

## > Échantillon représentatif et data

Les enquêtes marketing et opinion sont traditionnellement basées sur une interrogation d'un **échantillon représentatif** de la population étudiée.

- Un échantillon, obtenu par une méthode statistique éprouvée, est dit représentatif lorsque le profil des participants sur les critères clé pour l'enquête (sociodémographiques, comportementaux...) reflète celui de l'ensemble de la cible, dans l'objectif que les résultats observés sur l'échantillon puissent fournir une estimation à faible biais pour l'ensemble de la cible.
- La méthodologie doit notamment définir cette cible de manière exhaustive : l'échantillon peut être représentatif de la population nationale (« national représentatif ») adulte ou utiliser d'autres tranches d'âge (15-64 ans...), ou encore concerner d'autres populations d'individus (internauts, consommateurs d'un produit, électeurs, salariés...).
- La cible peut également être définie par des unités statistiques autres que l'individu, par exemple le foyer ou ménage, ou concerner des entreprises ou établissements, des avis, tweets ou likes sur les réseaux sociaux notamment.
- La méthodologie doit aussi fixer les critères de représentativité pertinents pour l'enquête, et les moyens utilisés lors des étapes d'échantillonnage ou de redressement pour obtenir sur ces critères des distributions proches de celles de la cible.
- Pour un échantillon d'individus, les critères de représentativité habituels en France peuvent être le sexe, l'âge, la région et la profession/catégorie socioprofessionnelle. Mais certaines problématiques peuvent nécessiter d'intégrer d'autres informations comme le niveau d'éducation ou la taille du foyer, le précédent vote (sondage politique), le revenu et patrimoine (placements financiers), l'équipement et les abonnements (médias).

L'enquête par échantillon représentatif n'est pas la seule méthode utilisée par les professionnels des études.

- La masse de **data déjà existante** sur le web, les réseaux sociaux, les mobiles et l'IoT est une source d'enrichissement pour plusieurs analyses.
- Le dialogue direct avec les membres des réseaux sociaux et des **communautés en ligne** permet également d'obtenir rapidement des données supplémentaires sans provenir d'échantillon représentatif.

Avec la nouvelle data, l'intérêt des échantillons représentatifs est parfois remis en cause. En fait, les experts de la mesure apprécient de disposer d'un large éventail de méthodes pour choisir la plus pertinente en fonction de la population à étudier, des biais de sélection et du sens qui peut être donné aux résultats en particulier.

Ainsi, pour la mesure de la **satisfaction à chaud**, où l'on sollicite l'ensemble des clients suite à une interaction avec un service client, l'échantillon final présente un biais dû notamment à l'appétence plus ou moins forte à répondre aux sollicitations d'enquête de plus en plus nombreuses. Cependant, si les biais sont constants, les variations observées sur la mesure en flux continu peuvent s'avérer utiles pour détecter des tendances et piloter au plus près

la relation client. Cette analyse « à chaud » peut être complétée par une analyse « à froid » intégrant une correction des biais par redressement, et des analyses quantitatives approfondies apportant des éléments de réflexion structurante pour l'organisation de la relation client par exemple. La recherche peut également intégrer l'analyse du verbatim de sources hybrides comme les questions ouvertes des questionnaires, les réclamations ou les avis sur le Web et les réseaux sociaux qui enrichissent les résultats et aident à interpréter les tendances détectées à chaud.

Dans le domaine **politique**, l'échantillon représentatif a plusieurs fois démontré son intérêt par rapport aux méthodes utilisant une masse de données intégrant de forts biais de sélection. Dès les élections américaines de 1936, le statisticien George H. Gallup, fondateur du premier institut de sondage d'opinion, a montré qu'un échantillon de quelques milliers d'interviews a été plus pertinent que les 2,3 millions de votes de paille du magazine *Literary Digest*. L'élection présidentielle française de 2017 a notamment confirmé que les sondages politiques sur échantillon représentatif, malgré l'incertitude des taux de participation et les avis indécis, demeurent plus fiables pour mesurer l'opinion de l'ensemble de la population que les seules analyses des réseaux sociaux. La veille des médias sociaux n'en reste pas moins utile pour repérer les sujets de prédilection des citoyens, les nouvelles tendances et les signaux faibles de l'opinion.

Ainsi, **les méthodes classiques et nouvelles se complètent** pour enrichir l'analyse. Les sondages sur échantillon représentatif sont en particulier utiles pour des estimations quantitatives précises pour l'ensemble de la population, tandis que les nouvelles sources fournissent des enseignements qualitatifs et des métriques des comportements digitaux réels de cibles particulières. Le code déontologique international ICC/ESOMAR a ainsi intégré en 2016 cette nouvelle data et les analytiques dans le périmètre des professionnels des études marketing et opinion.