

# Atelier de formation sur l'évaluation d'impact contrefactuelle (EIC)

*3ème édition*

**Analyse de données pour le suivi et l'EIC**

**C4ED – EUTF**

Octobre 2023

**Bienvenue au Jour 1 de  
l'Atelier de formation sur l'évaluation d'impact contrefactuelle (EIC)  
Analyse de données pour le suivi et l'EIC**

Le contenu de cet atelier a été produit avec le soutien financier de l'Union Européenne. Son contenu relève de la seule responsabilité du C4ED et ne reflète pas nécessairement les opinions de l'Union Européenne.



# Introduction



**BOUTON DE  
SOURDINE**



**QUESTIONS**



**RETOUR  
D'INFORMATION**



**BOUTON DE  
SOURDINE**



**QUESTIONS**



**RETOUR  
D'INFORMATION**



**BOUTON DE  
SOURDINE**




**QUESTIONS**



**RETOUR  
D'INFORMATION**



# Poser des questions

- Veuillez écrire vos questions dans l'espace de discussion
- “Likez”  les questions des autres pour que nous sachions qu'elles sont particulièrement pertinentes pour vous aussi
- Carolin gardera un oeil sur les questions pour que nous essayions de répondre à toutes
- Profitez des pauses plus longues pour poser davantage de questions



**BOUTON DE  
SOURDINE**



**QUESTIONS**



**RETOUR  
D'INFORMATION**





# Questions et retour d'expérience

- N'hésitez pas à faire des suggestions
- N'hésitez pas à nous faire part de vos commentaires
- Plus de réactions et de questions (en particulier pour la séance de questions et réponses)



# Aperçu du Jour 1



- Tout d'abord, nous passerons brièvement en revue les *bases de l'évaluation d'impact contrefactuel (EIC)*, les méthodes courantes d'**identification de l'impact** et l'importance de disposer de *données de haute qualité pour l'EIC*
- La prochaine session portera sur les *statistiques descriptives de base* - comment les calculer, les présenter et les interpréter.
- Enfin, le dernier thème de la journée sera celui des *tests statistiques*
- Nous partagerons des ressources externes utiles et des études de cas sur les EIC.
- <https://europa.eu/capacity4dev/>



# Jour 1 – Ordre du jour

10:00 – 10:30	<b>Accueil par EUTF et C4ED</b>
10:30 – 11:20	<b>Session 1:</b> Récapitulatif de l'année 1 (méthodes EIC) et de l'année 2 (collecte de données)
11:20 – 11:30	Pause (10 minutes)
11:30 – 12:30	<b>Session 2:</b> Statistiques descriptives pour le suivi et l'EIC
12:30 – 12:45	<b>Questions – Réponses (Q&amp;R)</b>
12:45 – 13:45	Pause déjeuner (1 heure)
13:45 – 14:45	<b>Session 3:</b> Tests statistiques
14:45 – 15:00	Pause (15 minutes)
15:00 – 15:30	<b>Travail en groupe – Partie 1</b>
15:30 – 15:50	<b>Quiz 1</b>
15:50 – 16:00	<b>Q&amp;R et Clôture du Jour 1</b>

# **Session 1 : Récapitulatif de l'année 1 (méthodes EIC) et de l'année 2 (collecte de données)**

**C4ED - EUTF**

Octobre 2023



# RECAP ANNÉE 1

---

## Méthodes d'évaluation de l'impact contrefactuel (EIC)



- Examiner le "pourquoi", le "quoi" et le "comment" de l'évaluation d'impact contrefactuelle (EIC)
- Revoir l'intuition des méthodes d'évaluation expérimentale → Essais contrôlés randomisés
- (brièvement) Examiner l'intuition de deux méthodes quasi-expérimentales importantes (l'appariement et la double différence).

## Pourquoi réaliser une évaluation d'impact contrefactuelle ?

- Déterminer si une intervention a un **effet causal** sur les variables de résultat, **comment (le mécanisme causal)** et dans quelle **mesure**.
- Pour **savoir** quelle stratégie d'intervention fonctionne le mieux
- Pour aider à **prendre des décisions fondées sur des données probantes**



## Qu'est-ce qu'une évaluation d'impact contrefactuelle ?

- **Impact** : l'effet sur les variables de résultat d'intérêt que le programme/la politique *provoque* directement et qui peut être directement *attribué* au programme → lien de **causalité**
  - **Contrefactuel** : le résultat qui aurait été observé/mesuré pour les bénéficiaires du programme s'ils *n'avaient pas* bénéficié du programme.
- Problème fondamental : il est impossible de mesurer ou d'observer le contrefactuel.
- Les cibles du programme reçoivent le programme ou non, nous ne pouvons pas les observer dans les deux scénarios en même temps
- Solution : **utiliser un groupe de contrôle/comparaison** pour simuler la situation contrefactuelle.





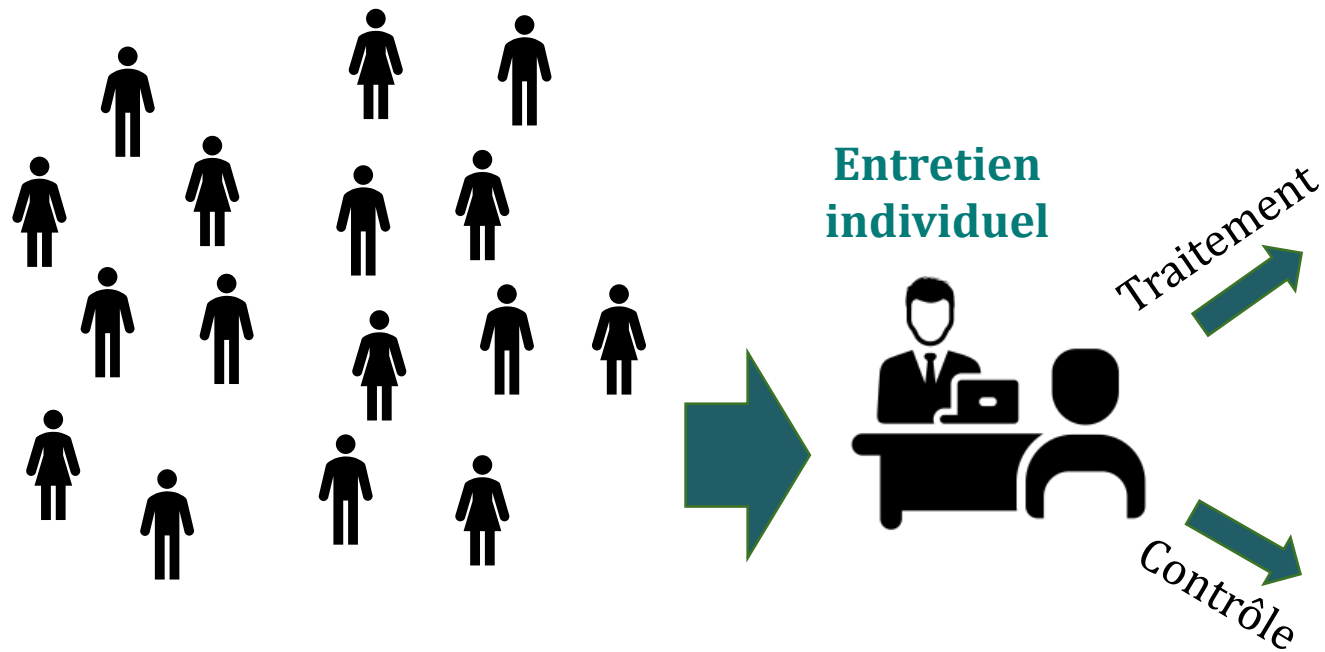
## Comment est conçue une EIC ?

**Objectif** : reproduire la situation contrefactuelle avec un groupe témoin

- Le groupe de comparaison doit :
  - Avoir les mêmes caractéristiques (en moyenne) que le groupe de traitement
  - Ne pas être exposé au programme
  - Réagir au programme de la même manière que le groupe de traitement (s'il y participait)
- L'approche la plus appropriée est sélectionnée en fonction de la conception et du contexte de l'intervention, du calendrier, de la disponibilité des données et du budget :
  - Méthodes expérimentales
  - Méthodes quasi-expérimentales



### Processus de sélection pour un programme de formation professionnelle des jeunes



Candidats : 2 000 (1 000 hommes / 1 000 femmes)

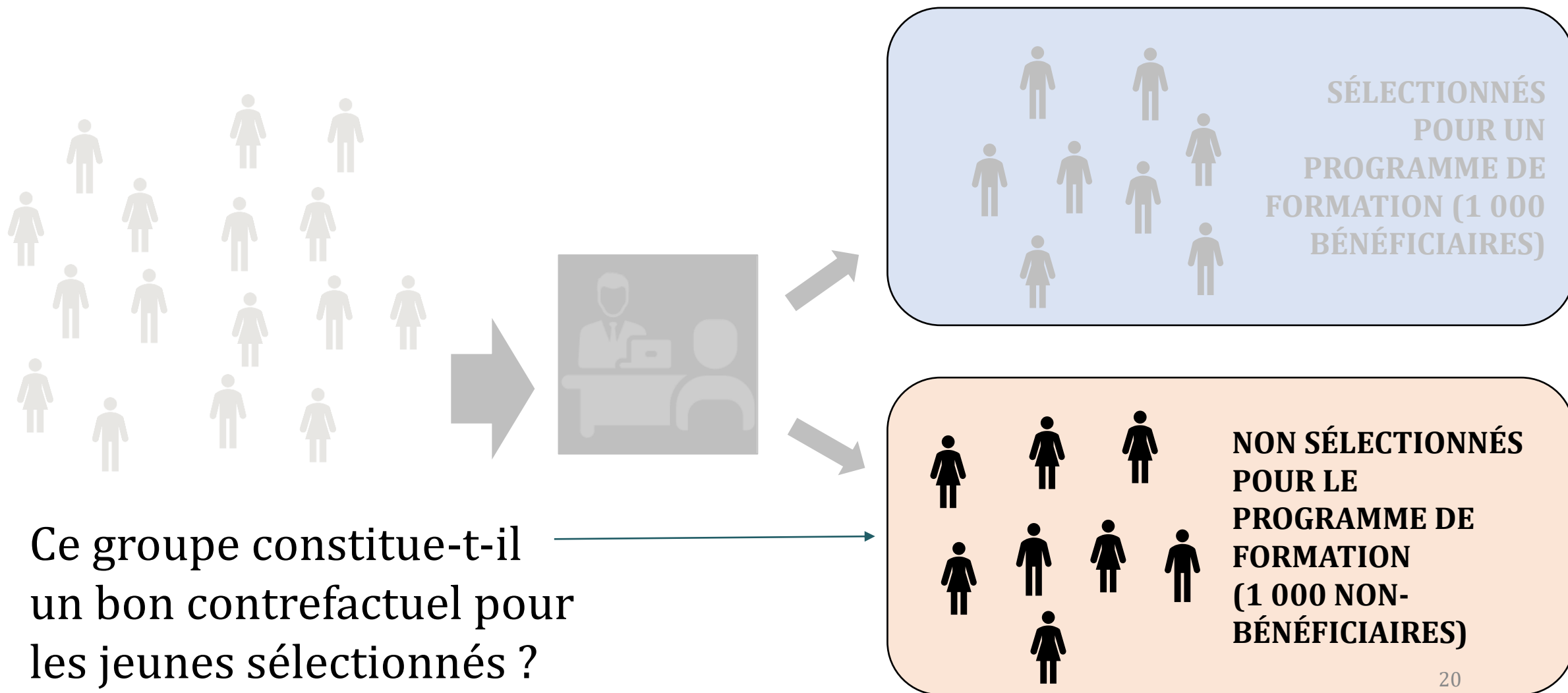
**SÉLECTIONNÉS  
POUR UN  
PROGRAMME DE  
FORMATION  
(1 000  
BÉNÉFICIAIRES)**

**NON SÉLECTIONNÉS  
POUR LE  
PROGRAMME DE  
FORMATION  
(1 000 NON-  
BÉNÉFICIAIRES)**



# Simulation d'un contrefactuel

## Exemple





## **Les candidats rejetés constituent-ils un bon groupe de comparaison/contrefactuel ?**

- Il est probable que les « meilleurs » candidats soient choisis
  - Le scénario contrefactuel est composé de candidats « moins bons » et susceptibles d'être très différents de ceux qui ne le sont pas
- Pour un coordonnateur de programme, il peut être souhaitable que les candidats les plus aptes s'inscrivent au programme
- Cependant, pour un évaluateur cherchant à mesurer l'impact du programme, ces différences signifient que les groupes ne sont pas facilement comparables
- **Compte tenu du processus de sélection, ils ne représenteraient pas un bon contrefactuel. Pourquoi ?**



# Comparaison des groupes

## Caractéristiques observées

- Supposons que la variable de résultat qui nous intéresse soit l'emploi ou le revenu
- Les différences de résultats entre des groupes présentant des caractéristiques différentes ne peuvent pas être *attribuées* au programme avec une certitude de 100 %



Variable	Moyenne
Age	30
Années de scolarité	10
A déjà eu un emploi	60%
Revenu des parents	5 000



Variable	Moyenne
Age	24
Années de scolarité	6
A déjà eu un emploi	46%
Revenu des parents	3 400



# Comparaison de groupes

## Caractéristiques non observées

→ Même si nous prenons en compte les différences observables, comment tenir compte des différences non observables (ou non mesurables) ?



Motivation



Confiance en soi



Détermination



Motivation



Confiance en soi



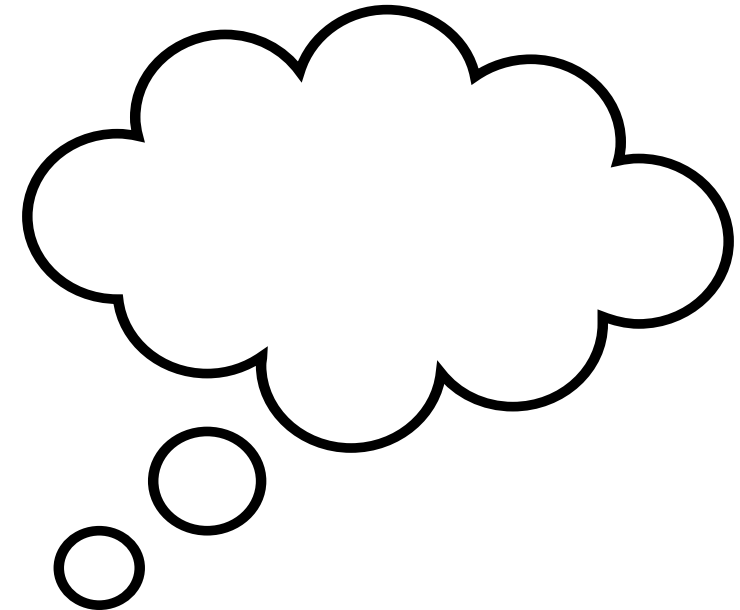
Détermination





# Choisir le contrefactuel

- Dans l'exemple précédent, le groupe des candidats rejetés ne constitue pas un bon contrefactuel
- Comment pourrions-nous adapter/modifier le processus de sélection des candidats afin de pouvoir utiliser les candidats rejetés comme contrefactuel ?





Center for Evaluation  
and Development



# Essai Contrôlé Randomisé (ECR)

---

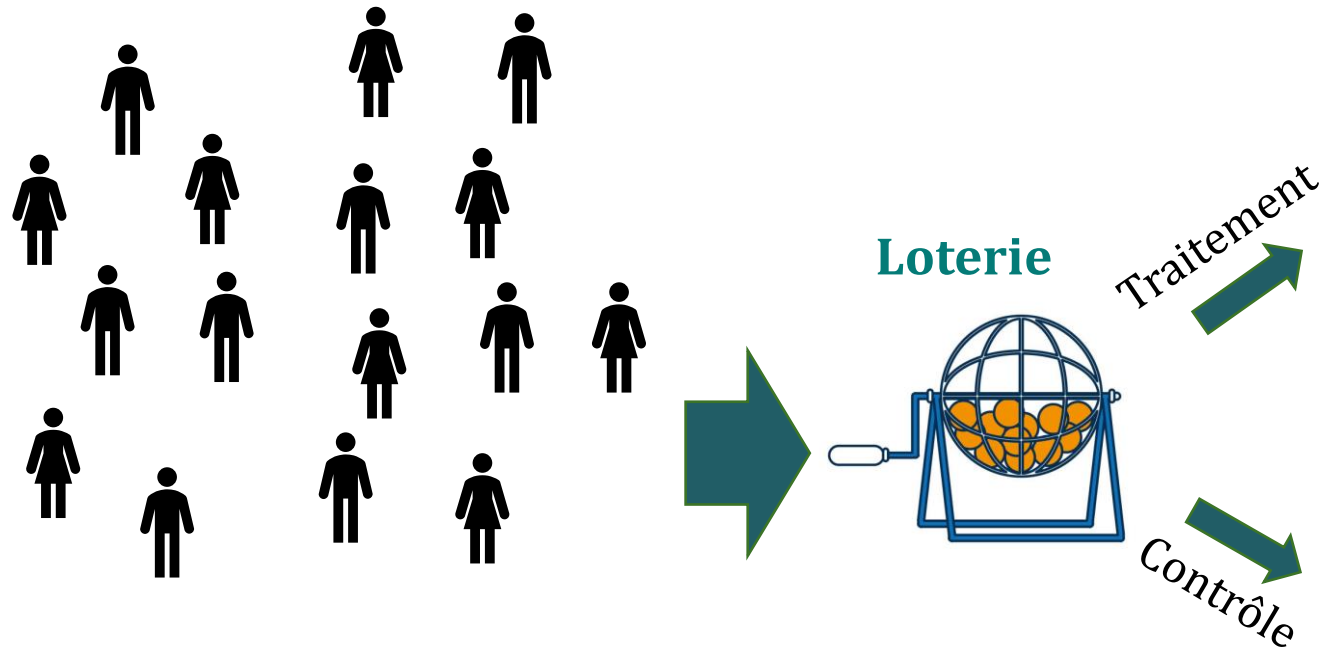




- Si l'évaluation est intégrée à la mise en œuvre du programme, il est possible de créer un scénario contrefactuel en recourant à une loterie pour décider qui est sélectionné
  - C'est ce qu'on appelle l'**assignment aléatoire**
- Pourquoi cela est-il utile ?
  - En supposant que le nombre de candidats soit suffisamment important, les deux groupes choisis de manière aléatoire seront similaires (en moyenne)



### Processus de sélection pour un programme de formation professionnelle des jeunes



Candidats : 2 000 (1 000 hommes / 1 000 femmes)

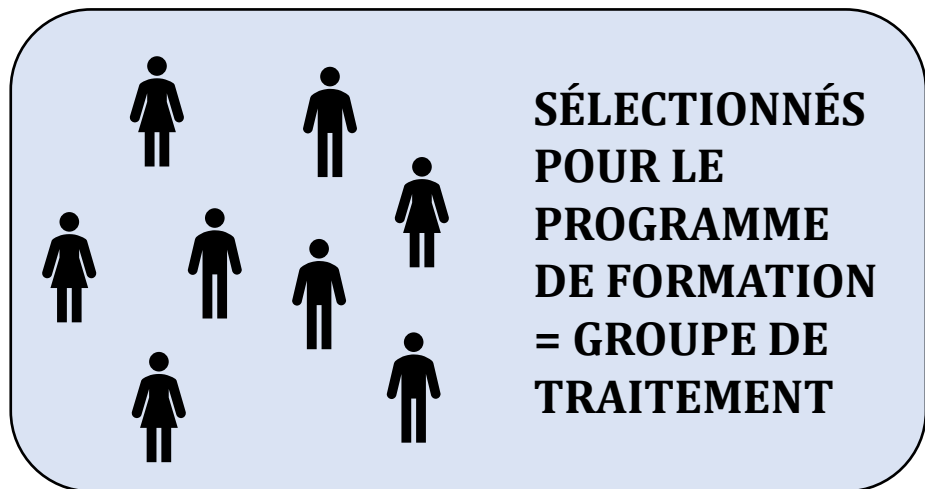
**SÉLECTIONNÉS  
POUR UN  
PROGRAMME DE  
FORMATION  
(1 000  
BÉNÉFICIAIRES)**

**NON SÉLECTIONNÉS  
POUR LE  
PROGRAMME DE  
FORMATION  
(1 000 NON-  
BÉNÉFICIAIRES)**

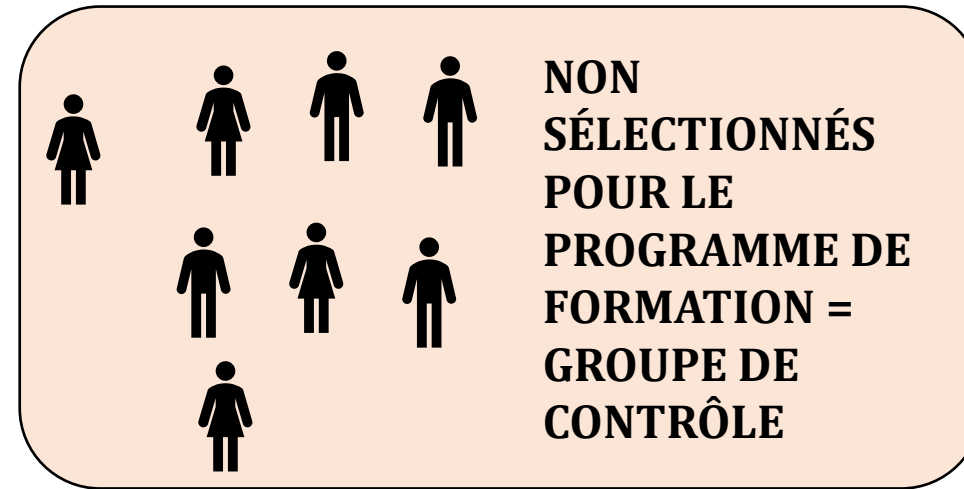


# Comparaison de groupes

## Assignation aléatoire – Caractéristiques observées



Variable	Moyenne
L'âge	27
Années de scolarité	8
A déjà eu un emploi	52%
Revenu des parents	4 300

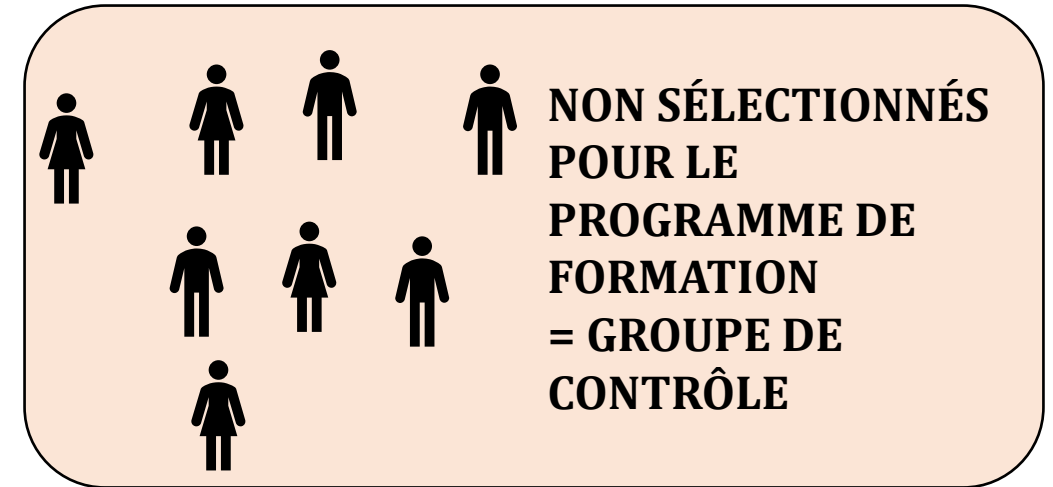
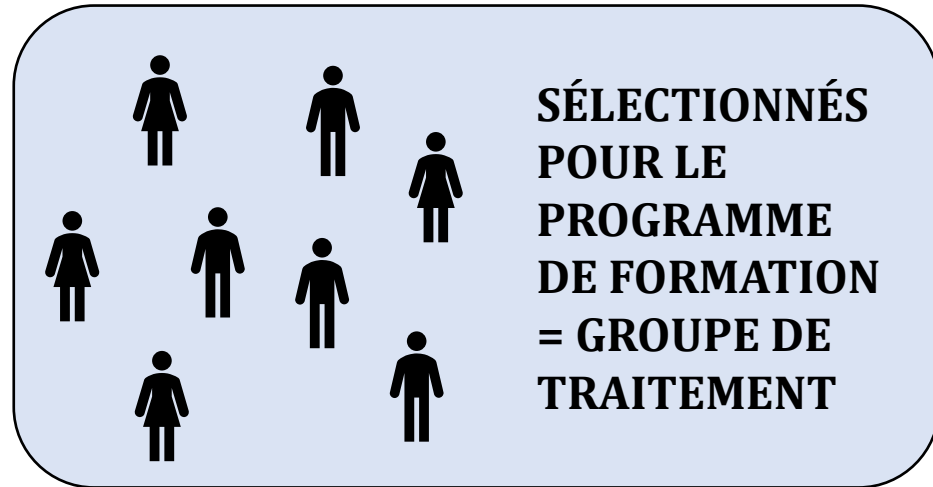


Variable	Moyenne
L'âge	27
Années de scolarité	8
A déjà eu un emploi	54%
Revenu des parents	4 100



# Comparaison de groupes

## *Assignation aléatoire – Caractéristiques non observées*



Motivation



Confiance en soi



Détermination



Motivation



Confiance en soi



Détermination

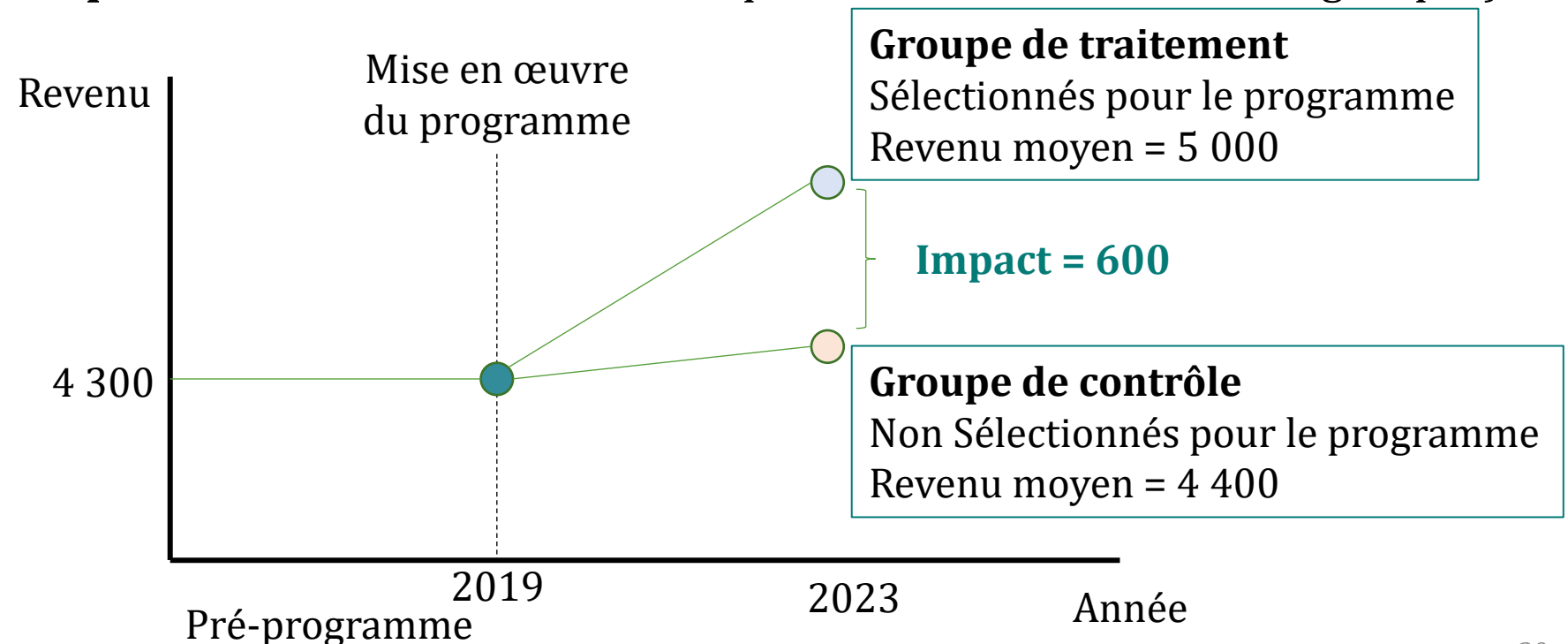




# Mesurer l'impact

## Essai contrôlé randomisé (ECR)

- Étant donné que l'assignation aléatoire crée deux groupes qui sont (en moyenne) comparables au début du programme, l'**impact** peut être mesuré simplement comme la différence de résultats entre les deux groupes après le programme.
- En d'autres termes, les différences de résultats entre les groupes peuvent être *attribuées* au programme (**parce que** toutes les autres caractéristiques sont similaires entre les groupes).





## **Que faut-il** randomiser ?

- Tout aspect du programme que l'équipe de mise en œuvre *contrôle entièrement*
- Il faut souvent faire preuve de créativité et avoir une connaissance approfondie du programme → La préparation d'un ERC prend du temps !

## **Quand** randomiser ?

- *Avant le* début du programme → l'ERC doit être inclus dans la mise en œuvre du programme

## **Comment** randomiser ?

- Loterie simple
- Groupes de traitement multiples → pour tester différentes modalités de traitement
- Mise en place progressive → traitement retardé pour une partie des bénéficiaires du programme
- Encouragement → Tous les individus éligibles ont accès au programme, mais certains bénéficiaires sont activement encouragés à y participer



# Et si l'on ne peut pas randomiser?

## *Méthodes quasi-expérimentales*



- Il n'est parfois pas possible de procéder à une assignation aléatoire des programmes :
  - La randomisation peut ne pas être socialement ou politiquement acceptable
  - La randomisation n'est pas toujours possible
  - L'EIC n'est conçue qu'après le début de la mise en œuvre
- Il peut être possible d'utiliser des **méthodes quasi-expérimentales** pour construire un scénario contrefactuel.
- Nous présentons ici l'intuition de base de deux méthodes dites quasi-expérimentales, à savoir l'**appariement** et la **double différence**.



Center for Evaluation  
and Development



# Méthodes quasi-expérimentales

---

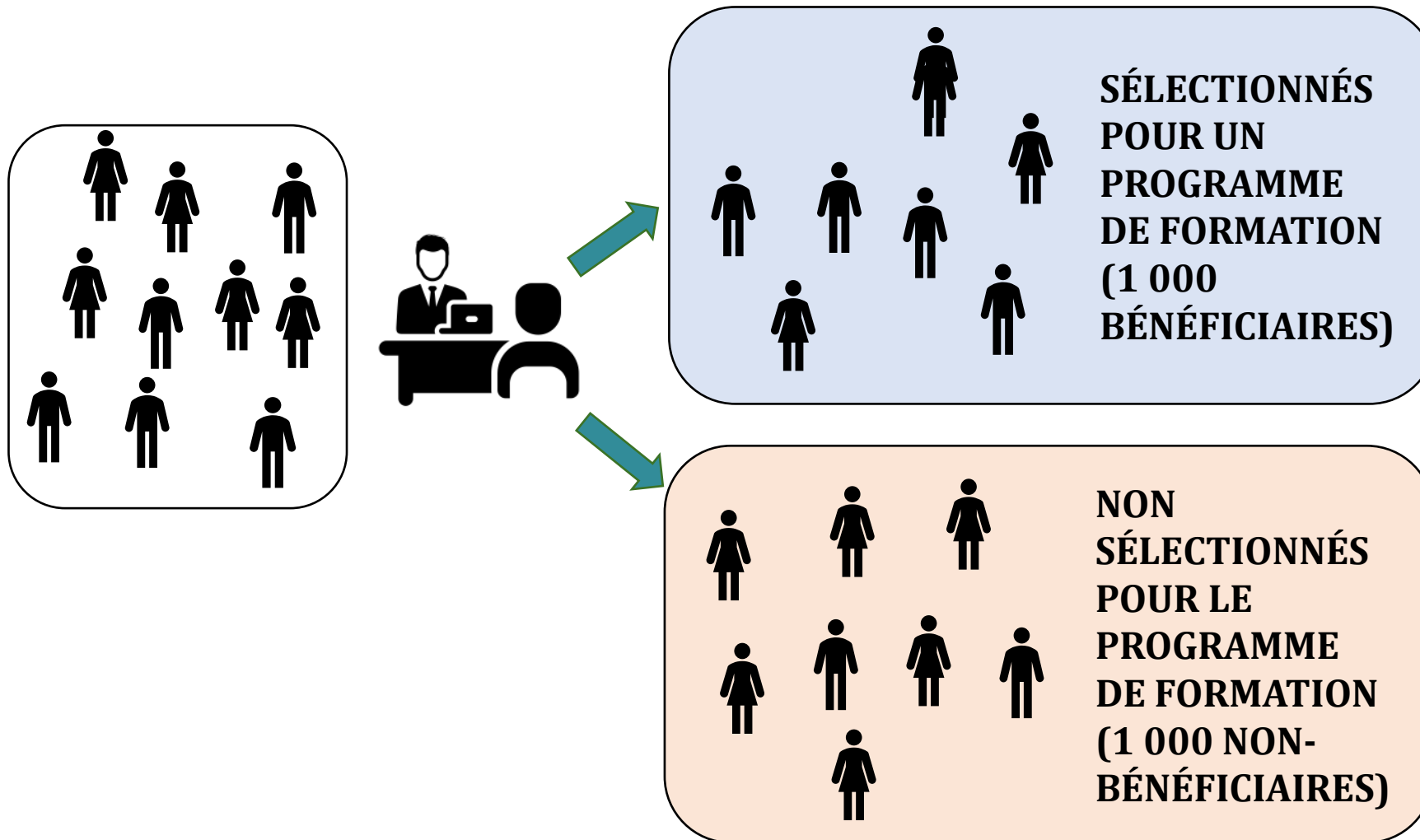
## Appariement





# Sélection non aléatoire, les groupes sont différents

Nombre de candidats : 2 000 (1 000 hommes / 1 000 femmes)

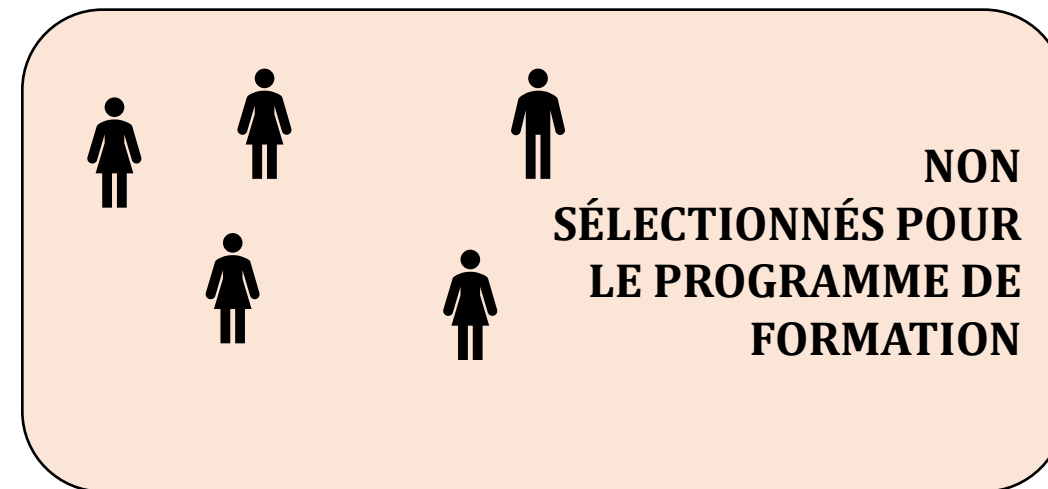


Variable	Moyenne
Age	30
Années de scolarité	10
A déjà eu un emploi	60%
Revenu des parents	5 000

Variable	Moyenne
Age	24
Années de scolarité	6
A déjà eu un emploi	46%
Revenu des parents	3 400

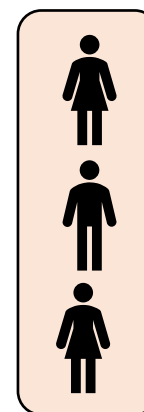
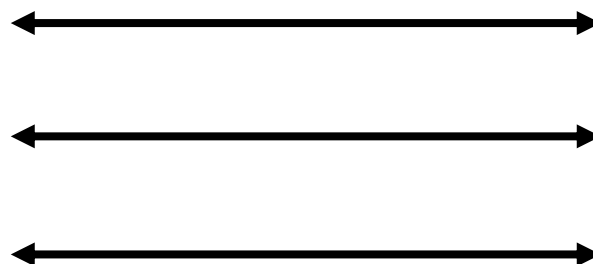


# Appariement



Rechercher des paires  
de personnes ayant  
des caractéristiques  
similaires

TRAITEMENT



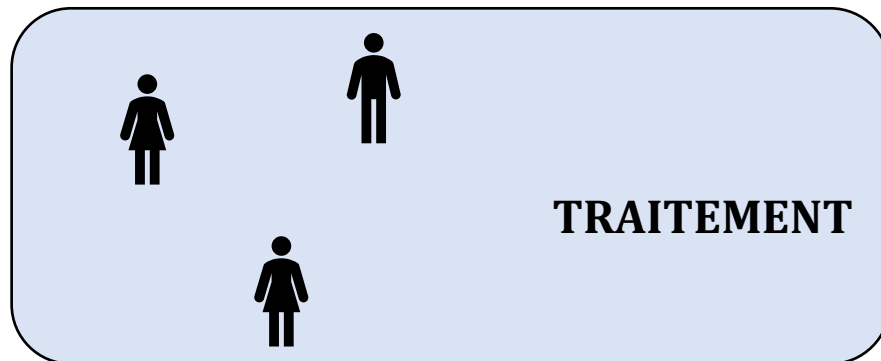
COMPARAISON

**Ne doit prendre en compte que les caractéristiques antérieures au programme ou les caractéristiques qui ne changent pas au fil du temps.**

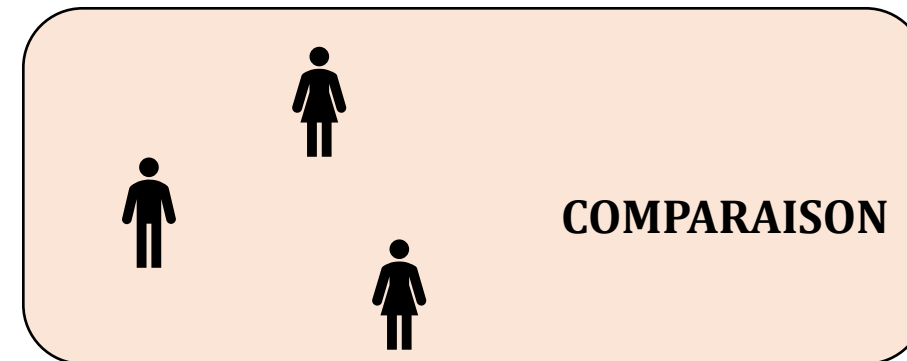


## POINTS CLÉS

- Les variables d'appariement doivent être mesurées *avant le* programme (ou ne pas changer au fil du temps).
- Les opérations d'appariement ne portent que sur les caractéristiques **observées**



Variable	Moyenne
Age	28
Années de scolarité	8
A déjà eu un emploi	54%
Revenu des parents	4 200



Variable	Moyenne
L'âge	27
Années de scolarité	7
A déjà eu un emploi	53%
Revenu des parents	4 300



Center for Evaluation  
and Development



# Méthodes quasi-expérimentales

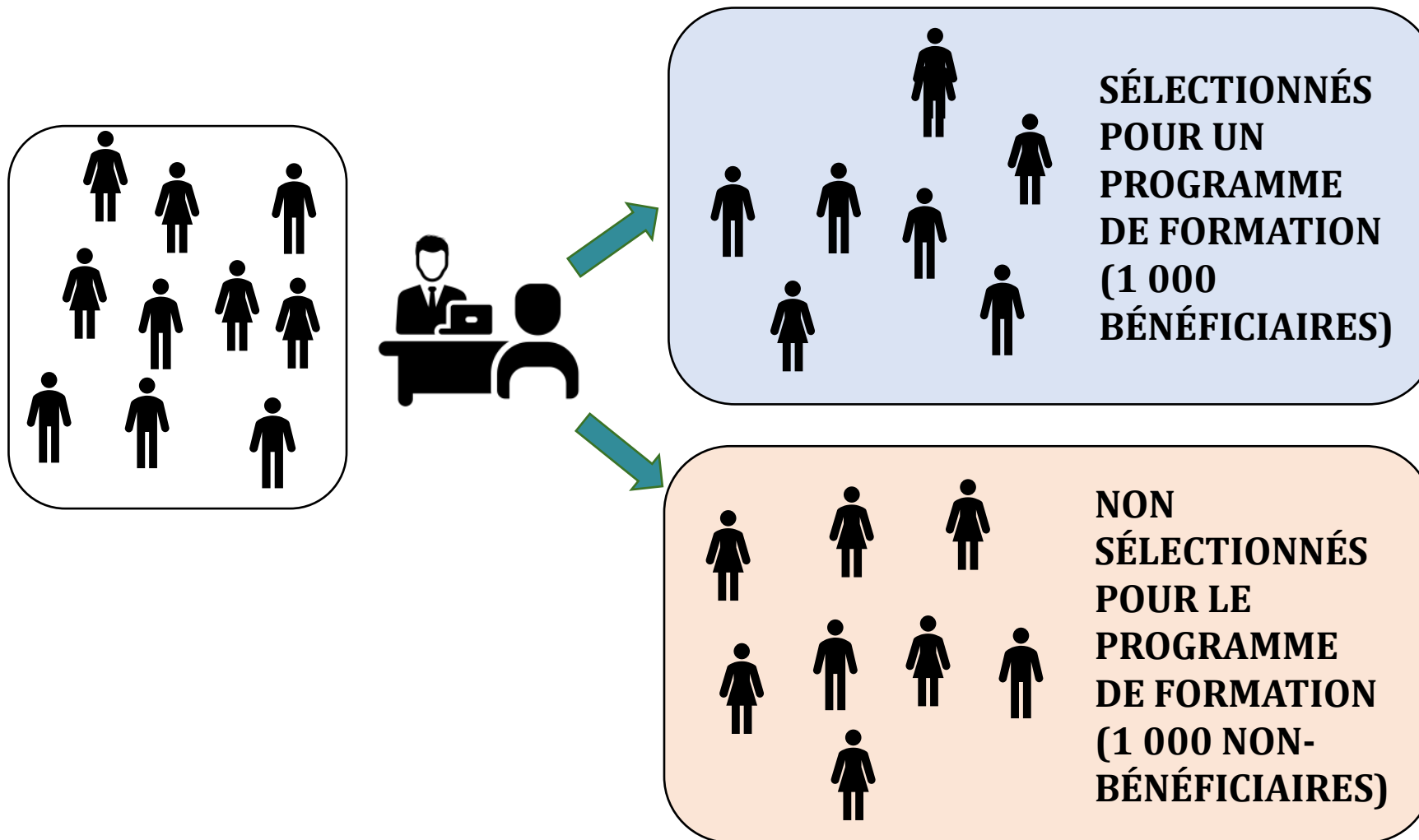
---

## Double différence



# Sélection non aléatoire, les groupes sont différents

Nombre de candidats : 2 000 (1 000 hommes / 1 000 femmes)



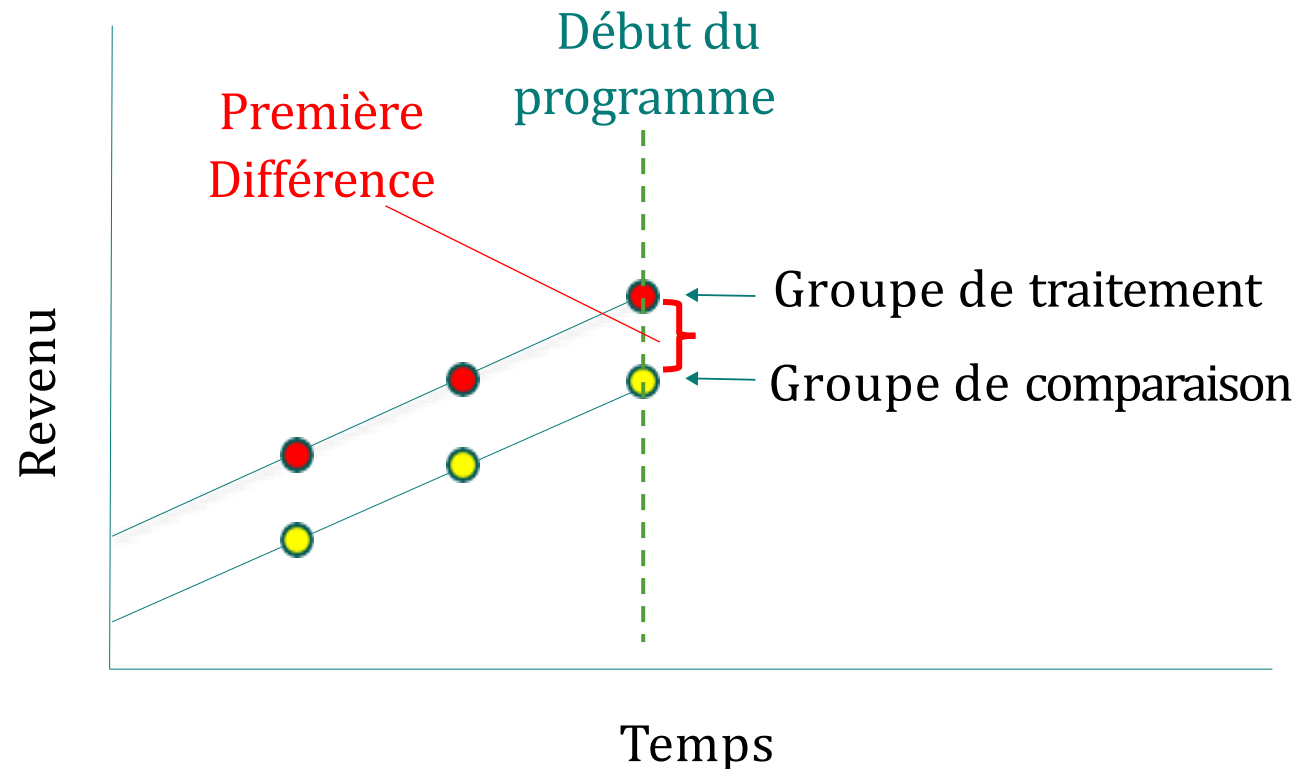
Variable	Moyenne
Age	30
Années de scolarité	10
A déjà eu un emploi	60%
Revenu des parents	5 000

Variable	Moyenne
Age	24
Années de scolarité	6
A déjà eu un emploi	46%
Revenu des parents	3 400



# Double Différence

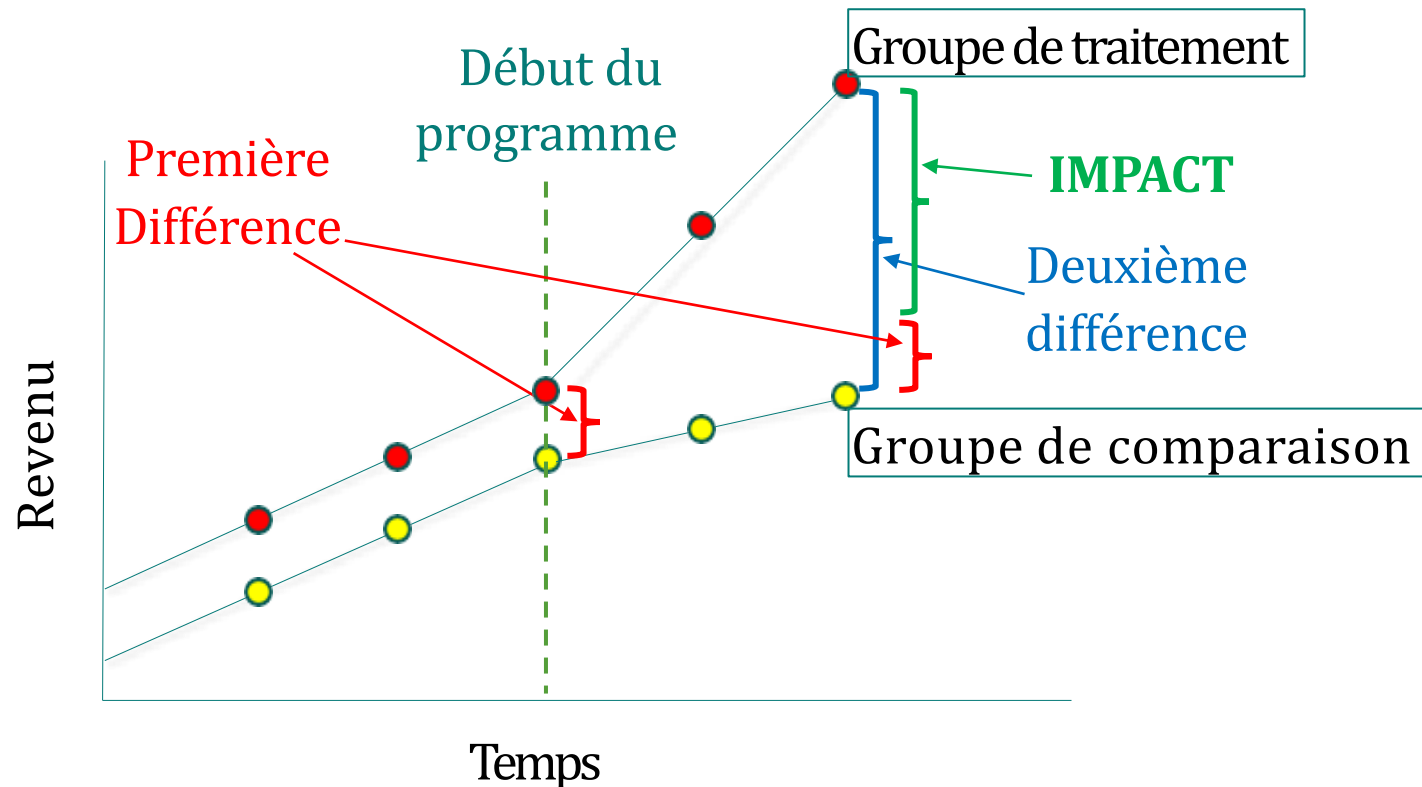
- Dans l'approche de la double différence, on accepte que les groupes de traitement et de comparaison soient *différents*.
- **IMPORTANT** : cette approche nécessite de disposer de données sur les deux groupes *avant le début* du programme.





# Double Différence

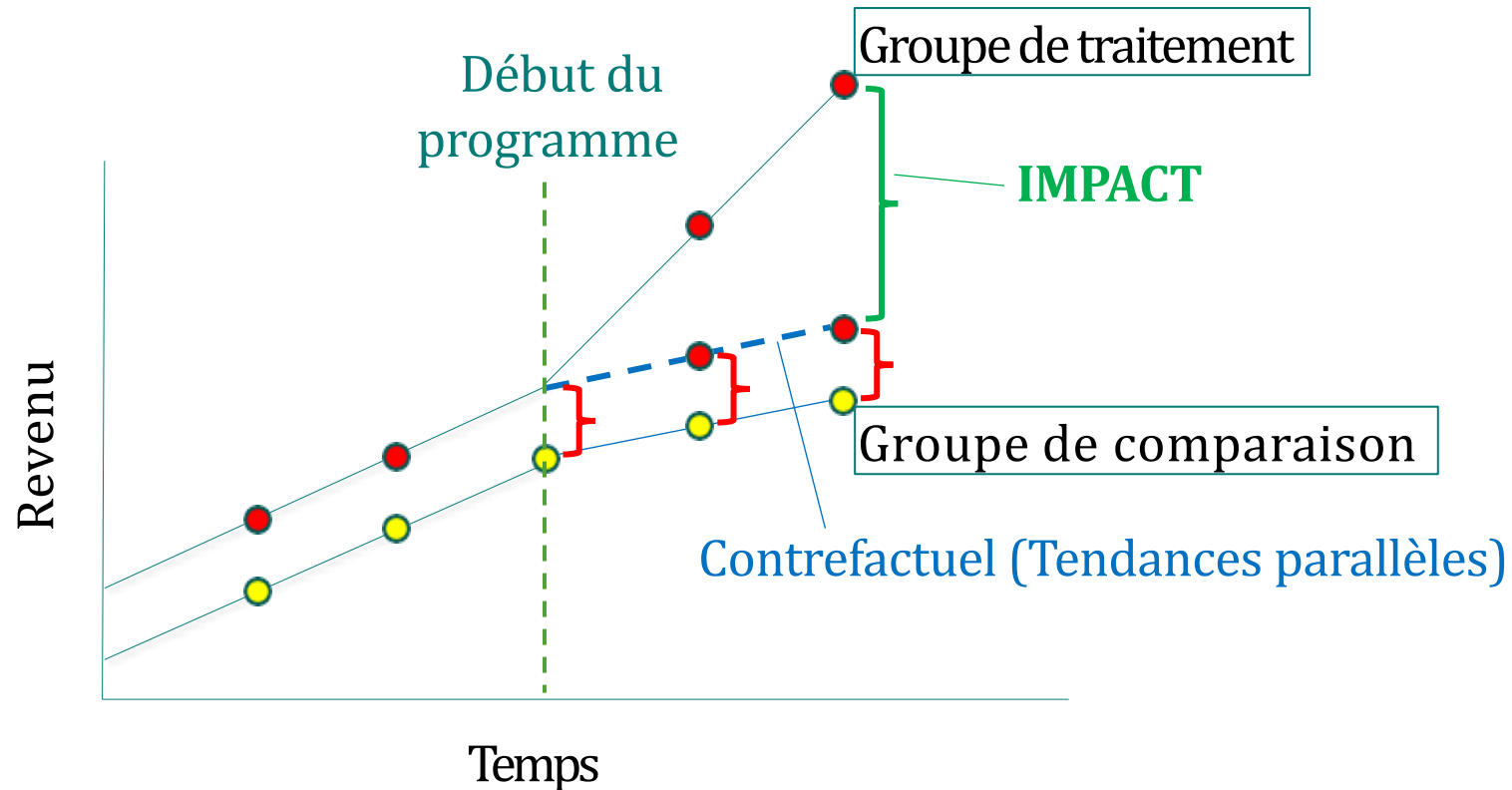
- Des données sur les deux groupes doivent aussi être collectées ultérieurement, *après* que le programme a fonctionné pendant un certain temps
- La différence observée après le début du programme est ajustée en soustrayant la première différence observée *avant* le programme pour obtenir l'estimation de l'impact





# Double Différence

- L'approche de la double différence repose sur l'**hypothèse de tendances parallèles** :  
→ *Nous supposons que le groupe de traitement aurait évolué de la même manière que le groupe de comparaison s'il n'avait pas bénéficié du programme (ligne bleue en pointillés)*







D'autres approches quasi-expérimentales de l'EIC sont également possibles :

- Variable instrumentale
- Discontinuité de la régression



- [Atelier de formation sur l'évaluation d'impact contrefactuelle \(EIC\) - PowerPoint Slides](#)

## **Livres**

- [Banque mondiale, Impact Evaluation in Practice - Second Edition \(Livre\)](#)

## **Vidéos**

- [InterAction, Introduction à l'évaluation d'impact](#)
- [Esther Duflo, Essais contrôlés randomisés et élaboration des politiques dans les pays en développement](#)

## **Podcasts**

- [IEU Talks Episode 2 : Le pouvoir de l'évaluation d'impact dans la coopération au développement](#)
- [Démontrer l'impact \(parties 1+2\)](#)

## RECAP ANNÉE 2

---

### Collecte de données (CD) pour l'EIC



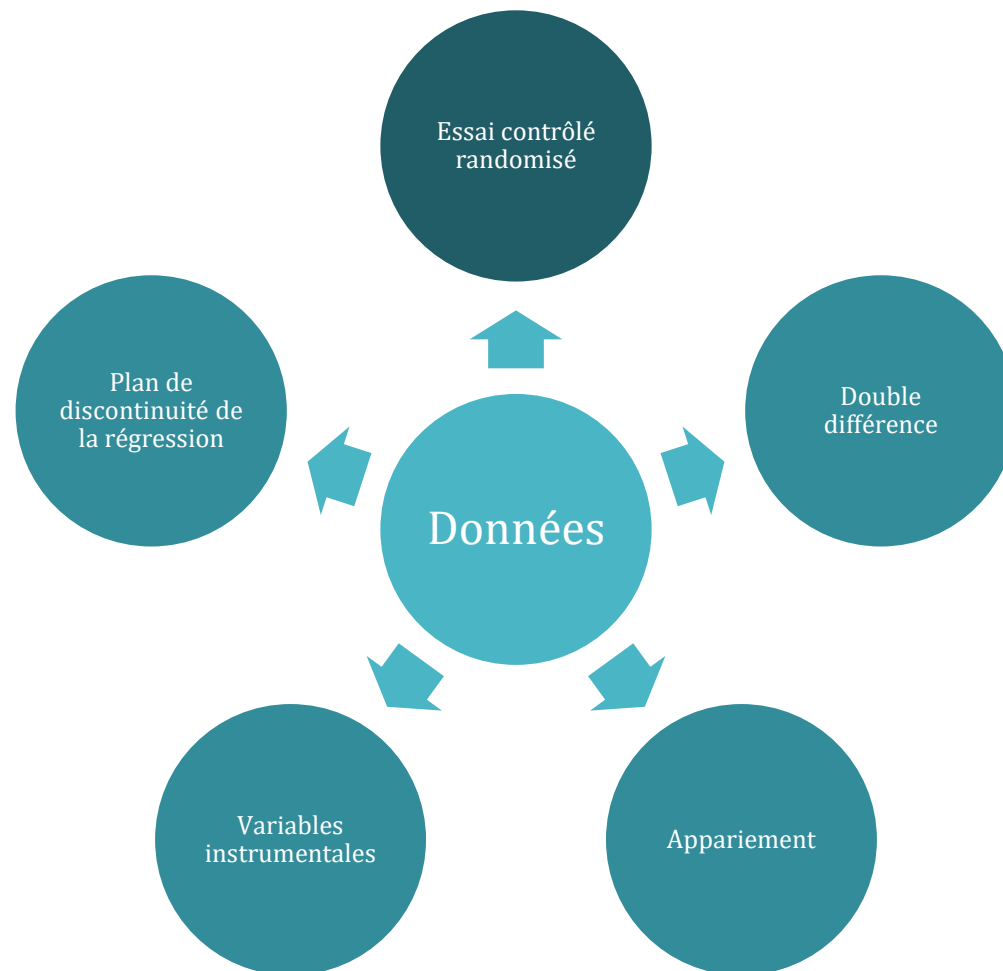
# Récapitulation de l'année 2 - Objectifs



- Examiner les aspects clés de la conceptualisation/préparation de la collecte de données
- Revoir les idées clés sur l'échantillonnage et les bases d'échantillonnage
- Examiner les considérations pratiques telles que la qualité des données et l'éthique de la recherche



# Pertinence de la collecte de données dans cadre de l'EIC





- La collecte et la qualité des données sont essentielles pour l'EIC car elles permettent la construction d'un contrefactuel valide et fiable.
- Des données de qualité sont essentielles pour répondre aux questions d'évaluation et mesurer l'impact des programmes.

“Garbage in, garbage out”



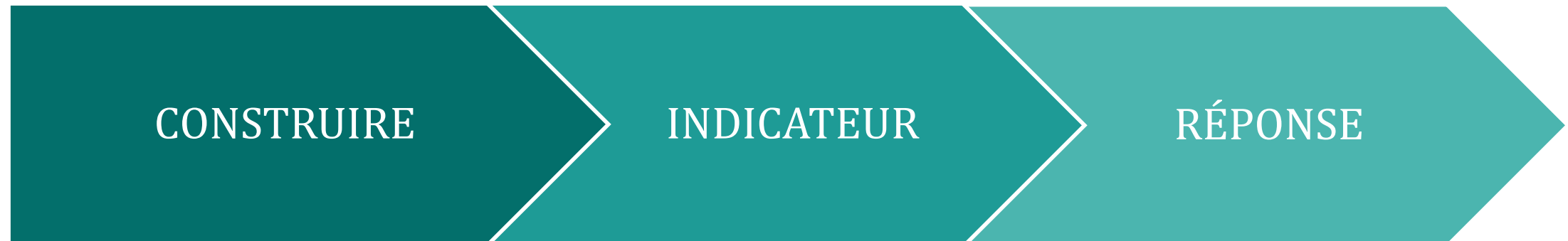
Your analysis is as good as your data.



- Les données relatives à l'EIC peuvent être obtenues à l'aide de méthodes mixtes :
  - Données quantitatives et qualitatives
- Les méthodes mixtes permettent d'approfondir les connaissances :
  - Les expériences et les motivations des bénéficiaires, des responsables de la mise en œuvre et des parties prenantes
  - Efficacité du programme et facteurs contextuels qui déterminent l'impact
- Les données pour l'EIC vont au-delà des résultats/impacts : données de suivi, données secondaires pour l'étalonnage et variables influençant les résultats.
- Les données disponibles déterminent le choix de l'approche de l'EIC



*.....Pour être bien préparé lors de la conception de la collecte de données, il faut connaître les questions et les objectifs de la recherche.....*



Définir les concepts  
de la question  
d'évaluation

Comment pouvons-  
nous essayer de  
mesurer cela dans  
le monde réel ?

Questions au sein  
d'une enquête





## CONSTRUIRE

Quels sont les effets des interventions sur les moyens de subsistance en termes de **bien-être économique** des réfugiés et des communautés d'accueil ?

## INDICATEUR

- Revenu moyen
- **Statut de l'emploi**
- Sécurité de l'emploi
- Propriété de l'entreprise
- Performance des entreprises
- Propriété des actifs

## RÉPONSE

1. Un employé rémunéré
2. Un travailleur rémunéré sur l'exploitation agricole du ménage
3. Un employeur
4. Travailleur non rémunéré
5. Stage
6. Aucune de ces réponses



- Le développement d'outils est essentiel pour la collecte de données en vue d'une EIC fiable.
  - Exploiter la littérature et les outils existants
  - Prétester les outils (au bureau et sur le terrain) et affiner les questions et les réponses avant de les utiliser dans les collectes de données.
- Important pour éviter les erreurs de mesure
  - Questions et enquête mal conçues
  - Défis cognitifs pour répondre à la question
  - Biais de désirabilité sociale



- **Population cible** = le groupe pour lequel les données de l'enquête sont utilisées pour obtenir des informations
- **Base d'échantillonnage** = listes ou procédures utilisées pour identifier toutes les unités de la population cible
- **Échantillon** = groupe d'unités sélectionnées dans le cadre de l'échantillonnage à partir duquel la mesure sera effectuée.
- **Répondants** = éléments mesurés avec succès à partir de l'échantillon



- Dans la plupart des EIC, les données ne peuvent pas être collectées auprès de toutes les unités. → se concentre sur un *échantillon*.
- Le cadre d'échantillonnage comprend idéalement toutes les unités de la population sur laquelle porte l'évaluation → recensement.
- Pour une EIC, la base de sondage est généralement une liste de toutes les unités qui.. :
  - Recevoient le programme
  - N'ont pas bénéficié du programme et sont identifiés comme le groupe contrefactuel
- L'échantillon est tiré de la base de sondage





# Cadre d'échantillonnage

## Le rôle des données du programme



- Lorsque l'échantillonnage n'est pas effectué correctement, il conduit à une **erreur de couverture**.  
→ Il est important de disposer d'un cadre d'échantillonnage **complet, valide et fiable**.
- Les bases d'échantillonnage pour le programme EIC s'appuient souvent sur les **données du programme** :
  - Base de données des candidatures et informations sur la sélection
  - Informations sur les remplaçants ou les abandons du programme
  - Coordonnées des participants
- S'assurer que l'évaluateur peut **utiliser efficacement les données du programme pour constituer une base d'échantillonnage** :
  - S'intégrer dans un système de suivi centralisé afin de rassembler toutes les données recueillies sur le terrain dans une seule base de données (en ligne).
  - Créer des liens reproductibles entre divers documents/sources de données (par exemple, un identifiant unique pour chaque participant).
  - Tenir à jour autant que possible
  - Inclure des contrôles d'assurance qualité de base - pas de doublons, correspondance des totaux



# Échantillonnage



- Le choix de l'échantillon dépend des questions auxquelles vous essayez de répondre avec vos données.
  - Si vous essayez de répondre quantitativement à des questions sur la population ou le programme - par ex :
    - Quel est le niveau d'éducation typique des participants à mon programme ?
    - Quel est le taux d'emploi des participants après l'obtention de leur diplôme ?
    - Quel est l'impact du programme sur le revenu ?
- L'objectif est de tirer un échantillon non biaisé afin d'obtenir les meilleures estimations pour la population.
- En règle générale, plus la taille de l'échantillon est importante, meilleure est l'estimation.



- Types d'échantillonnage :
  - Les méthodes d'**échantillonnage probabiliste** réduisent la possibilité de biais, car la possibilité qu'une personne soit sélectionnée pour faire partie de l'échantillon repose entièrement sur le hasard.
    - Échantillonnage aléatoire simple ; Échantillonnage aléatoire stratifié ; Échantillonnage par grappes
    - Approches les mieux adaptées à la partie quantitative d'un EIC
  - **L'échantillonnage non aléatoire** signifie que la sélection de l'échantillon n'est pas le fruit du hasard → échantillonnage de commodité, échantillonnage raisonné (parfois utilisé pour les études qualitatives)
    - L'étude et l'échantillon peuvent être guidés par les résultats
    - Peut être utile si le cadre d'échantillonnage n'est pas clair
- La taille de l'échantillon est essentielle pour augmenter les chances d'identifier correctement les impacts.



# Taille de l'échantillon - Concepts clés



- Compromis entre la taille de l'échantillon, la taille de l'effet (impact) qui peut être mesuré (MDE) et la précision statistique

	<i>Pour une...</i>	<i>Si...</i>	<i>Alors...</i>
	Taille de l'échantillon	MDE ↑	Précision statistique ↑
	MDE	Taille de l'échantillon ↑	Précision statistique ↑
	Précision statistique	Taille de l'échantillon ↑	MDE ↓

- **Les compromis ne sont généralement pas linéaires** - par exemple, si l'impact attendu du programme (MDE) est divisé par 2, la taille de l'échantillon nécessaire pour le mesurer avec précision augmentera de plus de 2 !
- **La taille égale du groupe de traitement et du groupe de contrôle** améliore l'EDM et la précision statistique.
- **Le regroupement a une incidence** sur la taille de l'échantillon (les plans EIC regroupés nécessitent généralement une taille d'échantillon plus importante).





- **Non-réponse : Absence d'obtention des informations souhaitées de la part des répondants**
  - **Attrition**
    - Par exemple, les personnes interrogées qui faisaient partie du programme ont déménagé et changé de numéro de téléphone, de sorte que les équipes d'enquête ne peuvent pas les retrouver.
    - Les enseignants qui ont quitté l'enseignement depuis la base de référence et qui ne font donc plus partie de la population concernée par un programme de formation des enseignants.
  - Refus
  - Mauvaise conception du questionnaire
- **Non-réponse systématique des répondants de l'échantillon**
  - S'il n'est pas aléatoire, il peut conduire à des données biaisées.
  - Perte de la taille de l'échantillon et de la puissance de l'analyse



## Comment garantir la qualité des données ?

- Conception minutieuse du questionnaire
- Méthodologie de collecte des données
  - Protocoles de travail sur le terrain
  - Formation du personnel de terrain
  - Méthode d'administration de l'enquête
    - Entretiens personnels papier-crayon (PAPI)
    - Entretien personnel assisté par ordinateur (CAPI)
- Suivi de la collecte des données
  - Contrôles quotidiens (automatisés) pour identifier les données potentiellement problématiques et/ou les agents recenseurs potentiellement peu performants.



**Survey  
Solutions**

**SurveyCTO**

**KoboToolbox**

Développé par la Banque mondiale

L'utilisateur doit installer/avoir son propre serveur local ou en nuage.

Des paradata extrêmement approfondis

Conception simple de questionnaires complexes

Développé par Doherty à l'aide de l'outil ODK Open source

Frais d'abonnement

Gamme de plug-ins développés pour améliorer les enquêtes

Nécessite de solides compétences en programmation pour les questionnaires complexes

Développé par KoBo Inc. à l'aide de l'outil ODK Open source

Gratuit (avec des limites d'utilisation)

Des plug-ins moins développés

Nécessite de solides compétences en programmation pour les questionnaires complexes

# Tirer parti de la technologie pour obtenir des données de haute qualité pour l'EIC et le MIS

- Tirer parti de la technologie pour collecter des données de haute qualité et mettre en place des systèmes de suivi :
  - Entretien personnel assisté par ordinateur (CAPI)
  - Interview téléphonique assistée par ordinateur (CATI)
    - Enquêtes téléphoniques
  - Systèmes d'information géographique (SIG)
    - Échantillonnage
    - Contrôle de la qualité des données



- Le suivi et l'évaluation ne sont pas la même chose que l'EIC
- Suivi de l'évaluation = Le programme/l'intervention fonctionne-t-il(elle) comme prévu ?
- Comme pour l'EIC, le suivi de l'évaluation dépend fortement de la collecte de données
- Les systèmes de suivi sont essentiels pour l'évaluation de l'impact
- Les systèmes de suivi fournissent des informations sur les ressources disponibles, les résultats et la nécessité d'un soutien et d'une correction.



- Aucune EIC ne vaut la peine de risquer la sécurité des participants aux collectes de données.
- La protection des participants doit être le principe directeur ultime de toutes les collectes de données.
- Obtenir le consentement éclairé avant la collecte des données
- Satisfaire à toutes les exigences éthiques du comité d'éthique compétent et obtenir les approbations et les autorisations de l'IRB avant le début de la collecte des données sur le terrain.
- La sécurité et la protection des données sont essentielles pour toutes les collectes de données



- Atelier de formation sur l'évaluation d'impact contrefactuelle (EIC) - PowerPoint Slides
- Atelier de formation sur la collecte de microdonnées dans les zones difficiles d'accès - Diapositives PowerPoint

## Livres

- [Banque mondiale, Impact Evaluation in Practice - Second Edition \(Livre\)](#)

## Vidéos

- [InterAction, Introduction à l'évaluation d'impact](#)
- [Esther Duflo, Essais contrôlés randomisés et élaboration des politiques dans les pays en développement](#)

## Podcasts

- [IEU Talks Episode 2 : Le pouvoir de l'évaluation d'impact dans la coopération au développement](#)
- [Démontrer l'impact \(parties 1+2\)](#)

**FIN DE LA SESSION 1**



# **Session 2 : Statistiques descriptives pour le suivi et l'évaluation**

**C4ED - EUTF**

**Octobre 2023**



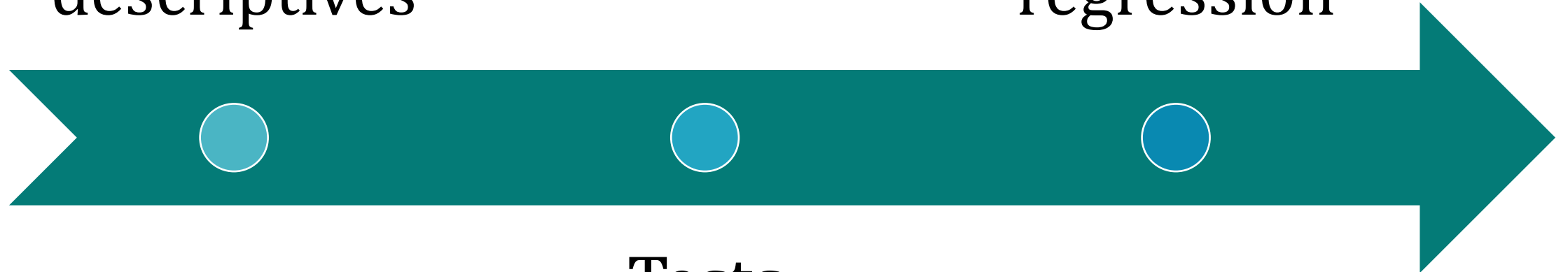
# Vue d'ensemble

- La première année s'est concentrée sur l'intuition des méthodes EIC, et la deuxième année sur les aspects pratiques de la collecte de données pour l'EIC et le suivi.
- L'atelier de cette année se concentre sur l'**utilisation des données**
- Nous passerons en revue les concepts/méthodes de base de l'analyse des données qui permettent d'informer et de mener une EIC
- L'objectif est de se faire une idée du fonctionnement des différentes analyses à un niveau intuitif
  - Aucune connaissance préalable des statistiques n'est requise



Statistiques  
descriptives

Analyse de  
régression

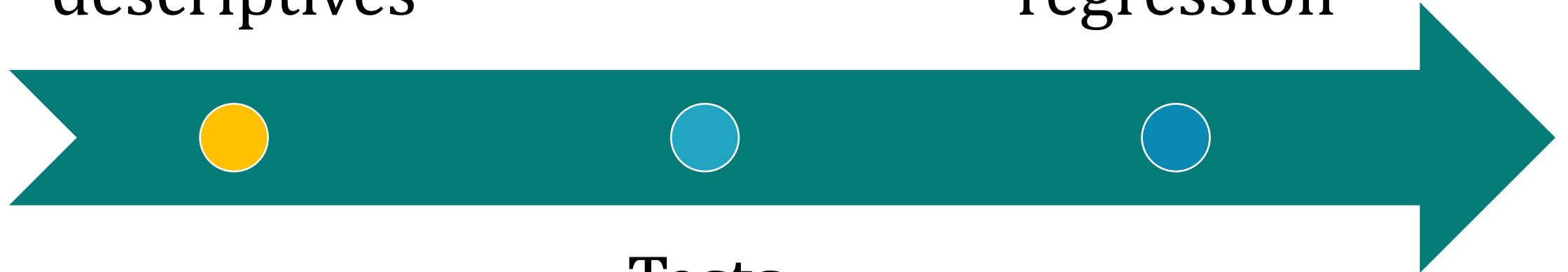


Tests  
statistiques



Statistiques  
descriptives

Analyse de  
régression



Tests  
statistiques



Center for Evaluation  
and Development

# Statistiques descriptives pour le suivi et l'évaluation

---



# Vue d'ensemble

- Statistiques descriptives
- Données catégorielles et continues
- Tabulation (univariée), tableau croisé (bivariée)
- Mesures de tendance centrale → moyenne, médiane
- Mesures de dispersion → min-max, variance/écart-type
- Valeurs aberrantes → comment les identifier et comment les traiter (suppression, imputation, winsorisation).
- Statistiques inférentielles (introduction / intuition)
- ANNEXE : Distribution statistique et asymétrie, Mode, Intervalle interquartile

Nous nous concentrerons sur l'intuition plutôt que sur les aspects techniques.



# Qu'est-ce que les statistiques descriptives ?

---

- Les statistiques descriptives fournissent une vue d'ensemble/un résumé des informations quantitatives complexes contenues dans de grands ensembles de données
- Même des statistiques descriptives de base peuvent être utiles pour :
  - Comprendre les principales caractéristiques des bénéficiaires du programme
  - Confirmer si les objectifs d'un programme sont atteints
  - Repérer les schémas potentiellement problématiques ou inhabituels dans les données



## **Données catégorielles**

- Informations pouvant être enregistrées en termes de catégories exclusives → Sexe ; statut professionnel ; possession d'actifs

## **Données continues**

- Variables pouvant prendre n'importe quelle valeur réelle dans l'éventail des valeurs possibles
  - En pratique, les variables continues sont souvent limitées à 0
    - Âge → ne peut être inférieur à 0, ne peut être infini
    - Revenu ; superficie des terres possédées → ne peuvent être inférieurs à 0
- Différents types de données requièrent différentes statistiques





### Données catégorielles

- Agrégation
  - Tabulation
- } Univarié (une seule variable)
- Tableaux croisés
  - Désagrégation
- } Bivarié (deux variables)

### Données continues

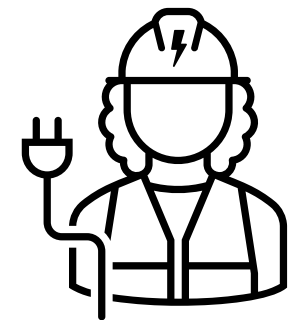
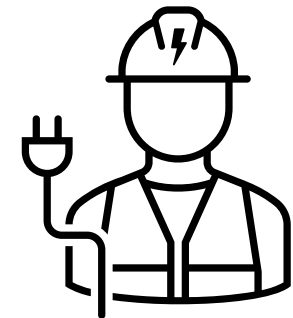
- Mesures de tendance centrale (moyenne, médiane)
- Mesures de dispersion (min-max, écart-type)



# Statistiques descriptives pour le suivi et l'évaluation

## *Exemple – Enoncé*

- Une organisation (votre client) conçoit et met en œuvre un programme d'éducation et formation techniques et professionnels (EFTP) dans des centres d'EFTP.
- **Objectif général** : donner aux jeunes défavorisés les moyens économiques de s'engager dans des stratégies d'emploi et de subsistance
- **Objectifs spécifiques** :
  - Inscrire 1 000 jeunes à la formation professionnelle
  - Assurer une répartition hommes/femmes de 50/50 entre tous les participant
  - Atteindre un taux d'obtention de diplôme de 90%
  - Revenu mensuel moyen de 800 unités pour les diplômés six mois après la formation





# Statistiques descriptives pour le suivi et l'évaluation

## *Exemple – Objectif*

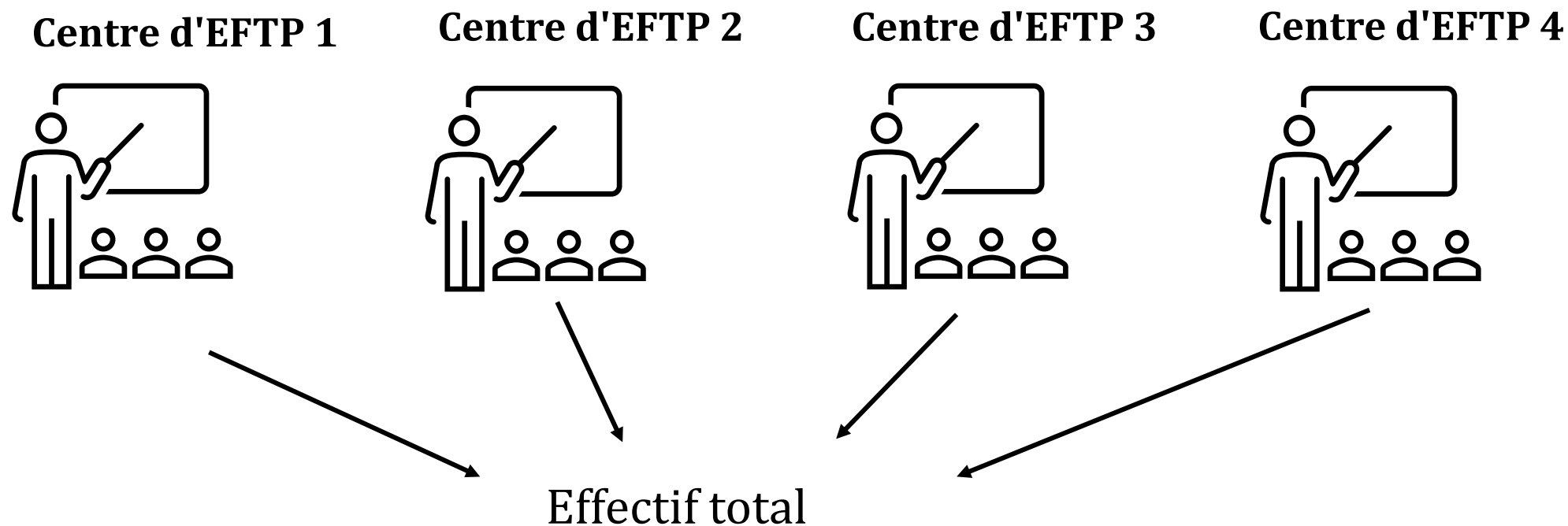
- Vous disposez d'un ensemble de données de suivi contenant les informations suivantes sur tous les participants :
    - Sexe
    - Centre EFTP d'inscription
    - Statut de fin de formation (diplômé ou non)
    - Revenu moyen des diplômés six mois après la fin de la formation
  - Dans les étapes suivantes, nous utiliserons la boîte à outils des statistiques descriptives pour déterminer si le programme a atteint ses objectifs
- Selon les critères d'évaluation de l'OCDE, le programme a-t-il été *efficace* ?



# Statistiques descriptives pour le suivi et l'évaluation

## Exemple – Effectifs

- Objectif : inscrire 1 000 jeunes à la formation professionnelle





# Statistiques descriptives pour le suivi et l'évaluation

## Exemple – Agrégation / Tabulation

Center for Evaluation  
and Development

- Objectif : faire participer 1 000 jeunes à la formation professionnelle

Centre d'EFTP	Nombre de participants
Centre d'EFTP 1	115
Centre d'EFTP 2	122
Centre d'EFTP 3	109
Centre d'EFTP 4	116
<b>Total</b>	<b>462</b>

- Le programme n'a pas atteint son objectif
  - Une enquête plus approfondie (éventuellement qualitative) est justifiée pour en comprendre les raisons (problèmes dans la conception du programme ? ou goulets d'étranglement dans la mise en œuvre ?)



- Objectif : Assurer une répartition hommes/femmes de 50/50 entre tous les participants

Sexe	Centre d'EFTP 1	Centre d'EFTP 2	Centre d'EFTP 3	Centre d'EFTP 4	Total	%
Homme	60	79	42	51	232	50.2%
Femme	55	43	67	65	230	49.8%
<b>Total</b>	<b>115</b>	<b>122</b>	<b>109</b>	<b>116</b>	<b>462</b>	<b>100%</b>

- Le programme a presque atteint une répartition parfaitement égale entre les sexes (50:50)
  - On ne peut cependant pas en dire autant de chaque centre d'EFTP



- Objectif : taux d'obtention de diplôme de 90%

Centre d'EFTP	Diplômé		Non Diplômé		Total
	Fréquence	Pourcentage	Fréquence	Pourcentage	
Centre d'EFTP 1	108	94%	7	6%	115
Centre d'EFTP 2	114	93%	8	7%	122
Centre d'EFTP 3	92	84%	17	16%	109
Centre d'EFTP 4	108	93%	8	7%	116
<b>Total</b>	<b>422</b>	<b>91%</b>	<b>40</b>	<b>9%</b>	<b>462</b>



Selon vous, y a-t-il quelque chose en particulier qui ressort de ce tableau ?



# Statistiques descriptives pour le suivi et l'évaluation

## Exemple – Agrégation / Tabulation

Centre d'EFTP	Diplômé		Abandon		Total
	Fréquence	Pourcentage	Fréquence	Pourcentage	
Centre d'EFTP 1	108	94%	7	6%	115
Centre d'EFTP 2	114	93%	8	7%	122
Centre d'EFTP 3	92	84%	17	16%	109
Centre d'EFTP 4	108	93%	8	7%	116
<b>Total</b>	<b>422</b>	<b>91%</b>	<b>40</b>	<b>9%</b>	<b>462</b>

- Le taux global d'obtention de diplômes est supérieur à 90 % → Le programme atteint son objectif !
  - **Toutefois**, le taux d'obtention de diplômes dans le centre d'EFTP 3 semble nettement inférieur à celui des autres centres
- Creusons un peu plus en utilisant des **tableaux croisés**





- Les tableaux croisés et la désagrégation peuvent être extrêmement utiles lors de l'exploration des données.
  - Peut aider à vérifier/détecter certaines tendances sous-jacentes aux données/statistiques agrégées
- Point de départ d'une analyse plus approfondie



# Statistiques descriptives pour le suivi et l'évaluation

## Exemple - Désagrégation / Tableau croisé

- Examinons le taux moyen d'obtention de diplôme par sexe et par centre d'EFTP avec un **tableau croisé**

Genre	Taux d'obtention du diplôme	
	Homme	Femme
Centre d'EFTP 1	95%	93%
Centre d'EFTP 2	94%	93%
Centre d'EFTP 3	100%	75%
Centre d'EFTP 4	96%	91%
<b>Moyenne</b>	<b>96%</b>	<b>87%</b>

- Les hommes ont un taux d'obtention de diplôme plus élevé que les femmes dans l'ensemble
- La différence est particulièrement marquée dans le centre d'EFTP 3



- Objectif : revenu mensuel moyen de 800 unités pour les diplômés six mois après la fin de la formation
- Jusqu'à présent, nous avons examiné des données qui pouvaient être facilement résumées à l'aide de fréquences et de proportions
- Les données telles que le revenu ne peuvent pas être facilement résumées de cette manière
- Nous avons besoin d'autres outils, à savoir les **mesures de tendance centrale**. Rappelons ce qu'elles sont !



Center for Evaluation  
and Development

# Mesures de tendance centrale

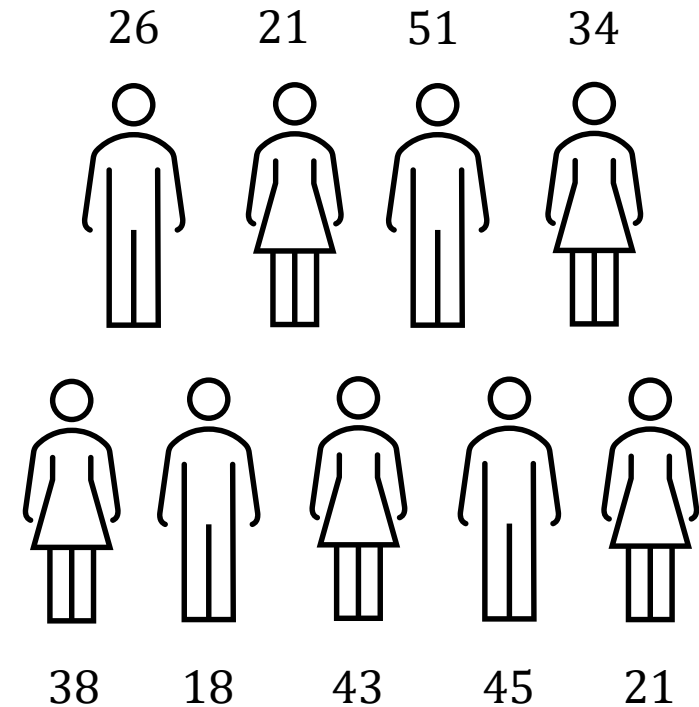
---



# Mesures de tendance centrale

## Moyenne

- **Moyenne** : La somme de toutes les valeurs divisée par le nombre total de valeurs dans l'ensemble de données
- Exemple :
  - Nous avons des mesures d'âge pour 9 personnes
  - Valeur moyenne de l'âge :  
 $(26+21+51+34+38+18+43+45+21)/9 = 33$

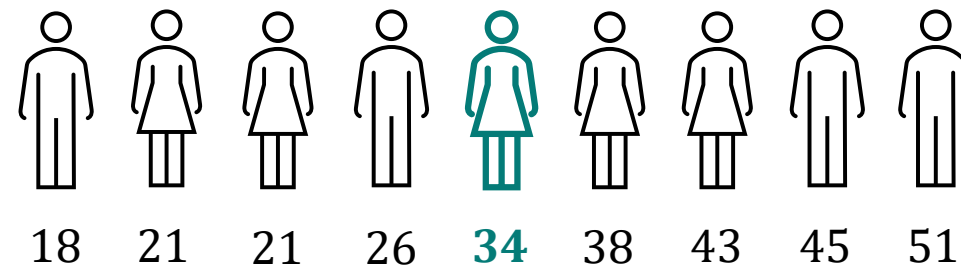




# Mesures de tendance centrale

## Médiane

- **Médiane** = Pour un ensemble donné de valeurs, la médiane est la valeur qui divise l'ensemble exactement au milieu, c'est-à-dire qu'exactement la moitié des valeurs sont inférieures/supérieures à la médiane.
- Elle est calculée en classant les valeurs par ordre croissant (de la plus faible à la plus élevée) et en trouvant la valeur située exactement au milieu.
- Dans notre exemple → médiane = 34



*Remarque:* La médiane appartient à la famille plus large des **percentiles**. De même que la médiane partage l'ensemble de données en 2, les percentiles partagent l'ensemble des données en 4, en 10, en 100, etc. ensembles de même taille (voir détails en Annexe).



# Statistiques descriptives pour le suivi et l'évaluation

Center for Evaluation  
and Development

## Exemple – Moyenne

- Objectif : revenu mensuel moyen de 800 euros pour les diplômés six mois après la fin de la formation

[NB : il n'est pas pratique d'écrire le revenu mesuré pour l'ensemble des 462 participants, nous nous concentrons donc sur un échantillon de 10 participants].

Participant	Revenu mensuel
Diplômé 1	200
Diplômé 2	350
Diplômé 3	400
Diplômé 4	600
Diplômé 5	100
Diplômé 6	700
Diplômé 7	300
Diplômé 8	0
Diplômé 9	6 000
Diplômé 10	200
<b>Total</b>	<b>8 850</b>



# Statistiques descriptives pour le suivi et l'évaluation

Center for Evaluation  
and Development

## Exemple – Moyenne

- Objectif : revenu mensuel moyen de 800 euros pour les diplômés six mois après la fin de la formation
  - Moyenne =  $8\ 850 / 10 = 885$
  - Le programme a atteint son objectif
- Seriez-vous prêt à annoncer que le revenu moyen des diplômés est de 885 euros par mois ?



Participant	Revenu mensuel
Diplômé 1	200
Diplômé 2	350
Diplômé 3	400
Diplômé 4	600
Diplômé 5	100
Diplômé 6	700
Diplômé 7	300
Diplômé 8	0
Diplômé 9	6 000
Diplômé 10	200
<b>Total</b>	<b>8 850</b>





# Statistiques descriptives pour le suivi et l'évaluation

Center for Evaluation  
and Development

## *La moyenne et les valeurs aberrantes*

**Traduction:**  
"Le problème  
des moyennes.  
Profondeur  
moyenne = 1m."

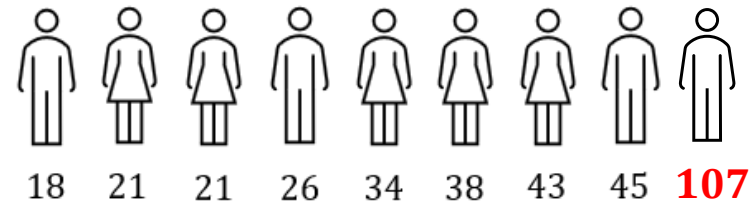
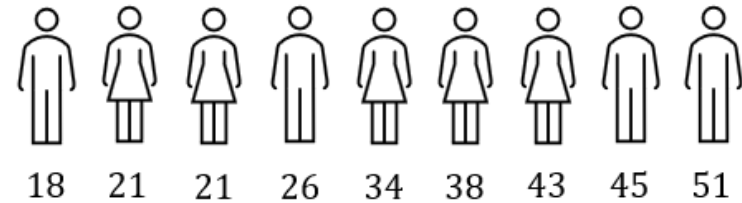




# Statistiques descriptives pour le suivi et l'évaluation

## La moyenne et les valeurs aberrantes

- Le problème de la moyenne est qu'elle peut être affectée par des valeurs très grandes ou très petites par rapport aux autres
- Exemple précédent, âge moyen = 33 ans  
→ 4 personnes plus jeunes, 5 plus âgées, la moyenne semble donc donner une bonne idée de la tendance centrale
- Imaginez que l'une des personnes est très âgée  
→ La moyenne est maintenant de 40 ans  
→ 6 personnes plus jeunes, 3 plus âgées, de sorte que la moyenne n'est peut-être pas la meilleure mesure de la tendance centrale dans ce cas (on dit qu'elle est "biaisée vers le haut")
- Ces valeurs extrêmes sont appelées "**valeurs aberrantes**" (voir plus loin)





- Le problème de la moyenne est qu'elle peut être affectée par des valeurs très grandes ou très petites par rapport aux autres.  
→ Ces valeurs extrêmes sont appelées "**valeurs aberrantes**" (voir plus loin).
- En statistiques, nous dirions que les données sont **asymétriques**
  - **Le coefficient d'asymétrie** est une mesure de la symétrie d'une distribution statistique, c'est-à-dire qu'elle indique si les observations sont uniformément réparties autour de la moyenne
  - Les distributions symétriques ont un coefficient d'asymétrie de 0 ; les distributions asymétriques peuvent avoir une asymétrie négative ou positive.
  - Plus l'asymétrie est grande, plus la moyenne est une mauvaise mesure de la tendance centrale !

*[Pour plus de détails sur l'asymétrie, voir l'annexe].*



Avez-vous une idée de variables dont la distribution pourrait être asymétrique ?





# Statistiques descriptives pour le suivi et l'évaluation

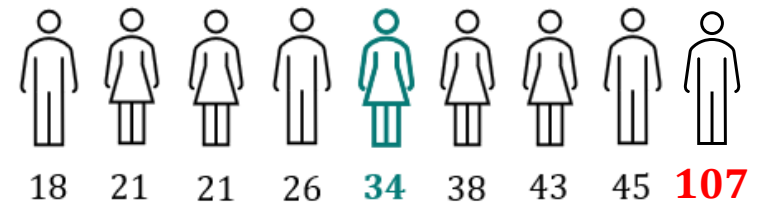
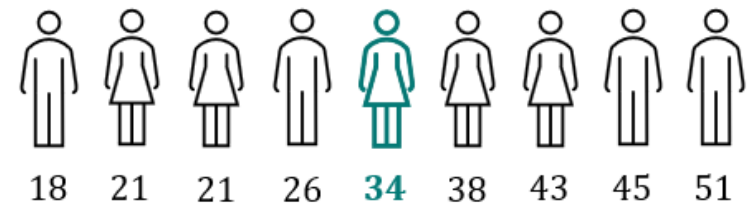
Center for Evaluation  
and Development

## Médiane – Robuste aux valeurs aberrantes

- Contrairement à la moyenne, la médiane est moins sensible aux valeurs extrêmes
- On dit que la médiane est **robuste** aux valeurs aberrantes

[RAPPEL : la médiane est calculée en classant les valeurs par ordre croissant (de la plus faible à la plus élevée) et en trouvant la valeur qui se situe exactement au milieu].

- Dans l'exemple précédent, l'âge médian était de 34 ans
- Si la personne la plus âgée a maintenant 107 ans au lieu de 51, l'âge médian est toujours de 34 !





# Statistiques descriptives pour le suivi et l'évaluation

Center for Evaluation  
and Development

## Exemple – Médiane, robuste aux valeurs aberrantes

- Objectif : revenu mensuel moyen de 800 euros pour les diplômés six mois après la fin de la formation
  - Moyenne = 885 → "biaisée" en raison de valeurs aberrantes
  - On peut également présenter le revenu médian
  - Classer les valeurs par ordre croissant et prendre la valeur du milieu.
    - Note : avec un nombre *pair* d'observations, la médiane est le point médian des deux valeurs moyennes.
  - Médiane =  $(300+350)/2 = 325$
- Nous savons que la médiane est une mesure non biaisée de la tendance centrale en présence de valeurs aberrantes, mais suffit-elle à décrire les données ?

Participant	Revenu mensuel
Diplômé 8	0
Diplômé 5	100
Diplômé 1	200
Diplômé 10	200
Diplômé 7	300
Diplômé 2	350
Diplômé 3	400
Diplômé 4	600
Diplômé 6	700
Diplômé 9	6,000





- Dans cet exemple, nous voyons les limites de l'examen des seules mesures qui décrivent la moyenne/tendance centrale
- Outre les mesures de la tendance centrale, vous pouvez examiner les mesures de la **dispersion**
  - En d'autres termes, il ne s'agit pas seulement de la valeur moyenne ou de la tendance centrale, mais aussi de la façon dont les données sont dispersées autour de cette valeur.



Center for Evaluation  
and Development

# Mesures de dispersion

---





- Une mesure simple de la dispersion consiste à examiner l'éventail des valeurs dans les données, c'est-à-dire les valeurs minimales et maximales.
- Nous constatons que nos mesures de revenus mensuels vont de 0 à 6 000 euros.

Participant	Revenu mensuel
Diplômé 8	0
Diplômé 5	100
Diplômé 1	200
Diplômé 10	200
Diplômé 7	300
Diplômé 2	350
Diplômé 3	400
Diplômé 4	600
Diplômé 6	700
Diplômé 9	6 000



# Statistiques descriptives pour le suivi et l'évaluation

## Mesures de dispersion – Variance et écart-type

- On utilise des mesures plus sophistiquées de la dispersion, telles que la variance et l'écart-type.
  - **Variance** = mesure la variabilité globale des données
  - **Écart-type (ET)** = mesure la distance entre chaque valeur et la moyenne.
    - Écart-type = racine carrée de la variance
  - Plus la variance/écart-type est élevée, plus les données sont dispersées autour de la moyenne
- La variance et l'écart-type sont particulièrement importants pour les tests statistiques (prochaine session)



# Statistiques descriptives pour le suivi et l'évaluation

## Exemple – Mesures de dispersion

- Objectif : revenu mensuel moyen de 800 euros pour les diplômés six mois après la fin de la formation

→ Moyenne = 885

→ Écart-type = 1 810

Participant	Revenu mensuel
Diplômé 1	200
Diplômé 2	350
Diplômé 3	400
Diplômé 4	600
Diplômé 5	100
Diplômé 6	700
Diplômé 7	300
Diplômé 8	0
Diplômé 9	6 000
Diplômé 10	200



# Statistiques descriptives pour le suivi et l'évaluation

Center for Evaluation  
and Development

## Exemple – Mesures de dispersion

Moyenne = 885 ; écart-type = 1 810

- En d'autres termes, le revenu d'un diplômé est en moyenne inférieur ou supérieur de 1 810 unités au revenu moyen (885).
- La moyenne montre que le programme a atteint son objectif
- L'écart-type nuance la moyenne et indique que les revenus tendent à être assez dispersés – certains étant très inférieurs et d'autres très supérieurs à la moyenne

Participant	Revenu mensuel
Diplômé 1	200
Diplômé 2	350
Diplômé 3	400
Diplômé 4	600
Diplômé 5	100
Diplômé 6	700
Diplômé 7	300
Diplômé 8	0
Diplômé 9	6 000
Diplômé 10	200



# Statistiques descriptives pour le suivi et l'évaluation

Center for Evaluation  
and Development

## Exemple – L'écart-type en complément de la moyenne

Comme vous pouvez le constater, les diplômés de notre programme gagnent en moyenne 885 par mois, ce qui dépasse largement notre objectif de 800 par mois !



Montrez-nous vos écarts types !

Participant	Revenu mensuel
Diplômé 1	200
Diplômé 2	350
Diplômé 3	400
Diplômé 4	600
Diplômé 5	100
Diplômé 6	700
Diplômé 7	300
Diplômé 8	0
Diplômé 9	6 000
Diplômé 10	200



# Statistiques descriptives pour le suivi et l'évaluation

## Exemple – L'écart-type en complément de la moyenne

- Idéalement, lors de la présentation des données (séminaire, rapport), vous auriez la possibilité de contextualiser la moyenne avec la mesure de la dispersion.
  - Inclure l'écart-type dans vos résultats
  - Discuter des résultats pour fournir un contexte et les expliquer
- Cependant, les gens s'attendent souvent à ce que la réponse à la question de savoir si un objectif a été atteint soit oui ou non...

Revenu après 6 mois		
Nombre de diplômés	Moyenne	Écart-type
10	885	1 810



Que pouvez-vous faire pour traiter les données asymétriques (c'est-à-dire les données comportant des valeurs aberrantes), de manière à ce que les statistiques descriptives soient plus robustes ?



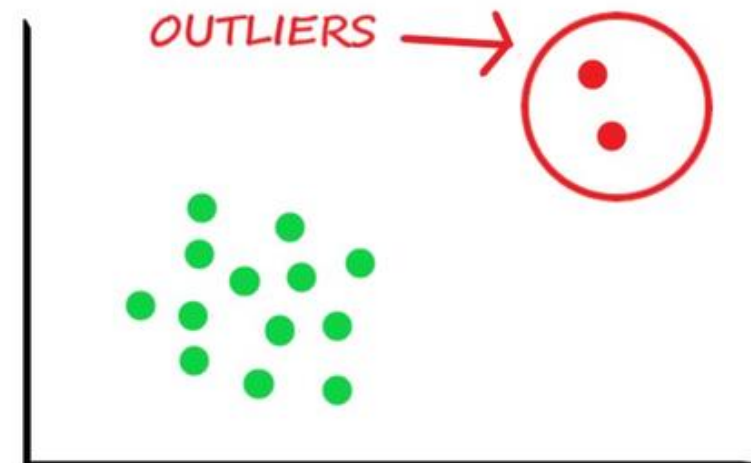


# Statistiques descriptives pour le suivi et l'évaluation

## Traitement des données – Valeurs aberrantes

Que pouvez-vous faire pour traiter des données asymétriques (c'est-à-dire des données comportant des valeurs aberrantes) ?

- On traite généralement les **valeurs aberrantes** lors de la préparation des données avant l'analyse.
  - Étape 1 : identifier les valeurs aberrantes
  - Étape 2 : "traiter" les valeurs aberrantes
- Cela peut permettre d'obtenir des statistiques descriptives plus informatives/plus robustes.
- C'est également important pour une analyse plus complexe à un stade ultérieur !







Center for Evaluation  
and Development

# Valeurs aberrantes

---



# Statistiques descriptives pour le suivi et l'évaluation

Center for Evaluation  
and Development

## Exemple – Valeurs aberrantes

- Une valeur aberrante diffère significativement des autres observations
- Dans notre exemple de données, le revenu du diplômé 9 pourrait être décrit comme une **valeur aberrante**
- Pas de définition fixe pour identifier les valeurs aberrantes
- Règle empirique conventionnelle en sciences sociales → valeur aberrante = valeur située à au moins 2,5 à 3 écarts types de la moyenne (au-dessus ou au-dessous)

Participant	Revenu mensuel
Diplômé 1	200
Diplômé 2	350
Diplômé 3	400
Diplômé 4	600
Diplômé 5	100
Diplômé 6	700
Diplômé 7	300
Diplômé 8	0
Diplômé 9	6 000
Diplômé 10	200



# Statistiques descriptives pour le suivi et l'évaluation

Center for Evaluation  
and Development

## Exemple – Valeurs aberrantes

Moyenne = 885 ; écart-type (ET) = 1 810

- Règle : une valeur aberrante est une valeur d'au moins 2,5 ET (4 525) inférieure/supérieure à la moyenne
- Considérer les revenus inférieurs à 2,5 ET (- 3 640) comme des valeurs aberrantes.
  - Si nous supposons que les revenus ne peuvent être inférieurs à 0 → pas de valeurs aberrantes au bas de l'échelle
- Considérer les revenus supérieurs à 2,5 ET (5 410) comme une valeur aberrante.
  - Le revenu du diplômé 9 est une valeur aberrante

Participant	Revenu mensuel
Diplômé 1	200
Diplômé 2	350
Diplômé 3	400
Diplômé 4	600
Diplômé 5	100
Diplômé 6	700
Diplômé 7	300
Diplômé 8	0
Diplômé 9	6 000
Diplômé 10	200



- Que peut-on faire des valeurs aberrantes ?

→ Dépend de leur origine, c'est-à-dire si elles proviennent ou non d'*erreurs de mesure*.



- Que pouvons-nous faire avec les valeurs aberrantes ?
- Les valeurs aberrantes en tant qu'erreurs de mesure (Rappel de la 2<sup>ème</sup> année du séminaire sur la collecte de données)
  - Erreur dans la réponse (intentionnelle ou non)
  - Erreur d'enregistrement - par exemple, saisie accidentelle de 6 000 au lieu de 600
- Si on est certain qu'il s'agit d'une erreur de mesure – par ex., si une personne est déclarée âgée de 200 ans – on peut :
  - Corriger les données (si possible)
  - Supprimer l'observation



- Que peut-on faire des valeurs aberrantes ?
- Si rien ne prouve que les valeurs aberrantes sont dues à des erreurs de mesure, 3 options sont possibles :
  - Option 1 : Supprimer
  - Option 2 : Imputer
  - Option 3 : « Winsorisation »



- Que pouvons-nous faire avec les valeurs aberrantes ?
  - **Option 1 : Supprimer**
    - Généralement, *suppression complète du cas* →, c'est-à-dire que l'observation est entièrement supprimée de l'analyse.
    - Dans notre exemple, le diplômé 9 a été identifié comme une valeur aberrante en termes de revenus.
    - *La suppression complète du cas* signifie que nous éliminerions complètement le diplômé 9 de l'analyse, même s'il n'est pas une valeur aberrante par rapport à d'autres variables



- Que pouvons-nous faire avec les valeurs aberrantes ?
  - **Option 2 : Imputer**
    - *L'imputation* consiste à remplacer la valeur aberrante par une valeur jugée « représentative »
    - On peut utiliser par exemple la moyenne ou la médiane, mais dans la pratique, les méthodes d'imputation sont souvent plus sophistiquées

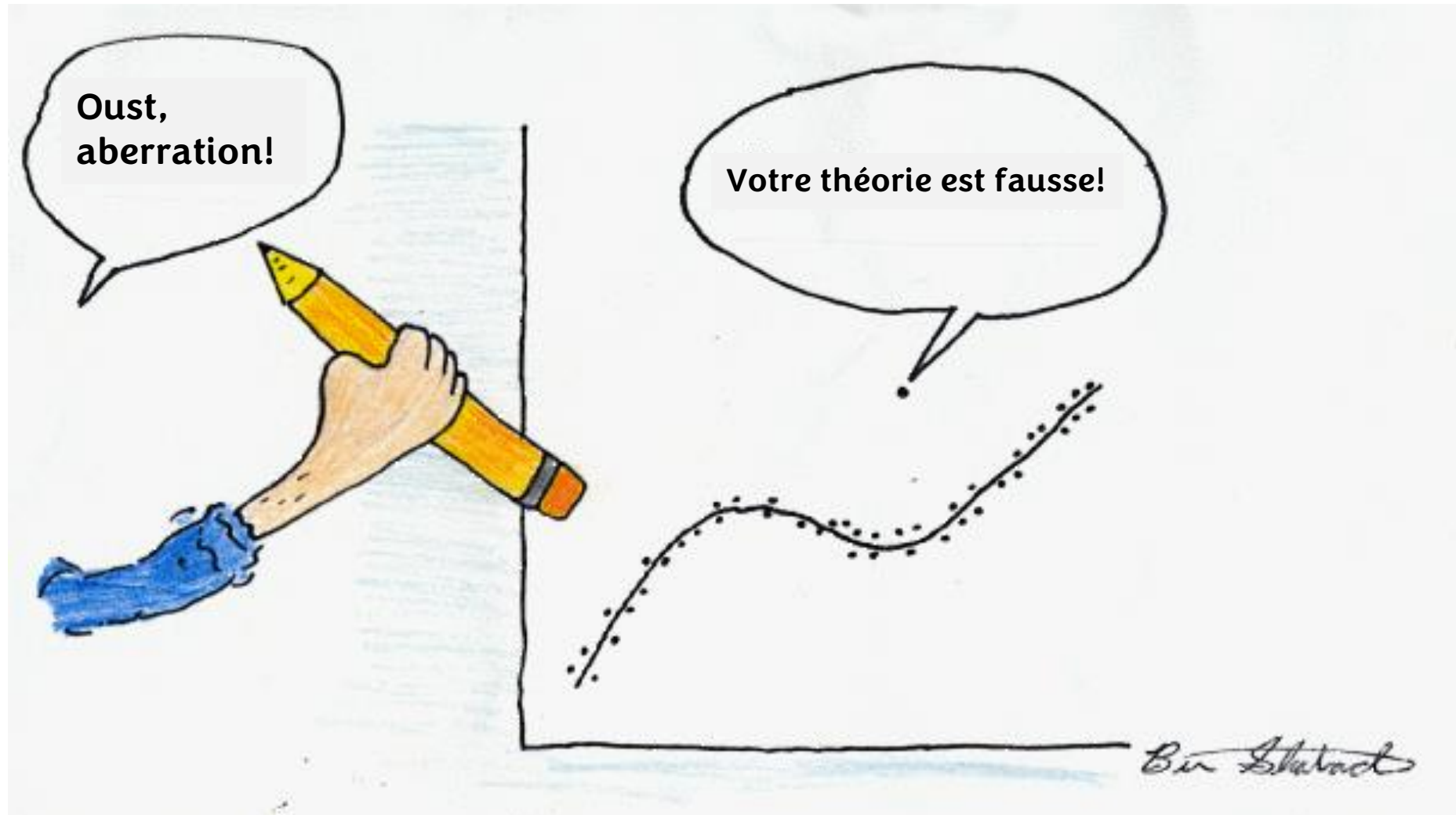




- Que pouvons-nous faire avec les valeurs aberrantes ?
  - **Option 3 : « Winsorisation »**
    - Définir une valeur minimale/maximale acceptable - par exemple, le percentile 1% ou 5% (P1 ou P5) comme minimum, le percentile à 99% ou 95% (P99 ou P95) comme maximum.
    - Remplacer les valeurs inférieures (supérieures) à la valeur minimale (maximale) choisie par la valeur minimale (maximale)
    - Cette approche vise à trouver un équilibre entre le maintien des valeurs aberrantes et leur suppression
    - Protège vos données contre les valeurs aberrantes les plus extrêmes qui causent des problèmes



- Que pouvons-nous faire avec les valeurs aberrantes ?
- Si rien ne prouve que les valeurs aberrantes sont dues à des erreurs de mesure, 3 options sont possibles :
  - Option 1 : Supprimer
  - Option 2 : Imputer
  - Option 3 : « Winsorisation »
- Ces options devraient faciliter l'utilisation et l'interprétation de vos données
- **Mais** certaines valeurs aberrantes font naturellement partie des données et fournissent des informations importantes sur la variance/imprévisibilité des données





### Remarque 1

- Certaines valeurs aberrantes apportent des informations réelles et doivent être conservées dans les données.
  - Il peut être difficile de distinguer une « vraie » valeur extrême d'une valeur aberrante en raison de l'erreur de mesure → toujours un compromis
  - Par exemple, avec la winsorisation, on peut fixer un seuil/plafond plus élevé (percentile à 99 % ou 99,5 %) pour conserver certaines valeurs aberrantes
  - Il existe d'autres approches qui visent à trouver un équilibre entre la stabilisation des données et la conservation des valeurs extrêmes informatives



### Remarque 2

- Nous avons dit que la médiane est résistante aux valeurs aberrantes, alors pourquoi ne pas toujours indiquer la médiane plutôt que la moyenne ?
  - Mesure de dispersion lors de la déclaration de la médiane → intervalle interquartile
  - Les statistiques ont été largement développées autour des moyennes (tests statistiques, analyse de régression) → Les statistiques descriptives ont tendance à se concentrer sur les moyennes parce que l'analyse ultérieure se concentre généralement sur les moyennes.



Center for Evaluation  
and Development

# Statistiques inférentielles – Intro

---



- Si vous disposez d'un système de suivi solide, vous pouvez disposer de données sur **tous les** participants à une intervention.
  - Toutefois, dans la pratique, il arrive souvent que nous ne disposions pas de toutes les informations sur tous les participants
  - Du coup, comment pouvons-nous connaître les caractéristiques de *tous* les participants au programme ?
- Nous pouvons prélever un *échantillon représentatif* de la population.
- Veuillez vous référer aux diapositives de l'année 2 pour des informations sur l'échantillonnage



- Si nous utilisons un échantillon, nous devons estimer dans quelle mesure nous sommes **sûrs** de pouvoir appliquer nos conclusions à l'ensemble de la population
  - C'est ce que l'on appelle notre capacité à faire des déductions – ou « inférer » – sur la base de notre échantillon.
- Nous pouvons également utiliser des statistiques inférentielles pour estimer le niveau de confiance lorsqu'on mesure un impact causal





# Statistiques descriptives pour le suivi et l'évaluation

## Statistiques inférentielles – Intro

- Imaginons que 2 000 personnes se soient portées candidates à notre programme et que 1 000 d'entre elles aient été sélectionnées de manière aléatoire.
  - C'est à dire que nous avons un ECR rigoureux → Toute différence entre les deux groupes doit être attribuable au programme !
- Les données montrent que ceux qui ont bénéficié de la formation professionnelle gagnent 150 unités de plus par mois que ceux qui n'en ont pas bénéficié

	Nombre de personnes	Revenu moyen
Formation professionnelle	1 000	1 200
N'a pas reçu de formation professionnelle	1 000	1 050



- Le programme a-t-il eu un *impact* sur les revenus ? Dans quelle mesure sommes-nous convaincus que cette différence reflète l'*impact réel* du programme ?

**FIN DE LA SESSION 2**



Center for Evaluation  
and Development

## **Session 2 : Statistiques descriptives pour le suivi et l'évaluation**

---

Annexe



## **Session2 - Annexe**

---

### Distribution statistique



## Session 2 – Annexe

### *Distribution statistique – Définition*

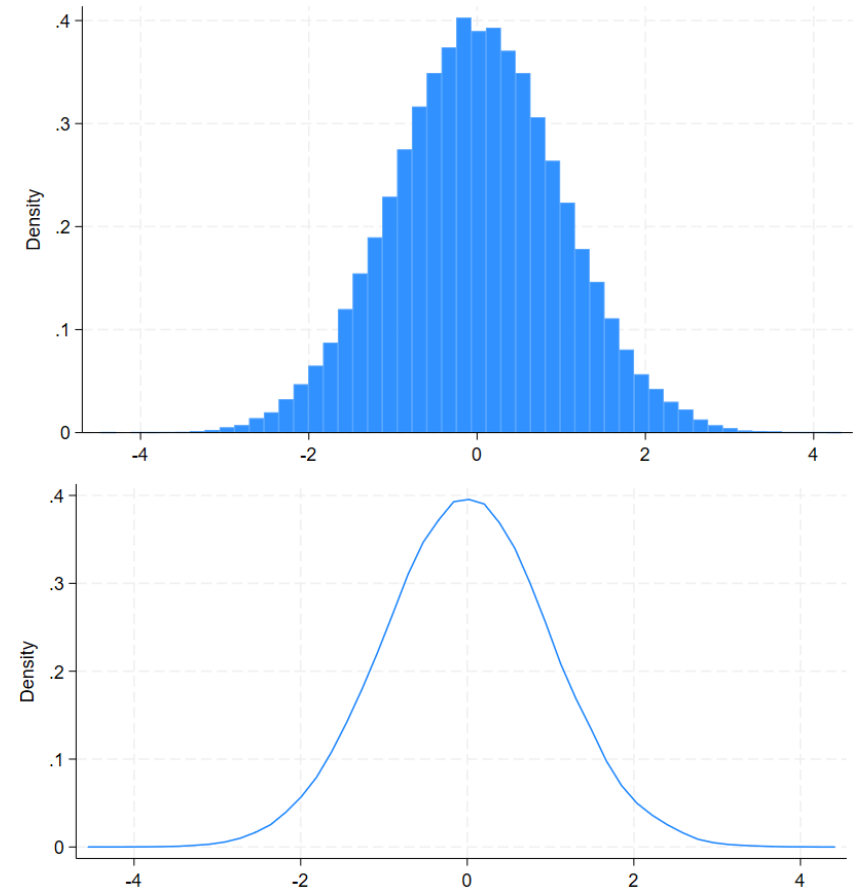
- Une variable continue peut prendre plusieurs valeurs possibles
  - Par exemple, l'âge mesuré en années peut prendre n'importe quelle valeur entre 0 et 123 (la personne la plus âgée est décédée à l'âge de 122,5 ans)
- La **distribution statistique** (ou distribution de probabilité) d'une variable montre essentiellement quelles sont les valeurs courantes (que l'on peut s'attendre à observer souvent) et les valeurs plus rares
  - Par exemple, si nous mesurons l'âge de 100 personnes (sélectionnées au hasard), nous nous attendons à ce qu'environ  $\frac{1}{4}$  d'entre elles aient moins de 14 ans, qu'environ 50 % aient entre 25 et 65 ans et qu'environ 10 % aient plus de 65 ans



## Session 2 – Annexe

### *Distribution statistique – Visualisation*

- La **distribution statistique** (ou distribution de probabilité) d'une variable indique quelles sont les valeurs communes ou plus rares
- Souvent représentée par la **fonction de densité de probabilité** :
  - Histogramme (diagramme à barres verticales) ou courbe
  - Axe des X = plage de valeurs possibles pour la variable
  - Axe des Y = la *densité de probabilité*
    - Attention : la densité n'est **PAS** une probabilité
    - Intuition : pour une valeur donnée  $x$ , plus la densité est élevée, plus la probabilité que la variable – lorsqu'elle est mesurée – prenne une valeur proche de  $x$  est *élevée*.



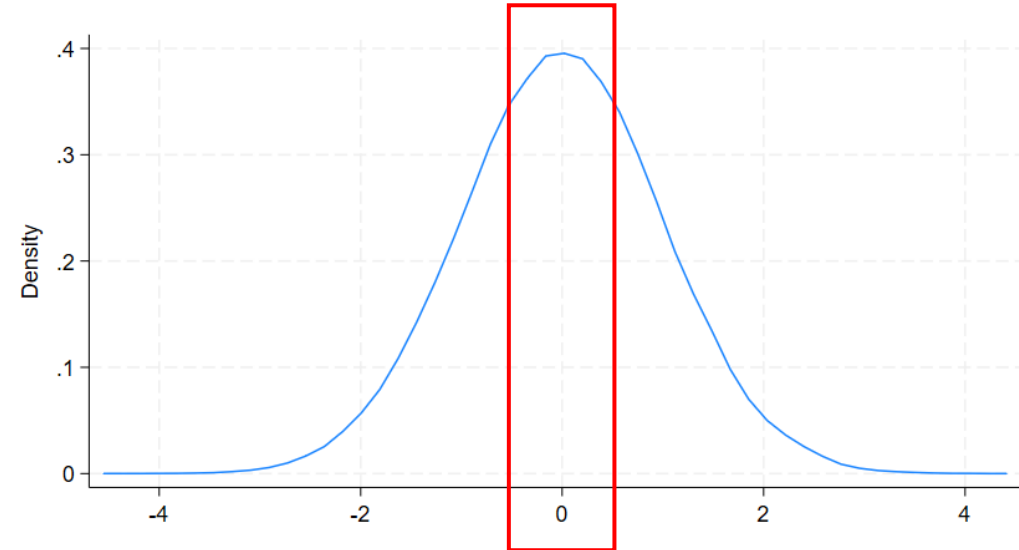
*Exemple : Fonction de densité d'une distribution normale*



## Session 2 – Annexe

### *Distribution statistique et tendance centrale*

- Une variable continue peut prendre plusieurs valeurs possibles
- Les mesures de **tendance centrale** nous aident à nous faire une idée de la distribution sans avoir à parcourir toutes les différentes valeurs mesurées pour la variable
- Plus la densité est élevée, plus la variable est susceptible de prendre cette valeur
- On se concentre sur la **tendance centrale** car, dans la plupart des cas, les valeurs les plus courantes tendent à se situer autour du “centre” de la distribution.





## **Session2 – Annexe**

---

### Valeurs aberrantes et données asymétriques



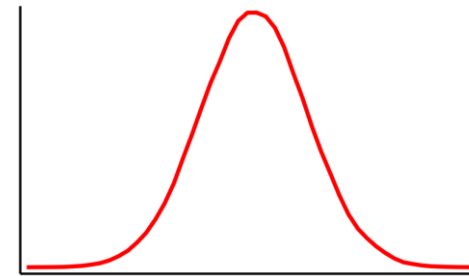


## Session 2 - Annexe

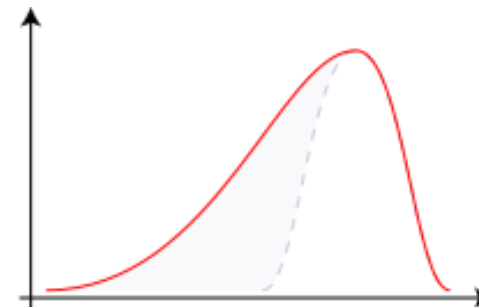
Center for Evaluation  
and Development

# Asymétrie (« skew » ou « skewness » en anglais)

- Le problème de la moyenne est qu'elle peut être affectée par des valeurs très grandes ou très petites par rapport aux autres.
- Ces valeurs extrêmes sont appelées "**valeurs aberrantes**"
- En statistiques, nous dirions que les données sont **asymétriques**
  - **L'asymétrie** (« skew » ou « skewness » en anglais) est une mesure de la symétrie d'une distribution.
  - Plus l'asymétrie est grande, plus la moyenne est une mauvaise mesure de la tendance centrale.

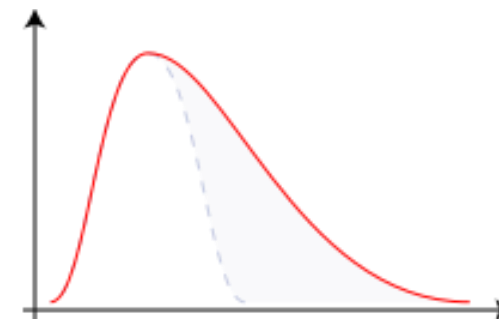


Distribution normale (idéal théorique), parfaitement symétrique → asymétrie = 0



Negative Skew

Exemple d'asymétrie négative → Les valeurs faibles sont plus fréquentes que prévu (asymétrie "à gauche").



Positive Skew

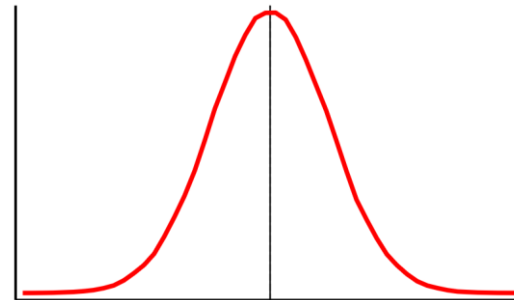
Exemple d'asymétrie positive → Les valeurs élevées sont plus fréquentes que prévu (asymétrie "à droite").



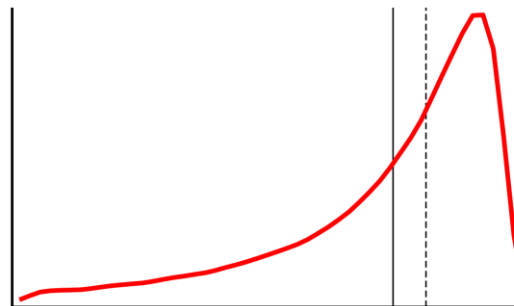
## Session 2 – Annexe

### Center for Evaluation and Development *Asymétrie, Moyenne et Médiane*

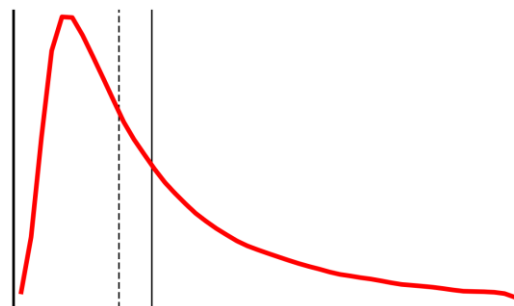
- **L'asymétrie** est une mesure de la symétrie d'une distribution.
  - Plus l'asymétrie est grande, plus la moyenne est une mauvaise mesure de la tendance centrale.
- Plus les données sont asymétriques, plus la moyenne et la médiane sont éloignées l'une de l'autre.



Pas d'asymétrie  
→ Moyenne = Médiane



Asymétrie négative/“à gauche”  
→ Moyenne < Médiane



Asymétrie positive/“à droite”  
→ Moyenne > Médiane

——— moyenne      - - - - - médiane



## **Session 2 – Annexe**

---

### Médiane et percentiles



## Session 2 – Annexe

### *Médiane et percentiles – Remarque*

- La médiane appartient à la famille plus large des ***percentiles***
- Percentiles = valeurs qui répartissent les données dans des proportions données
  - Par exemple, la médiane divise les données en deux, c'est-à-dire que 50 % des valeurs sont supérieures à la médiane et 50 % sont inférieures.
  - La médiane est le 50<sup>ème</sup> percentile et est parfois notée P50.
- De même, nous pouvons définir, par exemple, le 10<sup>ème</sup> percentile (P10) = valeur telle que 10 % des valeurs lui sont inférieures et 90 % supérieures.
- Les pourcentages qui divisent les données en 4 ensembles de valeurs de même taille sont appelés ***quartiles***
  - Il existe 3 quartiles (P25, P50 et P75) que l'on peut appeler les 1<sup>er</sup>, 2<sup>ème</sup> et 3<sup>ème</sup> quartiles; la médiane est le 2<sup>ème</sup> quartile
- Les pourcentages qui divisent les données en 10 ensembles de même taille sont appelés ***déciles***.
  - Il y a 9 déciles (P10, P20, P30, P40, ..., P80, P90) ; la médiane est le 5<sup>ème</sup> décile.



## Session 2 – Annexe

### *Médiane et percentiles – Remarque*

Percentile	Nom	Équivalent
P1	1 <sup>er</sup> percentile	Percentile inférieur
P2	2 <sup>ème</sup> percentile	-
...	...	...
P10	10 <sup>ème</sup> percentile	1 <sup>er</sup> décile / Décile inférieur
...	...	...
P25	25 <sup>ème</sup> percentile	1 <sup>er</sup> quartile / quartile inférieur
...	...	...
P50	50 <sup>ème</sup> percentile	2 <sup>ème</sup> quartile / 5 <sup>ème</sup> décile / Médiane
...	...	...
P75	75 <sup>ème</sup> percentile	3 <sup>ème</sup> quartile / Quartile supérieur
...	...	...
P90	90 <sup>ème</sup> percentile	9 <sup>ème</sup> décile / Décile supérieur
...	...	...
P98	98 <sup>ème</sup> percentile	-
P99	99 <sup>ème</sup> percentile	Percentile supérieur



## Session2 - Annexe

---

Mode et intervalle interquartile



Les diapositives suivantes présentent :

- Une autre mesure de la dispersion → l'***intervalle interquartile***.
  - Utile pour mesurer la dispersion avec la médiane, car l'écart-type n'est pertinent que pour la moyenne
- Une autre mesure de la tendance centrale → le ***mode***
  - Utile pour les variables catégorielles



## Session 2 – Annexe

Center for Evaluation  
and Development

# Mesures de dispersion – Intervalle interquartile

- **Intervalle interquartile** = intervalle des valeurs comprises entre le 1<sup>er</sup> et le 3<sup>ème</sup> quartile, c'est-à-dire entre P25 et P75.  
→ En d'autres termes, la différence de valeurs après avoir retiré les 25 % de valeurs les plus basses et les 25 % de valeurs les plus hautes
- P25 = 25<sup>ème</sup> percentile = 200
- P75 = 75<sup>ème</sup> percentile = 550  
→ Intervalle interquartile =  $550 - 200 = 350$
- Les valeurs intermédiaires (les 50 % des valeurs comprises entre P25 et P75) se situent dans une fourchette de seulement 350 unités

Participant	Revenu mensuel
Diplômé 8	0
Diplômé 5	100
Diplômé 1	200
Diplômé 10	200
Diplômé 7	300
Diplômé 2	350
Diplômé 3	400
Diplômé 4	600
Diplômé 6	700
Diplômé 9	6,000



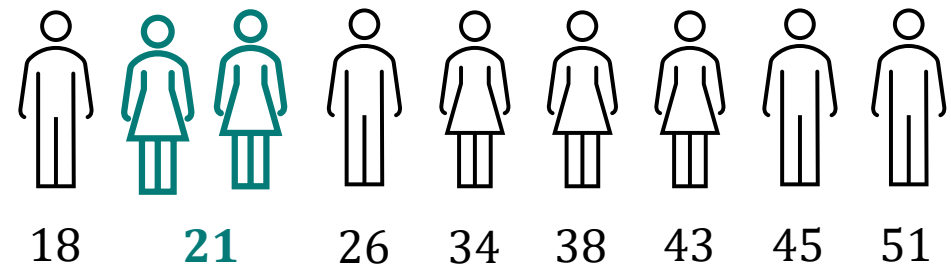


## Session 2 – Annexe

### *Mesures de tendance centrale – Mode*

- **Mode** : La valeur qui apparaît le plus fréquemment dans un ensemble de données

→ Plus utile pour les variables catégorielles



- Dans notre exemple, → mode = 21

# **Session 3 : Tests statistiques en EIC**

**C4ED - EUTF**

**Octobre 2023**



# Statistiques Descriptives – Résumé

- Les **statistiques descriptives** offrent une présentation compacte de données complexes
- Données catégorielles → Fréquences, proportions, tableaux (une variable) ou tableaux croisés
- Données continues → Mesures de tendance centrale:
  - Moyenne → intuitif, mais problématique avec des données asymétriques (càd en présence de valeurs aberrantes)
  - Médiane → robuste aux valeurs aberrantes



# Statistiques Descriptives – Résumé (suite)

- En présence de valeurs aberrantes (données asymétriques), les mesures de tendance centrale peuvent donner une image erronée de la distribution.
  - Utiliser des **mesures de dispersion** en complément
- Mesures de dispersion → Min-max; Variance et Ecart-type
  - Variance/Ecart-type élevés → beaucoup de variabilité dans les données
  - Ecart-type = écart moyen à la moyenne pour tous les points de données
- Variance/Ecart-type → ingrédient clé en statistique inférentielle
  - Dans quelle mesure pouvons-nous être sûrs que la moyenne d'un échantillon représente la moyenne de l'ensemble de la population ?
  - Dans quelle mesure pouvons-nous être sûrs que l'impact mesuré pour un programme reflète l'impact *réel* ?



Statistiques  
descriptives

Analyse de  
régression



Tests  
statistiques



Center for Evaluation  
and Development

# Statistiques inférentielles – Intro

---



- Si vous disposez d'un système de suivi solide, vous pouvez disposer de données sur **tous les** participants à une intervention.
  - Toutefois, dans la pratique, il arrive souvent que nous ne disposions pas de toutes les informations sur tous les participants
  - Du coup, comment pouvons-nous connaître les caractéristiques de *tous* les participants au programme ?
- Nous pouvons prélever un *échantillon représentatif* de la population.
- Veuillez vous référer aux diapositives de l'année 2 pour des informations sur l'échantillonnage



- Si nous utilisons un échantillon, nous devons estimer dans quelle mesure nous sommes **sûrs** de pouvoir appliquer nos conclusions à l'ensemble de la population
  - C'est ce que l'on appelle notre capacité à faire des déductions – ou « inférer » – sur la base de notre échantillon.
- Nous pouvons également utiliser des statistiques inférentielles pour estimer le niveau de confiance lorsqu'on mesure un impact causal





# Tests Statistiques

- 2 000 candidats au programme → 1 000 sélectionnés au hasard
  - ECR rigoureux → Toute différence entre les deux groupes doit être attribuable au programme !
- Les données montrent que:

	Nombre de personnes	Revenu moyen
Formation professionnelle	1 000	1 200
Pas de formation professionnelle	1 000	1 050



- Le programme a-t-il eu un *impact* sur les revenus? Dans quelle mesure peut-on être sûrs que cette différence reflète l'*impact réel* du programme?



# Tests Statistiques

---

- Nous allons maintenant aborder la théorie des probabilités et des méthodes statistiques plus avancées
- Les tests statistiques commencent par une **hypothèse** que l'on souhaite tester
- Cette hypothèse sera guidée par la (les) question(s) de recherche/évaluation qui vous intéresse(nt)
- Les tests statistiques nous permettent de dépasser les hypothèses et les anecdotes et de voir si des preuves quantitatives confirment notre théorie



Nous allons aborder les concepts suivants:

- Hypothèse de test et Hypothèse Nulle
- Sources d'incertitude dans les statistiques inférentielles
  - Incertitude due à l'utilisation d'échantillons → Niveau de confiance/Significativité
  - Erreur d'échantillonnage → Erreur Type
- Les étapes générales des tests statistiques
- Test d'égalité des moyennes → test  $t$  (test de Student)
- Décision sur un test → valeur  $p$  et significativité statistique
- Au-delà des tests statistiques
- ANNEXE : Détails sur les valeurs critiques, l'erreur standard, les formules de test  $t$  et le processus général des tests statistiques



Center for Evaluation  
and Development

# L'Hypothèse de Test

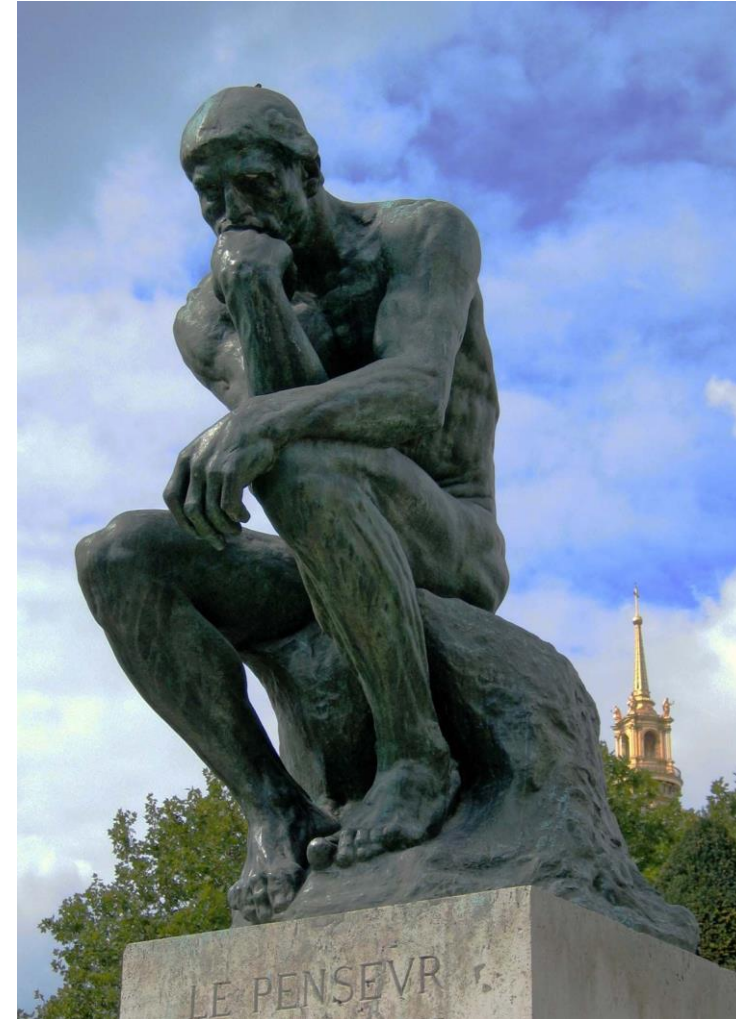
---



# Tests Statistiques

## *Hypothèse*

- Le meilleur point de départ pour les tests statistiques est de réfléchir à une hypothèse
- Hypothèse = proposition/explication servant de point de départ à des recherches plus approfondies

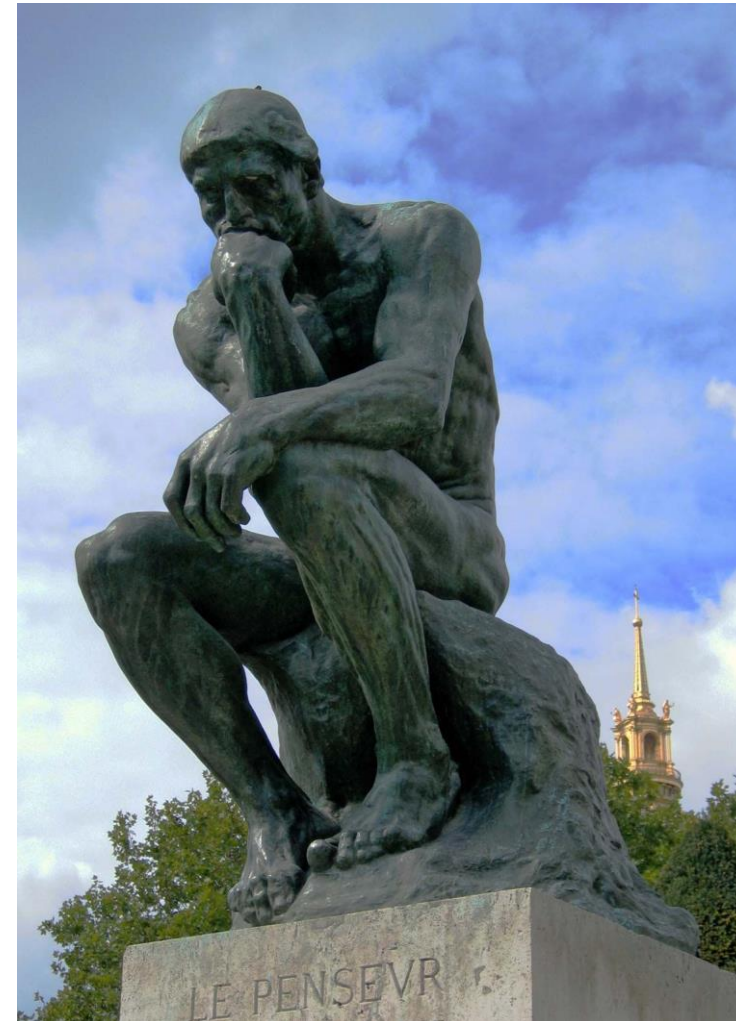




# Tests Statistiques

## *Hypothèse*

Plus le niveau  
d'éducation d'une  
personne est élevé,  
plus son revenu sera  
élevé...





# Tests Statistiques

## *Hypothèse Nulle ( $H_0$ )*

- Strictement parlant, en statistique, nous ne testons pas directement notre hypothèse
- Formellement, nous formulons une hypothèse nulle (notée  $H_0$ ) et testons si nous pouvons la *rejeter*
- L'hypothèse nulle est généralement formulée en ces termes : "il n'y a pas d'effet/de différence«
- Par exemple :
  - Notre hypothèse initiale : « Plus le niveau d'éducation d'une personne est élevé, plus son revenu sera élevé... »
  - L'hypothèse nulle associée : « La différence de revenu entre les personnes ayant terminé l'école primaires et celles ne l'ayant pas terminée est de 0. »



# Tests Statistiques

## *Hypothèse Nulle ( $H_0$ )*

- Hypothèse Nulle ( $H_0$ ): « La différence de revenu entre les personnes ayant terminé l'école primaires et celles ne l'ayant pas terminée est de 0. »
- En d'autres termes, le test statistique nous indique dans quelle mesure nous pouvons être *sûrs* de *rejeter* l'hypothèse nulle
- Si nous pouvons rejeter l'hypothèse nulle avec suffisamment de confiance, nous pouvons conclure que la différence de revenu entre les deux groupes est ***statistiquement significative***, c'est-à-dire que nous sommes convaincus que la différence observée reflète la réalité
- Bien que la distinction entre hypothèse principale et  $H_0$  soit importante d'un point de vue formel, par souci de simplicité nous parlerons en général du test de l'hypothèse principale



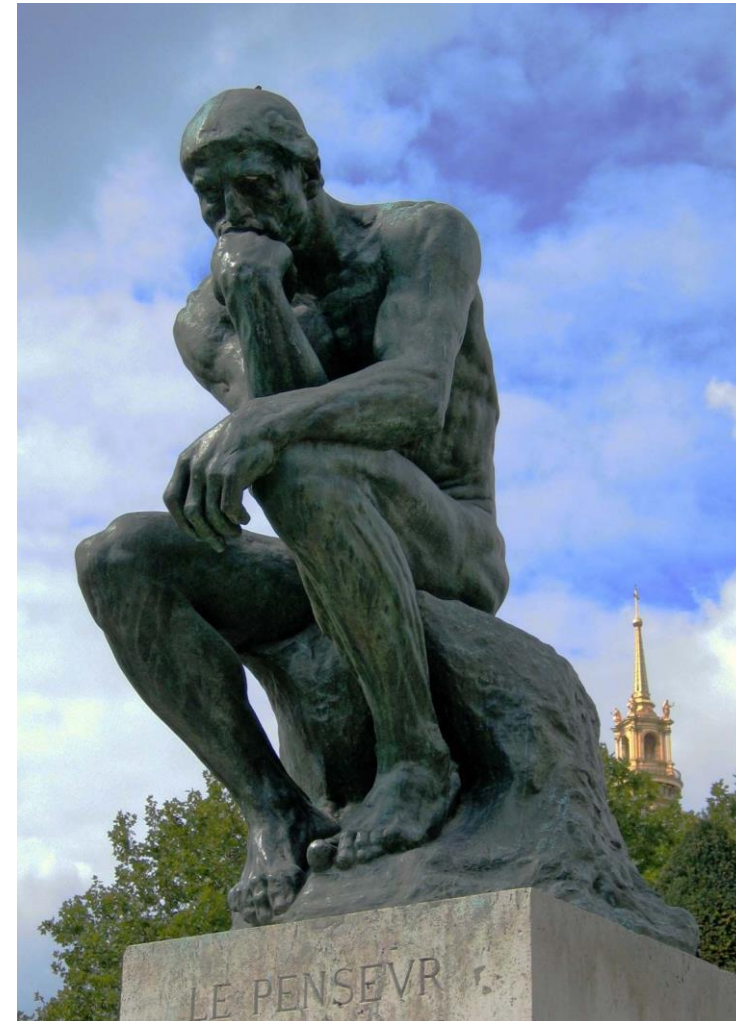


# Tests Statistiques

## *Hypothèse*

Plus le niveau  
d'éducation d'une  
personne est élevé,  
plus son revenu sera  
élevé...

- Comment pourrions-nous tester cette hypothèse ?





Center for Evaluation  
and Development

# Hypothèse, Population et Echantillon

---



# Tests Statistiques

## *Hypothèse, Population, Echantillon*

- L'hypothèse que nous voulons tester se réfère généralement à une *population* spécifique d'intérêt
  - **Population** = ensemble complet de tous les objets ou personnes d'intérêt. Elle peut être très vaste.
- En pratique, on dispose rarement d'informations sur l'ensemble de la population.
- On sélectionne un échantillon *représentatif*
  - **Échantillon** = sous-ensemble de la population concernée. En statistique inférentielle, les chercheurs utilisent généralement un échantillon pour tirer des conclusions sur la population.
- Veuillez vous référer au matériel de formation de l'année 2 pour plus de détails sur les différentes approches d'échantillonnage.



# Tests Statistiques

## *Hypothèse, Population, Echantillon*

---

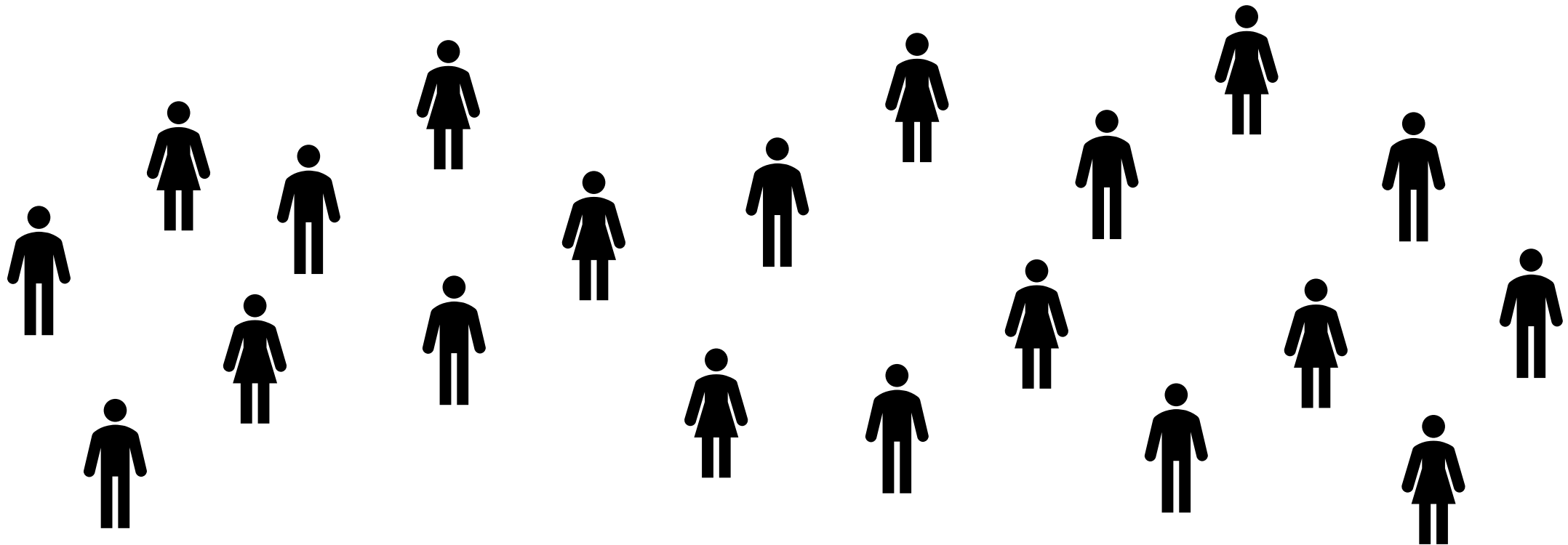
- Hypothèse: « Plus le niveau d'éducation d'une personne est élevé, plus son revenu sera élevé... »  
→ Population d'intérêt= toutes les personnes en âge de travailler
- Il est peu probable que nous puissions recueillir des données sur l'éducation et le revenu pour l'ensemble de la population concernée !  
→ On sélectionne un échantillon représentatif et on réalise une enquête pour recueillir les informations requises



# Tests Statistiques

## *Hypothèse, Population, Echantillon*

- Supposons, pour simplifier, qu'il s'agisse d'un pays extrêmement petit de 20 habitants seulement

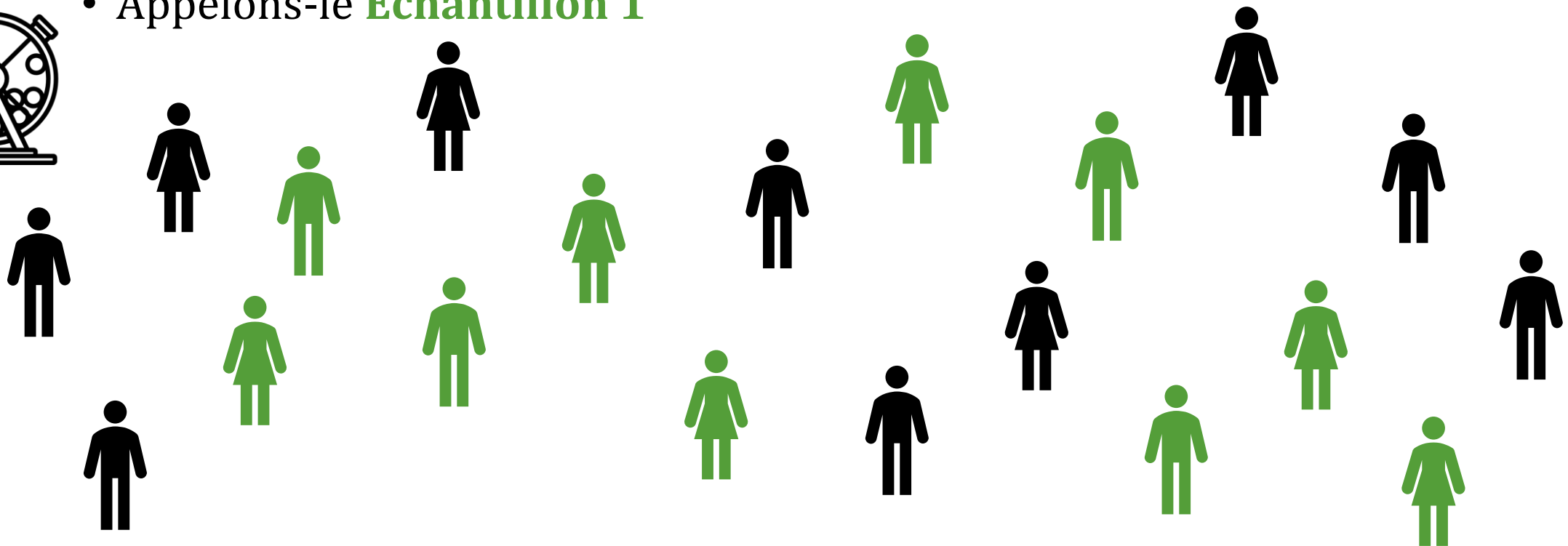




# Tests Statistiques

## *Hypothèse, Population, Echantillon*

- On sélectionne un échantillon aléatoire de 10 personnes et leur demande si elles ont terminé l'école primaire - notre mesure de l'éducation - et leur revenu mensuel...
- Appelons-le **Echantillon 1**





# Tests Statistiques

Les informations  
recueillies par le biais  
d'une enquête sur un  
sous-ensemble  
sélectionné de la  
population peuvent  
être appelées  
**données d'enquête**  
ou **données**  
**d'échantillon**

Nom	A terminé l'école primaire	Revenu mensuel
David	Oui	1 400
Michael	Oui	1 100
Anna	Non	1 000
Monica	Oui	1 200
Emma	Non	900
April	Oui	1 200
Frank	Non	1 300
Daniel	Oui	1 500
Jennifer	Non	800
Jodie	Non	950



# Tests Statistiques

## *Comment faire un test statistique?*

---

- Comment tester l'hypothèse selon laquelle les personnes ayant un niveau d'éducation plus élevé - c'est-à-dire ayant terminé l'école primaire - gagnent un revenu mensuel plus important ?
- Commençons par créer deux groupes - ceux qui ont terminé l'école primaire et ceux qui ne l'ont pas terminée - et comparons simplement le revenu mensuel moyen entre les deux groupes





# Tests Statistiques

## *Comment faire un test statistique?*

### Ont terminé l'école primaire

Nom	A terminé l'école primaire	Revenu mensuel
David	Oui	1 400
Michael	Oui	1 100
Monica	Oui	1 200
April	Oui	1 200
Daniel	Oui	1 500
	Moyenne	1 280

### N'ont pas terminé l'école primaire

Nom	A terminé l'école primaire	Revenu mensuel
Anna	Non	1 000
Emma	Non	900
Frank	Non	1 300
Jennifer	Non	800
Jodie	Non	950
	Moyenne	990

- Les données de l'échantillon montrent que le revenu mensuel moyen des personnes ayant terminé l'école primaire (1 280) est plus élevé que celui des personnes ne l'ayant pas terminée (990).
- Ces éléments sont-ils suffisants pour affirmer que notre hypothèse est vraie ?



Center for Evaluation  
and Development

# Sources d'incertitude en statistique inférentielle

---



# Tests Statistiques

## *Incertitude dans les tests statistiques*

Les personnes qui ont terminé l'école primaire gagnent en moyenne plus que celles qui ne l'ont pas terminée

Revenu Mensuel Moyen – Par groupe

Ecole primaire	Pas d'école primaire	<i>Différence</i>
1 280	990	290

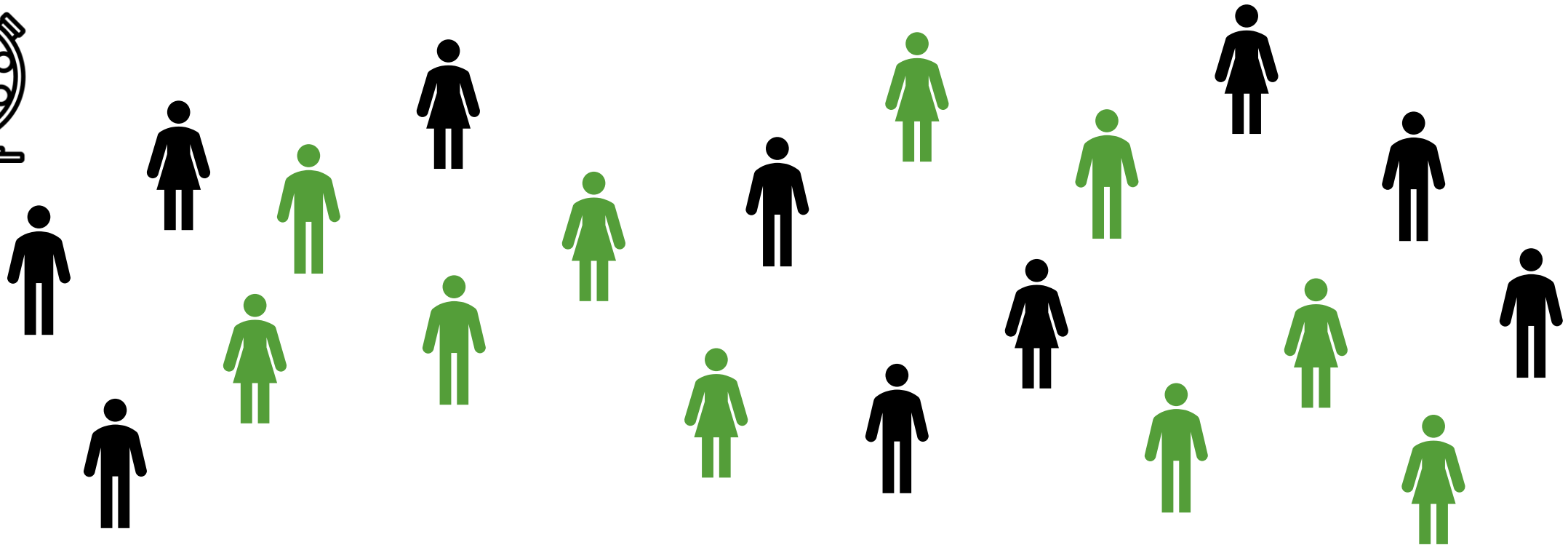
- Cette affirmation est vraie pour notre échantillon - c'est-à-dire un sous-ensemble de la population - mais...
- ...nous ne savons pas si les *moyennes calculées dans l'échantillon* sont une mesure précise des *véritables revenus moyens dans la population*
  - Et si cette différence n'existait que dans cet échantillon spécifique et non dans l'ensemble de la population ? Autrement dit, cette différence pourrait-elle être due au hasard ?
  - **Source d'incertitude n° 1** : essayer de tirer des conclusions sur *l'ensemble de la population* à partir d'*informations provenant d'un échantillon*



# Tests Statistiques

## *Echantillonnage et incertitude*

- Et si, par hasard, nous n'avions pas sélectionné ces 10 personnes et avons choisi un autre échantillon de 10 personnes du même pays ?

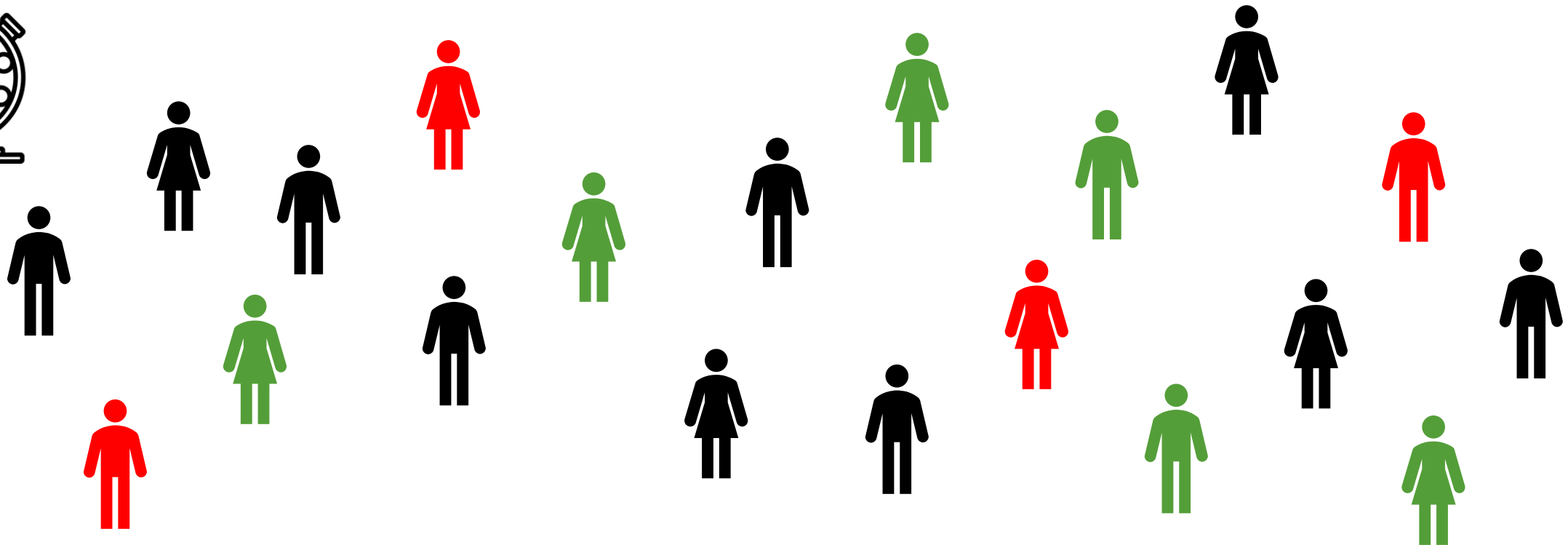




# Tests Statistiques

## *Echantillonnage et incertitude*

- Et si, par hasard, nous n'avions pas sélectionné ces 10 personnes et avons choisi un autre échantillon de 10 personnes du même pays ?
- Appelons-le **Echantillon 2**





# Tests Statistiques

## *Echantillonnage et incertitude*

Voici les nouvelles  
données de  
l'échantillon

Nom	A terminé l'école primaire	Revenu mensuel
John	Oui	900
Michael	Oui	1 100
Anna	Non	1 000
Monica	Oui	1 200
Barbara	Non	1 300
April	Oui	1 200
Frank	Non	1 300
Frederic	Oui	950
Jennifer	Non	800
Claire	Non	1 400



# Tests Statistiques

## *Echantillonnage et incertitude*

- Comparons de nouveau le revenu moyen dans les deux groupes

### Ont terminé l'école primaire

Nom	A terminé l'école primaire	Revenu mensuel
John	Oui	900
Michael	Oui	1 100
Monica	Oui	1 200
April	Oui	1 200
Frederic	Oui	950
	Moyenne	<b>1 070</b>

### N'ont pas terminé l'école primaire

Nom	A terminé l'école primaire	Revenu mensuel
Anna	Non	1 000
Barbara	Non	<b>1 300</b>
Frank	Non	1 300
Jennifer	Non	800
Claire	Non	<b>1 400</b>
	Moyenne	<b>1 160</b>



Que s'est-il passé par rapport à l'échantillon précédent?



# Tests Statistiques

## *Echantillonnage et incertitude*

### Ont terminé l'école primaire

Nom	A terminé l'école primaire	Revenu mensuel
John	Oui	900
Michael	Oui	1 100
Monica	Oui	1 200
April	Oui	1 200
Frederic	Oui	950
	Moyenne	<b>1 070</b>

### N'ont pas terminé l'école primaire

Nom	A terminé l'école primaire	Revenu mensuel
Anna	Non	1 000
Barbara	Non	1 300
Frank	Non	1 300
Jennifer	Non	800
Claire	Non	1 400
	Moyenne	<b>1 160</b>

- Dans le nouvel échantillon, le revenu mensuel moyen des personnes ayant terminé l'école primaire (1 070) est ***inférieur*** à celui des personnes n'ayant pas terminé l'école primaire (1 160)





# Tests Statistiques

## *Echantillonnage et incertitude*

### Revenu Mensuel Moyen – Par groupe

Echantillon	Ecole primaire	Pas d'école primaire	<i>Différence</i>
Echantillon 1	1 280	990	290
Echantillon 2	1 070	1 160	-90

- Les deux échantillons donnent des conclusions opposées !
- Cet exemple est assez extrême, mais il illustre bien le propos :  
→ **Source d'incertitude n° 2** : le *choix d'un échantillon* génère une incertitude - c'est ce qu'on appelle l'**erreur d'échantillonnage**.



# Tests Statistiques

## *Tests statistiques et incertitude*

- En résumé, il y a 2 sources d'incertitude:
  - **Source d'incertitude n° 1** : essayer de tirer des conclusions sur *l'ensemble de la population* à partir d'*informations provenant d'un échantillon*
  - **Source d'incertitude n° 2** : le *choix d'un échantillon* génère une incertitude - c'est ce qu'on appelle l'**erreur d'échantillonnage**.
- L'objectif de la statistique inférentielle est de tenir compte de ces deux types d'incertitude lors du test d'une hypothèse
- Voyons comment cela fonctionne



Center for Evaluation  
and Development

# Incertitude dans les tests statistiques

---



# Tests Statistiques

## *Tests statistiques et incertitude*

- Utilisons les données de l'échantillon 1 et concentrons-nous sur le revenu mensuel moyen (indépendamment du niveau d'éducation)
- Moyenne de l'échantillon = 1 135
- Rappelons notre problème principal :
  - Nous ne savons pas si la *moyenne calculée dans l'échantillon* est une mesure précise de la *véritable moyenne dans la population*
- En d'autres termes, dans quelle mesure le revenu moyen réel de la population est-il proche de 1 135 ?
- Pour répondre, on peut utiliser un **intervalle de confiance**

Nom	Revenu Mensuel
David	1 400
Michael	1 100
Anna	1 000
Monica	1 200
Emma	900
April	1 200
Frank	1 300
Daniel	1 500
Jennifer	800
Jodie	950



# Tests Statistiques

## *Intervalle de Confiance*

- Formellement, on calcule un intervalle de confiance comme suit:

$$\textit{Intervalle de Confiance (IC)} = \textit{Moyenne d'échantillon} \pm Z * \textit{Erreur Type}$$

- **Intervalle de Confiance** = consiste à calculer une borne inférieure et une borne supérieure pour la moyenne de la population à partir d'informations provenant d'un *échantillon*
- L'IC donne une fourchette dans laquelle nous nous attendons à ce que la "vraie" moyenne de la population se situe - avec un certain *niveau de confiance*
- **Erreur Type** = une mesure de l'écart auquel nous nous attendons entre la moyenne calculée dans un échantillon et la "vraie" moyenne dans l'ensemble de la population
- Elle tient compte de l'incertitude liée à l'**erreur d'échantillonnage**

$$\textit{Erreur Type} = \frac{\textit{Ecart - type}}{\textit{Racine carrée de la taille de l'échantillon}}$$



# Tests Statistiques

## *Intervalle de Confiance*

- Formellement, on calcule un intervalle de confiance comme suit:

$$\text{Intervalle de Confiance (IC)} = \text{Moyenne d'échantillon} \pm Z * \text{Erreur Type}$$

- La valeur de **Z** dans la formule contrôle le *niveau de confiance* que nous voulons pour notre fourchette
  - Elle prend en compte *l'incertitude liée à l'utilisation d'échantillons*
  - Le *niveau de confiance* le plus commun en sciences sociales 95% – notons que 90% et 99% sont aussi communs
- Voyons un exemple des différents concepts statistiques introduits là



# Tests Statistiques

## *Intervalle de Confiance – Exemple*

- Moyenne de l'échantillon = 1 135 ;
- Écart-type = 226,1 ;
- Taille de l'échantillon = 10 ;
- Erreur Type =  $226,1 / \sqrt{10} = 71,5$
  
- Pour calculer l'IC avec un niveau de confiance de 95 %, on fixe Z à 1,96 dans la formule :
  - Limite basse = 994,86
  - Limite haute = 1 275,14
  
- **Nous sommes sûrs à 95 % que le revenu mensuel moyen de *l'ensemble de la population* se situe entre 995 et 1 275**

Nom	Revenu Mensuel
David	1 400
Michael	1 100
Anna	1 000
Monica	1 200
Emma	900
April	1 200
Frank	1 300
Daniel	1 500
Jennifer	800
Jodie	950



# Tests Statistiques

## *Intervalle de Confiance – Exemple*

- Moyenne de l'échantillon = 1 135; Erreur Type = 71,5

Niveau de Confiance	Z	Limite basse	Limite haute	Risque de se tromper
90%	1,645	1 017,4	1 252,6	10%
95%	1,96	994,9	1 275,1	5%
99%	2,575	950,9	1 319,1	1%

- Plus le niveau de confiance est élevé, plus l'intervalle est large !
- En d'autres termes, plus on veut être sûr de la conclusion, plus on doit accepter de l'*imprécision* (càd un intervalle plus large)
- Vice versa: si on veut augmenter la précision (intervalle plus étroit), on doit accepter un risque plus élevé de se tromper...

Hmmm...  
Remarquez-vous  
quelque chose?







# Tests Statistiques

## *Intervalle de Confiance – Exemple*

- Moyenne de l'échantillon = 1 135; Erreur Type = 71,5

Niveau de Confiance	Z	Limite basse	Limite haute	Risque de se tromper
90%	1,645	1 017,4	1 252,6	10%
95%	1,96	994,9	1 275,1	5%
99%	2,575	950,9	1 319,1	1%

- Plus on veut être sûr de la conclusion, plus on doit accepter de *l'imprécision* (càd un intervalle plus large)
- Vice versa: si on veut augmenter la précision (intervalle plus étroit), on doit accepter un risque plus élevé de se tromper...

**Mais... Peut-être y a-t-il un moyen d'obtenir un intervalle plus étroit sans augmenter le risqué de se tromper?**





# Tests Statistiques

## *Taille d'échantillon et Incertitude*

- Peut-on obtenir un intervalle plus étroit sans augmenter le risque de se tromper?

*Intervalle de Confiance (IC) = Moyenne d'échantillon  $\pm$  Z \* Erreur Type*

$$\text{Intervalle de Confiance (IC)} = \text{Moyenne d'échantillon} \pm Z * \frac{\text{Ecart - type}}{\sqrt{\text{Taille d'échantillon}}}$$

Hmmm...  
Y a-t-il un paramètre  
que l'on pourrait  
influencer ici?





# Tests Statistiques

## *Taille d'échantillon et Incertitude*

- Peut-on obtenir un intervalle plus étroit sans augmenter le risque de se tromper?

$$\text{Intervalle de Confiance (IC)} = \text{Moyenne d'échantillon} \pm Z * \frac{\text{Ecart - type}}{\sqrt{\text{Taille d'échantillon}}}$$

- Oui, en **augmentant la taille de l'échantillon!**
- Pour un niveau de confiance *donné*:
  - ↑ taille d'échantillon ⇒ ↓ erreur type ⇒ ↑ Limite basse de l'IC et ↓ Limite haute de l'IC
    - **Intervalle plus étroit pour le même niveau de confiance, c'est le même risque de se tromper**



# Tests Statistiques

## *Taille d'échantillon et Incertitude – Exemple*

- Combinons les données sur les revenus de l'échantillon 1 et de l'échantillon 2, ce qui nous donne un échantillon de 14 personnes

- Moyenne de l'échantillon = 1 135,7
- Écart-type = 223,1
- Taille de l'échantillon = 14

→ IC 95% = [1 019 - 1 252,6]

Nom	Revenu Mensuel
David	1 400
Michael	1 100
Anna	1 000
Monica	1 200
Emma	900
April	1 200
Frank	1 300
Daniel	1 500
Jennifer	800
Jodie	950
John	900
Barbara	1 300
Frederic	950
Claire	1 400



# Tests Statistiques

## *Taille d'échantillon et Incertitude – Exemple*

	<b>Exemple 1</b>	<b>Exemple 2</b>
Moyenne de l'échantillon	1 135	1 135,7
Ecart-type	226,1	223,1
Taille d'échantillon	10	14
<b>Erreur Type</b>	<b>71,5</b>	<b>59,6</b>
<b>Intervalle de Confiance 95%</b>	<b>[995 – 1 275]</b>	<b>[1 019 – 1 252,6]</b>

- Comme prévu, en prenant un plus grand échantillon de la population, l'erreur type diminue et l'intervalle est plus étroit
- Grâce aux données supplémentaires, nous sommes maintenant sûrs à 95 % que le véritable revenu moyen de la population se situe entre 1 019 et 1 253



# Tests Statistiques

## *Incertitude en Statistique Inférentielle – Résumé*

- En résumé, il y a 2 sources d'incertitude:
  1. Incertitude parce que nous voulons tirer des conclusions sur *l'ensemble de la population concernée* à partir d'*informations provenant d'un échantillon*
  2. Incertitude due à *l'échantillonnage* → calculer le revenu moyen dans différents échantillons donnera des réponses différentes.
- L'objectif des statistiques inférentielles est de prendre en compte les deux types d'incertitude



# Tests Statistiques

## *Incertitude en Statistique Inférentielle – Résumé*

- Les intervalles de confiance sont un exemple de statistiques inférentielles :
  1. Incertitude due à *l'utilisation d'échantillons*
    - Elle est exprimée par le **niveau de confiance**
  2. Incertitude due à *l'échantillonnage*
    - Elle est prise en compte par **l'erreur type**.
- Pour l'essentiel, les tests statistiques en général utilisent les mêmes ingrédients - erreur type, niveau de confiance - et une intuition similaire à celle des intervalles de confiance.



Center for Evaluation  
and Development

# Les tests statistiques en pratique

---





# Tests statistiques

## *Tests statistiques en pratique*

- Les étapes principales d'un test statistique :
  1. Formuler l'hypothèse nulle ( $H_0$ )
  2. Utiliser les données d'échantillon pour calculer une **statistique de test**
    - Prend en compte l'incertitude due à l'erreur d'échantillonnage
  3. Pour un niveau de confiance donné, trouver la **valeur critique (Z précédemment)**
    - Prend en compte l'incertitude due à l'utilisation d'échantillons pour inférer sur les caractéristiques de la population
  4. Prendre la décision comme suit:
    - Statistique de test  $<$  valeur critique  $\rightarrow$  nous ne pouvons pas rejeter  $H_0$
    - Statistique de test  $>$  valeur critique  $\rightarrow$  nous pouvons rejeter  $H_0$



# Tests statistiques

## *Tests statistiques en pratique – La valeur $p$*

---

- En pratique, il n'est pas nécessaire de trouver la valeur critique et de la comparer à la statistique de test soi-même
  
- Les logiciels statistiques le font pour nous et calculent la **valeur  $p$**  du test
  
- **Valeur  $p$**  = probabilité de se tromper lorsque l'on rejette l'hypothèse nulle



# Tests statistiques

## *Remarques sur valeur p*

- *Remarque 1* : La valeur p est une probabilité, elle est donc toujours comprise entre 0 et 1.
- *Remarque 2* : La valeur p est comparée au **niveau de significativité**.
- *Remarque 3* : Le niveau de significativité et le niveau de confiance (dont nous avons parlé précédemment) sont étroitement liés
  - Niveau de confiance = 100% – niveau de significativité

Niveau de Confiance	Z	Limite basse	Limite haute	Risque de se tromper
90%	1,645	1 017,4	1 252,6	10%
95%	1,96	994,9	1 275,1	5%
99%	2,575	950,9	1 319,1	1%



# Tests statistiques

## *Valeur $p$ et décision*

- Les niveaux de significativité conventionnels en sciences sociales sont 10%, 5% et 1%, correspondant respectivement à 90%, 95% et 99% de confiance

Niveau de significativité	Valeur $p$	Interprétation	Conclusion du test
10%	0,1; 0,09; 0,08;...	Il y a (au plus) 10% de chances de se tromper si l'on rejette l'hypothèse nulle	On peut rejeter l'hypothèse nulle avec un degré de confiance de 90%
5%	0,05; 0,045; 0,04;...	Il y a (au plus) 5% de chances de se tromper si l'on rejette l'hypothèse nulle	On peut rejeter l'hypothèse nulle avec un degré de confiance de 95%
1%	0,01; 0,0099; 0,0098;...	Il y a (au plus) 1% de chances de se tromper si l'on rejette l'hypothèse nulle	On peut rejeter l'hypothèse nulle avec un degré de confiance de 99%



# Tests statistiques

## *Tests statistiques en pratique – Test t*

- Dans notre exemple, l'hypothèse nulle ( $H_0$ ) est:  
« La différence de revenu entre les personnes ayant terminé l'école primaires et celles ne l'ayant pas terminée est de 0. »
- Le test statistique le plus commun pour comparer la moyenne entre deux groupes s'appelle le **test t de Student** (souvent appelé simplement **test t**)
- Pour réaliser un test t, nous calculons la statistique t (c'est-à-dire la statistique associée au test t) et nous utilisons la valeur p
- Voyons un exemple



# Tests statistiques

## *Test $t$ et valeur $p$ – Exemple*

Plus le niveau  
d'éducation d'une  
personne est élevé, plus  
son revenu sera  
important...

Nom	A terminé l'école primaire	Revenu mensuel
David	Oui	1 400
Michael	Oui	1 100
Monica	Oui	1 200
April	Oui	1 200
Daniel	Oui	1 500
	Moyenne	1 280

Nom	A terminé l'école primaire	Revenu mensuel
Anna	Non	1 000
Emma	Non	900
Frank	Non	1 300
Jennifer	Non	800
Jodie	Non	950
	Moyenne	990



# Tests statistiques

## *Test t et valeur p – Exemple*

### Revenu Mensuel Moyen – Par groupe

Ecole primaire	Pas d'école primaire	Différence	Valeur p du test t
1 280	990	290	<b>0,032</b>

- On réalise un test t sur notre échantillon et on trouve une valeur p de **0,032**
- Quelle est la conclusion du test?





# Tests statistiques

## *Test t et valeur p – Exemple*

### Revenu Mensuel Moyen – Par groupe

Ecole primaire	Pas d'école primaire	Différence	Valeur p du test t
1 280	990	290	<b>0,032</b>

→ Quelle est la conclusion du test?

- On peut dire qu'il y a 3,2 % de chances que la différence entre les groupes soit due au hasard et n'existe pas dans la population
- En d'autres termes, nous pouvons affirmer avec une confiance d'au moins 95 % que les personnes qui terminent l'école primaire ont un revenu plus élevé dans la *population*
- Dans un rapport : « La différence de revenu moyen entre les deux groupes est statistiquement significative au niveau (de significativité) de 5 %. »





# Tests statistiques

## *Résumé*

- Les tests statistiques exploitent les informations d'un échantillon pour tirer des conclusions (c'est-à-dire inférer) sur la population
- Il faut définir l'hypothèse de test et l'hypothèse nulle associée
- Il faut fixer le niveau de significativité/confiance souhaité
- On utilise la valeur  $p$  pour conclure → Plus la valeur  $p$  est petite, plus nous sommes sûrs de nous quand on rejette l'hypothèse nulle



# Tests statistiques

## *Remarques*

---

- *Remarque 1* : Nous nous sommes focalisés sur un test de comparaison de moyennes, mais de nombreux autres tests existent en fonction du type de données (continues, catégorielles) et de ce que l'on veut comparer (variance, médiane, etc.).
  
- *Remarque 2* : La logique présentée ici est valable pour les tests statistiques autres que le test t. En d'autres termes, si vous connaissez l'hypothèse nulle et la valeur p, vous pouvez conclure !



- *Remarque 3* : nous avons vu que l'augmentation de la taille de l'échantillon pouvait réduire l'incertitude et améliorer les intervalles de confiance. La même intuition vaut pour le test t :
- Une taille d'échantillon plus grande  $\Rightarrow$  des informations sur une plus grande partie de la population  $\Rightarrow$  on réduit l'incertitude due à l'utilisation d'échantillons
- Une taille d'échantillon plus grande  $\Rightarrow$  on réduit l'incertitude due à l'échantillonnage
- Globalement, plus la taille de l'échantillon est grande  $\Rightarrow$  plus nous sommes sûrs de notre conclusion



Center for Evaluation  
and Development

# Au-delà des tests statistiques

---



### ATTENTION – Partie 1

- Un test statistique est une formule, il ne connaît pas le **contexte** de l'analyse, ni où ou comment vous avez obtenu vos données
- Les chercheurs doivent travailler dur pour fournir à la formule des données solides et contextualiser les résultats
- La première étape pour apporter de la crédibilité/confiance à vos tests est de sélectionner soigneusement l'échantillon !
- (Veuillez vous référer au matériel de formation de l'année 2 sur l'échantillonnage)



**WARNING**

**Statistical  
testing**

**Handle  
with care**

### ATTENTION – Partie 2

- Le simple fait de tester la différence de revenu moyen indique qu'il existe une différence entre deux groupes
- Le test *ne dit pas* que l'enseignement primaire mène à des revenus plus élevés ou que la différence de revenus est uniquement due à l'éducation
- Un test statistique *ne permet pas* à lui seul d'établir un lien de causalité



**WARNING**

**Statistical  
testing**

**Handle  
with care**



# Tests statistiques

## *Tests statistiques et causalité*

Ceux qui  
terminent l'école  
primaire ont un  
revenue moyen  
plus élevé



Terminer l'école  
primaire  
augmentera votre  
revenu

- Pouvons-nous prouver la deuxième hypothèse avec notre analyse actuelle ? → NON
- Pertinent pour l'évaluation des programmes car on s'intéresse à la causalité ! → EIC



# Tests statistiques

## *Tests statistiques et causalité*

- Dans notre échantillon, le groupe qui a terminé l'école primaire compte 3 hommes et 2 femmes
- Le groupe qui n'a pas fini l'école primaire compte 4 femmes et seulement 1 homme
- Faut-il en tenir compte lors de la vérification de notre hypothèse ?



Nom	A terminé l'école primaire	Revenu mensuel
David	Oui	1 400
Michael	Oui	1 100
Monica	Oui	1 200
April	Oui	1 200
Daniel	Oui	1 500
	Moyenne	1 280

Nom	A terminé l'école primaire	Revenu mensuel
Anna	Non	1 000
Emma	Non	900
Frank	Non	1 300
Jennifer	Non	800
Jodie	Non	950
	Moyenne	990





# Tests statistiques

## *Tests statistiques et causalité*

- Armés uniquement de notre test t, nous ne pouvons pas confirmer que la différence de revenus observée n'est pas due à des facteurs autres que l'éducation, tels que les phénomènes liés au genre (par exemple, les inégalités de salaires entre hommes et femmes).
- Comment y remédier ?
  - Conception minutieuse de l'EIC (année 1) et échantillonnage minutieux (année 2)
  - Analyse de régression → notre prochain sujet

**FIN DE LA SESSION 3**



Center for Evaluation  
and Development

# Session 3: Tests Statistiques en EIC

---

Annexe



## Session 3 – Annexe

---

### Etapas générales d'un test statistique



- Les étapes principales d'un test statistique :
  1. Formuler l'hypothèse nulle ( $H_0$ )
  2. Utiliser les données d'échantillon pour calculer une **statistique de test**
    - Prend en compte l'incertitude due à l'erreur d'échantillonnage
  3. Pour un niveau de confiance donné, trouver la **valeur critique (Z)**
    - Prend en compte l'incertitude due à l'utilisation d'échantillons pour inférer sur les caractéristiques de la population
  4. Prendre la décision comme suit:
    - Statistique de test  $<$  valeur critique  $\rightarrow$  nous ne pouvons pas rejeter  $H_0$
    - Statistique de test  $>$  valeur critique  $\rightarrow$  nous pouvons rejeter  $H_0$



## Session 3 – Annexe

### *Tests statistiques en pratique – Test t*

---

- Dans notre exemple, l'hypothèse nulle est :  
« La différence de revenu entre les personnes ayant terminé l'école primaires et celles ne l'ayant pas terminée est de 0. »
- Le test statistique le plus commun pour comparer la moyenne entre deux groupes s'appelle le **test t de Student** (souvent appelé simplement **test t**)
- La statistique de test associée s'appelle la **statistique t** (ou **stat-t**), ou **valeur t**



## Session 3 – Annexe

### *Tests statistiques en pratique – Test t*

- On utilise le **test t** pour comparer la moyenne entre deux groupes, pour lequel on calcule la **stat t**
- Exemple:
  - Hypothèse nulle: « La différence de revenu entre les personnes ayant terminé l'école primaires et celles ne l'ayant pas terminée est de 0. »
  - Stat  $t <$  valeur critique  $\rightarrow$  On ne peut pas rejeter l'hypothèse nulle  
 $\rightarrow$  Intuition: pas de preuves suffisantes pour dire avec certitude que les gens qui ont terminé l'école primaire ont un niveau de revenu (moyen) différent de ceux qui ne l'ont pas terminée
  - Stat  $t >$  valeur critique  $\rightarrow$  On peut rejeter l'hypothèse nulle  
 $\rightarrow$  Intuition: on a des preuves suffisantes pour dire avec certitude que les gens qui ont terminé l'école primaire ont un niveau de revenu (moyen) différent de ceux qui ne l'ont pas terminée



## Session 3 – Annexe

---

Erreur type





## Session 3 – Annexe

### *Echantillonnage et incertitude*

---

- Le problème:
  - On ne sait pas si la *moyenne calculée dans l'échantillon* est une *mesure précise de la moyenne dans la population*
- On peut obtenir une idée de cette précision en calculant un **intervalle de confiance** – c`ad, en utilisant des données d'échantillon, dans quelle mesure peut-on être sûr que la "vraie" moyenne dans la population est comprise dans un certain intervalle de valeurs
- Un des éléments clés des intervalles de confiance est **l'erreur type**



## Session 3 – Annexe

### *Echantillonnage et incertitude – Erreur type*

---

- **Erreur type** = elle mesure le degré de divergence auquel on peut s'attendre entre la moyenne calculée dans un échantillon et la “vraie” moyenne dans la population
- Intuition:
  - Imaginez qu'on calcule la moyenne dans plusieurs échantillons différents issus de la même population
  - On s'attend à ce que la moyenne calculée diffère entre les échantillons – comme dans notre exemple
  - *L'erreur type* indique le degré de *variabilité* auquel on peut s'attendre entre les moyennes calculées pour différents échantillons



## Session 3 – Annexe

### *Echantillonnage et incertitude – Erreur type*

---

- **Erreur type** = elle mesure le degré de divergence auquel on peut s'attendre entre la moyenne calculée dans un échantillon et la “vraie” moyenne dans la population
- Formellement, l'erreur type de la moyenne est calculée comme suit:

$$\textit{Erreur type} = \frac{\textit{Ecart – type}}{\textit{Racine carrée de la taille d'échantillon}}$$

- Voyons un exemple



## Session 3 – Annexe

### *Erreur type – Exemple*

- On utilise les données de l'échantillon 1 sur le revenu

Nom	Revenu Mensuel
David	1 400
Michael	1 100
Anna	1 000
Monica	1 200
Emma	900
April	1 200
Frank	1 300
Daniel	1 500
Jennifer	800
Jodie	950

Moyenne de l'échantillon = 1 135

Ecart-type = 226,1

Taille d'échantillon = 10

→ **Erreur type = 71,5**



## Session 3 – Annexe

### *Erreur type – Exemple*

- Moyenne de l'échantillon = 1 135
- Erreur type = 71,5
- L'erreur type signifie que:
  - Si on prend de nombreux échantillons (aléatoires) de 10 personnes issus de la même population, et calcule la moyenne dans chaque échantillon, celle-ci sera égale à  $1\ 135 \pm 71,5$
  - En d'autres termes, la moyenne calculée dans n'importe quel échantillon (aléatoire) de 10 personnes issu de cette population sera comprise entre 1 063,5 et 1 206,5
- L'erreur type est utile en statistiques inférentielles pour calculer les intervalles de confiance ou les statistiques de test



## Session 3 – Annexe

---

### Valeurs critiques et formules de test t



## Session 3 - Annexe

### *Valeur Critique Z et Intervalles de Confiance*

*Intervalle de Confiance (IC) = Moyenne d'échantillon  $\pm$  Z \* Erreur Type*

- **Z** = valeur critique
- La valeur critique est issue d'une distribution statistique connue (théorique) et change avec le niveau de significativité/confiance désiré
- Pour les **test t**, dans les petits échantillons ( $N < 30$ ), les valeurs critiques sont issues de la **distribution t de Student**, qui prend en compte la taille limitée de l'échantillon
- Pour les échantillons où  $N > 30$ , les valeurs critiques sont issues de la distribution Normale (centrée réduite)
  - Nous n'utilisons pas les valeurs critiques du t de Student dans nos exemples car, en pratique, on a rarement des échantillons comptant moins de 30 observations, et les valeurs critiques issues de la distribution Normale sont des nombres communs/connus que je voulais vous montrer



## Session 3 – Annexe

### *Formule du test t à un échantillon*

- Test t à *un échantillon* → utilisé pour tester si la moyenne dans la population est égale à une valeur  $\mu$

$$t = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

$\bar{x}$  = moyenne de l'échantillon

$\mu$  = valeur hypothétique pour la moyenne de la *population* (peut être 0)

$\sigma$  = écart-type (de la moyenne de l'échantillon)

$n$  = taille de l'échantillon

$\sigma / \sqrt{n}$  = erreur type





## Session 3 – Annexe

### *Formule du test t à deux échantillons*

- Test t à *deux échantillons* → utilisé pour tester si les moyennes dans deux groupes sont statistiquement significativement différentes
  - Note: peut être entre 2 populations ou 2 groupes de la même population

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$\bar{x}_i$  = moyenne dans l'échantillon  $i = 1,2$

$\sigma_i^2$  = variance (càd le carré de l'écart-type) pour l'échantillon  $i = 1,2$

$n_i$  = taille de l'échantillon  $i = 1,2$

$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$  = une estimation de *l'erreur type* pour les deux échantillons combinés