

UNIVERSITE DE LA MEDITERRANEE AIX - MARSEILLE II  
ECOLE DOCTORALE N°372 DE "SCIENCES ECONOMIQUES ET DE GESTION D'AIX-MARSEILLE"  
FACULTÉ DES SCIENCES ECONOMIQUES ET DE GESTION

**CENTRE DE RECHERCHE SUR LES DYNAMIQUES ET POLITIQUES ECONOMIQUES  
ET L'ECONOMIE DES RESSOURCES**

**C E D E R S**

## **OBSERVATION DU TOURISME**

### **ANALYSE ECONOMIQUE DES RECETTES TOURISTIQUES INTERNATIONALES EN FRANCE**

RAPPORT FINAL  
JUN 2003

ETUDE RÉALISÉE POUR LA DIRECTION DU TOURISME



14, avenue Jules Ferry - 13621 Aix-en-Provence cedex - Tél (33) 04.42.91.48.33 Tcp (33) 04.42.91.48.29  
E.Mail : [ceders@univ-aix.fr](mailto:ceders@univ-aix.fr) - Site Internet : <http://www.ceders.org>

Lors des dernières réunions à la Direction du Tourisme, les différents acteurs (ORT, représentants divers...) sont arrivés à la conclusion quasi-unanime que la quantification du tourisme en France et en particulier du tourisme étranger, se heurte à plusieurs difficultés compte tenu des nouvelles contraintes auxquelles l'observation touristique doit faire face.

Un groupe de travail<sup>1</sup> s'est réuni le 8 janvier 2003 pour analyser les pistes méthodologiques envisageables pour faire évoluer l'observation touristique, à la lumière de différentes expériences conduites en France et à l'étranger.

Cette note a pour objet de faire le point sur cette question.

En premier lieu, les fortes disparités régionales ne facilitent pas la mise en œuvre d'une méthodologie d'enquête nationale cohérente avec les approches régionales.

La méthode des flux, fondée notamment sur les comptages routiers, utilisée dans cinq régions françaises comporte des limites pour l'évaluation du tourisme dans les régions de fort transit, les régions frontalières, ou dans les territoires où la composante touristique n'est pas dominante.

---

<sup>1</sup> Composé notamment de Jean-Michel BLANC (ORT Auvergne), Patrick VECE (ORT Riviera), Gilles NANCY (CEDERS), Jérôme JOUBERT (CEDERS), MM. VASSILLE et LHUILLIER (Direction du Tourisme), Melle SYLVAINIER et Jackie BENESSE (Direction du Tourisme).



Par exemple, la région parisienne, et l'Alsace, espaces dans lesquels une méthode de comptages routiers augmente considérablement le risque de comptabiliser comme touristes des migrations domicile-travail. Mais cette remarque vaut également pour les grandes agglomérations : Lyon, Lille, Marseille...

Par ailleurs, les enquêtes aux frontières sont devenues difficilement praticables depuis la disparition des barrières douanières (espace Schengen) ; de même les enquêtes cordon sont devenues délicates en raison du désengagement des forces de l'ordre.

A l'heure actuelle, le socle méthodologique commun à toutes les Régions est constitué, principalement, par les enquêtes de fréquentation hôtelière et para-hôtelière qui, grâce au partenariat entre Direction du Tourisme, ORT et les professionnels du tourisme, permettent l'établissement d'une base de données fiable même si la quantité d'information produite est plutôt faible.

Ce type d'enquêtes ne permet pas d'évaluer le tourisme non-marchand et ne fournit donc qu'une partie de l'information requise ; elles pourraient être néanmoins réintégrées dans le dispositif suggéré dans cette note, basé sur le croisement de plusieurs sources d'information.

Les enquêtes sur sites, pratiquées dans certains observatoires régionaux, sont également une composante de l'observation touristique ; on ne peut cependant pas ignorer les biais que ces approches introduisent quand il s'agit d'estimer la population touristique, en particulier les touristes hébergés dans le secteur non-marchand.

Le recours aux enquêtes sur sites suppose la mise au point de techniques de redressement des échantillons, que nous présentons en annexes.

*Sur le fond, la question centrale consiste à savoir si, dans le cas où on ne dispose pas d'une base de sondage, fournie par exemple par la méthode des flux, il est possible de remonter à la population totale à partir de sources d'informations multiples.*

Différents travaux ont ainsi été réalisés en France et à l'étranger afin d'établir, en l'absence d'une base de sondage, une méthodologie d'enquête, permettant à la fois de minimiser les biais d'échantillon et d'estimer la population totale.

L'idée principale de la méthode candidate, pour remplacer dans certains cas la méthode des flux, serait de remonter à la population totale (pour laquelle on ne dispose *a priori* que d'informations ponctuelles) en réalisant des croisements de différentes sources d'information.

A cet égard, la technique développée par l'INSEE dans l'enquête SDF (2001), dont l'objet était d'estimer la population totale des Sans Domiciles Fixes en France, et le dispositif d'observation élaboré par la région des Asturies (Nord de l'Espagne) apparaissent comme complémentaires. Ces deux approches reposent sur l'identification des individus sur des lieux produisant des services à destination des usagers (les SDF pour l'INSEE, les touristes dans les Asturies).

Dans la mesure où la méthode permet d'estimer la population totale, à partir des liens que les individus tissent avec des lieux producteurs de services, l'analyse des caractéristiques qualitatives et quantitatives de la population d'intérêt se fera, dès lors, à partir d'enquêtes « *in situ* ».

## **1 – ESTIMATION DE LA POPULATION TOURISTIQUE TOTALE EN L'ABSENCE DE BASE DE SONDAGE.**

Quelle que soit la méthode d'analyse de l'activité touristique, les données collectées doivent fournir des informations sur les motivations du séjour (connaissance de la région...), son type (organisation du voyage, moyen de transport utilisé, hébergement, dépenses réalisées, sites visités...), l'évaluation globale de la région (accueil, hébergement, excursions...) ainsi que sur le profil des touristes (âge, pays d'origine...).

La fiabilité des estimations repose sur la connaissance de la base de sondage, c'est à dire du nombre et des caractéristiques des individus composant la population étudiée.

En l'absence de base de sondage, la représentativité des échantillons analysés dépend de la possibilité de produire une estimation de la population totale à partir d'une identification partielle de la base (population accueillie dans les hébergements marchands).

Il est possible de ranger les touristes dans cinq catégories d'hébergement :

- Touristes hébergés dans le secteur hôtelier et para-hôtelier,
- Touristes hébergés dans les gîtes, fermes auberges...

- Touristes hébergés dans des structures collectives de type camps de vacances,
- Touristes hébergés dans le secteur non-marchand (parents, amis, ...),
- Visiteurs d'un jour.

A l'exception des hébergements non-marchands, les autres modes d'hébergement peuvent faire l'objet d'enquêtes quasi-exhaustives.

Il est nécessaire de disposer d'au moins deux méthodes complémentaires, l'une permettant d'identifier la population des hébergements non-marchands (enquête sur sites), l'autre celle des hébergements marchands (enquête de fréquentation des structures d'hébergement collectif). Une troisième technique est indispensable pour redresser l'échantillon.

*1.1– Estimation de la population touristique hébergée dans le secteur marchand : les enquêtes de fréquentation<sup>2</sup>*

L'idée générale de la méthode consiste, dans un premier temps, à estimer la population totale des hébergements marchands à partir des taux d'occupation.

Dans un second temps, une enquête est réalisée dans chaque catégorie d'hébergement marchand, permettant de définir des profils de clientèle (par CSP, nationalités, âges...).

Les questionnaires sont redressés en tenant compte du poids d'une catégorie donnée d'hébergement dans l'hébergement marchand total mesuré par le nombre de lits.

Le plan de sondage doit tenir compte du caractère saisonnier de la fréquentation touristique.

Il convient donc, dans un premier temps, de déterminer la distribution touristique par période (saisons) ; l'activité touristique est très inégale durant la totalité d'une année, chaque période doit donc être étudiée en fonction de ses caractéristiques.

Le nombre d'interviews réalisées ne peut être le même sur toute une année d'étude, certaines saisons sont propices au tourisme (hautes saisons), d'autres le sont moins (basses saisons).

Le nombre d'interviews qui devront être réalisées durant l'année doit donc dépendre des taux d'occupation (observés par exemple les années précédentes) pour chaque saison.

---

<sup>2</sup>Nous nous référons à la méthodologie utilisée par les Asturies (région du Nord de l'Espagne), notamment aux travaux du SITA (*Sistema de Informacion Turistica de Asturias – Tourist Information System of Asturias*).

E.TORRES, I.SUSTACHA, J.M. MENENDEZ, L.VALDES "A solution to problems and disadvantages in statistical operations of surveys of visitors at accommodation establishments and at popular visitor places".

6<sup>th</sup> International Forum on Tourism Statistics, Budapest 25-27 September 2002.

Il semble aussi raisonnable de partitionner l'échantillon par jour d'interview. Par exemple, la distinction entre tourisme de week-end et tourisme de semaine permettrait d'isoler le tourisme d'affaires (qui, bien souvent biaise les résultats d'enquêtes). Effectivement, il est fortement probable qu'un individu interviewé en basse saison, un jour de la semaine, séjourne dans le pays pour des raisons professionnelles (tourisme d'affaire principalement du lundi au vendredi).

La structuration de l'échantillon (par période, type d'hébergement...) permet d'établir les différentes composantes de la clientèle touristique du secteur marchand et d'en établir l'évaluation. Cette structuration se fait *a posteriori* : les proportions observées sont représentatives de la clientèle touristique du secteur marchand.

### 1.2– Estimation de la population touristique hébergée dans le secteur non-marchand : les enquêtes sur sites<sup>3</sup>

Les enquêtes sur sites permettent d'estimer les composantes de la population touristique non observables dans les enquêtes de fréquentation hôtelière et para-hôtelière : population touristique ayant recours à un type d'hébergement non-marchand.

Le principe d'une telle enquête est d'échantillonner des prestations et d'interroger les individus qui les consomment pour remonter des prestations (différents services offerts aux touristes) aux individus.

Le passage d'une pondération des prestations échantillonnées à une pondération des individus consommateurs permet d'estimer des paramètres caractéristiques à la population étudiée.

#### **1.2.1 – Méthodologie de l'enquête**

**Les sites touristiques sondés, sont définis comme des lieux de prestations à destination des touristes.** Ils doivent être préalablement définis de manière à ce que

---

<sup>3</sup> Nous pouvons nous référer à la méthodologie employée par l'INSEE pour la réalisation d'une enquête visant à quantifier la population Sans Domicile Fixe en France (2001).

« Hébergement et distribution de repas chauds : le cas des sans-domicile » INSEE Première N°823, Janvier 2002.

l'échantillon sondé soit le plus représentatif possible de la population étudiée, tant sur le plan temporel que spatial (forte diversité de sites<sup>4</sup>).

L'impossibilité de collecter des données pour chaque lieu d'enquêtes, sur toute la période considérée, nous pousse à sonder les sites sur des périodes particulièrement propices au tourisme (hautes saisons) mais également de choisir des semaines de basse-saison.

Il convient aussi de déterminer, par sites enquêtés, le nombre d'interviews qui devront y être réalisées (nombre proportionnel à la fréquentation touristique du site sondé) : par exemple, le nombre d'interviews effectuées entre un musée et une plage au mois d'août ne sera pas le même, la majeure partie de la clientèle se trouvant sur la plage.

**La population d'intérêt** est composée de personnes ayant fréquenté au moins une prestation du champ de l'enquête (un service à destination des touristes) pendant la période de référence. Les prestations sont fournies sur des bases temporelles qui varient selon leur nature, dans un lieu précis (que l'on appellera *site*), à chaque lieu correspond un ou plusieurs types de services.

L'unité statistique échantillonnée (*prestation*) sera définie comme un triplet (*service, jour, horaire*) ; le service sera une prestation type donnée dans un site donné, l'horaire un intervalle de temps, un jour donné.

Dans un premier temps, des bassins touristiques (définis en collaboration avec les professionnels du tourisme) sont tirés. On obtient une base de sondage de sites, pour chacun on connaît la capacité moyenne, le type de service rendu et d'autres informations.

Pour des raisons pratiques, il n'est pas possible d'enquêter l'ensemble des sites et de maintenir sur le terrain un enquêteur une journée entière sur un site donné, il convient donc d'échantillonner des sites dans les différents bassins, des intervalles de temps (à calibrer de façon à ce qu'un individu ne puisse pas fréquenter deux prestations différentes durant cet intervalle), et des prestations au sein d'un (*site, intervalle de temps*) tirés.

---

<sup>4</sup> alentours des stations balnéaires, musées, centres culturels, parcs et réserves naturels, stations de sports d'hiver, lacs aménagés, réserves de pêche, aéroports, gares, parkings, stations-essence, cinémas, marchés, supermarchés, théâtres, restaurants, boutiques de souvenirs...

Les unités pertinentes à considérer seront les couples  $(s,t)$  correspondant au croisement d'un site et d'un intervalle de temps (élimination de sites fermés durant le créneau horaire  $t$  considéré).

L'échantillonnage des prestations peut se faire de deux manières :

- comptage pendant la période d'échantillonnage du nombre  $N$  de prestations servies (détermine la probabilité de tirage des prestations échantillonnées), ce rôle pourrait être confié à un « dénombreur » ;
- parallèlement, tirage de type systématique, ce qui pourrait constituer le rôle d'un superviseur.

Deux types de tirage sont dès lors possibles : soit on tire  $n$  prestations,  $n$  étant fixé avant l'enquête, soit on tire les prestations avec un taux de sondage  $f$  (dépendant du nombre de prestations attendues  $\check{N}$  et le nombre de prestations que l'on désire échantillonner  $n$ , la taille  $N$  de l'échantillon n'est pas connu avant la fin de la période d'échantillonnage).

### **1.2.2 - Estimations**

Un intervalle de temps est défini de façon à ce qu'aucune personne ne puisse fréquenter simultanément deux prestations différentes ; pour chaque individu qui a fréquenté un des sites de la base pendant l'intervalle de temps considéré, ne correspond qu'une seule prestation.

La population d'intérêt sera donc déterminée par l'ensemble des personnes se présentant au cours d'une unité de temps (jour et intervalle temps donnés) dans l'ensemble des sites touristiques de la base.

En définissant les liens existant entre les prestations de la base et les individus qui les fréquentent, nous pouvons déterminer un système de poids. Le poids d'un individu dans la population d'intérêt sera déterminé par le rapport entre somme des poids affectés aux prestations qui ont servi à l'attraper et nombre de liens de la base de sondage.

Une fois ces poids définis, on est en mesure d'estimer la population d'intérêt (cf. démonstration Annexe 1).



Dans la méthodologie présentée, la connaissance des liens entre les personnes et l'univers des prestations est indispensable mais en pratique, ces relations ne sont pas connues :

- la collecte étant étalée dans le temps, un individu interrogé en début de période ne peut prévoir les services qu'il va fréquenter après la date d'interview,
- la mémoire des personnes interrogées fait défaut au-delà de certains jours ; de plus, le repérage des prestations par l'enquêteur peut s'avérer problématique.

Ainsi, le touriste interviewé en début de séjour, qui n'a pas une vision complète des activités à venir, sera affecté d'un poids inférieur à celui du touriste interviewé en fin de séjour.

La méthode de redressement consiste à affecter une pondération au jour d'enquête, par exemple, la part du jour  $j$  dans la valeur ajoutée hebdomadaire du site enquêté.

Pour prendre en compte l'incertitude des interviews de début de séjour, il convient d'inclure des questions de type : « où allez-vous être hébergés, que comptez vous faire pendant votre séjour? », le problème qui se pose alors est celui de la véracité des réponses et du taux de non-réponse.

L'intuition est d'estimer une probabilité de fréquentation d'autres sites pendant le reste du séjour : les liens que l'on pourrait ainsi définir sembleraient une bonne approximation des vrais liens.

### 1.3- Méthodologie de redressement de l'échantillon : la correction des biais.

- **La correction de non-réponse** doit être prise en compte dans l'enquête, la correction devra se faire par sous-groupes homogènes.

Pour constituer des sous-groupes homogènes, d'autres variables de la base peuvent être utilisées : jour de la semaine, période du jour, bassins touristiques...

Par exemple, dans le cas de l'estimation de la population touristique hébergée dans le secteur non-marchand, les prestations répondantes peuvent être repondérées<sup>5</sup>.

---

<sup>5</sup>  $\lambda_i = \frac{1}{\delta_i \pi_i}$  avec  $\pi_i$  la probabilité d'inclusion de la prestation  $i$  dans l'échantillon et  $\delta_i$  la probabilité *a posteriori* que la prestation donne lieu à une réponse.

Certaines non-réponses peuvent venir du fait qu'un même individu est échantillonné plusieurs fois, on gagnerait ainsi en qualité à identifier l'individu non-répondant et la raison de sa non-réponse : la « vraie valeur » pouvant être récupérée dans un questionnaire déjà rempli.

Les doublons sont très difficiles à évaluer, il est par ailleurs complexe de vérifier la véracité des dires d'un individu sur son interview passée (l'individu déclare avoir déjà été interrogé mais l'a-t-il déjà vraiment été, si oui est-ce par la même enquête ?).

- **La minimisation des biais** (biais de durée, biais de fréquence<sup>6</sup>) se réalise essentiellement grâce à la pondération de chaque interview par son poids représentatif dans la population (coefficient correcteur).

Les coefficients correcteurs sont construits à partir de trois variables fondamentales :

- Le nombre de personnes participant au séjour, noté par  $n$ .  
« Les compagnons de route » auront un comportement identique en ce qui concerne le mode de transport, le type d'hébergement, les motivations du séjour... En supposant une interview par « famille de touristes », les réponses à l'enquête sont élargies au nombre de participants ( $n$ ),
- La durée du séjour, en nombre de nuitées  $k$  (évaluation de la population touristique hébergée dans le secteur marchand),
- La fréquence de passage sur sites : nombre de passages d'un touriste sur un point de collecte de données (évaluation de la population touristique hébergée par le secteur non-marchand). Ceci permet de minimiser le biais de durée du séjour et empêche les doublons au niveau de la collecte.

Le redressement statistique de l'échantillon est ici opéré par une structuration de la population d'intérêt par période d'enquête et par type d'hébergement. Dans le cadre de l'évaluation de la composante marchande de la population touristique, le redressement peut être tout d'abord réalisé pour chaque sous-groupe, un redressement intra-groupe peut ensuite être proposé (cf. Annexe 2).

---

<sup>6</sup> L'individu séjournant quatre jours sur un même lieu a deux fois plus de probabilités d'être interviewé que l'individu y séjournant deux jours.

La réalisation parallèle de ces trois méthodologies complémentaires générerait une base de données fiable, quasi-exhaustive.

En effet, la combinaison des deux types d'enquêtes (de fréquentation et sur sites) permet de couvrir la totalité de la population d'intérêt et d'obtenir les informations nécessaires à l'évaluation de la population touristique recourant à un hébergement non-marchand (fréquemment omise des enquêtes traditionnelles).

Le redressement statistique de l'échantillon par des pondérations pour chaque individu interviewé (par la représentativité de son comportement dans la population), méthode simple de minimisation des principaux biais, offre une certaine justesse dans les estimations.

Certains questionnements d'ordre pratique subsistent pour la mise en application potentielle de la méthode candidate.

La sélection déterminante des sites à sonder, ainsi que le choix du questionnaire adéquat doivent faire l'objet de quelques prérogatives, afin d'optimiser les résultats escomptés.

## **2 - LES ÉTAPES DE LA MÉTHODE**

### *2.1 - Choix des sites touristiques*

Comme nous l'avons souligné, pour la méthodologie d'analyse de la composante non-marchande de la population touristique (enquêtes sur sites) proposée, le choix des sites touristiques à sonder est crucial.

Les sites sont définis comme des lieux de prestations, de services à destination des touristes.

Il convient donc de prendre en compte des sites larges et diversifiés (supermarchés, plages, musées, magasins de souvenirs, sites naturels et culturels...) afin d'établir un échantillon d'individus à interviewer le plus représentatif possible de la population d'intérêt.

Une des difficultés majeures pour la réalisation de ce type d'enquête est directement liée à la caractéristique intrinsèque de la population touristique : il n'existe pas de sites ni de prestations à destination exclusive des touristes, la population d'intérêt se mélange à la population des autochtones. Il convient de ne pas le perdre de vue lors du processus de dénombrement des sites à sonder.

De plus, le risque d'omission de certains sites lors du dénombrement, ainsi que le décalage temporel entre le recensement et l'enquête ne tenant pas compte des possibles apparitions de nouvelles structures, sont des problèmes importants pour la mise en pratique de la méthodologie candidate.

En effet, les individus fréquentant des sites non définis par l'inventaire ne fréquenteraient aucun service de la base de sondage : les passages sur des lieux non recensés ne sont pas comptabilisés dans les liens, biaisant ainsi l'estimation de la population d'intérêt.

Le découpage des différentes régions en bassins touristiques faciliterait néanmoins un repérage quasi-exhaustif : ceci nous permettrait de recouper les informations dont disposent les professionnels du tourisme (ORT, Offices du Tourisme...).

La collaboration avec ces professionnels, plus proches des bassins touristiques dont ils ont charge, permettrait de repérer les sites susceptibles de générer les collectes de données les plus intéressantes par rapport à la population d'intérêt.

## *2.2 - Le plan de sondage*

Avant l'enquête, il convient de réaliser un premier degré de sondage dont l'objectif est d'extraire des sites touristiques. Il faudra ainsi établir une liste complète des différents bassins d'attraction des touristes dans la (ou les) région(s) ciblée(s) pour les besoins de l'enquête<sup>7</sup>.

La réalisation de ce répertoire nous permet de caractériser chaque lieu par une pondération (déterminée, par exemple, en fonction des enquêtes de fréquentation hôtelière ou des résultats issus d'anciennes enquêtes cordon).

Un tirage de type aléatoire permet, par la suite, d'extraire certains lieux à partir desquels un second degré de sondage est réalisé.

A partir de ces sites, la liste complète des distributeurs de prestations à destination des touristes doit être établie. A chaque distributeur est associée une pondération dépendant principalement du nombre de prestations estimées sur la période d'enquête<sup>8</sup>.

---

<sup>7</sup> Villes, stations balnéaires, parcs naturels...

<sup>8</sup> Un entretien téléphonique avec chaque distributeur est alors nécessaire afin de connaître le nombre de prestations attendues.

Un troisième degré de sondage consiste à tirer arbitrairement (tirage aléatoire) un certain nombre de prestations, ne seront interrogés que les individus fréquentant ces prestations durant le champ de l'enquête.

*Une prestation retenue pourrait être, par exemple, la consommation d'un « repas du jour » à 25 euros dans un restaurant d'Aix en Provence, durant la première quinzaine d'Août. Il conviendrait d'aller interviewer l'ensemble des personnes consommant ce type de repas entre midi et quatorze heures, durant la semaine considérée.*

*A cette prestation, préalablement extraite d'une base de sondage, correspond un poids (déterminé par le nombre de couverts moyen réalisé dans ce restaurant lors de la semaine d'enquête), ce poids nous permet de donner une pondération à chaque touriste consommateur (cf. annexe 1, Exemple simplifié d'application de la méthode du partage des poids).*

Parallèlement à l'échantillonnage des sites, il convient, donc, de tirer purement arbitrairement des intervalles de temps. Les intervalles sont définis de sorte qu'une même personne ne puisse recevoir deux prestations durant l'intervalle de temps défini.

### 2.3 - Le choix du questionnaire

Le questionnaire élaboré par la Direction du Tourisme pour l'enquête sur les visiteurs étrangers<sup>9</sup> (enquête EVE) est satisfaisant pour le cadre de notre méthodologie candidate.

Ce questionnaire (pouvant être réalisé sur les sites) permet d'obtenir des informations sur le nombre de personnes participant au séjour, le motif du séjour (professionnel, personnel), le mode d'hébergement, les activités réalisées, l'évaluation globale du pays...

Nous pouvons cependant, proposer une implémentation à l'aide du questionnaire réalisé pour la quantification du tourisme, par le Canada, en collaboration avec l'Espagne, la Suède, la France, la Finlande, l'Italie, le Mexique, les Etats-Unis et l'Organisation Mondiale du Tourisme (2000/01 cf. Annexe 4).

Le questionnaire type est assez complet, édité dans une langue internationale ; la France pourrait retenir les points forts de cette expérience, par exemple éditer les formulaires dans plusieurs langues afin de faciliter la compréhension des touristes étrangers interrogés.

Les dernières questions correspondent aux coordonnées de la personne interviewée, ceci peut faciliter la constitution d'un échantillon type de touristes étrangers, habitués, fidélisés, susceptibles d'être suivis et régulièrement enquêtés.

---

<sup>9</sup> Cf. Annexe 3.

#### 2.4 - Estimation des dépenses

Les différentes caractéristiques de l'industrie touristique peuvent être énoncées comme suit :

- les touristes peuvent consommer différents types de produits à différentes périodes et différents endroits,
- l'ensemble des biens et services produits pour les touristes est un agrégat complexe qui ne peut être défini *a priori* et une bonne fois pour toutes,
- peu de biens et services peuvent être répertoriés comme étant uniquement pour les touristes, une majeure partie de la production sert avant tout les résidents.

Retracer le comportement de dépenses d'un touriste est donc très complexe. Cependant, il est nécessaire pour chaque pays soucieux d'établir des statistiques fiables sur le rôle du tourisme dans l'économie nationale, de disposer d'estimateurs clairs et synthétiques sur les dépenses.

Pour retracer le comportement de dépenses des touristes séjournant en France, nous proposons de remettre aux touristes dès leur arrivée, un « carnet de bord » de dépenses<sup>10</sup>.

Lors des interviews, sur les sites touristiques ou sur les lieux d'hébergement, on se propose de remettre aux touristes un questionnaire enregistrant plusieurs classes de dépenses potentielles (dépenses alimentaires, tabac, vêtements, services médicaux et paramédicaux, dépenses d'hébergement, transports locaux...).

Les touristes doivent être incités à tenir correctement ce « cahier de bord », en conservant les pièces justificatives de chaque dépense réalisée (notes, factures...). Le questionnaire pourra être rendu par le touriste sur un des points de collecte (sites touristiques ou structures d'hébergement) les derniers jours de son séjour.

---

<sup>10</sup> nous nous référons à l'application de cette méthode dans la région vénitienne, travaux de recherche du Ciset - *International Center of Studies on the Tourism Economy* Université de Venise, en collaboration avec l'observatoire du tourisme de la région vénitienne (*Tourism Department of the Veneto Region*). M.MANENTE et V.MINGHETTI "The collection of international tourism expenditure statistics on open area: methodology and applications. The case of the Veneto region", *Tour. Hosp. Manag.* Vol. 4, N°1, Pp. 93-106.

Il est évidemment souhaitable de mettre en place un système d'incitations, récompensant le taux effectif de réponse (inciter le touriste à remettre ses factures, tickets de caisse, notes...) et la bonne tenue de ce carnet.

L'ORT Riviera Côte D'Azur a fait des propositions qui semblent être un bon compromis entre l'approche exhaustive de la région vénitienne et les méthodes appliquées par les différents ORT utilisant la méthode des enquêtes cordon.

Il conviendrait de se rallier à l'approche de cet observatoire pour l'estimation des dépenses<sup>11</sup>.

Une méthode alternative aux méthodologies traditionnelles pour la quantification du tourisme en France pourrait ainsi s'énoncer comme suit :

- Réalisation d'enquêtes de fréquentation dans les structures d'hébergement collectif : évaluation de la composante marchande de la population touristique.
- Réalisation d'enquêtes sur sites (lieux de prestations à destination des touristes) : évaluation de la composante non-marchande de la population touristique ; en complément de l'interview traditionnelle, les individus devront être interrogés sur leur moyen d'hébergement.
- Redressement de l'échantillon statistique. Etablissement de poids pour chaque interview, permettant la minimisation des biais et l'étude de la composition de la population d'intérêt permettant de remonter à l'estimation de la population totale.

L'obtention des informations nécessaires à l'évaluation des caractéristiques quantitatives et qualitatives du tourisme se fait par le croisement de plusieurs sources. Ainsi, les informations issues du travail des professionnels du tourisme permettraient, grâce à leur partenariat avec la Direction du Tourisme, un découpage des régions par bassins touristiques facilitant le choix des sites et le redressement statistique de l'échantillon.

---

<sup>11</sup> Cf. Annexe 5.

Le questionnaire à adopter pourrait comporter deux parties :

- une première, fidèle au questionnaire de l'enquête EVE, quelques variantes peuvent être réalisées à la lumière du questionnaire de l'Organisation Mondiale du Tourisme (langue du questionnaire...),
- une seconde partie concernant les dépenses réalisées, conçue comme un cahier de dépenses. Ce carnet peut être remis aux touristes lors de leur premier passage sur un point de collecte (lieu d'hébergement ou site), les touristes devant le restituer lors de leur dernier passage sur un de ces points.

Afin d'évaluer la faisabilité et les résultats générés par cette méthode candidate, il conviendrait de pouvoir la mettre en place dans un ou deux départements, susceptibles d'accepter un partenariat (notamment financier) avec la Direction du Tourisme et les différents professionnels désireux d'y participer.



## ANNEXE 1

### Estimation de la population touristique hébergée dans le secteur non-marchand

Nous reprenons les notations utilisées dans les travaux de l'INSEE :

- les différents jours d'enquête sont repérés par l'indice  $j$ ,
- $S_N$  : sera l'ensemble des sites touristiques ouverts le jour  $j$ , repérés par l'indice  $n$  ;
- $\Pi_S$  : représente l'ensemble des prestations servies dans le site  $s$ , repérés par l'indice  $i$ .

On définit un intervalle de temps  $t$  de façon à ce qu'aucune personne ne puisse fréquenter deux prestations différentes durant cet intervalle de temps.

Il en ressort donc qu'à chaque individu  $k$  appartenant à la population des touristes qui ont fréquenté un des sites de la base pendant l'intervalle de temps considéré, correspond une seule prestation  $i$ .

Pour un triplet  $(S_{n,j}, t)$  donné, la correspondance entre prestation et individu est donc biunivoque.

Si  $P_{n,j,t}$  représente l'ensemble des personnes se présentant au cours de l'unité de temps  $(j, t)$  dans le centre  $S_n$ , la population d'intérêt est alors définie par :  $P(J) = \bigcup_{n,j,t} P_{n,j,t}$ .

La population d'intérêt croît donc avec le temps mais son effectif croît moins vite : d'un jour à l'autre, on retrouve les mêmes personnes dans les centres. La population d'intérêt dépend fondamentalement de la période de collecte.

La population touristique ne fréquente, à un moment donné, qu'épisodiquement les sites de la base ; pour la couvrir, il convient d'enquêter sur une période de temps où toutes les personnes de cette population ont au moins une fois recours à des services (semaine, mois).

De plus, ces populations se renouvellent dans le temps, d'une année sur l'autre des entrées et sorties, nombreuses, interviennent.

Il convient dans la méthodologie de différencier deux types de variables :

- Des variables qui varient au cours du temps : nombre de sites fréquentés le jour de l'enquête (nombre de sites visités, nombre d'excursions...),

- Des variables fixes au cours du temps (propres aux caractéristiques des individus : nationalité des touristes, âge, profession...).

On définit une application  $K$ , qui à toute prestation  $i$  servie durant la période de référence  $J$  dans l'ensemble des sites touristiques du champ de l'enquête, associe l'individu bénéficiaire de cette prestation :

$$K : \{\text{prestations}\} \rightarrow \{\text{individus}\}$$

$$i \rightarrow K\{i\}$$

On cherche dans un premier temps à estimer un total relatif à une variable  $Y$  (âge auquel l'individu a terminé ses études...) définie sur la population  $P(J)$  :

$$T_j = \sum_{k \in P(j)} Y_k .$$

La population d'intérêt  $P(J)$  est l'image par  $K$  de l'ensemble des prestations servies durant la période de référence dans l'ensemble des sites du champ de l'enquête ; pour tout  $k \in P(j)$ , on définit :  $R_k(j) = \text{card}(K^{-1}(k))$ , le nombre d'antécédents de  $k$  au cours de la période d'enquête (le nombre de services utilisés par l'individu  $k$  durant la période  $J$  de l'enquête dans l'ensemble des sites de la base de sondage – *nombre de liens*).

Nous avons :

$$T_j = \sum_{k \in P(J)} Y_k = \sum_{j=1}^J \left( \sum_{s \in S_N} \left( \sum_{i \in \Pi_s} \frac{Y_{K(i)}}{R_{K(i)}(J)} \right) \right) = \sum_{k \in P(J)} \frac{Y_k}{R_k(J)} \left[ \sum_{j=1}^J \left( \sum_{s \in S_N} \left( \sum_{i \in \Pi_s; K(i)=k} \right) \right) \right]$$

La quantité entre crochets est le nombre de prestations servies à l'individu  $k$  durant la période  $J$  dans l'ensemble des sites du champ de l'enquête<sup>12</sup> (soit  $R_k(J)$ ).

Le problème posé n'est alors rien d'autre qu'un problème d'estimation du total de la variable sur les prestations<sup>13</sup>. L'application de la méthode du partage des poids<sup>14</sup> doit permettre l'estimation du total de la population d'intérêt.

On suppose deux populations  $U$  et  $V$ , les unités de  $U$  seront les prestations dans le champ de l'enquête, les unités de  $V$  seront les personnes ayant bénéficié d'au moins une prestation pendant la période de l'enquête ( $V = P(J)$ ).

<sup>12</sup> La variable  $Y$  prenant la même valeur pour toutes les prestations  $i$  concernant l'individu  $k$ , nous obtenons l'égalité précédemment présentée.

<sup>13</sup>  $Y_{K(i)}$  est rattaché à la prestation  $i$ , nous pouvons ainsi la noter  $Y_i$  ; de même,  $R_{K(i)}(J)$  peut se noter  $R_i(J)$ .

<sup>14</sup> P.ARDILLY et D.LE BLANC, *Sampling and weighting a survey of homeless persons : a french example*, Survey Methodology, Vol.27, No.1, pp.109-118, Statistics Canada, June 2001.

De plus, on pose  $\text{card}(V) \leq \text{card}(U)$  (chaque unité de  $U$  ayant au moins un lien avec une unité de  $V$ ).

On définit une matrice de liens reliant les prestations aux individus qui les ont fréquentées (liens entre les unités des populations  $U$  et  $V$ ) :

- $r_{ik}=1$  si la personne  $k$  a fréquenté la prestation  $i$ ,
- $r_{ik} = 0$  sinon.

Nous nous intéressons au total d'une variable d'intérêt  $y$  sur  $V$  (nationalité de l'individu interrogé, âge...), nous avons :

$$Y = \sum_{k=1}^m y_k \quad \text{et} \quad r_k = \sum_{i \in U} r_{ik}$$

L'identité  $Y = \sum_{i \in U} \sum_{k \in V} \frac{r_{ik}}{r_k} y_k$  permet de définir pour tout  $i \in U$  la variable :

$$z_j = \sum_{k \in V} \frac{r_{ik}}{r_k} y_k \quad \text{avec} \quad Z = \sum_{i \in U} z_j = \sum_{k \in V} y_k = Y$$

On suppose que l'on dispose d'un échantillon  $S_U$  issu de la population  $U$  auquel est associé un jeu de poids  $(w_i)_{i \in S_U}$ <sup>15</sup>, ceci nous permet de définir indirectement et implicitement un échantillon prélevé dans  $V$  – ensemble des personnes ayant bénéficié d'au moins une prestation :  $S_V$ , tel que :

$$S_V = \{k \in V; \exists i \in S_U, r_{ik} = 1\}$$

On collecte les  $r_{ik}$  respectivement sur  $S_U$  et  $S_V$ , le total est estimé par :  $\hat{Z} = \sum_{S_U} w_i z_i = \hat{Y}$

Si les poids sont sans biais,  $\hat{Z}$  est sans biais,  $\hat{Y}$  estime sans biais  $Y$ .

On peut réécrire  $\hat{Z} = \sum_{S_U} w_i \sum_{k \in V} \frac{r_{ik}}{r_k} y_k = \hat{Y} = \sum_{S_V} y_k \left( \sum_{S_U} \frac{w_i r_{ik}}{r_k} \right) = \sum_{S_V} y_k \tilde{w}_k$  avec  $\tilde{w}_k = \frac{1}{r_k} \sum_{S_U} w_i r_{ik} \quad \forall k \in S_V$ .

$r_k$  est le nombre de services fréquentés par la personne interrogée durant la période de référence de l'enquête. Le poids d'un individu est égal à la somme des poids des prestations qui ont servi à « l'attraper » divisée par le nombre de liens qu'il possède avec la base de sondage (nombre total de prestations consommées par l'individu  $k$ ).

<sup>15</sup> Les  $w_i$  sont les pondérations relatives aux prestations (obtenues par le croisement des trois degrés de sondage de la base, cf. 2.3. *Le plan de sondage*). Ces pondérations sont données par l'inverse de la probabilité d'inclusion de la prestation dans la base de sondage.

L'individu  $k$  représentera  $x$  individus représentatifs, c'est donc la somme des poids individuels qui nous donne l'estimation de la population totale.

***Exemple simplifié d'application de la méthode du partage des poids<sup>16</sup> :***

*Pendant la période de référence de l'enquête, a été servi un total de dix prestations, cinq relatives par exemple à des services de vente à destination des touristes notées V1 à V5, et cinq relatives à des services de restauration, notées R1 à R5.*

*On a tiré au sort cinq de ces prestations, V1, V2, R2, R3 et R4, toutes ont une même probabilité de tirage égale à  $5/10 = 1/2$ .*

*Quatre personnes, dont l'une a été touchée deux fois, ont été enquêtées au titre de ces cinq prestations : P1 au titre V1, P2 au titre de V2 et R2, P3 au titre de R3, P4 au titre de R4. Cette dernière personne P4 a déclaré avoir bénéficié, pendant la période de référence de l'enquête, de deux autres prestations.*

*Au total, les quatre personnes enquêtées ont donc bénéficié de sept prestations.*

*En appliquant la méthode du partage des poids, le poids d'un individu est égal à la somme des poids des prestations (donnés par l'inverse des probabilités d'inclusion) à partir desquelles il a été enquêté, divisé par le nombre total de prestations qui lui ont été servies. Sous cette hypothèse et puisque toutes les prestations enquêtées sont affectées d'un même poids, en l'occurrence  $10/5 = 2$ , les poids de P1 et P3 sont égaux à 2 de même que celui de P2 (égal à  $4/2$ ). Le poids de P4 s'établit quant à lui à  $2/3$ .*

*Le nombre total de personnes ayant bénéficié d'au moins une prestation pendant la période de référence de l'enquête sera ainsi estimé à  $2 + 2 + 2 + 2/3 = 6,67$ . Ce résultat est à rapprocher d'un maximum théorique de 7 (cas où les 3 prestations restantes auraient été servies à 3 personnes différentes) et d'un minimum théorique de 5 (cas où les 3 prestations restantes auraient été servies à une seule et même personne).*

---

<sup>16</sup> C.BROUSSE, B.de la ROCHERE, E.MASSE, *L'enquête de l'INSEE auprès des usagers des services d'hébergement et de distribution des repas chauds : une méthodologie inédite pour étudier la population des sans-domicile*, Courrier des Statistiques n°104, décembre 2002.

## ANNEXE 2

### Redressement statistique de l'échantillon : évaluation de la composante marchande de la population touristique

#### - Redressement statistique par sous-groupe :

Tous les questionnaires sont corrigés par un poids  $\gamma_{i,j,m}$  donné par l'inverse de la fréquence de passage aux points de collecte<sup>17</sup>, multiplié par le nombre de participants au voyage.

$$\gamma_{i,j,m} = \frac{1}{h_{i,j,m}} \times n_{i,j,m}$$

Avec  $h_{i,j,m} = k_{i,j,m} + 1$  pour les structures d'hébergement.

Nous notons :

- i**, les quatre trimestres de l'année étudiée,
- j**, le type d'hébergement,
- m**, la m<sup>ième</sup> interview réalisée durant le trimestre i pour les touristes de type j,
- k<sub>i,j,m</sub>** représente le nombre de nuitées passées par le touriste de la m<sup>ième</sup> interview réalisée durant le trimestre i pour les touristes de type j,
- h<sub>i,j,m</sub>** est la fréquence de passage aux différents points de collecte du touriste de la m<sup>ième</sup> interview réalisée durant le trimestre i pour les touristes de type j, égal à  $k_{i,j,m} + 1$  dans les structures d'hébergement.
- n<sub>i,j,m</sub>** représente le nombre de « compagnons de route » de la m<sup>ième</sup> interview réalisée durant le trimestre i pour les touristes de type j.

En supposant l'interview individuelle mais extensible aux « compagnons de route », chaque questionnaire correspond à une seule personne, représentante de l'ensemble des participants au séjour ; nous pouvons donc définir la moyenne d'une variable type (motivations, moyen de transport, hébergement, durée du séjour et dépenses moyennes) :

$$\bar{X}_{i,j} = \frac{\sum_m x_{i,j,m} \times \gamma_{i,j,m}}{\sum_m \gamma_{i,j,m}}$$

---

<sup>17</sup> Durée du séjour en nuitées pour l'hébergement.

où  $x_{i,j,m}$  est la valeur de la variable X dans le m<sup>ième</sup> questionnaire réalisé durant le trimestre i pour les touristes de type j<sup>18</sup>.

La fréquence de passage sur les différents points de collecte (nombre de nuitées) est la réponse d'un seul individu (agissant comme le représentant de ses compagnons de route), par conséquent, la fréquence moyenne de passage pour chaque sous-population peut être calculée comme une moyenne pondérée (les coefficients correcteurs sont donnés par les  $\gamma_{i,j,m}$ ) :

$$\bar{H}_{i,j} = \frac{\sum_m h_{i,j,m} \times \gamma_{i,j,m}}{\sum_m \gamma_{i,j,m}} = \frac{\sum_m n_{i,j,m}}{\sum_m \frac{n_{i,j,m}}{h_{i,j,m}}}$$

Cette expression n'est rien d'autre que la moyenne harmonique de la fréquence de passage, elle représente la tendance centrale avec une forte sensibilité aux faibles valeurs de la variable.

- **Redressement statistique intra-groupe :**

Les enquêtes de fréquentation opérées dans les différentes structures d'hébergement, permettent la collecte de données relatives à la moyenne des nuitées des touristes (taux d'occupation de chaque strate) :

$$\bar{K}_{i,j} = \frac{\sum_{i,j,m} k_{ij,m} \times \gamma_{i,j,m}}{\sum_{i,j,m} \gamma_{i,j,m}}$$

Le taux net d'occupation de lits (TNO<sub>ij</sub>) relate la capacité potentielle des centres d'hébergement et leur taux d'utilisation effectif :

$$TNO_{i,j} = \frac{PT_{i,j}}{b_{i,j} \times d_{i,j}} \times 100$$

Ce taux nous est donné par le rapport entre le nombre total de nuitées passées par les touristes dans la structure (PT<sub>ij</sub>) et le produit du nombre de lits offerts (b<sub>ij</sub>, capacité d'accueil) et du

---

<sup>18</sup> Si l'interview était réalisée de manière collective par l'ensemble des compagnons de route, nous obtiendrions une équation similaire, à une variante près : les poids  $\gamma_{i,j,m}$  étant remplacés par les poids  $\gamma'_{i,j,m} = \frac{\gamma_{i,j,m}}{n_{i,j,m}}$ .

nombre de jours de leur réelle disponibilité pour un même groupe de structures d'hébergement ( $d_{ij}$ ) ; le résultat est multiplié par 100 pour son expression en pourcentage.

En considérant le transit d'un touriste via plusieurs structures d'hébergement durant le séjour, le nombre de touristes hébergés dans les structures d'hébergement du secteur marchand, sera donné par :

$$NT_{i,j} = \frac{TNO_{i,j} \times b_{i,j} \times d_{i,j}}{K_{i,j}}$$

Pour chaque trimestre et chaque type de structures d'hébergement, nous pouvons donc calculer le nombre de touristes hébergés par les structures collectives.

C'est à partir de ces différentes fréquences qu'un poids est donné à chaque sous-groupe étudié. Ce traitement par strates nous permet non seulement de minimiser les biais (de répétition, durée...) mais aussi de remonter vers l'estimation sans biais de la population d'intérêt.