

# Séance 5 : Relier des variables entre elles

Application : L'expérience d'Esther Duflo et al. au Ghana

**Mattéo Lanoë**

Printemps 2026

# De la description à la relation

## Où en sommes-nous ?

- ▶ Jusqu'à présent, nous avons **décrit** des variables isolées (Tendance centrale, dispersion).

# De la description à la relation

## Où en sommes-nous ?

- ▶ Jusqu'à présent, nous avons **décrit** des variables isolées (Tendance centrale, dispersion).
- ▶ Aujourd'hui, nous cherchons à les **relier**.

# De la description à la relation

## Où en sommes-nous ?

- ▶ Jusqu'à présent, nous avons **décrit** des variables isolées (Tendance centrale, dispersion).
- ▶ Aujourd'hui, nous cherchons à les **relier**.
- ▶ L'objectif : passer de la simple description à l'**explication** des phénomènes sociaux.

## Retours sur l'interprétation de l'écart-type

## Retours sur l'interprétation de l'écart-type

Groupe (Classe)	Prix Moyen du Billet	Écart-Type (Dispersion)
1ère Classe	84 £	78 £ ( <i>Fort</i> )
2ème Classe	20 £	13 £ ( <i>Fort</i> )
3ème Classe	13 £	11 £ ( <i>Fort</i> )

## Retours sur l'interprétation de l'écart-type

Groupe (Classe)	Prix Moyen du Billet	Écart-Type (Dispersion)
1ère Classe	84 £	78 £ ( <i>Fort</i> )
2ème Classe	20 £	13 £ ( <i>Fort</i> )
3ème Classe	13 £	11 £ ( <i>Fort</i> )

### Interprétation sociologique :

- ▶ **Le Niveau (Moyenne)** : Il y a un gouffre financier. La 1ère classe paie en moyenne 6 fois plus cher que la 3ème.

## Retours sur l'interprétation de l'écart-type

Groupe (Classe)	Prix Moyen du Billet	Écart-Type (Dispersion)
1ère Classe	84 £	78 £ ( <i>Fort</i> )
2ème Classe	20 £	13 £ ( <i>Fort</i> )
3ème Classe	13 £	11 £ ( <i>Fort</i> )

### Interprétation sociologique :

- ▶ **Le Niveau (Moyenne)** : Il y a un gouffre financier. La 1ère classe paie en moyenne 6 fois plus cher que la 3ème.
- ▶ **La Cohésion (Écart-type)** :

## Retours sur l'interprétation de l'écart-type

Groupe (Classe)	Prix Moyen du Billet	Écart-Type (Dispersion)
1ère Classe	84 £	78 £ ( <i>Fort</i> )
2ème Classe	20 £	13 £ ( <i>Fort</i> )
3ème Classe	13 £	11 £ ( <i>Fort</i> )

### Interprétation sociologique :

- ▶ **Le Niveau (Moyenne)** : Il y a un gouffre financier. La 1ère classe paie en moyenne 6 fois plus cher que la 3ème.
- ▶ **La Cohésion (Écart-type)** :
  - ▶ Les pauvres ont payé des prix allant de 0£ à 70£. La différence est énorme. Si on regarde par quartiles, 75% des passagers de 3ème classe payaient moins de 15£. Les 25% restants ont payé des prix beaucoup plus élevés, allant jusqu'à 70£.

## Retours sur l'interprétation de l'écart-type

Groupe (Classe)	Prix Moyen du Billet	Écart-Type (Dispersion)
1ère Classe	84 £	78 £ ( <i>Fort</i> )
2ème Classe	20 £	13 £ ( <i>Fort</i> )
3ème Classe	13 £	11 £ ( <i>Fort</i> )

### Interprétation sociologique :

- ▶ **Le Niveau (Moyenne)** : Il y a un gouffre financier. La 1ère classe paie en moyenne 6 fois plus cher que la 3ème.
- ▶ **La Cohésion (Écart-type)** :
  - ▶ Les pauvres ont payé des prix allant de 0£ à 70£. La différence est énorme. Si on regarde par quartiles, 75% des passagers de 3ème classe payaient moins de 15£. Les 25% restants ont payé des prix beaucoup plus élevés, allant jusqu'à 70£.
  - ▶ Les riches (1ère) sont très différents entre eux (écart fort) : il y a des "petits riches" et des multimilliardaires.

# Représenter les variables Y et X sur un graphique

Avant de relier des variables, il faut savoir les représenter correctement.

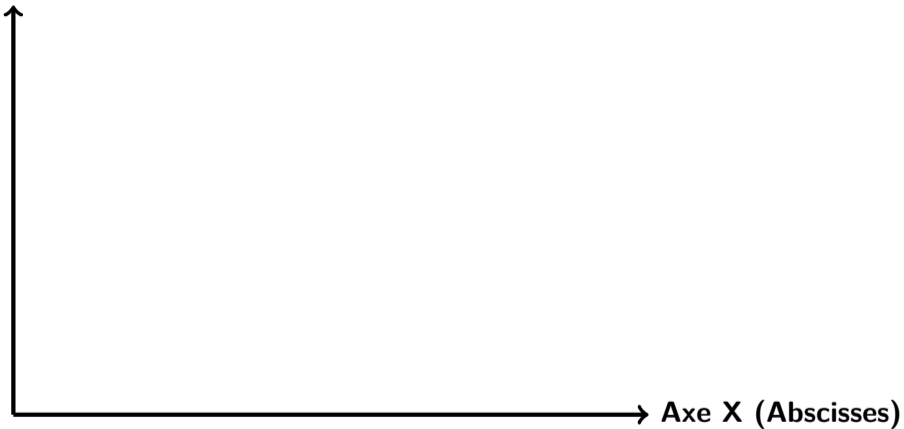
## Une convention universelle

En sciences sociales (et naturelles), la place d'une variable sur un graphique n'est pas due au hasard.

- ▶ L'axe horizontal (**Abscisses**) = La cause supposée.
- ▶ L'axe vertical (**Ordonnées**) = L'effet observé.

## Visualiser la convention

Axe Y (Ordonnées)



## Visualiser la convention

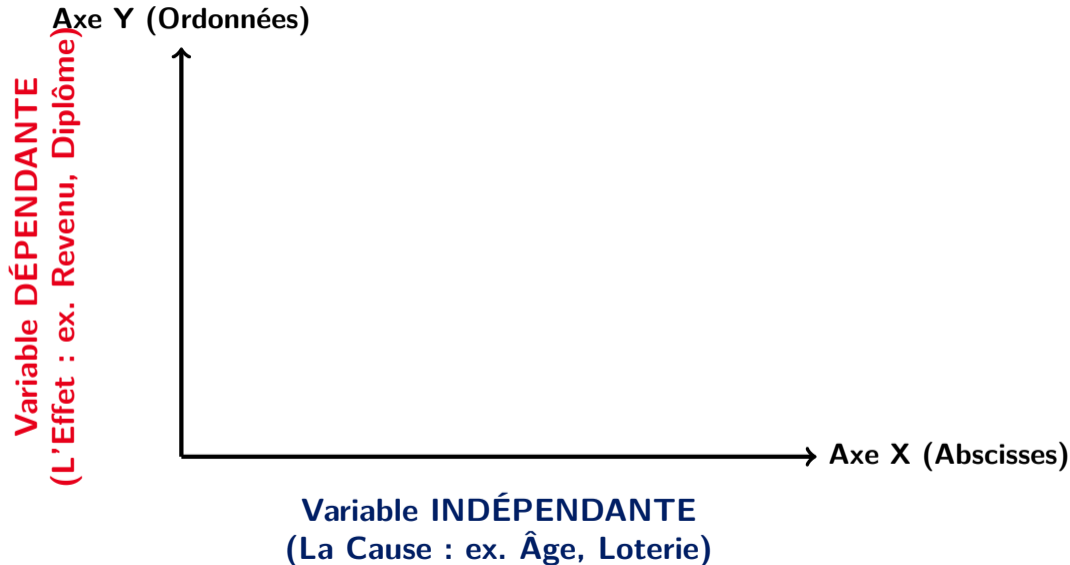
Axe Y (Ordonnées)



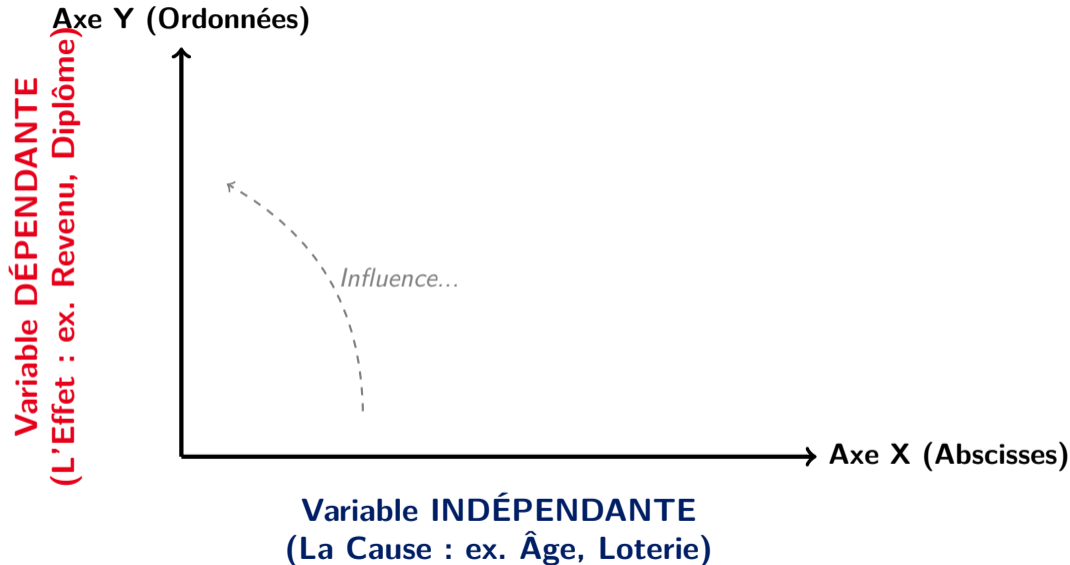
Axe X (Abscisses)

**Variable INDÉPENDANTE**  
(La Cause : ex. Âge, Loterie)

## Visualiser la convention



## Visualiser la convention



## Une application concrète : L'expérience au Ghana

Nous allons étudier l'article de :

**Esther Duflo, Pascaline Dupas, Michael Kremer (2021).**

- ▶ E. Duflo et M. Kremer ont reçu le **Prix Nobel d'Économie** (2019).
- ▶ Leur but : évaluer scientifiquement les politiques de lutte contre la pauvreté.

*Esther Duflo, économiste*



# Le point de départ de la recherche

## La question de recherche

Est-ce que rendre l'école secondaire (le lycée) **gratuite** permet d'améliorer la vie des jeunes dans les pays en développement ?

# Le point de départ de la recherche

## La question de recherche

Est-ce que rendre l'école secondaire (le lycée) **gratuite** permet d'améliorer la vie des jeunes dans les pays en développement ?

## Les effets attendus (Variables dépendantes) :

- ▶ Plus d'années d'études validées ?

# Le point de départ de la recherche

## La question de recherche

Est-ce que rendre l'école secondaire (le lycée) **gratuite** permet d'améliorer la vie des jeunes dans les pays en développement ?

### Les effets attendus (Variables dépendantes) :

- ▶ Plus d'années d'études validées ?
- ▶ Un meilleur accès à l'emploi (notamment public) ?

# Le point de départ de la recherche

## La question de recherche

Est-ce que rendre l'école secondaire (le lycée) **gratuite** permet d'améliorer la vie des jeunes dans les pays en développement ?

### Les effets attendus (Variables dépendantes) :

- ▶ Plus d'années d'études validées ?
- ▶ Un meilleur accès à l'emploi (notamment public) ?
- ▶ Une baisse des grossesses précoces chez les jeunes femmes ?

## Le problème de la méthode classique

Pourquoi ne peut-on pas **simplement comparer** ceux qui vont au lycée et ceux qui n'y vont pas ?

## Le problème de la méthode classique

Pourquoi ne peut-on pas **simplement comparer** ceux qui vont au lycée et ceux qui n'y vont pas ?

### Le Biais de Sélection

Les enfants qui vont au lycée ne sont pas les mêmes que ceux qui n'y vont pas :

- ▶ Ils viennent souvent de familles plus aisées.
- ▶ Leurs parents sont parfois plus instruits.
- ▶ Ils habitent plus près des villes.

## Le problème de la méthode classique

Pourquoi ne peut-on pas **simplement comparer** ceux qui vont au lycée et ceux qui n'y vont pas ?

### Le Biais de Sélection

Les enfants qui vont au lycée ne sont pas les mêmes que ceux qui n'y vont pas :

- ▶ Ils viennent souvent de familles plus aisées.
- ▶ Leurs parents sont parfois plus instruits.
- ▶ Ils habitent plus près des villes.

→ Si on observe qu'ils ont de meilleurs salaires plus tard, est-ce grâce au lycée... ou grâce à la richesse de leurs parents ?

## La solution : L'expérience scientifique

Pour prouver que c'est bien la **gratuité** qui cause la réussite, il faut créer une situation artificielle parfaite.

## La solution : L'expérience scientifique

Pour prouver que c'est bien la **gratuité** qui cause la réussite, il faut créer une situation artificielle parfaite.

### L'Essai Randomisé Contrôlé (RCT)

On va distribuer des bourses d'études **par tirage au sort** (une loterie) parmi des jeunes qui ont tous le niveau pour aller au lycée mais pas les moyens.

## Pourquoi utiliser le hasard ?

Le tirage au sort est la seule méthode qui garantit l'égalité parfaite.

- ▶ La machine choisit au hasard.

## Pourquoi utiliser le hasard ?

Le tirage au sort est la seule méthode qui garantit l'égalité parfaite.

- ▶ La machine choisit au hasard.
- ▶ Elle ne regarde ni le revenu des parents, ni la motivation de l'élève.

## Pourquoi utiliser le hasard ?

Le tirage au sort est la seule méthode qui garantit l'égalité parfaite.

- ▶ La machine choisit au hasard.
- ▶ Elle ne regarde ni le revenu des parents, ni la motivation de l'élève.
- ▶ **Résultat** : Les deux groupes (ceux qui gagnent et ceux qui perdent) sont **statistiquement identiques** au départ.

## Pourquoi utiliser le hasard ?

Le tirage au sort est la seule méthode qui garantit l'égalité parfaite.

- ▶ La machine choisit au hasard.
- ▶ Elle ne regarde ni le revenu des parents, ni la motivation de l'élève.
- ▶ **Résultat** : Les deux groupes (ceux qui gagnent et ceux qui perdent) sont **statistiquement identiques** au départ.
- ▶ La seule différence dans leur vie sera l'intervention.

# Le protocole de l'expérience

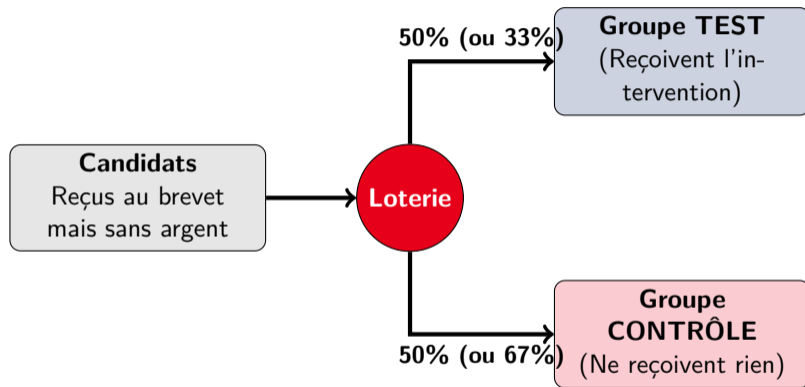
## Candidats

Reçus au brevet  
mais sans argent

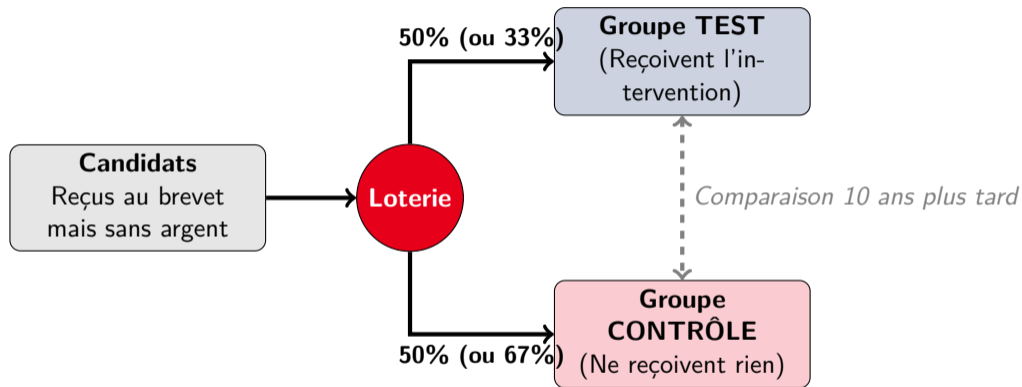


Loterie

## Le protocole de l'expérience



## Le protocole de l'expérience



## De l'expérience au tableau de données

Comment cette expérience se traduit-elle dans nos fichiers Excel / R?

# De l'expérience au tableau de données

Comment cette expérience se traduit-elle dans nos fichiers Excel / R?

## La Variable Indépendante (La Cause)

C'est la variable Test (Le résultat de la loterie).

- ▶ 1 = A gagné la loterie (Groupe Test).
- ▶ 0 = A perdu la loterie (Groupe Contrôle).

*C'est une variable qualitative binaire.*

# De l'expérience au tableau de données

Qu'allons-nous mesurer ?

## Les Variables Dépendantes (Les Effets)

- ▶ `education` : Le nombre d'années d'études validées (Quantitative).
- ▶ `grossesse` : A eu un enfant prématurément (Qualitative : 1 = Oui, 0 = Non).
- ▶ `emploi_public` : Travaille pour l'État (Qualitative : 1 = Oui, 0 = Non).

## Votre mission : Étape 1 (Le test d'équilibrage)

**Fichier** : `expe_lycee_Ghana_avantloterie`

Avant de crier victoire, il faut vérifier que le tirage au sort a bien fonctionné. Les deux groupes devaient être égaux **avant** l'intervention.

## Votre mission : Étape 1 (Le test d'équilibrage)

**Fichier** : `expe_lycee_Ghana_avantloterie`

Avant de crier victoire, il faut vérifier que le tirage au sort a bien fonctionné. Les deux groupes devaient être égaux **avant** l'intervention.

- ▶ **Action** : Calculez la moyenne des notes au brevet (BECE) pour le groupe Test = 1, puis pour le groupe Test = 0.
- ▶ **Objectif** : Les deux moyennes doivent être presque identiques.

## Ta mission : Étape 2 (Mesurer l'impact)

**Fichier** : `expe_lycee_Ghana_aprestloterie`

C'est le croisement **Quali** × **Quanti** (vu en Séance 4).

## Ta mission : Étape 2 (Mesurer l'impact)

**Fichier** : `expe_lycee_Ghana_aprestloterie`

C'est le croisement **Quali** × **Quanti** (vu en Séance 4).

**Marche à suivre** :

1. Calculer la moyenne des années d'études (`education`) validées par le groupe `Test = 1`.
2. Calculer cette même moyenne pour le groupe `Test = 0`.
3. Soustraire les deux moyennes pour obtenir **l'effet causal** de la bourse.

## 14.1 Pourquoi la soustraction donne l'effet causal ?

Pour mesurer un impact réel, il faudrait théoriquement cloner un individu pour observer sa vie avec et sans la bourse.

### Le rôle du groupe de contrôle

Puisque le clonage est impossible, le groupe Contrôle ( $\text{Test} = 0$ ) sert de copie conforme au groupe Test ( $\text{Test} = 1$ ).

## 14.1 Pourquoi la soustraction donne l'effet causal ?

Pour mesurer un impact réel, il faudrait théoriquement cloner un individu pour observer sa vie avec et sans la bourse.

### Le rôle du groupe de contrôle

Puisque le clonage est impossible, le groupe Contrôle (Test = 0) sert de copie conforme au groupe Test (Test = 1).

La loterie garantit que les deux groupes ont exactement les mêmes caractéristiques de départ (origines sociales, niveau scolaire, etc.).

Le groupe Contrôle incarne le **contrefactuel** : ce qui se serait passé pour le groupe Test s'il n'avait pas reçu l'argent.

## 14.2 L'isolement de l'effet

Puisque les deux groupes sont des clones statistiques au départ, leurs trajectoires de vie devraient évoluer exactement de la même manière.

## 14.2 L'isolement de l'effet

Puisque les deux groupes sont des clones statistiques au départ, leurs trajectoires de vie devraient évoluer exactement de la même manière.

### La déduction logique

Toute différence finale (comme le nombre d'années d'études) ne peut provenir que d'une seule cause : l'intervention elle-même.

## 14.2 L'isolement de l'effet

Puisque les deux groupes sont des clones statistiques au départ, leurs trajectoires de vie devraient évoluer exactement de la même manière.

### La déduction logique

Toute différence finale (comme le nombre d'années d'études) ne peut provenir que d'une seule cause : l'intervention elle-même.

### Le calcul :

1. Prendre le niveau final du groupe Test.
2. Retirer le niveau final du groupe Contrôle (qui représente l'évolution naturelle sans la bourse).
3. Le reste correspond à l'effet pur et exclusif de la bourse.

## 15.1 L'incertitude scientifique

En sciences sociales, on n'interroge presque jamais **toute** la population, mais seulement un **échantillon** (ex : 2000 lycéens au Ghana).

## 15.1 L'incertitude scientifique

En sciences sociales, on n'interroge presque jamais **toute** la population, mais seulement un **échantillon** (ex : 2000 lycéens au Ghana).

### Le problème du hasard

Si on tirait au sort un autre échantillon de 2000 jeunes, on trouverait un résultat légèrement différent. Il y a donc toujours un "bruit" statistique, une incertitude.

## 15.1 L'incertitude scientifique

En sciences sociales, on n'interroge presque jamais **toute** la population, mais seulement un **échantillon** (ex : 2000 lycéens au Ghana).

### Le problème du hasard

Si on tirait au sort un autre échantillon de 2000 jeunes, on trouverait un résultat légèrement différent. Il y a donc toujours un "bruit" statistique, une incertitude.

*Exemple de la vie courante* : Les sondages politiques.

"Le candidat A est estimé à 51% des voix, **plus ou moins 2%**."

Ce "plus ou moins", c'est notre marge d'erreur.

## 15.2 Calculer la Marge d'Erreur Statistique (MES)

Comment calcule-t-on ce "flou" statistique ? Voici la formule utilisée dans votre fichier de données :

## 15.2 Calculer la Marge d'Erreur Statistique (MES)

Comment calcule-t-on ce "flou" statistique ? Voici la formule utilisée dans votre fichier de données :

La Formule (pour un niveau de confiance à 95%)

$$\text{MES} = 1,96 \times \text{Écart-Type} \times \left( \frac{1}{\sqrt{n_1}} + \frac{1}{\sqrt{n_0}} \right)$$

## 15.2 Calculer la Marge d'Erreur Statistique (MES)

Comment calcule-t-on ce "flou" statistique ? Voici la formule utilisée dans votre fichier de données :

La Formule (pour un niveau de confiance à 95%)

$$\text{MES} = 1,96 \times \text{Écart-Type} \times \left( \frac{1}{\sqrt{n_1}} + \frac{1}{\sqrt{n_0}} \right)$$

**Que signifient ces lettres ?**

- ▶ **1,96** : Le seuil mathématique standard pour être sûr à 95%.
- ▶ **Écart-Type** : La dispersion de notre variable (ex : l'éducation).
- ▶  **$n_1$  et  $n_0$**  : Le nombre de personnes dans le groupe Test ( $n_1$ ) et dans le groupe Contrôle ( $n_0$ ). Plus on a de monde, plus la MES rétrécit !

## 15.3 La Règle d'Or : Que faire de cette MES ?

Avoir la MES ne suffit pas. Il faut obligatoirement la comparer à l'**effet mesuré** de notre politique (la bourse).

## 15.3 La Règle d'Or : Que faire de cette MES ?

Avoir la MES ne suffit pas. Il faut obligatoirement la comparer à l'**effet mesuré** de notre politique (la bourse).

**Étape 1** : Calculer la différence entre les deux moyennes.

→ *Différence = Moyenne(Groupe Traité) - Moyenne(Groupe Contrôle)*

## 15.3 La Règle d'Or : Que faire de cette MES ?

Avoir la MES ne suffit pas. Il faut obligatoirement la comparer à l'**effet mesuré** de notre politique (la bourse).

**Étape 1** : Calculer la différence entre les deux moyennes.

→  $Différence = Moyenne(Groupe Traité) - Moyenne(Groupe Contrôle)$

### Étape 2 : La Comparaison

- ▶ Si la **Différence**  $>$  **MES** → L'effet dépasse le "bruit" statistique. Le résultat est **statistiquement significatif** (solide).

## 15.3 La Règle d'Or : Que faire de cette MES ?

Avoir la MES ne suffit pas. Il faut obligatoirement la comparer à l'**effet mesuré** de notre politique (la bourse).

**Étape 1** : Calculer la différence entre les deux moyennes.

→  $Différence = Moyenne(Groupe Traité) - Moyenne(Groupe Contrôle)$

### Étape 2 : La Comparaison

- ▶ Si la **Différence**  $>$  **MES** → L'effet dépasse le "bruit" statistique. Le résultat est **statistiquement significatif** (solide).
- ▶ Si la **Différence**  $<$  **MES** → Le "flou" est plus gros que l'effet observé. Le résultat est **non significatif** (impossible de conclure).

## 15.4 Comment formuler son interprétation ?

Voici comment rédiger vos conclusions dans un article ou un rapport :

## 15.4 Comment formuler son interprétation ?

Voici comment rédiger vos conclusions dans un article ou un rapport :

**Exemple de résultat Significatif (Différence  $>$  MES) :**

*"La bourse a augmenté la scolarisation de 1,5 an en moyenne. La marge d'erreur n'étant que de 0,5 an, cette différence est statistiquement significative. La politique est donc efficace."*

## 15.4 Comment formuler son interprétation ?

Voici comment rédiger vos conclusions dans un article ou un rapport :

### **Exemple de résultat Significatif (Différence $>$ MES) :**

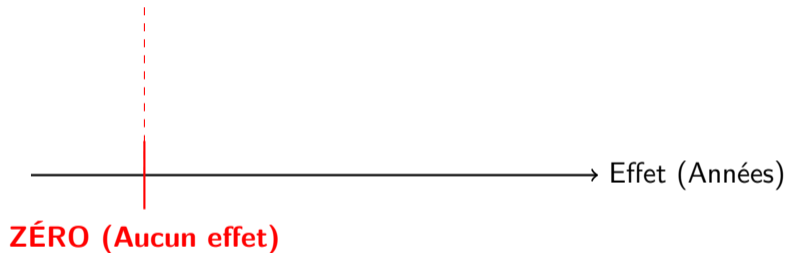
*"La bourse a augmenté la scolarisation de 1,5 an en moyenne. La marge d'erreur n'étant que de 0,5 an, cette différence est statistiquement significative. La politique est donc efficace."*

### **Exemple de résultat Non Significatif (Différence $<$ MES) :**

*"Le groupe traité gagne en moyenne 50€ de plus que le groupe contrôle. Cependant, la marge d'erreur étant de 80€, cette différence n'est pas statistiquement significative. Nous ne pouvons pas affirmer que la politique a un impact réel."*

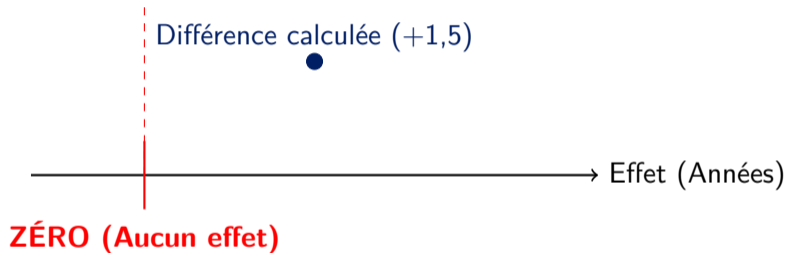
## 15.5 Visualiser un résultat SIGNIFICATIF

**Différence mesurée** : +1,5 an.    **MES** :  $\pm 0,5$  an.



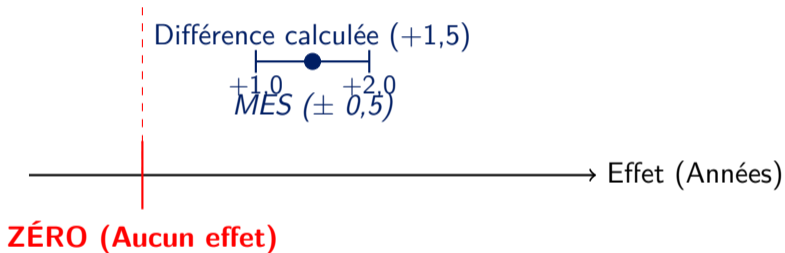
## 15.5 Visualiser un résultat SIGNIFICATIF

**Différence mesurée** : +1,5 an.    **MES** :  $\pm 0,5$  an.



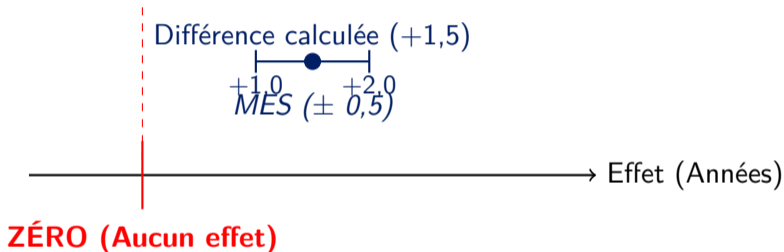
## 15.5 Visualiser un résultat SIGNIFICATIF

**Différence mesurée** : +1,5 an.    **MES** :  $\pm 0,5$  an.



## 15.5 Visualiser un résultat SIGNIFICATIF

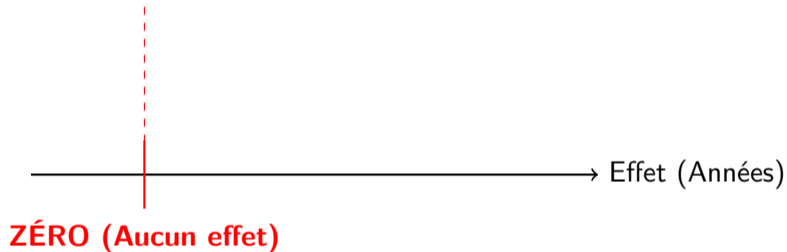
**Différence mesurée** : +1,5 an.    **MES** :  $\pm 0,5$  an.



**Conclusion** : La fourchette va de +1,0 à +2,0. Elle ne touche pas le Zéro.  
→ L'effet positif est prouvé.

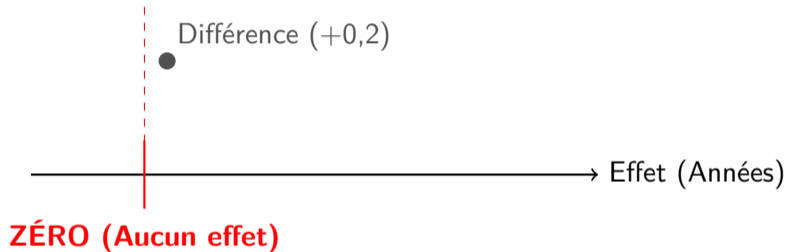
## 15.6 Visualiser un résultat NON SIGNIFICATIF

**Différence mesurée** : +0,2 an.    **MES** :  $\pm 0,5$  an.



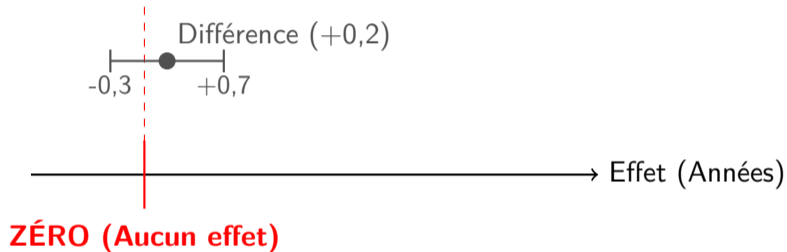
## 15.6 Visualiser un résultat NON SIGNIFICATIF

**Différence mesurée** : +0,2 an.    **MES** :  $\pm 0,5$  an.



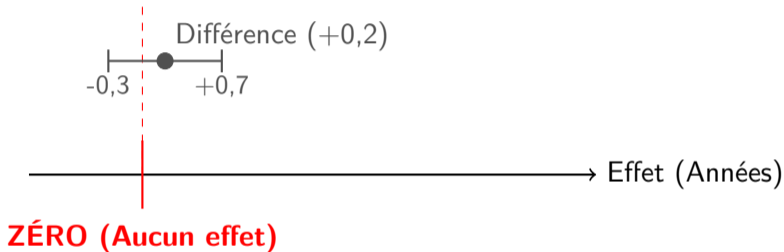
## 15.6 Visualiser un résultat NON SIGNIFICATIF

**Différence mesurée** : +0,2 an.    **MES** :  $\pm 0,5$  an.



## 15.6 Visualiser un résultat NON SIGNIFICATIF

**Différence mesurée** : +0,2 an.    **MES** :  $\pm 0,5$  an.



**Conclusion** : La fourchette croise le Zéro (l'effet réel pourrait être nul, ou même négatif).

→ La différence mesurée est "noyée" dans la marge d'erreur.

## 15.7 Application sur vos données

Maintenant que l'on comprend la mécanique, retournons au fichier Excel.

### Étape 3 : Valider votre résultat sur le Ghana

1. Ouvrez le fichier excel `expe_lycee_Ghana_MES`.
2. Regardez la **Différence** d'années d'études que vous avez calculée.
3. Regardez la **MES** indiquée dans la case correspondante.
4. Posez-vous la question : **La Différence est-elle supérieure à la MES ?**

*Si oui, vous venez de prouver statistiquement l'efficacité de la politique d'Esther Duflo !*

## Votre mission : Étape 3 (Valider le résultat)

**Fichier** : excel expe\_lycee\_Ghana\_MES

- ▶ **Action** : Reprenez la différence que vous avez calculée à l'Étape 2.
- ▶ **Vérification** : Est-elle supérieure à la Marge d'Erreur Statistique fournie dans le tableau ?
- ▶ **Conclusion** : L'article de Duflo est-il scientifiquement prouvé ?

# Conclusion

## Ce qu'il faut retenir de cette séance :

1. Placer correctement ses variables (Abscisses = Cause, Ordonnées = Effet).
2. Le tirage au sort (loterie) permet d'éliminer le biais de sélection pour prouver la causalité.
3. Relier une variable Quali (Loterie) à une variable Quanti (Années d'études) permet de mesurer concrètement l'impact d'une politique publique.