

Quelques questions en ce qui concerne la
détermination de valeurs de référence et la mise
au point d'une méthode diagnostique

M.L. Delignette-Muller

19 novembre, 2018

Détermination de valeurs de référence

Détermination d'un intervalle de référence

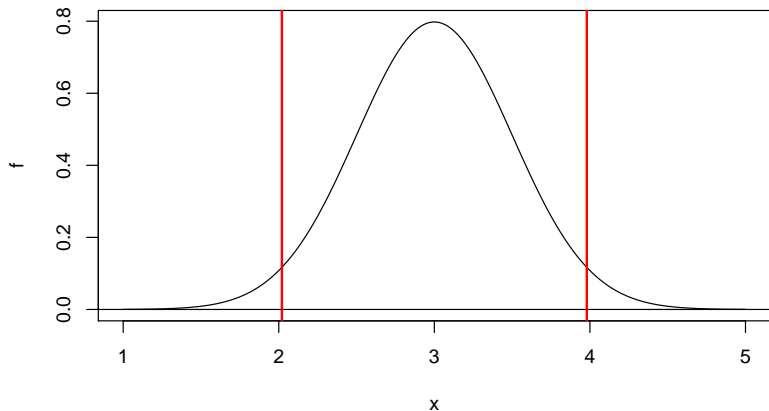
L'intervalle de référence pour une variable biologique est par définition l'**intervalle comprenant 95% des valeurs observées sur des animaux sains** (population de référence en bonne santé).

On l'appelle parfois l'intervalle de variabilité à 95%.

Il ne faut pas le confondre avec l'intervalle de confiance à 95% sur la moyenne (incertitude sur la moyenne de la distribution)

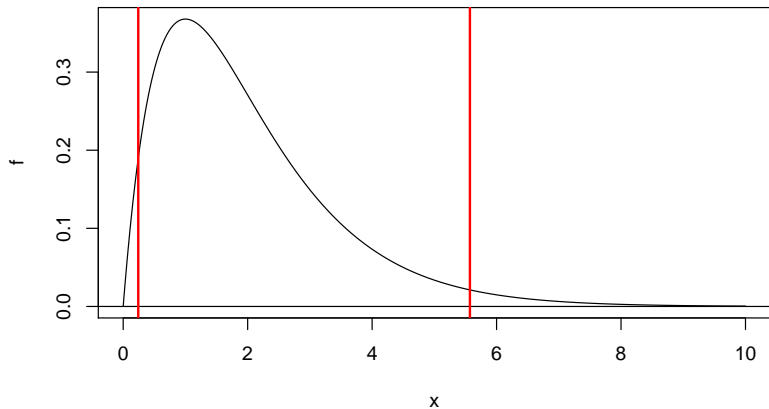
!

Visualisation d'un intervalle de référence pour une loi normale



Pour une loi normale, intervalle correspondant à $m \pm 2SD$.

Visualisation d'un intervalle de référence pour une loi non normale



Comment estimer l'intervalle de référence à partir d'un échantillon de la population saine ?

Méthode paramétrique

On ajuste une distribution (loi normale sur données brutes ou transformées ou autre loi) sur les données observées et on calcule les quantiles à 2.5 et 97.5% associés.

- ▶ le résultat peut être très dépendant de la loi ajustée.
- ▶ il convient de s'assurer que cette loi décrit très bien les données.

Méthode non paramétrique

On calcule les quantiles à 2.5 et 97.5% directement sur les données observées.

- ▶ **avantage** : on ne fait pas d'hypothèse quant à la forme de la distribution.
- ▶ **inconvénient** : le calcul de quantiles extrêmes sur une loi empirique (données observées) est imprécis et peu robuste si on ne dispose pas d'un grand nombre de données.

Illustration de la méthode paramétrique sur un petit échantillon - 20 observations

Histogram and theoretical densities

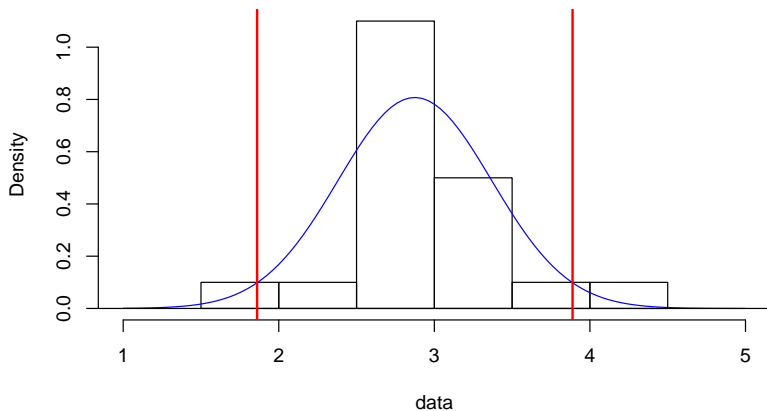


Illustration de la méthode paramétrique sur un petit échantillon - en fréquences cumulées

Distributions empirique et ajustée

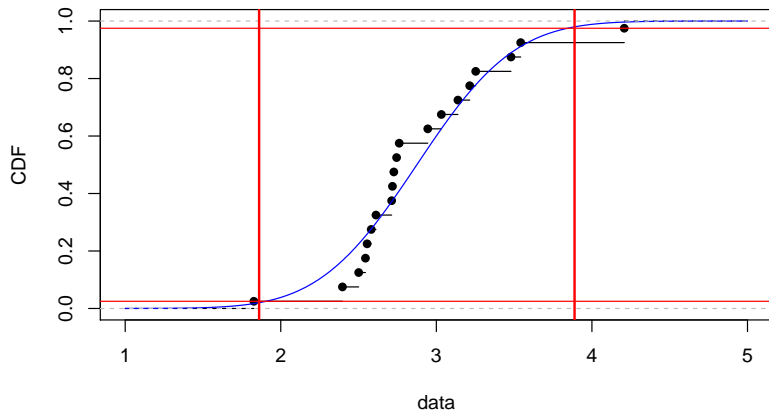


Illustration de la méthode non paramétrique sur un petit échantillon

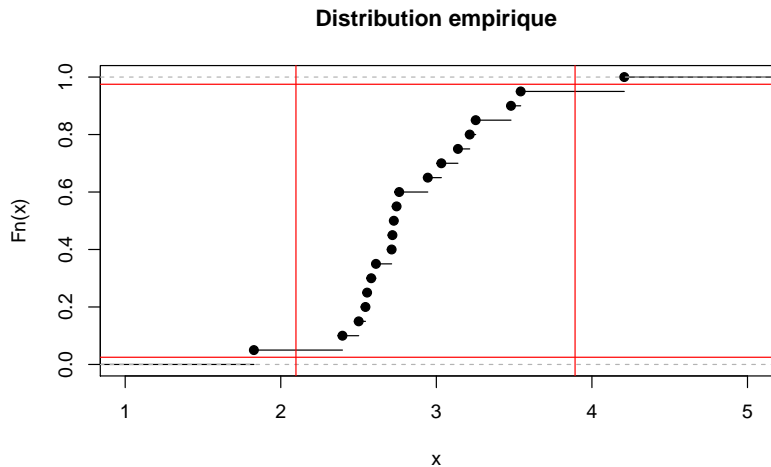


Illustration de la méthode paramétrique sur un grand échantillon - 300 observations

Distributions empirique et ajustée

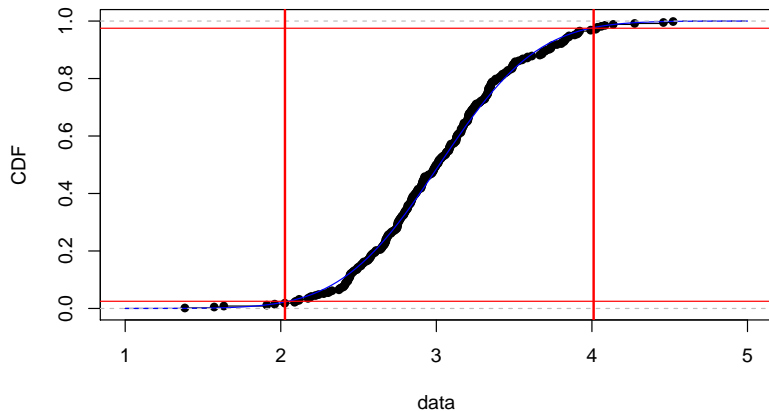


Illustration de la méthode non paramétrique sur un grand échantillon

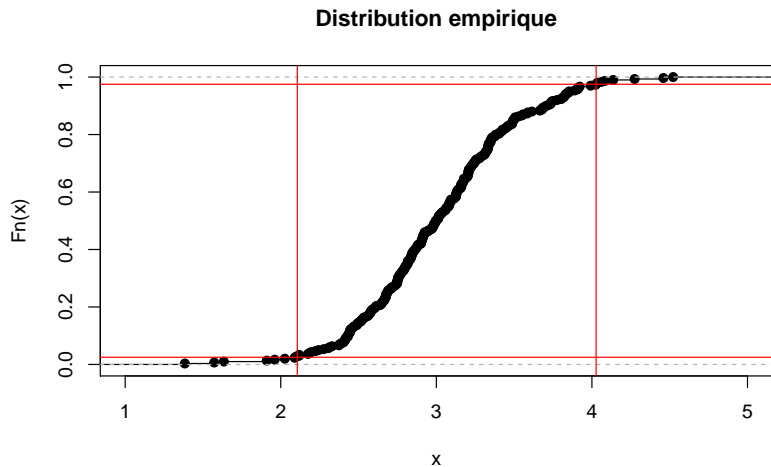
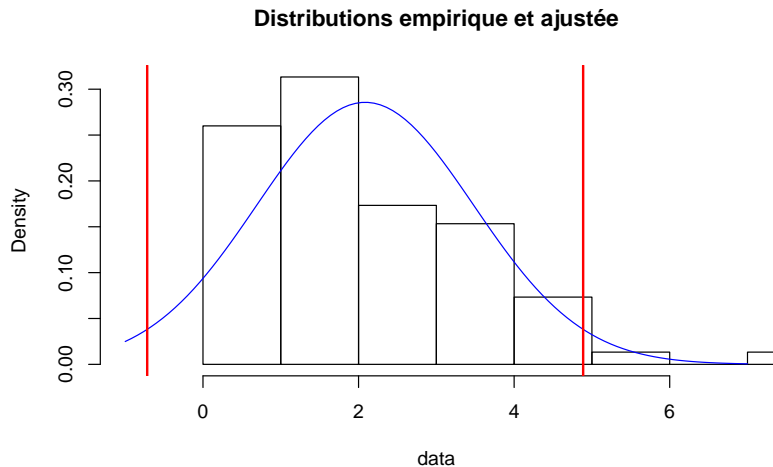


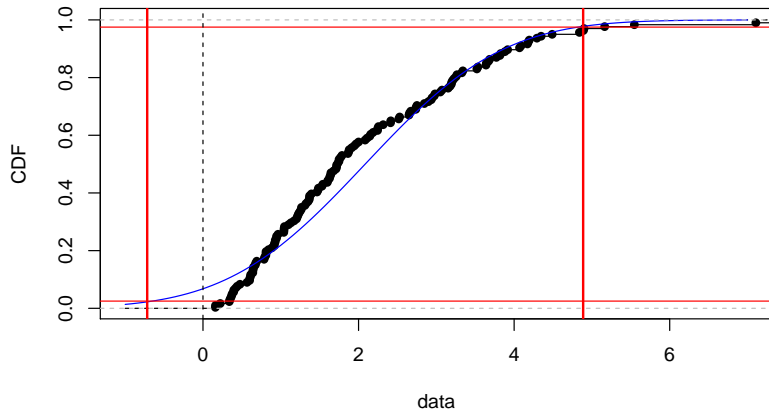
Illustration de la méthode paramétrique avec un mauvais choix de distribution ajustée



Ici valeur de référence basse peu raisonnable !

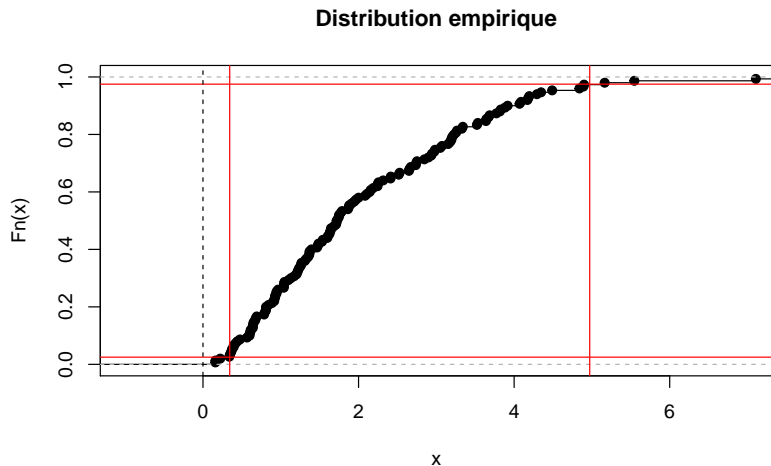
Illustration de la méthode paramétrique avec un mauvais choix de distribution ajustée - en fréquences cumulées

Distributions empirique et ajustée



Ici valeur de référence basse peu raisonnable !

Illustration de la méthode non paramétrique sur le même échantillon



Valeurs de référence plus raisonnables !

Recommandations

- ▶ **Grands échantillons** ($n \geq 120$) : méthode **non paramétrique**.
- ▶ **Echantillons de taille moyenne** ($20 \leq n \leq 120$) : méthode **paramétrique** si loi normale ou normalisable par transformation de variable ou si une autre loi ajuste bien les données (**bien vérifier l'ajustement de la loi aux données**).
- ▶ **Petits échantillons** ($n \leq 20$) : on ne calcule **pas d'intervalle de référence**.

Importance, dans tous les cas, de la **représentation graphique de la distribution observée** et ajustée (diagramme en boîte, histogramme si possible) si méthode paramétrique utilisée.

Mise au point d'un test diagnostique

Si une des valeurs de référence est prise comme seuil de positivité ?

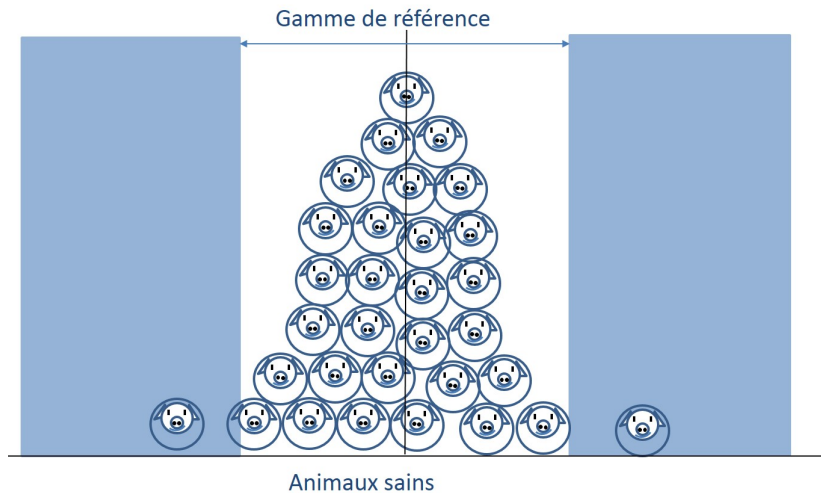
En supposant que les animaux malades ont une distribution décalée vers la droite par ex.,

si l'on choisit la valeur de référence haute comme étant le seuil de positivité,

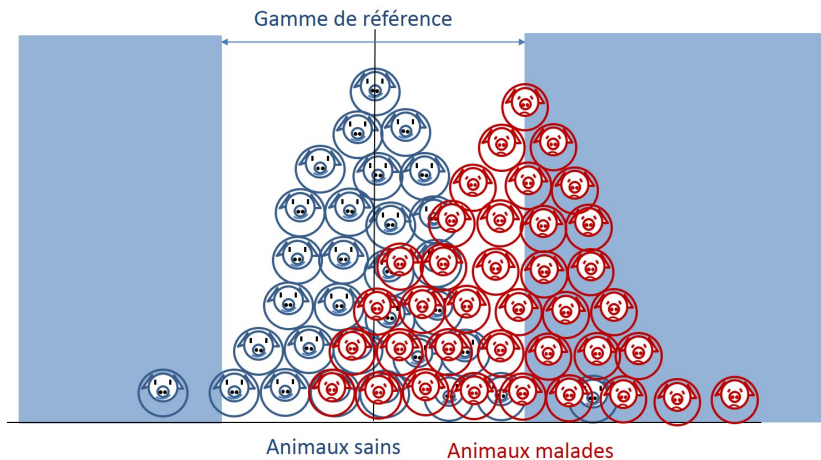
on aura une spécificité de 97.5% et la sensibilité dépendra de la distributions des animaux malades.

A noter que l'observation d'une différence significative entre les moyennes des deux populations n'est absolument pas une garantie de la possibilité de mettre au point un test diagnostique satisfaisant sur la variable étudiée.

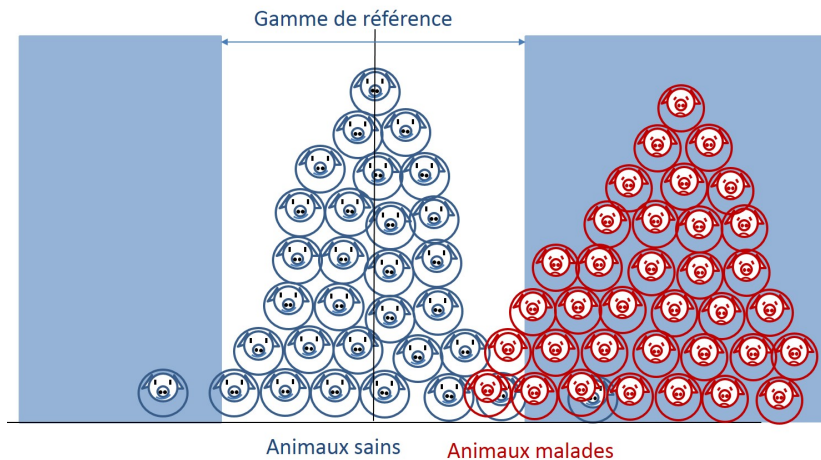
Un exemple



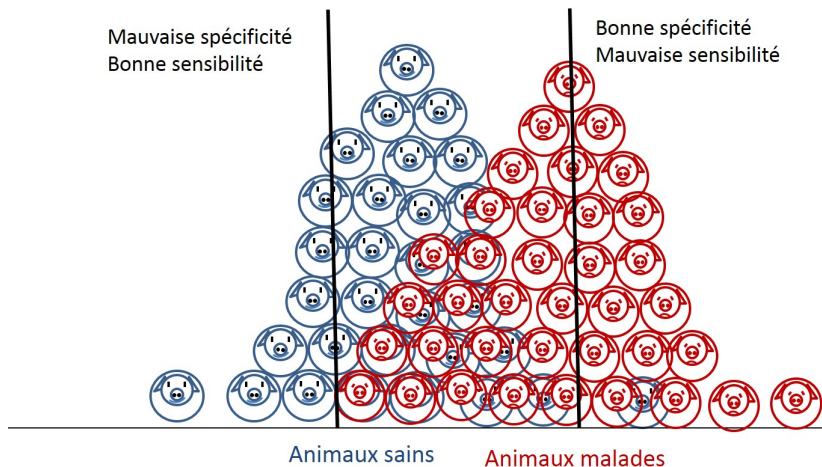
Sensibilité de l'ordre de 50% dans ce cas



Sensibilité de l'ordre de 90% dans ce cas



Seuil d'un test diagnostique = compromis sensibilité / spécificité



Mise en évidence d'une différence significative entre sains et malades et test diagnostique

La mise en évidence d'une différence significative de la valeur moyenne d'une variable quantitative entre les sains et les malades n'induit pas forcément la possibilité d'établir une bonne méthode diagnostique à partir de cette variable.

Examinons l'exemple très controversé de l'ostéoporose.

La définition de l'ostéoporose

Extrait de Kanis 2002, Diagnosis of osteoporosis and assessment of fracture risk.

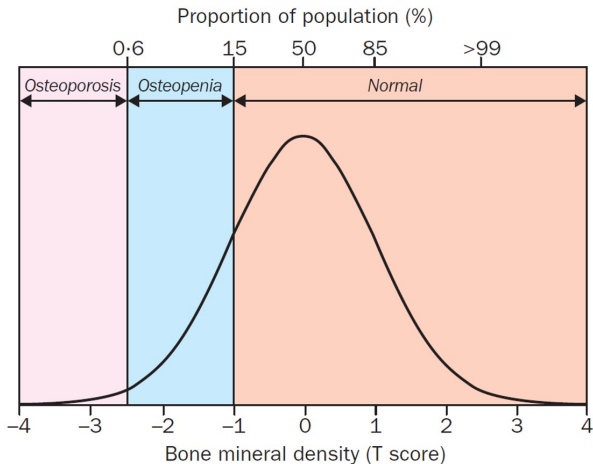


Figure 1: **Distribution of bone mineral density in healthy women aged 30-40 years**

Lien entre la densité osseuse et le risque de fractures

Le T score de densité osseuse mesure un écart entre la densité osseuse d'un individu et la moyenne de jeunes adultes du même sexe, exprimée en écart type de la distribution chez les jeunes de même sexe.

Il a été montré qu'une perte d'un SD du T score était associé à une multiplication du risque par 1.4. Par conséquent en cas d'ostéoporose (T score moins de -2.5) le risque est multiplié par 2.3 ($1.4^{2.5}$)

(Johnell 2005, Predictive Value of BMD for Hip and Other Fractures).

Risque de fractures chez les femmes suivant leur statut (ostéoporose ou non)

Une étude prospective sur 4 ans a montré que les femmes en post ménopause atteintes d'ostéoporose avaient un **risque de fracture sur 4 ans d'environ 20%**.

(Cummings, 1998, Effect of Aledronate on risk of fracture in women with low bone density but without vertebral fractures).

Nous considérerons dans la suite que le risque chez les femmes en post ménopause **non atteinte d'ostéoporose** ont un **risque** plus de 2.3 fois inférieur : 8% (un peu approximatif, mais c'est pour se faire une idée).

L'étude de Cummings a montré que le traitement à l'**Alendronate** permettait de **réduire de 20% à 13% le risque de fractures** sur 4 ans chez les femmes atteintes d'ostéoporose

Que permet le diagnostic de l'ostéoporose : quelques calculs simples sur 1000 femmes en post ménopause

Prévalence de l'ostéoporose chez des femmes en post ménopause : environ 20%

(Kanis 2005, Assessment of fracture risk)

200 avec ostéoporose

- ▶ **40 fractures (20%) (avec alendronate : 26 fractures)**
- ▶ 160 non fractures (potentiellement traitées pour rien)

800 sans ostéoporose

- ▶ **64 fractures (8%, sur des femmes non identifiées à risque)**
- ▶ 736 non fractures

$$\text{Sensibilité} = \frac{40}{40+64} = 38\% \quad \text{Spécificité} = \frac{736}{736+160} = 82\%$$

Le diagnostic de l'ostéoporose : un sujet très controversé

Extrait de Moynihan et al., 2002, "Selling sickness : the pharmaceutical industry and disease mongering", BMJ.

The link between bone density and fracture risk is also the subject of scientific controversy, with reviewers pointing out that while bone mineral density is associated with fracture, it is not a sufficiently accurate predictor of an individual's risk of fracture to be used as a guide to therapy. A recent evaluation by the University of British Columbia concluded that "Research evidence does not support either whole population or selective . . . bone mineral density testing of well women at or near menopause as a means to predict future fractures."

Conclusion sur cet exemple

Cet exemple illustre bien le fait que la mise en évidence d'une corrélation significative entre une variable quantitative et le risque d'une pathologie n'induit pas forcément la possibilité d'établir un diagnostic efficace de cette pathologie.

Actuellement le diagnostic de fractures sur la base de l'ostéoporose est remis en cause et il est conseillé de prendre en compte d'autres facteurs de risque dans le diagnostic et la décision ou non de traiter, afin

- ▶ de mettre en garde un plus grand nombre de femmes et de leur préconiser des moyens simples de prévention
- ▶ et de ne pas traiter inutilement un grand nombre de femmes.

Introduction aux courbes ROC sur un exemple

Diagnostic de la trypanosomiase bovine par une méthode Elisa (recherche des anticorps dans le sérum sanguin)

Jeu de données extrait de Greiner, et al. 2000. Principles and practical application of the **receiver-operating characteristic** analysis for diagnostic tests. Prev vet med, 45, 23-41.

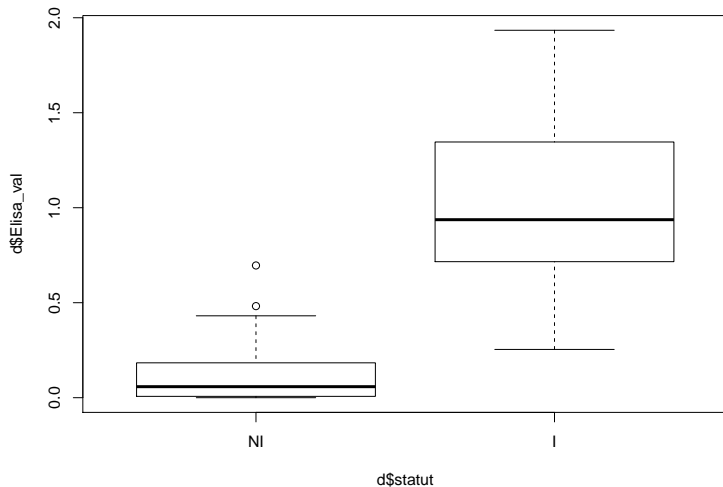
```
require(pROC)
d <- read.table("DATA/trypano.txt", header = TRUE)
str(d)
```

```
## 'data.frame':    100 obs. of  2 variables:
## $ statut      : Factor w/ 2 levels "I","NI": 2 2 2 2 2 2 2 2 2 2 ...
## $ Elisa_val: num  0 0 0 0 0 0 0 0 0.001 0.001 0.001 ...
```

```
# on met la modalité non infecté (NI) en première modalité
d$statut <- factor(d$statut, levels = c("NI", "I"))
```

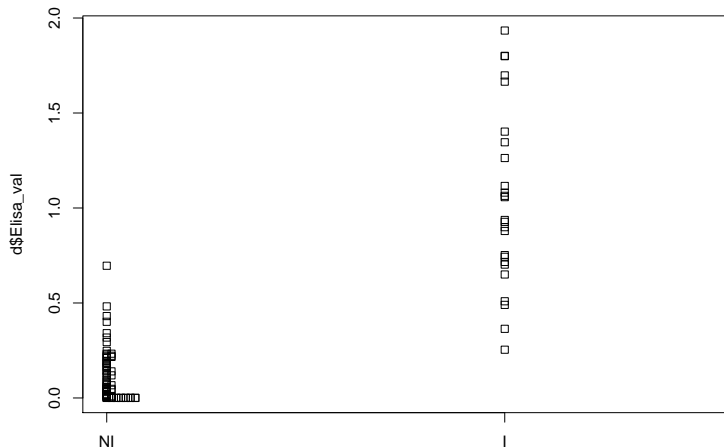
Visualisation des données en boîtes à moustache

```
plot(d$Elisa_val ~ d$statut)
```



Visualisation de tous les points observés

```
stripchart(d$Elisa_val ~ d$statut, vertical = TRUE,  
           method = "stack")
```



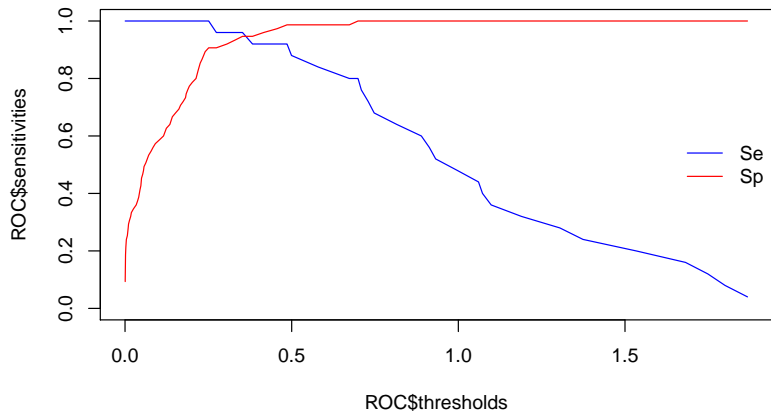
Calcul des sensibilités et spécificités pour différents seuils

```
ROC <- roc(d$statut, d$Elisa_val)
ROC

##
## Call:
## roc.default(response = d$statut, predictor = d$Elisa_val)
##
## Data: d$Elisa_val in 75 controls (d$statut NI) < 25 cases
## Area under the curve: 0.993
```

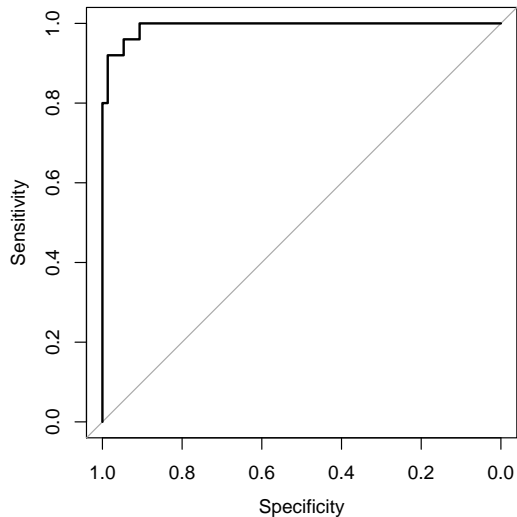

Tracé des sensibilités et spécificités pour différents seuils

```
plot(ROC$thresholds, ROC$sensitivities, type = "l", col = "blue", lty = 1)  
lines(ROC$thresholds, ROC$specificities, type = "l", col = "red", lty = 1)  
legend("right", legend = c("Se", "Sp"), lty = 1, col = c("blue", "red"))
```



Tracé de la courbe ROC (Se fonction de 1 - Sp)

```
plot(ROC)
```



Utilisation de la courbe ROC

- ▶ **Définition d'un seuil optimal.** Si on ne sait pas quoi privilégier entre Se et Sp on peut maximiser un indicateur comme l'indice de Youden = $Se + Sp - 1$. Mais il est important de prendre aussi en compte la prévalence de la pathologie dans la population cible pour estimer les valeurs prédictives positives et négatives et de tenir dans le choix du seuil des conséquences d'un faux positif et d'un faux négatif. Pour aller plus loin, package `OptimalCutpoints`.
- ▶ **Comparaison de différentes courbes ROC** via leur aire sous la courbe (AUC) : plus elle est élevée plus la variable quantitative est intéressante en terme diagnostique.

Rappel sur le calcul des valeurs prédictives en fonction de la prévalence (cours d'épidémiologie)

- ▶ Valeur prédictive positive :

$$VPP = \frac{VP}{FP+VP} = \frac{Se \times p}{Se \times p + (1-Se) \times (1-p)}$$

- ▶ Valeur prédictive négative :

$$VPN = \frac{VN}{FN+VN} = \frac{Sp \times (1-p)}{Sp \times (1-p) + (1-Sp) \times p}$$

Ex. pour un test considéré donné avec $Se = 0.98$ et $Sp = 0.98$

Test utilisé sur une population de prévalence 1 cas sur 10

```
Se <- 0.98; Sp <- 0.98; p <- 0.1  
(VPP <- Se*p / (Se*p + (1 - Sp)*(1 - p)))
```

```
## [1] 0.845
```

```
(VPN <- Sp*(1 - p) / ( Sp*(1 - p) + (1 - Se)*p))
```

```
## [1] 0.998
```

Que se passe-t-il si on diminue la prévalence ?

Ex. pour un test considéré donné avec $Se = 0.98$ et $Sp = 0.98$

Test utilisé sur une population de prévalence 1 cas sur 100

```
Se <- 0.98; Sp <- 0.98; p <- 0.01  
(VPP <- Se*p / (Se*p + (1 - Sp)*(1 - p)))
```

```
## [1] 0.331
```

```
(VPN <- Sp*(1 - p) / ( Sp*(1 - p) + (1 - Se)*p))
```

```
## [1] 1
```

Pourquoi est-il souvent illusoire de faire un diagnostic à grande échelle sur une population à faible risque ?

Ex. pour un test considéré donné avec $Se = 0.98$ et $Sp = 0.98$

Test utilisé sur une population de prévalence 1 cas sur 1000

```
Se <- 0.98; Sp <- 0.98; p <- 0.001  
(VPP <- Se*p / (Se*p + (1 - Sp)*(1 - p)))
```

```
## [1] 0.0468
```

```
(VPN <- Sp*(1 - p) / ( Sp*(1 - p) + (1 - Se)*p))
```

```
## [1] 1
```

Parce que la plupart de positifs seraient des faux positifs.

Pour résumer cette deuxième partie

- ▶ Un **seuil de positivité** permet d'établir un diagnostic, de **discriminer entre deux populations** (saine et malade).
- ▶ Une différence significative entre la valeur moyenne d'un indicateur quantitatif entre les sains et les malades ne suffit pas à garantir un bon pouvoir diagnostique.
- ▶ Avant de définir un seuil de positivité il convient de définir la **population à laquelle s'adresse le test diagnostique** et si possible d'avoir une idée de la **prévalence de la maladie dans cette population**.
- ▶ Il convient aussi d'envisager les impacts respectifs des deux erreurs de diagnostic possibles (faux négatif ou faux positif) et de déterminer si le test sera utilisé pour un diagnostic individuel ou de troupeau.
- ▶ En fonction des différents éléments **le choix du seuil de positivité sera toujours un compromis entre gain de sensibilité et gain de spécificité** (courbes ROC).

Quelques références méthodologiques pour aller plus loin

- ▶ Friedrichs, K. R., Harr, K. E., Freeman, K. P., Szladovits, B., Walton, R. M., Barnhart, K. F., & Blanco-Chavez, J. (2012). **ASVCP reference interval guidelines**: determination of de novo reference intervals in veterinary species and other related topics. *Veterinary Clinical Pathology*, 41(4), 441-453.
- ▶ Geffré, A., Friedrichs, K., Harr, K., Concordet, D., Trumel, C., & Braun, J. P. (2009). **Reference values: a review**. *Veterinary clinical pathology*, 38(3), 288-298.
- ▶ Greiner, M., Pfeiffer, D., & Smith, R. D. (2000). **Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests**. *Preventive veterinary medicine*, 45(1), 23-41.
- ▶ López-Ratón, M., Rodríguez-Álvarez, M. X., Cadarso-Suárez, C., & Gude-Sampedro, F. (2014). **OptimalCutpoints: an R package for selecting optimal cutpoints in diagnostic tests**. *Journal of statistical software*, 61(8), 1-36.