

MINISTERE DE L'ECONOMIE, DU PLAN ET DE LA COOPERATION

AGENCE NATIONALE DE LA STATISTIQUE ET DE LA DÉMOGRAPHIE



ÉCOLE NATIONALE DE LA STATISTIQUE ET DE L'ANALYSE
ÉCONOMIQUE



TP: ANALYSE DE DONNEES

ANALYSE MULTIDIMENSIONNELLE DE LA
PAUVRETÉ INFANTILE AU SENEGAL

Réalisé par

DIALLO Alpha Oumar

GUEYE Pape Tidiane

Supervisé par

Dr. DIAKITE

Avril 2020

Sommaire

1	Présentation des données et de la méthodologie	2
1.1	Source de données	2
1.1.1	Présentation de la base EDS	2
1.1.2	Critère de choix des individus de l'échantillon et des variables	2
1.2	Méthodologie de l'étude	4
1.3	Procédures d'élaboration de l'ACM et de la classification	4
1.3.1	Analyse des Correspondances Multiples (ACM)	4
1.3.2	La Classification	5
1.4	Procédures de conception de l'Indice	6
1.4.1	Définition de concepts	6
1.4.2	Forme fonctionnelle de l'indice	7
2	Présentation des résultats	8
2.1	Description des variables	8
2.2	Interprétation de l'ACM	9
2.2.1	Apurement	9
2.2.2	Choix des axes	10
2.2.3	Caractérisation des axes	10
2.3	Nuage des individus	12
2.4	Classification des individus	14
2.4.1	Choix des axes	14
2.4.2	Caractérisation des classes	15
2.5	Construction de l'indice composite de la pauvreté infantile	17
2.5.1	Profil de pauvreté multidimensionnelle	17
2.5.2	Détermination du seuil de pauvreté	19

Table des figures

2.1	Histogramme des valeurs propres	10
2.2	Graphique des corrélations : variables-axes	11
2.3	Nuage des modalités de l'ACM	12
2.4	Nuage des points individus	13
2.5	Histogramme des indices de niveau	14
2.6	Dendrogramme	15

Liste des tableaux

1.1	Listes des variables retenues	3
2.1	Description des variables de l'étude	9
2.2	Caractéristiques de la première classe	16
2.3	Caractéristiques de la deuxième	17
2.4	Tableau descriptif de l'IPCI	18
2.5	Description de l'IPCI selon les régions	19
2.6	Commodité de base	20
2.7	Taux de pauvreté	20

Introduction

« Un bon dessin vaut mieux qu'un long discours », disent les statisticiens. Ainsi se justifie la quintessence de l'utilisation des méthodes d'analyse de données (ADD) qui permettent de transformer, de la manière la plus adéquate possible, de gros tableaux en images synthétiques, plus parlantes qui en dégagent les principales structures. Il faut préciser que l'ADD diffère de la statistique descriptive qui cherche à représenter de manière simplifiée les informations statistiques.

Les méthodes d'analyse de données sont plus puissantes et leur usage dépend tant bien de la nature de la problématique que des données. Ainsi, lorsqu'on dispose de variables quantitatives, il faut utiliser l'Analyse en Composants Principales (ACP). Par ailleurs, en face de deux variables qualitatives, il faut recourir à l'Analyse Factorielle des Correspondances (AFC). Avec plus de deux variables qualitatives, c'est l'Analyse des Composantes Multiples (ACM) qui s'applique. Cette dernière est d'ailleurs plus maniable dans la mesure où elle peut être utilisée pour analyser des variables quantitatives par transformation de celles-ci en variables qualitatives. On peut également procéder dans l'analyse des données, à des regroupements de groupes d'individus à priori différents en des classes homogènes : c'est la classification. En outre, on peut également citer l'Analyse Factorielle Discriminante (AFD) qui, comme l'indique son nom, sert à discriminer des individus décrits par un ensemble de variables quantitatives, par des variables qualitatives.

Pour notre analyse qui cherche à déceler une méthodologie de construction d'un indicateur de la pauvreté des enfants au Sénégal, nous allons utiliser l'ACM. Cette dernière sera suivie d'une partie réservée à la classification.

Pour mieux étayer notre travail, nous scinderons notre travail en deux parties. Dans la première, il sera question de la présentation de la source de données et des méthodologies à adopter dans la suite de l'analyse, la procédure d'élaboration de l'indice incluse. Dans la seconde partie, il s'agira de présenter les résultats d'abord sous forme d'analyse descriptive avant de finir avec ceux de l'analyse multidimensionnelle des données qui s'accompagne d'une application de l'indice aux données dont on dispose.

Présentation des données et de la méthodologie

1.1 Source de données

1.1.1 Présentation de la base EDS

Les données utilisées dans le cadre de notre étude sont celles des Enquêtes Démographiques et de Santé (EDS), et plus précisément celles de la base EDS 2017.

Pour cette enquête, la méthode d'échantillonnage est probabiliste, basée sur un sondage aréolaire stratifié et tiré à deux degrés. Au premier degré, 400 grappes ont été tirées à partir de la liste des Zones de Dénombrement (ZD) établies au cours du Recensement Général de la Population et de l'Habitat de, l'Agriculture et de l'Élevage (RGPHAE) réalisé en 2013, en procédant à un tirage systématique avec probabilité proportionnelle à la taille, la taille de l'UPS étant le nombre de ménages. Un dénombrement des ménages dans chacune de ces grappes a fourni une liste des ménages à partir de laquelle a été tiré, au second degré, un échantillon de 22 ménages par grappe, aussi bien en milieu urbain qu'en milieu rural avec un tirage systématique à probabilité égale. Un total de 8 800 ménages (4 092 en milieu urbain et 4 708 en milieu rural) a été sélectionné.

1.1.2 Critère de choix des individus de l'échantillon et des variables

1.1.2.1 Choix des individus

Notre sujet portant sur les enfants dont on s'intéresse à l'étude de la pauvreté, il s'avère nécessaire de bien définir la population des enfants. Autrement dit, nous sommes amenés à préciser qui, dans la base, sont considérés comme enfants. En effet, selon la

Convention Internationale des Droits de l'Enfant (CIDE) ratifiée par les pays de l'ONU en 1989, est enfant tout individu âgé de moins de 18 ans. Sa mise en application est assurée par l'UNICEF qui est en charge de l'éducation des enfants. Donc, en se basant sur ladite convention, nous avons restreint la base de données aux individus ayant un âge inférieur à 18ans.

1.1.2.2 Choix des variables

La pauvreté infantile est un phénomène multidimensionnel qui s'appréhende à travers les privations multiples dont sont victimes les enfants. Ainsi notre étude circonscrit le phénomène autour de 11 variables. Le choix de ces variables est guidé par le fait qu'elles traduisent les besoins prioritaires pour l'enfance ; qu'elles reflètent des dimensions standard pour caractériser la pauvreté au sens large et qu'au regard de la base de données elles sont bien renseignées (présentes moins de valeurs manquantes). Ces dimensions se présentent comme suit :

- ✦ **Le niveau de vie** : celui-ci s'identifie à travers les variables décrivant les dispositifs matériels présents dans le ménage auquel l'enfant s'identifie ;
- ✦ **La situation démographique et sanitaire** : ce bloc prend en compte les privations relatives au droit civil et à l'accès au service sanitaire ;
- ✦ **La situation éducative** : cette dimension, comme son nom le suggère, tient compte des privations des droits à l'éducation.

Ainsi les variables de l'étude sont :

Tableau 1.1 – Listes des variables retenues

Variables	Libellés
mil-res	Milieu de résidence
electricite	Présence d'électricité dans la maison
television	Dotation en télévision
mstiquaire	L'enfant dort sous une moustiquaire
mere-vit	Orphelin de mère
pere-vit	Orphelin de père
niv-educ	Niveau d'éducation
tailleMen	Taille du ménage
newsource-eau-boire	Source d'eau buvable
new-soin-santé	Utilisation des soins de santé durant l'année

Source : Base-EDS Sénégal

1.2 Méthodologie de l'étude

1.3 Procédures d'élaboration de l'ACM et de la classification

1.3.1 Analyse des Correspondances Multiples (ACM)

Nous avons des variables qui sont toutes qualitatives, après transformation. De ce fait, nous allons utiliser la méthode de l'ACM pour le traitement. L'Analyse des Correspondances Multiples (ACM) est une méthode d'analyse multidimensionnelle qui étudie les tableaux de données d'individus décrits par plusieurs variables qualitatives. Il s'agit d'une généralisation de l'Analyse Factorielle des Correspondances (AFC). Elle vise à cet effet à étudier les liaisons qui existent entre plusieurs variables qualitatives. Le procédé de l'ACM repose d'abord sur le choix des axes factoriels. Pour ce faire, nous allons utiliser le critère du taux d'inertie. Ce dernier est défini par :

Taux d'inertie

$$r_q = \frac{\sum_{j=1}^q \lambda_j}{\sum_{j=1}^K \lambda_j}$$



Le taux d'inertie consiste en une indication concernant la part d'information restituée par les q premiers axes. La valeur de q est choisie de sorte que r_q soit supérieure à une valeur seuil fixée a priori par les praticiens. Et, eu égard à la décroissance des valeurs propres, nous allons adjoindre à cette méthode la règle du coude de Cattell qui va nous permettre de décider du nombre d'axes à suivre, suite à une chute notable au niveau de l'histogramme des valeurs propres.

On interprétera la proximité entre individus en termes de ressemblance. Cette interprétation se fera avec la construction du nuage des individus, dans lequel les individus seront scindés suivant la variable considérée comme supplémentaire dans l'étude, à savoir le milieu de résidence.

Par ailleurs, concernant les variables, nous allons analyser le nuage des modalités construit sur la base de leurs contributions respectives. En effet :

Contribution des modalités

$$CTR(j) = \frac{\frac{1}{p}(1 - \frac{n_{.j}}{n})}{I_{Gq}}$$



On remarque, avec la formule, que la contribution d'une modalité j dans la construction d'un axe augmente si sa fréquence d'apparition diminue. Voilà pourquoi, dans notre étude, nous nous sommes fixé l'idée de réduire les modalités des variables de la base à 3 au maximum puisqu'en réalité on y retrouvait des modalités à faible valeur. En effet, nous avons choisi de considérer une modalité comme étant faible si elle représente moins de 5% des effectifs. Par ailleurs, le choix du nombre de modalités 3 découle du fait que la contribution d'une variable à l'inertie est fonction de son nombre de modalités. Ainsi, faire le maximum possible pour avoir le même nombre de modalités pour les variables rendrait plus judicieuse l'analyse qui s'ensuivra. Il faut noter la présence de variables dichotomiques, ce qui rend impossible de ne considérer que 2 modalités pour les variables et pour les autres variables regroupées en trois, nous avons constaté le regroupement impertinent en-dessous de ce nombre donné de modalités.

1.3.2 La Classification

L'ACM, vue précédemment, nous permet d'étudier les relations entre les variables. Cela permettra en même temps de procéder à la construction d'un indicateur composite dont la méthodologie sera présentée dans la partie développée après la classification.

Néanmoins, compte tenu du nombre élevé de variables, il serait intéressant de compléter l'analyse factorielle qu'est l'ACM par une typologie en classes homogènes. En effet, cela permettra de regrouper les individus en classes homogènes de sorte que ceux d'une même classe soient semblables et ceux de classes différentes, très différenciés : c'est le principe des méthodes de classification. Il existe quatre (4) principales méthodes de classification, à savoir la méthode ascendante hiérarchique, la méthode descendante hiérarchique, la méthode non hiérarchique de partitionnement et la méthode de segmentation.

Dans notre analyse, nous effectuerons une Classification Ascendante Hiérarchique (CAH). En premier lieu, il sera question du choix du nombre de classes optimales (partition en k classes optimales). En second lieu, avec le nombre de classes fixé au niveau de la partie précédente, nous nous adonnerons à une caractérisation par les variables avec comme outil de base le critère mod/class.

✦ Classification Ascendante Hiérarchique

Le principe de la CAH est de partir d'une population où chaque individu forme sa propre classe avec une inertie interclasse maximale et celle intra minimale. Normalement, à la fin du processus, on aura une seule classe où l'inertie intra est maximale et celle inter nulle. Cependant, l'algorithme s'arrêtera à un niveau où l'inertie inter et intra ne varie plus si l'on regroupe les individus. Les étapes de la CAH sont décrites ci-dessous :

Etape 1 : Nous calculons la distance entre tous les individus pris deux à deux. Nous regroupons les deux individus les plus proches. Ensuite, les deux individus sont remplacés par leur centre de gravité.

Etape 2 : Le tableau des distances est actualisé. Nous regroupons après cela les centres de gravité les plus proches (individus ou centre de gravité). Dans cette phase la distance la plus généralement utilisée est celle de Ward car possédant certaines caractéristiques comme la minimisation de la perte d'information par l'inertie intra-classes. Cette distance s'écrit sous la forme suivante :

Indice d'agrégation

$$D(A, B) = \frac{P(A)*P(B)}{P(A)+P(B)} * d(g(A), g(B))$$
 , avec $g(A)$, $g(B)$, $P(A)$ et $P(B)$ sont respectivement les centres de gravité des classe A et B et les proportions d'individus dans ces classes.

Etape k : On calcule la distance entre toutes les classes. On agrège les deux centres de classe les proches.

Etape n-1 : Etant la dernière étape, on agrège les deux classes qui se sont dégagées et on les remplace par leurs centres de gravité.

✦ Détermination de la partition optimale

Concernant le nombre optimal de classes à retenir, nous utilisons la distance de Ward définie plus haut. Cette distance a la particularité de faciliter l'interprétation du niveau des nœuds d'agrégation en termes de coût informationnel. Rappelons qu'un nœud de l'agrégation de deux classes A et B représentées par leurs centres de gravité respectifs $g(A)$ et $g(B)$, le point $g(A \cup B)$ et que le niveau de ce nœud est la valeur de l'inertie d'agrégation qui est égale l'inertie interne de la classe $A \cup B$ enlevée des inerties internes respectives de A et de B. Les partitions les plus intéressantes a priori sont celles obtenues en coupant les nœuds dont les niveaux forment un grand décrochement dans l'histogramme des niveaux sachant ces niveaux forment une suite croissante. Graphiquement, la partition optimale est obtenue en coupant l'histogramme des indices de niveaux à un endroit où on a un décrochage brutal.

1.4 Procédures de conception de l'Indice

1.4.1 Définition de concepts

En plus de donner une description multivariée de la pauvreté infantile, notre étude consistera aussi en la construction de l'Indice Composite de la Pauvreté Infantile. Cet indice sera un instrument de mesure du niveau cette pauvreté en faisant intervenir plusieurs

dimensions.

1.4.2 Forme fonctionnelle de l'indice

Bien vrai que plusieurs approches conduisent à la construction d'un indice (*approche par la théorie des ensembles flous, approche par entropie, ...*), le cadre pédagogique nous guide directement vers l'approche de l'inertie qui, contrairement aux autres approches, exclut le choix arbitraire des pondérations. Ainsi posé, cette approche fera intervenir l'ACM qui est la méthode d'analyse factorielle la plus adaptée à nos données (qualitatives). La forme fonctionnelle de l'ICPI est dans ce cas :

Indice Composite de Pauvreté Infantile (ICPI)

$$IPCI_i = \frac{\sum_t \sum_{jk} \omega_{jk}^t I_{ijk}^k}{T}$$

- $IPCI_i$: valeur de l'indice pour l'individu i ;
- ω_{jk}^t : poids de la modalité jk sur l'axe t ;
 - $\omega_{jk}^t = \frac{coord_{jkt}^2}{\lambda_t}$, sur SPSS;
 - $\omega_{jk}^t = \frac{coord_{jkt}}{\sqrt{\lambda_t}}$, sur SPAD;
 - λ_t : l'inertie par rapport à l'axe t ;
 - $coord_{jk}$: coordonnées de la modalité jk .
- T : le nombre d'axes.

Il est aussi important de noter que le poids ω_{jk}^t de la modalité jk sera affecté d'un signe moins lorsque cette modalité contribue négativement au bien être de l'enfant.

Présentation des résultats

2.1 Description des variables

Les variables de l'étude ont été retravaillées de sorte à avoir au plus trois modalités par variables. Le tableau ci-dessous montre comment ces modalités ont été renseignées.

Tableau 2.1 – Description des variables de l'étude

Variables	Modalités	Fréquences
Télévision	Non	46,4
	Oui	53,6
	Total	100
Electricité	Non	42,69
	Oui	57,31
	Total	100
Source d'eau à boire	Eau canalisée	69,45
	Eau de source (puits, forage, surface)	25,73
	Citerne ou eau en bouteille	4,82
	Total	100
Milieu de résidence	Urbain	39,02
	rural	60,98
	Total	100
Taille du ménage	Petite taille et Moyenne	66,71
	Grande taille	33,29
	Total	100
Mère vivante	Non	1,76
	Oui	98,24
	Total	100
Père vivant	Non	4,40
	Oui	95,60
	Total	100
Dort sous une moustiquaire	Non	9,71
	Oui	90,29
	Total	100
Niveau d'étude	Préscolaire ou sans niveau	62,95
	Primaire	28,49
	Secondaire	8,56
	Total	100
Soin de santé	Non	65,4
	Oui	34,96
	Total	100

Source : Calcul des auteurs.

2.2 Interprétation de l'ACM

2.2.1 Apurement

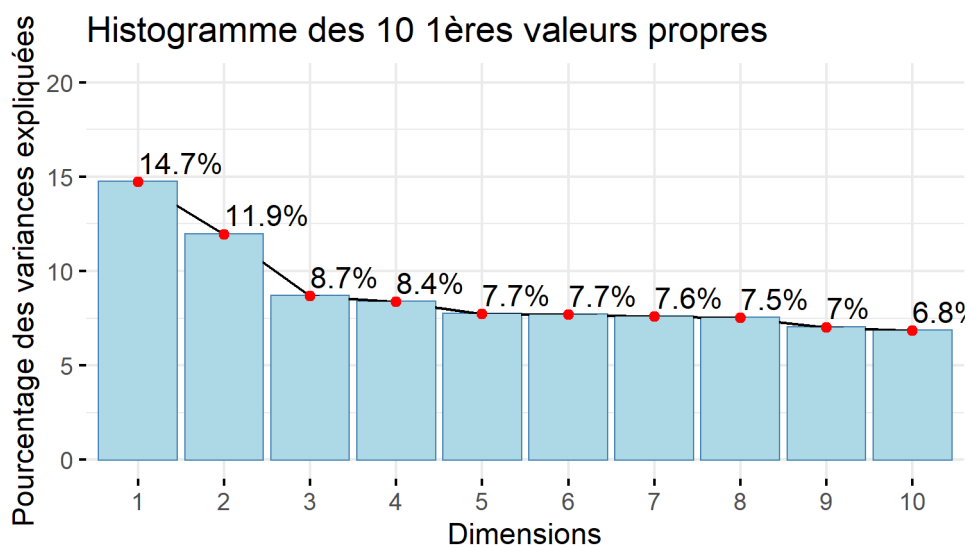
Avant d'entamer l'ACM deux traitements majeurs sont à préciser. D'abord, la contribution d'une modalité étant d'autant plus forte que sa fréquence d'apparition est faible,

nous avons donc insérer les modalités de faible fréquence dans les modalités les plus semblables (i.e : pour la variable taille du ménage, nous avons réuni petite taille et taille moyenne pour donner petite et moyenne). Pour rappel, nous avons considéré comme une fréquence faible toute fréquence inférieure à 5% et cette transformation ne concerne que les variables ayant plus de deux modalités. Ensuite, sachant que la contribution d'une variable est fonction de ses nombres de modalités, nous avons donc essayé de ramener toutes les variables au maximum à trois modalités.

2.2.2 Choix des axes

Il existe diverses méthodes de détermination du nombre d'axes à retenir. Dans notre analyse, l'utilisation de la règle du coude de Cattell basé sur l'histogramme des valeurs propres donne les résultats suivants :

Graphique 2.1 – Histogramme des valeurs propres



Source : Calcul des auteurs.

D'après ce graphe, nous allons retenir deux (2) axes factoriels. En effet, on remarque sur cet histogramme une chute brutale dès la deuxième valeur propre : c'est le critère du coude Cattell.

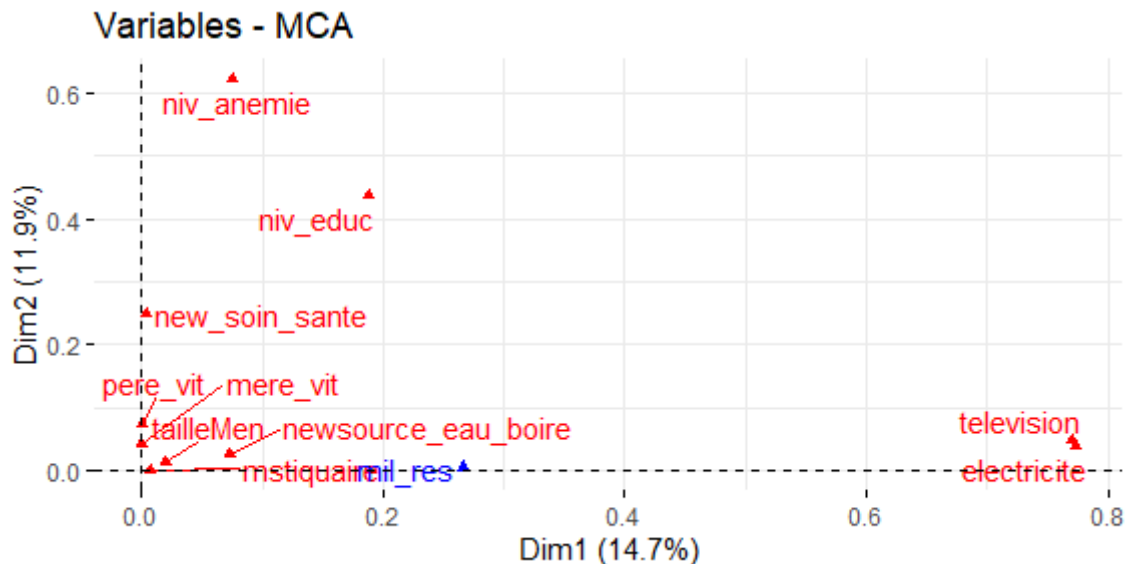
2.2.3 Caractérisation des axes

Les deux axes retenus pour l'analyse des correspondances multiples donnent 25,9% de l'information contenue dans les données, avec 14,7% pour le premier axe et 11,2% pour le second axe. A travers les variables puis les modalités retenues, nous obtenons ce que représente chaque axe.

2.2.3.1 Caractérisation par les variables

Le graphique des carrés des corrélations ci-dessous donne une représentation des variables en fonction de leur liaison avec les axes. Chaque variable est repérée par un système de deux coordonnées représentant les corrélations de la variable avec les deux axes. On remarque, d'abord, que les variables télévision, électricité. L'axe 1 est donc déterminé par des variables de **dimension économique**, c'est-à-dire celles qui montrent le niveau de vie à travers les avoirs des ménages. Ensuite, l'axe 2, quant à lui, est très corrélé avec les variables telles que niveau d'anémie, niveau d'éducation, usage des soins de santé et dans une moindre mesure présence du père et/ou de la mère. Ceci traduit que l'axe 2 est caractérisé par les variables **socio-démographiques, sanitaires et éducatifs**.

Graphique 2.2 – Graphique des corrélations : variables-axes



Source : Calcul des auteurs.

2.2.3.2 Caractérisation par les modalités

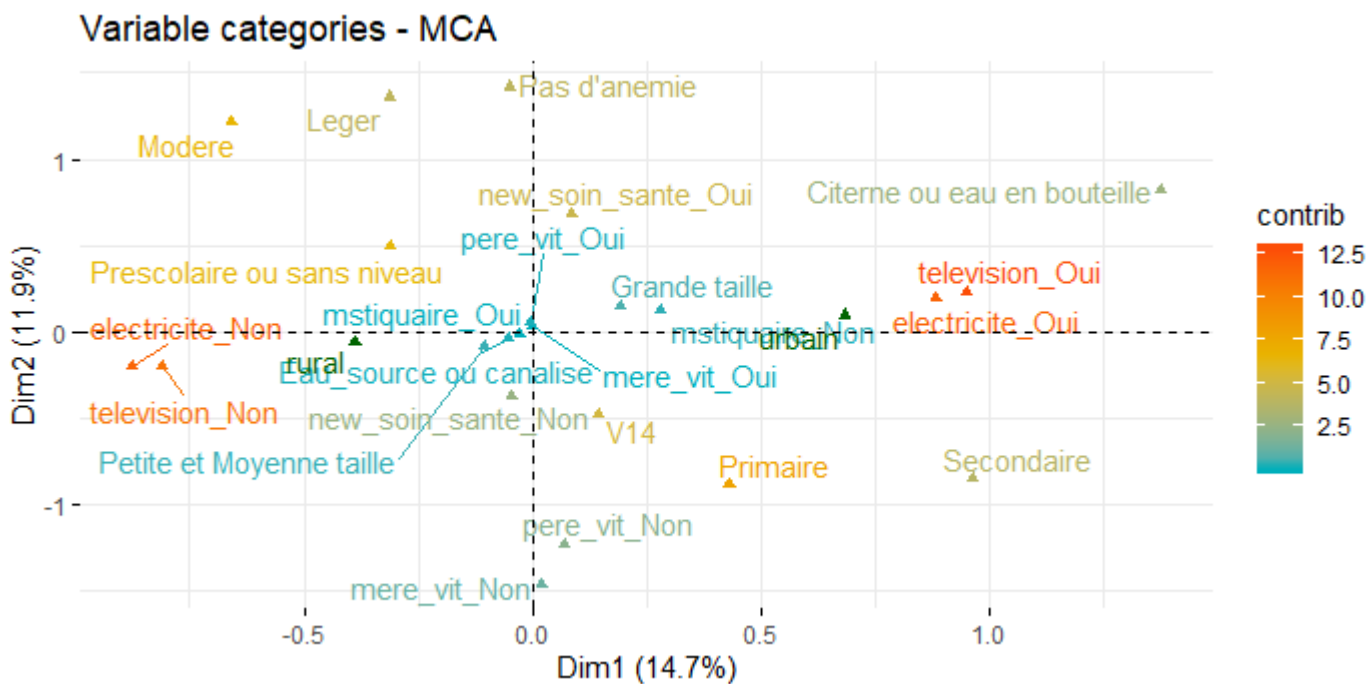
Le graphique ci-dessous présente le nuage des modalités qui apporte plus d'explication que le graphique de corrélation des variables. Avec une coloration en fonction des contributions à la formation des axes, le graphique permet de déceler les oppositions, associations et ressemblances entre les individus.

Sur l'axe 1, les modalités les plus contributives à la formation de l'axe sont : *télévision-oui/non*, *électricité-oui/non*, *anémie-moderée*, *prescolaire ou sans niveau*, *niveau-education-primaire* et enfin *eau-potable-citerne/bouteille*. Ainsi au regard de ces modalités et de leur emplacement vis-à-vis du centre, nous pouvons déduire que l'axe 1 traduit **la dotation économique et le niveau d'étude**. En termes d'association, nous pouvons remarquer

que les enfants privés d'**eau potable**, d'**électricité**, de **télévision**, d'**éducation** et présentant des signes d'**anémie modérée** sont plus fréquents en **milieu rural**.

L'axe 2, caractérisé par les variables socio démographique et sanitaire, est principalement constitué par les modalités pas d'anémie, anémie-modérée, soin-de-santé-oui, niveau d'éducation primaire, père-vit-non, mère-vit-non. L'axe 2 traduit, ainsi l'impact de la présence des parents sur les enfants. En effet, nous constatons une forte association entre les orphelins de père et de mère et que groupe des enfants non orphelins s'associent plus au groupe n'ayant pas d'anémie et bénéficiant de soin de santé, signe d'une meilleure prise en charge par les parents.

Graphique 2.3 – Nuage des modalités de l'ACM

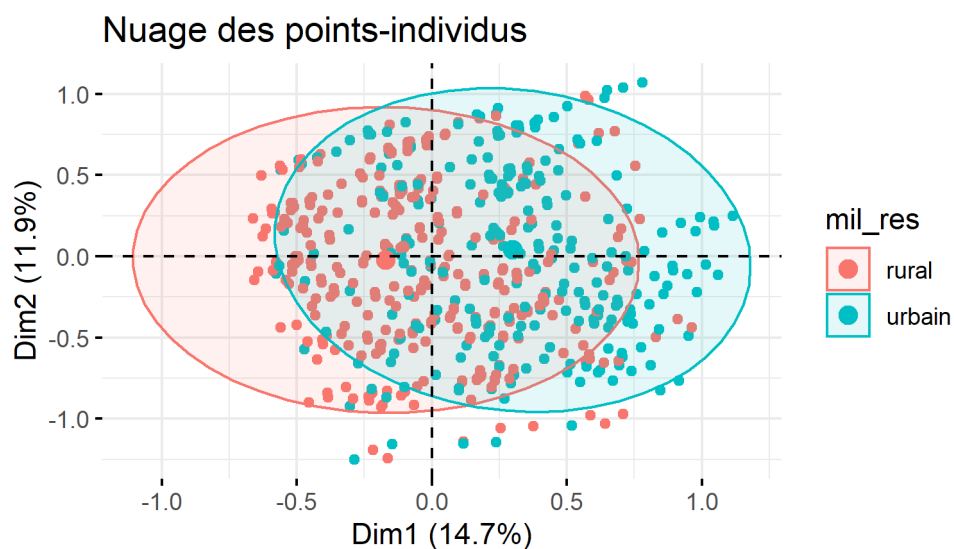


Source : Calcul des auteurs.

2.3 Nuage des individus

Cette partie concerne la présentation des individus sous forme de nuage de points. Nous allons scinder les individus en fonction de la variable supplémentaire qui est le milieu de résidence. La représentation graphique sous forme d'ellipses est la suivante :

Graphique 2.4 – Nuage des points individus



Source : Calcul des auteurs.

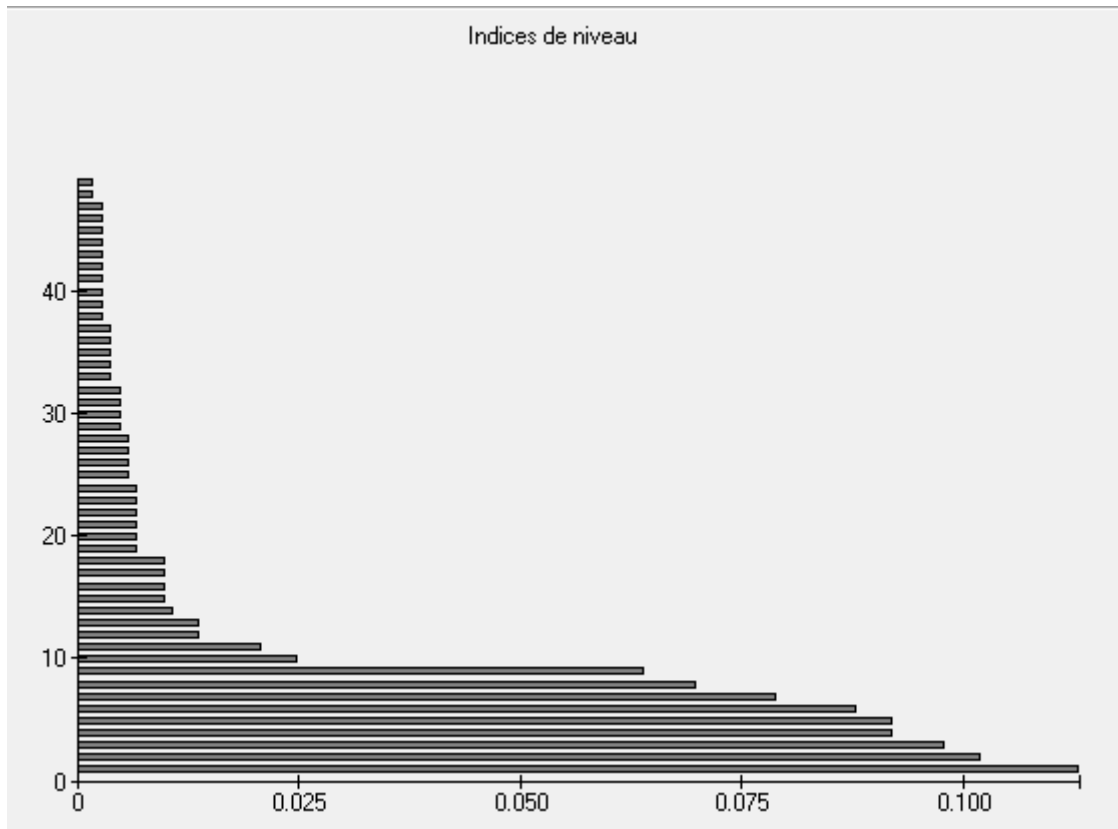
On voit que les individus sont semblables dans un milieu donné et éparpillés d'un milieu à un autre, et ce de part et d'autre de l'axe 1. D'une part, nous remarquons que les populations urbaines sont caractérisées par de fortes valeurs des variables économiques et éducatives telles que les dotations en électricité, en télévision ou en eau potable (citerne ou eau en bouteille) et le niveau d'éducation. D'autre part, nous avons les populations rurales qui ont de faibles valeurs pour ces variables. Ainsi, l'axe1 oppose les individus du milieu urbain et ceux du milieu rural, comme le montre le graphe précédent.

2.4 Classification des individus

2.4.1 Choix des axes

Au terme de l'algorithme de classification hiérarchique, nous obtenons, ci-dessous, l'histogramme des indices de niveau pour les 50 premiers nœuds d'agrégation. L'histogramme montre d'abord une chute assez importante entre le neuvième et le dixième nœud d'agrégation et une autre assez remarquable entre le dernier et l'avant-dernier nœud. Pour des commodités d'analyse nous nous limiterons à la coupure entre le dernier et l'avant-dernier nœud, ce qui a pour effet un partitionnement en deux classes.

Graphique 2.5 – Histogramme des indices de niveau



Source : Calcul des auteurs.

La coupure du dendrogramme entre le dernier et l'avant-dernier nœud occasionne une partition en deux classes qui seront caractérisées dans la suite. Il existe plusieurs outils pour caractériser les classes : caractérisation par les modalités (classe/mod ; mod/classe), par les variables ainsi que par les axes.

Graphique 2.6 – Dendrogramme



Source : Calcul des auteurs.

2.4.2 Caractérisation des classes

La coupure précédemment effectuée aboutit sur une partition de la population en 2 classes décrites et caractérisées ci-dessous. Cette caractérisation se fait en fonction des modalités les plus importantes pour la classe, c'est-à-dire celles ayant une grande valeur-test en valeur absolue pour la classe. Le critère mod/classe qui désigne le nombre d'individus ayant la modalité dans la classe divisée par l'effectif de la classe sera utilisé pour déterminer les caractéristiques des classes.

2.4.2.1 Première classe

La première classe obtenue concentre 22302, soit 54,3% de la population. Le tableau ci-dessous présente les modalités les plus importantes pour la première classe, rangées par ordre décroissant selon la V-test. On remarque que cette classe présente de forte précarité. En effet, 94,57% des enfants de cette classe sont privés de télévision et 88,95% vivent sans électricité. De surcroît, plus de 7 enfants sur 10 sont dans le préscolaire ou n'ont jamais fréquenté l'école et n'ont pas bénéficié de soins santé durant toute l'année. Ainsi, au vu des tendances dégagées ci-dessus, cette classe regroupe les enfants dits **pauvres**.

Tableau 2.2 – Caractéristiques de la première classe

Variables	Modalités	mod/classe	classe/mod	V-Test
television	Non	94,57	94,78	192,55
television	Oui	5,43	4,71	192,52
electricite	Non	88,95	96,92	185,628
electricite	Oui	11,05	8,97	185,609
mil_res	rural	89,99	68,62	112,533
mil_res	urbain	10,01	11,93	112,508
niv_educ	Prescolaire ou sans	74,97	55,38	45,3726
niv_educ	Secondaire	3,45	18,73	34,368
new_soin_sante	Non	69,44	53,27	32,7709
new_soin_sante	Oui	30,56	36,08	32,7614
newsource_eau_boire	Citerne / Bou-teille	1,35	13,04	31,171
newsource_eau_boire	Eau_source ou canal	98,65	48,20	31,1124
niv_educ	Primaire	21,58	35,23	27,813
mstiquaire	Oui	94,43	48,64	25,8056
mstiquaire	Non	5,57	26,66	25,7918
niv_anemie	Modere	14,91	62,75	22,2264
tailleMen	Grande taille	28,78	40,20	17,3374
tailleMen	Petite et Moyenne taille	71,22	49,65	17,3253
niv_anemie	Pas d'anemie	6,58	39,78	7,53449
mere_vit	Oui	97,99	46,39	3,30918
pere_vit	Non	4,74	50,13	2,98503
pere_vit	Oui	95,26	46,34	2,96076

Source : Calcul des auteurs.

2.4.2.2 Deuxième classe

Cette deuxième classe regroupe 45,7% de la population et se présente comme l'opposé de la précédente. En effet, elle compte 95,48% et 97,55% d'enfants respectivement dotés en télévision et en électricité. En sus de cela, 47,06% d'entre eux bénéficient des services de santé. Dans cette les enfants ayant au moins fait la classe primaire font 47,51%. Les enfants de cette classe seront alors considérés comme **riches**.

Tableau 2.3 – Caractéristiques de la deuxième

Variables	Modalités	mod/classe	classe/mod	V-Test
télévision	Oui	95,48	95,29	192,567
télévision	Non	4,53	5,22	192,505
électricité	Oui	97,55	91,04	185,629
électricité	Non	2,46	3,08	185,616
mil_res	urbain	64,23	88,07	112,516
mil_res	rural	35,77	31,37	112,516
niv_educ	Prescolaire ou sans	52,50	44,61	45,3585
niv_educ	Secondaire	13,01	81,26	34,3711
new_soin_sante	Oui	47,06	63,92	32,7696
new_soin_sante	Non	52,95	46,72	32,7577
newsourc_eau boire	Citerne/Bouteille	7,85	86,96	31,1146
newsourc_eau boire	Eau_source/canal	92,15	51,80	31,1146
niv_educ	Primaire	34,50	64,77	27,8194
mstiquaire	Non	13,31	73,34	25,8322
mstiquaire	Oui	86,69	51,36	25,8089
niv_anemie	Modere	7,70	37,25	22,2561
tailleMen	Grande taille	37,21	59,80	17,3226
tailleMen	Petite et Moyenne taille	62,79	50,35	17,3104
niv_anemie	Pas d'anemie	8,65	60,22	7,5767
mere_vit	Oui	98,45	53,61	3,3862
pere_vit	Oui	95,90	53,66	3,00889
pere_vit	Non	4,10	49,87	2,98281

Source : Calcul des auteurs.

2.5 Construction de l'indice composite de la pauvreté infantile

2.5.1 Profil de pauvreté multidimensionnelle

Pour pouvoir calculer l'indice, nous nous sommes rabattus sur l'élaboration du tableau disjonctif complet qui nous permet de recueillir les valeurs prises par les individus pour chaque variable construite en modalités dichotomiques, codées 0 et 1. Rappelons que

l'ICPI est par définition une moyenne pondérée des poids des modalités des différentes variables. Pour le poids, au lieu de prendre les contributions sur les racines des valeurs propres, nous allons élever le rapport au carré car nous avons constaté que les contributions sont tantôt positives, tantôt négatives. De surcroît, pour ramener les valeurs de l'indice à 1, nous avons procédé à une normalisation suivant la formule :

Formule de normalisation

$$\frac{I_i - I_{min}}{I_{max} - I_{min}}$$

Donc, on poursuit l'étude avec l'indice normalisé. Les résultats de ce dernier sont consignés dans le tableau suivant :

Tableau 2.4 – Tableau descriptif de l'ICPI

Moyenne	Ecart-type	Médiane	Coefficient de variation
0,57	0,178	0,549	0,312

Source : Calcul des auteurs.

De ce tableau, il ressort que près de la moitié des enfants se retrouvent dans des ménages ayant un niveau de dotations (ou niveau de vie) évalué par un indicateur inférieur ou égal à la médiane, égale à 0,549. La moyenne égale à 0,570 est supérieure à la médiane, donc on peut dire que distribution est telle que la plupart des enfants ont des niveaux de inférieur à la moyenne. Par ailleurs, la valeur relativement faible du coefficient de variation, soit 31,2%, montre que la distribution de cet indicateur de niveau de vie est homogène autour de la moyenne.

Pour une appréhension du niveau de pauvreté au niveau de chaque région, nous allons nous servir du tableau suivant :

Source : Calcul des auteurs.

On constate que la plupart des régions ont des ICPI inférieurs à la moyenne, soit 8 régions. Cela veut dire que plus de la moitié des enfants du pays sont plus pauvres que d'autres. Il faut noter que l'ICPI le plus faible est détenu par les régions de Kolda (0,495) et Kaffrine (0,485). Par ailleurs, on constate une faible dispersion des enfants pauvres dans le pays, si on se réfère au coefficient de variation. Cela veut dire que les enfants pauvres sont présents dans toutes les régions, et cela traduit une homogénéité plus ou moins vérifiée au niveau de la pauvreté des enfants concernant les quatorze (14) régions du pays.

Tableau 2.5 – Description de l'IPCI selon les régions

Régions	ICPI				
	Moyenne	Ecart-type	Minimum	Maximum	CV
Dakar	0,746	0,0823	0,19	0,996	0,11
Ziguinchor	0,613	0,166	0,792	0,996	0,271
Diourbel	0,659	0,208	0,19	0,996	0,316
Saint-Louis	0,599	0,172	0,115	0,853	0,287
Tambacounda	0,499	0,157	0,045	0,853	0,315
Kaolack	0,57	0,178	0,034	0,996	0,312
Thiès	0,633	0,161	0,034	0,961	0,254
Louga	0,583	0,172	0,034	0,959	0,295
Fatick	0,566	0,179	0,034	0,996	0,316
Kolda	0,495	0,152	0,034	0,853	0,307
Matam	0,549	0,159	0,159	0,079	0,29
Kaffrine	0,485	0,14	0,079	0,8532	0,289
Kédougou	0,527	0,166	0,041	0,959	0,315
Sédhiou	0,491	0,151	0	0,853	0,308

2.5.2 Détermination du seuil de pauvreté

L'objectif principal de la construction d'un indice de pauvreté est de pouvoir définir une valeur de référence telle que tout individu ayant un indice en dessous de cette valeur puisse être considéré comme pauvre et les individus étant au-dessus de cette valeur sont éligibles dans la classe des riches.

Deux méthodes conduisent à la détermination d'un seuil de pauvreté. La première consiste à calculer l'indice à partir des modalités définies comme des commodités de base pour un individu et le seuil sera la moyenne de l'ensemble des individus possédant toutes ces modalités. La seconde approche repose sur la classification des individus en deux classes (riche et pauvre). Selon cette méthode l'indice est calculé selon la formule suivante :

Approche classification

$$S = P_r \cdot \text{Min}(ICPI_{riche}) + P_p \cdot \text{Max}(ICPI_{pauvre})$$



Dans le cadre de notre étude, nous nous sommes référés à la première méthode car cette dernière était moins aléatoire et les modalités choisies sont :

Tableau 2.6 – Commodité de base

1	L'enfant vit dans un ménage électrifié
2	Dort sous une moustiquaire
3	A accès à l'eau potable
4	A accès au soin de santé
5	A accès à une télévision (pour le loisir par exemple)

Source : Calcul des auteurs.

Ainsi établie, cette méthode donne un seuil de pauvreté de **0,7**. Cela veut donc dire que tout enfant ayant un indice de pauvreté inférieur à **0,7** est alors considéré comme pauvre. Par suite nous obtenons un taux de pauvreté de 69,29%.

Tableau 2.7 – Taux de pauvreté

Taux de pauvreté	
IPCI normalisé	Fréquence
Pauvre	69,29%
Riche	30,71%
Total	100%

Source : Calcul des auteurs.

On constate que le seuil défini par l'indice apporte une légère correction sur les effectifs des classes obtenues plus haut. En effet, la classe des pauvres s'élargit et passe de 54,3% à 69,29% et celle des riches se rétracte en passant de 45,7% à 30,71%.

Conclusion

Ce travail nous a permis de mettre en pratique les notions théoriques acquises au cours d'ADD. Ainsi, nous avons pu parcourir une bonne partie des méthodes d'analyse multidimensionnelle des données, allant de l'ACM qui est une méthode factorielle à la classification. Par ailleurs, cela nous a permis d'avoir un aperçu de la pauvreté des enfants au Sénégal à travers l'Indice Composite de la Pauvreté Infantile (ICPI).

S'agissant des tâches effectuées, nous avons eu à scinder le travail en deux parties majeures : une partie méthodologique et une partie analytique. Dans la première, il a d'abord été question de présenter la source de données et de parler du procédé de choix des individus devant représenter notre population-cible. Ensuite, nous avons eu à décrire la méthodologie à adopter au niveau de l'analyse des données, à savoir l'Analyse des Correspondances Multiples (ACM), la classification, plus particulièrement la classification ascendante hiérarchique. Enfin, nous avons étalé la procédure de calcul de l'ICPI basée sur l'ACM. La seconde partie, quant à elle, a été réservée aux résultats de l'analyse. Entre autres, nous avons eu à construire les nuages de modalités et d'individus, pour l'ACM. Par la suite, le recours à la classification, qui vient après l'ACM, nous a permis de classer les individus, de déterminer le nombre de partitions optimales et de caractériser les classes par les variables. Et pour finir, nous avons aussi donné les résultats du calcul de l'indice tout en présentant le seuil de pauvreté. Ce dernier a été calculé d'une part sur la base de certaines commodités jugées élémentaires, et d'autre part grâce aux résultats de la classification ascendante hiérarchique.

Toutefois, il faut préciser que l'analyse des données se verrait donner une étude plus puissante si elle était associée à une étude économétrique sur les variables qualitatives pour mieux expliquer certains phénomènes ou certaines interactions.