



THE WORLD BANK



# Session VII

# Echantillonnage et puissance

Christel Vermeersch

janvier 2008

Human Development  
Network

Latin American and  
the Caribbean Region

Finance, Private Sector,  
and Infrastructure  
Department

World Bank Institute  
Evaluation Group

# Introduction

---

- ❑ Objectif : évaluer les programmes et les interventions  
( Ont-ils un effet ou pas ?)
- ❑ Nous avons besoin de données suffisantes pour être en mesure de détecter des changements imputables au programme
  - Partie 1 : discuter de la puissance
- ❑ Nous avons besoin de données représentatives :
  - Partie 2 : discuter de l'échantillonnage

# PARTIE 1 :

## Avoir « suffisamment » de données

- ❑ Evaluer un programme est un test de signification
- ❑ Exemple :  $y$  est une variable dénotant les résultats à une épreuve scolaire. Nous essayons d'estimer les effets du doublement du budget de l'école sur les résultats des étudiants à l'épreuve.

$$H_0: E(y|\text{Traitement}) = E(y|\text{Témoin})$$

C.àD.: le programme n'a pas d'effet

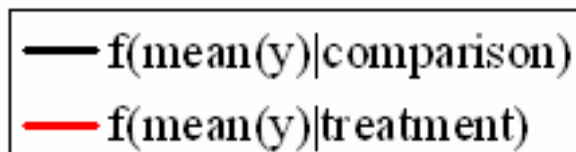
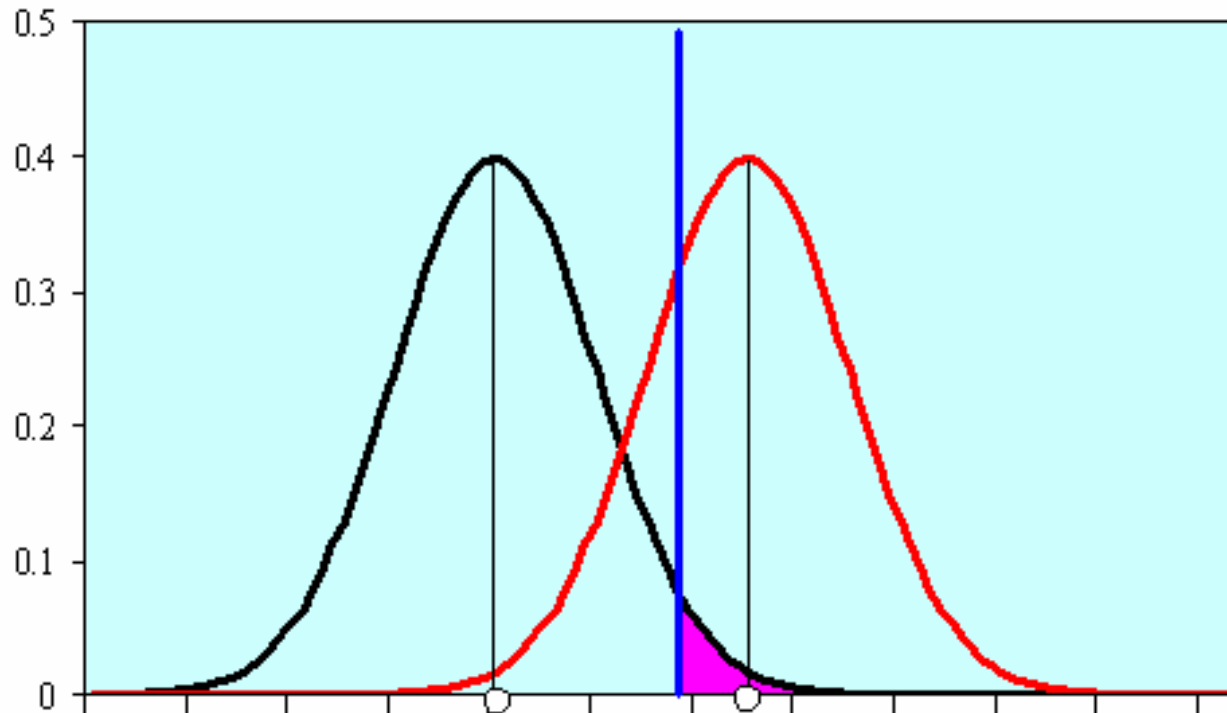
$$H_1: E(y|\text{Traitement}) > E(y|\text{Témoin})$$

C.àD.: le programme a un effet positif

		L'intervention a un effet (en réalité)	
		Non	Oui
Test statistique	Nous rejetons $H_0$	Erreur Type I	OK
	Nous ne rejetons pas $H_0$	OK	Erreur Type II

# Erreur de Type I

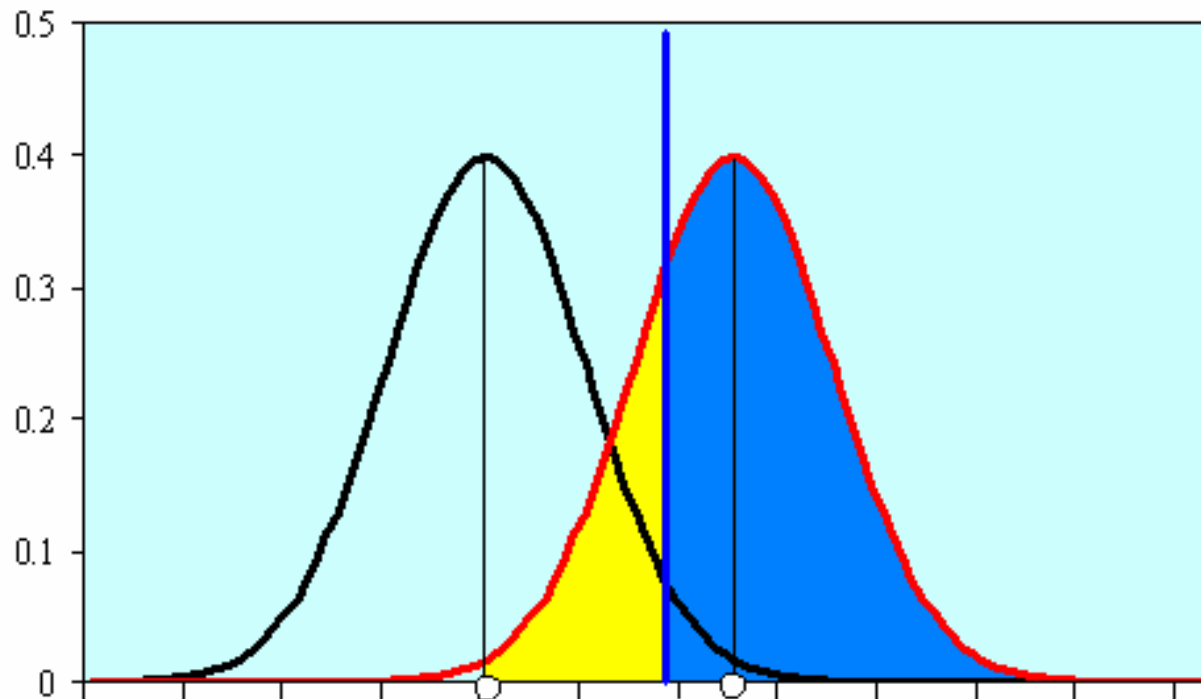
## Type I error



# Erreur de Type II et puissance

Type II error

Power



—  $f(\text{mean}(y)|\text{comparison})$   
—  $f(\text{mean}(y)|\text{treatment})$

# Erreurs de Type I et Type II

- Erreur de Type I
  - = rejeter l'hypothèse nulle quand elle est vraie
  - **Niveau de signification  $\alpha$  = probabilité que nous allons conclure que l'intervention a un effet quand en réalité elle n'en a pas**
  - Valeurs typiques pour  $\alpha$  : 0.01, 0.05, 0.1
- Erreur de Type II
  - = échouer à rejeter l'hypothèse nulle quand l'hypothèse nulle est fautive
- Puissance =  $1 - \beta$ 
  - Probabilité que nous rejeterons l'hypothèse nulle quand l'hypothèse nulle est fautive
  - = Probabilité que nous rejeterons l'hypothèse nulle quand l'hypothèse alternative est vraie
  - **= probabilité que nous allons conclure que l'intervention a un effet quand elle a réellement un effet**
- Objectif : maximiser la puissance, pour un niveau donné de  $\alpha$

# Qu'est-ce qui affecte la puissance d'un test?

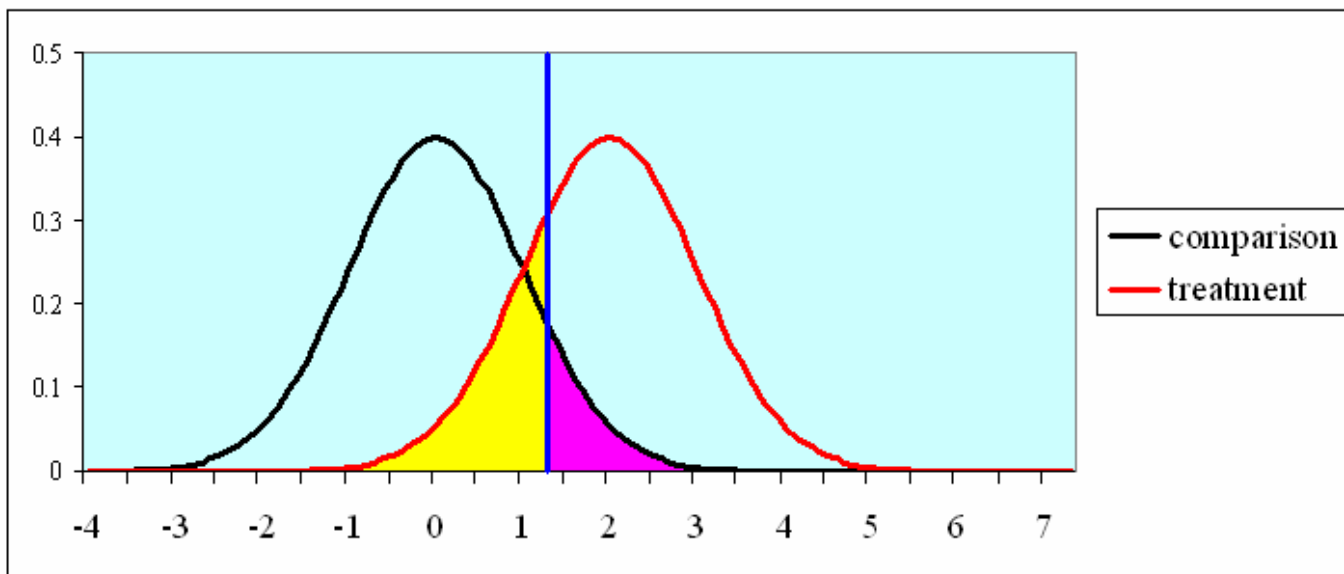
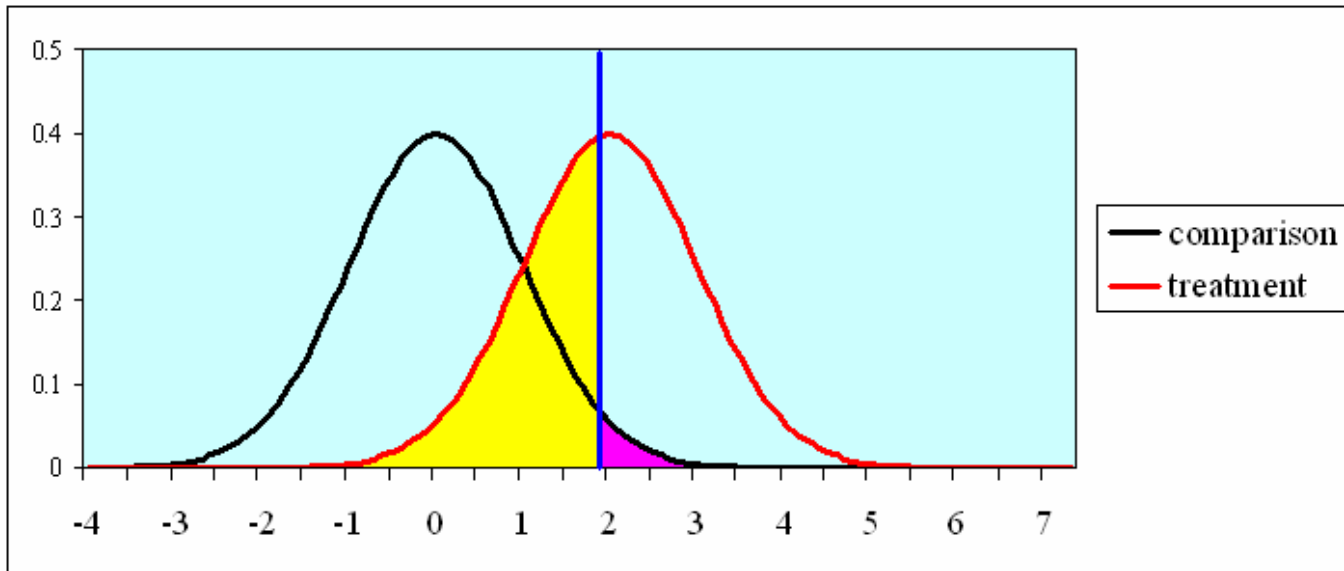
La puissance d'un test augmente quand...

- ❑  $\alpha$  augmente (nous nous accommodons de plus d'erreur de type I)
- ❑ Quand l'effet attendu est plus grand
- ❑ Quand la variance de l'indicateur augmente... comment cela se passe-t-il???
- Se rappeler que la variance d'une moyenne est proportionnelle à  $1/n$ , où  $n$  est la taille de l'échantillon
- Ainsi : augmenter la taille de l'échantillon diminuera la variance d'une moyenne (de nombreux indicateurs sont des moyennes !)

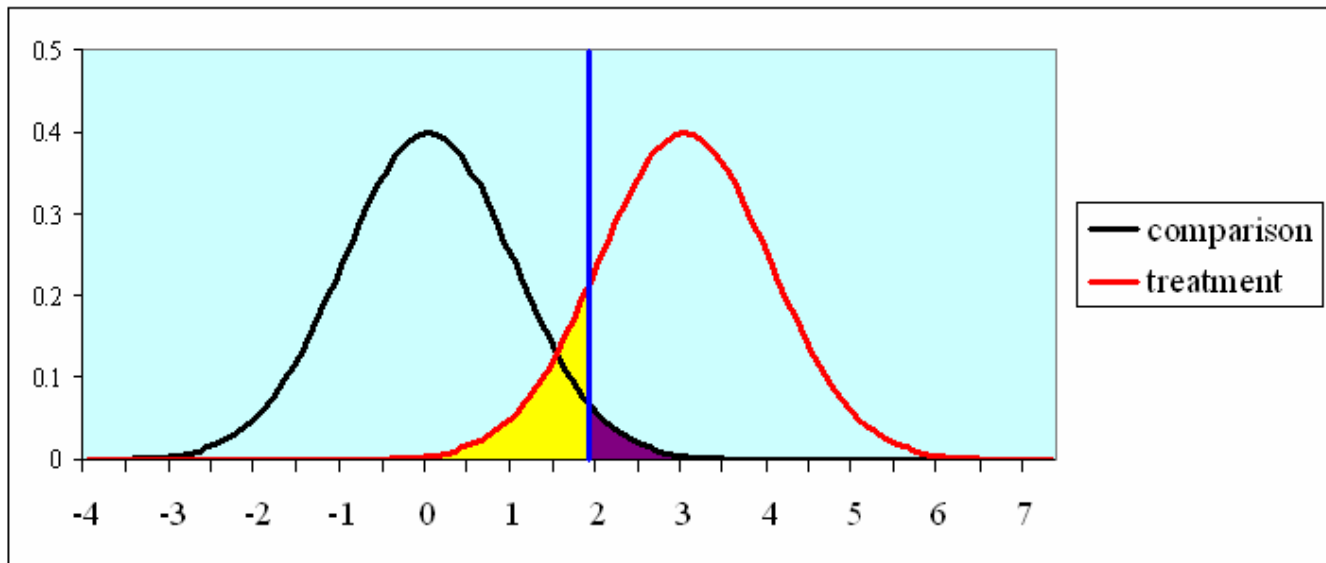
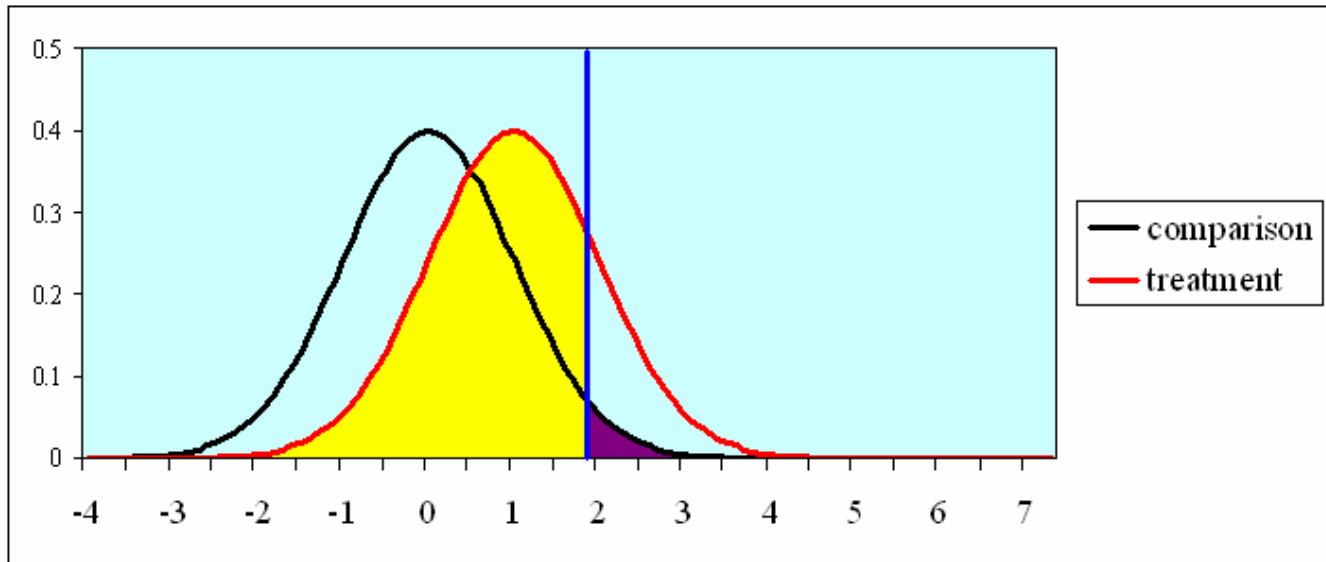
⇒ Pour augmenter la puissance d'un échantillon, quand la taille de l'effet est fixe et l'erreur de type I est fixe, nous aurons à augmenter la taille de l'échantillon.

⇒ Dans la pratique, nous calculons la taille de l'échantillon qui est nécessaire pour détecter un changement  $X$  dans l'indicateur, avec une erreur de Type I de 5 % et une puissance de 80 ou 90 %.

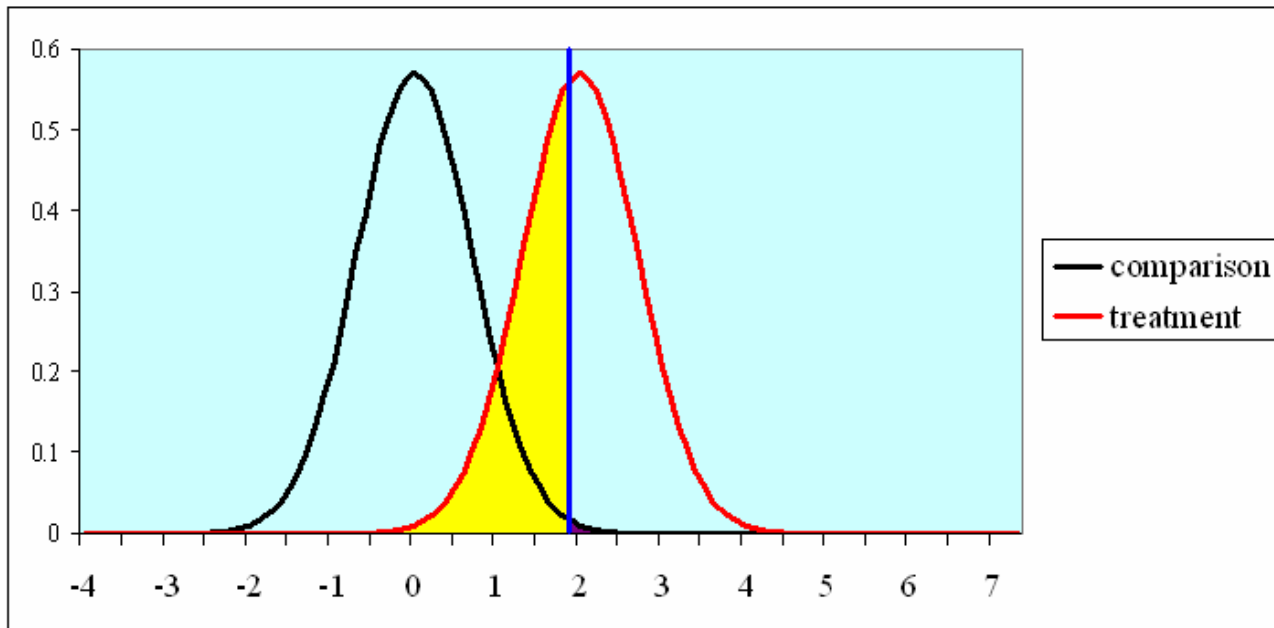
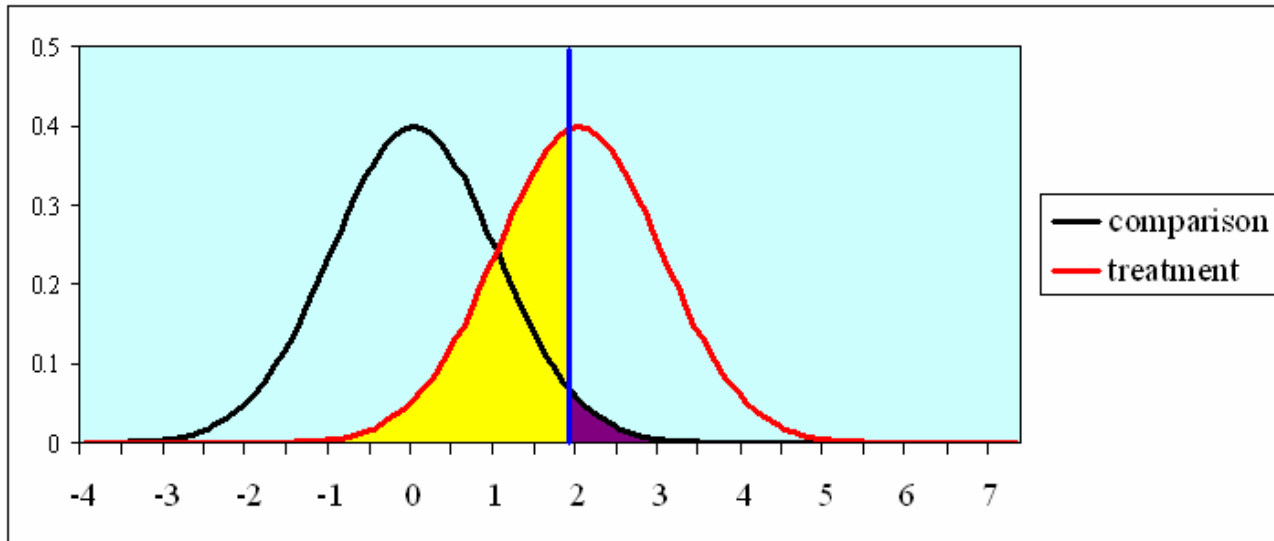
# Que se passe-t-il quand nous augmentons l'erreur de Type I ?



# Que se passe-t-il quand nous augmentons la taille attendue de l'impact ?



# Que se passe-t-il quand nous augmentons la taille de l'échantillon ?



# Exercice à faire à domicile : quelques simulations en Stata

**\*calculer une taille de l'échantillon;**

```
sampsi 130 135, alpha (0.05) power(0.8) sd1 (15)  
sd2 (18) onesided;
```

**\*Augmentons la taille attendue de l'effet==> un  
échantillon plus petit sera suffisant!;**

```
sampsi 130 145, alpha(0.05) power(0.8) sd1 (15)  
sd2 (18) onesided;
```

**\*Augmentons la puissance requise ==> nous avons  
besoin d'un échantillon plus grand;**

```
sampsi 130 135, alpha (0.05) power(0.9) sd1 (15)  
sd2 (18) onesided;
```

# Exercice à faire à domicile : d'autres simulations en Stata

**\*Calculons la puissance d'un test;**

```
sampsi 130 135, alpha (0.05) sd1(15) sd2(18)  
n1(100) n2(100) onesided;
```

**\*Augmentons la taille attendue de l'effet ==>  
la puissance augmente;**

```
sampsi 130 145, alpha (0.05) sd1(15) sd2(18)  
n1(100) n2(100) onesided;
```

**\*Augmentons la taille de l'échantillon => la  
puissance augmente;**

```
sampsi 130 135, alpha (0.05) sd1(15) sd2(18)  
n1(200) n2(200) onesided;
```

# Partie 2 : Echantillonnage (→ Données représentatives)

## □ Enquêtes représentatives

- Leur but : apprendre sur une population entière
  - Ex. LSMS/ enquête nationale sur les ménages
- Leur échantillon : doit être représentatif de la population nationale

## □ Evaluation d'impact

- Son but : mesurer les changements dans les indicateurs clés pour la population ciblée par une intervention
- Dans la pratique : mesurer la différence dans les indicateurs entre groupes de bénéficiaires et groupes témoins
- Nous échantillonnons stratégiquement afin d'avoir un échantillon représentatif dans les groupes de bénéficiaires et les groupes témoins
- Ce qui n'est pas nécessairement le même qu'un échantillon représentatif de la population nationale

# Pouvons-nous utiliser une enquête générale sur les ménages pour l'évaluation d'impact ?

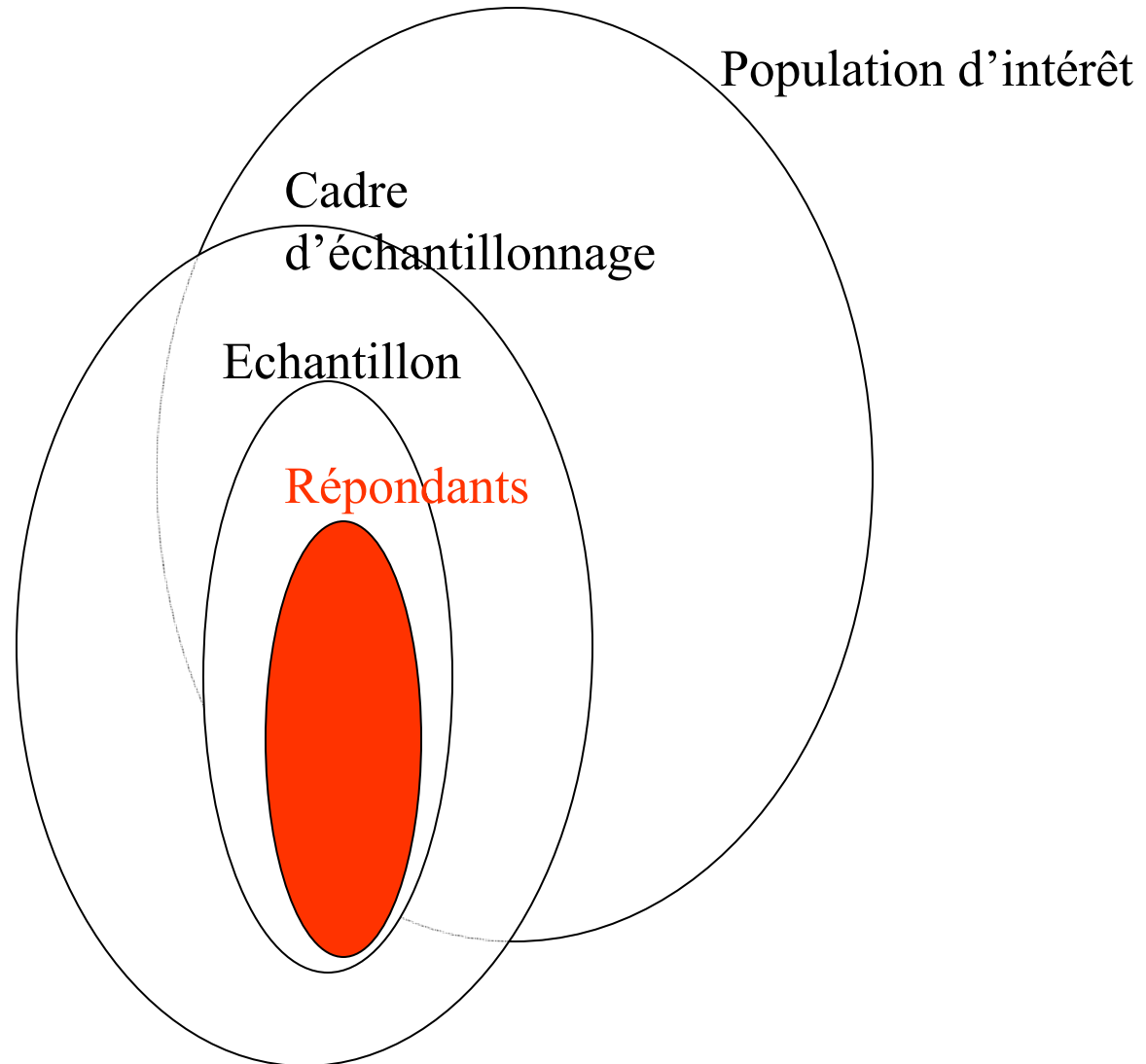


- Seulement si...
  - L'enquête générale sur les ménages est représentative des groupes de bénéficiaires et des groupes témoins (représentativité)
  - Il y a suffisamment d'observations dans les groupes de bénéficiaires et de témoins. (puissance)
  
- Exemple: Nous évaluons l'effet d'une campagne d'information pour les mères sur l'alimentation des enfants de moins d'un an, 10 districts témoins, 10 districts de bénéficiaires. Pouvons-nous utiliser la DHS (enquête démographique et de santé) comme enquête de base (baseline) ?
  - Beaucoup de femmes dans la DHS n'auront pas un enfant âgé de moins d'1 an → nous perdons beaucoup d'observations dans l'échantillon
  - La DHS n'est pas représentative au niveau de district.

# Définitions

- ❑ **Unité d'analyse** : Le type d'entité pour laquelle nous voulons rassembler l'information (personnes, villages, écoles,...)
- ❑ **Population**: L'ensemble d'unités d'analyse pour lesquelles nous voulons savoir si l'intervention fonctionne ou pas
- ❑ **Cadre d'échantillonnage** : La liste physique d'unités d'analyse de laquelle nous tirerons l'échantillon (Quel est le scénario du meilleur cas?)
- ❑ **Echantillon** : Les unités d'analyse que nous tirons du cadre d'échantillonnage et sur lesquelles nous voulons collecter l'information
- ❑ **Répondants** : Les unités d'analyse pour lesquelles nous obtenons des données (ex. les personnes qui répondent aux questions du sondage)
- ❑ **Taux de réponse** : pourcentage d'unités d'analyse dans l'échantillon, pour lesquelles nous collectons l'information (ex., qui répondent aux questions du sondage)

# Graphiquement...



# Le taux de réponse ...

- ❑ Scénario du meilleur cas : c'est 100 %
- ❑ Quand le taux de réponse est faible, il y a un risque de biais de sélection, par exemple quand le groupe témoin ne veut pas répondre à l'enquête.
  - « biais de non réponse »
- ❑ Méthodes empiriques :
  - Ne pas croire aux résultats si le taux de réponse est inférieur à 70 % pour le groupe de bénéficiaires ou le groupe témoin
  - Il ne devrait pas y avoir de différences significatives dans les taux de réponse entre le groupe de bénéficiaires et le groupe témoin
- ❑ Tout cela c'est bien beau.... Mais que faisons-nous quand le taux de réponse est faible ???

# Quand le taux de réponse est faible ...

- Pouvons nous « augmenter la taille de l'échantillon à l'avance parce que nous savons que certaines personnes ne vont pas répondre » ?
  - Cela ne marchera pas parce que tout biais de non réponse sera également présent dans un échantillon plus grand.
- Meilleures solutions :
  - Des efforts supplémentaires pour collecter l'information sur les non répondants (e.g. 3 visites à domicile au lieu d'une)
  - Sous-échantillonnage des non répondants pour vérifier s'ils sont différents des répondants.
  - Dans une enquête de suivi : Vérifier les caractéristiques de base /départ des répondants et des non répondants pour vérifier si elles sont similaires.
- Impératif: tout rapport d'évaluation reçu d'un consultant doit faire état du taux de réponse obtenu!

# Méthode d'échantillonnage

- Echantillonnage aléatoire (choix au hasard)
  - Simple : chaque unité dans le cadre d'échantillonnage a la même probabilité d'être sélectionnée dans l'échantillon
  - Stratifié : nous divisons d'abord le cadre d'échantillonnage en strates (groupes), et ensuite nous faisons un simple échantillon aléatoire dans chaque strate
  - Par grappe/ conglomérat : nous échantillonnons des grappes d'unités : e.g., des villages avec toutes les personnes qui y habitent
  - Stades multiples – une combinaison de deux ou plus méthodes d'échantillonnage aléatoire de ci-dessus, par exemple : échantillon aléatoire stratifié de villages, ensuite simple échantillon aléatoire de personnes dans les villages échantillonnés
- Echantillonnage choisi à dessein (systématique)
  - Ex : Etudes de cas : attention : nous ne pouvons pas généraliser les résultats de ces échantillons

# L'effet d'un échantillonnage par grappe/ à stades multiples sur la puissance des tests

- ❑ Lorsque nous échantillons par grappe, cela affecte la puissance du test de signification.
  - ❑ Exemple : estimons la taille des enfants scolarisés de 7 ans en Egypte
    - **Première méthode**: enquête sur 200 enfants, choisis au hasard parmi tous les enfants scolarisés de 7 ans en Egypte
    - **Deuxième méthode** : choisir d'abord au hasard 20 écoles parmi toutes les écoles égyptiennes, ensuite choisir au hasard 10 enfants dans chacune de ces 10 écoles
- Quel échantillon nous donne le plus d'information ??

# Le tirage par grappe change la variance de l'indicateur

- ❑ Les personnes au sein d'une grappe tendent à être plus similaires l'une à l'autre que les personnes dans un autre groupe :
  - 20 enfants dans la même école tendent à venir d'origines plus similaires que 20 enfants d'écoles différentes
- ❑ Par conséquent : l'écart type sur la taille moyenne de 200 enfants dans 20 écoles (échantillonnage aléatoire stratifié) est plus grand que si nous choisissons 200 enfants au hasard (simple échantillonnage aléatoire)

$$Var_{Clustered} = Var_{SRS} * \underbrace{(1 + \rho * (k - 1))}_{\text{design effect}}$$

- $\rho$  est une mesure d'homogénéité entre les unités (enfants) appartenant à la même grappe (école)
- $k$ =nombre d'unités (enfants) échantillonnées dans chaque groupe

# Qu'est-ce que cela veut dire pour la taille de l'échantillon?

- ❑ Plus grande variance → puissance plus faible pour une taille donnée d'échantillon et erreur de type I (alpha)!
- ❑ La perte de puissance due au tirage par grappe peut être compensée en augmentant la taille globale de l'échantillon
- ❑ De combien devons-nous augmenter la taille de l'échantillon quand nous faisons un tirage par grappe ?

# Déterminer la taille de l'échantillon quand il y a tirage par grappe

- Etape 1 : utiliser des méthodes standards pour déterminer l'échantillon en assumant un simple échantillonnage aléatoire.  $n_{SRS}$  (voir la première partie de la présentation)
- Etape 2 : utiliser la formule suivante pour ajuster la taille de l'échantillon pour les groupes (cluster=grappe)

$$\begin{aligned}n_{Cluster} &= n_{SRS} * \text{design effect} \\ &= n_{SRS} * (1 + \rho * (k - 1))\end{aligned}$$

- Vous aurez à estimer  $\rho$  des enquêtes existantes...
- Utilisez Stata!

# Quand nous nous préoccupons *vraiment* de la taille de l'échantillon ?



- ❑ Quand nous avons de très petits échantillons *au niveau où nous faisons le tirage au hasard !*
  - P.ex. les bénéficiaires dans 20 écoles bénéficiaires et les témoins dans 20 écoles témoin
    - ❑ Mais il y a 400 enfants dans chaque école
  - Ceci reste *malgré tout* un petit échantillon parce que l'attribution au groupe témoin/ traitement ce fait au niveau de l'école
- ❑ Quand notre évaluation a un plan expérimental intercroisé
  - La taille de l'échantillon nécessaire augmente de manière exponentielle lorsque nous augmentons le nombre de traitement intercroisés
- ❑ Quand nous utilisons la discontinuité de la régression comme notre stratégie d'évaluation

# Conclusions

- ❑ Malheureusement dans de nombreux cas nous NE pouvons PAS utiliser les enquêtes générales comme les LSMS et DHS comme enquête de base.
- ❑ Le cadre d'échantillonnage et la méthode de sélection de l'échantillon déterminent la représentativité de l'échantillon pour la population d'intérêt.
- ❑ Quand nous voulons mesurer de très petites différences dans les indicateurs entre le groupe bénéficiaire et le groupe témoin, nous avons besoin d'un échantillon très grand.
- ❑ Quand il y a des grappes dans l'échantillonnage, nous avons besoin d'un échantillon plus grand !