

An aerial photograph showing a wide, light-colored landscape, likely sand or dry mud, with a prominent blue-grey winding channel or riverbed cutting through it. Several small figures of people are scattered across the landscape, appearing as dark silhouettes against the lighter terrain.

# Biais de sélection

Vincent Arel-Bundock

# Biais de sélection

Concept ambigu. En sciences sociales, deux problèmes distincts:

## 1. Sélection dans l'analyse

- Certains individus ou certaines variables ne sont pas observés par l'analyste.
- L'échantillon étudié n'est pas représentatif de la population.

Exemples:

- Les partisans de partis politiques d'extrême droite refuse de répondre à des sondages téléphoniques.
- Les patients qui souffrent d'effets secondaires refusent de participer à une expérience.

# Biais de sélection

Concept ambigu. En sciences sociales, deux problèmes distincts:

## 1. Sélection dans l'analyse

- Certains individus ou certaines variables ne sont pas observés par l'analyste.
- L'échantillon étudié n'est pas représentatif de la population.

Exemples:

- Les partisans de partis politiques d'extrême droite refusent de répondre à des sondages téléphoniques.
- Les patients qui souffrent d'effets secondaires refusent de participer à une expérience.

## 2. Sélection dans le traitement

- Le processus qui détermine qui reçoit le traitement (ou la valeur du traitement) est associé à la variable dépendante.
- Très similaire au biais par variable omise.

Exemples:

- Les étudiants de baccalauréat les plus forts poursuivent à la maîtrise.
- Les gouvernements "gentils" signent des traités pour la protection des droits humains.

# Plan

- Sélection dans l'analyse
  - Analyse graphique
  - Sélection sur la variable dépendante
  - Sélection sur une collision
  - Sélection sur le descendant d'collision
- Sélection dans le traitement
  - Analyse algébrique
  - Exemples



A dark hallway featuring a row of doors. All doors are dark grey except for one, which is a vibrant red. The doors are set against a dark wall and sit on a dark floor.

# Analyse graphique Sélection dans l'analyse

# Sélection dans l'analyse

Le biais de sélection dans l'analyse peut survenir pour plusieurs raisons:

- Le choix du chercheur d'étudier certaines régions, organisations, ou phénomènes plutôt que d'autres.
- Le choix des individus de participer ou non à une expérience.
- La difficulté d'observer ou de mesurer certains objets d'étude.

Analyse à l'aide de graphes orientés acycliques.

# Sélection dans l'analyse

Objectif:

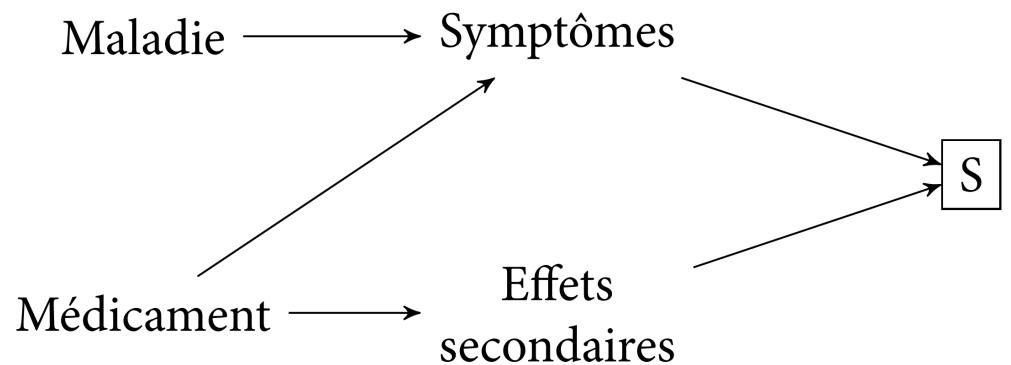
- Estimer l'effet d'un nouveau médicament sur les symptômes de la maladie de Lyme.

Méthodologie:

- Expérience aléatoire: 50% placebo, 50% médicament.

Biais de sélection:

- Lorsqu'ils souffrent d'effets secondaires, certains participants se retirent de l'étude.
- Ceux-ci sont exclus de la banque de données et de l'analyse.



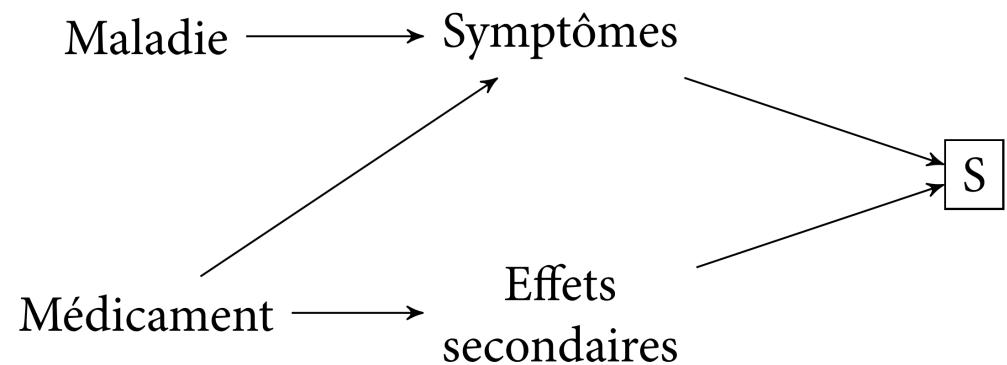
# Sélection dans l'analyse

Variable  $\boxed{S}$ :

- Résultat du processus de sélection.
- Les facteurs qui déterminent si un individu est inclus dans l'analyse causent  $\boxed{S}$ .

Problème:

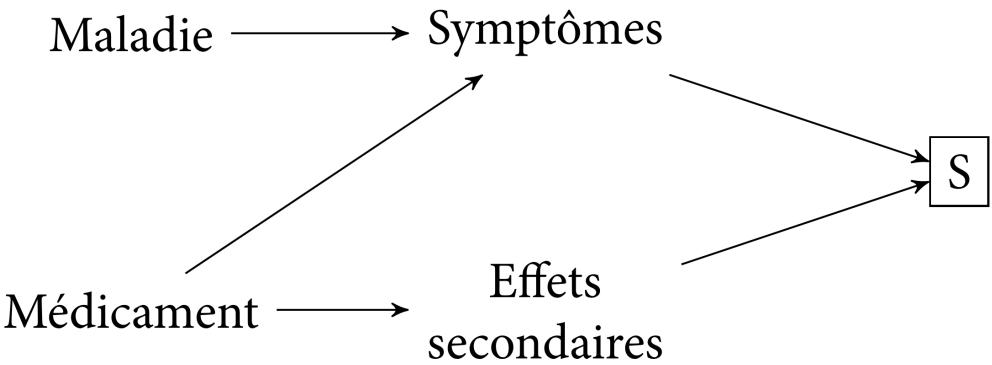
- Exclure des observations de l'échantillon équivaut à "contrôler"  $\boxed{S}$ .
- Ceci peut ouvrir un chemin non-causal.



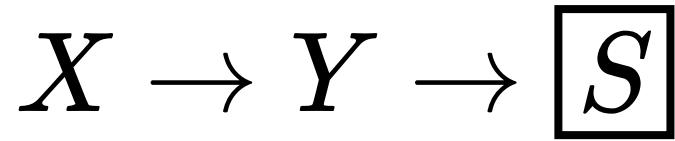
# Sélection dans l'analyse

Problèmes:

- $S$  est un descendant de la cause qui nous intéresse. Ce processus de sélection viole donc la première condition de l'identification causale.
- Contrôler  $S$  ouvre un chemin non causal entre la variable indépendante et la variable dépendante.



# Sélection sur la variable dépendante

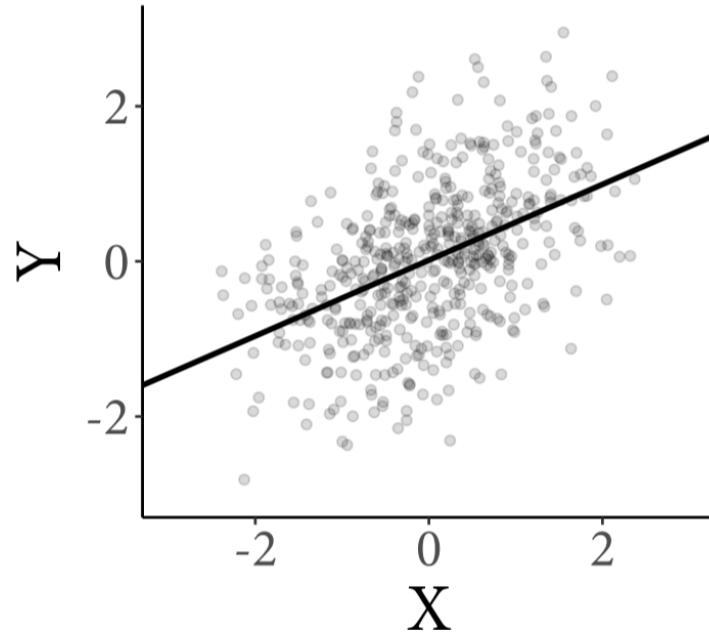


Problèmes:

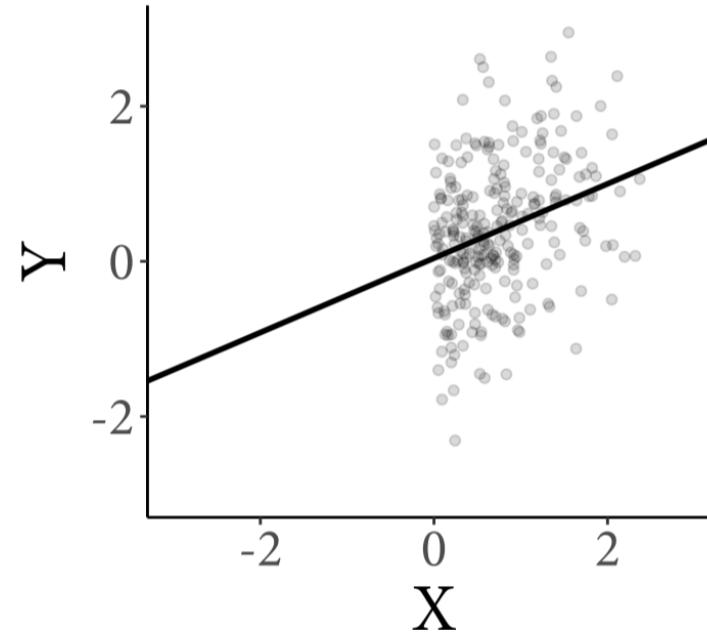
- Contrôler un descendant viole la première condition de l'identification causale.
- Choisir l'échantillon en fonction des valeurs de la variable dépendante introduit un biais de sélection.

# Sélection sur la variable dépendante

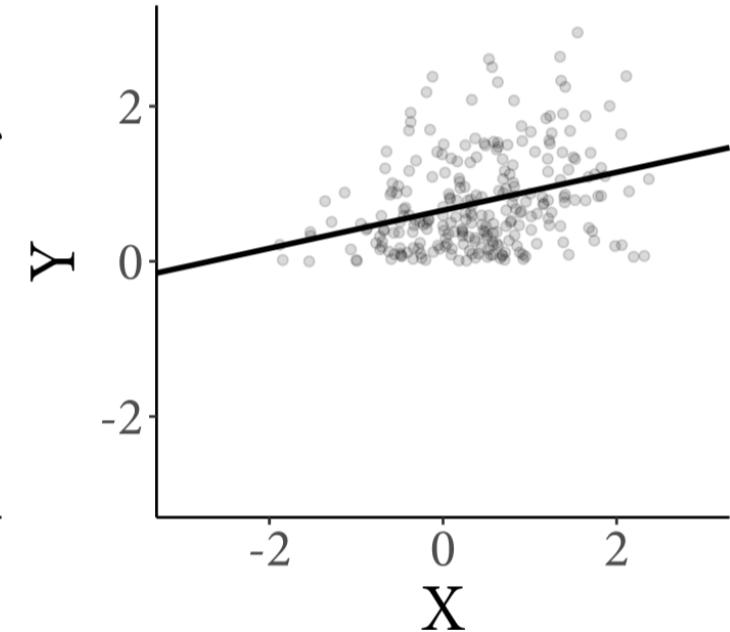
Échantillon complet



Sélection sur X



Sélection sur Y



# Sélection sur la variable dépendante

## American dream

Google, Apple, Microsoft, Amazon, Disney et Hewlet-Packard ont toutes été dirigées à partir de garages résidentiels.

## Hypothèse

- L'accès au capital n'a pas d'effet sur le succès d'une jeune entreprise.



# Sélection sur la variable dépendante

Problèmes:

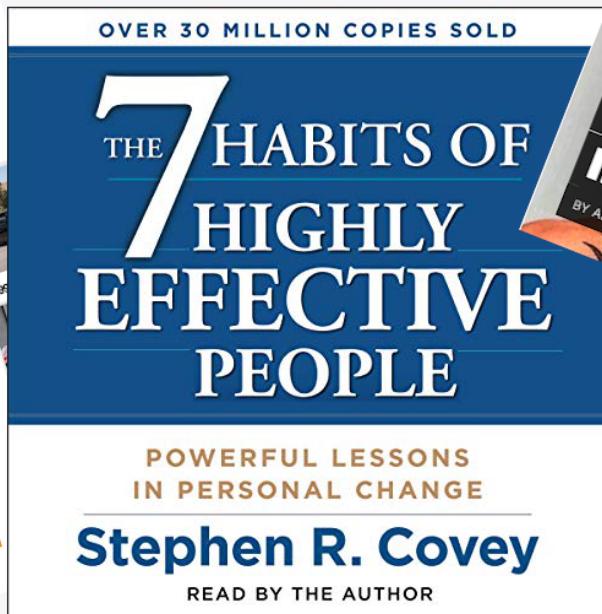
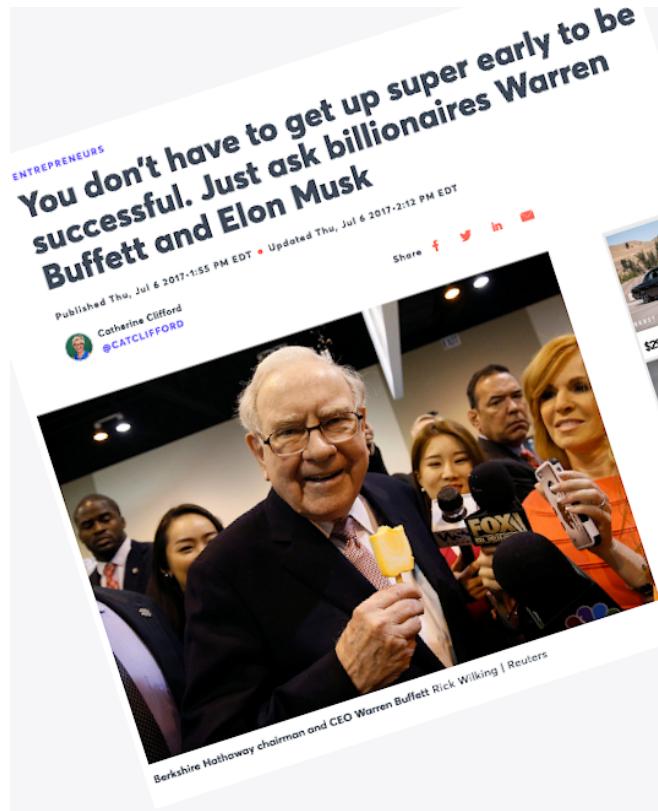
- "Contrôler" un descendant de la cause qui nous intéresse
- Considérer seulement les entreprises à succès
- Ignorer toutes les start-up qui ont échoué dans un garage



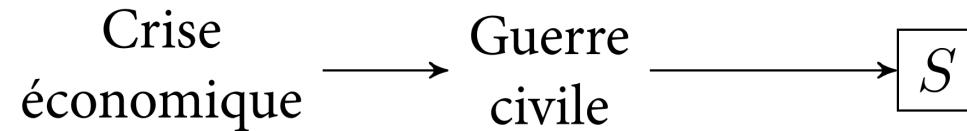
Biais de sélection!

# Sélection sur la variable dépendante

C'est partout!



# Sélection sur la variable dépendante



Étudier seulement des pays (ou des périodes historiques) frappés par la guerre civile néglige tous les cas de pays qui subissent une crise économique sans pour autant tomber en guerre civile.

# Sélection sur une collision

## Biais de publication

Les revues scientifiques prestigieuses comme *Nature* ou *Science* publient une petite proportion des articles qui leur sont soumis.

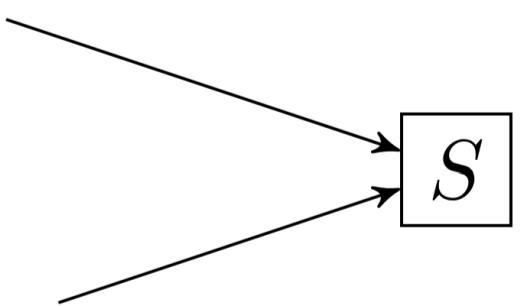
Un article est plus susceptible d'être choisi pour publication si les résultats sont intéressants ou et si la méthodologie est rigoureuse.

Aucune flèche entre intérêt et rigueur. Aucune relation causale entre intérêt et rigueur.

Résultat intéressant

Méthodologie rigoureuse

$S$



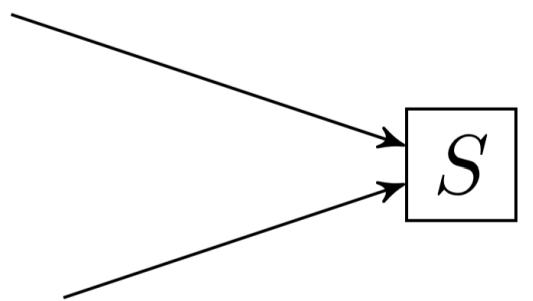
# Sélection sur une collision

Aucune relation causale entre intérêt et rigueur,  
mais

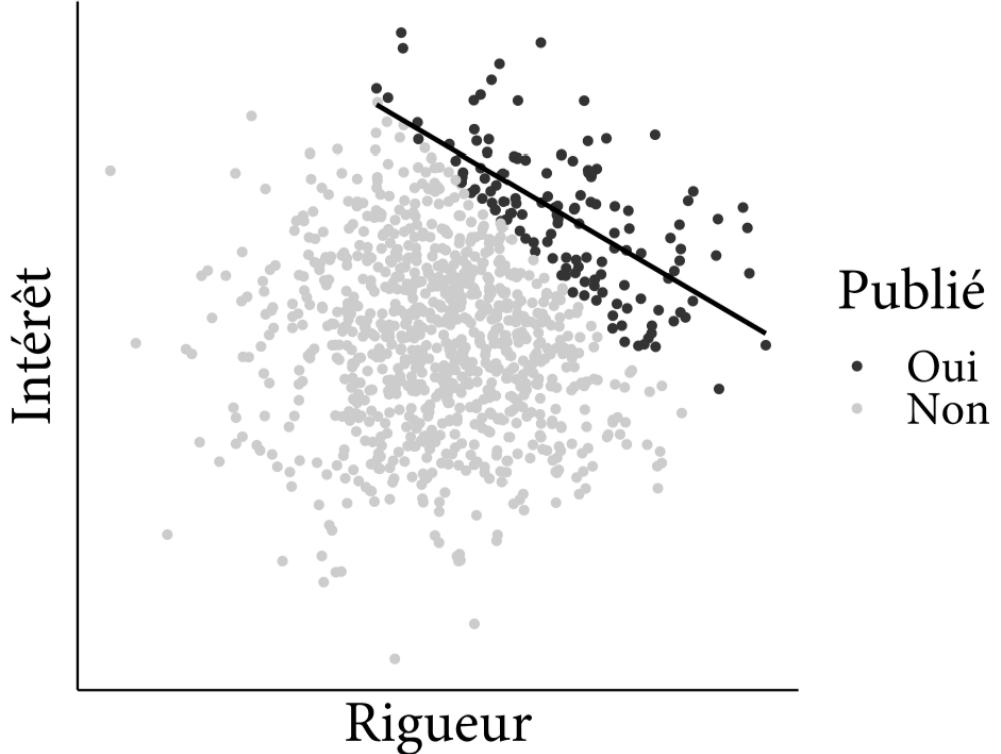
- Contrôler  $\boxed{S}$  ouvre un chemin factice
  - Résultat  $\rightarrow \boxed{S} \leftarrow$  Méthodologie
- Biais de sélection

Résultat  
intéressant

Méthodologie  
rigoureuse



# Sélection sur une collision



Résultat  
intéressant

Méthodologie  
rigoureuse

$S$

# Sélection sur une collision

## Tests d'aptitudes

L'admission aux études en droit ou doctorat est souvent déterminée par les résultats au *Law School Admissions Test* ou *Graduate Record Examination*.

Certains chercheurs ont remis en cause cette pratique : les résultats aux tests standardisés ne sont pas associés au succès (p.ex. taux de graduation ou nombre de publications dans des revues scientifiques).

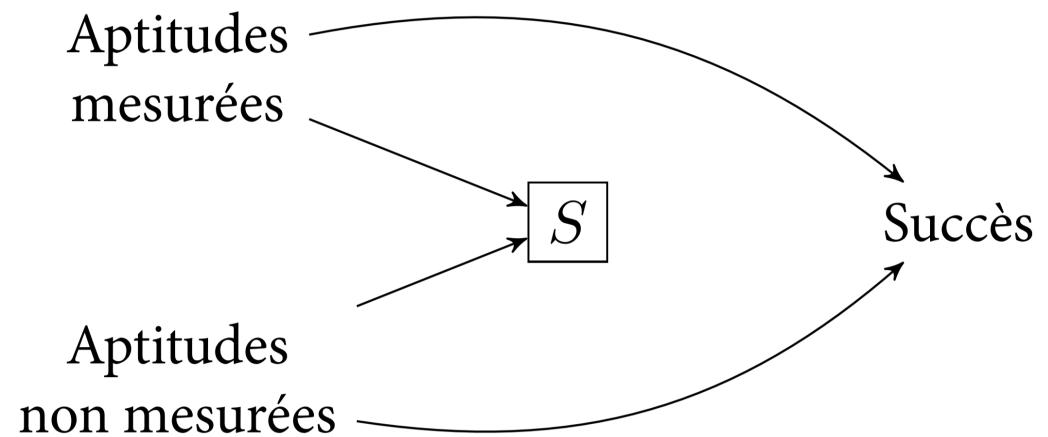
Devons-nous conclure que les compétences mesurées par les tests standardisés n'ont aucun effet causal sur la performance des étudiants? Pas nécessairement.

# Tests d'aptitudes

On distingue deux types d'aptitudes:

1. Les aptitudes mesurées par le test standardisé.
2. Les aptitudes qui ne sont pas mesurées par ce test, mais qui influencent néanmoins l'admission.

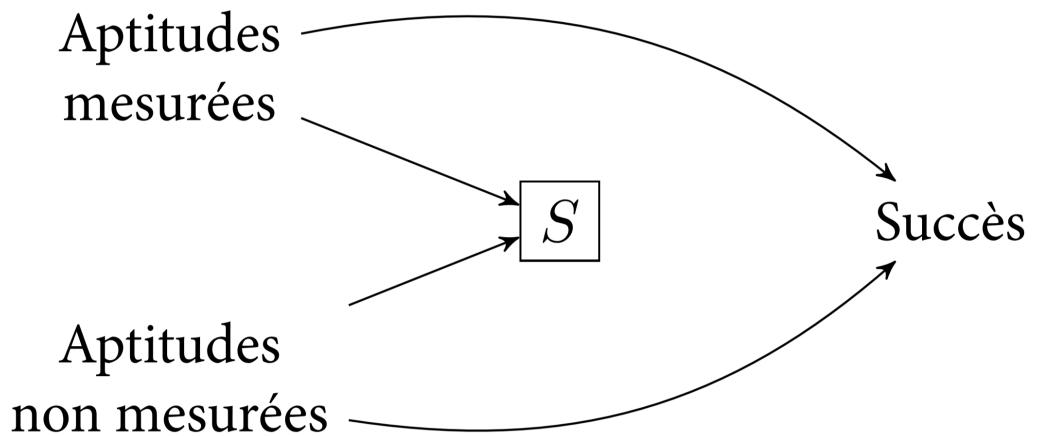
**S** représente l'admission à un programme de doctorat.



# Tests d'aptitudes

Quel est l'effet causal des compétences mesurées par les tests standardisés sur le succès dans les études doctorales?

Si l'échantillon se limite aux étudiants qui ont été admis aux études doctorales, on contrôle implicitement pour la variable  $S$ , qui est descendante de la cause.

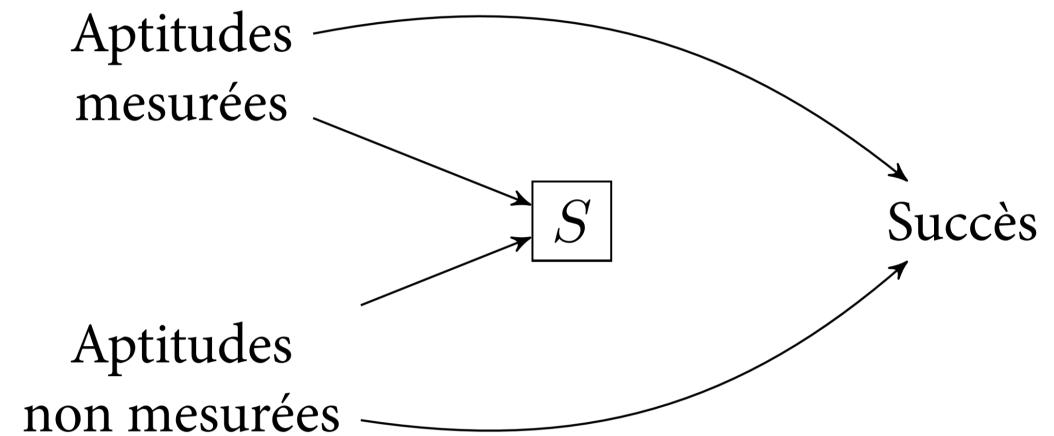


# Tests d'aptitudes

Les candidats qui sont admis malgré leur faible performance au test standardisé ont probablement de très fortes compétences intangibles.

Les compétences intangibles leur permettent de bien réussir, malgré leurs déficiences dans d'autres domaines.

Dans ce cas, il serait incorrect de conclure que les compétences mesurées par le test sont inutiles.



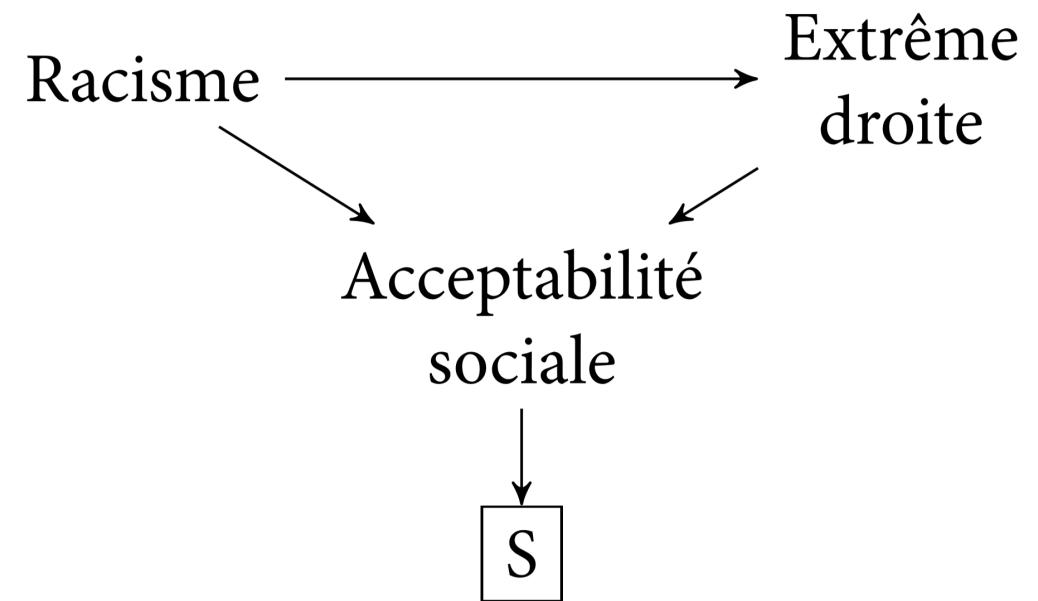
# Sélection sur le descendant d'une collision

Contrôler pour le descendant d'une collision...

- Effet causal du racisme sur l'appui aux partis politiques d'extrême droite.
- Sondage téléphonique et échelles psychométriques.

Problème:

- Partisans d'extrême droite et les racistes savent que l'acceptabilité sociale pour leurs opinions est faible.
- Ceci les rend moins susceptibles de répondre au sondage téléphonique (**S**).
- Biais de sélection



A dark hallway featuring a row of doors. All doors are dark grey except for one, which is a vibrant red. The red door is slightly ajar, emitting a bright red glow that washes over the floor in front of it. The floor is a light-colored polished stone.

# Analyse algébrique Sélection dans le traitement

# Analyse algébrique: Sélection dans le traitement

Définition:

La probabilité de recevoir le traitement et/ou la valeur de la variable indépendante sont associées aux résultats potentiels.

Analogue au biais par variable omise:

$$X \rightarrow Y$$
$$\uparrow$$
$$Z$$

# Sélection dans le traitement

Rappel: Problème fondamental de l'inférence causal

L'effet de traitement individuel est *toujours* impossible à estimer:

$$\kappa = Y_{i1} - Y_{i0}$$

Postulats:

1. Indépendance:  $X_i \perp Y_{i0}, Y_{i1}$
2. SUTVA
3. Positivité

Effet de traitement moyen:

$$E[Y_{i1} - Y_{i0}] = E[Y_{i1}|X_i = 1] - E[Y_{i0}|X_i = 0]$$

# Sélection dans le traitement

Rappel: Problème fondamental de l'inférence causal

L'effet de traitement individuel est *toujours* impossible à estimer:

$$\kappa = Y_{i1} - Y_{i0}$$

Postulats:

1. Indépendance:  $X_i \perp Y_{i0}, Y_{i1}$
2. SUTVA
3. Positivité

Effet de traitement moyen:

$$E[Y_{i1} - Y_{i0}] = E[Y_{i1}|X_i = 1] - E[Y_{i0}|X_i = 0]$$

Réarranger et substituer donne:

$$Y_{i1} = \kappa + Y_{i0}$$

$$E[\kappa + Y_{i0}|X_i = 1] - E[Y_{i0}|X_i = 0]$$

Règles de l'espérance:

$$\kappa + E[Y_{i0}|X_i = 1] - E[Y_{i0}|X_i = 0]$$

# Sélection dans le traitement

$$\widehat{\kappa} = \underbrace{E[Y_{i0}|X_i = 1] - E[Y_{i0}|X_i = 0]}_{\text{Non observable}} + \underbrace{E[Y_{i0}|X_i = 1] - E[Y_{i0}|X_i = 0]}_{\text{Observable}}$$

Biais de sélection

# Direction du biais

$$\overbrace{\kappa}^{\text{Effet causal}} + \underbrace{E[Y_{i0}|X_i = 1] - E[Y_{i0}|X_i = 0]}_{\begin{array}{l} \text{Non observable} \\ \text{Observable} \end{array}} \quad \text{Biais de sélection}$$

Expérience de pensée:

Quel aurait été le résultat dans le groupe de traitement si les membres n'avaient *pas* reçu le traitement?

# Direction du biais

$$\underbrace{\kappa}_{\text{Effet causal}} + \underbrace{E[Y_{i0}|X_i = 1] - E[Y_{i0}|X_i = 0]}_{\begin{array}{l} \text{Biais de sélection} \\ \text{Non observable} \quad \quad \quad \text{Observable} \end{array}}$$

Expérience de pensée:

Quel aurait été le résultat dans le groupe de traitement si les membres n'avaient pas reçu le traitement?

$$E[Y_{i0}|X_i = 1] = E[Y_{i0}|X_i = 0]$$

Aucun biais

$$E[Y_{i0}|X_i = 1] > E[Y_{i0}|X_i = 0]$$

La différence des moyennes surestime l'effet causal.

$$E[Y_{i0}|X_i = 1] < E[Y_{i0}|X_i = 0]$$

La différence des moyennes sous-estime l'effet causal.

# Est-ce que les études X augmentent le salaire Y?

- Traitement:
  - $X_i = 1$  si études universitaires
- Résultat:
  - $Y$  revenu

Question hypothétique:

Quels seraient les revenus du groupe de traitement si ils n'avaient pas été à l'université?

$$\widehat{\kappa} \quad + \underbrace{E[Y_{i0}|X_i = 1]}_{\text{Non observable}} - \underbrace{E[Y_{i0}|X_i = 0]}_{\text{Observable}}$$

Biais de sélection

# Est-ce que les études X augmentent le salaire Y?

- Traitement:
  - $X_i = 1$  si études universitaires
- Résultat:
  - $Y$  revenu

Question hypothétique:

Quels seraient les revenus du groupe de traitement si ils n'avaient pas été à l'université?

Si les compétences qui rendent les études universitaires sont prisées sur le marché de l'emploi, alors:

$$E[Y_{i0}|X_i = 1] - E[Y_{i0}|X_i = 0] > 0$$

Une comparaison du revenu moyen des universitaires et des non universitaires *surestime* l'effet causal des études sur le revenu.

$$\widehat{\kappa} \quad + \underbrace{E[Y_{i0}|X_i = 1]}_{\text{Non observable}} - \underbrace{E[Y_{i0}|X_i = 0]}_{\text{Observable}}$$

Biais de sélection

# Est-ce signer un traité des droits de la personne améliore le comportement du gouvernement?

- Traitement:
  - $X_i = 1$  si traité signé
- Résultat:
  - $Y$  qualité du traitement des population

Question hypothétique:

Quel serait le comportement du gouvernement dans le groupe de signataire n'avaient pas signé le traité?

$$\widehat{\kappa} \quad \text{Effet causal} + \underbrace{E[Y_{i0}|X_i = 1]}_{\text{Non observable}} - \underbrace{E[Y_{i0}|X_i = 0]}_{\text{Observable}}$$

Biais de sélection

# Est-ce signer un traité des droits de la personne améliore le comportement du gouvernement?

- Traitement:
  - $X_i = 1$  si traité signé
- Résultat:
  - $Y$  qualité du traitement des population

Question hypothétique:

Quel serait le comportement du gouvernement dans le groupe de signataire n'avaient pas signé le traité?

Si les gouvernements "gentils" signent plus facilement les traités:

$$E[Y_{i0}|X_i = 1] - E[Y_{i0}|X_i = 0] > 0$$

Une comparaison des droits de la personne dans les pays signataires et non-signataires *surestime* l'effet des traités.

Effet causal

$$\widehat{\kappa} + \underbrace{E[Y_{i0}|X_i = 1]}_{\text{Non observable}} - \underbrace{E[Y_{i0}|X_i = 0]}_{\text{Observable}}$$

Biais de sélection

# Merci!

