Université de Thessalie
Département d'Aménagement,
D'Urbanisme et Développement Régional

Enseignant : As. Pr. Marie-Noelle Duquenne

II.Les Méthodes de Classification

Ces méthodes ont pour objectif de parvenir à définir une partition des individus en un nombre restreint de classes homogènes, permettant une typologie des individus considérés. Les classes sont obtenues à l'aide d'algorithmes formalisés et elles doivent décrire des types de comportements. Il s'agit en d'autres termes de faire apparaître des profils – types.

Les données se présentent sous la forme d'un tableau de n individus et k variables.

Nous travaillerons ici sur les méthodes de classification automatique et plus particulièrement sur les méthodes hiérarchiques et non hiérarchiques.

Dans tous les cas, il nous faut définir (a) un **critère de distance** qui reflète la dissemblance entre les individus, c'est-à-dire leur dissimilarité grâce auquel on pourra regrouper les individus et (b) une stratégie de classification.

2.1. La Classification Hiérarchique

Cette méthode repose sur le principe de la formation de partitions emboitées. Elle a pour but de former un ensemble de partitions de l'ensemble des n individus en classes de moins en moins fines. Ces classes sont obtenues par regroupements successifs des individus, en commençant par les individus les plus proches. A chaque étape de regroupement, le principe sera le même, réunir dans une classe, les individus les plus proches (objectif d'homogénéité des classes). Certes plus on avance dans le regroupement des individus et plus les individus présenteront des différences. Il faut donc tenir compte de cet aspect pour délimiter le nombre de classes finalement retenu.

Il ne s'agit pas d'une méthode manuelle mais d'une classification basée sur l'algorithme qui utilise un critère d'agrégation (regroupement des individus dans diverses classes), basé sur un critère de **distance**.

1. Définition du critère de distance

Le plus souvent on utilise la distance euclidienne usuelle. Supposons que nous observions k caractères (X_1, X_2 ... X_k) sur un ensemble d'individus, la distance euclidienne entre deux des individus du groupe (individus A et B) sera donnée par :

$$d^2_{A,B} = \sum_{j=1}^{k} (X_{A,j} - X_{B,j})^2$$

Considérons les 4 grandes villes de Thessalie pour lesquelles nous avons calculé leur densité ainsi que le nombre de résidents en 2001 qui vivaient déjà dans la même ville en 1995 pour 100 résidents en 2001. (Par exemple, pour 100 résidents à Larissa en 2001, 83 y vivaient déjà en 1995). Lorsque l'on regarde les données brutes (tableau de gauche), il est clair que les échelles de mesure sont très différentes et donc que les différences concernant la deuxième variable ne joueront quasiment aucun rôle dans le calcul de la distance. C'est pourquoi avons-nous centrés et réduites les deux variables (standardisation) pour prendre en compte les deux variables.

Variables initiales

Villes	Densite	Part en 95
Karditsa	376	86,2
Larissa	1083	83,2
Volos-Nea Ionia	570	84,6
Trikala	837	84,5

Variables centrées réduites

Villes	Densite	Part en 95
Karditsa	-1,27	1,50
Larissa	1,37	-1,32
Volos-Nea Ionia	-0,55	-0,04
Trikala	0,45	-0,14

A partir des deux variables centrées - réduites, nous pouvons calculer 2 à 2 les distances entre les villes.

La distance entre Karditsa et Larissa est[1] :

$$d^2_{K,L} = (-1{,}27 - 1{,}37)^2 + (1{,}50 - (-1{,}32))^2 = 14{,}9$$

Cette distance en elle-même ne nous dit rien, il faut la comparer avec les autres distances pour comprendre les similarités et dissimilarités.

La distance entre Karditsa et Volos est :

$$d^2_{K,V} = (-1{,}27 - (-0{,}55))^2 + (1{,}50 - (-0{,}04))^2 = 2{,}9$$

Calculées deux à deux les distances peuvent être retranscrites au travers de la matrice carrée dite matrice des proximités :

Villes	Karditsa	Larissa	Volos-Nea Ionia	Trikala
Karditsa	0,0	14,9	2,9	5,6
Larissa	14,9	0,0	5,3	2,2
Volos-Nea Ionia	2,9	5,3	0,0	1,0
Trikala	5,6	2,2	1,0	0,0

[1] Si nous n'avions pas normé les variables, la distance entre Karditsa et Larissa, basée sur les deux variables, serait de 499830 alors que la même distance en ne prenant en compte que la 1ère variable serait de 499821. Le rôle de la seconde variable est alors annulé. Ainsi, avec les variables brutes, on pourrait montrer facilement que la plus petite distance concerne Karditsa et Volos. Cela est dû au fait que l'impact de la 2ème variable n'est en fait pas pris en compte lorsque les variables ne sont pas normées.

Conduite de la Classification

La procédure comprend un ensemble d'étapes successives qui vont se répéter.

Etape initiale = Etape 0

a) On associe à chaque individu un poids. On considère généralement que tous les individus au départ, ont une même poids (=1).
b) Au départ, chaque individu forme une classe, donc la partition initiale correspond au nombre d'individus que l'on examine.
c) Puisque chaque classe initiale a un seul individu, le centre de gravité de chacune des classes initiales se confond l'unique point qui les compose et le poids de chaque classe k est noté $m_k = 1$
d) On calcule alors des distances entre les individus 2 à 2, ce qui revient à calculer les distances entre les centres de gravité des clases initiales.
e) On forme alors une nouvelle partition en réunissant dans une classe, les deux premiers individus qui sont les plus proches, au sens de la plus petite distance euclidienne (critère de Ward), ce qui signifie de la plus faible perte d'inertie (information).

D'autant plus les individus sont proches l'un à l'autre, d'autant plus les individus se ressemblent et donc en les réunissant dans une même classe, on perd d'autant moins d'information.

Le critère d'agrégation retenu est donc le critère de WARD = Minimisation de la perte d'inertie.

Si nous avons n individus initialement, donc une partition initiale en n classes, à la fin de l'étape 0, nous avons désormais n-1 classes. La nouvelle classe comprenant les deux individus les plus proches a donc un poids $m_k = 2$ alors que tous les autres classes ont toujours un poids $m_k = 1$.

1ère Etape

a) puisque la partition obtenue a permis de définir de nouvelles classes en nombre réduit, les centres de gravités se sont modifiés et il faut alors calculer ces nouveaux centres.
b) Puis on calcule les distances entre chacun de n-1 nouveaux centres de gravité, c'est-à-dire entre chacune des n-1 classes de la partition.
c) On forme une nouvelle partition en réunissant dans une même classe, les individus / classes qui sont les plus proches, ceux qui font perdre le moins d'inertie.
d) Selon ce critère de minimisation de la perte d'inertie, nous obtenons alors n-2 classes.

$2^{ème}$ Etape

a) on procède exactement de la même façon qu'à l'étape 1.
b) On obtient alors une nouvelle partition avec n-3 classes.
c) Si le nombre de classes obtenues est supérieur à 1, on continue, sinon la procédure est terminée.

On peut alors, représenter cette agrégation des individus par un arbre de classification hiérarchique, sur lequel apparaissent les nœuds successifs de réunion des classes, la valeur de chaque nœud correspondant à la perte d'inertie provoquée par la réunification des classes.

<u>Présentation de l'algorithme à l'aide d'un exemple simple</u>

Nous considérons 5 individus caractérisés par une variable qui prend les valeurs suivantes :

Individus	A	B	C	D	E
Variable	5	9	8	3	1

Etape 0

Tableau des distances initiales d^2 (i,j)= $(xi-xj)^2$

	A	B	C	D	E
A	0	16	9	4	16
B		0	1	36	64
C			0	25	49
D				0	4
E					0

$d^2(A,B) = (5-9)^2=16$

Calcul de la perte d'inertie par la réunion de 2 classes k et k'

$$P = \frac{m_k . m_{k'}}{m_k + m_{k'}} d^2(k,k')$$

Formule 1

Chaque classe a un poids = 1 (m_k). Lorsque l'on réunit 2 points en un seul, la perte d'information correspond donc à la distance entre les deux points, pondérés par le poids relatif de ces 2 points.

Tableau des pertes d'inertie possible par réunion de classes

	A	B	C	D	E
A	0	8	4.5	2	8
B		0	**0.5**	18	32
C			0	12.5	24.5
D				0	2
E					0

Résultat :
La plus faible d'inertie est de 0,5 par la réunion de B et C.
1er Nœud est entre B et C avec une perte de 0,5.

Nouvelle partition :

	A	D	E	F=B+C
Poids $=m_k$	1	1	1	2

Etape 1

Nous devons maintenant calculer les nouvelles distances prenant en compte la réunification de B et C. La distance entre A et D, A et E ainsi que D et E ne change pas.

Tableau des distances Etape 1

	A	D	E	F
A	0	4	16	12.25
D		0	4	30.25
E			0	56.25
F				0

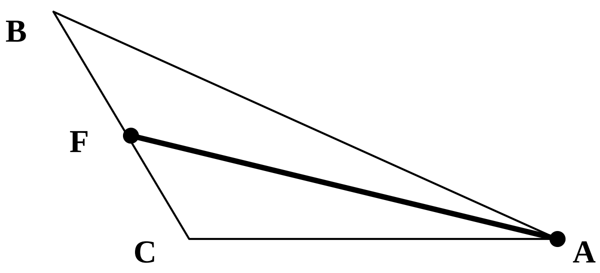
Il a fallu calculer les nouvelles distances selon le critère d'agrégation de Ward ; pour les réunions de A à F, D à F et E à F.

Le **calcul de la distance** de A à F (= B+C) se fait ainsi :

$$d^2(A,F) = \frac{1}{m_B + m_C}\left[m_B d^2(B,A) + m_C d^2(C,A) - \frac{m_B.m_C}{m_B + m_C} d^2(B,C) \right]$$ **Formule 2**

De la même façon, on calcule d2 (D,F) et d2 (E,F)

Quel est le sens de cette formule ?



La distance entre la classe A et la nouvelle classe F qui est la réunification de B et C n'est autre que la distance de A et B plus la distance de A à C moins la distance de
B à C, ces distances étant pondérées par le poids relatif de chaque classe.

Tableau des pertes d'inertie en Etape 1

	A	D	E	F
A	0	2	8	8.17
D		0	2	20.17
E			0	37.5
F				0

Pour cela, on utilise la première formule, en tenant compte des poids de chaque classe. En fait, les classes A, D et E ont 1 individu donc leur poids =1 tandis que la classe F a deux individus et son poids = 2.

Perte d'inertie due à la réunion de A et F (B+C) $= \frac{1\times 2}{1+2}\times 12{,}25 = \frac{2}{3}\times 12{,}25 = 8{,}17$

La perte minimale est égale à 2 soit par réunion de A et D
soit par réunion de D et E

Choisissons la réunion de D et E. Nous avons une nouvelle classe G composée de 2 individus.

Le 2$^{\text{ème}}$ Nœud est entre D et E avec une perte de 2.

Nouvelle partition

	A	F	G=D+F
Poids = m $_k$	1	2	2

Etape 2

Nous devons à nouveau calculer les nouvelles distances prenant en compte la réunification de D et F

Tableau des distances Etape 2

	A	F	G=D+E
A	0	12,25	9
F		0	42,25
G			0

La distance entre A et F ne change pas puisqu'il s'agit toujours de 2 individus uniques. Il faut calculer les nouvelles distances selon le critère d'agrégation de Ward pour les réunions de A à G et F à G.

Calcul de la distance de A à G (=D+E)

$$d^2(A,G) = \frac{1}{m_D + m_E}\left[m_D d^2(D,A) + m_E d^2(E,A) - \frac{m_D . m_E}{m_D + m_E} d^2(D,E)\right]$$ **Formule 2**

De la même façon, on calcule d² (D,G)

Tableau des partes d'inertie en Etape 2

	A	F	G=D+E
A	0	8,1	6
F		0	42,25
G			0

Pour cela, on a utilisé la première formule, en tenant compte des poids de chaque classe. En fait, la classe A contient 1 individu donc son poids =1 tandis que les classes F et G ont deux individus et leur poids = 2.

La perte minimale est égale à 6. Le 3ème nœud est entre A et G avec une perte de 6

Nouvelle partition :

	F	H= A + G
Poids = m_k	2	3

Etape 3

Il ne reste que deux classes dans la partition précédente, que signifie que cette étape est la dernière. Il suffit de calculer la perte d'inertie qui subviendra lorsque les deux dernières classes F et H vont s'agréger.

La nouvelle est distance est donnée par la formule 2 soit :

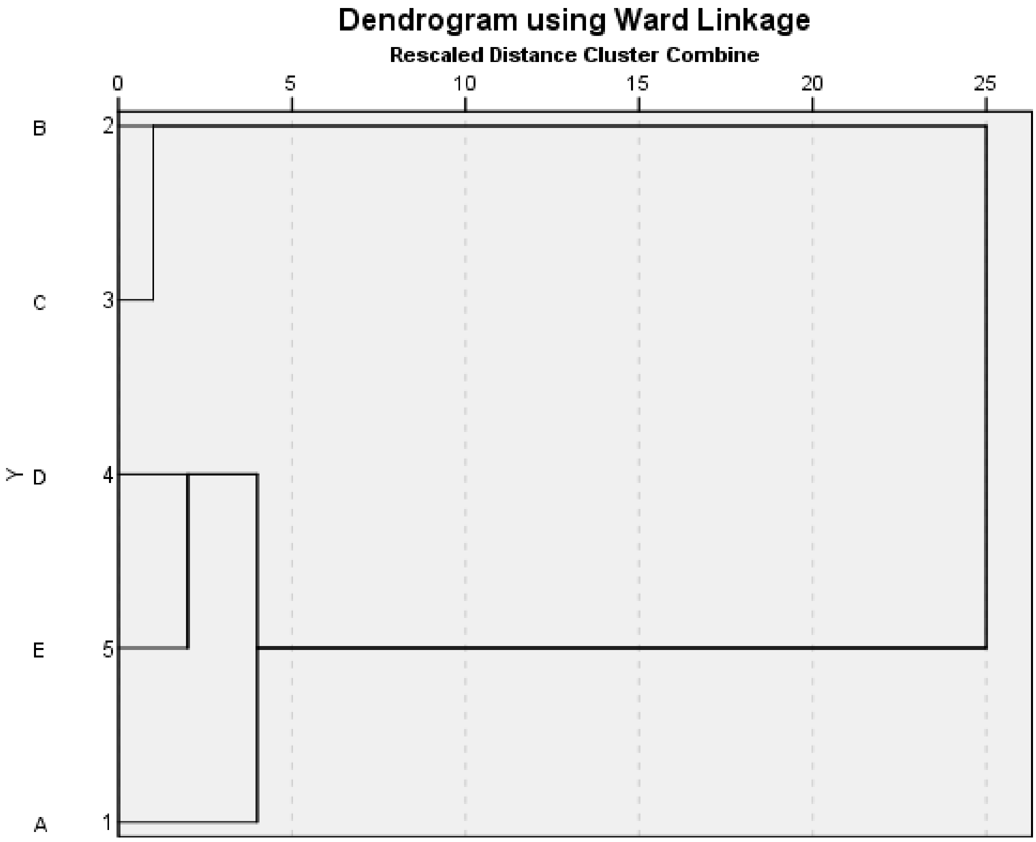
$$d^2(F,H) = \frac{1}{m_A + m_G}\left[m_A d^2(A,F) + m_G d^2(G,F) - \frac{m_A.m_G}{m_A + m_G} d^2(A,G)\right]$$

$D^2(F, H) = 30.25$

Le 4ème Nœud réunissant F et H engendre une perte d'inertie égale à :

$$P = \frac{m_F.m_H}{m_F + m_H} d^2(F,H) = \frac{2,3}{2+3} * 30,25 = 36,3$$

		Critère d'agrégation = niveau du noeud	Taux	Taux cumulés
Nœud 4	A,B,C,D,E	36,3	81,0%	81,0%
Nœud 3	A,D,E	6	13,4%	94,4%
Nœud 2	D,E	2	4,5%	98,9%
Nœud 1	D,C	0,5	1,1%	100%
		44,8	100%	


Dendrogram using Ward Linkage
Rescaled Distance Cluster Combine
0
5
10
15
20
25
B 2
C 3
Y D 4
E 5
A 1

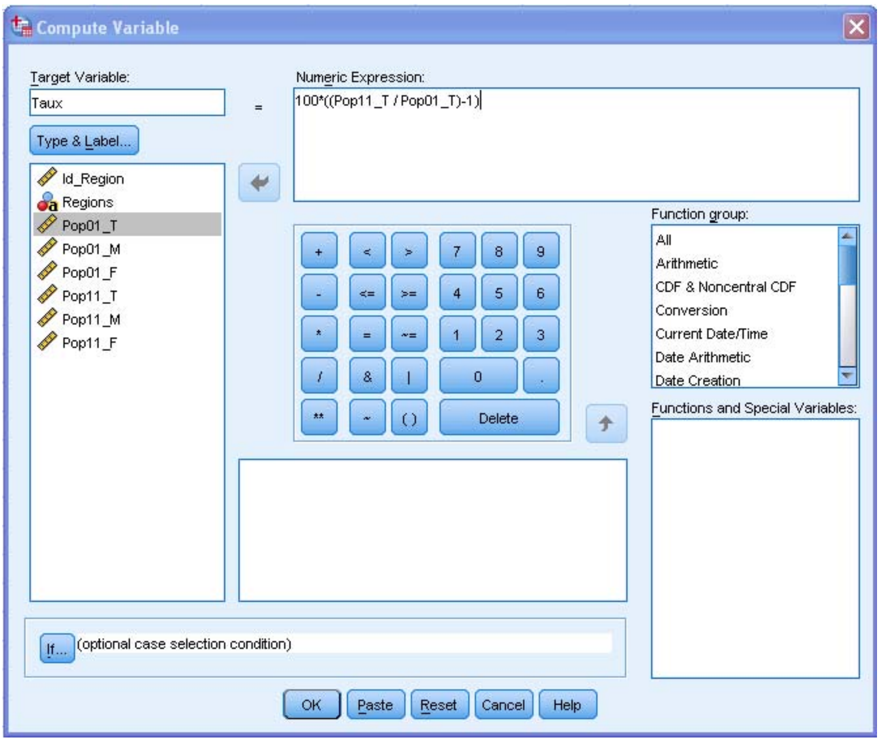
Exemple de déroulement d'une Classification Hiérarchique à l'aide du logiciel SPSS

Les données portent sur les **13 régions de Grèce** et plus précisément sur leur population totale (ainsi que population Hommes et Femmes) lors des deux derniers recensements 2001 et 2011. L'objectif de l'analyse est de produire **<u>une partition des 13 régions en fonction du taux de variation intercensitaire de leur population</u>**.

Id_Region	Regions	Pop01_T	Pop01_M	Pop01_F	Pop11_T	Pop11_M	Pop11_F
111	Macedoine de l'Est et Thrace	607162	303974	303188	606170	299100	307070
112	Macedoine Centrale	1874597	920997	953600	1874590	910270	964320
121	Macedoinde de l'Ouest	294317	148389	145928	282120	141260	140860
122	Epire	336392	166878	169514	336650	165890	170760
231	Thessalie	740115	366585	373530	730730	361900	368830
232	Grece Centrale	558144	284669	273475	546870	278160	268710
241	Iles Ioniennes	209608	104219	105389	206470	102020	104450
242	Grece de l''ouest	721541	365223	356318	680190	341400	338790
243	Peloponnese	597622	303851	293771	581980	294910	287070
351	Attique	3894573	1885889	2008684	3812330	1842680	1969650
461	Nord de l'Egee	205235	106001	99234	197810	99520	98290
462	Sude de l'Egee	298462	154599	143863	308610	155990	152620
471	Crete	594368	300191	294177	621340	308760	312580

Nous commençons donc par créer la variable **Taux** = taux de variation de la population entre 2001 et 2011, à l'aide de la commande : **Transform – Compute variable**. On utilisera l'une ou l'autre formule ci-dessous.

$$Taux = 100 * \left(\frac{Pop11\_T - Pop01\_T}{Pop01\_T} \right) = 100 * \left(\left(\frac{Pop11\_T}{Pop01\_T} \right) - 1 \right)$$



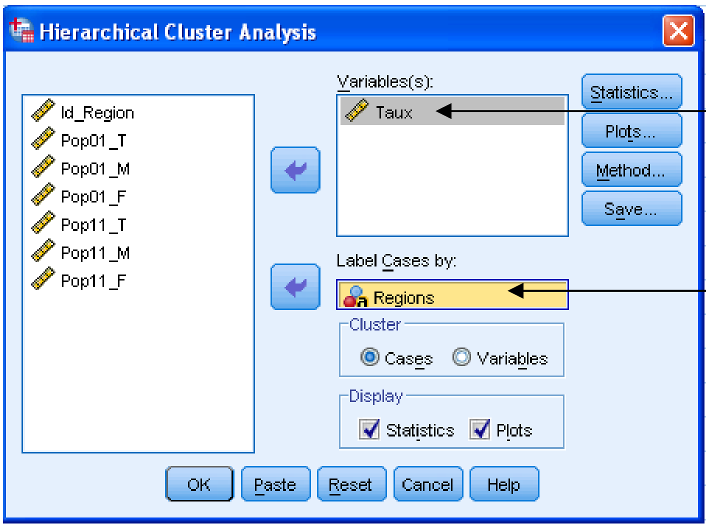
La nouvelle variable **Taux** est ainsi créée et apparait à la dernière colonne du fichier de données.

Id_Region	Regions	Taux
111	Macedoine de l'Est et Thrace	-,16
112	Macedoine Centrale	,00
121	Macedoinde de l'Ouest	-4,14
122	Epire	,08
231	Thessalie	-1,27
232	Grece Centrale	-2,02
241	Iles Ioniennes	-1,50
242	Grece de l' Ouest	-5,73
243	Peloponnese	-2,62
351	Attique	-2,11
461	Nord de l'Egee	-3,62
462	Sud de l'Egee	3,40
471	Crete	4,54

L'observation des valeurs prises par la variable Taux met en évidence - selon les régions considérées - des évolutions relativement disparates entre 2001 et 2011.
La plupart des régions ont du mal à maintenir leur population, certaines mêmes subissent une perte relativement importantes, seules deux régions présentent un taux relativement positif (> 3,4%)

Nous allons réaliser une simple classification hiérarchique afin de mettre en évidence les partitions successives (regroupements successifs des 13 régions) et proposer une classification de ces régions.

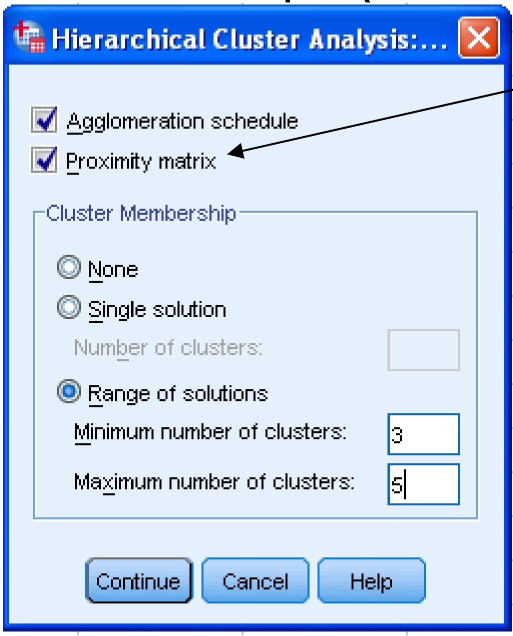
Commande utilisée : Analyze – Classify – Hierarchical Cluster



Choix de la variable à partir de laquelle sera effectuée la classification

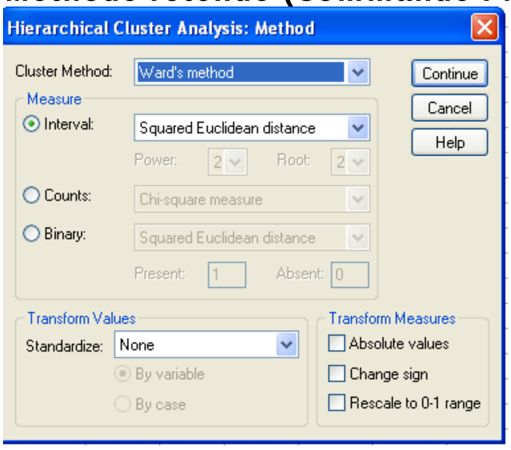
Sélection de la variable d'identification de chaque région de façon à mieux lire les résultats

Résultats Statistiques (Commande : Statistics)



Lorsque le nombre d'individus est limité, il est intéressant de produire la matrice de proximité qui n'est autre que la matrice initiale des distances 2 à 2 entre les individus. Attention, il s'agit d'une matrice carrée donc plus le nombre d'individus est élevé et plus cette matrice sera grande. Vous pouvez néanmoins la copier dans Excel pour faciliter sa lecture.
Vous pouvez demander la production d'un certain nombre de partitions de façon à lire le regroupement des individus par classe, en fonction du nombre de classes que vous aurez choisi. Ici nous avons choisi de produire les résultats de trois partitions, celles correspondant à 3, 4 et 5 classes. Nous aurons ainsi, un tableau qui nous permettra de mieux suivre les regroupements.

Méthode retenue (Commande : Method)

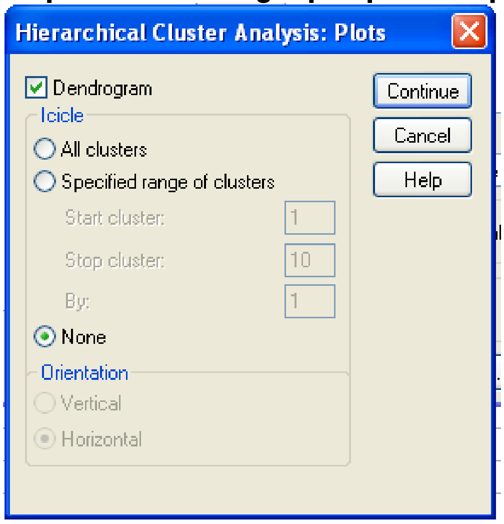


La méthode de **Ward** est l'une des méthodes les plus habituelles et communément utilisées

La mesure choisie pour regrouper les individus doit être la **distance euclidienne.**

Comme nous n'avons qu'une variable, la question de la normalisation ne se pose pas

Représentation graphique des partitions (Commande Plots)



Se limiter à l'arbre des partitions car c'est la représentation la plus explicite

Une fois toutes les sélections effectuées, les résultats apparaissent sur une nouvelle feuille de travail (feuille output). Ces résultats nous donnent :

(a) *un résumé de la méthode appliquée* : nombre d'individus retenus, nombre d'individus pour lesquels nous détenons des données valides, critère de regroupement.

Case Processing Summary[a]

Cases					
Valid		Missing		Total	
N	Percent	N	Percent	N	Percent
13	100,0	0	,0	13	100,0

a. Ward Linkage

(b) la *matrice initiale de proximité* qui n'est lisible que si le nombre d'individus est limité

Proximity Matrix

Case	Squared Euclidean Distance												
	1:Macedoine de l'Est et Thrace	2:Macedoine Centrale	3:Macedoinde de l'Ouest	4:Epire	5:Thessalie	6:Grece Centrale	7:Iles Ioniennes	8:Grece de l' Ouest	9:Pelopon nese	10:Attique	11:Nord de l'Egee	12:Sud de l'Egee	13:Crete
1:Macedoine de l'Est et Thrace	,000	,027	15,847	,058	1,220	3,447	1,779	30,998	6,022	3,796	11,933	12,698	22,102
2:Macedoine Centrale	,027	,000	17,171	,006	1,607	4,079	2,240	32,839	6,849	4,458	13,086	11,563	20,596
3:Macedoinde de l'Ouest	15,847	17,171	,000	17,816	8,272	4,512	7,007	2,518	2,331	4,131	,277	56,916	75,379
4:Epire	,058	,006	17,816	,000	1,808	4,396	2,477	33,729	7,258	4,789	13,649	11,045	19,903
5:Thessalie	1,220	1,607	8,272	1,808	,000	,565	,052	19,917	1,821	,712	5,521	21,792	33,709
6:Grece Centrale	3,447	4,079	4,512	4,396	,565	,000	,273	13,772	,357	,008	2,553	29,376	43,005
7:Iles Ioniennes	1,779	2,240	7,007	2,477	,052	,273	,000	17,925	1,255	,378	4,497	23,982	36,421
8:Grece de l' Ouest	30,998	32,839	2,518	33,729	19,917	13,772	17,925	,000	9,694	13,099	4,465	83,376	105,449
9:Peloponnese	6,022	6,849	2,331	7,258	1,821	,357	1,255	9,694	,000	,256	1,001	36,210	51,198
10:Attique	3,796	4,458	4,131	4,789	,712	,008	,378	13,099	,256	,000	2,268	30,380	44,218
11:Nord de l'Egee	11,933	13,086	,277	13,649	5,521	2,553	4,497	4,465	1,001	2,268	,000	49,251	66,516
12:Sud de l'Egee	12,698	11,563	56,916	11,045	21,792	29,376	23,982	83,376	36,210	30,380	49,251	,000	1,295
13:Crete	22,102	20,596	75,379	19,903	33,709	43,005	36,421	105,449	51,198	44,218	66,516	1,295	,000

Nous observons une très forte proximité entre quelques régions, tandis que la Grèce de l'Ouest, la Crète et le Sud de l'Egée présentent en général de fortes distances avec les autres régions. Par contre, la Crète et le Sud de l'Egée ont une relativement faible distance entre elles.

(c) *l'historique de la classification*, c'est-à-dire la présentation des étapes de regroupement. Puisque nous avons 13 individus, il y a donc 12 étapes successives jusqu'à ce que le dernier individu (le plus éloigné de tous) ait été regroupé. Si nous avons N individus, nous avons donc n-1 nœuds successifs.
Le tableau qui fait suite, reprend en détail la formation des 12 nœuds.

Agglomeration Schedule

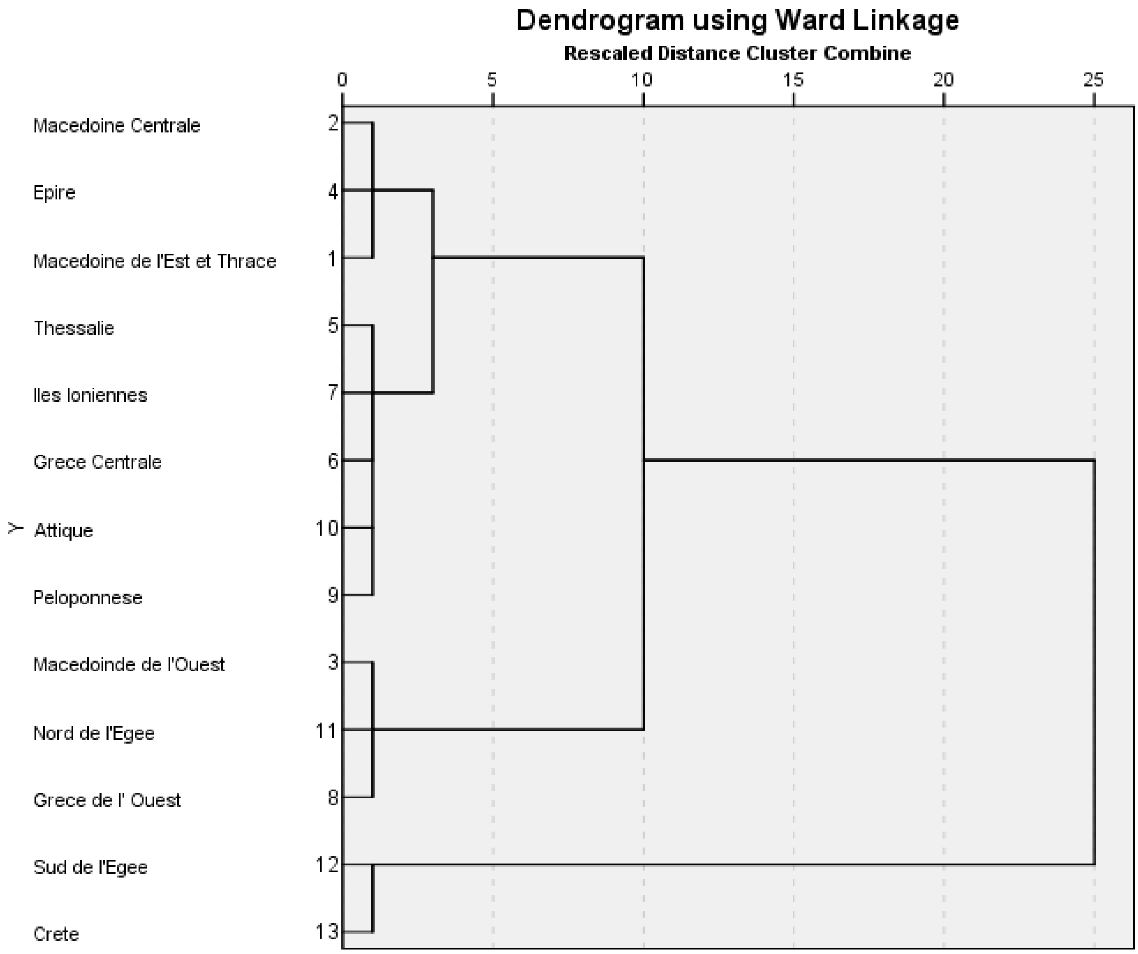
Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	2	4	,003	0	0	4
2	6	10	,007	0	0	6
3	5	7	,033	0	0	8
4	1	2	,060	0	1	10
5	3	11	,199	0	0	9
6	6	9	,402	2	0	8
7	12	13	1,049	0	0	12
8	5	6	1,951	3	6	10
9	3	8	4,233	5	0	11
10	1	5	10,816	4	8	11
11	1	3	34,540	10	9	12
12	1	12	96,862	11	7	0

1er nœud : Les 2 régions les plus proches sont les régions 2 et 4 (Macédoine Centrale et Epire).
Taux pour Région 2 = 0,00 et Taux pour Région 4 = 0,08. Leur distance euclidienne est donc : $(0{,}00-0{,}08)^2 = 0{,}0064$. En formant ce premier nœud, la perte d'information est alors de = 0,0032 *(voir Formule 1).* Il faut attendre le 4ème nœud (colonne next stage) pour que ces deux régions soient regroupées avec une autre région.
2ème nœud : Régions 6 (Grèce Centrale) et 10 (Attique) dont la distance est de 0,008, entrainant une perte totale d'information de 0,007 (0,0032 au 1er nœud + 0,004 au 2eme). Il faut attendre le 6ème nœud pour que ces 2 régions soient regroupées à une autre (colonne next stage). Au 6eme nœud apparait en effet l'information selon laquelle, la région 6 fut rattachée à la région 10, lors du 2eme nœud.
3ème nœud : Régions 5 et 7 (Thessalie et Iles Ioniennes) dont la distance est de : $(-1{,}27-(-1{,}50))^2 = 0{,}053$. La perte d'information à ce 3ème nœud, toujours selon la même formule, est de 0,026, soit donc une perte totale de 0,033.
4ème nœud : il réunit la région 1 à la région 2, plus exactement la région 1 aux régions 2 et 4 qui ont formé le 1er nœud. Il faudra attendre le 10eme nœud pour que d'autres régions soient rattachées aux trois régions 1, 2 et 4.

Au total, après la réunion de toutes les régions, la perte cumulée d'information est de 96,862. A partir du tableau qui suit, on observe que la perte d'information totale reste très limitée jusqu'au 9eme nœud. Elle est de moins de 5%. Par contre au 10eme nœud, nous avons effectivement un saut non négligeable puisque l'on passe à 11%. Quant au 11eme nœud, il correspond a une perte totale de près de 36%, ce qui est un niveau non acceptable. On admet en général qu'une bonne partition ne doit pas entrainer une perte supérieure à 25%.

La lecture du dendrogramme nous permet de lire plus aisément l'historique des partitions. Certes, le nombre limité de régions facilite largement cette lecture, ce qui ne sera plus le cas lorsque le nombre d'individus statistiques ser relativement grand.

Nœuds	Perte cumulée	%	
1	0,003	0,0	
2	0,007	0,0	
3	0,033	0,0	
4	0,06	0,1	
5	0,199	0,2	
6	0,402	0,4	
7	1,049	1,1	
8	1,951	2,0	
9	4,233	4,4	Quatre partitions avec très faible perte d' information
10	10,816	11,2	Trois partitions acceptables
11	34,54	35,7	Deux partitions non acceptables
12	96,862	100,0	Une seule partition comprenant les 13 régions



A la lecture du tableau et du dendrogramme, il semble que la partition en 4 groupes soit la meilleure. Le tableau qui suit, nous permet de lire la formation des groupes selon que l'on en retient 5, 4 ou 3.

Cluster Membership

Case	5 Clusters	4 Clusters	3 Clusters
1:Macedoine de l'Est et Thrace	1	1	1
2:Macedoine Centrale	1	1	1
3:Macedoinde de l'Ouest	2	2	2
4:Epire	1	1	1
5:Thessalie	3	3	1
6:Grece Centrale	3	3	1
7:Iles Ioniennes	3	3	1
8:Grece de l' Ouest	4	2	2
9:Peloponnese	3	3	1
10:Attique	3	3	1
11:Nord de l'Egee	2	2	2
12:Sud de l'Egee	5	4	3
13:Crete	5	4	3

La partition en 4 groupes que nous privilégierons ici, revient à considérer :

a) Les régions bénéficiant d'une croissance démographique supérieure à 3,4% (Sud de l'Égée et Crète)
b) Les régions qui parviennent tout juste à maintenir leur population (croissance nulle) : Macédoine de l'Est et Thrace, Macédoine Centrale et Épire
c) Les régions marquées par une légère baisse de population. Elles sont au nombre de cinq, la perte variant entre -1,27% (Thessalie) et -2,62% (Péloponnèse)
d) Les régions subissant la plus forte perte de population (supérieure a 3,6%). Il s'agit du Nord de l'Égée, de la Macédoine de l'Ouest et enfin de la Grèce de l'Ouest.

Cette méthode peut être utilisée en considérant non pas une mais plusieurs variables. Dans ce dernier cas, il faudra prendre soin <u>de standardiser les variables</u> afin d'éviter les problèmes que peuvent engendrer les différences d'échelle de mesure entre les diverses variables.