

# Analyses de biologie médicale - aspects statistiques -

- I. Terminologie : biologie médicale
- II. Terminologie : tests binaires
- III. Définition des bornes
- IV. Conséquences en pratique clinique



# Terminologie

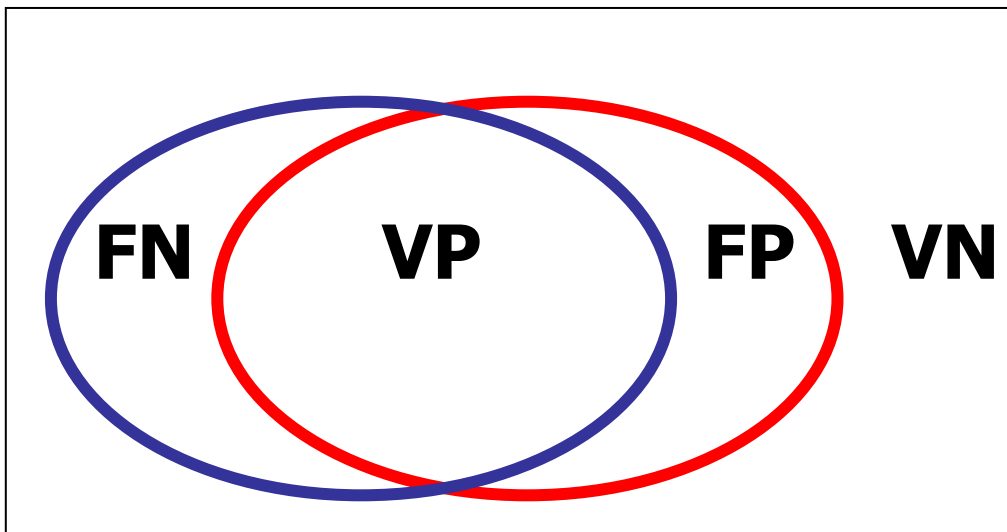


- Définitions, en pratique clinique
  - Pour un médecin « la biologie » signifie « les analyses de biologie médicale » ou « les résultats de... »
  - Dans ce cours : résultats de mesures réalisées sur des prélèvements sur une personne physique, à des fins diagnostiques et pronostiques
- Différents types (non limitatif)
  - Hématologie biologique (numération, hémostase...)
  - Biochimie (électrolytes, enzymes, hormones, toxiques...)
  - Microbiologie (bactériologie, virologie, parasitologie...)
  - Etc.
- Dans ce cours :
  - Examens produisant des résultats numériques, dont la valeur normale n'est pas nulle
  - Quantification de composés habituellement présents dans le sang
  - Exemples types : ionogramme sanguin, bilan rénal, bilan hépatique, bilan glucidique et métabolique, numération de la formule sanguine, dosages hormonaux...

# Rappel de terminologie : tests diagnostiques binaires



Ensemble **bleu** : Malades/non malades M/NM  
Ensemble **rouge** : Test positif/négatif T+/T-



$$FN = T- \cap M$$

$$VP = T+ \cap M$$

$$FP = T+ \cap NM$$

$$VN = T- \cap NM$$

# Rappel de terminologie : tests diagnostiques binaires



	M	NM
T+	# <b>VP</b>	# <i>FP</i>
T-	# <i>FN</i>	# <b>VN</b>

Sensibilité

$$Se = P(T^+ | M)$$
$$\widehat{Se} = VP / (VP + FN)$$

Spécificité

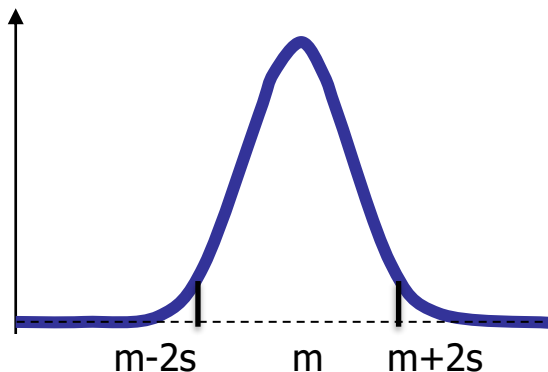
$$Sp = P(T^- | NM)$$
$$\widehat{Sp} = VN / (VN + FP)$$

# Paramètres biologiques

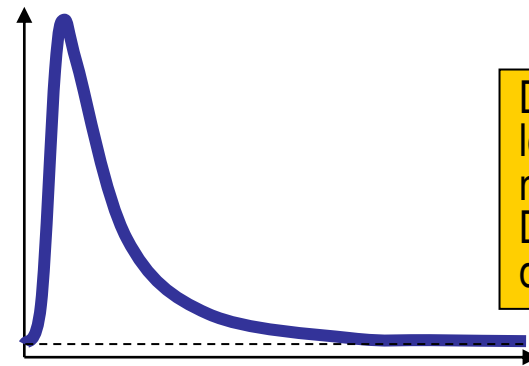


- Prélèvement dans divers milieux (sang, urine, LCR...)
- Nous nous intéressons aux examens retournant une variable quantitative, dosant un élément normalement présent (majoritaires en pratique clinique)
- Distribution de la variable connue chez les sujets sains :

Souvent distribution normale



Souvent distribution log-normale,  
« distribution de Galton »



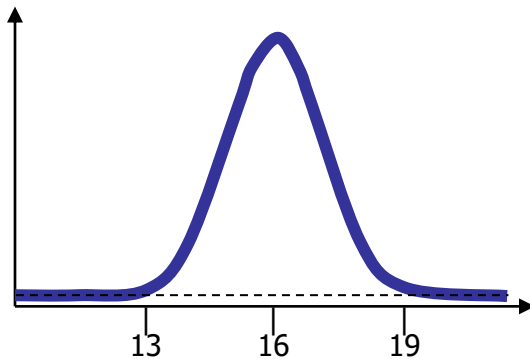
Distribution dont le logarithme suit une loi normale.  
Donc exponentielle d'une loi normale

# Exemples de distributions chez les sujets sains

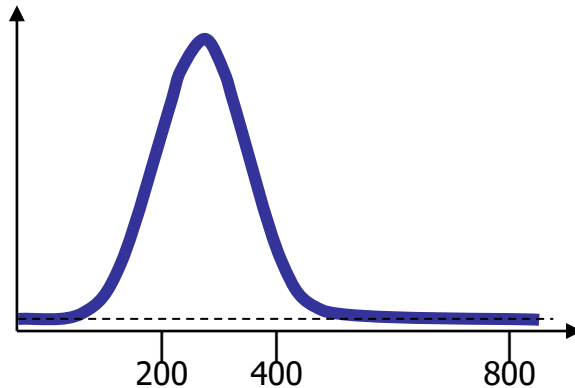


## ■ Distributions normales :

Taux sanguin d'hémoglobine en g/dl

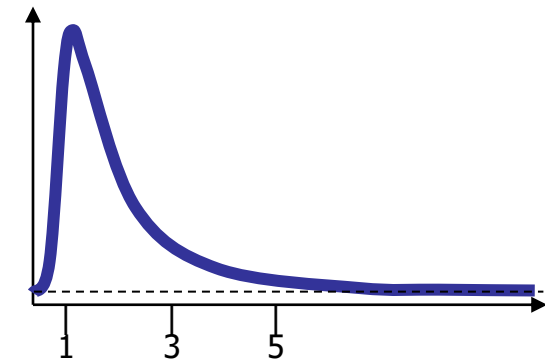


Taux sanguin de plaquettes en  $10^9/l$

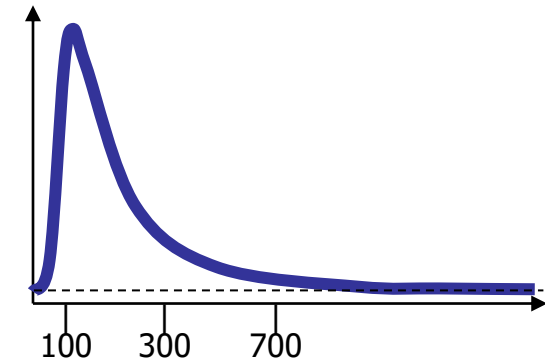


## ■ Distributions lognormales :

Concentration plasmatique en TSH en mUI/l



Concentration plasmatique en ferritine en ng/ml



# Définition de la normalité des paramètres biologiques



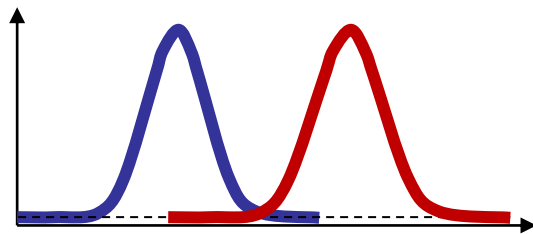
- Les bornes de normalités sont définies par un intervalle contenant 95% des valeurs normales des sujets sains
  - Variable normale :  $[\mu - 1,96 \times \sigma ; \mu + 1,96 \times \sigma]$
  - (autre méthode mais similaire pour les variables modélisables autrement qu'avec une loi normale)
  - Autres variables non modélisables : intervalle défini par les quantiles  $[F^{-1}(0.025) ; F^{-1}(0.975)]$
- Conséquence importante à retenir : sur un paramètre donné, **5% des sujets sains ont des valeurs « anormales » !**
- Pour simplifier, on présentera ci-après la situation où une seule des deux bornes est utilisée

# Application d'un test biologique pour détecter des sujets malades



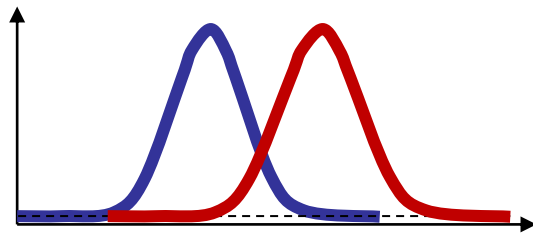
Distribution comparée du paramètre cher les **sujets sains** et les **sujets malades**.

Exemples de situations :



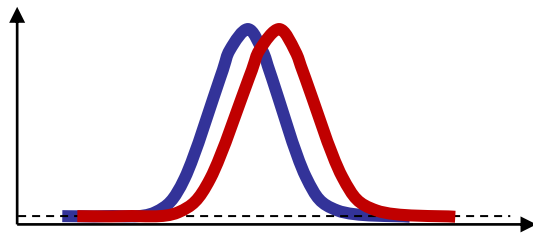
Situation idéale :

Il est facile de trouver un seuil qui discrimine sans erreur de classement les malades



Recouvrement partiel :

On peut choisir un seuil, mais erreur de classement dans la zone de recouvrement



Recouvrement important :

Test sans intérêt diagnostique



# Rappel : risques alpha et bêta appliqués à un test diagnostique



- Risque alpha (risque de 1<sup>ère</sup> espèce) :
  - Risque de déclarer un sujet malade alors qu'il est sain
  - =1-Spécificité
- Risque bêta (risque de 2<sup>ème</sup> espèce) :
  - Risque de déclarer un sujet sain alors qu'il est malade
  - =1-Sensibilité
- Puissance (1- $\beta$ ) :
  - Probabilité de déclarer un sujet malade à juste titre
  - =Sensibilité

**Réalité**

**Décision  
retenue  
au vu du  
test**

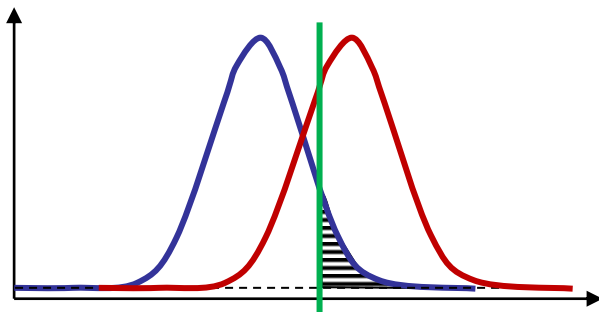
	Malade	Non malade
Test anormal (valeur $\notin$ intervalle)	Pas d'erreur	<b>Risque <math>\alpha</math></b>
Test normal (valeur $\in$ intervalle)	<b>Risque <math>\beta</math></b>	Pas d'erreur

# Risque alpha, risque bêta, puissance

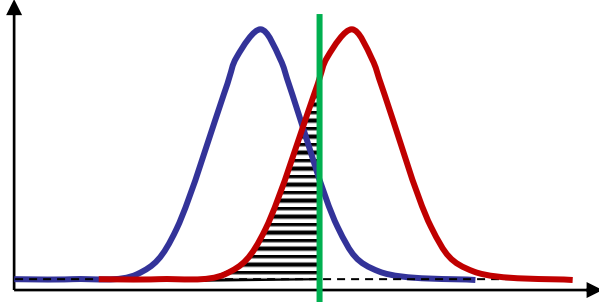


Distribution comparée du paramètre chez les **sujets sains** et les **sujets malades**, positionnement du **seuil choisi**.

Zone du risque Alpha

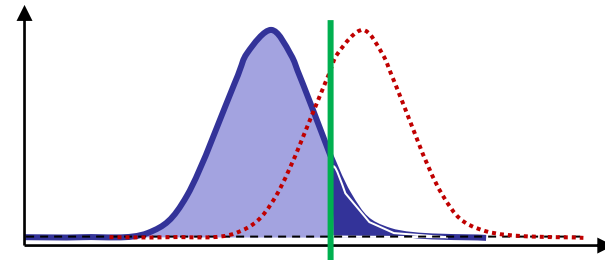


Zone du risque Bêta



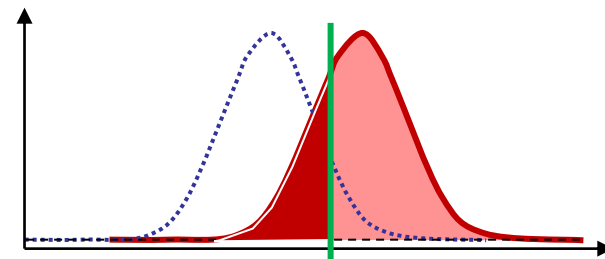
Dans cet exemple les sujets malades ont des valeurs plus élevées que les sains. Ce n'est pas toujours le cas !

Calcul du risque Alpha (probabilité associée)



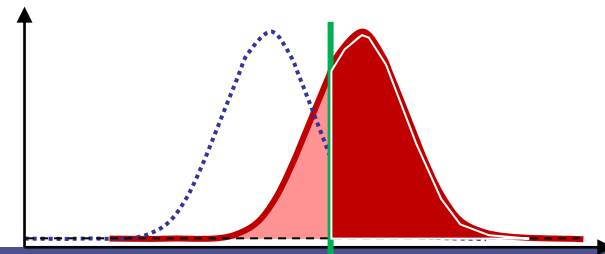
Probabilité de déclarer malades des sujets sains  
 $= 1 - \text{Spe}$

Calcul du risque Bêta (probabilité associée)



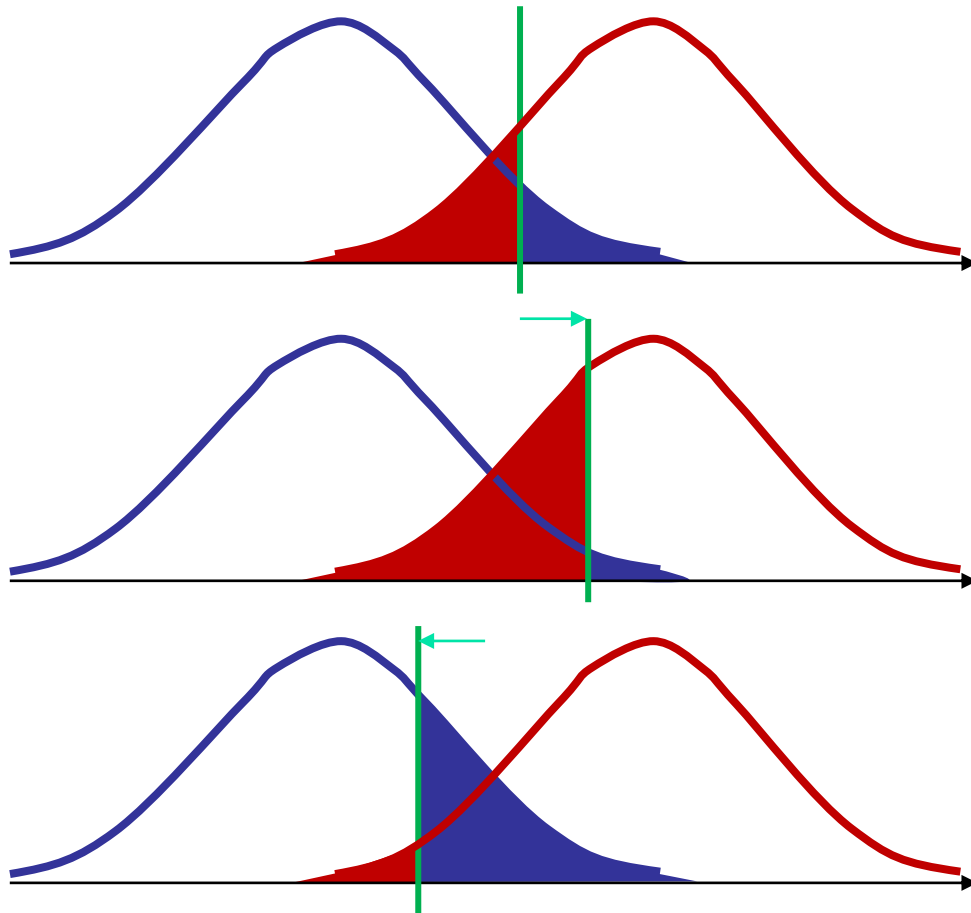
Probabilité de déclarer sains des sujets malades  
 $= 1 - \text{Se}$

Calcul de la Puissance (probabilité)



Probabilité de déclarer malades les sujets malades  
 $= \text{Se}$

# Influence du choix du seuil sur les risques alpha et bêta



Dans cet exemple (car les sujets malades ont des valeurs plus élevées que les sujets sains) :

Élévation du seuil =>

- Diminution du risque alpha
- Augmentation du risque bêta
- Diminution de la puissance

Abaissement du seuil =>

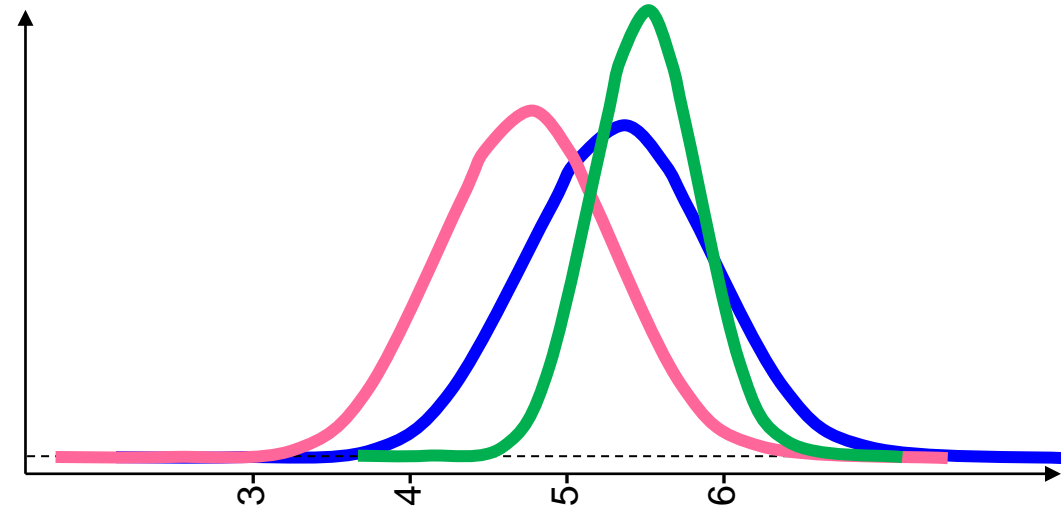
- Augmentation du risque alpha
- Diminution du risque bêta
- Augmentation de la puissance

# La distribution du paramètre peut varier selon la sous-population étudiée



- Exemple: concentration sanguine en globules rouges (en  $10^6/\mu\text{l}$ )
- Forte variation des valeurs normales entre 3 populations
- Les bornes utilisées pour le diagnostic doivent en tenir compte

Population	Borne basse	Borne haute
Adulte H	4.5	6.2
Adulte F	4	5.4
Nouveau-né	5	6



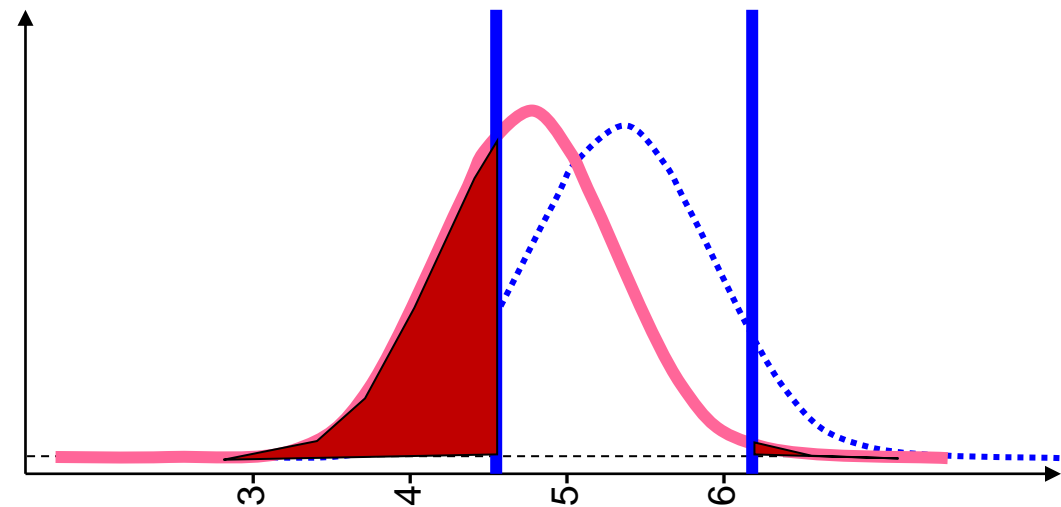
# La distribution du paramètre peut varier selon la sous-population étudiée



- Si par exemple on appliquait les bornes des hommes adultes aux femmes adultes, on obtiendrait :
- Pour la détection des valeurs trop élevées : risque bêta  $\uparrow$ , risque alpha  $\downarrow$
- Pour la détection des valeurs trop basses : risque alpha  $\uparrow$ , risque bêta  $\downarrow$

Population	Borne basse	Borne haute
Adulte H	4.5	6.2
Adulte F	<del>4</del>	<del>6.5</del>

- Pour certains paramètres, les seuils de normalité dépendent de la population étudiée.



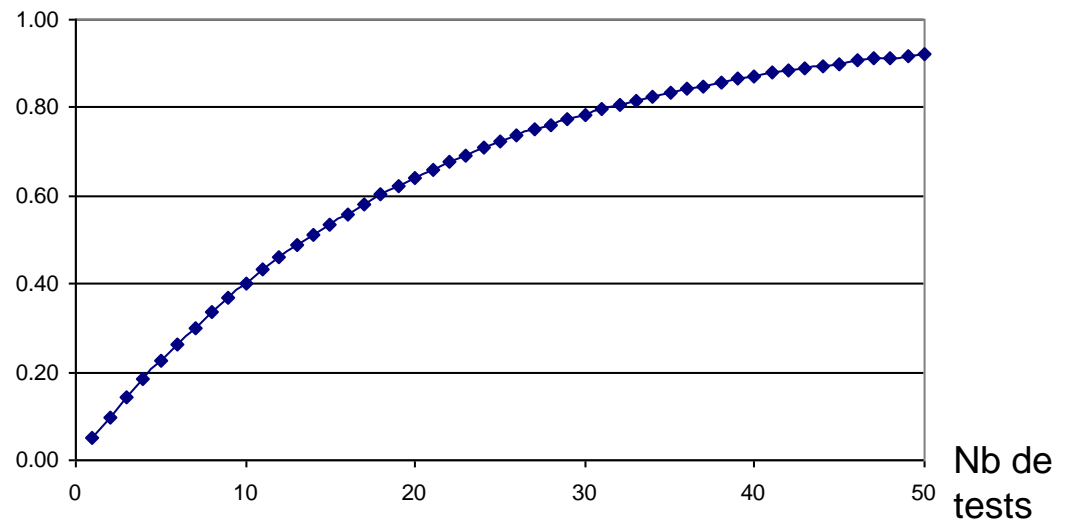
# Réalisation d'un grand nombre de tests chez le sujet sain



- Énoncé : on réalise K tests biologiques indépendants chez un sujet sain. Le risque alpha de chaque test s'élève à 5%. Quelle est la probabilité d'avoir **au moins un résultat anormal**, selon le nombre de tests ?
- Résolution : Pour chaque test, probabilité que le test soit positif  $p=0.05$ .  
Probabilité d'avoir au moins un résultat positif sur k tests  
= 1 - probabilité de n'avoir aucun résultat positif  
=  $1 - (1-p)^k$

k	Probabilité
1	0.05
2	0.10
5	0.23
10	0.40
20	0.64
50	0.92

Probabilité d'avoir au moins un test positif



# Conclusion



- Tests de biologie médicale : réalisation très fréquente
- Tests dosant un élément normalement présent dans le milieu :
  - Bornes définies chez les sujets sains : les 2,5% de valeurs les + basses et les + hautes sont déclarées « anormales »
  - Par définition, 5% des valeurs sont donc anormales
  - En cas de tests multiples, forte probabilité d'avoir au moins une valeur anormale
- Conduite à tenir :
  - Ne demander un test que si motivation clinique
  - Traiter par le mépris des anomalies isolées et modérées