

Méthodes quantitatives d'analyse (POL 2809)

Séance 10, 13 novembre 2019

Enseignante: Florence Vallée-Dubois

Bureau: C-3114

Dispos: mercredis, 10h-11h30

florence.vallee-dubois@umontreal.ca

Aujourd'hui

Retour sur le Devoir 4

Suite des GOA (lien avec la régression, simulations dans R)

Les différents types de biais

Rappel: Graphiques orientés acycliques (GOA)

Manière d'illustrer des relations causales (des théories).

Répondre à la question: Est-il possible d'estimer l'effet causal?

Savoir pour quelles variables contrôler ou pas dans une analyse.

Rappel: marche à suivre pour savoir quelles variables inclure (ou pas)

1. NE PAS contrôler pour un descendant de X.
2. Fermer les "chemins par la porte arrière".

Rappel: Régression linéaire (ou OLS)

$$Y = \hat{\alpha} + \hat{\beta} * X + \varepsilon$$

Rappel: Régression linéaire multiple

$$Y = \hat{\alpha} + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \varepsilon$$

Rappel: Régression linéaire multiple

$$Y = \hat{\alpha} + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \varepsilon$$

Ou, pour "k" variables indépendantes:

$$Y = \hat{\alpha} + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_k X_k + \varepsilon$$

"Contrôler"

Quand il s'agit d'un modèle de régression multiple (avec plus d'un X),...

Le modèle utilise la partie de X_1 qui n'est pas expliquée par X_2 pour prédire Y .

"Contrôler"

Quand il s'agit d'une modèle de régression multiple (avec plus d'un X),...

Le modèle utilise la partie de X_1 qui n'est pas expliquée par X_2 pour prédire Y.

Rappelez-vous les petits cercles! (page 83 du manuel)

"Contrôler"

L'interprétation de $\hat{\beta}_1$ devient donc:
"L'effet de X_1 sur Y après avoir contrôlé pour X_2 ".

"Contrôler"

L'interprétation de $\hat{\beta}_1$ devient donc:
"L'effet de X_1 sur Y après avoir contrôlé pour X_2 ".

Ou: "L'effet de X_1 sur Y en tenant X_2 constant".

"Contrôler"

L'interprétation de $\hat{\beta}_1$ devient donc:
"L'effet de X_1 sur Y après avoir contrôlé pour X_2 ".

Ou: "L'effet de X_1 sur Y en tenant X_2 constant".

Quand on interprète l'effet d'un X , on contrôle (on tient constantes) les autres variables.

Le lien avec les GOA

Les GOA nous indiquent quelles variables inclure (ou pas) dans l'analyse.

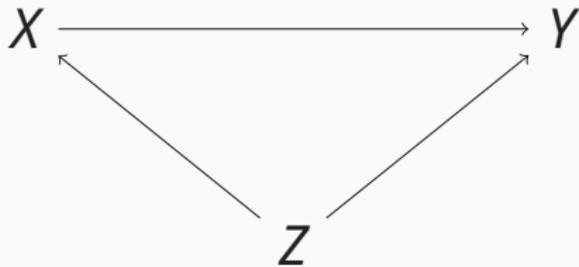
Le lien avec les GOA

Les GOA nous indiquent quelles variables inclure (ou pas) dans l'analyse.

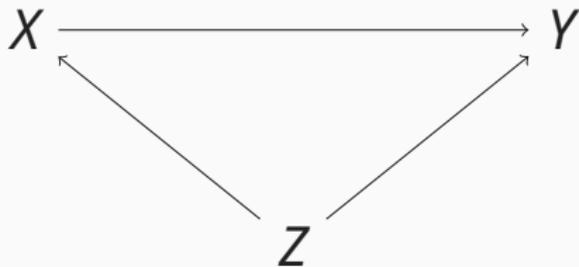
En d'autres mots, ils nous indiquent quelles variables inclure dans notre modèle de régression!

En plus de notre variable explicative principale.

Exemple

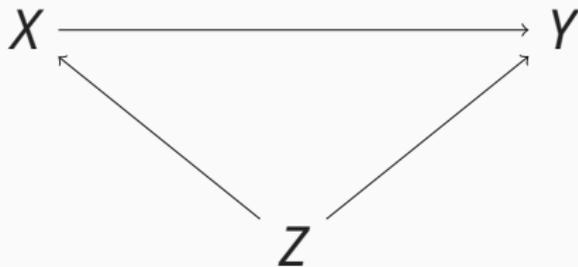


Exemple



Le bon modèle est: $Y = \hat{\alpha} + \hat{\beta}_1 * X + \hat{\beta}_2 * Z + \varepsilon$

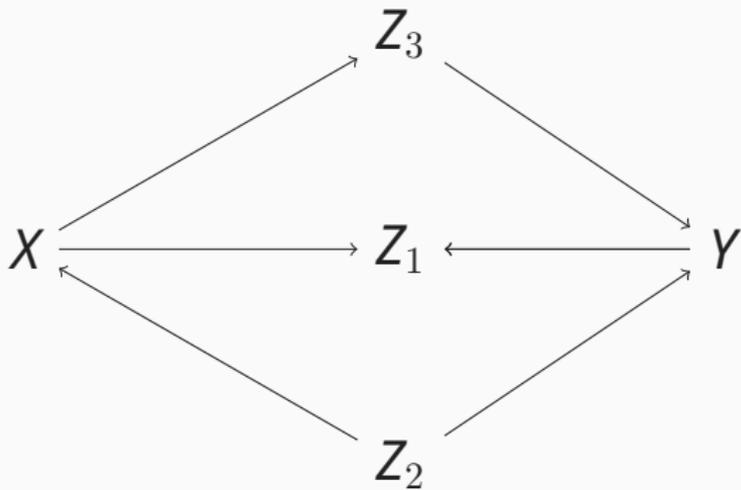
Exemple



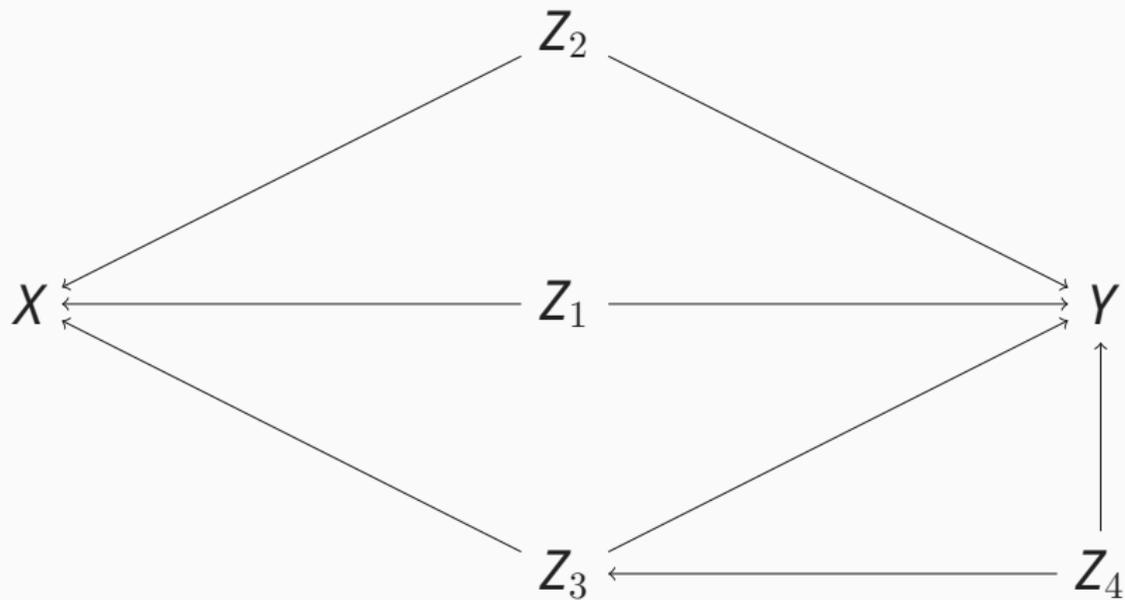
Le bon modèle est: $Y = \hat{\alpha} + \hat{\beta}_1 * X + \hat{\beta}_2 * Z + \varepsilon$

On pourra ensuite interpréter l'effet de **X** sur **Y**, en contrôlant pour **Z**.

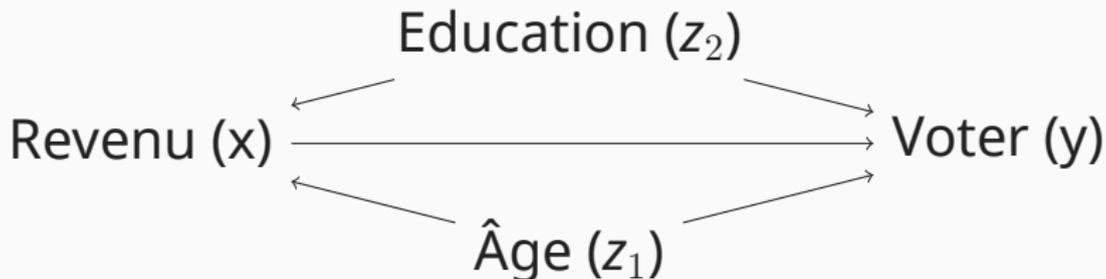
Exercise 1



Exercise 2



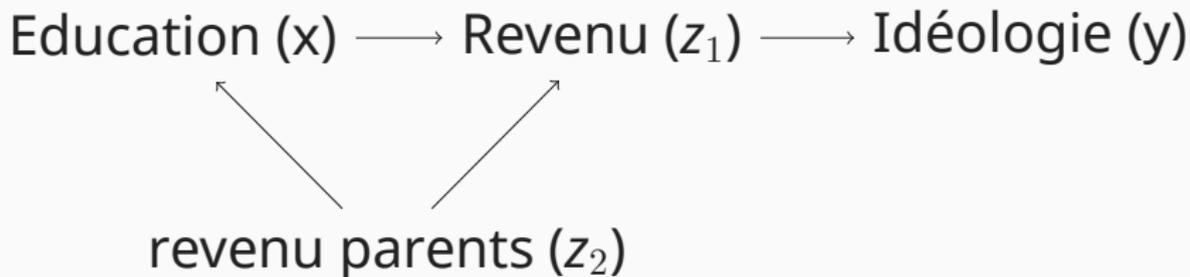
Exemples avec des "vraies" relations



Pour quelles variables faut-il contrôler?

Traduisez ce GOA en équation de régression linéaire complète.

Exemples avec des "vraies" relations



Pour quelles variables faut-il contrôler?

Traduisez ce GOA en équation de régression linéaire complète.

Pour illustrer le lien entre GOA et régression, faisons quelques simulations dans R (code disponible sur Studium et R Studio Cloud).

Interprétation des variables de contrôle

Dans un modèle de régression multiple, il faut généralement **éviter** d'interpréter les coefficients associés aux variables de contrôle comme des effets causaux.

Exemple.



Attention!

Attention!

L'analyse graphique (les GOA) nous permettent d'encoder notre théorie.

Attention!

L'analyse graphique (les GOA) nous permettent d'encoder notre théorie.

Ils nous aident à déterminer si l'effet causal est identifiable.

Attention!

L'analyse graphique (les GOA) nous permettent d'encoder notre théorie.

Ils nous aident à déterminer si l'effet causal est identifiable.

Mais même si on pense avoir contrôlé pour les bonnes variables dans une régression, on peut toujours en avoir oublié d'autres.

Attention!

La régression ne remplace pas les expériences aléatoires!

Attention!

La régression ne remplace pas les expériences aléatoires!

Les expériences aléatoires demeurent le "gold standard" de la recherche de la causalité.

Questions?

Questions?

C'est la pause!



Jusqu'ici

Jusqu'ici

On a vu comment bien estimer l'effet de X sur Y en se basant sur un GOA.

Jusqu'ici

On a vu comment bien estimer l'effet de X sur Y en se basant sur un GOA.

Maintenant: on va identifier les différents biais qui peuvent affecter les relations qui nous intéressent.

Types de biais

Biais par variable omise

Biais de sélection

Biais de mesure

Biais de simultanéité

Biais par variable omise

Se produit quand la relation entre X et Y est en partie (ou totalement) expliquée par une troisième variable, qui cause X et Y en même temps.

Exemples au tableau

Connaître le sens du biais

		Relation entre A et Y	
		+	-
Relation entre A et X	+	Biais positif	Biais négatif
	-	Biais négatif	Biais positif

A = variable omise

Connaître le sens du biais

		Relation entre A et Y	
		+	-
Relation entre A et X	+	Biais positif	Biais négatif
	-	Biais négatif	Biais positif

A = variable omise

Il s'agit d'une règle approximative.

Solutions

Expériences aléatoires: le traitement (X) est distribué aléatoirement, donc pas de risque de variable omise.

Contrôler pour les bonnes variables dans une régression: méthode imparfaite, mais parfois la seule possible à réaliser

Méthodes quasi-expérimentales, variables instrumentales, etc.

Biais de sélection

Deux types:

Biais de sélection

Deux types:

Sélection dans l'analyse

Sélection dans le traitement

Biais de sélection

Deux types:

Sélection dans l'analyse

Sélection dans le traitement

On va se concentrer sur le premier type.

Biais de sélection dans l'analyse

Quand on n'inclut pas tous les cas.

Biais de sélection dans l'analyse

Quand on n'inclut pas tous les cas.

Imaginez une expérience où on donne un médicament pour réduire les symptômes d'une maladie.

Biais de sélection dans l'analyse

Quand on n'inclut pas tous les cas.

Imaginez une expérience où on donne un médicament pour réduire les symptômes d'une maladie.

Or, le médicament a des effets secondaires, ce qui pousse certains à se retirer. Si cela se produit, l'effet estimé sera biaisé.

Sélection dans l'analyse: sélection sur la variable dépendante

Certaines valeurs de la variable dépendante ont plus de chances de se retrouver dans l'analyse.

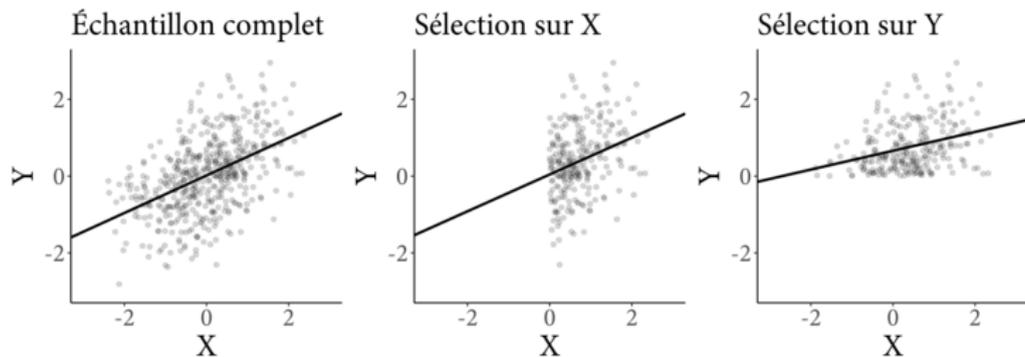
Sélection dans l'analyse: sélection sur la variable dépendante

Certaines valeurs de la variable dépendante ont plus de chances de se retrouver dans l'analyse.

Exemples: Silicon Valley; la relation entre choc économique et guerre civile; éviter de répondre à un sondage.

Sélection dans l'analyse: sélection sur la variable dépendante

Trois extraits de la même banque de données. Les lignes représentent le résultat d'une régression bivariée avec Y comme variable dépendante et X comme variable indépendante. Lorsque les observations sont sélectionnées en fonction de la variable dépendante, les résultats sont biaisés.



Sélection dans l'analyse: sélection sur une collision

Si on sélectionne sur une variable au centre d'une collision, on peut introduire un biais dans notre analyse.

Sélection dans l'analyse: sélection sur une collision

Si on sélectionne sur une variable au centre d'une collision, on peut introduire un biais dans notre analyse.

Exemple: les tests d'entrée dans les programmes d'études supérieures.



Solutions

Solutions

Réfléchir sérieusement à la sélection des cas

Solutions

Réfléchir sérieusement à la sélection des cas

Tracer un GOA pour s'aider.

Solutions

Réfléchir sérieusement à la sélection des cas

Tracer un GOA pour s'aider.

Dire explicitement si on pense qu'il y a un biais de sélection (ou pas)

Solutions

Réfléchir sérieusement à la sélection des cas

Tracer un GOA pour s'aider.

Dire explicitement si on pense qu'il y a un biais de sélection (ou pas)

Être transparent et honnête si on ne peut pas conclure à un effet causal

Questions?

Questions?

C'est la pause!

Biais de mesure

Un biais omniprésent en sciences sociales.

Biais de mesure

Un biais omniprésent en sciences sociales.

Souvent dû à la difficulté d'obtenir des données qui mesurent vraiment ce qu'on voudrait mesurer.



Qu'est-ce qu'une bonne mesure?

Qu'est-ce qu'une bonne mesure?

Une mesure qui est valide et fiable.

Qu'est-ce qu'une bonne mesure?

Une mesure qui est valide et fiable.

Validité: Adéquation entre un concept et la mesure employée pour l'opérationnaliser.

Qu'est-ce qu'une bonne mesure?

Une mesure qui est valide et fiable.

Validité: Adéquation entre un concept et la mesure employée pour l'opérationnaliser.

Fiabilité: Capacité de mesurer le concept sans faire d'erreurs accidentelles ou aléatoires.

Erreurs de mesure dépendantes

Quand la mesure choisie pour opérationnaliser X et de Y peut être simultanément affectée par un biais.

Exemple: Ne pas répondre la vérité dans un sondage, par peur d'être jugé.

Erreurs de mesure différentielles

Quand une mesure de Y peut être affectée directement par X , ou vice-versa.

Exemple: prise de drogues et comportements violents (mesurés dans un sondage)



À se rappeler

À se rappeler

Les mesures peuvent introduire des biais.

À se rappeler

Les mesures peuvent introduire des biais.

Il est donc important de réfléchir aux meilleures mesures possible pour nos concepts.

À se rappeler

Les mesures peuvent introduire des biais.

Il est donc important de réfléchir aux meilleures mesures possible pour nos concepts.

Dans votre travail d'analyse, vous n'avez pas le choix de travailler avec les mesures fournies: les données du CSES ou de Quality of Governance.

À se rappeler

Les mesures peuvent introduire des biais.

Il est donc important de réfléchir aux meilleures mesures possible pour nos concepts.

Dans votre travail d'analyse, vous n'avez pas le choix de travailler avec les mesures fournies: les données du CSES ou de Quality of Governance.

Mais vous pouvez quand même discuter les limites associées à ces mesures!

Prochain cours

Suite des biais.

Dernier devoir

A close-up photograph of a giant panda sitting on a tree branch. The panda's mouth is wide open, showing its pink tongue and teeth. Its black and white fur is clearly visible. The panda's left paw is raised, showing its black fur and light-colored pads. The background is a soft, out-of-focus green, suggesting a forest setting.

**À la semaine
prochaine!**