

## 4. QUATRIÈME PARTIE : ASPECTS STATISTIQUES

**Auteur : Emmanuel Massé**

L'objectif de cette partie est de présenter l'essentiel de la méthodologie statistique utilisée au cours de l'enquête SD2001. Cette opération présente un grand nombre d'innovations. L'absence de base de sondage des personnes sans-domicile (voir première partie) a conduit à organiser un mode d'échantillonnage de type indirect. Ne pouvant échantillonner directement les individus, il a été décidé d'échantillonner des repas et des nuitées d'hébergement au cours de la période d'enquête puis d'interroger les personnes correspondantes (utilisatrices de la prestation repas ou nuitée). Pour obtenir des statistiques sur les individus (et non sur les services qu'ils utilisent), on utilise alors la méthode dite du « partage des poids ». Celle-ci s'appuie sur la connaissance qu'ont les personnes interrogées de leur fréquentation des services. Le calcul des pondérations nécessite outre le calcul des probabilités d'inclusion découlant du plan de sondage et l'application de la méthode du partage des poids, l'application d'un certain nombre d'opérations de redressement. Il faut en particulier tenir compte de l'obsolescence de la base de sondage<sup>40</sup>. On observe au final une dispersion importante des pondérations des individus et donc des variances élevées sur les estimateurs.

Quelles sont les limites ou inconvénients de la méthode du partage des poids ?

D'une part le champ de l'enquête se limite aux usagers des services de la période de référence (c'est-à-dire la période sur laquelle on calcule le nombre de prestations utilisées par les individus interrogés), ce qui conduit à sous-estimer (voir partie 5) le nombre de personnes sans-domicile. D'autre part plus la période de référence utilisée est longue (ce qui permet d'augmenter le nombre de personnes dans le champ de l'enquête), plus le mode de collecte rétrospectif des niveaux de fréquentation des services produit un biais (du fait du renouvellement permanent des populations dans les structures d'aide, le passé reflète incorrectement l'utilisation qui serait faite des services pendant la période de référence).

### 4.1 Le plan de sondage des prestations individuelles

La première étape de la méthodologie adoptée pour l'enquête SD2001 consiste à réaliser un plan d'échantillonnage des nuitées et des repas fournis par les structures d'accueil pendant la période d'enquête.

#### 4.1.1 Caractéristiques de l'échantillon

Les caractéristiques principales de l'échantillon de l'enquête auprès des personnes sans-domicile sont :

- Le champ de l'enquête est constitué des personnes recourant à au moins un service d'aide aux personnes sans-domicile (les hébergements et les distributions de repas chauds aussi appelées services de restauration, voir partie 1) pendant la période d'enquête. En outre, on se limite à la partie du milieu urbain constituée par les agglomérations de plus de vingt mille habitants (au sens du RP 1990).
- La taille de l'échantillon initialement tiré est de 4400 unités statistiques, l'objectif était d'obtenir environ 4000 questionnaires remplis. Au vu des tests, compte tenu des taux d'acceptation des structures et de leurs usagers et de la pratique des remplacements, on anticipait en effet environ 10 % de pertes parmi les services.
- La période de collecte s'est étendue du 15 janvier au 12 Février 2001 (hors semaine complémentaire).

#### 4.1.2 Plan de sondage

Les notions de structure, de service et de visite ont été définies dans le paragraphe 2.4.2.

L'unité de tirage finale de l'enquête est la **prestation individuelle**, c'est-à-dire l'utilisation de l'un des services un jour donné par une personne, il peut donc s'agir d'un repas ou d'une nuitée dans une structure.

Le plan de sondage des prestations individuelles est à trois degrés :

1. Unité primaire : agglomérations de plus de 20000 habitants
2. Unité secondaire : couples (service, jour)
3. Unité tertiaire : prestations individuelles

<sup>40</sup> Celle-ci a été constituée au cours de l'année 2000 alors que l'enquête a eu lieu en janvier et février 2001.

Regardons maintenant plus en détail chacun des degrés du plan de sondage. Gardons à l'esprit qu'on ne parle pas ici des probabilités de tirage des individus mais bien des probabilités de tirage des prestations individuelles au cours du mois d'enquête. Ce qui signifie en particulier que plusieurs prestations échantillonnées au cours de cette période pourront être consommées par un même individu (ce sont des « doublons »). Une même personne peut être sélectionnée le 15 janvier dans un centre d'hébergement et le 12 février dans une distribution de repas chauds. En outre, pour éviter une trop grande dispersion des pondérations (voir section 6 de cette partie), on a comme objectif d'avoir pour toutes les prestations des probabilités de tirage les plus proches possibles.

#### 4.1.2.1 Unité primaire

L'objectif d'un tirage d'agglomérations au premier degré était d'éviter d'avoir à réaliser un inventaire des structures sur l'ensemble de la France, donc de limiter le nombre d'enquêtes préliminaires auprès des services d'aide. Les agglomérations ont été échantillonnées par un tirage proportionnel à une estimation du nombre de prestations individuelles servies dans l'agglomération.

Plus précisément, avec les notations suivantes :

- $T_i$  le nombre journalier de prestations individuelles dans l'agglomération  $i$  un jour donné
- $\Pi_i$  la probabilité de tirage de l'agglomération  $i$  ( $i$  allant de 1 à 80)
- $a$  le nombre d'agglomérations échantillonnées ( $a = 80$ )

On souhaiterait avoir comme probabilité de tirage d'une agglomération le cas d'un tirage proportionnel à la taille, c'est-à-dire :

$$\Pi_i = a \frac{T_i}{\sum_i T_i}$$

En fait, l'information sur le nombre des prestations individuelles servies aux sans-domicile n'était pas disponible au moment du tirage des agglomérations ; les effectifs ont donc été estimés à partir de deux variables auxiliaires :

1. la population des agglomérations lors du RP de 1990 ;
2. une approximation d'une partie des capacités d'hébergement (à partir du fichier Finess)<sup>41</sup>

La probabilité de tirage finalement obtenue est alors<sup>42</sup> :

$$\Pi_i^* = a \frac{T_i^*}{\sum_i T_i^*}$$

où les  $T_i^*$  sont des valeurs approchées des  $T_i$ .

On trouvera en annexe la liste des agglomérations de la base de sondage et les probabilités de tirage associées.

<sup>41</sup> Fichier des établissements sanitaires et sociaux

<sup>42</sup> L'écart entre les probabilités de tirage que l'on souhaitait obtenir et celles obtenues sera compensé au deuxième degré de tirage.

### **Encadré : Les informations initiales sur les prestations servies dans les agglomérations : deux fichiers partiels d'hébergement**

Il n'existe pas de base exhaustive de tous les distributeurs de prestations. Deux fichiers (Finess et Fnars) plus ou moins complets sur les capacités d'hébergement pouvaient être cependant mobilisés : l'hébergement conventionné pour Finess, et le fichier Fnars couvrant un champ un peu plus large (dont du hors champ ex : foyers de jeunes travailleurs). La confrontation de ces deux fichiers donne une différence de capacité de 1 à 2, ce qui semble raisonnable à ce stade d'évaluation grossière du nombre de sans-domicile par agglomération. Cependant dans les agglomérations de plus de 100 000 habitants, l'écart entre les deux sources va de 1 à 30 pour 10000 habitants, ce qui est assez important.

Cependant ces deux sources ne recourent ni l'intégralité des points d'hébergement, ni les points soupes, ni les centres d'accueil. En l'absence d'une collecte à un niveau départemental voire communal de l'information précise concernant la présence (et un ordre de grandeur de la taille) des structures de distributions de prestations, nous ne pouvons pas évaluer l'importance de la « partie manquante ».

On peut faire l'hypothèse que la partie manquante est proportionnelle à la partie officielle, en considérant que certaines des villes contiennent en effet beaucoup plus de sans-abri que d'autres (certaines municipalités ne sont pas très accueillantes pour les sans-abri, migration, histoire économique, etc.). Cependant, les agglomérations avec peu de capacité d'accueil ont une probabilité plus faible d'être sélectionnées : on ne pourra donc pas vérifier qu'elles ne contiennent pas des structures non répertoriées. Ce peut être le cas de grosses villes (ex: Clermont-Ferrand). En dessous de 80 000 habitants, certaines agglomérations n'ont pas de structures dans la base et ne seraient donc pas susceptibles d'être tirées, alors qu'il est probable qu'elles abritent quand même des sans-domicile.

Nous pouvons aussi faire l'hypothèse que la partie manquante et non officielle a tendance à compléter les lacunes des structures officielles, et donc que la taille de la population est la plus pertinente. Mais cela a pour inconvénient de ne pas tirer certaines villes où nous sommes sûrs, au travers des fichiers d'hébergement mentionnés, qu'il y a beaucoup de sans-abri, et donc de gaspiller l'information.

La réalité étant sans doute entre les deux, nous avons choisi d'établir une sorte de moyenne entre les deux critères de taille. Par exemple, de prendre la capacité d'accueil pour 10 000 habitants (qui varie de 0 à 200, avec une moyenne de 6), d'y ajouter un nombre arbitraire, 5 par exemple, puis de multiplier par la population. Ainsi, toutes les agglomérations auront une chance raisonnable d'être tirées, et celles où nous sommes sûrs d'y trouver une forte capacité seront sur-représentées. Les tests ultérieurs dans les quelques agglomérations ainsi tirées permettent éventuellement de refaire un tirage en modifiant ce nombre arbitraire en fonction de l'inventaire.

La répartition des structures d'hébergement, selon les fichiers Finess et Fnars, montre qu'elles sont davantage concentrées dans les centres administratifs (ex: préfecture), que dans des agglomérations proches d'autres agglomérations de taille supérieure. Aussi, il aurait été envisageable de calculer des poids non pas uniquement d'après la taille de l'agglomération, mais aussi d'après la taille de la zone d'attraction (bassin d'emploi, découpages administratifs).

La comparaison des données du fichier avec un répertoire papier partiel sur les lieux d'accueil de jour montre qu'à population égale, les lieux d'accueil de jour sont assez concentrés dans les agglomérations qui possèdent aussi des structures d'hébergement, mais qu'il en existe quand même quelques-uns dans des agglomérations absentes des deux fichiers d'hébergement. 20% des agglomérations sans structure d'hébergement ont quand même des centres d'accueil. Cela nous a conforté dans le choix de notre critère de taille.

#### 4.1.2.2 Unité secondaire

On distingue plusieurs niveaux de stratification :

- par type de service (selon la nomenclature utilisée à l'enquête téléphonique de recensement des structures) ;
- puis pour les services d'hébergement selon les caractéristiques démographiques du public accueilli (on distingue les services pour les hommes seuls et avec enfants, les services pour les femmes seules et avec enfants, et les services mixtes).

On se place maintenant à l'intérieur d'une strate particulière. Les tirages des couples (service x jour), aussi appelés visites, se font proportionnellement au nombre de prestations déclarées. L'enquête téléphonique a permis de déterminer le nombre de prestations servies en moyenne quotidiennement, déflaté par la probabilité d'inclusion de l'agglomération. Plus précisément, si  $\prod_i^*$  est la probabilité de tirage de l'agglomération, la

probabilité de tirage du couple (service  $x$  jour) sera proportionnelle à  $\frac{T_{i,j}}{\prod_i^*}$  ( $T_{i,j}$  étant l'effectif journalier

annoncé lors de l'inventaire des services, celui-ci peut donc être entaché d'erreurs). L'objectif de cette approche est de maintenir pour les prestations individuelles des probabilités d'inclusion finales proches les unes des autres.

On utilise une méthode de tirage à pas fixe. On distingue deux catégories de stratification, les stratifications implicites qui sont obtenues par simple tri du fichier et les stratifications explicites où le fichier est décomposé en plusieurs sous-fichiers. La stratification par le type de public des prestations est implicite. La stratification par type de service est implicite dans les petites agglomérations et explicite dans les grandes.

Il n'a pas été procédé à des tirages de plages horaires car l'information concernant les fréquentations horaires, obtenue à l'enquête téléphonique, s'avère de mauvaise qualité (même si ce degré de tirage aurait été souhaitable théoriquement, voir Ardilly, Leblanc, 2001). En outre, l'existence de plages horaires n'aurait pas facilité la collecte sur le terrain. En effet, les enquêteurs doivent être en mesure de déterminer l'heure d'arrivée d'une personne pour pouvoir appliquer les méthodes de tirage<sup>43</sup>. Des problèmes similaires se posaient pour les accueils de jour, ceci expliquant en partie qu'ils aient été exclus du champ de l'enquête. La méthode de collecte retenue finalement consiste plutôt à essayer de comptabiliser l'ensemble des personnes utilisant la prestation nuitée sans tenir compte de l'heure réelle d'arrivée, en envoyant les enquêteurs sur le terrain au moment le plus propice (de 19h à 21h pour les sites d'hébergement avec liste, de 21h à 23h pour les hébergements d'urgence, et exceptionnellement le matin).

Explicitons maintenant les probabilités de tirage. Ne connaissant pas au moment du tirage de l'échantillon de degré 2 les fréquentations des services, on a utilisé comme proxy les déclarations de fréquentation obtenus lors de l'enquête téléphonique.

Le deuxième degré de tirage étant stratifié par type de service, on peut simplement se préoccuper de ce qui se passe à l'intérieur de chaque strate  $s$ .

L'unité de tirage est le couple (service, jour). On impose deux contraintes :

1. La probabilité de tirage  $\prod_{i|j}$  doit être proportionnelle au nombre de prestations servies dans un service un jour donné :

$$\prod_{i|j} = s_{i,j} \frac{T_{i,j}}{\sum_{j \in s} T_{i,j}}$$

où :

- $j$  est un service de type  $s$  ;
- $T_{i,j}$  est le nombre de prestations individuelles journalières dans le service  $j$  de l'agglomération  $i$  ;
- $\prod_{i|j}$  est la probabilité de tirer le couple  $j$  dans l'unité primaire  $i$  ;
- $s_{i,j}$  est le nombre de couples (service, jour) tirés dans l'unité primaire  $i$  et le type de service  $s$ .

2. La probabilité de tirage  $\prod_{i|j}$  doit être construite de telle sorte que les prestations individuelles aient toutes la même probabilité finale d'être tirées. Pour réaliser cette condition, nous verrons qu'il est suffisant d'avoir l'équation supplémentaire suivante :

$$\prod_{i|j} = \alpha_s \frac{T_{i,j}}{\prod_i^*}$$

<sup>43</sup> Ce qui aurait pu s'avérer difficile dans la mesure où il n'y a en général pas de lieu de passage unique.

c'est-à-dire que la probabilité de tirage de deuxième degré doit être proportionnelle à la taille du service déflatée par la probabilité de tirage de l'agglomération ( $\alpha_s$  est un coefficient de proportionnalité - dépendant du type de service - que nous déterminerons ultérieurement). Le facteur de déflation augmente le nombre de couples (service, jour) à enquêter dans les moyennes et petites agglomérations et donc augmente la probabilité de tirage des prestations échantillonnées dans celles-ci. Ainsi au final, on aura théoriquement les mêmes probabilités de tirage pour toutes les prestations, même si les agglomérations ont été tirées avec des probabilités différentes.

A l'aide des deux conditions ci dessus, on peut calculer  $s_{i,j}$  (la nombre de services de type  $s$  à tirer dans l'agglomération  $i$ ), on a :

$$s_{i,j} = \alpha_s \frac{\sum_{j \in s} T_{i,j}}{\Pi_i^*}$$

Le nombre de couples tirés dans l'unité primaire  $i$  dépend donc à la fois du nombre de prestations, du type de service  $s$ , de l'agglomération  $i$  et du poids attribué à cette agglomération.

**Remarque** : si, lors du tirage des unités primaires, la fréquentation par service n'était pas connue, ce n'est plus le cas lors du deuxième degré de tirage (un proxy du nombre de prestations individuelles par service ayant été obtenu lors de l'inventaire téléphonique).

#### 4.1.2.3 Unité tertiaire

On tire alors quatre prestations individuelles dans chaque unité secondaire tirée à l'exception des prestations en chambre d'hôtel où on se contente de deux prestations individuelles. En effet les effectifs annoncés pour les chambres d'hôtel sont souvent peu élevés et ne permettent pas en général de collecter quatre questionnaires. L'objectif est de réaliser un sondage aléatoire simple. La méthode de tirage sur le terrain dépend de l'existence ou non de listes des personnes fréquentant le service (voir partie II).

Si  $\Pi_{k|i,j}$  est la probabilité de tirer la prestation  $k$ , on a :

$$\Pi_{k|i,j} = \frac{n_s}{T_{i,j}}$$

où  $n_s$  est le nombre de prestations individuelles tirées par service ( $n_s = 2$  ou  $n_s = 4$ ).

Les structures sont donc échantillonnées avec une probabilité correspondant à leur taille. Or le nombre réel d'individus est souvent différent de la capacité annoncée. Dans ce cas, on peut envisager deux solutions :

\* solution théorique ; on tire les individus en utilisant le même taux de sondage que celui fixé a priori. Le nombre d'individus sélectionné est donc variable ( et peut être différent de 4).

avantages : - chaque individu a la même probabilité d'être sélectionné quelle que soit le structure, donc les poids sont concentrés.

- le tirage des individus est plus facile, car les enquêteurs ont pour consigne d'en prendre 1 sur n (n fixé avant d'aller sur le terrain).

- le tirage peut être fait au fur et à mesure des arrivées, sans connaître le nombre total final d'individus, chaque individu ayant la même chance d'être sélectionné qu'il arrive au début ou à la fin de la tranche temporelle échantillonnée.

\* solution pratique ; on sélectionne un nombre d'individus fixé à l'avance. Le taux de sondage est donc variable selon ce qu'on trouve sur le terrain.

avantages : - la taille de l'échantillon est fixe. Si toutes les structures sont en sous-effectif par rapport à ce qu'on attend, la solution théorique conduirait à une enquête de taille réduite

- cela permet de s'assurer que chaque enquêteur aura une charge de travail constante. Dans la solution théorique, des enquêteurs peuvent se retrouver s'être déplacés pour rien, ou au contraire être en sous effectif et ne pouvoir interroger toutes les personnes échantillonnées.

- l'organisation de l'enquête rend difficile un nombre variable de questionnaires par enquêteur. Pendant que les enquêteurs passent leur questionnaire, ils ne peuvent pas échantillonner les individus.

La solution pratique a donc été adoptée .

#### 4.1.2.4 Synthèse

Calculons maintenant la probabilité finale de tirage d'une prestation individuelle. On a par la formule des probabilités conditionnelles :

$$P_k = \prod_{k|i,j} \prod_{j|i} \Pi_i^*$$

soit :

$$P_k = \frac{n_s}{T_{i,j}} s_{i,s} \frac{T_{i,j}}{\sum_{j \in s} T_{i,s}} \Pi_i^*$$

On voit alors que si l'on souhaite obtenir des probabilités égales à  $P$  pour toutes les prestations individuelles, il est nécessaire de choisir pour les  $s_{i,s}$  :

$$s_{i,s} = \frac{P \sum_{j \in s} T_{i,s}}{n_s \Pi_i^*}$$

On a alors :

$$\alpha_s = \frac{P}{n_s}$$

Si on note  $E$  le nombre total de questionnaires à réaliser durant l'enquête, fixé de manière exogène, on a alors l'équation supplémentaire :

$$E = \sum_i \sum_s n_s s_{i,s}$$

soit :

$$\begin{aligned} E &= \sum_i \sum_s n_s \frac{P \sum_{j \in s} T_{i,j}}{n_s \Pi_i^*} \\ &= P \sum_i \frac{T_i}{\Pi_i^*} \end{aligned}$$

**Remarque** : si les poids attribués aux unités primaires avaient pu être les  $\Pi_i$ , on aurait eu la formule :

$$\begin{aligned} E &= P \sum_i \frac{T_i}{a \frac{T_i}{T}} \\ &= PT \end{aligned}$$

c'est-à-dire une formule équivalente à un tirage aléatoire simple à 1 degré.

Par analogie notons  $\tilde{T} = \sum_i \frac{T_i}{\Pi_i^*}$

En résumé, on a alors :

- $\Pi_i^*$  la probabilité de tirer une agglomération donnée,  $a = 80$  le nombre d'agglomération tirées
- $\Pi_{j|i}$  la probabilité de tirer un couple (service, jour) (de type de service fixé) qui est donnée par :

$$\Pi_{j|i} = \frac{E T_{i,j}}{n_s \tilde{T} \Pi_i^*}$$

- le nombre de couples (service, jour) par unité primaire est alors donné par :

$$\begin{aligned} s_i &= \sum_s s_{i,s} \\ &= \frac{E}{\Pi_i^* \tilde{T}} \sum_s \frac{\sum_{j \in s} T_{i,j}}{n_s} \end{aligned}$$

- jusqu'à maintenant, on a supposé que le proxy du nombre de prestations servies donné par l'enquête téléphonique était correct. En fait, la probabilité de tirage conditionnelle des prestations individuelles n'est pas tout à fait donnée par  $\Pi_{k|i,j} = \frac{n_s}{T_{i,j}}$ , il faut prendre en compte le fait que lors

de la visite, l'échantillonneur mesure le nombre réel de personnes fréquentant le service le jour sélectionné. La probabilité de tirage de troisième degré s'écrit donc :

$$\Pi_{k|i,j} = \frac{n_s}{\hat{T}_{i,j}}$$

avec  $\hat{T}_{i,j}$  qui représente le nombre de prestations individuelles effectivement servies le jour de la visite dans le service. La probabilité de tirage n'est alors plus la même pour toutes les prestations individuelles. Evidemment, il n'est pas possible d'anticiper les écarts (entre les estimations données par les responsables lors de l'enquête téléphonique et les résultats du dénombrement des enquêteurs échantillonneurs) dans la mesure où les effectifs réels ne sont connus que le jour de passage des questionnaires.

#### Aménagements :

1. Dans certaines agglomérations (des régions Ile-de-France et Nord Pas-de-Calais) les capacités de collecte étaient insuffisantes pour réaliser l'ensemble des questionnaires (calculé avec la méthode exposée précédemment). Dans ces agglomérations, le nombre de questionnaires à réaliser a été diminué arbitrairement pour correspondre à la taille du réseau d'enquêteurs (cf partie III).
2. Dans certaines agglomérations de petite taille, quand  $\Pi_i^*$  est faible, le nombre de questionnaires à réaliser peut être du même ordre de grandeur, voire supérieur au nombre de prestations individuelles servies chaque jour. Il est donc nécessaire pour ces agglomérations de limiter le nombre de questionnaires en fixant un seuil au taux de sondage journalier des prestations individuelles. Le seuil a été choisi à 20%, soit :

$$\frac{\sum_s n_s s_{i,s}}{T_i} < 0,2$$

Théoriquement, il aurait fallu découper chaque jour en unités temporelles précises et relativement fines pour éviter d'avoir un même individu utilisant deux prestations différentes dans l'unité temporelle considérée. Pendant cette unité temporelle, il faudrait contrôler l'ensemble des personnes qui ont été présentes dans le lieu considéré.

Dans certains cas, il peut être difficile de limiter notre champ à des plages horaires précises. La question peut se poser par exemple dans certaines structures d'hébergement où arrivent des individus par le Samu Social la nuit, alors qu'aucune enquête n'est faite à ces heures là. Théoriquement, une personne (interrogée à un autre moment)

ayant rejoint une structure d'hébergement à 3 heures du matin ne devrait pas compter comme un lien, car c'est hors champ. Dans la pratique, cela concerne peu de personnes, et semble difficile à identifier.

## 4.2 Base de sondage et échantillon

Une fois le tirage de premier degré réalisé, l'étape suivante consiste à réaliser une base de sondage des services dans les 80 agglomérations sélectionnées. On réalise ensuite la mise en œuvre effective du tirage puis des aménagements pour tenir compte du fait que dans certains services de grande taille les visites sont trop nombreuses. L'ensemble des programmes associés au tirage de l'échantillon sont disponibles sur CD-ROM. Ils ont été écrits en Visual Basic. Les tables étant de taille relativement modeste, les opérations ont été complètement réalisées sur micro-ordinateur. Le choix de l'écriture d'un programme de tirage sous Excel n'est pas nécessairement judicieux. Mais il a été en grande partie contraint par le fait que les opérations à réaliser sur les tables sont nombreuses avant la mise en œuvre effective du tirage (celle-ci sont délicates à réaliser avec le langage macro SAS et le langage de requêtes SQL, puisqu'il s'agit de décomposer en plusieurs lignes des enregistrements). Une grande partie du programme correspond à des opérations de manipulation des tables (par région, par agglomération, par type de service). La souplesse de manipulation d'Excel s'avère particulièrement utile dans les regroupements réalisés après le tirage « théorique ».

### 4.2.1 La base de sondage de degré 2

La base de sondage de degré 2 a été constituée à partir de l'enquête téléphonique auprès des responsables des services d'hébergement et de restauration (partie 2).

#### 4.2.1.1 Constitution des tables de services

Cette enquête fournissant des informations par structure et par service, il était nécessaire de constituer une base de sondage par type de service. Les types de service sont ceux définis dans la partie 2, c'est-à-dire que l'on distingue 4 types d'hébergement (hébergement dispersé en logement, hébergement en hôtel incluant aussi les places d'urgence dans les FJT, FTM et résidences sociales<sup>44</sup>, hébergement regroupé en urgence ou hébergement regroupé hors urgence) et 4 types de restauration (restauration le midi et le soir qui peut être fixe ou itinérante).

#### 4.2.1.2 Estimation des effectifs pendant la période d'enquête

Pour effectuer le tirage des couples (service, jour), il est nécessaire de connaître une estimation de la fréquentation des services pendant la période de l'enquête. Celle-ci a été fournie par l'enquête téléphonique. Dans le cas où les effectifs n'ont été renseignés que par tranche, il a été décidé de prendre la moyenne des bornes de la tranche comme estimateur<sup>45</sup>.

Pour estimer le nombre de prestations servies pendant le mois d'enquête, il a fallu tenir compte des jours d'ouverture des services (certaines structures en particulier de restauration ne sont ouvertes que très épisodiquement). Lors du tirage, seuls les jours d'ouverture déclarés des services pouvaient être échantillonnés.

#### 4.2.1.3 La stratification par agglomération

On stratifie le tirage en distinguant les quatre grandes agglomérations (Paris, Lyon, Marseille, Lille), les agglomérations de taille moyenne et les petites agglomérations. Dans les grandes agglomérations, on stratifie explicitement par type de services ; dans les moyennes, cette stratification n'est qu'implicite ; les petites agglomérations sont, elles, regroupées dans une catégorie unique. L'objectif de ces regroupements est d'éviter d'avoir des strates de trop petite taille.

### 4.2.2 Echantillon de degré 2

#### 4.2.2.1 La semaine de réserve

Pour tenir compte d'éventuelles disparitions de services ou d'incidents de collecte, le tirage n'a pas été effectué sur 4 semaines mais sur 5, cette semaine complémentaire servant à réaliser dans les agglomérations où cela s'avérerait nécessaire des questionnaires supplémentaires (voir la section 5 de cette partie). Au final, le tirage a donc été réalisé sur 5 semaines avec un total de 5500 questionnaires. Dans un premier temps, seules les 4 premières semaines étaient communiquées aux enquêteurs. Le déclenchement d'éventuelles visites complémentaires était à la discrétion des concepteurs.

<sup>44</sup> Par la suite par souci de simplicité, on écrira simplement hébergement en chambre d'hôtel.

<sup>45</sup> Le cas de la tranche supérieure (plus de 350 personnes) ne s'est pas présenté.



#### 4.2.2.2 Le tirage de l'échantillon

Pour chaque strate, un fichier de couples (service, jour) est constitué en précisant l'effectif déclaré lors de l'enquête téléphonique. On calcule alors l'effectif déflaté par la probabilité de tirage de l'agglomération. Les services sont alors échantillonnés proportionnellement à ce coefficient.

Le fichier de tirage se présente donc de la façon suivante :

DATE	NOMI	QIDENT	EFFECTIFS	PROBAAGGLO	Ti/Pi
15/01/2002	ENTRAIDE ALBIGEOISE CCAS D'ALBI	2975M	30	0,259273094	115,708111
15/01/2002	ENTRAIDE ALBIGEOISE	2977M	39	0,259273094	150,420545
...	...	...	...	...	...
16/01/2002	ENTRAIDE ALBIGEOISE CCAS D'ALBI	2975M	30	0,259273094	115,708111
16/01/2002	ENTRAIDE ALBIGEOISE	2977M	39	0,259273094	150,420545
...	...	...	...	...	...
...	...	...	...	...	...
19/02/2002	ENTRAIDE ALBIGEOISE CCAS D'ALBI	2975M	30	0,259273094	115,708111
19/02/2002	ENTRAIDE ALBIGEOISE	2977M	39	0,259273094	150,420545

Le tableau ainsi constitué ne tient compte que des jours d'ouverture déclarés dans l'enquête téléphonique, et ceci pour éviter l'échantillonnage de jours où les services seraient éventuellement fermés.

**Remarque** : Les différents lieux de distribution d'une même restauration itinérante sont considérés comme des services à part entière.

On utilise pour l'échantillonnage la méthode du tirage systématique, qui se décompose de la façon suivante :

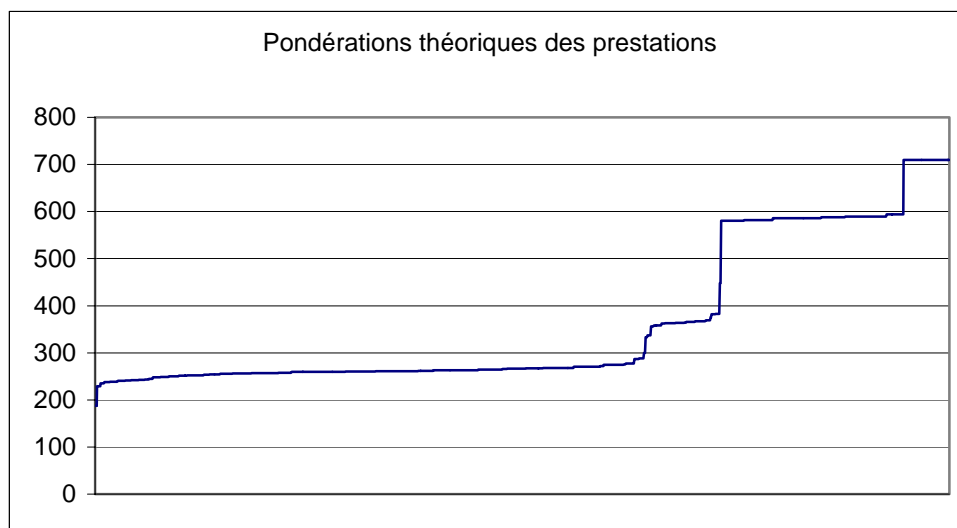
1. on effectue un tri aléatoire des enregistrements du fichier en tirant pour chaque ligne un nombre aléatoire et en triant le fichier suivant ce nombre ;
2. on tire un nombre aléatoire qui détermine la première visite échantillonnée ;

3. on applique le pas de tirage calculé à partir du total des  $\frac{T_{i,j}}{\prod_i^*}$  sur les 5 semaines et du nombre de visites à réaliser dans chaque strate<sup>46</sup>.

Les procédures de tirage ont permis de déterminer le nombre de couples (service, jour) à échantillonner dans chacune des strates. Ces nombres théoriques ont été arrondis au nombre entier le plus proche. Dans certaines directions régionales (Ile-de-France et Nord Pas-de-Calais), les capacités de collecte étaient insuffisantes pour réaliser l'ensemble des questionnaires. Une partie des visites prévues dans ces régions ont donc été reportées dans les autres strates.

On a, au final, échantillonné 1450 visites sur 5 semaines. Les pondérations théoriques (sans tenir compte de l'effectif réel mesuré sur le terrain) des prestations individuelles ne sont pas rigoureusement égales en raison de toutes les adaptations. Le graphique ci-dessous donne les pondérations théoriques des prestations dans l'ordre croissant.

<sup>46</sup> Pour plus de détail, se reporter à la partie théorique 4.1.



On observe un premier palier à 700 correspondant aux petites agglomérations (pour éviter un effet de saturation, elles ont été largement sous-échantillonnées) ; le deuxième palier aux environs de 600 correspond à la région Ile-de-France. A 350 environ, on a la région Nord Pas-de-Calais et pour finir aux environs de 250 les autres agglomérations de taille supérieure à 20 000 habitants. Les effets d'arrondis expliquent que les pondérations des prestations ne soient pas rigoureusement égales sur chacun des paliers.

Si l'on se restreint au 4 premières semaines, on a un total de 1157 visites échantillonnées (dans 867 services différents soit 701 structures), soit 4380 questionnaires initialement prévus.

#### 4.2.2.3 Pas de stratification selon les jours de la semaine

Le nombre et les caractéristiques des personnes fréquentant les services d'hébergement et de restauration étant susceptibles de varier au cours de la semaine, il pourrait sembler légitime de stratifier le tirage suivant les jours de la semaine. En fait, cette stratification implicite peut introduire des périodicités dans le fichier, et conduire à des résultats déroutants (cela s'est produit lors de la première tentative de tirage). Il peut en effet arriver que, dans certaines agglomérations où le pas de tirage est un diviseur de l'effectif total des services de l'agglomération, les mêmes services soient visités de nombreuses fois des jours différents. L'échantillonnage ainsi réalisé n'est pas biaisé, mais la variance risque d'augmenter considérablement. Compte tenu de cette difficulté, le tirage n'a pas été stratifié suivant les jours de la semaine.

#### 4.2.2.4 Regroupements de visites

Si la plupart des services sont échantillonnés une ou deux fois, quelques-uns le sont plus de 4 fois. Ainsi le tirage contraignait à se rendre 14 fois (9 fois le midi et 5 fois le soir) au Grand Ramier, une structure toulousaine qui sert plus de 600 repas par jour.

Répartition des services et structures en fonction du nombre de visites

Nombre de visites par service ou structure	Nombre de services concernés	Nombre de structures concernées
1	690	460
2	118	146
3	31	46
4	14	19
5	9	14
6	1	9
7	2	3
8	1	0
9	1	0
10	0	2
11	0	0
12	0	1
13	0	0
14	0	1

Source : enquête auprès des usagers des services d'hébergement ou de distribution de repas chauds, Insee, 2001

Evidemment, sur le terrain, il semblait délicat de venir un trop grand nombre de fois dans une même structure et a fortiori dans un même service. On craignait en effet, d'une part la lassitude de certains usagers et à l'opposé un nombre trop important d'enquêtés volontaires motivés par l'existence du « cadeau de remerciement ». A terme, une présence trop régulière des enquêteurs pouvait conduire à des perturbations dans le fonctionnement du service.

Il a donc été décidé de réaliser des regroupements de visites sous les deux contraintes suivantes :

1. un maximum de trois visites par service
2. un maximum de cinq visites par structure

Le nombre d'enquêteurs à mobiliser lors de certaines visites était trop important en région Ile-de-France, ces règles n'ont donc été que partiellement appliquées.

Répartition des services et structures en fonction du nombre de visites après regroupement

Nombre de visites après regroupements	Nombre de services concernés	Nombre de structures concernées
1	709	458
2	119	148
3	32	67
4	1	16
5	3	6
6	0	1
7	1	2

Source : enquête auprès des usagers des services d'hébergement ou de distribution de repas chauds, Insee, 2001.

D'autres modifications plus marginales ont été introduites :

- dédoublement pour quelques services sur deux jours de visites dans de petites agglomérations (les enquêteurs souhaitant quelquefois étaler les visites pour limiter la non-réponse) ;
- limitation du nombre de questionnaires dans les très gros services ;
- élimination des deux visites à réaliser dans la ville de Eu (une dizaine d'agglomérations différentes ayant été échantillonnée en Bretagne ou Basse-Normandie le réseau d'enquêteurs devait déjà réaliser de très nombreux déplacements. Pour ne pas trop augmenter la charge de travail et les temps de déplacement, à la demande de la direction régionale de la Bretagne, il a été décidé de retirer la ville de Eu de l'échantillon).

Au final, après ces modifications, il restait 4336 questionnaires à réaliser.

Répartition des visites selon le nombre de questionnaires à réaliser

Nombre de questionnaires par visite	Nombre de visites
2	116
4	891
8	52
12	9
16	1

Source : enquête auprès des usagers des services d'hébergement ou de distribution de repas chauds, Insee, 2001.

Nombre de visites et de questionnaires par type d'échantillon

	Echantillon initial (y compris réserve)			Echantillon final		
	Echantillon principal (4 semaines)	Echantillon de réserve (5 <sup>ème</sup> semaine)	Ensemble	Echantillon principal (4 semaines) après regroupement de visites et aménagements	Echantillon complémentaire tiré dans la réserve (5 <sup>ème</sup> semaine)	Ensemble
Nombre de visites	1157	293	1450	1069	78	1147
Nombre de questionnaires	4380	1108	5488	4336	296	4622

Source : enquête auprès des usagers des services d'hébergement ou de distribution de repas chauds, Insee, 2001.

### 4.3 Echantillonnage de degré 3

Les enquêteurs échantillonneurs étaient chargés de réaliser, lors des visites des services, l'échantillonnage de troisième degré, c'est-à-dire le tirage des personnes utilisant le service le jour échantillonné. Les méthodes de tirage ont déjà été exposées dans la partie 3. Que le tirage s'effectue sur liste ou sans liste, il est nécessaire d'avoir une table de tirage.

Les tables de tirage sont constituées de telle sorte que le caractère aléatoire de l'échantillonnage soit conservé sachant les conditions de collecte difficiles. Les contraintes étaient les suivantes :

1. Tirage de 2, 4, 8, 12 ou 16 questionnaires suivant la visite. Une table de tirage a été réalisée pour chacun de ces cas.
2. Dans le cas des services sans liste, pour éviter que les personnes sélectionnées n'aient à patienter trop longtemps (en attendant les enquêteurs interviewers déjà occupés avec une personne précédemment échantillonnée), le passage des questionnaires doit se répartir sur l'ensemble de la durée du service. Pour que les enquêteurs puissent plus aisément gérer les contacts avec les enquêtés, les tables de tirage ont été construites de sorte que la moitié des nombres de la table soient plus petits que  $N/2$  et la moitié plus grand ( $N$  désigne le nombre de prestations servies).
3. En cas de refus, la table de tirage doit permettre les remplacements et donc éviter un engorgement des tâches à réaliser par l'échantillonneur (cette situation risquait d'être particulièrement fréquente dans les services de restauration itinérante, où l'enquêteur échantillonneur avait de nombreuses missions à remplir, voir partie 2). Les numéros des tables de tirage ont donc été choisis de sorte que ceux-ci soient espacés afin de pouvoir réaliser au moins un remplacement (le pas de remplacement dépend du nombre de prestations servies lors de la visite, voir partie 2).
4. Quand les enquêteurs ont à réaliser plusieurs visites successives dans un même service, dans le cas des services avec liste et lorsque la liste n'a pas évolué entre les visites, avec une unique table de tirage, les enquêteurs auraient été conduits à réinterroger les mêmes personnes d'une visite à une autre<sup>47</sup>. Même si la liste a légèrement changé, le risque de corrélation entre les tirages est important. Trois tables de tirage ont donc été construites, pour que l'enquêteur échantillonneur puisse utiliser des tables différentes en cas de multivisites dans un même service.
5. En absence de liste, les effectifs devaient être estimés avant la passation des questionnaires (en particulier dans les services de restauration), pour éviter de perdre le dernier questionnaire, on anticipait une légère sur-estimation des effectifs. Quand l'enquêteur prévoyait par exemple 100 prestations servies, le tirage s'effectuait sur une base de 95 (c'est-à-dire une marge de sécurité de 5%). Cette marge a été choisie relativement faible, de sorte à limiter un éventuel biais dû au fait que les personnes en fin de file d'attente pourraient présenter des caractéristiques différentes des autres.

Les tables de tirage ont été réalisées à partir d'un programme écrit en Visual Basic (macro Excel). La liste des nombres de 1 à  $N$  est séparée en deux moitiés où l'on tire aléatoirement un nombre et on écarte tous les nombres interdits (les nombres sélectionnés et les nombres qui pourraient éventuellement servir pour des remplacements). Les tables de tirage sont alors constituées en itérant le processus.

---

<sup>47</sup> Dans le cas où la liste n'a pas du tout évolué (cas fréquent des services de longue durée), les enquêteurs auraient été amenés à toujours enquêter les quatre mêmes personnes lors de chacune des visites.

Le tableau suivant permet d'illustrer cette technique dans le cas de la constitution d'une table de tirage de 4 individus parmi 21. La table de tirage est finalement constituée des nombres 1, 5, 11 et 19

Liste des nombres	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
nombre tiré 5	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
nombre tiré 1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
nombre tiré 11	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
nombre tiré 19	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

Pour vérifier que l'impact des contraintes est relativement faible, on a procédé à une simulation de Monte-Carlo : 20 tirages de quatre prestations pour chacune des tables de tirage avec N compris entre 1 à 50 (pour les tables de taille inférieure à 4, tous les nombres disponibles sont sélectionnés). Le graphique ci-dessous (Comparaison des fréquences simulées et théoriques) donne les fréquences d'apparition des différents nombres lors de ces tirages. Ces fréquences sont comparées aux fréquences théoriques que l'on aurait obtenues si l'on avait réalisé un vrai tirage aléatoire sans remise. Plus précisément, dans un tirage aléatoire sans remise de 4 éléments parmi N, la probabilité d'inclusion vaut 4/N.

On définit alors la variable aléatoire  $X_N^j$  : nombre de fois qu'un élément j apparaît lors de 20 tirages successifs de 4 éléments parmi N. Cette variable n'est définie que pour N supérieur à j et suit une loi Binomiale :

$$X_N^j \rightarrow B\left(20, \frac{4}{N}\right)$$

On définit alors la variable aléatoire :

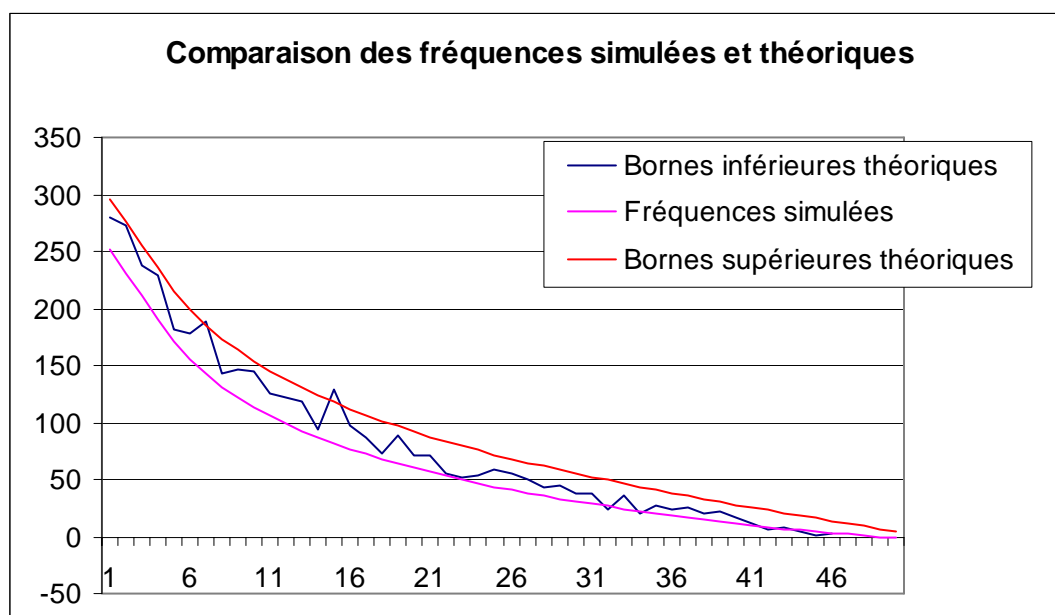
$$X^j = \sum_{i=j}^N X_i^j$$

La moyenne et la variance de  $X^j$  sont alors données par :

$$E(X^j) = 80 \sum_{i=j}^N \frac{1}{i}$$

$$\text{var}(X^j) = 20 \sum_{i=j}^N \frac{4}{i} \left(1 - \frac{4}{i}\right)$$

On observe dans le graphique ci-dessous que dans les tables de tirage, les fréquences d'apparition des nombres sont compatibles avec les fréquences que l'on aurait obtenues avec un vrai sondage aléatoire simple sans remise.



**Remarque** : pour tenir compte du fait que dans certains services (services sans liste), le nombre de prestations servies pouvait être sous-estimé lors de la pré-visite, les enquêteurs avaient la possibilité de réaliser jusqu'à deux questionnaires supplémentaires. Cette option offrait deux avantages :

- Limiter la dispersion des pondérations (voir section 6 de cette partie)
- Eviter un éventuel biais qui proviendrait de différences entre les individus utilisant les prestations servies au début et à la fin.

Cette option a été peu utilisée par les enquêteurs échantillonneurs, qui dans de nombreuses situations avaient déjà des difficultés à réaliser les questionnaires demandés.

#### 4.4 La méthode du partage des poids

La méthode du partage des poids est un élément essentiel de cette enquête. Elle permet en effet de passer d'estimateurs sur les prestations (obtenus à partir des pondérations du plan de sondage) à des estimateurs sur les individus utilisateurs des services qui offrent ces prestations. Pour une description détaillée de la méthode, on peut se référer aux articles suivants : Ardilly, Leblanc, 2001, Echantillonnage et pondération d'une enquête auprès de personnes sans-domicile : un exemple français, *Techniques d'enquête*, vol. 37, n°1, pp. 117-127 ; Lavallée, 1995, Pondération transversale des enquêtes longitudinales menées auprès des individus et des ménages à l'aide de la méthode du partage des poids, *Techniques d'enquête*, vol. 21, pp. 27-35.

##### 4.4.1 Présentation théorique

Considérons  $U$  une population de prestations et  $V$  une population d'individus. Nous supposons qu'il existe une surjection<sup>48</sup>  $\psi$  de  $U$  dans  $V$ , c'est-à-dire que chaque individu de la population  $V$  a au moins un antécédent (il a utilisé au moins une des prestations de la population  $U$ ). L'objectif est d'obtenir l'estimateur d'un total sur  $V$  à partir d'un échantillon tiré sur  $U$ .

On s'intéresse au total d'une variable d'intérêt  $Y$  :

$$Y = \sum_{k \in V} y_k$$

<sup>48</sup> La difficulté vient du fait que l'application n'est pas une bijection.

On considère le nombre de prestations de U utilisées par chaque individu k de la population V :

$$r_k = \sum_{i \in U} r_{i,k}$$

avec  $r_{i,k} = 1$  si l'individu k a utilisé la prestation i, 0 sinon.

On a alors l'égalité :

$$\sum_{i \in U} \frac{r_{i,k}}{r_k} y_k = y_k$$

d'où :

$$\begin{aligned} Y &= \sum_{k \in V} y_k \\ &= \sum_{i \in U} \sum_{k \in V} \frac{r_{i,k}}{r_k} y_k \end{aligned}$$

En posant :

$$z_i = \sum_{k \in V} \frac{r_{i,k}}{r_k} y_k$$

on en déduit les égalités :

$$Y = \sum_{k \in V} y_k = \sum_{i \in U} z_i$$

Par l'introduction d'une nouvelle variable d'intérêt, on est en mesure de réécrire une somme sur les individus en une somme sur les prestations.

Soit un échantillon  $(S_U, (w_i)_{i \in S_U})$  obtenu à partir de la population U (sans hypothèse sur la nature du plan de sondage) l'image par la surjection  $\psi$  de l'échantillon  $S_U$ , définissant un sous-ensemble de V, est notée :

$$S_V = \{k \in V \mid \exists i \in U; r_{i,k} = 1\}.$$

On suppose que pendant l'enquête, on est en mesure de collecter les  $\psi^{-1}(\psi(i))$  (ou au moins les  $r_{i,k}$ ) sur  $S_U$  (c'est-à-dire que l'on peut connaître pour chaque prestation échantillonnée, les autres prestations utilisées par la personne usagère de la prestation échantillonnée). L'estimateur du total sur les prestations est alors donné par Horvitz-Thomson :

$$\hat{Y} = \sum_{i \in S_U} w_i z_i$$

On peut alors réécrire cette somme comme une somme sur  $S_V$  :

$$\begin{aligned} \hat{Y} &= \sum_{S_U} w_i \sum_{k \in V} \frac{r_{i,k}}{r_k} y_k \\ &= \sum_{S_V} \tilde{w}_k y_k \end{aligned}$$

avec :

$$\tilde{w}_k = \frac{1}{r_k} \sum_{S_U} w_i r_{i,k}$$

#### 4.4.2 Les limites de la méthode du partage des poids

##### 4.4.2.1 Le problème de l'invariance par translation dans le temps

Le méthode du partage des poids suppose que l'on est en mesure de calculer le nombre de services fréquentés par chaque individu au cours de la période de collecte (ici les  $r_{i,k}$ ). Ce n'est généralement pas le cas, le semainier ne collecte les liens que sur la semaine précédant le jour d'enquête. On fait donc l'hypothèse que les liens mesurés sur la semaine passée sont un bon estimateur des liens de chaque semaine d'enquête. L'estimateur



donné par la méthode du partage des poids n'est sans biais que dans la mesure où chaque individu utilise le même nombre de prestations d'une semaine sur l'autre.

L'exemple suivant illustre ce qui peut se produire dans le cas où cette hypothèse n'est pas vérifiée.

Considérons un service d'hébergement où l'ensemble de la population change lors de la journée de jeudi. Plus précisément la population 1 est dans le centre depuis plusieurs semaines et est remplacé au cours de la journée de jeudi par une population nouvelle.

Les semainiers se présentent sous la forme suivante :

	Mardi	Mercredi	Jeudi	Vendredi	Samedi	Dimanche	Lundi	Mardi	Mercredi	Jeudi	Vendredi	Samedi	Dimanche	Lundi	Mardi	Mercredi	Jeudi	
<b>Population 1</b>	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
<b>Population 2</b>							■	■	■	■	■	■	■	■	■	■	■	■

La population de type 1 occupe toutes les places disponibles dans le service jusqu'au mercredi soir inclus où elle est complètement remplacée par la population de type 2. Nous allons appliquer la méthode du partage de poids en supposant que l'on a échantillonné la période allant du lundi au samedi (partie grisée du graphique). On suppose que le même nombre de prestations ont été échantillonnées chaque jour. On a aisément le nombre de liens, en fonction du jour d'interrogation :

	Lundi	Mardi	Mercredi	Jeudi	Vendredi	Samedi
Nbre de Liens	6	6	6	1	2	3

La formule du partage des poids nous donne alors comme estimateur de la population :

$$Y_{total} = \frac{n_p}{6} \frac{N_p}{n_p} \left( \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + 1 + \frac{1}{2} + \frac{1}{3} \right)$$

$$= 2,33 \frac{N_p}{6}$$

Où  $n_p$  est le nombre de prestations échantillonnées et  $N_p$  le nombre de prestations totales. Evidemment, l'estimation ainsi obtenue est biaisée, puisqu'on s'attend à obtenir :  $2 \frac{N_p}{6}$ .

**Remarque** : si l'on avait réussi à collecter les liens non pas sur la semaine précédente mais sur la semaine échantillonnée, chaque prestation aurait eu 3 liens, et on aurait retrouvé le résultat général que l'estimateur du partage des poids est non biaisé.

L'importance du biais sur l'estimateur dépend de la validité de l'hypothèse d'invariance par translation dans le temps d'une période de référence. C'est-à-dire que la période de temps échantillonnée est le reflet d'un comportement régulier d'utilisation des services. Plus la période de référence est longue, plus cette hypothèse semble fragile. Sur une période d'un mois par exemple, on peut facilement imaginer qu'une partie de la population est composée d'entrants (des personnes fréquentant nouvellement des services d'hébergement ou de restauration) et que donc, le nombre de leurs liens (déclarés sur le mois passé) sous-estime le nombre de liens sur le mois échantillonné. A l'opposé, on surestimera le nombre de liens de certains futurs sortants.

Pour corriger ce biais dans la pondération, on pourrait utiliser la méthode suivante :

- Estimation par un modèle de la distribution du temps passé dans chaque type de service (utilisation de données individuelles) ;
- Réalisation pour chaque prestation de l'échantillon, d'une simulation de Monte-Carlo du nombre de liens conditionnellement au temps déjà passé dans le service ;
- Application de la méthode du partage des poids (pour chaque simulation).

On peut en déduire la distribution d'un estimateur.

Le tableau ci-dessous donne pour les différents types de service d'hébergement la durée écoulée depuis l'arrivée de l'individu.

	Effectifs parmi les usagers hebdomadaires des services	% de personnes arrivées depuis moins d'une semaine	% de personnes arrivées depuis moins d'un mois
Hébergement dispersé	<b>13341</b>	<b>2,6%</b>	<b>13,9%</b>
Hébergement en hôtel, FJT	<b>3372</b>	<b>7,0%</b>	<b>29,2%</b>
Hébergement regroupé (urgence)	<b>8880</b>	<b>29,0%</b>	<b>54,1%</b>
Hébergement regroupé (hors urgence)	<b>18563</b>	<b>3,8%</b>	<b>15,6%</b>

Pour le calcul des pondérations avec comme période de référence le jour ou la semaine, ce biais n'a pas été pris en compte.

#### 4.4.2.2 Le problème de la corrélation entre les services échantillonnés et la variable estimée

Quand la variable d'intérêt est corrélée avec la fréquentation des services servant à l'échantillonnage, l'estimateur du partage des poids peut là encore être biaisé. Pour illustrer cela, on considère une population de personnes fréquentant des services de restauration ; l'objectif est d'estimer le nombre de personnes dormant dans la rue un soir moyen de la semaine. On suppose que la population se scinde en deux, et a le comportement donné par le schéma ci-dessous, c'est-à-dire que les personnes dorment dans la rue les jours où elles fréquentent les services de restauration (la question posée étant : « Où allez-vous dormir ce soir »).

	Lundi	Mardi	Mercredi	Jeudi	Vendredi	Samedi	Dimanche
Trajectoire (type 1)							
Restauration							
Rue							
Trajectoire (type 2)							
Restauration							
Rue							

Si l'on échantillonne les prestations (qui sont ici uniquement les repas servis dans les centres de restauration) et on applique la méthode du partage des poids à l'estimateur du nombre de personnes dormant dans la rue, on obtient :

$$5 * \left(\frac{1}{7} \frac{N}{n}\right) * \frac{1}{5} + 2 * \left(\frac{1}{7} \frac{N}{n}\right) * \frac{1}{2} = \frac{2}{7} \frac{N}{n}$$

c'est-à-dire que tous les utilisateurs des services de restauration dorment dans la rue.

Dans ce calcul, la variable « dormir dans la rue » vaut toujours 1, les personnes échantillonnées dormant tous les soirs dans la rue. En fait, on devrait obtenir que seulement  $5 * \left(\frac{1}{7} \frac{N}{n}\right) * \frac{1}{5} * \frac{5}{7} + 2 * \left(\frac{1}{7} \frac{N}{n}\right) * \frac{1}{2} * \frac{2}{7} = \frac{1}{7} \frac{N}{n}$  des utilisateurs

des distributions de repas dorment dans la rue. Par l'application de la méthode du partage des poids on surestime d'un facteur 2 le nombre de personnes dormant dans la rue.

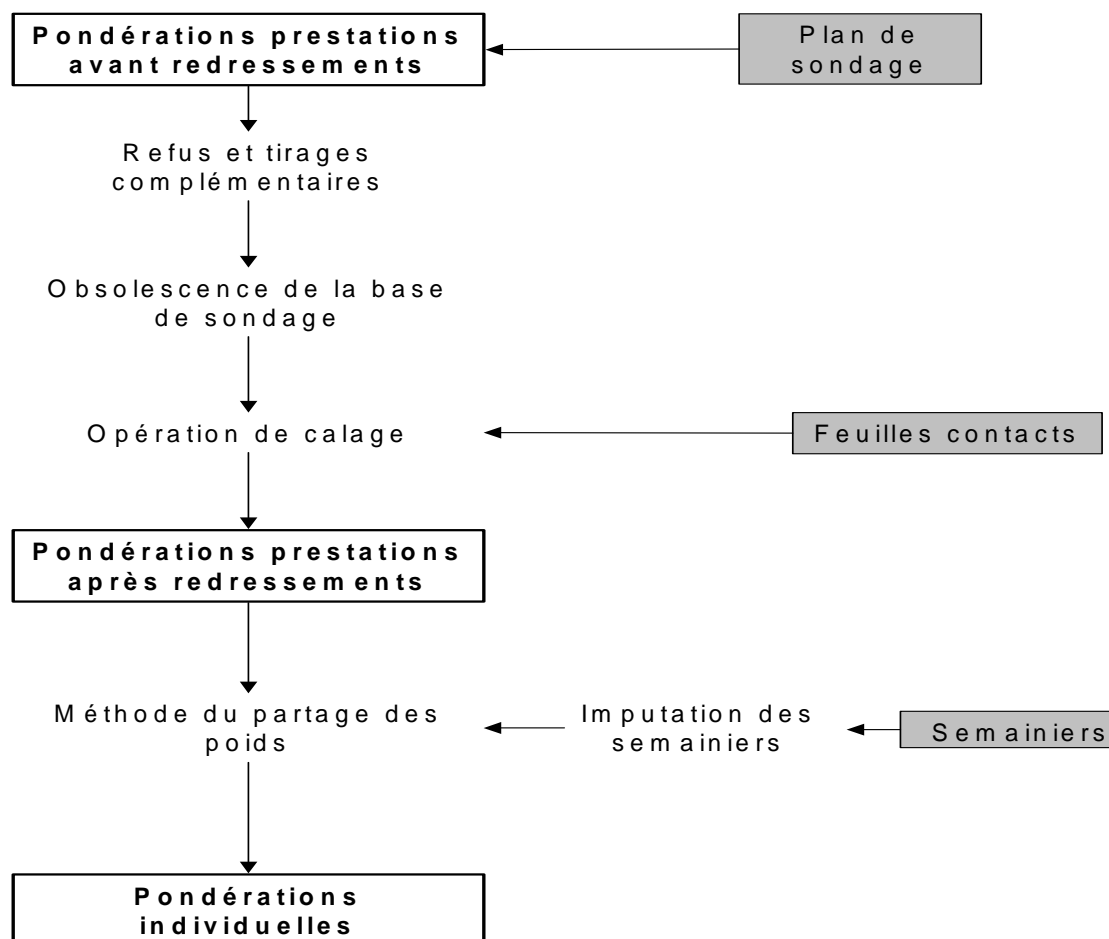
La corrélation entre la variable d'intérêt et la variable de fréquentation des services conduit donc à biaiser le résultat de l'estimation.

Pour éviter ce type de problème, on pourrait par exemple, ne pas regarder la situation le jour même ou la veille mais par exemple 7 jours plus tôt, de sorte que les variables ne soient plus que faiblement corrélées. Ce biais a été négligé dans le calcul des pondérations (la corrélation entre la fréquentation des services de restauration et la

ne n'étant pas évident). Il serait néanmoins souhaitable de regarder cette question plus précisément en particulier dans le cas des services d'hébergement.

## 4.5 Les opérations de redressement

Le calcul des pondérations individuelles nécessite de réaliser un certain nombre de traitements. On distingue dans la figure 1 les différentes opérations à réaliser sur les pondérations. Nous allons maintenant les détailler.



### 4.5.1 Tirages complémentaires et refus

Certains services ont refusé de participer à l'enquête et dans d'autres cas plus nombreux, aucun questionnaire n'a pu être réalisé. En outre, pour pallier la diminution du nombre de questionnaires finalement réalisés, il était prévu de réaliser dans les deux semaines suivant la période d'enquête initiale des visites complémentaires, l'objectif étant de limiter la dispersion des pondérations d'une agglomération à une autre.

#### 4.5.1.1 Les tirages complémentaires

Pour prendre en compte ces tirages complémentaires, on fait l'hypothèse dans le calcul des pondérations que les semaines 5 et 6 sont comparables aux 4 semaines initiales d'enquête. Ce que l'on traduit en réallouant aléatoirement à chaque visite complémentaire une des quatre semaines d'enquête (le jour de semaine étant supposé le même).

Les tirages complémentaires sont alors assimilés à un échantillonnage en deux temps. Chaque service pouvant être échantillonné soit directement, soit dans un second temps lors des tirages complémentaires. Il est alors

possible d'estimer les probabilités de tirage des visites. Pour simplifier, le calcul a été réalisé sous l'hypothèse de probabilités uniformes au deuxième tirage :

$$\begin{aligned}\pi_i' &= P(\text{tirage1}) + P(\text{tirage2}, \neg\text{tirage1}) \\ &= \pi_i + P(\text{tirage2} | \neg\text{tirage1})P(\neg\text{tirage1}) \\ &= \pi_i + \frac{n_2}{N - n_1}(1 - \pi_i)\end{aligned}$$

$N$  représente le nombre de visites dans la base de sondage,  $n_1$  et  $n_2$  respectivement le nombre de visites échantillonnées lors du premier et du second tirage,  $\pi_i$  la probabilité initiale de tirage et  $\pi_i'$  la probabilité obtenue après prise en compte des tirages complémentaires.

#### Calcul annexe :

Si l'on tient compte du fait que le sondage est à probabilités inégales, la probabilité d'inclusion s'écrit :

$$\pi_i' = \pi_i + (1 - \pi_i)E\left(\frac{\pi_i n_2}{\sum_{k \in U/s_1} \pi_k} \mid i \notin s_1\right)$$

Soit négligeant le biais, la formule approchée :

$$\begin{aligned}\pi_i' &\approx \pi_i + (1 - \pi_i) \frac{\pi_i n_2}{E\left(\sum_{k \in U/s_1} \pi_k \mid i \notin s_1\right)} \\ &\approx \pi_i + (1 - \pi_i) \frac{\pi_i n_2}{\sum_{\substack{k \in U/s_1 \\ k \neq i}} \pi_k (1 - \pi_k) + \pi_i} \\ &\approx \pi_i \left(1 + \frac{n_2}{\sum_{k \in U/s_1} \pi_k}\right)\end{aligned}$$

Dans cette dernière égalité les termes d'ordre deux sont négligés.

#### 4.5.1.2 Les refus

Un petit nombre de structures ont refusé de participer à l'enquête alors que les services associés avaient été échantillonnés (voir le paragraphe 3.1). On parle alors de refus de visite. En outre, dans plusieurs services, aucun questionnaire n'a pu être réalisé, par exemple parce que l'ensemble des usagers échantillonnés refusaient de répondre aux questionnaires. Par souci de simplification, on a assimilé les services où aucun questionnaire n'a été réalisé à des refus de visite. Le nombre de refus de visite s'élève alors à 28. Pour tenir compte de ces refus, on fait l'hypothèse (probablement abusive, mais acceptable dans la mesure où le niveau de non-réponse est très faible) que les services ayant refusé de participer sont comparables aux services participant à l'enquête. La non-réponse lors des visites est donc assimilée à un tirage en deux phases des visites, avec probabilités uniformes lors de la deuxième phase. Ce qui nous donne la pondération suivante après prise en compte de la non-réponse des services :

$$\pi_i'' = \pi_i' \left( 1 - \frac{n_{\text{refus}}}{n} \right)$$

où  $n_{\text{refus}}$  est le nombre de visites échantillonnées aboutissant à un refus et  $n$  le nombre de visites échantillonnées initialement.

#### 4.5.2 Correction de l'obsolescence<sup>49</sup> de la base de sondage

La base de sondage utilisée pour l'échantillonnage des visites a été achevée en octobre 2000 à partir de l'enquête téléphonique auprès des structures dans les 80 agglomérations échantillonnées au premier degré de tirage (voir paragraphe 2.1). Entre la période de passation des questionnaires en janvier et février 2001, et le tirage de l'échantillon des visites, il s'est donc écoulé une période d'environ 4 mois. Il est nécessaire dans le calcul des pondérations de tenir compte de cette obsolescence de la base de sondage. L'obsolescence de la base de sondage se traduit par :

- La disparition de services (2,7% des services échantillonnés n'existent plus<sup>50</sup>)
- De nouveaux services sont apparus (ou avaient été oubliés lors de la constitution de la base de sondage).

Il a en effet été possible de détecter à partir des semainiers d'éventuels oublis dans la base de sondage (certains individus fréquentant à la fois des services dans la base et d'autres nouvellement créés ou oubliés initialement). On estime à environ 5% le nombre de lacunes repérées à partir des semainiers. Evidemment ce taux est une borne inférieure car certains services oubliés ou nouvellement créés pouvant vivre en « autarcie », ils ne sont alors pas décelables par l'intermédiaire des semainiers.

Pour tenir compte des lacunes de la base de sondage, on effectue une post-stratification uniforme des services. N'ayant aucune information sur les effectifs des services repérés dans les semainiers (il a finalement été décidé de ne pas procéder à une étude complémentaire auprès de ces services pour ne pas retarder la publication des résultats), on fait l'hypothèse que les effectifs de ces services sont statistiquement comparables aux effectifs des services déjà présents dans la base de sondage. Pour tenir compte de la volonté de diffuser des résultats spécifiquement sur l'agglomération parisienne, on distingue les 6 strates suivantes :

Taux de lacunes par strate	Paris	Province
Hébergement dispersé	0%	0%
Hébergement regroupé	2%	3%
Restauration	16%	10%

On observe que le taux de lacunes est beaucoup plus important pour la distribution de repas chauds que pour l'hébergement.

L'hypothèse de post-stratification est probablement excessive, on peut en effet supposer qu'à l'intérieur de chaque strate, les services oubliés sont de taille plus modeste que la moyenne des autres services.

#### 4.5.3 Les opérations de calage des prestations

##### 4.5.3.1 Calage sur la structure par genre

On ne dispose pas a priori de données exogènes de calage. En effet, si l'on excepte les données du recensement (dont on ne peut assurer la totale fiabilité), il n'existe aucune source donnant une estimation du nombre total de sans-domicile.

On peut néanmoins réaliser de modestes opérations de calage en exploitant les informations contenues dans les feuilles contacts. Pour pallier des taux parfois élevés de non-réponse, il avait été décidé de remplacer les

<sup>49</sup> On a assimilé l'obsolescence et les erreurs de la base de sondage.

<sup>50</sup> Y compris les services qui se révèlent hors champ.

individus refusant de répondre à l'enquête (voir partie 3). Le profil des répondants pourraient a priori être différent de celui des personnes refusant de participer, il est souhaitable de recalculer les pondérations. Les informations contenues dans les feuilles contacts sont modestes (sexe de la personne ayant refusé de participer et une estimation de son âge).

On réalise alors le calage en deux étapes :

1. Estimation au sein des personnes échantillonnées (qu'elles aient ou non accepté de répondre à l'enquêteur) un taux d'hommes et de femmes ainsi qu'une répartition par structure d'âge.
2. Post-stratification uniforme pour les 6 strates (type de service et Paris versus province).

En fait les données portant sur l'âge de la personne ayant été assez mal renseignées, en particulier en restauration où les conditions de collecte étaient parfois délicates, le calage n'a été réalisé que pour la structure par genre. Vu les taux de non-réponse élevés observés dans les services de restauration, il serait peut-être souhaitable de parvenir à un calage sur la structure par âge dans une prochaine enquête.

Pour réaliser la première des deux étapes (l'estimation par service du taux d'hommes), on a utilisé un petit modèle de log-vraisemblance. On suppose que les hommes et les femmes ont des probabilités différentes d'accepter de participer à l'enquête, et que le rang du premier répondant suit une loi géométrique.

La vraisemblance du modèle s'écrit alors :

$$\prod_{j \in s} (pq)^{\sum_{i=1}^{n_j-1} x_i^{(j)}} ((1-p)r)^{n_j-1-\sum_{i=1}^{n_j-1} x_i^{(j)}} (p(1-q))^{x_{n_j}^{(j)}} ((1-p)(1-r))^{1-x_{n_j}^{(j)}}$$

En notant :

- $p$  la proportion d'hommes dans la population
- $q$  la probabilité d'acceptation pour les hommes
- $r$  la probabilité d'acceptation pour les femmes
- $n$  le nombre de tentatives pour aboutir à un questionnaire
- $x_i = 1$  pour un homme et  $x_i = 0$  pour une femme
- $j$  désignant le numéro du contact au sein du service

$$\hat{p} = \frac{\sum_{j \in s} \sum_{i=1}^{n_j} x_i^{(j)}}{\sum_{j \in s} n_j}$$

On a alors le maximum de vraisemblance pour le paramètre  $p$  qui est donné par :

#### 4.5.3.2 Estimation du nombre de non-francophones

Dans la mesure où les personnes fréquentant les services peuvent être francophones ou non, si l'on applique lors du calcul des pondérations le dénombrement effectué par les enquêteurs sur le terrain, on obtient un jeu de pondérations portant sur l'ensemble des prestations servies (ces pondérations supposent que le comportement des non-francophones est comparable au comportement des francophones). Pour obtenir des pondérations ne portant que sur la population francophone (champ de l'enquête initialement défini), il est nécessaire d'estimer pour chaque visite échantillonnée, un taux de personnes non-francophones ayant utilisé une prestation. A partir de ce taux, le nombre de prestations distribuées aux francophones peut être aisément estimé.

On dispose là encore pour estimer le nombre de non-francophones des feuilles contacts qui nous indiquent dans quel cas les refus sont consécutifs à une inaptitude à parler français. A partir de ces informations, on réalise un modèle de log-vraisemblance. Il s'agit d'un modèle assez proche de celui réalisé pour le calage sur le genre.

Voici, les taux de non-francophones obtenus pour quelques grandes agglomérations :

<b>NOM</b>	<b>% non-francophone usagers des services d'aide</b>	<b>Population RP99</b>
<i>Paris</i>	22%	9480707
<i>Lyon</i>	10%	1309596
<i>Marseille</i>	10%	1263562
<i>Lille</i>	6%	980503
<i>Toulouse</i>	9%	741120
<i>Bordeaux</i>	7%	735337
<i>Nantes</i>	7%	544932
<i>Nice</i>	29%	529241
<i>Strasbourg</i>	16%	410346

Source : enquête auprès des usagers des services d'hébergement ou de distribution de repas chauds, Insee, 2001. Recensement de la population, 1999.

Ces résultats sont à prendre avec précaution, en particulier dans les agglomérations de plus petites tailles, l'effet d'échantillonnage pouvant être important.

#### **4.5.4 Bilan des opérations de redressement**

Les opérations présentées ci-dessus ont été réalisées sur les pondérations prestations. Le tableau suivant résume les facteurs correctifs appliqués aux probabilités d'inclusion :

<b>Strates</b>	<b>Taux Lacunes (En %)</b>	<b>Calage Hommes (En %)</b>	<b>Calage Femmes (En %)</b>
Paris hébergement dispersé	1,00	1,04	0,95
Paris hébergement regroupé	0,98	0,95	1,09
Paris restauration	0,84	1,04	0,72
Province hébergement dispersé	1,00	1,04	0,96
Province hébergement regroupé	0,97	0,95	1,10
Province restauration	0,90	1,01	0,97

## **4.6 Le problème de la dispersion des pondérations**

La dispersion des pondérations représente une des difficultés majeures de l'enquête SD2001. Contrairement aux enquêtes ménages traditionnelles de l'INSEE<sup>51</sup>, les pondérations peuvent être très différentes entre deux individus.

### **4.6.1 Les facteurs de dispersion des pondérations**

Les principaux facteurs de dispersion des pondérations sont les suivants :

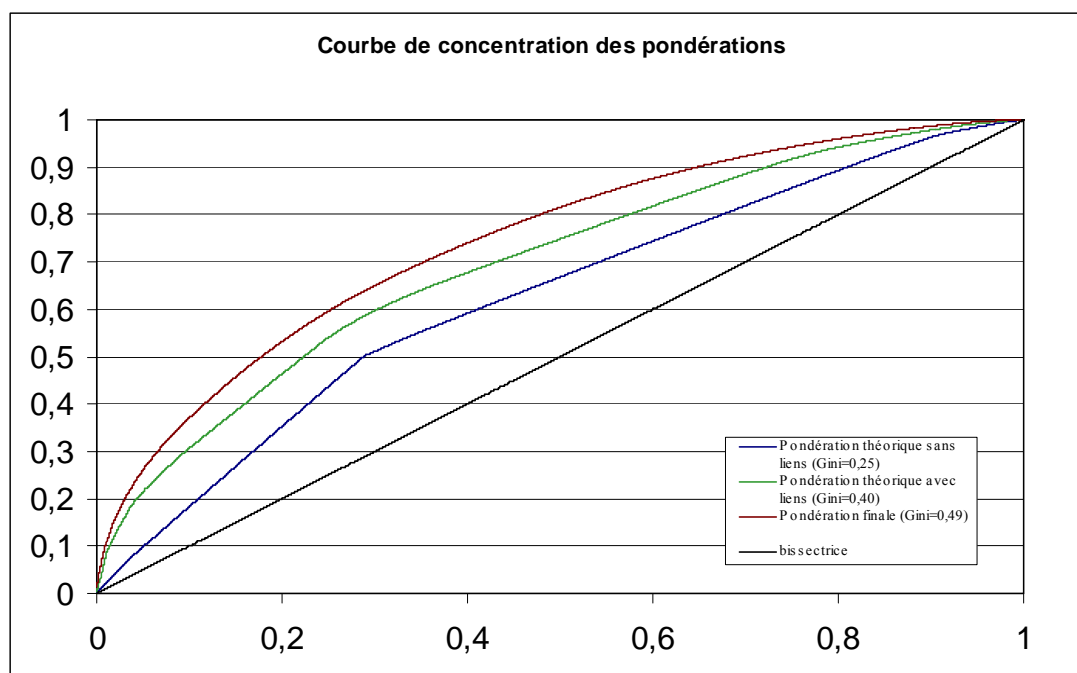
- Les prestations n'ont pas toutes la même probabilité théorique de tirage.
- Les fréquentations des services données par l'enquête téléphonique et utilisées pour le calcul théorique des pondérations peuvent être sensiblement différentes des fréquentations obtenues lors du

<sup>51</sup> Les procédures de calage dispersent en général faiblement les pondérations.

dénombrement des prestations. Pour une explication des facteurs expliquant les erreurs d'anticipation des effectifs, on peut se reporter à la partie 5.

- Les opérations de redressement conduisent dans certains cas à disperser les pondérations. L'impact reste cependant relativement faible
- La méthode du partage des poids pour passer des pondérations prestations aux pondérations individuelles peut avoir un impact considérable sur la dispersion des pondérations. Plus la période de référence est grande (le champ de l'enquête est constitué de l'ensemble des individus ayant fréquenté au moins une fois un service pendant la période de référence), plus les pondérations des individus se dispersent. Pour une pondération journalière, l'impact du partage des poids sur les pondérations est au maximum d'un facteur 3 (on suppose qu'une personne peut au maximum utiliser 3 prestations). Le facteur passe à 21 pour les pondérations hebdomadaires, et à 84 pour d'éventuelles pondérations mensuelles.

Le graphique ci-dessous donne les courbes de concentration des pondérations pour différents jeux de pondérations.



#### 4.6.2 Impact de la dispersion des pondérations

Des pondérations très dispersées détériorent la qualité de l'enquête, un petit nombre de questionnaires ayant un impact important sur les estimateurs. La variance d'un estimateur est d'autant plus importante que les pondérations sont dispersées<sup>52</sup>.

On se trouve donc devant le dilemme général suivant. D'une part pour étendre le champ de l'enquête (c'est-à-dire le champ des usagers des prestations), il est souhaitable de choisir un grand nombre de types de services et d'étendre le plus possible la période de référence. D'autre part plus le nombre de liens possibles est grand, plus les pondérations risquent d'être dispersées. Dans l'enquête française auprès des personnes fréquentant les services d'hébergement et de restauration, le champ des services est relativement réduit (on n'a pas cherché à inclure les accueils de jour, les maraudes, les services hospitaliers, ...) et la période de référence modeste (une semaine). Il faut toujours garder à l'esprit que l'extension du champ des usagers peut conduire, si les pondérations sont trop dispersées, à une enquête difficilement exploitable.

#### 4.6.3 Comment diminuer la dispersion des pondérations

Pour une éventuelle future enquête, il paraît souhaitable de tenter de limiter la dispersion des pondérations pour diminuer la variance des estimateurs. Plusieurs pistes peuvent être envisagées :

<sup>52</sup> sous réserve que la variable d'intérêt ne soit pas corrélée avec les pondérations.



1. tirage des agglomérations à partir des estimateurs de la population calculés sur l'enquête SD2001.
2. sur-échantillonnage des agglomérations et des types de service où les personnes sans-domicile ont peu de liens.
3. meilleure anticipation des écarts entre les fréquentations données par l'enquête téléphonique et la mesure sur le terrain.
4. rendre plus flexible le nombre de questionnaires à réaliser par visite, en corrélant ce nombre à la fréquentation le jour échantillonné.

Le tableau ci-dessous donne le nombre moyen de liens individuels par type de service (on a utilisé les pondérations hebdomadaires) :

Type de service	Population	Nombre moyen de liens
Hébergement dispersé	13341	5,10
Hébergement en chambre d'hôtel	3565	5,26
Restauration itinérante le midi	284	3,02
Restauration itinérante le soir	6470	3,44
Restauration fixe le midi	14993	4,09
Hébergement regroupé hors urgence	18567	5,11
Hébergement regroupé en urgence	9412	5,13
Restauration fixe le soir	4182	6,46
TOTAL	70814	4,82

On a de même le nombre moyen de liens pour les grandes agglomérations :

Agglomérations	Population	Nombre moyen de liens
Lyon	1929	5,30
Paris	25487	4,80
Marseille	3886	4,29
Lille	1991	5,74
Reste de la France	37521	4,82
TOTAL	70814	4,82

**Remarque** : Le nombre de liens par usager des services est d'environ 4,8 (sur 15 liens potentiels, cinq jours ouverts et 3 prestations par jour). Cela provient du fait que dans de très nombreuses structures, les services de restauration (le midi et le soir) avaient le même public que l'hébergement associé, et donc n'ont pas été échantillonnés.

**Encadré : le plan de sondage de l'enquête de l'INED auprès des personnes sans-domicile utilisatrices de services d'hébergement et de distribution de nourriture à Paris (13 février - 10 mars 1995).**

### 1. Champ de l'enquête

Utilisateurs des services d'hébergement et de distribution de nourriture à Paris intra-muros + 2 centres de banlieue parisienne où les sans-domicile sont conduits depuis Paris.

### 2. Caractéristiques de l'enquête

Trois types de services :

- les centres d'hébergement d'urgence (36) : prestation = nuitée ;

- les centres d'hébergement et de réadaptation sociale (CHRS) et les centres de longue durée hors CHRS (46) : prestation = nuitée ;
- les centres de distribution de repas gratuits et points-soupes (58) : prestation = repas (midi ou soir).

Durée de l'enquête = 4 semaines, 4 jours/semaine, soit 16 jours.

L'unité d'enquête est une prestation individuelle.

Objectif : enquêter 6 centres par jour, soit 96 centres × jours, avec 6 enquêtes par centre, soit environ 600 enquêtes :

- 225 en hébergements d'urgence ;
- 150 en hébergement de longue durée ;
- 225 en centres de distribution de repas.

### 3. Plan de sondage

Plan de sondage à deux degrés :

- unité primaire (UP) = centre × jour
- unité secondaire (US) = prestation

#### 3.1 Tirage des UP

Stratification par type : urgence, longue durée, repas.

Tirage de centres × jours à probabilités proportionnelles à la "taille" = capacité d'accueil hebdomadaire (théorique) en nombre de prestations (les repas de midi et ceux du soir sont des prestations différentes). A ce stade, les jours sont indifférenciés.

Algorithme de tirage : tirage systématique sur fichier (base de sondage) trié, pour bénéficier d'une stratification implicite :

- pour les centres d'hébergement d'urgence : par catégorie de population accueillie (centres pour hommes seuls ; hommes et femmes ; hommes, femmes et couples avec enfants ; femmes avec enfants ; femmes seules), puis par taille décroissante ;
- pour les centres d'hébergement de longue durée : idem ;
- pour les centres de distribution de repas : par taille décroissante.

→ 96 centres-jours tirés, correspondant à 56 centres différents.

Tirage des jours : pour chacune des 4 semaines d'enquête, 4 des 5 journées ouvrées ont été tirées au sort, et affectées par choix raisonné aux centres tirés.

#### 3.2 Tirage des US

Tirage aléatoire simple de 6 prestations.

"Algorithme de tirage" : protocole adapté au contexte local et le moins éloigné possible du tirage théorique (tirage systématique dans une "file d'attente" ; tirage d'avance dans une liste pour les CHRS, procédure qui a présenté quelques inconvénients).

Enregistrement des non-réponses sur des feuilles de contact : refus (des raisons invoquées), "inaptes" (langue, maladie,...).

→ 6 interviews réalisées en moyenne.

### 4. Bilan

591 questionnaires, sur 997 contacts (59%) :

219 en hébergements d'urgence, sur 407 (54%) ;

137 en hébergement de longue durée sur 153 (90%);

235 en centres de distribution de repas sur 437 (54%).

Difficultés de la collecte :

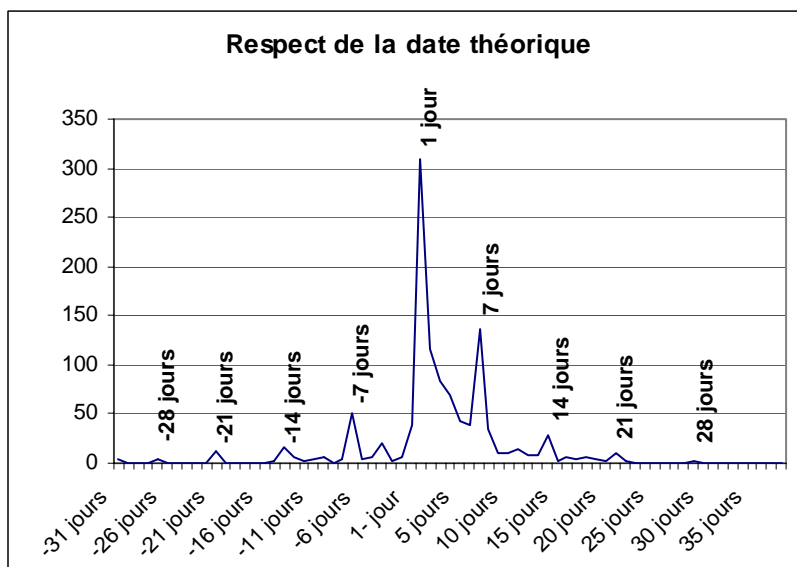
- centres se révélant ne pas appartenir au champ de l'enquête (foyers de travailleurs), ayant refusé l'enquête, ne s'adressant pas à la catégorie de population prévue par le critère de stratification de la liste des centres → ils ont été remplacés par le suivant dans la base de sondage ;
- tailles réelles des centres différentes des tailles théoriques → corrections sur les pondérations ;
- nombre d'enquêtes réalisées dans une UP différent de 6 → corrections sur les pondérations ;
- sélection "non aléatoire" des personnes interrogées → enquêtes non prises en compte.

#### 4.6.3.1 Le respect des dates de tirage

Contrairement aux enquêtes ordinaires, l'enquête présentait la particularité de reposer sur un tirage des jours d'enquête. Dès le mois de décembre, les dates de visites avaient été communiquées aux enquêteurs. La sélection aléatoire des dates d'enquête répondait à deux objectifs : éviter les biais qui auraient pu provenir de la saisonnalité dans la fréquentation des services, aider les enquêteurs à se coordonner à l'avance, le calendrier des visites étant connu dès le mois de décembre. La journée du 22 janvier étant consacrée à des réunions de bilans intermédiaires avec les enquêteurs, aucune visite n'avait été prévue et la collecte principale se terminait le lundi 12 février au soir. Malgré quelques écarts par rapport aux dates théoriques, le calendrier d'enquête a été bien respecté. Les communications téléphoniques ont été le moyen le plus utilisé pour confirmer les dates et les heures des visites. Les interviewers ont dû se rendre disponibles pour que l'échantillonneur puisse organiser le planning des entretiens, ce qui supposait que leur charge de travail ne soit pas trop importante. C'est la raison pour laquelle le programme de travail des enquêteurs a été allégé pendant la durée de cette opération.

#### 4.6.3.2 Comparaison avec la date théorique de l'entretien

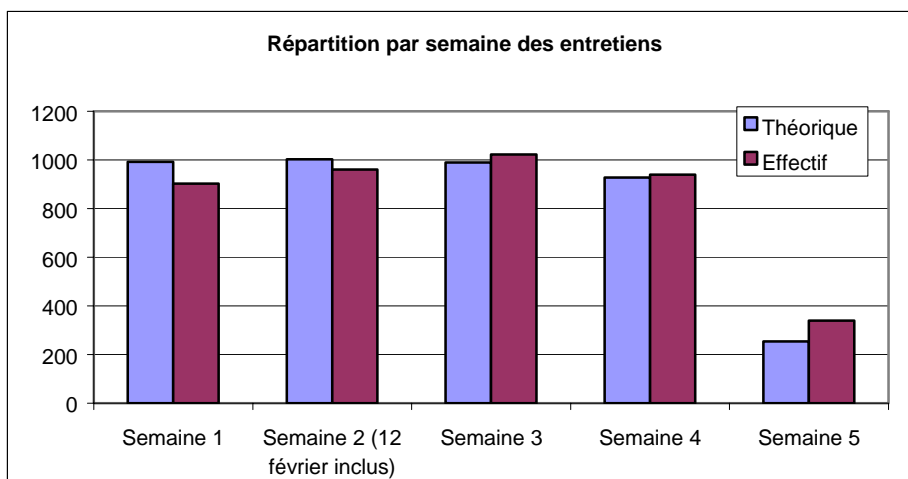
Il est arrivé que le calendrier initial soit modifié quand des structures demandaient des reports de dates ou quand des personnes hébergées en chambre d'hôtel ou en logement dispersé n'avaient pu être jointes. Ainsi, 5% des entretiens ne se sont pas déroulés à la date prévue : 163 ont eu lieu après la date fixée initialement et 53 avant. La plupart des enquêteurs ont réalisé leur interview dans les 3 jours suivants la date théorique, 1/4 entre 6 et 9 jours après et quelques-uns 14 et 21 jours après. Certains ont également avancé la date. Conformément aux indications qui leur avaient été données, les enquêteurs se sont donné 24 heures pour joindre les personnes hébergées en logement éclaté et ont essayé de respecter le jour donné de la semaine dans les autres cas : on constate des pics pour les nombres multiples de 7, c'est-à-dire les entretiens réalisés une ou plusieurs semaines avant ou après, le jour prévu (graphique). Les prises de rendez-vous ont été particulièrement fastidieuses lorsque les visites ont été organisées au domicile des personnes (logement ou chambre d'hôtel) et non pas au siège de l'association. Si les interviewers avaient pris contact directement avec les enquêtés comme pour une enquête ménage ordinaire, la procédure aurait peut-être été simplifiée. Dans ce cas précis, la présence de l'échantillonneur rajoutait inutilement un niveau de décision. Cette difficulté a été aggravée quand les rendez-vous manqués se sont multipliés. C'est un point qu'il faudrait revoir dans une future enquête.



Source : enquête auprès des usagers des services d'hébergement ou de distribution de repas chauds, Insee, 2001.  
 Note : Les entretiens réalisés le jour prévu n'ont pas été représentés sur le graphique.

#### 4.6.3.3 La répartition des entretiens par semaine et par jour

Les enquêteurs ayant assez bien respecté les dates de tirage, l'équilibre des entretiens par jour et par semaine a été satisfaisant. Le nombre d'interviews est relativement bien réparti entre les 4 semaines de collecte. On constate cependant qu'au cours de la première semaine 903 entretiens ont été réalisés sur les 992 prévus. Les enquêteurs ont repoussé la date des premiers entretiens, qu'ils n'étaient pas en mesure de faire à la date souhaitée.



Source : enquête auprès des usagers des services d'hébergement ou de distribution de repas chauds, Insee, 2001.

#### Les 4 semaines

Jours	Effectif	%
Lundi + 12/02	729	19.1
Mardi	792	20.7
Mercredi	742	19.4
Jeudi	766	20.0
Vendredi	739	19.3
Samedi	58	1.5

#### Toutes semaines confondues

Jours s	Effectif	%
Lundi	771	18.5
Mardi	879	21.1
Mercredi	834	20.0
Jeudi	839	20.1
Vendredi	779	18.7
Samedi	64	1.5

#### Jours théoriques du 15/01 au 19/02

Effectif	%
836	20.1
949	22.8
814	19.5
802	19.2
757	18.2
8	0.2

Source : enquête auprès des usagers des services d'hébergement ou de distribution de repas chauds, Insee, 2001.