

METHODE POUR LE CROISEMENT DE VARIABLES QUALITATIVES DANS LES ENQUETES A PLUSIEURS NIVEAUX

PIRON M.

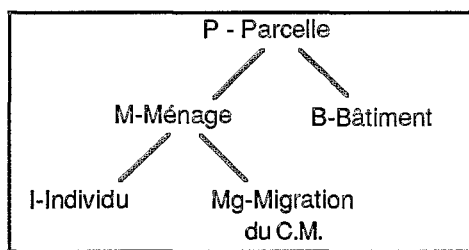
RESUME

Une enquête comportant plusieurs niveaux de collecte emboîtés fournit un ensemble d'informations hiérarchisées. C'est sur le croisement des informations et l'étude simultanée des niveaux que repose la richesse de l'enquête. L'exploiter, c'est :

- Concevoir un tableau rectangulaire qui croise l'ensemble des variables définies à plusieurs niveaux et qui traite une même population statistique. Pour cela, il s'agit de pondérer toutes les unités statistiques et nous mettons en évidence un système de pondération cohérent donné par une population statistique de référence.
- Envisager l'étude d'un ensemble des niveaux emboîtés qui constitue un système d'échelles. Il s'agit de dégager les structures spécifiques de chaque niveau et celles que l'on perd dans le transfert d'échelle. Nous proposons une méthode basée sur l'analyse des correspondances et qui consiste en une succession d'analyses inter-classes et intra-classes qui jouent le rôle d'aide à l'interprétation du système d'échelles.

INTRODUCTION

Il est fréquent de réaliser des enquêtes qui font intervenir différents niveaux de collecte ; par exemple, la parcelle, les ménages, les individus composant le ménage, les bâtiments obtenus à partir d'un plan de sondage à plusieurs degrés (stratifié, en grappe).



Ces niveaux sont emboîtés et l'on arrive alors à une "structure hiérarchique" de l'information: chaque individu est inclus dans un ménage, chaque ménage dans une parcelle.

L'objectif est de dégager les structures de cet univers composite. Autrement dit, on désire croiser les variables définies à plusieurs niveaux afin de conserver et de confronter toute l'information contenue à tous les niveaux. Pour cela, l'analyse multivariée et en l'occurrence l'analyse des correspondances multiples est adaptée puisqu'il s'agit, ici, de traiter des variables qualitatives issues de données d'enquête. Pour mettre en route une analyse multivariée, une grande partie du travail réside dans la conception du tableau qui doit d'une part être un tableau rectangulaire de données homogènes (travailler sur une même population statistique) et d'autre part respecter la structure hiérarchique de l'information.

Par ailleurs, on dispose de plusieurs échelles de perception de l'information et l'on peut envisager une approche ascendante (de l'individu à la parcelle), ou au contraire descendante (de la parcelle à l'individu) en considérant l'échelle intermédiaire (le ménage). Autrement dit, décrire un ensemble de variables définies à plusieurs niveaux revient à comprendre le système d'échelles que constitue l'univers d'enquête.

Deux questions sont alors posées et structurent l'exposé :

- Comment croiser des variables se trouvant à des niveaux différents et qui correspondent à des populations statistiques différentes ?

Ce qui revient à se demander quelle structure donner au tableau traitant d'une information hiérarchisée ?

- Comment observer les modifications de la perception d'un phénomène d'un niveau à un autre ?

Autrement dit, comment étudier un système d'échelles défini comme un ensemble de niveaux emboîtés ?

1 LA STRUCTURE DU TABLEAU

Dans le cadre d'une enquête à plusieurs niveaux, on définit plusieurs types d'unités statistiques, sur lesquelles sont effectuées des observations. Il s'agit alors de gérer et de traiter la complexité des relations entre ces unités. Le traitement statistique impose de travailler sur un tableau rectangulaire et sur une population homogène. A partir de là on cherche à croiser les variables définies à différents niveaux. On précise qu'il s'agit de variables qualitatives.

1.1. Le problème :

Nous nous plaçons dans le cadre de trois niveaux consécutifs d'observations :

P - Parcelle ---> 850 parcelles et 15 observations

|

M - Ménage ---> 1000 ménages et 50 observations

|

I - Individu ---> 5500 individus et 10 observations

A un niveau correspond une population statistique. Par conséquent, nous avons trois populations d'étude possibles: les Parcelles ou les Ménages ou les Individus, et trois séries d'observations effectuées sur chaque population c'est-à-dire trois groupes de variables J_P , J_M , J_I qu'il s'agit de croiser entre eux.

Soit trois variables prises sur les différents niveaux :

- au niveau Parcelle ; ELC, l'électricité éclatée en deux modalités suivant si elle est installée ou non sur la parcelle : "elect" et "non el" On a le tableau logique noté KP traitant la population Parcelle :

identificateur	var. Parcelle : J _p		
	ELC		
n°parc.	elect	non el	
1	1	0	***
2	0	1	
3	0	1	
***	***		

* au niveau Ménage ; STOC, le statut d'occupation ayant trois modalités : "propriétaire", "locataire" et "hébergé". On a le tableau logique noté KM traitant la population ménage :

	var. Ménage : J _M					
	STOC					
n°mén.	pro loc heb					n°parc.
1		1	0	0		1
2		0	0	1		1
3		0	0	1		1
4	- - - -	1	0	0		2
5		0	0	1		2
6	- - - -	0	1	0		3
....						

* au niveau Individu ; Sexe, le sexe ayant deux modalités : "masculin" et "féminin". On a le tableau logique noté KI traitant la population Individu :

identificateur		var. Individu : J _I					
n°ind.		SEXE				n°	n°
		mas		fem		parc.	mén.
1		1	0			1	1
2		1	0		----	1	1
3		0	1			1	2
4		1	0			1	3
5		0	1			1	3
6		0	1		--	2	4
7		1	0			2	4
8		1	0			2	4
9		0	1		----	2	5
10		0	1			3	6

On se rend compte que, face à la structure des données, **on dispose de plusieurs populations statistiques potentielles et de plusieurs échelles de perception de l'information**. Par exemple, si on choisit de travailler sur la population Ménage, on peut exploiter les variables définies à tous les niveaux à l'échelle du Ménage mais aussi à celle de la Parcelle ou de l'Individu.

Suivant les objectifs de l'étude, il convient alors de se demander sur quelle population statistique et à quelle échelle le traitement des variables doit être effectué. Compte tenu des choix à faire, il s'agit de construire un tableau juxtaposant l'ensemble des variables définies à plusieurs niveaux. Pour cela, nous allons explorer la structure hiérarchique de l'information.

1.2.Principe de la construction des tableaux :

Il existe deux types de tableaux pour exploiter des variables qualitatives:

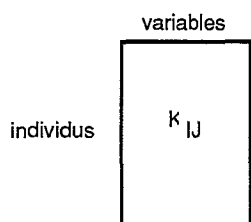


tableau juxtaposant les variables

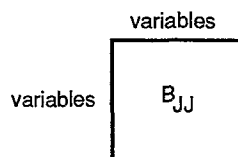


tableau croisant les variables

- Soit un tableau (individus, variables) qui juxtapose les variables tels que les tableaux disjonctif complet ou de correspondances.
- Soit un tableau (variables, variables) qui croise les variables entre elles tel que le tableau de Burt. Ce sont ceux là que l'on utilise généralement pour les traitements classiques d'enquête.

On passe du premier au deuxième par la formule :

$$B_{jj'} = \sum_i \frac{K_{ij} K_{ij'}}{K_i}$$

C'est-à-dire on comptabilise le nombre d'individus qui ont répondu à la fois à la variable en ligne et à la variable en colonne.

1.3.La structure hiérarchique de l'information :

Les niveaux sont emboîtés : une parcelle p définit un ensemble de ménages m et un ensemble d'individus i ; un ménage m définit un ensemble d'individus i :

$$\forall i \in I, \exists m \in M \text{ et } p \in P \text{ tel que } p \supset m \supset i$$

Si n est le nombre d'individus, n_m le nombre d'individus dans le ménage m et n_p le nombre d'individus résidant sur la parcelle p alors:

$$n_m = \sum_{i \in m} 1 ; n_p = \sum_{m \in p} n_m$$

et l'on a naturellement

$$n_p = \sum_{i \in p} 1 = \sum_{m \in p} \sum_{i \in m} 1 .$$

Par conséquent,

$$n = \sum_p n_p = \sum_m n_m = \sum_i n_i = \sum_p \sum_{m \in p} \sum_{i \in m} n_i .$$

On a de la même manière, pour m désignant le nombre de ménages et m_p le nombre de ménages résidant sur la parcelle p :

$$m_p = \sum_{m \in p} 1 \text{ et } m = \sum_p m_p$$

Le lien entre les informations est traduit par l'emboîtement entre les niveaux et est matérialisé par un identificateur affecté à chaque unité statistique p , m ou i . Un niveau est, ici, assimilé à une partition ce qui nous amène à considérer **une série de partitions emboîtées** engendrées par les populations des niveaux les moins fins sur les populations des niveaux les plus fins.

1.4. La population statistique :

Afin de travailler sur une population homogène, il est nécessaire de pondérer toutes les unités statistiques. La méthode proposée consiste à se fixer une population statistique de référence R parmi I , M , P , et on donne la pondération p_r à chaque unité de cette population ($p_r = 1$ s'il n'y a aucun redressement à faire). Compte tenu de l'emboîtement des unités statistiques, la pondération des unités des autres populations dépend de la pondération de référence et l'on doit avoir :

- si l'étude porte sur les Individus, $R=I$, on pose $p_r = p_i$ et l'on a :

$$p_p = \sum_{i \in p} p_i \text{ et } p_m = \sum_{i \in m} p_i$$

- si l'étude porte sur les Ménages, $R=M$, on pose $p_r = p_m$ et l'on a :

$$p_p = \sum_{m \in p} p_m \text{ et } p_t = \frac{p_m}{\sum_{i \in m} 1}$$

- si l'étude porte sur les Parcelles, $R=P$, on pose $p_r = p_p$ et l'on a :

$$p_m = \frac{p_p}{\sum_{m \in p} 1} \text{ et } p_t = \frac{p_m}{\sum_{i \in m} 1}$$

où $\sum_{i \in m} 1 = n_m$ et $\sum_{m \in p} 1 = m_p$. On vérifie que

$$\sum_r p_r = \sum_a p_a = \sum_b p_b = \sum_c p_c$$

et est égale à l'effectif de la population de référence si $p_r = 1$.

On met ainsi en évidence **un système de pondération** cohérent pour toutes les unités statistiques emboîtées de la hiérarchie et qui dépend :

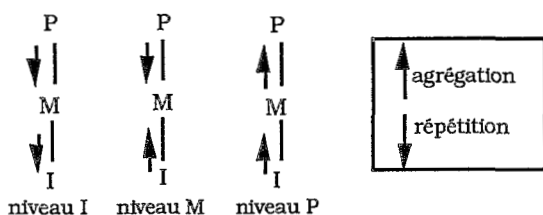
- de l'unité statistique de référence
- de l'emboîtement de niveaux.

1.5.L'échelle de perception :

Une fois fixée la population d'étude, il s'agit de construire à chaque échelle un tableau rectangulaire juxtaposant l'ensemble des variables. Pour cela, nous avons deux cas de figures :

- lorsque l'échelle de perception est plus fine que le niveau de définition des variables, ces variables sont répétées. Il y a **répétition** des informations des unités supérieures sur les unités inférieures qu'elles définissent.

- lorsque l'échelle de perception est moins fine que le niveau de définition des variables, celles-ci sont cumulées, autrement dit, elles sont agrégées. Il y a **agrégation** des informations des unités inférieures appartenant à une unité supérieure qui sont assimilées au centre de gravité de la classe définie par cette unité supérieure. Par conséquent, il y a une perte d'information.



On arrive alors à une structure de tableau hiérarchisé qui consiste en une série de tableaux HP, HM, HI, définis par chaque échelle et qui abordent une même population statistique donnée par un système de pondération défini précédemment.

Afin de minimiser la perte d'information, il est naturel d'examiner les phénomènes au niveau le plus fin et de "remonter" l'information vers les niveaux les moins fins.

• **à l'échelle de l'Individu**, on construit d'abord le tableau disjonctif complet pondéré $HI(I,J)$ défini sur $I*J$:

$$HI_{ij} = \begin{cases} p_i KI_{ij} & \forall j \in J_I \text{ et } \forall i \in I \\ p_i KM_{mj} & \forall j \in J_M \text{ et } \forall i \in m \\ p_i KP_{pj} & \forall j \in J_P \text{ et } \forall i \in p \end{cases}$$

Et l'on a par cumul les tableaux de correspondances suivants :

• **à l'échelle du Ménage**, le tableau $HM(M,J)$ défini sur $M*J$ par:

$$\forall j \in J: \quad HM_{mj} = \sum_{i \in m} HI_{ij}$$

• **à l'échelle de la Parcelle**, le tableau $HP(P,J)$ défini sur $P*J$ par:

$$\forall j \in J: \quad HP_{pj} = \sum_{i \in p} HI_{ij} = \sum_{m \in p} \sum_{i \in m} HI_{ij} = \sum_{m \in p} HM_{mj}$$

A trois populations statistiques potentielles correspondent par conséquent trois séries de trois tableaux données par la figure n°1.

SERIES DE TABLEUX JUXTAPOSANT DES VARIABLES DEFINIES A PLUSIEURS NIVEAUX

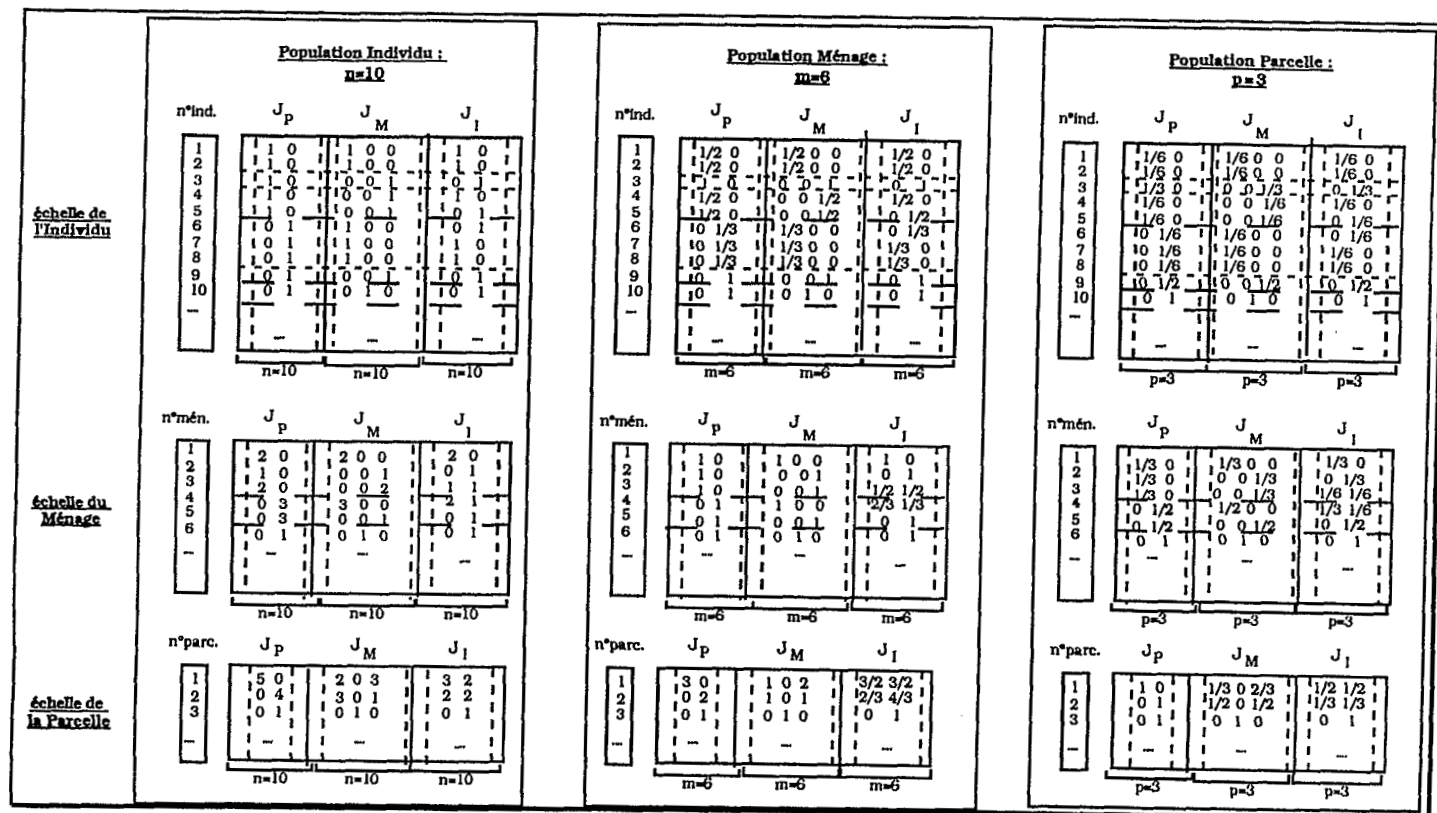
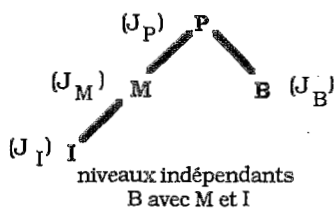


Figure n°1

1.6.Des niveaux indépendants :

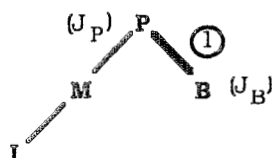
Nous avons envisagé jusqu'à présent des niveaux emboîtés mais l'on peut également concevoir des niveaux indépendants schématisés par:



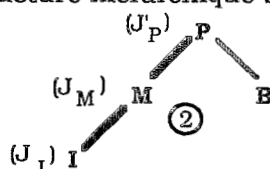
où B est l'ensemble des bâtiments de la parcelle et il n'y a pas de relation directe entre un ménage et un bâtiment. Les niveaux M et B ou I et B sont indépendants. Dans cette situation on procède en deux étapes qui nécessitent deux systèmes de pondération:

- On agrège l'information d'un des niveaux indépendants (Bâtiment par exemple) au niveau supérieur commun, la Parcelle. On augmente les variables de P et l'on pose alors

$$J'_P = J_P \cup J_B :$$



- On se ramène au cas précédent des niveaux emboîtés en considérant la structure hiérarchique suivante :



et l'on aboutit à la série de tableaux hiérarchisés suivante :

	J_P	J_B	J_M	J_I
P				
	agrégation au niveau P			

échelle de la Parcelle

	J_P	J_B	J_M	J_I
M				
	agrégation au niveau P			
	répartition au niveau M			

échelle du Ménage

	J_P	J_B	J_M	J_I
I				
	agrégation au niveau P			
	répartition au niveau I			

échelle de l'Individu

1.7. Un exemple sur deux niveaux :

L'exemple suivant considère les variables définies aux niveaux Parcelle et Ménage. Nous avons le choix de travailler soit sur la population des Ménages soit sur celle des Parcelles.

1.7.1. La population des Ménages :

L'unité statistique de référence est le Ménage. **On comptabilise le nombre de ménages** et l'on a le système de pondération :

$$p_m = 1 \text{ et } p_p = \sum_{m \in p} 1.$$

• A l'échelle du Ménage :

On construit un tableau à l'échelle la plus fine compte tenu des variables c'est-à-dire au niveau Ménage. On ne perd pas d'information. Les variables Parcelle sont répétées autant de fois qu'il y a de ménages sur la parcelle.

n° mén.	var. Parcelle			var. Ménage			n° parc.
	elect	non el	...	prop	loc	heb	
1	1	0	...	1	0	0	1
2	1	0	...	0	0	1	1
3	1	0	...	0	0	1	1
4	1	0	...	0	0	1	2
5	0	1	...	0	1	0	3
6	1	0	...	1	0	0	3
7	1	0	...	0	1	0	3
8	0	1	...	0	0	1	3
9	0	1	...	0	0	1	4
10	0	1	...	0	0	1	4

	var. Parcelle			var. Ménage		
	elect	non el	...	prop	loc	heb
var. Parcelle	6	0	...	2	1	3
	0	4	...	0	1	3
var. Ménage	...	...	...	2	0	0
	...	...	...	0	2	0
	...	...	...	0	0	6

A l'examen du tableau croisé, on obtient, au croisement des deux modalités "elect" et "heb", le nombre de ménages hébergés qui ont l'électricité sur la parcelle.

On réalise une analyse des correspondances sur ce tableau.

• A l'échelle de la Parcelle :

Mais on peut aussi concevoir une exploitation des variables à l'échelle de la Parcelle. Par exemple si l'on veut cartographier les résultats. On agrège alors l'information et on recrée un tableau (Parcelle, Variables) où ce sont les variables descriptives du ménage et de la parcelle.

n° parc	var. Parcelle			var. Ménage		
	elect	non el	...	prop	loc	heb
1	4	0	...	1	0	3
2	0	1	...	0	1	0
3	2	0	...	1	1	0
4	0	3	...	0	0	3

	var. Parcelle			var. Ménage		
	elect	non el	...	pro	loc	heb
var. Parcelle	6	0	...	2	1	3
	0	4	...	0	1	3
var. Ménage			...	0,8	0,5	0,7
			...	0,5	1,5	0

Au croisement des modalités "elect" et "heb", on a comme précédemment le nombre de ménages hébergés qui ont l'électricité sur la parcelle. Mais au croisement des modalités "prop" et "heb", on a le nombre de ménages propriétaires et hébergés sachant qu'ils résident sur une même parcelle. Il s'agit d'une moyenne conditionnelle.

Réaliser une analyse des correspondances sur ce tableau revient à effectuer une analyse inter-classes où les classes sont les parcelles.

1.7.2. La population des Parcelles :

On envisage, maintenant l'étude à partir de la population des parcelles qui devient la population de référence. **On compte ici des parcelles.** Le système de pondération est le suivant :

$p_p=1$ et $p_m = \frac{1}{\sum_{m \in p} 1}$. Les tableaux sont construits comme précédemment et l'on obtient.

• A l'échelle du Ménage :

n° mén.	var. Parcelle			var. Ménage			n° parc.
	elect	non el	...	prop	loc	heb	
1	1/4	0	...	1/4	0	0	1
2	1/4	0	...	0	0	1/4	1
3	1/4	0	...	0	0	1/4	1
4	1/4	0	...	0	0	1/4	2
5	0	1	...	0	1	0	3
6	1/2	0	...	1/2	0	0	3
7	1/2	0	...	0	1/2	0	3
8	0	1/3	...	0	0	1/3	3
9	0	1/3	...	0	0	1/3	4
10	0	1/3	...	0	0	1/3	4

	var. Parcelle			var. Ménage		
	elect	non el	...	pro	loc	heb
var. Parcelle	2	0	...	0,7	0,6	0,7
	0	2	...	0	1	1
var. Ménage			...	0,75	0	0
			...	0	1,5	0

Au croisement des deux modalités "elect" et "heb", on a le nombre de parcelles qui ont l'électricité et qui abritent des hébergés.

• A l'échelle de la Parcelle:

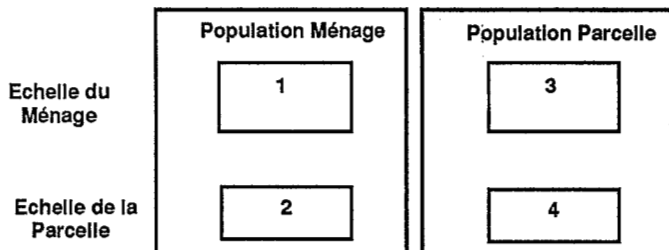
n°parc	var. Parcelle			var. Ménage		
	elect	non el	...	prop	loc	heb
1	1	0		1/4	0	3/4
2	0	1		0	1	0
3	1	0		1/2	1/2	0
4	0	1		0	0	1

	var. Parcelle		var. Ménage			
	elect	non el	pro	loc	heb	
var. Parcelle	elect	2	0		0,7	0,6
	non el	0	2		0	1
var. Ménage	pro			0,3	0,25	0,2
	loc			0,25	1,2	0
	heb			0,2	0	1,6

Au croisement des modalités "prop" et "heb", on compte le nombre de parcelles qui abritent en même temps des propriétaires et des hébergés. Il s'agit d'une moyenne conditionnelle.

1.8.Conclusion :

Pour deux niveaux, on a deux populations possibles pour exploiter les variables. Une fois fixer la population de référence, on a au moins autant de tableaux qu'il y a de niveaux.



On aboutit à la construction de quatre tableaux qui croisent l'ensemble des variables définies à plusieurs niveaux ce qui élargit les possibilités d'analyse puisqu'en général on étudie la population correspondant au niveau le plus fin compte tenu des variables considérées. C'est le premier tableau que l'on exploite pour le traitement de données d'enquête ou le second dès que l'on veut cartographier des résultats. Les analyses sont effectuées séparément les unes des autres.

2. L'ETUDE DU SYSTEME D'ECHELLES

L'analyse d'un des tableaux décrit précédemment permet l'étude d'une population à une échelle de perception donnée. Il est intéressant de savoir ce qui se passe dans le transfert d'échelle. Par conséquent, l'idée, ensuite, n'est pas de travailler sur un niveau mais sur tous les niveaux envisagés dans l'étude et d'une manière simultanée. Il s'agit d'expliquer la formation des niveaux et les passages entre les

niveaux. C'est ce qui contribue à mettre en évidence toute l'organisation du système d'information.

2.1. le principe :

Un système d'échelles se compose d'une part d'un ensemble de niveaux et d'autre part d'un lien entre ces niveaux qui est traduit par une série de niveaux emboîtés. On l'aborde par :

- Une analyse par niveau, c'est-à-dire l'exploitation du premier élément du système, qui donne la structuration du phénomène étudié sous des formes différentes. Chaque niveau fournit un point de vue particulier sur cette structuration.
- Une analyse du passage, c'est-à-dire l'exploitation du deuxième élément du système, qui établit la jonction entre ces différents points de vue. Elle fournit les structures que l'on perd dans le transfert et complète ainsi la connaissance.

Cette étude doit permettre de répondre aux questions suivantes :

- Quel niveau choisir, autrement dit, quelle est l'influence du découpage, géographique ou non, sur la perception d'un phénomène donné ?
- A quel niveau une variable n'intervient plus ? Trouve-t-on les mêmes facteurs aux différentes échelles ? Si oui, quels sont-ils suivant l'échelle considérée ?

Ce qui nous amène également à définir des niveaux d'intervention des variables et à réfléchir sur un critère d'homogénéité des niveaux.

2.2. La décomposition de l'inertie pour une série de partitions emboîtées:

Pour comprendre le système d'échelles, on va utiliser la décomposition de l'inertie suivant le principe de Huygens : l'inertie d'un nuage de points I muni d'une partition se décompose en inertie inter-classes et en inertie intra-classes. On démontre alors que, lorsque le nuage est muni d'une série de partitions emboîtées (P et M), cette décomposition devient :

$$In(I) = Inter(P) + Intm(MP) + Intra(M)$$

où $\text{Inter}(P)$ = inertie inter-classes associées à la partition P sur I
 $\text{Intra}(M)$ = inertie intra-classes associées à la partition M sur I
 $\text{Intm}(MP)$ = inertie intra-classes pour la partition engendrée par P sur l'ensemble des centres de gravités des classes associées à la partition M sur I .

Et l'on a :

$$\text{In}(I) = \underbrace{\text{In.inter}(P)}_{\text{In.inter}(M)} + \text{In.intm}(MP) + \overbrace{\text{In.intra}(P)}^{\text{In.intra}(M)}$$

On généralise cette formule en considérant l'inertie inter-classes associées à la partition P sur I , ($\text{Inter}(P)$), comme l'inertie intra-classes associées à la partition réduite à une seule classe $\{I\}$ sur A . La décomposition de l'inertie peut être envisagée comme la somme des inerties intra-classes entre deux niveaux, et l'on a pour une série de partitions emboîtées $(P_0, \dots, P_N)$, $P_0=I$ et $P_N=\{I\}$:

$$\text{In}(I) = \sum_I^N \text{In.intm}(P_{n-1}P_n)$$

2.3.L'analyse des niveaux :

L'analyse des niveaux revient à réaliser **une succession d'analyses inter-classes** et à dégager les structures induites par niveau. On saisit ainsi l'influence des niveaux sur la perception des phénomènes.

L'analyse inter-classes étudie l'écart entre le profil moyen d'une classe et le profil moyen du nuage. Un individu i d'une classe (d'une parcelle) p est assimilé au centre de gravité de p et n'est étudié qu'au travers de celui-ci. Elle traduit la dispersion inter et est réalisée sur l'ensemble des barycentres des classes (ce qui est équivalent à l'analyse du cumul de l'information par classe). On réalise par conséquent une analyse des correspondances sur les tableaux HM (ou BM) et HP (ouBP) pour les analyses des niveaux respectivement Ménage et Parcelle.

2.4.L'analyse des passages entre les niveaux :

L'analyse inter-classes examine les structures communes aux classes. Par conséquent, les variations internes aux classes données

par l'analyse intra-classes sont éliminées. Ce sont les structures perdues dans le transfert vers un niveau supérieur. Elles définissent les caractéristiques du passage d'un niveau à un autre et donnent les éléments qui diffèrent entre eux et se regroupent dans les mêmes classes au niveau inférieur. Il s'agit d'une analyse intra-classes à partir du niveau inférieur c'est-à-dire d'une analyse inter-classes. L'analyse des passages revient alors à réaliser une succession d'analyses intra-classes associées à la partition définie par le niveau le moins fin sur l'ensemble des classes défini par le niveau le plus fin.

L'analyse intra-classes étudie la dispersion des éléments à l'intérieur des classes. De ce fait, les profils sont analysés non pas par rapport au profil moyen du nuage mais par rapport au profil moyen de leur classe. On réalise alors une analyse des correspondances sur les tableaux :

- pour l'analyse du passage entre les niveaux Individu et Ménage

$$HI^* = HI - HM + (\text{produit des marges})$$

- pour l'analyse du passage entre les niveaux Ménage et Parcelle

$$HM^* = HM - HP + (\text{produit des marges})$$

2.5. L'analyse du système d'échelles :

Effectuer des analyses inter ou intra-classes permet **d'identifier ce qui est conservé (analyse inter) ou ce qui est perdu (analyse intra) dans le transfert d'échelle.**

Intuitivement, pour un nombre d'individus donné, plus une partition est fine, plus le nombre de classes d'une partition est élevé (faible effectif d'individus dans les classes) et plus ces classes sont homogènes.

On sait que si les classes sont homogènes, pour une partition donnée, c'est-à-dire s'il existe un regroupement d'éléments distributionnellement proches, il est sensiblement équivalent de réaliser l'analyse à un niveau supérieur déterminant ces classes. En effet, partant du principe de l'équivalence distributionnelle, on peut alors agréger les éléments d'une même classe, autrement dit effectuer un changement d'échelle, sans que l'analyse soit trop perturbée : *"l'intérêt de l'analyse des correspondances est que, du fait de l'équivalence distributionnelle, cette méthode est peu sensible au détail de la partition adoptée"*¹.

¹BENZECRI et coll, 1973, L'analyse des données, Tome IIA n°1 p21.

ANALYSE DU SYSTEME D'EHELLES

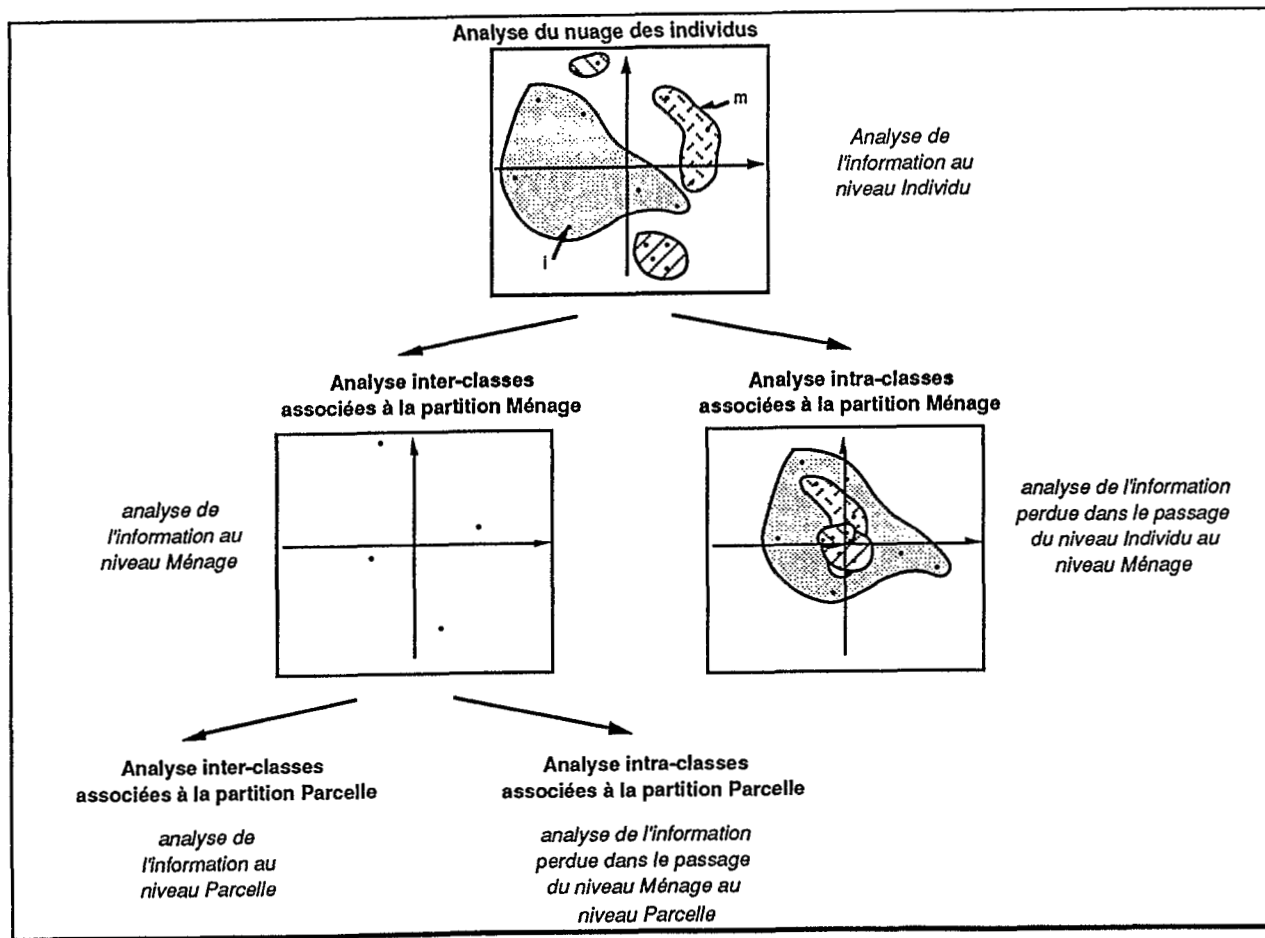


Figure n°2

Mais, la démarche d'un système d'échelles suppose que les classes soient déjà établies. Par conséquent, on ne cherche donc pas, comme pour une classification, à construire des classes ne comprenant que des individus semblables entre eux, suivant des critères d'agrégation restant à définir (pour un tableau de correspondances, celui-ci maximise l'inertie inter-classes). Ici, ce critère est imposé "naturellement" par le lien hiérarchique et une classe constitue alors un agrégat d'individus plus ou moins homogène.

Par conséquent, si l'inertie intra-classes définie entre deux niveaux est faible (peu de variation dans le passage), on peut ramener l'étude au niveau supérieur d'agrégation qui conserve alors la stabilité du phénomène observé au niveau inférieur. Les classes sont globalement homogènes par rapport aux classes du niveau inférieur.

De plus, le choix de l'échelle n'est pas neutre. Pour deux systèmes d'échelles de découpage en niveaux identiques, on peut avoir des comportements différents : les classes définies par niveau sont globalement plus ou moins homogènes et une variable peut intervenir pour un seul des niveaux. Mais il est délicat de fixer un critère d'homogénéité à partir des taux d'inertie.

On peut, cependant, donner une importance d'homogénéité du niveau par l'examen des deux séries d'analyses : l'analyse inter-classes c'est-à-dire l'analyse des niveaux et l'analyse intra-classes c'est-à-dire l'analyse des passages entre les niveaux. En effet :

- Si l'analyse du niveau le moins fin rend compte au mieux des mêmes structures que celle du niveau le plus fin alors l'analyse du passage entre ces deux niveaux dégage une perte d'information due à l'agrégation mais qui ne bouleverse pas l'organisation des données : l'inertie intra-classes est faible et le niveau le moins fin est globalement homogène (il rend compte de classes homogènes).

- Si l'analyse du niveau le moins fin, par contre, ne se structure pas comme celle du niveau le plus fin alors l'analyse du passage rend compte d'une même organisation des données que celle du niveau inférieur : l'inertie intra-classes est élevée et le niveau le moins fin est globalement hétérogène. Il s'est passé quelque chose dans le transfert d'échelle qu'il peut être intéressant de connaître

L'analyse de l'inertie la plus faible rend compte au mieux, de structures qui ne peuvent apparaître dans les autres analyses et complète ainsi la connaissance du phénomène étudié.

La figure n°2 représente la décomposition de l'inertie pour une série de partitions emboîtées (Parcelle et Ménage) à partir du nuage d'individus N(I).

2.6. Niveaux d'intervention des variables :

Sachant que l'inertie du nuage est la somme des inerties des variables, on décompose l'inertie d'une variable, v , suivant le même principe de la décomposition pour une série de partitions emboîtées :

$$In(v) = Inter(v,P) + Intm(v,MP) + Intra(v,M)$$

En relativisant ces formes d'inerties à l'inertie totale de la variable, on obtient la contribution de l'inertie décomposée de la variable v à l'inertie de la variable. Par exemple :

- pour le passage des niveaux Ménage et Parcelle,

$$CTR(v,MP) = \frac{Intm(v,MP)}{In(v)}$$

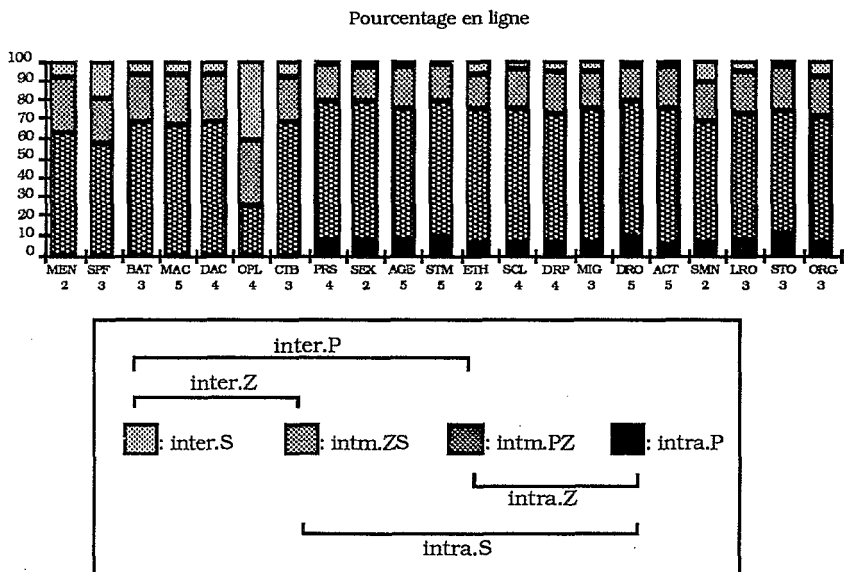
- au niveau M,

$$CTR(v,M) = \frac{Inter(v,M)}{In(v)}$$

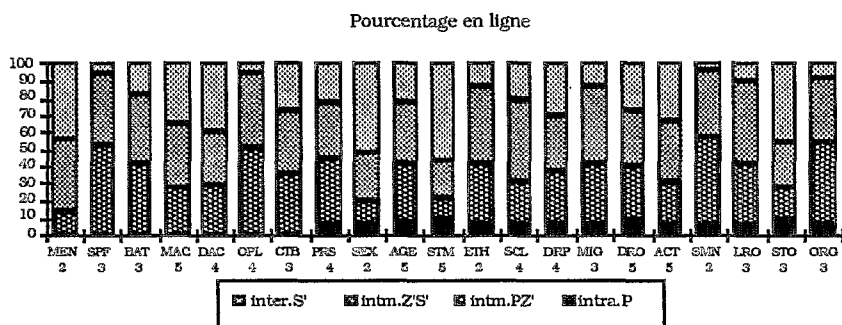
Cela nous autorise à construire un graphique qui permet une lecture simultanée des différents niveaux d'intervention des variables.

Nous présentons deux graphiques issus d'une étude urbaine sur Ouagadougou au Burkina Faso.

Le premier graphique est issu d'un système d'échelles Ménage-Parcelle-Zone de dénombrement-Secteur :



Le deuxième est issu d'un système d'échelles dont les niveaux ont le même nombre de classes que le système d'échelle précédent mais il est obtenu à partir d'une classification ascendante hiérarchique dont le critère d'agrégation est fondé sur la minimisation de l'inertie intra-classes. Par conséquent, ce système d'agrégation rend compte au mieux de niveaux homogènes :



L'allure générale de ces deux graphiques est très différente. Le dernier dénote un nombre plus important de variables qui interviennent aux niveaux supérieurs contribuant alors à enrichir les typologies de ces niveaux. Le premier système d'échelles démontre l'hétérogénéité des niveaux supérieurs Zone de dénombrement et Secteur (forte dispersion entre les niveaux Parcelle et Zone de dénombrement).

Ces graphiques ne permettent pas, évidemment, de donner des associations de variables. Leur intérêt est de définir le niveau d'intervention des variables : plus l'inertie inter-classes associée à une partition est importante, plus la variable contribue à caractériser le niveau correspondant. Inversement, plus l'inertie intra-classes est élevée, plus la variable contribue à caractériser le niveau inférieur considéré dans le passage.

2.7. Conclusion

Décrire un ensemble d'informations hiérarchisées revient à appréhender un système d'échelles comportant un nombre variable de niveaux. Le sens que l'on donne, ici, à l'échelle doit être élargi et relativisé.

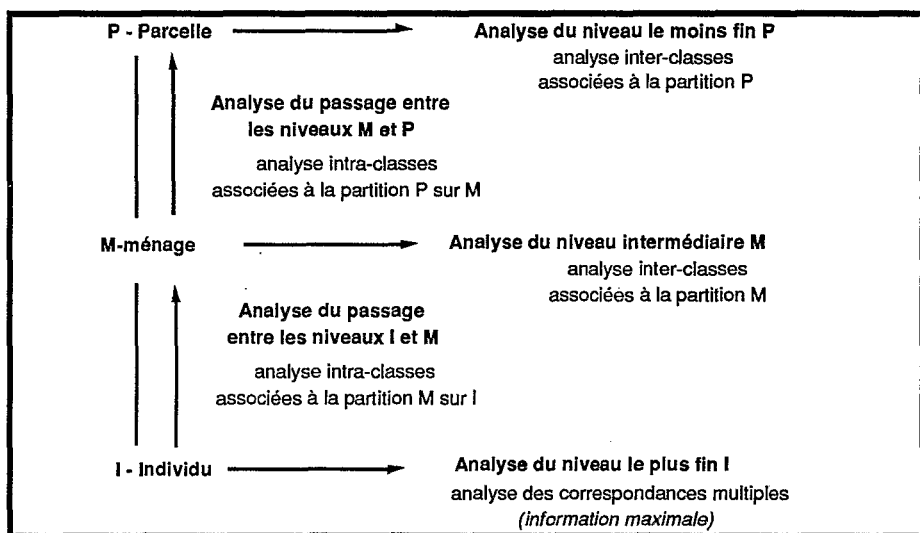
Pour exploiter l'information contenue dans un système d'échelles, la démarche à suivre est la suivante :

- 1- on réalise une analyse au niveau le plus fin et on obtient une organisation des données du maximum d'information.
- 2- puis, au niveau supérieur immédiat on réalise une double analyse :

- l'analyse du niveau qui donne l'organisation des données à ce niveau.
- l'analyse du passage qui donne l'information contenu au niveau inférieur qui est perdue dans le passage et qui rend compte de la variabilité interne de ce niveau.

3- on répète cette série de deux analyses aux niveaux supérieurs suivants .

Le processus de l'analyse d'un système d'échelles revient à faire des "va et vient" d'une échelle à une autre pour comprendre ce qui se passe à toutes les étapes du système. Celles-ci sont schématisées de la manière suivante :



On peut rapprocher le processus d'un système d'échelles de celui d'une classification ascendante hiérarchique mise à part que :

• **La classification ascendante hiérarchique est un processus automatique qui cherche à organiser des données en créant un lien** (défini suivant un critère d'agrégation de la maximisation de l'inertie inter-classes). Le principe est de s'intéresser à un ensemble de classes pour une partition fixée.

• **Le système d'échelles est un processus "naturel" qui explique une organisation des données déterminée par l'existence d'un lien hiérarchique** tel que nous l'avons défini (c'est-à-dire traduit par une série de partitions emboîtées) et auquel il s'agit de donner un sens en terme d'homogénéité. Le principe est de saisir une ensemble de partitions c'est-à-dire de niveaux et à étudier leur influence pour un phénomène donné.

Le rapprochement s'effectue non pas dans la formation des classes mais dans leur interprétation.