

Causalité

Analyse tabulaire multivariée

Visseho Adjewanou, PhD.

Département de Sociologie - UQAM

28 Janvier 2019

Plan de présentation

- ① Questions causales en sciences sociales et terminologie
- ② Effets causaux et contrefactuel
- ③ Essais contrôlés randomisés (*Randomized controlled trials*) et causalité
- ④ Causalité à partir des données observationnelles

Introduction

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.

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**

Exemple 1

- ① 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?

Exemple 2

② 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)?

Exemple 3

- ③ 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?

Exemple 4

- ④ Quelle est l'importance des questions souverainistes dans la victoire de François Legault?
- Au cours de ces élections, la question souverainiste a été laissé de côté et François Legault a gagné.
 - François Legault aurait-il gagné les élections si ces questions étaient présentes?

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

Terminologie

② Variable indépendante, facteur 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
 - Fumer
 - Questions souverainistes

③ Variables de contrôle

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 tableau dit de **contingence** ou **tableau croisé** ou d'un graphique

Association

- Exemple - Existe-t-il une association entre le degré d'ouverture d'un pays et l'attitude face à la violence contre les femmes?

Pierotti, Rachel. (2013). "Increasing Rejection of Intimate Partner Violence: Evidence of Global Cultural Diffusion." *American Sociological Review*, 78: 240-265.

Nous utilisons les données des enquêtes démographiques et de santé (EDS), qui représentent un ensemble de plus de 300 enquêtes représentatives à l'échelle nationale, régionale et résidentielle menées dans des pays en développement du monde entier depuis 1992.

Association

Name	Description
beat_burnfood	Pourcentage de femmes dans chaque pays qui pensent qu'un mari a le droit de battre sa femme si elle brûle la nourriture (quantitative)
beat_burnfood_cat	Variable beat_burnfood en 4 catégories (qualitative)
no_media	Pourcentage de femmes dans chaque pays qui ont rarement accès un journal, une radio ou une télévision (quantitative)
no_media_cat	Variable no_media en 3 catégories (qualitative)
country	pays

Association

```
## [1] "/Users/visseho/OneDrive - UQAM/Cours/SOC2206_UQAM/S
```

Association

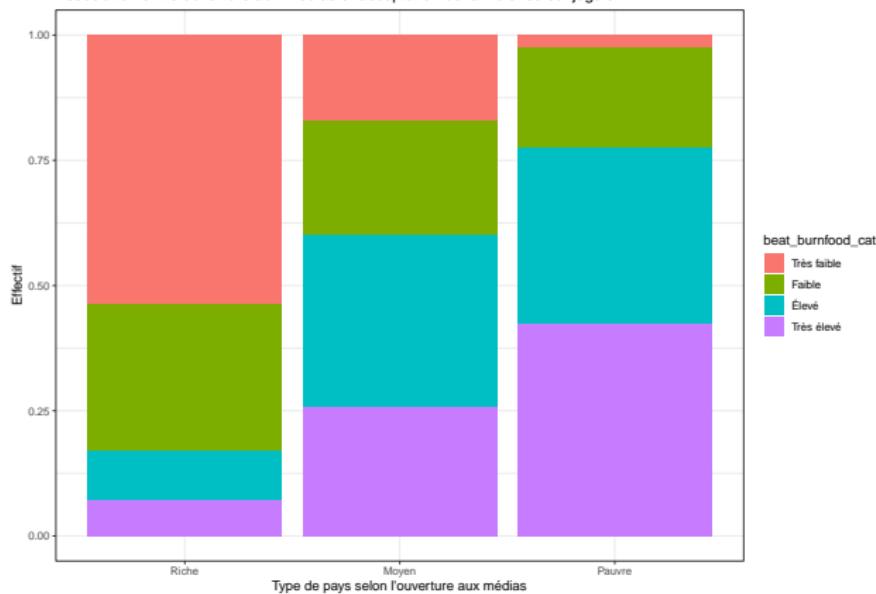
- Résumé de l'information contenue dans la base de donnée

```
##          ...1      beat_burnfood      beat_goesout      sec_
##  Min.   : 1.00   Min.   : 0.10   Min.   : 0.30   Min.
##  1st Qu.: 40.50  1st Qu.: 4.50   1st Qu.:11.85  1st Qu.
##  Median : 79.00  Median :11.85   Median :28.10  Median
##  Mean   : 80.53  Mean   :15.04   Mean   :28.60  Mean
##  3rd Qu.:119.50 3rd Qu.:22.25   3rd Qu.:42.08  3rd Qu.
##  Max.   :160.00  Max.   :64.50   Max.   :82.70  Max.
##                  NA's   :31      NA's   :27      NA's
##          no_media      country      year      re_
##  Min.   : 0.80   Length:151      Min.   :1999  Length
##  1st Qu.:11.25  Class :character 1st Qu.:2004  Class
##  Median :29.15  Mode   :character Median :2007  Mode
##  Mean   :28.40
```

Association

- Association entre deux variables qualitatives

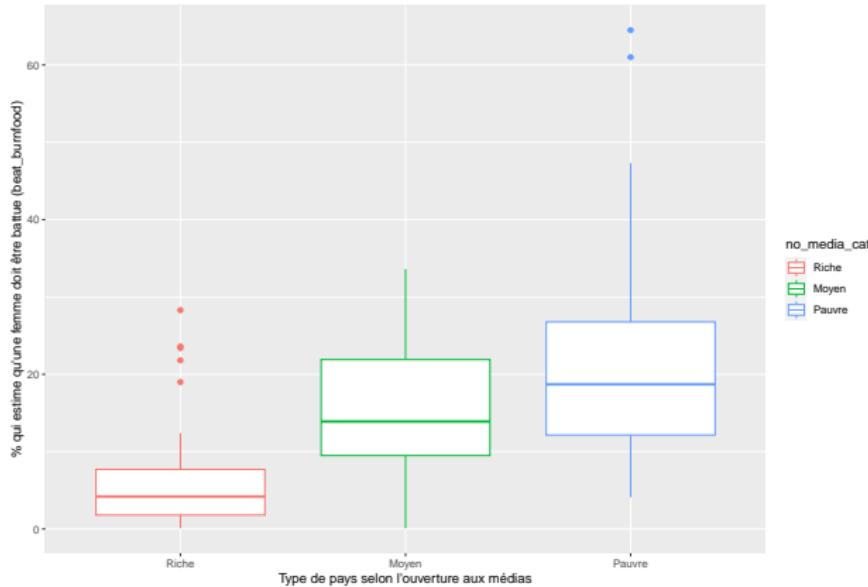
Association entre ouverture aux médias et acceptation de la violence conjugale



Association

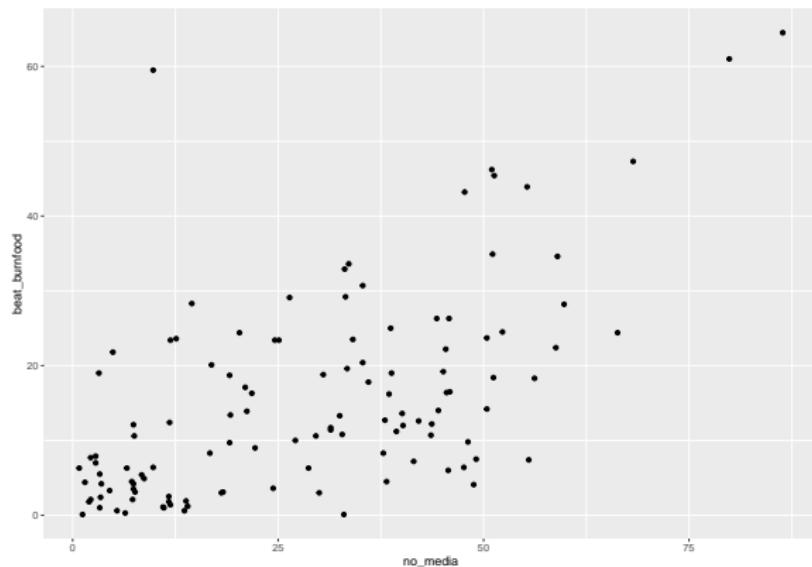
- Représentation graphique (boxplot) dans le cas d'une variable qualitative et d'une variable quantitative

Association entre ouverture aux médias et acceptation de la violence conjugale

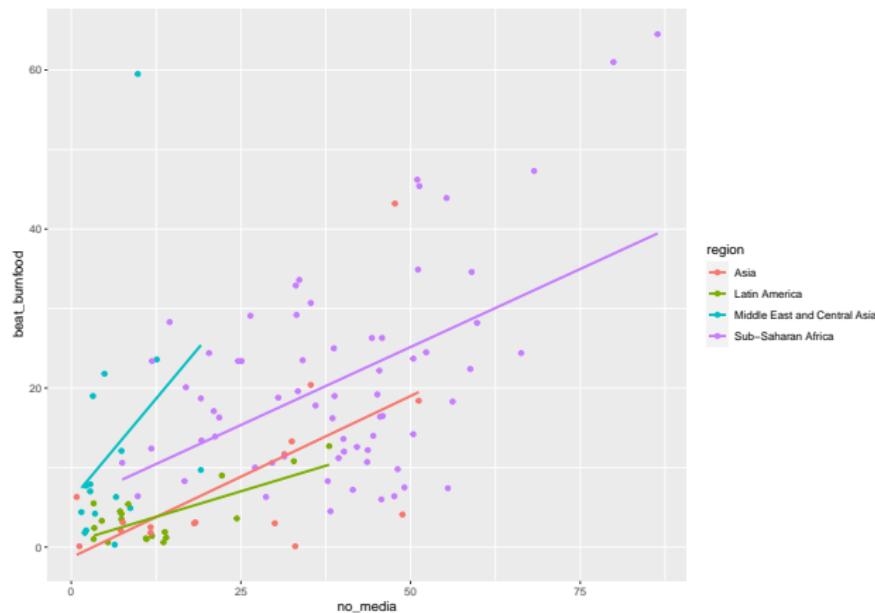


Association

- Représentation linéaire (scatterplot) et de calcul d'indicateurs (corrélation de Pearson): dans le cas de deux variables quantitatives.



Association



Relation associative

- 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*:

Relation causale

- L'association est une **condition nécessaire** à 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

Relation causale

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

Essais contrôlés randomisés (Expérimentation)

Essais contrôlés randomisés

- Idée clé: la **randomisation** du traitement rend les groupes de **traitement** et de **contrôle** en moyenne «identiques»

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

- Essais contrôlés randomisés comme **norme d'excellence** (*Gold Standard*)

Essais contrôlés randomisés

- 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**)

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

- Essais contrôlés randomisés comme **norme d'excellence** (Gold

Essais contrôlés randomisés

- 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

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

- Essais contrôlés randomisés comme **norme d'excellence** (Gold

Essais contrôlés randomisés

- 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

Essais contrôlés randomisés

- 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

① Forces

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

Essais contrôlés randomisés

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

Applications

Exemple 1 discrimination raciale sur le marché du travail

1 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?

Exemple 1 discrimination raciale sur le marché du travail

② 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 consonance caucasienne
- Les autres informations sont inchangées

③ Variable dépendante

- Taux de rappel

Exemple 1 discrimination raciale sur le marché du travail

- Unité d'analyse: Individus

Exemple 1 discrimination raciale sur le marché du travail

- **Unité d'analyse:** Individus
- **Variable de traitement** (variable d'intérêt causal) **T**: Nom à consonance afro-américain

Exemple 1 discrimination raciale sur le marché du travail

- **Unité d'analyse:** Individus
- **Variable de traitement** (variable d'intérêt causal) **T**: Nom à consonance afro-américain
- **Groupe de traitement** (unités traitées): Afro-américains

Exemple 1 discrimination raciale sur le marché du travail

- **Unité d'analyse:** Individus
- **Variable de traitement** (variable d'intérêt causal) **T**: Nom à consonance afro-américain
- **Groupe de traitement** (unités traitées): Afro-américains
- **Groupe de contrôle** (unités non traitées): Caucasiens

Exemple 1 discrimination raciale sur le marché du travail

- **Unité d'analyse:** Individus
- **Variable de traitement** (variable d'intérêt causal) **T**: Nom à consonance 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

Exemple 1 discrimination raciale sur le marché du travail

- **Unité d'analyse:** Individus
- **Variable de traitement** (variable d'intérêt causal) **T**: Nom à consonance 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**”?

Exemple 1 discrimination raciale sur le marché du travail

- **Unité d'analyse:** Individus
- **Variable de traitement** (variable d'intérêt causal) **T**: Nom à consonance 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

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**
 - Est-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 (T_i)	Appelé pour interview?		Age	Niveau d'éducation
		$Y_i(1)$	$Y_i(0)$		
1	1	1	?	25	Collège
2	0	?	0	55	Secondaire
3	0	?	1	40	Collège
n	1	0	?	62	Secondaire

Exemple 1 discrimination raciale sur le marché du travail

```
##   firstname    sex  race call
## 1 Allison female white    0
## 2 Kristen female white    0
## 3 Lakisha female black   0
## 4 Latonya female black   0
## 5 Carrie  female white   0
## 6 Jay     male   white   0
```

Exemple 1 discrimination raciale sur le marché du travail

```
freq(resume$sex)
```

```
## Frequencies
## resume$sex
## Type: Character
##
##          Freq  % Valid  % Valid Cum.  % Total
## ----- -----
##   female    3746    76.92      76.92    76.92
##   male      1124    23.08     100.00    23.08
##   <NA>         0
##   Total     4870    100.00     100.00   100.00
```

Exemple 1 discrimination raciale sur le marché du travail

```
freq(resume$race)
```

```
## Frequencies
## resume$race
## Type: Character
##
##          Freq  % Valid  % Valid Cum.  % Total
## black      2435    50.00    50.00    50.00
## white      2435    50.00   100.00    50.00
## <NA>           0
## Total     4870   100.00   100.00   100.00
```

Exemple 1 discrimination raciale sur le marché du travail

```
freq(resume$call)
```

```
## Frequencies
## resume$call
## Type: Integer
##
##          Freq   % Valid  % Valid Cum.  % Total
## ----- -----
##      0    4478    91.95    91.95    91.95
##      1     392     8.05   100.00     8.05
## <NA>      0
## Total  4870    100.00   100.00   100.00
```

Y'a-t-il discrimination ou pas?

```
#ctable(resume$race, resume$call)
tab <- table(resume$race, resume$call)
```

```
round(prop.table(tab, 1)*100, 2)
```

```
##
##          0      1
## black 93.55 6.45
## white 90.35 9.65
```

- SATE = 9,65 - 6,45 = 3,2%

Est-ce que les deux groupes étaient similaires au début?

```
ctable(resume$race, resume$sex)
```

```
## Cross-Tabulation, Row Proportions
```

```
## race * sex
```

```
## Data Frame: resume
```

```
##
```

##	sex	female	male	To
## race				
## black	1886	(77.5%)	549	(22.5%)
## white	1860	(76.4%)	575	(23.6%)
## Total	3746	(76.9%)	1124	(23.1%)
##	-----	-----	-----	-----

Causalité à partir des données observationnelles

Données observationnelles

- Souvent, nous ne pouvons pas randomiser le traitement pour des raisons éthiques et logistiques:
- par exemple, tabagisme et cancer du poumon
- Études observationnelles: traitement naturellement attribué
- Plans d'observation passifs ou plans corrélationnels
- Pas d'assignation aléatoire, pas de groupe de contrôle...

Données observationnelles

- Meilleure validité externe pour la généralisation au-delà de l'expérience
 - Validité interne plus faible:
 - les variables pré-traitement peuvent différer entre les groupes (traitement et contrôle)
- ① **biais de confusion (Confounding bias)** dû à ces différences :
Une variable de prétraitement associée aux variables de traitement et de résultat s'appelle un facteur de confusion et constitue une source de biais de confusion dans l'estimation de l'effet du traitement.

Données observationnelles

- Meilleure validité externe pour la généralisation au-delà de l'expérience
 - Validité interne plus faible:
 - les variables pré-traitement peuvent différer entre les groupes (traitement et contrôle)
- 1 biais de confusion (Confounding bias)** dû à ces différences : Une variable de prétraitement associée aux variables de traitement et de résultat s'appelle un facteur de confusion et constitue une source de biais de confusion dans l'estimation de l'effet du traitement.
- 2 biais de confusion non observée (Unobserved confounding)** constitue la menace la plus importante car il est inobservé.

Données observationnelles

- ③ **biais de sélection (selection bias)** de l'auto-sélection au traitement: Le biais de confusion dû à l'auto-sélection dans le groupe de traitement s'appelle un biais de sélection. Un biais de sélection apparaît souvent dans les études d'observation car les chercheurs n'ont aucun contrôle sur le destinataire du traitement.
- Exemple: 103 084 Parisiens se sont exprimés et 89,03 % d'entre eux ont voté pour la fin des trottinettes en libre-service à Paris. (<https://www.paris.fr/pages/pour-ou-contre-les-trottinettes-en-libre-service-23231#:~:text=103%20084%20Parisiens%20se%20sont,au%20long%20de%20cette%20consultation>)

Données observationnelles

- ③ **biais de sélection (selection bias)** de l'auto-sélection au traitement: Le biais de confusion dû à l'auto-sélection dans le groupe de traitement s'appelle un biais de sélection. Un biais de sélection apparaît souvent dans les études d'observation car les chercheurs n'ont aucun contrôle sur le destinataire du traitement.
- Exemple: 103 084 Parisiens se sont exprimés et 89,03 % d'entre eux ont voté pour la fin des trottinettes en libre-service à Paris. (<https://www.paris.fr/pages/pour-ou-contre-les-trottinettes-en-libre-service-23231#:~:text=103%20084%20Parisiens%20se%20sont,au%20long%20de%20cette%20consultation>)
- Contrôle statistique devient alors nécessaire

Données observationnelles

- Confounding bias
- Self selection bias
- Reverse causation bias
- Unobserved confounding
- Simultaneity

Endogeneity bias

Statistical control

Conditions de la causalité

- ① La variable indépendante X doit “survenir” avant la variable dépendante Y
- ② Les variables X et Y doivent être associées l'une à l'autre
- ③ L'association entre les valeurs X et Y ne doivent pas être due à un troisième facteur, une variable antécédente.

Explications

① X vient avant Y

- Est-ce que fumer cause une maladie coronarienne?
- Mais, est-ce que cette seule condition suffit?
- Non, X peut venir avant Y et il n'y a pas causalité.
Méfiez-vous des cas d'anticipation
- Sortir avec un parapluie vient avant la pluie. Est-ce que le parapluie cause la pluie?
- D'où la vérification de la deuxième condition

Explications

② X et Y sont correlées

- les maladies coronariennes sont plus fréquentes chez les fumeurs
- le jour où il pleut, il y a beaucoup de gens qui sortent leur parapluie
- Les élèves des classes de petite taille réussissent mieux que les élèves des classes de grande taille
- Nous avons appris une méthode particulière pour établir cette association quand les deux variables sont catégorielles
 - Avec deux variables ratio ou intervalle: nuage de point et test de corrélation
 - Avec une VD ratio ou intervalle et une VI catégorielle: Boxplot et analyse de la variance

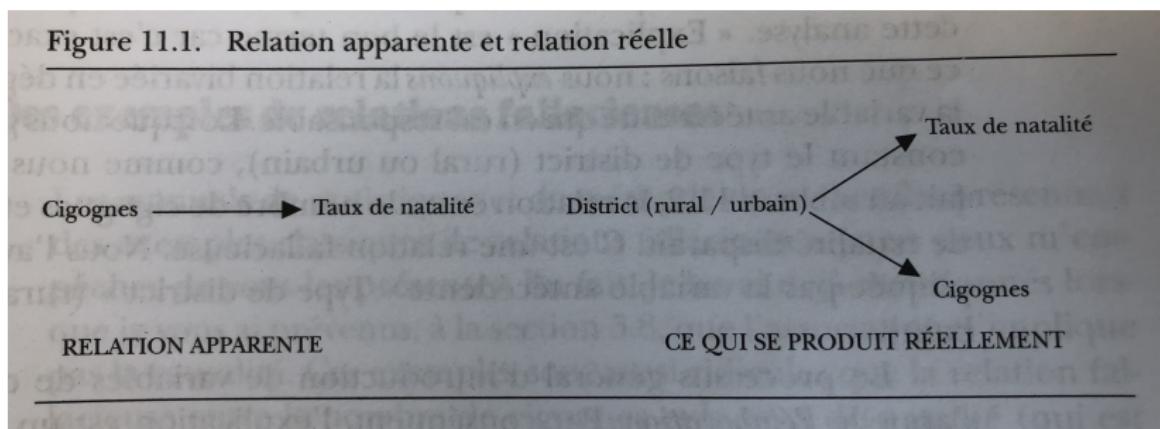
Explications

- ③ Il n'y a pas une explication alternative de la relation entre X et Y
 - Un troisième facteur ne vient pas expliquer la relation entre X et Y: c'est la condition la plus difficile à établir avec les données observationnelles.
 - Une variable antécédente est une variable qui agit avant la variable indépendante dans une chaîne causale.
 - Si c'est une telle variable qui fait que X et Y sont associées, alors X et Y ne sont pas reliées de façon causale

Explications

- ③ Il n'y a pas une explication alternative de la relation entre X et Y

Figure 11.1. Relation apparente et relation réelle



Processus d'élaboration

① Relation initiale

- les deux premières conditions sont réunies

Tableau 11.1. Le taux de natalité selon le
nombre de cigognes
(en pourcentages)

Taux de natalité	Nombre de cigognes	
	Peu	Beaucoup
Élevé	44	62
Bas	56	38
Total	(100)	(100)

Processus d'élaboration

2 Explication alternative: explication

Tableau 11.2. Le taux de natalité selon le nombre de cigognes, contrôlant l'effet du type de district (en pourcentages)

Taux de natalité	District			
	Rural		Urbain	
	Nombre de cigognes		Nombre de cigognes	
Taux de natalité	Peu	Beaucoup	Peu	Beaucoup
Élevé	80	80	20	20
Bas	20	20	80	80
Total	100	100	100	100
(N)	(40)	(70)	(60)	(30)

$\chi^2 = 0,00 ; \text{n.s.}$ $\chi^2 = 0,00 ; \text{n.s.}$
 $G = 0$ $G = 0$

Processus d'élaboration

② Explication alternative

Tableau 11.3. Le nombre de cigognes
selon le district
(en pourcentages)

Nombre de cigognes	District	
	Rural	Urbain
Beaucoup	64	33
Peu	36	67
Total (N)	100 (110)	100 (90)

$\chi^2 = 18,18$; $p < 0,001$; $G = -0,56$

Tableau 11.4. Le taux de natalité
selon le district
(en pourcentages)

Taux de natalité	District	
	Rural	Urbain
Élevé	80	20
Faible	20	80
Total (N)	100 (110)	100 (90)

$\chi^2 = 71,54$; $p < 0,001$; $G = -0,88$

- On voit clairement qu'il y a une association entre le nombre de cigognes et le district

Processus d'élaboration

- ③ Explication alternative : reproduction
- Mais le processus d'élaboration peut conduire à soutenir l'hypothèse de base

Tableau 11.5. Le taux de natalité selon le nombre de cigognes, contrôlant l'effet du type de district (en pourcentages)

Taux de natalité	District			
	Rural		Urbain	
	Nombre de cigognes	Peu	Nombre de cigognes	Peu
Élevé		45	61	43
Bas		55	39	57
Total		100	100	100
(N)		(40)	(70)	(60)

$\chi^2 = 2,78 ; p > 0,05$ $\chi^2 = 3,20 ; p > 0,05$

Processus d'élaboration

Remarque: la reproduction n'est pas une preuve qu'il n'y a pas d'explications alternatives

- Vous n'avez juste pas mis la main sur la bonne variable antécédente
- Sachez que vous ne pouvez pas mettre la main sur l'ensemble des variables antécédentes
 - la plupart vous sont données par la théorie
 - Plusieurs ne sont pas directement mesurables

Processus d'élaboration

④ Explication alternative : spécification

Tableau 11.7. Le taux de natalité selon le nombre de cigognes, contrôlant l'effet du type de district (en pourcentages)

Taux de natalité	District			
	Rural		Urbain	
	Nombre de cigognes		Nombre de cigognes	
	Peu	Beaucoup	Peu	Beaucoup
Élevé	48	71	42	40
Bas	52	29	58	60
Total	100	100	100	100
(N)	(40)	(70)	(60)	(30)

$\chi^2 = 6,23 ; p < 0,05$ $\chi^2 = 0,02 ; \text{n.s.}$
 $G = 0,47$ $G = 0,03$

Variables dissimulatrices

- Les variables antécédentes qui révèlent une association entre deux variables X et Y qui n'existaient pas sans l'introduction de cette variable

Variables intermédiaires

VI (X) —> Variable intermédiaire (Z) —> VD (Y)

3 résultats peuvent arriver:

- ① La relation entre X et Y disparaît: Z relie vraiment X et Y —> processus d'interprétation

Variables intermédiaires

VI (X) —> Variable intermédiaire (Z) —> VD (Y)

3 résultats peuvent arriver:

- ① La relation entre X et Y disparaît: Z relie vraiment X et Y —> processus d'interprétation
- ② La relation entre X et Y demeure pour les catégories de la variable Z: Z n'intervient pas vraiment dans la relation entre X et Y. Z ne fait pas partie de la chaîne causale.

Variables intermédiaires

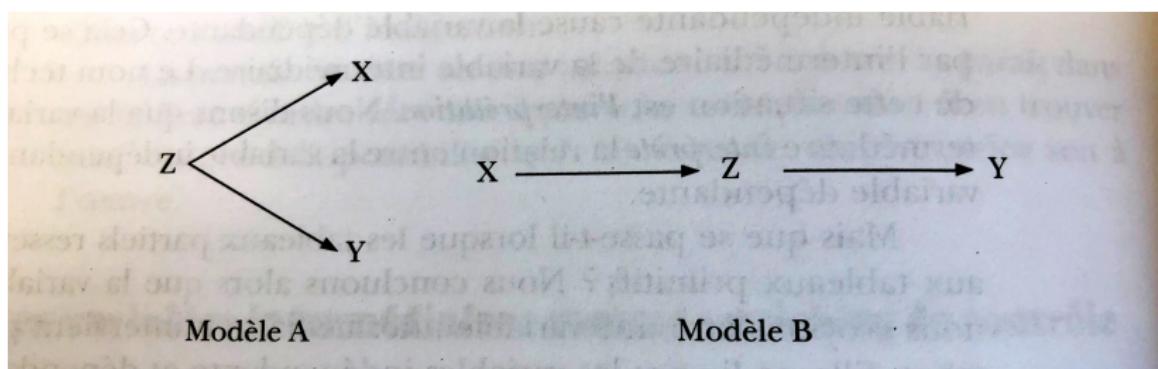
VI (X) —> Variable intermédiaire (Z) —> VD (Y)

3 résultats peuvent arriver:

- ① La relation entre X et Y disparaît: Z relie vraiment X et Y —> processus d'interprétation
- ② La relation entre X et Y demeure pour les catégories de la variable Z: Z n'intervient pas vraiment dans la relation entre X et Y. Z ne fait pas partie de la chaîne causale.
- ③ Z interprète une partie seulement de la relation entre X et Y

Remarque sur les variables antécédentes et intermédiaires

- L'analyse statistique ne peut distinguer entre les deux: seule la théorie et le raisonnement sociologique peuvent vous guider



Calcul de Gamma

- Gamma partiel est l'extension de Gamma aux tableaux multivariés

$$G_p = \frac{\sum \text{Semblables} - \sum \text{Opposées}}{\sum \text{Semblables} + \sum \text{Opposées}}$$

Conclusion

- L'établissement de la causalité entre une variable indépendante X et une variable dépendante Y à partir des données observationnelles ressemble au travail du limier.
- Le sociologue est un “inspecteur” des données pour :
 - lui faire révéler la vérité
 - ou pour valider ces théories

Conclusion

- Où aller à partir de maintenant?
- Relire et refaire les devoirs

Conclusion

- Où aller à partir de maintenant?
- Relire et refaire les devoirs
- Continuer l'apprentissage du logiciel

Conclusion

- Où aller à partir de maintenant?
- Relire et refaire les devoirs
- Continuer l'apprentissage du logiciel
- Prendre un boulot d'été dans un département où vous pouvez analyser les données

Conclusion

- Où aller à partir de maintenant?
- Relire et refaire les devoirs
- Continuer l'apprentissage du logiciel
- Prendre un boulot d'été dans un département où vous pouvez analyser les données
- Prendre un boulot pro-bono avec moi pour travailler sur de vraies données au Labo du département.