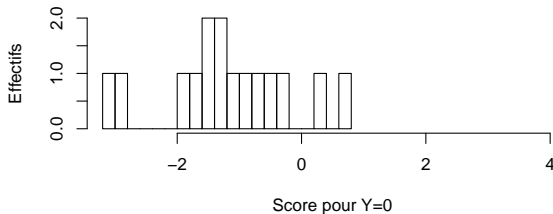
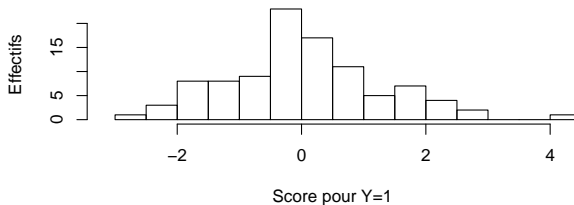
    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-2.59100 -0.68480 -0.04071  0.03929  0.62630  4.15000

> summary(S[rem==0])

    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-3.1560 -1.6460 -1.2140 -1.2010 -0.5997  0.6492

> par(mfrow=c(2,1))
> hist(S[rem==1], xlab="Score pour Y=1", ylab="Effectifs",
+      nclass=20, xlim=range(S), main="")
> hist(S[rem==0], xlab="Score pour Y=0", ylab="Effectifs",
+      nclass=20, xlim=range(S), main="")
>
```

Exemple : Rémission d'une leucémie



- Sélection des gènes à partir de p-values (MTP) : probabilité que la différence observée soit due au hasard. Ce n'est pas un indicateur des capacités pronostiques d'un gène.
- Les gènes pourraient être initialement sélectionnés à partir de l'OR univarié : mais l'OR n'est pas un bon indicateur pronostique.
- Les p-values du modèle logistique multivarié final sont fausses car elles ne prennent pas en compte la répétition initiale des tests.
- Problème des colinéarités entre beaucoup de gènes.
- Comment évaluer les capacités pronostiques de la signature finale ?
- ...

[†]. Limitations of the Odds Ratio in Gauging the Performance of a Diagnostic, Prognostic, or Screening Marker. American Journal of Epidemiology. Vol. 159, No. 9. 2004.

Plan

- Notion très courante en médecine :
ex : Probabilité d'une maladie en fonction d'un signe clinique.
- Notion très courante en génétique :
ex : Probabilité d'un certain phénotype en fonction de l'expression d'un gène.
- $Pr(A|B)$: Probabilité de l'événement A conditionnellement à la réalisation de l'événement B
 - ⇔ Proportion de la caractéristique A parmi tous les individus ayant la caractéristique B .
 - ex** : Proportion de malades parmi tous les sujets ayant ce signe clinique.
- Définition de la probabilité conditionnelle :

$$Pr(A \text{ et } B) = Pr(A|B)Pr(B)$$

- ex** : La probabilité d'être malade et de présenter le signe clinique est égale à la probabilité d'être malade sachant que le signe est présent multipliée par la probabilité que le signe soit présent.

Si A ne dépend pas de B , alors :

$$Pr(A|B) = Pr(A|\bar{B}) = Pr(A)$$

ex : Si être malade ne dépend pas du signe clinique, la probabilité d'être malade sachant que le signe est présent est égale à la probabilité d'être malade sachant que le signe clinique est absent.

Ces probabilités sont simplement égales à la probabilité d'être malade.

Théorème des probabilités totales

$$Pr(A) = Pr(A \text{ et } \bar{B}) + Pr(A \text{ et } B)$$

ex : La proportion d'individus malades est égale à la proportion d'individus malades avec un signe clinique plus la proportion d'individus malades sans signe clinique.

- Soit N sujets répartis de la manière suivante :

		Test diagnostique		Total
		T^+	T^-	
Statut de référence	M^+	a	b	$a + b$
	M^-	c	d	$c + d$
Total		$a + c$	$b + d$	N

- Le statut de référence est défini par le **gold standard**.
- Deux types d'erreurs :
 - b faux négatifs : sujets ayant un test négatif alors qu'ils sont malades.
 - c faux positifs : sujets ayant un test positif alors qu'ils sont sains.
- Prévalence de la maladie : $Pr(M^+) = (a + b)/N$

- Dépister une maladie non apparente cliniquement dans une population où la prévalence est généralement faible.
- Priorité : Eviter de laisser repartir un individu alors qu'il est malade.
- Réduire le nombre de **faux négatifs**, même si plus de faux positifs.
- Possibilité de refaire des examens pour confirmer/infirmier les tests positifs.

- Diagnostiquer une maladie suspectée avec une prévalence assez élevée.
- Priorité : Eviter de traiter un individu alors qu'il est sain.
- Réduire le nombre de **faux positifs**, même si plus de faux négatifs.
- Possibilité de refaire des examens pour confirmer/infirmes les tests négatifs.

Plan

		Test diagnostique		Total
		T^+	T^-	
Statut de référence	M^+	a	b	$a + b$
	M^-	c	d	$c + d$
Total		$a + c$	$b + d$	N

- Capacité du test à bien détecter les malades
- Probabilité d'un test positif sachant que le sujet est malade
- Proportion de tests positifs chez les malades

$$Se = Pr(T^+ | M^+) = a / (a + b)$$

- Se diminue avec l'augmentation de b , le nombre de faux négatifs
- Se parfaite et égale à 100% si $b = 0$.

		Test diagnostique		Total
		T^+	T^-	
Statut de référence	M^+	a	b	$a + b$
	M^-	c	d	$c + d$
Total		$a + c$	$b + d$	N

- Capacité du test à bien détecter les non-malades
- Probabilité d'un test négatif sachant que le sujet est sain
- Proportion de tests négatifs chez les non-malades

$$Sp = Pr(T^- | M^-) = d / (c + d)$$

- Sp diminue avec l'augmentation de c , le nombre de faux positifs
- Sp parfaite et égale à 100% si $c = 0$.

- Une clinique londonienne a comparé les performances de OraQuick[®] (test oral non-invasif) à celles du test de référence (*gold standard*) sur 820 sujets présentant un risque élevé d'infection à VIH/sida. Les résultats ont été les suivants :

		OraQuick		Total
		T^+	T^-	
Statut de référence	VIH^+	44	3	47
	VIH^-	1	772	773
Total		45	775	820

- Prévalence observée : $P(VIH^+) = 47/820 = 5,7\%$
 → Contexte d'un test diagnostique

‡. An evaluation of the performance of OraQuick ADVANCE Rapid HIV-1/2 Test in a high-risk population attending genitourinary medicine clinics in East London, UK ; J Zelin, et al. ; International Journal of STD & AIDS. 2008 ; 19 :665-667.

		OraQuick		Total
		T^+	T^-	
Statut	VIH^+	44	3	47
de référence	VIH^-	1	772	773
Total		45	775	820

- $Se = 44/47 = 93,6\%$; $Sp = 772/773 = 99,9\%$
- Très bon test diagnostique (à relativiser selon le contexte).
 - Parmi les sujets malades, 93,6% ont un test positif
 - Parmi les sujets non-malades, 99,9% ont un test négatif

Problème

En pratique, la présence ou l'absence de la maladie n'est pas connue. La prise de décision dépend du résultat du test. Quelles sont les probabilités d'erreurs en fonction du résultat du test ?

Valeur prédictive positive

		Test diagnostique		Total
		T^+	T^-	
Statut de référence	M^+	a	b	$a + b$
	M^-	c	d	$c + d$
Total		$a + c$	$b + d$	N

- Probabilité qu'un sujet soit malade sachant que le test est positif
- Proportion de vrais malades parmi les sujets ayant un test positif

$$VPP = Pr(M^+ | T^+) = a / (a + c)$$

- VPP diminue avec l'augmentation de c , le nombre de faux positifs
- VPP parfaite et égale à 100% si $c = 0$.

Valeur prédictive négative

		Test diagnostique		Total
		T^+	T^-	
Statut de référence	M^+	a	b	$a + b$
	M^-	c	d	$c + d$
Total		$a + c$	$b + d$	N

- Probabilité qu'un sujet ne soit pas malade sachant que le test est négatif
- Proportion de vrais sains parmi les sujets ayant un test négatif

$$VPN = Pr(M^- | T^-) = d / (b + d)$$

- VPN diminue avec l'augmentation de b , le nombre de faux négatifs
- VPN parfaite et égale à 100% si $b = 0$.

		OraQuick		Total
		T^+	T^-	
Statut de référence	VIH^+	44	3	47
	VIH^-	1	772	773
Total		45	775	820

- $VPP = 44/45 = 97,8\%$; $VPN = 772/775 = 99,6\%$
- Dans cette population :
 - 97,8% des sujets ayant un test positif sont réellement malades.
 - 99,6% des sujets ayant un test négatif sont réellement sains.

Problème

Les valeurs prédictives dépendent de la prévalence de la maladie dans la population d'étude : $Pr(M^+)$.

Relation entre la VPP et la prévalence

$$\begin{aligned} VPP &= Pr(M^+ | T^+) \\ &= Pr(M^+ \text{ et } T^+) / Pr(T^+) \\ &= Pr(T^+ | M^+) Pr(M^+) / \{Pr(T^+ \text{ et } M^+) + Pr(T^+ \text{ et } M^-)\} \\ &= Pr(T^+ | M^+) Pr(M^+) / \{Pr(T^+ | M^+) Pr(M^+) + Pr(T^+ | M^-) Pr(M^-)\} \end{aligned}$$

(Théorème de Bayes)

d'où

$$VPP = Se \times Pr(M^+) / \{Se \times Pr(M^+) + (1 - Sp) \times (1 - Pr(M^+))\}$$

$$\begin{aligned}
 VPN &= Pr(M^- | T^-) \\
 &= Pr(M^- \text{ et } T^-) / Pr(T^-) \\
 &= Pr(T^- | M^-) Pr(M^-) / \{Pr(T^- \text{ et } M^+) + Pr(T^- \text{ et } M^-)\} \\
 &= Pr(T^- | M^-) Pr(M^-) / \{Pr(T^- | M^+) Pr(M^+) + Pr(T^- | M^-) Pr(M^-)\}
 \end{aligned}$$

(Théorème de Bayes)

d'où

$$VPN = Sp \times (1 - Pr(M^+)) / \{(1 - Se) \times Pr(M^+) + Sp \times (1 - Pr(M^+))\}$$

Dans l'étude londonienne, la prévalence était de 5,7%. Doit-on conseiller ce test pour une stratégie de dépistage de masse en France ? (On suppose une prévalence égale à 5 pour 10000 français).

$$VPP = 0,936 \times 0,0005 / \{0,936 \times 0,0005 + (1 - 0,999) \times (1 - 0,0005)\} = 26,6\%$$

$$VPN = 0,999 \times (1 - 0,0005) / \{(1 - 0,936) \times 0,0005 + 0,999 \times (1 - 0,0005)\} \approx 100\%$$

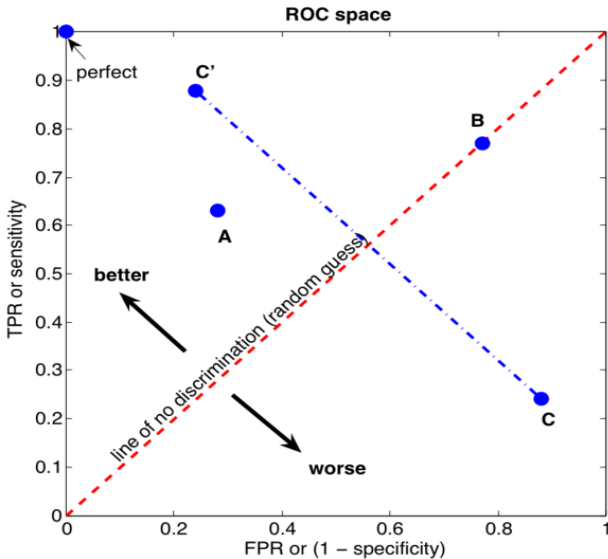
	Etude Londonnienne	Dépistage en France
<i>Se</i>	93,6%	93,6%
<i>Sp</i>	99,9%	99,9%
<i>Pr(M⁺)</i>	0,057	0,0005
<i>VPP</i>	97,8%	26,6%
<i>VPN</i>	99,6%	≈100%

Si le test est négatif, on a la quasi-certitude de l'absence de maladie. Si le test est positif, des examens complémentaires sont nécessaires. OraQuick semble constituer un bon test de dépistage.

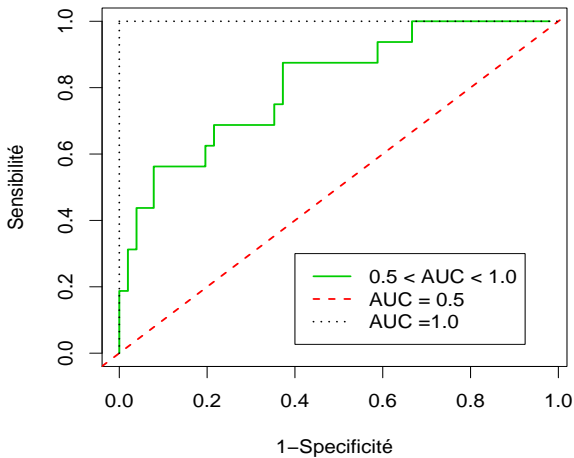
Plan

- Soit X la valeur du marqueur et G le groupe à diagnostiquer
ex : $G = 1$ si malade et $G = 0$ sinon
- Les patients sont à risque si X est supérieur à un seuil c .
 - Sensibilité : $SE = P(X > c | G = 1)$.
 - ex : Proportion de tests positifs chez les patients malades.
- Les patients ne sont pas à risque si X est inférieur à c .
 - Spécificité : $SP = P(X \leq c | D = 0)$.
 - ex : Proportion de tests négatifs chez les patients non-malades.
- Il existe un test diagnostique parfait si il existe une valeur de c pour laquelle on a $SE = SP = 1$.
- Courbes ROC : pour tous les seuils c possibles, on calcul $1 - SP$ et SP .

Courbes ROC



Courbes ROC



- Si $AUC = 0,5$: Aucune discrimination
- Si $0,7 \leq AUC < 0,8$: Discrimination acceptable
- Si $0,8 \leq AUC < 0,9$: Discrimination excellente
- Si $AUC \geq 0,9$: Discrimination hors du commun

Plan

Plan

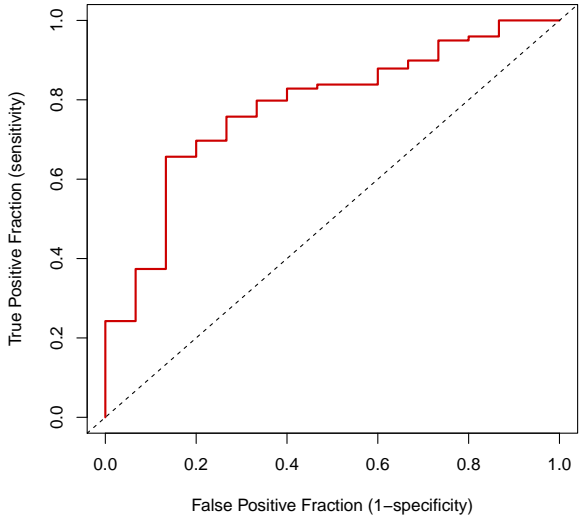
Exemple : Rémission d'une leucémie

```
> logit.model$coef

(Intercept)          X1          X2
2.5000572 -0.7980853  0.8972820

> S <- logit.model$coef[2] * X1 + logit.model$coef[3] * X2
> se<-function(x) { return(sum(1*(S[rem==1]>x))/sum(rem==1)) }
> sp<-function(x) { return(sum(1*(S[rem==0]<=x))/sum(rem==0)) }
> temp<-sort(unique(S))
> resultats<-data.frame(
+   sp=sapply(temp, sp),
+   sp1=1-sapply(temp, sp),
+   se=sapply(temp, se), seuils=temp)
> plot(c(1, resultats$sp1, 0), c(1, resultats$se, 0),
+   xlab="False Positive Fraction (1-specificity)",
+   ylab="True Positive Fraction (sensitivity)",
+   type="n")
> lines(c(1, resultats$sp1, 0), c(1, resultats$se, 0),
+   type="s", lwd=2, col="red3")
> abline(0,1, lty=2)
>
```

Exemple : Rémission d'une leucémie



```
> resultats[1:3,]
      sp      sp1      se      seuils
1 0.06666667 0.9333333 1.000000 -3.156476
2 0.13333333 0.8666667 1.000000 -2.872796
3 0.13333333 0.8666667 0.989899 -2.591430

> resultats<-resultats[order(resultats$sp1, resultats$se),]
> resultats[dim(resultats)[1]+1,1]<-1
> resultats[dim(resultats)[1],2]<-1
> resultats[dim(resultats)[1],3]<-resultats[dim(resultats)[1]-1,3]
> resultats[1:3,]

      sp sp1      se      seuils
114  1   0 0.00000000 4.149742
113  1   0 0.01010101 2.569103
112  1   0 0.02020202 2.553506

> sum( (resultats$sp1[2:length(resultats$sp1)] -
+ resultats$sp1[1:(length(resultats$sp1)-1)]) *
+ (resultats$se[2:length(resultats$se)]) )

[1] 0.7811448

>
```

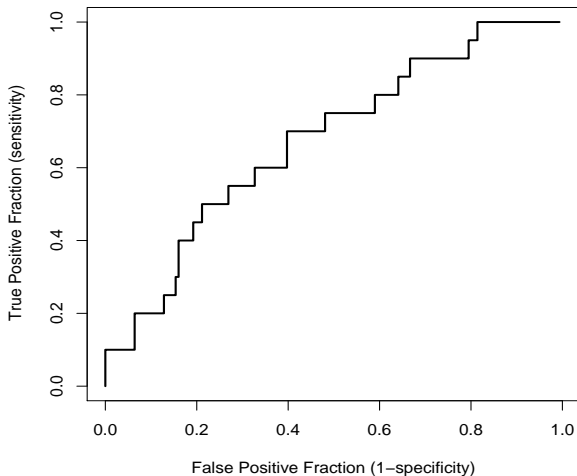
Attention : Sur-estimation des capacités pronostiques.

Plan

- Objectif : Evaluer si Tbet est un marqueur discriminant de patients transplantés rénaux stables et de patients diagnostiqués avec un rejet chronique humorale actif. §
 - L'échantillon est composé de 176 patients
 - 20 patients CAMR
 - 156 stables
 - Les valeurs observées sont comprises entre 0,07 et 9,64.
 - La moyenne sur tout l'échantillon est de 1,69
 - 1,08 chez les patients CAMR
 - 1,77 chez les stables
 - La distribution de Tbet semble différente entre les patients CAMR et les stables ($p=0,0116$, test de Mann-Withney).
- ⇒ Conclure que Tbet est un marqueur du rejet serait une erreur.

§. J. Asthon-Chess, et al. Projet Genhomme. PHRC 2003.

Exemple 2 : Tbet en transplantation rénale



Exemple 2 : Tbet en transplantation rénale

- L'AUC est égal à 0,67 (mauvaise discrimination).
- L'intervalle de confiance de bootstrap à 95% de l'AUC est compris entre 0,55 et 0,79.