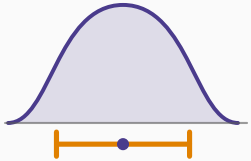
Statistique inférentielle

Chapitre 14 : Les deux risques et la puissance



Jaouad Madkour

Faculté d'Économie et de Gestion de Tanger · 27 août 2026

Plan du chapitre

Deux façons de se tromper

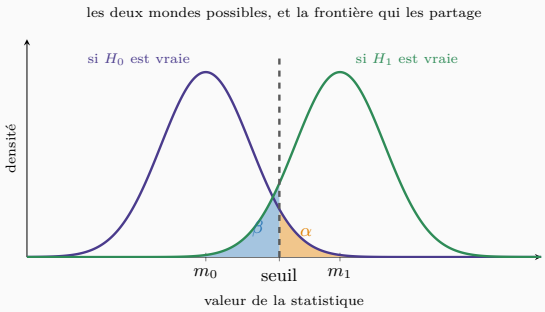
La puissance

Les mésusages

Ce qu'il faut retenir

Deux façons de se tromper

Les deux mondes possibles



Les deux risques ne sont pas symétriques

α — **erreur de première espèce**
rejeter H_0 alors qu'elle est vraie.
On la fixe : 5 %, 1 %.

β — **erreur de seconde espèce**
ne pas rejeter H_0 alors qu'elle est fausse.
On la subit : elle dépend de m_1, qu'on ignore.

Déplacer le seuil ne supprime rien
vers la droite : α baisse, β monte.
vers la gauche : l'inverse.

Le seul moyen de réduire les deux à la fois est d'**augmenter** n : les deux cloches se resserrent et se séparent.

Les deux erreurs

$$\alpha = P(\text{rejeter } H_0 \mid H_0 \text{ vraie}) \qquad \beta = P(\text{ne pas rejeter } H_0 \mid H_0 \text{ fausse})$$

α est l'erreur de **première espèce**, β celle de **seconde espèce**.

La dissymétrie fondamentale

α **se fixe** : c'est une décision, prise avant de voir les données. On choisit $5\,\%$, $1\,\%$.

β **se subit** : il dépend de la vraie valeur du paramètre, que l'on ignore. On ne peut que le calculer *pour un écart donné*.

Le seuil ne fait que déplacer le problème

Un arbitrage, pas une optimisation

Déplacer le seuil	Effet sur α	Effet sur β
vers la droite	diminue	augmente
vers la gauche	augmente	diminue

Le seul moyen de gagner sur les deux

Augmenter n. Les deux distributions se resserrent, se chevauchent moins, et les deux risques baissent ensemble.

C'est la seule action qui améliore vraiment un test — et elle coûte de l'argent. Tout le reste n'est qu'un déplacement de la frontière.

La puissance

Puissance d'un test

$$\text{puissance} = 1 - \beta = P(\text{rejeter } H_0 \mid H_0 \text{ fausse})$$

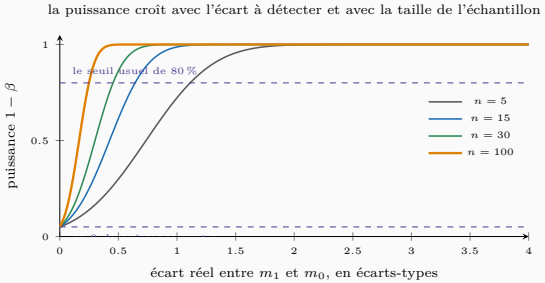
C'est la capacité du test à **détecter un écart qui existe**.

La question qu'elle permet enfin de poser

« Mon test n'a rien trouvé » est une information inutilisable tant qu'on ne sait pas ce qu'il était capable de trouver.

La puissance transforme cette phrase en énoncé exploitable : *mon échantillon détectait un écart de trois cents dirhams dans huit cas sur dix ; un écart de cinquante lui échappait presque toujours.*

Ce dont elle dépend



Trois lectures

Un petit écart est difficile à détecter. À 0,5 écart-type, il faut cent observations pour avoir une chance raisonnable de le voir.

En zéro, la puissance vaut α. Quand H_0 est vraie, on la rejette dans exactement 5 % des cas : c'est la définition du seuil.

La conséquence pratique
Un test non significatif sur un petit échantillon ne dit rien. Il faut calculer la puissance *avant* de collecter — c'est ainsi que se fixe la taille d'échantillon.

Ce qui fait monter la puissance

Levier	Effet
écart réel plus grand	monte — un gros effet se voit
n plus grand	monte — la seule action du gestionnaire
σ plus petit	monte — moins de bruit
α plus grand	monte — mais on rejette plus à tort

Deux lectures immédiates de la courbe

En zéro, la puissance vaut exactement α. Quand H_0 est vraie, on la rejette dans $5\,\%$ des cas : c'est la définition même du seuil.

À un demi écart-type, il faut environ cent observations pour atteindre $80\,\%$. **Les petits effets exigent de gros échantillons**, et aucune astuce ne contourne cela.

Le seuil de quatre-vingts pour cent

Une convention de planification

On dimensionne usuellement une étude pour atteindre une puissance de $80\,\%$ sur l'écart que l'on juge économiquement significatif.

Cela signifie accepter de manquer un vrai effet une fois sur cinq.

Le calcul se fait avant, jamais après

Une étude sous-dimensionnée est doublement perdante : si elle ne trouve rien, on ne peut rien conclure ; si elle trouve quelque chose, l'effet estimé est probablement exagéré.

Une étude sans puissance suffisante n'est pas une étude prudente : c'est une étude gaspillée.

Les mésusages

Ce que « non significatif » ne veut pas dire

Trois conclusions abusives

1. « Les deux groupes sont identiques. » — Non : on n'a pas montré de différence.
2. « L'effet est nul. » — Non : il est peut-être réel et trop petit pour ce dispositif.
3. « L'hypothèse est validée. » — Non : elle n'est pas contredite, ce qui est très différent.

Notre comparaison hommes-femmes

L'écart observé sur la population entière vaut $0{,}28$ dirham, avec $t \approx 0$.

On ne rejette pas H_0. Mais la conclusion correcte n'est pas « les salaires sont égaux » : c'est « **aucun écart détectable** avec ces données ». Avec cet effectif, un écart réel de trois cents dirhams aurait été vu ; un écart de cinquante ne l'aurait pas été.

Significatif ne veut pas dire important

Le piège inverse, et il est aussi fréquent

Sur un million d'observations, un écart de deux dirhams devient statistiquement significatif.

Il n'a aucune portée économique. **La significativité statistique mesure la fiabilité de la détection, pas l'ampleur de l'effet.**

La bonne pratique, en deux temps

Toujours accompagner un test d'une **estimation de l'effet** et de son **intervalle de confiance**.

Le test dit *si l'on peut affirmer quelque chose* ; l'intervalle dit *quoi*. Publier l'un sans l'autre est une information tronquée.

Ce qu’il faut retenir

Le résumé du chapitre

Six points

1. α : rejeter H_0 à tort. β : ne pas rejeter alors qu'elle est fausse.
2. α **se fixe**, β **se subit** : la dissymétrie est fondamentale.
3. Déplacer le seuil échange un risque contre l'autre ; seul **augmenter** n réduit les deux.
4. **Puissance** $= 1 - \beta$: la capacité à détecter un écart réel. En zéro, elle vaut α.
5. Convention de planification : $80\,\%$ de puissance sur l'écart économiquement significatif.
6. « Non significatif » ne prouve pas l'égalité ; « significatif » ne prouve pas l'importance.

La suite

Reste l'outil que tous les logiciels affichent et que presque personne ne définit correctement : la p-value.

Le chapitre 15 la définit, la calcule, et détaille les trois contresens qu'elle provoque.