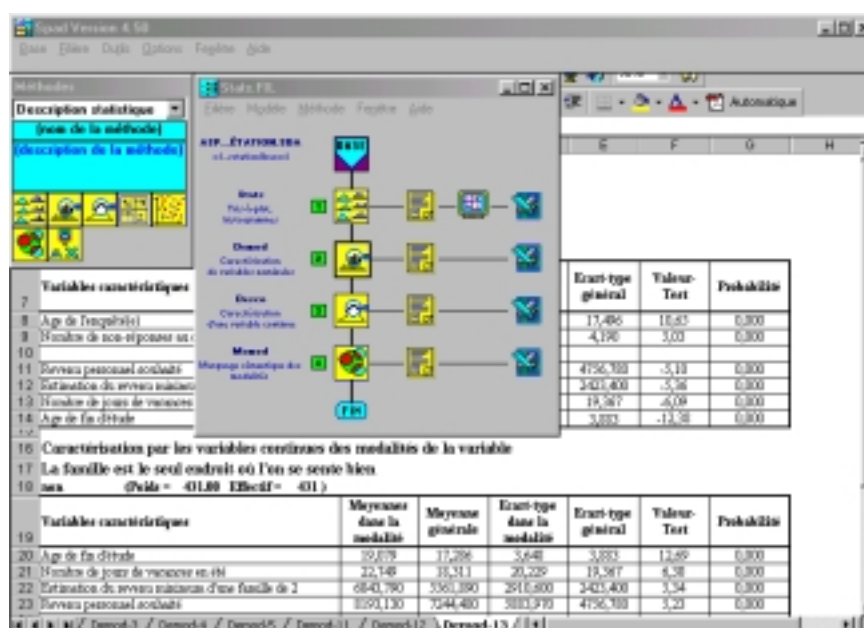


SYSTÈME SPAD POUR WINDOWS

Version **5.0**

SPAD•Base Aide à l'interprétation



CISIA-CERESTA - 261 rue de Paris - 93556 MONTREUIL Cedex
tel : +33 (0)1 55 82 15 15 - Fax : +33 (0)1 43 63 21 00
e-mail : cisia@fr.inter.net - Web : <http://www.cisia.com>

SPAD•Base

Aide à l'interprétation

SPAD•Base®

*Le **logiciel** décrit dans le manuel est diffusé dans le cadre d'un accord de licence d'utilisation et de non divulgation, et ne peut être utilisé ou copié qu'en conformité avec les stipulations de l'accord. Toute copie du programme sur cassette, disque ou autre support à des fins autres que l'usage personnel du programme par le licencié est interdite par la loi. Les informations figurant dans ce **manuel** sont sujettes à révision sans préavis et ne présentent aucun engagement de la part du CISIA.*

© copyright 2001, CISIA•CERESTA. Tous droits réservés

ISBN 2-906711-31-4

Centre International de Statistique et d'Informatique Appliquées

261 rue de Paris, 93556 Montreuil Cedex (France)

Tel : 01 55 82 15 15 – Fax : 01 43 63 21 00

e-mail : cisia@fr.inter.net – Web : <http://www.cisia.com/>

Sommaire

AVANT-PROPOS	8
---------------------------	----------

<u>LES MÉTHODES DE DESCRIPTION STATISTIQUE</u>	9
---	----------

TRIS À PLAT / HISTOGRAMMES (STATS)	10
---	-----------

1. Histogrammes.....	10
2. Variables continues	11
3. Variables nominales	12
4. Variables groupées.....	12
5. Discrétisation	13

TABLEAUX CROISÉS (TABLE)	15
---------------------------------------	-----------

1. Tableau de contingence	15
2. Tableau de moyennes	16
3. Tableaux de fréquence.....	17

CARACTÉRISATION AUTOMATIQUE D'UNE VARIABLE NOMINALE (DEMODO) 18
--

1. DEMODO-1 : Caractérisation de la variable par les variables nominales.....	18
2. DEMODO-2 : Tableaux croisant la variable à décrire avec les variables nominales.....	20
3. DEMODO-3 : Caractérisation d'un groupe d'individus par les variables nominales	21
4. DEMODO-4 : Caractérisation d'une variable nominale par les modalités	22
5. DEMODO-5 : Caractérisation d'un groupe d'individus par les modalités (Tri par valeurs-tests)	23
6. DEMODO-6 : Caractérisation d'un groupe d'individus par les modalités (Tri par % de la modalité dans le groupe).....	26
7. DEMODO-7 : Caractérisation d'un groupe d'individus par les modalités (Tri par % du groupe dans la modalité)	27
8. DEMODO-11 : Caractérisation d'une variable nominale par les variables continues.....	29
9. DEMODO-12 : Statistiques par modalité des variables continues (Tableaux de moyennes).....	30
10. DEMODO-13 : Caractérisation d'un groupe d'individus par les variables continues	31

CARACTÉRISATION AUTOMATIQUE D'UNE VARIABLE CONTINUE (DESCO). 33
--

1. DESCO-1 : Caractérisation par les modalités	33
2. DESCO-2 : Caractérisation par les variables nominales.....	34
3. DESCO-3 : Statistiques sommaires des variables continues.....	35
4. DESCO-4 : Caractérisation par les variables continues.....	35

ANALYSE BIVARIÉE (BIVAR)	37
---------------------------------------	-----------

1. BIVAR-1 : Statistiques sommaires des variables continues actives	37
2. BIVAR-2 : Moyennes et valeurs-test des modalités sur le plan défini par les variables nominales illustratives	38

3. BIVAR-3 : Corrélations entre les variables continues illustratives et les variables continues définissant le plan	39
--	----

MARQUAGE SÉMANTIQUE DE MODALITÉS DE VARIABLES NOMINALES (MSMOD).....	41
---	-----------

LES ANALYSES FACTORIELLES 45

ANALYSE EN COMPOSANTES PRINCIPALES (COPRI) 46

1. Les données et les objectifs	46
2. Choix du thème actif	47
3. COPRI-1 : Statistiques sommaires des variables continues	48
4. COPRI-2 : Matrice des corrélations	49
5. COPRI-3 : Matrice des valeurs-tests	50
6. COPRI-4 : Tableau des valeurs propres	50
7. COPRI-5 : Intervalles laplaciens d'Anderson (seuil: 0.95).....	51
8. COPRI-6 : Coordonnées, Corrélations, Anciens axes unitaires	52
8.1 Coordonnées des variables actives	52
8.2 Corrélations des variables actives avec les facteurs	53
8.3 Anciens axes unitaires	54
8.4 Note sur les Cosinus Carrés et les Contributions	54
9. COPRI-7 : Coordonnées, Corrélations des variables illustratives avec les 5 premiers axes	56
10.COPRI-8 : Matrice des corrélations permutée suivant le premier facteur	57
11.COPRI-9 : Coordonnées, Contributions, Cosinus Carrés des individus actifs	57
11.1 Coordonnées des individus actifs	57
11.2 Contributions des individus actifs	58
11.3 Cosinus Carrés des individus actifs	60
12.COPRI-10 : Coordonnées, Cosinus Carrés des individus illustratifs.....	61
12.1 Coordonnées des individus illustratifs.....	61
12.2 Cosinus Carrés des individus illustratifs	62
13.COPRI-11 : Coordonnées des modalités illustratives	63
14.COPRI-12 : Valeurs-tests des modalités illustratives	64

ANALYSE DES CORRESPONDANCES BINAIRES (CORBI) 66

1. Présentation du tableau analysé	66
2. CORBI-1 : Tableau des valeurs-propres	66
3. CORBI-2 : Coordonnées, Contributions, Cosinus Carrés des fréquences actives	67
3.1 Coordonnées des fréquences actives.....	67
3.2 Contributions des fréquences actives	68
3.3 Cosinus carrés des fréquences actives	69
4. CORBI-4 : Coordonnées, Contributions, Cosinus Carrés des individus actifs	70
4.1 Coordonnées des individus actifs	70
4.2 Contributions des individus actifs	71
4.3 Cosinus carrés des individus actifs.....	72
5. Interprétation générale.....	73

ANALYSE DES CORRESPONDANCES MULTIPLES (CORMU) 74

1. Présentation du thème analysé	74
--	----

2. CORMU-1 : Tris à plat des variables actives.....	75
3. CORMU-2 :Tableau de Burt	77
4. CORMU-3 : Profils horizontaux du tableau de Burt.....	77
5. CORMU-4 : Tableau des valeurs propres	78
6. CORMU-5 :Coordonnées des modalités actives.....	79
7. CORMU-6 :Contributions des modalités actives	80
8. CORMU-7 : Cosinus carrés des modalités actives	82
9. CORMU-8 : Coordonnées des modalités actives et illustratives.....	84
10.CORMU-9 : Valeurs-tests des modalités actives et illustratives	85
11.CORMU-10 : Coordonnées, Contributions, Cosinus Carrés des individus actifs.....	87
11.1 Coordonnées des individus actifs	87
11.2 Contributions des individus actifs	87
11.3 Cosinus Carrés des individus actifs.....	88
12.CORMU-11 : Coordonnées, Cosinus Carrés des individus illustratifs	90
13.CORMU-12 : Corrélations entre les variables continues et les facteurs	90

ANALYSE DES CORRESPONDANCES MULTIPLES AVEC CHOIX DES MODALITÉS ACTIVES (COREM)..... 92

DESCRIPTION DES AXES FACTORIELS (DEFAC)..... 93

1. DEFAC-1 : Description de l'axe par les individus actifs et illustratifs	93
2. DEFAC-2 : Description de l'axe par les modalités actives et illustratives	94
3. DEFAC-3 : Description de l'axe par les variables continues illustratives	96

CLASSIFICATION ET TYPOLOGIE 97

CLASSIFICATION SUR FACTEURS (RECIP / SEMIS) 98

1. Justification du passage aux coordonnées factorielles	98
2. Les techniques de classification	98
3. La Classification Ascendante Hiérarchique – RECIP	99
3.1 CLASSIF-3 : Description des 30 noeuds d'indices les plus élevés	100
3.2 CLASSIF-4 : Description des 30 nouveaux éléments terminaux	101
3.3 CLASSIF-5 : Coordonnées et valeurs-tests des noeuds terminaux.....	101
3.4 CLASSIF-6 : Description des noeuds de la hiérarchie (au dessus des noeuds terminaux)	103
4. La Classification Mixte – SEMIS.....	104
4.1 Paramètres de fonctionnement de SEMIS.....	105
4.2 CLASSIF-1 : Classification mixte sur les 14 premiers axes factoriels.....	106
4.3 CLASSIF-2 : Résultats du croisement	107

COUPURE DE L'ARBRE ET DESCRIPTION DES CLASSES (PARTI/DECLA)... 110

1. La procédure PARTI	110
1.1 PARTI-1 : Coupure 'a' de l'arbre en 9 classes	112
1.2 PARTI-2 : Coordonnées et valeurs-tests des centres de classes avant consolidation	112
1.3 PARTI-3 : Avant consolidation - Liste des individus actifs dans les classes de la partition	114
1.4 PARTI-4 : Avant consolidation - Appartenance des individus actifs dans les classes de la partition	114
1.5 PARTI-5 : Consolidation de la partition.....	115
1.6 PARTI-6 : Décomposition de l'inertie calculée sur 10 axes	115
1.7 PARTI-8 : Coordonnées et valeurs-tests des centres de classes après consolidation	116

1.8	PARTI-9 : Avant consolidation - Liste des individus actifs dans les classes de la partition	116
1.9	PARTI-10 : Avant consolidation - Appartenance des individus actifs dans les classes de la partition.....	116
1.10	PARTI-11 : Edition des parangons	117
1.11	PARTI-12 : Edition des individus à contribution maximale.....	117
2.	La procédure DECLA.....	118

CARACTÉRISATION DES CLASSES DE TYPOLOGIES (CLASS-MINER).....	119
---	------------

MARQUAGE SÉMANTIQUE DES CLASSES D'UNE PARTITION (MSCLA)	120
--	------------

GLOSSAIRE.....	121
-----------------------	------------

BIBLIOGRAPHIE	126
----------------------------	------------

Avant-propos

*Dans la version 5.0 de SPAD, les résultats des différentes méthodes sont mis en forme dans EXCEL. Pour obtenir ces résultats mis en forme, vous devez cocher l'option « **Fichier pour application tableur** » dans l'onglet « **Paramètres** » de la méthode, puis après exécution de la méthode cliquer sur l'icône EXCEL à droite de la méthode.*

Cet habillage dans EXCEL génère des classeurs contenant des feuilles qui pour une méthode donnée contiennent toujours les mêmes types de résultats. L'aide à l'interprétation proposée est faite par rapport à ces feuilles EXCEL.

Les aides à l'interprétation ne contiennent pas de longs développements statistiques, si vous le souhaitez il est préférable de consulter les ouvrages scientifiques diffusés par le CISIA•CERESTA, en particulier :

Statistiques Exploratoire Multidimensionnelle

L. Lebart, A. Morineau, M. Piron (DUNOD)

Auquel il est fait référence dans l'aide en ligne du logiciel, à la rubrique « Objet » de la méthode.

Les Méthodes de Description Statistique



Les méthodes de « Description statistiques » sont les premières étapes indispensables à l'analyse d'un corpus de données.

Les méthodes sont les suivantes :

STATS : Tris à plat, histogrammes, statistiques sur les variables continues.

TABLE : Tableaux croisés (effectifs & pourcentages, moyennes, fréquences)

DEMODO : Description automatique des modalités d'une variable nominale.

DESCO : Description automatique d'une variable continue.

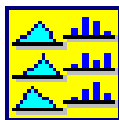
BIVAR : Analyse bivariable (représentation graphique des liaisons entre 2 variables continues)

MSMOD : Marquage sémantique des modalités d'une variable nominale

Après l'indispensable procédure STATS, nous vous conseillons de consulter les sorties des procédures **DEMODO**, **DESCO**, **MSMOD** qui figurent parmi les méthodes les plus puissantes en matière d'analyse exploratoire. La procédure BIVAR est surtout utilisée pour sa représentation graphique.

On travaille sur des données extraites d'une enquête d'opinion réalisée en 1978. Le fichier contient les réponses de 1000 individus sur environ 50 questions extraites d'un questionnaire qui porte sur les conditions de vie et les aspirations des français.

Le fichier de données utilisé est fourni avec le logiciel sous le nom **ASPI1000.SBA**



Tris à plat / Histogrammes (STATS)

Cette procédure fournit, dans un classeur Excel, l'ensemble des statistiques élémentaires sur les variables nominales et continues. Le rapport obtenu constitue la « base statistique » de l'enquête, à laquelle on pourra se référer à tout moment au cours de l'exploitation statistique approfondie qui suivra.

Fichier utilisé : ASPI1000.SBA

Paramétrage de la méthode : On sélectionne toutes les variables nominales pour les tris à plats et toutes les variables continues pour les histogrammes et statistiques sommaires.

1. Histogrammes

Estimation du salaire mensuel d'un ingénieur

Bornes	Poids
Inf. strict. à 3000.00	2
3000.00 - 6000.00	80
6000.00 - 9000.00	438
9000.00 - 12000.00	216
12000.00 - 15000.00	27
15000.00 - 18000.00	29
18000.00 - 21000.00	9
21000.00 - 24000.00	0
24000.00 - 27000.00	0
27000.00 - 30000.00	1
Sup. à 30000.00	4

Chaque variable continue est découpée en N classes d'amplitudes égales. Le nombre de classes est modifiable dans les paramètres. Les données manquantes sont exclues des histogrammes.

Bornes : cette colonne donne les bornes inférieures et supérieures de chacune des classes : la limite inférieure est incluse dans la classe, la limite supérieure est exclue.

Poids : cette colonne donne le poids des individus dans chaque classe.

Si les individus ont un poids uniforme, le poids de la classe est égal au nombre d'individus dans cette classe (effectif).

Il y a 438 individus qui estiment que le salaire mensuel d'un ingénieur est compris entre 6000 et 9000 Francs (Rappel : Il s'agit d'une enquête effectuée en 1978).

2. Variables continues

Pour les variables continues, la procédure STATS édite un tableau rassemblant les principales statistiques dans la feuille Excel « Variables continues ».

Libellé	Effectif	Poids	Moyenne	Ecart-type	Minimum	Maximum
Estimation du salaire mensuel d'un ingénieur	806	806,00	8 478,73	3 668,95	2 000,00	70 000,00
Estimation du revenu mensuel d'un médecin	713	713,00	19 383,90	12 608,80	3 500,00	99 000,00
Age de l'enquête(e)	1 000	1 000,00	42,68	17,50	18,00	90,00
Nombre de non-réponses au questionnaire	1 000	1 000,00	4,05	4,19	0,00	47,00
Age de fin d'étude	997	997,00	17,29	3,88	6,00	39,00
Revenu personnel souhaité	915	915,00	7 244,48	4 756,78	0,00	70 000,00
Estimation du revenu minimum d'une famille de 2 enfants	897	897,00	5 561,89	2 423,40	1 500,00	40 000,00
Nombre de jours de vacances en été	1 000	1 000,00	18,31	19,37	0,00	99,00
coefficient de pondération	1 000	1 000,00	1,00	0,09	0,92	1,20

Tous les calculs sont effectués hors données manquantes (194 individus n'ont pas donné d'estimation du salaire mensuel d'un ingénieur). Les moyennes et les écart-types tiennent compte du poids des individus.

- **Libellé** : Libellé complet de la variable continue.
- **Effectif** : Effectif des individus qui ont répondu.
- **Poids** : Poids des individus qui ont répondu. L'effectif et le poids sont identiques lorsque l'on n'utilise pas de pondération.
- **Moyenne** : Moyenne pondérée de chacune des variables.
- **Ecart-type** : Ecart-type pondéré de chacune des variables.
- **Minimum** : Valeur minimale non pondérée de chacune des variables.
- **Maximum** : Valeur maximale non pondérée de chacune des variables.

Libellé	Effectif	Poids	Moyenne	Ecart-type	Minimum	Maximum
Estimation du salaire mensuel d'un ingénieur	806	806,00	8 478,73	3 668,95	2 000,00	70 000,00
Estimation du revenu mensuel d'un médecin	713	713,00	19 383,85	12 608,83	3 500,00	99 000,00
Age de l'enquête(e)	1 000	1 000,00	42,68	17,50	18,00	90,00
Nombre de non-réponses au questionnaire	1 000	1 000,00	4,05	4,19	0,00	47,00
Age de fin d'étude	997	997,00	17,29	3,88	6,00	39,00
Revenu personnel souhaité	915	915,00	7 244,48	4 756,78	0,00	70 000,00
Estimation du revenu minimum d'une famille de 2 enfants	897	897,00	5 561,89	2 423,40	1 500,00	40 000,00
Nombre de jours de vacances en été	1 000	1 000,00	18,31	19,37	0,00	99,00

Par défaut, les statistiques contiennent deux décimales dans Excel.

3. Variables nominales

Les tris à plat des variables nominales apparaissent dans la feuille Excel « Variables Nominales ». Chaque tri à plat est constitué de 4 composantes :

- **Le libellé complet de la variable** (60 caractères au plus).

Puis pour chaque modalité

- **Effectif** : L'effectif absolu de la modalité, c'est-à-dire le nombre d'individus qui ont choisi cette modalité.
- **% / Total** : La part de l'effectif de la modalité dans l'ensemble des observations y compris les données manquantes.
- **% / Expr.** : La part de l'effectif de la modalité dans l'ensemble des données exprimées c'est-à-dire sans les données manquantes.

Regardez-vous la télévision ...

	Effectif	% / Total	% / Expr.
tous les jours	419	41,90	41,90
assez souvent	226	22,60	22,60
pas très souvent	231	23,10	23,10
jamais	124	12,40	12,40
Total	1 000	100,00	100,00

Pour que la société change, faut-il ...

	Effectif	% / Total	% / Expr.
réformes progres.	490	49,00	63,06
changements radicaux	258	25,80	33,20
ne sait pas	29	2,90	3,73
Total	777	77,70	100,00

Interprétation

Pour la variable « Regardez-vous la télévision », la part des modalités par rapport au total est identique à la part des modalités par rapport aux données exprimées. Cette variable ne contient pas de données manquantes. La variable suivante contient des données manquantes et les pourcentages ne sont pas égaux.

On note que 49% des personnes interrogées pensent que, pour changer la société, il faut des réformes progressives. Mais parmi les réponses exprimées, cette opinion est partagée en fait par 63.06% des individus.

Par défaut, les données manquantes ne sont pas recodées : elles n'apparaissent pas dans une modalité explicite (modifiez le paramétrage si besoin).

Dans le cas de l'utilisation d'une variable de pondération, les sorties dans Excel n'éditent que les résultats pondérés, les sorties fournies dans l'éditeur contiennent les résultats pondérés et non pondérés, permettant de mesurer l'impact de la pondération sur la distribution de chaque variable.

4. Variables groupées

De façon courante dans les enquêtes, on trouve des batteries de variables admettant la même liste de modalités : Possédez-vous (oui / non) un instrument de musique, un ordinateur, un téléphone portable, etc. Ou encore : Etes-vous satisfait (tout à fait / un peu / pas du tout) par la qualité du service, la rapidité de la livraison, l'accueil, etc.

L'option « Variables groupées » est une option d'édition (Tris édités par modalités dans le paramétrage) qui donne les résultats (effectifs et pourcentages) d'une modalité choisie pour chacune des variables. Cette édition peut-être triée.

Le tri par modalités concerne des variables nominales ayant les mêmes intitulés de modalités. Par exemple des variables qui ont 2 modalités : oui et non.

Distribution de la réponse :

oui

Pour les variables :	Effectif	% / Total	% / Expr.
Possession ou usage d'une télévision couleur	373	37,30	37,41
Possession ou usage d'une machine à laver la vaisselle	211	21,10	21,10
Possédez vous des valeurs mobilières ?	121	12,10	12,10
Possédez vous des biens immobiliers ?	81	8,10	8,11

Distribution de la réponse :

non

Pour les variables :	Effectif	% / Total	% / Expr.
Possédez vous des biens immobiliers ?	918	91,80	91,89
Possédez vous des valeurs mobilières ?	879	87,90	87,90
Possession ou usage d'une machine à laver la vaisselle	789	78,90	78,90
Possession ou usage d'une télévision couleur	624	62,40	62,59

Pour chaque modalité, on obtient la liste des fréquences pour l'ensemble des variables nominales choisies.

Le listage des résultats se compose de 4 colonnes :

- **L'intitulé complet de la variable.**
- **Effectif** : l'effectif absolu de la modalité, c'est-à-dire le nombre d'individus qui ont choisi cette modalité.
- **% / Total** : la part de l'effectif de la modalité dans l'ensemble des observations y compris les données manquantes : % / TOTAL.
- **% / Expr** : la part de l'effectif de la modalité dans l'ensemble des données exprimées, c'est-à-dire sans les données manquantes.

L'interface de paramétrage permet de regrouper « à chaud » des modalités.

5. Discrétisation

Pour les variables continues, l'interface de paramétrage vous permet d'en demander soit les « Statistiques sommaires et histogrammes », soit la discrétisation. Cette option consiste à considérer une variable continue comme une variable nominale, chaque valeur de la variable continue étant considérée comme une modalités. Cette option est intéressante pour détecter les valeurs aberrantes et déterminer les bornes de découpages en classes des variables continues. Les résultats sont dans la feuille « Variables discrétisées ». Dans l'exemple on discrétise la variable Age.

Age de l'enquêté(e)

	Effectif	% / Total	% / Expr.	% Cum.
18.000000	19	1,90	1,90	1,90
19.000000	21	2,10	2,10	4,00
20.000000	15	1,50	1,50	5,50
21.000000	21	2,10	2,10	7,60
22.000000	24	2,40	2,40	10,00
23.000000	21	2,10	2,10	12,10
24.000000	29	2,90	2,90	15,00
25.000000	28	2,80	2,80	17,80
78.000000	5	0,50	0,50	97,30
79.000000	10	1,00	1,00	98,30
80.000000	5	0,50	0,50	98,80
81.000000	4	0,40	0,40	99,20
82.000000	3	0,30	0,30	99,50
83.000000	1	0,10	0,10	99,60
84.000000	2	0,20	0,20	99,80
86.000000	1	0,10	0,10	99,90
90.000000	1	0,10	0,10	100,00
Total	1 000	100,00	100,00	

Pour chaque valeur de la variable, on obtient l'effectif (ou le poids si le calcul est pondéré), le pourcentage par rapport à l'ensemble de l'échantillon (**% / Total**), le pourcentage par rapport à ceux qui ont donné une réponses (**% / Expr.**). La colonne % Cum donne le pourcentage cumulé .

Dans notre exemple, 28 personnes ont 25 ans soit 2,8 % de l'échantillon et 17,8 % des personnes ont entre 18 et 25 ans (% cum).



Tableaux Croisés (TABLE)

La procédure TABLE est conçue pour le calcul et l'édition massive de tableaux croisés.

Cette procédure fournit, dans un classeur Excel, l'ensemble des tableaux croisés demandés par l'utilisateur. On peut à partir de cette procédure obtenir des tableaux de contingence, des tableaux de moyenne ou encore des tableaux de fréquence.

Tous les tableaux croisés apparaissent dans la feuille « Tableaux croisés ».

1. Tableau de contingence

Les tableaux de contingence peuvent contenir le poids des individus de chaque case, les pourcentages lignes et les pourcentages colonnes (optionnels). On choisit ici d'éditer ces trois composantes dans un même tableau.

Dans cet exemple, nous croisons la variable V11 (en ligne) avec la variable V25 (en colonne).

Les individus ont un poids uniforme et les individus présentant une donnée manquante ont été abandonnés (option par défaut).

Les marges du tableau apparaissent par défaut.

Dans chaque case, on retrouve les trois statistiques suivantes :

- **Effectifs** : effectif pondéré correspondant au nombre d'individus présentant les deux modalités croisées.
- **% ligne** : pour chaque case d'une ligne, il correspond à l'effectif pondéré de la case sur l'effectif total de la ligne. Pour chaque ligne, on obtient un pourcentage ligne total de 100%.
- **% colonne** : pour chaque case d'une colonne, il correspond à l'effectif de la case sur l'effectif total de la colonne. Pour chaque colonne, on obtient un pourcentage colonne total de 100%.

En ligne En colonne	Sexe de la personne interrogée Votre travail présente-t-il des risques pour la santé ?			
Effectifs % ligne % colonne	beaucoup de risques	peu de risques	aucun risque	ENSEMBLE
masculin	77 23,4% 71,3%	123 37,4% 64,1%	129 39,2% 46,7%	329 100,0% 57,1%
féminin	31 12,6% 28,7%	69 27,9% 35,9%	147 59,5% 53,3%	247 100,0% 42,9%
ENSEMBLE	108 18,8% 100,0%	192 33,3% 100,0%	276 47,9% 100,0%	576 100,0% 100,0%

Interprétation

77 hommes pensent que leur travail présente beaucoup de risque pour la santé.

Le tableau montre également que 424 individus ont été abandonnés puisque le tableau comptabilise 576 réponses alors que le fichier contient 1000 enquêtés.

D'après les pourcentages en ligne, on note que 23.4% des hommes qui ont répondu aux deux questions pensent que leur travail présente beaucoup de risque pour la santé contre seulement 12.5% chez les femmes.

De même, d'après les pourcentages en colonne, on remarque que parmi les personnes qui pensent que leur travail présente beaucoup de risque pour la santé, il y a 71.3% d'hommes et 28.7% de femmes.

2. Tableau de moyennes

Les tableaux de moyennes peuvent contenir le poids des individus de chaque case, la moyenne pondérée et l'écart-type pondéré de la variable continue dans chaque case. Ces trois composantes sont éditées dans le même tableau.

Dans cet exemple, nous souhaitons visualiser l'âge moyen en fonction du sexe et du risque pour la santé que représente le travail. Nous croisons donc les variables V11 et V25 et nous demandons la moyenne de la variable V37.

On a choisi de faire apparaître les données manquantes dans ce tableau (optionnel). Leur présence n'influe pas sur les moyennes et les écarts-types mais permet d'obtenir de l'information sur les individus qui n'ont pas répondu à la question.

Chaque case du tableau contient les trois statistiques suivantes :

- **Moyenne** : moyenne pondérée calculée à partir des valeurs prises sur les individus présentant les deux modalités croisées.
- **Ecart-type** : écart-type pondéré calculé à partir des valeurs prises sur les individus présentant les deux modalités croisées.
- **Effectif** : effectif ou poids correspondant au nombre d'individus présentant les deux modalités croisées.

En ligne	Sexe de la personne interrogée				
En colonne	Votre travail présente-t-il des risques pour la santé ?				
Moyennes de	Age de l'enquêté(e)				
Moyennes Ecart-type Effectifs	beaucoup de risques	peu de risques	aucun risque	reponse manquante	ENSEMBLE
masculin	38,30 11,90 77	38,76 12,33 123	38,16 12,67 129	54,21 21,77 140	43,13 17,35 469
féminin	37,52 13,82 31	37,57 11,97 69	36,71 13,41 147	46,83 19,67 284	42,28 17,61 531
ENSEMBLE	38,07 12,48 108	38,33 12,22 192	37,39 13,09 276	49,27 20,68 424	42,68 17,50 1 000

Interprétation

Sur ce tableau, on note que l'âge moyen des 147 femmes qui pensent que leur travail ne présente aucun risque pour la santé est de 36.7 ans. Elles sont sensiblement plus jeunes que les autres.

La moyenne d'âge des enquêtés qui n'ont pas répondu à la question sur les risques du travail est nettement plus élevée que celle des autres enquêtés : 54.2 ans pour les hommes et de 46.8 ans pour les femmes. On peut supposer que ces cases contiennent notamment des retraités et des personnes sans activité professionnelle.

3. Tableaux de fréquence

Les tableaux de fréquence peuvent contenir en plus de la fréquence, le poids des individus de chaque case, les pourcentages lignes et les pourcentages colonnes. Toutes ces composantes sont éditées dans le même tableau.

Dans cet exemple, nous souhaitons visualiser la fréquence du nombre de jours de vacances en été en fonction du sexe et du risque que représente le travail pour la santé. Nous croisons donc les variables V11 et V25 et nous demandons la fréquence de la variable V49.

On a choisi de faire apparaître les données manquantes dans ce tableau (optionnel).

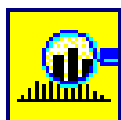
Chaque case du tableau contient les quatre statistiques suivantes :

- **Fréquence** : somme du nombre de jours de vacances pris en été par les individus présentant les deux modalités croisées.
- **Effectif** : effectif ou poids correspondant au nombre d'individus présentant les deux modalités croisées.
- **% ligne** : pour chaque case d'une ligne, il correspond à l'effectif pondéré de la case sur l'effectif total de la ligne. Pour chaque ligne, on obtient un pourcentage ligne total de 100%.
- **% colonne** : pour chaque case d'une colonne, il correspond à l'effectif de la case sur l'effectif total de la colonne. Pour chaque colonne, on obtient un pourcentage colonne total de 100%.

En ligne	Sexe de la personne interrogée				
En colonne	Votre travail présente-t-il des risques pour la santé ?				
Fréquences de	Nombre de jours de vacances en été				
Fréquences Effectifs % ligne % colonne	beaucoup de risques	peu de risques	aucun risque	reponse manquante	ENSEMBLE
masculin	1 033,0 77 13,2% 64,3%	1 981,0 123 25,3% 60,2%	1 927,0 129 24,6% 43,2%	2 894,0 140 36,9% 32,3%	7 835,0 469 100,0% 42,8%
féminin	574,0 31 5,5% 35,7%	1 308,0 69 12,5% 39,8%	2 536,0 147 24,2% 56,8%	6 058,0 284 57,8% 67,7%	10 476,0 531 100,0% 57,2%
ENSEMBLE	1 607,0 108 8,8% 100,0%	3 289,0 192 18,0% 100,0%	4 463,0 276 24,4% 100,0%	8 952,0 424 48,9% 100,0%	18 311,0 1 000 100,0% 100,0%

Interprétation

Les 77 hommes, pour lesquels le travail présente beaucoup de risques pour la santé, ont pris en tout 1033 jours de vacances en été.



Caractérisation automatique d'une variable nominale (DEMODO)

L'intérêt de cette procédure est de caractériser une variable nominale particulière en explorant automatiquement l'ensemble des liaisons qu'elle entretient avec toutes les autres variables du fichier, quelque soit leur type.

Elle est particulièrement adaptée en vue de réaliser un score, une segmentation ou encore une analyse discriminante.

Cette procédure permet l'édition de nombreux tableaux de résultats dans un classeur Excel, qui par défaut, ne sont pas tous édités.

Le tableau suivant résume les possibilités de caractérisation statistique proposées par la procédure DEMODO.

Eléments à caractériser	Eléments caractérisants
<ul style="list-style-type: none">Des groupes d'individus (définis par les modalités de la variable nominale à caractériser) <i>En d'autres termes, on décrit chaque modalité de la variable à caractériser par l'ensemble des éléments caractérisants.</i>	<ul style="list-style-type: none">Les modalitésLes variables nominalesLes variables continues
<ul style="list-style-type: none">La variable nominale à caractériser <i>On recherche parmi tous les éléments caractérisants ceux dont la liaison avec la variable nominale à caractériser est la plus significative.</i>	<ul style="list-style-type: none">Les modalitésLes variables nominalesLes variables continues

Un groupe d'individus est défini par une modalité de la variable à caractériser. Par exemple, les enquêtés qui ont répondu « oui » à la question « La famille est le seul endroit où l'on se sent bien » constituent un groupe d'individus. Il y a donc autant de groupes d'individus que de modalités dans la variable à caractériser. On parlera aussi de *classe* pour faire la distinction entre cette modalité à caractériser et les modalités des variables caractérisantes.

Fichier utilisé : ASPI1000.SBA

Paramétrage de la méthode : Dans cet exemple, la variable à caractériser est la variable V1 « La famille est le seul endroit où l'on se sent bien ? ». Toutes les autres variables du fichier, nominales et continues, sont sélectionnées comme caractérisantes à l'exception du coefficient de pondération (V50).

1. DEMODO-1 : Caractérisation de la variable par les variables nominales

La caractérisation d'une variable nominale par les autres variables nominales s'appuie sur les tableaux croisant la variable à caractériser et les autres variables. On calcule la statistique du Khi-2 associée au croisement des deux variables et la probabilité de dépasser la valeur calculée (test du Khi-2). On associe une valeur-test à cette probabilité. La valeur-test est la valeur de la loi normale centrée réduite qui a la même probabilité d'être dépassée.

Le paramètre *Nominales caractéristiques ordonnées par valeurs-tests* permet de lister les variables nominales par ordre décroissant des valeurs-tests associées (les variables les plus caractéristiques sont en tête). Sinon, elles seront listées dans l'ordre d'apparition dans le fichier de base.

Le tableau ci-après permet d'évaluer les liaisons entre la variable nominale à décrire et les autres variables nominales.

Ce tableau est constitué de 6 colonnes :

- **Libellé de la variable** : libellé complet de chacune des variables nominales.
- **Khi-2** : statistique du Khi-2 associée au croisement de deux variables nominales.
- **Nb. de degrés de liberté** : le nombre de degrés de liberté de la loi du Khi-2 résulte de la formule suivante : $(\ell - 1) * (c - 1)$ où ℓ et c sont les nombres de modalités de la variable caractérisante et de la variable nominale à décrire (y compris les modalités formées par les réponses manquantes).
- **Effectifs théoriques inférieurs à 5** : nombre de cases du tableau comportant un effectif théorique inférieur à 5 (l'effectif théorique est l'effectif que l'on aurait eu en cas d'indépendance entre les 2 variables).
- **Valeur-test** : valeur-test associée à chaque variable nominale caractérisante.
- **Probabilité** : probabilité relative au test du Khi-2.

Le tableau suivant a été tronqué.

Caractérisation par les questions de la variable

La famille est le seul endroit où l'on se sente bien

Libellé de la variable	Khi-2	Nb. de degrés de liberté	Effectifs théoriques inférieurs à 5	Valeur-Test	Probabilité
La famille est le seul endroit où l'on se sente bien	2000,00	4	5	99,99	0,000
Opinion à propos du mariage	323,75	8	7	99,99	0,000
La mère au foyer est un mode de garde ...	235,18	10	7	99,99	0,000
Diplôme de l'enquêté(e) en 5 classes	167,66	8	5	11,74	0,000
Diplôme d'enseignement général le plus élevé obtenu	177,89	12	9	11,65	0,000
Etes-vous gêné par les bruits ?	138,33	6	6	10,78	0,000
Taille d'agglomération (en nombre d'habitants)	135,22	8	5	10,34	0,000
Regardez-vous la télévision ...	125,55	6	4	10,19	0,000
Age de l'enquêté(e) en 5 classes	127,77	8	5	9,99	0,000
Age et sexe de l'enquêteur	84,47	6	4	8,05	0,000
La crèche est un mode de garde ...	91,43	10	8	7,81	0,000

...

Interprétation

La première ligne du tableau correspond au croisement de la variable nominale avec elle-même. Les lignes suivantes indiquent que de toutes les variables, c'est l'opinion à propos du mariage qui est la plus liée avec la variable à caractériser. On trouve ensuite l'opinion sur le mode de garde de la mère au foyer. Les valeurs-tests associées à ces deux variables sont maximales (99.99).

Les résultats de ces tests sont à prendre avec précaution. Les tableaux croisant la variable nominale à décrire avec les deux variables précédentes comportent 7 cases d'effectifs théoriques inférieurs à 5. On pourra vérifier avec l'édition des tableaux qu'il s'agit des cases définies par les données manquantes.

La caractérisation d'une variable nominale par les variables nominales permet d'obtenir l'édition de tableaux de contingence édités dans la feuille Excel intitulée DEMODO-2.

2. DEMODO-2 : Tableaux croisant la variable à décrire avec les variables nominales

Dans l'onglet de paramétrage, il existe cinq items permettant de commander l'édition de ces tableaux :

L'item *Non* spécifie qu'aucun tableau ne sera édité (option par défaut).

L'item *Tous dans l'ordre initial* permet d'éditer tous les tableaux croisés dans l'ordre des variables nominales sur le fichier de base.

L'item *Tous dans l'ordre des valeurs-tests* permet d'éditer ces tableaux dans l'ordre des valeurs-tests décroissantes, c'est-à-dire du plus caractéristique au moins caractéristique.

L'item *Nombre* (50 par défaut) fixe le nombre de tableaux croisés retenus. Les tableaux édités seront ceux dont la valeur-test associée est la plus élevée.

L'item *Seuil* donne en pourcentage le seuil de probabilité (1% par défaut) au-dessous duquel les tableaux seront édités.

Le tableau suivant est un exemple de tableau croisé entre la variable nominale à décrire « La famille est le seul endroit où l'on se sente bien » et une variable nominale caractérisante « Opinion sur le mariage ».

Opinion à propos du mariage

Effectif				
% en ligne	oui	non	*Reponse manquante*	Ensemble
% en colonne				
union indissoluble	196 84,8 34,9	33 14,3 7,7	2 0,9 25,0	231 100,0 23,1
dissout si pb. grave	226 66,1 40,3	114 33,3 26,5	2 0,6 25,0	342 100,0 34,2
dissout si accord	127 32,8 22,6	259 66,9 60,1	1 0,3 12,5	387 100,0 38,7
ne sait pas	12 30,8 2,1	25 64,1 5,8	2 5,1 25,0	39 100,0 3,9
Reponse manquante	0 0,0 0,0	0 0,0 0,0	1 100,0 12,5	1 100,0 0,1
Ensemble	561 56,1 100,0	431 43,1 100,0	8 0,8 100,0	1000 100,0 100,0

Ce tableau donne les pourcentage lignes et colonnes ainsi que l'effectif de chaque case. Cet effectif est la somme des poids des individus de la case quand les individus ont des poids différents. Les pourcentages sont calculés sur les poids des individus des cases (cf. page 15).

Interprétation

On lit dans ce tableau qu'il existe une sur-représentation des enquêtés pensant que l'union est indissoluble parmi ceux qui pensent que la famille est le seul endroit où l'on se sent bien (34.9% contre 23.1% dans la population). D'autre part, les enquêtés estimant que le mariage doit être dissous en cas d'accord entre les deux parties sont nettement sous-représentés (22.6% contre 38.7% dans la population).

3. DEMODO-3 : Caractérisation d'un groupe d'individus par les variables nominales

Pour caractériser un groupe d'individus par les variables nominales on procède de la manière suivante. On compare, à l'aide de la statistique du Khi-2, le profil de la variable dans le groupe d'individus au profil global de la variable dans la population. A la probabilité critique relative au test, on associe une valeur-test. Celle-ci est le fractile de la loi normale centrée réduite correspondant à cette probabilité.

Les variables nominales sont ensuite rangées par ordre décroissant des valeurs-tests. Plus la valeur-test est grande, plus la variable nominale est caractéristique du groupe d'individus.

Trois modes d'édition sont proposés dans les paramètres :

- L'item *Toutes* permet de lister l'ensemble des variables nominales caractérisantes dans l'ordre des valeurs-tests décroissantes, c'est-à-dire de la plus caractéristique à la moins caractéristique.
- L'item *Nombre* (50 par défaut) fixe le nombre de variables nominales à éditer pour chacune des modalités à caractériser. Ce sont les variables dont la valeur-test est la plus élevée (en valeur absolue).
- L'item *Seuil* (paramètre choisi dans l'exemple) fixe le seuil de probabilité en pourcentage (1% par défaut) au-dessous duquel les variables nominales sont éditées. Cela correspond à une valeur-test supérieure à 2.32 en valeur absolue. Pour allonger la liste des variables à éditer, il suffit d'augmenter ce seuil (2% par exemple).

Le tableau ci-après est constitué de 5 colonnes :

- **Variables caractéristiques** : libellé complet de la variable nominale caractérisante.
- **Khi-2** : il s'agit de la valeur de la statistique du Khi-2 associée au test de comparaison des deux profils. Cette statistique du Khi-2 dépend du poids des individus.
- **Nb. de degrés de liberté** : cette colonne donne le nombre de degrés de liberté de la loi du Khi-2. Ce nombre s'obtient en prenant le nombre de modalités de la variable nominale caractérisante (y compris la modalité formée par les éventuelles réponses manquantes) moins un.
- **Valeur-test** : cette colonne indique la valeur-test associée à chacune des variables.
- **Probabilité** : cette colonne donne la probabilité associée à la statistique du Khi-2.

La première ligne du tableau donne le libellé de la modalité à caractériser, ainsi que l'effectif et le poids des individus dans cette modalité.

Dans cet exemple, nous présentons uniquement les résultats de la modalité Oui. Rappelons que la variable à caractériser est la variable V1 « La famille est le seul endroit où l'on se sent bien ? ».

Caractérisation par les questions des modalités de la variable**La famille est le seul endroit où l'on se sente bien**

oui (Poids = 561.000 Effectif = 561)

Variables caractéristiques	Khi-2	Nb. de degrés de liberté	Valeur-Test	Probabilité
La famille est le seul endroit où l'on se sente bien	439,00	2	99,99	0,000
Opinion à propos du mariage	82,52	4	8,30	0,000
Diplôme de l'enquêté(e) en 5 classes	70,06	4	7,55	0,000
Diplôme d'enseignement général le plus élevé obtenu	74,54	6	7,45	0,000
Taille d'agglomération (en nombre d'habitants)	56,77	4	6,66	0,000
Regardez-vous la télévision ...	50,66	3	6,44	0,000
Age de l'enquêté(e) en 5 classes	51,72	4	6,29	0,000
La mère au foyer est un mode de garde ...	42,99	5	5,38	0,000
Age et sexe de l'enquêteur	36,94	3	5,34	0,000
Heure de coucher	39,75	6	4,89	0,000
Participation à une action de défense de l'environnement	15,24	1	3,73	0,000
Statut d'occupation du logement en 4 classes	19,99	3	3,58	0,000
Type d'emploi	21,97	4	3,54	0,000
La préservation de l'environnement est une chose ...	21,51	4	3,48	0,000
Vous arrive-t-il d'inviter des amis à déjeuner ?	16,27	2	3,44	0,000
Statut d'occupation du logement	21,00	4	3,42	0,000
Avez-vous des conflits (travail et vie personnelle) ?	13,45	2	3,04	0,001
Profession de l'enquêté(e) en 7 classes	23,24	7	2,96	0,002
L'enquêté(e) s'est-il (elle) montré(e) intéressé(e) ?	14,12	3	2,78	0,003
Possession ou usage d'une télévision couleur	11,59	2	2,74	0,003

Interprétation

La première variable de ce tableau est la variable de la modalité à caractériser. Sa valeur-test est très grande. Elle est ramenée à un seuil maximal 99.99. Ceci est dû au fait que la modalité à caractériser appartient à cette variable.

La variable sur l'opinion à propos du mariage est celle dont le profil dans la classe est le plus différent du profil global dans l'échantillon. C'est la variable nominale qui a la plus grande valeur-test. Cela signifie que les personnes qui pensent que la famille est le seul endroit où l'on se sente bien n'ont pas la même opinion à propos du mariage que l'ensemble des enquêtés de l'échantillon.

4. DEMOD-4 : Caractérisation d'une variable nominale par les modalités

On compare, pour chaque modalité caractérisante, le profil global de la variable à caractériser au profil de cette variable pour les individus de la modalité. Cette comparaison est effectuée à l'aide de la statistique du Khi-2. On associe une valeur-test à la probabilité associée à la statistique du Khi-2 et on range les modalités par ordre décroissant des valeurs-tests. La valeur-test est le fractile de la loi normale centrée réduite correspondant à cette probabilité.

L'édition des modalités les plus caractéristiques suit la même logique que dans le cas de la caractérisation par les variables nominales. Trois items sont proposés : *Toutes*, *Nombre* et *Seuil* (option choisie dans l'exemple).

Le tableau suivant liste les modalités qui caractérisent le plus la variable nominale « La famille est le seul endroit où l'on se sente bien ». Afin de limiter la longueur de ce tableau, on ne publie ici que les modalités dont la valeur-test est supérieure à 5.

Ce tableau est constitué de 6 colonnes :

- **Libellé des variables** : libellé complet de la variable.
- **Modalités caractérisantes** : libellé complet de la modalité caractérisante.

- **Khi-2** : statistique du Khi-2 associée au test de comparaison des deux profils. La statistique du Khi-2 est calculée avec les poids des individus. Lorsque les individus ont un poids uniforme, on utilise les effectifs.
- **Valeur-test** : valeur-test de chacune des modalités caractérisantes.
- **Probabilité** : probabilité associée à la statistique du Khi-2.
- **Poids** : poids des individus de la modalité caractérisante.

Tous les Khi-2 du tableau ont 2 degrés de liberté. En effet la variable « La famille est le seul endroit où l'on se sente bien » possède deux modalités (« oui » et « non »), et les réponses manquantes forment une troisième modalité.

Caractérisation par les modalités de la variable
La famille est le seul endroit où l'on se sente bien

Libellé des variables	Modalités caractéristiques	Khi-2	Valeur-Test	Probabilité	Poids
La famille est le seul endroit où l'on se sente bien	non	569,00	99,99	0,000	431,00
La famille est le seul endroit où l'on se sente bien	oui	439,00	99,99	0,000	561,00
Opinion à propos du mariage	dissout si accord	89,79	9,14	0,000	387,00
Opinion à propos du mariage	union indissoluble	78,54	8,51	0,000	231,00
Taille d'agglomération (en nombre d'habitants)	Paris	70,26	8,01	0,000	326,00
Diplôme d'enseignement général le plus élevé obtenu	université,gde école	63,09	7,56	0,000	142,00
Diplôme de l'enquêté(e) en 5 classes	Université,gde école	61,19	7,44	0,000	150,00
Regardez-vous la télévision ...	tous les jours	59,22	7,31	0,000	419,00
Age de l'enquêté(e) en 5 classes	65 ans et plus	43,08	6,13	0,000	169,00
Age et sexe de l'enquêteur	femme plus 38 ans	41,10	5,97	0,000	338,00
Diplôme d'enseignement général le plus élevé obtenu	baccalauréat (1/2)	40,69	5,94	0,000	162,00
Age de l'enquêté(e) en 5 classes	Moins de 25 ans	40,35	5,91	0,000	150,00
Taille d'agglomération (en nombre d'habitants)	moins de 2.000	39,59	5,85	0,000	83,00
La mère au foyer est un mode de garde ...	assez satisfaisant	37,61	5,68	0,000	129,00
Diplôme d'enseignement général le plus élevé obtenu	aucun	35,88	5,53	0,000	189,00
Diplôme de l'enquêté(e) en 5 classes	Aucun	35,88	5,53	0,000	189,00
Diplôme de l'enquêté(e) en 5 classes	CEP ou fin études	35,49	5,49	0,000	321,00
Diplôme d'enseignement général le plus élevé obtenu	CEP ou fin études	35,49	5,49	0,000	321,00
Nombre d'enfants considéré comme idéal	aucun	35,36	5,48	0,000	51,00
Regardez-vous la télévision ...	jamais	34,88	5,44	0,000	124,00
Diplôme de l'enquêté(e) en 5 classes	Bac - Brevet sup.	33,31	5,30	0,000	182,00
Participation à une action de défense de l'environnement	oui	30,67	5,05	0,000	126,00

Interprétation

Les deux premières lignes du tableau concernent les propres modalités de la variable à caractériser. Ce sont bien sûr les plus caractéristiques (valeurs-tests bornées à la valeur maximale 99.9).

Les modalités « dissous si accord » et « union indissoluble » de la variable « Opinion à propos du mariage » sont ensuite les modalités qui caractérisent le plus la variable « La famille est le seul endroit où l'on se sente bien ».

5. DEMODO-5 : Caractérisation d'un groupe d'individus par les modalités (Tri par valeurs-tests)

On recherche les modalités des variables caractérisantes (par exemple le sexe de la personne interrogée) qui caractérisent le plus la modalité à décrire (par exemple le groupe des individus qui ont répondu « oui » à la question sur la famille). On examine par exemple s'il y a une sur-représentation des hommes chez les enquêtés qui ont répondu « oui » à la question sur la famille.

Il s'agit d'évaluer l'écart entre le pourcentage de la modalité caractérisante dans la classe (i.e. la modalité à caractériser) et le pourcentage de cette modalité dans la population globale. Pour chacune des modalités des variables nominales, le logiciel calcule la probabilité d'observer un écart au moins aussi grand que celui qui s'est réalisé, dans l'hypothèse où la modalité serait distribuée dans la classe comme dans la population. Cette probabilité évalue en quelque sorte l'importance de l'écart entre les deux pourcentages. Plus la probabilité est faible, plus l'écart est jugé significatif et plus la valeur-test associée à cette probabilité est forte (la valeur-test est le fractile de la loi normale correspondant à la même probabilité).

Dans le menu Options, on trouve les deux paramètres suivants : le *Poids relatif minimal d'une modalité* et les *Modalités conservées*.

Le paramètre *Poids relatif minimal d'une modalité* permet d'écarter de l'édition les modalités dont le poids en pourcentage est jugé trop faible (2% par défaut). Seules les modalités dont le poids est supérieur à ce pourcentage seront utilisables. Si l'on choisit le paramètre *Pourcentage de la modalité dans le groupe* comme critère de tri des modalités, cette option ne s'applique pas.

L'item *Si sur-représentées* du paramètre *Modalités conservées* permet de ne retenir que les modalités dont le pourcentage dans le groupe est plus élevé que le pourcentage dans la population totale. L'item *Si sur ou sous-représentées* (item choisi ici) permet de retenir les modalités sur et sous-représentées. En effet, les modalités rares d'un groupe peuvent avoir de l'intérêt : on caractérise le groupe par les items qui y sont sous-représentés.

Le tableau donnant la caractérisation d'un groupe d'individus par les modalités varie selon le critère de tri et de sélection des modalités adopté dans le paramétrage. Lors du paramétrage, il convient de choisir l'un des critères de tri suivants : *Valeur-test*, *Pourcentage de la modalité dans le groupe* (Sortie DEMOD-6) ou *Pourcentage du groupe dans la modalité* (Sortie (DEMOD-7).

Le paramètre *Valeur-test* permet d'éditer les modalités caractéristiques les plus marquantes pour chaque modalité de la variable nominale à caractériser. Par défaut, le logiciel sélectionne les modalités dont la valeur-test est supérieure à 2 en valeur absolue (seuil de 1% ou probabilité critique 0.01).

Ce tableau se décompose en 8 colonnes :

- **Libellé de la variable** : libellé complet de la variable dont la modalité caractérisante est issue.
- **Modalité caractérisante** : libellé complet de la modalité caractérisante.
- **% de la modalité dans la classe** : équivaut à l'effectif de la modalité dans la classe divisé par l'effectif de la classe.
- **% de la modalité dans l'échantillon** : équivaut à l'effectif de la modalité dans la population globale divisé par l'effectif de l'ensemble de la population.
- **% de la classe dans la modalité** : équivaut à l'effectif de la modalité dans la classe divisé par l'effectif de la modalité dans la population globale.

Les effectifs sont des poids si les individus sont pondérés.

En réalité, les pourcentages précédents sont calculés à partir du poids des individus. Les résultats seront différents si le poids des individus est uniforme (comme c'est le cas ici) ou si les individus ont des poids différents.

- **Valeur-test** : cette colonne donne la valeur-test associée à chacune des modalités.

Lorsque la valeur-test est positive, cela signifie que la modalité est sur-représentée dans la classe. La modalité est sous-représentée si la valeur-test est négative. Si la variable caractérisante a seulement 2 modalités et pas de données manquantes, les valeurs-tests associées sont opposées et égales en valeur absolue.

- **Probabilité** : cette colonne donne la probabilité associée à la valeur-test.
- **Poids** : il s'agit du poids des individus dans la modalité caractérisante : effectif si le poids des individus est uniforme (comme c'est le cas ici) et somme des poids des individus si ceux-ci ont des poids différents.

Les modalités caractérisantes sont rangées par valeur-test décroissante (option par défaut). Les modalités les plus caractéristiques pour lesquelles la valeur-test est la plus grande (et où la probabilité est la plus faible) seront éditées en premier (pour réduire l'encombrement, seules les modalités dont la valeur-test est supérieure à 6 en valeur absolue ont été conservées).

Caractérisation par les modalités des classes de la variable

La famille est le seul endroit où l'on se sente bien

Classe: oui (Effectif: 561 - Pourcentage: 56.10)

Libellés des variables	Modalités caractéristiques	% de la modalité dans la classe	% de la modalité dans l'échantillon	% de la classe dans la modalité	Valeur-Test	Probabilité	Poids
La famille est le seul endroit où l'on se sente bien	oui	100,00	56,10	100,00	36,81	0,000	561
Opinion à propos du mariage	union indissoluble	34,94	23,10	84,85	10,44	0,000	231
Regardez-vous la télévision ...	tous les jours	55,79	41,90	74,70	10,15	0,000	419
La mère au foyer est un mode de garde ...	très satisfaisant	89,48	78,60	63,87	9,46	0,000	786
Age et sexe de l'enquêteur	femme plus 38 ans	44,21	33,80	73,37	7,92	0,000	338
Diplôme d'enseignement général le plus élevé obtenu	CEP ou fin études	41,53	32,10	72,59	7,26	0,000	321
Diplôme de l'enquête(e) en 5 classes	CEP ou fin études	41,53	32,10	72,59	7,26	0,000	321
Age de l'enquête(e) en 5 classes	65 ans et plus	24,06	16,90	79,88	6,99	0,000	169
Taille d'agglomération (en nombre d'habitants)	moins de 2.000	13,37	8,30	90,36	6,97	0,000	83
Diplôme d'enseignement général le plus élevé obtenu	aucun	25,85	18,90	76,72	6,43	0,000	189
Diplôme de l'enquête(e) en 5 classes	Aucun	25,85	18,90	76,72	6,43	0,000	189
Diplôme de l'enquête(e) en 5 classes	Bac - Brevet sup.	11,41	18,20	35,16	-6,20	0,000	182
Statut d'occupation du logement	locataire	43,49	52,30	46,65	-6,26	0,000	523
Statut d'occupation du logement en 4 classes	locataire	43,49	52,30	46,65	-6,26	0,000	523
La mère au foyer est un mode de garde ...	assez satisfaisant	6,77	12,90	29,46	-6,45	0,000	129
Age de l'enquête(e) en 5 classes	Moins de 25 ans	8,38	15,00	31,33	-6,55	0,000	150
Diplôme d'enseignement général le plus élevé obtenu	baccalauréat (1/2)	9,09	16,20	31,48	-6,82	0,000	162
La préservation de l'environnement est une chose ...	très importante	56,51	65,70	48,25	-6,94	0,000	657
Age et sexe de l'enquêteur	femme moins 39 ans	42,60	52,60	45,44	-7,13	0,000	526
Diplôme de l'enquête(e) en 5 classes	Université,gde école	6,77	15,00	25,33	-8,20	0,000	150
Diplôme d'enseignement général le plus élevé obtenu	université,gde école	6,06	14,20	23,94	-8,32	0,000	142
Taille d'agglomération (en nombre d'habitants)	Paris	19,25	32,60	33,13	-10,15	0,000	326
Opinion à propos du mariage	dissout si accord	22,64	38,70	32,82	-11,81	0,000	387
La famille est le seul endroit où l'on se sente bien	non	0,00	43,10	0,00	-35,71	0,000	431
Profession de l'enquête(e) (ou dernière exercée)	salarié agricole	0,00	0,00	0,00	-99,99	0,000	0
Age et sexe de l'enquêteur	inconnu	0,00	0,00	0,00	-99,99	0,000	0

Interprétation

Tout d'abord, on s'intéresse aux modalités sur-représentées dans le groupe d'individus définis par la modalité « oui » de la question sur la famille.

La modalité qui caractérise le mieux ce groupe d'individus est la modalité à caractériser elle-même. La valeur-test associée fixe la limite supérieure de toute valeur-test dans le contexte de cette variable.

On note que la modalité « union indissoluble » de la variable « Opinion à propos du mariage » est celle qui caractérise le mieux la variable à décrire (valeur-test la plus élevée). Parmi les enquêtés qui pensent que la famille est le seul endroit où l'on se sente bien, 34.9% (% de la modalité dans la classe) pensent que l'union est indissoluble. Seulement 23.1% de l'échantillon (% de la modalité dans l'échantillon) a cette opinion à propos du mariage. De plus, 84.8% (% de la classe dans la modalité) des enquêtés qui pensent que le mariage est indissoluble se trouvent dans cette classe.

Si l'on regarde les modalités des variables caractérisantes qui sont sous-représentées, on remarque que celle qui caractérise le mieux la modalité à décrire (valeur-test la plus négative) est la modalité « dissous si accord » de la question sur l'opinion à propos du mariage. Seulement 22.6% des enquêtés estiment que la famille est le seul endroit où l'on se sente bien pensent que le mariage doit être dissous si accord contre 38.7% dans l'échantillon interrogé (Valeur-test = -11.81).

6. DEMODO-6 : Caractérisation d'un groupe d'individus par les modalités (Tri par % de la modalité dans le groupe)

Cette sortie est en tout point identique à la sortie DEMODO-5 à l'exception du mode de tri des modalités caractérisantes.

Le paramètre *Pourcentage de la modalité dans le groupe* permet de lister les modalités les plus *consensuelles* de la classe. Ce sont les modalités qui, parmi les modalités caractéristiques, sont les mieux représentées dans la classe (recouvrement maximum de la classe).

Une façon de typer un groupe d'individus est en effet de regarder les modalités les plus fréquentes dans la classe (pourcentage de la modalité dans la classe). En effet, si tous les individus d'un groupe possèdent une certaine modalité, on peut s'attendre à ce que cette modalité soit une caractéristique de ce groupe. On liste les modalités par ordre décroissant de leur importance dans le groupe d'individus à décrire.

On prendra garde cependant qu'une modalité peut être très fréquente dans un groupe d'individus sans pour autant être caractéristique de ce groupe. Une modalité qui est possédée par 90% de la population globale et par 90% du groupe considéré n'est évidemment pas caractéristique du groupe. C'est pourquoi il convient de se restreindre aux modalités déclarées caractéristiques au vu du critère de la valeur-test (celles qui ont par exemple une valeur-test supérieure à 2 en valeur absolue).

Ce tableau se décompose en 7 colonnes :

- **Libellé de la variable** : libellé complet de la variable dont la modalité caractérisante est issue.
- **Modalité caractérisante** : libellé complet de la modalité caractérisante.
- **% de la modalité dans la classe** : équivaut à l'effectif de la modalité dans la classe divisé par l'effectif de la classe.
- **% de la modalité dans l'échantillon** : équivaut à l'effectif de la modalité dans la population globale divisé par l'effectif de l'ensemble de la population.

Les effectifs sont des poids si les individus sont pondérés.

En réalité, les pourcentages précédents sont calculés à partir du poids des individus. Les résultats seront différents si le poids des individus est uniforme (comme c'est le cas ici) ou si les individus ont des poids différents.

- **Valeur-test** : cette colonne donne la valeur-test associée à chacune des modalités.

Lorsque la valeur-test est positive, cela signifie que la modalité est sur-représentée dans la classe. La modalité est sous-représentée si la valeur-test est négative. Si la variable caractérisante a seulement 2 modalités et pas de données manquantes, les valeurs-tests associées sont opposées et égales en valeur absolue.

- **Probabilité** : cette colonne donne la probabilité associée à la valeur-test.

- **Poids** : il s'agit du poids des individus dans la modalité caractérisante : effectif si le poids des individus est uniforme (comme c'est le cas ici) et somme des poids des individus si ceux-ci ont des poids différents.

On s'est limité aux modalités dont le pourcentage dans la classe est supérieur à 50%.

La famille est le seul endroit où l'on se sente bien

La classe: oui (Effectif: 561 - Pourcentage: 56.10) contient ...

Libellés des variables	Modalités caractéristiques	% de la modalité dans la classe	% de la modalité dans l'échantillon	Valeur-Test	Probabilité	Poids
La famille est le seul endroit où l'on se sente bien	oui	100,00	56,10	36,81	0,000	561
Participation à une action de défense de l'environnement	non	92,87	87,40	5,80	0,000	874
Possédez vous des biens immobiliers ?	non	92,87	91,80	1,28	0,101	918
Faites-vous partie d'une association confessionnelle ?	non	92,34	93,10	-0,95	0,170	931
La mère au foyer est un mode de garde ...	très satisfaisant	89,48	78,60	9,46	0,000	786
A souffert d'état dépressif ces quatre dernières semaines :	non	89,30	87,40	1,95	0,026	874
Possédez vous des valeurs mobilières ?	non	88,24	87,90	0,27	0,393	879
Possession ou usage d'une machine à laver la vaisselle	non	78,97	78,90	0,02	0,493	789
A souffert de nervosité ces quatre dernières semaines :	non	76,47	72,60	3,02	0,001	726
La société française a-t-elle besoin de se transformer ?	oui	71,30	75,90	-3,80	0,000	759
Etes-vous gêné par les bruits ?	pas du tout	65,24	60,60	3,33	0,000	606
Comparée aux personnes de votre âge, votre santé est ...	satisfaisante	62,39	60,00	1,68	0,047	600
La vue sur l'extérieur vous plaît-elle ?	beaucoup	57,22	51,60	3,96	0,000	516
La préservation de l'environnement est une chose ...	très importante	56,51	65,70	-6,94	0,000	657
Regardez-vous la télévision ...	tous les jours	55,79	41,90	10,15	0,000	419
Possession ou usage d'une télévision couleur	non	55,44	62,40	-5,10	0,000	624
A qui incombent les travaux ménagers et les soins enfants ?	homme et femme	55,08	59,90	-3,46	0,000	599
L'enquêté(e) s'est-il (elle) montré(e) intéressé(e) ?	assez	55,08	54,20	0,57	0,285	542
Opinion sur le cadre de vie quotidien	satisfait	54,19	54,90	-0,45	0,328	549
Vous arrive-t-il d'inviter des amis à dîner ?	souvent	52,76	60,60	-5,71	0,000	606
Les découvertes scientifiques améliorent-elles la vie ?	oui, un peu	52,23	50,90	0,89	0,188	509
Sexe de la personne interrogée	féminin	51,34	53,10	-1,20	0,115	531
Vous imposez-vous régulièrement des restrictions ?	oui	51,34	56,90	-3,96	0,000	569
Appartenance à au moins une association	non	50,62	46,40	2,97	0,002	464
Les dépenses de logement sont pour vous ...	pas de gros problème	50,45	44,40	4,30	0,000	444

Interprétation

On compare la colonne % de la modalité dans la classe avec la colonne % de la modalité dans l'échantillon.

Le groupe d'individus à caractériser est constitué des enquêtés qui ont répondu « oui » à la question sur la famille (la première ligne du tableau indique que 100% des individus de ce groupe ont répondu « oui » à cette question).

La seconde ligne du tableau se lit de la façon suivante. Il y a 92.9% des individus du groupe ont répondu « non » à la question « Participation à une action de défense de l'environnement » alors qu'on en trouve en moyenne seulement 87.4% dans l'ensemble de la population. C'est la modalité la plus consensuelle. La valeur-test de 5.80 associée à cette variable mesure de l'importance de cet écart entre les pourcentages.

7. DEMOD-7 : Caractérisation d'un groupe d'individus par les modalités (Tri par % du groupe dans la modalité)

Cette sortie est en tout point identique aux deux sorties précédentes (DEMODO-5 et DEMODO-6) à l'exception du mode de tri des modalités caractérisantes.

Le paramètre *Pourcentage du groupe dans la modalité* permet de lister les modalités les plus exclusives de la classe. Ce sont les modalités qui, parmi les modalités caractéristiques, sont essentiellement dans cette classe et peu dans les autres (débordement minimum hors de la classe).

Une façon de typer un groupe d'individus est de regarder les modalités qui sont bien représentées dans le groupe et peu en dehors de ce groupe : ce sont les exclusivités de la classe. On liste les modalités par ordre décroissant de leur degré d'exclusivité : le pourcentage de la classe dans la modalité.

Il n'est pas suffisant qu'une modalité soit majoritaire dans un groupe d'individus (donc minoritaire à l'extérieur) pour être une exclusivité caractéristique de ce groupe. Avec par exemple seulement 40% dans le groupe d'individus, la modalité peut être caractéristique de ce groupe lorsque les 60% restant sont dispersés au hasard dans 3 autres groupes à raison de 20% par groupe. C'est pourquoi, dans le tableau suivant, on se restreint aux modalités caractéristiques, sélectionnées par les valeurs-tests. Elles permettent de contrôler l'importance de la modalité correspondante pour caractériser la classe.

Ce tableau se décompose en 6 colonnes :

- **Libellé de la variable** : libellé complet de la variable dont la modalité caractérisante est issue.
- **Modalité caractérisante** : libellé complet de la modalité caractérisante.
- **% de la classe dans la modalité** : équivaut à l'effectif de la modalité dans la classe divisé par l'effectif de la modalité dans la population globale.

Les effectifs sont des poids si les individus sont pondérés.

- **Valeur-test** : cette colonne donne la valeur-test associée à chacune des modalités.

Lorsque la valeur-test est positive, cela signifie que la modalité est sur-représentée dans la classe. La modalité est sous-représentée si la valeur-test est négative. Si la variable caractérisante a seulement 2 modalités et pas de données manquantes, les valeurs-tests associées sont opposées et égales en valeur absolue.

- **Probabilité** : cette colonne donne la probabilité associée à la valeur-test.
- **Poids** : il s'agit du poids des individus dans la modalité caractérisante : effectif si le poids des individus est uniforme (comme c'est le cas ici) et somme des poids des individus si ceux-ci ont des poids différents.

Le tableau suivant donne les modalités des variables caractérisantes dont la part dans le groupe est supérieure à 70% (% de la classe dans la modalité). On s'est également limité aux modalités dont la valeur-test est supérieure à 2 (seuil de 1%).

Caractérisation par les modalités des classes de la variable

La famille est le seul endroit où l'on se sente bien

Appartient à la classe: oui (Effectif: 561 - Pourcentage: 56.10)

Libellés des variables	Modalités caractéristiques	% de la classe dans la modalité	Valeur-Test	Probabilité	Poids
La famille est le seul endroit où l'on se sente bien	oui	100,00	36,81	0,000	561
Profession de l'enquêté(e) (ou dernière exercée)	exploitant agricole	90,63	4,11	0,000	32
Taille d'agglomération (en nombre d'habitants)	moins de 2.000	90,36	6,97	0,000	83
Age et sexe de l'enquêteur	homme plus 38 ans	85,71	3,61	0,000	35
Heure de coucher	21h. ou avant	84,93	5,31	0,000	73
Opinion à propos du mariage	union indissoluble	84,85	10,44	0,000	231
A qui incombent les travaux ménagers et les soins enfants ?	incombent à la femme	83,33	3,64	0,000	42
Age de l'enquêté(e) en 5 classes	65 ans et plus	79,88	6,99	0,000	169
La préservation de l'environnement est une chose ...	peu importante	77,78	2,57	0,005	36
Type d'emploi	Autres	77,08	2,93	0,002	48
Diplôme d'enseignement général le plus élevé obtenu	aucun	76,72	6,43	0,000	189
Diplôme de l'enquêté(e) en 5 classes	Aucun	76,72	6,43	0,000	189
Vous arrive-t-il d'inviter des amis à déjeuner ?	jamais	75,00	4,46	0,000	120
Regardez-vous la télévision ...	tous les jours	74,70	10,15	0,000	419
Profession de l'enquêté(e) (ou dernière exercée)	ouvrier spécialisé	74,49	3,84	0,000	98
Age et sexe de l'enquêteur	femme plus 38 ans	73,37	7,92	0,000	338
Profession de l'enquêté(e) en 7 classes	ex. agr.-art-commer	72,63	3,37	0,000	95
Diplôme de l'enquêté(e) en 5 classes	CEP ou fin études	72,59	7,26	0,000	321
Diplôme d'enseignement général le plus élevé obtenu	CEP ou fin études	72,59	7,26	0,000	321
Statut d'occupation du logement	propriétaire	70,69	5,95	0,000	290
Statut d'occupation du logement en 4 classes	propriétaire	70,69	5,95	0,000	290
Evolution du niveau de vie des français depuis 10 ans	beaucoup mieux	70,51	2,59	0,005	78
La crèche est un mode de garde ...	ne sait pas	70,50	3,65	0,000	139
La préservation de l'environnement est une chose ...	assez importante	70,47	5,97	0,000	298
Taille d'agglomération (en nombre d'habitants)	2.000 - 20.000	70,11	2,68	0,004	87

Interprétation

On note sur la première ligne que 100% des enquêtés qui ont répondu « oui » à la question sur la famille se trouvent dans cette classe puisque le groupe est constitué par les individus qui ont choisi la modalité « oui » à cette question.

La seconde ligne du tableau se lit de la façon suivante. 90.6% des enquêtés exploitants agricoles sont rassemblés dans cette classe. Les autres sont répartis en dehors de la classe. C'est la modalité la plus exclusive. La valeur-test 4.11 permet de juger de l'intérêt de cette profession pour caractériser cette classe.

On lit sur la sixième ligne du tableau que 84.8% des individus qui ont choisi l'item « union indissoluble » à la question « opinion sur le mariage » sont rassemblés dans le groupe d'individus. Les autres sont répartis en dehors de ce groupe. La valeur-test 10.44 montre l'intérêt de cet item pour caractériser le groupe (plus caractéristique bien que moins exclusive que la précédente).

8. DEMODO-11 : Caractérisation d'une variable nominale par les variables continues

Pour chaque variable continue caractérisante, on calcule la statistique de Fisher associée à l'analyse de la variance où la variable continue est la variable à expliquer et la variable nominale à décrire est le facteur. Pour chacune de ces statistiques de Fisher, on calcule la probabilité associée. La valeur-test est la valeur de la loi normale centrée réduite qui a la même probabilité d'être dépassée. L'analyse de la variance où la variable continue est « la mieux prévisible » à l'aide du facteur correspond à celle où la statistique de Fisher est la plus significative.

Le paramètre *Continues ordonnées par valeur-test décroissante* permet de lister les variables continues caractérisantes par ordre décroissant des valeurs-tests associées (c'est-à-dire, les plus caractéristiques en tête). Sinon, elles sont listées dans l'ordre d'apparition dans le fichier de base.

Le tableau est constitué de 5 colonnes :

- **Libellé de la variable** : numéro, libellé complet et abrégé de chaque variable continue.
- **Fisher** : statistique de Fisher associée à l'analyse de la variance où la variable à expliquer est la variable continue caractérisante et le facteur est la variable nominale à caractériser.
- **Nb. de degrés de liberté** : cette colonne indique le nombre de degrés de liberté du dénominateur. Ce nombre correspond au second paramètre de la loi de Fisher. Il s'agit du nombre d'individus qui se sont exprimés sur la variable continue moins le nombre de modalités de la variable nominale à décrire (y compris la modalité formée par les réponses manquantes).
- **Valeur-test** : valeur-test associée à chaque variable continue.
- **Probabilité** : probabilité associée à la statistique de Fisher.

Le tableau suivant liste les variables continues les plus caractéristiques de la variable nominale à décrire, ici « La famille est le seul endroit où l'on se sent bien ».

Caractérisation par les variables continues de la variable**La famille est le seul endroit où l'on se sente bien**

Libellé de la variable	Fisher	Nb. de degrés de liberté	Valeur-Test	Probabilité
Age de fin d'étude	96,13	994	13,00	0,000
Age de l'enquêté(e)	67,91	997	10,99	0,000
Nombre de non-réponses au questionnaire	35,65	997	7,93	0,000
Nombre de jours de vacances en été	20,88	997	5,96	0,000
Estimation du revenu minimum d'une famille de 2 enfants	14,81	894	4,91	0,000
Revenu personnel souhaité	14,12	912	4,78	0,000
Estimation du salaire mensuel d'un ingénieur	1,32	803	0,63	0,264
Estimation du revenu mensuel d'un médecin	1,31	710	0,62	0,266

Le premier paramètre de la loi de Fisher (numérateur) est identique pour toutes les variables continues puisqu'il s'agit du nombre de modalités de la variable à décrire (y compris la modalité formée par les réponses manquantes) moins un.

Interprétation

Les variables « Age de fin d'études » et « Age de l'enquête » sont donc les deux variables continues les plus liées à l'opinion « La famille est le seul endroit où l'on se sente bien ».

Le paramètre *Tableaux de moyennes* permet de demander l'édition des statistiques principales (moyenne et écart-type notamment) de la variable continue. Il y aura autant de tableaux édités que de variables continues caractérisantes.

9. DEMODO-12 : Statistiques par modalité des variables continues (Tableaux de moyennes)

L'édition de ces tableaux est commandée par les mêmes paramètres que l'édition des tableaux croisés. On ne présentera ici que le tableau relatif à la variable « Age de l'enquêté(e) ».

Ce tableau est constitué de 7 colonnes :

- **Modalités** : libellé complet de chacune des modalités de la variable nominale à décrire (y compris celle formée par les réponses manquantes).
- **Effectif** : effectif des individus dans la modalité pour les individus qui ont une valeur sur la variable continue.
- **Poids** : effectif pondéré des individus dans la modalité pour les individus qui ont une valeur sur la variable continue.
- **Moyenne** : moyenne pondérée de la variable continue dans la modalité.
- **Ecart-type** : écart-type pondéré de la variable continue dans la modalité.
- **Minimum** : minimum de la variable continue dans la modalité.
- **Maximum** : maximum de la variable continue dans la modalité.

La dernière ligne du tableau donne les statistiques calculées sur l'ensemble de l'échantillon.

Age de l'enquêté(e)

Modalités	Effectif	Poids	Moyenne	Ecart-type	Minimum	Maximum
oui	561	561,00	47,897	17,591	18,000	90,000
non	431	431,00	35,729	14,704	18,000	84,000
Reponse manquante	8	8,00	51,375	18,069	27,000	81,000
Ensemble	1000	1000,00	42,680	17,496	18,000	90,000

Interprétation

Les personnes qui pensent que la famille est le seul endroit où l'on se sente bien ont en moyenne 47.9 ans, contre 42.7 ans dans l'ensemble de l'échantillon. Les personnes dont l'opinion est inverse ont en moyenne 35.7 ans.

10. DEMOD-13 : Caractérisation d'un groupe d'individus par les variables continues

Une variable continue caractérise un groupe d'individus si sa moyenne dans la modalité de la variable à décrire diffère significativement de la moyenne générale de la variable dans la population. Le test mis en œuvre pour évaluer l'écart entre les deux moyennes est un test non-paramétrique de comparaison de moyennes. Plus la valeur-test associée est grande, plus l'écart est significatif et plus la variable est caractéristique (voir l'Annexe A consacrée aux valeurs-tests). Les variables continues sont classées dans l'ordre décroissant des valeurs-tests.

Pour l'édition des moyennes caractéristiques d'un groupe, les paramètres sont les mêmes que dans le cas de la caractérisation d'un groupe d'individus par les variables nominales : *Toutes*, *Nombre* et *Seuil* (option choisie dans l'exemple).

Le tableau suivant liste les variables continues pour lesquelles les moyennes dans la classe (définie par la modalité à décrire) sont les plus différentes des moyennes générales dans l'échantillon.

Caractérisation par les variables continues des modalités de la variable**La famille est le seul endroit où l'on se sente bien**

oui (Poids = 561.00 Effectif = 561)

Variables caractéristiques	Moyennes dans la modalité	Moyenne générale	Ecart-type dans la modalité	Ecart-type général	Valeur-Test	Probabilité
Age de l'enquêté(e)	47,897	42,680	17,591	17,496	10,65	0,000
Nombre de non-réponses au questionnaire	4,643	4,054	4,195	4,190	5,03	0,000
Revenu personnel souhaité	6537,470	7244,480	3512,310	4756,780	-5,10	0,000
Estimation du revenu minimum d'une famille de 2 enfants	5169,370	5561,890	1847,980	2423,400	-5,36	0,000
Nombre de jours de vacances en été	15,013	18,311	17,930	19,367	-6,09	0,000
Age de fin d'étude	15,937	17,286	3,496	3,883	-12,38	0,000

Ce tableau se compose de 5 colonnes :

- **Variables caractéristiques** : libellé complet des variables continues caractéristiques.
- **Moyennes dans la modalité** : moyenne pondérée du groupe d'individus pour chaque variable continue.
- **Moyenne générale** : moyenne pondérée dans l'échantillon pour chaque variable continue.

- **Ecart-type dans la modalité** : écart-type pondéré des individus du groupe.
- **Ecart-type général** : écart-type pondéré dans l'échantillon.
- **Valeur-test** : valeur-test associée à la variable continue.
- **Probabilité** : probabilité issue du test de comparaison de moyennes.

La première ligne du tableau donne le libellé complet de la modalité à caractériser, ainsi que l'effectif, le poids des individus dans la modalité.

Interprétation

On note que les variables « Age de l'enquêté » et « Age de fin d'études » sont les deux variables continues les plus caractéristiques de ce groupe d'individus. Cette classe se caractérise par des personnes plus âgées que la moyenne et dont l'âge de fin d'études est plus faible que la moyenne. Ainsi, les individus de cette classe ont 47.9 ans en moyenne contre 42.7 ans dans l'échantillon. En moyenne, ils ont arrêté leurs études à 15.9 ans contre 17.3 pour l'ensemble de l'échantillon.



Caractérisation automatique d'une variable continue (DESCO)

La procédure DESCO permet d'obtenir la caractérisation d'une ou plusieurs variables continues en explorant l'ensemble des liaisons qu'elle entretient avec toutes les autres variables du fichier quelque soit leur type.

Pour les réponses manquantes relatives à une variable continue, les individus sont éliminés de l'analyse.

Fichier utilisé : ASPI1000.SBA

Paramétrage de la méthode : on caractérise la variable Age de fin d'études (V45). On sélectionne toutes les variables nominales et continues caractérisantes à l'exception des variables nominales redondantes (V49, V51, V52 et V54) et du coefficient de pondération (V50).

1. DESCO-1 : Caractérisation par les modalités

La caractérisation par les modalités permet de détecter les modalités où la moyenne de la variable continue à décrire est notablement différente de la moyenne générale.

Le logiciel effectue les calculs correspondant à un test de comparaison de moyennes. Plus la valeur-test associée à ce test est grande, plus la modalité est caractéristique de la variable continue (ici l'âge de fin d'études).

Le tableau est constitué de 7 colonnes :

- **Libellé de la variable** : libellé complet de la variable nominale associée à la modalité.
- **Modalité caractéristique** : libellé complet de la modalité.
- **Moyenne** : moyenne pondérée de la variable continue pour les individus appartenant à cette modalité.
- **Ecart-type** : écart-type pondéré de la variable continue pour les individus appartenant à cette modalité.
- **Valeur-test** : valeur-test issue du test de comparaison de moyennes.
- **Probabilité** : probabilité issue du test de comparaison de moyennes.
- **Poids** : poids de la modalité ou effectif si le poids des individus est uniforme.

Dans ce tableau, on a choisi de n'éditer que 10 modalités. Il s'agit de celles qui ont la plus grande valeur-test en valeur absolue.

Caractérisation par les modalités de la variable**Age de fin d'étude****Individus actifs (997.000)**

Libellé de la variable	Modalité caractéristique	Moyenne	Ecart-type	Valeur-Test	Probabilité	Poids
	Ensemble	17,29	3,88			997,00
Diplôme d'enseignement général le plus élevé obtenu	université,gde école	23,32	2,97	19,98	0,000	142,00
La famille est le seul endroit où l'on se sente bien	non	19,08	3,65	12,69	0,000	430,00
Profession de l'enquêté(e) (ou dernière exercée)	cadre supérieur	22,83	3,82	12,28	0,000	69,00
Age de l'enquêté(e) en 5 classes	25 à 34 ans	18,94	3,75	8,49	0,000	284,00
Diplôme d'enseignement général le plus élevé obtenu	baccalauréat (1/2)	19,63	2,08	8,39	0,000	162,00
Heure de coucher	entre 23h. et 24h.	19,79	4,20	8,02	0,000	134,00
Regardez-vous la télévision ...	tous les jours	15,74	3,35	-10,64	0,000	416,00
Diplôme d'enseignement général le plus élevé obtenu	aucun	14,40	2,98	-11,22	0,000	186,00
La famille est le seul endroit où l'on se sente bien	oui	15,94	3,50	-12,38	0,000	559,00
Diplôme d'enseignement général le plus élevé obtenu	CEP ou fin études	14,82	1,77	-13,80	0,000	321,00

La première ligne du tableau donne la moyenne pondérée, l'écart-type pondéré et le nombre d'individus renseignés (poids) pour l'âge de fin d'études.

Interprétation

On note que les enquêtés diplômés d'une université ou d'une grande école ont un âge moyen de fin d'études (23.3 ans) très nettement supérieur à la moyenne (17.3 ans). A l'inverse, les enquêtés titulaires d'un certificat d'études primaires (CEP) ou de fin d'études ont un âge moyen de fin d'études (14.8 ans) nettement inférieur à la moyenne (17.3 ans).

2. DESCO-2 : Caractérisation par les variables nominales

Cette caractérisation permet de déterminer les variables nominales liées à la variable continue à décrire.

Le logiciel effectue toutes les analyses de variance où la variable à expliquer est la variable continue à caractériser et le facteur est chaque variable nominale prise successivement. Pour comparer entre elles les statistiques de Fisher issues de ces analyses, on les transforme en valeurs-tests. Ces valeurs-tests sont les fractiles de la loi normale centrée réduite correspondant aux mêmes probabilités critiques. La variable nominale sera d'autant plus caractéristique que la valeur-test sera grande.

Ce tableau se compose de 5 colonnes :

- **Libellé de la variable** : libellé complet de la variable nominale.
- **Nombre de degrés de liberté** : cette colonne indique un nombre de degrés de liberté.

Ces degrés de liberté correspondent au second paramètre de la loi de Fisher. Il s'agit du nombre d'individus qui se sont exprimés sur la variable nominale et qui possèdent une valeur pour la variable continue, moins le nombre de modalités de cette variable.

Le premier paramètre de cette loi est le nombre de modalités de la variable nominale caractérisante moins un.

- **Fisher** : statistique de Fisher associée à chacune des variables nominales.
- **Valeur-test** : valeur-test de la variable nominale, transformation de la statistique de Fisher issue de l'analyse de la variance.

- **Probabilité** : cette colonne donne pour chaque variable nominale la probabilité associée à la statistique de Fisher.

Dans le tableau de résultats suivant, on n'édite que les 10 variables nominales les plus caractéristiques de la variable « Age de fin d'études ».

Caractérisation par les variables nominales de la variable Age de fin d'étude

Libellé de la variable	Nombre de Degrés de liberté	Fisher	Valeur-Test	Probabilité
Diplôme d'enseignement général le plus élevé obtenu	990	308,65	31,84	0,000
Profession de l'enquêté(e) (ou dernière exercée)	979	30,96	18,85	0,000
La famille est le seul endroit où l'on se sente bien	994	96,13	13,00	0,000
Age de l'enquêté(e) en 5 classes	992	40,94	11,68	0,000
Heure de coucher	990	28,55	11,64	0,000
Regardez-vous la télévision ...	993	48,81	11,21	0,000
Vous arrive-t-il d'inviter des amis à déjeuner ?	994	37,43	8,14	0,000
La préservation de l'environnement est une chose ...	992	15,84	6,99	0,000
Participation à une action de défense de l'environnement	995	47,64	6,82	0,000
Taille d'agglomération (en nombre d'habitants)	992	14,28	6,58	0,000

Interprétation

Il existe naturellement une forte liaison entre l'âge de fin d'études et le diplôme d'enseignement le plus élevé. En fonction de la profession exercée, il semble que l'on observe également des différences significatives au niveau de l'âge moyen de fin d'études.

On note également sur ce tableau les liens qui existent entre l'âge de fin d'études et l'heure habituelle de coucher ou avec les réponses à la question « La famille est le seul endroit où l'on se sent bien ? ».

3. DESCO-3 : Statistiques sommaires des variables continues

Les statistiques sommaires des variables continues sélectionnées pour l'analyse sont éditées dans un tableau identique à celui décrit page 11 pour les variables continues de la méthode STATS.

Statistiques sommaires des variables continues

Libellé de la variable	Effectif	Poids	Moyenne	Ecart-type	Minimum	Maximum
Estimation du salaire mensuel d'un ingénieur	804	804,00	8482,39	3672,43	2000,00	70000,00
Estimation du revenu mensuel d'un médecin	711	711,00	19414,50	12613,10	3500,00	99000,00
Age de l'enquêté(e)	997	997,00	42,61	17,47	18,00	90,00
Nombre de non-réponses au questionnaire	997	997,00	4,05	4,18	0,00	47,00
Age de fin d'étude	997	997,00	17,29	3,88	6,00	39,00
Revenu personnel souhaité	912	912,00	7245,29	4761,59	0,00	70000,00
Estimation du revenu minimum d'une famille de 2 enfants	895	895,00	5563,15	2425,96	1500,00	40000,00
Nombre de jours de vacances en été	997	997,00	18,34	19,38	0,00	99,00

4. DESCO-4 : Caractérisation par les variables continues

Pour caractériser la variable continue par les autres variables continues, le logiciel s'appuie sur la corrélation.

A chaque test de corrélation est associée une probabilité critique associée à l'hypothèse de nullité. La probabilité critique est ensuite transformée en valeur-test.

Le tableau est constitué de 5 colonnes :

- **Libellé de la variable** : libellé complet des variables continues caractérisantes.
- **Corrélation** : corrélation entre la variable à caractériser et chacune des variables continues. Il faut noter que la corrélation est ici calculée après élimination des données manquantes.
- **Valeur-test** : valeur-test associée à la probabilité du test de corrélation nulle.
- **Probabilité** : probabilité associée au test de corrélation.
- **Poids** : effectif sur lequel est calculée la corrélation si le poids des individus est uniforme et le poids des individus si les individus ont des poids différents.

Ici, on a choisi de n'éditer que les corrélations dont la probabilité associée est inférieure à 1%.

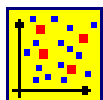
Corrélations avec les variables continues de la variable

Age de fin d'étude

Libellé de la variable	Corrélation	Valeur-Test	Probabilité	Poids
Age de fin d'étude	1,000	99,90	0,000	997,0
Revenu personnel souhaité	0,288	8,94	0,000	912,0
Nombre de jours de vacances en été	0,213	6,81	0,000	997,0
Estimation du revenu minimum d'une famille de 2 enfants	0,113	3,38	0,000	895,0
Nombre de non-réponses au questionnaire	-0,238	-7,64	0,000	997,0
Age de l'enquêté(e)	-0,365	-12,08	0,000	997,0

Interprétation

On note que la corrélation de la variable « Age de fin d'études » avec elle-même est logiquement de 1. La valeur-test associée dépasse le seuil maximum fixé à 99.90.



Analyse bivariée (BIVAR)

L'analyse bivariée est une procédure essentiellement graphique permettant de visualiser les liaisons que deux variables continues prises ensemble entretiennent avec les autres variables d'un fichier. De façon schématique, les deux variables continues seront les axes x et y du graphique, se coupant au point moyen des variables. Les autres variables continues seront des directions dans le plan (comme dans une analyse en composantes principales). Les individus seront des points correspondant à leurs coordonnées dans ce plan. Les modalités des variables nominales seront les points moyens des groupes correspondants d'individus (comme en analyse des correspondances multiples).

Fichier utilisé : ASPI1000.SBA

Paramétrage de la méthode : on croise l'âge de fin d'études (V45) avec le revenu personnel souhaité (V46). Les autres variables du fichier sont toutes sélectionnées comme illustratives à l'exception du coefficient de pondération.

1. BIVAR-1 : Statistiques sommaires des variables continues actives

Le tableau suivant présente les statistiques sommaires des deux variables continues actives de l'analyse.

Ce tableau est constitué de 6 colonnes :

- **Libellé de la variable** : libellé complet de chaque variable continue active.
- **Effectif** : effectif des individus renseignés.
- **Poids** : effectif des individus renseignés après pondération. L'effectif et le poids sont identiques lorsque l'on n'utilise pas de pondération.
- **Moyenne** : cette colonne donne la moyenne pondérée de chacune des variables.
- **Ecart-type** : cette colonne donne l'écart-type pondéré de chacune des variables.

La moyenne et l'écart-type sont calculés sur les enquêtés qui se sont exprimés. Les individus qui ont une donnée manquante ne sont pas pris en compte.

- **Coefficient de variation** : coefficient de variation de chaque variable continue. Le coefficient de variation (C.V) est l'écart-type divisé par la moyenne.

Statistiques sommaires des variables continues actives

Libellé de la variable	Effectif	Poids	Moyenne	Ecart-type	Coefficient de variation
Age de fin d'étude	997	997,00	17,29	3,88	0,22464
Revenu personnel souhaité	915	915,00	7244,48	4756,78	0,65661

Interprétation

On note que le revenu personnel souhaité par les enquêtés se caractérise par un écart-type qui représente près des deux tiers de sa moyenne (65,6%). Cette variable est beaucoup plus dispersée que l'âge de fin d'études.

2. BIVAR-2 : Moyennes et valeurs-test des modalités sur le plan défini par les variables nominales illustratives

Dans le cas de l'analyse bivariée, la description de l'échantillon par les variables nominales donne les tableaux suivants :

Moyennes et valeurs-test des modalités sur le plan défini par les variables:

V46 - Age de fin d'étude

V47 - Revenu personnel souhaité

La famille est le seul endroit où l'on se sente bien

Libellé	Effectif	Poids	Valeurs-Test Axe: V46	Valeurs-Test Axe: V47	Moyennes Axe: V46	Moyennes Axe: V47
oui	561	561,00	-12,38	-5,10	15,94	6595,45
non	431	431,00	12,69	5,23	19,07	8109,49
Reponse manquante	8	8,00	-1,58	-0,68	15,13	6155,56

Opinion à propos du mariage

Libellé	Effectif	Poids	Valeurs-Test Axe: V46	Valeurs-Test Axe: V47	Moyennes Axe: V46	Moyennes Axe: V47
union indissoluble	231	231,00	-4,63	-4,90	16,25	5958,01
dissout si pb. grave	342	342,00	-2,93	2,09	16,79	7661,51
dissout si accord	387	387,00	7,09	2,48	18,38	7694,38
ne sait pas	39	39,00	-0,43	-0,98	17,03	6544,16
Reponse manquante	1	1,00	-0,85	1,70	14,00	15000,00

A qui incombent les travaux ménagers et les soins enfants ?

Libellé	Effectif	Poids	Valeurs-Test Axe: V46	Valeurs-Test Axe: V47	Moyennes Axe: V46	Moyennes Axe: V47
incombent à la femme	42	42,00	-3,70	-2,05	15,12	5831,75
plutôt à la femme	336	336,00	-3,59	2,16	16,67	7681,89
homme et femme	599	599,00	4,65	-1,17	17,75	7107,18
ne sait pas	19	19,00	0,93	-0,20	18,11	7038,60
Reponse manquante	4	4,00	0,50	-0,16	18,25	6875,00

Opinion sur le cadre de vie quotidien

Libellé	Effectif	Poids	Valeurs-Test Axe: V46	Valeurs-Test Axe: V47	Moyennes Axe: V46	Moyennes Axe: V47
très satisfait	259	259,00	-1,15	0,36	17,05	7331,37
satisfait	549	549,00	0,42	-0,33	17,33	7201,99
peu satisfait	145	145,00	0,59	-1,59	17,46	6688,13
pas du tout satisf.	46	46,00	0,23	2,28	17,41	8738,89
Reponse manquante	1	1,00	1,22	2,80	22,00	20000,00

Chaque tableau est constitué de 7 colonnes :

- **Libellé** : libellé complet de chaque modalité,
- **Effectif** : effectif de la modalité, c'est-à-dire le nombre d'individus qui ont choisi cette modalité.

- **Poids** : effectif de la modalité après pondération.
- **Valeur-test axe V46** : cette colonne donne les valeurs-tests associées aux différences de moyennes dans chaque modalité pour la variable V46, à savoir l'âge de fin d'étude.
- **Valeur-test axe V47** : cette colonne donne les valeurs-tests associées aux différences de moyennes dans chaque modalité pour la variable V47 : le revenu personnel souhaité.

Ces valeurs-tests évaluent dans quelle mesure les moyennes des variables continues calculées dans chaque modalité s'écarte de la moyenne générale de ces variables. Elles sont calculées par un test de comparaison de moyennes.

Une valeur-test positive signifie que la moyenne de la variable continue pour les individus de la modalité est supérieure à la moyenne globale de la variable (inversement quand la valeur-test est négative).

- **Moyennes axe V46** : moyennes pondérées de la variable V46 (Age de fin d'étude) pour toutes les modalités des variables nominales sélectionnées.
- **Moyennes axe V47** : moyennes pondérées de la variable V47 (Revenu personnel souhaité) pour toutes les modalités des variables nominales sélectionnées.

Comme précédemment, la moyenne et l'écart-type sont calculés sur les enquêtés de la modalité qui se sont exprimés.

Il faut noter que le poids des individus intervient dans le calcul des moyennes et des valeurs-tests.

Interprétation

Si l'on retient les modalités qui ont les plus fortes valeurs-tests (supérieures à 10 en valeur absolue par exemple), on note que les enquêtés qui ont répondu oui ou non à la variable « La famille est le seul endroit où l'on se sente bien » ont un âge moyen de fin d'études significativement différent de l'âge moyen global de fin d'études. Cet âge moyen est respectivement de 15.94 ans pour les réponses « oui » et 19.07 ans pour les « non » pour une moyenne générale de 17.29 ans dans l'échantillon (voir tableau des statistiques sommaires : Bivar-1).

3. BIVAR-3 : Corrélations entre les variables continues illustratives et les variables continues définissant le plan

L'analyse bivariée donne également la description de l'échantillon par les variables continues illustratives :

Corrélations entre les variables continues illustratives et les variables continues définissant le plan

V46 - Age de fin d'étude

V47 - Revenu personnel souhaité

Numéro	Libellé de la variable	Effectif	Poids	Moyenne	Ecart-type	Corrélation avec: V46	Corrélation avec: V47	Corrélation partielle avec: V46	Corrélation partielle avec: V47
46	Age de fin d'étude	997	997,00	17,29	3,88	1,00	0,28		
47	Revenu personnel souhaité	915	915,00	7244,48	4550,13	0,28	1,00		
15	Estimation du salaire mensuel d'un ingénieur	806	806,00	8478,73	3293,89	-0,03	0,11	-0,07	0,13
19	Estimation du revenu mensuel d'un médecin	713	713,00	19383,90	10646,80	-0,01	0,06	-0,03	0,07
37	Age de l'enquêté(e)	1000	1000,00	42,68	17,50	-0,36	-0,12	-0,35	-0,02
43	Nombre de non-réponses au questionnaire	1000	1000,00	4,05	4,19	-0,24	-0,17	-0,20	-0,11
48	Estimation du revenu minimum d'une famille de 2 enfants	897	897,00	5561,89	2295,21	0,11	0,40	0,00	0,39
49	Nombre de jours de vacances en été	1000	1000,00	18,31	19,37	0,21	0,18	0,17	0,13

Ce tableau est constitué de 10 colonnes :

- **Numéro** : numéro de chaque variable continue correspondant à l'ordre d'apparition dans le fichier de base.
- **Libellé de la variable** : libellé complet de chaque variable continue.
- **Effectif** : effectif des individus renseignés pour chaque variable.
- **Poids** : effectif pondéré des individus renseignés pour chaque variable.
- **Moyenne** : moyenne pondérée de chaque variable continue.
- **Ecart-type** : écart-type pondéré de chaque variable continue.
- **Corrélation avec V46** : coefficient de corrélation entre les variables disponibles et la variable V46 : Age de fin d'étude.
- **Corrélation avec V47** : coefficient de corrélation entre les variables disponibles et la variable V47 : Revenu mensuel souhaité.
- **Corrélation partielle avec V46** : coefficient de corrélation partielle entre les variables disponibles et la variable de base V46.
- **Corrélation partielle avec V47** : coefficient de corrélation partielle entre les variables disponibles et la variable de base V47.

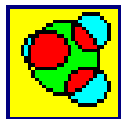
Le coefficient de corrélation partielle d'une variable de base avec une variable disponible est le coefficient de corrélation obtenu après avoir ôté l'influence linéaire de la seconde variable de base.

Interprétation

On note que la corrélation entre les deux variables de base (46 et 47) n'est pas négligeable (0.28). Plus l'âge de fin d'études est grand, plus les enquêtés ont tendance à désirer un revenu personnel élevé.

Les corrélations entre les variables continues de base et les variables continues illustratives sont parfois assez faibles. La corrélation entre les variables 48 et 47 atteint une valeur importante (- 0.40).

En ce qui concerne les coefficients de corrélation partielle, on remarque que ceux-ci ont parfois fortement évolués par rapport aux coefficients de corrélation simple, ce qui s'explique par la corrélation non négligeable entre les deux variables de base.



Marquage sémantique de modalités de variables nominales (MSMOD)

Cette procédure permet de caractériser les modalités d'une variable nominale par des marquages sémantiques.

C'est une généralisation de la procédure DEMOD existante, ou encore la généralisation d'arbres de segmentation.

Un marquage sémantique est une conjonction logique des modalités des variables caractérisantes (Exemple: les hommes de moins de 25 ans habitant en région parisienne).

Si vous utilisez des variables continues comme variables caractérisantes, elles seront découpées à chaud en N classes (paramétrage à spécifier) et considérées comme des variables nominales.

Pour piloter la recherche des marquages sémantiques d'une modalité, il faut définir 2 paramètres fondamentaux :

- **Le seuil (en pourcentage) de débordement admis pour chaque marquage.** S'il est de 15%, tout marquage sémantique dont plus de 15% des individus sont extérieurs à la modalité ou à la classe que l'on caractérise sera rejeté.

Plus le débordement d'un marquage est petit et son recouvrement de la modalité, ou de la classe, est important, plus le marquage est caractéristique.

- **Le seuil (en pourcentage) de recouvrement cumulé souhaité pour la modalité que l'on caractérise.** Si ce seuil vaut 75%, cela veut dire que vous cherchez à recouvrir 75% de la modalité ou de la classe que vous caractérisez. Au delà, le processus de recherche des marquages s'arrête.

A l'exception du premier marquage, le Nième marquage peut recouvrir une partie déjà recouverte par les N-1^{ième} premiers marquages et une partie nouvelle (Ajout du marquage N). Le cumul des recouvrements est l'union des N marquages.

Si le seuil n'est pas atteint, ceci signifie que les autres paramètres de recherche des marquages sont trop contraignants (le seuil de débordement, le seuil de la valeur-test pour une modalité, le seuil de recouvrement nouveau, le seuil de la valeur-test pour un marquage).

- D'autres paramètres permettent d'affiner encore la qualité des marquages (onglet paramètres).

Fichier utilisé : ASPI1000.SBA

Paramétrage de la méthode : on caractérise la variable « La société française a-t-elle besoin de se transformer ? » (V10). On sélectionne ensuite toutes les variables nominales et continues restantes comme caractérisantes à l'exception des variables nominales redondantes (V36, V50, V51, V52 et V54) et du coefficient de pondération (V50).

Dans l'exemple suivant, nous caractérisons les 759 enquêtés ayant répondu Oui à la question « La société française a-t-elle besoin de se transformer ? » par leurs 4 premiers marquages sémantiques.

Les marquages sémantiques d'un groupe sont classés par ordre d'importance à l'aide du critère de "valeur-test" auquel est associé une probabilité : plus la valeur-test est grande (plus la probabilité est faible), plus le marquage est caractéristique.

Description de la modalité: oui

De la variable: La société française a-t-elle besoin de se transformer ?

	Poids	Pourcentage	Valeur-Test	Modalité	Variable
Définition du marquage numéro 1	256	25,60	7,770		
			7,770	très mauvais	Opinion sur le fonctionnement de la justice en 1979
Recouvrement	237	31,23			
Ajout de recouvrement	237	31,23			
Cumul de recouvrement	237	31,23			
Débordement	19	7,42			

	Poids	Pourcentage	Valeur-Test	Modalité	Variable
Définition du marquage numéro 2	108	10,80	5,568		
			5,568	beaucoup moins bien	Evolution du niveau de vie des français depuis 10 ans
Recouvrement	103	13,57			
Ajout de recouvrement	44	5,80			
Cumul de recouvrement	281	37,02			
Débordement	5	4,63			

	Poids	Pourcentage	Valeur-Test	Modalité	Variable
Définition du marquage numéro 3	200	20,00	4,208		
			4,379	locataire	Statut d'occupation du logement
			3,279	Et non	Possédez vous des biens immobiliers ?
			2,497	Et assez mauvais	Opinion sur le fonctionnement de la justice en 1979
Recouvrement	174	22,92			
Ajout de recouvrement	155	20,42			
Cumul de recouvrement	436	57,44			
Débordement	26	13,00			

	Poids	Pourcentage	Valeur-Test	Modalité	Variable
Définition du marquage numéro 4	161	16,10	3,208		
			3,279	non	Possédez vous des biens immobiliers ?
			2,181	Et 34.0000 45.00	Age de l'enquêté(e)
Recouvrement	138	18,18			
Ajout de recouvrement	69	9,09			
Cumul de recouvrement	505	66,53			
Débordement	23	14,29			

Chaque marquage est présenté dans un tableau constitué de 6 colonnes. La lecture de chaque tableau ou marquage s'effectue en trois étapes :

1. Lecture de la première ligne qui présente le marquage de façon générale
2. Lecture de la conjonction de modalités afin d'identifier le marquage
3. Lecture des caractéristiques de recouvrement, d'ajout, de cumul et de débordement.

La **première ligne** de chaque tableau indique le numéro du marquage, son poids dans l'ensemble de l'échantillon, le pourcentage global ainsi que sa valeur-test. Le poids du

marquage correspond au nombre d'individus dans l'échantillon total qui présentent la conjonction de modalités décrites dans la deuxième partie du tableau, après pondération.

Une valeur-test élevée signifie que le marquage est sur-représenté dans la modalité à décrire.

La **conjonction de modalités** définissant le marquage est ensuite décrite dans le tableau à l'aide des trois colonnes suivantes :

- **Valeur-test** : cette colonne donne la valeur-test associée à chacune des modalités.
Lorsque la valeur-test est positive, cela signifie que la modalité caractérisante est sur-représentée dans la modalité à décrire. La modalité caractérisante est sous-représentée si la valeur-test est négative.
Cette valeur-test équivaut à celle que l'on obtient avec la procédure DEMOD (sortie Excel DEMOD-5) en croisant la variable à décrire avec les mêmes variables caractérisantes que celle sélectionnée dans cette procédure.
- **Modalité** : libellé complet de la modalité.
- **Variable** : libellé complet de la variable dont la modalité est issue.

Les caractéristiques de recouvrement, d'ajout, de cumul et de débordement sont décrites ci-après :

- **Recouvrement** : représente dans cet exemple le nombre d'individus inclus dans le marquage et qui ont répondu oui à la question « La société française a t'elle besoin de se transformer ? ».
- **Ajout de recouvrement** : correspond au recouvrement auquel on soustrait les individus déjà inclus dans un marquage précédent. Pour le premier marquage, le recouvrement et l'ajout sont identiques.
- **Cumul de recouvrement** : correspond à la somme des ajouts de recouvrement de tous les marquage précédents plus l'ajout du marquage en cours.
- **Débordement** : représente le nombre d'individus inclus dans le marquage mais qui n'ont pas répondu oui à la question « La société française a t'elle besoin de se transformer ? ».

Le poids du marquage est égal à la somme du recouvrement et du débordement.

Interprétation

Dans cet exemple, nous caractérisons les 759 enquêtés ayant répondu Oui à la question « La société française a t'elle besoin de se transformer ? » par 4 marquages sémantiques.

Le **premier marquage** regroupe les 256 enquêtés de l'étude qui ont répondu que le fonctionnement de la justice en 1979 était très mauvais. Ce marquage représente 25.6% de l'ensemble de l'échantillon. Ce marquage est défini par une seule modalité, ce qui explique le fait que la valeur-test du marquage et celle de la modalité soient identiques.

Cette valeur élevée (7.8) signifie que la proportion d'enquêtés très critiques sur le fonctionnement de la justice est significativement plus élevée parmi ceux qui pensent que la société française a besoin de se transformer que chez les autres.

Ce marquage, qui regroupe en tout 256 enquêtés, recouvre 237 individus parmi les 759 enquêtés qui pensent que la société française a besoin de se transformer, soit 31.23%. Comme il s'agit du premier marquage, l'ajout et le cumul sont égaux au recouvrement.

Parmi les 256 enquêtés concernés par ce marquage, 19 personnes (256-237) ont une opinion qui diffère sur la question des transformations à apporter à la société française, soit 7.42% (inférieur au seuil de débordement toléré par défaut : 15%).

Le **second marquage** concerne les enquêtés qui pensent que le niveau de vie des Français a fortement régressé depuis 10 ans. Ce marquage rassemble en tout 108 individus.

La valeur-test (5.6) de ce marquage, représenté par une seule modalité, signifie que l'opinion très pessimiste sur l'évolution du niveau de vie des français est plus représentée en proportion chez les enquêtés qui pensent que la société a besoin de se transformer que chez les autres.

Ce marquage recouvre en tout 108 individus dont 103 qui pensent que la société française doit se transformer, soit un débordement minimal de 5 individus.

En revanche ce marquage n'ajoute que 44 individus par rapport au précédent. Les 57 autres individus étaient en fait déjà identifiés par le premier marquage.

Au niveau du cumul des deux premiers marquages, 281 individus sur les 759 initiaux ont été identifiés, soit 37.02%.

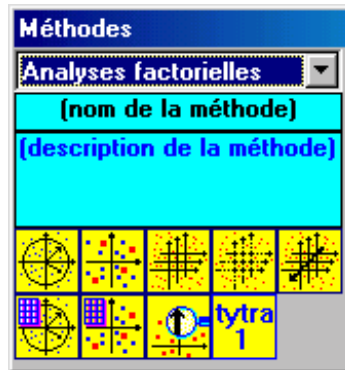
Le **troisième marquage** regroupe 200 enquêtés présentant les trois caractéristiques suivantes dans l'échantillon total :

- Ils sont tous locataires,
- **Et** ils ne possèdent aucun bien immobilier,
- **Et** ils ont une assez mauvaise opinion du fonctionnement de la justice.

Les modalités sont ordonnées dans l'ordre décroissant de leur valeur-test. Dans ce marquage, la plus forte caractéristique est liée au statut d'occupation du logement. En effet, les locataires sont en proportion plus nombreux chez les enquêtés qui pensent que la société française doit se transformer.

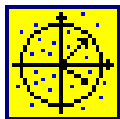
Sur les 200 individus de ce marquage, 174 pensent que la société française doit se transformer. Par rapport aux deux marquages précédent, celui-ci permet d'identifier 155 nouveaux individus, soit un cumul sur les trois marquages de 436 individus en tout.

Les Analyses Factorielles



Les méthodes factorielles établissent des représentations synthétiques de vastes tableaux de données, en général sous forme de représentations graphiques. Ces méthodes ont pour objet de réduire les dimensions des tableaux de données de façon à représenter les associations entre individus et entre variables dans des espaces de faibles dimensions.

Les méthodes d'analyse factorielle consistent à rechercher des sous-espaces de faibles dimensions qui ajustent au mieux le nuage de points des individus et le nuage de points des variables. Les proximités mesurées dans ces sous-espaces doivent refléter au mieux les proximités réelles. L'espace de représentation obtenu est appelé espace factoriel. Les méthodes diffèrent selon la nature des variables analysées : il peut s'agir de variables continues, de variables nominales ou de catégories dans le cas des tableaux de contingences. Les lignes peuvent être des individus ou des catégories. Pour plus de précisions, nous renvoyons l'utilisateur à l'ouvrage de Lebart, Morineau et Piron (1995).



Analyse en composantes principales (COPRI)

L'**Analyse en Composantes Principales** est une technique de description statistique conduisant à des représentations graphiques approchées (mais en un certain sens optimales) du contenu d'un tableau de données: description simultanée des liaisons entre variables et des similitudes entre individus. C'est aussi un outil de réduction de la dimensionnalité d'un ensemble de variables continues, utilisable comme intermédiaire de calcul en vue d'analyses ultérieures.

Dans ce chapitre, nous verrons que l'ACP est un outil exploratoire qui permet de visualiser et de découvrir les phénomènes tels qu'ils sont décrits par les données.

Fichier utilisé : VILLES.SBA

Paramétrage de la méthode : les variables V29 à V40 sont sélectionnées comme actives. Toutes les autres variables continues sont illustratives tout comme la variable nominale Région du monde. Pour la sélection des individus, nous effectuons un filtre logique sur la variable V1 « Année » en ne prenant que « l'édition 1994 ».

1. Les données et les objectifs

L'utilisateur se trouve confronté à un problème qui met en jeu diverses variables continues mesurées sur un grand nombre d'individus et souhaite tirer des enseignements des observations qu'il a faites. Il s'agira par exemple d'étudier un indice qui mesure la puissance économique des individus; ou bien déterminer un sous-ensemble de points adéquats pour contrôler la pollution dans une zone géographique; ou bien, classer les individus en fonction des notes de préférence données à un groupe de produits concurrentiels sur le marché.

Dans d'autres situations, l'Analyse en Composantes Principales servira seulement, ou servira aussi, comme étape intermédiaire de calcul avant une analyse ultérieure: régression, discrimination, ou classification.

Enfin, l'ACP pourra être utilisée comme technique de compression des données.

Les données sont fournies pour une série d'individus, sur lesquels nous avons mesuré un ensemble de variables continues et relevé éventuellement d'autres variables qualitatives.

Les dimensions du tableau de données sont telles qu'il n'est pas possible de repérer directement les individus qui se ressemblent ni les variables qui ont le même comportement sur l'ensemble des individus.

Le système des liaisons entre variables et la configuration des similitudes et différences entre individus, qui feraient la synthèse de l'information contenue dans le tableau de données, sont cachés. Nous verrons comment l'Analyse en Composantes Principales permet, dans une certaine mesure, de les révéler.

Nous raisonnerons de façon concrète sur un exemple. Il ne s'agit pas d'un problème réel mais d'un exercice qui possède certaines vertus pédagogiques sans être trop éloigné d'une situation réelle.

Le problème sera de comparer un certain nombre de grandes villes *selon le niveau moyen des salaires dans une douzaine de professions* afin de vérifier la cohérence de la description par rapport à nos connaissances économiques globales. Ces données sont extraites du fichier Villes.sba.

On s'intéresse à 51 villes. Les données recueillies ne font pas seulement référence aux salaires mais elles constituent un ensemble plus vaste de 40 variables concernant aussi les prix et quelques autres indicateurs essentiellement économiques¹. Les villes sont réparties dans 10 régions du monde (variable numéro 2) et les observations sont connues à deux dates (1991 et 1994: variable n° 1) bien que pour le moment on ne considère que les données de l'édition 1994.

Le tableau des données se présente sous la forme d'un tableau à 40 colonnes et deux fois 51 lignes: chaque ville apparaît deux fois, une première fois pour les observations de 1991, une autre fois pour celles de 1994 (le tableau est ici constitué de l'*empilement* des données des deux périodes).

Le tableau suivant fournit la liste des 40 variables relevées sur chaque ville. Les libellés, bien que peu explicites, suffiront pour l'analyse que l'on souhaite faire.

Libellé des variables	Type
Année	2 Modalités
Region du monde	10 Modalités
I_prix sans loyer	Continue
I_prix avec loyer	Continue
I_salaires bruts	Continue
I_salaires nets	Continue
Heures travail annuelles	Continue
Vacances annuelles payees	Continue
Pouvoir d'achat brut	Continue
Pouvoir d'achat net	Continue
Kg pain=temps de travail	Continue
Hamb=temps de travail	Continue
Denrées alimentaires	Continue
Panier complet	Continue
Dames vêtements	Continue
Hommes vêtements	Continue
4pièces appart meublé	Continue
3pièces appart non meublé	Continue
Loyer normale	Continue
Appareils ménagers	Continue
Bus tram ou metro	Continue
Taxi	Continue
Voitures	Continue
Restaurant	Continue
Nuit d'hôtel	Continue
Services diverses	Continue
Impôts et cotisations sociales en % salaire brut	Continue
Salaire horaire net	Continue
Instituteur	Continue
Chauffeur d'autobus	Continue
Mécanicien autos	Continue
Manoeuvre du bâtiment	Continue
Tourneur	Continue
Cuisinier chef	Continue
Chef de service	Continue
Ingenieur	Continue
Caissier de banque	Continue
Secrétaire direction	Continue
Vendeuse	Continue
Ouvrière du textile	Continue

2. Choix du thème actif

Le tableau de données contient une assez grande quantité de variables, ce qui est une caractéristique habituelle des études appliquées. En fait les variables peuvent se regrouper selon des *thèmes*. Par exemple, on trouve une série de variables relatives aux dépenses sur différents postes (vêtement, loyer, véhicule, services, etc.) dont l'ensemble doit concourir à donner une idée de la cherté de la vie dans chaque ville.

¹ Données recueillies par l'Union de Banques Suisses, dans sa publication "Prices and Earnings Around the Globe", 1991 et 1994

D'autres variables informent sur le niveau des salaires, répartis selon 12 professions. Enfin, d'autres variables donnent plutôt une idée du mode de vie, comme les congés payés, les jours de travail, etc.

Pour comparer les villes entre elles, il est certes possible de prendre en compte toutes les variables disponibles. Cela conduira ici à comparer les villes en tenant compte simultanément du niveau des prix, des salaires, des impôts, des heures nécessaires pour acquérir un hamburger, etc. Les différences observées entre les villes sont alors difficiles à interpréter car elles peuvent avoir des causes multiples et de nature très différentes.

Il est plus "sage" de sélectionner un groupe de variables, ce groupe étant homogène par rapport à un thème bien défini, et cohérent avec l'objectif propre de l'étude. Un thème, c'est-à-dire un groupe de variables, définit un certain *point de vue* choisi par l'utilisateur pour comparer les individus (ici les villes). Ce faisant, l'interprétation des proximités entre villes sera plus facile à faire.

Les variables choisies, appelées **variables actives**, constituent donc les seuls éléments utilisés pour comparer les villes entre elles. Cela ne signifie pas que le reste de l'information soit abandonné: il servira ensuite à illustrer ou peut-être suggérer des "explications" pour les similitudes et différences observées entre les villes. C'est pour cela que l'on appelle **illustratives** ou **supplémentaires** les autres variables.

Toute analyse statistique commence donc par le choix des éléments actifs.

Dans notre exemple, nous décidons de prendre comme variables actives l'ensemble des revenus nets perçus dans les 12 professions retenues. Deux villes seront proches si les rémunérations sont analogues dans l'ensemble de ces 12 professions, indépendamment de ce qui peut les différencier par ailleurs (taille, densité, altitude, etc.). La liste des variables actives est donnée dans le tableau suivant :

Variables actives	Type
Instituteur	Continue
Chauffeur d'autobus	Continue
Mécanicien autos	Continue
Manoeuvre du bâtiment	Continue
Tourneur	Continue
Cuisinier chef	Continue
Chef de service	Continue
Ingenieur	Continue
Caissier de banque	Continue
Secrétaire direction	Continue
Vendeuse	Continue
Ouvrière du textile	Continue

Les autres variables seront utilisées comme illustratives et serviront, comme on le verra, dans la phase d'interprétation des résultats. Nous mettons en œuvre une analyse normée.

3. COPRI-1 : Statistiques sommaires des variables continues

Le tableau suivant présente les statistiques sommaires de l'ensemble des variables continues actives et illustratives de l'analyse dans un tableau constitué de 7 colonnes.

Les variables actives sont listées en premier. Un saut de ligne permet de les différencier des variables continues illustratives.

L'édition du tableau a été tronquée. (CF page 11 pour le descriptif des colonnes qui composent le tableau).

Statistiques sommaires des variables continues

Libellé de la variable	Effectif	Poids	Moyenne	Ecart-type	Minimum	Maximum
Instituteur	51	51,00	16802,000	13243,400	600,000	56800,000
Chauffeur d'autobus	51	51,00	14311,800	10819,600	400,000	46100,000
Mécanicien autos	51	51,00	12384,300	8520,830	700,000	30500,000
Manoeuvre du bâtiment	51	51,00	10343,100	8239,820	200,000	28000,000
Tourneur	51	51,00	15145,100	10244,300	800,000	38700,000
Cuisinier chef	51	51,00	15615,700	8768,420	500,000	33900,000
Chef de service	51	51,00	30933,300	21250,600	1500,000	95000,000
Ingenieur	51	51,00	24664,700	14019,100	1600,000	59700,000
Caissier de banque	51	51,00	18749,000	13413,800	1200,000	58800,000
Secrétaire direction	51	51,00	13311,800	7569,800	1400,000	31500,000
Vendeuse	51	51,00	9658,820	6064,530	400,000	24700,000
Ouvrière du textile	51	51,00	9247,060	6429,780	300,000	23800,000
I. prix sans loyer	51	51,00	68,061	19,355	30,400	132,000
I. prix avec loyer	51	51,00	70,739	19,885	30,200	142,700
I. salaires bruts	51	51,00	41,192	27,781	2,100	100,000
I. salaires nets	51	51,00	39,494	25,596	2,300	100,000
Heures travail annuelles	51	51,00	1920,250	158,688	1669,000	2302,000

4. COPRI-2 : Matrice des corrélations

Le tableau suivant présente la matrice des corrélations entre les variables actives de l'analyse. Dans le cas d'une analyse non normée, nous obtenons la matrice des variances/covariances.

Le tableau se présente donc sous la forme d'une matrice carrée. Seules la partie inférieure et la diagonale de la matrice sont renseignées puisque celle-ci est symétrique.

Matrice des corrélations

	Instituteur	Chauffeur d'autobus	Mécanicien autos	Manoeuvre du bâtiment	Tourneur	Cuisinier chef	Chef de service	Ingenieur	Caissier de banque	Secrétaire direction	Vendeuse	Ouvrière du textile
Instituteur	1,00											
Chauffeur d'autobus	0,96	1,00										
Mécanicien autos	0,84	0,89	1,00									
Manoeuvre du bâtiment	0,83	0,88	0,95	1,00								
Tourneur	0,91	0,94	0,93	0,93	1,00							
Cuisinier chef	0,75	0,76	0,80	0,72	0,76	1,00						
Chef de service	0,78	0,74	0,64	0,59	0,69	0,82	1,00					
Ingenieur	0,81	0,82	0,74	0,70	0,80	0,82	0,87	1,00				
Caissier de banque	0,82	0,80	0,70	0,64	0,72	0,79	0,89	0,85	1,00			
Secrétaire direction	0,92	0,93	0,88	0,86	0,92	0,80	0,80	0,87	1,00			
Vendeuse	0,88	0,89	0,89	0,86	0,88	0,85	0,79	0,85	0,85	0,94	1,00	
Ouvrière du textile	0,88	0,92	0,89	0,92	0,94	0,71	0,65	0,81	0,73	0,93	0,89	1,00

Le coefficient de corrélation linéaire indique la force de la liaison **linéaire** entre deux variables continues. Ce coefficient prend des valeurs comprises entre -1 et 1 .

Plus le coefficient prend une valeur proche de 1 , plus la liaison linéaire entre les variables X et Y est forte et croissante. A une valeur élevée de X , on associera une valeur élevée de Y .

Plus le coefficient prend une valeur proche de -1 , plus la liaison linéaire entre les variables X et Y est forte et décroissante. A une valeur élevée de X , on associera une valeur faible de Y .

Si le coefficient prend une valeur proche de 0 , cela indique qu'il n'y a pas de liaison linéaire entre les variables X et Y . La relation entre X et Y peut être d'un autre ordre mais en tout cas, elle n'est pas linéaire. Attention, cela ne veut pas dire que les variables sont indépendantes.

Interprétation

Dans cette matrices de corrélations, nous lisons la force des associations linéaires entre les variables deux à deux.

Si un salaire est élevé dans une ville, les autres salaires le sont également. Les corrélations sont toutes positives et en général fortes, avec des valeurs comprises entre 0.59 et 0.96 .

5. COPRI-3 : Matrice des valeurs-tests

Cette matrice est directement liée à la matrice précédente. En effet, on retrouve la même matrice carrée croisant les variables continues actives deux à deux. L'objectif de cette matrice est de répondre à la question suivante : "Est-ce que la valeur du coefficient de corrélation observée entre 2 variables permet de dire que leur liaison est significative ?".

Dans un souci d'homogénéité, SPAD a retranscrit le test sous-jacent de nullité de la corrélation, en terme de valeur-test. Plus la valeur test sera élevée et plus la liaison linéaire sera forte. On peut également affirmer qu'une valeur-test inférieure à 2 (en valeur absolue) indique qu'il n'y a pas de liaison linéaire entre les variables.

Matrice des valeurs-tests

	Instituteur	Chauffeur d'autobus	Mécanicien autos	Manoeuvre du bâtiment	Tourneur	Cuisinier chef	Chef de service	Ingenieur	Caissier de banque	Secrétaire direction	Vendeuse	Ouvrière du textile
Instituteur	99,99											
Chauffeur d'autobus	13,87	99,99										
Mécanicien autos	8,82	10,07	99,99									
Manoeuvre du bâtiment	8,51	9,69	12,74	99,99								
Tourneur	10,82	12,31	11,91	11,90	99,99							
Cuisinier chef	6,89	7,09	7,85	6,50	7,03	99,99						
Chef de service	7,47	6,79	5,44	4,88	6,10	8,34	99,99					
Ingenieur	8,11	8,28	6,81	6,22	7,76	8,37	9,64	99,99				
Caissier de banque	8,33	7,86	6,19	5,46	6,51	7,64	10,27	9,10	99,99			
Secrétaire direction	11,12	11,73	9,79	9,18	11,18	7,86	7,93	9,63	9,62	99,99		
Vendeuse	9,72	10,20	10,20	9,14	9,96	9,08	7,74	9,02	8,94	12,15	99,99	
Ouvrière du textile	9,89	11,58	10,19	11,29	12,22	6,35	5,58	8,11	6,65	11,83	10,32	99,99

Interprétation

Dans cet exemple, aucune valeur-test n'est inférieure à 2. Nous sommes donc en présence de variables continues linéairement dépendantes. Nous pouvons tout de même établir une hiérarchie dans la force des liaisons.

6. COPRI-4 : Tableau des valeurs propres

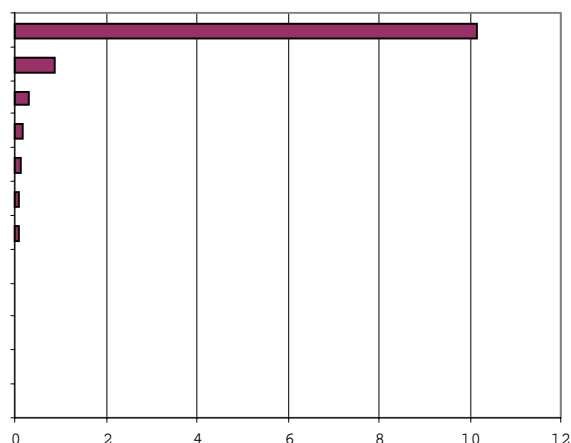
Cette sortie présente l'ensemble des valeurs propres de l'analyse.

Le tableau est constitué des éléments suivants :

- **Trace de la matrice** : correspond à la valeur de l'inertie totale du nuage de points étudié. Cette trace est égale à la somme des variances de chaque variable active. En analyse normée (ici), la variance associée à chaque variable active étant égale à 1, la trace de la matrice est égale au nombre de variables actives, c'est à dire 12.
- **Numéro** : correspond au numéro de l'axe factoriel. En ACP, le nombre d'axes factoriels est toujours égal au nombre de variables actives.
- **Valeur propre** : correspond à l'inertie interceptée par l'axe factoriel. La variance associée à la projection des individus actifs sur le premier facteur vaut 10,1390. La somme des valeurs propres est égale à l'inertie globale du nuage des villes, et elle est égale à 12, le nombre des variables actives, car nous réalisons une ACP normée
- **Pourcentage** : indique la part de l'inertie totale prise en compte par chaque axe factoriel. Nous voyons que le premier axe porte près de 85% de l'inertie du nuage de points qui est initialement contenu dans un espace de dimension 12. Ce pourcentage extrêmement élevé s'explique par le fait que nous sommes en présence de variables fortement corrélées linéairement.
- **Pourcentage cumulé** : édite la part de l'inertie totale prise en compte par tous les axes factoriels en sommant les pourcentages comprenant l'axe étudié et ses prédécesseurs.

Tableau des valeurs propres**Trace de la matrice: 12.00000**

Numéro	Valeur propre	Pourcentage	Pourcentage cumulé
1	10,1390	84,49	84,49
2	0,8612	7,18	91,67
3	0,3248	2,71	94,38
4	0,1715	1,43	95,80
5	0,1484	1,24	97,04
6	0,0973	0,81	97,85
7	0,0682	0,57	98,42
8	0,0525	0,44	98,86
9	0,0505	0,42	99,28
10	0,0332	0,28	99,55
11	0,0309	0,26	99,81
12	0,0226	0,19	100,00

**Interprétation**

Nous obtenons une première composante principale très dominante, expliquant près de 85% de l'inertie totale. Avec seulement les trois premiers axes factoriels, nous avons près de 95% de l'inertie (ou de la dispersion) de tout le nuage de points.

Combien d'axes faut-il retenir ?

Avec les premières composantes principales, nous recomposons la majeure partie de la dispersion entre les points. Mais combien d'axes faut-il retenir?

La question est difficile et n'admet certainement pas de réponse définitive. Il faut d'abord savoir à quoi l'on destine les axes que l'on veut retenir. Prenons quelques exemples:

On peut vouloir retenir les axes pour faire des représentations graphiques qu'il faudra examiner visuellement. Dans ce cas le nombre d'axes le plus confortable est 2, car on regardera les nuages sur une feuille de papier: ce sont les axes F_1 et F_2 . On peut essayer de représenter le "volume" dans l'espace (F_1 , F_2 et F_3), mais les réalisations graphiques sont souvent difficiles à lire. On ne peut rien voir au delà de trois dimensions. On peut imaginer regarder des projections telles que (F_2 , F_3) ou (F_1 , F_4) mais là encore, l'effort "intellectuel" est énorme pour visualiser les configurations de points où les distances lues sont le résultat de compressions dans des directions qui ont disparu. S'il s'agit de visualisations planes donc, on se contentera le plus souvent du premier plan factoriel. Pour "voir" au delà de ce plan, on aura recours en général à la complémentarité apportée par les méthodes de classification.

Si on retient des axes factoriels comme support pour réaliser une classification, le problème est différent. On cherche à retenir beaucoup d'axes (pour rester aussi proche que possible des données observées), mais on supprimera les directions correspondant aux plus petites valeurs propres (fluctuations aléatoires des données, ne correspondant à aucune liaison réelle entre les variables) pour assainir l'information utilisée pour comparer les individus.

7. COPRI-5 : Intervalles laplaciens d'Anderson (seuil: 0.95)

Le logiciel édite ici les bornes supérieures et inférieures des intervalles de confiance approchés au seuil 95% des valeurs propres correspondant à chaque axe factoriel.

L'ampleur de l'intervalle donne une indication sur la stabilité de la valeur propre vis-à-vis des fluctuations dues à l'échantillonnage. L'empiètement des intervalles de deux valeurs propres consécutives suggèrera donc l'égalité de ces valeurs propres. Les axes correspondants sont alors définis à une rotation près. Ainsi, l'utilisateur pourra éviter d'interpréter un axe instable selon ce critère.

Ce tableau est constitué de 4 colonnes :

- **Numéro** : correspond au numéro de l'axe factoriel. En ACP, le nombre d'axes factoriels est toujours égal au nombre de variables actives.
- **Borne inférieure** : valeur inférieure de l'intervalle de confiance.
- **Valeur propre** : correspond à l'inertie interceptée par l'axe factoriel. La variance associée à la projection des individus actifs sur le premier facteur vaut 10,1390.
- **La borne supérieure** : valeur supérieure de l'intervalle de confiance.

Intervalles laplaciens d'Anderson (seuil: 0.95)

Numéro	Borne inférieure	Valeur propre	Borne supérieure
1	6,1645	10,1390	14,1135
2	0,5236	0,8612	1,1988
3	0,1975	0,3248	0,4521
4	0,1042	0,1715	0,2387
5	0,0902	0,1484	0,2066

Interprétation

On est sûr à 95% que le premier axe de l'analyse interceptera une inertie comprise entre [6,16 ; 14,11].

Les intervalles de confiance d'Anderson concernent aussi bien les valeurs propres des matrices de covariances que des matrices de corrélations.

8. COPRI-6 : Coordonnées, Corrélations, Anciens axes unitaires

8.1 Coordonnées des variables actives

Le tableau ci-après donne les coordonnées des variables actives. Par défaut, le logiciel édite ces valeurs sur les 5 premiers axes.

Ce tableau se compose de 6 colonnes :

- **Libellé** : libellé complet de la variable
- **Axe 1 – Axe 5** : coordonnées des variables actives sur les premiers axes. On remarque que, dans le cas d'une analyse normée, la coordonnée d'une variable sur un axe coïncide avec la corrélation de cette variable avec la composante principale (projection des individus sur l'axe factoriel de même rang).

Coordonnées des variables actives

Libellé de la variable	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5
Instituteur	-0,94	-0,04	-0,21	0,16	0,13
Chauffeur d'autobus	-0,96	-0,13	-0,15	0,08	0,09
Mécanicien autos	-0,92	-0,27	0,19	0,07	-0,03
Manoeuvre du bâtiment	-0,90	-0,37	0,11	0,01	-0,02
Tourneur	-0,95	-0,24	-0,02	-0,01	0,11
Cuisinier chef	-0,87	0,24	0,40	0,05	0,06
Chef de service	-0,84	0,49	-0,01	0,04	0,12
Ingenieur	-0,90	0,27	-0,03	-0,30	0,09
Caissier de banque	-0,88	0,38	-0,13	0,11	-0,20
Secrétaire direction	-0,97	0,00	-0,10	-0,03	-0,11
Vendeuse	-0,96	0,01	0,08	0,00	-0,17
Ouvrière du textile	-0,94	-0,25	-0,10	-0,17	-0,06

Interprétation

Les coordonnées des variables ont toutes le même signe sur le premier axe et sont comprises entre -0,84 et -0,97. Ce sont des valeurs extrêmement élevées en valeur absolue. Il est clair que le nuage des points-variables n'est pas centré.

Dans cet exemple, la première composante principale naît de la forte corrélation entre les variables actives. La première composante se situe alors graphiquement au milieu du faisceau des variables actives.

Cette situation peut s'interpréter de la façon suivante. D'une façon générale pour toute ville, si un salaire est élevé sur un poste, il est élevé sur l'ensemble de postes; s'il est faible sur un poste, il est faible sur l'ensemble. C'est la généralité de cette observation sur l'ensemble du tableau qui constitue la structure la plus forte des données et engendre le premier facteur.

Cette caractéristique, fréquemment rencontrée en ACP sur le premier axe se nomme *facteur de taille*.

Le premier axe de l'analyse oppose d'une part les villes dont le salaire moyen est élevé aux villes dont le salaire moyen est plus faible.

Le second axe oppose d'autre part les villes qui payent relativement bien le chef de service, le caissier de banque et le cuisinier chef des villes qui payent relativement mieux les professions socialement moins considérées comme Manoeuvre du bâtiment, Mécanicien Auto, Tourneur et Ouvrière du textile.

Facteur de taille

Le fait le plus marquant de cette analyse est le facteur taille, très dominant, porté par la première composante. Ce facteur reflète pratiquement exclusivement la disparité des villes quand au niveau moyen des salaires. Les autres facteurs sont en quelque sorte "écrasés" par la force de ce phénomène dans le tableau des données.

Dans ce cas, il peut être intéressant de reprendre l'analyse en cherchant à éliminer des données cette connaissance que nous avons sur les salaires des villes.

On y parvient, par exemple, en divisant les salaires de chaque profession par le salaire moyen de la ville.

8.2 Corrélations des variables actives avec les facteurs

Une corrélation élevée indique que la disposition des individus sur l'axe factoriel est semblable à l'ordonnance des individus selon la variable (la corrélation unité signifierait que

la composante principale est fonction linéaire de la variable). Une corrélation proche de zéro indique qu'il n'y a pas d'association linéaire entre la composante principale et la variable.

Corrélations des variables actives avec les facteurs

Libellé de la variable	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5
Instituteur	-0,94	-0,04	-0,21	0,16	0,13
Chauffeur d'autobus	-0,96	-0,13	-0,15	0,08	0,09
Mécanicien autos	-0,92	-0,27	0,19	0,07	-0,03
Manoeuvre du bâtiment	-0,90	-0,37	0,11	0,01	-0,02
Tourneur	-0,95	-0,24	-0,02	-0,01	0,11
Cuisinier chef	-0,87	0,24	0,40	0,05	0,06
Chef de service	-0,84	0,49	-0,01	0,04	0,12
Ingenieur	-0,90	0,27	-0,03	-0,30	0,09
Caissier de banque	-0,88	0,38	-0,13	0,11	-0,20
Secrétaire direction	-0,97	0,00	-0,10	-0,03	-0,11
Vendeuse	-0,96	0,01	0,08	0,00	-0,17
Ouvrière du textile	-0,94	-0,25	-0,10	-0,17	-0,06

Interprétation

Comme il s'agit d'une analyse normée, les corrélations coïncident avec les coordonnées. On peut donc se référer à l'interprétation des coordonnées.

8.3 Anciens axes unitaires

Les anciens axes unitaires sont les coefficients de la liaison linéaire entre les variables et les axes.

Anciens axes unitaires

Libellé de la variable	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5
Instituteur	-0,30	-0,05	-0,37	0,39	0,34
Chauffeur d'autobus	-0,30	-0,14	-0,27	0,19	0,25
Mécanicien autos	-0,29	-0,29	0,34	0,16	-0,07
Manoeuvre du bâtiment	-0,28	-0,40	0,19	0,01	-0,04
Tourneur	-0,30	-0,26	-0,04	-0,03	0,28
Cuisinier chef	-0,27	0,25	0,71	0,12	0,16
Chef de service	-0,26	0,53	-0,02	0,11	0,30
Ingenieur	-0,28	0,30	-0,05	-0,72	0,22
Caissier de banque	-0,28	0,41	-0,23	0,26	-0,53
Secrétaire direction	-0,31	0,00	-0,17	-0,08	-0,29
Vendeuse	-0,30	0,02	0,14	0,00	-0,44
Ouvrière du textile	-0,29	-0,27	-0,18	-0,40	-0,16

On peut ainsi dire que :

$$\text{Axe 1} = -0,30 \left(\frac{\text{Instituteur} - \text{Moy}(\text{Instituteur})}{ET(\text{Instituteur})} \right) - 0,30 \left(\frac{\text{Chauffeur d'autobus} - \text{Moy}(\text{Chauffeur d'autobus})}{ET(\text{Chauffeur d'autobus})} \right) - \dots$$

8.4 Note sur les Cosinus Carrés et les Contributions

La sortie COPRI-6 ne donne pas les cosinus carrés et les contributions des variables continues actives sur les 5 premiers axes. Cependant, à partir des résultats précédents, on peut aisément les obtenir.

Comme pour les points-lignes, on peut donc définir les cosinus carrés et les contributions pour les variables.

Les cosinus carrés se définissent comme le quotient entre la distance projetée sur un axe et la distance à l'origine (les deux au carré). On sait que la distance d'une variable à l'origine est égale à sa variance :

$$COS^2(j, \alpha) = \frac{\varphi_{j\alpha}^2}{\text{var}(j)} \text{ avec } \varphi_{j\alpha} \text{ la coordonnée de la variable } j \text{ sur l'axe } \alpha.$$

La somme des cosinus carrés pour tous les axes est toujours égale à l'unité :

$$\sum_{\alpha=1}^p COS^2(j, \alpha) = 1$$

En ACP normée les variances sont égales à l'unité. Les cosinus carrés coïncideront avec le carré des coordonnées des variables :

$$COS^2(j, \alpha) = \varphi_{j\alpha}^2 \text{ en ACP normée}$$

Et dans tous le cas:

$$COS^2(j, \alpha) = \text{CORR}^2(\text{VARIABLE}, \text{FACTEUR})$$

La contribution d'une variable à l'inertie d'un axe est la part de l'inertie de l'axe due à la variable. L'inertie sur un axe s'écrit :

$$\lambda_{\alpha} = \sum_{j=1}^p \varphi_{j\alpha}^2$$

La contribution d'une variable à la construction de l'axe est:

$$CTR(j, \alpha) = \frac{\varphi_{j\alpha}^2}{\lambda_{\alpha}} = \frac{(\sqrt{\lambda_{\alpha}} \frac{u_{j\alpha}}{\lambda_{\alpha}})^2}{\lambda_{\alpha}} = u_{j\alpha}^2$$

Or $u_{j\alpha}$ est la coordonnée de l'ancien axe unitaire porteur de la variable j sur l'axe factoriel α . On a donc le résultat suivant:

$$CTR(j, \alpha) = (\text{ANCIEN AXE UNIT.})^2$$

Pour connaître la contribution d'une variable à la formation d'un axe, il suffit d'élever au carré chaque composante du vecteur \mathbf{u} . Ces contributions indiquent quelles variables sont responsables de la formation des axes. L'addition de toutes les contributions sur un axe est égale à 1 (ou 100 en pourcentage).

$$\sum_{j=1}^p CTR(j, \alpha) = 100$$

Les composantes de \mathbf{u} donnent également les combinaisons linéaires des variables d'origine qui définissent les nouvelles variables de variance maximale. Par exemple, l'expression de la première composante sera:

$$\begin{aligned} \Psi_1 = & -0.30z_{INST} - 0.30z_{CHAU} - 0.29z_{MECA} - 0.28z_{MANQ} - 0.30z_{OUTI} \\ & - 0.27z_{CUI}S - 0.26z_{CHEF} - 0.28z_{INGE} - 0.28z_{CAIS} - 0.31z_{SECR} \\ & - 0.30z_{VEND} - 0.29z_{OUVR} \end{aligned}$$

Où z indique la valeur de la variable centrée et réduite (car nous sommes en ACP normée).

La première composante est définie ici par des coefficients similaires entre eux, et définit donc une variable proche de la moyenne des salaires.

Ces composantes $u_{j\alpha}$ définissent aussi la projection des anciens axes unitaires sur les nouveaux axes obtenus.

9. COPRI-7 : Coordonnées, Corrélations des variables illustratives avec les 5 premiers axes

Cette sortie présente les coordonnées factorielles des variables continues illustratives ainsi que leur corrélation avec les 5 premiers axes de l'analyse. Du fait de l'analyse normée, les corrélations coïncident avec les coordonnées. Nous ne présenterons donc dans le tableau suivant que les corrélations des variables illustratives avec les 5 premiers axes.

Les variables continues illustratives sont positionnées en utilisant les mêmes formules que pour les variables actives. Dans le cas de l'ACP normée, on utilise la corrélation avec les composantes principales.

Dans le cas d'une ACP non normée, il suffit de multiplier la corrélation par l'écart-type de la variable.

La position des variables par rapport aux axes factoriels s'interprète comme pour les variables actives.

La position d'une variable illustrative dans le plan factoriel permet de visualiser la relation de la variable avec l'ensemble des variables actives par l'intermédiaire des axes factoriels.

On notera cependant qu'aucune distance n'a été définie entre deux variables illustratives. Leurs positions relatives ne présupposent aucune corrélation entre ces variables. Cependant, dans la mesure où les variables illustratives sont bien représentées sur le premier plan factoriel et proches l'une de l'autre, on peut attendre que la similitude de leurs corrélations avec les axes (similitude de leurs coordonnées) soit la conséquence d'une forte corrélation entre elles.

Corrélations des variables illustratives avec les facteurs

Libellé de la variable	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5
I_prix sans loyer	-0,68	0,05	-0,02	-0,09	-0,27
I_prix avec loyer	-0,68	0,10	-0,04	-0,15	-0,17
I_salaires bruts	-0,95	-0,19	-0,04	-0,04	-0,07
I_salaires nets	-0,98	-0,13	-0,02	-0,01	-0,02
Heures travail annuelles	0,33	0,30	0,14	0,10	0,16
Vacances annuelles payées	0,03	0,27	-0,24	0,13	-0,08
Pouvoir d'achat brut	-0,87	-0,26	0,01	-0,08	0,05
Pouvoir d'achat net	-0,90	-0,20	0,03	-0,05	0,11
Kg pain=temps de travail	0,55	-0,03	-0,12	0,07	-0,03
Hamb=temps de travail	0,52	-0,03	-0,12	0,15	-0,01
Denrées alimentaires	-0,58	0,06	-0,01	0,04	-0,25
Panier complet	-0,69	0,05	-0,02	-0,10	-0,27
Dames vêtements	-0,50	0,24	0,12	-0,11	-0,19
Hommes vêtements	-0,54	0,30	0,28	-0,18	-0,10
4pièces appart meublé	-0,14	0,24	-0,07	-0,21	0,24
3pièces appart non meublé	-0,01	0,11	-0,19	-0,17	0,18
Loyer normale	-0,62	0,14	-0,05	-0,09	0,14
Appareils ménagers	0,11	0,09	-0,17	0,10	-0,45
Bus tram ou metro	-0,69	-0,21	-0,01	-0,21	-0,19
Taxi	-0,72	-0,14	-0,20	-0,01	-0,23
Voitures	0,02	-0,05	-0,18	-0,23	0,04
Restaurant	-0,49	0,01	0,02	-0,25	0,03
Nuit d'hôtel	-0,42	0,19	-0,24	-0,13	-0,31
Services diverses	-0,78	0,00	-0,02	-0,21	-0,23
Impôts et cotisations sociales en % salaire brut	-0,41	-0,44	-0,09	-0,18	-0,25
Salaire horaire net	-0,98	-0,14	-0,02	-0,01	-0,03

Interprétation

Sur le premier axe factoriel, même si les corrélations avec les variables "indice des prix" sont relativement plus faibles que les corrélations avec les "indices de salaire", ce sont ces mêmes villes, les plus chères, qui ont le meilleur pouvoir d'achat.

10. COPRI-8 : Matrice des corrélations permutée suivant le premier facteur

On édite dans cette sortie la matrice des corrélations entre les variables actives ordonnées selon leur coordonnée sur le premier axe factoriel de l'analyse.

Matrice des corrélations permutée suivant le premier facteur

	Secrétaire direction	Vendeuse	Chauffeur d'autobus	Tourneur	Instituteur	Ouvrière du textile	Mécanicien autos	Ingenieur	Manoeuvre du bâtiment	Caissier de banque	Cuisinier chef	Chef de service
Secrétaire direction	1,00											
Vendeuse	0,94	1,00										
Chauffeur d'autobus	0,93	0,89	1,00									
Tourneur	0,92	0,88	0,94	1,00								
Instituteur	0,92	0,88	0,96	0,91	1,00							
Ouvrière du textile	0,93	0,89	0,92	0,94	0,88	1,00						
Mécanicien autos	0,88	0,89	0,89	0,93	0,84	0,89	1,00					
Ingenieur	0,87	0,85	0,82	0,80	0,81	0,81	0,74	1,00				
Manoeuvre du	0,86	0,86	0,88	0,93	0,83	0,92	0,95	0,70	1,00			
Caissier de banque	0,87	0,85	0,80	0,72	0,82	0,73	0,70	0,85	0,64	1,00		
Cuisinier chef	0,80	0,85	0,76	0,76	0,75	0,71	0,80	0,82	0,72	0,79	1,00	
Chef de service	0,80	0,79	0,74	0,69	0,78	0,65	0,64	0,87	0,59	0,89	0,82	1,00

11. COPRI-9 : Coordonnées, Contributions, Cosinus Carrés des individus actifs

11.1 Coordonnées des individus actifs

Le tableau ci-après donne les coordonnées des individus actifs. Par défaut, le logiciel édite ces valeurs sur les 5 premiers axes.

Ce tableau se compose de 8 colonnes :

- **Identificateur** : identifiant de l'individu
- **Poids relatif** : indique le poids donné à chaque ville, dans ce cas uniforme ($1.96=100/51$).
- **Distance à l'origine** : carré de la distance de la ville au centre de gravité de toutes les villes. Cette colonne permet de trouver facilement quelles sont les villes les plus "moyennes" (les plus proches du centre de gravité), comme Helsinki, et celles qui sont les plus "originales" (celles qui sont à plus grande distance du centre de gravité), comme Zurich ou Tokyo. La distance au centre est en quelque sorte un critère "d'originalité" de l'élément.
- **Axe 1 – Axe 5** : coordonnées obtenues par projection des villes sur les axes factoriels. La représentation sur le premier plan factoriel est obtenue à partir des coordonnées sur les deux premiers facteurs (axe1 et axe2).

Coordonnées des individus actifs

Identificateur	Poids relatif	Distance à l'origine	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5
AbuDhabi94	1,96	27,16	-2,17	4,61	0,75	-0,42	0,18
Amsterdam94	1,96	3,22	-1,58	-0,36	-0,58	-0,22	0,12
Athènes94	1,96	4,24	1,91	-0,51	-0,08	0,17	-0,04
Bangkok94	1,96	9,87	2,97	0,82	-0,22	-0,10	0,25
Bogor94	1,96	7,14	2,47	0,66	0,14	-0,31	0,26
Bombay94	1,96	21,11	4,56	-0,31	-0,26	0,28	0,02
Bruxelles94	1,96	0,74	-0,61	-0,17	0,25	-0,12	-0,26
Budapest94	1,96	17,74	4,19	-0,24	-0,16	0,28	0,01
BuenosAires94	1,96	5,39	0,89	1,23	-0,31	-0,20	-1,59
Caracas94	1,96	18,07	4,24	-0,06	-0,01	0,04	0,07
Chicago94	1,96	23,64	-4,42	-1,25	0,41	0,61	1,08
Copenhague94	1,96	7,43	-2,37	-0,83	0,74	-0,42	-0,49
Dublin94	1,96	0,79	0,27	-0,19	-0,63	-0,22	-0,10
Dusseldorf94	1,96	8,32	-2,72	0,24	-0,64	0,08	0,21
Frankfurt94	1,96	10,12	-3,05	0,63	-0,25	-0,01	0,15
Geneve94	1,96	42,20	-6,36	-0,30	-0,75	0,76	-0,40
Helsinki94	1,96	0,49	-0,03	-0,51	-0,03	-0,05	-0,04
Hongkong94	1,96	3,61	1,03	0,54	-0,57	-0,73	0,49
Houston94	1,96	15,21	-3,45	-0,78	1,37	-0,23	0,04
Jakarta94	1,96	16,92	4,08	-0,20	-0,11	0,21	0,14
Johannesburg94	1,96	4,88	2,08	-0,02	-0,12	-0,20	0,12
Lagos94	1,96	23,54	4,81	-0,43	-0,34	0,25	-0,04
Lisboa94	1,96	5,00	2,17	-0,28	-0,14	0,09	-0,21
London94	1,96	0,76	0,02	-0,63	-0,19	-0,21	0,18
LosAngeles94	1,96	18,89	-3,64	-1,80	-0,34	-1,09	0,33
Luxembourg94	1,96	32,79	-5,24	0,69	-1,91	-0,09	0,78
Madrid94	1,96	0,89	0,06	0,00	0,32	0,09	-0,27
Manama94	1,96	7,17	0,82	2,05	0,70	0,81	0,62
Manila94	1,96	16,51	4,05	-0,03	-0,10	0,19	0,12
Mexico94	1,96	8,63	2,83	-0,07	0,12	-0,32	0,25
Milan94	1,96	0,69	-0,02	-0,34	0,22	-0,14	-0,37
Montreal94	1,96	5,68	-2,17	-0,77	0,40	-0,09	-0,20
Nairobi94	1,96	23,45	4,82	-0,26	-0,20	0,25	0,03
NewYork94	1,96	23,01	-4,60	-0,30	0,99	-0,35	0,35
Nicosia94	1,96	3,56	1,78	-0,27	-0,27	0,26	0,10
Oslo94	1,96	3,98	-1,66	-0,73	0,02	-0,37	-0,38
Panama94	1,96	5,97	2,22	0,62	0,02	0,20	-0,33
Paris94	1,96	5,31	-1,65	1,41	0,07	-0,53	-0,05
Prague94	1,96	18,69	4,29	-0,31	-0,13	0,36	0,01
RioDeJaneiro94	1,96	12,22	3,40	0,28	-0,41	0,02	-0,05
SaoPaulo94	1,96	10,33	3,18	-0,01	-0,13	0,04	0,11
Seoul94	1,96	0,69	0,61	-0,04	0,01	0,21	-0,32
Singapore94	1,96	2,64	1,16	-0,16	-0,32	-0,69	0,18
Stockholm94	1,96	2,16	-0,67	-0,98	0,13	-0,19	-0,51
Sidney94	1,96	0,62	-0,03	-0,21	0,32	0,20	-0,33
Taipei94	1,96	6,07	-1,64	-0,27	1,36	1,02	0,10
Tel-Aviv94	1,96	3,35	1,41	0,00	0,98	0,15	0,25
Tokyo94	1,96	46,73	-6,72	0,72	0,13	0,05	-0,47
Toronto94	1,96	4,86	-1,77	-1,06	0,57	0,16	0,20
Vienna94	1,96	4,07	-1,86	-0,11	0,36	-0,61	0,02
Zurich94	1,96	65,45	-7,90	0,29	-1,18	1,12	-0,33

Interprétation

Le premier axe oppose les villes Zurich, Tokyo, Genève et Luxembourg d'une part aux villes Manila, Jakarta, Budapest, Caracas, Prague, Bombay, Lagos et Nairobi. Les villes dont les coordonnées sont proches de zéro pour l'axe 1 sont proches du centre de gravité et peuvent donc être considérées comme des villes moyennes de cet axe d'information.

Remarque : le sens d'un axe factoriel est arbitraire : seule la direction compte. On change le sens d'un axe en inversant les signes de toutes les coordonnées. Graphiquement, cela signifie que toutes les symétries sont autorisées : l'utilisateur choisit la représentation qui lui convient le mieux.

11.2 Contributions des individus actifs

Le tableau ci-après donne les contributions des individus actifs à la création des cinq premiers axes factoriels.

Ce tableau se compose de 8 colonnes :

- **Identificateur** : identifiant de l'individu
- **Poids relatif** : indique le poids donné à chaque ville, dans ce cas uniforme (1.96=100/51).
- **Distance à l'origine** : carré de la distance de la ville au centre de gravité de toutes les villes. Cette colonne permet de trouver facilement quelles sont les villes les plus "moyennes" (les plus proches du centre de gravité), comme Helsinki, et celles qui sont les plus "originales" (celles qui sont à plus grande distance du centre de gravité), comme Zurich ou Tokyo. La distance au centre est en quelque sorte un critère "d'originalité" de l'élément.

- **Axe 1 – Axe 5** : contributions (en pourcentage) des villes à l'inertie portée par chaque axe. Nous pouvons mesurer la part d'inertie d'un point-ligne par le quotient :

$$CTR(i, \alpha) = \frac{p_i \psi_{i\alpha}^2}{\lambda_\alpha} \times 100$$

Il s'agit de la contribution du point i à la fabrication de l'axe α avec : P_i le poids de i (1/51), $\Psi_{i\alpha}$ la coordonnées de l'individu i sur l'axe α et λ_α la valeur propre de l'axe α .

Identificateur	Poids relatif	Distance à l'origine	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5
AbuDhabi94	1.96	27.16	0.91	48.42	3.41	2.06	0.43
Amsterdam94	1.96	3.22	0.48	0.29	2.02	0.56	0.20
Athènes94	1.96	4.24	0.71	0.58	0.03	0.35	0.03
Bangkok94	1.96	9.87	1.71	1.53	0.29	0.11	0.84
Bogota94	1.96	7.14	1.18	0.99	0.11	1.11	0.88
Bombay94	1.96	21.11	4.03	0.22	0.39	0.91	0.00
Bruxelles94	1.96	0.74	0.07	0.07	0.38	0.16	0.88
Budapest94	1.96	17.74	3.39	0.13	0.16	0.89	0.00
BuenosAires94	1.96	5.39	0.15	3.43	0.58	0.45	33.49
Cairnes94	1.96	18.07	3.48	0.01	0.00	0.02	0.07
Chicago94	1.96	23.64	3.77	3.55	1.03	4.23	15.44
Copenhague94	1.96	7.43	1.09	1.58	3.27	2.03	3.11
Dublin94	1.96	0.79	0.01	0.08	2.39	0.56	0.14
Düsseldorf94	1.96	8.32	1.43	0.13	2.47	0.08	0.57
Frankfurt94	1.96	10.12	1.80	0.90	0.37	0.00	0.31
Genève94	1.96	42.20	7.82	0.21	3.38	6.55	2.13
Helsinki94	1.96	0.49	0.00	0.60	0.01	0.02	0.02
Hongkong94	1.96	3.61	0.21	0.66	1.98	6.05	3.22
Houston94	1.96	15.21	2.30	1.37	11.29	0.60	0.02
Jakarta94	1.96	16.92	3.22	0.09	0.07	0.51	0.28
Johannesburg94	1.96	4.88	0.84	0.00	0.09	0.44	0.18
Lagos94	1.96	23.54	4.47	0.42	0.70	0.72	0.02
Las Vegas94	1.96	5.00	0.91	0.18	0.12	0.09	0.57
London94	1.96	0.76	0.00	0.91	0.21	0.50	0.44
Los Angeles94	1.96	18.89	2.57	7.39	0.72	13.71	1.43
Luxembourg94	1.96	32.79	5.30	1.07	22.14	0.09	8.03
Madrid94	1.96	0.89	0.00	0.00	0.61	0.10	0.96
Malmö94	1.96	7.17	0.13	9.60	2.92	7.47	5.08
Mamilla94	1.96	16.51	3.17	0.00	0.06	0.43	0.18
Mexico94	1.96	8.63	1.55	0.01	0.09	1.20	0.81
Milan94	1.96	0.69	0.00	0.26	0.29	0.22	1.85
Montreal94	1.96	5.68	0.91	1.34	0.97	0.09	0.52
Nairobi94	1.96	23.45	4.50	0.15	0.23	0.72	0.01
New York94	1.96	23.01	4.09	0.20	5.96	1.43	1.66
Nicosia94	1.96	3.56	0.61	0.16	0.44	0.80	0.14
Oslo94	1.96	3.98	0.53	1.21	0.00	1.57	1.89
Panama94	1.96	5.97	0.96	0.88	0.00	0.44	1.45
Paris94	1.96	5.31	0.53	4.50	0.03	3.26	0.03
Prague94	1.96	18.69	3.56	0.22	0.11	1.49	0.00
RioDeJaneiro94	1.96	12.22	2.24	0.18	1.00	0.00	0.03
Sao Paulo94	1.96	10.33	1.96	0.00	0.11	0.02	0.17
Seoul94	1.96	0.69	0.07	0.00	0.00	0.53	1.35
Singapore94	1.96	2.64	0.26	0.06	0.63	5.45	0.41
Stockholm94	1.96	2.16	0.09	2.20	0.10	0.43	3.49
Sydney94	1.96	0.62	0.00	0.10	0.64	0.48	1.40
Taipei94	1.96	6.07	0.52	0.17	11.16	11.93	0.13
Tel-Aviv94	1.96	3.35	0.38	0.00	5.76	0.27	0.84
Tokyo94	1.96	46.73	8.72	1.18	0.11	0.03	2.89
Toronto94	1.96	4.86	0.60	2.53	1.95	0.28	0.52
Vienna94	1.96	4.07	0.67	0.03	0.80	4.29	0.00
Zurich94	1.96	65.45	12.08	0.20	8.44	14.28	1.44

Interprétation

Les contributions des points-lignes servent à détecter les villes qui contribuent le plus à la formation des axes factoriels.

Si toutes les villes avaient la même contribution, celle-ci devrait osciller autour de 2% (100/51). Les villes avec des contributions supérieures à 2% ont une influence supérieure à la moyenne.

Quand la valeur d'une contribution est-elle "élevée"?

La réponse n'est pas simple. Une contribution sera jugée élevée quand, en la comparant au reste des contributions, elle aura une valeur "anormalement" grande.

Par exemple, la ville qui contribue le plus au second axe est Abu Dhabi (48%). Presque la moitié de l'inertie de cet axe est due à une seule ville. Cette ville est clairement influente pour la fabrication de cet axe.

Tous les points actifs interviennent dans la fabrication d'un axe. On vérifie donc que l'addition de toutes les contributions sur un axe est 100.

$$\sum_{i=1}^n CTR(i, \alpha) = 100$$

11.3 Cosinus Carrés des individus actifs

Le tableau ci-après donne les cosinus carrés des individus actifs pour les cinq premiers axes factoriels.

Ce tableau se compose de 8 colonnes :

- **Identificateur** : identifiant de l'individu
- **Poids relatif** : indique le poids donné à chaque ville, dans ce cas uniforme (1.96=100/51).
- **Distance à l'origine** : carré de la distance de la ville au centre de gravité de toutes les villes. Cette colonne permet de trouver facilement quelles sont les villes les plus "moyennes" (les plus proches du centre de gravité), comme Helsinki, et celles qui sont les plus "originales" (celles qui sont à plus grande distance du centre de gravité), comme Zurich ou Tokyo. La distance au centre est en quelque sorte un critère "d'originalité" de l'élément.
- **Axe 1 – Axe 5** : cosinus carrés, utilisables pour apprécier la qualité du positionnement des points en représentation factorielle comparé à leur configuration réelle. Un cosinus carré égal à 1 indique que la ville se trouve située sur l'axe. Un cosinus carré égal à 0 indique que la ville est dans une direction orthogonale à l'axe.

Identificateur	Poids relatif	Distance à l'origine	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5
Abidjan94	1.96	27.16	0.17	0.78	0.02	0.01	0.00
Amsterdam94	1.96	3.27	0.77	0.04	0.10	0.02	0.00
Athènes94	1.96	4.24	0.86	0.06	0.00	0.01	0.00
Bangkok94	1.96	9.87	0.89	0.07	0.00	0.00	0.01
Bogotá94	1.96	7.14	0.86	0.06	0.00	0.01	0.01
Bombay94	1.96	21.11	0.99	0.00	0.00	0.00	0.00
Bruxelles94	1.96	0.74	0.50	0.04	0.08	0.02	0.09
Budapest94	1.96	17.74	0.99	0.00	0.00	0.00	0.00
BuenosAires94	1.96	5.39	0.15	0.28	0.02	0.01	0.47
Calcutta94	1.96	18.07	1.00	0.00	0.00	0.00	0.00
Chicago94	1.96	23.64	0.82	0.07	0.01	0.02	0.05
Copenhague94	1.96	7.43	0.76	0.09	0.07	0.02	0.03
Dublin94	1.96	0.79	0.10	0.05	0.50	0.06	0.01
Düsseldorf94	1.96	8.37	0.89	0.01	0.05	0.00	0.01
Frankfurt94	1.96	10.12	0.92	0.04	0.01	0.00	0.00
Geneve94	1.96	42.20	0.96	0.00	0.01	0.01	0.00
Helsinki94	1.96	0.49	0.00	0.53	0.00	0.00	0.00
Hongkong94	1.96	3.61	0.20	0.08	0.09	0.15	0.07
Houston94	1.96	15.21	0.78	0.04	0.12	0.00	0.00
Jakarta94	1.96	16.92	0.98	0.00	0.00	0.00	0.00
Johannesburg94	1.96	4.88	0.88	0.00	0.00	0.01	0.00
Kobe94	1.96	27.54	0.98	0.01	0.00	0.00	0.00
Ljubljana94	1.96	5.00	0.94	0.02	0.00	0.00	0.01
London94	1.96	0.76	0.00	0.52	0.05	0.06	0.04
LosAngeles94	1.96	18.89	0.70	0.17	0.01	0.06	0.01
Lyons94	1.96	22.79	0.84	0.01	0.11	0.00	0.02
Madrid94	1.96	0.89	0.00	0.00	0.11	0.01	0.08
Mannheim94	1.96	7.17	0.09	0.59	0.07	0.09	0.05
Mumbai94	1.96	16.51	0.99	0.00	0.00	0.00	0.00
Mexico94	1.96	8.63	0.93	0.00	0.00	0.01	0.01
Milano94	1.96	0.69	0.00	0.17	0.07	0.03	0.20
Montreal94	1.96	5.68	0.83	0.10	0.03	0.00	0.01
Nairobi94	1.96	23.45	0.99	0.00	0.00	0.00	0.00
NewYork94	1.96	23.01	0.92	0.00	0.04	0.01	0.01
Norwich94	1.96	3.56	0.89	0.02	0.02	0.02	0.00
Osaka94	1.96	3.98	0.69	0.13	0.00	0.03	0.04
Parana94	1.96	5.97	0.83	0.06	0.00	0.01	0.02
Paris94	1.96	5.31	0.51	0.57	0.00	0.05	0.00
Prague94	1.96	18.69	0.98	0.01	0.00	0.01	0.00
Roskilde94	1.96	12.22	0.95	0.01	0.01	0.00	0.00
SaoPaulo94	1.96	10.33	0.98	0.00	0.00	0.00	0.00
Seoul94	1.96	0.69	0.54	0.00	0.00	0.07	0.15
Singapore94	1.96	2.64	0.51	0.01	0.04	0.18	0.01
Stockholm94	1.96	2.16	0.21	0.45	0.01	0.02	0.12
Sidney94	1.96	0.62	0.00	0.07	0.17	0.07	0.17
Taipei94	1.96	6.07	0.45	0.01	0.30	0.17	0.00
Tek-Ary94	1.96	8.35	0.59	0.00	0.29	0.01	0.02
Tokyo94	1.96	46.73	0.97	0.01	0.00	0.00	0.00
Toronto94	1.96	4.86	0.64	0.23	0.07	0.01	0.01
Vienne94	1.96	4.07	0.85	0.00	0.03	0.09	0.00
Zurich94	1.96	65.45	0.95	0.00	0.02	0.02	0.00

Interprétation

Les images obtenues (plans factoriels) sont des approximations de la configuration réelle. Il y aura des distances entre couples de points bien représentées, tandis que d'autres ne reflèteront pas fidèlement la distance réelle entre les points.

Si deux points sont proches du plan factoriel, alors la distance représentée sera une bonne approximation à la distance réelle. Mais si au moins un point est éloigné du plan de projection, alors la distance réelle peut être différente de celle représentée sur le plan.

Cette proximité du plan factoriel de projection est mesurée par les cosinus carrés de chaque point avec les axes factoriels.

$$\cos^2(i, \alpha) = \frac{\psi_{i\alpha}^2}{d^2(i, G)}$$

Un cosinus carré égal à 1 indique que la ville se trouve située sur l'axe (l'angle ω est nul). Un cosinus carré égal à 0 indique que la ville est dans une direction orthogonale à l'axe.

En additionnant les cosinus carrés par rapport aux p axes factoriels, nous obtiendrons l'unité, car il faut utiliser tous les axes factoriels pour positionner exactement le point dans l'espace complet.

$$\sum_{\alpha=1}^p \cos^2(i, \alpha) = 1$$

L'addition des cosinus carrés d'un point sur différents axes donne, en pourcentage, la "qualité" de la représentation du point sur le sous-espace défini par ces axes.

A partir de quelle valeur des cosinus carrés dira-t-on qu'un point est "bien représenté" sur le plan factoriel?

Comme pour les contributions, la question n'est pas simple. On appréciera une valeur par rapport à l'ensemble des autres valeurs. Nous devons comparer les cosinus carrés (ou leur somme sur les deux premiers axes si nous travaillons sur le premier plan factoriel) pour pouvoir répondre à cette question.

Dans notre exemple, les villes sont en général bien représentées dans les deux premiers axes (avec une somme des cosinus carrés proche de 1). Seules Dublin, Madrid, Sidney ou Milan, proches du centre, sont mal représentées sur le premier plan. Bombay ou Caracas et d'une façon générale les points excentrés, sont bien représentés

Les villes les moins bien représentées dans le premier plan factoriel sont des villes "moyennes" (proches du centre). On ne peut interpréter avec sécurité les proximités entre les villes que si elles sont bien représentées dans le plan factoriel.

12. COPRI-10 : Coordonnées, Cosinus Carrés des individus illustratifs

12.1 Coordonnées des individus illustratifs

Le tableau ci-après donne les coordonnées des individus actifs. Par défaut, le logiciel édite ces valeurs sur les 5 premiers axes.

Ce tableau se compose de 8 colonnes :

- **Identificateur** : identifiant de l'individu
- **Poids relatif** : indique le poids donné à chaque ville, dans ce cas uniforme (1.96=100/51).
- **Distance à l'origine** : carré de la distance de la ville au centre de gravité de toutes les villes. Cette colonne permet de trouver facilement quelles sont les villes les plus "moyennes" (les plus proches du centre de gravité), comme Helsinki, et celles qui sont les plus "originales" (celles qui sont à plus grande distance du centre de gravité), comme Zurich ou Tokyo. La distance au centre est en quelque sorte un critère "d'originalité" de l'élément.
- **Axe 1 – Axe 5** : coordonnées obtenues par projection des villes sur les axes factoriels. La représentation sur le premier plan factoriel est obtenue à partir des coordonnées sur les deux premiers facteurs (axe1 et axe2).

Le tableau suivant a été tronqué.

Coordonnées des individus illustratifs

Identificateur	Poids relatif	Distance à l'origine	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5
AbuDhabi91	1,96	0,00	0,00	0,00	0,00	0,00	0,00
Amsterdam91	1,96	0,59	-0,19	-0,46	-0,07	-0,22	-0,17
Athenes91	1,96	4,19	1,81	-0,55	0,41	0,09	0,30
Bangkok91	1,96	0,00	0,00	0,00	0,00	0,00	0,00
Bogota91	1,96	12,34	3,41	0,16	0,61	-0,09	0,34
Bombay91	1,96	20,98	4,55	-0,33	-0,24	0,24	0,02
Bruxelles91	1,96	0,41	-0,23	0,02	0,31	-0,13	-0,09
Budapest91	1,96	0,00	0,00	0,00	0,00	0,00	0,00
BuenosAires91	1,96	11,95	3,25	0,67	-0,15	0,03	0,32
Caracas91	1,96	14,47	3,71	0,05	0,61	0,34	0,18
Chicago91	1,96	7,58	-2,20	-1,04	0,45	-0,25	0,79
Copenhague91	1,96	3,63	-1,53	-0,77	0,45	-0,52	-0,37
Dublin91	1,96	0,75	0,48	-0,13	-0,50	-0,22	-0,05
Dusseldorf91	1,96	4,55	-1,76	0,37	-0,16	-0,55	0,71
Frankfurt91	1,96	3,63	-1,34	0,20	-0,76	-0,07	0,51
Geneve91	1,96	30,75	-5,36	-0,31	0,04	0,39	0,35
Helsinki91	1,96	3,11	-1,52	-0,80	0,23	-0,05	-0,07
Hongkong91	1,96	1,53	1,01	-0,14	-0,05	0,32	0,22
Houston91	1,96	1,17	0,25	-0,71	0,48	-0,12	0,15
Jakarta91	1,96	18,96	4,32	0,21	0,09	0,30	0,18
Johannesburg91	1,96	5,41	2,19	-0,17	0,01	0,10	0,21

Interprétation

L'interprétation des coordonnées des individus illustratifs se fait de la même manière que pour les individus actifs.

Dans le tableau ci-dessus, on constate que les villes Abudhabi, Bangkok et Budapest présentent une distance nulle à l'origine.

En effet, ces villes n'étaient pas renseignées pour les variables actives sélectionnées. SPAD remplace automatiquement toute donnée manquante par la moyenne de la variable correspondante. Ces villes se sont donc vues affecter la valeur moyenne pour chaque variable active de l'analyse. Il est donc normal qu'elles soient confondues avec le centre de gravité du nuage.

12.2 Cosinus Carrés des individus illustratifs

Le tableau ci-après donne les cosinus carrés des individus actifs pour les cinq premiers axes factoriels.

Ce tableau se compose de 8 colonnes :

- **Identificateur** : identifiant de l'individu
- **Poids relatif** : indique le poids donné à chaque ville, dans ce cas uniforme ($1,96=100/51$).
- **Distance à l'origine** : carré de la distance de la ville au centre de gravité de toutes les villes. Cette colonne permet de trouver facilement quelles sont les villes les plus "moyennes" (les plus proches du centre de gravité), comme Helsinki, et celles qui sont les plus "originales" (celles qui sont à plus grande distance du centre de gravité), comme Zurich ou Tokyo. La distance au centre est en quelque sorte un critère "d'originalité" de l'élément.
- **Axe 1 – Axe 5** : cosinus carrés, utilisables pour apprécier la qualité du positionnement des points en représentation factorielle comparé à leur configuration réelle. Un cosinus carré égal à 1 indique que la ville se trouve située sur l'axe. Un cosinus carré égal à 0 indique que la ville est dans une direction orthogonale à l'axe.

Le tableau suivant a été tronqué.

Cosinus carrés des individus illustratifs

Identificateur	Poids relatif	Distance à l'origine	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5
AbuDhabi91	1,96	0,00	0,00	0,00	0,00	0,00	0,00
Amsterdam91	1,96	0,59	0,06	0,35	0,01	0,08	0,05
Athenes91	1,96	4,19	0,79	0,07	0,04	0,00	0,02
Bangkok91	1,96	0,00	0,00	0,00	0,00	0,00	0,00
Bogota91	1,96	12,34	0,94	0,00	0,03	0,00	0,01
Bombay91	1,96	20,98	0,99	0,01	0,00	0,00	0,00
Bruxelles91	1,96	0,41	0,12	0,00	0,23	0,04	0,02
Budapest91	1,96	0,00	0,00	0,00	0,00	0,00	0,00
BuenosAires91	1,96	11,95	0,88	0,04	0,00	0,00	0,01
Caracas91	1,96	14,47	0,95	0,00	0,03	0,01	0,00
Chicago91	1,96	7,58	0,64	0,14	0,03	0,01	0,08
Copenhague91	1,96	3,63	0,65	0,16	0,06	0,07	0,04
Dublin91	1,96	0,75	0,31	0,02	0,33	0,07	0,00
Dusseldorf91	1,96	4,55	0,68	0,03	0,01	0,07	0,11
Frankfurt91	1,96	3,63	0,49	0,01	0,16	0,00	0,07
Geneve91	1,96	30,75	0,94	0,00	0,00	0,00	0,00
Helsinki91	1,96	3,11	0,74	0,20	0,02	0,00	0,00
Hongkong91	1,96	1,53	0,67	0,01	0,00	0,06	0,03
Houston91	1,96	1,17	0,05	0,43	0,19	0,01	0,02
Jakarta91	1,96	18,96	0,99	0,00	0,00	0,00	0,00
Johannesburg91	1,96	5,41	0,88	0,01	0,00	0,00	0,01

Interprétation

On procède de la même manière que pour les individus actifs.

Nous sommes en présence d'un cas particulier avec les villes confondues avec l'origine : Abudhabi, Bangkok et Budapest. On ne peut pas calculer leurs cosinus carrés puisque leur distance à l'origine est nulle tout comme leurs coordonnées (par définition). Or le cosinus carré s'obtient par le quotient de la coordonnée au carré sur la distance à l'origine au carré.

13. COPRI-11 : Coordonnées des modalités illustratives

Le tableau ci-après donne les coordonnées des modalités illustratives. Par défaut, le logiciel édite ces valeurs sur les 5 premiers axes.

Ce tableau se compose de 8 colonnes :

- **Libellé** : libellé complet de la modalité.
- **Effectif** : effectif de la modalité.
- **Poids absolu** : poids de chaque modalité. Le poids des individus est ici uniforme et égal à 1.
- **Distance à l'origine** : carré de la distance de la modalité au centre de gravité de toutes les villes. Cette colonne permet de trouver facilement quelles sont les modalités les plus "moyennes" (les plus proches du centre de gravité), comme l'Europe du Nord ou l'Europe du Sud, et celles qui sont les plus "originales" (celles qui sont à plus grande distance du centre de gravité), comme l'Europe Centrale, l'Afrique et l'Europe Orientale. La distance au centre est en quelque sorte un critère "d'originalité" de la modalité.
- **Axe 1 – Axe 5** : coordonnées des modalités illustratives sur les 5 premiers axes.
Une modalité illustrative est positionnée comme point moyen des individus actifs qui la composent. Ainsi, sur l'axe 1, la coordonnée de la modalité Europe du Nord correspond à la moyenne pondérée des coordonnées des 6 villes qu'elle englobe : Dublin, Stockholm, London, Copenhague, Oslo, Helsinki. Dans le cas présent, le poids des ville est uniforme.

Coordonnées des modalités illustratives

Libellé	Effectif	Poids absolu	Distance à l'origine	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5
Région du monde								
NORD EUROPE	6	6,00	1,17305	-0,74	-0,65	0,01	-0,24	-0,22
CENTRAL EUROPE	9	9,00	12,19330	-3,44	0,26	-0,51	0,04	0,03
SUD EUROPE	5	5,00	1,61787	1,18	-0,28	0,01	0,10	-0,16
AFRIQUE	3	3,00	15,41930	3,90	-0,24	-0,22	0,10	0,04
EST ASIE	5	5,00	1,43674	-1,11	0,16	0,12	-0,03	0,00
SUD ASIE ET AUSTRALI	5	5,00	9,83115	3,13	0,01	-0,07	0,16	0,04
NORD AMERIQUE	7	7,00	7,19935	-2,46	-0,86	0,50	-0,19	0,29
SUD AMERIQUE	6	6,00	7,83858	2,74	0,45	-0,12	-0,04	-0,25
PROCHE ORIENT	3	3,00	5,76941	0,02	2,22	0,81	0,18	0,35
EUROPE ORIENTAL	2	2,00	18,20720	4,24	-0,28	-0,15	0,32	0,01

Interprétation

L'axe 1 oppose les régions Europe Orientale, Afrique, Asie du Sud et Australie, Amérique du Sud aux régions Europe Centrale et Amérique du Nord.

L'axe 2 oppose la région Proche Orient à toutes les autres.

14. COPRI-12 : Valeurs-tests des modalités illustratives

Le tableau ci-après donne les coordonnées des modalités illustratives. Par défaut, le logiciel édite ces valeurs sur les 5 premiers axes.

Ce tableau se compose de 8 colonnes :

- **Libellé** : libellé complet de la modalité.
- **Effectif** : effectif de la modalité.
- **Poids absolu** : poids de chaque modalité. Le poids des individus est ici uniforme et égal à 1.
- **Distance à l'origine** : carré de la distance de la modalité au centre de gravité de toutes les villes. Cette colonne permet de trouver facilement quelles sont les modalités les plus "moyennes" (les plus proches du centre de gravité), comme l'Europe du Nord ou l'Europe du Sud, et celles qui sont les plus "originales" (celles qui sont à plus grande distance du centre de gravité), comme l'Europe Centrale, l'Afrique et l'Europe Orientale. La distance au centre est en quelque sorte un critère "d'originalité" de la modalité.
- **Axe 1 – Axe 5** : valeur-test associée au positionnement de la modalité illustrative sur l'axe par rapport au centre de gravité du nuage.

Pour un axe donnée et une modalité, on s'intéresse à l'écart entre la moyenne des coordonnées des individus (de la modalité) et la moyenne générale (qui est nulle ici). Plus cet écart est "significatif", plus la modalité a une position intéressante dans le plan. Plus la v-test est grande (en valeur absolue), plus il est clair que le groupe d'individus occupe une position significative et caractérise la zone du graphique où il se trouve.

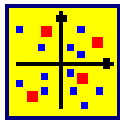
Au seuil 5%, on considère qu'un groupe d'individus défini par une modalité est significativement différent des autres individus lorsque sa valeur-test est supérieure à 2 en valeur absolue.

Valeurs-Tests des modalités illustratives

Libellé	Effectif	Poids absolu	Distance à l'origine	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5
Région du monde								
NORD EUROPE	6	6,00	1,17305	-0,60	-1,80	0,03	-1,52	-1,50
CENTRAL EUROPE	9	9,00	12,19330	-3,54	0,90	-2,95	0,33	0,23
SUD EUROPE	5	5,00	1,61787	0,86	-0,70	0,04	0,54	-0,96
AFRIQUE	3	3,00	15,41930	2,17	-0,45	-0,68	0,43	0,17
EST ASIE	5	5,00	1,43674	-0,81	0,40	0,50	-0,15	-0,02
SUD ASIE ET AUSTRALI	5	5,00	9,83115	2,29	0,04	-0,29	0,89	0,25
NORD AMERIQUE	7	7,00	7,19935	-2,18	-2,61	2,49	-1,29	2,15
SUD AMERIQUE	6	6,00	7,83858	2,22	1,26	-0,53	-0,23	-1,71
PROCHE ORIENT	3	3,00	5,76941	0,01	4,23	2,51	0,77	1,61
EUROPE ORIENTAL	2	2,00	18,20720	1,90	-0,43	-0,37	1,10	0,04

Interprétation

Sur l'axe 1, on constate que les régions Europe Centrale, Afrique, Asie du Sud et Australie, Amérique du Nord, Amérique du Sud occupent une position significativement différente de la moyenne.



Analyse des correspondances binaires (CORBI)

Cette procédure effectue l'analyse des correspondances d'un tableau de contingence ou, de façon plus générale, de tout tableau de nombres non négatifs.

Les éléments en colonnes seront toujours appelées "fréquences" et les éléments en lignes des "individus" ou des "lignes".

Fichier utilisé : ALCOOL.SBA

Paramétrage de la méthode : Tous les individus (lignes) et toutes les fréquences (colonnes) sont utilisés.

1. Présentation du tableau analysé

Nous effectuons l'analyse du tableau suivant auquel on a ajouté les marges. Il s'agit d'étudier la perception de différentes boissons alcoolisées.

	PASTIS	WHISKY	MARTINI	SUZE	VODKA	GIN	MALIBU	BIERE	Total
Aime le goût	49	50	42	18	25	23	25	59	291
Avec des amis	83	83	76	60	69	68	69	74	582
Pour se détendre	61	61	51	32	38	39	39	72	393
Qui revient cher	60	88	42	41	75	70	61	19	456
Rafraîchissante, désaltérante	78	22	18	19	17	19	14	80	267
Peu élégante, peu distinguée	26	11	13	17	13	11	13	29	133
Produit sympathique	64	64	56	34	45	42	46	68	419
Bien avant les repas	88	79	85	64	45	46	37	41	485
Bien dans la journée	24	21	12	10	13	12	13	85	190
Bien dans la soirée	7	61	12	11	53	50	48	54	296
Toute l'année	83	87	85	79	83	82	80	90	669
Apprécie des jeunes	45	77	36	16	65	69	76	89	473
Volontiers avec invités	88	92	87	60	70	67	67	81	612
Vieillesse, dépassée	12	4	13	38	5	6	8	7	93
Aussi bien hommes que femmes	50	62	69	43	49	51	61	60	445
Très proche	38	41	27	11	16	18	17	49	217
Par habitude	36	30	24	16	19	19	17	40	201
Fait snob, m'as-tu vu ?	3	35	9	8	28	25	21	4	133
On peut mélanger	43	87	29	32	82	80	43	40	436
La nuit/Bar/Disco	12	91	27	16	84	81	72	67	450
Total	950	1146	813	625	894	878	827	1108	7241

2. CORBI-1 : Tableau des valeurs-propres

Le tableau des valeurs propres est constitué de 4 colonnes.

- **Numero** : numéro de l'axe factoriel.
- **Valeur propre** : la valeur propre (inertie) associée à l'axe.
- **Pourcentage** : indique la part de la valeur propre (ou inertie) prise en compte sur l'axe factoriel.
- **Pourcentage cumulé** : pourcentage cumulé des parts de valeurs propres sur les n premiers axes.

Le nombre d'axes factoriels est égal au minimum entre le nombre de lignes et le nombre de colonnes moins 1, soit $\text{Min}\{20, 8\} - 1 = 7$.

La somme des valeurs propres représente l'inertie totale du nuage. Elle correspond à la statistique du khi-2. Appliquée à un tableau de contingence, cette statistique mesure l'écart

entre les effectifs observés et les effectifs théoriques que l'on obtiendrait en moyenne si les deux variables étaient indépendantes.

Le logiciel édite l'ensemble des valeurs propres, soit 7 valeurs propres.

Tableau des valeurs propres
Trace de la matrice: 0.13452

Numéro	Valeur propre	Pourcentage	Pourcentage cumulé
1	0,0664	49,37	49,37
2	0,0449	33,34	82,72
3	0,0124	9,24	91,96
4	0,0069	5,14	97,09
5	0,0029	2,18	99,27
6	0,0008	0,63	99,90
7	0,0001	0,10	100,00

Interprétation

Le premier axe factoriel extrait près de la moitié de l'inertie totale. La part d'information expliquée par le second axe est importante puisqu'elle représente un tiers de l'inertie totale. Les trois premiers axes factoriels résument plus de 90% de l'inertie totale du nuage. On pourra se servir des plans factoriels 1-2 et 2-3 pour la suite de l'analyse.

3. CORBI-2 : Coordonnées, Contributions, Cosinus Carrés des fréquences actives

3.1 Coordonnées des fréquences actives

Le tableau ci-après donne les coordonnées des fréquences actives. Par défaut, le logiciel édite ces valeurs sur les 5 premiers axes.

Ce tableau se compose de 8 colonnes :

- **Libellé** : libellé complet de la fréquence active
- **Poids relatif** : poids relatif de la fréquence active. Le poids relatif se calcule de la façon suivante : $(n_q \cdot 100) / n$ avec n_q l'effectif de la fréquence active et n l'effectif total. Par exemple pour la fréquence active Pastis, $P.REL = (950 \cdot 100) / 7241 = 13,12$.
- **Distance à l'origine** : carré de la distance de la fréquence active à l'origine. Cette distance donne une idée du caractère périphérique de certains points (comme la fréquence active « SUZE » pour laquelle $DISTO = 0,2961$).
- **Axe 1 – Axe 5** : coordonnées des fréquences actives sur les premiers axes.

Coordonnées des fréquences actives

Libellé de la variable	Poids relatif	Distance à l'origine	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5
PASTIS	13,12	0,17452	0,36	0,05	-0,16	-0,11	0,04
WHISKY	15,83	0,05211	-0,19	-0,02	-0,09	0,02	-0,09
MARTINI	11,23	0,11073	0,17	0,21	-0,09	0,17	0,00
SUZE	8,63	0,29610	0,22	0,43	0,24	-0,05	-0,04
VODKA	12,35	0,09581	-0,30	0,00	0,01	-0,06	0,00
GIN	12,13	0,08348	-0,28	0,00	0,01	-0,06	0,01
MALIBU	11,42	0,06846	-0,21	-0,02	0,06	0,07	0,11
BIERE	15,30	0,23277	0,26	-0,39	0,10	0,02	-0,02

Le premier axe qui explique près de la moitié de l'information oppose les alcools PASTIS, BIERE, SUZE, MARTINI aux alcools forts VODKA, GIN, MALIBU et WHISKY.

Le second axe qui explique un tiers de l'inertie totale oppose essentiellement la SUZE et le MARTINI à la BIERE.

3.2 Contributions des fréquences actives

Le tableau ci-après donne les contributions des fréquences actives. Par défaut, le logiciel édite ces valeurs sur les 5 premiers axes.

Ce tableau se compose de 8 colonnes :

- **Libellé** : libellé de la fréquence
- **Poids relatif** : poids relatif de la fréquence active. Le poids relatif se calcule de la façon suivante : $(n_q \cdot 100) / n$ avec n_q l'effectif de la fréquence active et n l'effectif total. Par exemple pour la fréquence active Pastis, $P.REL = (950 \cdot 100) / 7241 = 13,12$.
- **Distance à l'origine** : carré de la distance de la fréquence active à l'origine. Cette distance donne une idée du caractère périphérique de certains points (comme la fréquence active « SUZE » pour laquelle $DISTO = 0,2961$).
- **Axe 1 – Axe 5** : contributions (en pourcentage) des fréquences actives sur les 5 premiers axes. Les contributions décrivent la part de chaque fréquence active dans l'inertie totale de l'axe. Pour l'interprétation d'un axe, on sélectionne les modalités qui ont les plus fortes contributions.

Nous pouvons mesurer la part d'inertie d'une fréquence active par le quotient :

$$CTR(j, \alpha) = \frac{p_j \psi_{j\alpha}^2}{\lambda_\alpha} \times 100$$

Il s'agit de la contribution de la fréquence j à la fabrication de l'axe α avec : P_j le poids de j ($950/7241$), $\psi_{j\alpha}$ la coordonnées de la fréquence j sur l'axe α et λ_α la valeur propre de l'axe α .

La somme des contributions de toutes les fréquences actives sur un axe est égale à 100% (somme en colonne).

Contributions des fréquences actives

Libellé de la variable	Poids relatif	Distance à l'origine	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5
PASTIS	13,12	0,17452	26,26	0,61	26,53	23,53	8,31
WHISKY	15,83	0,05211	8,38	0,13	9,73	0,56	39,51
MARTINI	11,23	0,11073	4,87	10,53	7,24	49,75	0,03
SUZE	8,63	0,29610	6,30	35,59	40,71	3,23	3,86
VODKA	12,35	0,09581	16,75	0,00	0,03	7,21	0,01
GIN	12,13	0,08348	14,28	0,01	0,06	5,89	0,70
MALIBU	11,42	0,06846	7,94	0,06	3,02	8,71	45,87
BIERE	15,30	0,23277	15,21	53,07	12,69	1,12	1,72

Interprétation

Si toutes les fréquences avaient la même contribution, celle-ci devrait osciller autour de 14,28% ($100\% / 7$). Les fréquences avec des contributions supérieures à 14,28% ont une influence supérieure à la moyenne.

Les fréquences actives PASTIS, VODKA, GIN et BIERE contribuent fortement à la création de l'axe 1 (72,51% en tout).

Sur l'axe 2, la BIERE explique plus de la moitié de l'inertie de l'axe, la SUZE en explique plus du tiers. En y ajoutant le MARTINI, on obtient une contribution cumulée de plus de 99% avec seulement trois fréquences actives.

3.3 Cosinus carrés des fréquences actives

Le tableau ci-après donne les cosinus carrés des fréquences actives. Par défaut, le logiciel édite ces valeurs sur les 5 premiers axes.

Ce tableau se compose de 8 colonnes :

- **Libellé** : libellé complet de la modalité
- **Poids relatif** : poids relatif de la fréquence active. Le poids relatif se calcule de la façon suivante : $(n_q \cdot 100) / n$ avec n_q l'effectif de la fréquence active et n l'effectif total. Par exemple pour la fréquence active Pastis, $P.REL = (950 \cdot 100) / 7241 = 13,12$.
- **Distance à l'origine** : distance de la fréquence active à l'origine au carré. Cette distance donne une idée du caractère périphérique de certains points (comme la fréquence active « SUZE » pour laquelle $DISTO = 0,2961$).
- **Axe 1 – Axe 5** : cosinus carrés (ou contributions relatives) de chacune des fréquences actives sur les premiers axes.

Les cosinus carrés concernent la qualité de représentation des fréquences sur l'axe. Pour l'interprétation, on sélectionne les fréquences qui ont les cosinus carrés les plus forts : elles sont les mieux représentées au sens où les distances sont les moins altérées par la projection.

Si la fréquence j a pour coordonnées $\Psi_{j\alpha}$ sur l'axe α et si le carré de la distance à l'origine de cette fréquence est $d^2(j, G)$, le cosinus carré (ou contribution relative) $COS^2(j, \alpha)$ de la fréquence j sur l'axe α s'écrit :

$$COS^2(j, \alpha) = \frac{\Psi_{j\alpha}^2}{d^2(j, G)}$$

On a par exemple pour la fréquence PASTIS sur le premier facteur : $0,36^2 / 0,17 = 0,76$.

La somme des cosinus carrés d'une modalité sur la totalité des axes est égale à 1 (somme en ligne étendue à tous les axes).

Cosinus carrés des fréquences actives

Libellé de la variable	Poids relatif	Distance à l'origine	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5
PASTIS	13,12	0,17452	0,76	0,01	0,14	0,07	0,01
WHISKY	15,83	0,05211	0,67	0,01	0,15	0,00	0,14
MARTINI	11,23	0,11073	0,26	0,38	0,07	0,28	0,00
SUZE	8,63	0,29610	0,16	0,62	0,20	0,01	0,00
VODKA	12,35	0,09581	0,94	0,00	0,00	0,04	0,00
GIN	12,13	0,08348	0,94	0,00	0,00	0,04	0,00
MALIBU	11,42	0,06846	0,67	0,00	0,05	0,08	0,17
BIERE	15,30	0,23277	0,28	0,67	0,04	0,00	0,00

Interprétation

Les alcools VODKA, GIN sont tellement bien représentés sur le premier axe qu'ils ne faut plus les prendre en compte pour l'interprétation des autres axes.

On peut dire que les alcools PASTIS, WHISKY, MALIBU sont également bien représentés sur l'axe 1.

En revanche, les alcools MARTINI SUZE et BIERE sont mieux représentés sur l'axe 2. Ce sont les seuls à être bien représentés sur cet axe.

4. CORBI-4 : Coordonnées, Contributions, Cosinus Carrés des individus actifs

4.1 Coordonnées des individus actifs

On entend par individus actifs les lignes du tableau de contingence que l'on utilise pour l'analyse.

Le tableau ci-après donne les coordonnées des individus actifs. Par défaut, le logiciel édite ces valeurs sur les 5 premiers axes.

Ce tableau se compose de 8 colonnes :

- **Libellé** : libellé complet de l'individu actif
- **Poids relatif** : poids relatif de l'individu actif. Le poids relatif se calcule de la façon suivante : $(n_q \cdot 100) / n$ avec n_q l'effectif de l'individu actif et n l'effectif total. Par exemple pour l'individu actif « Aime le goût », $P.REL = (291 \cdot 100) / 7241 = 4,02$.
- **Distance à l'origine** : carré de la distance de l'individu actif à l'origine. Cette distance donne une idée du caractère périphérique de certains points (comme l'individu actif « Vieillotte, dépassée » pour laquelle $DISTO = 1,40646$).
- **Axe 1 – Axe 5** : coordonnées des individus actifs sur les premiers axes.

Coordonnées des individus actifs

Identificateur	Poids relatif	Distance à l'origine	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5
Aime le goût	4,02	0,07708	0,21	-0,10	-0,12	0,08	-0,06
Avec des amis	8,04	0,01368	0,04	0,10	0,00	0,01	0,04
Pour se détendre	5,43	0,02518	0,14	-0,04	-0,04	0,04	-0,02
Qui revient cher	6,30	0,11797	-0,25	0,19	-0,09	-0,11	0,03
Rafraîchissante, désaltérante	3,69	0,47791	0,56	-0,30	-0,07	-0,25	0,07
Peu élégante, peu distinguée	1,84	0,13704	0,32	-0,03	0,12	-0,11	0,08
Produit sympathique	5,79	0,01458	0,10	0,00	-0,05	0,04	0,01
Bien avant les repas	6,70	0,13546	0,18	0,30	-0,11	0,03	-0,06
Bien dans la journée	2,62	0,68617	0,43	-0,66	0,25	0,04	-0,11
Bien dans la soirée	4,09	0,24599	-0,40	-0,26	0,12	0,01	-0,03
Toute l'année	9,24	0,02156	0,02	0,11	0,08	0,01	0,03
Appréciée des jeunes	6,53	0,08716	-0,17	-0,22	0,02	0,03	0,09
Volontiers avec invités	8,45	0,01595	0,06	0,10	-0,03	0,04	0,01
Vieillotte, dépassée	1,28	1,40646	0,46	0,84	0,68	-0,11	-0,08
Aussi bien hommes que femmes	6,15	0,03105	0,01	0,09	0,02	0,14	0,06
Très proche	3,00	0,11469	0,22	-0,19	-0,13	0,05	-0,10
Par habitude	2,78	0,05316	0,21	-0,08	-0,06	-0,02	-0,03
Fait snob, m'as-tu vu ?	1,84	0,39872	-0,61	0,09	-0,03	-0,02	-0,09
On peut mélanger	6,02	0,13185	-0,31	0,03	-0,03	-0,16	-0,07
La nuit/Bar/Disco	6,21	0,22897	-0,44	-0,18	0,07	0,02	-0,01

Sur le premier axe factoriel, on remarque une opposition entre les appréciations au niveau des boissons alcoolisées :

- Bien avant les repas, Aime le goût, Par habitude, Très proche, Peu élégante, peu distinguée, Bien dans la journée, Vieillotte dépassée, Rafraîchissante, désaltérante.

S'oppose à :

- Fait snob, m'as-tu vu ?, La nuit/Bar/Disco, Bien dans la soirée, On peut mélanger, Qui revient cher, Appréciée des jeunes.

Au niveau des fréquences actives, le premier axe oppose les alcools PASTIS, BIERE, SUZE, MARTINI aux alcools VODKA, GIN, MALIBU et WHISKY.

Le second axe factoriel marque les oppositions suivantes :

- Vieillotte, dépassée, Bien avant les repas, Qui revient cher,

S'oppose à :

- Bien dans la journée, Rafraîchissante, désaltérante, Bien dans la soirée, Appréciée des jeunes, Très proche, La nuit/Bar/Disco,

Au niveau des fréquences actives, le second axe qui explique un tiers de l'inertie totale oppose essentiellement la SUZE et le MARTINI à la BIERE.

Pour l'interprétation des axes, il est nécessaire de prendre en compte les contributions et les cosinus carrés.

L'appréciation « Vieillotte, dépassée » qui présente pourtant une coordonnée très forte sur l'axe 1 comparativement aux autres, ne contribue que très faiblement à sa création. Elle est de plus l'une des appréciations les plus mal représentées sur cet axe. On ne la prendra pas en compte pour l'interprétation du premier axe.

4.2 Contributions des individus actifs

Le tableau ci-après donne les contributions des individus actifs. Par défaut, le logiciel édite ces valeurs sur les 5 premiers axes.

Ce tableau se compose de 8 colonnes :

- Libellé** : libellé de l'individu actif
- Poids relatif** : poids relatif de l'individu actif. Le poids relatif se calcule de la façon suivante : $(n_q \cdot 100) / n$ avec n_q l'effectif de l'individu actif et n l'effectif total. Par exemple pour l'individu actif « Aime le goût », $P.REL = (291 \cdot 100) / 7241 = 4,02$.
- Distance à l'origine** : carré de la distance de l'individu actif à l'origine. Cette distance donne une idée du caractère périphérique de certains points (comme l'individu actif « Vieillotte, dépassée » pour laquelle $DISTO = 1,40646$).
- Axe 1 – Axe 5** : contributions (en pourcentage) des individus actifs sur les 5 premiers axes. Les contributions décrivent la part de chaque individu actif dans l'inertie totale de l'axe. Pour l'interprétation d'un axe, on sélectionne les individus qui ont les plus fortes contributions.

Nous pouvons mesurer la part d'inertie d'un individu actif par le quotient :

$$CTR(i, \alpha) = \frac{p_i \psi_{i\alpha}^2}{\lambda_\alpha} \times 100$$

Il s'agit de la contribution de l'individu i à la fabrication de l'axe α avec : P_i le poids de i ($291/7241$), $\psi_{i\alpha}$ la coordonnées de l'individu i sur l'axe α et λ_α la valeur propre de l'axe α .

La somme des contributions de toutes les individus actifs sur un axe est égale à 100% (somme en colonne).

Contributions des individus actifs

Identificateur	Poids relatif	Distance à l'origine	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5
Aime le goût	4,02	0,07708	2,56	0,91	4,40	3,91	5,17
Avec des amis	8,04	0,01368	0,15	1,93	0,01	0,22	3,93
Pour se détendre	5,43	0,02518	1,63	0,21	0,69	1,13	0,72
Qui revient cher	6,30	0,11797	5,69	5,03	4,16	10,14	1,83
Rafraîchissante, désaltérante	3,69	0,47791	17,54	7,35	1,51	33,00	6,92
Peu élégante, peu distinguée	1,84	0,13704	2,87	0,03	2,00	3,02	3,91
Produit sympathique	5,79	0,01458	0,85	0,00	1,21	1,61	0,23
Bien avant les repas	6,70	0,13546	3,12	13,05	6,74	0,78	8,53
Bien dans la journée	2,62	0,68617	7,15	25,12	12,99	0,52	10,62
Bien dans la soirée	4,09	0,24599	10,01	6,03	5,06	0,02	0,93
Toute l'année	9,24	0,02156	0,07	2,67	4,23	0,27	3,69
Apprécie des jeunes	6,53	0,08716	2,83	7,02	0,24	0,71	17,51
Volontiers avec invités	8,45	0,01595	0,46	1,72	0,73	2,24	0,17
Vieilloté, dépassée	1,28	1,40646	4,15	20,22	47,52	2,28	2,95
Aussi bien hommes que femmes	6,15	0,03105	0,01	1,17	0,28	16,28	6,55
Très proche	3,00	0,11469	2,17	2,29	4,15	0,97	9,58
Par habitude	2,78	0,05316	1,78	0,38	0,76	0,15	0,61
Fait snob, m'as-tu vu ?	1,84	0,39872	10,45	0,36	0,12	0,11	4,62
On peut mélanger	6,02	0,13185	8,61	0,13	0,54	22,30	11,43
La nuit/Bar/Disco	6,21	0,22897	17,90	4,40	2,68	0,32	0,12

Interprétation

Si toutes les individus avaient la même contribution, celle-ci devrait osciller autour de 5% (100% / 20). Les individus avec des contributions supérieures à 5% ont une influence supérieure à la moyenne.

Les individus actifs « La nuit/Bar/Disco », « Rafraîchissante, désaltérante », « Fait snob, m'as-tu vu ? », « Bien dans la soirée », « On peut mélanger », « Bien dans la journée », « Qui revient cher », contribuent fortement à la création de l'axe 1 (77,36% cumulé).

Sur l'axe 2, les appréciations « Bien dans la journée », « Vieilloté, dépassée », « Bien avant les repas », « Rafraîchissante, désaltérante », « Apprécie des jeunes », « Bien dans la soirée », « Qui revient cher » expliquent près de 85% de l'inertie de l'axe, les appréciations « Bien dans la journée » et « Vieilloté, dépassée » en expliquent près de la moitié à elles seules.

Pour l'interprétation des deux premiers axes factoriels, on prendra en compte les appréciations citées précédemment. Il faut maintenant s'assurer qu'elles sont bien représentées sur ces axes.

4.3 Cosinus carrés des individus actifs

Le tableau ci-après donne les cosinus carrés des individus actifs. Par défaut, le logiciel édite ces valeurs sur les 5 premiers axes.

Ce tableau se compose de 8 colonnes :

- **Libellé** : libellé complet de la modalité
- **Poids relatif** : poids relatif de l'individu actif. Le poids relatif se calcule de la façon suivante : $(n_q \cdot 100) / n$ avec n_q l'effectif de l'individu actif et n l'effectif total. Par exemple pour l'individu actif « Aime le goût », $P.REL = (291 \cdot 100) / 7241 = 4,02$.
- **Distance à l'origine** : carré de la distance de l'individu actif à l'origine. Cette distance donne une idée du caractère périphérique de certains points (comme l'individu actif « Vieilloté, dépassée » pour laquelle $DISTO = 1,40646$).
- **Axe 1 – Axe 5** : cosinus carrés (ou contributions relatives) des individus actifs sur les premiers axes.

Les cosinus carrés concernent la qualité de représentation des individus sur l'axe. Pour l'interprétation, on sélectionne les individus qui ont les cosinus carrés les plus forts : ils sont les mieux représentés au sens où les distances sont les moins altérées par la projection.

Si l'individu i a pour coordonnée $\psi_{i\alpha}$ sur l'axe α et si le carré de la distance à l'origine de cette individu est $d^2(i,G)$, le cosinus carré (ou contribution relative) $COS^2(i,\alpha)$ de la individu i sur l'axe α s'écrit :

$$COS^2(i,\alpha) = \frac{\psi_{i\alpha}^2}{d^2(i,G)}$$

La somme des cosinus carrés d'une modalité sur la totalité des axes est égale à 1 (somme en ligne étendue à tous les axes).

Cosinus carrés des individus actifs

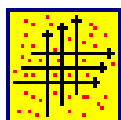
Identificateur	Poids relatif	Distance à l'origine	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5
Aime le goût	4,02	0,07708	0,55	0,13	0,18	0,09	0,05
Avec des amis	8,04	0,01368	0,09	0,79	0,00	0,01	0,10
Pour se détendre	5,43	0,02518	0,79	0,07	0,06	0,06	0,02
Qui revient cher	6,30	0,11797	0,51	0,30	0,07	0,09	0,01
Rafraîchissante, désaltérante	3,69	0,47791	0,66	0,19	0,01	0,13	0,01
Peu élégante, peu distinguée	1,84	0,13704	0,76	0,01	0,10	0,08	0,05
Produit sympathique	5,79	0,01458	0,67	0,00	0,18	0,13	0,01
Bien avant les repas	6,70	0,13546	0,23	0,64	0,09	0,01	0,03
Bien dans la journée	2,62	0,68617	0,26	0,63	0,09	0,00	0,02
Bien dans la soirée	4,09	0,24599	0,66	0,27	0,06	0,00	0,00
Toute l'année	9,24	0,02156	0,02	0,60	0,26	0,01	0,05
Apprécie des jeunes	6,53	0,08716	0,33	0,55	0,01	0,01	0,09
Volontiers avec invités	8,45	0,01595	0,23	0,57	0,07	0,11	0,00
Vieillesse, dépassée	1,28	1,40646	0,15	0,50	0,33	0,01	0,00
Aussi bien hommes que femmes	6,15	0,03105	0,00	0,28	0,02	0,59	0,10
Très proche	3,00	0,11469	0,42	0,30	0,15	0,02	0,08
Par habitude	2,78	0,05316	0,80	0,11	0,06	0,01	0,01
Fait snob, m'as-tu vu ?	1,84	0,39872	0,95	0,02	0,00	0,00	0,02
On peut mélanger	6,02	0,13185	0,72	0,01	0,01	0,19	0,04
La nuit/Bar/Disco	6,21	0,22897	0,84	0,14	0,02	0,00	0,00

Interprétation

Les appréciations « Fait snob, m'as-tu vu ? », « La nuit/Bar/Disco », « Par habitude », « Pour se détendre », « Peu élégante, peu distinguée », « On peut mélanger », « Produit sympathique », « Bien dans la soirée », « Rafraîchissante, désaltérante », « Aime le goût », « Qui revient cher », « Très proche » sont bien représentés sur le premier axe factoriel. Elles sont toutes mieux représentées sur cet axe que sur le second axe factoriel. C'est l'inverse pour les autres appréciations.

Sur le second axe factoriel, les appréciations les mieux représentées sont : « Avec des amis », « Bien avant les repas », « Bien dans la journée », « Toute l'année », « Volontiers avec invités », « Apprécie des jeunes », « Vieillesse, dépassée ».

5. Interprétation générale



Analyse des correspondances multiples (CORMU)

Fichier utilisé : ASPI1000.SBA

Paramétrage de la méthode : on prend comme variables actives V11, V29, V39, V49, V51, V52 et V53. Toutes les autres variables nominales et continues sont sélectionnées comme illustratives à l'exception du coefficient de pondération.

Tous les individus sont conservés pour l'analyse. Nous modifions le seuil de ventilation pour les modalités d'effectifs faibles (5% au lieu de 2% par défaut).

1. Présentation du thème analysé

Dans l'exemple, on choisit d'étudier le thème relatif aux variables socio-démographiques, autrement dit le thème du *signalétique* (âge, sexe, catégorie socioprofessionnelle, etc.).

Les variables choisies, appelées variables actives, constituent les seuls éléments utilisés pour comparer les enquêtés entre eux. Les autres variables, appelées variables illustratives ou complémentaires, serviront à illustrer, compléter ou suggérer des explications pour les similitudes et les différences observées entre les enquêtés.

La liste des variables actives utilisées pour évaluer la similitude entre les individus apparaît dans le tableau ci-après. Toutes les variables actives de cette analyse sont du type nominal.

Variables actives de l'analyse	Nbre de modalités
Sexe de la personne interrogée.	2
Possédez-vous des valeurs mobilières ?	2
Taille d'agglomération (en nombre d'habitants).	5
Diplôme de l'enquêté(e) en 5 classes.	5
Statut d'occupation du logement en 4 classes.	4
Age de l'enquêté(e) en 5 classes.	5
Type d'emploi.	4

Les variables nominales illustratives sont constituées par l'ensemble des 39 autres variables nominales du fichier. Les 8 variables continues du fichier sont également utilisées en illustratives.

Pour cette analyse, on conserve l'ensemble des individus, avec un poids uniforme.

Dans cette analyse, on a choisi un seuil d'apurement de 5%.

L'apurement permet de s'affranchir (artificiellement) des modalités de faibles effectifs qui peuvent avoir des effets perturbateurs sur l'analyse. Les réponses appartenant à ces modalités peu fréquentes seront réparties aléatoirement entre les autres modalités de la variable. Par défaut, sont ventilées les modalités actives dont l'effectif est inférieur à 2%. Les modalités rares concernent souvent les mêmes individus qui forment alors un sous-nuage très concentré sur lui-même mais loin de tous les autres points. Ils peuvent rendre instables

les axes factoriels. L'apurement vise à rendre plus robuste l'analyse. Les modalités ne sont pas abandonnées pour autant. Elles seront positionnées en éléments supplémentaires.

2. CORMU-1 : Tris à plat des variables actives

Le tableau suivant présente le tri à plat des quatre premières variables actives de l'analyse dans un tableau constitué de 5 colonnes :

- **Libellé des modalités** : libellé complet de la modalité.
- **Effectif avant apurement** : effectif de la modalité avant la ventilation des modalités d'effectif inférieur à n% (ici 5%, soit 50 observations).
- **Poids avant apurement** : poids de la modalité avant la ventilation des modalités d'effectif inférieur à n%.
Dans le cas où le poids des individus est uniforme (comme ici), le poids des individus dans la modalité est identique à l'effectif.
- **Effectif après apurement** : effectif de la modalité après la ventilation des modalités d'effectif inférieur à n%. Les individus concernés par la ventilation d'une modalité se voient affecter aléatoirement une autre modalité de la même variable.
- **Poids après apurement** : poids de la modalité après la ventilation des modalités d'effectif inférieur à n%.

Tris à plat des variables actives (Seuil: 5.0 %)

Sexe de la personne interrogée

Libellé des modalités	Effectif avant apurement	Poids avant apurement	Effectif après apurement	Poids après apurement
masculin	469	469,00	469	469,00
féminin	531	531,00	531	531,00

Possédez vous des valeurs mobilières ?

Libellé des modalités	Effectif avant apurement	Poids avant apurement	Effectif après apurement	Poids après apurement
oui	121	121,00	121	121,00
non	879	879,00	879	879,00

Taille d'agglomération (en nombre d'habitants)

Libellé des modalités	Effectif avant apurement	Poids avant apurement	Effectif après apurement	Poids après apurement
moins de 2.000	83	83,00	83	83,00
2.000 - 20.000	87	87,00	87	87,00
20.000 - 100.000	175	175,00	175	175,00
plus de 100.000	329	329,00	329	329,00
Paris	326	326,00	326	326,00

Type d'emploi

Libellé des modalités	Effectif avant apurement	Poids avant apurement	Effectif après apurement	Poids après apurement
Ouvriers	263	263,00	276	276,00
Employés	335	335,00	344	344,00
Cadres	229	229,00	241	241,00
Autres	48	48,00	Ventilée	
Reponse manquante	125	125,00	139	139,00

Interprétation

Seule la variable Type d'emploi est concernée par la ventilation (Modalité « Autres » d'effectif 48).

On peut s'assurer que la somme des effectifs après ventilation est bien égale à 1000.

La répartition des 48 observations s'est faite de la manière suivante : 13 vers la modalité « ouvriers », 9 vers la modalité « employés », 12 vers la modalité « cadres » et 14 vers la modalité formée par les réponses manquantes.

Dans le cas où des modalités ont été ventilées, il est prudent de vérifier si les coordonnées des modalités actives sont proches de celles de ces mêmes modalités placées en illustratif à partir des données brutes (non-apurées). Si ces coordonnées ont des valeurs différentes, cela signifie que la ventilation a notablement affecté la répartition des individus. Il faudrait alors choisir un seuil d'apurement plus faible.

3. CORMU-2 :Tableau de Burt

Le tableau de Burt est un tableau symétrique d'ordre (p,p), où p est le nombre de modalités actives. Il s'agit du croisement deux à deux de toutes les variables actives de l'analyse (après les ventilations éventuelles).

Le tableau est constitué de $(p+1)*(p/2)$ tableaux, soit dans notre cas $8*7/2 = 28$ tableaux de contingence. On utilise l'identifiant court des modalités.

Tableau de BURT

	masc	fémi	vmo1	vmo2	agg1	agg2	agg3	agg4	agg5	emp1	emp2	emp3	49	die1	die2	die3	die4	die5	slo1	slo2	slo3	slo4	agc1	agc2	agc3	agc4	agc5
masc	469	0																									
fémi	0	531																									
vmo1	54	67	121	0																							
vmo2	415	464	0	879																							
agg1	42	41	4	79	83	0	0	0	0																		
agg2	40	47	9	78	0	87	0	0	0																		
agg3	81	94	22	153	0	0	175	0	0																		
agg4	161	168	36	293	0	0	0	329	0																		
agg5	145	181	50	276	0	0	0	0	326																		
emp1	196	80	8	268	16	29	50	106	75	276	0	0	0														
emp2	108	236	36	308	39	33	59	109	104	0	344	0	0														
emp3	125	116	53	188	14	19	36	69	103	0	0	241	0														
49	40	99	24	115	14	6	30	45	44	0	0	0	139														
die1	98	91	10	179	18	16	35	68	52	90	72	11	16	189	0	0	0	0									
die2	163	158	26	295	43	34	67	107	70	137	128	30	26	0	321	0	0	0									
die3	68	90	19	139	12	13	33	51	49	37	72	29	20	0	0	158	0	0									
die4	73	109	29	153	9	16	23	63	71	9	49	75	49	0	0	0	182	0									
die5	67	83	37	113	1	8	17	40	84	3	23	96	28	0	0	0	0	150									
slo1	62	58	11	109	7	20	27	48	18	38	39	34	9	16	45	15	30	14	120	0	0	0					
slo2	150	140	60	230	59	33	61	70	67	69	92	80	49	55	113	36	44	42	0	290	0	0					
slo3	224	299	44	479	10	29	79	191	214	155	188	112	68	103	147	97	91	85	0	0	523	0					
slo4	33	34	6	61	7	5	8	20	27	14	25	15	13	15	16	10	17	9	0	0	0	67					
agc1	67	83	11	139	4	8	24	66	48	36	39	16	59	23	17	35	52	23	9	17	108	16	150	0	0	0	
agc2	129	155	16	268	17	21	49	91	106	79	108	77	20	24	74	65	63	58	40	40	184	20	0	284	0	0	
agc3	102	107	28	181	25	25	40	60	59	64	65	61	19	39	87	20	32	31	45	65	89	10	0	0	209	0	
agc4	84	104	27	161	25	20	26	60	57	53	72	40	23	45	81	26	20	16	20	80	79	9	0	0	0	188	
agc5	87	82	39	130	12	13	36	52	56	44	60	47	18	58	62	12	15	22	6	88	63	12	0	0	0	169	

Par défaut, les valeurs prennent deux décimales.

Interprétation

Les blocs diagonaux sont les tris à plat des variables. Le poids est égal à l'effectif lorsque les individus ont un poids uniforme (comme ici).

On note par exemple que l'enquête concerne 469 hommes et 531 femmes.

Les blocs non diagonaux représentent les tris croisés de tous les couples de variables actives. Les valeurs sont des poids. Lorsque les individus ont des poids uniformes, les poids coïncident avec les effectifs.

Par exemple, 196 hommes sont ouvriers et 236 femmes sont employées.

4. CORMU-3 : Profils horizontaux du tableau de Burt

On édite ici le tableau des pourcentages.

Le tableau suivant n'est pas symétrique. Pour chacun des tableaux de contingence (28 dans le cas présent), ce tableau est constitué des profils lignes (au-dessous de la diagonale) et des transposés des profils colonnes (au-dessus de la diagonale).

Les pourcentages sont calculés sur les poids des individus (les valeurs seraient différentes si l'on avait opéré un redressement de l'échantillon).

Tableau de BURT - Profils horizontaux

	masc	fémi	vmo1	vmo2	agg1	agg2	agg3	agg4	agg5	emp1	emp2	emp3	49	die1	die2	die3	die4	die5	slo1	slo2	slo3	slo4	agc1	agc2	agc3	agc4	agc5
masc	46,9	0,0																									
fémi	0,0	53,1																									
vmo1	44,6	55,4	12,1	0,0																							
vmo2	47,2	52,8	0,0	87,9																							
agg1	50,6	49,4	4,8	95,2	8,3	0,0	0,0	0,0	0,0																		
agg2	46,0	54,0	10,3	89,7	0,0	8,7	0,0	0,0	0,0																		
agg3	46,3	53,7	12,6	87,4	0,0	0,0	17,5	0,0	0,0																		
agg4	48,9	51,1	10,9	89,1	0,0	0,0	0,0	32,9	0,0																		
agg5	44,5	55,5	15,3	84,7	0,0	0,0	0,0	0,0	32,6																		
emp1	71,0	29,0	2,9	97,1	5,8	10,5	18,1	38,4	27,2	27,6	0,0	0,0	0,0														
emp2	31,4	68,6	10,5	89,5	11,3	9,6	17,2	31,7	30,2	0,0	34,4	0,0	0,0														
emp3	51,9	48,1	22,0	78,0	5,8	7,9	14,9	28,6	42,7	0,0	0,0	24,1	0,0														
49	28,8	71,2	17,3	82,7	10,1	4,3	21,6	32,4	31,7	0,0	0,0	0,0	13,9														
die1	51,9	48,1	5,3	94,7	9,5	8,5	18,5	36,0	27,5	47,6	38,1	5,8	8,5	18,9	0,0	0,0	0,0	0,0									
die2	50,8	49,2	8,1	91,9	13,4	10,6	20,9	33,3	21,8	42,7	39,9	9,3	8,1	0,0	32,1	0,0	0,0	0,0									
die3	43,0	57,0	12,0	88,0	7,6	8,2	20,9	32,3	31,0	23,4	45,6	18,4	12,7	0,0	0,0	15,8	0,0	0,0									
die4	40,1	59,9	15,9	84,1	4,9	8,8	12,6	34,6	39,0	4,9	26,9	41,2	26,9	0,0	0,0	0,0	18,2	0,0									
die5	44,7	55,3	24,7	75,3	0,7	5,3	11,3	26,7	56,0	2,0	15,3	64,0	18,7	0,0	0,0	0,0	0,0	15,0									
slo1	51,7	48,3	9,2	90,8	5,8	16,7	22,5	40,0	15,0	31,7	32,5	28,3	7,5	13,3	37,5	12,5	25,0	11,7	12,0	0,0	0,0	0,0					
slo2	51,7	48,3	20,7	79,3	20,3	11,4	21,0	24,1	23,1	23,8	31,7	27,6	16,9	19,0	39,0	12,4	15,2	14,5	0,0	29,0	0,0	0,0					
slo3	42,8	57,2	8,4	91,6	1,9	5,5	15,1	36,5	40,9	29,6	35,9	21,4	13,0	19,7	28,1	18,5	17,4	16,3	0,0	0,0	52,3	0,0					
slo4	49,3	50,7	9,0	91,0	10,4	7,5	11,9	29,9	40,3	20,9	37,3	22,4	19,4	22,4	23,9	14,9	25,4	13,4	0,0	0,0	0,0	6,7					
agc1	44,7	55,3	7,3	92,7	2,7	5,3	16,0	44,0	32,0	24,0	26,0	10,7	39,3	15,3	11,3	23,3	34,7	15,3	6,0	11,3	72,0	10,7	15,0	0,0	0,0	0,0	
agc2	45,4	54,6	5,6	94,4	6,0	7,4	17,3	32,0	37,3	27,8	38,0	27,1	7,0	8,5	26,1	22,9	22,2	20,4	14,1	14,1	64,8	7,0	0,0	28,4	0,0	0,0	
agc3	44,8	51,2	13,4	86,6	12,0	12,0	19,1	28,7	28,2	30,6	31,1	29,2	9,1	18,7	41,6	9,6	15,3	14,8	21,5	31,1	42,6	4,8	0,0	0,0	20,9	0,0	
agc4	44,7	55,3	14,4	85,6	13,3	10,6	13,8	31,9	30,3	28,2	38,3	21,3	12,2	23,9	43,1	13,8	10,6	8,5	10,6	42,6	42,0	4,8	0,0	0,0	0,0	18,8	
agc5	51,5	48,5	23,1	76,9	7,1	7,7	21,3	30,8	33,1	26,0	35,5	27,8	10,7	34,3	36,7	7,1	8,9	13,0	3,6	52,1	37,3	7,1	0,0	0,0	0,0	0,0	

Interprétation

On note par exemple que le fichier est constitué de 46.9% d'hommes et de 53.1% de femmes. On remarque que 41.8% des hommes sont ouvriers contre seulement 15.1% des femmes. Parmi les personnes qui déclarent posséder des valeurs mobilières 44.6% sont des hommes et 55.4% sont des femmes.

5. CORMU-4 : Tableau des valeurs propres

Le tableau des valeurs propres est constitué de 4 colonnes.

- **Numéro** : numéro de l'axe factoriel.
- **Valeur propre** : la valeur propre (inertie) associée à l'axe.
- **Pourcentage** : indique la part de la valeur propre (ou inertie) prise en compte sur l'axe factoriel.
- **Pourcentage cumulé** : pourcentage cumulé des parts de valeurs propres sur les n premiers axes.

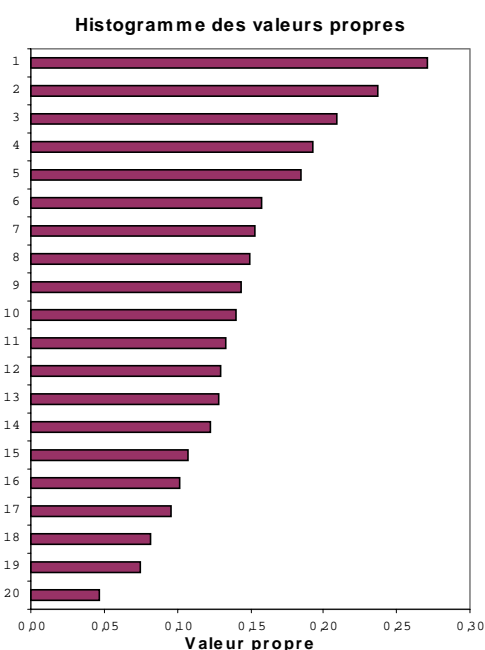
La somme des valeurs propres vaut $(J-Q) / Q$ avec J le nombre total de modalités actives après la ventilation ($J = 27$ dans l'exemple) et Q le nombre de variables actives ($Q = 7$ dans l'exemple). Ainsi dans l'exemple, la somme des valeurs propres a pour valeur : $(27-7) / 7 = 2.8571$.

Le logiciel édite l'ensemble des valeurs propres, soit $J-Q$ valeurs propres. Dans l'exemple, on obtient l'édition de $27-7 = 20$ valeurs propres.

Le tableau des valeurs propres peut aider à déterminer le nombre d'axes factoriels à archiver pour les utilisations ultérieures.

Tableau des valeurs propres
Trace de la matrice: 2.85714

Numéro	Valeur propre	Pourcentage	Pourcentage cumulé
1	0,2703	9,46	9,46
2	0,2369	8,29	17,75
3	0,2084	7,29	25,05
4	0,1922	6,73	31,77
5	0,1846	6,46	38,23
6	0,1578	5,52	43,76
7	0,1534	5,37	49,13
8	0,1493	5,23	54,35
9	0,1441	5,04	59,40
10	0,1398	4,89	64,29
11	0,1326	4,64	68,93
12	0,1300	4,55	73,48
13	0,1284	4,49	77,97
14	0,1222	4,28	82,25
15	0,1070	3,74	86,00
16	0,1015	3,55	89,55
17	0,0954	3,34	92,89
18	0,0821	2,87	95,76
19	0,0748	2,62	98,38
20	0,0462	1,62	100,00



Interprétation

On remarque que la décroissance des valeurs propres n'est pas régulière. Elle le serait sur un tableau purement aléatoire. L'histogramme des valeurs propres fait apparaître des différences d'inertie importantes entre la 5^{ème} et la 6^{ème} valeur.

La faible part de variance expliquée sur les premiers axes est une caractéristique de l'analyse factorielle des correspondances multiples (qui donne généralement des mesures pessimistes de l'information extraite).

Dans cet exemple, il convient d'archiver au moins les 5 premiers axes de l'analyse. Il est cependant conseillé d'utiliser en général au moins la moitié des axes factoriels. Ici, on archivera certainement les 14 premiers axes, le palier entre les axes 14 et 15 étant notablement marqué, pour une classification ultérieure.

6. CORMU-5 : Coordonnées des modalités actives

Le tableau ci-après donne les coordonnées des modalités actives. Par défaut, le logiciel édite ces valeurs sur les 5 premiers axes.

Ce tableau se compose de 8 colonnes :

- **Libellé** : libellé complet de la modalité
- **Poids relatif** : poids relatif la modalité. Le poids relatif se calcule de la façon suivante : $(n_q \cdot 100) / (n \cdot Q)$ avec n_q l'effectif de la modalité, n l'effectif total et Q le nombre de variables actives. Par exemple pour la modalité masculin, $P.REL = (469 \cdot 100) / (1000 \cdot 7) = 6.70$.

Il faut noter que lorsque les individus ont des poids différents, ce n'est pas l'effectif, mais le poids des individus qui est utilisé.

- **Distance à l'origine** : carré de la distance du Khi-2 à l'origine. Cette distance donne une idée du caractère périphérique de certains points (comme la modalité « logé gratuitement, autres » de la variable « Statut d'occupation du logement » pour laquelle $DISTO = 13.93$). Il s'agit généralement de points à faible masse. La distance à l'origine

(au centre de gravité G du nuage) ne dépend en fait que de l'effectif dans la modalité. La formule est la suivante :

$$d^2(j, G) = (n_j / n) - 1, \text{ avec } n_j \text{ l'effectif de la modalité } j \text{ et } n \text{ l'effectif total}$$

- **Axe 1 – Axe 5** : coordonnées des modalités actives sur les premiers axes.

Coordonnées des modalités actives

Libellé	Poids relatif	Distance à l'origine	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5
Sexe de la personne interrogée							
masculin	6,70	1,13220	0,29	-0,08	-0,43	0,47	0,25
féminin	7,59	0,88324	-0,26	0,07	0,38	-0,41	-0,22

Possédez vous des valeurs mobilières ?

oui	1,73	7,26446	-0,69	-1,46	0,25	0,23	-0,06
non	12,56	0,13766	0,10	0,20	-0,03	-0,03	0,01

Taille d'agglomération (en nombre d'habitants)

moins de 2.000	1,19	11,04820	1,06	-0,83	1,06	-0,75	0,06
2.000 - 20.000	1,24	10,49430	0,55	-0,26	-0,28	-0,80	0,61
20.000 - 100.000	2,50	4,71429	0,27	-0,07	0,17	-0,07	0,12
plus de 100.000	4,70	2,03951	0,04	0,40	-0,05	0,22	0,27
Paris	4,66	2,06748	-0,60	-0,08	-0,24	0,22	-0,52

Type d'emploi

Ouvriers	3,94	2,62319	0,88	0,47	-0,54	0,66	0,20
Employés	4,91	1,90698	0,19	0,20	0,38	-0,67	-0,63
Cadres	3,44	3,14938	-0,80	-0,89	-0,74	-0,02	0,14
Reponse manquante	1,99	6,19424	-0,80	0,12	1,41	0,38	0,91

Diplôme de l'enquêté(e) en 5 classes

Aucun	2,70	4,29101	0,70	0,23	0,23	0,93	-0,34
CEP ou fin études	4,59	2,11526	0,80	-0,08	-0,05	-0,29	0,07
BEPC-BE-BEPS	2,26	5,32911	-0,23	0,62	0,17	-0,47	-0,56
Bac - Brevet sup.	2,60	4,49451	-0,93	0,06	0,32	-0,26	0,95
Université, gde école	2,14	5,66667	-1,23	-0,84	-0,73	0,26	-0,27

Statut d'occupation du logement en 4 classes

en accession	1,71	7,33333	0,31	0,06	-0,85	-1,02	1,30
propriétaire	4,14	2,44828	0,44	-1,00	0,51	0,07	0,01
locataire	7,47	0,91205	-0,27	0,51	-0,15	0,15	-0,33
logé gratuit, autre	0,96	13,92540	-0,34	0,25	0,50	0,33	0,20

Age de l'enquêté(e) en 5 classes

Moins de 25 ans	2,14	5,66667	-0,81	0,98	0,89	0,68	0,80
25 à 34 ans	4,06	2,52113	-0,35	0,45	-0,63	-0,47	-0,41
35 à 49 ans	2,99	3,78469	0,33	-0,36	-0,41	-0,41	0,69
50 à 64 ans	2,69	4,31915	0,51	-0,30	0,42	-0,21	-0,25
65 ans et plus	2,41	4,91716	0,34	-0,84	0,32	0,93	-0,59

7. CORMU-6 :Contributions des modalités actives

Le tableau ci-après donne les contributions des modalités actives. Par défaut, le logiciel édite ces valeurs sur les 5 premiers axes.

Ce tableau se compose de 8 colonnes :

- **Libellé** : libellé complet de la modalité
- **Poids relatif** : poids relatif la modalité. Le poids relatif se calcule de la façon suivante : $(n_q \cdot 100) / (n \cdot Q)$ avec n_q l'effectif de la modalité, n l'effectif total et Q le nombre de variables actives. Par exemple pour la modalité masculin, $P.REL = (469 \cdot 100) / (1000 \cdot 7) = 6.70$.

Il faut noter que lorsque les individus ont des poids différents, ce n'est pas l'effectif, mais le poids des individus qui est utilisé.

- **Distance à l'origine** : carré de la distance du Khi-2 à l'origine. Cette distance donne une idée du caractère périphérique de certains points (comme la modalité « logé gratuitement, autres » de la variable « Statut d'occupation du logement » pour laquelle DISTO = 13.93). Il s'agit généralement de points à faible masse. La distance à l'origine (au centre de gravité G du nuage) ne dépend en fait que de l'effectif dans la modalité. La formule est la suivante :

$$d^2(j, G) = (n / n_j) - 1, \text{ avec } n_j \text{ l'effectif de la modalité } j \text{ et } n \text{ l'effectif total}$$

- **Axe 1 – Axe 5** : contributions absolues des modalités actives sur les 5 premiers axes.

Les contributions absolues, ou contributions, décrivent la part de chaque modalité dans l'inertie totale de l'axe. Pour l'interprétation d'un axe par exemple, on sélectionne les modalités qui ont les plus fortes contributions.

Si la modalité j a pour coordonnées sur l'axe ϕ_j et a une masse m_j , et si λ désigne la valeur propre de cet axe, la contribution absolue $ca(j)$ de la modalité s'écrit :

$$ca(j) = m_j \phi_j^2 / \lambda$$

La contribution de la modalité masculin sur le premier facteur a donc pour valeur : $6.7 * 0.29^2 / 0.2703 = 2.1$.

La somme des contributions de toutes les modalités actives sur un axe est égale à 100% (somme en colonne).

Contributions des modalités actives

Libellé	Poids relatif	Distance à l'origine	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5
Sexe de la personne interrogée							
masculin	6,70	1,13220	2,09	0,17	5,97	7,64	2,28
féminin	7,59	0,88324	1,84	0,15	5,28	6,75	2,02

Possédez vous des valeurs mobilières ?

oui	1,73	7,26446	3,08	15,48	0,51	0,49	0,03
non	12,56	0,13766	0,42	2,13	0,07	0,07	0,00

Taille d'agglomération (en nombre d'habitants)

moins de 2.000	1,19	11,04820	4,96	3,42	6,36	3,50	0,02
2.000 - 20.000	1,24	10,49430	1,41	0,35	0,47	4,16	2,50
20.000 - 100.000	2,50	4,71429	0,66	0,06	0,33	0,07	0,19
plus de 100.000	4,70	2,03951	0,03	3,16	0,05	1,16	1,87
Paris	4,66	2,06748	6,23	0,14	1,26	1,21	6,72

Type d'emploi

Ouvriers	3,94	2,62319	11,18	3,63	5,58	8,92	0,84
Employés	4,91	1,90698	0,63	0,80	3,47	11,40	10,45
Cadres	3,44	3,14938	8,25	11,43	8,98	0,01	0,37
Reponse manquante	1,99	6,19424	4,74	0,13	18,87	1,51	8,98

Diplôme de l'enquêté(e) en 5 classes

Aucun	2,70	4,29101	4,96	0,60	0,67	12,12	1,74
CEP ou fin études	4,59	2,11526	10,88	0,13	0,06	2,02	0,11
BEPC-BE-BEPS	2,26	5,32911	0,43	3,70	0,30	2,65	3,83
Bac - Brevet sup.	2,60	4,49451	8,31	0,04	1,28	0,88	12,63
Université, gde école	2,14	5,66667	12,06	6,45	5,54	0,76	0,84

Statut d'occupation du logement en 4 classes

en accession	1,71	7,33333	0,60	0,03	5,88	9,20	15,68
propriétaire	4,14	2,44828	2,96	17,56	5,16	0,11	0,00
locataire	7,47	0,91205	2,04	8,19	0,84	0,89	4,42
logé gratuit, autre	0,96	13,92540	0,40	0,26	1,16	0,55	0,22

Age de l'enquêté(e) en 5 classes

Moins de 25 ans	2,14	5,66667	5,18	8,73	8,18	5,19	7,41
25 à 34 ans	4,06	2,52113	1,87	3,42	7,83	4,76	3,68
35 à 49 ans	2,99	3,78469	1,19	1,64	2,46	2,59	7,65
50 à 64 ans	2,69	4,31915	2,57	1,04	2,28	0,60	0,94
65 ans et plus	2,41	4,91716	1,02	7,19	1,18	10,79	4,56

L'analyse compte après apurement 27 modalités actives. Pour chaque axe, le pourcentage d'inertie théorique moyen expliqué par chaque modalité est de 3,7% (100%/27). Or on constate sur l'axe 1 que les contributions varient de 0,03% à 12,06%.

Seules les modalités dont la contribution est élevée sont à considérer pour l'interprétation d'un axe. Les modalités à considérer pour l'interprétation de l'axe 1 sont dans l'ordre d'apparition dans le tableau : moins de 2.000 (4,96%), Paris (6,23%), Ouvriers (11,8%), Cadres (8,25%), Aucun (4,96%), CEP ou fin études (10,88%), Bac - Brevet sup. (8,31%), Université, gde école (12,06%), Moins de 25 ans (5,18%).

Ces 9 modalités représentent à elles seules 72% de l'inertie de l'axe 1.

Pour obtenir la contribution globale d'une variable à la création d'un axe, il suffit d'additionner les contributions de toutes ses modalités. Ainsi, sur l'axe 1, le sexe (3,93%), la possession de valeurs mobilières (3,50%) et le statut d'occupation du logement (6%) contribuent très faiblement à la création de l'axe 1.

Par contre, les variables Diplôme de l'enquêté en 5 classes (36,64%), Type d'emploi (24,8%) et Taille d'agglomération (13,29%) expliquent près de 75% de l'inertie sur l'axe 1. Pour chacune de ces 3 variables, on constate que ces fortes contributions globales sont dues essentiellement à quelques modalités et non à l'ensemble.

8. CORMU-7 : Cosinus carrés des modalités actives

Le tableau ci-après donne les cosinus carrés des modalités actives. Par défaut, le logiciel édite ces valeurs sur les 5 premiers axes.

Ce tableau se compose de 8 colonnes :

- **Libellé** : libellé complet de la modalité
- **Poids relatif** : poids relatif la modalité. Le poids relatif se calcule de la façon suivante : $(n_q \cdot 100) / (n \cdot Q)$ avec n_q l'effectif de la modalité, n l'effectif total et Q le nombre de variables actives. Par exemple pour la modalité masculin, $P.REL = (469 \cdot 100) / (1000 \cdot 7) = 6.70$.

Il faut noter que lorsque les individus ont des poids différents, ce n'est pas l'effectif, mais le poids des individus qui est utilisé.

- **Distance à l'origine** : carré de la distance du Khi-2 à l'origine. Cette distance donne une idée du caractère périphérique de certains points (comme la modalité « logé gratuitement, autres » de la variable « Statut d'occupation du logement » pour laquelle $DISTO = 13.93$). Il s'agit généralement de points à faible masse. La distance à l'origine (au centre de gravité G du nuage) ne dépend en fait que de l'effectif dans la modalité. La formule est la suivante :

$$d^2(j, G) = (n / n_j) - 1, \text{ avec } n_j \text{ l'effectif de la modalité } j \text{ et } n \text{ l'effectif total}$$

- **Axe 1 – Axe 5** : cosinus carrés (ou contributions relatives) de chacune des modalités actives sur les premiers axes.

Les contributions relatives, ou cosinus carrés, concernent la qualité de représentation des modalités sur l'axe. Pour l'interprétation, on sélectionne les modalités qui ont les cosinus carrés les plus forts : elles sont les mieux représentées au sens où les distances sont les moins altérées par la projection.

Si la modalité j a pour coordonnées ϕ_j sur l'axe et si le carré de la distance du Khi-2 à l'origine de cette modalité est $d^2(j)$, le cosinus carré (ou contribution relative) $cr(j)$ de la modalité s'écrit :

$$cr(j) = \phi_j^2 / d^2(j)$$

On a par exemple pour la modalité masculin sur le premier facteur : $0.29^2 / 1.13^2 = 0.07$.

La somme des cosinus carrés d'une modalité sur la totalité des axes est égale à 1 (somme en ligne étendue à tous les axes).

Cosinus carrés des modalités actives

Libellé	Poids relatif	Distance à l'origine	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5
Sexe de la personne interrogée							
masculin	6,70	1,13220	0,07	0,01	0,16	0,19	0,06
féminin	7,59	0,88324	0,07	0,01	0,16	0,19	0,06

Possédez vous des valeurs mobilières ?

oui	1,73	7,26446	0,07	0,29	0,01	0,01	0,00
non	12,56	0,13766	0,07	0,29	0,01	0,01	0,00

Taille d'agglomération (en nombre d'habitants)

moins de 2.000	1,19	11,04820	0,10	0,06	0,10	0,05	0,00
2.000 - 20.000	1,24	10,49430	0,03	0,01	0,01	0,06	0,04
20.000 - 100.000	2,50	4,71429	0,02	0,00	0,01	0,00	0,00
plus de 100.000	4,70	2,03951	0,00	0,08	0,00	0,02	0,04
Paris	4,66	2,06748	0,18	0,00	0,03	0,02	0,13

Type d'emploi

Ouvriers	3,94	2,62319	0,29	0,08	0,11	0,17	0,01
Employés	4,91	1,90698	0,02	0,02	0,08	0,23	0,21
Cadres	3,44	3,14938	0,21	0,25	0,17	0,00	0,01
Reponse manquante	1,99	6,19424	0,10	0,00	0,32	0,02	0,13

Diplôme de l'enquêté(e) en 5 classes

Aucun	2,70	4,29101	0,12	0,01	0,01	0,20	0,03
CEP ou fin études	4,59	2,11526	0,30	0,00	0,00	0,04	0,00
BEPC-BE-BEPS	2,26	5,32911	0,01	0,07	0,01	0,04	0,06
Bac - Brevet sup.	2,60	4,49451	0,19	0,00	0,02	0,01	0,20
Université, gde école	2,14	5,66667	0,27	0,13	0,10	0,01	0,01

Statut d'occupation du logement en 4 classes

en accession	1,71	7,33333	0,01	0,00	0,10	0,14	0,23
propriétaire	4,14	2,44828	0,08	0,41	0,11	0,00	0,00
locataire	7,47	0,91205	0,08	0,28	0,03	0,03	0,12
logé gratuit, autre	0,96	13,92540	0,01	0,00	0,02	0,01	0,00

Age de l'enquêté(e) en 5 classes

Moins de 25 ans	2,14	5,66667	0,12	0,17	0,14	0,08	0,11
25 à 34 ans	4,06	2,52113	0,05	0,08	0,16	0,09	0,07
35 à 49 ans	2,99	3,78469	0,03	0,03	0,05	0,04	0,12
50 à 64 ans	2,69	4,31915	0,06	0,02	0,04	0,01	0,01
65 ans et plus	2,41	4,91716	0,02	0,14	0,02	0,17	0,07

Les contributions relatives, ou cosinus carrés, concernent la qualité de représentation des modalités sur l'axe. Pour l'interprétation, on sélectionne les modalités qui ont les cosinus carrés les plus forts : elles sont les mieux représentées au sens où les distances sont les moins altérées par la projection.

Si la modalité j a pour coordonnées ϕ_j sur l'axe et si le carré de la distance du Khi-2 à l'origine de cette modalité est $d^2(j)$, le cosinus carré (ou contribution relative) $cr(j)$ de la modalité s'écrit :

$$cr(j) = \phi_j^2 / d^2(j)$$

On a par exemple pour la modalité masculin sur le premier facteur : $0.29^2 / 1.13^2 = 0.07$.

La somme des cosinus carrés d'une modalité sur la totalité des axes est égale à 1 (somme en ligne étendue à tous les axes).

Interprétation

Dans cet exemple, les modalités à prendre en compte pour l'interprétation du premier axe doivent au minimum avoir un cosinus carré de 0,10.

9. CORMU-8 : Coordonnées des modalités actives et illustratives

Le tableau ci-après donne les coordonnées de toutes les modalités (actives en tête et illustratives ensuite) à partir des données brutes (avant ventilation). Ainsi, les modalités actives abandonnées pour la construction des axes du fait de la faiblesse de leur effectif sont repositionnées ici (exemple : modalité « autres » avant ventilation de la variable « Type d'emploi »).

Ce tableau se décompose de la manière suivante :

- **Libellé** : libellé complet de la modalité.
- **Effectif** : effectif de la modalité.
- **Poids absolu** : poids de chaque modalité. Le poids des individus est ici uniforme et égal à 1.
- **Distance à l'origine** : carré de la distance du Khi-2 à l'origine. Le calcul est identique à celui qui est fait pour les modalités actives.
- **Axe 1 – Axe 5** : coordonnées des modalités actives et illustratives sur les 5 premiers axes.

La comparaison des coordonnées des modalités actives sur les sorties CORMU-5 et CORMU-8 montre que les coordonnées de la variable « Type d'emploi » (dont l'une des modalités a été ventilée) ont des valeurs très légèrement différentes. Ceci confirme que le seuil d'apurement de 5% était acceptable.

L'édition du tableau a été tronquée.

Coordonnées des modalités actives et illustratives

Libellé	Effectif	Poids absolu	Distance à l'origine	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5
Sexe de la personne interrogée								
masculin	469	469,00	1,13220	0,29	-0,08	-0,43	0,47	0,25
féminin	531	531,00	0,88324	-0,26	0,07	0,38	-0,41	-0,22

Possédez vous des valeurs mobilières ?

oui	121	121,00	7,26446	-0,69	-1,46	0,25	0,23	-0,06
non	879	879,00	0,13766	0,10	0,20	-0,03	-0,03	0,01

Taille d'agglomération (en nombre d'habitants)

moins de 2.000	83	83,00	11,04820	1,06	-0,83	1,06	-0,75	0,06
2.000 - 20.000	87	87,00	10,49430	0,55	-0,26	-0,28	-0,80	0,61
20.000 - 100.000	175	175,00	4,71429	0,27	-0,07	0,17	-0,07	0,12
plus de 100.000	329	329,00	2,03951	0,04	0,40	-0,05	0,22	0,27
Paris	326	326,00	2,06748	-0,60	-0,08	-0,24	0,22	-0,52

Type d'emploi

Ouvriers	263	263,00	2,80228	0,86	0,51	-0,57	0,67	0,18
Employés	335	335,00	1,98507	0,16	0,22	0,38	-0,68	-0,63
Cadres	229	229,00	3,36681	-0,85	-0,86	-0,77	-0,01	0,12
Autres	48	48,00	19,83330	0,73	-0,75	0,50	0,03	0,47
Reponse manquante	125	125,00	7,00000	-0,96	0,20	1,39	0,42	0,91

Diplôme de l'enquêté(e) en 5 classes

Aucun	189	189,00	4,29101	0,70	0,23	0,23	0,93	-0,34
CEP ou fin études	321	321,00	2,11526	0,80	-0,08	-0,05	-0,29	0,07
BEPC-BE-BEPS	158	158,00	5,32911	-0,23	0,62	0,17	-0,47	-0,56
Bac - Brevet sup.	182	182,00	4,49451	-0,93	0,06	0,32	-0,26	0,95
Université,gde école	150	150,00	5,66667	-1,23	-0,84	-0,73	0,26	-0,27

Statut d'occupation du logement en 4 classes

en accession	120	120,00	7,33333	0,31	0,06	-0,85	-1,02	1,30
propriétaire	290	290,00	2,44828	0,44	-1,00	0,51	0,07	0,01
locataire	523	523,00	0,91205	-0,27	0,51	-0,15	0,15	-0,33
logé gratuit, autre	67	67,00	13,92540	-0,34	0,25	0,50	0,33	0,20

Age de l'enquêté(e) en 5 classes

Moins de 25 ans	150	150,00	5,66667	-0,81	0,98	0,89	0,68	0,80
25 à 34 ans	284	284,00	2,52113	-0,35	0,45	-0,63	-0,47	-0,41
35 à 49 ans	209	209,00	3,78469	0,33	-0,36	-0,41	-0,41	0,69
50 à 64 ans	188	188,00	4,31915	0,51	-0,30	0,42	-0,21	-0,25
65 ans et plus	169	169,00	4,91716	0,34	-0,84	0,32	0,93	-0,59

La famille est le seul endroit où l'on se sente bien

oui	561	561,00	0,78253	0,40	-0,12	0,10	-0,02	-0,02
non	431	431,00	1,32019	-0,53	0,16	-0,13	0,02	0,03
Reponse manquante	8	8,00	124,00000	0,11	-0,20	0,13	0,34	-0,54

Opinion à propos du mariage

union indissoluble	231	231,00	3,32900	0,46	-0,23	0,19	0,19	-0,02
dissout si pb. grave	342	342,00	1,92398	0,08	-0,15	0,08	-0,14	0,03
dissout si accord	387	387,00	1,58398	-0,35	0,25	-0,19	0,01	-0,01
ne sait pas	39	39,00	24,64100	0,05	0,21	0,00	0,06	-0,08
Reponse manquante	1	1,00	999,00000	-0,79	-0,77	0,81	-0,09	0,42

...

10. CORMU-9 : Valeurs-tests des modalités actives et illustratives

Le tableau ci-après donne les valeurs-tests de toutes les modalités (actives en tête et illustratives ensuite) à partir des données brutes (avant ventilation). Ainsi, les modalités actives abandonnées pour la construction des axes du fait de la faiblesse de leur effectif sont repositionnées ici (exemple : modalité « autres » avant ventilation de la variable « Type d'emploi »).

Ce tableau se décompose de la manière suivante :

- **Libellé** : libellé complet de la modalité.
- **Effectif** : effectif de la modalité.
- **Poids absolu** : poids de chaque modalité. Le poids des individus est ici uniforme et égal à 1.

- **Distance à l'origine** : carré de la distance du Khi-2 à l'origine. Le calcul est identique à celui qui est fait pour les modalités actives.
- **Axe 1 – Axe 5** : valeurs-tests de chacune des modalités sur les 5 premiers axes.

La valeur-test est d'autant plus forte que la modalité correspondante occupe une position significative sur l'axe. Une modalité est d'autant plus intéressante sur un axe que sa valeur-test est plus grande. Dans le cas usuel, on considère qu'une valeur-test supérieure à 2 en valeur absolue indique que la modalité correspondante est significativement différente du centre de gravité.

L'édition du tableau a été tronquée.

Valeurs-Tests des modalités actives et illustratives

Libellé	Effectif	Poids absolu	Distance à l'origine	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5
Sexe de la personne interrogée								
masculin	469	469,00	1,13220	8,62	-2,31	-12,80	13,90	7,45
féminin	531	531,00	0,88324	-8,62	2,31	12,80	-13,90	-7,45

Possédez vous des valeurs mobilières ?

oui	121	121,00	7,26446	-8,14	-17,08	2,92	2,74	-0,71
non	879	879,00	0,13766	8,14	17,08	-2,92	-2,74	0,71

Taille d'agglomération (en nombre d'habitants)

moins de 2.000	83	83,00	11,04820	10,11	-7,86	10,05	-7,16	0,58
2.000 - 20.000	87	87,00	10,49430	5,40	-2,52	-2,75	-7,82	5,95
20.000 - 100.000	175	175,00	4,71429	3,89	-1,08	2,42	-1,03	1,74
plus de 100.000	329	329,00	2,03951	0,87	8,83	-1,00	4,83	6,00
Paris	326	326,00	2,06748	-13,22	-1,84	-5,22	4,92	-11,35

Type d'emploi

Ouvriers	263	263,00	2,80228	16,14	9,65	-10,73	12,60	3,46
Employés	335	335,00	1,98507	3,64	4,96	8,50	-15,19	-14,21
Cadres	229	229,00	3,36681	-14,64	-14,87	-13,18	-0,20	2,11
Autres	48	48,00	19,83330	5,20	-5,30	3,55	0,24	3,31
Reponse manquante	125	125,00	7,00000	-11,44	2,40	16,60	4,99	10,85

Diplôme de l'enquêté(e) en 5 classes

Aucun	189	189,00	4,29101	10,75	3,50	3,46	14,17	-5,26
CEP ou fin études	321	321,00	2,11526	17,40	-1,75	-1,17	-6,32	1,47
BEPC-BE-BEPS	158	158,00	5,32911	-3,12	8,54	2,28	-6,50	-7,66
Bac - Brevet sup.	182	182,00	4,49451	-13,86	0,87	4,77	-3,81	14,11
Université, gde école	150	150,00	5,66667	-16,38	-11,21	-9,75	3,47	-3,58

Statut d'occupation du logement en 4 classes

en accession	120	120,00	7,33333	3,59	0,71	-9,87	-11,85	15,17
propriétaire	290	290,00	2,44828	8,88	-20,24	10,29	1,42	0,22
locataire	523	523,00	0,91205	-8,98	16,86	-5,06	5,01	-10,94
logé gratuit, autre	67	67,00	13,92540	-2,84	2,13	4,25	2,82	1,73

Age de l'enquêté(e) en 5 classes

Moins de 25 ans	150	150,00	5,66667	-10,73	13,04	11,84	9,06	10,61
25 à 34 ans	284	284,00	2,52113	-7,03	8,90	-12,62	-9,45	-8,14
35 à 49 ans	209	209,00	3,78469	5,34	-5,86	-6,74	-6,64	11,17
50 à 64 ans	188	188,00	4,31915	7,74	-4,61	6,40	-3,14	-3,86
65 ans et plus	169	169,00	4,91716	4,82	-11,97	4,55	13,21	-8,41

La famille est le seul endroit où l'on se sente bien

oui	561	561,00	0,78253	14,47	-4,36	3,55	-0,59	-0,54
non	431	431,00	1,32019	-14,56	4,47	-3,63	0,42	0,82
Reponse manquante	8	8,00	124,00000	0,31	-0,56	0,36	0,95	-1,53

Opinion à propos du mariage

union indissoluble	231	231,00	3,32900	7,94	-4,06	3,33	3,22	-0,33
dissout si pb. grave	342	342,00	1,92398	1,82	-3,42	1,76	-3,23	0,68
dissout si accord	387	387,00	1,58398	-8,71	6,37	-4,66	0,21	-0,19
ne sait pas	39	39,00	24,64100	0,30	1,32	0,03	0,39	-0,52
Reponse manquante	1	1,00	999,00000	-0,79	-0,77	0,81	-0,09	0,42

...

11. CORMU-10 : Coordonnées, Contributions, Cosinus Carrés des individus actifs

11.1 Coordonnées des individus actifs

Le tableau ci-après donne les coordonnées des individus actifs.

Ce tableau se compose de 8 colonnes :

- **Identificateur** : identificateur de l'individu.
- **Poids relatif** : poids relatif la modalité. Le poids relatif se calcule de la façon suivante : $(n_i/100) / n$ avec n_i le poids de l'individu, n le poids total. Dans cet exemple, les individus ont un poids uniforme. Ils ont donc le même poids relatif.
- **Distance à l'origine** : carré de la distance du Khi-2 à l'origine. Cette distance donne une idée du caractère périphérique de certains individus.
- **Axe 1 – Axe 5** : coordonnées des individus sur les 5 premiers axes.

L'édition du tableau a été tronquée.

Coordonnées des individus actifs

Identificateur	Poids relatif	Distance à l'origine	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5
1	0,10	2,44	-0,28	-0,01	-0,31	0,03	-0,42
2	0,10	2,45	-0,60	-0,66	-0,72	0,16	-0,26
3	0,10	3,72	-0,58	-1,37	-0,30	0,33	-0,23
4	0,10	3,21	-1,16	-0,66	-0,58	-0,01	-0,55
5	0,10	2,23	-0,79	-0,22	-0,93	0,19	-0,37
6	0,10	2,19	-0,94	-0,17	-0,67	-0,10	-0,53
7	0,10	2,41	-0,75	-0,62	-0,47	-0,13	-0,42
8	0,10	1,85	-0,59	0,41	0,01	-0,48	-0,38
9	0,10	1,98	0,00	-0,21	-0,38	0,09	-0,21
10	0,10	2,58	-0,28	-0,68	0,18	0,12	-0,73
11	0,10	1,97	-0,39	0,58	-0,04	-0,55	-0,88
12	0,10	1,69	0,08	0,13	-0,04	-0,47	-0,31
13	0,10	2,43	0,64	-0,17	-0,03	0,69	-0,21
14	0,10	2,58	0,38	-0,05	-0,05	0,23	-0,29
15	0,10	2,10	-0,05	0,61	-0,58	0,17	-0,45
16	0,10	1,95	0,20	0,50	-0,57	0,63	-0,38
17	0,10	3,52	0,11	-0,55	-0,22	0,25	0,00
18	0,10	3,96	-0,07	0,54	-0,38	0,23	-0,27
19	0,10	1,51	-0,11	0,37	-0,11	-0,49	-0,67
20	0,10	2,91	-0,38	0,55	0,52	0,91	0,26

11.2 Contributions des individus actifs

Le tableau ci-après donne les contributions des individus actifs à la création des cinq premiers axes factoriels.

Ce tableau se compose de 8 colonnes :

- **Identificateur** : identifiant de l'individu
- **Poids relatif** : indique le poids donné à chaque individu, dans ce cas uniforme (0,10=100/1000).
- **Distance à l'origine** : carré de la distance de l'individu au centre de gravité de tous les individus. Cette colonne permet de trouver facilement quelles sont les individus les plus proches du centre de gravité, et ceux qui sont les plus "originaux" (les individus qui sont à plus grande distance du centre de gravité). La distance au centre est en quelque sorte un critère "d'originalité" de l'élément.

- **Axe 1 – Axe 5** : contributions (en pourcentage) des individus à l'inertie portée par chaque axe. Nous pouvons mesurer la part d'inertie d'un individu par le quotient :

$$CTR(i, \alpha) = \frac{p_i \psi_{i\alpha}^2}{\lambda_\alpha} \times 100$$

Il s'agit de la contribution de l'individu i à la fabrication de l'axe α avec : P_i le poids de i (1/1000), $\psi_{i\alpha}$ la coordonnées de l'individu i sur l'axe α et λ_α la valeur propre de l'axe α .

Le tableau ci-après a été tronqué.

Identificateur	Poids relatif	Distance à l'origine	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5
1	0,10	2,44	0,03	0,00	0,05	0,00	0,10
2	0,10	2,45	0,13	0,18	0,25	0,01	0,04
3	0,10	3,72	0,12	0,79	0,04	0,06	0,03
4	0,10	3,21	0,50	0,18	0,16	0,00	0,17
5	0,10	2,23	0,23	0,02	0,41	0,02	0,08
6	0,10	2,19	0,33	0,01	0,22	0,01	0,15
7	0,10	2,41	0,21	0,16	0,10	0,01	0,09
8	0,10	1,85	0,13	0,07	0,00	0,12	0,08
9	0,10	1,98	0,00	0,02	0,07	0,00	0,02
10	0,10	2,58	0,03	0,19	0,02	0,01	0,29
11	0,10	1,97	0,06	0,14	0,00	0,16	0,42
12	0,10	1,69	0,00	0,01	0,00	0,11	0,05
13	0,10	2,43	0,15	0,01	0,00	0,25	0,03
14	0,10	2,58	0,05	0,00	0,00	0,03	0,04
15	0,10	2,10	0,00	0,16	0,16	0,01	0,11
16	0,10	1,95	0,02	0,10	0,15	0,20	0,08
17	0,10	3,52	0,00	0,13	0,02	0,03	0,00
18	0,10	3,96	0,00	0,12	0,07	0,03	0,04
19	0,10	1,51	0,00	0,06	0,01	0,12	0,25
20	0,10	2,91	0,05	0,13	0,13	0,43	0,04

Interprétation

Les contributions servent à détecter les individus qui contribuent le plus à la formation des axes factoriels.

Si tous les individus avaient la même contribution, celle-ci devrait osciller autour de 0,1% (100/1000). Les individus avec des contributions supérieures à 0,1% ont une influence supérieure à la moyenne. Dans cet exemple, 818 individus présentent une contribution supérieure à 0,1% sur le premier axe factoriel.

Dans le cas d'un grand nombre d'individus, il est préférable de visualiser la représentation graphique des individus actifs (avec une taille proportionnelle à la contribution) sur le plan factoriel 1-2 pour identifier les individus qui contribuent le plus à la création de ce plan.

Tous les points actifs interviennent dans la fabrication d'un axe. On vérifie donc que l'addition de toutes les contributions sur un axe est 100.

$$\sum_{i=1}^n CTR(i, \alpha) = 100$$

11.3 Cosinus Carrés des individus actifs

Le tableau ci-après donne les cosinus carrés des individus actifs pour les cinq premiers axes factoriels.

Ce tableau se compose de 8 colonnes :

- **Identificateur** : identifiant de l'individu
- **Poids relatif** : indique le poids donné à chaque individu, dans ce cas uniforme (0.1=100/1000).
- **Distance à l'origine** : carré de la distance de l'individu au centre de gravité.
- **Axe 1 – Axe 5** : cosinus carrés, utilisables pour apprécier la qualité du positionnement des points en représentation factorielle comparé à leur configuration réelle. Un cosinus carré égal à 1 indique que l'individu se trouve située sur l'axe. Un cosinus carré égal à 0 indique que l'individu est dans une direction orthogonale à l'axe.

Cosinus carrés des individus actifs

Identificateur	Poids relatif	Distance à l'origine	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5
1	0,10	2,44	0,03172	0,00	0,04	0,00	0,07
2	0,10	2,45	0,14496	0,18	0,21	0,01	0,03
3	0,10	3,72	0,08904	0,50	0,02	0,03	0,01
4	0,10	3,21	0,41780	0,14	0,11	0,00	0,10
5	0,10	2,23	0,28092	0,02	0,39	0,02	0,06
6	0,10	2,19	0,40413	0,01	0,21	0,00	0,13
7	0,10	2,41	0,23061	0,16	0,09	0,01	0,07
8	0,10	1,85	0,18559	0,09	0,00	0,12	0,08
9	0,10	1,98	0,00001	0,02	0,07	0,00	0,02
10	0,10	2,58	0,03118	0,18	0,01	0,01	0,21
11	0,10	1,97	0,07844	0,17	0,00	0,15	0,40
12	0,10	1,69	0,00355	0,01	0,00	0,13	0,06
13	0,10	2,43	0,16612	0,01	0,00	0,19	0,02
14	0,10	2,58	0,05573	0,00	0,00	0,02	0,03
15	0,10	2,10	0,00134	0,18	0,16	0,01	0,10
16	0,10	1,95	0,02114	0,13	0,16	0,20	0,07
17	0,10	3,52	0,00361	0,09	0,01	0,02	0,00
18	0,10	3,96	0,00126	0,07	0,04	0,01	0,02
19	0,10	1,51	0,00802	0,09	0,01	0,16	0,30

Interprétation

Les images obtenues (plans factoriels) sont des approximations de la configuration réelle. Il y aura des distances entre couples de points bien représentées, tandis que d'autres ne reflèteront pas fidèlement la distance réelle entre les points.

Si deux points sont proches du plan factoriel, alors la distance représentée sera une bonne approximation à la distance réelle. Mais si au moins un point est éloigné du plan de projection, alors la distance réelle peut être différente de celle représentée sur le plan.

Cette proximité du plan factoriel de projection est mesurée par les cosinus carrés de chaque point avec les axes factoriels.

$$\cos^2(i, \alpha) = \frac{\psi_{i\alpha}^2}{d^2(i, G)}$$

Un cosinus carré égal à 1 indique que l'individu se trouve sur l'axe (l'angle ω est nul). Un cosinus carré égal à 0 indique que l'individu est dans une direction orthogonale à l'axe.

En additionnant les cosinus carrés d'un individu sur tous les axes factoriels, nous obtenons l'unité, car il faut utiliser tous les axes factoriels pour positionner exactement le point dans l'espace complet.

$$\sum_{\alpha=1}^p \cos^2(i, \alpha) = 1$$

L'addition des cosinus carrés d'un point sur différents axes donne, en pourcentage, la "qualité" de la représentation du point sur le sous-espace défini par ces axes.

Dans le cas d'un grand nombre d'individus, il est préférable de visualiser la représentation graphique des individus actifs (avec une taille proportionnelle aux cosinus carrés) sur le plan factoriel 1-2 pour identifier les individus qui contribuent le plus à la création de ce plan.

12. CORMU-11 : Coordonnées, Cosinus Carrés des individus illustratifs

Dans cet exemple, nous n'avons sélectionné aucun individu illustratif. Les coordonnées et les cosinus carrés des individus illustratifs se présentent et s'interprètent de la même manière que pour les individus actifs.

Par définition, les individus illustratifs ne contribuent pas à la création des axes, c'est pourquoi le tableau des contribution n'apparaît pas dans ce listing.

13. CORMU-12 : Corrélations entre les variables continues et les facteurs

Ce tableau donne la corrélation entre les variables continues illustratives et les différents axes factoriels. Ce tableau est constitué de 10 colonnes :

- **Libellé des variables** : cette colonne donne le libellé de la variable continue.
- **Effectif** : effectif des individus renseignés.
- **Poids absolu** : effectif des individus renseignés après pondération. L'effectif et le poids sont identiques lorsque l'on n'utilise pas de pondération.
- **Moyenne** : cette colonne donne la moyenne pondérée de chacune des variables.
- **Ecart-type** : cette colonne donne l'écart-type pondéré de chacune des variables.
- **Axe1 – Axe5** : ces colonnes donnent les corrélation de chaque variable continue avec les 5 premiers axes factoriels.

Corrélations entre les variables continues et les facteurs

Libellé des variables	Effectif	Poids absolu	Moyenne	Ecart-type	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5
Estimation du salaire mensuel d'un ingénieur	806	806,00	8478,73	3668,95	0,04	-0,06	-0,04	0,01	-0,05
Estimation du revenu mensuel d'un médecin	713	713,00	19383,90	12608,80	0,05	-0,12	-0,02	-0,03	0,06
Age de l'enquêté(e)	1000	1000,00	42,68	17,50	0,40	-0,55	0,14	0,21	-0,28
Nombre de non-réponses au questionnaire	1000	1000,00	4,05	4,19	0,20	-0,12	0,20	0,08	-0,12
Age de fin d'étude	997	997,00	17,29	3,88	-0,69	-0,13	-0,24	-0,05	0,11
Revenu personnel souhaité	915	915,00	7244,48	4756,78	-0,26	-0,21	-0,15	-0,03	0,09
Estimation du revenu minimum d'une famille de 2 enfants	897	897,00	5561,89	2423,40	-0,19	0,01	-0,14	0,08	-0,14
Nombre de jours de vacances en été	1000	1000,00	18,31	19,37	-0,38	-0,02	-0,03	0,06	0,07

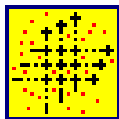
Interprétation

On note par exemple que la variable « Age de fin d'études » est corrélée négativement et de façon importante avec le premier axe (-0.69).

Les coefficients de corrélation des variables sont aussi les coordonnées de ces variables sur les différents axes factoriels.

Il est important de noter que les carrés des coefficients de corrélations sont des contributions relatives (cosinus carrés). Ils ne sont pas imprimés dans ces tableaux.

Les variables continues qui sont les plus corrélées aux axes factoriels sont donc aussi les mieux représentées.



Analyse des correspondances multiples avec choix des modalités actives (COREM)

Cette méthode est quasiment identique à l'analyse des correspondances multiples. L'unique différence réside dans le choix des modalités actives.

La méthode COREM est une alternative à la ventilation automatique des modalités d'effectif faible utilisée notamment dans l'analyse des correspondances multiples. Avec COREM, l'utilisateur sélectionne lui-même les modalités actives de son analyse.

La ventilation permet de s'affranchir automatiquement (en spécifiant un seuil en %) des modalités de faibles effectifs qui peuvent avoir des effets perturbateurs sur l'analyse. Les réponses appartenant à ces modalités peu fréquentes sont réparties aléatoirement entre les autres modalités de la variable.

Dans la méthode COREM, l'utilisateur définit lui-même les modalités à ventiler en ne les sélectionnant pas comme actives. Ces modalités qui ne sont pas sélectionnées comme actives sont automatiquement positionnées comme illustratives. Les individus concernés par ces modalités illustratives se voient affecter aléatoirement une autre modalité de la même variable. C'est cette technique qui est utilisée pour la ventilation de modalités d'effectifs faibles.

En termes de résultats, les sorties de cette méthode sont strictement identiques à celles de l'analyse des correspondances multiples, présentée dans le chapitre précédent.



Description des axes factoriels (DEFAC)

Cette procédure constitue une aide à l'interprétation des facteurs issus d'une procédure d'analyse factorielle.

Un facteur (i.e. axe factoriel) peut ainsi être rapidement et clairement décrit par les éléments les plus significatifs. Ces éléments pourront être les individus, les modalités, les variables continues, les fréquences, utilisés en tant qu'éléments actifs ou illustratifs dans l'analyse préalable.

Les éléments caractéristiques sont classés selon leurs coordonnées. Les modalités peuvent être classées selon le critère statistique appelé "valeur-test".

La description de chaque facteur est faite dans un document indépendant. Dans cet exemple, nous avons demandé à visualiser au plus 20 éléments caractéristiques par description : pour les individus actifs, les modalités actives et illustratives ainsi que les variables continues.

Nous effectuons la description des axes issus de l'analyse des correspondances multiples réalisée sur le fichier ASPI1000.SBA (CF page 74).

1. DEFAC-1 : Description de l'axe par les individus actifs et illustratifs

Cette sortie présente dans un premier temps les individus actifs de l'analyse, ordonnés dans l'ordre croissant de leur coordonnée factorielle sur l'axe.

Dans le cas d'un grand nombre d'individus, il est préférable de modifier le critère de sélection des éléments pour la description des axes (par défaut 50%). Dans le cas présent, nous avons demandé la sélection de 20 éléments.

Pour les individus actifs et illustratifs, le tableau se décompose en trois colonnes :

- **Identificateur de l'individu** : identifiant de l'individu.
- **Coordonnées** : coordonnée factorielle de l'individu sur l'axe étudié. Dans cet exemple, il s'agit du premier axe factoriel de l'analyse des correspondances multiples. Les individus sont triés dans l'ordre croissant de leur coordonnée factorielle. La ligne du tableau intitulée « Zone Centrale » permet de délimiter les coordonnées négatives des coordonnées positives.
- **Poids** : poids des individus dans l'analyse. Dans ce cas uniforme, tous les individus ont un poids identique : l'unité.

Description de l'axe 1
Par les INDIVIDUS ACTIFS

Identificateur de l'individu	Coordonnée	Poids
4	-1,16	1,00
188	-1,13	1,00
336	-1,09	1,00
297	-1,08	1,00
158	-1,08	1,00
24	-1,07	1,00
200	-1,07	1,00
187	-1,07	1,00
284	-1,07	1,00
334	-1,05	1,00
ZONE CENTRALE		
697	0,95	1,00
674	0,98	1,00
775	1,04	1,00
871	1,04	1,00
858	1,05	1,00
753	1,07	1,00
787	1,09	1,00
952	1,12	1,00
841	1,12	1,00
750	1,12	1,00

Interprétation

Dans cet exemple, l'identificateur des individus correspond à un numéro de questionnaire. Il est donc difficile d'interpréter directement ces résultats.

2. DEFAC-2 : Description de l'axe par les modalités actives et illustratives

Cette sortie présente dans un premier temps les modalités actives de l'analyse. Par défaut, ces modalités sont ordonnées dans l'ordre croissant de leur valeur-test sur l'axe mais il est possible de les ordonner dans l'ordre croissant de leur coordonnée factorielle en modifiant le paramétrage.

Pour les modalités actives et illustratives, le tableau se décompose en 4 colonnes :

- **Libellé de la variable** : libellé complet de la variable.
- **Libellé de la modalité** : libellé complet de la modalité.
- **Valeur-test** : valeur-test de la modalité liée à sa coordonnée sur l'axe factoriel. La valeur-test est d'autant plus forte que la modalité correspondante occupe une position significative sur l'axe. Une modalité est d'autant plus intéressante sur un axe que sa valeur-test est plus grande.

Dans le cas usuel, on considère qu'une valeur-test supérieure à 2 en valeur absolue indique qu'une modalité est significativement différente du centre de gravité. Les modalités sont ordonnées dans l'ordre croissant des valeurs-tests. La ligne du tableau intitulée « Zone Centrale » permet de délimiter les valeurs-tests négatives des valeurs-tests positives.

- **Poids** : poids des modalités dans l'analyse.

Description de l'axe 1

Par les MODALITES ACTIVES

Libellé de la variable	Libellé de la modalité	Valeur-Test	Poids
Diplôme de l'enquêté(e) en 5 classes	Université,gde école	-16,38	150,000
Type d'emploi	Cadres	-14,64	229,000
Diplôme de l'enquêté(e) en 5 classes	Bac - Brevet sup.	-13,86	182,000
Taille d'agglomération (en nombre d'habitants)	Paris	-13,22	326,000
Type d'emploi	*Reponse manquante*	-11,44	125,000
Age de l'enquêté(e) en 5 classes	Moins de 25 ans	-10,73	150,000
Statut d'occupation du logement en 4 classes	locataire	-8,98	523,000
Sexe de la personne interrogée	féminin	-8,62	531,000
Possédez vous des valeurs mobilières ?	oui	-8,14	121,000
Age de l'enquêté(e) en 5 classes	25 à 34 ans	-7,03	284,000
Z O N E C E N T R A L E			
Age de l'enquêté(e) en 5 classes	35 à 49 ans	5,34	209,000
Taille d'agglomération (en nombre d'habitants)	2.000 - 20.000	5,40	87,000
Age de l'enquêté(e) en 5 classes	50 à 64 ans	7,74	188,000
Possédez vous des valeurs mobilières ?	non	8,14	879,000
Sexe de la personne interrogée	masculin	8,62	469,000
Statut d'occupation du logement en 4 classes	propriétaire	8,88	290,000
Taille d'agglomération (en nombre d'habitants)	moins de 2.000	10,11	83,000
Diplôme de l'enquêté(e) en 5 classes	Aucun	10,75	189,000
Type d'emploi	Ouvriers	16,14	263,000
Diplôme de l'enquêté(e) en 5 classes	CEP ou fin études	17,40	321,000

Interprétation

Sur l'axe 1, on note la présence de valeurs-tests extrêmement élevées en valeur absolue, indiquant la présence de modalités significativement différentes du centre de gravité.

On remarque une nette opposition au niveau des diplômes, de l'emploi et de la taille d'agglomération. On observe ensuite une opposition moins marquée au niveau des classes d'âges, du statut d'occupation du logement, du sexe et des valeurs mobilières.

Cet axe oppose des personnes plus jeunes (moins de 25 ans jusqu'à 34 ans), assez bien diplômées (Bac, Brevet sup., Université gde école) où les femmes sont plus représentées à des personnes plus âgées (de 35 à 64 ans) et beaucoup moins diplômées (aucun diplôme, CEP ou fin d'études).

Dans le même sens, on constate une opposition entre Paris où le statut de locataire est prépondérant, et des communes de moins de 20 000 habitants dans lesquelles on retrouve davantage de propriétaires.

On observe également une opposition entre les cadres qui semblent posséder des valeurs mobilières et les ouvriers qui n'en possèdent pas ou moins.

Nous présentons maintenant la description de l'axe 1 par les 20 modalités illustratives les plus caractéristiques.

Description de l'axe 1

Par les MODALITES ILLUSTRATIVES

Libellé de la variable	Libellé de la modalité	Valeur-Test	Poids
Diplôme d'enseignement général le plus élevé obtenu	université,gde école	-15,80	142,000
La famille est le seul endroit où l'on se sente bien	non	-14,56	431,000
Diplôme d'enseignement général le plus élevé obtenu	baccalauréat (1/2)	-13,22	162,000
Profession de l'enquêté(e) en 7 classes	*Reponse manquante*	-11,44	125,000
Profession de l'enquêté(e) (ou dernière exercée)	*Reponse manquante*	-11,44	125,000
Profession de l'enquêté(e) en 7 classes	prof. lib.-cad. sup.	-10,13	84,000
Profession de l'enquêté(e) (ou dernière exercée)	cadre supérieur	-9,23	69,000
Profession de l'enquêté(e) (ou dernière exercée)	cadre moyen	-9,13	135,000
Statut d'occupation du logement	locataire	-8,98	523,000
Opinion à propos du mariage	dissout si accord	-8,71	387,000
Z O N E C E N T R A L E			
Opinion à propos du mariage	union indissoluble	7,94	231,000
Age et sexe de l'enquêteur	femme plus 38 ans	8,77	338,000
Statut d'occupation du logement	propriétaire	8,88	290,000
Profession de l'enquêté(e) (ou dernière exercée)	ouvrier spécialisé	9,05	98,000
Diplôme d'enseignement général le plus élevé obtenu	aucun	10,75	189,000
Profession de l'enquêté(e) (ou dernière exercée)	ouvrier qualifié	11,31	152,000
Regardez-vous la télévision ...	tous les jours	12,14	419,000
La famille est le seul endroit où l'on se sente bien	oui	14,47	561,000
Profession de l'enquêté(e) en 7 classes	ouvriers	16,14	263,000
Diplôme d'enseignement général le plus élevé obtenu	CEP ou fin études	17,40	321,000

On retrouve les variables initiales pour la profession et le diplôme d'enseignement le plus élevé. Sur ces variables, on constate les mêmes oppositions que pour les modalités actives.

On note tout de même une différence d'opinion sur le mariage. Concernant « les plus âgés et les moins diplômés », on note qu'ils sont plus nombreux à regarder la télévision tous les jours et qu'ils pensent en proportion supérieure à la moyenne que la famille est le seul endroit où ils se sentent bien.

3. DEFAC-3 : Description de l'axe par les variables continues illustratives

Ce tableau ordonne les variables continues illustratives dans l'ordre de leur coordonnée factorielle croissante sur chaque axe. Ce tableau est constitué de 3 groupes de colonnes :

- **Libellé des variables** : cette colonne donne le libellé de la variable continue.
- **Coordonnée** : coordonnée de la variable continue sur l'axe étudié. La coordonnée équivaut en fait à la corrélation entre la variable continue illustrative et l'axe factoriel (CF page 90).
- **Poids absolu** : effectif des individus renseignés après pondération. L'effectif et le poids sont identiques lorsque l'on n'utilise pas de pondération.
- **Moyenne** : cette colonne donne la moyenne pondérée de chacune des variables.
- **Ecart-type** : cette colonne donne l'écart-type pondéré de chacune des variables.

Description de l'axe 1

Par les VARIABLES CONTINUES ILLUSTRATIVES

Libellé de la variable	Coordonnée	Poids	Moyenne	Ecart-type
Age de fin d'étude	-0,69	997,00	17,286	3,883
Nombre de jours de vacances en été	-0,38	1000,00	18,311	19,367
Revenu personnel souhaité	-0,26	915,00	7244,480	4756,780
Estimation du revenu minimum d'une famille de 2 enfants	-0,19	897,00	5561,890	2423,400
ZONE CENTRALE				
Estimation du salaire mensuel d'un ingénieur	0,04	806,00	8478,730	3668,950
Estimation du revenu mensuel d'un médecin	0,05	713,00	19383,900	12608,800
Nombre de non-réponses au questionnaire	0,20	1000,00	4,054	4,190
Age de l'enquêté(e)	0,40	1000,00	42,680	17,496

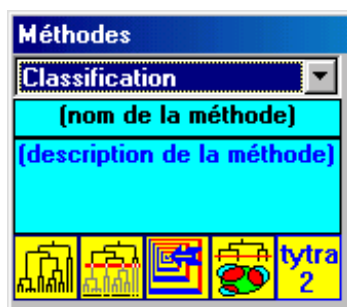
Interprétation

Ces résultats nous confortent dans l'analyse de la description de l'axe 1 par les modalités actives.

Les fortes valeurs pour l'âge de fin d'études, le nombre de jours de vacances en été et le revenu personnel souhaité concernent davantage les jeunes, les cadres, les plus diplômés, habitant Paris avec un statut de locataire, etc.

Les fortes valeurs pour l'âge concernent les plus âgés et les moins diplômés.

Classification et Typologie



Les techniques de classification font partie de la statistique exploratoire multidimensionnelle. Elles ont pour but d'expliciter la structure d'un ensemble de données importantes, permettant ainsi de formuler des hypothèses à vérifier dans une étape ultérieure. Elles sont à distinguer des méthodes de classement qui ont un but explicatif ou prédictif.

La typologie se sert de la classification pour fournir une autre forme de synthèse des données qu'une analyse factorielle. Cette synthèse de l'information contenue dans le tableau de données est présentée de la façon suivante. Les individus sont regroupés de façon automatique en un petit nombre de classes. Les individus qui se ressemblent au niveau des variables actives sont rassemblés dans une même classe. Les classes sont calculées pour que, lorsque l'on passe d'une classe à une autre, on passe d'une catégorie particulière d'individus à une catégorie différente.

La synthèse de toute l'information contenue dans le tableau de données se ramène alors à la caractérisation de ce petit nombre de classes homogènes. On aura ainsi fabriqué et typé des classes d'individus. La description en clair de ces types résume l'essentiel de l'information contenue dans les données.

Dans ce chapitre, nous réalisons une typologie à partir des résultats issus de l'analyse des correspondances multiples réalisée précédemment (CF page 74).

Pour cela, on va procéder à une classification. Il s'agit d'une classification sur facteurs. On se sert donc des facteurs issus de l'analyse des correspondances multiples.

Pour cette première étape, on utilise la procédure RECIP/SEMIS. Cette procédure permet de choisir les niveaux de coupure de l'arbre hiérarchique (aussi appelé dendrogramme). Il s'agit de choisir une ou plusieurs partitions au vu des résultats fournis en sortie.

La seconde étape, effectuée par la procédure PARTI-DECLA, permet d'obtenir une ou plusieurs partitions des individus. Il est possible de réaliser plusieurs partitions simultanément. Cette procédure propose aussi la description statistique des partitions choisies. On peut caractériser soit chacune des classes d'une partition, soit globalement l'ensemble de la partition. Les modalités des variables nominales, les variables nominales elles-mêmes, les variables continues et les axes factoriels peuvent intervenir dans les caractérisations statistiques.



Classification sur facteurs (RECIP / SEMIS)

1. Justification du passage aux coordonnées factorielles

La méthode RECIP/SEMIS de SPAD permet d'effectuer une classification à partir de coordonnées factorielles issues d'une analyse préalable.

Il est équivalent d'effectuer une classification des individus à partir d'un ensemble de p variables ou à partir de l'ensemble des p facteurs issus de l'analyse factorielle. En effet, en passant des variables initiales aux facteurs, sans en réduire leur nombre et ce, malgré leur obtention dans l'ordre décroissant de la variance expliquée, on ne perd aucune information. Il s'agit mathématiquement d'un changement de repère des individus (changement de base).

On peut, néanmoins, ne prendre en compte qu'un sous-espace factoriel de dimension q avec q inférieur à p , et effectuer une classification sur les q premiers axes factoriels. Cela présente l'avantage d'éliminer des fluctuations aléatoires qui constituent en général l'essentiel de la variance prise en compte par les $(p-q)$ derniers axes.

Le fait d'abandonner les derniers facteurs revient à « lisser » les données, ce qui en général améliore la partition en produisant des classes plus homogènes.

Les axes factoriels qui sont à conserver pour la classification sont ceux qui engendrent un sous-espace dans lequel le nuage des individus à classer est stable. En pratique on garde généralement un peu plus de la moitié des axes, même si un « coude » apparaît au bout de quelques axes à l'examen de l'histogramme des valeurs propres associées à ces axes.

Dans le paramétrage de cette méthode, vous pouvez définir le nombre de coordonnées factorielles à prendre en compte pour l'agrégation (10 par défaut).

Ainsi quelque soit le tableau de données initiales, on se ramènera toujours à un tableau de données quantitatives à partir duquel sera effectuée la classification des individus. Une seule distance, *la distance euclidienne usuelle*, sera utilisée pour calculer les ressemblances entre individus, et un seul critère d'agrégation, *la perte d'inertie minimum* (critère de Ward) sera utilisé pour calculer l'écart entre deux sous-ensembles disjoints.

2. Les techniques de classification

Les techniques proposées dans SPAD sont la classification ascendante hiérarchique (CAH, RECIP dans SPAD) qui fournit une hiérarchie de partitions, et la méthode d'agrégation autour de centres mobiles qui conduit directement à une seule partition.

Une utilisation conjointe de ces deux types de méthodes (classification mixte) permettra de consolider la partition et d'obtenir une **partition fiable** sinon optimale (SEMIS).

Les deux types de méthode – CAH et centres mobiles – présentent les inconvénients respectifs suivants :

- la CAH fournit un grand nombre de partitions parmi lesquelles on doit en choisir une : il n'est souvent pas aisé de choisir la coupure significative. D'autre part, l'arbre hiérarchique obtenu n'est pas un arbre optimal puisque la partition construite à un niveau donné dépend de la partition obtenue à l'étape précédente.
- dans la méthode des centres mobiles, le nombre de classes doit être fixé au départ, et la partition obtenue dépend du tirage initial des centres provisoires des classes.

Pour remédier en partie à ces inconvénients et pour essayer de s'approcher le plus possible de la partition optimale si elle existe, on peut avoir recours à l'utilisation conjointe de la CAH et de la CCM : c'est l'objet de la classification mixte appelée SEMIS dans SPAD.

Une première utilisation conjointe des deux techniques de classification est la suivante : on effectue une classification (CCM) autour d'un nombre important de centres mobiles et on construit ensuite un arbre hiérarchique à partir des classes formées dans cette CCM.

Cependant, cette méthode est relativement instable sur des échantillons de petite taille. Nous vous conseillons d'utiliser la procédure RECIP (CAH) sur des échantillons de moins de 10000 individus. Au delà, la méthode SEMIS permet de réduire les temps d'exécution et fournit des partitions stables.

Nous présentons dans ce chapitre les méthodes RECIP et SEMIS à partir des données issues de l'analyse des correspondances multiples présentées page 74.

3. La Classification Ascendante Hiérarchique – RECIP

La procédure RECIP agrège les éléments suivant un algorithme de classification hiérarchique utilisant le critère d'agrégation de Ward (critère de la variance). Elle agrège les éléments de façon à minimiser la variance interne de chaque classe (inerties intra-classes) et à maximiser la variance entre les classes (inertie inter-classes). Ces inerties sont calculées à partir des coordonnées des éléments à classer sur les axes factoriels sauvegardés pour la typologie. Pour plus de précisions, nous renvoyons à l'ouvrage [Lebart, Morineau et Piron (1995)].

Coordonnées utilisées pour l'agrégation

Ce paramètre indique le nombre de facteurs pris en compte pour calculer les distances entre les individus et pour effectuer la partition en classes. Par la suite, la construction des partitions et des classes ainsi que le calcul des différentes aides à l'interprétation seront effectués dans le sous-espace factoriel choisi. Par défaut, 10 facteurs sont pris en compte. Il faut noter que le nombre maximum d'axes possibles est celui choisi pour l'analyse factorielle.

C'est au moment du choix de ce paramètre que l'utilisateur devra se référer aux résultats de l'analyse factorielle et notamment la présentation des valeurs propres.

Dans l'exemple, la partition en classes s'est faite sur les 14 premiers facteurs. On a choisi ici le nombre d'axes factoriels de manière à ce que au moins 80% de l'inertie totale soit conservée (CF page 78).

Il n'y a pas de règle simple pour le choix du nombre d'axes. On peut conseiller en général de conserver au moins la moitié des axes et souvent les deux tiers.

Sauvegarde partielle de l'arbre

La sauvegarde partielle de l'arbre fixe le nombre d'éléments terminaux qui seront sauvegardés pour l'édition des différents résultats (50 par défaut). Seule la partie de l'arbre au-dessus de cette coupure sera conservée. Le nombre d'individus actifs limite le nombre d'éléments terminaux qui peuvent être conservés.

Dans le cas où l'on sauvegarde trop d'éléments terminaux, le logiciel n'édite pas la composition de ces éléments, pas plus que leurs coordonnées et leur valeur-test. Ici, on ne conserve que 30 éléments terminaux.

3.1 CLASSIF-3 : Description des 30 noeuds d'indices les plus élevés

La description des noeuds est constituée de 6 composantes.

- **Numéro** : donne les numéros des noeuds. Ces numéros sont attribués en séquence.
- **Aîné** : un des deux noeuds terminaux agrégés (avec le benjamin) à une étape donnée pour former le noeud Numéro.
- **Benjamin** : noeud terminal agrégé avec l'aîné pour former le nouveau noeud identifié par son numéro.

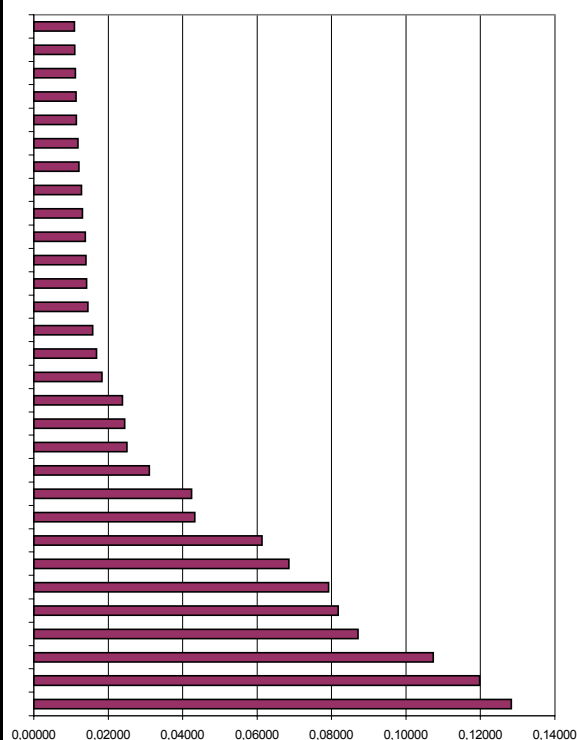
L'aîné et le benjamin sont les deux éléments dont le regroupement entraîne une perte d'inertie inter-classes minimale.

- **Nb d'éléments terminaux du noeud** : donne le nombre d'éléments terminaux dans le noeud créé. Equivaut à la somme des éléments terminaux contenus dans l'aîné et le benjamin.
- **Poids du noeud** : poids du noeud qui dépend de celui de chaque élément terminal. Dans ce cas uniforme, le poids et le nombre d'éléments terminaux sont identiques.
- **Indice de niveau** : donne la valeur de l'indice d'agrégation (ou indice de niveau) de chaque noeud. Cet indice d'agrégation au sens du critère de Ward mesure la dissimilarité entre les classes. L'indice de niveau d'un noeud est en fait la perte d'inertie inter-classes occasionnée par la formation du noeud. Ce noeud est formé par le regroupement des deux noeuds dont les numéros sont donnés par les colonnes Aîné et Benjamin.

Par défaut, les 50 premiers noeuds sont édités. Ici, la publication des 30 premiers est suffisante.

Description des 30 noeuds d'indices les plus élevés

Numéro	Aîné	Benjamin	Nb d'éléments terminaux du noeud	Poids du noeud	Indice de niveau
1970	1936	1938	78	78,00	0,01097
1971	1946	1953	60	60,00	0,01101
1972	1964	1943	50	50,00	0,01116
1973	1928	1951	52	52,00	0,01136
1974	1940	1958	49	49,00	0,01139
1975	1969	1966	79	79,00	0,01188
1976	1968	1952	84	84,00	0,01213
1977	1944	1954	49	49,00	0,01285
1978	1965	1878	68	68,00	0,01304
1979	1923	1904	42	42,00	0,01387
1980	1930	1901	53	53,00	0,01397
1981	1972	1920	63	63,00	0,01417
1982	1963	1947	101	101,00	0,01450
1983	1967	1959	73	73,00	0,01582
1984	1979	1955	76	76,00	0,01687
1985	1935	1950	54	54,00	0,01835
1986	1974	1962	110	110,00	0,02384
1987	1980	1970	131	131,00	0,02446
1988	1971	1985	114	114,00	0,02504
1989	1982	1976	185	185,00	0,03104
1990	1973	1984	128	128,00	0,04241
1991	1983	1987	204	204,00	0,04329
1992	1977	1986	159	159,00	0,06131
1993	1989	1990	313	313,00	0,06857
1994	1993	1992	472	472,00	0,07922
1995	1978	1981	131	131,00	0,08172
1996	1988	1994	586	586,00	0,08705
1997	1975	1996	665	665,00	0,10724
1998	1995	1997	796	796,00	0,11979
1999	1991	1998	1000	1000,00	0,12828
Somme des indices de niveau					1,83681



Le nœud 1970, par exemple, est formé des nœuds 1936 et 1938. Il est constitué de 78 éléments de poids 78. La valeur de l'indice d'agrégation est de 0.01097.

On observe sur l'histogramme des indices de niveau des sauts importants pour une coupure en quatre classes (trois dernières barres) et en neuf classes (huit dernières barres). On optera pour une partition en neuf classes.

3.2 CLASSIF-4 : Description des 30 nouveaux éléments terminaux

Le tableau ci-après édite la composition des 30 éléments terminaux (50 par défaut). Cela permet de connaître la répartition des individus dans les éléments terminaux qui se définissent comme les classes d'individus obtenues lorsque l'on coupe l'arbre d'agrégation hiérarchique en un certain niveau.

Ce tableau est constitué de 7 colonnes :

Numéro : rang de l'élément terminal. Celui-ci est attribué séquentiellement.

Noeud : numéro du nœud terminal.

Indice : valeur de l'indice de niveau du nœud terminal.

Aîné : premier nœud qui compose le nœud terminal.

Benjamin : second nœud qui compose le nœud terminal.

Effectif : effectif du nœud terminal.

Identificateur : donne l'identificateur de l'ensemble des individus qui sont agrégés dans le nœud terminal.

Dans le cas de cette enquête, les individus sont anonymes et l'édition de la composition des éléments terminaux a peu d'intérêt. On se limite à la publication de 5 nœuds parmi les nouveaux éléments terminaux formés par l'agrégation.

Description des 30 nouveaux éléments terminaux
Eléments terminaux formés de plusieurs individus

Numéro	Noeud	Indice	Aîné	Benjamin	Effectif	Identificateur	Identificateur	Identificateur	Identificateur	Identificateur	Identificateur	Identificateur	Identificateur
1	1962	0,00888	1915	1948	61	1	90	201	36	249	14	234	185
						230	233	383	859	607	992	822	837
						779	857	590	271	51	91	290	250
						209	327	298	229	47	177	329	206
						255	93	65	40	11	39	723	722
						384	618	845	539	897	368	619	485
						892	980	405	114	232	238	184	306
						276	269	252	15	248			
2	1958	0,00754	1933	1891	27	254	647	599	805	851	896	270	911
						890	500	724	366	616	409	693	929
						694	883	676	698	968	704	281	408
						930	552	555					
3	1940	0,00476	1868	1875	22	548	549	739	901	696	902	951	683
						707	277	759	962	933	936	964	538
						20	283	97	125	935	96		
4	1954	0,00650	1870	1911	21	133	107	512	160	214	461	612	418
						820	463	378	435	400	632	385	483
						335	334	285	879	563			
5	1944	0,00525	1850	1888	28	577	889	315	910	482	645	484	157
						131	313	49	212	167	396	818	908
						456	623	988	987	375	727	509	600
						397	535	833	855				

3.3 CLASSIF-5 : Coordonnées et valeurs-tests des noeuds terminaux

Cette sortie édite dans un premier temps le tableau des coordonnées des éléments terminaux sur les 5 premiers axes (issus de l'analyse des correspondances multiples), puis le tableau des valeurs-tests.

Le tableau des coordonnées est composé de 9 colonnes :

- **Numéro** : donne le rang du nœud
- **Identificateur** : numéro du nœud terminal,
- **Poids** : poids du nœud terminal.
- **Effectif** : effectif du nœud terminal.

Dans l'exemple, le poids des individus est uniforme, donc le poids et l'effectif des nœuds sont identiques.

- **Axe1 – Axe5** : valeurs-tests des nœuds terminaux sur les axes demandés par l'utilisateur (les cinq premiers par défaut).

Coordonnées des nœuds terminaux

Numéro	Identificateur	Poids	Effectif	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5
1	1962	61,00	61	-0,41	0,95	-0,20	-0,39	-1,15
2	1958	27,00	27	-0,51	1,07	0,88	0,17	0,09
3	1940	22,00	22	0,51	0,95	-0,15	0,98	0,30
4	1954	21,00	21	-0,58	-0,40	1,42	0,28	1,02
5	1944	28,00	28	-1,44	1,08	1,48	0,79	1,86
6	1955	34,00	34	0,37	0,58	0,17	-1,05	-0,73
7	1904	18,00	18	-0,35	-0,07	0,80	-0,67	1,17
8	1923	24,00	24	1,21	0,00	-0,85	-0,19	1,00
9	1951	29,00	29	-0,60	-0,12	-1,31	-0,70	1,24
10	1928	23,00	23	0,40	0,25	-0,64	-1,52	1,14
11	1952	24,00	24	0,64	1,05	-0,08	1,64	0,19
12	1968	60,00	60	0,65	0,57	-1,18	0,63	0,21
13	1947	43,00	43	0,33	0,78	0,02	-0,25	-0,32
14	1963	58,00	58	0,64	0,14	0,33	-0,06	-0,61
15	1950	29,00	29	0,85	-0,64	1,04	0,69	-0,46
16	1935	25,00	25	-0,24	-1,70	0,91	0,46	-0,53
17	1953	28,00	28	0,94	-0,11	-0,07	1,77	-0,48
18	1946	32,00	32	0,24	-0,23	0,19	0,93	-1,24
19	1966	31,00	31	0,13	-0,51	-0,75	-0,93	0,47
20	1969	48,00	48	1,03	-0,10	-0,15	-0,85	0,56
21	1920	13,00	13	-1,16	0,99	1,60	0,51	1,30
22	1943	18,00	18	0,66	0,56	0,06	1,16	0,05
23	1964	32,00	32	-0,50	-0,11	0,24	-0,23	-0,12
24	1878	12,00	12	0,60	0,01	0,64	-2,10	-0,73
25	1965	56,00	56	1,37	-0,95	1,16	-0,53	0,06
26	1970	78,00	78	-1,07	0,38	-0,32	-0,37	-0,38
27	1901	19,00	19	-1,69	-0,15	0,35	0,71	0,12
28	1930	34,00	34	-1,39	-0,80	-1,48	-0,02	-0,53
29	1959	35,00	35	-0,67	-1,34	-0,24	-0,13	0,80
30	1967	38,00	38	-1,00	-2,04	-0,66	0,75	-0,41

Le tableau des valeurs-tests comprend les colonnes suivantes :

- **Numéro** : donne le rang du nœud
- **Identificateur** : numéro du nœud terminal,
- **Poids** : poids du nœud terminal.
- **Effectif** : effectif du nœud terminal.

Dans l'exemple, le poids des individus est uniforme, donc le poids et l'effectif des nœuds sont identiques.

- **Axe1 – Axe5** : donne les coordonnées des nœuds terminaux sur les axes demandés par l'utilisateur (les cinq premiers par défaut).

Valeurs-Test des noeuds terminaux

Numéro	Identificateur	Poids	Effectif	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5
1	1962	61,00	61	-3,28	7,66	-1,64	-3,18	-9,27
2	1958	27,00	27	-2,67	5,62	4,64	0,89	0,48
3	1940	22,00	22	2,43	4,49	-0,73	4,64	1,45
4	1954	21,00	21	-2,67	-1,83	6,59	1,29	4,73
5	1944	28,00	28	-7,73	5,78	7,96	4,25	9,96
6	1955	34,00	34	2,20	3,42	1,03	-6,23	-4,34
7	1904	18,00	18	-1,51	-0,29	3,42	-2,85	5,02
8	1923	24,00	24	6,00	0,00	-4,21	-0,93	4,94
9	1951	29,00	29	-3,25	-0,66	-7,17	-3,80	6,76
10	1928	23,00	23	1,94	1,22	-3,11	-7,36	5,55
11	1952	24,00	24	3,18	5,20	-0,42	8,15	0,96
12	1968	60,00	60	5,18	4,55	-9,44	5,01	1,64
13	1947	43,00	43	2,24	5,26	0,11	-1,68	-2,13
14	1963	58,00	58	5,04	1,08	2,59	-0,45	-4,79
15	1950	29,00	29	4,62	-3,52	5,68	3,77	-2,53
16	1935	25,00	25	-1,22	-8,60	4,58	2,34	-2,68
17	1953	28,00	28	5,03	-0,62	-0,39	9,48	-2,57
18	1946	32,00	32	1,40	-1,34	1,11	5,35	-7,11
19	1966	31,00	31	0,71	-2,89	-4,26	-5,27	2,67
20	1969	48,00	48	7,33	-0,74	-1,04	-6,05	3,97
21	1920	13,00	13	-4,19	3,58	5,79	1,85	4,73
22	1943	18,00	18	2,83	2,39	0,27	4,96	0,21
23	1964	32,00	32	-2,88	-0,61	1,38	-1,31	-0,69
24	1878	12,00	12	2,09	0,05	2,21	-7,30	-2,56
25	1965	56,00	56	10,53	-7,34	8,92	-4,04	0,44
26	1970	78,00	78	-9,83	3,53	-2,96	-3,41	-3,46
27	1901	19,00	19	-7,43	-0,66	1,54	3,13	0,54
28	1930	34,00	34	-8,27	-4,76	-8,76	-0,11	-3,16
29	1959	35,00	35	-4,02	-8,09	-1,42	-0,79	4,81
30	1967	38,00	38	-6,26	-12,84	-4,15	4,69	-2,59

Cette sortie n'est pas éditée par défaut. Il permet de juger, à travers les coordonnées, de « l'originalité » des éléments terminaux sur les premiers axes factoriels, et à travers les valeurs-tests de la significativité de ces éléments terminaux ces axes.

Le tableau des coordonnées et des valeurs-tests des classes de la partition que l'on verra ultérieurement est souvent suffisant.

3.4 CLASSIF-6 : Description des noeuds de la hiérarchie (au dessus des noeuds terminaux)

Ce tableau donne les caractéristiques des nœuds situés au-dessus des nœuds terminaux. Il décrit de quelle façon s'agrègent ces nœuds. On a 8 colonnes.

- **Rang du nœud** : rang du nœud étudié au dessus des nœuds terminaux.
- **Indice du nœud** : indice de niveau du nœud étudié.

Dans l'exemple, on a sauvegardé les 30 derniers nœuds terminaux. Le premier nœud formé par l'agrégation de deux nœuds terminaux est donc de rang 31. Le second est de rang 32, etc.

- **Aîné** : rang du premier nœud qui compose le nœud étudié.
- **Benjamin** : rang du second nœud qui compose le nœud étudié.

Le nœud de rang 42 est formé des nœuds de rang 13 et de rang 14, qui sont donc deux des 30 nœuds terminaux.

- **Effectif** : effectif du nœud terminal.
- **Poids** : poids du nœud terminal.

Dans l'exemple, le poids des individus est uniforme, donc le poids et l'effectif des nœuds sont identiques.

- **Rang du premier nœud terminal** : donne le rang du premier nœud terminal compris dans le nœud étudié.

- **Rang du dernier nœud terminal** : donne le rang du dernier nœud terminal compris dans le nœud étudié.

Description des nœuds de la hiérarchie (au dessus des nœuds terminaux)
Indices en pourcentage de la somme des indices (1.12563)

Rang du nœud	Indice du nœud	Ainé	Benjamin	Effectif	Poids	Rang du premier nœud terminal	Rang du dernier nœud terminal
31	0,98	18	17	60	60,00	17	18
32	0,99	23	22	50	50,00	22	23
33	1,01	10	9	52	52,00	9	10
34	1,01	3	2	49	49,00	2	3
35	1,06	20	19	79	79,00	19	20
36	1,08	12	11	84	84,00	11	12
37	1,14	5	4	49	49,00	4	5
38	1,16	25	24	68	68,00	24	25
39	1,23	8	7	42	42,00	7	8
40	1,24	28	27	53	53,00	27	28
41	1,26	32	21	63	63,00	21	23
42	1,29	14	13	101	101,00	13	14
43	1,41	30	29	73	73,00	29	30
44	1,50	39	6	76	76,00	6	8
45	1,63	16	15	54	54,00	15	16
46	2,12	34	1	110	110,00	1	3
47	2,17	40	26	131	131,00	26	28
48	2,22	31	45	114	114,00	15	18
49	2,76	42	36	185	185,00	11	14
50	3,77	33	44	128	128,00	6	10
51	3,85	43	47	204	204,00	26	30
52	5,45	37	46	159	159,00	1	5
53	6,09	49	50	313	313,00	6	14
54	7,04	53	52	472	472,00	1	14
55	7,26	38	41	131	131,00	21	25
56	7,73	48	54	586	586,00	1	18
57	9,53	35	56	665	665,00	1	20
58	10,64	55	57	796	796,00	1	25
59	11,40	51	58	1000	1000,00	1	30

Ainsi, le nœud de rang 57 est le regroupement des nœuds terminaux de rang 1 à 20.

4. La Classification Mixte – SEMIS

Les algorithmes de classification sont plus ou moins bien adaptés au traitement d'un nombre important d'individus à classer. La méthode d'agrégation autour des centres mobiles offre des avantages puisqu'elle permet d'obtenir une partition sur un volume de données important. Elle présente cependant l'inconvénient de produire des partitions dépendant des premiers centres choisis et celui de fixer a priori le nombre de classes.

A l'inverse, la classification hiérarchique est une méthode que l'on peut qualifier de « déterministe » puisqu'elle donne toujours les mêmes résultats à partir des mêmes données. Cette méthode utilise des algorithmes qui donnent des indications sur le nombre de classes à retenir mais elle est mal adaptée aux ensembles volumineux de données.

La classification autour des centres mobiles peut aussi être utilisée en parallèle avec la classification hiérarchique (classification mixte). En fournissant des partitions sur un volume de données important, elle permet de réduire la dimension de l'ensemble des éléments à classer en effectuant des regroupements préalables. Ainsi, l'algorithme de classification qui paraît bien adapté au partitionnement d'un ensemble comprenant des milliers ou des dizaines de milliers d'individus ou davantage est un algorithme mixte. Il repose sur la combinaison des deux algorithmes : l'algorithme d'agrégation autour des centres mobiles et l'algorithme d'agrégation hiérarchique.

L'algorithme de classification mixte est effectué par la procédure SEMIS sur les individus décrits par leurs premières coordonnées factorielles.

Cet algorithme procède en trois étapes :

1. Partitionnement préliminaire

L'ensemble des éléments à classer subit tout d'abord un partitionnement préliminaire (autour des centres mobiles).

Cette première étape consiste à obtenir une partition des n objets en k classes homogènes, où k sera choisi largement plus élevé que le nombre réel de classes désirées, et largement plus petits que le nombre total d'objets à classer.

Pour ce partitionnement préliminaire en quelques dizaines ou quelques centaines de classes homogènes, on utilise l'algorithme d'agrégation autour des centres mobiles (ou nuées dynamiques). Cet algorithme augmente l'inertie entre les classes à chaque itération et produit une partition en un nombre de classes fixé au préalable. Ce nombre dépend du choix initial des centres qui sont tirés au hasard ou choisis par l'utilisateur.

L'optimum ne peut pas être atteint, mais la partition obtenue peut être améliorée à partir de groupements stables. Ces groupes d'individus qui apparaissent toujours dans les mêmes classes seront les éléments de base de l'étape suivante.

2. Agrégation hiérarchique des classes obtenues

La seconde étape vise à effectuer une classification ascendante hiérarchique où les éléments terminaux de l'arbre sont les k classes de la partition préliminaire. Certains de ces groupements peuvent être proches les uns des autres. Ils correspondent à un groupe « réel » qui aurait été coupé artificiellement par l'étape précédente.

L'étape précédente crée, en général, plusieurs petits groupes ne contenant parfois qu'un seul élément. L'objectif de l'étape d'agrégation hiérarchique est de reconstituer les classes qui ont été fragmentées et d'agréger des éléments apparemment dispersés autour de leur centre d'origine. L'arbre est construit en utilisant le critère de Ward. Cette étape est semblable à la classification hiérarchique présentée précédemment. Le dendrogramme suggérera éventuellement le nombre de classes finales à retenir.

3. Partition finale et consolidation

La partition finale de la population est définie par coupure de l'arbre de classification hiérarchique. L'homogénéité des classes obtenues peut être optimisée par réaffectations par la technique des centres mobiles (voir la consolidation de la partition).

Les deux premières étapes sont effectuées par la procédure SEMIS. La dernière est réalisée par les procédures PARTI-DECLA qui effectuent également la description statistique des classes obtenues.

4.1 Paramètres de fonctionnement de SEMIS

Les 4 paramètres de fonctionnement soumis à l'utilisateur sont présentés ci-après.

Coordonnées utilisées pour l'agrégation

Ce paramètre indique le nombre de facteurs pris en compte pour calculer les distances entre les individus et pour effectuer la partition autour des centres mobiles. Par la suite, la construction des partitions et des classes ainsi que le calcul des différentes aides à l'interprétation seront effectués dans le sous-espace factoriel choisi. Par défaut, 10 facteurs sont pris en compte. Il faut noter que le nombre maximum d'axes est choisi dans les

paramètres de fonctionnement de l'analyse. Dans l'exemple, la partition utilise les 14 premiers facteurs.

Partition(s) de base

Trois méthodes de classification sont disponibles.

La première méthode consiste à chercher les classes stables par croisement de plusieurs partitions de base construites autour d'individus tirés au hasard. L'item *Nombre* définit le nombre de partitions construites (2 par défaut) et l'item *Taille* détermine le nombre d'individus tirés au hasard pour chaque partition. Ce sont les centres initiaux de chacune des partitions.

Les deux autres méthodes consistent à construire une seule partition par l'algorithme des centres mobiles, autour de N centres choisis par l'utilisateur (bouton Choix) ou tirés au hasard dans l'ensemble de la population.

L'item *Nombre d'itérations pour la formation* indique le nombre d'itérations du processus à centres mobiles (ou de nuées dynamiques). Par défaut, ce nombre est de 7.

Groupelements stables à conserver

Ce paramètre détermine le nombre de classes stables à conserver et à utiliser pour la classification hiérarchique.

L'item *Tous* permet de retenir l'ensemble des classes stables.

L'item *Sélection par seuil de poids (en %)* permet de ne conserver que les classes stables dont le poids, en pourcentage, dépasse le seuil fixé par l'utilisateur (par défaut, le seuil est de 0%).

L'item *Les N plus lourds* sélectionne les N classes stables ayant les effectifs les plus élevés.

Création d'une classe résiduelle

Une fois déterminé le nombre de classes stables à conserver pour la classification, il reste à reclasser les individus qui n'appartiendraient pas aux classes retenues.

Si on choisit l'item *Non*, les individus sont affectés à la classe stable la plus proche. En sélectionnant l'item *Oui*, ils sont regroupés au sein d'une classe dite résiduelle.

Dans l'exemple, nous avons choisi de construire une partition de base sur 10 centres mobiles tirés au hasard et 2 partitions croisées de taille 10 chacune.

4.2 CLASSIF-1 : Classification mixte sur les 14 premiers axes factoriels

4.2.1 Résultats obtenus avec la partition de base sur 10 centres mobiles tirés au hasard

Le tableau suivant comporte 4 colonnes :

- **Numéro de la classe** : rang de chaque classe.
- **Numéro d'individu** : numéro du centre mobile de chaque classe, tiré au hasard ou choisi.
- **Poids** : indique le poids des individus de la classe.
- **Effectif** : effectif de la classe.

Le nombre de classes dépend de la taille (ou du nombre de centres mobiles) adoptée pour la partition (10 dans l'exemple).

Ces classes sont censées être des morceaux de classes « réelles » que l'algorithme d'agrégation autour des centres mobiles a éclatées.

Classification mixte sur les 14 premiers axes factoriels
Centres mobiles à partir de 10 individus-source.

Partition

Numéro de la classe	Numéro d'individu	Poids	Effectif
1	98	66,00	66
2	232	122,00	122
3	392	112,00	112
4	618	104,00	104
5	668	101,00	101
6	699	89,00	89
7	759	102,00	102
8	813	101,00	101
9	896	97,00	97
10	942	106,00	106

4.2.2 Résultats obtenus avec 2 partitions croisées de taille 10 chacune

Les deux tableaux suivants comportent 4 colonnes :

- **Numéro de la classe** : rang de chaque classe.
- **Numéro d'individu** : numéro du centre mobile de chaque classe, tiré au hasard ou choisi.
- **Poids** : indique le poids des individus de la classe.
- **Effectif** : effectif de la classe.

Classification mixte sur les 10 premiers axes factoriels
A partir de 2 partitions de base avant croisement

Partition 1

Numéro de la classe	Numéro d'individu	Poids	Effectif
1	98	66,00	66
2	232	122,00	122
3	392	112,00	112
4	618	104,00	104
5	668	101,00	101
6	699	89,00	89
7	759	102,00	102
8	813	101,00	101
9	896	97,00	97
10	942	106,00	106

Partition 2

Numéro de la classe	Numéro d'individu	Poids	Effectif
1	61	135,00	135
2	163	87,00	87
3	179	78,00	78
4	248	64,00	64
5	386	148,00	148
6	402	81,00	81
7	454	107,00	107
8	456	86,00	86
9	619	111,00	111
10	945	103,00	103

4.3 CLASSIF-2 : Résultats du croisement

Dans cette sortie, la procédure SEMIS fournit le tableau des résultats du croisement des partitions de base, avant et après des réaffectations éventuelles (à centres mobiles).

Les réaffectations consistent à améliorer l'homogénéité interne des classes en réaffectant les éléments d'une classe à une autre classe dont ils sont plus proches. Ce processus s'effectue par itérations successives à centres mobiles de façon analogue à la consolidation de la partition effectuée par la procédure PARTI-DECLA.

4.3.1 Résultats obtenus avec la partition de base sur 10 centres mobiles tirés au hasard

Le tableau est composé des colonnes suivantes :

Poids avant : indique le poids de la classe avant les réaffectations éventuelles.

Pourcentage cumulé : donne le pourcentage cumulé des poids des individus des classes avant les réaffectations.

Numéro de la classe : rang de la classe. Chaque classe est considérée comme un nœud ou un élément terminal.

Poids après : représente le poids des individus de la classe après les réaffectations éventuelles.

Résultats du croisement

Poids avant	Pourcentage cumulé	Numéro de la classe	Poids après
122,00	12,20	&01&	122,00
112,00	23,40	&02&	112,00
106,00	34,00	&03&	106,00
104,00	44,40	&04&	104,00
102,00	54,60	&05&	102,00
101,00	64,70	&06&	101,00
101,00	74,80	&07&	101,00
97,00	84,50	&08&	97,00
89,00	93,40	&09&	89,00
66,00	100,00	&10&	66,00
0,00	100,00	Classe résiduelle	

Dans le cas d'une partition de base unique autour de N centres mobiles (10 dans l'exemple), on retrouve les classes (que l'on considère stables) obtenues par l'algorithme d'agrégation autour des centres mobiles :

4.3.2 Résultats obtenus avec 2 partitions croisées de taille 10 chacune

Le tableau ci-après présente le croisement entre les partitions de base obtenues précédemment (2 dans l'exemple). On retient comme classes finales les classes stables constituées par les individus classés ensemble dans les partitions de base.

Le tableau est composé des colonnes suivantes :

Poids avant : indique le poids de la classe avant les réaffectations éventuelles.

Pourcentage cumulé : donne le pourcentage cumulé des poids des individus des classes avant les réaffectations.

Numéro de la classe : rang de la classe. Chaque classe est considérée comme un nœud ou un élément terminal.

Poids après : représente le poids des individus de la classe après les réaffectations éventuelles.

Une partie du tableau a été tronquée.

Résultats du croisement

Poids avant	Pourcentage cumulé	Numéro de la classe	Poids après
85,00	8,50	&01&	85,00
82,00	16,70	&02&	82,00
72,00	23,90	&03&	72,00
65,00	30,40	&04&	65,00
58,00	36,20	&05&	58,00
49,00	41,10	&06&	49,00
49,00	46,00	&07&	49,00
43,00	50,30	&08&	43,00
43,00	54,60	&09&	43,00
39,00	58,50	&10&	39,00
...			
1,00	99,60	&55&	1,00
1,00	99,70	&56&	1,00
1,00	99,80	&57&	1,00
1,00	99,90	&58&	1,00
1,00	100,00	&59&	1,00
0,00	100,00	Classe résiduelle	

Dans l'exemple, on remarque qu'à partir des 2 partitions de 10 classes, on obtient 59 classes stables. Ces classes seront ensuite agrégées par l'algorithme d'agrégation hiérarchique dont les résultats sont rassemblés dans le tableau de la description des nœuds d'indices les plus élevés. Les 10 classes obtenues dans l'autre exemple sont agrégées de la même manière.

La procédure SEMIS fournit ensuite respectivement dans les sorties **CLASSIF-3** et **CLASSIF-6** la description des nœuds d'indices les plus élevés puis la description des nœuds de la hiérarchie. Ces résultats sont analogues à ceux produits par la procédure RECIP.



Coupure de l'arbre et description des classes (PARTI/DECLA)

La procédure **PARTI-DECLA** permet de couper à un niveau convenable l'arbre obtenu par la procédure RECIP ou SEMIS pour construire une partition des individus. Il est possible de réaliser plusieurs partitions simultanément et de décrire statistiquement les partitions choisies.

Dans cet exemple, nous reprenons les résultats issus de la procédure RECIP présentée page 99. Le fichier utilisé est toujours ASPI1000.sba fourni avec le logiciel.

1. La procédure PARTI

La procédure **PARTI** construit des partitions par coupure de l'arbre d'agrégation obtenu précédemment avec la procédure RECIP/SEMIS. Elle constitue les partitions demandées par l'utilisateur ou par recherche automatique des meilleures partitions, en les améliorant éventuellement par des itérations à centres mobiles (consolidation). Les partitions ainsi créées seront ensuite caractérisées de façon automatique par la procédure DECLA.

La procédure PARTI fournit une série de listages (une par partition demandée). Chaque listage possède la même structure : on obtient d'abord une description sommaire de la partition obtenue par coupure de l'arbre, indiquant pour chaque classe formée quels éléments terminaux la composent.

Si elle a été demandée, l'édition des coordonnées et valeurs-tests de ces classes apparaît ensuite. On obtient également, à la demande, la correspondance entre les classes et les individus.

Dans le cas d'une "consolidation" de cette partition par des itérations à centres mobiles, on trouve ensuite la chronique de progression de l'inertie inter-classes au cours des itérations.

Dans tous les cas, on trouve ensuite la décomposition de l'inertie en inertie inter-classes et inerties intra-classe, divisée en "avant et après consolidation" dans le cas où des itérations ont été effectuées. Si la consolidation a eu lieu et que cette édition est demandée, on trouve alors à nouveau les coordonnées et valeurs-tests des classes de la partition, puis la correspondance entre les classes et les individus.

Egalement à la demande, vient ensuite pour chaque classe, la liste des parangons (points les plus proches du centre de gravité des classes) et celle des points qui contribuent le plus à l'inertie intra-classe.

Deux paramètres de fonctionnement sont proposés.

- **Itérations de consolidation :**

L'objectif de la consolidation est de réaffecter les éléments d'une classe à une autre classe dont ils sont plus proches de façon à améliorer l'homogénéité à l'intérieur des classes. Ce processus s'effectue par des itérations successives à centres mobiles. Les centres mobiles sont initialement les centres de gravité des classes obtenues par coupure de l'arbre. Les calculs sont arrêtés dès que l'accroissement de l'inertie inter-classes devient trop faible d'une itération à l'autre. Il faut donc choisir un nombre d'itérations de consolidation (10 par défaut).

Si le nombre d'itérations est fixé à 0, il n'y aura pas de consolidation : les résultats de la coupure de l'arbre sont conservés tels quels.

Si le nombre d'itérations est positif, le logiciel édite l'affectation des individus seulement après la consolidation.

Si le nombre d'itérations est négatif, le logiciel édite l'affectation des individus avant et après la consolidation.

Nous avons fixé le nombre d'itérations à -10 dans le cadre de cet exemple.

- **Affectation des individus illustratifs aux classes les plus proches :**

Ce paramètre commande l'affectation des individus illustratifs à la classe dont le centre de gravité est le plus proche au sens de la distance euclidienne. Par défaut, le logiciel ne réalise pas cette affectation et considère les individus illustratifs comme des données manquantes.

Trois paramètres d'édition sont soumis à l'utilisateur :

- **Correspondance classes-individus :**

La correspondance peut-être demandée pour les individus actifs et/ou illustratifs. Mais la correspondance de ces derniers n'est possible que si ceux-ci ont été affectés aux centres de classe les plus proches (voir paramètre ci-dessus).

Ce paramètre permet d'une part de lister les individus de chacune des classes (item Composition de chaque classe) et d'autre part d'éditer le numéro de la classe à laquelle chaque individu appartient (item Classe d'appartenance des individus).

Les individus peuvent être repérés soit par leur identificateur court (4 caractères), soit par leur identificateur long (identificateur en entier).

Chaque édition sera donnée deux fois (avant et après la consolidation) si le paramètre relatif au nombre d'itérations de consolidation prend une valeur négative.

- **Coordonnées éditées pour les classes :**

Ceci permet de lister les coordonnées et les valeurs-tests des classes. Par défaut, ces caractéristiques sont éditées sur les cinq premiers axes. Mais on peut en demander l'édition sur l'ensemble des axes pris en compte dans l'agrégation des individus.

Dans le cas où le paramètre relatif au nombre d'itérations de consolidation est négatif, cette édition sera donnée deux fois (avant et après la consolidation).

- **Edition des parangons :**

Nombre indique que l'on souhaite éditer les n individus les plus caractéristiques de chaque classe, c'est-à-dire les n individus les plus proches du centre de gravité de chaque classe.

Les N plus proches et les N plus éloignés signifient que l'on désire éditer les N individus les plus caractéristiques de chaque classe ainsi que les N individus les plus éloignés (ceux qui contribuent le plus à l'inertie intra-classe).

Cette liste n'est fournie qu'après la consolidation (s'il y en a une).

Nous optons pour une partition en 9 classes. Les résultats de la procédure PARTI vous sont présentés ci-après.

1.1 PARTI-1 : Coupure 'a' de l'arbre en 9 classes

Ce tableau, constitué de 4 colonnes, décrit sommairement les classes avant la consolidation.

- **Classe** : identificateur de la classe.
- **Effectif** : nombre d'individus de la classe.
- **Poids** : poids des individus de la classe.
- **Contenu** : rangs des nœuds terminaux de la classe. Un nœud terminal est un ensemble d'individus réunis au cours du processus d'agrégation. Dans l'exemple, on, avait choisi 30 nœuds terminaux.

Coupure 'a' de l'arbre en 9 classes
Formation des classes (sur les individus actifs)
Description sommaire

Classe	Effectif	Poids	Contenu
Classe 1	110	110,00	1 à 3
Classe 2	49	49,00	4 à 5
Classe 3	128	128,00	6 à 10
Classe 4	185	185,00	11 à 14
Classe 5	114	114,00	15 à 18
Classe 6	79	79,00	19 à 20
Classe 7	63	63,00	21 à 23
Classe 8	68	68,00	24 à 25
Classe 9	204	204,00	26 à 30

Interprétation

La partition en 9 classes, adoptée pour les 1000 individus, donne deux grandes classes de 204 et 185 éléments, trois classes allant de 110 à 128 éléments et quatre petites classes (49 à 79 éléments).

Les poids des classes sont identiques aux effectifs puisque tous les individus ont le même poids égal à un.

1.2 PARTI-2 : Coordonnées et valeurs-tests des centres de classes avant consolidation

Cette sortie donne les coordonnées et les valeurs-tests des classes avant la consolidation. Dans le listage PARTI-8, nous obtenons la même sortie après consolidation.

1.2.1 Coordonnées des centres de classes avant consolidation

Ce tableau se compose de 4 grandes colonnes :

- **Libellé** : identificateur de la classe.
- **Effectif** : nombre d'individus de la classe.
- **Poids absolu** : poids des individus de la classe.
- **Distance à l'origine** : carré de la distance du Khi-2 de l'origine au centre de gravité de la classe mesuré dans le sous-espace factoriel choisi (14 axes ici).
- **Axe 1 – Axe 5** : coordonnées des centres de gravité sur les premiers axes.

Coordonnées des centres de classes après consolidation
Coupure 'a' de l'arbre en 9 classes

Libellé	Effectif	Poids absolu	Distance à l'origine	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5
CLASSE 1 / 9	147	147,00	0,54511	-0,15826	0,42521	-0,00168	-0,26465	-0,37854
CLASSE 2 / 9	97	97,00	1,18427	-0,63481	0,27568	0,59712	0,21892	0,50340
CLASSE 3 / 9	105	105,00	0,63120	0,20758	-0,00686	-0,26380	-0,39956	0,41959
CLASSE 4 / 9	164	164,00	0,44963	0,33236	0,35903	-0,18770	0,27021	-0,04286
CLASSE 5 / 9	124	124,00	0,70303	0,28163	-0,33038	0,21092	0,41367	-0,26417
CLASSE 6 / 9	74	74,00	1,37837	0,38592	-0,08714	-0,16999	-0,41457	0,24967
CLASSE 7 / 9	56	56,00	1,37230	-0,12985	0,13429	0,17477	0,13547	0,04521
CLASSE 8 / 9	78	78,00	1,45062	0,62320	-0,47910	0,50629	-0,28731	-0,02969
CLASSE 9 / 9	155	155,00	0,78910	-0,62116	-0,45254	-0,40029	0,06138	-0,10417

Interprétation

D'après les carrés des distances à l'origine, la classe 8 est la plus « atypique ». Elle contient des individus dont les caractéristiques sont très différentes de « l'individu moyen » de l'échantillon. La classe 4 est la plus proche de cet « individu moyen » (du moins dans l'espace des 14 premiers axes factoriels).

1.2.2 Valeurs-tests des centres de classes avant consolidation

Ce tableau se compose de 4 grandes colonnes :

- **Libellé** : identificateur de la classe.
- **Effectif** : nombre d'individus de la classe.
- **Poids absolu** : poids des individus de la classe.
- **Distance à l'origine** : carré de la distance du Khi-2 de l'origine au centre de gravité de la classe mesuré dans le sous-espace factoriel choisi (14 axes ici).
- **Axe 1 – Axe 5** : valeurs-tests des centres de gravité sur les premiers axes. Une classe est d'autant plus intéressante sur un axe que sa valeur-test est plus grande en valeur absolue.

Valeurs-Tests des centres de classes après consolidation
Coupure 'a' de l'arbre en 9 classes

Libellé	Effectif	Poids absolu	Distance à l'origine	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5
CLASSE 1 / 9	147	147,00	0,54511	-3,99381	11,46270	-0,04819	-7,92101	-11,56100
CLASSE 2 / 9	97	97,00	1,18427	-12,64750	5,86745	13,55000	5,17313	12,13810
CLASSE 3 / 9	105	105,00	0,63120	4,32206	-0,15265	-6,25602	-9,86737	10,57320
CLASSE 4 / 9	164	164,00	0,44963	8,94839	10,32650	-5,75582	8,62876	-1,39647
CLASSE 5 / 9	124	124,00	0,70303	6,44107	-8,07183	5,49435	11,22150	-7,31210
CLASSE 6 / 9	74	74,00	1,37837	6,63172	-1,59957	-3,32705	-8,44959	5,19243
CLASSE 7 / 9	56	56,00	1,37230	-1,92248	2,12404	2,94713	2,37894	0,81009
CLASSE 8 / 9	78	78,00	1,45062	11,01860	-9,04914	10,19560	-6,02513	-0,63542
CLASSE 9 / 9	155	155,00	0,78910	-16,17190	-12,58610	-11,87000	1,89545	-3,28235

Interprétation

L'intérêt de ce tableau est de juger à travers la statistique des valeurs-tests de l'intérêt des classes dans les différentes directions factorielles. On voit par exemple que toutes les classes sont bien différenciées sur le premier plan factoriel sauf la classe 7 qui a besoin des trois axes suivants pour trouver sa place dans l'espace.

1.3 PARTI-3 : Avant consolidation - Liste des individus actifs dans les classes de la partition

La correspondance classe-individu est pertinente lorsque l'on s'intéresse de près aux individus. Dans cette enquête, comme dans la majorité des cas, les individus sont anonymes et la correspondance classe-individu présente peu d'intérêt.

Le tableau contient 3 colonnes :

- **Numéro d'ordre** : numéro de l'individu dans la classe.
- **Numéro dans la base** : indique le numéro de la ligne où apparaît l'individu dans la base. Pour chaque classe, les individus sont triés dans l'ordre d'apparition dans la base.
- **Libellé de l'individu** : identifiant court ou long (voir onglet Paramètres) de l'individu. Dans cet exemple, les individus sont anonymes, SPAD leur affecte comme libellé le numéro de la ligne sur laquelle ils se trouvent dans la base.

Le tableau suivant a été tronqué :

Avant consolidation - Liste des individus actifs dans 1
Coupure 'a' de l'arbre en 9 classes
CLASSE 1 / 9

Numéro d'ordre	Numéro dans la Base	Libellé de l'individu
1	1	1
2	11	11
3	14	14
4	15	15
5	20	20
6	36	36
7	39	39
8	40	40

Interprétation

Les individus 1, 11, 14, 15, etc. appartiennent à la classe 1 de la partition en 9 classes.

1.4 PARTI-4 : Avant consolidation - Appartenance des individus actifs dans les classes de la partition

La correspondance individu-classe se présente sous la forme d'un tableau à 3 colonnes :

- **Numéro d'ordre** : numéro de l'individu dans la liste.
- **Libellé de l'individu** : identifiant court ou long (voir onglet Paramètres) de l'individu. Dans cet exemple, les individus sont anonymes, SPAD leur affecte comme libellé le numéro de la ligne sur laquelle ils se trouvent dans la base.
- **Classe** : indique le numéro de la classe à laquelle appartient l'individu.

Le tableau suivant a été tronqué :

Avant consolidation - Appartenance des individus a
Coupure 'a' de l'arbre en 9 classes

Numéro d'ordre	Libellé de l'individu	Classe
1	1	1
2	2	9
3	3	9
4	4	9
5	5	9
6	6	9
7	7	9
8	8	9
9	9	4
10	10	5

Interprétation

L'individu 9 appartient à la classe 4.

1.5 PARTI-5 : Consolidation de la partition

Dans ce tableau, on suit le processus de consolidation de la partition autour des centres de classes. Ce tableau sert à analyser l'évolution de l'inertie inter-classes au fur et à mesure des itérations.

Il se compose de 4 colonnes :

- **Itération** : numéro de l'itération.
- **Inertie totale** : est l'inertie totale du nuage des individus. Elle recalculée et doit rester constante.
- **Inertie inter-classes** est l'inertie inter-classes (à maximiser). Elle mesure l'hétérogénéité entre les classes de la partition.
- **Quotient** : part de l'inertie inter-classes dans l'inertie totale.

L'intérêt de la consolidation est de réaffecter les éléments d'une classe dans une autre classe dont ils sont plus proches de façon à améliorer l'homogénéité interne des classes. Ce processus s'effectue par des itérations successives à centres mobiles. Les centres sont initialement les centres de gravité des classes obtenues par coupure de l'arbre. Ces centres évoluent lorsque les individus passent d'une classe à l'autre.

Consolidation de la partition
réalisée par 10 itérations à centres mobiles

Itération	Inertie totale	Inertie inter-classes	Quotient
0	1,83681	0,73318	0,39916
1	1,83681	0,82054	0,44672
2	1,83681	0,82550	0,44942
3	1,83681	0,82669	0,45007
4	1,83681	0,82870	0,45116
5	1,83681	0,83069	0,45225
6	1,83681	0,83326	0,45364
7	1,83681	0,83580	0,45503
8	1,83681	0,83639	0,45535
9	1,83681	0,83650	0,45541

Interprétation

La consolidation de la partition a permis de faire passer la part de l'inertie inter-classes de 39.9% à 45.5%. Les classes sont globalement devenues plus homogènes.

L'accroissement de l'inertie inter-classes est pratiquement nul après la neuvième itération. Les calculs sont alors arrêtés.

En principe, un nombre important d'itérations (au-delà de 10) témoigne d'une classification peu stable.

1.6 PARTI-6 : Décomposition de l'inertie calculée sur 10 axes

Ce tableau de décomposition de l'inertie présente avant et après la consolidation les 5 composantes suivantes :

- **Inerties** : il s'agit des intitulés de différentes inerties éditées par le tableau : inertie inter-classes, inerties intra-classe et inertie totale.

- **Inertie avant/après** : donne l'inertie inter-classes, les inerties intra-classe ainsi que l'inertie totale (somme de l'inertie inter-classes et des inerties intra-classe), avant et après la consolidation.
- **Effectifs avant/après** : contient les effectifs de classes avant et après la consolidation.
- **Poids avant/après** : poids de chaque classe avant et après la consolidation.
- **Distances avant/après** : le carré de la distance du Khi-2 de l'origine au centre de gravité de la classe, avant et après la consolidation. Cette distance est mesurée dans le sous-espace vectoriel choisi précédemment (14 axes factoriels dans l'exemple).

Décomposition de l'inertie calculée sur 10 axes

Inerties	Inerties avant	Inerties après	Effectifs avant	Effectifs après	Poids avant	Poids après	Distances avant	Distances après
Inter-classes	0,73318	0,83650						
Intra-classe								
Classe 1 / 9	0,11137	0,13140	110	147	110,00	147,00	0,63534	0,54511
Classe 2 / 9	0,04592	0,10067	49	97	49,00	97,00	1,38326	1,18427
Classe 3 / 9	0,16249	0,11491	128	105	128,00	105,00	0,46112	0,63120
Classe 4 / 9	0,16870	0,13483	185	164	185,00	164,00	0,34437	0,44963
Classe 5 / 9	0,12852	0,13783	114	124	114,00	124,00	0,66692	0,70303
Classe 6 / 9	0,09443	0,08427	79	74	79,00	74,00	1,32930	1,37837
Classe 7 / 9	0,09149	0,07506	63	56	63,00	56,00	1,36612	1,37230
Classe 8 / 9	0,06626	0,06968	68	78	68,00	78,00	1,52292	1,45062
Classe 9 / 9	0,23445	0,15165	204	155	204,00	155,00	0,50056	0,78910
Totale	1,83681	1,83681						
Quotient (I. inter / I. totale)	0,39916	0,45541						

Interprétation

On constate que la consolidation améliore globalement la partition. Cependant, l'inertie intra-classe ne diminue pas dans toutes les classes. Si les classes 3 et 9 voient leur inertie intra-classe fortement diminuée, la classe 2 voit quant à elle son inertie augmenter (en même temps que son poids).

1.7 PARTI-8 : Coordonnées et valeurs-tests des centres de classes après consolidation

Ces tableaux s'interprètent de la même manière que ceux obtenus avant consolidation dans la sortie PARTI-2 (CF page 112).

1.8 PARTI-9 : Avant consolidation - Liste des individus actifs dans les classes de la partition

Ces tableaux s'interprètent de la même manière que ceux obtenus avant consolidation dans la sortie PARTI-3 (CF page 114).

1.9 PARTI-10 : Avant consolidation - Appartenance des individus actifs dans les classes de la partition

Ces tableaux s'interprètent de la même manière que ceux obtenus avant consolidation dans la sortie PARTI-4 (CF page 114).

1.10 PARTI-11 : Edition des parangons

Les parangons sont les individus les plus « caractéristiques » de chaque groupe au sens suivant : ce sont les individus les plus proches du centre de gravité (du point moyen) de la classe.

Par défaut, le logiciel édite les n individus (10 par défaut) situés le plus près du centre de gravité de chaque classe.

Mais il est possible de demander l'édition des n individus dont la contribution à l'inertie de la classe est la plus grande. Lorsque ces derniers ont un poids uniforme, ce sont les individus les plus éloignés du centre de gravité de chaque classe.

Le tableau suivant concerne les 10 parangons de la première classe. Pour chaque parangon on trouve 3 colonnes :

- **Rang** : rang de l'individu. Plus le rang de l'individu est grand, plus celui-ci se trouve éloigné du centre de gravité de sa classe.
- **Distance au centre de la classe** : carré de la distance entre l'individu et le centre de gravité de sa classe. Cette distance du Khi-2 est mesurée dans le sous-espace vectoriel choisi (14 axes dans l'exemple).
- **Libellé** : identificateur de l'individu.

Edition des parangons

Classe 1 / 9 Effectif: 147

Rang	Distance au centre de la classe	Libellé
1	0,29766	237
2	0,36841	888
3	0,40372	485
4	0,40372	619
5	0,41069	184
6	0,41069	232
7	0,41069	238
8	0,42478	980
9	0,45426	546
10	0,46176	271

Interprétation

On note que dans la classe 1, l'élément le plus proche du centre de gravité est l'individu 237 dont la distance au centre de gravité de la classe est de 0.29766.

1.11 PARTI-12 : Edition des individus à contribution maximale

Cette sortie n'est pas éditée par défaut. Pour l'obtenir, il suffit de demander l'édition des n individus les plus proches (on obtient la sortie PARTI-11) et des n plus éloignés, c'est à dire ceux dont la contribution à l'inertie de la classe est la plus grande.

On présente pour chaque classe les 10 individus qui ont le plus contribué à l'inertie de la classe. Pour chaque individu on trouve 3 colonnes :

- **Rang** : rang de l'individu. Plus le rang de l'individu est grand, plus sa contribution à l'inertie de la classe est élevée.
- **Inertie** : contribution à l'inertie de la classe de chacun des individus.
- **Libellé** : identificateur de l'individu.

Edition des individus à contribution maximale

Classe 1 / 9 Effectif: 147

Rang	Inertie	Libellé
1	2,25671	676
2	2,23244	901
3	2,10290	883
4	1,94831	693
5	1,83489	739
6	1,67668	362
7	1,67668	878
8	1,66126	555
9	1,55139	764
10	1,45083	678

Interprétation

Dans l'exemple, les individus ont des poids uniformes. Les individus dont la contribution à l'inertie de la classe est la plus grande sont donc les individus les plus éloignés du centre de gravité de la classe.

On note que dans la classe 1, l'élément qui a la contribution à l'inertie la plus forte est l'individu 676. Compte tenu de l'uniformité des poids, il s'agit de l'individu le plus éloigné du centre de gravité de la classe 1.

2. La procédure DECLA

La procédure **DECLA** permet de décrire les partitions déterminées par la procédure PARTI.

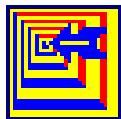
On peut caractériser soit chaque classe d'une partition, soit globalement la partition elle-même.

Cette procédure est en tous points identique à la procédure DEMOD présentée page 18. Dans DECLA, la variable à caractériser est la partition et les modalités sont les classes d'appartenance des individus. Dans cet exemple, la partition contient 9 classes donc 9 modalités.

Tous les éléments disponibles (actifs et illustratifs, sélectionnés pour l'analyse factorielle préalable) peuvent intervenir dans la caractérisation : les modalités des variables nominales, les variables nominales elles-mêmes, les variables continues, les fréquences (seulement lorsque l'analyse factorielle préalable est une AFC) et les axes factoriels.

Les éléments caractéristiques sont classés par ordre d'importance à l'aide d'un critère statistique ("valeur-test") auquel est associé une probabilité : plus la valeur-test est grande, plus la probabilité est faible, plus l'élément est caractéristique.

Dans le cas de la description des classes par les modalités des variables nominales, une option permet de classer les modalités caractéristiques soit suivant les valeurs-tests, soit suivant les pourcentages.



Caractérisation des classes de typologies (CLASS-MINER)

La procédure **CLASS-MINER** permet de décrire les partitions déterminées par la procédure PARTI.

La partition est créée par la procédure PARTI-DECLA, la procédure CLASS-MINER doit être insérée dans la filière après cette méthode.

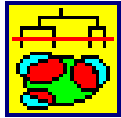
En termes de résultats, les procédures DECLA et CLASS-MINER sont en tous points identiques.

Dans la procédure DECLA, on caractérise la partition obtenue à partir des variables actives et illustratives sélectionnées dans l'analyse factorielle préalable.

CLASS-MINER vous permet de ne sélectionner qu'une partie des variables du fichier pour la caractérisation de la partition.

Cela vous permet entre autres de différencier les variables actives des variables illustratives.

Si plusieurs thèmes ressortent de votre étude, pour chaque thème, on utilise CLASS-MINER afin d'obtenir des tableaux de résultats pour chaque thème.



Marquage sémantique des classes d'une partition (MSCLA)

Cette procédure permet de caractériser les classes d'une partition par des marquages sémantiques.

La partition est créée par la procédure PARTI-DECLA, la procédure MSCLA doit être insérée dans la filière après cette méthode.

L'interprétation des résultats de cette méthode est en tous points identique à l'interprétation des résultats issus de la procédure MSMOD (CF page 41). Dans MSMOD, la variable nominale à caractériser est sélectionnée par l'utilisateur, tandis que dans MSCLA la variable nominale à caractériser est la partition obtenue avec la procédure PARTI-DECLA.

A

ALEATOIRE (sondage ou tirage...) :

Tirage au sort pour sélectionner un *échantillon*. Chaque *unité statistique* a une *probabilité* connue d'appartenir à cet *échantillon*. Pour constituer ce dernier, il est nécessaire de disposer d'une *base de sondage* de façon à connaître la totalité des *caractères* applicables aux *individus* de la *population* sur laquelle porte l'enquête.

ANALYSE DES CORRESPONDANCES MULTIPLES :

Analyse factorielle qui s'applique à des tableaux dont les lignes sont des *individus* ou des observations et dont les colonnes sont des *modalités* de variables *nominales*.

ANALYSE EN COMPOSANTES PRINCIPALES :

Analyse factorielle qui s'applique à des tableaux rectangulaires dont les colonnes sont des variables à valeurs numériques (ou variables *continues*) et dont les lignes représentent les *individus* sur lesquels sont mesurées ces variables.

ANALYSE FACTORIELLE :

Méthode statistique dont le but est de représenter graphiquement des objets et de décrire des liaisons entre ces objets. Parmi ces méthodes, on retient notamment l'*analyse en composantes principales*, l'*analyse des correspondances simples* et l'*analyse des correspondances multiples*.

ANALYSE FACTORIELLE DES CORRESPONDANCES:

Analyse factorielle qui s'applique à des *tableaux de contingence* afin d'étudier les relations existant entre deux variables *nominales*.

ATYPIQUE :

Individu ou donnée pour laquelle les informations sont jugées trop éloignées de la réalité ou de leurs valeurs habituelles pour en tenir compte.

C

CARACTERISTIQUES (ou caractères) :

Critères permettant de rendre compte de la diversité des *individus* d'une *population*. On distingue les caractères qualitatifs (sexe, profession, diplôme...), des caractères quantitatifs continus (âge, revenu...) et des caractères quantitatifs discrets (nombre d'enfants, nombre de voiture dans le foyer...). Le statisticien parle souvent de variables *nominales* et de variables *continues*.

CHOIX DE L'ECHANTILLON :

Il peut être effectué à l'aide de différentes techniques de *sondage* parmi lesquelles les *échantillons* de type probabiliste (*échantillons aléatoires*), les *échantillons* d'unités-types et les *échantillons* par *quotas*.

CLASSIFICATION (voir typologie).

COEFFICIENT DE VARIATION (CV) :

C'est le rapport entre l'*écart-type* d'une *variable* et sa *moyenne*. Il est le plus souvent exprimé en pourcentage.

CONTINUE (variable...) :

On parle de *variable* continue quand la mesure faite sur l'*individu* est de type quantitatif. Pour fixer les idées, une variable est continue quand la notion de *moyenne* a un sens.

CONTRIBUTION (ou contribution absolue) :

Mesure de la participation d'un élément (*modalité*, *variable*, *fréquence* ou *individu*) à la construction d'un axe factoriel.

CORRELATION :

Elle mesure la dépendance entre deux *variables continues*. La corrélation est exprimée par un coefficient de corrélation linéaire, partielle, multiple, ou encore par une droite de régression, etc.

COSINUS (ou contribution relative) :

Mesure de la qualité de représentation d'un élément (*modalité*, *variable*, *fréquence* ou *individu*) sur un axe factoriel.

D

DEONTOLOGIE :

Règles que les *instituts de sondage* et d'études de marché s'engagent à respecter. Le Code International CCI/ESOMAR de pratiques loyales concerne les études de marché et d'*opinion*. Le Code Syntec/U.D.A. de pratiques loyales concerne les *panels* et les *enquêtes* répétitives. Les *sondages* politiques doivent suivre les directives de la *Commission des sondages*.

DISPERSION :

Critère permettant de mesurer la plus ou moins *grande dispersion des observations autour de la moyenne* (voir *Ecart-type*, *Variance*).

Dans un *sondage*, la dispersion de l'*échantillon* est une qualité nécessaire pour avoir une bonne représentativité.

DISTRIBUTION STATISTIQUE :

Dans le cas d'une *variable nominale*, la distribution statistique est la répartition des *individus* d'une *population* dans les *modalités* d'une *variable* (effectifs).

Dans le cas d'une *variable continue*, la distribution est caractérisée en particulier par sa *moyenne* et son *écart-type*.

E

ECART-TYPE :

Critère de *dispersion* d'une *distribution*. C'est la racine carrée de la *variance*. Il est fonction de la *taille de l'échantillon*.

ECHANTILLON :

Partie de la *population* analysée (*base de sondage*) et choisie selon un *plan de sondage* défini au préalable.

ECHANTILLONNAGE :

Procédure de sélection d'un *échantillon* où les *individus* représentent la *population* étudiée. Il existe l'échantillonnage *aléatoire*, par *quotas*, par *grappes*, etc.

EQUIPONDERATION (ou équirépartition) :

Procédure consistant à attribuer un *poids* identique à différents objets, événements ou *caractères*.

ERREURS D'OBSERVATION :

Ce sont les erreurs de mesure liées à la méthodologie, au *questionnaire* ou à l'enquêteur, mais aussi les erreurs de réponse des enquêtés dues à une défaillance de la mémoire ou à un manque de franchise.

F

FOURCHETTE :

Terme statistique qui représente l'intervalle entre deux valeurs extrêmes.

FREQUENCE :

Il s'agit d'un nombre de quelque chose tel que le nombre d'observations d'un événement ou d'un phénomène. La fréquence peut être en pourcentage ou en effectif.

G

GAUSS (courbe de...) :

Courbe en cloche ou courbe de la *loi normale* qui donne la *probabilité* d'une *variable aléatoire continue*. Cette courbe est symétrique et elle est définie par sa *moyenne* et son *écart-type*.

I

INDIVIDU :

Unité statistique tel qu'une personne, un ménage, une entreprise ou encore une ville auprès de laquelle sont recueillies des informations. L'ensemble de ces individus représente la *base de sondage* ou la *population* de référence.

INTERVALLE DE CONFIANCE :

Intervalle (appelé marge d'erreur) dans lequel se trouve la valeur cherchée avec une *probabilité* fixée. La longueur de cet intervalle dépend du niveau de confiance, du taux de la réponse considérée et de la *taille de l'échantillon*. Le calcul n'est justifié que pour les *sondages aléatoires*. Il ne peut pas être déterminé dans le cas des sondages par *quotas* mais on considère qu'elle est proche de celle des sondages aléatoires.

K

KHI-2 (test du ...)

Le test (basé sur la *distribution* statistique d'une *variable aléatoire* suivant une loi du Khi-2) permet de tester la similitude entre une distribution observée et une distribution théorique et de tester l'indépendance entre deux *variables qualitatives*.

L

LISSAGE :

Calculs permettant de réduire l'influence de facteurs accidentels et de variations saisonnières observées sur des données chronologiques.

LOI NORMALE :

Loi de *probabilité* représentée par une courbe symétrique (courbe en cloche). Elle intervient dans le calcul de la précision d'un *sondage aléatoire*, en particulier pour la détermination d'*intervalles de confiance*. Une loi normale de *moyenne* nulle et d'*écart-type* 1 est dite centrée réduite.

M

MEDIANE :

Paramètre de *position* de la valeur centrale d'une *distribution*. Elle scinde la *population* en deux parties égales (50% des observations de part et d'autre de cette valeur).

MODALITE :

Les modalités sont les valeurs que prend une *variable nominale*. La *variable* sexe a deux modalités. On peut dire aussi : groupe, catégorie, classe.

MODE :

Paramètre de *position* de la valeur centrale d'une *distribution*. Il s'agit de la valeur, d'une distribution où l'effectif est le plus important.

MOYENNE :

Paramètre de *position* d'une *distribution* le plus

souvent utilisé pour caractériser la valeur centrale (moyenne arithmétique, moyenne pondérée).

N

NOMINALE (variable...) :

La *variable* est nominale quand les valeurs qu'elle prend sont des catégories. C'est le cas du sexe, de l'âge (s'il est découpé en classes), de la catégorie socioprofessionnelle. On parle aussi de *variable qualitative* ou catégorielle.

NON-REPONSE :

Nom donné à une réponse lorsque la personne interrogée ne répond pas à une question. On parle de non-réponse totale lorsqu'une personne interrogée ne répond à aucune des questions et de non réponse partielle lorsqu'une personne interrogée répond seulement à certaines questions.

O

OPINION :

Avis, idée, sentiment, jugement émis par une personne interrogée sur des faits ou des phénomènes.

P

POIDS :

Part d'un groupe d'*individus* par rapport à l'ensemble de la *population* (exemple : 48% d'hommes et 52% de femmes). Il s'agit d'un coefficient affecté à chacun des individus d'un *échantillon* (poids de *redressement*).

PONDERATION :

Traitement statistique dont l'objet est d'affecter un *poids* à des catégories d'*individus* notamment dans le but de redresser l'*échantillon*.

POPULATION :

Il s'agit d'*unités statistiques* telles que des personnes, des ménages ou des entreprises. Elles servent de *base de sondage* pour constituer un *échantillon*.

POSITION :

Critère permettant de mesurer la tendance centrale des observations d'une *distribution*. On distingue la *moyenne*, la *médiane* et le *mode*.

PREVISION :

Traitement statistique dont le but est de prévoir des *variables* de comportement ou d'*opinion* selon différents facteurs.

PROBABILITE

Procédure consistant à attribuer à tout événement un nombre compris entre 0 et 1. Ce nombre reflète la chance qu'un événement a de se réaliser. Une probabilité nulle correspond à un événement impossible alors qu'une probabilité égale à 1 concerne un événement certain. La théorie des probabilités est à la base de la statistique et des *sondages*.

PROPORTIONNEL (échantillon) :

Procédure d'*échantillonnage* où les catégories de *population* sont représentées dans l'*échantillon* au prorata de leur *poids* dans la *population* de référence.

Q

QUALITATIVE (voir variable nominale).

QUANTITATIVE (voir variable continue).

QUANTILE (ou fractile) :

Valeur partageant une *distribution* en parties d'effectifs identiques. La *médiane* divise la distribution en deux parties, les *quartiles* en quatre parties, les *déciles* en dix parties et les *centiles* en cent parties.

R**REDRESSEMENT :**

Opération dont l'objet est de corriger le *bias* d'*échantillonnage* observé à la suite d'une *enquête*. Cette correction est effectuée par rapport au *plan de sondage* prévu ou par rapport au *recensement* par exemple. Le redressement permet d'améliorer la représentativité d'un *échantillon* notamment lorsqu'il n'a pas été possible d'interroger exactement le nombre prévu de personnes d'une certaine catégorie.

REPARTITION :

Mot utilisé dans le sens de *distribution* d'une part

et dans le sens de structure d'autre part.

T**TABEAU DE DONNEES :**

C'est un tableau rectangulaire composé de n lignes et de p colonnes. Dans chaque case (intersection d'une ligne avec une colonne), se trouve une donnée. On parle aussi de matrice des données ou de tableau « *individus x variables* ».

TABEAU DE FREQUENCES :

Tableau à double entrée dont chaque case contient la somme des valeurs des *individus* sur une troisième variable de type *continue*.

TABEAU DE CONTINGENCE :

Tableau à double entrée présentant la *distribution* des effectifs croisés de deux variables *nominales*.

TEST DE STUDENT :

Opération statistique dont l'objectif est de vérifier si les différences observées entre deux résultats de *sondages* sont significatives (généralisable à l'ensemble de la *population* de référence) ou si elles sont dues aux seuls hasards de l'*échantillonnage*.

TEST STATISTIQUE :

Opération statistique dont l'objectif est de vérifier des hypothèses, des estimations ou des ajustements (exemple : test du *Khi-2*, test de Student, etc.).

TRI :

Opération qui consiste à déterminer la *répartition* des *individus* dans les *modalités* des variables *nominales* ainsi que les statistiques de base (*moyenne*, *écart-type*, minimum et maximum) des variables *continues*.

TRI CROISE :

Opération qui consiste à croiser deux variables *nominales* sous forme de tableau numérique ou de graphique. Ceux-ci présentent la *distribution* d'effectif conjoint des deux variables *nominales*

TPOLOGIE :

Procédure d'analyse des données visant à construire des groupes d'*individus* ou de phénomènes les plus homogènes possibles en fonction d'un ensemble de variables.

U

UNITE STATISTIQUE :

L'unité statistique est l'*individu* si le *questionnaire* concerne l'individu interrogé.
L'unité statistique est le ménage si le questionnaire concerne le ménage.

V

VARIABLES :

Ce sont les *caractéristiques* de l'*unité statistique*.
Les variables permettent de décrire ou de caractériser son comportement. Pour chaque *unité statistique*, on fait les mêmes mesures, les mêmes relevés, on pose les mêmes questions. Au lieu de variables, on peut dire mesure ou paramètre, ou question.

VARIABLES ACTIVES :

Variables qui participent à la construction des axes de l'*analyse factorielle*.

VARIABLES ILLUSTRATIVES :

Toutes les *variables* qui n'ont pas participé à la construction des axes, mais permettent d'illustrer les *analyses factorielles*.

VARIANCE :

Critère de *dispersion* d'une *distribution*. Il s'agit plus précisément de la *moyenne* des carrés des écarts entre les valeurs de la distribution et sa moyenne. Sa racine carrée est appelée *écart-type*.

Bibliographie

- Aluja-Banet T., Morineau A. (1998) – *Analyse en composantes principales*. Cisia-Ceresta.
- Benzécri, J.P. (1976) – L'Analyse des Données. Tome 1 : La Taxinomie. Tome 2 : L'Analyse des Correspondances. *Dunod*.
- Benzécri, J.P. (1982) – Construction d'une classification ascendante hiérarchique par la recherche en chaîne de voisins réciproques. *Cahiers d'Analyse des Données*, 7, 209-218.
- Benzécri, J.P., Benzécri, F. (1985) – Introduction à la Classification Ascendante hiérarchique d'après un exemple de Données Economiques. *J. Soc.Stat. de Paris*, 1, 14-34
- Bouroche J.-M., Saporta G. (1980) – *L'analyse des données*. Coll « Que sais-je », n°1854, PUF, Paris.
- Brossier G., Dussaix A.-M. (1999), *Enquêtes et sondages – Méthodes, modèles, applications, nouvelles approches*. Dunod, Paris.
- Confais J., Nakache J.P. (2000) – *Méthodes de classification*. Cisia-Ceresta.
- Grangé D., Lebart L. (1993) - *Traitements statistiques des enquêtes*. Dunod, Paris.
- Lebart L. (1975 a) – L'orientation du dépouillement de certaines enquêtes par l'analyse des correspondances multiples. *Consommation*, 2, p 73-96. Dunod.
- Lebart L. (1989), Stratégies du traitement des données d'enquêtes, *La revue de Modula*, numéro 3, p 21-29.
- Lebart L., Fénélon J.P. (1971) – *Statistique et informatique appliquées*. Dunod, Paris.
- Lebart L., Morineau A., Piron M. (1995) – *Statistique exploratoire multidimensionnelle*. Dunod, Paris.
- Lebart L., Morineau A., Tabard N. (1977) – *Technique de la description statistique*. Dunod, Paris.
- Morin S., Morineau A. (2000) – *Pratique du traitement des enquêtes*. Cisia-Ceresta.
- Morineau A. (1984) – Note sur la caractérisation statistique d'une classe et les valeurs-tests, *Bull. Techn. du Centre de Statist. et d'Infor. Appl.*, 2 , p20-27.
- Saporta G. (1990) – *Probabilités, analyse des données et statistiques*. Technip, Paris.
- Tenenhaus M. (1994) – *Méthodes statistiques en gestion*. Dunod, Paris.