

SECTION: Psychologie - Master 2ème année

N.B. Calculatrice et documents autorisés

ANALYSE MULTIDIMENSIONNELLE DES DONNEES

Sujet donné en 2006-2007 et indications de correction

Dans le cadre d'une recherche nommée "Identité humaine – Identité animale"¹, un groupe de travail en sciences sociales de l'Université de Lausanne a cherché à mieux cerner les caractéristiques perçues généralement comme typiques des êtres humains ou typiques des animaux, ainsi que les caractéristiques perçues comme communes aux deux espèces.

Les chercheurs ont identifié 43 traits susceptibles de caractériser les êtres humains, les animaux ou les deux espèces. La liste de ces traits est la suivante : *affectueux, agressif, altruiste, amical, chaleureux, combatif, communicatif, compétitif, cruel, curieux, égoïste, habile, imaginatif, imprévisible, impulsif, inconstant, indépendant, individualiste, ingénieux, insensé, intelligent, intuitif, inventif, libre, logique, loyal, malin, moral, paresseux, querelleur, raisonnable, raisonneur, rationnel, réfléchi, sauvage, sociable, solidaire, solitaire, spirituel, spontané, têtu, tolérant, violent.*

Un questionnaire a été mis au point et a été distribué durant l'année académique 2000-2001 à des étudiants de première année en architecture à l'Ecole Polytechnique Fédérale de Lausanne (E.P.F.L.). Au total, 50 questionnaires ont été analysés: 22 ont été remplis par des femmes et 26 par des hommes (2 valeurs manquantes). L'âge moyen de la population est de 21 ans (avec un âge minimum de 18 ans et un âge maximum de 36 ans).

Après une première page d'introduction qui présente l'expérience sous le titre «Identité humaine – Identité animale», les sujets sélectionnent, parmi les 43 traits, les 5 traits qui leur semblent être «les plus typiques des êtres humains» et – sur la même page – les 5 traits qui leur semblent être «les plus typiques des animaux».

Chaque sujet, sans exception, a sélectionné 5 traits qu'il jugeait être typiques des êtres humains et 5 traits qu'il jugeait être typiques des animaux. Au total, 250 sélections ont été effectuées pour décrire chacune des deux espèces. Les variables indépendantes considérées dans l'analyse des résultats sont: le sexe des sujets (sex1 : homme, sex2 = femme), le mot stimulus (sti1 : êtres humains, sti2 : animaux) et la catégorie d'âge des sujets (age1 : 18-19 ans, age2 : 20-21 ans, age3 : 22 et +). Seuls les traits qui ont été cités au minimum cinq fois ont été considérés dans l'analyse. De ce fait, l'analyse porte sur un total de 33 traits sur les 43 proposés.

Les résultats observés ont été rassemblés dans le tableau ci-dessous et soumis à une méthode d'analyse des données multivariées.

¹ Ricciardi Joos Paola, DES THEORIES IMPLICITES DE L'HOMME. Comparer l'être humain et l'animal dans des contextes intra- et inter-espèces. Thèse présentée à la Faculté des sciences sociales et politiques de l'Université de Lausanne (Suisse), en cotutelle avec l'Université de Paris X-Nanterre (France), Lausanne, 2004

	sex1	sex2	sti1	sti2	age1	age2	age3	Total
affectueux	3	4	1	6	2	2	3	21
combatif	4	2	1	5	3	2	1	18
communicatif	6	8	9	7	4	6	4	44
compétitif	12	7	11	10	5	7	7	59
cruel	9	9	15	5	3	9	6	56
curieux	11	16	14	14	13	9	5	82
égoïste	5	5	10	1	3	4	3	31
habile	7	4	1	10	2	4	5	33
imaginatif	7	7	14	0	6	5	3	42
imprévisible	8	7	8	8	7	3	5	46
impulsif	9	7	3	15	5	8	3	50
inconstant	3	8	10	1	5	2	4	33
indépendant	3	6	4	5	3	3	3	27
individualiste	2	3	6	0	2	2	1	16
ingénieux	10	2	5	7	3	4	5	36
intelligent	6	10	13	4	5	8	3	49
intuitif	13	13	1	26	9	11	6	79
inventif	10	5	15	1	3	5	7	46
libre	13	6	4	15	6	8	5	57
logique	8	5	7	7	4	5	4	40
loyal	3	6	2	7	1	5	3	27
malin	3	6	0	10	6	2	1	28
moral	10	7	16	1	5	9	3	51
paresseux	3	3	3	3	3	1	2	18
querelleur	2	5	3	4	1	4	2	21
raisonneur	4	5	8	1	4	3	2	27
rationnel	9	7	11	5	2	10	4	48
sauvage	18	11	1	29	11	11	7	88
sociable	6	8	11	4	4	5	5	43
solidaire	6	1	0	7	5	1	1	21
solitaire	3	1	2	3	3	1	0	13
spirituel	11	7	17	1	4	9	5	54
spontané	16	5	0	21	9	4	8	63
Total	243	206	226	243	151	172	126	1367

Résultats de l'analyse de données multivariées :

Val. Propres & stat. associées.

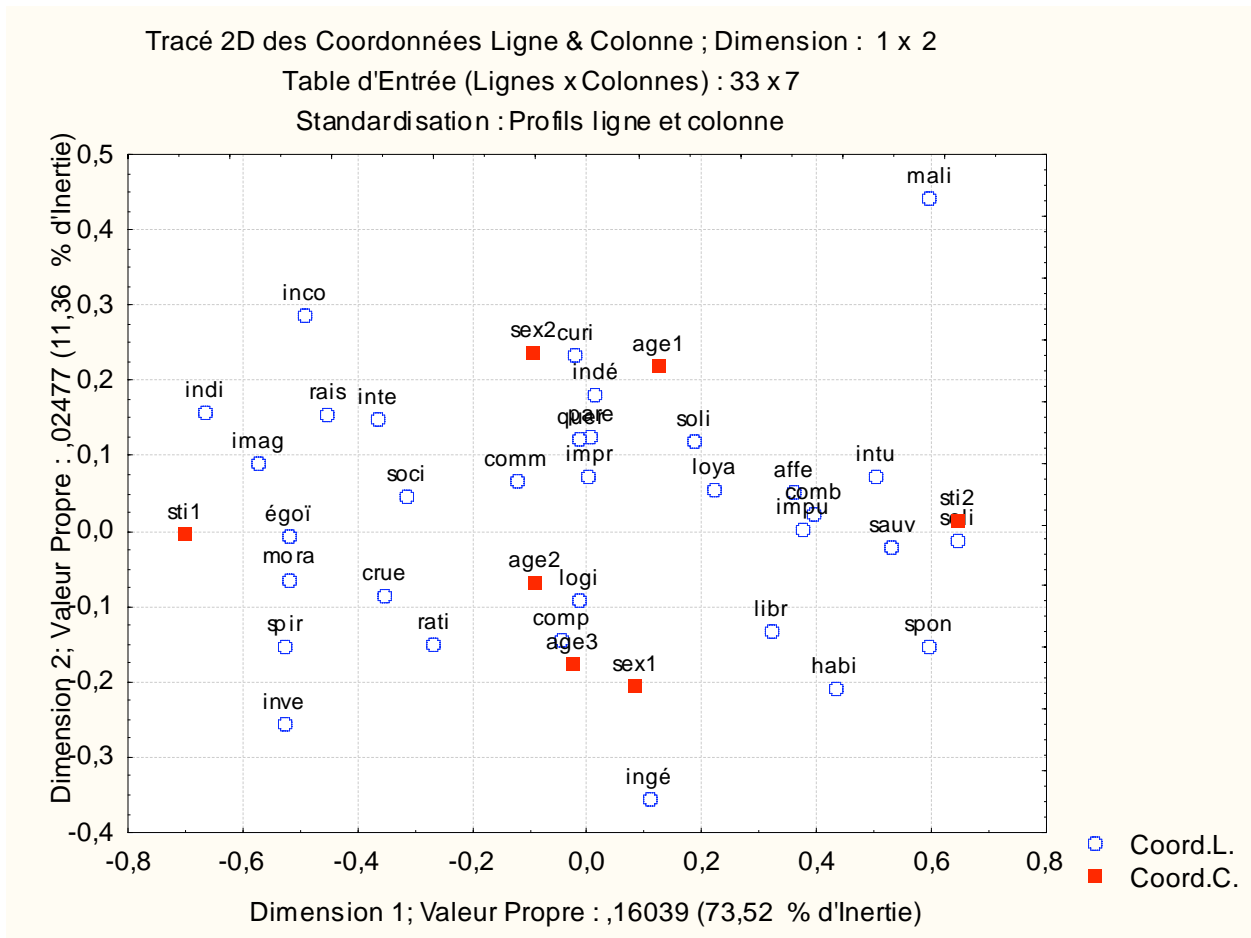
Nombre de Dims.	Valeurs Propres et Inertie de toutes les Dimensions (Enquete-EPFL) Table d'Entrée (Lignes x Colonnes) : 33 x 7 Inertie Totale = ,21815 Chi? = 298,21 dl = 192 p = ,00000				
	ValSing.	ValProp.	%age Inertie	%age Cumulé	Chi?
1	0,400	0,160	73,52	73,52	219,25
2	0,157	0,025	11,36	84,88	33,86
3	0,147	0,022	9,90	94,78	29,53
4	0,104	0,011	4,98	99,76	14,86
5	0,023	0,001	0,24	100,00	0,70
6	0,000	0,000	0,00	100,00	0,00

Scores, contributions et qualités de représentation des individus lignes

NomLigne	Coordonnées Ligne et Contributions à l'Inertie (Enquete-EPFL)											
	Table d'Entrée (Lignes x Colonnes) : 33 x 7 Standardisation : Profils ligne et colonne											
	Coord. Dim.1	Coord. Dim.2	Coord. Dim.3	Masse	Qualité	Inertie Relative	Inertie Dim.1	Cos? Dim.1	Inertie Dim.2	Cos? Dim.2	Inertie Dim.3	Cos? Dim.3
affectueux	0,36	0,05	-0,13	0,015	0,735	0,014	0,013	0,642	0,002	0,014	0,012	0,079
combatif	0,39	0,02	0,15	0,013	0,911	0,012	0,013	0,797	0,000	0,002	0,013	0,112
communicatif	-0,12	0,07	-0,10	0,032	0,928	0,005	0,003	0,478	0,006	0,143	0,014	0,307
compétitif	-0,05	-0,15	0,02	0,043	0,851	0,005	0,001	0,075	0,037	0,767	0,001	0,009
cruel	-0,35	-0,09	-0,13	0,041	0,993	0,028	0,032	0,830	0,012	0,048	0,033	0,115
curieux	-0,02	0,23	0,05	0,060	0,985	0,016	0,000	0,007	0,131	0,933	0,007	0,045
égoïste	-0,52	-0,01	0,04	0,023	0,998	0,028	0,038	0,992	0,000	0,000	0,001	0,005
habile	0,43	-0,21	-0,13	0,024	0,910	0,030	0,028	0,692	0,042	0,160	0,018	0,058
imaginatif	-0,57	0,09	0,16	0,031	0,990	0,051	0,063	0,899	0,010	0,022	0,036	0,069
imprévisible	0,00	0,07	0,17	0,034	0,709	0,007	0,000	0,000	0,007	0,110	0,045	0,599
impulsif	0,37	0,00	-0,12	0,037	0,899	0,029	0,032	0,810	0,000	0,000	0,026	0,089
inconstant	-0,49	0,29	0,10	0,024	0,852	0,044	0,036	0,615	0,080	0,209	0,012	0,028
indépendant	0,01	0,18	-0,11	0,020	0,672	0,006	0,000	0,003	0,026	0,492	0,011	0,177
individualiste	-0,67	0,16	0,08	0,012	0,987	0,026	0,032	0,924	0,012	0,051	0,003	0,012
ingénieux	0,11	-0,35	0,12	0,026	0,983	0,019	0,002	0,076	0,134	0,817	0,017	0,090
intelligent	-0,36	0,15	-0,10	0,036	0,961	0,028	0,029	0,771	0,031	0,127	0,018	0,062
intuitif	0,50	0,07	-0,13	0,058	0,995	0,073	0,091	0,917	0,012	0,019	0,043	0,059
inventif	-0,53	-0,26	0,10	0,034	0,947	0,058	0,059	0,746	0,090	0,176	0,014	0,024
libre	0,32	-0,13	-0,00	0,042	0,954	0,024	0,027	0,815	0,029	0,138	0,000	0,000
logique	-0,01	-0,09	0,03	0,029	0,970	0,001	0,000	0,022	0,010	0,880	0,001	0,069
loyal	0,22	0,05	-0,41	0,020	0,976	0,021	0,006	0,217	0,002	0,013	0,155	0,746
malin	0,60	0,44	0,07	0,020	0,998	0,052	0,045	0,639	0,161	0,349	0,005	0,010
moral	-0,52	-0,06	0,03	0,037	0,887	0,053	0,063	0,872	0,006	0,013	0,001	0,002
paresseux	0,01	0,13	0,19	0,013	0,710	0,005	0,000	0,001	0,008	0,210	0,023	0,499
querelleur	-0,01	0,12	-0,38	0,015	0,991	0,011	0,000	0,001	0,009	0,090	0,104	0,901
raisonneur	-0,45	0,15	0,13	0,020	0,996	0,022	0,025	0,836	0,019	0,096	0,014	0,064
rationnel	-0,27	-0,15	-0,20	0,035	0,884	0,025	0,016	0,463	0,032	0,148	0,068	0,273
sauvage	0,53	-0,02	-0,02	0,064	0,990	0,084	0,113	0,988	0,001	0,001	0,001	0,001
sociable	-0,32	0,05	-0,03	0,031	0,885	0,017	0,020	0,858	0,003	0,018	0,002	0,010
solidaire	0,65	-0,01	0,46	0,015	0,966	0,046	0,040	0,639	0,000	0,000	0,152	0,327
solitaire	0,19	0,12	0,42	0,010	0,729	0,014	0,002	0,114	0,005	0,046	0,078	0,568
spirituel	-0,53	-0,15	0,00	0,040	0,976	0,056	0,069	0,900	0,037	0,076	0,000	0,000
spontané	0,59	-0,15	0,18	0,046	0,974	0,089	0,102	0,838	0,044	0,056	0,071	0,079

Scores, contributions et qualités de représentation des individus colonnes

Nom C	Coordonnées Colonne et Contributions à l'Inertie (Enquete-EPFL)											
	Table d'Entrée (Lignes x Colonnes) : 33 x 7 Standardisation : Profils ligne et colonne											
	Coord. Dim.1	Coord. Dim.2	Coord. Dim.3	Masse	Qualité	Inertie Relative	Inertie Dim.1	Cos? Dim.1	Inertie Dim.2	Cos? Dim.2	Inertie Dim.3	Cos? Dim.3
sex1	0,08	-0,21	0,12	0,178	0,958	0,053	0,008	0,105	0,302	0,641	0,114	0,212
sex2	-0,09	0,24	-0,14	0,151	0,953	0,061	0,008	0,101	0,338	0,632	0,135	0,220
sti1	-0,70	-0,00	0,05	0,165	0,998	0,374	0,505	0,993	0,000	0,000	0,017	0,004
sti2	0,64	0,01	-0,04	0,178	0,998	0,341	0,461	0,993	0,001	0,000	0,015	0,004
age1	0,13	0,22	0,28	0,110	0,986	0,075	0,011	0,110	0,215	0,328	0,413	0,549
age2	-0,09	-0,07	-0,23	0,126	0,725	0,051	0,006	0,091	0,025	0,054	0,301	0,580
age3	-0,02	-0,18	-0,03	0,092	0,317	0,044	0,000	0,005	0,118	0,303	0,004	0,009



1) Quelle méthode l'auteur utilise-t-il ici ? A quel type de tableau cette méthode s'applique-t-elle généralement ? Pour les données étudiées ici, quelle interprétation peut-on donner des sommes des lignes ? des sommes des colonnes ? Que représente le profil-ligne moyen ?

2) Comment peut-on qualifier la décroissance des valeurs propres ? Selon les critères généralement utilisés, combien de valeurs propres semblerait-il pertinent de retenir ?

3) Les colonnes du tableau des résultats se regroupent naturellement selon les trois variables Sexe, Stimulus et Age. Quels sont les poids respectifs de chacune de ces variables ? Quelles en sont les inerties relatives ? Expliquer pourquoi l'inertie de la variable Stimulus est largement supérieure à celles de chacune des deux autres variables.

4) Examiner les qualités de représentation des modalités colonnes sur chacun des trois axes retenus dans les tableaux de résultats.

5) Etude de la première variable factorielle

a) Quels sont les individus-colonnes dont la contribution est supérieure à la moyenne ? Pour chacun d'eux, préciser le signe de la coordonnée correspondante.

b) Quels sont les individus-lignes dont la contribution est supérieure à la moyenne ? Pour chacun d'eux, préciser le signe de la coordonnée correspondante.

6) Décrire les résultats précédents en termes d'opposition entre individus-colonnes, puis en termes

d'association entre individus-lignes et individus-colonnes.

7) Les données observées auraient pu être étudiées à l'aide d'une AFC, en se limitant aux variables "trait retenu" et "stimulus". Combien de valeurs propres aurait-on alors obtenu ?

8) On aurait également pu étudier les données à l'aide d'une ACM, en utilisant les 4 variables "trait retenu", sexe, stimulus, âge.

a) Comment aurait-il fallu organiser les données pour réaliser cette ACM ?

b) L'une des caractéristiques des données rend l'analyse en composantes multiples peu pertinente. Laquelle ? Pourquoi ?

INDICATIONS DE REPONSES

1) La méthode utilisée ici est une AFC (analyse factorielle des correspondances). En général, cette méthode s'applique à un tableau de contingence. Ici, la structure du tableau est un peu différente ("tableau lexical des questions" selon P. Cibois). En effet, en première colonne, nous retrouvons les modalités de la variable "trait cité", et en première ligne, les modalités de 3 variables : la variable d'intérêt "stimulus" (2 modalités), et deux facteurs : "sexe" (2 modalités) et "âge" (3 modalités).

Chaque citation d'un trait constitue un individu statistique. L'effectif "théorique" est de 480 (48 sujets x 10 traits par sujet). Cet effectif se retrouve - approximativement du fait des non réponses et de l'élimination des traits les moins fréquents - comme somme des colonnes sex1 et sex2 (449), des colonnes sti1 et sti2 (469) ou des colonnes age1 à age3 (449).

L'effectif de chaque ligne représente (approximativement du fait des non réponses) trois fois l'effectif de chaque modalité de trait cité. Chaque case du tableau est l'effectif d'une combinaison de modalités "citation" - modalité d'une autre variable. La somme générale représente trois fois le nombre de citations relevées, et le profil ligne moyen, ou plutôt trois fois ce profil ligne, représente les fréquences des modalités sex1, sex2, sti1, sti2 (environ 50% pour chacune, par construction), age1, age2 et age3.

2) La méthode produit 6 valeurs propres (minimum du nombre de lignes et du nombre de colonnes, moins 1). La dernière valeur propre est nulle (ou presque) en raison des propriétés particulières du tableau (Col 1 + Col 2 = Col 3 + Col 4 = Col 5 + Col 6 + Col 7). Le pourcentage d'inertie moyen par axe est $100\% / 6 = 16,6\%$ ou $100\% / 5 = 20\%$. Seule la première valeur propre dépasse ce seuil.

Autrement dit, l'essentiel de la variation se retrouve sur le premier axe factoriel. Nous pourrions limiter l'étude à cet axe, mais il est commode de pouvoir faire des diagrammes plans. Il n'y a alors aucune raison de couper entre la 2^e et la 3^e valeurs propres, qui sont voisines. On observe par contre, une brusque décroissance entre la 3^e valeur propre et les suivantes.

3) Les poids et inerties des modalités des variables peuvent être lus dans les colonnes "masse" et "inertie relative" du tableau des scores, contributions et qualités de représentation des individus colonnes.

Les poids des 3 variables sont donnés dans le tableau suivant :

Variable	Poids des modalités	Poids de la variable
Sexe	0,178 + 0,151	0,329
Stimulus	0,165+0,178	0,343
Age	0,110 + 0,126 + 0,092	0,328

Les inerties relatives des 3 variables sont quant à elles, les suivantes :

Variable	Inerties des modalités	Inertie de la variable
Sexe	0,053+0,061	0,114
Stimulus	0,374+0,341	0,715
Age	0,075+0,051+0,044	0,170

On constate que les poids diffèrent peu de 0,33, car le tableau est formé en juxtaposant 3 variables. Les écarts par rapport à 0,33 ne sont dus qu'aux non réponses et aux traits peu fréquents éliminés de l'étude.

En revanche, les inerties diffèrent considérablement : la variable stimulus représente plus des 2/3 de l'inertie, la variable sexe n'en représente que 11%, l'âge 17%. Ces résultats traduisent le fait que, pour les variables sexe et âge, les observations s'écartent peu du profil moyen. Le choix de tel ou tel trait est relativement indépendant du sexe ou de l'âge du sujet. En revanche, il existe une forte dépendance entre le choix d'un trait et la nature du stimulus. Il existe de gros écarts par rapport à un profil moyen dans lequel le nombre de choix de chaque trait se répartirait de manière égale entre stimulus 1 (êtres humains) et stimulus 2 (animaux).

4) Les qualités de représentation de chacune des modalités des variables colonnes peuvent être lues dans les colonnes \cos^2 relatives aux 3 axes étudiés, dans le tableau de résultats relatif aux individus colonnes.

Les deux modalités de la variable "stimulus" sont très bien représentées sur le premier axe (\cos^2 égaux à 0,993) et n'interviennent pratiquement pas dans les axes suivants.

Les deux modalités de la variable "sexe" sont mal représentées sur le premier axe (\cos^2 de l'ordre de 0,10), assez bien représentées sur l'axe 2 (\cos^2 de l'ordre de 0,6) et bien représentées dans le plan formé par les axes 2 et 3 (\cos^2 de l'ordre de 0,85).

Pour la variable âge, la représentation est mauvaise sur le premier axe. La modalité âge1 est bien représentée dans le plan (axe 2, axe 3). La modalité âge2 est mal représentée sur les axes 1 et 2, assez bien représentée sur l'axe 3. Enfin la modalité âge3 reste mal représentée dans l'espace formé par les 3 premiers axes. Il faudrait sans doute étudier le 4^e axe pour obtenir une qualité de représentation acceptable.

5) On relève dans la colonne "inertie dimension 1" du tableau des individus colonnes, les modalités correspondant à plus de $100\% / 7 = 14\%$ d'inertie, tout en notant le signe de la coordonnée relevée dans la colonne "Coord. dim. 1". On obtient :

-	+
Sti1 (50%)	Sti2 (46%)

b) Un travail analogue sur les lignes, en fixant cette fois le seuil à $100\% / 33 = 3\%$ donne :

-	+
Spirituel (7%)	Sauvage (11%)
Imaginatif (6%)	Spontané (10%)
Moral (6%)	Intuitif (9%)
Inventif (6%)	Malin (4%)
Egoïste (4%)	Solidaire (4%)
Inconstant (4%)	Impulsif (3%)
Cruel (3%)	
Individualiste (3%)	

6) L'axe 1 oppose clairement la modalité sti1 (êtres humains, partie négative de l'axe) à la modalité sti2 (animaux, partie positive de l'axe). Les traits associés aux êtres humains sont ceux qui caractérisent l'intelligence (spirituel, inventif, imaginatif), mais aussi ceux qui caractérisent l'agressivité "raisonnée" : cruel, individualiste... Les traits associés aux animaux sont liés au comportement instinctif (spontané, sauvage, impulsif), avec aussi quelques traits relatifs au comportement social : solidaire, par exemple. Il convient de noter la position particulière de "malin", cité exclusivement pour les animaux, et essentiellement par les femmes (2/3 des citations) d'où sa position particulière en haut à droite du graphique.

7) Une AFC se limitant aux variables "trait retenu" et "stimulus" n'aurait produit qu'une seule valeur propre différente de 1, car "stimulus" comporte deux modalités seulement. Il aurait été impossible de construire un diagramme selon deux axes tel que celui présenté dans l'énoncé. Toutefois, les résultats n'auraient sans doute pas été très différents de ceux que l'on a obtenu. L'introduction des variables "sexe" et "âge" a essentiellement le mérite "d'étaler" le graphique sur une deuxième dimension, ce qui en rend la lecture plus facile. Mais, il n'était pas évident a priori que les deux variables "sexe" et "âge" auraient aussi peu d'effet sur les choix des sujets.

8) a) Pour réaliser une ACM, avec les 4 variables "trait retenu", "sexe", "stimulus" et "âge", les 4 variables auraient dû jouer des rôles analogues, par exemple toutes les 4 en colonne. Dans un tableau protocole, chaque ligne représente alors une citation d'un trait. Par exemple :

Observation	Trait	Sexe	Stimulus	Age
1	Affectueux	Masculin	Animaux	18-19 ans

Il aurait cependant fallu gérer le problème des valeurs manquantes, vraisemblablement en ajoutant une modalité "NR" à chaque variable, ou "autre" à la variable "trait".

b) Cependant, on sait que dans une ACM, l'importance prise par une variable dépend étroitement du nombre de ses modalités. Ici, les nombres de modalités des différentes variables (33, 2, 2 et 3) sont trop différents les uns des autres pour que l'ACM puisse produire des résultats intéressants.