

Sommaire

1	Analyse en Composantes Principales (ACP)	1
1.1	Données - notations - exemples	1
1.2	Objectifs	2
1.2.1	Étude des individus	2
1.2.2	Étude des variables	3
1.2.3	Relation entre les deux études	5
1.3	Étude des individus	5
1.3.1	Nuage des individus	5
1.3.2	Ajustement du nuage des individus	7
1.3.3	Représentation des variables	11
1.4	Étude des variables	13
1.4.1	Nuage des variables	13
1.4.2	Ajustement du nuage des variables	14
1.5	Relation entre les représentations de N_I et de N_K	16
1.6	Aides à l'interprétation	17
1.6.1	Indicateurs numériques	17
1.6.2	Éléments supplémentaires	19
1.6.3	Description automatique des axes	23
1.7	Mise en œuvre avec FactoMineR	25
1.8	Compléments	26
1.8.1	« Test » de significativité des axes	26
1.8.2	Résultats sur les variables : loadings <i>versus</i> corrélation	26
1.8.3	Représentation simultanée : graphe biplot	27
1.8.4	Données manquantes	27
1.8.5	Jeu de données de grandes dimensions	28
1.8.6	Rotation varimax	28
1.9	Exemple : données dépenses des ménages	29
1.9.1	Description des données	29
1.9.2	Problématique	29
1.9.3	Choix de l'analyse	31
1.9.4	Mise en œuvre de l'analyse	33
1.10	Exemple : données températures	44

1.10.1	Description des données - problématique	44
1.10.2	Choix de l'analyse	45
1.10.3	Mise en œuvre avec FactoMineR	47
1.11	Exemple : données génomiques	53
1.11.1	Description des données - problématique	53
1.11.2	Choix de l'analyse	53
1.11.3	Mise en œuvre	54
2	Analyse Factorielle des Correspondances (AFC)	61
2.1	Données - notations	61
2.2	Objectifs et modèle d'indépendance	64
2.2.1	Objectifs	64
2.2.2	Modèle d'indépendance et test du χ^2	64
2.2.3	Modèle d'indépendance et AFC	67
2.3	Les nuages et leur ajustement	67
2.3.1	Nuage des profils lignes	67
2.3.2	Nuage des profils colonnes	69
2.3.3	Ajustement des nuages N_I et N_J	70
2.3.4	Exemple : attitude des femmes à l'égard du travail féminin	71
2.3.5	Représentation superposée des lignes et des colonnes	74
2.4	Aides à l'interprétation	79
2.4.1	Inerties associées aux axes (valeurs propres)	79
2.4.2	Contribution d'un point à l'inertie d'un axe	82
2.4.3	Qualité de représentation d'un point par un axe ou un plan	83
2.4.4	Distance et inertie dans l'espace initial	84
2.5	Éléments supplémentaires (= illustratifs)	85
2.6	Mise en œuvre avec FactoMineR	88
2.7	AFC et traitement de données textuelles	89
2.8	Exemple : données Jeux Olympiques	93
2.8.1	Description des données	93
2.8.2	Problématique	94
2.8.3	Choix de l'analyse	96
2.8.4	Mise en œuvre de l'analyse	96
2.9	Exemple : 10 vins blancs du Val de Loire	103
2.9.1	Description des données - problématique	103
2.9.2	Marges	105
2.9.3	Inerties	106
2.9.4	Représentation sur le premier plan	107
2.10	Exemple : causes de mortalité des Français	109
2.10.1	Description des données - problématique	109
2.10.2	Marges	111
2.10.3	Inerties	112
2.10.4	Premier axe factoriel	115
2.10.5	Plan 2-3	117

2.10.6	Projection d'éléments supplémentaires	121
2.10.7	Conclusion	125

3 Analyse des Correspondances Multiples (ACM) 127

3.1	Données - notations	127
3.2	Objectifs	128
3.2.1	Étude des individus	129
3.2.2	Étude des variables et des modalités	129
3.3	Distance entre individus et distance entre modalités	130
3.3.1	Distance entre individus	130
3.3.2	Distance entre modalités	130
3.4	AFC sur le tableau disjonctif complet	132
3.4.1	Lien entre ACM et AFC	132
3.4.2	Nuage des individus	133
3.4.3	Nuage des variables	134
3.4.4	Nuage des modalités	135
3.4.5	Relations de transition	138
3.5	Aides à l'interprétation	140
3.5.1	Indicateurs numériques	140
3.5.2	Éléments supplémentaires	142
3.5.3	Description automatique des axes	144
3.6	Mise en œuvre avec FactoMineR	144
3.7	Compléments	148
3.7.1	Analyse d'une enquête	148
3.7.2	Description d'une variable qualitative et d'une sous-population	150
3.7.3	Le tableau de Burt	154
3.7.4	Données manquantes	155
3.8	Enquête sur la perception des OGM	156
3.8.1	Description des données - problématique	156
3.8.2	Choix de l'analyse et mise en œuvre	159
3.8.3	Analyse du premier plan	160
3.8.4	Projection de variables supplémentaires	162
3.8.5	Conclusion	163
3.9	Exemple : catégorisation	163
3.9.1	Description des données - problématique	163
3.9.2	Choix de l'analyse	165
3.9.3	Représentation des individus sur le premier plan	165
3.9.4	Représentation des modalités	166
3.9.5	Représentation des variables	167

4	Classification	169
4.1	Données - problématique	169
4.2	Formalisation de la notion de ressemblance	173
4.2.1	Ressemblance entre individus	173
4.2.2	Ressemblance entre groupes d'individus	176
4.3	Construction d'une hiérarchie indicée	177
4.3.1	Algorithme classique de construction ascendante	177
4.3.2	Hiérarchie et partition (figure 4.6)	177
4.4	Méthode de Ward	178
4.4.1	Qualité d'une partition	179
4.4.2	Agrégation par l'inertie	181
4.4.3	Deux propriétés de l'indice d'agrégation	182
4.4.4	Analyse d'une hiérarchie, choix d'une partition	183
4.5	Recherche d'une partition par agrégation autour des centres mobiles	184
4.5.1	Données - problématique	184
4.5.2	Principe	185
4.5.3	Méthodologie	186
4.6	Partitionnement et classification hiérarchique	186
4.6.1	Consolidation d'une partition	187
4.6.2	Algorithme mixte	187
4.7	Classification et analyse factorielle	187
4.7.1	Analyse factorielle en amont d'une CAH	188
4.7.2	Analyse simultanée d'un plan factoriel et d'une hiérarchie .	189
4.8	Classification et données manquantes	189
4.9	Exemple : données températures	189
4.9.1	Description des données - problématique	189
4.9.2	Choix de l'analyse	190
4.9.3	Mise en œuvre	190
4.10	Exemple : données thé	195
4.10.1	Description des données - problématique	195
4.10.2	Construction de la CAH	196
4.10.3	Description des classes	197
4.11	Exemple : découpage en classes de variables quantitatives	200
4.11.1	Découpage en classes d'une variable	200
4.11.2	Découpage automatique de plusieurs variables	202
5	Visualisation de données issues d'une analyse factorielle	203
5.1	Données - problématiques	203
5.2	Visualisation de données issues d'une ACP	204
5.2.1	Sélection d'un sous-ensemble d'objets - nuage des individus	204
5.2.2	Sélection d'un sous-ensemble d'objets - nuage des variables	206
5.2.3	Intégration d'informations supplémentaires	206
5.3	Visualisation de données issues d'une AFC	208

5.3.1	Sélection d'un sous-ensemble d'objets, nuage des lignes ou des colonnes	209
5.3.2	Intégration d'informations supplémentaires	210
5.4	Visualisation de données issues d'une ACM	211
5.4.1	Sélection d'un sous-ensemble d'objets - nuage des individus	211
5.4.2	Sélection d'un sous-ensemble d'objets - nuage des modalités	212
5.4.3	Sélection d'un sous-ensemble d'objets - nuages des variables	213
5.4.4	Intégration d'informations supplémentaires	213
5.5	Alternatives aux fonctions graphiques du package FactoMineR . . .	214
5.5.1	Le package Factoshiny	214
5.5.2	Le package factoxtra	214
5.6	Amélioration des graphes par les arguments communs aux fonctions graphiques de FactoMineR	216
A	Annexe	219
A.1	Pourcentage d'inertie expliqué par un axe et par un plan	219
A.2	Le logiciel R	224
A.2.1	Présentation générale	224
A.2.2	Le package Rcmdr	228
A.2.3	Le package FactoMineR	230
	Bibliographie	235
	Index	237