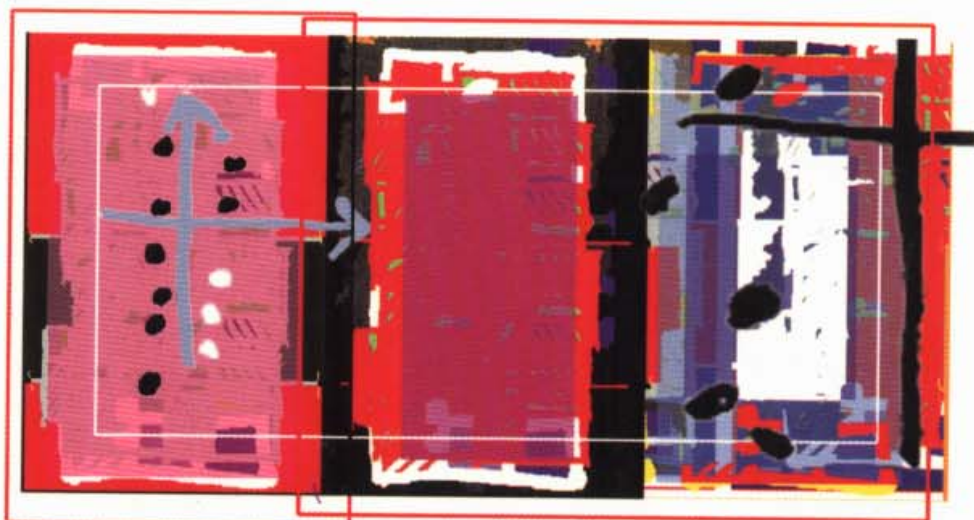


Brigitte Escofier Jérôme Pagès

2^e CYCLE • ÉCOLES D'INGÉNIEURS

Analyses factorielles simples et multiples

Objectifs, méthodes et interprétation



3^e édition

DUNOD

Table des matières

1 Analyse en Composantes Principales

1.1 Données et objectifs de l'étude	7
1.2 Transformation des données	9
1.3 Nuage des individus	10
1.4 Nuage des variables	11
1.5 Ajustement du nuage des individus	13
1.6 Ajustement du nuage des variables	14
1.7 Dualité et formules de transition en ACP	16
1.7.1 Inerties	16
1.7.2 Facteurs	16
1.7.3 Relations de transition	18
1.7.4 Représentation superposée	19
1.7.5 Projection des vecteurs unitaires de la représentation des individus	19
1.8 Schéma général de l'ACP	20
1.9 Aides à l'interprétation	22
1.9.1 Définitions	22
1.9.2 Exemple numérique	23

2 Exemple d'Analyse en Composantes Principales et de Classification Ascendante Hiérarchique

2.1 Données et problématique	25
2.1.1 Description des données	25
2.1.2 Problématique	26
2.2 Résultats de l'ACP	28
2.2.1 Inertie des facteurs	28
2.2.2 Interprétation du premier facteur	29
2.2.3 Interprétation du deuxième facteur	31
2.2.4 Premier plan factoriel	33
2.2.5 Conclusion	35

2.3 Introduction à la méthode de Ward	35
2.3.1 Construction et description d'un arbre	35
2.3.2 Arbre et partition	37
2.3.3 Qualité d'une partition	37
2.3.4 Algorithme de Ward	39
2.3.5 Utilisation des indices d'agrégation	40
2.4 Caractérisation directe d'une classe d'individus	42
2.4.1 Problématique	42
2.4.2 Notion de la valeur-test	43
2.4.3 Synthèse : le tableau de caractérisation	45
2.4.4 Valeur-test et probabilité associée	45
2.4.5 Probabilité associée et test d'hypothèse	47
2.5 Interprétation simultanée d'un plan factoriel et d'un arbre hiérarchique	48
2.5.1 Graphiques	48
2.5.2 Indicateurs	49
2.5.3 Mise en œuvre conjointe d'une ACP et d'une CAH	51
2.6 Construction et amélioration d'une partition : l'algorithme d'agrégation	51
autour des centres mobiles	51
2.6.1 Principe	51
2.6.2 Utilisation à l'issue d'une classification hiérarchique.	53
2.6.3 Utilisation en amont d'une classification hiérarchique	53
3. Analyse factorielle des correspondances	
3.1 Données, notations, hypothèse d'indépendance	55
3.2 Objectifs	57
3.3 Transformations des données en profils	58
3.4 Ressemblance entre profils : distance du χ^2	59
3.5 Les deux nuages	60
3.5.1 Nuage des profils-lignes	60
3.5.2 Nuage des profils-colonnes	61
3.6 Ajustement des deux nuages	61
3.6.1 Ajustement du nuage des profils-lignes	61
3.6.2 Ajustement du nuage des profils-colonnes	63
3.6.3 Un aspect technique du centrage en AFC	63
3.7 La dualité	64
3.7.1 Statistique du χ^2 et inertie des deux nuages N_I et N_J	64
3.7.2 Dualité entre les facteurs sur I et les facteurs sur J	65
3.7.3 Interprétation de l'inertie des axes	67
3.7.4 Formule de reconstitution des données	68
3.8 Aides à l'interprétation et éléments supplémentaires	69
3.9 Schéma général de l'AFC	69
3.10 Conclusion	71

4 Analyse des Correspondances Multiples

4.1 Données et notations	73
4.1.1 Données	73
4.1.2 Codage condensé	73
4.1.3 Tableau disjonctif complet	74
4.1.4 Hypertableau de contingence	74
4.1.5 Tableau de Burt	75
4.2 Objectifs	76
4.2.1 Etude des individus	76
4.2.2 Etude des variables	77
4.2.3 Etude des modalités	77
4.2.4 Conclusion sur les objectifs	77
4.3 AFC appliquée à un Tableau Disjonctif Complet	78
4.3.1 ACM et AFC	78
4.3.2 Nuage des individus	79
4.3.3 Nuage des modalités	79
4.3.4 Relations de transition et représentation simultanée	80
4.3.5 Les variables à travers leurs modalités	82
4.3.6 Synthèse des variables qualitatives	83
4.3.7 Représentation des variables en ACM	84
4.4 Analyse des Correspondances d'un tableau de Burt	85
4.4.1 Tableau de Burt et Tableau Disjonctif Complet	85
4.4.2 Analyse des liaisons binaires et décomposition des χ^2	86
4.5 Codage des variables qualitatives	87
4.5.1 Pourquoi transformer des variables continues en variables qualitatives ?	87
4.5.2 Choix du nombre de classes	88
4.5.3 Choix des classes	89
4.6 Conclusion	90

5 Calculs et dualité en Analyse Factorielle

5.1 Introduction	91
5.2 Calcul des axes d'inertie et des facteurs d'un nuage de points	91
5.2.1 Notations : les matrices X , M et D	91
5.2.2 Projection d'un nuage sur un axe u	92
5.2.3 Inertie du nuage projeté	92
5.2.4 Calcul des axes d'inertie maximum. Cas de la métrique identité	93
5.2.5 Calcul des axes d'inertie maximum pour une métrique quelconque	93
5.2.6 Calcul des facteurs et de leur inertie	94
5.2.7 Définition du nuage des colonnes de X	94
5.3 Nuages des lignes et des colonnes en ACP et en AFC	95
5.3.1 Matrices X , M , D en ACP	95
5.3.2 Matrices X , M , D en AFC	95
5.3.3 Matrices X , M , D en ACM	97

5.4 Dualité	97
5.4.1 Relations entre les axes d'inertie et les facteurs des deux nuages	97
5.4.2 Le schéma de dualité	99
5.4.3 Dualité des objectifs en ACP	100
5.4.4 Formules de transition	101
5.4.5 Analyse factorielle de tableaux de distances ou de similarités	102
5.5 Mise en œuvre des calculs	103
5.5.1 Simplification en AFC	103
5.5.2 Diagonalisation d'une matrice non symétrique particulière	103
5.5.3 Enchaînement des calculs	104
5.6 Reconstitution des données et approximation de X	104
5.7 Une équivalence en ACM	106

6 Exemple de traitement de tableau multiple par ACM et AFC

6.1 L'enquête Ouest-France	109
6.2 Analyse simultanée de plusieurs groupes de variables	111
6.3 Le problème des réponses manquantes	112
6.3.1 Les non-réponses dans les variables supplémentaires	112
6.3.2 Les non-réponses dans les variables actives	112
6.3.3 Les non-réponses dans les tableaux croisés	113
6.4 Première analyse : ACM des rubriques	114
6.4.1 Plan des deux premiers facteurs	114
6.4.2 Facteurs 3 et 4	120
6.5 Deuxième analyse : ACM du signalétique	120
6.5.1 Les modalités actives : le signalétique	121
6.5.2 Les modalités supplémentaires : les rubriques	122
6.5.3 Contradiction apparente entre les deux analyses	122
6.5.4 Perte des dispersions spécifiques des profils de lecture	123
6.6 Une analyse non satisfaisante : ACM des rubriques et du signalétique	123
6.6.1 Typologie des individus	123
6.6.2 Typologie des modalités	123
6.6.3 Indices concernant les groupes	123
6.6.4 L'Analyse Factorielle Multiple, alternative de cette ACM	124
6.7 Troisième analyse : AFC du tableau croisant signalétique et rubriques	124
6.7.1 Plan des deux premiers facteurs	125
6.7.2 Intérêt et limites de cette analyse	127
6.8 Conclusion	127

7 L'Analyse Factorielle Multiple à partir de deux applications

7.1 L'exemple des vins	129
7.1.1 Présentation des données	129
7.1.2 Description de la problématique	130
7.1.3 Pondération des groupes de variables	132

7.1.4 Typologie des vins et principaux facteurs de variabilité	133
7.1.5 Représentation superposée des vins décrits par chaque groupe de variables séparément	136
7.1.6 Facteurs communs	139
7.1.7 Typologie des groupes	140
7.1.8 Projections des composantes principales de chaque groupe	142
7.2 AFM appliquée aux données de l'enquête Ouest-France	143
7.2.1 Premier facteur : les modalités	144
7.2.2 Deuxième facteur : les modalités	145
7.2.3 AFM et ACM des rubriques	145
7.2.4 Représentation superposée des individus	146
8 Aspects théoriques et techniques de l'Analyse Factorielle Multiple	
8.1 Données et notations	149
8.2 L'Analyse Factorielle Multiple dans R^K : les individus	150
8.2.1 Influence de la pondération des groupes sur les J nuages N_j^I	151
8.2.2 Influence de la pondération des groupes sur le nuage N_I associé à toutes les variables	151
8.2.3 Représentation des J nuages N_j^I dans R^K et nuage moyen	152
8.2.4 Représentation du nuage moyen	152
8.2.5 Représentation superposée des J nuages définis par chaque groupe de variables	153
8.3 L'Analyse Factorielle Multiple dans R^I : les variables	156
8.3.1 Influence de la pondération des groupes sur les nuages des variables	157
8.3.2 Représentation des variables	158
8.3.3 Représentation des composantes principales de chaque groupe	158
8.3.4 Recherche de facteurs communs aux groupes de variables : AFM et analyse multicanonique	159
8.3.5 Aides à l'interprétation	164
8.4 L'Analyse Factorielle Multiple dans R^{I^2} : les groupes de variables	164
8.4.1 Le nuage N_j des groupes de variables	165
8.4.2 Influence de la pondération des groupes sur le nuage N_j	165
8.4.3 Interprétation du produit scalaire entre deux groupes	165
8.4.4 Etude du nuage des groupes de variables	167
8.5 AFM et modèle INDSCAL	169
8.5.1 Interprétation du modèle INDSCAL dans R^K	170
8.5.2 Interprétation du modèle INDSCAL dans R^I	170
8.5.3 Interprétation du modèle INDSCAL dans R^{I^2}	171
8.5.4 Estimation des paramètres du modèle INDSCAL par l'AFM	172
8.5.5 Cas des tableaux de distances et de similarités	172
8.6 Cas des variables qualitatives et des tableaux mixtes	173
8.6.1 Equivalence entre ACM et ACP pondérée des indicatrices	173
8.6.2 Variables qualitatives et tableaux mixtes en AFM	175

8.7	Éléments supplémentaires	177
8.7.1	Individus supplémentaires	177
8.7.2	Groupes de variables supplémentaires	177
8.8	Mise en œuvre de l'Analyse Factorielle Multiple	178
9	Méthodologie de l'AFM	
9.1	Tactique méthodologique	181
9.1.1	AFM et analyses séparées	181
9.1.2	Cas dans lequel les variables sont homologues d'un groupe à l'autre	182
9.1.3	Définition et statut des groupes de variables	185
9.2	Aides à l'interprétation	187
9.2.1	Mesures globales de liaison entre deux groupes de variables	188
9.2.2	Aides relatives à la représentation des groupes de variables	188
9.2.3	Qualité de représentation du nuage des variables d'un groupe	191
9.2.4	Aides relatives aux facteurs partiels	192
9.2.5	Décomposition de l'inertie associée à la représentation superposée des nuages partiels	194
10	Comparaison de tableaux de fréquence binaire	
10.1	Données et problèmes	197
10.1.1	Notations	197
10.1.2	Exemples	198
10.1.3	Réduction à des tableaux binaires	200
10.1.4	Quelques questions sur la comparaison de tableaux binaires	200
10.1.5	Conclusion	201
10.2	Etude des marges binaires	202
10.3	Première analyse : les tableaux en supplémentaire dans l'AFC de leur somme	203
10.3.1	Principe	203
10.3.2	Interprétation du tableau somme	205
10.3.3	Interprétation des tableaux supplémentaires (hommes et femmes)	210
10.3.4	Bilan	212
10.4	Deuxième analyse : AFC de variables croisées ou de tableaux juxtaposés	213
10.4.1	Tableau traité et problèmes spécifiques aux tableaux composés de sous-tableaux	213
10.4.2	Comparaison avec l'analyse de la somme et décomposition de l'inertie	215
10.4.3	Indices de contribution à l'inertie de la marge et des sous-tableaux	216
10.4.4	Indices de qualité de représentation des différents nuages	220
10.4.5	Indices de comparaison des facteurs des différents tableaux	221
10.4.6	Calcul des différents indices et AFC par sous-tableaux	224
10.4.7	AFC de tableaux juxtaposés et AFM	225
10.4.8	Bilan	227

10.5 Troisième analyse : analyse intra	229
10.5.1 Problématique et principes de l'analyse intra	229
10.5.2 Généralisation de l'AFC	230
10.5.3 Modèle de l'analyse intra	232
10.5.4 Interprétation des formules de l'analyse intra	233
10.5.5 Commentaires de l'exemple	235
10.5.6 Bilan	236
10.6 Conclusion	238
11 Interprétation des résultats d'une analyse factorielle	
11.1 Prolégomènes	241
11.2 Interprétation d'une ACP	244
11.2.1 Etude de l'inertie des facteurs	244
11.2.2 Interprétation des facteurs	246
11.2.3 Cas de l'ACP non normée	250
11.3 Interprétation d'une AFC	251
11.3.1 Valeurs propres	251
11.3.2 Contributions des lignes et des colonnes	252
11.3.3 Coordonnées des éléments actifs	252
11.3.4 Cas de modalités ordonnées ou partitionnées	252
11.3.5 Cas dans lesquels le tableau analysé n'est pas un tableau de contingence	253
11.4 Interprétation d'une ACM	253
11.4.1 Inertie de facteurs	253
11.4.2 Contributions des individus et des modalités	254
11.4.3 Contributions des variables	254
11.4.4 Coordonnées des modalités et des individus	255
11.5 Interprétation d'une AFM	255
11.5.1 Résultats de l'analyse séparée de chaque groupe	255
11.5.2 Valeurs propres de l'analyse globale	256
11.5.3 Relations entre les facteurs de l'analyse globale et les groupes	256
11.5.4 Projections des variables et du nuage moyen des individus	258
11.5.5 Représentations superposées des individus et des modalités	259
11.5.6 Cas où tous les groupes comprennent les mêmes variables	259
11.6 Quelques types de facteurs	259
11.6.1 Facteur dû à quelques éléments aberrants	260
11.6.2 Facteur d'opposition	260
11.6.3 Facteur mettant en évidence un groupe	261
11.6.4 Facteur associé à une partition	261
11.6.5 Facteur d'échelle	261
11.6.6 L'effet taille en ACP	262
11.6.7 L'effet Guttman en AFC ou ACM	262
11.6.8 Le nom d'un facteur	263

12 Fiches techniques

12.1 Fiche 1 : moyenne et barycentre, variance et inertie	265
12.1.1 Cas d'une variable	265
12.1.2 Cas de deux variables	267
12.1.3 Cas d'un nombre quelconque K de variables	268
12.2 Fiche 2 : représentation des variables dans R^I	269
12.2.1 Espace et métrique	269
12.2.2 Centrage	269
12.2.3 Réduction	270
12.2.4 Coefficient de corrélation	270
12.3 Fiche 3 : distance, norme et produit scalaire	271
12.3.1 Espace vectoriel et espace euclidien	271
12.3.2 Distance	272
12.3.3 Norme	272
12.3.4 Produit scalaire	273
12.3.5 Angles	273
12.3.6 Orthogonalité, théorème de Pythagore et projection orthogonale	274
12.3.7 Expression matricielle	274
12.3.8 Produit scalaire et base orthonormée	274

Bibliographie	277
----------------------	-----

Index systématique	279
---------------------------	-----

Numérotations et renvois

Les sections, tableaux et figures sont numérotés par chapitre. **Tableau 2.5** : cinquième tableau du chapitre 2.

SCIENCES SUP

Brigitte Escofier • Jérôme Pagès

ANALYSES FACTORIELLES SIMPLES ET MULTIPLES

Objectifs, méthodes et interprétation

Les méthodes d'analyse des données ont démontré leur efficacité dans l'étude des grandes masses complexes d'informations permettant la confrontation simultanée des données – ce qui est infiniment plus fructueux que leur étude séparée – elles fournissent un outil de synthèse remarquable.

Débordant largement de leurs champs traditionnels, ces méthodes sont maintenant appliquées dans tous les domaines où l'on accumule d'importants fichiers de données.

Cette troisième édition, entièrement revue et corrigée, a été enrichie des thèmes nouveaux suivants :

- une présentation de la méthode de classification automatique utilisée couramment en parallèle à l'analyse factorielle ;
- un chapitre dédié à la méthodologie de l'analyse factorielle multiple, méthode désormais bien répandue et disponible sur plusieurs grands logiciels.

Brigitte Escofier était l'une des fondatrices de l'École française d'analyse des données. Docteur d'État ès sciences mathématiques, professeur à l'université de Rennes, elle a enseigné à l'Institut national des sciences appliquées (INSA) de Rennes et à l'IUT de Vannes. Brigitte Escofier s'est éteinte en juillet 1994.

Jérôme Pagès est ingénieur agronome, docteur en économie et habilité à diriger des recherches en statistique. Professeur de l'enseignement supérieur agronomique, il dirige actuellement l'unité de génie statistique de l'Institut supérieur de formation en agro-alimentaire (INSFA).



9 782100 041275

ISBN 2 10 004127 4
Code 044127



3^e édition



BRIGITTE ESCOIFIER
ET JÉRÔME PAGÈS

ont travaillé en équipe pendant douze années. On leur doit de nombreux travaux de recherches et un ouvrage d'initiation aux traitements statistiques.



MATHÉMATIQUES



PHYSIQUE



CHIMIE



PHYSIQUE APPLIQUÉE



INFORMATIQUE



SCIENCES DE LA NATURE
ET DE LA VIE

