Le tirage au sort: pourquoi ça marche

Luc Behaghel

PSE, INRA et J-PAL

Les termes du problème

- La population d'intérêt

Les termes du problème

- La population d'intérêt
  - les demandeurs d'emploi

Les termes du problème

- La population d'intérêt
  - les demandeurs d'emploi

- Le traitement

Les termes du problème

- La population d'intérêt
  - les demandeurs d'emploi
- Le traitement
  - l'accompagnement renforcé ($T_i = 1$) plutôt que l'accompagnement standard ($T_i = 0$)

Les termes du problème

- La population d'intérêt
  - les demandeurs d'emploi
- Le traitement
  - l'accompagnement renforcé ($T_i = 1$) plutôt que l'accompagnement standard ($T_i = 0$)
- Les résultats potentiels

Les termes du problème

- La population d'intérêt
  - les demandeurs d'emploi

- Le traitement
  - l'accompagnement renforcé ($T_i = 1$) plutôt que l'accompagnement standard ($T_i = 0$)

- Les résultats potentiels
  - statut d'emploi avec accompagnement renforcé : $Y_i(1)$

Les termes du problème

- La population d'intérêt
  - les demandeurs d'emploi
- Le traitement
  - l'accompagnement renforcé ($T_i = 1$) plutôt que l'accompagnement standard ($T_i = 0$)
- Les résultats potentiels
  - statut d'emploi avec accompagnement renforcé : $Y_i(1)$
  - statut d'emploi sans accompagnement renforcé : $Y_i(0)$

L'effet du traitement

$$\Delta_i = Y_i(1) - Y_i(0).$$

Solution 1 : le clonage

"But economics is not a cloning science." (A. Everyone)

(*Journal of Common Sense*, 2017, p. 31)

Solution 2 : le tirage au sort

Un peu d'équations

$$E(X|T=1) = E(X|T=0) = E(X).$$

$$E(\text{motivation}|\,T = 1) = E(\text{motivation}|\,T = 0) = E(\text{motivation}).$$

$$E(\text{réseaux}|T = 1) = E(\text{réseaux}|T = 0) = E(\text{réseaux}).$$

Ca marche aussi pour les résultats potentiels

$$E(Y(0)|T=0)=E(Y(0)|T=1)=E(Y(0)).$$

Ca marche aussi pour les résultats potentiels

$$E(Y(0)|T = 0) = E(Y(0)|T = 1) = E(Y(0)).$$

et

$$E(Y(1)|T = 0) = E(Y(1)|T = 1) = E(Y(1)).$$

Et on déroule...

$$D = E(Y|T=1) - E(Y|T=0).$$

Et on déroule...

$$D = E(Y|T=1) - E(Y|T=0).$$

$$Y = \begin{cases} Y(1) \text{ si } T = 1 \\ Y(0) \text{ si } T = 0. \end{cases}$$

Et on déroule...

$$D = E(Y|T=1) - E(Y|T=0).$$

$$Y = \begin{cases} Y(1) \text{ si } T = 1 \\ Y(0) \text{ si } T = 0. \end{cases}$$

$$D = E(Y(1)|T=1) - E(Y(0)|T=0).$$

$$D = E(Y(1)|T = 1) - E(Y(0)|T = 0)$$

avec

$$\begin{aligned} E(Y(0)|T = 0) &= E(Y(0)) \\ E(Y(1)|T = 1) &= E(Y(1)) \end{aligned}$$

$$D = E(Y(1)|T=1) - E(Y(0)|T=0)$$

avec

$$\begin{aligned} E(Y(0)|T=0) &= E(Y(0)) \\ E(Y(1)|T=1) &= E(Y(1)) \end{aligned}$$

donc

$$\begin{aligned} D &= E(Y(1)) - E(Y(0)) \\ &= E(Y(1) - Y(0)) \\ &= \Delta. \end{aligned}$$

Le tirage au sort: comment ça marche

Luc Behaghel

PSE, INRA et J-PAL

Rappel : Deux ingrédients

- Un tirage au sort (“loterie”, “randomisation”)
- De grands échantillons

Questions qui restent

1. Comment tirer au sort ?
2. Comment savoir si les échantillons sont suffisamment grands ?
3. Quid si les individus tests et témoins ne suivent pas le tirage au sort ?
4. Quid si les individus tests influencent les résultats des témoins ?
5. Et l'éthique dans tout ça ?

Questions qui restent

1. Comment tirer au sort ?
2. Comment savoir si les échantillons sont suffisamment grands ?
3. Quid si les individus tests et témoins ne suivent pas le tirage au sort ?
4. Quid si les individus tests influencent les résultats des témoins ?
5. Et l'éthique dans tout ça ?

Aujourd'hui : Comment tirer au sort

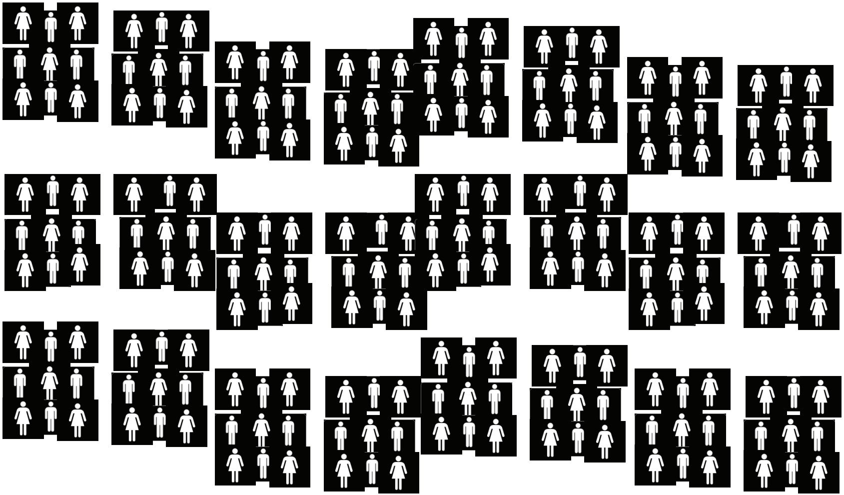
1. Un préalable : le choix de l'unité de tirage
2. Quatre méthodes de base
3. Quelques critères de choix

I. Choisir l'unité de tirage

- Tirer les individus un par un = tirage simple
  exp. : tirer parmi des élèves

- Tirer les individus par groupes = tirage en grappes
  exp. : tirer parmi des classes, des niveaux, ou des écoles

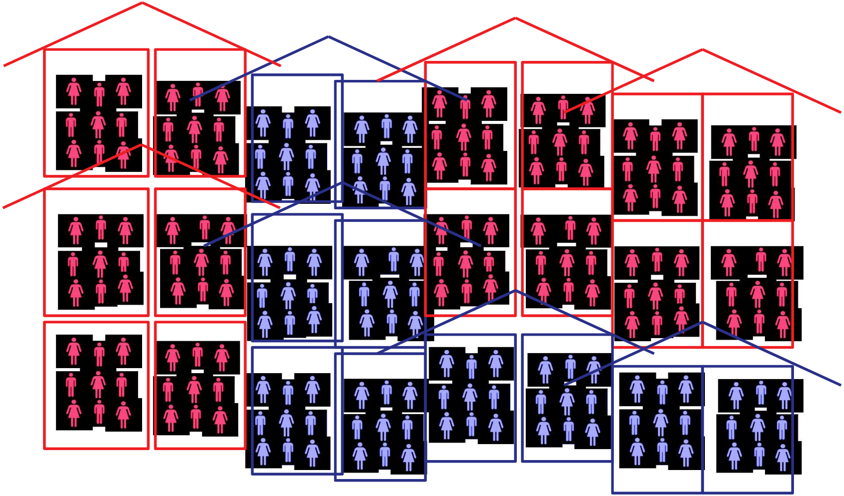
Population de départ



Tirage d'élèves

Tirage de classes

Tirage d'écoles



Unité de tirage $\geq$ unité d'intervention

L'unité de tirage ne peut être plus petite que l'unité d'intervention

Intervention	unité de tirage
Soutien individuel	Elève, classe ou école
Réduction taille des classes	Classe ou école
Hausse du budget	Ecole

II. Quatre méthodes

1. Le tirage au sort simple

- Référence des essais cliniques
  - médicament ou placebo, selon l'ordre de recrutement dans l'essai

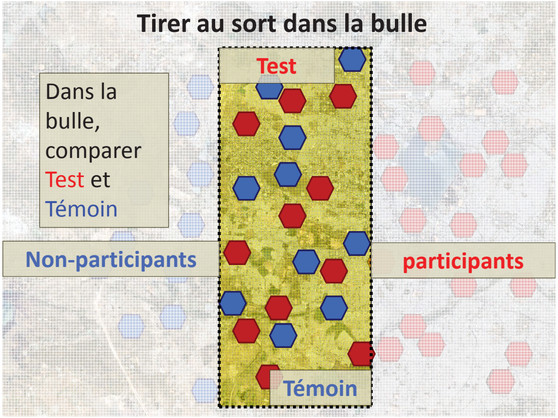
- En sciences sociales : particulièrement indiqué en cas d'excès de candidats éligibles, sans priorité claire ("oversubscription design")

2. Le tirage au sort “dans la bulle”

Excès de candidats mais...

- Candidats “absolument prioritaires”
- Candidats “pas du tout prioritaires”

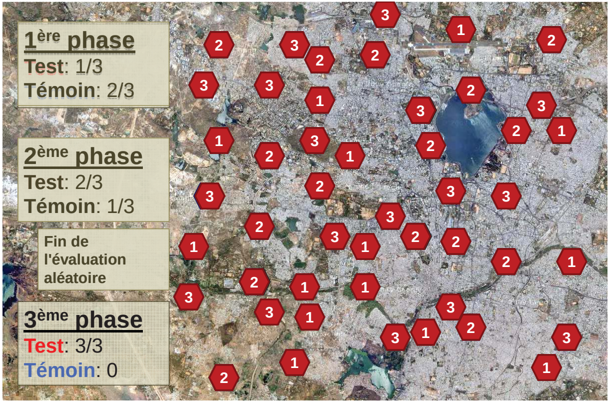
- Candidats “entre les deux”


Tirer au sort dans la bulle
Test
Dans la bulle, comparer Test et Témoin
Non-participants
participants
Témoin

3. Mise en place progressive

De la place pour tous mais pas tout de suite

- tirer au sort l'ordre de déploiement du programme ("randomized phase-in")


1ère phase
Test: 1/3
Témoin: 2/3
2ème phase
Test: 2/3
Témoin: 1/3
Fin de l'évaluation aléatoire
3ème phase
Test: 3/3
Témoin: 0

4. Encouragement

Pas possible ou pas acceptable de tirer au sort l'accès à un programme
▶ mais possible de tirer au sort l'effort supplémentaire d'encouragement à participer (incitations, information,...)

Exp. : réunions sur l'orientation scolaire

- Information habituelle ▶ groupe témoin
- Information personnalisée et relance ▶ groupe test

III. Quelle méthode de tirage choisir ?

1. Faisabilité
2. Acceptabilité
3. Portée de l'évaluation
4. Capacité de détection

Choisir la taille des échantillons

Luc Behaghel

PSE, INRA et J-PAL

Retour sur un exemple

- Type A: motivé

$$Y(0) = 1, \qquad Y(1) = 1.$$

Retour sur un exemple

- Type A: motivé

$$Y(0) = 1, \qquad Y(1) = 1.$$

- Type B: pas motivé

$$Y(0) = 0, \qquad Y(1) = 1.$$

$$E(Y(1) - Y(0)) = \frac{1}{2}E(Y(1) - Y(0)|A) + \frac{1}{2}E(Y(1) - Y(0)|B)$$

$$
\begin{aligned}
E(Y(1) - Y(0)) &= \frac{1}{2}E(Y(1) - Y(0)|A) + \frac{1}{2}E(Y(1) - Y(0)|B) \\
&= \frac{1}{2} \times 0 + \frac{1}{2} \times 1
\end{aligned}
$$

$$
\begin{aligned}
E(Y(1) - Y(0)) &= \frac{1}{2}E(Y(1) - Y(0)|A) + \frac{1}{2}E(Y(1) - Y(0)|B) \\
&= \frac{1}{2} \times 0 + \frac{1}{2} \times 1 \\
&= \frac{1}{2}.
\end{aligned}
$$

Expérience randomisée avec un élève test et un élève témoin

$$\hat{\Delta} = \bar{Y}^{T=1} - \bar{Y}^{T=0}.$$

Expérience randomisée avec un élève test et un élève témoin

$$\hat{\Delta} = \bar{Y}^{T=1} - \bar{Y}^{T=0}.$$

Quatre tirages possibles:

1. (A, A): $\hat{\Delta} = 1 - 1 = 0.$

Expérience randomisée avec un élève test et un élève témoin

$$\hat{\Delta} = \bar{Y}^{T=1} - \bar{Y}^{T=0}.$$

Quatre tirages possibles:

1. (A, A): $\hat{\Delta} = 1 - 1 = 0.$
2. (A, B): $\hat{\Delta} = 1 - 0 = 1.$

Expérience randomisée avec un élève test et un élève témoin

$$\hat{\Delta} = \bar{Y}^{T=1} - \bar{Y}^{T=0}.$$

Quatre tirages possibles:

1. (A, A): $\hat{\Delta} = 1 - 1 = 0.$
2. (A, B): $\hat{\Delta} = 1 - 0 = 1.$
3. (B, A): $\hat{\Delta} = 1 - 1 = 0.$

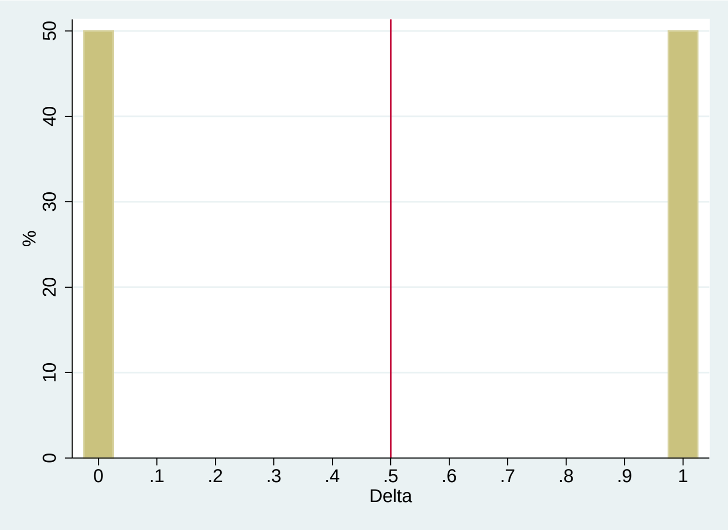
Expérience randomisée avec un élève test et un élève témoin

$$\hat{\Delta} = \bar{Y}^{T=1} - \bar{Y}^{T=0}.$$

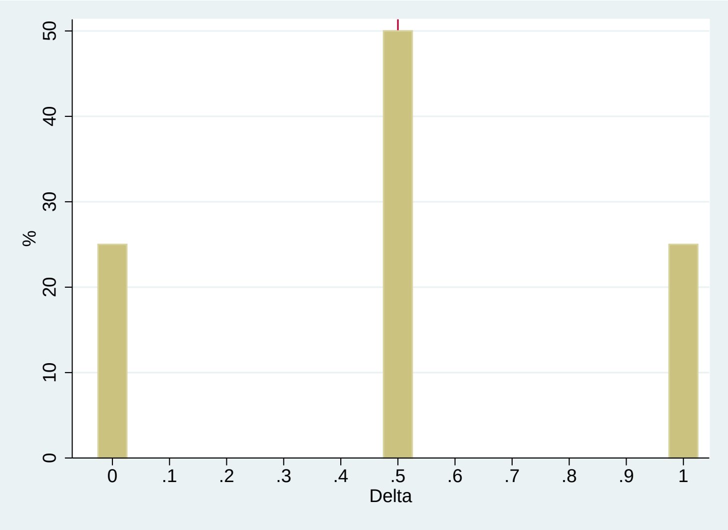
Quatre tirages possibles:

1. (A, A): $\hat{\Delta} = 1 - 1 = 0.$
2. (A, B): $\hat{\Delta} = 1 - 0 = 1.$
3. (B, A): $\hat{\Delta} = 1 - 1 = 0.$
4. (B, B): $\hat{\Delta} = 1 - 0 = 1.$

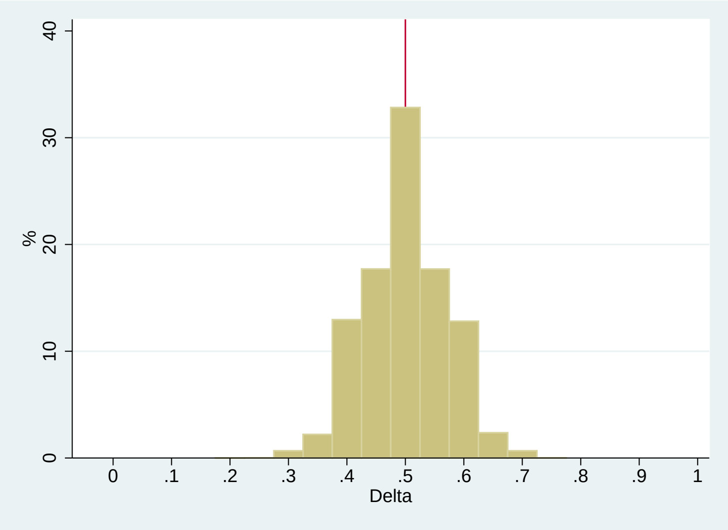
Distribution d'échantillonnage de l'effet: $n = 2$



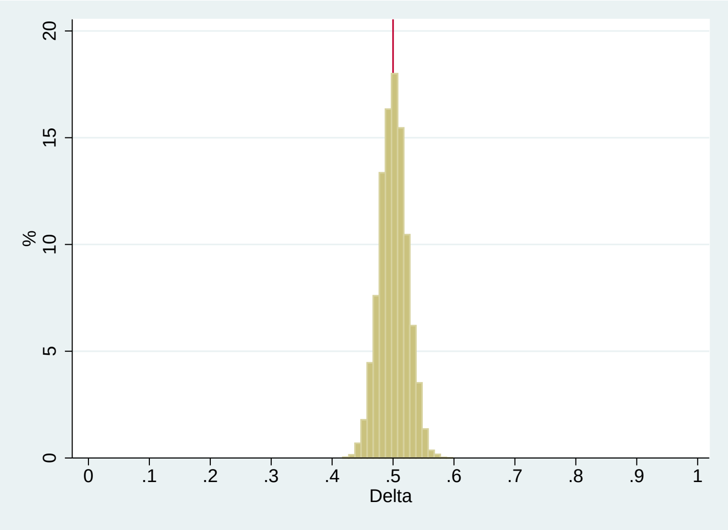
Distribution d'échantillonnage de l'effet: $n = 4$



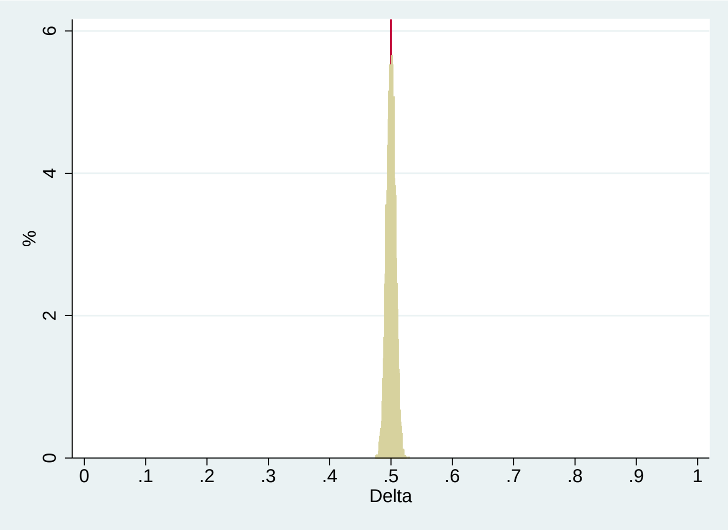
Distribution d'échantillonnage de l'effet: $n = 100$



Distribution d'échantillonnage de l'effet: $n = 1000$



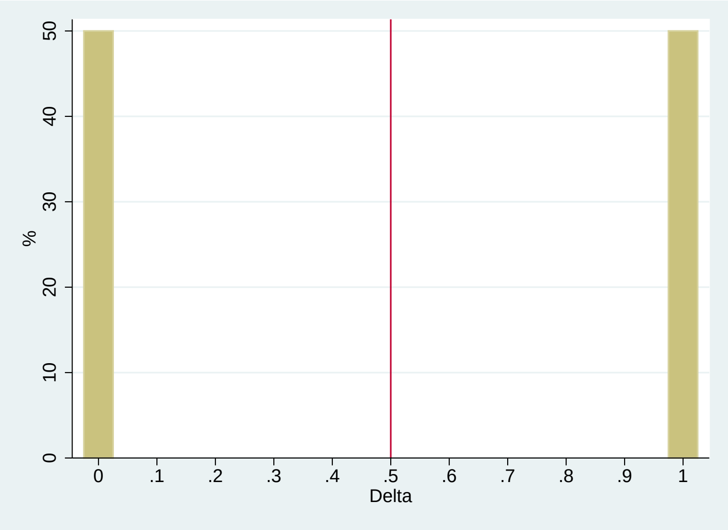
Distribution d'échantillonnage de l'effet: $n = 10000$



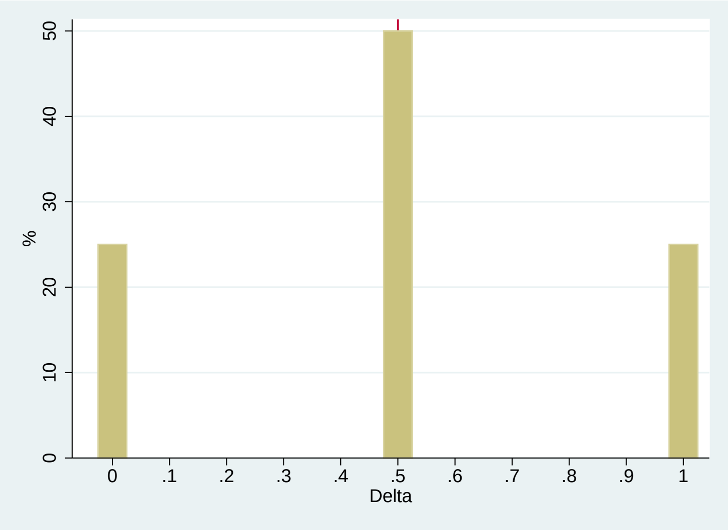
Un résumé utile: l'écart-type de l'estimateur

- distance moyenne entre les différentes estimations ($\hat{\Delta}$) et la vraie valeur (Δ).

Distribution d'échantillonnage de l'effet: $n = 2$



Distribution d'échantillonnage de l'effet: $n = 4$



Intervalles de confiance

Le véritable effet a 95% de chances de se trouver dans l'intervalle

$$CI_{95} = \left[\hat{\Delta} - 1.96 \times S ; \hat{\Delta} + 1.96 \times S\right]$$

Test de significativité

$\hat{\Delta} \neq 0 \Rightarrow ?$

Test de significativité

$\hat{\Delta} \neq 0 \Rightarrow$?

1. $\Delta \neq 0$?

Test de significativité

$\hat{\Delta} \neq 0 \Rightarrow$?

1. $\Delta \neq 0$?
2. $\Delta = 0$, mais fluctuations d'échantillonnage?

Test de significativité

$\hat{\Delta} \neq 0 \Rightarrow$?

1. $\Delta \neq 0$?
2. $\Delta = 0$, mais fluctuations d'échantillonnage?

$$t = \frac{\hat{\Delta}}{S}.$$

- l'écart entre $\hat{\Delta}$ et 0 est-il grand par rapport à l'écart créé en moyenne par les fluctuations d'échantillonnage (S)?

Test de significativité

$\hat{\Delta} \neq 0 \Rightarrow$?

1. $\Delta \neq 0$?
2. $\Delta = 0$, mais fluctuations d'échantillonnage?

$$t = \frac{\hat{\Delta}}{S}.$$

▶ l'écart entre $\hat{\Delta}$ et 0 est-il grand par rapport à l'écart créé en moyenne par les fluctuations d'échantillonnage (S)?

$|\hat{\Delta}/S| > 1.96$ est rare si $\Delta = 0$: risque $< 5\%$
▶ effet "statistiquement significatif" à 5%.

Effet minimum détectable

Taille nécessaire de l'effet pour avoir 80% de chances de rejeter l'hypothèse d'effet nul avec un risque de 5%?

Effet minimum détectable

Taille nécessaire de l'effet pour avoir 80% de chances de rejeter l'hypothèse d'effet nul avec un risque de 5%?

$$EMD = 2.8 \times S.$$

Intervalles de confiance

Le véritable effet a 95% de chances de se trouver dans l'intervalle

$$CI_{95} = \left[\hat{\Delta} - 1.96 \times S; \hat{\Delta} + 1.96 \times S\right]$$

Test de significativité

$\hat{\Delta} \neq 0 \Rightarrow ?$

Test de significativité

$\hat{\Delta} \neq 0 \Rightarrow$?

1. $\Delta \neq 0$?

Test de significativité

$\hat{\Delta} \neq 0 \Rightarrow$?

1. $\Delta \neq 0$?
2. $\Delta = 0$, mais fluctuations d'échantillonnage?

Test de significativité

$\hat{\Delta} \neq 0 \Rightarrow$?

1. $\Delta \neq 0$?
2. $\Delta = 0$, mais fluctuations d'échantillonnage?

$$t = \frac{\hat{\Delta}}{S}.$$

Test de significativité

$\hat{\Delta} \neq 0 \Rightarrow$?

1. $\Delta \neq 0$?
2. $\Delta = 0$, mais fluctuations d'échantillonnage?

$$t = \frac{\hat{\Delta}}{S}.$$

$|\hat{\Delta}/S| > 1.96$ est rare si $\Delta = 0$: risque $< 5\%$

▶ effet "statistiquement significatif" à 5%.

Effet minimum détectable

Taille nécessaire de l'effet pour avoir 80% de chances de rejeter l'hypothèse d'effet nul avec un risque de 5%?

Effet minimum détectable

Taille nécessaire de l'effet pour avoir 80% de chances de rejeter l'hypothèse d'effet nul avec un risque de 5%?

$$EMD = 2.8 \times S.$$

Anticiper la capacité de détection d'une évaluation – et l'améliorer

1. Tirage simple
2. Tirage en grappes
3. Tirage stratifiés
4. Le rôle de l'enquête initiale

a. Tirage au sort simple

$$S = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_0^2}{n_0}},$$

σ_1: écart-type de $Y_i(1)$, σ_0: écart-type de $Y_i(0)$, n_1: nombre d'individus tests; n_0: nombre d'individus témoins.

a. Tirage au sort simple

$$S = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_0^2}{n_0}},$$

σ_1: écart-type de $Y_i(1)$, σ_0: écart-type de $Y_i(0)$, n_1: nombre d'individus tests; n_0: nombre d'individus témoins.

$$\begin{aligned} D &= \frac{1}{n_1} \sum_{T_i=1} Y_i - \frac{1}{n_0} \sum_{T_i=0} Y_i \\ Var(D) &= \ldots \end{aligned}$$

Si $\sigma_0 = \sigma_1 = \sigma$,

$$S = \frac{\sigma}{\sqrt{n(1-p)p}}.$$

Si $\sigma_0 = \sigma_1 = \sigma$,

$$S = \frac{\sigma}{\sqrt{n(1-p)p}}.$$

1. n: taille de l'échantillon.

Si $\sigma_0 = \sigma_1 = \sigma$,

$$S = \frac{\sigma}{\sqrt{n(1-p)p}}.$$

1. n: taille de l'échantillon.
2. p: équilibre groupe test / témoin.

Si $\sigma_0 = \sigma_1 = \sigma$,

$$S = \frac{\sigma}{\sqrt{n(1-p)p}}.$$

1. n: taille de l'échantillon.
2. p: équilibre groupe test / témoin.
3. σ: dispersion du résultat auquel on s'intéresse.

Si $\sigma_0 = \sigma_1 = \sigma$,

$$S = \frac{\sigma}{\sqrt{n(1-p)p}}.$$

1. n: taille de l'échantillon.
2. p: équilibre groupe test / témoin.
3. σ: dispersion du résultat auquel on s'intéresse.

- Estimer σ.

b. Tirage au sort par grappes

- Qui se ressemble s'assemble...
- Les fluctuations d'échantillonnage s'accroissent.

Effet de design

$$S_{\text{grappe}} = S_{\text{simple}} \times \sqrt{1 + \rho(m-1)},$$

ρ: coefficient de corrélation intraclasse; m: nombre d'élèves par classe.

Effet de design

$$S_{\text{grappe}} = S_{\text{simple}} \times \sqrt{1 + \rho(m-1)},$$

ρ: coefficient de corrélation intraclasse; m: nombre d'élèves par classe.

- estimer ρ.

c. Tirage stratifié

Idée: Regrouper les individus qui se ressemblent en “strates”, et tirer au sein de chaque strate des individus tests et témoins.

- garantit que chaque strate est bien représentée dans les groupes tests et témoins.
- réduit les fluctuations d’échantillonnage.

Stratifier par école

d. L'enquête initiale

1. Pour stratifier.

d. L'enquête initiale

1. Pour stratifier.
2. Pour vérifier – et tirer de nouveau!

d. L'enquête initiale

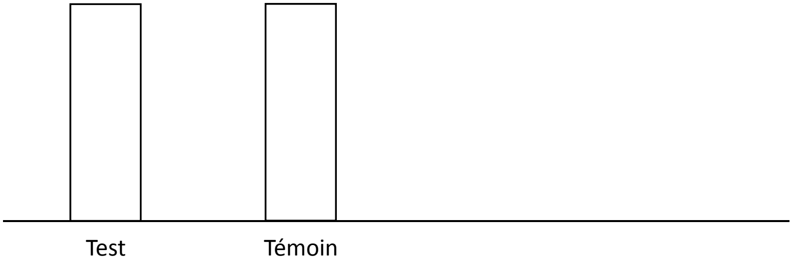
1. Pour stratifier.
2. Pour vérifier – et tirer de nouveau!
3. Pour réduire σ.

Difficultés et solutions

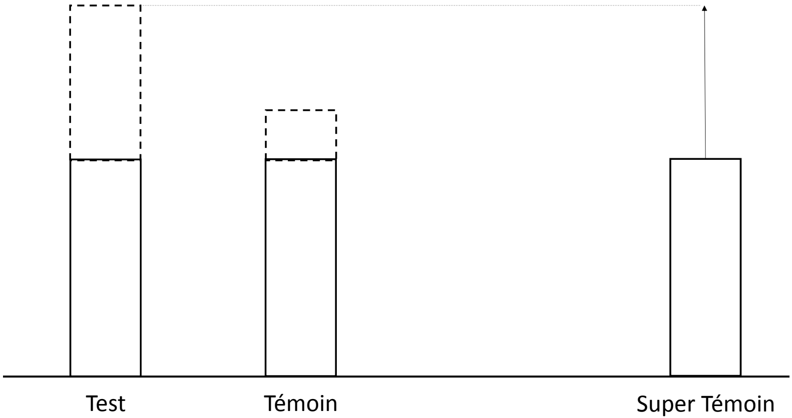
Luc Behaghel

PSE, INRA et J-PAL

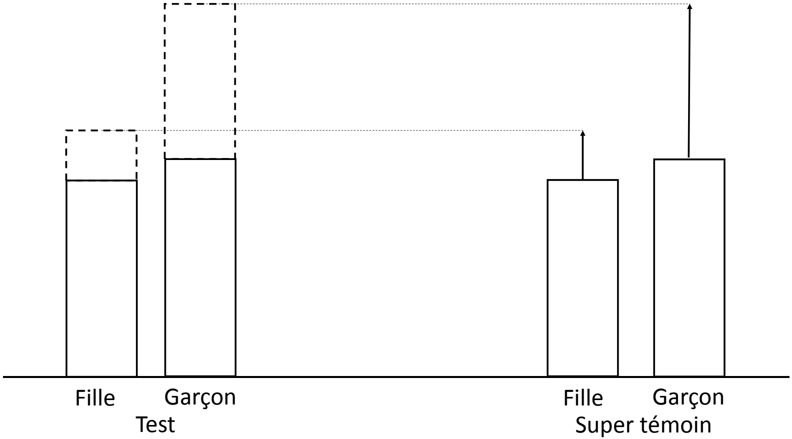
Menace n° 1: le groupe témoin est affecté



Solution: les “super témoins”



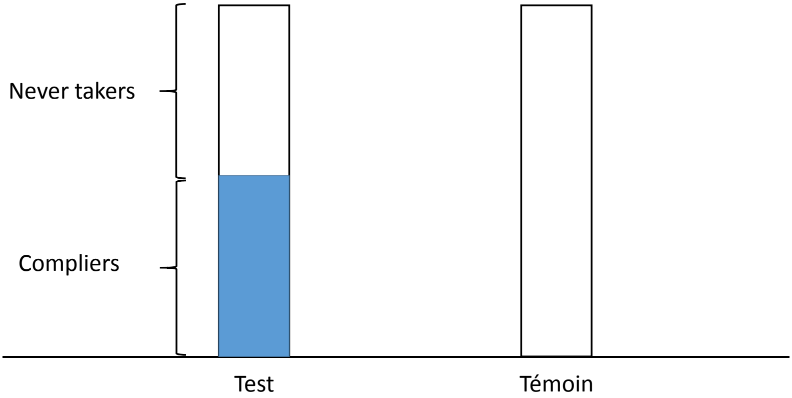
Super témoins et externalité au sein des écoles

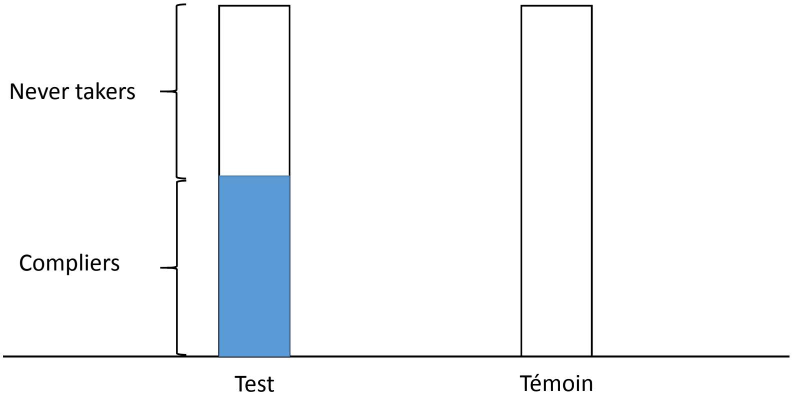


Autres exemples

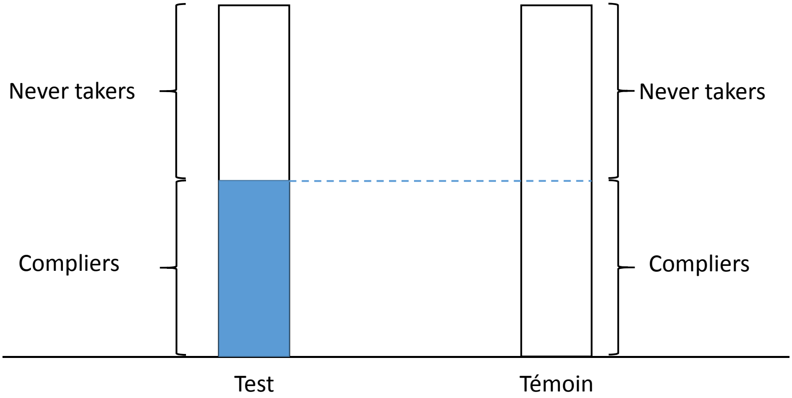
- Accompagnement des demandeurs d'emploi: un jeu à somme nulle? (Crépon, Duflo, Gurgand, Rathelot et Zamora, *Quarterly Journal of Economics* 2013)
- Effets de pairs et implication des parents (Avvisati, Gurgand, Guyon et Maurin, *Review of Economic Studies* 2014)

Menace n° 2: des individus test déclinent le traitement




Never takers
Compliers
Test
Témoin

La beauté du tirage au sort: compliers à droite, compliers à gauche



La beauté des équations?

$$
\begin{aligned}
\Delta_{ITT} &= 10 \\
&= \frac{1}{2} \times \Delta_{Never-takers} + \frac{1}{2} \times \Delta_{Compliers} \\
&= \frac{1}{2} \times 0 + \frac{1}{2} \times \Delta_{Compliers} \\
&= \frac{1}{2} \times \Delta_{Compliers}
\end{aligned}
$$

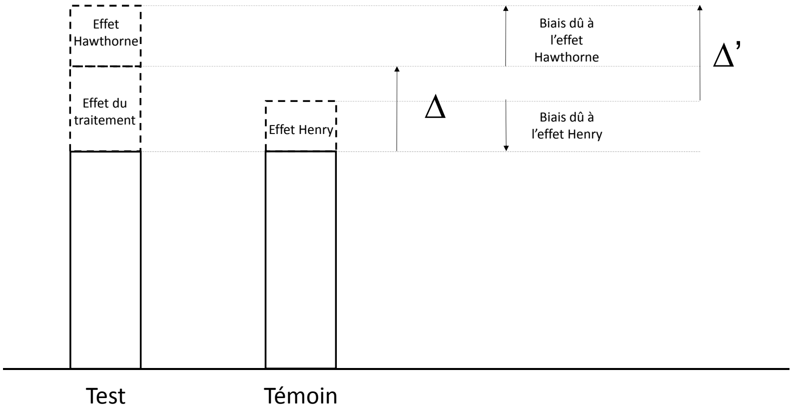
$$
\begin{aligned}
\Delta_{ITT} &= \frac{1}{2} \times \Delta_{Never-takers} + \frac{1}{2} \times \Delta_{Compliers} \\
&= \frac{1}{2} \times \Delta_{Compliers}
\end{aligned}
$$

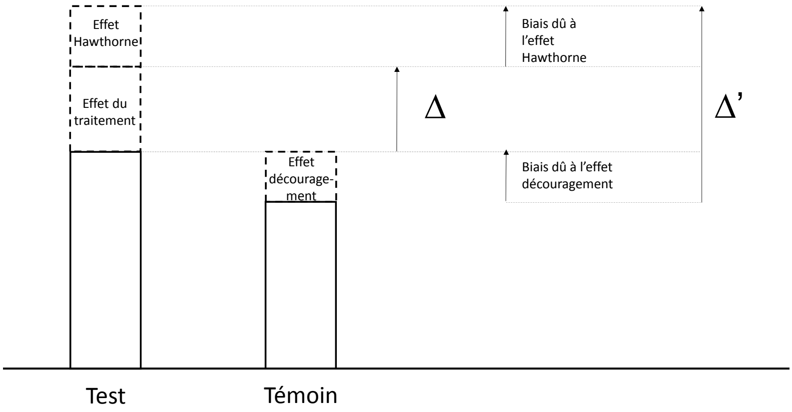
$$
\begin{aligned}
\Delta_{Compliers} &= 2 \times \Delta_{ITT} \\
&= 2 \times 10 \\
&= 20
\end{aligned}
$$

Pour en savoir plus

Angrist, Imbens and Rubin, " Identification of Causal Effects Using Instrumental Variables", *Journal of the American Statistical Association* 1996.

Menace n° 3: effet Placebo & Cie




Effet Hawthorne
Effet du traitement
Effet découragement
Biais dû à l'effet Hawthorne
Δ
Δ'
Biais dû à l'effet découragement
Test
Témoin

> *When a foreign agency comes in with its heavy boots and suitcases of dollars to administer a 'treatment,' whether through a local NGO or government or whatever, there is a lot going on other than the treatment.*

Drèze (2016), quoted in Deaton and Cartwright (2016)

Que faire?

1. Placebos
2. Tests en aveugle
3. Privilégier des variables de résultats "dures", moins susceptibles de manipulation (consciente ou inconsciente)
4. Evaluer le risque de biais avec des données complémentaires

Exemple de 3 et 4: évaluation du CV anonyme (Behaghel, Crépon et Le Barbanchon, 2015)

Menace n° 4: des résultats non transportables

Theorem

Les résultats d'un contexte donné ne sont jamais transposables sans hypothèses additionnelles à un autre contexte.

(Banerjee, Chassang et Snowberg, 2016)

Solution 1: réplication

Mais...

> *L'homme qui a nourri le poulet tous les jours de sa vie finit par lui tordre le cou, montrant par là qu'il eût été bien utile audit poulet d'avoir une vision plus subtile de l'uniformité de la nature.*

(Russell, 1912)

Solution 2: “simple theory” (Deaton et Cartwright, 2016)

- Kant: pour que l’expérimentation soit possible, il faut que la causalité existe.
- Deaton: exemple de la “théorie simple des prix”
  … mais les “lois” de l’offre et de la demande ne tiennent pas toujours.

Solution 3: "structured speculation" (Banerjee, Chassang et Snowberg, 2016)

Une section à part, systématique, où l'auteur engage sa crédibilité

- sur des prévisions précises et falsifiables
- mobilisant sa perception (non démontrée) des mécanismes, l'hétérogénéité des effets dans les différentes sous-populations, ou des modèles théoriques
- reconnaissant une part de subjectivité

Exemple de spéculation structurée (Dupas, 2014)

Spéculation à partir d'une expérimentation de l'effet de subventions de court terme sur l'adoption à long terme des moustiquaires imprégnées d'insecticide
▶ quel serait l'effet de subventions similaires sur d'autres technologies?

- L'effet dépend *en théorie*
  - de la vitesse à laquelle on peut apprendre sur la technologie
  - de la vitesse à laquelle les bénéfices et les coûts de l'adoption se matérialisent
- Selon les *a priori* qu'on a sur ces paramètres, on peut prévoir l'effet de la subvention sur l'adoption de réchauds écologiques, de dispositifs de purification de l'eau, etc.