

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	26
1.8.4	Données manquantes	27
1.8.5	Jeu de données de grandes dimensions	27
1.8.6	Rotation varimax	28
1.9	Exemple : données dépenses des ménages	28
1.9.1	Description des données	28
1.9.2	Problématique	30
1.9.3	Choix de l'analyse	30
1.9.4	Mise en œuvre de l'analyse	32
1.10	Exemple : données températures	45

1.10.1	Description des données - problématique	45
1.10.2	Choix de l'analyse	45
1.10.3	Mise en œuvre avec FactoMineR	47
1.11	Exemple : données génomiques	52
1.11.1	Description des données - problématique	52
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	94
2.8.1	Description des données	94
2.8.2	Problématique	95
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	110
2.10.1	Description des données - problématique	110
2.10.2	Marges	112
2.10.3	Inerties	113
2.10.4	Premier axe factoriel	116
2.10.5	Plan 2-3	118

2.10.6	Projection d'éléments supplémentaires	122
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
3.8	Enquête sur la perception des OGM	155
3.8.1	Description des données - problématique	155
3.8.2	Choix de l'analyse et mise en œuvre	158
3.8.3	Analyse du premier plan	159
3.8.4	Projection de variables supplémentaires	161
3.8.5	Conclusion	162
3.9	Exemple : catégorisation	162
3.9.1	Description des données - problématique	162
3.9.2	Choix de l'analyse	164
3.9.3	Représentation des individus sur le premier plan	164
3.9.4	Représentation des modalités	165
3.9.5	Représentation des variables	166
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)	178
4.4	Méthode de Ward	179
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	183
4.4.4	Analyse d'une hiérarchie, choix d'une partition	183
4.5	Recherche d'une partition par agrégation autour des centres mobiles	185
4.5.1	Données - problématique	185
4.5.2	Principe	185
4.5.3	Méthodologie	186
4.6	Partitionnement et classification hiérarchique	187
4.6.1	Consolidation d'une partition	187
4.6.2	Algorithme mixte	187
4.7	Classification et analyse factorielle	188
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	Exemple : données températures	189
4.8.1	Description des données - problématique	189
4.8.2	Choix de l'analyse	189
4.8.3	Mise en œuvre	190
4.9	Exemple : données thé	195
4.9.1	Description des données - problématique	195
4.9.2	Construction de la CAH	196
4.9.3	Description des classes	197
4.10	Exemple : découpage en classes de variables quantitatives	200
4.10.1	Découpage en classes d'une variable	200
4.10.2	Découpage automatique de plusieurs variables	202
A	Annexe	203
A.1	Pourcentage d'inertie expliqué par un axe et par un plan	203
A.2	Le logiciel R	208
A.2.1	Présentation générale	208
A.2.2	Le package Rcmdr	212
A.2.3	Le package FactoMineR	214
	Bibliographie	219
	Index	221