

# *La démographie appliquée à la gestion publique et des entreprises*

*Actes du séminaire de l'AIDELF en Calabre (Cosenza, avril 1995),  
et de la session spéciale du Congrès de l'EAPS (Cracovie, juin 1997)*

*Édité par :*

*Giuseppe De Bartolo  
et Michel Poulain*



ASSOCIATION INTERNATIONALE DES DÉMOGRAPHES DE LANGUE FRANÇAISE  
**AIDELF**

# Estimation de population locale pour la gestion publique et le marketing

---

Giuseppe De BARTOLO, Rosina CHIAPPETTA

Département d'Économie Politique  
Université de Calabre

## Introduction

Comme on le sait, les événements démographiques sont étroitement liés aux facteurs socio-économiques, par suite, dans le cadre de la programmation des interventions dans ce domaine, on se trouve souvent obligé de lier les décisions à la population de référence. La définition du type et des dimensions de ces interventions nécessite, au préalable, la connaissance de la situation présente aussi bien que l'évaluation des besoins futurs<sup>1</sup>. Tout cela implique une conscience de la dimensions de la population mais surtout la connaissance de sa structure selon certains caractères ; en outre, dans le but de saisir les besoins futurs, il est nécessaire de faire des prévisions démographiques.

Souvent, aussi bien l'analyse que la prévision doivent être réalisées sur des zones très limitées, à la suite de l'autonomie de plus en plus forte des entités subrégionales, comme les provinces, les communes, les unités sanitaires de base et les subdivisions scolaires. Ces entités, pour bien programmer les investissements de leur ressources, ont l'exigence de trouver un équilibre socio-démographique et par conséquent il leur est nécessaire, dans leurs choix et décisions, d'avoir les informations les plus fiables au sujet de la démographie du territoire de leur compétence.

A ce niveau territorial très restreint, les analyses et les prévisions démographiques ne sont pas faciles à réaliser à cause de la difficulté d'obtenir les données de base nécessaires ; et si la désagrégation est trop poussée, on ne connaît même pas la dimension démographique. Pour tout cela, on se trouve obligé d'utiliser des méthodes d'estimation et de prévision qui sont conçues pour faire face aux exigences susdites.

Les sources qui peuvent fournir des données pour estimer la population des petites zones sont d'abord des sources classiques : recensement, état civil ; mais on peut exploiter aussi les fichiers de certains organismes publics ou privés pour obtenir des données symptomatiques (comme les inscriptions à l'école primaire, les immatriculations des automobiles, les données sur les permis de construire, les dépôts bancaires, etc.) à travers lesquelles on peut évaluer la dimension démographique.

## 1. Caractères des méthodes d'estimation pour micro-zones

Nous passerons en revue certaines méthodes d'estimation de la population qui pourraient être employées pour les petites zones. Même si, parmi ces méthodes, certaines ont été construites pour combler des situations particulières (par exemple, pour les États-Unis), elles sont reportées ici car elles pourraient être adaptées aux autres situations<sup>2</sup>. Les estimations concernant la population se distinguent en :

---

<sup>1</sup> Cfr De Rita, G., Ruberto, A., L'utilizzazione delle previsioni demografiche nella programmazione regionale degli interventi sociali, *Rivista Italiana di Economia, Demografia e Statistica*, V, XXII, n. 1-4, 1968.

<sup>2</sup> Aux États-Unis, en effet, comme il n'existe pas, comme dans de nombreux pays européens, un registre de la population, la nécessité s'est révélée, plus qu'ailleurs, d'estimer la population durant les années intercensitaires au niveau de grandes et petites aires. Pour faire face à cette exigence, on a élaboré des méthodes non démographiques, de plus en plus perfectionnées avec le temps, qui emploient des variables symptomatiques du niveau de la population.

- Estimations inter-censitaires, appelées ainsi car elles tentent d'estimer le chiffre de la population dans les années comprises entre deux recensements, prenant en considération justement les informations de ces derniers.
- Estimations post-censitaires, qui utilisent les données du recensement précédent.
- Estimations acensitaires c'est-à-dire non basées sur les données censitaires.

Elles peuvent être classifiées aussi par rapport à d'autres caractéristiques : par exemple, elles peuvent être réalisées pour un pays entier, pour les subdivisions géographiques plus importantes ou bien pour une variété d'aires d'un même territoire (aires urbaines et rurales, typologies de villes) ; elles peuvent se référer à toute la population ou à certaines sous-populations différenciées par sexe, âge, état civil, à la population résidente ou à la population présente.

Pour les petites zones, la méthodologie dépend fortement des données disponibles et souvent, elle doit être adaptée à la situation concrète : une méthode conçue pour une certaine configuration de données et pour une situation donnée n'est pas nécessairement applicable à une autre typologie de données et à un autre milieu géographique. En outre, au niveau de petites zones, la qualité inégale de données réduit fortement la précision des estimations. Des méthodes d'estimation construites avec un input peu flexible sont peu adaptables aux champs des données très variables des petites zones. Il faut aussi souligner qu'à la faiblesse de la méthodologie, dont on a conscience, ne s'oppose pas toujours une prudence accrue de la part des utilisateurs des estimations et des prévisions.

## 2. Quelques méthodes d'estimation

En négligeant les techniques qui emploient des fonctions mathématiques spéciales, et qui pourraient être employées pour faire des interpolations et des extrapolations, dont les caractéristiques et les limites sont bien connues, nous examinerons ici, très brièvement, certaines méthodes qui emploient des variables considérées comme symptomatiques des variations de la population. Ces variables, étant souvent disponibles aussi pour des zones très petites, peuvent être utilisées pour permettre l'estimation de la population d'une zone très restreinte<sup>3</sup>.

### Rapport population/variable

Sans spécifier le nom de la variable symptomatique, l'estimation de la population d'une zone au temps  $t_2$ , en connaissant la valeur de la population et de la variable à un temps  $t_1$  (par exemple au moment du recensement), pourrait être obtenue par :

$$P_{t_2} = \left( S_{t_1} + U \right) \times \frac{P_{t_1}}{S_{t_1}}$$

où  $P_{t_2}$  est l'estimation de la population au temps  $t_2$ ,  $S_{t_1}$  est la valeur de la variable à une occasion précédente et  $U$  est la variation nette de la variable dans la période entre la dernière évaluation et le moment de l'estimation.

Étant donné que les valeurs de la variable au moment de l'estimation sont employées pour estimer la population, la précision de l'estimation est fonction de la stabilité dans le temps du rapport variable/population ou de la connaissance des changements de ces rapports.

Souvent les données symptomatiques sont produites pour des raisons très variées et elles ont une couverture géographique qui ne correspond pas à celle de la zone dont on veut estimer

<sup>3</sup> A ce sujet voir notamment : Murdock, S.H., Ellis, D.R., *Applied Demography, An Introduction to Basic Concepts, Methods and Data*, Westview Press, Boulder, 1991.

la population. Par conséquent, soit l'échelle temporelle, soit la couverture géographique des variables doivent être convenablement modifiées. Naturellement ces précautions sont extensibles à tous types d'estimation examinés ici. Comme exemple une liste des variables symptomatiques sera donnée après.

### Méthode basée sur les habitations

Dans cette méthode, on utilise les données sur les habitations. Les variables plus fréquemment employées sont le nombre et les variations des permis de construire et le nombre et les variations des consommateurs d'électricité. Dans ce cas, l'estimation de la population assume la forme suivante :

$$P_{t_2} = (H_{t_1} + U) \times \frac{P_{t_1}}{H_{t_1}}$$

où  $P_{t_2}$  est la population de la zone estimée à un temps postérieur au recensement,  $H_{t_1}$  est le nombre des habitations occupées dans la zone considérée au dernier recensement,  $U$  est la variation nette de la variable pendant l'intervalle recensement - moment de l'estimation,

$\frac{P_{t_1}}{H_{t_1}}$  est le nombre moyen des personnes par habitation au dernier recensement.

Pour utiliser cette méthode de la façon la plus correcte, il est nécessaire de connaître les données sur les démolitions et aussi sur les nouveaux permis, les premières étant plus fiables que les deuxièmes. Si l'on utilise les données sur les consommateurs d'électricité, on a besoin d'avoir des informations sur la qualité de ces données, sur les usages multiples et sur le nombre d'habitations corrélées à ces usages. Les avantages qu'on a avec cette méthode, c'est que la corrélation entre le nombre d'habitations et la population est forte ; les désavantages sont dus au fait qu'il est difficile d'évaluer les variations du nombre moyen des personnes occupant une habitation. Il a été prouvé que la qualité de cette méthode est susceptible d'amélioration<sup>4</sup>.

### Rapports élémentaires

Si l'on indique par  $R_1, R_2, \dots, R_n$  les rapports de certaines variables symptomatiques entre le moment de l'estimation et une date précédente, la moyenne de ces rapports peut être égale au rapport entre les populations aux mêmes dates. C'est-à-dire :

$$\frac{P_{t_2}}{P_{t_1}} = \frac{\sum_{i=1}^n R_i}{n} \Rightarrow P_{t_2} = \frac{\sum_{i=1}^n R_i}{n} \times P_{t_1}$$

Par exemple, si l'on utilise comme variables les naissances et les décès on a :

$$P_{t_2} = P_{t_1} / \left( \left( \frac{n_{t_2}}{n_{t_1}} + \frac{m_{t_2}}{m_{t_1}} \right) \div 2 \right)$$

### Méthode des taux bruts

Cette méthode consiste à rapporter, à deux dates différentes  $n_1$  et  $n_2$ , les taux bruts de natalité (ou de mortalité) d'une sous-zone et à égaliser tout cela au même type de rapport pour la zone plus grande. Après, en connaissant les événements (par exemple les naissances) de la sous-

<sup>4</sup> Smith, S. K., A Review and Evaluation of the Housing Unit Method of Population Estimation, *Journal of American Statistical Association*, v. 81, n. 394, 1986.

zone au moment  $n_2$ , on peut estimer la population totale de la sub-zone au même moment. En indiquant par  $Q^{sub}$  un taux brut de la sous-zone et par  $Q^{sup}$  le même taux pour la zone, on peut écrire :

$$\frac{Q^{sub} n_2}{Q^{sub} n_1} \frac{Q^{sup} n_2}{Q^{sup} n_1} \Rightarrow \dots Q^{sub} n_2 \quad Q^{sub} n_1 \times \frac{Q^{sup} n_2}{Q^{sup} n_1}$$

$$\Rightarrow \text{Population totale sous - zone au moment } n_2 \quad \frac{n \text{ événements sous - zone au moment } n_2}{Q^{sub} n_2}$$

### Méthode des proportions

Cette méthode consiste à égaliser les rapports entre la population de la sous-zone et celle de la zone dans deux temps différents, dans l'hypothèse que le rapport sous-zone /zone ne varie pas dans la période considérée. Pour deux temps  $t_1$  et  $t_2$  on a :

$$\frac{P_{t_2}^{sub}}{P_{t_2}^{sup}} = \frac{P_{t_1}^{sub}}{P_{t_1}^{sup}}$$

dont on peut obtenir l'estimation de la population de la sous-zone à l'époque  $t_2$ .

### Méthode composée

Ici la population totale est obtenue comme somme des estimations de la population des différentes classes d'âge, estimations obtenues au moyen des variables symptomatiques liées aux classes mêmes. Dans le tableau suivant, on a reporté quelques-uns de ces variables.

Classes d'âges	Variables
0- 5	Naissances des six dernières années
6- 10	Inscriptions à l'école primaire
11- 13	Inscriptions à l'école moyenne
14-49 (sexe féminin)	Taux global de fécondité générale
14-49 (sexe masculin)	Rapport entre les sexes
50-65	Taux de mortalité des personnes de 50 ans et plus
65+	Données sur les personnes retraitées

### Méthode de régression

Cette méthode peut être vue comme un instrument d'estimation dans lequel on emploie des indicateurs symptomatiques multiples avec poids variables. Voici une liste des variables symptomatiques les plus souvent utilisées :

- naissances
- décès
- inscriptions à l'école primaire
  
- usagers de téléphone privés
- usagers électriques
- données sur les votes
- immatriculations des automobiles

– dépôts bancaires

qui peuvent servir pour estimer une relation de régression multiple du type :

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_nx_n$$

où y est la valeur de la population, les  $x_i$  représentent les variables symptomatiques,  $b_i$  les poids obtenus sur la base de l'expérience passée.

La relation précédente est évaluée pour une période antécédente et est supposée valable pour un moment donné ; par conséquent, la connaissance des valeurs des variables symptomatiques permet d'obtenir une estimation de la population. Pour que cette méthode puisse être employée, il faut connaître les valeurs des variables pour chaque année.

La méthode de régression est utilisée pour évaluer les rapports entre la valeur de la population d'une sous-zone et celle de la zone (Ratio-Correlation Method), plutôt que pour faire une estimation directe de la population<sup>5</sup>. Plus précisément le RCM est une technique de régression multivariée où les variables indépendantes sont en général des rapports entre indices de « composition », calculés en temps différents.

Si pour simplifier on se limite à une seule variable symptomatique, qu'ici nous avons indiquée par S, l'équation de régression à estimer aura la forme suivante :

$$\frac{P_{t_2}^{sub} \div P_{t_2}^{sup}}{P_{t_1}^{sub} \div P_{t_1}^{sup}} = b_0 + \frac{S_{t_2}^{sub} \div S_{t_2}^{sup}}{S_{t_1}^{sub} \div S_{t_1}^{sup}}$$

La méthode Ratio-Correlation parmi les techniques d'estimation de population est la plus employée parce qu'elle utilise une pluralité d'indicateurs symptomatiques des variations de la population et en plus elle permet d'employer les estimations de la zone plus grande, qui sont en général robustes. Sa faiblesse au contraire est due au fait qu'on doit accepter passivement toutes les hypothèses du modèle de régression. Différentes variantes de cette méthode ont été proposées, mais aucune ne s'est révélée la meilleure dans toutes les situations.

De l'examen des méthodes d'estimation de la population que nous avons considérées, sommairement, les plus simples, il est possible de tirer quelques leçons. Les techniques symptomatiques sont très utiles pour estimer la population d'une petite aire, soit qu'on ait, souvent, la disponibilité immédiate de nombreuses variables, soit parce que ces techniques sont désormais expérimentées depuis longtemps. La plus grande faiblesse de ces méthodes réside dans la difficulté d'identifier les changements qui interviennent dans les relations entre les variables et la population. Aucune méthode n'est la meilleure dans toutes les situations à cause des grandes différences qui existent entre pays, dans les sources et dans la qualité des données. Par exemple, aux U.S.A, on a observé que les données sur les inscriptions scolaires sont des variables très sensibles des groupes d'âges juvéniles, la même chose vaut pour les décès et l'âge sénile. Dans d'autres pays, ces relations peuvent ne pas être aussi fortes. Le choix des sources et des méthodes pourrait varier en fonction de la nature des statistiques et des estimations voulues. Nous voyons enfin que le choix d'une méthode plutôt que d'une autre est

---

<sup>5</sup> Pour un exemple d'estimation de la population italienne qui utilise la régression multiple, voir De Bartolo, G. , *Stima della popolazione presente delle provincie italiane mediante il metodo della regressione multipla*, Saggi di Statistica Metodologica ed Applicata, Contributi del Centro Interdipartimentale di Studi Statistici, v. I, Università degli Studi della Calabria, 1976. Dans cette étude, cinq variables symptomatiques ont été employées, naissances, décès, mariages, inscriptions scolaires élémentaires, relevées au niveau provincial. La relation a concerné la variation de la population dans la période 1961-71, et les variations, à la même période, des variables symptomatiques adoptées. L'équation de régression estimée a été donc utilisée pour évaluer la population présente des provinces italiennes en 1974.

une opération très délicate car, quelque soit la meilleure méthode, cela ne sera clair qu'au moment où l'événement se sera vérifié<sup>6</sup>.

### 3. Les ordures comme source de connaissance en démographie L'expérience du Garbage Project<sup>7</sup>

Les ordures représentent une source d'information non seulement dans le domaine scientifique mais aussi dans le domaine commercial, social et en particulier dans le domaine démographique, car elles sont un document direct, minutieux et incontestable des habitudes de celui qui les a produites, au-delà même de ses convictions<sup>8</sup>.

En examinant des ordures solides, on peut saisir les préférences d'achat d'un ménage. Leur quantité permet aussi de parvenir à une estimation de la population et donc, c'est un moyen pour évaluer la taille démographique d'une très petite zone, comme un hameau, un quartier et pour évaluer aussi les caractéristiques démographiques des ménages de ces zones.

L'étude des ordures comme moyen de connaissance de la composition sociale d'une population est une discipline relativement nouvelle, qui s'est développée aux États-Unis depuis 1973 et qui est l'oeuvre d'un groupe d'anthropologues de l'Université de l'Arizona.

En se référant à cette expérience, très fructueuse et de longue durée, et en particulier au Garbage Project (GP) développé dans cette université, il a été mis en évidence que, par l'examen des ordures solides, il est possible d'avoir une idée précise du schéma de consommation de certains groupes particuliers (Latino-américains, Anglo-saxons). En outre, on a pu constater que ces types de corrélation sont très stables et qu'ils ont une validité générale soit pour ce qui concerne le comportement d'achat, soit pour ce qui concerne la liaison entre masse de population et masse d'ordures solides.

En 1986 le Bureau du Recensement américain a demandé le soutien du Garbage Project pour tenter de résoudre le problème chronique de la sous-estimation de la population dans certaines zones critiques : ghettos, zones à forte immigration clandestine, etc. Un des buts du GP a été d'évaluer si, en général, il était possible de reconstruire la structure d'un quartier par sexe et par âge sur la base de la quantité et de la qualité d'ordures solides. Ce projet a été confié à des archéologues qui sont familiarisés avec ce genre de recherche. Dans ce cas particulier, ils avaient comme but la détermination d'un nombre « magique » qui, appliqué à certaines quantités de certaines types d'ordures solides, aurait pu permettre une bonne estimation de la population. Il faut souligner que les chercheurs du GP avaient l'avantage d'avoir à leur disposition toutes les informations sur la consommation alimentaire et, surtout, sur la production d'ordures de 63 familles observées sur une période de temps suffisamment longue. Grâce à cette base de données, il a été possible d'avoir une idée correcte du schéma de production d'ordures d'une famille américaine type.

Le premier objectif était de passer de la quantité d'ordures ou de la quantité d'un certain type d'ordures à l'estimation de la population totale : c'est-à-dire déterminer le nombre « magique » qui, multiplié par le poids total d'ordures produites, ou multiplié par le poids d'un des seize types d'ordures classifiées par un certain groupe de familles, puisse permettre une estimation de la population. Pendant cette recherche, on s'est aperçu que les catégories « total ordures solides » et « plastique » ont le plus fort pouvoir prédictif. Réduite à l'essentiel, voici la relation trouvée qui lie la production d'ordures de plastique à la population :

<sup>6</sup> Purcell, N. J., and Kish, L., Postcensal Estimates for Local Areas (or Domains), *International Statistical Review*, 48 (1980), p. 7.

<sup>7</sup> Cfr Rathje, W. and Murphy, C., *Garbage Demographics*, *American Demographics*, May 1992. -Rubbish, *The Archaeology of Garbage*, Harper Collins Publisher, 1992.

<sup>8</sup> Viale, G., *Un mondo usa e getta*, Feltrinelli, Milano, 1994, pp. 28-36.

$$Y = 0,2815 X$$

où **X** représente la quantité de plastique en livres, produite pendant cinq semaines et **Y** la taille du ménage.

Les chercheurs reconnaissent que l'estimation de la population totale obtenue soit en employant la catégorie « total ordures solides », soit la catégorie « plastique » a un écart maximal de plus ou moins 2,5% par rapport à la valeur réelle de la population.

La recherche, visant à estimer la structure par sexe et par âge, a été plus difficile. La classe d'âge qu'on peut plus facilement estimer est celle des enfants. Les couches pour nouveau-nés sont un bon indicateur de cette tranche d'âge parce que la régularité et le volume de ce genre de déchets aident à éliminer les distorsions statistiques.

Voici la relation qui a été trouvée et qui relie la population au nombre d'enfants :

$$Y = 0.01506 X$$

où **X** représente le nombre de couches pour nouveau-nés trouvées dans la poubelle pendant une semaine et **Y** la population infantile correspondante.

En réalité, il y a des déchets « exclusifs » qui pourraient permettre d'estimer la proportion entre les sexes, lier les enfants aux vieux, mais ils n'ont pas la même force prédictive que les couches pour nouveau-nés, car ils sont produits en quantité et fréquence insuffisantes pour ce type d'analyse.

Les chercheurs du Projet ont réussi à déterminer une estimation de la proportion des nouveau-nés sur la base des paramètres suivants :

- taux moyen d'ordures jouets et boîtes : 2,52 par enfant par semaine ;
- taux moyen d'ordures vestimentaires et boîtes : 0,87 par enfant par semaine.

Il a été possible également d'obtenir la proportion de femmes adultes dans une population sur la base des catégories suivantes d'ordures féminines :

- produits hygiéniques féminins : 1,58 par femme par semaine ;
- produits de beauté : 0,86 par femme par semaine ;
- articles d'habillement : 0,62 par femme par semaine.

L'objectif du Projet qui voulait trouver des indicateurs robustes de la présence masculine dans les ordures solides a échoué : en effet, les femmes peuvent boire et manger comme les hommes ; elles fument des cigarettes comme les hommes. Les mégots de cigares ou les boîtes de lingerie masculine pourraient être des indicateurs exclusifs pour le sexe masculin, mais leur présence dans les poubelles n'est pas en quantité suffisante pour établir une relation forte.

Pour résoudre le problème de la détermination numérique de la population masculine, une voie indirecte a été indiquée : de la population totale on soustrait d'abord les nouveau-nés, ensuite les enfants et enfin les adultes de sexe féminin. L'estimation ainsi obtenue est accréditée avec un écart de plus ou moins 10% vis-à-vis de la vraie valeur.

Le Bureau de Census a jugé qu'il était dangereux, pour un bureau gouvernemental d'employer des gens pour examiner les ordures et il a donc renoncé, à ce moment là, à utiliser cette méthodologie de correction. Mais ce projet pourrait être appliqué en l'an deux mille.



## Conclusions

Dans cette communication, nous avons souligné l'importance de la composante démographique dans la programmation des interventions sociales et économiques. Cette activité de programmation, qui est aujourd'hui de plus en plus fréquente au niveau local (communes, unités sanitaires de base, districts scolaires) ne nécessite pas seulement la connaissance de la population, mais aussi, souvent, de sa structure selon certaines caractéristiques. A ce niveau, les données pour faire une analyse démographique adéquate aux objectifs, sont souvent insuffisantes et tout cela affaiblit considérablement les efforts pour ancrer la programmation à des bases solides. Ainsi, nous avons examiné certaines méthodes d'estimation de la population qui nous paraissent utiles pour les petites zones, et dont nous avons fait une analyse critique. Enfin, nous avons souligné les potentialités informatives des ordures pour saisir la composition sociale, démographique d'une population, et aussi pour connaître son comportement d'achat. Cette approche pourrait avoir des utilisations plus générales dans un future proche.