

## Pour voir mieux voir peu

« *Un bon sondage vaut mieux qu'un mauvais recensement* ». L'affirmation mérite quelques explications. Dans l'histoire, les statisticiens, qui sont des gens méticuleux, ont longtemps résisté à l'emploi de la méthode des sondages dans leurs travaux. Examinons deux cas de figure extrêmes, autour d'applications d'actualité :

### **Les sondages électoraux**

S'ils sont réalisés avec soin, quelques milliers d'interviews cernent avec une précision étonnante la réalité électorale, qu'il s'agisse des sondages effectués peu de temps avant le vote et des sondages dits de « *sortie des urnes* » réalisés à l'instant où les électeurs viennent de voter<sup>1</sup>. Ce ne sont pas des arguments théoriques – les fondements du calcul de probabilités remontent XVII<sup>ème</sup> siècle - qui ont convaincu de l'intérêt des sondages, c'est plutôt un fait médiatique, l'élection américaine de 1936 (sondage Gallup notamment) : une confrontation entre un sondage effectué sur une base solide, représentative du corps électoral, et une base biaisée, fondée sur le lectorat nombreux d'un journal.

Tout résultat statistique est entaché d'une incertitude. Dans tous les cas, une incertitude liée à la qualité plus ou moins bonne de la collecte. Dans le cas des sondages, s'ajoute l'incertitude propre aux enquêtes sur échantillon, mais celle-ci est mesurable. Un bon sondage pourra être préféré à un recensement exhaustif car il coûte le plus souvent beaucoup moins cher.

### **Les recensements de la population**

Longtemps la France a utilisé la méthode du recensement exhaustif pour compter sa population : par définition, tout le monde doit être observé. Mais pour recenser 65 millions d'habitants selon la méthode traditionnelle, il faut mettre en place une organisation complexe pour un temps de collecte sur le terrain limité à quelques jours. A chaque recensement, soit une fois tous les 7 ou 8 ans, il fallait former environ 100 000 enquêteurs ; la mémoire collective se perdait d'une collecte à l'autre. On ne pouvait éviter des défaillances dans la collecte ici ou là, malgré tous les efforts pour en assurer le contrôle.

C'est pourquoi la France a décidé d'appliquer la méthode des sondages<sup>1</sup> pour compter sa population : par facilité de langage on dira que désormais le recensement se fait par sondage.

---

<sup>1</sup>Une observation annuelle de toute la population (donc un « recensement » au sens classique) habitant les communes de moins de 10000 habitants tirées au sort chaque année à raison de 1 commune sur 5.

Par conséquent, aujourd'hui, le terme recensement renvoie plutôt à l'opération administrative qu'à la méthode. La perspective d'une telle « révolution » a longtemps effrayé les utilisateurs traditionnels du recensement, les universitaires surtout. Qu'en est-il aujourd'hui, maintenant que la première série des cinq campagnes annuelles successives s'est déroulée sans incident ? Les critiques sont moins audibles aujourd'hui, bien que l'on entende nombre de praticiens du terrain se plaindre de certaines faiblesses, notamment dans la mesure des flux migratoires<sup>2</sup>.

Lorsque le journal *Le Monde* (1<sup>er</sup> août 2012) s'est fait l'écho de certaines critiques venues des maires qui voyaient la population légale de leur commune évoluer en dents de scie d'une année à l'autre<sup>3</sup>, l'Insee dans un communiqué s'est justifié, en arguant que l'incertitude de la mesure était de 0,02%<sup>4</sup>. Cette affirmation, assurément optimiste, mériterait d'être mise en débat. En effet, que faut-il en penser ?

- 1) Affirmer que la précision des résultats est de 0,02% (soit une incertitude de plus ou de moins 15000 personnes pour l'ensemble du pays) peut apparaître présomptueux. Ce pourcentage ne concerne que la précision théorique qui dépend du plan de sondage. Sa détermination rigoureuse est difficile, mais on n'a pas a priori à remettre en cause un résultat scientifique : on fait volontiers crédit aux théoriciens sur leur capacité à mener à bien leurs calculs. En revanche, ce qu'il aurait fallu dire préciser, c'est que ce pourcentage n'est valable *que pour l'ensemble de la France*. Mais pour une ville donnée, la précision est bien moindre. Elle dépend de sa taille. Ainsi pour une commune de 10 000 habitants, qui ne représente que un sur 625 de la population totale, l'incertitude sur le chiffre de population totale serait, en pourcentage, approximativement 25 fois plus grande, « doigt mouillé » de l'ordre de 0,5%<sup>5</sup>. Ce n'est qu'un ordre de grandeur. Une telle incertitude est tout à fait acceptable si l'on s'intéresse à la population totale, mais il faut savoir qu'elle sera beaucoup plus grande s'il s'agit d'estimer des effectifs par catégories : sexe, âge et condition de logement par exemple.

Pour les communes de moins de 10000 habitants, interrogées de façon exhaustive, la qualité est meilleure qu'auparavant à la fois parce que la collecte est plus fréquente, mais aussi parce que les données sont mises à jour chaque année par des informations administratives.

---

*Une observation par échantillon (tirage au sort à raison de 8% des logements) pour les communes de plus de 10000 habitants.*

*Au bout de 5 ans, 70% de la population se trouve donc recensée. C'est en principe suffisant pour répondre aux besoins de connaissance de la société. Noter que le coût global ne change pas (mieux, il est réparti sur 5 ans, ce qui est plus favorable à une gestion optimale des finances publiques).*

<sup>2</sup> Laurent Chalard, géographe : <http://lecerclerlesechos.fr/economie-societe/politique-eco-conjoncture/territoires/221188613/recensement-population-francaise-a->

<sup>4</sup> La réponse du Monde : [http://www.lemonde.fr/idees/article/2012/08/13/population-les-calculs-de-l-insee-sont-fiabiles\\_1745688\\_3232.html](http://www.lemonde.fr/idees/article/2012/08/13/population-les-calculs-de-l-insee-sont-fiabiles_1745688_3232.html)

- 2) En revanche on ne dit rien sur les erreurs d'observation : omissions, refus de réponse, doubles comptes, déclarations erronées.... Il ressort de divers témoignages que certaines situations rendent la collecte difficile. Quel est alors l'impact réel sur la qualité des résultats ? On ne le sait pas. En effet, l'Insee ne communique pas sur ce sujet et c'est regrettable puisque les inquiétudes, légitimes, s'alimentent du silence de l'Institution.

On est cependant fondé à imaginer que la qualité de la collecte est meilleure aujourd'hui que ce qu'elle était hier, à la fois parce que le système de collecte repose sur une base d'immeubles, le RIL (Répertoire d'immeubles localisés) et parce qu'une partie importante des enquêteurs gardent d'une année à l'autre la mémoire de leur expérience.

<sup>i</sup> Les élections présidentielles de 2012. Les écarts entre les sondages (moyenne des 8 instituts) et les résultats de l'élection, à différentes dates :

candidat	20 avril	... J - 3	... J - 7	... J - 15	... J - 30	... J - 90
Mélenchon	13,8	-0,1	-0,4	-0,4	1,4	5,7
Hollande	28,1	0,2	0,5	0,2	-0,2	-1,3
Joly	2,5	0,1	0,1	0,3	0,2	-0,5
Bayrou	10,1	-0,2	-0,1	-0,3	-1,9	-3,2
Sarkozy	26,3	-0,2	-0,6	-1,6	-1,7	1,9
Le Pen	16,0	0,2	0,4	1,0	1,1	-1,3
autres	3,3	0,1	0,2	0,9	1,1	-1,3

Mode de lecture : Hollande : 28,1 % des voix au premier tour, Sarkozy, 26,3% En regardant la série des sondages, on constate que depuis au moins un mois, la moyenne des 8 instituts ne s'est jamais écartée de plus d'un demi-point pour Hollande, ce qui n'est pas le cas pour Sarkozy : son électorat a sans doute été plus difficile à cerner.

Ce tableau est extrait d'un document mis en ligne sur le site de l'APR :

<http://www.apr-strasbourg.org/membre/2054/aubry-bernard/document/588.html>

A consulter également :

<http://www.apr-strasbourg.org/membre/2054/aubry-bernard/document/587.html>

