

ANALYSES DISCRIMINANTES REGULARISEES VIA LA REGRESSION PLS ET L'ANALYSE EN RESULTANTES COVARIANTES

Xavier Bry

Professeur au Département de Statistique de l'E.N.E.A.,
BP 5420, Dakar Fann, Sénégal. E-mail : bryxavier@yahoo.fr

Résumé :

Ce travail part du fait que l'Analyse Factorielle Discriminante (AFD) est un cas particulier de deux familles de méthodes : l'une comprenant l'Analyse Canonique (AC) et la régression PLS, et l'autre étant celle de l'Analyse en Résultantes Covariantes (ARC). Ces méthodes, en élargissant le cadre instrumental de l'Analyse Discriminante, permettent d'obtenir des facteurs discriminants privilégiant les structures fortes présentées par les variables explicatives, et par conséquent plus robustes, et d'interprétation plus riche.

On montre qu'à l'instar de l'AFD, les adaptations de PLS et de l'ARC à l'Analyse Discriminante peuvent être considérées comme des Analyses en Composantes Principales du nuage des centres de classes, avec une métrique et un système de poids particuliers pour chaque méthode.

Le formalisme élargi fourni par chaque famille de méthodes permet en outre d'étendre l'AFD à la discrimination de plusieurs variables nominales, en tenant compte des liaisons que celles-ci présentent. On obtient alors une Analyse Multi-Discriminante.

Mots-clefs : *Analyse Discriminante, Analyse Multi-Discriminante, Analyse en Résultantes Covariantes, Régression PLS, Analyse Canonique, Analyse des Correspondances Multiples, Robustesse.*

Introduction

L'Analyse Factorielle Discriminante, recherchant ses facteurs discriminants dans le sous-espace engendré par les variables explicatives, accorde *a priori* la même importance à l'information résiduelle qu'aux structures de corrélation que ces variables contiennent.

On cherche ici, de façon empirique, à obtenir des facteurs régularisés, i.e. plus axés sur les structures explicatives fortes, donc moins vulnérables au bruit et plus facilement interprétables.

Nous redonnons ci-dessous une présentation rapide des deux programmes fondamentaux dont les méthodes découlent. Une présentation plus complète pourra être trouvée dans [BRY01].

1. Les programmes

1.1. le programme Q :

$$Q : \underset{\substack{u^T Mu = 1 \\ v^T Nv = 1}}{\text{Max}} (YMu | ZNv)$$

La quantité à maximiser est un produit de trois facteurs : $\|YMu\|$, $\|ZNv\|$ et $\cos(YMu, ZNv)$. Les deux premiers mesurent la part de l'information des groupes Y et Z captée respectivement par les facteurs F et G . Le troisième mesure la corrélation linéaire entre les facteurs F et G .

Les conditions du premier ordre fournissent deux systèmes équivalents caractérisant les vecteurs solutions (cf. [BRY01]):

$$\text{caractérisation (a) : } (1) MY^N ZNv = \sqrt{\eta} Mu ; \quad (1') NZ^M YMu = \sqrt{\eta} Nv$$

$$\text{caractérisation (b) : } (2) MY^M R_Z^N YMu = \eta Mu ; \quad (2') NZ^N R_Y^M ZNv = \eta Nv$$

Les produits $Y(1)$, $Z(1')$, $Z(2')$ et $Y(2)$ permettent d'en déduire deux caractérisations équivalentes des facteurs solutions:

$$\text{caractérisation (a') : } (3) R_Y^M G = \sqrt{\eta} F ; \quad (3') R_Z^N F = \sqrt{\eta} G$$

$$\text{caractérisation (b') : } (4) R_Y^M R_Z^N F = \eta F ; \quad (4') R_Z^N R_Y^M G = \eta G$$

où η est la plus grande valeur propre des opérateurs $R_{Z^N}^N R_{Y^M}^M$ et $R_{Y^M}^M R_{Z^N}^N$.

- Les matrices dont les F (resp. G) sont les vecteurs propres ne sont pas systématiquement symétriques. Par conséquent, les F (resp. G) ne sont pas toujours orthogonaux entre eux. Dans le cas où l'on cherche des facteurs décorrélés, il peut alors être nécessaire, une fois obtenu le premier couple (F_1, G_1) , de reprendre le programme de maximisation sous contrainte d'orthogonalité des solutions F et G aux facteurs F_1 et G_1 respectivement, et réitérer le procédé à chaque étape¹.
- Pour l'interprétation des résultats, on formera les graphiques (F_α, F_β) et (G_α, G_β) et on examinera la qualité de leur recollement. On pourra aussi former les graphiques (F_α, G_α) , afin de mettre en évidence les individus qui renforcent la corrélation entre ces deux facteurs, et ceux qui l'affaiblissent.

1.2. Le programme **P** (Analyse en Résultantes Covariantes)

- Rappelons que la résultante d'un facteur H sur un groupe X au sens d'une métrique A , égale à $XAX'H$ et notée $R_X^A H$, mesure l'accord de H avec X de deux façons : sa direction donne la dimension de cet accord dans le groupe, et sa norme en mesure l'intensité (cf.[BRY01]).
- Dès lors, on considère le programme :

$$\mathbf{P} : \underset{\substack{F^M F=1 \\ G^N G=1}}{\text{Max}} \langle R_Y^M F | R_Z^N G \rangle$$

Comme pour **Q**, la quantité à maximiser est un produit de trois facteurs : $\|R_Y^M F\|$, $\|R_Z^N G\|$ et $\cos(R_Y^M F, R_Z^N G)$. Les deux premiers mesurent la part de l'information des groupes Y et Z captée respectivement par les facteurs F et G . Le troisième mesure la corrélation linéaire entre ces parts d'information. On parvient aisément à deux caractérisations équivalentes des solutions (cf. [BRY01]):

¹ On peut également, ce qui n'est pas toujours équivalent, remplacer dans l'étape courante de ce programme les groupes Y et Z par leur projection sur le sous-espace orthogonal aux facteurs précédemment obtenus.

Vecteurs u et v :

$$\text{caractérisation (c) : } (5) \quad MY^N R_Z^N ZNv = \sqrt{\eta} Mu ; \quad (5') \quad NZ^M R_Y^M YMu = \sqrt{\eta} Nv$$

$$\text{caractérisation (d) : } (6) \quad MY^N R_Z^N R_Z^N R_Y^M YMu = \eta Mu ; \quad (6') \quad NZ^M R_Y^M R_Y^M R_Z^N ZNv = \eta Nv$$

Soit, pour les facteurs F et G :

$$\text{caractérisation (c') : } (7) \quad R^{M,y} R^{N,z} G = \sqrt{\eta} F ; \quad (7') \quad R^{N,z} R^{M,y} F = \sqrt{\eta} G$$

$$\text{caractérisation (d') : } (8) \quad R_Y^M R_Z^N R_Z^N R_Y^M F = \eta F ; \quad (8') \quad R_Z^N R_Y^M R_Y^M R_Z^N G = \eta G$$

avec η maximale.

- Les facteurs F (resp. G) étant vecteurs propres de matrices symétriques, ils sont 2 à 2 orthogonaux.
- Pour l'interprétation des résultats, on procédera de même que pour le programme **Q**.
- En outre, les graphiques couplant les résultantes de même ordre ($R_Y^M F_\alpha, R_Z^N G_\alpha$) montrent quelles sont les observations qui renforcent la corrélation entre résultantes, et celles qui l'affaiblissent.

1.3. Choix des métriques M et N

1.3.1. Les groupes Y et Z ne sont pas structurés en sous-groupes

a) Il est montré dans [TEN98] pour le programme **Q** et [BRY01] pour **P** que :

- Si l'on veut faire intervenir pleinement la structure de corrélation interne du groupe Y (resp. Z), on peut prendre $M = I_J$ (resp. $N = I_K$). Les variables de chaque groupe étant réduites, l'opérateur de résultante correspondant est la somme des projecteurs orthogonaux sur chacune des variables du groupe.
- Si l'on veut gommer entièrement la structure interne d'un groupe Y (resp. Z), on prendra $M = (Y'Y)^{-1}$ (resp. $N = (Z'Z)^{-1}$). L'opérateur de résultante correspondant est le projecteur orthogonal sur le sous-espace engendré par les variables du groupe.

b) On notera qu'il existe une gradation entre ces deux attitudes extrêmes. Prenons le groupe Y et un entier k compris entre 1 et J . On peut former un opérateur de résultante R_k en sommant les projecteurs orthogonaux sur les sous-espaces engendrés respectivement par tous les k -uples de variables de Y . Le choix $k = 1$ correspond à la prise en compte maximale de la structure interne de Y , et plus k est grand, moins le rôle de cette structure est important, jusqu'à s'annuler pour $k=J$.

Supposons par exemple que les J variables soient composées de $J-1$ variables de structure et d'une variable de bruit b . La résultante R_k implique des projecteurs faisant intervenir le bruit b , et d'autres calculés sans lui. Le nombre des projecteurs faisant intervenir b est celui des $k-1$ -uples formables à partir des variables de structure, soit: $\binom{k-1}{J-1}$. Le nombre total des projecteurs composant la résultante est celui des k -uples de variables, soit: $\binom{k}{J}$. La proportion des projecteurs faisant intervenir b est donc $\binom{k-1}{J-1} / \binom{k}{J} = \frac{k}{J}$. L'influence du bruit ainsi mesurée est minimale pour $k = 1$, et maximale pour $k = J$.

La métrique correspondant à l'opérateur de résultante R_k se déduit très facilement de l'expression de l'opérateur, mais son écriture est lourde.

La possibilité évoquée ici n'est que théorique, sa mise en œuvre étant limitée par une explosion combinatoire.

1.3.2. Les groupes sont structurés en sous-groupes

Formellement, on considère ici que les groupes Y et Z sont composés de sous-groupes. On notera ces décompositions en sous-groupes: $Y = (Y_I, \dots, Y_h, \dots, Y_H)$; $Z = (Z_I, \dots, Z_l, \dots, Z_L)$. On peut envisager deux possibilités: faire intervenir chaque sous-groupe comme sous-espace ou comme structure de corrélations.

1.3.2.1. Chaque sous-groupe n'intervient dans l'analyse que par le sous-espace qu'il engendre.

a) La structure de corrélations interne de chaque sous-groupe est donc gommée. Les métriques pondérant les variables de ces sous-groupes dans les programmes **P** et **Q** sont alors les métriques bloc-diagonales:

$$M = \begin{pmatrix} (Y_1' Y_1)^{-1} & & \\ & \ddots & \\ & & (Y_H' Y_H)^{-1} \end{pmatrix} ; \quad N = \begin{pmatrix} (Z_1' Z_1)^{-1} & & \\ & \ddots & \\ & & (Z_L' Z_L)^{-1} \end{pmatrix}$$

Dans la suite de cet article, seul ce traitement des sous-groupes sera considéré.

b) Ce qui précède fournit immédiatement un traitement naturel des variables qualitatives. Une variable qualitative étant représentée par le sous-espace engendré par ses indicatrices, il suffira de considérer l'ensemble de ses indicatrices comme un sous-groupe.

N.B. : Le sous-espace engendré par les indicatrices d'une même variable contient toujours le vecteur (noté e) dont toutes les composantes valent 1. En réalité, seul le projeté de ce sous-espace sur $\langle e \rangle^\perp$ est à considérer. On utilisera donc les indicatrices centrées. Comme ces dernières ne sont alors plus linéairement indépendantes, il faut ôter une indicatrice par sous-groupe pour pouvoir calculer les projecteurs sur le sous-espace correspondant.

1.3.2.2. Chaque sous-groupe intervient dans l'analyse comme structure de corrélations

On tient ici compte des liaisons internes à chaque sous-groupe. Dans ce cas, les contributions des différents sous-groupes doivent être harmonisées, afin notamment que les sous-groupes les plus fournis ne pèsent pas *a priori* plus lourd sur les résultats de l'analyse. Cette démarche est au fondement de l'Analyse Factorielle Multiple ([ESC90]). On pondérera chaque sous-groupe Y_n (resp. Z_i) par l'inverse de la première valeur propre λ_1^n (resp. μ_1^i) de son ACP (cf. [BRY01],[ESC90]). On obtient ainsi $M = \text{diag}(\text{diag}(\lambda_1^n))_n$; et $N = \text{diag}(\text{diag}(\mu_1^i))_i$.

2. Analyses Discriminantes

Dans ce paragraphe, Z est réduit à un seul sous-groupe formé des indicatrices d'une variable qualitative.

Lemme : pour les programmes **P** et **Q**, dès lors que les variables de Y sont centrées et que le groupe Z n'intervient que par le sous-espace qu'il engendre, il est indifférent de prendre les indicatrices de Z centrées ou non.

Dans la suite, en particulier pour la détermination des ACP équivalentes sur les centres de gravité, il sera commode de prendre les indicatrices non centrées.

Preuve : Lorsque $N = (Z'Z)^{-1}$, le programme **Q** : $\underset{\substack{u'Mu=1 \\ v'Nv=1}}{\text{Max}} \langle YMu | ZNv \rangle$ équivaut à $\underset{\substack{u'Mu=1 \\ w'Z'Zw=1}}{\text{Max}} \langle YMu | Zw \rangle$

(il suffit pour le voir de poser $w = (Z'Z)^{-1}v$). Supposons les indicatrices de Z non centrées. Le sous-espace $\langle Z \rangle$, qui contient e , peut être écrit comme somme de deux sous-espaces orthogonaux: $\langle Z \rangle = \langle e \rangle \oplus (\langle Z \rangle \cap \langle e \rangle^\perp)$. Par conséquent, Zw s'écrit de manière unique $\xi e + Z^*w$, où Z^* est la projection orthogonale de Z sur $\langle Z \rangle \cap \langle e \rangle^\perp$, i.e. la matrice des indicatrices centrées. Dès lors : $\langle YMu | Zw \rangle = \langle YMu | \xi e + Z^*w \rangle = \langle YMu | Z^*w \rangle$ car $Y \perp e$. Or, si $\xi \neq 0$: $\|\xi e + Z^*w\|^2 = \|\xi e\|^2 + \|Z^*w\|^2 = 1$ implique $\|Z^*w\|^2 < 1$. Le critère $\langle YMu | Zw \rangle = \langle YMu | Z^*w \rangle$ est donc amélioré strictement par $\left\langle YMu \left| \frac{Z^*w}{\|Z^*w\|} \right. \right\rangle$. Ceci montre que le vecteur Zw solution du programme est un vecteur normé de $\langle Z^* \rangle$.

Il en va exactement de même pour le programme **P** qui, lorsque $N = (Z'Z)^{-1}$, s'écrit:

$\underset{\substack{F'F=1 \\ G'G=1}}{\text{Max}} \langle R_Y^M F | \Pi_{\langle Z \rangle} G \rangle$. La solution G est trivialement un vecteur normé de $\langle Z \rangle$, et la solution F un

vecteur de $\langle Y \rangle$. A l'aide du raisonnement précédent, on montre que $G \in \langle Z^* \rangle$, sinon le critère peut être strictement amélioré.

2.1. Analyse Factorielle Discriminante

Le groupe Y est lui aussi considéré comme un seul sous-groupe. Il peut contenir des variables continues comme des indicatrices de variables qualitatives (dans ce dernier cas, toutefois, il est nécessaire d'éliminer les multicollinéarités en ôtant une modalité par variable qualitative).

Les opérateurs de résultante sur Y et Z sont alors réduits aux projecteurs orthogonaux sur $\langle Y \rangle$ et $\langle Z \rangle$, respectivement.

Les programmes **P** et **Q** ont la même solution (cf. [BRY01]), car ils se réduisent tous deux à l'Analyse Canonique de Y et Z , i.e. à l'Analyse Factorielle Discriminante.

On sait que l'AFD équivaut à l'ACP du nuage des centres de gravité des classes dans R^J muni de la métrique $(Y'Y)^{-1}$. Cette métrique a pour effet de neutraliser la structure de corrélation interne du groupe Y .

2.2. AFD régularisées

On choisit ici de prendre en compte la structure de corrélations interne au groupe Y .

Une des premières idées qui viennent à l'esprit est de réaliser une Analyse Factorielle préliminaire du groupe Y afin de séparer la structure du bruit, et d'utiliser les q premiers facteurs de cette AF en remplacement du groupe Y (cf. [CEL94] pour une présentation détaillée). Une telle démarche a pour avantage pratique de fournir une meilleure connaissance du groupe Y avant de l'utiliser pour la discrimination de Z . Mais elle a pour inconvénient de remplacer la hiérarchie des structures de Y par un effet de seuil assez brutal et arbitraire: le q -ième facteur retenu y a exactement la même importance que le premier, tandis qu'à partir du $q+1$ -ième, ils sont réduits au silence.

Ceci incite à utiliser directement une méthode de discrimination incluant l'aspect structurel dans son critère à optimiser. Nous utiliserons alternativement les programmes **Q** et **P**, en y considérant chaque variable de Y comme un sous-groupe. Ainsi, la métrique M réduit chaque variable continue et représente chaque variable qualitative par le sous-espace de ses indicatrices.

2.2.1. Programme **Q**

- Comme $N = (Z'Z)^{-1}$, on a : $R_Z^N = \Pi_{\langle Z \rangle}$. Les facteurs solutions de **Q** sont alors caractérisés par

$$\text{caractérisation (a')} : \quad (3) \quad R_Y^M G = \sqrt{\eta} F \quad ; \quad (3') \quad \Pi_{\langle Z \rangle} F = \sqrt{\eta} G$$

$$\text{ou caractérisation (b')} : (4) \quad R_Y^M \Pi_{\langle Z \rangle} F = \eta F \quad ; \quad (4') \quad \Pi_{\langle Z \rangle} R_Y^M G = \eta G$$

Comme $G \in \langle Z \rangle$, on a : $(4') \Leftrightarrow \Pi_{\langle Z \rangle} YMY' \Pi_{\langle Z \rangle} G = \eta G$. Cette dernière équation est celle de l'Analyse des Redondances Maximales de Y sur Z et fournit G_1 associé à la plus grande valeur propre. Le premier facteur discriminant recherché est alors F_1 , donné par (3).

Dans le cas où M est telle que la résolution de **Q** ne fournit pas d'emblée des facteurs F deux à deux orthogonaux, on projette le groupe Y sur l'orthogonal de F_1 , et on réitère la démarche, qui fournit alors F_2 . Et ainsi de suite. C'est notamment le cas pour le choix $M = I$, caractéristique de PLS.

Dans la suite, la méthode obtenue pour $M = I$ sera notée AD PLS, pour *Analyse Discriminante PLS*.

- L'équation (3') montre qu'à une constante multiplicative près, le facteur G s'obtient en faisant, pour chaque observation, la moyenne des valeurs de F pour les observations ayant la même modalité de Z . Le graphique (G_α, G_β) est identique à la projection des centres de gravités des classes sur le plan (F_α, F_β) .
- A l'étape k , on dispose de k facteurs discriminants robustes. On peut alors reconduire une AFD classique sur ces seuls facteurs, afin d'optimiser le pouvoir discriminant des premiers axes.
- Il a été noté dans [BRY01] et [CAZ77] que si les variables de Y sont toutes qualitatives et que chacune est considérée comme un sous-groupe intervenant par le sous-espace qu'il engendre, le programme **Q** conduit à l'AFC du croisement du groupe Y avec Z , dite *Analyse Discriminante Barycentrique* (cf. [CEL94] [TEN98]). La prise en compte par l'ADB des liaisons internes entre variables de Y , soulignée dans [CEL94], apparaît très simplement *via* le programme **Q**. L'avantage de cette présentation est de faire également apparaître un lien entre l'ADB et l'ACM de Y (cf. §3).
- *ACP équivalente des centres de gravité des classes :*

On considère ici que les indicatrices de Z ne sont pas centrées. On remarque alors que:

$$(2) \Leftrightarrow MY' \Pi_{\langle Z \rangle} YMu = \eta Mu \Leftrightarrow MY' Z(Z'Z)^{-1} Z' YMu = \eta Mu$$

$$\Leftrightarrow MY' Z(Z'Z)^{-1} (Z'Z)(Z'Z)^{-1} Z' YMu = \eta Mu \Leftrightarrow M \Gamma' P_\Gamma \Gamma Mu = \eta Mu$$

avec $\Gamma = (Z'Z)^{-1} Z' Y$ (matrice dont les lignes sont les centres de gravité des classes dans R^J) et $P_\Gamma = Z'Z$ (matrice diagonale des effectifs de classe, i.e. des poids associés aux centres de gravité).

La technique proposée ici revient par conséquent à faire l'ACP du nuage des centres de gravité dans R^J muni de la métrique M . Ceci montre qu'il y aura au plus $K-1$ facteurs discriminants. Ces facteurs engendrent le même sous-espace que ceux de l'AFD classique.

2.2.2. Programme \mathbf{P} (ARC)

• Les facteurs solutions de \mathbf{P} sont ici déterminés par :

$$\text{caractérisation (c')} : \quad (7) \quad R_Y^M \Pi_{\langle Z \rangle} G = \sqrt{\eta} F ; \quad (7') \quad \Pi_{\langle Z \rangle} R_Y^M F = \sqrt{\eta} G$$

$$\text{ou caractérisation (d')} : \quad (8) \quad R_Y^M \Pi_{\langle Z \rangle} R_Y^M F = \eta F ; \quad (8') \quad \Pi_{\langle Z \rangle} R_Y^M R_Y^M \Pi_{\langle Z \rangle} G = \eta G$$

avec η maximale.

• L'équation (7') montre qu'à une constante multiplicative près, le facteur G s'obtient en faisant, pour chaque observation, la moyenne des valeurs de $R_Y^M F$ pour les observations ayant la même modalité de Z . Le graphique (G_α, G_β) est identique à la projection des centres de gravités des classes sur le plan des résultantes $(R_Y^M F_\alpha, R_Y^M F_\beta)$. Si le facteur F_α est proche d'une composante principale du groupe Y , i.e. d'un vecteur propre de R_Y^M , la résultante $R_Y^M F_\alpha$ sera presque colinéaire à F_α . Le graphique (G_α, G_β) ressemblera alors à la projection des centres de gravités des classes sur le plan (F_α, F_β) .

• ACP équivalente des centres de gravité des classes :

Dans le cas présent, et les indicatrices de Z n'étant pas centrées :

$$(6) \Leftrightarrow MY' \Pi_{\langle Z \rangle} YMY' YMu = \eta Mu \Leftrightarrow MY' Z(Z'Z)^{-1} Z' YMY' YMu = \eta Mu$$

$$\Leftrightarrow M\Gamma' P_\Gamma \Gamma MY' YMu = \eta Mu .$$

Soit, en prémultipliant par $MY'Y$, et en notant $\Omega = MY'YM$: $\Omega\Gamma' P_\Gamma \Gamma \Omega u = \eta \Omega u$

La technique proposée revient donc à faire l'ACP du tableau Γ des centres de gravité munis de leurs poids naturels, dans \mathbb{R}^J muni de la métrique Ω . Cette métrique a pour effet de renforcer l'intervention de la structure de corrélation interne de Y . En effet : $\|u\|_\Omega^2 = u' MY' YMu$.

Pour $\|u\|_M^2 = 1$, $\|u\|_\Omega^2$ est l'inertie des individus de Y dans la direction $\langle u \rangle$ de l'espace (\mathbb{R}^J, M) .

Calculons les coordonnées des individus de Y dans l'ACP de Γ avec la métrique Ω :

$$Y\Omega u = YMY' YMu = R_Y^M F$$

Il s'agit de la résultante du facteur F .

2.3. Applications comparées

Nous avons utilisé un petit jeu de données plutôt appétissant, puisqu'il s'agit de 55 recettes tirées au hasard du manuel culinaire [BER87]. Pour chaque recette, nous avons relevé les caractéristiques suivantes (les données sont fournies en annexe):

- la présence d'une phase de marinade (1=oui / 0=non)
- la difficulté (1=élémentaire ; 2=facile ; 3=difficile)
- le coût (1=bon marché ; 2=raisonnable ; 3=cher)
- la possibilité d'utiliser des restes dans la préparation (1=oui / 0=non)
- le caractère complet de la préparation, permettant de l'utiliser comme plat du jour (1=oui / 0=non)
- le caractère raffiné de la recette, permettant de l'utiliser pour une réception (1=oui / 0=non)
- la durée de préparation, cuisson comprise (en minutes)
- le nombre d'étapes de la recette (pris comme indicateur de complexité)
- la place de la préparation dans le repas (1=entrée ; 2=plat de résistance ; 3=dessert ; 4=garniture)
- l'aliment de base² de la préparation (1=oeuf ; 2=poisson/fruits de mer ; 3=viande ; 4=volaille/lapin ; 5=légume/fruit)

On se propose d'abord de chercher quelles sont les grandes différences induites, dans la préparation de la recette et son emploi, par le choix de l'aliment de base. Cette dernière variable sera donc la variable Z à discriminer, et les autres formeront le groupe Y des variables discriminantes. Les variables ordinales (*difficulté*, *coût*) ainsi que les indicatrices (*marinade*, *reste*, *plat du jour*, *réception*, mais aussi les indicatrices de la variable *place* : *entrée* ; *plat de résistance* ; *dessert* ; *garniture*) seront considérées comme quantitatives.

² Pour certaines préparations faisant intervenir plusieurs ingrédients dans des proportions équilibrées, la désignation de l'aliment de base est assez arbitraire.

On soumet ces données à l'AFD, à la régression PLS des indicatrices de Z, à l'AD PLS et à l'ARC, afin d'en comparer les résultats³.

N.B. Tous les facteurs ont été *réduits* pour les représentations graphiques, de sorte à permettre la comparaison des facteurs homologues fournis par les différentes techniques.

2.3.1. Comparaison numérique

Comparons numériquement les résultats des différentes techniques. On examine d'abord les corrélations entre facteurs de même rang (**tableaux 1 à 4**).

Tableau 1 : Corrélations entre facteurs de rang 1

	F1 AFD	F1 PLS indic	F1 ADPLS	F1 ARC
F1 AFD	1	0,912	0,925	0,919
F1 PLS indic		1	0,996	0,999
F1 ADPLS			1	0,999
F1 ARC				1

Le **tableau 1** montre des corrélations entre facteurs de rang 1 toutes supérieures à 0.9. Les moins fortes sont celles du facteur 1 d'AFD avec les facteurs 1 des autres méthodes, lesquels sont pratiquement identiques.

Tableau 2 : Corrélations entre facteurs de rang 2

	F2 AFD	F2 PLS indic	F2 ADPLS	F2 ARC
F2 AFD	1	0,888	0,891	0,899
F2 PLS indic		1	0,996	0,998
F2 ADPLS			1	0,998
F2 ARC				1

Il en va de même pour les facteurs de rang 2 (**tableau 2**), à ceci près que les corrélations ont très légèrement baissé.

³ On a également employé l'ARC itérative, i.e. l'algorithme d'extraction de facteurs utilisé par PLS, mais maximisant à chaque étape le critère du programme P à la place de celui du programme Q. Cette méthode donne ici des résultats pratiquement indistinguables de ceux de l'AD PLS.

Tableau 3 : Corrélations entre facteurs de rang 3

	F3 AFD	F3 PLS indic	F3 ADPLS	F3 ARC
F3 AFD	1	0,029	0,118	0,638
F3 PLS indic		1	0,979	0,436
F3 ADPLS			1	0,329
F3 ARC				1

Tableau 4 : Corrélations entre facteurs de rang 4

	F4 AFD	F4 PLS indic	F4 ADPLS	F4 ARC
F4 AFD	1	0,242	0,355	0,675
F4 PLS indic		1	0,988	0,496
F4 ADPLS			1	0,371
F4 ARC				1

Les facteurs de rang 3 et 4 s'avèrent au contraire très instables d'une méthode à l'autre (tableaux 3 et 4). Une exception toutefois : l'AD PLS reste très proche de la régression PLS des indicatrices. Les différences entre l'AFD et les autres méthodes étaient attendues *a priori*. Plus étonnantes sont les divergences entre facteurs de PLS et de l'ARC, puisque ces deux types de méthodes partent d'un principe commun (et donnent le plus souvent, de ce fait, des résultats proches). La corrélation entre le facteur 3 de PLS et F3 ARC est même nettement plus faible que celle de F3 ARC et F3 AFD. L'explication est simple. Le plan (F1,F2) est pratiquement le même pour l'ARC et PLS : les facteurs de rang supérieur à 2 sont donc assujettis à la « même » contrainte d'orthogonalité dans les deux cas. Enfin, les centres de gravité des 5 classes à discriminer engendrant un sous-espace de dimension 4, les facteurs F3 et F4 des deux méthodes se trouveront pratiquement dans le même plan. Or, il apparaît dans le **tableau 5** que les pouvoirs discriminants de F3 et F4 de l'AD PLS (resp. de l'ARC) sont presque égaux, et que les parts de variance de *Y* qu'ils captent respectivement sont très proches également. Ceci rend ces facteurs très faiblement déterminés, individuellement : seul leur plan est fortement déterminé.

Tableau 5 : Pouvoir discriminant et part de variance de Y captée par le facteur

N.B. : Le pouvoir discriminant est le rapport de la variance entre classes à la variance totale.

	F1 AFD	F2 AFD	F3 AFD	F4 AFD
<i>Discrim</i>	0,805	0,420	0,251	0,142
<i>%varY</i>	0,185	0,135	0,057	0,092
	F1 PLS indic	F2 PLS indic	F3 PLS indic	F4 PLS indic
<i>Discrim</i>	0,679	0,368	0,157	0,167
<i>%varY</i>	0,245	0,181	0,109	0,105
	F1 ADPLS	F2 ADPLS	F3 ADPLS	F4 ADPLS
<i>Discrim</i>	0,691	0,351	0,165	0,167
<i>%varY</i>	0,232	0,180	0,104	0,099
	F1 ARC	F2 ARC	F3 ARC	F4 ARC
<i>Discrim</i>	0,685	0,362	0,167	0,158
<i>%varY</i>	0,234	0,176	0,095	0,100

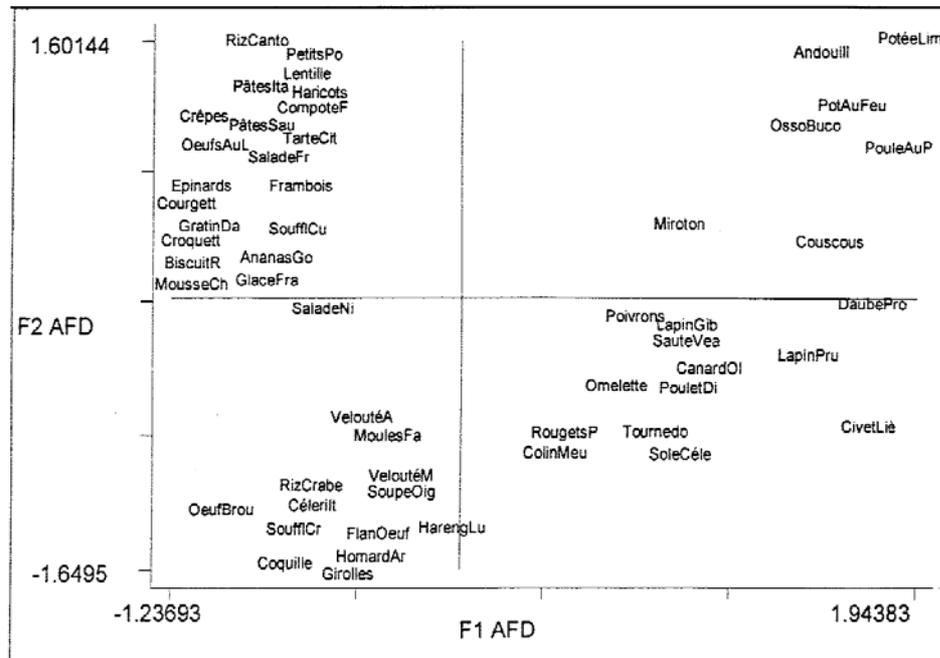
Les meilleurs pouvoirs discriminants sont évidemment ceux de l'AFD, dont ils constituent le critère à maximiser. La baisse de pouvoir discriminant des autres méthodes est compensée par une plus grande part de variance de Y captée par les facteurs.

Les pouvoirs discriminants des axes de même rang fournis par PLS et l'ARC sont très proches. Il en va de même des parts de variance de Y captées par ces facteurs.

2.3.2. Comparaison graphique

Comparons à présent les plans factoriels homologues de l'AFD, de l'AD PLS et de l'ARC. Les facteurs F1 et F2 étant pratiquement identiques pour l'ARC et AD PLS, on ne montrera ici que le plan (1,2) de l'ARC.

Figure 1 : plan (F1,F2) de l'AFD



Les différents aliments sont assez bien discriminés par ce plan, à l'exception des spécialités à base d'œuf, qui sont éparpillées entre les recettes végétariennes et celles de la mer, et de quelques spécialités excentriques qu'on détaille plus loin.

Figure 2 : corrélations des variables discriminantes avec les facteurs F1 et F2 de l'AFD

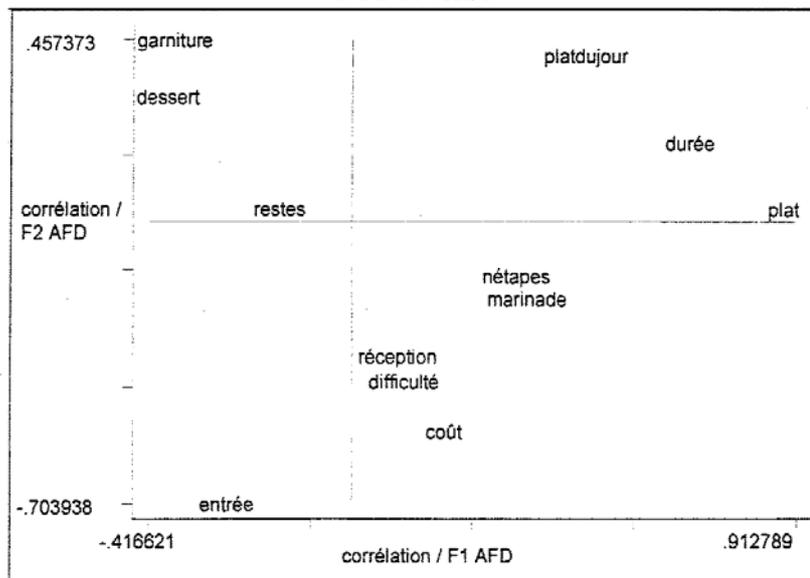
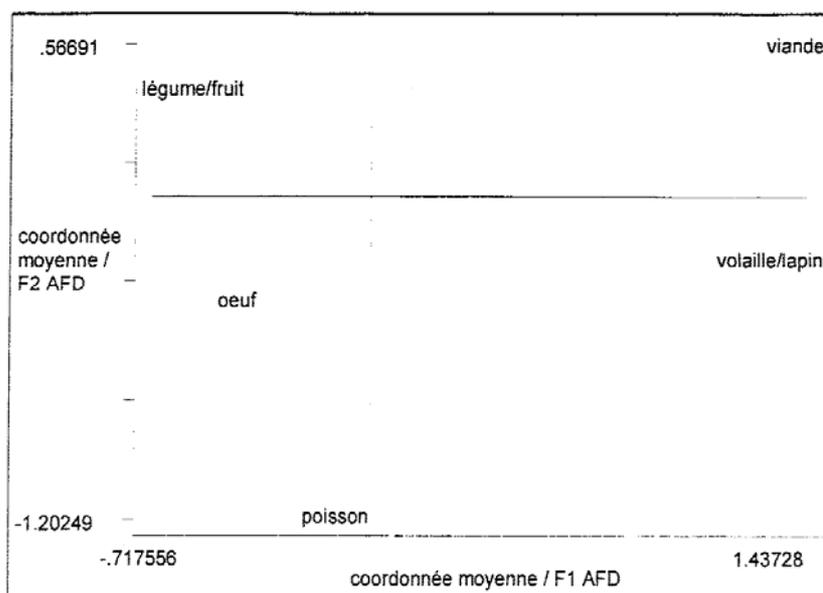


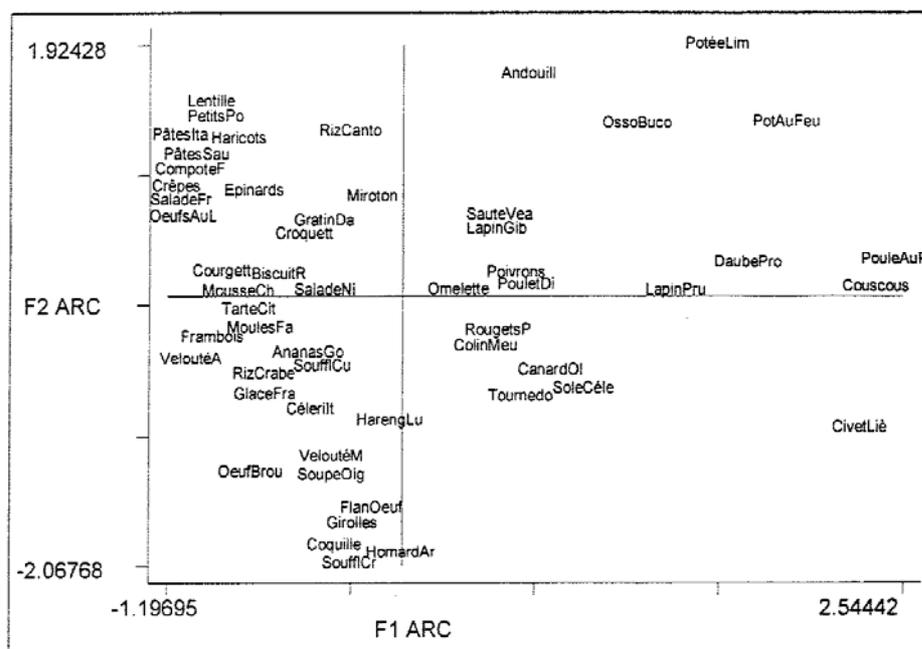
Figure 3 : Projection des centres de classes sur le plan (F1,F2) de l'AFD

Les figures 2 et 3 permettent de donner un contenu à la discrimination. La variable la plus discriminante est clairement l'emploi : ce sont ses modalités qui occupent les positions les plus périphériques. Globalement, les préparations de viandes ont le temps de préparation le plus long et sont utilisées comme plat (voire plat du jour). Les poissons et fruits de mer prennent moins de temps, leur coût est plus élevé, et ils sont servis en entrée. Les recettes de volaille ou lapin occupent une position intermédiaire. Les préparations à base de fruits ou de légumes sont servies en dessert ou garniture. Les recettes à base d'œuf, tantôt entrée, tantôt dessert, se mélangent aux préparations végétariennes et aux recettes de poisson.

On comprend également les spécificités des plats excentriques, tels que les poivrons farcis au riz, les girolles, le céleri à l'italienne, les soupes (assez dissemblables des autres préparations végétariennes, en particulier par leur emploi en plat ou entrée au lieu d'accompagnement ou dessert), et le Tournedos Clamart (différent des autres recettes de viande par son temps de préparation plus court, son coût et sa sophistication plus importants).

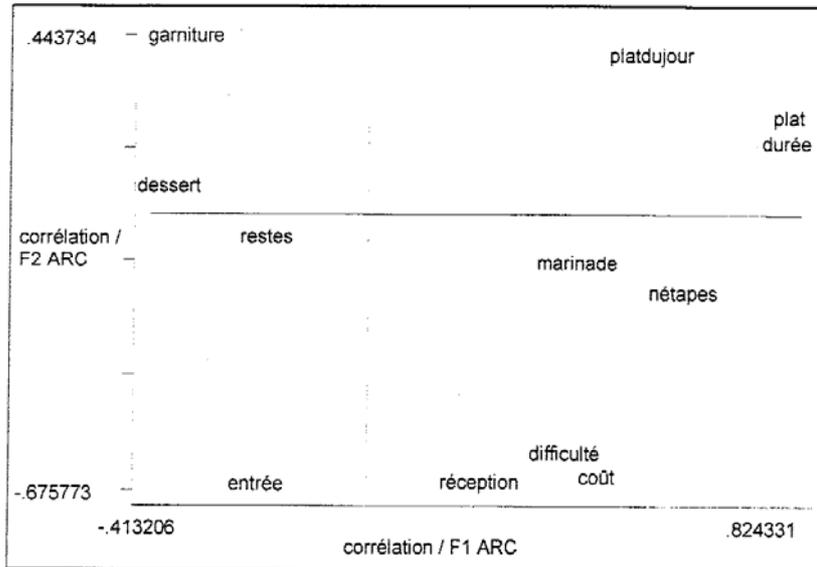
ARC :

La **figure 4** montre une configuration du nuage globalement semblable à celle qu'a donnée l'AFD. Cette similitude va de pair avec celle des **figures 3** et **6**. Quelques variations marginales font la différence : le Miroton s'est séparé des autres recettes de viande ; andouillette, potée limousine, osso buco et pot-au-feu se sont rapprochés du riz cantonnais ; les spécialités de volaille et de lapin occupent une position plus centrale (moins proche des viandes) à l'exception notable de la poule au pot et du civet de lièvre. En bas du graphique, certaines entrées sont un peu mieux distinguées, comme le velouté de moules et la soupe à l'oignon, qui ont pris quelque distance avec le velouté d'asperges, le riz au crabe et les moules farcies.

Figure 4 : plan (F1,F2) de l'ARC

La **figure 5** permet d'expliquer ces variations. Les modalités de la variable *emploi* ne sont plus les seules en position périphérique : du fait des corrélations internes aux variables discriminantes, la *durée* de préparation est venue s'accoler à la modalité *plat*, et les variables

Figure 5 : corrélations des variables discriminantes avec les facteurs F1 F2 de l'ARC



réception, difficulté et coût indiquant une cuisine plus luxueuse se sont excentrées vers le bas. L'interprétation des axes s'en trouve enrichie, le prix en étant une discrimination un peu plus spécifique (les plats à préparation longue sont mis en relief parmi l'ensemble des plats, et les entrées de luxe parmi l'ensemble des entrées).

Figure 6 : Projection des centres de classes sur le plan (F1,F2) de l'ARC

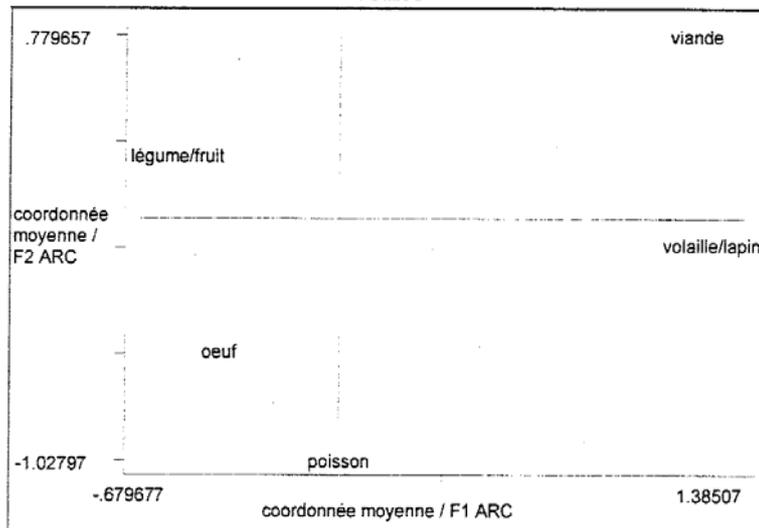
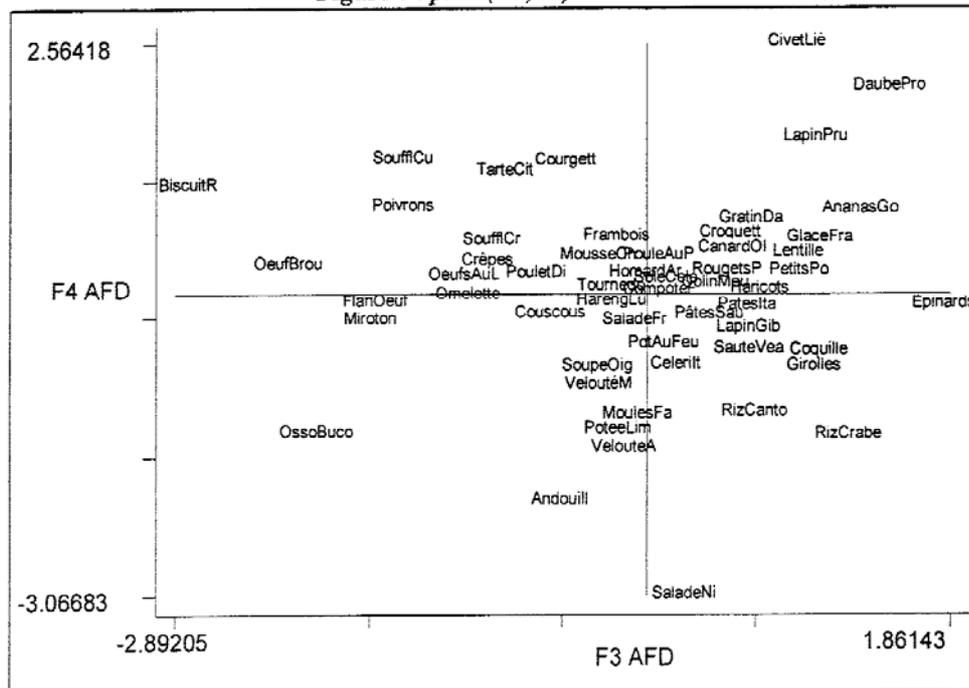
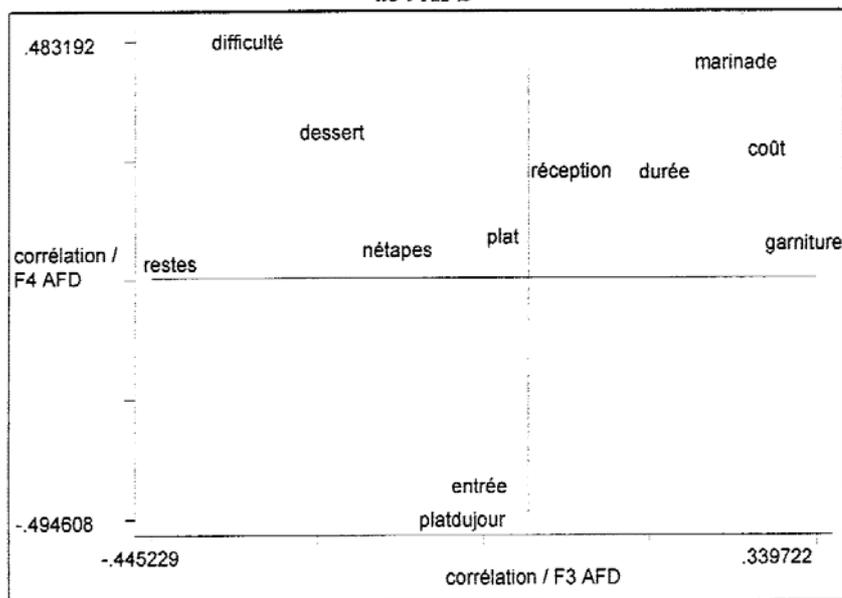


Figure 7 : plan (F3,F4) de l'AFD



Sur le plan (3,4) de l'AFD (fig.7), le mélange des genres est à peu près général : seule la catégorie des préparations à base d'œuf ressort un peu, mais elle se trouve brouillée par des plats tels que poivrons farcis, Miroton...

Figure 8 : corrélations des variables discriminantes avec les facteurs F3 F4 de l'AFD



Les corrélations des variables discriminantes avec les facteurs F3 et F4 sont toutes faibles (fig. 8). Il n'y a pas de variable illustrant correctement l'un des axes ou le plan. L'interprétation des facteurs à partir de ces variables est donc difficile.

aliments dans nos recettes : les préparations à base d'œuf sont ici plus fréquemment des recettes de dessert, et elles permettent plus souvent l'utilisation de restes⁴.

Figure 11 : corrélations des variables discriminantes avec les facteurs F3 F4 de l'ARC

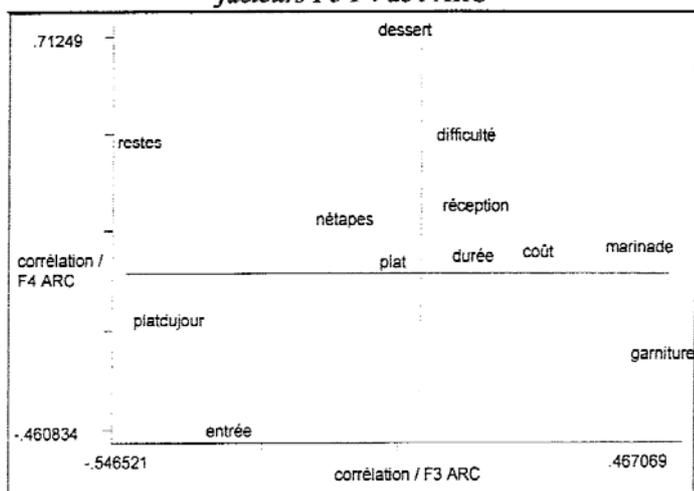
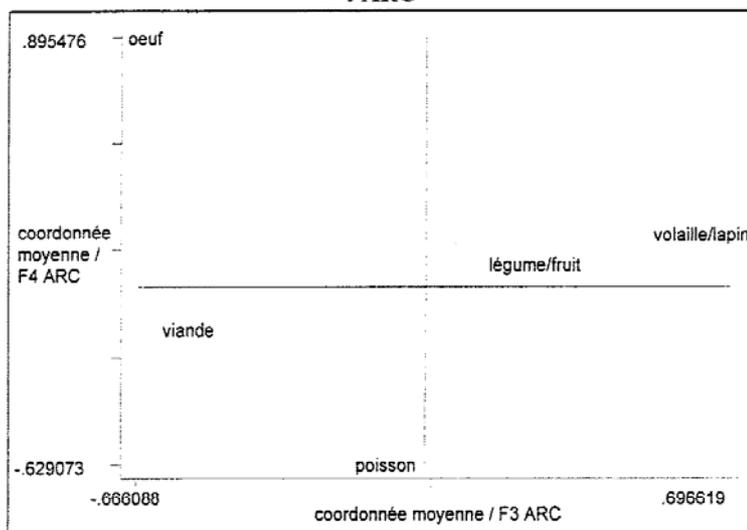


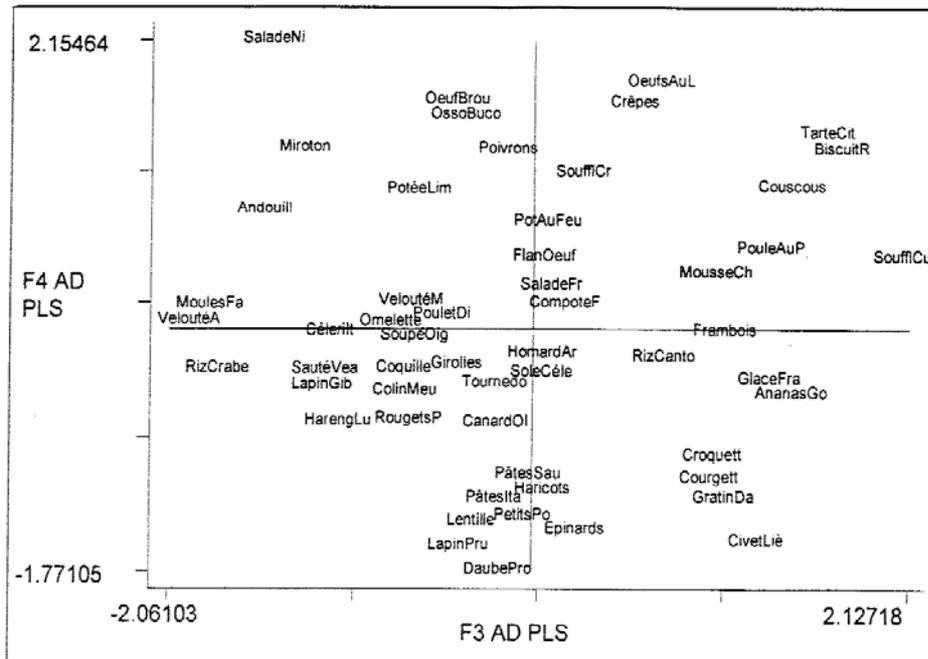
Figure 12 : projection des centres de classes sur le plan (F3,F4) de l'ARC



⁴ Les cuisiniers n'ignorent pas qu'il s'agit des blancs, les jaunes entrant dans la préparation de maintes sauces ; les logiciens objecteront que le jaune également n'est que le reste du blanc, tout dépendant de la priorité : sauce ou dessert ?

AD PLS :

Figure 13 : plan (F3,F4) de l'AD PLS



La comparaison détaillée des facteurs F3 et F4 de l'ARC avec ceux de l'AD PLS s'impose : bien qu'on sache qu'ils engendrent des plans très proches, les corrélations entre facteurs de même rang de ces deux analyses sont très faibles. Examinons les corrélations croisées de ces facteurs :

$$\rho(F3 \text{ ARC}, F4 \text{ AD PLS}) = -0,927 \quad ; \quad \rho(F4 \text{ ARC}, F3 \text{ AD PLS}) = 0,761$$

Très grossièrement, il y a eu interversion des axes entre l'ARC et l'AD PLS. Cela n'a rien d'étonnant, compte tenu de la proximité des pouvoirs discriminants des facteurs de ce plan, ainsi que des parts de variance de Y qu'ils captent. Si l'on examine plus en détail les graphiques, on s'aperçoit que la catégorie *œuf* est mieux discriminée des autres par l'ARC que par l'AD PLS : dans cette dernière, elle occupe une position moins excentrée, est davantage étalée, et se trouve mélangée avec plusieurs autres plats (e. g. *couscous*, *poule au pot*...).

La **figure 14** montre que la variable la plus illustrative de l'axe 4 est *restes* (mais l'illustration n'est pas très bonne). La variable *dessert* est surtout corrélée avec F3, mais intervient également un peu le long de F4. Ces variables sont légèrement moins illustratives de ce plan que du plan homologue de l'ARC.

Figure 14: corrélations des variables discriminantes avec les facteurs F3 F4 de l'AD PLS

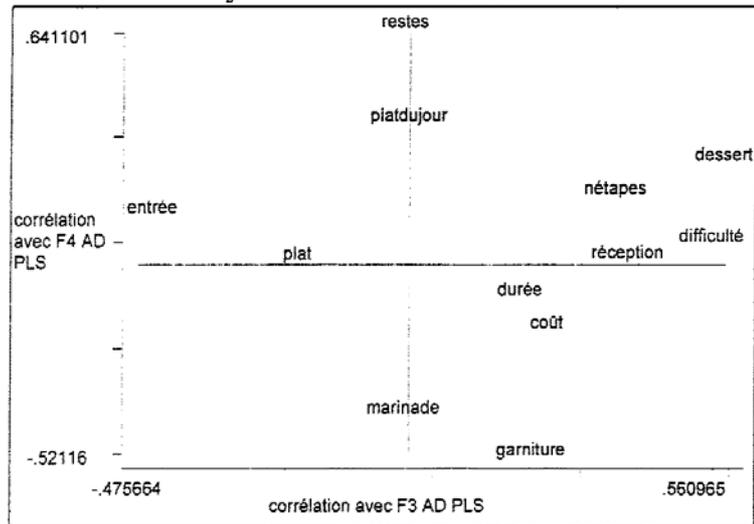
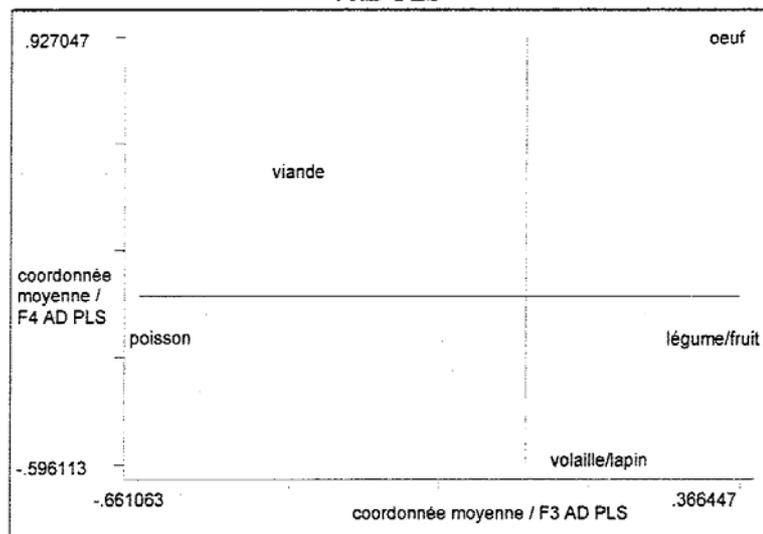


Figure 15 : projection des centres de classes sur le plan (F3,F4) de l'AD PLS



3. Analyses Multi-Discriminantes⁵

• L'utilisation des programmes **P** et **Q** permet de généraliser très simplement le problème de l'AFD au cas où l'on désire discriminer plusieurs variables qualitatives en tenant compte de leurs correspondances.

Le groupe Z est dans ce cas composé de plusieurs variables qualitatives. Chacune de ces variables intervient en tant que sous-espace engendré par ses indicatrices ; l'ensemble des indicatrices de chaque variable est donc considéré comme un sous-groupe. La résultante R_Z^N est la somme des projecteurs orthogonaux sur ces sous-espaces.

Que l'on utilise **P** ou **Q** pour l'analyse, nous allons voir que les correspondances entre les variables de Z sont prises en compte par la résultante du facteur G sur ce groupe.

Concernant le programme **P** : $\text{Max}_{\|F\|=1, \|G\|=1} \langle R_Y^M F | R_Z^N G \rangle$, il est montré dans [BRY01] que $\text{Max}_{\|G\|=1} \|R_Z^N G\|$ conduit à l'Analyse Canonique Généralisée de Z , c'est-à-dire à son Analyse des Correspondances Multiples.

Concernant **Q** : $\text{Max}_{\substack{u' Mu=1; v' Nv=1}} \langle YMu | ZNv \rangle$, on s'intéresse au programme $\text{Max}_{v' Nv=1} \|ZNv\|$. Ce programme est celui de l'ACP de Z avec la métrique N , et sa solution G est caractérisée par : $ZNZ'G = \eta G$. On est donc conduit à diagonaliser R_Z^N , ce qui correspond là encore à l'ACM de Z .

• Les graphiques utilisant les facteurs F permettent de donner un contenu à ces facteurs « explicatifs ». Les graphiques qui utilisent les facteurs G reproduisent la partie des correspondances multiples de Z qu'on a pu mettre en rapport avec Y .

⁵ Les Analyses Multi-Discriminantes proposées dans la suite seront collectivement regroupées sous le sigle AMD. Plus spécifiquement, l'analyse fondée sur le programme **Q** pourra être notée PLSMD (PLS Multi-Discriminante), tandis que son homologue fondée sur **P** sera notée ARCMD (ARC Multi-Discriminante).

3.1. AMD non régularisées

Qu'on utilise le programme **P** ou **Q**, le groupe explicatif Y est considéré ici comme un seul sous-groupe. On y cherche donc des C.L. discriminant Z , mais ne tenant pas compte des structures internes à Y .

3.2. AMD régularisées

- Ici, chaque variable du groupe explicatif Y est considérée comme un sous-groupe, de sorte que les structures de corrélation de Y influent sur les facteurs explicatifs.

Les AMD réalisent un compromis entre trois analyses : l'AF du groupe explicatif (ACP si les variables sont considérées comme quantitatives, ACM si elles sont qualitatives), l'ACM du groupe à expliquer, et l'AC des deux groupes.

- Il a été noté dans [BRY01] et [CAZ77] que si les variables de Y sont toutes qualitatives, le programme **Q** conduit à l'AFC du croisement des groupes Y et Z .

3.3. Exemple d'ARCMD

Nous avons scindé l'ensemble des variables en deux : un groupe de variables qualitatives décrivant les types de recettes, formé des variables : *aliment, emploi, réception, plat du jour*, et un groupe représentant les contraintes de fabrication, et contenant les variables : *marinade, durée de préparation, nombre d'étapes, difficulté, coût, restes*.

On procède d'abord à l'ACM des types de recettes, à l'ACP des contraintes, ainsi qu'à l'AC des deux groupes. Les résultats de ces analyses sont ensuite comparés à ceux de l'ARCMD des deux groupes. On observera notamment les structures dépistées par l'ACM et l'ACP qui apparaissent, subsistent ou disparaissent dans l'ARCMD.

3.3.1. ACM des types de recettes : aliment, emploi, réception, plat du jour

Figure 16 : ACM des types : modalités

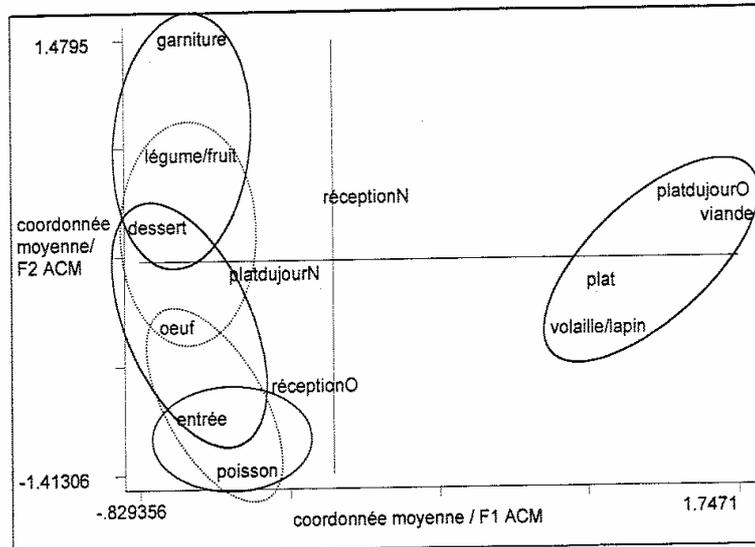
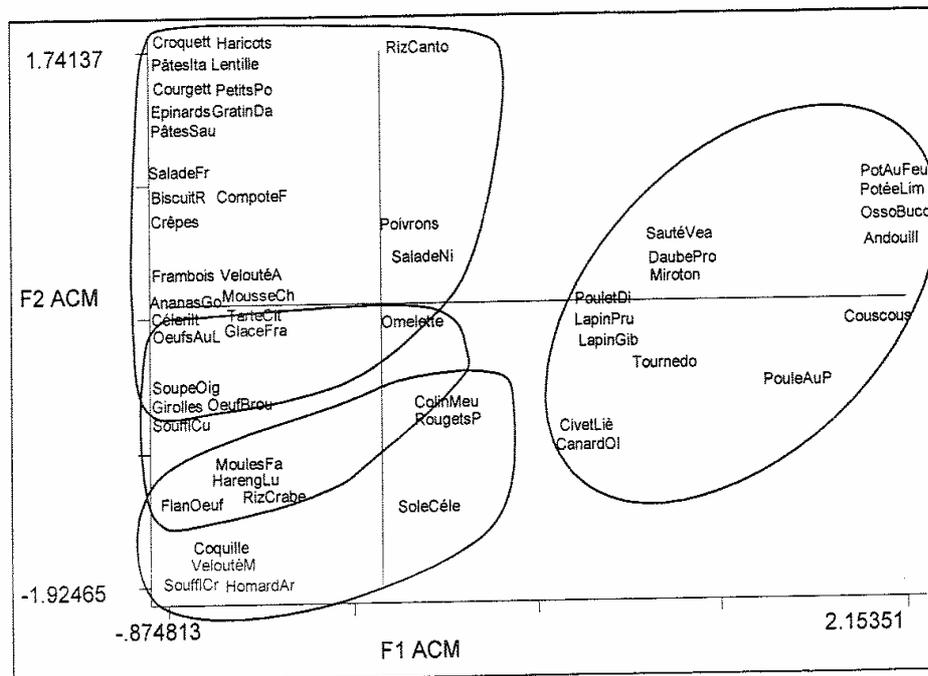


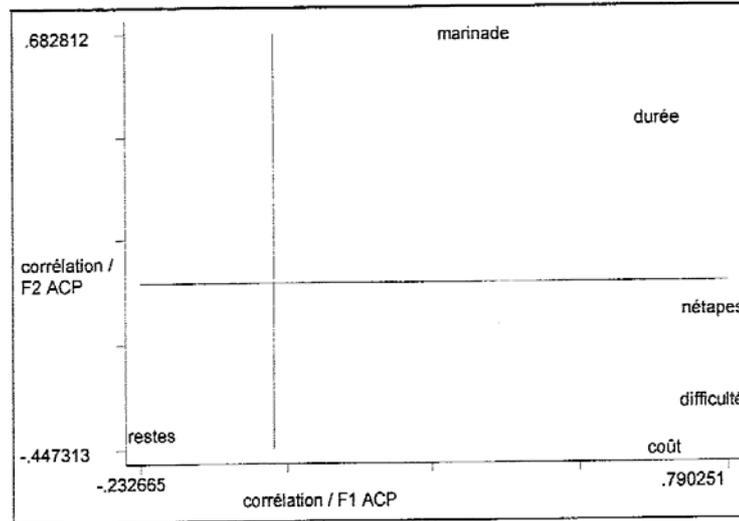
Figure 17 : ACM des types : observations



Sur le plan (1,2) de l'ACM (fig. 16), on constate que les grands types de recettes qui émergent sont structurés par la correspondance entre *aliment*, *emploi* et *plat du jour*. La variable *réception* n'y joue pas un grand rôle.

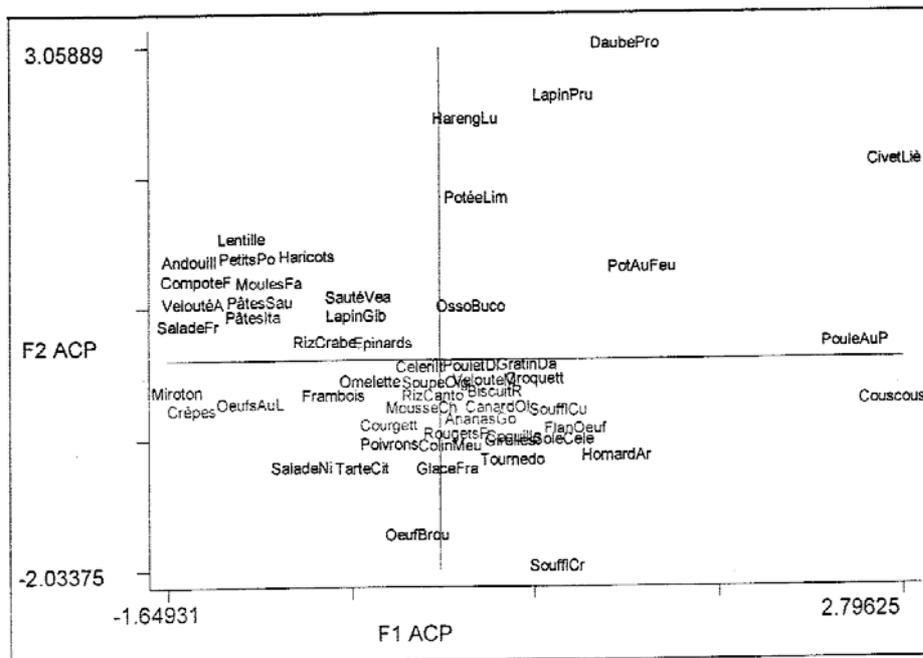
3.3.2. ACP des contraintes : marinade, restes, difficulté, coût, nombre d'étapes, durée de préparation

Figure 18 : ACP des contraintes : variables



Sur le plan (1,2) de cette ACP (fig. 18), les variables sont bien représentées, à l'exception de *restes*. L'axe 1 est bien illustré par *durée*, *nombre d'étapes* et *difficulté*, tandis que l'axe 2 est assez mal illustré, mais distingue *marinade* des autres variables.

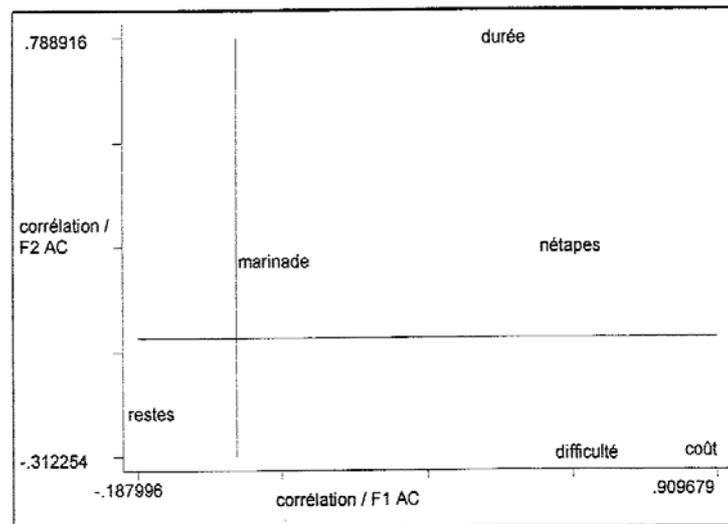
Figure 19 : ACP des contraintes : observations



Les observations situées dans le quadrant Nord-Est ressortent clairement. On y reconnaît essentiellement (à l'exception du *Hareng sauce Lucas*) des recettes à base de viande, volaille ou lapin, servies en plat de résistance. *A contrario*, les observations des trois autres quadrants forment un conglomérat peu lisible car assez concentré vers l'intersection des axes. Les types de recettes y sont mal discriminés.

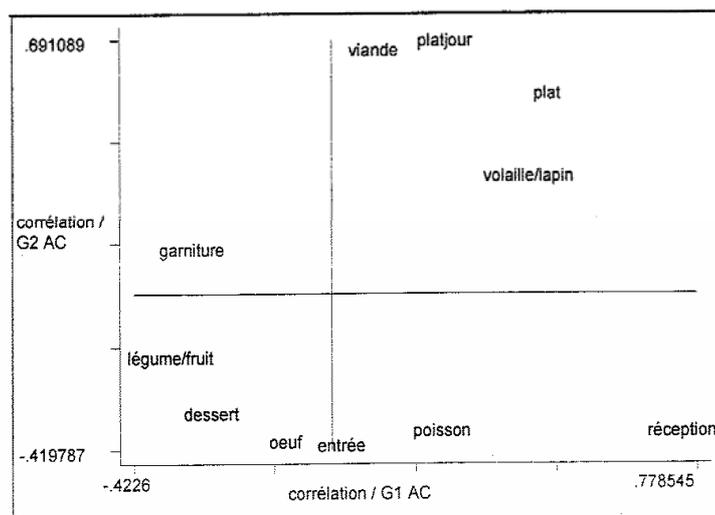
3.3.3. AC des deux groupes

Figure 20 : AC des types et contraintes : contraintes



La figure 20 montre que, par rapport à l'ACP des contraintes, la variable *coût* acquiert dans l'AC une importance particulière (au point de « réclamer » le premier axe), alors que *marinade* n'en a plus aucune. La première a donc un rapport étroit avec certaines variables décrivant les types de recettes, alors que la seconde ne leur est globalement pas liée.

Figure 21 : AC des types et contraintes : types



Par rapport à l'ACM des types de recettes, l'AC met en relief la variable *réception* (fig. 21).

Cela signifie que certaines variables de contrainte peuvent bien la discriminer.

Figure 22 : AC des types et contraintes : observations / contraintes

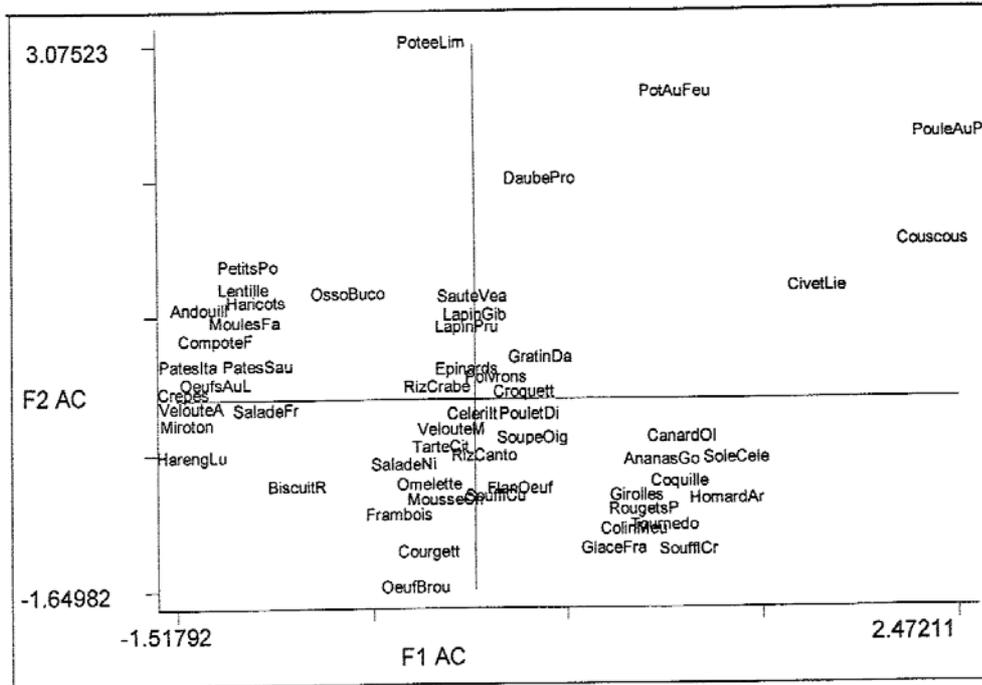
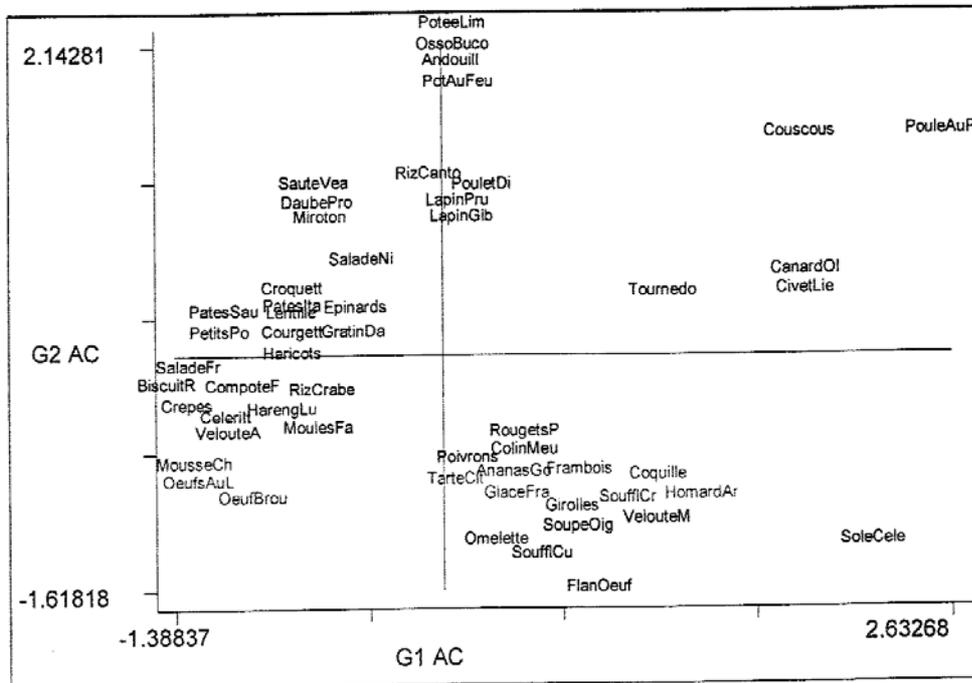


Figure 23 : AC des types et contraintes : observations / types

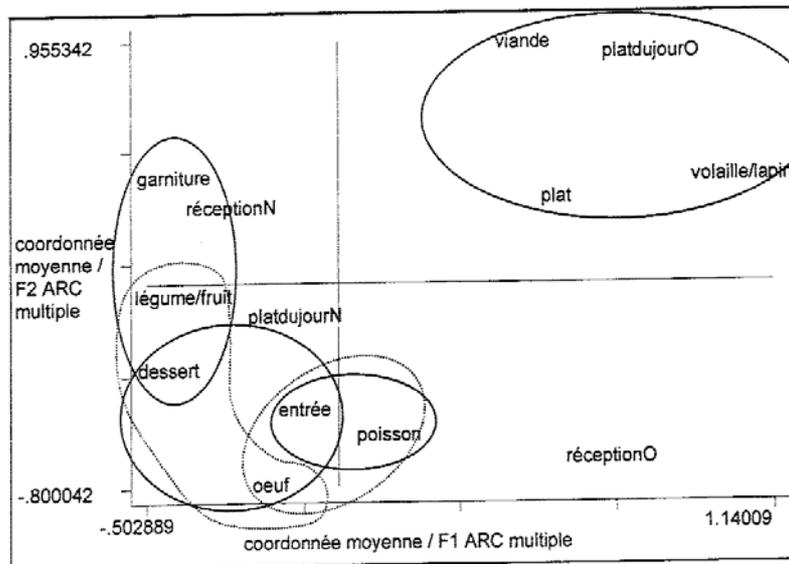


Sur le plan (G1,G2) de l'AC (fig. 23), les différents types de recettes sont moins clairement

délimités que dans l'ACM.

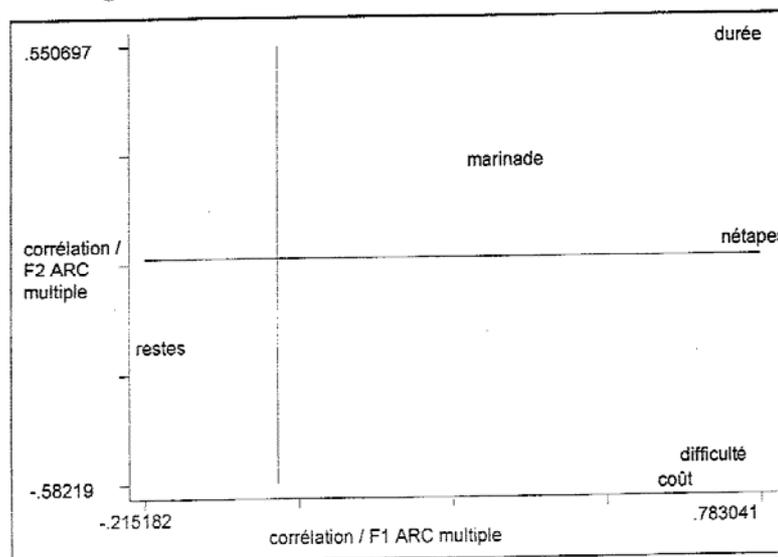
3.3.4. ARCMD

Figure 24 : ARCMD des types et contraintes : types



Les grands types de préparations montrés par l'ACM restent bien visibles sur le plan (1,2) de l'ARCMD (fig. 24): la structure d'ensemble des types est conservée. Mais la variable *réception* occupe dans l'ARCMD une position excentrée. Elle apparaît donc discriminable à partir des variables de contraintes.

Figure 25 : ARCMD des types et contraintes : contraintes



La figure 25 montre des variables de contraintes bien représentées sur le plan (1,2) de l'ARCMD à l'exception de *restes*, mais aussi de *marinade*. Cette dernière variable, assez

types) et l'AC des deux groupes. La mise en rapport des types de cuisine et des contraintes s'est appuyée sur les structures fortes de chaque groupe, de sorte que les photographies de ces rapports sont facilement lisibles. La recherche des liaisons inter-groupes a en outre permis de mettre en relief un type de cuisine jusque là mal visible : la cuisine de luxe.

BIBLIOGRAPHIE

- [BRY01] Bry X. (2001) : Une autre approche de l'Analyse Factorielle: l'Analyse en Résultantes Covariantes, RSA n°3, vol. XLIX, pp. 5-38.
- [BER87] Bernard F. (1987) : Les recettes faciles de Françoise Bernard, Hachette
- [CAZ77] Cazes P., Baumerder A., Bonnefous S., Pagès J.-P. (1977) : Codage et analyse des tableaux logiques, introduction à la pratique des variables qualitatives, Cahiers du BURO n°27
- [CEL94] Celeux G., Nakache J.-P. (1994) : Analyse discriminante sur variables qualitatives, Polytechnica
- [ESC90] Escoffier B., Pagès J. (1990) : Analyses factorielles simples et multiples, Dunod.
- [LEB95] Lebart L., Morineau A., Piron M. (1995) : Statistique Exploratoire Multidimensionnelle, Dunod.
- [TEN98] Tenenhaus M. (1998) : La régression PLS, théorie et pratique, Technip.

ANNEXE

ANNEXE

recette	marinade 0 = non 1 = oui	difficulté 1=élémen- taire 2=facile 3=difficile	coût 1=bon marché 2=raison- nable 3=cher	restes 0 = non 1 = oui	plat du jour 0 = non 1 = oui	réception 0 = non 1 = oui	durée de préparation minutes	nb étapes	emploi 1=entrée 2=plat 3=dessert 4=garniture	aliment 1=oeuf 2=poisson 3=viande 4=volaille/lapin 5=légume / fruit
Ananas en Gondole	0	2	3	0	0	1	60	3	3	5
Andouillette Braisée PdT	0	1	1	0	1	0	60	2	2	3
Biscuit Roulé	0	3	1	0	0	0	30	5	3	5
Canard aux Olives	0	2	3	0	0	1	70	3	2	4
Céleri à l'Italienne	0	2	2	0	0	0	60	3	1	5
Civet de Lièvre	1	3	3	0	0	1	180	6	2	4
Colin Meunière	0	2	3	0	0	0	25	3	2	2
Compote de Fruits	0	1	1	0	0	0	45	2	3	5
Coquilles St Jacques au Gratin	0	2	3	0	0	1	45	4	1	2
Courgette en Beignet	0	3	2	0	0	0	30	2	4	5
Couscous	0	3	3	0	1	1	180	8	2	3
Crêpes	0	1	1	1	0	0	30	3	3	5
Croquettes de Pommes de Terre	0	2	2	0	0	0	60	5	4	5
Daube Provençale	1	2	2	0	0	0	190	3	2	3
Epinards à la Crème	0	1	2	0	0	0	45	4	4	5
Flan Oeuf tomate	0	3	2	0	0	1	50	5	1	1
Framboises Melba	0	2	2	0	0	1	15	2	3	5
Girolles à la Crème	0	2	3	0	0	1	40	4	1	5
Glace à la Fraise	0	2	3	0	0	1	20	3	3	5
Gratin Dauphinois	0	2	2	0	0	0	80	5	4	5
Hareng sauce Lucas	1	2	1	0	0	0	30	4	1	2
Haricots Verts Sautés	0	1	1	0	0	0	60	3	4	5
Homard à l'Armoricaine	0	3	3	0	0	1	75	4	1	2
Lapin aux Pruneaux	1	2	2	0	0	0	120	3	2	4
Lapin en Gibelotte	0	1	2	0	0	0	75	3	2	4
Lentilles au Naturel	0	1	1	0	0	0	70	2	4	5
Miroton	0	1	1	1	0	0	30	2	2	3
Moules Farcies	0	1	1	0	0	0	50	3	1	2
Mousse au Chocolat	0	2	2	0	0	0	15	4	3	5
Oeuf Brouillé aux Foies de Volaille	0	3	2	1	0	0	25	3	1	1
Oeufs Au Lait	0	1	1	1	0	0	30	4	3	1
Omelette Forestière	0	2	2	0	0	0	20	3	2	1
Osso Buco	0	2	1	0	1	0	90	5	2	3
Pâtes à l'Italienne	0	1	1	0	0	0	25	2	4	5
Pâtes Sautées aux Champignons	0	1	1	0	0	0	30	3	4	5
Petits Pois Paysanne	0	1	1	0	0	0	75	2	4	5
Poivrons Farcis au Riz	0	2	2	1	0	0	90	4	2	5
Pot-Au-Feu	0	2	2	0	1	0	210	5	2	3
Potée Limousine	0	1	1	0	1	0	180	5	2	3
Poule Au Pot	0	3	3	0	1	1	240	6	2	4
Poulet à la Diabie	0	2	2	0	0	0	60	4	2	4

Riz au Crabe	0	1	2	0	0	0	40	3	1	2
Riz Cantonnais	0	2	2	0	1	0	40	4	4	5
Rougets Provençale	0	2	3	0	0	0	35	3	2	2
Salade de Fruits Frais	0	1	1	0	0	0	20	2	3	5
Salade Niçoise	0	1	2	1	1	0	15	4	1	5
Sauté de Veau à la Crème d'Estragon	0	1	2	0	0	0	75	3	2	3
Soie Célestine	0	2	3	0	0	1	50	5	2	2
Soufflé au Crabe	0	3	3	1	0	1	60	4	1	2
Soufflé au Curaçao	0	3	2	0	0	1	50	4	3	1
Soupe à l'Oignon	0	2	2	0	0	1	45	4	1	5
Tarte Meringuée au Citron	0	2	2	1	0	1	60	3	3	5
Tournedos Clamart	0	2	3	0	0	1	20	4	2	3
Velouté d'Asperge	0	1	1	0	0	0	15	2	1	5
Velouté de Moules	0	2	2	0	0	1	45	4	1	2

