

JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

RAYMOND LÉVY-BRUHL

Enquêtes récentes par sondage sur l'emploi

Journal de la société statistique de Paris, tome 92 (1951), p. 108-125

http://www.numdam.org/item?id=JSFS_1951__92__108_0

© Société de statistique de Paris, 1951, 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/>

VII

ENQUÊTES RÉCENTES PAR SONDAGE SUR L'EMPLOI

MONSIEUR LE PRÉSIDENT, MES CHERS COLLÈGUES,

Les problèmes relatifs à l'emploi de la main-d'œuvre, à la durée du travail, au chômage, aux effectifs qui travaillent dans les différentes branches d'activité sont certainement parmi les plus importants que pose aux pouvoirs publics, aux économistes, aux syndicats ou aux organisations professionnelles l'évolution économique d'un pays. La seule évocation de la misère immédiate qu'entraîne pour un foyer le chômage ou la diminution d'activité du chef de la famille suffit pour comprendre l'intérêt humain et considérable que présente la connaissance de ces problèmes. D'autre part, et en considérant maintenant le niveau général, dans un pays donné, de l'emploi et du chômage, il n'est besoin que de rappeler l'accent mis sur la nécessité de réalisation du plein emploi par les théories économiques modernes pour se rendre compte de l'obligation devant laquelle se trouvent ceux qui ont la responsabilité économique d'un pays d'être renseignés avec précision et rapidité sur les variations du niveau de l'emploi.

Or, quels sont en France les moyens d'information dont on dispose pour la connaissance de ces problèmes fondamentaux?

Trois méthodes d'investigation paraissent utilisables.

1^o On peut tout d'abord s'adresser aux employeurs et leur demander de fournir périodiquement des informations sur leurs effectifs, la durée du travail dans leurs établissements, la répartition des personnes employées suivant l'âge, le sexe, la qualification, etc... C'est cette approche du problème qu'utilise le Ministère du Travail dans son enquête trimestrielle qui touche environ 25.000 entreprises qui occupent plus de 10 salariés remplissant un questionnaire qui leur est transmis par l'Inspecteur du Travail de leur circonscription. Cette enquête, dont l'intérêt extrême en raison des informations qu'elle fournit sur la durée du travail, les effectifs employés, le niveau des salaires dans les diffé-

entes branches d'activité, est encore accru du fait de sa périodicité (trimestre), et de la rapidité de la collecte des questionnaires et de leur dépouillement mécanographique (6 semaines entre la date de l'enquête et l'élaboration des résultats), ne permet cependant pas d'obtenir un tableau complet de l'activité et de l'emploi dans la nation. L'enquête ne porte en effet que sur les établissements occupant plus de 10 salariés. Une extension de l'enquête aux petits établissements (occupant de 1 à 10 salariés) est actuellement à l'étude.

De plus une certaine proportion d'ailleurs variable suivant la taille et le domaine d'activité des entreprises interrogées ne retourne pas le questionnaire à l'Inspection Départementale du Travail. On ne peut donc être certain de la parfaite représentativité de l'échantillon soumis à l'enquête. L'enquête enfin ne couvre pas l'ensemble des activités. Sont exclus de son champ les services publics, l'agriculture ainsi que le gaz et l'électricité. Elle ne touche ni les travailleurs isolés ni le personnel domestique. Ses résultats ne permettent donc pas d'estimer les mouvements de main-d'œuvre qui se produisent entre le secteur agricole et le secteur industriel et commercial ni, à l'intérieur de ce dernier secteur, entre petites et grandes entreprises, artisans et salariés. Elle n'est donc pas susceptible de fournir un bilan d'ensemble de l'état de la main-d'œuvre ni des indications globales sur le mouvement de la main-d'œuvre.

Les enquêtes périodiques effectuées par les organisations professionnelles patronales auprès de leurs ressortissants, à la demande du Ministère de l'Industrie et du Commerce, ont une valeur très variable suivant les secteurs. Elles ne sont en tous cas, en aucune manière, susceptibles de fournir des indications valables sur l'ensemble de la population active française et sur les variations de celle-ci. Elles ne touchent que quelques branches d'activité et d'une manière fragmentaire.

Il est enfin inutile de préciser que toute enquête auprès des employeurs est incapable de fournir une indication quelconque sur la composition de la population non active.

2^o Une seconde méthode d'investigation consiste à profiter d'une opération administrative d'enregistrement nécessairement effectuée par chaque individu actif ou chômeur d'une manière périodique pour établir un bilan de la main-d'œuvre employée, du chômage, etc...

C'est la méthode qui est utilisée en Grande-Bretagne où le renouvellement des cartes de l'Assurance Nationale permet un recensement annuel de l'emploi et où le relevé des inscriptions des chômeurs dans les bureaux locaux de la main-d'œuvre permet l'obtention d'une statistique mensuelle du chômage. En France, il n'est malheureusement pas possible d'envisager, dans l'immédiat, l'utilisation des dossiers ou des cartes de Sécurité Sociale pour effectuer un recensement de la population active. En revanche les statistiques actuelles du chômage sont établies d'après le nombre d'individus inscrits pour toucher des secours à des fonds de chômage et le nombre des demandes d'emploi enregistrées dans les bureaux officiels de placement qui n'ont pu être satisfaites au dernier jour du mois. Ces données sont cependant très insuffisantes pour permettre de connaître exactement l'état et le mouvement du chômage en France. L'inscription à un fonds de chômage est en effet assortie de nombreuses conditions restrictives qui font que le nombre des chômeurs inscrits

dans ces fonds est obligatoirement inférieur au nombre réel de chômeurs. Parmi ces conditions, indiquons qu'il faut qu'il existe un fonds de chômage dans la commune et il n'en existe actuellement que 450 environ dans toute la France, qu'il faut avoir travaillé pendant au moins six mois ce qui exclut tous les jeunes qui recherchent du travail pour la première fois, qu'il faut résider dans la commune depuis au moins trois mois et avoir des ressources totales inférieures à un maximum. De même les demandes d'emploi non satisfaites au dernier jour du mois dans les offices de placement ne fournissent pas plus une valeur exacte du nombre des chômeurs à cette date. Nombreux sont, en effet, ceux qui recherchent du travail par d'autres moyens qu'en s'inscrivant comme demandeurs d'emploi dans les offices publics de placement.

3^e Il est enfin une troisième méthode d'investigation qui consiste à s'adresser directement aux individus formant la population pour connaître leur situation personnelle relative aux problèmes de l'activité, la durée du travail, la branche professionnelle à laquelle ils appartiennent, etc... C'est la méthode qui est utilisée lors des recensements périodiques de la population. Elle permet, à condition d'utiliser un questionnaire compréhensible de la part de tous les individus de la population, de disposer de bons agents recenseurs et de pouvoir compter sur une certaine bonne volonté de la part des personnes recensées, l'établissement de statistiques détaillées et précises relatant l'état, au moment du recensement, de la population étudiée. De tels recensements sont cependant lourds, onéreux, longs à dépouiller. On ne peut envisager de les effectuer avec une courte périodicité (semestre ou année par exemple). C'est placé devant ces difficultés de coût et de lenteur du recensement, qu'on a été amené à envisager l'utilisation des méthodes modernes de sondage qui ont précisément pour principal avantage d'être d'un prix de revient faible et de fournir des résultats rapides. Le premier pays qui ait utilisé la méthode des sondages pour effectuer une enquête périodique sur les problèmes de l'emploi, c'est les États-Unis. Pendant les années de crise de 1930-1933 on n'a disposé d'aucune mesure du chômage. On pense qu'il y a eu à certains moments plus de 10 millions de chômeurs et probablement moins de 20 millions et cette ignorance ne facilita pas la tâche des Pouvoirs Publics. De plus dans ce pays où la main-d'œuvre est beaucoup plus mobile qu'en France, il est indispensable entre les recensements décennaux, d'être renseigné sur la répartition géographique et professionnelle de la population active. Le bureau du Census a mis au point une méthode de sondage portant sur 25.000 ménages comptant 75.000 personnes âgées de 14 ans et plus qui utilise les techniques les plus modernes du sondage au hasard. Les 3.067 comtés des États-Unis sont groupés en 2.000 unités primaires. Les 2.000 unités primaires sont réparties en 68 strates suivant divers critères (géographie, industrie, urbain ou rural). Dans chaque strate une unité est choisie au hasard. Cette unité est à son tour divisée en petites zones (quartiers de villes, zones rurales) dont une certaine proportion est choisie au hasard. Dans ces zones les recenseurs font la liste de tous les ménages y résidant et une certaine proportion de ces ménages est choisie au hasard pour faire partie de l'échantillon. L'enquête a lieu tous les mois et 1/5 de l'échantillon est renouvelé chaque mois. Dans chacune des 68 unités échantillons, se trouve une équipe d'enquêteurs dirigée par un fonctionnaire de l'administration fédérale qui

assure leur formation et leur contrôle. Cette enquête rend les plus grands services, fournissant en particulier chaque mois la répartition de la population active, non active et en chômage par sexe, âge et activité professionnelle (pour ceux qui en possèdent).

Des principes analogues sont appliqués pour des enquêtes trimestrielles au Canada et au Japon. Une enquête mensuelle a lieu à Porto-Rico.

L'intérêt de posséder en France de telles enquêtes étant apparu certain, l'I. N. S. E. E. a tenté d'élaborer un plan d'enquête qui tout en ne sacrifiant rien à la précision nécessaire, ait le prix de revient minimum. Pour pouvoir mesurer la précision des résultats on a effectué uniquement des tirages au hasard qui seuls permettent d'appliquer les lois du calcul des probabilités; pour réduire au maximum le coût on a tenté de grouper les unités interrogées afin de réduire le plus possible les déplacements des enquêteurs. Notre collègue Thionet, profitant de l'expérience acquise lors de l'exécution d'enquêtes par sondage au cours de ces dernières années et de sa connaissance tant des expériences étrangères que de la théorie sur laquelle elles reposent, a élaboré un plan d'enquête qui utilise au maximum les possibilités offertes par l'organisation et la documentation administratives françaises.

L'unité interrogée est le ménage occupant un logement désigné. Les principaux avantages présentés par cette unité sont les suivants : possibilité d'obtention de données relatives aux ménages telles que nombre de personnes travaillant par ménage, etc..., économies de déplacement pour l'enquêteur qui peut interroger plusieurs personnes à la fois. On peut en revanche craindre que le fait de tirer non un individu mais dans le même tirage un groupe d'individus vivant ensemble, réduise la précision de l'enquête en raison de la corrélation qui pourra exister entre les réponses de certains de ces individus. Il est évident que pour certaines questions cette corrélation existe et que l'information tirée d'une enquête auprès de N individus tirés par grappes de 2, 3 ou 4 personnes est inférieure à l'information que fournissent N individus tirés un à un au hasard de l'urne totale comprenant l'ensemble de la population. Cependant, pour les problèmes de main-d'œuvre étudiés dans une enquête sur l'emploi, la corrélation entre les réponses des individus d'un même ménage est faible et la légère diminution d'information ainsi introduite est largement compensée par la réduction de coût ainsi obtenue qui permet, pour un même prix de revient total, d'interroger un nombre N bien plus grand d'individus. Pour certaines questions même, la grappe d'individus vivant ensemble apparaît dans cette enquête plus efficace que l'individu isolé.

On ne peut évidemment envisager un tirage direct dans une urne où figuraient tous les logements de France de quelques milliers de logements où s'effectuerait l'enquête. Il est évident qu'une telle méthode amènerait une dispersion géographique de l'échantillon telle qu'elle rendrait toute enquête impossible. On est donc amené à grouper les ménages échantillons dans l'espace. L'unité la plus logique, étant donné l'organisation administrative française, paraît être soit le canton soit la commune soit le hameau ou la section de commune. Mais pour cette dernière subdivision il n'existe aucune statistique et on a été obligé d'avoir recours à l'unité commune malgré l'extrême hétérogénéité, au point de vue de la taille, que présente cette unité. Des études pré-

liminaires sont en cours en vue d'examiner si l'unité canton ne pourrait pas être substituée à l'unité commune d'une manière fructueuse.

Actuellement le plan d'enquête se présente donc sous forme d'un sondage à deux degrés. Au premier degré on tire un certain nombre de communes. Au second degré on tire un certain nombre de logements dans ces communes tirées. Le premier tirage est un tirage au hasard dans un univers stratifié avec probabilité proportionnelle à la taille de chaque unité, le second tirage est un tirage systématique.

Il apparaît en effet, et la théorie des sondages le confirme, que plutôt que de tirer, mettons 200 communes dans une grande urne où se trouveraient mélangées les 38.000 communes de France, Paris voisinant avec une commune rurale du Midi de 15 habitants, il était préférable de répartir toutes les communes en un certain nombre d'urnes, chaque urne ayant des caractères communs soit par la proximité géographique soit par la taille ou en raison de critères économiques. A l'intérieur de chacune de ces urnes on tire une ou plusieurs communes qui représentent l'ensemble des communes de cette urne. De plus, étant donné les tailles très différentes de communes françaises en nombre d'habitants, on donne dans ce tirage qui est effectué au hasard (au moyen de tables de nombres au hasard, dites tables de Tippett) à chaque commune une chance d'être tirée proportionnelle à son nombre d'habitants. Ainsi une commune de 600 ménages se trouvant dans la même urne qu'une commune de 100 ménages aura 6 fois plus de chance d'être tirée que cette dernière, 10 logements étant par exemple ensuite tirés dans la commune choisie. Cette méthode donnera à chaque logement de France la même probabilité de figurer dans l'échantillon et par conséquent à chaque individu puisqu'un individu habite dans un logement et un seul.

Il paraît intéressant de décrire brièvement comment ont été constituées les urnes de communes. On a commencé par répartir les communes des 90 départements dans 8 grandes urnes, chacune de ces grandes urnes contenant un certain nombre de départements. On a ainsi délimité 8 grandes régions : Région Parisienne, Nord, Ouest, Nord-Est, Centre, Centre-Est, Sud-Est et Sud-Ouest. Il n'était pas possible d'utiliser les circonscriptions des Directions Régionales de l'I. N. S. E. ni un regroupement de celles-ci en raison de certaines anomalies dans leur délimitation (la Somme par exemple se trouve certainement mieux placée dans l'urne où sont les départements du Nord que dans celle des départements normands, or elle dépend de la Direction de Rouen). De même on comprendra tout de suite que pour un sondage qui ne porte que sur quelques milliers de ménages, le département forme une urne trop petite. En revanche les 8 grandes régions délimitées sont assez grandes pour qu'on puisse envisager de dépouiller certains résultats valables de l'enquête séparément pour chacune d'elles.

A l'intérieur de ces grandes urnes on a réparti les communes, quels que soient les départements auxquels elles appartiennent entre 10 petites urnes selon les critères suivants : dans les 4 premières urnes se sont trouvées toutes les communes rurales c'est-à-dire toutes les communes dont la population agglomérée au chef-lieu était lors du recensement de 1946 inférieure à 2.000 habitants. La proportion de la population vivant de l'agriculture dans la commune

déterminait ensuite dans laquelle de ces 4 urnes, appelées en termes de sondage des strates, elle devait être placée. C'est ainsi qu'on a, pour chaque grande région :

dans la 1^{re} urne, les communes rurales dont plus de 60 % de la population vit de l'agriculture;

dans la 2^e urne, les communes rurales dont la proportion de population vivant de l'agriculture est comprise entre 40 et 60 %;

dans la 3^e urne, les communes rurales dont la proportion de population vivant de l'agriculture est comprise entre 20 et 40 %;

dans la 4^e urne, les communes rurales dont la proportion de population vivant de l'agriculture est inférieure à 20 %.

Les 6 autres urnes contiennent les communes urbaines classées suivant leurs populations :

dans la 5^e urne, se trouvent les communes urbaines ayant moins de 5.000 habitants; dans la 6^e urne, de 5.000 à 10.000 habitants; dans la 7^e urne, de 10.000 à 30.000 habitants; dans la 8^e urne, de 30.000 à 100.000 habitants; dans la 9^e urne, de 100.000 à 150.000 habitants; dans la 10^e urne, plus de 150.000 habitants.

De plus, certaines communes de banlieue, véritable prolongement de la ville dont elles sont des satellites, ont été placées dans la même urne que cette ville. Enfin étant donné le grand nombre de villes qui se trouvent dans certaines urnes, on a créé de plus petites urnes. Ainsi dans la 10^e urne de la région parisienne, les villes ont été réparties entre 10 petites urnes, appelées sous-strates avant le tirage. Des critères économiques (communes résidentielles bourgeoises, banlieues industrielles) ou politiques (votes lors des dernières élections) ont été utilisés en vue de ce classement. Des études sur les agglomérations urbaines que l'Institut entreprend actuellement permettront dans un proche avenir d'améliorer certainement cette stratification. D'autre part, certaines des urnes sont vides. Ainsi il n'existe aucune ville de 100 à 150.000 habitants (9^e urne) dans la région du Sud-Ouest.

En général on a tiré, avec une probabilité proportionnelle au nombre des ménages de chaque commune, comme il a déjà été indiqué, deux communes dans chaque urne ce qui permet de calculer une estimation de la variance entre communes des caractères à mesurer. Dans certains cas cependant, on a pu être amené, pour éviter qu'un enquêteur ait, dans une commune à visiter, trop ou trop peu de personnes à interroger, à tirer dans une strate donnée soit une seule commune, soit trois communes. Au total de l'ensemble des strates couvrant la France entière on a tiré 184 communes où s'effectue l'enquête, ce tirage étant fait au hasard sauf dans les strates où toutes les communes de la strate sont soumis à l'enquête et où par conséquent on n'effectue pas de tirage (Marseille et Nice par exemple ou Toulon).

Dans chaque commune tirée, il y a lieu ensuite de choisir les ménages à enquêter. Le plan adopté consiste dans un tirage systématique d'un certain nombre de feuilles de ménage remplies par les chefs de ménage lors du recensement de 1946 dans le tas des feuilles de ménages de la commune, tas qui est entreposé dans les locaux de la Direction Régionale de l'Institut dans le ressort duquel se trouve la commune. Les feuilles de ménage d'une même

commune étant rangées généralement dans un ordre quelconque, un tirage à intervalles équidistants dans l'ensemble de ces feuilles équivaut à un tirage au hasard. Lorsque les feuilles de ménage sont rangées dans un ordre géographique par hameaux, rues, etc..., ce tirage systématique a en outre l'avantage de se faire dans un ensemble stratifié géographiquement.

Comme les phénomènes à mesurer présentent une plus grande homogénéité dans les campagnes un taux d'échantillonnage variable a été adopté suivant les régions. Dans les strates 4 à 10 (communes urbaines et communes rurales ayant moins de 20 % de la population vivant de l'agriculture qui peuvent être assimilées aux précédentes) le taux d'échantillonnage a été de 1/1500 ménages. Dans la strate 3, il a été de 1/3000 ménages et dans les strates 1 et 2, il a été, suivant les régions de 1/3000 et 1/6000. Dans chacune des urnes précédemment délimitées on a donc calculé le nombre de ménages à interroger en appliquant ces taux au nombre total des ménages de toutes les communes de l'urne. Ce nombre de ménages à interroger a été divisé, en parts égales, entre les 1, 2 ou 3 communes tirées de l'urne et a déterminé l'intervalle de tirage des feuilles de ménage dans ces différentes communes.

Ce plan d'échantillonnage, qui a été dicté à la fois par des considérations de coût et par le désir d'utiliser au maximum la documentation déjà existante, présente certains inconvénients. Notre but est d'améliorer notre documentation de disposer de crédits plus importants afin de pouvoir rendre plus précis encore l'instrument que nous utilisons.

Le plan d'enquête utilisé présente les caractéristiques suivantes :

1^o Le plan d'enquête basé sur un tirage de feuilles de ménages exclut du champ de l'enquête toute la partie de la population qui ne vit pas dans les logements normaux pour lesquels une feuille de ménage a été remplie en 1946. Cette population, appelée population comptée à part, comprend les communautés, casernes, internats, prisons, etc... Son inclusion dans l'enquête exigerait l'élaboration d'un plan d'enquête spécial;

2^o L'échantillon tiré étant un échantillon de logements pour lesquels en 1946 des feuilles de ménage ont été remplies, les logements construits depuis cette date ne sont pas représentés dans cet échantillon. Il y a là une source d'erreurs qui iront en s'accentuant à mesure que le dernier recensement s'éloignera dans le temps. On s'est donc préoccupé de réunir une documentation permettant le tirage d'un échantillon de logements neufs qui compléterait l'échantillon précédemment tiré.

3^o Le ménage interrogé est celui qui occupe le logement tiré même si ce n'est plus le même ménage que celui qui habitait ce même logement en 1946. En fait 15 % des logements ont changé d'occupant entre mars 1946 et avril 1950 et 17 % entre mars 1946 et octobre 1950.

4^o Il est indispensable que les enquêteurs obtiennent des réponses pour tous les ménages occupant les logements tirés au hasard sans pouvoir effectuer de substitution. Cette condition oblige parfois l'enquêteur à retourner 2, 3, 4 fois ou plus visiter le même logement dont la porte est toujours close. Il n'est pas possible cependant de négliger ces ménages difficiles à joindre. Toutefois afin de réduire au maximum ces visites multiples, il a été jugé indispensable d'établir un questionnaire tel qu'un membre du ménage puisse répondre aux diffé-

rentes questions pour tous les membres du ménage. On a pu ainsi obtenir 70 % des questionnaires remplis après une visite et 90 % après une ou deux visites.

5^o Quelle que soit la conscience des enquêteurs et leur habileté, il n'en reste pas moins que certains enquêtés refusent de répondre, que des locaux qui étaient occupés en 1946 sont maintenant vides, que des logements habités sont cependant vides pendant la durée de l'enquête, leurs occupants étant par exemple en voyage. Toutes les précautions sont prises pour réduire au maximum le taux des refus (lettre aux enquêtés, publicité, intervention de contrôleurs, etc...). Ce taux a pu être maintenu aux environs de 2 % (à Paris et à Lyon 7 %). Au total les questionnaires collectés étaient de 13 % inférieurs au nombre des logements tirés. On peut envisager deux méthodes pour corriger l'erreur ainsi introduite. La première méthode qui était utilisée jusqu'à l'an dernier dans les enquêtes par sondage de l'I. N. S. E. E. consiste à effectuer dans chaque commune un tirage supplémentaire dans les mêmes conditions d'un certain nombre de feuilles de ménage. Si un ménage de l'échantillon principal est défaillant, on le remplace par un ménage de la liste complémentaire. Outre qu'elle a tendance à favoriser la paresse de l'enquêteur qui considérera facilement des ménages successifs comme défaillants, cette méthode présente le grave inconvénient de remplacer systématiquement les ménages difficiles à joindre (célibataire, personnes ayant une durée hebdomadaire de travail exceptionnellement longue, etc...) par les ménages faciles à joindre de la liste complémentaire (familles ayant des enfants en bas âge, oisifs, etc...). La seconde méthode qui a finalement paru préférable, et qui a été utilisée pour les 2 enquêtes sur l'emploi de 1950 ne comporte aucun remplacement des ménages défaillants. Les X questionnaires renseignés pour une strate ou un groupe de strates donné d'une grande région donnée et une taille de ménages donnée sont considérés comme représentatifs des n questionnaires qui auraient dû être remplis par tous les ménages à enquêter de cette catégorie. On se borne donc à créer n — X questionnaires en recopiant n — X des questionnaires tirés au hasard parmi les X questionnaires remplis. Ainsi si au lieu des 29 questionnaires on en a recueilli 26, 3 questionnaires parmi ces 26 sont tirés au hasard et doublés. On opère de cette façon pour chaque taille de ménage de chaque groupe de strates de chaque région. Pour les questionnaires qui doivent être doublés on crée deux cartes perforées à partir de chacun d'eux.

Au total environ 5.700 ménages comprenant 13.500 personnes âgées de 14 ans et plus ont été effectivement interrogées au cours de chacune des deux enquêtes.

6^o Enfin un dernier problème que pose l'élaboration du plan d'échantillonnage est extrêmement important. Convient-il à chaque enquête de changer les communes où s'effectue l'enquête (unité du 1^{er} degré) et si les mêmes communes sont choisies, convient-il de changer les ménages interrogés (unités du 2^e degré) ou est-il préférable de conserver intégralement le même échantillon d'une enquête à l'autre ou enfin, suivant l'exemple de l'enquête américaine, la meilleure solution ne consiste-t-elle pas en un renouvellement partiel de l'échantillon à chaque enquête? Il apparaît évident, et la théorie le confirme, qu'étant donné la corrélation que présentent les mesures du même phénomène

sur le même échantillon à 2 époques différentes, la précision maximum qu'on peut atteindre dans la mesure des variations de ce phénomène dans le temps correspondra à un échantillon qui n'est pas du tout renouvelé. Cependant l'enquête se déroulant périodiquement il n'est pas possible de conserver indéfiniment les mêmes enquêtés qui, soit se lasseraient, soit deviendraient des enquêtés professionnels. En 1950, l'échantillon a été, pour ces raisons psychologiques et parce qu'il s'agissait d'enquêtes de mise au point à la fois du plan de sondage, du questionnaire et des instructions totalement renouvelé entre les enquêtes. On envisage au contraire pour l'avenir un renouvellement partiel de l'échantillon.

Tout aussi importants pour la précision de l'enquête que le plan d'échantillonnage sont le questionnaire et les instructions données aux enquêteurs.

Un grand soin a donc été apporté à la confection du questionnaire. De multiples essais de modèles successifs de questionnaires ont été effectués dans diverses Directions Régionales de l'Institut, à la fois dans le milieu urbain et dans le milieu rural puisque le questionnaire devait pouvoir s'appliquer à ces deux milieux. En vue d'obtenir des réponses sincères et précises, on s'est attaché :

1^o à ne poser qu'un nombre réduit de questions à chaque individu, ces questions étant très simples et pouvant être comprises par tous;

2^o à éliminer toute question qui puisse être jugée indiscrète par la personne interrogée (salaires par exemple);

3^o à se borner à des questions dont la réponse puisse être fournie par n'importe quel membre du ménage pour tous les autres membres ce qui simplifie le travail de l'enquêteur.

On a pu juger de l'importance de l'énoncé des questions en notant les variations considérables qu'une légère modification de l'énoncé pouvait entraîner. Par exemple on demandait en avril aux personnes qui ne travaillaient plus si c'était parce qu'elles étaient en retraite (trop âgées pour travailler) et 10 % de ces personnes se sont rangées dans cette catégorie. En revanche, en octobre, on leur a demandé si leur âge avait été la cause principale de leur cessation de travail et cette fois on a trouvé 22 % des personnes inactives qui se sont trouvées dans ce cas.

L'importance de définitions précises dans les instructions écrites données aux enquêteurs est pleinement apparue au cours de ces enquêtes. Il a donc fallu définir, parfois avec quelque arbitraire, quand une femme était active ou non, ce qui caractérisait la recherche de travail, les différentes situations dans la profession qu'on désirait distinguer, etc...

Outre ces instructions écrites, les enquêteurs, qui sont des agents des Directions Régionales de l'Institut, ayant pour la plupart effectués depuis plusieurs années des enquêtes par sondage (quelques-uns même depuis 1942) et par conséquent spécialisés, reçoivent une formation orale nécessaire avant chaque enquête. Tout enquêteur est par ailleurs appelé, à la fin de l'enquête, à fournir un rapport où sont relatées toutes les difficultés qu'il a rencontrées au cours de cette enquête. Ceci s'est révélé fort instructif, pour une enquête périodique, en permettant de se rendre compte des questions délicates, mal comprises, etc... à modifier lors de l'enquête suivante et de contrôler par l'examen des

difficultés rencontrées que l'instruction de l'enquêteur avait été correctement assurée.

Chaque ménage enquêté a reçu une lettre l'avisant de la prochaine visite d'un enquêteur (cette mesure n'a touché que les ménages occupant le même logement qu'en 1946). De plus en avril une publicité par presse et radio, qui se révèle peu efficace, avait été entreprise.

149 enquêteurs plus 11 étudiants, ayant une longue expérience des enquêtes par sondage, ont formé le corps d'enquêteurs. Des contrôles de leur travail ont été effectués par lettre, par téléphone ou par une contre-visite. Il y a lieu de souligner que dans la presque-totalité des cas l'enquêteur a été aimablement reçu, les questions ont paru intéresser tous les milieux enquêtés et les réponses ont été données de bonne grâce.

L'enquête a eu lieu du 24 avril au 6 mai et du 16 octobre au 28 octobre.

Les questions peuvent être divisées en deux groupes correspondant grossièrement aux 2 buts de l'enquête :

1^o le sondage met à jour pour certains points très importants à intervalles rapprochés des données que fournit un recensement : on peut ranger dans cette catégorie les questions relatives au sexe, à l'âge, au nombre d'enfants à charge, à l'activité ;

2^o le sondage permet de connaître ou d'approfondir des problèmes qu'il n'est pas possible de poser à toute la population lors d'un recensement : certaines questions sont posées chaque fois : durée hebdomadaire du travail, durée et raisons de l'inactivité ; d'autres une fois : mode de rémunération, distance du lieu de travail au domicile, stabilité de l'emploi, horaire de travail, journées de repos hebdomadaire.

Il n'y a pas de questions relatives au ménage sauf quant au nombre d'enfants. Étant donné que le plan d'échantillonnage adopté est tel qu'il fournit un échantillon représentatif de ménages, aussi bien que d'individus, on pense introduire quelques questions très simples s'appliquant au logement du ménage lors de prochaines enquêtes. Quant aux résultats qui ont été publiés dans le bulletin mensuel de statistique (1), il faut, pour comprendre leur valeur, se souvenir toujours qu'ils sont tirés d'un sondage et qu'on ne peut par conséquent avoir des résultats que pour des groupes importants de population. On aura à partir de ce sondage la répartition de la population par sexe et par groupes décennaux d'âge, on n'aura pas le nombre de garçons de 19 ans à Montélimar.

Il ne saurait être question dans le cadre de cette communication d'analyser tous les résultats de cette enquête mais simplement de souligner les catégories d'information que permet de recueillir cette forme de collecte de renseignements et les difficultés rencontrées.

On peut tout d'abord obtenir une répartition de la population âgée de 14 ans et plus, à l'exception de la population comptée à part, suivant le sexe, l'âge, le nombre d'enfants à charge, etc...

Ce qui est particulièrement intéressant est la répartition de la population en 3 catégories :

1^o population active ayant un emploi ;

(1) *Bulletin mensuel de Statistique*. Supplément, Janvier-Mars 1951.

2^o population active sans emploi et en recherchant;

3^o population non active.

Si la première et la troisième catégories paraissent assez faciles à délimiter, il n'en est pas de même de la seconde.

Les résultats bruts de l'enquête font en effet apparaître, comme on pouvait s'y attendre, qu'un grand nombre de personnes, quand on les interroge, ont tendance à déclarer qu'en effet elles n'ont pas d'emploi mais qu'elles en recherchent soit parce qu'elles sont réellement des chômeurs, soit parce que, quoique ne recherchant pas activement un emploi, si on leur en offrait un particulièrement intéressant, elles l'accepteraient, soit pour ne pas passer pour un oisif vis-à-vis de l'enquêteur. Il est donc nécessaire, si on veut mesurer l'importance des chômeurs de poser une définition arbitraire de ce qu'on entendra sous ce terme. On a été ainsi amené à distinguer : 1^o les individus ayant déjà travaillé, aptes physiquement à un travail (âge, durée du chômage) pour lequel ils sont qualifiés et en recherchant activement; 2^o les individus inaptes à un travail et en recherchant d'une autre part, et enfin 3^o les individus ne recherchant pas activement de travail ou n'ayant jamais travaillé qui, d'après la définition adoptée, doivent être réintégrés dans la population non active à laquelle ils appartiennent en fait. Il a été nécessaire pour cela d'introduire une question sur l'aptitude au travail et sur le mode de recherche du travail. On peut ainsi établir un bilan de la population âgée de 14 ans et plus (à l'exception de la population comptée à part) aux dates des 2 enquêtes.

Répartition de la population âgée de 14 ans et plus en % de cette population.

	POPULATION ACTIVE										POPULATION non active	TOTAL		
	Occupée		N'ayant pas d'emploi et cherchant du travail											
			Bénéficiant d'allocation de chômage		Apte au travail et ne recevant pas d'allocation		Inapte au travail et ne recevant pas d'allocation							
	Avril	Octobre	Avril	Octobre	Avril	Octobre	Avril	Octobre	Avril	Octobre	Avril	Octobre		
Hommes . .	82,0	82,3	0,3	0,1	0,8	0,5	0,7	0,3	16,2	16,8	100	100		
Femmes . .	39,8	40,3	0,1	0,1	0,6	0,4	0,9	0,5	58,6	58,7	100	100		
Ensemble . .	59,4	59,9	0,2	0,1	0,7	0,5	0,8	0,4	38,9	39,1	100	100		

On aboutit ainsi à des estimations du nombre de chômeurs d'environ 290.000 en avril et 190.000 en octobre. Les nombres de chômeurs secourus trouvés sont, à 2.000 près, les mêmes que ceux du Ministère du Travail. Une telle précision ne peut être attribuée qu'à un bienheureux hasard.

On a établi pour l'ensemble de la population non occupée et recherchant du travail sa structure par sexe et par âge, la répartition suivant le mode de recherche du travail (hommes : 33 %, office public de placement et 40 % relations personnelles. Femmes : 14 %, office public de placement, 55 % relations personnelles) la cause de l'inactivité (congédiement : hommes 54 %, femmes 36 %), la durée de l'inactivité et l'activité précédemment exercée (très faible dans l'agriculture, 56 % des hommes recherchant du travail exerçaient auparavant une activité industrielle).

De même l'étude de la population non active a porté sur sa structure par sexe et âge, la cause de l'inactivité et la durée de cette inactivité.

Mais c'est pour la population ayant un emploi que l'étude a été la plus poussée. Outre la répartition par sexe et âge, on a effectué une répartition suivant la situation dans la profession (où on a rencontré des difficultés de classement en particulier pour les cadres et les agents de maîtrise, de définition en particulier pour les aides familiaux) et suivant la branche d'activité professionnelle.

L'étude de la répartition de la population occupée suivant la durée hebdomadaire du travail est particulièrement intéressante. Les enquêtes auprès des employeurs ne fournissent en effet que la moyenne par établissement de la durée du travail, ce qui élimine les très faibles ou très fortes durées du travail sauf lorsqu'elles sont observées par tous les travailleurs de l'entreprise. Au contraire l'enquête par sondage fournit une répartition individuelle qui est par conséquent beaucoup plus étalée (8,6 % moins de 40 heures et 31,5 % de plus de 48 heures pour l'ensemble des activités non agricoles contre 1,8 % moins de 40 heures et 10,5 % plus de 48 heures dans l'enquête du Ministère du Travail). En revanche il n'y a que 13,2 % de personnes travaillant exactement 40 heures contre 23 % dans l'enquête du Ministère du Travail.

Parmi les personnes occupées 1,5 % ont déclaré en octobre qu'elles cherchaient un autre emploi que celui qu'elles occupent.

On a dans les 2 enquêtes demandé aux enquêtés s'ils exercent une activité secondaire. Il apparaît que seuls ceux qui exercent réellement 2 activités ont répondu à cette question mais que les nombreuses personnes qui se bornent à quelques heures de travail, non déclaré en général au point de vue fiscal, dans une autre activité que la leur préfèrent passer ce fait sous silence. 2,7 % des personnes actives ont déclarer exercer une activité secondaire à laquelle elles consacrent en moyenne 13,7 heures par semaine. Les résultats des 2 enquêtes sont concordants et il ne paraît pas nécessaire de continuer à poser cette question dans les enquêtes ultérieures.

Enfin certaines questions ont varié au cours des 2 enquêtes et ne seront pas posées d'une manière périodique. Elles permettent simplement d'être renseignés sur certains problèmes qu'aucune autre approche ne permet de mesurer.

En avril une question a porté sur la stabilité de l'emploi. Elle a permis de souligner la grande permanence de l'emploi occupé en France — 96,5 % des personnes ont déclaré que leur emploi était permanent, 2,3 % qu'il était occasionnel et 0,9 % qu'il était saisonnier. On note simplement que le pourcentage des emplois occasionnels croît légèrement pour les manœuvres (6,5 %) ou les personnes travaillant dans les services domestiques ou l'hygiène (5,3 %).

Une autre question portait sur le mode de rémunération et fournit une intéressante répartition des salariés suivant qu'ils sont payés à l'heure, au mois, au pourboire, au pourcentage des ventes, à la tâche ou en nature. On note naturellement de très importantes variations suivant l'activité collective exercée et surtout suivant la situation dans la profession (manœuvres 87,4 % sont payés sur une base horaire, cadres 2,8 %). Enfin une question sur la distance du lieu de travail au domicile permet de souligner l'importance de l'activité exercée et de la situation dans la profession sur cette distance (dis-

tance en temps quel que soit le mode de locomotion utilisé). Il semble que ce problème, étant donnée son importance sociale et économique, mérite d'être approfondi et il est souhaitable qu'on puisse, au cours d'une enquête ultérieure, être renseigné en outre non plus sur le coût en heure mais sur le coût en francs de ces déplacements journaliers.

En octobre une première question a porté sur le type d'horaire journalier de travail et a montré l'extrême importance en France de l'horaire en 2 vacances avec un arrêt de 2 heures exactement pour le déjeuner — 45,6 % des salariés appliquent cet horaire, 45,6 % des non-salariés l'appliquent aussi. On n'a pas noté de différences considérables suivant les activités. On a déterminé aussi l'heure moyenne du début du travail, elle varie en octobre de 7 h. 08 pour les agriculteurs à 8 h. 33 pour les membres des professions administratives ou libérales.

Enfin une seconde question était relative aux jours de repos hebdomadaire. Elle montre que 5 % des personnes occupées ne prenaient aucun repos (surtout des commerçants), 10 % ont une demi-journée de repos, 50 % ont un jour complet, 13 % ont un jour et demi et 20 % ont deux jours. On a pu aussi obtenir la répartition de la population suivant les jours où elle se repose, données qui peuvent être intéressantes par exemple pour les organismes de tourisme, sport, spectacles, etc... La prédominance du repos le dimanche est écrasante. Sur 100 personnes occupées, 77 % ont repos le dimanche entièrement (avec ou sans un autre jour de repos) et 12 % ont repos une demi-journée le dimanche (avec ou sans un autre jour de repos).

Il me reste encore une dernière question à évoquer devant vous et je ne crois pas que ce soit la moins importante. C'est celle des erreurs dont sont affectés les résultats de ces enquêtes. Le plan d'échantillonnage qui a été exposé rend possible le calcul des erreurs d'échantillonnage dues au fait que seul un échantillon a été interrogé. Le tirage à deux degrés qui a été effectué rend le calcul plus long mais nullement impossible. L'aide de l'appareil mécanographique qui effectue une partie des calculs est dans ce domaine particulièrement précieux. Si ce calcul n'est pas rigoureux puisque les variances doivent être estimées d'après l'échantillon et non pas d'après la distribution du caractère dans la population, il donne cependant une estimation très suffisante de l'erreur. On est amené pour chaque donnée pour laquelle on désire connaître l'erreur d'échantillonnage à calculer la variance du caractère à l'intérieur des communes et celle qui existe entre les communes pour chaque strate de chaque région. C'est dire qu'on ne peut effectuer un calcul d'erreur pour toutes les données de l'enquête.

Je sais aussi qu'il peut paraître vain de vouloir chiffrer cette erreur d'échantillonnage alors que de nombreuses autres erreurs d'origine psychologique ou autre interviennent et sont parfois supérieures et non mesurables. Je citerai l'erreur due à une réaction de l'enquêté vis-à-vis de l'enquêteur, l'erreur due à une mauvaise interprétation des questions ou des réponses, l'erreur due à l'imprécision des définitions ou à des fautes de l'enquêteur dans l'interprétation des instructions, etc... sans parler des erreurs de chiffrement ou de calcul. Il ne paraît pas que, parce que de telles causes d'erreur existent, il soit vain de calculer l'erreur d'échantillonnage ou de tenter de la réduire. Ce qui est

indispensable c'est qu'en même temps toutes les précautions soient prises pour réduire au maximum toutes les causes d'erreur non mesurables et j'ai tenté de montrer quelques-unes des mesures prises dans ce sens.

Sur le nombre de personnes déclarant être sans emploi et rechercher du travail par exemple, qui est mettons de 300.000 d'après l'enquête, on trouve un coefficient de variation d'environ 10,6 %, c'est-à-dire qu'on a 95 chances sur 100 de ne pas se tromper en disant que le nombre de personnes recherchant du travail est compris entre $300.000 - 21,2\% (63.600)$ et $300.000 + 21,2\% \text{ soit entre } 236.400 \text{ et } 363.600$!

Ceci montre qu'avec l'échantillon actuellement utilisé et en négligeant les erreurs autres que celles d'échantillonnage on ne pourra déceler de faibles variations du chômage que quand celui-ci atteindra une très forte proportion de la population active !

Au contraire le calcul d'erreur effectué sur l'estimation du nombre total de personnes de 14 ans et plus fait apparaître un coefficient de variation de 0,9 %. Pour chacune des 8 grandes régions distinguées ce coefficient de variation varie de 2 à 3 % sauf pour le Nord-Est où il atteint 3,7 %. L'erreur sur le nombre global d'heures de travail fournies par la population active est plus élevée : le coefficient de variation est de 3,3 % pour l'ensemble de la France. On a enfin calculé l'erreur sur le quotient des heures travaillées par l'effectif de la population. On obtient un coefficient de variation de 2,9 % (portant sur 28 heures).

En conclusion, il apparaît que cette approche de la mesure des problèmes d'emploi et d'activité se révèle extrêmement féconde tant pour les renseignements périodiques qu'elle permet d'obtenir rapidement que par des informations uniques qu'elle fournit sur des sujets qui n'ont jamais fait l'objet d'une enquête générale. Il faut cependant se souvenir qu'il s'agit d'un sondage et ne pas vouloir lui faire donner la variation des effectifs dans les industries extractives ou des chômeurs secourus dans la Seine. Seul un recensement exhaustif permet de tels détails.

Cette enquête sera reprise cette année. De nouvelles questions remplaceront les anciennes. Le plan d'échantillonnage sera amélioré. On peut espérer qu'un recensement général fournira à la fois une documentation nécessaire à un meilleur plan d'échantillonnage et une base d'extrapolation des résultats.

Vos critiques et vos suggestions enfin nous aideront à rendre l'outil mieux adapté à nos besoins.

Raymond Lévy-BRÜHL.

DISCUSSION

M. RIVET est heureux de rendre un hommage public à ceux de ses collaborateurs et particulièrement à M. Lévy-Bruhl, qui ont mis au point cette enquête par sondage sur l'emploi. Ce travail illustre bien l'exposé du professeur Zwick : en effet, sa préparation a exigé à la fois une culture mathématique étendue, des qualités psychologiques profondes, et une bonne connaissance des conditions économiques et sociales du pays. Il a conduit à des résultats qui consti-

tuent un grand progrès sur la documentation antérieure, instituée essentiellement par le relevé des chômeurs et l'enquête trimestrielle du ministère du Travail sur l'activité économique.

Certes, beaucoup reste encore à faire; l'Institut national de la Statistique et des Études économiques aura à tirer profit des enseignements des deux premières enquêtes, mais à cet égard, on peut avoir confiance dans les spécialistes responsables pour apporter aux futures enquêtes toutes les améliorations désirables.

Pour répondre à plusieurs demandes, M. Rivet apporte également quelques précisions sur le degré de confiance à accorder à certains des résultats obtenus, et sur l'intérêt des résultats même approximatifs.

M. STOETZEL s'associe à M. Rivet pour féliciter le conférencier et lui demander approximativement le coût d'une telle enquête.

M. LÉVY-BRÜHL les évalue approximativement à 500.000 francs pour les frais de déplacements et à 2.000.000 pour les heures de travail, sommes qui apparaissent extrêmement faibles quand on les compare aux crédits consacrés ou envisagés pour des enquêtes du même genre dans les pays étrangers (projet d'enquête du Danemark, par exemple : 10 millions de francs pour 6.000 ménages enquêtés).

M. STOETZEL émet le vœu que lorsqu'une étude statistique est présentée, il soit précisé quel en a été son coût.

M. THIONET précise qu'à son avis le travail des administrateurs représente en plus une dépense de 500.000 francs environ

M. FÉRON. — 1^o J'ai demandé s'il « est possible de combiner l'emploi de questionnaires écrits avec les enquêtes orales ».

2^o Dans une seconde intervention, survenue à propos des différences de point de vue qui opposent M. Stötzsel à M. Lévy-Bruhl, j'ai soutenu le fait que de ne pas interroger les gens au hasard, mais de poser nos questions à tous les individus d'un ménage devait certainement diminuer la précision de l'enquête.

Les paroles que j'ai prononcées étaient donc extrêmement vagues. On peut préciser ma pensée de la manière suivante : « Dans le schéma proposé par M. Lévy-Bruhl on tire d'abord le ménage. Et on interroge ensuite tous les individus des ménages ainsi choisis (On a la probabilité 1 de choisir un individu donné quand le ménage auquel il appartient a été choisi). Donc si π est la probabilité qu'à un ménage d'être tiré, chaque individu d'un ménage de n_k individus a aussi la probabilité $\pi \times 1 = \pi$ d'être tiré. Si donc nous mesurons par ce procédé les valeurs prises par les variables aléatoires $X_1 \dots X_N$ au cours des N épreuves la quantité $\bar{x} = \frac{\sum x_i}{N}$ sera bien une estimation sans biais de la moyenne des espérances mathématiques de tous les éléments de la population.

Par contre la précision du résultat est diminuée par cette méthode de sondage. En effet soient σ_i l'écart type de X_i et ρ_{ij} le coefficient de corrélation entre X_i et X_j . Un calcul simple montre que :

$$\sigma_{\bar{x}}^2 = \frac{1}{N^2} \sum_{i=1}^n \sigma_i^2 + \frac{2}{N^2} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \sigma_i \sigma_j \rho_{ij} \quad (4)$$

Si en particulier pour les n individus d'un ménage $\sigma_i^2 = \sigma^2 \varphi_i = \varphi$ nous avons pour la moyenne \bar{x}_k de ces n individus

$$\sigma_{\bar{x}_k}^2 = \frac{\sigma^2}{n} + \frac{n-1}{n} \sigma^2 \varphi. \quad (2)$$

si nous supposons que les φ_i sont nuls pour des individus appartenant à des ménages différents.

Aussi, si N est le nombre d'individus et F le nombre de familles interrogées, la précision de l'expérience sera comprise entre les précisions qu'on obtiendrait en interrogeant au hasard F ou N personnes, les cas extrêmes étant obtenus pour $\varphi = 1$ (toutes les personnes d'un même ménage répondent exactement la même chose) et pour $\varphi = 0$ (indépendance).

Il convient de remarquer que le problème traité par M. Lévy-Bruhl est différent de celui qui consiste à évaluer la proportion des familles de p enfants en demandant à chaque individu combien il a de frères et sœurs. L'expérimentateur aurait commis une erreur analogue à celle signalée par M. Stötzl s'il avait tiré au hasard les *individus* dans la population et leur avait demandé des renseignements sur les ménages dont ils font partie.

MM. LÉVY-BRÜHL et THIONET estiment qu'il y aurait eu trop peu de réponses, et que celles-ci auraient été peu valables, les enquêteurs n'étant pas sûrs de la bonne compréhension du questionnaire.

M. BATICLE se fait confirmer que les enquêtes du ministère du Travail effectuées, non par sondage, mais à 100 % ont permis d'attester par recouplements de la valeur de la méthode des sondages.

M. DE PORTZAMPARC demande au Dr Zwick comment sont organisées les statistiques en Allemagne : celui-ci précise que c'est le ministère du Travail qui détient toutes les statistiques. Il déclare que malheureusement ce ministère « a la tête carrée et qu'il est difficile d'entrer dedans ».

M. BARRIOL demande que les questions posées ne puissent donner lieu à aucune hésitation ; il vient de recevoir le questionnaire relatif au recensement du Canada et deux questions l'ont embarrassé.

A propos du Canada, il rappelle que le recensement (qui est fait tous les dix ans) est effectué par 20.000 enquêteurs, bien formés à leur tâche qui se rendent dans chaque maison pour recueillir les renseignements et reviennent si cela paraît nécessaire.

La méthode analogue employée par la Statistique générale paraît donner d'excellents résultats.

M. POUDEYROUX demande pourquoi avoir fait des sondages par ménage plutôt que par individu ?

M. STOETZEL signale le problème difficile que constitue le calcul des probabilités lorsque les sondages se font dans des groupes. Il rappelle que le regretté Halbwachs avait déjà signalé ces difficultés dans une enquête portant sur le nombre des membres des familles, la question posée concernant plus précisément le nombre des frères et sœurs des personnes enquêtées.

M. HENON signale deux autres cas voisins, dans lesquels la répartition est différente suivant que le prélèvement a lieu individu par individu ou par groupe.

En premier lieu dans l'industrie textile pour le nombre des fils, et en second lieu la répartition des chiffres d'affaires.

M. POUDEROUX reprenant ces questions demande si l'on a tenu compte des mouvements particuliers de population provoqués par exemple par la construction des barrages.

M. THIONET, pour répondre à M. Pouderoux, rappelle que la population comptée à part est exclue de l'enquête et que la définition officielle de celle-ci mentionne non seulement les militaires, les religieux, les pensionnaires, les aliénés, etc..., mais encore les ouvriers des chantiers temporaires. Au recensement de 1946, ces derniers étaient 27.000 (sur 1.040.000 personnes comptées à part). Il est possible que ce poste de la population comptée à part se soit notablement accru depuis; néanmoins il présente peu d'importance vis-à-vis des erreurs d'échantillonnage.

M. GRONDEIN fait observer que la méthode de sondage employée avait les défauts inhérents à tout sondage. Il est sans doute difficile d'en appliquer une autre, mais le fait que certains résultats obtenus concordent avec les statistiques du ministère du Travail, ne prouve pas son excellence. D'ailleurs l'expérience s'est étendue sur une trop faible période pour qu'on puisse tirer de cette comparaison des conclusions bien solides.

Ce serait trop se hâter que de considérer les taux obtenus comme reflétant exactement la situation de l'ensemble de la population. Il y a trop de causes d'erreurs pour qu'on cherche à déterminer la 4^e décimale alors qu'on n'est pas certain de la 1^{re}. Ces taux, à mon sens, n'ont pas de valeur absolue, mais ont un caractère indicatif, pouvant permettre, par comparaison d'un sondage à l'autre, de déceler les variations importantes et de supputer la tendance.

MM. LÉVY-BRÜHL et THIONET répondent comme suit aux diverses questions posées :

L'unité de sondage adoptée, aussi bien que l'unité statistique, est le ménage, pour un grand nombre de raisons pratiques : possibilité d'obtenir le type de renseignements demandés en n'interrogeant qu'une personne du ménage; intérêt de posséder les réponses relatives à tous les membres d'un ménage; existence de questions propres au ménage (au logement) et non à chaque personne; intérêt d'un échantillon de logements, car un répertoire des logements est beaucoup plus facilement tenu à jour qu'un répertoire de personnes. Il importe d'indiquer ici que les États-Unis et le Canada emploient pour leur « Man Power Survey » un échantillon de ménages et que tous les pays d'Europe qui préparent un tel sondage s'y sont ralliés.

La remarque de M. Stoetzel ne s'applique pas au cas présent. En effet, les difficultés auxquelles il fait allusion sont relatives au tirage d'un échantillon de ménages au moyen de listes d'individus, par exemple, de listes de salariés (feuilles de paie) un ménage ayant alors autant de chances d'être tiré qu'il compte de salariés. Inversement, si l'on tire un échantillon de ménages avec d'égales probabilités (et faute d'employer des coefficients de pondération) le calcul des probabilités montre qu'il y a lieu d'inclure dans l'échantillon tous les membres de ces ménages.

Ceci dit, il n'a jamais été question de prétendre qu'un échantillon de 18.000 personnes, groupées par « grappes » en 6.000 ménages, donnait des résul-

tats aussi précis qu'un échantillon de 18.000 personnes tirées individuellement. De toute façon, la précision des estimations reposant sur l'un ou l'autre de ces 2 échantillons est inférieure à celle qu'on calculerait par la formule élémentaire puisque l'échantillon a été tiré à 2 degrés et non par tirages indépendants. En général, l'échantillon de 6.000 ménages donne des résultats moins précis que celui de 18.000 personnes prises individuellement; mais il existe cependant des cas où les résultats sont plus précis : répartition des personnes suivant le sexe, suivant l'âge et en personnes actives et non actives. C'est la théorie de l'échantillon en grappes (voir HANSEN et HURWITZ, *Relative Efficiencies of Various Sampling Units in Population Inquiries*; *Journal of the American Statistical Society*, mars 1942, p. 89-94.)

A ce propos l'une des observations présentée par écrit par M. Féron n'est pas entièrement valable : le coefficient de corrélation intraclasse (moyenne des ρ_{ij}) peut prendre, non seulement des valeurs positives ou nulles, mais encore des valeurs négatives (sans toutefois atteindre — 1).

Si m est le nombre de personnes par grappe (ce nombre étant supposé commun à toutes les grappes), la corrélation intraclasse peut atteindre $-1/m - 1$. Ce sont précisément à ces valeurs négatives de la corrélation intraclasse que correspondent les échantillons en grappes *plus efficaces* que les échantillons d'individus isolés.
