



# Journées Nationales APMEP TOULOUSE 2014

## Les tests d'équivalence et de non-infériorité

J. Faisant<sup>1</sup>

20 octobre 2014

---

<sup>1</sup>IREM de Caen - BN, groupe « Histoire des probabilités et de la statistique »

[http ://www.math.unicaen.fr/irem/spip.php?rubrique43/](http://www.math.unicaen.fr/irem/spip.php?rubrique43/)  
[http ://jacques.faisant.pagesperso-orange.fr/JN2014/](http://jacques.faisant.pagesperso-orange.fr/JN2014/)

# Pourquoi s'intéresser à ce type de tests ?

- Ce qui m'a motivé :

- ▶ L'affaire Séralini en 2012 (évoquée incidemment lors des JN de Metz)
- ▶ Les réactions (*antérieures*) de Monsieur Marc Lavielle, INRIA Saclay, dans son document de 2009 :  
« *Quelques commentaires sur les méthodologies statistiques utilisées dans l'analyse des études de toxicité des OGM* »

Une alternative pertinente :

## Test d'équivalence

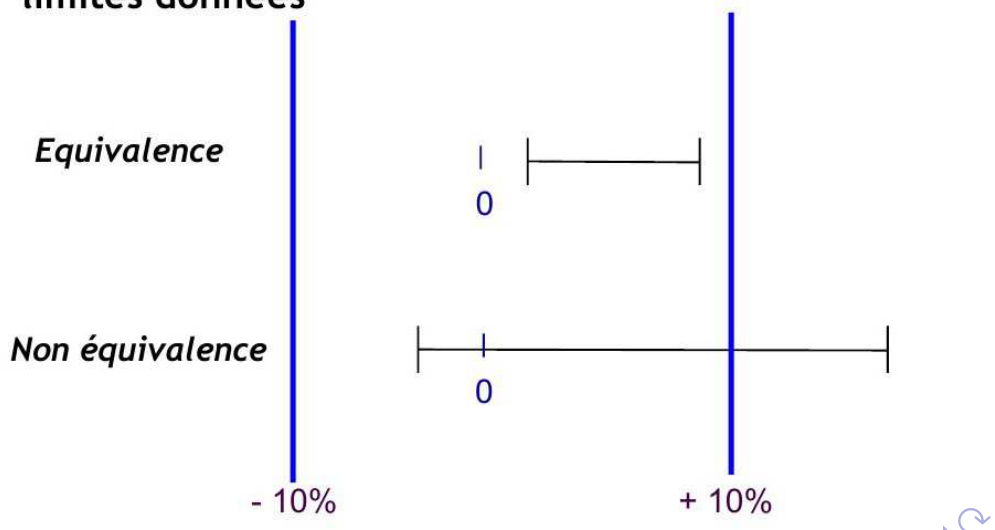
### Hypothèse de référence :

“il existe une différence biologiquement significative entre les 2 groupes”

*C'est alors aux données de démontrer l'équivalence des 2 régimes.*

**Remarque** : ce test est très largement utilisé dans le domaine du médicament

On conclut à l'équivalence des 2 plantes si tout l'intervalle de confiance est compris entre 2 limites données



# Quatre images de la conférence du 31/01/2013 à Nancy

1. OGM, PGM, HCB,...

2. rôle et limites de la statistique dans l'évaluation des risques

3. quand la statistique est maltraitée

4. « l'affaire » NK603

Les pièces de monnaies produites par un fabricant sont elles bien équilibrées ?

Si les pièces sont équilibrées, Probabilité(FACE) = 50%

Pour tester cette hypothèse, on réalise  $N$  lancers et on compare la proportion observée de FACE à la proportion attendue sous l'hypothèse « nulle » (équilibre parfait).

Si le résultat obtenu est « invraisemblable » sous l'hypothèse d'équilibre parfait, alors on rejettera cette hypothèse.

Intuitivement,

- si on observe 4 FACE sur 10 lancers, on aura du mal à conclure que la pièce est truquée,
- si on observe 400 FACE sur 1000 lancers, on aura du mal à conclure que la pièce est équilibrée...

- Au sujet du MON863, l'AFSSA conclut

*"Considérant qu'aucune différence significative n'ayant été observée entre les résultats obtenus pour le maïs MON 863 et les autres variétés de maïs, on peut donc conclure à l'équivalence alimentaire de la nouvelle plante".*  
(AFSSA, Saisine 2003-SA-0215, p 6)

- Au sujet du MON810, MONSANTO conclut

*"MON 810 was shown to be substantially equivalent to conventional maize, except for the introduced lepidopteran protection trait, and as safe and as nutritious as any other, commercially available maize."*  
(MON810 Technical dossier p 99)

Le test d'équivalence

Hypothèse de référence :

"il existe un effet OGM biologiquement significatif"

- L'OGM est "présumé coupable"
- C'est alors aux données de démontrer le contraire !  
(i.e. de démontrer l'équivalence des 2 régimes)

*Dans l'exemple concernant les pièces de monnaie, il s'agit, si on a suffisamment d'arguments, de rejeter l'idée qu'elles sont bien équilibrées.*

*Si, finalement, on manque d'arguments pour rejeter cette idée, **selon l'énoncé précédent**, elle n'en sera **pas** pour autant **démontrée**.*

# Hypothèses ...

**Le test d'équivalence**

**Hypothèse de référence :**

“il existe un effet OGM biologiquement significatif”

- L'OGM est “présumé coupable”
- C'est alors aux données de démontrer le contraire !  
(i.e. de démontrer l'équivalence des 2 régimes)

Marc Lavielle indique ici aux fabricants que,

- *pour démontrer statistiquement que le produit OGM dont ils demandent la mise sur le marché n'est pas biologiquement différent du produit naturel,*
- ils doivent fournir une étude permettant de **rejeter l'hypothèse** de référence<sup>2</sup> :

▶ « cet OGM **est** biologiquement différent du produit naturel »

<sup>2</sup>plus généralement désignée par l'expression « l'hypothèse nulle »

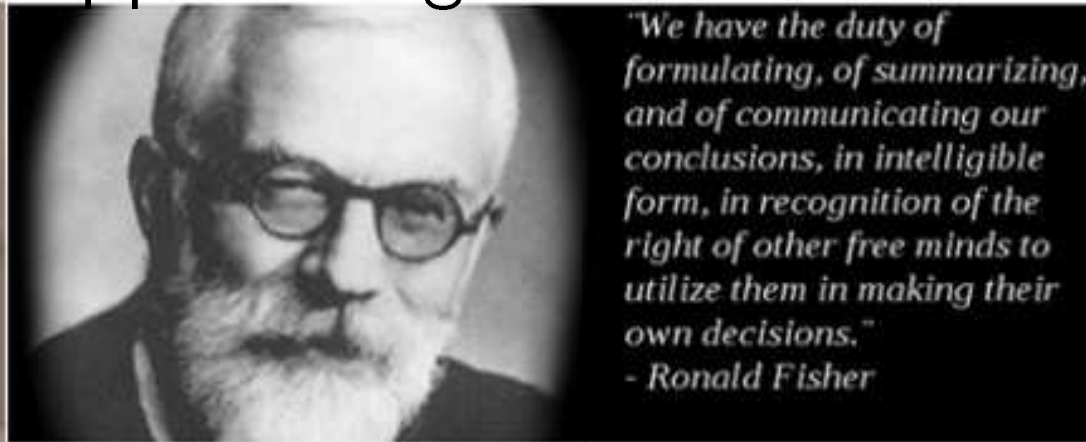
# Plan de l'exposé

- Une historique abrégée de la théorie des tests d'hypothèses
- Les définitions et théorèmes fondamentaux
- Un peu de pratique
- Démontrer l'hypothèse nulle ?  
Dans quels domaines particuliers cette question se pose-t-elle ?
- Les tests de non-infériorité
  - ▶ Méthodes
  - ▶ et exemples
- Les tests d'équivalence
  - ▶ Méthodes
  - ▶ et exemples
- Pour en savoir plus ...
- Annexes

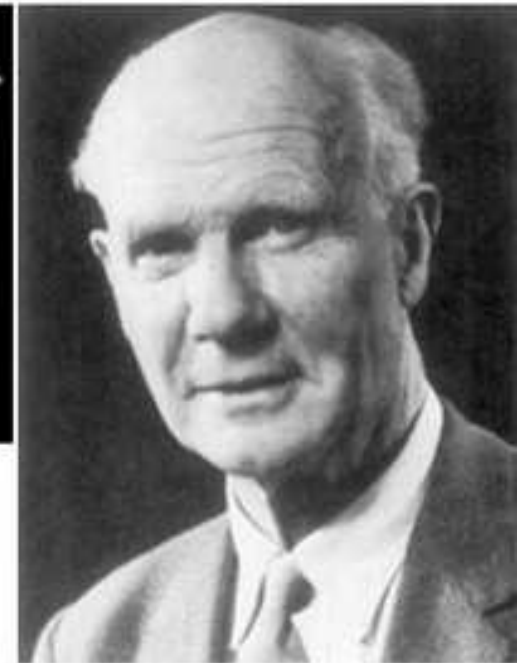
# Portraits de cinq personnages essentiels



Karl Pearson



Ronald Fisher



Egon Pearson



Jerzy Neyman



Erich Leo Lehmann



# Historique de la théorie des tests d'hypothèses

Un résumé en cinq personnages :

- Karl Pearson (1857 – 1936) : études à Cambridge et en Allemagne, professeur de mathématiques (1881) et d'eugénisme (1911) à Londres, un des fondateurs de *Biometrika* (1901).
- Ronald Aylmer Fisher (1890 – 1962) : études à Cambridge, agriculteur (1917), statisticien agronome (1919), auteur de *Statistical Methods for Research Workers* (1925), professeur d'eugénisme à Londres (1933) ...
- Egon Sharpe Pearson (1895 – 1980) : fils de Karl, prend partiellement la place de son père en 1933, travaille avec Neyman de 1925 à 1938.
- Jerzy Spława-Neyman (1894 – 1981), études à Kharkov (1912), Varsovie (1921), stage à Londres (1925) puis Paris (1926), crée un laboratoire de biométrie à Varsovie, crise en 1932, poste à Londres en 1933, développe avec E. Pearson la théorie des tests en introduisant l'hypothèse alternative, la puissance du test ... Berkeley en 1938.
- Erich Leo Lehmann (1917 – 2009) : né à Strasbourg, études à Zurich (1933), Londres (1938) puis Berkeley (1941-1946 avec J. Neyman), enseignant à Berkeley (1947), auteur de *Testing Statistical Hypotheses* (1959), livre dans lequel il prolonge largement le lemme de Neyman-Pearson. Livre posthume : *Fisher, Neyman, and the Creation of Classical Statistics*.

# Tests d'hypothèses : les débuts

- Dans son article publié en 1900 dans Philosophical Magazine Series, Karl Pearson définit la loi du  $\chi^2$  puis, dans l'illustration VII, il l'utilise pour rejeter une hypothèse :
  - ▶ « ... si, sur la cible, les tirs étaient distribués selon une loi normale, alors une observation telle que celle que M. Merriman décrit pourrait seulement apparaître, en moyenne, environ quinze fois sur 10 millions d'essais. »
- Ronald Fisher développa le test t de Student, créa l'analyse de la variance, répandit l'usage de 5% ou 1% comme niveaux pour le risque de rejeter par erreur l'hypothèse de référence ... C'est lui qui a diffusé la méthode des tests d'hypothèses dans de nombreux domaines de la science.
  - ▶ Notons qu'il indiqua explicitement que le fait que l'hypothèse nulle **n'était pas rejetée** par un test **ne** constituait **aucunement une démonstration** de celle-ci.



# L'hypothèse alternative

- Egon Pearson ne pouvait se contenter de l'autorité de R. A. Fisher pour justifier l'emploi des tests ( $t$ ,  $\chi^2$ ,  $F$ , anova) proposés dans le livre de 1925 : **n'existait-il pas de meilleurs tests ?** Et, d'ailleurs, comment **comparer deux tests** ?
- Pearson et Neyman montrèrent d'abord que le principe du *quotient des vraisemblances* conduisait aux tests ci-dessus. Mais cela ne satisfaisait pas Neyman.
- « J'ai trouvé ce qui suit : nous cherchons un contour tel que :
  - ① la probabilité déterminée par  $H$  qu'un échantillon soit situé dans le contour (= rejet) soit égale à  $\varepsilon$ , par exemple 0,01 , et
  - ② que la probabilité déterminée par *une autre hypothèse*  $H'$  qu'un échantillon soit situé dans le contour soit **maximale**. »

(Lettre de J. Neyman à E. Pearson, 20 février 1930.)

# Définitions fondamentales pour les tests d'hypothèses

- On dispose d'une variable aléatoire *vectorielle*  $\mathbf{X}$  dont on sait que la loi de probabilité appartient à  $\mathcal{P} = \{\mathbb{P}_\theta, \theta \in \Omega\}$ ,  $\theta$  étant un paramètre de la loi.
  - On appelle fonction vraisemblance la fonction  $L : (\mathbf{x}, \theta) \mapsto f_\theta(\mathbf{x})$  où  $f_\theta : \mathbf{x} \mapsto f_\theta(\mathbf{x})$  est la fonction densité de la loi de probabilité  $P_\theta$ .  
(On emploie souvent le logarithme de la vraisemblance  $\ln \circ L : (\mathbf{x}, \theta) \mapsto \ln(f_\theta(\mathbf{x}))$ . En effet,  $\mathbf{X}$  est en principe un  $n$ -uplet de variables aléatoires réelles indépendantes; la densité de  $\mathbf{X}$  est donc le **produit** de leurs densités.)

$\Omega$  est partitionné en  $\Omega_{H_0}$  et  $\Omega_{H_1}$ .

- L'**hypothèse**  $H_0$  est  $\theta \in \Omega_{H_0}$  et l'**hypothèse**  $H_1$  est sa négation.  $\Omega_{H_1}$  est l'ensemble des hypothèses alternatives. On note  $S_1$  et on appelle **région critique du test** l'ensemble des valeurs possibles de  $\mathbf{X}$  pour lesquelles la décision de rejeter  $H_0$  serait prise, et  $S_0$  l'ensemble complémentaire.
- On choisit un réel  $\alpha$  compris entre 0 et 1, appelé le niveau du test ou borne du *risque de première espèce*, et on impose, si possible, que
  - pour tout  $\theta \in \Omega_{H_0}$ ,  $\mathbb{P}_\theta\{\mathbf{X} \in S_1\} = \mathbb{P}_\theta(S_1) \leq \alpha$
  - et que, pour tout  $\theta \in \Omega_{H_1}$ ,  $\mathbb{P}_\theta\{\mathbf{X} \in S_1\}$ , appelé puissance du test contre l'alternative  $\theta$ , soit maximisé *du point de vue du choix* de  $S_1$ .
- Ceci définit la fonction puissance du test  $\bar{\beta} : \theta \mapsto \bar{\beta}(\theta) = \mathbb{P}_\theta(\mathbf{X} \in S_1)$  et le *risque de seconde espèce en*  $\theta$ ,  $\beta(\theta) = 1 - \bar{\beta}(\theta)$

# Suite

- L'hypothèse  $H_0$  est dite simple si  $\Omega_{H_0}$  n'a qu'un élément ; sinon, on dit qu'elle est composite. De même pour  $H_1$ .
- Si  $H_1$  est composite et s'il existe un test vérifiant les deux conditions ci-dessus, on dira que ce test est uniformément plus puissant (UPP, *UMP en Anglais*).
- On appelle test trivial de niveau  $\alpha$  le test qui ne rejette jamais  $H_0$ , sauf si un événement extérieur à l'étude et de probabilité  $\alpha$  s'est réalisé. (C'est un *test aléatoire*.) Par construction, ce test est de niveau  $\alpha$ . Sa puissance est constante et égale à  $\alpha$ .  $\beta(\theta) = 1 - \alpha$  : le risque de seconde espèce est égal à 0,95 si  $\alpha=0,05$  !
- On dit qu'un test est non biaisé si
  - ▶ pour tout  $\theta \in \Omega_{H_0}$ ,  $\mathbb{P}_\theta\{\mathbf{X} \in S_1\} = \mathbb{P}_\theta(S_1) \leq \alpha$
  - ▶ et si, pour tout  $\theta \in \Omega_{H_1}$ ,  $\mathbb{P}_\theta\{\mathbf{X} \in S_1\} \geq \alpha$
- Un test UPP est non biaisé (comparaison avec le test trivial). Un test biaisé peut, cependant, être puissant pour certaines valeurs de  $\theta$ . On appelle test UPPNB (UMPU) un test qui vérifie (2) ci-dessus seulement par rapport aux tests non biaisés.

# Le lemme fondamental de Neyman-Pearson et les théorèmes de Lehmann

- *Lemme* : lorsque  $H_0$  et  $H_1$  sont toutes les deux simples, notons  $\Omega_{H_0} = \{\theta_0\}$  et  $\Omega_{H_1} = \{\theta_1\}$ . Alors, quel que soit  $\alpha \in [0; 1]$  il existe un test UPP de niveau  $\alpha$  (UPP : de puissance maximale).  
Ce test est défini par la région critique  $S_1 : S_1 = \{\mathbf{x}, \frac{L(\mathbf{x}, \theta_0)}{L(\mathbf{x}, \theta_1)} \leq k\}$  où  $k$  est choisi de sorte que  $\mathbb{P}_{\theta_0}\{\mathbf{X} \in S_1\} = \mathbb{P}_{\theta_0}(S_1) = \alpha$ .
- Les *théorèmes de Lehmann* prolongent le lemme fondamental, mais seulement lorsque  $\mathcal{P} = \{\mathbb{P}_\theta, \theta \in \Omega\}$  est une **famille à vraisemblance monotone**, c'est-à-dire lorsqu'il existe une statistique réelle  $T$  ( $T = g \circ \mathbf{X}$ ) telle que  $\frac{L(\mathbf{x}, \theta')}{L(\mathbf{x}, \theta)}$  avec  $\theta' > \theta$ , soit une fonction croissante au sens large de  $T$ .

► Exemple :  $X_i$  de loi  $\mathcal{N}(\theta; 1)$  ;  $\frac{L(\mathbf{x}, \theta')}{L(\mathbf{x}, \theta)} = e^{\frac{-n\theta'^2 + n\theta^2}{2}} \cdot e^{(\theta' - \theta) \cdot \sum_{i=1}^n x_i}$  ;  
pour  $T$ , prendre  $\sum x_i$  ou  $\bar{\mathbf{X}}$ .

- 1 *Théorème de Lehmann unilatéral à droite* :  $H_0 : \boxed{\theta \leq \theta_0}$  contre  $H_1 : \boxed{\theta > \theta_0}$  ; région critique  $S_1 = \{\mathbf{x}, T(\mathbf{x}) > k\}$  où  $k$  est défini par  $\mathbb{P}_{\theta_0}(S_1) = \alpha$  : UPP.

# Théorèmes de Lehmann, suite

Ce sont de beaux théorèmes, très simples à énoncer et d'énoncés similaires.  
(*Rappel : valables seulement pour une famille à vraisemblance monotone.*)

- ② Th. unilatéral à gauche, analogue au précédent :  $H_0 : \theta \geq \theta_0$  contre  $H_1 : \theta < \theta_0$  ; région critique  $S_1 = \{\mathbf{x}, T(\mathbf{x}) < k\}$  où  $k$  est défini par  $\mathbb{P}_{\theta_0}(S_1) = \alpha$  ; *UPP*.
- ③ Th. pour  $H_0 : \theta \leq \theta_1 \text{ ou } \theta \geq \theta_2$  contre  $H_1 : \theta_1 < \theta < \theta_2$  (où  $\theta_1 < \theta_2$ ) ; région critique  $S_1 = \{\mathbf{x}, k_1 < T(\mathbf{x}) < k_2\}$  où  $k_1$  et  $k_2$  sont définis par  $\mathbb{P}_{\theta_1}(S_1) = \alpha$  et  $\mathbb{P}_{\theta_2}(S_1) = \alpha$  ; *UPP*.
- ④ Th. pour  $H_0 : \theta_1 \leq \theta \leq \theta_2$  contre  $H_1 : \theta < \theta_1 \text{ ou } \theta > \theta_2$  ; région critique  $S_1 = \{\mathbf{x}, T(\mathbf{x}) < k_1 \text{ ou } T(\mathbf{x}) > k_2\}$  où  $k_1$  et  $k_2$  sont définis par  $\mathbb{P}_{\theta_1}(S_1) = \alpha$  et  $\mathbb{P}_{\theta_2}(S_1) = \alpha$  ; ce test est seulement *UPPSB* ; *il n'existe pas de test UPP*.
- ⑤ Th. pour  $H_0 : \theta = \theta_0$  contre  $H_1 : \theta \neq \theta_0$  ; région critique  $S_1 = \{\mathbf{x}, T(\mathbf{x}) < k_1 \text{ ou } T(\mathbf{x}) > k_2\}$  où  $k_1$  et  $k_2$  sont définis par  $\mathbb{P}_{\theta_0}(S_1) = \alpha$  ; ce test, cas particulier du précédent, est seulement *UPPSB* ; *il n'existe pas de test UPP*.

Remarque : les régions critiques obtenues sont conformes à l'intuition ! ▶

# Test « d'égalité » de deux pourcentages

Dans ce qui suit, nous nous concentrons sur des tests **asymptotiques** de comparaison (ou homogénéité) de deux pourcentages.

Les variables aléatoires  $X_i$  et  $Y_j$  suivent les lois de Bernoulli  $\mathcal{B}(1, \pi_X)$  et  $\mathcal{B}(1, \pi_Y)$ . Donc  $\sum_{i=1}^{n_X} X_i$  suit  $\mathcal{B}(n_X, \pi_X)$  et  $\sum_{j=1}^{n_Y} Y_j$  suit  $\mathcal{B}(n_Y, \pi_Y)$ .

On note  $H_0 : \ll \pi_X = \pi_Y \gg$  et  $H_1 : \ll \pi_X \neq \pi_Y \gg$ . Comme statistique

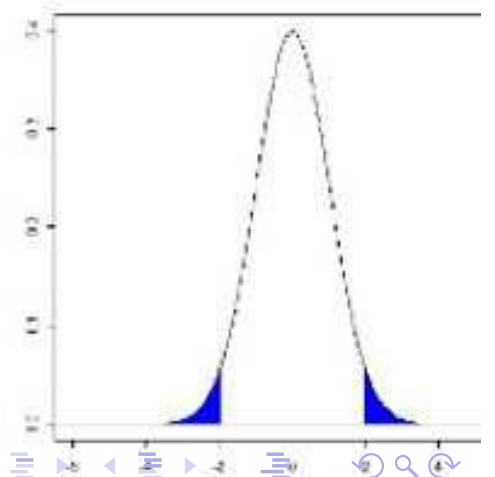
de test, on emploie  $T = \frac{\frac{\sum_{i=1}^{n_X} X_i}{n_X} - \frac{\sum_{j=1}^{n_Y} Y_j}{n_Y}}{\sqrt{P_g \cdot (1 - P_g) \cdot \left(\frac{1}{n_X} + \frac{1}{n_Y}\right)}}$  où  $P_g = \frac{\sum_{i=1}^{n_X} X_i + \sum_{j=1}^{n_Y} Y_j}{n_X + n_Y}$ .

$T$  a, **asymptotiquement**, comme loi de probabilité dans le cadre de  $H_0$ , la loi  $\mathcal{N}(0, 1)$ .

La zone de rejet est déterminée par :

$$\alpha = \mathbb{P}_{H_0}(\text{Rejet de } H_0) = \mathbb{P}_{H_0}(|T| > c)$$

où  $c$  est donc le fractile d'ordre  $1 - \frac{\alpha}{2}$  de la loi  $\mathcal{N}(0, 1)$ .





# Exemple

Une étude randomisée a inclus 200 patients hospitalisés avec une fracture du col du fémur pour comparer la mortalité à 5 ans entre les patients traités chirurgicalement et les patients traités orthopédiquement (*traitement orthopédique : une traction collée peut être posée à titre antalgique. Ce traitement impose un lever au fauteuil et l'absence d'appui sur le membre inférieur pendant trois à six mois.* ).

$$H_0 : \ll \pi_x = \pi_y \gg \text{ contre } H_1 : \ll \pi_x \neq \pi_y \gg.$$

	Décès à 5 ans		
Traitement	Oui	Non	Total
Chirurgical x	15 (15%)	85 (85%)	100 (50%)
Orthopédique y	25 (25%)	75 (75%)	100 (50%)
Total	40 (20%)	160 (80%)	200 (100%)

On trouve que  $T$  a valu  $t = \frac{0,15-0,25}{\sqrt{0,2 \cdot 0,8 \cdot (\frac{1}{100} + \frac{1}{100})}} \cong -1,77$  ; or, **asymptotiquement**, la zone de rejet de  $H_0$  au risque  $\alpha = 0,05$  est bien connue :  $]-\infty; -1,96[ \cup ]1,96; +\infty[$  environ. Donc on ne rejette pas  $H_0$ , ce qui ne démontre absolument pas  $H_0$ . Que faire ?

# La fonction puissance

- Approximativement,  $1 - \beta \cong 1 - \phi \left( z_{1-\frac{\alpha}{2}} - \frac{e}{\sqrt{p_g \cdot (1-p_g)} \cdot \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \right)$  où  $e$  est l'écart entre les proportions,  $p_g$  est ou bien une valeur crédible pour la proportion globale ou bien  $\frac{1}{2}$ ,  $z_a$  est le fractile d'ordre  $a$  de  $\mathcal{N}(0,1)$  et  $\phi$  est sa fonction de répartition.
- Donc la probabilité de repérer un écart de 18% (par exemple) est  $1 - \beta \cong 1 - \phi \left( 1,96 - \frac{0,18}{\sqrt{0,2 \cdot (1-0,2)} \cdot \sqrt{\frac{1}{100} + \frac{1}{100}}} \right) \cong 90\%$  alors que la probabilité de repérer un écart de 10% n'est que 42%.
- Il est possible d'utiliser la formule pour choisir à l'avance la **taille des échantillons**; si on a, pour simplifier,  $n_X = n_Y = n$ , on peut prendre  $n \cong \frac{2 \cdot p_g \cdot (1-p_g) \cdot (z_{1-\frac{\alpha}{2}} + z_{1-\beta})^2}{e^2}$ . Cette formule permet, après avoir *intuité*  $p_g$ , de déterminer l'effectif  $n$  des échantillons avec lequel la puissance  $1 - \beta$  est approximativement obtenue pour un choix de l'écart  $e$  entre les proportions inconnues. Par exemple, pour  $e = 0,1$ ,  $\alpha = 0,05$  et  $1 - \beta = 0,9$ , prendre  $n_X = n_Y = 336$ .

# Les tests unilatéraux

- On a vu précédemment la difficulté principale, qui concerne le non-rejet de  $H_0$ .
- Un autre problème apparaît lors de la comparaison d'un traitement expérimental avec un placebo :

il n'est pas intéressant de s'apercevoir que le traitement est  
« moins ou plus » efficace que le placebo.

Dans ce cas, un *test unilatéral* convient mieux.

# Essai randomisé d'un corticoïde, par infiltration, sur les lombosciatiques

	Succès/échec à J+20		
Traitement	Succès	Échec	Total
Corticoïde x	18 (41,9%)	25 (58,1%)	43(50,6%)
Placebo y	10 (23,8%)	32 (76,2%)	42(49,4%)
Total	28 (32,9%)	57 (67,1%)	85 (100%)

On est en présence de  $H_0 : \ll \pi_x \leq \pi_y \gg$  et  $H_1 : \ll \pi_x > \pi_y \gg$ .

On utilise la même statistique de test,

$$T = \frac{\frac{\sum_{i=1}^{n_X} X_i}{n_X} - \frac{\sum_{j=1}^{n_Y} Y_j}{n_Y}}{\sqrt{P_g \cdot (1 - P_g) \cdot \left( \frac{1}{n_X} + \frac{1}{n_Y} \right)}}$$

mais seules les valeurs *trop* positives de  $T$  permettent ici de rejeter utilement  $\ll \pi_x = \pi_y \gg$ .

Les deux hypothèses sont composites ; ce test sera de niveau  $\alpha$ .

# Essai randomisé d'un corticoïde, par infiltration, sur les lombosciatiques

	Succès/échec à J+20		
Traitement	Succès	Échec	Total
Corticoïde x	18 ( <b>41,9%</b> )	25 (58,1%)	<b>43</b> (50,6%)
Placebo y	10 ( <b>23,8%</b> )	32 (76,2%)	<b>42</b> (49,4%)
Total	28 ( <b>32,9%</b> )	57 (67,1%)	85 (100%)

$$T = \frac{\frac{\sum_{i=1}^{n_X} X_i}{n_X} - \frac{\sum_{j=1}^{n_Y} Y_j}{n_Y}}{\sqrt{P_g \cdot (1 - P_g) \cdot \left( \frac{1}{n_X} + \frac{1}{n_Y} \right)}}$$

La zone de rejet à considérer est  $]c; +\infty[$  pour le cas unilatéral à droite ( $p_x - p_y > 0$  comme ici),  $c$  étant le fractile d'ordre  $1 - \alpha$  de la loi  $\mathcal{N}(0, 1)$ .

(À gauche, ce serait  $]-\infty; -c[$ )

$T$  a valu  $t \cong \frac{0,419 - 0,238}{\sqrt{0,329 \cdot 0,671 \cdot \left( \frac{1}{43} + \frac{1}{42} \right)}} \cong 1,78$  et, asymptotiquement, la zone

de rejet au niveau  $\alpha = 5\%$  est  $]1,65; +\infty[$ . On rejette  $H_0$ .

# L'asymétrie des deux hypothèses

- Nous avons vu que le risque de rejeter par erreur l'hypothèse  $H_0$  est borné par le niveau  $\alpha$ , choisi à l'avance, alors que le risque de ne pas se décider pour  $H_1$  quand il le faut est, certes, minimisé mais ne peut pas être connu, puisqu'il dépend de la valeur inconnue de  $\theta$  ; il peut cependant atteindre  $1 - \alpha$ . (*On n'a pas trouvé mieux comme théorie!*)
- Ne pas rejeter  $H_0$  n'en constitue donc pas (*a priori*) une justification statistique, alors que rejeter  $H_0$  constitue une justification statistique de  $H_1$ .
- La théorie des tests de Neyman et Pearson a été conçue pour détecter des différences (placebo vs substance active, ...) et est adaptée à ce cas.

Mais ...



# Arrive-t-il qu'on doive démontrer, non pas une différence, mais une équivalence ?

- Des exemples :

1



2



3



- Donc

► Une question va se poser :

- ★  $H_0$  étant une hypothèse d'**équivalence**, peut-on échanger  $H_0$  et  $H_1$ , c'est-à-dire : peut-on « démontrer statistiquement » l'hypothèse nulle  $H_0$  ?

# Citation : les médicaments génériques, lever l'opacité

ANSM, Agence nationale de sécurité du médicament et des produits de santé

- **Les études de bioéquivalence** permettent de s'assurer que le devenir du principe actif dans l'organisme est *superposable au princeps*.
- L'étude est réalisée chez des volontaires sains. Elle consiste à comparer, après administration du médicament générique ou du médicament princeps, la concentration plasmatique du produit actif. [...] des prélèvements sanguins permettent de mesurer l'évolution de la concentration dans le sang du principe actif au cours du temps.
- Une courbe comportant en abscisse le temps et en ordonnée la concentration plasmatique en principe actif est établie [...] sont calculés : la concentration maximale en principe actif (notée  $C_{max}$ ) [...] et l'aire sous la courbe des concentrations plasmatiques en principe actif au cours du temps (notée AUC).
- Les courbes obtenues pour le médicament générique et pour le princeps sont ensuite comparées par une *étude statistique*, dont les résultats établissent ou non la bioéquivalence entre les deux spécialités comparées.

# Citation : essais d'équivalence et de non-infériorité

(Caroline ELIE, Service de biostatistique et informatique médicale, faculté de médecine Paris-Descartes, et al.)

- Dans beaucoup de pathologies, l'existence de traitements de référence efficaces fait que les bénéfices escomptés en terme d'efficacité avec de nouveaux traitements sont souvent réduits et donc la démonstration statistique de leur supériorité plus difficile.
- De plus, l'intérêt potentiel du nouveau traitement, outre son éventuelle efficacité thérapeutique, est fondé sur d'autres critères comme
  - ▶ son coût,
  - ▶ sa facilité d'utilisation,
  - ▶ sa voie d'administration,
  - ▶ ses effets indésirables immédiats ou à long terme.
- Dans ce contexte, la problématique devient celle d'une recherche de l'équivalence ou de la non-infériorité du nouveau traitement par rapport au traitement de référence déjà performant.

# Citation : Anton Wellek, tests préliminaires

Tests préliminaires pour contrôler les conditions d'application d'autres méthodes d'inférence statistique

- En consultant des manuels de statistique en tous genres, il est difficile d'en trouver qui ne donnent pas au moins quelques brèves indications sur des méthodes permettant de contrôler les **conditions d'application** sur lesquelles les tests d'hypothèses les plus fréquemment utilisés sont basés.
- Toutes ces indications prennent ce problème de la même façon : on teste l'hypothèse nulle selon laquelle ces conditions d'application sont remplies contre l'hypothèse alternative selon laquelle elles ne le sont pas ...
- Comme ce que l'utilisateur de statistique a en tête est de disposer d'une preuve de l'**exactitude** des conditions requises, il est clair que les tests d'équivalence sont exactement ce que,  **finalement**, il faut faire !
- Il faut donc disposer de tests pour justifier statistiquement la « goodness of fit », et non pas le « lack of fit ».

# Peut-on « démontrer statistiquement » l'hypothèse nulle ?

Exemple du test de l'espérance nulle pour une loi normale d'écart-type 1

Dans le but de *démontrer* que  $\mu = 0$ , existe-t-il un test utilisable de

$$H_0 : \boxed{\mu \neq 0} \text{ contre } H_1 : \boxed{\mu = 0} ?$$

- Selon la définition de Neyman et Pearson, pour  $S_1$ , région critique du test,

on aurait  $\forall \mu \in \mathbb{R} \setminus \{0\} \mathbb{P}_\mu(S_1) = \frac{1}{(\sqrt{2\pi})^n} \iiint \dots \int_{S_1} e^{-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2}} dx_1 dx_2 \dots dx_n \leq \alpha$

- La fonction puissance du test se réduirait au réel

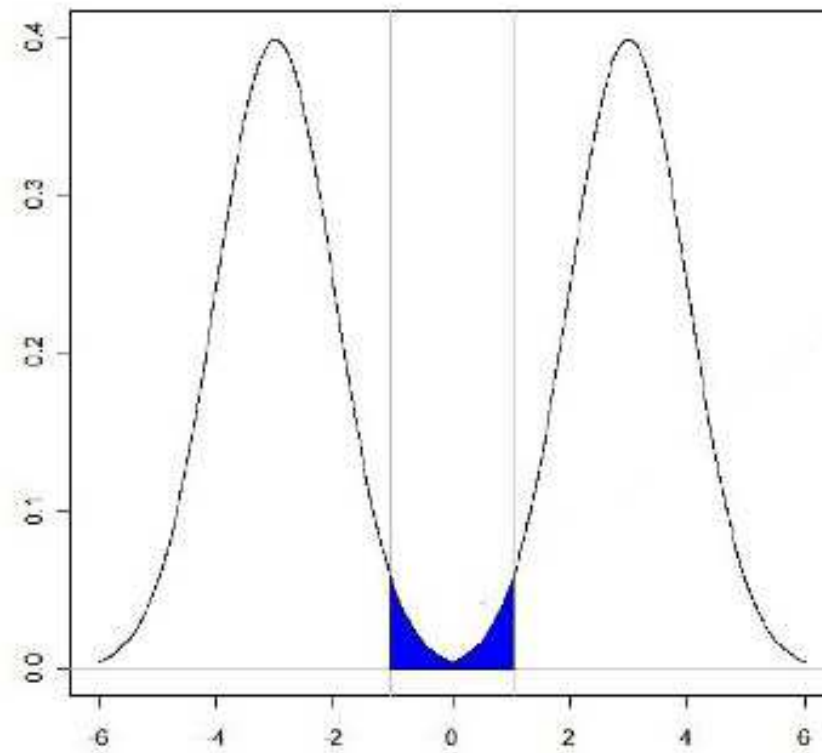
$$\bar{\beta}(0) = \mathbb{P}_0(S_1) = \frac{1}{(\sqrt{2\pi})^n} \iiint \dots \int_{S_1} e^{-\frac{\sum_{i=1}^n x_i^2}{2}} dx_1 dx_2 \dots dx_n.$$

- Pour chaque  $i$ , la fonction  $(x_i, \mu) \mapsto e^{-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2}}$  est continue et dominée par 1 ; donc  $\bar{\beta}(0) = \lim_{\mu \rightarrow 0} \mathbb{P}_\mu(S_1) \leq \alpha$ .

Ainsi la puissance d'un tel test serait au maximum celle du test trivial !

Donc il est généralement impossible de démontrer statistiquement  $H_0$  lorsque cette hypothèse est simple et  $H_1$  est composite (ce qui est cependant très fréquent) ...

# Présentation des tests d'équivalence ou de non-infériorité dans le domaine des essais cliniques





# Tests d'équivalence

- Pour traiter une maladie, on dispose d'un traitement **standard** et, de plus, on étudie un nouveau traitement, **expérimental**.
- On juge la qualité de chacun des traitements à l'aide du pourcentage des patients guéris (ou bien soulagés, par exemple).

Le non-rejet de  $H_0 : \ll \pi_e = \pi_s \gg$  **n'est pas une démonstration** de  $H_0$ .

L'échange littéral de  $H_0 : \ll \pi_e = \pi_s \gg$  et  $H_1 : \ll \pi_e \neq \pi_s \gg$  n'étant pas possible, on se contentera d'**imposer un maximum d'écart** entre l'efficacité du traitement standard et l'efficacité du traitement expérimental en prenant

- $H_0 : \ll |\pi_e - \pi_s| \geq \delta \gg$       • (c'est la non-équivalence)
- et  $H_1 : \ll |\pi_e - \pi_s| < \delta \gg$       • (c'est l'équivalence),

le choix de la valeur de  $\delta$  relevant, certes, de l'équipe médicale et/ou du statisticien, mais essentiellement des « bonnes pratiques » préconisées par les organismes de contrôle, nationaux et internationaux.

# Test de non-infériorité

En fait, pour les études de médicaments avec *contrôle actif*, on utilise quasi exclusivement la version unilatérale de ce test. On note  $\pi_s$  la probabilité de succès avec la substance active de référence,  $\pi_e$  celle qui correspond au produit en cours d'essai.

- $H_0 : \ll \pi_e - \pi_s \leq -\delta \gg$
- $H_0 : \ll \pi_s - \pi_e \geq \delta \gg$
- et  $H_1 : \ll \pi_e - \pi_s > -\delta \gg$ .
- et  $H_1 : \ll \pi_s - \pi_e < \delta \gg$

Rejeter  $H_0$  signifie ici que

- $\pi_e > \pi_s$  si  $\pi_s - \pi_e < 0$
- ou, sinon, que  $\pi_e$  n'est pas trop inférieur à  $\pi_s$  :  $\pi_s \geq \pi_e > \pi_s - \delta$ .

*« The type I error rate for these tests is now the likelihood that we conclude the treatment group is not inferior to the control, when in fact it really is inferior.*

*When we control the type I error for these noninferiority hypotheses, we are protecting against incorrectly concluding noninferiority for the treatment group compared to the control group. »*

# Mise en oeuvre d'un test de non-infériorité

Comme on l'a vu, on définit

- $H_0 : \ll \pi_s - \pi_e \geq \delta \gg$
- et  $H_1 : \ll \pi_s - \pi_e < \delta \gg$ .

et, donc, on peut utiliser comme valeur de la réalisation de la statistique de test

$$t = \frac{p_s - p_e - \delta}{\sqrt{\left( \frac{p_s \cdot (1-p_s)}{n_s} + \frac{p_e \cdot (1-p_e)}{n_e} \right)}}$$

(translation des données de  $\delta$  vers la gauche).

- Ce sont les valeurs ***trop* négatives** de  $t$  qui permettent de rejeter  $H_0$ .
- La zone de rejet est donc  $] -\infty ; -c[$ ,  
 $c$  étant le fractile d'ordre  $1 - \alpha$  de la loi  $\mathcal{N}(0, 1)$ ,  
ce qui donne  $] -\infty ; -1,65[$  au risque  $\alpha = 5 \%$ .

# Exemple

L'allogreffe de cellules souches hématopoïétiques est utilisée pour le traitement de certaines maladies hématologiques. Une étude de 2007 a eu pour but de comparer l'emploi de cellules sources du sang périphérique (PBSC) avec celui de la moëlle osseuse (BM). [..]

On voit que l'avantage du sang périphérique [...] beaucoup moins invasif pour le donneur ; de plus, des études indiquent qu'il a une meilleure efficacité que BM dans le cas des donneurs **jumeaux des receveurs**. Cependant, dans le cas des donneurs non apparentés aux receveurs, le risque de réaction du greffon contre l'hôte semble accru.

On prend donc ici BM comme traitement expérimental, PBSC comme traitement de contrôle (de référence) et on cherche à savoir si BM n'est, au pire, que peu inférieur à PBSC pour la survie à 6 mois. Le choix a été fait de  $\delta = 0,1 = 10\%$ .

# Mise en oeuvre du test

Données	n	mortalité liée au traitement
BM	583	187
PBSC	328	95

- La statistique de test vaut donc

$$t = \frac{\frac{233}{328} - \frac{396}{583} - 0,1}{\sqrt{\left(\frac{\frac{233}{328} \cdot (1 - \frac{233}{328})}{328} + \frac{\frac{396}{583} \cdot (1 - \frac{396}{583})}{583}\right)}} \cong -2,18$$

- Or la zone de rejet pour un « z-test » unilatéral à gauche au risque  $\alpha = 5\%$  est environ  $] -\infty ; -1,65[$ .
- Donc, à ce risque, on rejette l'hypothèse  $H_0 : \ll \pi_s - \pi_e \geq \delta \gg$  et on déclare que BM est non-inférieur à PBSC du point de vue de la mortalité liée au traitement,
- (en plus de sa plus faible tendance à la réaction du greffon contre l'hôte).

## Autre exemple

Objectif de l'étude NECT : **comparaison** de la combinaison thérapeutique d'eflornithine en administration I.V. (deux fois par 24 h, sept jours) et de nifurtimox par voie orale (10 jours) **au** traitement de référence d'eflornithine I.V. (quatre fois par 24 h, 14 jours) en termes d'efficacité thérapeutique pour le traitement des patients atteints de trypanosomiase TBG en phase méningo-encéphalique.

Données	n	nombre d'échecs
EN	47	3
E	51	5

La statistique de test vaut donc ici

$$t = \frac{\frac{46}{51} - \frac{44}{47} - 0,1}{\sqrt{\left( \frac{\frac{46}{51} \cdot (1 - \frac{46}{51})}{51} + \frac{\frac{44}{47} \cdot (1 - \frac{44}{47})}{47} \right)}} \cong -2,45$$

Même conclusion : au risque  $\alpha = 5\%$ , on rejette l'hypothèse  $H_0$  :

«  $\pi_s - \pi_e \geq \delta$  » et on déclare que EN est non-inférieur à E du point de vue de l'efficacité, et plus simple à mettre en oeuvre.

Défaut : les données sont telles que l'approximation normale n'est pas totalement justifiée ici ...



# Test de non-infériorité et intervalle de confiance

En admettant que l'approximation normale soit justifiée, un intervalle de confiance de niveau  $1 - \alpha$  de la différence  $\pi_X - \pi_Y$  est

$$\left[ p_X - p_Y - z_{1-\frac{\alpha}{2}} \cdot \sqrt{v} ; p_X - p_Y + z_{1-\frac{\alpha}{2}} \cdot \sqrt{v} \right] \text{ où } v = \frac{p_X \cdot (1-p_X)}{n_X} + \frac{p_Y \cdot (1-p_Y)}{n_Y}$$

et  $z_p$  est le fractile d'ordre  $p$  de  $\mathcal{N}(0,1)$ .

On définit aussi « intervalle de confiance à gauche de niveau  $1 - \alpha$  » :

$$\left[ p_X - p_Y - z_{1-\alpha} \cdot \sqrt{v} ; +\infty \right[ ; \text{ de même à droite.}$$

Or, dans un test de non-infériorité,

$H_0 : \ll \pi_e - \pi_s \leq -\delta \gg$  est rejetée si et seulement si

$$\frac{p_e - p_s + \delta}{\sqrt{\left( \frac{p_e \cdot (1-p_e)}{n_e} + \frac{p_s \cdot (1-p_s)}{n_s} \right)}} > z_{1-\alpha}$$

$$\text{c'est-à-dire } p_e - p_s - z_{1-\alpha} \cdot \sqrt{\left( \frac{p_s \cdot (1-p_s)}{n_s} + \frac{p_e \cdot (1-p_e)}{n_e} \right)} > -\delta.$$

On voit donc que  $H_0$  est rejetée si et seulement si l'intervalle de confiance à gauche de  $\pi_e - \pi_s$  de niveau  $1 - \alpha$  est inclus dans  $]-\delta ; +\infty[$ .

# Test de non-infériorité et intervalle de confiance symétrique

- Si on ne veut pas utiliser d'intervalle de confiance non symétrique, on peut aussi dire que  $H_0$  est rejetée si et seulement si
  - ▶ l'intervalle de confiance de  $\pi_e - \pi_s$  de niveau  $1 - 2.\alpha$  a sa borne gauche **incluse** dans  $]-\delta ; +\infty[$ .

# Tests d'équivalence et inclusion d'intervalles

Test proposé par Westlake en 1972 pour les hypothèses :

$$H_0 : \boxed{\theta \leq \theta_0 - \varepsilon_1 \text{ ou } \theta \geq \theta_0 + \varepsilon_2} \text{ contre } H_1 : \boxed{\theta_0 - \varepsilon_1 < \theta < \theta_0 + \varepsilon_2}$$

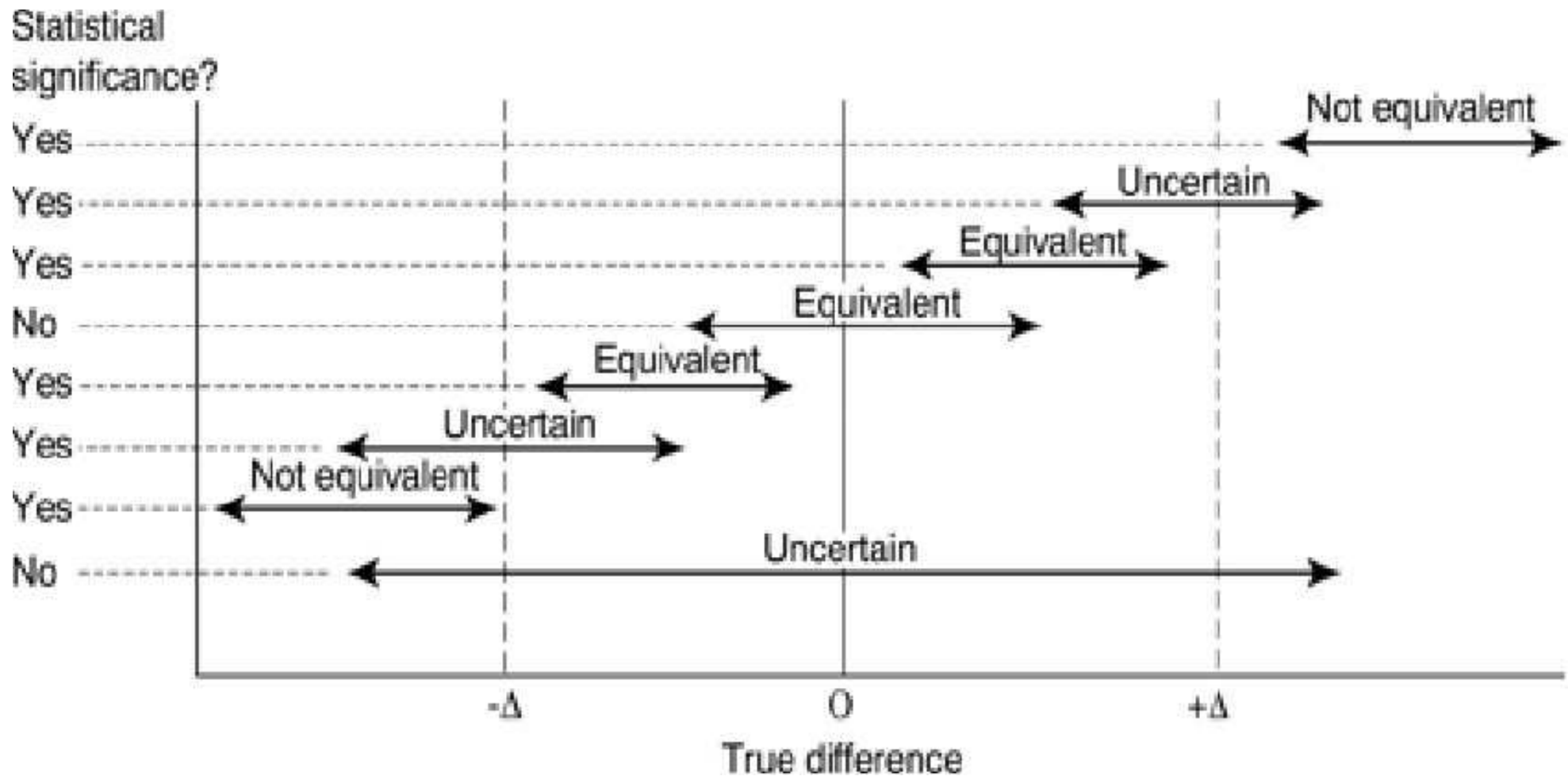
où  $\theta_0$  est le nombre correspondant exactement au cas de l'équivalence, (*par exemple 0 pour la différence des valeurs de deux indicateurs, 1 pour leur quotient*),  $\varepsilon_1$  et  $\varepsilon_2$  étant généralement égaux à  $\delta$ .

- $\underline{\theta}(\mathbf{x}; \alpha)$  étant la borne gauche d'un intervalle de confiance à gauche de niveau  $\alpha$  pour  $\theta$  et  $\bar{\theta}(\mathbf{x}; \alpha)$  étant la borne droite d'un intervalle de confiance à droite de niveau  $\alpha$  pour  $\theta$ ,
- Ce test rejette  $H_0$  si et seulement si l'intervalle aléatoire  $[\underline{\theta}(\mathbf{x}; \alpha); \bar{\theta}(\mathbf{x}; \alpha)]$  est complètement recouvert par l'intervalle  $]\theta_0 - \varepsilon_1; \theta_0 + \varepsilon_2[$

(Cela équivaut à  $\boxed{\underline{\theta}(\mathbf{x}; \alpha) > \theta_0 - \varepsilon_1 \text{ et } \bar{\theta}(\mathbf{x}; \alpha) < \theta_0 + \varepsilon_2}$ .)

Or  $[\underline{\theta}(\mathbf{x}; \alpha); \bar{\theta}(\mathbf{x}; \alpha)]$  est un intervalle de confiance bilatéral pour  $\theta$  de niveau  $1 - 2\alpha$ . Il n'est donc pas indispensable d'utiliser la notion d'intervalle de confiance *unilatéral* pour définir ce test d'équivalence.

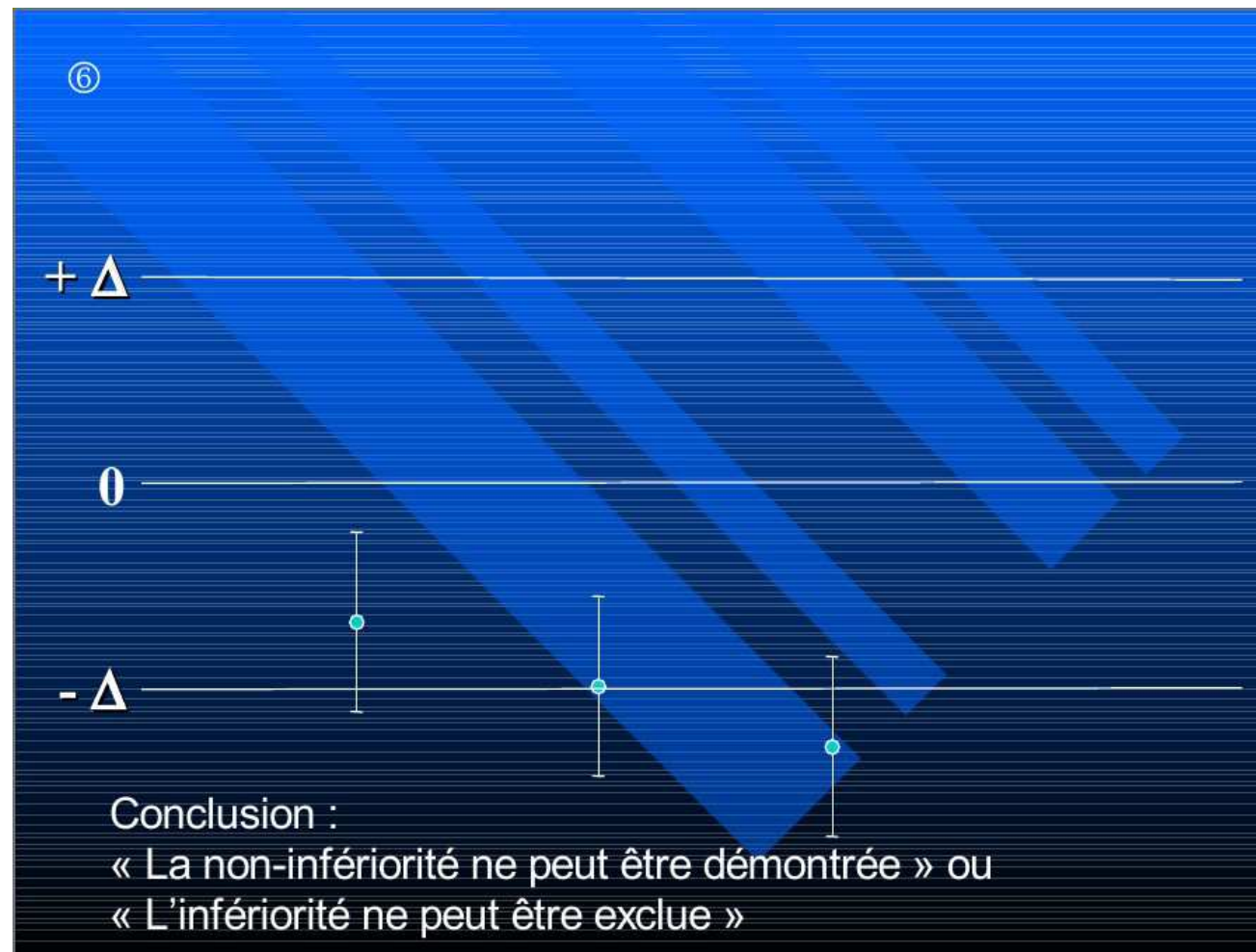
# Tests d'équivalence : un graphique



Ce graphique figure parmi les documents destinés aux étudiants de psychologie de l'université du Nord Texas.

Il donne la vision que ces étudiants sont censés avoir au sujet des tests d'équivalence.

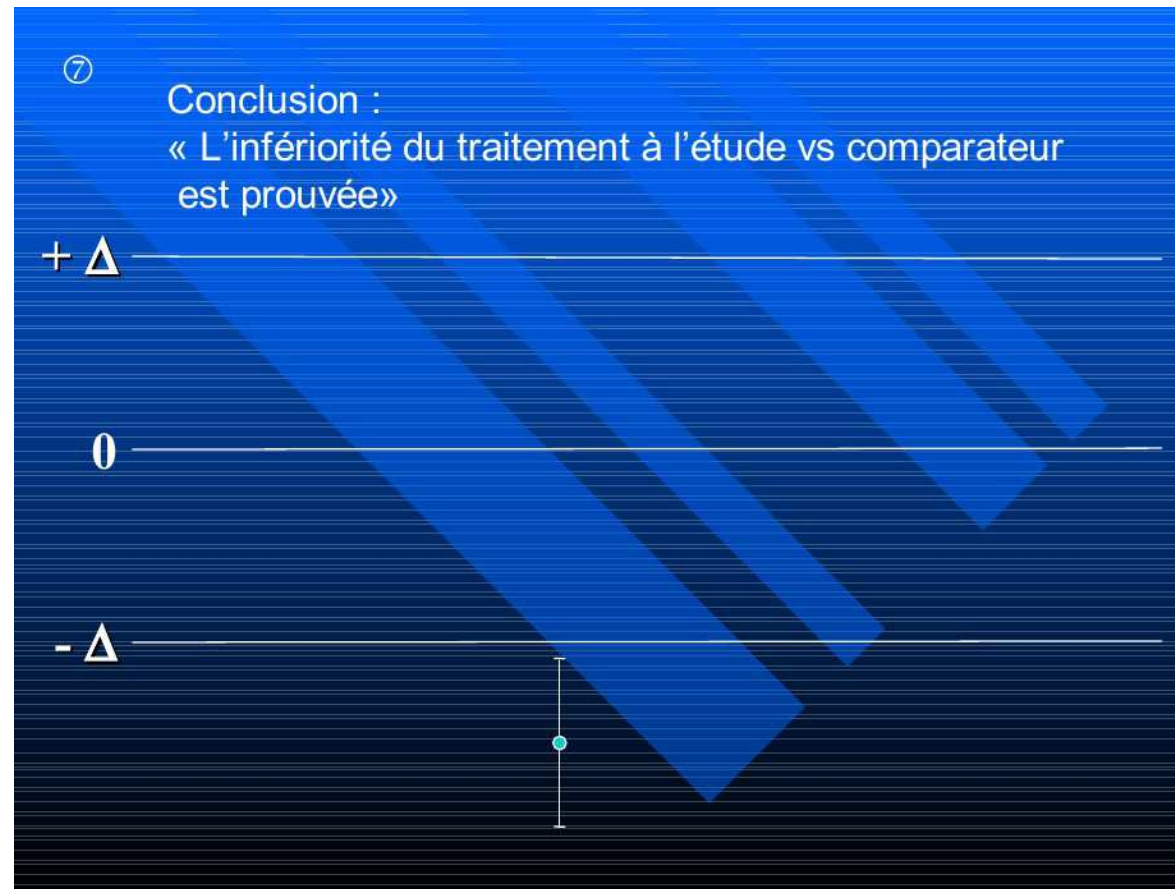
# Tests d'équivalence : un graphique plus local



Ce graphique-ci est extrait, parmi une dizaine d'autres du même genre, d'un polycopié de médecine de l'Université de Rouen.

Le but est le même.

# Suite



Comme dans l'avant-dernier graphique, le cas où l'intervalle de confiance est situé complètement en-dehors de  $]\theta_0 - \varepsilon_1; \theta_0 + \varepsilon_2[$  (*ici, en-dessous de cet intervalle*) est présenté. Et, en effet,  $p_e - p_s + z\sqrt{v} < -\delta$  équivaut à  $\frac{p_s - p_e - \delta}{\sqrt{v}} > z$ , ce qui est la condition de rejet de  $H_0 : \pi_s - \pi_e \leq \delta$  en faveur de  $H_1 : \pi_s - \pi_e > \delta$  »

# Un exemple de test d'équivalence

Staszewski et al. (JAMA. 2001) reported the results of a clinical trial in 562 HIV patients designed to demonstrate equivalence between an abacavir-lamivudine-zidovudine therapy and an indinavir-lamivudine-zidovudine therapy. The primary endpoint was the proportion of patients having an HIV RNA level of 400 copies/ml or less at week 48. Based on discussions with researchers, clinicians, and the FDA, the equivalence margin for the difference in proportions was set at  $\delta=12$  percentage points [...] the response rates were 50.8% and 51.3% for abacavir and indinavir, respectively. The reported 95% confidence interval for the difference in the response rates was  $(-9, 8)$  and, since it is included in  $(-12, 12)$ , the two therapies could be declared equivalent at the 0.025 significance level.

*Esteban Walker, PhD and Amy S. Nowacki, PhD, Department of Quantitative Health Sciences, Cleveland Clinic, Cleveland, OH, USA. 2010.*

	A	B	C	D	E	F	G
1		Effectif	Echec	Succès	%		
2	abacavir	262	129	133	0.508		
3	indinavir	265	129	136	0.513		
4	+ ou -	527	258	269	-0.00557		
5	Pas pris part	35					
6		562					
7	Z	1.960	LOI.NORMALE.STANDARD.INVERSE(0.975)				
8	Borne g :	-0.0909	E2-E3-B7*RACINE(E3*(1-E3)/B3+E2*(1-E2)/B2)				
9	Borne d :	0.0798	E2-E3+B7*RACINE(E3*(1-E3)/B3+E2*(1-E2)/B2)				

```

> x<-c(133,136)
> n<-c(262,265)
> prop.test(x,n,conf.level=0.95)

2-sample test for equality of proportions
with continuity correction
data:  x out of n
X-squared = 0.0017, df = 1, p-value = 0.9674
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.09472809  0.08358017
sample estimates:
 prop 1    prop 2 
0.5076336 0.5132075

```

# Test t pairé d'équivalence UPPI, selon A. Wellek

Lors d'une étude (1992) sur la pression intracrânienne chez les lapins, un test a été fait pour s'assurer que, préalablement, les mesures de flux restaient stables pendant 15 minutes. Dans un échantillon de  $n = 23$  animals, la variation moyenne a été de  $\bar{D} = 0,16$  [ml/min/100g de masse corp.]. Pour l'écart-type du changement intraindividuel, on a trouvé  $S_D = 3.99$ . Les limites d'équivalence pour la variation **standardisée** du flux **moyen** ont été choisies égales à  $\pm\delta = \pm 0,50$ . D'après ces données, la statistique standardisée  $T$  a valu  $\frac{0,16}{\frac{3,99}{\sqrt{23}}} \cong 0,1923$ .

D'après Wellek, la zone de rejet pour ce test de niveau  $\alpha = 0,05$  de  $H_0$  :

la moyenne des changements n'est pas nulle contre  $H_1$ , son contraire, est

donnée par  $-C < T < C$  où  $C = \sqrt{F^{-1}(\alpha)}$ ,  $F$  étant la fonction de répartition de la loi  $F$  de Fisher-Snedecor à 1 et  $n - 1$  degrés de liberté, **décentrée de**  $n \cdot \delta^2 = 23 \cdot 0,5^2 = 5,75$ . Commande R pour obtenir la valeur de  $C$  :

`> sqrt(qf(0.05,1,22,5.75))` [1] 0.7594576

On a  $-0.7595 < 0,1923 < 0.7594$  , donc on rejette  $H_0$  :

We can conclude that at the 5% level, the experimental data of the present example contain sufficient evidence in support of the hypothesis that the cortical microflow of rabbits does not change to a relevant extent over a time interval of 15 minutes during which no active treatment is administered.



# Un peu de naïveté ...

*La démonstration donnée par Wellek pour le test précédent est assez ardue. Pour changer, soyons naïfs : confondons une variance inconnue et dépendant de paramètres avec un nombre connu et constant ...*

Muni de ces **redoutables approximations**, reprenons l'exemple de l'étude concernant le VIH, pour laquelle les taux de réponses positives étaient  $\frac{133}{262}$

et  $\frac{136}{265}$  ;  $\pm\delta = \pm 0,12$ . Notons  $\sigma = \sqrt{\left(\frac{\frac{133}{262} \cdot (1 - \frac{133}{262})}{262} + \frac{\frac{136}{269} \cdot (1 - \frac{136}{269})}{269}\right)} \cong 0,04340$ .

Pour appliquer ici le troisième théorème de Lehmann dans le cas d'une loi normale, Tassi dit de déterminer un intervalle  $[\beta_1; \beta_2]$  de centre  $\frac{\delta - (-\delta)}{2 \cdot \sigma} \cong 2.7652$  tel que  $\mathbb{P}(\beta_1 \leq U < \beta_2) = \alpha = 0,05$ , l'intervalle de rejet (Lehmann) étant alors  $]k_1; k_2[$  où  $k_1 = \beta_1 \cdot \sigma - \delta$  et  $k_2 = \beta_2 \cdot \sigma - \delta$ . On écrit une fonction R qui donne directement  $k_1$  et  $k_2$  :

```
> tassi(-0.12,0.12,0.0434,0.05)
```

```
[1] -0.0486348
```

```
[1] 0.0486348
```

Comme la statistique de test a

valu  $\frac{133}{262} - \frac{136}{265} \cong -0.0056$ , on

pourrait rejeter  $H_0$  en disant que

les deux combinaisons médicamenteuses ont une efficacité équivalente ... si  
*ce qui précède était plus sérieux !*

# Pour en savoir plus

## Références :

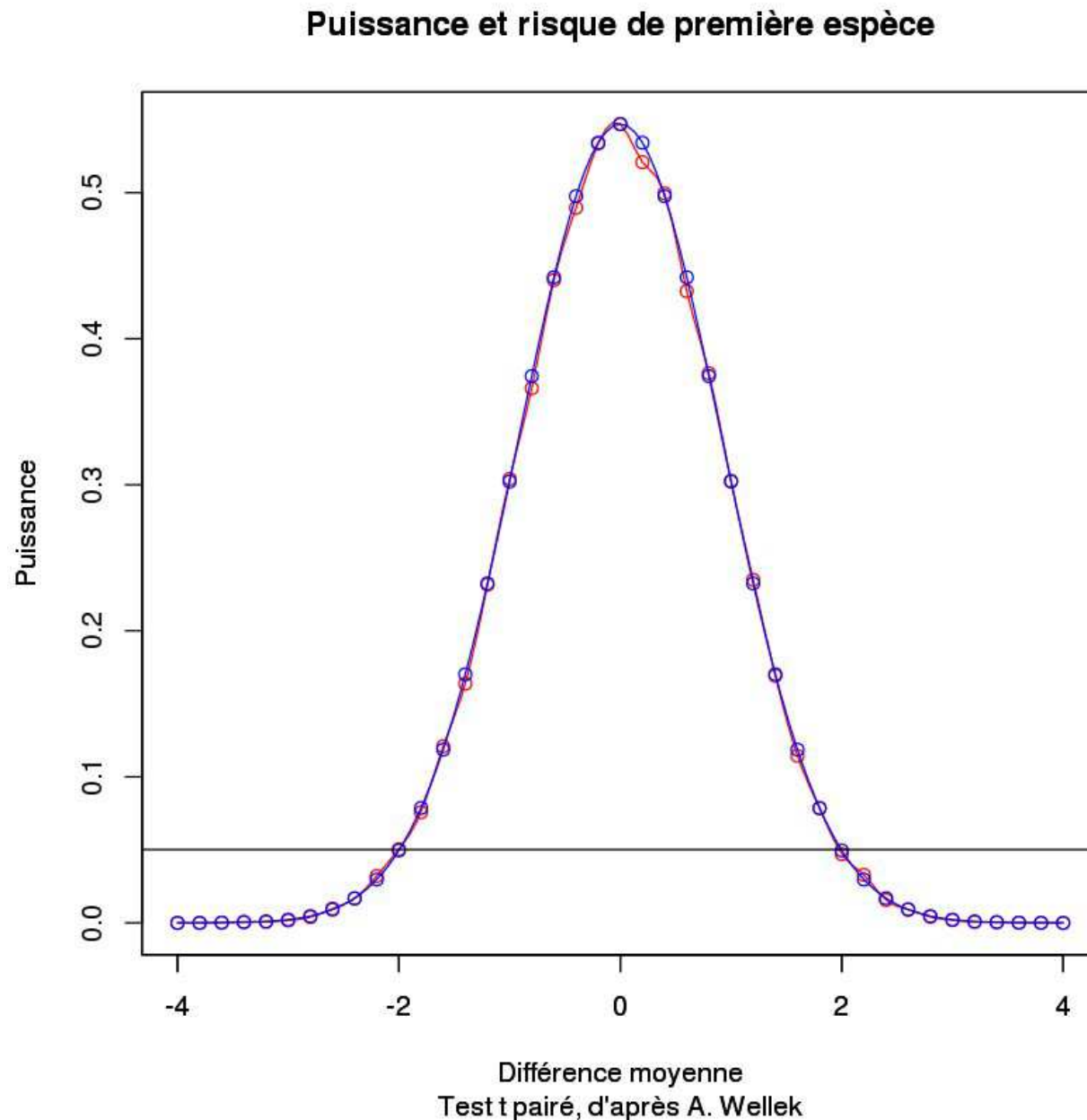
- Statistical Methods for Research Workers, Ronald Aylmer Fisher, 1925, Oliver & Boyd  
(en ligne : [http ://krishikosh.egranth.ac.in/bitstream/1/2028664/1/G3681.pdf](http://krishikosh.egranth.ac.in/bitstream/1/2028664/1/G3681.pdf))
- Testing Statistical Hypotheses, Erich Leo Lehmann, 1959, John Wiley & Sons
- Méthodes statistiques, Philippe Tassi, 1985, Economica
- Testing Statistical Hypotheses of Equivalence and Noninferiority, Second Edition, Stefan Wellek, 2010, CRC Press - Chapman & Hall
- Fisher, Neyman, and the Creation of Classical Statistics, Erich L. Lehmann, 2011, Springer  
(en ligne : [http ://en.bookzz.org/book/2050616/72a379](http://en.bookzz.org/book/2050616/72a379) )

# Annexe

## Plan

- Simulations donnant la puissance et le risque de première espèce pour les tests précédents
- Le code de fonctions R utilisées pour la simulation
- La méthode de Tassi
- Test classique d'égalité de deux espérances
- Test de non-infériorité pour deux espérances
- Année post-bac : comment les tests d'hypothèses sont-ils présentés ?

# Test t pairé selon Anton Wellek



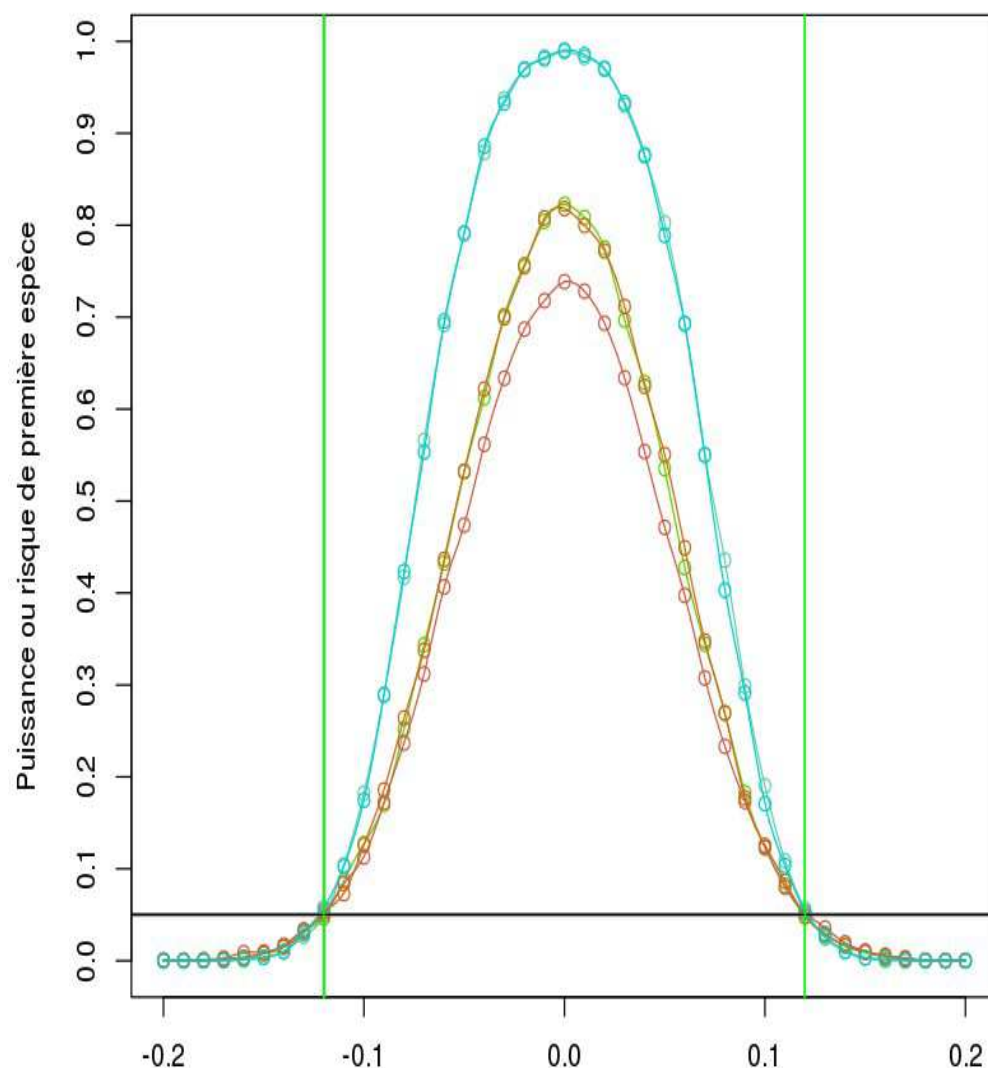
Le risque de seconde espèce,  $\beta(\theta)$ , est la probabilité de ne pas rejeter  $H_0$  alors que le paramètre vaut  $\theta$ , disons à tort.

La puissance en  $\theta$  est  $1 - \beta(\theta)$ , la probabilité du contraire, rejeter  $H_0$ .

Le risque de première espèce, de majorant  $\alpha$ , est la probabilité de rejeter  $H_0$  alors que le paramètre vaut  $\theta$ , disons à tort.

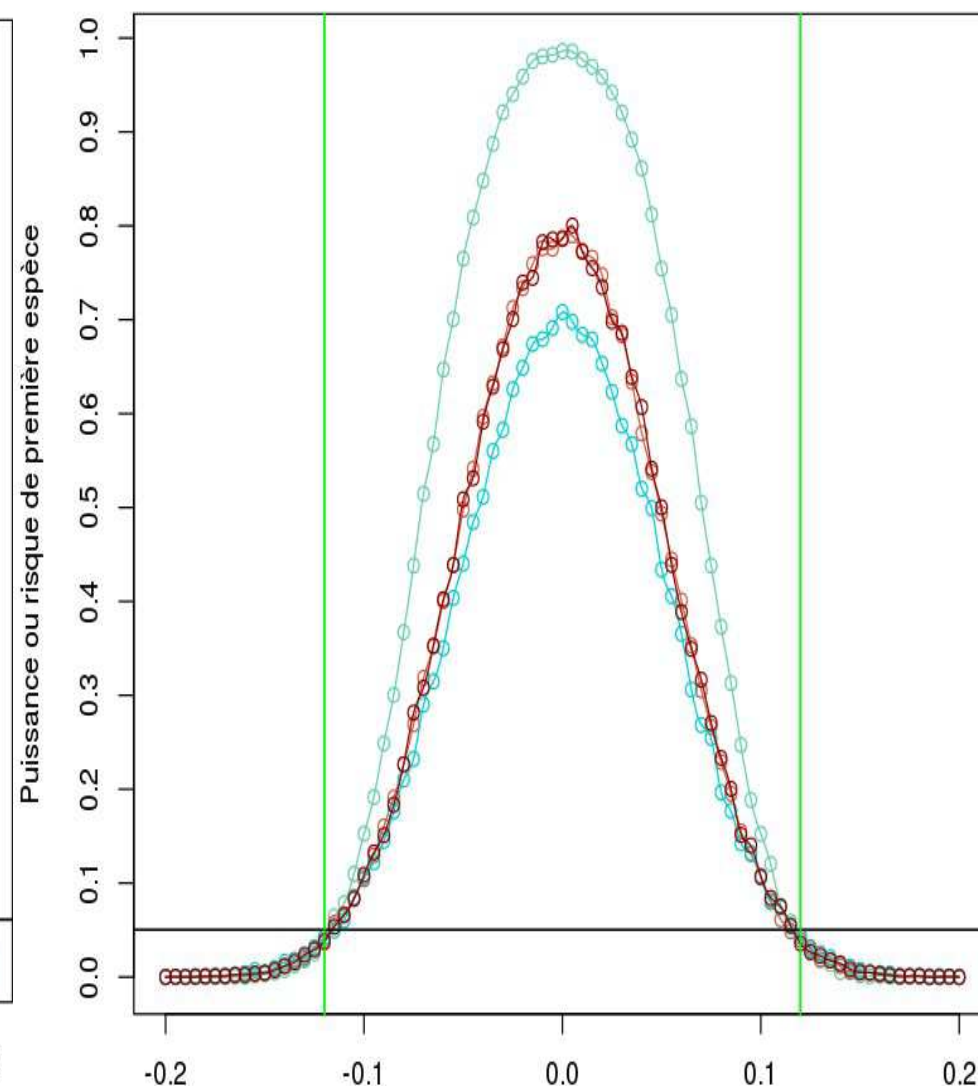
# Suivant une méthode inspirée de P. Tassi

Puissance et risque de première espèce



Différence des pourcentages  
Test bricolé, d'après P. Tassi

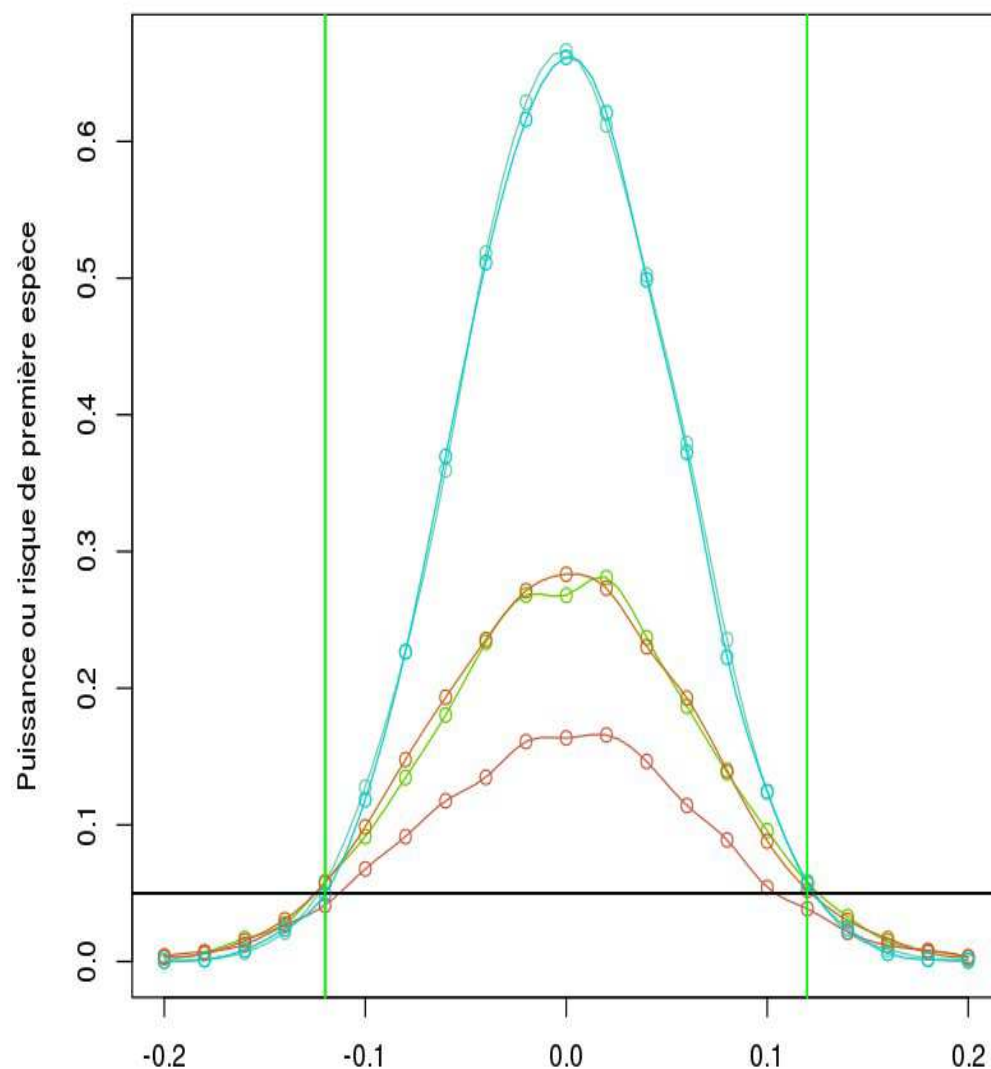
Puissance et risque de première espèce



Différence des pourcentages  
Test classique, données abacavir contre indinavir

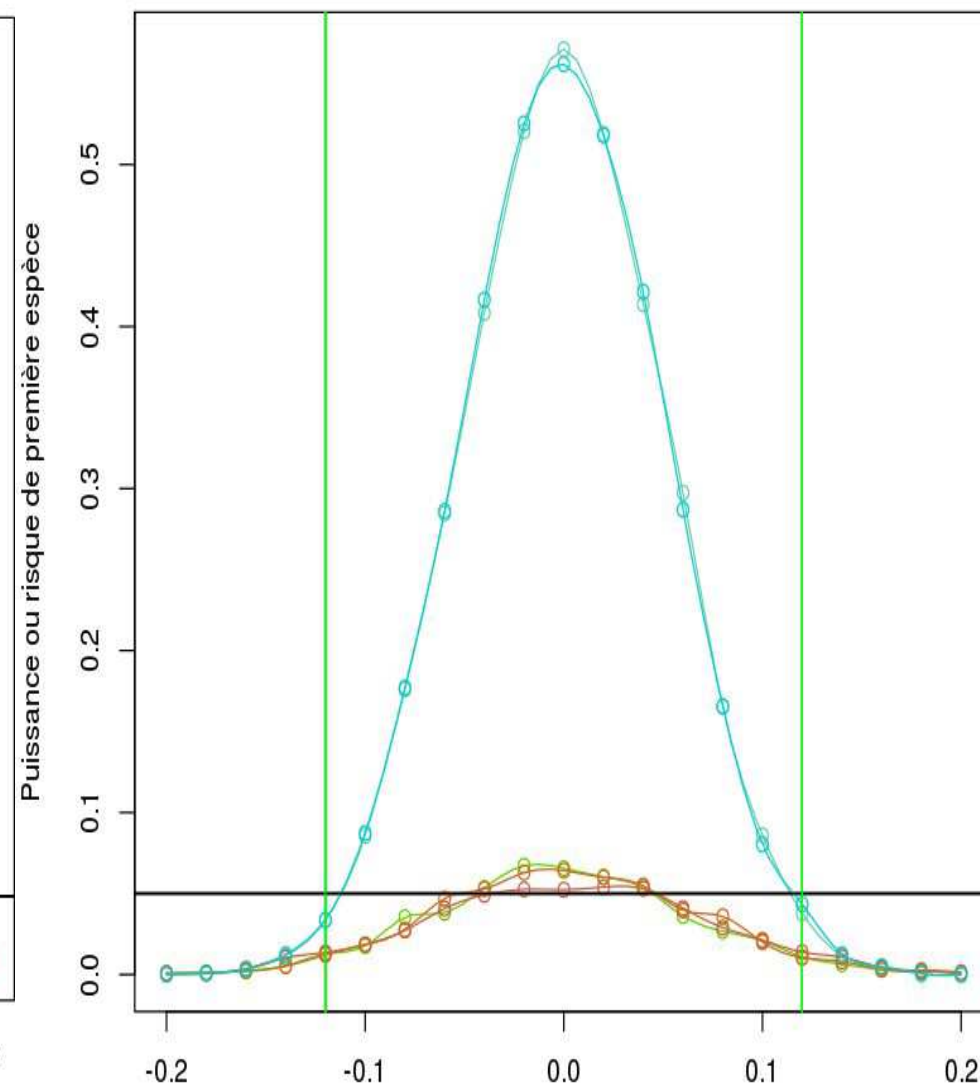
# Idem avec des effectifs de 100 et 103

Puissance et risque de première espèce



Différence des pourcentages  
Test bricolé, d'après P. Tassi

Puissance et risque de première espèce



Différence des pourcentages  
Test classique, données abacavir contre indinavir

# La fonction R simul (wellek)

```

simul <- function(d,ec,n,delta,alpha,nb){
# Commentaire.  $d$  : différence des espérances à étudier,  $ec$  : écart-type,  $n$  : taille
de l'échantillon,  $delta$  : borne d'équivalence,  $alpha$  : niveau du test,  $nb$  : nombre
de répétitions pour la simulation
  compteur <- 0
# Cette variable contiendra le nombre de fois où  $H_0$  aura été rejetée
  b <- sqrt(qf(alpha,1,n-1,n*delta^2))
# Cette variable contient la borne droite de l'intervalle de rejet
  for (i in 1 :nb){
# Il s'agit d'une boucle qui sera effectuée  $nb$  fois
    data <- rnorm(n,d,ec)
# Tirage aléatoire d'un échantillon de taille  $n$  suivant la loi  $\mathcal{N}(d;ec)$ 
    t <- sqrt(n)*mean(data)/sd(data)
# Standardisation de la moyenne des nombres de l'échantillon
    if ((t<b)&(t>-b)){
# Cette moyenne standardisée est-elle située entre les bornes  $-b$  et  $b$  ?
      compteur <- compteur+1
# Si oui, c'est une fois de plus, le rejet de  $H_0$ 
    }
  }
}

```

# La fonction R simul (wellek) : suite

```
}  
  v <- c(compteur/nb,pt(b,n-1,ncp=d*sqrt(n)/ec)-pt(-b,n-1,ncp=d*sqrt(n)/ec))  
# Cette variable vectorielle contient la proportion de rejets de  $H_0$  estimée par  
simulation et aussi calculée théoriquement  
  return(v)  
# Renvoi de ce vecteur à la fonction appelante  
}
```



# La fonction R rep\_simul (wellek)

```
rep_simul <- function(ec,n,delta,alpha,nb,max,pas){
# Commentaire. ec : écart-type, n : taille de l'échantillon, delta : borne d'équivalence,
alpha : niveau du test, nb : nombre de répétitions pour la simulation, max : maximum
(en valeur absolue) de la différence des espérances à étudier, pas : pas !
  j <- 1
  resS <- numeric(0)
  resT <- numeric(0)
# vecteurs vides (dimension zéro) dans lesquels seront stockés les résultats, resS : par
simulation, resT : en employant la loi t de student décentrée
  valeursx <- seq(-max,max,pas)
# contient les valeurs de ces différences, par exemple de -0,9 à +0,9 par pas de 0,1
  for (i in valeursx){
# La variable de boucle est i ; i contient la différence des espérances
    ret <- simul(i,ec,n,delta,alpha,nb);
# La fonction simul renvoie un couple de nombres
    resS[j] <- ret[1]
    resT[j] <- ret[2]
    j <- j+1
# j compte le numéro d'ordonnée pour remplir resS et resT
  }
```

# La fonction R rep\_simul (wellek) : suite

```

plot(valeursx,resS,type="n",main="Puissance et risque de première
espèce",sub="Test t pairé, d'après A. Wellek",xlab="Différence
moyenne",ylab="Puissance ou risque de première espèce")
# Préparation du graphique
points(valeursx,resS,type="p",col="red")
# Affichage des points connus sur la courbe "simulation"
lines(spline(valeursx, resS),col="red")
# Joindre ces points
points(valeursx,resT,type="p",col="blue")
lines(spline(valeursx, resT),col="blue")
# de même pour la courbe "loi t décentrée"
abline(h=alpha)
abline(v=-delta,col="green")
abline(v=delta,col="green")
# tracer des droites utiles (une horizontale et deux verticales)
}
# Exemple d'appel de la fonction : rep_simul(3.99,23,0.5,0.05,10000,0.9,0.1)

```

# La fonction R tassi

```
function(m1,m2,sigma,alpha){
  lambda <- (m1-m2)/sigma;
  c <- -lambda/2;
  epsi <- 1e-6;
  cont <- TRUE;
  pas <- c;
  x <- c;
  while(cont) {
    aire <- abs(pnorm(c+x)-pnorm(c-x));
    while(aire<alpha){
      x <- x+pas;
      aire <- abs(pnorm(c+x)-pnorm(c-x));
    }
    x <- x-pas;
    pas <- pas/10;
    cont <- (abs(aire-alpha))>=epsi;
  }
  k1 <- m1+sigma*min(c+x,c-x)/sqrt(n);
  k2 <- m1+sigma*max(c+x,c-x)/sqrt(n);
  print(k1); print(k2) }
```

*(Il aurait été plus efficace de procéder par dichotomie.)*

# La méthode de Tassi

## III – RÉPONSE À UN PROBLÈME DE TEST $\theta \leq \theta_1$ OU $\theta \geq \theta_2$ CONTRE $\theta_1 < \theta < \theta_2$

### 1. Détermination du test UPP

On se place toujours dans le cas d'une famille  $P_\theta$  à paramètre réel et à RVM en la statistique  $T$ .

**Théorème 3 :** Soit le problème de test :

$$(P.3) \quad \begin{cases} (H_0) \theta \leq \theta_1 \text{ ou } \theta \geq \theta_2 \\ (H_1) \theta_1 < \theta < \theta_2 \end{cases}$$

Il existe un test UPP au niveau  $\alpha$  déterminé par :

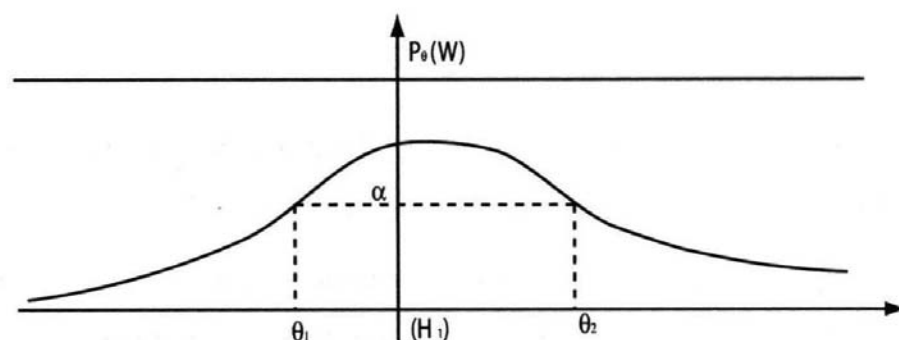
$$\phi = 1 \text{ si } x \in W = \{x/k_1 < T(x) < k_2\}$$

$$\phi = 0 \text{ si } T(x) < k_1 \text{ ou } T(x) > k_2$$

$$\phi = \gamma_1 \text{ ou } \gamma_2 \text{ si } T(x) = k_1 \text{ ou } T(x) = k_2$$

Les constantes  $k_1$  et  $k_2$  sont déterminées par  $P_{\theta_1}(W) = \alpha$  et  $P_{\theta_2}(W) = \alpha$ .

La fonction  $P_\theta(W), \theta \in \Theta$ , passe donc par un maximum  $\alpha$  sur la frontière de  $H_0$ , en  $\theta_1$  et  $\theta_2$ .



### 2. Applications

Les difficultés pratiques de résolution d'un tel test viendront fréquemment de la détermination des constantes  $k_1$  et  $k_2$ , ainsi que tendent à le montrer les deux exemples suivants :

#### a) Test de l'espérance d'une loi normale

Soit  $X \sim N(m, 1)$  dont on a un  $n$  - échantillon indépendant. On teste au

$$\begin{cases} (H_0) m \leq m_1 \text{ ou } m \geq m_2 \\ (H_1) m_1 < m < m_2 \end{cases}$$

La région critique est de la forme :

$$W = \{x \in \mathcal{H}^n / k_1 < \bar{X} < k_2\}$$

Les constantes  $k_1$  et  $k_2$  sont solutions du système :

$$P_{m_1}(W) = \alpha$$

$$P_{m_2}(W) = \alpha$$

Après avoir centré et réduit, on a :

$$P(\sqrt{n}(k_1 - m_1) < U < \sqrt{n}(k_2 - m_1)) = \alpha$$

$$P(\sqrt{n}(k_1 - m_2) < U < \sqrt{n}(k_2 - m_2)) = \alpha$$

où  $U$  suit la loi  $N(0, 1)$ .

Posons  $\beta_1 = \sqrt{n}(k_1 - m_1)$ ,  $\beta_2 = \sqrt{n}(k_2 - m_1)$

$$\sqrt{n}(k_1 - m_2) = \beta_1 + \sqrt{n}(m_1 - m_2)$$

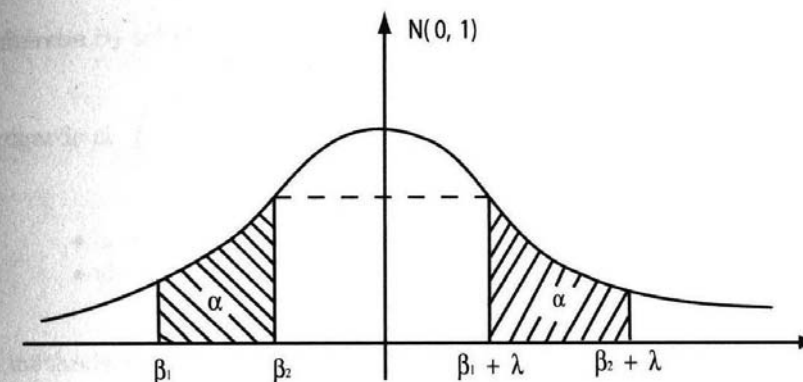
$$\sqrt{n}(k_2 - m_2) = \beta_2 + \sqrt{n}(m_1 - m_2)$$

Posons  $\lambda = \sqrt{n}(m_1 - m_2)$ ;  $n, m_1, m_2$  et donc  $\lambda$  sont connus.

On cherche à résoudre en  $\beta_1$  et  $\beta_2$  le système :

$$P(\beta_1 < U < \beta_2) = \alpha$$

$$P(\beta_1 + \lambda < U < \beta_2 + \lambda) = \alpha$$



La symétrie de la loi normale entraîne que l'intervalle  $[\beta_1 + \lambda, \beta_2 + \lambda]$  doit être le symétrique de  $[\beta_1, \beta_2]$  par rapport à l'axe des ordonnées, puisque les masses de probabilité réparties sur ces deux intervalles sont les mêmes.

On en déduit :

# Tassi : fin du calcul

$$\begin{aligned}\beta_1 + \lambda &= -\beta_2 \\ \beta_2 + \lambda &= -\beta_1\end{aligned}$$

D'où la relation :

$$\begin{cases} \sqrt{n}(k_1 - m_1) = \beta_1 \\ \sqrt{n}(k_1 - m_2) = -\beta_2 \end{cases} \quad \begin{cases} \sqrt{n}(k_2 - m_1) = \beta_2 \\ \sqrt{n}(k_2 - m_2) = -\beta_1 \end{cases}$$

$$k_1 = m_1 + \frac{\beta_1}{\sqrt{n}} = m_2 - \frac{\beta_2}{\sqrt{n}}$$

$$\sqrt{n} \frac{(m_2 - m_1)}{2} = \frac{\beta_1 + \beta_2}{2} = -\frac{\lambda}{2} \text{ connu}$$

Le centre de l'intervalle  $(\beta_1, \beta_2)$  est donc fixe : la résolution se fait alors pas à pas, en recherchant les valeurs  $\beta_1$  et  $\beta_2$  de moyenne connue telles que :

$$\int_{\beta_1}^{\beta_2} dN(0, 1) = \int_{\beta_1 + \lambda}^{\beta_2 + \lambda} dN(0, 1) = \alpha$$

# Test classique « d'égalité » de deux espérances, écarts-type inconnus mais égaux

Les variables aléatoires  $X$  et  $Y$  suivent les lois  $\mathcal{N}(\mu_1, \sigma^2)$  et  $\mathcal{N}(\mu_2, \sigma^2)$ .  
On se pose la question de l'égalité de  $\mu_1$  et  $\mu_2$ .

- On note  $H_0$  : «  $\mu_1 = \mu_2$  » (*hypothèse simple*) et
- $H_1$  : «  $\mu_1 \neq \mu_2$  » (*hypothèse composite*).

(On voit ici la **dissymétrie** entre  $H_0$  et  $H_1$ . On sait que, cependant,  $H_0$  et  $H_1$  peuvent être *toutes les deux composites*.)

Comme statistique de test, on emploie  $T = \frac{\bar{X} - \bar{Y}}{S \cdot \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}}$  où les deux échantillons indépendants sont représentés par  $X_1, X_2, X_3, \dots, X_{n_X}$  et  $Y_1, Y_2, Y_3, \dots, Y_{n_Y}$

# Notations, loi et règle de décision

$$\dots \text{ en notant } \bar{X} = \frac{\sum_{i=1}^{n_X} X_i}{n_X}, \bar{Y} = \frac{\sum_{i=1}^{n_Y} Y_i}{n_Y}, S_X = \sqrt{\frac{\sum_{i=1}^{n_X} (X_i - \bar{X})^2}{n_X - 1}}, S_Y = \sqrt{\frac{\sum_{i=1}^{n_Y} (Y_i - \bar{Y})^2}{n_Y - 1}} \text{ et}$$

$$S = \sqrt{\frac{(n_X - 1) \cdot S_X^2 + (n_Y - 1) \cdot S_Y^2}{n_X + n_Y - 2}} = \sqrt{\frac{\sum_{i=1}^{n_X} (X_i - \bar{X})^2 + \sum_{i=1}^{n_Y} (Y_i - \bar{Y})^2}{n_X + n_Y - 2}}.$$

On montre que la loi de cette statistique dans le cadre de  $H_0$  est la loi de Student à  $n_X + n_Y - 2$  degrés de liberté.

Pour la règle de décision à prendre, le choix naturel est, par construction de  $T$ , de rejeter  $H_0$  au profit de  $H_1$  («  $\mu_1 \neq \mu_2$  ») lorsque la réalisation de  $T$  obtenue est trop grande en valeur absolue, c'est-à-dire lorsque cette valeur absolue appartient à la **zone de rejet** qu'on détermine comme suit :

$$\alpha = \mathbb{P}_{H_0}(\text{Rejet de } H_0) = \mathbb{P}_{H_0}(|T| > c)$$

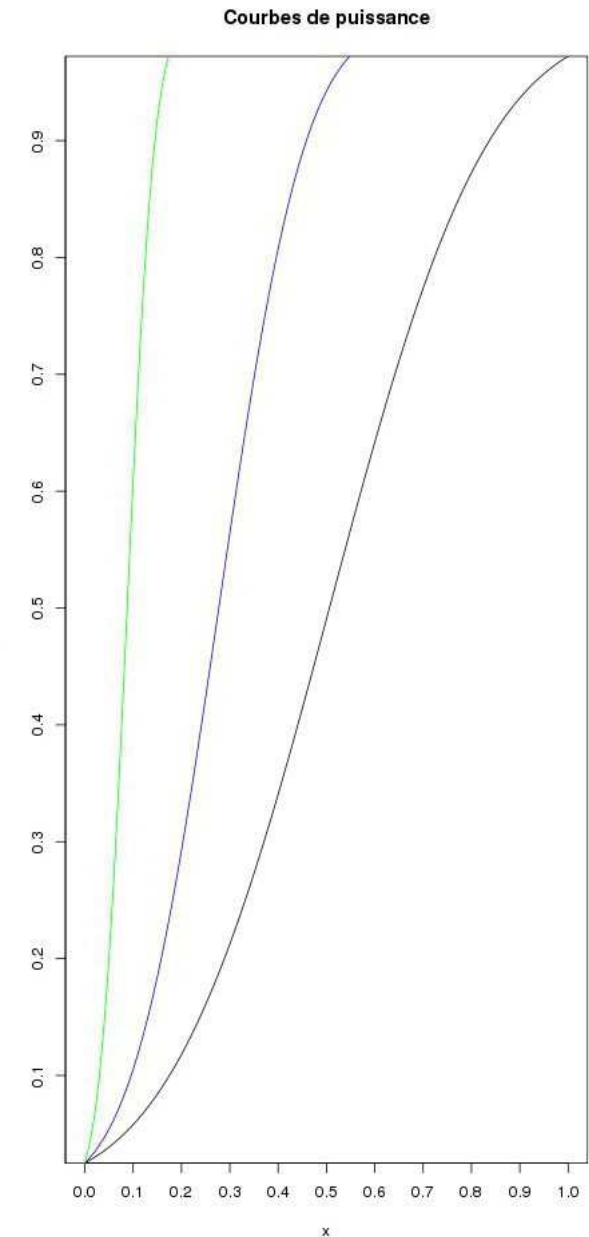
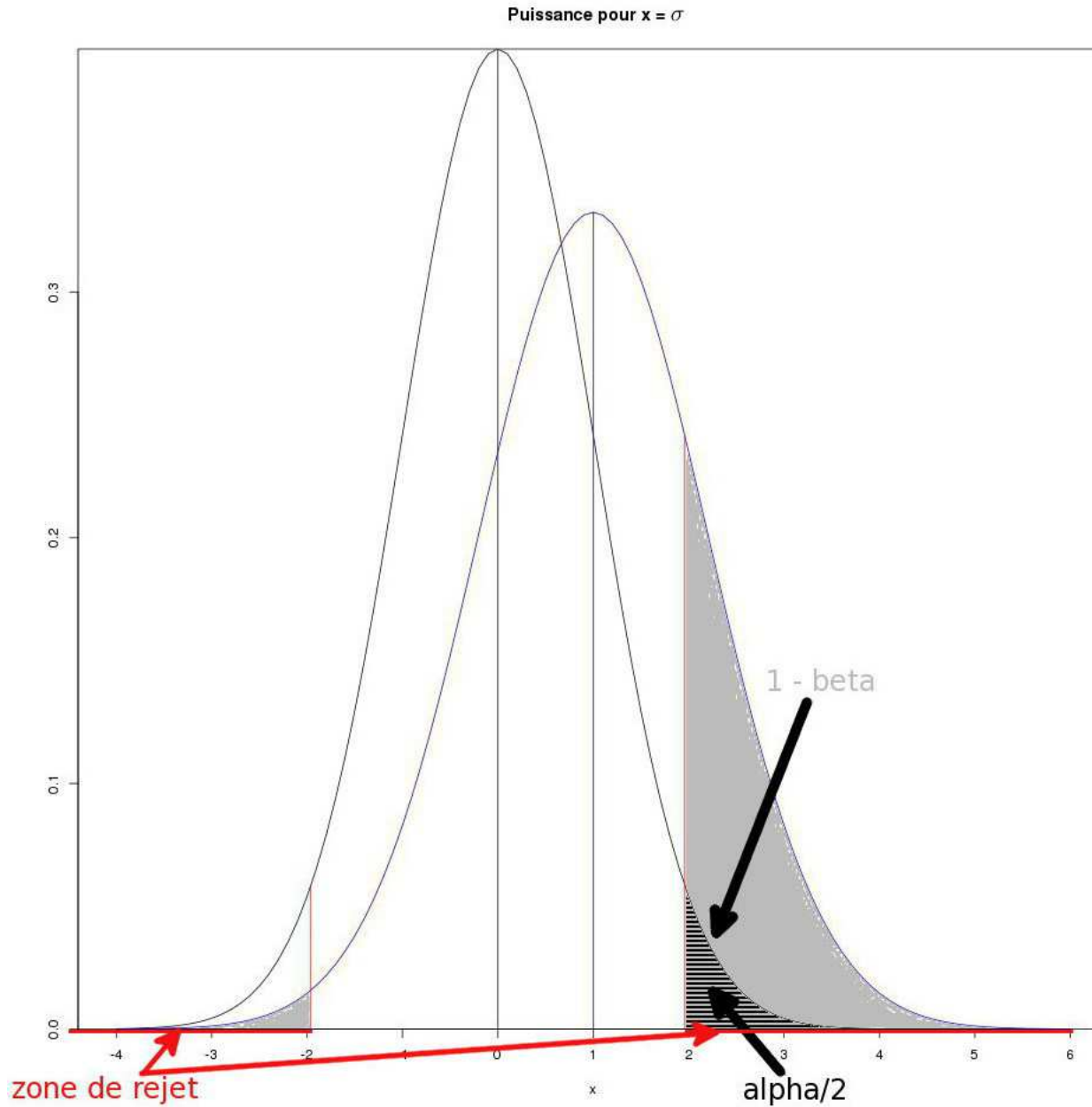
ce qui, pour une zone de rejet prise symétrique par rapport à 0, équivaut à  $c = t_{1-\frac{\alpha}{2}}$  où  $t_p$  est le fractile d'ordre  $p$  de la loi de Student à  $n_X + n_Y - 2$  degrés de liberté, défini par  $\mathbb{P}(T \leq c) = p$ .

# Remarques

- Remarque 1 : une autre écriture de la zone de rejet est  $] -\infty; -c[ \cup ]c; +\infty[.$
- Remarque 2 : ce test est considéré comme robuste, c'est-à-dire que le résultat obtenu lorsque les variables aléatoires ne suivent pas exactement une loi normale ou que leurs variances ne sont pas franchement égales reste relativement utilisable.
- Remarque 3 : si les effectifs des échantillons sont suffisamment grands, on pourra utiliser comme loi de  $T$  dans le cadre de  $H_0$  la loi  $\mathcal{N}(0,1)$ .



# La puissance du test



# La puissance du test

## Calcul approximatif

Pour déterminer la zone de rejet, utilisons  $\mathcal{N}(0,1)$  comme loi approximative pour  $T$  puis disons que, en fait,

$\mu_x = \mu_y + e$  avec  $e > 0$  et que  $S$  n'est pas aléatoire et a  $\sigma$  comme valeur.

Alors,  $1 - \beta = \mathbb{P}_e(T > z_{1-\frac{\alpha}{2}})$  et

$$\begin{aligned}\beta &\cong \mathbb{P}_e(T \leq z_{1-\frac{\alpha}{2}}) \cong \mathbb{P}_e\left(\bar{X} - \bar{Y} \leq z_{1-\frac{\alpha}{2}} \cdot \sigma \cdot \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}\right) \cong \\ &\mathbb{P}_e\left(\bar{X} - \bar{Y} - e \leq z_{1-\frac{\alpha}{2}} \cdot \sigma \cdot \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}} - e\right) \cong \\ &\mathbb{P}_e\left(\frac{\bar{X} - \bar{Y} - e}{\sigma \cdot \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \leq z_{1-\frac{\alpha}{2}} - \frac{e}{\sigma \cdot \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}}\right).\end{aligned}$$

Or, comme, ici, l'écart entre les espérances inconnues est égal à  $e$ ,

$\frac{\bar{X} - \bar{Y} - e}{\sigma \cdot \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}}$  suit approximativement  $\mathcal{N}(0,1)$ . Donc  $z_\beta \cong z_{1-\frac{\alpha}{2}} - \frac{e}{\sigma \cdot \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}}$

et la puissance est  $1 - \beta \cong 1 - \phi\left(z_{1-\frac{\alpha}{2}} - \frac{e}{\sigma \cdot \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}}\right)$  où  $z_p$  est le fractile d'ordre  $p$  de  $\mathcal{N}(0,1)$  et où  $\phi$  est sa fonction de répartition.

# La taille nécessaire pour les échantillons

## La formule

- Dans le cas où  $n_x = n_y = n$ , on a  $z_\beta \cong z_{1-\frac{\alpha}{2}} - \frac{e}{\sigma \cdot \sqrt{\frac{2}{n}}}$ , soit
- $-z_{1-\beta} \cong z_{1-\frac{\alpha}{2}} - \frac{e \cdot \sqrt{n}}{\sigma \cdot \sqrt{2}}$ , c'est-à-dire  $n \cong \frac{2 \cdot \sigma^2 (z_{1-\frac{\alpha}{2}} + z_{1-\beta})^2}{e^2}$ .
- Cette formule permet, après avoir *intuité*  $\sigma$ , de déterminer l'effectif  $n$  des échantillons avec lequel la puissance  $1 - \beta$  est approximativement obtenue pour un choix de l'écart  $e$  entre les espérances inconnues.

# Test de non-infériorité pour deux espérances

De même ; exemple : on compare un antihypertenseur expérimental, A, avec un antihypertenseur de référence, B.

- Le critère de jugement est la variation de tension artérielle avant/après traitement (mmHg).
- On considère que ce caractère suit une loi normale, de même variance pour les deux traitements. La statistique de test est

$$T = \frac{\overline{X}_B - \overline{X}_A - \delta}{S \cdot \sqrt{\frac{1}{n_B} + \frac{1}{n_A}}} \text{ avec } S = \sqrt{\frac{(n_B - 1) \cdot S_B^2 + (n_A - 1) \cdot S_A^2}{n_B + n_A - 2}}.$$

- La borne d'équivalence  $\delta$  choisie est 2 mmHg.

Données :

$n_A$	$\overline{X}_A$	$S_A$	$n_B$	$\overline{X}_B$	$S_B$
140	13,1	7,8	138	12,0	8,0

$$T \text{ a donc valu } t \cong \frac{-12 + 13,1 - 2}{\sqrt{\frac{137 \cdot 8^2 + 139 \cdot 7,8^2}{276}} \cdot \sqrt{\frac{1}{138} + \frac{1}{140}}} \cong -3,27.$$

- La zone de rejet au risque  $\alpha = 5\%$  étant environ  $]-\infty; -1,65[$ , on rejette  $H_0$  et on affirme que le nouvel antihypertenseur n'est pas inférieur au contrôle employé.

# Test de non-infériorité pour deux espérances

Autre exemple : L'hypothèse principale à tester était la suivante : essai d'équivalence unilatérale ou de non infériorité de l'harpagophyton par rapport à la diacerhéine sur la douleur spontanée mesurée à l'aide d'une Echelle Visuelle Analogique (100 mm). Ce test d'équivalence devait être mené en situation unilatérale avec un risque à 0,05.

En posant l'hypothèse que la différence vraie entre les traitements était nulle avec un  $\delta$  de 10 mm, un test d'équivalence unilatéral avec un risque  $\alpha$  à 0,05, un risque  $\beta$  à 0,10 et un écart type de 18 mm, nécessitait l'inclusion de 56 patients par groupe.

Tests d'équivalence basés sur les intervalles de confiance de la différence entre les groupes  
(douleur à l'EVA)

	Moyenne des différences par rapport à la valeur de base		Différence entre les groupes (H-D)	Intervalle de confiance à 90 % de la différence (H-D)	Hypothèse : Harpagophyton est au moins aussi bon que la diacerhéine
	Harpagophyton	Diacerhéine			
Douleur à j120	-30,6 $\pm$ 3,3	-25,5 $\pm$ 3,6	-5,1	-13,1 ; 3,0	Acceptée

## Suite de l'article

Vérifions les résultats indiqués ; «  $\pm 3,3$  » indique ou bien l'estimation de l'écart-type ou bien le produit de celui-ci par 1,96 ; disons que  $s_{harpa} \cong 3,3$  , etc ... La valeur -5,1 est évidemment bonne ; l'estimation d'écart-type à utiliser est  $\sqrt{3,3^2 + 3,6^2} \cong 4,88$  ; comme  $z_{0,90} \cong 1,6449$  , la demi-largeur d'un  $IC_{0,90}$  pour H-D est 8,03. La borne droite de l'IC de H-D est donc incluse dans  $]-10 ; +10[$ . OK, puisque les meilleures valeurs du critère d'efficacité sont les plus petites.

- Les anti-inflammatoires non stéroïdiens sont utilisés depuis longtemps pour lutter contre la douleur des patients arthrosiques.
- Leur efficacité n'est pas contestable, attestée par une multitude d'essais thérapeutiques. Ces traitements symptomatiques d'action rapide ont démontré leur utilité dans la poussée douloureuse d'arthrose, mais n'ont pas prouvé leur action, bénéfique, délétère ou neutre au long cours sur le cartilage.
- Plusieurs médicaments, les anti-arthrosiques d'effet différé et prolongé, ont récemment montré leur efficacité symptomatique dans l'arthrose sans démontrer pour le moment un effet structural.

# Suite de l'article

- Ils sont utiles comme traitement de fond des patients arthrosiques souffrant de façon régulière et chez lesquels les mesures non médicamenteuses ne suffisent plus à maîtriser les symptômes. Leur intérêt principal est de réduire la consommation d'antalgiques ou d'anti-inflammatoires ...
- Il n'a pas été inclus de groupe placebo dans cette étude car il est souvent difficile de recruter des patients lorsqu'ils sont informés du risque (égal à 50 %) d'être dans un groupe placebo, et ce d'autant plus que la durée de traitement est particulièrement longue (quatre mois).
- Par ailleurs, l'appréciation de l'intensité de l'effet placebo dans l'arthrose a récemment été étudiée.
  - ▶ Une revue portant sur 457 patients traités par placebo dans le cadre de six essais contrôlés d'anti-arthrosiques symptomatiques d'action lente a montré que sous placebo, la baisse moyenne de la douleur à l'EVA était de l'ordre de 10 à 16 mm.
  - ▶ Par conséquent, la baisse observée dans le groupe harpagophyton (30,6 mm) est nettement supérieure à celle habituellement observée sous placebo.

# Année post-bac : comment les tests d'hypothèses sont-ils présentés ?

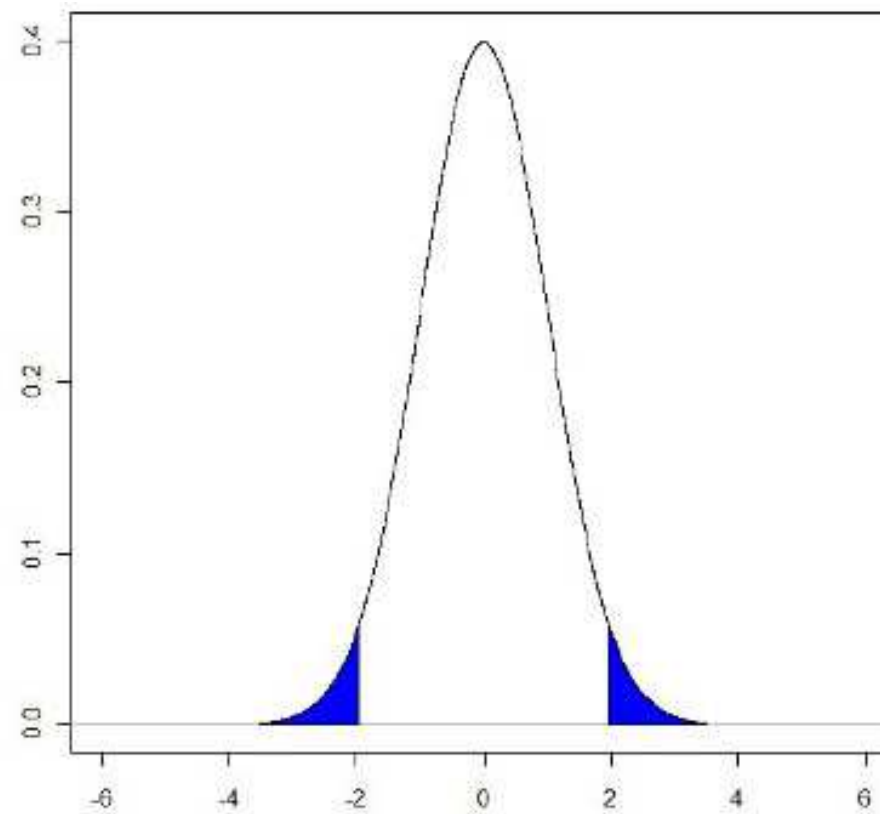
Le programme actuel de Terminale ne permet pas d'aborder ce qui suit,

car il lui manque (au moins) les notions simples suivantes :

- indépendance de deux variables aléatoires réelles
- variance de la somme de deux variables aléatoires réelles
- variance du quotient d'une variable aléatoire réelle par une constante (non aléatoire).



# La méthode



- On souhaite étudier une caractéristique numérique  $\theta \in \Theta$  d'une population.
- On définit une hypothèse concernant  $\theta$ , appelée hypothèse nulle et notée  $H_0$ ,
- ainsi qu'une hypothèse contradictoire avec  $H_0$ , appelée hypothèse alternative et notée  $H_1$ .
- Souvent,  $H_0$  est «  $\theta = \theta_0$  » et  $H_1$  est «  $\theta \neq \theta_0$  ».
- Le but du test est de déterminer si les données  $x_1, x_2, x_3, \dots, x_n$ , recueillies à partir d'un échantillon dont on dispose, permettent ou non d'affirmer **raisonnablement** que  $H_0$  est fausse et, en conséquence, que  $H_1$  est vraie.

# Les deux erreurs

Pour cela, on utilise une statistique  $T = f(X_1, X_2, X_3, \dots, X_n)$

- dont la valeur apporte, si possible, des renseignements ***optimaux*** sur  $H_0$
- et dont la loi de probabilité ***dans le cadre de  $H_0$***  est connue (?), au moins asymptotiquement.

La statistique  $T$  nous permet d'instaurer une règle de décision statistique, qui peut conduire à deux types d'erreurs :

- l'erreur de première espèce : **rejeter  $H_0$  alors qu'elle est vraie ;**
- l'erreur de deuxième espèce : **ne pas rejeter  $H_0$  alors qu'elle est fausse.**

On appelle risque de première espèce d'un test et on note  $\alpha$  la probabilité, ***dans le cadre de  $H_0$*** , que la règle de décision de ce test conduise à rejeter  $H_0$ . Comme la loi de  $T$  est connue dans ce cadre, il est possible (?) de choisir  $\alpha$ , souvent avec  $\alpha = 0,05$ .

Si on sait seulement que le risque de première espèce **maximal** d'un test est égal au nombre  $\alpha$ , on dit que ce test est de ***niveau  $\alpha$*** .

# La fonction risque de deuxième espèce

- Le risque de deuxième espèce, noté  $\beta(\theta_1)$ , où  $\theta_1 \in \Theta$  a une valeur contradictoire avec  $H_0$ , est la probabilité, dans le cadre de  $\theta = \theta_1$ , de ne pas rejeter  $H_0$ .
- On voit que le **risque  $\beta$**  (risque de deuxième espèce) est une **fonction** et n'est pas aussi simple à maîtriser que le risque  $\alpha$ .
- On appelle puissance du test la fonction  $\theta_1 \mapsto 1 - \beta(\theta_1)$ . Sa valeur intéressante dépend, évidemment, de la valeur du *paramètre* inconnu  $\theta$ .