

Introduction

OBJET DU LIVRE

La critique de la statistique a déjà donné lieu à la production de nombreux ouvrages et articles dont Desrosières est la personnalité centrale en France. Ce livre s'inscrit dans le prolongement de cette critique de l'usage de la statistique mais avec un axe légèrement différent. En effet, le terme *statistique* couvre deux sens différents. Selon le sens commun, la statistique renvoie à des données chiffrées ou des indicateurs statistiques, c'est-à-dire chiffres exhibés par des études socio-économiques ou les médias. C'est d'ailleurs sur cet aspect que la critique est la plus connue. Cependant, il y a un autre sens du terme *statistique* qui renvoie quant à lui à ce que l'on appelle des *modèles statistiques*, généralement considérés comme un champ des mathématiques (on parle de *statistique-mathématique*). Ce mot *statistique* est à comprendre ici dans le sens de valider/prouver des modèles (simples ou complexes). Il s'avère qu'il y a peu de textes critiquant les modèles statistiques et ceux-ci sont souvent peu accessibles au grand public.

Nous sommes passés, à partir du XVIII^e siècle, d'un mode d'appréhension de la réalité qui s'exprimait par la description verbale, à un mode d'appréhension où la quantification, la mesure et les nombres font office de langage explicatif et en conséquence,

l'argument de preuve dominant est devenu l'argument statistique. Ceci est bien sûr en lien avec ce que l'on appelle la *modernité* et la recherche de rationalité¹. La *statistisation* de la pensée a été, dès son origine, une démarche politique impulsée par les États (XVIII^e siècle). L'État s'est mis à recourir de plus en plus aux statistiques pour structurer son intervention, pour connaître la population sur laquelle il avait autorité, pour éventuellement impulser des changements, surveiller et modeler les comportements. Aujourd'hui, les politiques publiques sont ainsi basées sur des indicateurs purement quantitatifs dits d'efficience. Cette statistisation a aussi gagné (depuis longtemps) la pensée scientifique et constitue actuellement la principale source d'autorité scientifique. La mathématisation puis l'usage de la statistique mathématique, sont rapidement devenus centraux en physique. À partir de là, le développement de la quantification et l'usage de modèles statistiques se sont étendus à tous les champs scientifiques. Même les études qui n'intègrent que des « statistiques » au sens de « données chiffrées » (comme des pourcentages ou des moyennes) renvoient de façon sous-jacente à des outils de la statistique-mathématique. La moyenne par exemple est un résultat intuitif mais qui est formalisable mathématiquement et a des propriétés mathématiques qui permettent souvent d'extrapoler des résultats ; c'est la base des sondages. Donc, toutes les études quantitatives basées sur des sondages/enquêtes/protocoles expérimentaux... qui touchent à l'humain vont utiliser des modèles mathématiques sous-jacents. Ces modèles renvoient d'une part à des hypothèses mathématiques parfois fortes et d'autre part, à une vision de la réalité sociale inscrite dans un cadre philosophique/idéologique

1. REY Olivier, *Quand le monde s'est fait nombre*, Paris, Stock, 2016 ; DESROSIÈRES Alain, *La politique des grands nombres*, Paris, La Découverte, 2010 ; PORTER Theodore, *La confiance dans les chiffres. La recherche de l'objectivité dans la science et dans la vie publique*, Paris, Les Belles lettres, 2017.

particulier. Peu de scientifiques sont éveillés à l'épistémologie² de la statistique car les formations universitaires n'offrent que très rarement des cours sur ce champ. En conséquence, nous développerons une critique épistémologique des modèles statistiques souvent ignorée y compris par des scientifiques.

Ce livre s'adresse aussi bien aux non-spécialistes des statistiques qu'à celles et ceux qui les utilisent. Son objectif est de développer une critique de l'usage des modèles statistiques en sciences humaines entendues ici comme des sciences dévolues à l'analyse de l'humain. Il se veut être un outil militant contre *l'argument statistique* des modèles quantitatifs en sciences humaines. Dès lors, cette critique portera sur plusieurs champs : critique de l'usage (ou mésusage) des modèles statistiques, critique des soubassements idéologiques des modèles statistiques et critique de la fonction politique de ces modèles statistiques. Ce livre s'inscrit dans un courant général de critique de l'usage de la quantification et de la statistique. Il est destiné à des personnes non expertes en statistique et/ou en sciences humaines mais il est également destiné aux personnes qui utilisent les statistiques dans leurs activités professionnelles ou extraprofessionnelles (peu importe le domaine de cette utilisation) car la critique développée dans cet ouvrage peut s'étendre bien au-delà des sciences humaines.

Notre développement sera parcellaire car il existe énormément de possibilités de modéliser en statistique et concrètement il faudrait analyser chaque modèle dans chaque situation où il est utilisé pour pouvoir en poser les limites. Comme nous ne pouvons prétendre discuter de tous les types de modèles statistiques en sciences humaines, nous centrerons notre critique sur les modèles ayant pour objectif de comprendre l'humain par des données individuelles. Ces modèles considèrent implicitement les

2. Le terme *épistémologie* sera utilisé pour nommer la manière philosophique/idéologique d'aborder les problèmes en sciences.

individus comme interchangeables et n'intègrent pas de dimensions macrosociales. C'est ce que l'on appelle *l'individualisme méthodologique*³ dominant en sciences humaines lorsque l'on a recours à des modèles statistiques. Cette critique pourra cependant s'étendre aux modèles statistiques qui prétendent avoir une démarche plus holistique que les modèles basés sur l'individualisme méthodologique, comme ce que l'on appelle *l'analyse de données exploratoires* très utilisée en sociologie. Ces approches dites exploratoires n'utilisent pas d'hypothèses formelles associées aux modèles statistiques (au sens d'hypothèses probabilistes associées aux données). Elles ont pour objectif d'analyser/décrire des structures relationnelles au-delà de la personne. Ces approches exploratoires se basent sur une codification de données, c'est-à-dire des catégories statistiques dont la critique a déjà été faite par de nombreux travaux⁴. Nous pouvons reprendre à notre compte ce que dit Porter au sujet des « drôles de nombres » qui servent à synthétiser et en conséquence simplifient et aplatissent le réel. Nombres qui sont généralement présentés comme une « dure réalité de faits objectifs » alors qu'ils entraînent « la folie raisonnante, c'est-à-dire raisonner de façon logique à partir de postulats biaisés ou erronés » (Porter⁵).

PLAN DU LIVRE

Pour permettre aux personnes non expertes en statistique de suivre le développement de ce livre, celui-ci sera essentiellement construit sur des exemples. La bibliographie renverra à des travaux académiques afin de compléter nos chapitres et permettre aux

3. En entendant ce terme non pas de la manière dont Boudon l'a développé mais bien comme représentant les approches méthodologiques basées sur les individus, voir : GRENIER Jean-Yves, GRIGNON Claude et MENER Pierre-Michel, *Le modèle et le récit*, Paris, Maison des Sciences de l'homme, 2001.

4. Voir la bibliographie commentée à la fin de cet ouvrage.

5. PORTER Theodore, *La confiance dans les chiffres...*, *op. cit.*

lecteurs/lectrices d'aller plus loin s'ils/elles le souhaitent. Dans la bibliographie, nous avons mis en avant certains livres qui nous apparaissent centraux, avec un résumé pour chacun d'entre eux. Il peut être intéressant de lire ces résumés lorsque nous ferons référence à ces ouvrages.

Le premier chapitre se centrera sur différents problèmes méthodologiques usuels des sondages ainsi que certaines études académiques. Le second chapitre développera par des paradoxes des erreurs usuelles en statistique inférentielle. Aucune connaissance en statistique n'est exigée. Il cherchera à convaincre que la « preuve statistique » n'existe pas. Pour éclairer les lectrices et lecteurs, nous introduirons à la fin de ce chapitre une section qui explique les fondements de l'inférence statistique et les problèmes associés. Le troisième chapitre étudiera les *leurre*s des modèles statistiques dits causaux en sciences humaines. Lui non plus ne nécessitera pas de connaissances particulières même s'il fera référence à des modèles usuels en statistique (le fond du raisonnement devrait être accessible même sans connaissance en statistique). La discussion portera sur les fondements épistémologiques et idéologiques des modèles statistiques en sciences humaines. Le quatrième chapitre discutera de problèmes souvent méconnus même par les équipes de recherche. Pour utiliser des modèles statistiques, il faut pouvoir associer des valeurs aux objets analysés. C'est-à-dire les mesurer. Nous discuterons donc de la mesure en psychologie car (presque) tous les modèles statistiques utilisés en sciences humaines renvoient aux fondements épistémologiques de la mesure en psychologie. Les modèles statistiques en psychologie modélisent des attributs psychologiques qui sont quantifiés. Quelle est la légitimité d'attribuer des valeurs numériques à des attributs psychologiques ? Concrètement, qu'elle est la nature d'un attribut psychologique (ce qu'on appelle son *ontologie*) ? Ce chapitre parlera en conséquence des neurosciences de façon critique. Il cherchera à montrer que la plupart des articles scientifiques qui usent de

modèles statistiques, faisant appel à des mesures sur l'humain, souffrent de faiblesses voire d'erreurs. Le chapitre suivant sera une analyse de la fonction politique et policière des modèles statistiques en sciences humaines. En effet, la critique exposée dans les chapitres précédents est connue et développée par différents auteurs/autrices depuis bien longtemps dans un certain nombre d'articles académiques ou livres⁶. Nous expliquerons les raisons sociales et politiques de l'usage actuel des modèles statistiques et pourquoi celles et ceux qui utilisent les résultats des études quantitatives en sciences humaines n'entendent pas cette critique. Enfin, le dernier chapitre développera deux exemples caractéristiques de tout le propos de ce livre. Le premier exemple portera sur l'utilisation politique d'un article économique dont les analyses statistiques sont complètement erronées. Le second exemple développera une analyse critique de l'usage des statistiques ayant justifié la politique sanitaire en France durant la crise du Covid.

L'objectif de cet ouvrage est de donner *in fine* des armes critiques contre l'usage politique des modèles statistiques en sciences humaines. Des exemples serviront de support et il n'y aura pas de développements théoriques en mathématique. En conséquence, certains raisonnements seront parfois un peu simplifiés afin de permettre aux personnes non expertes en mathématique de les comprendre.

AXE DU LIVRE

Si ce livre s'inspire et prolonge les travaux cités dans la bibliographie et qui, à des degrés divers et sur des axes différents, posent une critique de l'usage de la statistique, il n'est pas une attaque de la statistique en tant qu'outil intellectuel. Il se veut une critique de la pratique de la modélisation statistique et de son

6. Voir la bibliographie.

usage politique. La critique de la modélisation statistique ne doit pas se confondre avec le refus d'user de tels modèles statistiques. Il y a une contradiction dans les modèles statistiques qui amène une contradiction dans leur critique. Dans un modèle statistique, on analyse le collectif, on cherche des tendances au-delà de l'individu isolé sans intégrer les dimensions macrosociales. Dès lors, un modèle statistique simplifie en rapportant le collectif à un portrait-robot unique, c'est-à-dire à un idéaltype qui n'existe pas mais qui peut nourrir notre réflexion. Critiquer toute modélisation statistique est donc critiquable. Il faut bien essayer de donner du sens au réel et les modèles statistiques sont un des moyens pour y parvenir. Ne pas user de modèles globalisant le réel serait en rester à une démarche sans ressort ni dynamique collective dans certains schémas explicatifs. Raison pour laquelle les données et les modèles statistiques sont des armes utilisées depuis longtemps par la critique sociale⁷. Un modèle statistique permet dans certaines situations d'avoir un regard plus global et permet des généralisations, avec prudence toutefois.

L'objectif de cet ouvrage n'est pas de considérer qu'il faut s'interdire d'utiliser des modèles statistiques en sciences humaines. Ils sont parfois nécessaires pour essayer de donner du sens à des phénomènes étudiés. Comme tout outil intellectuel, ils ne sont cependant jamais épurés de la matrice idéologique de la personne qui les utilise. Notre travail sert à alerter sur l'usage des modèles statistiques car ils sont toujours inscrits dans des tensions sociales, politiques et idéologiques. Les modèles statistiques sont des armes politiques qui s'appuient sur des choix idéologiques sous-jacents, d'où le titre : *Stat Wars. Le côté obscur de la force des statistiques*.

7. BRUNO Isabelle, DIDIER Emmanuel et PRÉVIEUX Julien, *Statactivisme*, Paris, Zones, 2014.