

JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

P. THIONET

Quelques données sur les sondages de l'I.N.S.E.E. en 1952-1953

Journal de la société statistique de Paris, tome 95 (1954), p. 165-196

http://www.numdam.org/item?id=JSFS_1954__95__165_0

© Société de statistique de Paris, 1954, tous droits réservés.

L'accès aux archives de la revue « Journal de la société statistique de Paris » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

III

VARIÉTÉ

Quelques données sur les sondages de L'I. N. S. E. E. en 1952-1953.

INTRODUCTION

Après le recensement général de 1954, les enquêtes par sondage de l'Institut National de la Statistique et des Études Économiques vont s'engager dans des voies quelque peu nouvelles. Avant eu jusqu'à ce recensement (et depuis la fin de 1947) la responsabilité des techniques employées, il nous a semblé qu'il y avait lieu de présenter aux lecteurs du *Journal de la Société de Statistique*

de Paris certains renseignements inédits, d'ordre méthodologique, concernant ces sondages. Il s'agit de deux petites études qui répondaient à un *besoin interne*, mais qui paraissent à présent présenter un intérêt général suffisant pour que leur publication ait lieu. Elles concernent :

— d'une part, l'enquête de l'I. N. S. E. E. sur l'Emploi, faite à la fin de 1952 (1);

— d'autre part, les études des marchés de la chaussure et des assurances effectuées en 1953 par l'I. N. S. E. E. pour le compte des organismes chargés de promouvoir la productivité dans ces deux secteurs.

Il semblerait qu'il s'agisse de deux études particulières et indépendantes; il n'en est rien. L'I. N. S. E. E. a procédé, en 1952 et 1953, à une série d'enquêtes par sondage, établies suivant un plan d'échantillonnage commun (les échantillons variant d'une enquête à l'autre). Ce sont :

— trois enquêtes sur l'Emploi (de 1952, du 1^{er} et du 2^e semestre 1953);

— quatre enquêtes sur la consommation de textiles (une pour chaque trimestre de 1953);

— une enquête auprès du public sur le marché de la chaussure;

— deux enquêtes sur le marché des assurances (respectivement : en mai 1953, pour les villes de plus de 30.000 habitants; en novembre 1953, pour les autres localités);

— en outre, une enquête sur l'écoute radiophonique, fin 1952, reprise début 1954 (les deux enquêtes-radio présentent toutefois des caractères techniques assez notablement différents de ceux des autres sondages de l'I. N. S. E. E.).

Les deux études qui vont suivre concernent donc, en fait, l'ensemble de ces enquêtes. Au préalable, il y a lieu de rappeler brièvement leur plan d'échantillonnage commun (à l'exclusion toutefois des enquêtes radio).

* * *

Plan d'échantillonnage

Le *champ* de l'enquête est la France métropolitaine (Corse exclue).

L'enquête a lieu sur un *échantillon à deux degrés*.

Unités du 1^{er} degré : les communes (à l'intérieur de certaines « strates »).

Unités du 2^e degré : les ménages (à l'intérieur de certaines « sous-strates »).

Rappelons qu'on appelle ménage l'ensemble des personnes vivant dans un même logement.

Stratification

La « stratification » des communes a été effectuée (dans ses grandes lignes) en croisant un découpage géographique du territoire (en grandes régions) et un classement des communes par grandes catégories.

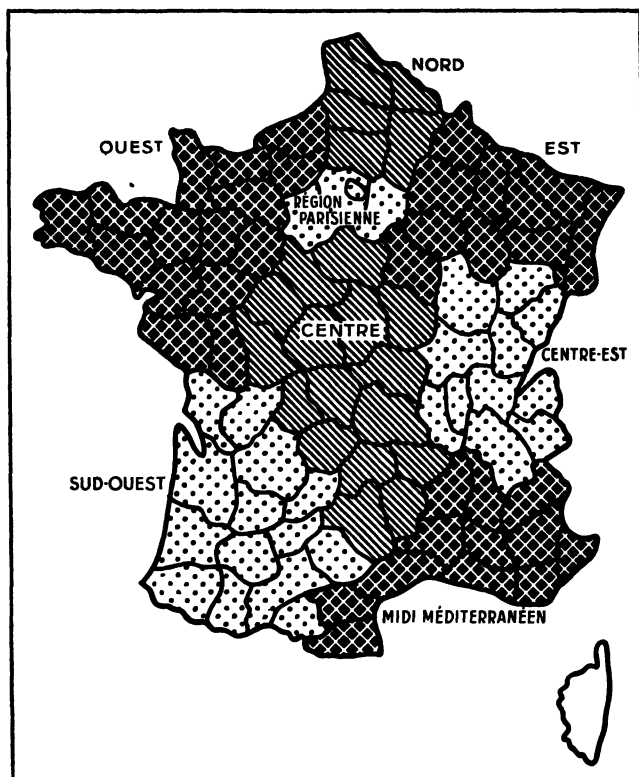
(1) En 1950 et 1951, l'I. N. S. E. E. avait déjà effectué *chaque semestre*, une enquête sur l'Emploi.

Voir au *Journal de la Société de Statistique de Paris*, d'avril-juin 1951, p. 108, la communication de R. Levy-Bruhl, Enquêtes récentes par sondage sur l'Emploi.

a) *Les grandes régions :*

Le territoire a été découpé en huit grandes régions, dont la composition est donnée par la carte 1 ci-dessous.

CARTE 1 : *Les huit grandes régions*



b) *Les catégories de communes :*

On a utilisé le classement des communes en communes rurales et communes urbaines, auquel on a superposé :

— pour les premières : un classement suivant la population vivant de l'agriculture;

— pour les secondes : un classement suivant la population totale.

Au tableau 1 on trouvera la définition des dix catégories de communes ainsi retenues.

TABLEAU 1

CATÉGORIES DE COMMUNES	NUMÉRO
<i>Communes rurales :</i>	
Proportion de la population vivant de l'agriculture :	
Plus de 60 %	1
40 à moins de 60 %	2
20 à moins de 40 %	3
Moins de 20 %	4
<i>Communes urbaines :</i>	
Moins de 5.000 habitants.	5
De 5.000 à 10.000 habitants.	6
De 10.000 à 30.000 habitants.	7
De 30.000 à 100.000 habitants.	8
De 100.000 à 150.000 habitants.	9
Plus de 150.000 habitants.	10

Mode de tirage de l'échantillon

Unité du 1^{er} degré. — Les communes-échantillon ont été désignées par tirage au sort en attribuant aux communes des probabilités d'être tirées proportionnelles à leur taille (1).

Unité du 2^e degré. — A l'intérieur d'une commune-échantillon, les ménages-échantillon ont été tirés au sort avec d'égales probabilités.

PROCÉDÉS PRATIQUES

Tirage au 1^{er} degré

On possède des listes de communes de chaque strate où, en regard des noms des communes, figurent les tailles de celles-ci et les totaux cumulés de ces tailles. On désigne les communes-échantillon en repérant des nombres aléatoires (prélevés sur les *Tables of Random Numbers* de Fisher et Yates) dans la colonne des totaux cumulés.

Tirage au 2^e degré

Les logements des ménages-échantillon sont désignés par tirage systématique dans les liasses de feuilles de ménage du recensement de 1946; si le logement a changé d'occupants, on interroge les nouveaux occupants.

Les logements disparus depuis 1946 s'éliminent d'eux-mêmes.

On tire, en outre, un échantillon adéquat de logements neufs (habités depuis 1946) à partir du fichier des nouveaux logements établi en principe pour chaque commune-échantillon à l'aide des permis de construire.

* * *

QUELQUES EXPLICATIONS AU SUJET DE CE PLAN

Voici le principe de la méthode d'échantillonnage suivie par l'I. N. S. E. E., à la suite du *Bureau of the Census* (États-Unis d'Amérique), de l'*Indian Statistical Institute*, etc... L'échantillon est obtenu, non par un choix raisonné, mais par *des tirages au sort* à l'intérieur d'un plan de sondage qui, lui, a été *choisi* en vue de réduire les erreurs d'échantillonnage. Cette méthode est *mathématique*, c'est-à-dire se prête à l'application correcte du calcul des probabilités.

Mais, alors que la méthode du *Bureau of the Census* consiste à tirer au sort des petites aires (dessinées sur des cartes ou des plans de ville), la méthode de l'I. N. S. E. E., comme d'ailleurs celle du *Social Survey* britannique, repose

(1) Par taille d'une commune on entend :
le nombre de ménages pour les communes urbaines et les communes rurales où moins de 20 % de la population vit de l'agriculture;
le nombre d'habitants pour les communes rurales où, au moins, 20 % de la population vit de l'agriculture.

sur des tirages au sort d'adresses dans des fichiers ou sur des listes préétablis. La différence essentielle entre le procédé de l'I. N. S. E. E. et celui du *Social Survey* est que les administrations britanniques sont en mesure de fournir au *Social Survey* des listes d'adresses adaptées au but visé; tandis qu'en France, après avoir essayé (Enquête Emploi 2^e semestre 1951 et Enquête Loyers 1952) le tirage au sort dans les calepins de l'administration des Contributions Directes, l'I. N. S. E. E. a préféré s'en tenir à l'emploi de sa documentation personnelle (les feuilles de ménage du recensement 1946) complétée par un échantillon de nouveaux logements (grâce à l'institution du permis de construire).

PREMIÈRE PARTIE

ÉTUDE DE L'ÉCART ENTRE LES ASPECTS THÉORIQUE ET PRATIQUE DE L'ÉCHANTILLONNAGE

Les ménages auprès desquels a lieu l'enquête sont donc tirés au sort; et il n'est pas évident qu'ils accepteront de répondre aux questions des enquêteurs. Si beaucoup s'y refusent, ou si l'enquêteur ne peut les joindre, l'échantillon soumis à l'enquête devient plus petit (ce qui n'est pas en soi un vice rédhibitoire) et surtout *se déforme*; il importe donc de savoir si l'échantillon réellement interrogé ne cesse pas d'être représentatif. En outre, il faut s'assurer que beaucoup d'enquêtés ne répondent pas avec tellement de mauvaise grâce que la sincérité de leurs réponses puisse être elle-même mise en doute.

Telle est bien d'ailleurs la vérité pour des enquêtes *difficiles* comme celles sur les budgets de famille, pour lesquelles l'I. N. S. E. E. a renoncé (après plusieurs tentatives) à tirer au sort ses échantillons et s'est décidé à suivre une méthode non mathématique, la *méthode des quotas*, c'est-à-dire celle qu'adoptent, en général, les Instituts d'opinion publique comme l'Institut Gallup (aux États-Unis) ou l'Institut français d'Opinion Publique.

Mais en revanche pour ses sondages de type courant, l'I. N. S. E. E. obtient avec les tirages au sort des résultats satisfaisants, qui ne peuvent que l'inciter à persévérer dans l'emploi des méthodes mathématiques. C'est surtout le cas pour le *Sondage sur l'Emploi* effectué périodiquement par l'I. N. S. E. E.; la présente étude a été faite à propos de l'enquête qui eut lieu à la fin de 1952; celle prévue pour le 1^{er} semestre 1952 avait dû être annulée faute de crédits. Elle avait été précédée, en 1950 et 1951, d'enquêtes semestrielles sur le même sujet, qui avaient permis de mettre au point le détail des procédés utilisés. En revanche, les deux enquêtes Emploi de 1953 se sont peut-être déroulées dans de moins bonnes conditions, en raison de l'afflux d'enquêtes que l'I. N. S. E. E. a connu cette année-là : notamment pour l'enquête Emploi du 1^{er} semestre qui fut jumelée à l'enquête Textile du 2^e trimestre, laquelle eut lieu en même temps et dans les mêmes localités (auprès de ménages différents).

En résumé, les résultats qui suivent correspondent probablement à des conditions un peu meilleures que la moyenne.

CHAPITRE I

L'ACCUEIL DU PUBLIC

L'accueil que le public réserve aux enquêteurs de l'I. N. S. E. E. présente déjà un intérêt en soi, en dehors de toute considération sur la validité des méthodes de sondage. Cet accueil est connu de l'échelon central par deux voies :

— d'une part, les comptes rendus que tout enquêteur doit fournir, sa mission terminée;

— d'autre part, l'indication portée par lui sur *chaque questionnaire*.

Voici les résultats du dépouillement des questionnaires de l'enquête Emploi 1952, en ce qui concerne l'accueil.

Le tableau 2 fait connaître comment l'accueil reçu varie avec la région où l'on opère.

Au tableau 2 on a distingué, pour la région de Paris, deux catégories d'enquêteurs :

les *agents* de l'I. N. S. E. E. proprement dits, qui n'effectuent qu'une partie des enquêtes;

les *enquêteurs extérieurs* (rémunérés à la tâche), qui sont chargés de la quasi-totalité du travail à Paris même et dans une partie de la proche banlieue (en fait, on leur confie la visite des adresses qui se trouvent les plus proches de leur domicile).

On remarquera que l'accueil fait aux enquêteurs extérieurs serait moins bon que celui réservé aux agents de l'I. N. S. E. E. Il est possible, effectivement, que les premiers soient moins expérimentés que les seconds; mais il est également vraisemblable que la différence tient pour une part au fait que les enquêteurs extérieurs sont plus exigeants que des fonctionnaires sur ce qu'il y a lieu d'appeler un bon accueil.

TABLEAU 2

Variation de l'accueil reçu par les enquêteurs suivant la région

DIRECTIONS RÉGIONALES de l'I.N.S.E.E.	PROPORTIONS % D'ACCUEILS				
	Très bons	Bons	Passables	Mauvais	
Paris { Agents I.N.S.E.E. . .	24,4	62,2	11,1	2,3	100
Paris { Enquêteurs extérieurs .	19,9	71,4	7,1	1,6	100
Lille	10,5	85,9	4,8	0,8	100
Rouen	9,3	80,6	8,8	1,3	100
Reims	21,8	69,7	7,6	0,9	100
Nancy	7,0	82,8	9,7	0,5	100
Rennes	6,8	87,3	5,4	0,5	100
Nantes	17,7	72,1	10,2	—	100
Orléans	5,5	88,4	4,8	1,3	100
Dijon	15,6	80,3	3,3	0,8	100
Limoges	7,9	84,2	7,9	—	100
Clermont	28,8	62,8	7,8	0,6	100
Lyon	21,1	70,8	6,9	1,2	100
Bordeaux	32,2	64,4	3,4	—	100
Montpellier	52,4	44,7	2,9	—	100
Toulouse	16,8	77,6	9,6	1,0	100
Marseille	9,4	88,4	1,9	0,3	100
Poitiers	16,5	77,3	5,1	1,1	100
Strasbourg	53,7	31,5	14,8	—	100
ENSEMBLE	17,2	75,4	6,5	0,9	100

De même les écarts observés entre les enquêteurs des diverses directions régionales correspondent partie à des différences réelles et partie à des différences d'appréciation. Quoi qu'il en soit on retiendra qu'il n'y a guère eu que 1 % d'accueils vraiment mauvais (signalons qu'il y eut moins de 2 % de refus de répondre).

Le tableau 3 montre que l'accueil varie beaucoup avec la classe sociale des enquêtés. Reçoivent assez mal les enquêteurs : les artisans et petits commerçants et aussi les salariés agricoles.

En revanche, les meilleurs accueils proviennent des professions intellectuelles, cadres supérieurs et cadres moyens. Les groupes les plus nombreux (agriculteurs, ouvriers, employés) réservent aux enquêteurs un accueil assez bon.

TABEAU 3
*Variation de l'accueil reçu par les enquêteurs
suivant la catégorie sociale des enquêtés*

PROPORTION % d'accueils	CATÉGORIE SOCIO-PROFESSIONNELLE DU CHEF DE MÉNAGE									
	Agriculteurs	Petits patrons	Patronat	Intellectuels Cadres supérieurs	Techniciens Cadres moyens	Employés	Ouvriers Maîtrise	Salariés agricoles	Services	Retraités Rentiers
Tres bons.	15,1	20,3	20,8	36,7	31,1	21,6	13,0	9,8	16,7	12,9
Bons.	77,3	68,8	73,8	61,1	65,3	71,2	80,0	80,4	75,0	79,3
Passables.	7,6	10,9	5,4	2,2	3,6	7,2	7,0	9,8	8,3	7,8
Mauvais.										
TOTAL	100	100	100	100	100	100	100	100	100	100

Le tableau 4 fait ressortir une amélioration régulière de l'accueil reçu quand on quitte Paris pour la province, et en province quand on passe des grandes villes aux petites et des petites villes aux communes rurales.

Le tableau 4 bis reprend en détail les résultats résumés au tableau 4; les communes rurales sont ventilées en quatre groupes définis au tableau 1.

On voit que l'accueil serait assez médiocre dans les communes de la catégorie 4 qui sont souvent des communes de banlieue des grandes villes.

La différence d'accueil qui apparaît entre la banlieue de Paris et Paris lui-même est accentuée (on l'a vu) par le fait que les enquêteurs de Paris sont presque tous des enquêteurs extérieurs.

TABEAU 4
*Variation de l'accueil reçu par les enquêteurs
suivant le lieu de l'enquête*

PROPORTION % D'ACCUEILS	COMMUNES rurales	VILLES de moins de 30.000 habitants	VILLES de plus de 30.000 habitants	ENSEMBLE de la province	PARIS ville et banlieue
Tres bons.	13,6	16,0	20,1	15,7	21,8
Bons.	78,9	77,1	72,3	77,0	69,8
Passables.	6,9	6,1	6,4	6,6	7,3
Mauvais.	0,6	0,8	1,2	0,7	1,1
TOTAL	100	100	100	100	100

TABEAU 4 bis
*Variation de l'accueil reçu par les enquêteurs
suivant le lieu d'enquête*

PROPORTION % d'accueils	COMMUNES RURALES				VILLES DE PROVINCE DE						BAN- LIEUE de Paris	PARIS ville
	Cat. 1	Cat. 2	Cat. 3	Cat. 4	Moins de 5.000 habi- tants	5.000 à 10.000 habi- tants	10.000 à 30.000 habi- tants	30.000 à 100.000 habi- tants	100.000 à 150.000 habi- tants	Plus de 150.000 habi- tants		
Très bons.	13,2	12,8	17,9	10,2	14,7	20,0	13,6	23,6	13,5	19,8	16,2	27,1
Bons.	80,3	80,1	74,5	79,6	77,4	73,5	80,2	68,7	75,4	74,0	72,6	66,9
Passables.	6,3	6,8	6,7	8,3	7,0	6,3	5,2	6,1	10,1	5,3	9,2	5,5
Mauvais.	0,2	0,3	0,8	1,0	0,9	0,2	1,0	1,6	1,0	0,9	2,0	0,5
TOTAL	100	100	100	100	100	100	100	100	100	100	100	100

Corrélation entre l'accueil et les réponses des enquêtés

Les questionnaires recueillis au cours de l'enquête Emploi comportent une certaine proportion de lacunes; on s'est occupé plus spécialement de ceux où manquait :

- soit l'activité principale;
- soit le nombre d'heures de travail,

et on a recherché si cette absence de renseignement ne coïncidait pas avec un mauvais accueil de la part de l'enquêté. Ce n'est absolument pas le cas (tableau 5).

TABEAU 5
Proportion % d'accueils

	Très bons	Bons	Passables	Mauvais	
Ensemble des enquêtés. . . .	17,2	75,4	6,5	0,9	100,0
Enquêtés pour lesquels l'en- quête n'a pas pu recueillir :					
l'activité principale	18,1	75,8	5,3	0,8	100,0
le nombre d'heures de travail	17,9	76,0	5,3	0,8	100,0

Mais on a constaté qu'une confusion s'était établie dans le *chiffrement des réponses* entre les notations *Sans objet* et *Sans réponse*. Et il y a lieu de limiter le dépouillement des « Sans réponse » (pour ces deux questions) aux seules personnes actives de l'échantillon; ici encore il faut distinguer deux cas, suivant qu'on a affaire à des personnes actives de l'échantillon ayant ou n'ayant pas d'emploi. On constate que les « Sans réponse » correspondent pour moitié à des chômeurs (pour lesquels on devait indiquer la profession la plus récente).

TABEAU 5 bis

PERSONNES ACTIVES	PROPORTION % D'ENQUÊTÉS dont l'activité principale est non-déclarée
Ayant un emploi	0,7 % — 74 personnes sur 9.804
N'ayant pas d'emploi et en cherchant un (autrement dit en chômage)	29 % — 68 personnes sur 233

On ne peut finalement discerner aucune différence significative dans la qualité de l'accueil, entre les personnes ayant un emploi et celles qui sont en chômage. L'accueil des personnes actives serait légèrement moins bon que celui des personnes non actives (Tabl. 5 *ter*).

TABLEAU 5 *ter*

PERSONNES N'AYANT PAS FAIT CONNAÎTRE LEUR ACTIVITÉ PRINCIPALE					
Personnes actives	ayant un emploi	89	%	d'accueils	bons ou très bons
	sans emploi	91	%	d'accueils	bons ou très bons
Personnes non actives		94	%	d'accueils	bons ou très bons
Ensemble		92,6	%	d'accueils	bons ou très bons

CHAPITRE 2

CAS DES PERSONNES AYANT EMMÉNAGÉ DEPUIS 1946

Le bon accueil général, qui vient d'être signalé, tient en grande partie à ce que la méthode suivie ces dernières années par l'I. N. S. E. E. consiste à avertir, à l'avance les intéressés, par lettre officielle, de la visite de l'enquêteur. Celui-ci est donc reçu sans difficulté si l'enquête n'est pas hostile *a priori* à l'administration et à l'objet de l'enquête. En revanche, dans les rares cas où l'enquête est hostile, on aura plus de difficulté à se faire recevoir que s'il n'y avait pas eu de lettre d'avis. En fait, les refus et accueils mauvais ne viennent pas de là; ils résultent de ce que les destinataires de certaines lettres d'avis ont changé d'adresse; l'enquêteur doit alors interroger les ménages qui les remplacent dans les mêmes logements (1); la lettre d'avis n'a pas joint ces ménages et l'enquêteur a une tâche beaucoup plus difficile.

Effectivement, on voit (tableaux 6 et 6 *bis*) que les enquêteurs ont alors un succès très relatif (la quasi-totalité des refus se produisent dans ce cas). En outre, l'enquêteur a beaucoup de mal à trouver ces personnes à leur domicile.

TABLEAU 6

*Résultat de la visite des enquêteurs
dans le cas des logements dont le ménage occupant a changé depuis 1946*

PROPORTION %	PROVINCE			ENSEMBLE de la province	PARIS ville et banlieue
	Communes rurales	Ville de moins de 30.000 habitants	Ville de plus de 30 000 habitants		
D'acceptations.	75	70	73	73	24
De refus.	5	8	10	8	25
D'absences de longue durée. .	19	18	10	15	41
D'inconnus.	2	4	7	4	10
TOTAL	100	100	100	100	100

(1) On pourrait croire que c'est là un procédé incorrect, le nouvel occupant pouvant être très différent de l'ancien. Il n'en est rien pourtant et la méthode est des plus correctes, car l'unité de sondage que nous adoptons est le *logement* (et toutes les personnes y demeurant à l'époque du passage de l'enquêteur). On a tiré un *échantillon de logements*.

TABLEAU 6 bis

*Résultat de la visite des enquêteurs
dans le cas des logements dont le ménage occupant a changé depuis 1946*

PROPORTION %	COMMUNES RURALES				VILLES DE PROVINCE DE						PARIS ban- lieue	PARIS ville
	Cat. 1	Cat. 2	Cat. 3	Cat. 4	Moins de 5.000	5.000 à 10.000	10.000 à 30.000	30.000 à 100.000	100.000 à 150.000	Plus de 150.000		
D'acceptations	89	80	52	73	74	78	57	85	59	69	31	16
De refus	1	8	12	6	4	2	18	4	20	10	19	33
D'absences de longue durée	10	17	32	17	22	18	13	8	13	12	39	43
D'inconnus	—	—	4	4	—	2	12	3	8	9	11	8
TOTAL	100	100	100	100	100	100	100	100	100	100	100	100

Pour les enquêtes plus récentes (mais antérieures au recensement de 1954), on s'était efforcé de réduire les difficultés en mettant sur les lettres d'avis la mention (destinée aux P. T. T.) : « à retourner à l'expéditeur en cas de changement d'adresse », — puis en faisant rechercher le nom du nouvel occupant du local pour lui adresser également une lettre d'avis. La difficulté était que le nombre de tels ménages allait toujours croissant (faute d'un nouveau recensement), malgré les efforts faits pour mettre à jour les listes d'adresses tirées (consultation des secrétaires de mairie, des listes électorales, des bottins, etc.).

En définitive, les personnes ayant emménagé depuis 1946 échappaient plus au sondage que les autres. Cette déformation de l'échantillon n'est pas totalement négligeable car on constate, par exemple, qu'il s'agit de personnes *plus jeunes* que les autres (tableau 7). Toutefois, elles présentent une structure socio-professionnelle remarquablement voisine de celle de la population totale (sauf pour les agriculteurs, plus stables que le reste de la population).

TABLEAU 7

Cas des logements dont le ménage occupant a changé depuis 1946.

Répartition de leurs occupants suivant le sexe, l'âge (en %)

	SEXE		GROUPE D'ÂGE (ANNÉES)									
	Masc.	Fém	14-19	20-24	25-29	30-39	40-49	50-54	55-59	60-64	65-69	70 et plus
Ensemble des logements	47	53	10,7	7,5	7,6	15,4	20,5	10,3	7,7	6,0	5,3	9,0
Logements dont le ménage a changé.	46	54	9,1	9,4	15,9	22,5	20,3	9,4	3,9	3,2	4,5	4,8

Répartition de leurs occupants suivant la catégorie socio-professionnelle

	AGRICULTEURS	PETITS patrons	PATRONAT	PROFESIONS intellectuelles	TECHNICIENS Cadres moyens	EMPLOYES	OUVRIERS et Maîtrise	OUVRIERS agricoles	SERVICES	RETRAITÉS Rentiers
Ensemble des logements.	16,6	9,1	1,3	1,8	2,9	6,5	19,1	3,1	3,9	35,7
Logements dont le ménage a changé. . . .	11,3	11,1	0,9	1,9	3,2	6,4	20,8	2,9	4,6	36,9

CHAPITRE 3

Le tirage au sort des adresses et les visites à domicile

Le travail des enquêteurs, déjà pénible en cas de mauvais accueil, est rendu aussi assez fatigant par le nombre de démarches qu'il comporte : l'enquêteur doit se faire recevoir par des ménages dont l'adresse (tirée au sort par les Directions régionales de l'I. N. S. E. E.) lui est strictement imposée.

Si personne ne lui ouvre ou si la personne rencontrée au domicile n'est pas qualifiée pour fournir les renseignements demandés, l'enquêteur doit effectuer à la même adresse de nouvelles visites (1). Il est fréquent que trois visites soient nécessaires pour une seule adresse, surtout dans les grandes villes où le public travaille rarement à son domicile ou à proximité, et ceci entraîne des déplacements supplémentaires pour l'enquêteur, une sérieuse perte de temps et un surcroît de fatigue (l'enquêteur n'ayant jamais de voiture à sa disposition).

Le nombre de visites effectuées pour chaque « interview » est mentionné à la fois sur l'état des visites de l'enquêteur et sur le questionnaire lui-même. Voici ce que donne le dépouillement des questionnaires (sondage Emploi 1952) en cumulant les nombres de questionnaires remplis à la première visite, aux deux premières visites, aux trois premières visites, etc...

1^o Échantillon obtenu après une seule visite

Pourquoi ne pas limiter le travail de l'enquêteur à une seule visite, couronnée ou non de succès? C'est que l'échantillon ainsi obtenu ne serait pas représentatif. Pour s'en assurer il suffit de se reporter aux tableaux ci-après et de comparer la colonne « 1^{re} visite » aux dernières colonnes de chaque tableau (on verra plus loin ce qu'il faut entendre par « Résultats redressés »).

On voit d'abord au tableau 8 que l'échantillon qui aurait été ainsi obtenu, eût compris une proportion trop faible de ménages de 1 personne, contre un excès de ménages de 5 personnes et plus; la proportion finale aurait été à peu près respectée pour les ménages de 2 à 4 personnes.

(1) La technique d'enquête comporte d'ailleurs le dépôt d'un avis de passage dans la boîte aux lettres des enquêtés et la recherche d'heures de passage plus favorables à une rencontre.

TABEAU 8

Répartition de 1.000 ménages suivant leur nombre de personnes

NOMBRE DE PERSONNES par ménage	RÉSULTATS BRUTS					RÉSULTATS redressés
	1 ^{re} visite	1 ^{re} et 2 ^e visites	1 ^{re} , 2 ^e et 3 ^e visites	1 ^{re} , 2 ^e , 3 ^e et 4 ^e visites	Toutes visites	
1.	131	140	144	145	145	157
2.	271	273	272	272	272	277
3.	196	190	200	201	201	202
4.	158	156	156	156	155	151
5.	116	112	111	110	111	105
6 et plus.	128	118	117	116	116	109
ENSEMBLE	1.000	1.000	1.000	1.000	1.000	1.000

L'échantillon obtenu à la 1^{re} visite comprend nettement plus d'agriculteurs, de salariés agricoles et de personnes âgées que l'échantillon final; et c'est le contraire en ce qui concerne la proportion d'ouvriers, d'employés, cadres divers, étudiants, etc. (tableau 9).

La première visite fournit encore un excès de personnes de 14 à 19 ans (sexe féminin) et de plus de 60 ans (deux sexes), un déficit très net de personnes de 40 à 50 ans (tableau 10).

Elle fournit surtout un excès de personnes travaillant à domicile (tableau 11).

TABEAU 9

Répartition de 1.000 enquêtés suivant la catégorie socio-professionnelle

CATÉGORIE SOCIO-PROFESSIONNELLE	RÉSULTATS BRUTS					RÉSULTATS redressés
	1 ^{re} visite	1 ^{re} et 2 ^e visites	1 ^{re} , 2 ^e et 3 ^e visites	1 ^{re} , 2 ^e , 3 ^e et 4 ^e visites	Toutes visites	
Agriculteurs.	193	179	178	177	177	166
Artisans.	35	35	34	34	34	34
Petits commerçants.	58	58	57	57	57	57
Industriels.	2	3	3	3	3	3
Commerçants.	5	5	5	5	5	4
Professions libérales.	6	5	6	6	6	6
Clergé.	2	2	2	2	2	2
Cadres intellectuels.	3	3	3	3	3	3
Ingénieurs.	2	3	3	3	3	3
Cadres supérieurs.	8	9	9	10	10	10
Autres intellectuels.	11	11	11	11	11	11
Autres techniciens.	6	6	6	6	7	7
Cadres adm. moyens.	8	9	10	10	10	10
Employés.	45	47	49	49	49	52
Employés de commerce.	12	13	12	12	12	13
Contremaîtres.	7	8	8	8	8	8
Manœuvres.	28	30	30	30	30	31
Apprentis.	12	13	13	13	13	12
Ouvriers s.a.l.	117	128	130	130	130	135
Mineurs.	6	6	6	6	6	5
Salariés agricoles.	37	38	32	32	32	31
Gens de maison.	10	10	10	10	10	11
Femmes de ménage.	9	10	10	10	10	11
Huissiers et gardiens.	6	5	5	5	5	6
Autres gens de service.	5	5	5	5	5	6
Armée, police.	5	5	5	5	5	5
Étudiants.	36	37	38	38	38	38
Retraités.	41	42	42	42	42	42
Personnes âgées s.a.l.	76	72	71	71	71	71
Inaptes au travail.	22	20	21	21	20	21
Ménagères.	181	182	180	180	180	180
Autres personnes inactives.	6	6	6	6	6	6
TOTAL	1.000	1.000	1.000	1.000	1.000	1.000

TABEAU 10
Répartition de 1.000 enquêtés suivant l'âge

Sexe masculin

AGE	RÉSULTATS BRUTS					RÉSULTATS redressés
	1 ^{re} visite	1 ^{re} et 2 ^e visites	1 ^{re} , 2 ^e et 3 ^e visites	1 ^{re} , 2 ^e , 3 ^e et 4 ^e visites	Toutes visites	
14-19 ans.	118	118	118	118	118	115
20-29 ans.	150	149	148	148	148	149
30-39 ans.	154	155	156	157	156	155
40-49 ans.	204	210	210	210	210	211
50-59 ans.	181	181	181	181	181	184
60-69 ans.	110	106	106	106	106	106
70 ans et plus.	83	81	81	80	81	81
	1.000	1.000	1.000	1.000	1.000	1.000

Sexe féminin

AGE	RÉSULTATS BRUTS					RÉSULTATS redressés
	1 ^{re} visite	1 ^{re} et 2 ^e visites	1 ^{re} , 2 ^e et 3 ^e visites	1 ^{re} , 2 ^e , 3 ^e et 4 ^e visites	Toutes visites	
14-19 ans.	106	102	102	102	102	106
20-29 ans.	154	157	156	156	156	155
30-39 ans.	156	153	154	154	154	153
40-49 ans.	189	196	198	198	198	199
50-59 ans.	170	173	173	173	173	175
60-69 ans.	126	120	120	120	120	120
70 ans et plus.	99	99	97	97	97	98
	1.000	1.000	1.000	1.000	1.000	1.000

TABEAU 11
*Répartition de 1.000 adultes enquêtés
suivant la distance de leur domicile à leur lieu de travail*

DISTANCE	RÉSULTATS BRUTS					RÉSULTATS
	1 ^{re} visite	1 ^{re} et 2 ^e visites	1 ^{re} , 2 ^e et 3 ^e visites	1 ^{re} , 2 ^e , 3 ^e et 4 ^e visites	Toutes visites	
Nulle.	513	483	476	475	475	457
500 m au plus.	107	116	117	117	118	119
De 600 m à 1 km.	34	35	36	36	36	37
De 1 km à 1,9 km.	83	90	92	92	92	94
De 2 km à 4,9 km.	124	138	141	142	141	147
De 5 km à 9,9 km.	73	74	74	74	74	79
De 10 km à 19,9 km.	34	33	33	33	33	35
De 20 km et plus.	17	16	16	16	16	17
Variable { petites distances. . .	9	9	9	9	9	9
{ grandes distances. . .	6	6	6	6	6	6
	1.000	1.000	1.000	1.000	1.000	1.000

2° Variation du nombre de visites nécessaires

Le tableau 12 fait apparaître que dans les grandes villes, on ne remplit guère que la moitié des questionnaires à la première visite; en revenant plusieurs fois aux mêmes adresses, on n'arrive guère qu'à 75 % de succès à Paris, contre 90 % dans les autres grandes villes. Dans les communes rurales, on réussit dès la première visite environ 70 fois sur 100; avec des visites répétées, on obtient 95 % de succès.

TABLEAU 12

*Proportion pour 1.000 de questionnaires remplis
d'après le nombre de visites nécessaires pour chaque catégorie de communes*

CATÉGORIE DE COMMUNES où a lieu l'enquête		1 ^{re} visite	1 ^{re} et 2 ^e visites	1 ^{re} , 2 ^e et 3 ^e visites	1 ^{re} , 2 ^e , 3 ^e et 4 ^e visites	TOUTES visites
Communes rurales	Catégorie 1	765	965	970	970	970
	— 2	698	889	952	952	952
	— 3	703	898	925	926	928
	— 4	671	908	946	950	950
Villes de moins de 5.000 habitants.		719	918	938	940	940
Communes de 5.000 à 10.000 habitants.		675	905	951	956	956
— 10.000 à 30.000 habitants.		627	850	901	916	920
— 30.000 à 100.000 habitants.		525	853	946	957	962
— 100.000 à 150.000 habitants.		594	783	870	875	875
Plus de 150.000 habitants.		518	804	890	904	906
Paris ville.		513	707	740	750	750
Paris banlieue		591	745	774	774	776

3° Le redressement de l'échantillon

On vient de voir que malgré les visites réitérées des enquêteurs, il subsistait un « déchet », parfois notable 25 % à Paris) échappant à l'enquête par suite, soit d'un refus de répondre, soit surtout d'une absence de longue durée, soit aussi (et c'est surtout le cas pour les personnes seules) d'une présence à domicile à des heures très irrégulières.

On trouvera, au tableau 13, la répartition par régions et par catégories de communes, du « déchet » observé pour l'enquête Emploi 1952.

ÉTUDE DU REDRESSEMENT DE L'ÉCHANTILLON

Répartition des ménages manquants (et de ceux qu'on leur a substitués)

TABLEAU 13
Par grande région

GRANDES RÉGIONS	RÉGION pari- sienne	NORD	OUEST	NORD- EST	CENTRE	CENTRE- EST	MIDI méditer- ranéen	SUD- OUEST	TOTAL
Proportion (pour mille) de ménages manquants (1)	527	41	114	64	38	105	79	32	1.000

(1) La proportion de ménages substitués aux manquants est identique à celle-ci.

TABLEAU 13 bis

Par catégorie de communes

CATÉGORIE DE COMMUNES	COMMUNES RURALES				COMMUNES URBAINES DE PROVINCE ¹						PARIS	
	Catégorie 1	Catégorie 2	Catégorie 3	Catégorie 4	Moins de 5 000 habit.	De 5 000 à 10 000 habit.	De 10 000 à 30 000 habit.	De 30 000 à 100 000 habit.	De 100 000 à 150 000 habit.	Plus de 150 000 habit.	Ville	Banlieue
Proportion (pour mille) de ménages manquants (1). . .	28	32	73	39	47	36	73	34	47	92	265	125

(1) La proportion de ménages substitués aux manquants est identique à celle-ci.

On préfère ne pas ignorer les déformations de l'échantillon résultant de ce « *déchet* » et essayer de les réduire par des méthodes empiriques, c'est-à-dire où le calcul des probabilités est en défaut. L'une de ces méthodes consiste à se servir de coefficients de pondération convenablement choisis; par exemple, s'il y a 25 % de « *déchet* » à Paris et 5 % à la campagne, on multipliera les résultats de Paris par 1,33 et ceux des campagnes par 1,05. Cette méthode complique beaucoup le travail d'élaboration des résultats, elle n'est guère employée. On songe plutôt à rendre à l'échantillon une structure meilleure, en remplaçant les ménages du « *déchet* » par d'autres ménages *leur ressemblant*.

Dans certains sondages antérieurs, les remplacements avaient lieu sur le terrain et incombaient aux enquêteurs; mais on s'est vite aperçu que ce procédé était incompatible avec les prescriptions rigoureuses données par ailleurs à l'enquêteur de ne procéder à aucune substitution d'un ménage à un autre pour convenance personnelle.

Le procédé actuellement en vigueur consiste donc à remplacer le « *déchet* » par certains ménages déjà visités, en doublant les cartes perforées qui les représentent. Le même procédé est employé au Canada, mais avec des « *déchets* » beaucoup moins grands, car il accroît certainement les erreurs d'échantillonnage; il risque même de conduire à n'importe quel résultat, dès que le « *déchet* » est grand.

Les choix des critères de ressemblance entre les ménages est laissé à l'appréciation du statisticien, pour chaque enquête, dans la mesure où il s'agit bien de critères connus pour chaque ménage du « *déchet* ». Par exemple, si l'enquêteur connaît et indique la *catégorie sociale* du chef de ménage pour le « *déchet* », on peut songer à effectuer dans chaque commune-échantillon des remplacements qui conservent ce critère, du moins *dans la mesure du possible*. C'est ce qui a eu lieu effectivement pour l'enquête Emploi 1952, comme on le voit au tableau 14.

TABLEAU 14

Répartition (pour 1.000) des ménages manquants et de ceux qu'on leur a substitués, suivant la catégorie socio-professionnelle du chef de ménage

CATÉGORIE SOCIO-PROFESSIONNELLE du chef de ménage	MÉNAGES MANQUANTS dans l'échantillon		MÉNAGES SUBSTITUÉS aux ménages manquants
	(1)	(2)	
Agriculteurs	23	28	23
Petits patrons	79	97	103
Patronat	28	34	21
Professions intellectuelles	30	37	34
Techniciens, cadres moyens	52	64	60
Employés	118	145	148
Ouvriers et contremaîtres	251	307	358
Ouvriers agricoles	21	25	26
Services	60	74	56
Retraités. Rentiers	154	189	171
Inconnue	184	—	—
TOTAL	1.000	1.008	1.000

(1) Proportion pour 1.000 ménages manquants.

Proportion pour 1.000 ménages manquants dont la catégorie socio-professionnelle est connue.

L'opération était rendue quelque peu arbitraire par le fait qu'on ignorait la catégorie sociale du ménage défaillant dans 18 % des cas. Il semble qu'il en soit résulté finalement un accroissement du nombre d'ouvriers dans l'échantillon.

Un critère pour redresser l'échantillon — employé pour d'autres sondages de l'I. N. S. E. E. — est le nombre de personnes par ménage. Il y a encore quelques années, on se contentait du nombre figurant sur la base de sondage, c'est-à-dire vrai en 1946. Cette façon de faire ne tenait pas compte du fait que, dans notre pays où le nombre de logements diminue et le nombre d'habitants augmente, le nombre de personnes par ménage tend à s'accroître. C'est pourquoi on a prescrit, en 1952, aux enquêteurs de chercher à connaître le *nombre actuel* de personnes. Ceci fut un échec dans plus de la moitié des cas, et il n'a donc pas été possible de redresser l'échantillon suivant ce critère. Le tableau 15 donne la répartition des déchets et des ménages substitués suivant le nombre de personnes par ménage.

Malgré le doute résultant des ménages d'effectifs inconnus, il semble que les ménages substitués de 3 et 4 personnes soient plutôt en excédent.

TABLEAU 15

Répartition (‰) des ménages manquants et de ceux qu'on leur a substitués suivant le nombre de personnes du ménage

NOMBRE TOTAL DE PERSONNES DU MÉNAGE	MÉNAGES MANQUANTS dans l'échantillon		MÉNAGES SUBSTITUÉS aux ménages manquants
	(1)	(2)	
1 personne	156	371	268
2 —	122	290	330
3 —	69	185	223
4 —	38	89	101
5 —	28	67	47
6 — et plus	7	18	31
Inconnu	580	—	—
TOTAL	1.000	1.000	1.000

(1) Remarque analogue à celle du tableau 14.

(2) Remarque analogue à celle du tableau 14.

Quelles sont les conséquences du « *redressement d'échantillon* » ainsi effectué? Il faut, pour s'en rendre compte, se reporter aux dernières colonnes « résultats redressés » des tableaux 8 et 11 ci-dessus. On constate que le redressement conduit à un résultat « *extrapolé* » lorsque l'accroissement du nombre de visites entraîne l'évolution régulière d'une proportion (ou de toute autre statistique).

On s'est demandé si une *extrapolation graphique* donnerait les mêmes résultats, l'abscisse étant la proportion de ménages « interviewés », l'ordonnée étant la statistique dont il s'agit d'évaluer la valeur correspondant à 100 % de ménages interviewés.

L'extrapolation graphique se fait à l'aide des données ci-dessous en abscisses :

VISITES N°	1	1-2	1-2-3	1-2-3-4	1-2-3-4-5	BUT VISÉ
Proportions (%) de ménages atteints	64,5	86,1	90,5	91,0	91,1	100

Une première méthode consiste à porter toute une répartition (proportions cumulées) sur le même graphique où l'on extrapole (par exemple : répartition par âge, tableau 16 figure n° 1). On obtient par exemple :

FIG. 1. — Première méthode d'extrapolation graphique
Application à la répartition par âge de 1.000 enquêtés du sexe féminin

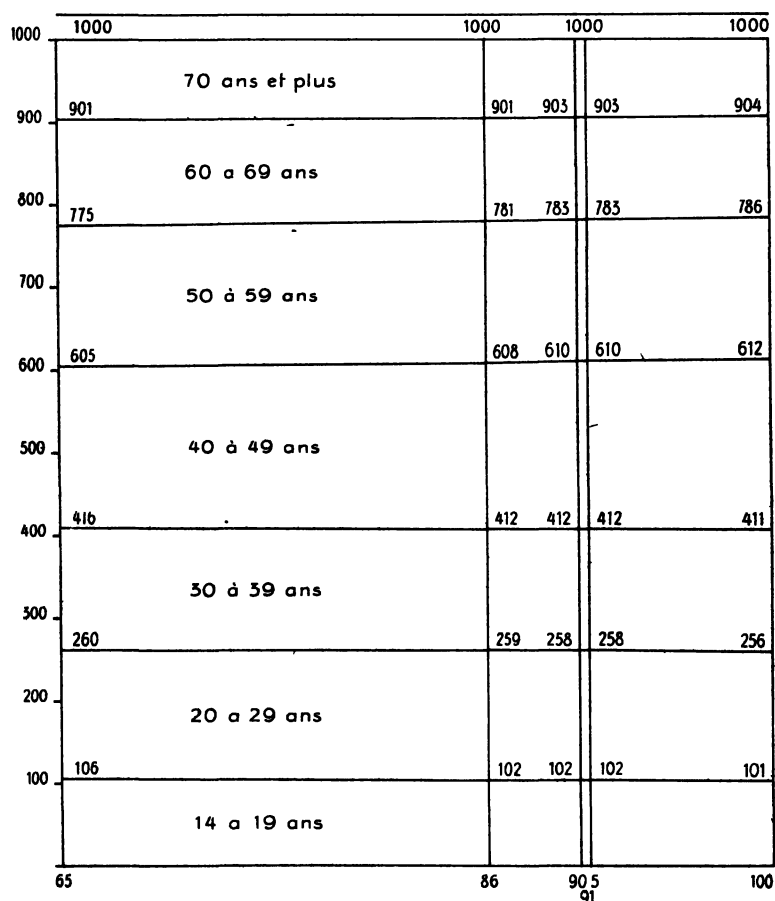


TABLEAU 16

Répartition de 1.000 enquêtés du sexe féminin suivant l'âge
(groupes d'âges cumulés). Résultats de l'extrapolation

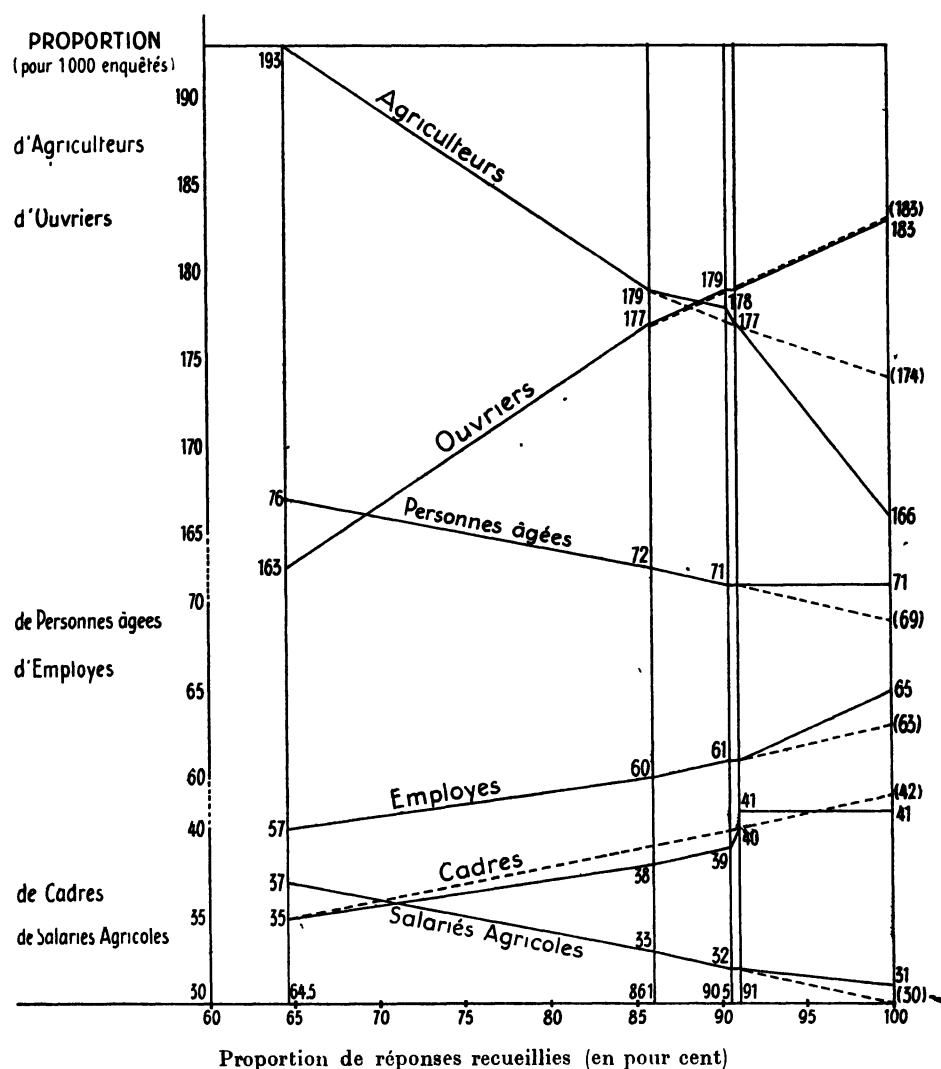
AGES	14-19	14-29	14-39	14-49	14-59	14-69	14 et plus
Par redressement	100	255	408	607	782	802	1.000
Par extrapolation.	101	256	411	612	786	904	1.000

Une seconde méthode consiste à extrapoler chaque groupe de la répartition indépendamment des autres (cf. fig. 2 : extrapolations en pointillés, redressements en trait plein).

FIG. 2

Seconde méthode d'extrapolation graphique

Application à la répartition par groupes socio-professionnels de 1.000 enquêtés



On l'a employée pour la *répartition socio-professionnelle* des enquêtés (pour laquelle l'ordre des professions reste arbitraire). La difficulté est de s'assurer que le total des proportions pour 1.000 reste égal à 1.000 après des extrapolations indépendantes.

La répartition socio-professionnelle comprend quelques groupes pratiquement *constants*; leur ensemble représente 439, 441, 440, 439 pour 1.000 du total (après la 1^{re}, la 2^e, la 3^e visite, après toutes les visites).

Les autres groupes représentent, au contraire, des proportions très sensibles au nombre de visites (tableau 17).

TABLEAU 17

Répartition de 1.000 enquêtés par catégories socio-professionnelles
Résultats de l'extrapolation graphique

PROPORTION (‰) APRÈS.		La 1 ^{re} visite	TOUTES visites	REDRES- SEMENT	EXTRA- POLATION
Tendance à décroître	Agriculteurs.	193	177	166	(174)
	Salariés agricoles.	37	32	31	30
	Personnes âgées.	76	71	71	69
		306	280	268	278
Tendance à croître	Ouvriers (toutes catégories).	163	179	183	183
	Employés <i>Idem</i>	57	61	65	68
	Cadres <i>Idem</i>	35	41	41	42
		255	281	289	288
Stabilité	Ensemble des autres catégories. . . .	439	439	443	439
	TOTAL	1.000	1.000	1.000	1.000

REMARQUE 1. — Le redressement avait conduit à une anomalie : les groupes « constants » s'étaient trouvés portés de 439 à 443 ‰; l'extrapolation les maintient au contraire à 439.

REMARQUE 2. — Pour les ouvriers, le redressement et l'extrapolation conduisent au même résultat. En revanche, pour les agriculteurs, le redressement semble exagérer leur chute, mais l'extrapolation ne donne rien de sûr. On a donc choisi la proportion de 174 ‰ pour assurer un total général égal à 1.000.

En résumé, nous préférons, en principe, le procédé « d'extrapolation » au « redressement »; mais il faut tenir compte des facteurs suivants :

1^o extrême commodité du dépouillement après redressement; lenteur des extrapolations;

2^o arbitraire possible des extrapolations (qu'on peut vouloir effectuer à l'intérieur de chaque grande région, de chaque strate, etc...).

DEUXIÈME PARTIE

SUR LES ERREURS D'ÉCHANTILLONNAGE

On rappellera, tout d'abord, que l'emploi des méthodes mathématiques de sondage permet, en principe, d'estimer la *variance* de chaque estimation fournie par le sondage, et que cette variance donne une idée de la précision du résultat. La racine carrée de la variance est appelée *erreur-type* (1), la valeur relative de l'erreur-type est appelée *coefficient de variation* et s'exprime généralement en %.

Dire qu'un résultat est évalué sur échantillon avec un coefficient de variation de 7 % par exemple, ne signifie pas que l'erreur d'échantillonnage est de 7 %, ni qu'elle est inférieure à 7 %. Cela signifie seulement qu'environ 95 fois sur 100 l'erreur d'échantillonnage est inférieure à 14 % (soit deux fois l'erreur type).

Bien entendu, pour le même sondage, chaque résultat a sa variance propre.

Il convient d'ajouter que le calcul de la variance est toujours long et peu précis; de sorte que, au lieu de calculer la variance de chaque évaluation, on est obligé (faute de moyens de calcul) de se limiter aux variances des résultats les plus importants; et s'il en est ainsi dans tous les pays, c'est encore plus le cas pour les sondages de l'I. N. S. E. E. (qui ne possède aucun calculateur étudiant pour la licence de mathématiques, comme il s'en rencontre dans les laboratoires de statistique).

Comme de toute façon l'erreur-type d'échantillonnage sera assez mal connue, il paraît peu utile d'employer à son calcul des formules absolument rigoureuses faisant intervenir des grandeurs inconnues; et on utilise dans tous les pays des procédés assez approximatifs.

A la suite des enquêtes sur l'Emploi de 1950, nous avons procédé à des calculs d'erreurs, dont les résultats ont été publiés au Supplément trimestriel du B. M. S. (avril-juin 1951, page 22) et qui ont servi à orienter les changements à apporter aux plans d'enquêtes ultérieures; pratiquement, il a été tenu compte de ces résultats pour l'enquête sur l'Emploi de 1952 et toutes les enquêtes de 1953. Celles-ci font à leur tour l'objet de calculs d'erreur, etc.

L'idéal serait de pouvoir publier côte à côte les résultats de l'enquête et des calculs d'erreur.

Malheureusement, ceci retarderait notablement la publication des résultats; et il a fallu y renoncer une fois pour toutes à l'I. N. S. E. E. Les calculs d'erreur sont donc effectués dans les temps morts, tant pour la partie mécanique que pour la partie manuelle du travail.

* * *

Voici à présent les résultats que nous avons obtenus sur les erreurs d'échantillonnage des études de marché de l'I. N. S. E. E. en 1953. Ces calculs mettent

(1) Ou erreur quadratique moyenne, parmi toutes les erreurs d'échantillonnage possibles avec les échantillons de même taille tirés de la même population avec le même plan d'échantillonnage.

en évidence les faiblesses de notre plan d'échantillonnage et sont par conséquent destinés à provoquer des modifications dans les méthodes de l'I. N. S. E. E. (Ceci fera l'objet du chapitre 1 ci-après).

Nous avons saisi d'ailleurs cette occasion pour expérimenter deux méthodes indirectes de calcul d'erreurs-types, inspirées par F. Yates et notablement plus simples que la méthode directe que nous employions jusqu'ici. (Ceci fera l'objet du chapitre 2). Nous comparerons les résultats des deux méthodes.

CHAPITRE 1

LE CALCUL « DIRECT » DES ERREURS-TYPES

I. — La méthode

Nous avons rétabli les formules concernant le calcul d'erreurs, pour un sondage du type étudié ici, dans l'Étude théorique n° 6 de l'I. N. S. E. E. (Application des méthodes de sondage aux enquêtes statistiques, pages 171-4) (1); et leur application numérique directe fait, en outre, l'objet d'une plaquette ronéo de l'I. N. S. E. E. (Calculs d'erreur dans un sondage à deux degrés). On n'y reviendra donc pas ici. Indiquons simplement que nous évaluons *séparément* deux termes appelés, l'un " variance à l'intérieur des communes ", l'autre " variance entre communes-échantillon ", — qu'il faut additionner pour évaluer la variance proprement dite.

Avec les notations suivantes dans chaque strate :

Unité i du 1 ^{er} degré	}	Vraie moyenne du caractère $x : \bar{x}_i$	}	(Moyennes par unité du 1 ^{er} degré)
		Moyenne de x sur l'échantillon : $\bar{X}_i$		
Moyenne des $\bar{x}_i$	}	Population $\bar{\bar{x}}$	}	Échantillon $\bar{\bar{x}}$
— des $\bar{X}_i$		$\bar{\bar{X}}$		$\bar{\bar{X}}$

la variance de l'estimation $\bar{\bar{X}}$ dans la strate est :

$$\begin{aligned} E(\bar{X} - \bar{\bar{x}})^2 &= E(\bar{X} - \bar{\bar{X}})^2 + E(\bar{\bar{X}} - \bar{\bar{x}})^2 \\ &= E(\bar{X} - \bar{x})^2 + E(\bar{x} - \bar{\bar{x}})^2. \end{aligned}$$

C'est la première de ces deux décompositions qui est effectivement employée ici. Pour un système de strates (h) on aura :

$$V(\bar{X}) = \sum_h E(\bar{X}_h - \bar{\bar{x}}_h)^2.$$

Nous estimons séparément (par évaluation directe des paramètres qu'elles renferment) les expressions :

$$\begin{aligned} \sum E(\bar{X} - \bar{\bar{X}})^2 &\text{ variance entre les communes-échantillon} \\ \sum E(\bar{\bar{X}} - \bar{\bar{x}})^2 &\text{ variance à l'intérieur des communes} \end{aligned}$$

(1) Voir aussi P. G. GRAY et T. CORLETT, *Sampling for the Social Survey* (formules 5 et 7) *Journal of the Roy. Stat. Soc.*, 1950, II, p. 150-206.

II. — *Les résultats* (voir tableau 18)

On a limité les calculs aux caractères *x* suivants de l'enquête *Chaussures* :

- Nombre d'hommes,
- Nombre de femmes,
- Nombre de paires de souliers d'hommes achetées,
- Nombre de paires de souliers de femmes achetées.

En outre, on s'est occupé des *deux achats moyens* :

- Nombre moyen de paires de souliers d'hommes, par homme,
- de femmes, par femme.

Pour abrégé, on a admis dans ce dernier calcul, que le coefficient de corrélation entre le nombre de paires de souliers achetées et le nombre de personnes correspondantes *d'une commune* était de 75 % pour toutes les strates; de sorte que la précision avec laquelle sont évalués les achats moyens n'est que très grossièrement connue.

Les sommes de carrés et de produits (qui permettent d'estimer les divers écarts-types et coefficients de corrélation intervenant dans les formules) ont été calculées par la mécanographie (cartes perforées). On avait exclu au préalable les cartes en double (créées pour redresser l'échantillon).

Nous n'avons pu faire effectuer aucun calcul analogue pour l'enquête Assurances, en raison de l'absence de l'indication des communes-échantillon dans les cartes perforées servant au dépouillement (indication supprimée pour rester dans le cadre des 80 colonnes de la carte).

TABLEAU 18

Calcul direct — Coefficient de variation (en %) de l'estimation

	DU NOMBRE		DU NOMBRE de paires de chaussures		DE L'ACHAT MOYEN PAR		
	d'hommes	de femmes	d'hommes	de femmes	homme	femme	
France entière.	1,0	1,0	3,4	3,5	2,8	2,8	
<i>Par grande région :</i>							
Région parisienne.	2,6	2,2	5,0	5,5	3,6	4,2	
Nord	3,5	3,7	9,2	6,4	6,9	4,4	
Ouest.	3,5	2,2	9,7	10,2	7,5	8,7	
Est.	3,9	3,0	9,4	7,7	7,0	5,7	
Centre.	3,3	3,3	14,8	9,8	12,5	7,7	
Centre-Est.	3,3	3,2	6,2	7,3	4,4	5,4	
Midi Méditerranéen.	4,3	2,6	8,5	6,6	6,8	5,0	
Sud-Ouest.	2,8	3,0	14,7	16,6	12,8	14,5	
<i>Par catégorie de communes :</i>							
<i>Population agricole</i>							
Communes rurales	Plus de 80 %	2,4	2,0	11,5	14,8	9,9	13,4
	40 à 60 %	4,8	4,7	9,8	8,9	6,9	6,2
	20 à 40 %	5,2	2,4	14,2	10,0	10,9	8,4
	moins de 20 %	3,8	3,7	13,3	12,8	11,1	10,3
<i>Habitants</i>							
Communes urbaines	Moins de 5.000.	3,6	3,5	10,6	7,6	8,4	5,6
	5 à 10.000.	4,6	3,9	11,8	10,4	8,8	7,8
	10 à 30.000.	4,6	2,2	11,1	9,9	8,2	8,4
	30 à 100.000.	4,1	4,3	11,9	11,3	9,8	8,7
	100 à 150.000.	5,8	4,6	10,6	8,8	8,1	8,1
	Plus de 150.000.	2,4	2,0	4,1	4,3	2,8	4,2

Au sujet des deux composantes de la Variance

Il paraît utile de signaler le rôle extrêmement différent de chacune des composantes :

$$\Sigma E (\bar{X} - \bar{\bar{X}})^2$$

est prépondérante pour les catégories de communes rurales; son rôle diminue progressivement quand on passe aux strates de petites et de grandes villes, mais reste important à cause de la banlieue de Paris.

$$\Sigma E (\bar{X} - \bar{x})^2$$

joue bien entendu le rôle inverse; en particulier elle est absolument négligeable pour les strates de petites communes et n'est prépondérante que pour les villes de 100.000 à 150.000 habitants. Les grandes régions pour lesquelles elle représente une part importante de la variance sont la *Région Parisienne* et dans une moindre mesure le Centre et le Midi Méditerranéen (voir tableau 19),

TABLEAU 19

Importance (en %) de $\Sigma E (\bar{X} - \bar{x})^2$ dans $V(\bar{X})$

	ACHAT		PAR CATÉGORIE DE COMMUNES	ACHAT		
	Hommes	Femmes		Hommes	Femmes	
France entière.	4,7	4,0				
Par grandes régions :						
Région parisienne.	32,0	21,0	Com- munes rurales	Pop. agricole :		
Nord	3,2	6,1		Plus de 60 %	0,01	0,05
Ouest.	1,6	1,0		40 à 60 %	0,5	0,4
Est.	3,3	2,4		20 à 40 %	0,1	0,1
Centre	1,2	2,3		Moins de 20 %	0,3	0,2
Centre-Est.	12,5	6,5	Com- munes urbaines	Population		
Midi Méditerranéen.	11,5	16,7		Moins de 5.000 h.	0,6	1,0
Sud-Ouest.	1,6	1,2		5.000 à 10.000 h.	1,2	1,0
				10.000 à 30.000 h.	2,7	2,6
				30.000 à 100.000 h.	8,7	4,9
				100.000 à 150.000 h.	70,7	77,1
				Plus de 150.000 h.	44,6	31,7

CHAPITRE 2

Essai de deux méthodes indirectes de calcul d'erreur (F. Yates)

Chacune des études de marché de l'I. N. S. E. E. en 1953 présentait cette particularité de comporter *deux modèles distincts de questionnaires*.

Pour le sondage Assurances, les pages 1 des deux questionnaires étaient identiques, mais les pages 2, 3 et 4 différaient totalement. Cette technique permet, avec un questionnaire (relativement bref) de quatre pages, de recueillir des données sur des questions qui exigeraient autrement un questionnaire de sept pages, c'est-à-dire un « interview » fatigant pour l'enquêteur et l'enquêté (d'où une baisse de qualité des réponses).

Pour le sondage Chaussures, on utilisait également *deux types de questionnaires*, certaines questions étant posées de façon différente (ou dans un ordre différent) en vue de tester la validité de certaines réponses.

Dans chacun des cas, les instructions données aux enquêteurs tendaient à répartir aussi exactement que possible les deux types de questionnaires entre les enquêtés. L'échantillon se compose donc de deux sous-échantillons comparables l'un à l'autre; et il est possible de se servir de cette singularité pour chercher à évaluer les erreurs d'échantillonnage relatives aux *questions communes aux deux types* de questionnaires.

On trouve à cet effet dans le manuel de Frank Yates (*Sampling Methods for Censuses and Surveys*, 1949, § 7.25, page 242 de l'édition anglaise) deux formules de calcul d'erreur :

— l'une, réputée peu précise, mais valable en tout état de cause, — sera désignée ici sous le nom de *formule a*;

— l'autre, réputée plus précise mais dont la validité suppose une certaine hypothèse de structure, — sera appelée *formule b*.

Nous allons donner d'abord les résultats de ces calculs (indirects) d'erreurs. Nous examinerons ensuite quelle valeur présentent ces résultats.

I. — Résultats des calculs

A. — Étude du marché de la chaussure

Les calculs faits ne concernent que la question suivante :

(3) Combien de paires de chaussures ont été achetées en 1952 pour les besoins de votre famille? (en distinguant : chaussures de cuir et autres, chaussures d'hommes et de femmes).

TABLEAU 20

Coefficients de variation (en %) du nombre de paires de chaussures achetées

Formule employée.	CHAUSSURES D'HOMMES				CHAUSSURES DE FEMMES			
	En cuir		Autres		En cuir		Autres	
	a	b	a	b	a	b	a	b
France entière.	1,7	1,9	2,7	2,4	1,4	1,6	2,9	2,3
<i>Par grande région</i>								
Région parisienne.	2,3	5,8	4,3	3,4	1,5	3,0	5,6	6,2
Nord.	4,5	4,8	6,3	5,0	3,7	4,4	7,4	8,2
Ouest.	5,1	4,1	4,7	3,8	5,4	5,6	3,6	3,4
Est.	3,8	4,7	8,7	7,6	3,6	3,2	5,9	5,6
Centre.	6,4	6,6	4,9	7,0	6,7	6,6	7,0	7,2
Centre-Est.	4,6	4,6	10,3	10,2	3,3	3,8	8,3	6,6
Midi Méditerranéen.	6,5	5,9	8,6	8,6	3,6	2,8	9,8	10,8
Sud-Ouest.	6,8	5,3	8,4	7,4	6,1	7,2	10,8	6,6
<i>Par catégorie de communes :</i>								
Population agricole :								
Com- Plus de 60 %	4,0	4,4	6,5	5,4	5,1	5,6	8,4	7,6
munes 40 à 60 %	9,8	11,1	5,2	3,6	5,5	5,4	7,0	7,0
rales 20 à 40 %	4,9	3,1	10,5	8,2	6,2	5,8	7,4	9,6
Moins de 20 %	5,1	6,6	8,5	8,8	7,0	6,4	8,9	7,8
Habitants								
Com- Moins de 5.000.	6,1	6,4	7,7	7,0	7,7	9,0	8,5	6,4
munes 5 à 10.000.	6,1	5,8	9,4	7,8	5,0	5,8	3,4	3,4
10 à 30.000.	3,6	3,9	8,1	11,8	6,0	5,6	10,4	10,2
urbaines 30 à 100.000.	7,3	7,7	11,0	13,6	3,7	3,6	3,5	2,8
100 à 150.000.	11,4	10,8	14,7	4,8	4,1	3,4	17,4	16,2
Plus de 150.000.	1,7	2,9	3,7	5,6	0,9	1,2	4,6	4,4

Les coefficients de variation ont été calculés pour chacune de ces quatre catégories de chaussures, avec les formules (a) et (b) de Yates. On les a établis non seulement pour la France entière, mais aussi :

- pour chaque grande région, d'une part,
- pour chaque catégorie de communes, d'autre part.

Ils sont donnés au tableau 20 qui fait apparaître :

- d'une part, l'importance des erreurs affectant les estimations par grande région et par catégorie de communes (alors que les estimations France entière ont une précision satisfaisante);
- d'autre part, la concordance générale entre les résultats des formules (a) et (b).

B. — *Étude du marché des Assurances*

Les calculs faits concernent deux questions posées au chef de famille : les *Retraites* et l'*Assurance-Vie*. D'une part, après avoir demandé la profession détaillée, la question suivante était posée (à la page 1) :

Cette profession vous donne-t-elle droit à une retraite? Oui — Non. Si oui, quel régime?

Code des réponses : 0 *Pas de retraite.*

- | | | |
|---|---|--|
| — | 1 | <i>Retraite de fonctionnaire.</i> |
| — | 2 | — <i>de militaire.</i> |
| — | 3 | — <i>des cadres.</i> |
| — | 4 | — <i>d'artisans, commerçants, professions libérales.</i> |
| — | 5 | — <i>mines, É. D. F., S. N. C. F.</i> |
| — | 6 | — <i>Sécurité sociale, Vieux Travailleurs.</i> |
| — | 7 | — <i>Autres retraites.</i> |
| — | 8 | — <i>Retraites s. a. i.</i> |

La seconde question faisant l'objet d'un calcul d'erreur était plus délicate, car on la posait sous une forme légèrement différente dans les deux questionnaires.

Le questionnaire du type 1 comportait (page 2) une étude détaillée de l'*Assurance-Vie* et notamment la question suivante :

Question 5. — *Êtes-vous titulaire d'une (ou plusieurs) police personnelle d'assurance sur la vie, souscrite par vous-même auprès d'une (ou plusieurs) société d'Assurances?*

Réponse : *Oui (une Société). Oui (plusieurs). Non.*

(Cette question faisait suite à une autre relative à l'*Assurance Groupe*).

En revanche, le questionnaire de type 2 ne comportait (page 3) qu'une brève question sur l'*Assurance-Vie*.

Question 17. — *Avez-vous souscrit d'autres polices auprès d'une société d'Assurances?*

a : *Assurance personnelle sur la vie? Oui — Non.*

b, c, d...

(Cette question faisait suite aux questions détaillées sur les *Assurances Incendie, Vol, Dommages aux tiers*).

On a extrait en définitive des réponses aux questions 5 (type 1) ou 17 (type 2) la réponse à la question unique :

Avez-vous une (ou plusieurs) polices d'assurance personnelle sur la Vie? Oui
— *Non.*

Et on a comparé les résultats des questionnaires type 1 et type 2.

On ne devra pas perdre de vue que l'estimation des erreurs d'échantillonnage qui en résulte pêche par excès, puisque la différence entre les façons de poser la question aux deux types d'enquête tend à exagérer le désaccord entre les deux groupes de réponses (en particulier, avec le type 2, risques d'oubli assez grands et aussi risques de confusion entre l'assurance-groupe et l'assurance personnelle).

Les coefficients de variation calculés sont donnés au tableau 21.

On a limité les calculs à la première enquête, celle dans les villes de plus de 30.000 habitants (environ 3.000 ménages interrogés).

On remarquera l'ampleur considérable de l'erreur d'échantillonnage relative aux retraites *assez rares* (retraites militaires, retraites des cadres), alors que les retraites très fréquentes (vieux travailleurs ou Sécurité sociale) n'ont bien entendu qu'une faible erreur aléatoire.

En outre, restent valables les remarques faites à propos de l'étude du marché de la chaussure, concernant soit les résultats par grande région ou par catégorie de communes, soit l'emploi des formules (a) et (b).

II. — EN QUOI CONSISTAIT LE CALCUL INDIRECT — SA VALIDITÉ

A. — Les formules employées

Étant donné l'existence de deux échantillons superposés (qu'on désignera par un ou deux accents), soit par exemple $\mathcal{Y}$ le nombre de personnes qui ne sont pas assurées sur la vie dans la *région parisienne*, avec

$\mathcal{Y}_1$ pour la ville de Paris, $\mathcal{Y}_2$ pour la banlieue

$$\mathcal{Y} = \mathcal{Y}_1 + \mathcal{Y}_2$$

Les deux échantillons fournissent les estimations (1) suivantes (fraction sondée = 1/1.500) :

Avec le :

	1 ^{er} échantillon	2 ^e échantillon	D'où :
Pour $\mathcal{Y}_1$	$Y'_1 = 590 \times 1.500$	$Y'_1 = 637 \times 1.500$	$Y'_1 - Y''_1 = -47 \times 1.500$
Pour $\mathcal{Y}_2$	$Y'_2 = 568 \times 1.500$	$Y''_2 = 585 \times 1.500$	$Y'_2 - Y''_2 = -17 \times 1.500$

On estime finalement $\mathcal{Y}$ par $Y = Y_1 + Y_2$, c'est-à-dire

$$, \frac{Y'_1 + Y'_2}{2} + \frac{Y''_1 + Y''_2}{2}$$

(1) On a tenu compte du « déchet » (voir 1^{re} partie) en admettant qu'il possède les mêmes caractères que l'échantillon et en majorant proportionnellement les estimations Y'_1, Y'_2, Y''_1, Y''_2 . (ce qui est tout à fait arbitraire).

TABLEAU 21

Enquête assurances. Villes de plus de 30.000 habitants

Coefficients de variation (en %) des proportions d'enquêtés ayant répondu comme suit :

RUBRIQUE	Pas de RETRAITE	RETRAITE										ASSURANCE-VIE								
		de fonction- naires		des militaires		des cadres		artisans, commerçants, professions libérales		E.D.F., S.N.C.F., etc.		des Vieux Travailleurs et S. S.		autres retraites		non assurés		assurés		
		a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b	
Formule employée	a	b															a	b	a	b
France entière	4,4	5,0															1,3	1,2	6,8	4,8
Par grande région :																				
Région parisienne	0,2	0,1	8,3	8,6	45,6	45,6	21,9	19,4	11,5	11,8	28,3	28,2	4,3	4,2	6,6	7,2	2,1	2,1	15,2	16,0
Nord	16,3	17,6	17,3	15,4	46,0	48,8	14,0	27,2	17,8	18,2	16,0	20,6	14,2	17,6	27,2	30,4	0,4	0,3	0,7	0,6
Ouest	18,8	18,4	15,4	18,4	60,5	57,8	17,1	16,6	29,7	28,8	13,1	12,2	5,7	5,6	62,0	45,8	6,1	6,4	10,5	11,6
Est	24,7	24,4	30,6	37,2	61,0	57,8	17,6	17,2	9,4	17,4	29,1	28,4	11,4	11,2	36,2	20,8	5,0	5,0	24,2	25,0
Centre	15,7	18,8	8,7	8,2	34,5	40,4	70,7	70,6	21,5	23,6	14,4	15,6	11,4	11,2	37,9	37,0	3,5	5,6	10,8	12,2
Centre-Est	10,9	10,8	27,0	26,4	0	0	35,9	39,8	22,9	18,6	21,1	23,0	6,4	9,8	3,1	4,2	3,7	3,0	12,8	10,8
Midi Méditerranéen	27,9	22,8	7,3	10,8	21,2	26,6	21,2	27,4	19,3	18,2	23,2	23,8	7,3	9,6	35,5	29,0	2,5	2,2	12,9	10,8
Sud-Ouest	14,8	17,8	33,6	29,0	74,0	70,6	39,2	44,4	9,4	9,6	33,5	37,8	23,0	21,8	14,3	7,2	4,2	3,9	15,6	14,6
Par catégorie de communes :																				
30.000 à 100.000 habitants	12,1	11,8	5,1	5,8	32,2	29,8	28,8	28,4	12,2	12,4	11,2	13,2	9,0	9,8	20,9	17,0	2,7	3,0	7,6	7,6
100.000 à 150.000 habitants	16,5	17,0	15,3	18,0	50,1	44,6	25,8	29,2	30,1	22,8	20,3	20,4	7,3	9,2	50,2	30,0	2,8	2,8	7,1	7,8
Plus de 150.000 habitants	4,8	6,0	7,6	9,2	29,6	23,8	17,0	9,4	7,8	7,0	15,5	13,2	3,4	3,6	6,1	7,6	1,6	1,4	9,8	7,8

(a) On pourra estimer la variance de Y par la formule (a)

$$[(Y'_1 - Y''_1)^2 + (Y_2 - Y''_2)^2] / 4,$$

et son coefficient de variation par

$$\sqrt{(Y'_1 - Y''_1)^2 + (Y_2 - Y''_2)^2} / (Y'_1 + Y''_1 + Y_2 + Y''_2)$$

c'est-à-dire : $(2.209 + 289)^{1/2} / 2.380 = 49,98/2.380 = 2,1 \%$

(b) On peut encore faire l'hypothèse de structure suivante : les variantes $V(Y_1)$, $V(Y_2)$, etc... seraient proportionnelles à $\mathcal{Y}_1$, $\mathcal{Y}_2$, etc...

On a :

$$Y = Y_1 + Y_2$$

$$V(Y) = V(Y_1) + V(Y_2)$$

et $V(Y)$ serait proportionnelle à $\mathcal{Y}$. Les données fournissent deux estimations du rapport de proportionnalité, à savoir : d'une part

$$\frac{(Y'_1 - Y''_1)^2}{4 Y_1} = \frac{(Y_1 - Y''_1)^2}{2 (Y_1 + Y''_1)} = \frac{2.209}{2 \times 1.227} = \frac{1,8003}{2}$$

d'autre part

$$\frac{(Y'_2 - Y''_2)^2}{4 Y_2} = \frac{(Y_2 - Y''_2)^2}{2 (Y_2 + Y''_2)} = \frac{289}{2 \times 1.153} = \frac{0,2507}{2}$$

et une estimation de $\mathcal{Y}$:

$$\frac{1.227 + 1.153}{2} = 1.190.$$

De sorte qu'on peut estimer $V(Y)$ par la formule (b)

$$\frac{1,8003 + 0,2507}{4} \times 1.190 = \frac{2,0510 \times 1.190}{4}$$

et le coefficient de variation par la formule :

$$\frac{1}{2} \sqrt{\frac{2,0510}{1.190}} = 2,08 \%$$

La formule algébrique (b) est d'ailleurs :

$$\text{estimation de } V(Y) = \left[\frac{(Y'_1 - Y''_1)^2}{Y_1 + Y''_1} + \frac{(Y'_2 - Y''_2)^2}{Y_2 + Y''_2} \right] \cdot (Y'_1 + Y''_1 + Y'_2 + Y''_2) \cdot \frac{1}{8}$$

Généralisation : On a calculé $(Y'_1 - Y''_1)^2$ pour chacune des catégories de communes de chaque grande région formant un tableau rectangulaire, et on les a regroupés : horizontalement par catégories de communes, verticalement par grande région. D'où les résultats donnés plus haut.

(1) Cf. *The Danish Labour Force Surveys, by the Statistical Department. Copenhagen (Kungl. Boktr. Stklm., 1953, pp. 259-267).*

B. — Formule (a) ou formule (b)?

La formule (b) repose sur une hypothèse de structure qui ne s'impose pas, mais n'est pas personnelle à M. Yates; on en fait usage dans plusieurs méthodes de *calcul approché* des erreurs d'échantillonnage (1). Elle correspond en fait à un usage plus ou moins justifié de la loi de Poisson.

On sait qu'une telle loi est la limite de celle suivie par le nombre X de boules blanches tirées lorsqu'on fait n tirages indépendants dans une urne où la proportion de blanches est p , si n devient très grand et p très petit, $m = np$ restant constant.

Alors la moyenne de X est m et son écart-type est $\sqrt{m}$.

Toute variable $Y = kX$, proportionnelle à X , a pour moyenne km et écart-type $k\sqrt{m}$; sa moyenne est proportionnelle à la racine carrée de son écart-type.

Considérons un caractère relativement rare, comme le fait d'avoir droit à une retraite des cadres. Si, dans une strate donnée, on interroge n personnes tirées au sort indépendamment les unes des autres, il est permis d'admettre que le nombre X de celles qui comptent bénéficier d'une retraite des cadres suivra une loi de Poisson. En revanche, pour un caractère relativement courant (retraite des Vieux Travailleurs) on revient à une « loi binomiale » dont la variance est inférieure à celle d'une loi de Poisson. Si l'on passe à une statistique d'un autre type, comme celle du nombre de paires de chaussures achetées, il faut tenir compte du fait qu'un ménage tiré au sort peut avoir acheté dans l'année, par exemple, plusieurs paires de chaussures d'homme en cuir, ce qui introduit un « effet de grappe » modifiant la variance. En outre, l'achat de chaussures n'est pas un événement tellement rare que la loi de Poisson puisse paraître s'imposer.

Ceci dit, on remarque que l'hypothèse des tirages indépendants n'est pas valable. En fait, les ménages sont tirés au sort à l'intérieur des communes-échantillon préalablement désignées, dont chacune joue le rôle d'une strate, la variance se trouvant un peu réduite (par rapport à celle d'une loi de Poisson).

En définitive, l'emploi de la formule (b) n'est pas strictement correct, mais ne doit pas entraîner d'erreur trop grande.

En pratique, on constate ici une étroite concordance entre les résultats des deux formules pour l'enquête Assurances et une concordance moins nette pour l'enquête Chaussures.

Le désaccord entre les deux formules est parfois assez net, et ceci se produit lorsque l'évaluation du coefficient de variation dans une strate varie beaucoup entre les strates d'un même domaine d'étude (par exemple entre les catégories de communes d'une même grande région). En effet (a) et (b) sont deux moyennes (l'une simple, l'autre pondérée) des expressions de la forme

$$\frac{1}{2} (Y_1 - Y_1'')^2, \frac{1}{2} (Y_2 - Y_2'')^2, \dots$$

et les deux résultats ne peuvent être bien différents que si les termes soumis à ces moyennes diffèrent eux-mêmes profondément.

NOTA. — La formule (a) est celle des deux qui se calcule le plus rapidement.

C. — Les formules (a) et (b) ne donnent qu'une partie de la variance

Les indications assez brèves du manuel M. Yates pourraient faire croire que les formules (a) ou (b) fournissent une estimation de la variance complète. Il n'en est rien. Les calculs du chapitre 2 ne permettent d'estimer qu'une partie de la variance :

Variance (entre ménages) à l'intérieur de l'échantillon de communes.

Autrement dit, avec les notations du chapitre 1, les formules (a) ou (b) permettent d'estimer :

$$\Sigma E (\bar{X} - \bar{x})^2$$

et il manque pour obtenir $V(\bar{X})$ un deuxième terme, soit

$$\Sigma E (\bar{X} - \bar{x})^2.$$

On observera que cette décomposition de $V(\bar{X})$ n'est pas celle que nous avons employée pour les calculs du chapitre 1 :

$$\Sigma E (\bar{X} - \bar{\bar{X}})^2 + \Sigma E (\bar{\bar{X}} - \bar{x})^2.$$

Pour se faire une idée de l'importance du terme négligé, on a calculé également avec la formule (a) les coefficients de variation relatifs aux nombres de paires de chaussures d'hommes ou de femmes achetées (quelle que soit la matière avec laquelle elles sont fabriquées). Les résultats figurent au *tableau 22* qu'il faut comparer à présent au *tableau 18*.

TABLEAU 22

Formule (a). Achats de chaussures de toutes catégories
Coefficients de variation (en %)

	CHAUSSURES			CHAUSSURES	
	d'hommes	de femmes		d'hommes	de femmes
France entière.	1,6	1,5	Par catégorie de communes :		
Par grande région :			Population agricole :		
Région parisienne.	2,1	2,2	Com- { Plus de 60 % . . .	4,3	6,0
Nord	4,3	3,6	munes { 40 à 60 % . . .	6,2	2,8
Ouest.	4,6	2,5	rurales { 20 à 40 % . . .	5,3	6,0
Est.	3,2	4,3	Moins de 20 % . . .		
Centre	4,8	2,9	Population :		
Centre-Est.	5,9	4,6	Moins de 5.000 h.	4,9	3,6
Midi Méditerranéen.	5,7	3,3	5.000 à 10.000 h.	6,5	3,1
Sud-Ouest.	1,6	7,8	10.000 à 30.000 h.	2,5	5,8
			30.000 à 100.000 h.	7,3	2,5
			100.000 à 150.000 h.	6,5	5,1
			Plus de 150.000 h.	2,4	2,0

On observe que ces coefficients de variation sont nettement plus faibles que ceux donnés au chapitre 1. Et le terme négligé paraît être presque toujours plus grand que celui dont on tient compte, comme on le voit au *tableau 23*.

TABLEAU 23

Achats de chaussures. Valeur du rapport $\sum E (\bar{x} - \bar{x})^2 / \sum E (\bar{X} - \bar{x})^2$

	CHAUSSURES			CHAUSSURES	
	d'hommes	de femmes		d'hommes	de femmes
France entière.	3,7	4,3	Par catégorie de communes :		
Par grande région :			Population agricole :		
Région parisienne.	4,2	5,4	Com-munes { Plus de 60 %	6,2	5,1
Nord.	3,5	3,2	{ 40 à 60 %	1,5	9,0
Ouest.	8,5	15,8	{ 20 à 40 %	6,2	1,8
Est.	7,7	2,2	{ Moins de 20 %	6,7	2,0
Centre.	8,5	10,4	Population :		
Centre-Est.	0,3	1,6	Moins de 5.000 h.	3,7	3,6
Midi Méditerranéen.	1,2	3,0	Com-munes { 5 000 à 10 000 h.	2,3	10,2
Sud-Ouest.	5,7	3,5	{ 10 000 à 30.000 h.	18,6	1,9
			{ 30 000 à 100.000 h.	1,7	19,6
			{ 100 000 à 150.000 h.	1,7	1,6
			{ Plus de 150.000 h.	1,9	3,7

Bien entendu, il faut se rappeler que l'on compare des variances établies par des méthodes différentes, dont une [indirecte avec formule (a)] est réputée peu précise et dont l'autre (directe) ne l'est guère plus, en raison des difficultés rencontrées pour évaluer sur échantillon les paramètres entrant dans les formules. On peut néanmoins admettre que le terme négligé est, en gros, quatre fois plus grand que le terme obtenu par les méthodes « indirectes », et retenir que le coefficient de variation ainsi calculé doit (par conséquent) être plus que doublé.

Il serait intéressant de connaître une méthode d'évaluation indirecte de $\sum E (\bar{x} - \bar{x})^2$ car il ne fait aucun doute que les calculs indiqués par F. Yates sont beaucoup plus simples et rapides que ceux nécessités par l'application directe des formules.

CONCLUSION

Pour terminer, nous formulons le vœu que des recherches analogues puissent être effectuées à propos des nouvelles enquêtes de l'I. N. S. E. E. ; nous souhaitons qu'on se rende compte de l'absolue nécessité de ce genre de spéculation (pour une administration qui entend employer des méthodes scientifiques) et que, par conséquent, on sache y consacrer les moyens nécessaires.

P. THIONET.

BIBLIOGRAPHIE

Publications de l'Institut National de la Statistique et des Études Économiques.

A. — SUR LA QUESTION DES ENQUÊTES PAR SONDAGE :

- La méthode des sondages. (Impr. I.N.S.E.E.).
- Le manuel des enquêteurs par sondage. —
- Études théoriques } n° 5. La théorie des sondages,
- (Imprimerie Nle) } n° 6. Application des méthodes des sondages aux enquêtes statistiques.

B. — SUR LES « ENQUÊTES EMPLOI » DE L'I. N. S. E. E. :

1. Plans d'échantillonnage des « Enquêtes Emploi » 1950 à 1953 (brochure ronéo).

2. Méthodes et résultats : voir les publications suivantes :

Bulletin hebdomadaire de statistique du 30 janvier 1951 (supplément); du 7 mars 1953 (variété statistique); du 19 décembre 1953 (variété statistique).

Bulletin mensuel de statistique : octobre 1951 (page 46); octobre 1952 (page 46).

Supplément trimestriel du Bulletin mensuel de statistique : janvier-mars 1951 (page 1).

avril-juin 1951 (page 22); octobre-décembre 1952 (page 1); avril-juin 1953 (page 1).

Voir aussi le *Journal de la Société de Statistique de Paris*, d'avril-juin 1951 (page 108) :

R. Lévy-Bruhl. Enquêtes récentes par sondage sur l'emploi.

C. — SUR LES ÉTUDES DU MARCHÉ DE LA CHAUSSURE ET DU MARCHÉ DES ASSURANCES (1953) :

Plan d'échantillonnage de l'enquête chaussure (brochure ronéo).

Plan d'échantillonnage de l'enquête assurances (brochure ronéo).

Jean Desabie : Le marché français de la chaussure vu par le consommateur (publication du Centre d'Études et de Mesure de Productivité).

Les résultats de ces deux sondages ont été communiqués aux organismes ayant financé les enquêtes, dont ils sont dans une certaine mesure la propriété.

D. — SUR LES SONDAGES RADIO DE L'I. N. S. E. E. :

— Plan d'échantillonnage de l'enquête radio 1952 (brochure ronéo).

— — — — — 1954 — — — — —

— Supplément trimestriel du *Bulletin mensuel de Statistique*, janvier-mars 1954 (pages 31-58); juillet-septembre 1954 (pages 22-24). Une enquête par sondage sur l'auditoire radio-phonique).
