

Le voisinage : proposition d'indicateurs

Bernard AUBRY*

Présentation générale

La méthode proposée présente l'avantage de la simplicité et de son adéquation à l'information existante dans la mesure où l'on utilise l'une des spécificités des fichiers des recensements, à savoir le classement des logements selon un critère de proximité. Cela vaut pour les recensements classiques (c'est-à-dire jusqu'en 1999), mais aussi peut-être, dans une certaine mesure seulement, pour les enquêtes annuelles de recensement telles qu'elles sont réalisées depuis 2004. La méthode doit pouvoir s'appliquer à chaque fois que l'on dispose d'informations à l'adresse, ce qui devrait être de plus en plus fréquent avec le traitement, sous forme cryptée, des données administratives (par exemple, celles figurant dans les fichiers des déclarations annuelles des données sociales (DADS), des caisses d'allocations familiales, etc.).

Résumée en quelques mots, la méthode consiste à comparer une distribution observée à une distribution théorique, comme on le fait classiquement en statistique. L'unité statistique est la grappe formée de deux logements voisins. La variable d'intérêt est le nombre de logements de la grappe disposant d'un caractère donné. Si la variable d'intérêt est distribuée sur le territoire de façon aléatoire, la variance de la distribution n'est autre que la variance d'une loi binomiale. Si les ménages ont tendance à se regrouper, il se crée une « *distance* » entre la distribution observée et la distribution théorique de référence. C'est cette « *distance* » qui est à la base des indicateurs qui seront utilisés pour caractériser cet « entre soi ».

* INSEE Alsace au moment de la rédaction de cet article.

Ce travail n'aurait pu être réalisé sans le précieux concours de Claire Glénat et de Jacques Postic qui ont bien voulu en cas de nécessité me porter secours et assistance. Merci à Typhaine Aunay pour avoir accepté de relire le texte et pour avoir rendu plus lisible la formulation mathématique.

La méthode diffère sensiblement de celle utilisée pour l'analyse du voisinage des jeunes d'origine étrangère, qui a été présentée lors d'un récent colloque (Aubry, Tribalat, 2008). En effet dans cette étude-ci, l'unité statistique est l'enfant de moins de 18 ans (et non pas la grappe de deux logements déterminés par la personne de référence du ménage qui l'occupe) et pour chaque enfant on calcule le nombre des enfants résidant dans les logements voisins, les uns et les autres étant repérés par les caractéristiques propres à leurs parents. Cette analyse sur les enfants est, comme on le verra, plus riche en terme d'information potentielle.

Précisions de vocabulaire

Il y a **concentration** d'un caractère sur un territoire dès lors que la distribution du caractère n'est pas uniforme. Si par exemple au plan national les cadres supérieurs représentent 10 % des effectifs et si dans la ville A ils en forment 30 %, on parlera d'une concentration élevée des cadres dans la ville A.

Le terme de concentration est donc ici employé dans le sens d'une fréquence ou d'une probabilité (mais non au sens d'un indicateur du type Gini).

L'objet de l'analyse porte sur la composition du voisinage. On cherche à créer des indicateurs mesurant « **l'entre soi** » pour exprimer, sur un territoire, la plus ou moins grande propension des individus à vivre les uns à côté des autres. Différents indicateurs sont envisageables et il existe une littérature sur cette question.

Dans notre étude on caractérisera le voisinage par deux indicateurs qui, bien que liés entre eux par une relation très simple, ont une signification différente. Ces indicateurs sont notés ES_{var} et ES_{prob} avec $ES_{prob} = (1-p)ES_{var}$ où p est la fréquence de la catégorie étudiée sur le territoire.

Le premier indicateur renvoie à la **variance** de la distribution. Il caractérise l'écart de la distribution observée par rapport à la distribution théorique. C'est en quelque sorte un indicateur adjoint à la fréquence. L'indicateur serait un peu à la fréquence ce que l'écart-type serait à la moyenne, à savoir un indicateur complémentaire à la moyenne. C'est un indicateur d'**écart** par rapport à la valeur centrale, en l'occurrence la fréquence.

On écrit $ES_{var} = \frac{\sigma_{obs}^2}{\sigma_{th}^2} - 1$ où σ^2 est la variance (respectivement de la distribution observée et de la distribution théorique).

Le second indicateur apparaît plutôt comme une **probabilité additionnelle**. Soit p la proportion du caractère étudié. Si on appelle p' la probabilité conditionnelle, c'est-à-dire, pour un logement disposant du caractère d'étude, la probabilité que le logement d'à-côté ait le même caractère, on écrit $p' = p + \Delta p$. On démontre aisément que⁽¹⁾:

$$ES_{prob} = (1 - p)ES_{var} = \Delta p$$

Un exemple fictif : une ville A

Soit un territoire donné, une ville notée A (ce pourrait être une unité urbaine, un quartier, voire même un îlot) à la date d'un recensement. Soit p la fréquence d'un caractère.

On considère toutes les grappes formées de deux logements voisins⁽²⁾. Chaque grappe peut contenir 0, 1 ou 2 logements disposant de la variable d'intérêt. La démarche consiste à étudier la distribution du nombre de logements disposant du caractère, distribution que l'on compare à la loi de probabilité correspondante.

Tableau 1. Distribution observée et distribution théorique

<i>Nombre de logements « cadres » dans la grappe</i>	<i>Effectif observé</i>	<i>Effectif théorique</i>	<i>Fréquence observée</i>	<i>Fréquence théorique</i>
0	8 300	8 100	83 %	81 %
1	1 400	1 800	14 %	18 %
2 (grappes « pleines »)	300	100	3 %	1 %
Ensemble	10 000	10 000	100 %	100 %

Mode de lecture : on forme tous les couples possibles de grappes de logements voisins (ici, 10 000 grappes de taille 2). Dans la ville A, 10 % des logements sont occupés par des cadres. Si la distribution des cadres sur le territoire de la ville était aléatoire, on devrait avoir 1 % de grappes « pleines ». Or on en dénombre 3 % (300), soit beaucoup plus que prévu : cela signifie donc qu'il y a chez les cadres une propension à vivre entre soi.

La question qui se pose est alors de mesurer l'écart entre les deux distributions. Cet écart, cette « distance », renseigne sur la propension,

(1) Cf. annexe 2.

(2) Pour qu'il y ait 10 000 grappes, il faut 10 001 logements.

voulue ou non, des personnes d'un même groupe à vivre à proximité les unes des autres.

- Premier indicateur : on fait le rapport entre la variance observée et la variance théorique. On obtient après calcul⁽³⁾ :

$$ES_{var} = \frac{\bar{w} - p^2}{p(1-p)} = 0,22$$

Dans cette formule, p est la proportion des ménages disposant de la variable d'intérêt dans le territoire d'étude (10 %), \bar{w} est la proportion de grappes pleines (3 % de grappes formées de deux logements possédant le caractère d'étude). ES_{var} s'exprime sous la forme d'un pourcentage (ici 22 %).

- Second indicateur : on remarque que \bar{w} , proportion de grappes pleines, peut être considéré comme le produit de deux probabilité (pp'), à savoir :
 - p , la probabilité de tomber sur un logement occupé par un cadre ;
 - p' , la probabilité d'avoir un voisin cadre, quand le logement de référence est lui-même celui d'un cadre (probabilité conditionnelle).

Ainsi pour reprendre l'exemple (10 % de chances pour un logement d'être occupé par un cadre), la probabilité p' pour le logement voisin d'être également occupé par un cadre n'est pas de 10 %, mais elle est beaucoup plus grande, en l'occurrence de 30 %. En effet :

$$ES_{prob} = (1-p)ES_{var} = 0,2 = \Delta p ; \text{ ainsi } p' = p + \Delta p = 0,3 \text{ ou } 30 \%$$

Propriétés des indicateurs

- a) Les valeurs prises par l'indicateur ES_{var} s'échelonnent d'une valeur minimale négative à 1. Celles de l'indicateur ES_{prob} , probabilité additionnelle, varient d'une valeur négative à 0,5 (valeur maximale).
- b) Quand la population se distribue de façon aléatoire sur le territoire, les indicateurs ES_{var} et ES_{prob} prennent la valeur 0, quel que soit p .
- c) Quand le nombre de grappes pleines est maximal, l'indicateur ES_{var} prend la valeur 1, l'indicateur ES_{prob} prend la valeur $(1-p)$. Ce serait le cas

(3) Cf. annexe 2.

si l'ensemble des membres du groupe étaient réunis sur une portion du territoire (ségrégation absolue).

d) Il peut arriver que les deux indicateurs prennent des valeurs inférieures à 0, mais c'est alors l'exception. Les limites inférieures dépendent de p (cf. annexe 2).

Tableau 2. Sept cas de figure

Nombre de logements « cadres » dans la grappe	I	II	III	IV	V	VI	VII
0	8 000	8 100	8 200	8 300	8 500	8 700	9 000
1	2 000	1 800	1 600	1 400	1 000	600	0
2	0	100	200	300	500	700	1000
ensemble	10 000	10 000	10 000	10 000	10 000	10 000	10 000
moyenne	0,1	0,1	0,1	0,1	0,1	0,1	0,1
ES_{var}	-0,11	0,00	0,11	0,22	0,44	0,67	1,00
ES_{prob}	-0,1	0,0	0,1	0,2	0,4	0,6	0,9
p'	0,0	0,1	0,2	0,3	0,5	0,7	1,0

Le tableau envisage sept cas de figure. Les sept distributions ont en commun d'avoir, par hypothèse, la même moyenne (0,1). On fait varier le nombre de grappes pleines. On vérifierait aisément que ce nombre ne peut varier par construction que de 0 (cas n° I) à 1 000 (cas n°VII). Le tableau montre comment les deux indicateurs augmentent de façon régulière, dans le même rapport 0,9 ($1-p$).

Un exemple sur trois communes, Nancy, Strasbourg et Mulhouse

Soit trois villes de l'Est. La proportion de cadres diffère : Nancy, 13,5 %, Mulhouse, 7,2 %. Entre les deux, Strasbourg avec 12,6 %.

On calcule les deux indicateurs définis ci-dessus et l'on reporte les résultats sous la forme de deux graphiques. La figure 1 met en regard pour chaque commune la fréquence et l'indicateur ES_{var} . Il apparaît que la propension à vivre entre soi est beaucoup plus grande à Mulhouse qu'à Nancy. La figure 2 fait ressortir les probabilités additionnelles ES_{prob} et il en ressort la même conclusion : les cadres vivent davantage entre eux à Mulhouse. Toutefois la figure 2 apporte quelque chose de plus, à savoir que la probabilité, pour un cadre, d'avoir près de lui un autre cadre n'est pas très différente dans les trois villes. Il ne s'agit pas de généraliser ce

cas particulier, il s'agit de montrer comment chacun des indicateurs traduit un aspect particulier du problème étudié.

Figure 1. L'effet de voisinage vu comme un indicateur adjoint à la fréquence

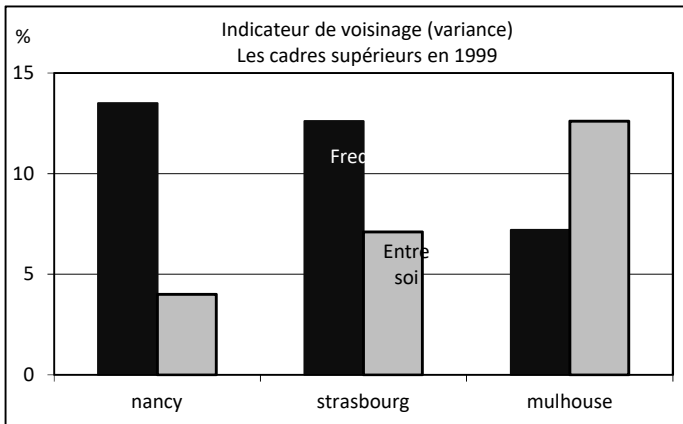
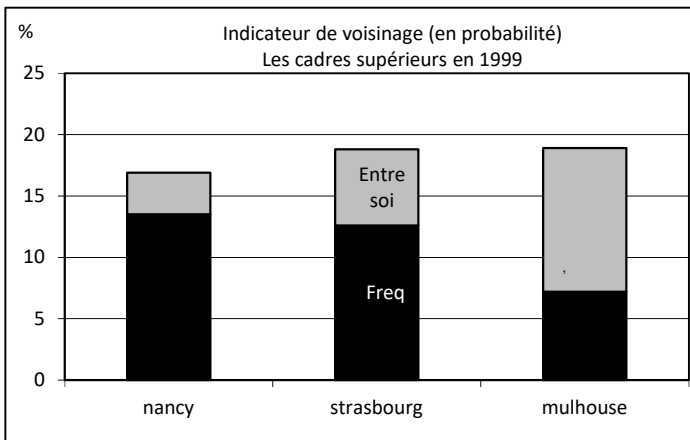


Figure 2. L'effet de voisinage vu comme une probabilité (ou fréquence) additionnelle



L'« entre soi » est représenté de deux façons. La figure 2 exprime peut-être mieux la signification du phénomène mesuré et les deux composantes sont homogènes, puisque ce sont deux probabilités. En revanche, les deux éléments de la figure 1 ne sont pas de même nature (une fréquence et un écart de variance).

Une propriété intéressante de l'indicateur ES_{var} est à noter : ce dernier est du même ordre de grandeur pour une catégorie et son complément (les

cadres et les non-cadres par exemple). Il suffit pour s'en rendre compte de se reporter à la formule et de remplacer p par $(1-p)$, \bar{w} étant alors la proportion de grappes vides (ne comprenant aucun cadre). Si les deux indicateurs ne sont pas strictement égaux, ils le sont en espérance mathématique et les différences qui apparaissent sont le fait des fluctuations aléatoires dans les distributions observées (rôle joué par la classe intermédiaire : grappes comprenant 1 cadre et un seul).

On a choisi par convention de présenter dans cette étude l'indicateur ES_{var} sous forme de tableaux et l'indicateur ES_{prob} sous forme de graphiques⁽⁴⁾.

Quelques résultats

On dispose d'une base de données communales qui couvre la période 1968-1999 pour l'ensemble du territoire de la métropole (pour Paris, Lyon et Marseille, les résultats sont disponibles par arrondissement). On a aussi la possibilité de décomposer les résultats selon la densité de l'habitat, définie comme le nombre moyen de logements par immeuble. Chaque îlot est ainsi caractérisé par un indicateur qui varie de 1 (îlot formé uniquement de maisons individuelles) à 40 (grands ensembles). Les résultats sont actuellement disponibles pour quatre catégories de population, à savoir :

- l'âge, en 6 classes ;
- le diplôme, en 5 classes ;
- les catégories socio-professionnelles, en 8 classes ;
- l'origine, en 14 classes, regroupées de différentes manières.

Il n'est pas difficile d'enrichir la base, soit en faisant appel à d'autres variables (le lieu de naissance, de résidence antérieure, etc.), soit en augmentant le nombre de modalités d'une variable déjà prise en compte (en prenant par exemple l'âge quinquennal ou la catégorie socio-professionnelle à 2 chiffres). Cependant les résultats deviendront plus fragiles au fur et à mesure que l'on procède à des décompositions puisqu'alors le nombre de grappes pleines devient souvent très faible ou

(4) Cf. annexe 1.

même nul. Seules alors des exploitations statistiques portant sur une série d'unités territoriales ont un sens.

Les chiffres en sortie des machines ont la particularité de ne pas « parler » spontanément, contrairement aux chiffres qui sortent des exploitations effectuées sur les voisins des enfants. Il n'existe pas d'étalonnage des résultats qui pourrait permettre de juger du caractère plus ou moins élevé de la propension d'une catégorie à vivre entre soi. En revanche il semble que l'on puisse donner un sens aux indicateurs de rang, qui permettent de situer telle commune par rapport à un ensemble d'autres communes. Il semble aussi que l'on puisse accorder du crédit aux évolutions temporelles. En revanche il n'est pas sûr que l'on puisse sans précaution comparer les résultats d'une ville aux indicateurs nationaux correspondants.

Pour chaque modalité étudiée, on présente deux tableaux issus d'une analyse statistique portant sur les communes de plus de 10 000 habitants au recensement de 1999 (environ 850 communes). Les variables sont la fréquence du caractère et l'indicateur ES_{var} . La médiane renseigne sur la valeur centrale, le 95^{ème} centile donne une idée des valeurs maximales rencontrées localement, l'intervalle interquartile (P75 - P25) renseigne sur la dispersion des indicateurs. L'analyse porte sur les seules communes pour lesquelles le nombre de grappes pleines est supérieur à 5, de sorte que certaines catégories qui ne sont pas représentées de façon uniforme sur le territoire (les immigrés d'origine turque) ne portent que sur un nombre de communes beaucoup plus restreint (minimum 61 pour l'origine Turquie)⁽⁵⁾.

Dans le tableau 3, les différentes catégories de population sélectionnées sont classées selon l'importance de l'« entre soi ». Cette présentation permet ainsi d'appréhender les ordres de grandeur que l'on rencontre et forme en quelque sorte un étalonnage de l'indicateur.

La valeur de la médiane la plus élevée concerne les ménages d'origine turque avec un indicateur égal à 17,5 %, devançant largement les ménages d'origine marocaine (10,9 %). C'est à peu de choses près la valeur observée pour les 20-29 ans, elle-même supérieure à celle des personnes sans diplôme. Plus au-dessous, on trouve les cadres

(5) Cette inégalité du nombre des communes prise en compte dans l'analyse conduit probablement à des biais. On peut pallier à cet inconvénient en affinant les comparaisons.

supérieurs avec 6,9 %, valeur nettement supérieure à celle des ouvriers (5,0 %). Parmi les catégories les plus proches d'une répartition aléatoire on rencontre les ménages d'âge intermédiaire (ici 40-49 ans).

Tableau 3. La gamme des valeurs rencontrées

Catégorie (origine, âge, PCS, diplôme)	Concentration (fréquence)			Entre soi (ES_{var})		
	Médiane	P95	Q3-Q1	Médiane	P95	Q3-Q1
Turquie	3,8	10,6	4,0	17,5	38,1	8,7
Maghreb	8,7	26,2	7,9	11,8	30,2	11,1
Maroc	5,1	15,6	4,6	10,9	32,8	14,0
20-29 ans	11,2	20,5	5,1	10,1	25,2	8,7
Immigrés	22,0	51,9	15,6	9,5	24,1	9,5
Non diplômés	43,9	61,6	13,7	8,1	18,7	6,9
70 ans et plus	17,3	29,9	7,0	8,0	18,1	6,4
Algérie	6,4	21,0	5,4	7,8	19,4	7,1
Bac + 4	18,9	41,8	11,4	7,4	18,5	6,3
Cadres supérieurs	10,9	26,6	8,8	6,9	16,2	5,8
Non actifs	37,1	52,3	10,7	6,6	17,9	6,4
Portugal	5,6	12,0	3,8	6,5	15,9	5,3
Europe du sud	8,2	19,5	6,3	5,3	14,7	5,7
Ouvriers	16,5	27,5	8,6	5,0	13,3	5,4
Chômeurs	8,3	14,2	3,5	4,4	11,4	4,5
Employés	11,8	17,9	4,3	4,4	16,0	6,2
Prof. interm.	14,0	20,2	4,8	2,9	9,8	4,4
40 - 49 ans	20,6	27,5	4,5	2,7	10,9	4,7
Communes-centres et communes de banlieue : immigrés						
centre	19,6	43,5	15,9	10,1	19,8	7,7
banlieue	22,5	56,0	16,4	6,0	14,4	5,7
Communes-centres et communes de banlieue : cadres supérieurs						
centre	8,6	16,4	4,9	6,0	13,5	3,9
banlieue	12,9	28,3	10,6	5,9	14,4	5,6

Champ : environ 840 unités

Les catégories ont été classées selon le niveau décroissant de la médiane de l'« entre soi ».

Il est possible de trouver des indicateurs plus élevés. La colonne P95 (seuil à 95 %, ce qui signifie qu'une commune sur 20 dépasse la valeur

indiquée) renseigne à ce sujet. Il n'est donc pas rare de trouver des valeurs supérieures à 30 % dans certaines communes et si l'on regarde vers le passé, on atteignait 50 % chez certaines catégories d'immigrés récemment arrivées sur le territoire métropolitain (tableau 4). On peut aussi rencontrer des valeurs de l'ordre de 40 % dans la catégorie « policiers et militaires » (PCS=53).

Tableau 4. Quelques catégories importantes - Evolution 1968-1999

Catégorie (origine, âge, PCS)	Concentration (fréquence)			Entre soi (ESvar)		
	Médiane	P95	Q3-Q1	Médiane	P95	Q3-Q1
20 - 29 ans						
1968	13	21	5	7	18	7
1975	18	28	6	9	22	8
1982	16	25	6	10	20	8
1990	14	21	5	9	20	7
1999	11	21	5	10	25	9
70 ans et plus						
1968	13	20	5	6	17	6
1975	14	23	6	9	21	7
1982	14	24	7	10	25	8
1990	15	28	7	10	22	7
1999	17	30	7	8	18	6
Cadres sup.						
1968	9	21	6	10	23	8
1975	10	26	7	10	23	10
1982	10	26	8	9	24	8
1990	11	27	9	9	20	7
1999	11	27	9	7	16	6
Portugal						
1968	3	14	6	23	54	13
1975	6	16	6	13	29	8
1982	6	15	5	10	26	7
1990	6	14	4	8	19	5
1999	6	12	4	7	16	5
Orig. Algérie						
1968	4	13	4	28	48	19
1975	5	19	4	21	39	14
1982	5	19	4	15	33	12
1990	5	19	4	11	27	10
1999	6	21	5	8	19	7

Champ : environ 840 communes

Comme tenu de la contrainte des seuils, le nombre de communes n'est pas identique pour chaque catégorie (et notamment d'un recensement à l'autre).

A travers le tableau 4, on a surtout voulu montrer non seulement la variété des situations rencontrées en matière d'évolution de l'« entre

soi », mais aussi une certaine régularité des évolutions dans le temps pour une même catégorie. Les à-coups qui pourraient *a priori* rendre suspecte la validité des résultats sont rares.

Chez les plus jeunes, à savoir les ménages de 20 à 29 ans, l'augmentation de l'indicateur est très sensible. Il semble que la plupart des villes soient touchées, mais c'est surtout dans les villes universitaires que la poussée est la plus forte. Il n'est rien de semblable chez les ménages les plus âgés (70 ans et plus) alors même que l'hypothèse d'une ségrégation de plus en plus forte envers les personnes âgées n'était pas exclue. Ce n'est manifestement pas le cas, puisque l'indicateur, après être passé par un maximum de 10,2 % au recensement de 1982, est redescendu à 8,0 % au recensement de 1999.

L'indicateur est également à la baisse chez les cadres supérieurs, mais la diminution, assez brutale et récente, rompt avec une stabilité remarquable qui avait caractérisé les vingt années précédentes.

Chez les immigrés, la tendance à la diminution de l'« entre soi » est évidente. Elle l'est certes dans la population originaire de l'Europe du sud, mais elle l'est aussi dans la population originaire du Maghreb. Elle est même spectaculaire chez les ménages originaires d'Algérie puisqu'on passe de 27,9 % en 1968 à 7,8 % en 1999.

Limites et perspectives

Des contraintes sur les effectifs

La formule rend l'indicateur directement dépendant du nombre de grappes pleines et du carré de la fréquence du caractère observé. Si l'on veut que la crédibilité de l'indicateur ne soit pas affaiblie par des aléas trop importants, il faut donc qu'il y ait suffisamment de grappes disposant du caractère étudié.

Soit une commune ayant 100 000 habitants. Le nombre de logements dans le fichier au quart⁽⁶⁾ est donc d'environ 10 000. S'il y a 10 % de cadres, cela conduit à attendre une centaine de grappes pleines. Les lois de la statistique sont alors en mesure de fonctionner normalement. Il

(6) On peut aussi faire des travaux sur l'entre soi à partir des exploitations exhaustives (dites principales), mais les variables disponibles sont moins nombreuses que pour les exploitations par sondage (ou complémentaires).

n'en est plus de même si l'on considère une minorité formant 1 % de la population (les Turcs par exemple). On ne pourra attendre au mieux que quelques grappes pleines (probabilité de 1 pour 10 000). L'indicateur n'aura donc plus aucune fiabilité.

Il faut donc être vigilant dans l'analyse des résultats, notamment quand on concentre son attention sur une ville donnée. En revanche on ne doit pas craindre des résultats qui portent sur des ensembles plus importants.

Les ruptures territoriales

L'hypothèse implicite est que deux logements voisins dans le fichier informatique sont également voisins sur le terrain. C'est sûrement vrai, en moyenne. Les logements sont repérés par un numéro d'îlot, un numéro d'immeuble et un numéro de logement. En principe, des logements ayant le même numéro d'îlot/immeuble et un numéro voisin de logement sont proches l'un de l'autre.

Cette hypothèse ne se vérifie pas absolument. De toute évidence il existe des ruptures dues au fait qu'une même grappe comprend deux logements parfois éloignés l'un de l'autre. Une façon de pallier à cet inconvénient consiste à supprimer du champ de l'étude les grappes qui appartiennent à deux îlots, voire même pour certaines investigations spécifiques à deux immeubles différents.

Tous les résultats figurant dans la base correspondent à des grappes de logements appartenant à un même îlot.

Notons que cette critique ne remet pas en cause la méthode. Par analogie, observons que chaque fois que l'on réalise des travaux statistiques qui font référence à la distance entre communes, on est confronté à des critiques du même type : si l'on prend la distance à vol d'oiseau comme on le fait généralement, celle-ci sous-estime la réalité, et cela de façon variable puisque la morphologie du terrain intervient dans la distorsion d'avec la réalité. Du reste, la « distance » entre communes dépend de nombreux facteurs autres que ceux de la distance physique (accessibilité, temps de parcours, etc.).

Il ne faut donc pas surestimer les conséquences de ce handicap. C'est le lot commun à toute statistique que de s'appuyer sur des conventions. Il reste que les contraintes de la collecte ont conduit à créer des partitions

communales qui sont de taille toujours plus ou moins équivalente d'un recensement à l'autre.

Quoi qu'il en soit, par prudence, on devra vérifier que les différences entre les résultats selon le choix des limites ne changent pas les conclusions que l'on peut tirer de l'analyse.

Le pas de tirage

Les fichiers des recensements utilisés pour le calcul des indicateurs d'« entre soi » sont ceux des exploitations par sondage (ou exploitations complémentaires). On souhaiterait pouvoir travailler sur les données des exploitations principales (exploitations exhaustives), mais en ce cas la possibilité d'obtenir des séries historiques est bien moindre. Elles sont limitées à des variables basiques telles que l'âge, le diplôme, la nationalité (en deux postes, Français ou non). On ne peut rien faire sur la catégorie socio-professionnelle ni sur l'origine géographique des immigrés.

Si on se contente d'observations portant sur les deux derniers recensements (ceux de 1990 et de 1999), le choix offert est cependant plus vaste.

L'utilisation de fichiers par sondage présente au moins deux propriétés, qui peuvent être des inconvénients. Tout d'abord, le voisin n'est pas exactement le ménage d'à-côté (L+1), mais le ménage situé à (L+4), ce qui n'est pas bien grave quand on limite les investigations à des zones où les immeubles collectifs sont nombreux, mais qui peut poser des problèmes plus sérieux localement. En second lieu, le fait que l'on ne puisse retenir qu'un logement sur quatre crée une incertitude aléatoire qui n'est pas sans nuire à la fiabilité des résultats.

On a cherché à mettre en évidence l'effet « pas » en traitant, sur la région Alsace, le lien entre la valeur des indicateurs d'« entre soi » et le pas de tirage, c'est-à-dire la proximité entre deux logements successifs. On admet que dans le cas d'un tirage exhaustif, la proximité sur le terrain de deux logements conduit par construction à la proximité dans le fichier.

L'exhaustif sert de référence. On calcule l'indicateur en augmentant le pas de tirage. D'abord deux logements de suite (pas de 1), puis un logement sur deux, et ainsi de suite jusqu'à un pas de 8. Plus le pas est

grand, donc plus la distance est grande entre les logements, et plus l'indicateur est faible, ce qui paraît assez naturel. Mais s'il y a de fortes différences entre le pas 2 et le pas 1, puis entre le pas 3 et le pas 2, il semble bien qu'au-delà l'indicateur ne soit plus très sensible au pas (tableau 5).

Tableau 5. L'effet du pas de tirage dans le cas de l'âge (indicateur ES_{var})

Pas de tirage	20 - 29 ans		70 ans et plus	
	Pas de tirage (exhaustif)	Sondage au 1/4	Pas de tirage (exhaustif)	Sondage au 1/4
1	21,5		12,6	
2	19,0		11,3	
3	17,8		10,1	
4	17,5	17,3	9,9	8,0
5	16,7		9,5	
6	15,9		9,2	
7	15,5		9,1	
8	15,4	17,5	8,6	8,0

L'observation porte sur les seules données de l'Alsace.

Les pas 4 et 8 à l'exhaustif sont confrontés au pas (L+4) et (L+8) du sondage. On observe entre les deux exploitations des différences qui ne sont pas négligeables⁽⁷⁾, mais elles s'expliquent sans doute par le seul effet des fluctuations aléatoires. Il serait donc souhaitable que ces travaux soient poursuivis.

Les indicateurs inférieurs à 0

Il peut arriver que l'indicateur prenne des valeurs inférieures à 0. Il y a deux raisons à cela. D'une part il peut s'agir de l'effet des fluctuations aléatoires dont on sait qu'elles peuvent être importantes dès lors que la taille du territoire d'observation est un tant soit peu modeste. D'autre part, un indicateur négatif peut provenir d'une propension chez certaines catégories de population à prendre des distances par rapport à leurs semblables, une forme de « ségrégation négative ». Ce serait le cas

(7) Rappelons que l'on parle d'un sondage au 1/4, alors que ce n'est qu'une image. Il est plus correct, au moins au recensement de 1999, de parler d'exploitation complémentaire. En réalité, une partie du territoire, certes petite, a été traitée dans le complémentaire de façon exhaustive (ces cas particuliers ont été redressés dans les exploitations faites dans cette étude).

si les personnes d'un même groupe en quelque sorte s'évitaient, se fuyaient mutuellement. Il va de soi que pour pouvoir valider un tel comportement, il faudrait s'assurer que le phénomène se répète un grand nombre de fois.

L'effet de l'espace géographique

L'interprétation des résultats doit tenir compte de l'espace sur lequel les indicateurs ont été calculés. Prenons un cas limite. Un pays est fait d'une région (A), où les cadres sont très nombreux, et du reste du territoire (B) où les cadres sont peu nombreux. Faisons l'hypothèse qu'en A et B les cadres se distribuent de façon parfaitement aléatoire. Au niveau national l'indicateur calculé ne sera pas nul. En effet l'indicateur se réfère à la fréquence observée dans le pays et au nombre de grappes pleines. Or ces grappes pleines, très nombreuses dans la région A, contribueront à pousser l'indicateur vers le haut.

Cet inconvénient peut rendre délicat l'interprétation des résultats, à l'instar de l'effet de structure qui perturbe parfois les discours que l'on tient par exemple sur des écarts de salaires entre villes ou des taux de mortalité entre départements. On pourrait envisager des moyens de contourner cet inconvénient en corrigeant les résultats et ce en les comparant « toutes choses égales par ailleurs ».

Pour aller plus loin

Ces différentes remarques conduisent à suggérer des analyses complémentaires. Ainsi à l'intérieur des grandes communes, on pourrait suivre des indicateurs par quartiers. Quand des découpages homogènes existent, par exemple à Strasbourg⁽⁸⁾, des séries historiques sur l'« entre soi » devraient avoir un sens. Il faudrait aussi développer des travaux sur la sensibilité des résultats, ne serait-ce que pour valider ou non les chiffres obtenus pour les villes. On pourrait aussi envisager d'améliorer la connaissance en travaillant non plus sur des grappes de 2 logements, mais de 3, voire de 4 logements ou plus, ne serait-ce que pour examiner de plus près les problèmes locaux d'hyper-concentration ou de ségrégation de population (ce que ne permettent pas les grappes de deux logements). Par ailleurs, au lieu de comparer des communes entre

(8) Cf. dans ce même volume (pp. 115-148) le travail de Bénédicte Gérard sur les grands ensembles strasbourgeois.

elles, il serait parfois plus judicieux de comparer entre eux des ensembles de communes, dès lors que la politique de l'habitat est partagée sur un territoire multi-communal.

On sait que des démarches nouvelles pour étudier le voisinage sont rendues possibles grâce à la généralisation des SIG (systèmes d'information géographique). Des travaux de nature différente sont d'ores et déjà engagés qui permettent d'observer les questions de voisinage. Actuellement ces méthodes ne s'appliquent pas aux recensements de la population en perspective historique du fait que les informations ne sont pas disponibles à l'adresse.

Enfin, serait-il possible d'aller plus loin et de dégager des indicateurs synthétiques qui pourraient mesurer, en combinant les différents indicateurs, des marqueurs de cohésion sociale qui permettraient de suivre sur plusieurs décennies les caractéristiques des villes et de faire de ces indicateurs un instrument d'évaluation des politiques de cohésion sociale (Putman, 2007) ?

Les données : un fichier historique des recensements (Saphir)

L'étude s'appuie sur une exploitation spécifique du fichier SAPHIR (Système d'analyse de la population par l'historique des recensements), fichier mis au point par l'INSEE-Alsace. La version utilisée contient les données brutes et harmonisées pour les cinq recensements classiques couvrant la période 1968 à 1999.

Le traitement spécifique élaboré pour cette étude a conduit à produire trois fichiers thématiques (tableau 6) ayant chacun une unité statistique de base différente. Ces unités statistiques ont en commun un code d'identification qui permet de les enrichir facilement si besoin est. Ce code d'identification appelé NUM⁽⁹⁾ est la concaténation de trois éléments : l'année du recensement, la commune de résidence et le numéro d'ordre du logement dans la commune (numéro attribué dans l'ordre où sont classés les logements dans le fichier informatique).

(9) NUF pour le numéro de famille (composé de NUM et de NF, numéro d'ordre de la famille dans le ménage).

Tableau 6. Variables actuellement disponibles dans le fichier Saphir
(période 1968-1999)

Variables	Fichier Logement	Fichier Individu	Fichier Famille		
			Homme	Femme	Enfant
Identification	NUM	NUF		NUF	
Rang dans le logement		RGM	RGH	RGF	RGE
Taille du ménage	TMEN	TMEN			
Taille de la famille		TFAM			
Lien avec le chef de ménage		ZLCM			
Lien avec le chef de famille		ZLCF			
Catégorie de population		CPX			
Sexe		S			
Age	AGM	AGE	AH	AF	AE
État matrimonial		M	MH	MF	
Cohabitation		COHA			
Diplôme	DPM	DPX			
Catégorie socio-prof.	CSM	CSX	CSH	CSF	CSE
Statut	STM	STX	STH	STF	
Type d'activité		TAX			
Origine	ORIM	ORIG	ORIH	ORIF ORIZ	ORIE
Année d'arrivée en métropole	ARRIM	ARRIV			
Nationalité	NAM	NAX	NAH	NAF	NAE
Pays de naissance	PNM	PN2	PNH	PNF	PNE
Département de naissance	DNM	DN2			
Pays de résid. ant.	PRAM	PRA2			
Département de résid. ant.	DRAM	DRA2			
Commune de résid. ant.	CRAM	CRA2			
Type de logement	TLX				
Nombre de log. dans l'immeuble	NLX				
Nombre d'étages de l'immeuble	ETX				
Année d'achèvement	AAX				
Nombre de pièces	NPX				
Statut d'occupation	SOX				
HLM	HLMX				
Voiture	VX				
Nombre de logements par immeuble	DS1				
Nombre d'îlots dans la commune	NKK1				
Nombre de logements dans l'îlot	NKK2				

Année d'arrivée en métropole : disponible pour les seules années 1968 et 1999

On dispose donc de trois fichiers de travail :

- un fichier logement (WMEN)
- un fichier individu (WIND)
- un fichier famille (WFAM)

Le contenu de ce tableau peut être aisément modifié par introduction de nouvelles variables, notamment de variables reconstruites. Ainsi le code ORIZ (origine de la famille) est obtenu à partir de la nationalité et du pays de naissance de l'homme et de la femme.

Enfin, précisons qu'à côté du numéro d'ordre NUM/NUF, il existe d'autres codes d'identification. Tout d'abord le code figurant dans le fichier Saphir original (IDMENX/IDFAMX) dont NUM/NUF est une variante. Par ailleurs NK1 est le numéro d'ordre de l'îlot dans la commune (NKK1 est le nombre d'îlots dans la commune) et NK2 est le numéro d'ordre du logement dans l'îlot (NKK2, nombre de logements dans l'îlot). Enfin on dispose pour chaque logement des variables brutes (codes îlot, numéro d'immeuble et numéro de logement dans l'immeuble).

Références bibliographiques

AUBRY Bernard, TRIBALAT Michèle, 2008, *Les voisins des jeunes d'origine étrangère*, Colloque de l'AIDELF « Démographie et cultures », Québec, 25-29 août 2008.

AUBRY Bernard, TRIBALAT Michèle, 2007, *Importance et évolution des concentrations en zone gendarmerie de 1968 à 1999*, Centre de Prospective de la Gendarmerie Nationale.

LE TOQUEUX Jean-Luc, 2007, « Ségrégation ou mixité socio-spatiale : De quoi parle-t-on ? », in : MATTEI Marie-Flore, PUMAIN Denise, *Données urbaines*, n° 5, Economica, Anthropos, coll. Villes.

MAURIN Eric, 2004, *Le ghetto français, enquête sur le séparatisme social*, Le Seuil-La république des idées.

PUTMAN Robert D., 2007, « E Pluribus Unum : Diversity and Community in the Twenty-first Century », *Scandinavian Political Studies*, Vol. 30, N° 2, 137-174.

Annexe 1. Graphiques

Les possibilités de représentation graphique à partir des données obtenues au niveau communal sont très nombreuses. On a donc procédé à une sélection permettant de comparer deux catégories distinctes, mais dont l'importance mesurée par la fréquence du caractère au plan national est de même ordre de grandeur. Il s'agit des cadres supérieurs et des immigrés. Pour chaque ensemble d'unités territoriales retenues, on présente donc deux graphiques.

Les confrontations ont été faites au recensement de 1999 pour différents territoires. Pour les trois grandes villes que sont Paris, Lyon et Marseille, les données sont produites au niveau de l'arrondissement. La diversité des situations en banlieue parisienne est illustrée par les résultats portant sur les communes de plus de 30 000 habitants de trois départements de l'Ile-de-France, à savoir les Hauts-de-Seine (92), la Seine-Saint-Denis (93) et le Val-de-Marne (94). Enfin, des graphiques font apparaître les diversités inter et intra-régionales pour les communes de plus de 50 000 habitants. Le territoire métropolitain hors Ile-de-France a ainsi été divisé en grandes régions repérées par le premier chiffre du code régional (ex : 41, 42, 43 pour les trois régions de l'Est : la Lorraine, l'Alsace et la Franche-Comté). Ces sept grandes régions ont pour nom : le Bassin parisien (2), le Nord (3), l'Est (4), l'Ouest (5), le Sud-ouest (7), le Centre-est (8) et la Méditerranée (9).

a) *Les arrondissements*

Les arrondissements sont présentés selon le rang décroissant de la fréquence du caractère. Les deux classements se font manifestement dans un ordre inversé : les arrondissements classés en tête dans une catégorie sont à la queue dans l'autre.

- *Les cadres supérieurs* (figure a1). Des trois villes, c'est celle de Lyon qui fait apparaître les plus grands contrastes, tant en terme de taux (de 9 % à 25 %) qu'en terme d'« entre soi » (de 3 % à 15 % en probabilité additionnelle). A Paris, les taux sont plus proches les uns des autres (sauf dans trois arrondissements où les cadres sont relativement peu nombreux et cependant marqués par l'« entre soi »). Dans trois arrondissements, le 1^{er}, le 5^{ème} et le 6^{ème}, la distribution des cadres est proche d'une distribution aléatoire. A Marseille où la proportion de cadres est

assez faible, l'« entre soi » est très élevé dans le 11^{ème}. Lyon se caractérise par une très grande variabilité non seulement s'agissant des taux, mais aussi des indicateurs, notamment là où les cadres sont peu nombreux.

- *Les immigrés* (figure a2). On retrouve un peu la même physionomie des graphiques et la même hiérarchie entre les trois communes, quoique inversée. A Paris, cependant, l'indicateur d'« entre soi » est également fort là où il était élevé chez les cadres supérieurs.

b) *Trois départements de l'Île-de-France*

Ces trois départements ont été choisis car ils contiennent la majorité des communes de banlieue dont la population est supérieure au seuil retenu (30 000 habitants).

- *Les cadres supérieurs* (figure b1). Si la Seine-Saint-Denis est relativement homogène (faible pourcentage de cadres dans la plupart des communes), en revanche les deux autres départements sont partagés entre communes à forte proportion de cadres et les autres. Les communes qui se caractérisent par un indicateur élevé d'« entre soi » sont Chatenay-Malabry et Nanterre dans le 92, Pantin, Drancy et Sevran dans le 93 et Maisons-Alfort dans le 94.
- *Les immigrés* (figure b2). La proportion d'immigrés est en général forte, parfois très forte, avec des différences dans les taux très grandes à l'intérieur de chaque département. L'« entre soi » est le plus élevé à Issy-les-Moulineaux (92), commune à faible taux d'immigrés et à Aulnay-sous-Bois (94). Parfois il arrive que l'indicateur prenne une valeur proche de zéro. C'est le cas pour la commune de Neuilly-sur-Seine (92) et pour celle de Bobigny (93).

c) *Les villes de province*

Le classement par grande région fait ressortir des physionomies de courbes très intéressantes.

- *Les cadres supérieurs* (figure c1). Les différences sont les plus frappantes dans les villes du Nord ; l'« entre soi » y est partout important. Les communes de Roubaix et de Tourcoing ont de

loin les taux de cadres les plus faibles alors que les quatre autres villes ont des taux plutôt supérieurs à la moyenne. Toujours pour ce qui concerne les taux, des écarts importants apparaissent au sein des différentes régions du sud : Pessac, Grenoble, Aix-en-Provence ont des résultats nettement supérieurs aux autres villes. L'« entre soi » semble presque partout très marqué, à l'exception de nombre de villes du sud. Ainsi, Arles et Hyères seraient proches d'une distribution aléatoire des cadres sur le territoire de leur commune.

- *Les immigrés* (figure c2). Les contrastes territoriaux sont extrêmement forts : entre le Nord et l'Ouest, mais aussi entre le Centre-est et la Méditerranée. Entre villes d'une même région, il peut l'être également : c'est ce que l'on observe dans le Nord, mais aussi dans le Centre-est. L'« entre soi » est toujours fort dans les villes du Bassin parisien. Il l'est moins dans l'Ouest (exception notable pour Cholet). Dans le Sud, les contrastes sont marqués, ainsi entre des communes comme Avignon, Fréjus, Nîmes, où l'« entre soi » est sensible et d'autres, comme Antibes ou Hyères, où il est faible, parmi les plus bas du territoire métropolitain.

Figure a1. L'« entre soi » à Marseille, Lyon et Paris en 1999 : les cadres supérieurs

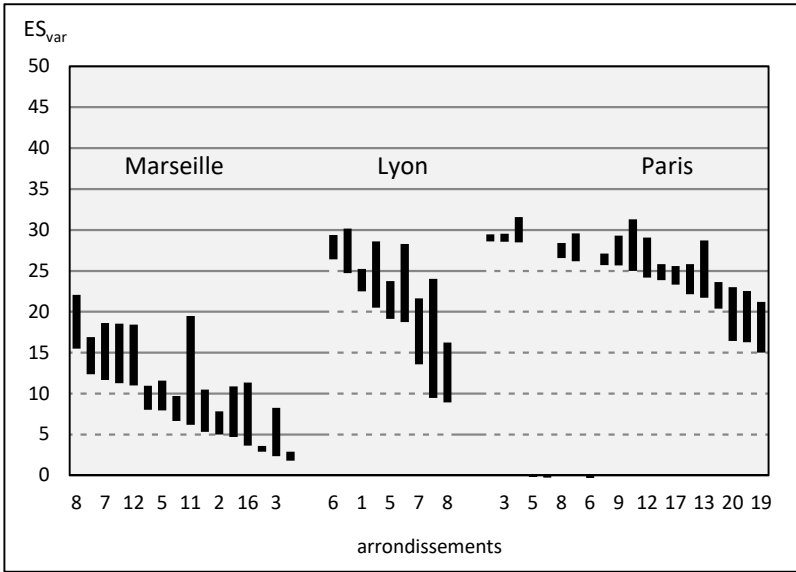


Figure a2. L'« entre soi » à Marseille, Lyon et Paris en 1999: les immigrés

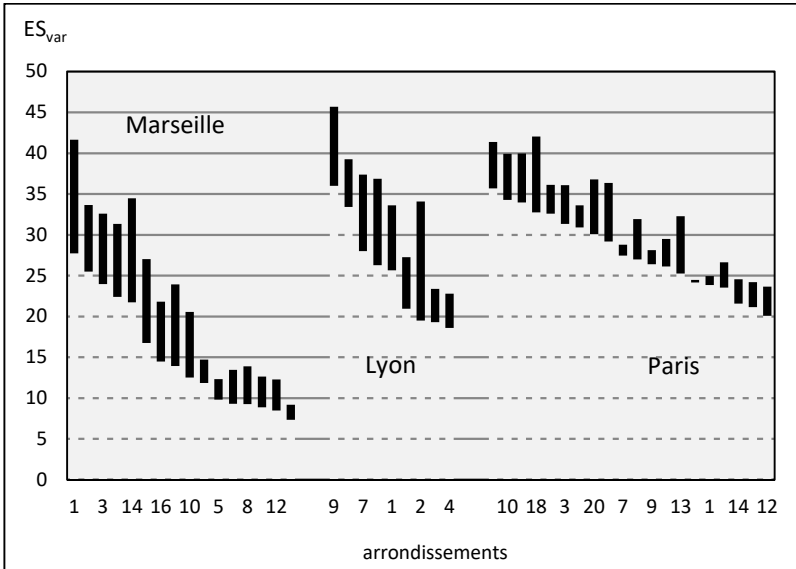


Figure b1. L'« entre soi » dans les départements 92, 93 et 94 :
les cadres supérieurs

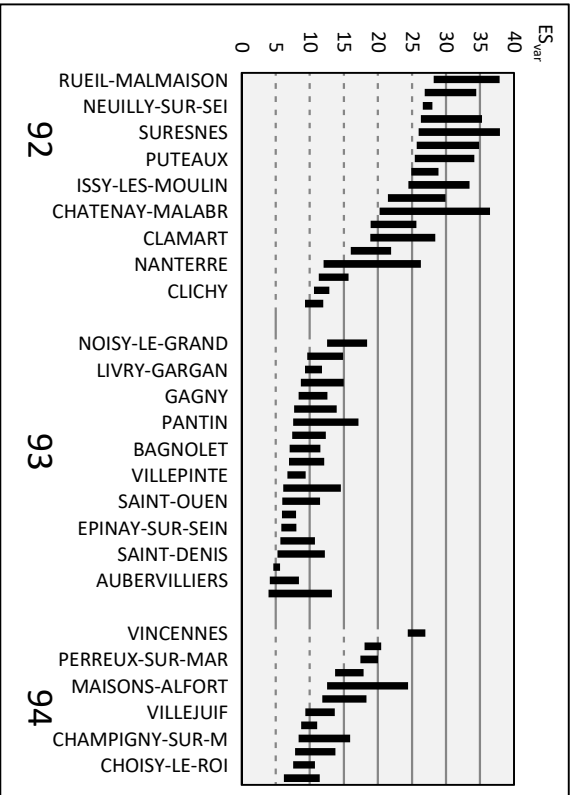


Figure b2. L'« entre soi » dans les départements 92, 93 et 94 : les immigrés

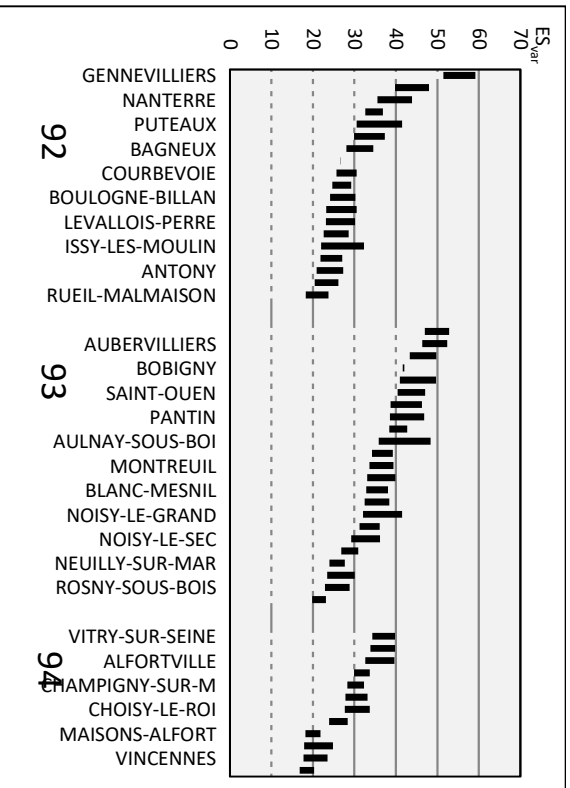


Figure c1. L'« entre soi » dans les villes de province : les cadres supérieurs

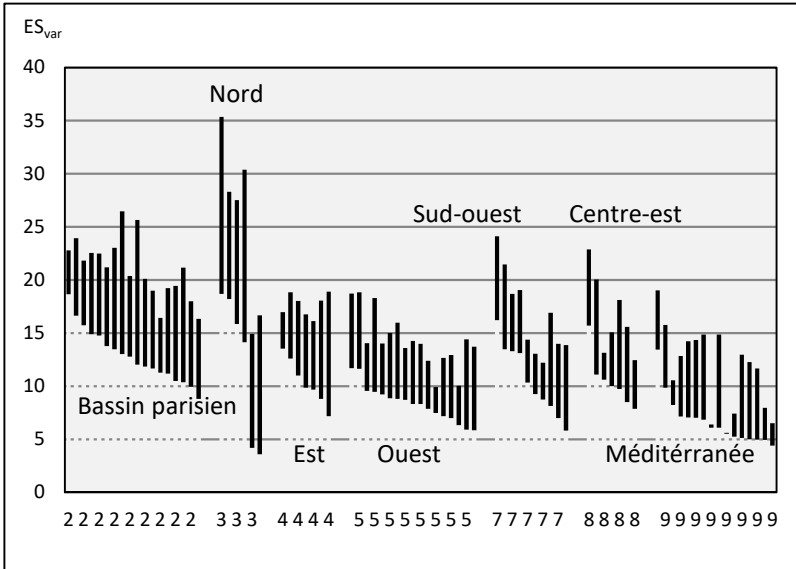
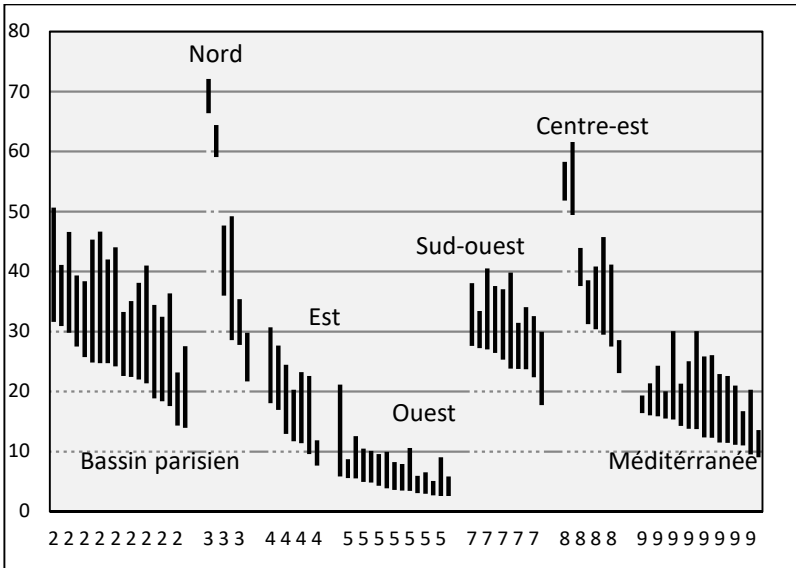


Figure c2. L'« entre soi » dans les villes de province : les immigrés



Annexe 2. Les grappes de taille 2 : Formulation mathématique

Soit N le nombre de grappes de taille 2 formées de deux logements voisins. On note p la proportion de logements disposant du caractère. Un premier tableau (tableau a1) conduit à la moyenne de la distribution : nombre moyen de logements disposant du caractère.

La première partie conduit au calcul des deux indicateurs d'« entre soi » utilisés dans le texte. Ces calculs sont approximatifs, dans la mesure où une moyenne calculée sur un ensemble de N grappes n'est pas, en toute rigueur, égal à une moyenne calculée sur $N+1$ logements nécessaires à la détermination des N grappes (en raison des extrémités). Le calcul rigoureux de la variance théorique de la loi de probabilité est donné en seconde partie. Ce calcul a été effectué grâce au concours de René Padiou, inspecteur général de l'INSEE.

Tableau a1. Grappes de taille 2.
Calcul de la moyenne de la distribution

<i>Variable d'intérêt</i> x_i	<i>Effectif</i> n_i	<i>Contribution à la moyenne</i> $n_i x_i$
0	n_0	0
1	n_1	n_1
2	n_2	$2n_2$
Ensemble	$n_0+n_1+n_2 = N$	$n_1+2n_2 = 2 Np$

Il ressort de l'examen de ce tableau que l'on peut poser un système de deux équations :

$$n_0+n_1+n_2=N$$

$$n_1+2n_2=2Np$$

La première équation exprime une contrainte sur les effectifs, la seconde sur la moyenne. La connaissance d'un seul effectif détermine la connaissance des deux autres. On fixe par convention n_2 , que l'on notera dans la suite ω et que l'on appellera nombre de « grappes pleines ».

Le tableau a2 donne les éléments du calcul de la variance de la distribution en fonction du nombre des grappes pleines.

Tableau a2. Grappes de tailles 2. Calcul de la variance (pour ω fixé)

Nombre de logements dans la grappe x_i	Effectif n_i	Contribution à la moyenne $n_i x_i$	Ecart quadratique $e_i = (x_i - 2p)^2$	Contribution à la variance $p_i e_i^2$
0	$N - 2pN + \omega$	0	$4p^2$	$4p^2(N - 2pN + \omega)$
1	$2(pN - \omega)$	$2(Np - \omega)$	$(1 - 2p)^2$	$2(pN - \omega)(1 - 2p)^2$
2	ω	2ω	$4(1 - p)^2$	$4\omega(1 - p)^2$
Ensemble	N	$2Np$		$-4Np^2 + 2Np + 2\omega$

Il ressort des calculs que la variance, conformément à la formule :

$$\sigma^2 = \frac{\sum_{i=1}^N n_i (x_i - \bar{x})^2}{N}$$

s'exprime sous la forme :

$$\sigma_{obs} = \frac{-4Np^2 + 2Np + 2\omega}{N}$$

De son côté, la variance théorique σ^{2th} est celle de la loi binomiale classique, à savoir $2pq$ (les grappes sont de taille 2).

Les indicateurs d'« entre soi »

Un premier indicateur, noté ES_{var} , est défini comme le rapport entre les deux variances $\frac{\sigma_{obs}^2}{\sigma_{th}^2} - 1$. Au terme d'un calcul, on aboutit au résultat

suisant :
$$ES_{var} = \frac{\bar{\omega} - p^2}{p(1-p)}$$

- Cas particuliers (application numérique pour $p = 0,1$) :
 - Aucune grappe n'est pleine ($\omega = 0$) : $ES_{var} = \frac{p}{p-1}$, soit $ES_{var} = -0,11$.
 - Le nombre de grappes pleines est maximal ($\bar{\omega} = 2p^2$), soit $ES_{var} = 1$.

Un second indicateur apparaît comme une probabilité additionnelle. On considère $\bar{\omega}$ comme le produit de deux probabilités : $\bar{\omega} = pp'$. Dans

cette expression, p' est une probabilité **conditionnelle**, à savoir la probabilité qu'un ménage disposant du caractère d'étude ait pour voisin un autre ménage disposant également du même caractère.

$$\text{En ce cas on écrit } ES_{var} = \frac{\overline{w} - p^2}{p(1-p)} = \frac{pp' - p^2}{p(1-p)} = \frac{p' - p}{1-p}$$

$$\text{En posant } ES_{prob} = \Delta p = p' - p$$

$$\text{On parvient à la relation : } \boxed{ES_{prob} = (1-p)ES_{var}}$$

Calcul rigoureux de la variance théorique

a) Hypothèses et notations

- p : probabilité pour un logement de disposer de la variable d'intérêt ;
- N : nombre de grappes de taille 2 (soit $N+1$ logements) ;
- $x_i = u_i + u_{i+1}$: nombre de logements disposant de la variable d'intérêt (0, 1 ou 2) ;
- On fait l'hypothèse d'indépendance des logements :
 $E(u'_i u'_j) = 0$ si $i \neq j$.

b) Propriétés de la variable u (loi binomiale)

$$E(u) = p, \quad E(u') = 0 \text{ si } u' \text{ est la variable centrée : } u'_i = u_i - p$$

$$V(u) = E(u'^2) = \sigma^2 = p(1-p)$$

c) Propriétés de la variable x (loi binomiale)

$$E(x) = 2p, \quad E(x') = 0 \text{ si } x' \text{ est la variable centrée : } x'_i = x_i - 2p$$

$$V(x) = 2\sigma^2 = 2p(1-p), \quad E(u') = (u - p) = 0$$

(propriété d'additivité de la variance du fait de l'indépendance)

d) Calcul de la variance à partir de la formule de la définition

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$$

- *Premier cas : N grappes*

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$$

$$V = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 = \frac{1}{N} \sum_i \left[u'_i + u'_{i+1} - \frac{1}{N} \sum_{j=1}^N u'_j - \frac{1}{N} \sum_{k=2}^{N+1} u'_k \right]^2$$

La partie entre crochets est un polynôme de 4 termes (a+b+c+d). Elevé au carré, il comprend dix termes :

$$(a^2 + b^2 + c^2 + d^2 + 2ab + 2ac + 2ad + 2bc + 2bd + 2cd)$$

$V = \frac{1}{N} \sum_{i=1}^N (a^2 + b^2 + c^2 + d^2 + 2ab + 2ac + 2ad + 2bc + 2bd + 2cd)$		
$\sum_{i=1}^N u_i^2$		
$\sum_{i=1}^N u_{i+1}^2$	$2 \sum_{i=1}^N u'_i u'_{i+1}$	
$\frac{1}{N} \left[\sum_{j=1}^N u'_j \right]^2$	$-\frac{2}{N} \sum_{i=1}^N u'_i \sum_{j=1}^N u'_j$	$-\frac{2}{N} \sum_{i=1}^N u'_{i+1} \sum_{j=1}^N u'_j$
$\frac{1}{N} \left[\sum_{k=2}^{N+1} u'_k \right]^2$	$-\frac{2}{N} \sum_{i=1}^N u'_i \sum_{k=2}^{N+1} u'_k$	$-\frac{2}{N} \sum_{i=1}^N u'_{i+1} \sum_{k=2}^{N+1} u'_k$
		$\frac{2}{N} \sum_{j=1}^N u'_j \sum_{k=2}^{N+1} u'_k$

V est une variable aléatoire. On s'intéresse à son espérance mathématique. Comme on a fait l'hypothèse de l'indépendance, les éléments s'écrivent aussi :

$N\sigma^2$		
$N\sigma^2$	0	
σ^2	$-2\sigma^2$	$-2\frac{N-1}{N}\sigma^2$
σ^2	$-2\frac{N-1}{N}\sigma^2$	$-2\sigma^2$
		$2\frac{N-1}{N}\sigma^2$

Soit après simplification :

$$V = 2\sigma^2 \left[1 - \frac{2}{N} + \frac{1}{N^2} \right]$$

- *Deuxième cas : le nombre de grappes est infini ($\bar{x} = 2$)*

Variance de la variable aléatoire x , ayant pour espérance $2p$ (moment de premier ordre par rapport à p) :

$$V^* = \frac{1}{N} \sum_{i=1}^N (x_i - 2p)^2 = \frac{1}{N} \sum_i (u'_i + u_{i+1})^2 = \frac{1}{N} \sum_{j=1}^N (u_i'^2 + u_{i+1}^2 + 2u'_i u'_{i+1})$$

$$E V^* = \frac{1}{N} (N\sigma^2 + N\sigma^2 + 0) = 2p(1 - p)$$

Bernard AUBRY

Le voisinage : proposition d'indicateurs

Les travaux réalisés à ce jour sur le voisinage sont peu nombreux, du moins en Europe. On s'intéresse presque toujours à la connaissance de la population à travers les individus qui la composent, les familles ou les ménages, très rarement à travers celle des voisins. Or la propension qu'ont souvent les personnes appartenant à un même groupe social à se rapprocher, quelle qu'en soit la cause, n'est pas sans rapport avec la notion de cohésion sociale qui renvoie à celle d'une distribution harmonieuse sur le territoire de la population dans ses différentes composantes. L'un des objectifs d'une bonne gouvernance est d'éviter les ségrégations spatiales qui peuvent conduire parfois au rejet dans certains quartiers d'une population marginale.

C'est l'objet de cette communication que d'examiner, à partir d'une analyse spécifique des fichiers des recensements de la population, la composition du voisinage par un examen des propriétés statistiques des grappes formées de deux logements proches l'un de l'autre. La question peut être ainsi posée : en moyenne un ménage dont la personne de référence est un cadre (respectivement, un immigré, une personne âgée, etc.) a-t-il une probabilité plus ou moins forte d'avoir autour de lui un autre ménage appartenant à sa propre catégorie ?

Ce travail est l'un des volets d'une série d'investigations effectuées en collaboration avec Michèle Tribalat, directrice de recherches à l'INED, investigations qui s'appuient sur des exploitations originales du fichier historique des recensements (Saphir).