



HAL
open science

Perte de précision liée au tirage d'un ou plusieurs individus Kish

Nathalie Caron

► **To cite this version:**

| Nathalie Caron. Perte de précision liée au tirage d'un ou plusieurs individus Kish. 1999. <hal-05569760>

HAL Id: hal-05569760

<https://insee.hal.science/hal-05569760v1>

Preprint submitted on 27 Mar 2026

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

INSTITUT NATIONAL DE LA STATISTIQUE ET DES ÉTUDES ÉCONOMIQUES
Série des Documents de Travail
Méthodologie Statistique

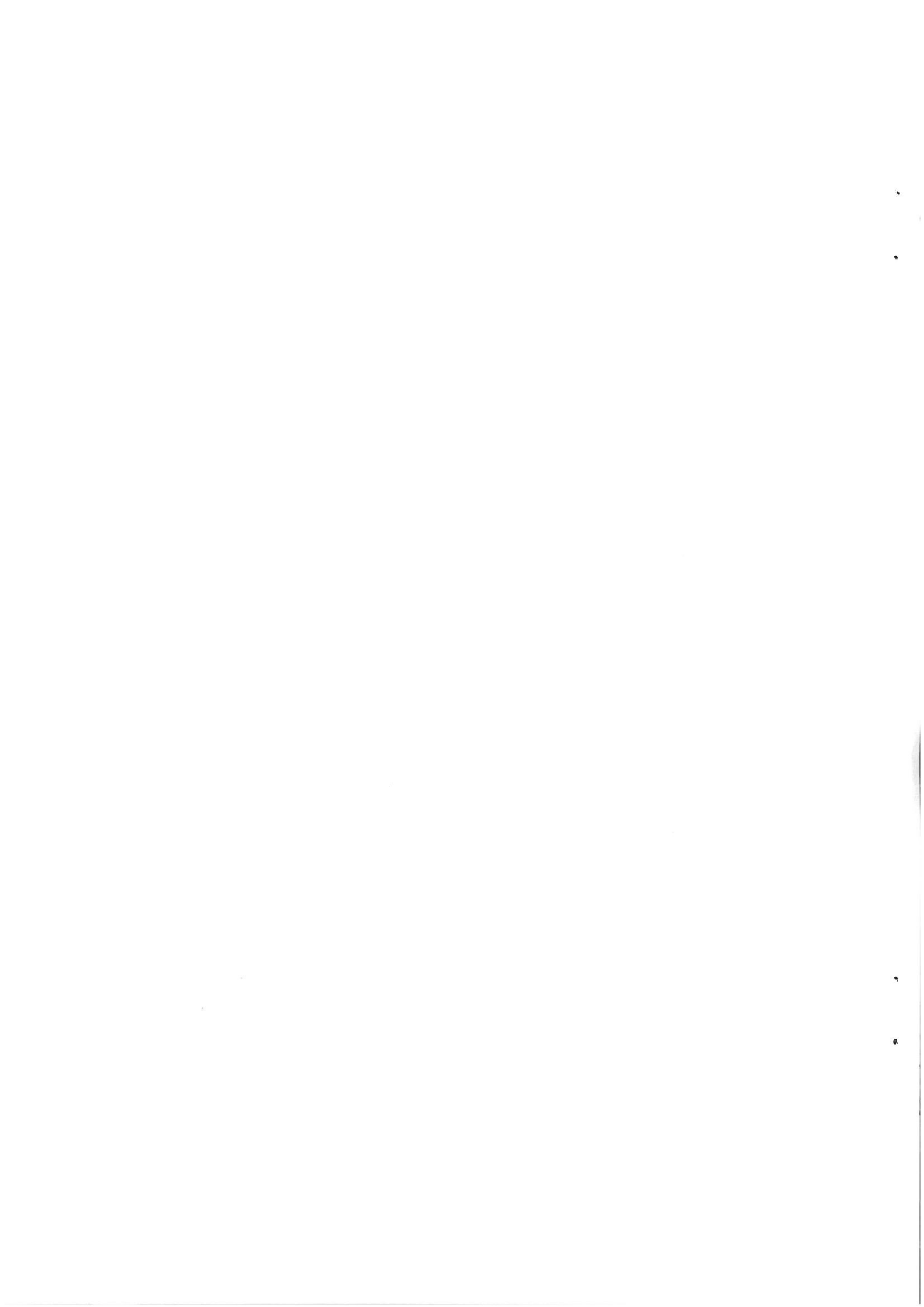
N° 9901

**Perte de précision liée au tirage d'un ou
plusieurs individus Kish**

N. Caron

Février 1999

Ces documents de travail ne reflètent pas la position de l'INSEE et n'engagent que leurs auteurs.
Working papers do not reflect the position of INSEE but only their authors views.



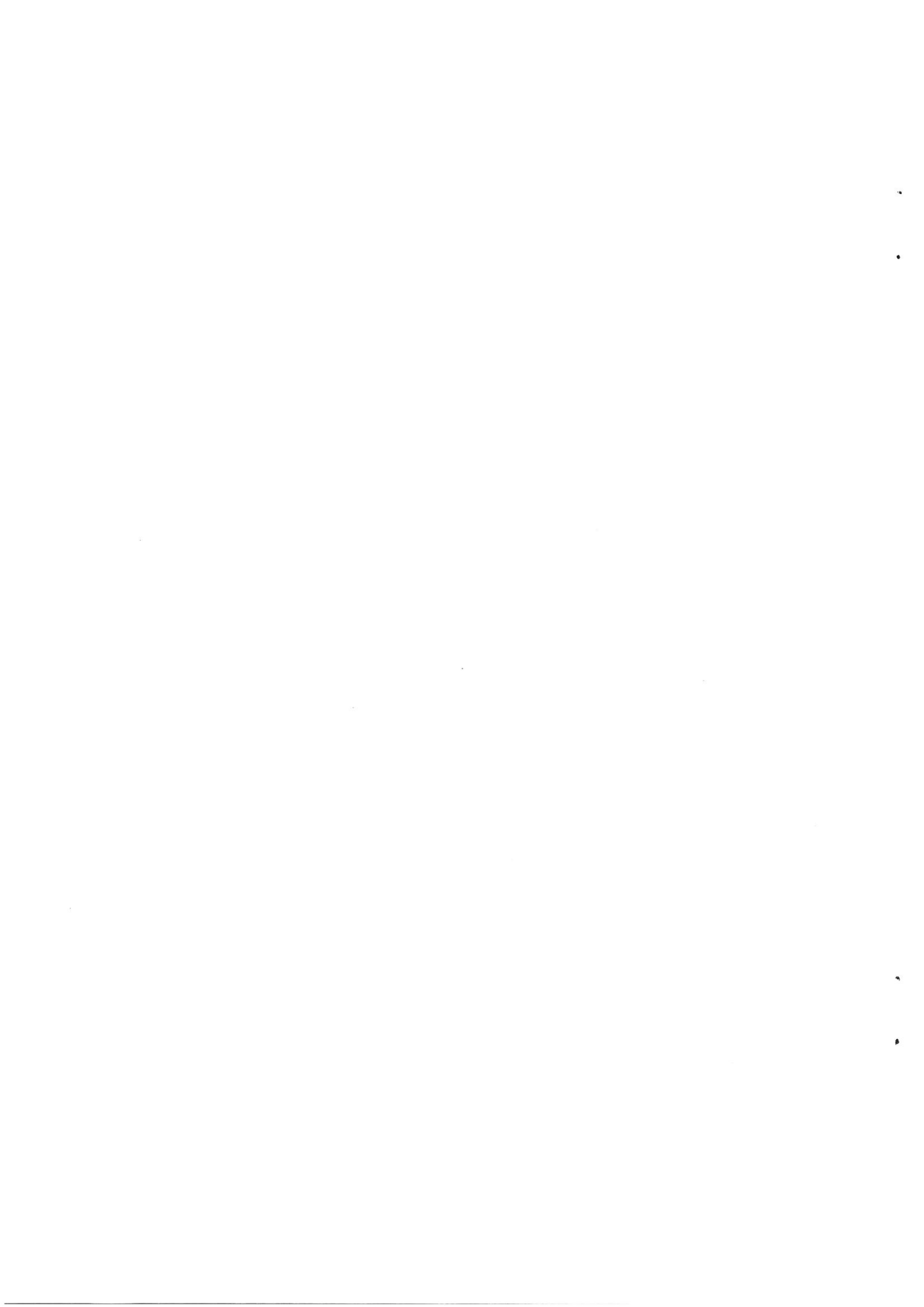
Nathalie CARON

INSEE, CREST - Rennes, Laboratoire de Statistique d'Enquêtes

RÉSUMÉ

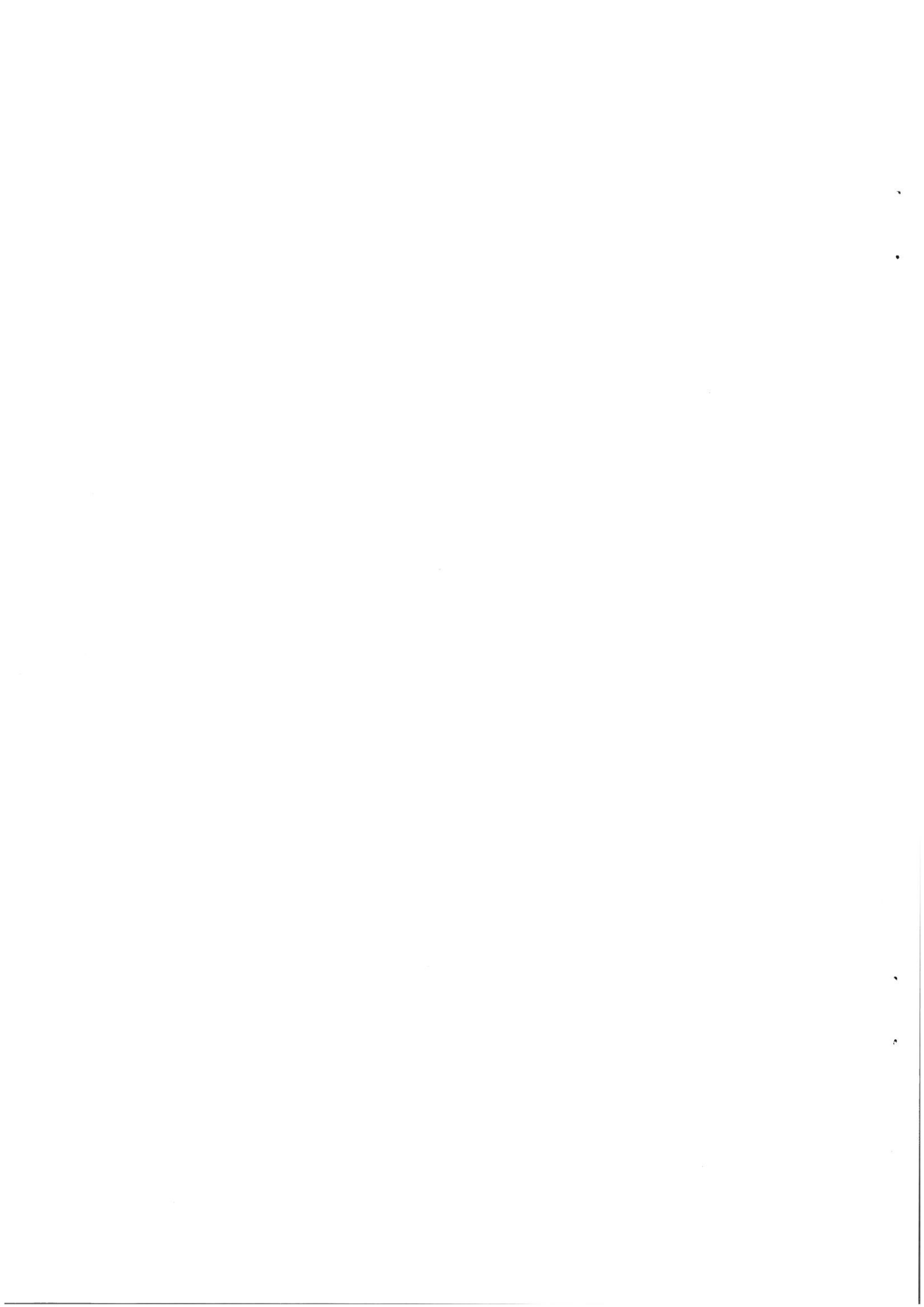
L'unité échantillonnée pour les enquêtes sociales menées à l'Insee est le ménage. Or, il est possible qu'une partie de l'enquête ne s'adresse qu'à un seul individu par ménage. Afin de pouvoir extrapoler correctement les résultats obtenus sur cette partie de questionnaire, une méthode de tirage, appelée tirage Kish, est proposée à l'enquêteur et peut être mise en œuvre immédiatement une fois qu'il connaît le nombre de personnes composant le ménage. Cette méthode, dispersant par nature les poids des individus, introduit une variance supplémentaire. Dans ce document, nous estimons la part de la variance totale liée directement au tirage d'un individu Kish. Ensuite, pour une enquête au niveau individu, nous réalisons une analyse coût-précision de deux modes de collecte possibles : interroger un seul individu Kish par ménage ou interroger l'ensemble des individus éligibles dans un nombre plus restreint de ménages. Cependant, la seconde optique, aussi séduisante soit-elle, conduit à une perte de précision pour des statistiques calculées au niveau ménage. Nous terminerons donc ce document en recherchant des méthodes permettant d'améliorer la méthode Kish traditionnelle. Ces améliorations porteront sur le nombre optimal d'individus à interroger suivant la taille du ménage.

MOTS CLÉS : Individu Kish ; précision ; méthode Kish optimale.



Sommaire

<i>I. Introduction</i>	3
<i>II. Comparaison de la variance obtenue en interrogeant l'ensemble des éligibles et de celle obtenue en sélectionnant un seul individu Kish par ménage.</i>	4
<i>II.1. Méthodologie</i>	4
<i>II.2. Application à une enquête</i>	5
<i>III. Comparaison budgétaire de deux approches différentes conduisant à la même précision</i>	8
<i>IV. Recherche d'une méthode optimale de type Kish.</i>	10
<i>V. Conclusion</i>	12
<i>BIBLIOGRAPHIE</i>	13
<i>ANNEXE</i>	14



I. Introduction

Les enquêtes sociales menées à l'Insee sont le plus souvent réalisées en face à face par des enquêteurs. Les unités échantillonnées sont les ménages qui occupent les logements sélectionnés de façon aléatoire parmi l'ensemble des logements du recensement. Or, en général, une partie de l'enquête s'adresse à un seul individu éligible à un certain titre par ménage. Afin de pouvoir extrapoler correctement les résultats obtenus, il faut que le choix de l'individu au sein du ménage soit lui-même aléatoire et ne se porte pas systématiquement sur les personnes les plus faciles à joindre dans leur logement. Une méthode de tirage, appelée tirage Kish, est proposée à l'enquêteur et peut être mise en œuvre immédiatement une fois qu'il connaît le nombre de personnes composant le ménage.

Apparemment d'emploi facile, cette méthode n'est cependant pas sans reproche. En effet, nous constatons en général une déformation de la population des individus Kish interrogés par rapport à celle des Kish qui auraient dû être interrogés. Cette déformation s'explique principalement par une différence entre la définition théorique des éligibles et celle considérée comme opérationnelle sur le terrain (voir Berthier, Caron et Néros (1998)). De plus, la méthode Kish a comme principal défaut de disperser les poids des individus et par conséquent d'augmenter arbitrairement la variance. Une solution alternative consiste à enquêter l'ensemble des personnes éligibles des ménages quitte à diminuer le nombre de ménages enquêtés. C'est d'ailleurs plutôt cette alternative que préconise Leslie Kish lui-même, l'inventeur de la méthode (Kish, 1995).

Dans les deux premières parties de ce document, nous comparons ces deux méthodes. Ainsi, à partir de l'enquête Permanente sur les Conditions de Vie réalisée en mai 1995 auprès des ménages, nous évaluons la perte de précision due au fait que l'on n'interroge qu'un seul individu par ménage plutôt que d'interroger tous les individus éligibles des ménages (ce qui correspond à un plan de sondage en grappe au niveau ménage). Cette comparaison est basée sur des variables disponibles sur l'ensemble des éligibles et par conséquent sur l'individu Kish. Elle nous permettra d'en déduire un nombre moyen de ménages que l'on doit visiter en supposant que toutes les personnes éligibles sont interrogées pour obtenir la même précision qu'avec des renseignements sur l'individu Kish. Nous terminerons la comparaison en essayant d'estimer le budget nécessaire à la réalisation des deux solutions pour un niveau de précision donné.

Pour établir les premiers résultats, nous avons supposé implicitement que les variables d'intérêt ne concernent que le niveau individu. Cependant, le concepteur d'enquête est en général aussi intéressé par des statistiques construites au niveau ménage. S'il adopte la solution consistant à interroger l'ensemble des personnes éligibles des ménages en diminuant le nombre de ménages enquêtés, les statistiques bâties à partir du questionnaire ménage seront moins précises que dans l'autre méthode. Par conséquent, nous proposons dans la dernière partie de ce document des méthodes permettant d'améliorer la méthode Kish. Ces améliorations porteront sur le nombre optimal d'individus à sélectionner suivant la taille de ménages.

II. Comparaison de la variance obtenue en interrogeant l'ensemble des éligibles et de celle obtenue en sélectionnant un seul individu Kish par ménage.

A partir de l'enquête Permanente sur les Conditions de Vie (PCV) réalisée en mai 1995, nous évaluons dans cette partie la perte de précision due au fait que l'on n'interroge qu'un seul individu par ménage plutôt que d'interroger tous les individus éligibles des ménages (ce qui correspond à un plan de sondage en grappe au niveau ménage).

II.1. Méthodologie

J.C. Deville (1998) montre qu'on peut obtenir facilement une estimation correcte de variance pour des estimateurs construits à partir de variables recueillies sur l'individu Kish. Cette estimation de variance intègre en particulier l'aléa relatif au choix du Kish dans les ménages sélectionnés. Pour cela, il suffit de « remonter » au niveau des ménages la variable observée Y pour l'individu Kish en prenant comme nouvelle variable Z définie pour chaque ménage m par $z_m = n_m y_{mK}$ où n_m est le nombre d'individus éligibles pour le ménage m et y_{mK} est la valeur de la variable Y pour l'individu Kish du ménage m . En réalité, l'estimation de variance obtenue par cette méthode sous-estime la vraie variance mais cette sous-estimation est négligeable dès que les probabilités d'inclusion sont faibles (hypothèse que l'on admettra par la suite). De plus, pour des variables uniquement recueillies sur l'individu Kish, il n'est pas possible d'isoler directement la part due au tirage du Kish. En revanche, si les variables sont recueillies sur l'ensemble de tous les éligibles, cela devient possible en estimant aussi la variance pour des estimateurs construits sur l'ensemble des éligibles. Pour cela, les variables du niveau individu sont « remontées » au niveau ménage, c'est-à-dire que pour chaque ménage nous avons créé pour chacune des variables relatives à tous les individus une variable qui correspond à la somme de cette variable pour l'ensemble des éligibles du ménage. Le calcul d'estimation de variance a été effectué sur cette variable.

La variance totale en présence d'un individu Kish peut s'écrire sous la forme :

$$V(\hat{Z}) = V_S(E(\hat{Z}|s)) + E_S(V(\hat{Z}|s))$$

où $E(\hat{Z}|s)$ et $V(\hat{Z}|s)$ représentent respectivement l'espérance et la variance conditionnelles à l'échantillon s de l'estimateur \hat{Z} du total Z de la variable Z .

Conditionnellement à s , les $z_m = n_m y_{mK}$ sont des variables indépendantes d'espérance $y_m = \sum_{i \in k} y_{mi}$, total de la variable Y pour l'ensemble des éligibles du ménage m . En notant θ_m la variance statistique de z_m , la variance devient :

$$\begin{aligned} V(\hat{Z}) &= V_S\left(\sum_{m \in s} w_m y_m\right) + E_S\left(\sum_{m \in s} w_m^2 \theta_m\right) & (I) \\ &= \textcircled{1} \quad + \quad \textcircled{2} \end{aligned}$$

où w_m est le poids d'extrapolation du ménage m .

La première partie de la variance (terme ①) correspond à la variance que l'on aurait obtenue si on avait interrogé l'ensemble de tous les éligibles des ménages. La seconde partie (terme ②) correspond au supplément de variance imputable au tirage de l'individu Kish. La variance totale obtenue pour des estimateurs construits à partir de variables recueillies sur l'individu Kish est par conséquent supérieure à celle qu'on aurait obtenue avec des estimateurs construits sur l'ensemble des éligibles.

En ne sélectionnant qu'un seul individu Kish par ménage, la variance statistique θ_k de la variable Z au sein du ménage k et par conséquent le terme ② ne peuvent pas être estimés directement. Cependant, en réalisant les calculs d'estimation de variance pour des variables connues pour l'ensemble des tous les individus éligibles et par conséquent pour l'individu Kish, on en déduit une estimation du terme ① et de la variance totale en présence d'un Kish (terme ①+ terme ②). Ainsi, par différence de ces deux estimations, la part de la variance imputable directement à l'individu Kish c'est-à-dire le terme ② peut être estimée.

II.2. Application à une enquête

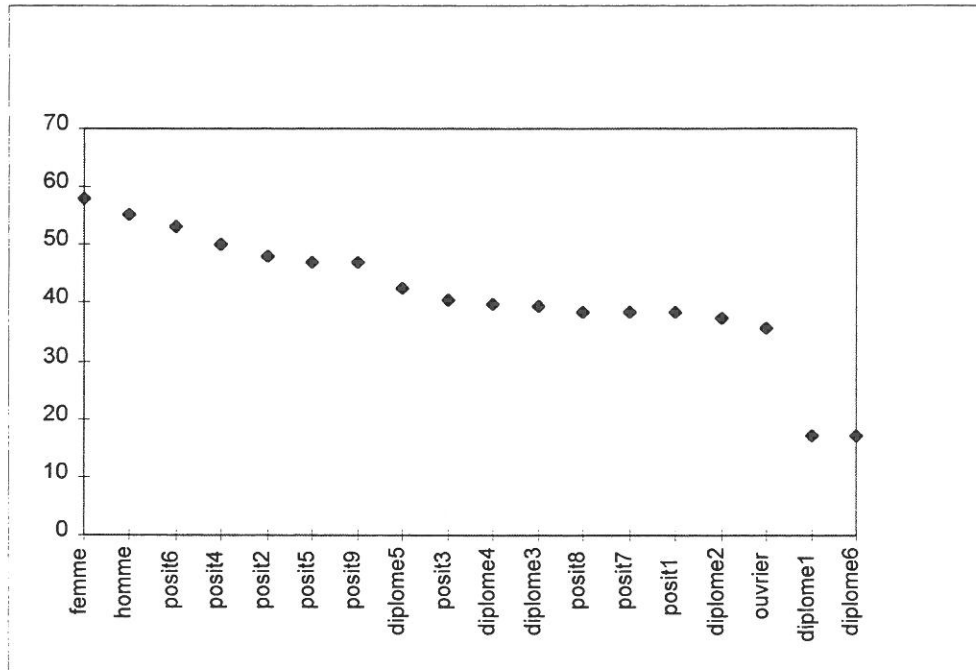
Nous avons comparé la précision d'un certain nombre de variables connues simultanément sur l'ensemble des éligibles et par conséquent sur l'individu Kish : le diplôme (en 6 modalités), la position professionnelle (en 9 modalités), le sexe et le fait d'être ou non ouvrier. La définition exacte de chacune des modalités est précisée en annexe. Le fichier sur lequel nous avons travaillé comporte 12333 éligibles et 5954 individus Kish.

Les ménages interrogés pour l'enquête PCV ont été extraits de l'échantillon maître 1990 (EM90) c'est-à-dire qu'ils ont été sélectionnés par un plan de sondage stratifié à plusieurs degrés selon la taille de l'unité urbaine. La précision obtenue par le plan de sondage de l'EM90 a été calculée en utilisant la méthode simplifiée programmée dans le logiciel POULPE (voir Caron, Deville, Sautory, (1998)), qui consiste à approximer le plan de sondage initial à plusieurs degrés en un plan de sondage dont le tirage des unités primaires a été réalisé avec remise.

Nous constatons que la variance obtenue dans le cas d'un tirage Kish est supérieure à celle obtenue en interrogeant l'ensemble de tous les éligibles pour les variables étudiées (voir tableau 1). Le rapport de variance varie entre 1.2 pour l'estimateur du nombre de personnes sans diplôme (variable DIPLOME1) et 2.4 pour l'estimateur du nombre de femmes (variable FEMME) ce qui situe la part de variance imputable directement au tirage de l'individu Kish entre 17% et 57% pour les mêmes variables (voir graphique 1). Comme les ménages sont en majorité composés de deux éligibles formant un couple, la personne de référence et son conjoint, l'effet grappe induit par l'appartenance au même ménage est très faible pour les estimateurs du nombre d'hommes et du nombre de femmes. Il en est aussi de même pour toutes les variables fortement corrélées avec le sexe (par exemple, plus de 75% des ouvriers qualifiés (variable POSIT2) et des agents de maîtrise (variables POSIT3 et POSIT4) sont des hommes). Ainsi, nous constatons empiriquement à partir du graphique 1 que la part imputable au tirage Kish est d'autant plus importante que l'effet grappe est faible. Ce résultat peut se

déduire directement de la formule de variance (I) . En effet, la part imputable directement au tirage du Kish s'écrit $\sum_{m \in S} w_m^2 \theta_m$; elle est donc d'autant plus élevée que la variance statistique θ_m de z_m est importante et par conséquent que l'effet grappe est faible.

Graphique 1 : Part de la variance imputable au tirage Kish (en %) ¹



Cette étude suggère aussi que l'on peut obtenir la même précision que celle obtenue avec un seul Kish par ménage en interrogeant l'ensemble des individus éligibles des ménages pour un nombre plus réduit de ménages (c'est-à-dire en interrogeant moins de 5954 ménages). Ainsi, par exemple, pour l'estimation du nombre de femmes, la variance estimée avec l'optique Kish est 2.4 fois plus importante que celle obtenue en interrogeant l'ensemble des 12333 éligibles. En considérant que les variances sont inversement proportionnelles au nombre de ménages sélectionnés, on en déduit que l'on peut obtenir la même précision qu'avec le plan de sondage optique Kish avec un échantillon de $5954/2.4=2480$ ménages dont tous les individus éligibles sont enquêtés. Par contre, le même calcul pour l'estimation du nombre de personnes ayant un diplôme supérieur (diplome6) conduit à une taille d'échantillon d'environ 4930. La réduction de la taille d'échantillon dépend fortement des variables : plus l'effet de grappe est important, moins le gain est élevé. Les différentes tailles d'échantillon sont précisées pour chaque variable dans la colonne « Taille fictive (Nbre de ménages) » du tableau 1. La taille moyenne est de l'ordre de 3500 ménages.

¹ Les variables sont classées par ordre décroissant de la part de la variance imputable au tirage Kish qui est égale à $(\text{est. de variance optique Kish} - \text{est. de variance optique individus}) / \text{est. de variance optique Kish}$.

Variables	Estimation de variance totale optique Individus	Estimation de variance totale optique Kish	Taille fictive
<i>diplome1²</i>	8,6E+10	1,04E+11	4923
<i>diplome2</i>	3,7E+10	5,9E+10	3734
<i>diplome3</i>	2,3E+10	3,8E+10	3604
<i>diplome4</i>	5,6E+10	9,3E+10	3585
<i>diplome5</i>	3,1E+10	5,4E+10	3418
<i>diplome6</i>	5,3E+10	6,4E+10	4931
<i>posit1³</i>	2,59E+10	4,2E+10	3672
<i>posit2</i>	2,7E+10	5,2E+10	3091
<i>posit3</i>	5,4E+09	9,1E+09	3533
<i>posit4</i>	1,7E+09	3,4E+09	2977
<i>posit5</i>	8,8E+09	1,66E+10	3156
<i>posit6</i>	9,9E+09	2,11E+10	2794
<i>posit7</i>	1,17E+10	1,9E+10	3666
<i>posit8</i>	5,78E+09	9,4E+09	3661
<i>posit9</i>	4,3E+10	8,1E+10	3161
<i>homme</i>	6,2E+10	1,39E+11	2656
<i>femme</i>	5,6E+10	1,33E+11	2507
<i>ouvrier</i>	5,6E+10	8,7E+10	3833

Tableau 1 : Comparaison des estimations de variances

² diplome1----diplome6 représentent les 6 modalités de la variable diplôme décrite en annexe.

³ posit1----posit9 représentent les 9 modalités de la variable position professionnelle décrite en annexe.

III. Comparaison budgétaire de deux approches différentes conduisant à la même précision

L'objet de cette partie est d'estimer grossièrement le budget nécessaire à la réalisation d'une enquête en deux parties dont la seconde partie ne concerne qu'un seul individu choisi par la méthode Kish parmi l'ensemble des éligibles (1ère optique) et de le comparer à celui obtenu en considérant que l'on choisit d'interroger tous les éligibles (2ème optique) pour un nombre réduit de ménages conduisant à une précision comparable (3500 ménages répondants contre 5934 ménages répondants). Pour chiffrer le budget, nous n'avons considéré que les frais directs liés à la collecte. En particulier, les économies réalisées sur le nombre d'heures manuelles, le nombre d'heures de saisie et la gestion de l'enquête n'ont pas été prises en compte.

Pour le calcul, nous avons retenu les hypothèses suivantes :

- 8000 logements ont été sélectionnés pour l'enquête PCV et 5954 ménages ont effectivement répondu à cette enquête. Le taux de déchet est approximativement de 25.5%. Nous appliquons le même taux de déchet dans les deux optiques. Ainsi pour obtenir 3500 ménages répondants, il faut en sélectionner environ 4700.
- On suppose que l'enquête PCV est composée de deux questionnaires de 20 minutes. Le premier concerne l'ensemble du ménage et le second est un questionnaire individuel. Ce dernier doit être rempli par une seule personne choisie au hasard par la méthode Kish dans la première optique et par l'ensemble des individus éligibles dans la seconde.
- Dans l'enquête PCV, un ménage contient en moyenne 2 individus éligibles (i.e. de plus de 15 ans).
- Dans l'optique Kish, le questionnaire ménage est passé lors de la première visite. Le questionnaire individu est passé dans 80% des cas lors de la première visite et pour 20% des cas lors d'une seconde visite. Cette hypothèse est sans doute très optimiste. Dans l'optique individus, le questionnaire ménage est passé lors de la première visite. Comme on estime à deux le nombre de personnes éligibles, on supposera que le 1er questionnaire individu est passé avec la personne ayant répondu au questionnaire ménage lors de la première visite. Par contre, le second questionnaire est passé au cours d'une seconde visite.

	Optique individus⁴	Optique Kish
Frais déplacements	4700 * FC	8000 * FC
Frais Déchets	(4700-3500)*FD	(8000-5954)* FD
Frais Questionnaire ménage	[51+26+14]*3500	[51+26+14]*5954
Frais Questionnaire(s) individu	[10+26+14]*3500 + [26+26+14]*3500	[10+26+14]*5954*0.80 + [26+26+14]*5954*0.20

⁴ Pour évaluer le budget, nous avons repris les éléments suivants du barème de rémunération (notes n°121/C504/DB/AG et n°066/C504/FD/MF) :

- * Les frais de déplacement sont fixés à 35 F (notés FC) et les déchets sont rémunérés 25F (notés FD).
- * Pour une première visite, la rémunération de base du premier questionnaire est de 51F, celle du questionnaire suivant 10F. Les 10 premières minutes sont rémunérées 26F et les minutes suivantes 1F40.
- * Pour les autres visites, la rémunération de base du premier questionnaire est de 26F. Les 10 premières minutes sont rémunérées 26F et les minutes suivantes 1F40.

Le budget nécessaire selon l'optique Kish est, avec les hypothèses envisagées, de 1 189 717 Francs. Par contre, selon l'optique individus le budget nécessaire est avec les hypothèses envisagées de 919 000 Francs

Ainsi, avec une précision identique pour les variables du niveau individu et sous les hypothèses indiquées ci-dessus, il est plus intéressant financièrement de réaliser une enquête auprès de l'ensemble de tous les éligibles plutôt qu'un seul individu sélectionné par un tirage Kish. Le gain estimé est de 200 000 Francs, ce qui représente environ 1/6 du budget relatif à la collecte.

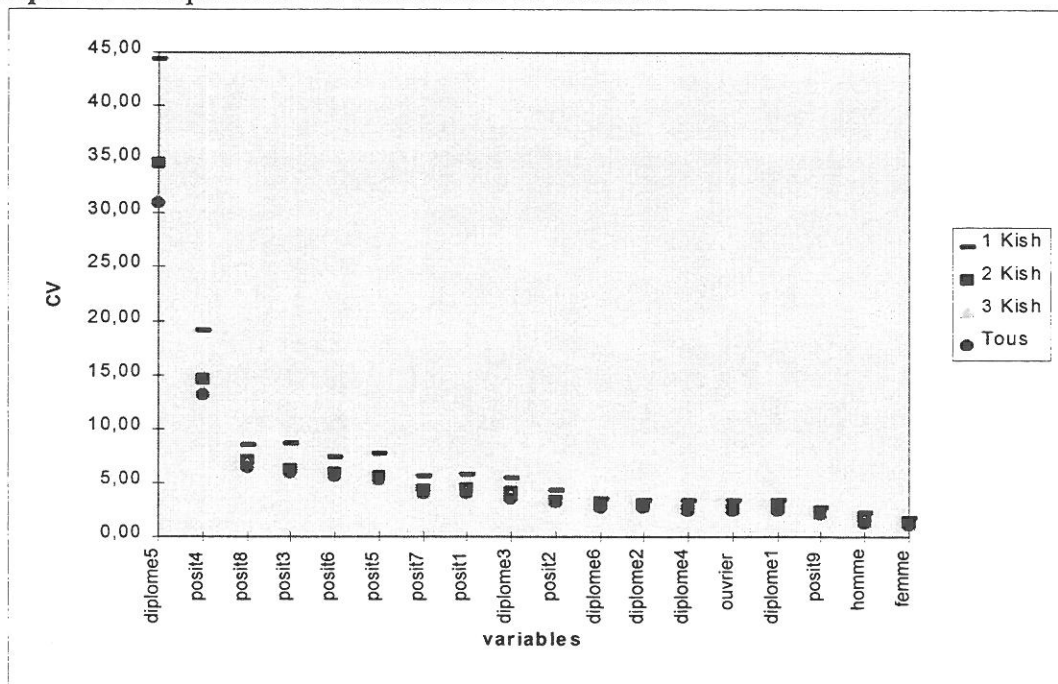
IV. Recherche d'une méthode optimale de type Kish.

Dans les parties précédentes, nous avons considéré que le principal objectif de l'enquête est de recueillir des renseignements au niveau des individus ; par conséquent, l'entité ménage n'intervient qu'en tant qu'unité échantillonnée dans la base de sondage. Cependant, le concepteur d'enquête est en général aussi intéressé par des statistiques construites au niveau ménage. Ainsi, s'il adopte la solution consistant à interroger l'ensemble des personnes éligibles des ménages en diminuant le nombre de ménages enquêtés, les statistiques bâties à partir du questionnaire ménage seront moins précises que dans l'autre solution. Notons qu'une partie du gain estimé dans la partie précédente peut servir à interroger plus de ménages pour ne pas trop dégrader les statistiques obtenues sur des ménages.

Dans cette partie, nous proposons des méthodes permettant d'améliorer la méthode Kish élémentaire qui consiste à sélectionner un seul individu par ménage à probabilités égales. Les améliorations proposées portent sur le nombre d'individus à sélectionner suivant la taille du ménage. Nous envisagerons deux possibilités : l'une consistant à choisir deux individus Kish par ménage (ou plus exactement l'ensemble des individus éligibles pour les ménages d'un ou deux éligibles et deux éligibles sélectionnés au hasard parmi l'ensemble des éligibles pour les ménages d'au moins trois éligibles) et l'autre consistant à choisir trois éligibles par ménage (ou plus exactement l'ensemble des individus éligibles pour les ménages d'un, deux ou trois éligibles et trois éligibles sélectionnés au hasard parmi l'ensemble des éligibles pour les ménages d'au moins quatre éligibles).

Nous nous intéressons au gain obtenu au niveau de la précision des estimateurs en sélectionnant deux individus Kish, trois individus Kish puis l'ensemble des éligibles par rapport au tirage d'un seul individu Kish. Les coefficients de variation (correspondant au rapport de l'écart-type sur l'estimation) permettent de faire directement des comparaisons pour tout type de variable. Par conséquent, nous avons comparé les coefficients de variation obtenus pour ces différentes hypothèses pour l'ensemble des variables utilisées dans la partie II.

Graphique 2 : Comparaison des coefficients de variation⁵



⁵ Les variables sont classées par ordre décroissant du coefficient de variation (pour l'ensemble des individus).

Le gain en terme de précision est très appréciable lorsqu'on passe de 1 à 2 individus Kish par ménage et est nettement plus faible lorsque l'on passe de 2 à 3 individus Kish (voir graphique 1). D'autre part, il est d'autant plus important que la statistique est estimée de façon peu précise dans l'enquête (comme le nombre de personnes ayant le niveau bac ou le nombre d'agents de maîtrise). De plus, les estimations de variance obtenues dans le cas d'un tirage de trois éligibles sont très proches de celles obtenues en considérant l'ensemble des éligibles.

Ces résultats s'expliquent en grande partie par la répartition du nombre de ménages selon le nombre d'éligibles (voir tableau 2).

Nombre d'éligibles	Répartition des ménages	Répartition cumulée des ménages
1	29.2 %	29.2%
2	49.1%	78.3%
3	13.7%	92.0%
4	6.2%	98.2%
5	1.4%	99.6%
6	0.3%	99.9%
7 et plus	0.1%	100%

Tableau 2 : Répartition des ménages selon le nombre d'éligibles

En ne sélectionnant qu'un seul individu par ménage, nous constatons ainsi que le tirage est réalisé dans plus des $\frac{3}{4}$ des ménages. Par contre, en enquêtant deux individus Kish par ménage, dans 78% des cas, il n'y a aucun tirage puisque l'on interroge l'ensemble des éligibles. De même, les ménages de plus de trois éligibles sont en proportion négligeable dans le fichier traité (8% des ménages).

Finalement, ces résultats indiquent qu'il ne semble pas nécessaire de prévoir un budget permettant d'interroger trois individus Kish par ménage. On obtient quasiment la même précision en ne sélectionnant que deux individus par ménage. Par contre, il serait souhaitable d'interroger plus d'un individu Kish par ménage en diminuant légèrement le nombre de ménages sélectionnés.

V. Conclusion

A partir de l'étude réalisée sur l'enquête Permanente sur les Conditions de Vie, nous constatons que la perte de précision due au fait que l'on n'interroge qu'un seul individu par ménage plutôt que d'interroger l'ensemble de tous les individus éligibles des ménages peut être relativement importante. Ainsi, la part de variance imputable directement au tirage de l'individu Kish varie entre 17% et 57% sur les différentes variables étudiées. Cette étude suggère ainsi que l'on peut obtenir la même précision que celle obtenue avec un seul Kish par ménage en interrogeant tous les individus éligibles des ménages en sélectionnant moins de ménages que dans la première optique. De plus, le budget nécessaire est inférieur.

Dans un second temps, nous proposons des méthodes permettant d'améliorer la méthode Kish élémentaire qui consiste à sélectionner un seul individu par ménage à probabilités égales. Nous avons ainsi comparé la précision obtenue en sélectionnant un, deux ou trois individus Kish ou l'ensemble de tous les individus éligibles. Nous constatons que **le gain** en terme de précision est **très appréciable lorsqu'on passe de 1 à 2 individus Kish** par ménage et est nettement plus faible lorsque l'on passe de 2 à 3 individus Kish. Ainsi, deux individus Kish sélectionnés dans chaque ménage suffisent, la précision marginale apportée par le troisième individu étant négligeable. De plus, le gain est d'autant plus important que la statistique est estimée de façon peu précise dans l'enquête (comme le nombre de personnes ayant le niveau bac ou le nombre d'agents de maîtrise).

BIBLIOGRAPHIE

BERTHIER, C., CARON, N., NEROS, B. (1998) : "le Kish : les problèmes de réalisation du tirage et de son extrapolation", série Méthodologie Statistique, n°9810, INSEE.

CARON, N., DEVILLE, J.-C., SAUTORY (1998) : "le logiciel POULPE, Aspects Méthodologiques", série Méthodologie Statistique, n°9806, INSEE.

DEVILLE, J.-C.(1998) : "Pour essayer d'en finir avec l'individu Kish", document de travail n°F9803 de la série Méthodologie Statistique, INSEE.

DEVILLE, J.-C.(1998) : "Variance et estimation de variance en cas d'erreurs de mesure non corrélées ou de l'intrusion d'un individu Kish", série Méthodologie Statistique, n°9805, INSEE.

KISH, L.(1965) : *Survey sampling*, New-York, Wiley.

KISH, L.(1995) : Questions/Answers from the survey statistician, IASS.

SÄRNDAL, C.-E., SWENSSON, B., WRETMAN, J.(1993) : *Model assisted Survey Sampling*, New-York, Springer-Verlag.

ANNEXE

La variable DIPLOME décrivant le diplôme le plus élevé obtenu est en 6 modalités :

- * *diplome1* : Aucun diplôme
- * *diplome2* : Certificat d'études primaires (CEP), diplôme de fin d'études obligatoires (DFEO)
- * *diplome3* : Brevet d'études du 1er cycle (BEPC), brevet des collèges.
- * *diplome4* : Diplôme de l'enseignement court technique ou professionnel (CAP, CAPA, BEP, BEPA, EFAA, autres diplômes de niveau CAP ou BEP)
- * *diplome5* : Diplôme de niveau BAC y compris technique ou professionnel.
- * *diplome6* : Diplôme de l'enseignement supérieur y compris technique supérieur

La variable POSIT décrivant la position professionnelle de l'emploi est en 9 modalités :

- * *posit1* : Manœuvre ou ouvrier spécialisé
- * *posit2* : Ouvrier qualifié ou hautement qualifié
- * *posit3* : Agent de maîtrise dirigeant des ouvriers, maîtrise administrative ou commerciale
- * *posit4* : Agent de maîtrise dirigeant des techniciens ou d'autres agents de maîtrise
- * *posit5* : Technicien, dessinateur, VRP (non cadre)
- * *posit6* : Instituteur, assistant social, infirmier et personnel de catégorie B de la fonction publique
- * *posit7* : Ingénieur ou cadre
- * *posit8* : Professeur et personnel de catégorie A de la fonction publique
- * *posit9* : Employé de bureau, employé de commerce, agent de service, aide soignant, gardienne d'enfants, personnel des catégories C ou D de la fonction publique.

Homme : nombre d'hommes

Femme : nombre de femmes

Ouvrier : nombre d'ouvriers

Série des Documents de Travail
'Méthodologie Statistique'

9601 : 'Une méthode synthétique, robuste et efficace pour réaliser des estimations locales de population'

G. DECAUDIN, J.-C. LABAT

9602 : 'Estimation de la précision d'un solde dans les enquêtes de conjoncture auprès des entreprises'

N. CARON, P. RAVALET, O. SAUTORY

9603 : 'La procédure FREQ de SAS[®] - Tests d'indépendance et mesures d'association dans un tableau de contingence'

J. CONFAIS, Y. GRELET, M. LE GUEN

9604 : 'Les principales techniques de correction de la non-réponse et les modèles associés'

N. CARON

9605 : 'L'estimation du taux d'évolution des dépenses d'équipement dans l'enquête de conjoncture : analyse et voies d'amélioration'

P. RAVALET

9606 : 'L'économétrie et l'étude des comportements. Présentation et mise en œuvre de modèles de régression qualitatifs. Les modèles univariés à résidus logistiques ou normaux (LOGIT, PROBIT)'

S. LOLLIVIER, M. MARPSAT, D. VERGER

9607 : 'Enquêtes régionales sur les déplacements des ménages : l'expérience de Rhône-Alpes'

N. CARON, D. LE BLANC

9701 : 'Une bonne petite enquête vaut-elle mieux qu'un mauvais recensement ?'

J.C. DEVILLE

9702 : 'Modèles univariés et modèles de durée sur données individuelles'

S. LOLLIVIER

9703 : 'Comparaison de deux estimateurs par le ratio stratifiés et application aux enquêtes auprès des entreprises'

N. CARON, J.C. DEVILLE

9704 : 'La faisabilité d'une enquête auprès des ménages

1. au mois d'août. 2. à un rythme hebdomadaire'

C. LAGARENNE, C. THIESSET

9705 : 'Méthodologie de l'enquête sur les déplacements dans l'agglomération toulousaine'

P. GIRARD

9801 : 'Les logiciels de désaisonnalisation TRAMO & SEATS : philosophie, principes et mise en œuvre sous SAS'

K. ATTAL-TOUBERT, D. LADIRAY

9802 : 'Estimation de variance pour des statistiques complexes : technique des résidus et de linéarisation'

J.C. DEVILLE

9803 : 'Pour essayer d'en finir avec l'individu Kish'

J.C. DEVILLE

9804 : 'Une nouvelle (encore une !) méthode de tirage à probabilités inégales'

J.C. DEVILLE

9805 : 'Variance et estimation de variance en cas d'erreurs de mesure non corrélées ou de l'intrusion d'un individu Kish'

J.C. DEVILLE

9806 : 'Estimation de précision de données issues d'enquêtes : document méthodologique sur le logiciel POULPE'

N. CARON, J.C. DEVILLE, O. SAUTORY

9807 : 'Estimation de données régionales à l'aide de techniques d'analyse multidimensionnelle'

K. ATTAL-TOUBERT, O. SAUTORY

9808 : 'Matrices de mobilité et calcul de la précision associée'

N. CARON, C. CHAMBAZ

9809 : 'Echantillonnage et stratification : une étude empirique des gains de précision'

J. LE GUENNEC

9810 : 'Le Kish : les problèmes de réalisation du tirage et de son extrapolation'
C. BERTHIER, N. CARON, B. NÉROS

9811 : 'Vocabulaire statistique Français - Chinois - Anglais'
LIU Xiaoyue, CUI Bin

9901 : 'Perte de précision liée au tirage d'un ou plusieurs individus Kish'
N. CARON

