

Power Calculation | Calcul de Puissance

Vin Arceneaux and/et Nahomi Ichino

21 May/mai 2025

Key points for this lecture | *Points clés du cours*

- ▶ Informally: Power is the ability of our experiment to detect treatment effects, if they in fact exist
- ▶ Power matters: for practical reasons and for interpretation
- ▶ You can increase power by strengthening the intervention, reducing noise, and increasing sample size
- ▶ De manière informelle: La puissance est la capacité de notre expérience à détecter des effets de traitement, s'ils existent réellement
- ▶ La puissance est importante : pour des raisons pratiques et pour l'interprétation
- ▶ Vous pouvez augmenter la puissance en renforçant l'intervention, en réduisant le bruit et en augmentant la taille de l'échantillon

Power | *La puissance*

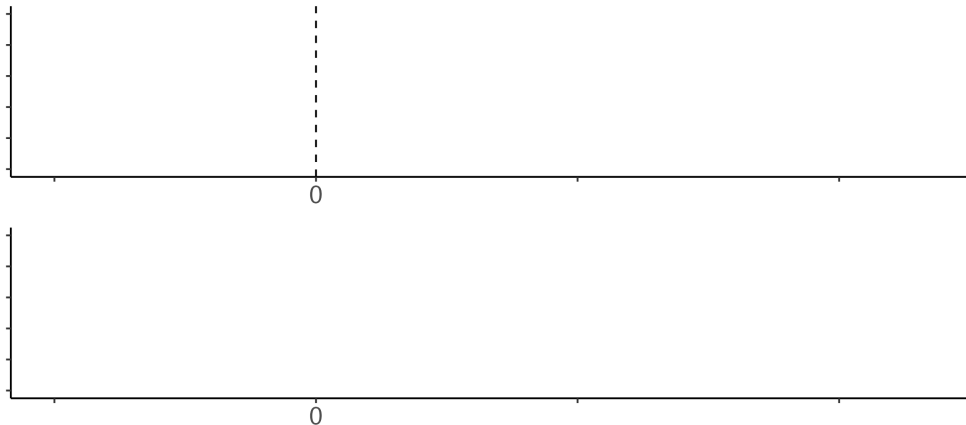
- ▶ Main idea: The ability of the experiment to detect statistically significant treatment effects, if the effect really exists.
 - ▶ More specifically: The probability of correctly rejecting a null hypothesis when the alternative is true. (The probability of avoiding a Type II error.)
 - ▶ Power is a probability between 0 and 1. We generally prefer more power.
- ▶ L'idée principale : La capacité de notre expérience à détecter des effets de traitement statistiquement significatifs, s'ils existent réellement.
 - ▶ Plus précisément: La probabilité de rejeter correctement une hypothèse nulle lorsque l'alternative est vraie. (La probabilité de l'expérience à éviter de commettre une erreur de type II.)
 - ▶ La puissance est une probabilité, un nombre compris entre 0 et 1. Nous préférons généralement plus de puissance.

Power Analysis | *L'analyse de la puissance*

- ▶ We do a power analysis *before* we run a study. It is part of the *design* stage.
- ▶ Goal: To discover whether our planned design has enough power to detect effects if they exist. Then improve the design or drop the study.
- ▶ Nous faisons l'analyse de puissance *avant* de mener une étude. Cela fait partie de la phase de *conception*.
- ▶ Objectif : Déterminer si la conception de l'étude a suffisamment de puissance pour détecter les effets s'ils existent. Puis améliorer ou abandonner la conception.

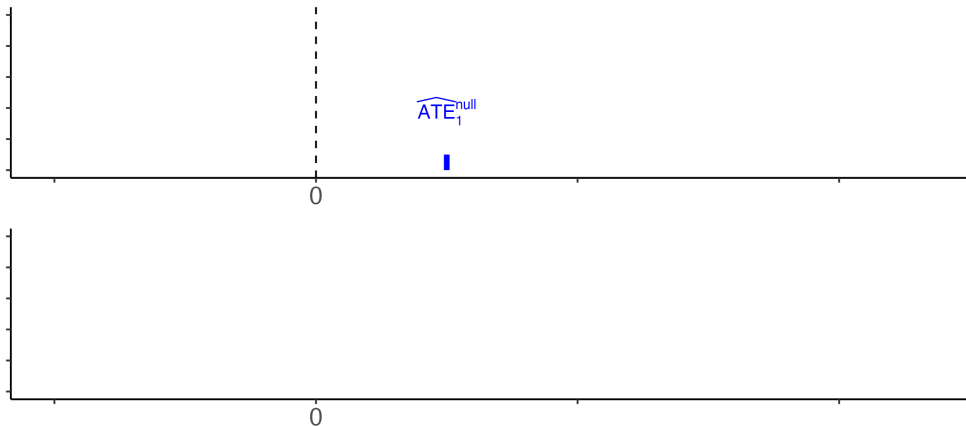
Review: Hypothesis testing | *Rappel : les tests d'hypothèse*

Distribution of \widehat{ATE}^{null} when true ATE = 0 and SD = 1



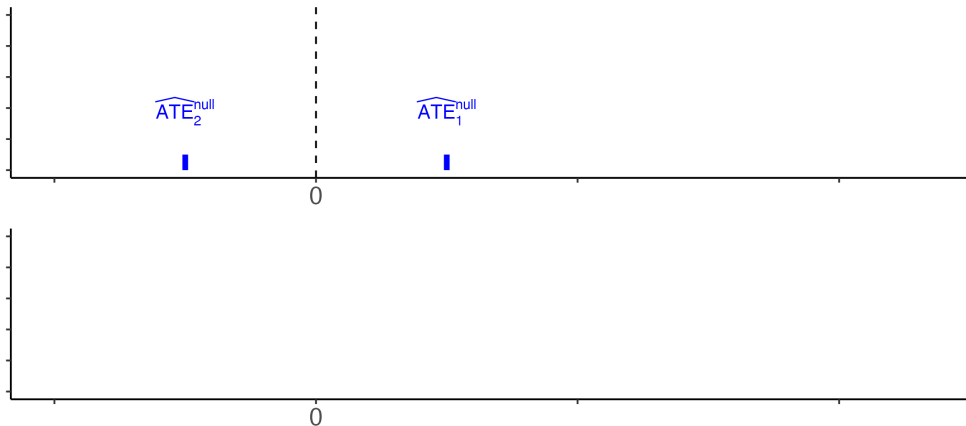
Review: Hypothesis testing | *Rappel : les tests d'hypothèse*

Distribution of \widehat{ATE}^{null} when true ATE = 0 and SD = 1



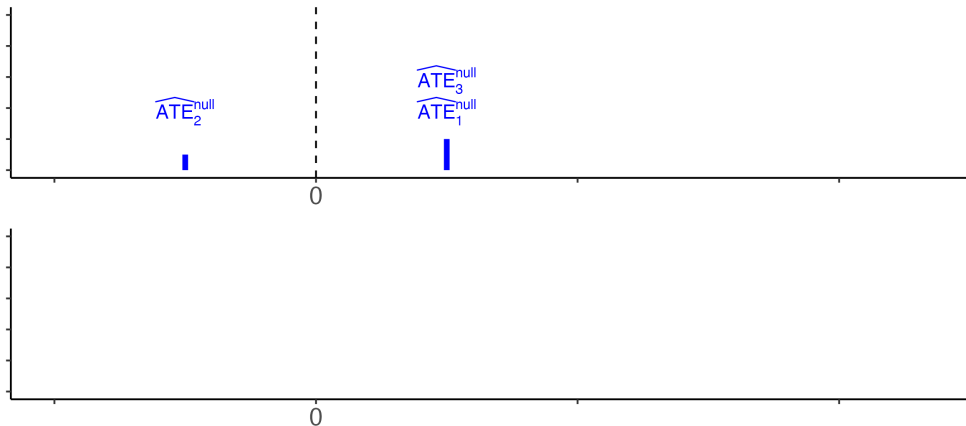
Review: Hypothesis testing | *Rappel : les tests d'hypothèse*

Distribution of \widehat{ATE}^{null} when true ATE = 0 and SD = 1



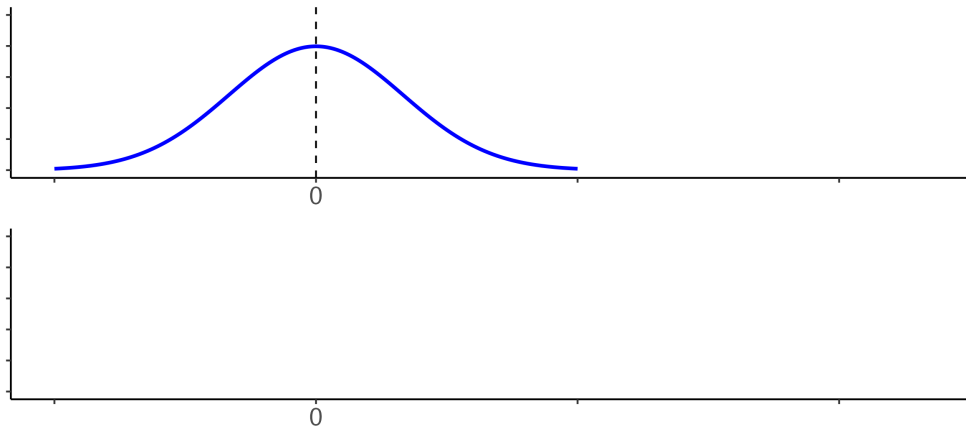
Review: Hypothesis testing | *Rappel : les tests d'hypothèse*

Distribution of \widehat{ATE}^{null} when true ATE = 0 and SD = 1



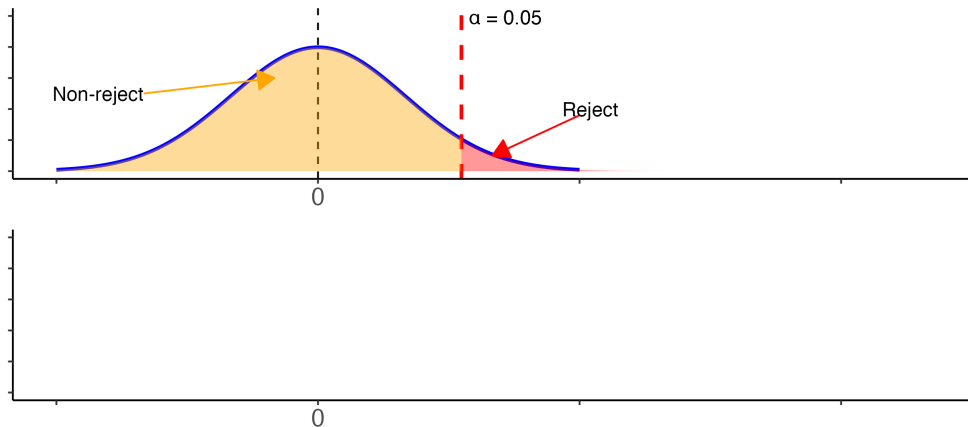
Review: Hypothesis testing | *Rappel : les tests d'hypothèse*

Distribution of \widehat{ATE}^{null} when true ATE = 0 and SD = 1



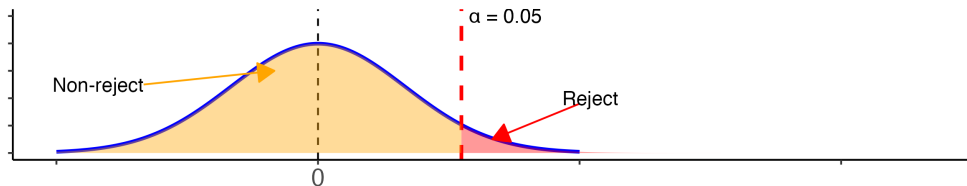
Review: Hypothesis testing | *Rappel : les tests d'hypothèse*

Distribution of \widehat{ATE}^{null} when true ATE = 0 and SD = 1

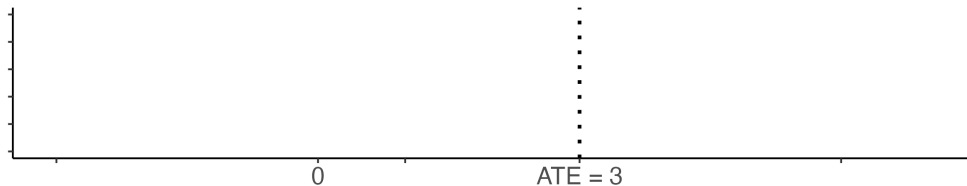


Power | *La puissance*

Distribution of \widehat{ATE}^{null} when true ATE = 0 and SD = 1

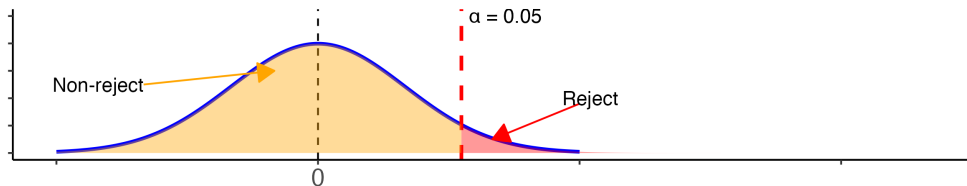


Distribution of \widehat{ATE} when true ATE = 3 and SD = 1

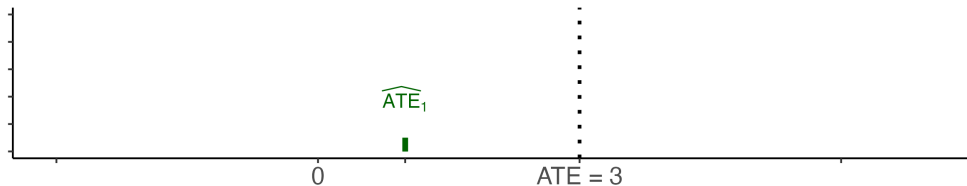


Power | *La puissance*

Distribution of \widehat{ATE}^{null} when true ATE = 0 and SD = 1

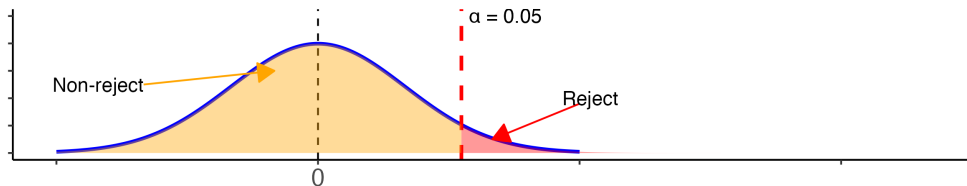


Distribution of \widehat{ATE} when true ATE = 3 and SD = 1

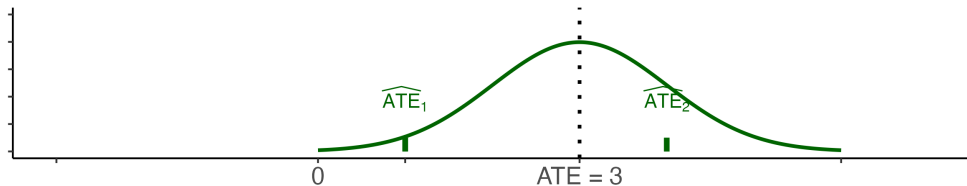


Power | *La puissance*

Distribution of \widehat{ATE}^{null} when true ATE = 0 and SD = 1

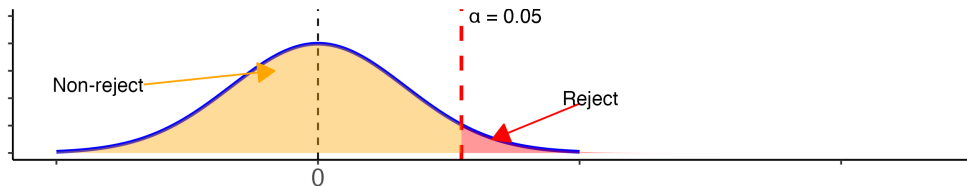


Distribution of \widehat{ATE} when true ATE = 3 and SD = 1

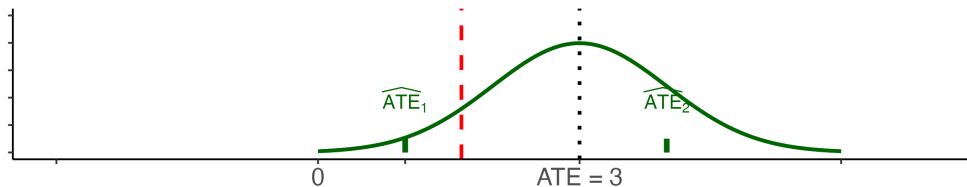


Power | *La puissance*

Distribution of \widehat{ATE}^{null} when true ATE = 0 and SD = 1

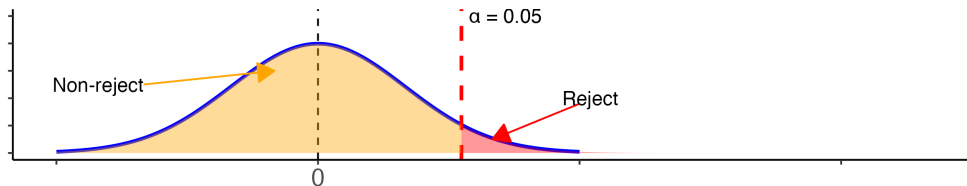


Distribution of \widehat{ATE} when true ATE = 3 and SD = 1

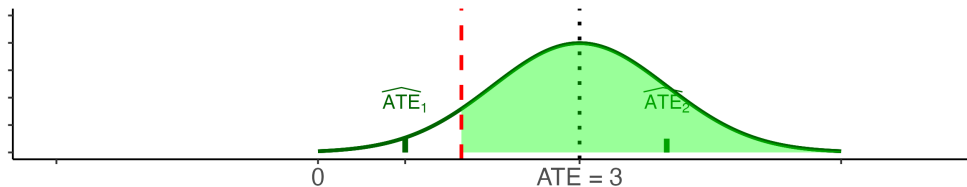


Power | *La puissance*

Distribution of \widehat{ATE}^{null} when true ATE = 0 and SD = 1

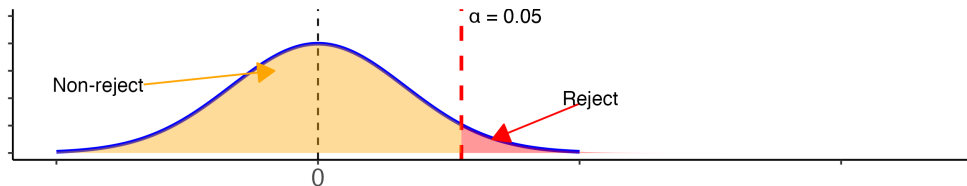


Distribution of \widehat{ATE} when true ATE = 3 and SD = 1

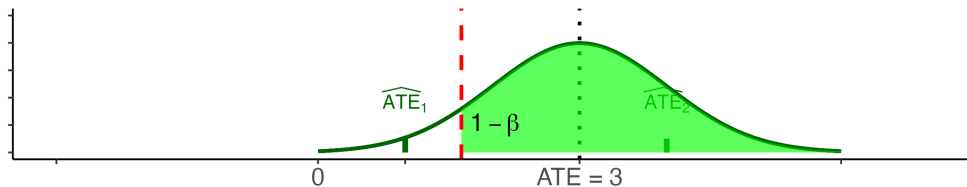


Power | *La puissance*

Distribution of \widehat{ATE}^{null} when true ATE = 0 and SD = 1

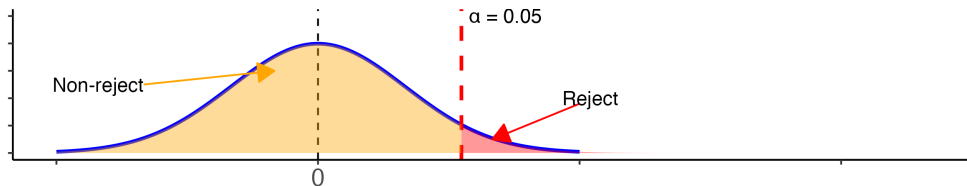


Distribution of \widehat{ATE} when true ATE = 3 and SD = 1

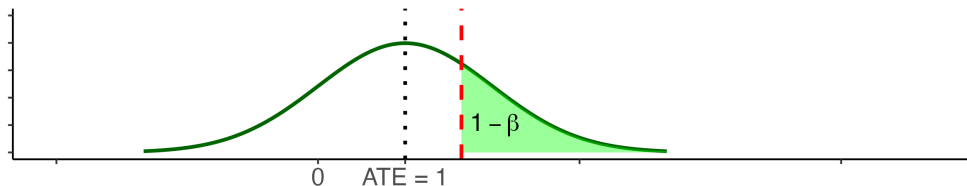


Power with a smaller ATE | *La puissance avec un ATE plus petit*

Distribution of \widehat{ATE}^{null} when true ATE = 0 and SD = 1

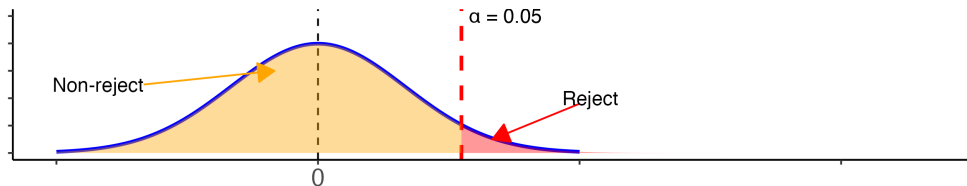


Distribution of \widehat{ATE} when true ATE = 1 and SD = 1

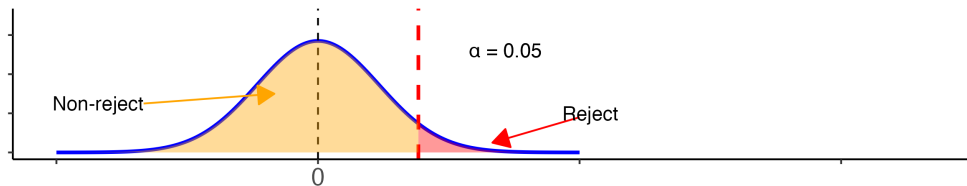


Power with less noise | *La puissance avec moins de bruit*

Distribution of \widehat{ATE}^{null} when true ATE = 0 and SD = 1

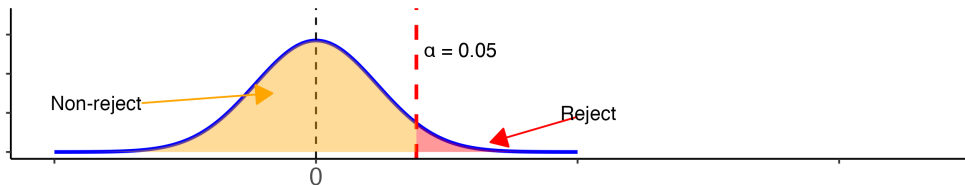


Distribution of \widehat{ATE}^{null} when true ATE = 0 and SD = 0.7

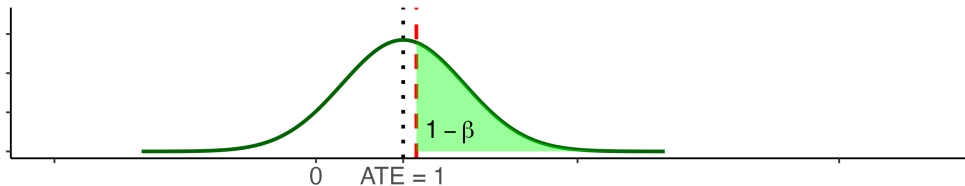


Power with less noise | *La puissance avec moins de bruit*

Distribution of \widehat{ATE}^{null} when true ATE = 0 and SD = 0.7

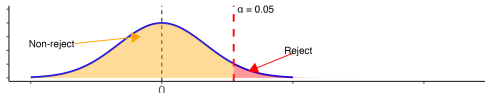


Distribution of \widehat{ATE} when true ATE = 1 and SD = 0.7

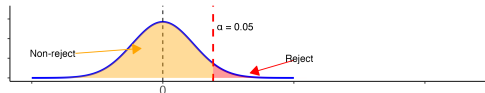


Power with less noise | *La puissance avec moins de bruit*

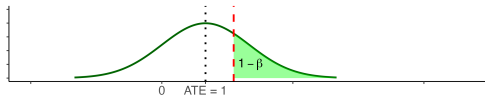
Distribution of \widehat{ATE}^{null} when true ATE = 0 and SD = 1



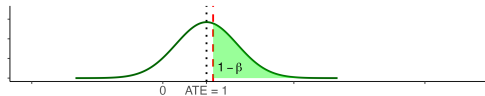
Distribution of \widehat{ATE}^{null} when true ATE = 0 and SD = 0.7



Distribution of \widehat{ATE} when true ATE = 1 and SD = 1



Distribution of \widehat{ATE} when true ATE = 1 and SD = 0.7



Under-powered studies | *Les études de faible puissance*

- ▶ With a study that does not have enough power, we can't determine a null finding (no statistically significant effect) means:
 - ▶ There is no effect, or
 - ▶ We just can't detect an effect that exists.
- ▶ Dans le cas d'une étude dont la puissance est insuffisante, il est impossible de dire si un résultat nul (pas d'effet statistiquement significatif) signifie:
 - ▶ Qu'il n'y a pas d'effet, ou
 - ▶ Si nous ne pouvons tout simplement pas détecter un effet existant.

Under-powered studies | *Les études de faible puissance*

- ▶ We don't know whether we should adopt this treatment because we don't know whether it's effective.
- ▶ Perhaps running this study is not a good use of our resources.
- ▶ Nous ne savons pas si nous devons adopter ce traitement car nous ne savons pas s'il est efficace.
- ▶ Peut-être que la réalisation de cette étude n'est pas une bonne utilisation de nos ressources.

Compare with hypothesis testing | *La comparaison avec les tests d'hypothèse*

Power Analysis	Hypothesis Test
1. Do before the experiment <i>Faire avant l'expérience</i>	1. Do after the experiment <i>Faire après l'expérience</i>
2. No data yet <i>Pas encore de données</i>	2. Use data from experiment <i>Utiliser les données de l'expérience</i>
3. Need assumptions about a world in which: <i>Besoin d'hypothèses sur un monde dans lequel:</i>	
a. there is no treatment effect <i>il n'y a pas d'effet de traitement</i>	a. there is no treatment effect <i>il n'y a pas d'effet de traitement</i>
b. there is a treatment effect <i>il existe un effet de traitement</i>	

What are the three main inputs into statistical power? | *Quels sont les trois principaux éléments de la puissance statistique ?*

- ▶ Treatment-effect size
 - ▶ Noisiness of the outcome variable (σ)
 - ▶ Sample size
- ▶ Taille de l'effet du traitement.
 - ▶ La dispersion de la variable de résultat (σ)
 - ▶ Taille de l'échantillon

The Power Formula | *The Power Formula (1)*

$$Power = \Phi\left(\frac{|\tau|\sqrt{N}}{\sqrt{2}\sigma} - \Phi^{-1}(1 - \alpha)\right) \quad (1)$$

- ▶ Standard 2-arm experiment
- ▶ 2 equal sized groups with known, equal variances
- ▶ One-sided alternative hypothesis
- ▶ Normal distribution of the outcome
- ▶ Expérience standard à 2 bras
- ▶ 2 groupes de taille égale avec des variances connues et égales
- ▶ Hypothèse alternative unilatérale
- ▶ Distribution normale du résultat

The Power Formula | *The Power Formula (2)*

- ▶ Φ is the cumulative density function of the normal distribution (FIXED)
 - ▶ τ is the effect size
 - ▶ N is the sample size
 - ▶ σ is the standard deviation of the outcome
 - ▶ α is the significance level (FIXED by convention)
- ▶ Φ est la fonction de densité cummulative de la distribution normale (FIXE)
 - ▶ τ est la taille de l'effet
 - ▶ N est la taille de l'échantillon
 - ▶ σ est l'écart-type du résultat
 - ▶ α est le niveau de signification (FIXE par convention)

Three main inputs into statistical power 1: Sample size | *Trois éléments principaux de la puissance statistique 1 : La taille de l'échantillon*

- ▶ More observations → more power
- ▶ Add observations!
- ▶ Problems?
- ▶ Plus d'observations → plus de puissance
- ▶ Ajoutez des observations !
- ▶ Problèmes éventuels ?

Three main inputs into statistical power 2: Noisiness of outcome measure | *Les trois principaux facteurs de la puissance statistique 2 : La dispersion de la variable dépendante*

- ▶ Less noise → more power
 - ▶ Reduce noise. How?
 - ▶ Blocking — conduct experiments among subjects that look more similar;
 - ▶ Collect baseline covariates — background information about experimental units;
 - ▶ Problems?
- ▶ Faible dispersion → plus de puissance
 - ▶ Comment réduire la dispersion ?
 - ▶ Stratifier - mener des expériences avec des sujets qui se ressemblent le plus.
 - ▶ Recueillir des covariables de base - informations de base sur les unités expérimentales;
 - ▶ Problèmes éventuels ?

Three Main Inputs into Statistical Power 3: Size of Treatment Effect | *Trois éléments principaux de la puissance statistique 3 : Taille de l'effet du traitement*

- ▶ Bigger effect → more power;
 - ▶ Boost dosage / avoid very weak treatments;
 - ▶ Problems?
- ▶ Plus d'effet → plus de puissance;
 - ▶ Augmenter le dosage / éviter les traitements très faibles;
 - ▶ Problèmes éventuels ?

Power is the Art of Tweaking! | La puissance est l'art du réglage !

We tweak different parts of our design up front to make sure that our experiment has enough power to detect effects (assuming they exist)

Nous modifions les différentes parties de notre conception dès le départ afin de nous assurer que notre expérience a suffisamment de puissance pour détecter les effets (en supposant qu'ils existent).

An Alternative Perspective: Minimum Detectable Effect | *Une autre perspective : L'effet minimal détectable*

Difficulty with Power Analysis |

- ▶ Hardest part of power analysis is plugging in treatment effect : how can we possibly know before experiment has been run?
- ▶ La partie la plus difficile de l'analyse de puissance consiste à introduire l'effet du traitement - comment pouvons-nous le savoir avant que l'expérience n'ait été réalisée?

Ask two questions: | *Posez deux questions :*

1. For a given set of inputs, what's the smallest effect that my study would be able to detect?
2. Would this effect-size be "satisfactory"?
 - ▶ Cost-effectiveness
 - ▶ Disciplinary rules of thumb (e.g. 0.2 SD effects in education research)
 - ▶ Other studies which had similar goals to yours

1. Pour un ensemble donné de données, quel est le plus petit effet que mon étude pourrait détecter ?
2. Cette taille d'effet serait-elle "satisfaisante" ?
 - ▶ Coût-efficacité
 - ▶ Règles empiriques disciplinaires (par exemple, effets de 0,2 SD dans la recherche sur l'éducation)
 - ▶ Autres études ayant des objectifs similaires aux vôtres

Minimum Detectable Effect | *L'effet minimal détectable*

- ▶ After we answer these two questions, we design the study that will allow us to detect an effect with that size.
- ▶ Après avoir répondu à ces deux questions, nous concevons l'étude qui nous permettent de détecter un effet de cet ampleur.

Special Case: Cluster-Randomized Designs | *Cas particulier : Designs d'une randomisation par grappe*

Special Case: Clustered-Randomized Designs | *Cas particulier : Designs d'une randomisation par grappe*

- ▶ Randomize treatment at the cluster level and measure outcome at the unit level.
 - ▶ We often only sample a few people in each cluster.
 - ▶ Examples?
- ▶ Randomiser le traitement au niveau des grappes et mesurer le résultat au niveau des unités.
 - ▶ Nous n'échantillonnons souvent que quelques personnes au sein de ces grappes.
 - ▶ Exemples ?

Special issues for power | *Challenges particuliers liés à la puissance*

- ▶ Number of individuals sampled per cluster
- ▶ Intra-cluster correlation: how similar are units that are in the same cluster as compared with other clusters
- ▶ Nombre d'individus échantillonnés par grappe
- ▶ Correlation intra-grappe: le degré de similarité des unités qui sont dans la même grappe par rapport aux autres grappes

Example | *Exemple*

- ▶ 2,000 bureaucrats, divided into 200 municipalities of 10 workers each; 100 municipalities in treatment and 100 municipalities in control.
- ▶ When the intracluster correlation is 0, bureaucrats within the same municipalities are not similar;
- ▶ It's like assigning 2,000 bureaucrats to treatment or control!
- ▶ 2000 bureaucrats, répartis en 200 communes de 10 bureaucrates par communes ; 100 communes dans le groupe de traitement et 100 communes dans le groupe de contrôle.
- ▶ Lorsque la corrélation intracluster est égale à 0, les bureaucrates d'une même municipalités ne sont pas semblables.
- ▶ Cela revient à assigner 2000 bureaucrates qui seront répartis dans le groupe de traitement et dans le groupe de contrôle!

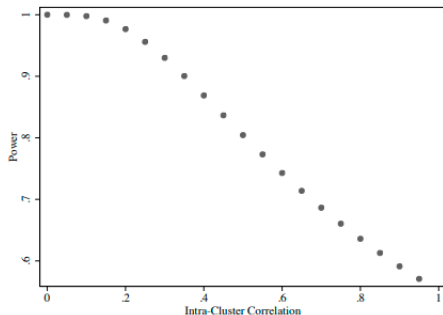
Example | *Exemple*

- ▶ When the intracluster correlation is 1, everyone within a municipality acts the same, and so you effectively have 200 independent observations
- ▶ Implications for power?
- ▶ Lorsque la corrélation intra grappe est de 1, tous les individus d'une même municipalité agissent de la même manière et vous disposez donc de 200 observations indépendantes.
- ▶ Implications pour la puissance ?

Tweak Intra-Cluster Correlation: How Does Power Respond? | *Modifier la corrélation intra grappe : Comment la puissance réagit-elle ? (1)*

► Number of clusters = 140; 10 sampled per cluster

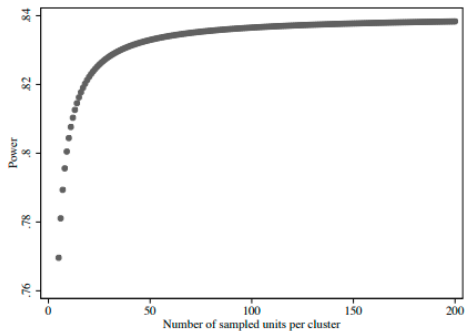
► Nombre de grappes = 140 ; 10 échantillonnées par grappe



Tweak Number of Units Per Cluster: How Does Power Respond? | *Modifier le nombre d'unités par grappe: Comment la puissance réagit-elle ? (2)*

- ▶ Another choice we have to make in cluster designs is how many units within clusters to sample
- ▶ Surely we want to sample as many as possible, right?
- ▶ Hmm..
- ▶ Un autre choix à faire dans les conceptions par grappe est celui du nombre d'unités à échantillonner à l'intérieur des grappes
- ▶ Nous voulons certainement en échantillonner le plus possible, n'est-ce pas ?
- ▶ Hmm. . .

Tweak Number of Units Per Cluster: How Does Power Respond? | Modifier le nombre d'unités par grappe : Comment la puissance réagit-elle ?



Golden Rule of Cluster-Randomized Designs | *Règle d'or des plans randomisés en grappes*

- ▶ Unless intra-cluster correlation is very small, it's always better to add more clusters than to sample more people within the clusters
- ▶ A moins que la corrélation intra-grappe ne soit très faible, il est toujours préférable d'ajouter des grappes plutôt que d'échantillonner davantage de personnes à l'intérieur des grappes.

Your Turn! | *À vous de jouer !*

- ▶ Go to <https://egap.org/resource/egap-app-power-calculator/>
 - ▶ Click box which says “Clustered Design?”
 - ▶ Set Significance Level at Alpha = 0.05
 - ▶ Set Treatment Effect Size at 5
 - ▶ Standard Deviation of Outcome Variable at 10
 - ▶ Set Power Target at 0.8
 - ▶ Set Maximum Number of Subjects at 2000
- ▶ Aller sur <https://egap.org/resource/egap-app-power-calculator/>
 - ▶ Cliquez sur la case qui dit “Clustered Design ?”
 - ▶ Fixez le niveau de signification à Alpha = 0,05
 - ▶ Fixez la taille de l'effet de traitement à 5
 - ▶ Écart-type de la variable de résultat à 10
 - ▶ Fixer l'objectif de puissance à 0,8
 - ▶ Fixer le nombre maximum de sujets à 2000

Your Turn! | *À vous de jouer !* (1)

1. Fix Number of Clusters per Arm at 40. How many subjects do I need if my Intra-cluster Correlation is 0.6 in order for my experiment to have 80% power? What about Intra-cluster Correlation of 0.3? 0?
1. Fixer le nombre de grappes par traitement à 40. De combien de sujets ai-je besoin si ma corrélation intra-groupe est de 0,6 pour que mon expérience ait une puissance de 80 % ? Qu'en est-il d'une corrélation intra-groupe de 0,3 ? 0 ?

Your Turn! | *À vous de jouer !* (2)

2. Fix Intra-cluster Correlation at 0.5.
How many subjects do I need if the Number of Clusters per Arm is 100 in order for my experiment to have 80% power? What if the Number of Clusters per Arm is 50? 20?

2. Fixer la corrélation intra-grappe à 0,5.
De combien de sujets ai-je besoin si le nombre de grappes par bras est de 100 pour que mon expérience ait une puissance de 80 % ? Quel serait le nombre d'individus par grappe à échantillonner si le nombre de grappe par traitement est de 50 ? 20 ?

Summary | Résumé (1)

- ▶ Informally: Power is the ability of our experiment to detect treatment effects, if they in fact exist
- ▶ Power matters: for practical reasons and for interpretation
- ▶ You can increase power by strengthening the intervention, reducing noise, and increasing sample size
- ▶ De manière informelle: La puissance est la capacité de notre expérience à détecter des effets de traitement, s'ils existent réellement
- ▶ La puissance est importante : pour des raisons pratiques et pour l'interprétation
- ▶ Vous pouvez augmenter la puissance en renforçant l'intervention, en réduisant le bruit et en augmentant la taille de l'échantillon

Summary | Résumé (2)

- ▶ In cluster-randomized designs, almost always better to add more clusters rather than interview more people within clusters
- ▶ Always run a power analysis before committing to a final design
- ▶ But beware that it involves some guesswork; be skeptical and vary assumptions
- ▶ Dans les expériences randomisées par grappe, il est presque toujours préférable d'ajouter des grappes plutôt que d'interroger davantage de personnes au sein des grappes.
- ▶ Effectuez toujours une analyse de puissance avant de vous engager dans une conception finale.
- ▶ Mais attention, cela implique des suppositions ; soyez sceptiques et variez les hypothèses.

- ▶ Some of these slides are not original. Material is borrowed from several sources:
 - ▶ Gareth Nellis, Learning Days slides (on power analysis)
 - ▶ Cyrus Samii, NYU slides (on minimum detectable effects)
 - ▶ Tara Slough, Columbia slides (graphs on the sensitivity of effects)
- ▶ Certaines de ces slides ne sont pas originales. Le matériel est extrait de plusieurs sources:
 - ▶ Gareth Nellis, Slides des Journées d'Apprentissages de EGAP (Analyse de puissance)
 - ▶ Cyrus Samii, NYU diapositives (sur les effets minimaux détectables)
 - ▶ Tara Slough, Columbia diapositives (graphiques sur la sensibilité des effets)