

Num 57134

2F

1C

1T

upl N

Olivier Barbary

Allocataire M.R.E.S.

E.H.E.S.S. - ORSTOM (dépt. SDU)

U.R. : "systèmes d'accumulation des hommes et des richesses"

MISE AU POINT DE PLANS DE SONDAJE POUR L'ESTIMATION DE
CARACTERES DEMOGRAPHIQUES A MARSEILLE

Evaluation des précisions permises par l'emploi de diverses
techniques de sélection, de stratification et d'estimation.

Quito, Janvier 1988.

AVERTISSEMENT

L'un des objectifs principaux du travail présenté ici est l'évaluation des avantages et inconvénients de quatre techniques de sondage. Les formulaires qui leurs correspondent sont exposés complètement au chapitre 4 et intégrés au programme informatique de calcul VARECH.

Lors de la rédaction, en Octobre 86, d'une première version très incomplète de ce rapport, nous ne disposions pas dans ce programme de l'option "estimation par le ratio sous plan à probabilités inégales". Pour des raisons ayant trait au calendrier des travaux à mener à Quito avant le lancement de l'enquête finale, je n'ai pas eu le temps de répéter les calculs sur les données issues du recensement de la population à Marseille pour ce quatrième plan de sondage. Les résultats correspondants manquent donc toujours pour que ce compte rendu soit complet et, pour ne pas retarder plus encore sa livraison, ils ne figureront, totalement exposés et analysés, que dans la thèse qui clôturera ce travail.

Je prie les lecteurs de m'en excuser, cependant les travaux effectués depuis la rédaction de ce rapport sur les données provenant des enquêtes préliminaires réalisées à Quito permettent d'affirmer que les conclusions pratiques qui peuvent être tirées de la lecture du présent rapport ne seront pas significativement modifiées par l'introduction de cette quatrième technique.

PLAN

Pages

7	INTRODUCTION
7	. Antécédents et objectif du rapport, insertion dans le programme ORSTOM
9	. La méthode, justification d'une démarche empirique
11	. Plan simplifié du rapport
13	1. CADRE GEO-DEMOGRAPHIQUE DE L'ETUDE.
	Les indicateurs démographiques intégrés dans la base de données sur Marseille : définition et variations selon les types morphologiques
13	Introduction
13	1.1. Définitions utilisées par l'INSEE
13	1.1.1. Les différentes catégories de population
15	1.1.2. Les caractéristiques individuelles
17	1.1.3. Les immeubles et logements
17	1.2. Stratification morphologique et caractéristiques de la population marseillaise
18	1.2.1. L'immigration étrangère
18	1.2.2. Le chômage
18	1.2.3. La taille des ménages
20	1.2.4. La structure par âge

Pages

25	2. ETUDE STATISTIQUE PRELIMINAIRE.
	Caractères démographiques et stratification morphologique
25	Introduction
26	2.1. Rappel sur la stratification et la base de sondage
27	2.2. Description statistique des caractères suivant les strates
32	2.3. Variance inter-strate et intra-strate des caractères, prévision du gain apporté par la stratification
38	3. MISE AU POINT DES OUTILS ET DE LA METHODE DU TEST.
	Sondages équiprobables, estimation sans biais d'effectifs et de densités, stratification morphologique synthétique
38	Introduction
38	3.1. Base et plan de sondage, caractéristiques des échantillons
41	3.2. Taux de sondage
41	3.3. Formulaire du plan équiprobable sans remise
41	3.3.1. Sans stratification
45	3.3.2. Avec stratification
46	3.3.3. Allocation optimale de NEYMAN
47	3.4. Résultats numériques et commentaires
48	3.4.1. Résultats sur l'ensemble de la base
49	3.4.2. Résultats par strate
54	3.5 Conclusion, perspectives

Pages

54	3.5.1. Précision des estimations, ordre de grandeur des intervalles de confiance, effet de la répartition géographique de l'échantillon
54	3.5.2. Apport de la stratification, estimation globale et estimations partielles, domaines géographiques de précision
56	3.5.3. Comparaison entre les estimateurs de l'effectif et de la densité
57	3.5.4. Conclusion
62	4. FORMULAIRE, PROGRAMME INFORMATIQUE
62	Introduction
63	4.1. Formulaire : estimateur sans biais, estimateur par le ratio, plan équiprobable, plan à probabilités inégales.
63	4.1.1. Préliminaire
64	4.1.2. Conventions
65	4.1.3. Plan équiprobable, estimateur sans biais
66	4.1.4. Plan équiprobable, estimateur par le ratio
67	4.1.5. Plan à probabilités inégales, estimateur sans biais
69	4.1.6. Plan à probabilités inégales, estimation d'un ratio ou estimation d'un total par le ratio
69	. L'estimateur
69	. Expression des biais
71	. Variance des estimations
72	. Allocation optimale
73	4.2. Le programme VARECH associé au logiciel dbaseIII

Pages	
73	4.2.1. Matériel, système gestionnaire de fichiers et base de données sur Marseille
75	4.2.2. Le programme VARECH (variance d'échantillonnage), architecture générale, fonctionnement et domaines d'application
83	5. ESTIMATIONS STRATIFIEES DE CARACTERES DEMOGRAPHIQUES : RESULTATS
83	Introduction
83	5.1. Comparaison de diverses estimations d'une même variable
88	5.2. Comparaison des estimations de plusieurs caractères démographiques
88	5.2.1. Résultats sur l'ensemble de la base : estimateurs et caractères démographiques
88	. Estimation d'effectifs
94	. Estimation de ratios
98	5.2.2. Résultats par strate : variabilité des caractères démographiques et précision des estimations partielles
105	5.3. Conclusion
107	6. STRATIFICATIONS MORPHOLOGIQUES OBTENUES A PARTIR DE L'INFORMATION SATELLITAIRE : RESULTATS
107	Introduction
108	6.1. Provenance et élaboration des nouvelles stratifications
109	6.2. Résultats, commentaires et perspectives
110	6.2.1. Comparaison des stratifications selon la densité du bâti en trois et six classes

Pages

111	6.2.2. Contribution du caractères morphologique "densité du bâti", à une stratification morphologique synthétique
117	7. TENTATIVE D'EVALUATION DE L'EFFICACITE DES TIRAGES AREOLAIRES SYSTEMATIQUES ASSURANT UNE BONNE REPARTITION GEOGRAPHIQUE DE L'ECHANTILLON
117	Introduction
118	7.1. La méthode : approximation du tirage systématique aréolaire par une stratification de la base de sondage sur un critère de localisation
119	7.2. Les données nécessaires au test, élaboration des stratifications et remarques sur l'interprétation des résultats
121	7.3. Résultats et conclusions
128	CONCLUSION
128	. Bilan de l'expérience menée à Marseille : objectifs, déroulement, résultats.
132	. Questions en suspens, perspectives de travail à Quito
134	BIBLIOGRAPHIE
139	ANNEXES
140	. Listage du programme VARECH
181	. Publications et notes de travail de l'équipe ORSTOM

INTRODUCTION

Ce rapport rend compte du travail effectué d'Août 86 à Mai 87 en vue, principalement, d'évaluer le gain de précision qu'apportent divers types de stratification à l'estimation par sondage de caractères démographiques en milieu urbain et de sélectionner des plans de sondage et des estimateurs adaptés au problème. Dans cette introduction, après avoir situé cet objectif dans le cadre du projet de recherche de l'ORSTOM, nous présentons rapidement la méthode empirique qui a été appliquée et nous donnons quelques arguments pour sa justification. On trouvera pour finir un plan sommaire de l'exposé.

. Antécédants et objectif du rapport, insertion dans le programme de l'ORSTOM

Le travail dont l'exposé va suivre s'inscrit dans un programme de recherche initié par F. DUREAU en 1985 au sein du département "Urbanisation et socio-systèmes urbains" de l'ORSTOM et dont l'intitulé est : "Utilisation de la télédétection pour l'observation des populations urbaines". Un des objectifs principaux de ce programme est la mise au point d'une méthode de collecte par sondage d'informations démographiques en milieu urbain, et son optimisation grâce à l'emploi d'images satellites. Ces dernières, outre qu'elles fournissent la base de sondage spatiale nécessaire à l'opération de sélection de l'échantillon, permettent une stratification de l'espace urbain sur la base d'une information morphologique exhaustive. Une des hypothèses fondatrices du projet et qu'il va s'agir ici de vérifier, est que l'emploi d'une telle stratification améliore grandement la précision de l'estimation par sondage des informations recherchées. De plus, comme nous le verrons par la suite, la technique du tirage aréolaire systématique sur document satellitaire induit un plan de sondage particulier dont la précision des estimations qu'il fournit doit être évaluée.

Afin de permettre la mise au point de cette méthode, la première étape (juin 85 - avril 86) fut la constitution, sur la ville de Marseille choisie comme premier site d'expérimentation, d'une base de données morphologiques et démographiques localisées. Cette base comprend des informations issues des photographies aériennes, l'image satellite Thematic Mapper et une sélection d'indicateurs démographiques fournis par le recensement général de la population effectué par l'INSEE à Marseille en 1982. L'ensemble de ces informations est géré informatiquement par le système d'information géographique TIGRE développé par l'Unité d'Infographie de l'ORSTOM.

Un travail d'analyse typologique des données morphologiques issues des photographies aériennes a ensuite permis d'obtenir une stratification de l'agglomération marseillaise, ceci pour satisfaire deux objectifs principaux :

1. Permettre le tirage d'un échantillon de districts INSEE représentatif des divers types de morphologie urbaine présents à Marseille sur lequel s'effectue l'étude précise des relations entre la morphologie observée au sol (ou sur photographie aérienne) et l'information spectrale satellitaire. Cette étude doit conduire les spécialistes en télédétection de l'équipe à produire, directement d'après l'image satellite, des stratifications utilisables pour réaliser des enquêtes par sondage.

2. Permettre, à partir des données du RGP 82 à Marseille, d'évaluer l'efficacité d'une stratification morphologique de la ville pour estimer par sondage les caractéristiques démographiques des habitants. Ces mesures seront répétées pour divers plans de sondages, différents soit par la probabilité de sélection des unités statistiques, soit par l'estimateur employé. Le but final est ici de sélectionner, après avoir évalué sa précision, le plan de sondage "optimal" que l'on devra utiliser lors de la phase d'application de la méthode à Quito, second site test du programme.

Le présent rapport traite du second de ces objectifs. Dans un premier temps (chapitres 1 à 3), pour cerner le problème en introduisant la méthode et un formulaire simple, nous ne considérons que des plans de sondage équiprobables. Ensuite (chapitre 4 à 7), nous examinons des plans où le tirage est à probabilités inégales : plus précisément, la probabilité de tirage sera proportionnelle à la surface des unités statistiques.

Il ne faut en effet pas perdre de vue que, dans la situation concrète d'application de la méthode, on ne disposera pas, a priori, d'une liste des unités composant la base de sondage. Le tirage s'effectuera à partir d'une grille de points plaquée sur une image satellite; cette image ne comportera que les limites externes de l'espace urbain et celles des strates morphologiques. Dans ce cas de figure et cela quelque soit

l'unité spatiale considérée, le tirage s'effectue avec des probabilités proportionnelles à la surface de ces unités. Il s'agit donc ici, entre autres, de comparer ce plan de sondage, imposé par la technique d'échantillonnage que l'on devra appliquer à Quito, à son concurrent classique, le plan équiprobable sans remise.

. La méthode, justification d'une démarche empirique

On s'apercevra très vite, à la lecture de ce compte rendu, que le travail effectué ici procède de la méthode la plus empirique. Il s'agit en effet, pour l'essentiel, de vérifier, par expérimentation sur des données réelles, l'hypothèse de départ du projet : la prise en compte, dans la stratification de la base de sondage, des différences morphologiques intra-urbaines, permet un gain appréciable de précision lors de l'estimation des caractéristiques démographiques des habitants. Cette vérification, ainsi que l'ensemble des résultats du présent travail, sont obtenus grâce au programme VARECH (voir chapitre 4) qui, en appliquant le formulaire des principaux plans de sondage classiques, permet le calcul exact sur les données exhaustives du recensement de population des variances des différents estimateurs. La démarche expérimentale, suivie ici a donc pour objectif principal l'évaluation chiffrée d'un ensemble d'alternatives possibles, cela afin d'effectuer en connaissance de cause les choix techniques conduisant à un sondage efficace en milieu urbain et pratiquement réalisable sur document satellitaire. Accumuler le maximum de constatations pratiquement exploitables lors de l'application à Quito, sans toujours en chercher l'explication ou une formalisation mathématique du phénomène qui les sous-tend, tel est le but d'un travail qui se veut avant tout appliqué quitte à rester exploratoire. En particulier, on ne trouvera ici aucun développement théorique nouveau, ni même de références aux avancées récemment effectuées dans le domaine de l'estimation des "variables régionalisées", pour reprendre le vocabulaire de G. MATHERON. Les lecteurs mathématiciens s'en trouveront probablement déçus, mais pour qu'ils ne croient pas qu'il s'agit là d'un a priori rétrograde face à ces nouvelles théories, il ne paraît pas inutile de tenter de justifier brièvement cette absence.

Une des voies importantes du développement récent des techniques de sondage, tant du point de vue théorique que dans des applications plus pratiques, est l'étude mathématique des caractéristiques spatiales des variables que l'on cherche à estimer et leur modélisation. Dans les sciences de la nature, en géostatistique ou dans le domaine des statistiques forestières, cette étude se réfère le plus souvent à la

théorie des processus stochastiques ponctuels ou fonctions aléatoires (1). J. DELINCE en évoque ainsi l'objet : "plutôt qu'un test du caractère aléatoire de la population, ces techniques cherchent à décrire la structure (du processus) à différentes échelles et proposent des tests d'adéquation vis à vis de modèles quelquefois fort complexes" (2).

Outre son éventuelle complexité, l'utilisation de ces modèles nécessite généralement qu'on fasse des hypothèses sur la variation spatiale du phénomène étudié. Par exemple, GIUDICELLI, LANLY, OUAKAM et PIETRI, dans un article publié dans les Annales des Sciences Forestières (3), examinent le problème suivant : "L'estimation de l'erreur statistique dans les dispositifs de sondage systématique, largement utilisés en forêt, est un problème qui n'est pas soluble dans le strict cadre de la théorie classique des sondages. Conformément à une suggestion faite par COCHRAN et d'autres, les auteurs proposent une solution en considérant la forêt comme une réalisation d'un processus aléatoire stationnaire et isotrope". Plus loin, ils justifient ainsi les hypothèses faites :

"Dans le cas d'une forêt :

- l'hypothèse de stationnarité peut se justifier par une certaine stabilité, dans toute la forêt, des conditions d'apparition et de développement des arbres; s'il semble que cette hypothèse n'est pas réalisée - par exemple lorsqu'une certaine tendance apparaît due à la modification de facteurs écologiques (reliefs, sol, microclimat) - l'hypothèse de stationnarité peut être acceptée au niveau de strates judicieusement choisies.

- L'hypothèse d'isotropie, vérifiée si la corrélation entre les valeurs de la caractéristique cherchée dans deux parcelles situées à une distance donnée est la même dans toutes les directions, n'est satisfaite dans une forêt que s'il n'apparaît pas une ou plusieurs directions privilégiées de la répartition de l'essence étudiée."

(1) G. MATHERON (ed. 1965) : Les variables régionalisées et leur estimation, une application de la théorie des fonctions aléatoires aux sciences de la nature. MASSON, PARIS.

(2) J. DELINCE (1983) : Estimations robustes de densité dans le plan, étude comparée de techniques d'échantillonnage, application en forêt. Université Catholique de Louvain, Faculté des Sciences Agronomiques.

(3) X. GUIDICELLI, J.P. LANLY, J.B. OUAKAM et M. PIETRI (1972) (ISUP) : Application de la théorie des processus aléatoires à l'estimation de la précision d'un inventaire forestier par échantillonnage systématique. Annales des Sciences Forestières, Institut National de la Recherche Agronomique.

Malheureusement, dans le cadre qui nous occupe - l'étude des paramètres démographiques en milieu urbain - il apparaît peu raisonnable de penser que ces hypothèses (stationnarité et isotropie) sont vérifiées au niveau de l'agglomération entière. Les déterminants de la variation spatiale du peuplement humain des villes nous semble en effet trop complexes et encore trop mal connus pour que l'on puisse faire, avec quelques chances de succès, des hypothèses aussi fortes que celles qui peuvent se justifier dans le cas des peuplement des milieux naturels ou dans celui de la géostatistique.

De la même manière, si l'on cherche à définir des strates "judicieuses" à l'intérieur desquelles ces hypothèses seraient plus plausibles, on se heurte très vite au fait que l'hétérogénéité des caractéristiques démographiques des habitants reste élevée, même au sein de quartiers morphologiquement homogènes.

Pour en finir avec cette tentative de justification de la méthode, il est un dernier argument que l'on peut avancer. L'objet des statisticiens spécialistes de sondages dans le domaine des sciences de la nature est souvent de perfectionner en la raffinant la méthode de l'estimation par échantillonnage aréolaire systématique, laquelle, dans ce milieu et pour le type d'information recherchée, a déjà fait ses preuves.

Notre objectif en revanche, dans un domaine où l'on en est qu'aux premières tentatives, est de mettre au point cette même méthode du sondage spatial en milieu urbain et d'évaluer la précision qu'elle permet d'atteindre, pour un coût donné, dans l'estimation de données démographiques en ville.

Pour l'ensemble de ces raisons, on s'en tiendra donc dans l'ensemble de ce rapport aux méthodes déjà classiques du point de vue des théories mathématiques qui les sous-tendent mais qui, comme on le verra, conduisent néanmoins à des résultats intéressants.

. Plan simplifié du rapport

Les deux premiers chapitres forment une étude préliminaire succincte du cadre géo-démographique du travail : la ville de Marseille et le recensement général de la population de 1982 (chap. 1), les distributions des variables que l'on va chercher à estimer dans la suite (chap. 2).

Les chapitres 3 et 4 présentent les outils statistiques (plans de sondages et estimateurs) et informatiques (programme de calcul des variances d'estimations) qui, appliqués aux résultats du recensement de 1982 à Marseille, permettent d'effectuer l'évaluation des différentes techniques d'échantillonnage envisagées.

Le chapitre 5 fournit les résultats obtenus pour plusieurs indicateurs démographiques à l'aide de divers plans de sondages et évalue l'intérêt d'une stratification morphologique de la base de sondage.

Dans le chapitre 6, on s'intéresse à des stratifications proches de celles qui pourront être obtenues directement de l'information satellitaire et on mesure le gain de précision qu'elles apportent.

Enfin, le chapitre 7 examine l'apport spécifique du caractère systématique du sondage, la bonne répartition spatiale des unités sondées, et tente par une méthode d'approximation d'évaluer le gain de précision attaché à ce type de sondage.

La conclusion résume les principaux résultats, définit les acquis qu'a permis ce travail pour la définition de la technique de sondage à mettre en oeuvre à Quito et trace les grandes lignes du travail restant à effectuer.

CHAPITRE 1

CADRE GEO-DEMOGRAPHIQUE DE L'ETUDE LES INDICATEURS DEMOGRAPHIQUES INTEGRES DANS LA BASE DE DONNEES SUR MARSEILLE : DEFINITIONS ET VARIATIONS SELON LES TYPES MORPHOLOGIQUES (*)

Introduction

Les données démographiques utilisées sur Marseille dans le cadre du programme sur "L'utilisation de la télédétection pour l'observation des populations urbaines" sont issues du Recensement Général de la Population, réalisé par l'INSEE en Mars 1982.

Avant d'aborder l'exploitation de ces données pour l'axe de recherche "sondages spatiaux en démographie" (chapitres 2 à 7), il importe de rappeler les définitions utilisées par l'INSEE pour cette opération de collecte ; ce sera l'objet de la première partie de ce chapitre.

Dans un deuxième temps, nous nous intéresserons aux quelques indicateurs socio-démographiques intégrés dans la base de données sur Marseille, à leurs variations selon les strates morphologiques.

1. 1. DEFINITIONS UTILISEES PAR L'INSEE

1.1. 1. Les différentes catégories de population

L'INSEE distingue quatre catégories de population, selon "le type de logement et le mode de vie, centré sur la famille ou organisé de façon communautaire" (1) :

- **La population des ménages ordinaires** : "Un ménage ordinaire est constitué de l'ensemble des occupants d'une unité d'habitation privé (local séparé et indépendant) occupée comme résidence principale."

- La population "Hors ménage ordinaire", qui comprend :

- **La population des ménages collectifs** : "Un ménage collectif est un

(*) Ce chapitre est intégralement repris de l'annexe 12 du Rapport ATP CNRS/CNES d'Août 1987, annexe co-rédigée par F. DUREAU et O. BARBARY en Décembre 1986.

(1) Toutes les citations entre guillemets du chapitre 1. 1 sont de l'INSEE (Recensement général de la population de 1982. Guide d'utilisation).

Population des ménages ordinaires
vivant effectivement à leur résidence
personnelle dans la commune A

Population des habitations
mobile et recensée dans la commune
A où ils se trouvaient le 4 mars 1982

Population des ménages collectifs
de la commune A

- 1) Infirmiers, personnel de service d'un établissement hospitalier
- 2) Professeurs, surveillants, personnel de service d'un établissement public ou privé avec internat
- 3) Personnel de service d'un hôtel
- 4) Membres d'une communauté religieuse
- 5) Malades ou pensionnaires d'un établissement de soins ou de convalescence
- 6) Etudiants logés en cité universitaire ou foyer d'étudiants
- 7) Travailleurs logés dans un foyer
- 8) Vieillards vivant dans une maison de retraite ou un hospice
- 9) Autres cas

Ayant une résidence personnelle
dans la commune B de l'étab.

Ayant une résidence personnelle
dans une autre commune A de
France métropolitaine

Sans résidence personnelle en
France métropolitaine

POPULATION MUNICIPALE
de la commune A

Population comptée à part de la commune B

- 1) Militaires de l'armée de terre, de l'air ou de la marine logés dans des casernements, camps ou assimilés
- 2) Elèves internes (lycées, collèges, écoles normales, grandes écoles etc)
- 3) Personnes en traitement dans un sanatorium, un préventorium, ou un aérium
- 4) Personnes logées sur un chantier temporaire

- 5) Personnes en traitement dans un hôpital psychiatrique
- 6) Détenus dans des établissements pénitentiaires, mineurs confiés à un établissement d'éducation surveillée, personnes recueillies

Population comptée à part de la commune A

- 1) Militaires de l'armée de terre, de l'air ou de la marine logés dans des casernements, camps ou assimilés
- 2) Elèves internes (lycées, collèges, écoles normales, grandes écoles etc)
- 3) Personnes en traitement dans un sanatorium, un préventorium, ou un aérium
- 4) Personnes logées sur un chantier temporaire

- 5) Personnes en traitement dans un hôpital psychiatrique
- 6) Détenus dans des établissements pénitentiaires, mineurs confiés à un établissement d'éducation surveillée, personnes recueillies dans un centre d'hébergement ou d'accueil.

Ayant une résidence personnelle
dans la commune A de l'étab

Ayant une résidence personnelle
dans une autre commune B de
France métropolitaine

Sans résidence personnelle en
France métropolitaine

+ Population municipale de la commune A
= POPULATION TOTALE de la commune A

POPULATION SANS DOUBLES COMPTES de la commune A = Population totale de la commune A - Population comptée à part des catégories 1 à 4 qui n'ont pas leur résidence personnelle dans la commune

groupement de personnes vivant en communauté, logeant dans certains établissements en chambres individuelles ou collectives et prenant souvent leur repas en commun." Le tableau ci-contre précise les 9 catégories de personnes constituant les ménages collectifs.

- **La population comptée à part** : cette catégorie est composée de 6 types de population, répertoriés dans le tableau ci-contre.

- **La population des habitations mobiles** : il s'agit "des personnes itinérantes logeant en caravane ou roulotte".

A partir de ces catégories de population, l'INSEE calcule la population des communes selon trois définitions :

- population municipale
- population totale (ou légale)
- population totale sans doubles comptes.

Dans le tableau ci-contre, sont indiqués les modes de calcul de ces trois types de population.

1. 1. 2. Les caractéristiques individuelles

Nous ne rappellerons ici que les définitions des caractéristiques individuelles que nous avons retenues pour l'intégration dans le système d'information géographique TIGRE (dont la liste figure dans le tableau ci-dessous).

Tableau 1 - Liste des attributs intégrés dans la base de données sur Marseille

1. Population municipale
2. Population sans double compte
3. Nombre de ménages ordinaires
4. Nombre de ménages collectifs
5. Population comptée à part
6. Nombre de résidences principales
7. Nombre de résidences secondaires
8. Nombre de logements vacants
9. Nombre d'immeubles
10. Pourcentages d'étrangers (Hommes + femmes)
11. Pourcentage d'étrangers (Hommes > 19 ans)
12. Pourcentage de chômeurs (Hommes > 19 ans)
13. Taille moyenne des ménages
14. Pourcentage de ménages d'une personne
15. Pourcentage de ménages de 6 personnes et plus
16. Pourcentages de 0-4 ans révolus (Hommes + femmes)
17. Pourcentage de 60 ans et plus (Hommes + femmes)

Age

Dans les statistiques du recensement de 1982, "l'âge indiqué est l'âge atteint au cours de l'année 1982 (différence entre 1982 et l'année de naissance) qui est également l'âge en années révolues au 31 Décembre 1982".(1)

C'est à dire que les personnes classées "10 ans" au recensement de 1982 avaient en fait entre 9 ans 2 mois et 10 ans 2 mois, à la date de référence du recensement, soit le 4 mars 1982.

Nationalité

La population est répartie en "Français de naissance", "Français par acquisition" (classés selon leur nationalité antérieure), et "Etrangers" (classés selon leur nationalité).

Parmi les étrangers présents en France au moment du recensement de 1982, n'ont été recensés que "ceux qui ont leur résidence permanente en France et ceux qui y travaillent ou étudient (ainsi, le cas échéant, que leur famille)", y compris "le personnel étranger (administratif, technique ou de service) des ambassades résidant de façon permanente en France".

N'ont pas été recensés les "travailleurs saisonniers et les travailleurs frontaliers" ainsi que "les étrangers membres du corps diplomatique".

Type d'activité

Toutes les personnes recensées en 1982 atteignant leur quinzième anniversaire l'année du recensement sont classées selon le type d'activité, c'est à dire dans une des rubriques suivantes :

- ACTIFS : "La population active comprend la population active ayant un emploi et les chômeurs".
- INACTIFS : cette catégorie regroupe :
 - les anciens actifs (retraités n'ayant pas d'emploi et n'en cherchant pas
 - les étudiants ou élèves (sans activité professionnelle)
 - les militaires du contingent
 - les autres inactifs

Dans le cadre du recensement, la catégorie "chômeurs", qui, comme nous venons de le voir font partie des "actifs", regroupe les personnes qui :

- ne sont pas étudiants ou stagiaires rémunérés
- et - n'ont pas répondu aux questions 12 à 15 relatives à la profession exercée actuellement (voir bulletin individuel du recensement ci-joint)
- et - ont répondu à la question 16a "être chômeur, ou mère de famille, ou retraité" (ou pas de réponse) et à la question 16b "rechercher actuellement un travail".

(1) Toutes les citations entre guillemets des chapitres 1.2 et 1.3 sont de l'INSEE (Recensement général de la population de 1982. France métropolitaine).

Une personne sans travail n'en recherchant pas est classée "inactive".
Un étudiant ou élève recherchant un travail est classé "étudiant ou élève".
Par contre, un retraité recherchant un travail est classé "chômeur".

1. 1. 3. Immeubles et logements

Est considéré comme immeuble :

- " toute construction habitée au moment du recensement, quelle que soit sa nature, les matériaux utilisés, y compris les immeubles vétustes, les habitations de fortune, et les immeubles en cours de construction partiellement habités ;
- toute construction à usage d'habitation, même sans occupant à l'époque du recensement, à l'exception des immeubles en démolition ou devenus totalement inhabitables par vétusteté ;
- tout ensemble de bâtiments administratifs, industriels ou commerciaux, même inhabités."

On appelle "Logement", tout local qui satisfait aux trois conditions suivantes :

- "séparé, c'est à dire complètement fermé par des murs et cloisons sans communication avec un autre local, si ce n'est sur les parties communes de l'immeuble"
- "indépendant, c'est à dire qu'il doit posséder une entrée indépendante où l'on a directement accès sur l'extérieur ou sur les parties communes de l'immeuble sans avoir à traverser un autre local"
- "utilisé pour l'habitation, en tenant compte de sa destination actuelle et non de sa destination primitive".

L'INSEE distingue plusieurs catégories de logements :

- les résidences principales : "logements occupés de façon permanente et à titre principal par un ménage"
- les logements vacants : "logements sans occupants"
- les résidences secondaires : logements "soit utilisés par le propriétaire, sa famille ou ses amis, soit loués à des particuliers, pour des séjours temporaires, en général pendant les vacances, fins de semaine...etc"

1. 2. STRATIFICATION MORPHOLOGIQUE ET CARACTERISTIQUES SOCIO-DEMOGRAPHIQUES DE LA POPULATION MARSEILLAISE

Le Tableau 2 fournit les moyennes arithmétiques des indicateurs de structure

démographique intégrés dans la base de données gérée par le système TIGRE calculées sur l'ensemble des districts de recensement (à ne pas confondre avec la moyenne du caractère dans chaque strate) ; il confirme l'intérêt de la stratification en types morphologiques pour l'appréhension des structures de population à Marseille.

1. 2. 1. L'immigration étrangère

La ville de Marseille est bien connue pour sa forte proportion d'étrangers ; la division de l'agglomération en types morphologiques montre de très fortes différences dans la répartition des étrangers dans la ville. Le phénomène de concentration de la population étrangère dans le centre ville dense et les quartiers industriels du nord de la ville ressort clairement de l'examen des deux premières lignes du tableau 2 : le pourcentage d'étrangers (hommes et femmes confondus) oppose le centre ville dense (12 %) et les quartiers industriels (9 %) au reste de la ville, où il est partout inférieur à 6 %. Les taux d'étrangers les plus faibles (entre 3 et 4,5 %) se rencontrent dans les strates "petits immeubles" et "pavillonnaire".

1. 2. 2. Le chômage

Pour le taux de chômeurs parmi les hommes de plus de 19 ans, les différences entre moyennes par strate morphologique et moyenne générale sur la zone d'étude sont beaucoup moins importantes que précédemment : pour une moyenne générale de 7,55 %, les valeurs par strate s'échelonnent entre 5,31 et 9,67 %. La répartition des chômeurs à Marseille semble beaucoup plus homogène que celle des étrangers. La faible sensibilité de cet indicateur de chômage aux types de morphologie urbaine pourrait s'expliquer par la nature même du phénomène : tandis qu'une caractéristique individuelle permanente telle que la nationalité a un effet ségrégoire certain, le chômage constitue une caractéristique temporaire qui, si elle ne se prolonge pas trop longtemps, n'a pas d'effet systématique et immédiat sur le lieu et le type de résidence des personnes.

En dépit de cette relative homogénéité des taux de chômeurs par strate, on note toutefois que les taux les plus élevés se rencontrent dans le centre ville dense et les grands immeubles d'habitation.

1. 2. 3. La taille des ménages

Taille moyenne des ménages et taux de ménages d'une personne apparaissent négativement corrélés et créent trois sous-ensembles dans l'espace marseillais :

- le centre ville et la strate 8 (immeubles élevés denses) sont caractérisés par une faible taille moyenne des ménages (2,2 personnes par ménage), avec un fort pourcentage de ménages d'une personne (37 %).

- dans les quartiers industriels, le pavillonnaire et la strate 7, les indicateurs de taille des ménages sont semblables à ceux observés sur l'ensemble de la zone d'étude : 2,5 personnes par ménage, et 28 % de ménages d'une personne.

- les grands ensembles, les petits immeubles et le péri-urbain ont des ménages un peu plus grands que la moyenne (2,7 personnes par ménage), et comptent peu de ménages d'une personne (20 %).

Tableau 2 - Indicateurs de structure de population selon les strates de morphologie urbaine (ensemble des districts INSEE de la base de sondage - moyennes des valeurs observées par district))

Strate	1	2	3	4	5	6	7	8	9	TOTAL
%Etrangers (H + F)	12,18	9,11	5,76	3,29	3,50	4,48	5,05	4,36	3,70	7,27
%Etrangers (H > 19 ans)	13,79	9,77	5,87	3,94	3,94	4,75	5,88	5,17	4,09	8,09
%Chomeurs (H > 19 ans)	9,67	6,72	7,92	6,28	5,65	5,31	7,10	5,85	6,22	7,55
Taille moy. des ménages	2,18	2,48	2,86	2,68	2,55	2,69	2,35	2,11	3,03	2,48
%Ménages d' 1 personne	37,68	28,75	20,51	20,36	23,70	20,50	29,54	36,69	13,60	28,28
%Ménages >= 6 personnes	2,13	3,11	5,38	3,03	2,55	3,65	1,64	0,88	3,40	2,93
% 0-4 ans révolus	3,93	4,02	4,93	4,20	3,31	3,28	3,46	3,52	3,33	3,89
% >= 60 ans exacts	29,56	24,94	19,94	22,05	29,64	28,33	26,92	32,85	19,32	27,01

(Source : Recensement Général de la Population de 1982, INSEE, Résultats par district de recensement)

Le découpage induit par le pourcentages de ménages de 6 personnes et plus, s'il présente des similitudes avec le précédent, ne lui est pourtant pas strictement équivalent. Les grands ensembles s'isolent avec un pourcentage de familles très nombreuses (5,4 %) presque deux fois supérieur à la moyenne générale. Les strates péri-urbain, quartiers industriels et petits immeubles d'habitation ont des valeurs très proches de la moyenne (3 %) ; enfin, le centre ville et le pavillonnaire comptent peu de ménages nombreux, respectivement 2,1 et 2,5 % de ménages de 6 personnes et plus.

1. 2. 4. La structure par âge

Si les deux indicateurs de structure par âge retenus (% de moins de 5 ans et % de plus de 60 ans) ne varient que relativement peu d'une strate à l'autre (respectivement 3,3 à 5 %, et 19 à 33 %), ces variations semblent toutefois significatives, surtout sur des ensembles où les effectifs de population sont importants.

On observe une bonne corrélation négative entre ces deux indicateurs de la structure par âge. Les grands ensembles d'habitation sont les quartiers où les jeunes enfants sont les plus présents, suivi des petits immeubles d'habitation, puis du centre ville dense et des quartiers industriels, où les pourcentages sont comparables à ceux de l'ensemble de la zone d'étude. A l'opposé, les strates pavillonnaire et péri-urbain comptent peu de jeunes enfants et beaucoup de personnes âgées.

L'intérêt des quelques lignes qui précèdent ne réside pas, bien évidemment, dans l'embryon d'analyse socio-démographique de la ville de Marseille basée sur une division en neuf strates morphologiques. Par contre, l'analyse du tableau 1 montre que les variations des indicateurs de structure sont suffisamment significatives du point de vue numérique, et cohérentes avec ce que l'on sait sur ce sujet, pour rendre *intéressante l'introduction d'une stratification morphologique dans les plans de sondage d'enquêtes ne visant pas seulement l'estimation d'effectifs totaux de population, mais aussi l'analyse des structures et comportements démographiques.*

BORDEREAU DE MAISON

REPUBLIQUE FRANCAISE
RECENSEMENT GÉNÉRAL DE LA POPULATION DE 1982

IMPRIMÉ NUMFRO 4	Ce bordereau sera rempli par l'agent recenseur	Numéro du district de recensement	Cachet de la Mairie
		Numéro d'ordre de l'immeuble <small>Pour attribuer ce numéro d'ordre voir le paragraphe 5.1 du Manuel de l'agent recenseur</small>	

Établissez un bordereau de maison

- Pour toute construction habitée, quelle que soit sa nature, quels que soient les matériaux utilisés, y compris les immeubles vétustes, les habitations de fortune; y compris également les immeubles en cours de construction partiellement habités.
- Pour toute construction à usage d'habitation, même sans occupants à l'époque du recensement (logements vacants et résidences secondaires), à l'exception des immeubles en démolition ou devenus totalement inhabitables par vétusté.
- Pour tout ensemble de bâtiments administratifs, industriels ou commerciaux, même inhabités.

Ne remplissez pas ce bordereau pour les caravanes, roulottes et autres habitations mobiles.

Adresse précise: N° _____ Rue (ou lieudit) _____
Commune _____ Département: _____
(Pour Paris Lyon Marseille préciser l'arrondissement)
Donnez toutes autres indications permettant de repérer l'immeuble _____

LISTE DES LOGEMENTS D'HABITATION

- Cette liste sera établie après achèvement de la collecte, à raison d'une ligne pour chaque logement d'habitation des catégories 1 à 7 énumérées sur la première page de la feuille de logement (reportez-vous au paragraphe 5.1 du Manuel de l'agent recenseur)
- S'il n'y a aucun logement d'habitation dans l'immeuble, inscrivez "néant" sur la première ligne.
- N'inscrivez pas les locaux occupés par de la population comptée à part ou par des ménages collectifs (les imprimés n° 3 et n° 1 bis correspondants ne seront pas classés dans les bordereaux de maison)

Nombre de d'après la colonne 4 ci-dessous	résidences principales (Catégories 1 à 5)	<input type="text"/>
	logements vacants (Catégorie 6)	<input type="text"/>
	résidences secondaires (Catégorie 7)	<input type="text"/>
	Nombre total de logements d'habitation	<input type="text"/>

A remplir par l'agent recenseur				A remplir par la Mairie				
Numéro du logement	Localisation dans l'immeuble (escalier, étage, situation sur le palier, numéro du logement ou de la chambre, etc.)	Nom de l'occupant	Catégorie de logement (1 à 7), voir FL	Nombre d'imprimés n° 2		Nombre de volets B réintégrés		
				supprimés (raison double emploi avec des volets B)	ajoutés (autres recueils pour des personnes en déplacement et ne faisant pas double emploi)			
1	2	3	4	5	6	7	8	
01								
02								
03								
04								
05								
06								
07								
08								
Total				partiel (à reporter en haut de la page 2)	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
					général	<input type="text"/>	<input type="text"/>	<input type="text"/>

PAGE 4

CARACTÉRISTIQUES DE L'IMMEUBLE

1 TYPE D'IMMEUBLE

- Ferme bâtiments agricoles 1
- Habitation de fortune (telle que wagon, baraque de "bidonville", bâtiment habité bien ou en ruines, etc.) ou construction provisoire à usage d'habitation 2
- Hôtel, pension de famille garni (occupant la totalité ou la plus grande partie de l'immeuble) 3
- Maison individuelle ou immeuble collectif entièrement ou principalement utilisé pour l'habitation 4
- Immeuble principalement à usage industriel, commercial, administratif ou public; un tel immeuble peut comprendre un ou plusieurs logements d'habitation pour personnel logé (c'est souvent le cas du directeur, du gardien, etc.) Précisez la nature exacte de l'immeuble 5

Exemples: usine, atelier, magasin, dépôt, grand magasin, banque, immeuble aménagé en bureaux, Hôpital, clinique, maison de repos, sanatorium, dispensaire, Ecole, internat, collège, Immeuble administratif, mairie, caserne, gare, bureau de poste, église, stade, etc.

Si vous hésitez sur un cas particulier, décrivez-le

- 2 Cet immeuble appartient-il à un organisme H.L.M. (société ou office)? OUI 1 NON 0
Si oui, précisez le nom de l'organisme propriétaire: _____

3 ANNÉE D'ACHÈVEMENT DE LA CONSTRUCTION

Si les différentes parties ne sont pas de la même époque, indiquez l'année d'achèvement de la partie habitée, ou de la partie habitée la plus importante.

- Avant 1871 1
 - de 1871 à 1914 2
 - de 1915 à 1948 3
 - de 1949 à 1960 4
 - de 1961 à 1967 5
 - de 1968 à 1974 6
 - 1975 ou après 7
 - Immeuble en cours de construction partiellement habité 8
- Dans ce cas, précisez l'année d'achèvement

4 NOMBRE D'ÉTAGES AU-DESSUS DU REZ-DE-CHAUSSÉE:

Y compris les étages mansardés habitables. Ne comptez pas les caves et les sous-sols; le rez-de-chaussée surélevé est assimilé à un rez-de-chaussée ordinaire, mais l'entresol compte pour un étage. S'il y a plusieurs corps de bâtiment, indiquez le nombre d'étages du corps de bâtiment le plus élevé

5 ASCENSEUR

- Y a-t-il un ascenseur dans l'immeuble? OUI 1 NON 0

6 ORIGINE DE L'EAU POUR USAGES DOMESTIQUES

- Immeuble raccordé à un réseau collectif de distribution (ville, commune, service des eaux, usine, etc.) 1
- Immeuble raccordé par une canalisation à un point d'eau (puits, citerne, source, etc.) 2
- L'eau courante n'est pas installée dans l'immeuble 3

7 GAZ

- Immeuble raccordé à un réseau public de distribution 1
- Immeuble alimenté par une citerne de gaz 2
- Immeuble alimenté par une installation fixe de bouteilles de gaz 3
- Ni raccordement à un réseau public ni alimentation par une installation fixe (citerne ou bouteilles) 4

8 ÉVACUATION DES CABINETS D'AISANCES (W.-C.) ET DES EAUX MÉNAGÈRES

- A Cabinets d'aisances:
- Raccordement à l'égout 1
 - Autres cas (fosse septique, fosse fixe, etc.) 2
- B Eaux ménagères (évier, lessive, toilette, etc.)
- Raccordement direct à l'égout 1
 - Autres cas (caniveau, fossé, puisard, etc.) 2

9 CHAUFFAGE CENTRAL

- Chauffage urbain 1
- Chauffage collectif pour un groupe d'immeubles 2
- Installation collective de chauffage central propre à l'immeuble (desservant la totalité ou la plupart des logements de l'immeuble), maison individuelle disposant du chauffage central 3
- Pas d'installation de chauffage central (même si certains logements disposent d'installations individuelles) 4

Si vous avez marqué la case 2 ou la case 3, indiquez le combustible utilisé.

Charbon	<input type="checkbox"/> 1
Bois	<input type="checkbox"/> 2
Marou (fuel)	<input type="checkbox"/> 3
Gas	<input type="checkbox"/> 4
Électricité	<input type="checkbox"/> 5
Autres	<input type="checkbox"/> 6
Précisez _____	

10 EXPLOITATIONS AGRICOLES

- Cet immeuble est-il le siège d'une exploitation agricole? OUI 1 NON 0

Si oui:

A Superficie agricole utilisée: _____ hectares, _____ ares
(Ne pas tenir compte des bois, étangs, terrains à bâtir, parcs et jardins d'agrément, landes et friches improductives, bâtiments et cours)

- B Orientation des productions agricoles (cochez une seule case)
- Exploitation avec une production principale:
 - Polyculture (cultures de terres labourables) 1
 - Marâchage ou horticulture 2
 - Vignes ou arbres fruitiers 3
 - Élevage d'herbivores (bovins ovins) 4
 - Élevage de granivores (porcs volailles) 5
 - Autre cas:
 - Polyculture - élevage 6
 - Élevage d'herbivores et de granivores 7
 - Autres 8

- C Une personne au moins travaille-t-elle à plein-temps sur l'exploitation (chef d'exploitation, non membre de la famille ou salarié agricole)? OUI 1 NON 0

1

Cet imprimé sera rempli pour tout logement d'habitation, occupé ou non.

Numero du district de recensement

Numero d'ordre de l'immeuble

Numero du logement ou de l'habitation

POURQUOI LE RECENSEMENT ?

Le recensement général de la population est une opération à laquelle procèdent régulièrement tous les pays. Il a pour but de déterminer le nombre d'habitants de chaque commune, de connaître la répartition de la population selon l'âge, la profession, la branche d'activité, d'analyser les migrations, la composition et l'équipement des immeubles et logements. Ces informations sont indispensables à l'étude de tout problème démographique, économique ou social dans le cadre national ou communal.

Prescrit par le décret du 28 avril 1981 (Journal Officiel du 30 avril 1981), le recensement général de 1982 — le 31^e depuis 1801 — est obligatoire pour toutes les personnes résidant en France.

Sur le secret statistique garanti que les renseignements individuels figurant sur les questionnaires sont utilisés uniquement à des fins statistiques. Ces questionnaires confidentiels sont remis à la Direction Nationale de la Statistique et des Etudes Economiques (I.N.S.E.E.); il est interdit d'en prendre copie sous peine de sanctions prévues à l'article 44 de la loi n° 13-1 du 6 juin 1978 (voir la notice explicative).

COMMENT REMPLIR LES IMPRIMES ?

Remplissez d'abord le présent imprimé en commençant par les pages 2 et 3 (liste des personnes habitant dans le logement) puis en continuant par la page 4 (composition et équipement du logement). Ensuite, pour chacune des personnes inscrites dans la liste A de la page 2, vous remplirez un bulletin individuel (imprimé n° 2).

Vous devez répondre à une question. Lisez les explications qui figurent sur les imprimés, lorsque ces explications ne suffisent pas pour répondre à une situation particulière, reportez-vous à la notice explicative. En cas de difficulté, adressez-vous à l'agent recenseur.

En observant ces quelques recommandations et en suivant les indications qui vous sont fournies par l'agent recenseur, vous faciliteriez le travail de l'I.N.S.E.E. qui vous en remercie.

CADRE A REMPLIR PAR L'AGENT RECENSEUR

Logement occupé par M. _____

Adresse très précise _____

N° _____ Rue (ou hendid) _____

Commune _____

Département _____

Précisez la localisation du logement dans l'immeuble (escalier, étage, situation sur le palier, numéro du logement ou de la chambre, etc.): _____

CATEGORIE DE LOGEMENT

1 Logement ordinaire

2 Logement foyer pour personnes âgées

3 Pièce(s) indépendante(s) (exemple: chambre de domestique; louée(s) sous-louée(s) ou prêtée(s) à des particuliers. Indiquez le logement principal auquel-elle(s) se rattache(nt). _____

4 Chambres meublées dans un hôtel, une pension de famille, un garni, etc.

5 Habitation de fortune ou construction provisoire à usage d'habitation. Précisez: _____

6 Logement vacant (sans occupants, disponible ou non pour la vente ou pour la location). _____

7 Résidence secondaire ou logement loué (ou à louer) pour des séjours touristiques. _____

8 Caravane; roulotte; habitation mobile (ne pas remplir la page 4, ne pas établir de bordereau de maison; à classer à part).

5. vous êtes en présence d'un cas que vous hésitez à classer dans la liste ci-dessus, décrivez-le: _____

Dans le cas d'un ménage collectif, établissez une feuille de ménage collectif (imprimé n° 1bis)

1 AVEZ-VOUS UNE CUISINE ?

OUI → Quelle est sa surface?
 moins de 7 m² 1
 de 7 à 12 m² 2
 plus de 12 m² 3

NON → Avez-vous cependant dans votre logement une installation pour faire la cuisine?
 OUI 5
 NON 6

Ne répondez "OUI" que si cette installation comprend un évier permettant l'évacuation des eaux usées.

2 Indiquez, s'il y a lieu, le nombre de pièces de logement ayant un usage exclusif professionnel.

(Par exemple: bureau d'homme d'affaires, cabinet de médecin ou d'avocat, atelier de tailleur en appartement, etc.)

3 NOMBRE DE PIÈCES D'HABITATION non compris la cuisine ni les pièces à usage exclusivement professionnel.

Comptez comme pièces d'habitation les pièces telles que: chambre à coucher, salle à manger, salle de séjour, etc., quelle que soit leur surface, ainsi que les chambres de domestique et les mansardes habitables.

Ne comptez pas comme pièces d'habitation les pièces telles que: cuisine, couloir, salle de bains, armoire, W.-C., buanderie, etc.

Cas particulier des pièces indépendantes (par exemple, chambres de domestique séparées du logement proprement dit):

Comptez ces pièces parmi vos pièces d'habitation si vous en disposez vous-même (soit pour y loger un membre de votre ménage: enfant, domestique, etc.; soit comme débarras, etc.)

Ne les comptez pas si elles sont louées, sous-louées ou prêtées à d'autres personnes; dans ce cas, elles constituent, au sens du recensement, un logement distinct et leurs occupants rempliront une feuille de logement distincte.

4 ÊTES-VOUS:

Propriétaire de votre logement ou de la maison où se trouve votre logement? (y compris les différentes formes d'accession à la propriété, dont la location-vente et la location-attribution). 1

Logé par votre employeur (à titre gratuit ou onéreux) pour la durée de votre fonction ou de votre contrat de travail? 2

Logé à titre gracieux, par exemple par des parents? (y compris le cas des personnes occupant un logement qu'elles ont vendu en viager ou dont elles ont la jouissance par usufruit). 3

Locataire ou sous-locataire d'un local loué vide? 4

Locataire ou sous-locataire d'un local loué meublé, d'une chambre d'hôtel, d'un garni? 5

Si vous êtes dans un cas non prévu ci-dessus, décrivez-le: _____

5 SI VOUS ÊTES PROPRIÉTAIRE DE VOTRE LOGEMENT (réponse 1 à la question 4)

Pour l'achat de ce logement avez-vous fait un emprunt que vous n'avez pas encore entièrement remboursé? 1 OUI 0 NON

Répondez "OUI" en cas de location-vente ou de location attribution.

6 ALIMENTATION EN EAU

Eau courante dans le logement
 Eau chaude par installation individuelle (propre au logement) 1
 Eau chaude par installation collective 2
 Eau froide seulement 3

Pas d'eau courante dans le logement. 4

7 INSTALLATIONS SANITAIRES

Avez-vous une baignoire ou une douche installée avec eau courante et évacuation des eaux usées?
 Baignoire 1
 Douche 2
 Ni baignoire ni douche 3

8 CABINETS D'AISANCES (W.-C.)

Situés à l'intérieur du logement 1

Situés hors du logement mais réservés à ses seuls occupants
 avec chasse d'eau 2
 sans chasse d'eau 3

Autres cas 4

9 MODE DE CHAUFFAGE

Pour les maisons individuelles disposant du chauffage central cochez la case 1

Chauffage central individuel avec une chaudière propre à votre logement (y compris le chauffage électrique intégré et le chauffage par air pulsé) 1

Chauffage central collectif (commun à la totalité ou à la plupart des logements de l'immeuble) 2

Autres modes de chauffage (poêle, cuisinière, radiateur électrique à accumulation, etc.) quel que soit le combustible utilisé 3

10 COMBUSTIBLE UTILISÉ POUR LE CHAUFFAGE

Quel est le combustible (ou la source d'énergie) utilisé pour le chauffage de votre logement?
 Charbon 1
 Bois 2
 Mazout (fuel) 3
 Gaz 4
 Electricité 5
 Autres 6

11 TÉLÉPHONE

Est-il installé dans le logement? 1 OUI 0 NON

12 Les habitants du logement disposent-ils d'une ou plusieurs voitures de tourisme?

OUI 1
 NON 0

Combien? 1 / 2 ou plus 2

2 A remplir après la feuille de logement (imprimé n° 1). Cet imprimé sera rempli pour TOUTE PERSONNE inscrite dans la liste A de la feuille de logement.

Cadres à remplir par l'agent recenseur

Nombre du DISTRICT de recensement

Numero d'ordre de l'immeuble

Numero du logement (ou de l'habitation mobile, ou du ménage collectif)

Cadres de la Mairie

1 NOM ET PRÉNOMS
Ecrivez le nom en capitales, une femme ajoutera son nom de jeune fille (exemple MAURIN née ALLARD, Marie, Lucie).

2 ADRESSE

3 SEXE Masculin 1
Féminin 2

4 SITUATION DE FAMILLE Célibataire 1
Mariée 2
Veuf(ve) 3
Divorcé(e) 4

5 DATE ET LIEU DE NAISSANCE
Née le _____
(Jour, mois, année)

à (commune) : _____
(Pour Paris, Lyon, Marseille, précisez l'arrondissement)

6 NATIONALITÉ

- Français de naissance (y compris par réintégration) 1
- Devenu français par naturalisation, mariage, déclaration ou option 2
- Indiquez votre nationalité antérieure : _____
- Étranger 3
- Indiquez votre nationalité : _____

7 OU HABITIEZ-VOUS LE 1^{er} JANVIER 1975 ?
(Pour toute personne née avant le 1^{er} janvier 1975)

Si, le 1^{er} janvier 1975, vous étiez militaire ou élève interne ou en traitement dans un établissement de soins, indiquez l'adresse de votre résidence personnelle à cette date et non celle de l'établissement (caserne, internat, sanatorium, etc.)

- Dans le même logement que maintenant 1
- Dans la même commune (ou le même arrondissement pour Paris, Lyon, Marseille) 2
- Dans une autre commune (ou un autre arrondissement pour Paris, Lyon, Marseille) 3

Indiquez cette autre commune :

Commune : _____
(Pour Paris, Lyon, Marseille, précisez l'arrondissement)

Département : _____
(Pays pour l'étranger, territoire pour les T.O.M.)

POUR LES PERSONNES DE PASSAGE (voir le cadre C, page 3 de l'imprimé n° 1), adresse de la résidence habituelle :

N° _____ Rue (ou lieu-dit) : _____ Commune : _____ Département : _____
(Pour Paris, Lyon, Marseille, précisez l'arrondissement)

8 POUR TOUT ENFANT DE 2 A 14 ANS :
L'enfant va-t-il à l'école ? OUI 1
(y compris école maternelle) NON 0

POUR TOUTE PERSONNE DE 14 ANS OU PLUS

9 ÉTES-VOUS ACTUELLEMENT :

- Élève ou étudiant ? 1
- Stagiaire rémunéré ? (Stage pratique en entreprise ou stage de formation professionnelle, y compris A.F.P.A.) 2
- Dans une autre situation ? (y compris apprenti sous contrat, contrat emploi-formation, cours du soir ou cours par correspondance : C.N.T.E.) 3

10 INDIQUEZ TOUTS LES DIPLOMES QUE VOUS POSSEDEZ :

A ENSEIGNEMENT GÉNÉRAL PRIMAIRE OU SECONDAIRE :

- Certificat d'études primaires (C.E.P.), Diplôme de fin d'études obligatoires (D.F.E.O.) 1
- Brevet d'études du 1^{er} cycle (B.E.P.C.), Brevet élémentaire (B.E.) ou Brevet d'enseignement primaire supérieur (B.E.P.S.) 2
- Baccalauréat (1^{re} partie, probatoire ou 2^e partie), non compris les séries F, G et H, Brevets supérieurs; Certificat de fin d'études secondaires (C.F.E.S.) 3

B ENSEIGNEMENT TECHNIQUE ET FORMATION PROFESSIONNELLE DE NIVEAU SECONDAIRE :

- Certificat d'aptitude professionnelle (C.A.P.), Brevet d'enseignement professionnel (B.E.P.), Examen de fin d'apprentissage artisanal (E.F.A.A.), Brevets agricoles (B.A.A., B.P.A.), Certificat de fin de stage de la F.P.A. 1^{er} degré 1
- Brevet professionnel (B.P.), Brevet de maîtrise, Certificat de fin de stage de la F.P.A. 2^e degré 2
- Brevet d'enseignement agricole (B.E.A.), commercial (B.E.C.), hôtelier (B.E.H.), industriel (B.E.I.), social (B.E.S.); Brevet d'agent technique agricole (B.A.T.A.) 3
- Baccalauréat de technicien (Séries F, G ou H), Brevet de technicien (B.T., B.T.A.), Étève breveté des E.N.P. ou d'un lycée technique d'État, Brevet supérieur d'enseignement commercial (B.S.E.C.), Capacité en droit 4

C ENSEIGNEMENT SUPÉRIEUR :

- Diplôme des professions de la santé et des professions sociales, sage-femme, infirmière, kinésithérapeute, assistant(e) social(e), éducateur spécialisé 1
- Brevet de technicien supérieur (B.T.S.), Diplôme universitaire de technologie (D.U.T.), Diplôme d'études supérieures techniques (D.E.S.T.) 2
- Diplôme universitaire du 1^{er} cycle (Propédeutique, D.U.E.L., D.U.E.S., D.E.U.G., P.C.E.M. à l'exclusion des D.U.T.) Certificat de fin d'études normales, Certificat d'aptitude pédagogique 3
- Diplôme universitaire du 2^e ou 3^e cycle (licence, maîtrise, doctorat, etc.), C.A.P.E.S., C.A.P.E.T 4
- Diplôme de sortie d'une grande école publique ou privée, diplôme d'une école d'ingénieur 5
- AUTRE DIPLOME. Précisez : _____ 0

11 A QUEL ÂGE avez-vous cessé de suivre régulièrement les cours d'un établissement scolaire (y compris professionnel ou technique) ou universitaire ? _____ ans

• Vous exercez une activité professionnelle : répondez aux questions 12 à 15.
y compris : si vous aidez un membre de votre famille dans son travail, même à temps partiel.
(- si vous êtes apprenti sous contrat ou stagiaire rémunéré)

• Vous n'exercez pas actuellement d'activité professionnelle ou vous êtes en chômage : répondez à la question 16

12 Indiquez la profession ou le métier que vous exercez actuellement.
Soyez précis. Exemples : ouvrier électricien d'entretien, chauffeur de poids lourds, dessinateur d'études en électricité, ingénieur chimiste, vendeur en électroménager, employé de comptabilité, etc.

13 Aidez-vous un membre de votre famille dans son travail ? (Exploitation agricole, artisanale : commerce : profession libérale ; etc.) OUI 1
NON 0

13 Exercez-vous cette profession comme :

- Employeur ou travailleur indépendant (chef d'exploitation agricole ou coopérative, artisan, commerçant, industriel, membre d'une profession libérale, etc.) 1
- Aide familial non salarié (enfant ou autre membre de la famille d'un agriculteur, d'un commerçant, etc.) 2
- Apprenti sous contrat 3
- Salarié 4

Employez-vous des salariés ?
Ne comptez ni les apprentis ni les gens de maison.
Dans l'agriculture, comptez seulement les salariés permanents.

OUI Combien ? 1
 2
 3
 4
NON 0

Êtes-vous travailleur à domicile pour le compte d'une ou plusieurs entreprises ? OUI 1
NON 0

14 OÙ TRAVAILLEZ-VOUS ?

A ADRESSE de votre lieu de travail :
N° _____ Rue (ou lieu-dit) : _____
Commune et dépt : _____
(Pour Paris, Lyon, Marseille, précisez l'arrondissement)

B NOM (ou raison sociale) de l'établissement (industriel, commercial, administratif, agricole, etc.) qui vous emploie ou que vous dirigez : _____
Commune et dépt : _____
(Pour Paris, Lyon, Marseille, précisez l'arrondissement)

C ACTIVITÉ de cet établissement :
Soyez précis. Exemples : commerce de vins en gros, fabrication de charpentés métalliques, filature de coton, transport routier de voyageurs, etc.

D Adresse de cet établissement, si elle est différente de celle déclarée à la question 14 a.
N° _____ Rue (ou lieu-dit) : _____
Commune et dépt : _____
(Pour Paris, Lyon, Marseille, précisez l'arrondissement)

15 POUR LES SALARIÉS :

A Indiquez la catégorie professionnelle de votre emploi actuel :

- Ouvrier : manoeuvre ou manoeuvre spécialisé 1
ouvrier spécialisé (OS, OI, O2, O3...) 2
ouvrier qualifié (PI, P2, P3, TA, OP, OQ...) 3
- Employé 4
- Technicien, dessinateur 5
- Agent de maîtrise : dirigeant des ouvriers ou des employés 6
dirigeant des agents de maîtrise ou des techniciens 7
- Autres cas 8

Ingénieur ou cadre (les employés, techniciens, agents de maîtrise n'ayant pas la qualité de cadre ne devront pas se classer ici, même s'ils cotisent à une caisse de retraite des cadres)

Donnez toute précision complémentaire : position hiérarchique, coefficients, niveau, échelon. Exemples : chef d'équipe, mécanicienne en confection catégorie C, tourneur coefficient 155, grade classe IV, etc.

B Si vous êtes agent de l'État, d'une collectivité locale ou d'un service public (E.D.F., S.N.C.F., etc.) ou militaire de carrière, précisez votre grade. Exemples : contrôleur du Trésor, receveur P.T.T. de 4^e classe, agent des services hospitaliers

C Si vous êtes ingénieur, cadre, agent de maîtrise ou technicien, précisez votre fonction principale dans l'entreprise ou l'organisme qui vous emploie :

- Directeur général ou un de ses adjoints directs 1
- Fonction administrative, financière ou comptable 2
- Fonction commerciale ou technico-commerciale 3
- Production, fabrication, chantiers 4
- Entretien, travaux neufs, maintenance, dépannage 5
- Études, essais, méthodes, recherche 6
- Autres fonctions 7

Précisez (informatique, sécurité, santé...)

16 SI VOUS N'EXERCEZ PAS ACTUELLEMENT D'ACTIVITÉ PROFESSIONNELLE OU SI VOUS ÊTES EN CHÔMAGE :

A Êtes-vous :

- Mère de famille, femme au foyer 1
- Retraité, retiré des affaires (ancien commerçant, ancien agriculteur, etc.) 2
- Élève ou étudiant 3
- Chômeur 4
- Dans un autre cas ; précisez : _____ 5

Quelle a été votre profession principale ? _____

Avez-vous déjà travaillé ? OUI 1
NON 0

B Cherchez-vous actuellement du travail ? OUI 1
NON 0

Quel est votre métier ? _____

CHAPITRE 2

ETUDE STATISTIQUE PRELIMINAIRE :

CARACTERES DEMOGRAPHIQUES ET STRATIFICATION MORPHOLOGIQUE

Introduction

Dans l'ensemble des chapitres suivants, on va s'intéresser à l'estimation de statistiques très simples, attachées aux variables démographiques étudiées : totaux de caractères comme la population sans doubles comptes ou la population étrangère ou ratios entre deux caractères, qui décrivent la structure de ces populations comme, par exemple, le ratio population des hommes à la recherche d'un emploi / population des hommes de plus de 19 ans (1) .

Avant de passer à l'estimation proprement dite, ce chapitre préliminaire décrit "l'univers" dans lequel se déroulera le test : la base de sondage, la première stratification élaborée. Il donne un résumé des distributions de certains caractères tant sur la base entière qu' à l'intérieur de chacune des strates; enfin, il tente de fournir une première indication sur les gains à attendre de cette stratification en calculant les variances inter et intra-strate de ces caractères . Comme il serait fastidieux de répéter ici la description de tous les caractères considérés dans la suite et que, d'autre part, de fortes ressemblances existent entre les différentes distributions, nous nous limitons à l'examen de trois variables définies sur l'ensemble des districts INSEE : la population sans doubles comptes, la surface (2) et la densité de population sans doubles comptes.

(1) : Voir la définition des caractères démographiques au sens de l'INSEE au chapitre 1.

(2) : Bien que n'étant pas un caractère démographique, la surface des districts jouera par la suite un rôle très important dans la définition des plans de sondages utilisés.

2.1. Rappel sur la stratification et la base de sondage

La stratification considérée ici est une classification des districts INSEE qui composent la base de sondage d'après des critères de morphologie urbaine. Le travail d'analyse typologique n'a pas porté directement sur les districts INSEE, sur lesquels nous ne disposons pas de l'information morphologique exhaustive et précise. Cette information a été définie sur un ensemble de zones morphologiquement homogènes, déterminées lors de la photo-interprétation d'une mosaïque de photographies aériennes au 1/23000 (3).

Une fois obtenue une classification de cet ensemble de zones, le type morphologique de chaque zone a été affecté aux districts qu'elle contenait, ou, plus exactement, à ceux dont elle contenait la majorité de la surface. Bien sûr, l'homogénéité attribuée à une zone à l'échelle d'observation du 1/23000 ne se retrouve pas nécessairement lorsqu'on considère des échelles plus grandes. Ceci explique, par exemple, que l'on trouve dans des zones où la densité du bâti est dans l'ensemble très faible (classées dans la strate "non bâti"), certains districts de petite taille très peuplés.

La stratification des districts INSEE doit donc être manipulée avec prudence. On doit plutôt la considérer comme une information d'appartenance à une zone plus vaste que le district, de caractéristiques morphologiques globalement homogènes, que comme une donnée décrivant la morphologie du district lui-même. Ces précautions étant prises, on va constater que cette stratification opère des regroupements de districts intéressants du point de vue démographique.

Enfin, pour obtenir la base de sondage qui va nous intéresser ici, on a ramené le nombre des districts de la zone d'étude de 3534 à 3461, ceci à cause des exclusions suivantes :

1. 68 districts INSEE qui ne sont pas entièrement inclus dans la zone d'étude et dont on ne connaît que partiellement les caractéristiques morphologiques, qui constituent la strate 10.
2. cinq districts correspondant à des zones non décrites morphologiquement (chemin de fer, eau, digue, port), qui constituent la strate 11.

(3) : Pour plus de détails sur le travail ayant permis l'intégration des données et l'obtention de la stratification, se reporter aux annexes 6 et 7 du rapport ATP CNRS/CNES Télédétection spatiale, d'août 1987.

L'ensemble des résultats cités dans ce chapitre (ainsi que dans le suivant) ont été établis sur cet ensemble de 3461 districts. Par la suite, dans les trois derniers chapitre, la base de sondage subira une nouvelle modification lors de l'exclusion des districts de caractérisation morphologique mixte (c.a.d. ceux que leur situation géographique place à cheval sur des zones de morphologie différente) et des districts constitués de rond-points, terre-plein au milieu de voies larges ou autres zones pouvant être identifiées comme innoccupées dès l'interprétation de l'image satellite (on n'a donc supprimé, au cours de cette phase, que des districts dont la surface dépassait l'équivalent de quatre pixels T.M.). Ces suppressions, qui ramèneront le nombre de districts de la base à 2619, ont été effectuées parce qu'il n'y a pas de raison de penser que la base de sondage, constituée d'après l'image lors de la phase d'application, présente ce genre d'imperfections.

2.2. Description statistique des caractères suivant les strates

Dans les tableaux 1, 2 et 3 qui suivent, on donne successivement, pour la population sans doubles comptes, la surface et la densité des districts INSEE, les principaux paramètres statistiques des variables calculés sur l'ensemble de la base de sondage et pour chacune des strates morphologiques.

De l'examen des trois tableaux on peut tirer les conclusions suivantes quant à l'intérêt de la stratification obtenue et à son emploi dans la suite du programme.

La répartition de la population sans doubles comptes dans les différentes strates et les variations de la densité d'une strate à l'autre nous laissent espérer une certaine efficacité de la stratification sur la précision des estimations ; en effet, d'une part 77,7 % de la population est concentrée dans les strates 3,1 et 5, c'est à dire sur la réunion des grands ensembles d'habitation, du centre ville dense et du pavillonnaire. Ces trois strates ne totalisent que 58 % de la surface totale de la zone d'étude. Compte tenu des intitulés qui décrivent synthétiquement la morphologie des strates, ces résultats peuvent paraître triviaux, mais il est déjà rassurant, concernant l'efficacité supposée de la stratification, que la population soit bien majoritairement là où on l'attendait.

D'autre part, exception faite des strates 7 et 8, qui sont très marginales en surface (respectivement 3,3 et 0,9 % de l'ensemble de la zone) et qui regroupent à elles deux moins de 9% de la population totale, le sous-ensemble "peuplé" décrit plus haut regroupe bien les zones de densités les plus fortes : 3,02 h/are pour la strate 1, 1,30 h/are dans la strate 3 et 0,61 h/are dans la strate 5.

Tableau 1 : Surface des districts INSEE (statistiques d'ensemble et par strate.)

Strate	Somme (km ²)	part du total (%)	Minimum (m ²)	Maximum (m ²)	Moyenne (m ²)	Ecart type (m ²)
1 Centre ville dense	6,879	9,4	90	75573	6163	6764
2 Quartiers industriels	9,283	12,6	303	309038	3375	48576
3 Grands ensembles d'habitation	20,347	27,7	458	292264	33192	42330
4 Petits immeubles d'habitation	5,022	6,8	175	334615	34160	52976
5 Pavillonnaire	15,41	21,0	37	382529	16684	28879
6 Péri-urbain peu dense	6,710	9,1	1265	368221	75387	83243
7 Reliqua morphologique	2,432	3,3	747	113218	14306	15561
8 Dense: grands bâtiments élevés	0,664	0,9	341	76370	9486	13256
9 Non bâti	6,811	9,3	728	689231	119489	139181
Total	73,563	100	37	689231	21255	41682

Tableau 2 : Population sans doubles comptes des districts INSEE (statistiques d'ensemble et par strate.)

Strate	Somme	part du total (%)	Minimum	Maximum	Moyenne	Ecart type
1 Centre ville dense	207739	28,5	0	1034	186,15	156,31
2 Quartiers industriels	40680	5,6	0	1909	147,93	260,09
3 Grands ensembles d'habitation	264764	36,3	0	2964	431,92	416,40
4 Petits immeubles d'habitation	28750	3,9	0	1931	195,58	287,15
5 Pavillonnaire	93937	12,9	0	1460	101,66	141,33
6 Péri-urbain peu dense	11648	1,6	0	941	130,88	209,58
7 Reliquat morphologique	40534	5,6	0	1319	238,44	255,00
8 Dense: grands bâtiments élevés	23788	3,3	0	1358	339,83	287,36
9 Non bâti	17032	2,3	0	3407	298,80	549,40
Total	728872	100	0	3407	211,00	277,94

Tableau 3 : Densité de population sans doubles comptes dans les districts INSEE (statistiques d'ensemble et par strate.)

Strate	Minimum (h/are)	Maximum (h/are)	Moyenne (h/are)	Densité moyenne dans la strate (h/are)	Ecart type	Coéf. de var. (%)
1 Centre ville dense	0	261,12	4,338	3,020	8,516	196
2 Quartiers industriels	0	17,78	1,291	0,438	2,442	189
3 Grands ensembles d'habitation	0	102,66	2,313	1,301	4,668	202
4 Petits immeubles d'habitation	0	22,48	1,050	0,573	2,059	196
5 PAVILLONNAIRE	0	22,32	0,944	0,609	1,195	126
6 Péri-urbain peu dense	0	3,64	0,416	0,174	0,601	144
7 Reliquat morphologique	0	10,40	2,172	1,667	2,067	95
8 Dense: grands bâtiments élevés	0	12,60	4,695	3,583	2,838	60
9 Non bâti	0	3,01	0,342	0,250	0,576	168
Total	0	261,12	2,426	0,991	5,559	229

En résumé, la stratification morphologique permet donc d'isoler trois sous-ensembles principaux de densités bien différentes qui, chacun, regroupe une part importante de la population :

1. Les grands ensembles d'habitation : 36,6 % de la population sur 27,7 % de la surface, $d_3 = 1,301$ h/are.
2. Le centre ville dense : 28,5 % de la population sur 9,4 % de la surface. $d_1 = 3,020$ h/are.
3. Le pavillonnaire : 12,9 % de la population sur 21 % de la surface. $d_5 = 0,609$ h/are.

Il faut également remarquer que du point de vue de la densité de population, les strates morphologiques sont assez homogènes puisque la totalité des coefficients de variation à l'intérieur des strates sont nettement inférieurs au même coefficient sur l'ensemble de la base ($cdv = 229$). Cette différence est particulièrement importante dans le pavillonnaire ($cdv_5 = 126$), les strates peu denses (péri-urbain, $cdv_6 = 144$, et non bâti, $cdv_8 = 168$) et surtout dans les deux strates "marginales" 7 et 8 : respectivement, $cdv_7 = 95$ et $cdv_8 = 60$. En revanche, dans les quatre premières strates, la densité de population est plus variable.

Du point de vue de la stratégie de sondage, un tel résultat nous porte à espérer un gain important de précision dû à l'emploi d'un plan stratifié, ceci grâce à la variance inter-strate des caractères démographiques.

Par ailleurs, en considérant les colonnes qui donnent l'écart-type des caractères "population sans doubles comptes" et "densité de population sans doubles comptes", on note des différences importantes suivant les strates : en ce qui concerne l'effectif, les strates "pavillonnaire" et "centre ville dense" ont des écarts types très faibles (resp. 141 et 156,3) alors que les strates "non bâti" et "grands ensembles d'habitation" sont celles où la variation du caractère est la plus forte (resp. 549,4 et 416,4). Pour la densité, les différences sont plus importantes encore entre, d'une part, les strates peu denses, où les écarts types sont faibles : "non bâti", $\sigma_8 = 0,58$, et "péri-urbain peu dense", $\sigma_6 = 0,60$, et, d'autre part, les strates de densité plus forte où la variance est importante : "centre ville dense", $\sigma_1 = 8,5$ et "grands ensembles d'habitation", $\sigma_3 = 4,7$.

Ainsi, il apparaît probable qu'une optimisation de l'allocation à chaque strate soit elle aussi payante grâce aux différences importantes de variance intra-strate des

caractères d'une strate à l'autre (l'allocation optimale de Neyman est en effet proportionnelle à l'écart-type intra-strate (4)).

On le voit, les gains de précision qu'on peut attendre de la stratification d'abord, puis de l'optimisation de l'allocation, peuvent être prévus et quantifiés grâce au calcul des variances inter et intra-strate du caractère à estimer. Rappelons d'abord rapidement les définitions et les principaux résultats qui sous-tendent ces calculs.

2.3. Variance inter-strate et intra-strate des caractères, tentative de prévision du gain apporté par les stratification

Il est classique en taxinomie mathématique ou en classification automatique (5) de caractériser une partition Q d'un ensemble I par le taux d'inertie, $t(Q)$, ou de variance, $t(V)$, dont elle rend compte. Pour préciser ces notions, nous adoptons les notations suivantes qui sont celles de J.P. BENZECRI :

- I est une partie finie d'un espace euclidien E
- $\|i-i'\|^2$ est le carré de la distance euclidienne entre deux éléments i et i' de I
- $m_i = \{m_i / i \in I\}$ est un système de masses positives sur I
- $m = \sum_{i \in I} m_i$ est la masse totale de I
- On note $g(I) = g = 1/m \cdot \sum_{i \in I} m_i \cdot i$, le centre de gravité du système.

Le moment centré d'ordre 2 ou moment d'inertie du système est :

$$M^2(I) = \sum_{i \in I} m_i \cdot \|i-g\|^2$$

(4) : Voir J. DESABIE : Théorie et pratique des sondages (DUNOD 1966) P. 145 : L'échantillon de NEYMAN ou échantillon optimal.

(5) : Voir par exemple : J.P. BENZECRI et Collaborateurs : L'analyse des données, 1. La taxinomie, P. 186 et 187.

Soit Q une partition de I en classes non vides deux à deux disjointes, et q le centre de gravité de la classe C_q , tandis que $mc_q = \sum_{i \in C_q} m_i$ est la masse totale de la classe C_q . le moment

d'ordre 2 :

$$M^2(Q) = \sum_{C_q \in Q} mc_q \cdot ||q-g||^2, \text{ est l'inertie inter-classe de la}$$

partition. La variance inter-classe est :

$$V^2(Q) = \sum_{C_q \in Q} (mc_q/m) \cdot ||q-g||^2.$$

Pour la classe C_q , on a de même : $M^2(C_q) = \sum_{i \in C_q} m_i \cdot ||i-q||^2$

$$\text{et } V^2(C_q) = \sum_{i \in C_q} (m_i/mc_q) \cdot ||i-q||^2$$

On montre alors facilement (théorème de Huygens) que :

$$M^2(I) = M^2(Q) + \sum_{C_q \in Q} M^2(q),$$

ou, ce qui revient au même, que :

$$V^2(I) = V^2(Q) + \sum_{C_q \in Q} (m_q/m) \cdot V^2(q)$$

Cela signifie en clair que la variance totale $V^2(I)$ est égale à la somme de la variance entre les classes $V^2(Q)$ (ou variance inter-classe), et de la moyenne pondérée par les m_q des variances intérieures aux classes (variance intra-classe).

Le taux de variance qui caractérise la partition est alors :

$$t(Q) = V^2(Q)/V^2(I) = \text{var. inter-classe} / \text{var. totale}$$

L'application au problème qui nous intéresse est simple. L'espace euclidien E est la droite réelle sur laquelle l'ensemble I est l'ensemble des valeurs prises par le caractère dont on cherche à estimer le total ou la moyenne (population sans doubles comptes ou densité) sur l'ensemble des districts INSEE. Le système de masses quant à lui, est défini par la probabilité de tirage de chaque district dans le plan de sondage considéré. La partition Q est pour nous la stratification.

Bien sûr l'éventail des plans de sondage utilisables pour l'estimation d'effectifs de population est assez vaste, compte tenu des possibilités de choix dans la définition des unités spatiales à enquêter, des probabilités d'inclusion dans l'échantillon et de la stratification ; le formulaire des

variances inter et intra-classe s'écrit cependant très facilement dans le cas du plan équiprobable sans remise. C'est dans ce cadre que nous nous plaçons maintenant.

On a alors $m_i = 1$ pour tout i et, si N est le cardinal de la base :

$$m = N, g(I) = g = 1/N \sum_{i=1}^N y_i = \bar{y}, V(I) = (1/N) \sum_{i=1}^N (y_i - \bar{y})^2 = \sigma_y^2$$

où y_i est la valeur du caractère y pour le district i , \bar{y} la moyenne de y sur la base, σ_y^2 est donc la variance classique de y sur la base.

De même, le centre de gravité d'une strate h , de cardinal N_h , est la moyenne de y dans la strate :

$$y_h = (1/N_h) \sum_{i=1}^{N_h} y_i,$$

et la variance interne à la strate :

$$V_y(h) = \sigma_{y,h}^2 = (1/N_h) \sum_{i=1}^{N_h} (y_i - y_h)^2$$

L'égalité entre variance totale et somme des variances inter-strate et intra-strate s'écrit, pour un caractère y donné :

$$V_y(I) = \sigma_y^2 = \sum_{h=1}^K (N_h/N) \cdot (y_h - \bar{y})^2 + \sum_{h=1}^K (N_h/N) \cdot \sigma_{y,h}^2$$

où K est le nombre de strates. Les résultats des calculs, effectués pour les deux caractères, population sans doubles comptes et densité de cette même population, sont consignés dans les tableaux 4 et 5.

L'observation des deux tableaux permet de faire les remarques suivantes concernant la suite du test, c'est à dire la comparaison des variances d'estimation proprement dite.

1. Pour chacun des deux caractères, on calculera donc les variances des estimateurs, avec et sans stratification (estimateur du total pour l'effectif et de la moyenne pour la densité de population). On devrait, d'après les deux colonnes qui donnent les variances inter-strates, obtenir un gain significatif en stratifiant l'estimateur du total de population; mais, en revanche, l'estimateur de la moyenne des densités ne profitera que très peu de la stratification : 20 % de la variance de l'effectif est due à la variance inter-strates alors qu'elle ne représente que 5 % de la variance totale de la densité de population.

Tableau 4 : Variances inter-strate et intra-strate de la population sans doubles comptes des districts INSEE.

Strate	N_h/N	$\sum p_h$	$\sum \frac{N_h}{N} (p_h - p)^2$ (var. inter.)	$\sum \sigma^2_{p,h}$	$\sum \frac{N_h}{N} \sigma^2_{p,h}$ (var. intra.)	% de var. intra.
1 Centre ville dense	0,322	186,15	199,18	24431	7877,96	12,3
2 Quartiers industriels	0,079	147,93	316,09	67648	5375,14	8,4
3 Grands ensembles d'habitation	0,177	431,92	8643,90	173387	30709,81	7,8
4 Petits immeubles d'habitation	0,042	195,58	10,10	82457	3502,24	5,5
5 Pavillonnaire	0,267	101,66	3191,56	21705	5794,89	9,0
6 Péri-urbain peu dense	0,026	130,88	165,09	43923	1129,49	1,8
7 Reliquat morphologique	0,049	238,44	36,97	65025	3193,95	5,0
8 Dense: grands bâtiments élevés	0,020	339,83	335,68	82574	1670,09	2,6
9 Non bâti	0,016	298,81	126,98	301840	4971,06	7,7
Total	1,000		13025,54 (16,86 %)		64224,62 (83,14 %)	100

Tableau 5 : Variance inter-strate et intra-strate de la densité de population sans doubles comptes des districts.

Strate	N_h/N	d_h	$\frac{N_h}{N} \frac{(d_h - d)^2}{N}$ (var. inter.)	$\sigma^2_{d,h}$	$\frac{N_h}{N} \sigma^2_{d,h}$ (var. intra.)	% de var. intra.
1 Centre ville dense	0,322	3,020	1,328	72,522	23,385	81,57
2 Quartiers industriels	0,079	0,438	0,024	5,963	0,474	1,65
3 Grands ensembles d'habitation	0,177	1,301	0,017	21,791	3,860	13,46
4 Petits immeubles d'habitation	0,042	0,573	0,007	4,239	0,180	0,63
5 Pavillonnaire	0,267	0,609	0,039	1,427	0,381	1,33
6 Péri-urbain peu dense	0,025	0,174	0,017	0,361	0,009	0,03
7 Reliquat morphologique	0,049	1,667	0,022	4,274	0,210	0,73
8 Dense: grands bâtiments élevés	0,020	3,582	0,136	8,054	0,163	0,57
9 Non bâti	0,016	0,250	0,009	0,331	0,005	0,02
Total	1,000		1,600 (5,29%)		28,667 (94,71%)	100

Soulignons que cette conclusion n'est valable que dans le cadre du plan de sondage équiprobable sans remise, avec un estimateur sans biais qui ne fasse pas intervenir la surface des districts INSEE, et seulement dans le cas de l'allocation proportionnelle.

2. On recommencera les calculs de variances dans le cas stratifié en adoptant l'allocation optimale de NEYMAN, le taux dans chaque strate sera proportionnel à l'écart-type du caractère dans la strate.

Comme on pouvait s'y attendre, la contribution des strates à la somme des variances intra-strate est très variable (voir les colonnes " % de variance intra-strate" des tableaux 4 et 5), et ceci, particulièrement dans le cas de la densité de population sans doubles comptes. Donc ici, on s'attend à ce que l'allocation optimale profite beaucoup à l'estimation de la moyenne de cette densité.

CHAPITRE 3

MISE AU POINT DES OUTILS ET DE LA METHODE DU TEST.

Sondages équiprobables, estimations sans biais d'effectifs et de densités, stratification morphologique syntétique.

Introduction

Avant d'élargir le champ de l'étude en modifiant les plans de sondages, les estimateurs et les caractères démographiques (voir chapitre 5) et de considérer d'autres stratifications (chapitres 6 et 7), dans ce qui va suivre, nous rendons compte des étapes de la mise au point du test : il sera question successivement du plan de sondage, des taux, puis nous donnerons le formulaire simple correspondant à l'estimation sans biais dans le cas du plan équiprobable, enfin nous examinerons les premiers résultats obtenus à partir des chiffres du recensement général de la population à Marseille en 1982.

3.1. Base et plan de sondage, caractéristiques des échantillons

La base de sondage est la réunion des neuf premières strates morphologiques. Dans ce qui suit, on considère des échantillons obtenus par tirage systématique dans la liste des districts INSEE. Ce mode de tirage sera considéré, dans un premier temps, comme une approximation d'un plan équiprobable sans remise (1).

Pour le plan stratifié qui a servi au tirage de l'échantillon de travail (2), la même procédure a été appliquée pour chaque

(1) : Pour une justification de cette approximation, voir par exemple, J. DESABIE : Théorie et pratique des sondages (DUNOD 1966), ou C. GOURIEROUX : Théorie des sondages (ECOMICA 1981, P. 115).

(2) : A propos de l'échantillon de travail, son utilité dans le cadre du programme et la technique de tirage utilisée, voir l'annexe 10 du rapport ATP CNRS/CNES d'Août 1987.

strate; la base de la suite arithmétique de raison 44, correspondant à un taux approximatif de 2,31%, est choisie aléatoirement et indépendamment dans chaque strate. L'assimilation à un tirage équiprobable sans remise étant admise, le plan stratifié qui résulte de ce tirage peut être vu sous deux angles :

- il s'agit d'un sondage à deux degrés si l'on considère les districts INSEE comme individus statistiques. Les unités primaires sont les strates et le tirage au premier degré est un recensement. Au second degré, on a tiré les districts INSEE selon un plan équiprobable sans remise au taux approximativement constant de $1/44$ (les variations éventuelles du taux dans les strates provenant des choix aléatoires des premiers districts tirés dans chaque strate).

- comme un plan à trois degrés dont les unités primaires sont les strates, les unités secondaires les districts et les unités tertiaires les unités de surface élémentaires (ares ou m^2) constituant les districts. Les tirages aux premier et troisième degrés sont alors des recensements et il s'agit d'un tirage stratifié de grappes (3) .

Les résultats sont, bien sûr, équivalents dans les deux cas. Cette distinction apparaît donc purement formelle et, dans le formulaire comme dans la pratique, on utilisera directement les formules stratifié sans faire référence aux degrés du sondage.

Soulignons enfin que, lors du tirage effectué pour l'obtention de l'échantillon de travail, la liste était triée sur l'identifiant alphanumérique attribué par l'INSEE à chaque district. Comme le remarque F. DUREAU dans une note sur le tirage de cet échantillon (4), ce tri assure une bonne répartition géographique des districts tirés au sein de l'aire urbaine étudiée. En effet, l'identifiant INSEE regroupe les districts quartier par quartier, puis arrondissement par arrondissement. Il est clair, pour les urbaniste et démographe de l'équipe, que ce classement n'est "neutre", ni du point de vue morphologique, ni du point de vue démographique ; des deux points de vue, les districts qui sont de rangs voisins dans la liste, auront tendance à se ressembler, autrement dit, ce classement apporte certainement une information sur les caractères que nous cherchons à estimer. Dans ce cas, on sait que l'assimilation du tirage systématique à un tirage aléatoire équiprobable sans remise, conduit à une sur-estimation des variances des estimateurs qui peut être

(3) : Voir C. GOURIEROUX, Théorie des sondages, P. 122.

(4) : Voir F. DUREAU 1986 - Marseille : tirage de l'échantillon de travail. Quito, 9 p.

importante (voir DESABIE : théorie et pratique des sondages, p 99, qui prend d'ailleurs le même exemple, ou GOURIEROUX, p 115). L'évaluation de cette sur-estimation est possible par assimilation du tirage systématique, dans l'ensemble de la base ou à l'intérieur de chaque strate, à un tirage stratifié de paires (voir DESABIE p 171 ou GOURIEROUX p 117). Mais, comme nous l'avons déjà dit, une procédure de tirage sur liste ne peut pas être à la base de la méthode applicable aux cas concrets de plans de sondage bâtis d'après images satellite. Il ne nous semble donc pas utile de mener ici plus avant l'évaluation du gain de précision qu'amène le tri sur l'identifiant INSEE. Nous nous contenterons de deux remarques :

1. l'information amenée par ce tri est certainement en partie redondante avec celle qu'apporte la stratification morphologique. On peut donc raisonnablement supposer que si les variances des estimations non stratifiées sont largement sur-estimées, à l'inverse, les variances des estimations stratifiées sont, elles, à peu de chose près réalistes. Il se pourrait toutefois que certaines strates géographiquement dispersées et démographiquement hétérogènes, profitent également du tri.

2. Dans le cas concret d'application, un problème semblable se posera pour évaluer l'apport du tirage aréolaire sur grille de points, qui, lui aussi, assure une bonne répartition géographique de l'échantillon. Il sera donc de nouveau question de ce problème dans l' chapitre 7 de ce rapport.

Remarque : Le plan de sondage P, défini ci-dessus, respecte les surfaces en espérance, c'est à dire au sens suivant :

$$E_P \left(\sum_{i=1}^n s_i \right) = t.S$$

où $E_P()$ désigne l'espérance mathématique calculée sous la loi de probabilité définie par P, s_i les surfaces des districts de l'échantillon, t le taux de sondage, n le nombre de districts tirés et S la surface totale de la base de sondage.

$$\begin{aligned} \text{En effet,} \quad E_P \left(\sum_{i=1}^n s_i \right) &= E_P \left(\sum_{i=1}^N s_i \cdot 1(i \in e) \right) \\ &= \sum_{i=1}^N s_i \cdot E_P(1(i \in e)) = \sum_{i=1}^N s_i \cdot z_i \end{aligned}$$

Où N est le cardinal de la base de sondage, $1(i \in e)$ désigne la fonction indicatrice de l'appartenance du district i à l'échantillon (noté ici e) et $z_i = E_P(1(i \in e))$, la probabilité d'inclusion du district i dans l'échantillon (toujours sous la loi du plan P).

Le plan étant équiprobable sans remise, on a $z_i = n/N = t$.

$$D' \text{ où } E_p \left(\sum_{i=1}^N s_i \right) = \sum_{i=1}^N s_i \cdot t = t \cdot \sum_{i=1}^N s_i = t \cdot S .$$

3.2. Taux de sondage

Comme on le vérifiera tout au long des formulaires qui suivent, pour l'ensemble des estimateurs étudiés, la variance comprend toujours, dans son expression, un facteur $(N-n)/n$ (ou $(N_h - n_h)/n_h$ dans les expressions stratifiées), qui représente l'influence du taux de sondage sur les variances d'estimation. Il est donc très simple d'étudier les gains de précision dûs aux augmentations du taux pour chacun des estimateurs que nous proposons dans la suite, la multiplication par le facteur

$$\frac{t_1(1-t_2)}{t_2(1-t_1)}$$

permettant de passer de la variance d'une estimation

au taux $t_1 = n_1/N$ à la variance au taux $t_2 = n_2/N$.

L'hyperbole de la figure 1 donne la valeur du facteur $(N-n)/n$ en fonction du taux n/N (valable pour toute valeur de N). Sur la figure 2, le facteur $(N-n)/n$ est rapporté à la valeur qu'il a au taux de 1% et exprimé en % ; elle permet d'évaluer comment la variance des estimations va décroître dans la région de variation du taux qui nous intéresse. Enfin le tableau 3 donne la variance pour les taux de 4, 5, 6% en fonction de la variance V obtenue au taux de 2,31% (80/3461) qui correspond à l'échantillon de travail.

3.3. Formulaire du plan équiprobable sans remise

Les résultats suivants, classiques en théorie des sondages (voir par exemple DESABIE ou GOURIEROUX) sont rappelés sans démonstration. Les notations sont les suivantes : la variable estimée est toujours notée y , les valeurs qu'elle prend sur les unités statistiques y_i , son total Y , la variable définissant la probabilité z et ses valeurs z_i .

3.3.1. Sans stratification

Le résultat général utilisé ici est le suivant : y étant le caractère à estimer et les probabilités d'inclusion z_i étant fixées, les seuls estimateurs sans biais du total et de la moyenne de Y , de la forme : $\sum_{i \in e} a_i y_i$, sont ceux dit d'Horwitz - Thompson.

Fig. 1

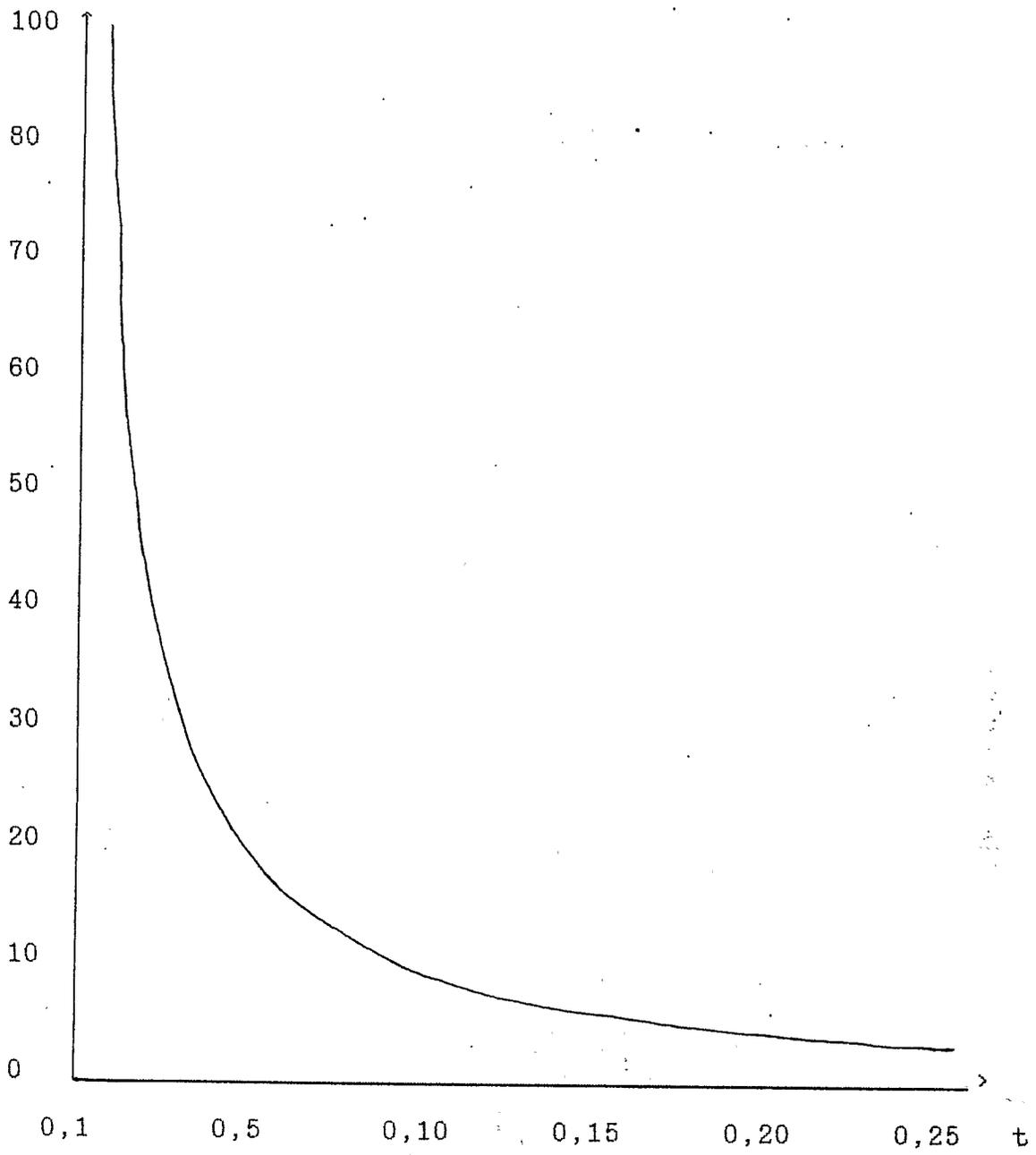


Fig 2

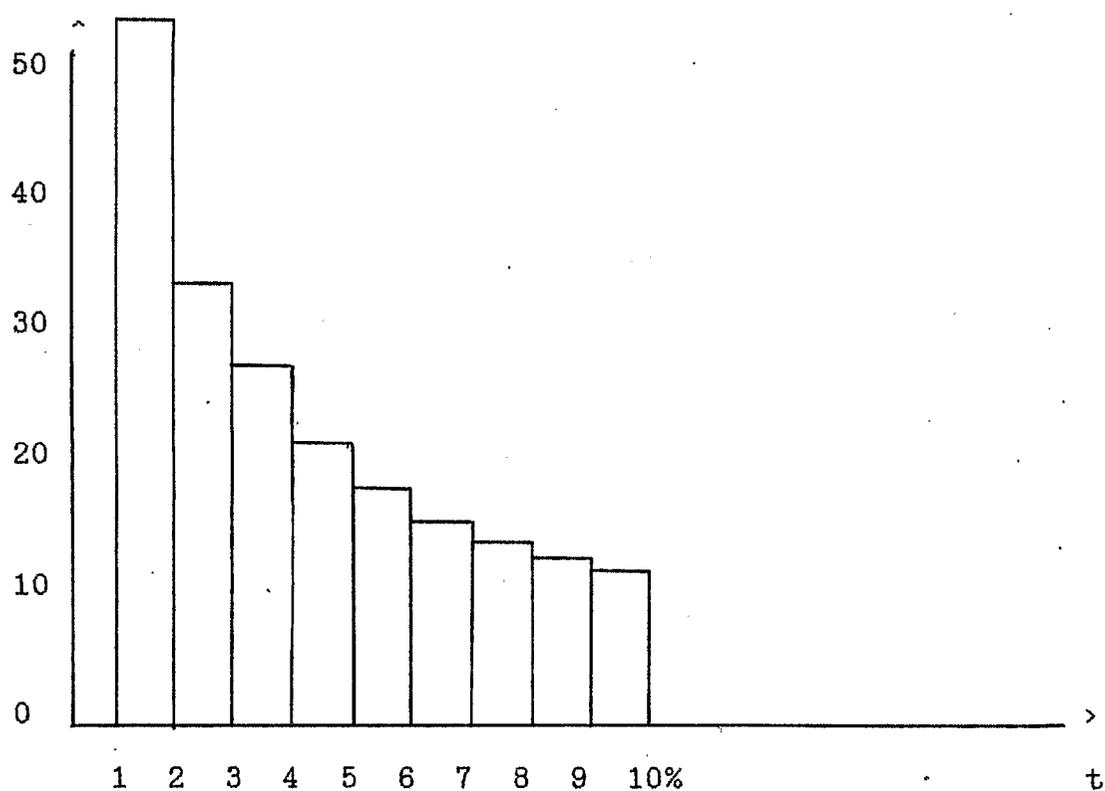


tableau 3

taux	2,31%	4%	5%	6%
variances	V	0,5675.V	0,4493.V	0,3705.V

Par exemple, l'estimateur du total de y , est donné par :

$$\hat{T}(Y) = \sum_{i \in e} \frac{y_i}{z_i} = \sum_{i=1}^N (y_i / z_i) \cdot (1(i \in e)),$$

où e est l'échantillon, N le cardinal de la base de sondage et z_i , la probabilité d'appartenance du district i à l'échantillon. Le plan étant pour l'instant équiprobable, on a $z_i = n/N$ pour tout i et par conséquent, les estimateurs et leurs variances ont les expressions suivantes (5):

- estimation du total

$$\hat{T}(Y) = N \sum_{i=1}^n y_i / n = N \cdot \bar{y} \quad V(\hat{T}(Y)) = \frac{N \cdot (N-n)}{n} \cdot S^2$$

- estimation de la moyenne

$$\hat{M}(Y) = \sum_{i=1}^n y_i / n = \bar{y}, \quad V(\hat{M}(Y)) = \frac{N-n}{n \cdot N} \cdot S^2,$$

où \bar{y} désigne donc la moyenne empirique du caractère sur l'échantillon, $V()$ la variance sous la loi du plan P , S^2 la variance à $N-1$ degrés de liberté sur l'ensemble de la base :

$$S^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i - \bar{Y})^2 = (N-1) \cdot \sigma^2 / N,$$

σ^2 étant la variance classique sur l'ensemble de la base.

Application : estimation de l'effectif total de population et de la moyenne des densités.

Les quantités à estimer sont donc : $P = \sum_{i=1}^N p_i$, total de

population sur l'ensemble de la base et $\bar{d} = \frac{1}{N} \sum_{i=1}^N d_i$,

moyenne sur l'ensemble de la base des densités de districts.

- Total des effectifs : P est estimé par

$$\hat{T}(P) = N \cdot (1/n) \sum_{i=1}^n p_i = N \cdot \bar{p}, \quad \text{où } p_i \text{ désigne l'effectif de}$$

population sans doubles comptes du district courant i et \bar{p} la moyenne empirique de cet effectif sur l'échantillon.

$N \cdot \bar{p}$ estimera donc sans biais l'effectif total et ceci avec la variance :

(5) : Voir, par exemple, C. GOURIEROUX, Théorie des sondages, chap. III, P. 77 à 81.

$$\hat{V}(T(P)) = \frac{N^2 \cdot (N-n)}{(N-1) \cdot n} \cdot \sigma_p^2 \quad \text{où } \sigma_p^2 \text{ est la variance de}$$

l'effectif de population sans doubles comptes sur l'ensemble de la base :

$$\sigma_p^2 = \frac{1}{N} \sum_{i=1}^N (p_i - \bar{p})^2 \quad \text{où } \bar{p} = \frac{1}{N} \sum_{i=1}^N p_i$$

Le plan étant équiprobable sans remise et l'estimateur celui d'Horwitz-Thompson, la loi de l'estimateur est asymptotiquement normale (théorème central limite) ce qui permet de construire l'intervalle de confiance à 95% de l'estimation du total :

$$I_p = [T(P) - 2N \left(\frac{N-n}{n \cdot (N-1)} \cdot \sigma_p^2 \right)^{1/2}, T(P) + 2N \left(\frac{N-n}{n \cdot (N-1)} \cdot \sigma_p^2 \right)^{1/2}]$$

- Moyenne de densités des districts

La moyenne de densités des districts est estimée sans biais par :

$$\hat{M}(\bar{d}) = \frac{1}{n} \cdot \sum_{i=1}^n d_i$$

La variance sera :

$$\hat{V}(M(\bar{d})) = \frac{N-n}{n \cdot (N-1)} \cdot \sigma_d^2, \quad \text{où } \sigma_d^2 = \frac{1}{N} \cdot \sum_{i=1}^N (d_i - \bar{d})^2$$

D'où l'intervalle de confiance à 95% :

$$I_a = [\bar{d} - 2 \left(\frac{N-n}{n \cdot (N-1)} \cdot \sigma_d^2 \right)^{1/2}, \bar{d} + 2 \left(\frac{N-n}{n \cdot (N-1)} \cdot \sigma_d^2 \right)^{1/2}]$$

3.3.2 Avec stratification

- Effectifs

L'estimateur du total s'écrit :

$$\hat{T}(P) = \sum_{h=1}^K \frac{N_h}{n_h} \cdot \left(\sum_{i=1}^{n_h} p_{h,i} \right), \quad \text{où } K \text{ est le nombre de}$$

strates, N_h le cardinal de la strate h et n_h le nombre de districts tirés dans la strate h .

Cet estimateur a pour variance :

$$V(T(P)) = \sum_{h=1}^K \frac{N_h(N_h - n_h)}{n_h(N_h - 1)} \cdot \sigma_{P,h}^2,$$

où $\sigma_{P,h}^2 = \frac{1}{N_h} \sum_{i=1}^{N_h} (p_{h,i} - \bar{p}_h)^2$, \bar{p}_h étant la moyenne du

caractère "population sans doubles comptes" dans la strate h , $\sigma_{P,h}^2$ sa variance sur la strate h .

L'intervalle de confiance à 95% s'écrit :

$$I = [T(P) - 2 \cdot V(Th.t.(P))^{1/2}, T(P) + 2 \cdot V(Th.t.(P))^{1/2}]$$

- Moyenne des densités

Estimateur

$$M(\bar{d}) = \sum_{h=1}^K (N_h/N) \cdot \sum_{i=1}^{n_h} d_{h,i}/n_h$$

Variance de l'estimateur

$$V(M(\bar{d})) = \sum_{h=1}^K \frac{N_h(N_h - n_h)}{N^2(N_h - 1)n_h} \cdot \sigma_{d,h}^2$$

Intervalle de confiance à 95%

$$I = [\bar{d} - 2 \cdot V(M(\bar{d}))^{1/2}, \bar{d} + 2 \cdot V(M(\bar{d}))^{1/2}]$$

3.3.3 Allocation optimale de NEYMAN

L'allocation proportionnelle retenue lors du tirage de l'échantillon de travail (taux uniforme dans toutes les strates) a été choisie pour que, dans l'échantillon, tous les types morphologiques figurent avec une surface approximativement proportionnelle à celle qu'ils occupent dans l'univers. Cet objectif n'est pas compatible avec l'optimisation de l'estimation des caractères démographiques.

Pour l'ensemble de la base de sondage et pour chacune des strates, les résultats du test mené sont consignés dans les tableaux 4.1 (estimation du total de population sans doubles comptes) et 4.2 (estimation de la moyenne du caractère "densité de la population s.d.c."). Le commentaire qui suit s'appuie sur ces tableaux qu'on peut consulter à la fin du chapitre.

3.4.1. Résultats sur l'ensemble de la base

Les résultats sur l'ensemble de la base appellent les commentaires suivants.

1. Pour l'ensemble des plans testés, stratifiés ou non, l'estimateur du total de population est toujours nettement plus précis que celui de la moyenne des densités. Ce fait est en accord avec la différence importante de variabilité des deux caractères sur l'ensemble de la base : la population sans doubles comptes des districts a un coefficient de variation de 132 % alors que pour la densité, il est de 229 %. Nous reviendrons sur cette question lors de l'examen des résultats par strate, mais on peut dire tout de suite que l'impératif qu'a l'INSEE, de constituer des districts dont la population ne dépasse pas un certain seuil, a tendance à diminuer artificiellement la variance de l'effectif de population des districts, surtout dans les endroits très densément peuplés. On ne peut donc probablement pas conclure de ce résultat qu'un estimateur du total de population par l'effectif soit toujours préférable à un estimateur par la densité.

2. L'apport de la stratification est intéressant pour les deux caractères étudiés : au taux de 5 %, il permet de faire diminuer la variance de l'estimateur du total de population de 16,6 % en adoptant une allocation proportionnelle, et de 31,8 % en optimisant l'allocation. Quant à l'estimation de la densité moyenne, elle profite beaucoup moins de l'allocation proportionnelle (-5,6 %) mais beaucoup plus de l'allocation optimale (-38,9 %). Au total, les gains sont donc comparables, dans leur ordre de grandeur, pour les deux caractères, mais différemment répartis entre allocation proportionnelle et optimale. Reprenons plus en détail l'examen des gains dus à la stratification et à l'augmentation du taux pour chacun des caractères.

3. Estimation du total de population

Quelle que soit la fraction de sondage, la variance de l'estimation obtenue sans stratification est multipliée par un facteur 0,834 lors de l'introduction de la stratification à allocation proportionnelle, par un facteur 0,682 si l'on choisit l'allocation de NEYMAN. Ceci permet de ramener le coefficient de variation de 10,0 % (sans stratification) à 8,3 % (allocation optimale) au taux de 5%. L'intervalle de

confiance est ramené de +/- 146270 à +/- 120759, soit un gain de 17 % .

Les gains dus à la variation du taux de sondage sont, eux, beaucoup plus importants, surtout dans l'intervalle 2,31 % à 5 % (voir fig. 1 et 2). Par contre, au passage de 5 à 6 %, le gain est du même ordre que celui dû à la stratification.

4. Estimation de la moyenne des densités

Ici le gain qu'apporte la stratification est presque entièrement dû à l'introduction de l'allocation optimale : La variance est multipliée par 0,929 quand on emploie l'allocation proportionnelle, puis par 0,615 dans le cas de l'allocation optimale, soit 80 % du gain lors de son introduction.

Au taux de 5 %, le coefficient de variation passe de 17,42 % sans stratification à 13,66 % avec l'allocation optimale.

En conclusion, l'estimation de la moyenne des densités profite donc un peu mieux de la stratification que celui de l'effectif total.

5. Au taux où s'est réellement fait le tirage (80/3461), l'échantillon obtenu fournit des estimations beaucoup plus précises que ne laissent prévoir les intervalles de confiance à 95 %. L'estimation de la densité ne s'écarte que de 21,4 % de la vraie valeur tandis que le total de population est obtenu à 0,073 % près (estimations stratifiées). On a donc eu beaucoup de chance lors du tirage, en particulier du point de vue de l'estimation du total de population sans doubles comptes. Il est, bien sûr, exclu d'en tirer des conclusions quant à l'efficacité de la méthode, puisque seules les variances des estimateurs sont réellement significatives. Notons tout de même qu'une part de l'explication de ce résultat "inespéré" peut résider dans le fait que le calcul de la variance ne prend pas en compte le tri sur l'identifiant INSEE, mais nous verrons plus tard que le gain que l'on peut en attendre n'est pas de cet ordre de grandeur.

3.4.2. Résultats par strates

Les commentaires des résultats par strates tiennent en quatre points principaux .

1. Naturellement, la précision des estimations à l'intérieur des strates est partout très nettement inférieure à celle des estimations sur l'ensemble de la base. Le Tableau 5 ci-dessous compare, pour les deux caractères et dans le cas de l'allocation proportionnelle, la précision des estimations obtenues sur l'ensemble de la base à celle à l'intérieur des strates : le coefficient de variation à l'intérieur des strates est compris entre 1,1 fois (estimation de la moyenne de densité dans la strate pavillonnaire) et 9,7 fois (estimation du total de population dans la strate non bâti) le coefficient de variation sur l'ensemble de la base.

Tableau 5 : coefficients de variation pour les estimations globales et les estimations par strate.

Estimateurs	Estimat- -ions globales	Estimation par strates	
		Mini.	Maxi.
1. Du total de population sans doubles comptes	9,17 %	11,25 %	109,88 %
2. De la moyenne de densité de population	16,79 %	18,62 %	100,47 %

2. Plus haut, nous remarquons que l'estimateur du total était, sur l'ensemble de la base, plus précis que l'estimateur de la moyenne de la densité. Précisons maintenant ce qu'il en est au niveau de chaque strate. Le tableau 6 donne, pour chaque strate, la précision (coefficient de variation) des estimations pour les deux caractères.

Tableau 6 : Coefficients de variation des deux caractères estimés à l'intérieur de chaque strate (au taux de 5%, estimation stratifiée à allocation proportionnelle)

Strates	estimation du total de population (%)	estimation de la moyenne de densité (%)
1: Centre ville dense	11,25	26,29
2: Quartiers industriels	47,50	51,12
3: Grands ensembles d'habitation	17,43	36,48
4: Petits immeubles d'habitation	54,34	72,55
5: Pavillonnaire	21,33	18,62
6: Péri-urbain peu dense	76,34	68,83
7: Reliquat morphologique	36,79	32,74
8: Dense : grands bâtiments élevés	45,53	32,54
9: Non bâti	109,88	100,47

L'estimateur du total est beaucoup plus précis dans les strates 1 et 3, "centre ville dense" et "grands ensembles d'habitation" (environ deux fois plus) ; il est également plus précis dans la strate 4, "petits immeubles d'habitation" (1,34 fois).

Les deux estimateurs sont à peu près équivalents dans les strates 2 ("quartiers industriels"), 5 ("pavillonnaire"), 6 ("péri-urbain peu dense"), 7 ("reliquat morphologique") et 9 ("non bâti"). Pour ces quatre dernières, l'estimateur de la densité est même légèrement meilleur : 1,1 fois plus précis.

Enfin, dans la strate 8, l'estimateur de la moyenne des densités est 1,4 fois plus précis que celui du total de population. La précision globalement meilleure des estimations du total de population s'explique donc par les meilleurs résultats qu'elle fournit sur des sous-ensembles importants, en population et en surface, de l'agglomération marseillaise : centre ville, grands ensembles et, à un moindre degré, petits immeubles. Par ailleurs, la différence est peu importante, entre les deux estimateurs dans la troisième strate dominante, le pavillonnaire.

3. Examinons maintenant la précision des estimations obtenues pour chaque strate. Les figures 7 et 8 donnent, par strate et pour chacun des deux estimateurs (du total et de la moyenne), les intervalles de confiance en pourcentage de l'espérance de l'estimateur dans la strate. Les deux types d'allocations, proportionnelle et optimale, y sont considérés.

Figure 7 : Précision des estimations partielles du total de population sans doubles comptes au taux de 5 % .

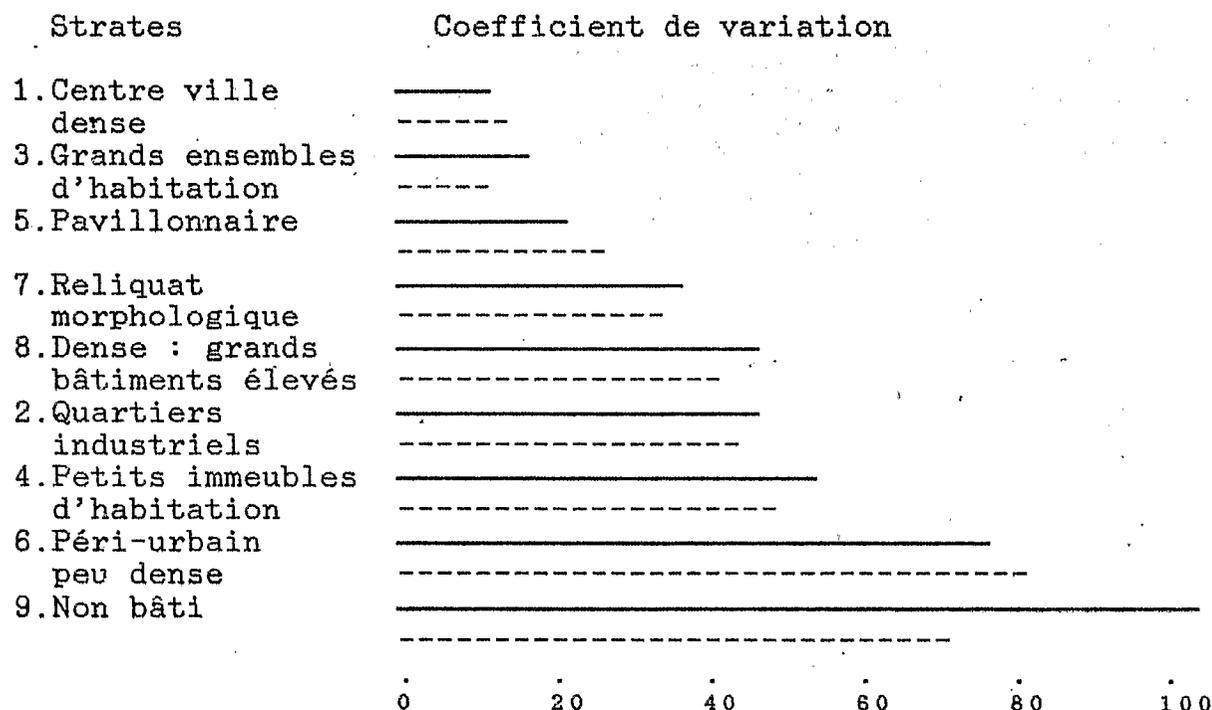
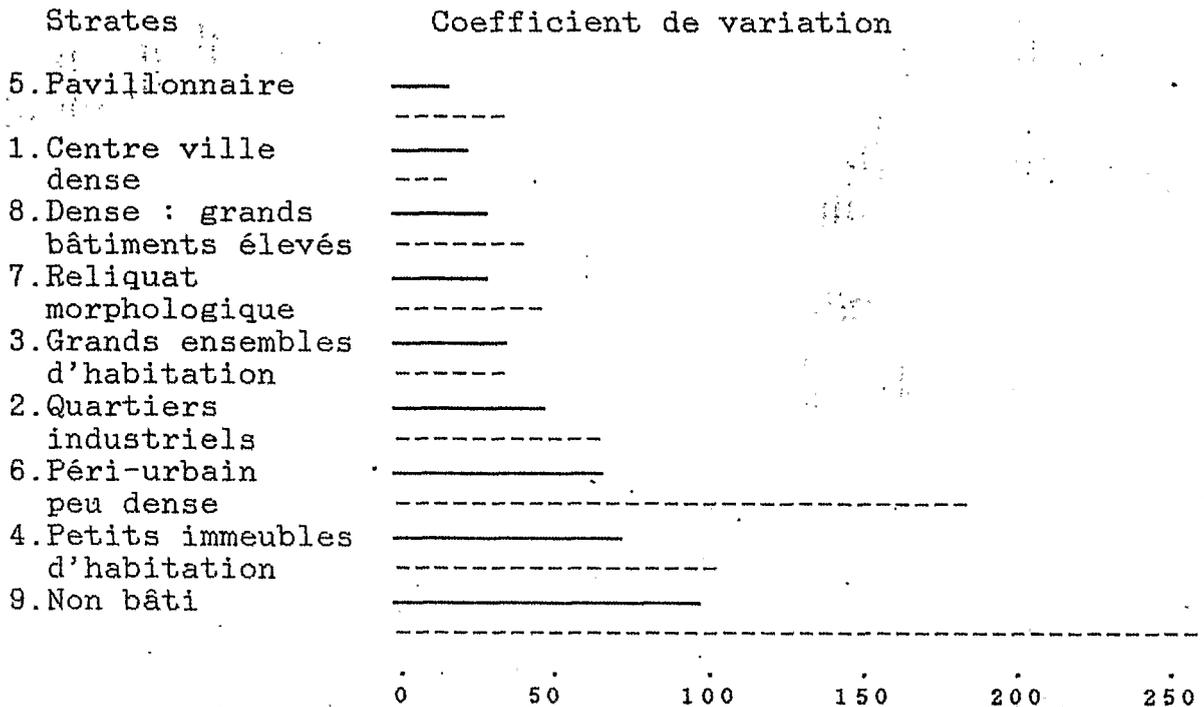


Figure 8 : précision des estimations partielles de la moyenne de la densité de population sans doubles comptes au taux de 5 %



figures 7 et 8 : — : allocation proportionnelle
 - - - - - : allocation optimale

Il se dégage clairement, de l'examen des deux figures, trois groupes de strates :

a/ Les strates dont la morphologie est bien homogène, "centre ville", "grands ensembles" et "pavillonnaire", bénéficient des meilleurs précisions, tant pour l'estimation de l'effectif que pour celle de la densité : l'estimation du total de population sans doubles comptes se fait dans un intervalle de confiance compris entre +/- 20 % et +/- 50 % de la vraie valeur; pour l'estimation de la moyenne de densité des districts, cet intervalle varie entre +/- 40 % et +/- 70 %. Les meilleurs résultats sont obtenus, pour l'estimation du total, dans les strates 1 (+/- 21%) et 3 (+/- 34 %), et pour l'estimation de la moyenne de densité, dans les strates 5 (+/- 36,5%) et 1 (+/- 51%).

b/ Les strates marginales (très minoritaires en surface), "reliquat morphologique" et "dense : grands bâtiments élevés". Pour ces deux strates, la précision des estimations est moyenne dans le cas du total et assez bonne pour la densité du moins si l'on s'en tient à l'allocation proportionnelle (intervalles de confiance de l'ordre de +/- 75 % pour l'estimation du total et +/- 65 % pour celle de la moyenne des densités).

c/ Les strates globalement peu denses et hétérogènes du point de vue morphologique et démographique : "quartiers industriels", "petits immeubles d'habitation", "péri-urbain peu dense" et "non bâti", pour lesquelles les estimations des caractères démographiques sont très imprécises : les intervalles de confiance se situent entre +/- 100 % (industriel) et +/- 220 % (non bâti) pour l'estimation du total, et entre +/- 100 % (industriel) et +/- 200 % (non bâti), pour l'estimation de la densité.

Ces résultats ne surprendront ni les urbanistes ni les démographes; ils justifient à nos yeux l'emploi des critères morphologiques pour la stratification et, a posteriori, la méthode d'analyse typologique employée pour sa mise au point. En effet, il est clair d'une part, que cette stratification a permis d'isoler des sous-ensembles urbains où les caractères démographiques ont des variances différentes, et d'autre part, que les zones les plus homogènes morphologiquement bénéficient de la meilleure précision d'estimation. Rappelons, à ce propos, que les groupes "centre ville", "grands ensembles" et "pavillonnaire" s'isolaient en premier lieu lors des classifications et analyses factorielles (voir rapport intermédiaire d'avril 86 ou annexe 6 du rapport ATP CNRS/CNES d'août 87).

4. Enfin, il nous faut examiner dans quelle mesure l'optimisation des résultats globaux grâce à l'allocation de NEYMAN est compatible avec la précision des estimations partielles.

Les figures 7 et 8 sont également très éloquentes sur ce point :

- dans le cas de l'estimation du total de population, l'allocation optimale n'affecte qu'assez peu la précision des estimations du total de chaque strate. Elle l'améliore même sensiblement dans les strates 3 : "grands ensembles" et 9 : "non bâti".

- il en va tout autrement pour l'estimation de la moyenne des densités : la précision des estimations diminue de manière très sensible pour l'ensemble des strates, à la seule exception de la strate "centre ville dense". En particulier, les deux groupes de strates globalement peu denses (4,5,6 et 9) et morphologiquement marginales (7 et 8), pâtissent beaucoup de ce type d'allocation. Cela constitue évidemment un argument supplémentaire en la défaveur de cet estimateur.

3.5. Conclusions, perspectives

3.5.1. Précision des estimations, ordre de grandeur des intervalles de confiance, effet de la répartition géographique de l'échantillon

En assimilant le tirage systématique pratiqué à un tirage équiprobable sans remise, une fois introduite la stratification et optimisée l'allocation aux strates, on obtient les précisions suivantes (mesurées par l'intervalle de confiance à 95 %) pour les estimations, sur l'ensemble de la base de sondage, du total de population sans doubles comptes et de la moyenne de densité des districts.

Total de population s. d. c. : +/- 16,6 % au taux de 5 %

Moyenne des densités : +/- 27,3 % au taux de 5 %

Comme nous le remarquons déjà plus haut, une bonne répartition géographique des unités statistiques de l'échantillon au sein de l'agglomération étudiée ainsi qu'à l'intérieur de chaque strate, est assurée, dans le cas de l'expérience marseillaise, par le tri de la base de sondage sur l'identifiant INSEE. Elle le sera mieux encore dans les cas concrets d'application, grâce au tirage d'unités aréolaires sur grille de points.

Cette répartition géographique apporte certainement un gain de précision notable à l'estimation des caractères démographiques. Ce gain n'est pas pris en compte dans les chiffres donnés ci-dessous: on peut donc les considérer comme des évaluations pessimistes des précisions réelles.

Le premier objectif qu'on peut fixer à la suite de ces tests est donc la quantification de ce gain et son introduction dans les calculs de précision des estimations (voir chapitre 7).

3.5.2. Apport de la stratification, estimations globale et estimations partielles, domaines géographiques de précision

Pour les estimations globales - on entend par là celles qui s'étendent à l'ensemble de la base de sondage considérée - l'introduction de la stratification, et de son corollaire, l'allocation optimale, permet un gain sur la variance des estimations compris, suivant les cas, entre 32 et 39 % de la variance non stratifiée.

D'autre part, l'introduction de la stratification permet des estimations partielles (estimation des caractères démographiques à l'intérieur de chaque strate). Les résultats obtenus dessinent dans l'espace urbain marseillais certains

domaines de bonne précision et d'autres d'imprécision de ces estimations partielles (par "bonne précision" ou "imprécision", il faut entendre des chiffres comparables où non, dans leur ordre de grandeur, à ceux obtenus au même taux sur l'ensemble de la base) :

Domaine de bonne précision : strates "centre ville", "grands ensembles d'habitation" et "pavillonnaire". Les estimations se font, au taux de 5%, avec des intervalles de confiance compris entre +/- 20 et +/- 50 %, pour celle du total de population s.d.c., entre +/- 40 et +/- 70 %, pour la moyenne de la densité. Rappelons que près de 80 % de la population totale de la zone d'étude est regroupée dans ces trois strates.

Domaine d'imprécision : strates "quartiers industriels", "petits immeubles d'habitation", "péri-urbain peu dense" et "non bâti"; ici les intervalles de confiance, compris entre +/- 100 % et +/- 170 %, ne permettent que des estimations très imprécises des caractères démographiques.

En résumé, le principe d'une stratification de la base de sondage à l'aide de critères morphologiques ainsi que la méthode d'analyse typologique qui a servi à son élaboration sont à retenir pour deux raisons :

1. parce qu'ils permettent une amélioration significative des estimations globales.
2. parce qu'ils autorisent des estimations partielles avec une précision correcte dans les strates de caractéristiques morphologiques homogènes et, ce qui n'est pas moins important, permettent de quantifier l'imprécision qui affecte ces mêmes estimations ailleurs. Ce résultat représente, à notre avis, l'acquis majeur de cette première phase de test.

Pour la suite, s'agissant de stratification, nous considérerons comme acquise l'utilisation des caractères morphologiques, mais nous testerons des méthodes de stratification plus proches de celle qu'on pense pouvoir appliquer à l'information satellitaire. En procédant analytiquement cette fois (par opposition au caractère synthétique des résultats d'analyse typologique), on s'attachera à évaluer séparément, l'apport des principaux caractères morphologiques que l'on pense pouvoir observer sur l'image (densité du bâti, taille et hauteur des bâtiment ...etc), en stratifiant la base de sondage d'après chacun d'eux. Bien entendu, on continuera à mesurer le gain apporté par la stratification synthétique. On peut d'ailleurs la considérer comme une référence à laquelle on peut comparer ce qu'il sera possible de faire d'après l'image satellite (voir les chapitres 5 et 6).

3.5.3. Comparaison entre les estimateurs de l'effectif et de la densité.

Tout au long du test mené, la comparaison entre l'estimateur du total de population sans doubles comptes et celui de la moyenne de la densité des districts s'est conclu en faveur du premier. Rappelons rapidement les avantages qu'il possède dans le cadre d'un plan équiprobable qui fut le nôtre ici.

1. Il est nettement plus précis que celui de la densité quand on s'intéresse à l'estimation globale; le gain sur l'amplitude de l'intervalle de confiance, dans le cas d'estimations stratifiées à allocation optimale, est de 40 %.

2. Il est plus précis, ou au moins autant, si l'on considère les estimations partielles dans le sous-ensemble où elles sont les meilleures : nettement plus précis dans le centre ville et les grands ensemble (gain approximatif de 55 % dans les deux cas), et à peu près équivalent à l'estimateur de la moyenne dans le pavillonnaire.

3. L'allocation optimale ne perturbe que faiblement les résultats par strate obtenus avec cet estimateur, ce qui permet une bonne précision de l'estimation globale comme des estimations partielles. Dans le cas de l'estimation de la moyenne des densités, ces deux objectifs apparaissent difficilement conciliables.

Il nous reste à faire plusieurs remarques qui relativisent beaucoup l'intérêt de ce résultat.

Premièrement, il est probablement dû en partie à la réduction artificielle de la variance de l'effectif de population induite par le découpage des districts de recensement de l'INSEE. A l'examen des fonds de cartes utilisés lors du recensement, il apparaît que ce facteur a pu jouer dans les strates où le réseau de voirie étant assez lâche, un certain nombre de grands îlots ont été partagés en plusieurs districts de recensement, ayant chacun des effectifs de population "raisonnables". En revanche, ce ne peut pas être le cas dans le centre ville dense, où la taille des îlots ne justifie jamais un tel découpage.

Deuxièmement, s'il est vrai, au moins dans des sous-ensembles importants de l'agglomération, que l'effectif de population des districts est moins variable que leur densité, il ne faut pas perdre de vue que la mesure de cette variabilité qui sert à déterminer la précision de l'estimation est la variance du caractère, calculée sous la loi de probabilité correspondante au tirage. Ceci signifie que cette mesure varie suivant le plan de sondage et, avec elle, la variance de l'estimateur. Il n'est donc pas acquis que l'estimation du total de population soit toujours plus précise que celle de la densité pour un plan de sondage à probabilités inégales.

Troisièmement, ce résultat n'est valable que tant que l'unité statistique enquêtée reste le pâté de maisons ou peut lui être comparé. Il faut également que dans la ville considérée, existent des sous-ensembles urbain, du type "centre ville dense" ou "grands ensembles d'habitation", semblables à ceux rencontrés à Marseille et qui soient prédominants en effectifs de population.

Enfin et surtout, l'intérêt de la comparaison pratiquée jusqu'à présent est très limité puisque ses deux termes n'estiment pas la même quantité : le premier fournit les totaux de population dans les strates ou l'ensemble de la base, alors que le second est la moyenne arithmétique des densités des districts, quantité qui n'a d'ailleurs, que peu d'intérêt pratique.

3.5.4. Conclusion

Cette étape aura donc essentiellement permis de mettre au point la séquence d'un test que nous allons étendre maintenant à des plans de sondage à probabilités inégales, à des estimateurs par le ratio faisant intervenir des variables exogènes, à d'autres caractères démographiques et enfin à d'autres stratifications. Les chapitres 3 à 7 sont consacrés à ces différentes phases de l'expérience.

Tableau 4.1 : Précision de l'estimation du total de population sans doubles comptes, plan équiprobable, estimateur sans biais au taux de 5 %, stratification morphologique synthétique en 9 classes.

Strates		Total	Vari- -ance x 10 ⁻⁸	Coef. de varia- -tion (%)	Intervalle de confiance à 95 %	Evolution de la var. avec l'al- -location optimale (%)
1 Centre ville	*	207739	5.458	11.25	+/- 46724.79	0.00
	**	207739	8.010	13.62	+/- 56604.29	46.76
2 Quartiers Indus- -triels	*	40680	3.734	47.50	+/- 38648.41	0.00
	**	40680	3.289	44.58	+/- 36270.93	-11.92
3 Grands ensembles	*	264764	21.292	17.43	+/- 92286.60	0.00
	**	264764	11.725	12.93	+/- 68484.79	-44.93
4 Petits immeubles	*	28750	2.441	54.34	+/- 31246.44	0.00
	**	28750	1.944	48.50	+/- 27886.19	-20.36
5 Pavil- -lonnaire	*	93937	4.016	21.33	+/- 40077.76	0.00
	**	93937	6.252	26.62	+/- 50006.84	55.68
6 Peri- urbain	*	11648	0.791	76.34	+/- 17784.44	0.00
	**	11648	0.861	79.66	+/- 18557.84	8.85

* : Allocation proportionnelle

** : Allocation optimale

Tableau 4.1 (suite) : Précision de l'estimation du total de population sans doubles comptes, plan équiprobable, estimateur sans biais au taux de 5 %, stratification morphologique synthétique en 9 classes.

Strates		Total	Vari- -ance x 10 ⁻⁸	Coef. de varia- -tion (%)	Intervalle de confiance à 95 %	Evolution de la var. avec l'al- -location optimale (%)
7	*	40534	2.224	36.79	+/- 29825.73	0.00
Reliquat	**	40534	1.996	34.85	+/- 28253.18	-10.25
8	*	23788	1.173	45.53	+/- 21659.08	0.00
Dense, Grands batiments	**	23788	0.930	40.54	+/- 19286.52	-20.72
9	*	17032	3.502	109.88	+/- 37429.50	0.00
Non bâti	**	17032	1.450	70.70	+/- 24084.38	-58.60
Ensemble (non strati- -fié)		728872	53.488	10.03	+/-146270.83	0.00
Ensemble (stratifié, allocation prop.)		728872	44.631	9.17	+/-133612.26	-16.56
Ensemble (stratifié, allocation optimale)		728872	36.457	8.28	+/-120759.19	-31.84

* : Allocation proportionnelle

** : Allocation optimale

Tableau 4.2 : Précision de l'estimation de la moyenne de densité de population sans doubles comptes, plan équiprobable, estimateur sans biais au taux de 5 %, stratification morphologique synthétique en 9 classes.

Strates		Total	Vari- -ance x 10 ⁻⁸	Coef. de varia- -tion (%)	Intervalle de confiance à 95 %	Evolution de la var. avec l'al- -location optimale (%)
1 Centre ville	*	4.338	1.30	26.29	+/- 2.28	0.00
	**	4.338	0.67	18.81	+/- 1.63	-48.46
2 Quartiers Indus- -triels	*	1.291	0.44	51.12	+/- 1.32	0.00
	**	1.291	0.78	68.43	+/- 1.77	77.30
3 Grands ensembles	*	2.313	0.71	36.48	+/- 1.69	0.00
	**	2.313	0.66	35.22	+/- 1.63	-7.00
4 Petits immeubles	*	1.050	0.58	72.55	+/- 1.52	0.00
	**	1.050	1.22	105.38	+/- 2.21	110.30
5 Pavil- -lonnaire	*	0.944	0.03	18.62	+/- 0.35	0.00
	**	0.944	0.11	35.29	+/- 0.67	266.60
6 Peri- urbain	*	0.416	0.08	68.83	+/- 0.57	0.00
	**	0.416	0.58	183.78	+/- 1.53	625.00

* : Allocation proportionnelle

** : Allocation optimale

Tableau 4.2 (suite) : Précision de l'estimation de la moyenne de densité de population sans doubles comptes, plan équiprobable, estimateur sans biais au taux de 5 %, stratification morphologique synthétique en 9 classes.

Strates		Moyen- -ne	Vari- -ance x 10 ⁻⁸	Coef. de varia- -tion (%)	Intervalle de confiance à 95 %	Evolution de la var. avec l'al- -location optimale (%)
7	*	2.172	0.51	32.74	+/- 1.42	0.00
Reliquat	**	2.172	1.07	47.58	+/- 2.07	109.80
8	*	4.695	2.33	32.54	+/- 3.06	0.00
Dense, Grands batiments	**	4.695	3.56	40.18	+/- 3.77	52.80
9	*	0.342	0.12	100.47	+/- 0.69	0.00
Non bâti	**	0.342	0.89	276.22	+/- 1.89	641.67
Ensemble (non strati- -fié)		2.426	0.18	17.42	+/- 0.85	0.00
Ensemble (stratifié, allocation prop.)		2.426	0.17	16.79	+/- 0.81	-5.56
Ensemble (stratifié, allocation optimale)		2.426	0.11	13.66	+/- 0.66	-38.89

* : Allocation proportionnelle

** : Allocation optimale

CHAPITRE 4

FORMULAIRE, PROGRAMME INFORMATIQUE

Introduction

On a déjà souligné précédemment que le tirage d'échantillons d'après documents satellitaires se fera de manière systématique, à l'aide d'une grille de points placée sur la base de sondage. Cette contrainte définit un plan de sondage où la probabilité de sélection d'une unité statistique dans l'échantillon est proportionnelle à une mesure de la taille de cette unité : surface totale de l'unité si la procédure sélectionne tous les points de la grille, surface bâtie si elle ne retient que les points situés au dessus d'une surface bâtie. Outre qu'il s'agit pour nous d'un impératif pratique (1), de tels plans de sondage modifient la précision des estimations ; mais cette précision dépend également des estimateurs et des stratifications employées.

COCHRAN (2), dans "Sampling technics" et DESABIE (3) dans "Théorie et pratique des sondages", recensent les différentes possibilités d'utiliser une information exogène relative à la taille des unités statistiques; COCHRAN dit, par exemple : "Quand les unités sont de taille inégale, il y a le choix entre aux moins quatre techniques (en supposant connues la

1 : Rappelons que, dans la situation typique d'application de la méthode, nous ne pensons pas pouvoir disposer d'information qui permettent de constituer une autre base de sondage que celle, spatiale, issue de l'image satellite. Dans une telle situation, la constitution d'une liste d'unités statistiques spatiales, qui seule, permet d'envisager des tirages équiprobables et, plus généralement de contrôler complètement la probabilité de sortie des unités, n'est pas toujours possible. Elle réclamerait, de toute façon, un important travail d'identification et de vérification des unités spatiales, préjudiciable à la rapidité de mise en oeuvre de l'enquête.

2 : W.G. COCHRAN : Sampling technics, WILEY 1977, troisième édition, chap. 9, P. 255.

3 : J. DESABIE : Théorie et pratique des sondages, DONOD 1966, Chap. 12, P. 241.

taille de chaque unité si ces techniques le nécessitent).", et il dresse la liste suivante :

1. Plan équiprobable, estimation linéaire sans biais (bien que figurant dans la liste de COCHRAN, cette option ne permet pas d'utiliser l'information exogène, elle sert en fait de référence).

2. Plan équiprobable, estimation par le ratio relatif à la taille des unités.

3. Plan à probabilités proportionnelles à la taille des unités, estimation sans biais.

4. Plan équiprobable, stratification de la base suivant la taille des unités, estimation stratifiée usuelle (sans biais).

En fait, seule la dernière de ces techniques nécessite, avant le tirage, la connaissance sur l'ensemble de la base de la taille des unités. Elle est donc pratiquement difficile à mettre en oeuvre, de plus notre recherche repose sur l'hypothèse qu'une stratification sur des critères morphologiques lui est préférable. Il ne sera donc plus question de cette technique dans la suite.

Dans les chapitres suivants, on tente de mesurer l'effet sur la précision des estimations, des différentes possibilités d'intégration d'informations exogènes dans la technique de sondage : plans à probabilités inégales, estimation par le ratio et stratification. Pour cela, nous reprenons les trois premières idées exposées par COCHRAN, en leur adjoignant l'estimation par le ratio sous plan à probabilités proportionnelles à la taille des unités.

Ce chapitre décrit les outils statistiques et informatiques qui vont permettre d'effectuer le test, on donne le formulaire correspondant aux quatre possibilités d'estimation, puis un exposé de l'architecture et du mode d'emploi du programme informatique.

4.1 Formulaire : estimateurs sans biais, estimateurs par le ratio, plans équiprobables, plans à probabilités inégales.

4.1.1 Préliminaire

Lors de la phase exploratoire exposée dans le chapitre 3, le formulaire correspondait à un plan de sondage sans remise, dont la théorie, développée en 1952 par Horwitz et Thompson trouve une application simple dans le cas de plans équiprobables. Lorsque les probabilités de sélection sont inégales, le tirage sans remise, qui reste bien entendu préférable, conduit à des calculs complexes qui supposent

d'expliciter les probabilités d'appartenance à l'échantillon, celles ci se modifiant au fur et à mesure du tirage. Or, le but de l'exposé qui va suivre est, rappelons le, triple :

- comparer des plans de sondages équiprobables à des plans où la probabilité est proportionnelle à la taille des unités d'échantillonnage ;
- comparer les estimateurs sans biais aux estimateurs par le quotient ;
- évaluer l'apport d'une stratification morphologique de la base de sondage à la précision d'estimations démographiques.

Relativement à ces objectifs, s'il importe que la procédure de tirage considérée soit la même dans tous les cas, on peut par contre utiliser indifféremment un tirage avec ou sans remise. L'hypothèse qui est alors faite est que le tirage sans remise améliore chaque estimation approximativement dans les mêmes proportions. C'est la raison pour laquelle on donnera ici les formules simples correspondant à des tirages avec remise. La même option étant prise au sein du programme de calcul des variances d'estimation, ces dernières sont donc surévaluées dans une proportion approximativement égale au taux de sondage : 5 à 7 % suivant la base de sondage et le taux retenus(4).

4.1.2 Conventions

1. Dans ce qui suit, on se contente dans la plupart des cas, d'énoncer les expressions des estimateurs, de leur variance, etc..., sans démonstrations. Ces résultats sont en effet désormais classiques et leurs démonstrations figurent dans la plupart des ouvrages généraux cités dans la bibliographie. Seule l'estimation par le ratio sous plan à probabilités inégales est exposée en détail, cet exposé étant beaucoup moins fréquent dans la littérature sur les sondages.

2. Comme on l'a constaté dans le chapitre 3, les formules stratifiées s'obtiennent sans difficultés par sommation sur l'ensemble des strates des formules établies pour l'ensemble de la base. Elles n'apparaîtront généralement pas ici ; seul le raisonnement menant à l'allocation optimale part de l'expression explicite de la variance dans le cas stratifié.

3. Les notations seront identiques à celle adoptées dans le précédent chapitre : la variable estimée est toujours notée y , les valeurs prises sur les unités statistiques y_i , son total Y , etc... . Lorsqu'il est fait, dans les formules relatives à

(4) : Cette estimation du gain dû au tirage sans remise provient de la comparaison des expressions des variances des estimateurs sans biais obtenues avec et sans remise dans le cas d'un plan de sondage équiprobable : la variance sans remise est alors inférieure à la variance avec remise d'un facteur $(N - n)/N$ égal au taux.

l'estimation par le ratio, référence à la variable exogène utilisée comme dénominateur, celle ci est notée x, ses valeurs x_i , son total X, etc... La variable servant à définir la probabilité dans les plans à probabilités inégales est notée z, sa valeur z_i et son total Z.

4.1.3 Plan équiprobable, estimateur sans biais

On a, de même qu'au chapitre 3, les estimateurs :

$$\hat{T}_1(Y) = N \cdot \bar{y} = N \cdot \sum_{i=1}^n y_i / n \quad (413.1 \text{ a}) \text{ et,}$$

$$\hat{M}_1(y) = \bar{y} = \sum_{i=1}^n y_i / n \quad (413.1 \text{ b}) , \text{ mais les variances ont des}$$

expressions plus simples :

$$V(\hat{T}_1(Y)) = \frac{N^2 \cdot \sigma_y^2}{n} \quad \text{et} \quad V(\hat{M}_1(y)) = \frac{\sigma_y^2}{n} \quad (413.2 \text{ a \& b}) .$$

La détermination de l'allocation optimale se fait par minimisation de la variance stratifiée sous la contrainte $\sum n_h = n$. La solution du problème est obtenue grâce à l'utilisation des multiplicateurs de Lagrange.

on recherche $\text{Min} \left(\sum_{h=1}^K (N_h \cdot \sigma_h^2 / n_h) \right)$, dont le Lagrangien

$$\text{s'écrit : } L = \sum_{h=1}^K (N_h \cdot \sigma_h^2 / n_h) + \left(n - \sum_{h=1}^K n_h \right) \cdot \beta,$$

ou β est un multiplicateur de LAGRANGE.

Les conditions du premier ordre sont : $\frac{dL}{dn_h} = 0$ pour tout h,

$$\text{on a donc : } N_h \cdot \sigma_h^2 / n_h^2 = \beta \Leftrightarrow n_h = (1/\sqrt{\beta}) \cdot N_h \cdot \sigma_h .$$

En écrivant la contrainte $\sum_{h=1}^K n_h = n$, on trouve :

$$1/\sqrt{\beta} = n / \sum_{h=1}^K N_h \cdot \sigma_h .$$

D'ou l'allocation $n_h = n \cdot (N_h \cdot \sigma_h / \sum_{h=1}^K N_h \cdot \sigma_h)$ (413.3) pour la strate h.

4.1.4 Plan équiprobable, estimateur par le ratio.

La connaissance du total, sur l'ensemble de la base, d'une variable exogène jamais nulle (dans nos applications, il s'agira de la taille des districts : surface totale dans un premier temps, surface bâtie par la suite), permet d'utiliser, pour estimer le total Y de y ou pour estimer directement le quotient Y/X, l'estimateur naturel du ratio Y/X = R :

$$\hat{R} = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i} \quad (414.1)$$

L'estimateur du total Y s'écrit alors : $T_2(Y) = X \cdot \hat{R}$, celui de la moyenne $M_2(y) = \bar{x} \cdot \hat{R}$.

Ces estimateurs sont légèrement biaisés, le calcul de ce biais se fait en utilisant le développement limité d'ordre un à l'origine de $1/(1 + \epsilon)$, où ϵ représente la quantité

$(x - X)/X$, proche de zéro dans notre cas. On obtient, pour un plan équiprobable avec remise, l'expression suivante pour le biais sur l'estimation du ratio R (5).

$$E(\hat{R} - R) \approx - \frac{N^2}{n \cdot X^2} \cdot \left(\frac{\sigma_{x,y}^2}{x \cdot y} - R \cdot \frac{\sigma_x^2}{x} \right) \quad (414.2 \text{ a}).$$

Le biais sur l'estimation d'un total est :

$$E(T_2(Y) - Y) \approx - \frac{N^2}{n \cdot X^2} \cdot \left(\frac{\sigma_{x,y}^2}{x \cdot y} - R \cdot \frac{\sigma_x^2}{x} \right), \quad (414.2 \text{ b})$$

celui sur l'estimation d'une moyenne :

$$E(M_2(y) - \bar{y}) \approx - \frac{N}{n \cdot X} \cdot \left(\frac{\sigma_{x,y}^2}{x \cdot y} - R \cdot \frac{\sigma_x^2}{x} \right). \quad (414.2 \text{ c})$$

La variance de l'estimateur $T_2(Y)$ s'écrit :

$$V(T_2(Y)) = (N^2/n) \cdot \left(\frac{\sigma_y^2}{y} + R^2 \cdot \frac{\sigma_x^2}{x} - 2R \cdot \frac{\sigma_{x,y}^2}{x \cdot y} \right) \quad (414.3 \text{ a}), \quad \text{celle}$$

(5): Pour un exposé complet du calcul, voir par exemple J.M. GROSBAS : Méthodes statistiques des sondages, Pages 131 & 132.

de l'estimateur $M_2(y)$:

$$V(M_2(y)) = (1/n) \cdot (\sigma_y^2 + R^2 \cdot \sigma_x^2 - 2R \cdot \sigma_{x,y}) \quad (414.3 \text{ b}).$$

Le risque associé aux deux estimateurs s'obtient ensuite en sommant la variance et le carré du biais.

Pour déterminer l'allocation optimale, il faut minimiser la variance totale stratifiée, dont l'expression est :

$$V = \sum_{h=1}^K (N_h/n_h) \cdot (\sigma_{hy}^2 + R_h^2 \sigma_{hx}^2 - 2R_h \sigma_{hxy}) + \sum_{h=1}^K (N_h/n_h X_h) \cdot (\sigma_{hxy}^2 - R_h \sigma_{hx}^2),$$

où K est le nombre de strates et σ la variance. En appliquant la méthode des multiplicateurs de Lagrange, on abouti au système d'équations en n_h :

Pour tout h ,

$$(N_h/n_h) \cdot (\sigma_{hy}^2 + R_h^2 \sigma_{hx}^2 - 2R_h \sigma_{hxy}) + (2N_h/n_h X_h) \cdot (\sigma_{hxy}^2 - R_h \sigma_{hx}^2) = c$$

En négligeant le second terme, d'ordre $1/n_h^3$, devant le premier, d'ordre $1/n_h^2$, les conditions s'écrivent :

pour tout h ,

$$N_h/n_h = C \cdot (\sigma_{hy}^2 + R_h^2 \sigma_{hx}^2 - 2R_h \sigma_{hxy})^{1/2}, \text{ puis en écrivant la}$$

contrainte sur la taille de l'échantillon : $\sum_{h=1}^K n_h = n$, on obtient l'expression des n_h :

$$n_h = \frac{N_h \cdot (\sigma_{hy}^2 + R_h^2 \sigma_{hx}^2 - 2R_h \sigma_{hxy})^{1/2}}{\sum_{h=1}^K N_h \cdot (\sigma_{hy}^2 + R_h^2 \sigma_{hx}^2 - 2R_h \sigma_{hxy})^{1/2}} \quad (414.4)$$

4.1.5 Plan à probabilités inégales, estimateur sans biais.

Les probabilités d'inclusion π_i sont définies proportionnelles à une variable exogène z de total Z ; il s'agit pour nous de la

mesure de la taille z_i associée à chaque unité statistique de la base. On a donc $\pi_i = n z_i / Z$ pour tout i et l'estimateur sans biais du total Y s'écrit (6) :

$$\hat{T}_3(Y) = (Z/n) \cdot \sum_{i=1}^n (y_i / z_i) , \text{ celui de la moyenne :}$$

$$\hat{T}_3(y) = (Z/(N \cdot n)) \cdot \sum_{i=1}^n (y_i / z_i) \quad (415.1 \text{ a \& b})$$

Pour les variances, on a :

$$V(\hat{T}_3(Y)) = (1/n) \cdot (Z \cdot \sum_{i=1}^N (y_i^2 / z_i) - Y^2) \quad (415.2 \text{ a})$$

$$V(\hat{T}_3(y)) = (1/n \cdot N^2) \cdot (Z \cdot \sum_{i=1}^N (y_i^2 / z_i) - Y^2) \quad (415.2 \text{ b})$$

Le calcul de l'allocation optimale se fait à partir de la variance stratifiée :

$$V(\hat{T}_3(Y)) = \sum_{h=1}^K (1/n_h) \cdot (Z_h \cdot \sum_{i=1}^{N_h} (y_i^2 / z_i) - Y_h^2)$$

En appliquant à cette fonction la même méthode qu'au paragraphe précédent, on obtient les conditions :

$$\text{Pour tout } h, (1/n_h) \cdot (Z_h \cdot \sum_{i=1}^{N_h} (y_i^2 / z_i) - Y_h^2) = c$$

$$\text{D'où } n_h = C \cdot (Z_h \cdot \sum_{i=1}^{N_h} (y_i^2 / z_i) - Y_h^2)^{1/2} \quad \text{et, puisque } \sum_{h=1}^K n_h = n,$$

$$n_h = \frac{n \cdot (Z_h \cdot \sum_{i=1}^{N_h} (y_i^2 / z_i) - Y_h^2)^{1/2}}{\sum_{h=1}^K (Z_h \cdot \sum_{i=1}^{N_h} (y_i^2 / z_i) - Y_h^2)^{1/2}} \quad (415.3)$$

(6) : Voir DESABIE : Théorie et pratique des sondages, chap. 11, P. 227.

4.1.6 Plan à probabilités inégales; estimation d'un ratio ou d'un total par le ratio

. L'estimateur

De nouveau, les probabilités de tirage π_i sont définies par les z_i , avec : $\pi_i = n \cdot z_i / Z$; x étant une variable exogène (en principe distincte de z), de total X , on s'intéresse à l'estimation du ratio $R = Y/X$, ainsi qu'à celle du total de y par le truchement de R puisque $Y = RX$.

Sous le plan π , ainsi défini, les estimateurs sans biais des totaux X et Y sont connus :

$$\hat{T}_3(Y) = (Z/n) \cdot \sum_{i=1}^n (y_i/z_i), \quad \hat{T}_3(X) = (Z/n) \cdot \sum_{i=1}^n (x_i/z_i)$$

L'estimateur naturel du ratio $R = Y/X$ est donc :

$$\hat{R} = \frac{\sum_{i=1}^n (y_i/z_i)}{\sum_{i=1}^n (x_i/z_i)} \quad (416.1), \text{ on estimera } Y \text{ par :}$$

$\hat{T}_4(Y) = \hat{X} \cdot \hat{R}$, et \hat{y} par : $\hat{T}_4(y) = (\hat{X}/N) \cdot \hat{R}$. Mais ces estimateurs sont légèrement biaisés.

. Expression des biais

La différence entre le ratio R et son estimateur s'écrit :

$$\hat{R} - R = \frac{\sum_{i=1}^n (y_i/z_i)}{\sum_{i=1}^n (x_i/z_i)} - Y/X = (\hat{Y}/\hat{X}) - (Y/X)$$

ou encore : $\hat{R} - R = \frac{\hat{Y} - RX}{\hat{X}} = \frac{\hat{Y} - RX}{X(1 + (\hat{X} - X)/X)}$, où la quantité

$(\hat{X} - X)/X$ est proche de zéro, surtout lorsque n grandit puisque \hat{X} est un estimateur sans biais de X . On peut donc obtenir, par développement limité du premier ordre, une expression approchée de cette différence :

$$\hat{R} - R \approx \frac{\hat{Y} - RX}{X} (1 - (\hat{X} - X)/X), \text{ à partir de laquelle on}$$

calcule la valeur approchée du biais sur l'estimation du ratio :

$$E(\hat{R} - R) \approx E\left(\frac{\hat{Y} - R\hat{X}}{X}\right) - 1/X^2 \cdot E((\hat{Y} - R\hat{X})(\hat{X} - X)).$$

Or, $E(\hat{Y} - R\hat{X}) = E(\hat{Y}) - RE(\hat{X}) = Y - RX = 0$, par définition de \hat{X} et \hat{Y} , donc,

$$E(\hat{R} - R) \approx -1/X^2 \cdot E(\hat{X}(\hat{Y} - R\hat{X})) = -1/X^2 \cdot (E(\hat{X}\hat{Y}) - RE(\hat{X}^2)), \text{ ou}$$

encore, grâce à des relations connues :

$$E(\hat{R} - R) \approx -1/X^2 \cdot (\text{Cov}(\hat{X}, \hat{Y}) - R \cdot \text{Var}(\hat{X})). \text{ Exprimons } \text{Cov}(\hat{X}, \hat{Y})$$

en fonction des variables aléatoires x, y et z :

$$\begin{aligned} \text{Cov}(\hat{X}, \hat{Y}) &= \text{Cov}\left(\frac{Z}{n} \cdot \sum_{i=1}^n (x_i/z_i), \frac{Z}{n} \cdot \sum_{i=1}^n (y_i/z_i)\right) \\ &= (Z^2/n^2) \cdot \text{Cov}\left(\sum_{i=1}^n (x_i/z_i), \sum_{i=1}^n (y_i/z_i)\right). \end{aligned}$$

Si l'on pose, pour la commodité de l'écriture, $t_1 = x/z$ et $t_2 = y/z$, on a :

$$\text{Cov}(\hat{X}, \hat{Y}) = (Z^2/n^2) \cdot \text{Cov}\left(\sum_{i=1}^n (t_{1i}), \sum_{i=1}^n (t_{2i})\right) \text{ et, comme les}$$

tirages sont indépendants avec remise, les variables aléatoires t_{1i} et t_{2i} sont égales à t_1 et t_2 quelque soit i .

On obtient alors facilement :

$$\text{Cov}\left(\sum_{i=1}^n (t_{1i}), \sum_{i=1}^n (t_{2i})\right) = n \cdot \text{Cov}(t_1, t_2) = n \cdot \text{Cov}(x/z, y/z), \text{ d'ou,}$$

$$\text{Cov}(\hat{X}, \hat{Y}) = (Z^2/n) \cdot \text{Cov}(x/z, y/z). \quad (416.2 \text{ a})$$

$$\begin{aligned} \text{De même, } \underset{\pi}{\widehat{\text{Var}}(X)} &= \underset{\pi}{\text{Var}}\left(\frac{Z}{n} \cdot \sum_{i=1}^n (x_i / y_i)\right) = \frac{Z^2}{n^2} \cdot \underset{\pi}{\text{Var}}\left(\sum_{i=1}^n t_{1i}\right) \\ &= \frac{Z^2}{n} \underset{\pi}{\text{Var}}(t_1) = \frac{Z^2}{n} \underset{\pi}{\text{Var}}(x/z). \quad (416.2 \text{ b}) \end{aligned}$$

D'où l'expression du biais en fonction des variances, sous le plan π , des aléatoires x/z et y/z :

$$\underset{\pi}{E}(R - \hat{R}) = - \frac{Z^2}{nX^2} (\underset{\pi}{\text{Cov}}(x/z, y/z) - R \cdot \underset{\pi}{\text{Var}}(x/z)) \quad (416.2 \text{ c})$$

La formule (416.2 c) montre que si n tend vers l'infini, $E(R - \hat{R})$ tend vers zéro. \hat{R} est donc un estimateur asymptotiquement sans biais de R . A distance finie, le biais est un infiniment petit d'ordre $1/n$. Les biais sur les estimations du total et de la moyenne de y s'obtiennent en multipliant respectivement par X et par X/N , le biais sur l'estimation du ratio.

. Variance des estimations

La variance de l'estimateur du ratio s'écrit :

$$\underset{\pi}{V}(\hat{R}) = \underset{\pi}{E}(\hat{R} - R)^2 \approx \underset{\pi}{E}\left(\left(\frac{\hat{Y} - R\hat{X}}{X} \cdot \left(1 - \frac{\hat{X} - X}{X}\right)\right)^2\right).$$

Si l'on néglige le terme en $\left(\frac{\hat{X} - X}{X}\right)^2$, il reste à évaluer :

$$\underset{\pi}{V}(\hat{R}) \approx \underset{\pi}{E}\left(\left(\frac{\hat{Y} - R\hat{X}}{X}\right)^2\right) = \frac{1}{X^2} \cdot \underset{\pi}{E}(Y^2 + R^2 X^2 - 2RXY)$$

$$\text{D'où } \underset{\pi}{V}(\hat{R}) \approx \frac{1}{X^2} \cdot (\underset{\pi}{\text{Var}}(\hat{Y}) + R^2 \cdot \underset{\pi}{\text{Var}}(\hat{X}) - 2R \cdot \underset{\pi}{\text{Cov}}(\hat{X}, \hat{Y})).$$

En utilisant les relations 1.6.2 a et 1.6.2 b établies au paragraphe précédent, on obtient :

$$\underset{\pi}{V}(\hat{R}) = \frac{Z^2}{nX^2} \cdot (\underset{\pi}{\text{Var}}(y/z) + R^2 \cdot \underset{\pi}{\text{Var}}(x/z) - 2R \cdot \underset{\pi}{\text{Cov}}(x/z, y/z)) \quad (416.3 \text{ a})$$

Puis, pour l'estimation du total Y et de la moyenne y , respectivement :

$$\underset{\pi}{V}(T_4(Y)) = \frac{Z^2}{n} \cdot (\underset{\pi}{\text{Var}}(y/z) + R^2 \cdot \underset{\pi}{\text{Var}}(x/z) - 2R \cdot \underset{\pi}{\text{Cov}}(x/z, y/z)) \quad (416.3 \text{ b})$$

$$V(T_4(\bar{y})) = \sum_{\pi} (Z_h^2 / N n_h) \cdot (\sum_{\pi} \text{Var}_h(y/z) + R_h^2 \cdot \sum_{\pi} \text{Var}_h(x/z) - 2R_h \cdot \sum_{\pi} \text{Cov}_h(x/z, y/z))$$

(416.3 c).

. Allocation optimale

La fonction à minimiser est la variance totale, c'est à dire incluant le carré du biais, de l'estimateur stratifié qui s'écrit :

$$V_{tot} = \sum_{\pi} \sum_{h=1}^K (Z_h / n_h)^2 (\sum_{\pi} \text{Var}_h(y/z) + R_h^2 \cdot \sum_{\pi} \text{Var}_h(x/z) - 2R_h \cdot \sum_{\pi} \text{Cov}_h(x/z, y/z))$$

$$+ \sum_{h=1}^K (Z_h / n_h \cdot X_h) (\sum_{\pi} \text{Cov}_h(x/z, y/z) - R_h \cdot \sum_{\pi} \text{Var}_h(x/z))^2$$

La dérivation, par rapport aux n_h , de cette expression fournit les conditions suivantes. Pour tout h ,

$$\sum_{\pi} (Z_h / n_h)^2 (\sum_{\pi} \text{Var}_h(y/z) + R_h^2 \cdot \sum_{\pi} \text{Var}_h(x/z) - 2R_h \cdot \sum_{\pi} \text{Cov}_h(x/z, y/z))$$

$$+ (2Z_h / (n_h \cdot X_h)) (\sum_{\pi} \text{Cov}_h(x/z, y/z) - R_h \cdot \sum_{\pi} \text{Var}_h(x/z)) = c.$$

Le second terme, d'ordre $1/n_h^3$, peut être négligé par rapport au premier, d'ordre $1/n_h^2$. On obtient alors que, pour tout h , n_h doit être proportionnel à :

$$Z_h (\sum_{\pi} \text{Var}_h(y/z) + R_h^2 \cdot \sum_{\pi} \text{Var}_h(x/z) - 2R_h \cdot \sum_{\pi} \text{Cov}_h(x/z, y/z))^{1/2},$$

puis, en écrivant la contrainte $\sum n_h = n$, l'expression des n_h .

$$n_h = n \cdot \frac{Z_h (\sum_{\pi} \text{Var}_h(y/z) + R_h^2 \cdot \sum_{\pi} \text{Var}_h(x/z) - 2R_h \cdot \sum_{\pi} \text{Cov}_h(x/z, y/z))^{1/2}}{\sum_{h=1}^K Z_h (\sum_{\pi} \text{Var}_h(y/z) + R_h^2 \cdot \sum_{\pi} \text{Var}_h(x/z) - 2R_h \cdot \sum_{\pi} \text{Cov}_h(x/z, y/z))^{1/2}}$$

(416.4)

4.2. Le programme VARECH associé au logiciel dbaseIII

Pour obtenir, avec les plans de sondages et les estimateurs qui viennent d'être décrits, les résultats souhaités, il nous fallait réitérer les calculs de variance, pour l'ensemble du formulaire, sur plusieurs variables démographiques, plusieurs définitions de la probabilité et plusieurs types de stratifications. Il était donc nécessaire de disposer d'un outil informatique comprenant d'une part, un logiciel gestionnaire de fichiers, au sein duquel puisse être intégré et géré l'ensemble des données nécessaires et, d'autre part, d'un programme de calcul des variances des différents estimateurs, dans lequel soient paramétrés les choix des estimateurs, des probabilités et des stratifications. Nous donnons ici rapidement, les caractéristiques du matériel, de la base de données et du logiciel qui permet sa gestion, ainsi que l'architecture générale du programme, l'ensemble formant un outil portable sur d'autres sites ou à d'autres données. Le système demandera ensuite à être complété pour fournir, d'après les données recueillies par sondage, des estimations et leur précision.

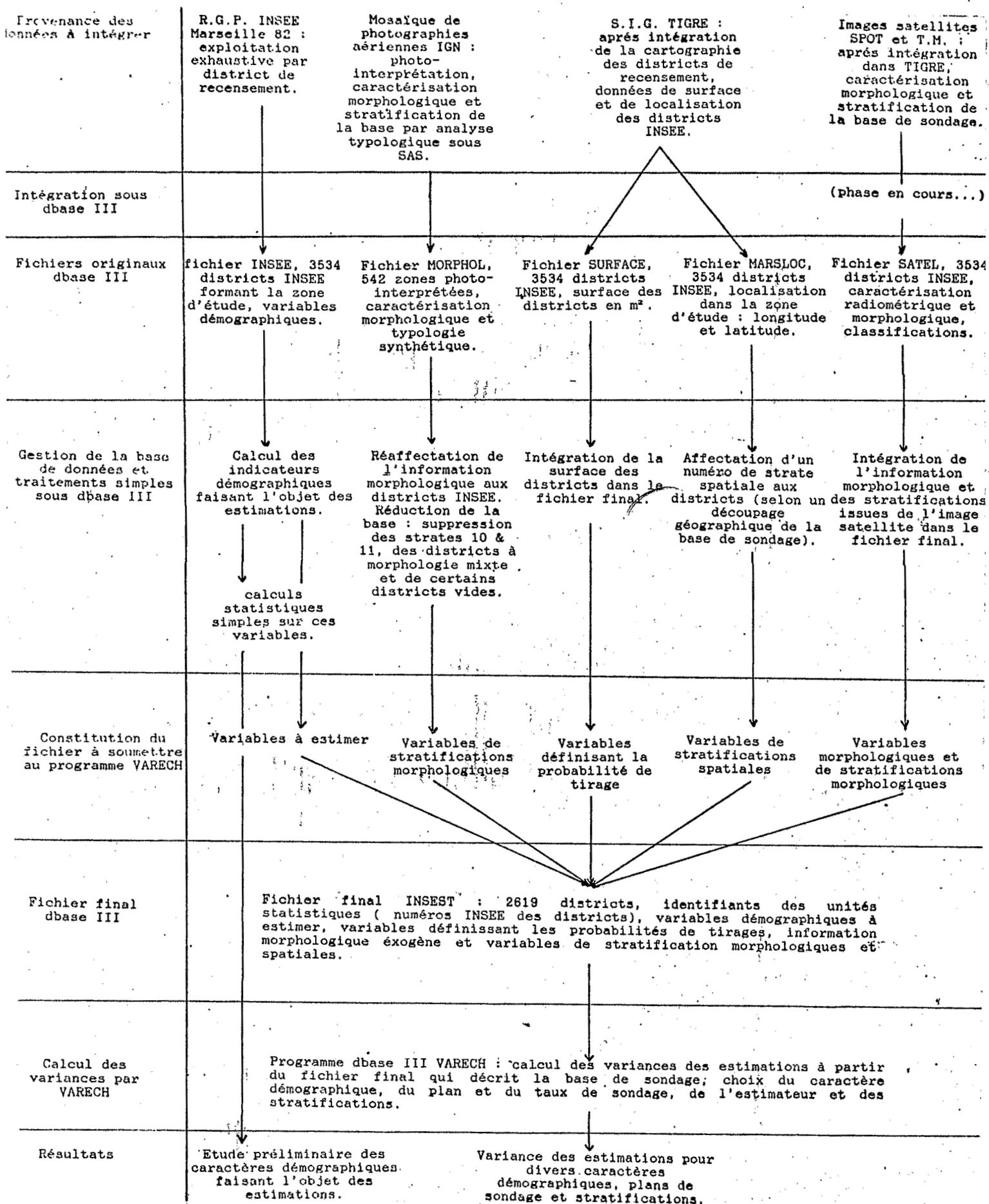
4.2.1 Matériel, système gestionnaire de fichiers et base de données sur Marseille.

Le matériel utilisé depuis notre arrivée à Quito est un micro ordinateur IBM PC-AT, équipé d'un disque dur de 20 Mega octets, l'ensemble géré par le système d'exploitation DOS (version 3.1). Sur ce système a été implanté le S.G.B.D. dbase III qui offre, outre de multiples facilités pour la gestion des fichiers, un langage de programmation intégré, à partir duquel a été développé le programme de calcul.

Pour disposer, sous dbase III, de la base de données qui sert de support aux tests, nous avons constitué un ensemble de fichiers correspondants à des données de provenance diverse : recensement général de la population de 1982 à Marseille, photo-interprétation d'une mosaïque de photographies aériennes couvrant la zone d'étude, données de surface et de localisation des districts INSEE de la base, obtenues grâce au système d'information géographique TIGRE, développé par l'unité d'infographie de l'ORSTOM (7). Par la suite, nous y intégrerons également un quatrième type de données : celles issues de l'interprétation des images satellites fournissant une caractérisation morphologique et des stratifications de la base de sondage. La phase d'intégration et de croisement de ces données est résumée dans la figure 1.

(7) : Voir, à propos des divers types de données utilisées, l'annexe 7 du rapport ATP CNRS/CNES d'Août 1987.

Figure 1 : l'intégration et les croisements de données sous dbase III



4.2.2 Le programme VARECH (VARIance d'EChantillonnage), architecture générale, fonctionnement et domaine d'application.

Une fois constitué, sous dbase III, le fichier contenant les informations nécessaires, le programme VARECH permet d'effectuer les calculs de variance correspondant à l'estimateur, au plan de sondage et à la stratification choisis. Ecrit sous dbase III, ce programme se présente comme un ensemble de commandes dbase, organisées en procédures qui calculent, pour un plan de sondage, un estimateur, une stratification et un taux donnés, la variance qu'aura l'estimation du total (ou de la moyenne) du caractère démographique considéré. L'organigramme général du programme est donné dans la figure 2, ainsi que la liste des procédures dont il est constitué (figure 3).

L'ensemble des informations nécessaires au programme doivent se trouver dans un même fichier dbase III, appelé fichier de données. Une fois le calcul effectué, le programme écrit leurs résultats dans un fichier que l'on devra éditer ensuite.

Après avoir entré le nom du fichier résultat, l'utilisateur fait, grâce au menu principal, le choix du plan de sondage et de l'estimateur; puis, dans le menu secondaire, il choisit l'estimation, stratifiée ou non, d'un total ou d'une moyenne. L'utilisateur indique ensuite au programme, le nom du fichier de données et les noms des champs contenant les différentes variables : variable à estimer, variable exogène figurant au dénominateur (dans le cas de l'estimation d'un ratio), variable définissant la probabilité (dans le cas d'un plan à probabilités inégales), variable de stratification si il y a lieu. Il doit également choisir un taux de sondage et préciser le nombre de strates que contient la base de sondage. Dans sa version finale, le programme comprendra une procédure d'édition du fichier résultat qui permettra de sélectionner et d'imprimer les résultats. Des exemples de déroulement du programme sont donnés sous forme de copies d'écrans dans la figure 4.

Soulignons que le programme actuel est à l'état de maquette. Pourtant on peut, dès maintenant, donner le domaine d'application qu'auront les différentes versions opérationnelles :

- Dans son état de développement actuel, VARECH permet un travail d'évaluation, à partir de données exhaustives existant sur la base de sondage choisie (celles-ci proviennent généralement de recensements), des possibilités d'estimation par sondage de ces données. Intégrant diverses possibilités quant aux choix de plans de sondage, estimateurs et stratifications, il fournit la précision réelle des estimations et conduit ainsi à la sélection d'une ou plusieurs

Figure 2. : Organisation générale du programme VARECH.

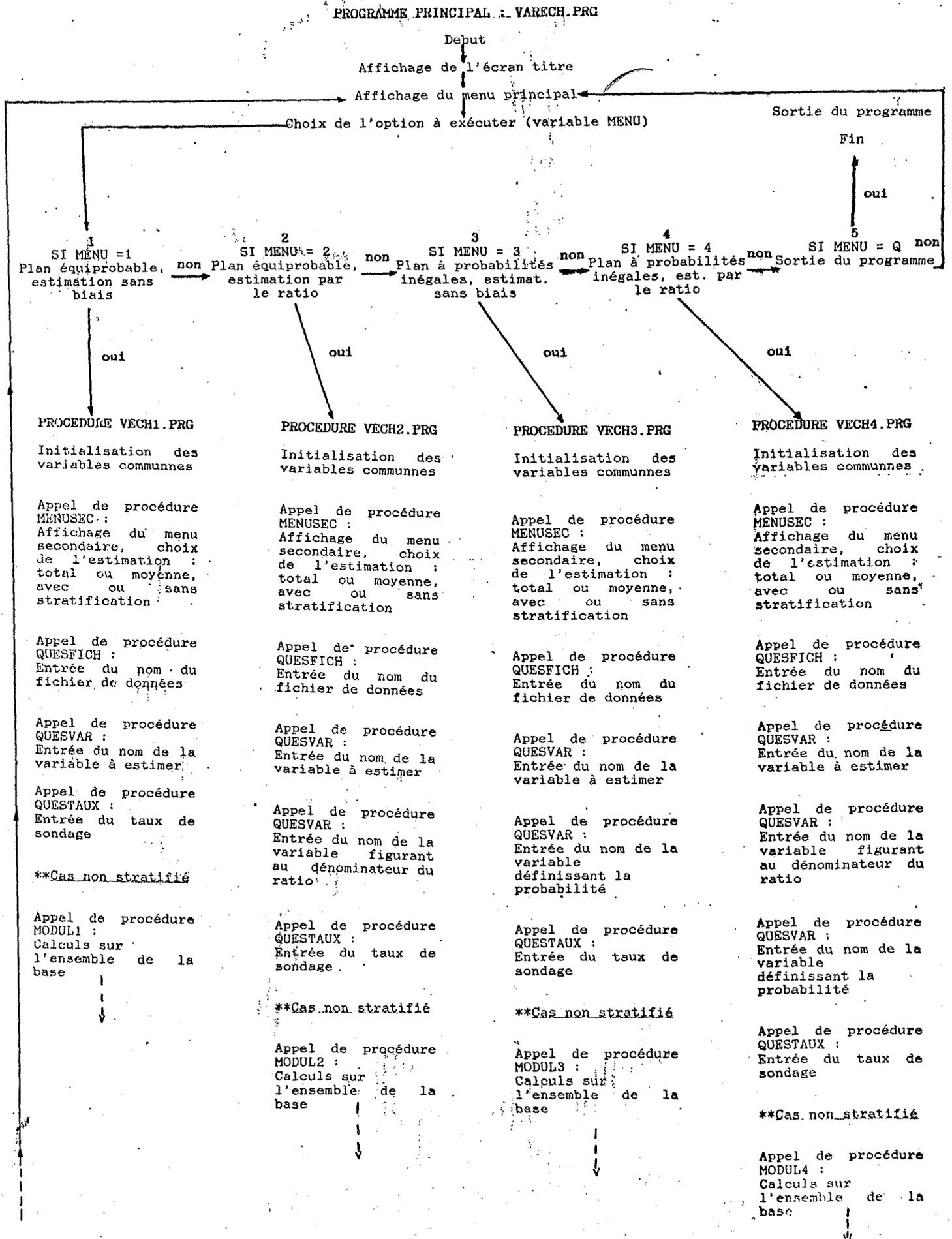


Figure 2. (suite) : Organisation générale du programme VARECH.

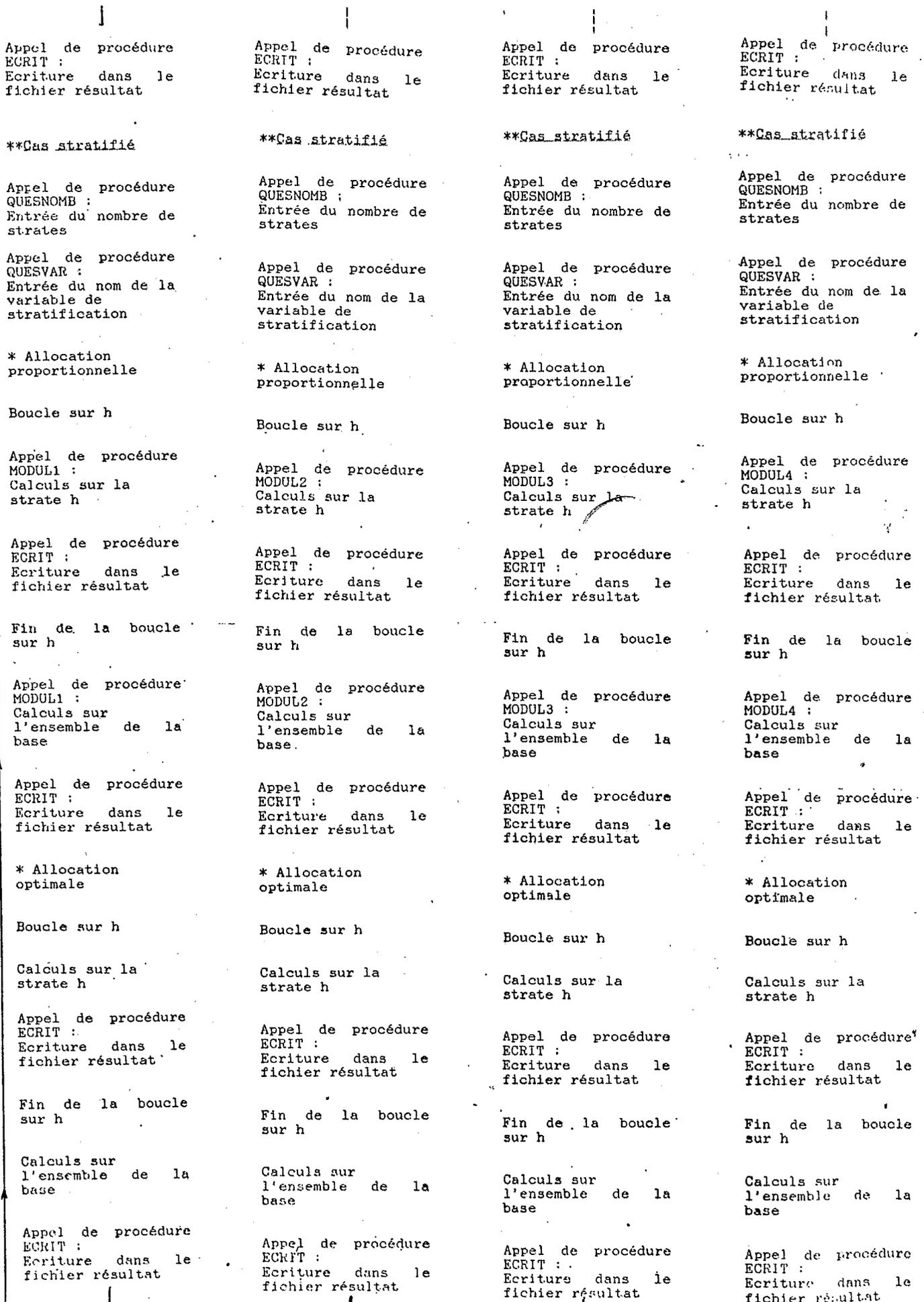


Figure 3. : Liste des procédures du programme VARECH.

Niveaux	Noms	Fonctions
-1- Programme principal	VARECH	Programme principal : affichage de la page de titre, menu principal, appel de procédures principales (VECH1, VECH2, VECH3, VECH4), sortie du programme.
-2- Procédures principales calcul des variances	VECH1	Procédure de calcul des variances : cas du plan équiprobable, estimation sans biais. Appel de la procédure MODUL1 et des procédures utilitaires.
	VECH2	Procédure de calcul des variances : cas du plan équiprobable, estimation par le ratio. Appel de la procédure MODUL2 et des procédures utilitaires.
	VECH3	Procédure de calcul des variances : cas du plan à probabilités inégales, estimation sans biais. Appel de la procédure MODUL3 et des procédures utilitaires.
	VECH4	Procédure de calcul des variances : cas du plan à probabilités inégales, estimation par le ratio. Appel de la procédure MODUL4 et des procédures utilitaires.
-3.1- Procédures de calculs pour les procédures principales	MODUL1	Calcul des variances, écarts type, coefficients de variation pour VECH1. Appel de CALVECH1.
	CALVECH1	Sommation pour le calcul de l'effectif de la base et de la variance du caractère Y sur la base, pour MODUL1. Probabilités uniformes.
	MODUL2	Calcul des variances, écarts type, coefficients de variation pour VECH2. Appel de CALVECH2.
	CALVECH2	Sommation pour le calcul de l'effectif de la base et de la variance des caractères Y et X sur la base, pour MODUL2. Probabilités uniformes.

Figure 3. (suite) : Liste des procédures du programme VARECH.

Niveaux	Noms	Fonctions
	MODUL3	Calcul des variances, écarts type, coefficients de variation pour VECH3. Appel de CALVECH3.
	CALVECH3	Sommation pour le calcul de l'effectif de la base et de la variance des caractères Y et Z sur la base, pour MODUL3. Probabilités inégales, proportionnelles à Z.
	MODUL4	Calcul des variances, écarts type, coefficients de variation pour VECH4. Appel de CALVECH4.
	CALVECH4	Sommation pour le calcul de l'effectif de la base et de la variance des caractères Y, X et Z sur la base, pour MODUL4. Probabilités inégales, proportionnelles à Z.
-3.2- Procédures utilitaires pour les procédures principales	MENUSEC	Procédure du menu secondaire pour les procédures principales : affichage du menu secondaire, choix de l'estimation (stratifiée, non stratifiée, estimation d'un total ou d'une moyenne).
	QUESFICH	Procédure d'entrée d'un nom de fichier dans le programme VARECH.
	QUESVAR	Procédure d'entrée d'un nom de variable dans le programme VARECH.
	QUESNOMB	Procédure d'entrée du nombre de strates dans le programme VARECH.
	QUESTAUX	Procédure d'entrée du taux de sondage dans le programme VARECH.
	QUESNEYM	Procédure de choix de l'allocation optimale de NEYMAN dans le programme VARECH.
	ECRIT	Procédure d'écriture des résultats de calcul du programme dans le fichier résultat.

Figure 4. : Déroulement du programme VARECH.

Varech, version 5.

Jeudi 25 Février 1988

ECRAN 1

MENU PRINCIPAL

- <1> : Plan équiprobable, estimateur sans biais.
- <2> : Plan équiprobable, estimateur d'un/par le ratio.
- <3> : Plan à probabilités inégales, estimateur sans biais.
- <4> : Plan à probabilités inégales, estimateur d'un/par le ratio.
- <Q> : Quitter le programme

Choisissez une option :1

Jeudi 25 Février 1988

ECRAN 2

MENU SECONDAIRE (2)

- <1> : Estimation non stratifiée.
- <2> : Estimation stratifiée.

Choisissez une option :2

Jeudi 25 Février 1988

ECRAN 3

MENU SECONDAIRE (1)

- <1> : Estimation d'un total.
- <2> : Estimation d'une moyenne.
- <R> : Retour au menu principal.

Choisissez une option :1

Figure 4. (suite) : Déroulement du programme VARECH.

ECRAN 4

FICHER DE DONNEES

 Entrez le nom du fichier, sans extention ni chemin d'accès.
 (8 caractères maximum).
 il doit être dans le repertoire dbaseiii du disque dur.

nom : QMANEST1
 VARIABLE A ESTIMER

Entrez le nom de la variable. (12 caractères maximum) : EFFECTRES
 VARIABLE EXOGENE AU DENOMINATEUR DU RATIO

Entrez le nom de la variable. (12 caractères maximum) : NBMENAGE
 VARIABLE EXOGENE DEFINISSANT LA PROBABILITE

Entrez le nom de la variable. (12 caractères maximum) : SURFACE

PLAN EQUIPROBABLE, ESTIMATION BIAISEE : ECRAN 5
 taux 5
 Estimation stratifiée fichier : QMANEST1.DBF

Estimation du quotient : EFFECTRES/NBMENAGE

STRATE N° : 1

Calcul en cours, ... patientez.

techniques de sondage adaptées au site et au type des données recherchées.

Il s'agit donc, à l'heure actuelle, d'un outil de mise au point méthodologique, qui trouve sa principale utilité lors d'opérations d'évaluation des possibilités d'utilisation des techniques de sondage.

- Les développements futurs devront conduire à des versions assurant ou facilitant la mise en pratique de la méthode d'échantillonnage et d'estimation sélectionnée. Ces versions comprendront des procédures de sélection (ou d'aide à la sélection) des unités de l'échantillon, d'extrapolation des résultats du sondage aux domaines d'études et d'estimation des précisions.

Le domaine d'utilisation de la version finale s'étendra donc de l'évaluation à la mise en oeuvre de méthode de sondage, sur des sites où existent des données permettant une approche rigoureuse des choix techniques à opérer.

CHAPITRE 5

ESTIMATIONS STRATIFIEES DE CARACTERES DEMOGRAPHIQUES :

RESULTATS

Introduction

Ce chapitre présente les résultats de l'application du programme VARECH à une série d'indicateurs démographiques issus des chiffres du recensement général de la population de 1982 à Marseille.

Comme il a été dit en avertissement, les résultats dont nous disposons au moment de la rédaction de ce rapport ne portent que sur les trois premiers estimateurs considérés dans le chapitre précédent. La comparaison effectuée ici n'inclue donc pas l'estimation par le ratio sous plan à probabilités inégales ; ces conclusions sont susceptibles d'être révisées lorsque le test aura été complété. Cependant, nous ne pensons pas que l'introduction de probabilités inégales, proportionnelles aux surfaces des districts, modifie énormément la précision des estimations de ratios. En effet, comme nous allons le voir maintenant, l'introduction d'une telle probabilité pour l'estimation sans biais de totaux, change relativement peu de choses.

En conclusion, on compare les précisions obtenues avec les différents plans de sondage (plans équiprobables et plan à probabilité proportionnelle à la taille), on opère une sélection parmi les estimateurs, enfin on examine les précisions obtenues pour les différents caractères démographiques et les gains qu'apporte, dans chaque cas, l'introduction d'une stratification morphologique.

5.1. comparaison de diverses estimations d'une même variable.

Le tableau 1.1 fournit les coefficients de variation des trois estimateurs dans les cas stratifiés et non stratifiés, pour l'ensemble de la base et pour chacune des neuf strates morphologiques. Le graphique 1.2 visualise ces résultats pour l'estimateur sans biais sous plan à probabilités inégales, dans le cas de l'allocation optimale.

La conclusion à tirer de la comparaison des trois estimateurs tient en quatre points.

1. La surface des unités statistiques (districts INSEE), prise ici comme mesure de leur taille, n'est pas suffisamment corrélée à l'effectif de population pour que son introduction dans la formule d'estimation par l'intermédiaire du ratio P/S, améliore la précision. Au contraire, cet estimateur s'avère moins précis que les deux autres, tant pour l'ensemble de la base, qu'au niveau de chaque strate. Il sera donc écarté sous cette forme et nous chercherons, pour la suite, une meilleure mesure de la taille des unités.

2. L'estimation sans biais se fait avec une précision équivalente sous les deux plans de sondage considérés : plan équiprobable et plans à probabilités proportionnelles à la surface des districts. Pour une estimation non stratifiée, il y a un léger avantage au premier (5,75 % de la variance en moins), mais cet avantage devient négligeable dans le cas stratifié avec l'allocation optimale (1,3 % de la variance). L'équivalence des deux plans est un résultat important pour la suite puisqu'en pratique, le tirage de l'échantillon sur documents issus de l'image satellite nous contraint à utiliser un plan à probabilités inégales. Il reste, pour conclure cette phase de l'expérience, à tester un plan où les probabilités soient déterminées par la surface bâtie dans l'unité et non plus par sa surface totale. Une mesure exacte de cette surface est certainement impossible sur l'image satellite. Néanmoins l'obtention d'un coefficient qui permette de l'approcher dans chaque strate, alliée à une technique de tirage des unités de l'échantillon qui garantisse une probabilité approximativement égale à leur surface bâtie, devrait déjà apporter aux estimations, un gain de précision sensible.

3. Les domaines de précision et d'imprécision des estimations dans l'espace urbain, tels qu'ils ont été identifiés lors de l'examen des premiers résultats (annexe 2), demeurent identiques quels que soient le plan de sondage et l'estimateur retenus :

- centre ville, grands ensembles et pavillonnaire (strates 1,3 et 5) : domaine de bonne précision où les intervalles de confiance sont compris entre +/- 20% et +/- 40%

- les strates 7 et 8, "reliquat morphologique" et "grands bâtiments élevés", malgré leur faibles effectifs, conservent une précision acceptable

- le domaine d'imprécision couvre les strates 2, 4, 6 et 9, c'est à dire les quartiers industriels, les petits immeubles d'habitation et les zones périphériques peu denses ou non bâties. Ici les intervalles de confiances dépassent +/- 100%, ce qui rend très hasardeux toute estimation. Notons que le résultat catastrophique obtenu dans la strate 9 s'explique en partie par son effectif trop faible : 7 districts. D'autre part ne perdons pas de vue le fait que ces strates sont peu peuplées dans leur ensemble.

Tableau 1.1 : Résultats du test mené avec le programme VARECH. Estimation du total de population sans doubles comptes : deux plans de sondage, trois estimateurs, neuf strates morphologiques synthétiques, taux de 6,6% appliqué a une base de 2619 districts (R.G.P. Marseille 1982).

Estimations	(1)		(2)		(3)	
	*	**	*	**	*	**
Strates						
1. Centre ville dense	9.90	10.85	9.71	10.79	10.42	11.96
2. Quartiers industriels	55.71	49.66	76.04	58.21	70.51	60.43
3. Grands ensembles d'habitation	21.06	13.92	21.61	14.15	24.46	16.07
4. Petits immeubles d'habitation	69.30	69.91	57.64	63.98	82.45	84.86
5. Pavillonnaire	16.02	23.17	14.25	21.93	21.85	29.40
6. Péri-urbain peu dense	59.81	107.45	113.11	148.27	125.34	214.74
7. Reliquat morphologique	37.14	32.91	31.70	30.51	33.38	33.56
8. Dense : grands bâtiments élevés	38.04	33.24	34.41	31.72	69.21	48.49
9. Non bâti	370.09	598.17	320.98	559.00	454.80	912.63
Ensemble de la base stratifiée	8.83	7.62	9.08	7.67	11.35	8.92
Ensemble de la base non stratifiée.	9.84		10.13		13.67	

(1) : Estimation sans biais, plan équiprobable.

(2) : Estimation sans biais, plan à probabilités inégales.

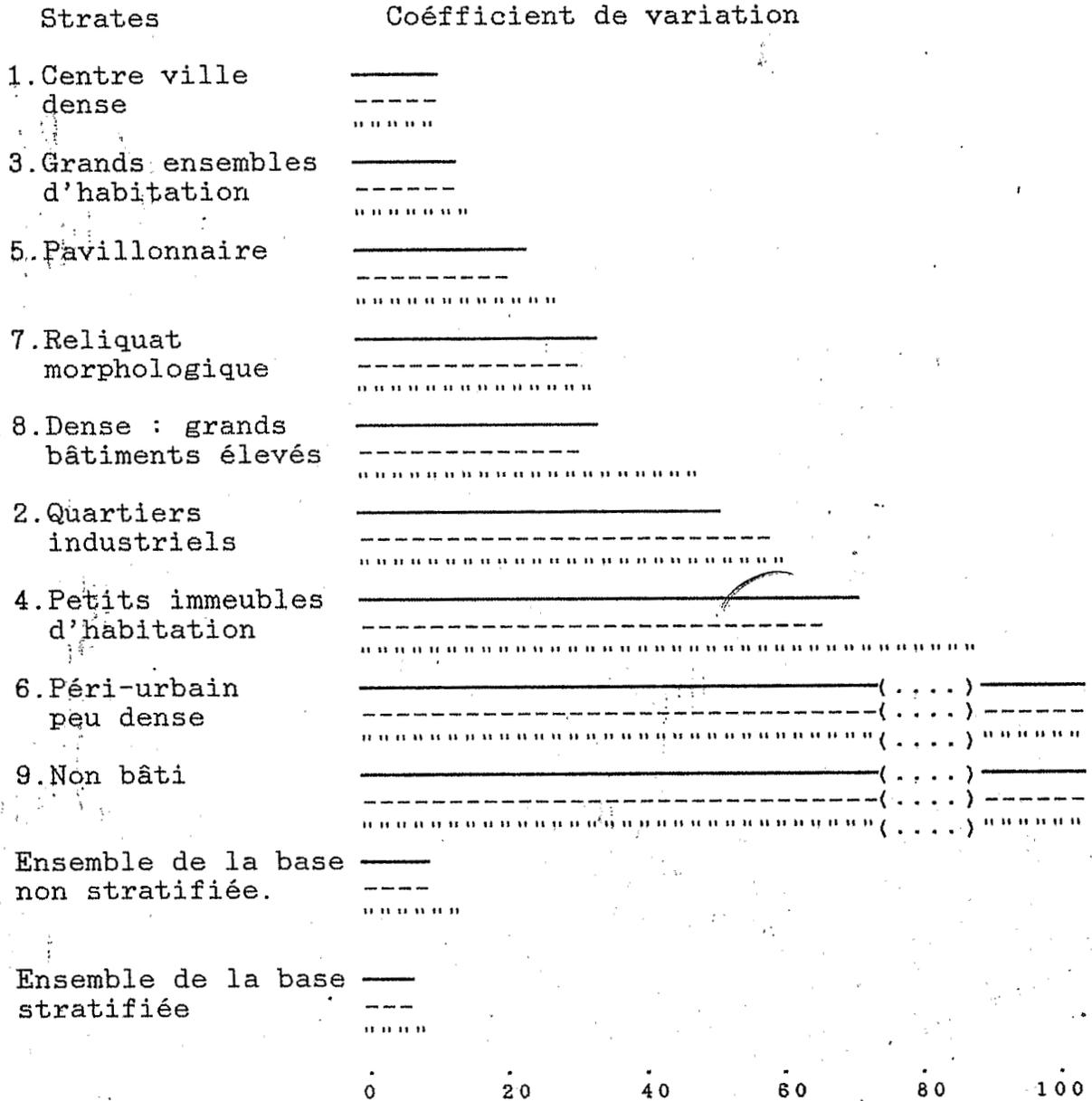
(3) : Estimation par le ratio, plan équiprobable.

* : Allocation proportionnelle

** : Allocation optimale

Graphique 1.2 : Résultats du test mené avec le programme VARECH

Estimation du total de population sans double comptes.
 Comparaison graphique des coefficients de variation pour les trois techniques de sondages (allocation optimale).



_____ : Estimation sans biais, plan équiprobable.
 ----- : Estimation sans biais, plan à probabilités inégales.
 " " " " " " : Estimation par le ratio, plan équiprobable.

4. Les gains qu'apportent la stratification morphologique puis, l'allocation optimale, sont également toujours du même ordre. Ils sont consignés dans le tableau suivant qui donne la diminution de la variance en pour-cent de cette dernière dans le cas non stratifié.

	Plan équiprob. est. sans biais	Plan équiprob. est. du ratio	Plan à proba. inégale est. sans biais
estimation stratifiée à allocation proport.	19,4 %	31,2 %	19,7 %
estimation stratifiée à allocation optimale	40,0 %	57,5 %	42,6 %

Soulignons, au vu du tableau, l'équivalence des gains dans les deux cas : équiprobable et probabilités inégales. Ce résultat confirme l'intérêt de la stratification morphologique dans le cadre d'une technique d'échantillonnage d'après document satellitaire, technique qui contraint à pratiquer un plan à probabilités inégales.

5.2. Comparaison des estimations de plusieurs caractères démographiques

Disposant d'un outil de test qui fournit les précisions associées à différents plans de sondage, estimateurs et stratifications, nous l'appliquons maintenant à divers indicateurs démographiques. Deux questions principales sous-tendent ce travail :

1. Existe-t-il des types de variables démographiques qui, du fait des caractéristiques de leur distribution (au sens statistique comme au sens spatial), se prêtent mieux que d'autres aux estimations par sondage ?

2. Les choix des probabilités de tirage, des formules d'estimation et des stratifications, qui définissent le sondage, ont ils les mêmes conséquences sur la précision des estimations, quelles que soient les variables considérées ?

Pour tenter d'y répondre, nous avons soumis au programme le fichier contenant, pour les 2619 districts de la base, la donnée de 28 indicateurs démographiques, dont la liste figure en fin d'annexe, dans le tableau 2.1. Parmi eux, deux catégories principales sont à distinguer. D'une part des mesures d'effectifs comme le nombre d'individus ou de ménages du district répondant à une certaine caractéristique (15 variables), d'autre part les ratios mesurant l'importance relative au sein du district de catégories d'individus ou de ménages (14 variables).

Nous nous intéresserons, tout d'abord, aux estimations concernant l'ensemble de la base, successivement, estimations d'effectifs et de ratios, puis, après sélection de 15 caractères parmi les 28 de départ, aux estimations partielles dans chacune des strates. Enfin, en conclusion, nous tenterons de donner réponse aux deux questions formulées ci-dessus.

5.2.1. Résultats sur l'ensemble de la base : estimateurs et caractères démographiques

1) Estimation d'effectifs

Le tableau 2.2 donne, pour les 15 variables retenues, les coefficients de variation ($C = \text{écart type}/\text{quantité à estimer}$, en %) des trois estimateurs non stratifiés et stratifiés de l'effectif total sur l'ensemble de la base (dans le cas stratifié, c'est l'allocation optimale qui est choisie). La dernière colonne mesure le gain de précision dû à la stratification par le pourcentage de variance qu'elle permet d'éliminer. Enfin le graphique 2.3 fournit une traduction visuelle des précisions obtenues avec chacun des trois sondages, dans le cas de l'allocation optimale. De l'ensemble des résultats se dégagent quatre conclusions :

1. La précision des estimations est du même ordre pour la plupart des variables considérées, avec des coefficients de variation de 7 à 8% lorsqu'on choisit le meilleur estimateur et que l'allocation aux strates est optimale. Il n'y a d'ailleurs rien d'étonnant à cela puisque les variables considérées sont, pour la plupart d'entre elles, fortement corrélées. Notons tout de même que, lorsqu'il s'agit d'estimer les totaux de catégories nettement minoritaires, la précision se dégrade assez sensiblement. C'est le cas pour la population de nationalité étrangère, celle des hommes de plus de 19 ans au chômage ou encore pour le nombre de ménages de plus de six personnes et, dans une moindre mesure, pour la population des moins de 5 ans. Les coefficients de variation les meilleurs sont alors compris entre 11 et 21%.

2. Comme le montre le graphique 2.3, en général, les deux estimateurs sans biais ont une précision équivalente alors que l'estimation de ces totaux par le ratio faisant intervenir la surface des districts est moins précise. Ceci confirme l'idée déjà exprimée selon laquelle la surface totale des districts ne constitue pas une mesure efficace de la taille des unités lorsqu'on s'intéresse à l'estimation de caractères démographiques.

3. On constate également sur le graphique 2.3 que lorsque deux catégories de population sont complémentaires, au sein d'une population totale (populations française et étrangère par exemple), les estimations concernant la catégorie la plus nombreuse sont généralement les plus précises. L'exception est ici la population des plus de 60 ans, dont l'estimation est plus précise que celle des moins de 60 ans. Nous verrons plus loin que ce phénomène s'accroît lors de l'estimation des ratios correspondants.

4. Enfin, l'examen de la dernière colonne du tableau 2.2 permet de chiffrer à environ 40 % la diminution moyenne de la variance due à l'introduction de la stratification morphologique. Ce gain peut atteindre 50 à 55 % dans le cas des caractères dont la variance inter-strate est importante : population étrangère, hommes de plus de 19 ans au chômage, nombre de ménages d'une personne. A l'inverse, il n'est que de 20 % lors de l'estimation de la population âgée de 60 ans et plus.

Tableau 2.2 : Résultats du test mené avec le programme VARECH

Coefficients de variation des estimations stratifiées et non stratifiées de 15 caractères démographiques (effectifs).
Mesure du gain dû à la stratification.

Estimateurs:	(1)		(2)		(3)		(4) Gain de précision
	*	**	*	**	*	**	
Caractères estimés :	*	**	*	**	*	**	
Population totale	9.83	7.62	10.15	7.67	13.67	9.95	42.9
Population totale des ménages ordinaires	9.88	7.64	10.25	7.76	13.77	9.02	42.7
Population de nationalité étrangère	30.81	21.39	37.83	26.58	32.62	22.32	50.6
Population de nationalité française	9.59	7.59	9.43	7.18	13.39	8.78	42.0
Population des hommes de plus de 19 ans au chômage	14.92	12.34	22.04	14.64	18.25	14.71	55.9
Population des hommes de plus de 19 ans ayant un emploi	9.19	7.61	10.87	8.03	13.56	9.01	45.4
Population âgée de moins de cinq ans	16.27	10.73	14.76	11.67	18.12	11.68	37.5

(1) : Plan équiprobable, estimateur sans biais.

(2) : Plan à probabilités inégales, estimateur sans biais.

(3) : Plan équiprobable, estimateur par le ratio.

(4) : Gain de précision dû à la stratification, en % de la variance non stratifiée.

* : Estimation non stratifiée.

** : Estimation stratifiée, allocation aux strates optimale.

Tableau 2.2 (suite) : Résultats du test mené avec le programme VARECH

Coefficients de variation des estimations stratifiées et non stratifiées de 15 caractères démographiques (effectifs).
Mesure du gain dû à la stratification.

Estimateurs:	(1)		(2)		(3)		(4) Gain de précision
	*	**	*	**	*	**	
Caractères estimés :							
Population âgée de cinq ans et plus	9.66	7.56	10.07	7.59	13.60	8.87	43.3.
Population âgée de soixante ans et plus	8.57	7.45	8.75	7.76	14.12	9.07	21.3
population âgée de moins de soixante ans	11.02	8.15	10.66	8.21	14.15	9.30	40.7
Nombre de ménages ordinaires	8.70	7.20	10.30	7.43	13.66	8.74	48.0
Nombre de ménages d'une personne	9.44	7.98	12.81	8.70	15.21	9.51	53.9
Nombre de ménages de deux personnes ou plus	9.26	7.43	9.97	7.53	13.50	8.89	43.0
Nombre de ménages de six personnes ou plus	29.98	17.59	24.43	18.88	30.40	18.11	40.5
Nombre de ménages de moins de six personnes	8.65	7.22	10.36	7.46	13.73	8.78	48.1

(1) : Plan équiprobable, estimateur sans biais.

(2) : Plan à probabilités inégales, estimateur sans biais.

(3) : Plan équiprobable, estimateur par le ratio.

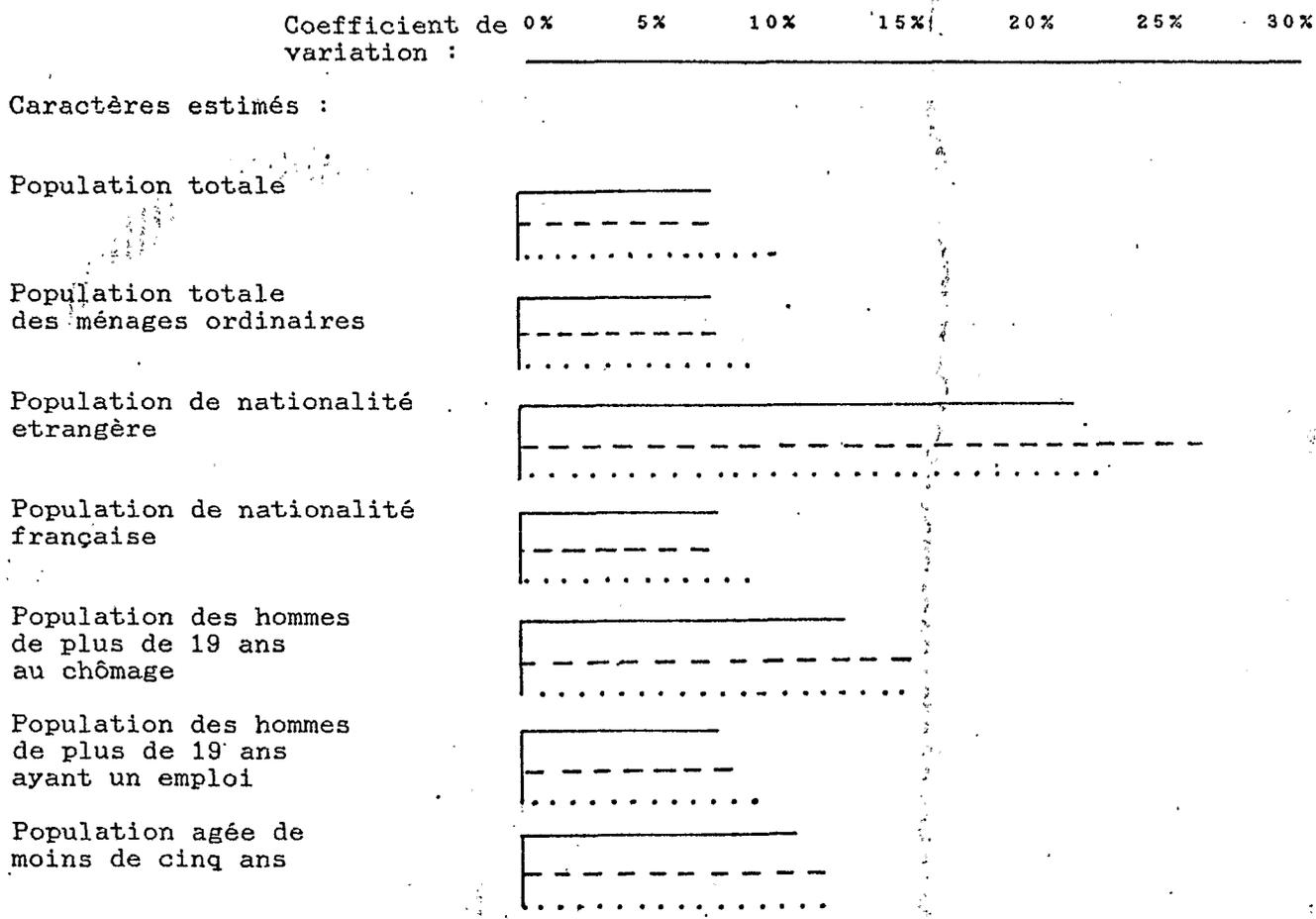
(4) : Gain de précision dû à la stratification, en % de la variance non stratifiée.

* : Estimation non stratifiée.

** : Estimation stratifiée, allocation aux strates optimale.

Graphique 2.3 : Résultats du test mené avec le programme VARECH

Comparaisons, pour trois sondages, des coefficients de variation des estimateurs stratifiés de 15 mesures d'effectifs. Allocation optimale aux strates.



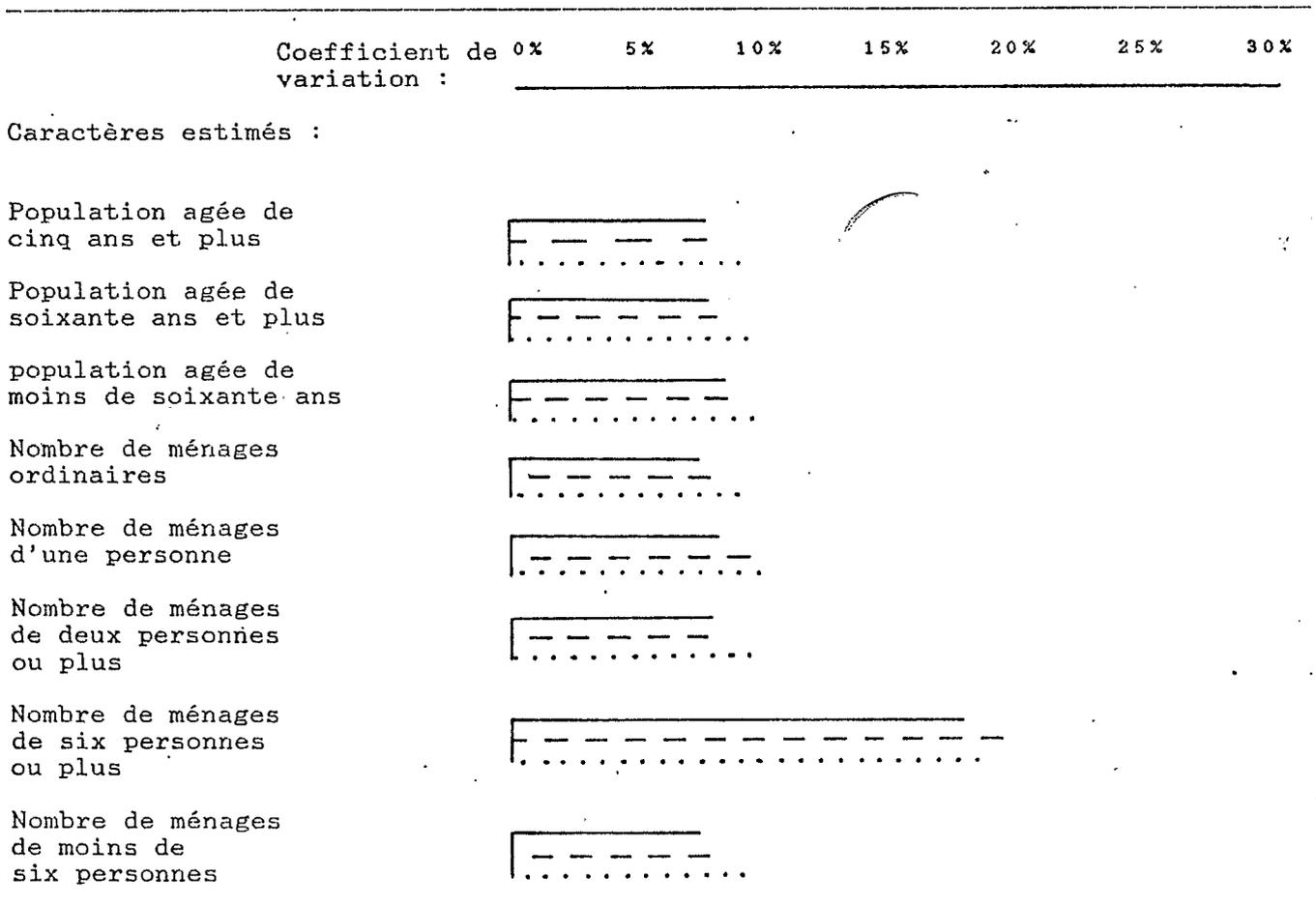
————— : Plan équiprobable, estimateur sans biais.

- - - - : Plan à probabilités inégales, estimateur sans biais.

..... : Plan équiprobable, estimateur par le ratio.

Graphique 2.3 (suite) : Résultats du test mené avec le programme VARECH

Comparaisons, pour trois sondages, des coefficients de variation des estimateurs stratifiés de 15 mesures d'effectifs. Allocation optimale aux strates.



_____ : Plan équiprobable, estimateur sans biais.
 . - - - : Plan à probabilités inégales, estimateur sans biais.
 : Plan équiprobable, estimateur par le ratio.

2) Estimations de ratios

Si l'estimateur par le ratio n'a pas, comme on l'a montré, amélioré les estimations d'effectifs, son intégration dans le programme conserve néanmoins un grand intérêt puisqu'elle permet d'estimer directement les valeurs des quotients décrivant les structures de la population. Le tableau 2.4 fournit les coefficients de variation de ces estimations et une mesure du gain de précision obtenu grâce à la stratification pour 13 de ces quotients. Son examen conduit à trois conclusions :

1. Ce qui frappe tout d'abord, c'est que dans l'ensemble, les coefficients de variation sont de beaucoup inférieurs à ceux qu'on obtenait lors de l'estimation d'effectifs : sept des ratios sont estimés avec des coefficients inférieurs à 2,5 % et on atteint même, dans le cas du quotient population de 4 ans et plus / population totale, 0,29 %. Ce résultat ne doit pas surprendre : ces ratios ont, par nature, une variabilité bien moindre que celle des effectifs. Nous verrons d'ailleurs plus loin que, même avec des précisions de cet ordre, le sondage ne donne qu'une image floue des différences qui existent, à l'intérieur de l'agglomération, entre les strates morphologiques.

2. A nouveau, et de manière plus systématique encore quand il s'agit d'estimer un quotient, si le numérateur est une partie nettement minoritaire de la population décrite au dénominateur, mieux vaut alors estimer le quotient complémentaire. Par exemple, le quotient "population de nationalité française / population totale", est estimé plus précisément que le quotient "population étrangère / population totale". Le gain ainsi réalisé s'échelonne, suivant les cas, d'un facteur 2 à un facteur 32 sur le coefficient de variation ; plus la catégorie du numérateur est minoritaire, plus le gain obtenu en passant à son complémentaire est important. Bien sûr, le désavantage dont souffrent ici les catégories minoritaires n'est en fait qu'apparent puisque, dans le cas des ratios, la précision obtenue sur l'estimation du "ratio majoritaire" s'étend tout à fait légitimement au ratio complémentaire : dans l'exemple cité, l'estimation du ratio population étrangère / population totale profitera de la même précision que celle du ratio population française / population totale.

Pour les habitués des techniques de sondages, ces résultats vont peut être de soi, mais étant donnée leur réelle importance pratique, nous avons considéré leur exposé utile.

3. La diminution de variance due à la stratification est, en moyenne, légèrement supérieure à celle obtenue lors des estimations d'effectifs : entre 30 et 60 % de la variance non stratifiée.

Tableau 2.4 : Résultats du test mené avec le programme VARECH

Coefficients de variation des estimations stratifiées et non stratifiées de 13 caractères démographiques (ratios). Mesure du gain dû à la stratification.

Quotients estimés :	Estimation non stratifiée	Estimation stratifiée allocation optimale	Gain de précision en % de la var. non stratifiée
Population de nationalité étrangère/ Population totale	27.02	19.02	50.0
Population de nationalité française/ Population totale	3.01	2.12	50.4
Population des hommes de plus de 19 ans au chômage/ Population des hommes de plus de 19 ans	9.89	8.31	29.4
Population des hommes de plus de 19 ans ayant un emploi/ Population des hommes de plus de 19 ans	0.99	0.82	31.4
Population âgée de moins de cinq ans/ Population totale	9.71	6.06	61.1
Population âgée de cinq ans ou plus/ Population totale	0.47	0.29	61.9
Population âgée de soixante ans ou plus/ Population totale	6.33	4.10	58.0

Tableau 2.4 (suite) : Résultats du test mené avec le programme VARECH

Coefficients de variation des estimations stratifiées et non stratifiées de 13 caractères démographiques (ratios). Mesure du gain dû à la stratification.

Quotients estimés :	Estimation non stratifiée	Estimation stratifiée allocation optimale	Gain de précision en % de la var. non stratifiée
Population âgée de moins de soixante ans/ Population totale	2.06	1.34	57.7
Population des ménages ordinaires/ Nombre de ménages ordinaires	3.72	2.28	62.3
Nombre de ménages d'une personne/ Nombre de ménages ordinaires	4.90	3.59	46.3
Nombre de ménages de plus d'une personne/ Nombre de ménages ordinaires	2.28	1.67	46.4
Nombre de ménages de six personnes ou plus/ Nombre de ménages ordinaires	18.99	16.14	27.8
Nombre de ménages de moins de six personnes/ Nombre de ménages ordinaires	0.86	0.50	66.2

5.2.2. Résultats par strates : variabilité des caractères démographiques et précision des estimations partielles.

Les conclusions de l'examen des résultats sur l'ensemble de la base permettent, en écartant les variables décrivant des catégories très minoritaires au profit de leurs complémentaires, de ne retenir, pour l'examen plus détaillé des résultats par strates, que huit effectifs et six ratios, qui constituent un échantillon restreint mais significatif de caractères démographiques. A travers cet échantillon, outre la mesure de l'effectif total de population (au sens de la population totale (1)), quatre thèmes seront abordés.

- La population étrangère : effectif et ratio population française / population totale.
- Le chômage : effectif de la population des hommes de plus de 19 ans ayant un emploi et ratio par rapport à la population totale.
- La structure par âge de la population totale : effectifs des populations de 5 ans et plus et de moins de 60 ans, ainsi que les ratios correspondant par rapport à la population totale.
- La taille des ménages : nombre de ménages ordinaires, nombre de ménages de plus d'une personne et nombre de ménages de moins de six personnes, taille moyenne des ménage, c'est à dire le quotient population des ménages ordinaires / nombre de ménages ordinaires et enfin les deux ratios donnant la proportion de ménages de plus d'une personne et de moins de six personnes.

Le tableau 2.5 donne, pour chaque caractère démographique, le coefficient de variation sur l'ensemble de la base et dans chacune des neuf strates morphologiques (estimateurs sans biais sous plan à probabilités inégales pour les effectifs, estimateurs de ratios sous plan équiprobable pour les quotients). L'examen des chiffres conduit aux conclusions suivantes :

1. Les domaines de précision (strates 1,3,5,7 et 8) et d'imprécision (strates 2, 4, 6 et 9), sont les mêmes quelque soit le caractère considéré.
2. L'ensemble des effectifs soumis au test, sont estimés avec des précisions très voisines tant pour la totalité de la base de sondage qu'au niveau de chaque strate. Par contre, en ce qui concerne les ratios, la précision varie beaucoup plus d'un caractère à l'autre. Ce phénomène trouve son explication dans

(1) : Voir la définition retenue par l'INSEE au chapitre 1.

Tableau 2.5 (1) : Résultats du test mené avec le programme VARECH

Coefficient de variation, sur l'ensemble de la base et pour les neuf strates morphologiques, de 15 caractères démographiques. (estimateurs sans biais sous plan à probabilités inégales pour les effectifs, estimateurs de ratios sous plan équiprobable pour les quotients, allocation au strates optimale).

Caracteres :	Population totale	Population de nationalité française	Pop. des hommes de + de 19 ans ayant un empl.	Population de plus de cinq ans
Strates				
1. Centre ville	10.80	9.73	11.70	10.66
2. Quartiers industriels	58.05	55.78	58.79	57.68
3. Grands ensembles	14.17	13.50	15.16	14.10
4. Petits immeubles	63.85	59.92	67.72	63.60
5. Zonnes pavillonnaires	21.61	20.15	21.52	21.33
6. Péri urbain peu dense	149.05	138.78	152.23	149.22
7. Reliquat morphologique	30.53	28.71	30.55	30.14
8. Dense, grands batiments	31.68	29.37	32.63	31.31
9. Non bâti	561.80	518.17	582.91	550.01
Total (non stratifié)	10.15	9.43	10.87	10.07
Total (stratifié, allocation optimale)	7.67	7.18	8.03	7.59

Tableau 2.5 (2) : Résultats du test mené avec le programme VARECH

Coefficient de variation, sur l'ensemble de la base et pour les neuf strates morphologiques, de 15 caractères démographiques. (estimateurs sans biais sous plan à probabilités inégales pour les effectifs, estimateurs de ratios sous plan équiprobable pour les quotients, allocation au strates optimale).

Caractères :	Population de moins de 60 ans	Nombre de ménages ordinaires	Nombre de menages de plus de une pers.	Nombre de ménages de moins de 6 pers.
Strates				
1. Centre ville	12.26	9.46	10.48	9.33
2. Quartiers industriels	59.14	59.23	57.45	59.93
3. Grands ensembles	14.17	15.46	14.24	15.99
4. Petits immeubles	65.68	69.32	62.50	69.65
5. Zonnes pavillonnaires	23.20	22.50	20.84	22.51
6. Péri urbain peu dense	155.79	156.01	141.40	156.98
7. Reliquat morphologique	32.61	30.53	30.37	30.36
8. Dense, grands batiments	34.74	29.31	30.69	29.02
9. Non bâti	527.03	639.32	576.92	641.42
Total (non stratifié)	10.66	10.30	9.97	10.36
Total (stratifié, allocation optimale)	8.21	7.43	7.53	7.46

Tableau 2.5 (3) : Résultats du test mené avec le programme VARECH

Coefficient de variation, sur l'ensemble de la base et pour les neuf strates morphologiques, de 15 caractères démographiques. (estimateurs sans biais sous plan à probabilités inégales pour les effectifs, estimateurs de ratios sous plan équiprobable pour les quotients, allocation au strates optimale)

Quotients :	Population Français/ Pop. totale	Population ayant un emp./ 5 ans/ population des hommes	Pop. + Pop. de moins de 60 ans / Pop. totale	Pop. de moins de 60 ans / Pop. totale
Strates				
1. Centre ville	3.55	1.39	0.38	2.05
2. Quartiers industriels	10.08	4.15	1.79	9.02
3. Grands ensembles	4.58	1.54	0.65	2.15
4. Petits immeubles	13.97	5.08	2.61	10.88
5. Zonnes pavillonnaires	2.68	1.63	0.69	3.87
6. Péri urbain peu dense	14.82	10.99	4.92	25.30
7. Reliquat morphologique	3.80	2.43	0.94	6.05
8. Dense, grands batiments	5.13	2.13	0.78	6.59
9. Non bâti	664.10	292.83	97.52	379.66
Total (non stratifié)	3.01	0.99	0.47	2.06
Total (stratifié, allocation optimale)	2.12	0.82	0.29	1.34

Tableau 2.5 (4) : Résultats du test mené avec le programme VARECH

Coefficient de variation, sur l'ensemble de la base et pour les neuf strates morphologiques, de 15 caractères démographiques. (estimateurs sans biais sous plan à probabilités inégales pour les effectifs, estimateurs de ratios sous plan équiprobable pour les quotients, allocation au strates optimale)

	Quotients :	Pop. des men. ord./ Nombre des men. ord.	Nombre de men. de + de 1 pers./ Nb. de ménages ordinaires	Nombre de men. de + de 1 pers./ Nombre de ménages ordinaires
Strates				
1. Centre ville		2.86	2.64	0.50
2. Quartiers industriels		15.84	10.16	3.83
3. Grands ensembles		5.05	2.84	1.51
4. Petits immeubles		16.25	12.01	3.59
5. Zonnes pavillonnaires		4.87	4.40	0.96
6. Péri urbain peu dense		31.24	23.66	7.55
7. Reliquat morphologique		8.21	7.94	0.91
8. Dense, grands batiments		6.22	5.94	1.00
9. Non bâti		790.59	627.92	236.42
Total (non stratifié)		3.72	2.28	0.86
Total (stratifié, allocation optimale)		2.28	1.67	0.50

les différences que présentent les distributions des caractères : une différence de variance entre deux variables se répercute évidemment dans la variance de leurs estimations. On observe, en effet, une plus grande variabilité des distributions au sein de la famille des variables de type ratios que pour celle des effectifs.

3. Le tableau 2.6, reprend les résultats du test pour trois caractères démographiques en donnant les quantités à estimer dans chaque strate et les intervalles de confiance des estimations correspondantes. En l'examinant, on prend conscience que les mesures de précisions fournies par les coefficients de variation sont très relatives. De manière générale, les degrés de précision atteints sont encore insuffisants pour fournir des images nettes des variations démographiques à l'intérieur de l'agglomération. De plus, dans le cas des ratios caractérisant la structure de la population, des coefficients de variation de l'ordre de 2 ou 3%, ou même inférieurs à 1%, ne doivent pas faire illusion : il s'agit en effet de caractères dont les variations d'une strate à l'autre se situent précisément dans cet ordre de grandeur (2 ou 3%, parfois moins, autour de la moyenne générale). Les intervalles de confiance associés aux estimations dans chaque strate se chevauchent alors beaucoup trop pour que le sondage à ce taux renseigne de façon fiable sur les variations inter-strates de ces indicateurs.

Tableau 2.6 : Résultats du test mené avec le programme VARECH

Quantités à estimer et intervalles de confiance à 95% par strate pour trois caractères démographiques : population totale, ratio population de plus de 5 ans / population totale, ratio population française / population totale.

Caractère estimé : POPULATION TOTALE	RATIO POP. DE PLUS DE CINQ ANS / POPULATION TOTALE		RATIO POP. FRANCAISE / POPULATION TOTALE	
	Intervalle de confiance Total à 95 % (*)	Valeur de confiance en % à 95 % (**)	Intervalle de confiance Valeur de confiance en % à 95 % (**)	Intervalle de confiance à 95 % (**)
Strates				
1. Centre ville	181508 [142302;220714]	96.03 [95.30;96.76]	86.67 [80.52;92.83]	
2. Quartiers industriels	20830 [0;45014]	94.78 [91.39;98.17]	89.42 [71.40;100]	
3. Grands ensembles	142480 [102101;182859]	93.91 [92.7;95.1]	89.36 [81.2;97.5]	
4. Petits immeubles	8581 [0;19539]	95.41 [90.4;100]	94.70 [68.2;100]	
5. Zonnes pavillonnaires	54298 [30830;77766]	96.49 [95.16;97.80]	96.50 [91.3;100]	
6. Péri urbain peu dense	2287 [0;9105]	95.10 [85.70;100]	96.24 [67.7;100]	
7. Reliquat morphologique	26679 [10389;42969]	96.50 [94.7;98.3]	96.50 [89.2;100]	
8. Dense, grands batim.	18282 [6699;29865]	96.50 [94.9;98.00]	96.22 [86.3;100]	
9. Non bâti	183 [0;2239]	98.36 [0;100]	98.91 [0;100]	
Total non stratifié	455128 [356820;553436]	95.39 [94.5;96.3]	89.97 [84.6;95.4]	
Total stratifié alloc. optim.	455128 [385311;524945]	95.39 [94.84;95.95]	89.97 [86.2;93.8]	

(*) : Plan de sondage à probabilités proportionnelles à la surface des districts, estimation sans biais

(**) : Plan de sondage équiprobable, estimation de ratios.

5.3 Conclusion

Au terme des comparaisons entre variables démographiques que l'on vient d'effectuer, on peut répondre aux deux questions formulées au début de ce chapitre. Rappelons les tout d'abord.

Questions :

1. Existe-t-il des types de variables démographiques qui, du fait des caractéristiques de leur distribution (au sens statistique comme au sens spatial), se prêtent mieux que d'autres aux estimations par sondage ?
2. Les choix des probabilités de tirage, des formules d'estimation et des stratifications, qui définissent le sondage, ont-ils les mêmes conséquences sur la précision des estimations, quelles que soient les variables considérées ?

Réponses :

1. L'estimation des variables de type ratio bénéficie dans l'ensemble d'une précision sensiblement meilleure que celle des effectifs, du moins si l'on s'en tient aux valeurs des coefficients de variation. Ce résultat ne doit pas surprendre puisque, par nature, les premières ont des variances nettement inférieures aux secondes. C'est également ce qui explique que ces coefficients de variation plus faibles ne permettent pas pour autant de saisir de manière précise les variations des ratios au sein de l'agglomération Marseillaise. Il reste donc beaucoup à faire pour atteindre une précision satisfaisante de ce type de sondage.

2. Les conclusions établies lors de l'examen des résultats de l'estimation de l'effectif de population sans doubles comptes, qui concernent le plan de sondage et la formule d'estimation, demeurent valables pour l'ensemble des caractères démographiques considérés par la suite. Il en est de même pour l'ordre de grandeur du gain amené par la stratification ainsi que pour la définition des domaines de "bonne ou de mauvaise précision". La démarche d'amélioration du sondage présentée jusqu'ici semble donc bien robuste lorsqu'on change la variable à estimer. Ce résultat important ne peut que nous encourager à la perfectionner.

A ce propos, et afin de relativiser l'impression négative que peut produire l'examen des intervalles de confiances rapportés dans le tableau 2.6. dont les amplitudes restent élevées, soulignons que la technique de sondage à laquelle ils correspondent est encore très frustrante puisqu'elle ne comprend qu'un seul degré. On perd évidemment beaucoup à enquêter exhaustivement tous les districts de l'échantillon, mais l'évaluation du gain possible grâce à un deuxième degré de sondage ne sera possible qu'à partir des données recueillies à Quito.

Tableau 2.1 : liste des caractères démographiques ayant fait l'objet du test.

EFFECTIFS	QUOTIENTS
1. Population totale	1. Population de nationalité étrangère / Population totale
2. Population des ménages ordinaires	2. Population de nationalité française / Population totale
3. Population de nationalité étrangère	3. Population des hommes de plus de 19 ans au chômage / Population des hommes de plus de 19 ans
4. Population de nationalité française	4. Population des hommes de plus de 19 ans ayant un emploi / Population des hommes de plus de 19 ans
5. Population des hommes de plus de 19 ans au chômage	5. Population âgée de moins de cinq ans / Population totale
6. Population des hommes de plus de 19 ans ayant un emploi	6. Population âgée de cinq ans ou plus / Population totale
7. Population âgée de moins de cinq ans	7. Population âgée de soixante ans ou plus / Population totale
8. Population âgée de cinq ans ou plus	8. Population âgée de moins de soixante ans / Population totale
9. Population âgée de soixante ans ou plus	9. Population des ménages ordinaires / Nombre de ménages ordinaires
10. Population âgée de moins de soixante ans	10. Nombre de ménage (ordinaires) d'une personne / Nombre de ménages ordinaires
11. Nombre de ménages ordinaires	11. Nombre de ménages (ordinaires) de plus d'une personne / Nombre de ménages ordinaires
12. Nombre de ménages (ordinaires) d'une personne	12. Nombre de ménages (ordinaires) de six personnes ou plus / Nombre de ménages ordinaires
13. Nombre de ménages (ordinaires) de plus d'une personne	13. Nombre de ménages (ordinaires) de moins de six personnes / Nombre de ménages ordinaires
14. Nombre de ménages (ordinaires) de six personnes et plus	
15. Nombre de ménages (ordinaires) de moins de six personnes	

CHAPITRE 6

STRATIFICATIONS MORPHOLOGIQUES OBTENUES A PARTIR DE

L'INFORMATION SATELLITAIRE :

RESULTATS

Introduction

Nous venons de mesurer l'intérêt que peut avoir, pour améliorer la précision du sondage, la stratification morphologique de la base. Mais il ne faut pas perdre de vue que la stratification testée jusqu'à présent représente une synthèse de l'information morphologique issue de photographies aériennes, contrôlée par des vérifications au sol. Elle ne peut donc pas servir de base à la mesure qui nous intéresse réellement, le gain dû à une stratification construite d'après l'image satellite. C'est précisément cette mesure que nous abordons dans ce chapitre. Suivant en cela le travail fourni par l'équipe de télédétection, nous aurons désormais une démarche analytique : tester une à une les stratifications basées sur les différents descripteurs de la morphologie urbaine jugés pertinents pour une observation démographique. Le premier, et jusqu'à maintenant le seul, des descripteurs morphologiques étudié est la mesure de la densité du bâti ; c'est donc aux stratifications basées sur la seule densité du bâti que nous allons nous intéresser maintenant.

Nous exposons rapidement la démarche qui a conduit aux stratifications étudiées, qui, comme on va le voir, ne sont pas encore directement issues de l'image satellite. Puis nous commentons les résultats obtenus pour ce descripteur, et évoquons les étapes suivantes du travail qui, en étendant le champ de l'étude à d'autres descripteurs morphologiques, doivent conduire à une stratification morphologique efficace, issue de l'image satellite.

6.1. Provenance et élaboration des nouvelles stratifications.

Dans un rapport remis en avril 87 (1), Alain MICHEL résume ainsi le travail qu'il a effectué :

" Le descripteur densité du bâti, (...) caractérise l'ensemble de la zone à l'aide des six modalités suivantes : NON BÂTI, TRES FAIBLE, FAIBLE, MOYENNE, FORTE, TRES FORTE. Le but de ce travail est de générer, à partir de l'image Landsat T.M. six canaux, une image qui rende compte de la densité du bâti observée sur mosaïque de photographies aériennes et vérifiée sur le terrain. De façon plus pragmatique, il s'agit donc de classer les images Landsat T.M. afin de se rapprocher le plus possible de l'interprétation visuelle de la densité du bâti réalisée sur la mosaïque."

La démarche adoptée est donc la suivante :

1. Identification d'un descripteur morphologique et d'une stratification de référence ; ici, il s'agit de la stratification en six classes définie par le descripteur photo-interprété : "densité du bâti".

2. Recherche d'une classification de l'image satellite la plus proche possible de la stratification de référence.

De notre côté, nous allons suivre la même démarche pour évaluer le gain de précision attaché à chaque stratification :

1. Mesure du gain amené par la stratification de référence (six classes de densité de bâti définies lors de la photo-interprétation).

2. Mesure du gain amené par la stratification correspondante de l'image satellite.

Pratiquement, se pose alors le problème de la réaffectation à la base de sondage (ici, les 2619 districts INSEE), de stratifications définies sur d'autres unités spatiales : les zones photo-interprétées pour la première, les pixels de l'image satellite pour la seconde. La première réaffectation est facile puisque la base de sondage actuelle est constituée de districts entièrement contenus dans une même zone morphologique. Elle fournit donc une stratification de la base de sondage en six classes :

1 : Alain MICHEL, Télédétection et observation suivie de la morphologie et de la démographie des villes des pays en développements. Mise au point des méthodes d'analyse des images satellite à haute résolution et évaluation des informations fournies par ces images. Etude statistique du descripteur "densité du bâti, présentation des classifications.

ORSTOM, dept D, U.R. 406 et EHESS. Avril 1987.

- 1 : NON BATI
- 2 : TRES FAIBLE
- 3 : FAIBLE
- 4 : MOYENNE
- 5 : FORTE
- 6 : TRES FORTE.

C'est le système TIGRE (2) qui permet de mener à bien la seconde réaffectation en mettant en correspondance l'image issue de la classification des pixels avec celle de la base de sondage. Pour des raisons d'encombrement du matériel informatique de l'ORSTOM, cette opération n'a pu être réalisée dans des délais compatibles avec la rédaction de ce rapport. Nous ne disposons donc pas des résultats concernant cette stratification, mais il ressort du travail d'Alain MICHEL (voir rapport d'avril 87, P. 72), qu'elle peut être approximée par la stratification que l'on obtient en regroupant deux à deux les modalités du descripteur photo-interprété pour obtenir trois classes :

- 1 : NON BATI et TRES FAIBLE
- 2 : FAIBLE et MOYENNE
- 3 : FORTE et TRES FORTE.

6.2. Résultats, commentaires et perspectives

Pour comparer ces stratifications, nous avons retenu 11 caractères démographiques, choisis dans l'ensemble de ceux qui ont servi aux tests des estimateurs et des plans de sondage : six effectifs et cinq ratios dont la liste est donnée dans le tableau 1., en fin de chapitre. Conformément aux conclusions de l'étape précédente (3), le présent test a été mené avec l'estimateur sans biais sous plan à probabilités proportionnelles à la surface des districts, du moins dans la mesure du possible, c'est à dire quand il s'agissait d'estimer les effectifs ; pour les quotients, il s'agit de l'estimateur par le ratio sous plan équiprobable.

Le tableau 2 (fin de chapitre) résume les résultats : pour chacun des caractères étudiés, il donne le coefficient de variation de l'estimation non stratifiée (col.1), puis celui obtenu avec trois stratifications distinctes :

- la stratification morphologique synthétique en neuf classes utilisée jusqu'à présent (col.2)

2 : Le système d'information géographique TIGRE est un logiciel de gestion de bases de données localisées, organisé suivant le schéma relationnel, mis au point par l'unité d'infographie de l'ORSTOM.

3 : voir chapitre 5 : Estimations stratifiées de caractères démographiques (5.1. Comparaison de plusieurs estimations d'une même variable).

- la stratification en six classes selon la densité du bâti qu'a permis d'obtenir la photo-interprétation (col.3)
 - la stratification en trois classes de densité de bâti, obtenue par regroupement, et considérée ici comme une approximation de celle qu'on pourra tirer de l'image satellite (col.4). Dans le même tableau, on trouve ensuite le gain de précision apporté par chacune des stratifications par rapport à l'estimation non stratifiée (col. 5, 6 et 7). Ce gain est mesuré par la différence entre les écarts-type avant et après stratification, rapportée à l'écart-type de l'estimation non stratifiée et exprimé en %. Il s'agit donc d'une mesure du gain obtenu sur l'amplitude des intervalles de confiance à 95%, et non sur les variances des estimations, comme nous le pratiquons jusqu'à présent. Enfin, la figure 3 (fin de chapitre) donne une représentation graphique de ces gains.

Les conclusions qui suivent sont relatives à deux comparaisons distinctes effectuées sur l'ensemble des caractères :

- La stratification en trois classes de densité du bâti est confrontée à sa référence, la classification en six classes issue de la photo-interprétation, afin d'évaluer dans quelle mesure une classification de l'image satellite sur ce thème permet d'approcher une certaine réalité terrain (qu'apporte la photo-interprétation), dont on sait que la prise en compte améliore le sondage.

- Les stratifications basées sur la seule densité du bâti sont comparées à la stratification morphologique synthétique afin de mesurer l'apport spécifique de ce descripteur à une future stratification synthétique, construite cette fois à partir de l'information satellitaire.

6.2.1. Comparaison des stratifications selon la densité du bâti en trois et six classes.

Le tableau suivant donne, pour chaque caractères, le gain de précision dû à la stratification en trois classes (G1) en pourcentage du gain dû à celle en six classes (G2).

Caractères : -1- -2- -3- -4- -5- -6- -7- -8-

 G1/G2 en % : 88.4 93.0 82.6 88.3 93.0 88.7 32.2 16.7

Caractères : -9- -10- -11-

 G1/G2 en % : 100 93.8 100

DÉPARTEMENT
SOCIÉTÉ

Hormis pour les deux ratios relatifs à la nationalité (population française/population totale) et à l'emploi (population des hommes de plus de 19 ans ayant un emploi/population des hommes de plus de 19 ans), le gain apporté par la stratification en trois classes est très proche de celui induit par celle qui a servi de référence au travail sur l'image satellite. Sous réserve que la stratification obtenue par classification des pixels de l'image T.M soit effectivement comparable au regroupement en trois classes opéré, on peut d'ores et déjà conclure à la pertinence de la télédétection comme moyen de stratification d'une base de sondage urbaine sur un critère de densité du bâti. Nous allons constater maintenant qu'une telle stratification améliore déjà de façon notable la précision de l'ensemble des estimations.

6.2.2. Contribution du caractère morphologique "densité du bâti" à une stratification morphologique synthétique.

A nouveau, le tableau ci-dessous donne le gain dû au descripteur "densité du bâti" (G1) par rapport à celui de la stratification synthétique (G2).

Caractères :	-1-	-2-	-3-	-4-	-5-	-6-	-7-	-8-

G1/G2 en % :	62.6	57.8	82.7	61.8	78.6	64.1	31.5	70.6

Caractères :	-9-	-10-	-11-					

G1/G2 en % :	51.4	44.4	38.9					

A l'examen du tableau, on constate que la contribution du descripteur "densité du bâti", au gain qu'apporte une stratification morphologique synthétique varie entre 30 et 80% du gain total et se situe, pour la plupart des variables démographiques étudiées, au dessus de 50% (les seules exceptions étant les ratios 7, 10 et 11 concernant la nationalité, la taille des ménages et la structure par âge). Rappelons ici que la densité du bâti fut l'un des descripteurs morphologiques dont on a augmenté l'importance par pondération lors de l'analyse typologique. Il ne faut donc pas s'étonner qu'il soit responsable d'une grande part du gain de précision

que permet la stratification synthétique. Il se pourrait d'ailleurs que la somme des contributions de chacun des descripteurs morphologiques pris isolément, soit supérieure à 100 %. Peut être même obtiendrait-on, par superposition des classifications fondées sur chacun d'eux, une stratification meilleure (au sens du sondage à but démographique), que celle qui nous sert pour l'instant de référence.

C'est pourquoi il nous paraît intéressant de poursuivre le travail analytique qui a été commencé ici, dans la direction suivante : il faudra soumettre au même type de test, chacun des autres descripteurs morphologiques issus de la photo-interprétation, construire avec eux des stratifications et sélectionner les plus efficaces d'entre elles en mesurant les gains de précision qu'elles permettent d'obtenir. Ce travail effectué, l'équipe de télédétection produira une série d'images, relatives à chacun des descripteurs sélectionnés, en les identifiant à l'aide de l'information spectrale. Les résultats qu'ils obtiendront alors, sous forme de classifications thématiques de l'image satellite, seront ensuite réaffectés à la base de sondage et testés comme de nouvelles stratifications. A l'issue de la démarche, nous disposerons d'un ensemble de stratifications de la même base de sondage, chacune représentant un thème morphologique, qu'il suffira alors de superposer pour obtenir la stratification finale.

Figure 1. : liste des caractères démographiques considérés dans le chapitre

A) effectifs

N°	Nom du caractère	Abréviation utilisée
-1-	Population sans doubles comptes	POPSDC
-2-	Population totale de nationalité française	POPFRANC
-3-	Population totale des hommes de plus de 19 ans ayant un emploi	POPHEMP
-4-	Population totale des ménages ordinaires	POPMEN
-5-	Nombre de ménages ordinaires de moins de 6 personnes	NBMEN-6
-6-	Population totale de plus de cinq ans	POP+5

B) ratios

N°	Nom du caractère	Abréviation utilisée
-7-	Population totale de nationalité française ----- Population totale	POPFRANC ----- POPTOTAL
-8-	Population totale des hommes de plus de 19 ans ayant un emploi ----- Population totale des hommes de plus de 19 ans	POPHEMP ----- POPH+19
-9-	Population totale des ménages ordinaires ----- Nombre de ménages ordinaires	POPMEN ----- NBMENAG
-10-	Nombre de ménages ordinaires de moins de 6 personnes ----- Nombre de ménages ordinaires	NBMEN-6 ----- NBMENAG
-11-	Population totale de plus de cinq ans ----- Population totale	POP+5 ----- POPTOTAL

Tableau 2. : coefficients de variation des estimations non stratifiées et stratifiées selon trois critères morphologiques et gains de précision correspondant pour 11 caractères démographiques.

A) Effectifs

Caractères estimés	Coef. de variation des estimations				Gain de précision (en % du coef. de variation sur l'estimation non stratifiée)		
	(0)	(1)	(2)	(3)	(1)	(2)	(3)
1 POPSDC	10.13	7.67	8.59	8.77	24.28	15.20	13.43
2 POPFRANC	9.43	7.18	8.13	8.22	23.86	13.79	12.83
3 POPHEMP	10.87	8.03	8.52	8.93	26.13	21.62	17.85
4 POPMEN	10.25	7.76	8.71	8.89	24.29	15.02	13.27
5 NBMEN-6	10.36	7.46	8.08	8.24	27.99	22.01	20.46
6 POP+5	10.07	7.59	8.48	8.66	24.63	15.79	14.00

(0) : Sans stratification

(1) : Stratification morphologique synthétique en 9 classes

(2) : Stratification sur la densité du bâti en 6 classes

(3) : Stratification sur la densité du bâti en 3 classes

Tableau 2. (suite) : coefficients de variation des estimations non stratifiées et stratifiées selon trois critères morphologiques et gains de précision correspondant pour 11 caractères démographiques.

B) Ratios

Caractères estimés	Coef. de variation des estimations				Gain de précision (en % du coef. de variation sur l'estimation non stratifiée)		
	(0)	(1)	(2)	(3)	(1)	(2)	(3)
7 POPFRANC ----- POPTOTAL	3.01	2.12	2.73	2.92	29.57	9.30	2.99
8 POPHEMP ----- POPH+19	0.99	0.82	0.87	0.97	17.17	12.12	2.02
9 POPMEN ----- NBMENAG	3.72	2.28	2.98	2.98	38.71	19.89	19.89
10 NBMEN-6 ----- NBMENAG	0.86	0.50	0.70	0.71	41.86	18.60	17.44
11 POP+5 ----- POPTOTAL	0.47	0.29	0.40	0.40	38.30	14.89	14.89

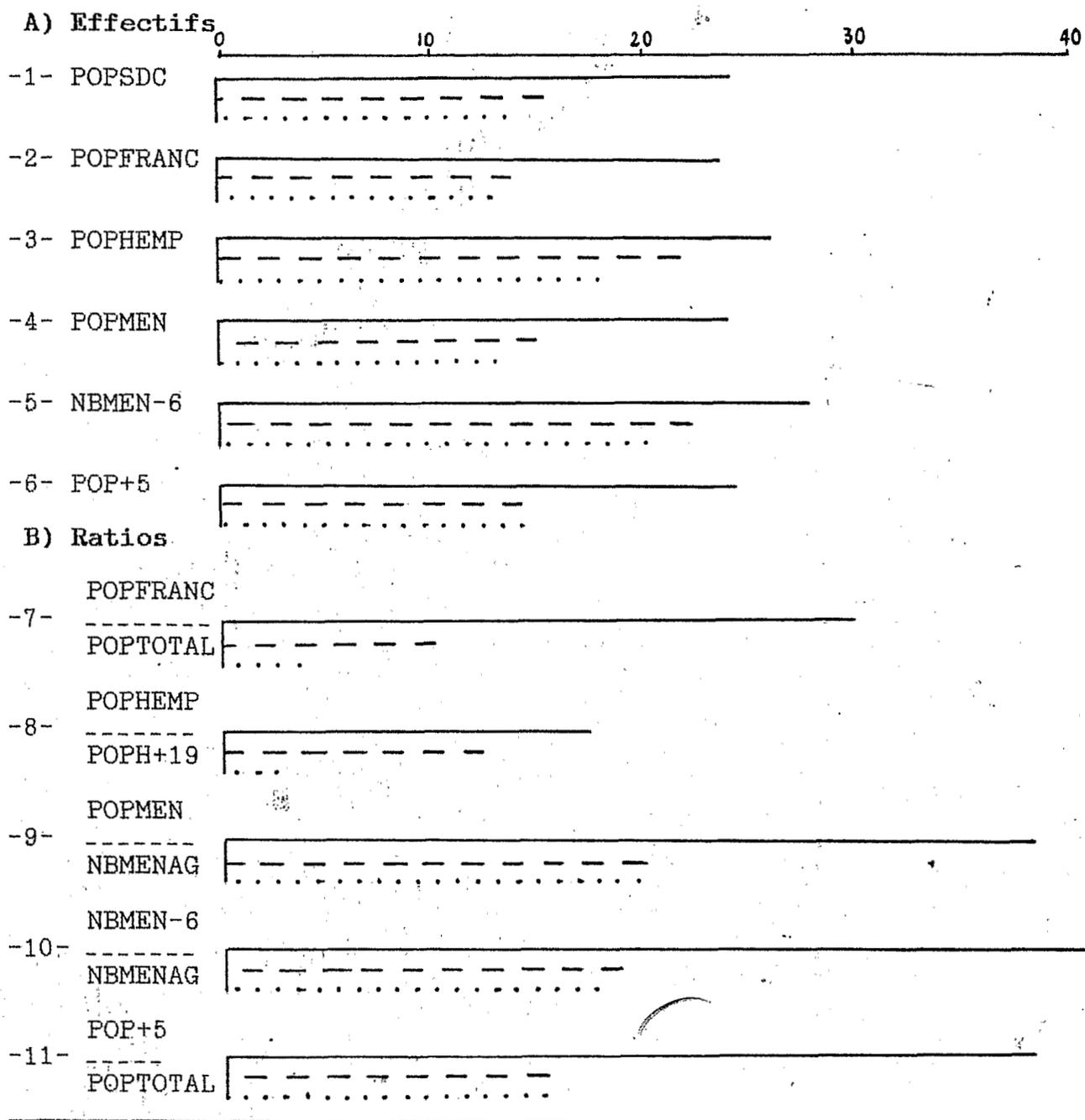
(0) : Sans stratification

(1) : Stratification morphologique synthétique en 9 classes

(2) : Stratification sur la densité du bâti en 6 classes

(3) : Stratification sur la densité du bâti en 3 classes

Graphique 3. : Comparaison des gains de précision apportés par trois stratifications morphologiques à l'estimation de 11 caractères démographiques (en % de la variance non stratifiée)



Stratification synthétique morphologique : (1) —————
(9 classes)

Stratification sur la densité du bâti : (2) - - - - -
(6 classes)

Stratification sur la densité du bâti : (3)
(3 classes)

CHAPITRE 7

TENTATIVE D'EVALUATION DE L'EFFICACITE DES TIRAGES AREOLAIRES
ASSURANT UNE BONNE REPARTITION GEOGRAPHIQUE DES UNITES SONDEES

Introduction

Dans la méthode que nous tentons de mettre au point, la collecte se fera par sondage aréolaire systématique à partir de documents satellitaires ; il s'agit d'une contrainte pratique dont on a déjà abordé les conséquences sur les probabilités de sélection des districts ou, plus généralement, des unités d'échantillonnage, et donc, sur la définition du plan de sondage. Mais ce type de sondage possède une autre caractéristique importante : il fournit toujours un échantillon géographiquement bien réparti au sein de la base de sondage. Or, il n'est un secret pour personne et surtout pas pour les citoyens eux-mêmes, que dans la plupart des villes les caractéristiques démographiques des quartiers ne sont pas indépendantes de la localisation. Marseille et Quito, nos deux sites d'expérience, n'échappent pas à cette règle, même si l'organisation concentrique de la première s'oppose, par certains aspects, à la structure linéaire de la seconde. C'est sur cette évidence que repose l'idée directrice du travail dont l'exposé va suivre : les caractères démographiques que nous cherchons à estimer n'étant pas indépendants de la localisation, comment évaluer le gain de précision qu'amène la bonne répartition spatiale de l'échantillon obtenue quand la technique de sondage prend en compte la localisation des unités statistiques ?

D'autre part, il n'y a pas non plus indépendance entre les spécificités morphologiques et la localisation : à un type morphologique donné correspond une localisation plus ou moins précise. Donc, si l'on considère, comme c'est le cas ici, morphologie et localisation comme deux variables "explicatives" de la démographie, il y a très probablement une certaine redondance entre les deux et les gains permis par les stratifications construites à partir de chacun des critères ne s'additionnent pas totalement lorsqu'on les prend en compte simultanément.

Pourtant cette redondance n'est sûrement pas totale dans la mesure où, au sein d'une même strate morphologique, les différences de localisation peuvent à nouveau "expliquer" une part de la variance intra-strate des caractères démographiques.

Dans le cadre de notre sondage stratifié, peut-on chiffrer l'apport, non redondant avec la morphologie, de la localisation telle que la prend en compte un tirage spatial systématique ?

Pour répondre à ces questions, la première partie de ce chapitre propose une méthode d'approximation de la précision du tirage systématique aréolaire par assimilation à un tirage stratifié sur un critère de localisation. Puis, on dresse la liste des données nécessaires au test et on expose comment sont obtenues les stratifications employées. Enfin, après avoir présenté les résultats, on tire les conclusions de cette expérience.

7.1. La méthode : approximation du tirage systématique aréolaire par une stratification de la base de sondage sur un critère de localisation.

La méthode exposée ici s'inspire très largement d'une proposition faite par DESABIE dans "Théorie et pratique des sondages" à propos du tirage systématique sur liste (1), dont nous reprenons l'idée centrale pour l'adapter au tirage aréolaire utilisant une grille de points. DESABIE assimile le tirage systématique sur liste au tirage suivant (M est le cardinal de la base de sondage, m celui de l'échantillon).

" L'univers est découpé en $m/2$ strates égales, chacune d'entre elles contenant $2(M/m)$ unités consécutives de la population. Dans chaque strate on tire au sort deux unités" et l'estimateur de la moyenne du caractère Y est :

$$Y = (2/m) \cdot \sum_{h=1}^{m/2} Y_h, \text{ où}$$

$$Y_h = 1/2 (Y_{2h-1} + Y_{2h}).$$

L'idée est en fait d'utiliser les formules d'estimations et de variances liées au sondage stratifié, en stratifiant la base de sondage à l'aide de l'ordre de la liste, et cela en créant le nombre maximal de strates. Ce nombre n'est limité que par la contrainte qu'on se donne de pouvoir calculer un estimateur de la variance de l'estimation dans chaque strate ; l'échantillon doit donc contenir au moins deux individus de chaque strate.

L'adaptation au tirage d'unités aréolaires à l'aide d'une grille de points se fait en considérant comme strates les mailles, carrées et de surface égale, d'une seconde grille placée sur le document satellitaire : chaque strate est définie par la longueur de son côté et les coordonnées géographiques d'un de ses sommets.

1 : J. DESABIE, "Théorie et pratique des sondages" (DUNOD 1966) chapitre 7,7.7 : Le tirage systématique envisagé comme un sondage stratifié, P. 171.

Ici apparaît l'unique problème posé par l'adaptation de la méthode au sondage aréolaire : si la grille de tirage a une maille constante sur l'ensemble de la base, chaque strate définie par la grille de stratification contient le même nombre de districts échantillonnés, mais le taux de sondage varie alors suivant les strates puisqu'il est fonction du rapport entre la surface de la maille qui sert au tirage et la surface moyenne des unités de la strate. L'approximation, par cette méthode, du tirage aréolaire à l'aide d'une grille uniforme nécessite donc le calcul de taux spécifiques à chaque strate. Le programme actuel de calcul des variances d'estimation ne le permettant pas, nous avons préféré procéder, dans un premier temps, à l'approximation de la variance qu'aurait un sondage aréolaire à taux constant sur l'ensemble de la base. Ce sondage correspond à la procédure de tirage suivante : la taille de la grille, redéfinie pour chaque sous-ensemble de la base au sein duquel les districts ont une taille voisine, est proportionnelle à la surface moyenne des unités du sous-ensemble, ce qui permet d'obtenir un taux à peu près constant. Une telle procédure nécessite en revanche un premier zonage de la base de sondage, effectué d'après la taille des unités. Les résultats qui vont suivre sont donc une approximation de la précision attachée à une telle procédure.

Enfin, la méthode appelle une dernière remarque : pour optimiser l'approximation du sondage systématique, si l'on suit l'idée de DESABIE, le nombre de strates et donc leur surface, devrait en principe être défini de manière à ce que chaque strate contienne deux points de la grille de tirage. En pratique, aux taux de sondages que nous envisageons, ce nombre de strates (environ 90) ralentit énormément l'exécution du programme de calcul des variances d'estimation. Nous avons donc retenu des maillages de stratification plus larges, comportant suivant les cas 6 ou 15 strates, ce qui permet des temps de calculs raisonnables tout en gardant un découpage spatial suffisamment fin.

7.2. Les données nécessaires au test : élaboration des stratifications et remarques sur l'interprétation des résultats

L'information permettant l'élaboration des stratifications spatiales dont il va être question provient du système de gestion de base de données localisées TIGRE. Les fonds cartographiques concernant l'ensemble des districts de la base de sondage ayant été intégrés dans ce système, nous disposons, grâce à lui, d'un fichier contenant la latitude et la longitude des centres de ces districts. A partir de ces coordonnées géographiques, un programme affecte chacun des 2619 districts de la base de sondage à une strate définie par un intervalle de longitude et de latitude. Trois stratifications ont été produites qui contiennent respectivement six, huit et quinze classes. Les résultats du test ayant fait apparaître une très faible différence entre

les stratifications en six et huit classes, nous n'avons finalement retenues que les stratification en six et quinze classes schématisées sur la figure 1.

Les stratifications spatiales permettent, lorsqu'on les utilise seules, d'approcher la précision du sondage systématique non stratifié. Pour approximer celle d'un sondage stratifié selon un critère morphologique, nous croisons deux à deux les stratifications morphologique et spatiale pour obtenir les six stratifications du tableau 2.

Figure 1. : schéma du nombre de districts par strate dans les deux stratifications spatiales.

Stratification en 15 classes			stratification en 6 classes	
159	86	43		
			386	275
222	306	108		
			1138	458
592	506	81		
			155	206
100	207	20		
73	97	18		

Tableau 2. : croisement des stratifications spatiales et morphologiques, nombre de classes des stratifications résultantes.

Localisation :	Grille 6 classes	grille 15 classes
Morphologie :		
Stratif. synthétique 9 classes	43 classes	80 classes
Stratif. sur densité du bâti 6 classes	31 classes	62 classes
Stratif. sur densité du bâti 3 classes	18 classes	38 classes

les stratifications en six et huit classes, nous n'avons finalement retenues que les stratification en six et quinze classes schématisées sur la figure 1.

Les stratifications spatiales permettent, lorsqu'on les utilise seules, d'approcher la précision du sondage systématique non stratifié. Pour approximer celle d'un sondage stratifié selon un critère morphologique, nous croisons deux à deux les stratifications morphologique et spatiale pour obtenir les six stratifications du tableau 2.

Figure 1. : schéma du nombre de districts par strate dans les deux stratifications spatiales.

Stratification en 15 classes			stratification en 6 classes	
159	86	43		
			386	275
222	306	108		
			1138	458
592	506	81		
			155	206
100	207	20		
73	97	18		

Tableau 2. : croisement des stratifications spatiales et morphologiques, nombre de classes des stratifications résultantes.

Localisation :	Grille 6 classes	grille 15 classes
Morphologie :		
Stratif. synthétique 9 classes	43 classes	80 classes
Stratif. sur densité du bâti 6 classes	31 classes	62 classes
Stratif. sur densité du bâti 3 classes	18 classes	38 classes

Ce principe des stratifications croisées appelle deux remarques :

1. Dans certaines strates contenant un très petit nombre d'unités, l'échantillon ne contient qu'une fraction de district. De ce fait, un tel tirage est impossible à mettre en oeuvre pratiquement. Ce n'est d'ailleurs pas l'objectif de la méthode qui se propose simplement de fournir une approximation de la précision du tirage systématique qui est, lui, praticable.

2. Au sein du programme de calcul des variances d'estimation, la détermination de l'allocation optimale se fait en minimisant l'expression de la variance stratifiée : il s'agit d'une somme sur l'ensemble des strates issues du croisement des stratifications spatiales et morphologiques. En toute rigueur, ce calcul n'est pas légitime puisqu'il conduit à faire varier le taux de sondage selon la localisation alors que la seule "vraie" stratification est la stratification morphologique. On aurait donc du déterminer l'allocation optimale uniquement par rapport à elle. Là encore, l'état actuel du programme de calcul nous en empêchait et les résultats relatifs à l'allocation optimale ne sont que des indications sur l'ordre de grandeur du gain amené par la procédure de tirage sur grille.

7.3. Résultats et conclusions

Pour ce test, nous n'avons retenu que trois caractères démographiques qui sont le total de population sans doubles compte et deux ratios par rapport à la population totale : population de nationalité française et population de 5 ans et plus. Ce choix est dirigé vers les caractères dont l'estimation profite le plus des stratifications morphologiques.

Le tableau 3 présente les gains de précision obtenus grâce aux deux stratifications spatiales et les compare aux gains dus à la stratification morphologique en 9 classes. Sont donnés successivement pour chacun des caractères, le coefficient de variation de l'estimation, puis le gain réalisé sur l'amplitude de l'intervalle de confiance en pour cent de l'intervalle non stratifié (allocation proportionnelle dans tous les cas).

Ce qui frappe d'abord à l'examen du tableau, c'est la variabilité des gains permis par les deux stratifications spatiales, suivant le caractère démographique que l'on estime. On obtient une réduction de 60 % de l'intervalle de confiance dans le cas du ratio population de plus de cinq ans / population totale, alors que le gain n'est que de 2 ou 4 %, suivant le degré de finesse de la stratification, quand on s'intéresse à la proportion de population française.

D'autre part, l'amélioration obtenue en passant de 6 à 15 strates est elle aussi variable, mais cette fois, ce sont l'estimation du total de population sans doubles comptes et celle de la proportion de français qui en profitent de façon significative.

Le tableau montre également que, dans certains cas, le gain que permet la bonne répartition spatiale de l'échantillon peut être égal et même supérieur à celui dû à la stratification morphologique.

Deux questions se posent maintenant :

1. Du point de vue de l'information qu'ils apportent, les deux types de stratifications, morphologique et spatiale, sont en partie redondants. Il est donc probable, comme on l'a déjà dit, que les gains liés à chacun d'eux ne s'additionnent pas entièrement lorsqu'on construit les stratifications croisées. On peut, en plus, se demander si l'amélioration permise par le tirage systématique, représenté ici par la stratification spatiale, est la même quelque soit la stratification morphologique employée.

2. La variabilité, selon les caractères estimés, des gains permis par la stratification spatiale peut avoir, à notre sens, deux explications :

- premièrement, certains caractères démographiques sont plus liés que d'autres à la localisation des districts.

- deuxièmement, le niveau de finesse de la partition spatiale retenue comme stratification peut suffire à rendre compte des variations dans l'espace urbain de certains caractères, alors que pour d'autres, ces variations ont lieu pour une part importante à l'intérieur de la maille employée. Dans ce cas, on devrait en stratifiant plus, augmenter le gain de précision de manière sensible.

On le voit, le gain dû au sondage systématique semble devoir varier suivant le caractère estimé et également suivant le degré de finesse de la stratification morphologique de départ. En examinant les résultats obtenus lorsqu'on croise les deux types de stratification, nous tentons maintenant de préciser ce qui vient d'être dit. Le tableau 4. donne les coefficients de variation correspondant aux six croisements de stratifications possibles (3 stratifications morphologiques de départ et 2 niveaux de stratification spatiale).

Tout d'abord, on constate que le gain amené par la stratification spatiale est effectivement du même ordre quelque soit la stratification morphologique. Par contre, il demeure très variables selon le caractère démographique que l'on cherche à estimer. L'effectif de population sans doubles comptes ne profite que très peu de la stratification spatiale

alors que l'estimation des deux ratios est améliorée sensiblement (30 à 50 %).

Il est particulièrement intéressant de noter que le ratio "population de nationalité française / population totale", dont l'estimation ne bénéficiait que très peu de la stratification purement spatiale, profite maintenant beaucoup mieux des stratifications plus fines. Pour cette variable, on constate donc un premier gain lors de l'introduction d'une stratification morphologique, puis un second gain du même ordre que le premier lorsqu'on stratifie spatialement (environ 40 % du coefficient de variation dans les deux cas). Il semble donc que certains caractères démographiques présentent plusieurs niveaux de variation spatiale appréhendables, au niveau de l'amélioration d'un sondage, soit par la stratification morphologique, soit par le tirage systématique aréolaire. Le découpage de l'espace urbain en sous-ensembles morphologiquement homogènes, rend compte d'une partie de la variation, tandis qu'à l'intérieur de ces strates, le tirage systématique à l'aide d'une grille de points permet d'appréhender les variations plus locales.

La dernière constatation à faire sur le tableau 4. est que le gain obtenu grâce aux stratifications croisées est équivalent à la division spatiale soit faite en six ou en quinze classes. On peut donc penser que l'on parvient, grâce à la méthode retenue, à une approximation correcte du gain qu'amène le tirage systématique.

Pour conclure, résumons les leçons que l'on peut tirer du test.

En premier lieu, il confirme qu'un tirage aréolaire systématique, en assurant une bonne répartition géographique de l'échantillon, permet une amélioration sensible des estimations.

Deuxièmement, il apparaît que le gain à attendre d'un tel tirage est variable suivant le type de caractère que l'on cherche à estimer : il se situe, dans nos exemples, entre 10 et 50 % du coefficient de variation correspondant au tirage strictement aléatoire. Dans les cas stratifiés, il sera évidemment d'autant plus important que le caractère variera localement à l'intérieur des strates morphologiques.

Par contre, pour un caractère démographique donné, le gain dû au tirage systématique a été le même pour les trois stratifications morphologiques testées. On peut, nous semble-t-il, conclure à une certaine robustesse de cette technique quand on change de critère de stratification, du moins en gardant un niveau de finesse des partitions comparable.

Enfin, la méthode d'assimilation à un tirage stratifié d'après critère spatial semble efficace pour évaluer le gain apporté par le tirage systématique que nous pratiquerons. Pour une ville de la taille de Marseille ou de Quito, il semble qu'un découpage de l'agglomération en une dizaine de zones géographiques soit suffisant pour obtenir une approximation satisfaisante de ce gain.

Tableau 3. : coefficients de variation et gain de précision obtenus avec les stratifications morphologiques et spatiales pour trois caractères démographiques (allocation proportionnelle).

Caractères	POPSDC		POP+5 ----- POPTOTAL		POFFRANC ----- POPTOTAL	
	(*)		(**)		(**)	
	(1)	(2)	(1)	(2)	(1)	(2)
Stratifications						
Non stratifié	10.13	0.00	0.78	0.00	3.03	0.00
Stratif. morpho. synthétique (9 classes)	9.08	10.40	0.31	60.30	2.57	15.20
Stratif. sur la localisation (6 classes)	9.70	4.20	0.33	57.70	2.96	2.30
Stratif. sur la localisation (15 classes)	9.25	8.70	0.30	61.50	2.91	4.00

(*) : Estimateur sans biais sous plan à probabilité proportionnelle à la surface des districts.

(**) : Estimateur du ratio sous plan équiprobable.

(1) : coefficient de variation (%)

(2) : gain sur le coefficient de variation en pour cent du coefficient de l'estimateur non stratifié.

Tableau 4. : coefficients de variation et gains de précision obtenus sur l'estimation de trois caractères démographiques à partir de stratifications croisant un critère morphologique et un critère de localisation.

(six croisements, allocations proportionnelle et optimale)

Carac- tères	POPSDC			POP+5 ----- POPTOTAL			POPFRANC ----- POPTOTAL		
	(*)			(**)			(**)		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Sans strat.	10.13	0.00	0.00	0.78	0.00	0.00	3.03	0.00	0.00

1) Croisements avec la stratification morphologique syntétique en neuf classes.

Strat. morph. (9 cl.)	9.08	7.67	24.30	0.31	0.24	69.20	2.57	1.89	37.60
morph. (9 cl.) x loc. (6 cl.)	8.96	7.16	29.30	0.25	0.19	75.60	2.00	1.55	48.80
morph. (9 cl.) x loc. (15 cl.)	8.93	6.83	32.60	0.25	0.19	75.60	2.03	1.50	50.50

(*) : Estimateur sans biais sous plan à probabilité proportionnelle à la surface des districts.

(**) : Estimateur du ratio sous plan équiprobable.

(1) : coefficient de variation (%)

(2) & (3) : gain sur le coefficient de variation en pour cent du coefficient de l'estimateur non stratifié.

Tableau 4. (suite) : coefficients de variation et gains de précision obtenus sur l'estimation de trois caractères démographiques à partir de stratifications croisant un critère morphologique et un critère de localisation.

(six croisements, allocations proportionnelle et optimale)

Carac- tères	POPSDC			POP+5 ----- POPTOTAL			POPFRANC ----- POPTOTAL		
	(*)			(**)			(**)		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Sans strat.	10.13	0.00	0.00	0.78	0.00	0.00	3.03	0.00	0.00

2) croisement avec la stratification en six classes de densité du bâti.

Strat. dens. (6 cl.)	8.81	8.59	15.20	0.38	0.36	53.80	2.73	2.24	26.10
Dens. (6 cl.) x loc. (6 cl.)	8.57	7.98	21.20	0.26	0.23	70.50	2.00	1.56	48.50
Dens. (6 cl.) x loc. (15 cl.)	8.37	7.48	26.20	0.25	0.21	73.10	2.00	1.58	47.90

(*) : Estimateur sans biais sous plan à probabilité proportionnelle à la surface des districts.

(**) : Estimateur du ratio sous plan équiprobable.

(1) : coefficient de variation (%)

(2) & (3) : gain sur le coefficient de variation en pour cent du coefficient de l'estimateur non stratifié.

Tableau 4. (fin) : coefficients de variation et gains de précision obtenus sur l'estimation de trois caractères démographiques à partir de stratifications croisant un critère morphologique et un critère de localisation.

(six croisements, allocations proportionnelle et optimale)

Carac- tères	POPSDC			POP+5			POPFRANC		
	(*)			(**)			(**)		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Sans strat.	10.13	0.00	0.00	0.78	0.00	0.00	3.03	0.00	0.00

3) croisement avec la stratification en trois classes de densité du bâti.

Strat. dens. (3 cl.)	8.90	8.77	13.43	0.41	0.39	50.00	2.92	2.60	14.20
Dens. (3 cl.) x loc. (6 cl.)	8.74	8.29	18.16	0.26	0.24	69.20	2.33	2.00	34.00
Dens. (3 cl.) x loc. (15 cl.)	8.85	7.91	21.90	0.26	0.22	71.80	2.37	1.98	34.70

(*) : Estimateur sans biais sous plan à probabilité proportionnelle à la surface des districts.

(**) : Estimateur du ratio sous plan équiprobable.

(1) : coefficient de variation (%)

(2) & (3) : gain sur le coefficient de variation en pour cent du coefficient de l'estimateur non stratifié.

CONCLUSION

Le corps de ce rapport a été consacré au compte rendu détaillé de la deuxième étape de l'expérience commencée à Marseille il y a maintenant plus de deux ans et dont l'un des objectifs est le test et l'évaluation de méthodes de sondages spatiaux à but démographique en milieu urbain. En conclusion de ce travail, et pour en faire la synthèse, il nous semble utile, après avoir rappelé ses principaux objectifs et résumé les étapes de son déroulement, de tenter de proposer des réponses claires, reposant sur les résultats acquis, aux questions que nous nous posions à son début. Enfin, pour finir, nous essaierons d'identifier les principaux problèmes restants à résoudre et nous donnerons la liste des travaux à réaliser à Quito devant conduire à leurs solutions.

. Bilan de l'expérience menée à Marseille : objectifs, déroulement, résultats

Pour la recherche que nous menons actuellement sur les sondages spatiaux en milieu urbain, l'expérience réalisée à Marseille avait trois objectifs principaux :

1. Vérifier l'hypothèse fondamentale selon laquelle une stratification morphologique de la base de sondage issue d'une image satellite améliore de façon significative la précision de l'estimation des informations démographiques et quantifier ce gain de précision.
2. Sélectionner, après évaluation des différentes alternatives, un plan de sondage efficace et adapté aux contraintes qui surgissent de son application à la collecte de données démographiques en milieu urbain et de l'utilisation de l'image satellite comme base de sondage.
3. Mettre en évidence les principales spécificités du plan de sondage défini et analyser les conséquences qu'elles auront sur la précision des estimations et sur la mise en oeuvre pratique de l'enquête.

Dans le déroulement de l'expérience on peut distinguer quatre étapes dont les trois premières fournissent des résultats principalement méthodologiques, lesquels peuvent d'ailleurs intéresser d'autres domaines que celui de la mise au point de sondages spatiaux, tandis que la dernière, dont l'exposé a fourni la matière de ce rapport, permet de préciser ce que pourrait être un sondage à but démographique en milieu urbain utilisant comme base une image satellite stratifiée.

Première étape : constitution et gestion par le système d'information géographique TIGRE, d'une base de données sur la ville de Marseille. Cette étape a permis d'importantes avancées dans le domaine de l'intégration dans un système de gestion de bases de données d'informations géographiques et satellitaires. Ces résultats, qui intéressent plus spécialement les géographes et les informaticiens sont consignés dans leur quasi totalité dans les deux rapport rendus en Avril 86 et Août 87 pour l'ATP Télédétection spatiale CNRS/CNES.

Deuxième étape: établissement, d'après des informations issues de photographies aériennes, d'une stratification morphologique de l'espace urbain marseillais. Le résultat principal de cette étape est bien entendu la stratification elle-même et son utilisation pour la sélection d'un échantillon de travail indispensable au travail des spécialistes de télédétection et à la mise au point des techniques de sondages spatiaux. Elle laisse également certains acquis dans le domaine de l'application des techniques d'analyse des données aux problèmes d'analyse typologique et de classification automatique des données de morphologie urbaine, ses possibilités et ses limites. On trouvera également l'exposé de cette étape et de ses résultats dans le rapport ATP CNRS/CNES d'avril 86.

Troisième étape : développement du programme VARECH qui permet l'évaluation, d'après des données exhaustives, des performances de différents plans de sondage et de différentes stratifications. Cet outil statistique et informatique est bien sur applicable à d'autres sites géographiques urbains ou non et quelque soit la demande thématique, démographique ou autre.

Quatrième étape : application du programme VARECH aux données du recensement général de la population de 1982 à Marseille pour l'évaluation de plans de sondage et de stratifications.

Nous ne répéterons pas ici les résultats obtenus dans cette quatrième étape puisqu'ils figurent, exposés en détail, dans les conclusions de chacun des chapitres de ce rapport. Il nous semble néanmoins utile d'en faire maintenant une synthèse et de montrer en quoi ils ont permis d'atteindre les objectifs fixés à l'expérience.

La vérification de l'hypothèse fondamentale

Il est acquis, à l'issue de cette expérience, qu'une stratification de l'espace urbain construite par synthèse d'un ensemble de critères morphologiques permet une amélioration significative de la précision de l'estimation par sondage des caractéristiques démographiques des habitants. On peut quantifier cette amélioration en disant qu'elle représente une diminution de 40 %, en moyenne, de la variance des estimateurs de totaux et de 50 % de celle des estimateurs de ratios. D'autre part le critère de densité de l'espace bâti s'avère jouer un rôle déterminant dans ce gain de précision : à lui, seul, il permet une amélioration égale à 50 %, voir 75 % pour certaines estimations, de celle due à la stratification synthétique. Ce résultat ne doit pas surprendre dans la mesure où ce critère a justement été privilégié lors de l'élaboration de la stratification synthétique. Par ailleurs, l'équipe de télédétection a montré que l'on pouvait produire des classifications de l'image satellite qui rendent compte de façon très satisfaisante de la densité du bâti en milieu urbain.

On peut donc penser, à l'issue de ce travail, pouvoir disposer rapidement de stratifications des images satellite qui permettent de réduire au moins de moitié la variance des estimations par sondage de caractères démographiques et cela quelque soit le plan de sondage considéré.

Sélection d'un plan de sondage efficace et adapté

Comme il a déjà été dit, l'utilisation de l'image satellite comme base de sondage spatiale implique quasi fatalement un plan de sondage à probabilités inégales. La technique de sélection des unités spatiales à enquêter la plus simple à mettre en oeuvre est celle du tirage systématique à l'aide d'une grille de points plaquée sur l'image. La probabilité d'inclusion est alors proportionnelle à la surface de chaque unité spatiale.

Les tests réalisés à partir des données de Marseille permettent d'affirmer qu'avec ce plan de sondage, les estimateurs sans biais ont à peu près la même précision que ceux associés à un plan équiprobable. De plus il s'avère que l'introduction des stratifications améliore les précisions dans les mêmes proportions pour les deux plans.

D'autre part il apparaît également que la surface totale des districts n'est pas suffisamment corrélée aux variables que nous cherchons à estimer pour que les estimateurs par le ratio faisant intervenir cette surface dans leurs dénominateurs soient meilleurs que les estimateurs sans biais d'Horwitz-Thompson.

Enfin, comme l'ensemble de ces résultats est quasiment invariant lorsqu'on change le caractère à estimer, on peut, au vu de ces conclusions, considérer que le sondage que nous cherchons à définir doit avoir les caractéristiques suivantes :

1. La base de sondage est l'ensemble de l'aire métropolitaine matérialisée sur l'image satellite par la limite de la zone considérée comme urbaine.

2. Cet espace est stratifié suivant un ensemble de critères morphologiques représentés par diverses classifications de l'image satellite. Parmi ces critères, le premier et certainement le plus important est la densité du bâti.

3. La sélection de l'échantillon s'effectue grâce à un tirage systématique où la probabilité d'appartenance à l'échantillon d'une unité spatiale donnée est proportionnelle à une mesure de sa taille. La plus "naturelle" de ces mesures est évidemment la surface totale de l'unité, c'est aussi la plus facile à mettre en oeuvre lors du tirage mais ce n'est bien sûr pas la seule possible.

4. Les estimateurs sont sans biais lorsqu'il s'agit d'estimer des totaux (estimateurs classiques d'Horwitz-Thompson) mais les estimateurs par le ratio restent utiles pour estimer directement des quotients de totaux.

Analyse des spécificités du plan défini

Par rapport aux sondages les plus couramment pratiqués, par exemple les sondages aléatoires équiprobables, le sondage que nous proposons présente deux particularités principales :

1. En tant que sondage stratifié à probabilités d'inclusion inégales, il nécessite, pour que l'on puisse extrapoler correctement les données recueillies sur l'échantillon au domaine d'estimation choisit, la connaissance d'informations exogènes :

- Sur chacune des unités spatiales enquêtées, outre l'ensemble de l'information correspondant aux variables que l'on cherche à estimer, il faudra connaître la taille de l'unité puisque cette grandeur entre dans le calcul de l'estimateur. Il s'agira très probablement de sa surface totale et il faut donc disposer d'un moyen fiable de la mesurer.

- Il faut également connaître la taille de chaque strate, quand cette taille est la surface totale, sa mesure s'effectue sans problèmes sur l'image satellite.

2. Le caractère systématique du sondage, en assurant une bonne répartition géographique de l'échantillon, améliore la précision des estimations. Il est malheureusement impossible de mesurer rigoureusement ce gain puisqu'il n'existe pas, dans le stricte cadre de la théorie des sondages, d'expression de la variance des estimateurs sous un tel plan. Les tests dont il est question au chapitre 7 ont néanmoins permis de le quantifier approximativement. En tant que tel, le tirage

systematique fait diminuer la variance de 20 à 30 % par rapport au tirage strictement aléatoire au même taux. Si l'on considère l'effet cumulé du tirage systématique et de la stratification, l'allocation aux strates étant optimale, le gain total se situe, suivant les caractères estimés, entre 45 et 95 % de la variance de départ (tirage aléatoire non stratifié).

. Questions en suspens, perspectives de travail à Quito

L'expérience menée à Marseille aura donc permis de préciser un certain nombre des caractéristiques que doit avoir un sondage à but démographique en milieu urbain utilisant comme base de sondage une image satellite stratifiée. Cependant, d'une part la description de ce plan de sondage reste encore incomplète et d'autre part il reste à vérifier la stabilité des résultats acquis à Marseille lorsqu'on change de site d'application. En conséquence, trois questions principales encore en suspens peuvent être formulées :

1. L'usage de l'image satellite comme base de sondage aréolaire et support de la sélection de l'échantillon implique, on l'a vu, un plan de sondage où la probabilité d'appartenance d'une unité à l'échantillon est proportionnelle à la taille de cette unité. Reste à définir quelle est la mesure de cette taille la plus efficace au sens de la précision du sondage. Concrètement, compte tenu de l'information actuellement disponible grâce à l'interprétation des images satellite, deux alternatives paraissent possibles : soit la mesure de la taille des unités spatiales est leur surface totale, soit il s'agit d'une approximation de leur surface bâtie.

Pratiquement, la question est donc : a-t-on intérêt, du point de vue de la précision des estimations produites, à choisir une probabilité de sélection proportionnelle à la surface bâtie des unités plutôt qu'à leur surface totale? Dans l'hypothèse affirmative, existe-t-il une stratégie de sélection de ces unités praticable sur l'image satellite qui garantisse une telle probabilité?

2. L'expérience marseillaise a permis de définir en partie le plan de sondage applicable à des unités spatiales du type districts de recensement INSEE. Ces unités correspondent en gros aux pâtés de maison physiques identifiables sur l'image satellite. Or, l'unité statistique prise en compte lors des enquêtes démographiques est soit l'individu soit, plus généralement, le ménage. Le plan de sondage complètement défini doit donc comprendre plusieurs degrés, le premier degré correspond à la sélection des pâtés de maison et le dernier à celle de l'unité statistique enquêtée, l'individu ou le ménage.

Il reste donc à fixer le nombre de degrés intermédiaires du sondage, les taux et les techniques de sélection à pratiquer pour chaque degré. Ces choix seront faits de manière à optimiser la précision du sondage pour un coût de mise en oeuvre donné c'est à dire pour un nombre de ménages à enquêter fixé à l'avance.

3. Quel degré de stabilité offrent les résultats acquis à Marseille lorsque l'on change de site? La méthode définie ici nécessite-t-elle des adaptations pour être efficace ailleurs ?

A l'issue de l'expérience menée à Marseille, trouver des réponses aux questions formulées ci-dessus constitue l'objectif principal de la phase suivante que nous réalisons actuellement à Quito. Son déroulement comprend les étapes suivantes :

1. Sélection d'un échantillon de "manzanas" (pâté de maison à Quito) et collecte d'informations morphologiques et démographiques sur cette échantillon.

2. Application des outils statistiques et informatiques mis au point à Marseille à ce corpus de données et réponse aux questions restant posées.

3. Définition complète du plan de sondage à mettre en oeuvre lors de l'enquête démographique finale à Quito et des techniques adéquates pour sa réalisation sur document satellitaire.

4. Sélection de l'échantillon et réalisation de l'opération de collecte, contrôle, codification et saisie des données collectées.

5. Développement du programme informatique permettant l'extrapolation des données de l'échantillon aux domaines d'estimations retenus : sous-strates, strates, ensemble de l'agglomération.

6. Exploitation de l'enquête, production d'estimations de caractères démographiques et d'indicateurs de migrations sur la ville de Quito, estimation de la précision des données produites, conclusion sur l'efficacité réelle de la méthode mise au point.

Actuellement, les quatre premières étapes ont été réalisées avec succès. Au rythme actuel d'avancement des travaux, les deux dernières devraient trouver leur conclusion avant le mois de juin 1988, date à laquelle nous pourrions fournir le dernier rapport à l'ORSTOM contenant les conclusions finales de ce travail.

BIBLIOGRAPHIE

. Organisation et exploitation du recensement de 1982

- INSEE Recensement général de la population de 1982. Guide d'utilisation. Tome 1 : Présentation générale, organisation, dénombrement; exploitation statistique, publications. Paris.
- INSEE Recensement général de la population de 1982. France métropolitaine. Structure de la population totale. Paris.
- CUEUGNIET C., DION J.F., FORTIER M., 1984. Recensement général de la population de 1982 : l'exploitation statistique exhaustive. INSEE, Courrier des statistiques n° 30. Paris.
- LERY A., 1983. Recensement général de la population de 1982 : aperçu de l'exploitation en cours. Courrier des statistiques n° 26, pp 43-50. Paris.

. Démographie, Marseille

- AGENCE D'URBANISME DE L'AGGLOMERATION MARSEILLAISE. Les tendances générales 62-68-75. Analyse comparative : communes, département, région, France. Cahiers du recensement n° 1. Marseille.
- AGENCE D'URBANISME DE L'AGGLOMERATION MARSEILLAISE. La population active ayant un emploi. Analyse par branche d'activité, puis par arrondissement (emplois au lieu de résidence). Cahiers du recensement n° 2. Marseille.
- AGENCE D'URBANISME DE L'AGGLOMERATION MARSEILLAISE. Les catégories socio-professionnelles à Marseille depuis 1954. Cahiers du recensement n° 3. Marseille.
- AGENCE D'URBANISME DE L'AGGLOMERATION MARSEILLAISE. Les conditions de logement des ménages marseillais. Tendances générales. Cahiers du recensement n° 4. Marseille
- DONZEL A., 1983. Marseille : politiques urbaines et société locale. Université d'Aix en Provence. Thèse de troisième cycle de sociologie. Aix en Provence

OBSERVATOIRE ECONOMIQUE PROVENCE ALPES COTE D'AZUR, 1983. Le recensement général de la population de 1982. Marseille par arrondissement et par quartier: Dénombrement de la population et des logements. Marseille.

. Utilisation des données de télédétection aérienne et spatiale pour des estimations démographiques

- ADENIYI P. O., 1983. An aerial photographic method for estimating urban population. Photogrametric engineering and remote sensing, vol. 49, n° 4, pp 545-560.
- BALLUT A., GAUTHIER M., 1983. Une méthode pour la production de données socio-économiques spatialisées dans le cadre d'un schéma directeur d'aménagement et d'urbanisme. IAURIF. Paris.
- GREEN N.E., 1956. Aerial photographic analysis of residential neighbourhoods : an evaluation of data accuracy. Social forces, n° 35, pp. 142-247.
- HOLZ R.K., HUFF D.L., MAYFIELD R.C., 1969. Urban spatial structure based on remote sensing imagery. Sixth International Symposium Remote Sensing Environment, vol. 2, pp. 819 - 830.
- HSU S.Y., 1971. Population estimation. Photogrametric engineering and remote sensing, vol. 37, n° 5, pp. 819 - 830.
- IAURIF, ROC, SIDAU, SEE, 1983. Les données urbaines. Système permanent de production rapide et économique. Recherche et synthèse. Paris.
- KRAUS S.P., SENGER L.W. RYERSON J.M. 1974. Estimating population from photographically determined residential land use types. Remote sensing of environment, n° 3, pp. 35 - 42.
- LO C.P., WELCH R., 1977. Chinese urban population estimates. Annals of the Association of American Geographers, vol. 67, n° 2, pp 246 - 250.
- PORTER P.W., 1956. Population distribution and land use in liberia. London School of Economics and Political Sciences. Thèse de Ph. D.
- VERNIERE M., 1978. Méthode d'analyse quantitative de la croissance urbaine dans l'espace et dans le temps. Exemple d'une banlieue de Dakar (Sénégal). Photo interprétation, n° 1, pp. 34 - 55.

WATKINS J.F., 1985. Small area production estimates using aerial photography. Photogrametric engineering and remote sensing, vol. 51, n° 2, pp. 1933 - 1935.

. Télédétection spatiale en milieu urbain

BALLUT A., 1980. L'étude de l'occupation du sol par satellite. Cahiers de l'IAURIF, vol. 61. Paris.

BOQUET E., 1980. Réactualisation du mode d'occupation du sol par l'utilisation des simulations SPOT (81 et 83). Mémoire de maîtrise de géographie. Université Paris VII. Paris.

CNRS, 1982. Débats des journées télédétection en milieu urbain (6 - 7 mai 1982). CERG, Paris.

IGN, 1984. Recherches méthodologiques sur l'utilisation de la télédétection spatiale en milieu urbain. IGN, Paris.

THIBAUT C. BIANCALE M. DELAVIGNE R., 1983. La télédétection par satellite appliquée au milieu urbain : test de validité de l'indice de végétation dans la proche couronne. IAURIF, Paris.

WILMET J., SOYER J., 1982. Lumbumbashi et le sud-est du haut Shaba : interprétation de données LANDSAT. Bulletin de la Société Belge d'Etude Géographiques, pp. 87 - 100.

. Analyse des données

BENZECRI J.P. et coll., 1973. L'analyse des données. Tome 1 : La taxinomie, Tome 2 : L'analyse des correspondances. DUNOD, Paris.

BENZECRI J.P. et coll., 1980. La pratique de l'analyse des données. vol. 2 : abrégé théorique, étude de cas modèles. DUNOD, Paris.

CHANDON J.L., PINSON S., 1981. Analyse typologique, théories et applications. MASSON, Paris.

DIDAY E. et coll., 1980 : Optimisation en classification automatique (2 vol.). INRIA, Paris.

JAMBU M., 1978. Classification automatique pour l'analyse des données, vol. 1 : Méthodes et algorithmes, vol. 2 : Logiciels. DUNOD, Paris.

LEBART L., MORINEAU A., FENELON J.P., 1982. Traitement des données statistiques, méthodes et programmes. DUNOD, Paris.

VOLLE M., 1981. L'analyse des données. ECONOMICA, Paris.

. Théorie des sondages et variables régionalisées

ASSELIN L.M., 1984. Techniques de sondage avec application à l'Afrique. GAETAN MORIN, Québec.

COCHRAN W.G., 1977. Sampling techniques (3ème édition). J. WILLEY, New York.

DELINCE J., 1983. Estimations robustes de densité dans le plan, étude comparée de techniques d'échantillonnage, application en forêt. Université Catholique de Louvain, Faculté des Sciences Agronomiques. Louvain.

DEMING W.E., 1950. Some theory of sampling. J.WILLEY, New York.

DEROO M., DUSSAIX A.M., 1980. Pratique et analyse des enquêtes par sondage. PUF, Paris.

DESSABIE J., 1966. Théorie et pratique des sondages. DUNOD, Paris.

DEVILLE J.C., ROY G., 1984. L'échantillon maître fait peau neuve. Courriers des statistiques n° 29, Paris.

GOURIEROUX Ch., 1981, Théorie des sondage. ECONOMICA, Paris.

GROSBRAS J.M. 1987. Méthodes statistiques des sondages, ECONOMICA, Paris

GUIDICELLI X., LANLY J.P., OUAKAM J.B., PIETRI M., 1972 (ISUP).
Application de la théorie des processus aléatoires à l'estimation de la précision d'un inventaire forestier par échantillonnage systématique. Annales des Sciences Forestières, Institut National de la Recherche Agronomique. Paris.

MATHERON G., 1965. Les variables régionalisées et leur estimation, une application de la théorie des fonctions aléatoires aux sciences de la nature. MASSON, Paris.

RAJ D., 1968. Sampling Theory. MC GRAW HILL, New York.

RAJ D., 1972. The design of sample surveys. MC GRAW HILL, New York.

- RAO J.N.K., HARTLEY H.O., COCHRAN W.G., 1962. On a simple procedure of unequal probability sampling without replacement. Journal of Royal Statistical Society B24, pp. 482 - 491.
- SUKHATME P.V., SUKHATME B.V., 1970. Sampling theory of surveys with applications. FAO, Rome.
- THIONET P., 1958. La théorie des sondages. INSEE, Paris.
- YATES F., 1981. Sampling methods for censuses and surveys (4ème édition). GRIFFIN, Londres.

. Informatique, programmation

- COHEN R., 1986. Programmer en dbase III et dbase III plus. éditions du PSI, Paris.

. Logiciels utilisés

- DBASE III PLUS (ASTHON TATE) : Gestion de base de données.
- SPSS PC + (SPSS inc.) : Traitements statistiques.
- STAT ITCF (Institut Technique des Céréales et Fourages) : Traitements statistiques.
- ADAD MICRO (ADAD) : Analyse des données.
- WORD (MICROSOFT) : Traitement de textes.

ANNEXES

ANNEXE 1 : Listage du programme VARECH (*)

ANNEXE 2 : Publications de l'équipe ORSTOM

(*) : Cet annexe, qui contient le listage complet du programme dbaseIII ayant permis les calculs de variance d'estimateurs effectués dans ce rapport, est disponible à l'ORSTOM, aux archives du département SDU.

ANNEXE 2 : PUBLICATION ET NOTES DE TRAVAIL DE L'EQUIPE ORSTOM
(peuvent être consultés au Département D de l'ORSTOM)

Articles

- DUREAU F. et GUILLAUME A., 1984. La démographie depuis l'espace : un nouveau système d'observation ? Paris, STATECO n° 38, pp 5-46.
- DUREAU F., 1986. A propos du traitement informatique de données localisées. Une expérience en cours : télédétection et observation des populations urbaines. Paris, ORSTOM, collection colloques et séminaires, numéro sur le traitement des données localisées, pp.263-286.
- DUREAU F. et BARBARY O., 1987. Pour une méthode de collecte démographique en ville intégrant la télédétection spatiale. A paraître dans la revue informatique et sciences humaines, EHESS, 42 p.
- MICHEL A., DUREAU F., LORTIC B., SOURIS M., 1987. Mise au point des méthodes d'analyse des images satellites à haute résolution et évaluation des informations fournies par ces images. 1 : Etude statistique du descripteur "densité du bâti". Présentation des classifications. 2 : Mise en évidence des réseaux routiers sur une image SPOT panchromatique. Etude de faisabilité sur Quito (Equateur). Paris, service technique de l'urbanisme, Bilan sur la télédétection urbaine en France. 10 p. et photos.
- SOURIS M., 1986. Système d'information géographique et bases de données. Paris, ORSTOM, collection colloques et séminaires. Numéro sur le traitement des données localisées, pp.29-87.

Communication à des colloques

- DUREAU F. et GUILLAUME A., 1985. La population dans l'espace. Télédétection et observations démographique des villes des pays en développement. Communication au XX^e congrès général de l'Union Internationale pour l'Etude Scientifique de la Population, Florence (Italie), 5-12 Juin 1985.

- DUREAU F., LORTIC B., MICHEL A., SOURIS M., 1987.
Télédétection et système d'information géographique. Communication au Forum International de l'Instrumentation et de l'Information Géographique, Lyon, (France), 10-13 Juin 1987, 15 p.
- DUREAU F., LORTIC B., MICHEL A., SOURIS M., 1987.
Informatique, télédétection et observation des populations urbaines. Une recherche en cours à Quito (EQUATEUR). Communication à la "primera conferencia Latinoamericana sobre informatica en geografia", San Jose; Costa Rica, 5-9 Octobre 1987, 29 p.
- MICHEL A., EBERHARD J.M., LORTIC B., DUREAU F., 1987. El uso de la teledeteccion para la observacion de las poblaciones urbanas. Investigacione en curso en Quito (ECUADOR). Communication au deuxième "Simposio Latinoamericano sobre sensores remotos", Bogota, Colombie, 16-20 Novembre 1987, 34 p.
- MICHEL A., EBERHARD J.M., LORTIC B., DUREAU F., 1987.
Utilisation de la télédétection pour l'observation de populations urbaines. Une recherche en cours à Quito (EQUATEUR). Communication au colloque SPOT, Paris, 23-27 Novembre 1987, 9 p.
- SOURIS M., 1987. A geographic information system with relational architecture. Principles and exemple of use of the TIGER system. Communication à la "primera conferencia Latinoamericana sobre informatica en geografia", San Jose, Costa Rica, 5-9 Octobre 1987.

Notes de travail non publiées

- Equipe ORSTOM (UR 406, ATOB, Unité d'Infographie), 1986.
Intégration des données de télédétection dans un système d'information géographique : suivi de la morphologie et de la démographie d'une ville. Avril 1986, Paris, 190 p.
- Equipe ORSTOM (UR 406, ATOB, Unité d'Infographie), 1986.
Intégration des données de télédétection dans un système d'information géographique : suivi de la morphologie et de la démographie d'une ville.
Rapport final, Août 1987, Paris, tome 1 : 53 p., tome 2 (annexes) 538p.

- BARBARY O., 1985. Projet de thèse de mathématiques appliquées aux sciences de l'homme, sous la direction de M. BARBUT, EHESS. Sujet : "Elaboration de plans de sondage pour l'estimation de données spatio-temporelles prenant en compte les informations fournies par la télédétection. Application aux données démographiques en milieu urbain des pays en développement". Paris, 20 p.
- BARBARY O., 1986. Apport d'une stratification morphologique des districts INSEE de recensement dans l'estimation par sondage des effectifs de population à Marseille. Rapport intermédiaire à l'ORSTOM, Quito, 47 p.
- BARBARY O., 1988. Mise au point de plans de sondage pour l'estimation de caractères démographiques à Marseille. Evaluation des précisions permises par l'emploi de diverses techniques de sélection, de stratification et d'estimation. Rapport intermédiaire à l'ORSTOM, Quito, 185 p.
- DUREAU F., 1986. Marseille : tirage de l'échantillon de travail. Quito, 9 p.
- DUREAU F., MICHEL A., 1986. Quito : enquête morphologie urbaine, instructions aux enquêteurs. (2 versions, Français et espagnol), Quito, 19 p.
- DUREAU F., 1987. Quito : enquête bâti/population, instructions aux enquêteurs. (2 versions, Français et espagnol), Quito, 27 p.
- DUREAU F., 1987. Quito : encuesta demografía/migraciones, instrucciones para los encuestadores, Quito, 25 p.
- EBERHARD J.M., 1987. Télédétection urbaine : éléments de caractérisation géographique des quartiers de Quito à l'aide d'une image SPOT. Mémoire de DEA, EHESS, 98 p.
- MICHEL A., 1986. Projet de thèse de géographie, sous la direction de J.P. RAISON, EHESS. Sujet : "Télédétection et observation suivie de la morphologie et de la démographie des villes des pays en développement. Mise au point des méthodes d'analyse des images satellite à haute résolution et évaluation des informations fournies par ces images.", Paris, 9 p.

- MICHEL A., 1986 (et version actualisée en 1987). Le point sur la télédétection urbaine en 1986 à travers la bibliographie française et anglo-saxonne sur le sujet. Recueil des fiches de dépouillement des articles et ouvrages. Paris, ORSTOM, 49 p.
- MICHEL A., 1987. Mise au point des méthodes d'analyse des images satellites à haute résolution : étude statistique du descripteur "densité du bâti", présentation des classifications (image TM Marseille, 1983), Paris, 99 p.
- MICHEL A., 1987. Influence des facteurs météorologiques sur la probabilité d'obtention d'une image satellite SPOT ou Thématic Mapper. Etude sur le site de Quito, Equateur. Quito, 14 p.

Rapports de missions

- BARBARY O., DUREAU F., 1986. Rapport de mission Montpellier 18-19 Février 1986, CNUCS et Maison de la Géographie, Paris, 6 p.
- BARBARY O., DUREAU F., LORTIC B., MICHEL A., 1985. Rapport de mission à Marseille, Paris, 3p et annexes.
- DUREAU F., 1987. Rapport de mission à Bogota (Colombie), 13-20 Novembre 1987, Quito, 4p et annexes.
- LORTIC B., 1987. Rapport de mission à Quito, 31 Décembre 86 - 18 Janvier 87, Paris, 5 p.
- MICHEL A., 1986. Rapport de mission à Quito, 16 Septembre - 17 Novembre 1986, Quito, 7p et annexes.
- MICHEL A., 1987. Rapport de mission à Quito, 1 Septembre - 10 Octobre 1987, Quito, 9p et annexes.
- SOURIS M., 1986. Rapport de mission à Quito, 8-22 Décembre 1986, Quito, 5 p.
- SOURIS M., DUREAU F., 1985. Rapport de mission en Equateur, 15-29 Mai 1985 (M. SOURIS) et 19-26 Mai 1985 (F. DUREAU), Paris, ORSTOM, 6 p.