THESIS – SS 142501

POPULATION ANALYSIS OF DISABLED CHILDREN BY DEPARTMENTS IN FRANCE

DIAH MEIDATUZZAHRA
NRP 1314 201 003

ADVISORS
Nicolas Pech
Amélie Etchegaray

MASTER PROGRAM
DEPARTMENT OF STATISTICS
FACULTY OF MATHEMATICS AND NATURAL SCIENCES
SEPULUH NOPEMBER INSTITUTE OF TECHNOLOGY
SURABAYA
2017

ANALYSE DE LA POPULATION DES ENFANTS HANDICAPE PAR DEPARTEMENT EN FRANCE

Réalisé par : **Diah MEIDATUZZAHRA**

Encadrant CREAI : **Amélie ETCHEGARAY**

Encadrant universitaire : **Nicolas PECH**

Rapport de stage

M2 MASS POP

Aix-Marseille Université

2015 /2016

DOCUMENT FOR SEPULUH NOPEMBER INSTITUTE OF TECHNOLOGY (ITS), INDONESIA

This document is research report held in Aix-Marseille University (France).
This report collection aims to meet the master graduation in ITS (Indonesia).

POPULATION ANALYSIS OF DISABLED CHILDREN BY DEPARTMENTS IN FRANCE

In partial fulfillment of the requirements for the degree of Master of Sciences
In
Sepuluh Nopember Institute of Technology

By :
DIAH MEIDATUZZAHRA
NRP 1314 201 003

Date : September 2016
Graduation Period : March 2017

Program Committee :
1. Nicolas Pech
2. Amélie Etchegaray

Approved by :
Director of Graduate Program
Sepuluh Nopember Institute of Technology

Prof. Ir. Djauhar Manfaat, M.Sc., Ph.D.
NIP.19601202 198701 1 001

an. Direktur Program Pascasarjana
Asisten Direktur

Prof. Dr. Ir. Tri Widjaja, M.Eng.
NIP. 19611021 198603 1 001

POPULATION ANALYSIS OF DISABLED CHILDREN BY DEPARTMENTS IN FRANCE

Name : Diah Meidatuzzahra

NRP : 1314 201 003

Advisors : Nicolas Pech
Amélie Etchegaray

ABSTRACT

In this study, a statistical analysis is performed by model the variations of the disabled about 0-19 years old population among French departments. The aim is to classify the departments according to their profile determinants (socioeconomic and behavioural profiles). The analysis is focused on two types of methods: principal component analysis (PCA) and multiple correspondences factorial analysis (MCA) to review which one is the best methods for interpretation of the correlation between the determinants of disability (independent variable). The hierarchical cluster analysis (HCA) can be used to classify the departments according to their profile determinants. Analysis of variance or ANOVA is performed to know difference the between cluster and within cluster variances of two proxy data (AEEH and EN3-EN12). The PCA reduces 14 determinants of disability to 4 axes, keeps 80% of total information, and classifies them into 7 clusters. The MCA reduces the determinants to 3 axes, retains only 30% of information, and classifies them into 4 clusters. The ANOVA of the proxy data by department cluster are difference significant between cluster and the variance within of cluster is not difference significant, the cluster are homogeneous.

Keywords : Disability of Children, Principal component analysis, Multiple Coresspondences Analysis, Hierarchical Cluster Analysis, Analysis of Variance.

ACKNOWLEDGEMENTS

In the name of Allah SWT, the beneficent and merciful. All praise is merely to The Mightiest Allah SWT, for the gracious mercy and tremendous blessing that enable me to accomplish this script. This research report entitled "Population Analysis of Disabled Children by Departments in France", is submitted to fulfill one of the requirements in accomplishing the S-2 Degree Program at the Department of Statistics, Faculty of Mathematics and Natural Sciences, Sepuluh Nopember Institue of Technology, Surabaya (ITS).

There are many individuals who have generously suggested to improve this research report. First of all the writer would like to express his sincere gratitude and respect to his first advisor, Nicolas Pech and Amélie Atchegaray, who have contributed and given their valuable evaluations, comments, and suggestions during the completion and accomplishing of this research report. The writer also would like to express his deepest gratitude and respect to all professor of MASS Aix-Marseille Univercity (AMU) and the associate of CREAI PACA et Corse. And then, to coordinator of master program in statistics ITS, Dr.rer.pol Heri Kuswanto, thank you for supporting me to study abroad.

Last but not least, special gratitude and indebtedness are dedicated to beloved mother, Suhartini and beloved father, M. Syuhairi, and also all of family, who always give their loves, prayers, supports, and encouragements for every single path the writer chooses. Special thankfullness is also due to beloved friends, Statistics ITS 2014, the colleagues of MASS POP AMU 2015, and PPI (Perhimpunan Pelajar Indonesia) in Marseille, for their loves, prayers and supports.

Hopefully, this script would give a positive contribution to the educational development or those who want to carry out further research.

Surabaya, January 2017

Author

TABLE OF CONTENTS

LIST OF TABLES

LIST OF DIAGRAM

LIST OF ATTACHEMENTS

I. INTRODUCTION

The term “disability” is defined as a limitation of a person's ability to interact with their environment, due to a permanent disability or non-permanent that leads to stress and moral disorder, intellectual, physical or social. Disability is the consequence of an impairment that may be physical, cognitive, mental, sensory, emotional, developmental, or some combination of these. A disability may be present from birth, or occur during a person's lifetime. Disability comes in multiple forms and ambiguous definition. It is to be distinguished from the disease or the accident, which can be the disability origin.

In France, the definition disability is governed by French law dated 11 February 2005 on the opportunities and the same rights, and the participation and citizenship of disability people, that “Disability of activity limitations or restrictions on participation in the social life suffered by a person, due to substantial continuing modification of one or more functions”.

One of the institutions cared for the disabled is CREAI branch PACA et Corse. CREAI collaborated with Population Environment Development Laboratory (LPED) Aix-Marseille University propounds a project "Geography of Disability". The purpose is to estimate of the population and establishment medical social service. Therefore, statistical analysis is required to estimate the disabled population of multiple databases which were defined as disability determinants.

In this study, statistical analysis performed to model the variations of disabled children about 0-19 years old population among French department consisting of 14 determinant variables of children disabilities based on six categories, namely, the professional category of social (CSP) of their parents, the level education of their parents, the premature rates, the tax of revenues, alcohol consumption, facilities and services of medical social for disabled. The aim is to clusterify departments according to their profiles determinants (socioeconomic and behavioral profiles). Two types applied of methods: principal components analysis (PCA) and multiple correspondence analysis (MCA) to review which one

is the best methods for the interpretation of the correlation between the determinants of disability (independent variable). And then, hierarchical clustering can be used to clusterify the departments according to their profile determinants. After that, analysis of variance or ANOVA is performed to know the between cluster and within cluster variances of two proxy data (AEEH and EN3-EN12).

II. CONTEXT AND PRESENTATION of DATA

2.1 Centre inter-Régional d'Etudes, d'Action et d'Information (CREAI)

CREAI PACA et Corse or Central interregional of studies, action and information branch Provence-Alpes-Côte d'Azur (PACA) and Corse in France was founded in 1965 and belongs to the National Association of CREAI (ANCREAI) for a person with the condition of vulnerability. The CREAI is a private organizations and non-profit status established by statute law in 1901 which is subsidized by the State to optimize information sharing, collaboration and develop synergies of the technical experts to reflection and observation in the sectors of social action and medico-social.

The main tasks entrusted to the Creai include:

- Observation of the needs and expectations of populations
- The carrying out of studies and observations on the specific phenomena centered of disability (Regional Health Agency), in the region or the departments
- Technical expertise through the internal evaluation of actions (accompanying the internal evaluation of institutions and medico-social services through training actions)
- Training professionals on topics such as violence, wellness, personalized project
- Animation notably to facilitate exchanges between the actors concerned by a thematic, but also in order to improve collaborations and encourage innovations
- Publication of publications, studies and work on their website.

The team names of CREAI workers:

DIRECTION	- Serge DAVIN (President) - Dr Monique PITEAU-DELORD (Directress)
Studies-Observations-Expertise	- Sophie BOURGAREL (Technical Advisor) - Céline MARIVAL (Technical Advisor) - Amélie ETCHEGARAY (Technical Advisor) - Philippe PITAUD (Scientific
FORMATION	- Hélène CATTANEO - Emilie GIRARD (Secretariat of the training center)
ADMINISTRATION	- Christiane CHAZOT (Executive Management) - Agnès DESBIEF-BLANJOUE (Accounting)
DOCUMENTATION, COMMUNICATION	- Patricia FIORENTIN (Documentation) - Thomas ROSSELET (Communications Officer)
INFORMATICS SERVICE	- Benjamin CAYRE (Computer scientist)

2.2 The Project "Geography of the disabled population"

CREAI collaborated with Population Environment Development Laboratory (LPED) of Aix-Marseille University propounds a project "Geography

of Disability" between 0 and 59 years by 96 departments in France. The purpose is to estimate of the population and establishment medical social service.

2.3 The Determinants of Disability

The hypothesis of the project "Geography of Disability" is the distribution unusual of the disabled population in the region. The distribution of this population related by multiple factors, for example, economic factors, education, environment, lifestyle, etc., called the determinants of disability. Six groups of determinants identified:

2.3.1 The professional category of social (CSP)

The CSP based on the data of INSEE. Labor force of 15 years and more having a job by gender, age, and the CSP are divided into 6 positions, is show on table 2.1.

Table 2.1. The position of CSP

Number	The professional category of social
1.	Farmer
2.	Artisan, craftsman and trader
3.	Manager and high professions
4.	Intermediate professions
5.	Employee
6.	Labor

2.3.2 The education level of their parents

HSM survey is identified 33% of the disabled population aged 20-59 years is not graduation (without diplome) (Espagnacq, 2015). The database is used BTX_TD_FOR2_2012 of INSEE. It is clusterify 4 of education level, is show on table 2.2.

Table 2.2. The clusterify of education level

Number	The level of education
1.	Without diplome
2.	BEPC / BEP / CAP (Diplome or certificate of professional)
3.	BAC (Baccalaureat)
4.	BAC^{+}

2.3.3 The Premature Rates

The EPIPAGE study showed the importance of disabling sequelae preterm infants, before 33 weeks of amenorrhea (WA), and among those born between 33 and 36 WA of age. According EPIPAGE, if the preterm birth is increased, the risk of disability is also. The 8th day certificate (Cs8) data used as database from 2010 to 2012.

2.3.4 The Tax of Revenues

The report of revenues tax is derived from local INSEE Social and Tax File (Philosophy) data. The first quartile of income report is the average wage in the department below which is 25% of wages (CREAI,2010).

2.3.5 The Consumption of Alcohol

The consumption of alcohol among women is unknown, therefore the number of premature deaths due to overdose of alcohol (it cause alcoholic psychoses and alcoholic cirrhosis of the liver) in women under the 65 years old were selected.

2.3.6 The facilities and services of medical social for disabled

The facilities and services of medical social for 1000 children disabled from age 0-19 years old. According the CREAI data (2016), Paris and PACA region are the least equipped regions mainly located (below 8 places for 1000 children). The Lozère is the most departments equipped in medico-social places with 22 places for

1000 children and also the Orne (18 places for 1000 children) and the Creuse (17 places for 1000 children).

2.4 The data of Proxy

The proxy data is concern and count of the population with a disability in France. Two databases are considered to be proxy data for children, namely the number of beneficiaries of the Allowance for the Education of Handicapped Children (AEEH) and the number of disabled children enrolled in national education (EN3-12).

2.4.1 The data of the Allowance for the Education of Handicapped Children (AEEH)

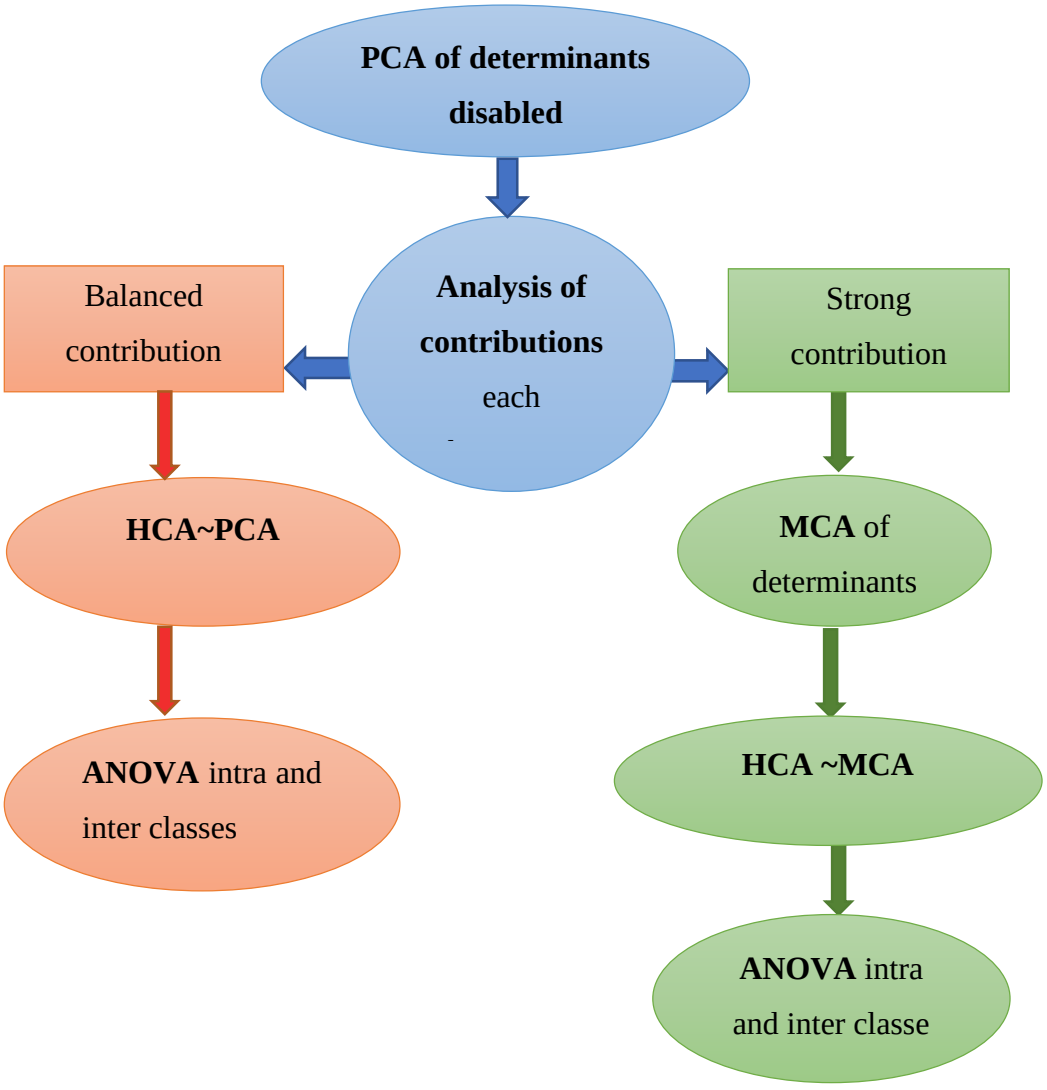
The AEEH is designed to help parents who assume the responsibilities of disabled children, regardless of their resources. It is awarded to families with disabled children who have a disability rate recognized by the Commission for the Rights and Autonomy of Persons with Disabilities (CDAPH) within the Departmental Houses for Persons with Disabilities (MDPH). According to the socioprofessional category of the household, the AEEH can be paid by the Caisse d'Allocation Familiale (CAF), the Mutuelles Sociale Agricole (MSA) and the Régime Social des Indépendants (RSI).

2.4.2 Survey No. 3 and 12 on the enrollment of pupils with disabilities in primary and secondary education (EN 3-12)

The Ministry of Education has set up surveys No. 3 and 12 on the enrollment of pupils with disabilities in primary and secondary education (Public and private) for a thorough knowledge of pupils with disabilities. The aim is to set up the schooling policy of disabled children and adolescents. These surveys are carried out annually with the Directorate for Evaluation, Foresight and Performance DEPP (Office for Student Statistical Studies) and the General Directorate for School Education DGESCO (Office for the personalization of school and the schooling of handicapped pupils).

III. TRAITEMENT OF STATISTICS

The aim of this section is to determine, using Principal Component Analysis (PCA) and hierarchical ascending clusterification (HCA), groups of departments with the same determinant profiles (socio-economic and behavioral profiles) And to know the « between » and « within » of cluster variances on the proxy by analysis of variances (ANOVA). If certain departments have too much contribution, it will be necessary to go through a factor analysis of multiple correspondences (MCA) and apply the same methodology (HCA and ANOVA).



3.1 Principal Components Analysis (PCA)

A principal component analysis (PCA) is concerned with explaining the variance-covariance structure of a set of variables through a few linear combinations of these variables. PCA is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. Its general objectives are reduction and interpretation of dimension (axes) data without reducing significantly the characteristics of the data. PCA is also often used to avoid problems of multicollinearity between independent variables in a multiple regression model.

The number of principal components is less than or equal to the number of original variables. This transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal to the preceding components. The resulting vectors are an uncorrelated orthogonal basis set. PCA is sensitive to the relative scaling of the original variables.

As part of this process, the PCA is involved in the interpretation of the relationship between the determinants of disability, interdependent variables. Its main purpose is to condense the information given by the determinants into a smaller number of independent fundamental variables that can not be directly observed.

3.1.1 The Correlation matrix

The table 3.1 already identifies the most interrelated determinants of disability. The highest correlation is between the people who works as a manager and the rate of people with a diploma above the bac (r = 0.96). The rate of labor is correlated with the rate of persons without a diploma or BEPC, CAP or BEP (r = 0.79). Conversely, the rate of people with a BEPC, CAP or BEP is correlated negative with the rate of professional categories social which manager (r = -0.89).

Table 3.1 Correlation matrix

	Taux_alcool	taux_prematurite	Taux_ESMS	Tx_sans_dipl	Tx_BEPC	Tx_BAC	Tx_BAC_PLUS	Taux_agri	Taux_artisans	Taux_cadres	Taux_prof_int	Taux_employes	Taux_ouvriers	Taux_revenus
Taux_alcool	1.0000	0.2768	0.1678	0.3372	0.2880	-.2966	-.3144	0.0037	-.2240	-.2418	-.1125	0.2250	0.3499	-.2301
taux_prematurite	0.2768	1.0000	0.0496	0.0938	-.1033	-.3036	0.0664	-.2887	-.5931	0.1750	0.1319	-.1374	0.1514	0.0421
Taux_ESMS	0.1678	0.0496	1.0000	0.3033	0.4178	-.1877	-.3909	0.5480	0.0903	-.4190	-.3210	0.1852	0.3021	-.1466
Tx_sans_dipl	0.3372	0.0938	0.3033	1.0000	0.5592	-.4793	-.8534	0.3876	0.0170	-.7204	-.5871	0.5047	0.7772	-.6076
Tx_BEPC	0.2880	-.1033	0.4178	0.5592	1.0000	-.2280	-.8778	0.4529	0.2013	-.8926	-.3180	0.4928	0.7919	-.1333
Tx_BAC	-.2966	-.3036	-.1877	-.4793	-.2280	1.0000	0.2145	0.0102	0.5840	0.0426	0.2669	0.2904	-.5506	-.0354
Tx_BAC_PLUS	-.3144	0.0664	-.3909	-.8534	-.8778	0.2145	1.0000	-.5091	-.2558	0.9614	0.4984	-.6694	-.8272	0.4737
Taux_agri	0.0037	-.2887	0.5480	0.3876	0.4529	0.0102	-.5091	1.0000	0.4433	-.5650	-.6740	0.2092	0.2759	-.2506
Taux_artisans	-.2240	-.5931	0.0903	0.0170	0.2013	0.5840	-.2558	0.4433	1.0000	-.4143	-.2914	0.4472	-.1484	-.3372
Taux_cadres	-.2418	0.1750	-.4190	-.7204	-.8926	0.0426	0.9614	-.5650	-.4143	1.0000	0.4238	-.6789	-.7516	0.4435
Taux_prof_int	-.1125	0.1319	-.3210	-.5871	-.3180	0.2669	0.4984	-.6740	-.2914	0.4238	1.0000	-.2428	-.4362	0.3848
Taux_employes	0.2250	-.1374	0.1852	0.5047	0.4928	0.2904	-.6694	0.2092	0.4472	-.6789	-.2428	1.0000	0.2485	-.6154
Taux_ouvriers	0.3499	0.1514	0.3021	0.7772	0.7919	-.5506	-.8272	0.2759	-.1484	-.7516	-.4362	0.2485	1.0000	-.2051
Taux_revenus	-.2301	0.0421	-.1466	-.6076	-.1333	-.0354	0.4737	-.2506	-.3372	0.4435	0.3848	-.6154	-.2051	1.0000

3.1.2 The Eigenvalue of the Correlation Matrix

Table 3.2 and diagram 3.1 below provide the inertia of each axis. By looking at the eigenvalues and especially that greater than 1 then axis is selected four. The first axis contains 42% of the diversity from the original data with eigenvalue is 5.95. The second axis allows to restore 19% of the total inertia with eigenvalue is 2.68. The third axis contains 10% of the information and the fourth 8%. These four axis thus make it possible to retain 80% of the information (Rule of Kaiser).

Table 3.2 The eigenvalues of the correlation matrix

	Eigenvalues	Difference	Proportion	Cumulative
1	5.94759063	3.26410811	0.4248	0.4248
2	2.68348252	1.24735361	0.1917	0.6165
3	1.43612891	0.30611838	0.1026	0.7179
4	1.13001053	0.22447436	0.0807	0.7998

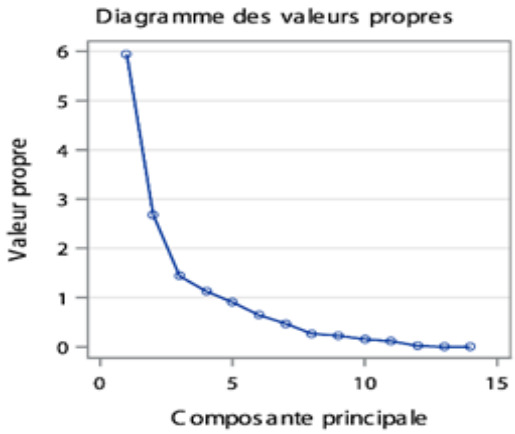


Diagram 3.1 Eigenvalue

3.1.3 Variables Results

Diagram 2 allows analyzing the contribution of the variables to the axe$_k$ ($Prin_k$). It is comparing the absolute value of variables with:

$$\frac{1}{\sqrt{P}} \tag{3.1}$$

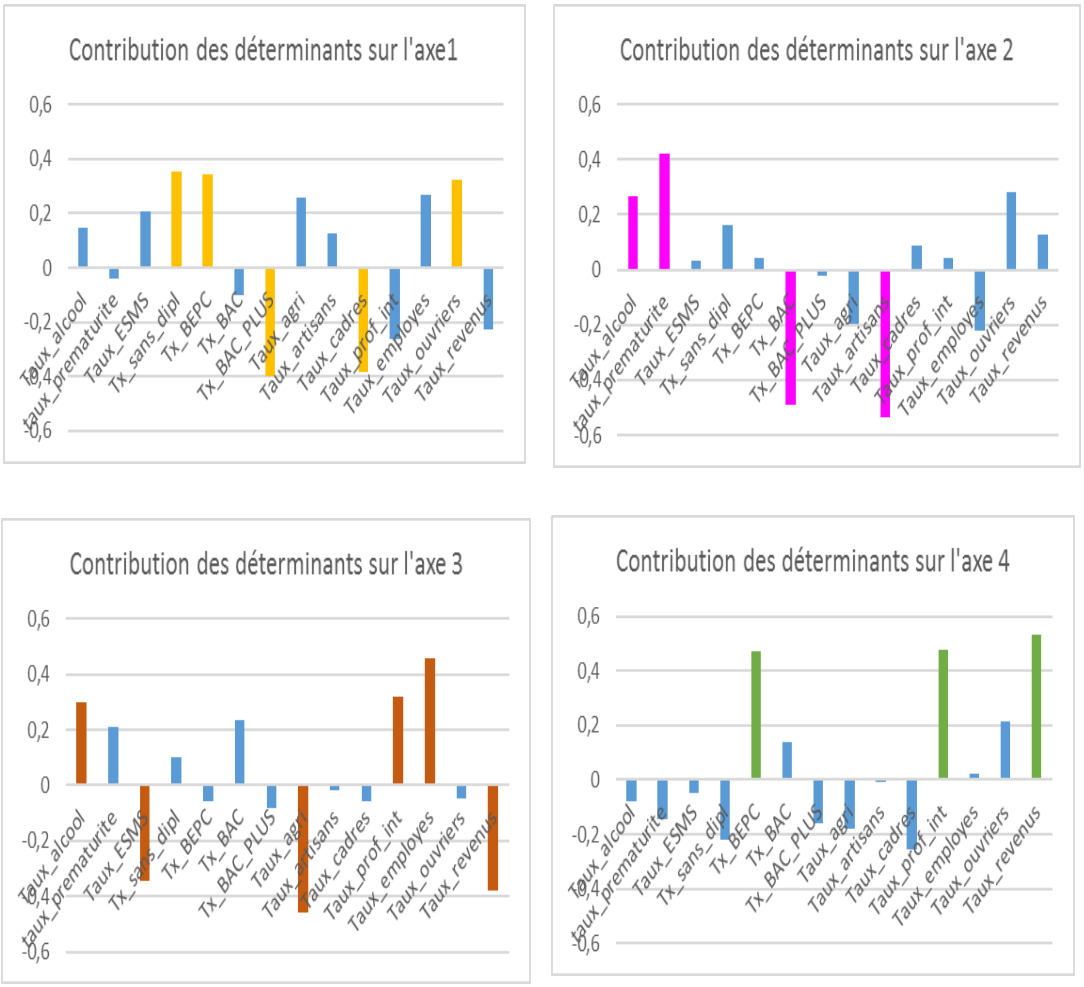
Where , $p = 14$

The higher values corresponded to the variables which most contribute formation of the axis. The positive or negative symbol is showed high or low contribution of variable.

The first axis is correlated positive with the labor variable and the level of education without a diploma, BEPC, CAP or BEP. Conversely, it is negative correlation with the manager variable as well as the level of education with a diploma above the bac. The second axis is contrasts with the rate of artisans-craftsmen-traders and the education level of people with the baccalaureate and the rate of preterm, the rate of consumption of alcohol and the rate of workers.

The third axis is the rate of alcohol consumption, the rate of intermediate professions and the rate of employees with the rate of farmers, the rate of facilities and services of medical social and the tax of revenues. The fourth axis is positive correlation with the education level of people with BEPC, CAP or BEP, the rate of intermediate professions and the tax of revenues.

Diagram 3.2 The contribution of determinants



3.1.4 The Departments Results

The ACP also calculated the coordinates of the individuals on the axes and their contributions to the dispersion according to each of these axes with the formula :

$$\frac{(prin_{ki})^2}{n \times \lambda_k} \tag{3.2}$$

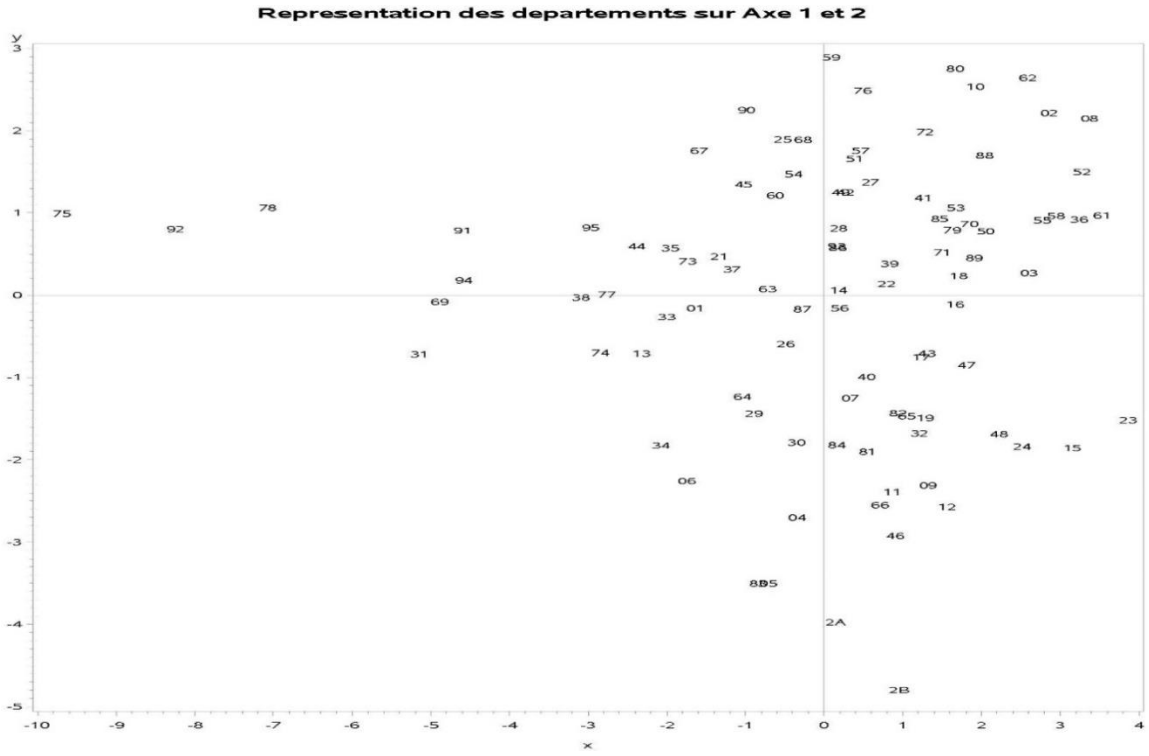
Where, λ_k = eigenvalue

The contributions of the departments according to axes 1, 2, 3 or 4 were calculated. If the contribution of the departments was homogeneous then this would be 1.04% (100/96). The axe 1 is coordinates -9.68 for Paris (75) and +3.86

for the Creuse (23). The departments of Haute Garonne (31), Rhone (69), Hauts de Seine (92), Yvelines (78), Val de Marne (94) and Essonne (91) have coordinates less than -4.

Five departments have large contributions 4% and account for almost 46% of the variance: They are predominant in the definition of axis 1. The 5 departments are Paris (75) with 16% contribution, Hauts de Seines (92) 12%, Les Yvelines (78) 9%, Haute Garonne (31) 5% and the Rhone (69) 4%.

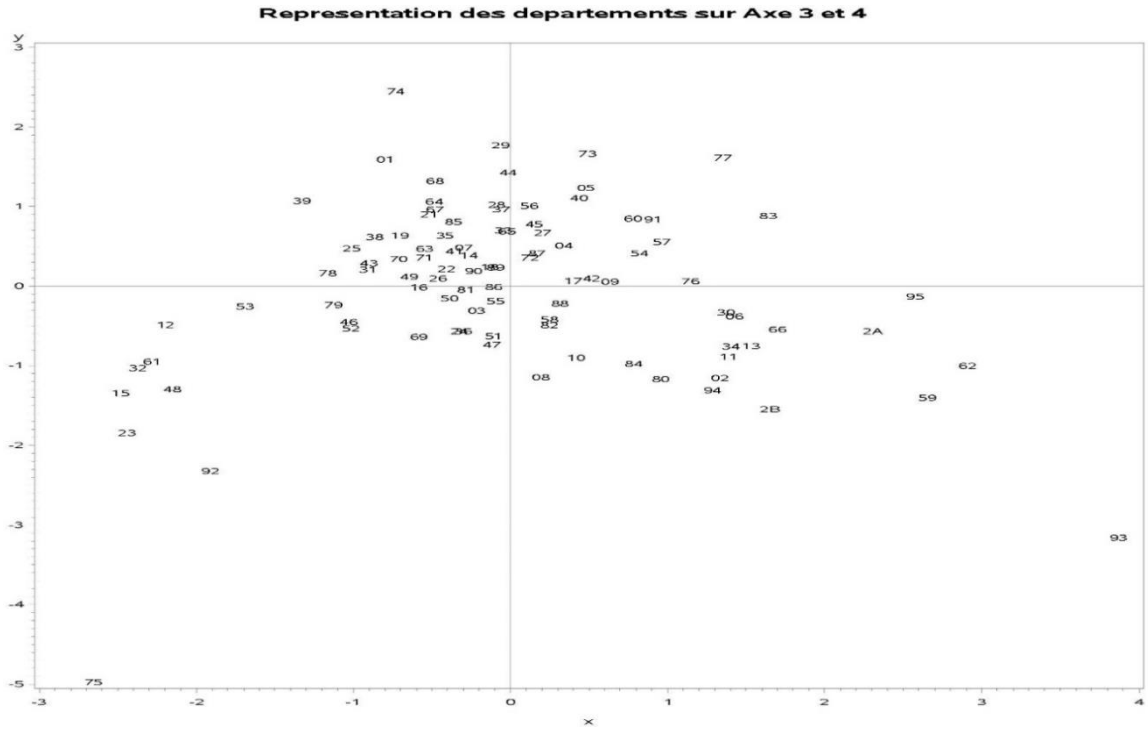
Diagram 3.3 Representation of departments on axis 1and 2



In axes 3 and 4, the variations in the coordinates of the departments are poor. Coordinates ranged from -2.65 from Paris (75) to 3.87 Seine Saint Denis (93) for axis3 and -4.95 from Paris (75) to 2.47 Haute-Savoie (74) for axe4. In the axe 3, 44% of the variance is held by 8 départements: Les Hauts de Seine (92) with 11%, Pas de Calais (62) with 6%, Nord (59), Paris Val d'Oise (95) with 5%, and Cantal (15), Creuse (23) and Gers (32) with 4%. In the axis4, 43% of the variance is held by 4 départements: Paris (75) with 23%, La Seine Saint Denis

(93) with 9%; The Haute Savoie (74) with 6% and the Hauts de Seine (92) with 5%. Details can be found in the diagram below.

Diagram 3.4 Representation of departments on axis 3 and 4



3.1.5 Conclusion of PCA

The PCA reduces 14 determinants of disability to 4 axes, keeps 80% of total information. PCA is also calculating the coordinates of the department and the contribution of the determinants variables per axis. It is not balanced because some departments have too strong contribution.

3.2 The Balanced Contribution of departments

3.2.1 Hierarchical Cluster Analysis (HCA) of the departments using the axes of the PCA

A hierarchical cluster analysis method is a procedure which represents the data as a nested sequence of partitions. An example of the corresponding graphical representation, called a dendrogram. It is important to note that the height of a node is proportional to the distance between groups it links. Consequently, the shape of a dendrogram gives information on the number of clusters in a data-set. Thus, cutting a dendrogram horizontally engineers a clustering (7 clusters appear in the diagram 3.5). Numerous methods have been proposed to determine the best cutting point, to automatically find the number of clusters [mil88].

A. The choice of the number of clusters

The number of clusters is selected by three criteria. The diagram below represents the Cubic Clastering Criterion (CCC), pseudo F and pseudo t2 as a function of the number of clusters.

- The Cubic Clustering Criterion (CCC)

  CCC values greater than 2 indicate good classification. The peak in the CCC between 0 and 2 indicate is a possible classification.

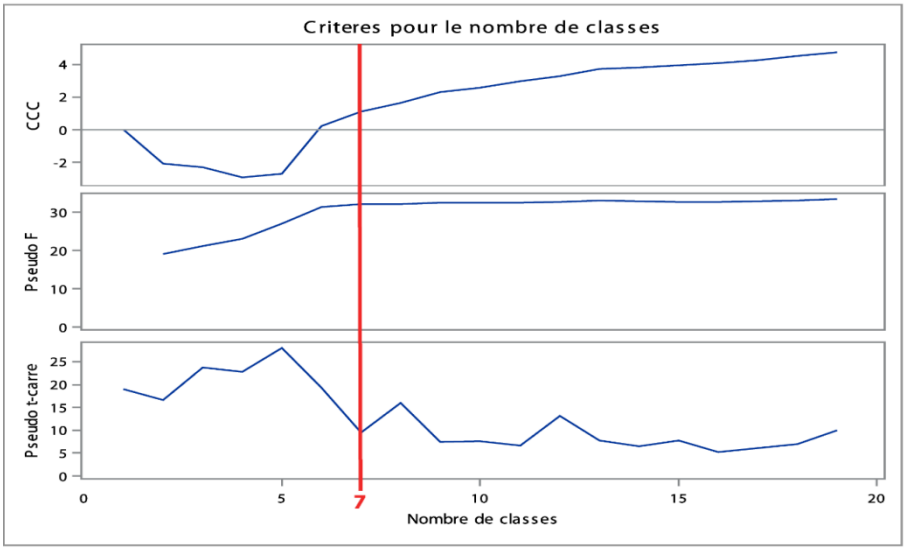
- The Pseudo F

  As a general rule, the higher this statistic, the better the score.

- The pseudo t^2

The pseudo t^2 must be weak and followed by a strong t^2 at the following aggregation.

Diagram 3.5 The criteria of cluster number (PCA)



The number of clusters is must satisfied above the criteria. The seven clusters are selected of departments.

b. The Characterization and localization of clusters

The table 3.3 below is showed the average of axis in each cluster.

Table 3.3 The average of axis

	classe						
	1	2	3	4	5	6	7
	moyenne	moyenne	moyenne	moyenne	moyenne	moyenne	moyenne
Prin1	1.40	-2.88	-0.45	0.04	2.16	1.82	-8.96
Prin2	-0.05	0.06	-2.63	1.65	-2.02	2.28	0.92
Prin3	-0.38	0.00	1.34	0.13	-2.11	1.77	-2.28
Prin4	0.14	0.83	-0.32	0.46	-1.05	-1.40	-3.63

The characterization of the seven clusters is as follows:

- Cluster 1

Cluster 1 is consists mainly of axis 1 of the ACP, that the rate of persons without a diploma or a BEPC, CAP or BEP level and the rate of workers are important. It is the cluster with the most departments, 35 departments are spread all of France. The Côtes d'Armor, Morbihan, Calvados, Manche, Orne, Mayenne, Meuse, Haute Marne, Haute Saône, Jura and Saône rivers. Loire, Yonne, Nièvre, Allier, Cher, Loir et Cher, Indre, Vienne, Deux Sèvres, Vendée, Charente

Maritime, Charente, The Haute Vienne, the Dordogne, the Corrèze, the Puy de Dôme, the Haute Loire, the Ardèche, the Drome, the Landes, Lot and Garonne, Tarn, Tarn et Garonne Pyrenees and the Ariège.

- Cluster 2

Cluster 2 is composed mainly of the inverse of axis 1 of the PCA, i.e. a high rate of executives as well as a rate of people with a diploma superior to the bac. The positive average 0.83 of the axis 4 corresponds to a population having mainly as a level of education the patent, CAP or BEP, a high level of intermediate professions and a first quartile of the median income. This cluster regroups 18 départements spread all over France. These include Ain, Côte-d'Or, Finistère, Haute-Garonne, Gironde, Ille-et-Vilaine, Indre-et-Loire, Isère, Loire-Atlantique, Pyrénées-Atlantiques, Rhône, Savoie, Haute-Savoie, Seine-et-Marne, Yvelines, Essonne, Val-de-Marne and Val-d Oise.

- Cluster 3

Cluster 3 includes all the departments of the Mediterranean arc (Hautes Alpes, Alpes-de-Haute-Provence, Alpes-Maritimes, Vars, Bouches-du-Rhône, Vaucluse, Gard, Hérault, Aude, Pyrénées-Orientales Corsica "Corse-du-Sud and Haute-Corse"). They are characterized by a negative mean for axis 2. This means that these are departments composed of people with a tray level and a craftsman status. The mean of 1.34 of axis 3 means that this region is also characterized by a high level of intermediate and employee professions and alcohol consumption.

- Cluster 4

Cluster 4 is characterized by axis 2 of the PCA. These are regions with a high rate of premature births, premature deaths linked to alcohol and also workers. 16 departments make up this cluster and are located in the northern half of France. They are on the outskirts of the Parisian crown (the Oise, the Eure, the Seine Maritime, the Eure and the Loire, the Loiret and the Marne) to the west (Sarthe

and Maine et Loire) East of France (the Meurthe et Moselle, the Lower Rhine and the Upper Rhine, the Vosges, the Territoire de Belfort and the Doubs).

- Cluster 5

Cluster 5 consists mainly of axis 1, that is, a population without a diploma or a BEPC, CAP, BEP or a baccalauréat. The dominant CSPs are the working clusters. These departments are mainly located in the Massif Central: Aveyron, Cantal, Creuse, Lot and Lozère. The Gers is also attached to this cluster.

- Cluster 6

Cluster 6 is the cluster of workers, employees and intermediate professions who do not have a diploma or the BEPC, CAP or BEP. The rate of premature deaths related to alcohol is high as premature births. The departments are located in the Nord Pas de Calais region and its surroundings (Somme, Aisne and Ardenne and Aube). The Seine Saint Denis also belongs to this cluster.

- Cluster 7

Cluster 7 consists of only two departments: Paris and the Hauts de Seine. These are departments represented by an extreme average of the inverse of axis 1: a population with a level of education higher than the bac and a socio-professional category of executives.

B. Conclusion Ascending Hierarchical Clustering (HCA)

The Hierarchical Ascending Clustering (HCA) issued by the PCA made it possible to group the departments into 7 clusters. With 35 departments, cluster 1 is the cluster with the most individuals. On the contrary, cluster 7 contains only two.

The number of clusters resulting from the HCA is therefore not homogeneous. This may be due to the fact that the contributions of the departments to the PCA are not balanced.

3.2.2 Analysis of Variance (ANOVA)

Analysis of Variance (ANOVA) is a statistical method used to analyze the differences among group means and their associated procedures (such as "variation" among and between groups).

In the case study, ANOVA is performed to the difference of variable disabled determinants in proxy data (AEEH and EN3-12) by the clustering of departments (between and within cluster).

A. ANOVA of the data AEEH

- Analysis Between Clusters

On the basis of the HCA analysis, ANOVA is performed to difference between the 7 clusters on the AEEH data. The hypothesis is:

H_0: $\mu_1 = \mu_2 = \cdots = \mu_k$; $where\ k = the\ number\ of\ cluster\ (1, 2, 3, 4, 5, 6, 7)$

(There is no significant difference between the averages of the 7 clusters)

H_1: At least one $\mu_i \neq \mu_k$; $where\ i = 1,2, \dots, k$

(At least one significant difference between the averages of the 7 clusters)

Table 3.4 ANOVA of between clusters AEEH

Source	DDL	Somme des carrés	Carré moyen	Valeur F	Pr > F
Model	6	186.182733	31.030455	2.87	0.0133
Error	89	961.791155	10.806642		
Corrected Total	95	1147.973888			

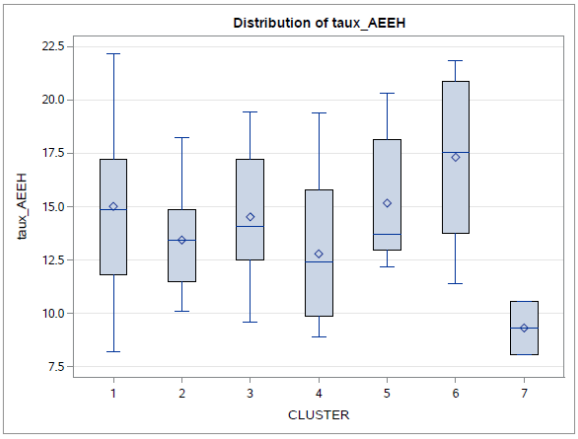
In the table 3.4, the value of probability Pr> F = 0.0133 is less than $\alpha = 0{,}05$. The hypothesis is rejected, so, at least a significant difference between the averages of the 7 clusters of the AEEH data.

In this study, three methods used to compare the averages and identify the clusters that are significantly different, namely Bonferroni, Hochberg, GT2 and Tukey. The Bonferroni and Hochberg-GT2 methods showed no significant

difference between the averages. But the Tukey method is revealed significant differences between clusters 6 and 4 and clusters 6 and 7.

As a reminder, clusters 4 and 6 are characterized by a high rate of premature births and consumption of alcohol, the workers, artisan and agricultural clusters. Cluster 6 is also characterized by a low ESMS rate and the lowest quartile of tax revenue. On the contrary, cluster 7 comprises only two departments, Paris and Hauts de Seine, departments with a population of the graduate level.

Diagram 3.6 Box Plot of 7 clusters AEEH



In box-plot diagram 3.6, the distribution of cluster 7 is very different from other clusters. Only two departments in cluster 7, Paris and the Hauts de Seine, departments already very different at the socio-economic level.

Therefore, a new ANOVA is realized on the 6 clusters by dismissing the cluster 7. The hypothesis is :

$$H_0\colon\ \mu_1 = \mu_2 = \cdots = \mu_j\ ;\ where\ k = the\ number\ of\ cluster\ (1, 2, 3, 4, 5, 6)$$

(There is no significant difference between the averages of the 6 clusters)

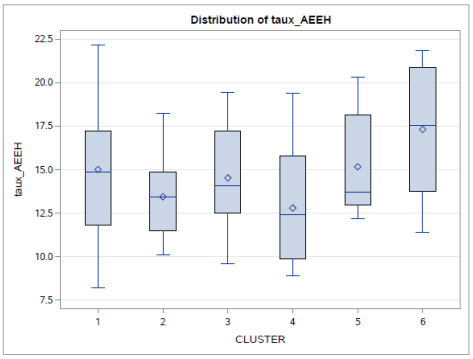
$$H_1\colon \text{At least one } \mu_i \neq \mu_j\ ;\ where\ i = 1,2, \dots, j$$

(At least one significant difference between the averages of the 6 clusters)

Table 3.5 ANOVA of 6 clusters

Source	DDL	Somme des carrés	Carré moyen	Valeur F	Pr > F
Model	5	134.445140	26.889028	2.47	0.0385
Error	88	958.654368	10.893800		
Corrected Total	93	1093.099508			

Diagram 3.7 Box Plot of 6 clusters AEEH



In table 3.5 ANOVA, the significant value is 0.0385 and allows us to reject H_0 for a significance level of 5%. The tests of Bonferroni, Tukey and Hochberg-GT2 showed the same result: there is a significant difference in the means of clusters 6 and 4.

The values of F on the anova 7 clusters and the anova 6 clusters are both significant and less than $\alpha = 5\%$. In other words, cluster 7 which consists of two departments (Paris and the Hauts de Seine) does not influence the average differences between the other clusters.

- Analysis Within Clusters

The ANOVA within clusters is done from the residual or error data of the previous ANOVA model. Before analyzing this data, test the normality of the data performed by univariate analysis. The following assumptions:

H_0: Data is a normal distribution

H_1: Data is not normal distribution

On the base of the univariate analysis, the Kolmogorov-Smirnov test will be able to show whether the data follow a normal distribution or not. The P-value of the Kolmogorov-Smirnov test is greater than $\alpha = 5\%$ or 0.05 (P_Value:> 0.15), then it can't reject H_0, the data follow a normal distribution. And then, ANOVA is proceed to analyze the difference of the mean which the

absolute values of the residuals between the 7 clusters on the AEEH data intra-cluster ANOVA. The following assumptions:

H_0: $\sigma_1 = \sigma_2 = \cdots = \sigma_k$; $where\ k = the\ number\ of\ cluster\ (1, 2, 3, 4, 5, 6, 7)$

(There is no significant difference between the variances within of the 7 clusters)

$$H_1: \text{At least one } \sigma_i \neq \sigma_k \ ;\ where\ i = 1,2, \dots, k$$

(At least one significant difference between the variances within of the 7 clusters)

Table 3.6 ANOVA of within clusters AEEH

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	20.1998961	3.3666493	0.96	0.4553
Error	89	311.2888800	3.4976279		
Corrected Total	95	331.4887761			

Table 3.7 of the ANOVA gives a value of and the probability that Pr> F = 0.4553 is greater than. It can't reject H_0, there is no significant difference in the variances within each of the 7 clusters. Moreover, it is supported by the Bartlette test which is also shows the value of P-value> 0.05 and ensures the homogeneity of the variances each of the 7 clusters.

B. ANOVA of the EN3-12 data

The second ANOVA is the analysis of variance by the number of children with disabilities in the national education (EN3 and EN12). First, analyze the mean difference between the 7 clusters and second step is analyze the difference of variance within each of these clusters on EN3-12 data.

- Analysis Between Clusters

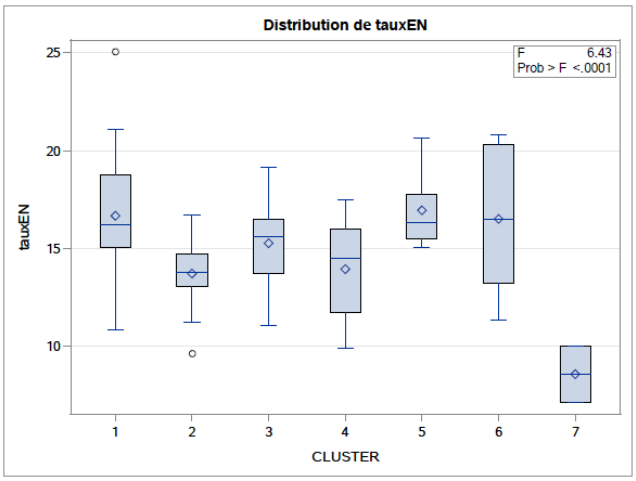
Like the AEEH data, table 3.8 showed that ANOVA enters cluster on data EN3-12 is rejects H0. The value P-value <0.0001 is less than a threshold of 5%.

Table 3.7 ANOVA of between 7 clusters EN3-12

Source	DDL	Somme des carrés	Carré moyen	Valeur F	Pr > F
Model	6	256.4020093	42.7336682	6.43	<.0001
Error	89	591.4285076	6.6452641		
Corrected Total	95	847.8305168			

The Bonferroni, Hochberg-GT2 and Tukey tests are showed the same results: there is a significant difference between the averages of the following clusters: 5-7; 1-4; 1-2; 1-7; 6-7; and 3-7.

Diagram 3.8 Box Plot of clusters 7 EN3-12



In the diagram 3.8, the distribution of cluster 7 is very different from the other clusters. Therefore, as like as the AEEH, the 6 clusters is analyzed by discarding cluster 7, the hypothesis is thus:

$$H_0\colon\ \mu_1 = \mu_2 = \cdots = \mu_j\ ;\ where\ k = the\ number\ of\ cluster\ (1, 2, 3, 4, 5, 6)$$

(There is no significant difference between the averages of the 6 clusters)

$$H_1\colon \text{At least one } \mu_i \neq \mu_j\ ;\ where\ i = 1,2, \dots, j$$

(At least one significant difference between the averages of the 6 clusters)

Table 3.8 ANOVA of between 6 clusters EN3-12

Source	DDL	Somme des carrés	Carré moyen	Valeur F	Pr > F
Model	5	163.5899418	32.7179884	4.90	0.0005
Error	88	587.1650885	6.6723306		
Corrected Total	93	750.7550303			

Diagram 3.9 Box Plot of 6 clusters EN3-12

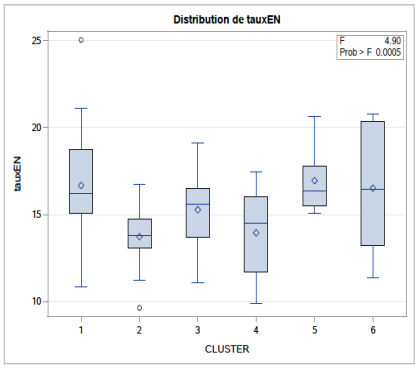


Table 3.8 of the ANOVA gives a value and the probability Pr> F = 0.0005 is less than α = 5%. The hypothesis is rejected; there is a significant difference between the averages of the number of children with disabilities in the national education according to the 6 clusters of the CAH.

The Bonferroni, Tukey and Hochberg-GT2 tests offer one common result: there is a significant difference in the averages of the rates of handicapped children enrolled in national education between grades 1-4 and 1-2.

• Analysis Within Clusters

As for the AEEH, before analyzing this data, test the normality of the data performed by univariate analysis. The Kolmogorov-Smirnov test will be able to show whether the data follow a normal distribution or not. The P-value of the Kolmogorov-Smirnov test is greater than α = 5% or 0.05 (P_Value:> 0.15), then it can't reject H0, the data is a normal distribution.

And then, ANOVA is proceed to analyze the difference of the mean of the absolute values of the residuals between the 7 clusters on the EN3 and EN12 data. The following assumptions:

H_0: $\sigma_1 = \sigma_2 = \cdots = \sigma_k$; $where\ k = the\ number\ of\ cluster\ (1, 2, 3, 4, 5, 6, 7)$

(There is no significant difference between the variances within of the 7 clusters)

H_1: At least one $\sigma_i \neq \sigma_k$; $where\ i = 1,2, \ldots, k$

(At least one significant difference between the variances within each the 7 clusters)

Table 3.9 ANOVA of within 7 clusters EN3-12

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	20.2425524	3.3737587	1.60	0.1563
Error	89	187.5865199	2.1077137		
Corrected Total	95	207.8290723			

The table 3.10 of the ANOVA gives a value probability Pr> F = 0, 7563 is greater than. α = 5% or 0.05 H_0 is not rejected, then there is no significant difference in the variances within each of the 7 clusters. Moreover, it is supported by the Bartlette test which also shows the value of P-value> 0.05 and ensures the homogeneity of the variances each of the 7 cluster.

C. Conclusion Analysis of variance (ANOVA)

The different averages of the proxy data (AEEH and EN 3-12) issued HCA of PCA by department clusters are significantly different between clusters. On the contrary, the difference of variance within clusters is not significant. The groups are homogeneous. The best method to identify the two-averages differences is Tukey. The Tukey test is generally more effective to testing a large number of pairs averages.

3.3. Strong Contribution of Certain Departments

3.3.1 Multiple Component Analysis (MCA)

Multiple correspondence analysis (MCA) is the factorial method which adapted to tables. It is a set of individuals which described by several qualitative variables. It can be presented in many different ways. In France, following the work of L. Lebart, the most common is to focus on the similarities with correspondence analysis, a method designed to study the relationship between two qualitative variables. As a PCA, the aim of multiple component analysis (MCA) is to read the information contained in a multidimensional space by a reduction of the dimension. The MCA is permitted to answers the following questions:

- Which departments resemble each other? Which are different?
- Are there homogeneous groups of individuals? Is it possible to identify a typology of individuals?

A. Construction of the table data

In this study, the clusterification of variable is selected into four clusters. The boundaries of the clusters being defined by the quartiles and renamed with readily identifiable labels as shown in Table 3.11 below.

The table is describes each department by the rate of the social professional category (CSP), the level education, the premature rates, the tax of revenues, alcohol consumption, facilities and services of medical social (ESMS) for disabled.

Table 3.10 The identification labels of determinants variables

	--	1er quartile	-	2nd quartile	+	3ème quartile	++
Alcool	--	<3.3%	-	<4,2%	+	<5.5%	++
Agriculteur	--	<1.3	-	<2.7	+	<4.1	++
Artisans	--	<5.9	-	<7.2	+	<8.8	++
Cadres	--	<10.5	-	<12.7	+	<15.9	++
Profession intermédiaire	--	<23.8	-	<25.1	+	<26.8	++
Employés	--	<26.8	-	<28.1	+	<29.3	++
Ouvriers	--	<19.4	-	<22.5	+	<25.7	++
Revenus	--	< 12 341€	-	< 12 887€	+	<13 716€	++
Prématurité	--	<0.6%	-	<0.7%	+	<0.8%	++
ESMS	--	<8.5%	-	<10.3%	+	<11.3%	++
Sans diplôme	--	<26.3%	-	<29.5%	+	<32.1%	++
BEPC	--	<30.6%	-	<32.2%	+	<33.7%	++
Bac	--	<14.9%	-	<15.9%	+	<17.1%	++
Bac plus	--	<19.4%	-	<22.3%	+	<25.6%	++

B. Eigenvalue and Numbers of Axis

The number of axis is determined by the formula :

$$(\frac{M}{m} - 1) \quad (3.3)$$

Where,

m = 14 disability determinant variables

M = 56 sum of the modalities of the active variables

By attachment 4, the eigenvalue is decided three axis with 30% of the total inertia.

C. Results on variables

The table 3.12 below is gives an analysis of the contribution the variables to the axek. The variables is provided to the most information which a contribution higher than the average contribution (1 / M with M = 56 are the number of variables), the sense of contribution depends on the sign of the coordinate.

Table 3.11 The contribution of determinants variables

Déterminant	Axe 1		Axe 2		Axe 3	
	+	-	+	-	+	-
ALCOOL						
AGRICULTEUR	--	++		-	++	-
ARTISAN			++	--	++	-
CADRES	++	--	+/-	++	--	+
PROF_INT	++	--	+		--	+
EMPLOYES	--				++	
OUVRIERS	--	++	-	++	--	
SANS DIPLÔME	--	++	+/-	++/--	+	
BEPC	--	++	-		--	-
BAC		--	++	--	++	-
BAC PLUS	++		+/-	++/--		+
ESMS					--	+
PREMATURITE			--	++		
REVENUS	++			++	--	+

The first axe is opposed the departments which manager have a the education level of bac and a first quartile of revenus tax, and the departments which workers without a diploma or a BEPC / CAP / BEP level. The second axe is defined the departments with artisan, craftsman and trader. The manager and the premature rates are low. On the contrary, the character of departments with a large population is not only manager but also workers, a first quartile of revenue tax and a high rate of premature.

The third axe is included the departments with a high rate of farmers, craftsmen and employees who do not have a high income quartile. The ESMS rate is poor too. On the other hand, the departments which is a large number of managers and intermediate professions with a higher or lower quartile of tax revenue and a high ESMS rate.

D. Results on the departments

The MCA is calculated the coordinates of the individuals on the axes and their contributions. As for the variables, some departments have a contribution higher than the average contribution (1 / n with n = 96 are the numbers of individuals), the direction of the contribution depends on the sign of the

coordinate. The MCA chart is difficult to interpret. The superposition of the 56 variables and the 96 departments do not allow the identification of departments.

E. Conclusion Multiple component analysis (MCA)

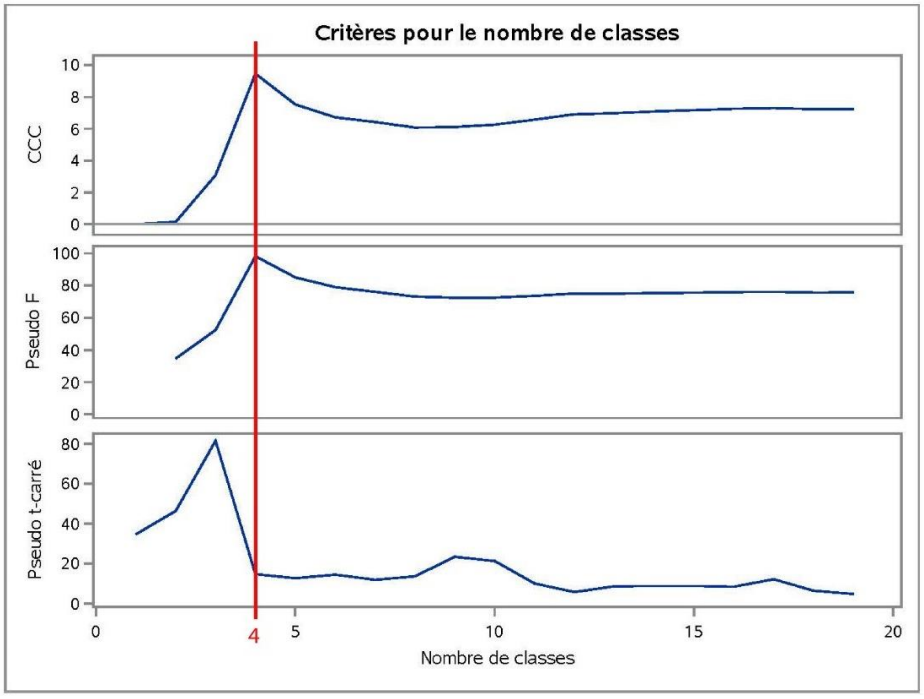
The MCA reduces the determinants to 3 axes, retains only 30% of total information. However, the determinant of the alcohol consumption does not contribute to these axes.

3.3.2 Hierarchical Cluster Analysis (HCA) of the departments using the axes of the MCA

As previously, the hierarchical cluster analysis will classify the departments into homogeneous clusters.

A. The choice of the number of clusters

Diagram 3.10 The criteria of cluster number (MCA)



The number of clusters is must satisfied above the criteria. The seven clusters are selected of departments.

b. The Characterization and localization of clusters

The table 3.13 below is showed the average of axis in each cluster.

Table 3.12 The average of clusters by axe

	classe			
	1	2	3	4
	moyenne	moyenne	moyenne	moyenne
Dim1	-0.83	0.09	-0.08	0.96
Dim2	-0.45	0.13	0.64	-0.45
Dim3	0.11	-0.53	0.40	0.24

The characterization of the four clusters is as follows:

- Cluster 1

Cluster 1 is consists mainly of axis 1 which a high rate of workers without a diploma or BEPC, BEP or CAP. There is little presence of senior or intermediate level professions at the Baccalaureate level. Cluster 1 is comprised 23 departments in the center and north of France: Haute-Marne, Mayenne, Meuse, Nièvre, Orne, Pas-de-Calais, Haute-Marne, Saône, Saône-et-Loire, Deux-Sèvres, the Ardennes, Vendée, Vosges, Yonne.

- Cluster 2

Cluster 2 is composed to the axe 3 or regions with a population of managers and intermediate professions, some farmers and artisan, craftsman and trader. The first quartile of the revenue tax is a little high and the rate of equipment in ESMS also. Cluster 2 is the cluster which 31 departments in the northern of France except the Pyrénées Atlantiques (64).

- Cluster 3

Cluster 3 is characterized by axe 2 of the MCA with the regions artisan, craftsman and trader, managers and intermediate professions. The level of education mainly the bac and premature births are poor. Cluster 3 also includes 23 departments in the southern of France. They are Aude, Aveyron, Charente-Maritime, Corrèze, Dordogne, Southern Corsica, Haute-Corse, Gard, Gers, Alpes-

de-Haute- Hautes-Pyrénées, Pyrénées-Orientales, Ardèche, Tarn, Tarn-et-Garonne, Var, Vaucluse, Ariège.

- Cluster 4

Cluster 4 is the largest on axis 1. It is characterized by a large population of managers and intermediate professions with a higher education level of bac. Thus, the first quartile of tax revenue is strong. There are few employees, workers and farmers. The 19 departments are Paris and suburbs to Rennes, Nantes, Bordeaux, Toulouse, Montpellier, Nice, Lyon, Grenoble and Strasbourg.

C. Conclusion of Hierarchical Cluster Analysis (HCA)

The HCA issued by the MCA put the 96 departments into 4 clusters. In cluster 1, there are 23 departments and cluster 2 is the biggest cluster because there are 31 departments. Cluster 3 includes 23 departments and in cluster 4 only 19 departments.

3.3.3 Analysis of Variance (ANOVA) of proxies

Like ANOVA resulting from the HCA-PCA, an analysis of variance is carried out to determine the average differences in the AEEH and EN 3-12 data by the clusterification of departments (between and within cluster).

A. ANOVA of the data AEEH

- Analysis Between Clusters

On the basis of the CAH analysis, ANOVA is performed to difference between the 4 clusters on the AEEH data. The hypothesis is:

$$H_0\colon\ \mu_1 = \mu_2 = \cdots = \mu_n\ ;$$

$$where\ n = the\ number\ of\ cluster\ on\ MCA\ \ (1, 2, 3, 4)$$

(There is no significant difference between the averages of the 4 clusters)

$$H_1\colon \text{At least one } \mu_i \neq \mu_n\ ;\ where\ i = 1,2, \dots, n$$

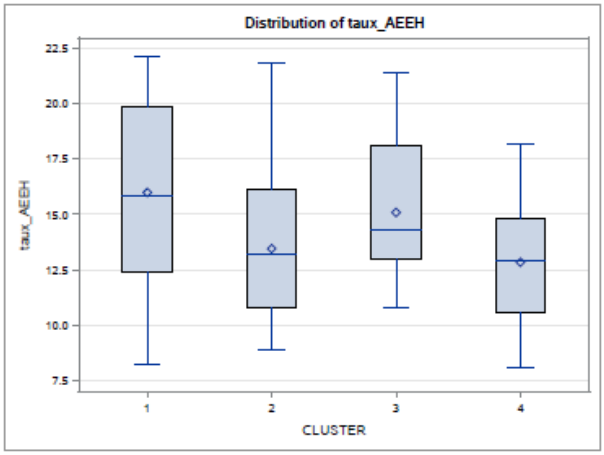
(At least one significant difference between the averages of the 4 clusters)

Table 3.13 ANOVA of MCA between clusters AEEH

Source	DDL	Somme des carrés	Carré moyen	Valeur F	Pr > F
Model	3	142.907852	47.635951	4.36	0.0064
Error	92	1005.066035	10.924631		
Corrected Total	95	1147.973888			

In Table 3.14 the value of probability Pr> F = 0.0064 is less than $\alpha = 0{,}05$. The hypothesis is rejected, so, at least a significant difference between the averages of the 4 clusters of the AEEH data. The Bonferroni, Tukey and Hochberg-GT2 tests offer a common result: there is a significant difference in the average of the beneficiary rates of the AEEH of clusters 1 and 2 and 1 and 4.

Diagram 3.11 Box Plot of 4 clusters AEEH



In diagram 3.11, the averages of each cluster are in the same rank. The clusters are not very different.

- ANOVA analysis within clusters

The ANOVA analysis within clusters is done from the residual or error data of the previous ANOVA model. Before analyzing this data, test the normality of the data performed by univariate analysis. The following assumptions:

H_0: Data is a normal distribution

H_1: Data is not normal distribution

On the base of the univariate analysis, the Kolmogorov-Smirnov test will be able to show whether the data follow a normal distribution or not. The P-value of the Kolmogorov-Smirnov test is greater than α = 5% or 0.05 (P_Value:> 0.15), then it can't reject H_0, the data follow a normal distribution. And then, ANOVA is proceed to analyze the difference of the mean which the absolute values of the residuals between the 4 clusters on the AEEH data intra-cluster ANOVA. The following assumptions:

$$H_0: \sigma_1 = \sigma_2 = \cdots = \sigma_n ;$$

$$where\ n = the\ number\ of\ cluster\ on\ MCA\ (1, 2, 3, 4)$$

(There is no significant difference between the variances within each on the 4 clusters)

$$H_1: \text{At least one } \sigma_i \neq \sigma_n ;\ where\ i = 1,2, \dots, n$$

(At least one significant difference between the variances within of the 4 clusters)

Table 3.14 ANOVA of MCA within clusters AEEH

Source	DDL	Somme des carrés	Carré moyen	Valeur F	Pr > F
Model	3	7.3192994	2.4397665	0.73	0.5365
Error	92	307.3679347	3.3409558		
Corrected Total	95	314.6872340			

The table 3.15 of the ANOVA gives a value probability Pr> F = 0.5365 is greater than. α = 5% or 0.05 H_0 is not rejected, then there is no significant difference in the variances within each of the 4 clusters. Moreover, it is supported by the Bartlette test which also shows the value of P-value> 0.05 and ensures the homogeneity of the variances each of the 4 cluster.

B. ANOVA of the EN3-12 data

- Analysis Between Clusters

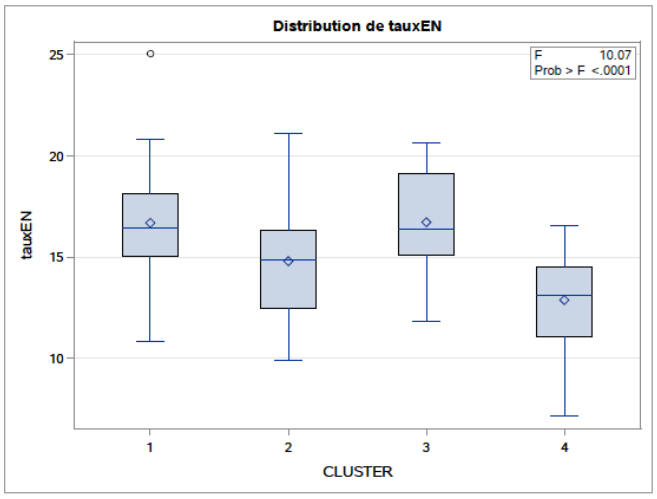
Like the AEEH data, table 3.16 is showed ANOVA enters cluster on data EN3-12 is rejects H0. The value P-value <0.0001 is less than a threshold of 5%.

Table 3.15 ANOVA of MCA between clusters EN3-12

Source	DDL	Somme des carrés	Carré moyen	Valeur F	Pr > F
Model	3	209.5120930	69.8373643	10.07	<.0001
Error	92	638.3184238	6.9382437		
Corrected Total	95	847.8305168			

The methods Bonferroni and Hochberg-GT2 tests are showed the same results: there is a significant difference between the averages of the following clusters: clusters 3 and 4, and clusters 1 and 4. The Tukey method is revealed significant differences between clusters 3 and 2, clusters 3 and 4, clusters 1 and 4.

Diagram 3.12 Box Plot of 4 clusters EN3-12



In diagram 3.12, the averages of each cluster are in the same rank. The clusters are not very different.

• ANOVA analysis within clusters

Table 3.16 ANOVA of MCA within clusters EN3-12

Source	DDL	Somme des carrés	Carré moyen	Valeur F	Pr > F
Model	3	1.8497813	0.6165938	0.24	0.8703
Error	92	239.3291916	2.6014043		
Corrected Total	95	241.1789728			

The table 3.16 of the ANOVA gives a value probability Pr> F = = 0, 8703 is greater than. α = 5% or 0.05 H_0 is not rejected, then there is no significant difference in the variances within each of the 4 clusters. Moreover, it is supported by the Bartlette test which also shows the value of P-value> 0.05 and ensures the homogeneity of the variances each of the 4 cluster.

C. Conclusion Analysis of variance (ANOVA)

The different averages of the proxy data (AEEH and EN 3-12) issued HCA of MCA by department clusters are significantly different between clusters and the variance within of cluster is not difference significantly, the clusters are homogeneous.

IV. SUMMARY OF RESULTS

Statistical analysis performed to model the variations of disabled children about 0-19 years old population among French department consisting of 14 determinant variables of children disabilities based on six categories, namely, the professional category of social (CSP) of their parents, the level education of their parents, the premature rates, the tax of revenues, alcohol consumption, facilities and services of medical social for disabled. The aim is to clusterify departments according to their profiles determinants (socioeconomic and behavioral profiles).

The PCA reduces 14 determinants of disability to 4 axes, keeps 80% of total information. PCA is also calculating the coordinates of the department and the contribution of the determinants variables per axis. It is not balanced because some departments have too strong contribution. The Hierarchical Ascending Clusterification (HCA) issued by the PCA made it possible to group the departments into 7 clusters. With 35 departments, cluster 1 is the cluster with the most individuals. On the contrary, cluster 7 contains only two. The different averages of the proxy data (AEEH and EN 3-12) issued HCA of PCA by department clusters are significantly different between clusters. On the contrary, the difference of variance within clusters is not significant. The groups are homogeneous. The best method to identify the two-average differences is Tukey. The Tukey test is generally more effective to testing a large number of pair's averages.

The MCA reduces the determinants to 3 axes, retains only 30% of total information. However, the determinant of the alcohol consumption does not contribute to these axes. The HCA issued by the MCA put the 96 departments into 4 clusters. In cluster 1, there are 23 departments and cluster 2 is the biggest cluster because there are 31 departments. Cluster 3 includes 23 departments and in cluster 4 only 19 departments. The different averages of the proxy data (AEEH and EN 3-12) issued HCA of MCA by department clusters are significantly different between clusters and the variance within of cluster is not difference significantly, the clusters are homogeneous.

V. CONCLUSION

Based on the results and the discussion on the methods, the PCA is the best method for interpretation of the correlation between the determinants of disability (independent variable). Although, a more balanced clusterification of the department is clusterification on MCA. The different averages of the proxy data (AEEH and EN 3-12) by department clusters are significantly different between clusters and the variance within of cluster is not difference significantly, the clusters are homogeneous.

REFERENCES

ACP sous SAS : http://www.univ-mrs.fr/~reboul/ACPsas.doc, consulté le 1er juin 2016.

Analyse factorielle multiple des correspondances (AFCM) : http://www.math.univ-toulouse.fr/~besse/Wikistat/pdf/st-m-explo-afcm.pdf, le 1er juin 2016.

Baccini, A. (2010), Statistique descriptive multidimensionnelle, Publications de l'institut de mathématiques de Toulouse, page 37.

Bressoux, P. (2010), Modélisation statistique appliquée aux sciences sociales, De boeck, 464p.

Dagnelie, P. (1975), Analyse statistique à plusieurs variables, Les presses agronomiques de gembloux, page 362.

Espagnacq M. (2015), Populations à risque de handicap et restrictions de participation sociale Dossiers solidarité et santé, Drees. n°68, page 18.

INSEE, The National Institute of Statistics and Economic Studies collects, analyses and disseminates information on the French economy and society

Laidier, L. (2015), A l'école et au collège, les enfants en situation de handicap constituent une population fortement différenciée scolairement et socialement. Note d'information DEPP n°4, page 4.

Le Centre interregional s'Etudes, d'Action et d'Informations (CREAI) en faveur des personnes en situation de vulnérabilité

Mormiche, P. (2003), Handicap et Inégalités sociales. Revue française des affaires sociales 2003/1-2 (n° 1-2) , page11-29.

_____ (2011) Organisation Mondiale de la Santé (OMS), France.

Pech, N., Les Guides SAS - L'analyse des données : La procédure cluster, Document du cours MASS POP, 2015/2016, page 160-252.

Rakotomalala, R. (2011), Tests de normalité Techniques empiriques et tests statistiques, Université Lumière Lyon, vers no.2, page 59.

Rican, S. (2011), Désavantages locaux et santé : construction d'indices pour l'analyse des inégalités sociales et territoiriales de santé en France et leurs évolutions. Env Risque Santé Vol10 n°3., page 211-215

Rican, S. (2014), Inégalités géographiques de santé en France Les tribunes de la santé n°43, page 39-45.

Rican, S., Jougla, E., and Salem, G. (2003), Inégalités socio-spatiales de mortalité en France Bulletin épidémiologique hebdomadaire INVS. n°30-31.

Ringuede, S. (2014), SAS 3e édition Introduction au décisionnel : du data management au reporting, Pearson, page 550.

Roubaud, Mc., Analyses de variance et covariance, Document du cours MASS POP, 2015/2016, p.age 37-56.

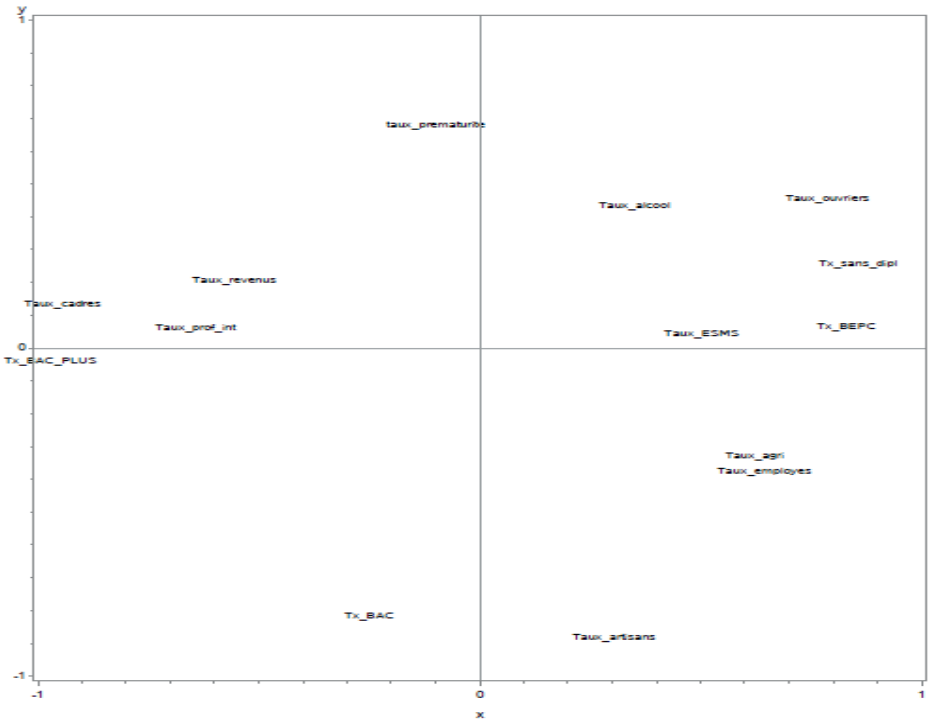
Roubaud, Mc., Information hiérarchique : Analyse multiniveau, Document du cours MASS POP, 2015/2016, page 1-17.

Salem, G. (2000), Dynamiques territoiriales, dynamiques sanitaires : de la description à l'action. Conférence préliminaire l'observation locale en santé, 2009. G. SALEM, S. RICAN, ML KURZINGER, Atlas de la santé de France, volume 2 : comportements et maladies. John Libbey, page 222.

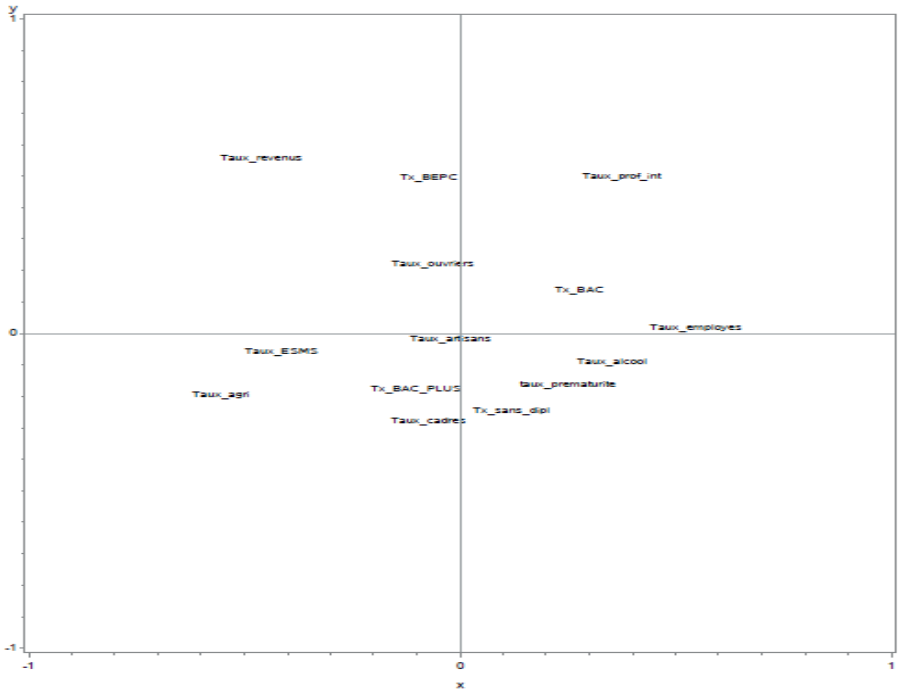
Vigneron, E. (2013), Inégalités de santé, inégalités de soins dans les territoires français, Etat des lieux et voies de progrès. Elsevier Masson, 2011, 194p E. VIGNERON, Inégalités de santé, inégalités de soins dans les territoires français Les tribunes de la santé n°38, page 41-53.

Yan, X., (2009), Linear Regression Analysis : Theory and Computing, World Scientific, page 384.

Attachements 1 : Representation of variables on axes 1 et 2

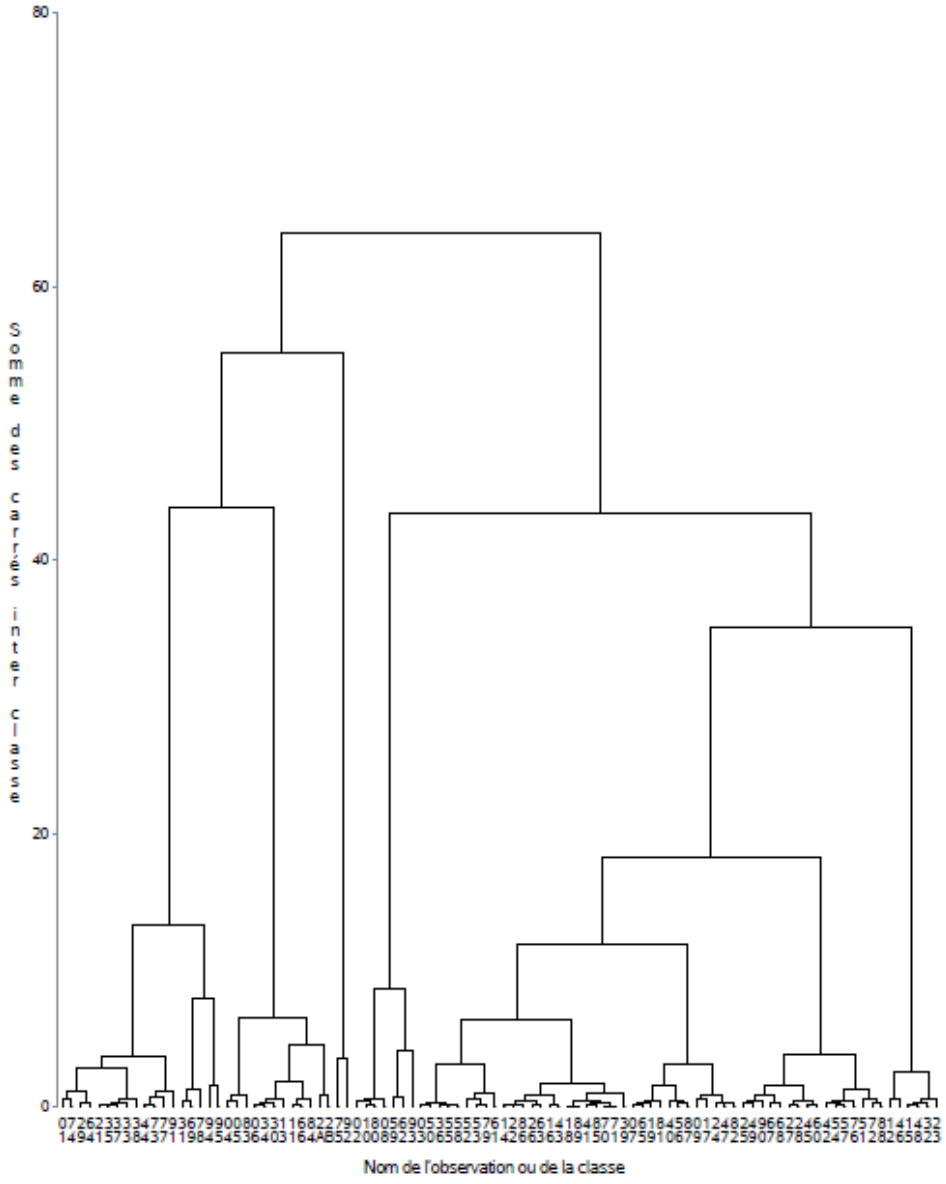


Attachements 2 : Representations of variables on axes 3 et 4



Source: Issued by SAS of the proceedings The PCA

Attachements 3 : Analyse of clusterifications on PCA - HCA



Source: Issued by SAS of the proceedings The HCA of PCA

Attachements 4 : Student test of Tukey (HSD) in AEEHdata

Comparaisons significatives au niveau 0.05 indiquées par ***.				
CLUSTER Comparaison	Différence Entre les moyennes	Simultané 95% Intervalle de confiance		
6 - 5	2.1524	-3.3650	7.6698	
6 - 1	2.3068	-1.7993	6.4129	
6 - 3	2.7970	-1.9195	7.5136	
6 - 2	3.8821	-0.5354	8.2995	
6 - 4	4.5301	0.0360	9.0242	***
6 - 7	8.0078	0.0564	15.9592	***
5 - 6	-2.1524	-7.6698	3.3650	
5 - 1	0.1544	-4.2276	4.5364	
5 - 3	0.6447	-4.3139	5.6032	
5 - 2	1.7297	-2.9453	6.4047	
5 - 4	2.3777	-2.3698	7.1252	
5 - 7	5.8554	-2.2419	13.9527	
1 - 6	-2.3068	-6.4129	1.7993	
1 - 5	-0.1544	-4.5364	4.2276	
1 - 3	0.4902	-2.8273	3.8078	
1 - 2	1.5753	-1.3011	4.4517	
1 - 4	2.2233	-0.7695	5.2161	
1 - 7	5.7010	-1.5091	12.9111	
3 - 6	-2.7970	-7.5136	1.9195	
3 - 5	-0.6447	-5.6032	4.3139	
3 - 1	-0.4902	-3.8078	2.8273	
3 - 2	1.0850	-2.6109	4.7809	
3 - 4	1.7330	-2.0541	5.5202	
3 - 7	5.2107	-2.3636	12.7851	
2 - 6	-3.8821	-8.2995	0.5354	
2 - 5	-1.7297	-6.4047	2.9453	

Comparaisons significatives au niveau 0.05 indiquées par ***.				
CLUSTER Comparaison	Différence Entre les moyennes	Simultané 95% Intervalle de confiance		
2 - 1	-1.5753	-4.4517	1.3011	
2 - 3	-1.0850	-4.7809	2.6109	
2 - 4	0.6480	-2.7595	4.0555	
2 - 7	4.1257	-3.2661	11.5175	
4 - 6	-4.5301	-9.0242	-0.0360	***
4 - 5	-2.3777	-7.1252	2.3698	
4 - 1	-2.2233	-5.2161	0.7695	
4 - 3	-1.7330	-5.5202	2.0541	
4 - 2	-0.6480	-4.0555	2.7595	
4 - 7	3.4777	-3.9602	10.9156	
7 - 6	-8.0078	-15.9592	-0.0564	***
7 - 5	-5.8554	-13.9527	2.2419	
7 - 1	-5.7010	-12.9111	1.5091	
7 - 3	-5.2107	-12.7851	2.3636	
7 - 2	-4.1257	-11.5175	3.2661	
7 - 4	-3.4777	-10.9156	3.9602	

Source : Issued by SAS of the proceedings The ANOVA of PCA in AEEH data

Attachements 5 : Students Test of modulus maximum (GT2) on AEEH data

Comparaisons significatives au niveau 0.05 indiquées par ***.				
CLUSTER Comparaison	Différence Entre les moyennes	95% Intervalle de confiance		
6 - 5	2.1524	-3.5436	7.8484	
6 - 1	2.3068	-1.9322	6.5458	
6 - 3	2.7970	-2.0722	7.6663	
6 - 2	3.8821	-0.6784	8.4425	
6 - 4	4.5301	-0.1095	9.1696	
6 - 7	8.0078	-0.2010	16.2166	
5 - 6	-2.1524	-7.8484	3.5436	
5 - 1	0.1544	-4.3694	4.6782	
5 - 3	0.6447	-4.4744	5.7637	
5 - 2	1.7297	-3.0966	6.5560	
5 - 4	2.3777	-2.5234	7.2789	
5 - 7	5.8554	-2.5040	14.2148	
1 - 6	-2.3068	-6.5458	1.9322	
1 - 5	-0.1544	-4.6782	4.3694	
1 - 3	0.4902	-2.9346	3.9151	
1 - 2	1.5753	-1.3943	4.5448	
1 - 4	2.2233	-0.8664	5.3130	
1 - 7	5.7010	-1.7425	13.1444	
3 - 6	-2.7970	-7.6663	2.0722	
3 - 5	-0.6447	-5.7637	4.4744	
3 - 1	-0.4902	-3.9151	2.9346	
3 - 2	1.0850	-2.7305	4.9006	
3 - 4	1.7330	-2.1767	5.6428	
3 - 7	5.2107	-2.6088	13.0303	
2 - 6	-3.8821	-8.4425	0.6784	
2 - 5	-1.7297	-6.5560	3.0966	

Comparaisons significatives au niveau 0.05 indiquées par ***.			
CLUSTER Comparaison	Différence Entre les moyennes	95% Intervalle de confiance	
2 - 1	-1.5753	-4.5448	1.3943
2 - 3	-1.0850	-4.9006	2.7305
2 - 4	0.6480	-2.8697	4.1658
2 - 7	4.1257	-3.5054	11.7568
4 - 6	-4.5301	-9.1696	0.1095
4 - 5	-2.3777	-7.2789	2.5234
4 - 1	-2.2233	-5.3130	0.8664
4 - 3	-1.7330	-5.6428	2.1767
4 - 2	-0.6480	-4.1658	2.8697
4 - 7	3.4777	-4.2009	11.1563
7 - 6	-8.0078	-16.2166	0.2010
7 - 5	-5.8554	-14.2148	2.5040
7 - 1	-5.7010	-13.1444	1.7425
7 - 3	-5.2107	-13.0303	2.6088
7 - 2	-4.1257	-11.7568	3.5054
7 - 4	-3.4777	-11.1563	4.2009

Source : Issued by SAS of the proceedings The ANOVAof PCA on AEEH

Attachements 6 : T Tests of Bonferroni (Dunn) in AEEH

Comparaisons significatives au niveau 0.05 indiquées par ***.			
CLUSTER Comparaison	Différence Entre les moyennes	Simultané 95% Intervalle de confiance	
6 - 5	2.1524	-3.5680	7.8727
6 - 1	2.3068	-1.9503	6.5639
6 - 3	2.7970	-2.0930	7.6871
6 - 2	3.8821	-0.6979	8.4620
6 - 4	4.5301	-0.1293	9.1895
6 - 7	8.0078	-0.2361	16.2517
5 - 6	-2.1524	-7.8727	3.5680
5 - 1	0.1544	-4.3887	4.6976
5 - 3	0.6447	-4.4963	5.7856
5 - 2	1.7297	-3.1172	6.5766
5 - 4	2.3777	-2.5444	7.2998
5 - 7	5.8554	-2.5397	14.2506
1 - 6	-2.3068	-6.5639	1.9503
1 - 5	-0.1544	-4.6976	4.3887
1 - 3	0.4902	-2.9493	3.9298
1 - 2	1.5753	-1.4069	4.5575
1 - 4	2.2233	-0.8796	5.3262
1 - 7	5.7010	-1.7743	13.1762
3 - 6	-2.7970	-7.6871	2.0930
3 - 5	-0.6447	-5.7856	4.4963
3 - 1	-0.4902	-3.9298	2.9493
3 - 2	1.0850	-2.7468	4.9169
3 - 4	1.7330	-2.1934	5.6595
3 - 7	5.2107	-2.6422	13.0637
2 - 6	-3.8821	-8.4620	0.6979
2 - 5	-1.7297	-6.5766	3.1172

Comparaisons significatives au niveau 0.05 indiquées par ***.			
CLUSTER Comparaison	Différence Entre les moyennes	Simultané 95% Intervalle de confiance	
2 - 1	-1.5753	-4.5575	1.4069
2 - 3	-1.0850	-4.9169	2.7468
2 - 4	0.6480	-2.8848	4.1808
2 - 7	4.1257	-3.5380	11.7894
4 - 6	-4.5301	-9.1895	0.1293
4 - 5	-2.3777	-7.2998	2.5444
4 - 1	-2.2233	-5.3262	0.8796
4 - 3	-1.7330	-5.6595	2.1934
4 - 2	-0.6480	-4.1808	2.8848
4 - 7	3.4777	-4.2337	11.1891
7 - 6	-8.0078	-16.2517	0.2361
7 - 5	-5.8554	-14.2506	2.5397
7 - 1	-5.7010	-13.1762	1.7743
7 - 3	-5.2107	-13.0637	2.6422
7 - 2	-4.1257	-11.7894	3.5380
7 - 4	-3.4777	-11.1891	4.2337

Source : Issued by SAS of the proceedings The ANOVA of PCA onAEEH

Attachements 7: Proc univariate residual of 7 Clustere AEEH

Moments			
N	96	**Somme des poids**	96
Moyenne	0	**Somme des**	0
Ecart-type	3.1818418	**Variance**	10.124117
Skewness	0.3176851	**Kurtosis**	-
Somme des carrés non	961.79115	**Somme des carrés**	961.79115
Coeff Variation	.	**Std Error Mean**	0.3247453

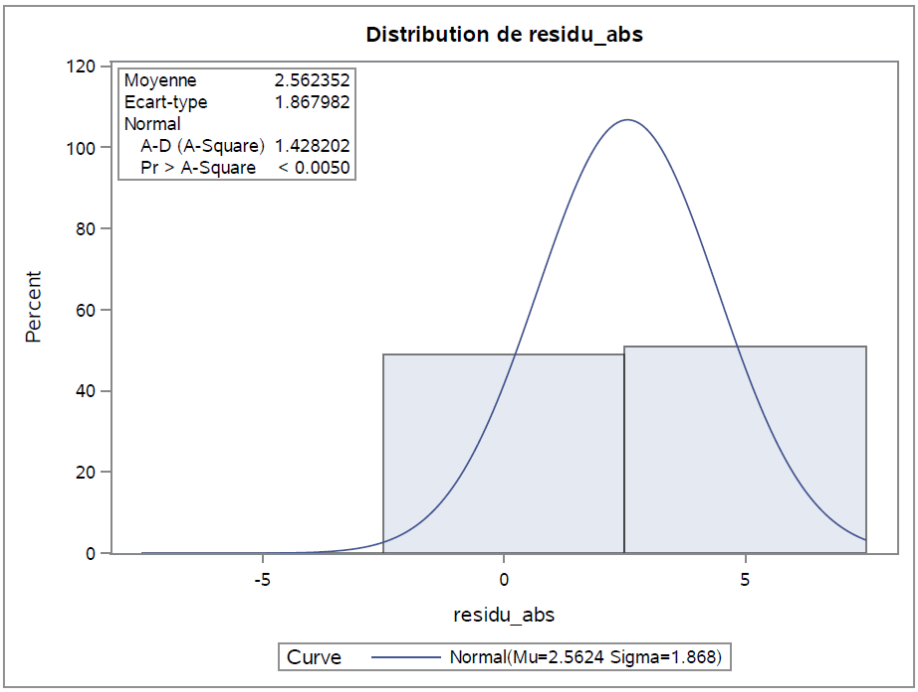
Mesures statistiques de base			
Emplacement		**Variabilité**	
Moyenn	0.00000	**Ecart-type**	3.18184
Médiane	-0.29720	**Variance**	10.12412
Mode	.	**Intervalle**	13.96108
		Ecart	5.02605

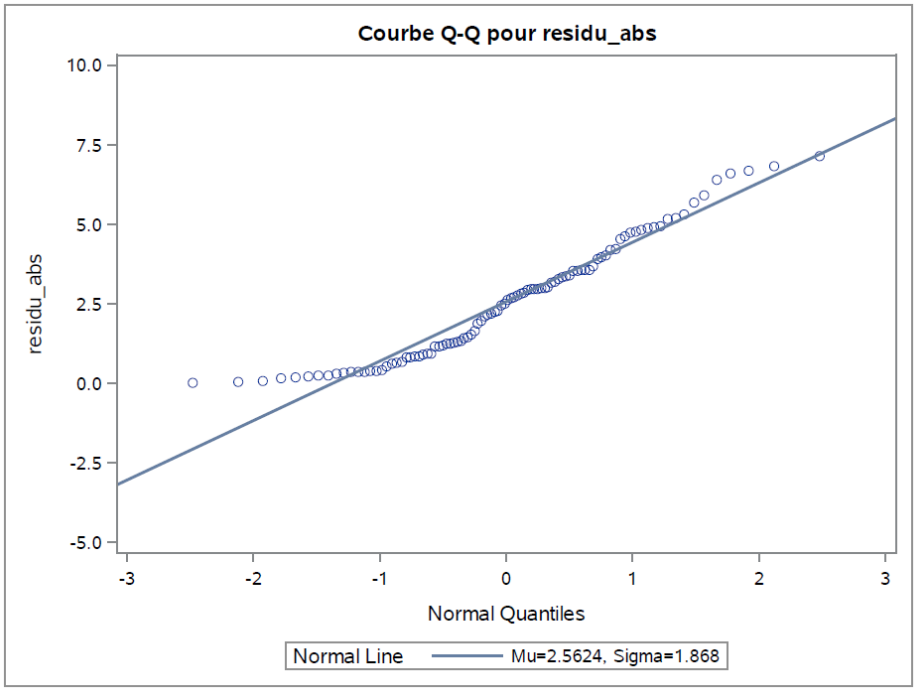
Tests de tendance centrale : Mu0=0				
Test	**Statistique**		**P-value**	
t de	**t**	0	**Pr > |t|**	1.0000
Signe	**M**	-5	**Pr >=**	0.3584
Rang signé	**S**	-89.5	**Pr >= |S|**	0.7455

Tests de normalité				
Test	**Statistique**		**P-value**	
Shapiro-Wilk	**W**	0.975658	**Pr < W**	0.0708
Kolmogorov-	**D**	0.073907	**Pr > D**	>0.1500
Cramer-von Mises	**W-Sq**	0.12445	**Pr > W-Sq**	0.0520
Anderson-Darling	**A-Sq**	0.78528	**Pr > A-Sq**	0.0418

Source : Issued by SAS of the proceedings the ANOVA of PCA on AEEH

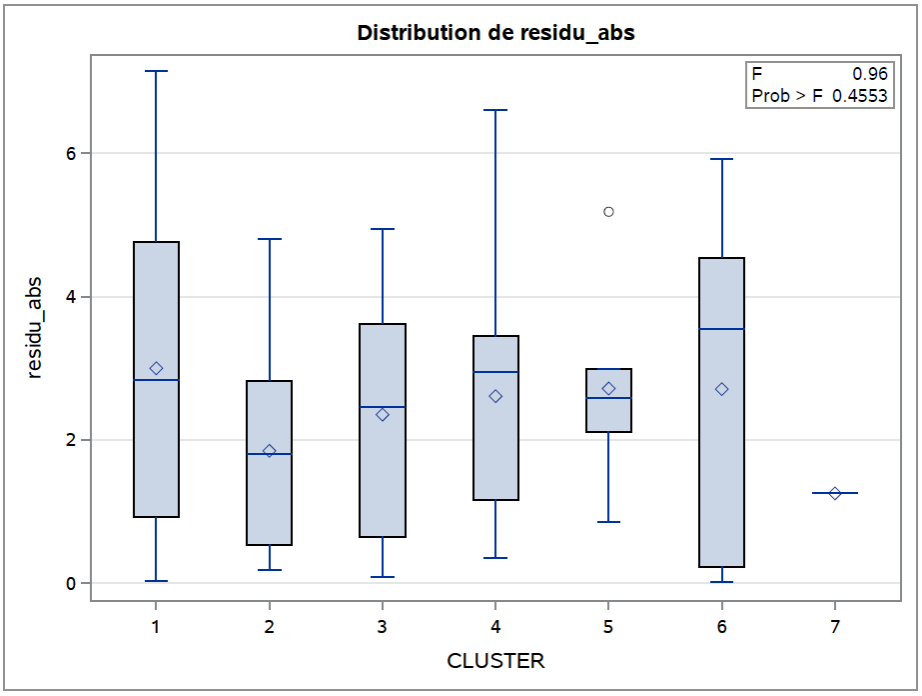
Attachements 8 : Distribution and Courbe Q-Q of absolut residuals on AEEH





Source : Issued by SAS of the proceedings the ANOVA of PCA on AEEH

Attachements 9 : Plot of absolut residuals on AEEH



Source : Issued by SAS of the proceedings the ANOVA of PCA on AEEH

Attachements 10 : Student Test of Tukey (HSD) on EN-12

Comparaisons significatives au niveau 0.05 indiquées par ***.				
CLUSTER Comparaison	Différence Entre les moyennes	Simultané 95% Intervalle de confiance		
5 - 1	0.2779	-3.1583	3.7141	
5 - 6	0.4368	-3.8898	4.7634	
5 - 3	1.6785	-2.2098	5.5669	
5 - 4	2.9998	-0.7231	6.7226	
5 - 2	3.2253	-0.4407	6.8913	
5 - 7	8.3628	2.0131	14.7125	***
1 - 5	-0.2779	-3.7141	3.1583	
1 - 6	0.1589	-3.0610	3.3788	
1 - 3	1.4006	-1.2009	4.0021	
1 - 4	2.7218	0.3750	5.0687	***
1 - 2	2.9474	0.6917	5.2030	***
1 - 7	8.0849	2.4309	13.7388	***
6 - 5	-0.4368	-4.7634	3.8898	
6 - 1	-0.1589	-3.3788	3.0610	
6 - 3	1.2417	-2.4569	4.9403	
6 - 4	2.5629	-0.9612	6.0871	
6 - 2	2.7885	-0.6756	6.2525	
6 - 7	7.9260	1.6907	14.1612	***
3 - 5	-1.6785	-5.5669	2.2098	
3 - 1	-1.4006	-4.0021	1.2009	
3 - 6	-1.2417	-4.9403	2.4569	
3 - 4	1.3212	-1.6486	4.2910	
3 - 2	1.5467	-1.3515	4.4450	
3 - 7	6.6843	0.7447	12.6238	***
4 - 5	-2.9998	-6.7226	0.7231	
4 - 1	-2.7218	-5.0687	-0.3750	***

Comparaisons significatives au niveau 0.05 indiquées par ***.				
CLUSTER Comparaison	Différence Entre les moyennes	Simultané 95% Intervalle de confiance		
4 - 6	-2.5629	-6.0871	0.9612	
4 - 3	-1.3212	-4.2910	1.6486	
4 - 2	0.2255	-2.4465	2.8976	
4 - 7	5.3630	-0.4695	11.1956	
2 - 5	-3.2253	-6.8913	0.4407	
2 - 1	-2.9474	-5.2030	-0.6917	***
2 - 6	-2.7885	-6.2525	0.6756	
2 - 3	-1.5467	-4.4450	1.3515	
2 - 4	-0.2255	-2.8976	2.4465	
2 - 7	5.1375	-0.6589	10.9339	
7 - 5	-8.3628	-14.7125	-2.0131	***
7 - 1	-8.0849	-13.7388	-2.4309	***
7 - 6	-7.9260	-14.1612	-1.6907	***
7 - 3	-6.6843	-12.6238	-0.7447	***
7 - 4	-5.3630	-11.1956	0.4695	
7 - 2	-5.1375	-10.9339	0.6589	

Source : Issued by SAS of the proceedings the ANOVA of PCA on EN3-12

Attachements 11 : Test modulus maximum (GT2) on EN3-12

Comparaisons significatives au niveau 0.05 indiquées par ***.				
CLUSTER Comparaison	Différence Entre les moyennes	95% Intervalle de confiance		
5 - 1	0.2779	-3.2695	3.8254	
5 - 6	0.4368	-4.0298	4.9034	
5 - 3	1.6785	-2.3357	5.6928	
5 - 4	2.9998	-0.8436	6.8431	
5 - 2	3.2253	-0.5594	7.0099	
5 - 7	8.3628	1.8076	14.9180	***
1 - 5	-0.2779	-3.8254	3.2695	
1 - 6	0.1589	-3.1652	3.4830	
1 - 3	1.4006	-1.2851	4.0863	
1 - 4	2.7218	0.2990	5.1447	***
1 - 2	2.9474	0.6187	5.2760	***
1 - 7	8.0849	2.2479	13.9218	***
6 - 5	-0.4368	-4.9034	4.0298	
6 - 1	-0.1589	-3.4830	3.1652	
6 - 3	1.2417	-2.5766	5.0600	
6 - 4	2.5629	-1.0753	6.2012	
6 - 2	2.7885	-0.7877	6.3646	
6 - 7	7.9260	1.4889	14.3631	***
3 - 5	-1.6785	-5.6928	2.3357	
3 - 1	-1.4006	-4.0863	1.2851	
3 - 6	-1.2417	-5.0600	2.5766	
3 - 4	1.3212	-1.7447	4.3871	
3 - 2	1.5467	-1.4453	4.5388	
3 - 7	6.6843	0.5524	12.8161	***
4 - 5	-2.9998	-6.8431	0.8436	
4 - 1	-2.7218	-5.1447	-0.2990	***

Comparaisons significatives au niveau 0.05 indiquées par ***.				
CLUSTER Comparaison	Différence Entre les moyennes	95% Intervalle de		
4 - 6	-2.5629	-6.2012	1.0753	
4 - 3	-1.3212	-4.3871	1.7447	
4 - 2	0.2255	-2.5330	2.9840	
4 - 7	5.3630	-0.6583	11.3844	
2 - 5	-3.2253	-7.0099	0.5594	
2 - 1	-2.9474	-5.2760	-0.6187	***
2 - 6	-2.7885	-6.3646	0.7877	
2 - 3	-1.5467	-4.5388	1.4453	
2 - 4	-0.2255	-2.9840	2.5330	
2 - 7	5.1375	-0.8466	11.1216	
7 - 5	-8.3628	-14.9180	-1.8076	***
7 - 1	-8.0849	-13.9218	-2.2479	***
7 - 6	-7.9260	-14.3631	-1.4889	***
7 - 3	-6.6843	-12.8161	-0.5524	***
7 - 4	-5.3630	-11.3844	0.6583	
7 - 2	-5.1375	-11.1216	0.8466	

Source : Issued by SAS of the proceedings the ANOVA of PCA on EN3-12

Attachements 12 : T Tests of Bonferroni (Dunn) on EN3-12

Comparaisons significatives au niveau 0.05 indiquées par ***.				
CLUSTER Comparaison	Différence Entre les moyennes	Simultané 95% Intervalle de confiance		
5 - 1	0.2779	-3.2847	3.8405	
5 - 6	0.4368	-4.0489	4.9225	
5 - 3	1.6785	-2.3529	5.7099	
5 - 4	2.9998	-0.8600	6.8595	
5 - 2	3.2253	-0.5755	7.0261	
5 - 7	8.3628	1.7796	14.9460	***
1 - 5	-0.2779	-3.8405	3.2847	
1 - 6	0.1589	-3.1794	3.4972	
1 - 3	1.4006	-1.2966	4.0978	
1 - 4	2.7218	0.2886	5.1550	***
1 - 2	2.9474	0.6088	5.2859	***
1 - 7	8.0849	2.2230	13.9467	***
6 - 5	-0.4368	-4.9225	4.0489	
6 - 1	-0.1589	-3.4972	3.1794	
6 - 3	1.2417	-2.5929	5.0763	
6 - 4	2.5629	-1.0908	6.2167	
6 - 2	2.7885	-0.8030	6.3799	
6 - 7	7.9260	1.4614	14.3906	***
3 - 5	-1.6785	-5.7099	2.3529	
3 - 1	-1.4006	-4.0978	1.2966	
3 - 6	-1.2417	-5.0763	2.5929	
3 - 4	1.3212	-1.7578	4.4002	
3 - 2	1.5467	-1.4581	4.5516	
3 - 7	6.6843	0.5262	12.8423	***
4 - 5	-2.9998	-6.8595	0.8600	
4 - 1	-2.7218	-5.1550	-0.2886	***

Comparaisons significatives au niveau 0.05 indiquées par ***.				
CLUSTER Comparaison	Différence Entre les moyennes	Simultané 95% Intervalle de confiance		
4 - 6	-2.5629	-6.2167	1.0908	
4 - 3	-1.3212	-4.4002	1.7578	
4 - 2	0.2255	-2.5448	2.9958	
4 - 7	5.3630	-0.6841	11.4101	
2 - 5	-3.2253	-7.0261	0.5755	
2 - 1	-2.9474	-5.2859	-0.6088	***
2 - 6	-2.7885	-6.3799	0.8030	
2 - 3	-1.5467	-4.5516	1.4581	
2 - 4	-0.2255	-2.9958	2.5448	
2 - 7	5.1375	-0.8721	11.1471	
7 - 5	-8.3628	-14.9460	-1.7796	***
7 - 1	-8.0849	-13.9467	-2.2230	***
7 - 6	-7.9260	-14.3906	-1.4614	***
7 - 3	-6.6843	-12.8423	-0.5262	***
7 - 4	-5.3630	-11.4101	0.6841	
7 - 2	-5.1375	-11.1471	0.8721	

Source : Issued by SAS of the proceedings the ANOVA of PCA on EN3-12

Attachements 13 : Proc univariate of 7 Clustere EN3-12

Moments			
N	96	**Somme des poids**	96
Moyenne	0	**Somme des observations**	0
Ecart-type	2.49510786	**Variance**	6.22556324
Skewness	0.17414871	**Kurtosis**	0.31476895
Somme des carrés non corrigée	591.428508	**Somme des carrés corrigée**	591.428508
Coeff Variation	.	**Std Error Mean**	0.25465588

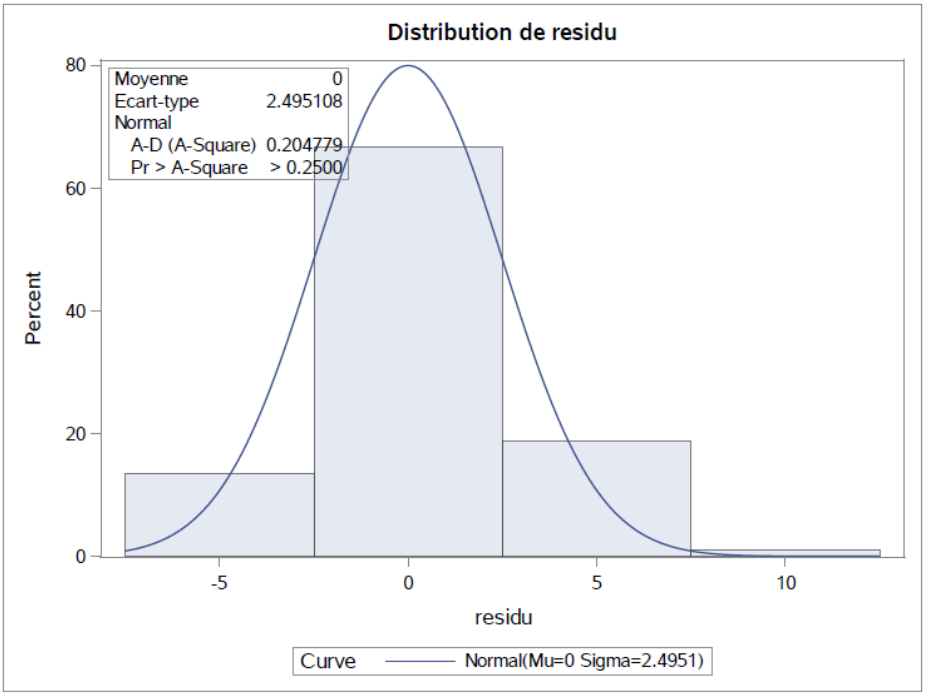
Mesures statistiques de base			
Emplacement		**Variabilité**	
Moyenne	0.00000	**Ecart-type**	2.49511
Médiane	-0.05687	**Variance**	6.22556
Mode	.	**Intervalle**	14.18677
		Ecart interquartile	3.57770

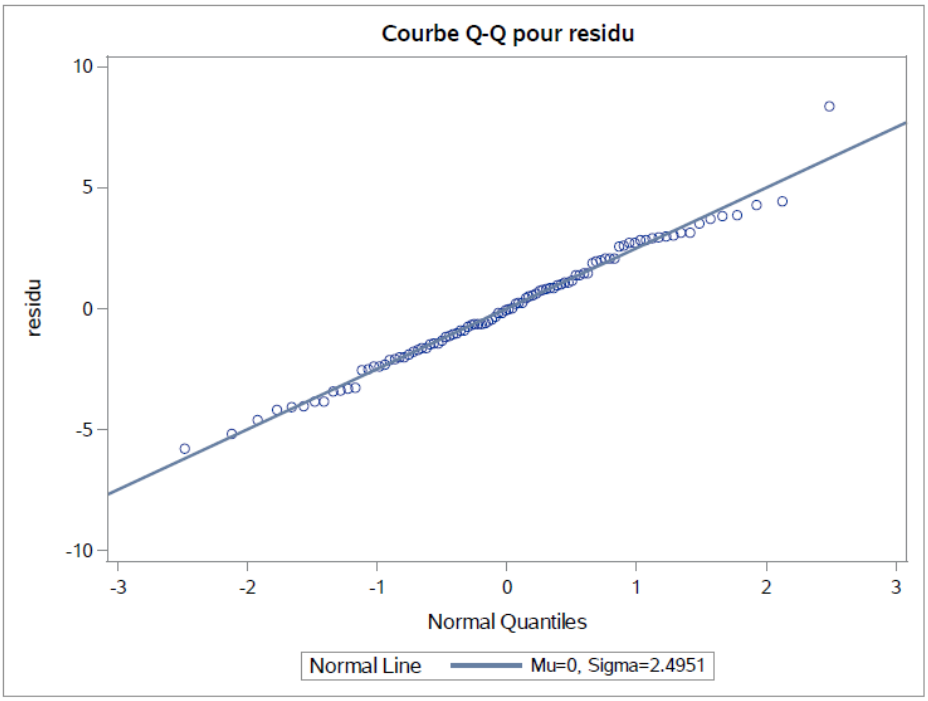
Tests de tendance centrale : Mu0=0				
Test	**Statistique**		**P-value**	
t de Student	**t**	0	**Pr > |t|**	1.0000
Signe	**M**	-1	**Pr >= |M|**	0.9188
Rang signé	**S**	-10.5	**Pr >= |S|**	0.9696

Tests de normalité				
Test	**Statistique**		**P-value**	
Shapiro-Wilk	**W**	0.988066	**Pr < W**	0.5430
Kolmogorov-Smirnov	**D**	0.044181	**Pr > D**	>0.1500
Cramer-von Mises	**W-Sq**	0.02077	**Pr > W-Sq**	>0.2500
Anderson-Darling	**A-Sq**	0.204779	**Pr > A-Sq**	>0.2500

Source : Issued by SAS of the proceedings the ANOVA of PCA on EN3-12

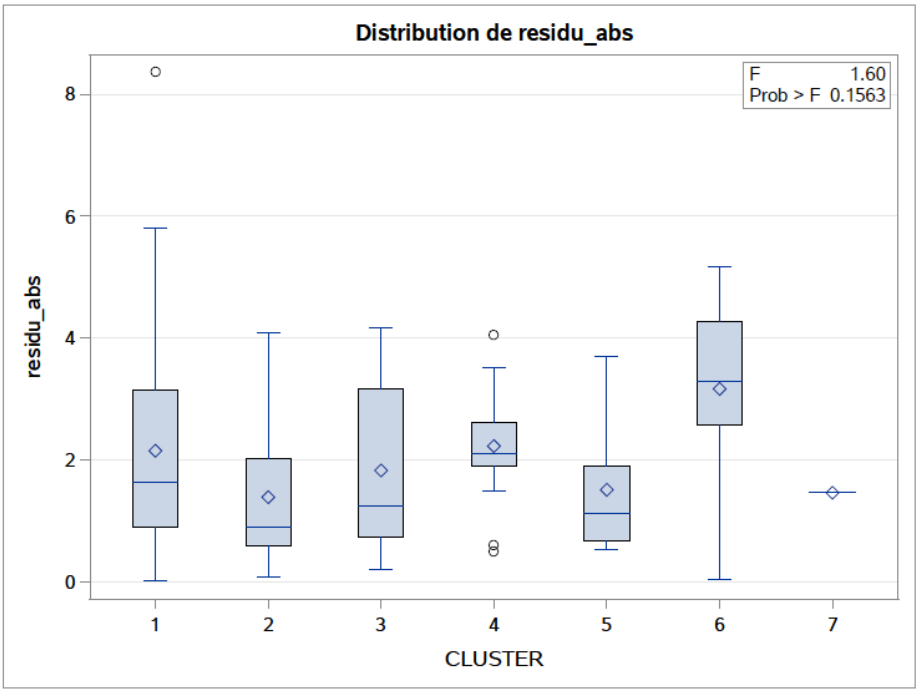
Attachements 14 : Distribution eand Courbe Q-Q of absolut residuals on EN3-12





Source : Issued by SAS of the proceedings the ANOVA of PCA on EN3-12

Attachements 15 : Plot of absolut residuals on EN3-12

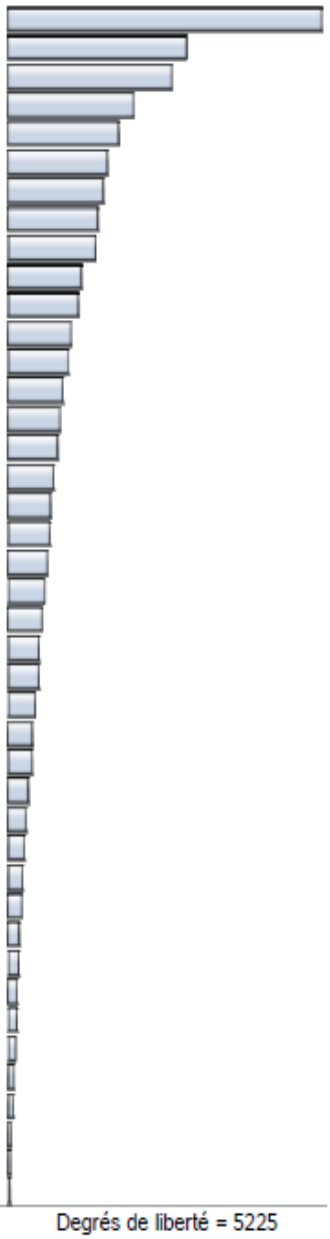


Source : Issued by SAS of the proceedings the ANOVA of PCA on EN3-12

Attachements 16 : The inertie of MCA

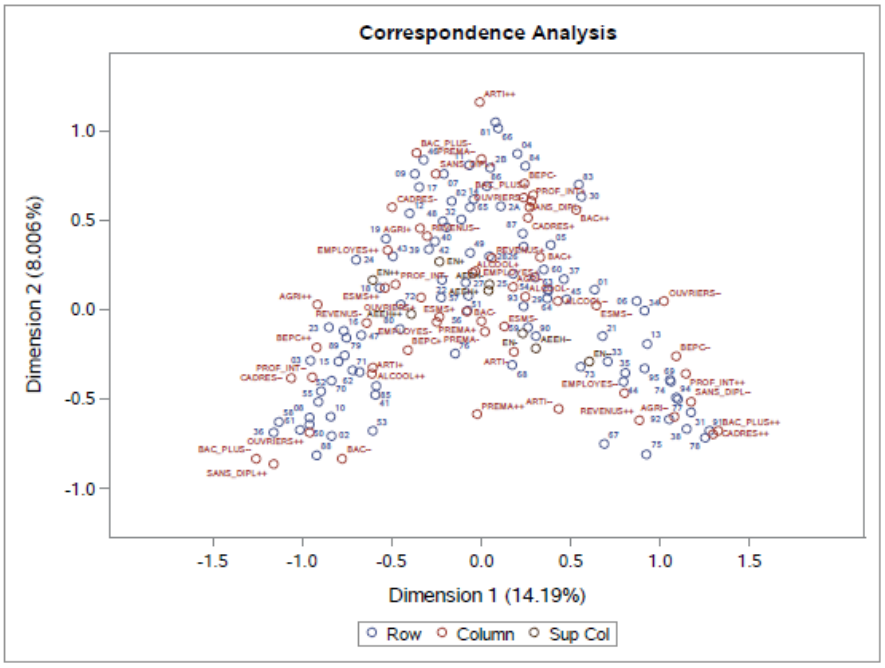
Décomposition de l'inertie et du Khi-2

Valeur singulière	Inertie principale	Khi-2	Pourcentage	Pourcentage cumulé
0.65255	0.42582	572.30	14.19	14.19
0.49009	0.24019	322.81	8.01	22.20
0.47166	0.22246	298.99	7.42	29.62
0.41321	0.17074	229.48	5.69	35.31
0.38619	0.14914	200.45	4.97	40.28
0.36688	0.13460	180.90	4.49	44.76
0.35875	0.12870	172.97	4.29	49.05
0.34908	0.12186	163.77	4.06	53.12
0.34316	0.11776	158.26	3.93	57.04
0.31634	0.10007	134.49	3.34	60.38
0.30718	0.09436	126.82	3.15	63.52
0.29282	0.08574	115.24	2.86	66.38
0.28558	0.08155	109.61	2.72	69.10
0.27255	0.07428	99.84	2.48	71.58
0.26584	0.07067	94.98	2.36	73.93
0.25943	0.06731	90.46	2.24	76.18
0.24888	0.06194	83.25	2.06	78.24
0.24107	0.05811	78.10	1.94	80.18
0.23717	0.05625	75.60	1.87	82.05
0.23056	0.05316	71.44	1.77	83.82
0.22138	0.04901	65.87	1.63	85.46
0.21590	0.04661	62.65	1.55	87.01
0.20418	0.04169	56.03	1.39	88.40
0.20346	0.04140	55.64	1.38	89.78
0.19005	0.03612	48.54	1.20	90.98
0.18302	0.03350	45.02	1.12	92.10
0.18063	0.03263	43.85	1.09	93.19
0.16665	0.02777	37.33	0.93	94.11
0.15804	0.02498	33.57	0.83	94.95
0.15083	0.02275	30.57	0.76	95.71
0.14115	0.01992	26.78	0.66	96.37
0.13803	0.01905	25.61	0.64	97.00
0.12420	0.01543	20.73	0.51	97.52
0.11920	0.01421	19.10	0.47	97.99
0.11426	0.01306	17.55	0.44	98.43
0.10707	0.01146	15.41	0.38	98.81
0.09892	0.00978	13.15	0.33	99.14
0.08966	0.00804	10.80	0.27	99.40
0.08453	0.00715	9.60	0.24	99.64
0.06539	0.00428	5.75	0.14	99.78
0.06136	0.00377	5.06	0.13	99.91
0.05195	0.00270	3.63	0.09	100.00
	3.00000	4032.00	100.00	

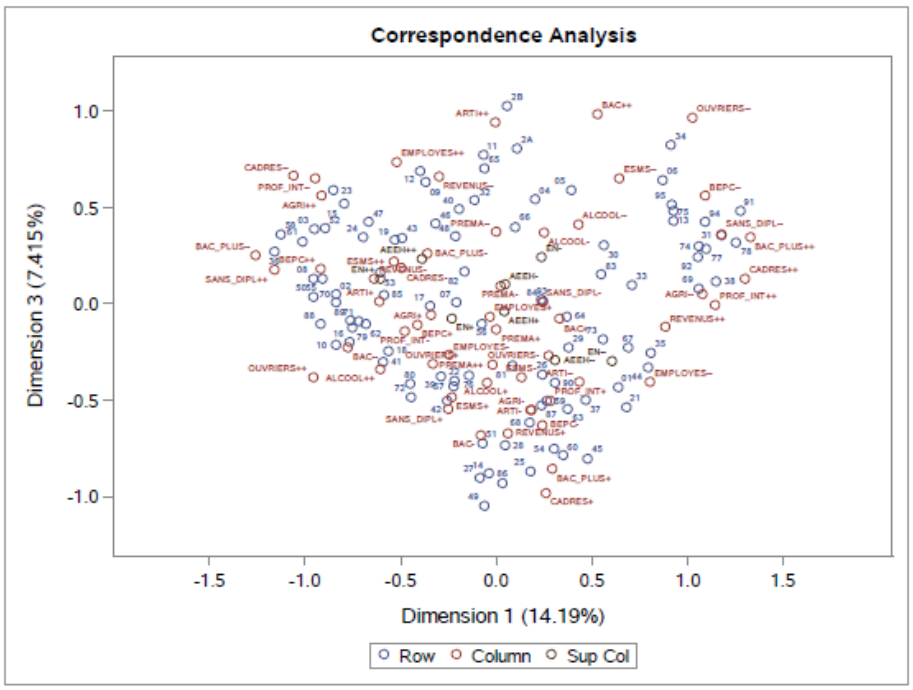


Source : Issued by SAS of the proceedings the MCA

Attachements 17 : The Individuals and variables on axes 1 and 2

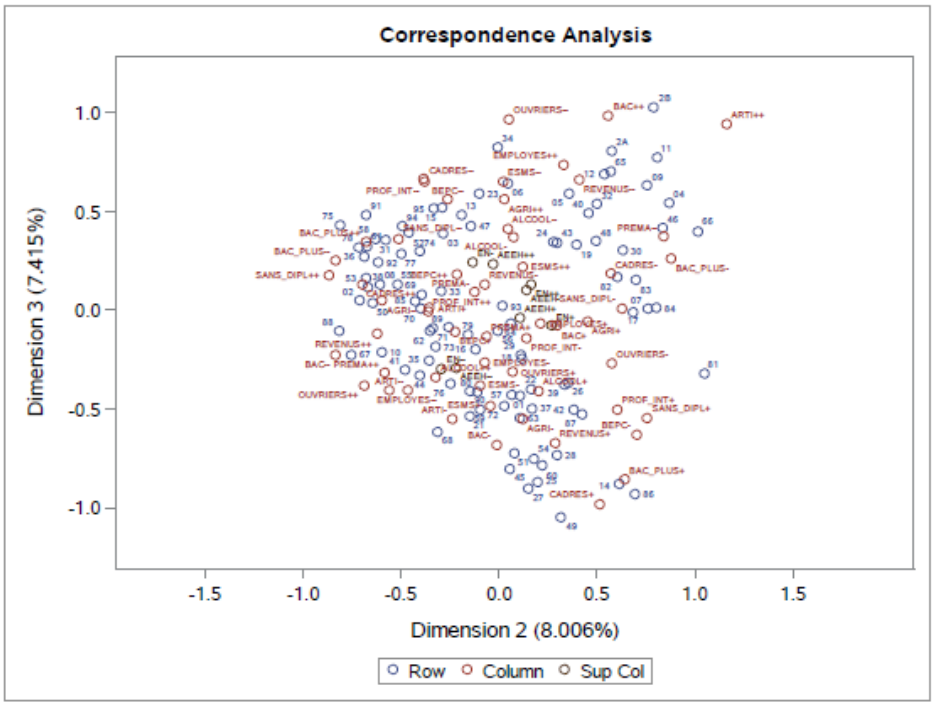


Attachements 18 : The Individuals and variables on axes 1 and 3



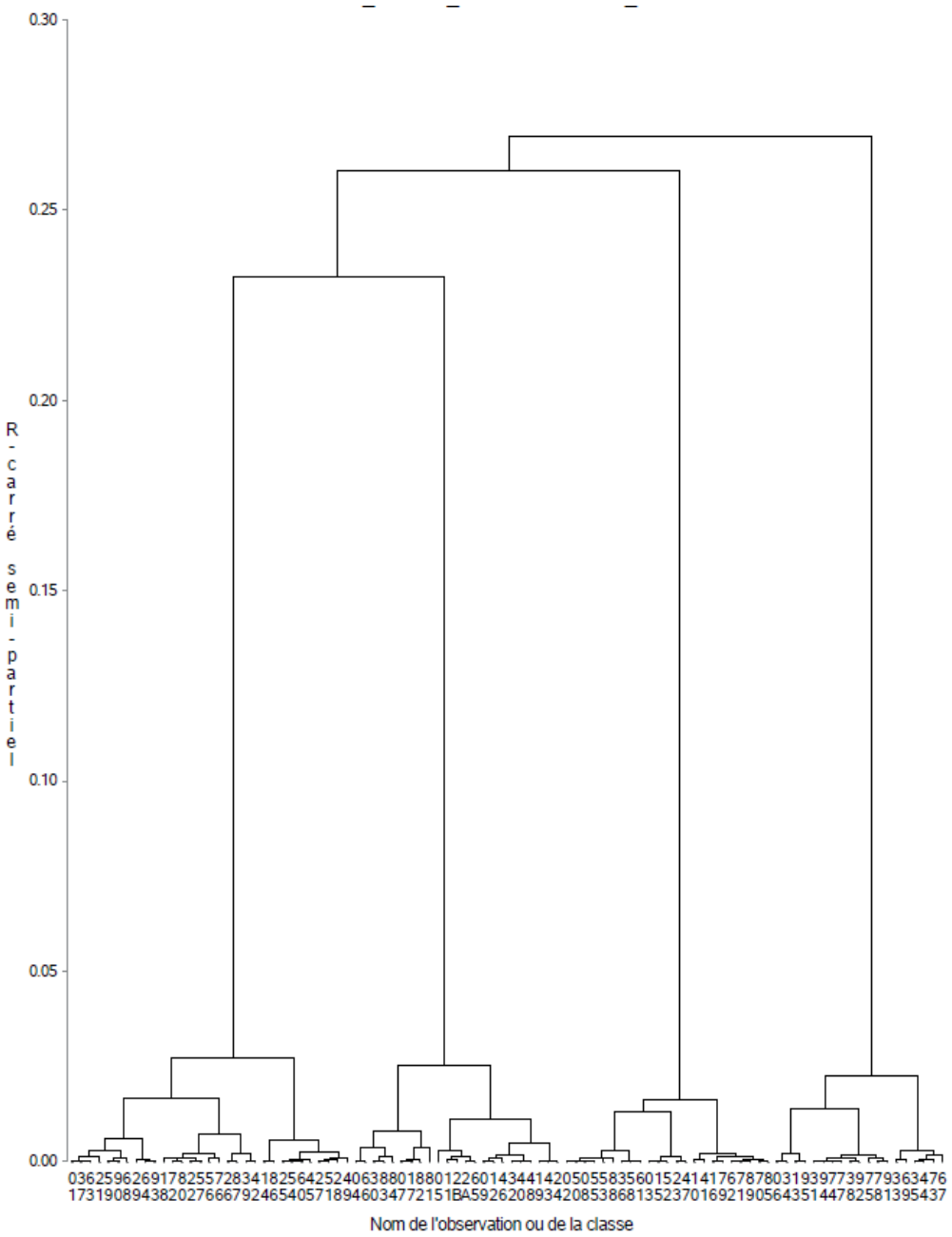
Source : Issued by SAS of the proceedings the MCA

Attachements 19 : The Individuals and variables on axes 2 and 3



Source : Issued by SAS of the proceedings the MCA

Attachements 20 : Analyse of clusterifications



Source : Issued by SAS of the proceedings the HCA of MCA

Attachements 21 : Clusterification of MCA

CLUSTER=1

dep	Dim1	Dim2	Dim3
02	-0.83785	-0.71013	0.04833
03	-0.95180	-0.28652	0.38910
08	-0.95703	-0.60722	0.13085
10	-0.83909	-0.59907	-0.21217
15	-0.79852	-0.29111	0.51830
16	-0.76822	-0.12151	-0.19934
23	-0.85309	-0.09971	0.58998
36	-1.16200	-0.68814	0.27158
41	-0.59278	-0.48067	-0.30207
47	-0.66868	-0.14351	0.42633
50	-0.95939	-0.64745	0.03660
52	-0.89484	-0.46050	0.39339
53	-0.60597	-0.67816	0.16227
55	-0.91008	-0.52047	0.12865
58	-1.12897	-0.62969	0.36007
61	-1.01316	-0.67573	0.32425
62	-0.68171	-0.35107	-0.10537
70	-0.83742	-0.40125	0.00582
71	-0.72027	-0.33972	-0.09079
79	-0.75282	-0.15808	-0.12381
85	-0.58737	-0.42934	0.04751
88	-0.91937	-0.81777	-0.10333
89	-0.76127	-0.25763	-0.08560

CLUSTER=2

dep	Dim1	Dim2	Dim3
01	0.63572	0.10805	-0.43455
14	-0.04165	0.61369	-0.88107
18	-0.56650	0.11707	-0.24499
21	0.68041	-0.15070	-0.53415
22	-0.21909	0.16164	-0.40209
25	0.17641	0.19567	-0.87043
26	0.23915	0.35120	-0.36659
27	-0.08544	0.15060	-0.90289
28	0.04536	0.29574	-0.73500
29	0.37539	0.11120	-0.22866
37	0.46497	0.16829	-0.49694
39	-0.29054	0.33451	-0.37860
42	-0.25570	0.38180	-0.50413
45	0.47609	0.05539	-0.80326
49	-0.06153	0.31647	-1.04879
51	-0.07317	0.07584	-0.72318
54	0.29974	0.17855	-0.74979
56	-0.07565	-0.00747	-0.10257
57	-0.22535	0.06360	-0.42958
59	0.26216	-0.09950	-0.50140
60	0.34909	0.22067	-0.78687
63	0.37431	0.10556	-0.54633
64	0.36785	0.05960	-0.06508
68	0.17564	-0.31454	-0.61733
72	-0.44694	0.02580	-0.48583
76	-0.14416	-0.25071	-0.37112
80	-0.45165	-0.11154	-0.41210
86	0.03195	0.69060	-0.92989
87	0.23320	0.42248	-0.52537
90	0.30418	-0.14901	-0.41146
93	0.23683	0.01788	0.02326

CLUSTER=3

dep	Dim1	Dim2	Dim3
04	0.20174	0.86791	0.54273
05	0.39165	0.35846	0.58764
07	-0.21026	0.75689	0.00832
09	-0.37189	0.75589	0.63256
11	-0.06838	0.80753	0.77078
12	-0.39990	0.53655	0.68991
17	-0.34711	0.68118	-0.01226
19	-0.53157	0.39655	0.32930
24	-0.69747	0.27775	0.34617
2A	0.10784	0.57380	0.80286
2B	0.05337	0.78977	1.02611
30	0.55923	0.63169	0.30195
32	-0.11287	0.49981	0.53832
40	-0.19423	0.45860	0.49210
43	-0.49450	0.29493	0.34289
46	-0.31922	0.83388	0.41459
48	-0.21354	0.49255	0.34857
65	-0.06210	0.57085	0.70219
66	0.09671	1.01307	0.39921
81	0.08256	1.04706	-0.32268
82	-0.16557	0.60269	0.16807
83	0.54757	0.69764	0.15146
84	0.24558	0.80059	0.01085

CLUSTER=4

dep	Dim1		Dim2	Dim3
06	0.86977		0.04669	0.64002
13	0.92727		-0.19261	0.48146
31	1.17616		-0.57894	0.35403
33	0.70618		-0.29447	0.09470
34	0.91230		-0.00682	0.82411
35	0.80551		-0.35616	-0.25722
38	1.14923		-0.66965	0.11790
44	0.79436		-0.40591	-0.32799
67	0.68814		-0.75385	-0.22507
69	1.05659		-0.39658	0.07637
73	0.55726		-0.32333	-0.18295
74	1.05931		-0.40357	0.29971
75	0.92554		-0.81339	0.42925
77	1.09801		-0.50169	0.28391
78	1.25378		-0.71699	0.31875
91	1.27548		-0.67847	0.47917
92	1.05305		-0.61722	0.24099
94	1.09016		-0.49292	0.42304
95	0.91492		-0.33372	0.51441

The average

	cluster			
	1	**2**	**3**	**4**
	moyenne	**moyenne**	**moyenne**	**moyenne**
Dim1	-0.83	0.09	-0.08	0.96
Dim2	-0.45	0.13	0.64	-0.45
Dim3	0.11	-0.53	0.40	0.24

Source : Issued by SAS of the proceedings the HCA of MCA

Attachements 22 : Test of Tukey (HSD) sur taux_AEEH

Comparaisons significatives au niveau 0.05 indiquées par ***.				
CLUSTER Comparaison	Différence Entre les moyennes	Simultané 95% Intervalle de confiance		
1 - 3	0.9021	-1.6482	3.4524	
1 - 2	2.5407	0.1606	4.9208	***
1 - 4	3.1516	0.4704	5.8328	***
3 - 1	-0.9021	-3.4524	1.6482	
3 - 2	1.6386	-0.7415	4.0187	
3 - 4	2.2495	-0.4317	4.9307	
2 - 1	-2.5407	-4.9208	-0.1606	***
2 - 3	-1.6386	-4.0187	0.7415	
2 - 4	0.6109	-1.9089	3.1308	
4 - 1	-3.1516	-5.8328	-0.4704	***
4 - 3	-2.2495	-4.9307	0.4317	
4 - 2	-0.6109	-3.1308	1.9089	

Attachements 23 : Test of modulus maximum (GT2) on AEEH

1.3.1

Comparaisons significatives au niveau 0.05 indiquées par ***.				
CLUSTER Comparaison	Différence Entre les moyennes	95% Intervalle de confiance		
1 - 3	0.9021	-1.7159	3.5201	
1 - 2	2.5407	0.0974	4.9840	***
1 - 4	3.1516	0.3992	5.9040	***
3 - 1	-0.9021	-3.5201	1.7159	
3 - 2	1.6386	-0.8047	4.0819	
3 - 4	2.2495	-0.5029	5.0019	
2 - 1	-2.5407	-4.9840	-0.0974	***
2 - 3	-1.6386	-4.0819	0.8047	
2 - 4	0.6109	-1.9758	3.1977	
4 - 1	-3.1516	-5.9040	-0.3992	***
4 - 3	-2.2495	-5.0019	0.5029	
4 - 2	-0.6109	-3.1977	1.9758	

Source : Issued by SAS of the proceedings the ANOVA of MCA on AEEH

Attachements 24 : T Tests of Bonferroni (Dunn) on AEEH

Comparaisons significatives au niveau 0.05 indiquées par ***.				
CLUSTER Comparaison	Différence Entre les moyennes	Simultané 95% Intervalle de confiance		
1 - 3	0.9021	-1.7261	3.5303	
1 - 2	2.5407	0.0879	4.9934	***
1 - 4	3.1516	0.3886	5.9147	***
3 - 1	-0.9021	-3.5303	1.7261	
3 - 2	1.6386	-0.8142	4.0913	
3 - 4	2.2495	-0.5135	5.0125	
2 - 1	-2.5407	-4.9934	-0.0879	***
2 - 3	-1.6386	-4.0913	0.8142	
2 - 4	0.6109	-1.9858	3.2077	
4 - 1	-3.1516	-5.9147	-0.3886	***
4 - 3	-2.2495	-5.0125	0.5135	
4 - 2	-0.6109	-3.2077	1.9858	

Attachements 25 : Proc univariate residual of 4 Clustere AEEH

Moments			
N	96	Somme des poids	96
Moyenne	0	Somme des observations	0
Ecart-type	3.25263171	Variance	10.579613
Skewness	0.29398059	Kurtosis	-0.5305621
Somme des carrés non corrigée	1005.06324	Somme des carrés corrigée	1005.06324
Coeff Variation	.	Std Error Mean	0.33197033

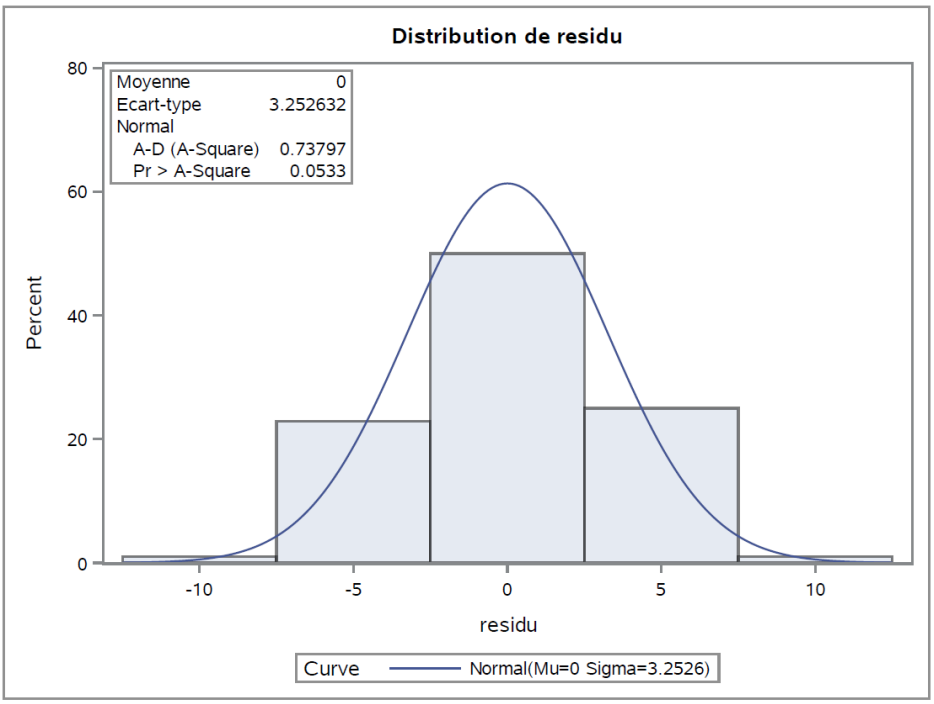
Mesures statistiques de base			
Emplacement		Variabilité	
Moyenne	0.00000	Ecart-type	3.25263
Médiane	-0.42878	Variance	10.57961
Mode	.	Intervalle	16.20728
		Ecart interquartile	5.05594

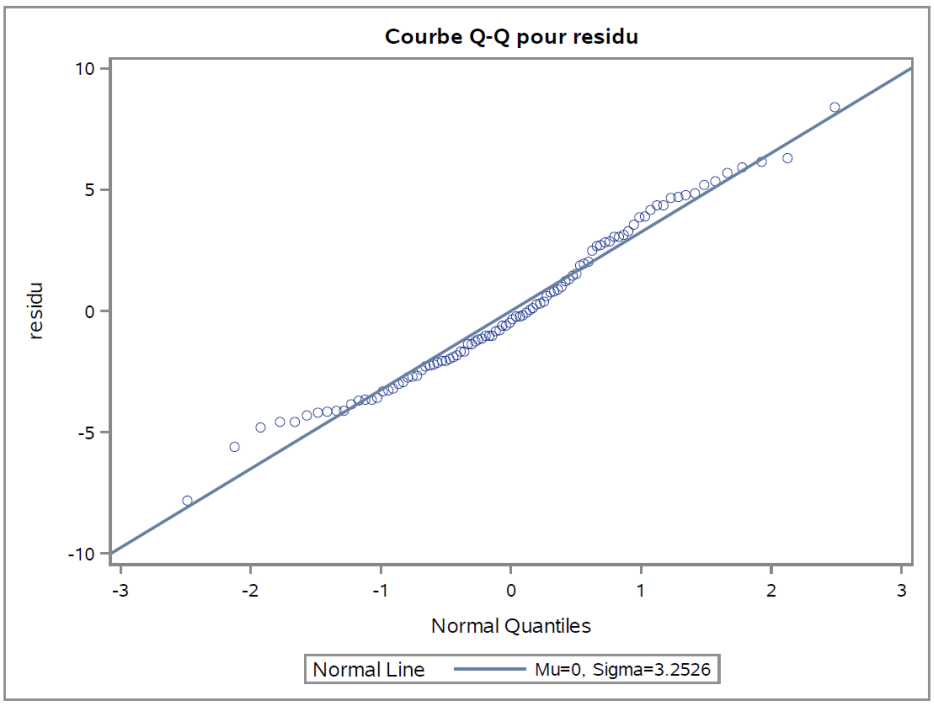
Tests de tendance centrale : Mu0=0				
Test	Statistique		P-value	
t de Student	t	0	Pr > |t|	1.0000
Signe	M	-5	Pr >= |M|	0.3584
Rang signé	S	-60	Pr >= |S|	0.8278

Tests de normalité				
Test	Statistique		P-value	
Shapiro-Wilk	W	0.979244	Pr < W	0.1320
Kolmogorov-Smirnov	D	0.071309	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.120211	Pr > W-Sq	0.0617
Anderson-Darling	A-Sq	0.73797	Pr > A-Sq	0.0533

Source : Issued by SAS of the proceedings the ANOVA of MCA on AEEH

Attachements 26 : Distribution and Courbe Q-Q of absolut residuals on AEEH





Source : Issued by SAS of the proceedings the ANOVA of MCA on AEEH

Attachements 27 : Plot of absolut residuals on AEEH

Source : Issued by SAS of the proceedings the ANOVA of MCA on AEEH

Attachements 28 : Test of Tukey (HSD) on EN3-12

Comparaisons significatives au niveau 0.05 indiquées par ***.				
CLUSTER Comparaison	Différence Entre les moyennes	Simultané 95% Intervalle de confiance		
3 - 1	0.0366	-1.9958	2.0690	
3 - 2	1.9247	0.0279	3.8215	***
3 - 4	3.8404	1.7037	5.9771	***
1 - 3	-0.0366	-2.0690	1.9958	
1 - 2	1.8881	-0.0087	3.7849	
1 - 4	3.8038	1.6670	5.9405	***
2 - 3	-1.9247	-3.8215	-0.0279	***
2 - 1	-1.8881	-3.7849	0.0087	
2 - 4	1.9157	-0.0925	3.9238	
4 - 3	-3.8404	-5.9771	-1.7037	***
4 - 1	-3.8038	-5.9405	-1.6670	***
4 - 2	-1.9157	-3.9238	0.0925	

Attachements 29 : Test of modulus maximum (GT2) on EN3-12

Comparaisons significatives au niveau 0.05 indiquées par ***.				
CLUSTER Comparaison	Différence Entre les moyennes	95% Intervalle de confiance		
3 - 1	0.0366	-2.0498	2.1230	
3 - 2	1.9247	-0.0224	3.8718	
3 - 4	3.8404	1.6469	6.0338	***
1 - 3	-0.0366	-2.1230	2.0498	
1 - 2	1.8881	-0.0590	3.8352	
1 - 4	3.8038	1.6103	5.9972	***
2 - 3	-1.9247	-3.8718	0.0224	
2 - 1	-1.8881	-3.8352	0.0590	
2 - 4	1.9157	-0.1458	3.9771	
4 - 3	-3.8404	-6.0338	-1.6469	***
4 - 1	-3.8038	-5.9972	-1.6103	***
4 - 2	-1.9157	-3.9771	0.1458	

Source : Issued by SAS of the proceedings the ANOVA of MCA on EN3-12

Attachements 30 : T Tests of Bonferroni (Dunn) on EN3-12

Comparaisons significatives au niveau 0.05 indiquées par ***.				
CLUSTER Comparaison	Différence Entre les moyennes	Simultané 95% Intervalle de confiance		
3 - 1	0.0366	-2.0579	2.1311	
3 - 2	1.9247	-0.0300	3.8794	
3 - 4	3.8404	1.6384	6.0423	***
1 - 3	-0.0366	-2.1311	2.0579	
1 - 2	1.8881	-0.0666	3.8428	
1 - 4	3.8038	1.6018	6.0057	***
2 - 3	-1.9247	-3.8794	0.0300	
2 - 1	-1.8881	-3.8428	0.0666	
2 - 4	1.9157	-0.1538	3.9851	
4 - 3	-3.8404	-6.0423	-1.6384	***
4 - 1	-3.8038	-6.0057	-1.6018	***
4 - 2	-1.9157	-3.9851	0.1538	

Attachements 31 : Proc univariate residuals of 4 Clustere EN3-12

Moments			
N	96	Somme des poids	96
Moyenne	0	Somme des observations	0
Ecart-type	2.5921256	Variance	6.7191151
Skewness	0.24487385	Kurtosis	0.51382877
Somme des carrés non corrigée	638.315935	Somme des carrés corrigée	638.315935
Coeff Variation	.	Std Error Mean	0.26455771

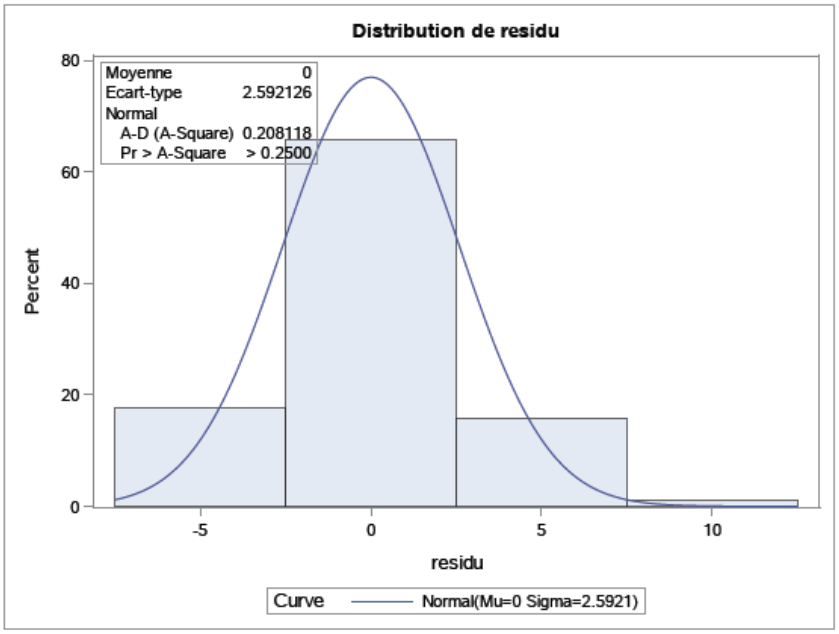
Mesures statistiques de base			
Emplacement		Variabilité	
Moyenne	0.00000	Ecart-type	2.59213
Médiane	-0.04283	Variance	6.71912
Mode	.	Intervalle	14.18670
		Ecart interquartile	3.21054

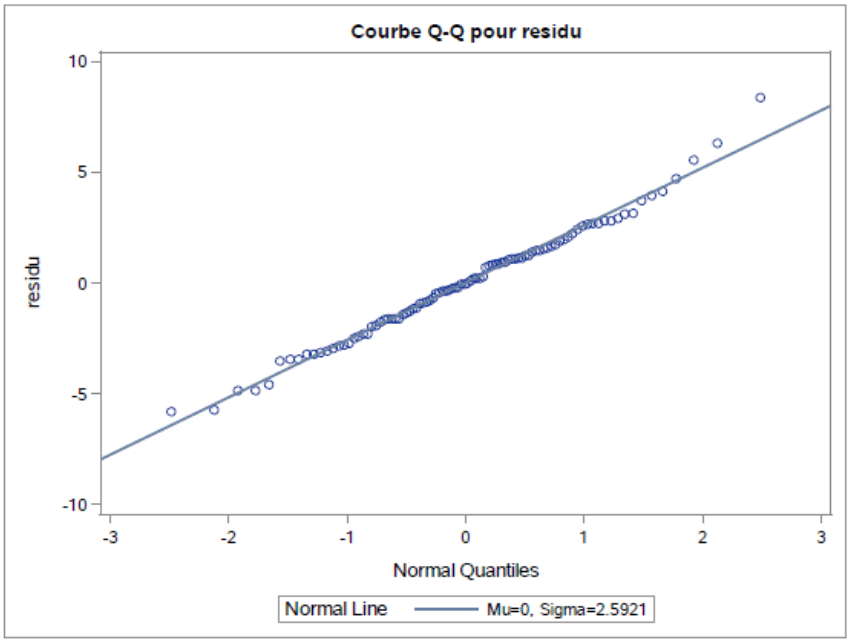
Tests de tendance centrale : Mu0=0				
Test	Statistique		P-value	
t de Student	t	0	Pr > |t|	1.0000
Signe	M	-1	Pr >= |M|	0.9188
Rang signé	S	-47	Pr >= |S|	0.8647

Tests de normalité				
Test	Statistique		P-value	
Shapiro-Wilk	W	0.989427	Pr < W	0.6466
Kolmogorov-Smirnov	D	0.044733	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.025604	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.208118	Pr > A-Sq	>0.2500

Source : Issued by SAS of the proceedings the ANOVA of MCA on EN3-12

Attachements 32 : Distribution and Courbe Q-Q of absolut residuals on EN3-12





Source : Issued by SAS of the proceedings the ANOVA of MCA on EN3-12

Attachements 33 : Plot of absolut residuals on EN3-12

Source : Issued by SAS of the proceedings the ANOVA of MCA on EN3-12

Attachements 34 : The Departments in France

Observation	Number of Region	Name of Departments
1	01	Ain
2	02	Aisne
3	03	Allier
4	04	Alpes-de-Haute-Provence
5	05	Hautes-Alpes
6	06	Alpes-Maritimes
7	07	Ardèche
8	08	Ardennes
9	09	Ariège
10	10	Aube
11	11	Aude
12	12	Aveyron
13	13	Bouches-du-Rhône
14	14	Calvados
15	15	Cantal
16	16	Charente
17	17	Charente-Maritime
18	18	Cher
19	19	Corrèze
20	21	Côte-d'Or
21	22	Côtes-d'Armor
22	23	Creuse
23	24	Dordogne
24	25	Doubs
25	26	Drôme
26	27	Eure
27	28	Eure-et-Loir
28	29	Finistère
29	2A	Corse-du-Sud
30	2B	Haute-Corse
31	30	Gard
32	31	Haute-Garonne
33	32	Gers
34	33	Gironde
35	34	Hérault
36	35	Ille-et-Vilaine
37	36	Indre

Observation	Number of Region	Name of Departments
38	37	Indre-et-Loire
39	38	Isère
40	39	Jura
41	40	Landes
42	41	Loir-et-Cher
43	42	Loire
44	43	Haute-Loire
45	44	Loire-Atlantique
46	45	Loiret
47	46	Lot
48	47	Lot-et-Garonne
49	48	Lozère
50	49	Maine-et-Loire
51	50	Manche
52	51	Marne
53	52	Haute-Marne
54	53	Mayenne
55	54	Meurthe-et-Moselle
56	55	Meuse
57	56	Morbihan
58	57	Moselle
59	58	Nièvre
60	59	Nord
61	60	Oise
62	61	Orne
63	62	Pas-de-Calais
64	63	Puy-de-Dôme
65	64	Pyrénées-Atlantiques
66	65	Hautes-Pyrénées
67	66	Pyrénées-Orientales
68	67	Bas-Rhin
69	68	Haut-Rhin
70	69	Rhône
71	70	Haute-Saône
72	71	Saône-et-Loire
73	72	Sarthe
74	73	Savoie
75	74	Haute-Savoie

Observation	Number of Region	Name of Departments
76	75	Paris
77	76	Seine-Maritime
78	77	Seine-et-Marne
79	78	Yvelines
80	79	Deux-Sèvres
81	80	Somme
82	81	Tarn
83	82	Tarn-et-Garonne
84	83	Var
85	84	Vaucluse
86	85	Vendée
87	86	Vienne
88	87	Haute-Vienne
89	88	Vosges
90	89	Yonne
91	90	Territoire de Belfort
92	91	Essonne
93	92	Hauts-de-Seine
94	93	Seine-Saint-Denis
95	94	Val-de-Marne
96	95	Val-d'Oise

BIOGRAPHY

Diah Meidatuzzahra was born in Mataram, Lombok on May 26, 1990. In 1996-2003, author studied in SDN 41 Primary School Mataram, and then 2003 to 2005, author studied junior high school in SMP Negri 1 Mataram. In 2005-2008, author studed High School in SMAN 5 Mataram.

Author really love mathe-matics, because mathematics is the basic knowledge to the other fields. Therefore, author majored in mathe-matics at Faculty of Mathematics and Natural Sciences (FMIPA), Mataram University (2008-2013). Then, author worked as a teacher in mathe-matics agency for primary school called Sinau Mataram (October 2013 to April 2014) and also worked as a presenter on local television, Lombok Post TV. Then, in 2014, author continue the study in FMIPA, Sepuluh Nopember Institute of Technology, Surabaya (ITS) with majors master Statistics.

At ITS, author follows the program of Double Degree Indonesia France (DDIP). It is a joint degree between ITS and several campus in France. Author studied at Faculty of Science, Aix-Marseille University, with majors Mathematics Application and Social Science (MASS) 2015/2016.

Alhamdulillah, in 2017, author has completed a study in ITS by collecting thesis entitled "Population Analysis of Disabled Children by Departments in France", which is the thesis of raport held in Aix-Marseille University

If you have question, suggestion, or need more discussion, you can contact via email : meyda.azzahra@gmail.com