REVUE DE STATISTIQUE APPLIQUÉE

J. DESABIE

Méthodes empiriques d'echantillonnage

Revue de statistique appliquée, tome 11, n° 1 (1963), p. 5-24

<http://www.numdam.org/item?id=RSA_1963__11_1_5_0>

MÉTHODES EMPIRIQUES D'ECHANTILLONNAGE

J, DESABIE
Administrateur à l'Institut National de la Statistique
et des Études Économiques

I - SONDAGES PAR "CHOIX RAISONNE"

Parmi les procédés qui viennent à l'esprit lorsqu'on se propose de représenter un ensemble par un échantillon des unités qui le constituent, le plus naturel, et de beaucoup, consiste à construire un échantillon qui ressemble à la population dont il est issu, qui en soit, au sens courant du mot : représentatif.

La désignation de l'échantillon résultera d'un choix "raisonné" d'où le nom de la méthode ("puposive Sampling" en anglais(*)).

Plusieurs techniques procèdent de cette idée. Nous parlerons de deux d'entre elles, d'importance fort inégale : la méthode des unités-types et la méthode des quota.

I.I - La méthodes des unités-types

I.I.I - Principe

Cette méthode repose sur l'idée que les différentes variables attachées à un individu n'étant pas indépendantes entre elles, un individu qui se situe dans la moyenne de la population pour un certain nombre de caractères importants sera également peu différent de la moyenne de la population en ce qui concerne les autres caractères.

La méthode consiste donc à diviser la population en un certain nombre de sous-ensembles relativement homogènes et à représenter chacun de ces sous-ensembles par une "unité-type".

Le choix des unités-types peut être plus ou moins systématisé :

On peut :

(*) Nous préférons parler de "sondages par choix raisonné" plutôt que de "sondages représentatifs", comme il est dit quelquefois, l'adjectif représentatif étant employé avec plusieurs sens différents en théorie des sondages. Il est d'ailleurs particulièrement malheureux d'opposer "représentatif" à "aléatoire" puisqu'en toute rigueur, la désignation par tirage au sort est la seule qui assure le caractère représentatif de l'échantillon.

a) se fier entièrement au jugement du statisticien ou d'un expert consulté, ou

b) retenir un certain nombre de variables importantes, calculer les valeurs moyennes par unité pour chacune de ces variables et choisir une unité pour laquelle les valeurs de la variable s'écartent peu des valeurs moyennes. Cette deuxième méthode n'est pas nécessairement la meilleure. Elle comporte, elle aussi, une forte part d'arbitraire : choix des variables contrôlées, importance relative attachée à chacune d'entre elles ; en effet, il se peut fort bien qu'aucune unité ne satisfasse aux conditions énoncées plus haut ; il faudra alors rechercher l'unité la plus "proche" du point moyen - ce qui est un très délicat problème statistique.

c) en pratique, on adoptera une méthode intermédiaire.

Les unités d'une sous-population homogène étant généralement peu nombreuses (cf. conclusion), on établira une fiche pour chacune d'elles ; sur cette fiche, on notera les valeurs des variables essentielles et quelques autres renseignements ; l'unité-type sera choisie de manière assez empirique.

Exemple : les "cantons-type" de l'I.N.S.E.E.

En 1942, on procéda - sous la Direction de l'I.N.S.E.E. - à un découpage de la France en 600 régions agricoles environ.

Dans chaque région agricole - supposée homogène par définition - fut désigné un "canton-type". L'idée était intéressante, puisqu'elle permettait d'établir les statistiques agricoles courantes pour l'ensemble de la France - et par région, tout en réduisant le coût de collecte dans le rapport de 5 à 1 (il y a en tout environ 3 000 cantons en France).

Malheureusement cette expérience n'ayant pas été poussée à fond, il n'est pas possible d'en tirer tout l'enseignement qu'elle aurait pu comporter. Il semble toutefois que les experts agricoles régionaux aient assez fréquemment cédé à la tentation d'"accentuer les constrastes" - en désignant dans une région boisée un canton encore plus boisé ; dans une région viticole un canton encore plus viticole.

I.I.2 - Avantages et inconvénients de la méthode

La méthode des unités-type comporte une part d'arbitraire que rien ne saurait éliminer ; le choix d'une unité-type parfaite impliquerait une connaissance parfaite de la population et d'ailleurs l'existence d'une unité-type parfaite n'est pas du tout certaine : un canton-type peut être très représentatif de sa région en ce qui concerne la répartition du sol entre les cultures et être tout-à-fait aberrant en ce qui concerne les modes de faire-valoir pratiqués, les engrais utilisés ; croire le contraire serait admettre un déterminisme des plus grossiers.

La méthode doit donc être employée avec une extrême prudence ou par des personnes ayant une connaissance approfondie de la population étudiée ; le choix des unités-type doit être repensé avant chaque étude.

Ces réserves étant faites, la méthode des unités-type est susceptible de rendre de grands services lorsque, ne disposant pas de moyens matériels importants, on désire néanmoins obtenir des renseignements pour des sous-groupes relativement petits de la population. La méthode offre, en effet, des

avantages certains lorsque les extrapolations doivent être réalisées à partir d'un échantillon de très faible effectif(*).

L'exemple donné plus haut est, à cet égard, très instructif. Les statistiques agricoles doivent être publiées avec un détail géographique assez grand - ceci est une nécessité absolue. Il n'est pas douteux que pour chacune des 600 régions agricoles, les estimations fournies par le canton-type ont moins de chance de s'écarter fortement de la vérité que les estimations qui auraient pu être fournies par un échantillon aléatoire d'un seul canton.

Il n'en est pas moins vrai d'ailleurs qu'un échantillon aléatoire de 600 cantons (un par région) donnerait sans doute pour l'ensemble de la France de meilleures estimations que l'échantillon de 600 cantons-types. Ces deux propositions ne sont nullement contradictoires, ce qui mérite d'être médité comme illustration de la loi des grands nombres, fondement des sondages probabilistes : "l'erreur d'échantillonnage d'un sondage probabiliste tend vers zéro lorsqu'augmente la taille de l'échantillon" - propriété que les sondages aléatoires ne partagent avec aucun autre.

I.2 - La méthodes des "quota".

I.2.I - Principe de la méthode

La méthode des quota repose sur la proposition suivante : "les différents caractères que l'on peut observer dans une population n'étant pas indépendants entre eux, un échantillon identique à la population dans laquelle il est prélevé en ce qui concerne la distribution statistique de certains caractères importants sera également peu différent de la population, en ce qui concerne la distribution statistique des caractères qui ne sont pas contrôlés".

La méthode implique donc une bonne connaissance statistique de la population étudié - préalable à l'enquête.

On subdivise la population en classes, les statistiques font connaître l'effectif de chacune d'entre elles. Ces effectifs, multipliés par le taux de sondage, donnent "les quota" qui devront être respectés ; ainsi - prenons l'exemple d'une étude de la population des marseillais âgés de 16 ans ou plus - le Recensement de 1954 donne la répartition de cette population par sexe, classe d'âge et milieu social (d'après la profession du chef de ménage). L'effectif de chaque classe - en milliers - est donné dans le tableau cidessous :

En supposant que l'on désire un échantillon de 500 personnes environ (taux de sondage f = 1/1 000), le tableau ci-dessus donne également les "quota" de chaque classe.

Ces quota sont imposés aux enquêteurs - ainsi un enquêteur ayant reçu 50 questionnaires à remplir recevra le tableau de contrôle donné ci-dessous.

Le choix des individus-échantillons est - par ailleurs - laissé à l'initiative des enquêteurs auxquels on impose cependans parfois quelques restrictions supplémentaires : dispersion géographique, interdiction d'interro-

---

(*) Elle peut en particulier servir à désigner une ville pour y procéder à un essai dont les résultats seront extrapolés à une région.

ger des personnes se connaissant entre elles ou connaissant l'enquêteur ; ou encore d'interroger les passants dans la rue.

Sexe	Age	Milieu Social(*)
Masculin 236	16 à 24 ans 72	Patrons..... 75
	25 à 44 ans 187	Cadres supérieurs..... 31
Féminin 274	45 à 64 ans 178	Cadres moyens et employés . 83
	65 ans et plus 73	Ouvriers..... 200
		Rentiers-Retraités..... 121
510	510	510

Sexe	Age	Milieu social
M : 23	16 à 24 ans : 7	Patrons..... 7
F : 27	25 à 44 ans : 18	Cadres supérieurs..... 3
	45 à 64 ans : 18	Cadres moyens et employés . 6
	65 et plus : 7	Ouvriers..... 20
		Rentiers, retraités..... 12
50	50	50

I.2.2 - Mise en œuvre des principes

I.2.2.I - Choix des variables de contrôle.

Pour mériter d'être retenue comme "contrôle", une variable quantitative ou qualitative doit présenter simultanément les avantages suivants :

1/ Sa distribution statistique dans la population est bien connue.

2/ Son observation par les enquêteurs sur le terrain est facile et ne comporte pas de sérieux risques d'erreurs.

3/ Une corrélation étroite existe entre la variable contrôlée et la (ou les) variables(s) étudiée (s).

Les deux premières conditions doivent être remplies pour qu'il soit possible d'appliquer la méthode ; la troisième condition doit être remplie pour que la méthode soit efficece.

Les contrôles proposés en exemple présentent simultanément ces trois qualités - il n'en serait pas de même du revenu par exemple (lequel ne remplit pas les conditions 1 et 2) ; ni des catégories sociales A, B, C, D - fré-

---

(*) D'après la catégorie socio-professionnelle du chef de ménage.

quemment utilisées par les maisons d'études de marché, mais dont on voit mal comment elles sont définies, et sur lesquelles il n'existe aucune statistique officielle.

Ces conditions étant extrêmement restrictives, le choix des variables de contrôle est étroitement limité.

Exemple : pour un échantillon de personnes :

Région - habitat (d'après la population de la commune de résidence).

Sexe - age - catégorie socio-professionnelle(*).

Pour un échantillon de ménages :

Région - habitat - catégorie socio-professionnelle du chef de ménage - effectif du ménage.

Pour un échantillon de points de vente :

Région - catégorie de commune - nature de l'activité-statut juridique - nombre de salariés.

Ces mêmes caractères sont utilisés dans les sondages aléatoires stratifiés.

Bien entendu, la liste que nous avons donnée n'est en rien limitative. Pour certaines enquêtes, il pourra être très important de distinguer les ménages vivant dans un logement neuf, ou vivant en meublé ; - ou dont la femme travaille au dehors - on introduira comme contrôle les caractères correspondants.

Presque toujours plusieurs caractères sont simultanément retenus comme contrôles ; on fera en sorte que les différents caractères contrôlés soient aussi indépendants que possible.

Ainsi, on ne retiendra pas simultanément la catégorie socio-professionnelle individuelle et la catégorie socio-professionnelle du chef de ménage - il existe, en effet, une étroite corrélation entre ces deux caractères.

I.2.2.2 - Contrôles marginaux ou croisés

On se reportera utilement aux études sur la stratification.

a - Contrôles marginaux ou indépendants

On contrôle la distribution de chacune des variables séparément. Ainsi, pour en revenir à notre exemple d'une étude sur la population marseillaise adulte, on contrôle les distributions suivant le sexe, l'âge, le milieu social indépendamment les unes des autres. Ces distributions constituent les marges de la distribution à trois dimensions par sexe, âge et milieu social.

(*) Pour certaines enquêtes, on retiendra la catégorie socio-professionnelle individuelle de la personne interrogée, déterminée par la profession qu'elle exerce. Dans d'autres enquêtes, on retiendra la catégorie socio-professionnelle du chef de ménage. Dans le premier cas, une femme au foyer sera classée inactive ; dans le second cas, elle sera classée d'après la profession de son mari.

b - Contrôles croisés

Pour reprendre le même exemple, il aurait été concevable de contrôler la distribution de l'échantillon suivant ces trois variables conjointement. On fixerait, par exemple, le quota des femmes âgées de 25 à 44 ans appartenant au milieu "ouvrier".

Il serait nécessaire de disposer pour la ville de Marseille du tableau statistique à trois entrées : sexe × âge × CSC

2 × 4 × 5 = 40 cases

40 contrôles ou quota seraient donc imposés - chaque enquêteur recevrait un tableau de contrôle de la forme :

Milieu Social (CSC)	16 à 24 ans		25 à 44 ans		45 à 64 ans		65 ans et plus	
	M	F	M	F	M	F	M	F
Patrons								
Cadres supérieurs								
Cadres moyens et employés								
Ouvriers								
Rentiers - Retraités								

Cette solution - théoriquement meilleure - serait impraticable sous cette forme : elle exigerait notamment des statistiques trop détaillées généralement non disponibles et imposerait aux enquêteurs des contrôles trop stricts qu'ils respecteraient mal - voir plus loin : p. 21.

En pratique, il serait possible de retenir une solution intermédaire en contrôlant la distribution par sexe et âge d'une part, par sexe et milieu social d'autre part.

En notation symbolique S x (A + CSC).

I.2.2.3 - Echantillons à plusieurs degrés

Pour exposer le principe de la méthode, nous avons volontairement choisi l'exemple particulièrement simple d'une étude portant sur une seule localité.

En pratique, il n'en est généralement pas ainsi ; le domaine d'étude sera la France entière ou une région et comportera de nombreuses localités d'importance très inégale.

En pratique (et ceci serait également vrai de la méthode aléatoire), on procède généralement à un sondage à deux degrés. On désigne un échantillon de localités ; dans les localités-échantillon, un échantillon de personnes, de ménages, de points de vente, d'exploitations agricoles.

Le choix des localités - échantillon est d'une importance considérable. On répartit les 38 000 communes du territoire français en un certain nombre de "Strates", une strate comprenant toutes les communes de même catégorie appartenant à une même région.

La catégorie de commune est définie par le nombre d'habitants qui y ont été recensés - voir ci-dessous ; (les communes situées dans la banlieue d'une grande ville constituant avec celle-ci une agglomération traitée comme une unité).

Ainsi - en adoptant les classifications I.N.S.E.E.(*) on pourrait distinguer distinguer dans chacune des huit grandes régions sondages.

- les communes rurales (population municipale agglomérée au chef-lieu inférieure à 2 000 habitants)
- les petites villes : 2 à 10 000 habitants
- les villes ou agglomérations de : 10 à 50 000 habitants
- les villes ou agglomérations de : 50 à 100 000 habitants

définissant ainsi 8 4 = 32 strates, dans chacune desquelles on désignerait un échantillon à deux degrés.

Chacune des 25 agglomérations de plus de 100 000 habitants serait incorporée d'office à l'échantillon et pourrait donc être considérée comme constituant une strate à elle seule.

Poussons un peu plus loin pour faire apparaître certains problèmes intéressants.

Reprenons l'exemple d'un sondage sur la population adulte : (16 ans et plus) mais étendu à l'ensemble de la population française ;

le taux de sondage adopté sera plus faible : $f = \frac{1}{5\,000}$ par exemple.

Dans le Midi méditerranéen : région composée des départements suivants : Hautes-Alpes, Basses-Alpes, Alpes-Maritimes, Var, Bouches-du-Rhône, Vaucluse, Gard, Drôme, Ardècge, Hérault, Aude, Pyrénées-Orientales.

Les agglomérations de Toulon, Marseille, Nice qui comptent chacune plus de 1000 000 habitants seront toutes trois retenues dans l'échantillon - les quota par sexe, âge, milieu social seront calculés pour chacune d'elles - il n'y a pas de difficultés spéciales. L'exemple donné plus haut - pour Marseille - apparaît maintenant comme un fragment d'un ensemble. Pour obtenir les quota de la ville de Marseille, il suffit de diviser par 5 les nombres du tableau de la page 8 , les tableaux de contrôle remis aux enquêteurs n'étant d'ailleurs pas modifiés.

Considérons maintenant la strate des villes ou agglomérations de 50 à 100 000 habitants : il y en a sept :

Aix-en-Provence, Avignon, Cannes, Nîmes, Montpellier, Béziers, Perpignan.

(*) Et en les simplifiant un peu.

L'effectif de l'échantillon qui doit représenter ces sept villes ou agglomérations est égal au nombre de personnes de 16 ans ou plus recensées dans ces sept villes ou agglomérations multiplié par le taux de sondage $\frac{1}{5\,000}$, soit approximativement 80. Mais comment doit-on contrôler cet échantillon ? c'est ici que les difficultés commencent.

En effet, l'enquête n'aura pas lieu dans chacune des sept villes, mais seulement dans une, deux, peut-être trois d'entre elles - disons deux pour fixer les idées - les deux villes échantillon pourront être choisies(*) ou désignées par le sort. Pour simplifier, nous admettrons qu'elles sont tirées au sort, en accordant à chacune d'elles, une chance de sortie proportionnelle au nombre de personnes âgées de 16 ans ou plus qui l'habitent.

Il est alors correct de répartir également l'échantillon total entre les deux villes-échantillon - il y aura donc un échantillon de 40 personnes à désigner dans chacune des deux villes-échantillon - nous supposerons que le sort a désigné Cannes et Montpellier.

Comment établir les quota par sexe, âge, milieu social ?

1ère méthode - On établit la répartition pour chacun des trois caractères de contrôle pour l'ensemble des sept villes réunies - chacun des échantillons de 40 étant réparti proportionnellement à cette répartition générale. Ainsi si, au total, les sept villes comptent 110 000 personnes (âgées de 16 ans et plus) appartenant au milieu ouvrier, dans chacune des deux villes-échantillon le quota sera fixé pour ce groupe à :

$$\frac{1}{2} \times \frac{1}{5\,000} \times 110\,000 = 11$$

2ème méthode - Les quota sont déterminés séparément pour chacune des deux villes-échantillon proportionnellement aux effectifs recensés dans chacune d'elles. Ainsi, si $\frac{1}{5}$ de la population adulte appartient au milieu ouvrier à Montpellier et $\frac{1}{6}$ à Cannes - le quota des ouvriers sera $\frac{1}{5} \times 40$ à Montpellier ; $\frac{1}{6} \times 40$ à Cannes.

Les deux méthodes ne conduisent pas à des résultats très différents en ce qui concernent le sexe et l'âge ; mais la différence peut ne pas être négligeable en ce qui concerne le milieu social dont la répartition varie assez fortement d'une ville à l'autre, même lorsqu'il s'agit des villes d'une même "strate", à priori pas trop différentes les unes des autes.

Chacune des deux méthodes présente des avantages et des inconvénients ;

(*) Le choix s'apparentera alors beaucoup à la désignation d'"unités-type". Toutefois, on tiendra largement compte de la facilité d'accès.

la seconde méthode, d'application plus commode(*), s'éloigne moins que la première des conditions du sondage probabiliste - mais nous ne disposons pas d'expérience permettant d'assurer qu'elle est préférable. Naturellement, on imagine facilement une méthode intermédiaire entre les méthodes 1 et 2.

On remarquera que même dans une grande ville constituant une strate à elle seule, le problème se pose au niveau du quartier.

En effet, il y aurait 102 personnes à interroger à Marseille ; le travail pourrait fort bien être réparti entre deux enquêteurs se partageant la ville par moitié. Or, il n'est pas possible en pratique d'établir des quota différents(**) pour chacune des deux moitiés de la ville ; de sorte que si l'on impose aux deux enquêteurs les mêmes quota, chacune des deux moitiés sera représentée par un échantillon dont les contrôles sont valables pour l'ensemble de la ville et pour l'ensemble seulement.

Ces difficultés se poseraient de manière particulièrement aigüe en ce qui concerne les strates constituées par des communes rurales ou des petites villes - les données statistiques faisant toujours défaut au niveau de la commune ou de la ville - il n'est donc pas possible d'appliquer la méthode numéro 2. Certains organismes - l'E.T.M.A.R. en particulier - procèdent par tirage au sort à l'intérieur des communes rurales échantillon.

I.2.2.4 - Considérations diverses sur la mise en œuvre de la méthode des quota

La méthode des quota est empirique.

Il est, pour cette raison, très difficile de la décrire en détail.

Les conditions optima d'application de la méthode varient fortement en fonction de l'implantation du réseau d'enquêteurs.

1/ Une première distinction très nette doit être notamment entre :

- les organismes qui utilisent un réseau d'enquêteurs fixes - travaillant au voisinage de leur domicile - l'implantation du réseau d'enquêteurs impose en pratique le choix des localités-échantillon. Les quota seront établis séparément pour chaque localité(***).

En notation symbolique, les contrôles prennent la forme :

(Région X Catégorie de commune) X (Sexe + âge + milieu social).

- les organismes qui utilisent des enquêteurs itinérants partant de Paris (ou de quelques grandes villes) et dont chaque enquêteur couvre une large fraction du territoire.

(*) Il peut être difficile pour l'enquêteur opérant dans une ville de remplir le quota d'une classe relativement plus rare dans cette ville que dans la "Strate". On remarquera que la méthode n° 1 favorise les individus aberrants (appartenant à une classe peu représentée dans la ville où ils résident) et défavorisent ceux qui se trouvent dans la situation contraire.

(**) Plus exactement on ne disposera généralement pas de statistiques permettant de le faire en toute rigueur.

(***) Chaque enquêteur opérant, en principe, dans une seule localité, il est nécessaire d'établir des quota pour chaque localité-échantillon.

Les enquêteurs opèrent généralement en équipe de 4 ou 6 dirigés par un(e) moniteur(trice). Il est alors possible d'établir des quota pour une région entière - ainsi, pour reprendre l'exemple donné plus haut, on imposera à l'équipe d'enquêteurs le nombre d'enquêtes à réaliser sur le midi méditerranéen. On lui imposera également la répartition des enquêtes entre les localités :

102 à Marseille.....
40 à Montpellier..... .
40 à Cannes.....

par ailleurs, les quota par sexe, âge, milieu social seront établis pour l'ensemble de la région :

En notation symbolique, les contrôles prennent la forme :

Région X (Catégorie de commune + sexe + âge + milieu social).

Ce type d'organisation - plus coûteux (les frais de déplacements sont très élevés) - laisse une très grande initiative au chef d'équipe ; la qualité de l'échantillon dépendra dans une large mesure de lui.

Du point de vue de la qualité de l'échantillon, nous ne pouvons accorder nettement la préférence à l'un des deux types d'organisation ; le deuxième est toutefois plus souple et laisse une liberté de manœuvre plus grande,

- soit au responsable du plan de sondage : s'il choisit de fixer des quota par localité - ce qu'il peut faire ;

- soit au chef de l'équipe d'enquêteurs en cas contraire, ce qui semble préférable.

2/ Nous avons vu par ailleurs que la désignation de l'échantillon se faisait en deux temps :

- premier temps : désignation de localités-échantillon,

- deuxième temps : désignation d'individus échantillon dans les localités échantillon.

A chacun des deux temps, on peut utiliser le mode de désignation par choix à dessein, le mode de désignation par tirage au sort. Il y a donc quatre modes de désignation possible :

	Localités	Individus
1	Choix à dessein	Choix à dessein
2	Tirage au sort	Choix à dessein
3	Choix à dessein	Tirage au sort
4	Tirage au sort	Tirage au sort

Les modes de désignation 1, 2 et 3 sont en fait utilisés parfois concurremment par le même organisme - ainsi, à notre connaissance, l'E.T.M.A.R. utilise les modes de désignation 1 et 2 pour les catégories urbaines, et le mode de désignation 3 pour les communes rurales.

Le mode de désignation n° 4 doit évidemment être mis à part puisqu'il s'agit du tirage aléatoire.

3/ Bien entendu, il est possible d'adopter des taux de sondage variables suivant les catégories d'individus.

Ainsi, dans une enquête sur l'épargne, on prendra un taux de sondage plus élevé dans les classes supérieures. Pour obtenir des résultats d'ensemble, on adoptera au dépouillement des coefficients de pondération inverses des taux de sondage.

4/ Il est bon d'ajouter aux contrôles statistiques quelques conditions supplémentaires que l'on imposera aux enquêteurs : sur la dispersion géographique des enquêtes, sur les conditions de recrutement des personnes interrogées (ceci est fondamental comme on le verra plus loin). Sur le lieu de réalisation de l'interview (interdiction d'opérer dans la rue, ou sur le lieu de travail par exemple)(*) et s'agissant d'enquêteurs fixes : interdiction de revenir trop souvent interroger les mêmes personnes.

Des précautions de cette nature sont certainement - bien qu'entièrement empiriques - beaucoup plus efficaces que des raffinements sur la définition des quota. Lorsqu'on emploie une méthode empirique, il faut être empirique jusqu'au bout.

I.2.3 - Critique de la méthode

I.2.3.I. - Avantages

a) La méthode ne requiert pas l'existence d'une "base de sondage" énumérant sans omission ni répétition les individus constituant la population.

Il s'agit là d'un avantage tout à fait décisif qui impose l'emploi de la méthode des quota dans de nombreux cas.

b - Economie et rapidité

Les sondages par quota sont nettement moins coûteux que les sondages probabilistes. Ceci sans même parler du coût de la constitution de la base de sondage.

L'enquêteur a un rendement environ deux fois plus élevé lorsqu'on lui permet de choisir les individus échantillon que lorsqu'on lui impose de travailler sur une liste d'adresses.

On peut estimer en gros qu'un sondage aléatoire coûte en France 50 % de plus qu'un sondage par quota identique par ailleurs.

c - Adaptation aux échantillons de faibles effectifs

La méthode probabiliste repose sur la loi des grands nombres. Lorsque l'échantillon est - par nécessité - petit, la méthode des quota est préférable à la méthode aléatoire(**).

(*) On pourra d'ailleurs imposer aux enquêteurs un itinéraire comme dans la méthode "haphazard".

(**) Si toutefois on stratifie l'univers avant tirage au sort, la méthode aléatoire reprend l'avantage.

d - Adaptation aux enquêtes comportant de sérieuses difficultés d'observation

Ceci est - à notre avis - le deuxième avantage essentiel de la méthode des quota.

Lorsque le questionnaire comporte des questions très délicates, les erreurs d'observation peuvent l'emporter et de beaucoup sur les erreurs d'échantillonnage. Peut-être seront-elles moindres sur des individus choisis par l'enquêteur.

Par ailleurs, le nombre élevé des refus tend alors à faire partiellement(*) perdre à la méthode aléatoire ses avantages.

I.2.3.2 - Inconvénients

a - Le défaut fondamental de la méthode est qu'elle repose sur une pétition de principe

A savoir que la distribution des variables de contrôle a, b, c ... détermine la distribution de la variable étudiée y.

Rien ne permet d'affirmer la vérité générale de ce principe déterministe - l'on n'est jamais certain a priori que les contrôles choisis assurent effectivement le caractère représentatif de l'échantillon, compte tenu de l'objet de l'étude. Pour prendre un exemple volontairement caricatural, il est certain que l'enquêteur opérant à Marseille peut parfaitement respecter les quota qui lui sont imposés tout en se contentant d'interroger les personnes qui attendent aux portes des cinémas.

Bien entendu, un sabotage aussi catastrophique serait facile à déceler et à éviter. Il n'en rest pas moins que le principe sur lequel repose la méthode n'est pas fondé en théorie et ne se suffit pas à lui-même.

A posteriori, en revanche, on pourra se tranquilliser si l'échantillon fournit avec une exactitude suffisante la distribution d'une variable non contrôlée mais dont la distribution est connue par ailleurs(**).

Nous venons de donner un exemple évidemment catastrophique, nous n'y reviendrons pas ; signalons quelques risques d'erreurs de même nature bien que moins grossières.

Les échantillons désignés par quota comprendront à la campagne beaucoup trop de foyers situés à proximité des voies de communication.

En général, ils comprendront trop de personnes actives - connues - ayant des relations - (voir page 22).

Un danger assez grand est le suivant - les enquêteurs(ses) travaillent

---

(*) Partiellement mais non totalement comme on le prétend souvent à tort.

(**) Il ne s'agira cependant que d'une présomption en faveur du caractère représentatif de l'échantillon - pour en revenir à l'exemple donné plus haut la répartition de l'échantillon par état matrimonial ou niveau d'instruction pourra être à peu près satisfaisante. L'échantillon n'en sera pas moins très mauvais, surtout si l'on procède à une étude sur les loisirs.

tantôt sur quota ; tantôt sur listes - il s'agit généralement alors de listes fournies par l'entreprise cliente de l'organisme d'enquêtes.

Exemple : liste d'abonnés - ou liste de personnes ayant fait livrer tel objet à domicile - l'enquêteur(se) expérimenté conserve ces listes qui comportent d'utiles renseignements sur les personnes qui y figurent et est tenté d'y puiser à l'occasion des enquêtes par quota. Ceci peut être très grave. En effet, les personnes qui figurent sur une de ces listes ont évidemment quelque chose en commun : toutes lisent le magazine x, ou font de la photo en couleurs. On va évidemment à une catastrophe si l'objet de l'étude est la lecture, la photo ou une activité complémentaire ou concurrente de celles-là.

Ces exemples illustrent l'importance toute particulière des précautions énumérées ci-dessus : alors que des quota très détaillés et très étudiés ne sont d'aucune efficacité pour éviter de tels dangers. Au contraire, des contrôles trop rigides peuvent conduire les enquêteurs à se procurer des listes pour trouver les individus présentant les caractéristiques recherchées : exemple : homme de 45 à 64 ans appartenant au milieu employé et résidant dans telle localité - et nous venons de voir le danger des listes.

b - Difficulté d'obtenir des contrôles corrects

Les contrôles imposés à l'enquêteur conduiront celui-ci à déformer l'échantillon si le recensement à partir duquel ont été calculés les quota :

1/ est erroné

2/ n'est plus à jour

3/ est réalisé suivant des principes de classification différents de ceux adoptés par les enquêteurs.

Les deux premiers points sont évidents - le troisième est plus délicat - éclairons-le par un exemple : si l'enquêteur au cours du sondage classe "ouvriers" des individus qui au recensement ont été classés "employés", il en résultera une sous-représentation des employés (au sens du recensement), une sur-représentation des ouvriers (au sens du recensement).

On le verra facilement sur un exemple fictif.

Le recensement donne : 100 000 ouvriers
50 000 employés

d'où les quota 20 pour les ouvriers - 10 pour les employés.

Au sens où l'entendent les enquêteurs, il y a :

110 000 ouvriers(*) (dont 20 seront interrogés)
40 000 employés (dont 10 seront interrogés).

On remarquera que tout le mal provient de ce que l'enquêteur ne parle pas le même langage que le service ayant assuré le recensement - peu im-

---

(*) 10 000 personnes classées "employés" au recensement seraient classées "ouvriers" par l'enquêteur.

porte de savoir qui se trompe, tout irait bien si par extraordinaire l'enquêteur et le recensement commettaient les mêmes erreurs(*).

Cette critique est moins grave que la précédente ; il n'en reste pas moins que l'on doit choisir comme contrôles des caractères faciles à observer et pour lesquels il existe de bonnes statistiques.

c - La méthode des quota ne permet pas d'évaluer la précision des estimations

Il est impossible de connaître les chances qu'avait chaque individu d'appartenir à l'échantillon (pour beaucoup d'individus, ces chances sont sans doute nulles). Dans ces conditions, il est impossible de se fonder sur le calcul des probabilités pour affirmer par exemple : "qu'il y a 95 chances sur 100 pour que l'estimation soit approchée à moins de 10 %".

En effet, la variabilité de l'estimation est inconnue et l'on n'est jamais sûr que l'estimation ne soit pas affectée d'une erreur systématique.

Les "calculs d'erreur" réalisés par les utilisateurs de la méthode des quota sont donc dénués de fondement - il y a d'ailleurs lieu de penser qu'ils tendent à surestimer la précision des résultats.

La méthode des quota ne peut invoquer d'autre autorité que celle de l'expérience, ce qui n'est pas négligeable ; un énoncé légitime serait le suivant : "sur des sujets qui semblent être voisins et en employant des méthodes identiques, on a constaté, lorsque des recoupements étaient possibles, que l'on obtenait en génral des résultats approchés à moins de 10 %".

d - Le contrôle des enquêteurs est difficile

Lorsqu'une enquêteur travaille sur listes d'adresses, il est facile de vérifier qu'il se conforme à c'elles-ci ; il est beaucoup plus difficile de contrôler dans le cadre d'une enquête par quota la manière dont l'enquêteur choisit les personnes qu'il interroge - beaucoup d'organismes exigent des enquêteurs qu'ils notent le nom et l'adresse des personnes interrogées, ce qui est une sage précaution. De toute manière, l'initiative laissée aux enquêteurs dans le choix des personnes à interroger est une cause de variabilité non négligeable.

I.2.4 - Etude expérimentale de la méthode des quota

Les sondages par quota sont l'objet de critiques violentes des théoriciens ; par ailleurs, leur importance pratique est considérable et ils semblent donner très fréquemment des résultats satisfaisants - de ce fait, ils méritent une étude attentive.

Une expérience très intéressante a été réalisée en Angleterre par la division des etudes techniques de la "London School of Economics"(**).

Cette expérience avait pour objet non seulement de comparer les résultats des sondages par quota aux résultats des sondages aléatoires (ou des

---

(*) Pour définir avec précision un contrôle correct, disons que les enquêteurs, s'ils recommençaient le recensement, devraient retrouver les chiffres qui ont servi de base au calcul des quota.

(**) Pour des résultats détaillés, voir J.R.S.S. - série A - volume CXVI - partie IV 1953.

recensements lorsque des résultats étaient disponibles) mais encore de comparer différentes méthodes de sondages par quota.

Les points étudiés étaient les suivants :

- Y a-t-il intérêt à augmenter le nombre des caractères "contrôlés" ?

- Y a-t-il intérêt à utiliser des contrôles marginaux ou des contrôles croisés ?

- Quelle est l'influence des enquêteurs ?

Organisation de l'expérience

L'expérience a été réalisée dans trois villes : Birmingham, Bristol, Edimbourg, auprès d'une population de personnes adultes.

Quatre organisations utilisant habituellement la méthode des quota, participèrent à l'expérience :

1/ B.B.C. Audience Research Department
2/ British Market Research Bureau
3/ Research Services
4/ Market Information Services.

Trois caractères étaient systématiquement contrôlés :

- le sexe
- l'âge : 4 groupes : 20-29 ; 30-44 ; 45-64 ; 65 et plus
- la classe sociale : 3 classes sociales.

Les organisations 1 et 4 ont utilisé des contrôles marginaux. Les organisations 2 et 3 ont utilisé des contrôles croisés.

Par ailleurs, chaque organisation utilise :

- d'une part un plan I comportant les 3 contrôles énumérés ci-dessus,

- d'autre part un plan II comportant un contrôle supplémentaire.

Deux contrôles supplémentaires ont été utilisés alternativement :

a) l'activité collective (par les organisations 1 et 2).

Avec la classification suivante : commerce - transports - professions administratives et libérales - industrie - sans activité.

b) le quartier (pour les organisations 3 et 4) ; dans chaque ville, on a tiré six points de repère(*). Chaque enquêteur devait répartir l'échantillon qui lui était alloué par parties égales entre les six quartiers.

Des statistiques par sexe, âge, activités collectives existaient pour chaque ville ; en revanche, la répartition par classe sociale n'existait pas mais cette répartition n'existe pas davantage à l'échelle nationale : ce "caractère" est subjectif ; cependant les procédés de classification utilisés par les quatre organismes semblent peu différents. On a vu le caractère empirique de la définition du quartier.

(*) 6 "quartiers" ont ainsi été définis : un quartier étant l'ensemble des point accessibles à pied à partir d'un point de repère.

Le plan d'expérience(*) était le suivant pour chacune des villes :

Organisation	1		2		3		4	
Contrôle	marginaux		croisés		croisés		marginaux	
Plan	I	IIa	I	IIa	I	IIb	I	IIb
Sous-échantillons	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2

Les deux sous-échantillons identiques quant au mode de désignation étaient affectés à des enquêteurs différents. Dans chaque ville, chaque organisation utilisait donc 2 × 2 = 4 enquêteurs différents, chaque enquêteur devait effectuer 90 interview.

Dans chaque ville : 4 × 4 × 90 = 1 440 interview furent donc réalisées soit : 3 × 1 440 = 4 320 au total.

On remarquera que ce plan d'expérience est conçu pour donner le maximum de précision aux comparaisons entre les enquêteurs et entre les méthodes utilisées.

Les enquêteurs avaient reçu les instructions habituelles(**), mais il y a lieu de penser qu'ils ont travaillé avec un soin particulier.

Le sondage aléatoire témoin

360 individus ont été tirés au sort sur les listes électorales dans chaque ville ; chaque ville était divisée en trois zones, le sondage étant réalisé dans chaque zone par l'une des organisations 2, 3 ou 4 (l'organisation 1 n'ayant pu participer à cette partie de l'expérience). La répartition des zones entre les trois organisations a été réalisée par tirage au sort.

Le sondage "aléatoire" était effectué sans stratification, mais on aurait très bien pu stratifier par sexe et âge.

Le questionaire

Le questionnaire - très simple - comportait des questions démographiques, des questions relatives à la profession, au niveau d'instruction, au niveau de vie et à l'utilisation des loisirs. Tous ces caractères sont faciles à observer. L'expérience n'est donc pas troublée par d'importantes erreurs d'observation.

Les résultats

On peut dire que - en gros - la méthode des quota est sortie de l'expérience à son avantage surtout si l'on pense à son utilisation pour les études de marché : en effet, l'infériorité de la méthode des quota par rapport à la

---

(*) C'est un exemple intéressant des méthodes expérimentales.

(**) Les conditions de l'expérience étaient suffisamment proches des conditions habituelles pour ne pas avoir gêné les organismes d'études de marché dans leur travail.

méthode aléatoire est surtout manifeste en ce qui concerne l'étude des caractères démographiques et socio-professionnels.

Les résultats sont présentés avec un grand détail dans l'article du J.R.S.S., dont on ne saurait trop recommander la lecture.

On a constaté :

1/ que les enquêteurs ne respectaient pas absolument les quota qui leur étaient imposés : la distribution de l'échantillon n'est pas exactement contrôlée en ce qui concerne le sexe, l'âge, le milieu social.

2/ La distribution par groupes d'âge quinquennaux, à l'intérieur des quatre groupes d'âge dont l'effectif était (approximativement) contrôlé, est assez fortement perturbée.

Dans chacun des quatre groupes d'âge (sauf le premier), on constate une nette sous-représentation des groupes d'âge les plus élevés. Ces écarts ne sauraient être dus au hasard : l'échantillon est systématiquement "trop jeune".

Il y a même lieu de penser que les enquêteurs, ayant de la peine à trouver des personnes appartenant aux groupes d'âge supérieurs modifient le moins possible) l'âge des personnes qui acceptent de répondre ; ainsi une personne de 63 ans sera indiquée comme ayant 65 ans et comptera pour la classe d'âge supérieur. On constate, en effet, la présence dans les échantillons par quota d'un nombre nettement trop élevé (environ 50 %) de personnes dont le second chiffre de l'âge est un 5.

Cette constatation n'est pas sans gravité. Il est regrettable que la méthode conduise les enquêteurs à manquer de sincérité et de précision pour remplir leurs quota - et cela pour une variable aussi bien définie que l'âge. On peut légitimement se demander comment fonctionne le contrôle par classes sociales A, B, C. . .

3/ Le contrôle par "quota indépendants" semble avoir - dans l'ensemble - donné de meilleurs résultats que le contrôle par quota croisés. Cette dernière méthode impose à l'enquêteur des contrôles trop étroits et minutieux, ce qui nuit plutôt à la qualité de l'échantillon.

Bien entendu, lorsqu'on contrôle seulement les distributions marginales, la distribution de l'échantillon par sexe, âge, milieu social s'éloigne significativement de la distribution correcte (par exemple la distribution en quatre groupes d'âge est à peu près correcte pour l'ensemble mais pas pour chaque sexe). Toutefois, comme nous venons de le voir, cet inconvénient ne justifie pas l'emploi de quota croisés.

4/ La méthode des quota conduit - en l'absence d'un contrôle par catégorie socio-professionnelle ou par activité collective - à une sous-représentation massive des travailleurs de l'industrie d'accès plus difficile que les travailleurs de l'administration et du commerce.

Si l'on élimine l'échantillon contrôlé suivant l'activité collective, on constate qu'à Birminghan et Bristol pour les hommes, dans les trois villes pour les femmes, le nombre de travailleurs de l'industrie désigné par la méthode des quota est voisin de la moitié de ce qu'il devrait être. A Edimbourg toutefois, pour les hommes, l'échantillon est très correct sur ce point.

Il y a donc le plus grand intérêt à introduire un contrôle suivant la catégorie socio-professionnelle.

5/ La méthode des quota conduit à une sous-représentation, d'ailleurs légère, des classes les moins instruites de la population.

Ce résultat joint au précédent confirme le fait que les enquêteurs ont tendance à rechercher les personnes à interroger dans un milieu socialement proche du leur.

6/ Les résultats relatifs à l'utilisation des loisirs sont en accord très correct avec les résultats fournis par l'échantillon aléatoire témoin.

	Echantillon	
	Aléatoire	Par quota
Inscrits à une bibliothèque publique.....	26,6	28,7
Ayant été au cinéma au cours des quatre dernières semaines.....	42,1	47,9
Ayant participé au pari mutuel au cours de la saison.....	23,9	26,9
Fumeurs.....	52,8	55,5
Lecteurs d'au moins un des huit grands quotidiens énumérés.....	73,2	77,5
Lecteurs d'au moins un des journaux du dimanche énumérés.....	81,5	85,3

Quel que soit le type de loisirs étudié, les échantillons établis par quota donnent des estimations légèrement supérieures aux estimations fournies par les échantillons désignés par tirage au sort.

Ce qui peut s'interpréter ainsi : les individus interrogés ont en moyenne des activités sociales plus nombreuses lorsqu'ils sont choisis par les enquêteurs que lorsqu'ils sont désignés par le sort.

Il n'est d'ailleurs pas certain que la vérité soit du côté de la méthode aléatoire.

En effet : plus une personne connaît de monde, plus il y a de chances pour qu'un enquêteur travaillant par quota ait l'occasion de l'interroger (d'où risque des biais dans le sens d'une surestimation de l'activité au sens large).

Mais plus une personne exerce d'activités diverses, moins un enquêteur travaillant sur listes d'adresses a de chances de la rencontrer à domicile (d'où risque de biais dans le sens d'une sous-estimation de l'activité au sens large).

7/ Les échantillons ayant été systématiquement répartis par parts égales entre deux enquêteurs, il est possible d'estimer la variabilité des estimations obtenues par la méthode des quota (voir étude sur les erreurs d'observation) ; cette variabilité est assez considérable et les auteurs de l'article du J.R.S.S. pensent qu'elle est nettement supérieure à la variabilité des estimations obtenues par la méthode aléatoire.

En effet, une partie importante de la variabilité totale provient de la variabilité entre enquêteurs ; or celle-ci est la somme de deux termes :

- variabilité dans le mode de sélection de l'échantillon ;
- variabilité dans la manière d'interroger.

Le premier terme disparaît dans le cas des enquêtes aléatoires et l'on peut penser que c'est le plus important des deux.

L'estimation de la variabilité a pu être réalisée car les organisateurs de l'expérience avaient pris soin de constituer des réseaux superposés d'enquêteurs, ce qui ne se fait pas en pratique. Par ailleurs, cette estimation de la précision ne tient aucun compte des erreurs systématiques qui pourraient résulter d'un fonctionnement défectueux de la méthode.

I.2.5 - Conclusions sur la méthode des quota.

Il ressort de ce qui vient d'être dit, et plus particulièrement des résultats de l'expérience anglaise (I.2.4.), que la méthode des quota - bien que dénuée de fondements théoriques satisfaisants - est un instrument de travail très utilisable - et bien souvent même le seul instrument de travail dont on dispose en l'absence d'une base de sondage adéquate.

Cette méthode convient particulièrement lorsqu'on désire obtenir rapidement des résultats avec une large approximation et lorsque les caractères étudiés ne peuvent de toute manière être observés avec précision.

Il semble souhaitable de remplacer le contrôle assez flou suivant des classes sociales A, B, C ... mal définies, par un contrôle faisant intervenir la catégorie socio-professionnelle ; enfin les précautions énumérées p. 15 sont fondamentales, alors qu'il n'est sans doute pas utile - et peut-être même dangereux - d'imposer aux enquêteurs des contrôles trop détaillés.

Indiquons enfin que la méthode des quota a plutôt tendance à reculer au bénéfice de la méthode haphazard et de la méthode aléatoire.

2 - ECHANTILLONS DESIGNES SUIVANT LA METHODE "HAPHAZARD"

Le principal inconvénient de la méthode des quota est de laisser aux enquêteurs une trop grande initiative dans le choix des individus échantillon. Le principal avantage est de ne pas requérir l'existence d'une base de sondage. La méthode "haphazard" (dite encore méthode Politz, du nom de son inventeur ou méthode de randomization) tente de conserver cet avantage en supprimant cet inconvénient. Cette méthode utilisée surtout pour désigner des échantillons de ménages ou logements, consiste à imposer à chaque enquêteur un itinéraire défini jusqu'au moindre détail en lui indiquant exactement en quels points de son itinéraire il doit réaliser une interview. On est souvent conduit à concevoir dans chaque ville deux plans de sondage.

- l'un concernant les immeubles individuels,
- l'autre concernant les immeubles collectifs.

Pour l'enquêteur, les conditions de travail sont à peu près les mêmes que dans les enquêtes aléatoires : tout se passe comme si on lui imposait une liste de logements à visiter, chaque logement étant repéré par ses coordonnées géographiques.

Bien entendu, la méthode ne donne pas à tous les logements des chances égales d'être désignés.

On remarquera à cette occasion que le mot hasard a un sens très dif-

férent dans le langage courant et dans le langage probabiliste ; l'échantillon est désigné par le hasard au premier sens mais non pas au deuxième sens de ce mot. Le caractère de représentativité de l'échantillon dépend donc uniquement du discernement et des connaissances géographiques de la personne qui établit le plan de sondage.

On remarquera qu'il est en pratique essentiel que le travail de l'enquêteur puisse être contrôlé - les personnes qui ont l'expérience de cette méthode assurent qu'un enquêteur et un contrôleur consciencieux sont en général d'accord sur l'identité des ménages à interroger(*) ; on peut donc vérifier le sérieux avec lequel l'enquêteur respecte les consignes qui lui sont données.

Cette méthode - plus coûteuse que la méthode des quota - paraît également plus sûre - elle est de plus en plus employée en France par les organismes d'études de marché.

III - AUTRES METHODES EMPIRIQUES INTERESSANTES

3.1 - Dans son enquête sur le comportement sexuel aux U.S.A. le docteur Kinsey ne pouvait évidemment interroger que des volontaires. Mais il y avait lieu de craindre une liaison très étroite entre le comportement sexuel, objet de l'étude, et l'attitude d'acceptation ou de refus à l'égard de l'enquête, d'où le risque d'une erreur systématique considérable. La méthode des quota n'offrait donc - par elle-même - aucune sécurité.

Le Docteur Kinsey s'efforça d'obtenir l'accord non pas d'individus mais de groupes entiers (amicales d'anciens élèves, associations diverses...) dont tous les membres seraient invités à répondre à l'enquête. Le redoutable effet de sélection était ainsi partiellement évité.

Ce procédé peut être utilisé pour des sujets très différents mais également très délicats... tel que l'épargne par exemple.

3.2 - Un autre procédé, offrant moins de garanties, consiste à obtenir l'accord préalable d'un grand nombre de volontaires sur lesquels on note les principaux caractères socio-démographiques.

L'échantillon proprement dit est choisi parmi ce gros échantillon de volontaires de manière à être "représentatif" au sens de la méthode des quota.

(*) Ce qui n'est pas évident à priori, vu la structure peu cartésienne de la plupart des villes françaises.