

Dakar, Senegal

# IsDB-World Bank DIME Impact Evaluation Event

January 29-31, 2019

Transforming Development  
through Evidence-Based Policy

## Echantillonnage et calculs de puissance pour une évaluation d'impact

Guigonan Serge Adjognon

Economiste, DIME



# Introduction

- **Maintenant que vous savez comment construire les groupes de traitement et de contrôle dans la théorie, comment le faire dans la pratique?**
  - Quelle population (ou groupe) nous intéresse et où allons-nous la (les) trouver? **Sélection des personnes à interroger**
  - Dans cette population, quel est le nombre de personnes/d'unités à interroger/observer? **Taille de l'échantillon**
- **Apparemment trivial, mais «le diable est dans les détails»**
- **Exemple:** Supposons que nous voulions comprendre l'impact de la mise en place d'un programme de subvention d'intrants agricoles sur les rendements agricoles et la pauvreté.

Dakar, Senegal | January 29-31, 2019



# les échantillons plus grands sont plus précis

- Pensez à la taille de l'échantillon comme **à la précision d'un outil de mesure**
  - Plus vous avez des **observations**,
  - Plus précis est votre « **outil de mesure** »,Alors, plus vous êtes **confiant** en les conclusions de votre évaluation

J					E						
					L				I		
			M			C					

## les échantillons plus grands sont plus précis

- Augmentons le **nombre « d'observations »** (dans ce cas, des lettres)
  - C'est tellement plus facile

J		A			E						
L		E			L			T	I		N
D		I	M			C	T				

- Vous vous sentez plus à **l'aise** pour deviner
- Intuition : **plus compliquée** est la phrase, plus vous aurez besoin de lettres

## les échantillons plus grands sont plus précis

- Augmentons encore le **nombre « d'observations »** (dans ce cas, des lettres)
  - Encore plus facile

J	'	A		M	E						
L	'	E	V	A	L	U	A	T	I		N
D		I	M	P		C	T				

- Intuition : **plus compliquée** est la phrase, plus vous aurez besoin de lettres

## les échantillons plus grands sont plus précis

- Augmentons encore le **nombre « d'observations »** (dans ce cas, des lettres)
  - Encore plus facile

J	'	A	I	M	E						
L	'	E	V	A	L	U	A	T	I	O	N
D	'	I	M	P	A	C	T				

- Intuition : **plus compliquée** est la phrase, plus vous aurez besoin de lettres

---

# Quelle est la taille propice pour l'échantillon?

Dakar, Senegal | January 29-31, 2019



...

$$n = \left[ \frac{4\sigma^2 (z_{1-\alpha/2} + z_{1-\beta})^2}{D^2} \right] [1 + \rho(m-1)]$$

Dakar, Senegal | January 29-31, 2019





---

# Merci pour votre attention

Dakar, Senegal | January 29-31, 2019



## Quelle est la taille d' échantillon nécessaire ???

- La dérivation statistique de la taille de l'échantillon donne **une formule affreuse** :

$$N = \left[ \frac{4\sigma^2 (z_{\alpha/2} + z_{\beta})^2}{D^2} \right] [1 + \rho(H-1)]$$

- **Ingrédients fondamentaux**

1. Quel est **l'impact minimum** qui justifierait d'investir dans le programme ?
2. Quelle est la **variance** de l'indicateur de résultat qui vous intéresse ?
3. Le programme génère-t-il des **grappes** ?

# La « confiance » en statistiques

## ▪ Confiance :

- Cette fois, le jargon statistique et le français courant renvoient à la même notion
- Nous comprenons «confiance» pour signifier «avec un certain degré de certitude» ou «avec peu d'erreurs »
- La même chose est vraie dans le sens statistique du terme, cela implique de clarifier ce que l'on entend par « erreur ».
- Un échantillon suffisamment grand nous donne « confiance » dans la solidité de nos conclusions

# Comment déterminer la taille de l'échantillon?

## 1. Ingrédients pour déterminer la taille d'échantillon

1<sup>er</sup> ingrédient : effet minimum détectable

2<sup>e</sup> ingrédient : variance de l'indicateur/des indicateurs

3<sup>e</sup> ingrédient : grappes

## 2. Autres facteurs

- a) Multiples questions d'évaluation/ groupes expérimentaux
- b) Résultats désagrégés par groupe
- c) Taux de participation au programme
- d) Qualité des données

Dakar, Senegal | January 29-31, 2019



## Puissance et risque d'erreur de type II

“Erreur de type 2” : Le risque de ne pas conclure que le programme a un impact alors qu’il en a un.

Le risque d'erreur de type II est critique surtout en cas de petite différences

Cela pourrait conduire à décider de la suppression du programme, ce qui serait préjudiciable à la société ---- GROSSE ERREUR !

Une évaluation d'impact est dite **puissante** si **le risque de non détection de l'impact réel du programme alors qu'il y en a un** – autrement dit, le risque d'erreur de type 2 – **est minimal.**

# 1<sup>er</sup> ingrédient : effet minimum

- Nous ne savons pas l'effet de notre politique à l'avance . Nous voulons concevoir une façon précise de le mesurer
- Mais la précision n'est pas gratuite: besoin d'analyse coûts-bénéfices pour décider.
- **Premier ingrédient**: Impact minimal du programme que l'on souhaite détecter
  - **C'est-à-dire le plus petit effet pour lequel nous serions en mesure de conclure qu'il est statistiquement différent de zéro**
  - «**Détecter**» est utilisé dans un sens statistique
  - Si l'intervention est supposée accroître la productivité agricole d'au moins 30%, on peut décider qu'on ne cherche pas forcément à détecter avec précision les réductions inférieures à 5%

Dakar, Senegal | January 29-31, 2019



# Qui est le plus grand ?

- Plus grand est l'échantillon → plus précis est le dispositif de mesure → plus il est facile de détecter les plus petites différences
  - augmentation de la taille de l'échantillon → précision croissante (de notre appareil de mesure)



# Qui est le plus grand ?

- Plus grand est l'échantillon → plus précis est le dispositif de mesure → plus il est facile de détecter les plus petites différences
  - augmentation de la taille de l'échantillon → précision croissante (de notre appareil de mesure)



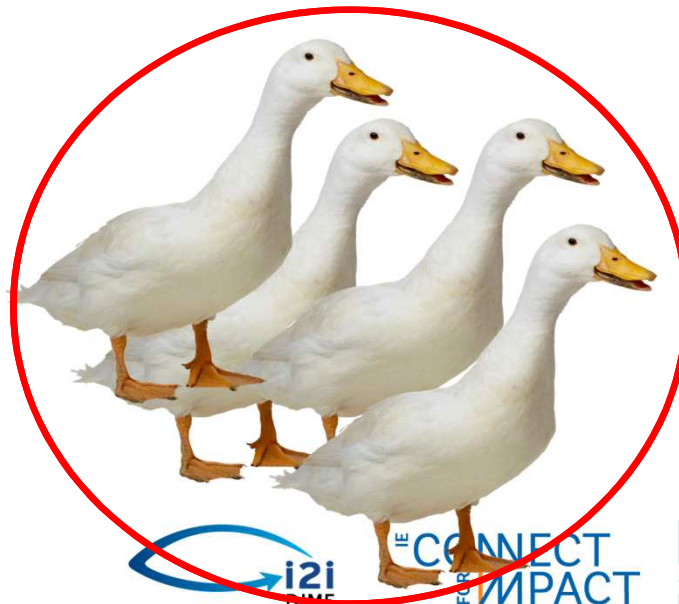
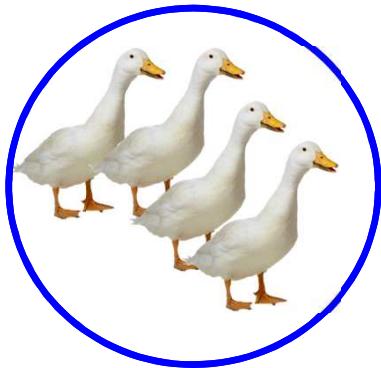


## 1<sup>er</sup> ingrédient : effet minimum

- L'analyse coûts-bénéfices nous guide dans la détermination de « l'impact minimum détectable » :
  - Cela pourrait être utile pour la politique
  - Cela pourrait justifier le coût d'une évaluation d'impacts, etc.
- Plus la différence (espérée) entre groupes de traitement & témoin sera faible ...
- ... plus l'outil utilisé pour la détecter devra être précis
  - ➔ Plus l'échantillon devra être grand

## 2<sup>ème</sup> ingrédient : variance de l'indicateur de résultats

- Comment la variance de l'indicateur de résultat affecte-t-elle notre capacité à détecter un impact ?
  - **Exemple** : Laquelle des deux populations (entourées) est plus grande ? Combien d'observations de chaque cercle faut-il pour y répondre ?



Dakar, Senegal | January 29-31, 2019

## 2<sup>ème</sup> ingrédient : variance de l'indicateur de résultats

- **Exemple** : quel groupe a les plus gros animaux en moyenne ?
- La comparaison est plus complexe, et nécessite plus d'informations (i.e. un échantillon plus large)
  - La réponse dépend desquels membres des groupes bleu et rouge sont observés



Dakar, Senegal | January 29-31, 2019

i2i  
DIME  
TRANSFORM DEVELOPMENT

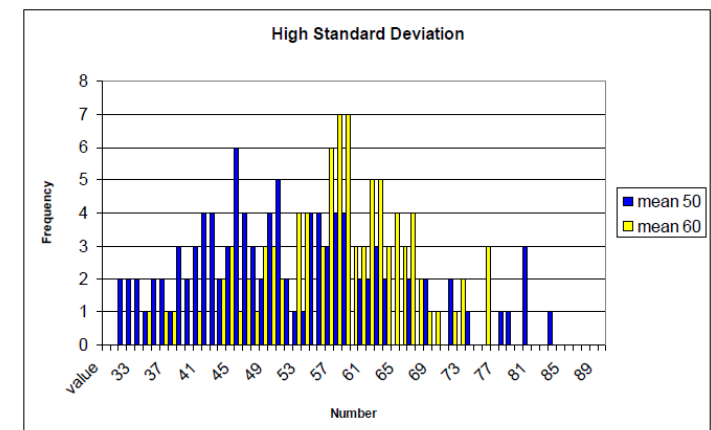
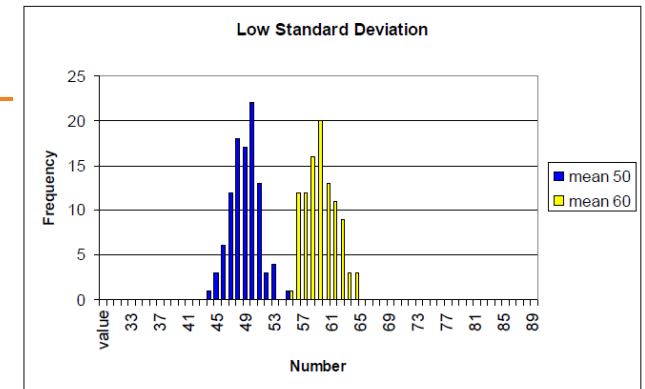
FOR  
CONNECT  
IMPACT

IsDB  
البنك الإسلامي للتنمية  
Islamic Development Bank

UKaid  
from the British people

## 2<sup>ème</sup> ingrédient : variance de l'indicateur de résultats

- **En somme :**
  - Plus de variance sous-jacente (hétérogénéité)
  - → plus difficile de détecter des différences
  - → besoin d'une plus grande taille d'échantillon
- **Délicat:** que savons-nous de l'hétérogénéité avant de décider de la taille de l'échantillon et de la collecte des données?
  - Idéal: données pré-existantes ... mais souvent inexistantes
  - Peut utiliser des données déjà existantes à partir d'une population similaire
  - Exemple: enquêtes auprès des communes, provinces, enquêtes sur la population active
  - Bon sens



Dakar, Senegal | January 29-31, 2019

# 3<sup>e</sup> ingrédient : grappes

- **Le programme génère-t-il des grappes?**
  - A quel **niveau les résultats sont-ils mesurés** ?
  - A quel **niveau le programme est-il mis en œuvre** ?
- **Exemple d'un programme de subvention agricole destiné à augmenter l'accès aux services gouvernementaux par les populations**
  - Le programme est mis en œuvre au niveau commune ou province?
    - La précision statistique dépend principalement du nombre de communes ou provinces
  - Mais l'impact est mesuré au niveau plus bas: ménage
    - Échantillonnage par grappe : d'abord échantillonner les communes, puis les ménages au sein de ces communes (en passant par villages)
- **Enjeux liés aux grappes :**
  - Les résultats des individus du même groupe sont susceptibles d'être corrélés (**corrélation intra-classe**)
- **Nécessite d'ajuster l'échantillon** : En termes de puissance, il vaut mieux ajouter 1 observation d'une nouvelle grappe, plutôt qu'1 observation d'une grappe existante
- **C'est le nombre de grappes qui détermine en grande partie la taille 'utile' de l'échantillon** (le nombre d'individus au sein des grappes importe moins)

Dakar, Senegal | January 29-31, 2019



# Autres facteurs

- Autres caractéristiques de la conception / des données qui peuvent avoir des implications pour la détermination de la **taille de l'échantillon**
  1. Multiples questions d'évaluation/traitements
  2. Résultats désagrégés par groupe
  3. Taux de participation au programme
  4. Qualité des données

# Traitements multiples

---

- Besoin alors de très grands échantillons
  - Plus on veut faire de comparaisons, plus l'échantillon doit être grand
- Si les traitement sont similaires, les différences entre groupes seront petites
  - On aura besoin d'échantillons encore plus grands

Dakar, Senegal | January 29-31, 2019



# Désagréger les résultats par groupe

- **Résultats désagrégés par groupe**
  - Les effets sont-ils différents pour les hommes et les femmes? Pour les différents secteurs?
  - Si ces groupes différents réagissent de façon similaire, estimer les différences d'impact de traitement exige également de très grands échantillons
- Afin d'assurer l'équilibre entre le traitement et les groupes de comparaison, il convient de diviser l'échantillon en **strates** avant d'assigner les strates de traitement
- **Strates**
  - Sous-populations
  - Strates souvent utilisées : géographie, sexe, secteur, valeurs de référence de variables de résultats
  - L'assignation de traitement (ou échantillonnage) se produit au sein de ces groupes (c'est-à-dire randomiser au sein des strates)



# La participation au programme

---

- Faible participation (taux) diminue la précision de nos comparaisons
  - Diminue la taille de l'échantillon

Dakar, Senegal | January 29-31, 2019



# Qualité des données

- Une mauvaise qualité des données augmente en réalité la taille requise de l'échantillon
  - Observations manquantes
  - Grande erreur de mesure
- Garantir des données de bonne qualité est souvent un bon investissement
  - Peut-être en partie résolue avec le coordinateur de terrain par le contrôle sur le terrain de la collecte des données

Dakar, Senegal | January 29-31, 2019



# Conclusion

Éléments :	Implication pour la Taille de l'Echantillon :
Les <b>effets minimaux</b> que l'on souhaite détecter	Plus la <b>taille de l'échantillon</b> devra être <b>grande</b>
Plus la <b>variance</b> sous-jacente est élevée	
Plus le niveau de mise en oeuvre est élevé ( <b>grappes</b> ) et la corrélation des résultats au sein d'une grappe sont élevés	
Plus on souhaite de <b>confiance/précision</b> (statistique)	
Plus la <b>nature des questions d'évaluation est complexe</b> - Traitements multiples - Intérêt porté à la comparaison entre sous-groupes	
Plus le <b>taux de participation</b> est faible	
Plus les <b>données</b> sont mauvaises	

## Points clés

- L'échantillonnage correspond au processus de tirage d'un échantillon d'unités d'une population, afin d'estimer les caractéristiques de cette population
- Les échantillons plus grands permettent d'estimer plus précisément les caractéristiques de la population
- Des petits échantillons créent des risques de tirer des conclusions politiques erronées
- Les calculs de puissance nous informent de la taille requise pour les échantillons. Des échantillons plus grands sont nécessaires pour estimer précisément un impact si nous nous attendons à un impact faible, ou si le programme génère des grappes,...

# Thank You



<http://dime.worldbank.org>  
<http://transport.worldbank.org>



[Dimewiki.worldbank.org](http://Dimewiki.worldbank.org)  
[DIME Wiki](#) | [ietoolkit](#) | [Stata GitHub](#)



@impacteval  
@WB\_transport



[blogs.worldbank.org/impactevaluations](http://blogs.worldbank.org/impactevaluations)  
[blogs.worldbank.org/transport](http://blogs.worldbank.org/transport)



[microdata.worldbank.org/index.php/catalog/impact\\_evaluation](http://microdata.worldbank.org/index.php/catalog/impact_evaluation)

## Research Design

Experimental Methods  
Quasi-Experimental Methods  
Research Ethics  
Sampling & Power Calculations

## Data Collection

Primary Data Collection  
Secondary Data Sources  
Field Management  
Questionnaire Design

## Analysis

Data Management  
Data Cleaning  
Data Analysis  
Software Tools

## Publication

Reproducible Research  
Publishing Data  
Collaboration Tools  
Dissemination

Dakar, Senegal | January 29-31, 2019

