

# Introduction

François Beck

---

En quelques décennies, être interrogé dans le cadre d'une enquête par sondage est devenu un événement courant, au point que la confrontation avec un enquêteur (ou une autre forme de sollicitation invitant à répondre) apparaît désormais banal. Pourtant, la participation des personnes contactées pour des enquêtes en population générale devient de plus en plus difficile à obtenir. Cette difficulté croissante rend la question de la représentativité, ainsi que des processus méthodologiques mis en œuvre pour l'atteindre, plus que jamais cruciale. Elle cristallise une bonne part de l'enjeu de telles enquêtes, dans une compétition entre objectifs de qualité et contraintes de coût. Une grande énergie est en effet régulièrement déployée par les instituts de sondage pour tenter de montrer qu'un échantillon de panélistes volontaires, aux caractéristiques bien choisies, contrôlées puis redressées, peut satisfaire des commanditaires d'études soucieux de réduire les coûts – souvent prohibitifs – qu'entraînent les protocoles d'enquête les plus scrupuleux. De nombreux travaux menés ces dernières années portant notamment sur des sujets sensibles, montrent toutefois les limites de tels arguments de « commodité ». La difficulté à joindre certaines personnes dans les échantillons aléatoires, ou le fait que celles-ci ne disposent pas d'un téléphone fixe (ce qui est souvent un signe de précarité sociale), ne dépendent pas seulement de quelques facteurs socioéconomiques comme l'âge ou le revenu ; ces caractéristiques ne sont donc pas aisément maîtrisables par un simple redressement. Les femmes et hommes qui acceptent encore de répondre aux enquêtes quantitatives peuvent-ils porter la parole de ceux qui ne croient plus aux sondages ou n'y ont jamais cru, et de ceux qui veulent bien y croire, mais pas suffisamment pour y consacrer un peu de leur propre temps ?

Dans l'univers des enquêtes quantitatives, la notion de représentativité ouvre en effet la voie de la légitimité au chiffre dans le débat public, sorte de garantie qui correspondrait à une valeur absolue. La généralisation de l'utilisation d'enquêtes, tant dans les travaux scientifiques que dans la société civile, s'accompagne d'une diffusion quotidienne de leurs résultats à des fins d'informations politiques, économiques et sociales, de marketing ou de connaissance de l'opinion sur les sujets les plus variés. Ces résultats sont régulièrement proclamés au nom de la représentativité de l'échantillon considéré. Celle-ci paraît par conséquent plutôt simple à mettre en œuvre et à garantir, voire aller de soi dès lors qu'il s'agit d'une

enquête quantitative. Elle demeure en réalité un concept mou qui offre prise à une très (trop) large variété d'acceptions : un échantillon serait « représentatif » comme un point de vue serait « intéressant » ou « sérieux ».

On peut au minimum considérer que la notion, dans son acception contemporaine, réside dans l'idée que, pour être représentative, une enquête doit reposer sur un échantillon qui est une réduction offrant une image conforme à celle de la population observée selon des critères bien définis. Le fondement de cette position s'appuie sur la *notion d'inférence*, qui entend que ce qui s'observe sur la partie est le reflet de ce qui serait visible sur le tout si l'observateur disposait des moyens nécessaires à une appréhension exhaustive de la population de référence. L'inférence est en cela parfois comparée à la synecdoque, figure de style qui consiste à prendre la partie pour le tout.

Mais il ne s'agit là que d'une esquisse de définition, qui ne doit pas laisser croire à un consensus au sein de la communauté scientifique. Les récents échanges qui ont eu lieu autour de la publication de l'ouvrage *Le raisonnement statistique en sociologie* (Marion Selz, Florence Maillachon, Puf, 2009), puis dans le cadre d'une séance *ad hoc* du séminaire de « Modélisation et méthodes statistiques en sciences sociales » du centre Maurice Halbwachs, à la suite d'une intervention d'Emmanuel Didier (CNRS-GSPM<sup>(1)</sup>) sur « la représentativité statistique : histoire et usages actuels », ont montré que les doutes autour de cette notion, pourtant fondamentale de la statistique et plus largement des sciences sociales, étaient largement partagés par les chercheurs en sciences sociales et les statisticiens.

La Société française de statistique s'est ensuite associée au centre Maurice Halbwachs pour proposer une réflexion approfondie en organisant, en février 2010, une journée ayant rassemblé plus de deux cents personnes à l'ENS-Ulm, témoignant ainsi de l'intérêt et des nombreuses interrogations, manifestés autour du sujet. Les communications délivrées lors de cette journée constituent les chapitres de cet ouvrage.

Les ambiguïtés autour du terme et de la notion de représentativité statistique découlent sans doute de deux raisons principales : d'une part, l'adjectif « représentatif », non spécifique du vocabulaire des statistiques, est polysémique ; d'autre part, l'expression tend à être galvaudée dans les médias ou par les instituts de sondage, dans des discours ou en association avec des pratiques parfois peu respectueuses des pré-requis élémentaires de la représentativité statistique.

Rappelons ici quelques usages de la notion de représentativité, utilisée dans bien d'autres contextes que celui des statistiques. Selon la définition du Petit Robert (2006) est représentatif ce « qui tient lieu de quelque chose ou a pour but de rendre sensible » ou « qui concerne la présentation du peuple par des personnes désignées (généralement élues), pour l'exercice du pouvoir » ; ici il s'agit en quelque sorte de porte-parole. Vient ensuite : « qui représente à l'esprit un objet dont il prend connaissance » ; c'est alors une fonction d'explicitation symbolique qui est évoquée. Enfin : « Propre à représenter un groupe, qui le représente bien » est sans doute le sens qui s'approche le plus de ce que

---

(1) Groupe de sociologie politique et morale.

désigne un échantillon « représentatif » même si pour un statisticien, aucune de ces définitions ne convient tout à fait.

Qu'en est-il de la définition théorique de la représentativité statistique ? En existe-t-il une qui soit précise, circonscrite et fasse consensus au sein même de la communauté des statisticiens ? Le premier enjeu de cet ouvrage est de retracer l'histoire ancienne et contemporaine de l'évolution du concept. Les deux premiers chapitres concernent ainsi la recherche d'une définition théorique de la représentativité statistique.

Le concept de représentativité a eu d'emblée du mal à s'imposer, comme nous le rappelle l'histoire de la statistique. Dans le chapitre 1, consacré à l'histoire du concept, Emmanuel Didier (CNRS-GSPM), présente la genèse de cette notion mise en perspective par les efforts que déploient les probabilistes pour en restreindre la grande pluralité de sens initiale. À cette phase fait suite celle de la renommée acquise avec les sondages d'opinion et ayant entraîné d'autres transformations de la notion de représentativité.

Olivier Sautory (Insee) montre, dans le chapitre 2, la diversité des significations que prend le mot « représentativité » au sein même des catégories de la statistique publique et en particulier de l'Insee, avant de préciser la théorie centrale sur laquelle repose la représentativité statistique et exposer les méthodes utilisées à l'Insee pour assurer cette représentativité à laquelle, souvent, la non-réponse vient malheureusement nuire.

La définition théorique de la représentativité est liée aux règles d'inférence qui précisent dans quelle mesure – c'est-à-dire avec quels intervalle et seuil de confiance –, les observations et mesures constatées sur l'échantillon sont généralisables à la population qu'il s'agit de représenter. Il reste ensuite, et ce n'est pas le moindre des défis, à déterminer sur quels critères et jusqu'à quel point l'échantillon doit ressembler à la population pour bien la représenter.

Autrement dit, en accord avec ces principes, il faut affronter la difficulté de la construction matérielle d'un échantillon qui pourrait correspondre à cette définition. De nombreuses questions se posent lorsque l'on ne dispose pas de base de sondage ou qu'elle n'est pas exhaustive, lorsque la population étudiée comporte une grande diversité, qu'elle est difficile à atteindre, que l'on s'intéresse en particulier en son sein à une minorité difficile à cerner, à circonscire, ou encore lorsque l'on suit des cohortes dont la composition se modifie au fil du suivi. D'autres dimensions moins souvent évoquées, bien que cruciales, entrent également en compte dans la détermination de la représentativité d'une enquête : la qualité des enquêteurs, la pertinence et l'ergonomie des questionnaires, les conditions de passation...

Diverses autres circonstances occasionnent des problèmes dont la résolution ne fait pas forcément consensus, tels que les comparaisons internationales ou le choix des sujets dans les enquêtes cas-témoins.

Une fois considérée la question de la définition même se pose donc souvent celle de la conduite pratique à tenir dans cette quête de la représentativité. Différentes méthodes permettent de s'approcher des objectifs visés avec plus ou moins de

réussite. Quelle est la pratique de ces méthodes que développent les producteurs de données ? Quelles différences observe-t-on, par exemple, entre les pratiques mises en œuvre par la statistique publique et par les instituts de sondages privés ? Et au-delà, la représentativité est-elle toujours nécessaire à l'établissement de résultats « valides » ?

Quatre chapitres sont consacrés à ces questions particulières de la mise en œuvre de la représentativité statistique dans différentes situations d'enquêtes.

Alain Tripier (Sereho) dévoile, dans le chapitre 3, les pratiques qui se nouent dans les échanges entre annonceurs et instituts de sondages, en particulier quant au choix des personnes à interviewer, de la taille de l'échantillon, de la façon d'appliquer la méthode des quotas et du mode d'enquête. Des problèmes concrets sont abordés au prisme de l'expérience des instituts de sondages privés, à partir de quelques exemples tels que la question des enquêtes sur Internet.

L'objectif de Maryse Marpsat (Insee-CMH) et de Nicolas Razafindratsima (Ined), dans le chapitre 4, est autre : ils cherchent à qualifier la recherche de la représentativité dans le cas des « populations difficiles à joindre » et présentent quelques méthodes permettant de tirer un échantillon « extrapolable », c'est-à-dire à partir duquel on peut produire une estimation sans biais de certaines variables sur toute la population d'intérêt. L'expérience française des enquêtes auprès des personnes sans domicile apparaît ici particulièrement éclairante.

Arnaud Bringé (Ined) et Patrick Festy (Ined) abordent, dans le chapitre 5, les problèmes de comparabilité internationale pour montrer en quoi celle-ci dépend avant tout de la comparabilité des bases de sondage disponibles dans les différents pays concernés. Ils posent notamment la question de la représentativité et de la significativité lors des différents types d'analyses de données internationales.

Enfin, dans le chapitre 6, Marcel Goldberg (Inserm) et ses collègues donnent un point de vue d'épidémiologistes en s'intéressant en premier lieu, autre challenge de taille, à la causalité de la survenue de phénomènes (maladies ou décès) pour laquelle ils estiment n'avoir pas toujours besoin de représentativité. Sont présentées les méthodes permettant d'approcher la notion de causalité. Plusieurs exemples provenant de cohortes françaises actuelles sont exposés pour illustrer les questions traitées. Les problèmes posés par l'attrition, généralement sélective, produite au cours des suivis de cohortes, sont mis en évidence.

Cet ouvrage dresse ainsi à la fois un panorama théorique en même temps qu'il illustre, par des exemples pratiques, les différents usages de la notion de représentativité. Il invite à une réflexion et une lecture sans cesse renouvelée de ces questions, face notamment à l'évolution des possibilités et des contraintes techniques.