

SECTION: Psychologie - Master 2ème année

N.B. Calculatrice et documents autorisés

ANALYSE MULTIDIMENSIONNELLE DES DONNEES
Sujet donné en 2009-2010 et indications de correction

Ref. Costalat-Founeau A.-M. et al., Représentation du corps et de l'alimentation chez une population de femmes de plus de 75 ans, in Papers on Social Representations, Textes sur les représentations sociales Volume 11, pages 4.1 - 4.20 (2002).

Dans une étude publiée en 2002, les auteurs s'intéressent aux représentations sociales du corps et de l'alimentation chez les femmes de plus de 75 ans. L'étude porte sur un échantillon de 52 sujets. Les participantes sont d'abord été interviewées dans le cadre d'un entretien semi-directif. A partir des thèmes recueillis au cours des entretiens, cinq propositions relatives à la représentation du corps ont été élaborées :

Mécanique: Le corps est avant tout une machine vivante qui obéit à une mécanique naturelle très précise. Chaque organe joue un rôle particulier et le moindre petit problème peut déboucher sur une maladie qui affectera tout le corps.

Séduction: Le corps est la première chose que l'on offre au regard des autres. C'est pour cela qu'il est important de l'entretenir, soit par des exercices, soit par des soins cosmétiques afin de conserver une apparence physique agréable.

Communication: Le corps est sans doute le premier moyen de communication de l'homme. Il nous permet d'abord d'établir des contacts physiques avec des personnes ou des animaux; nous les voyons, les touchons, les sentons. En fait, c'est grâce à nos cinq sens que nous percevons notre univers dans sa globalité.

Emotion: Le corps est avant tout le siège des émotions ; il nous permet de ressentir le plaisir ou bien la douleur, mais aussi d'extérioriser nos sentiments. Il est aussi le porte-parole de l'amour et de l'amitié ainsi que celui de la haine.

Maternité: La fonction première du corps de la femme est sans doute celle de pouvoir concevoir, porter et donner le jour à un enfant. Le corps de la femme est avant toute autre chose, le corps nourricier de la mère.

Dans une seconde phase, chacun des sujets devait classer les cinq propositions en fonction de son degré d'accord. De façon à faciliter l'interprétation dans les traitements qui suivent, le score 5 a été attribué à la proposition qui lui semble la plus juste et le score 1 à celle qui semble la moins satisfaisante.

Les données recueillies forment donc un tableau de 52 lignes (52 sujets) sur 5 colonnes (les scores recueillis par les 5 propositions). Les variables correspondant à ces 5 colonnes sont nommées CMECA, CSEDUC, CCOM, CEMOT et CMATER dans la suite du texte. Ces données sont traitées par une analyse en composantes principales normée, dans laquelle on place en individus supplémentaires les scores moyens observés en regroupant les sujets de l'échantillon selon diverses caractéristiques socio-économiques :

- le revenu, réparti en 4 classes (moins de 3000 F, de 3000 F à 6000 F, de 6 000F à 8 500 F, plus de 8 500 F),
- le niveau d'études, en deux classes (Certificat d'études primaires, Etudes Supérieures)
- le mode de vie (en couple, en institution spécialisée, seule, avec ses enfants),
- l'indice de masse corporelle, avec 3 catégories (maigre, normale, obèse),
- l'âge, avec deux classes (moins de 80 ans, plus de 80 ans).

Des traitements analogues menés sur les résultats obtenus lors d'une réplique de cette enquête, conduisent aux résultats suivants.

Extrait des données

Sujet	CMECA	CSEDUC	CCOM	CEMOT	CMATER
1	5	2	1	3	4
2	3	4	5	2	1
3	1	3	5	4	2
4	1	4	3	5	2
5	1	4	5	3	2
6	4	1	3	2	5

Individus supplémentaires

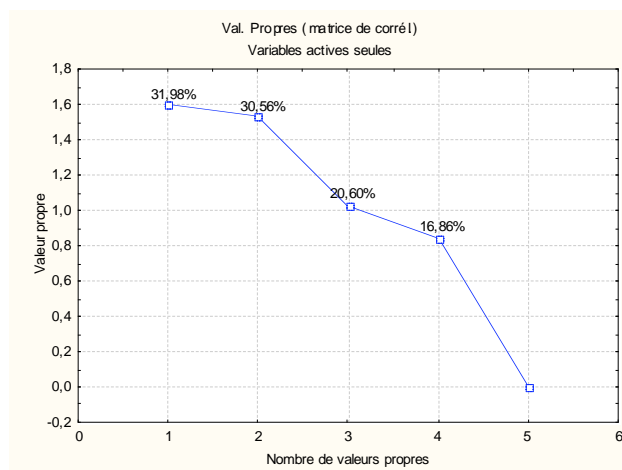
	CMECA	CSEDUC	CCOM	CMOT	CMATER	Effectif
R:0-3	3,65	3,30	2,40	2,86	2,89	10
R:3-6	3,07	2,44	2,94	3,32	3,20	16
R:6-8,5	2,38	3,67	2,94	2,77	3,41	13
R:8,5+	2,77	3,17	2,67	3,42	3,20	13
CEP	3,36	2,81	2,76	3,05	3,30	15
Etudes Sup	2,48	3,67	2,76	3,23	3,20	35
Couple	2,77	3,05	2,67	2,77	3,67	12
Institution	3,56	3,42	2,49	2,95	2,68	10
Seule	2,68	3,30	3,03	3,23	3,05	19
Avec enfants	3,16	2,81	2,76	3,14	3,37	11
Maigre	3,16	2,56	2,94	3,51	2,79	15
Normale	2,58	3,67	2,67	3,05	3,51	19
Obèse	3,26	3,17	2,76	2,95	3,20	18
Age:- 80 ans	3,16	2,93	2,49	3,14	3,51	23
Age:+80 ans	2,77	3,42	2,94	3,23	2,89	29

Corrélations entre les variables

	CMECA	CSEDUC	CCOM	CEMOT	CMATER
CMECA	1,00	-0,45	-0,36	-0,07	-0,16
CSEDUC	-0,45	1,00	-0,08	-0,28	-0,09
CCOM	-0,36	-0,08	1,00	-0,16	-0,40
CEMOT	-0,07	-0,28	-0,16	1,00	-0,43
CMATER	-0,16	-0,09	-0,40	-0,43	1,00

Val. Propres (matrice de corrél.) & stat. associées
Variables actives seules

	Val Propre	% Total variance	Cumul Val Propre	Cumul %
1	1,60	31,98	1,60	31,98
2	1,53	30,56	3,13	62,54
3	1,03	20,60	4,16	83,14
4	0,84	16,86	5,00	100,00



Coord. factorielles des variables, basées sur les corrélations

	Fact. 1	Fact. 2
CMECA	0,77	-0,28
CSEDUC	-0,74	0,01
CCOM	-0,35	0,68
CEMOT	0,52	0,51
CMATER	-0,24	-0,85

Contributions des variables, basées sur les corrélations

	Fact. 1	Fact. 2
CMECA	0,3755	0,0525
CSEDUC	0,3446	0,0001
CCOM	0,0760	0,3030
CEMOT	0,1687	0,1708
CMATER	0,0353	0,4735

Communautés, basées sur les corrélations

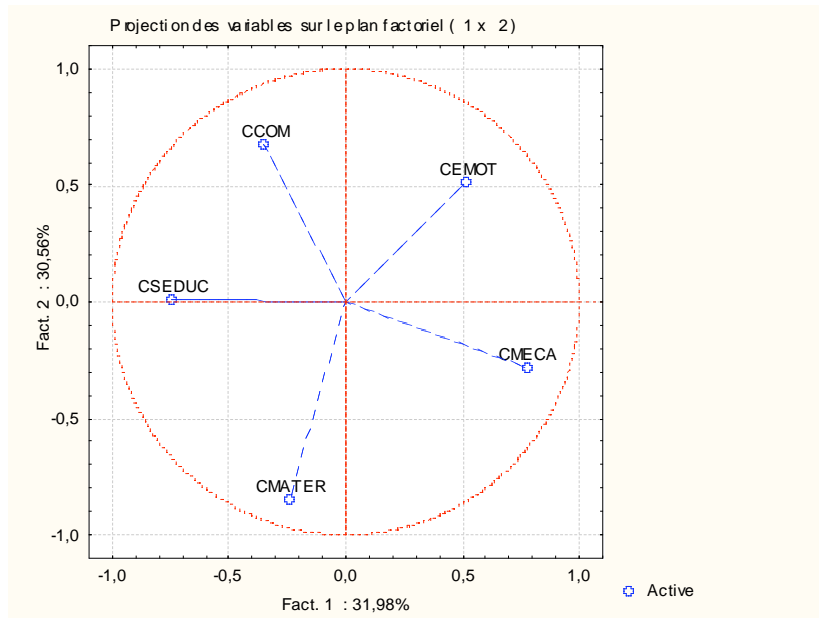
	Avec 1 facteur	Avec 2 facteurs
CMECA	0,60	0,68
CSEDUC	0,55	0,55
CCOM	0,12	0,58
CEMOT	0,27	0,53
CMATER	0,06	0,78

Coordonnées factorielles et cosinus carrés des individus supplémentaires

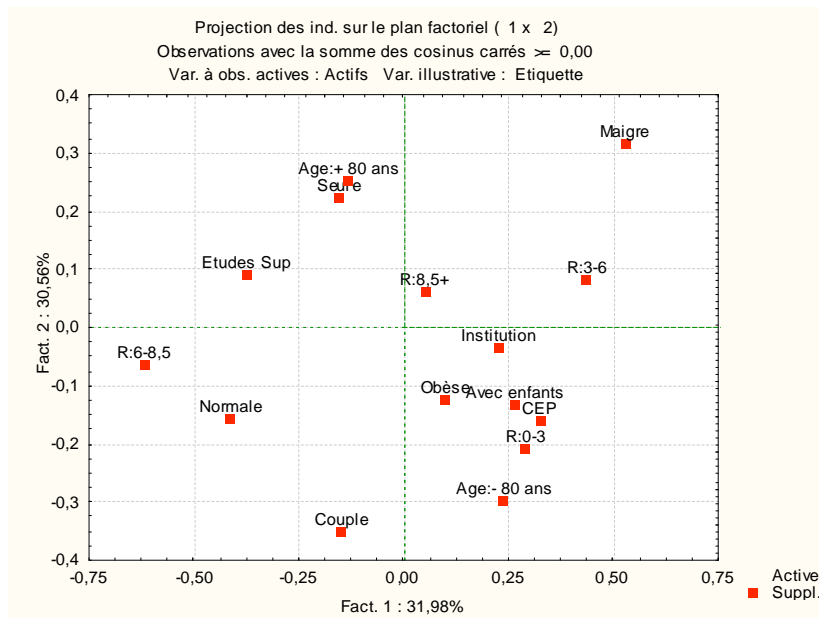
Etiquette	Fact. 1	Fact. 2	Cosinus carrés Fact. 1	Cosinus carrés Fact. 2
R:0-3	0,28984	-0,20946	0,17	0,09
R:3-6	0,43585	0,08165	0,57	0,02
R:6-8,5	-0,61743	-0,06367	0,96	0,01
R:8,5+	0,05093	0,06108	0,04	0,06
CEP	0,32757	-0,16291	0,54	0,13
Etudes Sup	-0,37417	0,08951	0,53	0,03
Couple	-0,14874	-0,35251	0,13	0,72
Institution	0,22860	-0,03539	0,11	0,00
Seule	-0,15430	0,22381	0,28	0,59
Avec enfants	0,26474	-0,13167	0,58	0,14

Maigre	0,52925	0,31356	0,69	0,24
Normale	-0,41074	-0,15928	0,60	0,09
Obèse	0,10076	-0,12511	0,09	0,14
Age:- 80 ans	0,24140	-0,29937	0,38	0,58
Age:+ 80 ans	-0,13140	0,25093	0,14	0,49

Projection des variables sur le plan factoriel 1 x 2



Projection des individus supplémentaires sur le plan factoriel 1 x 2



1) Commenter le tableau des corrélations entre les variables.

2) Etude des valeurs propres.

a) Le tableau numérique ne mentionne que 4 valeurs propres, alors que le graphique semble en indiquer une cinquième.

Quel est en général le nombre de valeurs propres produites par la méthode ?
Comment peut-on expliquer le phénomène constaté ici ?

b) Comment peut-on qualifier la décroissance des valeurs propres ?

c) Selon les critères généralement utilisés, combien de valeurs propres semblerait-il pertinent de retenir ?

d) Les auteurs indiquent :

" Nous arrêterons l'interprétation des axes quand la somme des variances expliquées par chacun des axes sera supérieure à 50%."

Est-ce la règle généralement appliquée pour déterminer le nombre d'axes à étudier ?

3) *Etude des deux premières composantes, du point de vue des variables*

a) Pour le premier axe factoriel, quelles sont les variables dont la contribution est supérieure à la moyenne ? Pour chacune d'elles, préciser le signe de la coordonnée correspondante. Caractériser cet axe en termes d'opposition entre variables.

b) Même question pour le deuxième axe.

c) Comment peut-on qualifier les qualités de représentation des variables dans le premier plan factoriel ? Quelle est la variable la mieux représentée, la plus mal représentée ?

4) *Etude des individus supplémentaires*

a) Pourquoi a-t-on préféré s'intéresser aux individus supplémentaires définis à partir des variables socio-économiques plutôt qu'aux sujets observés ?

b) Dans quelle mesure les résultats fournis ici permettent-ils de retrouver les résultats décrits dans l'article, résultats obtenus à partir des données originales :

"Si on approfondit davantage les résultats à travers la projection des variables, on constate que ce sont les faibles revenus qui sont associés à la représentation du "corps mécanique". Par opposition, la représentation du "corps séduction" est associée aux sujets au poids normal, de revenus élevés."

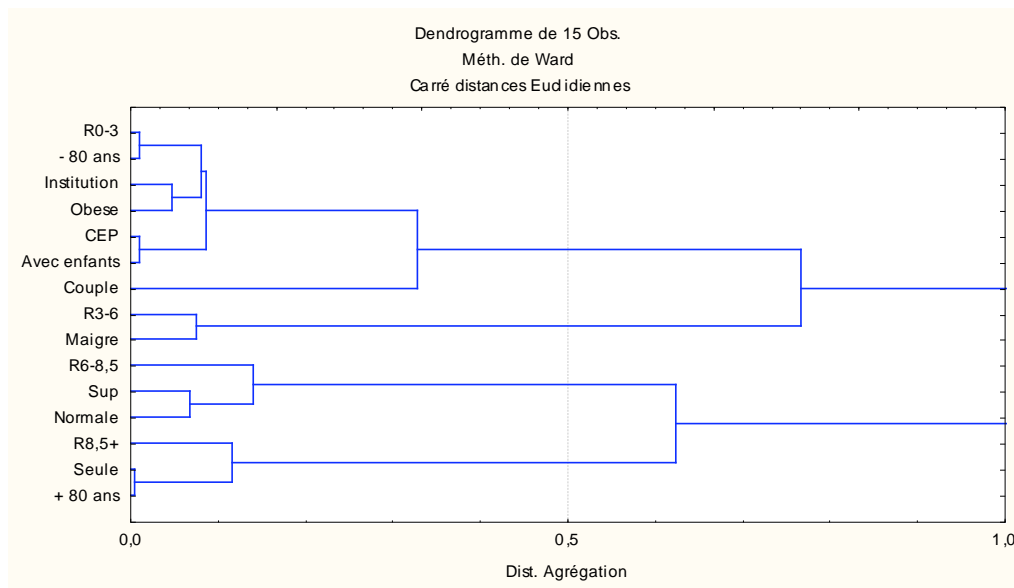
"Sur le deuxième axe, la représentation "corps-maternité" est associée aux variables couple avec enfants, obésité, études supérieures, et s'oppose au "corps-communication"."

5) On a réalisé ici une ACP normée.

Était-il strictement nécessaire de procéder au centrage et à la réduction des variables avant de réaliser l'ACP ?

Peut-on penser que les résultats obtenus à partir d'une ACP non normée auraient été équivalents ?

6) A partir des coordonnées des individus supplémentaires sur les 3 premiers axes factoriels, on a effectué une classification, qui a produit le résultat graphique suivant :



a) Quelle méthode de classification a-t-on utilisé ici ?

b) On choisit de conserver une partition en 4 classes.

Donner la composition de chacune d'elles.

Comparer les résultats obtenus par cette classification à ceux obtenus par l'ACP.

Indications de réponses

1) L'examen du tableau des corrélations montre que celles-ci sont, dans l'ensemble, peu intenses : de -0,07 (CMECA et CEMOT) à -0,45 (CMECA et CSEDUC). En revanche, il faut remarquer que ces corrélations sont toutes négatives : chacune des variables s'oppose, d'une certaine manière, à toutes les autres. Parmi les corrélations les plus intenses, on note :

- CMECA et CSEDUC (-0,45)
- CMATER et CEMOT (-0,43)
- CMATER et CCOM (-0,40)
- CMECA et CCOM (-0,36).

2) a) Le nombre de valeurs propres d'une ACP est en général le minimum du nombre de lignes et du nombre de colonnes du tableau des données (individus x variables). On a donc ici 5 valeurs propres. Cependant, la cinquième valeur propre est nulle, car les variables sont des rangs calculés sur les lignes, et elles vérifient donc une relation de dépendance ; pour chaque ligne, on a :

$$\text{CMECA} + \text{CSEDUC} + \text{CCOM} + \text{CEMOT} + \text{CMATER} = 15.$$

2) b) La décroissance des valeurs propres est faible, particulièrement entre la première et la deuxième valeurs propres, elle est un peu plus forte entre la deuxième et la troisième.

2) c) En général, on retient pour l'étude les valeurs propres supérieures à 1, ou (ce qui revient au même) celles qui représentent une inertie supérieure à la moyenne. Dans notre exemple, cela conduirait à étudier les trois premiers axes factoriels. Il faut cependant noter que la méthode du coude conduirait plutôt à étudier 2 ou 4 valeurs propres. En effet, les 3ème et la 4ème valeurs propres sont très proches, et il n'y a guère de justification à étudier l'une sans étudier l'autre.

2) d) Cette règle de 50% de variance expliquée n'est pas celle qui est généralement appliquée. Cependant, elle rejoint la méthode du coude, en justifiant de limiter l'étude aux deux premières valeurs propres.

3) a) On se limite aux variables qui ont contribué pour plus de 20% à l'inertie du premier axe. On obtient :

-	+
CSEDUC (34%)	CMECA (38%)

Ainsi le premier axe oppose le "corps séduction" au "corps mécanique" et dans une moindre mesure au "corps émotions".

b) En procédant de même pour le deuxième axe, on obtient :

-	+
CMATER (47%)	CCOM (30%)

Ainsi le premier axe oppose le "corps maternité" au "corps communication" et dans une moindre mesure au "corps émotions".

c) Les qualités de représentation sont lues dans la colonne "avec deux facteurs" du tableau des communautés. On voit que ces qualités de représentation varient de 0,53 pour CEMOT à 0,78 pour CMATER. Elles sont assez homogènes, aucune d'elles n'étant très forte ou très faible.

4) a) Ce ne sont pas les variations individuelles entre les sujets interrogés qui nous intéressent ici, mais bien plutôt les oppositions entre les groupes définis par les différents paramètres socio-économiques. En revanche, il était important de mener l'ACP sur l'ensemble des sujets interrogés afin de prendre en compte la variabilité totale constatée sur l'échantillon observé. Les positions des éléments supplémentaires correspondent aux centres de gravité des points correspondant aux sujets membres de ces groupes.

4) b) Pour la partie positive du premier axe (associée au "corps mécanique"), on trouve : Revenu de 3000 F à 6000 F, niveau d'études : CEP et IMC : maigre. A l'opposé, on trouve sur la partie négative de l'axe (associée au "corps séduction") : Revenu : de 6000 F à 8500 F, niveau d'études : Etudes Supérieures et IMC : normale. On retrouve donc les résultats formulés dans la première affirmation fournie.

Pour la partie positive du deuxième axe (associée au "corps communication"), on trouve : Mode de vie : Seule, Age : plus de 80 ans. A l'opposé, on trouve sur la partie négative de l'axe (associée au "corps maternité") : Mode de vie : Couple et Mode de vie : Avec ses enfants, IMC : normale et IMC : obèse, Age : moins de 80 ans. On retrouve en partie la deuxième affirmation (association "corps maternité" avec couple, avec enfants, obésité). Cependant, on ne retrouve pas le lien entre "corps maternité" et "Etudes supérieures" puisque "Etudes supérieures" a une coordonnée positive sur le deuxième axe.

5) Les cinq variables ont ici le même intervalle de variation : [1, 5]. Une ACP non normée aurait donc vraisemblablement donné des résultats analogues. Cependant, les résultats auraient été en partie différents, car les variables, avant réduction, ont sans doute des variances (dispersions) différentes, ce qui n'est plus le cas après réduction.

6) a) La méthode utilisée ici est une CAH (classification ascendante hiérarchique), pour laquelle la "distance" utilisée est le carré des distances euclidiennes, et la méthode d'agrégation est celle de Ward.

6) b) On coupe le dendrogramme, par exemple pour la distance d'agrégation 0,5. On obtient alors les classes suivantes :

- Classe 1 : { R0-3, - 80 ans, Institution, Obèse, CEP, Avec enfants, Couple }
- Classe 2 : { R3-6, Maigre }
- Classe 3 : { R6-8,5, Etudes Supérieures, Normale }
- Classe 4 : { R8,5+, Seule, + 80 ans }.

Ces quatre classes correspondent à quatre zones bien délimitées du premier plan factoriel, sans pour autant correspondre aux 4 quadrants définis par les deux premiers axes.

