

# **Echantillonnage aux fins d'évaluation**

**Mattias Lundberg, Banque Mondiale  
Atelier Regional du SIEF sur l'Evaluation d'Impact  
Accra, Ghana  
Mai 2010**

**Adapté des diapositives de Esther Duflo et Jed Friedman**

# Plan

---

- **Pourquoi faire un échantillonnage ?**
- **Cadre d' échantillonnage et méthodes d' échantillonnage**
- **Test d'hypothèse**
- **Taille et puissance de l' échantillon**
- **Questions qui affectent la puissance**
- **Exemple en utilisant le logiciel OD**

# Pourquoi utiliser des échantillons ?

---

- **Parce que nous ne sommes pas en mesure de cloner.**
- **Parce que nous n'avons pas assez d'argent ou de temps pour étudier toute la population.**
- **Parce que nous voulons comprendre l'impact d'un programme sur une population réelle.**

# Qu'est ce que nous voulons savoir?

---

- **Les expériences sont coûteuses. Est-ce que nous avons la chance d'identifier un impact?**
- **Quelle doit être la taille d'un échantillon en vue de pouvoir déceler un impact?**
- **Comment les différentes conceptions d'expérience influencent notre chance de trouver un impact ?**

# Cadre d'échantillonnage

- Commencer par la population concernée (le groupe sur lequel vous voulez avoir des connaissances ou auquel vous voulez appliquer des leçons générales):
  - La population en entier
  - Les ménages pauvres
  - Les jeunes femmes
- Définir une sous-population que pouvez étudier en principe (celles qui vous sont accessibles)
  - Les jeunes femmes dans la région x à un moment t...
- Dresser la liste de ceux que vous pouvez réellement contacter
  - Noms et informations de contact

***Voici votre cadre d'échantillon*** →

***Notez qu'il est probable que cet échantillon ne représente pas exactement la population concernée***

# Le cadre d'échantillonnage

---

- **Des listes récentes et toutes faites constituent un cadre d'échantillonnage idéal :**
  - Recensement
  - Annuaire de téléphone
  - Listes électorales

**Et s'il arrive que vous n'avez pas l'un de ces éléments**

- **Effectuez une énumération initiale – allez dans votre zone d'étude et dressez vos propres listes .**
- **Ou évitez de le faire – et vous aurez les conséquences .**

# Le cadre d'échantillonnage

Quelles sont donc les conséquences? En principe:

- Avec un cadre d'échantillonnage, vous pouvez tirer des leçons générales pour la population en entier;
- Sans le cadre d'échantillon, vous ne pourrez pas généraliser au delà de la population échantillon.
- ***NB: En tout cas, parfois nous le faisons.***

Par exemple: un programme de formation fait une grande publicité et organise une session d'information pour recruter des participants. La population de l'étude comprend ceux qui ont décidé de participer à cette session.

Nous devons tirer des leçons seulement au titre de ceux qui se sont rendus à la session d'information mais parfois nous tirons des conclusions plus élargies .

# Echantillonnage

## Choisir un échantillon dans un cadre d'échantillonnage :

- Echantillon avec probabilité ( avec des probabilités de sélection connues )
  - *Echantillon simple* (choisir au hasard d'une population en entier ).
  - *Echantillon systématique* (choisir au hasard et puis chaque nième personne de la liste )
$$\sum_1 n_i = N$$
  - *Echantillon stratifié* (diviser la population en des groupes unique ).



# Echantillonnage

Choisir un échantillon dans un cadre d'échantillonnage:

- **Echantillons sans probabilité**

- ***Commodité*** (ex. clients à une structure sanitaire en une journée)
- ***Intentionnel*** (ex. Populations difficile à atteindre, échantillons très probables, échantillon de contingent; échantillon d'hétérogénéité échantillons de boule de neige)

**Si vous évitez l'échantillonnage de probabilité, pensez à l'impact de vos méthodes de sélection non aléatoires sur vos conclusions et la généralisation de vos résultats**

# Test d'hypothèse

---

- Nous voulons tester l'hypothèse nulle que l'impact d'une intervention est zéro

$$H_0: \textit{impact} = 0;$$

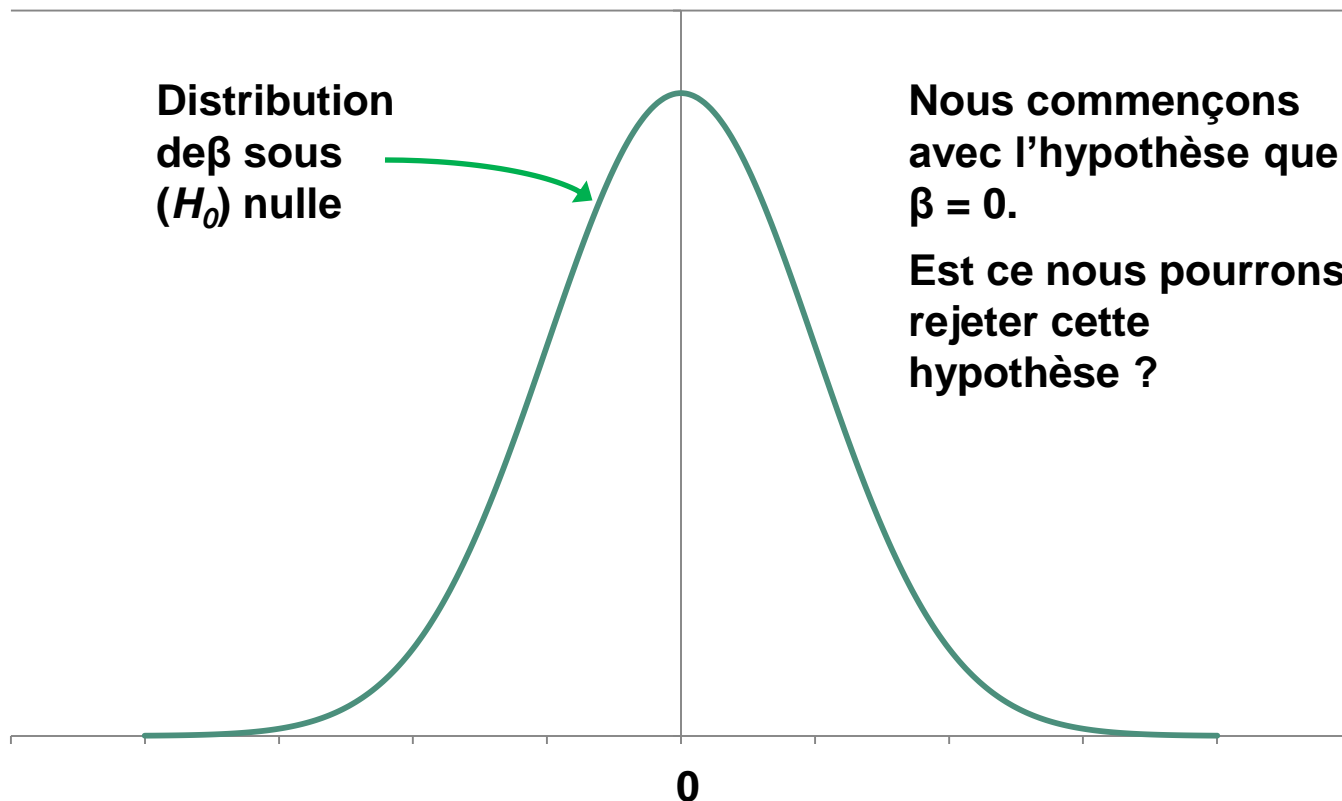
contre l'hypothèse alternative que l'impact de l'intervention n'est pas zéro

$$H_a: \textit{impact} \neq 0.$$

- On y parvient en comparant les résultats moyens d'un groupe traité avec les résultats moyens d'un groupe non traité .

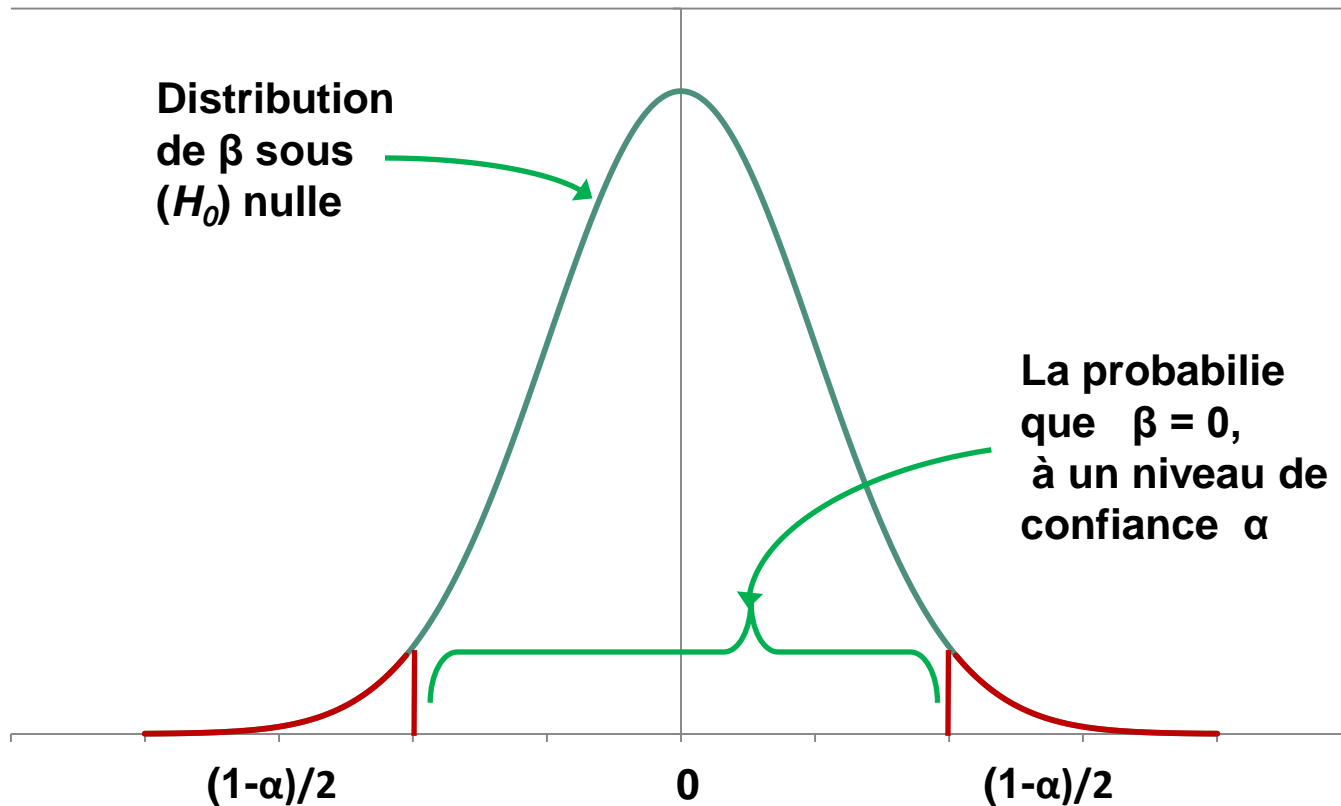
# Test d'hypothèse

## ■ La distribution attendue ( hypothèse nulle )



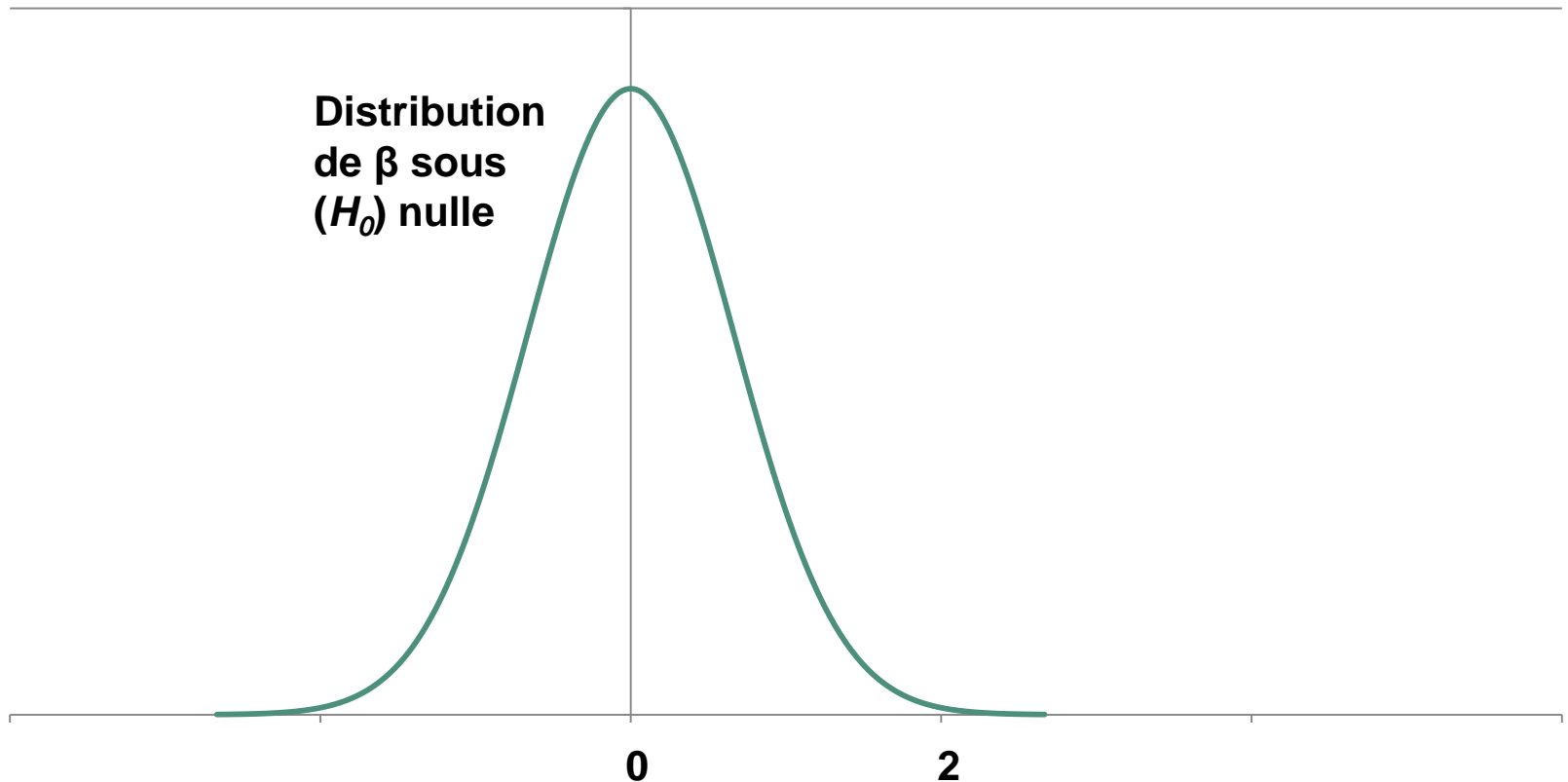
# Test d'hypothèse

- La distribution attendue ( hypothèse nulle )
- )



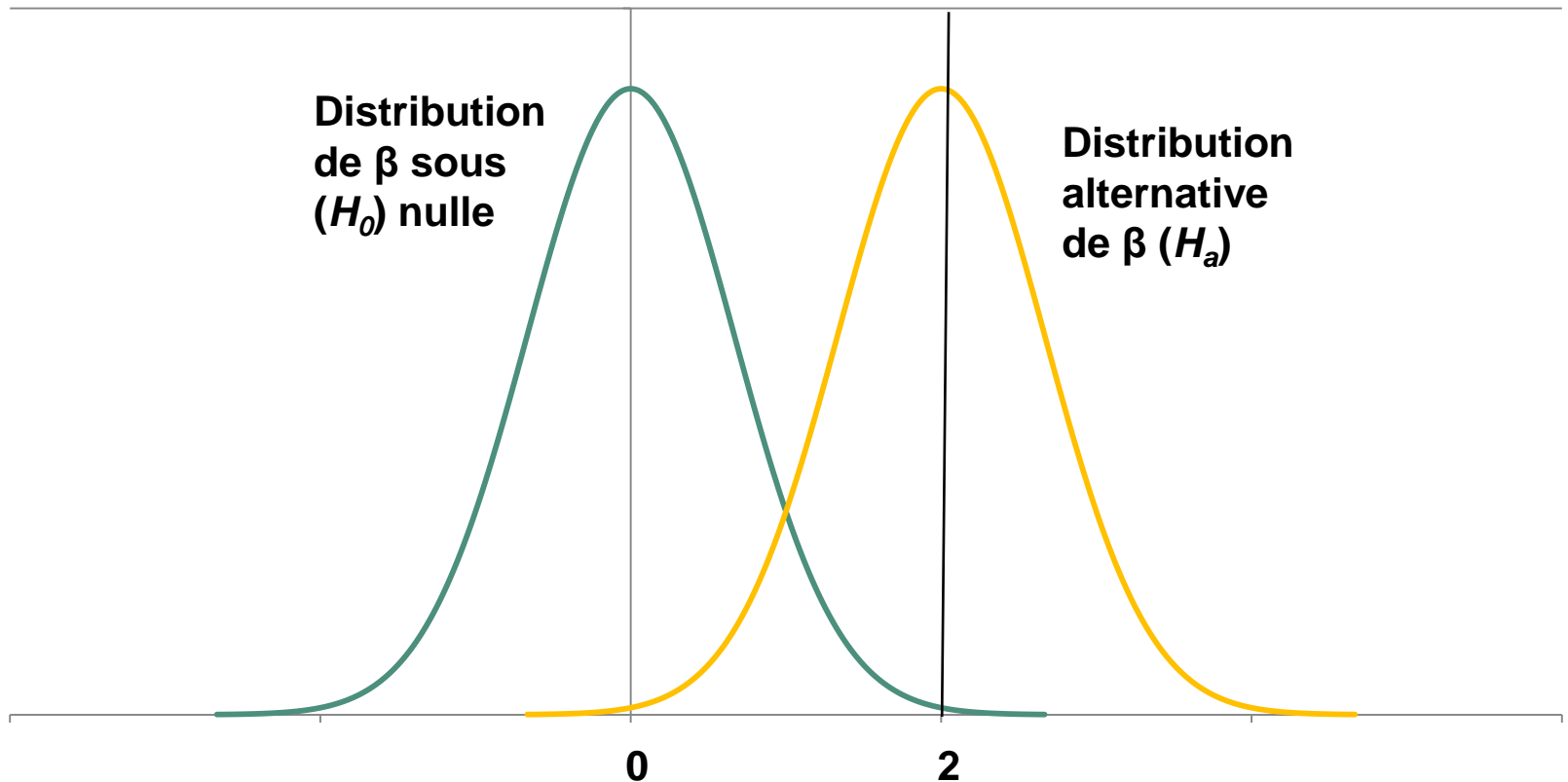
# Test d'hypothèse

## ■ La distribution alternative attendue



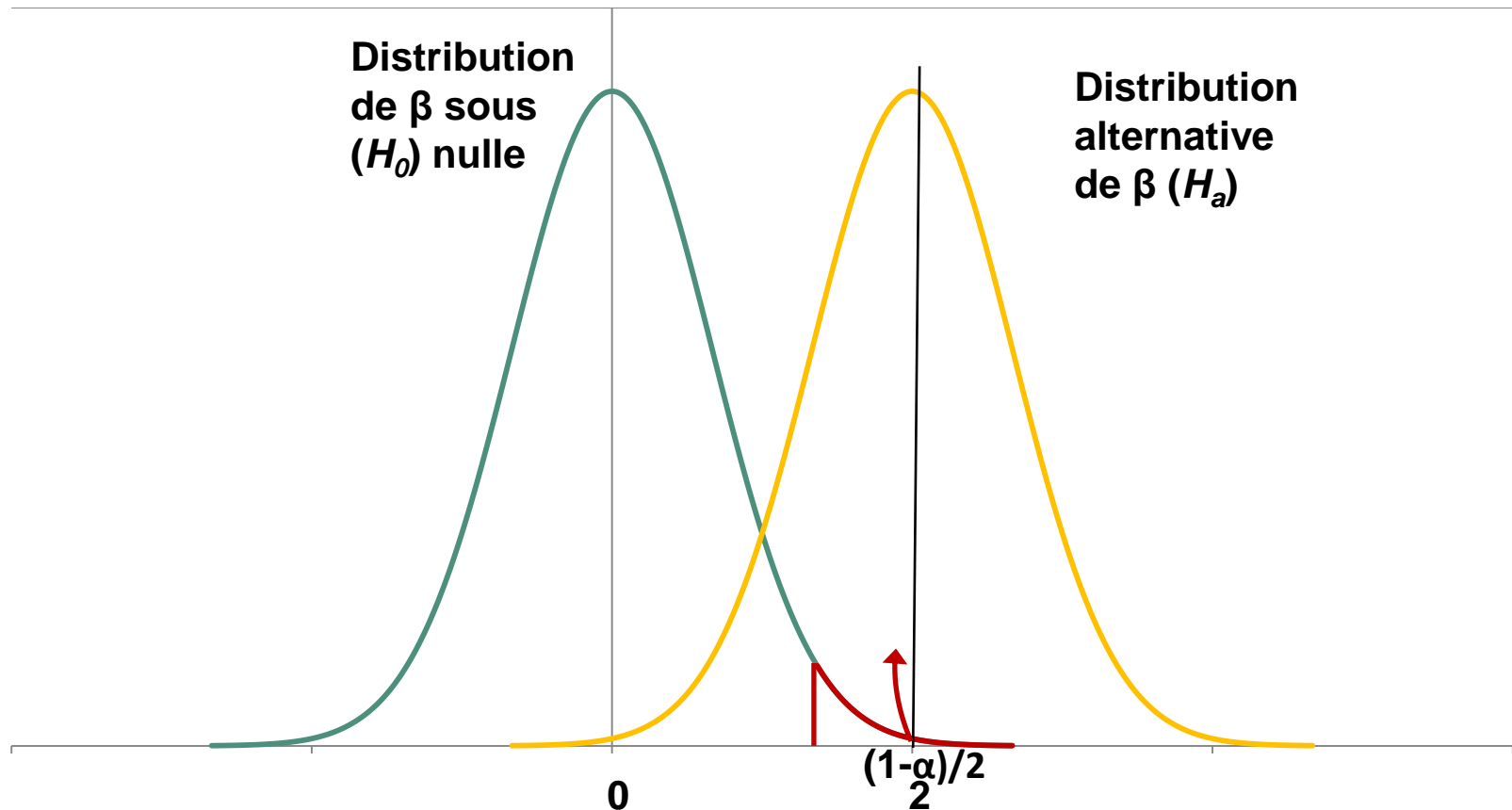
# Test d'hypothèse

## ■ La distribution alternative attendue



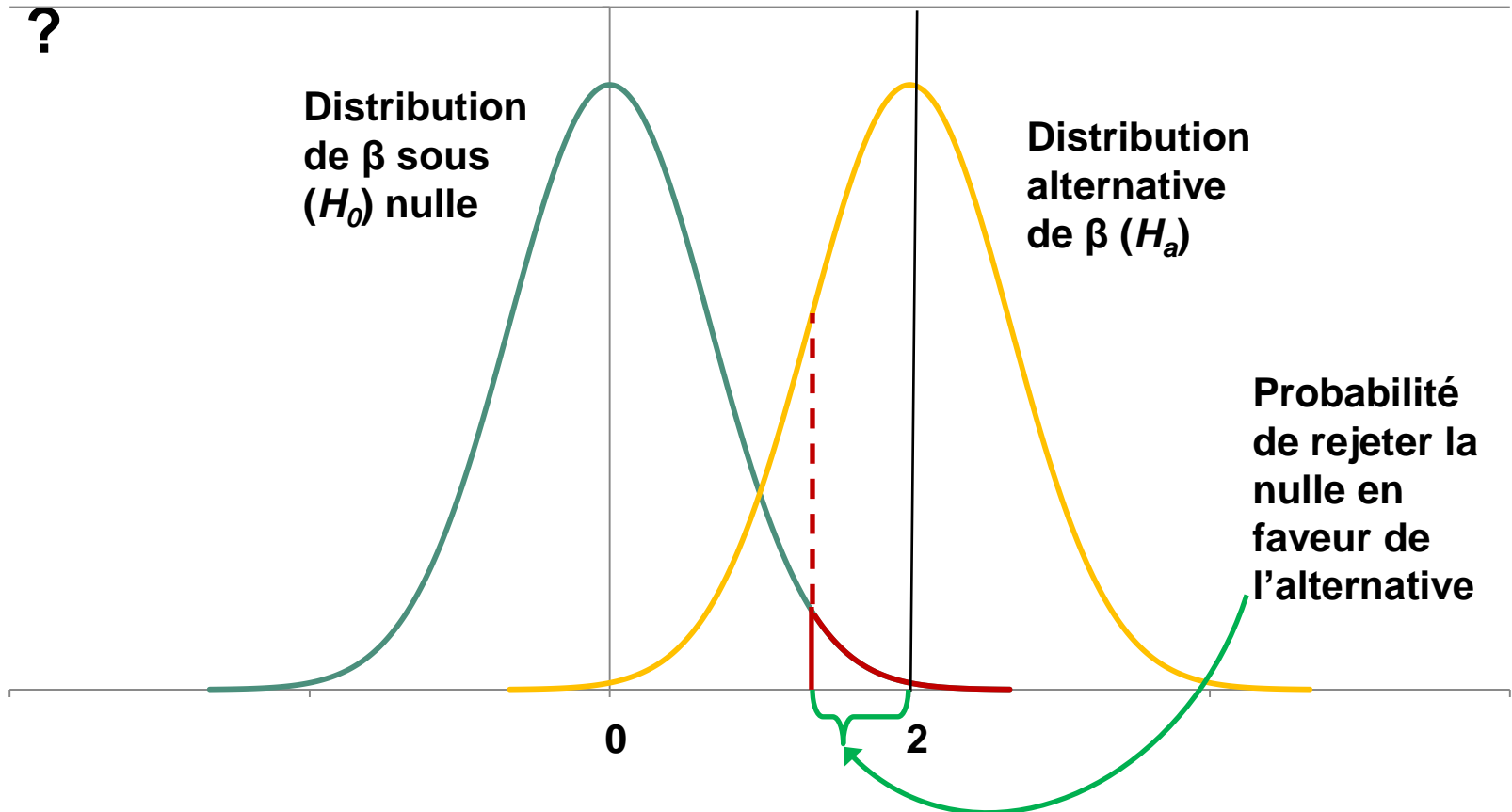
# Test d'hypothèse

- Est ce que  $H_0$  et  $H_a$  sont très différentes du point de vue statistique?



# Test d'hypothèse

- Est ce que  $H_0$  et  $H_a$  sont très différentes du point de vue statistique?
- ?





# Fiabilité, puissance, et deux types d'erreur

- Premier type d'erreur : conclure qu'il y a eu un effet lorsqu'en réalité il n'y pas eu d'effet (*faux positif*).
- Second type d'erreur : conclure qu'il n'y a pas eu d'effet lorsqu'en fait il y a eu d'effet (*faux négatif* )

		Etat du monde	
		$H_0$	$H_a$
Prévision	$H_0$	Acceptation correcte	Acceptation incorrecte (faux négatif, type-II, $\beta$ )
	$H_a$	Rejet incorrecte (faux positif, type-I, $\alpha$ )	Rejet correcte

# Fiabilité, puissance et deux types d'erreur

- ***Fiabilité*** décrit la capacité du test à minimiser les erreurs de type I (*faux positifs*).
- ***La puissance*** décrit la capacité du test à minimiser les erreurs de type II (*faux négatifs*).
- La règle est de se préoccuper des erreurs de type I plus que des erreurs de type II .

(ex, Nous préférons dire par erreur que quelque n'a pas marché lorsqu'elle a en fait marché au lieu de dire que quelque chose a marché lorsqu'en fait elle n'a pas marché)

- Nous voulons avoir une fiabilité d'au moins 90 ou 95 pour cent, mais on est prêt à accepter une puissance de 80 à 90 pour cent .

# Calculer la puissance

---

- Quelles sont les informations dont vous avez besoin?
  - **Taille de l'effet** (la différence minimale acceptable entre les moyennes des groupes traités et des groupes non traités or la différence attendue entre les deux)
  - **Variance de l'échantillon** (Ecart types entre les moyennes des cas traités et les estimations des cas non traités)
  - **Le niveau de fiabilité** acceptable (ex. 95%)
- Avec cette information vous pouvez estimer la taille de l'échantillon nécessaire pour réaliser une certaine puissance ou la puissance que vous pourrez obtenir avec une taille d'échantillon donnée .

# Calculer la puissance

---

**Trouver la taille d'effet et les prévisions de variance :**

- **Choisir le plus faible effet pour justifier l'adoption d'un programme:**
  - **Comparer le cout du programme et la valeur totale de ses avantages .**
- **Utiliser les preuves des études précédentes comparables.**
- **Les petits effets sont difficiles à trouver si le résultat est très variable dans la population ou s'ils ne sont pas bien mesurés .**

# Pourquoi la taille est importante ?

- Imaginez que nous voulons tester l'impact d'un traitement T sur un Résultat Y:

$$Y_i = \alpha + \beta T + \varepsilon_i$$

$\hat{\beta}$  est la mesure de l'effet moyen d'un traitement:

$$\hat{\beta} = \bar{Y}_T - \bar{Y}_0$$

- La variance de l'estimation de  $\hat{\beta}$  est:

$$\text{var}(\hat{\beta}) = \frac{1}{P(1-P)} \frac{\sigma^2}{N}$$

*P* étant la proportion traitée et *N* la taille de l'échantillon

# Pourquoi la taille est importante ?

$$\text{var}(\hat{\beta}) = \frac{1}{P(1-P)} \frac{\sigma^2}{N}$$

$\text{var}(\hat{\beta})$  diminue alors  $N$  augmente  $\sigma^2$  décroît  
, et  $P \rightarrow 0.5$ .

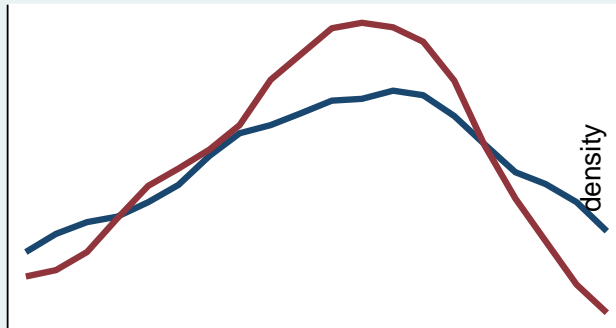
- Il est plus probable de trouver un effet **important** avec un grand  $\hat{\beta}$  et un effet moindre avec  $\text{var}(\hat{\beta})$  :

$$\frac{\hat{\beta}}{\sqrt{\text{var}(\hat{\beta})}} \sim t_n$$

(the familiar  
 $t$ -statistic)

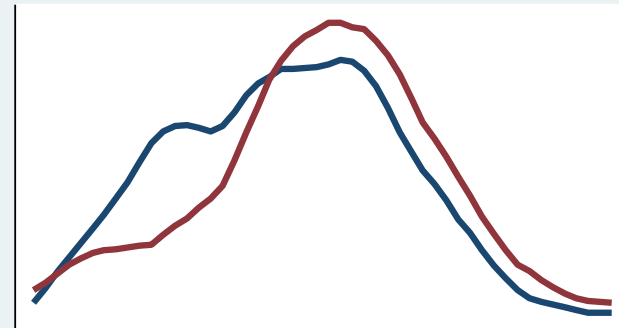
# Puissance et taille de l'échantillon

N = 10, t-test = .3, prob (.768)



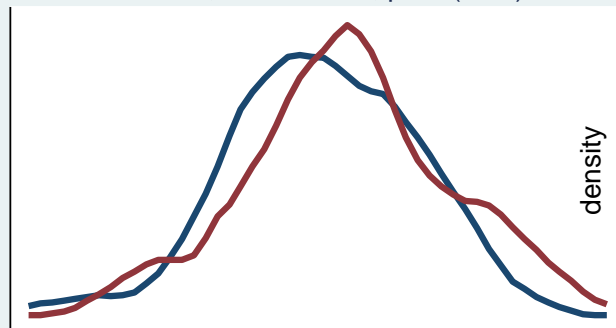
sample statistics: 107.99 (76.19), 98.54000000000001 (64.34)

N = 50, t-test = 1.71, prob (.09)



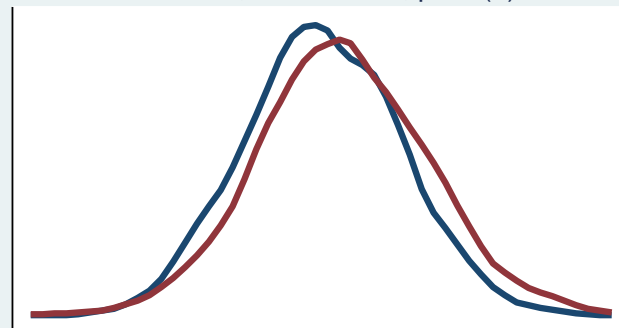
sample statistics: 104.73 (46.23), 121.1 (49.33)

N = 100, t-test = 1.68, prob (.095)



sample statistics: 99.89 (52.3), 112.91 (57.43)

N = 1000, t-test = 5.67, prob (0)

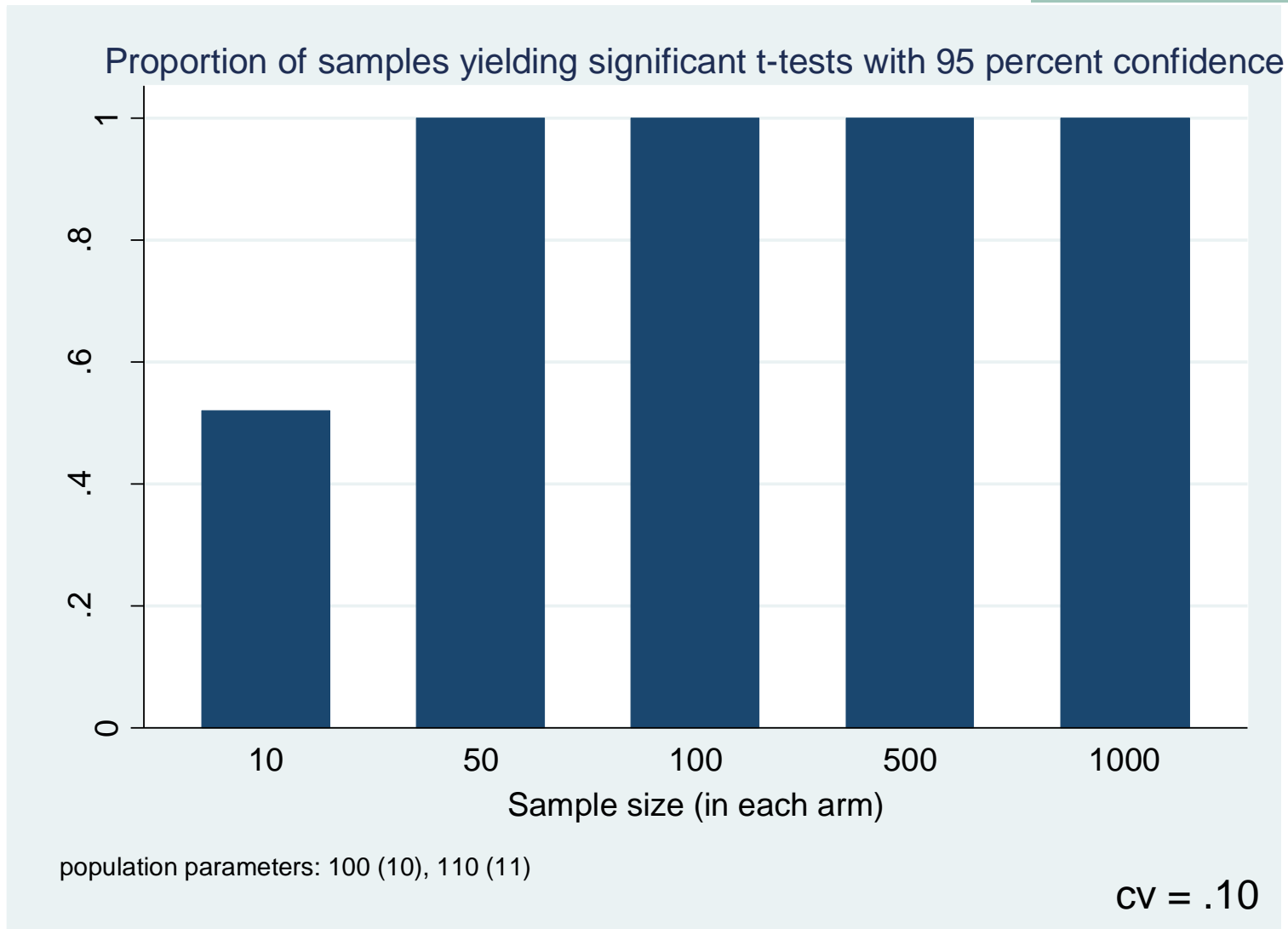


sample statistics: 98.82000000000001 (50.96), 112.34 (55.61)

population parameters: 100 (50), 110 (55)

# Puissance et variance

## Résultats mesurés avec précision

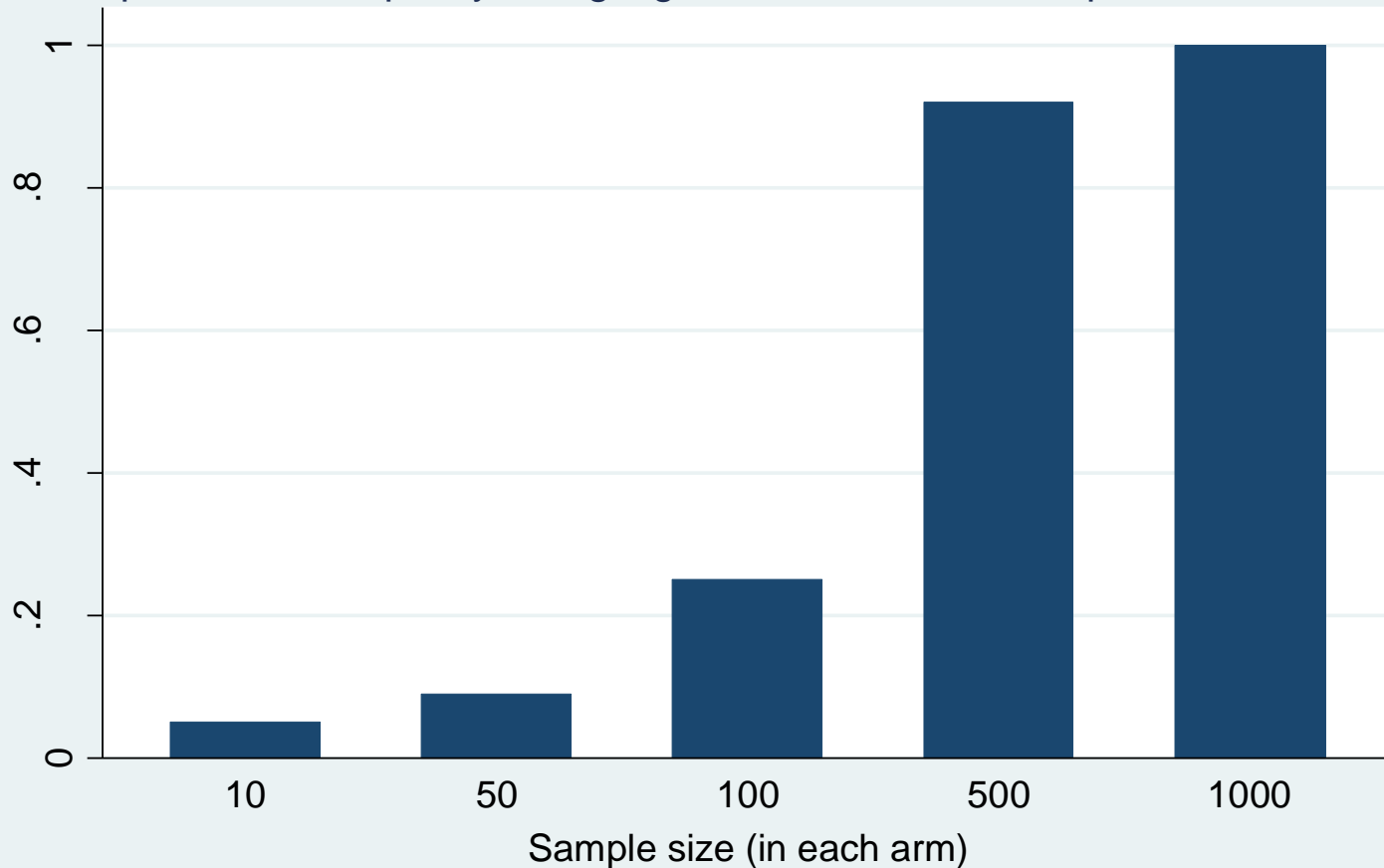




# Puissance et variance

## Résultats mesurés sans précision

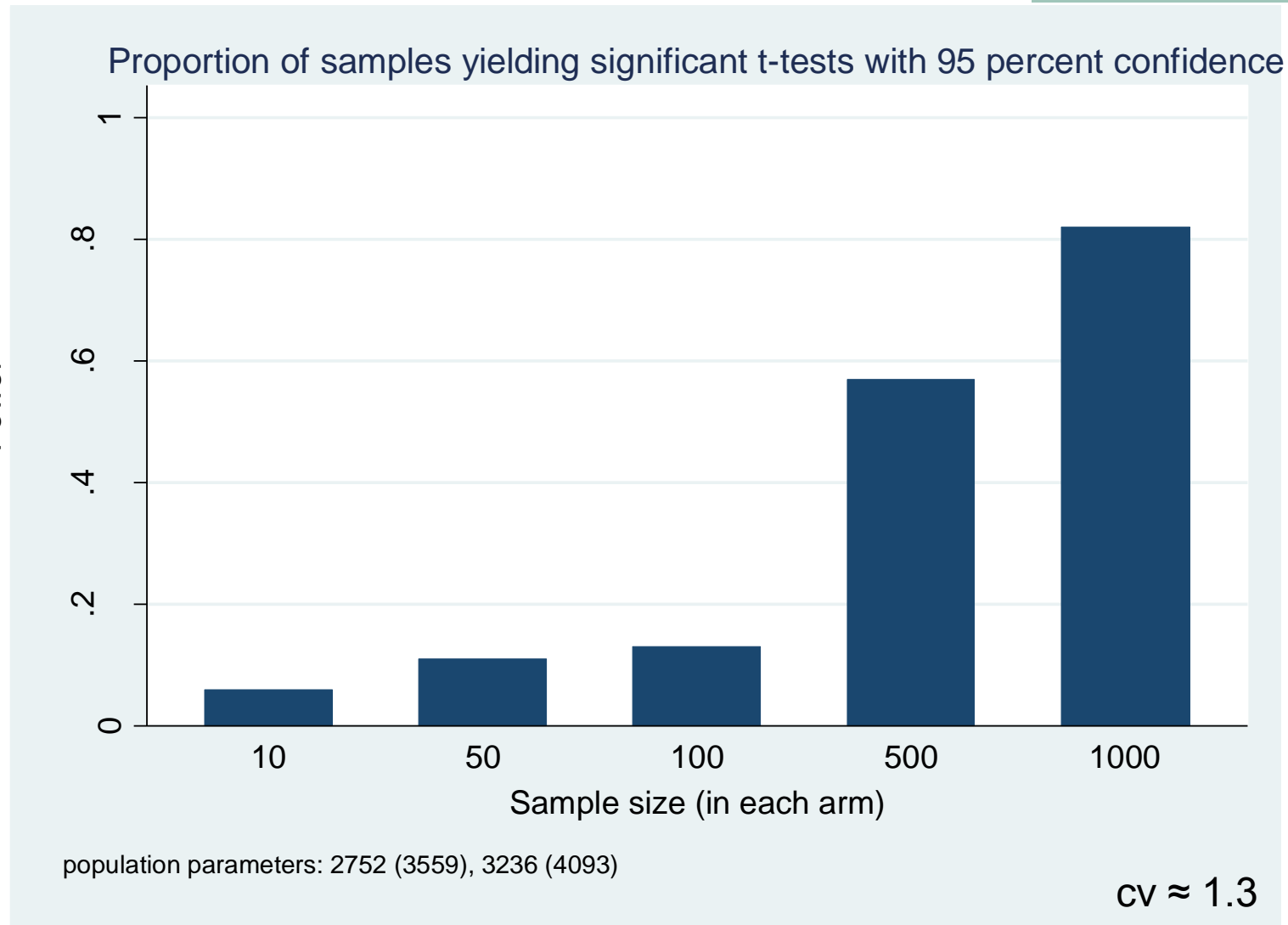
Proportion of samples yielding significant t-tests with 95 percent confidence



population parameters: 100 (50), 110 (55)

cv = .50

# Puissance et variance en utilisant les paramètres de la fiche (2007)



# **Puissance et variance en utilisant les paramètres de la fiche (2007)**

---

- **La fiche d'échantillon a 563 dans le groupe de contrôle et 786 dans le groupe de traitement – un ratio d'environ 1,4.**
- **Les résultats moyens étaient 2752 (3559) dans le groupe de contrôle et 3236 (4093) dans le groupe de traitement**
- **Ils ont eu la chance de trouver une différence.**
- **En vue d'assurer une mesure avec 95% de fiabilité , 80% de puissance, ils doivent avoir 826 dans le groupe de contrôle et 1157 dans le groupe de traitement (soit une augmentation de 50% ).**

# Echantillonnage de groupe

---

Les essais randomisés de groupe sont des expériences dans lesquelles des unités sociales ou des groupes plutôt que des individus sont assignés au hasard à des groupes d'intervention.

## Exemples:

---

<b>Intervention</b>	<b>Groupe</b>
<b>Transferts de fonds sous condition</b>	<b>Villages</b>
<b>Distribution des moustiquaires</b>	<b>Cliniques</b>
<b>Gestion des communautés</b>	<b>Schools</b>
<b>Soutien social</b>	<b>Familles</b>

---

# Pourquoi considérer l'usage de l'échantillonnage de groupe ?

---

- **Débordements, contagion ou contamination**
  - Une intervention affecte le groupe en entier (ex. la formation de l'enseignant affecte toute la classe, un programme de développement affecte tout le village )
  - Dans le cadre d'un programme antihelminthique, les écoles ont été choisies comme unité parce que les vers sont contagieux .
- **Considérations politiques ou sociales**
  - Le programme PROGRESA ne pourrait pas être mis en œuvre si quelques villages y étaient exposés et d'autres ne le l'était pas. .
  - Le programme de réintégration des anciens combattants devrait réduire la probabilité des conflits entre villages

# Implications du regroupement

---

- **Les résultats, les réponses à une intervention au niveau des individus dans un groupe peuvent être en corrélation :**
  - **Tous les villageois sont exposés au même climat.**
  - **Tous les patients partagent le même professionnel de la santé .**
  - **Tous les étudiants ont le même directeur d'école .**
  - **Les résidents du village entretiennent des relations les uns avec les autres.**
- **Les erreurs types prévus doivent être ajustées pour prendre en compte cette corrélation .**
- **La puissance de l'échantillon baisse au fur et à mesure que la corrélation au sein du groupe augmente (parce que chaque individu fournit des informations peu indépendantes)**

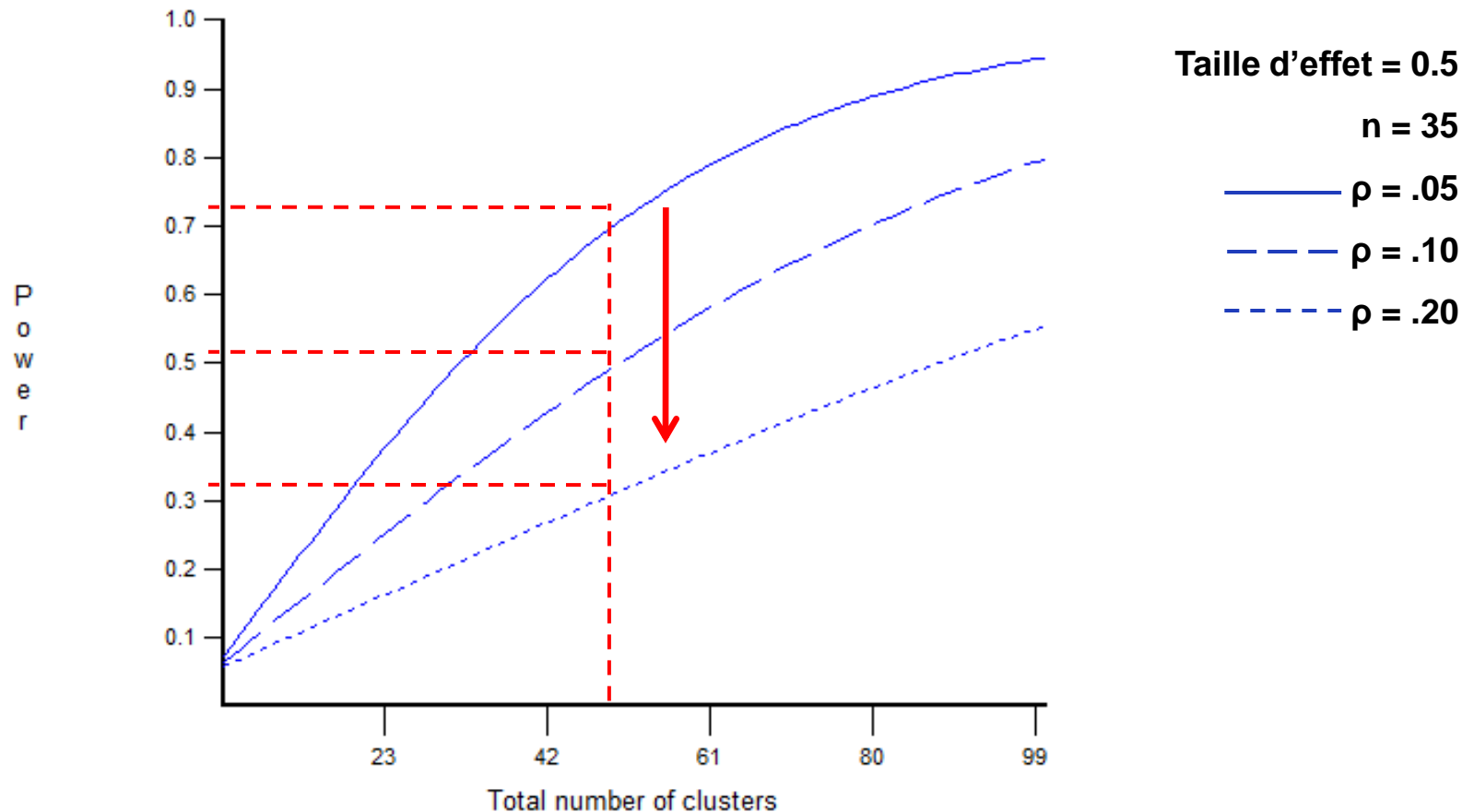
# Implications du regroupement

---

- **Rassurez-vous que vous avez un échantillon assez large .**
- **L'échantillon primaire est le nombre d'unités (les population, les familles, les écoles) au niveau où l'assignation au hasard se déroule. C'est différent du nombre de personnes enquêtées**
- **Par exemple, vous ne pouvez pas randomiser au niveau du district, avec un échantillon d'un district traité et un district de contrôle, même si chaque district a une population de 1000 personnes.**
- **Le nombre d'individus au sein d'un groupe n'est pas aussi important que le nombre de groupes (à cause des corrélations au sein du groupe)**

# La corrélation au sein du groupe réduit la puissance

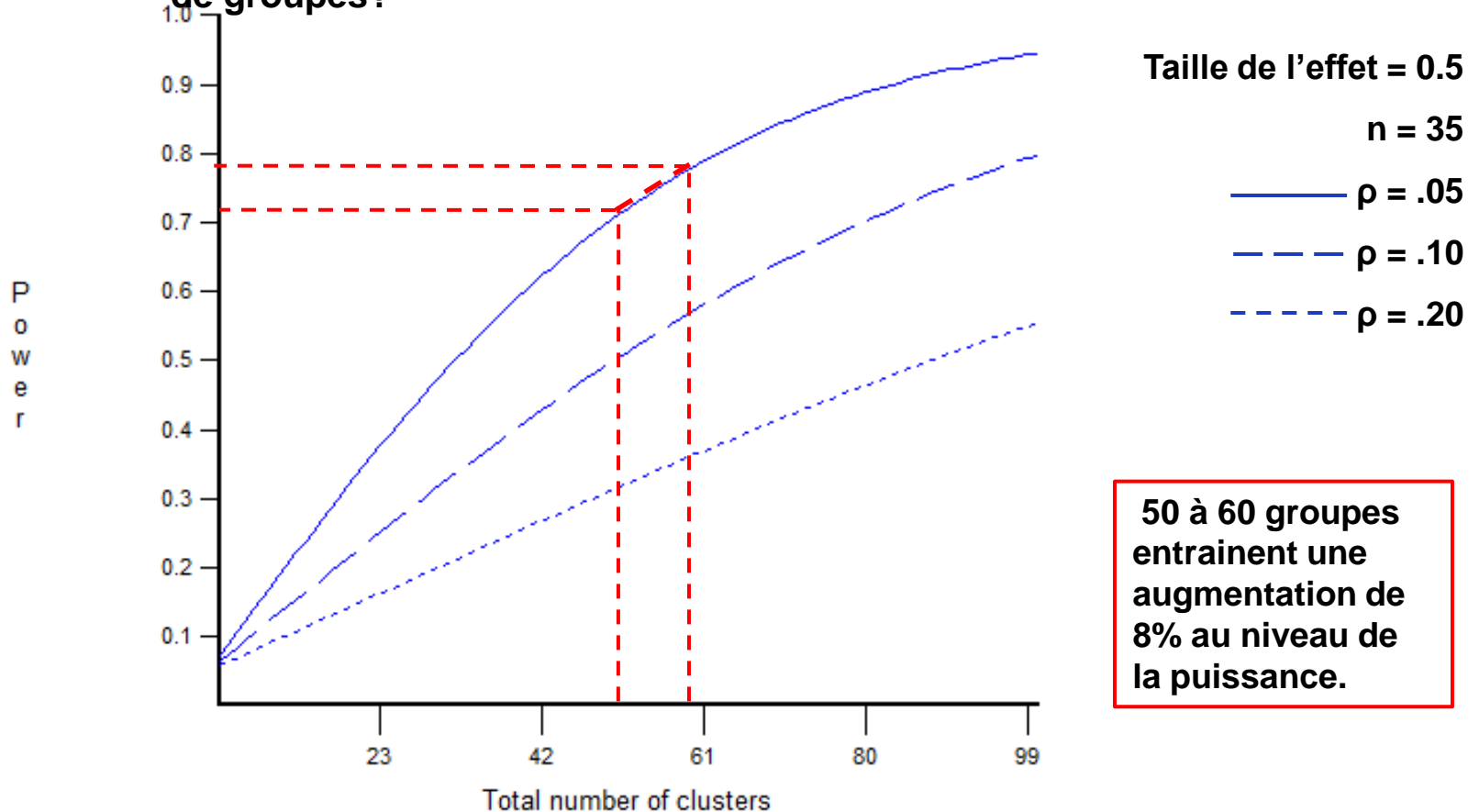
- A mesure qu'une corrélation au sein d'un groupe augmente, chaque constat au sein du groupe fournit des informations peu indépendantes.
- Avec 50 groupes, chacun avec 35 constats:





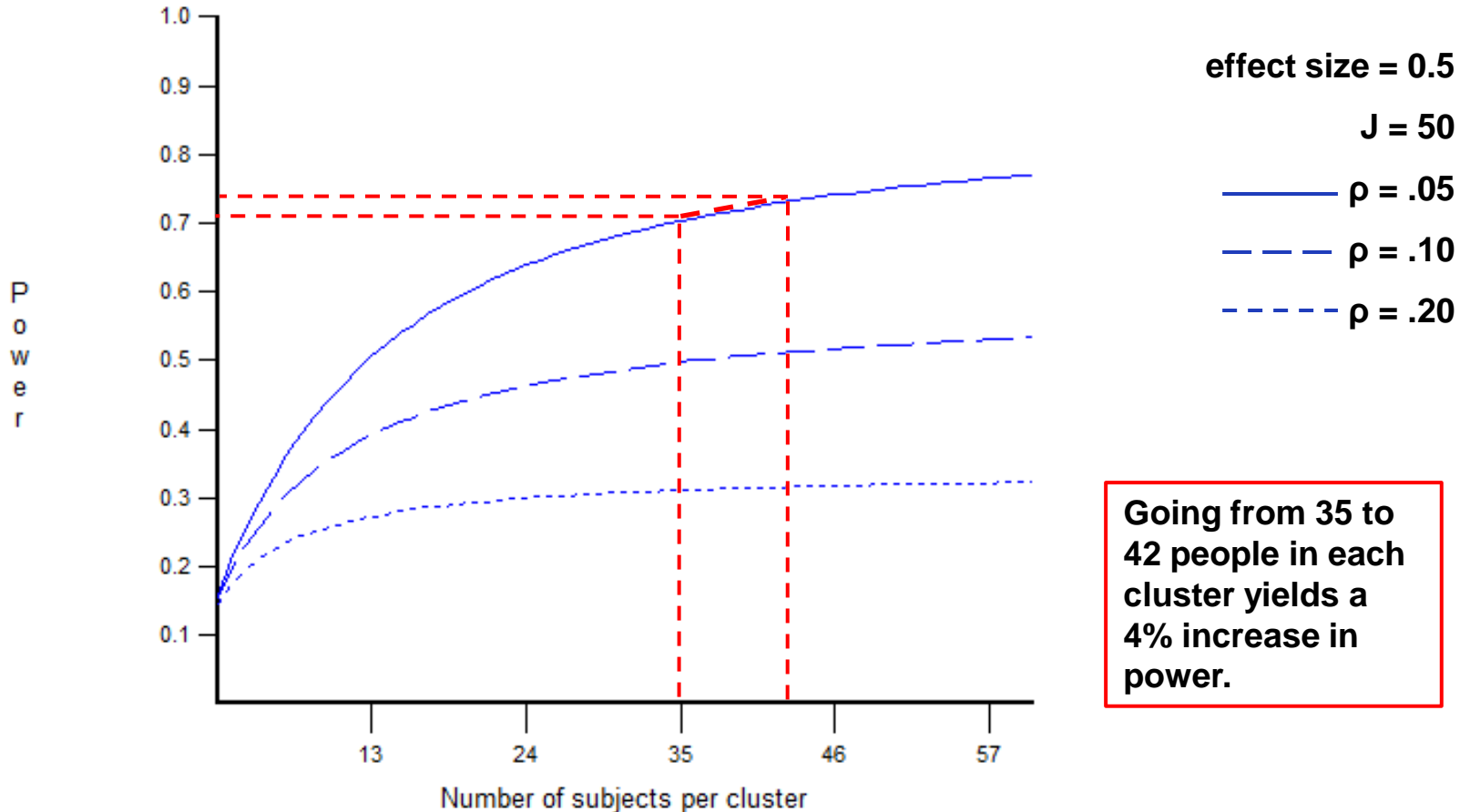
# La puissance peut augmenter rapidement avec le nombre de groupes

- Vous avez 50 groupes avec 35 personnes dans chacun (1750) et vous pouvez augmenter l'échantillon de 20% pour atteindre 2100
- Est-ce que vous augmenterez la taille de chaque groupe ou le nombre de groupes?



# than with the number of units within each cluster

- You have 50 clusters with 35 people in each (1750 people), and you can afford to increase the sample by 20%, to 2100.
- Do you increase the size of each cluster or the number of clusters?



# Echantillonnage avec une enquête de référence

---

- **Avantages d'une référence:**
  - Permet de vérifier si les groupes de traitement et de contrôle étaient les mêmes avant le traitement
  - Permet d'examiner les changements ainsi que les niveaux;
  - Réduit la taille de l'échantillon nécessaire, mais exige une enquête avant le démarrage de l'intervention;
  - Permet de stratifier et de former des sous-groupes .
- **Pour calculer la puissance à partir d'une référence, vous devez savoir la corrélation entre les mesures ultérieures du résultat ( par ex. la consommation mesurée sur période de deux ans)**

# D'autres questions qui affectent la taille et la sélection de l'échantillon

---

- Il y a t-il plus d'un traitement (plus d'une branche d'expérience)?
- Est-ce que vous vous intéressez aux écarts entre les groupes traités?
- Est-ce que vous intéressez aux interactions entre les traitements?
- Est-ce que vous cherchez à savoir si les effets sont différents dans les différentes sous-populations?
- Vous devez calculer la puissance pour chacune de ces comparaisons et non celle de l'intervention en entier.

# D'autres questions qui affectent la taille et la sélection de l'échantillon

---

- Est-ce que votre conception implique seulement une conformité partielle ? (e.g. conception d'encouragement?)
- Que deviendra la validité de votre expérience si quelques personnes abandonnent en chemin?
- Et si l'abandon n'est pas au hasard?
- Et si l'intervention n'est pas mis en œuvre comme prévu ?
- NB: Les calculs de la puissance sont conditionnels et se focalisent sur une variable à la fois; les calculs de la puissance multi variée (impliquant des distributions mixtes) sont extrêmement difficiles.

# Logiciel pour les calculs de la puissance

## De marque déposée

---

Stata	<a href="http://www.stata.com/">http://www.stata.com/</a>
SAS	<a href="http://www.sas.com/">http://www.sas.com/</a>
SPSS	<a href="http://www.spss.com/">http://www.spss.com/</a>
Splus	<a href="http://www.insightful.com/">http://www.insightful.com/</a>

---

## Gratuit

---

PS: Power and Sample Size Calculation	<a href="http://biostat.mc.vanderbilt.edu/twiki/bin/view/Main/PowerSampleSize">http://biostat.mc.vanderbilt.edu/twiki/bin/view/Main/PowerSampleSize</a>
Optimal Design Software	<a href="http://sitemaker.umich.edu/group-based/optimal_design_software">http://sitemaker.umich.edu/group-based/optimal_design_software</a>

---

## Web-based, interactive

---

<http://statpages.org/>

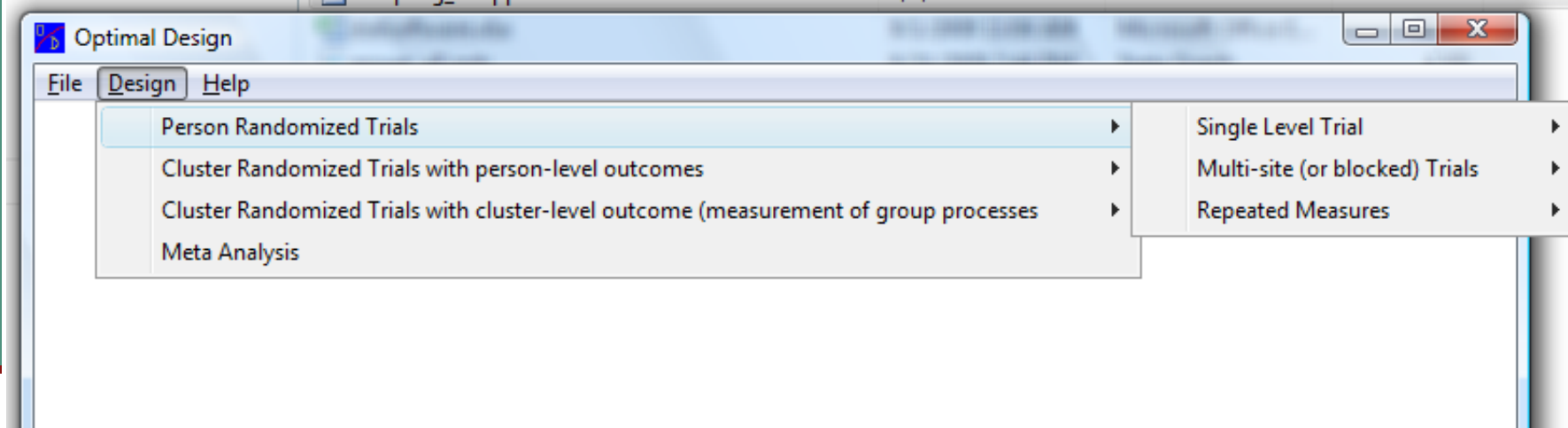
---

[http://www.dssresearch.com/toolkit/spcalc/power\\_a2.asp](http://www.dssresearch.com/toolkit/spcalc/power_a2.asp)

---

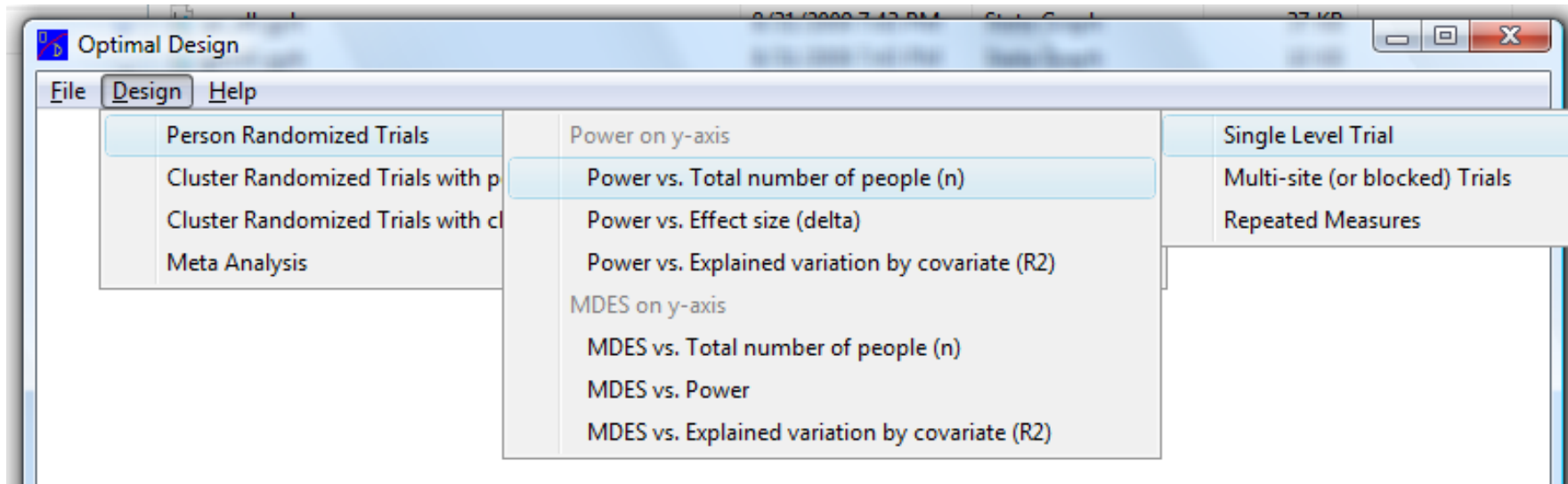
# Exemple avec le logiciel OD (1)

- Pour calculer la taille de l'échantillon nécessaire pour les test au niveau individuel test:



# Exemple avec le logiciel OD (1)

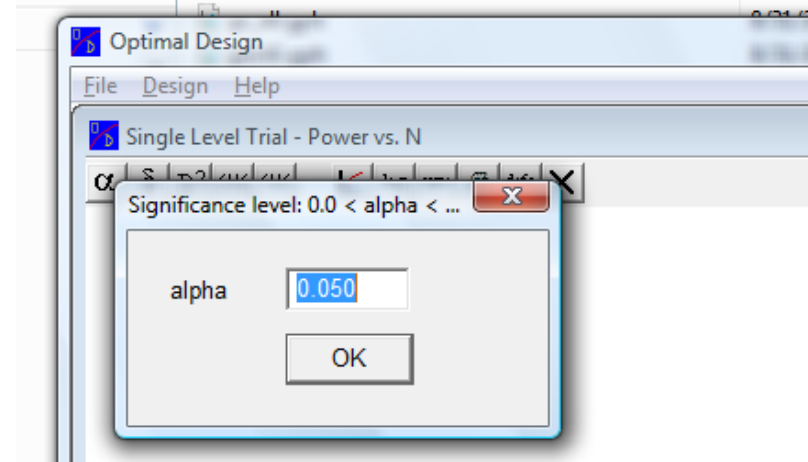
- OD peut exprimer la réponse en fonction de la puissance ou de la taille d'effet minimale décelable .
- Dans ce cas, recherchons la puissance et la taille de l'échantillon à une taille d'effet



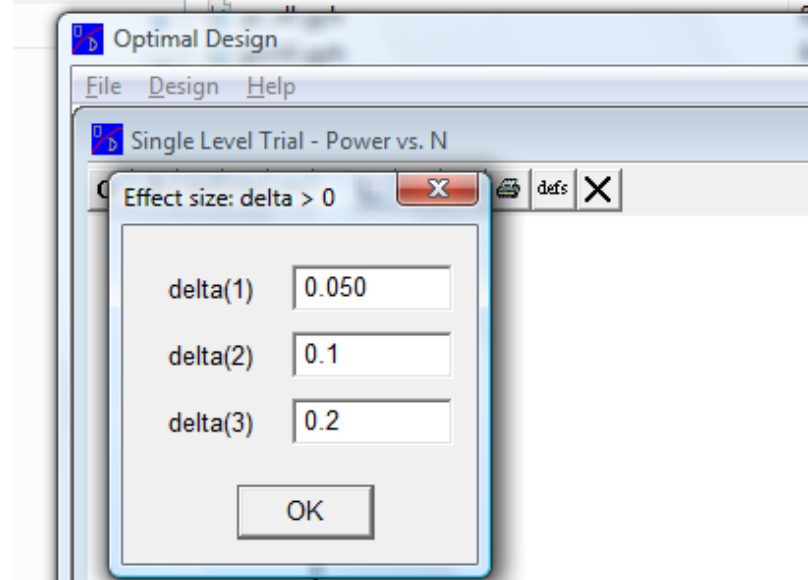


# Exemple avec le logiciel OD (1)

- Choisir en premier lieu le niveau d'importance  $\alpha$  (ex. 5%)



- Puis choisir la taille d'effet minimale acceptable  $\underline{\delta}$  (vous pouvez choisir 3 au maximum)

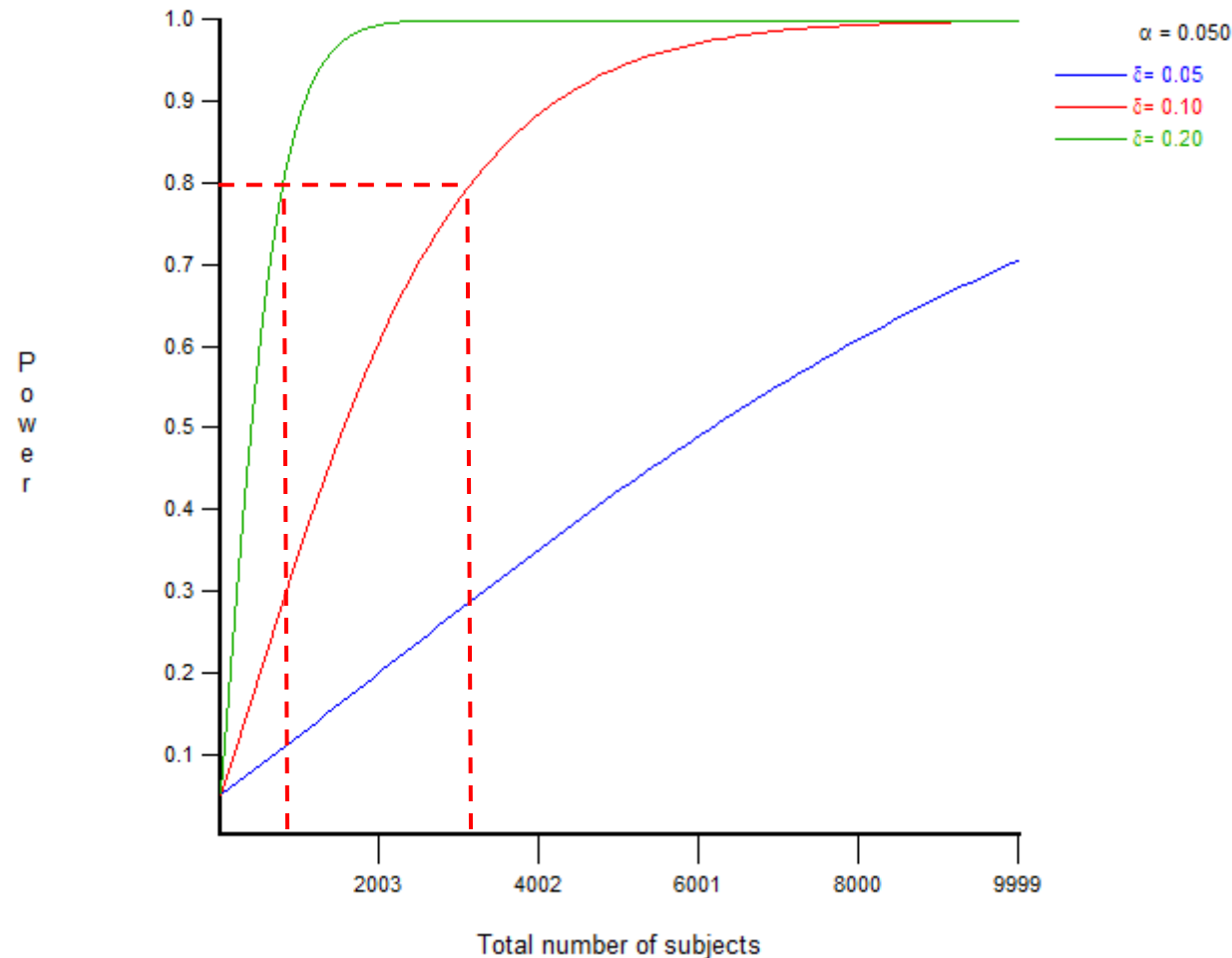


# Exemple avec le logiciel OD (1)

**Avec une taille d'effet de 0,2, une puissance de 80% est atteinte avec un échantillon de 800**

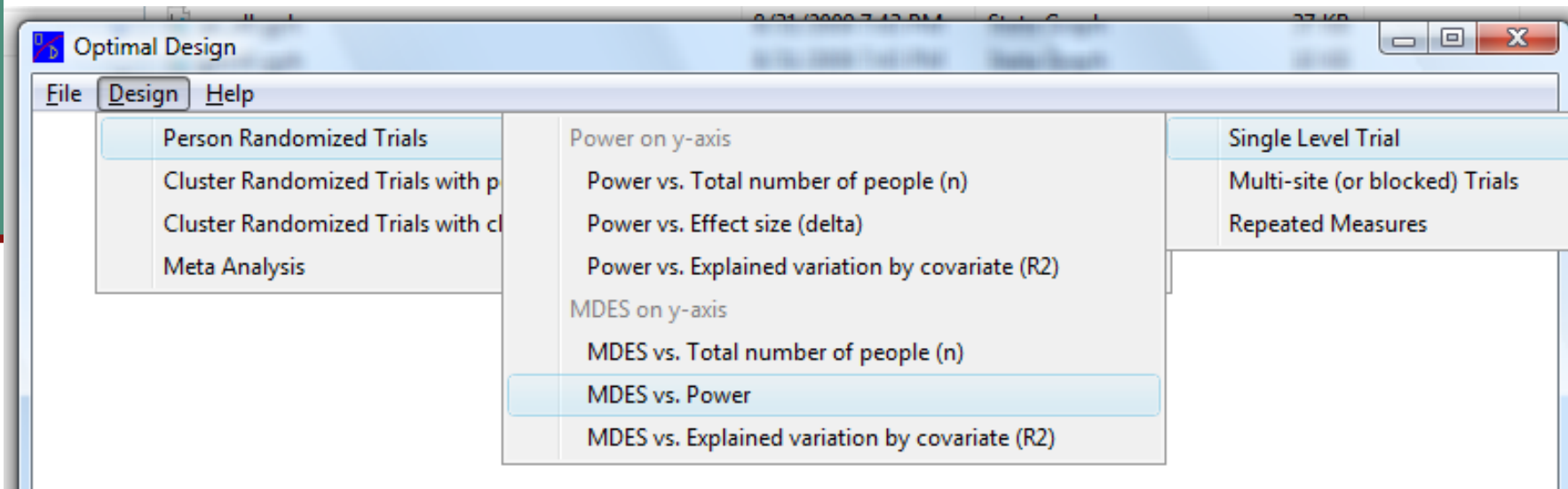
**Avec une taille d'effet de 0.1, vous devez avoir ~3100.**

**Les plus petits effets sont difficiles à vérifier**



# Exemple avec le logiciel OD (2)

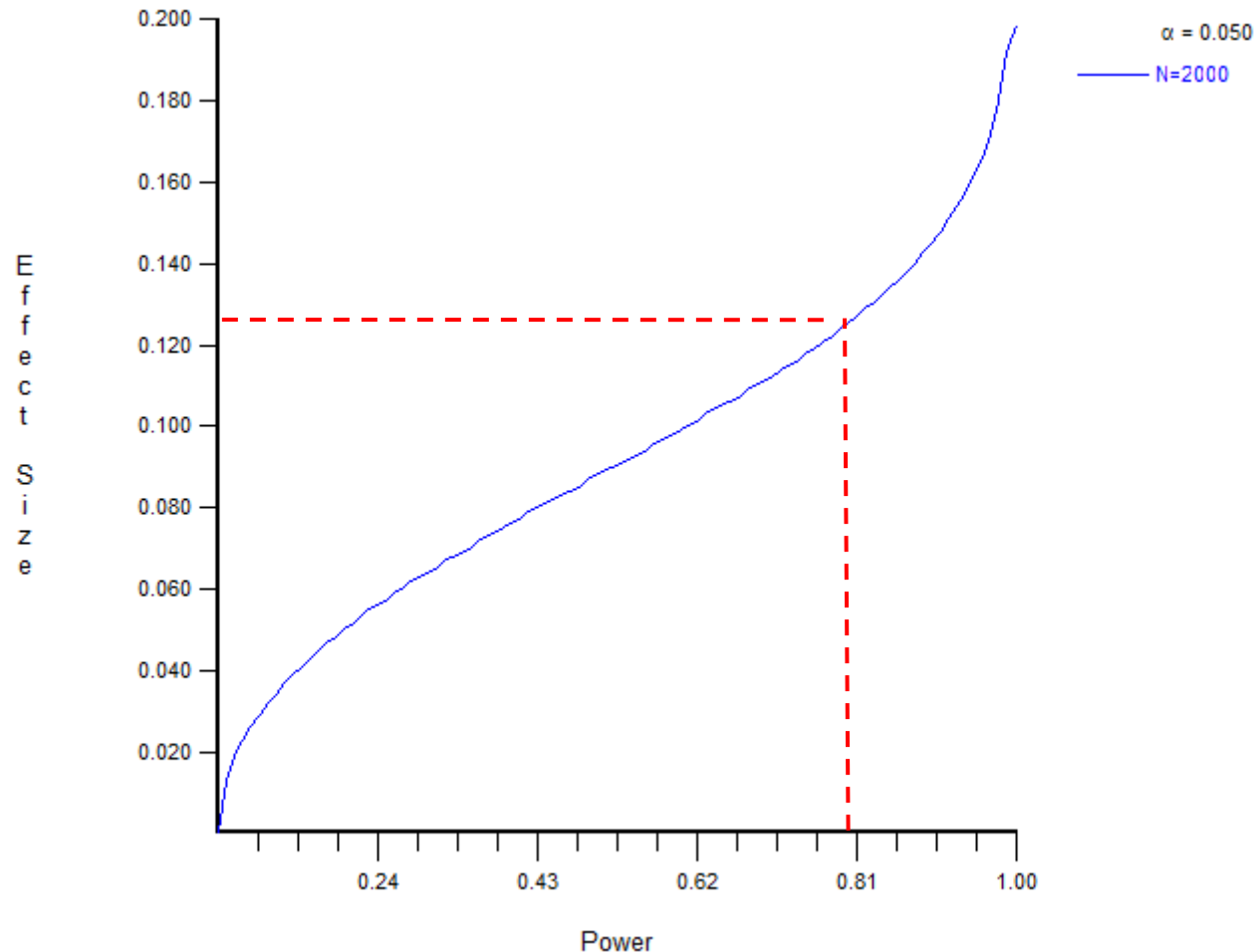
- Autrement, si vous pouvez vous permettre un échantillon de 2000, vous pouvez trouver la taille d'effet minimale décelable en fonction de cet échantillon:



# Exemple avec le logiciel OD (2)

Avec un échantillon de 2000, vous pouvez obtenir une puissance de 80% à une taille d'effet d'environ 0,125

Si vous attendez un petit effet, vous devez exiger un budget plus grand



# Conclusions: Puissance

## Calculs en pratique

---

- Les calculs de la puissance impliquent quelques hypothèses – nous n'avons pas à chaque fois des informations justes .
  - Cela vous permet d'éviter les pertes de temps et d'argent pour lancer des études qui n'auront aucune puissance.
  - Elles soutiennent l'argument que vous devez consacrer assez de temps et de ressources aux études que vous décidez de mener.
- Faites des prévisions; mettez en place des réserves ou des plans d'urgence; sachez que vous aurez à traiter les écarts par rapport à l'échantillonnage parfait; Cela va certainement se produire.

# Ressources sur l'échantillonnage et les statistiques aux fins d'évaluation

---

- **UNDESA (2005) *Household Sample Surveys in Developing and Transition Countries***  
[http://unstats.un.org/unsd/hhsurveys/pdf/Household\\_surveys.pdf](http://unstats.un.org/unsd/hhsurveys/pdf/Household_surveys.pdf)
- **Baker, J. (2000) *Evaluating the Impact of Development Projects on Poverty: A Handbook for Practitioners***  
<http://go.worldbank.org/8E2ZTGB0I0>
- **Bamberger, M. (2006) *Conducting Quality Impact Evaluations under Budget, Time and Data Constraints***  
[http://lnweb90.worldbank.org/oed/oeddoclib.nsf/DocUNIDViewForJavaSearch/757A5CC0BAE22558852571770059D89C/\\$file/conduct\\_qual\\_impact.pdf](http://lnweb90.worldbank.org/oed/oeddoclib.nsf/DocUNIDViewForJavaSearch/757A5CC0BAE22558852571770059D89C/$file/conduct_qual_impact.pdf)