

Section 5: Causalité à partir de données expérimentales

Visseho Adjiwanou, PhD.

04/10/2018

Plan du jour

- ① Travaux pratiques - Rappels statistiques et visualisations (90 minutes)
- ② Cours sur la causalité (90 minutes)
- ③ **Ressources**

- <https://www.rstudio.com/resources/cheatsheets/>
- <https://juba.github.io/tidyverse/04-bivarie.html>
- <http://r4ds.had.co.nz/>
- <http://r-statistics.co/Top50-Ggplot2-Visualizations-MasterList-R-Code.html>

Plan de présentation

- ➊ Questions causales en sciences sociales et terminologie
- ➋ Effets causaux et contrefactuel
- ➌ Essais contrôlés randomisés (*Randomized controlled trials*) et causalité
- ➍ Application
- ➎ Causalité à partir des données observationnelles (semaine prochaine)
- ➏ Lectures:
 - <https://www.bitbybitbook.com/fr/1st-ed/running-experiments/>
 - Sections 4.1, 4.2 et 4.3
 - <https://juba.github.io/tidyverse/04-bivarie.html>

Introduction

- Dans ce chapitre, nous considérons la causalité, l'un des concepts les plus centraux des sciences sociales quantitatives.
- Une grande partie de la recherche en sciences sociales s'intéresse aux effets causaux de diverses politiques et autres facteurs sociétaux.
- Par exemple:
 - les classes de petite taille augmentent-elles les résultats des tests standardisés des élèves?
 - Les soins de santé universels amélioreraient-ils la santé et les finances des pauvres?
 - L'éducation réduit-elle le nombre d'enfants?

Questions de recherche

- Une question de recherche est au cœur d'un projet de recherche, d'une étude ou d'une revue de littérature.
- Il concentre l'étude, détermine la méthodologie et guide toutes les étapes de la recherche, de l'analyse et de la production de rapports.
- Peut être **associatif** ou **causal**

Questions de recherche

- ❶ Le salaire minimum augmente-t-il le taux de chômage?
 - Le taux de chômage a augmenté après l'augmentation du salaire minimum
 - Le taux de chômage aurait-il augmenté si l'augmentation du salaire minimum n'avait pas eu lieu?
- ❷ La race/l'ethnie a-t-elle une incidence sur les perspectives d'emploi?
 - Mohamed a postulé pour un emploi mais ne l'a pas obtenu
 - Mohamed aurait-il trouvé un travail s'il était blanc (avait un nom européen)?
- ❸ Est-ce que fumer cause une maladie coronarienne?
 - Jean, fumeur, a eu une maladie coronarienne
 - Est-ce que Jean aurait eu la même maladie s'il n'était pas fumeur?

Questions de recherche

- ④ Quelle est l'importance des questions souverainistes dans la victoire de François Legault?
 - Au cours de ces élections, la question souverainiste a été laissée de côté et François Legault a gagné.
 - François Legault aurait-il gagné les élections si ces questions étaient présentes?
- ⑤ Maintenant, encore une fois, parler moi de vos questions de recherche.

Questions de recherche: Terminologie

1 Réponse ou variable dépendante, *outcome*

- C'est ce que nous voulons expliquer.
- *Exemples:*
 - Taux de chômage
 - Perspective d'emploi
 - Maladie coronarienne
 - Victoire de François Legault

2 Variable indépendante, facteurs de risque

- Tout facteur pouvant influencer la variable de réponse
- Peut être de différents niveaux
- Leur choix dépend de la théorie
- *Exemples:*
 - Salaire minimum
 - Ethnie / Race

Type de relation: Association

- On dit que deux variables A et B sont **associées** quand l'une se trouve plus communément en présence de l'autre.
- Se détecte souvent à partir d'un-e :
- tableau dit de **contingence** ou **tableau croisé** : dans le cas de deux variables qualitatives
- représentation graphique (boxplot) et de calcul d'indicateurs : dans le cas d'une variable qualitative et d'une variable quantitative
- représentation linéaire (scatterplot) et de calcul d'indicateurs (corrélation de Pearson): dans le cas de deux variables quantitatives.

Type de relation: Association

- Une association (linéaire) peut être
- **positive** si les deux variables vont dans le même sens (une augmentation de l'un est associée à une augmentation de l'autre);
- *Exemple* : éducation et revenu, durée de résidence et emploi
- **négative** si les deux variables vont dans des sens opposés (une augmentation de l'un est associée à une diminution de l'autre);
- *Exemple* : scolarisation et racisme, revenu et obésité, niveau de développement d'un pays et niveau de mortalité infantile
- **nulle** (Absence d'association).
- *Exemple*:
- Nous reviendrons sur les associations plus tard dans le cours.

Type de relation: Causalité

- L'association est une **condition nécessaire** de la causalité (Mais elle **n'est pas suffisante**).
- Toutes les associations ne sont pas causales. L'association peut arriver par hasard.
- **L'analyse statistique à elle seule ne peut constituer une preuve d'un lien de causalité**
- Comparaison entre **factuel** et **contrefactuel**
- Problème fondamental de l'inférence causale:
- il faut déduire des résultats contrefactuels
- Il n'y a pas de causalité sans manipulation: caractéristiques immuables

Type de relation: Causalité

- Contrefactuels ne sont pas observés, sauf dans les films.
- <https://www.youtube.com/watch?v=BvUbv4iwbDs&rel=0&modestbranding=1&autohide=1&showinfo=0>

Exemple 1 discrimination raciale sur le marché du travail

- **Question de recherche**
- La discrimination raciale existe-t-elle sur le marché du travail?
- Ou bien les disparités raciales dans le taux de chômage devraient-elles être attribuées à d'autres facteurs tels que les écarts raciaux dans le niveau d'instruction?
- **Expérimentation**
- En réponse aux annonces dans les journaux, les chercheurs ont envoyé les CV de candidats fictifs à des employeurs potentiels.
- Changé seulement le nom du demandeur d'emploi
 - Noms afro-américains
 - Noms à consonnance caucasienne
- Les autres informations sont inchangées
- **Variable dépendante**
- Taux de rappel

Exemple 1 discrimination raciale sur le marché du travail

Unité d'analyse: Individus

Variable de traitement (variable d'intérêt causal) **T**: Nom à consonnance afro-américain

Groupe de traitement (unités traitées): Afro-américains

Groupe de contrôle (unités non traitées): Caucasiens

Réponse (variable de réponse) **Y**: si un rappel a été effectué

- Que signifie “**T cause Y**”?
- Contrefactuels, “**Quoi si**” : Les Afro-Américains auraient-ils été rappelés s'ils n'avaient pas de noms afro-américains?

Exemple 1 discrimination raciale sur le marché du travail

Deux résultats possibles: $Y(1)$ et $Y(0)$

Effet causal: $Y(1) - Y(0)$

Problème fondamental d'inférence causale: un seul des deux résultats potentiels est observable

Importance du groupe de contrôle

Exemple 1 discrimination raciale sur le marché du travail

- Comment pouvons-nous comprendre les contrefactuels?
- L'association n'est pas un lien de causalité
- Trouvez une unité similaire! ==> **Matching**
- Es-ce-que Jamal n'a été rappelé à cause de sa race?
- Trouver une personne blanche qui ressemble à Jamal
- Le problème: on ne peut pas correspondre sur tout
- Facteurs de **confusion non observés**: variables associées au traitement et au résultat ==> **biais de sélection**

Exemple 1 discrimination raciale sur le marché du travail

- La clé pour comprendre la causalité est de penser au contrefactuel. L'inférence causale est une comparaison entre le factuel (ce qui s'est réellement passé) et le contrefactuel (ce qui se serait passé si une condition était différente).

CV (i)	Noms à consonnance afro-américain (Ti)	Appellé pour interview?		Age	Niveau d'éducation
		Yi(1)	Yi(0)		
1	1	1	?	25	Collège
2	0	?	0	55	Secondaire
3	0	?	1	40	Collège
n	1	0	?	62	Secondaire

Essais contrôlés randomisés (Expérimentation)

- Idée clé: la **randomisation** du traitement rend les groupes de **traitement** et de **contrôle** en moyenne «identiques»
- Les deux groupes sont similaires en termes de toutes les caractéristiques (**observées et non observées**)
- Peut attribuer les différences moyennes de résultats à la différence de traitement
- Effet du Traitement Moyen (Sample Average Treatment Effect, SATE)

$$SATE = \frac{1}{n} \sum_{i=1}^n \{Y_i(1) - Y_i(0)\}$$

- Essais contrôlés randomisés comme **norme d'excellence** (*Gold standard*)
- Expériences en double aveugle:
- Effets **placebo** et effets **Hawthorne**

Essais contrôlés randomisés (Expérimentation)

Dans un essai contrôlé randomisé (ECR), chaque unité est assignée de manière aléatoire au groupe de traitement ou au groupe de contrôle. La randomisation de l'assignation de traitement garantit que la différence moyenne de résultats entre les groupes de traitement et de contrôle peut être attribuée uniquement au traitement, car les deux groupes sont en moyenne identiques pour toutes les caractéristiques de prétraitement (observées et non observées).

Essais contrôlés randomisés (Expérimentation)

① Forces

- **Validité interne** - mesure dans laquelle les hypothèses de causalité sont satisfaites dans l'étude

② Limites

- **Validité externe** - mesure dans laquelle les conclusions peuvent être généralisées au-delà d'une étude particulière
- Explication causale faible
- Considérations éthiques
- Possibilité de contamination

Essais contrôlés randomisés (Expérimentation) : Exemple

Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomised screening trial



OPEN ACCESS

Anthony B Miller *professor emeritus*¹, Claus Wall *data manager*¹, Cornelia J Baines *professor emerita*¹, Ping Sun *statistician*², Teresa To *senior scientist*³, Steven A Narod *professor*^{1,2}

¹Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario M5T 3M7, Canada; ²Women's College Research Institute, Women's College Hospital, Toronto, Ontario M5G 1N8, Canada; ³Child Health Evaluative Services, The Hospital for Sick Children, Toronto, Ontario, Canada

Abstract

Objective To compare breast cancer incidence and mortality up to 25 years in women aged 40-59 who did or did not undergo mammography screening.

Design Follow-up of randomised screening trial by centre coordinators,

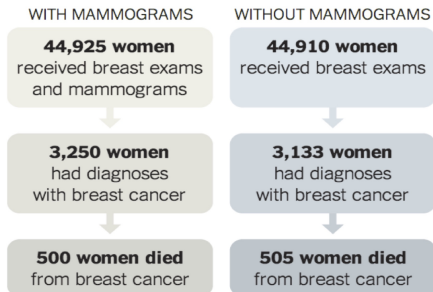
Conclusion Annual mammography in women aged 40-59 does not reduce mortality from breast cancer beyond that of physical examination or usual care when adjuvant therapy for breast cancer is freely available. Overall, 22% (106/484) of screen detected invasive breast cancers were over-diagnosed, representing one over-diagnosed breast cancer for every 424 women who received mammography screening in the trial.

Figure 2: A caption

Essais contrôlés randomisés (Expérimentation) : Exemple

Study Results

A large, 25-year study of Canadian women aged 40 to 59 found no benefit for women who were randomly assigned to have mammograms.



The death rate from breast cancer was the same in both groups, but **1 in 424 women** who had mammograms received unnecessary cancer treatment, including surgery, chemotherapy and radiation.

Pour la semaine prochaine

- Travaux Pratiques : Causalité : Pression sociale (sur Moodle demain)
- <https://www.bitbybitbook.com/fr/1st-ed/asking-questions/>
- Sections 3.1, 3.2 et 3.3
- <https://juba.github.io/tidyverse/04-bivarie.html>