

LE TRAITEMENT STATISTIQUE DE DONNEES SENSORIELLES

Jérôme Pagès

*Institut National Supérieur de Formation Agro-alimentaire (INSFA)
65, rue de Saint-Brieuc ; F-35042 Rennes Cedex*

Résumé. Les résultats de tests de dégustation peuvent donner lieu à de multiples traitements statistiques ; on présente ici quelques méthodologies utilisant des méthodes classiques appliquées à des données recueillies sous des formes elles aussi classiques. De nombreuses remarques sur la portée des résultats apparaissent au fil du texte.

L'évaluation (ou l'analyse) sensorielle est une discipline qui a connu ces dernières années un essor important dans le domaine agro-alimentaire. Elle a pour objet de décrire la perception de produits par des individus appelés juges, au cours d'épreuves appelées couramment « tests de dégustation ». L'analyse sensorielle est également très répandue dans le domaine des cosmétiques. Depuis peu, elle est présente dans des domaines moins traditionnels comme le secteur automobile.

Les traitements statistiques occupent une large place en évaluation sensorielle. Ceci est dû au fait que, dans tout ensemble de données sensorielles, de nombreuses variabilités interviennent : variabilités inter et intra juges ; variabilités inter et intra produits. Démêler ces variabilités implique le recueil de nombreuses données, selon un protocole de dégustation précis empruntant ses démarches au domaine des plans d'expériences, et un dépouillement statistique adapté.

Les problèmes statistiques rencontrés en évaluation sensorielle sont très variés. Nous nous restreignons ici :

- aux protocoles les plus simples ; c'est le cas lorsque J juges classent de façon hédonique I produits (le terme *hédonique* se réfère au plaisir : le juge indique les produits qui lui plaisent, ou ne lui plaisent pas, sans avoir à justifier ou commenter ce classement) ; c'est aussi le cas lorsque J juges évaluent I produits au moyen de K variables, appelées pour la circonstance *descripteurs* (on parle d'épreuve descriptive ; le questionnaire que chaque juge remplit pour un produit est appelé « fiche de dégustation ») ;
- à des méthodes de traitement classiques (analyse de variance à deux facteurs ; analyse en composantes principales).

En particulier, nous n'abordons pas :

- les épreuves discriminatives ; c'est le cas où l'on cherche à déceler une différence entre produits, différence si faible que l'on ne cherche pas à la qualifier ; le paradigme de l'épreuve discriminative est le fameux test triangulaire au cours duquel on propose à chaque individu trois produits dont deux sont identiques : l'individu doit alors déterminer lequel est différent des deux autres (Pagès 1994) ;

- la construction de protocoles expérimentaux : beaucoup de travaux ont été consacrés aux blocs incomplets (chaque juge ne déguste pas tous les produits), aux carrés latins (pour faire déguster chaque produit le même nombre de fois en premier, en second, etc.) et aux dispositifs équilibrés pour les voisinages (pour faire déguster chaque produit le même nombre de fois après chacun des autres) ; (MacFie et al 1989 ; Wakeling et al 1995)
- les méthodes d'analyse de tableaux multiples, Analyse Factorielle Multiple –AFM– (Escofier, Pagès, 1993), Analyse Procustéenne Généralisée –APG– (Dijksterhuis, 1991) et STATIS (Lavit, 1988).

Notre objectif est de mettre en évidence les principales problématiques rencontrées dans l'analyse de données sensorielles et les principaux traitements simples correspondant.

1. Analyse de classements hédoniques

1.1 Données

On soumet I produits au jugement hédonique de J juges.

Ces jugements sont généralement recueillis sous forme de classements, cas considéré dans ce texte ; mais la plupart des idées développées ici se transposent au cas de notes hédoniques, et ce d'autant plus facilement que, dans de nombreux calculs, les rangs sont en fait manipulés comme des notes.

Les données peuvent être regroupées dans un tableau dans lequel, à l'intersection de la ligne i et de la colonne j , se trouve x_{ij} , rang affecté au produit i par le juge j . Un exemple est fourni par le tableau 1.

	$J1$	$J2$	$J3$	$J4$	$J5$	$J6$	$J7$	$J8$	M
A	1	1	2	2	1	1	2	2	1.5
B	2	2	1	1	2	2	1	1	1.5
C	3	4	3	4	5	6	5	6	4.5
D	4	3	4	3	6	5	6	5	4.5
E	5	6	6	5	3	4	3	4	4.5
F	6	5	5	6	4	3	4	3	4.5

Tableau 1. Jeu de données 1. Classements de 6 produits ($A \dots F$) par 8 juges ($J1 \dots J8$).
Exemple : le juge $J3$ a classé le produit E au rang 6 ; M : rang moyen.

1.2 Problématique ; spécificité des jugements hédoniques

Dans le cas de descripteurs analytiques, l'attention est centrée sur les produits. Les variabilités liées aux juges sont considérées comme des bruits que l'on cherche à réduire ; tel est l'objet, par exemple, de l'entraînement du jury.

Dans le cas de jugements hédoniques, on s'intéresse aux couples juge-produit. Les variabilités liées aux juges sont parties intégrantes de l'objet de l'étude, ce qui va de pair avec l'utilisation d'un jury "naïf" (i.e. non entraîné) pour ce type de données.

Cette distinction est couramment admise actuellement ; mais, peut-être, la raison en est surtout sa simplicité, conceptuelle et de mise en oeuvre ; on peut penser qu'elle s'atténuera au fil du temps et que l'on cherchera de plus en plus à intégrer les variabilités liées aux juges dans les descriptions analytiques de produits (en disant, par exemple, que tel produit est l'objet de jugements homogènes et non tel autre). Ainsi, la spécificité de cette partie de l'exposé ne concerne pas tant la nature des données que le point de vue adopté pour les aborder (importance accordée aux variabilités liées aux juges). Le titre de cette partie, centré sur les classements hédoniques, est donc légèrement impropre mais a été choisi en référence aux pratiques courantes.

Cela étant, la prise en compte des variabilités liées aux juges contraint le recueil des données et alourdit (pour l'enrichir) l'analyse des résultats ; cette prise en compte se traduit généralement par la définition d'une typologie –ou segmentation– des juges (telle qu'à l'intérieur d'une classe –ou segment– les jugements sont homogènes), ce qui requiert beaucoup d'énergie et se justifie surtout pour les jugements hédoniques, en principe en prise directe avec les aspects marchands.

Dans l'étude d'un ensemble de classements, étudier la variabilité des juges, revient à étudier, parmi tous les classements possibles, la répartition des classements observés. Deux situations extrêmes servent utilement de références :

- la répartition uniforme : tous les classements possibles sont observés le même nombre de fois ;
- le consensus parfait : un seul classement est observé.

Entre ces deux extrêmes, on observe des répartitions plus ou moins régulières. Une prise en compte stricte de la variabilité des juges implique de considérer autant de segments de juges qu'il y a de classements observés. Ceci conduit en pratique à un nombre de segments trop grand et parallèlement à des effectifs par segment trop faibles pour qu'une telle segmentation soit utilisable. Une donnée importante du problème doit alors être prise en compte, à savoir que tous les classements possibles ne sont pas "équidistants" les uns des autres : ainsi, il est clair que les classements *ABCDEF* et *ABDCEF* sont plus "proches" entre eux qu'ils ne le sont de *FEDCBA*. D'où l'idée de rassembler des classements observés "proches" de façon à obtenir une segmentation des juges opérationnelle (avec un petit nombre de segments).

En résumé, on examine d'abord si l'ensemble des données peut être considéré comme l'expression d'un consensus. Si ce n'est pas le cas, on cherche une partition des individus telle que, à l'intérieur de chaque segment, les classements peuvent être considérés comme l'expression d'un consensus.

1.3 Présentation résumée des résultats pour l'ensemble des juges

1.3.1 Tableau produits \times rangs

Ce tableau croise les produits et les rangs. A l'intersection de la ligne *i* et de la colonne *l*, se trouve le nombre de fois que le produit *i* a été classé au rang *l* ; le tableau **2a** en est l'application au jeu de données 1.

La lisibilité de ce tableau est améliorée en rangeant les produits par rang moyen croissant. L'existence d'un consensus entre les juges se traduit par une structure en scalogramme (terme popularisé par Benzécri 1973) : les valeurs les plus grandes se trouvent alors sur ou proches de la diagonale (qui va du haut à gauche au bas à droite). Une structure bloc-diagonale suggère de considérer une partition de l'ensemble des produits plutôt que l'ensemble lui-même (tel est le cas du tableau **2a** dont les produits *A* et *B* sont toujours classés en premier). L'AFC de ce tableau est précieuse lorsque le nombre de produits est grand.

1.3.2 Rangs moyens

On peut calculer le rang moyen de chaque produit en considérant les classements comme des notes (cf. colonne *M* du tableau **1**). L'existence d'un consensus induit des différences entre les rangs moyens.

L'expression du degré de consensus par la variabilité des rangs moyens conduit à examiner le tableau **1** à l'aide de la grille de l'analyse de la variance. Du fait de la nature des classements, les moyennes des rangs des juges sont égales. La variabilité totale du tableau se décompose donc en une variabilité due aux produits (variabilité des rangs moyens) et une variabilité résiduelle (qui est en fait une interaction juge×produit).

Cette analyse de variance particulière est réalisée par le test de Friedman (le critère statistique de Friedman est proportionnel à la variance des rangs moyens). La probabilité associée à ce test peut être vue comme un indicateur de consensus au sens suivant : on considère le tirage au hasard dans un ensemble de classements dont la répartition est uniforme (chaque objet à classer occupe chaque rang possible avec la même probabilité) et donc telle que les rangs moyens sont égaux ; la probabilité fournie par le test de Friedman est celle d'obtenir, dans ce cadre, une variabilité entre les rangs moyens au moins aussi grande que celle effectivement observée. Deux cas doivent être distingués dans l'utilisation de cette probabilité :

- si les juges peuvent être considérés comme ayant été tirés au hasard dans une population, l'hypothèse H_0 associée au test spécifie que, *dans la population totale*, les rangs moyens sont égaux ; ainsi, dans ce cas, cette probabilité aide à décider de considérer ou non qu'il existe, *au niveau de la population*, des différences entre les rangs moyens ;
- si les juges ne peuvent pas être considérés comme étant tirés au hasard, l'hypothèse H_0 associée au test spécifie que, *en demandant aux juges de répéter à de nombreuses reprises leur classement*, on obtiendrait en final des rangs moyens égaux ; ainsi, dans ce cas, cette probabilité aide à décider, *au niveau du jury*, de considérer ou non comme fortuites les différences observées entre les rangs moyens.

La portée de cette démarche est limitée car :

- l'absence de différences entre les rangs moyens n'implique aucunement l'uniformité de la répartition des classements ;
- la présence de différences entre les rangs moyens n'implique aucunement un consensus tel qu'il ne soit pas intéressant de réaliser une segmentation des juges ; ainsi, dans le jeu de données 2, la probabilité associée au test de Friedman est .00005 : les écarts entre les rangs moyens ne peuvent être imputés au hasard, mais il est quand même intéressant de réaliser une segmentation des juges.

1.4 Segmentation (ou typologie) des juges

On peut réaliser une classification ascendante hiérarchique (CAH) des juges : concrètement, le tableau analysé est alors le transposé du tableau 1 (les juges constituent les lignes et sont traités comme des individus statistiques ; on peut aussi, moins classiquement –Couronne 1996–, considérer les juges comme des colonnes, ce qui permet d'introduire de façon élégante des contraintes sur l'homogénéité des groupes formés). Il convient de ne pas normer ce tableau. Prendre la distance euclidienne usuelle comme mesure de ressemblance entre classements est assez "naturel". L'inertie du nuage des juges, égale à l'inertie résiduelle de l'analyse de variance mentionnée ci-dessus, mesure l'écart au consensus. L'objectif usuel de construction d'une partition ayant un fort rapport *inertie inter/inertie intra* est adéquat : il vise à obtenir des jugements consensuels à l'intérieur des segments.

Le choix du nombre de segments dépend :

- des données elles-mêmes ;
- de l'homogénéité (inertie intra-segment) des segments, c'est-à-dire du degré de consensus intra-segment qu'il est raisonnable de supporter pour considérer un segment ;
- dans certaines études, du nombre de segments que l'on peut gérer (cette exigence peut être contradictoire avec la précédente et un compromis est alors à trouver).

L'ACP de ce tableau, qui met en évidence les principaux facteurs de la variabilité résiduelle, est une aide précieuse à l'interprétation de cette typologie.

Exemple : jeu de données 1

Les données du tableau 1 conduisent à distinguer les deux segments (*J1, J2, J3, J4*) et (*J5, J6, J7, J8*). L'homogénéité de ces segments apparaît dans les tableaux 2b et 2c.

	Rang					
	1	2	3	4	5	6
A	4	4				
B	4	4				
C			2	2	2	2
D			2	2	2	2
E			2	2	2	2
F			2	2	2	2

2a : ensemble

	Rang					
	1	2	3	4	5	6
A	2	2				
B	2	2				
C			2	2		
D			2	2		
E					2	2
F					2	2

2b : segment 1

	Rang					
	1	2	3	4	5	6
A	2	2				
B	2	2				
E			2	2		
F			2	2		
C					2	2
D					2	2

2c : segment 2

Tableau 2. *Jeu de données 1. Tableaux (produits×rangs) pour l'ensemble des juges et pour deux segments ; Exemple : parmi l'ensemble des juges, 4 ont classé le produit B au rang 1 ; l'ordre des produits diffère entre les deux segments.*

2. Nature et structure des données descriptives

2.1 Nature des variables

Généralement on demande aux juges de s'exprimer à l'aide d'échelles ; par exemple, on évalue l'acidité de *nulle ou très faible* (modalité codée 1) à *très forte* (modalité codée 5). Formellement les variables obtenues sont qualitatives à modalités ordonnées. En pratique elles sont traitées :

- en tant que variables quantitatives (cas de calculs de moyennes) ; chacune des n modalités est codée par un nombre, le plus souvent un entier de 1 à n ; les réponses sont alors considérées comme des notes ;
- en tant que variables qualitatives (cas de comptages comme, par exemple, le nombre d'individus qui ont attribué telle modalité à tel produit) ; l'ordre entre les modalités n'apparaît pas dans les calculs mais seulement dans la présentation des résultats.

Aucun de ces deux points de vue ne respecte strictement la nature des données recueillies ; le premier est plus synthétique mais le second plus riche (comparer, pour un descripteur, par exemple sa moyenne et la fréquence de ses modalités).

2.2 Structure a priori des données.

On adopte les notations suivantes : J juges évaluent I produits à l'aide de K descripteurs ; x_{ijk} est l'évaluation par le juge j du produit i pour le descripteur k .

Ces données constituent un "cube" qui peut être envisagé de plusieurs façons, illustrées par la figure 1.

Ainsi, le cube "brut" des données (noté $I \times J \times K$) peut être envisagé comme :

- une juxtaposition spatiale de K tableaux juges \times produits (notés chacun $(I \times J)_k$) ; ces tableaux sont analysés indépendamment les uns des autres dans les approches unidimensionnelles (essentiellement par analyse de variance) ;
- une juxtaposition spatiale de J tableaux produits \times descripteurs ; ce point de vue est celui de l'analyse multidimensionnelle des produits ; on peut étudier la moyenne de ces J tableaux (notée $I \times K$ et dont le terme général m_{ik} est la moyenne des notes attribuées au produit i pour le descripteur k) ou, de façon plus fine, l'une de leurs juxtapositions planes : $I \times (J \times K)$ ou $(I \times J) \times K$;
- une juxtaposition spatiale de I tableaux juges \times descripteurs ; ce point de vue est celui de l'analyse multidimensionnelle des juges ; on peut étudier la moyenne de ces I tableaux (notée $J \times K$ et dont le terme général m_{jk} est la moyenne des notes attribuées par le juge j pour le descripteur k) ou, plus finement, l'une de leurs juxtapositions planes : $J \times (I \times K)$ ou $(I \times J) \times K$;

Chacun de ces tableaux correspond à un point de vue sur les données. On peut d'emblée remarquer le rôle central joué par le tableau $(I \times J) \times K$, dont les lignes sont les fiches de dégustation (une fiche est remplie par un juge pour un produit), qui permet d'aborder, de façon multidimensionnelle, aussi bien l'étude des juges que celle de produits.

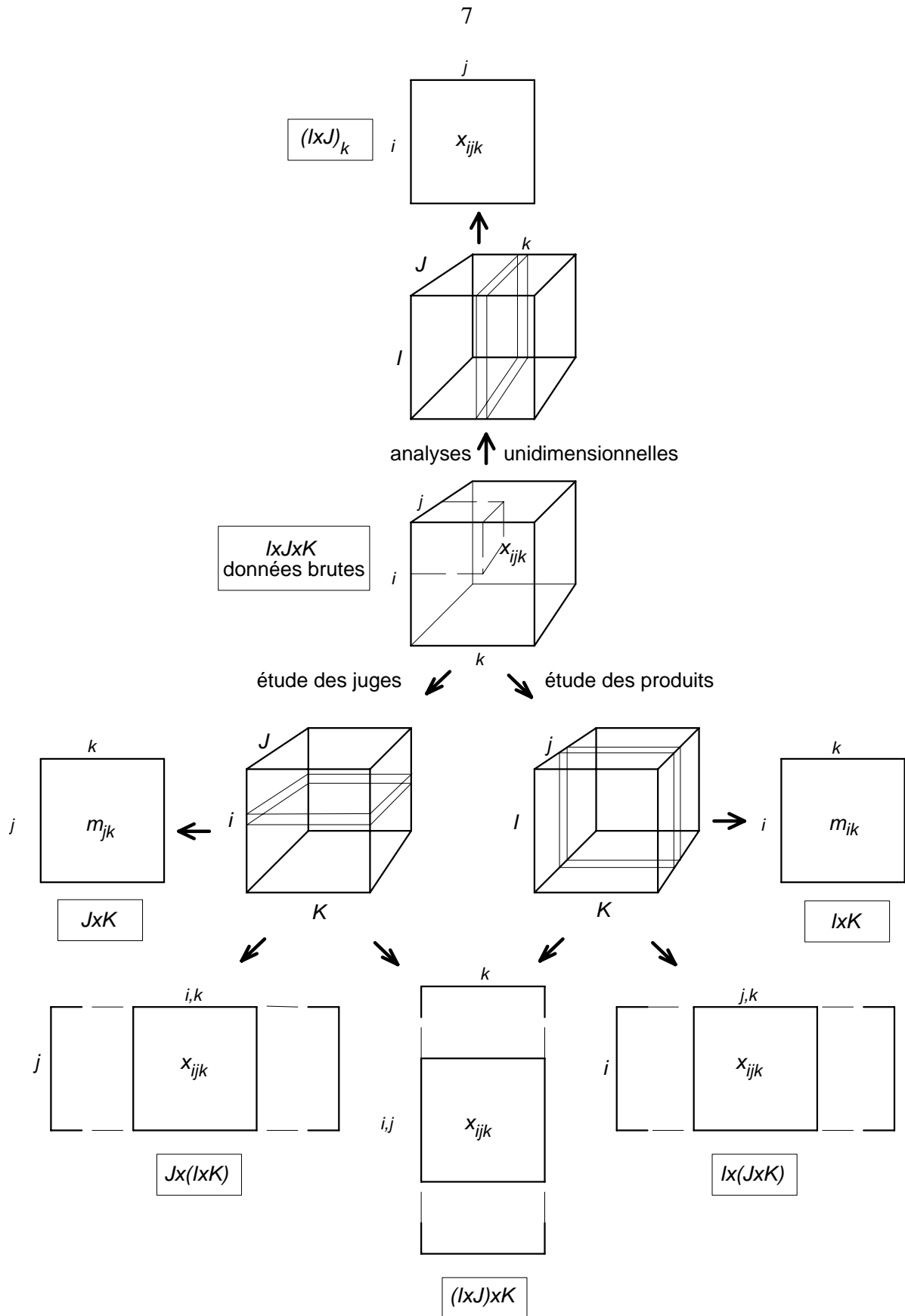


Figure 1. Données issues de l'évaluation de I produits par J juges selon K descripteurs. Au centre : cube brut des données ; autour : principaux tableaux de données extractibles du cube.

Exemple : jeu de données 2

Un jeu de données choisies sert d'illustration (cf. **Tab 3.**). 5 juges ($J1, \dots, J5$) ont évalué 4 produits ($P1, \dots, P4$) à l'aide de 3 descripteurs (*acidité, amertume, caractère salé*).

	<i>acidité</i>					<i>amertume</i>					<i>caractère salé</i>				
	<i>J1</i>	<i>J2</i>	<i>J3</i>	<i>J4</i>	<i>J5</i>	<i>J1</i>	<i>J2</i>	<i>J3</i>	<i>J4</i>	<i>J5</i>	<i>J1</i>	<i>J2</i>	<i>J3</i>	<i>J4</i>	<i>J5</i>
<i>P1</i>	1	0	2	0	2	1	0	2	0	2	4	5	4	5	2
<i>P2</i>	2	1	3	0	4	1	0	2	1	1	1	2	0	2	0
<i>P3</i>	3	1	5	3	3	4	3	5	4	4	3	4	3	4	1
<i>P4</i>	4	3	5	3	5	4	3	5	3	5	2	3	1	3	1

Tableau 3. Jeu de données 2 ; représentation de $I \times (J \times K)$. Exemple : le juge *J3* a attribué au produit *P2* la note 3 pour l'acidité ; remarque : les évaluations du juge *J1* coïncident avec les évaluations moyennes.

3. Approches unidimensionnelles des données descriptives

3.1 Données

Lorsque l'on considère le seul descripteur k , les données analysées présentent la même structure que les jugements hédoniques du tableau 1 (à ceci près que la somme des notes d'un même juge n'est pas constante). Dans ce tableau, noté $(I \times J)_k$ figure 1, à l'intersection de la ligne i et de la colonne j , se trouve x_{ijk} , valeur attribuée par le juge j au produit i (pour le descripteur k). A partir de ce tableau, on calcule la moyenne de chaque produit i (m_{ik}), et de chaque juge j (m_{jk}).

3.2 Comparaison globale des produits

Usuellement ces données sont analysées au travers de la grille de l'analyse de la variance. La variabilité de la moyenne des produits (resp. des juges) est une mesure brute de l'effet produit (resp. juge). L'équation de l'analyse de variance montre que, en mesurant les variabilités par des sommes de carrés d'écart (ou, dans un autre langage, par des inerties), les effets *produit* et *juge* sont des composantes additives de la variabilité totale. En effet :

$$\text{Variabilité totale} = \text{effet produit} + \text{effet juge} + \text{interaction produit} \times \text{juge}$$

On remarque qu'il n'apparaît pas de variabilité résiduelle au sens strict (l'approche de cette variabilité résiduelle nécessite des répétitions, c'est-à-dire que chaque juge évalue au moins deux fois les produits). L'interaction produit \times juge revêt une interprétation importante : elle représente le non-consensus entre les juges concernant l'appréciation des produits. Parallèlement, l'effet produit représente le consensus entre les juges.

Dans la pratique courante, on s'intéresse surtout à l'effet produit : on en teste globalement la signification par un test F fondé sur le rapport : [Carré Moyen produits] / [Carré Moyen interaction]. Intuitivement, compte tenu de l'interprétation de l'effet produit et de l'interaction en terme de consensus, ce rapport, et surtout la probabilité critique associée, mesure en

quelque sorte le degré de consensus entre les juges sur les produits. Plus précisément, une probabilité critique faible indique, par exemple, que le consensus pour trouver tel produit plus acide est bien plus fort que les différences d'appréciation inter-individuelles.

De façon plus formelle, deux cas doivent être distingués selon que l'on considère les juges pour eux-mêmes (dans la terminologie de l'analyse de variance on dit que l'effet juge est fixe) ou comme des représentants d'une population plus vaste (l'effet est dit « aléatoire »).

Le premier cas correspond à des juges très entraînés qui ne représentent qu'eux-mêmes ; faire le test précédent revient implicitement à supposer que l'interaction juge×produit est nulle (ce qui est conforme à un entraînement poussé) et que cette interaction est en fait une résiduelle ; les deux effets de l'analyse de variance sont fixes et doivent être testés par rapport à la résiduelle. La conclusion du test ne va pas au-delà du jury. Plus précisément, l'hypothèse H_0 associée au test spécifie que, *en demandant aux juges de répéter leurs évaluations à de nombreuses reprises*, on obtiendrait en final des notes moyennes égales ; ainsi, dans ce cas, cette probabilité aide à décider, *au niveau du jury*, de considérer ou non comme fortuites les différences de moyennes observées.

Dans le second cas, les juges sont considérés comme tirés au hasard d'une population plus vaste. L'effet juge est alors « aléatoire », l'effet produit lui étant fixe (le modèle de l'analyse de variance est dit alors mixte) ; dans ce cas, l'effet produit doit être testé par rapport à l'interaction (qu'il n'y a pas lieu dans ce cas de supposer nulle, le tirage au hasard de juges représentant ainsi une population vaste allant de pair avec un entraînement moins poussé) ; la conclusion du test va dans ce cas au-delà du jury et vaut pour la population qu'il représente. Plus précisément, l'hypothèse H_0 associée au test spécifie que, *dans la population totale*, il n'y a pas de différence entre les moyennes ; dans ce cas, la probabilité critique aide à décider ou non de généraliser à *la population* l'existence de différences entre les moyennes observées.

3.3 Comparaison des produits deux à deux

L'appréciation globale d'un effet produit n'est qu'une étape de l'analyse. Le résultat final ne peut être qu'une comparaison fine des produits entre eux. En pratique le résultat d'une étude descriptive consiste en la comparaison, pour chaque descripteur présentant un effet produit significatif, des produits deux à deux. Différents procédés graphiques permettent de synthétiser les résultats. Différents indicateurs peuvent être calculés.

Dans le test t comparant deux moyennes, on peut remplacer l'estimation de la variance résiduelle faite usuellement à partir des deux produits concernés par une estimation s'appuyant sur l'ensemble des produits (concrètement le carré moyen résiduel de l'analyse de variance). Cette idée est à la base du calcul de la Plus Petite Différence Significative (PPDS).

Cette amélioration n'empêche pas que les tests soient faits en grand nombre. Même si les juges donnent toutes leurs réponses au hasard, la probabilité est généralement très grande de voir apparaître des résultats significatifs. En d'autres termes, le risque *global* (probabilité de déceler au moins une différence alors qu'il n'y en a aucune) est alors beaucoup plus élevé que le risque *local*, relatif à chaque comparaison et fixé usuellement à 5%. Ce problème fait toujours l'objet de discussions ; une façon de diminuer le risque global consiste tout

simplement à diminuer les risques locaux (c'est ce que fait la procédure de Bonferroni dans laquelle le risque local usuel est divisé par le nombre $I(I-1)/2$ de comparaisons effectuées).

Quelle que soit la procédure utilisée, on ne pourra jamais se prémunir de résultats du type : $A=B$; $B=C$; $A \neq C$. Cette situation n'est pas simple mais correspond à bon nombre de cas réels : on est souvent en présence d'une gamme de produits dont les extrêmes sont bien différenciés mais entre lesquels se placent plusieurs produits intermédiaires.

Remarque : comparaison des produits deux à deux ou chacun par rapport à l'ensemble

Les comparaisons des produits deux à deux correspondent à une problématique simple et précise, mais deviennent inextricables dès que le nombre de produits est un tant soit peu grand. Il est plus commode de comparer chaque produit à l'ensemble des autres. Ce principe est celui de la procédure DEMOD du logiciel SPADN.

3.4 Puissance

Quel type de différence entre produits est-on susceptible de mettre en évidence dans une épreuve donnée ? Il n'y a pas de réponse générale à cette question. En revanche, on peut donner quelques points de repère. Le tableau suivant récapitule la puissance du test t de Student dans les cas suivants :

- taille du jury (n) : 25 16 et 9 juges ;
- écart-type résiduel (s) : 1, .7 et .5 ; ces valeurs sont couramment observées ; pour fixer les idées, des exemples de distributions correspondantes sont illustrés figure 2.
- différence à mettre en évidence (d) : 1 et 1/2 points ; une différence de 1/2 point peut s'interpréter comme : la moitié des juges évaluent la différence à 1 point, les autres ne faisant aucune différence.

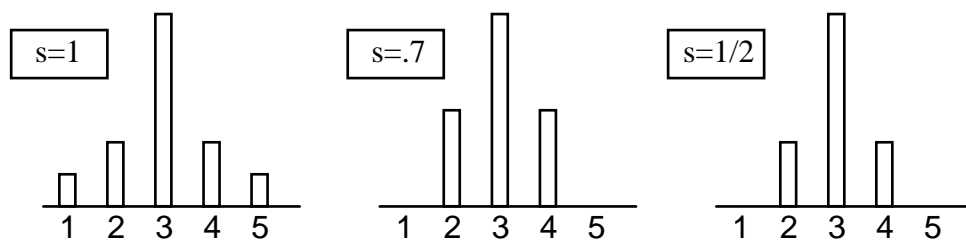


Figure 2. Exemples de distributions résiduelles avec $s=1$; $s=.7$; $s=1/2$. Exemple du cas $s=.7$: confronté au même produit, le juge attribue une fois sur deux la note 3 et une fois sur deux un point au-dessus ou au dessous.

n	s	s/\sqrt{n}	seuil	$d=1$	$d=.5$
25	1	.20	.41	.99	.67
	.7	.14	.29	.99	.93
	.5	.10	.21	.99	.99
16	1	.25	.53	.96	.45
	.7	.17	.37	.99	.76
	.5	.12	.26	.99	.96
9	1	.33	.71	.80	.27
	.7	.23	.49	.97	.52
	.5	.17	.35	.99	.80

Tableau 4. Puissance du test de t de Student (au seuil 5%) pour quelques valeurs de n (nombre d'observations), de s (écart-type résiduel), et de d (différence à mettre en évidence). Exemple : pour 16 juges ayant un écart-type résiduel de 1, la probabilité de mettre en évidence une différence de 1/2 point est .45

4. Approches multidimensionnelles de données descriptives

Nous donnons ici quelques principes généraux, illustrés par des traitements du jeu de données 2 (cf. **Tab. 3**). L'objectif essentiel est ici de montrer les problématiques différentes que l'on peut aborder par les méthodes factorielles, selon la façon dont on construit le tableau soumis à l'analyse. Par commodité, nous considérons les variables comme quantitatives et utilisons l'Analyse en Composantes Principales (ACP). La plupart des idées se transposent au cas "qualitatif" en ayant recours à l'Analyse des Correspondances Multiples (ACM).

4.1 Intérêt

Lors de l'analyse de données sensorielles, on met presque toujours en évidence des liaisons entre variables. Ces liaisons peuvent apparaître à différents niveaux.

- Niveau *inter-produits* ; il s'agit de liaisons entre effets produits. Ainsi, du fait du choix des produits étudiés simultanément, certains descripteurs peuvent apparaître liés. C'est le cas, par exemple, lorsque ce sont les mêmes produits qui sont évalués à la fois les plus acides et les plus amers. De telles liaisons peuvent correspondre à un facteur technique sous-jacent qui influence simultanément plusieurs caractéristiques sensorielles.
- Niveau *inter-juges* ; il s'agit de liaisons entre effets juges. Ainsi, pour certains critères, il se peut que quelques juges attribuent des notes systématiquement élevées et d'autres des notes systématiquement faibles.
- Niveau *intra-juge* ; il s'agit de liaisons entre variabilités résiduelles. En ce sens, deux descripteurs sont liés lorsque, pour chaque juge, le fait de donner une forte note pour l'un prédispose à donner une forte valeur pour l'autre.

Dans l'analyse de données sensorielles, il convient de détecter ces liaisons lorsqu'elles existent et ce dans une double perspective :

- les liaisons entre descripteurs présentent de l'intérêt en elles mêmes ; elles permettent une meilleure compréhension de la façon dont s'organisent les perceptions, par exemple en conjecturant l'existence de facteurs sous-jacents aux données ;
- l'existence de liaisons entre descripteurs incite à ne pas se contenter de juxtaposer des descriptions univariées des produits et à construire des variables synthétiques. L'intérêt de telles variables synthétiques pour décrire les produits dépend de leur interprétabilité.

4.2 Tableau $I \times K$ (produits \times descripteurs)

4.2.1 Tableau et analyse (cf. Figure 3 b)

A l'intersection de la ligne i et de la colonne k du tableau $I \times K$, on trouve m_{ik} , moyenne, sur tous les juges, des notes attribuées au produit i pour le descripteur k . Dans ce tableau, la variabilité provient des seuls effets produits. Les liaisons entre descripteurs s'interprètent ici en terme de liaisons entre effets produits.

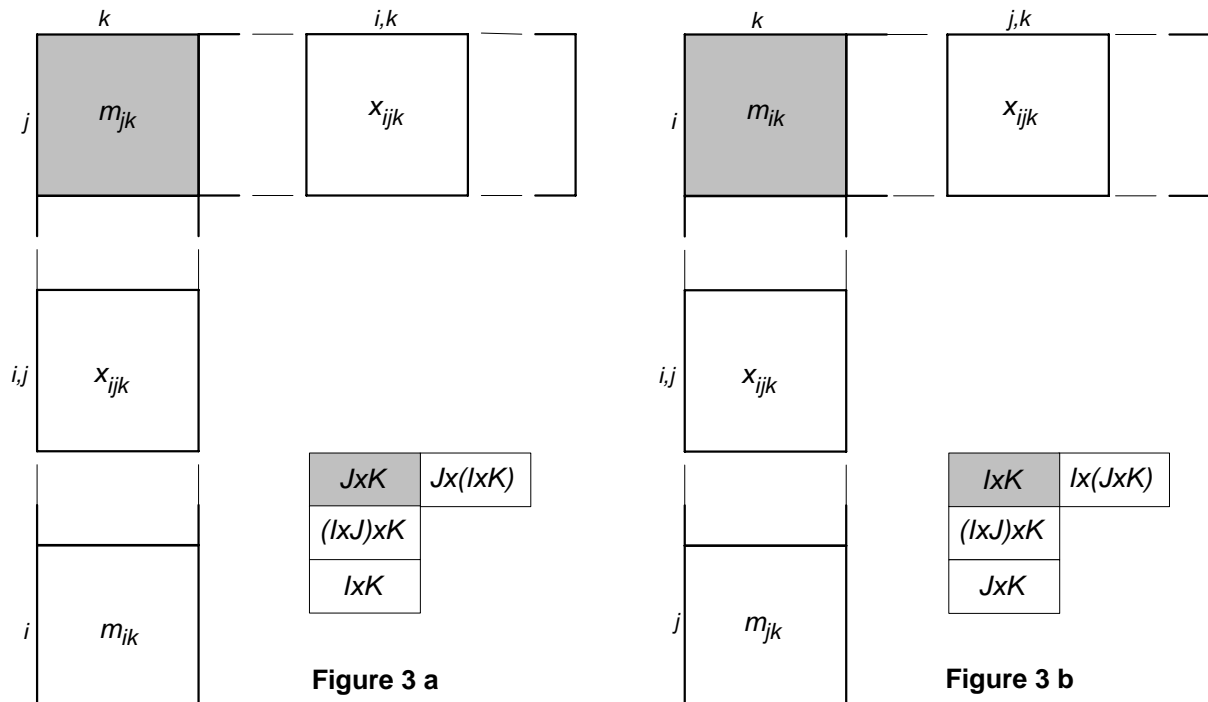


Figure 3. Tableaux soumis à l'ACP ; éléments actifs en grisé ; **3 a** correspond à l'ACP du § 4.2. et à la figure 5 ; **3 b** correspond à l'ACP du § 4.3. et à la figure 4.

L'analyse usuelle de ce tableau est une ACP. Réaliser une ACP normée revient à accorder le même poids à chaque descripteur. Réaliser une ACP non normée revient à accorder à chaque descripteur un poids proportionnel à son écart-type. Le choix entre ces deux ACP s'appuie sur le fait que l'écart-type d'un descripteur est une mesure brute de l'effet produit pour ce descripteur (certains préfèrent normer le tableau, arguant que la variabilité d'un descripteur pour un juge se réfère surtout à son aisance à s'exprimer à propos de ce descripteur, aisance sans rapport avec une importance relative des descripteurs).

Jeu de données 2 : on a choisi une ACP non normée. La relativement faible variabilité des écarts-types permet de penser que ce choix n'influence pas beaucoup les résultats. Le tableau analysé est schématisé figure **3b** ; les divers éléments supplémentaires apparaissent progressivement ci-après. Compte tenu du pourcentage d'inertie associé au premier plan (99%), les graphiques de la figure **4** suffisent pour l'interprétation.

4.2.2 Problématique associée aux éléments actifs (cf. Figure 4, éléments encadrés)

La problématique de l'ACP d'un tel tableau s'articule autour de deux thèmes principaux.

Bilan des liaisons entre descripteurs

Jeu de données 2 : *acide* et *amer* sont corrélés ; il existe des produits à la fois acides et amers et des produits à la fois peu acides et peu amers ; *amer* et *salé* sont non corrélés.

Bilan des ressemblances entre produits

Jeu de données 2 : le premier facteur de variabilité (F1) sépare d'une part *P1* et *P2* (produits peu acides et peu amers) et d'autre part *P3* et *P4* (produits acides et amers) ; le second (F2) sépare d'une part les produits *P1* et *P3* (salés) et d'autre part les produits *P2* et *P4* (peu salés).

4.2.3 Problématique associée aux éléments supplémentaires (cf. Figure 4)

Les interprétations précédentes concernent les évaluations moyennes, c'est-à-dire le jury considéré dans son ensemble. Il est légitime de se demander, dans un deuxième temps, si ces interprétations valent pour chacun des juges pris séparément. Pour cela, on introduit dans l'ACP différents éléments supplémentaires.

Point de vue des descripteurs

Les liaisons entre descripteurs établies à partir des évaluations moyennes sont-elles observables pour chaque juge ? Pour répondre à ce type de question, on introduit les évaluations individuelles en tant que colonnes supplémentaires ; cela revient (cf. **Fig. 3 b**) à juxtaposer au tableau $I \times K$ le tableau $I \times (J \times K)$; le tableau obtenu possède I lignes et $(K+JK)$ colonnes.

Jeu de données 2 : dans l'ensemble, les corrélations observées sur les moyennes du jury valent pour chaque juge. Ceci est particulièrement remarquable pour le juge *J1* dont les évaluations coïncident avec celles de l'ensemble du jury. Plus en détail, on peut remarquer la liaison étroite entre toutes les variables relatives au descripteur *salé* (même si leurs écarts-types diffèrent) ; en revanche, la liaison entre *acide* et *amer* est intense pour le juge *J3* et lâche pour le juge *J5*.

Point de vue des produits

Les variables synthétiques, qui donnent la meilleure représentation des produits moyens, donnent-elles la même représentation pour chacun des juges ? Pour répondre à ce type de question, on introduit les évaluations individuelles en tant que lignes supplémentaires. Cela revient (cf. **Fig. 3 b**) à juxtaposer au tableau $I \times K$ le tableau $(I \times J) \times K$. Par construction, la représentation précédente (P_i) du produit i apparaît au centre de gravité des représentations ($J_j P_i$) de ce produit i par chacun des juges.

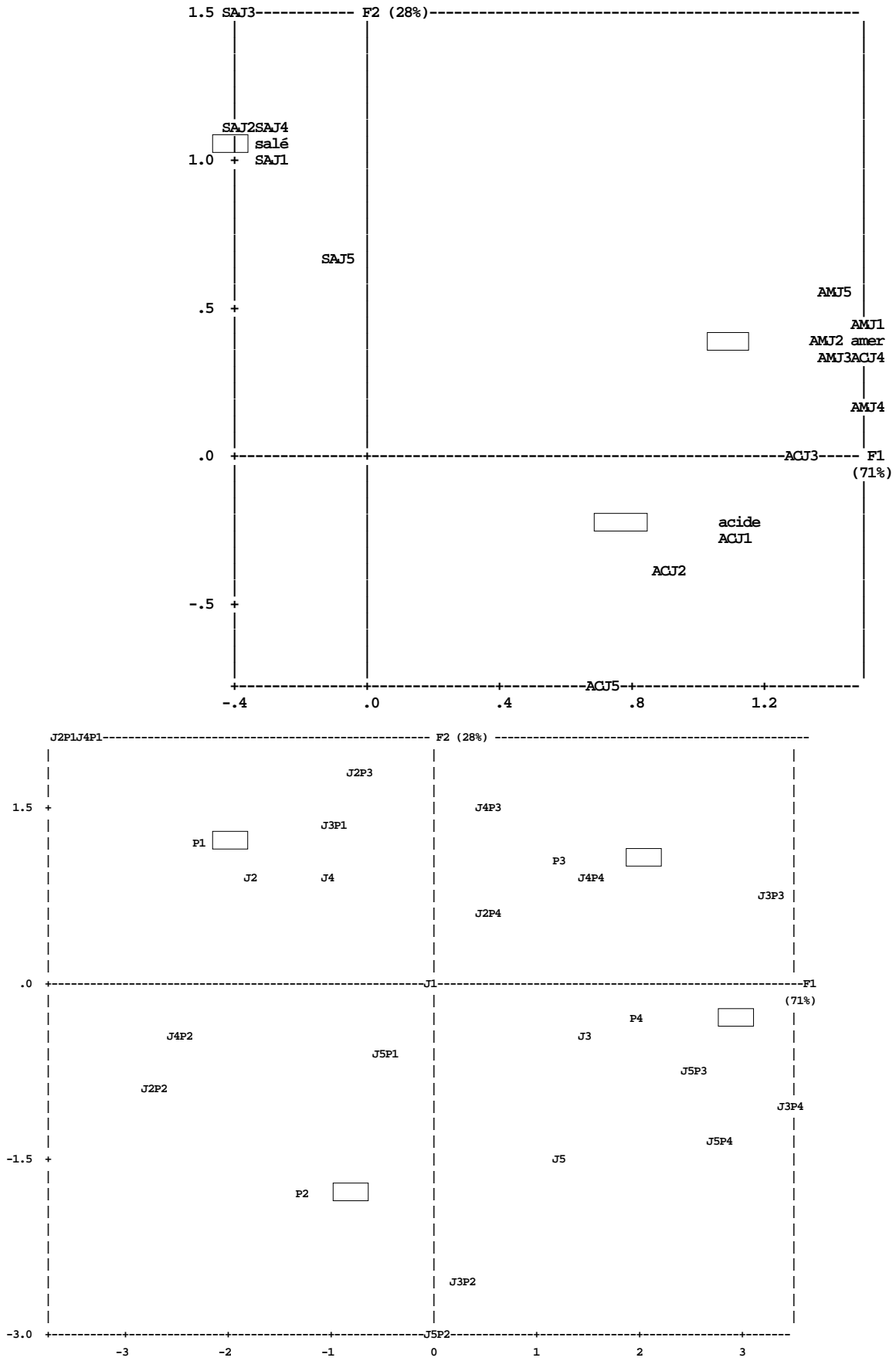


Figure 4. ACP non normée de $I \times K$ avec $(I \times J) \times K$ et $I \times (J \times K)$ en supplémentaires (cf. fig. 3b)

AMJ2 : variable amertume utilisée par le juge J2 ; J5P3 : produit P3 évalué par le juge J5.
Les éléments actifs sont encadrés.

Jeu de données 2 : les produits P3 et P4, évalués globalement acides et amers, ont été évalués très acides et très amers par le juge J3 et seulement moyennement acides et moyennement amers par le juge J2. La différence du caractère *salé* entre P1 et P2 est beaucoup plus importante pour le juge J3 que pour le juge J5.

Dans l'analyse de ce graphique, il est intéressant de relier d'une part les points relatifs à un même produit (mise en évidence de la variabilité des évaluations d'un même produit) et d'autre part ceux relatifs à un même juge (comparaison des représentations des produits données par chaque juge). La représentation des barycentres des points relatifs à un même juge aide cette analyse (elle s'effectue en introduisant le tableau $J \times K$ en supplémentaire) ; elle montre dans quelle mesure les principaux facteurs de variabilité des produits coïncident avec des effets juges.

Jeu de données 2 : l'alignement des juges le long de la seconde bissectrice montre que les deux principaux facteurs de variabilité des produits sont liés au même effet juge : les juges J3 et J5 évaluent l'ensemble des produits comme étant acides, amers et peu salés ; c'est l'inverse pour les juges J2 et J4.

4.3 Tableau $J \times K$ (juges \times descripteurs)

4.3.1 Tableau et analyse (cf. Figure 3 a)

A l'intersection de la ligne j et de la colonne k du tableau $J \times K$, on trouve m_{jk} , moyenne, sur tous les produits, des notes attribuées par le juge j pour le descripteur k . La variabilité de ce tableau provient des seuls effets juges. Les liaisons entre descripteurs s'interprètent ici en termes de liaisons entre effets juges.

L'analyse usuelle de ce tableau est une ACP. Réaliser une ACP normée revient à accorder le même poids à chaque descripteur. Réaliser une ACP non normée revient à accorder à chaque descripteur un poids proportionnel à son écart-type. Le choix entre ces deux ACP s'appuie sur le fait que l'écart-type d'un descripteur est une mesure brute de l'effet juge pour ce descripteur.

Jeu de données 2 : on a choisi une ACP non normée. La relativement faible variabilité des écarts-types permet de penser que ce choix n'influence pas beaucoup les résultats. Le tableau analysé est schématisé figure 3a ; les divers éléments supplémentaires apparaissent progressivement ci-après. Compte tenu du pourcentage d'inertie associé au premier plan (100%), les graphiques de la figure 5 suffisent pour l'interprétation.

4.3.2 Problématique associée aux éléments actifs (cf. Figure 5, éléments entourés)

La problématique de l'ACP d'un tel tableau s'articule autour de deux thèmes principaux.

Bilan des liaisons entre descripteurs

Jeu de données 2 : *acide* et *amer* sont corrélés positivement entre eux et négativement à *salé* ; il existe des juges qui évaluent globalement l'ensemble des produits comme étant à la fois acides, amers et peu salés et des juges pour lesquels c'est l'inverse ; l'interprétation de ce

phénomène en termes de seuils de perception, de références ou d'utilisations de l'échelle de notation ne peut être qu'extérieure à ces données.

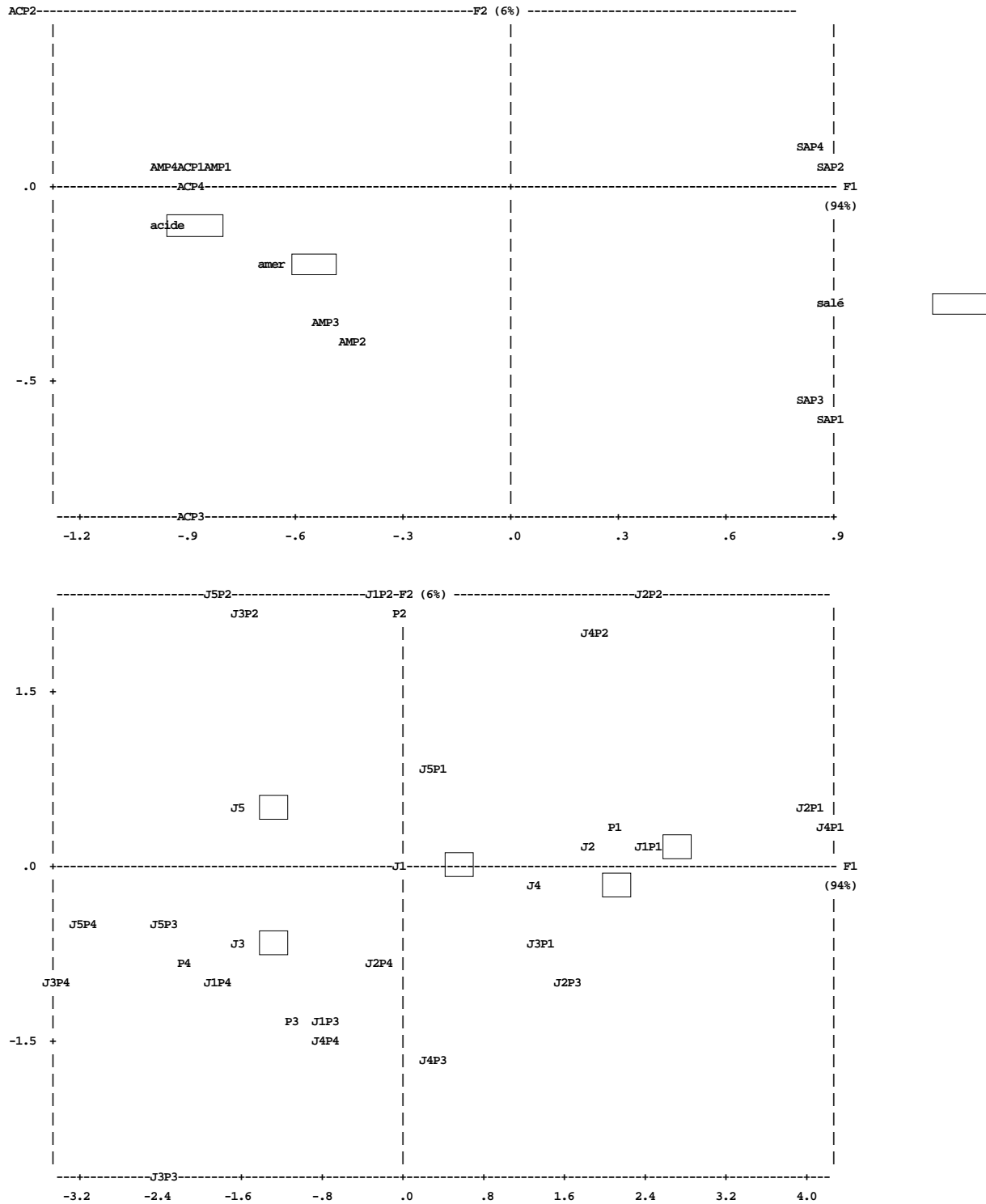


Figure 5. ACP non normée de $J \times K$ avec $(I \times J) \times K$ et $I \times (J \times K)$ en illustratifs (cf. fig. 3a).
 Variables : AMP2 : descripteur amertume utilisé dans l'évaluation du produit 2 ;
 Individus : P1 : produit 1 ; J1 : juge 1 ; J5P3 : produit P3 évalué par le juge J5.
 Les éléments actifs sont encadrés.

Bilan des ressemblances entre juges

Jeu de données 2 : le premier facteur de variabilité sépare d'une part les juges $J2$ et $J4$ (qui évaluent l'ensemble des produits comme étant salés, peu acides et peu amers) et d'autre part les juges $J3$ et $J5$ (qui évaluent l'ensemble des produits comme étant acides, amers et peu salés). Le juge $J1$ attribue des notes qui, en moyenne, sont les mêmes que celles de l'ensemble du jury.

4.3.3 Problématique associée aux éléments supplémentaires (cf. Figure 5)

Les interprétations de cette analyse concernent l'évaluation globale de l'ensemble des produits. Il est légitime de se demander, dans un deuxième temps, si ces interprétations valent pour chacun des produits pris séparément. On peut imaginer, par exemple, que certains produits sont plus propices que d'autres à l'expression d'un effet juge. Pour cela on introduit dans l'ACP différents éléments supplémentaires.

Point de vue des descripteurs

Les liaisons entre descripteurs, établies à partir de l'ensemble des produits, sont-elles observables pour chaque produit ? Pour répondre à ce type de question, on introduit les évaluations de chaque produit en tant que colonnes supplémentaires. Cela revient à juxtaposer (cf. **Fig. 3 a**) au tableau $J \times K$ le tableau $J \times (I \times K)$; le tableau obtenu possède J lignes et $(K+IK)$ colonnes.

Jeu de données 2 : les corrélations observées pour l'ensemble des produits valent globalement pour chaque produit. Plus en détail, on peut remarquer que la liaison globalement étroite entre *acide* et *amer* est assez lâche lorsque le jury évalue le produit $P2$. On peut imaginer, par exemple, qu'un produit présentant une caractéristique extrême est peu propice à l'expression d'un effet juge.

Point de vue des produits

Les variables synthétiques, qui donnent la meilleure représentation des juges du point de vue des produits considérés globalement, donnent-elles la même représentation lorsque l'on considère les produits séparément ? Pour cela, on introduit les évaluations individuelles en tant que lignes illustratives. Cela revient (cf. **Fig. 3a**) à juxtaposer au tableau $J \times K$ le tableau $(I \times J) \times K$. Par construction, la représentation précédente (J_j) du juge j apparaît au centre de gravité des représentations ($J_j P_i$) de ce juge j par chacun des produits.

Jeu de données 2 : la tendance des juges $J4$ et $J2$ à évaluer tous les produits comme étant salés, peu acides et peu amers, est particulièrement nette pour le produit $P1$. La parenté globale entre les juges $J4$ et $J2$ est très nette pour le produit $P1$ et moins pour le produit $P3$.

Comme dans l'analyse précédente (cf. § 4.2.3.), il est intéressant de relier d'une part les points relatifs à un même produit et d'autre part ceux relatifs à un même juge. La représentation des barycentres des points relatifs à un même produit aide cette analyse (elle s'effectue en introduisant le tableau $I \times K$ en supplémentaire) ; elle montre dans quelle mesure les principaux facteurs de variabilité des juges coïncident avec des effets produits.

Jeu de données 2 : les deux principaux facteurs de variabilité des juges sont liés à des effets produits. Ainsi, le produit $P1$ est jugé globalement salé, peu acide et peu amer.

4.4 Tableau ($I \times J$) $\times K$ (fiches \times descripteurs)

Les lignes de ce tableau représentent les fiches de dégustations, c'est-à-dire des perceptions (au sens multidimensionnel). L'objet de son analyse factorielle est donc :

- en premier lieu, la mise en évidence des principaux facteurs de variabilité des perceptions ; les liaisons entre variables peuvent provenir ici de n'importe lequel des trois niveaux (inter-produits, inter-juges et intra-juge) ;
- dans un second temps, la mise en relation de ces facteurs avec l'effet produit, l'effet juge et l'interaction juge \times produit.

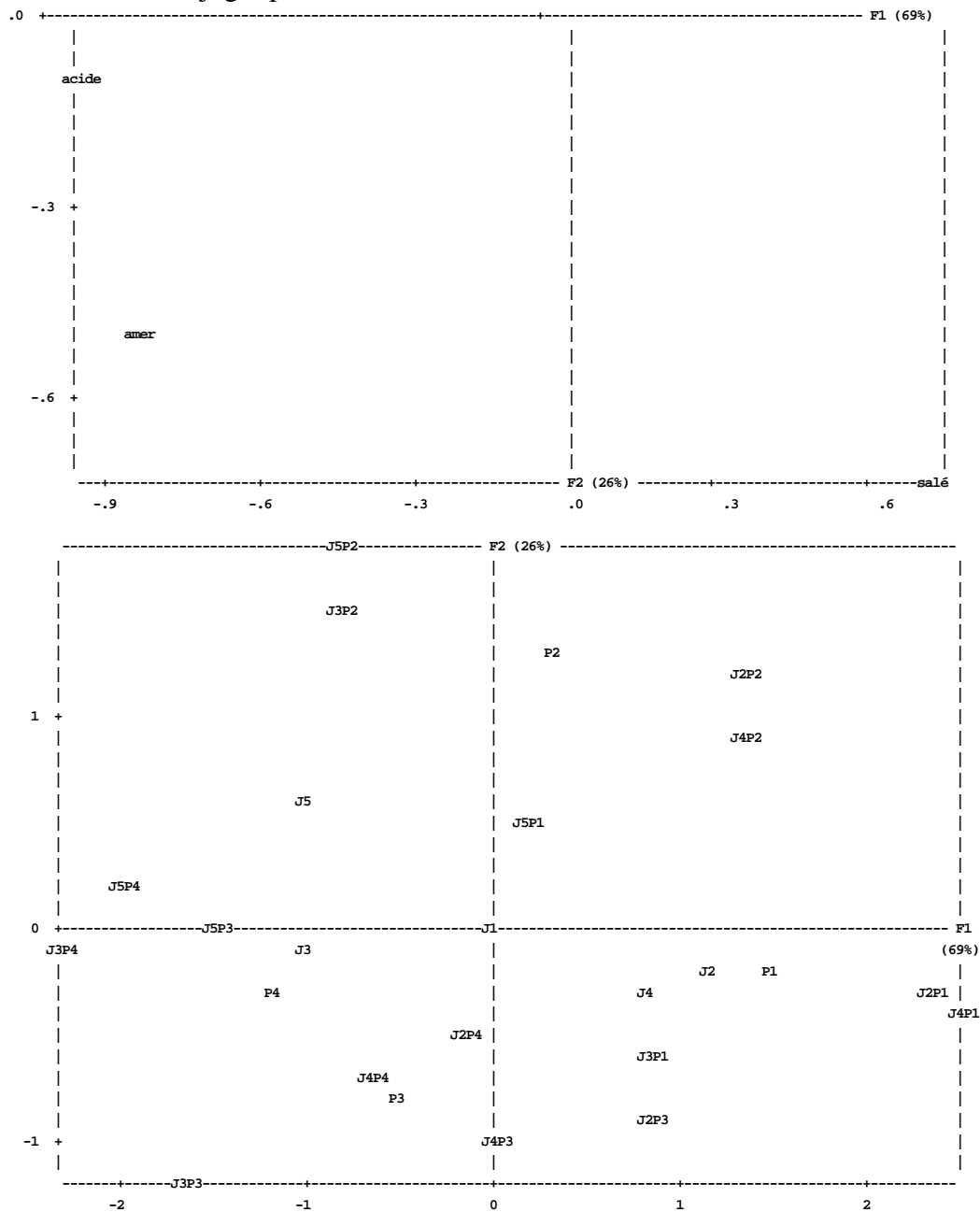


Figure 6. Premier plan de l'ACP non normée de ($I \times J$) $\times K$ (cf. figure 1).

P1 : produit 1 ; J1 : juge 1 ; J5P3 : produit P3 évalué par le juge J5.

Dans ce tableau, la variance d'un descripteur est la variabilité totale de l'analyse de variance associée à ce descripteur (cf. § 3.2.). Le choix entre une ACP normée ou non est à considérer dans chaque cas.

Nous ne commentons pas les résultats de cette analyse : les graphiques (cf. figure 6) montrent clairement que l'on retrouve les effets produits, les effets juges et les interactions des analyses précédentes. Soulignons simplement le grand intérêt de cette analyse qui permet, d'un point de vue multidimensionnel mais dans un unique référentiel, d'examiner les différents effets en jeu.

Remarque : dans cette analyse on projette le nuage des perceptions sur ses principaux facteurs de variabilité ; dans la première (resp. seconde) analyse, on projette ce même nuage sur les principaux facteurs de variabilité des points moyens des produits (resp. des juges).

Jeu de données 2 : selon la figure 6, le nuage des perceptions est pratiquement dans un plan (pourcentage d'inertie associé à ce plan : 95%) ; le nuage des points moyens des produits (cf. ACP du § 4.2.) et celui des points moyens des juges (cf. ACP du § 4.3.) sont donc inclus dans ce même plan ; ceci explique pourquoi, à des rotations des axes près, les deux premières ACP conduisent donc à un même plan.

4.5 Quelle(s) analyse(s) choisir ?

Les deux premières analyses ont été présentées de façon détaillée car elles permettent d'introduire successivement les différentes composantes des liaisons entre variables, et plus généralement les différents phénomènes que l'on peut mettre en évidence dans l'examen multidimensionnel de données sensorielles.

La troisième analyse, bien que non détaillée, illustre dans quelle mesure son point de vue permet de retrouver les principaux résultats des deux premières.

A ces analyses, on pourrait ajouter celle du tableau $I \times (J \times K)$ (avec $I \times K$ en supplémentaire) ou celle de $J \times (I \times K)$ (avec $J \times K$ en supplémentaire) ; ces deux analyses ont des objectifs proches des deux premières, en privilégiant toutefois l'examen des liaisons juge par juge ou produit par produit.

Toutes ces analyses ont des problématiques différentes ; celle du tableau $(I \times J) \times K$ constitue un bon compromis entre l'économie réalisée en examinant une seule analyse et la richesse des résultats obtenus. Elle est donc à effectuer en priorité. En complément on réalise une analyse :

- de $J \times K$ ou $J \times (I \times K)$ lorsque l'on souhaite obtenir une typologie des juges ;
- de $I \times K$ ou $I \times (J \times K)$ lorsque les effets juges sont très importants et soupçonnés de masquer des effets produits.

Remarque sur l'Analyse des Correspondances Multiples (ACM)

Pour illustrer cet exposé à l'aide d'un jeu de données restreint, il était nécessaire de considérer les variables comme quantitatives et de réaliser des ACP. On se souviendra que, dès qu'un tableau comporte beaucoup de lignes, il est avantageux de le soumettre à une ACM. Ceci est particulièrement vrai du tableau $(I \times J) \times K$, sur lequel une ACM peut être mise en oeuvre directement dès l'instant que l'échelle d'évaluation ne comporte pas trop de modalités.

5. Références

Cette article reprend un exposé réalisé aux journées technologiques sur l'évaluation et la gestion de la qualité dans l'industrie agro-alimentaire qui se sont déroulées au centre polytechnique (IUT +IUP) de Vannes les 13 et 14 juin 1994.

Les ACP ont été réalisées à l'aide du logiciel SPADN ; les calculs de probabilités ont été effectuées à l'aide du logiciel *LeProbabiliste*.

Benzécri J.P. (1972). *L'analyse des données ; tome 2 : correspondances*. Dunod 1973.

Cazes P. (1990). Quelques considérations sur l'utilisation des éléments supplémentaires en analyse factorielle. *Revue de Modulad*. n°5. INRIA.

Couronne T. (1996). Analyse statistique de la performance d'un jury. Thèse

Dagnélie P. (1986). *Théorie et méthodes statistiques*. Presses agronomiques de Gembloux.

Dijksterhuis, G., Gower, J. C. (1991). The interpretation of generalized Procrustes analysis and allied methods. *Food Quality and preference*, 3, 67-87.

Dijksterhuis, G. (1995). *Multivariate data analysis in sensory and consumer science*. Thèse. Rijksuniversiteit. Leiden.

Escoffier B., Pagès J. (1993). *Analyses factorielles simples et multiples ; objectifs, méthodes et interprétation*. Dunod. Paris.

Lavit Ch., (1988). *Analyse conjointe de tableaux quantitatifs*. Masson, Paris.

Macfie, H. J. H, Bratchell, N., Greenhoff, K., & Vallis, L. V. (1989). Designs to balance the effect of order of presentation and first-order carry-over effects in hall tests. *J. Sensory Studies*, 4 129-148.

Morineau A., Fayolle J.M. (1993). *LeProbabiliste*. Livre + logiciel diffusés par CISIA, 1, av. Herbillon 94160 Saint-Mandé.

Pagès J., (1994). Approche par estimation de l'essai triangulaire. *Journal de la société statistique de Paris*. tome 135, n°3, p21-44.

SPADN (1993). *Système Portable pour l'Analyse des Données*. Version 2.5. Logiciel diffusé par CISIA, 1, av. Herbillon 94160 Saint-Mandé.

Sztrygler F. et al., (1990). *Evaluation sensorielle ; manuel méthodologique*. Lavoisier. Paris.

Wakeling I., MacFie H. (1995). Designing consumer trials balanced for first and highr orders of carry over effect when only a subset of k samples from t may be tested. *Food Quality and Preference* 6 299-308.