

ECOLE NATIONALE DE LA STATISTIQUE
ET DE L'ANALYSE DE L'INFORMATION
1996

Groupe de statistique appliquée

L'ENQUETE « TRANSPORTS ET COMMUNICATIONS »
1993/94
LA MOBILITE DES MENAGES SUR LONGUE DISTANCE

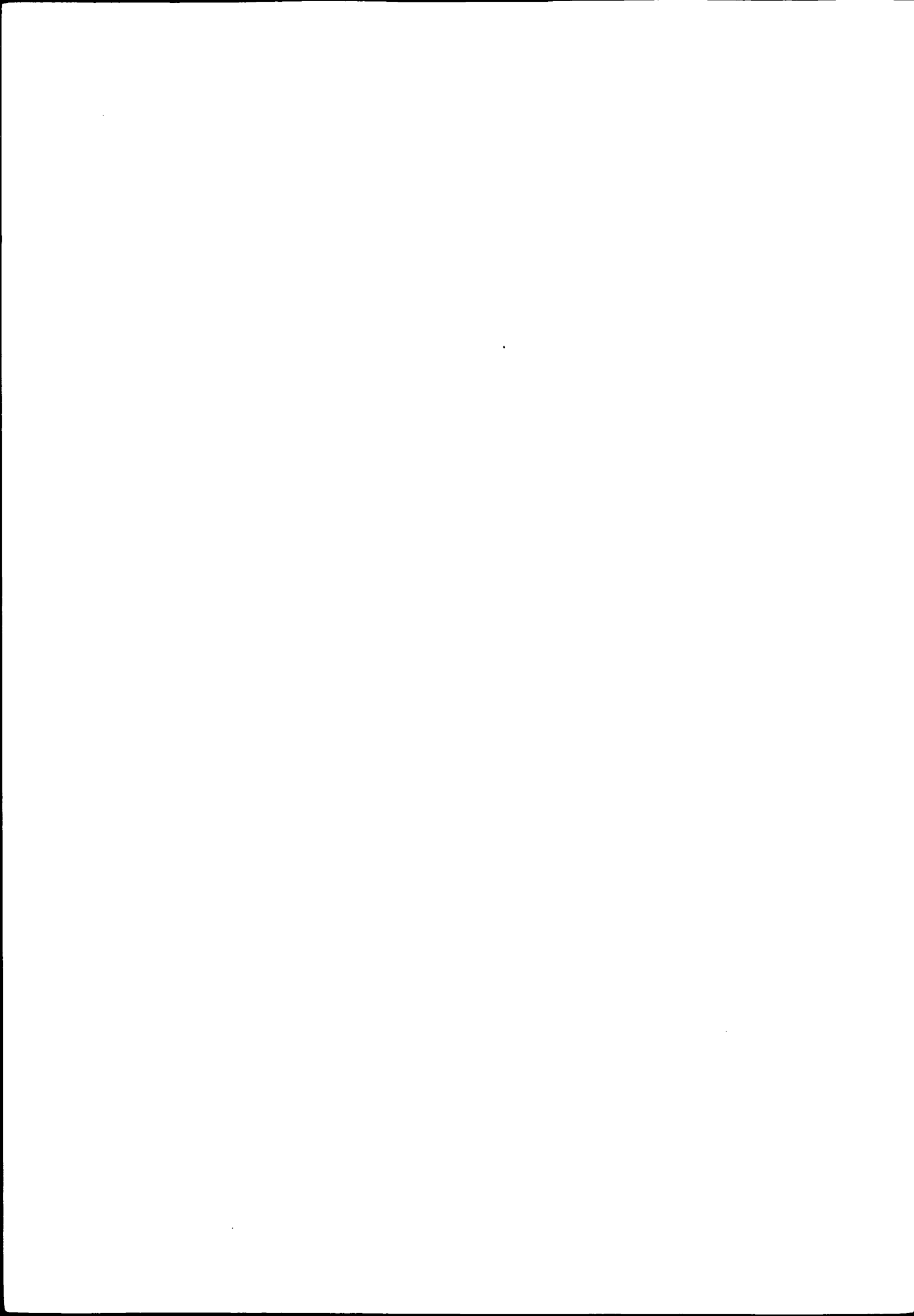
Rapport

Etudiants :

Pierrette BRIANT
Annie CHANUT
Brigitte GELEIN
François RAGEAU

Animateurs :

M. C. CALZADA
Mme N. GOUIDER
Service Economique et
Statistique de la DAEI



ECOLE NATIONALE DE LA STATISTIQUE
ET DE L'ANALYSE DE L'INFORMATION
1996

Groupe de statistique appliquée

L'ENQUETE « TRANSPORTS ET COMMUNICATIONS »
1993/94
LA MOBILITE DES MENAGES SUR LONGUE DISTANCE

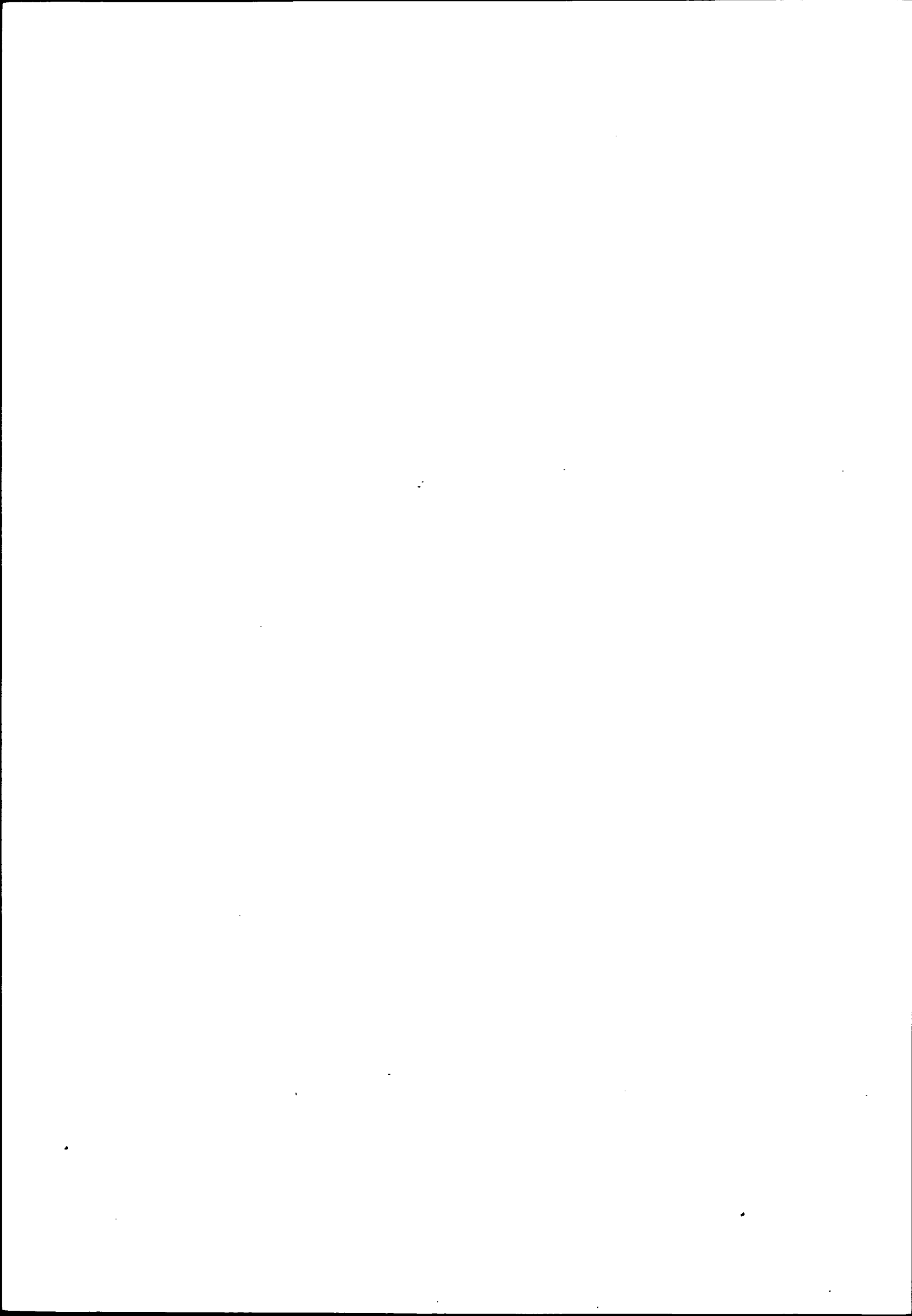
Note de synthèse

Etudiants

Pierrette BRIANT
Annie CHANUT
Brigitte GELEIN
François RAGEAU

Animateurs

M. C. CALZADA
Mme N. GOUIDER



L'enquête Transports et Communications de 1993/94 ¹ avait comme objectif de fournir une aide à la décision par l'étude des raisons qui amènent les ménages à se déplacer et la façon dont ils le font. Le but du travail réalisé dans le cadre du Groupe de Statistique Appliquée fut précisément de décrire l'intensité de mobilité et les modalités des déplacements, et d'en connaître les déterminants socio-économiques. Une approche non paramétrique du fichier -avec en préalable une présentation de l'enquête et des données -, puis une approche paramétrique ont été adoptées.

Les méthodes d'analyse non paramétrique que sont l'analyse des données, la segmentation et la classification, ont permis l'élaboration de typologies d'individus. Ainsi, l'analyse des données a offert la possibilité de créer des typologies d'individus en fonction de leur signalétique d'une part, et, d'autre part, en fonction des variables de déplacement. Les différentes Analyses des Correspondances Multiples ont notamment permis de montrer qu'un comportement de mobilité réduite est surtout le fait de personnes relativement peu favorisées socialement et économiquement (pas de permis de conduire, niveau d'étude égal au CEP faible revenu ...).

La segmentation, par contre, ne repose pas sur une différence de traitement entre variables de déplacement et variables de signalétique. Elle est fondée sur une distinction entre variable expliquée (le nombre de déplacements comme indicateur d'intensité de mobilité) et variables explicatives qui appartiennent tant au signalétique qu'au descriptif des déplacements. Son intérêt par rapport à l'ACM est de définir clairement l'appartenance d'un individu à une partie de la population. L'inconvénient majeur de cette méthode a été l'utilisation exclusive de variables de signalétique comme variables explicatives lorsque l'on n'a pas retenu un nombre de noeuds terminaux suffisamment élevé.

Enfin, la classification n'impose pas de distinction a priori entre les deux types de variables. De plus, elle a fait intervenir de façon équilibrée, dans la détermination des classes, les variables de mobilité et de signalétique. Elle a aussi permis une description des; personnes se déplaçant pour des motifs professionnels exclusivement - contrairement à la segmentation. Par ailleurs, la variable la plus déterminante dans la constitution des classes a été le mode de transport, ce qui est particulièrement intéressant, dans le cadre d'une analyse des choix modaux.

Les résultats obtenus par ces trois méthodes sont d'intérêt inégal. En effet, si l'ACM et la classification ont fourni des typologies assez complètes, cohérentes, et répondant bien à la problématique, la segmentation, elle, offre des résultats relativement pauvres et difficilement interprétables en terme d'intensité de mobilité.

Les méthodes non paramétriques appliquées présentent l'avantage de détecter des relations sans les fixer dans le cadre rigide d'une spécification particulière, à l'inverse des méthodes économétriques. Mais ce que ces dernières perdent ainsi en généralité, (elles nécessitent des hypothèses souvent contraignantes), elles le gagnent en possibilité d'isoler les effets propres des facteurs intervenant dans l'explication du phénomène. La 2ème partie

¹ Dont l'exploitation concernant les déplacements à longue distance fut confiée à l'ex-Observatoire Economique et Statistique, devenu depuis le Service Economique et Statistique.

s'inscrit dans cette perspective, en reprenant la problématique générale de l'analyse des comportements individuels et de ses modalités.

Deux approches ont été retenues pour l'analyse des facteurs de la mobilité. En la mesurant par le nombre de déplacements, l'on se trouve en présence d'un problème statistique de distribution tronquée, formalisé par le modèle TOBIT. Son estimation s'est effectuée en recourant à la méthode d'Heckman. La liste des variables explicatives a été suggérée par les statistiques descriptives préalablement établies. Celles-ci révèlent l'image d'une population socialement privilégiée parmi les mobiles, qui disposent de ce fait de possibilités de transports variés et flexibles leur permettant de se déplacer sans problème.

La première étape de la méthode d'Heckman semble indiquer que la mobilité est déterminée par trois types de facteurs :

- * un facteur d'ordre économique (revenu), car la mobilité de longue distance implique un coût financier ,
- * des facteurs d'ordre sociologique et culturel (capital relationnel, culturel accumulé par les catégories socialement privilégiées) ,
- * des facteurs liés à l'équipement du ménage et à son environnement de vie : l'équipement automobile est décisif dans un contexte où il assure près des trois quarts des déplacements sur longue distance (hors déplacements professionnels).

On peut affirmer ex-post que la procédure d'Heckman était justifiée, compte tenu de la significativité du ratio de Mill. Reprenant les variables explicatives de la première étape, nombre d'entre elles ne sont plus significatives suggérant une plus faible hétérogénéité au sein de la population des mobiles qu'entre les mobiles et les non mobiles.

Ces considérations sur l'importance de l'approche dichotomique par rapport à une approche quantitative pour analyser la mobilité nous ont incités à tester une voie médiane : un modèle d'intensité de mobilité trichotomique ordonné. Celui-ci permet de nuancer le modèle dichotomique, sans le remettre en cause. Par ailleurs, il est d'un "emploi" plus aisé que le modèle TOBIT, relativement lourd à mettre en oeuvre et vraisemblablement plus adapté au phénomène étudié, pour lequel il existe sans aucun doute des effets de seuil.

Lorsque les individus se sont déplacés, il est intéressant de savoir dans quelles conditions ils l'ont fait et plus particulièrement quel mode de transport ils ont utilisé. Cette analyse des choix modaux a été menée en deux étapes. La première étape distingue le choix de la voiture du choix d'un moyen de transport collectif. Les résultats du modèle LOGIT dichotomique indiquent un rôle important des variables indiquant le degré de motorisation du ménage, ce qui semble assez naturel. On remarque aussi que la voiture a la préférence des individus en activité, ayant le statut d'employé ou ouvrier et habitant une ville moyenne.

Les usagers des transports collectifs (avion, car et train) forment une population plus hétérogène dont nous avons voulu connaître plus en détail les comportements de choix. Toutefois la mise en oeuvre d'une modélisation polytomique non ordonnée s'est révélée plus fragile que pour les modèles LOGIT dichotomiques. Les résultats semblent plus délicats à interpréter, ce qui a motivé une discussion autour des conditions d'estimation de ce modèle, notamment de l'indépendance vis-à-vis des choix non pertinents.

SOMMAIRE

	Pages
Introduction	3
Présentation de l'enquête	5
1ère partie : ANALYSE NON PARAMETRIQUE	
Analyse factorielle, segmentation, classification	8
I. Analyse factorielle des données.....	8
1.1. Le signalétique en variables actives	9
1.2. Les déplacements en variables actives.....	11
II. Une analyse par arbre de discrimination : la segmentation.....	13
2.1. Segmentation sur la population totale	13
2.2. Segmentation sur les personnes mobiles.....	15
2.3. Segmentation sur la population des mobiles avec distinction des motifs de déplacement	15
III. Classification ascendante hiérarchique	18
2ème partie : ANALYSE PARAMETRIQUE	
Estimation micro-économétrique de la mobilité sur longue distance et des choix modaux de transport	20
I. Les déterminants de la mobilité privée et de son intensité	20
1.1. Définition de la mobilité et de son intensité.....	20
1.2. Les modélisations retenues.....	21
1.3. Profil de la population selon son degré de mobilité.....	22
1.4. Estimation du modèle TOBIT de mobilité.....	23
1.5. Modélisation trichotomique ordonnée de la mobilité	27
II. Déplacements privés : quel mode de transport ?	30
2.1. Voiture contre transports collectifs	31
2.2. Quel mode de transport collectif ?	34
Conclusion	38
Bibliographie	39
Annexes	40



INTRODUCTION

Cadre de travail

Cette étude repose sur les données de l'enquête *Transports et communications de 1993/94* dont une partie de l'exploitation (déplacements sur longue distance) a été confiée à l'Observatoire Economique et Statistique des Transports. Depuis le début de l'année 1996, cet observatoire est désormais le "Service Economique et Statistique" (SES) de la Direction des Affaires Economiques et Internationales (DAEI) au sein du Ministère de l'Equipement, du Logement, des Transports et du Tourisme (MELTT).

Aux missions traditionnelles sur le transport de l'ex-OEST se sont ajoutées de nouvelles missions en terme d'habitat et d'économie spatiale. Outre la production et la diffusion de l'information économique et statistique dans le domaine des transports, le service participe activement à tous les travaux de coordination et de réflexion visant à améliorer les systèmes d'information sur le secteur (schéma directeur d'aménagement du territoire).

Les deux animateurs du groupe de travail appartiennent au SES : M. CALZADA (chargé d'études au DEE¹) et Mme GOUIDER (responsable du pôle voyageurs à ST1). La mise à disposition des données sous forme informatique a été assurée par Mme DENIZOU-LETU (DIM3).

Thème de l'étude et problématique

Les transports représentent un secteur important de l'économie. Ils sont un passage obligé pour la plupart des activités de production et de consommation. Notre champ d'étude se restreindra aux trajets effectués par les ménages résidants et plus particulièrement à ceux qui les éloignent de plus de 100 kilomètres de leur domicile. Plusieurs dimensions peuvent en effet retenir l'attention comme le trafic engendré par ces déplacements ou le temps qu'ils occupent dans la vie des ménages.

Connaître les raisons qui amènent les ménages à se déplacer et la façon dont ils le font est une information précieuse pour plusieurs intervenants économiques tant publics que privés. Il s'agit d'outils d'aide à la décision pour tous les organismes intéressés par les questions d'aménagement du territoire (collectivités territoriales, DATAR, autres instances de l'Etat français et de l'Union Européenne) mais aussi pour les acteurs du marché des transports tels que la SNCF, le Comité des constructeurs français d'automobiles, les compagnies aériennes ou encore les sociétés chargées de la gestion des autoroutes.

Une des principales sources d'information structurelle sur le domaine est constituée par les enquêtes Transports réalisées environ tous les dix ans et dont l'INSEE est maître d'oeuvre. La dernière de ces enquêtes a eu lieu en 1993/94 et a déjà fait l'objet d'exploitations statistiques pour le domaine des déplacements de courte distance. Par contre, l'exploitation des données recueillies sur les longs trajets commence cette année. Cet aspect est d'autant plus important que les données de la précédente enquête de 1982 sur le sujet ont été peu utilisées et que la demande d'information est importante comme on a pu le voir précédemment.

¹ DEE : Département des Etudes Economiques

ST1 : Statistiques Transports Bureau 1

DIM3 : Département de l'Informatique et des Moyens

Dans ce cadre de travail, notre problématique sera précisément de décrire l'intensité et les modalités de déplacement et d'en déterminer les facteurs individuels socio-économiques (voire environnementaux).

Cependant, les concepts utilisés, la masse et l'articulation des informations recueillies font de l'enquête Transports et Communications une source de données particulièrement riche mais complexe à exploiter. Une réflexion méthodologique sur les outils statistiques les plus adéquats à "implémenter" est donc au coeur de notre démarche. Dans cette perspective et dans le cadre des possibilités techniques disponibles (SAS, allocation de mémoire limitée), deux orientations de travail radicalement différentes mais complémentaires ont été déterminées.

L'une s'appuie sur les méthodes d'analyse non paramétriques : analyse des données, classification et segmentation. En particulier, cette dernière, courante en analyse marketing, est relativement nouvelle au sein du SSP (Système Statistique Public). En présentant des profils de publics caractéristiques de comportement de mobilités, elle peut fournir un instrument d'aide à la décision pour certains commanditaires de l'enquête.

Le scoring (ou modèles de choix discrets) constitue en revanche une approche privilégiée des études menées au sein du SSP en raison de la vocation d'analyse de phénomènes multivariés et de leur confrontation sous une forme quantifiée.

PRESENTATION DE L'ENQUETE

Participants

Les enjeux décrits en introduction expliquent le nombre important d'associés pour cette opération : environ une vingtaine d'organismes ont participé à la préparation et à la réalisation de l'enquête. Aux partenaires traditionnels tels que l'INSEE, l'ex-OEST, l'INRETS (Institut National de Recherche et d'Etude sur les Transports et leur Sécurité) ou la DATAR, se sont ajoutés France-Télécom pour la prise en compte des besoins des ménages en matière de communication (d'où la double appellation de l'enquête), et EDF pour le marché potentiel du véhicule électrique.

Principaux concepts

Avant de présenter plus en détail l'enquête il est nécessaire de préciser les principaux concepts auxquels elle fait référence et qui serviront tout au long de l'étude.

déplacement

C'est le mouvement d'une personne d'un lieu de départ (origine) à un lieu d'arrivée (destination). La notion de déplacement est intimement liée à celle de motif : tout changement de motif entraîne un changement de déplacement.

voyage

Le voyage qui conduit l'enquêté à plus de 80 kilomètres à vol d'oiseau de son domicile (soit environ 100 km sur un réseau de transport) est un ensemble de déplacements effectués par une personne sans qu'elle ne revienne au domicile entre temps. Un voyage est toujours composé au minimum de deux déplacements : un aller et un retour.

motif

Le motif est la raison pour laquelle le déplacement a été effectué. On distingue deux grandes catégories de motifs : professionnels (se rendre à son lieu de travail ou à une conférence par exemple) et privés (visite à des parents ou vacances par exemple). Aller prendre un moyen de transport (ce que l'on appelle le rabattement) ne constitue pas en soi un motif de déplacement. Au sein d'un voyage, il y aura toujours un déplacement dont le motif est codé "retour au domicile" (a priori on rentre toujours chez soi).

moyen de transport principal

C'est le moyen de transport qui correspond à la plus grande distance parcourue lors du déplacement.

Objectifs

Comme pour les trois enquêtes Transports précédentes (1966/67, 1973/74 et 1981/82), l'objectif de l'enquête de 1993/94 est de fournir des informations structurelles sur la mobilité des ménages et leur usage des moyens de transports tant individuels que collectifs. Un des points forts de cette enquête est de fournir une vision complète des habitudes et des pratiques des ménages. En effet, y sont décrits tous les déplacements quels qu'en soient le motif, le mode de transport, la longueur et la période.

Les informations collectées comprennent, outre les données de cadrage du ménage et des individus interrogés, des éléments d'appréciation de l'équipement des ménages en véhicules, de leurs déplacements et des conditions dans lesquelles ils s'effectuent. L'accent a été mis sur l'utilisation de la voiture individuelle (remplissage d'un carnet voiture) compte tenu de l'importance croissante de ce moyen de transport. Les déplacements sur longue distance ont aussi fait l'objet d'une attention particulière avec un questionnaire spécial qui leur est consacré.

L'enquête n'aborde pas les questions de coûts liés aux transports.

Déroulement

La collecte a été répartie sur huit phases durant une année complète de mai 1993 à avril 1994. Cette durée de collecte visait à neutraliser les effets de saisonnalité influençant la mobilité des ménages particulièrement pour les déplacements pour motif privé.

L'enquêteur effectuait deux visites espacées d'au minimum huit jours (le carnet voiture, déposé lors de la première entrevue, concerne sept jours). Concernant les aspects liés aux déplacements sur longue distance, une seule personne a été interrogée. Elle devait décrire ses déplacements sur une période totale de six mois : pour les trois mois précédant le passage de l'enquêteur et les trois mois suivants. Pour chaque déplacement, les principales informations recueillies sont :

- lieux de départ et d'arrivée ;
- motif principal ;
- moyen de transport principal ;
- distance parcourue ;
- nombre de personnes accompagnant l'enquêté.

Le questionnaire concernant les trois mois suivant le passage de l'enquêteur devait être retourné par voie postale à la Direction régionale de l'INSEE concernée.

Echantillonnage et tirage des personnes interrogées

Un échantillon de 20 000 logements a été tiré dans la base du recensement de population de 1990 complétée par celle des logements neufs (construits depuis). Il s'agit d'un tirage à trois degrés avec probabilités inégales. En effet, afin de recueillir le maximum d'information sur les déplacements, les ménages possédant deux voitures et plus ont été sur-représentés car ils sont aussi les plus mobiles (sauf pour le cas de Paris où la forte densité constitue un frein à la motorisation). Cette distinction a été réalisée à partir des données du recensement de 1990 sur l'équipement des ménages en automobile. Bien que cette information figure depuis 1968 dans les recensements (RP), c'était la première fois qu'elle était utilisée pour stratifier un échantillon d'enquête sur les comportements de transport.

Pour le tirage du membre du ménage chargé de répondre au questionnaire "longue distance" la personne la plus mobile du ménage (ayant effectué le plus de déplacements à longue distance) a une probabilité de 2/3 d'être tirée. Les personnes susceptibles d'être tirées au sort doivent remplir les conditions suivantes :

- être âgées de six ans au moins (dans ce cas les proches peuvent aider à remplir le questionnaire) ;
- être présents sur leur résidence principale pendant la durée de l'enquête ;
- être aptes physiquement et intellectuellement à répondre.

Redressement, traitement des non-réponses

Sur l'échantillon initial de 20 053 fiches adresses, 13,3 % des adresses étaient hors champ (résidence secondaire ou occasionnelle, logement vacant) et on dénombrait 15,8 % de non-réponses totales. On aboutit ainsi à 14 213 ménages interrogés pour lesquels environ 60 % des questionnaires "longue distance" seront renvoyés.

Le traitement des non-réponses s'est fait par une méthode de repondération. En utilisant l'information auxiliaire issue du recensement de 1990, un modèle logit a permis de mettre en évidence les variables expliquant le mécanisme de réponse. Les variables ainsi retenues sont : la strate (taille de la commune d'habitation), le nombre de personnes du ménage, le nombre de voitures et l'âge de la personne de référence. Ces variables ont ensuite servi à post-stratifier l'échantillon de répondants. Les poids initiaux sont ainsi corrigés par les taux de réponse selon la catégorie à laquelle le ménage appartient.

L'erreur d'échantillonnage a été traitée par la méthode du calage sur marges (avec CALMAR) selon plusieurs variables : des variables socio-démographiques, selon les régions de résidence (22 régions) et suivant six équations de calage temporel.

On peut souligner que ce type d'enquête pose des problèmes de représentativité temporelle. Il convient en effet que chaque journée soit également représentée ce qui n'est pas le cas du fait, d'une part des poids différents attribués aux individus répondant au questionnaire "longues distances" et d'autre part, des irrégularités dans la répartition des jours d'enquête.

Données utilisées pour l'étude

La taille importante de l'échantillon nous contraint à travailler avec les outils fournis par le logiciel SAS (bien que d'autres outils existent notamment pour la segmentation). De plus, des contraintes en terme de taille de fichier exploitable, nous ont été imposées. Nous avons donc choisi de réduire le nombre de variables utilisées plutôt que la taille de l'échantillon (ce qui n'aurait eu aucun sens du point de vue de la segmentation). Trois tables SAS sont exploitées :

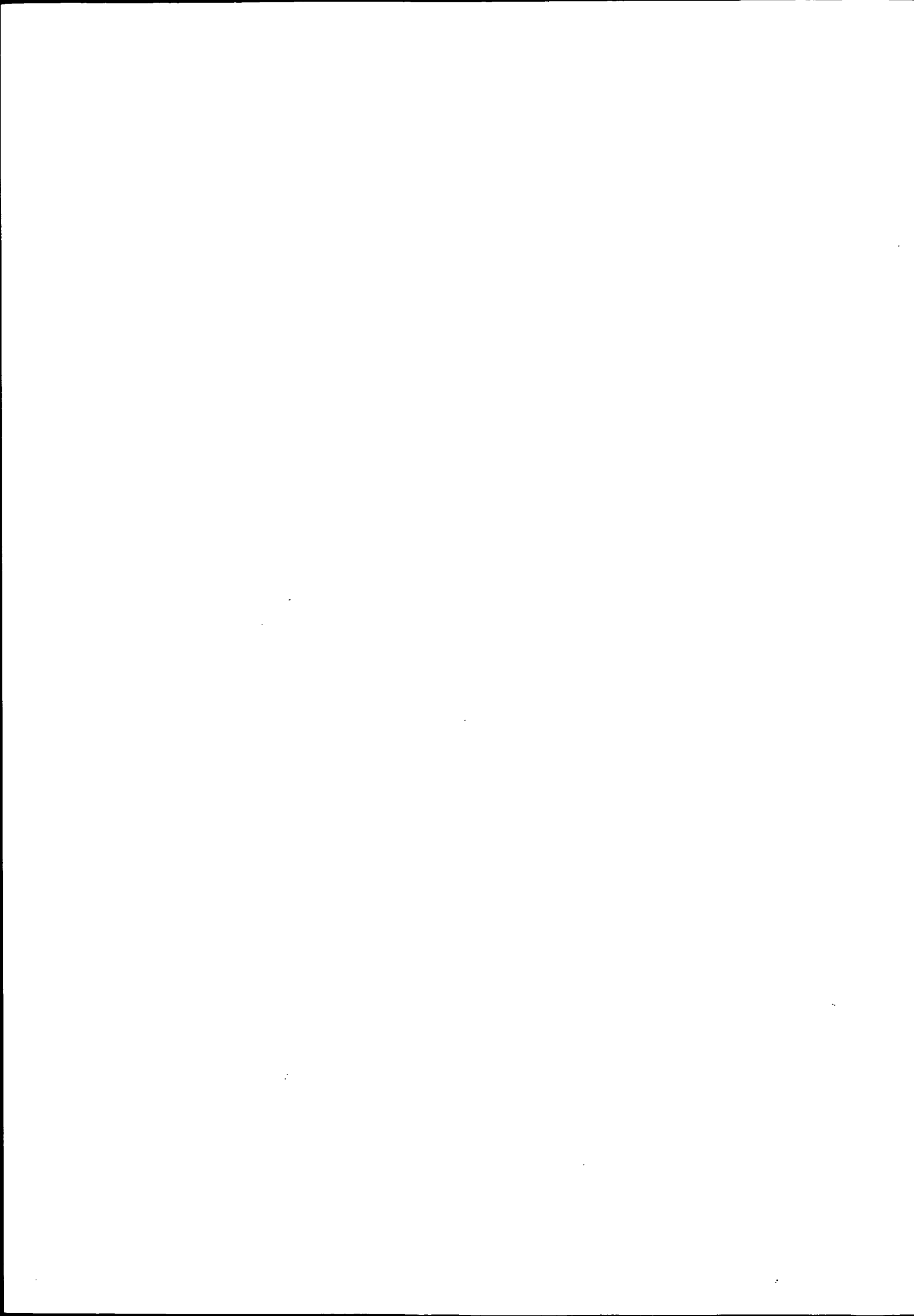
Nom	Contenu	Nombre d'observations	Nombre de variables
MENAGES	Caractéristiques du ménage	14213	9 sélectionnées parmi 195 collectées.
PERSO	Caractéristiques de la personne ayant répondu au questionnaire "longues distances"	14207	17 sélectionnées parmi 176 collectées.
LONGDIST	Caractéristiques des déplacements effectués lors d'un voyage de longue distance par la personne interrogée	41774	10 sélectionnées parmi 66 collectées.

Les tables peuvent être reliées entre elles par une variable qui leur est commune : l'identifiant individuel de la personne enquêtée.

Une première concaténation entre les tables MENAGES et PERSO est facilement réalisable du fait de la correspondance immédiate entre les deux. Le rapprochement de cette nouvelle table avec la table LONGDIST est plus délicate à opérer en raison de la différence de dimension : certains identifiants figurent plusieurs fois dans cette table, autant de fois que l'individu en question s'est déplacé, tandis que d'autres, ne s'étant pas déplacés n'y apparaissent pas. Puisque notre propos se place du point de vue de l'individu, il est naturel de ramener le fichier LONGDIST à la même dimension que celui contenant les variables relatives à l'individu et au ménage auquel il appartient.

Cette opération ne pose pas de problème pour les variables quantitatives (telles que les distances parcourues ou la durée des déplacements) qui peuvent être ramenées à une dimension unique grâce à des indicateurs statistiques simples (moyenne, maximum ou total par exemple). En revanche, synthétiser l'information contenue dans les variables qualitatives telles que le motif ou le mode de transport nécessite une démarche plus construite. L'idée retenue a consisté à sélectionner un déplacement particulier qui soit le plus représentatif et le plus porteur d'information possible au regard du comportement de l'individu.

La procédure de sélection associe deux critères : le choix du mode de transport et la distance parcourue. Dans une première étape, on a déterminé le moyen de transport le plus fréquemment utilisé. Dans une deuxième étape, parmi tous les déplacements pour lesquels l'individu a utilisé ce mode, on a retenu celui qui correspondait à la plus grande distance. Si, dès la première étape, il y avait égalité entre deux modes de transport, là aussi la sélection s'est faite par la distance maximale. Enfin, si des ex-aequo demeuraient lors de la deuxième étape, le déplacement type retenu était celui de dernier rang. Celui-ci, correspondant au déplacement le plus récent, est probablement le plus fiable en terme de qualité des informations restituées (effet mémoire).



1ère partie : ANALYSE NON PARAMETRIQUE

Analyse factorielle, segmentation, classification

I. ANALYSE FACTORIELLE DES DONNEES

L'ACM (Analyse des Correspondances Multiples) permet d'étudier une population d'individus décrits par des variables qualitatives, elle permet aussi l'étude de la liaison entre plusieurs variables qualitatives. Les variables disponibles peuvent être classées en deux groupes :

- celles décrivant les caractéristiques sociales de la personne interrogée (le signalétique),
- celles correspondant aux caractéristiques du déplacement moyen.

C'est pourquoi deux types d'ACM ont été réalisées, qui se complètent pour l'analyse du phénomène étudié afin de répondre à la question "qui se déplace avec quel comportement de mobilité ?".¹

D'une part, nous avons mis en oeuvre des ACM ayant comme variables actives celles qui caractérisent les déplacements et comme variables supplémentaires celles qui décrivent la personne interrogée. Cela permet de construire une typologie des individus selon leur profil de déplacements (1), d'étudier les liaisons entre les différentes variables concernant les déplacements et, avec les différentes variables supplémentaires, d'analyser la liaison entre chaque variable du signalétique, prise séparément, et les principaux facteurs de différenciation des profils de déplacement.

D'autre part, nous avons réalisé des ACM ayant comme variables actives celles décrivant la personne interrogée et comme variables supplémentaires celles portant sur les déplacements. Il s'agit cette fois d'obtenir une typologie des individus suivant leur signalétique (2), d'étudier les liaisons entre les différentes variables du signalétique, et, avec les variables supplémentaires d'étudier l'association entre le profil signalétique général et les caractéristiques de déplacements.

Nous avons, par ailleurs, pour chaque type d'ACM, réalisé trois analyses différentes, et ce dans l'objectif d'étudier plus précisément certains sous-groupes de la population :

Groupe A : individus non-mobiles et mobiles s'étant déplacés pour des motifs de retour au domicile (modalité mdom) ou autres motifs privés (modalités msoc, mlois),

Groupe B : individus mobiles avec les mêmes motifs que ci-dessus (mdom, msoc, lois),

Groupe C : individus mobiles s'étant déplacés pour des motifs professionnels (modalité mprof).

En fait, les résultats du groupe A sont très proches de ceux obtenus pour l'ensemble de la population (mobiles + non mobiles), ce qui s'explique par la faiblesse de la part des individus se déplaçant pour un motif d'ordre professionnel dans la population totale (environ 5%). Nous ne présenterons donc pas ici les résultats portant sur l'ensemble de la population.

Par ailleurs, la distinction faite entre motifs de déplacement d'ordre privé et motifs d'ordre professionnel répond à des logiques comportementales sous-jacentes différentes : les marges de manoeuvre, et donc de choix des individus sont plus contraintes dans le domaine des déplacements professionnels. Les résultats du groupe C figurent en annexe car ce groupe, d'effectif trop faible, n'a pas fait l'objet d'une analyse particulière ni en segmentation, ni en analyse paramétrique.

¹ Pour les aspects méthodologiques on pourra se reporter à l'annexe XX.

1.1. Le signalétique en variables actives

1.1.1. Sur les individus non mobiles et mobiles se déplaçant pour des motifs privés ou des retours au domicile

AXE 1 : axe social et familial

Les variables ayant le plus contribué à la constitution de l'axe sont : l'âge (CTR=10,5), l'occupation actuelle BIBIS (CTR=8,7), la possession du permis de conduire E1BIS (CTR=9,6), le nombre de voitures G1 (CTR=13,4), le nombre d'enfants NBENF (CTR=10,1), le nombre de permis de conduire du ménage NBPERVO (CTR=15,4), le type de ménage TYPMEN (CTR=12,8).

Pour les variables supplémentaires, chacune d'elles présente une ou deux modalités pour lesquelles la valeur V_{test} est très élevée en valeur absolue.

L'axe 1 oppose les personnes âgées (age4), inactives, sans permis de conduire, sans enfant, seules (coordonnées positives) aux personnes jeunes (age2), actives, possédant le permis de conduire, 2 voitures ou plus, ayant 2 enfants ou plus et vivant en couple (coordonnées négatives).

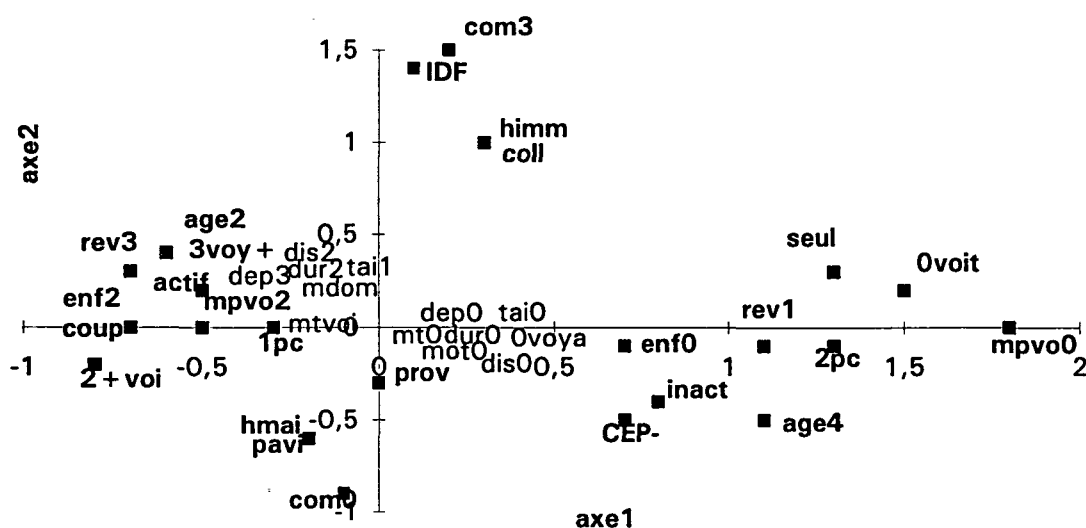
Cette deuxième catégorie voyage plus que la première (3 voyages ou plus en moyenne), se déplace davantage en voiture, effectue des déplacements plus nombreux avec une plus grande taille du groupe qui accompagne la personne interrogée.

AXE 2 : axe du domicile

Les variables actives ayant le plus contribué à la constitution de l'axe sont: le département du domicile DEPDOM (CTR=15,3), le type d'habitat du ménage Q1 (CTR=19,4), l'environnement du logement Q2 (CTR=18,4), la taille de la commune du domicile STRATE (CTR=19,8).

Contrairement au premier axe, l'axe 2 présente quelques modalités dont la valeur V_{test} est inférieure à 2 en valeur absolue : dc2 pour la variable DCTOT, 1voya pour la variable L02, mlois pour la variable MOTM. Mais pour chacune de ces variables les autres modalités peuvent avoir une valeur V_{test} très supérieure à 2 en valeur absolue.

Cet axe oppose les personnes résidant en province et en zone rurale, possédant une maison dans un quartier pavillonnaire (coordonnées négatives) à celles résidant en Ile de France, en immeuble, dans un quartier où les logements sont collectifs (coordonnées positives). Ces dernières se déplacent relativement plus, plus longtemps, davantage en transport en commun.



1.1.2. Sur les individus mobiles se déplaçant pour des motifs privés ou des retours au domicile

AXE 1 : effectif du ménage et automobile

Les variables actives ayant le plus contribué à l'axe sont le nombre de voitures du ménage G1 (CTR=13,3), le nombre d'enfants NBENF (CTR=10,9), le nombre de permis de conduire du ménage NBPERVO (CTR=15,3), le type du ménage TYPMEN (CTR=13,4). Cette prééminence des variables concernant l'automobile s'explique par le fait que les personnes mobiles utilisent surtout la voiture d'après la première ACM.

Les variables supplémentaires dont les modalités ont, dans l'ensemble, des valeurs Vtest bien supérieures aux autres (en valeur absolue), sont le mode de transport principal MTPM, le nombre de déplacements NBDEP et la taille du groupe qui accompagne la personne interrogée TAILLEM.

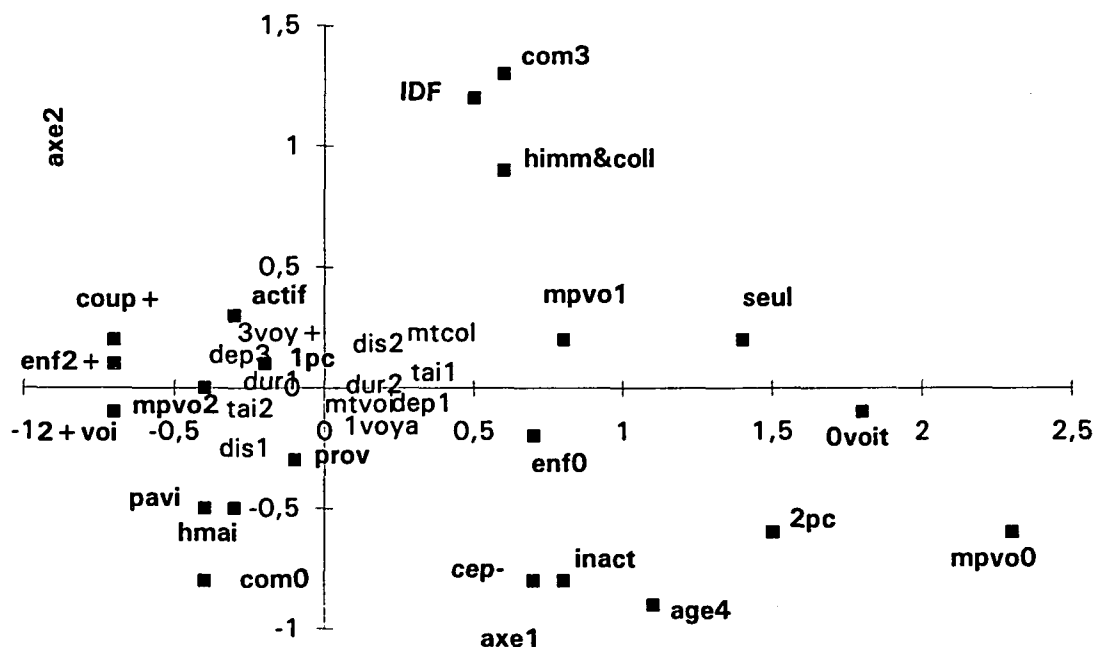
L'axe 1 oppose les personnes appartenant à un ménage ayant relativement plus de voitures, plus de permis de conduire et composé d'un nombre relativement plus grand de personnes (coordonnées négatives) aux personnes seules n'ayant pas de voiture ni de permis de conduire. La première catégorie se déplace bien plus et surtout en voiture, la taille du groupe qui accompagne est plus importante.

AXE 2 : domicile de la personne interrogée

Les variables actives qui contribuent le plus à la constitution de cet axe sont le département du domicile DEPDOM (CTR=13,3), le type d'habitat Q1 (CTR=15), l'environnement du domicile Q2 (CTR=14,5), la taille de la STRATE (CTR=16,8).

Les variables supplémentaires dont les modalités sont les plus significatives sont la durée (DUREM et DURTOT), la distance (DCTOT, DISTANM), le nombre de voyages L02, le mode de transport principal MTPM et le nombre de déplacements NBDEP.

L'axe 2 oppose les personnes résidant en province, en milieu rural, dans des quartiers pavillonnaires et ayant elles-mêmes une maison individuelle (coordonnées négatives), aux personnes résidant en Ile de France, dans des quartiers à habitations collectives et logeant elles-mêmes dans un immeuble (coordonnées positives). Ces dernières se déplacent relativement plus longtemps et sur une plus longue distance, font plus de voyages, utilisent davantage les transports en commun, et se déplacent plutôt seules.



1.2. Les déplacements en variables actives

1.2.1. Sur les individus non mobiles et mobiles se déplaçant pour des motifs privés ou des retour au domicile.

AXE 1 : axe de mobilité croissante

Les variables actives (MOTM, MTPM, DUREM, DISTANM, TAILLEM, NBDEP) contribuent toutes de façon égale à la constitution de l'axe. Les modalités de ces variables sont classées par ordre de mobilité croissante (en durée, distance, nombre de déplacements, taille du groupe accompagnant la personne interrogée). Les modalités mot0 de la variable MOTM (motif de déplacement) et mt0 de la variable MTPM (motif de transport), correspondant aux personnes non mobiles ont des coordonnées négatives contrairement aux autres modalités de ces mêmes variables.

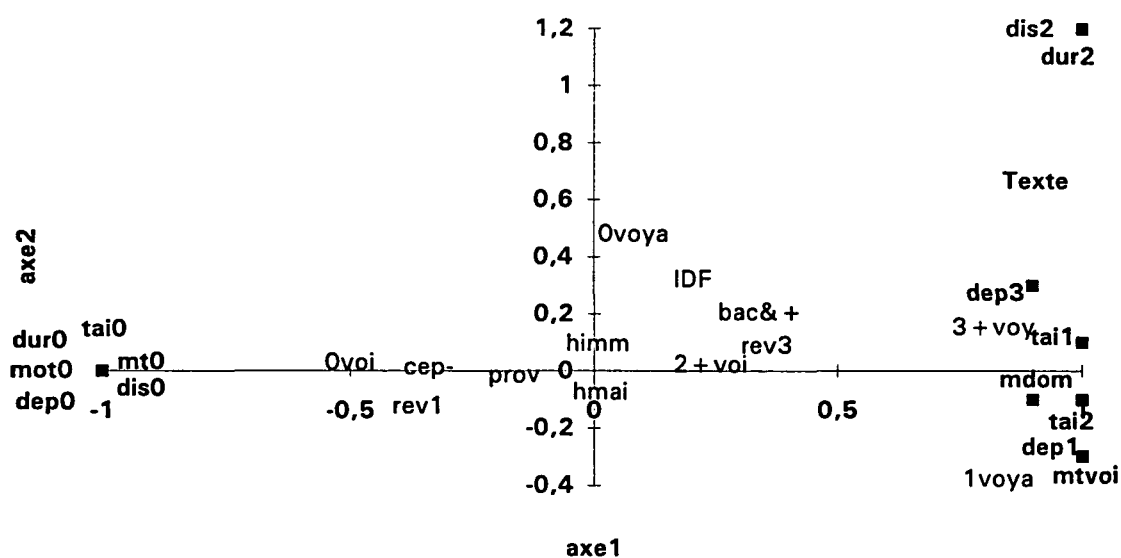
L'examen des variables supplémentaires montre que la mobilité réduite voire nulle est surtout le fait des inactifs, des employés, ouvriers ou des indépendants, n'ayant pas de permis de conduire, de niveau CEP ou inférieur, sans voiture, ni permis de conduire dans leur ménage et ayant un revenu faible (rev1) ; par opposition aux personnes plus favorisées socialement (actifs, cadres, possession du permis de conduire, niveau bac et supérieur, 2 voitures ou plus dans le ménage, revenu dans la tranche rev3).

Cet axe explique déjà 42% de l'inertie totale contre 13,31% pour l'axe 2 dont l'intérêt se situe surtout du point de vue des modalités supplémentaires.

AXE 2 : longueur des déplacements en temps et durée

Les variables actives ayant le plus contribué à la constitution de l'axe sont la distance DISTANM et la durée DUREM. La plupart des autres modalités actives illustrent bien l'axe. Les modalités sont classées sur cet axe, de nouveau dans le sens d'une mobilité croissante (pour les variables DISTANM, DUREM, NBDEP).

Le comportement de plus grande mobilité en termes de distance et de durée, est celui des personnes résidant en Ile de France, vivant en immeubles, bénéficiant de réductions tarifaires, ayant un niveau bac ou supérieur.



1.2.2. Sur les individus mobiles se déplaçant pour des motifs privés ou des retour au domicile.

AXE 1 : distance et durée

Les variables actives ayant le plus contribué à l'axe 1 sont la distance DISTANM (CTR=43,7) et la durée DUREM (CTR=40).

Les variables supplémentaires ayant les modalités avec les valeurs les plus fortes valeurs Vtest en valeur absolue sont ETUD, L02 (nombre de voyages), Q1 (type d'habitat), STRATE (strate de la commune).

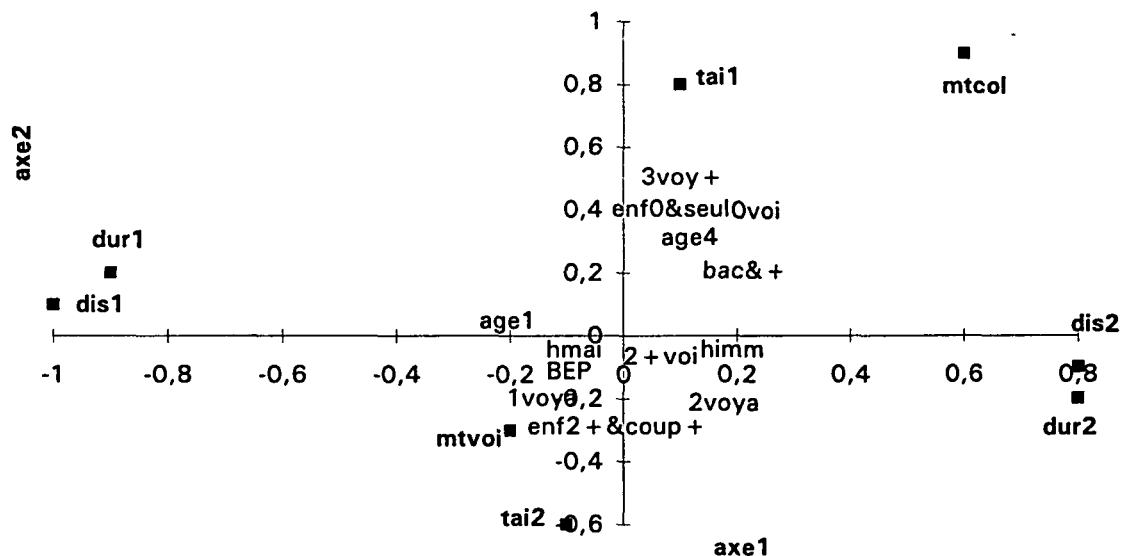
Cet axe oppose les déplacements de courte durée et de courte distance aux déplacements de longue durée et longue distance. Les premiers sont effectués le plus souvent par des personnes ayant le niveau BEP ou CEP, habitant une maison plutôt en zone rurale et ne faisant pas beaucoup de voyages (modalité 1voya). Le deuxième type de déplacements est plutôt réalisé par des personnes de niveau bac ou supérieur, résidant en immeuble à Paris et voyageant beaucoup.

AXE 2 : taille du groupe et moyen de transport principal

Les variables actives ayant le plus contribué à la constitution des axes sont le mode de transport principal MTPM (CTR=26,8) et la taille du groupe (CTR=47,9).

Les variables supplémentaires dont les modalités sont les plus significatives sont L02 (nombre de voyages), NBENF (nombre d'enfants) et TYPMEN (type de ménage).

Cet axe oppose les déplacements effectués en voiture et une taille du groupe élevée (tai2), aux déplacements effectués en transports collectifs par une personne seule (tai0) ou accompagnée d'une seule autre personne (tai1). D'après les modalités supplémentaires, le premier type de déplacements est relativement plus réalisé par des ménages ayant 2 enfants ou plus, où il y a un couple et résidant dans une maison. Les modalités actives dis2 et dur1 illustrent bien ce type de déplacements. Le deuxième type de déplacements est davantage le fait de personnes vivant seules et habitant en immeuble, ils sont de durée et de distance plus courte (dur1 et dis1).



AXE 3 : motif et nombre de déplacements

Les variables actives ayant le plus contribué à la constitution de l'axe sont le motif du déplacement MOTM (CTR=34) et le nombre de déplacements NBDEP (CTR=39,1).

Les variables supplémentaires ayant les modalités les plus significatives sont l'âge AGE, l'occupation B1BIS, le niveau d'études ETUD, le nombre de voyages L02, le nombre d'enfants NBENF.

Cet axe oppose les déplacements effectués dans la cadre d'un retour au domicile, relativement moins nombreux (dep1), aux déplacements effectués dans le cadre loisir-social (mlois-msoc) qui sont plus nombreux (dep3). Les variables qui illustrent bien l'axe et les variables supplémentaires montrent que le premier type de déplacements se fait généralement seul, en transport collectif, sur une durée (dur2) et une distance (dis2) longues et qu'il est surtout réalisé par des jeunes (jeune, age1), de niveau d'études CEP ou moins, ayant 2 enfants ou plus et faisant peu de voyages. Le deuxième type de déplacements se fait surtout accompagné (tail et tai2), en voiture (mtvoit), sur une distance (dis1) et une durée (dur1) plus courte (ces modalités actives illustrent bien l'axe). Les modalités supplémentaires montrent qu'ils sont effectués par des personnes qui ont entre 40 et 60 ans (age3), que ce sont des actifs ayant le niveau bac ou plus (bac&+), sans enfant (enf0) et effectuant de nombreux voyages (3voy+).

II. UNE ANALYSE PAR ARBRE DE DISCRIMINATION : LA SEGMENTATION

Afin de mieux connaître le comportement des personnes dans le domaine des transports, plusieurs segmentations ont été effectuées sur l'échantillon pour déterminer des segments de population se caractérisant par une mobilité particulière. On espère construire une typologie de la population selon son appartenance à différents niveaux d'intensité de mobilité mesurée par le nombre de déplacements (variable expliquée). Les variables explicatives retenues ont été choisies en fonction du lien que l'on pouvait supposer entre elles et la variable expliquée, à partir des analyses des correspondances multiples et de tableaux de contingence. Ces variables sont le statut du ménage et sa taille, la taille de sa commune de résidence, le nombre de voitures possédées, l'âge de la personne interrogée, son niveau d'études, sa PCS, le nombre de personnes ayant voyagé en sa compagnie, le motif du déplacement et le mode de transport utilisé en général pour se déplacer.

La segmentation a été effectuée à l'aide de la procédure TREEDISC du logiciel SAS. Cette procédure utilise comme critère de discrimination le chi-deux. Cela correspond à la recherche du χ^2 maximum sur le tableau de contingence croisant la variable à expliquer et chacune des variables explicatives candidates pour la coupure.

2.1. Segmentation sur la population totale

La première segmentation concerne l'ensemble de l'échantillon. C'est une discrimination entre les mobiles et les non mobiles. L'échantillon est assez bien équilibré puisqu'il comporte 46% de non mobiles et 54% de mobiles. Le critère d'arrêt choisi, outre la significativité du test du chi-deux, a été une taille du noeud supérieure ou égale à 10% de la taille de la population totale. La qualité de la discrimination obtenue ne pourra donc pas être très grande, mais cela permet d'avoir un nombre de noeuds terminaux pas trop élevé, ce qui facilite l'interprétation des résultats.

Lors de la première étape, la variable décrivant le type de ménage a été retenue comme permettant de séparer l'échantillon en quatre noeuds intermédiaires : couples, couples avec enfants, familles monoparentales et ménages divers, et enfin personnes seules. Une deuxième variable explicative a été retenue ensuite : la taille de la commune de résidence. Enfin, une dernière discrimination a été possible sur la variable occupation de l'individu, c'est-à-dire sa situation par rapport à une activité professionnelle.

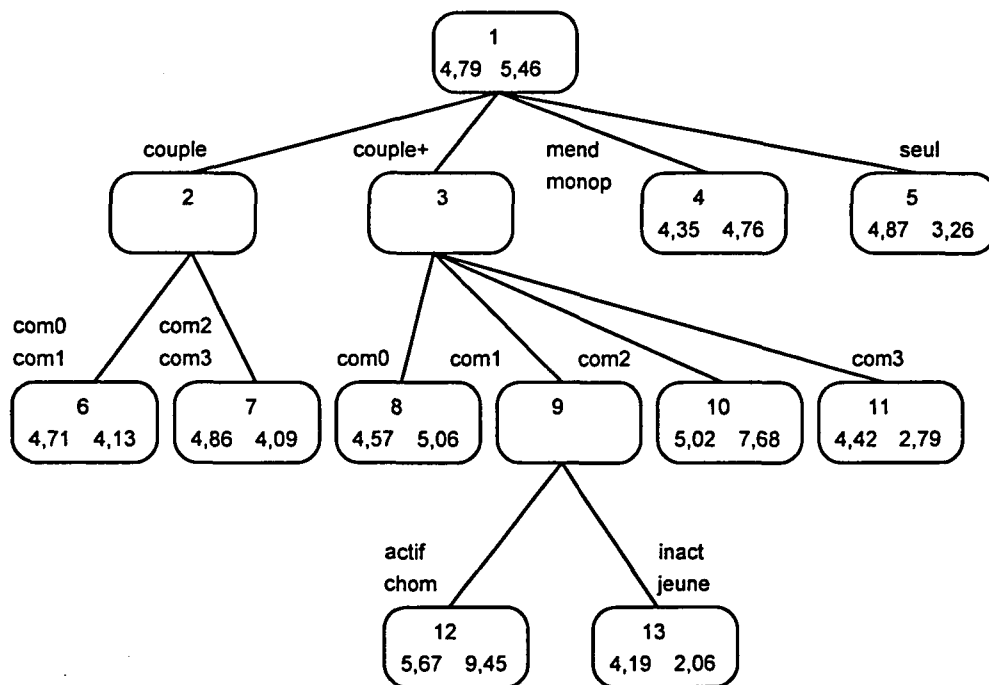
Tout d'abord, il faut constater que les résultats obtenus dans cette segmentation ne sont pas très bons. On n'obtient pas de noeuds terminaux très purs avec un nombre aussi réduit de divisions (on n'a pas de noeuds comprenant au moins 70% d'individus ayant la même modalité de déplacement. En réalité, on a rarement plus de 50% d'individus avec le même comportement de mobilité, cf annexe 1.3). Cependant, on peut tout de même remarquer que les personnes seules et les familles monoparentales

sont moins mobiles que la moyenne (les familles monoparentales sont 56% à ne pas s'être déplacées pendant l'enquête, alors qu'il n'y a que 46% de non mobiles dans la population totale). De même, les couples avec enfants, jeunes ou inactifs et vivant dans des communes de 2000 à 100000 habitants et les couples sans enfant résidant dans des communes de moins de 100000 habitants constituent des segments de population peu mobiles. On note que ce sont les ménages de jeunes ou d'inactifs avec enfants qui sont les moins mobiles (4,19 voyages en moyenne).

Par contre, les couples vivant dans les communes de plus de 100000 habitants, les couples de ruraux avec enfants et les couples d'actifs avec enfants vivant dans des communes de 2000 à 100000 habitants sont plutôt mobiles. Pour cette dernière catégorie, le fait d'être chômeur ne semble pas diminuer le comportement de mobilité. Ce sont les actifs urbains avec enfants qui sont les plus mobiles, avec 5,67 déplacements sur la période d'enquête.

Enfin le comportement de mobilité des personnes seules présente des particularités. En effet, ce segment a été classé comme non mobile, car il comprend 53% de non mobiles, c'est-à-dire plus que sur la population entière. Cependant, la moyenne du nombre de déplacements effectués par les ménages appartenant à cette catégorie est de 4,87 ce qui est élevé et supérieur à celle de plusieurs segments classés en mobiles. Cela peut peut-être s'expliquer par la présence dans ce segment de personnes âgées ne pouvant plus se déplacer, et de jeunes célibataires très actifs et très mobiles.

2.1 segmentation tous individus mobiles-non mobiles



Note: Dans chaque noeud terminal, correspondant à un segment de population, on donne à gauche la moyenne du nombre de déplacements pour le segment et à droite l'écart-type de cette moyenne.

Légende:

couple: couple sans enfant
 couple+: couple avec enfants
 monop: famille monoparentale
 mend: ménages divers
 seul: personne isolée

com0: commune de moins de 2000 habitants
 com1: commune de 2000 à 100000 habitants
 com2: commune de plus de 100000 habitants
 com3: Paris

actif: personne interrogée active
 chom: personne interrogée au chômage
 inact: personne interrogée inactive
 jeune: personne interrogée étudiant,
 militaire

2.2. Segmentation sur l'ensemble des personnes mobiles

Dans une deuxième segmentation, on introduit la notion d'intensité de mobilité. En effet, si on conserve toujours le groupe des non mobiles, on divise celui des mobiles en un groupe de peu mobiles (de 1 à 3 déplacements durant l'enquête) et un groupe de très mobiles (plus de 3 déplacements). Ainsi, si la population étudiée comprend toujours 46% de non mobiles, 23% est classée parmi les peu mobiles et 30% parmi les très mobiles.

L'arbre de discrimination obtenu est le même que celui donné par la segmentation entre les mobiles et les non mobiles. Mais on dispose d'informations supplémentaires sur les personnes mobiles concernant l'intensité de leur mobilité. Ainsi, on constate que pour le segment couples d'actifs ou de chômeurs avec enfants résidant dans des communes de 2000 à 100000 habitants, seule la classe des très mobiles augmente significativement son poids (34% contre 30% dans la population totale), les deux autres classes étant moins représentées. En outre, on remarque que le segment couples d'inactifs ou jeunes avec enfants résidant dans des communes de 2000 à 100000 habitants voit la catégorie des peu mobiles se renforcer. On a donc un segment de personnes peu ou pas mobiles.

Par la suite, les segmentations ont porté sur la population des mobiles, et selon le motif du déplacement.

2.3. Segmentations sur la population des mobiles avec distinction des motifs de déplacement (Graphiques page 16).

Trois segmentations ont été effectuées respectivement sur les personnes s'étant déplacées pour motif professionnel, pour motif privé ou pour un retour à domicile. Les retours à domicile représentent 86% des motifs déclarés², les motifs privés 12,8%. La segmentation sur les ménages s'étant déplacés pour motif professionnel a été abandonnée car elle ne concerne qu'un trop petit nombre de personnes (1,4%) ce qui ne permet pas de mener la discrimination de manière satisfaisante.

On étudie la fréquence des déplacements à longue distance selon les deux motifs, privé ou retour à domicile. Les deux modalités de la variable nombre de déplacements sont : peu mobiles, soit de 1 à 3 déplacements sur la période considérée, et très mobiles, soit 4 déplacements et plus.

La segmentation sur les ménages ayant effectué des déplacements pour motif privé permet de déterminer que les segments de population membres de familles monoparentales couples avec enfants, habitant Paris, ou une ville de 2000 à 100000 habitants ou une ville de plus de 100000 habitants lorsque la personne interrogée est au chômage, inactive ou jeune sont peu mobiles. Par contre, les couples sans enfant sont très mobiles, ainsi que les couples d'actifs avec enfants habitant les grandes villes sont très mobiles. De même, mais à un degré moindre, les personnes seules et les couples de ruraux avec enfants ont une mobilité assez élevée.

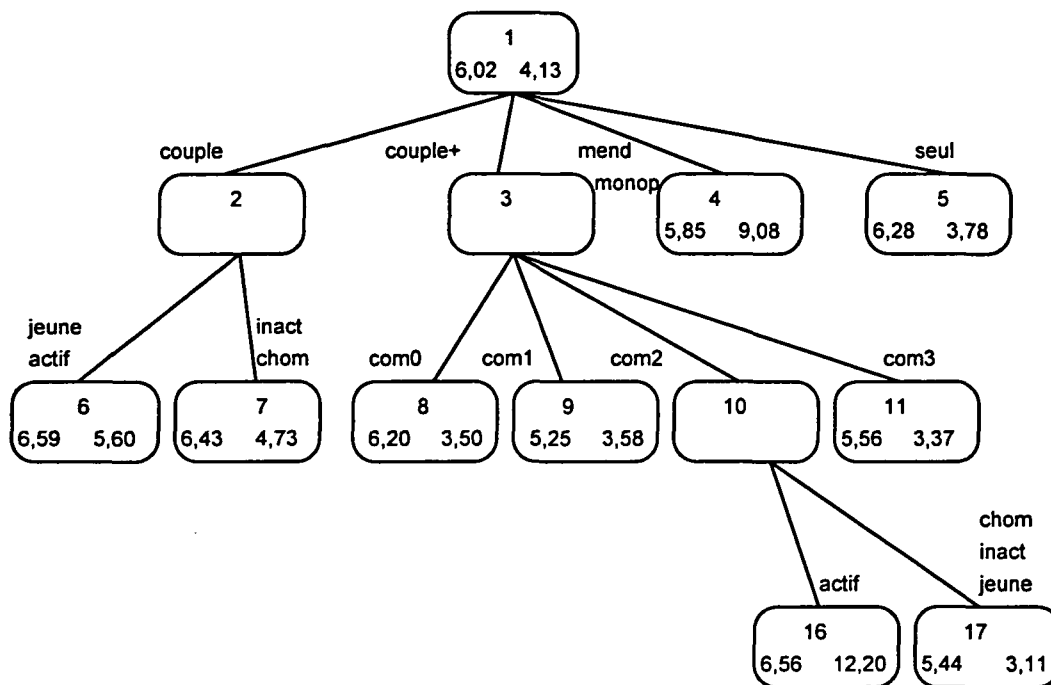
Le segment le plus mobile est celui des couples jeunes ou actifs sans enfant (6,56 déplacements), le moins mobile les couples avec enfants résidant dans des communes de 2000 à 100000 habitants (5,25).

Enfin, la segmentation sur la population des gens s'étant déplacés au titre du motif retour à domicile s'est effectuée à l'aide des variables explicatives type de ménage, taille de la commune de résidence et revenu total du ménage. On constate que les personnes seules et les familles monoparentales, les couples avec enfants résidant dans des communes de 2000 à 100000 habitants et les habitants de plus grandes villes avec un revenu faible sont peu mobiles. Les couples sans enfant, ceux avec enfants, disposant de revenus élevés et habitant dans les grandes villes sont eux très mobiles.

²Ce très fort pourcentage de motifs "retour à domicile" peut sembler surprenant. En fait, le traitement effectué pour déterminer un déplacement type, décrit en première partie du rapport, a augmenté de façon artificielle la part de ce motif.

La segmentation a donc permis d'avoir une première approche du comportement de mobilité des personnes. Cette étude n'apporte cependant que des informations très partielles sur ce comportement. Il serait nécessaire d'effectuer une analyse approfondie de chacun des segments déterminés. Cela n'a pas pu être réalisé dans le cadre de ce groupe de statistique appliquée, faute de temps.

2.2 segmentation sur les personnes mobiles motif privé



Note: Dans chaque nœud terminal, correspondant à un segment de population, on donne à gauche la moyenne du nombre de déplacements pour le segment et à droite l'écart-type de cette moyenne.

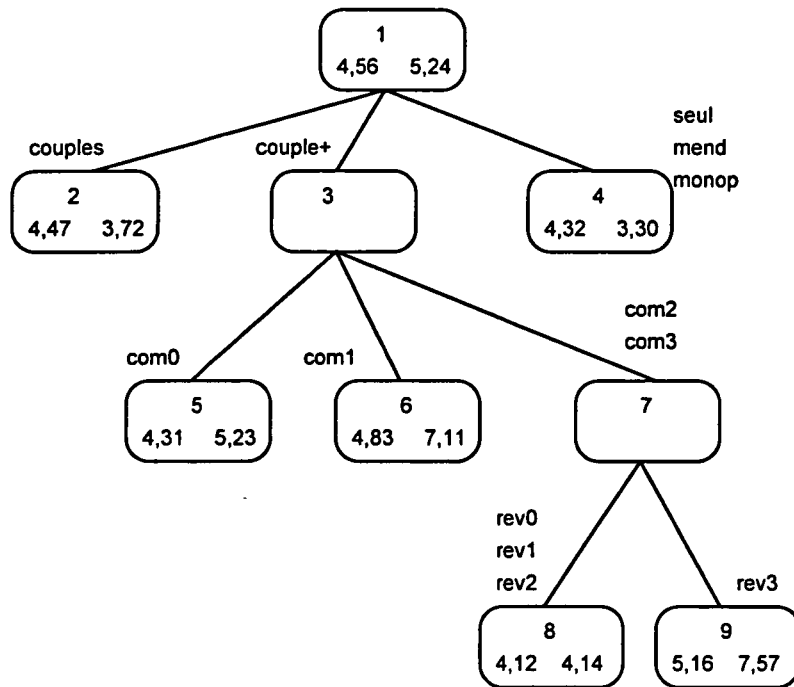
Légende:

couple: couple sans enfant
 couple+: couple avec enfants
 monop: famille monoparentale
 mend: ménages divers
 seul: personne isolée

com0: commune de moins de 2000 habitants
 com1: commune de 2000 à 100000 habitants
 com2: commune de plus de 100000 habitants
 com3: Paris

actif: personne interrogée active
 chom: personne interrogée au chômage
 inact: personne interrogée inactive
 jeune: personne interrogée étudiant, militaire

2.3 segmentation sur les mobiles retour à domicile



Note: Dans chaque noeud terminal, correspondant à un segment de population, on donne à gauche la moyenne du nombre de déplacements pour le segment et à droite l'écart-type de cette moyenne.

Légende:

couple: couple sans enfant
 couple+: couple avec enfants
 monop: famille monoparentale
 mend: ménages divers
 seul: personne isolée

com0: commune de moins de 2000 habitants
 com1: commune de 2000 à 100000 habitants
 com2: commune de plus de 100000 habitants
 com3: Paris

actif: personne interrogée active
 chom: personne interrogée au chômage
 inact: personne interrogée inactive
 jeune: personne interrogée étudiant,
 militaire

revenu annuel total du ménage:
 rev0: moins de 20000F
 rev1: de 20000F à 85000F
 rev2: de 85000F à 165000F
 rev3: plus de 165000F

III. CLASSIFICATION ASCENDANTE HIERARCHIQUE

Nous avons réalisé une classification ascendante hiérarchique en retenant les mêmes variables que celles choisies pour la segmentation afin de pouvoir comparer ces deux méthodes qui - bien que différentes - ont un même objectif : regrouper les individus en un nombre restreint de classes homogènes.

Tableau des 7 derniers noeuds de la hiérarchie
Inertie totale : 3.351843

Noeud	Effectif pondéré	Classes jointes	Perte d'inertie inter			Histogramme
			0/00	cum.	dif	
CL1	89683292.0	CL2-CL6	57	57	.	*****
CL2	59618841.0	CL3-CL5	36	94	21	*****
CL3	33593456.0	CL4-CL20	33	127	3	*****
CL4	28055991.0	CL9-CL7	28	155	5	*****
CL5	26025385.0	CL15-CL27	28	182	0	*****
CL6	30064451.0	CL8-CL62	26	208	1	*****
CL7	2359659.0	CL74-CL44	24	233	2	*****

L'examen de cet histogramme, en particulier de la perte d'inertie interclasse, nous suggérerait plutôt de ne retenir que 3 classes. La segmentation réalisée sur l'ensemble des individus aboutit à 9 noeuds terminaux donc, par souci d'homogénéiser un peu les deux méthodes, nous avons retenu 5 classes. L'analyse de la décomposition de l'inertie totale (voir tableau 1 en annexe) montre que la **classe la plus dispersée est la classe n°3**. La décomposition signée du Rho2 - distance entre le centre des classes et le centre de gravité du nuage - montre que la **classe n°5 est la plus atypique** (voir tableau 2 en annexe). On constate de plus que les variables permettant de discriminer au mieux les classes sont (par ordre de significativité décroissante): le mode de transport principal, le nombre de voitures dans le ménage, l'activité, la taille et le nombre de permis de conduire du ménage, le nombre de déplacements, le revenu, le motif du déplacement, le type de ménage, le niveau d'étude, la strate de la commune et la pcs (voir tableau 3 en annexe). La segmentation, par contre, avait retenu successivement les variables concernant le type de ménage, la strate de la commune et l'activité. Par ailleurs, chaque classe est significativement homogène par rapport à chaque variable (utilisation du test du χ_2).

Caractérisation de chaque classe (voir tableau 4 en annexe):

- **Classe 1**

Elle regroupe des individus qui, *plus que les autres*, présentent les caractéristiques suivantes: un permis de conduire dans le ménage, une taille du groupe accompagnateur égale à 1, des retours à domicile, une voiture dans le ménage, 2 ou 3 déplacements, surtout en voiture mais aussi en transports collectifs, revenu assez élevé (85000-165000), personnes seules ou famille monoparentale, bac ou études supérieures.

Elle regroupe des individus qui, *moins que les autres*, présentent la caractéristique suivante: l'immobilité (modalités motifm0, mtp0, ndep0, tai0), 2 voitures ou plus et 2 permis de conduire dans le ménage.

- **Classe 2**

Elle regroupe des individus qui, *plus que les autres*, présentent la caractéristique suivante: l'immobilité.

Elle regroupe des individus qui, *moins que les autres*, présentent les caractéristiques suivantes: les retours à domicile, en voiture et une taille du groupe accompagnateur supérieure à 2.

- **Classe 3**

Elle regroupe des individus qui, *plus que les autres*, présentent les caractéristiques suivantes: les retours à domicile, en voiture, la taille du groupe accompagnateur supérieure à 2, 2 voitures ou plus et 2 permis de conduire dans le ménage, plus de 4 déplacements, revenus supérieurs à 165000.

Elle regroupe des individus qui, *moins que les autres*, présentent les caractéristiques suivantes: l'immobilité, un seul permis de conduire dans le ménage, personnes seules.

- **Classe 4**

Elle regroupe des individus qui, *plus que les autres*, présentent les caractéristiques suivantes: ménage sans voiture ni permis de conduire, revenus compris entre 20000 et 85000, personne seule, CEP ou niveau d'études inférieur.

Elle regroupe des individus qui, *moins que les autres*, présentent les caractéristiques suivantes : 1 voiture ou plus et 2 permis de conduire dans le ménage, revenus supérieurs à 165000.

- **Classe 5**

Elle regroupe des individus qui, *plus que les autres*, présentent les caractéristiques suivantes : personnes seules, déplacements pour motifs professionnels, plus de 6 déplacements, transports collectifs, niveau bac, actif.

Elle regroupe des individus qui, *moins que les autres*, présentent les caractéristiques suivantes : l'immobilité, niveau CAP ou BEP, ouvrier.

Ces classes présentent toutes des profils bien déterminés et relativement complets, tant sur le plan du signalétique que de la mobilité. Les résultats bruts de la classification nous paraissent plus riches en enseignement que ceux de la segmentation. En effet, ils font intervenir de façon assez équilibrée les variables de signalétique et de mobilité alors que la segmentation ne retient que le type de ménage, la strate de la commune et l'activité pour la constitution des 9 noeuds terminaux. Les trois variables les plus déterminantes dans la construction des classes sont (dans l'ordre de significativité décroissante) : le mode de transport, le nombre de voitures et la situation eu égard à l'activité (variable occup). Or il est particulièrement intéressant, dans le cadre d'une analyse des choix modaux, que la variable mode de transports soit autant valorisée.

2ème partie : ANALYSE PARAMETRIQUE

Estimation micro-économétrique de la mobilité sur longue distance et des choix modaux de transport

L'analyse des comportements de mobilité sur longue distance et de ses modalités sous-tend l'ensemble de notre travail. Une telle démarche court le risque de demeurer incomplète si elle se limite à l'observation de tableaux croisés ventilant les pratiques selon différents critères. Certes, ces traitements sont indispensables dans une première étape et il semble s'en dégager des relations stables entre les caractéristiques individuelles et de déplacements. Ils seront présentés en introduction aux différents modèles économétriques.

Les méthodes non paramétriques appliquées - ACM, segmentation et classification - présentent l'avantage de détecter des relations sans les fixer dans le cadre rigide d'une spécification particulière, à l'inverse des méthodes économétriques. Mais ce que ces dernières perdent ainsi en généralité, (elles nécessitent des hypothèses souvent contraignantes), elles le gagnent en possibilité d'isoler les effets propres des facteurs intervenant dans l'explication du phénomène.

En matière d'étude de comportements individuels, les pratiques étudiées sont le plus souvent de nature discrète, qualitative. L'étude de la mobilité - pour motif privé - et de son intensité s'inscrit dans ce cadre. La restriction au motif privé a semblé souhaitable, car comme déjà mentionné, il est probable que la mobilité pour raisons professionnelles obéisse à des processus décisionnels d'une autre nature : les contraintes et les objectifs y sont spécifiques, comparés à la mobilité pour raisons privées. Les profils d'individus concernés par la mobilité professionnelle se différencient sensiblement du reste de la population de l'échantillon et inclinent donc à privilégier cette distinction. L'échantillon comporte un certain nombre d'indications relatives aux conditions de cette mobilité. Ainsi, après en avoir estimé les déterminants seront recherchés ceux des choix modaux de transport. Cette problématique est en effet d'un intérêt évident pour tous les opérateurs intervenant dans le champ des transports.

Un renvoi en annexe est proposé pour chacun des points méthodologiques abordés au fur et à mesure du développement de cette partie. Avant de présenter plus avant les deux axes d'étude, il est par exemple nécessaire de ce point de vue de préciser l'utilisation commune qu'il a été faite de la pondération personnelle initiale : pour chaque observation, elle a été divisée par sa moyenne (cf. annexe 2.1). De même, par souci de simplification et en l'absence d'un modèle théorique précis, les spécifications testées postulent des liaisons additives entre variables, alors que des phénomènes d'interaction peuvent se produire (cf. annexe 2.1).

I. LES DETERMINANTS DE LA MOBILITE PRIVEE ET DE SON INTENSITE

1.1. Définition de la mobilité et de son intensité

Comme pour les exploitations non paramétriques, le nombre de déplacements décomptés par l'enquête pendant les trois mois d'observation sera le critère privilégié de la définition de la mobilité. Est ainsi considérée comme "mobile" toute personne qui a effectué au moins deux déplacements pendant cette période (un déplacement hors du domicile est toujours associé à un retour au domicile, le nombre minimal de déplacements est donc de deux). Les "non mobiles" n'ont effectué aucun déplacement.

On pourrait, dans une première approche, se limiter à l'analyse des déterminants individuels selon cette distinction dichotomique. Néanmoins, il est possible de l'inclure dans une analyse à la fois plus générale et plus riche en terme d'information. Il faut alors faire appel au concept d'intensité de mobilité qui englobe tous les niveaux de mobilité, nul ou positifs, dont le nombre de déplacements fournit une mesure directe. Cette intensité peut être évaluée à l'aide de critères alternatifs, tels que le nombre de voyages effectués (inférieur ou égal au nombre de déplacements, puisque un voyage est un enchaînement de déplacements) ou la distance totale parcourue sur l'ensemble de la période. Bien entendu, un individu considéré non mobile du point de vue du nombre de déplacements le sera aussi du point de vue du nombre de voyages et de la distance totale parcourue.

1.2. Les modélisations retenues

1.2.1. le modèle TOBIT

Puisqu'il s'agit d'analyser les déterminants du nombre de déplacements à titre privé (ou de l'un des deux autres critères évoqués), l'on se trouve en présence d'un problème statistique de distribution tronquée.

Le nombre de déplacements a en effet pour borne inférieure la valeur 0 où est concentrée une forte minorité de l'échantillon (45% sur données non pondérées). Une estimation par les moindres carrés ordinaires (MCO) est a priori inadaptée dans la mesure où l'hypothèse habituelle de normalité des résidus n'est plus respectée. L'estimateur obtenu par les MCO sur l'ensemble des données est donc non convergent. Formellement, le problème peut s'exposer comme suit :

soit le modèle :

$$y_i^* = x_i' b + u_i \text{ pour l'individu } i$$

où y_i^* = niveau d'utilité retirée du nombre de déplacements, inobservable : y_i ,
le nombre de déplacements, est donc la manifestation visible d'une variable latente y_i^* non observable, qui est continue et non tronquée. L'événement "mobilité" se réalise, lorsque la variable latente dépasse le seuil d'utilité nulle.
L'on observe donc :

$$y_i = y_i^* \text{ si } y_i^* \geq 0 \\ = 0 \text{ si } y_i^* < 0,$$

avec x_i' = vecteur du signalétique de l'individu i ,
 b = vecteur des coefficients à estimer,
 u_i = résidu ; les u_i sont iid (indépendants et identiquement distribués) et suivent une loi $N(0, \sigma^2)$

On n'a plus $E(y_i/x_i) = x_i' b$ (espérance de y_i sachant x_i)

mais $E(y_i/x_i) = x_i' b \Phi(x_i' b / \sigma) + \sigma \phi(x_i' b / \sigma)$ (équation 1)

où Φ = fonction de répartition d'une loi normale centrée réduite,
 ϕ = fonction de densité d'une loi normale centrée réduite,
 σ = écart-type des résidus.

Si l'on recourt à l'estimateur des MCO sur les données complètes, c'est-à-dire telles que $y_i > 0$, il est toujours non convergent : $E(y_i/x_i) = x_i' b + \sigma [\phi(x_i' b / \sigma)] / [\Phi(x_i' b / \sigma)]$ (équation 2). Le biais est certes moindre que dans l'équation 1, mais il demeure.

La méthode d'Heckman articulant un modèle PROBIT à un modèle de régression linéaire par les MCO permet de résoudre ce problème (cf. annexe 2.2 pour l'exposé de son principe).

1.2.2. Le modèle LOGIT ordonné

Etant donné la forte concentration d'individus autour d'un nombre limité de déplacements (61,2% des "mobiles" ont effectué entre deux et quatre déplacements), une méthode alternative au modèle TOBIT consiste à catégoriser le nombre de déplacements en quelques tranches qui refléteraient des seuils d'intensité de mobilité. Compte tenu du caractère arbitraire de la méthode, plusieurs catégorisations peuvent être testées. La nouvelle variable ainsi créée prend le caractère d'une variable polytomique ordonnée, que la procédure LOGISTIC du logiciel SAS permet de traiter aisément. La fonction de répartition utilisée est celle de la loi logistique (assez proche dans ses résultats de la loi normale), définie par $F(w)=e^w/(1+e^w)=1/(1+e^{-w})$, soit $P(Y=1)=1/(1+e^{-xb})$ dans le cas d'un modèle dichotomique.

Dans le cas où Y prend un petit nombre de valeurs ordonnées de 1 à k+1, la forme du modèle devient: $P(Y \leq i) = F(a_i + X_i b)$ avec $1 < i < k$. Les a_i sont les k constantes et b est le vecteur des coefficients des variables explicatives, qu'on peut encore appeler "coefficients de pente". La constante a_i seule changeant avec i, b restant le même, ce modèle pose l'égalité des pentes quand i varie, ce qui paraît être une hypothèse assez forte mais testable.

1.3. Profil de la population selon son degré de mobilité.

Celui-ci est constitué à partir des indicateurs suivants :

Tableau 2.1 : Profil de l'échantillon des individus (déplacements professionnels exclus)			
Indicateurs de profil	total	mobiles	non mobiles
% femmes	51,6	49,8	53,6
% <26 ans / % >60 ans	30,0 / 20,3	32,1 / 14,8	27,5 / 26,6
% actifs / % chômeurs	40,4 / 6,5	45,7 / 5,1	34,3 / 8,0
% cadres / % employés+ouvriers	28,4 / 58,3	38,7 / 51,1	17,0 / 66,3
% niveau d'études ≥ baccalauréat	28,4	37,7	17,7
% personnes seules / % couples avec enfants	12,0 / 52,6	10,3 / 55,4	13,9 / 49,3
% com. <2000 hab / % com. >100000 hab	25,0 / 44,9	22,8 / 48,5	27,7 / 40,7
% résidant en Ile de France	18,9	22,1	15,2
% résidant dans une maison	64,6	64,0	65,1
% ayant une résidence secondaire	12,3	17,3	6,5
% ayant le permis de conduire(PC)	63,0	69,4	55,6
% n'ayant aucun PC dans son ménage	10,0	4,8	15,9
% n'ayant pas de voiture dans son ménage	16,5	9,7	24,3
% ayant une réduction transports	26,2	29,1	22,8
âge : moyenne (coefficient de variation en %)	39,7 (53, 4)	37,0 (52, 5)	42,8 (53)
revenu annuel du ménage : moyenne (CV%)	110025 (103)	125978 (102)	91694 (98)
nombre d'enfants : moyenne (CV%)	1,33 (103)	1,31 (97)	1,34 (109)
nbre de voitures du ménage : moyenne (CV%)	1,26 (63, 4)	1,4 (53, 6)	1,09 (75)
Effectif non pondéré	14069	7743	6318
Effectif pondéré	52,7 millions	28,2 millions	24,5 millions

Comparativement aux non mobiles, les mobiles forment une population :

- * légèrement moins féminisée et sensiblement plus jeune,
- * plus diplômée, plus engagée dans la vie active et à des niveaux hiérarchiques plus élevés (sur-représentation des cadres), ce qui leur confère en moyenne des revenus substantiellement plus élevés (près de 40% plus élevés),
- * plus urbanisée et en particulier plus concentrée en région parisienne,

- * disposant plus souvent d'une résidence secondaire, même si cela ne concerne qu'une fraction minoritaire,
- * vivant un peu plus souvent dans le cadre d'une famille constituée,
- * plus autonome du point de vue de ses possibilités de déplacement : seulement un ménage sur 10 ne possède pas de voiture contre près d'un sur 4 parmi les non mobiles. L'écart est également très net pour la proportion de ménages dans lesquels personne n'a son permis de conduire. Même sur le plan de l'utilisation de moyens de transports collectifs, la proportion de bénéficiaires de réductions est supérieure chez les mobiles, malgré leur relatif bon taux d'équipement automobile.

L'ensemble de ces caractéristiques opposant ces deux catégories renvoient une image assez contrastée : l'image d'une population socialement privilégiée parmi les mobiles, qui dispose de ce fait de possibilités de transports variés et flexibles - voiture et transports collectifs propres à un environnement urbanisé - leur permettant de se déplacer sans problème.

Cette différenciation des profils s'accroît avec l'intensité de la mobilité pour la plupart des critères (cf. annexe 2.3). Parmi les mobiles, l'accroissement du nombre de déplacements va en effet de paire avec :

- * une réduction de la part des femmes,
- * une progression de la situation sociale et de ses attributs habituels : revenu, part des cadres, part des études supérieures,
- * une urbanisation accrue sans que la part des Franciliens n'augmente,
- * une réduction du nombre d'enfants mais la part des couples avec enfants reste stable,
- * une progression de l'équipement automobile du ménage et une stabilité de la part des bénéficiaires de réductions de transport.

L'âge ne semble pas relié linéairement à l'intensité, ce que l'examen des ACM avait déjà révélé : nettement plus élevé parmi les non mobiles, il diminue pour des mobilités de faible intensité puis remonte ensuite, sans atteindre le niveau constaté pour les non mobiles.

L'examen des conditions de la mobilité en fonction de l'intensité de la mobilité (cf. annexe 2.4) laisse également présager quelques régularités statistiques : lorsque celle-ci s'accroît,

- * la distance moyenne de déplacement diminue ainsi que la durée moyenne,
- * la taille du groupe de personnes diminue, ce qui est cohérent avec le recours accru à un mode de transport collectif (mais celui-ci demeure très minoritaire) : la voiture est le mode de transport de groupe par excellence, en raison d'une constance des coûts fixes ; il est probable que pour un petit groupe, la voiture représente le mode le moins coûteux.

1.4. Estimation du modèle TOBIT de mobilité

1.4.1. résultats de l'estimation PROBIT (1ère étape de la méthode d'Heckman)

Les résultats sont dans l'ensemble conformes aux prédictions que les distributions statistiques permettaient de supposer :

- * Les actifs sont, toutes choses égales par ailleurs, plus mobiles que les inactifs eux mêmes plus mobiles que les chômeurs. Ainsi, à revenu égal à celui d'un actif, un chômeur, face à l'incertitude de rentrées financières futures, aura une attitude en retrait par rapport à l'action de se déplacer : il s'agit d'un poste budgétaire sur lequel des économies sont toujours possibles (la mobilité, impliquant une dépense, apparaît donc reliée positivement au revenu). Par ailleurs, des considérations plus sociologiques peuvent également jouer (une vie sociale moins active etc.).

Tableau 2.2 : Résultats du modèle PROBIT associé au modèle TOBIT ; variable dépendante : mobilité ; niveau 1 (mobiles)=5263 observations - niveau 2 (non mobiles)=4748 observations (2648 valeurs manquantes) valeur de la log-vraisemblance : -5907				
(1) variables	(2) DF	(3) coefficient	(4) probabilité critique	Guide de lecture
Constante	1	0,2807	**	La plupart des variables explicatives sont qualitatives et donc mises sous forme d'indicatrices.
Taille de la commune	4		ns	En colonne 2 est indiqué le nombre de degrés de liberté (DF).
communes rurales	1	-0,0722	*	L'estimation de l'effet de chacune des modalités des variables figure en colonne 3.
communes < 20000 habitants	1		ns	
comm. de 20 à 100000 hab.	1		ns	
comm > 100000 hab. sf Paris	0			
unité urbaine de Paris	1		ns	
<i>Régions (seules les modalités significatives sont présentées.</i>				La colonne 4 donne le seuil de significativité (ou probabilité critique) du test de $H_0 : b_k = 0$ contre $H_1 : b_k \neq 0$, relativement à la modalité de référence (ligne ombrée).
	21		***	Soit γ la valeur prise par la statistique de test de Wald dans l'échantillon : $\gamma = b_k^2 / \sigma_k^2$. γ suit une loi du Chi2 à 1 degré de liberté.
Alsace	1	-0,2103	*	La probabilité critique = $\Pr[\chi^2 > \gamma] = \alpha$.
Basse Normandie	1	0,2058	*	Plus α est faible, plus la probabilité de réalisation de H_0 est faible.
Bourgogne	1	0,216	*	Pour les variables qualitatives, on dispose également du seuil α du test de Wald de nullité simultanée de l'ensemble des b_k associés à ces variables.
Champagne	1	0,3633	***	$\alpha < 1\%$ représenté par ***.
Limousin	1	0,2415	*	$1\% < \alpha < 5\%$ représenté par **.
Nord- Pas de calais	1	-0,1755	*	$5\% < \alpha < 10\%$ représenté par *.
Midi-Pyrénées	1	0,2327	**	
Île de France	0			
<i>Occupation actuelle</i>				Les comparaisons entre modalités (hors modalité de référence) nécessitent le recours à la méthode de Scheffé. La statistique de Wald (à partir des estimations) associée est : $(b_j - b_k)^2 / (\sigma_j^2 + \sigma_k^2 - 2\sigma_{jk}^2)$.
actifs	1	0,0774	*	Cette statistique suit un Chi2 à (i-1) ddl où i = nombre de modalités de la variable.
chômeurs	1	-0,1023	*	Pour un seuil de significativité donné ($\alpha = 5\%$), la différence entre deux modalités est réputée significative si la valeur de la statistique est supérieure au quantile de cette loi.
inactifs	0			
<i>PCS</i>				Commentaires sur la spécification :
agriculteurs	1	-0,4175	***	Les variables quantitatives continues telles que le revenu, l'âge, le nombre d'enfants ont été introduites telles quelles, ce qui évite la perte d'informations liée à une catégorisation en tranches. En revanche une estimation à partir de tranches donne plus de souplesse au modèle, car elle suppose des élasticités (par rapport à la mobilité) différentes.
autres indépendants	1	-0,1526	***	
cadres supérieurs	1	0,2746	***	
professions intermédiaires	1		ns	
employés	0			
ouvriers	1	-0,2817	***	
<i>Niveau d'études</i>				Le pouvoir explicatif du modèle a été estimé à 19%. SAS ne fournit pas directement cette
CEP et moins	0			
BEPC	1	0,0986	**	
BEP-CAP	1		ns	
Bac	1	0,316	***	
Etudes supérieures	1	0,5149	***	
<i>Type d'habitat</i>				
maison	0			
immeuble	1	-0,0781	**	
<i>Résidence secondaire</i>				
non	0			
oui	1	0,4944	***	
<i>Type de ménage</i>				
seul	1	0,2313	***	
couple sans enfants	1	0,2236	***	
couple avec enfants	0			
familles monoparentales	1		ns	
autre ménage	1	-0,231	***	
<i>Sexe</i>				
homme	1		ns	
femme	0			
<i>Possession permis de cond.</i>				
oui	0			
non	1	-0,0942	*	
<i>Nombre de PC ds le ménage</i>				
aucun	1		ns	
un	1	-0,1242	***	
deux	0			

trois et plus	1	0,0866	*	mesure : celle ci est le résultat d'un algorithme qui repose sur la définition habituelle du R^2 , où la somme des carrés des résidus (SCR) est le produit de la variance estimée par le nombre de ddl. $R^2 = 1 - [(\sigma_1^2 \text{ ddl1} + \sigma_2^2 \text{ ddl2}) / \sigma_e \text{ ddle}]$ L'indice 1 : population des "mobiles" L'indice 2 : population des "non mobiles" l'indice e : ensemble de la population. (Algorithme dû à G.Forgeot, Insee)
<i>Nbre de voitures ds ménage</i>	3		***	
aucune voiture	1	-0,4141	***	
une voiture	0			
deux voitures	1	0,1567	***	
trois voitures et plus	1		ns	
<i>Abonnement/réduction</i>	1		***	
oui	1	0,3103	***	
non	0			
Nombre d'enfants ds ménage	1	-0,0589	***	
revenu (annuel) du ménage	1	5,21E7	***	
âge	1	-0,009	***	
POUVOIR EXPLICATIF			19%	

* La PCS exerce un impact significatif, montrant que des raisons autres que purement économiques jouent sur le comportement de mobilité. Par rapport aux employés, les cadres supérieurs sont en principe dotés d'un capital relationnel, social plus important, qui motive vraisemblablement une plus grande mobilité. En revanche, aucune différence n'a été détectée entre les professions intermédiaires (PI) et les employés, ce qui signifie ainsi un écart réel de comportement entre les cadres supérieurs et les PI.

Les agriculteurs et les indépendants ont une propension à se déplacer inférieure aux employés. Entre eux, il subsiste encore une différence (Statistique de Wald de 15,3 comparée au quantile du Chi2 à 5 ddl). Les agriculteurs présentent donc une probabilité inférieure aux indépendants (non agriculteurs). Alors qu'ils occupent une position dans la "hiérarchie sociale" considérée comme équivalente, les employés et les ouvriers se distinguent par une propension nettement inférieure pour les derniers. Dans le domaine de la mobilité, comme dans d'autres, les employés ont donc un style de vie qui se rapproche des cadres moyens (PI).

Certes ces remarques ne prennent pas en compte la composition du ménage en terme de PCS. Peut-être qu'une spécification introduisant une variable d'interaction sexe x PCS aurait été intéressante à tester.

* Plus le niveau de diplôme est élevé, plus la probabilité de mobilité est forte. Même entre deux catégories voisines, telles que les bacheliers et les diplômés de l'enseignement supérieur, une différence subsiste (statistique de Wald=12). Ainsi, le diplôme capture des effets que ne prend pas en compte la PCS. Il est certain qu'une dotation en capital culturel favorise la mobilité mais l'hypothèse d'un effet de structure lié à l'importance des étudiants parmi les diplômés (un quart) pourrait renforcer l'effet diplôme. Le statut d'étudiant est en effet incomplètement rendu par la modalité "inactivité" (dans laquelle figurent également des inactifs purs). Il s'agirait donc d'un type de mobilité particulier (retour au domicile familial pendant les congés par exemple) qui n'a que peu à voir avec le "capital culturel".

* La présence d'enfants (cf. la modalité de référence pour la variable type de ménage) semble freiner la mobilité. Les couples sans enfant et les personnes seules ont plus tendance à se déplacer sur longue distance (les deux étant à égalité). De plus, non seulement la présence d'enfants mais leur nombre influe aussi négativement sur la probabilité de se déplacer.

* la variable "taille de la commune" pouvait être considérée comme un proxy de l'offre en matière de transport collectif ; en effet, cette dernière varie de manière positive avec la taille de la commune. Néanmoins, cette variable se révèle non significative. L'introduction de la variable "région" était censée affiner cette mesure de l'offre de transport collectif. Cependant pour un non spécialiste en matière d'infrastructures de transport, les résultats sont assez difficilement interprétables. La variable est globalement significative.

L'interprétation de ces deux variables ne va pas de soi ; en effet, l'effet d'offre qui vient d'être exposé peut très bien être contrebalancé par l'effet de proximité dans l'offre de loisirs, de relations sociales etc. liées au caractère urbain d'un environnement. Ainsi, la mobilité est plutôt une mobilité de courte distance, qui se substitue à une mobilité de longue distance. L'effet de la taille de la commune sur la probabilité de se déplacer peut être négatif. Il est tout à fait possible que la significativité attachée au type d'habitation ressorte de cette explication, étant entendu que les habitations individuelles sont plus fréquentes en zone rurale ou modérément urbanisée.

* L'équipement automobile semble déterminant : ne pas avoir son permis de conduire ou de voiture dans le ménage diminuent la probabilité de se déplacer.

Un seul permis de conduire dans le ménage ne semble pas suffire pour entraîner une mobilité, l'effet reste le même qu'en l'absence de permis. Certes, les colinéarités avec la variable de possession du permis de conduire (sur le plan individuel) peuvent faire que les deux modalités possession du PC et un PC au sein du ménage entrent en concurrence dans l'explication du phénomène.

En résumé, la mobilité semble déterminée par trois types de facteurs :

- * **un facteur d'ordre économique (revenu et plus précisément de revenu "permanent"), car la mobilité de longue distance implique un coût financier qui peut être important,**
- * **des facteurs d'ordre sociologique et culturel : le capital relationnel, culturel accumulé par les catégories socialement privilégiées incite à la mobilité,**
- * **des facteurs liés à l'équipement du ménage et à son environnement de vie : l'équipement automobile est décisif dans un contexte où il assure près des trois quarts des déplacements sur longue distance (hors déplacements professionnels).**

Le statut des variables rendant compte de la localisation géographique et de la concentration urbaine est trop ambigu sans doute pour avancer une interprétation des résultats, sans recherche complémentaire. Néanmoins la variable type d'habitat donne à penser qu'une mobilité de proximité peut entrer en concurrence avec une mobilité de longue distance dans les grandes agglomérations.

1.4.2. Estimation du modèle de régression linéaire (2ème étape de la méthode d'Heckman)

La première étape de la méthode d'Heckman confirme la forte hétérogénéité dans la population en fonction de son comportement de mobilité considéré dans sa dimension dichotomique.

Aussi, une régression sur l'intensité de la mobilité par les MCO "standards" a toutes les chances d'être biaisée (pour un nombre d'observations grand, il y a équivalence entre la non convergence et l'existence d'un biais), compte tenu de la sélection opérée dans la population si l'on ne retient que les mobiles.

Le fait que le coefficient attaché au ratio de Mill soit significatif dans la régression du nombre de déplacements et du nombre de voyages en atteste ; il y a bien un biais de sélection attaché à cette population, que l'introduction de cette variable supplémentaire permet de corriger. Comparant deux régressions, l'une sans, l'autre avec le ratio de Mill, on constate que les résultats sont très différents, un certain nombre de variables n'étant plus significatives dans le deuxième cas.

Son absence correspond au cas de figure d'une variable (pertinente) omise : non seulement l'estimateur des MCO est alors biaisé, mais aussi on en accroît la précision. Aussi son introduction va de paire avec une réduction de la significativité des variables.

Trois variables dépendantes - le nombre de déplacements, le nombre de voyages et la distance totale parcourue - ont été tour à tour retenues dans l'estimation du modèle TOBIT (tableau 2.3).

Tableau 2.3 : Indication des seuils de significativité des variables explicatives introduites dans la modélisation de l'intensité de mobilité (estimation par les MCO) sur les données complètes (6278 observations) - 2ème étape du modèle TOBIT -						
variables dépendantes modèles n°	nombre de déplacements		nombre de voyages		distance totale	
	1	2	3	4	5	6
taille de la commune	ns	ns	ns	ns	***	***
région	***	***	***	***	ns	ns
occupation actuelle	ns	**	**	***	ns	ns
PCS	ns	***	**	***	ns	ns
niveau d'études	ns	ns	ns	ns	***	***
type d'habitation	***	***	***	***	ns	ns
possession d'une résid. secondaire	ns	ns	ns	***	ns	ns
type de ménage	*	ns	ns	ns	**	**
nombre d'enfants	ns	ns	**	*	ns	ns
revenu	***	***	*	***	***	***
sexe	***	***	***	***	***	***
possession du permis de conduire	ns	ns	ns	ns	ns	ns
nombre de PC dans le ménage	ns	ns	**	*	***	***
nombre de voitures dans le ménage	ns	ns	ns	**	**	***
abonnement/réduction	ns	***	ns	***	***	***
ratio de Mill	**		**		ns	
R ²	4,2%	4,2%	6%	5,9%	7,5%	7,5%

Ainsi dans le modèle n°1, les variables d'occupation, de PCS, de niveau d'études, de possession de résidence et d'équipement automobile du ménage ne sont plus significatives alors qu'elles l'étaient dans le modèle dichotomique de mobilité. Il semble donc qu'il subsiste des hétérogénéités dans la population des mobiles mais qu'elles soient moins marquées (limitées à un nombre restreint de critères) que dans l'ensemble de la population. Il n'est donc pas surprenant que ces hétérogénéités soient davantage apparentes sur le nombre de voyages qui constitue une mesure alternative de la mobilité (modèle 3), plus proche de la dimension dichotomique.

Il peut en effet être ici souligné l'aspect quelque peu factice d'une mesure de la mobilité en terme de déplacements, qui composent d'un voyage. Le concept de voyage se rattache donc davantage au fait de "sortir" de chez soi qui est aussi contenu dans l'aspect dichotomique de la mobilité.

Si comme le raisonnement précédent le suppose, la véritable frontière passe entre ceux qui restent chez eux et ceux qui "sortent", il n'est pas surprenant qu'aucun biais de sélection n'ait été détecté dans le modèle n°5. La partition selon l'intensité de mobilité mesurée par la distance totale parcourue sur l'ensemble de la période paraît trop fine pour discriminer réellement la population. De plus, l'intensité de mobilité ainsi mesurée n'est pas de la même nature que les deux autres : ces dernières renvoient à un concept de fréquence, alors que la distance totale est la résultante d'une fréquence et d'une distance moyenne, qui peuvent se compenser l'une l'autre et rendent donc de ce fait la mesure moins discriminante.

Ces considérations sur l'importance de l'approche dichotomique par rapport à une approche quantitative pour analyser la mobilité nous ont incités à tester une voie médiane : un modèle d'intensité de mobilité trichotomique ordonné.

1.5. Modélisation trichotomique ordonnée de la mobilité

Cette approche nécessite tout d'abord de déterminer un découpage relativement équilibré et le plus significatif possible entre un nombre restreint de catégories. Les conclusions émises à propos du modèle TOBIT nous conduisent à abandonner la variable de distance totale, trop floue dans sa définition par rapport à notre problématique et relativement peu discriminante. Pour le nombre de déplacements, deux découpages sont envisagés, les mêmes que pour les profils de mobilité (tableau 2.1). Les combinaisons étant plus limitées pour le nombre de voyages, un seul découpage est testé.

Le tableau suivant rappelle la distribution respective de ces différentes mesures de la mobilité dans la population de l'échantillon :

Tableau 2.4 : Description des différentes partitions possibles de la population selon son intensité de mobilité.		
nombre de déplacements (1) 0 : 44,4 % 2 ou 3 : 22,3% 4 et plus : 33,3%	nombre de déplacements (2) 0 : 44,4 % 2 à 4 : 34% 5 et plus : 21,6%	nombre de voyages 0 : 44,4 % 1 : 25% 2 et plus : 30,6%

Les variables explicatives introduites dans les modèles économétriques sont les mêmes que celles du modèle TOBIT, avec quelques différences de forme : les régions ont été agrégées en deux catégories : Ile de France et Province. Au sens des critères de mesure de la qualité générale du modèle (cf. annexe 2.5), il vaut mieux introduire le détail des régions.

Néanmoins, le gain est assez limité, aussi seule la version "allégée" est étudiée.

Quelque soit la variable dépendante, les résultats des estimations du modèle trichotomique sont très proches : on retrouve les mêmes modalités significatives, affectées des mêmes signes, à l'exception des professions intermédiaires qui ne se différencient plus des employés (modalité de référence) dans l'analyse du nombre de déplacements (2) et de la tranche d'âge 17-25 ans qui se révèle exercer une influence significative et négative sur le nombre de voyages.

Compte tenu de la similitude d'ensemble, la présentation de la régression du nombre de déplacements (2) qui paraissait plus discriminant du point de vue des distributions statistiques a été choisie.

Tableau 2.5 : Résultats du modèle LOGIT ordonné ; variable dépendante : nombre de déplacements (2). niveau 1 (5 déplacements et plus)=3008 observations - niveau 2 (2 à 4 déplacements)=4735 observations - niveau 3 (0 déplacement)=6181 observations.				
(1) variables	(2) DF	(3) coefficient	(4) probabilité critique	Guide de lecture
Constante 1	1	-2,2116	***	La procédure logistice de SAS modélise la variable d'intensité de mobilité (mesurée par le nombre de déplacements) sous l'hypothèse d'égalité des pentes.
Constante 2	1	-0,3778	***	
Taille de la commune	4			La différenciation des niveaux n'est exprimée que par la constante. Celle-ci est normée à 0 pour le niveau 3, et est estimée pour les deux autres niveaux : constante 1 pour le niveau 1 et constante 2 pour le niveau 2.
communes rurales	1		ns	
communes < 20000 habitants	1		ns	
comm. de 20 à 100000 hab.	1		ns	Leur valeur n'est en général pas interprétée, mais le fait qu'elles soient significatives indique que la partition effectuée l'est elle-même.
comm > 100000 hab. sf Paris	0	0		
unité urbaine de Paris	1	0,4785	***	
Localisation	1			
Ile de France	1	-0,3969	***	SAS fournit un test de l'égalité des pentes : Soit le modèle $P(Y \leq i) = F(a_i + Xb_i)$. $b_i = (b_{i1}, b_{i2}, \dots, b_{iS})$ le vecteur des s paramètres de pente et a_i la constante correspondant à la modalité i . Dans l'hypothèse d'égalité des pentes : $b_{1m} = b_{2m} = \dots = b_{km}$ pour tout $m=1$ à s , ($k+1$ =nombre de valeurs prises par la variable dépendante). On définit alors une statistique de score qui tend asymptotiquement vers une distribution du χ^2 à $s(k-1)$ degrés de liberté. L'hypothèse d'égalité des pentes est vérifiée si la statistique du score ne dépasse pas un seuil α . Dans cette estimation, $\alpha=0,0001$.
Province	0			
Occupation actuelle	2			
actifs	1	0,142	***	
chômeurs	1	-0,1819	**	
inactifs	0			
PCS	5			
agriculteurs	1	-0,6446	***	
autres indépendants	1	-0,3006	***	
cadres supérieurs	1	0,2979	***	
professions intermédiaires	1		ns	
employés	0			
ouvriers	1	-0,4304	***	
Niveau d'études	4			
CEP et moins	0			
BEPC	1		ns	
BEP-CAP	1	0,1776	***	
Bac	1	0,4518	***	
Etudes supérieures	1	0,7598	***	

<i>Type d'habitat</i>	1		
maison	0		
immeuble	1		ns
<i>Résidence secondaire</i>	1		
non	0		
oui	1	0,6703	***
<i>Type de ménage</i>	4		
seul	1	0,7427	***
couple sans enfants	1	0,4887	***
couple avec enfants	0		
familles monoparentales	1	0,1484	**
autre ménage	1		ns
<i>Sexe</i>	1		
homme	1	0,0926	***
femme	0		
<i>Possession permis de cond.</i>	1		
oui	1	0,2534	***
non	0		
<i>Nombre de PC ds le ménage</i>	3		
aucun	1	-0,3568	**
un	1	-0,2179	***
deux ⁷	0		
trois et plus	1	-0,1896	***
<i>Nbre de voitures ds ménage</i>	3		
aucune voiture	1	-0,5259	***
une voiture	0		
deux voitures et plus	1	0,1918	***
<i>Abonnement/réduction</i>	1		
oui	1	0,2343	***
non	0		
<i>Revenu</i>	4		
revenu annuel < 30000F	0		
revenu de 30 à 75000F	1	-0,3839	***
revenu de 75 à 125000F	1	-0,1178	**
revenu de 125 à 205000F	1		ns
revenu > 205000F	1	0,4122	***
<i>Age</i>	5		
< 17 ans	1	0,2467	**
17 à 25 ans	1		ns
26 à 35 ans	1		ns
36 à 50 ans	0		
51 à 65 ans	1	-0,1801	***
> 65 ans	1	-0,5428	***

En colonne 2 est indiqué le nombre de degrés de liberté (DF). L'estimation de l'effet de chacune des modalités des variables figure en colonne 3.

La colonne 4 donne le seuil de significativité (ou probabilité critique) du test de $H_0 : b_k = 0$ contre $H_1 : b_k \neq 0$, relativement à la modalité de référence (ligne ombrée). $\alpha < 1\%$ représenté par ***. $1\% < \alpha < 5\%$ représenté par **. $5\% < \alpha < 10\%$ représenté par *.

Commentaires sur la spécification :
L'âge a été introduit par tranches, afin de tenir compte de valeurs d'élasticité éventuellement différentes, ainsi que la description des profils selon l'intensité le laisse à penser. (le revenu également).

Par rapport à la modélisation dichotomique :

* On voit émerger l'impact d'habiter en région parisienne, positif, tandis que la modalité Ile de France est affectée d'un coefficient négatif. En s'inspirant des réflexions suggérées par la modélisation dichotomique, le premier résultat peut s'interpréter comme celui de l'offre de transports collectifs, qui devient plus sensible lorsqu'on s'intéresse à l'intensité de la mobilité qu'à la mobilité en général. L'accroissement de la part des moyens de transports collectifs avec celle de l'intensité de la mobilité milite pour cette interprétation.

Par ailleurs, le coefficient négatif attaché à la modalité Ile de France peut résulter de la concurrence entre mobilité de proximité et mobilité de longue distance, aux dépens de la seconde. Mais alors, peut-on se demander pourquoi pas l'inverse, dans la mesure où les deux modalités recouvrent des situations en grande partie identique. Tout semble se passer comme si l'offre de transport opposait davantage Paris aux autres communes que L'Ile de France à la Province et comme si la proximité opposait davantage l'Ile de France à la Province que Paris et les autres communes. Cette hypothèse demeure ouverte aux débats.

* Demeure essentiellement l'opposition entre cadres supérieurs et les autres catégories de PCS : les tests de Scheffé considèrent qu'il n'y a pas de différence entre agriculteurs et indépendants, entre indépendants et ouvriers.

* L'intensité de la mobilité n'évolue pas linéairement avec le niveau du revenu : elle est supérieure pour la tranche de revenus quasi-nuls aux tranches de revenus "modestes". Elle augmente ensuite significativement et nettement pour les tranches de revenus les plus élevées ; ceci peut s'expliquer en partie par la légère sur-représentation des jeunes dans cette tranche de revenus (30% alors qu'ils constituent 23% de la population). Ils forment certes des ménages à part, mais probablement bénéficient-ils d'un soutien financier familial. Il pourrait également s'agir d'un mode de vie propre aux jeunes, plus axés sur la mobilité, mais la variable d'âge est censée capter cet effet. Or, si elle est positive pour les moins de 17 ans, elle ne l'est pas dans la tranche 18-25 ans, susceptible de former des ménages autonomes. Il y certainement une concurrence entre ces deux modalités pour capter cet effet "mode de vie".

* La tranche d'âge 26-35 ans apparaît, avec les moins de 17 ans, avoir l'intensité de mobilité la plus élevée. Dans le cas de ces derniers, il s'agit d'enfants, selon toute vraisemblance, qui accompagnent leurs parents. Les plus de 65 ans sont les plus en retrait des phénomènes de mobilité.

* La présence d'enfants paraît toujours handicapante pour se déplacer, tandis que les personnes seules semblent avoir une intensité de mobilité supérieure aux couples sans enfant, alors que leur situation était équivalente au regard de la mobilité d'ensemble.

* Le sexe émerge comme une variable significative, comparé au modèle dichotomique : les hommes se déplaceraient plus fréquemment que les femmes.

En résumé, cette modélisation trichotomique permet de nuancer le modèle dichotomique, sans le remettre en cause. Par ailleurs, elle est d'un "emploi" plus aisé que le modèle TOBIT, relativement lourd à mettre en oeuvre et vraisemblablement plus adaptée au phénomène étudié, pour lequel il existe sans aucun doute des effets de seuils. La lecture du modèle LOGIT ordonné est de plus aussi abordable que celle du LOGIT ou PROBIT dichotomique.

II. DEPLACEMENTS PRIVÉS : QUEL MODE DE TRANSPORT ?

Comme dans l'étude de la mobilité, on proposera d'abord une première approche des populations distinguées selon leur mode de transport à l'aide de statistiques simples : moyenne d'âge des individus ayant choisi la voiture, proportion de ruraux parmi les personnes ayant choisi le voiture, etc. Cette approche ne donne qu'une vision parcellaire des choses. C'est pourquoi, dans un deuxième temps une méthode plus analytique sera utilisée, la modélisation, afin de déterminer quels sont les facteurs les plus discriminants dans le fait de choisir un mode de transport plutôt qu'un autre.

A ce stade de l'étude, il faut toutefois souligner le fait que cette approche reste imparfaite étant donné que l'on ne dispose pas de données sur l'offre pour ces modes de transport (infrastructures et coûts par exemple). Comme déjà évoqué précédemment la taille de la commune ou encore le fait d'être en Ile de France ou dans telle région peuvent en constituer des indicateurs mais qui restent très approximatifs. Par conséquent, toute une dimension des déterminants de choix des individus ne sera pas prise en compte.

Les personnes se déplaçant pour des motifs privés représentent une grande majorité de la population (98 %). Environ la moitié n'ont fait qu'un seul voyage. Dans un premier temps, nous allons d'abord

chercher à savoir ce qui distingue les personnes ayant choisi la voiture de celles ayant choisi les transports collectifs. La deuxième partie de l'analyse prolongera ces résultats en étudiant en détail les choix modaux pour les transports collectifs : car, train et avion.

2.1. Voiture contre transports collectifs

2.1.1. Première approche

Près des trois-quarts des individus ont choisi la voiture plutôt que les transports collectifs. Le tableau 2.6 présente quelques caractéristiques des profils des deux sous-populations selon le mode choisi.

Tableau 2.6 : Profil de l'échantillon selon le mode de transport choisi (déplacements privés)		
Caractéristiques	Voiture	Collectif
<i>Age*</i> moyenne (37,08) coefficient de variation** (52 %)	36,9 50,8 %	37,7 56,1 %
<i>Revenu</i> moyenne (126 753) coefficient de variation (101,53 %)	124 257 97,4 %	133 739 110,7 %
<i>Nombre d'enfants</i> moyenne (1,31) coefficient de variation (96,86 %)	1,34 92,5 %	1,22 109,7 %
<i>Taille du groupe accompagnateur</i> moyenne (4,4) coefficient de variation (225 %)	2,5 111,9 %	9,5 182,5 %
lecture : parmi les personnes ayant choisi la voiture, 7,8 % sont des personnes vivant seules et 58,6 % vivent en couple et ont des enfants		
femmes	49,4	50,6
personnes vivant seules	7,8	17,4
couples avec enfants	58,6	46,8
cadres	36,9	45,9
employés / ouvriers	52,5	44,9
inactifs et jeunes	45,7	57,6
actifs	49,3	37,9
personnes ayant le BAC ou plus	35,2	45,9
ménages habitant l'Île de France	18,9	30,5
ménages habitant en commune rurale	25,1	16,5
ménages habitant une maison	68,9	52,1
ménages où personne n'a le permis	2,5	11,1
effectifs non pondérés	5449	2026
effectifs pondérés	20,6 millions	7,4 millions

* les chiffres entre parenthèses se réfèrent à l'ensemble

** coefficient de variation : (rapport de l'écart-type à la moyenne) x par 100

Il ressort de l'examen de ces statistiques que :

* En moyenne, les usagers de transports collectifs sont un peu plus âgés, vivent dans des ménages ayant un revenu un peu plus élevé et ont moins d'enfants que les personnes utilisant la voiture. Cependant, il faut être prudent sur ces résultats. En effet, on remarque que la population ayant choisi le mode collectif est plus hétérogène : les coefficients de variation sont toujours plus élevés au sein de cette catégorie ce qui traduit une plus forte dispersion. Ainsi pour l'âge par exemple, l'examen complémentaire de tableaux de contingence montre que les deux catégories extrêmes, les

jeunes et les plus de 60 ans, sont relativement plus importantes parmi les gens ayant opté pour les transports collectifs que parmi ceux ayant choisi la voiture.

* Il y a peu de différence entre les hommes et les femmes ; ce résultat semble récurrent dans les différentes analyses menées. Parmi les individus ayant choisi la voiture, on note une forte majorité de personnes vivant en couple et d'employés/ouvriers. Les actifs ont plus privilégié la voiture que les inactifs en général (retraités et scolaires en majorité) ainsi que les personnes ayant un diplôme inférieur au BAC.

* Les personnes vivant à la campagne ou dans de petites villes sont relativement plus nombreuses à avoir utilisé la voiture. De même pour celles habitant une maison. Ces résultats traduisent sans doute pour une bonne part le fait que l'offre de transports collectifs soit moins importante ou moins accessible en milieu rural qu'en milieu urbain. Ainsi les personnes habitant dans des villes de plus de 100 000 habitants représentent près de 60 % de la population ayant opté pour un mode collectif.

2.1.2. Modélisation et estimation

Pour rechercher les variables les plus discriminantes dans le choix du mode de transport, le modèle LOGIT dichotomique suivant a été estimé :

$Y_i = 1$ si l'individu i a choisi la voiture

$Y_i = 2$ si l'individu a choisi un transport collectif (car, train ou avion)

La spécification du modèle et les résultats de l'estimation sont présentés dans le tableau 2.7. Pour rendre le modèle estimable, nous avons choisi comme modalités de référence les classes modales de chaque variable explicative. Les critères de qualité du modèle sont les suivants :

Statistique de test	modèle avec uniquement la constante	modèle complet	Chi-deux (prob.)
Akaike	8292.957	6537.543	
Schwartz	8299.876	6738.203	
Max. de vraisemblance	8290.957	6479.543	1811.41 (0.0001)
Score			1611.87 (0.0001)

Globalement, d'après les tests du rapport des maxima de vraisemblance et du score, le modèle est significatif. De plus, il y a concordance entre choix prédit par le modèle et choix observé dans 80 % des cas.

Dans le tableau 2.7, n'ont été retenus que les coefficients significatifs jusqu'à un seuil maximum de 10 %. Les variables apparaissant comme les plus discriminantes sont en premier lieu celles qui ont trait aux disponibilités du ménage quant à l'utilisation de la voiture (motorisation, possession du permis) ou à celle des transports collectifs, appréhendée ici par l'existence ou non de réductions ou d'abonnements.

Comme on pouvait s'y attendre, le degré de motorisation du ménage a en effet un rôle important. D'après l'Odds Ratio¹, la probabilité qu'un individu choisisse la voiture est trois fois plus élevée si le

¹ L'Odds Ratio est le rapport de probabilité entre la modalité active et la modalité de référence.

**Tableau 2.7 : Estimation du modèle dichotomique "voiture contre collectifs"
(motifs privés)**

Variables	DF	Coefficient	Probabilité	Odds ratio
<i>Constante</i>	1	2,6734		
<i>Taille de la commune</i>			ns	
communes rurales petites et moyennes	1	0,1908	**	1,210
grandes agglom. Paris			ns	
<i>PCS</i>				
agriculteurs & indépendants	1	-0,2400	*	0,787
cadres	1	-0,2080	**	0,812
employés/ouvriers				
<i>Occupation actuelle</i>				
actifs				
chômeurs			ns	
jeunes	1	-0,5871	***	0,556
inactifs			ns	
<i>Niveau d'études</i>				
CEP et moins	1	0,1807	*	1,198
BEPC-BEP	1	0,2334	***	1,263
BAC et plus				
<i>Usage d'une résidence secondaire</i>				
non				
oui	1	0,1402	*	1,151
<i>Abonnement / réduction</i>				
non				
oui	1	-0,7701	***	0,463
<i>Type d'habitat</i>				
maison				
immeuble	1	-0,2672	***	0,766
<i>Type de ménage</i>				
seul	1	-0,5457	***	0,579
couple			ns	
couple avec enfants				
monoparental			ns	
autres ménages	1	-0,7396	***	0,477
<i>Possession d'un permis</i>				
oui				
non	1	-0,3324	***	0,717
<i>Sexe</i>				
homme				
femme			ns	
<i>Nombre de voitures dans le ménage</i>				
aucune voiture	1	-1,0933	***	0,335
une voiture				
deux et plus			ns	
<i>Nombre de permis dans le ménage</i>				
pas de permis			ns	
un permis			ns	
deux permis				
trois et plus	1	-0,4279	***	0,652
<i>nombre d'enfants</i>			ns	
âge	1	-0,00589	*	0,994
revenu du ménage	1	-4,12E-7	*	1,000
distance parcourue	1	-0,00155	**	0,998

ligne ombrée : modalité de référence pour la variable considérée

*** significatif au seuil de 1 % - ** significatif au seuil de 5 % - * significatif au seuil de 10 %

ménage est motorisé. De même, lorsque l'on dispose d'une réduction ou d'un abonnement (ce qui vaut essentiellement pour les transports collectifs), la probabilité d'utiliser la voiture est divisée par deux.

Le fait que l'individu ne possède pas le permis a un effet légèrement réducteur sur la préférence pour la voiture. Par contre, en ce qui concerne le nombre de permis dans le ménage, la différence la plus marquée est entre les ménages ayant plus de trois permis et ceux n'en possédant que deux. Les premiers préfèrent moins la voiture que les seconds. A ce sujet il pourrait être intéressant de prolonger l'analyse en cherchant à savoir quelles sont les caractéristiques respectives de ces deux populations, cette différence dans les préférences provenant sans doute d'autres variables.

Pour les caractéristiques socio-démographiques, les coefficients du modèle confirment certains résultats déjà mentionnés dans les profils de population. Ainsi, on prend moins la voiture lorsque l'on vit seul plutôt qu'en couple ou que l'on est jeune et inactif plutôt qu'actif.

Les cadres et les indépendants semblent avoir le même comportement de choix puisque pour ces deux catégories la probabilité de prendre la voiture est moindre que pour les employés/ouvriers. On a cependant cherché à savoir s'il existait une différence significative entre le coefficient estimé pour les cadres (-0.2080) et celui des indépendants et agriculteurs (-0.24) par le test de Wald en utilisant la règle de Scheffé. Le test aboutit à accepter l'hypothèse que les coefficients sont égaux. Il n'existe donc pas de différence significative entre les comportements de ces deux catégories.

De même pour les deux catégories de personnes ayant un diplôme inférieur au BAC, les coefficients ne sont pas significativement différents. De façon équivalente, ces individus préfèrent plus la voiture que les personnes ayant le BAC ou plus. Enfin on utilise plus la voiture si l'on vit dans une ville moyenne plutôt que dans une grande ville et si l'on habite une maison.

Les trois variables socio-démographiques introduites sous forme continue (âge, revenu et nombre d'enfants) ne se distinguent presque pas en matière d'influence sur le choix du mode de transport : les coefficients et leur significativité sont très faibles. Un deuxième modèle a été estimé en introduisant ces variables sous forme qualitative. La qualité globale du modèle reste inchangée : les statistiques d'Akaike et de Schwarz sont semblables et le modèle reste significatif. Les coefficients des anciennes variables sont un peu moins forts en valeur absolue que pour le premier modèle mais restent de même signe. Cependant, parmi les nouvelles modalités seule la catégorie des plus de 60 ans a un coefficient significatif. D'après ce deuxième modèle, les personnes âgées ont une préférence moindre pour la voiture que les moins de 25 ans.

2.2. Quel mode de transport collectif ?

2.2.1. Première approche

Le train est le premier transport collectif choisi (46 %), l'avion et le car ont été pris respectivement par 30 % et 24 % des individus. La moitié des personnes interrogées n'ont fait qu'un seul voyage. Le tableau 2.8 fournit les principales caractéristiques des trois sous-populations.

Il ressort que :

* Parmi les personnes ayant choisi le car, les employés/ouvriers sont largement majoritaires (66 %) ainsi que les personnes logeant en maison individuelle (75 %). Les inactifs et les habitants d'une commune rurale sont proportionnellement plus nombreux à avoir utilisé le car plutôt qu'un des deux autres modes de transport. En moyenne, il s'agit d'individus ayant des revenus moins élevés que pour les deux autres catégories et ayant plus d'enfants. On remarque la très forte moyenne de la taille du groupe ayant voyagé avec l'interrogé. Cette indication et les résultats précédents laissent à penser

qu'une bonne part de ces déplacements se situent dans le cadre de voyages touristiques organisés et peut-être même peut-on avancer pour une clientèle de retraités.

* Dans la population des utilisateurs de l'avion, on remarque une forte proportion de cadres, d'actifs et de personnes ayant le BAC ou plus. On peut noter aussi que près des trois-quarts de cette population possèdent le permis voiture contre seulement 58 % pour les utilisateurs du train et encore un peu moins pour le car. La moyenne des revenus est la plus forte dans cette catégorie ce qui semble peu surprenant, l'avion étant un moyen de transport pour lequel des considérations de coûts sont encore souvent dissuasives.

* En général, la population des utilisateurs du train est moins contrastée que celles des usagers du car ou de l'avion. Si l'on fait abstraction du cas de l'avion pour lequel la contrainte de budget est importante, il semblerait que le train soit plus souvent un substitut à la voiture que le car. En effet, les individus appartenant à des ménages non motorisés ou à des ménages où personne n'a le permis sont plus fréquents parmi les utilisateurs du train. De même les individus ayant le permis y représentent presque 60 % contre 45 % pour le car. La proportion de personne vivant seules est plus importante pour le train que pour les deux autres modes.

Tableau 2.8 : Profil de l'échantillon selon le mode de transport collectif choisi (déplacements privés)			
Caractéristiques	Car	Train	Avion
<i>Age</i>			
moyenne	38,1	37,51	37,72
coefficient de variation*	63,5 %	55,86 %	49,62 %
<i>Revenu</i>			
moyenne	96 446	129 788	169 160
coefficient de variation	102,7 %	106,32 %	108,04 %
<i>Nombre d'enfants</i>			
moyenne	1,4	1,17	1,15
coefficient de variation	96,7 %	119,18 %	105,89 %
<i>Taille du groupe</i>			
moyenne	29,8	2,7	3,9
coefficient de variation	76,2 %	270,2 %	241,1 %
lecture : parmi les personnes ayant choisi le car 12,5 % sont des personnes vivant seules et 52,5 % vivent en couple et ont des enfants			
femmes	54,7	51,4	48,1
personnes vivant seules	12,5	21,6	15,22
couples avec enfants	52,5	41,1	19,6
cadres	18,4	47,4	59,3
employés / ouvriers	66,3	46,9	30,5
inactifs	36,5	25,5	21,7
actifs	24,0	34,5	51,3
personnes ayant le BAC ou plus	18,8	51,8	56,8
ménages habitant l'Ile de France	5,5	34,9	43,3
ménages habitant une commune rurale	35,2	10,9	10,8
ménages habitant une maison	74,8	42,9	47,9
ménages où personne n'a le permis	10,5	14,9	5,9
personne ayant le permis	44,7	57,9	74,4
ménages n'ayant pas de voiture	16,0	30,7	13,2
effectifs non pondérés	463	936	630
effectifs pondérés	1,7 millions	3,3 millions	2,2 millions

* coefficient de variation : (rapport de l'écart-type à la moyenne) x par 100

2.2.2. Modélisation

Pour modéliser le choix entre avion, train ou car nous avons utilisé un modèle LOGIT polytomique non ordonné mis en oeuvre avec la procédure CATMOD du logiciel SAS. Il peut s'exposer comme suit :

On modélise $Y_i =$

1	si l'individu a choisi l'avion
2	" le car
3	" le train

Notons $C=\{1,2,3\}$ l'ensemble des choix possibles de modes de transport.

Soit Y_{ih}^* l'utilité de l'individu i s'il choisit le mode h .

On suppose alors que $Y_i = h \Leftrightarrow Y_{ih}^* \geq Y_{ih'}^*$ quel que soit h' .

Le modèle s'écrit :

$$Y_i = 1 \text{ ou } 2 \text{ ou } 3$$

$$Y_{ih}^* = x_i' b_h + u_{ih} \quad \text{où } x_i \text{ représente les caractéristiques socio-démographiques de } i$$

Y_{ih}^* inobservable

$$Y_i = h \Leftrightarrow Y_{ih}^* \geq Y_{ih'}^* \text{ quel que soit } h'$$

u_{ih} indépendants et identiquement distribués, de fonction de répartition $F(z) = \exp(-\exp(-z))$

Il y a donc H (ici $H=3$) vecteurs de paramètres à estimer. On suppose souvent comme hypothèse identifiante que l'un des vecteurs est nul. La procédure CATMOD prend comme référence la modalité ayant l'indice le plus élevé, en l'occurrence le train.

L'estimation se fait par le maximum de vraisemblance. L'estimateur obtenu est asymptotiquement normal et asymptotiquement efficace.

Le modèle LOGIT polytomique non ordonné repose sur l'hypothèse d'indépendance vis-à-vis des alternatives non pertinentes (IIA pour Independence from Irrelevant Alternatives). Cette peut s'exprimer de la façon suivante : si l'on réduit l'ensemble des choix en excluant une éventualité jugée non pertinente, l'avion par exemple, il ne doit pas y avoir de modification dans les rapports de probabilités de choix pour les deux modes restant à savoir, le car et le train. Soit, en d'autres termes, le fait que l'individu choisisse plutôt le car que le train par exemple n'est pas influencé par le fait que l'avion figure ou non dans son éventail de choix.

Sous forme probabilisée, en reprenant les indices du modèle, on doit avoir :

$$P_{2/123} / P_{3/123} = P_{2/23} / P_{3/23}$$

Si les alternatives 3 (train) et 1 (avion) sont fortement substituables, cette hypothèse à la base du modèle LOGIT multinomial risque d'être violée.

Une des méthodes employées pour tester l'indépendance est d'appliquer le test d'Hausman et de McFadden. Si l'hypothèse IIA est vérifiée, les paramètres estimés pour les variables explicatives devraient rester inchangés lorsqu'on effectue l'estimation sur un sous-ensemble de choix R inclus dans C , ensemble des choix de départ. Donc :

Soit $R \subset C$.

$\hat{\delta}_C$ et \hat{V}_C sont les estimateurs du maximum de vraisemblance du paramètre δ et de sa matrice de variance-covariance sur C .

$\hat{\delta}_R$ et \hat{V}_R sont leurs estimateurs pour le sous-ensemble R .

Le test de l'hypothèse IIA revient à tester l'hypothèse nulle : $H_0 : \hat{\delta}_R - \hat{\delta}_C = 0$.

Ce test est basé sur la statistique suivante : $\hat{q} = (\hat{\delta}_R - \hat{\delta}_C)' (\hat{V}_R - \hat{V}_C)^{-1} (\hat{\delta}_R - \hat{\delta}_C)$

$\hat{q} \rightarrow \chi^2(p)$ avec p : nombre de variables explicatives

2.2.3. Estimation

La mise en oeuvre de l'estimation s'est révélée plus délicate que pour les modèles dichotomiques. Beaucoup de spécifications aboutissaient à des modèles non significatifs. En définitive, le modèle retenu, estimé sur 1556 observations pondérées, contient les variables suivantes :

Source	DF	Chi-deux	Probabilité
Constante	2	14.72	0.0006
Taille de la commune	6	126.80	0.0000
PCS	4	58.23	0.0000
Nombre d'enfants	4	11.99	0.0175
Nombre de voitures	4	56.45	0.0000
Occupation actuelle	6	38.93	0.0000
Significativité du modèle	404	444.61	0.0797

On remarque en particulier que le revenu n'apparaît pas, ce qui n'est pas conforme à nos attentes suite à l'étude des profils.

Le tableau 2.9 fournit les résultats de l'estimation (les coefficients significatifs figurent en gras).

Par rapport aux habitants des petites villes et des villes moyennes, les personnes habitant en commune rurale ont une probabilité plus importante de prendre le car plutôt que le train. Cette préférence se retrouve mais dans une moindre mesure pour les habitants des grandes villes. En outre ces derniers utilisent moins l'avion que train.

Les agriculteurs et indépendants ont une préférence tant pour le car que pour l'avion par rapport aux employés et ouvriers mais ils forment une catégorie peut être trop hétérogène. Par contre les cadres choisissent nettement moins le car, tandis que leur probabilité de prendre l'avion n'est pas très différente de celle de prendre le train, toutes choses égales par ailleurs. Autrement dit, leur sur-représentation dans les usagers de l'avion est à relier à d'autres caractéristiques que leur statut professionnel en lui-même.

On prend moins le car que le train lorsque l'on a un enfant par rapport aux personnes n'en ayant pas. Les personnes appartenant à des ménages non motorisés semblent préférer le train à tout autre mode de transports collectifs. Bien que significative dans son ensemble, la variable "occupation actuelle" ne ressort dans aucune de ses modalités par rapport à celle de référence à savoir, les actifs.

Tableau 2.9 : Estimation du modèle polytomique non ordonné "transports collectifs" (train, avion et car pour motifs privés)				
Le train est le mode de référence				
Variables	Coefficient		Significativité	
	avion	car	avion	car
<i>Constante</i>	-0.4992	-1.3505		
<i>Taille de la commune</i>				
communes rurales	-0.2597	1.3014	ns	***
petites et moyennes				
grandes	-0.3596	0.4610	***	***
agglomération de Paris	0.0908	0.0766	ns	ns
<i>PCS</i>				
agriculteurs & indépendants	0.5435	0.4435	***	**
cadres	-0.0649	-0.7694	ns	***
employés/ouvriers				
<i>Occupation actuelle</i>				
actifs				
chômeurs	0.3257	0.2163	ns	ns
"jeunes" (scolaires, militaires)	0.1267	-0.3611	ns	ns
inactifs	-0.3920	-0.9189	ns	ns
<i>Nombre d'enfants</i>				
pas d'enfant				
un enfant	-0.0049	-0.4326	ns	***
deux enfants et plus	0.1209	0.2029	ns	ns
<i>Nombre de voitures dans le ménage</i>				
aucune voiture	-0.8276	-0.2835	***	*
une voiture				
deux voitures et plus	0.0395	0.1585	ns	ns

ligne ombrée : modalité de référence pour la variable considérée

*** significatif au seuil de 1 % - ** significatif au seuil de 5 % - * significatif au seuil de 10 %

2.2.4. Remarques et discussion sur l'IIA

D'une façon générale, les résultats de l'estimation de ce modèle sont difficilement exploitables. A ce sujet, on peut formuler deux remarques. D'une part, l'on ne dispose pas des variables rendant compte des attributs propres aux moyens de transport utilisés (coût, confort, temps...). D'autre part, l'hypothèse d'indépendance vis-à-vis des choix non pertinents risque de ne pas être vérifiée. En effet, cette étude est réalisée sur des milieux très hétérogènes en matière d'offre de transports collectifs, (ce qui rejoint un peu la première remarque). Il est en effet évident que les déterminants de choix et les possibilités de substitutions entre les différents modes ne sont pas les mêmes selon que la personne interrogée habite une grande ville (ou près d'une grande ville) ou une commune rurale.

Le test de l'hypothèse IIA selon la méthode de Hausman McFadden n'est pas mis en oeuvre ici mais nous avons cependant cherché à nous en approcher en complétant l'analyse par deux modèles LOGIT dichotomiques (les résultats détaillés de ces estimations figurent en annexe 2.5).

Le modèle polytomique initial modélise la probabilité de prendre le car et celle de prendre l'avion par référence au train. Dans un premier modèle LOGIT dichotomique, on a estimé la probabilité de prendre le car par rapport au train lorsque l'avion est absent de l'univers des choix possibles. Les coefficients ainsi obtenus peuvent être comparés à ceux issus de l'estimation trichotomique. Si l'IIA est vérifiée ces coefficients devraient être peu différents. Or sur les 13 coefficients comparés 8 sont de signes opposés. Le deuxième modèle dichotomique estime la probabilité de prendre l'avion plutôt que le train lorsque le car est exclu. Cette fois 6 coefficients sur 13 sont de signes différents. Il semble donc délicat de considérer l'IIA comme vérifiée dans de telles conditions.

Dans ce cas, il pourrait être alors plus intéressant d'essayer d'affiner l'analyse quant aux populations étudiées. Une solution envisageable serait d'estimer ce même type de modèle mais de façon séparée, soit selon certaines caractéristiques de l'individu, soit selon certaines variables donnant une approximation de son environnement en matière d'offre de transport telles que la taille de la commune d'habitation.

Résumé sur les choix modaux

La première modélisation opposant la voiture aux transports collectifs montre le rôle important joué par le degré de motorisation du ménage. Mais elle indique aussi que la voiture a la préférence des individus en activité, employés ou ouvriers et habitant plutôt en ville.

Pour les choix modaux ne concernant que les transports collectifs, la modélisation renseigne un peu moins bien sur les différents comportements. Ces résultats, éclairés par les statistiques descriptives, font supposer que le car a, par rapport au train, la préférence des ruraux et des familles modestes. Certains résultats descriptifs laissent aussi à penser que le train semble le substitut le plus fréquent à la voiture.

CONCLUSION

Le bilan de cette étude peut être articulé autour de trois axes de réflexion. Du point de vue de la méthodologie statistique, une gamme variée d'outils a été utilisée, c'est-à-dire testées. Certes aucune de ces méthodologies n'a la prétention de constituer une recherche aboutie, toutes demandent à être prolongées.

Ainsi, en analyse non paramétrique (ACM, segmentation ou classification), la mise en évidence de clivages dessinés par les variables explicatives du phénomène de mobilité devrait être prolongée par une construction de typologies plus précises. En analyse paramétrique, il est également possible d'élaborer des modèles plus complets reposant sur des hypothèses d'interactions entre les variables. En effet, cette possibilité de faire intervenir des interactions constitue l'avantage décisif de cet outil par rapport aux tableaux croisés. Les modèles simples proposés apparaissent donc comme des étapes utiles susceptibles d'être perfectionnées. Il est néanmoins certain que ces méthodes statistiques, aussi performantes soient-elles en termes de capacité d'explication par rapport aux analyses non paramétriques (du fait de l'élimination des effets de structure) peuvent donner lieu à des interprétations erronées (existence d'hétérogénéités non observées parmi la population et utilisation de variables qui ne sont que des proxys des véritables effets que l'on souhaite faire intervenir etc.). Aussi, la confrontation aux méthodes non paramétriques est indispensable, au moins à titre de contrôle.

Les points de convergence entre les différentes méthodes sont nombreux : toutes indiquent que la mobilité est déterminée par trois types de facteurs (à nuancer quant aux conditions de sa réalisation) :

- * un facteur d'ordre économique (revenu), car la mobilité de longue distance implique un coût financier,

- * des facteurs d'ordre sociologique et culturel (capital relationnel, culturel accumulé par les catégories socialement privilégiées),

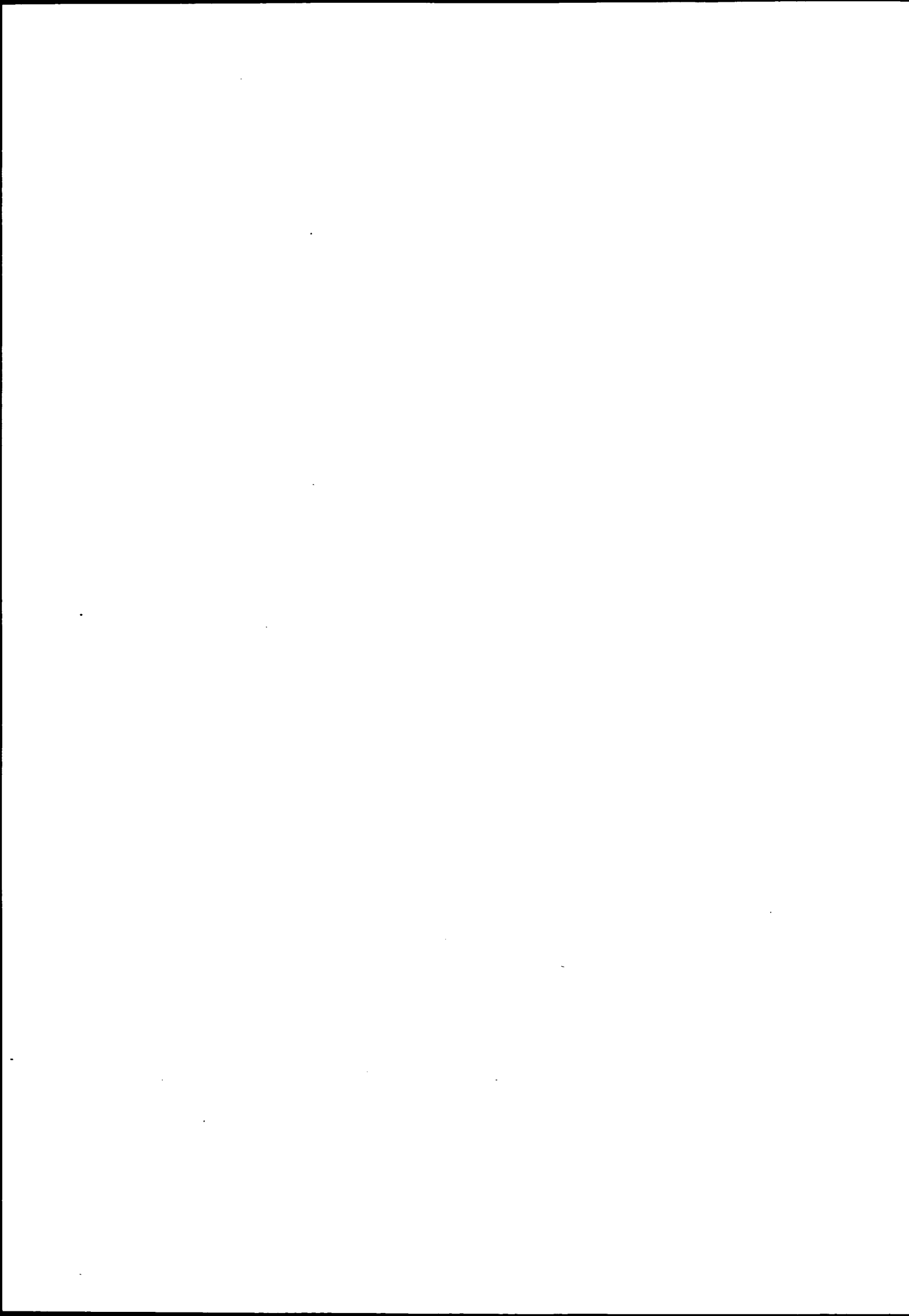
- * des facteurs liés à l'équipement du ménage et à son environnement de vie : l'équipement automobile est décisif dans un contexte où il assure près des trois quarts des déplacements sur longue distance (hors déplacements professionnels).

On pourrait objecter que la mise en oeuvre de techniques aussi lourdes est disproportionnée par rapport à la portée des résultats dégagés, vraisemblablement déjà connus des différents observateurs familiers du champ des transports. Mais ces techniques permettent de les conforter, et dans le cas de l'analyse paramétrique, de quantifier les impacts des facteurs explicatifs.

En dernier lieu, il faut souligner l'intérêt mais aussi les insuffisances de l'enquête qui sont apparues au fil des exploitations statistiques : elle fournit des données originales dans un domaine encore assez peu exploré. Néanmoins, à cet avantage correspond aussi un inconvénient : la difficulté de structurer une information aussi riche est apparue dès le départ, d'où le choix de redimensionner le fichier de déplacements selon des critères que l'on peut toujours discuter. Toujours est-il que l'on s'est ainsi privé d'une bonne partie de l'information collectée.

Le principe d'enregistrement du motif a également nui à son utilisation compte tenu de la prééminence des retours à domicile qui constituent une information assez tautologique. Or, définie plus précisément, cette variable pouvait constituer une information cruciale. De même, peut-on regretter l'absence d'informations relative aux préférences révélées et hiérarchisées des individus qui peuvent prendre en compte le caractère non répétitif des processus de choix par un modèle ordonné au lieu d'un modèle non ordonné dans la modélisation des choix des modes de transport

Il aurait également été intéressant, comme cela a déjà été évoqué, de disposer de véritables variables d'offre en matière de transport sachant qu'elles déterminent la demande de transport autant que les caractéristiques personnelles.



BIBLIOGRAPHIE

Références sur les enquêtes-transports

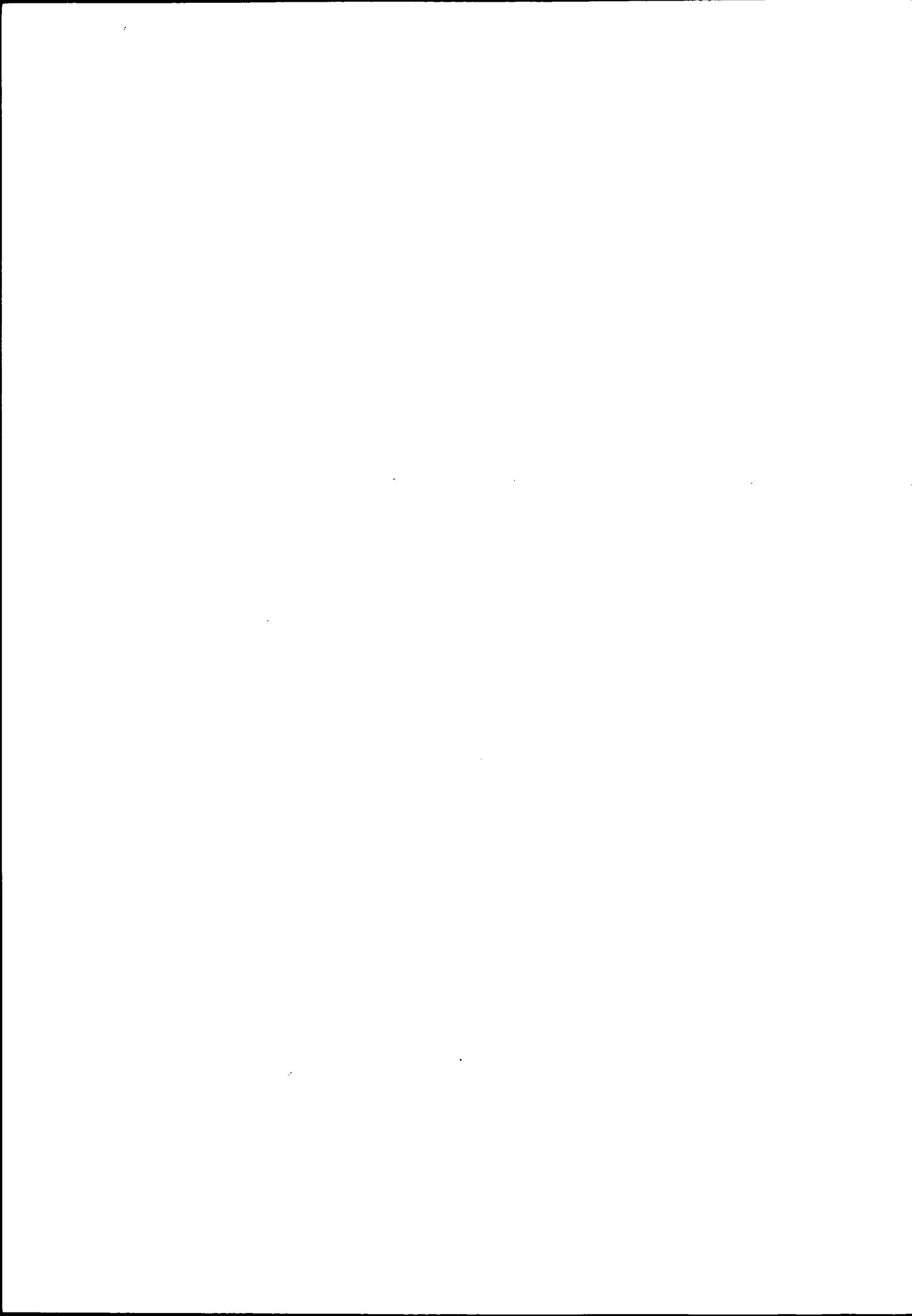
- Questionnaires de l'enquête 1993/94
- Instructions aux enquêteurs, INSEE, Division "Conditions de vie", mars 1993
- OEST, "L'OEST à votre service", plaquette de présentation
- MADRE J.L. et MAFFRE J., "L'enquête transports et communications 1993-1994", in Courrier des statistiques n° 69, mars 1994
- "Présentation générale et définition des concepts" [enquête 82]
- LEFOL J.F., MEY A. et MORELLET O., "Panorama général de la mobilité à longue distance" [enquête 82]

Références sur la segmentation, l'analyse factorielle et la classification

- BABAYOU P., CHATELAIN C. et CORDONNIER F., "Programmation et étude de la méthode CART appliquée à la détection des défaillances d'entreprises", Groupe de travail ENSAE 1993/1994
- CALZADA C., "Transport et méthodes marketing (1) : de la segmentation descriptive à la segmentation prédictive", 1995
- CELEUX G. et NAKACHE J.P., "Analyse discriminante sur variables qualitatives", Polytechnica
- CHEBBOUB K. et GAUTIER J.M., "Analyse comparative des logiciels de segmentation KS, CHAID, SPAD.S et AC2"
- GUEGEN A. et NAKACHE J.P., "Méthode discriminante basée sur la construction d'un arbre de décision binaire", in Revue de statistique appliquée, vol. 36, n° 1, 1988
- TOMASSONE R., DANZART M., DAUDIN J.J. et MASSON J.P., "Discrimination et classement", Ed. Masson
- "La segmentation des ménages utilisée dans le redressement de l'échantillon panel" extrait d'un document sur le panel TGV-Nord

Références sur la modélisation

- BILLOT A. et THISSE J.F., "Modèles de choix individuels discrets - Théorie et applications à la micro-économie", in Revue économique, vol. 46, n°3, mai 1995
- BOUYAUX P., "Une difficulté d'interprétation de l'approche Logit : l'exemple de l'économie des transports", in Economie et prévision, n° 91, mai 1989
- GHAZOUANI S. et GOAIED M., "Analyse micro-économétrique de la demande de transport urbain pour la ville de Tunis", in Economie et prévision, n° 108, février 1993
- GOURIEROUX C., "Econométrie des variables qualitatives", Economica, 1989
- INSEE, "L'économétrie et l'étude des comportements", Note 1, série des documents de travail de la Direction des statistiques démographiques et sociales
- JACQUOT A., "Modèles de la statistique - Econométrie, analyse de la variance", polycopié de cours, ENSAI, 1995/1996



ANNEXES

ANNEXE 1.1 : ACM - Points méthodologiques et résultats dans le cas des motifs professionnels

ANNEXE 1.2 : Point méthodologique sur la segmentation

ANNEXE 1.3 : Résultats détaillés pour la segmentation

ANNEXE 1.4 : Classification

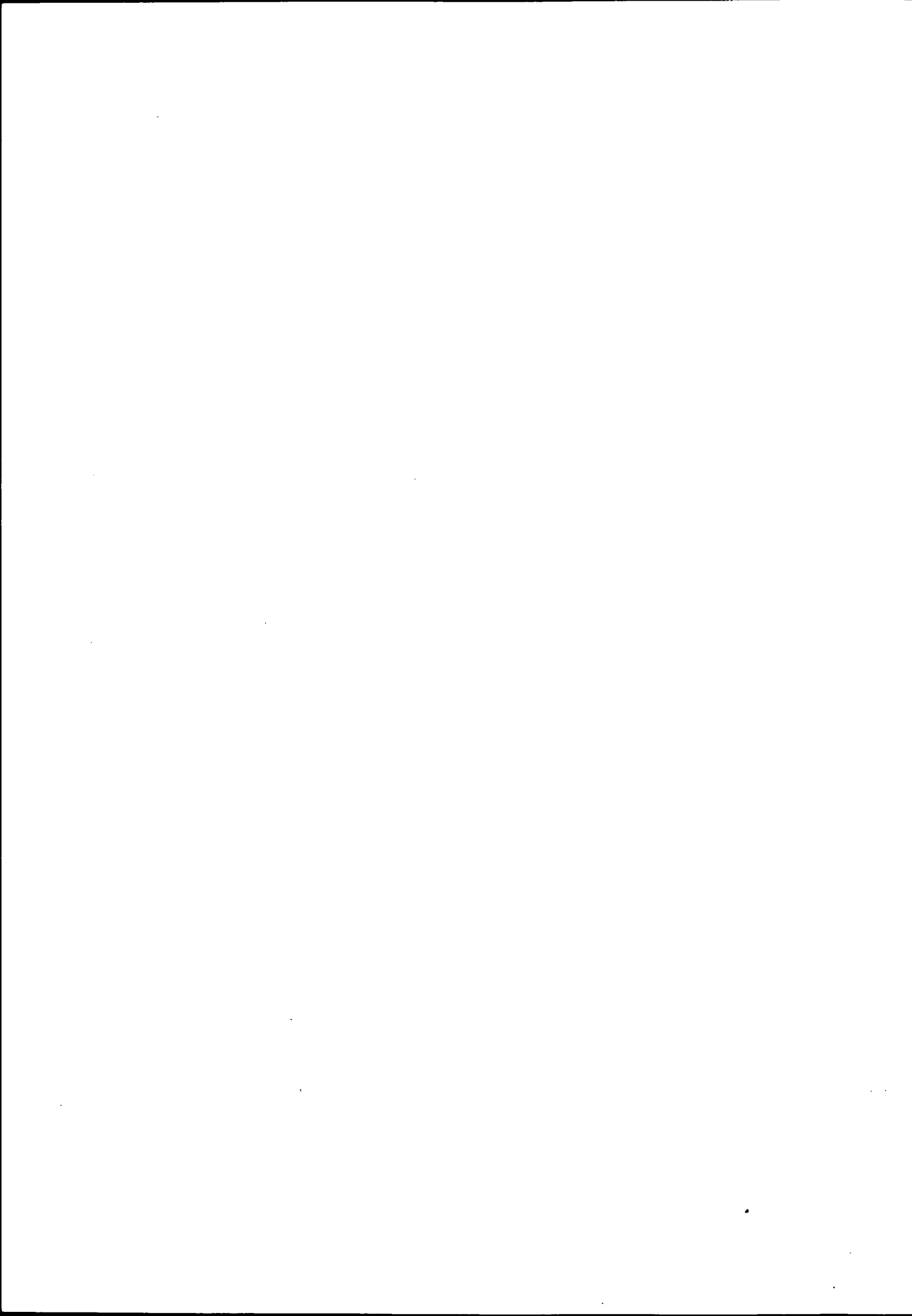
ANNEXE 2.1 : Préalables méthodologiques à propos des estimations économétriques

ANNEXE 2.2 : La méthode d'Heckman pour l'estimation des modèles TOBIT

ANNEXE 2.3 : Description des profils de la population et de leur pratiques de déplacements selon l'intensité de leur mobilité

ANNEXE 2.4 : Tests de validité générale des modèles qualitatifs fournis par la proc logistic de SAS

ANNEXE 2.5 : Modèles complémentaires pour discussion de l'IIA



ANNEXE 1.1 : ACM - Points méthodologiques et résultats dans le cas des motifs professionnels

• Points méthodologiques :

- ▷ La réalisation des différentes ACM a nécessité la construction d'un tableau disjonctif avec création de variables indicatrices (une pour chaque modalité de chaque variable).
- ▷ Par ailleurs, seules les projections du nuage des variables ont été étudiées. En effet, en ACM, on ne projette généralement pas les individus mais seulement les modalités des variables car il existe des relations barycentriques pour chaque axe:

$$\sqrt{\lambda} \times C_k = \frac{1}{M} \times \sum_{m=1}^M d_j^m(k)$$

avec C_k = coordonnées de l'individu k sur l'axe

d_j^m = coordonnée, sur l'axe, de la modalité j de la variable m prise par l'individu k

λ = valeur propre de l'axe = inertie expliquée par l'axe

M = nombre de variables

- ▷ En ce qui concerne les résultats, une modalité est interprétée comme barycentre des individus qui prennent cette modalité. Ainsi, sur le plan formé par deux axes principaux, la proximité des modalités de variables différentes signifie que ces modalités sont prises simultanément par un grand nombre d'individus. Quant à la proximité des modalités différentes d'une même variable, elle signifie que les individus prenant l'une ou l'autre de ces modalités ont un comportement semblable par rapport aux autres variables. De plus, une modalité active présentant une forte contribution à la constitution d'un axe et une valeur élevée pour le Cosinus est considérée comme **expliquant** bien l'axe. Dans le cas d'une faible contribution et d'un cosinus ou d'une qualité de représentation élevées, on a une modalité active qui **illustre** bien l'axe. Pour ce qui est des modalités supplémentaires, on ne retiendra que celles dont la valeur V_{test} est supérieure à 2 en valeur absolue. Enfin le nombre d'axes retenus sera fonction du critère du coude et de la possibilité de leur donner une interprétation particulière aux axes.
- ▷ Aucune ACM dans laquelle les deux groupes de variables sont en variables actives n'a été réalisée. Cette analyse aurait été délicate dans la mesure où les variables des deux groupes sont hétérogènes. Dans la typologie des individus, un des deux groupes peut prédominer sur l'autre et la typologie "mixte" est alors très proche de la typologie induite par ce groupe uniquement. En ce qui concerne la typologie des modalités, en cas de déséquilibre entre les deux groupes, les liaisons internes du groupe dominant seront mises en évidence aux dépens des liaisons internes de l'autre groupe et des liaisons inter-groupes.

• Résultats pour le groupe des individus mobiles se déplaçant pour des motifs professionnels

1 - Variables de signalétique en variables actives

AXE 1 : automobile et famille

Les variables actives ayant le plus contribué à la constitution de l'axe sont l'âge AGE (CTR=8,8), la possession du permis de conduire E1BIS (CTR=9,9), le nombre de voitures dans le ménage G1 (CTR=12), le nombre de permis de conduire dans le ménage NBPERVO (CTR=13,3), l'environnement du domicile Q2 (CTR=8,4), la strate de la commune (CTR=7,9), le type de ménage TYPMEN (CTR=10,3).

Les variables supplémentaires dont les modalités sont significatives d'après la valeur V_{test} sont la distance moyenne du déplacement DISTANM et le mode de transport principal MTPM.

Cet axe oppose les personnes âgées de 26 à 40 ans possédant un permis, ayant 2 voitures ou plus dans leur ménage, avec plus de deux permis de conduire dans leur ménage effectuant des déplacements de moins de 350 km, en voiture principalement, résidant en zone pavillonnaire dans une maison individuelle, vivant en couple avec des enfants (coordonnées négatives); aux personnes de moins de 25 ans n'ayant pas de permis de conduire, pas de voiture dans le ménage, vivant seules (cf. modalités : seul, enf0), dans un immeuble situé dans un quartier à logements collectifs, effectuant des déplacements de plus de 350 km et utilisant comme mode de transport principal les transports collectifs (coordonnées positives).

AXE 2: professionnel et domicile

Les variables actives ayant le plus contribué à cet axe sont l'occupation B1BIS (CTR=12,3), la PCS B4 (CTR=9,7), le département de domicile DEPDOM (CTR=9,9), le nombre de permis de conduire NBPERVO (CTR=9,3), la taille de la commune STRATE (CTR=9,2), le type de ménage TYPMEN (CTR=14,1).

Les variables supplémentaires ayant les modalités les plus significatives sont la distance totale DCTOT, la distance d'un déplacement moyen DISTANM, le nombre de voyages.

Cet axe oppose les personnes qui sont plutôt des actifs, cadres, résidant en Ile de France ou plus précisément à Paris, avec un ou deux permis de conduire dans le ménage, et vivant davantage seul ou en famille monoparentale - se déplaçant sur de longues distances et voyageant plus souvent - (coordonnées négatives); aux personnes qui sont des employés ou ouvriers plutôt au chômage, résidant en province voire dans des communes rurales. Ces personnes n'ont pas de permis de conduire, vivent en couple dans une maison en quartier pavillonnaire, effectuent des déplacements sur de petites distances et voyagent peu.

2 - Variables de déplacement en variables actives

AXE 1 : mobilité en termes de distance, durée et intensité

Les variables actives ayant le plus contribué à la constitution de l'axe sont la distance DISTANM (CTR=39,7), la durée DUREM (CTR=35,1), le nombre de déplacements NBDEP (CTR=21,9). Les modalités de ces variables sont classées par ordre décroissant.

Les variables supplémentaires montrent que le comportement de plus grande mobilité (coordonnées négatives) est surtout adopté par des personnes résidant en Ile de France, disposant d'une réduction tarifaire et voyageant beaucoup - par opposition à celles habitant en province, ne bénéficiant pas d'une réduction, voyageant peu et ayant adopté un comportement de faible mobilité en terme de distance, durée et intensité.

AXE 2 : moyen de transport principal

Les variables actives ayant le plus contribué à la constitution de l'axe sont le mode de transport MTPM (CTR=27,5), le nombre de déplacements NBDEP (CTR=42). Le nombre de déplacements était déjà pris en compte dans l'axe 1, le principal apport de l'axe 2 est la mise en valeur du moyen de transport principal. Un comportement de forte intensité de mobilité (dep3) est surtout lié à l'utilisation de la voiture (mtvoi).

Les variables supplémentaires ont presque toutes leurs modalités non significatives. Seule la variable L02 (nombre de voyages) présente des modalités vraiment significatives. Ceci permet seulement de vérifier que plus le nombre de voyages est élevé plus le nombre de déplacements est grand (un voyage est constitué de plusieurs déplacements)!

ANNEXE 1.2 : Point méthodologique sur la segmentation

Les méthodes de discrimination par arbre binaire dites segmentation sont des méthodes de discrimination non linéaires sur des variables qualitatives. Elles fournissent une règle de décision et permettent une étude descriptive de la population. La réalisation d'une discrimination par arbre binaire ou encore segmentation paraît tout à fait adaptée à notre enquête.

E étant l'ensemble des états possibles des J variables explicatives, un arbre A est construit par divisions successives de l'ensemble E en sous-ensembles appelés noeuds de l'arbre. A l'aide d'une variable, on divise un noeud en deux noeuds descendants gauche et droit. Au sommet de l'arbre, se trouve le noeud racine qui contient l'ensemble E tout entier. Un noeud est dit intermédiaire s'il est divisé, et terminal sinon. Ces noeuds sont des ensembles d'états possibles et les noeuds terminaux sont affectés à un des K groupes (chaque partie de E - partitionné en K parties - est affectée à un groupe).

Pour construire un arbre de discrimination, il nous faudra :

- pour chaque noeud, établir l'ensemble des divisions admissibles et définir un critère permettant de sélectionner la meilleure : on retiendra à chaque pas la division qui rend les noeuds descendants de plus en plus purs c'est-à-dire tendant à contenir des observations d'un seul groupe.
- définir une règle permettant de déclarer un noeud comme terminal ou intermédiaire :
 - 1 - On construit un arbre intentionnellement trop grand Amax, puis on l'élague pour obtenir une séquence de sous-arbres emboîtés.
 - 2 - On estime les taux de mal classés des arbres de la séquence pour sélectionner un arbre satisfaisant.
- affecter chaque noeud terminal à l'un des groupes.

-La méthodologie de la segmentation

La segmentation se déroule en deux grandes étapes, la construction de l'arbre de sélection, puis son élagage.

-Construction de l'arbre de sélection maximal.

L'arbre de sélection va être construit à partir de variables binaires. A chaque noeud, chaque observation sera affectée soit au noeud droit, soit au noeud gauche descendants selon qu'un critère binaire est satisfait ou non. Il faut auparavant avoir choisi la variable permettant la meilleure discrimination de la variable expliquée en fonction d'un critère non paramétrique, l'impureté. L'explication donnée ci-dessous pour deux groupes s'étend facilement à n groupes.

Soit n le nombre d'observations de l'échantillon de base

n_1 le nombre d'individus affectés au premier groupe

n_2 le nombre d'individus affectés au deuxième groupe

$n_i(t)$ le nombre d'individus du groupe i dans le noeud t

π_i la probabilité a priori qu'un individu appartienne au groupe i (cette probabilité sera un paramètre)

$p(t)$ la probabilité qu'un individu parvienne au noeud t

$p(i,t)$ la probabilité qu'un individu appartienne au noeud t et au groupe i

$p(t/i)$ la probabilité qu'un individu appartienne au noeud t sachant qu'il fait partie du groupe i

alors:

$$p(i, t) = \pi_i p(t / i) = \pi_i \frac{n_i(t)}{n_i}$$

$$p(i / t) = \frac{p(i, t)}{p(t)} = \pi_i \frac{p(t / i)}{p(t)}$$

$$p(t) = \pi_1 p(t / 1) + \pi_2 p(t / 2)$$

$R_i(t)$ est le risque de resubstitution d'un individu du groupe i dans un groupe j , risque dont le coût est c_{ij}

$$R_i(t) = c_{ij} p(j, t)$$

L'impureté d'un noeud t , $I(t)$ est l'espérance du risque de resubstitution de ce noeud.

$$I(t) = R_1(t) p(1 / t) + R_2(t) p(2 / t)$$

Cette fonction est d'autant plus faible que le noeud est homogène (l'impureté d'un noeud pur est nul). Il existe cependant d'autres fonctions d'impureté.

On choisit la variable (et le seuil qui a été utilisé pour rendre cette variable binaire) qui maximise la perte d'impureté entre deux niveaux de l'arbre:

$$\Delta I(t) = I(t) - I_g(t) - I_d(t)$$

$I_g(t)$ et $I_d(t)$ l'impureté des deux noeuds gauche et droit.

Le critère d'arrêt

On s'arrête lorsque un noeud a atteint un degré de pureté suffisant, c'est-à-dire en général lorsqu'il reste moins de cinq individus d'un des deux groupes.

Affectation des noeuds aux groupes

Il existe plusieurs règles d'affectation. On présente ici le critère de minimisation du coût d'erreur de classement $r(t)$ du noeud.

$$r(t) = \min_j c_{ij} p(j, t)$$

$$p(j, t) = \pi_j \frac{n_j(t)}{n_j}$$

On obtient donc un arbre de sélection maximale.

-L'élagage de l'arbre de sélection

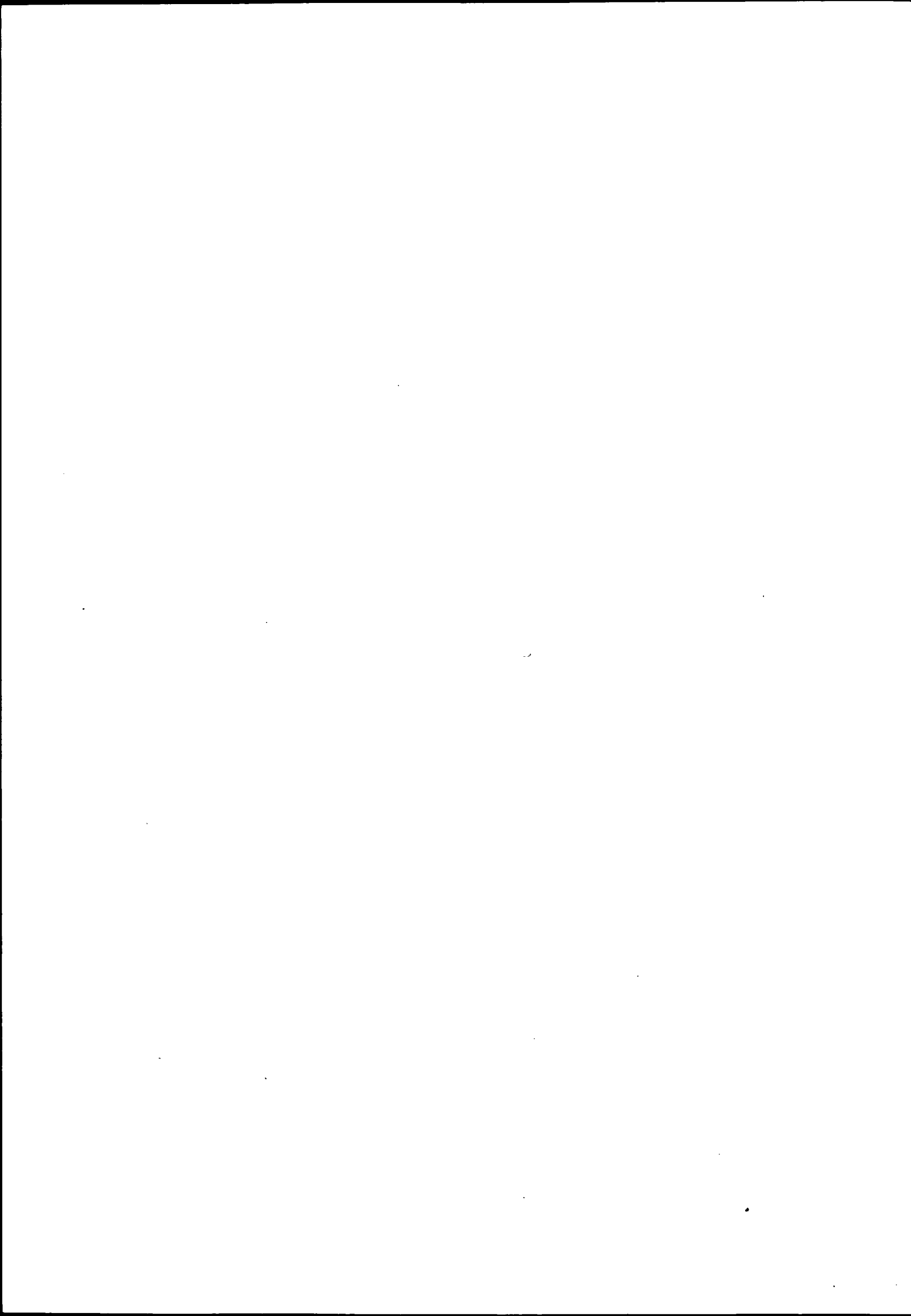
L'arbre de sélection obtenu est dit maximal puisqu'un noeud est séparé tant qu'il n'est pas assez pur. Mais on rencontre un problème dans cette procédure:

Les effectifs des noeuds terminaux, s'ils sont trop faibles, entraînent un manque de robustesse dans l'interprétation de la règle de décision. Pour y remédier, on procède à un élagage de l'arbre. On supprime un certain nombre de branches tout en conservant une bonne qualité de l'information. Celle-ci se mesure par le taux d'erreur de classement des observations de l'échantillon. De plus, on veut obtenir un nombre "raisonnable" de noeuds terminaux, ce qui permettra une interprétation plus facile de l'arbre de discrimination final.

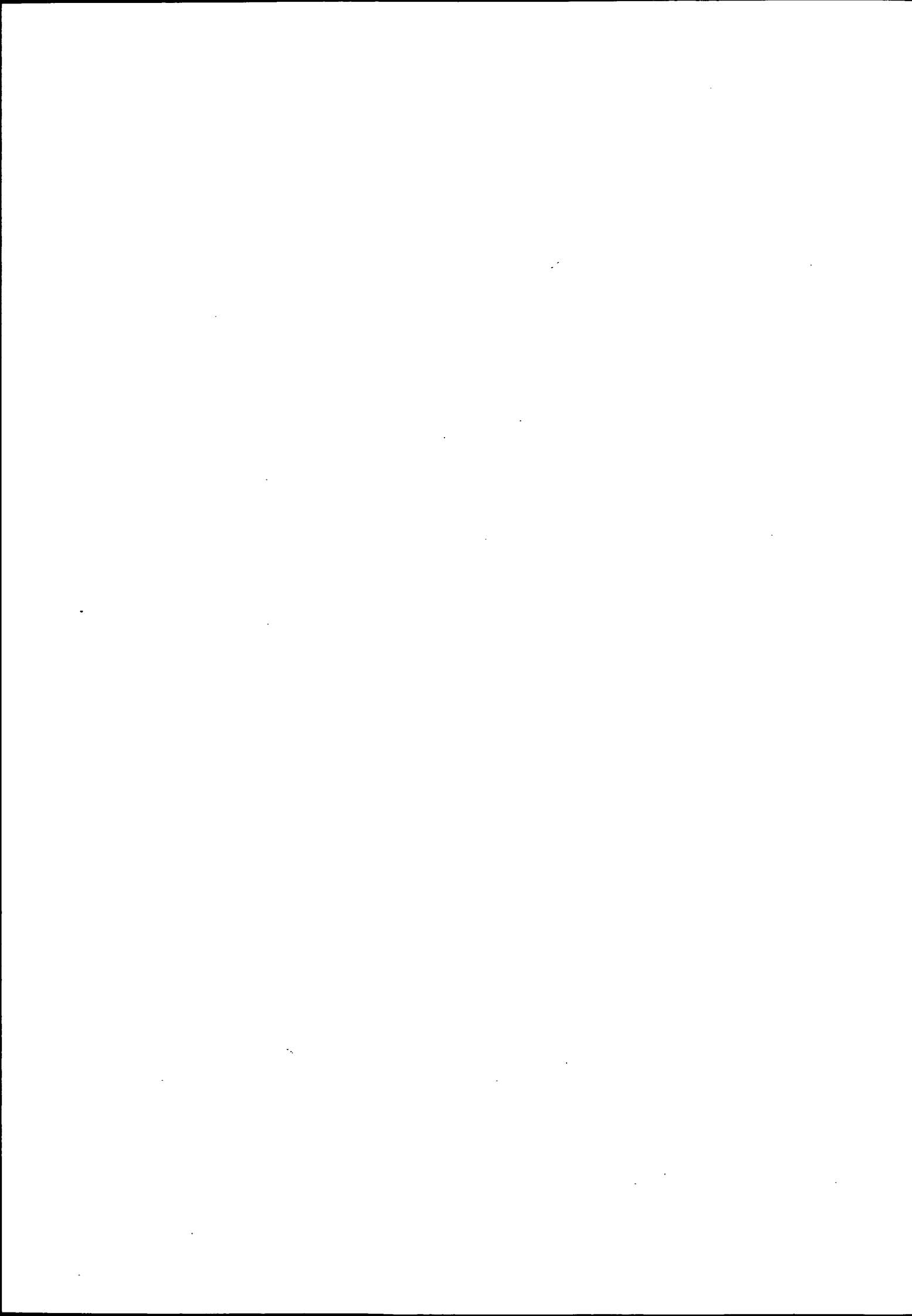
Sur l'échantillon de base, le taux d'erreurs de classement diminue avec l'accroissement du nombre de noeuds terminaux puisque l'on construit des dichotomies de manière à obtenir des noeuds de plus en plus purs.

Appliqué sur l'échantillon, lorsqu'on représente les estimations des risques associés aux arbres T_h par un graphique ayant en abscisse le nombre de leurs noeuds terminaux, dans la plupart des cas, on constate une décroissance rapide du taux d'erreur de classement quand le nombre de noeuds terminaux augmente, puis un plateau de risque minimal dans lequel le risque est assez stable, puis une lente remontée. Le risque minimal se produit pour un arbre de la zone "plateau", zone dans laquelle le nombre de noeuds terminaux peut varier considérablement, alors que le taux d'erreur reste pratiquement stable. On retient l'arbre pour lequel le nombre de noeuds terminaux est minimal.

Le logiciel CART fournit un algorithme permettant de déterminer rapidement l'arbre élagué minimisant le taux d'erreur de classement.



ANNEXE 1.3 : Résultats détaillés pour la segmentation



segmentation sur tous les individus
ndep0=0 ndep1=1-3, ndep2>3

OBS	_FROM_	_TO_	SPLIT_	SIZE_	ERRORS_	PCNODE_	VALUES_	NODE_	TXERS
1	1	5	NDEP	52718099.31	28309077.36	46% 23% 30%	0 1 2	1	0.53699
2	1	4	NDEP	52718099.31	28309077.36	46% 23% 30%	0 1 2	1	0.53699
3	1	3	NDEP	52718099.31	28309077.36	46% 23% 30%	0 1 2	1	0.53699
4	1	2	NDEP	52718099.31	28309077.36	46% 23% 30%	0 1 2	1	0.53699
5	2	7	TYPMENS	11869095.00	6676935.40	44% 23% 33%	COUP	2	0.56255
6	2	6	TYPMENS	11869095.00	6676935.40	44% 23% 33%	COUP	2	0.56255
7	6	.	STRATES	6471071.46	3397263.82	48% 22% 31%	com0 com1	6	0.52499
8	7	.	STRATES	5398023.54	3279671.58	39% 25% 36%	com2 com3	7	0.60757
9	3	11	TYPMENS	27743530.76	15716198.14	43% 25% 32%	COUP+	3	0.56648
10	3	10	TYPMENS	27743530.76	15716198.14	43% 25% 32%	COUP+	3	0.56648
11	3	9	TYPMENS	27743530.76	15716198.14	43% 25% 32%	COUP+	3	0.56648
12	3	8	TYPMENS	27743530.76	15716198.14	43% 25% 32%	COUP+	3	0.56648
13	8	.	STRATES	7481149.04	4102172.30	45% 24% 31%	com0	8	0.54833
14	9	13	STRATES	8557747.42	4516818.19	47% 23% 29%	com1	9	0.52780
15	9	12	STRATES	8557747.42	4516818.19	47% 23% 29%	com1	9	0.52780
16	12	.	BIBISS	4386126.12	2511385.79	43% 23% 34%	actif chom	12	0.57257
17	13	.	BIBISS	4171621.30	2005432.41	52% 24% 24%	inact jeune	13	0.48073
18	10	.	STRATES	7396276.30	4265340.96	42% 25% 32%	com2	10	0.57669
19	11	.	STRATES	4308358.01	2689406.98	34% 28% 38%	com3	11	0.62423
20	4	.	TYPMENS	6796556.03	2980020.94	56% 22% 22%	MEND MONOP	4	0.43846
21	5	.	TYPMENS	6308917.51	2935922.88	53% 19% 27%	SEUL	5	0.46536

note

from: noeud père

to: noeud fils

split: variable ayant conduit à la discriminationsur le noeud père

size: taille de la population du noeud père

errors: nombre de personnes n'appartenant pas au groupe majoritaire

pcnode:répartition de la population du noeud entre les groupes

values: modalité de la variable explicative présente dans le noeud

txers: taux de mal classés dans le noeud.

segmentation sur 'non mobiles' et 'mobiles'

OBS	_FROM_	_TO_	SPLIT_	SIZE_	ERRORS_	PCNODE_	VALUES_	NODE_	TXERS
1	1	5	NDEP	52718099.31	24409021.95	46% 54%	0 1	1	0.46301
2	1	4	NDEP	52718099.31	24409021.95	46% 54%	0 1	1	0.46301
3	1	3	NDEP	52718099.31	24409021.95	46% 54%	0 1	1	0.46301
4	1	2	NDEP	52718099.31	24409021.95	46% 54%	0 1	1	0.46301
5	2	7	TYPMENS	11869095.00	5192159.60	44% 56%	COUP	2	0.43745
6	2	6	TYPMENS	11869095.00	5192159.60	44% 56%	COUP	2	0.43745
7	6	.	STRATES	6471071.46	3073807.64	48% 52%	com0 com1	6	0.47501
8	7	.	STRATES	5398023.54	2118351.96	39% 61%	com2 com3	7	0.39243
9	3	11	TYPMENS	27743530.76	12027332.63	43% 57%	COUP+	3	0.43352
10	3	10	TYPMENS	27743530.76	12027332.63	43% 57%	COUP+	3	0.43352
11	3	9	TYPMENS	27743530.76	12027332.63	43% 57%	COUP+	3	0.43352
12	3	8	TYPMENS	27743530.76	12027332.63	43% 57%	COUP+	3	0.43352
13	8	.	STRATES	7481149.04	3378976.74	45% 55%	com0	8	0.45167
14	9	13	STRATES	8557747.42	4040929.22	47% 53%	com1	9	0.47220
15	9	12	STRATES	8557747.42	4040929.22	47% 53%	com1	9	0.47220
16	12	.	B1BISS	4386126.12	1874740.33	43% 57%	actif chom	12	0.42743
17	13	.	B1BISS	4171621.30	2005432.41	52% 48%	inact jeune	13	0.48073
18	10	.	STRATES	7396276.30	3130935.34	42% 58%	com2	10	0.42331
19	11	.	STRATES	4308358.01	1476491.32	34% 66%	com3	11	0.34270
20	4	.	TYPMENS	6796556.03	2980020.94	56% 44%	MEND MONOP	4	0.43846
21	5	.	TYPMENS	6308917.51	2935922.88	53% 47%	SEUL	5	0.46536

note

from: noeud père

to: noeud fils

split: variable ayant conduit à la discriminationsur le noeud père

size: taille de la population du noeud père

errors: nombre de personnes n'appartenant pas au groupe majoritaire

pcnode: répartition de la population du noeud entre les groupes

values: modalité de la variable explicative présente dans le noeud

txers: taux de mal classés dans le noeud.

ANNEXE 1.4 : Classification

• Point méthodologique :

⇒ Les méthodes de classification ont pour but de **regrouper les individus en un nombre restreint de classes homogènes** : la classification ascendante hiérarchique produit des suites de partitions en classes de plus en plus vastes. Pour un ensemble de n individus, le principe consiste à construire n puis $n-1$ puis $n-2$... **classes emboîtées les unes dans les autres**: la partition en k classes est obtenue en regroupant les deux classes les plus proches. Nous avons retenu comme critère de regroupement des deux classes celui de l'**inertie interclasse**. Nous avons en effet choisi d'appliquer la méthode de **Ward**. Dans une partition donnée, on fusionne les deux classes pour lesquelles la perte d'inertie interclasse est la plus faible. La mesure de dissimilarité entre les individus et les classes se fait à l'aide du χ_2 car les variables utilisées sont qualitatives.

⇒ Généralement on retient un nombre de classes tel que la perte d'inertie ne dépasse 95%.

⇒ Le nombre d'individus de l'échantillon étant supérieur à 14000, nous avons dû effectuer un tirage aléatoire (Sondage Aléatoire Simple) de 2000 individus de façon à permettre aux macros CAHQVAL, PARTQUAL et DESQUAL de fonctionner.

• Résultats :

Tableau 1 : Décomposition de l'inertie totale

	CLASS5	Inertie	En 0/00 de l'inertie totale	Effectif absolu	Effectif pondéré (en 0/00)
Inertie totale		3.3518	1000	1976	1000
Inertie inter		0.6649	198	.	.
	1	0.0879	26	323	137
	2	0.1745	52	696	400
	3	0.1537	46	698	356
	4	0.1502	45	200	80
	5	0.0986	29	59	26
Inertie intra		2.6870	802	.	.
	1	0.4631	138	323	137
	2	0.8655	258	696	400
	3	1.0002	298	698	356
	4	0.1946	58	200	80
	5	0.1636	49	59	26

Tableau 2 : Décomposition signée du RHO2

Tous les chiffres sont en millièmes, exceptés l'effectif absolu et le RHO2

Classe	Effectif absolu	Effectif pondéré	RHO2
1	323	137	0.63905
2	696	400	0.43634
3	698	356	0.43156
4	200	80	1.87598
5	59	26	3.74782

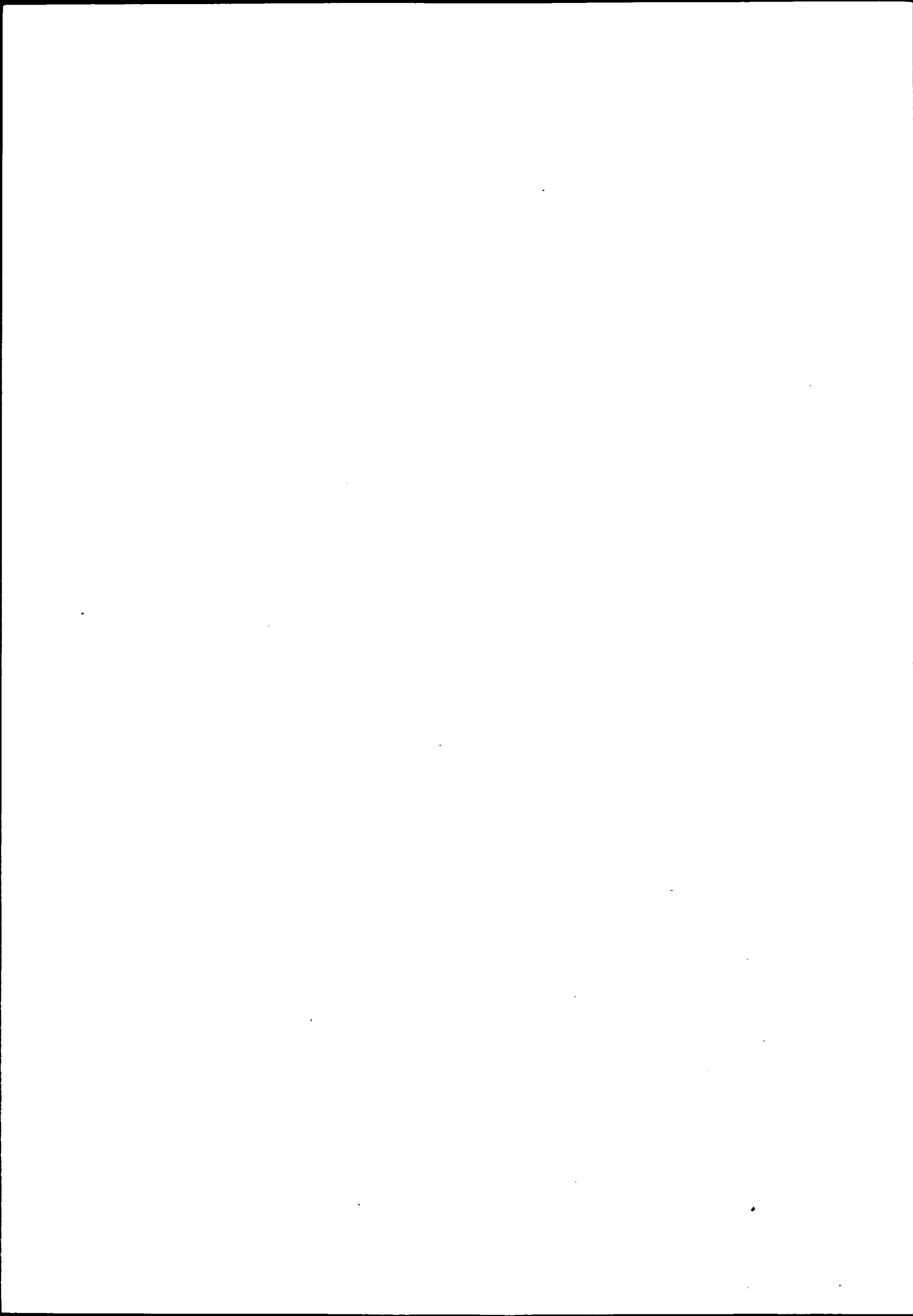
Tableau 3 : CHI2 des tableaux de contingence croisant la variable de classe et les variables analysées (triés par niveau de significativité)

Variable	Nbre de modalités	Effectif pondéré du tableau	Nbre de cases d'effectif <5	Nbre de cases d'effectif <5 (en %)	CHI2	Degrés de liberté	PROBA
mtp	3	7672564	5	33	7498192	8	0.0000
nbvoit	3	7672564	1	7	5544988	8	0.0000
occup	3	7672564	0	0	450824	8	0.0000
taille	4	7672564	8	40	1.294E7	12	0.0000
nbpervo	4	7672564	4	20	9591143	12	0.0000
nbdep	4	7672564	6	30	7363933	12	0.0000
revtr	4	7672564	0	0	1979306	12	0.0000
motif	5	7672564	8	32	9572604	16	0.0000
typmen	5	7672564	1	4	2678604	16	0.0000
etud	5	7672564	0	0	1015413	16	0.0000
strate	5	7672564	0	0	297361	16	0.0000
pcs	6	5273474	0	0	741267	20	0.0000

Tableau 4 : Caractérisation de chaque classe de la partition par les modalités des variables analysées (triées par niveau de significativité)

CLASSE	Modalités	Modalité	Variable	Effectif Pondéré	Fréquence dans la classe(%)	Fréquence dans la pop. (%)	PROBA	Val.Test
1	sur-représentées	NBPER1	nbpervo	867279	82.3	26.9	0.0000	1223.32
		TAI2	taille	531113	50.4	17.5	0.0000	96.61
		MOTIFM1	motif	855862	81.2	46.3	0.0000	795.48
		VOIT1	nbvoit	854467	81.0	47.0	0.0000	777.64
		NDEP1	nbdep	471660	44.7	21.1	0.0000	569.56
		MTP2	mtp	698200	66.2	40.6	0.0000	566.58
		MTP1	mtp	356203	33.8	13.7	0.0000	544.58
		TYP4	typmen	201458	19.1	5.6	0.0000	501.19
		TYP1	typmen	296304	28.1	11.4	0.0000	487.86
		REV2	revtr	479067	45.4	26.8	0.0000	434.77
ETUD5	etud	310410	29.4	15.0	0.0000	395.08		
1	sous-représentées	NBPER2	nbpervo	168780	16.0	51.6	0.0000	-823.58
		VOIT2	nbvoit	33796	3.2	38.9	0.0000	-998.07
		MOTIFM0	motif	698	0.1	45.7	0.0000	-1354.24
		MTP0	mtp	0	0.0	45.7	0.0000	-1386.41
		NDEP0	nbdep	0	0.0	45.7	0.0000	-1386.41
		TAI0	taille	0	0.0	45.8	0.0000	-1388.99
2	sur-représentées	MTP0	mtp	3064671	100.0	45.7	0.0000	2581.01
		NDEP0	nbdep	3064671	100.0	45.7	0.0000	2581.01
		MOTIFM0	motif	3064671	100.0	45.7	0.0000	2579.72
		TAI0	taille	3064671	100.0	45.8	0.0000	2576.98
		VOIT1	nbvoit	1796767	58.6	47.0	0.0000	527.23
		NBPER1	nbpervo	1069577	34.9	26.9	0.0000	400.85
		PCS6	pcs	793970	36.8	27.6	0.0000	391.07
		ETUD1	etud	1259181	41.1	34.7	0.0000	303.58
2	sous-représentées	TAI2	taille	0	0.0	17.5	0.0000	-1464.51
		NDEP3	nbdep	0	0.0	19.0	0.0000	-1523.64
		NDEP1	nbdep	0	0.0	21.1	0.0000	-1606.93
		TAI3	taille	0	0.0	34.8	0.0000	-2063.30
		MTP2	mtp	0	0.0	40.6	0.0000	-2230.19
		MOTIFM1	motif	0	0.0	46.3	0.0000	-2380.41
3	sur-représentées	MOTIFM1	motif	2414388	88.1	46.3	0.0000	1747.82
		MTP2	mtp	2226165	81.2	40.6	0.0000	1678.22
		TAI3	taille	1971633	72.0	34.8	0.0000	1559.15
		VOIT2	nbvoit	1845282	67.3	38.9	0.0000	1183.10
		NBPER2	nbpervo	2147318	78.4	51.6	0.0000	1121.95

		NDEP3	nbdep	970716	35.4	19.0	0.0000	831.69
		NDEP2	nbdep	783795	28.6	14.3	0.0000	810.53
		REV3	revtr	1223179	44.6	27.4	0.0000	774.47
3	sous-représentées	TYP1	typmen	0	0.0	11.4	0.0000	-1116.25
		NBPER1	nbpervo	39803	1.5	26.9	0.0000	-1443.38
		TAI0	taille	11052	0.4	45.8	0.0000	-2142.35
		MOTIFM0	motif	3480	0.1	45.7	0.0000	-2184.83
		MTP0	mtp	0	0.0	45.7	0.0000	-2236.08
		NDEP0	nbdep	0	0.0	45.7	0.0000	-2236.08
4	sur-représentées	VOIT0	nbvoit	613011	99.8	14.2	0.0000	1433.89
		NBPER0	nbpervo	573146	93.3	7.8	0.0000	1427.99
		REV1	revtr	394540	64.2	16.6	0.0000	810.69
		TYP1	typmen	326599	53.2	11.4	0.0000	766.26
		ETUD1	etud	384047	62.5	34.7	0.0000	456.35
4	sous-représentées	REV3	revtr	8690	1.4	27.4	0.0000	-662.79
		VOIT2	nbvoit	0	0.0	38.9	0.0000	-975.58
		VOIT1	nbvoit	1118	0.2	47.0	0.0000	-1028.02
		NBPER2	nbpervo	0	0.0	51.6	0.0000	-1124.34
5	sur-représentées	TAI1	taille	144490	72.4	1.9	0.0000	750.29
		MOTIFM4	motif	62489	31.3	0.8	0.0000	493.42
		NDEP3	nbdep	121469	60.9	19.0	0.0000	390.62
		MTP1	mtp	99465	49.9	13.7	0.0000	358.33
		ETUD5	etud	101848	51.1	15.0	0.0000	352.65
		OCCU1	occup	140130	70.3	37.9	0.0000	293.96
5	sous-représentées	PCS6	pcs	9220	5.6	27.6	0.0000	-253.34
		ETUD3	etud	8901	4.5	23.9	0.0000	-264.08
		MTP0	mtp	0	0.0	45.7	0.0000	-602.12
		NDEP0	nbdep	0	0.0	45.7	0.0000	-602.12
		MOTIFM0	motif	0	0.0	45.7	0.0000	-602.48
		TAI0	taille	0	0.0	45.8	0.0000	-603.24



ANNEXE 2.1 : Préalables méthodologiques à propos des estimations économétriques

Pondération des modèles économétriques :

Une note de l'INSEE "L'économétrie et l'étude des comportements" précise les termes du débat : "les avis divergent sur l'opportunité de pondérer. Traditionnellement, les économistes recommandent de ne pas pondérer car il s'agit de l'étude des comportements... Cette recommandation est justifiée par le fait qu'en règle générale, il n'y aurait pas de relation entre la façon dont les poids sont déterminés et le phénomène analysé..." Or, l'enquête "Transports - mobilité à longue distance" a organisé une pondération qui accorde un poids plus élevé aux personnes qui se sont déplacées. L'on se trouve donc précisément dans le cas de figure où la prise en compte de la pondération s'impose. De plus, ce choix correspond davantage à l'orientation praticienne à laquelle notre démarche aspire. Cette note évoque ainsi : "les praticiens qui utilisent les résultats du modèle pour estimer les probabilités d'émergence du phénomène étudié au sein des diverses strates de population souhaitent retomber au plus près sur les probabilités effectivement constatées dans la strate. C'est pourquoi ils préfèrent pondérer."

Se pose alors le problème du gain de précision artificiel dans les estimations, lorsque les pondérations sont appliquées telles quelles, liées à la reproduction des individus de l'échantillon. Il faut alors utiliser des pondérations normalisées, de moyenne 1, en divisant la variable de pondération par sa moyenne calculée sur l'ensemble du fichier. C'est donc cette solution qui a été appliquée à l'ensemble des traitements économétriques. En revanche, les statistiques de distribution des caractéristiques demeurent calculées à partir des pondérations individuelles initiales.

Modèles additifs ou multiplicatifs :

Spécification additive/spécification multiplicative :

L'ensemble des spécifications testées posent des liaisons additives entre variables. L'unicité du coefficient b_j attaché à la variable explicative signifie que l'on suppose que celle-ci agit de la même façon quelles que soient les configurations prises par les variables X_k , $k \neq j$. Une spécification complète nécessiterait d'introduire des interactions entre variables, c'est-à-dire d'ajouter de nouvelles variables qui soient des composées des variables de la spécification additive (par exemple, le comportement selon les tranches d'âge n'étant pas obligatoirement le même selon le sexe, une variable sexe x tranche d'âge serait intégrée). Néanmoins, cette approche n'a pas été retenue, en raison de la multiplicité des variables créées en jeu, sans qu'il ne soit aisé de sélectionner les plus pertinentes d'entre elles, sans modèle théorique précis. Si le phénomène est différent d'une "strate" à l'autre, la solution alternative consiste à réaliser des régressions indépendantes, une pour chacune d'elles. C'est dans cette optique qu'il faut comprendre la décision prise de dissocier les motifs d'ordre privé et professionnel dans les modèles, ce critère paraissant particulièrement discriminant dans les problématiques étudiées.

ANNEXE 2.2 : La méthode d'Heckman pour l'estimation des modèles TOBIT

En reprenant les notations du rapport (paragraphe 1.2.1 de la 2ème partie), le modèle TOBIT s'écrit :

$$y_i^* = x_i' b + u_i \text{ pour l'individu } i$$

où y_i^* = niveau d'utilité retirée du nombre de déplacements, inobservable : y_i ,
le nombre de déplacements, est donc la manifestation visible d'une variable latente y_i^* non observable, qui est continue et non tronquée. L'événement "mobilité" se réalise, lorsque la variable latente dépasse le seuil d'utilité nulle.
L'on observe donc :

$$y_i = y_i^* \text{ si } y_i^* \geq 0$$

$$= 0 \text{ si } y_i^* < 0,$$

avec x_i' = vecteur du signalétique de l'individu i ,
 b = vecteur des coefficients à estimer,
 u_i = résidu ; les u_i sont iid et suivent une loi $N(0, \sigma^2)$

On n'a plus $E(y_i/x_i) = x_i' b$ (espérance de y_i sachant x_i)
mais $E(y_i/x_i) = x_i' b \Phi(x_i' b / \sigma) + \sigma \varphi(x_i' b / \sigma)$ (équation 1)
et $E(y_i > 0/x_i) = x_i' b + \sigma [\varphi(x_i' b / \sigma)] / [\Phi(x_i' b / \sigma)]$ (équation 2)
où Φ = fonction de répartition d'une loi normale centrée réduite,
 φ = fonction de densité d'une loi normale centrée réduite,
 σ = écart type-des résidus.

La méthode d'Heckman articulante un modèle PROBIT à un modèle de régression linéaire par les MCO permet de résoudre le problème de perte de convergence de l'estimateur par les MCO :

soit $a = b/\sigma$; supposons qu'on dispose d'un estimateur \hat{a} convergent de a . Alors en estimant l'équation 2 sur les données complètes, on obtient un estimateur convergent de b et de σ .

Or, soit $z_i = 1_{\{y_i^* > 0\}}$; z_i suit un modèle PROBIT, (appelé modèle PROBIT associé au modèle TOBIT) dont l'estimation fournit une estimation convergente de a .

Après l'estimation du modèle PROBIT associé au modèle TOBIT (étape 1), on estime par les MCO l'équation 2, sur les données complètes (étape 2).

La première étape de la méthode d'Heckman utilise donc une approche dichotomique : dans le cas présent, il s'agit de modéliser la probabilité d'avoir effectué un nombre non nul de déplacements.

D'où $z_i = 0$ s'il n'y a pas de mobilité

et $z_i = 1$ s'il y a mobilité.

Le modèle correspondant peut s'écrire : $z_i^* = x_i'(b/\sigma) + v_i$, avec v_i iid et suivant $N(0, \sigma^2)$ et $z_i^* = y_i^*/\sigma$.

soit encore $z_i^* = x_i' a + v_i$

$P_i = P(z_i = 1) = P(z_i^* > 0) = P(x_i'(b/\sigma) > -v_i) = F((x_i'(b/\sigma))$, si on note F la fonction de répartition de la loi normale de $-v_i$.

L'estimation du rapport $[\varphi(x_i' a)] / [\Phi(x_i' a)]$ s'effectue en calculant $[\varphi(x_i' \hat{a})] / [\Phi(x_i' \hat{a})]$ connu sous le terme de ratio de Mill. Ce ratio est alors ajouté aux variables explicatives de l'intensité de la mobilité pour former le modèle de l'équation 2.

ANNEXE 2.3 : Description des profils de la population et de leur pratiques de déplacements selon l'intensité de leur mobilité

Profil de l'échantillon des individus mobiles selon le nombre de déplacements (professionnels exclus)				
nombre de déplacements	2 ou 3	4 et plus	2 à 4	5 et plus
% femmes	52	48,1	51,5	46,8
% <26 ans / % >60 ans	36,9 / 17,1	28,4 / 13,0	35,4 / 15,8	26,2 / 13,0
% actifs / % chômeurs	39,2 / 5,8	50,8 / 4,6	41,5 / 5,3	53,3 / 4,7
% cadres / % employés+ouvriers	28,1 / 60,3	45,8 / 44,9	32,6 / 56,0	48,5 / 43,1
% niveau d'études ≥ baccalauréat	28,3	44,9	32,0	47,9
% personnes seules / % couples avec enfants	9,8 / 55,7	10,7 / 55,3	9,9 / 55,6	11,1 / 55,1
% com. <20000 hab / % com. >100000 hab	23,0 / 47,4	22,5 / 49,3	23,7 / 47,6	21,0 / 50,1
% résidant en Ile de France	21,1	22,9	22,3	21,7
% résidant dans une maison	66,5	62,6	66,1	60,9
% ayant une résidence secondaire	12,6	20,9	14,4	22,6
% ayant le permis de conduire(PC)	60,4	76,4	64,0	79,3
% n'ayant aucun PC dans son ménage	7,2	3,0	6,1	2,4
% n'ayant pas de voiture dans son ménage	12,3	7,7	11,2	7,1
% ayant une réduction transports	28,9	29,2	29,0	29,3
âge : moyenne (coefficient de variation)	36,5 (57,5)	37,4 (48,4)	36,5 (55,8)	38 (46,3)
revenu annuel du ménage : moyenne (CV%)	112755 (98)	136192 (102)	116229 (102)	143570 (99)
nombre d'enfants : moyenne (CV%)	1,42 (94,5)	1,23 (98,5)	1,37 (95,9)	1,21 (98,3)
nbre de voitures du ménage : moy. (CV%)	1,34 (57,3)	1,46 (50,7)	1,37 (55,7)	1,48 (99,7)
Effectif non pondéré	3110	4633	4735	3008
Effectif pondéré	12,3 millions	15,9 millions	18,1 millions	10 millions

Conditions de déplacement des individus mobiles selon le nombre de déplacements (professionnels exclus)				
nombre de déplacements	2 ou 3	4 et plus	2 à 4	5 et plus
nbre de déplacements : moyenne (CV%)	2,06 (11,5)	6,8 (93,3)	2,69 (34,5)	8,5 (88,9)
nbre de voyages : moyenne (CV%)	1,07 (68,2)	3,5 (111,6)	1,42 (78,1)	4,3 (107,4)
distance totale (en km) : moyenne (CV%)	1210 (176,2)	2479 (127)	1358 (159)	2950 (119)
durée totale (en heures) : moyenne (CV%)	12h45 (2h14)	27h03 (1h36)	14h44 (1h56)	31h50 (1h31)
dist. moyenne d'un dépla. (en km) : moyenne	587	365	505	347
durée moyenne d'un dépla. (en h) : moyenne	6h12	3h59	5h28	3h45
taille du groupe : moyenne (CV%)	4,86 (219)	3,82 (229)	4,7 (222)	3,6 (226)
% se déplaçant en transport collectif	26,5	27,5	25,9	29,2

ANNEXE 2.4 : Tests de validité générale des modèles qualitatifs fournis par la proc logistic de SAS

le rapport de vraisemblance	le critère d'Akaike	le critère de Schwartz
L désignant la Log-vraisemblance, \hat{b} l'estimateur du maximum de vraisemblance (EMV), \hat{b}_0 l'EMV sous la contrainte : $Q'b=0$ (nullité simultanée de tous les coefficients).	$AIC = -2 \log L + 2k$ (k=nombre de paramètres à estimer).	$SC = -2 \log L + k \log I$ (I=nombre total d'observations)
On a : $LRT = 2(L(\hat{b}) - L(\hat{b}_0))$ qui suit asymptotiquement un χ^2 à q ddl (où q=nombre de coefficients (non compris la constante). L'hypothèse de nullité simultanée des coefficients est rejetée si la valeur de la statistique dépasse un seuil critique. En l'occurrence, toutes les estimations conduites présentent des seuils critiques très faibles (0,0001), aussi est-elle toujours rejetée.	Ces deux critères sont utiles pour choisir des modèles différents portant sur les mêmes données. On préférera le modèle pour lequel ces statistiques ont les valeurs les plus faibles.	
Résultats, avec pour variable dépendante :		
Le nombre de déplacements (1) (cf. 1.5 de la 2ème partie pour la définition)	avec le détail des régions : AIC=26695 avec la partition IdF/Prov : AIC=26852	avec le détail des régions : SC=27155 avec la partition IdF/Prov : SC=27161
Le nombre de déplacements (2) (cf. 1.5 de la 2ème partie pour la définition)	avec le détail des régions : AIC=26291 avec la partition IdF/Prov : AIC=26434	avec le détail des régions : SC=26751 avec la partition IdF/Prov : SC=26743
Le nombre de voyages	avec le détail des régions : AIC=26665 avec la partition IdF/Prov : AIC=26817	avec le détail des régions : SC=27125 avec la partition IdF/Prov : SC=27126

ANNEXE 2.5 : Modèles complémentaires pour discussion de l'IIA

Résultats du modèle "car contre train" et comparaison avec le modèle trichotomique "car-avion-train"

Statistique de test	modèle avec uniquement la constante	modèle complet	Chi-deux (prob.)
Akaike	1696.947	1383.907	
Schwartz	1702.191	1452.073	
Max. de vrais. Score	1694.947	1357.907	337.040 (0.0001) 287.111 (0.0001)

Variables	DF	Coefficients	Prob.	Odds Ratio	Comparaison
Constante	1	-0.5688	***	0.566	
Communes rurales	1	0.5524	***	1.737	
Com. petites et moyennes					
Com. grandes	1	-0.6427	***	0.526	signes ≠
Agglom. Paris	1	-2.3194	***	0.098	signes ≠
Agriculteurs/indépendants	1	0.3293	ns	1.390	
Cadres	1	-1.0277	***	0.358	
Employés/ouvriers					
Aucun enfant					
Un enfant	1	0.9710	***	2.641	signes ≠
Deux enfants	1	0.7584	***	2.135	
Aucune voiture	1	-0.6666	***	0.513	signes ≠
Une voiture					
Deux voitures & plus	1	-0.3118	**	0.732	signes ≠
Actifs					
Chômeurs	1	-0.3328	ns	0.717	signes ≠
Jeunes (scol./militaires)	1	0.0915	ns	1.096	signes ≠
Inactifs	1	0.9896	***	2.690	signes ≠

Résultats du modèle "avion contre train" et comparaison avec le modèle trichotomique "car-avion-train"

Statistique de test	modèle avec uniquement la constante	modèle complet	Chi-deux (prob.)
Akaike	1930.187	1811.971	
Schwartz	1935.543	1881.602	
Max. de vrais. Score	1928.187	1785.971	142.216 (0.0001) 134.810 (0.0001)

Variables	DF	Coefficients	Prob.	Odds Ratio	Comparaison
Constante	1	-0.6193	***		
Communes rurales	1	0.1168	ns	1.124	signes ≠
Com. petites et moyennes					
Com. grandes	1	0.1318	ns	1.141	signes ≠
Agglom. Paris	1	0.5880	***	1.800	
Agriculteurs/indépendants	1	0.7735	***	2.167	signes ≠
Cadres	1	0.3381	**	1.402	
Employés/ouvriers					
Aucun enfant					
Un enfant	1	0.2823	*	1.326	signes ≠
Deux enfants	1	0.00389	ns	1.004	
Aucune voiture	1	-0.8991	***	0.407	
Une voiture					
Deux voitures & plus	1	0.3733	***	1.453	
Actifs					
Chômeurs	1	-0.4164	ns	0.659	signes ≠
Jeunes (scol./militaires)	1	-0.5745	***	0.563	signes ≠
Inactifs	1	-0.2648	*	0.767	

