



Institut national de recherche sur  
les transports et leur sécurité

**Note d'avancement N°2**

**Projet : Mise en perspective des Enquêtes Nationales Transports 1973/74 – 1981/82 –  
1993/94 – 2007/08**

**Subvention Ministère – DRI : N°07 MT S018°:**

**Correction de la non-réponse dans l'Enquête Nationale sur les Transports et les  
Déplacements 2007-08**

**Sophie Roux et Jimmy Armoogum  
INRETS –DEST**



# SOMMAIRE

<b>SOMMAIRE</b> .....	<b>3</b>
<b>LISTE DES TABLEAUX</b> .....	<b>7</b>
<b>LISTE DES GRAPHIQUES</b> .....	<b>8</b>
<b>INTRODUCTION</b> .....	<b>9</b>
<b>CHAPITRE 1. Présentation de l'enquête Nationale sur les transports et les Déplacements 2007-08</b> .....	<b>13</b>
I. Une source de données unique.....	14
a) Les sources de données en « transport de voyageurs » autres que les ENTD .....	14
b) Les enquêtes « transport » à l'étranger .....	15
c) L'Enquête Nationale Transport (ENT) .....	15
II. La méthodologie d'enquête.....	17
a) La collecte des données.....	17
b) Les principaux thèmes abordés par le questionnaire .....	18
III. Le plan de sondage.....	20
a) La construction de l'échantillon des ménages .....	20
Le plan de sondage de « l'échantillon national ».....	20
Le plan de sondage des « extensions régionales » .....	20
Le plan de sondage de « l'extension locale ».....	21
b) Le tirage au sort des « Kish ».....	22
IV. Les nouveautés de la collecte.....	22

<b>CHAPITRE 2. Analyse de la non-réponse.....</b>	<b>24</b>
I. Les taux de réponse sur les différentes parties de l'enquête.....	24
II. Correction de la non-réponse de l'Enquête Nationale sur les Transports et les Déplacements 2007-08.....	25
1) Les trois grands types d'erreur dans les enquêtes par sondage.....	26
a) L'erreur d'échantillonnage.....	26
b) L'erreur d'observation.....	26
c) L'erreur de couverture et de non-réponse.....	27
2) La non-réponse.....	28
a) Définition de la non-réponse.....	28
b) Le taux de non-réponse.....	29
c) Les causes de la non-réponse.....	30
d) Les problèmes liés à la non-réponse.....	31
3) La correction de la non-réponse.....	32
a) Les étapes préalables à la correction de la non-réponse.....	32
b) Les différentes méthodes de correction de la non-réponse.....	32
Les principales méthodes de repondération.....	33
Les principales méthodes d'imputation.....	35
III. Les taux de réponse dans le champ de l'enquête – Logements recensés en 1999.....	37
 <b>CHAPITRE 3. Modélisation du mécanisme de réponse.....</b>	 <b>44</b>
I. Modélisation du mécanisme de réponse de la visite 1.....	44
II. Modélisation du mécanisme de réponse de la visite 2.....	46
III. La correction des erreurs d'échantillonnage et de la non-réponse par calage sur marges.....	50
a) Les variables de calage pour le niveau ménages-individus.....	52
b) Les variables de calage pour le niveau des individus « Kish ».....	55

<b>CONCLUSION</b> .....	<b>60</b>
<b>BIBLIOGRAPHIE</b> .....	<b>62</b>
Annexe 1 : Résultat de la modélisation du mécanisme de réponse de la première visite.....	66
Annexe 2 : Résultat de la modélisation du mécanisme de réponse de la deuxième visite.....	69



## LISTE DES TABLEAUX

Tableau 1: Taux de succès à la première visite .....	24
Tableau 2: Taux de réponse à la deuxième visite .....	24
Tableau 3: Les abandons au cours de l'enquête pour les ménages ayant au moins un véhicule .....	25
Tableau 4: Analyse descriptive des défauts de réponse pour la première visite.....	25
Tableau 5: Non-réponse à diverses enquêtes sur les conditions de vie .....	30
Tableau 6 : Répartition du taux d'échec en visite 1 selon la zone de résidence .....	37
Tableau 7 : Répartition du taux d'échec en visite 1 selon le revenu moyen des ménages de la commune du logement.....	38
Tableau 8 : Répartition du taux d'échec en visite 1 selon l'appartenance de l'immeuble à un organisme HLM.....	38
Tableau 9 : Répartition du taux d'échec en visite 1 selon la période d'achèvement de l'immeuble.....	38
Tableau 10 : Répartition du taux d'échec en visite 1 selon le type de bâtiment.....	39
Tableau 11 : Répartition du taux d'échec en visite 1 selon la présence d'un digicode ou d'un interphone.....	39
Tableau 12 : Répartition du taux d'échec en visite 1 selon le nombre d'ascenseur .....	39
Tableau 13 : Répartition du taux d'échec en visite 1 selon le nombre de pièces du logement.....	39
Tableau 14 : Répartition du taux d'échec en visite 1 selon la surface du logement.....	40
Tableau 15 : Répartition du taux d'échec en visite 1 selon le nombre de personnes dans le ménage.....	40
Tableau 16 : Répartition du taux d'échec en visite 1 selon l'âge de la personne de référence.....	40
Tableau 17 : Répartition du taux d'échec en visite 1 selon le sexe de la personne de référence.....	40
Tableau 18 : Répartition du taux d'échec en visite 1 selon la motorisation du ménage.....	41
Tableau 19 : Répartition du taux d'échec en visite 1 selon la position professionnelle de la personne de référence à la date du RP99.....	41
Tableau 20 : Répartition du taux d'échec en visite 2 selon la zone de résidence .....	47
Tableau 21 : Répartition du taux d'échec en visite 2 selon le type de ménage .....	48
Tableau 22 : Répartition du taux d'échec en visite 2 selon le revenu du ménage .....	48
Tableau 23 : Répartition du taux d'échec en visite 2 selon le sexe de l'individu « Kish » .....	48
Tableau 24 : Répartition du taux d'échec en visite 2 selon le nombre de kilomètres parcourus cours des 12 derniers mois (de la première voiture du ménage).....	48
Tableau 25 : Répartition du taux d'échec en visite 2 selon le nombre de voyages de longue distance de l'individu « Kish » au cours des 3 derniers mois .....	49
Tableau 26 : Répartition du taux d'échec en visite 2 selon la durée entre la date de la première visite et la date de la fin vague.....	49

## **LISTE DES GRAPHIQUES**

Graphique 1 : Schéma de la non-réponse totale et de la non-réponse partielle .....	28
Graphique 2 : Nombre d'enquêtes réalisées par semaine au cours de l'année .....	51

## INTRODUCTION

Quelles que soient les méthodologies d'enquêtes mises en place pour collecter des réponses auprès des ménages, individus ou entreprises, l'information collectée n'est jamais complète ou entièrement satisfaisante en grande partie à cause de la non-réponse. Même si l'enquête est obligatoire, certains ménages sont difficiles à joindre ou refusent de répondre. L'accès des logements à l'enquêteur est aujourd'hui rendu de plus en plus difficile car les ménages s'équipent davantage en interphone ou digicode. Une partie des refus peut également s'expliquer par l'augmentation du nombre d'enquêtes réalisées ces dernières années, notamment à but marketing. La demande croissante d'informations des instituts de sondage peut constituer un frein à l'acceptation de tous types d'enquêtes par les enquêtés. Quand le contact est établi et l'enquête acceptée, tous les questionnaires ne sont pas toujours remplis et retournés. Et même dans un questionnaire rempli, notamment lorsque les questions sont trop complexes, certaines informations peuvent manquer ou s'avérer imprécises (par exemple arrondies), inexactes ou incohérentes.

L'Enquête Nationale sur les Transports et les Déplacements (ENTD) 2007-08, menée en collaboration par l'Institut National de la Statistique et des Études Économiques (INSEE), le Ministère de l'Écologie, de l'Énergie, du Développement Durable et de l'Aménagement du Territoire (MEEDDAT) et l'Institut National de REcherche sur les Transports et leur Sécurité (INRETS), ne fait pas exception à la règle. En effet, comme toutes les enquêtes probabilistes, elle est entachée de non-réponse. D'un point de vue méthodologique, l'ENTD est complexe :

- outre la description du ménage et de ses membres, on interroge une personne sélectionnée à l'aide de la méthode 'Kish', sur ses déplacements quotidiens et ses voyages à longues distances ;
- elle comporte un carnet de trajets à remplir.

La non-réponse est, bien entendu, présente à tous les niveaux de l'enquête. Elle induit une perte d'information conduisant à une diminution de la précision des estimateurs et une apparition de biais à l'enquête.

L'objectif de ce rapport intermédiaire est de parcourir l'éventail des erreurs dans une enquête par sondage en partant de la non-réponse totale, pour aller vers les plus ponctuelles en l'occurrence la non-réponse partielle et les erreurs de mesures des personnes enquêtées. Dans chaque cas, nous montrerons comment remédier à ces lacunes de l'information collectée.

Nous définirons, au fur et à mesure que nous avancerons, toutes les catégories de non-réponse, proposerons des solutions pour pondérer l'Enquête Nationale sur les Transports et les Déplacements 2007-08 et enfin, appliquerons la méthode d'imputation la mieux adaptée pour corriger la non-réponse partielle et les erreurs de mesure.

Le premier chapitre proposera un rappel des objectifs de l'Enquête Nationale sur les Transports et les Déplacements 2007-08. Le plan de sondage, le questionnaire et les innovations de l'enquête seront également abordés ainsi que la méthodologie mise en place pour collecter les données.

La meilleure méthode pour corriger le défaut de réponse est, bien évidemment, d'essayer d'éviter les non-réponses ou du moins d'en avoir « le moins possible » tout en gardant une excellente qualité des données. Malgré tous les efforts fournis pour éviter la non-réponse, il en subsiste toujours après la phase de la collecte des données. Il est donc nécessaire de pouvoir l'identifier, de connaître les problèmes qu'elle peut engendrer et dans une dernière étape de la corriger avec des méthodes adaptées. Le second chapitre présentera les taux de réponse à l'enquête. Nous passerons également en revue les différentes méthodes permettant de corriger la non-réponse et les erreurs de mesure.

Il faut tenir compte de la non-réponse lors de l'inférence puisque les répondants et les non-répondants ont généralement un comportement différent.

Le troisième chapitre portera sur la correction des erreurs de non-réponse totale et d'échantillonnage à l'aide d'une repondération. La correction de ces erreurs par un calage sur des distributions connues dans la population implique que nous connaissions auparavant les caractéristiques des non-répondants pour limiter les problèmes de biais de l'enquête. Une analyse des variables discriminantes du mécanisme de réponse à la première et à la seconde visite sera donc étudiée. Ce n'est qu'une fois les comportements analysés que nous pourrons réaliser le calage à l'aide de "Calmar", logiciel développé à l'INSEE par O. Sautory. Au vu de l'avancée de

l'enquête, cette dernière partie ne sera abordée que du point de vue de la méthodologie et du choix de la méthode à mettre en œuvre.



## **CHAPITRE 1. Présentation de l'enquête Nationale sur les transports et les Déplacements 2007-08**

Tous les dix ans l'INSEE entreprend avec le Ministère de l'Ecologie, de l'Energie, du Développement Durable et de l'Aménagement du Territoire (MEEDDAT) une Enquête Nationale sur les Transports (ENT). Dans les années cinquante, le champ de l'enquête était restreint à l'automobile (enquête « Parc et utilisation de l'automobile », CREDOC, 1958-59). Depuis les années soixante, l'enquête s'emploie à décrire l'ensemble des déplacements des ménages. L'enquête Nationale sur les Transports et les Déplacements 2007-08 (ENTD 2007-08) est la cinquième de la lignée des enquêtes nationales transports (1966-67, 1973-74, 1981-82, 1993-94) réalisées sur le même sujet avec des objectifs ayant peu varié. L'Institut National de Recherche sur les Transports et leur Sécurité (INRETS) est le coordinateur scientifique de cette enquête depuis l'enquête de 1981-82. Son Département Economie et Sociologie des Transports (DEST) s'est fortement investi dans la préparation de l'ENTD 2007-08. Il joue un rôle central dans la collecte des données puisque la division Méthodologie des Grandes Enquêtes (MGE) a travaillé à la fois sur la conception du questionnaire, la correction de la non-réponse et le redressement de l'enquête de façon à la rendre représentative de l'ensemble de la population française métropolitaine. Le coût de l'ENTD 2007-08 est estimé à 2,5 millions d'euros. Elle est financée à la fois par des organismes publics et privés (Agence de l'Environnement et de la Maîtrise de l'Energie, Direction Générale de l'Aviation Civile, EDF, SNCF, Réseau Ferré de France, Association des Sociétés Françaises d'Autoroutes et d'Ouvrages à Péage, Comité des Constructeurs Français d'Automobiles, association des compagnies d'assurances, etc... ).

L'objectif de cette enquête est la connaissance des déplacements des ménages résidant en France et de leur usage des moyens de transport tant collectifs qu'individuels. L'ENTD 2007-08 s'inscrit dans le prolongement des précédentes enquêtes transports afin d'assurer la mesure des évolutions structurelles.

Les principaux thèmes abordés sont les déplacements, de courtes et de longues distances des ménages et de leurs membres, selon les différents modes de transports. L'enquête met également l'accent sur la connaissance du parc de véhicules (voitures, motos, vélos, quads), de leur usage et

l'accessibilité des individus aux transports collectifs (dont les abonnements et les réductions tarifaires).

L'Enquête Nationale sur les Transports est la seule sur la mobilité réalisée à cette échelle et qui décrit tous les déplacements, quels que soient le motif, la longueur, la durée, le mode de transport, la période de l'année ou le moment de la journée. Elle s'intéresse aussi aux possibilités d'accès aux transports collectifs et aux moyens de transports individuels dont disposent les ménages pour comprendre leurs comportements.

## **I. Une source de données unique**

L'Enquête Nationale sur les Transports et les Déplacements est la seule source de données fournissant une vision globale et cohérente de tous les modes et situations de transport de personnes (articulation mobilité quotidienne et longue distance) sur l'ensemble du territoire français.

### ***a) Les sources de données en « transport de voyageurs » autres que les ENTD***

Les « Enquêtes Ménages Déplacements (EMD) » (méthodologie du CERTU) permettent de recueillir les pratiques globales de déplacements d'une population urbaine et, par la connaissance de ces pratiques, elles constituent un outil indispensable pour élaborer et évaluer les politiques de transport dans les agglomérations. Ce sont des enquêtes réalisées auprès des ménages qui permettent d'avoir une « photographie » des déplacements réalisés par les habitants d'un périmètre donné, un jour moyen de semaine en automne ou en hiver, hors vacances scolaires, c'est-à-dire lorsque le volume de trafic est maximal. Elles ne recueillent que des pratiques globales de déplacements. Or, même si les territoires d'enquête tendent à s'élargir, ce dispositif ne permet pas de mesurer la mobilité en dehors des zones agglomérées (en périphérie et en zone rurale).

Les enquêtes « Suivi de la Demande Touristique (SDT) » (mise en place par le Service Economie Statistiques et Prospective du MEEDDAT) comptabilisent les voyages à plus de 100 kilomètres à vol d'oiseau du domicile. Ce sont des enquêtes réalisées auprès d'un panel dont l'objectif est la connaissance du tourisme et du suivi de la mobilité des personnes à longues distances. Les enquêtés sont interrogés sur leurs déplacements d'au moins une nuitée en dehors du domicile (interrogation mensuelle) et sur leurs déplacements dans la journée à plus de 100 kilomètres du domicile (interrogation trimestrielle).

L'enquête nationale transport est donc une source de données unique, complète et irremplaçable, à ce jour, dans les statistiques françaises en termes d'informations recueillies sur la mobilité des passagers.

### ***b) Les enquêtes « transport » à l'étranger***

On recense une douzaine de pays qui mènent de façon relativement régulière des enquêtes « transport » comparables aux ENT (dont Allemagne, Royaume-Uni, pays scandinaves, Suisse, Etats-Unis) (BONNEL, ARMOOGUM 2005). La périodicité de ces enquêtes est assez variable entre les pays. L'enquête est réalisée en continu au Royaume-Uni, aux Pays-Bas, en Suède et en Norvège et elle l'a été jusqu'en 2003 au Danemark. Ailleurs, l'intervalle est d'environ 7 ans. Les tailles d'échantillon sont généralement de l'ordre d'une dizaine de milliers de ménages (mais atteignent 66 000 aux Etats-Unis, pour l'enquête menée en 2001).

Les modalités d'enquête sont relativement comparables. Les systèmes assistés par ordinateur CATI/CAPI ou mixtes ont pris désormais la relève de la collecte classique par questionnaire papier. En général les enquêtés sont interrogés sur leurs déplacements de la veille. Pour la longue distance, lorsque ce type de déplacement est mesuré, la période oscille entre 2 semaines (Autriche, 1995) et 3 mois (France, 2007-08). Les Etats-Unis envisagent un questionnement assisté par récepteurs GPS pour leur enquête transport de 2007 sur un sous-échantillon de 10%.

Jusqu'à présent, les différents dispositifs nationaux restent indépendants les uns des autres. Souvent les définitions, même pour les phénomènes de base, diffèrent. Par exemple, un déplacement est défini comme à longue distance s'il est de plus de 50 miles dans l'enquête britannique et s'il comprend une nuit, hors domicile, dans l'enquête suisse. Eurostat n'a pour l'instant pas engagé de programme d'harmonisation et de coordination. De ce fait, les comparaisons entre enquêtes restent difficiles à mener.

### ***c) L'Enquête Nationale Transport (ENT)***

L'Enquête Nationale Transport est le seul outil qui permette d'observer les comportements des habitants de toutes les tailles d'agglomérations (notamment les plus petites) et d'estimer des indicateurs de mobilité tout au long de l'année. L'enquête est notamment fiable pour

l'appréhension des distances parcourues (véhicules\*kilomètres ou voyageurs\*kilomètres) ce qui est important pour les problématiques environnementales (émission de polluants locaux ou globaux, consommation d'énergie). Elle permet des comparaisons dans le temps (avec les enquêtes précédentes) et dans l'espace (avec les enquêtes menées à l'échelon local ou dans d'autres pays).

L'Enquête Nationale Transport permet également d'analyser de façon fine et précise le parc des véhicules automobiles dont disposent les ménages et leurs usages, ce qui est primordial pour un moyen de transport utilisé dans plus de 60% des déplacements, mais moins bien connu que d'autres. En effet, pour la circulation automobile, il n'existe pas de statistiques globales équivalentes à celles fournies par les entreprises exploitantes des secteurs ferroviaires et aériens et on ne dispose de comptages de véhicules et d'estimation des véhicules-kilomètres réalisés que sur le seul réseau routier national (composé des autoroutes et routes nationales). Le relevé du kilométrage (sur les fiches véhicules) réalisé sur les différents réseaux routiers sert donc, par exemple, de base aux études sur la mesure de l'exposition au risque d'accident de la route pour des groupes de population.

L'Enquête Nationale Transport constitue une base précieuse pour répondre aux nombreuses questions sur les trafics inter-régionaux et internationaux dont les enjeux sont particulièrement importants en matière d'investissements (autoroutes, TGV, transports aériens).

Comme les précédentes ENT, l'enquête 2007-08 est organisée autour des trois grands thèmes suivants :

- Description des déplacements :
  - La mobilité régulière : les déplacements habituels domicile-travail, domicile-lieu d'étude ou de garde des enfants.
  - La mobilité locale : les déplacements réalisés à l'occasion d'activités situées dans un rayon de 80 kilomètres autour du domicile ;
  - La mobilité à longues distances : les déplacements réalisés à plus de 80 kilomètres du lieu de sa résidence principale.
- Connaissance du parc de véhicules et de leurs usages :
  - Description des véhicules dont dispose le ménage et historique du parc du ménage (sur 3 ans) ;
  - Description des usages d'un véhicule pendant une semaine ;

- Détention de permis de conduire, pratique de la conduite et accidents de la circulation.
- Accessibilité aux transports collectifs (dont abonnements et réductions tarifaires).

Signalons aussi que cette enquête est utilisée pour des études sur une large gamme de sujets (concurrence entre modes de transport, spécialisation en fonction des motifs, déplacements régionaux dans l'aire d'attraction des grandes villes, consommations d'énergie et environnement, déplacements touristiques, ...).

## **II. La méthodologie d'enquête**

### *a) La collecte des données*

La phase de collecte des données de l'ENTD 2007-08 a été réalisée du 30 mars 2007 au 27 avril 2008 auprès des ménages ordinaires de France métropolitaine. Elle s'est déroulée en six vagues successives (d'environ huit semaines chacune) sur tout le territoire métropolitain français de façon à éliminer les effets de saisonnalité (voyage à plus de 100 kms plus nombreux en été, etc ...) qui affectent la mobilité, tout au moins privée, des ménages. Le ménage reçoit l'enquêteur au cours de deux visites espacées de huit jours. L'enquête est une interview en face-à-face d'une durée moyenne d'environ 2 heures. Les ménages sont interrogés dans leur résidence principale.

La première visite, orientée sur les facteurs explicatifs du comportement de la mobilité, autorise un proxy, c'est-à-dire qu'un membre du ménage peut répondre à la place d'un autre membre absent lors de la visite de l'enquêteur, mais le proxy n'est pas souhaité car il a tendance à amoindrir la qualité des réponses (voir encadré 1).

A la fin de cette visite, un tirage au sort est effectué pour désigner, d'une part, l'individu qui répondra aux questionnaires de la 2<sup>nd</sup> visite et, d'autre part, un véhicule. Le ou les utilisateurs de ce dernier doivent noter sur le carnet-véhicule tous les déplacements réalisés avec ce véhicule pendant sept jours consécutifs. Dans certains cas, l'individu tiré au sort, s'il est majeur et volontaire, reçoit un récepteur GPS à utiliser lors de tous ses déplacements au cours des sept prochains jours.

### **Encadré 1 : le proxy aurait tendance à amoindrir la qualité des réponses**

L'exploitation des bilans de collecte des enquêteurs après la vague 1 de l'ENTD 2007-08 laisse penser que les réponses fournies par le proxy ne sont pas fiables (ROUX, 2007). Aux questions « Selon vous [*les enquêteurs*], est-ce que les réponses du proxy sont moins fiables ? Si oui, quels sont les modules et/ou les questions les plus affectés ? » 94% des enquêteurs à avoir eu recours à des proxys (près de 2/3, généralement en visite 1), notent que, parfois, le proxy ne connaît pas les réponses, notamment aux questions sur les déplacements réguliers (temps et distance du trajet domicile-travail ou école, environnement du lieu de travail), les véhicules à décrire lorsque l'utilisateur n'est pas l'utilisateur principal (consommation en ville et sur route, kilométrages au cours des 12 derniers mois), mais aussi celles sur la date d'obtention du permis de conduire, voire les diplômes demandés au questionnaire du tronc commun des ménages, soit, en général, toutes questions qui demandent de la précision.

La seconde visite, orientée sur la description des comportements, n'autorise pas le proxy, sauf dans le cas d'enfant de moins de 12 ans ou de personne inapte à répondre. Au cours de cette visite, le carnet véhicule ainsi que le GPS sont récupérés par l'enquêteur. Un module de questions supplémentaires est alors posé aux utilisateurs du GPS pour connaître les motifs et les moyens utilisés pour les déplacements et dans le cas contraire, les raisons de l'immobilité.

#### ***b) Les principaux thèmes abordés par le questionnaire***

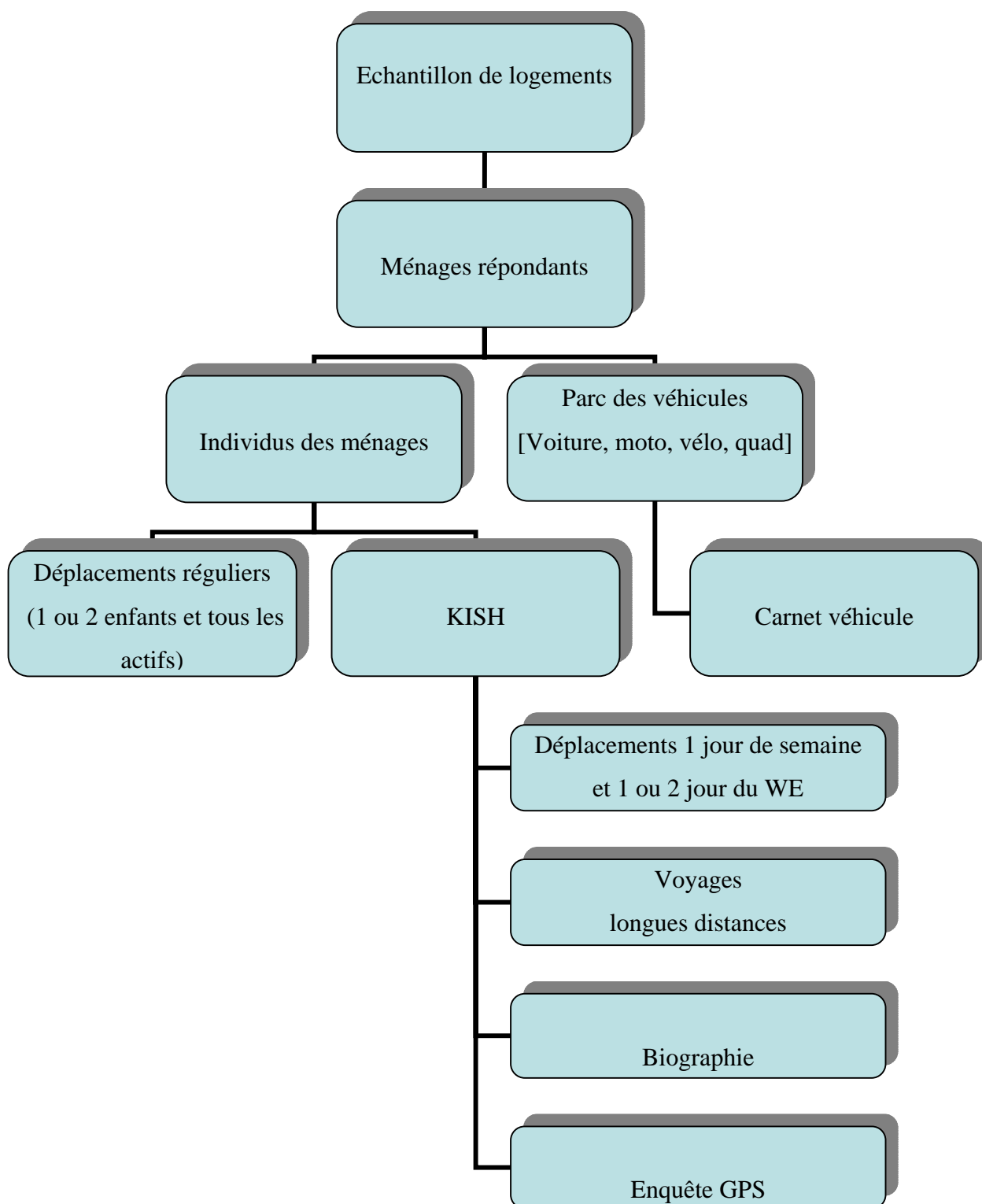
Lors de la visite initiale, l'enquête aborde, au niveau du ménage ou de l'ensemble de ses membres, les thèmes suivants :

- la description des caractéristiques socio-démographiques des individus qui composent le ménage ;
- les déplacements réguliers domicile-travail, domicile-lieu d'étude, domicile-lieu de garde des enfants ;
- le permis de conduire et pratique de la conduite, accidents de la circulation ;
- les abonnements et réductions dans les transports collectifs ;
- la description des véhicules dont dispose le ménage;
- l'environnement du logement.

Lors de la seconde visite, l'individu tiré au sort (appelé individu « Kish ») répond aux questions portant sur la description :

- de ses déplacements de la veille (le jour de semaine à décrire est le jour le plus proche de la deuxième visite durant lequel le « Kish » est sorti de chez lui) et du dernier week-end (seule une journée de week-end est décrite, sauf dans deux régions à extension : Ile de France et Pays de Loire. L'enquêté « Kish » répond à des questions sur un jour où il a été mobile, le samedi ou le dimanche, avec une probabilité égale de tirage s'il n'est pas resté chez lui l'un de ces deux jours) ;
- de ses déplacements à longues distances pendant les trois derniers mois précédents la date de l'enquête. La période d'observation est réduite à deux mois s'il y a plus de dix voyages à décrire. Cette description des voyages fait appel à une mémoire des enquêtés plus lointaine qui peut être source d'oubli ou d'erreur ;
- des modes de transports employés tout au long de sa vie à l'aide d'une grille biographique (seulement en vague 4 et 5).

**Encadré 2 : Architecture de l'Enquête Nationale sur les Transports et les Déplacements  
2007-08**



Source : INSEE Enquête Nationale sur les Transports et les Déplacements 2007-08

### **III. Le plan de sondage**

#### ***a) La construction de l'échantillon des ménages***

Le champ de l'enquête est celui des déplacements et du parc automobile des résidents de France métropolitaine. Les unités enquêtées sont les unités de vie telles qu'elles sont définies dans le tableau de composition du ménage (TCM) des enquêtes ménages de l'Insee, c'est-à-dire l'ensemble des personnes ayant une résidence habituelle commune et partageant un même budget.

Le plan de sondage de l'ENTD 2007-08 est réalisé en deux phases afin de collecter le plus d'informations possibles sur les déplacements. Un plan en deux phases est un échantillon issu d'un échantillon de la population U. La première phase consiste à tirer un échantillon dans la population U, et la deuxième consiste à tirer un échantillon à partir de l'échantillon de la première phase.

#### ***Le plan de sondage de « l'échantillon national »***

Au niveau national, l'échantillon de première phase de l'ENTD 2007-08 provient de l'échantillon-maître issu du recensement de la population de 1999 (RP 99), complété par la base de sondage des logements neufs achevés depuis le RP 99. Les ménages multi-motorisés étant les plus mobiles, ces derniers ont été sur-représentés lors de la seconde phase du plan de sondage, de même que ceux habitant dans des communes rurales (zone non-couverte dans les Enquêtes Ménages Déplacements).

En considérant un échantillon de deuxième phase pour le niveau national de 17 500 logements cela nous donne un taux de sondage moyen de 0,056%.

#### ***Le plan de sondage des « extensions régionales »***

En plus de l'échantillon national, cinq régions ont opté pour une extension régionale (Bretagne, Île-de-France, Midi-Pyrénées, Languedoc-Roussillon et Pays de la Loire). Les coefficients de sur-représentation sont identiques à ceux du plan de sondage national. Ainsi, les extensions régionales ont ajouté 10 308 logements à l'échantillon national avec la décomposition régionale suivante :

- 686 logements en Bretagne ;
- 7 002 logements en Ile de France ;
- 836 logements en Midi-Pyrénées ;
- 1 242 logements en Languedoc-Roussillon ;
- 542 logements en Pays de la Loire.

En pratique, l'échantillon national peut être complété avec les échantillons régionaux afin d'obtenir un échantillon représentatif plus conséquent.

### *Le plan de sondage de « l'extension locale »*

Le département de Loire-Atlantique souhaitait obtenir des estimations de mobilité pour les sous-populations suivantes :

- Communauté urbaine de Nantes (CUN)
- Schéma de cohérence et d'organisation du territoire (SCOT) de Nantes-Saint Nazaire
- Division territoriale d'aménagement (DTA)
- Le département de Loire-Atlantique (D44)

Ces territoires sont emboîtés, ce qui facilite la stratification de ces zones. Ainsi le pôle inter-régional d'ingénierie statistique ménages de l'Insee a tiré 2 066 logements supplémentaires par rapport aux tirages national et régional, dans l'ensemble du département de Loire-Atlantique (en visant 1 000 ménages répondant dans la CUN, en incluant tous les échantillons tirés (part de l'échantillon national tiré dans l'échantillon-maître et situé dans la CUN, part de l'échantillon EMEX appartenant à la CUN, échantillon complémentaire local).

La base de sondage était constituée du RP de 1999 et de la base de logements neufs (BLN) mise à jour en mai 2006, c'est-à-dire incluant les logements achevés à la fin du 1<sup>er</sup> trimestre 2006. L'échantillon a été réparti entre les deux au prorata du nombre de logements dans chaque base à l'intérieur de chaque zone.

En partant d'une allocation fixe dans la CUN, imposée par la DR et ses partenaires financiers, le reste de l'échantillon a été réparti dans les autres zones proportionnellement à leur population en nombre de logements.

- a) Dans les zones CUN et SCOT hors CUN, l'échantillon tiré dans le RP99 a été stratifié sur le croisement des variables : catégorie du logement en 1999 (principal+vacant, secondaire+occasionnel), taille du ménage de 1999 en 3 modalités et nombre de voitures du ménage. Le tirage est un sondage aléatoire simple dans les strates.

- b) Dans les autres zones (DTA hors SCOT, département hors DTA), l'échantillon a été stratifié par catégorie de logement en 1999 et nombre de voitures du ménage. On a fait un tirage systématique dans les strates, contrôlé sur la zone (DTA ou hors DTA), la taille de commune et le nombre de voitures du ménage de 1999.

Le pôle inter-régional d'ingénierie statistique ménages de l'Insee a appliqué le même principe de sur-représentation des ménages à 2 voitures et de sous-représentation des résidences secondaires ou occasionnelles en 1999 que dans l'échantillon national.

Au total, 30 165 fiches adresses ont été tirées de la base de sondage, dont 12 500 dans des échantillons de réserve. Au total, 20 267 ménages ont été interrogés.

#### ***b) Le tirage au sort des « Kish »***

En fin de 1<sup>ère</sup> visite, un individu du ménage (appelé individu « Kish ») est tiré au sort pour répondre aux questionnaires de la 2<sup>nd</sup> visite parmi les personnes éligibles du ménage (individus de plus de 5 ans, présents au moment de la 2<sup>nd</sup> visite et aptes à répondre). Ce tirage est à probabilités inégales afin d'interroger de préférence la personne qui fait le plus de voyages à longues distances.

Un second tirage est effectué pour désigner le véhicule (appelé véhicule « Kish ») qui recevra un carnet (appelé carnet-véhicule) à l'intérieur duquel tous les déplacements effectués avec ce véhicule seront consignés. Le tirage vise à privilégier d'abord les deux roues à moteur pour lesquels les informations sont rares, puis les vélos, puis les autres véhicules à moteur, sous la condition qu'ils aient été utilisés au cours du dernier mois. Les véhicules qui n'ont pas été utilisés au cours des quatre dernières semaines ont une probabilité plus faible d'être tirés.

### **IV. Les nouveautés de la collecte**

L'édition 2007-08 de l'ENTD se différencie des précédentes par la méthodologie mise en œuvre pour collecter les données et dont l'avantage principal est d'augmenter la qualité de l'enquête.

L'enquête a pu bénéficier, pour la première fois, du statut « obligatoire » délivré par la Commission Nationale Informatique et Liberté (CNIL), à l'exception de l'opération complémentaire impliquant l'usage d'un GPS, qui ne s'adresse qu'à des volontaires. Tous les ménages échantillonnés doivent donc obligatoirement répondre à l'enquête. Dans les faits, tous les enquêtés ne s'y soumettent pas mais cette mention permet d'obtenir des taux de réponses plus

élevés, et donc une meilleure qualité de l'enquête (BERTHIER, DUPONT, 1999). Deuxièmement, la collecte n'a pas été réalisée sur un support papier mais à l'aide d'un CAPI (Collecte Assistée Par Informatique). L'enquêteur rentre donc directement les réponses de l'enquêté sur son micro-ordinateur. Cette méthode de collecte permet de supprimer l'étape de la saisie des réponses du format papier vers un format numérique et ainsi de limiter les erreurs de codification. Le procédé CAPI utilisé pour la première fois par l'INSEE en 1992 pour l'enquête Emploi (depuis 1996 pour les enquêtes auprès des ménages) permet une automatisation des tests de cohérence mais n'a pas l'avantage, contrairement à ce que l'on pourrait penser, de diminuer significativement le délai entre collecte et disponibilité des résultats (GLAUDE, 2000).

Cette édition innove des précédentes par son utilisation d'outils de recueils dits « non traditionnels » aux ENT. L'emploi des nouvelles technologies, ainsi que de méthodes généralement réservées à la démographie ont permis de collecter toujours plus d'informations. En effet, parallèlement à la collecte informatisée, le suivi des déplacements de certains individus s'est fait à l'aide d'un récepteur GPS. Cette utilisation combinée au questionnaire « classique » permet de corriger l'oubli de certains trajets, et aussi de connaître avec une meilleure précision les heures de départ et les temps de transport renseignés. A noter que l'utilisation d'un récepteur GPS pour suivre les déplacements des individus est une première mondiale dans les enquêtes nationales transports.

Un volet rétrospectif a également été ajouté pour analyser les comportements passés des individus tout au long de leur vie ainsi que les changements survenus. Le recueil de ces informations est rendu possible par l'utilisation d'un questionnaire biographique. A l'aide d'une grille, les enquêtés récapitulent les changements intervenus dans leurs localisations, leurs moyens de transport et leurs pratiques de déplacement suivant les situations de leur vie. Les informations collectées portent sur la résidence principale, le nombre de personnes du ménage, le nombre de véhicules à moteur, l'activité et le moyen de transport habituel du déplacement. A partir de ces renseignements, il sera possible de comprendre comment ont changé les habitudes de transport suivant le lieu de résidence, l'activité et les véhicules à disposition et aussi de mieux connaître les façons de se déplacer en France dans le passé pour des périodes où aucune enquête n'existe, uniquement à partir d'un effort de mémoire des enquêtés (bien qu'il faudra relativiser les données en vue des erreurs de mémoire).

## CHAPITRE 2. Analyse de la non-réponse

L'information collectée lors d'une enquête n'est jamais complète. Même si, comme dans le cas de l'Enquête Nationale sur les Transports et les Déplacements 2007-08, l'enquête est obligatoire, certains ménages sont difficiles à joindre ou refusent de répondre. Quand le contact est établi, tous les questionnaires ne sont pas toujours remplis et retournés.

L'Enquête Nationale sur les Transports et les Déplacements 2007-08 n'échappe pas à la règle, la non-réponse est présente dans chaque partie du questionnaire. Ce chapitre présentera les taux de réponse des différentes parties de l'ENTD 2007-08, puis se focalisera sur la correction de la non-réponse.

*Les taux de réponse à l'enquête sont calculés en fonction des modalités de la variable RES\_MEN « résultat de la collecte des ménages », elle-même calculée directement par le logiciel CAPI selon certaines variables remplies par les enquêteurs. Les modalités prises en compte sont « enquête réalisée partiellement (ERP) » et « enquête réalisée entièrement, validée (ERV) ».*

### **I. Les taux de réponse sur les différentes parties de l'enquête**

**Tableau 1: Taux de succès à la première visite**

Fiches Etablies	Hors Champ	Résidences principales	Echecs par rapport à la 1 <sup>ère</sup> visite	Enquêtes réalisées 1 <sup>ère</sup> visite
30 165	4 272	25 893	5 626	20 267
100%	14,2%	85,8%	18,7%	67,2%

Source : INSEE Enquête Nationale sur les Transports et les Déplacements 2007-08

Le taux de réponse à la première visite est de : 78,3%

**Tableau 2: Taux de réponse à la deuxième visite**

Enquêtes réalisées 1 <sup>ère</sup> visite	Echecs par rapport à la 2 <sup>nd</sup> visite	Enquêtes réalisées 2 <sup>nd</sup> visite
20 267	1 627	18 648
100%	8,0%	92,0%

Source : INSEE Enquête Nationale sur les Transports et les Déplacements 2007-08

Entre la première et la deuxième visite, nous avons perdu 8% des enquêtés de la première visite, ce qui donne un taux de réponse globale pour l'enquête de : 72,0%.

**Tableau 3: Les abandons au cours de l'enquête pour les ménages ayant au moins un véhicule**

Nombre de ménage avec un véhicule	Nombre total de fiches véhicules	Nombre total de fiches décrites
17 586	33 609	33 596
86,7%	100,0%	99,9%

Source : INSEE Enquête Nationale sur les Transports et les Déplacements 2007-08

Seulement 13 fiches véhicules n'ont pas été remplies (il peut s'agir d'une erreur de saisie du nombre de véhicules, générant des fiches impossible à compléter).

**Tableau 4: Analyse descriptive des défauts de réponse pour la première visite**

Ménage	Logement neuf	Recensement de 1999	Total
Enquête réalisée	1 573	18 694	20 267
Echec dans le champ	370	5 256	5 626
Total	1 943	23 950	25 893
Taux de réponse	81,0%	78,1%	78,3%

Source : INSEE Enquête Nationale sur les Transports et les Déplacements 2007-08

Les ménages habitant un logement construit depuis 1999 répondent un peu plus souvent que les autres, mais on ne peut pas mieux les caractériser puisqu'on ne dispose pas d'informations autres que géographiques sur ces logements (cf. tableau).

## **II. Correction de la non-réponse de l'Enquête Nationale sur les Transports et les Déplacements 2007-08**

L'objectif premier de la technique des sondages est de « dégager des conclusions pour un ensemble (appelé univers) à partir d'informations qui ne sont connues que pour un sous-ensemble de l'univers, appelé échantillon »<sup>1</sup>. Cette technique faisant appel au calcul des probabilités a des inconvénients, telles que les erreurs d'échantillonnage, de non-réponse et de mesure qu'il est possible de corriger.

Il est donc nécessaire de pouvoir identifier ces erreurs, de connaître les problèmes qu'elles peuvent engendrer et de les corriger avec des méthodes adaptées.

---

<sup>1</sup> Encyclopédie Universalis

## **1) Les trois grands types d'erreur dans les enquêtes par sondage**

L'enquête par sondage est source d'erreurs multiples qu'il faut pouvoir mesurer. On distingue principalement trois types d'erreurs : l'erreur d'échantillonnage, l'erreur d'observation, et l'erreur de couverture et de non-réponse (ARDILLY, 2006).

### ***a) L'erreur d'échantillonnage***

Le choix de l'échantillon dans une enquête par sondage est primordial car son utilisation comporte un inconvénient majeur, celui de se retrouver avec des estimations sujettes à la variabilité d'échantillonnage. Les erreurs d'échantillonnage proviennent du fait que l'information désirée n'est observée que pour une partie de la population, il y a donc une perte de l'exactitude même si cette dernière est volontaire et planifiée. On estime l'erreur d'échantillonnage par la différence entre l'estimation calculée à partir d'une enquête par échantillon et la « vraie » valeur qui aurait été obtenue si un recensement auprès de la population entière avait été effectué dans les mêmes conditions.

### ***b) L'erreur d'observation***

Les erreurs d'observation, appelées aussi erreurs de mesure, correspondent à une information collectée qui diffère de la vraie valeur. Ce type d'anomalie est commun à toutes les enquêtes et est d'autant plus important à mesure que le sujet d'étude est dit sensible. En effet, les enquêtés auraient tendance à se valoriser vis-à-vis des enquêteurs et donc à fournir des réponses inexactes à des études délicates telles que les revenus, la sexualité ou encore la consommation de drogue. Par exemple, Caron (1993) a mis en évidence les erreurs de mesure de la variable « revenu » à partir de l'enquête de Conjoncture auprès des ménages. Pour ce sondage, chaque ménage est interrogé deux fois en un temps relativement court, et répond aux mêmes questions lors des deux visites. L'auteur a montré que seul 53,4% des réponses des ménages sur la variable « revenu » concordait d'une enquête à l'autre. Outre l'erreur de mesure, l'hypothèse que ces erreurs puissent venir d'un mauvais codage des réponses de l'enquêté n'est pas à écarter.

Les causes d'erreurs d'observation ne sont pas exclusivement volontaires, elles relèvent également d'une origine involontaire. On peut citer sans être exhaustif selon Ardilly (2006) :

- Les erreurs de bonne foi de l'enquêté lors par exemple d'une enquête de type rétrospective qui oblige le répondant à faire appel à sa mémoire ;

- Les erreurs introduites par l'enquêteur si ce dernier influence la réponse de l'enquêté (ce type d'erreur révèle également une anomalie dans la formation de l'enquêteur) ;
- Les erreurs de saisie de l'enquêteur (en se trompant de touche sur le clavier) ou d'écriture de l'enquêté (en se trompant de case en cochant) ;
- Une mauvaise compréhension de la question par l'enquêté ou une question mal formulée, ambiguë ;
- L'individu échantillonné a été remplacé par erreur ou volontairement par un autre individu.

La notion de « vraie » valeur n'est pas toujours évidente à définir ou à estimer et dépend pour beaucoup des variables étudiées. Il serait plus facile de rechercher la valeur exacte d'une date de naissance que celle d'un revenu. Toutefois, certains types d'erreur de mesure peuvent être observés à l'aide des méthodes de contrôles de cohérences des données. Une fois l'erreur de mesure repérée, la réponse peut être invalidée et être transformée en valeur manquante. Ceci créera donc « artificiellement » de la non-réponse partielle.

### *c) L'erreur de couverture et de non-réponse*

L'erreur de couverture de la base de sondage survient lorsque celle-ci ne couvre pas toute la population cible. Un exemple serait un sondage réalisé à l'aide d'une base de sondage issue de l'annuaire téléphonique (téléphone fixe) alors que la population cible serait la population des ménages (exclusion des ménages sans téléphone fixe). Il est relativement difficile de mesurer un défaut de couverture puisque l'on ne peut pas dénombrer l'effectif des individus ne se trouvant pas dans la base. Une base incomplète cause alors des biais pour deux raisons : d'une part, parce que l'information concernant les personnes manquantes est, par définition, inconnue et d'autre part parce que l'on ne connaît pas généralement s'ils ont des comportements différents de ceux figurant dans la base.

La seconde erreur, celle de non-réponse, provient du fait que l'information désirée n'est observée que pour une partie de l'échantillon seulement. Cette erreur cause également des problèmes de biais mais aussi de précision des estimateurs.

L'ensemble des erreurs dans les enquêtes par sondage peut se résumer de la façon suivante :

$\begin{aligned} \text{ERREUR TOTALE} &= \text{ERREUR D'ECHANTILLONNAGE} \\ &+ \text{ERREUR D'OBSERVATION} \\ &+ \text{ERREUR DE COUVERTURE ET DE NON-REPONSE} \end{aligned}$
---

Ces trois erreurs entraînent chacune des problèmes de biais et de précisions qu'il faut pouvoir réduire et mesurer.

Les erreurs d'observation étant transformées en non-réponse (lorsqu'il n'est pas possible de la corriger par une méthode d'imputation déductive), la suite de la réflexion portera sur l'erreur de non-réponse.

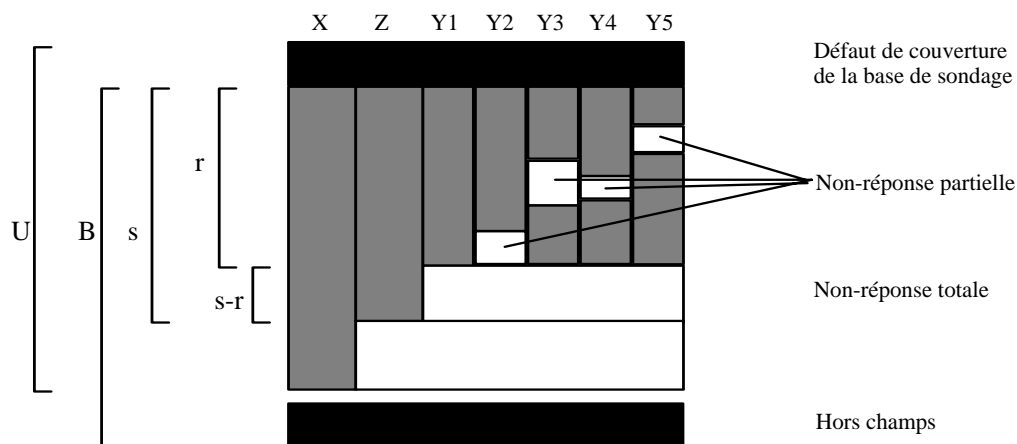
## 2) La non-réponse

Quelles que soient les méthodologies d'enquêtes mises en place pour collecter des réponses auprès des ménages, individus ou entreprises, l'information collectée n'est jamais complète. Cette seconde partie vise à définir et à mesurer la non-réponse, puis à identifier les causes et les problèmes liés à celle-ci.

### a) *Définition de la non-réponse*

La non-réponse est le fait de l'incapacité de mesurer sur toutes les unités de l'échantillon, toutes les variables d'intérêt (Graphique 1). Elle intervient dans tous les sondages probabilistes et tous les recensements et ce, même si on tente d'y remédier avant et pendant la collecte des données.

**Graphique 1 : Schéma de la non-réponse totale et de la non-réponse partielle**



Source : J.-C., Deville et F., Dupont, Insee Méthodes, N° 56-57-58, pp. 54.

Légende :

U : Univers

B : Base de sondage

s : Échantillon selon un tirage p(s)

r : Échantillon de répondant

X : Variables disponibles dans la base de sondage (utilisées soit pour fabriquer le plan de sondage, soit comme variables auxiliaires pour l'estimation)

Z : Variables 'faciles' à collecter sans non-réponse (exemple : type d'immeuble)

Y<sub>1</sub> : Variable à réponse totale (exemple : nombre de personnes vivant dans le ménage)

Y<sub>2</sub> à Y<sub>5</sub> : Variables entachées de données manquantes

On distingue deux catégories de non-réponses :

- La non-réponse totale : on ne dispose d'aucune information sur l'unité sélectionnée autre que celles présentes dans la base de sondage ;
- La non-réponse partielle : l'unité sélectionnée répond seulement à une partie de l'enquête mais pas à l'ensemble des variables d'intérêt.

La première mise en garde est de ne pas confondre non-réponse totale avec le concept d'unité hors champ. Une unité hors champs est une unité qui n'appartient pas au champ de l'enquête. L'un des premiers objectifs est donc de distinguer parmi les non-réponses celles qui appartiennent au champ de l'enquête (CARON, 2005). Le champ des enquêtes ménages de l'Insee, comme celui de l'Enquête Nationale sur les transports et les Déplacements 2007-08, correspond en général à l'ensemble des personnes qui vivent dans des résidences principales au moment de la réalisation de l'enquête (BRION, CARON, PIETRI-BESSY, 2005). L'échantillon des ménages provient des informations issues du recensement de la population, et peut être complété par la base de sondage des logements neufs. Les informations connues sur ces logements ne pouvant être actualisées, certaines des habitations identifiées s'avèrent par conséquent être des résidences secondaires, vacantes ou occasionnelles lors de la réalisation de l'enquête et constituent donc des logements hors champ. Le concepteur doit obligatoirement différencier les unités statistiques dans le champ de l'étude des autres puisque le taux de réponse à une enquête varie en fonction de cet effectif.

### ***b) Le taux de non-réponse***

Le taux de non-réponse totale est le rapport entre le nombre de non-répondants et le nombre d'individus échantillonnés qui font partie du champ de l'enquête. Il varie d'une enquête à l'autre selon le caractère obligatoire ou non de l'enquête, le sujet abordé ou la méthodologie mise en place pour recueillir les données. Selon la complexité de l'enquête (un ou plusieurs passages de l'enquêteur, remplissage ou non de carnet par l'enquêté, etc...) ce taux peut varier du simple au double (DEVILLE, DUPONT, 1996, Tableau 1). Une forte proportion de non-réponse totale

peut mettre en doute la représentativité de la population échantillonnée et est génératrice de problèmes de biais et de précisions.

**Tableau 5: Non-réponse à diverses enquêtes sur les conditions de vie**

Enquête	Non contacté (%)	Contacté + refus (%)	Accepté (%)		Total (%)
Loyers et charges (janvier 1989)	6,5	4,5	89		100
Conjoncture (mai 1989)	7,6	8,3	84,1		100
Emploi du temps (1985-1986)	7,5	8	84,5		100
			Complet 77	Abandon 7,5	
Budget de famille (1989)	8,3	12,5	79,4		100
			Complet 67,8	Abandon 11,6	
Logement (1984)	5	7	88		100
Actifs Financiers (1986)	7	14	79		100

Source : DEVILLE, DUPONT, 1996, « *Non-réponse : principes et méthodes* »

### *c) Les causes de la non-réponse*

Les causes de la non-réponse sont multiples et peuvent se différencier en fonction de la non-réponse totale et partielle (ARMOOGUM, 2002), (CARON, 2005).

Les causes invoquées pour expliquer la non-réponse totale se décomposent généralement ainsi :

- Les enquêtés ne retournent pas leur questionnaire (essentiellement pour les enquêtes postales) ;
- Les enquêtés refusent de répondre à toutes les enquêtes quelles qu'elles soient ;
- Les enquêtés sont impossibles à contacter ou absents de longue durée. En France, cette catégorie de non-réponse touche 5% à 9% des enquêtes ménages ;
- Les enquêtés refusent de répondre à une partie de l'enquête (carnet non rempli, etc...) ;
- Les enquêtés sont inaptes à répondre (barrière de la langue, sénilité, etc...).

Les principales causes de la non-réponse partielle s'articulent de la façon suivante :

- Les enquêtés ne comprennent pas une ou plusieurs questions ou sont dans l'impossibilité d'y répondre. Ces non-réponses résultent généralement d'une mauvaise formulation de la question ;
- Les enquêtés refusent de répondre à des questions sensibles. Le refus porte généralement sur le revenu des ménages (rappelons que pour limiter les réponses manquantes, cette variable est déjà découpée en tranches dans les enquêtes ménages de l'Insee) ;
- Le travail de l'enquêteur est de mauvaise qualité. Ce dernier peut commettre une erreur dans les filtres, ou la formation de l'enquêteur peut ne pas avoir été suffisante ;
- Les réponses de l'enquêté sont incohérentes et par conséquent invalidées.

Les causes liées à la non-réponse totale et partielle peuvent varier d'une enquête à l'autre en fonction du sujet d'étude et de la lourdeur du questionnaire. Mais quelles que soient les causes invoquées, les conséquences de la non-réponse sont semblables : perte de précision et apparition de biais à l'enquête.

#### *d) Les problèmes liés à la non-réponse*

La présence de données manquantes influe sur la qualité de l'inférence<sup>2</sup> notamment parce les répondants et les non-répondants ont généralement un comportement différent. La non-réponse introduit donc d'une part, une diminution de la précision des estimateurs à cause d'une perte de l'information et d'autre part, un biais à l'enquête qui peut être plus ou moins important selon que le comportement des non-répondants et des répondants est plus ou moins différent.

### **Le problème de la précision**

La non-réponse induit une perte d'information conduisant à une diminution de la précision des estimateurs. Lorsqu'on redresse un échantillon par pondération, on considère généralement que la non-réponse ajoute une phase supplémentaire à l'échantillonnage car elle réduit la taille de l'échantillon (la première phase consiste à tirer un échantillon dans la population U, et la deuxième consiste à tirer un échantillon à partir de l'échantillon de la première phase). La

---

<sup>2</sup> L'inférence statistique consiste à induire les caractéristiques inconnues d'une population à partir d'un échantillon issu de cette population.

réduction d'effectifs conduit alors à une réduction de la précision. On compense donc la perte de l'échantillon en augmentant les poids de sondage.

### **Le problème de biais**

Le taux de réponse ne mesure pas à lui seul l'erreur de non-réponse. Le biais de non-réponse est dû au fait que les répondants présentent des caractéristiques différentes de celles des non-répondants. Il dépend à la fois du nombre de non-répondants et de la différence pour une variable donnée entre la moyenne de l'ensemble des répondants et celle des non-répondants « corrigés ». Un taux de réponse élevé peut être très néfaste si les répondants et les non-répondants ont des comportements très différents en ce qui concerne les thèmes de l'enquête. Inversement, un taux de réponse faible n'est pas trop grave si ces comportements sont très voisins.

### **3) La correction de la non-réponse**

#### ***a) Les étapes préalables à la correction de la non-réponse***

Dans un premier temps, le statisticien doit définir précisément la population des non-répondants, c'est-à-dire une unité échantillonnée appartenant au champ de l'enquête pour lequel aucune réponse au questionnaire n'a été donnée. Il doit donc isoler les non-retours dans le champ de l'enquête de ceux hors du champ de l'enquête. De plus, avoir un questionnaire rempli entièrement ou en parti n'implique pas obligatoirement que les informations fournies sont correctes. La seconde étape est celle du contrôle de la cohérence des données. Ce contrôle est une étape préliminaire avant toutes corrections de la non-réponse car une réponse recueillie dans un questionnaire peut être invalidée et être transformée en valeur manquante. Ceci crée « artificiellement » de la non-réponse partielle qui se traite par les mêmes méthodes de correction que celle obtenue spontanément. Une dernière étape consiste à partager les questionnaires partiellement remplis entre ceux qui ne seront pas utilisés pour l'exploitation de l'enquête (ils seront alors corrigés par repondération) de ceux qui seront utilisés pour l'exploitation de l'enquête (ils seront alors corrigés par imputation).

#### ***b) Les différentes méthodes de correction de la non-réponse***

Il existe deux types de correction de la non-réponse : la repondération et l'imputation qui nécessitent toutes les deux de disposer d'informations sur les répondants et les non-répondants.

## ***Les principales méthodes de repondération***

Le principe des méthodes de repondération consiste à augmenter le poids de sondage des individus répondants pour compenser l'existence de la non-réponse. La repondération d'une enquête repose sur l'information dont dispose le statisticien sur les non-répondants, sur les modèles de comportement qui en découlent, et sur les sources extérieures. Il existe plusieurs stratégies pour corriger les biais dus à la non-réponse et ceux dus aux aléas d'échantillonnage, c'est au statisticien de trouver la meilleure méthode qu'il convient de mettre en œuvre lors du redressement de l'enquête.

Les méthodes proposées ici de correction de la non-réponse par repondération supposent que le nombre de non-répondants est connu, c'est-à-dire que l'on se place après la réalisation de l'enquête.

### **Le mécanisme de réponse globalement uniforme**

Il repose sur l'hypothèse peu réaliste que tous les individus de la population sont supposés avoir la même probabilité de répondre ainsi qu'un comportement indépendant les uns des autres. En utilisant ce modèle, l'estimateur du total d'une variable Y est sans biais. Ce modèle revient à calculer un estimateur en tenant compte uniquement des seuls répondants et en ignorant les non-répondants. Autrement dit : ne rien faire pour corriger la non-réponse revient à postuler un mécanisme de réponse uniforme, donc ne rien faire équivaut en réalité à faire quelque chose.

### **Le mécanisme de réponse homogène à l'intérieur de sous-populations**

Oh et Scheuren (1983) ont proposé de former des « groupes de réponse homogène » (GRH), en considérant un modèle de réponse un peu plus élaboré que celui du modèle de réponse naïf. Le mécanisme repose sur l'hypothèse que la population est divisée en sous-populations ayant des caractéristiques communes conduisant à un comportement de réponse proche. L'idée est de modéliser la non-réponse comme une phase de sondage supplémentaire au sondage initial. Les sous-populations sont constituées en analysant le critère répond/ne répond pas en fonction des variables connues pour les répondants et les non-répondants, tout cela conditionnellement à l'échantillon initial. Cette analyse se fait le plus souvent au moyen d'une régression logistique appliquée sur l'échantillon pour choisir les variables auxiliaires les plus explicatives de la non-

réponse ainsi que pour effectuer des regroupements adéquats de modalités. La première façon de constituer les sous-populations consiste à classer les individus selon leur probabilité de répondre et ensuite à faire une partition de la population ainsi classée en sous-populations. La seconde méthode consiste à définir comme sous-population celle qui correspond aux croisements de toutes les modalités des variables explicatives qui interviennent dans la régression logistique. Ces deux approches conduisent à des résultats proches en termes de réduction de biais. Une fois que les sous-populations sont constituées, la probabilité de réponse est supposée indépendante d'une sous-population à une autre et constante au sein d'une sous-population.

### **Le calage sur marges**

Il est également possible de redresser la non-réponse sans passer par les groupes de réponse homogène à l'aide d'un calage sur marges. Les techniques de calage sur marges ont été développées afin d'améliorer l'estimation de la variable d'intérêt, car nous réduisons la variance de notre estimateur lorsque les variables auxiliaires de calage sont corrélées avec la variable d'intérêt. Cette technique consiste à faire coïncider les marges de quelques variables de l'échantillon à celles de la population en modifiant la pondération. Lorsque les variables auxiliaires sont qualitatives, cette approche ne nécessite pas la connaissance dans la population du croisement de ces variables auxiliaires. La méthode de calage sur marges la plus connue en matière de sondage est la 'méthode itérative du quotient' ou Raking Ratio introduite par Deming et Stephan (1940) et Stephan (1942).

Deville et Särndal (1992) et Deville, Särndal et Sautory (1993) ont proposé une famille d'estimateurs par calage que l'on retrouve dans Calmar, logiciel de calage sur marges écrit par O. Sautory, couramment utilisé pour redresser les enquêtes dans plusieurs instituts nationaux de statistiques (SAUTORY, 1993).

Si la non-réponse affecte plusieurs variables du questionnaire (ce qui est généralement le cas), la repondération aura pour conséquence de donner un poids différent à chaque variable, ce qui n'est pas convenable lorsque l'on souhaite faire des tris croisés. De fait, la pondération sera généralement utilisée pour corriger la non-réponse totale. On appliquera des techniques d'imputation pour corriger la non-réponse partielle.

## ***Les principales méthodes d'imputation***

Le principe des méthodes d'imputation consiste à remplacer la réponse manquante ou invalide par une valeur plausible. L'objectif des procédures d'imputation est l'obtention d'une matrice de données complète (on parle dans ce cas de « clean data matrix »). Cette technique permet de traiter les données imputées comme des données réelles. L'inconvénient des méthodes d'imputation est qu'elles peuvent avoir des conséquences au niveau du calcul de la variance et de son estimateur (PLATEK, GRAY (1983), DEVILLE, SARNDAL (1994)). Le statisticien doit donc noter toutes les méthodes d'imputation qu'il a employé mais aussi créer des variables indicatrices qui permettraient de marquer dans le fichier les données imputées. Ceci laisserait alors la possibilité aux utilisateurs du fichier de juger de l'influence ou non des données imputées et de changer éventuellement de méthodologie d'imputation. Passons en revue les principales procédures d'imputation (LITTLE, RUBIN (1987), DUPONT (1994))

### **La méthode déductive**

La donnée manquante est déduite des réponses aux autres questions pour un même ménage. Cette méthode est souvent utilisée pour corriger des données jugées incohérentes ou invalides. On peut citer dans l'Enquête Nationale Transport l'exemple des déplacements vers un lieu de travail fixe. Prenons le cas d'un individu ayant renseigné son motif de déplacement par la modalité « lieu de travail fixe » mais pas le lieu de destination. Il sera alors possible de déduire la valeur manquante à l'aide du module « mobilité régulière » puisqu'auparavant il a été demandé à l'enquêté de renseigner le lieu de son travail fixe.

### **Le cold-deck**

Une imputation est de type cold-deck si on utilise une information extérieure à l'enquête relative au même ménage. Par exemple, lors des enquêtes sur les entreprises il n'est pas rare de remplacer l'effectif salarié lorsque cette donnée est manquante par celui qui a été déclaré par la même entreprise l'année précédente.

## **La prédiction par la moyenne**

Les données manquantes sont remplacées par la moyenne des répondants. Avec ce modèle, si l'on ajoute une proportion importante de valeurs égales à la moyenne, la distribution des valeurs est fortement modifiée et la variance ainsi calculée est inférieure à la variance réelle.

## **La prédiction par la moyenne par classe**

Pour limiter les effets négatifs de la méthode précédente sur la variance, la population enquêtée est éclatée en plusieurs classes. La donnée manquante des non-répondants d'une classe est alors remplacée par la moyenne des répondants de cette même classe.

## **La procédure Hot-deck**

L'imputation hot-deck est une procédure qui consiste à remplacer les réponses manquantes à certaines questions par des valeurs empruntées à d'autres répondants, on parle alors de « donneurs ». Il est préférable qu'un même enquêté reçoive des imputations d'un même donneur. Cette procédure se divise en plusieurs méthodes (CARON, 2005) :

- Le Hot-Deck d'ensemble : le donneur est choisi de façon aléatoire parmi l'ensemble des répondants de l'enquête. On distingue les cas où le tirage du donneur se fait avec ou sans remise, car la façon d'estimer la variance associée est différente ;
- Le Hot-Deck séquentiel : l'échantillon est classé dans un certain ordre dans le fichier. Si une unité est manquante, alors on lui impute la valeur renseignée par l'individu qui le précède dans le fichier. Cette procédure exige une valeur initiale ;
- Le Hot-Deck hiérarchisé : on remplace l'unité défaillante par une unité ayant les mêmes valeurs pour  $C_1, C_2, \dots, C_k$ . S'il n'en existe pas, on la remplace par une unité ayant les mêmes valeurs pour  $C_1, C_2, \dots, C_{k-1}$ , etc ;
- Le Hot-Deck métrique : on construit une distance entre individu en fonction de la corrélation entre les variables auxiliaires et les variables d'intérêts. Si l'individu est défaillant, on lui impute la valeur observée chez son plus proche voisin « donneur potentiel ». S'il y a ex-æquo, on impute une de ces valeurs possibles aléatoirement.
- Le Hot-Deck par classe : on construit des sous-populations sur le même principe que celui des groupes de réponses homogènes puis on applique l'une des méthodes du type

hot-deck présentées ci-dessus. On cherche le "donneur" dans la même classe que le non-répondant.

Les méthodes de repondération ne peuvent être utilisées que pour la correction de la non-réponse totale. Si la repondération était réalisée en fonction des variables non renseignées, un même individu pourrait avoir un poids de sondage différent pour chaque variable. Ce qui n'est pas admissible lorsque l'on souhaite faire des tris croisés ou des analyses multi-variées.

En revanche, l'imputation peut être appliquée à la fois sur la non-réponse totale et sur la non-réponse partielle. Selon Caron (2005), il n'existe pas de théorie satisfaisante pour choisir entre imputation et repondération concernant la non-réponse totale. A l'Insee, les deux approches sont utilisées : la non-réponse totale est corrigée par repondération pour des enquêtes réalisées auprès des ménages et par imputation pour des enquêtes réalisées auprès des entreprises.

### **III. Les taux de réponse dans le champ de l'enquête – Logements recensés en 1999**

Pour les 22 724 logements recensés qui étaient une résidence principale en 1999, nous disposons de quelques renseignements que nous pouvons mettre à profit afin d'analyser les taux d'échec, selon les modalités des variables suivantes.

**Tableau 6 : Répartition du taux d'échec en visite 1 selon la zone de résidence**

	Taux d'échec	Effectif total
Commune rurale	15,4%	7131
UU de moins de 20 000 habitants	20,1%	2707
UU de 20 000 à 99 999 habitants	22,4%	1815
UU de plus de 100 000 habitants	20,6%	4809
UU de Paris	29,7%	6262
Ensemble	21,5%	22 724

Source : INSEE Enquête Nationale sur les Transports et les Déplacements 2007-08

**Tableau 7 : Répartition du taux d'échec en visite 1 selon le revenu moyen des ménages de la commune du logement**

	Taux d'échec	Effectif total
Moins de 12000 €	17,4%	3333
Entre 12000 € et moins de 14000 €	21,0%	4150
Entre 14000 € et moins de 17000 €	19,1%	3797
17000 € et plus	22,6%	3398
Valeurs manquantes	24,3%	8046
Ensemble	21,5%	22 724

Source : INSEE Enquête Nationale sur les Transports et les Déplacements 2007-08

**Tableau 8 : Répartition du taux d'échec en visite 1 selon l'appartenance de l'immeuble à un organisme HLM**

	Taux d'échec	Effectif total
Immeuble d'un office HLM	26,9%	3080
Immeuble non HLM	21,3%	18013
Valeurs manquantes	14,7%	1631
Ensemble	21,5%	22 724

Source : INSEE Enquête Nationale sur les Transports et les Déplacements 2007-08

**Tableau 9 : Répartition du taux d'échec en visite 1 selon la période d'achèvement de l'immeuble**

	Taux d'échec	Effectif total
Immeuble datant d'avant 1915	22,9%	4240
Immeuble de 1915 à 1948	24,4%	2280
Immeuble de 1949 à 1967	24,2%	3355
Immeuble de 1968 à 1974	23,9%	3098
Immeuble de 1975 à 1981	18,8%	3000
Immeuble de 1982 à 1989	18,7%	2608
Immeuble d'après 1990	20,9%	2458
Immeuble en construction partiellement habité	18,5%	54
Valeurs manquantes	14,7%	1631
Ensemble	21,5%	22 724

Source : INSEE Enquête Nationale sur les Transports et les Déplacements 2007-08

**Tableau 10 : Répartition du taux d'échec en visite 1 selon le type de bâtiment**

	Taux d'échec	Effectif total
Maison individuelle, ferme	16,9%	14169
Logement dans un immeuble collectif	29,4%	8123
Logement-foyer pour personnes âgées	30,9%	81
Chambre d'hôtel	34,5%	29
Construction provisoire, habitation de fortune	27,3%	11
Pièce indépendante louée, sous-louée ou prêtée	30,7%	140
Logement dans un immeuble à usage autre que d'habitation	22,2%	171
Ensemble	21,5%	22 724

Source : INSEE Enquête Nationale sur les Transports et les Déplacements 2007-08

**Tableau 11 : Répartition du taux d'échec en visite 1 selon la présence d'un digicode ou d'un interphone**

	Taux d'échec	Effectif total
Immeuble d'un seul logement	17,4%	13363
Immeuble collectif avec digicode-interphone	30,9%	5496
Immeuble collectif sans digicode-interphone	28,6%	2234
Valeurs manquantes	14,7%	1631
Ensemble	21,5%	22 724

Source : INSEE Enquête Nationale sur les Transports et les Déplacements 2007-08

**Tableau 12 : Répartition du taux d'échec en visite 1 selon le nombre d'ascenseur**

	Taux d'échec	Effectif total
Pas d'ascenseur	17,4%	13363
Un ascenseur	31,0%	3604
2 ascenseurs ou plus	29,5%	4126
Valeurs manquantes	14,7%	1631
Ensemble	21,5%	22 724

Source : INSEE Enquête Nationale sur les Transports et les Déplacements 2007-08

**Tableau 13 : Répartition du taux d'échec en visite 1 selon le nombre de pièces du logement**

	Taux d'échec	Effectif total
0 et 1 pièce	37,9%	1054
2 et 3 pièces	27,0%	6824
4 et 5 pièces	18,6%	11045
6 pièces et plus	15,7%	3801
Ensemble	21,5%	22 724

Source : INSEE Enquête Nationale sur les Transports et les Déplacements 2007-08

**Tableau 14 : Répartition du taux d'échec en visite 1 selon la surface du logement**

	Taux d'échec	Effectif total
Moins de 40 m <sup>2</sup>	35,1%	1681
40 à 70 m <sup>2</sup>	27,0%	5223
70 à 100 m <sup>2</sup>	20,2%	7695
100 à 150 m <sup>2</sup>	17,1%	4941
150 m <sup>2</sup> et plus	16,9%	1553
Valeurs manquantes	14,7%	1631
Ensemble	21,5%	22 724

Source : INSEE Enquête Nationale sur les Transports et les Déplacements 2007-08

**Tableau 15 : Répartition du taux d'échec en visite 1 selon le nombre de personnes dans le ménage**

	Taux d'échec	Effectif total
1 personne	29,0%	5077
2 personnes	21,5%	7060
3 personnes	18,5%	4470
4 personnes et plus	17,7%	6117
Ensemble	21,5%	22 724

Source : INSEE Enquête Nationale sur les Transports et les Déplacements 2007-08

**Tableau 16 : Répartition du taux d'échec en visite 1 selon l'âge de la personne de référence**

	Taux d'échec	Effectif total
moins de 35 ans	26,8%	3866
Entre 35 et moins de 50 ans	20,4%	6905
Entre 50 et moins de 65 ans	19,1%	5618
65 ans et plus	24,1%	4704
Valeurs manquantes	14,7%	1631
Ensemble	21,5%	22 724

Source : INSEE Enquête Nationale sur les Transports et les Déplacements 2007-08

**Tableau 17 : Répartition du taux d'échec en visite 1 selon le sexe de la personne de référence**

	Taux d'échec	Effectif total
Homme	20,6%	16758
Femme	27,7%	4335
Valeurs manquantes	14,7%	1631
Ensemble	21,5%	22 724

Source : INSEE Enquête Nationale sur les Transports et les Déplacements 2007-08

**Tableau 18 : Répartition du taux d'échec en visite 1 selon la motorisation du ménage**

	Taux d'échec	Effectif total
Sans voiture	30,7%	3535
Une voiture	23,5%	8251
2 voitures ou plus	17,1%	10938
Ensemble	21,5%	22 724

Source : INSEE Enquête Nationale sur les Transports et les Déplacements 2007-08

**Tableau 19 : Répartition du taux d'échec en visite 1 selon la position professionnelle de la personne de référence à la date du RP99**

	Taux d'échec	Effectif total
Position professionnelle non précisée	26,7%	1026
Manœuvre, ouvrier spécialisé (OS1 à OS3, etc)	21,7%	1068
Ouvrier qualifié ou très qualifié (P1 à P3, TA, OQ, OQH, etc)	20,9%	2195
Agent de service, aide soignant(e), employé de maison	25,2%	412
Employé de commerce, employé de bureau, personnel administratif de catégorie C ou D de la fonction publique	22,6%	1344
Agent de maîtrise dirigeant des ouvriers, maîtrise administratif, commerciale, informatique	16,8%	709
Agent de maîtrise dirigeant des techniciens ou d'autres agents de maîtrise	18,2%	258
Technicien, dessinateur, VRP	20,3%	824
Instituteur, infirmier(e), travailleur social, technicien médical, personnel de catégorie B de la fonction publique	19,8%	691
Ingénieur, cadre d'entreprise	22,3%	1509
Personnel de catégorie A de la fonction publique et assimilés Non-salariés	18,3%	771
Indépendant	19,5%	1054
Employeur	20,4%	1102
Aide familial	23,1%	91
Valeurs manquantes	22,0%	9670
Ensemble	21,5%	22 724

Source : INSEE Enquête Nationale sur les Transports et les Déplacements 2007-08

Les refus à la première visite sont les plus nombreux dans les unités urbaines de plus de 20 000 habitants. Les ménages dont le revenu moyen est de moins de 12 000 € acceptent plus volontiers la première visite de l'enquêteur. Les ménages habitant dans des immeubles construits avant 1974 sont les plus réticents à recevoir l'enquêteur. Les échecs à la première visite sont plus importants pour les ménages habitant un logement de type immeuble collectif. Pour ce type d'immeuble, avoir un interphone ou digicode n'augmente pas le taux d'échec à la première visite. En revanche, les enquêtés refusent plus facilement l'enquête lorsque le logement de type

immeuble collectif appartient à un office HLM. Les ménages habitant dans un logement de moins de deux pièces sont les plus réticents à accepter la première visite. De même que ceux vivant dans un logement de moins de 40m<sup>2</sup> de surface. Les ménages composés de 3 personnes et plus sont les plus nombreux à répondre aux questions de la première visite. Ce sont les ménages dont la personne de référence était âgée entre 35 et 65 ans au RP99 qui répondent le plus à la visite 1 de l'enquête. Les femmes répondent moins souvent que les hommes. Les ménages dont la position professionnelle de la personne de référence à la date du RP99 n'est pas précisée ou agents de service, aide soignant(e)s, employés de maison sont les plus réticents à recevoir l'enquêteur. Suivent ensuite les employés de commerce, employés de bureau, personnels administratifs de catégorie C ou D de la fonction publique et les ingénieurs ou cadres d'entreprises. Et enfin, les échecs à la première visite sont plus importants chez les ménages n'ayant pas de voiture.

Avoir de la non-réponse dans une enquête est un facteur sur lequel le statisticien ne peut passer outre. Il est indispensable de la limiter au moment de la conception de l'enquête et de la corriger après la collecte des données. Mais quel que soit le procédé, appliquer une méthode de correction de la non-réponse revient à choisir un modèle de comportement des individus face à la non-réponse, c'est-à-dire faire des hypothèses sur les réponses des non-répondants. Ces hypothèses sont construites sur les informations des non-répondants et des répondants. Il est donc nécessaire d'avoir des informations pertinentes.



## **CHAPITRE 3. Modélisation du mécanisme de réponse**

Ce chapitre se focalisera sur la correction de la non-réponse totale. Dans l'ENTD 2007-08, la non-réponse totale est présente de deux façons : lorsque les enquêtés n'ont pas répondu à l'ensemble du questionnaire et lorsqu'ils ont répondu aux questions de la visite 1 mais pas à celles de la visite 2. Or en présence de données manquantes, le calcul d'estimations fondé sur les seuls répondants fournit en général des estimations, d'une part, biaisées puisque la plupart du temps les non-répondants n'ont pas les mêmes comportements que les répondants (BRILHAULT, CARON, 2004) et d'autre part, moins précises puisque la taille de l'échantillon est plus petite que celui initialement tiré. Pour l'ENTD 2007-08, il est donc indispensable de connaître les caractéristiques des non-répondants en fonction des deux sortes de non-réponse totale avant de passer à la correction.

Une première partie sera consacrée à l'étude de la non-réponse totale à travers l'analyse du mécanisme de réponse à la visite 1 et une seconde au mécanisme de réponse entre la visite 1 et 2. La dernière partie abordera la correction de la non-réponse d'un point de vue exclusivement théorique.

### **I. Modélisation du mécanisme de réponse de la visite 1**

Pour chacun des 22 724 des logements recensés qui étaient une résidence principale en 1999, nous disposons de quelques renseignements. Il convient donc de chercher les variables discriminantes du mécanisme de réponse à l'aide d'un modèle logit. Notons que le ménage occupant le logement au moment de l'enquête n'est pas nécessairement celui qui l'occupait au moment du recensement de la population de 1999 (RP99).

Les variables étudiées sont les suivantes :

- Zone de résidence
- Revenu moyen de la commune du logement au RP99
- Tranche d'unité urbaine de la commune en 1999
- Période d'achèvement de l'immeuble
- Appartenance de l'immeuble à un organisme HLM
- Type de bâtiment

- Digicode ou interphone
- Ascenseur
- Nombre de pièces du logement
- Surface du logement
- Nombre de personnes du ménage au RP99
- Age de la personne de référence au RP99
- Sexe de la personne de référence au RP99
- Position professionnelle de la personne de référence au RP99
- Motorisation du ménage au RP99

Le modèle logit permet de mettre en évidence les variables auxiliaires qui expliquent le mécanisme de réponse. Parmi les variables citées ci-dessus, on retient (résultat en annexe 1) :

- **Type de bâtiment** (maison ; ferme ; autres logements)
- **Nombre de pièces du logement** (0-1 pièce ; 2-3 pièces ; 4-5 pièces ; 6 pièces et plus)
- **Zone de résidence** (Commune rurale ; UU de moins de 20 000 habitants ; UU de 20 000 à 99 999 habitants ; UU de plus de 100 000 habitants ; UU de Paris)
- **Motorisation du ménage au RP99** (0 voiture, 1 voiture ; 2 voitures et plus)
- **Age de la personne de référence au RP99** (15-34 ans ; 35-49 ans ; 50-64 ans ; 65 ans et plus)
- **Appartenance de l'immeuble à un organisme HLM** (oui ; non)

On oppose ainsi en première analyse :

- Les ménages habitant une maison aux ménages résidant dans une habitation collective. Les échecs sont plus fréquents pour les logements collectifs (c'est probablement une question d'accessibilité du logement) ;
- Les ménages habitant un studio ou une chambre aux ménages résidant dans des logements ayant plusieurs pièces. Cette variable est corrélée avec le nombre de personnes vivant dans le ménage. Ainsi, une taille de ménage plus importante s'accompagne d'une probabilité plus grande de réaliser l'entretien ;
- Les logements situés en zone rurale et en agglomération de moins de 20 000 habitants à ceux situés dans l'agglomération de Paris. Les échecs sont d'autant plus nombreux qu'on progresse vers une plus grande urbanisation ;

- Les ménages n'ayant aucune automobile aux ménages motorisés. Les ménages non-équipés en automobile sont moins favorables à la réalisation des entretiens ;
- Les ménages dont la personne de référence a moins de 35 ans ou plus de 65 ans à ceux dont l'âge se situe entre 35 et 65 ans. Certainement pour des raisons différentes, les taux d'échec sont plus importants pour les ménages dont la personne de référence a moins de 35 ans et pour ceux dont l'âge de la personne de référence est supérieur à 65 ans. Pour les premiers cela souligne la difficulté des enquêteurs de joindre ces ménages et pour les seconds la réticence des personnes âgées à répondre à un long questionnaire ;
- Les ménages résidant dans un HLM aux autres. Les échecs sont plus nombreux pour les ménages habitant un HLM.

Les variables du RP99 dont nous disposons et qui expliquent le mécanisme de réponse de l'ENTD, soulignent souvent la difficulté de l'enquêteur à accéder aux logements, même si le fait d'avoir un digicode n'est pas un facteur aggravant le taux d'échec.

Un sérieux problème subsiste, surtout pour une enquête sur la mobilité, c'est le taux d'échec important chez les ménages plus jeunes, et probablement très mobiles.

## **II. Modélisation du mécanisme de réponse de la visite 2**

Entre les deux visites, nous avons perdu 8% des ménages. On dispose, pour l'ensemble des ménages répondant à la première visite, d'informations qu'il convient de mettre à profit afin d'étudier les profils des non-répondants à la deuxième visite. Les variables étudiées sont les suivantes :

- Zone de résidence
- Type de ménage
- Revenu du ménage
- La motorisation du ménage
- Age de la personne de référence
- Sexe de la personne de référence
- Age de l'individu « Kish »
- Sexe de l'individu « Kish »
- Nombre de kilomètres parcourus cours des 12 derniers mois (de la première voiture du ménage)
- Nombre de déplacements de longue distance de l'individu « Kish »

- Durée entre la date de la première visite et la date de la fin vague

Il convient donc de chercher les variables discriminantes du mécanisme de réponse (entre les 2 visites) à l'aide d'un modèle logit. Parmi les variables citées ci-dessus, celles qui expliquent le mécanisme de réponse sont les suivantes (résultat en annexe 2) :

- **Zone de résidence** (Commune rurale ; UU de moins de 20 000 habitants ; UU de 20 000 à 99 999 habitants ; UU de plus de 100 000 habitants ; UU de Paris)
- **Type de ménage** (ménage d'une personne ; couple sans enfant ; famille monoparentale ; couple avec un enfant ; couple avec plus de deux enfants ; autre)
- **Revenu du ménage** (moins de 1 499 € ; de 1 500 à 2 499 € ; de 2 500 à 3 399 € ; 3 400 € et plus ; valeur manquante)
- **Sexe de l'individu « Kish »**
- **Nombre de kilomètres parcourus cours des 12 derniers mois (de la première voiture du ménage)** (ménage non-motorisé ; moins de 2 499 km ; de 2 500 à 8 999 km ; de 9 000 à 14 999 km ; 15 000 km et plus ; valeur manquante)
- **Nombre de déplacements de longue distance de l'individu « Kish »** (0 ; 1 ; 2 ; 3 ; entre 4 et 9 ; plus de 10 voyages)
- **Durée entre la date de la première visite et la date de la fin vague** (Visite 1 réalisée après la date de fin de collecte, cas exceptionnels justifiés par la Division des Enquêtes Ménages et validés par l'Unité des Méthodes Statistiques ; entre 0 et 14 jours ; entre 15 et 28 jours ; entre 29 et 43 jours ; entre 44 et 57 jours ; 58 jours et plus)

Analyse du taux d'échec des variables expliquant le mécanisme de réponse entre la visite 1 et la visite 2 :

**Tableau 20 : Répartition du taux d'échec en visite 2 selon la zone de résidence**

	Taux d'échec	Effectif total
Commune rurale	7,4%	6885
UU de moins de 20 000 habitants	8,6%	2509
UU de 20 000 à 99 999 habitants	9,5%	1582
UU de plus de 100 000 habitants	9,5%	4325
UU de Paris	6,7%	4966
Total	8,0%	20267

Source : INSEE Enquête Nationale sur les Transports et les Déplacements 2007-08

**Tableau 21 : Répartition du taux d'échec en visite 2 selon le type de ménage**

	Taux d'échec	Effectif total
ménage d'une personne	7,7%	5364
couple sans enfant	9,5%	1494
famille monoparentale	7,8%	6209
couple avec un enfant	7,9%	6694
couple avec plus de deux enfants	9,9%	506
Total	8,0%	20267

Source : INSEE Enquête Nationale sur les Transports et les Déplacements 2007-08

**Tableau 22 : Répartition du taux d'échec en visite 2 selon le revenu du ménage**

	Taux d'échec	Effectif total
moins de 1500 €	8,1%	4660
de 1500 à 2499 €	6,8%	5423
de 2500 à 3399 €	5,9%	3853
3400 € et plus	6,6%	4761
Valeur manquante	21,0%	1570
Total	8,0%	20267

Source : INSEE Enquête Nationale sur les Transports et les Déplacements 2007-08

**Tableau 23 : Répartition du taux d'échec en visite 2 selon le sexe de l'individu « Kish »**

	Taux d'échec	Effectif total
Hommes	8,6%	8892
Femmes	6,5%	11253
Valeur manquante	100,0%	122
Total	8,0%	20267

Source : INSEE Enquête Nationale sur les Transports et les Déplacements 2007-08

**Tableau 24 : Répartition du taux d'échec en visite 2 selon le nombre de kilomètres parcourus cours des 12 derniers mois (de la première voiture du ménage)**

	Taux d'échec	Effectif total
Ménage non-motorisé	8,1%	3319
moins de 2500 km	7,9%	1615
de 2500 à 8999 km	5,7%	4391
de 9000 à 14999 km	5,9%	3547
15000 km et plus	7,9%	5227
Valeur manquante	16,1%	2168
Total	8,0%	20267

Source : INSEE Enquête Nationale sur les Transports et les Déplacements 2007-08

**Tableau 25 : Répartition du taux d'échec en visite 2 selon le nombre de voyages de longue distance de l'individu « Kish » au cours des 3 derniers mois**

	Taux d'échec	Effectif total
0 voyage	8,7%	8456
1 voyage	5,7%	4581
2 voyages	5,2%	2356
3 voyages	6,3%	1433
entre 4 et 9 voyages	6,4%	2144
plus de 10 voyages	10,1%	951
Valeur manquante	51,2%	346
Total	8,0%	20267

Source : INSEE Enquête Nationale sur les Transports et les Déplacements 2007-08

**Tableau 26 : Répartition du taux d'échec en visite 2 selon la durée entre la date de la première visite et la date de la fin vague**

Visite 1 réalisée :	Taux d'échec	Effectif total
après la date de fin de collecte	24,1%	58
entre 0 et 14 jours de la date de fin collecte	15,7%	2144
entre 15 et 28 jours de la date de fin collecte	8,0%	4552
entre 29 et 43 jours de la date de fin collecte	6,6%	6304
entre 44 et 57 jours de la date de fin collecte	6,8%	5965
58 jours et plus de la date de fin collecte	6,7%	1244
Total	8,0%	20267

Source : INSEE Enquête Nationale sur les Transports et les Déplacements 2007-08

Les ménages dont les revenus mensuels sont entre 1500 € et 3400 € acceptent plus volontiers la deuxième visite de l'enquêteur. Les ménages qui ont déclaré un kilométrage parcouru au cours des 12 derniers mois de moins de 2 500 km et plus de 15 000 km sont les plus réticents pour une deuxième visite. Les femmes répondent un peu mieux que les hommes. C'est dans les unités urbaines de plus de 20 000 habitants, hors celle de Paris, que les refus à la deuxième visite sont les plus nombreux. C'est lorsque le nombre de voyages de longue distance est compris entre 1 et 4 que les taux de succès sont les plus importants. Coté type de ménage, c'est avec les couples sans enfant et les couples avec plus de deux enfants que les enquêteurs ont plus de mal pour la seconde visite. La variable donnant le taux de succès en fonction du laps de temps entre la première visite et la date de la fin vague explique la difficulté des enquêteurs à obtenir des rendez-vous auprès des enquêtés. Remarquons que les enquêteurs avaient pour consigne de réaliser une première visite même lorsqu'il ne restait plus assez de temps pour réaliser la seconde. L'échec à la deuxième visite, dans ces cas, n'est donc pas tant le fait d'un refus des enquêtés que de leur indisponibilité pendant les premières semaines de la vague, ou de la

surcharge de travail de certains enquêteurs. Enfin, les échecs à la deuxième visite sont plus importants chez les ménages n'ayant pas répondu à certaines questions (revenu, kilométrage parcouru les 12 derniers mois par une voiture du ménage et nombre de déplacements à longue distance).

D'après cette modélisation, les non-répondants à la deuxième visite seraient, soit peu motivés par le sujet de l'enquête (et peu mobiles), soit peu disponibles pour recevoir une deuxième l'enquête (et probablement très mobiles).

Il est utile d'étudier le mécanisme de réponse pour les deux visites car l'on suppose que les personnes qui n'ont pas répondu en visite 1 ne l'ont pas fait pour les mêmes raisons que les personnes en visite 2. En effet, le mécanisme de réponse pour la visite 1 souligne la difficulté d'accès aux logements des enquêtés, alors que le mécanisme de réponse pour la visite 2 suppose que la mobilité des enquêtés est un facteur de non-réponse.

L'étape suivante consiste à corriger les erreurs d'échantillonnage et de non-réponse par la méthode du calage sur marges en utilisant, en autres, les variables explicatives du mécanisme de réponse.

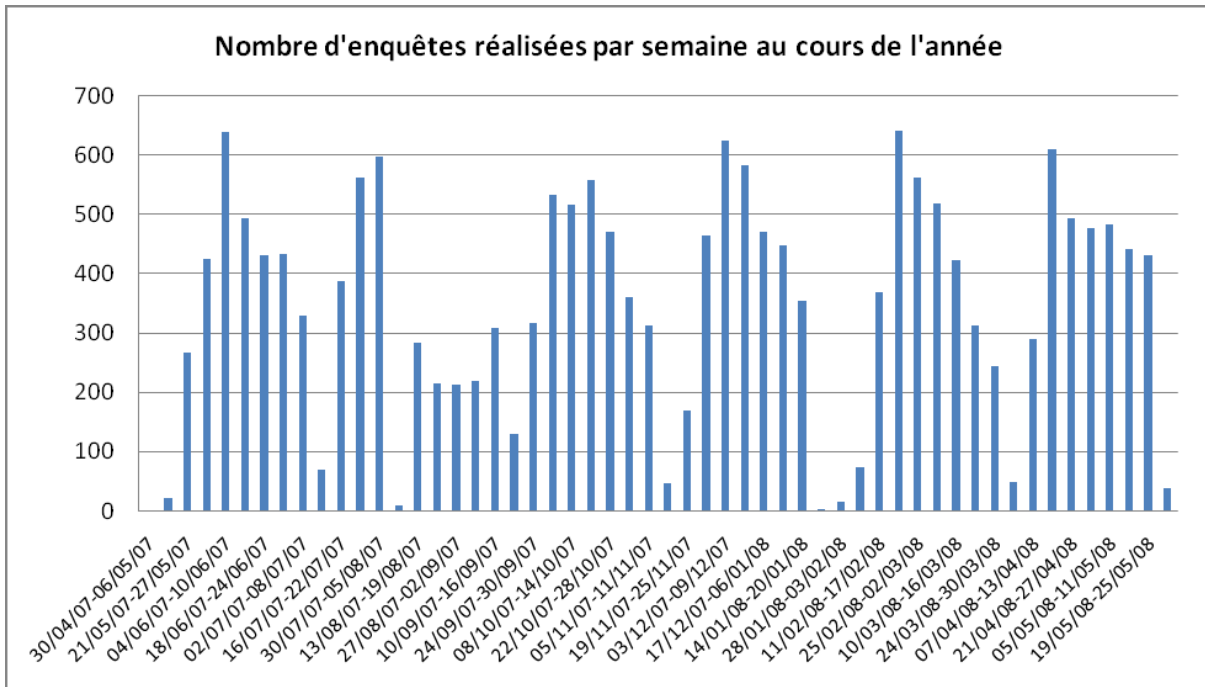
### **III. La correction des erreurs d'échantillonnage et de la non-réponse par calage sur marges**

Les techniques de calage sur marges ont été développées afin d'améliorer l'estimation de la variable d'intérêt, car nous réduisons la variance de notre estimateur lorsque les variables auxiliaires de calage sont corrélées avec la variable d'intérêt. Cette technique consiste à faire coïncider les marges de quelques variables de l'échantillon à celles de la population en modifiant la pondération. Lorsque les variables auxiliaires sont qualitatives, cette approche ne nécessite pas la connaissance, dans la population, du croisement de ces variables auxiliaires. La méthode de calage sur marges la plus connue en matière de sondage est la 'méthode itérative du quotient' ou Raking Ratio introduite par Deming et Stephan (1940) et Stephan (1942). Deville et Särndal (1992) ont généralisé l'approche des estimateurs par régression en introduisant les estimations par 'calage sur marges'. Nous pouvons ainsi corriger les différents types d'erreurs à condition d'utiliser les variables expliquant le mécanisme de la non-réponse lors du calage sur marges.

Cette étape est essentielle pour assurer une bonne représentativité de l'échantillon et la comparabilité avec d'autres sources statistiques (enquêtes de l'INSEE). Au cours de cette étape,

les contraintes temporelles relatives à l'ENTD 2007-08 seront introduites lorsque cela sera nécessaire (Graphique 2).

**Graphique 2 : Nombre d'enquêtes réalisées par semaine au cours de l'année**



Source : INSEE Enquête Nationale sur les Transports et les Déplacements 2007-08

Nous disposons, pour réaliser ce calage, de la macro CALMAR écrite dans les années 90 (DEVILLE, SARNDAL, SAUTORY 1993) et aussi de la nouvelle version de CALMAR2, développée en 2003 (LE GUENNEC, SAUTORY, 2003) qui proposent à l'utilisateur de nouvelles facilités pour réaliser un calage, et permettent de mettre en œuvre la méthode de traitement de la non-réponse par calage généralisé, proposée par Deville (1998).

L'idéal pour corriger la non-réponse et les erreurs d'échantillonnage serait de caler sur des variables expliquant le mécanisme de réponse et liées aux variables d'intérêt des sous-populations sur lesquelles nous souhaitons faire des estimations (exemple dans DEVILLE, SARNDAL, SAUTORY 1993).

Avant sa réalisation, il faut définir les variables d'intérêt retenues pour le calage. La visite 1 s'adressant aux ménages et à l'ensemble de ses membres, la visite 2 aux individus tirés au sort, les variables d'intérêt sont différentes mais liées.

### *a) Les variables de calage pour le niveau ménages-individus*

Afin de construire une pondération de la première visite, nous souhaiterions utiliser les variables de calage suivantes pour le niveau ménages – individus. Les marges peuvent être issues du RP ou de l'enquête emploi.

- Le nombre de ménages
  
- PCS de la personne de référence (en %)
  - Agriculteurs (actifs occupés et anciens actifs)
  - Indépendants ou professions libérales (actifs occupés et anciens actifs)
  - Professeurs, instituteurs et assimilés, professions intermédiaires de la santé, autres cadres ou professions intermédiaires (actifs occupés)
  - Professeurs, instituteurs et assimilés, professions intermédiaires de la santé, autres cadres ou professions intermédiaires (anciens actifs)
  - Employés administratifs, employés du commerce et des services aux personnes, ouvriers (actifs occupés)
  - Employés administratifs, employés du commerce et des services aux personnes, ouvriers (anciens actifs)
  - Autres (inactifs, chômeurs n'ayant jamais travaillé,...)
  
- Sexe et âge de la personne de référence (en %)
  - Hommes :
    - de moins de 30 ans
    - de 30 à 39 ans
    - de 40 à 49 ans
    - de 49 à 50 ans
    - de 60 à 69 ans
    - plus de 70 ans
  - Femmes :
    - de moins de 30 ans
    - de 30 à 39 ans
    - de 40 à 49 ans
    - de 49 à 50 ans

- de 60 à 69 ans
  - plus de 70 ans
- Type du ménage (en%)
    - 1 personne
    - Couple sans enfant
    - Famille monoparentale
    - Couple avec 1 enfant
    - Couple avec 2 enfants ou plus
    - Autre
  - Variable type d'immeuble
    - Immeuble individuel
    - Autre
  - Variable diplôme le plus élevé obtenu
    - Aucun diplôme ou CEP ou diplôme équivalent ou diplôme non déclaré
    - CAP, BEP ou autre diplôme de ce niveau ou BEPC seul
    - Bac ou brevet professionnel ou diplôme de ce niveau ou Bac+2 ans,
    - Diplôme supérieur
  - Nationalité
    - Française
    - Autre
  - Zone de résidence (en % des ménages)
    - Commune rurale (TUU=0)
      - Commune monopolarisée ou commune multipolarisée selon le ZAU 99
      - Espace à dominante rurale selon le ZAU 99
    - Commune urbaine (TUU>0) -
      - Ville-Centre
      - Banlieue

- Unité urbaine de moins de 19.999 habitants (1,2,3)
  - Ville-Centre
  - Banlieue
- Unité urbaine de 20.000 à 49.999 habitants (4)
  - Ville-Centre
  - Banlieue
- Unité urbaine de 50.000 à 99.999 habitants (5)
  - Ville-Centre
  - Banlieue
- Unité urbaine de 100.000 à 199.999 habitants (6)
  - Ville-Centre
  - Banlieue
- Unité urbaine de 200.000 à 1.999.999 habitants (7)
  - Ville-Centre
  - Banlieue
- Unité urbaine de Paris (8)
  - Ville-Centre
  - Banlieue
- Nombre d'individus
  - Nombre d'hommes
    - de moins de 25 ans
    - de 25 à 34 ans
    - de 35 à 49 ans
    - de 50 à 64 ans
    - de plus de 65 ans
  - Nombre de femmes
    - de moins de 25 ans
    - de 25 à 34 ans
    - de 35 à 49 ans
    - de 50 à 64 ans
    - de plus de 65 ans

- Motorisation du ménage (%)
  - 0 voiture
  - 1 voiture
  - 2 voitures
  - 3 voitures ou plus
- Variable temporelle : vague de l'enquête (6 vagues)

Concernant la pondération de la deuxième visite (taux d'échec de 8% entre les 2 visites), nous souhaiterions comparer un calage similaire et un calage en ajoutant quelques marges supplémentaires (total kilométrage, revenu, ...) issues de la première visite.

***b) Les variables de calage pour le niveau des individus « Kish »***

- Le nombre d'individus de 6 ans et plus appartenant à un ménage
- PCS de la personne « Kish » (population des individus de 6 ans et plus en %)
  - Agriculteurs (actifs occupés et anciens actifs)
  - Indépendants ou professions libérales (actifs occupés et anciens actifs)
  - Professeurs, instituteurs et assimilés, professions intermédiaires de la santé, autres cadres ou professions intermédiaires (actifs occupés)
  - Professeurs, instituteurs et assimilés, professions intermédiaires de la santé, autres cadres ou professions intermédiaires (anciens actifs)
  - Employés administratifs, employés du commerce et des services aux personnes, ouvriers (actifs occupés)
  - Employés administratifs, employés du commerce et des services aux personnes, ouvriers (anciens actifs)
  - Autres (inactifs, chômeurs n'ayant jamais travaillé,...)
  - Individus de 6 à 15 ans
- Sexe et âge de la personne (population des individus de 6 ans et plus en %)
  - Hommes
    - de 6 à 25 ans
    - de 25 à 34 ans
    - de 35 à 49 ans

- de 50 à 64 ans
- plus de 65 ans
- Femmes
  - de 6 à 25 ans
  - de 25 à 34 ans
  - de 35 à 49 ans
  - de 50 à 64 ans
  - plus de 65 ans
- Type du ménage (population des individus de 6 ans et plus en%)
  - 1 personne
  - Couple sans enfant
  - Famille monoparentale
  - Couple avec 1 enfant
  - Couple avec 2 enfants ou plus
  - Autre
- Type d'immeuble (population des individus de 6 ans et plus en %)
  - Immeuble individuel
  - Autre
- Diplôme le plus élevé obtenu (population des individus de 6 ans et plus en %)
  - Aucun diplôme ou CEP ou diplôme équivalent ou diplôme non déclaré
  - CAP, BEP ou autre diplôme de ce niveau ou BEPC seul
  - Bac ou brevet professionnel ou diplôme de ce niveau ou Bac+2 ans,
  - Diplôme supérieur
- Nationalité (population des individus de 6 ans et plus en %)
  - Française
  - Autre

- Zone de résidence (population des individus de 6 ans et plus en %)
  - Commune rurale (TUU=0)
    - Commune monopolarisée ou commune multipolarisée selon le ZAU 99
    - Espace à dominante rurale selon le ZAU 99
  - Commune urbaine (TUU>0) -
    - Ville-Centre
    - Banlieue
  - Unité urbaine de moins de 19.999 habitants (1,2,3)
    - Ville-Centre
    - Banlieue
  - Unité urbaine de 20.000 à 49.999 habitants (4)
    - Ville-Centre
    - Banlieue
  - Unité urbaine de 50.000 à 99.999 habitants (5)
    - Ville-Centre
    - Banlieue
  - Unité urbaine de 100.000 à 199.999 habitants (6)
    - Ville-Centre
    - Banlieue
  - Unité urbaine de 200.000 à 1.999.999 habitants (7)
    - Ville-Centre
    - Banlieue
  - Unité urbaine de Paris (8)
    - Ville-Centre
    - Banlieue
  
- Jour de la semaine
  
- Variable temporelle : vague de l'enquête (6 vagues)

Plusieurs pondérations sont donc à effectuer mais dans une optique de traitement de l'enquête, nous ne pouvons pas nous permettre d'avoir plusieurs jeux de pondérations sur une même table. Nous souhaiterions donc disposer d'une pondération unique, afin de n'utiliser qu'un seul poids quel que soit le niveau géographique sur lequel nous travaillerons (zone

d'extension de l'enquête ou non). La macro CALMAR de l'INSEE doit donc caler l'enquête en tenant compte à la fois de toutes les variables d'intérêts, des variables explicatives des 2 modèles de réponse, et du zonage géographique. Cette méthode de redressement permet de réduire la variance d'échantillonnage, et, dans certains cas, de réduire le biais dû à la non réponse totale.



## CONCLUSION

Dans toute enquête, pour éviter la non-réponse lors du recueil des données, il est crucial que les concepts utilisés soient clairs et compris de manière identique, tant par les personnes interrogées que par les enquêteurs. L'accès au logement des enquêtés étant de plus en plus difficile, de nouvelles méthodes de recueil des données doivent être développées. Nous devons aussi profiter des développements technologiques. L'utilisation du récepteur GPS dans l'Enquête Nationale sur les Transports et les Déplacements 2007-08 laissent présager la réalisation d'enquête presque totalement automatique ne nécessitant que peu de contact avec l'enquêteur. Les Enquêtes Nationales sur les Transports sont les seules à donner une description détaillée des déplacements des ménages résidant en France et de leur usage des moyens de transport tant collectifs qu'individuels. Les nouveautés méthodologiques et technologiques apportées pour à l'édition 2007-08 sont des atouts supplémentaires qui permettent d'augmenter la quantité et la qualité des informations recueillies.

Dans la théorie de la correction par repondération, la non-réponse totale ajoute une phase supplémentaire à l'échantillonnage : le mécanisme de réponse. Nous pouvons l'identifier à l'aide d'une modélisation qui fait apparaître les probabilités de réponse. Deville et Särndal (1992) ont généralisé les méthodes d'estimation utilisant de l'information auxiliaire à l'aide d'estimateurs par calage sur marges. Le calage sur marges permet de corriger les erreurs liées à la non-réponse et aussi d'augmenter la précision des estimateurs.

L'apurement de l'enquête se poursuivra ensuite par la correction de la non-réponse partielle et des erreurs de mémoire des enquêtés. Une erreur de mémoire est une omission ou une réponse erronée produite involontairement par personne interrogée. En effet, un volet rétrospectif est présent dans l'Enquête Nationale sur les Transports et les Déplacements 2007-08. Une personne est interrogée sur ses voyages à longue distance (plus de 80 kilomètres à vol d'oiseau) effectués entre la date de visite 1 et les trois derniers mois, en faisant appel à sa mémoire. Nous suspectons donc des erreurs de mémorisation du nombre de voyages faits et des erreurs de datation qu'il faudra corriger.

Les transports vont encore considérablement évoluer au XXI<sup>e</sup> siècle. L'étude de l'histoire des habitudes passées et des changements de comportement est indispensable pour éclairer les changements à venir. Les Enquêtes Nationales sur les Transports ont été menées au niveau national à cinq reprises donnant cinq photographies de la mobilité des Français à un moment donné. Elles sont riches en informations mais ne permettent pas, d'une part, de mesurer les changements individuels de comportement puisque les données sont recueillies environ tous les 10 ans et d'autre part, d'appréhender le récit d'une histoire de la mobilité. L'édition 2007-08 innove par rapport aux précédentes par sa volonté d'esquisser une perspective historique en proposant un questionnaire biographique, outil de recueil ordinairement réservé à la démographie.

L'analyse des biographies combinées à celles des ENT va nous permettre approfondir ces questions de changements de comportements.

# BIBLIOGRAPHIE

## OUVRAGES

- Ardilly, P., 2006, *Les techniques de sondage*, Technip, 675p.
- Armoogum, J., 2002, *Correction de la non-réponse et de certaines erreurs de mesures dans une enquête par sondage : Application à l'enquête Transports et Communications 1993-94*, Rapport INRETS, n 239, 173p.
- CERTU, 2008, *L'enquête ménages déplacements "Standard Certu"*, 204p.
- Little, R.J.A., Rubin, D.B., 1987, *Statistical Analysis with Missing Data*, John Wiley, New York, 304p.
- Platek R., Pierre-Pierre, F.K. et Stevens, P., 1985, *Élaboration et conception des questionnaires d'enquête*, Statistique Canada, Division des méthodes de recensement d'enquêtes-ménages, 83p.
- Richardson, A.J., Ampt, E.S. et Meyburg, A., 1995, *Survey Methods for Transport Planning*, Eucalyptus Press, Melbourne, ??p.

## ARTICLES

- Armoogum, J., Madre, J.-L., 1997, « Du redressement des non-réponses totales aux contrôles sur la cohérence des réponses », *Recherche Transport et Sécurité*, N° 57, pp.67-77.
- Bonnell P., Armoogum J., 2005, « National transport surveys –What can we learn from international comparisons », *European Transport Conference 2005*, Strasbourg.
- Berthier, C., Dupont F., 1999, « L'incidence du caractère obligatoire des enquêtes », *Insee Méthodes* N°69-70-71, pp. 131-146.
- Billocq, F., 1996, « Conception et évaluation de questionnaires », *Insee Méthodes* N°69-70-71, pp. 77-92.
- Brion, P., Caron, N., Pietri-Bessy, P., 2005, « Redresser la non-réponse totale dans les enquêtes auprès des entreprises : les pièges à éviter. Illustration avec l'enquête innovation », *Document INSEE, Actes des Journées de Méthodologie Statistique*, 9p.
- Brilhault, G., Caron, N., 2004, « Correction de la non-réponse totale : par imputation ou par repondération ? », *Documents INSEE - DSE*, N° E2004/01, 76p.

- Caron, N., 1993, « Réflexion sur les erreurs de mesure : l'exemple de l'enquête conjoncture auprès des ménages », Documents INSEE - DSDS, N° F9308.
- Caron, N., 2005, « La correction de la non-réponse par repondération et par imputation », Série Documents de Travail INSEE Méthodologie Statistique, N° M0502, 48p.
- Deming, W.E., Stephan, F.F., 1940, « On a least squares adjustment of a sampled frequency table when the exal totals are known », *Annals of Mathematical Statistics*, Vol. 11, pp. 427-444.
- Deville, J.-C., 1998, « La correction de la non-réponse par calage ou par échantillonnage équilibré », Actes du colloque de la Société Statistique du Canada, Sherbrooke, Canada.
- Deville, J.-C., Särndal, C.E., 1992, « Calibration estimators and generalized raking techniques in survey sampling », *Journal of the American Statistical Association*, Vol. 87, pp. 376-382.
- Deville, J.-C. , Särndal, C.E., Sautory, O., 1993, « Generalised raking procedures in survey sampling », *Journal of the American Statistical Association*, Vol. 88, pp. 1013-1020.
- Deville, J.-C. et Särndal, C.E. (1994), « Variance estimation for the regression imputed Horvitz-Thompson estimator », *Journal of Official Statistics*, Vol. 10, N° 4 pp. 381-394.
- Deville, J.C., Dupont, F., 1996, « Non-réponse : principes et méthodes », *Insee Méthodes* N°56-57-58, pp. 53-69.
- Desrosieres A., 2003, « Historiciser l'action publique. L'Etat, le marché et les statistiques », in Laborier, P., Trom, D. (dir.), *Historicités de l'action publique*, Paris, PUF, pp. 207-221.
- Dupont, F., 1994, « Imputation procedures for quantitative and qualitative variables », document INSEE - DSDS, N° F9406.
- Glaude, M., 2000, « Les enquêtes auprès des ménages à l'Insee : petit bilan et perspective », *Courrier des statistiques* N°95-96, pp. 39-52.
- Le Guennec, J. et Sautory, O., 2003, « La macro Calmar2, manuel d'utilisation », document interne INSEE.
- Oh, H. L., Scheuren, F. J., 1983, « Weighting Adjustment for Unit non-response », in W. G. Madow, I. Olkin, and D. B. Rubin (eds), *Incomplete data in Sample Surveys*, Vol. 2: Theory and Bibliographies, New York: Academic Press, pp. 143-184.
- Platek, R., Gray, G.B., 1983, « Imputation methodology : Total survey error », in W. G. Madow, I. Olkin, and D. B. Rubin (eds), *Incomplete Data in Sample Surveys*, Vol II : Theory and Bibliographies, New York: Academic Press, pp. 249-333.

Sautory, O., 1993, « La macro CALMAR : redressement d'un échantillon par calage sur marges », INSEE Document de travail, N° F9310, 56p.

Stephan, F. F., 1942, « An iterative method of adjusting sample frequency tables when expected marginal totals are known », Annals of Mathematical Statistics, Vol. 13, pp. 166-178.

## **SITES WEB**

Centre d'Études sur les Réseaux, les Transports, l'Urbanisme et les constructions publiques (CERTU), <http://www.certu.fr/>

Institut de REcherche sur les Transports et leur Sécurité (INRETS), <http://www.inrets.fr/>

Ministère de l'Écologie, de l'Énergie, du Développement Durable et de l'Aménagement du Territoire (MEEDDAT), <http://www.developpement-durable.gouv.fr/>

## **NOTES**

Roux S., 2007, « Exploitation des bilans de collecte des enquêteurs après la vague 1 de l'Enquête Nationale Transports et Déplacements », Note de l'Insee N°2406 / DG75-F340.

## **CIGLE**

CAPI : Collecte Assistée Par Informatique.

CNIL : Commission Nationale Informatique et Liberté.

DEST : Département Economie et Sociologie des Transports.

EMD : Enquêtes Ménages Déplacements.

ENT : Enquête Nationale Transports.

ENTD 2007-08 : Enquête Nationale Transports et Déplacements 2007-08.

INRETS : Institut National de Recherche sur les Transports et leur Sécurité.

MEEDDAT : Ministère de l'Ecologie, de l'Energie, du Développement Durable et de l'Aménagement du Territoire.

MGE : Méthodologie des Grandes Enquêtes.

RP 99 : Recensement de la Population de 1999.

SDT : Suivi de la Demande Touristique.

TCM : Tableau de Composition du Ménage.



## ANNEXES

### Annexe 1 : Résultat de la modélisation du mécanisme de réponse de la première visite

<b>Number of Observations Read</b>	22724
<b>Number of Observations Used</b>	21093

Response Profile		
Ordered Value	Rep_v1	Total Frequency
1	0	4657
2	1	16436

Summary of Stepwise Selection							
Step	Effect		DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq
	Entered	Removed					
1	TYPL		1	1	469.9722		<.0001
2	NPCE		3	2	90.9579		<.0001
3	STRATE		4	3	82.0877		<.0001
4	VOIT		2	4	33.5922		<.0001
5	AGERP		3	5	20.7371		0.0001
6	HLMI		1	6	4.9031		0.0268

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
AGERP	3	19.7828	0.0002
HLMI	1	4.9011	0.0268
TYPL	1	47.5877	<.0001
NPCE	3	32.7950	<.0001
STRATE	4	81.7738	<.0001
VOIT	2	22.5648	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-1.1730	0.0300	1530.0863	<.0001
AGERP	1	1	0.0721	0.0332	4.7007	0.0302
AGERP	2	1	-0.0438	0.0280	2.4455	0.1179
AGERP	3	1	-0.1113	0.0304	13.4195	0.0002
HLMI	1	1	-0.0550	0.0249	4.9011	0.0268
TYPL	0	1	0.1707	0.0247	47.5877	<.0001
NPCE	1	1	0.2816	0.0586	23.1218	<.0001
NPCE	2	1	0.0363	0.0315	1.3326	0.2483
NPCE	4	1	-0.1305	0.0313	17.3585	<.0001
STRATE	0	1	-0.2292	0.0354	41.9891	<.0001
STRATE	1	1	0.0307	0.0433	0.5038	0.4778
STRATE	2	1	0.0285	0.0481	0.3512	0.5534
STRATE	3	1	-0.0539	0.0377	2.0408	0.1531
VOIT	0	1	0.1208	0.0327	13.6209	0.0002
VOIT	1	1	0.0146	0.0239	0.3735	0.5411

<b>Odds Ratio Estimates</b>			
<b>Effect</b>	<b>Point Estimate</b>	<b>95% Wald Confidence Limits</b>	
<b>AGERP 1 vs 4</b>	0.989	0.889	1.100
<b>AGERP 2 vs 4</b>	0.881	0.801	0.969
<b>AGERP 3 vs 4</b>	0.823	0.745	0.910
<b>HLMI 1 vs 2</b>	0.896	0.813	0.987
<b>TYPL 0 vs 1</b>	1.407	1.277	1.550
<b>NPCE 1 vs 6</b>	1.598	1.329	1.922
<b>NPCE 2 vs 6</b>	1.251	1.107	1.413
<b>NPCE 4 vs 6</b>	1.058	0.953	1.176
<b>STRATE 0 vs 4</b>	0.636	0.574	0.705
<b>STRATE 1 vs 4</b>	0.824	0.732	0.929
<b>STRATE 2 vs 4</b>	0.823	0.724	0.934
<b>STRATE 3 vs 4</b>	0.758	0.686	0.836
<b>VOIT 0 vs 2</b>	1.292	1.156	1.444
<b>VOIT 1 vs 2</b>	1.162	1.071	1.260

<b>Association of Predicted Probabilities and Observed Responses</b>			
<b>Percent Concordant</b>	61.6	<b>Somers' D</b>	0.248
<b>Percent Discordant</b>	36.8	<b>Gamma</b>	0.252
<b>Percent Tied</b>	1.5	<b>Tau-a</b>	0.085
<b>Pairs</b>	76542452	<b>c</b>	0.624

## Annexe 2 : Résultat de la modélisation du mécanisme de réponse de la deuxième visite

Number of Observations Read	20275
Number of Observations Used	19921

Response Profile		
Ordered Value	rep_v2	Total Frequency
1	0	1442
2	1	18479

Model Fit Statistics		
Criterion	Intercept Only	Intercept And Covariates
AIC	10351.650	9701.128
SC	10359.549	9946.013
-2 Log L	10349.650	9639.128

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	710.5219	30	<.0001
Score	839.8323	30	<.0001
Wald	729.7963	30	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
13.1898	8	0.1055

Summary of Stepwise Selection							
Step	Effect		DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq
	Entered	Removed					
1	REVMEN		4	1	346.5998		<.0001
2	sfvkm1an		5	2	168.1596		<.0001
3	DIFF_V1_DF		5	3	160.8887		<.0001
4	SEXKKB		1	4	45.0773		<.0001
5	STRATE		4	5	54.3300		<.0001
6	Q3		5	6	42.2695		<.0001
7	NBVEHTOT		2	7	8.0448		0.0179
8	typmen5		4	8	10.7793		0.0292

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
STRATE	4	49.5971	<.0001
SEXKKB	1	43.2670	<.0001
typmen5	4	10.7494	0.0295
REVMEN	4	249.2900	<.0001
sfvkm1an	5	153.3712	<.0001
Q3	5	40.9228	<.0001
NBVEHTOT	2	8.7516	0.0126
DIFF_V1_DF	5	168.3241	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-2.2035	0.0866	646.8054	<.0001
STRATE	0	1	-0.1431	0.0520	7.5581	0.0060
STRATE	1	1	0.0834	0.0675	1.5240	0.2170
STRATE	2	1	0.1833	0.0786	5.4385	0.0197
STRATE	3	1	0.1963	0.0532	13.5935	0.0002
SEXKKB	1	1	0.1885	0.0287	43.2670	<.0001
typmen5	1	1	-0.1860	0.0669	7.7236	0.0055
typmen5	2	1	0.1062	0.0853	1.5498	0.2132
typmen5	3	1	-0.0905	0.0597	2.2980	0.1295
typmen5	4	1	0.0141	0.0622	0.0518	0.8200
REVMEN	1	1	-0.0256	0.0631	0.1643	0.6852
REVMEN	2	1	-0.2035	0.0529	14.7865	0.0001
REVMEN	3	1	-0.3981	0.0657	36.7581	<.0001
REVMEN	4	1	-0.3258	0.0651	25.0192	<.0001
sfvkm1an	1	1	-0.0392	0.0913	0.1839	0.6680
sfvkm1an	2	1	-0.4418	0.0735	36.1392	<.0001
sfvkm1an	3	1	-0.4281	0.0791	29.2820	<.0001
sfvkm1an	4	1	-0.0761	0.0673	1.2785	0.2582
sfvkm1an	5	1	0.5905	0.0711	69.0608	<.0001
Q3	0	1	0.1831	0.0528	12.0257	0.0005
Q3	1	1	-0.1992	0.0651	9.3793	0.0022
Q3	2	1	-0.2297	0.0852	7.2663	0.0070
Q3	3	1	-0.0572	0.0977	0.3427	0.5583
Q3	4	1	-0.0290	0.0824	0.1237	0.7251
NBVEHTOT	0	1	-0.3431	0.1573	4.7555	0.0292
NBVEHTOT	1	1	0.2304	0.0811	8.0634	0.0045
DIFF_V1_DF	0	1	1.1131	0.2696	17.0496	<.0001
DIFF_V1_DF	1	1	0.5073	0.0815	38.6993	<.0001
DIFF_V1_DF	2	1	-0.2535	0.0775	10.6875	0.0011
DIFF_V1_DF	3	1	-0.4316	0.0748	33.2653	<.0001
DIFF_V1_DF	4	1	-0.4179	0.0756	30.5179	<.0001

<b>Odds Ratio Estimates</b>			
<b>Effect</b>	<b>Point Estimate</b>	<b>95% Wald Confidence Limits</b>	
STRATE 0 vs 4	1.193	1.006	1.415
STRATE 1 vs 4	1.497	1.224	1.830
STRATE 2 vs 4	1.654	1.323	2.069
STRATE 3 vs 4	1.676	1.417	1.981
SEXEKB 1 vs 2	1.458	1.303	1.631
typmen5 1 vs 5	0.710	0.508	0.993
typmen5 2 vs 5	0.951	0.663	1.366
typmen5 3 vs 5	0.782	0.565	1.081
typmen5 4 vs 5	0.868	0.628	1.199
REVMEN 1 vs 6	0.376	0.312	0.452
REVMEN 2 vs 6	0.315	0.264	0.375
REVMEN 3 vs 6	0.259	0.212	0.317
REVMEN 4 vs 6	0.278	0.228	0.339
sfvkm1an 1 vs 6	0.648	0.402	1.045
sfvkm1an 2 vs 6	0.433	0.273	0.686
sfvkm1an 3 vs 6	0.439	0.276	0.700
sfvkm1an 4 vs 6	0.625	0.396	0.986
sfvkm1an 5 vs 6	1.216	0.769	1.923
Q3 0 vs 5	0.862	0.678	1.096
Q3 1 vs 5	0.588	0.455	0.760
Q3 2 vs 5	0.570	0.428	0.760
Q3 3 vs 5	0.678	0.498	0.923
Q3 4 vs 5	0.697	0.527	0.923
NBVEHTOT 0 vs 2	0.634	0.393	1.024
NBVEHTOT 1 vs 2	1.125	0.976	1.297
DIFF_V1_DF 0 vs 5	5.106	2.600	10.030
DIFF_V1_DF 1 vs 5	2.786	2.116	3.668
DIFF_V1_DF 2 vs 5	1.302	0.996	1.702
DIFF_V1_DF 3 vs 5	1.089	0.838	1.417
DIFF_V1_DF 4 vs 5	1.105	0.848	1.438

<b>Association of Predicted Probabilities and Observed Responses</b>			
<b>Percent Concordant</b>	68.8	<b>Somers' D</b>	0.390
<b>Percent Discordant</b>	29.8	<b>Gamma</b>	0.396
<b>Percent Tied</b>	1.4	<b>Tau-a</b>	0.052
<b>Pairs</b>	26646718	<b>c</b>	0.695

