



LES SONDAGES : UNE TECHNIQUE A MIEUX CONNAITRE

Mémo N°92

On n'en finirait plus de citer des cas où « les sondages nous ont trompé ». On pense alors aux sondages d'opinion, et surtout aux sondages politiques sur les intentions de vote. Mais ce ne sont là que des utilisations particulières d'une technique qui, comme beaucoup d'autres, peut être bien ou mal employée. Un rapide descriptif de ce qu'est la théorie des sondages et un survol de ses pratiques peuvent nous aider à exercer notre indispensable esprit critique. Sans naïveté ni dénigrement systématique. Et éventuellement à utiliser cet outil comme l'ont fait certains militants.

Des échecs cuisants et des succès remarquables

En février 2005, « les sondages » annonçaient un succès écrasant du Traité constitutionnel européen, donné aux alentours de 70% de Oui. Trois mois et un débat public plus tard, le Non l'emportait avec près de 55% des voix.

En 2016, presque tout le monde s'attendait à une défaite de Donald Trump à l'élection présidentielle étasunienne. « Les sondages » l'avaient affirmé. On connaît la suite.

La même année, en France, F. Fillon était situé en septembre à moins de 10 % dans les primaires de la droite. Il a finalement obtenu 44 % au premier tour en novembre et 66 % au second. Devenu alors, toujours selon les sondages d'intentions de vote, vainqueur presque certain de la présidentielle de 2017, il n'atteignait finalement même pas le second tour.

On pourrait multiplier les exemples d'échecs. Chacun d'entre eux appelle une analyse spécifique. La dynamique du débat public, et l'intérêt tardif d'une partie des citoyens, dans le cas de 2005. Les particularités du scrutin, la réticence de certaines populations à répondre et apparemment des erreurs d'analyse des chiffres en 2016 aux États-Unis. La dynamique des campagnes, là aussi, et sans doute l'existence d'un « biais de sélection » dans le cas de la primaire en France. Il y avait en effet un biais de sélection dans la mesure où les votants à la primaire étant apparemment plus sensibles aux arguments de F. Fillon que la population plus large, celle des électeurs de ce camp, qui avait été sondée. Quant à sa défaite en 2017, elle prouve que des événements inattendus peuvent modifier les intentions de votes.

Mais certaines réussites sont également remarquables.

La plus célèbre, qui est même souvent présentée comme fondatrice, se situe le 3 novembre 1936. Le grand magazine *Literary Digest* avait organisé une consultation auprès de ses millions de lecteurs. Il en résultait une nette victoire d'Alfred Landon. Pourtant F.D. Roosevelt a été réélu président des États-Unis. La société *Gallup*, à l'aide d'un sondage auprès de trois mille personnes, avait annoncé le bon résultat avec une assez bonne estimation du pourcentage obtenu.

C'est là une illustration de ce que la quantité n'est pas la qualité : **un bon sondage vaut mieux qu'un mauvais recensement**. Les lecteurs du LD, ou ceux d'entre eux qui ont répondu à ce « vote de paille », ne votaient pas comme l'électorat dans son ensemble. Le LD espérait-il influencer les suffrages ? Il n'a pas survécu au discrédit. Et les « sondages d'opinion » triomphent depuis. Au risque de susciter la rébellion, justement, de « l'opinion » qu'on prétend observer et qui se sent, parfois à juste titre, manipulée. Et avec périodiquement des « accidents » : Gallup lui-même s'est trompé en 1946 en annonçant la défaite de Truman à l'élection présidentielle US. Biais de sélection, là aussi : il avait utilisé l'annuaire téléphonique comme base de sondage. Et, on l'a su plus tard, les abonnés au téléphone ne votaient pas comme les autres citoyens.

Une histoire émaillée de controverses scientifiques et des pratiques diverses

Bien avant 1936, des statisticiens ont envisagé de n'observer qu'une partie de la population étudiée et de tenter d'en tirer des enseignements valables pour l'ensemble. Cela permet des économies et s'impose parfois à l'évidence. Par exemple pour prévoir les récoltes de blé avant la moisson, en ne prélevant que quelques épis « bien choisis ». Ou encore pour calculer un indice des prix à la consommation sans pouvoir observer tous les achats.

On parle alors d'échantillonnages ou de sondages. D'où d'âpres débats : comment sélectionner les échantillons ? Comment évaluer la qualité de la quantification obtenue ? Est-on proche de la réalité – quelle réalité ? Le débat se déploie depuis plus d'un siècle.

Il y a deux grandes manières de procéder à un échantillonnage, avec de nombreuses variantes, Les sondages aléatoires et les sondages par quotas.

Pour mettre en œuvre la première **méthode, dite de sondages aléatoires**, il faut disposer d'une bonne connaissance de la population étudiée dans son ensemble. On peut alors tirer des unités au hasard dans la population, ou dans une « base de sondage » qui la décrit correctement. La théorie des sondages, qui est une branche des mathématiques désormais bien développée, permet alors de définir des marges d'imprécision, ou marges d'erreurs, qu'on appelle aussi « intervalles de confiance » ou fourchettes.

Illustration : quand l'Insee annonce 20.000 habitants pour une commune à la suite d'un sondage, sa population est en fait cernée à 1,5 % près ; elle se situe donc probablement entre 19.700 et 20.300. Le hasard peut être capricieux, mais on peut quand même en chiffrer le risque : il y a 95 % de chances pour que la population de cette commune appartienne vraiment à cette fourchette. On a obtenu, grâce à ce sondage aléatoire, un bon ordre de grandeur de la population de cette commune. Il y a un risque infime qu'elle soit en réalité de 18.000 ou de 22.000... Sauf si la base de sondage (les logements) était vraiment mauvaise ou si on a commis de grosses erreurs d'observation (on y reviendra).

La « population » peut aussi être, par exemple, un ensemble d'épis de blé. On sait assez bien, dès le printemps, où il en a été semé. On dispose donc d'une base de sondage correcte. Une sélection, c'est le « plan de sondage », peut alors permettre de tirer des enseignements valables pour l'ensemble. Et d'obtenir une prévision de récolte convenable... sauf bien sûr accident climatique.

Mais, en général, on ne dispose pas d'un fichier de la population ou d'une base de sondage fiable. On utilise

alors des **méthodes empiriques d'échantillonnage**. Les plus connues sont les enquêtes par quotas comme celles qui sont pratiquées par les sociétés de sondages privées. On connaît assez bien le nombre de femmes de 40 à 50 ans dans la population (c'est à peu près 6,4 %). Dans un sondage auprès de 1.000 personnes, on cherchera alors à avoir 64 femmes de cette tranche d'âge. Si on n'en a que 60, on corrigera cette légère sous-représentation : c'est une forme de « redressement » des résultats bruts du sondage. Idem pour les hommes de plus de 60 ans etc. En toute rigueur, on ne peut en déduire des marges d'erreurs, comme dans les sondages aléatoires, mais on fait souvent comme si...

Des résultats d'élections connus avant 20 h, sauf si...

Les citoyens se demandent souvent, à juste titre, comment on peut annoncer à 20 H un résultat alors que le scrutin n'est pas clos dans les grandes villes ? L'électeur de grande ville, dont le bulletin n'a pas encore pu être dépouillé à 20 H, peut avoir l'impression de compter pour du beurre. On pourra le rassurer mais examinons rapidement les sondages sur lesquels se fondent ces estimations.

On peut appliquer ici la méthode des sondages de deux manières car on peut bien connaître ou approcher deux « populations » : l'ensemble des bulletins de vote d'une part ; les électeurs qui ont voté d'autre part.

On peut donc tirer au hasard, selon un « plan de sondage » bien déterminé, quelques bulletins de vote au fur et à mesure des ouvertures d'urnes. Et en tirer progressivement des enseignements pour l'ensemble du scrutin.

Mais cette méthode n'est pas sans risques. Par exemple si le vote des campagnes et celui des villes évoluent de manière différente. En outre, elle ne permet pas d'avoir une idée du vote des jeunes, des ouvriers, etc. puisque ce n'est pas indiqué sur le bulletin de vote.

On recourt pour ceci à des sondages auprès des électeurs eux-mêmes, à la sortie des bureaux de vote. Ce n'est pas un « sondage d'opinion » puisqu'on n'observe pas une intention de vote mais le vote lui-même. On quantifie ici des faits et non des intentions. Ici, la « base de sondage » est bien connue. On peut donc en principe maîtriser la marge d'erreur... à condition bien sûr que les sondés déclarent sincèrement ce qu'ils viennent de voter !

Comment rassurer l'électeur de grande ville ? Si l'élection se jouait à une voix près, les sondages toucheraient leurs limites car on ne pourrait se prononcer avant le dépouillement des derniers bulletins, dont le sien.

Mais ajoutons aussi que, dans certains pays où la sincérité du scrutin est très contestable, des sondages auprès des électeurs peuvent même être considérés comme meilleurs que le résultat officiel !

Et les erreurs d'observation ?

Mais, dans toutes les configurations, les statisticiens font face à des adversaires redoutables : les erreurs d'observation. Dans le cas du sondage à la sortie des bureaux de vote, tous les électeurs n'acceptent pas de répondre (ou ne déclarent pas forcément leur véritable vote). Si ce « biais de sélection » est important, les calculs de précision qu'autorise la théorie des sondages seront impuissants.

On peut alors, parfois, procéder à divers « redressements » : on tente alors de corriger les résultats bruts en utilisant des informations supposées fiables dont on dispose par ailleurs. Cela se complique un peu, mais on peut ainsi améliorer les estimations.

Par exemple : on sait que les petites entreprises répondent moins volontiers aux enquêtes statistiques que les moyennes ou grandes. Admettons que la moitié des PME aient refusé de répondre à une enquête : on pourra envisager de compter en double celles qui ont répondu. Mais, en agissant ainsi, on suppose que les PME qui n'ont pas répondu auraient fourni des indications similaires à celles qui ont répondu.

D'où un risque d'imprécision résiduelle qui ne peut pas toujours être diminué par d'autres moyens.

Le cas des sondages par internet en fournit une illustration : peut-on considérer qu'un internaute fournirait le même type de réponse qu'une autre personne de même âge, de même catégorie socioprofessionnelle, de même sexe habitant la même ville... ?

Tous ces redressements sont souvent nécessaires mais parfois acrobatiques et éventuellement insuffisants. La clarté sur les méthodes employées serait nécessaire pour dissiper l'impression d'opacité voire de magouilles qui en résulte. Mais certains statisticiens considèrent que « le profane ne peut pas comprendre » ou qu'il s'agit là de « secrets de fabrication des chiffres » qu'ils ne veulent pas livrer à la concurrence et à la critique. Le secret des affaires est caractéristique du secteur privé commercial mais, en principe, les statisticiens publics ne devraient pas (dé)raisonner ainsi...

Et les sondages d'opinion : une entreprise de manipulation des citoyens ?

Dans les paragraphes précédents, on a surtout envisagé l'observation des faits (bulletins remis dans les urnes, votes qu'on vient d'effectuer, épis de blé à sélectionner).

Imprécision inhérente à l'échantillonnage, erreurs d'observation, flou des définitions : les sondages qui portent sur des objets concrets ou sur des faits sont entachés d'inexactitudes plus ou moins bien maîtrisées.

Cela se corse quand on prétend évaluer des intentions (de votes, d'achats, de voyages...)

Ceux qui prétendent analyser « l'opinion » ou les opinions ajoutent à ces difficultés le caractère fluctuant de la matière observée. On l'a vu : **un débat public peut faire évoluer l'opinion. C'est même normalement son objectif !**

Des sociologues comme Pierre Bourdieu ou Patrick Champagne ont mis en évidence le caractère pour le moins douteux de questions posées à des gens qui ne se les posent pas (ou pas encore). Il ne s'agit plus d'observer une opinion mais de l'influencer et à la limite de la fabriquer, délibérément ou involontairement.

Et que dire d'une étude de marché qui porte sur des produits qui n'existent pas encore ? C'est pourtant la principale activité des sociétés de sondages privées. Un exemple cuisant a affecté la Ford Edsel dans les années 50. À coup de sondages, on avait prétendument défini la voiture idéale. Ce fut un échec commercial total.

Il s'y ajoute des biais intrinsèques à ces sondages commerciaux : le commanditaire peut refuser de publier les résultats s'ils ne lui conviennent pas. L'honnêteté et le professionnalisme de ces sociétés n'est pas en cause mais elles peuvent aussi avoir tendance à « tordre » leurs enquêtes par divers moyens afin de fournir à leurs clients des résultats qui les satisfont et ne pas risquer de perdre des marchés au profit de concurrents moins scrupuleux. Et les moyens de manipulation sont bien connus des spécialistes : les réponses sont notamment très sensibles au libellé des questions et même à l'ordre ou au contexte dans lequel elles sont posées.

En ce qui concerne la statistique publique (Insee, services statistiques ministériels), c'est une des missions du conseil national de l'information statistique que de veiller à l'opportunité et à la sincérité de ses enquêtes ainsi qu'à la mise à disposition des résultats pour le public et les chercheurs.

Il n'est certes pas impossible de scruter certaines opinions et d'utiliser la technique des sondages pour ce faire. Les enquêtes de la statistique publique auprès des entreprises et des ménages y recourent largement et peuvent contribuer à éclairer le pays sur la conjoncture économique. On peut aussi dresser un portrait social intéressant du pays sans avoir à interroger individuellement tous les habitants. Même le recensement de la population a renoncé à l'exhaustivité : il utilise largement, désormais, la technique des sondages.

Quelles utilisations syndicales de ces sondages ?

A l'automne 1986, un sondage auprès des étudiants, publié par Le Monde, concluait que ceux-ci n'avaient que des préoccupations matérielles et n'étaient « donc » nullement prêts à se mobiliser. En fait, on l'a su après, le mouvement qui allait aboutir au retrait de la loi Devaquet était déjà amorcé au moment où ce sondage a été publié. Mais ce mouvement n'était pas « statistiquement significatif ». Les sondages pré-électorales posent des problèmes similaires, on l'a vu. Ils deviennent même nocifs, d'un point de vue démocratique, quand ils conduisent à sélectionner les candidats... et à occulter les débats de fond.

Faut-il mettre tous les sondages dans le même sac et jeter ce sac ? Non. Ils offrent des moyens de connaissance qu'on aurait tort de rejeter sans nuances. Mais les sondages en général et surtout ceux qui prétendent quantifier la volonté et les « intentions » de la population doivent être l'objet d'une vigilance particulière de la part des citoyens et de leurs représentants politiques, associatifs et syndicaux. À défaut de quoi les manipulateurs de l'opinion disposeront d'une arme supplémentaire.

Et, après tout, pourquoi ne pas l'utiliser cette technique ? Les associations et les syndicats utilisent depuis longtemps des méthodes de consultation des populations dont l'opinion les intéresse particulièrement. Il s'agit à la fois d'observer des réalités, de tenter de discerner des intentions mais aussi de forger des outils de mobilisation. La technique des sondages peut y aider en évitant de se noyer dans des tentatives démesurées de consulter tout le monde mais en sélectionnant des échantillons bien choisis. C'est par exemple ce qu'a fait l'UGICT, avec la coopération de statisticiens professionnels, pour mettre en évidence les conditions dégradées dont le télétravail s'est développé en 2020. L'utilisation militante des stats en général, et de la technique des sondages en particulier, porte même désormais un nom : le stactactivisme, pour qui « un autre nombre est possible ».

Montreuil, le 17/03/2022

Pour plus de détails :

Revue de l'Institut d'Histoire sociale CGT n°100 décembre 2006 Les sondages, démocratie d'opinion

Analyses et documents économiques CGT n°123 de Janvier 2018. Les sondages nous trompent-ils toujours ?

Introduction à la théorie des sondages (Insee 2018)

<http://www.nc233.com/wp-content/uploads/2018/01/sondages-amphi-1.pdf>

La précision des résultats du recensement

Conseils pour l'utilisation des résultats statistiques (fiche thématique relativement simple)

<https://www.insee.fr/fr/information/2383177>

.. Et article plus approfondi :

<https://www.insee.fr/fr/statistiques/2017638>

Patrick Champagne « Faire l'opinion. Le nouveau jeu politique »

<https://www.cultura.com/faire-l-opinion-tea-9782707328793.html>

Site du conseil national de l'information statistique :

<https://www.cnis.fr/>

Statactivisme. Comment lutter avec les nombres

www.editionsladecouverte.fr/statactivisme-9782355220548