

Séance 8.1: Regression linéaire simple et multiple

Visseho Adjiwanou, PhD.

SICSS - Montréal

30 October 2022

Plan de présentation

- Exemples graphiques
 - Exemple 1: accès aux médias et attitude face à la violence
 - Exemple 2: Relation entre taille et poids
- Discuter de la meilleure manière d'estimer cette relation
- Présenter le modèle linéaire simple avec ces interprétations

Introduction

- lorsque la voie de l'assignation aléatoire est bloquée, nous cherchons d'autres voies vers la connaissance causale.
- Manipulés avec habileté, les outils métriques autres que l'assignation aléatoire peuvent avoir une grande partie du pouvoir de **révélation de la causalité** d'une véritable expérience.
- Le plus élémentaire de ces outils est la **régression**, qui permet de comparer les sujets traités et les sujets témoins qui présentent les **mêmes caractéristiques observées**.
- Les concepts de régression sont fondamentaux et ouvrent la voie aux outils plus élaborés.
- L'inférence causale basée sur la régression repose sur l'hypothèse selon laquelle, **lorsque les variables clés observées ont été rendues égales entre les groupes de traitement et de contrôle, le biais de sélection provenant de ce que nous ne pouvons pas voir est également**

Exemple : A tale of two colleges (L'histoire de deux universités)

Les étudiants qui ont fréquenté un collège privé de quatre ans en Amérique ont payé en moyenne environ 29 000 dollars de frais de scolarité et de frais d'inscription au cours de l'année scolaire 2012-2013. Ceux qui ont fréquenté une université publique dans leur État d'origine ont payé moins de 9 000 dollars. Un enseignement privé d'élite peut être meilleur à bien des égards : les classes sont plus petites, les installations sportives plus récentes, le corps enseignant plus distingué et les étudiants plus intelligents. Mais 20 000 dollars par année d'étude, c'est une grande différence. On peut se demander si la différence en vaut la peine.

Dans ce cas, la question est de savoir combien un diplômé de 40 ans, né dans le Massachusetts et ayant étudié à Harvard, aurait gagné s'il était allé à l'Université du Massachusetts (U-Mass).

L'argent n'est pas tout, mais, comme l'a observé Groucho Marx :

Matchmaker, mathmaker

Hélas, les revenus ne se limitent pas au sexe, aux écoles et aux scores SAT. Étant donné que les décisions de fréquentation des collèges ne sont pas attribuées au hasard, nous devons contrôler tous les facteurs qui déterminent à la fois les décisions de fréquentation et les revenus ultérieurs. Ces facteurs incluent les caractéristiques des étudiants, comme la capacité d'écriture, la diligence, les liens familiaux, etc. Le contrôle d'un tel éventail de facteurs semble intimidant : les possibilités sont pratiquement infinies et de nombreuses caractéristiques sont difficiles à quantifier. Mais Stacy Berg Dale et Alan Krueger ont trouvé un raccourci intelligent et convaincant. Au lieu d'identifier tout ce qui pourrait avoir de l'importance pour le choix du collège et les revenus, ils travaillent avec une mesure récapitulative clé : les caractéristiques des collèges auxquels les étudiants ont postulé et ont été admis.

Considérez à nouveau l'histoire d'Uma et Harvey: tous deux ont

table

Group	Etud.	Private			Publique			1
		Ivy	Leafy	Smart	A_state	T_state	Alt_state	
A	1							
	2							
	3							
B	4							
	5							