

- Résumé
- 1. Randomisation
- 2. Contrôle expérimental (homogénéité unitaire induite)
- 3. Expériences naturelles (randomisation supposée)
- 4. Comparaisons avant/après
- 5. Contrôle ex post I : régression
- 6. Contrôle ex post II: appariement et pondération
- 7. Variables instrumentales
- 8. La méthode de régression par discontinuité (regression discontinuity design, RDD)
- 9. Traçage de processus
- 10. Stratégies “porte d’entrée” (argument basé sur les mécanismes)

Résumé

Les expériences sont un moyen de déterminer si quelque chose cause quelque chose d’autre. L’idée de base, c’est d’essayer pour savoir. Ce qui est délicat, c’est d’essayer d’une manière qui permette d’avoir assez confiance pour croire les effets causaux. L’intervention randomisée est une stratégie qui occupe une place de choix dans la boîte à outils du chercheur. C’est la stratégie qui est au cœur de la plupart des recherches expérimentales menées par les membres de EGAP. Mais il existe d’autres stratégies qui sont parfois plus appropriées. Nous décrivons ici les dix stratégies les plus importantes pour déterminer les effets causaux. ¹

1. Randomisation

La stratégie utilisée dans les essais randomisés contrôlés (interventions randomisées, expériences randomisées) consiste à utiliser une forme de loterie pour déterminer qui, dans un groupe, aura ou n’aura pas accès à un traitement ou à un programme (ou peut-être qui aura accès en premier et qui aura accès plus tard, ou qui aura accès à une version et qui aura accès à une autre). L’élégance de l’approche est qu’elle utilise le hasard pour déterminer quels sont les effets systématiques d’un programme. Le caractère aléatoire réduit le risque qu’une relation observée entre le traitement et les résultats soit due à des “facteurs de confusion” - i.e. d’autres différences entre les groupes (par exemple, on pourrait craindre que les choses s’améliorent dans les zones de traitement précisément parce que les programmes choisissent une zone qui fonctionne déjà bien, mais savoir que la sélection est aléatoire efface complètement cette préoccupation). La randomisation est puissante car elle garantit qu’il n’y a pas de relation systématique entre le traitement et toutes les autres caractéristiques qui peuvent affecter les résultats, que vous en soyez conscient ou non. Pour cette raison, c’est souvent considéré comme l’étalon-or. Cependant, la randomisation ne peut être utilisée toujours et partout, à la fois pour des raisons éthiques et pratiques. Mais elle peut être utilisée dans beaucoup plus de situations que les gens ne le pensent. Voir Humphreys et Weinstein

(http://www.columbia.edu/~mh2245/papers1/HW_ARPS09.pdf) pour une discussion sur les forces et les limites de l’approche en économie politique du développement.

2. Contrôle expérimental (homogénéité unitaire induite)

Une deuxième stratégie plus utilisée en laboratoire et également en sciences physiques consiste à utiliser un contrôle expérimental pour s'assurer que deux unités sont identiques à tous les égards pertinents, à l'exception du traitement. Par exemple, si vous vouliez voir si une balle lourde tombe plus vite qu'une balle plus légère, vous pouvez vous assurer qu'elles ont la même forme et la même taille et les laisser tomber toutes les deux en même temps, dans les mêmes conditions météorologiques, et ainsi de suite. Vous attribuez ensuite toute différence de résultats à la caractéristique que vous n'avez pas maintenue constante entre les deux unités. Cette stratégie est fondamentalement différente de celle utilisée dans les essais randomisés. Dans les essais randomisés, vous abandonnez l'idée de tout garder fixe et cherchez plutôt à vous assurer que la variation naturelle - sur des variables que vous pouvez ou ne pouvez pas observer - ne produit pas de biais dans vos estimations ; en outre, vous cherchez à évaluer les effets moyens pour une plage de conditions de base plutôt que pour un ensemble fixe de conditions de base. Les mérites de l'approche du contrôle expérimental dépendent de votre confiance dans le contrôle effectif de tous les facteurs pertinents ; si le contrôle n'est pas effectif, alors une approche randomisée peut être meilleure.

3. Expériences naturelles (randomisation supposée)

Parfois, les chercheurs ne sont pas en mesure de randomiser, mais l'inférence causale est toujours possible car la nature a fait la randomisation pour vous. La caractéristique clé de l'approche "expérience naturelle" est que vous avez des raisons de croire que la variation de certains traitements naturels est "supposée aléatoire". Par exemple, disons que les places dans une école sont attribuées par tirage au sort. Ensuite, vous pourrez peut-être analyser les effets de la fréquentation scolaire comme s'il s'agissait d'un essai randomisé contrôlé. Une étude intelligente des effets des conflits sur les enfants par Annan et Blattman (<http://www.chrisblattman.com/documents/research/2010.Consequences.RESTAT.pdf>) a utilisé le fait que la Lord's Resistance Army (LRA) en Ouganda a enlevé des enfants de façon assez aléatoire. Une autre étude intelligente sur les programmes de Désarmement, Démobilisation et Réintégration (DDR) par Gilligan, Mvukiyehe et Samii (http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1911968) a utilisé le fait que les opérations d'une ONG ont été interrompues en raison d'un différend contractuel, ce qui a donné lieu à un groupe de contrôle "naturel" d'ex-combattants qui n'ont pas bénéficié de programmes de démobilisation. Voir le livre de Dunning (<http://www.cambridge.org/us/academic/subjects/politics-international-relations/research-methods-politics/natural-experiments-social-sciences-design-based-approach>) pour trouver et analyser des expériences naturelles.

4. Comparaisons avant/après

Souvent, la première chose vers laquelle les gens se tournent pour déterminer les effets causaux est la comparaison des unités avant et après le contrôle. Ici, vous utilisez le passé comme un contrôle pour le présent. L'idée de base est très intuitive : vous éteignez la lumière et vous voyez la lumière s'éteindre ; attribuer le changement de lumière à l'action semble facile même en l'absence de toute randomisation ou contrôle. Mais pour de nombreuses interventions sociales, l'approche n'est pas si fiable, en particulier dans des environnements changeants. Le problème est que les choses s'améliorent ou empirent pour de nombreuses raisons sans rapport avec les traitements ou les programmes qui vous intéressent. En fait, il est possible qu'en raison de toutes les autres choses qui changent, les choses peuvent empirer dans un domaine du programme même si les programmes ont eu un effet positif (donc les choses empirent mais ne sont toujours pas aussi mauvaises qu'elles l'auraient été sans le programme!). Une approche plus sophistiquée que la simple comparaison avant/après est appelée "méthode des doubles différences" – en gros, vous comparez la différence avant/après dans les zones de

traitement avec celles dans les zones de contrôle. C'est une bonne approche mais vous devez toujours être sûr que vous avez de bons groupes de contrôle et en particulier que les groupes de contrôle et de traitement ne sont pas susceptibles de changer différemment pour des raisons autres que le traitement.

5. Contrôle ex post I : régression

L'approche la plus courante de l'identification causale dans les travaux de statistique appliqués est peut-être l'utilisation de la régression multiple pour contrôler les facteurs de confusion possibles. L'idée est d'essayer d'utiliser toutes les informations dont vous disposez sur les raisons pour lesquelles les zones de traitement et de contrôle ne sont pas facilement comparables et d'ajuster statistiquement ces différences. Cette approche fonctionne bien dans la mesure où vous pouvez déterminer et mesurer les facteurs de confusion et savoir comment ils sont liés au traitement, mais elle n'est pas correcte si vous ne savez pas quels sont les facteurs de confusion. En général, nous ne savons tout simplement pas quels sont tous les facteurs de confusion et cela expose cette approche à toutes sortes de biais (en effet, si vous contrôlez les mauvaises variables, il est possible *d'introduire* un biais là où il n'y en avait pas auparavant).

6. Contrôle ex post II: appariement et pondération

Diverses approches alternatives cherchent à tenir compte des variables de confusion en associant soigneusement les unités de traitement à une ou plusieurs unités de contrôle. L'appariement présente certains avantages par rapport à la régression (par exemple, les estimations peuvent être moins sensibles aux choix de la forme fonctionnelle), mais l'idée de base est néanmoins similaire. En effet, les méthodes d'appariement peuvent être mises en œuvre dans un cadre de régression en utilisant des poids appropriés. Comme la régression, cette stratégie repose sur la conviction qu'il n'y a pas de variables de confusion importantes que le chercheur ignore ou est incapable de mesurer. Les méthodes spécifiques comprennent :

- appariement complet optimal (<http://dept.stat.lsa.umich.edu/~bbh/hansen2004.pdf>), voir le package `optmatch` (<https://github.com/markmfredrickson/optmatch>)
- appariement optimal avec équilibrage précis via une programmation mixte en nombres entiers (<https://projecteuclid.org/euclid.aoas/1396966284>), voir aussi le package `designmatch` (<https://cran.rstudio.com/web/packages/designmatch>) et l'article comparant les approches (<http://jrzubizarreta.com/evaluation.pdf>)
- appariement multi-niveaux optimal (pour les designs avec des écoles et des étudiants) (<https://projecteuclid.org/euclid.aoas/1536652962>)
- appariement creux optimal (<https://www.stat.berkeley.edu/~spi/software.html>)
- appariement généralisé complet (<https://arxiv.org/abs/1703.03882>)
- appariement grossier exact (<http://gking.harvard.edu/cem>)
- appariement génétique (<http://sekhon.berkeley.edu/papers/GenMatch.pdf>)
- équilibre d'entropie (<http://web.stanford.edu/~jhain/Paper/PA2012.pdf>)
- pondération inverse du score de propension (<http://pan.oxfordjournals.org/content/18/1/36.short>)
- pondération de l'équilibre stable (<http://www.columbia.edu/~jz2313/sbw.pdf>)
- contrôles synthétiques (<http://web.stanford.edu/~jhain/Paper/AJPS2015a.pdf>).

7. Variables instrumentales

Une approche très différente pour estimer les effets causaux peut être utilisée si les chercheurs peuvent trouver une caractéristique qui explique pourquoi un groupe donné a reçu un traitement mais qui n'est par ailleurs pas liée au résultat d'intérêt. Une telle caractéristique est appelée un instrument. Par exemple, disons que vous êtes intéressé par l'effet sur l'emploi d'un programme d'aide à la précarité, et que la plupart des personnes ont accédé à ce programme parce qu'elles avaient un lien de parenté avec un agent travaillant pour ce programme particulier. Ensuite, s'il n'y avait pas d'autres façons d'établir une relation entre le lien de parenté de cette personne et ses perspectives d'emploi, alors vous pouvez déterminer l'effet du programme en calculant l'effet d'être un parent de cette personne sur les perspectives d'emploi. Cela a été une approche assez populaire, mais l'enthousiasme pour cette approche est un peu retombé, essentiellement parce qu'il est difficile de trouver un bon instrument. Une application intelligente pour examiner les effets de la pauvreté sur les conflits a utilisé les précipitations en Afrique comme instrument de revenu/croissance. Bien que l'on s'inquiète que la corrélation entre conflit et pauvreté soit due au fait que le conflit pourrait causer la pauvreté, il ne semble pas plausible que le conflit provoque des précipitations ! Ainsi, l'utilisation des précipitations comme instrument ici a augmenté la certitude qu'il existe réellement une relation causale, et pas seulement une corrélation, entre pauvreté et conflit (<http://emiguel.econ.berkeley.edu/research/economic-shocks-and-civil-conflict-an-instrumental-variables-approach>).

8. La méthode de régression par discontinuité (regression discontinuity design, RDD)

La régression par discontinuité est l'une des approches les plus sous-utilisées, mais elle a un potentiel énorme. La stratégie fonctionne comme suit. Disons qu'un programme va être mis à la disposition d'un ensemble de bénéficiaires potentiels. Ces bénéficiaires potentiels sont tous classés selon un ensemble de critères pertinents, tels que le niveau d'éducation antérieur, le statut d'emploi, etc. Ces critères peuvent être quantitatifs ; mais ils peuvent également inclure des informations qualitatives telles que des évaluations issues d'entretiens. Ces critères individuels sont ensuite agrégés en un seul score et un seuil est identifié. Les candidats dont le score est supérieur à ce seuil sont admis au programme, tandis que les autres ne le sont pas. Les groupes "projet" et "comparaison" sont ensuite identifiés en sélectionnant des candidats proches de ce seuil de part et d'autre. En utilisant cette méthode, nous pouvons être sûrs que les unités traitées et témoins sont similaires, au moins autour du seuil. De plus, nous avons une mesure directe de la principale caractéristique sur laquelle ils diffèrent (leur score sur les critères de sélection). Cette information fournit la clé pour estimer l'effet d'un programme en comparant les résultats entre ces deux groupes. L'avantage de cette approche est que tout ce qui est nécessaire est que l'organisme en charge utilise un ensemble clair de critères (qui peuvent être transformés en un score) sur lesquels ils prennent la décision d'assignation au traitement. L'inconvénient est que des estimations vraiment fiables de l'impact ne peuvent être faites que pour les unités situées juste autour du seuil. Pour un aperçu de la régression par discontinuité, voir Skovron et Titiunik (<http://www-personal.umich.edu/~titiunik/papers/SkovronTitiunik2015.pdf>), Lee et Lemieux (<http://econ.sites.olt.ubc.ca/files/2014/02/Lee-Lemieux-rev.pdf>). Pour deux applications intéressantes, voir Manacorda et al. en Uruguay (<http://emiguel.econ.berkeley.edu/research/government-transfers-and-political-support>) et Samii au Burundi (<http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=8963107>).

9. Traçage de processus

Dans de nombreux travaux qualitatifs, les chercheurs essaient d'établir la causalité en cherchant à savoir si le fait d'être dans un programme est associé à de meilleurs résultats, mais aussi (a) en recherchant les étapes du processus qui montreraient que le programme a eu les effets que vous pensez qu'il a eu (b) et en recherchant des preuves d'autres résultats qui devraient être observés si (ou peut-être : si et seulement si) le programme était efficace. Par exemple, ils veulent savoir si les personnes participant à un programme d'aide à la précarité ont obtenu un emploi, mais aussi si elles ont été formées à quelque chose d'utile, ont obtenu de l'aide des personnes participant au programme pour trouver un employeur dans ce domaine, etc. Si toutes ces étapes sont présentes, cela donne l'assurance que la relation est causale et non fausse. Si un programme a été mis en œuvre mais que personne n'y a réellement participé, cela peut laisser penser que toute corrélation entre le traitement et les résultats est fausse. La difficulté avec cette approche est qu'il peut être difficile de savoir si un élément de preuve au sein du processus a une valeur probante. Par exemple, un programme peut avoir des effets positifs (ou négatifs) à travers de nombreux processus dont vous ne savez rien et des processus que vous jugez importants, peuvent ne pas l'être. Voir Humphreys et Jacobs (<http://www.columbia.edu/%7Emh2245/papers1/BIQQ.pdf>) pour une description de la logique bayésienne sous-jacente au traçage des processus et des illustrations sur la façon de le combiner avec d'autres approches statistiques.

10. Stratégies "porte d'entrée" (argument basé sur les mécanismes)

Une dernière approche, conceptuellement proche du traçage de processus, consiste à argumenter à partir des mécanismes. Supposons que vous sachiez que seul A peut causer C uniquement via B . De plus, vous savez que rien d'autre ne cause à la fois B et C (autre que, peut-être, via A) et que rien d'autre ne cause à la fois A et B . Ensuite, la covariation entre A et B et entre B et C peut être utilisée pour évaluer l'effet de A sur C . L'avantage est que la causalité peut être établie même en présence de facteurs de confusion — par exemple si des variables non observées causent à la fois A et C . La difficulté cependant est que la stratégie nécessite une confiance forte en vos croyances sur la structure des relations causales. Pour en savoir plus, voir Pearl (2000) (https://books.google.com/books?id=wnGU_TsW3BQC&hl=en).

-
1. Auteur d'origine : Macartan Humphreys. Révisions mineures : Winston Lin, 30 août 2016. Le guide est un document vivant et sujet à une mise à jour par les membres de EGAP à tout moment ; les contributeurs répertoriés ne sont pas responsables des modifications ultérieures. ←