
SECTION: Psychologie - Master 1ère année

**ANALYSES MULTIDIMENSIONNELLES ET APPLICATIONS
INFORMATIQUES - PREMIERE SESSION
Enoncé et Indications de réponses**

N.B. Calculatrices et résumés de cours autorisés

Source : Philippe TISON. La dépression du sujet âgé. Hypothèse de schémas cognitifs spécifiques. Thèse de doctorat de psychologie soutenue le 23 juin 2003. Accessible en ligne à l'adresse : http://documents.univ-lille3.fr/files/pub/www/recherche/theses/tison-philippe/html/these_body.html

Dans le cadre de l'étude citée ci-dessus, l'auteur considère trois instruments d'évaluation utilisés en ce qui concerne les pathologies dépressives : GDS, réellement spécifique aux sujets âgés, HDRS et MADRS qui sont d'une spécificité acceptable.

- La GDS (Geriatric Depression Scale) de Yesavage et Brink (1983) est un questionnaire d'évaluation en 30 items, à réponse oui-non avec possibilité d'auto et d'hétéro-évaluation. On la retrouve parfois traduite sous l'expression d'échelle de dépression gériatrique. Elle permet le dépistage de la dépression chez la personne âgée.
- L'HDRS (Hamilton Depression Rating Scale) de Hamilton (1967), est une des échelles les plus utilisées dans le monde pour l'hétéro-évaluation de la dépression. Cette échelle est un bon indicateur de l'intensité globale du syndrome dépressif, suffisamment sensible pour apprécier le changement sous traitement antidépresseur. Ce n'est pas un instrument de diagnostic.
- La MADRS (Montgomery and Asberg Depression Rating Scale) de Montgomery et Asberg (1979), est également un instrument souvent utilisé. Cette échelle fournit un bon indice de gravité globale de la dépression du fait de sa grande sensibilité de changements aux traitements. Cette échelle semble plus sensible et plus fidèle que l'HDRS.

Par ailleurs, l'auteur considère la triade cognitive de Beck. Pour Beck et coll. (1979), la dépression s'explique par la présence et l'activité constante de schémas cognitifs négatifs, qui orientent le traitement des informations externes dans un sens toujours négatif. En résultent donc des pensées négatives sur soi, sur le monde et sur le futur (triade cognitive de Beck), et par conséquent des émotions de tristesse. Selon l'auteur, de manière empirique, nous retrouvons en effet assez facilement la triade cognitive de Beck dans le discours des sujets âgés déprimés, mais nous trouvons également 2 cognitions qui n'entrent pas clairement dans cette classification : il s'agit de l'idée d'un changement impossible avec l'âge (CIA) « je suis trop vieux pour changer », et l'idée d'une banalisation des symptômes dépressifs avec l'âge (BSD) « c'est normal d'être déprimé à mon âge ». Ces 2 cognitions semblent fréquentes et paraissent dépasser un cadre anecdotique. Ces productions mentales étant bien issues de raisonnements, d'interprétations établies à partir de faits de l'environnement, donc correspondant bien à des pensées négatives de patients déprimés, il semblait alors logique qu'elles puissent être correctement réparties dans le système de la

triade cognitive de Beck (TCB). Or, CIA et BSD sont deux cognitions qui n'entrent pas clairement dans la TCB, et qui semblent se différencier légèrement des pensées typiques et habituelles de la TCB.

L'auteur a mené une étude sur un échantillon de 100 sujets âgés déprimés. Ces sujets sont évalués d'une part sur les 3 échelles GDS, HDRS, MADRS et d'autre part sur des échelles évaluant la TCB, le CIA et la BSD, à l'aide d'un questionnaire de cognitions spécifiques. Pour chacune de ces échelles, un score est d'autant plus élevé qu'il correspond à une dépression plus profonde ou à des cognitions plus affirmées.

L'auteur analyse ensuite les liens entre les 7 variables observées sur les sujets (les 6 échelles précédentes et la variable Age) à l'aide d'une analyse en composantes principales normée.

Les principaux résultats sont les suivants :

- Moyennes et écarts types pour les 7 variables :

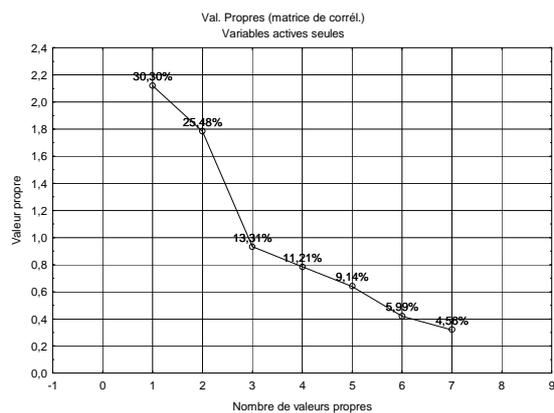
	N Actifs	Moyenne	Ecart-type
Age	100	79,46	8,25
GDS	100	18,56	4,19
HDRS	100	16,91	5,12
MADRS	100	22,58	5,92
CIA	100	4,41	1,54
BSD	100	3,36	1,63
TCB	100	3,59	1,57

- Corrélations entre les variables :

	Age	GDS	HDRS	MADRS	CIA	BSD	TCB
Age	1,00	-0,03	-0,13	-0,16	0,20	0,15	0,10
GDS	-0,03	1,00	0,14	0,40	0,32	0,44	0,39
HDRS	-0,13	0,14	1,00	0,60	-0,11	-0,18	-0,01
MADRS	-0,16	0,40	0,60	1,00	-0,02	-0,01	0,20
CIA	0,20	0,32	-0,11	-0,02	1,00	0,21	0,33
BSD	0,15	0,44	-0,18	-0,01	0,21	1,00	0,29
TCB	0,10	0,39	-0,01	0,20	0,33	0,29	1,00

- Valeurs propres & stat. associées

	Val Propre	% Total variance	Cumul Val Propre	Cumul %
1	2,12	30,30	2,12	30,30
2	1,78	25,48	3,90	55,78
3	0,93	13,31	4,84	69,09
4	0,78	11,21	5,62	80,30
5	0,64	9,14	6,26	89,44
6	0,42	5,99	6,68	95,44
7	0,32	4,56	7,00	100,00



- Saturations des variables sur les trois premiers axes

Coord. factorielles des var., basées sur les corrélations			
	Fact. 1	Fact. 2	Fact. 3
Age	-0,12	-0,48	0,76
GDS	-0,83	0,09	-0,22
HDRS	-0,20	0,81	0,32
MADRS	-0,50	0,74	0,12
CIA	-0,55	-0,39	0,26
BSD	-0,59	-0,40	-0,35
TCB	-0,70	-0,15	0,04

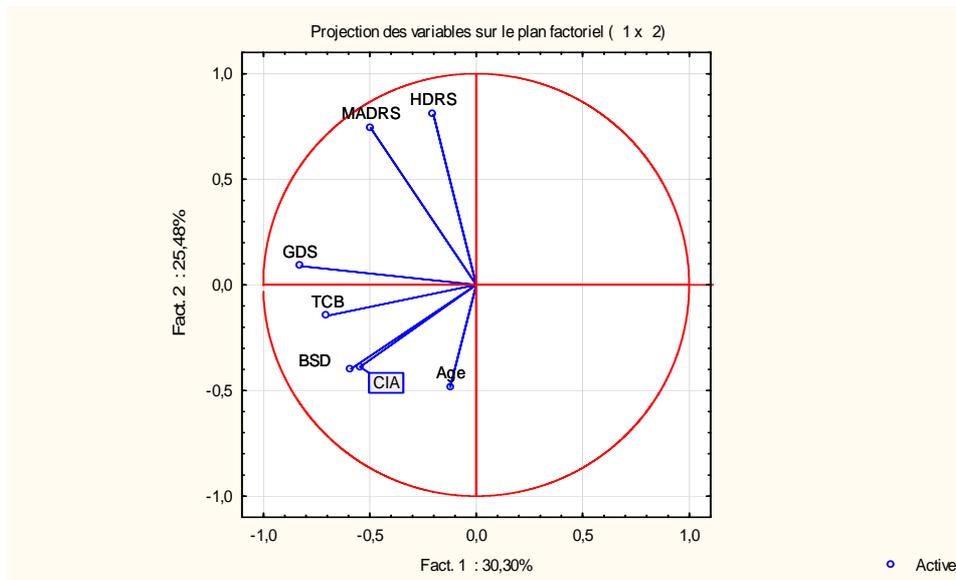
- Contributions des variables

Contributions des var., basées sur les corrélations			
	Fact. 1	Fact. 2	Fact. 3
Age	0,01	0,13	0,61
GDS	0,32	0,00	0,05
HDRS	0,02	0,37	0,11
MADRS	0,12	0,31	0,02
CIA	0,14	0,08	0,07
BSD	0,16	0,09	0,13
TCB	0,23	0,01	0,00

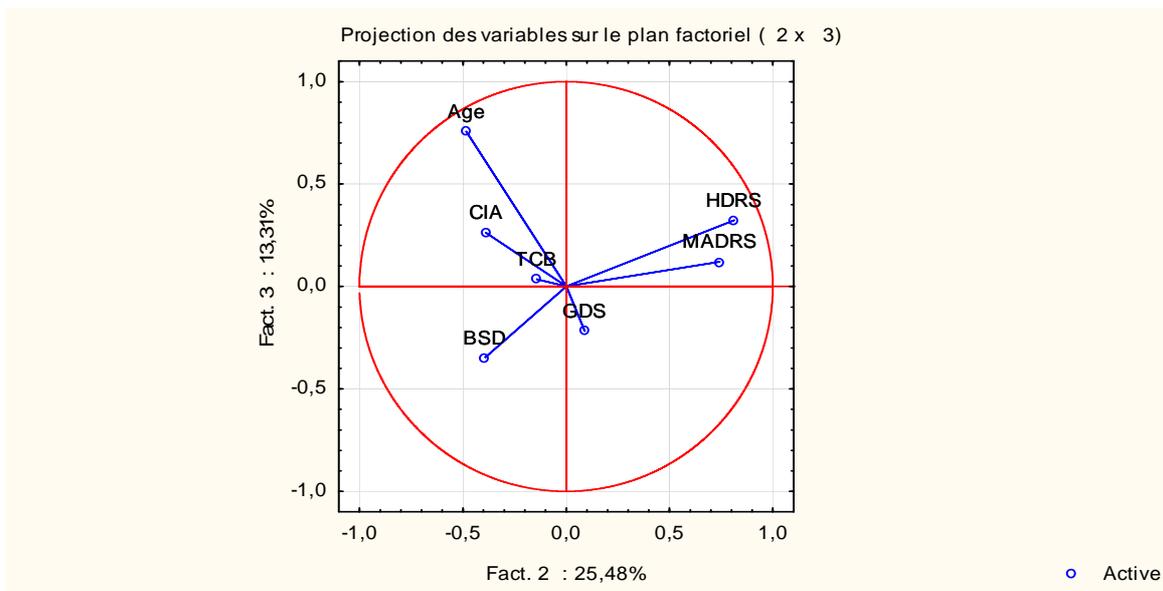
- Communautés

Communautés, basées sur les corrélations			
	Avec 1 facteur	Avec 2 facteurs	Avec 3 facteurs
Age	0,01	0,25	0,82
GDS	0,69	0,69	0,74
HDRS	0,04	0,70	0,81
MADRS	0,25	0,80	0,81
CIA	0,30	0,45	0,52
BSD	0,35	0,50	0,63
TCB	0,49	0,51	0,51

- Représentation graphique des variables selon les axes 1 et 2



- Représentation graphique des variables selon les axes 2 et 3



1) a) Examiner et commenter le tableau des corrélations entre les variables. Quelles sont les corrélations qui sont significatives, sachant que, pour un échantillon de taille 100, la valeur critique du coefficient de corrélation au seuil de 5% bilatéral est $r_{crit} = 0,1946$?

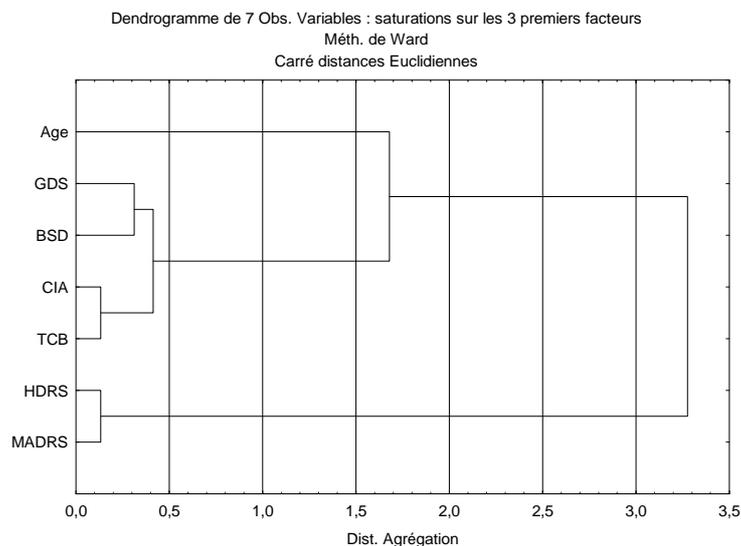
b) Préalablement à son étude, l'auteur a fait les hypothèses de recherche suivantes :

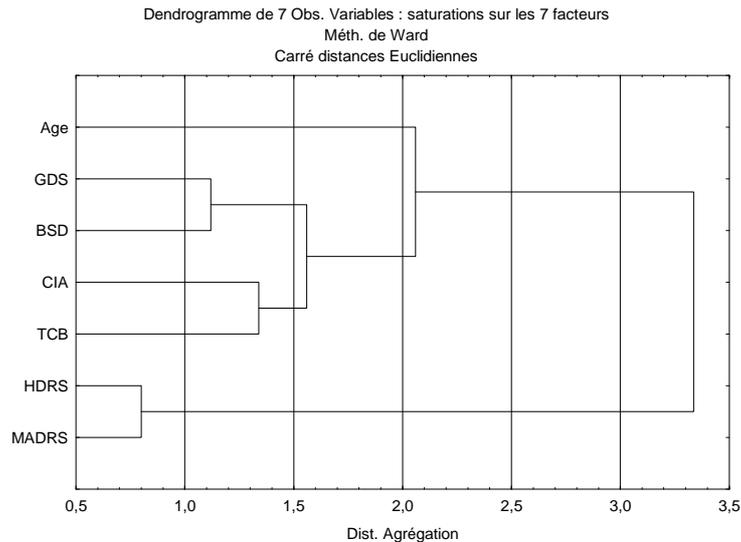
- L'adhésion aux cognitions de la Triade Cognitive de Beck (TCB), du « changement impossible avec l'âge » (CIA), et de la « banalisation des symptômes dépressifs avec l'âge » (BSD) est différente selon l'état dépressif des sujets âgés.
- Le niveau d'adhésion aux cognitions est fonction de la sévérité de la dépression estimée à l'HDRS.

Dans quelle mesure l'étude des corrélations confirme-t-elle ces hypothèses de recherche ?

2) Etude du tableau des valeurs propres

- a) Calculer la somme des valeurs propres. A quoi correspond cette valeur ?
 b) On choisit d'étudier les trois premières composantes principales. Discuter ce choix en analysant le tableau et le graphique des valeurs propres.
- 3) Etude des qualités de représentation des variables
 a) Quelles sont les variables le mieux représentées dans l'espace défini par les trois premiers facteurs. Quelles sont celles qui sont le plus mal représentées ?
 b) L'une des variables est quasiment indépendante du premier facteur. Laquelle ? Comment peut-on qualifier sa représentation dans le premier plan factoriel ?
- 4) Etude du premier axe principal, du point de vue des variables.
 a) Quelles sont les variables qui ont le plus contribué à la formation du premier axe principal ?
 b) Quelle remarque peut-on faire concernant le signe des saturations des variables sur le premier axe ? Comment appelle-t-on l'effet ainsi mis en évidence ? Quelle interprétation peut-on faire de ce premier axe ?
- 5) Etude du deuxième axe principal, du point de vue des variables.
 Quelles sont les variables dont la contribution à la formation de la deuxième composante principale est supérieure à la moyenne ? Pour chacune d'elles, préciser le signe de la saturation correspondante. Caractériser cet axe en référence aux variables étudiées.
- 6) Etude du troisième axe principal, du point de vue des variables.
 Quelle est la variable dont la contribution à la formation de la troisième composante principale est supérieure à la moyenne ? Un tel axe est souvent qualifié d'*unipolaire*. Pourquoi ? Que représente-t-il ?
- 7) A partir des résultats de l'ACP, on a réalisé deux classifications ascendantes hiérarchiques. Pour la première, on a utilisé les saturations des variables sur les trois premiers axes ; pour la seconde, on a pris en compte les saturations des variables sur les 7 axes factoriels. Les dendrogrammes obtenus sont les suivants :





- a) Pour la première CAH (celle utilisant les saturations sur les 3 premiers facteurs), calculer la "distance" entre les objets MADRS et HDRS. Comment peut-on retrouver ce résultat sur le graphique ?
- b) Pour chacun des deux dendrogrammes, indiquer la composition des classes dans le cas d'une partition en deux classes, en trois classes, puis en quatre classes.
- c) Quelles sont les deux premières classes formées par l'algorithme de classification pour la première CAH (celle utilisant les saturations sur les 3 premiers facteurs) ?
Indiquer de même quelles sont les deux premières classes formées par l'algorithme de classification pour la deuxième CAH.
- d) Si on réalisait une CAH sur les variables observées centrées et réduites, le résultat serait-il différent de ceux qui sont obtenus ici ? Lequel des deux dendrogrammes retrouverait-on ainsi ?

Indications de réponses

- 1) a) Le tableau des corrélations ne comporte pas de valeurs extrêmes (aucune corrélation au-dessus de 0,60 ou en dessous de -0,18). La majorité des corrélations sont positives ; on ne trouve que 8 corrélations négatives mais assez proches de 0. Les corrélations qui sont significatives sont toutes positives : Age n'est significativement corrélé qu'avec CIA, GDS est assez fortement corrélé avec MADRS, CIA, BSD et TCB, mais n'est pas corrélé avec HDRS. HDRS n'est corrélé significativement qu'avec MADRS. Outre les corrélations précédentes, MADRS est corrélé avec TCB. Quant aux trois cognitions (CIA, BSD, TCB) elles sont assez faiblement, mais significativement, corrélées entre elles et avec GDS ; TCB est également corrélée à MADRS.
- b) On a noté des corrélations significatives entre GDS et les variables mesurant les trois cognitions. La première hypothèse de recherche est ainsi confirmée, si on admet que l'état dépressif est correctement mesuré par GDS. En revanche, les corrélations entre HDRS et les variables mesurant les cognitions ne sont pas significatives, ce qui infirme la deuxième hypothèse. On pourrait également noter que ces corrélations sont négatives, mais cette remarque est de peu d'intérêt puisque ces corrélations ne sont pas significativement différentes de 0.
- 2) a) La somme des valeurs propres est égale à 7 (6,99 si on utilise le tableau fourni). Cette valeur est égale au nombre de variables, et également à l'inertie totale du nuage de points.

b) Une première règle consiste à n'étudier que les axes dont l'inertie est supérieure à la moyenne ($100\%/7=14,28\%$), ce qui revient à n'étudier que les axes correspondant à des valeurs propres supérieures à 1. Cependant, la troisième valeur propre est égale à 0,93. Elle est donc assez proche de 1. Par ailleurs, comme nous le verrons à la question 4, le premier axe représente un effet de taille et son interprétation est assez banale. Si on neutralise cet axe et qu'on ne considère que les axes 2 à 7, on a une inertie de $7-2,12=4,88$ et donc une moyenne de $4,88/6=0,81$ par axe. On peut donc choisir d'étudier les axes correspondant aux valeurs propres supérieures à 0,81, c'est-à-dire les 3 premiers axes. On notera malgré tout l'absence de rupture nette entre la 3e et la 4e valeurs propres.

3) a) On peut se référer à la colonne "avec 3 facteurs" du tableau des communautés. On voit que ces communautés se situent entre 0,51 et 0,82, et sont donc assez équilibrées. Aucune variable n'est véritablement ignorée par l'étude sur 3 axes. Les mieux représentées sont Age (0,82), HDRS (0,81) et MADRS (0,81). Les moins bien représentées sont TCB (0,51), CIA (0,52) et BSD (0,63).

b) Les saturations des variables sont aussi les coefficients de corrélation de ces variables avec les facteurs. En examinant la colonne "Fact. 1" du tableau des saturations, on constate que la corrélation la plus proche de 0 est celle liant le facteur 1 et Age. Cette valeur (-0,12) montre que ces deux variables sont indépendantes (la valeur critique donnée dans la question 1 peut encore être utilisée). La qualité de représentation de l'âge dans le premier plan est 0,25 (cf. tableau des communautés, colonne "Avec 2 facteurs"). Cette variable est donc plutôt mal représentée dans le premier plan (tout en étant bien représentée dans l'espace de dimension 3, comme nous l'avons déjà noté).

4) a) On peut se limiter aux variables dont la contribution à l'inertie du premier axe est supérieure à 14%. On obtient ainsi :

-	+
GDS (32%)	
TCB (23%)	
BSD (16%)	
CIA (14%)	

b) On constate que les saturations sont toutes de même signe (ici, le signe négatif, mais le sens de l'axe provient d'un choix arbitraire fait par le logiciel). Ce résultat met en évidence un effet de taille : le premier facteur représente une mesure globale, ou synthétique, de la dépression des sujets. Autrement dit, ce facteur rassemble toutes les échelles, dans ce qu'elles ont en commun. Un tel effet est classique lorsque les variables de départ sont toutes corrélées positivement entre elles. Ici, il n'était pas évident a priori qu'un tel effet serait mis en évidence, puisqu'il existe des corrélations négatives entre certaines variables. Mais nous avons déjà noté que les corrélations positives sont plus nombreuses et plus fortes (ce sont les seules à être significatives).

5) Si on se limite aux variables dont la contribution à l'inertie du premier axe est supérieure à 14%, on obtient :

-	+
	HDRS (37%)
	MADRS (31%)

Cependant, si on considère l'ensemble des variables, on voit que cet axe oppose les échelles GDS, MADRS, HDRS aux trois cognitions TCB, BSD, CIA et à l'âge. Alors que le premier facteur représente plutôt ce que ces six mesures ont en commun, le deuxième facteur représente plutôt ce en quoi elles diffèrent. Cet axe est plutôt bipolaire, opposant mesures de la dépression et cognitions.

6) La variable Age représente 61% de l'inertie de cet axe. Il s'agit pour l'essentiel d'un axe unipolaire classant les sujets selon leur âge.

7) a) La "distance" utilisée est ici le carré de la distance euclidienne, mesurée dans l'espace de dimension 3. En utilisant le tableau des saturations sur les 3 axes, on obtient :

$$d = (-0,20 + 0,50)^2 + (0,81 - 0,74)^2 + (0,32 - 0,12)^2 = 0,1349$$

Cette valeur correspond à celle qui est lue sur le graphique (entre 0,1 et 0,15).

b) Pour des partitions en 2, 3 ou 4 classes, les deux dendrogrammes fournissent les mêmes résultats, à savoir :

- en deux classes (on coupe les deux dendrogrammes à la distance d'agrégation 2,5) :

Classe 1 = {Age, GDS, BSD, CIA, TCB}

Classe 2 = {HDRS, MADRS}

- en trois classes (on coupe le premier dendrogramme à la distance d'agrégation 1,5 et le second à la distance d'agrégation 2) :

Classe 1 = {Age}

Classe 2 = {GDS, BSD, CIA, TCB}

Classe 3 = {HDRS, MADRS}

- en quatre classes (on coupe le premier dendrogramme à la distance d'agrégation 0,4 et le second à la distance d'agrégation 1,5) :

Classe 1 = {Age}

Classe 2 = {GDS, BSD}

Classe 3 = {CIA, TCB}

Classe 4 = {HDRS, MADRS}

c) La question porte ici sur les deux premières classes formées lors de l'exécution de l'algorithme de la CAH, donc sur les classes correspondant aux plus petites distances d'agrégation, dans chacune des deux classifications.

Pour la première classification (distances calculées à partir de 3 facteurs), les deux premières classes sont {HDRS, MADRS} ($d=0,1349$) et {CIA, TCB} ($d=0,1285$).

Pour la seconde classification (distances calculées à partir des coordonnées sur les 7 facteurs), les deux premières classes sont : {HDRS, MADRS} ($d=0,6$ environ) et {GDS, BSD} ($d=1,2$ environ).

La comparaison des deux dendrogrammes montre que l'organisation générale des variables est la même dans les deux cas (classifications en 2, 3 ou 4 classes). En revanche, en ne considérant que 3 dimensions au lieu de 7, on élimine une partie de l'écart entre les variables : les distances d'agrégation observées lorsqu'on réunit deux variables sont moins élevées dans le premier dendrogramme que dans le second, et cette "déformation" est variable selon le couple de variables considéré. CIA et TCB apparaissent ainsi plus proches l'une de l'autre que ne le sont GDS et BSD lorsqu'on mène le calcul sur trois facteurs, alors qu'on obtient le résultat inverse en calculant sur sept facteurs.

d) Lorsque l'ensemble des facteurs est pris en compte, l'ACP revient en fait à une rotation des axes de l'espace multidimensionnel. Les distances entre les objets (ici les variables) sont représentées sans déformation, et donc, les distances entre variables, calculées à partir des valeurs des variables centrées réduites ou à partir des saturations sur les 7 axes sont identiques. On obtiendrait donc un dendrogramme identique à celui obtenu à partir des 7 facteurs.