



HAL
open science

La performance en analyse sensorielle, une approche base de données

Nicolas Pineau

► **To cite this version:**

Nicolas Pineau. La performance en analyse sensorielle, une approche base de données. domain_other. Université de Bourgogne, 2006. Français. NNT : . tel-00125171

HAL Id: tel-00125171

<https://theses.hal.science/tel-00125171>

Submitted on 18 Jan 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITE DE BOURGOGNE

THESE

présentée à

**L'ECOLE NATIONALE DE BIOLOGIE APPLIQUEE A LA
NUTRITION ET A L'ALIMENTATION**

pour obtenir le diplôme de DOCTORAT
Spécialité : Science de l'alimentation

Par

Nicolas PINEAU

**La performance en analyse sensorielle
Une approche base de données**

Soutenance prévue le 13 décembre 2006 devant la commission d'examen :

M. SAPORTA Gilbert	Professeur, CNAM, Paris	Président du jury
M. DANZART Marc	Professeur, ENSIA, Massy	Rapporteur
M. LECALVE Georges	Professeur honoraire, Rennes II	Examineur
M. QANNARI El Mostafa	Professeur, ENITIAA, Nantes	Directeur de thèse
M. SCHLICH Pascal	Chargé de Recherches, CESG, Dijon	Codirecteur de thèse

Remerciements

Je remercie Messieurs Gilbert Saporta et Marc Danzart, rapporteurs de ce manuscrit, ainsi que monsieur Georges Lecalvé pour m'avoir fait l'honneur de juger ce travail de thèse.

J'exprime toute ma reconnaissance à El Mostafa Qannari, pour avoir accepté et assuré la direction de ce travail. Je le remercie en particulier pour le temps qu'il a su me consacrer. Cela m'a chaque fois permis de faire le point sur mes travaux et de m'orienter dans la bonne direction pour poursuivre ce travail.

Je remercie également Pascal Schlich, co-directeur et initiateur de ce projet de thèse. Je mesure toute la chance que j'ai eu de travailler sur un tel sujet, très novateur à mon goût, et sur les opportunités de développement qui m'ont été offertes. Merci de m'avoir fait confiance, de m'avoir permis de participer à autant de congrès scientifiques, nationaux et internationaux, et de m'avoir donné l'occasion de travailler avec différentes personnes du domaine au niveau international. Je le remercie sincèrement pour l'excellent point de départ qu'il m'a donné dans la vie professionnelle.

Je remercie les centres ACTIA ainsi que le conseil régional de Bourgogne, financeurs de ce travail de thèse. Je remercie également les centres ACTIA en tant que premiers participants du projet SensoBase. Mes remerciements s'adressent en particuliers à monsieur Didier Majou, président de cette association, ainsi qu'aux personnes des différents centres que j'ai pu rencontrer lors de nos nombreuses réunions. Travailler avec eux tous fût un réel plaisir !

Je tiens à remercier Graham Cleaver, du centre de recherche d'Unilever, pour sa collaboration dans le projet SensoBase, ainsi que son accueil lors de la semaine passée dans son laboratoire. Nos discussions concernant les possibilités d'analyse de données longitudinales de la performance m'ont été très utiles.

Mes remerciements s'adressent également à Per Brockhoff, heureux papa de l'Assessor Model. Ses travaux sur ce modèle sont à l'origine d'un chapitre de ce manuscrit. Les discussions que nous avons pu avoir à plusieurs reprises ont toujours été très instructives et bénéfiques pour l'avancée de mes travaux. Je suis très heureux d'avoir pu échanger des idées sur les méthodologies statistiques avec lui.

J'exprime toute ma reconnaissance à Claire Chabanet pour sa grande implication et ses conseils avisés en matière d'analyses statistiques, en particulier concernant l'étude de l'évolution de la performance. Je la remercie sincèrement pour m'avoir initié au modèle mixte et à l'étendue de ses possibilités.

Je tiens également à remercier toutes les personnes du LIRIS avec qui j'ai pu travailler.

Merci à Delphine Brajon pour son formidable travail statistique et technique sur la SensoBase. Si ce projet est une réussite aujourd'hui, c'est en grande partie grâce à elle. Merci à Julien Delaborde, la SensoBase lui doit également beaucoup concernant la convivialité et l'ergonomie du site. Merci à Bérengère Fournier enfin, qui a su apprivoiser l'Assessor Model et l'implémenter dans la SensoBase. Merci à eux trois, leur aide me fût précieuse.

Je remercie aussi Sylvie Jedi, qui a souvent su me communiquer la force. Elle était impliquée à l'origine dans le projet SensoBase est c'est aujourd'hui à elle que revient la tâche de faire grandir ce beau bébé. J'espère que Pépette saura lui donner un coup de main !

Je remercie également, pour leur aide, leur gentillesse et pour tous les bons moments que nous avons pu passer ensemble, les autres membres de l'équipe, des plus anciens aux derniers arrivés. Merci donc à Corinne, Malvina, Catherine et ses gâteaux, les Caro, les Christine, Emilie, Denis, Rita et tous ceux qui sont restés plus ou moins longtemps au LIRIS.

Je remercie enfin les autres personnes avec qui j'ai pu travailler aux cours de ces années dijonnaises, en particulier Sylvie Issanchou et toute l'équipe de l'INRA, Laurent Brondel pour m'avoir impliqué dans deux de ses projets et Catherine Dacremont, responsable du master2 d'analyse sensorielle dans lequel j'ai pu donner des travaux dirigés de statistiques.

Je ne saurais oublier mes amis et ma famille qui ont aussi largement contribué au bon déroulement de ces années de thèse. Comprenez qui pourra :

Spéciale dédicace à Apéri et Kube, Mirette, Kiki et sa moitié, notre cuisto ami des xenopes, Sieur Dardou et tous les rôlistes occasionnels. A vos côtés, j'ai appris à surveiller mon langage et ne pas employer trop souvent : "Dégage !!!".

Merci à mes partenaires de badminton, Christophe, Christine et en particulier Coach avec qui c'était toujours un régal de s'affronter. Et merci pour cette finale de double à Amboise.

J'ai également été très heureux d'organiser ces tournois annuels de badmintons auxquels vous êtes venus nombreux. Ce fût toujours un plaisir de préparer ce petit évènement!

Merci également à tous ceux qui sont venus s'essayer à l'ultimate frisbee au cours de ces années. Je garderai de nombreux souvenirs d'olivierades et de maevades dignes des plus belles acrobaties. Et mention spéciale aux fidèles, Jé et Ced, qui étaient là par tous les vents !

Je n'oublie pas mes partenaires de jorky-ball et/ou de squash, Bruno, Vincent, Arnaud et Ced avec qui le sport était pour le moins... physique !!

The « Special Guide of the Empire State Building » vous remercie également pour cette inoubliable promenade au cœur de Yosemite park. Et surtout, merci d'avoir "voler" en éclairage pendant ces trois années de thèse. Une partie de la qualité de ce travail te revient directement.

J'exprime enfin ma profonde gratitude à mes parents et à ma sœur, toujours présents malgré la distance. Tout ceci n'aurait jamais été possible sans vous. Je vous laisse maintenant contempler votre travail. Vous pouvez être fiers de vous !

Merci à tous...

RESUME

La statistique a proposé beaucoup de méthodes pour évaluer les performances d'un panel d'analyse sensorielle. En revanche, peu de méthodes ont été proposées pour suivre l'évolution de cette performance au fil du temps. De plus, les niveaux moyens de performance en fonction des différents contextes de l'analyse sensorielle (types de produits, de descripteurs, d'échelles ...) n'ont jamais été documentés. Enfin, les comparaisons des mérites respectifs des méthodes statistiques de traitement des données sensorielles prennent rarement en compte cette diversité des contextes d'application. Elles se limitent en général à l'analyse de deux ou trois études sensorielles seulement. En accumulant plusieurs centaines de jeux de données d'analyse sensorielle descriptive et en les organisant dans une base de données (SensoBase), ce travail aborde ces trois questions simultanément.

Le suivi longitudinal des performances du sujet et du panel est appréhendé à l'aide d'une adaptation du modèle mixte et des cartes de contrôle. Cette méthodologie est appliquée à deux séries de profils sensoriels extraites de la SensoBase. La première série rassemble 8 études réalisées par un même panel sur 4 années, alors que la seconde rassemble 43 études réalisées par un autre panel sur une année. Plus de 300 études de la SensoBase sont ensuite utilisées pour mesurer les niveaux des performances généralement atteints en analyse sensorielle. L'analyse de ces données permet également de déterminer quelles variables individuelles (âge, sexe, niveau d'éducation, ...) ou liées aux caractéristiques de l'étude (type d'échelle, pays d'origine, ...) influent sur les performances. Enfin, la SensoBase est utilisée pour évaluer la fréquence de l'existence d'une hétérogénéité entre les sujets d'un panel d'analyse sensorielle en termes de répétabilité et d'étendue d'utilisation de l'échelle. Cette approche est basée sur une comparaison entre le modèle de Brockhoff et l'analyse de la variance classique.

La base de données, que ce travail a contribué à fonder, a permis, d'une part, de mieux connaître les capacités et les limites d'un panel d'analyse sensorielle et, d'autre part, d'offrir un banc d'essai et de comparaison des méthodes d'analyse des données sensorielles. A la frontière entre l'acquisition de connaissances en analyse sensorielle et le choix de méthodes d'analyse statistique, ce travail est intrinsèquement une thèse de sensométrie.

Mots clés : analyse sensorielle, sensométrie, performance, base de données, étude longitudinale, modèle mixte, cartes de contrôle, méta-analyse, Assessor Model, SensoBase

SUMMARY

Statistics proposed numerous techniques for the evaluation of panel performances in sensory analysis. However, few methods are dedicated to the monitoring of performances' evolution over time. Moreover, the performance levels, according to the various situations of evaluation (types of products, attributes, scales ...) have never been established. Last, the comparisons of statistical techniques for the analysis of sensory data are most often based on two or three datasets and thus do not take into account the variability of the different contexts of sensory analysis. Thanks to the gathering of hundreds of sensory profiling datasets and thanks to their organization into a database (SensoBase), this work addresses these three issues simultaneously.

A longitudinal analysis of panelist and panel performances is developed using a mixed model of analysis of variance and a corresponding adaptation of charts used in quality control. The results from two longitudinal series of sensory studies extracted from the SensoBase are presented. The first one features 8 studies from a given panel over 4 years whereas the second one features 43 studies from another panel over one year. More than 300 studies from the SensoBase are then used to establish the performance levels generally obtained in sensory analysis. The analysis makes also possible to determine which individual (age, gender, educational level ...) or group (type of scale, country ...) variables have an effect on the performances. Last, the SensoBase data are used to evaluate how often panelists are heterogeneous towards repeatability and scaling. This approach is based on a comparison between the so-called "Brockhoff's Assessor Model" and the regular analysis of variance.

This work has contributed to the building of a database from which a better understanding of sensory panel performances is gained. In addition, this database can be used as a benchmark for comparing different statistical techniques proposed for the analysis of sensory data. At the border between gaining knowledge in sensory analysis and choosing and developing statistical methodologies, this work is precisely a sensometrics thesis.

Keywords : sensory analysis, sensometry, performance, database, longitudinal study, mixed model, control chart, meta-analysis, Assessor Model, SensoBase

SOMMAIRE

INTRODUCTION.....	1
-------------------	---

PARTIE I : Méthodologies statistiques pour l'analyse de la performance

CHAPITRE 1 : REVUE BIBLIOGRAPHIQUE DES METHODES D'ANALYSE DE LA PERFORMANCE.....	9
---	----------

1	CONTEXTE SCIENTIFIQUE.....	9
2	DEFINITION DE LA PERFORMANCE.....	10
2.1	<i>La répétabilité.....</i>	13
2.2	<i>L'accord.....</i>	13
2.3	<i>La discrimination.....</i>	14
3	ANALYSES UNIDIMENSIONNELLES.....	15
3.1	<i>L'analyse de la variance au niveau du groupe.....</i>	15
3.1.1	Facteur sujet en effet fixe ou en effet aléatoire.....	17
3.1.2	Le terme d'interaction.....	18
3.2	<i>Autres modèles d'ANOVA au niveau du groupe.....</i>	20
3.2.1	L'effet session.....	20
3.2.2	Données sans répétitions.....	21
3.3	<i>Indices de performances au niveau individuel.....</i>	22
3.3.1	Tests de contrastes.....	22
3.3.2	Autres tests basés sur le modèle d'ANOVA au niveau du groupe.....	23
3.3.3	ANOVA par sujet.....	23
3.3.4	Coefficients de corrélation.....	24
3.3.5	Coefficient de variation.....	25
3.4	<i>Présentation synthétique des résultats unidimensionnels.....</i>	26
3.4.1	Méthode GRAPES.....	27
3.4.2	Méthode CAP.....	30
3.5	<i>Bilan des méthodes unidimensionnelles.....</i>	32
4	ANALYSES MULTIDIMENSIONNELLES.....	32
4.1	<i>L'analyse en composante principale.....</i>	33
4.1.1	ACP par descripteur.....	36
4.1.2	ACP par sujet.....	38
4.1.3	Autres ACP.....	38
4.2	<i>Analyse de la variance multivariée.....</i>	38
4.3	<i>Analyse Procustéenne Généralisée.....</i>	41
4.4	<i>Méthode STATIS.....</i>	43
4.5	<i>Analyse Factorielle Multiple.....</i>	47
4.6	<i>Méthodes avec pondération par dimension.....</i>	48
4.7	<i>Bilan des méthodes multidimensionnelles.....</i>	49
5	CONCLUSION.....	49

CHAPITRE 2 : SELECTION DES METHODES D'ANALYSE DE LA PERFORMANCE ET AUTOMATISATION D'UN OUTIL DE TRAITEMENT DES DONNEES.....	50
--	-----------

1	INTRODUCTION.....	50
2	PERFORMANCES UNIDIMENSIONNELLES.....	51
2.1	<i>La table CAP.....</i>	51
2.1.1	Performances du panel.....	51
2.1.2	Performances des panélistes.....	53
2.2	<i>La CAPMAP.....</i>	56
2.2.1	Construction et lecture de la carte.....	56
2.2.2	Diagnostics des sujets et des descripteurs.....	58
3	PERFORMANCES MULTIDIMENSIONNELLES.....	61
3.1	<i>Accord multidimensionnel.....</i>	61
3.2	<i>Discrimination multidimensionnelle.....</i>	63
4	ANALYSES DES DIFFERENCES ENTRE PRODUITS.....	65
4.1	<i>L'Analyse en Variables Canoniques (AVC).....</i>	65
4.1.1	Construction et interprétation.....	65
4.1.2	Sélection du nombre de descripteurs.....	67
4.2	<i>La table FLASH.....</i>	67
5	CONCLUSION.....	70

PARTIE II : Construction de la SensoBase et exploitation sensorielle et sensométrique

CHAPITRE 3 : LE SYSTEME SENSOBASE	74
1 ARCHITECTURE GENERALE	74
2 LE FICHIER EXCEL® SENSOBASE	75
3 LA BASE DE DONNEES	79
4 LE SITE INTERNET	80
4.1 Insertion d'une étude de profil sensoriel dans la base de données	82
4.2 Demande d'analyse d'une étude sensorielle	83
4.3 Conclusion	85
CHAPITRE 4 : EVOLUTION DE LA PERFORMANCE AU COURS DU TEMPS.....	86
1 CONTEXTE ET OBJECTIFS.....	86
2 MATERIEL ET METHODES	87
2.1 Données longitudinales	87
2.2 Indice de répétabilité.....	89
2.3 Modèle mixte	90
2.4 Choix de la structure de covariance.....	91
2.5 Choix de la partie fixe	94
2.6 Paramètres individuels.....	94
2.7 Cartes de contrôle	95
3 RESULTATS	97
3.1 Première étude	97
3.1.1 Modèle mixte : choix de la structure de covariance.....	97
3.1.2 Modèle mixte : choix de la partie fixe	100
3.1.3 Modèle mixte : paramètres individuels	100
3.1.4 Cartes de contrôle	102
3.1.5 Vérification de l'hypothèse de normalité des résidus	108
3.1.6 Conclusion	109
3.2 Seconde étude	110
3.2.1 Modèle mixte : choix de la structure de corrélation	110
3.2.2 Modèle mixte : choix de la partie fixe	111
3.2.3 Modèle mixte : paramètres individuels	112
3.2.4 Cartes de contrôle	113
3.2.5 Effet du nombre de dégustations par séance sur la répétabilité.....	116
3.2.6 Vérification de l'hypothèse de normalité des résidus	118
3.2.7 Conclusion	119
4 DISCUSSION	120
5 CONCLUSION.....	121
CHAPITRE 5 : META-ANALYSE DE LA PERFORMANCE.....	122
1 INTRODUCTION	122
2 MATERIEL ET METHODE.....	126
2.1 Les données SensoBase	126
2.2 Variables intra-étude.....	128
2.2.1 Description des variables	128
2.2.2 Indices de performance.....	130
2.2.3 Modèle d'analyse.....	131
2.3 Variables inter-études	133
2.3.1 Description des variables	133
2.3.2 Indice de performance	135
2.3.3 Modèle d'analyse.....	136
3 RESULTATS - DISCUSSION	137
3.1 Variables intra-étude.....	140
3.1.1 Effet de l'âge	140
3.1.2 Effet du sexe	141
3.1.3 Effet du niveau d'expérience	142
3.1.4 Effet du niveau d'éducation.....	142
3.1.5 Effet du type de descripteur	143
3.1.6 Interaction entre le type de descripteur et les autres variables	144
3.2 Variables inter-études	145
3.2.1 Le facteur "fournisseur"	145
3.2.2 Le type d'échelle	147

3.2.3	Le pays d'origine	148
3.2.4	Laboratoire privé ou public	149
3.2.5	Type de produit.....	149
4	DISCUSSION - CONCLUSION	150
CHAPITRE 6 : ETUDE DE L'ASSESSOR MODEL A L'AIDE DES DONNEES DE LA SENSOBASE		152
1	INTRODUCTION	152
2	ETUDE DE L'ASSESSOR MODEL ET COMPARAISON AVEC L'APPROCHE CLASSIQUE.....	153
2.1	<i>Matériel et Méthodes</i>	153
2.1.1	Le modèle classique.....	153
2.1.2	L'Assessor Model.....	154
2.1.3	Comparaison des résultats des deux modèles	159
2.2	<i>Résultats</i>	161
2.2.1	Tests de l'Assessor Model	161
2.2.2	Comparaison des deux modèles.....	162
2.3	<i>Conclusion</i>	165
3	TRANSFORMATION PROPOSEE PAR BROCKHOFF ET MODIFICATION DU MODELE.....	167
3.1	<i>Méthode</i>	167
3.2	<i>Résultats</i>	169
3.3	<i>Conclusion</i>	169
4	DISCUSSION-CONCLUSION	170
DISCUSSION GENERALE - CONCLUSION.....		173
1	EVOLUTION DE L'OUTIL STATISTIQUE DE TRAITEMENT DES DONNEES DE PROFIL SENSORIEL.....	173
2	EVOLUTION DE LA PERFORMANCE	175
3	META-ANALYSE DES NIVEAUX DE PERFORMANCE	176
4	ASSESSOR MODEL	179
5	CONCLUSION GENERALE	180
REFERENCES BIBLIOGRAPHIQUES.....		182

INTRODUCTION

L'analyse sensorielle est à l'origine une technologie dont l'objectif est de caractériser les propriétés sensorielles des aliments. Son utilisation s'est d'abord répandue dans les entreprises agroalimentaires, puis elle s'est étendue à d'autres secteurs industriels comme les cosmétiques, le textile ou l'automobile. Schématiquement, cette discipline se scinde en deux parties : la mesure des préférences des consommateurs d'une part et l'analyse descriptive des différences de perception sensorielle entre produits d'autre part. Lors d'une épreuve d'analyse descriptive, les sujets ne doivent pas donner un jugement en fonction de leurs préférences ; ils doivent uniquement différencier les produits en fonction de leur perception sensorielle, à la manière d'un instrument de mesure. L'épreuve principale de l'analyse descriptive consiste à réunir un jury de sujets et à l'entraîner à la description des caractéristiques sensorielles d'un type de produit. Lors de la phase de mesure, les sujets doivent noter l'intensité de la perception de plusieurs descripteurs sensoriels pour différents produits d'une même catégorie. Les données recueillies permettent d'obtenir le profil descriptif de chaque produit de l'étude. Dans ce manuscrit, nous focaliserons notre attention uniquement sur ce type de données.

Dans le secteur industriel, certaines entreprises mettent en place leurs propres jurys de panélistes entraînés, également appelés panels, alors que d'autres font appel à des sociétés de service spécialisées dans ce domaine. Le panel est un véritable outil de mesure permettant de caractériser les propriétés sensorielles des aliments ou d'autres produits pour différents descripteurs. Comme pour tout autre appareil de mesure, il est important que les données produites soient fiables. La mesure de la performance d'un panel et de ses panélistes est donc une question centrale en analyse sensorielle. Cette notion de performance, largement abordée dans la littérature, est le plus souvent divisée en trois concepts : la répétabilité, la discrimination et l'accord entre les panélistes (ISO, 1993).

- La **répétabilité** traduit la capacité à donner des résultats très proches en réponse à un même stimulus. Celle-ci est appelée reproductibilité lorsque le même stimulus est présenté dans des conditions différentes (changement du lieu ou du moment de l'expérimentation, par exemple). Le contrôle du niveau de répétabilité du panel et des panélistes est indispensable afin de maîtriser la qualité des résultats obtenus.

- La **discrimination** est l'aptitude à détecter les différences sensorielles entre les produits. Cette caractéristique est au centre des préoccupations du contrôle de la performance puisqu'elle indique si les panélistes ont été capables de différencier les produits de l'étude ou non.

- L'**accord** entre les panélistes mesure l'homogénéité des réponses obtenues pour le même stimulus par les différents dégustateurs. Cette caractéristique est héritée du fait que le panel,

contrairement à la majorité des outils de mesure « classiques », est composé de plusieurs sous-unités de mesures indépendantes : les panélistes.

Ces trois notions relatives à la performance peuvent être abordées sous différents angles. L'étude peut s'effectuer au niveau monodimensionnel (un descripteur sensoriel à la fois) lorsque l'objectif est d'obtenir une information détaillée sur chaque descripteur sensoriel. Elle peut aussi s'effectuer au niveau multidimensionnel (plusieurs descripteurs simultanément) si le but est de construire une image plus globale de la performance du panel qui tienne compte des corrélations entre les descripteurs. L'analyse peut également se faire selon un autre niveau de lecture qui sépare la performance du panel tout entier de la performance de chaque panéliste. Pour chacun de ces cas, il existe de nombreuses méthodes de traitements statistiques qui proposent différents indices de performance (Næs et Solheim, 1991 ; Schlich, 1994 ; Couronne, 1997 ; Qannari, MacFie et al., 1999). Piggott et Hunter, 1999) effectuent une synthèse des différentes méthodes afin de les regrouper par type d'analyse, mais le choix est encore très vaste pour l'expérimentateur. Une sélection parmi les différentes analyses semble indispensable afin de réunir toutes les informations nécessaires à la caractérisation de la performance sans pour autant utiliser des indicateurs redondants. Certains articles s'intéressent à la comparaison entre différentes méthodes afin de mettre en avant les forces et les faiblesses de chaque approche et ainsi faciliter le choix de la méthode (Dijksterhuis, 1995; Hunter et Muir, 1995 ; Brockhoff, 1994; Couronne, 2001). Il faut toutefois noter qu'il s'agit le plus souvent de comparaisons théoriques ou basées sur un ou deux exemples utilisant des données sensorielles réelles. Du point de vue de l'animateur de panel, cela peut alors sembler insuffisant pour choisir une méthode au détriment d'une autre. L'abondance de méthodes d'analyse de la performance offre donc une grande diversité d'indices disponibles, mais le choix de ces indices de performance reste une question importante en analyse sensorielle.

Avec le développement de l'analyse sensorielle, de nouvelles situations apparaissent et posent de nouvelles questions autour de la notion de performance. Les panels recrutés puis entraînés dans les entreprises privées réalisent des profils sensoriels de plus en plus fréquemment et sur des périodes de plus en plus longues. La question de l'évolution de la performance au cours du temps est alors soulevée. Derrière cette problématique générale, l'animateur de panel s'intéresse à la fois à l'évolution globale du niveau de performance du groupe et à l'évolution de chaque panéliste afin de détecter d'éventuelles améliorations ou dégradations au fil des évaluations. Il existe de nombreuses études qui discutent de l'intérêt de l'entraînement du panel en analyse sensorielle (Roberts et Vickers, 1994; Moskowitz, 1996; Chollet et Valentin, 2001; Labbe, Rytz et al., 2004) mais, à notre connaissance, très peu abordent le problème de l'évolution de la performance des panélistes. La modélisation de ces comportements pourrait permettre de définir des durées optimales d'entraînement des panélistes et des durées optimales de maintien des sujets au sein du groupe. En outre, la détermination de l'impact de certains facteurs sur la performance, comme le sexe, l'âge ou le niveau d'entraînement, est aussi une question importante. L'influence de ces facteurs sur la performance est un sujet peu abordé dans la littérature et les articles qui s'intéressent à ces questions proposent des conclusions qui reposent

sur peu de jeux de données. Enfin, il n'existe pas à l'heure actuelle de niveau de performance de référence qui permettrait de savoir si le niveau atteint pour un indice donné est satisfaisant. Quelques travaux ont néanmoins pour objectif de comparer les performances de plusieurs jurys issus de laboratoires différents sur un même ensemble de produits (Martin, Molimard et al., 2000; Pagès et Husson, 2001; McEwan, Hunter et al., 2002). Cependant, les résultats ne permettent pas de définir des niveaux généraux de performance en analyse sensorielle. Ce type d'information ne peut être obtenu que si de nombreux tableaux de données provenant d'horizons divers sont rassemblés.

Afin de répondre à toutes ces questions avec un niveau de confiance satisfaisant, il semble indispensable d'utiliser un grand nombre de données. C'est pourquoi l'origine de ce travail repose sur la volonté de construire un système permettant de recueillir des profils descriptifs chez les différents acteurs de l'analyse sensorielle dans le monde, qu'il s'agisse de sociétés privées ou de laboratoires publics. L'agrégation d'un grand nombre de données permettrait de documenter les pratiques de l'analyse sensorielle, afin de déterminer, par exemple, le nombre de panélistes utilisés en général dans une étude et la durée de l'entraînement. La finalité de cette mise en commun d'information est également de définir des niveaux moyens de performance en fonction de différents facteurs, comme le type de produit ou le type de descripteur. L'animateur de panel pourrait alors utiliser ces informations pour calibrer ses propres expérimentations. Le dernier objectif est de permettre la comparaison de méthodes statistiques sur un grand nombre de jeux de données afin de tester la pertinence de chaque méthode sur des données sensorielles.

Suite à ces considérations, une base de données de profils sensoriels, appelée SensoBase, est née au Laboratoire d'Interface Recherche-Industrie-Sensométrie (LIRIS) de Dijon. Sa construction effective a débuté en 2004, pendant ce travail de thèse. Les centres ACTIA (Association des Centres Techniques des Industries Agroalimentaires), co-financeurs de la thèse avec la région Bourgogne, furent les premiers à abonder cette base avec des données de profils sensoriels. Depuis, de nombreux laboratoires d'origines diverses ont pris part à ce projet. Pour ces « fournisseurs » de données, l'intérêt de la participation à ce programme est de recevoir une analyse statistique gratuite de chacune de leurs études. Cette analyse est basée sur l'utilisation de plusieurs méthodologies statistiques sélectionnées à partir d'une analyse bibliographique et des pratiques du laboratoire du LIRIS en matière de contrôle des performances. Les résultats retournés aux fournisseurs sont accompagnés d'une documentation qui explique chaque analyse.

Les fournisseurs les plus réguliers envoient en outre des jeux de données, recueillis avec les mêmes panélistes et les mêmes attributs sensoriels, qui se succèdent chronologiquement. Ce type d'information permet alors d'étudier l'évolution de la performance au fil des évaluations. Après moins de deux années d'existence, la SensoBase a déjà recueilli plus de 450 ensembles de données de profil et continue à recevoir des dizaines d'études sensorielles tous les mois.

L'ensemble du travail présenté dans ce manuscrit se situe à l'interface entre l'analyse sensorielle et les statistiques. D'un côté, les utilisateurs de l'analyse sensorielle, qui génèrent de plus en plus de

données, sont demandeurs d'informations et d'indicateurs sur la qualité des données produites afin de valider leurs conclusions sur les différences entre produits. Ils ne possèdent cependant pas toujours les moyens statistiques adéquats, ni le recul que permettrait une vision globale de l'ensemble des résultats du domaine. De l'autre côté, de nombreuses méthodologies de traitement sont proposées par les statisticiens. Néanmoins, leurs applications restent souvent limitées à quelques études, par manque de diffusion des méthodes mises au point et d'adéquation avec les questions de l'analyse sensorielle. La SensoBase, dont le but est de réunir un grand nombre de données sensorielles, peut donc être perçue comme un outil de communication entre les deux domaines. Cette structure va permettre de documenter les pratiques de l'analyse sensorielle, de mesurer des niveaux de performance et de déterminer les facteurs qui peuvent influencer ces niveaux. Les études rassemblées permettront également d'éprouver des méthodes statistiques sur un grand nombre de données réelles afin de diriger la discussion sur le choix des méthodes en fonction des résultats obtenus dans la pratique. Pour que la communication entre les deux domaines soit efficace, un autre domaine de compétence, plus technique, doit intervenir : l'informatique. Cette discipline occupe une place importante dans la réalisation de ce travail car elle fournit les outils nécessaires à la construction et à la gestion de la base de données. Elle intervient également dans l'automatisation des calculs des méthodes statistiques de mesure de la performance. Elle permet enfin la communication entre les fournisseurs et le LIRIS via la construction d'un site internet.

Le travail de recherche, relaté dans ce manuscrit, consiste tout d'abord à définir la notion de performance puis à développer une méthodologie statistique pour l'analyse des données de profil sensoriel. L'agrégation d'un grand nombre de données permet ensuite de s'intéresser aux facteurs de variation de la performance à différents niveaux. L'étude de l'évolution de cette performance au cours du temps renseigne sur la variabilité des résultats pour un panel et un fournisseur donné. La détermination des variables individuelles (sexe et âge des sujets...) et méthodologiques (type d'échelle, par exemple) qui influencent les niveaux de performance implique en revanche les données de tous les fournisseurs. La comparaison de méthodes statistiques d'un point de vue sensoriel utilise également la totalité des données.

La première partie de ce manuscrit, divisée en deux chapitres, aborde les méthodologies statistiques de l'analyse sensorielle.

Le chapitre 1 présente d'abord les concepts de la notion de performance, puis décrit les méthodes d'analyses statistiques utilisées en analyse sensorielle. Celles-ci sont comparées afin d'établir une sélection de méthodes adaptée aux attentes de l'animateur de panel.

Le second chapitre est consacré à la description approfondie des méthodes retenues dans la première section. Elle contient la justification des choix effectués. Cette sélection de méthodes correspond aux analyses restituées aux fournisseurs de données lorsqu'ils déposent un jeu de données dans la base. Cette partie présente, pour chaque analyse, des exemples de résultats et les conclusions qui en découlent au niveau sensoriel.

Les chapitres basés sur le développement et l'utilisation de la SensoBase sont réunis dans la seconde partie.

Le chapitre 3 présente l'architecture de la base de données, son mode de fonctionnement ainsi que les protocoles d'échange de données et de résultats entre les fournisseurs de données et le LIRIS.

L'étude de l'évolution de la performance au cours du temps fait l'objet du chapitre 4. Cette section propose une démarche basée sur l'utilisation du modèle mixte en analyse de la variance pour modéliser l'évolution du panel et de ses panélistes au fil du temps pour chaque descripteur. Cette approche est complétée par l'utilisation de cartes de contrôle pour définir un intervalle de confiance de l'évolution du panel et/ou des panélistes. Les résultats obtenus pour deux séries de jeux de données de profil provenant de différents fournisseurs sont ensuite présentés.

Le chapitre 5 dresse l'état des lieux des connaissances apportées par l'étude de la base de données dans son ensemble. Elle décrit la diversité des caractéristiques des jeux de données actuellement présents dans la base ainsi que les facteurs qui semblent avoir une influence sur le niveau de la performance, comme, par exemple, l'âge ou le type de descripteur. L'étude d'un grand nombre de jeux de données relève du domaine de la méta-analyse. Ce terme, introduit par Gene Glass en 1976, représente une démarche, plus qu'une simple technique, qui a pour but de combiner les résultats de plusieurs essais réalisés dans des conditions variables (Glass, 1976). Elle met en avant les problèmes de représentativité et d'échantillonnage des données recueillies. Dans le cadre de ce travail, l'utilisation de cette démarche permet, en partie, de prendre en compte l'hétérogénéité des données pour mesurer les effets des différentes variables qui peuvent influencer la performance.

Le dernier chapitre présente la comparaison des résultats entre deux modèles d'analyse univariés : le modèle classique d'analyse de la variance de données de profil (produit + sujet + produit*sujet) d'une part et l'« Assessor model » proposé par Per Brockhoff en 1994 d'autre part. Cette méthode propose un modèle dans lequel l'hypothèse majeure est que les différences de notation entre les panélistes sont liées à des différences de dispersion des notes sur l'échelle et non au désaccord sur les différences entre produits. La comparaison porte sur la validité de ce modèle, la pertinence des éléments de réponse supplémentaires apportés par cette analyse et la confrontation des résultats obtenus avec ceux du modèle classique.

Cette thèse a fait l'objet des communications suivantes :

- Publications

Pineau, N., Chabanet, C., Schlich, P. (2006). Evolution of the repeatability of a panel in sensory analysis: a mixed model and control chart approach. (*Acceptée dans Journal of sensory Studies*).

Pineau, N., Brajon, D., Cordelle, S., Schlich, P. (2006). Meta analysis of panellist performances from hundreds of sensory profiling datasets. (*Soumise dans Food Quality and Preference*).

Fournier, B., Schlich, P., **Pineau, N.**, Brockhoff, P.B. (2006). Confronting the Brockhoff's Assessor Model to the usual ANOVA model on 264 datasets. (*en préparation*).

- Présentations orales

Pineau, N., Chabanet, C., Schlich, P. (2006). Utilisation du modèle mixte et des cartes de contrôle pour le suivi des performances d'un panel d'analyse sensorielle au cours du temps. *Proceedings of the "9èmes Journées Agro-Industries et Méthodes Statistiques" symposium*, Montpellier, January 24-27, pp61-67.

Schlich, P., **Pineau, N.**, Brajon, D., Cordelle, S., (2006). Le projet Sensobase: construire une base de profils sensoriels pour documenter les performances des panels d'analyse sensorielle. *Proceedings of the "9èmes Journées Agro-Industries et Méthodes Statistiques" symposium*, Montpellier, January 24-27, pp101-106.

Fournier, B., Schlich, P., **Pineau, N.**, Brockhoff, P.B. (2006). Confronting the Brockhoff's Assessor Model to the usual ANOVA model on 264 datasets. *8th Sensometrics meeting*, As, Norway, August 02-04. p10.

Pineau, N., Brajon, D., Cordelle, S., Schlich, P. (2005). Mining a database of sensory descriptive analyses to understand panellist performances. *6th Pangborn Sensory Science Symposium*, Harrogate, UK, August 07-11.

Pineau, N., Schlich, P. (2005). Présentation d'un outil de contrôle des performances en Analyse Sensorielle. *XI^{ème} Forum des Jeunes Chercheurs*, Dijon, June 09-10. P80.

Chabanet, C., **Pineau, N.** (2005). Using linear mixed models to handle variability of consumer's liking. *6th Pangborn Sensory Science Symposium*, Harrogate, UK, August 07-11.

Schlich, P., **Pineau, N.**, Brajon, D., Qannari, E.M., (2004). Multivariate Control of Assessor Performances. *7th Sensometrics meeting*, Davis, California, July 28-30. p37.

Schlich, P., Brajon, D., **Pineau, N.**, (2004). Le contrôle des performances en analyse sensorielle. *SAS Forum*, Paris, November 9-10.

- Présentations par affiche (disponibles à la fin des annexes)

Pineau, N., Schlich, P., Brajon, D., Cordelle, S. (2006). Meta-analysis of panel performances from a database of hundreds of datasets. *8th Sensometrics meeting*, As, Norway, August 02-04. p102.

Schlich, P., **Pineau, N.**, Brajon, D. (2005). Control of Assessor Performances: display of the CAP-V3 outputs. *6th Pangborn Sensory Science Symposium*, Harrogate, UK, August 07-11.

Communications réalisées en parallèle avec ce travail de thèse :

- Publications

Pineau, N., Cordelle, S., Mathonière, C., Etiévant, P., Issanchou, I., Imbert, A., Rogeaux, M., Schlich, P. (2006). Temporal Dominance of Sensations: a new technique to record several sensory attributes simultaneously over time. (*Soumise dans Food Quality And Preference*).

Perruchini, J.M., Rabec, C., **Pineau, N.**, Van Wymelbeke, V., Vassard, C., Jeannin, L. and Brondel, L. (2006). Effects of home-based respiratory maintenance program following conventional outpatient pulmonary rehabilitation program in COPD patients. Results of 9 months follow-up. (*Soumise dans Thorax*).

Brondel, L., Davenne, D., Callard, D., Guillard, J.C., **Pineau, N.** and Van Hoecke, J. (2006). Fluctuation of exercise induced hypoxemia during 24h cycling in elite endurance athletes. (*Soumise dans J. Appl. Physiol.*).

- Présentations orales

Pineau, N., Pessina, R., Cordelle, S., Imbert, A., Rogeaux, M., Schlich, P. (2004). Temporal Dominance of Sensations: comparison with conventional profiling and time-intensity. *7th Sensometrics meeting*, Davis, California, July 28-30. p40.

Pineau, N., Cordelle, S., Schlich, P. (2004). Dominance Temporelle des Sensations, une nouvelle méthode d'acquisition de données sensorielles temporelles. *X^{ème} Forum des Jeunes Chercheurs*, Besançon, June 09-10. p30.

Pineau, N., Cordelle, S., Schlich, P. (2004). Méthodes d'acquisition, de codage et d'analyse de profils sensoriels temporels. *Proceedings of the "8èmes Journées Agro-Industries et Méthodes Statistiques" symposium*, Rennes, March 10-12. pp87-93.

Pineau, N., Cordelle, S., Imbert, A., Rogeaux, M., Schlich, P. (2003). Dominance Temporelle des Sensations. Codage et analyse d'un nouveau type de données sensorielles. *Proceedings of the "XXXV^{èmes} Journées de Statistiques" symposium*, Lyon, June 2-6. pp777-780.

- Présentation par affiche

Pineau, N., Cordelle, S., Schlich, P. (2003). Temporal Dominance of Sensations: a new technique to record several sensory attributes simultaneously over time. *5th Pangborn Sensory Science Symposium*, Boston, July 20-24.

Partie I

Méthodologies statistiques pour l'analyse de la performance

Chapitre 1 : Revue bibliographique des méthodes d'analyse de la performance

1 Contexte scientifique

Lors de la réalisation d'un profil descriptif, la première étape consiste à sélectionner puis entraîner un panel de juges à l'évaluation d'un type de produit. Les protocoles standards d'entraînement (ASTM, 1981; AFNOR, 1993) préconisent plusieurs épreuves afin que les panélistes apprennent à utiliser correctement les échelles de notation, à se familiariser avec les termes descriptifs et à en donner une interprétation sensorielle en accord avec les autres panélistes. En parallèle avec cet apprentissage, l'animateur de panel établit la liste des descripteurs qui seront utilisés lors de la phase de mesure. Ce dernier se base sur les résultats obtenus pendant l'entraînement et sur la discussion menée avec les panélistes autour des différentes caractéristiques sensorielles des produits étudiés. La notation s'effectue sur des échelles dont les caractéristiques peuvent varier (continues ou discrètes, de longueur différentes, en présence de libellés plus ou moins détaillés).

L'intérêt de la phase d'entraînement a souvent été étudié en comparant les résultats de panels entraînés et de panels non entraînés, ou de panels avant et après entraînement. Certains articles ne mettent pas (ou peu) en évidence l'effet de l'entraînement sur le niveau de performance du panel. En particulier, la discrimination n'est pas toujours améliorée avec un panel entraîné par rapport à un panel de consommateurs naïfs (Roberts et Vickers, 1994; Gou, Guerrero et al., 1995; Moskowitz, 1996). Les auteurs émettent l'hypothèse que la signification des descripteurs n'est pas la même pour les sujets non entraînés et pour les sujets qui ont été entraînés à associer un terme verbal à une sensation gustative particulière. Dans certains cas, cela peut même conduire des sujets non entraînés à discriminer les produits pour un descripteur avec lequel un panel de sujets entraînés n'a trouvé aucune différence à cause de la confusion sur la signification du terme descriptif (Roberts et Vickers, 1994). Néanmoins, la majorité des études montrent que l'entraînement permet d'accroître le niveau de discrimination du panel (Clapperton et Piggott, 1979; Moskowitz et Jacobs, 1988; Sawyer, Cardello et al., 1988; Labbe, Rytz et al., 2004) et de détecter des différences entre les produits sur un plus grand nombre de descripteurs (Moskowitz, Jacobs et al., 1985; Wolters et Allchurch, 1994). L'entraînement

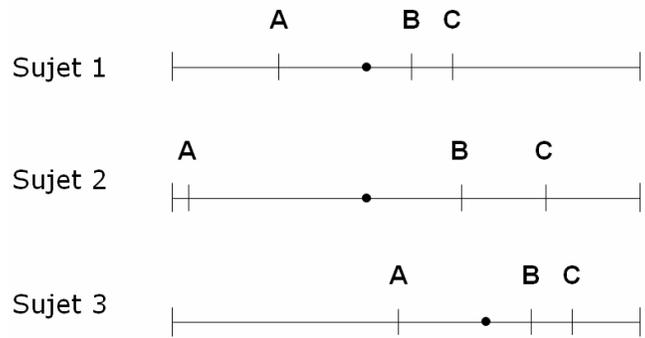
contribue également à augmenter la répétabilité des panélistes et l'accord entre les sujets (Roberts et Vickers, 1994; Chollet et Valentin, 2001).

Le contrôle de la performance est donc une question essentielle en analyse sensorielle. Cela permet d'orienter le travail à effectuer au cours de l'apprentissage et de valider la qualité du panel à la fin de l'entraînement. Cette question est tout aussi cruciale lors de la phase de mesure. Il s'agit alors de valider la qualité des données. La suite de ce chapitre définit la notion de performance et présente les outils statistiques disponibles pour analyser les performances d'un panel ayant réalisé un profil sensoriel.

2 Définition de la performance

Le panel d'analyse sensorielle peut être considéré comme un outil de mesure des propriétés sensorielles des produits testés. Comme tout autre outil, il doit donc fournir une mesure juste (Piggott et Hunter, 1999), c'est-à-dire que la note attribuée à un descripteur doit refléter l'intensité de la sensation dans le produit de la manière la plus exacte possible. Cependant, la mesure de l'intensité d'une caractéristique sensorielle dépend non seulement du produit testé mais également des capacités sensorielles du panéliste. Ces capacités étant intrinsèquement différentes, de par la variabilité du nombre de récepteurs sensoriels pour chaque molécule selon les individus (**Prutkin et al., 2000**), la perception sensorielle d'un stimulus donné est différente pour chaque panéliste. Il n'existe donc pas d'intensité « exacte » dans l'absolu. La justesse ne peut donc pas être évaluée à proprement parler lors de la réalisation d'un profil descriptif. De plus, des phénomènes d'interaction interviennent fréquemment entre les descripteurs sensoriels ce qui rend encore plus difficile l'estimation de l'intensité propre à chaque descripteur dans un produit (Pritchett-Mangan, 1992).

La présentation de références pour chaque descripteur sensoriel peut néanmoins permettre au panel de calibrer en partie sa notation (AFNOR, 1993). L'objectif de la calibration est de réduire les différences d'utilisation de l'échelle. Cette notion recouvre deux aspects : le niveau moyen de notation des produits et la dispersion des notes sur l'échelle. Sur la figure 1-1, le niveau de notation est le même pour les sujets 1 et 2 mais la dispersion des notes est différente. Pour les sujets 1 et 3, la dispersion des notes est identique mais les niveaux de notation sont différents. Malgré les différences d'utilisation de l'échelle, ces cas de figure reflètent la même perception sensorielle des différences entre les stimuli. Ces différences ne devraient donc pas intervenir dans la mesure de la performance du panel et de ses panélistes.

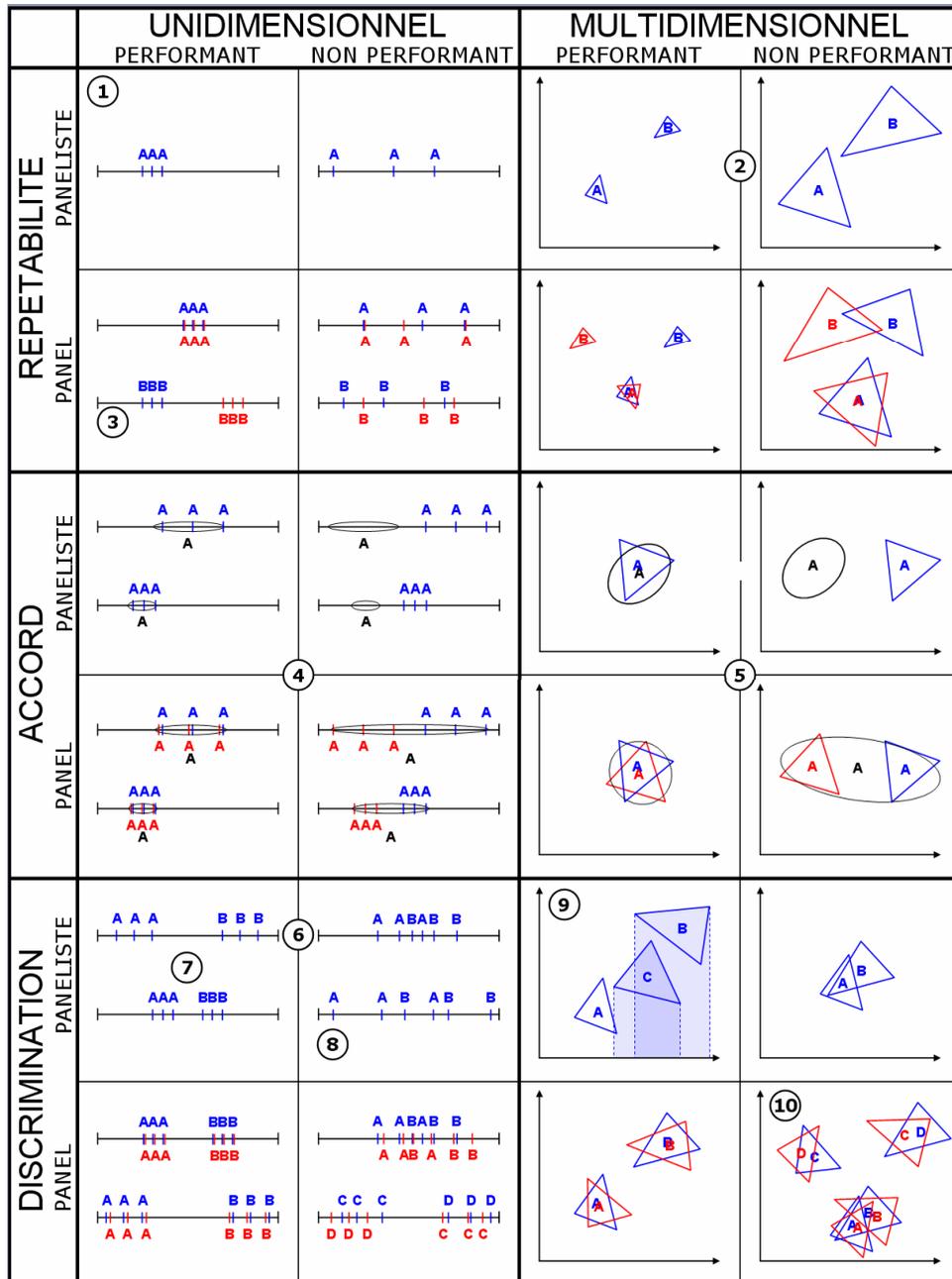


A, B et C : 3 produits testés; • : moyenne des trois produits
 Sujet 1 / 2 : dispersion des notes différente
 Sujet 1,2 / 3 : niveau moyen de notation différent

Figure 1-1 : Différences d'utilisation de l'échelle

La performance d'un panel d'analyse sensorielle se mesure en fonction de la précision des résultats : lors de la présentation répétée d'un même stimulus, les mesures doivent donc être aussi proches que possible les unes des autres. En analyse sensorielle, la précision de la mesure est également appelée la répétabilité. Enfin, la qualité d'un outil de mesure est aussi appréhendée par sa sensibilité, c'est-à-dire sa capacité à séparer des stimuli (produits) différents. En analyse sensorielle, la sensibilité correspond à la notion de discrimination entre les produits. Le panel, en tant qu'outil de mesure, possède en outre la particularité d'être composé de plusieurs juges. Cela implique que la répétabilité et la discrimination peuvent être évaluées au niveau du groupe ou au niveau de chaque panéliste. Cela implique également que la mesure de l'accord entre les juges constitue un aspect supplémentaire de la performance de cet outil. Enfin, ces notions peuvent être abordées au niveau unidimensionnel (étude d'un seul descripteur à la fois) ou multidimensionnel (plusieurs descripteurs simultanément).

La figure 1-2 propose une illustration des notions de répétabilité, discrimination et accord en fonction du niveau d'analyse adopté (panel ou panéliste, unidimensionnel ou multidimensionnel). Au niveau unidimensionnel, les échelles de notations sont représentées par des lignes horizontales bornées à leurs deux extrémités. Au niveau multidimensionnel, les deux axes représentent deux descripteurs de l'étude.



Produits : A, B, C, D. Sujets : bleu, rouge, groupe. Les représentations supposent que chaque sujet a évalué plusieurs produits et que les données ont ensuite été centrées par sujet. Chaque représentation ne fait pas apparaître tous les produits testés

- ① Répétabilité = notes des répétitions d'un même produit proche
- ② Répétabilité multidimensionnelle = petitesse de la taille des triangles (dont chaque sommet correspond à une répétition)
- ③ Panel répétable n'est pas synonyme de panel en accord
- ④ Accord = position identique des produits sur l'échelle (données centrées par sujet). Ellipses = intervalle de confiance de la moyenne du produit A pour le groupe.
- ⑤ Accord multidimensionnel = recouvrement des triangles et/ou des ellipses (données centrées par sujet). Ellipses = intervalle de confiance de la moyenne des notes des deux sujets.
- ⑥ Discrimination = séparation des produits. La répétabilité du sujet est la même dans les deux situations
- ⑦ Le niveau de discrimination est le même dans les deux situations mais les répétabilités (et les dispersions de produits) sont différentes
- ⑧ La non répétabilité est trop importante pour que le sujet puisse être discriminant
- ⑨ Discrimination multidimensionnelle = séparation des triangles dans l'espace. La discrimination multidimensionnelle n'implique pas la discrimination pour au moins un des descripteurs de l'étude.
- ⑩ Le fait que chaque sujet soit discriminant n'implique pas que le groupe entier est discriminant. Cela dépend de l'accord entre les sujets.

Figure 1-2 : Illustration des concepts de répétabilité, discrimination et accord

2.1 La répétabilité

La répétabilité est définie par l'AFNOR (1993) comme « l'étroitesse de l'accord entre les résultats successifs obtenus avec la même méthode sur une matière identique soumise à l'essai dans les mêmes conditions (même opérateur, même appareil, même laboratoire et court intervalle de temps) ». Cette information n'est donc accessible que lorsque des répétitions des mesures sont effectuées. Sur une échelle de notation, une répétabilité élevée se traduit par le fait que chaque observation est très proche de la moyenne des répétitions. Sur la figure 1-2, on suppose dans tous les cas présentés que le ou les sujets ont effectué trois répétitions de la dégustation d'un ou plusieurs produits (A, B, C et D) sur un (cas unidimensionnel) ou deux (cas multidimensionnel) descripteurs. Pour chaque cas, le critère de performance envisagé est meilleur pour la situation de gauche que pour la situation de droite. Un niveau élevé de répétabilité individuelle, illustrée dans la situation 1 de la figure 1-2, correspond à une situation où tous les scores donnés à un même stimulus se situent dans une région restreinte de l'échelle (niveau unidimensionnel). Au niveau multidimensionnel la répétabilité peut être assimilée à la taille des triangles délimités par les trois répétitions, les triangles les plus petits indiquant les sujets les plus répétables (situation 2). La répétabilité au niveau du groupe correspond à la moyenne des répétabilités individuelles. Elle ne dépend pas des différences de notation des produits entre les sujets. Par exemple, le niveau de répétabilité du groupe est identique dans les deux illustrations de la situation 3.

2.2 L'accord

L'accord correspond au degré de ressemblance qui existe entre les notations des sujets pour différents produits. Deux sujets sont d'autant plus en accord que leurs écarts entre produits sont similaires. Cependant, le niveau moyen de notation (sur tous les produits) propre à chaque sujet et l'étalement des notes sur l'échelle ne doivent pas être considérés comme des sources de désaccord. Sur la figure 1-2, les représentations concernant l'accord supposent que le niveau moyen de notation et l'étalement des notes sont les mêmes pour tous les sujets du panel. L'accord correspond au recouvrement des plages de notes sur l'échelle (niveau unidimensionnel, situation 4) ou des triangles dans l'espace des descripteurs (niveau multidimensionnel, situation 5) pour des produits identiques. Les ellipses représentent la variabilité de la notation entre les sujets pour un produit, donc le désaccord au niveau du groupe (leur taille est d'autant plus grande que le désaccord est grand).

Lors de la réalisation d'un profil descriptif, l'indice d'accord pour un sujet peut se concevoir comme la moyenne des accords de ce sujet avec chaque autre panéliste, comme l'accord du sujet avec le reste du groupe, ou encore comme l'accord du sujet avec le groupe (y compris lui-même). Cette dernière stratégie peut toutefois sembler moins satisfaisante puisque les notes du sujet considéré interviennent dans le calcul de la notation moyenne du groupe. Il n'y a donc pas indépendance entre les données du sujet et les données du groupe. La notion d'accord au niveau du groupe tout entier correspond à l'homogénéité de la notation des sujets. Elle peut se concevoir comme la moyenne des indices d'accord individuels ou être estimée directement en mesurant la variabilité des notes du groupe.

Si l'utilisation de l'échelle (niveau moyen et dispersion des notes sur l'échelle) est différente d'un sujet à l'autre, il est nécessaire de procéder à une transformation des données pour évaluer l'accord

correctement. En pratique, afin d'harmoniser les niveaux moyens de notation, il suffit de centrer les données par juge, mais il n'est pas aussi évident de réaliser la transformation adéquate pour homogénéiser l'étalement des notes sur l'échelle entre les sujets. Plusieurs approches décrites dans la littérature seront détaillées ultérieurement (paragraphe 3.1.2).

2.3 La discrimination

Le pouvoir discriminant peut se définir comme la capacité à détecter une différence d'intensité entre deux ou plus de deux échantillons pour un descripteur donné (Lesschaeve, 1997). Il est relié à la notion de répétabilité puisqu'il est d'autant plus difficile d'être discriminant que le niveau de répétabilité est faible. Mais il est également possible qu'un sujet moins répétable qu'un autre soit pourtant plus discriminant. Typiquement, une personne pour laquelle toutes les notes données étaient 0 est très répétable mais n'est pas du tout discriminante, alors qu'une personne peu répétable mais qui utilise la totalité de l'échelle pour effectuer sa notation peut être discriminante. Les illustrations de la figure 1-2 montrent d'une part qu'à niveau de répétabilité égal, un sujet peut être discriminant ou non (situation 6), et d'autre part que le niveau de discrimination peut être le même pour des niveaux de répétabilité différents (situation 7). La dernière échelle représentée montre en revanche que lorsque le niveau de répétabilité est vraiment très faible (très forte variabilité des notes), il n'est pas possible d'obtenir un bon niveau de discrimination (situation 8). Pour résumer, la discrimination peut se concevoir comme le rapport entre la variabilité des moyennes des notes pour les différents produits (variabilité inter-produits) et la variabilité des notes des répétitions de chaque produit (variabilité intra-produit). Au niveau multidimensionnel, la discrimination se traduit par la séparation des triangles (pour des produits différents) dans l'espace. Notons cependant que dans la situation 9 de la figure 1-2, le sujet discrimine bien les produits B et C (triangles séparés) au niveau multidimensionnel sans pour autant être discriminant sur l'un ou l'autre des descripteurs (niveau unidimensionnel). A l'inverse, un panéliste discriminant pour au moins un descripteur est obligatoirement discriminant au niveau multidimensionnel puisqu'il existe au moins une direction de l'espace pour laquelle les produits sont différenciés. Dans les analyses multidimensionnelles, il est donc rare que les sujets ou le panel ne soient pas du tout discriminants, car il suffit, en théorie, d'une seule variable ou combinaison de variables pour laquelle les produits sont différenciés pour que l'analyse multidimensionnelle indique que le panel ou les panélistes sont discriminants. Cependant, les calculs tiennent également compte de la taille de l'espace multidimensionnel et tendent à réduire en partie cet effet d'inflation du pouvoir discriminant. Comme pour le cas unidimensionnel, la discrimination est reliée à la répétabilité. Au niveau du groupe, même si chaque panéliste est discriminant individuellement, il est possible que le panel ne soit pas discriminant si les sujets ne sont pas en accord (produits C et D de la situation 10). Un panel discriminant dans son ensemble indique donc que les sujets sont suffisamment en accord pour qu'un consensus sur les différences entre produits soit obtenu.

Au niveau multidimensionnel, le nombre de descripteurs pour lesquels un sujet ou un panel est discriminant est également une information qui permet de caractériser la performance. Du point de vue de l'animateur de panel, il peut sembler plus judicieux d'avoir des panélistes relativement peu discriminants sur un grand nombre de descripteurs que très discriminants sur un ou deux descripteurs

uniquement ; cela indique que le sujet est capable de percevoir davantage la complexité du produit. Cette approche conduit à la notion de dimensionnalité de l'espace des descripteurs, qui peut être abordée de plusieurs façons. Une manière triviale de répondre à cette question serait de compter le nombre de descripteurs pour lesquels un sujet ou le panel est discriminant.

Répétabilité, discrimination et accord sont trois notions différentes qu'il est utile de prendre en compte conjointement pour caractériser la performance d'un sujet ou d'un panel dans son ensemble. Bien que ces notions soient différentes, elles ne sont pas complètement indépendantes, ce qui peut parfois poser des problèmes dans le calcul d'indices correspondants et dans leur interprétation. De nombreuses méthodologies statistiques ont été développées pour obtenir des éléments de réponse concernant un ou plusieurs aspects de la performance, au niveau unidimensionnel ou multidimensionnel, individuel ou sur le panel. La suite de ce chapitre propose une synthèse bibliographique et une comparaison de ces différentes approches au niveau unidimensionnel d'une part et multidimensionnel d'autre part.

3 Analyses unidimensionnelles

L'analyse des données au niveau unidimensionnel permet de s'intéresser aux résultats relatifs à chaque descripteur. Nous nous situons dans le cadre du traitement des données issues d'un profil descriptif pour lequel I produits ont été dégustés par J sujets lors de K répétitions et P descripteurs. Y_{ijk}^p , représente donc la note donnée au produit i par le sujet j lors de la répétition k pour le descripteur p.

3.1 L'analyse de la variance au niveau du groupe

La méthodologie statistique la plus employée pour l'analyse de données de profil est l'analyse de la variance (ANOVA). Elle est disponible dans la quasi-totalité des logiciels de statistiques.

Le modèle d'ANOVA le plus classique pour l'étude de données de profil au niveau du panel (pour un descripteur) est le suivant :

$$\begin{aligned}
 Y_{ijk} &= \mu + \alpha_i + B_j + C_{ij} + \varepsilon_{ijk} \\
 B_j &\sim N(0, \sigma_{\text{suj}}^2) \\
 C_{ij} &\sim N(0, \sigma_{\text{suj*prod}}^2) \\
 \varepsilon_{ijk} &\sim N(0, \sigma_c^2) \quad \textit{indépendants entre eux}
 \end{aligned}
 \tag{1-1}$$

Où μ représente l'effet moyen global, α_i correspond à l'effet fixe produit, B_j à l'effet aléatoire sujet, C_{ij} à l'interaction (aléatoire) sujet*produit et ε_{ijk} à l'erreur résiduelle (aléatoire) du modèle. Les variables aléatoires sont supposées être indépendantes deux à deux et distribuées selon une loi Normale,

chacune avec sa variance spécifique. Les calculs des degrés de liberté, des sommes des carrés ainsi que les statistiques de Fisher sont donnés dans le tableau 1-1.

Tableau 1-1 : DDL et SC des facteurs de l'ANOVA du groupe

	ddl	Somme des carrés	Statistique de Fisher
Produit (fixe)	I-1	$SCP = JK \sum_i (x_{i..} - x_{...})^2$	$F_{produit} = \frac{SCP / ddl_{produit}}{SCI / ddl_{interaction}}$
Sujet (aléatoire)	J-1	$SCS = IK \sum_j (x_{.j.} - x_{...})^2$	$F_{sujet} = \frac{SCS / ddl_{sujet}}{SCI / ddl_{interaction}}$
Sujet*Produit (aléatoire)	(I-1)(J-1)	$SCI = K \sum_i \sum_j (x_{ij.} - x_{i..} - x_{.j.} + x_{...})^2$	$F_{interaction} = \frac{SCI / ddl_{interaction}}{SCE / ddl_{résiduelle}}$
Erreur (aléatoire)	I J(K-1)	$SCE = \sum_i \sum_j \sum_k (x_{ijk} - x_{ij.})^2$	

Un effet sujet significatif indique que le niveau moyen de notation sur l'échelle n'est pas homogène entre les sujets. En pratique, cet effet est souvent significatif mais il n'apporte pas réellement d'information sur la performance des panélistes. L'effet produit indique si les moyennes des notes des produits sont différentes. Si l'effet est significatif, le panel sera déclaré discriminant pour ce descripteur. La statistique de Fisher de l'effet produit et la probabilité associée sont les indicateurs les plus utilisés en analyse sensorielle pour caractériser la discrimination au niveau unidimensionnel. Lorsque l'effet lié à l'interaction est significatif, cela implique que les sujets ne perçoivent pas les mêmes différences entre produits. L'interaction est souvent interprétée comme une mesure du désaccord entre les sujets, mais ce terme englobe en fait deux sources de variation de la notation : le désaccord au sens strict, c'est-à-dire le fait que tous les sujets ne donnent pas les mêmes écarts entre les produits, et les différences de dispersion des notes sur l'échelle (figure 1-1). Dans un cas extrême, il est possible que le terme d'interaction ne soit dû qu'à des différences de dispersion des notes entre les sujets, aussi appelé effet de "scaling", et non à du désaccord. Lors de l'utilisation du terme d'interaction comme indicateur de la mesure du désaccord il ne faut donc pas perdre de vue que ce terme reflète à la fois le désaccord au sens strict et l'effet de scaling. Enfin, dans le modèle présenté, le terme d'erreur peut s'interpréter comme la moyenne des écarts au carré entre les notes des répétitions de chaque produit. Une erreur résiduelle élevée indique que les écarts entre les répétitions sont grands, donc que le panel n'est pas répétable. Le terme d'erreur ne peut pas être testé comme les autres facteurs à l'aide d'un test de Fisher mais il donne, malgré tout, une information sur le niveau de répétabilité des sujets.

Le modèle d'ANOVA présenté permet donc d'obtenir une information sur la répétabilité (erreur résiduelle), la discrimination (effet produit) et en partie sur le désaccord entre les juges (effet de l'interaction) au niveau du panel.

3.1.1 Facteur sujet en effet fixe ou en effet aléatoire

La discussion sur le choix du facteur sujet en effet fixe ou en effet aléatoire a longtemps fait débat en analyse sensorielle (Lundahl et McDaniel, 1988). Dans le modèle présenté dans le paragraphe ci-dessus, le facteur sujet a été choisi comme facteur aléatoire parce que les sujets du panel ont été sélectionnés au hasard parmi une population de dégustateurs potentiels. Pour les partisans du facteur sujet comme effet fixe, l'hypothèse est que les sujets ont, certes, été choisis dans une vaste population de dégustateurs potentiels, mais qu'ils ont été sélectionnés en fonction de leurs aptitudes sensorielles. De plus, ils ont ensuite été entraînés à la dégustation. Le tirage des dégustateurs n'est donc pas aléatoire. Néanmoins, d'autres sujets auraient pu être sélectionnés pour la même étude sans pour autant en changer les objectifs, ce qui montre que les sujets ne constituent pas un facteur fixe de l'étude, au contraire du facteur produit.

Au niveau des concepts statistiques, le fait de choisir un effet fixe ou un effet aléatoire modifie l'espérance des carrés moyens (ou "carrés moyens attendus") des facteurs de l'ANOVA (tableau 1-2).

Tableau 1-2 : Espérance des carrés moyens selon la nature du facteur sujet

Facteur	Carrés moyens attendus	
	avec le facteur sujet en effet fixe	avec le facteur sujet en effet aléatoire
Sujet	$\frac{IK}{J-1} \sum_j (\beta_j - \beta_{..})^2 + \sigma_e^2$	$IK\sigma_{suj}^2 + K\sigma_{suj*prod}^2 + \sigma_e^2$
Produit	$\frac{JK}{I-1} \sum_i (\alpha_i - \alpha_{..})^2 + \sigma_e^2$	$\frac{JK}{I-1} \sum_i (\alpha_i - \alpha_{..})^2 + K\sigma_{suj*prod}^2 + \sigma_e^2$
Interaction	$\frac{K}{(I-1)(J-1)} \sum_i \sum_j (\gamma_{ij} - \gamma_{..})^2 + \sigma_e^2$	$K\sigma_{suj*prod}^2 + \sigma_e^2$
erreur	σ_e^2	σ_e^2

Sous l'hypothèse de nullité de l'effet produit dans le cas fixe, le carré moyen de l'effet produit devient égal au carré moyen de l'erreur (σ_e^2). La statistique de test de l'effet produit est donc :

$$F_{produit} = \frac{CM_{produit}}{CM_{erreur}} \quad (1-2)$$

Sous l'hypothèse nulle, cette statistique suit une loi de Fisher.

Lorsque l'effet sujet est aléatoire, l'interaction produit*sujet devient automatiquement une variable aléatoire. Puisque celle-ci dépend également du facteur produit, la variabilité due à l'interaction intervient dans l'expression du carré moyen de l'effet produit. Dans ce cas, sous l'hypothèse de nullité de l'effet produit, le carré moyen correspondant devient égal à celui de l'interaction et non de l'erreur. La statistique de test de l'effet produit est donc celle déjà présentée dans le tableau 1-1:

$$F_{produit} = \frac{CM_{produit}}{CM_{interaction}} \quad (1-3)$$

Ces considérations sur la nature de la variable sujet impliquent que les résultats obtenus dans le cas fixe ne s'appliquent qu'aux sujets et aux produits de l'étude (Næs et Langsrud, 1998). Il n'est donc pas possible de généraliser les conclusions à toute la population des dégustateurs potentiels avec ce modèle. Le but de la réalisation d'un profil descriptif étant le plus souvent d'obtenir une caractérisation des produits généralisable à toute la population, il est préférable de choisir le facteur sujet comme aléatoire dans la majorité des cas (Carlucci et Monteleone, 2001). Cependant, lorsque l'objectif est de caractériser uniquement les sujets de l'étude, notamment lors du contrôle de la performance pendant l'entraînement, il peut être plus judicieux de considérer l'effet sujet comme fixe. Dans ce cas précis, en poussant le raisonnement à l'extrême, l'effet produit pourrait même être considéré comme aléatoire puisque d'autres produits auraient pu être utilisés pour caractériser la performance des panélistes en question. Cependant, les produits ne sont en général pas tirés au sort mais judicieusement choisis en fonction des objectifs de l'étude, ce qui tend plutôt à conserver le facteur produit en effet fixe.

Le fait de considérer l'effet sujet comme aléatoire procure un autre avantage lorsque l'analyse des résultats montre que l'interaction et l'effet produit sont significatifs (ce cas de figure est relativement courant en analyse sensorielle). Si tous les facteurs sont fixes cela implique que les sujets ne sont pas en accord (interaction significative) mais que le panel fait malgré tout des différences communes entre les produits (effet produit significatif). Il est alors difficile de donner une interprétation globale qui tienne compte de ces deux résultats. En revanche, lorsque l'effet sujet est aléatoire, le carré moyen de l'effet produit est comparé au carré moyen de l'interaction, donc un effet produit significatif indique que les différences entre produits au niveau du groupe sont malgré tout plus importantes que le désaccord entre les sujets.

Cependant, il reste la situation réelle dans laquelle le carré moyen de l'interaction est plus petit que le carré moyen de l'erreur (l'interaction est donc non significative). Dans ce cas, il n'y a plus de raison de choisir le carré moyen de l'interaction plutôt que celui de l'erreur pour tester l'effet produit. En effet, puisque l'interaction n'est pas significative, le carré moyen de ce facteur conduit, comme le terme d'erreur, à une estimation de la variabilité résiduelle. Afin d'obtenir des résultats plus conservateurs, nous considérons donc qu'il serait préférable de choisir le facteur dont la variabilité est la plus élevée, à savoir le terme d'erreur. De plus, dans le type de modèle utilisé, les degrés de liberté associés au terme d'erreur sont plus élevés que ceux de l'interaction. L'estimation de l'erreur résiduelle utilisant le terme d'erreur est donc plus robuste que celle utilisant l'interaction.

3.1.2 Le terme d'interaction

Le terme d'interaction du modèle (1-1) peut se décomposer de la manière suivante (Couronne, 1997):

$$\begin{aligned}
 SCI &= K \sum_i \sum_j (x_{ij} - x_{i..} - x_{.j.} + x_{...})^2 \\
 &= K \sum_i \sum_j ((x_{ij} - x_{.j.}) - (x_{i..} - x_{...}))^2 \\
 &= K \sum_i \left(\sum_j d^2(\text{juge } j \text{ centré}, \text{juge moyen centré}) \right)
 \end{aligned}
 \tag{1-4}$$

L'expression $(x_{ij} - x_{.j})$ réalise le centrage des scores des produits pour chaque sujet et l'expression $(x_{i.} - x_{..})$ centre les scores moyens des produits au niveau du groupe. La somme des carrés de l'interaction correspond donc à la somme (sur les produits) des carrés des distances entre la note moyenne de chaque juge et la moyenne du groupe, après avoir centré les données par sujet. Cette interprétation est illustrée sur la figure 1-3.

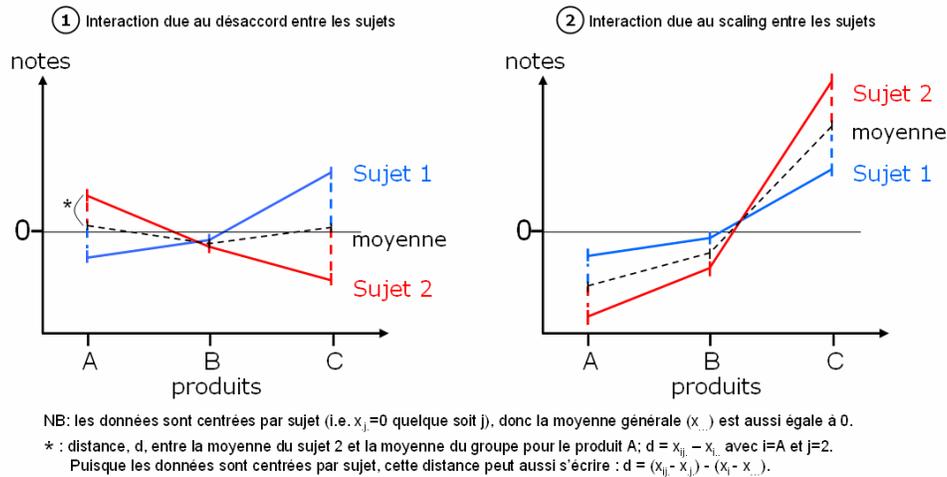


Figure 1-3 : Représentation graphique du terme d'interaction

Dans la situation 1 de la figure 1-3, l'interaction est élevée parce que les sujets n'ont pas classé les produits dans le même ordre. Le terme d'interaction reflète donc le désaccord entre les juges. En revanche, la situation 2 conduit à une interaction aussi élevée, mais elle est due à la différence d'utilisation de l'échelle (scaling) entre les deux sujets (le sujet 1 disperse moins ses notes que le sujet 2). Le classement des produits reste néanmoins le même pour les deux sujets.

Afin de tenir compte de l'effet de scaling, plusieurs stratégies ont été proposées dans la littérature. La plus intuitive consiste à réaliser un prétraitement des données en imposant que la variance des notes de chaque sujet soit égale à 1, en divisant les notes de chaque sujet par l'écart-type de l'ensemble des notes du sujet en question (Næs, 1990). Cette étape permet d'homogénéiser la variabilité des notes pour chaque sujet et de réduire ainsi les différences de scaling. Cette transformation peut conduire à une réduction importante du terme d'interaction et modifier les conclusions sur la nature des différences entre produits lors de son utilisation dans un cas concret (Carlucci et Monteleone, 2001). Un autre prétraitement consiste à déterminer les meilleurs coefficients multiplicateurs à appliquer aux notes de chaque sujet afin de minimiser la somme des carrés de l'interaction (Næs et Solheim, 1991). Cette technique vise donc à optimiser la dispersion des notes de chaque sujet afin de minimiser l'interaction. Une approche différente, proposée par Per Brockhoff en 1994, consiste à construire un modèle, baptisé « Assessor Model », qui intègre explicitement le scaling comme un effet à part entière (Brockhoff et Skovgaard, 1994; Brockhoff, 1998) :

$$Y = \alpha_j + \beta_j \cdot v_i + \varepsilon_{ij} \quad (1-5)$$

$$\varepsilon_{ij} \sim N(0, \sigma_j^2)$$

Où α_j correspond à l'effet sujet et v_i à l'effet produit qui est multiplié par le coefficient de scaling β_j du sujet j . De plus, ce modèle autorise l'hétérogénéité des variances des notes de chaque sujet, contrairement au modèle classique qui suppose l'homogénéité des variances des panélistes. Cette approche permet, entre autres, de tester l'existence des différences de dispersion des notes et de répétabilité entre les sujets. Elle permet également, par comparaison avec le modèle classique, de tester la présence de désaccord qui ne serait pas dû à un effet de scaling. Le chapitre 6 de ce manuscrit est entièrement dédié à la présentation et la comparaison de cette méthode avec le modèle d'ANOVA classique afin de mesurer les conséquences pratiques du choix de l'un ou l'autre des modèles sur les conclusions sensorielles avec des données réelles.

Certains auteurs décomposent la somme des carrés de l'interaction par sujet afin d'identifier les panélistes qui contribuent le plus à l'élévation de l'interaction et qui sont donc les plus éloignés du profil de notation de la moyenne du groupe (Schlich, 1994). Des tests de contrastes entre chaque sujet et le reste du panel peuvent aussi être utilisés afin d'isoler les sujets dont la notation des produits est différente de celle du groupe (Lundahl et Daniel, 1990). Les auteurs regardent ensuite si l'interaction du modèle devient non significative en enlevant les sujets responsables de la forte élévation de ce facteur. Cette stratégie nécessite d'enchaîner plusieurs modèles d'ANOVA mais elle permet d'obtenir un test pour chaque sujet sur son accord avec le groupe. Le but de cette démarche est de réduire la variabilité au sein du groupe et ainsi augmenter la significativité des différences entre produits. Cela implique cependant de retirer des panélistes du groupe *a posteriori*. Cette démarche est donc discutable car les résultats ainsi obtenus ne reflètent pas la perception du groupe mais seulement celle de la sélection des sujets qui maximisent les différences entre produits. L'expérimentateur s'autorise en quelques sortes à décréter que la perception de certains panélistes est fautive. Les conclusions formulées à la suite de ce type d'analyse doivent donc être prudentes.

Le modèle d'ANOVA présenté dans l'équation 1-1 décompose la variance en fonction des facteurs sujet, produit et de leur interaction. Ce modèle est particulièrement adapté lorsque les répétitions de chaque mesure ont été effectuées au cours d'une même séance. Cependant, la réalisation d'un profil sensoriel peut se réaliser dans des conditions différentes (répartition des séances de mesure sur plusieurs sessions, absence de répétition...), ce qui peut impliquer des modifications du modèle d'analyse de la variance.

3.2 Autres modèles d'ANOVA au niveau du groupe

3.2.1 L'effet session

Lorsque les dégustations des produits se déroulent sur plusieurs séances, l'effet session peut être ajouté dans le modèle (ainsi que les interactions qui en découlent) si celui-ci représente une source de variabilité que l'on souhaite identifier. C'est le cas par exemple lorsque des séances ont été réalisées avant et après une période de vacances. L'objectif est alors de savoir si le moment de la dégustation, et indirectement le temps de repos, a un impact sur la notation. Dans ce cas, le facteur

session est un effet fixe. Lorsque le facteur session correspond à une simple répétition de la mesure, celui-ci peut être considéré comme un effet aléatoire. Il peut alors sembler préférable de ne pas l'identifier comme tel et de laisser la variabilité due à cet effet dans l'erreur (en particulier si cet effet n'est pas significatif). Il est néanmoins possible de conserver cet effet en tant que tel dans le modèle afin d'obtenir une erreur résiduelle épurée de cet effet session (en particulier lorsque l'effet session est significatif). Dans tous les cas, lorsque l'expérimentation met en jeu plusieurs sessions, le plus important est que le plan de dégustation des produits entre chaque séance soit équilibré, au moins au niveau du groupe, afin que la comparaison des moyennes des notes entre les différents produits ne soit pas liée aux différences de niveaux de notation entre les sessions, et réciproquement.

3.2.2 Données sans répétitions

Lorsque que les données ne comportent pas de répétition, le modèle d'analyse de la variance au niveau du groupe est le suivant :

$$Y_{ij} = \alpha_i + B_j + \varepsilon_{ij}$$

$$B_j \sim N(0, \sigma_s^2) \quad (1-6)$$

$$\varepsilon_{ij} \sim N(0, \sigma_e^2) \quad \text{indépendants entre eux}$$

Tableau 1-3 : DDL et SC des facteurs de l'ANOVA du groupe (pas de répétition)

	ddl	Somme des carrés	Statistique de Fisher
Produit (fixe)	I-1	$SCP = J \sum_i (x_{i.} - x_{..})^2$	$F_{produit} = \frac{SCP / ddl_{produit}}{SCI / ddl_{erreur}}$
Sujet (aléatoire)	J-1	$SCS = I \sum_j (x_{.j} - x_{..})^2$	$F_{sujet} = \frac{SCS / ddl_{sujet}}{SCI / ddl_{erreur}}$
Erreur (aléatoire)	(I-1)(J-1)	$SCE = \sum_i \sum_j (x_{ij} - x_{i.} - x_{.j} + x_{..})^2$	

Au coefficient K près, les expressions des sommes des carrés des effets sujet et produit sont identiques à celles du modèle avec répétition et l'expression du terme d'erreur devient équivalente à celle de l'interaction (tableau 1-3). Le terme d'erreur donne alors une information sur le désaccord entre les panélistes et sur la variabilité résiduelle. La statistique de Fisher de l'effet produit est utilisée comme indice de discrimination. D'autres indices ont été proposés comme le coefficient d'accord r_{acc} (Finn, 1970; James, Demaree et al., 1984).

$$r_{acc} = 1 - \frac{\sigma_e^2}{\sigma_u^2} \quad (1-7)$$

Où σ_u^2 est la variance résiduelle obtenue dans le cas où tous les sujets auraient donné des notes aléatoires aux produits (selon une loi uniforme). Ce terme représente la variabilité théorique pour une distribution aléatoire des scores (hypothèse nulle). Le coefficient r_{acc} est interprété par les auteurs comme la proportion de variance de l'erreur qui n'est pas due au hasard, donc comme un indice du

degré de consensus entre les notations des différents panélistes. La significativité de ce coefficient peut ensuite être testée par un Chi² (Bi, 2003). Notons toutefois que la variance calculée pour une distribution uniforme ne correspond pas à la valeur maximale que peut atteindre l'erreur. Même si cela reste très improbable, celle-ci peut, en pratique, être plus grande et donc conduire à une valeur négative de ce coefficient.

Le coefficient de corrélation intra-classe (CCI), proposé à l'origine en génétique (Fisher, 1946), représente la part de variabilité due à l'effet produit par rapport à la variabilité totale :

$$CCI = \frac{\sigma_{prod}^2}{\sigma_{prod}^2 + \sigma_e^2} \quad (1-8)$$

Dans le cas présenté, l'effet produit est défini en facteur aléatoire mais le calcul et l'interprétation de cet indice ne sont pas modifiés si l'effet produit est fixe. En comparaison avec la statistique de Fisher de l'effet produit, ce coefficient de discrimination présente la particularité d'être borné entre 0 et 1, ce qui permet d'obtenir des valeurs plus faciles à interpréter lorsque l'erreur est particulièrement faible. Dans le cas extrême où la variance résiduelle est nulle, la statistique F de l'effet produit n'est pas calculable (à cause de la division par 0) alors que CCI=1.

3.3 Indices de performances au niveau individuel

Après l'examen des performances au niveau du groupe, l'analyse au niveau individuel permet d'identifier les différences de performances entre les panélistes. Cette étude peut se réaliser à l'aide de tests de contrastes estimés à partir du modèle d'ANOVA utilisé pour l'étude au niveau du groupe. Une autre solution consiste à définir un modèle d'ANOVA par sujet.

3.3.1 Tests de contrastes

Les tests de contrastes permettent, pour chaque facteur du modèle d'ANOVA, de spécifier les niveaux du facteur à comparer. Par exemple, cela permet de comparer uniquement le niveau de notation de deux juges sélectionnés parmi les sujets du panel. Latreille, Mauger et al. (2006) utilisent les contrastes pour définir plusieurs indices de performance individuelle :

- le pouvoir discriminant individuel est obtenu en calculant, pour chaque sujet, les tests de contraste pour chaque paire de produits (ce qui conduit au calcul de (I-1)/2 tests). Un résultat significatif indique que les deux produits ont été différenciés par le sujet considéré. Cette stratégie permet d'obtenir une information très fine sur la nature du pouvoir discriminant de chaque sujet mais elle présente l'inconvénient de devoir regarder les résultats de beaucoup de tests. L'interprétation globale des résultats peut s'avérer complexe lorsque le nombre de produits est important (Latreille, Mauger et al., 2006). Un test de contraste prenant en compte tous les produits simultanément pourrait palier ce problème. Ce test serait équivalent au test F de l'effet produit pour un sujet donné.
- La répétabilité individuelle est appréhendée grâce aux tests de contrastes de l'effet session par sujet. Une fois encore, les auteurs analysent les différences pour chaque paire de sessions afin d'obtenir l'information la plus fine possible. Le nombre de sessions étant en général relativement faible, le problème du nombre de tests à examiner est moins

crucial. Cependant, ce test suppose que tous les produits aient été dégustés dans chaque session. Dans le cas contraire, le test de l'effet session pourrait refléter des différences entre produits et non la répétabilité du sujet.

- Le désaccord de chaque panéliste avec le groupe se calcule grâce à un test de contraste basé sur le terme d'interaction sujet*produit : la moyenne de chaque produit pour un sujet donné est comparée à la moyenne de chaque produit pour le groupe. Ce test indique si le sujet considéré est responsable d'une proportion significative de l'interaction. Rappelons que le désaccord ainsi défini englobe également la notion de scaling.

3.3.2 Autres tests basés sur le modèle d'ANOVA au niveau du groupe

Rossi (2001), utilise des indices basés sur les variabilités individuelles issues du modèle d'ANOVA au niveau du groupe. Il applique au cas d'un panel des indices utilisés à l'origine pour le contrôle de la performance inter-laboratoires (Mandel, 1991). L'indice de répétabilité pour un sujet et un produit est défini comme le rapport entre la variabilité du sujet et la variabilité moyenne du groupe pour ce produit :

$$h_{ij} = \frac{\sigma_{ij}}{\sigma_i} \quad (1-9)$$

Ce coefficient est donc supérieur à 1 si le sujet j est moins répétable que le groupe pour le produit i. Le second indice mesure l'écart "standardisé" entre la notation d'un sujet et la notation du groupe pour chaque produit de la manière suivante :

$$K_{ij} = \frac{x_{ij} - x_{i..}}{\sigma_i} \quad (1-10)$$

Cet indice mesure l'écart entre le niveau de notation d'un sujet pour un produit et le niveau du groupe (numérateur), divisé par l'écart-type des niveaux des sujets (dénominateur). Un indice élevé, en valeur absolue, indique que le sujet note à un niveau vraiment différent de la moyenne du groupe. Cet indice se rapproche du calcul de la décomposition de la somme des carrés de l'effet juge pour chaque sujet (à un carré et un coefficient multiplicateur près).

3.3.3 ANOVA par sujet

L'analyse de la variance par sujet est une alternative simple à mettre en œuvre et qui permet de retrouver la plupart des résultats obtenus avec les tests de contrastes ; seuls les indices d'accord entre les sujets ne peuvent évidemment pas être calculés. Le modèle s'écrit de la manière suivante :

$$Y_{ik} = \alpha_i + \varepsilon_{ik} \quad (1-11)$$

$$\varepsilon_{ik} \sim N(0, \sigma_e^2)$$

Le test de l'effet produit, α_i , et la probabilité associée indiquent si le sujet est discriminant ou non. La variance résiduelle σ_e^2 renseigne sur la non répétabilité du sujet. Bien qu'une variance résiduelle très faible indique que le sujet est très répétable, cela peut néanmoins correspondre à une situation où le sujet n'est pas du tout discriminant (ex : toutes les notes égales à 0). Les deux indices doivent donc être étudiés simultanément en construisant par exemple le graphique de la probabilité associée à

l'effet produit en fonction de la variance résiduelle (Lea, Rødbotten et al., 1995). Les sujets ayant de faibles valeurs pour les deux indices sont considérés comme meilleurs que les autres ; ils sont à la fois plus discriminants et plus répétables. Toutefois, comme l'effet produit est égal au rapport de la variance produit sur la variance résiduelle, la statistique de Fisher de l'effet produit et la probabilité associée résumant déjà les deux notions. Si les effets produits sont identiques pour deux sujets, cela indique que le rapport de la variabilité inter-produits divisée par la variabilité intra-produit est le même pour les deux sujets. L'importance des différences entre produits est donc la même dans les deux cas, mais les notes peuvent être plus ou moins étalées selon les sujets (effet de scaling). La significativité de l'effet produit pourrait donc être conservée comme seul indice de performance, si l'information sur les différences de dispersion des notes selon les sujets est considérée comme secondaire.

3.3.4 Coefficients de corrélation

Puisque l'analyse de la variance par sujet ne permet pas de donner d'information sur l'accord du sujet avec le groupe, cette analyse est souvent accompagnée d'un test de corrélation entre les notes moyennes données par un sujet pour différents produits et celles du reste du groupe. Lorsque les notes sont données sur des échelles continues, le coefficient de Pearson est le plus utilisé (Labbe, Rytz et al., 2004). Lorsque les données sont discrètes, que la liaison entre les deux variables n'est pas linéaire, et/ou que l'on souhaite réaliser un test plus conservateur (les petites différences seront moins détectées), les coefficients de Spearman ou de Kendall sont utilisés (Schlich, 1997). Ces derniers n'utilisent pas directement les notes obtenues mais les données transformées en rang. En comparaison avec les indices qui utilisent l'interaction sujet*produit pour déterminer le désaccord entre les sujets, les coefficients de corrélation présentent l'avantage de ne pas être sensible aux différences de dispersion des notes sur l'échelle entre les sujets. Les seuils de significativité choisis pour ces corrélations sont en général moins conservateurs que les tests au niveau du groupe, dont le risque de première espèce est souvent fixé à 0,05. Ainsi, Schlich (1994) déclare qu'un individu est en accord avec le groupe lorsque le coefficient de Kendall est significatif au risque 0,20 ; Labbe et al. (2004) choisissent un seuil similaire pour un coefficient de corrélation de Pearson.

Une autre approche, connue sous le nom de « eggshell plot », qui utilise les corrélations sur les rangs a été développée (Næs, Hirst et al., 1994; Næs, 1998). Elle consiste à calculer le rang de chaque produit au niveau du groupe, puis à représenter les écarts entre le classement de chaque panéliste et le classement du groupe. La manière dont est représenté l'ordonnement du groupe et de chaque sujet est présentée sur la figure 1-4. Le nom de ce graphique (eggshell, dont la traduction signifie coquille d'œuf) renvoie à la forme des courbes de cette représentation. Lorsque la courbe d'un panéliste est très éloignée de la courbe du groupe, cela signifie que le sujet a classé les produits dans un ordre très différent de celui du groupe. Au contraire, si un sujet n'a fait que de petites inversions par rapport au classement du panel, sa courbe sera quasiment superposée à celle du groupe.

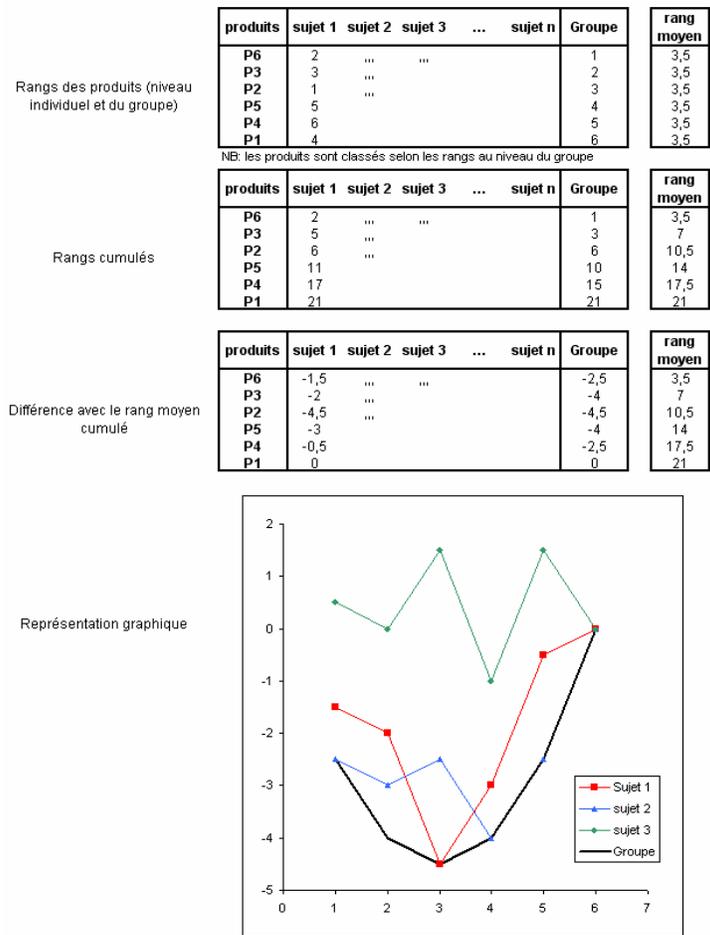


Figure 1-4 : Construction des graphiques « Eggshell plot »

Dans cet exemple, les sujets 1 et 2 ont classé les produits dans un ordre relativement proche de celui du groupe, contrairement au sujet 3. Les auteurs indiquent que ce genre de représentation est surtout adapté aux cas où le nombre de produits dégustés est élevé. De plus, le fait de travailler sur les rangs déforme les distances réelles entre les moyennes des produits. Afin de tenir compte de cet aspect dans la représentation, il est possible de préciser, sur l'axe des abscisses, quels sont les produits significativement différents les uns des autres d'après un test de comparaison multiple. Une autre solution consiste à remplacer les rangs en abscisse par les moyennes des produits, mais cela déforme alors les courbes par rapport au graphique initial (Lea, Rødbotten et al., 1995). En pratique, ce type de graphique est donc seulement adapté aux cas où la majorité des panélistes perçoivent des différences significatives entre les produits. Cette méthode est donc rarement applicable à tous les descripteurs d'une étude.

3.3.5 Coefficient de variation

Le coefficient de variation correspond à une variabilité (le plus souvent l'erreur résiduelle) divisée par la moyenne des valeurs correspondantes :

$$CV = \frac{\sigma_e}{\mu} \quad (1-12)$$

Ce rapport à la moyenne permet de tenir compte du fait que la précision d'une mesure est en général fonction du niveau à mesurer. Notamment, le coefficient de variation reste constant lorsque la variance est proportionnelle au carré de la moyenne. Ce coefficient est très souvent utilisé pour des variables non bornées à droite (une taille, un poids...) et peut servir d'indice de répétabilité pour des données physicochimiques (Karpinski, 1989). En analyse sensorielle, le coefficient de variation pourrait également être utilisé à la place de la variance résiduelle comme indice de répétabilité, afin de tenir compte du fait que plus la moyenne obtenue pour un produit est proche de 0, plus il est facile d'obtenir une faible variabilité entre les répétitions. Cependant, le même phénomène se produit à l'autre extrémité de l'échelle alors que la moyenne est très élevée. Le CV n'est donc pas adapté pour la partie supérieure de l'échelle. Un indice qui reproduirait le comportement du CV de manière symétrique aux deux extrémités de l'échelle serait donc préférable. En voici un exemple (Chabanet, communication personnelle):

$$CV_s = \frac{\sigma_e}{\sqrt{\mu(\text{rangeMax} - \mu)}} \quad (1-13)$$

Où le rangeMax correspond à l'étendue de l'échelle de notation.

De cette manière la variance résiduelle correspondant à une moyenne μ de 7 ou de 3 sur une échelle de 0 à 10 sera divisée par la même valeur.

3.4 Présentation synthétique des résultats unidimensionnels

Les analyses unidimensionnelles permettent d'obtenir un diagnostic détaillé de la performance du groupe et de chaque sujet pour chaque descripteur. En contrepartie, ces méthodes ont l'inconvénient majeur de générer un grand nombre de résultats : une analyse est menée par descripteur puis chaque sujet est en général étudié individuellement. Dans chaque cas, plusieurs indices de performance peuvent être calculés. La présentation synthétique des résultats est donc un aspect non négligeable de l'analyse afin que l'animateur de panel puisse extraire rapidement les informations essentielles. Latreille et al. (2006) réalise des tableaux de synthèse (de dimensions descripteur*sujet) pour chaque indice de performance. Il faut alors combiner les résultats de plusieurs tableaux pour avoir une vue globale de la performance d'un panéliste. Labbe et al. (2004), propose un tableau dans lequel sont présentés les indices d'accord choisis (coefficient de corrélation de Pearson) pour chaque sujet et chaque descripteur, seulement si le sujet est discriminant pour ce descripteur ($F_{\text{produit}} > 2,2$). Le coefficient est mis en gras lorsque l'accord avec le groupe est significatif.

La méthode GRAPES (Schlich, 1994) permet quant à elle de calculer six indices de performance. Elle propose ensuite une représentation synthétique des résultats en trois graphiques bidimensionnels par descripteur. Afin de limiter le nombre de figures à observer, le même auteur a ensuite proposé la méthode CAP (Schlich, 1997) dont les résultats sont synthétisés en un seul tableau pour tous les descripteurs. Ces méthodes sont décrites dans les paragraphes suivants.

3.4.1 Méthode GRAPES

La méthode GRAPES propose, pour chaque descripteur, six indices de performance pour chaque sujet et six indices équivalents au niveau du groupe dans son ensemble (Schlich, 1994). Pour chaque descripteur, l'auteur propose de réaliser le modèle (1-14) pour chaque sujet :

$$Y_{ik} = \alpha_i + session_k + \varepsilon_{ik} \quad (1-14)$$

$$\varepsilon_{ik} \sim N(0, \sigma_c^2)$$

Et le modèle (1-15) pour le panel tout entier :

$$Y_{ijk} = \alpha_i + session_k(\alpha_i) + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \quad (1-15)$$

$$\varepsilon_{ijk} \sim N(0, \sigma_c^2)$$

Le paramètre α_i correspond à l'effet produit, $session_k$ à l'effet session, $session_k(\alpha_i)$ à l'effet session hiérarchisé dans l'effet produit, β_j à l'effet sujet et γ_{ij} à l'interaction. Ces modèles se rapprochent des modèles (1-11) et (1-1) définis précédemment, à ceci près qu'ils prennent en compte explicitement l'effet session et que tous les facteurs sont déclarés en effet fixe. Les sommes de carrés (SC) et les carrés moyens (CM) des effets de ces modèles sont utilisés pour construire les indices de performance au niveau du sujet (tableau 1-4) et au niveau du groupe (tableau 1-5).

Tableau 1-4 : Définition des 6 indices de performances individuels de GRAPES
(d'après Schlich, 1994)

Statistique	Formule	Interprétation
LOCATION _j	$x_{.j}$	Note moyenne sur l'ensemble des produits
SPAN _j	$\frac{1}{K} \sum_k \sqrt{\sum_i \frac{(x_{ijk} - x_{.jk})^2}{I-1}}$	Moyenne des écarts entre produits pour chaque séance = dispersion des notes sur l'échelle
UNRELIABILITY _j	$\sqrt{CME_j} / SPAN_j$	Erreur pure
DRIFT-MOOD _j	$\sqrt{CMS_j} / SPAN_j$	Déviations des scores d'une séance à l'autre
DISCRIMINATION _j	CMP_j / CME_j	Statistique de Fisher de l'effet produit
DISAGREEMENT _j	$J \cdot SCInt_j / (I-1)(J-1)CME$	Contribution du juge j au terme d'interaction

SCX : somme des carrés du facteur X et ddlX, les degrés de liberté associés

CMX : Carré moyen du facteur X = SCX/ddlX

S : séance ; P : produit ; Int : interaction sujet*produit ; E : erreur.

LOCATION_j représente le niveau moyen de notation du sujet j, tous produits confondus. SPAN_j mesure, en moyenne sur les séances, la dispersion des notes des produits sur l'échelle. Ces deux indices donnent des informations sur la manière dont chaque sujet utilise l'échelle. UNRELIABILITY_j est le rapport entre la racine carrée du carré moyen de l'erreur associée au sujet j et la dispersion de ses notes (SPAN_j). Une valeur élevée pour ce coefficient indique que les erreurs de mesures commises par le sujet sont importantes par rapport aux différences que celui-ci est capable de faire

entre les produits au sein d'une séance. $DRIFT-MOOD_j$ (qui pourrait être traduit par « dérive selon l'humeur ») mesure la déviation du niveau moyen de notation entre les séances par rapport à la dispersion des notes. $UNRELIABILITY_j$ représente la variabilité intra-séance alors que $DRIFT-MOOD_j$ mesure la variabilité inter-séances. $DISCRIMINATION_j$ est la statistique de Fisher classique de l'effet produit (dans le modèle par sujet). $DISAGREEMENT_j$ mesure la contribution du sujet j à l'interaction sujet*produit du modèle (1-15). Une valeur élevée pour ce coefficient indique que la part d'interaction dont est responsable le sujet j est importante, donc que le désaccord de ce sujet avec le panel est fort. $DISCRIMINATION_j$ et $DISAGREEMENT_j$ donnent des informations sur l'importance et l'homogénéité des différences entre produits.

Chaque couple d'indices donne ensuite lieu à une représentation bidimensionnelle pour chaque descripteur où les observations sont les sujets (figures 1-5, 1-6 et 1-7).

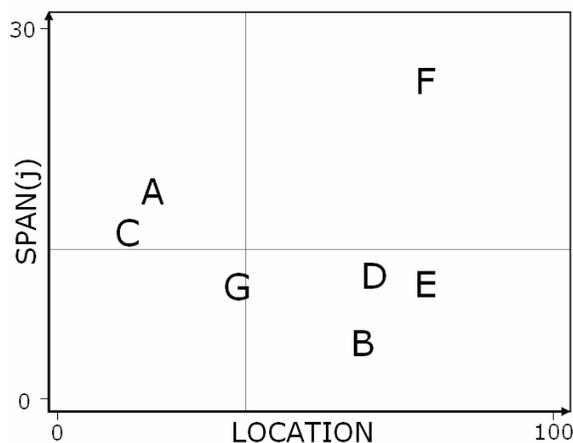


Figure 1-5 : Graphique des indices de GRAPES concernant l'utilisation de l'échelle

La figure 1-5 permet de mettre en évidence des sujets (représentés par des lettres) dont la dispersion des notes et/ou le niveau moyen de notation est particulièrement différent des autres panélistes. Sur cet exemple, les sujets A, C et G utilisent plutôt la partie gauche de l'échelle, au contraire des autres panélistes. De plus, le sujet F disperse davantage ses notes que les autres ($SPAN_j$ élevé).

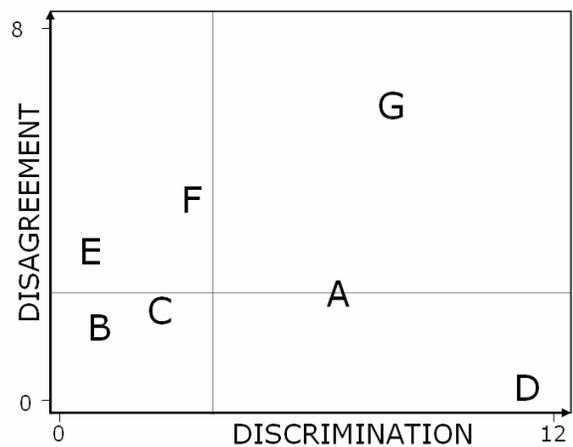


Figure 1-6 : Graphique des indices de GRAPES concernant les différences entre produit

Le graphique de la figure 1-6 représente les panélistes en fonction de leur niveau de discrimination et de leur désaccord avec le groupe. Le sujet D pourrait être qualifié de meilleur dégustateur puisqu'il combine un niveau de discrimination élevé et un faible désaccord. Le sujet G au contraire, présente un désaccord avec le groupe relativement prononcé, même si son pouvoir discriminant est élevé.

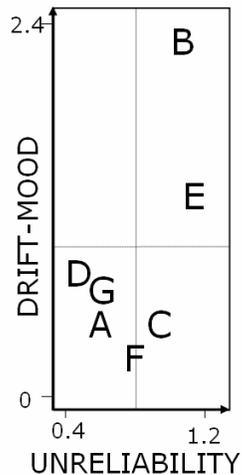


Figure 1-7 : Graphique des indices de GRAPES concernant la dispersion des répétitions

La figure 1-7 présente des résultats relatifs à la dispersion des notes à l'intérieur de chaque session (UNRELIABILITY) et entre les sessions (DRIFT-MOOD). Cet exemple met en évidence que le sujet E est moins répétable que les autres et que le niveau de notation du sujet B d'une séance à l'autre varie beaucoup plus que pour les autres sujets (par rapport à la dispersion de ses notes).

L'analyse au niveau du groupe conduit également au calcul de 6 indices de performance par descripteur (tableau 1-5).

Tableau 1-5 : Définition des 6 indices de performances au niveau du groupe de GRAPES (d'après Schlich, 1994)

Statistique	Formule	Interprétation
LOCATION	$x_{...}$	Moyenne générale
SPAN	$\sum_j SPAN_j / J$	Moyenne des SPAN individuels
UNRELIABILITY	$\sqrt{CME} / SPAN$	Erreur pure
DRIFT-MOOD	$\sqrt{CMS} / SPAN$	Déviations des scores d'une séance à l'autre
DISCRIMINATION	CMP / CME	Statistique de Fisher de l'effet produit
DISAGREEMENT	$CMInt / CME$	Statistique de Fisher de l'interaction produit*sujet

SCX : somme des carrés du facteur X et ddlX, les degrés de liberté associés

CMX : Carré moyen du facteur X = SCX/ddlX

S : séance ; P : produit ; Int : interaction sujet*produit ; E : erreur.

Comme au niveau individuel, trois graphiques regroupant à chaque fois deux indices sont produits. Les observations sont cette fois les descripteurs de l'étude. Sans correspondre explicitement à une approche multidimensionnelle, cette méthode permet néanmoins de comparer les performances obtenues pour chaque descripteur grâce à ces graphiques (non représentés).

La méthode GRAPES est complétée par un tableau qui récapitule pour chaque sujet et chaque indice, le nombre de descripteur pour lesquels la performance est significativement différente de celle du groupe. La performance d'un sujet est considérée comme différente de celle du groupe si elle se situe à l'extérieur de la région définie par la boîte à moustache de la figure 1-8.

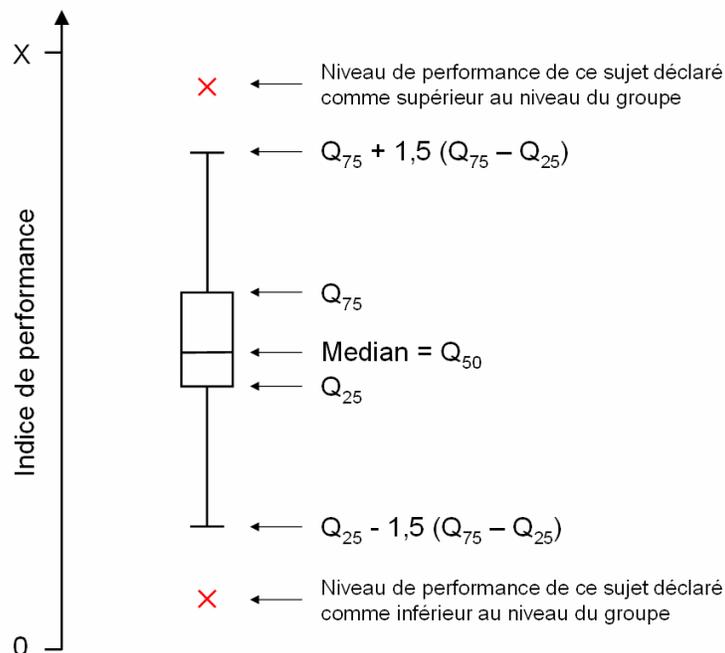


Figure 1-8 : Définition de la boîte à moustache construite pour chaque indice de performance de GRAPES

La méthode GRAPES donne une représentation précise et complète de la notion de performance pour chaque sujet puisqu'elle repose sur le calcul de six indices. Les représentations graphiques permettent ensuite de comparer les sujets entre eux et de synthétiser les résultats pour chaque descripteur en seulement trois figures (1-5, 1-6, 1-7). Néanmoins, ce nombre de figures est multiplié par le nombre de descripteur de l'étude, ce qui peut conduire dans certains cas à regarder un nombre de graphiques assez important. A ces graphiques illustrant la performance individuelle, s'ajoutent les trois graphiques concernant la performance du groupe qui permettent de comparer les niveaux de chaque descripteur.

3.4.2 Méthode CAP

La méthode CAP (Control of Assessor Performances) a été développée par Schlich (1997) quelques années après la méthode GRAPES afin de proposer des résultats sur la performance au niveau unidimensionnel sous une forme encore plus synthétique. Certains indices de cette méthode ont été repris de GRAPES et parfois modifiés, d'autres indices proviennent d'approches différentes.

Les résultats sont présentés sous la forme d'une seule table (tableau 1-6) dans laquelle chaque ligne correspond à un descripteur. Ceux-ci sont triés du plus discriminant au moins discriminant selon la valeur de la statistique de Fisher de l'effet produit. Les premières colonnes contiennent les indices de performance au niveau du groupe ; les colonnes suivantes correspondent chacune à un sujet. Les panélistes sont également classés du plus discriminant au moins discriminant en fonction de leur statistique de Fisher du test de MANOVA de l'effet produit.

Tableau 1-6 : Exemple de table CAP (d'après Schlich, 2000)

attribut	Fprod	Finter	Mean	RMSE	1	12	8	6	5	3	11	14	7	4	10	2	9	13
SWEET	9,49	1,1	42,4	21,9	+	+	+!	+	+	+	+	+	=	+	+!	=	=	=
CHESNUT	6,02	1,1	41,1	24,3	+	=!	+	-	+	+	+	=	+	=	+!	=!	+	=
PASTRY	4,42	1,51	19,2	20,1	+	+	-	-	+	+!	-	+	+	=	+	=	=	+!
HERBACE	3,23	1,55	30,5	21,2	-	=!	=	+	-	=!	+	+	=!	=	+	=	=	=!
BITTER	2,87	1,28	26,9	20,2	+	+	+	-	=	=	=	+	=!	+	=	+	=!	=
RAWPOTAT	2,81	1,09	22,7	21,9	+	=	=!	=	+	=	+	=	=!	=	+	+	=!	=
CEREAL	2,32	1,05	26,3	22,5	+	+	-!	=	=!	-	=	=!	=!	-	=	=	=	=
CELERY	1,86	1,38	16,9	15,9	=	+!	+!	=	=!	+	-	=	-!	=	=	-	=!	=
BUTTER	1,73	1,52	16,8	17,3	=!	+!	+!	=	-	=	+	=	+	=	=!	=	=	=
NUTS	1,63	1,65	16,3	16,5	+	+!	-!	-	=!	+	+	=	+!	=	=	=	=	=
CARDBOAR	1,45	1,3	8,07	13,8	-	-	=!	-	+	=	=	=	=!	-	=	=!	=	-
MUSHROOM	1,38	1,29	12	14,9	=	-	-!	-	=!	-	=	=	=!	=	=!	=	=	=
Fmanova					18,2	12,8	10,6	9,91	9,81	7,18	6,84	6,78	4,28	4,22	3,96	2,34	2,12	1,01

Fprod : valeur de la statistique de Fisher de l'effet produit dans le modèle produit + sujet + produit*sujet

Finter : valeur de la statistique de Fisher de l'interaction

Mean : moyenne générale tous produits confondus

RMSE : écart-type moyen entre les répétitions

Fmanova : valeur de la statistique de Fisher du test de l'effet produit dans les MANOVA individuelles

+, -, =! : diagnostic individuel qui dépend du pouvoir discriminant (discriminant : + ou -, non discriminant : =), de l'accord (en accord : +, désaccord : -) et de la répétabilité (non répétable : !)

Dans cette table, chaque sujet reçoit un diagnostic pour chaque descripteur qui permet d'indiquer s'il est discriminant, répétable et/ou en accord avec le reste du groupe. L'affectation du diagnostic est réalisée à l'aide de l'arbre de décision présenté sur la figure 1-9.

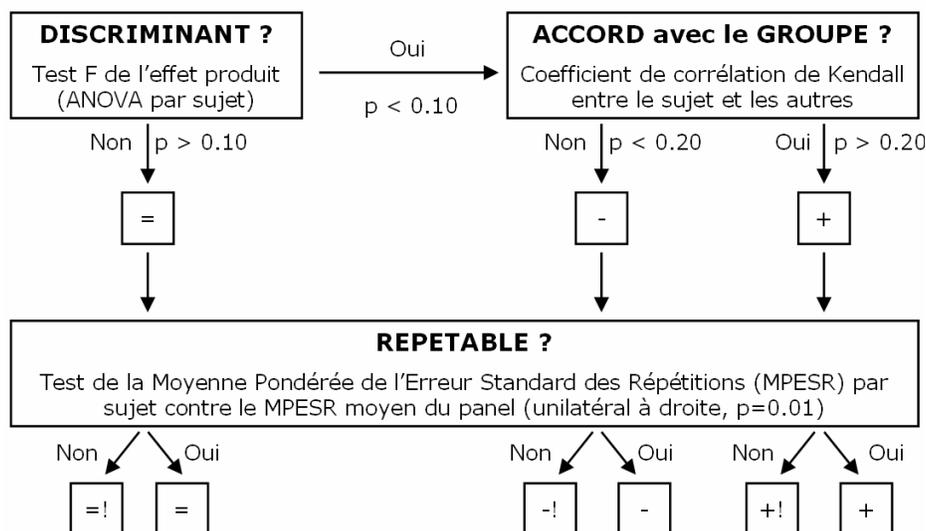


Figure 1-9 : Détermination du diagnostic CAP

Si le sujet n'est pas discriminant d'après la statistique de Fisher du test de l'effet produit dans le modèle individuel d'analyse de la variance : score = produit + erreur (probabilité associée supérieure à 10%), le sujet reçoit le diagnostic '='. Sinon, selon son accord avec le groupe, il reçoit le diagnostic '+'

(accord) ou '-' (désaccord). L'indice d'accord est le coefficient de corrélation de Kendall calculé entre les moyennes des notes des produits du sujet considéré et les notes moyennes du groupe constitué des autres panélistes discriminants. Le diagnostic des sujets dont le niveau de répétabilité, calculé à l'aide d'un test de contraste, est significativement plus faible que la moyenne du groupe (restreint aux sujets discriminants) est accompagné d'un point d'exclamation (+!, -! ou =!). Cette méthode sera décrite et discutée plus en détails dans le chapitre suivant.

3.5 Bilan des méthodes unidimensionnelles

Il existe de nombreuses manières d'aborder la performance lors de l'analyse de données de profil sensoriel. Le choix de la méthode dépend de la spécificité des données (présence ou non de répétitions des mesures, prise en compte de l'effet séance...) et de la question à laquelle l'expérimentateur souhaite répondre (détection des panélistes les moins performants, estimation du niveau moyen du panel...). Mais dans la majorité des cas, la méthode utilisée repose sur un modèle d'analyse de la variance (pour chaque descripteur) qui permet de décomposer la variabilité des mesures obtenues selon les différents facteurs de l'étude (sujet, produit...). Les indices de performances correspondent finalement à des estimations et/ou à des comparaisons de ces sources de variabilité afin de quantifier l'importance de chaque facteur, au niveau du panel dans son ensemble ou de chaque sujet. Comme la notion de performance recouvre différents aspects du style de notation des panélistes, la plupart des auteurs utilisent et recommandent l'utilisation de plusieurs indicateurs de la performance simultanément. Cela permet d'obtenir une image plus complète du niveau de chaque sujet et du panel (Lea, Rødbotten et al., 1995). Les analyses unidimensionnelles présentent l'avantage de donner des résultats détaillés pour chaque descripteur de l'étude. Cette information est indispensable pour l'animateur de panel qui souhaite vérifier que tous les termes employés ont été correctement utilisés. En contrepartie, ce type d'analyse présente l'inconvénient de produire beaucoup d'indices de performance et ne facilite pas la synthèse des résultats. De plus, l'analyse unidimensionnelle ne tient pas compte des éventuelles corrélations entre variables, ce que permet l'analyse multidimensionnelle.

4 Analyses multidimensionnelles

Les méthodes multidimensionnelles permettent d'analyser plusieurs variables simultanément et de tenir compte de leurs corrélations. Les données peuvent se présenter sous la forme d'un cube de données $X_{(I \times J \times P)}$ contenant les P descripteurs en colonne, les I produits en lignes et les J sujets dans la profondeur. Lorsque K répétitions sont présentes, elles peuvent former une quatrième dimension ou être associées à l'une des dimensions déjà existantes (par exemple, l'ensemble de données $X_{((I \times K) \times J \times P)}$ contient autant de lignes que de couples produit*répétitions). La matrice notée X_j , de dimension produit*descripteur ou (produit*répétition)*descripteur, contient l'ensemble des notes du juge j. Plusieurs familles de méthodes, très utilisées en analyse sensorielle, sont décrites dans les paragraphes suivants. Pour la plupart, l'objectif est de caractériser les différences entre produits en apportant également une information sur l'accord entre les juges et les corrélations entre descripteurs.

Dans la suite de ce manuscrit, le terme "individu" correspond aux lignes d'une matrice (et non aux sujets de l'étude) et les "variables" sont les colonnes de la matrice.

4.1 L'analyse en composante principale

L'ACP, acronyme d'analyse en composante principale (Joliffe, 1986), est la méthode la plus utilisée en analyse sensorielle. Sa vocation n'est pas de décrire la performance d'un panel mais sa présentation permet d'introduire les concepts généraux de l'analyse multidimensionnelle. De plus, elle fait généralement office de technique de référence dans les comparaisons de méthodes multidimensionnelles. L'ACP permet d'examiner les relations entre des variables corrélées. L'objectif est de résumer la variabilité entre les individus en un minimum de dimensions non corrélées (composantes principales) avec une perte d'information minimale. Cette méthode consiste à calculer, à partir du tableau $X_{(I \times P)}$ contenant les données de la moyenne du groupe pour chaque produit et chaque descripteur, la matrice des covariances entre variables : $V=X'X$. La recherche des valeurs propres et des vecteurs propres de V permet de déterminer les composantes principales, combinaisons linéaires des variables de X qui maximisent la variance entre les individus. Les résultats de l'ACP peuvent se présenter sous la forme de deux graphiques bidimensionnels : le plan factoriel de la projection des variables (figure 1-10) et le plan factoriel représentant les coordonnées des individus dans l'espace (figure 1-11). Les composantes principales suivantes (axes 3, 4...) peuvent être utilisées de la même manière pour former des graphiques similaires.

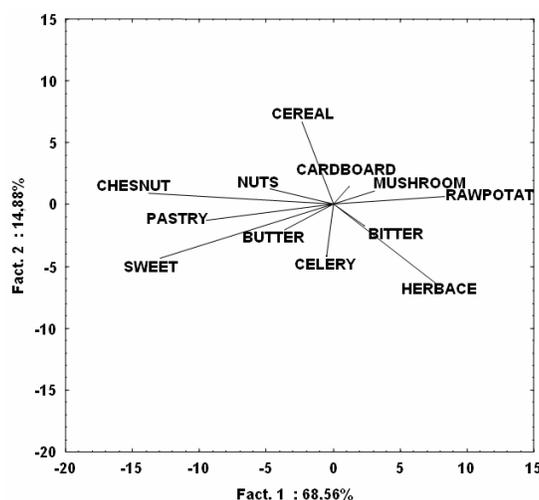


Figure 1-10 : Projection des variables sur le plan factoriel 1-2 (ACP)

Données : 12 descripteurs, 10 produits (moyenne des observations sur un panel de 16 sujets)

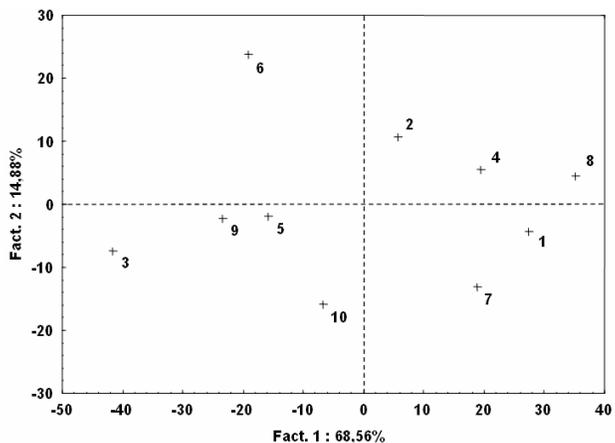


Figure 1-11 : Projection des individus sur le plan factoriel 1-2 (ACP)

Données : 12 descripteurs, 10 produits (moyenne des observations sur un panel de 16 sujets)

La figure 1-10 permet par exemple de voir que le premier axe oppose en particulier les descripteurs chesnut, sweet et pastry (à gauche) aux descripteurs rawpotat et herbage (à droite). Sur la carte des produits (figure 1-11), la configuration obtenue montre, entre autres choses, une opposition entre les produits 3 et 8.

Ces deux représentations peuvent également être réunies sur le même graphique pour former un « biplot » (Gabriel, 1971). Celui-ci permet d'expliquer directement la position des individus dans l'espace en fonction des variables : la projection orthogonale d'un individu sur une variable donne une indication sur l'intensité de la valeur obtenue par l'individu sur cette variable par rapport aux autres individus. La figure 1-12 montre par exemple, que le produit 3 obtient la coordonnée la plus élevée pour le descripteur SWEET. Cela laisse penser qu'il a obtenu une note très forte pour ce descripteur par rapport aux autres produits. A l'inverse, la note du produit 8 pour le descripteur SWEET devait être très faible.

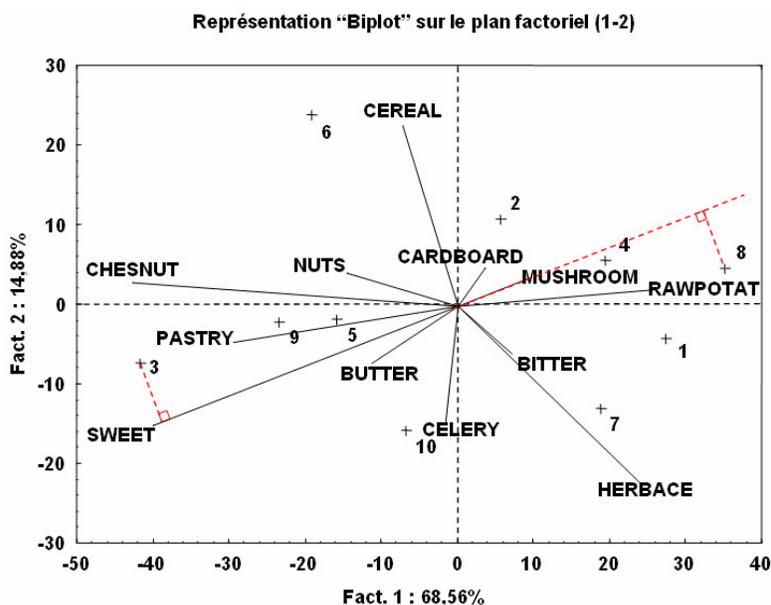


Figure 1-12 : Représentation « biplot » sur le plan factoriel 1-2 (ACP)

Il est parfois utile de standardiser les variables de l'analyse (donner une variance équivalente à toutes les variables) afin de ne pas tenir compte des différences entre les échelles de mesure. Par exemple, une ACP effectuée sur des variables physico-chimiques, qui ne sont pas toutes exprimées dans la même unité (longueur en mètre, concentration en g.mol⁻¹...), doit être réalisée sur les variables normalisées. Il s'agit alors d'une ACP dite « normée ». En revanche, en analyse sensorielle, les ACP « non normées » sont plus souvent utilisées car toutes les observations proviennent d'échelles de notation de même amplitude. Les variables pour lesquelles la dispersion des notes est plus faible auront alors un impact moins important dans la détermination des composantes principales.

Dans la majorité des cas, l'ACP est utilisée sur le tableau qui présente les produits en ligne (individus) et les descripteurs sensoriels en colonne (variables) ; chaque observation correspond à la moyenne des notes du groupe pour un produit et un descripteur. Cette analyse permet de mettre en évidence les différences entre produits et les corrélations entre descripteurs au niveau du groupe. Cependant, puisque cette représentation utilise seulement les notes moyennes des produits données par le groupe, elle ne fournit aucune indication sur la variabilité des notes entre les juges. Pour ajouter cette information, une possibilité consiste à projeter les évaluations de chaque sujet pour chaque produit (en tant qu'individus supplémentaires) sur la carte décrite précédemment (Monrozier et Danzart, 2001). Si tous les sujets ont évalué les produits d'une manière semblable pour tous les descripteurs, les observations individuelles seront proches du point moyen de chaque produit. Pour faciliter la lecture de la variabilité due aux sujets sur la carte d'ACP, il existe plusieurs possibilités de représentations graphiques (un marqueur différent par produit, connexion de chaque observation individuelle avec le point moyen...). D'après Monrozier et Danzart, la représentation la plus lisible consiste à tracer une ellipse de confiance de la moyenne de chaque produit. Ces ellipses sont définies par le critère de significativité suivant (Saporta, 1990), sous l'hypothèse de multi-normalité de la distribution :

$$\frac{n-p}{p}(x-\mu)'V^{-1}(x-\mu) = F_{(p;n-p)} \quad (1-16)$$

n : nombre d'observations par produit (i.e. nombre de sujets, dans ce cas)

p : nombre de dimensions de la carte factorielle (en général, 2 ou 3)

μ : moyenne des n scores individuels

V : matrice de covariance des scores individuels

F : valeur de la loi de Fisher à p et n-p degrés de liberté

En terme de caractérisation de la performance des panélistes, les ellipses procurent l'avantage de visualiser l'homogénéité des réponses du panel : l'ellipse autour d'un produit est d'autant plus petite que les réponses des sujets sont similaires. De plus, deux ellipses qui se chevauchent indiquent que les produits concernés ne sont pas perçus de manière significativement différente par le groupe. Les ellipses donnent donc à la fois une information sur l'accord du groupe (taille des ellipses) et sur la discrimination entre les produits (séparation des ellipses). Une alternative consiste à représenter la variabilité du point moyen en projetant des points moyens supplémentaires obtenus par ré-échantillonnage (Husson, Le et al., 2005). Cette technique consiste à créer un grand nombre de

nouveaux panels fictifs constitués de J juges tirés au sort (avec remise) parmi les J juges du panel initial. Le même juge peut donc intervenir plusieurs fois dans un panel « ré-échantillonné ». Les moyennes des produits sont ensuite calculées pour chaque panel ré-échantillonné et projetées en individus supplémentaires sur la carte. Des ellipses de confiance peuvent alors être dessinées pour chaque produit en utilisant la formule décrite précédemment (équation 1-16). Notons toutefois que les ellipses de ces deux représentations permettent d'ajouter une information sur la variabilité des notes des sujets, mais cette variabilité n'intervient pas dans la détermination des composantes principales de l'analyse.

Afin d'étudier la performance du panel, l'ACP peut être utilisée sur des tableaux de données présentés différemment. L'ACP par descripteur du tableau des sujets (en lignes) et des produits (en colonne) permet de détecter les juges qui ne perçoivent pas les différences entre produits de la même manière que les autres panélistes. L'ACP par juge de la matrice produit*descripteur donne une image des corrélations entre descripteurs et des différences de perception entre les produits.

4.1.1 ACP par descripteur

Afin d'obtenir une représentation des différences de notations des produits entre les sujets, il est possible d'effectuer l'ACP, pour chaque descripteur, du tableau dans lequel les produits sont les variables et les sujets sont les individus (Sinesio, Moneta et al., 1991/2; Dijksterhuis, 1995; Couronne, 1997). Les observations correspondent, pour chaque sujet et chaque produit, aux moyennes des notes sur les répétitions. Les données sont ensuite centrées en ligne (et en colonne, mais ce deuxième centrage est systématique avec l'ACP) de manière à supprimer les différences de niveaux de notations entre les sujets et entre les produits. Le tableau de données contient alors seulement les écarts de notation de chaque sujet pour chaque produit par rapport au groupe. En comparaison avec l'ANOVA, cette matrice contient les différences dues à l'interaction entre les facteurs sujet et produit. La représentation graphique (figure 1-13) permet alors de mettre en évidence les sujets qui contribuent le plus à l'élévation de l'interaction, c'est-à-dire les sujets dont le style de notation (ordonnancement des produits et dispersion des notes sur l'échelle) est le plus éloigné des autres panélistes. Bien que l'ACP soit une méthode multidimensionnelle, elle donne ici des résultats pour un seul descripteur sensoriel à la fois.

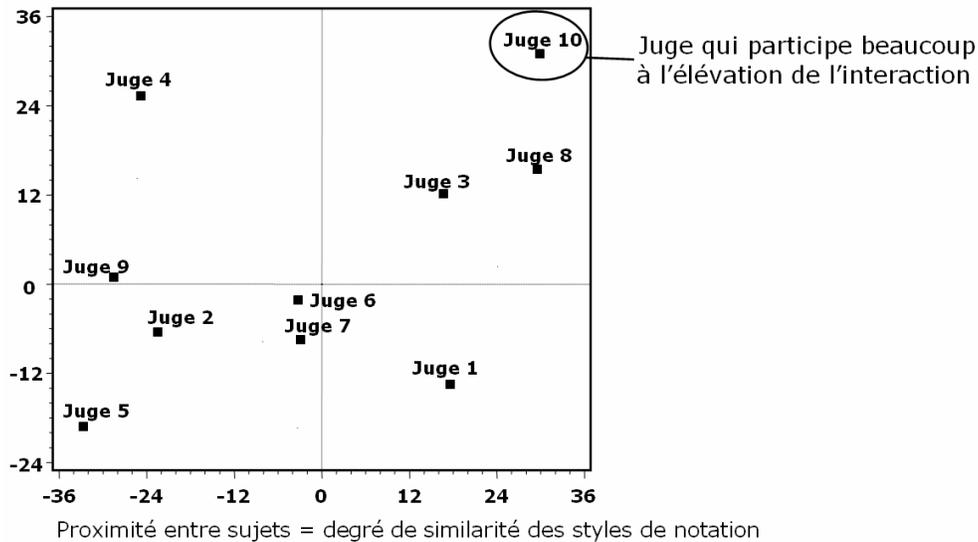


Figure 1-13 : ACP (pour un descripteur) des données centrées par juge

Certains auteurs proposent également de réaliser l'ACP par descripteur du tableau contenant les sujets en variables et les produits en individus (Dijksterhuis, 1995; Couronne, 1997; Couronne, 2001). Dans ce cas, les données sont seulement centrées par colonne (i.e. par sujet). Sur le graphique correspondant (figure 1-14), la longueur du vecteur d'une variable représente le pouvoir discriminant ; plus le vecteur est long, plus le sujet est discriminant. Le cosinus de l'angle entre deux vecteurs traduit la corrélation entre les notes des produits des deux panélistes ; plus les vecteurs pointent dans la même direction, plus les sujets ont noté les produits de la même façon. Les juges 3, 8 et 10 sont donc en accord. Ils sont en revanche en désaccord avec les sujets 2, 4, 5 et 9 qui pointent dans la direction opposée sur le premier axe.

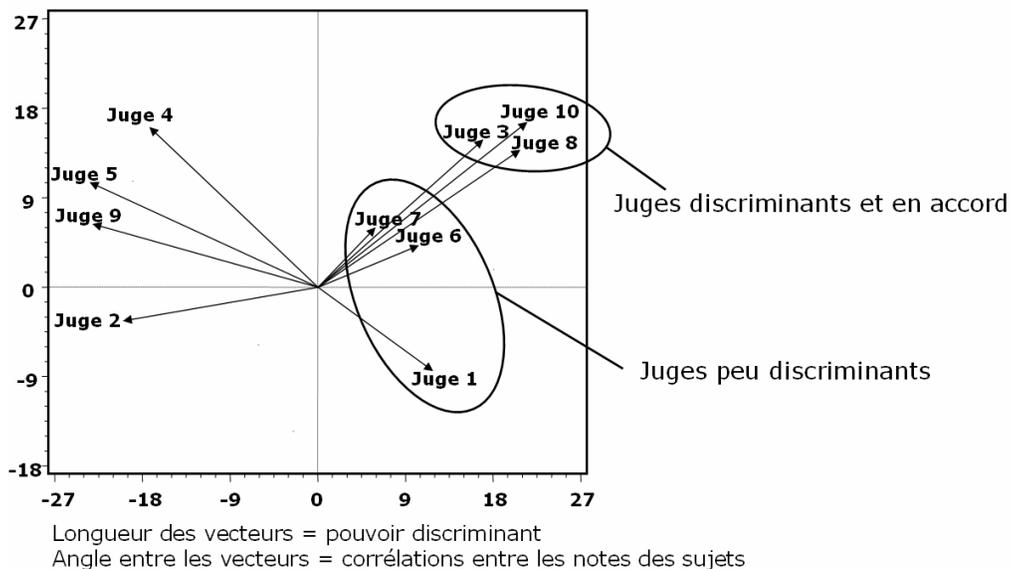


Figure 1-14 : ACP pour un descripteur avec les sujets en variables

4.1.2 ACP par sujet

L'ACP par sujet est effectuée sur le tableau contenant les produits en lignes et les descripteurs en colonnes. Les observations sont les moyennes des notes des répétitions de chaque produit. Ce type d'analyse conduit à une représentation du positionnement relatif des produits et des corrélations entre variables pour ce sujet (King, Hall et al., 2001). Pour obtenir une indication visuelle du niveau de répétabilité du sujet, l'ACP où les individus sont les répétitions de chaque produit peut être réalisée. La proximité des individus représentant les répétitions d'un même produit traduit le niveau de répétabilité du sujet.

4.1.3 Autres ACP

Certains auteurs réalisent l'ACP de la matrice qui considère les produits*séances comme individus et les descripteurs comme variables (Sinesio, Risvik et al., 1990; Risvik, Colwill et al., 1992; Vannier, Brun et al., 1999). Pour chaque séance, les valeurs correspondent à la moyenne sur le panel. Chaque produit intervient donc autant de fois que de répétitions. La répétabilité se traduit par la proximité des répétitions d'un même produit sur la carte d'ACP. Comme toutes les autres ACP, cette méthode reste descriptive et ne produit pas de test.

Il est également possible de considérer la matrice constituée de la juxtaposition horizontale des tableaux de dimension produit*descripteur de chaque sujet (Kunert et Qannari, 1999). Les auteurs réalisent au préalable une standardisation de la variance de chaque tableau individuel afin d'accorder le même poids aux données de chaque panéliste. Le nombre total de variables est donc égal à descripteur*sujet et le nombre d'individus correspond au nombre de produits. Contrairement à l'ACP du tableau des moyennes, cette approche permet d'obtenir une configuration des produits qui tient compte des observations de chaque sujet. Les corrélations entre variables homologues permettent en outre de vérifier si la compréhension des descripteurs est la même pour tous les sujets.

La matrice composée des mêmes tableaux individuels, juxtaposés verticalement cette fois, permet au contraire de représenter autant d'individus que de produits multipliés par le nombre de sujets dans l'espace des descripteurs (Dijksterhuis, 1998). La proximité des points concernant un produit (pour les différents sujets) donne une indication sur la similarité de la notation du produit par les panélistes.

4.2 Analyse de la variance multivariée

L'analyse de la variance multivariée (MANOVA) est la généralisation multidimensionnelle de l'analyse de la variance (ANOVA). Cette méthode permet de tenir compte des données de plusieurs descripteurs simultanément. L'estimation des différents effets n'utilise plus les carrés moyens de chaque facteur mais les matrices de covariance entre les descripteurs (figure 1-15).

	ANOVA	MANOVA
Intensité des différences entre produits	Carré moyen produit : CM_{prod}	Matrice de covariance des descripteurs entre produit : $V_B (p \times p)$
Différences entre sujets	Carré moyen de l'interaction : $CM_{prod*suj}$	Matrice de covariance des descripteurs produit*sujet : $V_W (p \times p)$
Discrimination du groupe entre les produits	$F_{prod} = CM_{prod}/CM_{prod*suj}$	Valeurs propres de $V_B V_W^{-1}$ (cf. CVA) Tests de Wilks, Pillai, Hotteling, Roy
Dimensions qui maximisent la discrimination		Vecteurs propres de $V_B V_W^{-1}$ (=variables canoniques de la CVA)

Figure 1-15 : parallèle ente ANOVA et MANOVA (d'après Schlich, 2004)

Plusieurs statistiques multidimensionnelles, qui suivent la même philosophie que le test de Fisher de l'analyse univariée, peuvent ensuite être calculées pour conclure sur la significativité de chaque effet. Le test le plus répandu est le lambda de Wilks (Everitt et Dunn, 1991; Polit, 1996). Il consiste à réaliser une approximation de la statistique de Fisher en utilisant les déterminants des matrices de corrélation et en tenant compte du nombre de descripteurs de l'étude. Les trois autres indices les plus connus sont le critère de Roy, la trace de Pillai et la trace de Hotelling-Lawley. Lorsque le nombre d'observations augmente, ces tests convergent asymptotiquement.

A partir d'un modèle de MANOVA déterminé, il est aussi possible d'obtenir une représentation graphique de chaque effet en utilisant l'analyse en variables canoniques (ou CVA, pour Canonical Variate Analysis). La CVA détermine les axes qui maximisent la dispersion des individus (i.e. les produits) dans l'espace des variables (comme l'ACP) tout en minimisant la variabilité autour de chaque individu. Cela revient à définir une matrice de covariance entre variables, W , qui contient les variabilités résiduelles associées aux valeurs de la matrice B . Le calcul des valeurs propres et des vecteurs propres de la matrice BW^{-1} donne alors les variables canoniques de la CVA (figure 1-15).

Porcherot et Schlich (2000), utilisent le modèle de MANOVA : descripteur = produit + sujet + produit*sujet + erreur. B est la matrice de covariance de l'effet produit et W est la matrice de covariance de l'interaction produit*sujet. La CVA ainsi obtenue maximise donc la séparation entre les produits en tenant compte de l'hétérogénéité du panel. Cette analyse est à mettre en parallèle avec le test de l'effet produit lorsque le facteur sujet est aléatoire ($F_{produit} = CM_{produit}/CM_{produit*sujet}$). Les auteurs ajoutent des ellipses de confiances autour de la moyenne de chaque produit en se basant sur les observations individuelles qui peuvent être projetées comme individus supplémentaires sur la carte (figure 1-16). Le calcul des ellipses utilise l'équation (1-16).

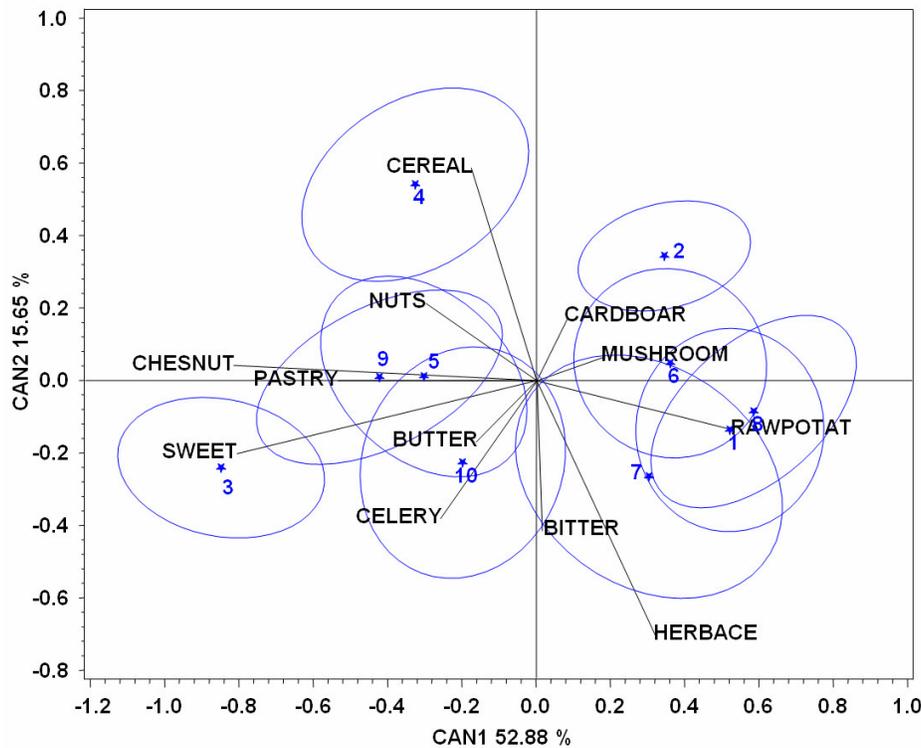


Figure 1-16 : Plan factoriel (axes 1 et 2) de la CVA

Comparée à l'ACP avec ellipses de confiance, la CVA présente l'avantage de tenir compte de la variabilité entre les sujets pour définir les axes principaux.

Monrozier et Danzart (2001) comparent les résultats de l'ACP (cartographie des trois premières dimensions) avec plusieurs cartes de CVA (tridimensionnelles) issues des modèles de MANOVA suivants :

- M1 : produit + erreur
- M2 : produit + sujet + erreur
- M3 : produit + sujet + répétition + erreur
- M4 : produit + sujet + produit*sujet + erreur
- M5 : produit + sujet + répétition + produit*sujet + erreur

Les auteurs construisent les CVA respectives où B est la matrice de covariance de l'effet produit et W est la matrice de covariance de l'erreur. Au niveau unidimensionnel, cela correspondrait au test de l'effet produit lorsque tous les facteurs sont fixes. Leurs résultats montrent que le plan d'ACP explique un pourcentage de variance un peu plus faible que les CVA. De plus, les ellipses de confiance montrent que les produits sont moins bien discriminés avec l'ACP. La position des produits sur les trois premières dimensions est également légèrement différente entre l'ACP et les CVA. Les cartes factorielles des CVA montrent en revanche des positionnements de produits et des corrélations entre variables similaires. Le modèle M1 conduit cependant à l'obtention d'ellipses de confiance plus réduites, indiquant une meilleure discrimination des produits dans ce modèle. Cette conclusion tend à suggérer que les variabilités contenues dans les facteurs sujet, répétition et produit*sujet étaient relativement faibles.

La CVA permet, grâce à la représentation des ellipses de confiance, de visualiser le niveau de discrimination entre les produits et le niveau d'accord entre les sujets. De plus, elle tient compte du désaccord entre les sujets pour déterminer les axes principaux. Cependant, cette analyse ne génère pas explicitement d'indice d'accord multidimensionnel des juges.

4.3 Analyse Procustéenne Généralisée

Historiquement, l'analyse procustéenne généralisée, plus connue sous l'acronyme GPA (Generalized Procrustes Analysis) a été introduite en analyse sensorielle pour permettre de traiter les données de profil libre choix (Gower, 1975; Arnold et Williams, 1986). Dans ce type d'épreuve, tous les sujets dégustent les mêmes produits mais chacun utilise sa propre liste de descripteurs sensoriels. Les données relatives au juge j sont rangées dans la matrice X_j à l lignes (nombre de produits) et p_j colonnes (nombre de descripteurs du sujet j). Le principe de la GPA est de déterminer un compromis C des configurations X_j en se basant uniquement sur les distances entre produits. Les descripteurs peuvent donc être différents d'un juge à l'autre. Cette méthode fonctionne également dans le cas particulier où les descripteurs sont les mêmes pour tous les sujets. Afin de faire coïncider au mieux les configurations des produits de chaque sujet entre elles, des transformations géométriques (rotations et dilatations/contractions) sont appliquées à chaque matrice X_j . La détermination de C revient à résoudre le problème d'optimisation suivant (Qannari, Courcoux et al., 1995; Qannari, Courcoux et al., 1997) :

$$\underset{\{C, \alpha_j, H_j\}}{\text{Min}} \sum \|\alpha_j X_j H_j - C\|^2 \quad (1-17)$$

Où H_j est la rotation qui ajuste la configuration X_j au compromis, C , et α_j est le facteur d'échelle associé à X_j . Gower met en place une procédure itérative qui ajuste de proche en proche les configurations individuelles à la configuration C , à l'aide de rotations et de dilatations/contractions. La figure 1-17 présente la démarche de la méthode GPA de manière géométrique.

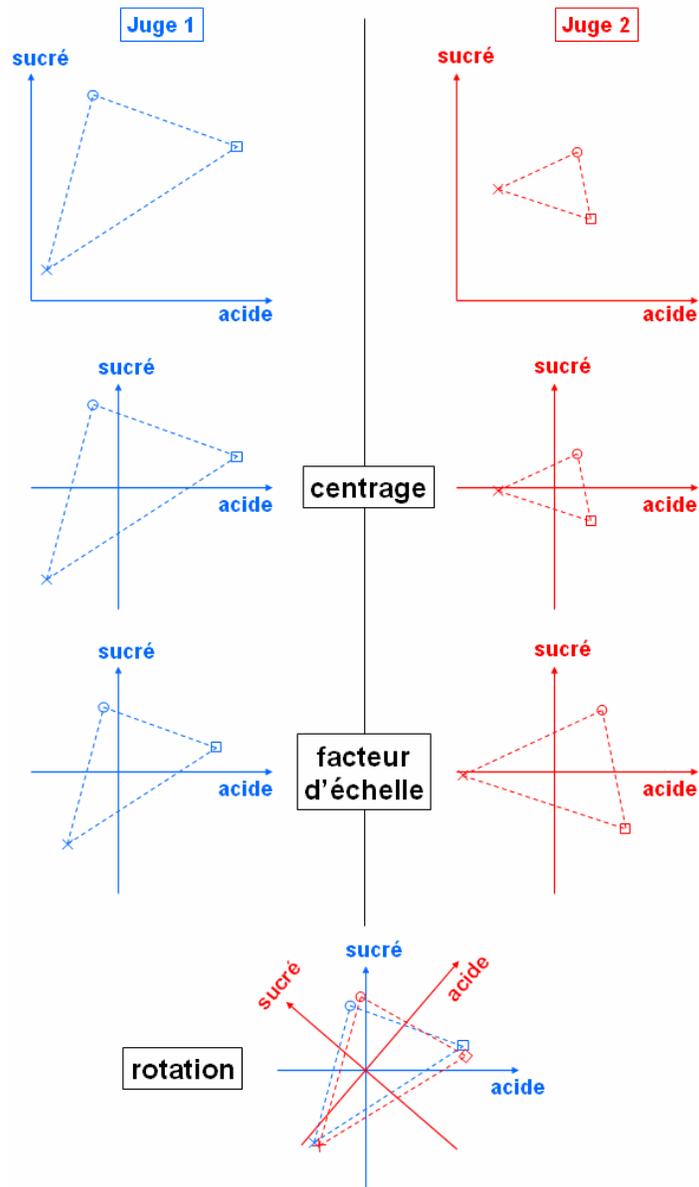


Figure 1-17 : Représentation géométrique de la démarche de la méthode GPA (exemple avec deux juges)

L'ACP de la matrice C permet alors de représenter la configuration des produits du compromis dans un espace de dimension réduite. Les axes peuvent être interprétés en fonction de leurs corrélations avec les descripteurs individuels. Par rapport à l'ACP sur les notes moyennes du groupe, la méthode GPA donne moins d'importance aux sujets qui ont un comportement atypique. Plusieurs auteurs ont comparé ces deux méthodes sur des données sensorielles. Il en ressort que la méthode GPA permet d'identifier les sujets qui ont un style de notation extrême et d'obtenir des informations plus détaillées sur les données (Sinesio, Risvik et al., 1990; Hunter et Muir, 1995) mais la carte des produits varie en général très peu selon la méthode (Risvik, Colwill et al., 1992).

L'inconvénient majeur de la méthode GPA réside dans le mode de calcul de la configuration consensus. Il s'agit d'un processus itératif qui ne conduit pas toujours à la solution optimale. De plus, cette méthode ne produit pas explicitement d'indice d'accord de chaque sujet avec la configuration moyenne. Qannari et Courcoux (1997) proposent une variante de la méthode GPA qui calcule explicitement un facteur de pondération de chaque sujet. Ce poids est d'autant plus élevé que le sujet est en accord avec le compromis. Les auteurs montrent ensuite que les résultats de la méthode classique et de sa variante, plus simple à mettre en œuvre, conduisent à des résultats très proches. Ces approches sont ensuite comparées avec la méthode STATIS, décrite dans le paragraphe suivant.

4.4 Méthode STATIS

La méthode de Structuration des Tableaux A Trois Indices de la Statistique (STATIS) (L' Hermier des Plantes, 1976; Lavit, 1988) recherche également un compromis entre les configurations individuelles, mais en ayant recours aux matrices des produits scalaires W_j de chaque juge (Tucker, 1966) :

$$W_j = X_j X_j' \quad (1-18)$$

Par rapport à GPA, cette méthode présente l'avantage de se dispenser de déterminer explicitement les rotations qui ajustent les configurations (Qannari, Courcoux et al., 1997), ce qui simplifie nettement le problème de détermination d'un compromis. La fonction à minimiser est alors :

$$\underset{\{w, \gamma_j\}}{\text{Min}} \sum \|\gamma_j^2 W_j - W\|^2 \quad (1-19)$$

Où W représente le compromis entre les matrices de produits scalaires W_j . Une manière de déterminer les poids γ_j consiste ensuite à calculer, pour chaque paire de sujet j et j' , le coefficient RV (Escoufier, 1973) entre les matrices W_j et $W_{j'}$:

$$RV_{jj'} = \frac{\text{trace}(W_j W_{j'})}{\sqrt{\text{trace}(W_j W_j)} \sqrt{\text{trace}(W_{j'} W_{j'})}} \quad (1-20)$$

Ce coefficient, compris entre 0 et 1, correspond exactement au coefficient de corrélation de Pearson qui serait calculé sur les tableaux vectorisés (tableaux dans lesquels les colonnes sont positionnées les unes sous les autres). Un coefficient RV de 1 signifie que les configurations des individus dans les tableaux X_j et $X_{j'}$ sont identiques à une rotation et un facteur d'échelle près. Le regroupement de tous les coefficients $RV_{jj'}$ forme une matrice de taille $J \times J$ d'indices de proximité entre tableaux (i.e. entre les sujets). Le premier vecteur propre de cette matrice des RV fournit alors les coefficients δ_j permettant de calculer les poids des sujets. En effet, un coefficient δ_j élevé pour un sujet indique que la configuration de ses produits était similaire à celle de beaucoup d'autres sujets. Son accord avec le groupe est donc élevé. Sa configuration aura donc un poids w_j plus important dans le calcul du compromis.

* La « trace » correspond à la somme des éléments diagonaux de la matrice considérée.

Cette pondération correspond à l'équation (1-21):

$$w_j = \frac{\delta_j}{\sum_{j=1}^J \delta_j} \quad (1-21)$$

L'opérateur compromis W se calcule alors de la manière suivante :

$$W = \sum_{j=1}^J w_j W_j \quad (1-22)$$

La diagonalisation de la matrice W donne accès à la représentation graphique de la configuration des produits pour le compromis, sur laquelle les descripteurs de chaque sujet peuvent être projetés en variables supplémentaires. Il est également possible d'obtenir une carte qui représente l'accord des sujets en réalisant la diagonalisation de la matrice des RV centrée en ligne et en colonne. Cette étape correspond à l'Analyse en Coordonnées Principales (PCO) de la matrice des RV, dont le principe est de diagonaliser une matrice de distances pour cartographier les écarts entre individus. Le double centrage de la matrice des RV permet de ne conserver que l'information sur la proximité relative entre les éléments ; une valeur positive indique, par exemple, que les RV entre les deux sujets considérés étaient plus élevés que la moyenne des RV de chacun des deux sujets considérés. Les vecteurs propres issus de la diagonalisation de cette matrice permettent d'obtenir les coordonnées des sujets sur chaque axe principal. La représentation graphique associée permet de visualiser les sujets sur une carte. Lorsque deux juges sont proches, cela signifie que les différences relatives entre les produits sont les mêmes pour ces deux sujets. La configuration des produits des juges à proximité du centre de la carte est très similaire à la configuration du compromis.

Afin d'obtenir un indice d'accord pour chaque sujet, il est possible de calculer la moyenne des coefficients RV individuels obtenus avec chaque autre juge :

$$\overline{RV}_j = \frac{1}{J-1} \sum_{\substack{j=1 \\ j \neq j'}}^J RV_{jj'} \quad (1-23)$$

Ils proposent également de calculer un indice d'accord pour le groupe, égal à la moyenne des coefficients individuels. Le coefficient RV est utilisé dans de nombreuses publications comme indice d'accord entre deux matrices. Afin de conclure sur la significativité de cet indice multidimensionnel, un test basé sur des permutations aléatoires a été développé (Schlich, 1993). Il consiste à permuter entre elles les lignes d'une des deux matrices utilisées dans la comparaison et à calculer un RV dans cette situation. Un grand nombre de simulations de ce genre sont réalisées afin d'obtenir une distribution des RV sous l'hypothèse de nullité de l'accord entre les deux matrices. La position de la valeur réelle du RV dans cette distribution permet de donner la probabilité d'obtenir une valeur supérieure ou égale à la valeur obtenue, sous l'hypothèse nulle d'absence d'accord. Cela permet également de calculer une valeur normalisée du RV qui pourra aisément être comparée aux valeurs de la loi Normale centrée réduite. Une alternative analytique permet de s'affranchir du calcul des permutations (Kazi-Aoual, 1993). En supposant la normalité de la distribution des RV, l'auteur propose des formules

permettant d'estimer l'espérance et la variance du coefficient RV sous l'hypothèse nulle. Un RV normalisé peut alors être calculé de la manière suivante :

$$NRV_{jj'} = \frac{RV_{jj'} - E(RV_{jj'})}{\sqrt{V(RV_{jj'})}} \quad (1-24)$$

La valeur obtenue permet de retrouver le quantile correspondant dans la loi Normale centrée réduite et ainsi définir une probabilité de nullité du coefficient. Par exemple, étant donné que le quantile d'ordre 95% (unilatéral à droite) de la loi Normale centrée réduite est 1.645, un coefficient NRV supérieur à cette valeur sera déclaré significatif au seuil de 5% ; l'accord entre les deux configurations testées sera déclaré significatif.

Une méthode dérivée de STATIS, appelée STATIS DUAL, n'utilise pas les matrices des produits scalaires, W_j , mais les matrices des corrélations ou des covariances entre les variables, V_j (Schlich, 1996) :

$$V_j = X_j' X_j \quad (1-25)$$

Le coefficient RVD s'écrit alors :

$$RVD_{jj'} = \frac{\text{trace}(V_j V_{j'})}{\sqrt{\text{trace}(V_j V_j)} \sqrt{\text{trace}(V_{j'} V_{j'})}} \quad (1-26)$$

La démarche est ensuite similaire à celle de STATIS mais la méthode renseigne sur les similarités des corrélations entre les variables d'un sujet à l'autre. La carte des descripteurs, issue du compromis V entre les situations individuelles, permettra ainsi de visualiser les variables les plus corrélées entre elles. De la même manière, le coefficient RV, alors appelé RVD, est calculé pour chaque paire de sujets. Une valeur élevée indique que les corrélations entre les descripteurs homologues sont similaires d'un juge à l'autre. Cette méthode peut donc s'appliquer seulement si chaque sujet a noté les mêmes descripteurs. La version normalisée, NRVD, de ce coefficient indique si l'accord entre les deux configurations de corrélation des descripteurs est significatif. Cependant, il n'existe pas de formule permettant d'estimer la variance du NRVD. Le calcul fait alors appel à un test basé sur un grand nombre de simulations dans lesquels les lignes d'un des deux tableaux à comparer sont permutées de manière aléatoire.

En résumé, STATIS s'intéresse à l'accord entre les configurations des produits de chaque sujet, indépendamment des variables de l'étude. Concrètement, lors de la réalisation d'un profil sensoriel portant sur deux descripteurs uniquement (acide et amer), si deux juges ont donné exactement les mêmes notes aux produits homologues mais que les descripteurs sont inversés (les notes d'acidité d'un juge sont les notes d'amertume de l'autre juge et inversement), le coefficient RV sera malgré tout égal à 1. Ainsi, le coefficient d'accord conserve la même valeur lorsque les sujets n'ont pas interprété les descripteurs de la même manière mais qu'ils font globalement les mêmes différences entre produits.

De manière réciproque, la méthode STATIS DUAL confronte les corrélations entre descripteurs pour chaque sujet sans tenir compte de la spécificité des produits. Dans ce cas, le coefficient RVD garde la même valeur tant que les corrélations entre variables sont identiques, même si les configurations des produits changent.

Dans le cas de l'analyse de données de profil sensoriel, la question est en général de savoir si les sujets ont trouvé les mêmes différences entre produits sur les mêmes descripteurs (cela suppose alors qu'ils interprètent les descripteurs de la même manière). Il est donc souhaitable de tenir compte à la fois des configurations des produits ET des corrélations entre variables. Dans cette optique, une nouvelle stratégie, développée par Ledauphin, Hannafi et al. (2006), appelée méthode GAMMA, a été mise en place. L'indice d'accord entre deux sujets est alors calculé de la manière suivante :

$$s_{jj'} = \frac{1 + \frac{\text{trace}(X_j' X_{j'})}{\sqrt{\text{trace}(X_j' X_j)} \sqrt{\text{trace}(X_{j'}' X_{j'})}}}{2} \quad (1-27)$$

Ce coefficient utilise directement les matrices contenant les notes des produits. L'expression $\text{trace}(X_j' X_{j'})$ correspond au calcul de la somme des covariances entre descripteurs homologues. La division par le produit des deux racines revient à standardiser la valeur obtenue entre -1 et 1. Le terme devient alors strictement équivalent à un coefficient de corrélation. Les derniers éléments de la formule ramènent ce coefficient entre 0 et 1 (comme les coefficients RV et RVD). Si la corrélation est égale à 1 pour chaque descripteur, le coefficient atteint sa valeur maximale : 1. La matrice des RV ou des RVD, obtenue respectivement avec STATIS et STATIS DUAL est alors remplacée par la matrice contenant ces coefficients de similarité s . La même démarche est ensuite utilisée. Après avoir déterminé une matrice compromis, C , de dimension produit*descripteur, les auteurs définissent un indice de performance, α , dont la valeur est d'autant plus grande que le sujet est en accord avec le compromis :

$$\alpha_j = \frac{\text{trace}(X_j' C)}{\sqrt{\text{trace}(C' C)}} \quad (1-28)$$

Ils définissent également γ , la moyenne des α_j , comme un indice global de performance du groupe. Plus cet indice est élevé, plus les réponses des panélistes étaient homogènes. Un test de permutation a également été développé afin de conclure sur la significativité de ces coefficients (Ledauphin, Hanafi et al., 2006). La figure 1-18 propose une comparaison théorique de ces trois méthodes afin de mettre en exergue les points communs et les différences. Une synthèse plus détaillée est disponible en annexe 1.

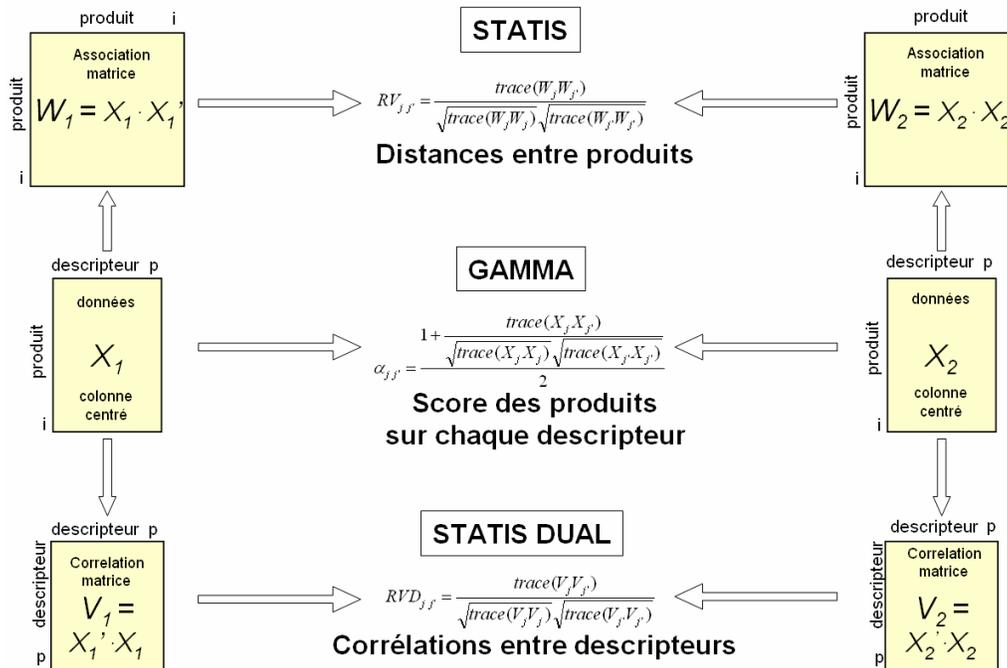


Figure 1-18 : Comparaison des méthodes STATIS, STATIS DUAL et GAMMA

Sur cette figure, la méthode développée par Ledauphin apparaît clairement à mi-chemin entre les deux méthodes STATIS et STATIS DUAL. Des exemples de résultats concernant ces trois méthodes seront discutés dans le chapitre 2.

4.5 Analyse Factorielle Multiple

L'analyse factorielle multiple (AFM) est une méthode multidimensionnelle appartenant au même ensemble de méthodes que GPA et STATIS. Elle a été introduite en analyse sensorielle par Escofier et Pagès dans les années 1980 (Escofier et Pagès, 1984; Escofier et Pagès, 1985; Escofier et Pagès, 1988). A l'instar de STATIS, elle repose sur une ACP pondérée du tableau constitué de la juxtaposition des matrices individuelles X_i (produit*descripteur). La différence essentielle se situe dans l'expression de la pondération affectée à chaque vision individuelle. Alors que STATIS utilise une pondération reflétant l'accord multidimensionnel du panéliste avec le groupe, l'AFM propose la pondération suivante :

$$w_i = \frac{1}{\lambda_1^i} \quad (1-29)$$

Chaque tableau X_i est donc pondéré par l'inverse de la première valeur propre de l'ACP de la matrice X_i . Géométriquement, cela revient à rendre égale à 1 l'inertie maximum de chaque juge sur le premier axe principal (Pagès, 1996). Les sujets dont la structure des données est fortement unidimensionnelle auront donc une importance moindre dans le calcul de la configuration moyenne du panel.

Cette méthode a souvent été appliquée à l'étude des différences entre produits (Pagès, Asselin et al., 1987; Le Quéré, Husson et al., 2006; Morand et Pagès, 2006) mais plus rarement à l'analyse de la performance des panélistes. Ce constat est vraisemblablement lié au fait que l'objectif de la

pondération utilisée n'est pas de tenir compte de l'accord entre les panélistes mais uniquement de leur dimensionnalité. L'ajout d'ellipses de confiance autour de la moyenne de chaque produit sur la représentation graphique de l'AFM permet cependant de représenter l'accord entre les panélistes (taille des ellipses) et leur niveau de discrimination (séparation des ellipses).

Ces dernières années, une version hiérarchique de l'AFM a été proposée (Le Dien et Pagès, 2003). Elle permet de réaliser le même type d'analyse en ajoutant une structure en sous-groupes de variables. Les sujets peuvent ainsi être regroupés en catégories et l'analyse tiendra compte de ce niveau de hiérarchisation supplémentaire. Des ellipses de confiance ont également été développées dans ce cas (Le Dien, Husson et al., 2004; Le Dien, Pagès et al., 2004).

4.6 Méthodes avec pondération par dimension

La PCAMDS (Qannari et Meyners, 2001), acronyme de Principal Component Analysis on Merged Data Sets, consiste à réaliser une ACP sur les tableaux individuels (produit*descripteur) concaténés horizontalement (comme Kunert et Qannari, 1999). Chaque tableau individuel est au préalable standardisé afin d'homogénéiser les variances individuelles. Chaque tableau individuel est ensuite régressé sur chaque composante principale de l'ACP afin de déterminer la corrélation entre chaque juge et chaque composante principale. Cette corrélation caractérise l'accord entre la configuration des produits du sujet et la configuration des produits, projetés sur la composante principale en question, pour le groupe. En comparant ces corrélations selon les dimensions de l'ACP pour un sujet donné, cela permet de savoir avec quelles dimensions du compromis le sujet est le plus en accord. Par rapport aux méthodes GPA ou STATIS, cela permet d'affiner la pondération des sujets pour chaque dimension. De plus, la moyenne des coefficients de tous les sujets pour chaque dimension permet de mesurer le niveau de représentativité de chaque composante principale de l'ACP par rapport aux configurations individuelles. Cette moyenne correspond également au pourcentage de variance expliquée par la composante principale. Cette méthode a l'avantage de mettre en œuvre des techniques simples (ACP et régression). Notons toutefois que celle-ci peut conduire à sélectionner plus de composantes principales que de descripteurs sensoriels puisque le nombre de variables de cette ACP est égal à sujet*descripteur. Une telle analyse a malgré tout du sens au niveau sensoriel : cela signifie que l'on ne considère pas que les descripteurs homologues ont exactement la même signification pour chaque sujet. Cependant, on peut se demander si cette méthode se comporte toujours bien étant donné le nombre important de variables par rapport au nombre d'individus lors de la réalisation de l'ACP. En pratique, le problème se pose rarement car ces variables sont, en principe, très fortement corrélées entre les sujets.

Dans l'optique d'obtenir, comme avec la PCAMDS, un indice d'accord de chaque sujet avec chaque dimension issue du calcul d'un compromis, l'ACCPS (Analyse des Composantes Communes et Poids Spécifiques) travaille sur les matrices des produits scalaires (dimension produit*produit) de chaque sujet et stipule l'existence d'une structure commune pour l'ensemble de ces matrices (Qannari, Wakeling et al., 2000; Hanafi, 2003). Cette structure peut s'exprimer d'une manière variable d'une dimension à une autre et d'un tableau à un autre. La détermination des poids est effectuée, dimension après dimension, à l'aide d'une procédure itérative. Cette méthode permet d'obtenir les

représentations graphiques de la typologie des produits, des corrélations entre variables et de l'accord des sujets avec le compromis sur chaque dimension. Cette méthode, très peu utilisée en analyse sensorielle, permet de caractériser les différences entre produits tout en tenant compte de la spécificité de chaque juge sur chaque dimension. Contrairement à la PCAMDS, le nombre de dimensions retenu ne peut pas dépasser le nombre de variables. Elle met cependant en jeu des notions et des techniques statistiques assez complexes.

4.7 Bilan des méthodes multidimensionnelles

Les analyses multidimensionnelles donnent la possibilité de tenir compte de plusieurs variables simultanément. En analyse sensorielle l'objectif est en général de résumer l'information contenue dans les descripteurs en un petit nombre de dimensions pour créer une carte des produits dans un plan. Afin de tenir compte de l'hétérogénéité des réponses de chaque sujet, différentes stratégies ont été mises au point. La plupart visent à calculer un compromis entre les configurations individuelles. Ces méthodes permettent de calculer un indice d'accord de chaque sujet avec le groupe (GPA, STATIS, STATIS DUAL, GAMMA). Dans le même ordre d'idée, certaines méthodes permettent aussi d'obtenir ce type d'indice pour chaque dimension du compromis (PCAMDS, ACCPS). D'autres méthodes sont issues de la généralisation de l'ANOVA au cas multidimensionnel (MANOVA, CVA). L'objectif est alors de trouver les dimensions qui maximisent les différences entre produits au niveau du groupe en tenant compte du désaccord entre les sujets. Afin de visualiser l'hétérogénéité de la notation des panélistes, plusieurs méthodes proposent d'ajouter des ellipses de confiance autour de la moyenne de chaque produit.

Toutes ces méthodes permettent en général d'obtenir des informations sur le pouvoir discriminant du groupe et l'accord entre les sujets. En revanche, elles sont rarement utilisées pour caractériser la répétabilité multidimensionnelle. Il est néanmoins possible dans la majorité des cas d'obtenir une information sur la répétabilité en considérant les répétitions, plutôt que les sujets, comme les individus ou comme les variables dans les analyses. Les résultats ainsi obtenus caractérisent alors la répétabilité d'un seul panéliste au niveau multidimensionnel (McEwan, Hunter et al., 2002).

5 Conclusion

Tant au niveau unidimensionnel que multidimensionnel, il existe un grand nombre de méthodes qui permettent de caractériser la performance. Les méthodes univariées conduisent à des analyses plus fines de chaque descripteur, tandis que les analyses multidimensionnelles permettent d'avoir une vision globale de l'espace sensoriel. De nombreuses méthodes présentées ici produisent des indices dont la signification est proche. Il n'est donc pas nécessaire d'utiliser toutes ces méthodes afin de caractériser la performance d'un panel. Chacune correspond à des situations et des points de vue légèrement différents. Pour l'animateur de panel dont l'objectif est de déterminer facilement et rapidement la qualité de son panel, il est nécessaire d'effectuer une sélection parmi ces méthodes. Le chapitre suivant présente la sélection que nous avons retenue, en fonction des objectifs qui semblaient les plus importants pour l'animateur de panel et en fonction des méthodes déjà existantes au sein du laboratoire.

Chapitre 2 : Sélection des méthodes d'analyse de la performance et automatisation d'un outil de traitement des données

1 Introduction

La sélection de méthodes présentée dans ce chapitre a pour objectif de rassembler des analyses statistiques qui permettent d'avoir une vision complète et synthétique des différents aspects de la performance. Des méthodes unidimensionnelles et multidimensionnelles sont donc utilisées afin de caractériser la performance du panel et de ses panélistes en terme de répétabilité, de discrimination et d'accord. La création d'un outil automatisé contenant ces méthodes est destinée aux animateurs de panels désireux de pouvoir établir un diagnostic rapide et efficace de la qualité de leur jury. Un effort particulier a donc été porté à la présentation synthétique des résultats et à la lisibilité des tableaux et des figures proposés. Afin de proposer un outil de traitement complet, certaines analyses, dont l'objectif est de caractériser les produits et non la performance des juges, ont également été incluses dans l'outil d'analyse de la performance. La sélection de méthodes s'appuie en majeure partie sur les outils d'analyse des données sensorielles déjà existants au laboratoire. Certaines analyses ont cependant été ajoutées afin de compléter l'outil de traitement proposé.

La sélection contient deux analyses de la performance au niveau unidimensionnel : la table "CAP" et la carte "CAPMAP". La première, décrite succinctement dans le chapitre précédent, rassemble dans un tableau des diagnostics individuels tenant compte du niveau de discrimination, d'accord et de répétabilité du panéliste. La carte "CAPMAP" utilise les indicateurs de performance calculés dans la table CAP et présente les résultats dans un espace bidimensionnel. Cela permet de positionner les sujets les uns par rapport aux autres et de visualiser les corrélations entre les descripteurs en terme de performance des panélistes. La sélection propose ensuite deux tableaux de résultats multidimensionnels. Le premier, basé sur la méthodologie des indices RV, caractérise l'accord entre les panélistes selon différentes approches ; le second renseigne sur le niveau de discrimination multidimensionnelle de chaque juge à l'aide de trois indices. Deux analyses, orientées vers la caractérisation des différences entre produits, viennent compléter l'outil d'analyse de la performance :

la table FLASH et l'analyse en variable canonique (AVC). La table FLASH propose une classification des descripteurs ainsi qu'une comparaison des moyennes des produits pour chaque descripteur. L'AVC présente une carte des différences entre produits (comme l'ACP) qui tient compte de la variabilité des réponses des panélistes.

Afin d'illustrer ces analyses, un ensemble de données (issues de Porcherot, Issanchou et al., 1998) contenant 16 produits (variétés de pommes de terre), 14 panélistes, 30 descripteurs (échelles de 0 à 100) et 2 répétitions de chaque mesure sera utilisé au cours de ce chapitre.

2 Performances unidimensionnelles

2.1 La table CAP

2.1.1 Performances du panel

La table CAP (Control of Assessor Performances) a été proposée par Schlich en 1997. Depuis sa première version, elle a subi plusieurs modifications mineures afin d'améliorer la pertinence et la présentation des résultats. Cette méthode repose essentiellement sur le modèle d'ANOVA au niveau du groupe, présenté dans le chapitre précédent (équation 1-1), et sur le modèle d'ANOVA par panéliste (équation 1-11), pour chaque descripteur. Ces deux équations sont rappelées ci-dessous :

$$\text{Modèle au niveau du groupe : } Y_{ijk} = \mu + \alpha_i + B_j + C_{ij} + \varepsilon_{ijk} \quad (1-1)$$

$$B_j \sim N(0, \sigma_{\text{suj}}^2), C_{ij} \sim N(0, \sigma_{\text{suj} \cdot \text{prod}}^2), \varepsilon_{ijk} \sim N(0, \sigma_e^2), \text{indépendants entre eux.}$$

$$\text{Modèle par panéliste : } Y_{ik} = \alpha_i + \varepsilon_{ik} \quad \varepsilon_{ik} \sim N(0, \sigma_{\text{ic}}^2) \quad (1-11)$$

Le paramètre α_i désigne le facteur produit, B_j correspond au facteur sujet et C_{ij} à l'interaction entre les facteurs produit et sujet. Les résultats issus de ces modèles sont présentés dans une table contenant les descripteurs en lignes et les sujets en colonnes (tableau 2-1).

Tableau 2-1 : exemple de table CAP

attribut	fprod	finter	mean	rmse	7	5	11	3	1	8	12	4	10	14	6	9	2	13
FIRM	17.17	1.51	37.5	18.8	+	+	+	+	+	+	+	+	+	+	!=	+	+	=
MASHABL	13.91	0.88	40.2	25.5	+	=	+	+	=	!=	=	=	+	=	+	+	=	=
MOIST	11.65	1.20	46.9	21.1	+	+	+	+	=	+	=	+	+	!=	+	+	=	=
GRAINY	11.57	1.10	26.9	20.6	+	+	+	+	!=	=	+	=	=	+	!=	+	=	=
MEALY	11.44	0.83	42.3	25.2	+	=	=	=	!=	=	=	=	=	=	=	=	=	+
SWEET	9.49	1.10	42.4	21.9	=	+	=	+	=	!=	=	+	!=	+	+	=	=	=
COMPACT	8.57	1.77	48.6	18.6	+	+	=	+	=	+	!=	+	+	+	!=	=	+	+
HOMOGEN	7.40	1.17	59.0	22.4	+	+	+	=	=	+	=	+	+	!=	=	+	+	+
CHESNUT	6.02	1.10	41.1	24.3	+	+	=	=	=	+	!=	=	+	!=	=	+	+	+
EARTHY	5.78	1.07	29.8	21.7	+	+	=	=	=	!=	=	=	=	=	=	+	+	+
GREASY	5.26	1.47	15.3	15.7	+	+	=	+	=	!=	=	=	+	=	=	+	+	+
PASTRY	4.42	1.51	19.2	20.1	=	+	=	+	=	+	+	+	+	=	=	+	+	+
PASTY	3.67	1.32	34.7	22.8	+	+	+	+	+	+	+	+	+	+	+	+	+	+
HERBACE	3.23	1.55	30.5	21.2	+	+	+	+	+	+	+	+	+	+	+	+	+	+
STICKY	2.95	1.30	27.6	22.5	-	+	=	=	=	+	=	+	-	!=	=	=	=	=
BITTER	2.87	1.28	26.9	20.2	+	=	=	=	+	=	+	=	=	=	=	=	=	=
RAWPOTA	2.81	1.09	22.7	21.9	+	=	+	=	+	+	+	+	+	+	+	+	+	+
ARTICHOK	2.62	0.95	39.7	24.7	=	+	=	=	=	+	=	=	+	=	=	=	=	=
ASTRING	2.58	1.00	29.3	21.9	-	+	+	+	+	+	+	+	+	+	+	+	+	+
CEREAL	2.32	1.05	26.3	22.5	+	+	=	=	+	+	+	+	+	+	+	+	+	+
CELERY	1.86	1.38	16.9	15.9	+	+	+	+	+	+	+	+	+	+	+	+	+	+
METALLIC	1.77	1.13	14.9	17.7	=	+	=	=	=	=	=	=	=	=	=	=	=	=
BUTTER	1.73	1.52	16.8	17.3	=	=	=	=	=	+	+	+	+	+	+	+	+	+
NUTS	1.63	1.85	16.3	16.5	+	+	+	+	+	+	+	+	+	+	+	+	+	+
STRAW	1.62	1.05	29.0	21.5	-	+	=	=	=	+	=	=	=	=	=	=	=	=
CARDBOA	1.45	1.30	8.07	13.8	+	+	=	=	+	+	+	+	+	+	+	+	+	+
TOASTEDB	1.43	1.06	8.71	15.8	=	=	=	=	=	=	=	=	=	=	=	=	=	=
MUSHROO	1.38	1.29	12.0	14.9	+	+	=	=	+	+	+	+	+	+	+	+	+	+
ACID	1.34	1.39	14.7	15.8	+	+	=	=	+	+	+	+	+	+	+	+	+	+
UMAMI	0.99	1.08	14.8	17.4	+	+	=	=	+	+	+	+	+	+	+	+	+	+
rankf					6.57	6.13	6.50	6.67	6.73	7.03	7.10	7.37	7.60	7.90	8.23	8.87	9.20	10.10

FPROD et FINTER : statistique de Fisher de l'effet produit et de l'interaction sujet*produit dans le modèle d'analyse de la variance descripteur = produit + sujet + produit*sujet (effet sujet aléatoire)
 MEAN : moyenne des notes, tous produits et sujets confondus
 RMSE : racine carré de l'erreur résiduelle dans le modèle défini ci-dessus
 RANKF : moyenne des rangs des effets produits individuels sur l'ensemble des descripteurs
 Les sujets et les descripteurs sont triés selon leur pouvoir discriminant (resp. RANKF et FPROD)
 + : panéliste discriminant et en accord avec le groupe ; - : panéliste discriminant et en désaccord
 = : panéliste non discriminant ; ! : panéliste moins répétable que le groupe

Les descripteurs sont triés selon le pouvoir discriminant du groupe. Celui-ci est mesuré par la valeur de la statistique de Fisher de l'effet produit, inscrit dans la colonne FPROD. La colonne suivante (FINTER) correspond au F de l'interaction ; elle renseigne donc sur la présence de désaccord (et/ou de différence de dispersion des notes sur l'échelle) entre les sujets. Pour ces deux colonnes, le code couleur est défini en fonction du risque associé à la statistique de Fisher :

Tableau 2-2 : couleurs associées aux risques des indices FPROD et FINTER

FPROD	FINTER
$\alpha < 0,05$	$\alpha < 0,01$
$0,05 < \alpha < 0,15$	$0,01 < \alpha < 0,10$
$0,15 < \alpha$	$0,10 < \alpha$

Les cases vertes de la colonne FPROD indiquent que l'effet produit est significatif au risque $\alpha = 0,05$. Le panel est donc discriminant. A l'inverse, les cases vertes de la colonne FINTER indiquent que l'interaction n'est pas significative ($p > 0,10$), donc que le désaccord entre les panélistes est négligeable. Le risque choisi pour l'interaction est plus faible afin de ne détecter que les situations où le désaccord est vraiment important. Les cases orange et rouges indiquent que le niveau de performance de chaque indice se dégrade. Dans la table CAP en exemple, le panel est discriminant pour tous les descripteurs excepté les sept derniers (de NUTS à UMAMI). Cependant, un désaccord persiste pour certains descripteurs discriminants comme GREASY et PASTRY (sur fond rouge), ou dans une moindre mesure STICKY et BITTER (sur fond orange). Rappelons toutefois que l'effet sujet est aléatoire dans le modèle d'ANOVA au niveau du groupe. Le dénominateur de la statistique de Fisher de l'effet produit est donc le carré moyen de l'interaction et non celui de l'erreur. Un effet produit significatif indique alors que le groupe perçoit des différences communes entre les produits malgré un éventuel désaccord entre les sujets. Par conséquent, quel que soit la valeur de FINTER, le panel est discriminant lorsque FPROD est significatif (valeur sur fond vert).

Les colonnes MEAN et RMSE contiennent respectivement la moyenne des observations de l'ensemble du panel et l'écart-type des résidus. L'information sur la moyenne permet de rendre compte du niveau de notation global afin de détecter les sensations très peu perçues parmi les produits. Par exemple, la moyenne de TOASTEDB est de 8,07 sur une échelle de 0 à 100, ce qui peut expliquer en partie le faible niveau de discrimination de ce descripteur (FPROD=1,45). L'écart-type des résidus (RMSE) correspond à l'écart-type entre les répétitions pour le modèle d'ANOVA utilisé. Sa valeur donne une indication sur le niveau de répétabilité du groupe : une valeur forte indique que le panel est moins répétable pour le descripteur considéré. Dans les données présentées, le RMSE varie entre 13,8 pour le descripteur MASHABL et 25,5 pour le descripteur MEALY, soit près du double. Cet écart peut en partie s'expliquer par la valeur des moyennes correspondantes, respectivement égales à 8,07 et 40,2. En effet, plus la moyenne des observations est proche d'une des bornes de l'échelle, plus la variabilité des notes autour de cette moyenne est réduite.

2.1.2 Performances des panélistes

Les colonnes suivantes de la table CAP (tableau 2-1) contiennent les diagnostics de la performance de chaque sujet pour chaque descripteur. Les sujets (i.e. les colonnes) sont triés selon leur RANKF. Cet indice représente une moyenne de rangs. Pour chaque descripteur, les sujets sont classés en fonction de la valeur de la statistique de Fisher obtenue pour l'effet produit dans le modèle d'ANOVA individuel note = produit + erreur. Le rang 1 est attribué au panéliste le plus discriminant. Le dernier panéliste reçoit le rang égal au nombre de sujets. Pour chaque panéliste, le RANKF correspond à la moyenne des rangs obtenus pour tous les descripteurs. Le calcul est détaillé dans la figure 2-1.

		F produit							
		Acide	Descr2	Descr3	Descr4	Descr5	Descr6	Descr7	Descr8
1. Tableau des F produits (ANOVAs individuelles)	Sujet 1	1,51	2,11	9,27	3,21	2,21	0,45	2,84	5,18
	Sujet 2	2,48	0,89	15,21	1,47	1,45	2,01	4,33	6,81
	Sujet 3	7,19	3,84	4,65	4,15	3,12	5,89	3,11	6,56
	Sujet 4	3,51	4,18	7,52	5,74	3,18	5,82	3,77	7,84
	Sujet 5	1,89	3,21	3,85	1,22	4,51	1,23	3,56	4,25
	Sujet 6	2,36	3,99	3,89	2,12	6,36	2,18	1,33	5,07

		rangs								RANKF
		Acide	Descr2	Descr3	Descr4	Descr5	Descr6	Descr7	Descr8	
2. Transformation des données en rangs pour chaque descripteur	Sujet 1	6	5	2	3	5	6	5	4	4,5
	Sujet 2	3	6	1	5	6	4	1	2	3,5
	Sujet 3	1	3	4	2	4	1	4	3	2,75
	Sujet 4	2	1	3	1	3	2	2	1	1,88
	Sujet 5	5	4	6	6	2	5	3	6	4,62
	Sujet 6	4	2	5	4	1	3	6	5	3,75

3. RANKF = moyenne des rangs de chaque sujet

Figure 2-1 : Calcul de l'indice RANKF

L'indice RANKF, borné entre 1 et le nombre de panélistes, est d'autant plus faible que le panéliste était souvent parmi les sujets les plus discriminants pour chaque descripteur. Dans la table CAP (tableau 2-1), le sujet 7 obtient le plus faible RANKF (5,57). Il est donc classé en première position parmi les sujets. Cet indice a été préféré au F de MANOVA de l'effet produit, obtenu, pour chaque panéliste, à partir du modèle multidimensionnel note = produit + erreur. Ce choix est lié à la nature de la statistique de Fisher étendue au cas multidimensionnel dont la valeur peut être très élevée alors que le panéliste n'est discriminant que pour un petit nombre de descripteurs. Dans le cas où un panéliste est extrêmement discriminant pour un seul descripteur et non discriminant pour les autres descripteurs, le F de MANOVA aura une valeur élevée. Au contraire, le RANKF de ce sujet sera médiocre puisqu'il sera classé parmi les meilleurs panélistes pour un seul descripteur. Le RANKF favorise donc les sujets plus discriminants que les autres sur un grand nombre de descripteurs.

Pour chaque descripteur, le diagnostic individuel (+, -, =, !) est établi en fonction de l'arbre de décision présenté dans le chapitre précédent (figure 1-9). L'indice de discrimination individuel est la statistique de Fisher de l'effet produit dans le modèle individuel note = produit + erreur. Lorsque le sujet est discriminant (risque $\alpha = 0,10$), son accord avec les autres panélistes discriminants est testé à l'aide du coefficient de Kendall. Ce coefficient de corrélation se calcule de la manière suivante :

$$\tau = \frac{Nc - Nd}{n(n-1)/2} \quad (2-1)$$

Nc désigne le nombre de paires de produits classés dans le même ordre entre le sujet considéré et la moyenne des autres sujets discriminants ; Nd est le nombre de paires discordantes. Le dénominateur correspond au nombre total de paires de produits (avec n, le nombre de produits testés). Lorsque toutes les paires de produits concordent entre le sujet considéré et les autres panélistes, les produits sont dans le même ordre et le coefficient de Kendall est égal à 1. Ce coefficient diminue avec le nombre de paires discordantes pour atteindre -1 lorsque l'ordre des produits est totalement inversé. Le test unilatéral de ce coefficient ($H0 : \tau = 0 / H1 : \tau > 0$), réalisé avec un risque α fixé à 0,20, indique si le sujet est en accord avec les autres sujets discriminants.

Le diagnostic final est alors :

- "+" si le sujet est discriminant et en accord
- "-" si le sujet est discriminant et en désaccord
- "=" si le sujet n'est pas discriminant.

Dans chaque cas, un dernier calcul évalue le niveau de répétabilité du sujet par rapport au groupe à l'aide d'un test de contraste dans lequel les observations sont les écart-types des répétitions. Le modèle d'ANOVA à deux facteurs (sujet et produit) avec interaction est utilisé. Lorsque le panéliste est significativement moins répétable que le groupe (risque $\alpha = 0.01$), le signe "!" est ajouté au précédent diagnostic. La signification de chaque diagnostic est rappelée dans le tableau 2-3.

Tableau 2-3 : Signification des diagnostics de la table CAP

Panéliste DISCRIMINANT ¹ ?	Panéliste en ACCORD ² ?	Panéliste REPETABLE ³ ?	diagnostic
OUI	OUI	OUI	+
		NON	+!
	NON	OUI	-
		NON	-!
NON	non testé	OUI	=
		NON	=!

1 : F produit de l'ANOVA individuelle score = produit + erreur ($\alpha=0,10$).

2 : coefficient de Kendall entre le sujet et le reste du groupe (observations = produits)

3 : test de contraste entre le sujet et la moyenne du groupe où les observations sont les écart-types des répétitions de chaque produit

Sur le table CAP (tableau 2-1), ces diagnostics mettent par exemple en évidence que la plupart des sujets sont discriminants et en accord pour le descripteur FIRM (12 diagnostics "+" sur 14 panélistes). Ces diagnostics montrent également que le sujet 13, positionné en dernière place dans le tableau, est discriminant pour quatre descripteurs seulement (COMPACT, HOMOGEN, ASTRING et STRAW).

Afin que cet outil puisse être utilisé dans les cas les plus divers, un autre type de diagnostic a été ajouté pour les études dans lesquelles certains sujets n'ont pas effectué de répétition. Dans ce cas, il n'est pas possible d'évaluer le pouvoir discriminant et la répétabilité de ce sujet. Le diagnostic sera alors "+?" si le sujet est en accord avec le groupe d'après le calcul du coefficient de Kendall et "-?" si le sujet est en désaccord avec le groupe. Dans de très rares cas, le diagnostic sera "=?" lorsque les notes de tous les produits seront identiques. Les sujets n'ayant pas effectué de répétitions seront systématiquement classés à la fin du tableau.

La table CAP permet de visualiser dans un seul tableau les performances unidimensionnelles issues des analyses de la variance au niveau du groupe et de chaque sujet. Cette synthèse permet de détecter rapidement les descripteurs pour lesquels le groupe est discriminant (FPROD sur fond vert), en accord (FINTER sur fond vert) et répétable (valeurs faibles du RMSE comparées à celles des autres descripteurs). Le classement des sujets et les diagnostics individuels donnent ensuite une information détaillée sur le niveau de performance de chaque sujet pour chaque descripteur.

2.2 La CAPMAP

Cette méthodologie a été développée au cours de ce travail de thèse afin de proposer une représentation graphique des informations contenues dans la table CAP. Elle permet de positionner les sujets dans l'espace des descripteurs en fonction de leur niveau de discrimination. De plus, cette carte permet, en cliquant sur chaque sujet ou descripteur, d'accéder à un diagnostic détaillé qui reprend les indices de la table CAP.

2.2.1 Construction et lecture de la carte

L'analyse sous-jacente est une ACP de la matrice $X_{(j,p)}$ de dimension sujet*descripteur dont les observations sont les racines carrées des F individuels de l'effet produit. L'observation x_{jp} est donc d'autant plus élevée que le niveau de discrimination du sujet j est élevé pour le descripteur p. La transformation par la racine carrée est appliquée à la statistique de Fisher (F) étant donné la nature de cette dernière. En effet, le F correspond à un rapport de carrés moyens, i.e. un rapport de deux valeurs qui représentent des distances au carré. Afin d'utiliser un indice qui reflète d'avantage un rapport de distances qu'un rapport de distances au carré, il nous a donc semblé préférable d'utiliser la racine du F.

Des informations supplémentaires sur la performance du panel et des panélistes sont ensuite ajoutées sur la carte obtenue (figure 2-2). Bien que l'analyse réalisée soit multidimensionnelle, elle utilise majoritairement les résultats de traitements unidimensionnels de la performance. De plus, en cliquant sur les éléments de la carte, l'utilisateur a accès à des tableaux individuels de résultats unidimensionnels. Elle est de ce fait considérée comme une analyse juxtaposant des résultats unidimensionnels plutôt que comme une réelle analyse multidimensionnelle. Cette classification reste discutable. La lecture de la CAPMAP s'effectue de la manière suivante :

La longueur des vecteurs représente l'amplitude de la variabilité inter-individuelle des niveaux de discrimination. Elle traduit donc l'hétérogénéité du pouvoir discriminant dans le groupe. La figure 2-2 montre par exemple que les performances en terme de discrimination sont très différentes d'un sujet à l'autre pour les descripteurs BUTTER et PASTY (vecteurs longs), alors que les niveaux de discrimination sont très similaires pour le descripteur SWEET (vecteur court). Il est important de noter que la longueur du vecteur ne traduit en rien le niveau global de discrimination du groupe.

Cette information est donnée par **la taille de la pointe du vecteur**, qui dépend directement de la valeur de FPROD (i.e. la statistique de l'effet produit dans le modèle d'ANOVA au niveau du groupe, cf. table CAP). La taille maximale est atteinte lorsque le risque de première espèce associé au FPROD est inférieur à 0,001. La figure montre par exemple que le panel est moins discriminant pour le descripteur BUTTER que pour le descripteur PASTY.

La couleur du vecteur représente l'accord entre les sujets du panel. Elle dépend à la fois des valeurs de FPROD et de FINTER. Lorsque FPROD est significatif (valeur sur fond vert), le panel perçoit des différences communes entre les produits malgré un éventuel désaccord (FINTER significatif). Du point de vue de l'animateur de panel, on peut alors considérer que le panel est suffisamment en accord pour différencier les produits, quel que soit la valeur de FINTER. La flèche représentant un descripteur est donc verte si le panel est discriminant (FPROD sur fond vert). Lorsque le panel n'est pas

discriminant (FPROD sur fond orange ou rouge), le niveau de significativité de l'interaction donne une information sur la raison éventuelle de ce résultat. Si l'interaction n'est pas significative, le désaccord entre les panélistes est négligeable. Le désaccord n'est donc pas la cause du faible niveau de discrimination du panel. On peut alors supposer que le panel n'est pas discriminant parce que les produits ne sont pas différents pour ce descripteur. Celui-ci sera donc orange sur la CAPMAP pour signifier que le panel n'est pas discriminant mais que le niveau d'accord n'est pas en cause. En revanche, si l'interaction est significative (FINTER sur fond orange ou rouge), le désaccord entre les panélistes est une raison potentielle du faible niveau de discrimination du groupe. Le vecteur correspondant à ce descripteur sera rouge. Ainsi, le descripteur MUSHROOM est rouge sur la CAPMAP car le panel n'est pas discriminant et en désaccord pour ce descripteur, alors que TOASTED est orange puisque le panel n'est pas discriminant mais que les panélistes ne sont pas en désaccord. La majorité des vecteurs sont verts, indiquant que le panel est discriminant.

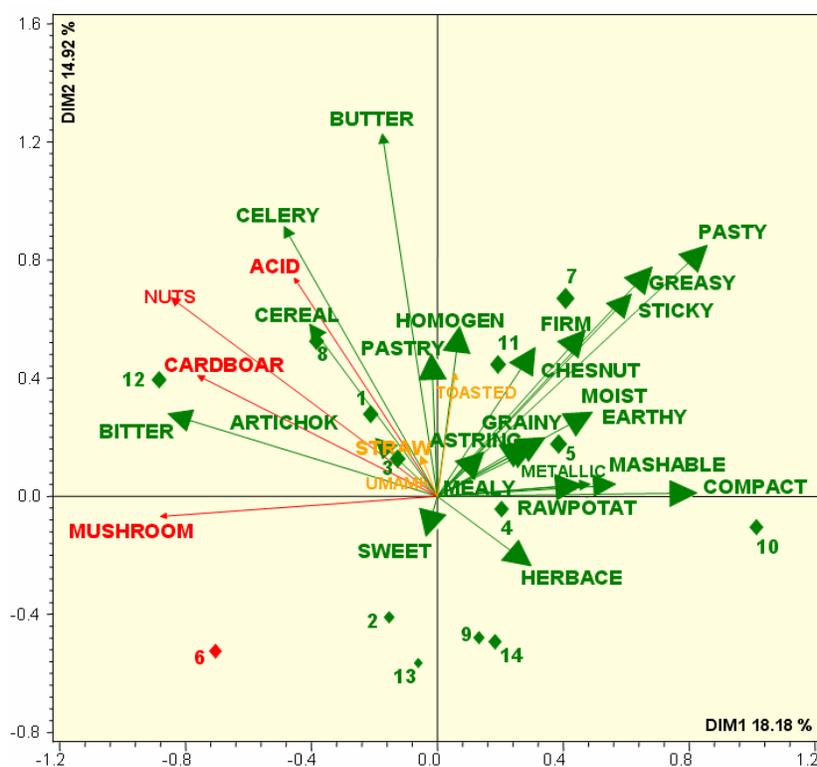


Figure 2-2 : Exemple de CAPMAP

La position des panélistes sur la carte dépend des niveaux de discrimination obtenus pour chaque descripteur. La projection d'un individu sur le vecteur d'un descripteur reflète son niveau de discrimination. Le sujet 10 est par exemple très discriminant sur le descripteur COMPACT par rapport aux autres panélistes, à l'inverse des sujets 6 et 12. Deux sujets sont proches lorsque leurs niveaux de discrimination sont semblables pour la plupart des descripteurs.

La taille du point est inversement proportionnelle à la valeur du RANKF du sujet. Elle traduit le niveau global de discrimination du panéliste pour tous les descripteurs. Dans l'exemple de la figure 2-2, le sujet 7 apparaît très discriminant par rapport au sujet 13, conformément aux résultats de la table CAP (tableau 2-1).

La couleur du point représente l'accord du panéliste avec le reste du groupe. Elle dépend de la valeur du coefficient ALPHA (différent du risque " α " couramment utilisé en statistiques) présenté dans le tableau "accord multidimensionnel" (paragraphe 3.1 de ce chapitre). Ce coefficient est d'autant plus élevé que l'accord entre le panéliste et le reste du groupe est important pour chaque descripteur de l'étude. Dans l'exemple présenté, l'accord avec le reste du groupe est satisfaisant pour tous les sujets à l'exception du juge 6 pour lequel l'accord multidimensionnel est plus faible.

2.2.2 Diagnostics des sujets et des descripteurs

La carte de la CAPMAP est interactive. Elle permet en effet d'accéder à des diagnostics détaillés de chaque descripteur et de chaque panéliste, en cliquant sur les items de la carte (symbole représentant chaque panéliste ou pointe de la flèche de chaque descripteur). Les diagnostics étant basés sur les mêmes types de calculs pour les descripteurs et les panélistes, seuls les résultats obtenus pour un panéliste ($n=7$) seront expliqués et commentés. Quelques précisions seront toutefois apportées dans le cas des descripteurs.

Le diagnostic de la figure 2-3 apparaît lorsque l'utilisateur clique sur le sujet 7 dans la carte CAPMAP.



Figure 2-3 : diagnostic individuel du juge 7 issu de la CAPMAP

Les trois boîtes à moustaches permettent de situer les niveaux de performance du sujet par rapport au groupe en terme de discrimination, de répétabilité et d'accord. Elles seront détaillées après le tableau de la figure 2-3. Ce dernier présente, pour chaque descripteur, la valeur des trois indices de performance qui permettent d'établir le diagnostic consigné dans la table CAP. Ce diagnostic est rappelé dans la seconde colonne du tableau.

La colonne Discrimination contient les valeurs de la statistique de Fisher de l'effet produit dans le modèle d'ANOVA individuel. La valeur est sur fond vert lorsque l'effet produit est significatif (risque associé inférieur à 0,10) ; le diagnostic correspondant sera alors "+" ou "-" selon le niveau d'accord. Si l'effet produit n'est pas significatif, la valeur de la statistique de Fisher est sur fond rouge et le diagnostic est "=" (tableau 2-3). Les descripteurs sont ordonnés selon cet indice afin de faciliter la lecture des résultats.

La colonne Répétabilité contient les valeurs "sudentisées" des écart-types entre les répétitions. Elles sont obtenues en réalisant, pour chaque descripteur, un test de contraste entre l'écart-type des répétitions d'un sujet et la moyenne du reste du panel. La valeur calculée est positive lorsque l'écart-type entre les répétitions d'un juge est supérieur à l'écart-type du groupe. Le sujet est donc moins répétable que le groupe. En fonction du risque fixé et du nombre de degrés de libertés, les valeurs positives élevées sont déclarées significativement supérieures à la moyenne du groupe. Le sujet est donc significativement moins répétable et la valeur correspondante est imprimée sur fond rouge dans le tableau de la figure 2-3. Le risque est fixé à 0.01 dans la table CAP afin de ne détecter que les panélistes dont le niveau de répétabilité est particulièrement moins satisfaisant que le reste du groupe. Le diagnostic inscrit dans la table CAP est alors accompagné du signe "!". Dans le cas contraire, le sujet n'est pas significativement moins répétable que le panel et la valeur de l'écart-type "sudentisé" est représentée sur fond vert ; le diagnostic n'est pas modifié.

La colonne Accord présente les valeurs du coefficient de corrélation de Kendall calculé entre les moyennes des notes de chaque produit pour le sujet considéré et les moyennes des notes du panel. Lorsque ce coefficient n'est pas significativement positif au risque de 0,20, le sujet est déclaré en désaccord avec le groupe et la valeur correspondante est écrite sur fond rouge dans le tableau de la figure 2-3. Le risque est fixé à 0,20 afin de ne détecter que les panélistes présentant un fort désaccord avec le groupe. Le choix du coefficient de corrélation de Kendall, plus conservateur que les autres coefficients de corrélation classiques (Pearson et Spearman) appuie cette volonté de ne détecter que les désaccords prononcés. Ce coefficient n'est calculé que lorsque le sujet est discriminant pour ce descripteur, considérant que le classement des produits d'un sujet non discriminant ne porte aucune information.

Les boîtes à moustaches de la figure 2-3 représentent chacune un des indices décrits ci-dessus (discrimination, répétabilité et accord). Elles représentent la dispersion des panélistes pour chaque indice de performance en tenant compte de tous les descripteurs de l'étude. En utilisant la même stratégie que pour le calcul du RANKF (figure 2-1), le rang moyen de chaque sujet est calculé pour chaque indice de performance. Cependant, dans ce calcul, le meilleur panéliste reçoit le rang le plus élevé, à l'inverse de la stratégie développée dans le calcul du RANKF, afin d'obtenir des scores

élevés pour les meilleurs juges. Notons que cette démarche conduit exactement au calcul de l'opposé du RANKF (par rapport au rang moyen) pour l'indice de discrimination. Chaque boîte à moustaches représente la dispersion des sujets pour un indice. La boîte centrale est bornée par les centiles 25% et 75% ; les moustaches peuvent atteindre 1,5 fois la taille de la boîte centrale ou s'arrêter à la dernière observation comprise dans cette zone. Le trait horizontal à l'intérieur de la boîte centrale représente la médiane. La petite croix rouge correspond à la moyenne. Par construction, la moyenne est identique pour les trois indices et correspond à la moyenne des rangs. En revanche, l'étendue des boîtes à moustaches est propre à chaque indice et reflète l'hétérogénéité du groupe.

La transformation de chaque indice de performance en rang permet de comparer l'étendue des boîtes à moustaches puisqu'elles utilisent la même échelle : le nombre de panélistes. L'exemple présenté dans la figure 2-3 montre ainsi une relative homogénéité des niveaux de discrimination, tous descripteurs confondus, (boîte centrale de faible amplitude) malgré quelques panélistes dont le niveau de discrimination s'éloigne de la moyenne du groupe (moustaches longues). A l'inverse, la boîte à moustaches de la répétabilité présente une hétérogénéité plus importante entre les panélistes (boîte centrale plus grande) mais moins de comportements extrêmes (moustaches courtes). Enfin, le niveau d'accord entre les panélistes est très homogène (boîte centrale petite et moustaches courtes)*.

La croix bleue représente la position du panéliste considéré par rapport au groupe. La figure 2-3 montre par exemple que le niveau de discrimination du sujet 7 est très élevé par rapport au groupe, contrairement au niveau de répétabilité de ce sujet qui se situe dans la partie basse de la boîte centrale correspondante. Le niveau d'accord se situe dans la moyenne du groupe.

Le diagnostic individuel présenté ci-dessus est également disponible pour chaque descripteur. Le même type de résultat est proposé en inversant le rôle des descripteurs avec celui des panélistes. Les lignes du tableau correspondent alors aux sujets, triés selon leur pouvoir discriminant. Les boîtes à moustaches représentent la dispersion des rangs moyens des descripteurs et la croix bleue positionne le descripteur sélectionné parmi l'ensemble des variables. Une boîte à moustaches présente une forte dispersion lorsque l'hétérogénéité des niveaux de performance des descripteurs est importante au niveau du groupe.

La CAPMAP est un outil de visualisation de l'hétérogénéité du panel en terme de discrimination des produits. De plus, la taille des éléments graphiques traduit le niveau de discrimination : la taille des symboles représentant les sujets traduit leur niveau de discrimination multidimensionnel, alors que la taille de la pointe des vecteurs représentant les descripteurs correspond au niveau de discrimination du groupe. Les couleurs ajoutées sur cette carte apportent une information sur l'accord de chaque panéliste avec le reste du groupe (couleur des juges) ainsi que sur le niveau d'accord du groupe pour chaque descripteur (couleur des descripteurs).

* : Rappelons que l'accord d'un panéliste avec le groupe n'est calculé que lorsque ce dernier est discriminant. Par conséquent, le nombre de sujets pris en compte dans le calcul des rangs des sujets selon le niveau d'accord peut varier d'un descripteur à l'autre. Cette baisse d'effectif par rapport aux deux autres indices de performance est en partie responsable de la faible dispersion des valeurs. De plus, cette variation influence le calcul des rangs de chaque panéliste et modifie la valeur du rang moyen. Afin de conserver un rang moyen identique aux autres boîtes à moustaches, nous avons choisi de centrer les rangs calculés pour chaque descripteur sur la moyenne des rangs obtenue pour les autres indices. Ce centrage permet de garder la même moyenne pour chaque indice.

Au cours du développement de cette méthode, la possibilité d'ajouter une information sur les niveaux de répétabilité a été envisagée en modifiant, par exemple, la forme des symboles représentant chaque panéliste en fonction du niveau de répétabilité (Schlich, Pineau et al., 2004). Cette proposition n'a finalement pas été retenue pour ne pas surcharger la représentation graphique.

Les tableaux de diagnostic individuel apportent une information très détaillée sur la performance de chaque sujet et de chaque descripteur. Ils peuvent s'avérer très utiles pour l'animateur de panel. En particulier, ces tableaux peuvent être distribués à chaque panéliste après les séances d'évaluation afin de leur permettre de se positionner par rapport au groupe et d'identifier les descripteurs pour lequel leur performance est moindre.

3 Performances multidimensionnelles

Les indices multidimensionnels procurent une vision plus globale de la performance des panélistes. Les notations utilisées dans les paragraphes suivants sont identiques à celles du premier chapitre : P désigne le nombre de descripteurs, I le nombre de produits et J le nombre de panélistes. Les lettres minuscules p, i et j désignent un élément de chaque ensemble.

3.1 Accord multidimensionnel

L'analyse de l'accord entre les panélistes au niveau multidimensionnel se base sur l'utilisation des méthodes STATIS, STATIS DUAL et GAMMA, présentées dans le paragraphe 4.4 du chapitre 1. Chaque méthode permet de calculer un indice d'accord de chaque sujet avec le groupe en se basant sur des approches différentes (annexe 1).

La méthode STATIS se base sur l'utilisation du coefficient RV. Celui-ci permet de mesurer la similarité des distances entre les produits de deux configurations individuelles (ou d'une configuration individuelle et de la configuration moyenne). La stratégie classique consiste ensuite à définir un compromis des configurations individuelles qui tienne compte de l'accord de chaque panéliste avec le groupe. La matrice contenant les coefficients RV de toutes les paires de sujets est alors utilisée pour déterminer les poids affectés à chaque panéliste. Ces poids reflètent le niveau d'accord de chaque sujet avec le groupe. Le compromis est obtenu en réalisant la moyenne des configurations individuelles pondérées par l'accord de chaque individu avec le groupe. Les sujets dont la configuration des produits s'éloigne de celle du groupe ont alors moins de poids dans la définition du compromis. Cette pondération est légitime lorsque l'objectif est d'obtenir une carte des produits qui s'affranchisse des comportements extrêmes. En revanche, dans l'optique de l'analyse des performances du panel, il semble préférable d'opter pour une stratégie qui accorde le même poids à chaque panéliste. De plus, notre objectif est uniquement de calculer un indice d'accord de chaque sujet avec le reste du groupe et non de produire une carte des différences entre produits. Pour chaque sujet, l'indice retenu dans notre cas sera donc simplement la moyenne des J-1 coefficients RV, calculés entre le sujet considéré et chaque autre panéliste :

$$RV_j = \frac{1}{J-1} \sum_{\substack{j' \\ (j' \neq j)}} RV_{jj'} \quad (2-2)$$

Le coefficient RV_j , calculé pour chaque panéliste, donne alors une indication sur le niveau d'accord de chaque sujet avec le groupe concernant la configuration des produits dans l'espace des descripteurs. De manière similaire, les indices d'accord RVD et ALPHA, issus respectivement des méthodes STATIS DUAL et GAMMA, sont utilisés pour obtenir des coefficients d'accord de chacun des sujets :

$$RVD_j = \frac{1}{J-1} \sum_{\substack{j' \\ (j' \neq j)}} RVD_{jj'} \quad (2-3)$$

$$ALPHA_j = \frac{1}{J-1} \sum_{\substack{j' \\ (j' \neq j)}} ALPHA_{jj'} \quad (2-4)$$

Rappelons que le coefficient RVD (RV DUAL) permet de mesurer la similarité des corrélations inter-descripteurs entre deux panélistes (ou un panéliste et le groupe). Le coefficient ALPHA mesure directement la corrélation entre les scores donnés par deux sujets pour les descripteurs homologues. Ces trois indices d'accord (RV, RVD et ALPHA) peuvent ensuite être « normalisés » (cf. paragraphe 4.4, chapitre 1) pour donner les coefficients NRV, NRVD et NALPHA. Ces indices peuvent alors être comparés aux valeurs obtenues d'après la distribution de la loi Normale centrée réduite. Un coefficient supérieur à 1,65 indique que l'accord mesuré pour un panéliste est significativement supérieur au hasard (test unilatéral), au risque $\alpha = 0.05$ de se tromper.

Le tableau de résultats proposé dans le cadre de la sélection de méthodes d'analyses de la performance est présenté ci-dessous :

Tableau 2-4 : Accord multidimensionnel

Subject	NRV	Nalpha	NRVD
1	2.68	3.45	3.59
10	2.92	3.73	5.96
11	1.94	3.18	5.41
12	2.57	3.15	5.49
13	1.79	2.81	6.32
14	3.32	4.02	4.00
2	3.75	4.34	6.35
3	3.23	3.74	7.23
4	3.17	3.74	6.16
5	3.20	3.88	7.55
6	0.97	0.91	4.29
7	3.22	4.24	6.50
8	2.36	2.92	5.42
9	2.06	3.21	5.03
Mean	2.66	3.49	5.66

Pour chaque sujet, les coefficients normalisés NRV, NALPHA et NRVD sont présentés. Lorsque leur valeur est supérieure à 2 (approximation de 1,96 ; le risque α est donc voisin de 0.025 pour un test unilatéral), l'accord entre le panéliste considéré et le groupe est déclaré significatif ; la valeur correspondante est écrite sur fond vert. Lorsque le coefficient est compris entre 1,5 et 2, l'accord du

panéliste est moins significatif et la valeur correspondante est écrite sur fond orange. En dessous de 1,5, la valeur est écrite sur fond rouge, considérant que le sujet n'est pas significativement en accord avec le groupe. L'exemple proposé dans le tableau 2-4 indique que la majorité des sujets sont en accord concernant les distances entre les produits de l'étude (colonne NRV). Seule la configuration du sujet 6 semble différente du reste du groupe (NRV=0,97). Le niveau d'accord des sujets 11 et 13 par rapport au groupe est également plus faible que celui des autres panélistes (valeurs sur fond orange). La colonne NRVD ne contient que des valeurs supérieures à 2 donc tous les sujets sont en accord concernant les corrélations entre descripteurs. Le résultat est quasiment identique concernant l'indice NALPHA. Seul le panéliste 6 ne s'accorde pas avec les autres juges concernant le classement des produits en tenant compte de tous les descripteurs.

Pour des raisons purement calculatoires, les indices normalisés ne peuvent pas être calculés à l'aide des formules analytiques lorsque les données contiennent moins de quatre produits. En effet, les valeurs $n-1$, $n-2$ et $n-3$ (n étant le nombre de produits) interviennent dans le dénominateur de certaines équations utilisées, provoquant une division par zéro qui rend le calcul impossible. La seconde méthode de calcul des indices normalisés nous semble également peu fiable lorsque le nombre de produits est réduit. Celle-ci repose sur la génération d'un grand nombre de données simulées obtenues par permutation des produits dans les tableaux individuels. Cependant, si les données ne comptent que 4 produits, il n'existe que $4! = 24$ permutations possibles. La distribution des coefficients sous l'hypothèse nulle repose alors sur peu de valeurs (seulement 24). Il nous semble donc que les indices normalisés NRV et NALPHA ne devraient pas être utilisés lorsque les données contiennent moins de cinq produits. De même, l'indice NRVD ne devrait pas être utilisé avec moins de cinq descripteurs, bien que le calcul de ce coefficient repose sur des permutations physiques et non sur des formules analytiques (puisque le nombre de possibilités de permutations physiques est tout aussi restreint). Lorsque ces situations se produisent, les indices proposés dans le tableau de l'accord multidimensionnel seront remplacés par les coefficients RV, ALPHA et/ou RVD.

3.2 Discrimination multidimensionnelle

Trois indices de performance sont calculés pour chaque sujet afin de caractériser la discrimination multidimensionnelle.

NATTSEL (Nombre d'ATTRibuts SElectionnés) est le résultat d'une étape de sélection de variables « pas à pas ». Cette méthode itérative, basée sur l'utilisation de la procédure STEPDISC de SAS®, recherche successivement les variables qui contribuent le plus à l'augmentation de la variance expliquée par l'effet produit. A chaque étape, l'algorithme ajoute le descripteur qui augmente le plus la variance expliquée tout en apportant l'information la moins redondante par rapport à celle des descripteurs déjà sélectionnés. De cette manière, un ensemble de descripteurs très corrélés ne sera représenté que par un seul descripteur dans la sélection de variables. Au cours de cette procédure, un descripteur sélectionné à une étape peut être retiré à la suivante si le gain d'information porté par ce descripteur n'est plus significatif (seuil fixé à 0,15). Lorsque plus aucun descripteur n'apporte suffisamment d'information sur les différences entre produits (au seuil de 0,15), l'algorithme est arrêté. Le seuil choisi est relativement élevé afin de conserver un maximum de descripteur dans la sélection.

NATTSEL correspond au nombre de descripteurs sélectionnés par cet algorithme. Sa valeur est d'autant plus grande que le panéliste exprime des différences entre produits non corrélées pour un grand nombre de descripteurs. Dans le tableau 2-5, le nombre de descripteurs sélectionnés (parmi 30) varie entre 3 pour le sujet 9 et 11 pour les sujets 3, 7 et 14.

Tableau 2-5 : Discrimination multidimensionnelle

subject	NATTSEL	F	NCAN
12	9	27,15	2
14	11	20,32	4
2	10	11,51	4
3	11	10,84	3
10	7	8,25	4
7	11	8,14	4
1	9	7,57	4
6	7	6,28	3
13	5	4,92	3
5	5	4,71	4
9	3	4,69	2
8	7	4,24	4
4	6	3,50	3
11	6	3,34	4

L'indice appelé F correspond à l'approximation de la statistique de Fisher de l'effet produit dans le modèle de MANOVA individuel descripteur = produit + erreur. Ce calcul est réalisé en ne conservant que les descripteurs sélectionnés d'après l'indice précédent (NATTSEL). Le F de MANOVA traduit le pouvoir discriminant du sujet dans l'espace des descripteurs sélectionnés. Les sujets du tableau 2-5 sont classés par ordre décroissant de F. La valeur de cet indice varie de 27,15 à 3,34 dans l'exemple proposé.

L'indice NCAN, enfin, correspond au nombre de dimensions canoniques significatives dans l'espace des descripteurs retenus après sélection. Ces dimensions sont déterminées à partir du modèle de MANOVA en exécutant une analyse en variables canoniques (AVC : paragraphe 4.2, chapitre 1). Un indice NCAN égal à 1 signifierait que le sujet perçoit les mêmes différences entre produits, quel que soit le descripteur. Un seul descripteur suffirait donc à exprimer l'ensemble des différences perçues par ce sujet. Lorsque la valeur de NCAN augmente, cela signifie que la complexité de la notation du sujet croît. Celui-ci a besoin d'un nombre plus important de dimensions pour exprimer toutes les différences entre produits. Dans le tableau 2-5, NCAN varie entre 2 et 4, indiquant que certains panélistes utilisent plus de dimensions de l'espace sensoriel que d'autres.

Les deux tableaux d'analyse multidimensionnelle de la performance, présentés dans le paragraphe 3, se focalisent respectivement sur l'accord entre les sujets et sur la discrimination des produits. La sélection de méthodes ne met pas en avant de méthode d'analyse multidimensionnelle de la répétabilité parce que la majorité des analyses multivariées nécessite que les matrices de données soient complètes. Or il est assez fréquent en analyse sensorielle que tous les sujets n'aient pas effectué le même nombre de répétitions de chaque produit, générant ainsi des données incomplètes

pour certains sujets. Cependant, lorsque les données sont parfaitement équilibrées quant au nombre de répétitions par sujet et par produit, un tableau supplémentaire est généré par l'outil statistique. Pour chaque sujet, l'analyse réalisée correspond au calcul de la moyenne des coefficients RV entre les répétitions. Cet indice est d'autant plus élevé que les distances entre produits (dans l'espace des descripteurs) sont similaires d'une répétition à l'autre. Le tableau de résultats propose alors, pour chaque sujet, la valeur du coefficient RV moyen et la valeur du coefficient RV normalisé.

4 Analyses des différences entre produits

L'objectif principal des deux méthodes présentées ici n'est pas de caractériser la performance du panel et de ses panélistes mais d'étudier les différences entre produits. Elles ont malgré tout été ajoutées afin de construire un outil statistique qui couvre l'ensemble des besoins de l'animateur de panel.

4.1 L'Analyse en Variables Canoniques (AVC)

4.1.1 Construction et interprétation

L'Analyse en Variables Canoniques (AVC) consiste à rechercher les dimensions qui maximisent les différences entre produits (comme l'ACP) en tenant compte de la variabilité des notes des panélistes (Schlich, 2004). Cette méthode se base sur l'utilisation d'un modèle d'analyse de la variance multidimensionnel (MANOVA). Le modèle retenu dans le cadre de cette sélection de méthodes est le suivant :

$$\text{Descripteur} = \text{produit} + \text{sujet} + \text{erreur} \quad (2-5)$$

Lorsque des répétitions des mesures ont été effectuées, elles sont moyennées au préalable. Ce modèle a été retenu afin de s'adapter au plus grand nombre d'études sensorielles, y compris les profils déséquilibrés ou réalisés sans répétitions. Dans les analyses de la variance multidimensionnelles, la variabilité liée à chaque facteur ne s'exprime plus en terme de carré moyen (cas de l'ANOVA) mais en terme de matrice de variance covariance de chaque effet. La variabilité de l'effet produit s'exprime par la matrice de variance covariance $B=X'X$, où $X_{(I \times P)}$ est le tableau des moyennes des produits pour chaque descripteur. Le tableau X est celui utilisé dans l'ACP. La variabilité de l'effet sujet s'exprime dans la matrice $S=A'A$, où $A_{(J \times P)}$ est la matrice des moyennes des sujets pour chaque descripteur. La variabilité résiduelle, enfin, s'exprime dans la matrice $W=Y'Y$, où $Y_{((I \times J) \times P)}$ est une matrice contenant $I \times J$ lignes (i.e. le nombre de sujets multiplié par le nombre de produits) et P colonnes (nombre de descripteurs). Les observations, $y_{i \times j, p}$, sont les moyennes des notes par couple produit*sujet pour chaque descripteur, centrées par sujet et par produit. Ce centrage des données de la matrice Y permet de ne conserver que la variabilité non expliquée par les effets produit et sujet. Dans le cadre du modèle de MANOVA utilisé, la matrice de variance covariance W (associée à Y) peut s'interpréter comme la matrice représentant le désaccord entre les panélistes.

L'analyse au niveau unidimensionnel se poursuit par le calcul de la statistique de Fisher de chaque effet. Le pouvoir discriminant du groupe est mesuré par le rapport du carré moyen de l'effet produit divisé par le carré moyen de l'erreur : $F = CM_{\text{produit}}/CM_{\text{erreur}}$. De manière analogue, l'effet produit au

niveau multidimensionnel est mesuré par la matrice BW^{-1} . La multiplication de B par l'inverse de W peut être comparée à une pondération des différences entre produits (B) par l'inverse du désaccord entre les panélistes (W^{-1}), afin de donner plus d'importance aux descripteurs pour lesquels les sujets sont en accord. L'AVC correspond à la diagonalisation de la matrice BW^{-1} et la représentation graphique associée (figure 2-4) permet de positionner les produits dans l'espace des premiers axes factoriels. En comparaison avec l'ACP, qui utilise seulement les données de la moyenne du groupe, cette méthodologie permet de tenir compte de l'hétérogénéité de la notation des sujets.

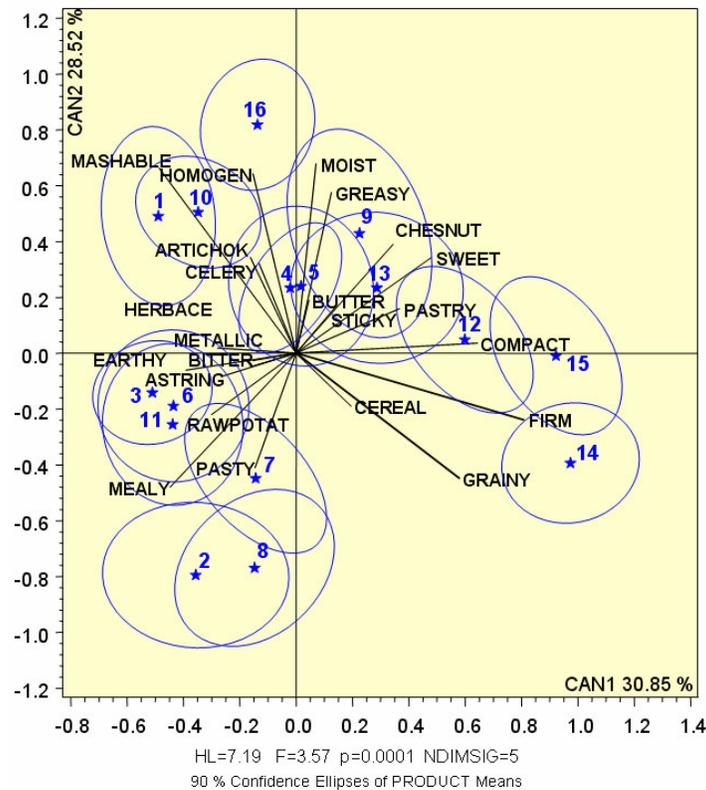


Figure 2-4 : plan factoriel de l'analyse en variable canonique (AVC)

La représentation graphique fournit également les ellipses de confiance (à 90%) autour de la moyenne de chaque produit. Elles sont calculées en utilisant l'équation (1-16) décrite dans le chapitre 1. Deux ellipses qui ne se chevauchent pas indiquent que les moyennes des produits concernés sont significativement différentes au niveau du groupe sur le premier plan factoriel. Sur la figure 2-4, le panel perçoit des différences entre les produits 12 et 14, à droite du graphique, mais le produit 15 n'est pas significativement différent de ces deux produits.

Ce graphique est accompagné de plusieurs informations multidimensionnelles. La trace de Hotelling-Lawley (HL), déjà évoquée dans le paragraphe 4.2 du chapitre précédent, correspond à la somme des valeurs propres des deux premières dimensions. L'approximation de la statistique de Fisher correspondante, F, et la probabilité associée, p, donne une indication sur la significativité multidimensionnelle des différences entre produits dans le plan. La valeur obtenue sur la figure 2-4 ($F=3,57$) indique que les différences entre produits sont significatives ($p=0,0001$). Le dernier indice proposé, noté NDIMSIG, est un indicateur de discrimination multidimensionnel. Ce test consiste tout

d'abord à projeter les valeurs des observations individuelles sur chaque axe. Une ANOVA à un facteur produit est ensuite réalisée pour chaque axe (les données sont centrées par sujet). NDIMSIG correspond alors au nombre d'axe pour lesquels l'effet produit est significatif. L'exemple présenté ci-dessus indique cinq axes pour lesquels des différences sont perçues (NDIMSIG=5). Cependant, d'après notre expérience, ce calcul a tendance à surestimer le nombre de dimensions significatives. Cet indice ne doit donc être utilisé que de manière comparative entre plusieurs AVC.

4.1.2 Sélection du nombre de descripteurs

D'un point de vue purement calculatoire, cette méthode est limitée par l'inversion de la matrice $W=Y'Y$. En effet, ce calcul est réalisable seulement si W est de rang plein. En pratique, le nombre descripteurs est alors limité à $(I-1)*(J-1)$, où I et J sont respectivement les nombres de produits et de juges. Lorsque le nombre de descripteurs est supérieur à cette limite, une sélection de variables en deux étapes est réalisée.

- étape 1 : les descripteurs dont l'effet produit (unidimensionnel) est non significatif au niveau du groupe sont exclus de l'analyse
- étape 2 : si la première étape ne permet pas de réduire suffisamment le nombre de variables, une sélection de variables « pas à pas » est réalisée. Cette méthode itérative, basée sur l'utilisation de la procédure STEPDISC de SAS, recherche successivement les variables qui contribuent le plus à l'augmentation de la variance expliquée, tout en apportant une information non redondante. Cette sélection aboutit à une combinaison de variables qui explique le maximum de variabilité avec un minimum de descripteur (cf. paragraphe 3.2, chapitre 2).

Ces deux étapes permettent d'accroître le nombre de cas pour lesquels l'AVC peut être utilisée. Les études comportant un nombre de descripteurs très important, ou très peu de produits, ne peuvent cependant pas être analysées avec l'AVC.

L'AVC a l'avantage de tenir compte de la variabilité des sujets pour définir les axes principaux, ce que ne permet pas l'ACP. Les représentations graphiques de ces deux méthodes peuvent en revanche proposer des ellipses de confiance autour des produits (Aubry, Schlich et al., 1999; Monrozier et Danzart, 2001; Husson, Bocquet et al., 2004). Dans le cadre de la construction d'un outil d'analyse de la performance, les informations données par le F de MANOVA et par le nombre de dimensions significatives permettent à l'animateur de panel de comparer rapidement le niveau de discrimination obtenu entre plusieurs études.

4.2 La table FLASH

La table Flash a été développée par Schlich en 1995 afin de synthétiser dans un unique tableau les différences entre produits observées pour chaque descripteur. En outre, les descripteurs sont classifiés en groupes de variables pour lesquels les différences entre produits sont similaires (Schlich et Degoud, 1995; Schlich, 1998; Porcherot et Schlich, 2000). Par rapport à la version initiale, des couleurs ont simplement été ajoutées afin d'améliorer la lisibilité des résultats. Dans l'exemple illustré

par le tableau 2-6, l'analyse est réalisée en utilisant les descripteurs discriminants uniquement. Il est cependant possible de faire intervenir l'intégralité des descripteurs si l'utilisateur le souhaite.

Tableau 2-6 : table FLASH

ATTRIBUTE	F	PROB	GMEAN	16	3	10	1	6	2	8	5	11	4	13	9	7	12	15	14	
FIRM	17,17	0,000	37	19-	21-	21-	22-	22-	24-							49+	63+	69+	77+	
COMPACT	8,57	0,000	49		29-		40-	27-	39-	36-			59+					60+	73+	73+
GRAINY	11,57	0,000	27	6-	14-	12-	15-	18-		52+								39+	51+	46+
PASTRY	4,42	0,000	19				10-			10-						4-	28+	44+		
SWEET	9,49	0,000	42	54+	30-	56+	27-	32-		29-	33-			54+	56+	22-	51+	69+		
BUTTER	1,73	0,047	17								25+								25+	9-
CHESNUT	6,02	0,000	41					29-	21-		30-		53+	58+				63+	58+	
CEREAL	2,32	0,004	26				17-													44+
ARTICHOK	2,62	0,001	40			52+			23-											28-
CELERY	1,86	0,029	17			29+									24+					
HERBACE	3,23	0,000	31					20-			49+					42+		20-	13-	
EARTHY	5,78	0,000	30				40+	42+		41+	43+		21-			47+	19-	14-		
ASTRING	2,58	0,001	29													44+	18-			
BITTER	2,87	0,000	27						18-	36+				17-		46+				
RAWPOTAT	2,81	0,001	23			32+				38+										10-
METALLIC	1,77	0,040	15				23+	26+												
PASTY	3,67	0,000	35				22-	47+	56+									23-	24-	
MEALY	11,44	0,000	42	29-	60+			54+	78+	55+	31-	54+		32-				26-	27-	31-
STICKY	2,95	0,000	28	37+	16-						17-				47+					
MASHABLE	13,91	0,000	40	69+		58+	66+		29-	25-	51+						26-	27-	16-	11-
HOMOGEN	7,40	0,000	59	78+		72+	70+		42-	31-	69+		70+		69+					45-
GREASY	5,26	0,000	15	26+	7-			8-	2-	7-	26+	7-	25+		25+	8-				
MOIST	11,65	0,000	47	68+		63+	69+	38-	17-	29-	58+	35-				33-	56+			34-

ATTRIBUTE : descripteurs classés par groupe de variables corrélées, le premier descripteur de chaque groupe est le plus représentatif du groupe ; **F** : statistique de Fisher de l'effet produit pour le groupe ; **PROB** : probabilité associée au F ; **GMEAN** : moyenne générale pour le descripteur considéré sur tous les produits ; **autres colonnes** = produits classés par ordre croissant selon le premier descripteur ; **case verte** : moyenne du produit significativement inférieure à GMEAN ; **case jaune** : moyenne du produit significativement supérieure à GMEAN.

Les lignes de ce tableau correspondent aux descripteurs, classés en groupes de variables fortement corrélées. Cette classification est réalisée en utilisant l'algorithme VARCLUS de SAS® (option CENTROÏD). La première étape de cette procédure consiste à réaliser une ACP du tableau des moyennes des produits ($B_{(I \times P)}$ de dimension produit*descripteur) afin de constituer deux groupes de variables. Les descripteurs corrélés positivement à la première composante principale de l'ACP forment le premier groupe et les descripteurs corrélés négativement le second. Une ACP des variables de chaque groupe est alors réalisée. Le pourcentage de variance expliquée par la première composante de chaque ACP est d'autant plus grand que les corrélations entre les variables de ce groupe sont fortes. Cette variance expliquée est ensuite rapportée au nombre total de variables de l'étude (et non au nombre de variables du groupe uniquement). Le pourcentage de « variance totale » expliquée par chaque groupe permet de mesurer la qualité de la représentativité de chaque groupe par rapport à l'ensemble des variables de l'étude. A ce stade, l'algorithme autorise les variables à changer de groupe afin de maximiser le pourcentage de « variance totale » expliquée par les deux groupes. Lorsque la composition des deux groupes est stable, l'algorithme recherche alors la meilleure partition des variables en trois groupes. Cette nouvelle étape s'initialise en prenant le groupe dont le pourcentage de variance totale expliquée est la plus faible afin de le diviser en deux nouveaux groupes de variables. Le même processus de séparation des variables et d'optimisation des groupes est réitéré pour un nombre de groupes supérieur.

Le choix du nombre de groupes est déterminé en fonction du pourcentage de variance expliquée par les groupes par rapport à la variance totale. La valeur seuil est fixée à 50% dans le cadre de ce travail

(valeur également préconisée dans les publications qui utilisent cette méthode). Dans le tableau 2-6, cette procédure conduit à la constitution de huit groupes de variables.

Ces groupes sont ensuite ordonnés. Le premier est celui dont la variance expliquée est la plus forte. Les groupes suivants sont classés par ordre décroissant de corrélation avec le premier. La similarité de l'ordonnement des produits avec le premier groupe se dégrade donc au fil de ce classement. A l'intérieur de chaque groupe, le descripteur placé en première position présente la corrélation à la moyenne du groupe la plus élevée. Ce descripteur est donc théoriquement le plus représentatif du groupe de variables. Notons que cette sélection n'est pas liée au pouvoir discriminant des variables. Les descripteurs suivants sont classés par ordre de corrélation décroissante avec le premier afin de réunir les variables pour lesquelles la similitude de la notation des produits est maximale.

Dans la table FLASH, chaque descripteur est accompagné de la valeur de la statistique de Fisher de l'effet produit (F) dans le modèle d'ANOVA descripteur = produit + sujet + produit*sujet + erreur (en modèle mixte), ainsi que de la probabilité associée (PROB). Cette indication permet de garder à l'esprit la valeur du pouvoir discriminant de chaque variable. La valeur GMEAN correspond à la moyenne des notes données par le groupe, tous produits confondus. Cela permet de détecter les descripteurs dont le niveau global de notation est particulièrement faible (par exemple, $GMEAN_{GREASY}=15$ sur une échelle de 0 à 100). Cette valeur est également utilisée comme référence dans les tests de comparaison des moyennes de chaque produit, présentés ci-après.

Les colonnes suivantes de la table FLASH correspondent chacune à un des produits de l'étude. Elles contiennent les moyennes des notes données par le panel et sont triées par ordre croissant des valeurs observées pour le premier descripteur du premier groupe (descripteur FIRM dans l'exemple du tableau 2-6). Un test d'égalité est réalisé entre la moyenne générale (GMEAN) et la moyenne de chaque produit. Lorsque la valeur observée pour un produit est significativement inférieure (seuil fixé à 0.10) à la moyenne, elle est accompagnée du signe "-" et affichée sur fond vert. Lorsque la valeur observée est significativement supérieure à la moyenne, elle est accompagnée du signe "+" et affichée sur fond jaune. Les valeurs, dont l'écart à la moyenne n'est pas significatif, ne sont pas imprimées. Ce test de comparaison à la moyenne générale a été préféré aux tests de comparaisons multiples qui comparent toutes les paires de produits. La solution retenue donne une information moins détaillée sur les différences entre produits, mais elle présente l'avantage de générer moins de tests et de faciliter l'interprétation des résultats.

Puisque les variables du premier groupe (tableau 2-6) sont fortement corrélées entre elles, la majorité des valeurs inférieures à la moyenne se retrouvent à gauche du tableau et les valeurs supérieures à la moyenne se retrouvent à droite. Etant donné que les groupes de variables sont classés selon leur corrélation avec le premier groupe, cette dichotomie s'estompe puis s'inverse en bas du tableau. Par exemple, le produit 16 a obtenu les moyennes les plus faibles pour les descripteurs FIRM (moyenne=19) et GRAINY (moyenne=6) dans le premier groupe alors qu'il obtient des valeurs

significativement supérieures à celles du panel pour les quatre descripteurs du dernier groupe de variables.

La table FLASH permet d'obtenir une synthèse des différences significatives entre produits pour chaque descripteur. La classification des variables en groupes de descripteurs fortement corrélés permet en outre de caractériser plus facilement l'espace produit. Cet outil peut également être utilisé pendant l'entraînement afin de sélectionner les variables les plus représentatives de l'ensemble des descripteurs.

5 Conclusion

L'outil statistique défini par les six analyses présentées dans ce chapitre rassemble des résultats unidimensionnels et multidimensionnels afin de caractériser les performances d'un panel et de ses panélistes en termes de discrimination, de répétabilité et d'accord. Chaque analyse correspond à un programme automatisé en langage SAS®. Les programmes concernant la table CAP, la table FLASH, l'AVC et le tableau des indices de discrimination multidimensionnels ont été conçus à l'origine par Pascal Schlich au cours des dix dernières années. Ces programmes ont été repris, principalement afin d'améliorer la présentation des résultats à l'aide des outils informatiques récents (résultats en couleurs, utilisation de pages html). Quelques améliorations ont également été apportées au niveau des calculs statistiques afin de prendre en compte des situations plus diverses. Par exemple, les études comportant des sujets n'ayant pas effectué de répétitions peuvent être prises en compte dans la table CAP. Le tableau de résultats multidimensionnels sur l'accord entre les panélistes a été développé en parallèle avec ce travail de thèse par Delphine Brajon dans le cadre de son stage de fin d'études (Master en statistiques). La CAPMAP a été développée au cours de la thèse, ainsi que l'automatisation de la mise en forme des données brutes et de l'enchaînement des analyses jusqu'à la production finale des résultats sous la forme d'un ensemble de pages html agencées autour d'un menu (figure 2-5).

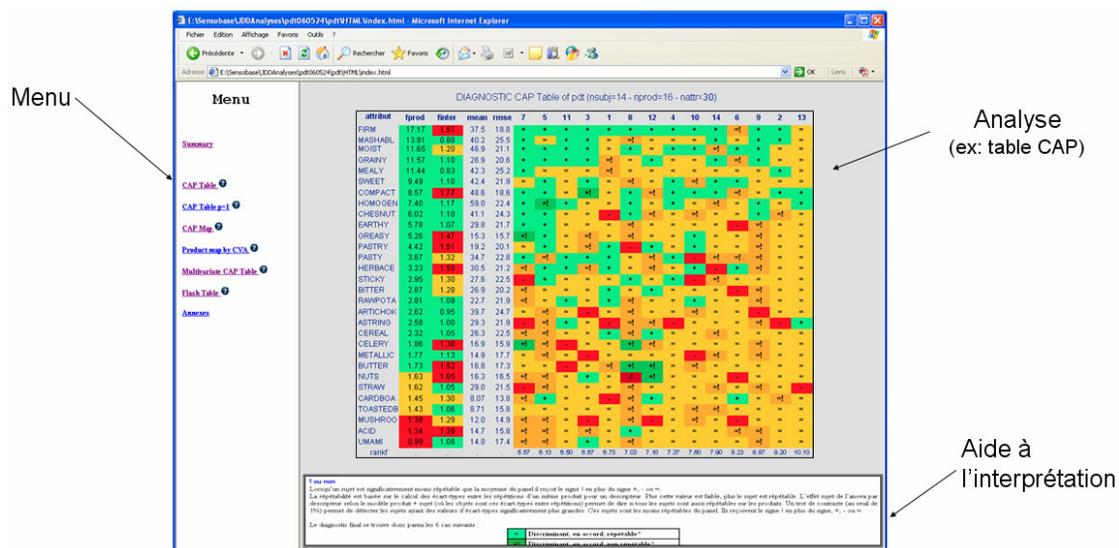


Figure 2-5 : Aperçu de la présentation des résultats issus du programme PERF

Cette restitution des analyses est accompagnée d'une notice explicative (aide à l'interprétation) destinée aux utilisateurs non statisticiens. Le texte intégral, rédigé au cours de ce travail de thèse, est disponible en annexe 2. L'ensemble de ce système est réuni dans un programme SAS® appelé PERF qui permet à l'utilisateur de régler les paramètres de chaque analyse (choix des analyses, risques des différents tests, options graphiques...). Une présentation du fonctionnement de ce programme est disponible en annexe 3.

En fonction des remarques des utilisateurs et des développements méthodologiques actuels au laboratoire, cet outil est amené à évoluer dans les années à venir afin de simplifier davantage la lecture de certains résultats ou apporter des informations complémentaires à d'autres analyses.

Partie II

**Construction de la SensoBase et exploitation
sensorielle et sensométrique**

De nombreuses méthodes statistiques permettent d'analyser la performance d'un panel lors d'une étude de profil sensoriel. Ces analyses sont précieuses car elles mettent en évidence les forces et les faiblesses du groupe, aussi bien que de chaque individu, pour les différentes facettes de la performance (répétabilité, discrimination et accord). Cependant, ces analyses, effectuées au niveau d'une seule étude, ne permettent pas de mettre en perspective les résultats obtenus par rapport aux performances généralement atteintes dans le domaine de l'analyse sensorielle. L'agrégation d'un grand nombre d'études permettrait de documenter les pratiques de l'analyse sensorielle, de définir des niveaux de performance moyens de mesurer l'influence de différentes variables sur les niveaux de performance. La mise en commun d'un grand nombre de données de profil sensoriel pourrait également permettre de comparer des méthodes statistiques, en observant si leur utilisation conduit à des résultats différents dans la pratique. La base de données de profils sensoriels créée dans ce but, appelée SensoBase, rassemble des données provenant de divers acteurs de l'analyse sensorielle dans le monde (industriels français et étrangers, laboratoires publics de recherche ...). Le principe de fonctionnement de cette base repose sur un échange entre données de profils sensoriels (envoyées par les fournisseurs de données) et analyse statistique de la performance (retournée par notre laboratoire, le LIRIS : Laboratoire d'Interface Recherche-Industrie et de Sensométrie). Certains fournisseurs envoient en outre des successions d'études sensorielles réalisées avec les mêmes panélistes au cours du temps. Cela donne alors la possibilité d'étudier l'évolution de la performance du panel au fil des évaluations.

La seconde partie de ce manuscrit envisage donc différentes situations dans lesquelles l'agrégation d'un nombre important d'études sensorielles permet de répondre à des questions nouvelles autour de la notion de performance. Ainsi, les trois derniers chapitres de cette partie aborderont respectivement :

- l'évolution de la performance d'un panel et de ses panélistes au fil des séances d'évaluation réalisées sur un même univers produit avec les mêmes descripteurs (chapitre 4) ;
- l'estimation des niveaux de performance atteints dans le domaine de l'analyse sensorielle et l'étude de plusieurs facteurs pouvant influencer ces niveaux de performances (chapitre 5) ;
- l'étude d'une méthode d'analyse statistique de la performance peu utilisée à l'heure actuelle ("Assessor model" de Brockhoff, développé en 1994) et l'impact de son utilisation sur les conclusions sensorielles (chapitre 6).

L'étude de ces thématiques est rendue possible par la construction et l'utilisation de la SensoBase, dont l'architecture et le système de fonctionnement sont présentés dans le chapitre 3.

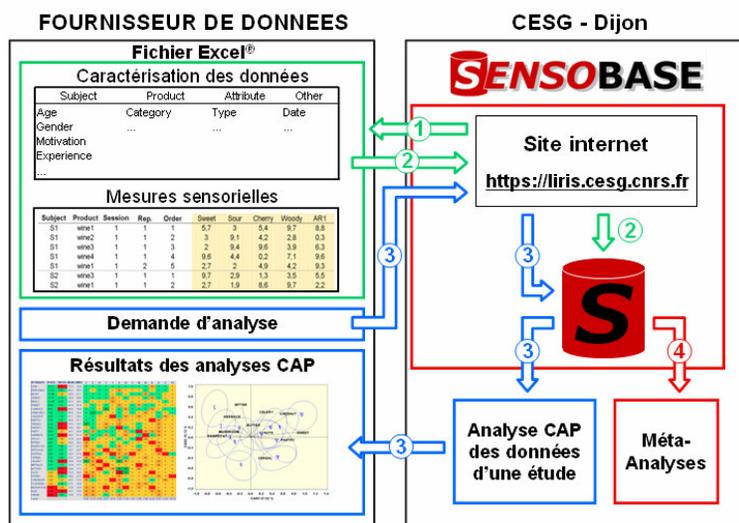
Chapitre 3 : Le système SensoBase

L'origine de la création de la SensoBase repose sur la volonté de concevoir un système permettant d'accumuler des données de profil sensoriel provenant de divers organismes pratiquant l'analyse sensorielle (français ou étrangers, privés ou publiques). En échange de l'envoi de chaque profil sensoriel, le fournisseur de données (i.e. un animateur de panel) recevra une analyse de la performance de son panel.

Le système créé doit donc être capable de réceptionner et de stocker dans une structure unique des études d'origine diverses, réalisées avec des protocoles différents, tout en garantissant la confidentialité des données apportées par chaque fournisseur. De plus, le système SensoBase doit permettre à chaque fournisseur (ou utilisateur) d'effectuer des demandes d'analyse de ses données. L'architecture de ce système a donc été pensée afin de répondre à ces objectifs.

1 Architecture générale

Le projet Sensobase s'articule autour d'un site Internet (<https://liris.cesg.cnrs.fr>) qui permet aux fournisseurs de données de s'inscrire à ce programme d'échange, d'envoyer leurs études de profil sensoriel, puis de demander des analyses statistiques. Afin de faciliter le stockage dans une structure unique, les données relatives à chaque étude doivent être saisies dans un fichier Excel[®] spécifiquement formaté afin de recevoir ce type de données. Ce fichier demande à l'utilisateur de fournir divers renseignements sur l'étude sensorielle (age et sexe des panélistes, type de descripteurs...) puis de saisir les données à proprement parler selon une structure précise. Le fichier est ensuite envoyé via le site Internet et stocké dans une base de données. Suite à l'insertion des données dans la base, l'utilisateur peut effectuer une demande d'analyse de ce jeu de données dans laquelle certains paramètres des analyses sont réglables. L'utilisateur reçoit ultérieurement les résultats par courrier électronique au format html. Un schéma du fonctionnement de ce système est présenté dans la figure 3-1.



- ① Téléchargement du fichier Excel[®] formaté pour accueillir des données SensoBase
- ② Envoi et insertion des données d'un profil sensoriel dans la SensoBase
- ③ Demande d'analyse sur un profil préalablement envoyé et réception des résultats
- ④ Méta-analyses sur l'ensemble des données rassemblées dans la SensoBase (évaluation des niveaux de performance, comparaison de méthodes sensorielles...)

Figure 3-1 : Schéma de fonctionnement de la SensoBase

Les trois organes essentiels de la SensoBase qui interagissent entre eux sont donc :

- le fichier Excel[®] permettant au fournisseur de saisir les données dans un format standardisé,
- la base de données qui recueille les données envoyées par les fournisseurs,
- le site Internet qui réalise l'interface entre les fournisseurs et notre laboratoire.

Chaque élément est détaillé dans un des paragraphes suivants.

2 Le fichier Excel[®] Sensobase

Le fichier Excel[®] permet de saisir les données relatives à la réalisation d'un profil descriptif. Ce format de fichier a été retenu car il est très largement répandu chez tous les acteurs de l'analyse sensorielle, contrairement à des formats spécifiques issus de logiciels dédiés à la gestion d'études sensorielles (FIZZ, Compusense...). Ce fichier se décompose en trois parties, correspondant à trois feuilles Excel[®]. Il est disponible en français et en anglais.

La première feuille donne les instructions nécessaires au renseignement de ce fichier (figure 3-2). Elle rappelle la définition d'un profil sensoriel au sens de la SensoBase, à savoir l'étude d'un ensemble de produits dégustés, de préférence plusieurs fois, par chaque panéliste. Un profil peut se dérouler sur plusieurs séances de mesures. Cette feuille donne également diverses informations pratiques et techniques et rappelle qu'une notice d'utilisation est disponible sur le site (cf. annexe 4)

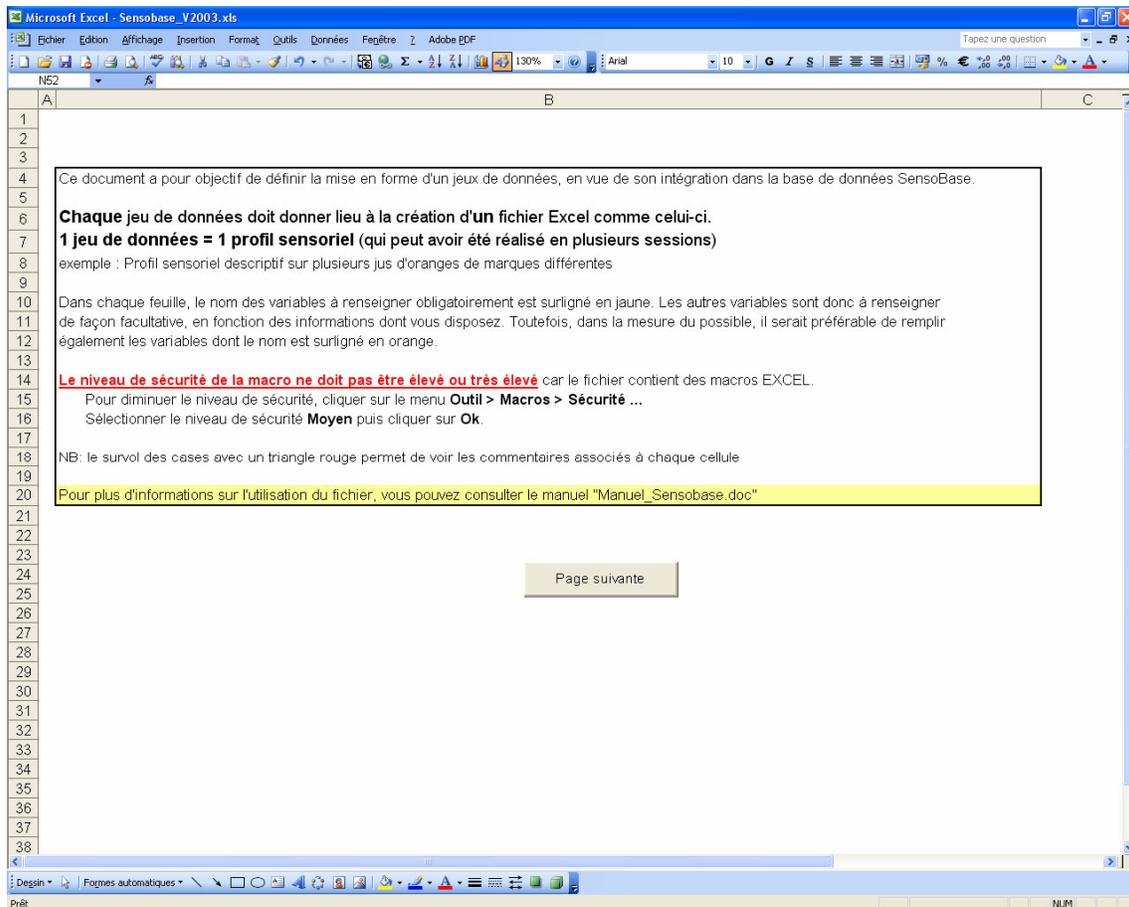


Figure 3-2 : Instructions générales pour renseigner le fichier Sensobase Excel2003® (feuille 1)

La seconde feuille est consacrée à la caractérisation des données de l'étude. Elle contiendra les renseignements sur les panélistes, les produits et les descripteurs de l'étude fournis par l'utilisateur (extrait présenté dans la figure 3-3). Les informations obligatoires ne concernent que l'âge et le sexe des sujets, la famille de produits (légumes frais, boissons alcoolisées, desserts lactés....) et le type de descripteur (saveur, arôme, texture...). L'utilisateur est également encouragé à fournir des informations supplémentaires, en particulier le niveau d'étude, le niveau d'expérience en analyse sensorielle et la motivation de la participation de chaque panéliste, le type de panel (interne/externe), la fréquence de consommation du produit de l'étude, le type d'échelle (continue, discrète) et le mode de gratification. Ces variables ont été spécifiquement identifiées parce qu'elle nous semblaient *a priori* pouvoir influencer le niveau de performance des panélistes. De plus, la collecte de ces informations est relativement simple pour l'utilisateur. D'autres variables peuvent également être renseignées si l'utilisateur en a la possibilité. La liste complète et la description de l'ensemble de ces items est disponible en annexe 5. Une explication détaillée du codage de chaque variable est accessible en arrêtant le curseur sur la cellule correspondante. La figure 3-3 montre, par exemple, la liste des types de descripteurs possibles et le codage numérique correspondant.

A la fin de la saisie des informations concernant la caractérisation des données, l'utilisateur doit cliquer sur un bouton pour accéder à la page suivante. L'activation de ce bouton entraîne également

l'exécution d'un ensemble de macros Excel® (langage Visual Basic pour Applications) qui vérifient l'intégrité des données. Cette étape assure que tous les champs obligatoires ont été correctement remplis et que le format des données saisies correspond au format attendu. Elle permet également à l'utilisateur de vérifier en partie qu'il n'a pas commis d'erreur de saisie en faisant apparaître un récapitulatif des informations saisies ou la liste des incohérences rencontrées, le cas échéant (figure 3-4).

	A	B	C	D	E	F	G	H	I	
1	DESCRIPTION DES DONNEES				Page Précédente	Vérifier la saisie et passer à l'étape suivante				
2	FOURNISSEUR									
3	Nom de la Société / login		CESG					obligatoire		
4								souhaitable		
5								facultatif		
6	SUJETS									
7	Code sujet			S1	S2	S3	S4	S5	S6	
8	Est-ce une Clé ?			0	0	0	0	0	0	0
9	Date de naissance			06/11/1920	21/10/1990	31/01/1920	31/01/1908	25/05/1982	25/05/1982	
10	Sexe		F	M	F	F	M	M		
11	Pays de naissance									
12	...									
19	Niveau d'études			2	3	3	2	3	1	
140	PRODUITS									
141	Code produit		AF1	AF2	KG	LX	YoSi			
142	Est-ce une Clé ?							0		
143	Famille d'aliments		121	121	121	121	121			
144	Produit du commerce ?									
145	...									
159	DESCRIPTEURS									
160	Code descripteur		Sweet	Bitter	AR2	Tx1	Tx2			
161	Est-ce une Clé ?							0		
162	Libellé du descripteur			Bitter	AR2	Tx1	Tx2			
163	Type de descripteur				3	5	4	4		
164	Valeur minimale de l'échelle				0	0	0	0		
165	Valeur maximale de l'échelle				10	10	10	10		
166	Type d'échelle				3	3	3	3		
167	Graduations									
168	Echelle discrète : précision									
169	Profil : définition du descripteur									
170	...									
176	PROTOCOLE DE L'ETUDE									
177	Date début étude			22/12/2006						
178	Date fin étude									
179	...									

Figure 3-3 : Extraits de la feuille de caractérisation des données du fichier SensoBase Excel2003 (feuille 2)

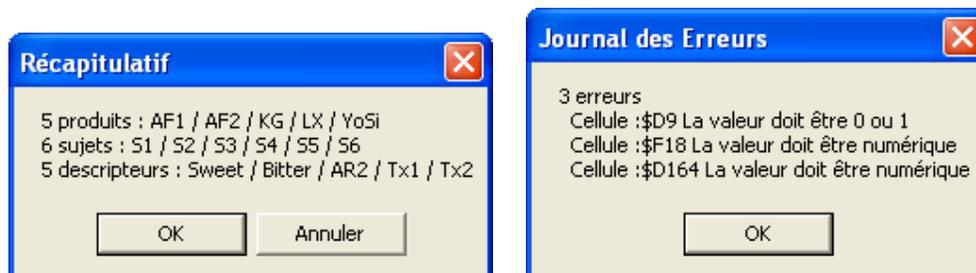


Figure 3-4 : Récapitulatif des données ou liste des erreurs affichée à la fin de la saisie des caractéristiques des données

Lorsque la vérification est terminée, l'utilisateur accède à la troisième feuille dans laquelle doivent être insérées les données de l'étude. La structure du tableau est illustrée dans la figure 3-5.

	A	B	C	D	E	F	G	H	I	J	K	L	M
	Code sujet	Code produit	Session	Répétition	Ordre	Sweet	Bitter	AR2	Tx1	Tx2			
1													
2	S1	AF1	1	1	1	9,1	0,2	6	2,7	3,6			
3	S1	LX	1	1	2	7,3	6,5	6,8	7,8	9,1			
4	S1	KG	1	1	3	3,1	0,7	0,3	9,6	6,1			
5	S1	AF2	1	1	4	6,5	8	9,4	5	6			
6	S1	YoSi	1	1	5	3,8	3,4	0,1	0,8	0,3			
7	S1	AF2	1	2	6	3,6	7,2	2,1	6,6	4,9			
8	S2	AF2	1	1	1			
9	S2	KG	1	1	2			
10	S2	LX	1	1	3			
11	S2	YoSi	1	1	4			
12	S2	AF1	1	1	5			
13	S2	LX	1	2	6			
14	S3	KG	1	1	1			
15	S3	YoSi	1	1	2			
16	S3	AF2	1	1	3			
17	S3	KG	1	2	4			
18	S3	LX	1	1	5			
19	S3	AF1	1	1	6			
20	S4	LX	1	1	1			
21	S4	AF1	1	1	2			
22	S4	YoSi	1	1	3			
23	S4	AF1	1	2	4			
24	S4	AF2	1	1	5			
25	S4	KG	1	1	6			
26	S5	YoSi	1	1	1			
27	S5	AF2	1	1	2			
28	S5	AF1	1	1	3			
29	S5	LX	1	1	4			
30	S5	KG	1	1	5			
31	S5	YoSi	1	2	6			
32	S1	AF2	2	1	1			
33	S1	KG	2	1	2			

Figure 3-5 : Structure des données dans le fichier SensoBase excel2003 (feuille 3)

Les première colonnes correspondent respectivement aux variables sujet, produit, session, répétition et ordre. Les colonnes suivantes correspondent chacune à un descripteur de l'étude. A la fin de la saisie, une étape de vérification des données est également effectuée. L'attention est en particulier portée sur la correspondance des noms des produits, sujets et descripteurs entre la feuille de caractérisation des données et la feuille contenant les données. L'utilisateur accède ensuite à un tableau croisé qui récapitule, pour chaque panéliste, le nombre de dégustations de chaque produit effectuées (figure 3-6). Une inspection rapide de ce tableau doit permettre à l'utilisateur de vérifier qu'il a bien fourni les données désirées.

	A	B	C	D	E	F	G
1							
2		Retour aux Données					
3							
4							
5							
6							
7	Nombre de Code sujet	Code produit					
8	Code sujet	AF1	AF2	KG	LX	YoSi	Total
9	S1	2	3	2	3	2	12
10	S2	2	2	2	3	3	12
11	S3	3	2	3	2	2	12
12	S4	3	2	3	2	2	12
13	S5	2	3	2	2	3	12
14	Total	12	12	12	12	12	60
15							
16							

Figure 3-6 : Tableau croisé du nombre de répétitions par sujet et par produit

Lorsque la procédure de saisie est terminée, un dernier bouton permet de déclencher une macro Excel[®] qui génère un fichier CSV (Comma Separated Value) contenant l'intégralité des données saisies. Contrairement à un fichier Excel[®], ce type de fichier ne contient que du texte. Il est donc facilement interprétable et utilisable par tous les systèmes d'exploitation, y compris les serveurs web. Ce passage d'une structure Excel[®] à une structure CSV va permettre l'insertion automatique des informations dans la base de données.

Pour des raisons techniques, le niveau de vérification maximal n'est garanti que pour la version 2003 d'Excel[®], les versions antérieures ne permettant pas de maîtriser, de manière aussi pointue, tous les paramètres de sécurité du fichier. Des fichiers SensoBase ont cependant été développés pour les versions 2000 et 2002 mais le niveau de vérification est moindre, ce qui ne permet pas de certifier que les données saisies sont intégralement conformes au type de données que peut recevoir la base. Le passage au fichier CSV n'est donc pas proposé pour ces versions et l'insertion des données dans la SensoBase sera en partie manuelle.

L'architecture initiale de ce fichier a été développée par Sylvie Cordelle, ingénieur d'étude du LIRIS, puis reprise pendant ce travail de thèse. Les programmes de vérification des données ont été développés par Delphine Brajon puis par Julien Delaborde, pendant leur période respective de travail et de stage, que nous avons encadrés au cours de ce travail de thèse.

3 La base de données

Le stockage des profils sensoriels envoyés par les fournisseurs s'effectue dans une base de données MySQL, gérée en utilisant le logiciel phpMyAdmin. Ce type de structure peut facilement être interrogée et modifiée en utilisant le langage PHP, langage dédié à la conception de sites Internet dynamiques. Par dynamique, nous entendons le fait que l'utilisateur puisse interagir avec le site en envoyant des informations (remplissage de formulaires, envoi de fichiers...) ou en sélectionnant des paramètres qui agissent sur le contenu des pages. Le langage php permet de formuler des requêtes en langage SQL afin d'interroger directement la base de données.

La base de données est constituée d'un ensemble de tables qui contiennent les informations envoyées par les fournisseurs. La structure de ces tables est définie de manière à minimiser la redondance de l'information. Ainsi, il est préférable de créer, par exemple, une table "produit" contenant toutes les informations relatives à chaque produit (nom, famille d'aliment à laquelle il appartient, quantité ingérée...) puis de faire référence à chaque produit à l'aide d'un code unique dans les autres tables, plutôt que de répéter la description du produit à chaque fois que celui-ci intervient dans une évaluation. Ce type de structure permet de minimiser l'espace occupé par la base de données et facilite les modifications qui pourraient être effectuées ultérieurement. En effet, si une caractéristique était erronée pour un produit et qu'il fallait la modifier, la mise à jour ne devrait être effectuée qu'à un seul endroit (dans la table "produit"), et non autant de fois que le produit est cité.

Le schéma structurel de la base, présenté dans la figure 3-7, montre que l'architecture de la SensoBase repose sur une table centrale, "DONNEES", contenant tous les scores des différents

profils sensoriels. Cette table est reliée à quatre tables contenant respectivement les caractéristiques des produits, des sujets, des descripteurs et les informations générales concernant chaque étude (date de réalisation, type de panel...). Cette dernière table (nommée "jdd", abréviation de "jeu de données") est en particulier reliée à la table "fournisseur", elle-même reliée à la table "contact" contenant les données sur l'animateur de panel du fournisseur en question. Les tables "protocole" et "etat_suj" ne sont pas présentées dans leur intégralité ; elles permettent de limiter la redondance de l'information.

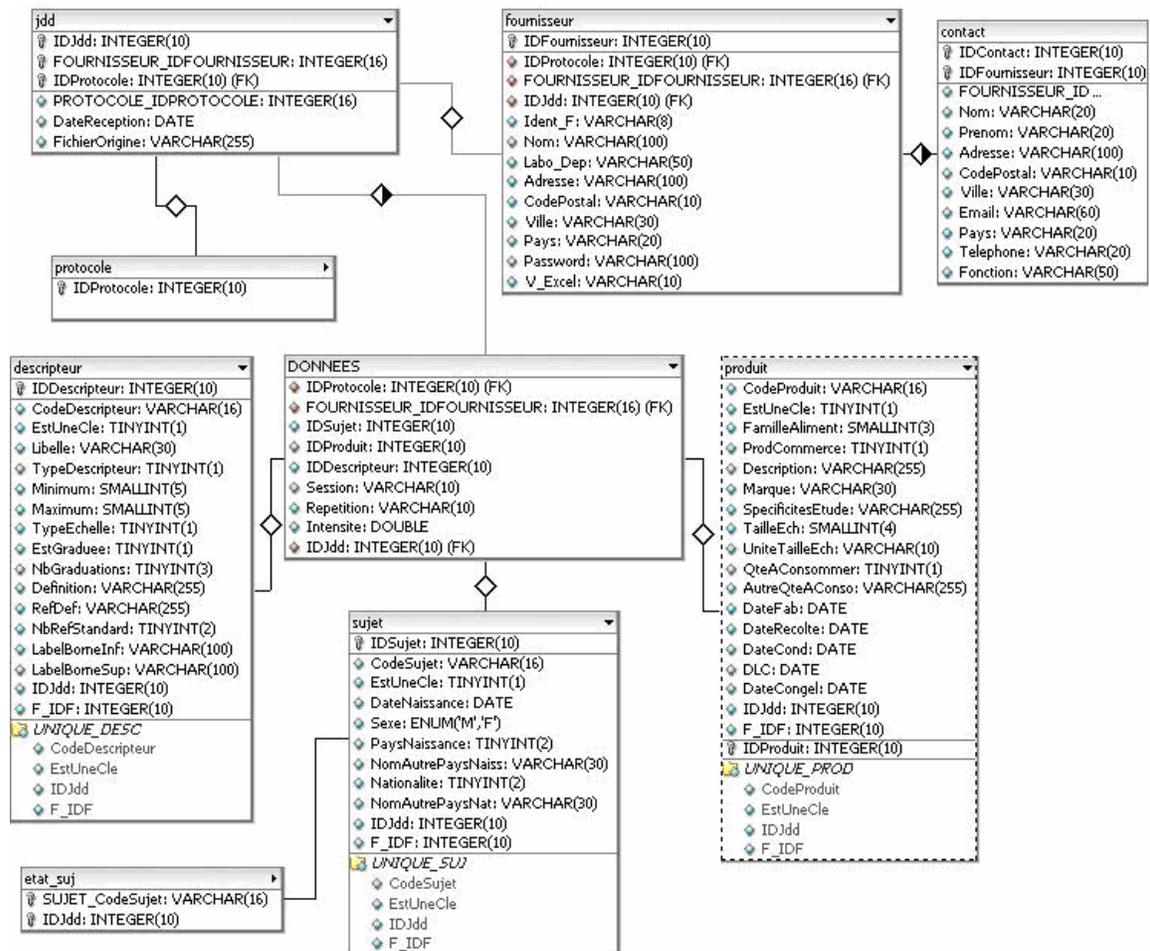


Figure 3-7 : Schéma structurelle de la base de données SensoBase

La base de données a été entièrement développée par Delphine Brajon. Seuls quelques éléments ont été modifiés depuis.

4 Le site Internet

Le site web de la SensoBase (<https://liris.cesg.cnrs>), dont la page d'accueil est présentée dans la figure 3-8, est l'interface dédié aux utilisateurs. Il a été majoritairement développé par Delphine Brajon puis mis à jour et partiellement modifié pendant ce travail de thèse. Ce site est constitué d'une partie accessible à tous les visiteurs et d'une partie réservée aux utilisateurs inscrits au programme SensoBase. La partie en libre accès présente le projet, le déroulement de l'envoi des données et la

restitution des analyses, un extrait de la notice explicative des résultats statistiques et le formulaire d'inscription (en cliquant sur les liens du menu, à gauche). Le site est disponible en français et en anglais.



Figure 3-8 : Page d'accueil du site de la SensoBase

La partie réservée aux personnes inscrites (figure 3-9) est accessible à l'aide d'un identifiant et d'un mot de passe fournis lors de l'inscription. L'utilisateur peut alors réaliser les tâches suivantes en accédant à son espace personnel :

- télécharger le fichier Excel[®] permettant d'envoyer des données de profil sensoriel en fonction de la version du logiciel Excel[®] de l'utilisateur (2000, 2002 ou 2003) ainsi que la notice explicative ;
- envoyer une étude sensorielle ;
- demander une analyse sur une étude déjà envoyée (ou sur une sélection des données de cette étude).

Ces deux dernières étapes constituent le cœur du programme SensoBase et seront détaillées ci-après.

Dans son espace personnel, l'utilisateur a également accès aux informations déjà saisies lors des précédentes études. Il peut ainsi afficher la liste des profils envoyés et les listes des sujets, produits et descripteurs avec leurs caractéristiques. Ces informations sont très utiles pour les fournisseurs qui

envoi des études contenant des éléments communs (des sujets qui participent à plusieurs études par exemple). Cela permet également de faire le lien entre les codes attribués aux études, générés par la base de données, et les noms explicites donnés par les fournisseurs.

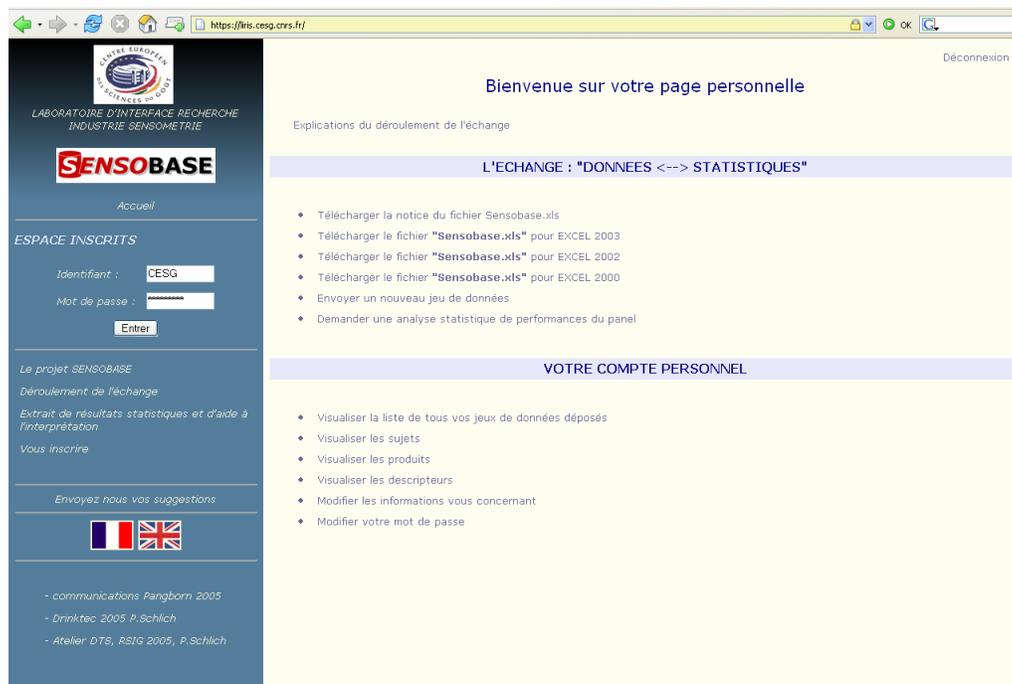


Figure 3-9 : Espace personnel accessible après inscription au projet Sensobase

4.1 Insertion d'une étude de profil sensoriel dans la base de données

Pour l'utilisateur, l'opération est simple. Il suffit d'envoyer le fichier CSV généré à partir du fichier Excel[®] (avec Excel[®] 2003 uniquement) via le site internet en accédant à la page "Envoyer un nouveau jeu de données". Le fichier est alors téléchargé sur le serveur de la Sensobase, puis un ensemble de procédures faisant appel aux informations de la base de données sont exécutées afin de vérifier la conformité du format des données. Cette étape procède également à la vérification de la concordance des informations entre la nouvelle étude envoyée et les études précédentes déjà insérées dans la base. Cette vérification est effectuée uniquement si l'utilisateur a explicitement déclaré, lors de la saisie des données dans le fichier Excel[®], que certains éléments pourront être utilisés dans d'autres études avec la même dénomination. Ainsi, si l'utilisateur a déclaré qu'un sujet (codé "101", par exemple) peut intervenir dans plusieurs études, la procédure vérifie l'existence de ce sujet chez ce fournisseur de données. Si ce sujet n'existe pas, il est créé. S'il existe déjà, la procédure vérifie que les informations concernant ce sujet (sexe, date de naissance...) sont cohérentes entre le fichier envoyé et la base de données. Si elles ne sont pas cohérentes, l'utilisateur est informé que l'insertion a échoué et précise les raisons de cet échec ; le message peut, par exemple, indiquer que la date de naissance du sujet 101 diffère entre les données de la base et celle du fichier CSV. Dans ce cas,

l'utilisateur doit corriger l'erreur dans le fichier Excel[®], générer à nouveau le fichier CSV et répéter l'envoi. Lorsqu'une étude est insérée correctement, l'utilisateur est informé du numéro que portera désormais cette étude dans la base de données. Il peut alors effectuer une demande d'analyse. Cette procédure d'insertion automatique des données depuis un fichier CSV a fait l'objet du stage de Licence2 de Julien Delaborde (10 semaines) que nous avons encadré.

L'insertion automatique des données est possible lorsque le fichier envoyé est au format CSV. En revanche, pour les fournisseurs ne disposant pas de la version 2003 d'Excel[®], le fichier CSV n'est pas généré. L'utilisateur envoie alors directement le fichier Excel[®] et la procédure d'insertion est semi-manuelle : une personne (administrateur de la base de données) doit exécuter successivement plusieurs procédures et vérifier à chaque étape la conformité des données. L'insertion est alors plus lente. L'utilisateur est en général informé de l'insertion de son étude une journée après son envoi.

4.2 Demande d'analyse d'une étude sensorielle

Après l'insertion d'un profil sensoriel dans la SensoBase, l'utilisateur peut effectuer une demande d'analyse depuis son espace personnel sur le site web. Une première page lui demande de sélectionner l'étude à analyser. Il est possible de sélectionner plusieurs études simultanément ; elles seront alors considérées comme faisant partie du même profil descriptif. L'utilisateur accède alors à la liste des sujets, produits et descripteurs de l'étude afin de lui permettre de retirer certains éléments de l'analyse s'il le désire. Il peut, par exemple, enlever un sujet qui n'a pas effectué de répétition ou un descripteur qui ne s'est pas révélé pertinent.

La page suivante propose à l'utilisateur de régler différents paramètres des analyses statistiques qui vont ensuite être effectuées (figure 3-10). Ces analyses correspondent à la sélection de méthodes présentées dans le chapitre 2. Seuls les principaux paramètres statistiques peuvent être modifiés ; les paramètres secondaires sont fixés pour toutes les analyses. Les valeurs par défaut correspondent à celles que nous recommandons lors d'une analyse standard. L'utilisateur doit également choisir la langue dans laquelle il souhaite recevoir les analyses et préciser l'adresse électronique où devront être envoyés les résultats.

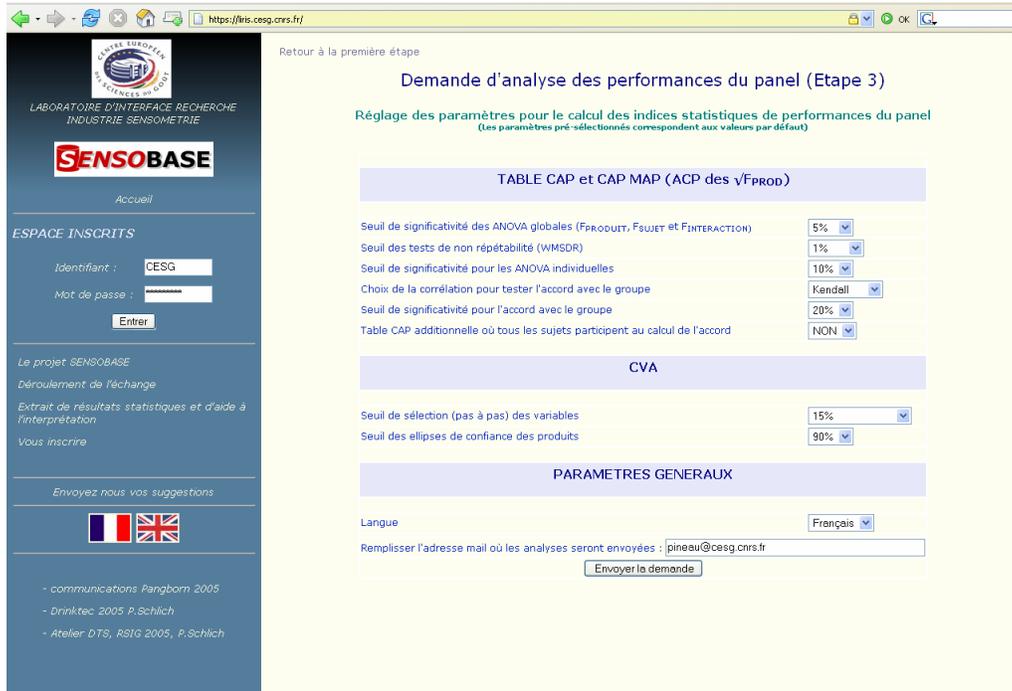


Figure 3-10 : page contenant les paramètres qui peuvent être modifiés lors d'une demande d'analyse

Enfin, les différents paramètres de la demande d'analyse sont récapitulés dans une dernière page (figure 3-11) et ces informations sont également envoyées par courrier électronique à l'utilisateur.

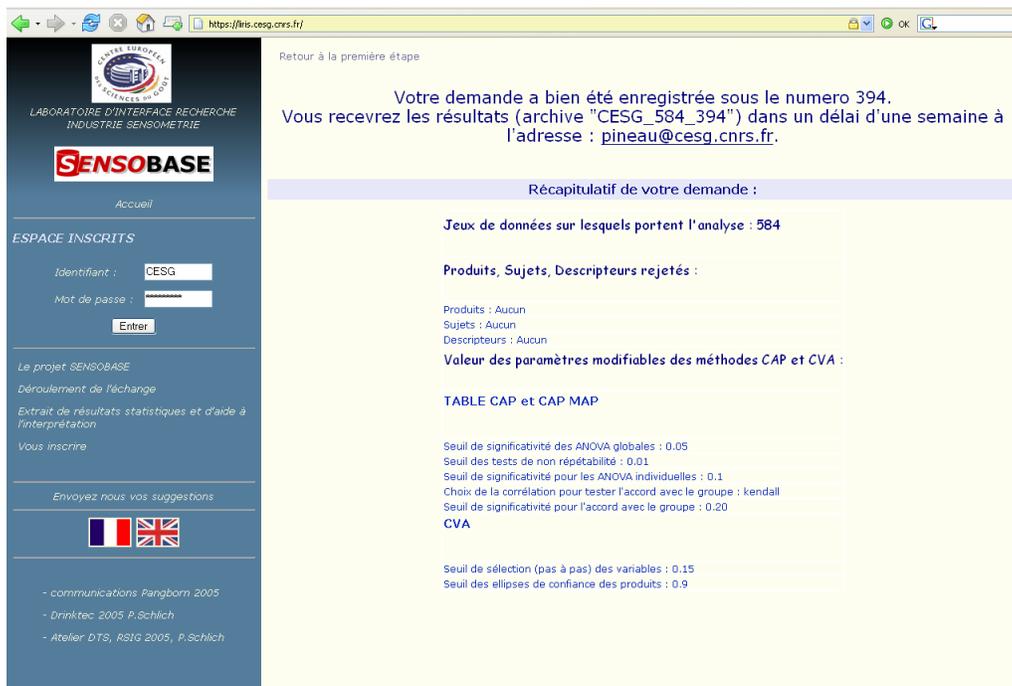


Figure 3-11 : Récapitulatif de la demande d'analyse

En parallèle, une procédure génère le code SAS qui devra être exécuté pour réaliser les analyses statistiques correspondantes et l'administrateur de la SensoBase reçoit un courrier lui indiquant qu'une demande vient d'être effectuée. Ce dernier doit alors manuellement exécuter le code SAS et vérifier que les résultats obtenus sont corrects. Les analyses, produites au format HTML, doivent ensuite être compressées (format ZIP) et envoyées manuellement par courrier électronique au fournisseur. Cette dernière étape de traitement des données et d'envoi des analyses n'est pas encore automatisée car il existe de nombreux cas pour lesquels certaines analyses ne peuvent pas être exécutées (absence de répétitions, plan de présentation trop déséquilibré...). Il faut alors expliquer à l'utilisateur la raison de ce problème. La grande diversité des cas à prévoir est donc un frein à la réalisation de cette étape. L'objectif du laboratoire est cependant d'automatiser complètement le processus d'échange de données contre des analyses statistiques dans les mois à venir.

4.3 Conclusion

Le système SensoBase offre une structure d'échange entre études de profils sensoriels et analyses statistiques. Son fonctionnement repose sur trois organes principaux (le site internet, le fichier Excel[®], et la base de données) qui assurent les différentes opérations depuis la saisie des données jusqu'à l'envoi des analyses. Outre la réalisation technique de chaque élément de ce système, le choix de chaque organe a constitué l'enjeu majeur du travail de conception de ce projet. Chaque élément doit en effet permettre de réaliser une tâche précise, avec ses contraintes propres, tout en permettant de communiquer facilement avec les autres éléments.

Aujourd'hui, le délai moyen d'attente entre l'envoi d'une étude sensorielle et la réception des analyses est inférieur à une journée. Chaque élément de la SensoBase fait régulièrement l'objet de modifications en fonction des attentes des fournisseurs et des développements que nous souhaitons ajouter. Les objectifs actuels sont d'automatiser l'étape d'analyse et d'envoi des résultats aux fournisseurs, ainsi que d'ajouter des options lors de la demande d'analyse afin de proposer des traitements plus adaptés à chaque situation.

Jusqu'à présent, le système proposé a permis de rassembler près de 500 études sensorielles. Cela représente approximativement 3 000 produits, 8 000 descripteurs, 1 500 panélistes, soit plus de trois millions d'observations. Les données sont décrites plus en détail dans le chapitre 5.

Chapitre 4 : Evolution de la performance au cours du temps

1 Contexte et objectifs

La réalisation de profils descriptifs sensoriels est aujourd'hui très répandue dans l'industrie agroalimentaire. Les entreprises peuvent constituer et entraîner leur(s) propre(s) panel(s) ou confier les études à des sociétés spécialisées en analyse sensorielle. Dans les deux situations, il n'est pas rare que les panels réalisent des profils sensoriels sur le même type de produits pendant de longues périodes (plusieurs mois, voir plusieurs années). Les centres ACTIA, co-financeur de ce travail de thèse disposent tous de panels qui effectuent régulièrement des profils sensoriels et leur motivation première dans ce projet était d'étudier l'évolution de la performance d'un panel et de ses panélistes au cours du temps. Les différents centres ACTIA partenaires de ce projet ont alors identifié des séries d'études sensorielles, réalisées avec le même panel et sur le même espace produit, et ont envoyé les données au LIRIS. Ces études correspondent aux premiers jeux de données insérés dans la SensoBase. Depuis, d'autres sociétés participant au projet SensoBase, comme le centre de recherche Unilever de Vlaardingen (Pays-bas), ont également apporté des séries d'études sensorielles.

Quelques études longitudinales de la performance ont déjà été proposées dans la littérature sur des thématiques voisines. Bitnes, Martens et al. (2005), s'intéressent à l'effet de l'âge, du niveau d'expérience et de la durée d'exposition à un produit sur le niveau de performance. Ils analysent pour cela les résultats de 21 panélistes ayant participé à des tests annuels d'identification de saveurs entre 1976 et 2003. Le nombre de mesures varie d'un sujet à l'autre (selon la durée de la participation à l'étude). Dans cette étude, l'analyse porte sur des tâches d'identification plutôt que sur des études de profil sensoriel et l'objectif n'est pas de caractériser l'évolution de la performance. Cependant, le fait que tous les panélistes n'aient pas effectué toutes les mesures est un élément important à prendre également en considération dans notre cas. Il est alors difficile d'utiliser des méthodes multidimensionnelles qui requièrent en général une structure de données sans observation manquante.

Vannier, Brun et al. (1999), étudient l'évolution de la qualité sensorielle d'un ensemble de produits (des champagnes). Ils utilisent pour cela une analyse en composante principale dans laquelle les individus sont les différents champagnes à chaque temps de mesure. Ils retracent alors, sur le premier

plan de l'ACP, l'évolution de la notation du panel en reliant chronologiquement les points représentant un même produit. Les auteurs précisent que l'évolution observée est sujette aux variations du style de notation des panélistes et conseillent de vérifier périodiquement que l'interaction entre les produits et les juges ne varie pas. Dans notre cas, le problème est inversé. L'objectif est de suivre l'évolution de la performance d'un panel et non l'évolution des produits, mais les variations de la qualité sensorielle des produits à chaque temps de mesure peuvent influencer le niveau de performance du panel. Les deux sources potentielles de variabilité de la performance (évolution de la notation des juges et produits différents à chaque temps de mesure) sont intimement liées et il est difficile de différencier les deux effets, ce qui peut compliquer l'interprétation des résultats.

Les études qui se focalisent spécifiquement sur la notion d'évolution de la performance utilisent le plus souvent des données obtenues, pour un panel, avant et après entraînement (Lesschaeve, 1997; Chollet et Valentin, 2001; Labbe, Rytz et al., 2004). Ces expérimentations permettent de mesurer l'effet de l'entraînement sur la performance en comparant les résultats des panélistes entre les deux temps de mesure. Notre objectif se situe dans le prolongement de ces études puisque nous cherchons à caractériser l'évolution de la performance d'un panel après l'entraînement, au fil des évaluations sensorielles.

L'objectif de ce chapitre est de proposer une méthodologie adaptée à l'analyse de données de profils sensoriels au cours du temps afin d'étudier l'évolution de la performance d'un panel. Plus précisément, l'enjeu de cette analyse est de permettre la détection d'une amélioration ou d'une baisse du niveau de la performance du panel ou de certains panélistes au fil du temps, ainsi que la détection de modifications brusques et ponctuelles du niveau de performance du groupe ou d'un panéliste.

L'approche méthodologique présentée dans ce chapitre utilise le modèle mixte pour caractériser l'évolution d'un indice de répétabilité : l'écart-type des scores entre les répétitions pour chaque sujet. Le modèle mixte a été retenu parce qu'il permet de modéliser simultanément l'évolution du panel et de chaque panéliste (Pinheiro et Bates, 2000). De plus, il peut être utilisé en présence de données manquantes. Cette analyse univariée est complétée par une représentation graphique basée sur l'utilisation des cartes de contrôle (Montgomery et Mastrangelo, 1991) afin d'identifier les temps de mesure auxquels le panel ou un panéliste s'éloigne significativement de la modélisation retenue. Cette méthodologie est mise en œuvre pour deux études présentant des caractéristiques différentes en terme de nombre de profils sensoriels et d'intervalle de temps entre chaque étude.

2 Matériel et méthodes

2.1 Données longitudinales

La première étude, fournie par le centre technique AERIAL, comprend huit profils sensoriels réalisés entre 1997 et 2001 (à intervalle de six mois). Pour chaque profil, quatre plats préparés congelés ont été dégustés deux fois et notés sur 30 descripteurs sensoriels par 9 à 11 panélistes parmi un ensemble de 13 dégustateurs (tableau 4-1). D'une étude à l'autre, les descripteurs étaient toujours les mêmes (excepté pour trois descripteurs qui n'ont pas été évalués lors des trois dernières études). En

revanche, les marques des produits étaient constantes mais les lots de production étaient différents. Les descripteurs étaient notés sur des échelles continues bornées entre 0 et 10. Parmi les 30 descripteurs, sept faisaient référence à l'apparence des produits (codés de ap_1 à ap_7), sept aux notes aromatiques (ar_1 à ar_7), sept aux odeurs (od_1 à od_7), trois aux saveurs (sa_1 à sa_3), deux aux sensations trigéminales (tr_1 et tr_2) et quatre à la texture (tx_1 à tx_4). Pendant la période de latence de six mois entre les études, les panélistes ont effectué d'autres études sensorielles sur des produits différents et ont participé à plusieurs séances d'entraînement.

Tableau 4-1 : Participation des panélistes selon les temps de mesure

panélistes	Profils sensoriels au cours du temps (années)								nbD**
	1	1.5	2	2.5	3	3.5	4	4.5	
101	X	X	X		X	X	X	X	7
102	X	X	X	X					4
103	X	X	X	X	X	X	X	X	8
104	X		X	X	X	X	X	X	7
105		X	X	X	X	X	X	X	7
106	X	X	X		X				4
107		X		X		X	X	X	5
108	X	X	X	X	X	X	X	X	8
109	X	X	X		X	X	X	X	7
111	X	X	X	X	X	X		X	7
112	X		X	X	X	X	X		6
113	X		X	X	X	X	X	X	7
115	X	X		X	X				4
Nbpan*	11	10	11	10	11	10	9	9	

*: nombre de panélistes à chaque temps de mesure

** : nombre de participation de chaque panéliste au cours de l'étude

La seconde étude présentée dans ce chapitre a été fournie par le centre de recherche Unilever de Vlaardingen. Elle est constituée de 43 profils répartis sur une année. D'une étude à l'autre, les nombres de produits, sujets et descripteurs pouvaient varier. Deux à huit sauces au soja étaient dégustées lors de chaque évaluation à l'exception des 32^{ème} et 38^{ème} profils pour lesquels le nombre de produits était particulièrement élevé (respectivement 16 et 24). Chaque dégustation était répétée au moins une fois (2 dégustations). Les sauces pouvaient varier au fil des études. Pour chaque profil, le nombre de panélistes était compris entre 7 et 12 parmi un ensemble de 13 juges. Les produits étaient notés sur un ensemble de 24 descripteurs (échelles en 16 points de 0 à 15) qui n'étaient pas systématiquement évalués à chaque temps de mesure. Ces descripteurs se répartissent en cinq catégories : 5 termes faisant référence à l'aspect visuel (apparence), 3 termes concernant les sensations trigéminales en bouche, 10 termes d'odeur, 4 termes de goût (3 saveurs et le "goût de brûlé") et 2 termes d'arrière goût. Cette seconde étude présente une situation moins équilibrée que la première mais plus proche de la réalité industrielle. Une description détaillée de cet ensemble de données est disponible en annexe 6. Dans cette étude, l'unité de temps retenue est le rang de chaque évaluation (de 1 à 43). Cet axe temporel représente le nombre d'expositions des panélistes aux produits de l'étude. Une autre possibilité serait de considérer les dates de réalisation de chaque mesure, mais cela revient à prendre en compte les variations d'intervalles de temps entre deux

mesures alors que l'apprentissage sensoriel n'a pas évolué. Cette question ne se pose pas pour la première étude dans laquelle les séances de mesures sont réparties uniformément au cours du temps.

La figure 4-1 propose une représentation schématique des données longitudinales que nous souhaitons analyser (pour une étude). A chaque temps de mesure, les observations des données brutes sont les évaluations (i.e. les notes sur les échelles de mesure) fournies par les panélistes à un ensemble de produits sur plusieurs descripteurs. Les juges évaluent chaque produit plusieurs fois. Chaque "tranche" du cube de données correspond aux évaluations d'un juge lors d'une répétition. Le nombre de produits, sujets et descripteurs peut varier d'un temps de mesure à l'autre et le dispositif peut présenter des données manquantes. L'analyse statistique ne porte pas directement sur ces données mais sur un indice de performance, calculé pour chaque panéliste et chaque descripteur sur l'ensemble des produits à chaque temps de mesure. Ces données longitudinales sur le niveau de performance sont ensuite regroupées par descripteur afin d'effectuer une analyse unidimensionnelle.

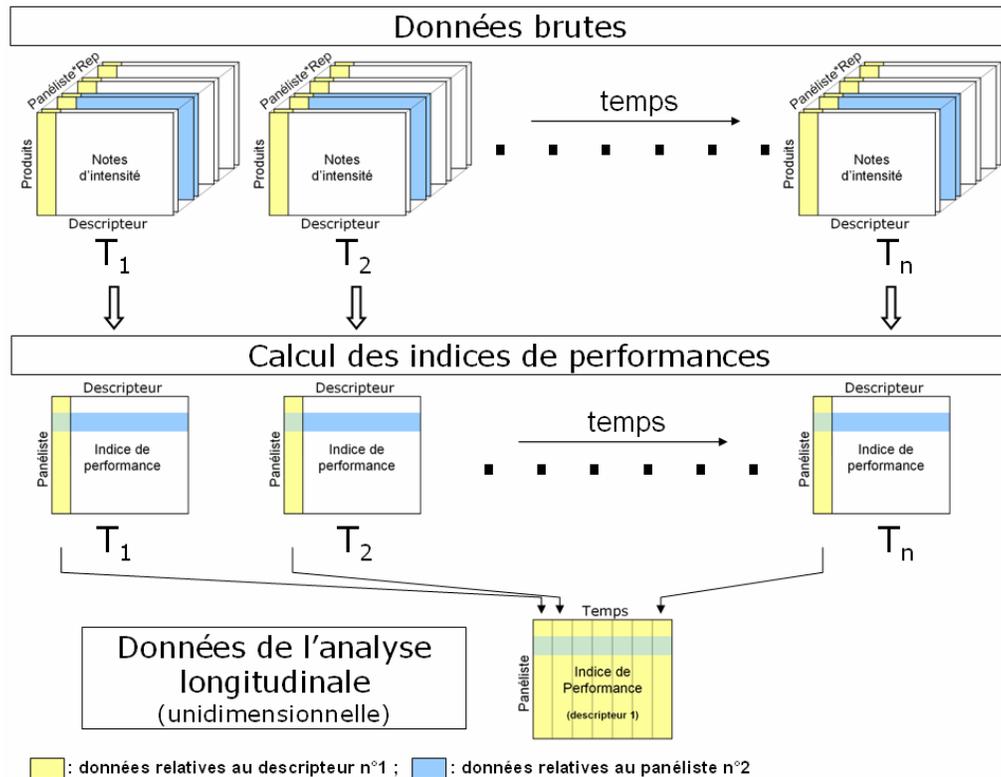


Figure 4-1 : Données utilisées pour étudier l'évolution de la performance

2.2 Indice de répétabilité

Afin de suivre l'évolution de la répétabilité au cours du temps, l'indice de performance retenu est la racine carrée du carré moyen résiduel obtenu à partir du modèle individuel d'ANOVA de l'effet produit à un facteur. Cet indice correspond à l'estimation de l'écart type entre les répétitions. Il est d'autant plus faible que le niveau de répétabilité est élevé. Cet indice a l'avantage d'être exprimé dans l'unité

de l'échelle de notation. Par exemple, un écart type des résidus égal à 1,2 sur une échelle de notation de 0 à 10 indique que, en moyenne, la note d'une répétition d'un produit s'éloigne de la moyenne de ce produit de 12% de la longueur de l'échelle. De plus, puisque la répétabilité mesure la variance intra produit, l'indice calculé ne dépend pas des différences entre les produits (variabilité inter produit). Par conséquent, la valeur de l'indice retenu ne dépend théoriquement pas de la variabilité des produits dans chaque étude. Cela n'aurait pas été le cas avec un indice de discrimination ou un indice d'accord entre les panélistes. Ceux-ci sont liés à la capacité à décrire l'espace produit et sont, de ce fait, dépendants des produits. Dans la suite de ce chapitre, l'indice retenu correspond à l'opposé de l'écart type entre les répétitions, appelé REP, afin d'obtenir des valeurs plus élevées pour les panélistes plus répétables. Pour chaque descripteur de l'étude, les données correspondent donc à des mesures répétées de cet indice de répétabilité au fil du temps.

2.3 Modèle mixte

La méthodologie du modèle mixte a été retenue pour analyser ce type de mesures répétées parce qu'elle offre la possibilité de prendre en compte simultanément l'évolution du panel dans son ensemble ainsi que l'évolution de chaque panéliste. De plus, le modèle mixte peut être utilisé malgré la présence de données manquantes. Pour chaque descripteur, le modèle mixte général peut s'écrire de la manière suivante :

$$y_{jt} = \mu_t + \text{panéliste}(j,t) + \varepsilon_{jt} \quad (4-1)$$

Où y_{jt} est la valeur de la répétabilité du panéliste j au temps t , μ_t est l'effet fixe temps, $\text{panéliste}(j,t)$ est l'effet aléatoire panéliste et ε_{jt} est le terme d'erreur.

Les effets aléatoires de tous les modèles présentés dans ce chapitre sont indépendants et distribués selon une loi normale centrée de variance constante. Pour chaque indice, la variance est différente. La partie fixe modélise l'évolution du panel dans son ensemble alors que la partie aléatoire modélise le comportement des panélistes. La phase de sélection de l'expression de la partie fixe permet de déterminer si la performance du panel croît, décroît, atteint un minimum ou un maximum, ou si elle reste constante au fil du temps. Ceci est réalisé en spécifiant un effet temps quadratique, linéaire ou aucun effet du temps dans la partie fixe du modèle. La sélection de la partie aléatoire permet de déterminer si le niveau de performance de certains panélistes augmente ou diminue plus rapidement que pour d'autres, ou si certains sujets progressent plus rapidement que le groupe au fil des premières études puis ralentissent leur progression par rapport au groupe. Cette étape correspond à la détermination de la structure de covariance de la partie aléatoire du modèle mixte (cf. paragraphe 2.4). Selon la structure de covariance choisie, le comportement de chaque panéliste peut plus ou moins s'écarter de l'évolution du groupe. La figure 4-2 montre la variété des modèles mixtes testés dans ce chapitre en fonction de la partie fixe et de la partie aléatoire retenue et propose une représentation graphique de chaque cas. La courbe noire représente l'évolution de l'ensemble du panel et chaque courbe grise représente l'évolution d'un panéliste. Par exemple, si l'effet fixe temps est quadratique (troisième ligne) et la structure de covariance de l'effet aléatoire est la "Compound Symmetry" (deuxième colonne), cela implique que la performance du panel atteint un maximum (ou

un minimum) et que l'évolution propre à chaque panéliste a la même forme que celle du groupe mais que le niveau de performance à l'origine diffère. Chaque sujet possède donc son propre potentiel de performance intrinsèque qu'il conserve au fil des profils sensoriels, tandis que le groupe évolue de manière quadratique.

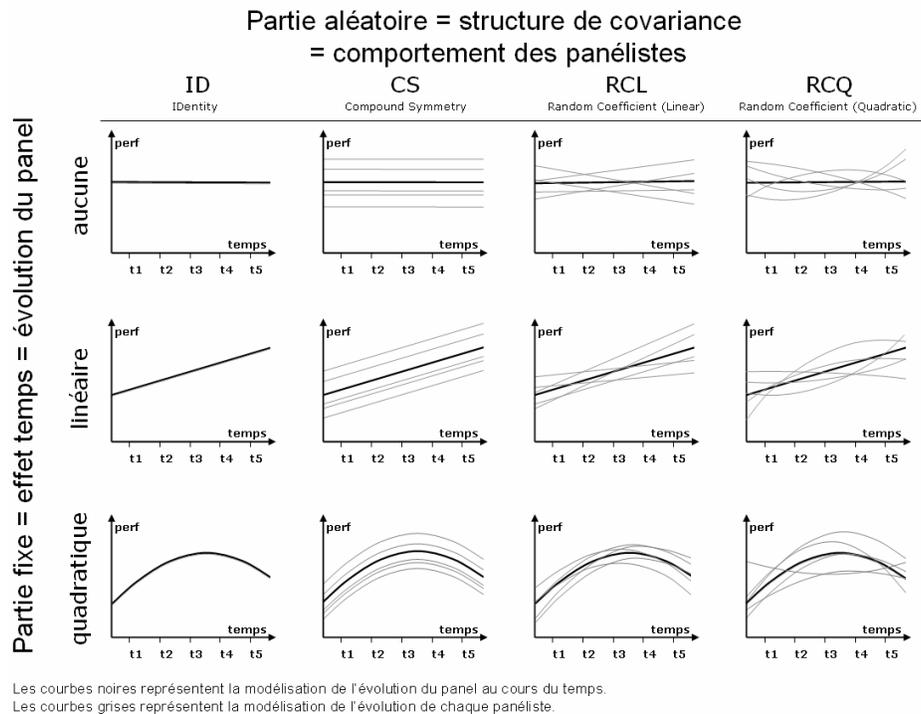


Figure 4-2 : Représentations graphiques des modèles mixtes en fonction de la partie fixe et de la partie aléatoire

Afin de sélectionner le modèle qui s'ajuste le mieux aux données, la première étape consiste à choisir la structure de covariance de la partie aléatoire du modèle (Pinheiro et Bates, 2000). La deuxième étape consiste à déterminer la partie fixe.

2.4 Choix de la structure de covariance

Le choix de la structure de covariance est lié au comportement "supposé" des panélistes par rapport à l'évolution du groupe. Lors de l'étape de sélection de la structure de covariance de la partie aléatoire, la partie fixe doit être identique dans chaque situation et doit être aussi complète que possible. L'effet temps est donc déclaré comme facteur catégoriel (et non comme régresseur). Parmi l'ensemble des structures de covariances existantes et adaptées aux mesures répétées (Wolfinger, 1996), quatre structures de covariances ont été retenues : les structures identité (ID), Compound Symmetry (CS), Linear Random Coefficient (RCL) et Quadratic Random Coefficient (RCQ). D'autres structures existent pour les situations dans lesquelles la corrélation entre les temps de mesure est plus forte (taux d'accroissement d'une plante, par exemple). La structure ID n'autorise aucune corrélation entre les temps de mesure. Ce choix implique qu'il n'existe aucune variabilité significative des niveaux de performance des panélistes au cours du temps. En d'autres termes, les panélistes

n'ont pas de comportement spécifique et sont supposés évoluer comme le groupe. Il n'y a donc pas d'effet panéliste dans le modèle mixte utilisé :

$$y_{jt} = \mu_t + \varepsilon_{jt} \quad (4-2)$$

$$\varepsilon_{jt} \sim N(0, \sigma_e^2)$$

La structure CS suppose qu'il existe une corrélation constante entre deux mesures, quel que soit l'intervalle de temps entre ces deux mesures. Lorsque cette structure s'ajuste aux données, cela signifie que le profil d'évolution de la performance de chaque panéliste est identique à celui du panel, mais le niveau initial est différent pour chaque sujet. En d'autres termes, la courbe d'évolution de chaque panéliste est la même mais l'ordonnée à l'origine est différente. Cela signifie que chaque panéliste est supposé avoir un niveau intrinsèque par rapport à l'évolution du niveau du groupe, qu'il conserve au cours du temps. Trois représentations graphiques faisant intervenir cette structure sont présentés dans la deuxième colonne de la figure 4-2. La structure CS correspond à un effet sujet aléatoire. L'expression du modèle mixte devient :

$$y_{jt} = \mu_t + \alpha_j + \varepsilon_{jt} \quad (4-3)$$

$$\alpha_j \sim N(0, \sigma_{\text{int}}^2)$$

$$\varepsilon_{jt} \sim N(0, \sigma_e^2)$$

Où α_j est l'effet sujet aléatoire, mesurant la valeur intrinsèque du panéliste j par rapport au panel.

La structure RCL est caractérisée par une ordonnée à l'origine aléatoire (comme pour la structure CS) et une pente aléatoire. Dans ce cas, la vitesse d'évolution de la performance de chaque panéliste peut être différente. Les graphiques correspondants à cette structure dans la figure 4-2 (troisième colonne) montrent que chaque sujet possède sa propre ordonnée à l'origine et sa propre pente. Le modèle mixte correspondant est :

$$y_{jt} = \mu_t + \alpha_j + \theta_{1j} \cdot \text{temps} + \varepsilon_{jt} \quad (4-4)$$

$$\alpha_j \sim N(0, \sigma_{\text{int}}^2)$$

$$\theta_{1j} \sim N(0, \sigma_{\text{pente}}^2)$$

$$\varepsilon_{jt} \sim N(0, \sigma_e^2)$$

Où α_j est l'effet sujet aléatoire et θ_{1j} est le paramètre de pente aléatoire, décrivant la vitesse d'évolution de la performance individuelle.

La structure RCQ ajoute un effet quadratique aléatoire. Les niveaux de performance individuels peuvent donc croître (resp. décroître) plus rapidement que le groupe au début de la série de mesures, puis décroître (resp. croître) à la fin par rapport à l'évolution du panel dans son ensemble.

Cette structure est montrée en exemple dans la quatrième colonne de la figure 4-2. Le modèle correspondant s'écrit :

$$y_{jt} = \mu_t + \alpha_j + \theta_{1j} \cdot \text{temps} + \theta_{2j} \cdot \text{temps}^2 + \varepsilon_{jt} \quad (4-5)$$

$$\alpha_j \sim N(0, \sigma_{\text{int}}^2)$$

$$\theta_{1j} \sim N(0, \sigma_{\text{pente}}^2)$$

$$\theta_{2j} \sim N(0, \sigma_{\text{courb}}^2)$$

$$\varepsilon_{jt} \sim N(0, \sigma_e^2)$$

Avec α_j l'effet sujet aléatoire, θ_{1j} le paramètre de pente aléatoire et θ_{2j} le paramètre de courbure aléatoire.

Puisque les structures de covariances sont organisées de manière hiérarchique, chaque modèle est testé par comparaison au modèle de degré supérieur afin de déterminer la structure de covariance qui s'ajuste le mieux aux données. Pour chaque comparaison de deux modèles emboîtés, l'hypothèse nulle est la nullité des variances (et covariances) des effets aléatoires ajoutés, ce qui correspond à l'égalité de l'ajustement des modèles. Le modèle le plus parcimonieux sera donc retenu (hypothèse nulle acceptée) à moins que le modèle plus complexe, utilisant un nombre plus important de paramètres, ne s'ajuste significativement mieux aux données (hypothèse nulle rejetée). Afin de réaliser ces comparaisons, des tests du rapport de vraisemblance, qui permettent de comparer des modèles emboîtés, sont effectués (Pinheiro et Bates, 2000). Pour chaque structure de covariance, la vraisemblance caractérise l'ajustement du modèle sélectionné aux données (Collett, 1996). Elle est calculée par "REstricted Maximum Likelihood" (REML), ou maximum de vraisemblance restreint. Cette méthode permet de corriger un biais existant dans le calcul du maximum de vraisemblance standard (pour de plus amples détails, voir Pinheiro et Bates, 2000). Sous l'hypothèse nulle d'équivalence entre deux structures, le logarithme du rapport de vraisemblance de deux modèles emboîtés avec la même partie fixe suit une distribution du Khi-deux, dans laquelle le nombre de degrés de liberté est la différence entre les nombres de paramètres de chaque modèle. Lorsque le test est significatif ($p < 0.05$), le modèle le plus complexe s'ajuste significativement mieux aux données.

Une autre approche pour comparer les modèles consiste à calculer un indice de vraisemblance comme le Critère d'Information d'Akaike (AIC) (Pinheiro et Bates, 2000) :

$$AIC = -2 \log(\text{vraisemblance}) + 2k \quad (4-6)$$

Avec k , le nombre de paramètres aléatoires du modèle. Cet indice réalise un compromis entre la qualité d'ajustement du modèle (vraisemblance) et la complexité de la structure de covariance (nombre de paramètres estimés). Il permet en outre la comparaison de modèles non emboîtés, ce qui n'est pas le cas du test utilisant les rapports des vraisemblances (Pinheiro et Bates, 2000; Chabanet et Pineau, 2005).

Les deux approches sont utilisées dans cette étude afin de comparer les modèles sélectionnés par chaque méthode pour chaque descripteur.

2.5 Choix de la partie fixe

L'objectif de la sélection de la partie fixe est de déterminer si l'évolution du panel tout entier au fil du temps est linéaire, quadratique ou si, au contraire, le niveau de performance reste constant au cours du temps. Le modèle mixte peut s'écrire de la manière suivante :

$$y_{jt} = \mu + \theta_1 \cdot \text{temps} + \theta_2 \cdot \text{temps}^2 + \text{panéliste}(j,t) + \varepsilon_{jt} \quad (4-7)$$

$$\varepsilon_{jt} \sim N(0, \sigma_e^2)$$

Avec $\text{panéliste}(j,t)$, l'expression générale de la partie aléatoire (effet sujet), spécifiée dans le paragraphe 2.4 "choix de la structure de covariance". Le paramètre μ est le terme constant (ou l'ordonnée à l'origine). Les paramètres θ_1 et θ_2 correspondent respectivement aux effets de pente et de courbure.

La significativité des effets linéaire et quadratique reliés aux paramètres θ_1 et θ_2 est testée à l'aide de tests de Fisher conditionnels (Pinheiro et Bates, 2000) avec des sommes de carrés de type I. Ces tests sont identiques à ceux réalisés classiquement en analyse de la variance mais ils utilisent les carrés moyens estimés à partir des paramètres du modèle mixte. Ils sont nommés "conditionnels" car ils dépendent de l'estimation (par REML) des paramètres aléatoires du modèle. Si le paramètre θ_1 est significativement différent de zéro, la performance du panel croît ou décroît au fil du temps selon le signe du paramètre estimé. Si le paramètre θ_2 est significatif, la performance du panel atteint un maximum ($\theta_2 < 0$) ou un minimum ($\theta_2 > 0$). Sur la figure 1, l'évolution du panel est représentée par la courbe noire. Les graphiques de la première ligne représentent des situations pour lesquelles l'effet temps n'est pas significatif ($\theta_1 = \theta_2 = 0$), la deuxième ligne montre des graphiques pour lesquels l'effet temps est linéaire ($\theta_1 \neq 0$ et $\theta_2 = 0$) et la troisième ligne représente des évolutions quadratiques de la performance du panel ($\theta_2 \neq 0$).

2.6 Paramètres individuels

Les solutions individuelles des effets aléatoires (correspondant aux estimations des paramètres α_j , θ_{1j} et θ_{2j} impliqués dans les modèles 4-3, 4-4 et 4-5) sont calculées afin d'apprécier l'évolution de chaque panéliste. Ces paramètres individuels permettent d'identifier les panélistes dont l'évolution est significativement différente de celle du panel entier. Les solutions individuelles de ces effets aléatoires sont estimés par "Best Linear Unbiased Predictors" (BLUP). Le mot "best" (meilleur), s'interprète en terme de carré moyen minimum (Pinheiro et Bates, 2000, S.A.S., 2003). Selon la structure de covariance sélectionnée dans le modèle mixte, le nombre de paramètres estimés est différent pour chaque descripteur. Il n'y a pas de paramètre avec la structure ID puisqu'il n'y a pas d'effet aléatoire (à l'exception de la variance résiduelle, estimée dans tous les modèles). L'ordonnée à l'origine est estimée dans les structures de covariance CS, RCL et RCQ, le paramètre de pente est estimé dans les structures RCL et RCQ et le paramètre de courbure est estimé uniquement dans la structure RCQ. La significativité de chaque paramètre individuel est évaluée en utilisant un test de Student :

$$T_{j,\omega} = \frac{\bar{\omega}_j - \bar{\omega}}{\sigma_{\bar{\omega}}} \quad (4-8)$$

$T_{j,\omega}$ est la statistique de test correspondant à la valeur ω_j du paramètre ω pour le sujet j (i.e. α_j, θ_{1j} ou θ_{2j}) et σ_ω est l'écart-type associé (i.e. σ_{int} , σ_{pente} ou σ_{courb}). Par construction, les valeurs de chaque paramètre sont centrées au niveau du groupe donc $\omega_j = 0$. Lorsque le test de nullité du paramètre est significatif, l'évolution du panéliste en question est significativement différente de celle du groupe. Par exemple, en supposant que la structure CS a été préalablement retenue, les panélistes dont l'ordonnée à l'origine est significativement supérieure à 0 sont des sujets pour lesquels le niveau de performance est plus élevé que pour les autres sujets au fil des évaluations sensorielles. Ces tests permettent de caractériser les performances individuelles au cours du temps.

2.7 Cartes de contrôle

Les cartes de contrôles permettent de représenter la valeur moyenne μ de la qualité d'une caractéristique et sa variabilité σ au fil du temps (Montgomery et Mastrangelo, 1991). Cette variabilité est utilisée pour calculer des limites de contrôle (LC) de cette caractéristique. L'expression générale des limites de contrôle est :

$$LC = \mu \pm 3\sigma \quad (4-9)$$

La valeur "3", couramment utilisée lors de la mise en œuvre de cartes de contrôles, représente un risque de 0.003 de rejeter à tort une observation dans un test bilatéral lorsque la variable est normalement distribuée. Cette valeur très conservative est utilisée pour les contrôles de qualité en milieu industriel afin de minimiser le risque de fausse alarme (Harris, Seppala et al., 1999). Cependant, le but de cette étude est d'identifier les temps de mesure auxquels la performance du panel ou d'un panéliste s'écarte de l'évolution globale et non pas seulement de détecter les dérives extrêmement fortes. La valeur sélectionnée pour définir les limites de contrôle est donc abaissée à 1,96 afin d'obtenir un intervalle de confiance à 0,95.

En général, μ et σ sont des paramètres inconnus estimés à partir d'une analyse de variance. Le modèle choisi dépend de la structure des données et de l'objectif de la carte de contrôle (Montgomery et Mastrangelo, 1991; Mortell et Runger, 1995). Dans cette étude, un modèle mixte a été ajusté au préalable sur les données. Les paramètres des cartes de contrôles sont donc estimés à partir des paramètres du modèle mixte. L'estimation des paramètres est fonction de l'information représentée sur la carte de contrôle. Lorsque l'objectif est de représenter la variabilité au niveau du panel, les données utilisées sont les moyennes de valeurs individuelles de l'indice REP à chaque temps de mesure. L'estimation de σ est σ_e / \sqrt{n} , la racine carrée de l'erreur résiduelle du modèle mixte divisée par la racine carrée du nombre de panélistes, n . L'estimation de μ , appelée $\hat{\mu}_t$, est issue de l'estimation des paramètres de l'effet temps dans le modèle mixte. Si l'effet temps est linéaire ou quadratique, $\hat{\mu}_t$ n'est pas constant au cours du temps. Les limites de contrôle de la répétabilité du panel sont alors :

$$LC_{panel} = \hat{\mu}_t \pm 1.96 \hat{\sigma}_e / \sqrt{n} \quad (4-10)$$

Avec :

$\hat{\mu}_t = \hat{\mu}$ si les effets linéaire et quadratique en fonction du temps ne sont pas significatif

$\hat{\mu}_t = \hat{\mu} + \hat{\theta}_1 \cdot temps$ si l'effet temps est significatif uniquement

$\hat{\mu}_t = \hat{\mu} + \hat{\theta}_1 \cdot temps + \hat{\theta}_2 \cdot temps^2$ si l'effet temps au carré est significatif

Les valeurs moyennes de l'indice REP (y_{jt}) situées à l'extérieur des limites de contrôle permettent de déterminer les temps de mesures auxquels le niveau de répétabilité du panel ne correspond pas à la modélisation de l'évolution de la répétabilité.

Lorsque l'objectif est de représenter la répétabilité au niveau du panéliste, les valeurs utilisées sont les valeurs individuelles de l'indice REP à chaque temps de mesure. L'estimation de μ est toujours $\hat{\mu}_t$, mais l'estimation de σ dépend de la structure de covariance sélectionnée dans le modèle mixte. Cependant, quel que soit le type de structure de covariance, la variabilité σ n'est pas divisée par \sqrt{n} parce que nous nous intéressons ici à l'intervalle de confiance autour des valeurs individuelles et non autour de la moyenne. Les limites de contrôles relatives à la répétabilité des panélistes sont donc :

$$LC_{panéliste} = \hat{\mu}_t \pm 1.96 \hat{\sigma}_t \quad (4-11)$$

Avec:

$\hat{\mu}_t$ est identique au terme de l'équation (4-10)

$\hat{\sigma}_t = \hat{\sigma}_e$ si la structure de corrélation est ID

$\hat{\sigma}_t = \sqrt{\hat{\sigma}_e^2 + \hat{\sigma}_{int}^2}$ si la structure de corrélation est CS

$\hat{\sigma}_t = \sqrt{\hat{\sigma}_e^2 + \hat{\sigma}_{int}^2 + temps^2 \cdot \hat{\sigma}_{pente}^2}$ si la structure de corrélation est RCL

$\hat{\sigma}_t = \sqrt{\hat{\sigma}_e^2 + \hat{\sigma}_{int}^2 + temps^2 \cdot \hat{\sigma}_{pente}^2 + temps^4 \cdot \hat{\sigma}_{courb}^2}$ si la structure de corrélation est RCQ

Les valeurs à l'extérieur des limites de contrôle mettent en évidence des panélistes pour lesquels le niveau de répétabilité est significativement différent de celui du groupe à un temps de mesure donné.

Bien que la méthodologie présentée dans ce chapitre soit unidimensionnelle, une présentation multidimensionnelle des résultats est proposée à l'aide d'une table de contingence dans laquelle les lignes correspondent aux sujet et les colonnes aux temps de mesure. Les valeurs reportées dans ce tableau sont les nombres de descripteurs pour lesquels la performance du panéliste est significativement plus faible que celle du panel (exemple dans le tableau 4-6). Cette table de

contingence met en évidence les panélistes aussi bien que les temps de mesure pour lesquels le nombre d'alertes sur le niveau de performance est élevé.

Toutes les analyses ont été réalisées avec le logiciel SAS® 9.1.3. Des programmes spécifiques ont été développés pour sélectionner automatiquement le meilleur modèle en fonction des données et pour produire les graphiques des cartes de contrôle correspondantes. La section du programme relative au paramétrage du modèle mixte est disponible en annexe 7.

3 Résultats

3.1 Première étude

3.1.1 Modèle mixte : choix de la structure de covariance

Pour chaque descripteur, les modèles mixtes correspondant aux quatre structures de covariances (ID, CS, RCL, RCQ) sont estimés et le choix de la structure de covariance est effectué en fonction des résultats des tests de modèles emboîtés (tableau 4-2).

Tableau 4-2 : Tests de différence des modèles emboîtés basés sur le logarithme du rapport de vraisemblance – étude n°1

DESCR	RCQ versus RCL			RCL versus CS			CS versus ID			choix
	ddl	logV	prob	ddl	logV	prob	Ddl	logV	prob	
ap_1	1	<0.01	0.9999	1	<0.01	0.9753	1	4.06	0.0439	CS
ap_2	1	<0.01	0.9999	1	0.40	0.527	1	0.33	0.5656	ID
ap_3	1	0.45	0.9297	1	1.99	0.1583	1	3.98	0.0461	CS
ap_4	1	<0.01	0.9785	1	0.03	0.9785	1	<0.01	0.9785	ID
ap_5	1	<0.01	0.9999	1	0.89	0.3454	1	10.68	0.0011	CS
ap_6	1	0.12	0.9893	1	1.45	0.2285	1	0.30	0.5838	ID
ap_7	1	0.23	0.9726	1	3.82	0.0506	1	4.62	0.0316	CS
ar_1	1	<0.01	0.9999	1	5.28	0.0215	1	19.15	<0.001	RCL
ar_2	1	<0.01	0.9999	1	3.55	0.0595	1	6.72	0.0095	CS
ar_3	1	<0.01	0.9999	1	<0.01	0.9999	1	5.74	0.0165	CS
ar_4	1	<0.01	0.9999	1	<0.01	0.9999	1	5.60	0.0179	CS
ar_5	1	<0.01	0.9999	1	1.14	0.2856	1	8.29	0.0039	CS
ar_6	1	<0.01	0.9999	1	<0.01	0.9999	1	4.75	0.0292	CS
ar_7	1	5.90	0.0151	1	0.01	0.9203	1	9.86	0.0016	RCQ
od_1	1	<0.01	0.9999	1	<0.01	0.9999	1	11.04	0.0008	CS
od_2	1	<0.01	0.9999	1	0.5	0.4795	1	0.61	0.4347	ID
od_3	1	0.03	0.9986	1	0.29	0.5902	1	23.30	<0.001	CS
od_4	1	0.76	0.859	1	3.33	0.068	1	21.09	<0.001	CS
od_5	1	<0.01	0.9999	1	0.81	0.3681	1	8.21	0.0041	CS
od_6	1	0.10	0.9918	1	<0.01	0.9999	1	10.20	0.0014	CS
od_7	1	5.70	0.0170	1	3.50	0.0614	1	6.90	0.0086	RCQ
ta_1	1	0.04	0.9978	1	0.49	0.4839	1	5.90	0.0151	CS
ta_2	1	3.03	0.3870	1	0.02	0.8875	1	8.04	0.0045	CS
ta_3	1	<0.01	0.9999	1	<0.01	0.9999	1	8.52	0.0035	CS
tr_1	1	8.84	0.0314	1	<0.01	0.9999	1	16.23	<0.001	RCQ
tr_2	1	<0.01	0.9999	1	0.04	0.8414	1	11.91	0.0005	CS
tx_1	1	13.56	0.0035	1	<0.01	0.9999	1	26.39	<0.001	RCQ
tx_2	1	<0.01	0.9999	1	<0.01	0.9999	1	45.51	<0.001	CS
tx_3	1	<0.01	0.9999	1	5.07	0.0243	1	7.83	0.0051	RCL
tx_4	1	<0.01	0.9999	1	1.59	0.2073	1	16.28	<0.001	CS

DESCR: nom des descripteurs, classés par type; **ddl** : nombre de degrés de liberté du test du Khi-deux ; **logV** : logarithme du rapport de vraisemblance entre les deux modèles comparés ; **prob** : probabilité associée au test, les valeurs inférieures à 0,05 sont surlignées en gras.

Dans cette table, chaque ligne correspond à un descripteur. Pour chaque test d'hypothèse, elle donne le nombre de degré de libertés de la statistique du Khi-deux (ddl), le logarithme du rapport de vraisemblance (logV) et la probabilité associée (prob). Lorsque la probabilité est inférieure à 0.05, le modèle le plus complexe est considéré comme mieux adapté aux données. La meilleure structure de covariance, sélectionnée d'après ces tests, est reportée dans la dernière colonne du tableau 4-2. La stratégie utilisée pour sélectionner la meilleure structure est expliquée à l'aide de la figure 4-3.

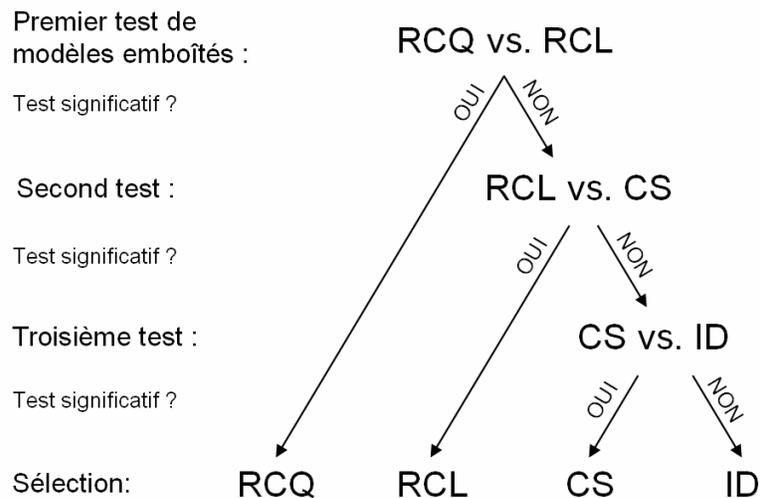


Figure 4-3 : Arbre de décision permettant d'établir la sélection de la structure de covariance de chaque modèle

La démarche consiste à tester le modèle le plus complexe (RCQ) par rapport au modèle emboîté de degré immédiatement inférieur (RCL). Si l'hypothèse nulle d'égalité d'ajustement des deux modèles est rejetée, le modèle complexe est conservé car il ne peut pas être approximé par un modèle plus simple. En revanche, si les deux modèles sont équivalents, le modèle simple est retenu et testé selon le même processus par rapport au modèle emboîté suivant, de degré immédiatement inférieur.

Pour la plupart des descripteurs, la structure de covariance sélectionnée est CS (20 descripteurs sur 30). Ce résultat implique que l'évolution de la performance de chaque panéliste est identique à celle du panel. Cependant, leur niveau initial est différent et cette différence persiste au fil des évaluations, quel que soit l'évolution du panel. Ce résultat peut néanmoins être modulé selon le type de descripteur. La structure CS est très souvent sélectionnée pour les descripteurs d'arôme, d'odeur et de saveur, mais pour les autres types de descripteurs, d'autres structures peuvent également être sélectionnées. Pour les descripteurs visuels, la structure ID est sélectionnée pour trois descripteurs (ap_2, ap_4 et ap_6) sur sept. De plus, les probabilités associées aux tests de modèles emboîtés entre les structures ID et CS pour les descripteurs ap_1 et ap_3 sont très proches de la limite de significativité de 0.05 ($prob_{ID-CS}=0.044$ pour le descripteur ap_1 et $prob_{ID-CS}=0.046$ pour le descripteur ap_3), donc la structure ID pourrait éventuellement être adaptée à deux descripteurs visuels supplémentaires. Dans ce cas, en considérant ID comme une structure potentielle pour les descripteurs visuels, cela signifierait que les niveaux de répétabilité individuels ne sont pas différents les uns des autres pour ce type de descripteur. Concernant les descripteurs trigéminaux et de texture, la structure CS n'est pas sélectionnée aussi souvent que pour les autres types de descripteurs : la

moitié des structures sélectionnées sont RCL ou RCQ. Ce résultat pourrait être un argument pour considérer que l'évolution de la performance des panélistes (par rapport au groupe) est plus complexe pour les descripteurs trigéminaux et de texture, puisque chaque panéliste possède sa propre pente et/ou sa propre courbure.

Le tableau 4-3 rassemble les valeurs d'AIC de chaque descripteur pour chaque structure de covariance. Cet indice est d'autant plus faible que le modèle est ajusté aux données. Le choix de la structure de covariance avec l'indice AIC conduit à une solution différente du choix réalisé avec les tests de modèles emboîtés pour quatre descripteurs seulement (ap_7, ar_2, ar_7, od_4). Pour ces descripteurs, la structure de corrélation retenue est plus complexe (RCL ou RCQ au lieu de CS). Notons toutefois que ces différences disparaissent lorsque le seuil de significativité des tests de modèles emboîtés est défini à 0,07 au lieu de 0,05. Le critère d'Akaike conduit donc à des résultats légèrement plus complexes dans certains cas mais les conclusions restent néanmoins très proches pour ces deux méthodes de sélection. Dans la suite de ce chapitre, nous retiendrons les résultats de la première méthode (tests de modèles emboîtés) afin de sélectionner des modèles aussi parcimonieux que possible.

Tableau 4-3 : indice AIC pour chaque descripteur en fonction de la structure de covariance

descripteurs	ID*	CS*	RCL*	RCQ*
ap_1	154.1	152.0	154.0	156.0
ap_2	130.2	132.2	134.2	136.2
ap_3	133.7	131.7	133.7	133.2
ap_4	117.6	119.6	119.5	121.6
ap_5	148.7	140.1	142.1	142.1
ap_6	188.7	190.4	190.9	188.8
ap_7	166.8	164.2	162.3	164.1
ar_1	185.5	168.4	165.1	167.1
ar_2	179.8	175.1	173.5	175.5
ar_3	188.8	185.1	187.1	189.1
ar_4	107.9	104.3	106.3	108.3
ar_5	116.1	109.8	110.7	112.7
ar_6	172.6	169.8	169.8	169.8
ar_7	180.6	172.7	174.7	170.9
od_1	152.5	143.4	145.4	147.4
od_2	163.1	165.1	167.1	169.1
od_3	98.3	77.0	78.7	80.7
od_4	184.6	165.5	164.2	165.4
od_5	157.3	151.1	152.3	152.3
od_6	164.1	155.9	155.9	157.8
od_7	207.9	203.1	201.6	197.9
ta_1	127.8	123.9	125.4	127.4
ta_2	153.9	147.9	149.9	148.8
ta_3	177.0	170.5	172.5	174.5
tr_1	153.5	139.3	141.3	132.5
tr_2	173.1	163.2	165.2	167.2
tx_1	191.0	166.7	169.7	155.1
tx_2	229.7	186.2	188.2	190.2
tx_3	173.3	167.5	164.4	166.4
tx_4	191.9	177.6	178.0	180.0
nb sel¹	4	17	5	4

*: meilleure structure en gras pour chaque descripteur (l'AIC est d'autant plus faible que l'ajustement est adéquat)

¹: nombre de sélection de chaque structure de covariance

3.1.2 Modèle mixte : choix de la partie fixe

Pour chaque descripteur, la meilleure structure de covariance est sélectionnée d'après les résultats du tableau 4-2. La partie fixe du modèle est ensuite sélectionnée (équation 4-7). L'effet du temps (linéaire ou quadratique) n'est significatif que pour cinq descripteurs. En général, le niveau de répétabilité du panel n'évolue donc pas au cours du temps ; la performance du groupe tend à garder le même niveau au fil des évaluations pour la majorité des descripteurs. Le tableau 4-4 présente les valeurs des paramètres estimés des descripteurs pour lesquels l'effet temps était significatif ($prob < 0,05$).

Tableau 4-4 : Estimation des paramètres de l'effet fixe temps
(seulement pour les descripteurs dont l'effet temps linéaire ou quadratique était significatif)

Descripteur	Paramètre	Estimation	Ecart-type	DDL	T de Student	Prob
ap_2	Ord. origine	-1.1751	0.08414	78	-13.97	<.0001
	Pente (temps)	-0.1206	0.04966	78	-2.43	0.0174
	Courbure (temps ²)	-0.00270	0.04890	78	-0.06	0.9561
ap_4	Ord. origine	-1.4287	0.08064	78	-17.72	<.0001
	Pente (temps)	0.08722	0.04760	78	1.83	0.0707
	Courbure (temps ²)	0.1130	0.04686	78	2.41	0.0183
ap_5	Ord. origine	-1.1611	0.1219	12	-9.52	<.0001
	Pente (temps)	0.1052	0.05156	66	2.04	0.0453
	Courbure (temps ²)	0.02277	0.04926	66	0.46	0.6454
od_2	Ord. origine	-0.7860	0.1088	78	-7.22	<.0001
	Pente (temps)	0.1703	0.06425	78	2.65	0.0097
	Courbure (temps ²)	-0.03292	0.06326	78	-0.52	0.6042
od_6	Ord. origine	-1.3022	0.1342	12	-9.70	<.0001
	Pente (temps)	0.1126	0.05490	66	2.05	0.0442
	Courbure (temps ²)	-0.00999	0.05237	66	-0.19	0.8493

Paramètre: paramètre estimé, variable correspondante entre parenthèses; **Estimation:** estimation du paramètre relatif à l'effet fixe; **Ecart-type:** Ecart-type de l'erreur liée à l'Estimation; **DDL:** nombre de degrés de liberté; **T de Student:** Valeur de la statistique de Student; **Prob:** probabilité associée à la statistique de Student.

Quatre descripteurs avaient un effet temps linéaire (ap_2, ap_5, od_2 et od_6). L'estimation des pentes des deux descripteurs visuels (ap_2 et ap_5) sont négatives donc le niveau de répétabilité décroît au cours du temps. Au contraire, les pentes des descripteurs d'odeur (od_2 et od_6) sont positives ; la performance du groupe augmente donc de manière significative. Un seul descripteur (ap_4) présente une évolution quadratique significative et le paramètre estimé correspondant est positif donc le niveau de répétabilité présente une courbure convexe (orientée vers le bas). De plus, le paramètre correspondant à la pente est à la limite de la significativité ($prob = 0,0707$) et son estimation est positive, indiquant la tendance à croître du niveau de répétabilité du panel. Précisons ici que lorsque l'effet quadratique est significatif, l'effet linéaire ne doit jamais être supprimé du modèle même si ce dernier n'est pas significatif, afin de maintenir la structure hiérarchique du modèle. Par conséquent, le niveau de répétabilité a plutôt tendance à rester constant lors des premières évaluations puis à augmenter progressivement. Le graphique de l'évolution du descripteur ap_4 sera présenté et commenté ultérieurement dans la partie sur les cartes de contrôle (figure 4-10).

3.1.3 Modèle mixte : paramètres individuels

En fonction du modèle mixte retenu pour chaque descripteur, les paramètres individuels correspondants sont estimés. Les résultats de trois descripteurs seulement sont présentés (ar_4, tx_3 et od_7), pour lesquels les structures de covariances sont respectivement CS, RCL et RCQ. Le tableau 4-5 présente les paramètres estimés pour chaque panéliste.

Tableau 4-5 : Estimation des paramètres individuels

panéliste	ar_4		tx_3		od_7	
	structure CS		structure RCL		structure RCQ	
101	Ord. O.	-0.131	Ord. O.	0.049	Ord. O.	0.003
			Pente	0.067	Pente	-0.03
					Courb.	-0.067
102	Ord. O.	-0.470*	Ord. O.	-0.337	Ord. O.	-0.001
			Pente	0.320^(*)	Pente	0.043
					Courb.	-0.475*
103	Ord. O.	0.352	Ord. O.	0.129	Ord. O.	-0.021
			Pente	-0.17	Pente	0.087
					Courb.	-0.169
104	Ord. O.	-0.243	Ord. O.	-0.226	Ord. O.	0.003
			Pente	0.269^(*)	Pente	-0.098
					Courb.	0.073
105	Ord. O.	0.257	Ord. O.	0.288	Ord. O.	-0.02
			Pente	0.23	Pente	-0.016
					Courb.	-0.344^(*)
106	Ord. O.	-0.233	Ord. O.	-0.267	Ord. O.	0.001
			Pente	-0.098	Pente	0.039
					Courb.	-0.374^(*)
107	Ord. O.	-0.121	Ord. O.	-0.081	Ord. O.	0.012
			Pente	0.054	Pente	0.035
					Courb.	0.222
108	Ord. O.	0.046	Ord. O.	0.114	Ord. O.	-0.016
			Pente	-0.055	Pente	-0.087
					Courb.	-0.02
109	Ord. O.	0.359	Ord. O.	0.441*	Ord. O.	0.035
			Pente	-0.052	Pente	0.022
					Courb.	0.407*
111	Ord. O.	-0.336	Ord. O.	-0.057	Ord. O.	-0.029
			Pente	-0.261^(*)	Pente	0.009
					Courb.	0.017
112	Ord. O.	0.214	Ord. O.	0.317	Ord. O.	0.015
			Pente	-0.04	Pente	0.031
					Courb.	0.142
113	Ord. O.	0.147	Ord. O.	-0.354	Ord. O.	0.036
			Pente	0.024	Pente	-0.009
					Courb.	0.316^(*)
115	Ord. O.	0.161	Ord. O.	-0.015	Ord. O.	-0.015
			Pente	-0.051	Pente	-0.026
					Courb.	0.273

Les estimations significatives sont en gras. *: $p < 0.05$; (*): $p < 0.10$

Concernant le descripteur ar_4, l'ordonnée à l'origine est significativement différente de la moyenne pour le sujet 102 uniquement. Etant donné que l'estimation de ce paramètre est négative (-0,47), le niveau de répétabilité de ce panéliste est globalement moins élevé que celui du reste du groupe au fil des évaluations. Pour le descripteur tx_3 (structure de covariance RCL), le paramètre de pente est significativement différent de la pente moyenne pour trois juges. Ce paramètre est positif pour les panélistes 102 et 104, donc leur répétabilité s'améliore au fil du temps par rapport au niveau du panel. A l'inverse, la pente du panéliste 111 est négative ; le niveau de répétabilité de ce sujet se détériore au cours du temps par rapport au groupe. Concernant le descripteur od_7 (structure de covariance RCQ), le paramètre de courbure aléatoire est significatif pour cinq panélistes mais le paramètre de pente n'est jamais significatif. La répétabilité des sujets 102, 105 et 106, pour lesquels l'estimation du paramètre de courbure est négative, augmente plus rapidement que le niveau de répétabilité du groupe lors des premières évaluations, puis décroît plus rapidement. Le phénomène inverse se

produit pour les panélistes 109 et 113 dont l'estimation du paramètre de courbure est positive. Ces résultats indiquent que le comportement de certains panélistes est sensiblement différent de l'évolution du panel. L'évolution moyenne ne reflète donc pas fidèlement le comportement de chaque panéliste.

3.1.4 Cartes de contrôle

En fonction du modèle mixte sélectionné pour chaque descripteur, les paramètres des cartes de contrôle correspondants sont calculés. Ce chapitre présente uniquement les cartes de contrôle de 4 descripteurs sur 30, dans les figures 4-4 à 4-11, correspondant respectivement aux descripteurs ar_4, tx_3, od_6 et ap_4. Les limites de contrôle supérieure (LCS) et inférieure (LCI) sont représentées en rouge. L'estimation de l'évolution de la moyenne du panel est représentée en noir et appelée EMP (Estimation Moyenne du Panel).

La carte de contrôle du descripteur ar_4 présente un exemple dans lequel la structure de covariance sélectionnée est CS et l'effet temps n'est pas significatif. Cette sélection correspond au modèle mixte retenu pour la plupart des descripteurs. Puisque l'effet temps n'est pas significatif, les limites de contrôles sont constantes au cours des évaluations. La figure 4-4 montre la carte de contrôle et l'évolution de la moyenne du panel. La performance du groupe n'est jamais à l'extérieur des limites de contrôles donc la répétabilité du panel est homogène au fil du temps. La figure 4-5 présente la carte de contrôle du niveau des panélistes. Celle-ci met en évidence que trois panélistes (102, 104 et 111) ont un niveau de répétabilité significativement moins élevé que les autres au temps t=2, bien que la performance du groupe ne soit pas en dehors des limites de contrôle à ce temps de mesure.

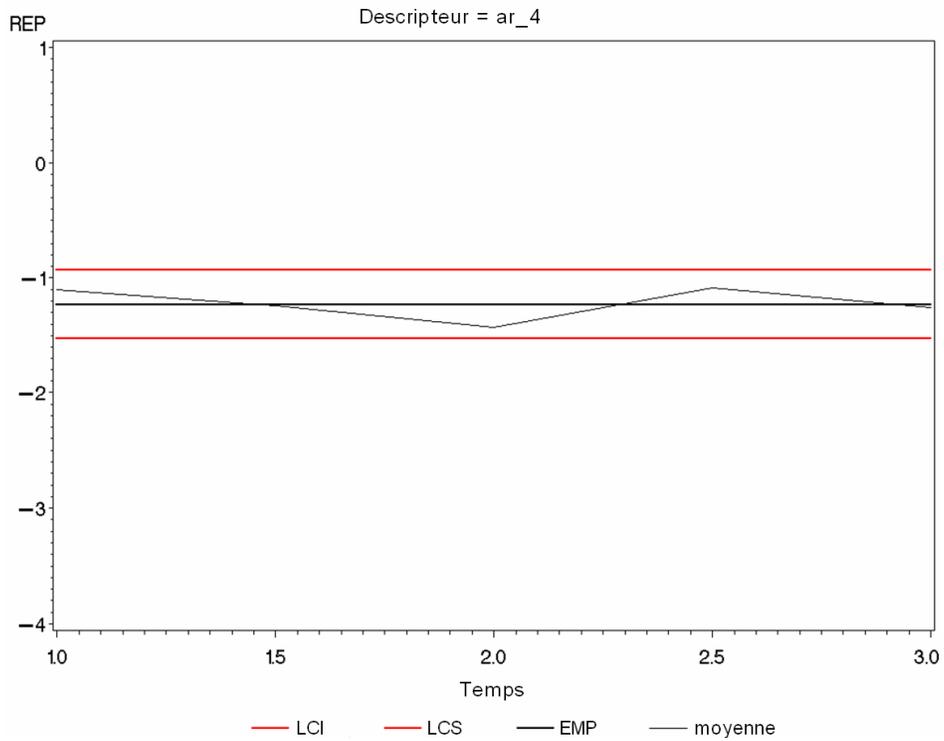


Figure 4-4 : carte de contrôle de l'évolution du panel pour le descripteur ar_4 (effet temps n.s.; aucune donnée hors des limites de contrôle)

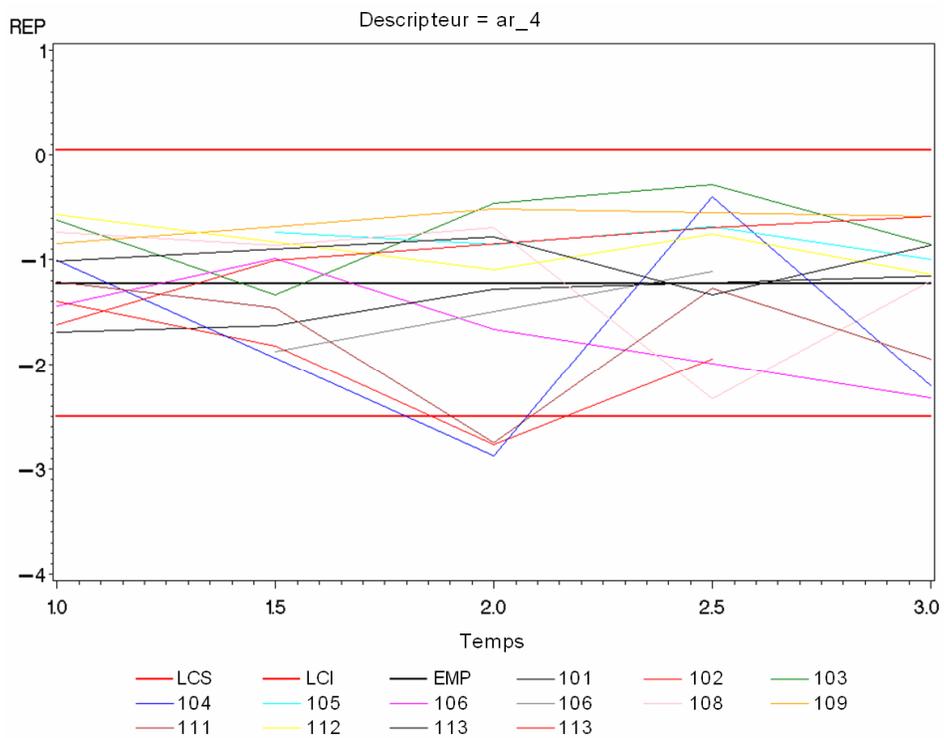


Figure 4-5 : carte de contrôle de l'évolution des panélistes pour le descripteur ar_4 (structure de covariance CS; effet temps n.s.; panélistes 102, 104 et 111 Hors des limites de contrôle au temps t=2)

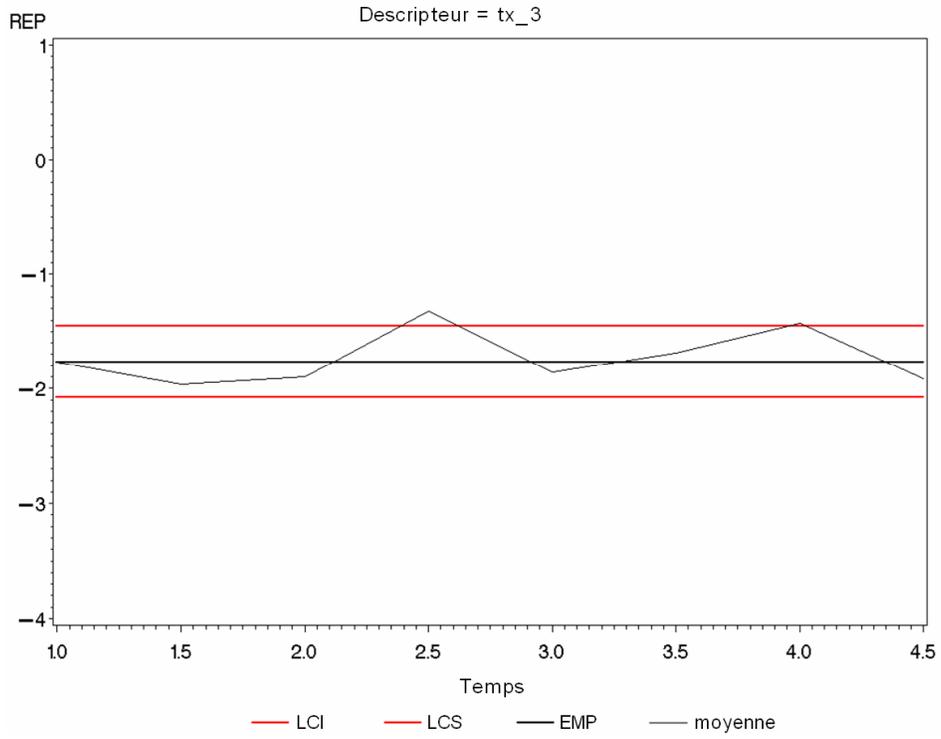


Fig. 3.a.

Figure 4-6 : carte de contrôle de l'évolution du panel pour le descripteur tx_3 (effet temps n.s.; performance du panel hors des limites de contrôle aux temps t=2,5 et t=4)

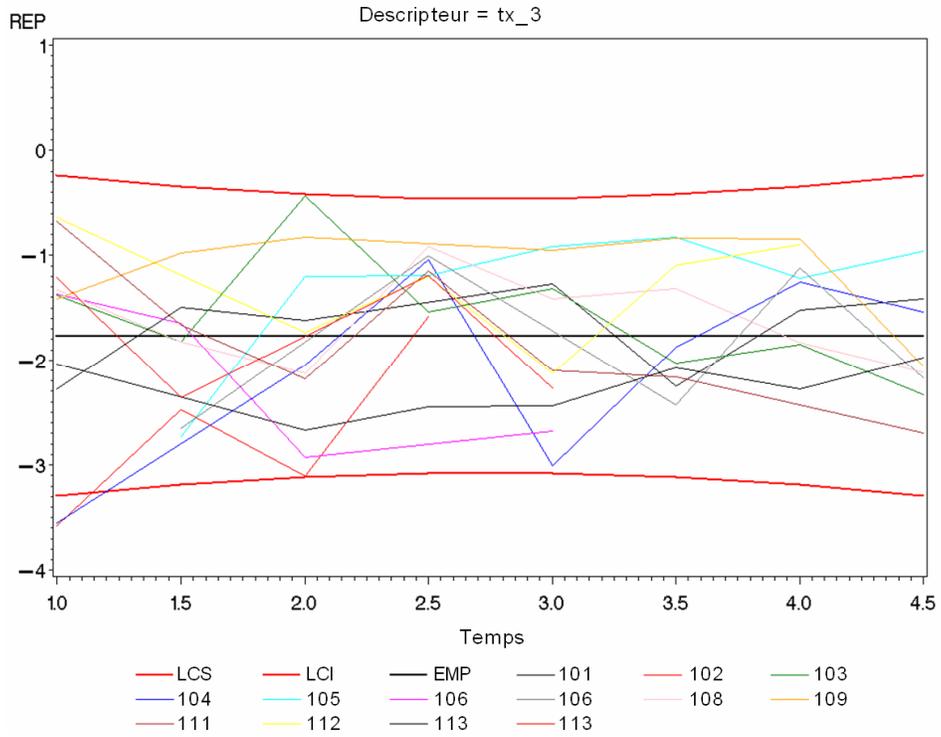


Figure 4-7 : carte de contrôle de l'évolution des panélistes pour le descripteur tx_3 (structure de covariance RCL; effet temps n.s.; panélistes 102 et 104 hors des limites de contrôle aux temps t=1)

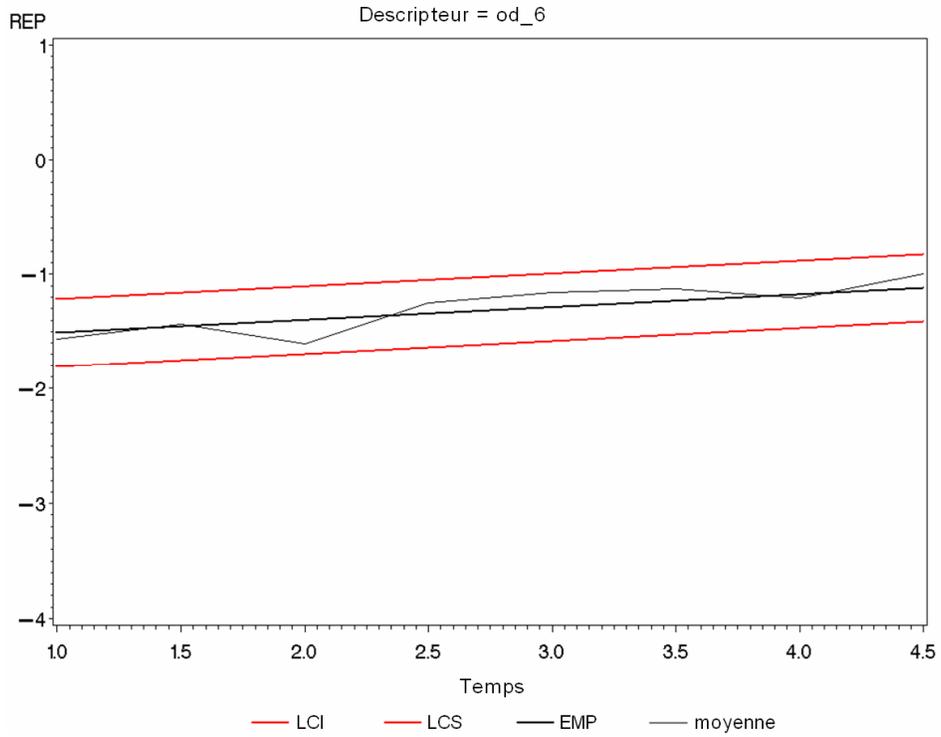


Figure 4-8 : carte de contrôle de l'évolution du panel pour le descripteur od_6 (effet temps linéaire significatif; aucune donnée hors des limites de contrôle)

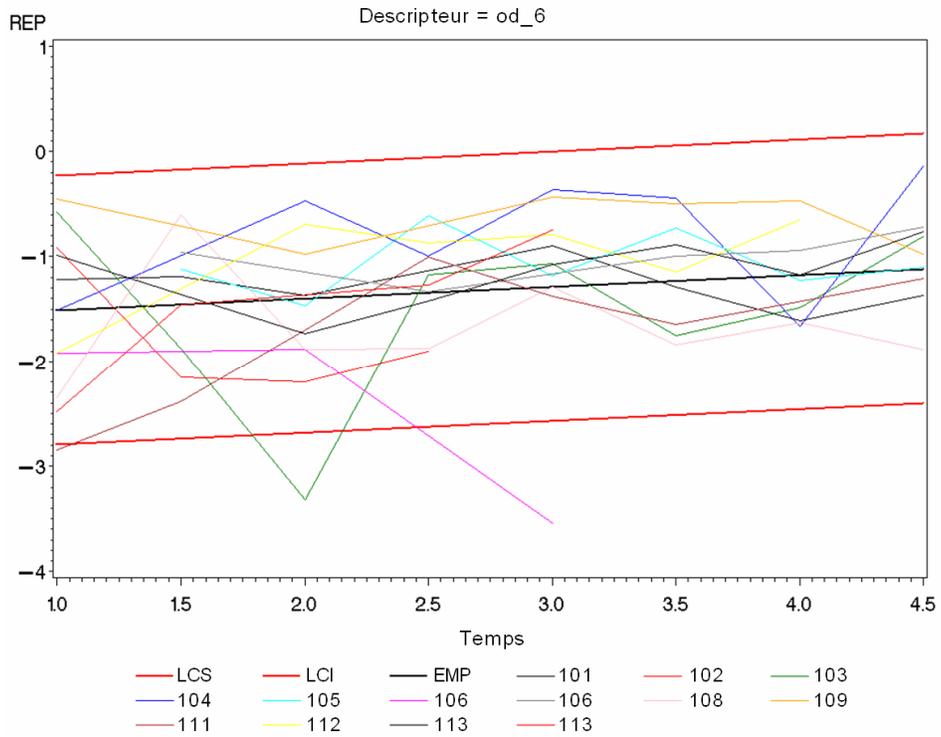


Figure 4-9 : carte de contrôle de l'évolution des panélistes pour le descripteur od_6 (structure de covariance CS; effet temps linéaire significatif ; panélistes 103, 106, 111 hors des limites de contrôle à une occasion)

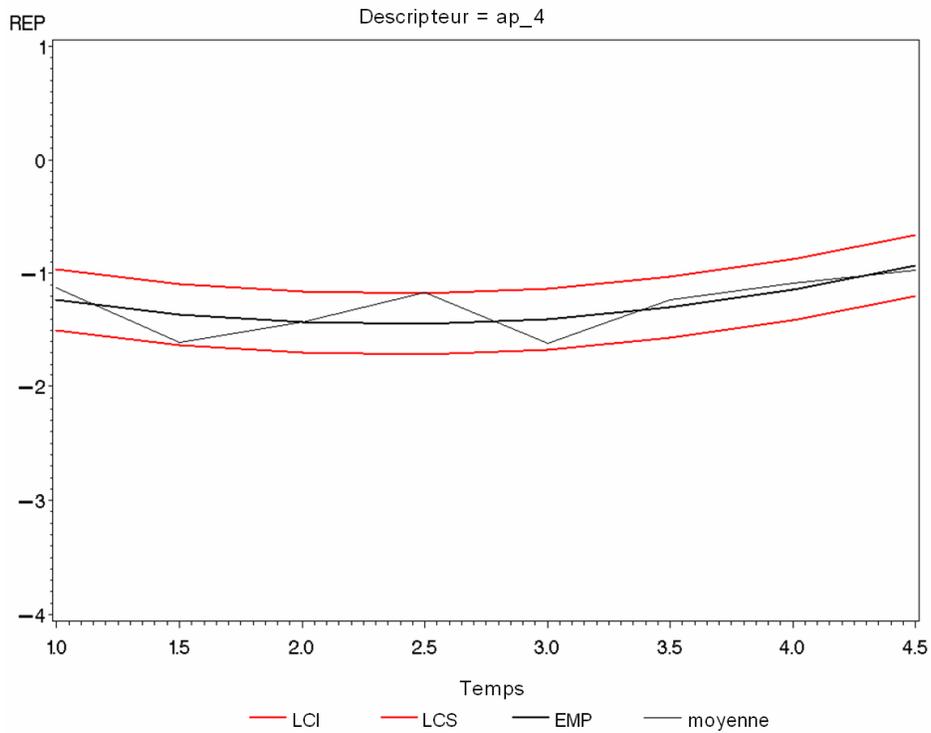


Figure 4-10 : carte de contrôle de l'évolution du panel pour le descripteur ap_4 (effet temps quadratique significatif; aucune donnée hors des limites de contrôle)

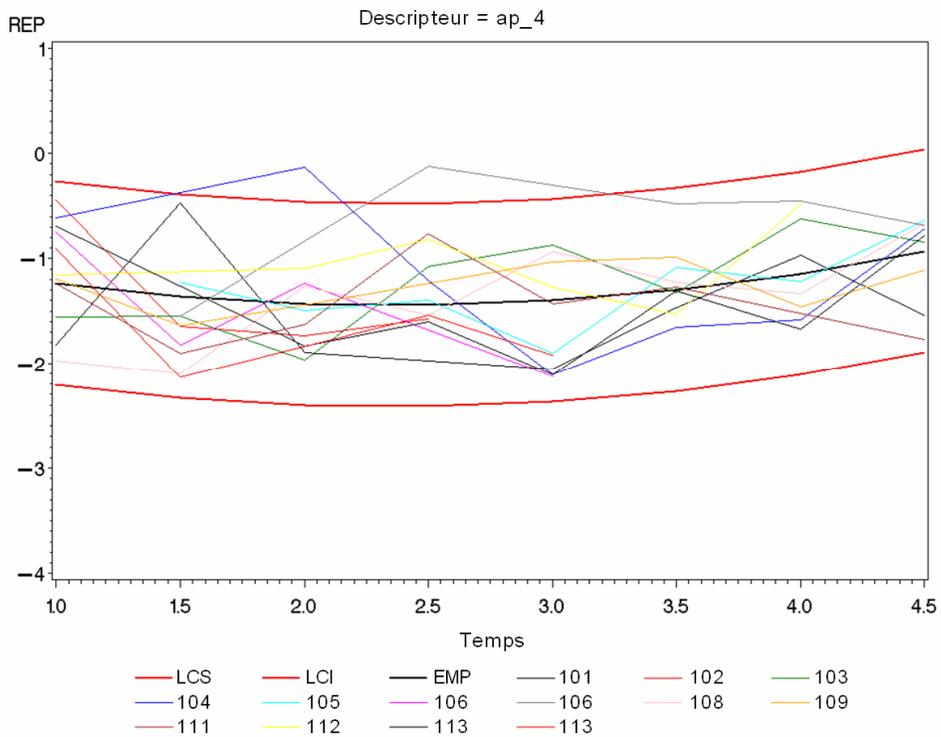


Figure 4-11 : carte de contrôle de l'évolution des panélistes pour le descripteur ap_4 (structure de covariance ID; effet temps quadratique significatif ; panélistes 104 et 107 hors des limites de contrôle à une occasion (limite supérieure))

Les figures 4-6 et 4-7 montrent les résultats du descripteur tx_3 pour lequel la structure de covariance retenue est RCL et l'effet temps n'est pas significatif. Contrairement au premier exemple, le niveau de répétabilité du panel est significativement plus élevée au temps t=2,5 et t=4 alors qu'aucun panéliste n'est significativement meilleur que les autres à ces deux temps de mesure. Cependant, le niveau de répétabilité augmente pour la plupart des sujets au temps t=2,5 sur la figure 4-7. Cette carte de contrôle met également en évidence deux panélistes (102 et 104) significativement moins répétables que le groupe à t=1. Puisque la variabilité utilisée dans le calcul des limites de contrôle dépend du temps lorsque la structure de covariance est RCL ou RCQ (équation 4-11), ces limites évoluent au fil du temps.

Les figures 4-8 et 4-9 présentent les résultats pour le descripteur od_6 dont la structure de covariance est CS et l'effet temps linéaire est significatif. Par conséquent, l'estimation de la répétabilité moyenne évolue au cours du temps ainsi que les limites de contrôle. L'évolution linéaire correspond à une augmentation significative du niveau de répétabilité au cours du temps (estimation de la pente : 0,113, p=0,042). Sur la figure 4-9, trois panélistes se situent hors des limites de contrôle à différents temps de mesure (sujets 103, 106 et 111).

Les résultats relatifs au descripteur ap_4, présentés dans les figures 4-10 et 4-11, utilisent la structure de corrélation ID et montrent un effet quadratique de l'évolution de la performance. Comme dans l'exemple précédent, l'estimation de la répétabilité du groupe évolue donc au cours du temps, mais, cette fois, selon un profil quadratique. Le niveau de répétabilité du groupe tend à décroître légèrement au début de la série de données, puis à augmenter plus nettement entre les temps de mesure 3 et 4,5. La répétabilité de deux panélistes est significativement plus élevée que celle du groupe à deux temps de mesure (t=2 pour le panéliste 104 et t=2,5 pour le panéliste 107).

Afin de réaliser une synthèse des alertes générées par les cartes de contrôle de tous les descripteurs, le tableau 4-6 rassemble les détections observées pour chaque panéliste lorsque le niveau de performance était significativement inférieur à celui du groupe.

Tableau 4-6 : Résumé des alertes de chaque panéliste en fonction des temps de mesure (performance significativement moins élevée que celle du groupe)

temps	102	106	104	111	108	103	115	112	101	105	113	109	107	Somme	Moyenne
1	6	2	2	7	2						1			20	1.82
1.5	2	2		1	4	2	2							13	1.30
2	5	1	1	3		2			1					13	1.18
2.5	4		5	1	5									15	1.50
3		6	2	1	1									10	0.91
3.5			3	2				1		1				7	0.70
4			4		1									5	0.56
4.5			1											1	0.11
Somme	17	11	18	15	13	4	2	1	1	1	1	0	0		
Moyenne	4.25	2.75	2.57	2.14	1.63	0.5	0.5	0.17	0.14	0.14	0.14	0	0		

Chaque cellule contient le nombre de descripteurs pour lesquels le niveau de répétabilité d'un panéliste était significativement inférieur à celui du groupe. Les cellules grisées indiquent les temps de mesures auxquels chaque panéliste était absent.

Certains panélistes sont souvent détectés (sujets 102 et 104, par exemple), alors que le niveau de répétabilité d'autres panélistes n'est jamais inférieur à celui du panel, quel que soit le temps de mesure et le descripteur considéré (sujets 107 et 109). Le bilan sur l'ensemble des descripteurs met donc en évidence des disparités entre les niveaux des sujets. En outre, le nombre d'alertes décroît au fil des évaluations indiquant que le niveau de répétabilité des panélistes a tendance à s'homogénéiser au fil du temps ; les écarts de performance se réduisent.

3.1.5 Vérification de l'hypothèse de normalité des résidus

L'utilisation du modèle mixte suppose la normalité de la distribution des résidus du modèle. Cette hypothèse peut être discutable dans notre cas puisque les données utilisées sont des écart-types. Le test d'ajustement de Kolmogorov-Smirnov montre que la normalité est vérifiée pour seulement 16 descripteurs sur 30. La légitimité de l'utilisation du modèle mixte est donc remise en question dans près de la moitié des situations. Cependant, la non normalité mise en évidence est due, dans la majorité des cas, aux individus hors contrôle. Lorsque les valeurs correspondantes sont écartées, la normalité des résidus est vérifiée pour 25 descripteurs sur 30. De plus, bien que la normalité de la distribution soit rejetée par le test d'ajustement pour certains descripteurs, les graphiques de la distribution des résidus et de la droite de Henry (figure 4-12) montrent que l'écart à la normalité existe mais qu'il est relativement faible. Les données représentées ci-dessous correspondent à un descripteur représentatif de la plupart des cas de rejet de la normalité, dans lesquels la distribution est décentrée sur la gauche.

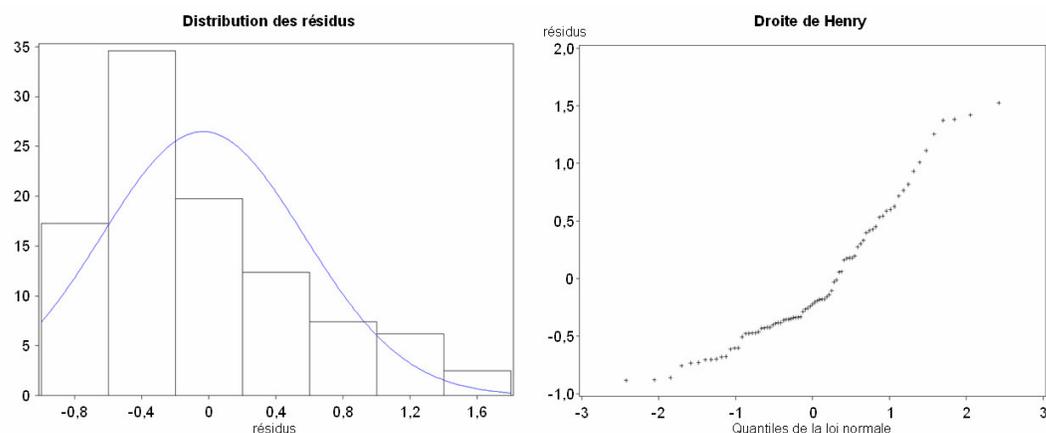


Figure 4-12 : représentations de la normalité des résidus pour le descripteur od_4 (modèle mixte avec effet fixe n.s. et structure de covariance RCL)

Afin d'éviter ce problème de non normalité des résidus, une autre possibilité consiste à transformer les données de façon à rendre la distribution normale, tout en présentant les résultats et les graphiques dans l'unité de mesure initiale. Dans le cas de l'indice étudié ici, la transformation de l'écart-type des résidus (opposé de l'indice REP), par sa racine carrée, avant de réaliser les analyses, permet de réduire à 5 le nombre de descripteurs pour lesquels la non normalité est mise en évidence. Cependant, l'interprétation des structures de covariance et des effets fixes sélectionnés est moins aisée parce qu'elle n'est pas directement reliée à l'indice de répétabilité. De plus, les analyses montrent que le modèle mixte sélectionné en utilisant l'indice de répétabilité transformé est rarement

différent de celui retenu sans transformation. La structure de corrélation est modifiée pour deux descripteurs (descripteur ap_1 : structure ID au lieu de la structure CS ; descripteur tx_4 : structure CS au lieu de la structure RCL) et la partie fixe du modèle change pour un seul descripteur (ap_5 : disparition de la pente significative). Les autres résultats sont identiques. Par conséquent, nous avons préféré garder les données non transformées afin de faciliter l'interprétation des résultats.

3.1.6 Conclusion

L'analyse de la partie fixe du modèle montre, sur ces données, que l'effet temps n'est pas significatif pour la majorité des descripteurs. Le groupe semble donc conserver un niveau de répétabilité constant au fil des évaluations. Ce résultat est probablement lié au fait que les premières évaluations des panélistes ne sont pas incluses dans les données étudiées parce que le protocole utilisé était différent pendant les deux premières années. En effet, les sujets n'avaient pas effectué de répétition des mesures donc la répétabilité ne pouvait pas être évaluée. Il est donc possible que l'amélioration du niveau du panel se soit principalement déroulée lors des premières sessions. Le niveau de répétabilité constant obtenu dans cette étude informe néanmoins sur l'évolution de la performance à long terme, montrant que le niveau atteint ne semble pas décroître, alors que des phénomènes de lassitude pouvaient être à craindre.

Le choix de la partie aléatoire du modèle correspond à la structure de covariance CS pour la majorité des descripteurs. Ce choix tend à indiquer que chaque panéliste possède son propre niveau de performance. Cependant, ces différences entre panélistes semblent disparaître pour les descripteurs visuels (structure ID) alors que la structure de covariance apparaît plus complexe pour les descripteurs trigéminaux et de texture (structures RCL et RCQ).

En outre, les choix de la structure de covariance et de l'effet du temps peuvent être discutés. Dans cette étude, le modèle le mieux ajusté est retenu pour chaque descripteur en fonction des résultats des tests de modèles emboîtés et des tests de Fisher. Ce type de sélection donne la possibilité de présenter les différents types de résultats qu'il est possible d'obtenir en utilisant le modèle mixte et les cartes de contrôles. Cependant, si un modèle différent est choisi pour chaque attribut, cela suppose que le panel et les panélistes peuvent évoluer selon un schéma totalement différent d'un descripteur à l'autre. Cette hypothèse forte semble assez peu réaliste en analyse sensorielle. Il serait, par exemple, difficile d'expliquer pourquoi la répétabilité des sujets évolue de la même manière pour un descripteur d'arôme (structure ID) alors que chaque panéliste possède sa propre pente et sa propre courbure (structure RCQ) pour un autre descripteur aromatique. La sélection d'un seul type de modèle pour tous les descripteurs, ou par type de descripteur, semble donc en général plus raisonnable.

Cette première étude longitudinale a permis d'illustrer la méthodologie proposée et de montrer la diversité des résultats envisageables. La deuxième étude repose sur des données couvrant une période de temps plus restreinte (environ 1 an) avec des évaluations plus rapprochées. Son analyse donne l'opportunité de présenter certains résultats différemment et d'approfondir la réflexion sur les causes éventuelles des évolutions observées.

3.2 Seconde étude

3.2.1 Modèle mixte : choix de la structure de corrélation

Pour chaque descripteur, les modèles mixtes correspondants aux quatre structures de covariances (ID, CS, RCL, RCQ) sont calculés. Le choix de la structure de covariance est effectué, soit en fonction des résultats des tests de modèles emboîtés (à l'aide de l'arbre de décision de la figure 4-3), soit en fonction du critère AIC. Les résultats complets sont disponibles en annexe 6. Une synthèse est présentée dans le tableau 4-7.

Tableau 4-7 : Synthèse du choix des structures de covariance selon les tests de modèles emboîtés et le critère AIC

DESCRIPTEUR	Probabilités des tests de modèles emboîtés			choix		Critère AIC
	RCQ vs. RCL	RCK vs. CS	CS vs. ID			choix
ap_brown	0,999	0,005	<0,001	RCL	=	RCL
ap_filmlayer	<0,001	0,014	<0,001	RCQ	=	RCQ
ap_red	0,998	0,254	<0,001	CS	=	CS
ap_transparent	0,784	0,100	<0,001	CS	≠	RCL
ap_yellow	0,006	<0,001	<0,001	RCQ	=	RCQ
m_astringent	0,100	0,527	<0,001	CS	=	CS
m_prickling	0,273	0,655	<0,001	CS	=	CS
m_saliva_stimul	0,094	0,752	<0,001	CS	≠	RCQ
o_alcohol	0,317	0,273	<0,001	CS	=	CS
o_animal	0,001	0,655	<0,001	RCQ	=	RCQ
o_cheese	<0,001	0,017	<0,001	RCQ	=	RCQ
o_ferment_fruit	0,168	0,040	<0,001	RCL	=	RCL
o_liquorice	0,206	0,273	<0,001	CS	=	CS
o_medecin	0,752	0,221	<0,001	CS	=	CS
o_mushroom	0,221	0,011	<0,001	RCL	=	RCL
o_roastpork	0,655	0,147	<0,001	CS	≠	RCL
o_ryebread	0,343	0,054	<0,001	CS	≠	RCL
o_treacle	<0,001	0,655	<0,001	RCQ	=	RCQ
t_burnt	0,752	0,002	<0,001	RCL	=	RCL
t_salt	0,752	0,089	<0,001	CS	≠	RCL
t_sour	0,480	0,752	<0,001	CS	=	CS
t_sweet	0,001	0,752	<0,001	RCQ	=	RCQ
ta_bitter	0,655	0,273	<0,001	CS	=	CS
ta_lavas	0,371	0,192	<0,001	CS	=	CS
				0	ID	0
				14	CS	9
				4	RCL	8
				6	RCQ	7

Ce tableau révèle tout d'abord que la structure ID n'est sélectionnée pour aucun descripteur. Les structures plus complexes (CS, RCL et RCQ) sont toujours nettement mieux ajustées aux données que la structure ID au regard des probabilités obtenues à chaque test de modèles emboîtés (<0.001). L'effet sujet est donc indiscutable dans cet ensemble de données, quel que soit le descripteur. Le choix de la nature de cet effet sujet est en revanche moins évident. Les structures de covariances CS, RCL et RCQ sont respectivement retenues pour 14, 4 et 6 des 24 descripteurs avec les tests de modèles emboîtés. Cela indique qu'un effet sujet constant (CS) serait mieux adapté aux données qu'un effet linéaire (RCL) ou quadratique (RCQ) pour la majorité des descripteurs. La structure de covariance ne semble en outre pas liée au type de descripteur, contrairement aux résultats de la première étude. D'après le critère AIC (dernière colonne du tableau 4-7), les structures de corrélations

retenues sont plus complexes que les structures retenues à l'aide des tests de modèles emboîtés pour cinq descripteurs. Ce résultat avait déjà été observé lors de la première étude. Les structures CS, RCL et RCQ sont alors respectivement sélectionnées pour 9, 8 et 7 descripteurs. Aucune structure ne semble donc faire l'unanimité selon l'indice AIC. Cependant, le graphique de la figure 4-13 montre que la diminution de la valeur de l'AIC (i.e. le gain d'ajustement du modèle aux données) est plus importante lors du passage de la structure ID à la structure CS que lors des évolutions suivantes de la structure de covariance. La structure CS est donc relativement bien adaptée aux données pour tous les descripteurs, bien qu'elle ne soit pas toujours la solution optimale. De plus, conformément aux remarques effectuées lors de l'analyse de la première étude, la sélection d'une seule structure pour tous les descripteurs est plus satisfaisant au niveau de l'interprétation sensorielle. Nous choisissons donc de sélectionner la structure CS pour tous les descripteurs afin de faciliter l'interprétation globale du comportement des panélistes sur l'ensemble des descripteurs. Il aurait néanmoins été possible de retenir une structure par descripteur afin d'analyser plus finement le comportement des panélistes dans chaque cas.

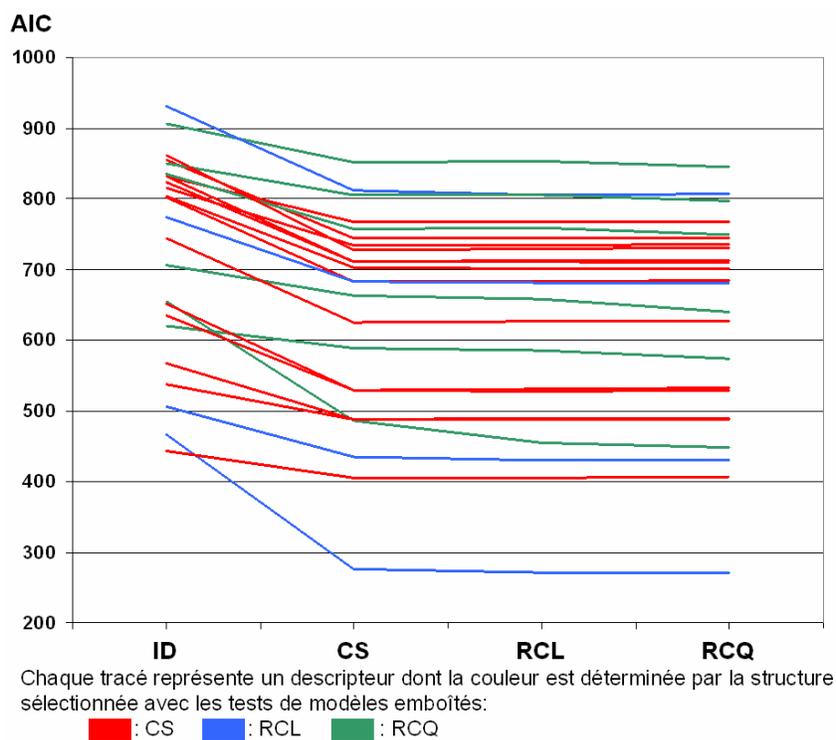


Figure 4-13 : Evolution de l'AIC de chaque descripteur selon la structure de corrélation

3.2.2 Modèle mixte : choix de la partie fixe

Pour chaque descripteur, la partie fixe du modèle est sélectionnée en utilisant l'équation 4-7. Cette partie du modèle permet de détecter la présence d'une évolution linéaire et/ou quadratique de la performance du groupe. Le meilleur modèle est sélectionné en fonction des résultats des tests de Fisher conditionnels. Les résultats complets sont disponibles en annexe 6 et une synthèse est proposée dans le tableau 4-8. Ce dernier classe les descripteurs en fonction de la significativité de leurs paramètres de pente et de courbure ainsi que du signe de ces deux paramètres. Dix-huit des 24

descripteurs présentent une évolution linéaire significative de la répétabilité et l'estimation de la pente est positive pour toutes ces variables. La répétabilité du groupe augmente donc au fil des évaluations pour la majeure partie des descripteurs. Le paramètre de courbure est significativement différent de zéro pour 19 descripteurs et son estimation est négative dans 18 situations, indiquant que la performance du groupe atteint (ou tend vers) un maximum. L'association d'une pente positive et d'une courbure convexe tend à montrer que le profil d'évolution le plus répandu (15 descripteurs sur 24) se caractérise par une amélioration du niveau de répétabilité au début de l'étude qui tend à ralentir puis à stagner ou à décroître au cours du temps. Une illustration de cette interprétation est présentée et commentée ultérieurement au cours de ce chapitre dans le paragraphe 3.2.4 (figure 4-14 et figure 4-15). La performance du panel reste stable au fil des évaluations pour deux descripteurs seulement (pente et courbure non significatives).

Tableau 4-8 : Résumé de la significativité et des signes des paramètres de l'effet fixe temps pour l'ensemble des descripteurs selon le modèle mixte retenu

		courbure					
		significative		non significative			
		positive	négative				
pente	significative	positive	0	15	3	18	18
		négative	0	0	0	0	
	non significative		1	3	2	6	
			1	18			24
		19		5			

NB : Chaque valeur représente le nombre de descripteurs correspondant à la situation

3.2.3 Modèle mixte : paramètres individuels

En fonction du modèle mixte sélectionné pour chaque descripteur, les paramètres individuels correspondants sont estimés. Puisque la structure CS a été retenue pour tous les descripteurs, seul l'ordonnée à l'origine de chaque panéliste est calculée. Une valeur positive indique que le niveau de répétabilité global d'un juge est supérieur à la moyenne du groupe sur l'ensemble des données de l'étude. Au contraire, si l'ordonnée à l'origine est négative pour un panéliste, le niveau de répétabilité de ce sujet est globalement moins élevé que le niveau du groupe. La significativité des paramètres individuels est mesurée en utilisant des tests de Student (équation 4-8). Le tableau 4-9 propose un récapitulatif des signes de ces paramètres pour chaque descripteur et chaque juge.

Tableau 4-9 : bilan de l'estimation des paramètres individuels

descripteur	45	36	5	220	50	203	32	73	4	219	2	37	1
ap_brown	++	++	-	-	-	+	++	-	+	-	-	--	--
ap_filmlayer	+	+	-	+	+	--	-	--	++	-	-	--	++
ap_red	++	+	++	--	-	+	+	+	+	-	-	--	+
ap_transparent	+	++	++	-	-	+	-	-	++	-	-	-	--
ap_yellow	+	++	++	+	-	++	+	--	++	-	-	--	--
m_astringent	++	++	+	+	-	++	++	-	--	-	-	--	-
m_prickling	++	++	+	+	+	--	+	+	--	+	-	--	-
m_saliva_stimul	++	++	+	+	+	+	--	-	--	+	-	--	--
o_alcohol	++	+	+	+	-	+	+	-	-	-	-	+	--
o_animal	+	+	++	+	+	+	-	-	+	-	-	+	--
o_cheese	++	+	+	+	-	-	-	+	+	-	+	+	--
o_ferment_fruit	++	++	++	-	-	-	+	+	+	-	+	-	--
o_liquorice	++	++	+	+	++	+	+	--	-	--	--	+	--
o_medecin	++	++	+	++	--	-	+	-	--	--	+	+	--
o_mushroom	++	++	++	-	++	-	--	-	-	-	-	++	--
o_roastpork	+	++	++	++	+	-	-	-	-	+	-	--	-
o_ryebread	++	+	++	+	+	+	+	--	-	-	-	-	--
o_treacle	++	+	++	+	+	-	+	-	-	-	--	-	--
t_burnt	++	++	+	++	+	++	+	-	-	--	--	--	--
t_salt	++	++	-	+	-	++	+	-	-	--	-	--	-
t_sour	++	++	-	-	-	++	++	-	--	-	--	-	+
t_sweet	++	++	-	+	-	++	+	+	-	-	--	-	-
ta_bitter	++	++	+	++	+	+	-	-	-	--	-	--	-
ta_lavas	++	++	++	++	+	+	-	-	-	-	--	--	-
++	19	17	10	5	2	6	3		3			1	1
+	5	7	9	13	10	10	12	5	5	3	3	5	2
-			5	5	11	6	7	15	11	16	12	6	7
--				1	1	2	2	4	5	5	9	12	14

++: estimation positive significative, +: estimation positive non significative
 --: estimation négative significative, -: estimation négative non significative

Pour chaque descripteur, ce tableau donne une information sur le niveau de répétabilité de chaque juge par rapport au groupe. Le diagnostic "+" indique que l'estimation du paramètre (ordonnée à l'origine) est positive pour un sujet. Dans le cas contraire, le diagnostic est "-". Lorsque le test du paramètre est significatif, le symbole du diagnostic est doublé ("++" ou "--"). Ces deux derniers diagnostics sont respectivement surlignés en vert et en rouge. Les panélistes sont triés en fonction du nombre de "--" (ordre croissant) puis du nombre de "++" (ordre décroissant). Cela permet de mettre en évidence les moins bons panélistes (à droite) et les meilleurs panélistes (à gauche) sur l'ensemble des descripteurs et des temps de mesure. Les juges 45, 36 et 5 sont plus répétables que le groupe pour un grand nombre de descripteurs, tandis que les panélistes 1, 37 et 2 obtiennent souvent un niveau de répétabilité significativement moins élevé que le panel sur l'ensemble de la durée de l'étude. Ce tableau permet également de mettre en évidence des comportements particuliers. Par exemple, le juge 1, dont le niveau de répétabilité est très souvent inférieur à celui du panel (14 diagnostics "--" et 7 diagnostics "-" sur 24 descripteurs), se situe parmi les dégustateurs les plus répétables concernant le descripteur ap_filmlayer.

3.2.4 Cartes de contrôle

Les paramètres utilisés pour construire les cartes de contrôle sont calculés en fonction du modèle mixte sélectionné pour chaque descripteur. Les résultats concernant un seul attribut sensoriel (ap_red) seront présentés. La partie fixe correspond à un effet quadratique et la structure de covariance est CS. La carte de contrôle au niveau du groupe montre l'évolution quadratique de la

performance (figure 4-14). Le niveau de répétabilité s'améliore lors des premières évaluations puis semble stagner à la fin de l'étude. Cette carte de contrôle permet également de faire apparaître les temps de mesure auxquels le niveau de performance du groupe est significativement différent de l'évolution modélisée par le modèle mixte. Pour le descripteur `ap_red`, le niveau de répétabilité du panel observé est significativement plus élevée que l'évolution modélisée à quatre temps de mesure ($t=4, 29, 30$ et 32) et significativement moins élevée à trois reprises ($t=13, 18$ et 34). Cependant, cette représentation ne réalise qu'une description de l'évolution de la performance. Les causes potentielles des variations de niveaux ne sont pas prises en considération. Une chute brutale du niveau de performance pourrait par exemple s'expliquer par la variabilité de l'espace produit ou par le changement des conditions expérimentales.

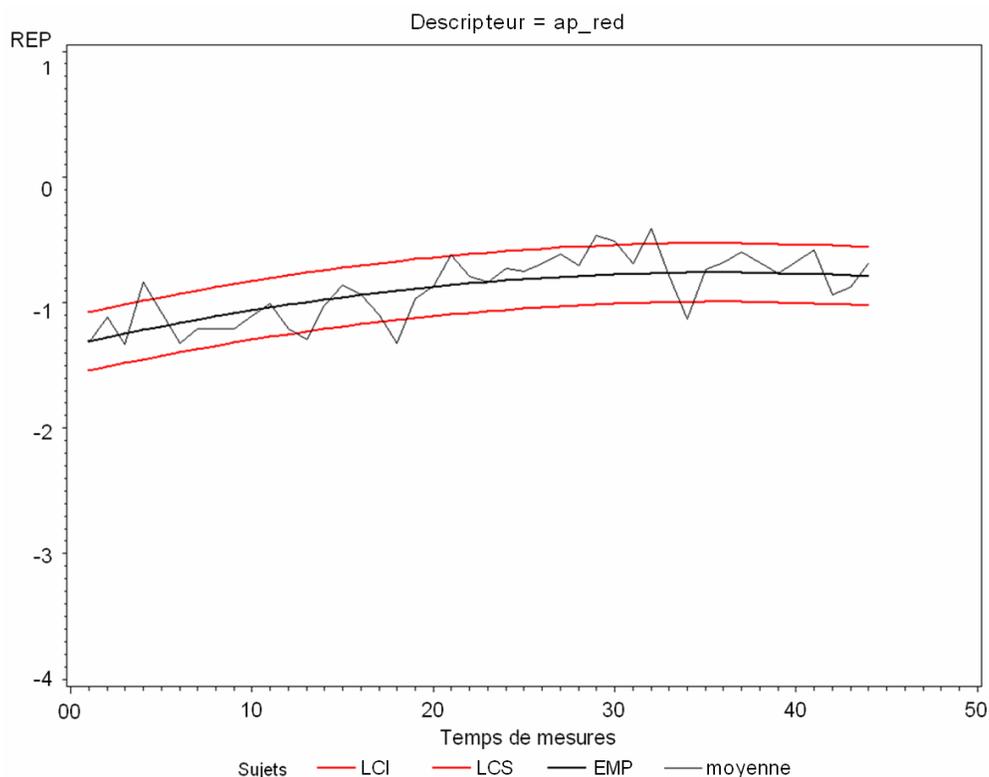
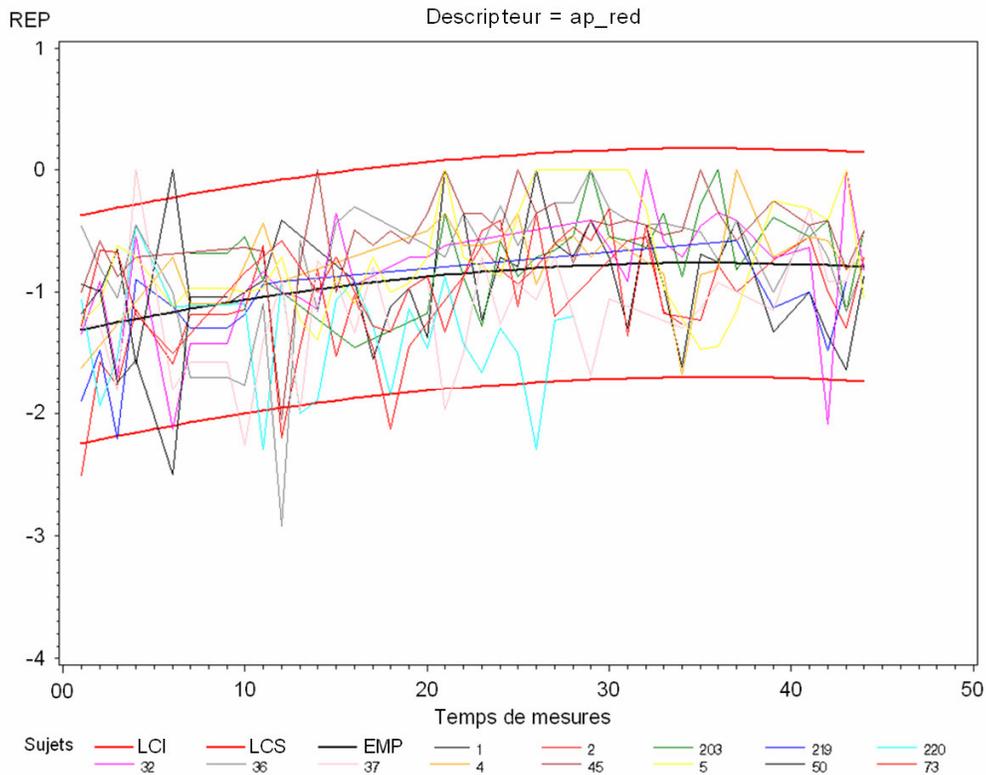


Figure 4-14 : carte de contrôle de l'évolution de la répétabilité du panel pour le descripteur `ap_red` (structure de covariance CS, effet temps quadratique)

La carte de contrôle des panélistes (figure 4-15) met en évidence les temps de mesure auxquels le niveau de répétabilité d'un ou plusieurs sujets est significativement différent de l'évolution du panel. L'intervalle de confiance à 95% dessiné sur la carte est relativement large (il s'étend de -0,4 à -2,3 à $t=0$). Cela signifie que les variations du niveau de répétabilité des panélistes sont importantes au fil des évaluations.



Afin de réaliser une synthèse des alertes générées par les cartes de contrôle, le tableau 4-10 indique, pour chaque panéliste et chaque temps de mesure, le nombre de descripteurs pour lesquels le niveau de performance était significativement plus faible que celui du groupe.

Le nombre de détection est très élevé pour certains panélistes (en particulier pour les sujets 1 et 37) et très réduit pour d'autres (panélistes 45, 36, 50 et 5). Le classement des juges effectué dans cette table est proche de celui du tableau 4-9. Par conséquent, les panélistes dont le niveau de performance général est inférieur à celui du groupe, correspondent aux sujets souvent considérés comme ponctuellement moins performants. Ce tableau permet également de détecter certains phénomènes particuliers. Par exemple, le sujet 32 est déclaré significativement moins répétable que le groupe pour 6 descripteurs au temps 41, alors que son nombre moyen d'alerte est de 0,62 descripteurs par temps de mesure. Le même phénomène se reproduit pour le sujet 2 au temps $t=12$. Cette information peut permettre à l'animateur de focaliser son attention sur ces données afin, éventuellement, d'expliquer ce résultat. Le nombre de détections à chaque temps de mesure varie entre 1 et 17 mais aucune structure spécifique ne se dégage. Cependant, on peut se demander si le nombre de détections, et de manière plus générale, si l'évolution de la répétabilité n'est pas lié à d'autres facteurs dépendants des produits ou des conditions de l'expérimentation à chaque temps de mesure. Par exemple, les manuels d'analyse sensorielle conseillent de présenter un nombre d'échantillons relativement réduit à chaque séance afin de ne pas altérer la concentration des panélistes (Amerine, Pangborn et al., 1965; AFNOR, 1999). Le nombre d'échantillons dégustés à

chaque séance est donc une source de variabilité potentielle, susceptible d'influencer le niveau de performance du panel. Bien que d'autres sources de variabilité puissent moduler le niveau de répétabilité, nous nous limiterons à l'étude de ce facteur dans le paragraphe suivant.

Tableau 4-10 : Résumé des alertes de chaque panéliste en fonction des temps de mesure (performance significativement moins élevée que celle du groupe)

temps	45	36	5	220	32	50	73	203	2	4	37	219	1	Somme	Moyenne
1				2					3		2	4		11	0,92
2				1							4			5	0,45
3			1						1		3	5	2	12	1,09
4				1	1	1	2		1		1	1	2	10	0,91
5				1	1	1	1		1	1	1	1	3	11	1,00
6					1	1		1				2	2	7	0,70
7					1	1		1				2	2	7	0,64
8					1	1		1				2	2	7	0,64
9				1			2	1	1	1	2	4	5	17	1,42
10				1				1	1		2	1	4	10	0,91
11	1	1		1	1			1	6	1	3	2		17	1,42
12		2		1							4			7	0,88
13			2	1	1		1		2		1			8	1,00
14					1				3		1			5	0,71
15				1				2	3		2		4	12	1,33
16							2	2	2		4		4	12	1,71
17				1					1		4		5	11	1,57
18											1		3	4	0,50
19									2		2		2	6	0,60
20			1		1		3	1			3			9	0,82
21							2	2		1	2		1	8	0,80
22								1					2	3	0,30
23			1							5	1		3	10	1,00
24									1				2	3	0,30
25				1				1	1	2	5		3	13	1,30
26									1		1		1	3	0,30
27				1						3	2		4	10	1,25
28					1						1		3	5	0,56
29					2			1			2		1	6	0,67
30					1			2						3	0,38
31									1				2	3	0,38
32		1			2		1	2	1	5			2	14	1,56
33							3	4		7	1		1	16	2,29
34								1	2		2		3	8	0,89
35		1	1					1		2	2		4	11	1,22
36									1			5	4	10	1,11
37									1	1	1	3	2	8	0,89
38													2	2	0,20
39										1				1	0,14
40										1			1	2	0,20
41					6				2	2	1		2	13	1,30
42								1	1	1	1	1	7	12	1,20
43													3	3	0,30
Somme	1	5	6	14	21	5	17	26	37	36	61	33	93		
Moyenne	0,03	0,13	0,15	0,58	0,62	0,63	0,65	0,84	0,90	1,16	1,61	1,74	2,38		

Chaque cellule contient le nombre de descripteurs pour lesquels le niveau de répétabilité d'un panéliste était significativement inférieur à celui du groupe. Les cellules grisées indiquent les temps de mesures auxquels les panélistes étaient absents.

3.2.5 Effet du nombre de dégustations par séance sur la répétabilité

Le nombre d'échantillons testés à chaque temps de mesure varie entre 6 et 48. Cependant, certains temps de mesure regroupent les données obtenues lors de plusieurs séances successives. Dans ce cas, la valeur prise en compte est le nombre moyen d'échantillons présentés lors d'une séance, afin d'obtenir un indice qui reflète la durée d'une épreuve. Cet indice varie entre 6 et 16. Afin de vérifier l'existence d'une relation entre le nombre de produits dégustés par séance et le niveau de répétabilité,

le coefficient de corrélation de Pearson est calculé entre ces deux variables. Pour chaque descripteur, les observations correspondent aux moyennes des valeurs obtenues à chaque temps de mesure. Les résultats (dernière colonne du tableau 4-11) montrent une corrélation négative entre les deux variables pour 19 des 24 descripteurs. Cette corrélation est significativement non nulle à 9 reprises. Parmi les 5 corrélations positives, une seule est significative. Il s'agit en outre de l'unique descripteur pour lequel le niveau de répétabilité tend à se détériorer au fil des évaluations. Il est par conséquent normal que la corrélation soit opposée. Le niveau de répétabilité a donc tendance, pour une partie des descripteurs, à diminuer lorsque le nombre de produits dégustés dans une séance augmente. L'hypothèse de la baisse de la concentration des sujets avec l'augmentation du nombre de dégustations est donc plausible. Afin de déterminer si cette relation peut parasiter l'évolution du niveau de répétabilité observée, le nombre de produits évalués à chaque séance est ajouté en covariable dans la partie fixe du modèle mixte. La structure CS est utilisée dans la partie aléatoire :

$$y_{ji} = \mu + nbobs + \theta_1 \cdot temps + \theta_2 \cdot temps^2 + \alpha_j + \varepsilon_{ji} \quad (4-12)$$

Où nbobs est déclaré en régresseur.

Puisque le calcul des carrés moyens est séquentiel (type I), ce modèle revient à déterminer la nature de l'effet du temps (quadratique, linéaire ou nul), après avoir épuré les données de l'effet du nombre d'observations. Le tableau 4-11 confronte, pour chaque descripteur, les estimations des effets fixes, lorsque la covariable "nbobs" est incluse ou non. Les descripteurs sont classés par ordre croissant de corrélation entre le nombre d'évaluations par séance et le niveau de répétabilité. Les valeurs surlignées en bleu et jaune indiquent respectivement les estimations significativement négatives et positives. Chaque colonne "fix" rappelle la nature de l'effet temps sélectionné (quadratique, linéaire ou constant). Les résultats montrent que les estimations des paramètres sont très proches dans les deux situations et que la modélisation de l'évolution de la répétabilité est modifiée pour trois descripteurs seulement (t_burnt, ta_bitter et o_mushroom). De plus, l'effet de la variable "nbobs" n'est pas significatif pour ces attributs. Les modifications engendrées sont donc discutables puisque la covariable ajoutée au modèle ne permet pas d'expliquer significativement les variations de niveaux de répétabilité. De manière générale, la prise en compte du nombre d'évaluation par séance ne modifie donc pas la modélisation de l'évolution de la performance. Notons enfin que l'effet de la covariable "nbobs" est significatif pour 5 descripteurs seulement. Par conséquent, bien que la corrélation entre le nombre de dégustations et la non répétabilité soit parfois significative, cela n'explique qu'une petite partie de la variabilité de la répétabilité et ne modifie pas l'évolution observée dans un premier temps.

Tableau 4-11 : Estimations des paramètres des modèles mixtes avec et sans la covariable "nbobs"

Descripteur	Modèle mixte sans covariable			Modèle mixte avec la covariable "nbobs"				corrélation*		
	fix**	Estimation			fix**	Estimation				
		Ord. O.	pente	courb.		nbobs	Ord. O.		pente	courb.
o_liquorice	t2	-0,961	0,022	-0,002	t2	-0,033	-0,775	0,016	-0,001	-0,379
o_treacle	t2	-1,273	0,011	-0,001	t2	-0,030	-0,983	0,009	-0,001	-0,377
o_ryebread	t2	-0,498	0,022	-0,003	t2	-0,025	-0,605	0,020	-0,002	-0,338
t_salt	t2	-1,071	0,010	0,000	t2	-0,015	-0,922	0,009	0,000	-0,293
o_cheese	t2	-0,263	0,014	-0,002	t2	-0,022	-0,581	0,008	-0,001	-0,342
m_prickling	t2	-1,196	0,014	-0,001	t2	-0,012	-1,070	0,013	-0,001	-0,336
o_saliva_stimul	t2	-1,192	0,012	0,000	t2	-0,016	-1,033	0,011	0,000	-0,317
o_ferment_fruit	t2	-0,517	0,014	-0,003	t2	-0,020	-0,800	0,013	-0,002	-0,268
ta_maggi_lavas	t2	-1,194	0,004	-0,001	t2	-0,013	-1,086	0,004	-0,001	-0,266
o_medecin	t2	-0,442	0,005	-0,003	t2	-0,005	-0,973	0,006	-0,001	-0,245
ap_filmlayer	t2	-0,808	0,013	-0,001	t2	-0,009	-0,921	0,011	0,000	-0,236
m_astringent	t2	-1,110	0,012	0,000	t2	-0,005	-1,057	0,011	0,000	-0,232
t_sour	t	-1,208	0,007		t	-0,010	-1,103	0,006		-0,230
o_alcohol	t2	-0,401	0,014	-0,003	t2	-0,026	-0,844	0,010	-0,002	-0,210
t_burnt	t2	-1,321	0,007	0,000	t	-0,007	-1,334	0,006		-0,196
ap_red	t2	-0,848	0,013	0,000	t2	0,003	-0,917	0,011	0,000	-0,102
o_animal	t2	-1,042	0,006	-0,001	t2	0,011	-1,360	0,009	-0,001	-0,055
ta_bitter	t	-1,348	0,005		cst	0,001	-1,369			-0,040
t_sweet	t2	-1,236	0,002	0,000	t2	0,008	-1,313	0,002	-0,001	-0,024
ap_brown	t	-0,907	0,006		t	0,006	-0,970	0,005		0,035
o_mushroom	t2	-0,609	0,014	-0,001	t	-0,007	-0,879	0,014		0,043
o_roastpork	cst	-1,318			cst	0,010	-1,409			0,114
ap_yellow	cst	-0,803			cst	0,000	-0,860			0,122
ap_transparant	t2	-1,025	-0,004	0,001	t2	0,052	-1,630	-0,003	0,001	0,395

* : coefficient de corrélation de Pearson entre le nombre moyen d'évaluations réalisées dans chaque séance par les panélistes et le niveau de répétabilité moyen du panel

** : choix de la partie fixe du modèle liée à l'effet du temps (cst=contants ; t=évolution linéaire ; t2=évolution quadratique)

■ : estimation positive significative

■ : estimation négative significative

3.2.6 Vérification de l'hypothèse de normalité des résidus

De même que lors de l'analyse de la première étude, la normalité de la distribution des résidus du modèle est évaluée à l'aide du test d'ajustement de Kolmogorov-Smirnov. Les résultats montrent que la normalité est très rarement vérifiée (seulement pour 4 descripteurs sur 24) et la suppression des individus hors contrôle améliore le résultat pour un seul descripteur. Cependant, l'aspect des distributions et des droites de Henry de chaque descripteur ne semble pas montrer de dissymétrie plus importante que dans la première étude (figure 4-16). La normalité de la distribution des résidus est pourtant rejetée pour une très large majorité des descripteurs, dont celui utilisé dans la représentation graphique. Cette augmentation du nombre de descripteurs qui ne vérifient pas la normalité de la distribution des résidus, peut être attribuée à une hausse de sensibilité du test d'ajustement, en partie liée au nombre de mesures utilisées dans chaque étude. En effet, dans la première étude, le nombre de données pour chaque descripteur était approximativement de 140, alors que la deuxième étude se base sur environ 400 observations pour chaque modèle.

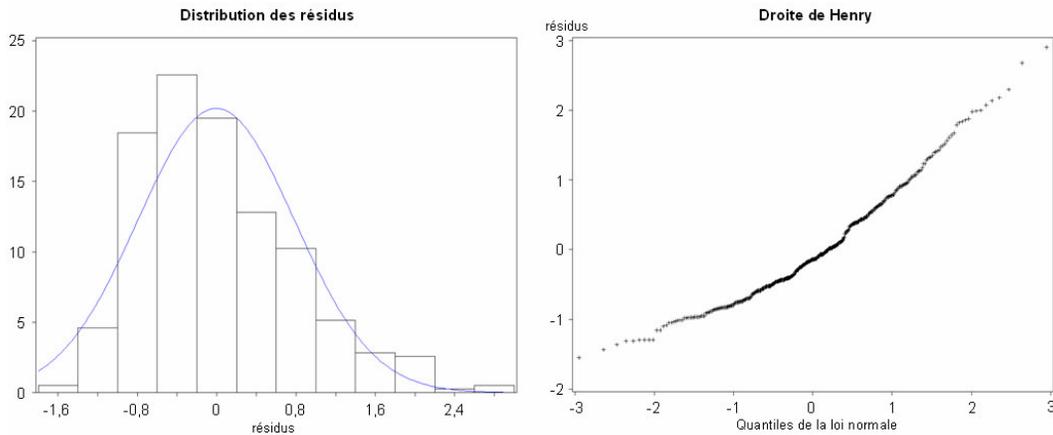


Figure 4-16 : représentations de la normalité des résidus pour le descripteur t_{burnt} (modèle mixte avec effet fixe quadratique et structure de covariance CS)

Comme pour la première étude, la transformation des données par leur racine carrée améliore sensiblement les résultats des tests de normalité et modifient peu le choix du modèle mixte de chaque descripteur.

3.2.7 Conclusion

Cette seconde étude met en évidence une évolution de la performance du groupe pour la majorité des descripteurs. Elle se caractérise généralement par une amélioration du niveau de répétabilité (pente positive) jusqu'à un maximum (paramètre de courbure négatif). Le niveau de performance reste ensuite constant ou diminue légèrement. Il est intéressant de noter que le profil d'évolution du groupe est relativement homogène entre les 24 descripteurs de l'étude. En effet, la pente est significativement positive dans 18 cas et la courbure est significativement négative pour 19 descripteurs. Ce profil d'évolution au niveau du groupe est donc relativement dominant, ce qui conforte l'existence réelle de cette structure. Cette remarque est aussi valable pour la première étude dans laquelle l'évolution du groupe, déterminée à l'aide du modèle mixte, était identique pour 25 des 30 descripteurs. En revanche, le résultat était très différent puisque le niveau de répétabilité restait constant au fil des évaluations pour ces descripteurs. L'étude de l'impact du nombre de dégustations par séance sur la répétabilité des panélistes montre l'existence d'une corrélation négative (significative pour plus d'un tiers des descripteurs) entre ces deux facteurs. Le nombre d'échantillons présentés par séance a donc tendance à diminuer de niveau de répétabilité. Cependant, ce facteur ne modifie pas l'évolution de la répétabilité du panel, exhibée à l'aide du modèle mixte.

La structure de covariance CS a été retenue pour l'ensemble des descripteurs parce qu'elle apportait un gain d'ajustement du modèle mixte aux données très important par rapport à la structure ID. Ce gain était ensuite plus faible lors du passage aux structures plus complexes (RCL ou RCQ). Cependant, la structure CS ne correspondait pas forcément à la structure la mieux adaptée d'après les tests de modèles emboîtés ou le critère AIC. Il existe donc un fort effet sujet, qui se maintient au fil des évaluations, mais la nature exacte de cet effet est nuancée selon les descripteurs (variations de pente ou de courbure des évolutions individuelles possibles).

Les estimations des paramètres individuels (ordonnée à l'origine), obtenues à l'aide du modèle mixte, mettent également en évidence des disparités entre les niveaux de répétabilité des panélistes. Le niveau global de certains sujets (45 et 36) est supérieur à la moyenne du panel quel que soit le descripteur considéré. A l'inverse, d'autres panélistes sont très fréquemment moins performants que le groupe (juges 37 et 1). Ces conclusions sont confirmées lors de l'analyse des cartes de contrôles ; les sujets déclarés peu performants d'après le modèle mixte correspondent aux juges dont le nombre d'alertes est élevé sur les cartes de contrôle.

4 Discussion

La méthodologie présentée dans ce chapitre est appliquée à l'évolution d'un indice de répétabilité, mais il est possible d'utiliser d'autres indicateurs. Notons cependant que la nature de l'indicateur sélectionné peut avoir un impact sur l'interprétation des résultats. Par exemple, un indice de discrimination ne dépend pas uniquement de la performance propre au panel mais également de la facilité à différencier les produits. Si l'espace produit est modifié au fil des évaluations, l'évolution du niveau de performance observé peut être attribuée indifféremment à la progression du panel ou à la variabilité des produits. Le suivi d'un indice d'accord pose un problème supplémentaire car le calcul de cet indice pour un panéliste implique les mesures des autres membres du jury. Les variations de la constitution du panel au cours du temps introduisent alors une autre source de variabilité. Il serait donc préférable de restreindre le suivi d'indices d'accord à des situations dans lesquelles la constitution du panel est relativement stable.

Lors de l'utilisation du modèle mixte, les deux exemples présentés dans ce chapitre montrent que la normalité des résidus n'est pas toujours respectée. Ce résultat est en partie lié à la nature de l'indice de performance retenu (écart-type des résidus d'un modèle d'ANOVA par sujet) dont la distribution ne suit pas la loi normale. La question est alors de savoir s'il est préférable de transformer les données afin d'obtenir une distribution satisfaisante des résidus, ou de conserver les données initiales afin d'obtenir une interprétation directe des résultats. Dans notre cas, l'indice retenu a la particularité d'être exprimé en unité d'échelle^{*}. Par conséquent, lorsque le modèle mixte sélectionné met en évidence un effet quadratique du temps pour cet indice, l'interprétation, par rapport aux notes fournies par les dégustateurs, est directe. En revanche, l'interprétation est plus complexe lorsque la même évolution est détectée sur des valeurs transformées. Dans notre cas, nous avons donc préféré ne pas transformer les données, en assumant le non respect de la normalité des résidus. Une autre possibilité consisterait à spécifier un modèle basé sur une loi de distribution mieux adaptée à l'indice utilisé.

Bien que ce chapitre ne présente que les résultats de deux études, des analyses ont également été menées sur trois autres ensembles de données longitudinales fournies par les centres ACTIA. Chacune présentait des caractéristiques différentes en terme de nombre de produits, de panélistes et

^{*} une valeur de 1,2, obtenue avec des données dont l'échelle de notation s'étendait de 0 à 10, signifie par exemple que chaque répétition s'écarte, en moyenne, de 1,2 points de la moyenne du produit correspondant, soit 12% de l'échelle (paragraphe 2.2 de ce chapitre)

de descripteurs impliqués. Chaque étude se déroulait sur un période de temps différente (entre trois mois et trois ans), avec des fréquences différentes (entre une semaine et trois mois). De manière générale, les résultats indiquent que le niveau de répétabilité tend à s'améliorer au fil du temps pour la majeure partie des descripteurs de chaque étude, au moins dans une première phase. Cet effet est plus marqué pour les études avec des temps de mesure rapprochés (<1mois). La corrélation entre le type de descripteur et la structure de covariance n'a été observée que lors de la première étude. Notons toutefois qu'il s'agissait de la seule étude dans laquelle les temps de mesures étaient aussi espacés (6 mois) et que l'univers produit était aussi constant au fil des évaluations. Ces deux particularités pourraient être liées aux résultats obtenus.

Enfin, cette méthodologie permet de décrire l'évolution du panel et des panélistes (modèle mixte), ainsi que de focaliser l'attention sur les données particulièrement éloignées de l'évolution générale (cartes de contrôles). En revanche, cette approche reste descriptive et ne permet pas de fournir d'explications aux variations observées. Des facteurs supplémentaires peuvent cependant être ajoutés dans la partie fixe du modèle mixte afin d'expliquer une partie de la variabilité observée. Ceci permet, par exemple, de vérifier que le niveau de répétabilité n'est pas lié au nombre de produits dégustés à chaque temps de mesure. Selon la disponibilité des informations à chaque temps de mesure, il serait également envisageable d'analyser l'effet de variables telles que la durée d'exposition des panélistes aux produits de l'étude, ou l'ordre de présentation des échantillons.

5 Conclusion

L'enjeu principal de ce chapitre était de mettre en valeur l'intérêt de l'utilisation du modèle mixte pour l'analyse de l'évolution de la répétabilité au cours du temps. Cette méthode tient compte simultanément des effets panel et panéliste dans un même modèle. Cette analyse est suivie par une représentation graphique basée sur la méthodologie des cartes de contrôle qui utilise les paramètres définis dans le modèle mixte. Cela permet en particulier de mettre en évidence des individus ponctuellement différents du groupe.

La méthodologie présentée dans ce chapitre permet de prendre en compte à la fois la variabilité au niveau du panel et au niveau de chaque panéliste dans le même modèle, en déterminant respectivement la partie fixe et la partie aléatoire. Pour chaque descripteur, le modèle mixte sélectionné donne des informations sur l'évolution du panel et sur le comportement des panélistes par rapport au groupe au fil des évaluations. Les cartes de contrôle de l'évolution du panel mettent en évidence les temps de mesure pour lesquels le niveau de répétabilité du groupe est significativement différent de l'évolution indiquée par le modèle mixte. Les cartes de contrôle au niveau du panéliste indiquent, à chaque temps de mesure, les panélistes dont le niveau de performance est significativement différent de celui du groupe.

La présentation de cette méthodologie a fait l'objet d'une publication acceptée dans *Journal of Sensory Studies* intitulées : " Evolution of the repeatability of a panel in sensory analysis: a mixed model and control chart approach" (Pineau, Chabanet et al., 2006). Elle présente la démarche proposée dans ce chapitre et les résultats relatifs à la première étude.

Chapitre 5 : Méta-analyse de la performance

1 Introduction

L'analyse simultanée d'un grand nombre d'études est appelée méta-analyse. Ce terme désigne une démarche générale d'analyse de vastes ensembles de données plutôt qu'une technique ou une méthodologie statistique particulière. Le mot "méta-analyse" a été introduit en sciences sociales par Gene Glass en 1976. Ce type de démarche s'est ensuite développée en médecine (Mullen et Suls, 1982) puis en sciences animales (Marchant et McGrew, 1991). A l'origine, la méta-analyse correspond à l'analyse simultanée des résultats de plusieurs études, publiées dans des revues scientifiques, qui portent sur le même type d'expérimentation. L'enjeu est de synthétiser des résultats, parfois contradictoires, pour obtenir une conclusion globale ayant une portée plus générale. Il peut s'agir de tester une hypothèse, comme l'effet d'un traitement médical, à partir des résultats de plusieurs expérimentations (Glass, 1976), afin de conclure valablement sur l'existence d'un effet. En analyse sensorielle, ce type d'analyse permettrait de définir des niveaux moyens de performance des panélistes et de quantifier l'effet de différents facteurs (type de produit, type de descripteur, temps d'entraînement...) sur ces niveaux. La méta-analyse peut aussi être utilisée pour mieux explorer et analyser collectivement des informations n'ayant qu'un intérêt secondaire à l'échelle d'une seule expérimentation. L'effet du sexe et l'âge des panélistes, ainsi que des variables "mineures" comme le niveau d'éducation peuvent alors être explorées.

L'analyse simultanée d'un grand nombre d'expérimentations requiert une attention particulière du fait de la nature des données. Une difficulté majeure réside dans le fait qu'il existe en général un grand nombre de données manquantes, ce qui réduit fortement la possibilité de recourir à des méthodes multidimensionnelles. D'autre part, contrairement à la planification d'une seule analyse, le dispositif rassemblant les méta-données n'a jamais été construit à l'avance. Il est donc très rarement équilibré ou orthogonal. Cela a parfois pour conséquence de ne pas pouvoir envisager les interactions entre certains facteurs. Par exemple, si tous les dégustateurs de moins de 25 ans correspondent exactement aux panélistes sans expérience en analyse sensorielle, l'interaction entre l'âge et le niveau d'expérience ne pourra pas être étudiée. De plus, la variabilité des observations entre les études (i.e. les profils sensoriels dans notre cas) est souvent plus importante que la variabilité induite

expérimentalement à l'intérieur des études (Sauvant, Schmidely et al., 2005). Il faut donc considérer cet élément avec attention lors de l'analyse de la variabilité intra-étude.

Après avoir défini une problématique, il est essentiel de s'intéresser aux quatre étapes suivantes :

- **Le codage des données**, de manière à pouvoir unifier la lecture des informations entre les différentes études. Il peut s'avérer utile de discrétiser certaines variables, comme l'âge, et de définir des classes.
- **Le filtrage des données**, afin de ne conserver que les études "recevables" en fonction des objectifs fixés. Par exemple, les études sans répétition seront exclues des calculs sur la répétabilité ; les études ne présentant pas, ou très peu d'information sur le facteur analysé seront également exclues.
- **La pondération des données**, qui permet d'accorder des poids plus ou moins forts à chaque étude en fonction de sa taille (nombre d'observations), de son adéquation avec la problématique ou de la qualité des données (équilibre des effectifs de chaque classe, représentativité des données...)
- **Le choix du modèle statistique**. La plupart des analyses utilisent le modèle linéaire (Hedges, Bushman et al., 1992). Les paramètres du modèle dépendent ensuite de la nature des données et de la problématique.

La représentation graphique des données permet en outre d'accompagner ces étapes et parfois de mettre en évidence des structures particulières qu'il faut alors prendre en considération.

Malgré son intérêt indéniable, la méta-analyse est à notre connaissance très peu développée en analyse sensorielle. La recherche des articles contenant le mot-clé "meta-analysis" dans la base de données ScienceDirect® donne une liste de 4224 articles parus depuis 1996. Lorsque le mot clé "sensory" est ajouté, le nombre d'articles est réduit à 16 et ne concerne que des études médicales. La recherche ne donne aucun article en combinant le mot clé "meta-analysis" avec "sensory analysis", "sensory science" ou "sensory study". Des études mettant en commun les données de plusieurs laboratoires existent néanmoins (Martin, Molimard et al., 2000; Pagès et Husson, 2001; McEwan, Hunter et al., 2002). Leur objectif est de comparer des panels qui ont réalisé des études sur un même ensemble de produits, dans des conditions ou des lieux différents. Ce type d'étude permet d'évaluer l'impact des conditions d'entraînement sur le niveau de performance. Il existe cependant un besoin de connaissance plus général sur les niveaux de performance moyens qui permettrait à l'animateur de panel de situer les performances de son panel par rapport à l'ensemble du domaine.

Le travail présenté dans ce chapitre applique les principes généraux évoqués ci-dessus à l'analyse des données de profil de la SensoBase. L'objectif est de définir des niveaux moyens de performance dans le domaine de l'analyse sensorielle pour plusieurs indices et d'étudier l'effet de plusieurs variables sur ces niveaux. Nous étudierons l'effet de cinq variables "intra-études" (âge, sexe, niveau scolaire, niveau d'expérience, motivation, type de descripteur), dont les niveaux varient à l'intérieur de chaque expérimentation d'un sujet à l'autre ou d'un descripteur à l'autre, et l'effet de cinq variables "inter-études" (fournisseur, type de produits, type d'échelle, pays d'origine, laboratoire public ou privé),

dont le niveau est constant à l'intérieur de chaque étude. Dans la littérature, la plupart de ces facteurs ont déjà fait l'objet d'expérimentations dédiées à chaque variable.

Age et sexe

Un grand nombre de travaux étudient d'effet du sexe et de l'âge sur les capacités de mémorisation et la sensibilité sensorielle. Dixon, Wahlin et al. (2004) mettent en évidence une diminution, lente mais significative, de la mémoire épisodique en fonction de l'âge, lors d'une étude longitudinale réalisée sur des personnes âgées de 54 à 94 ans. Les auteurs observent également une tendance pour les femmes à obtenir de meilleurs résultats que les hommes. Dans une expérience sur des tests de mémoire olfactive, Choudhury, Moberg et al. (2003) concluent que les performances obtenues sont plus élevées pour les femmes que pour les hommes, et que ces performances diminuent pour chaque sexe en fonction de l'âge. Larsson, Lovden et al. (2003) obtiennent des résultats similaires sur une expérience sur la mémoire olfactive et la mémoire verbale. Thomas-Danguin, Rouby et al. (2003) réalisent plusieurs tests olfactifs sur 121 personnes âgées en bonne santé. Ils n'observent aucun effet de l'âge ou du sexe lors des épreuves de détection supraliminaire et des épreuves de discrimination. En revanche, les performances des femmes se révèlent significativement plus élevées lors de deux tests de détection d'odeur et un test d'identification. Une diminution significative de la capacité d'identification avec l'âge est également détectée. Concernant la sensibilité gustative, Hummel, Futschik et al. (2003) montrent une baisse de la sensibilité trigéminal pour les personnes plus âgées mais aucun effet lié au sexe. Une autre étude portant sur la capacité à identifier des produits à l'aveugle met en évidence un effet de l'interaction $\text{age} \times \text{sexe}$ qui révèle une supériorité des femmes sur les hommes dans le groupe de personnes les plus âgées uniquement (Murphy, 1985). En résumé, le niveau de performance, mesuré pour différentes épreuves liées à la mémoire, semble meilleur chez les femmes que chez les hommes et tend à diminuer avec l'âge dans toutes les études. Concernant les capacités sensorielles, les seuils de sensibilité et les capacités d'identification de différents composés ont tendance à baisser avec l'âge. Le niveau de performance atteint dans ces épreuves est parfois plus élevé pour les femmes. Cependant, ces différences ne persistent pas dans les épreuves de détection supraliminaire et de discrimination. Notons en outre que la plupart de ces études sont réalisées avec des sujets relativement âgés et peu de données sont disponibles concernant les personnes entre 20 et 50 ans. L'effet de l'âge et du sexe sur les performances des panélistes en analyse sensorielle sera discuté dans ce chapitre, en utilisant les données de dégustateurs âgés de 16 à 81 ans. Il n'aurait pas été aisé de répondre à cette question avec un petit nombre de profils sensoriels en raison des nombreux facteurs psychologiques qui peuvent influencer le niveau de performance, comme la motivation, la concentration, le stress, la faim et la soif (Pangborn, 1981; Lawless et Heymann, 1998). Le grand nombre d'études collectées dans la SensoBase va permettre de limiter l'effet de ces co-facteurs et d'obtenir des résultats plus puissants que l'analyse d'un petit nombre d'études.

Niveau d'expérience

L'effet du niveau d'entraînement a souvent été étudié en analyse sensorielle lors de comparaisons de panels entraînés et non entraînés, ou de panels avant et après entraînement. Certaines publications ne mettent en évidence aucun effet positif de l'entraînement sur les performances. En particulier, le niveau de discrimination ne serait pas amélioré après une phase d'apprentissage sensoriel (Roberts et Vickers, 1994; Gou, Guerrero et al., 1995; Moskowitz, 1996). Les auteurs émettent l'hypothèse que la signification des descripteurs n'est pas identique pour des dégustateurs entraînés ou non. Les sujets non entraînés n'ont effectivement pas appris à associer une sensation gustative particulière à un terme descriptif précis. Ils peuvent alors confondre les descripteurs ou associer plusieurs stimuli sensoriels à un seul terme. Cependant, la majorité des études observent une augmentation significative du niveau de discrimination du groupe (Clapperton et Piggott, 1979; Moskowitz et Jacobs, 1988; Sawyer, Cardello et al., 1988; Labbe, Rytz et al., 2004) et du nombre de descripteurs pour lesquels une différence significative entre les produits est perçue (Moskowitz, Jacobs et al., 1985; Wolters et Allchurch, 1994). L'entraînement contribue aussi à accroître le niveau de répétabilité et d'accord des panélistes (Roberts et Vickers, 1994; Chollet et Valentin, 2001). Les panélistes, dont les données sont rassemblées dans la SensoBase, sont tous des sujets entraînés puisqu'ils ont au moins participé à une étude de profil, mais leur niveau d'expérience varie. Certains participent pour la première fois à une étude sensorielle tandis que d'autres ont intégré un panel depuis plusieurs années. L'analyse de l'effet du niveau d'expérience donne la possibilité de savoir si l'accroissement du niveau de performance se poursuit au fil des participations.

Type de descripteur

L'effet du type de descripteur sur la performance des panélistes est très rarement étudié. A notre connaissance, une seule étude a été publiée à ce sujet. Elle compare l'évolution du style de notation entre plusieurs séances d'évaluation en fonction du type de descripteur (Moskowitz, 2000). Les résultats suggèrent que les panélistes ont tendance à réduire leur plage de notation pour les descripteurs d'arôme et de saveur alors que ce n'est pas le cas pour des descripteurs de texture. La variété des conditions d'acquisition (nature de l'entraînement, type de produits, choix des descripteurs...) représentées dans la SensoBase offre la possibilité de mesurer l'influence du type de descripteur sur le niveau de performance avec un niveau de confiance relativement élevé.

Niveau d'éducation

Plusieurs études récentes s'intéressent à l'effet du niveau d'éducation sur la consommation de produits alimentaires (Magnusson et Koivisto Hursti, 2002; Aranceta, Perez-Rodrigo et al., 2003; Bäckström, Pirttilä-Backman et al., 2003; Trondsen, Braaten et al., 2004) ou sur les mécanismes d'intention d'achat (Mucci, Hough et al., 2004). En revanche, très peu d'analyses sont menées sur l'impact du niveau scolaire sur la performance lors d'une épreuve de profil sensoriel. Muller, Heymann et al. (2003) étudient l'influence du niveau d'alphabétisation lors de la réalisation d'une épreuve sensorielle descriptive. Ils observent qu'un niveau élevé d'alphabétisation permet de réduire le temps d'entraînement. Ce facteur n'influence cependant pas le niveau d'accord entre les panélistes.

L'utilisation de la SensoBase va permettre de vérifier ce type de résultat sur un ensemble de données plus vaste.

Variables inter-études

Les articles qui abordent l'analyse des variables que nous qualifions d' "inter-études" (fournisseur, type de produits, type d'échelle, pays d'origine et laboratoire public ou privé) sont assez rares car elles nécessitent l'agrégation de plusieurs profils sensoriels réalisés dans des conditions diverses. Par conséquent, les expérimentations dédiées à ce type d'analyse reposent généralement sur des dispositifs complexes et sont réalisées sur des périodes relativement longues.

Plusieurs études s'intéressent aux différences entre pays, mais elles concernent le plus souvent l'analyse des préférences alimentaires des consommateurs (Peterson et Jolibert, 1995; ESN, 1996; Rousset-Akrim, Martin et al., 1996) ou l'analyse des comportements d'achat (Li, Monroe et al., 1994; Verlegh et Steekamp, 1999). L'analyse de la performance des panélistes selon les pays est à notre connaissance très peu abordée. Une étude discute cependant la variabilité de la dispersion des notes entre le Japon, la Corée et les Etats-Unis (Yao, Lim et al., 2003). Les résultats montrent que les japonais utilisent une portion plus réduite de l'échelle que les américains. Cet article étudie également les différences de dispersion selon le type d'échelle. Les conclusions indiquent que la dispersion des notes est généralement plus importante sur les échelles continues que sur les échelles discrètes. Ce résultat n'est cependant pas vérifié dans tous les pays. Aucune étude concernant l'effet du type de laboratoire ou du type de produits ne semble avoir été publiée.

L'intérêt de la SensoBase est donc particulièrement notable pour répondre à des questions relatives à la variabilité inter-études. Elle permet en effet de rassembler des données d'horizons très diverses et ainsi de couvrir différents niveaux pour chaque variable étudiée.

2 Matériel et méthode

2.1 Les données SensoBase

Les données utilisées dans ce chapitre proviennent de la SensoBase (fonctionnement décrit dans le chapitre 3). Entre janvier 2005 et août 2006, 474 jeux de données provenant de 35 laboratoires sensoriels différents (26 français et 9 étrangers) ont été collectés. Cela représente 9 275 descripteurs, 1 502 panélistes différents et 3 153 produits, soit un total de 3 283 418 notes d'intensité. Parmi cet ensemble de données, les centres ACTIA (i.e. sept laboratoires français), partenaires de la thèse, ont fourni 115 études, ce qui représente un volume de 19,4% des notes enregistrées. En outre, il est important de noter que certains fournisseurs (centres ACTIA ou non) ont envoyé un nombre d'études très important par rapport aux autres fournisseurs. Afin d'obtenir des résultats qui ne soient pas trop influencés par quelques organismes seulement, il a donc été décidé de limiter le nombre d'études à 30 par fournisseur. Une autre approche serait de donner le même poids à chaque fournisseur dans les analyses, mais cela entraînerait une surpondération considérable des données des sociétés qui n'ont envoyé que quelques études. Nous avons donc décidé de garder un poids proportionnel au

nombre de données fournies dans la limite de 30 profils sensoriels par organisme. La sélection des sous-ensembles de 30 études s'effectue de manière aléatoire. Les analyses menées dans ce chapitre reposent donc sur une sélection de 329 profils sensoriels, dont les caractéristiques sont présentées dans la figure 5-1 et le tableau 5-1. Les distributions du nombre de panélistes, produits, descripteurs et répétitions par produit sont respectivement présentées dans les figures 5-1.a, 5-1.b, 5-1.c et 5-1.d. Une étude de profil sensorielle correspondant à la médiane des observations serait composée de 12 panélistes, 5 produits, 26 descripteurs et 2 répétitions. Les valeurs des moyennes sont supérieures aux médianes en raison d'un petit nombre d'études très volumineuses qui étirent les distributions vers la droite. Ces études sont donc en partie responsables de la dissymétrie des quatre distributions. Chaque distribution est unimodale, à l'exception de celle concernant le nombre de panélistes, où les valeurs 10 et 12 sont plus fréquentes que les valeurs 9, 11 ou 13. Ce résultats pourrait être lié aux recommandations écrites dans les textes normatifs qui préconisent en général la sélection de 10 ou 12 sujets (AFNOR, 1993; ISO, 1993).

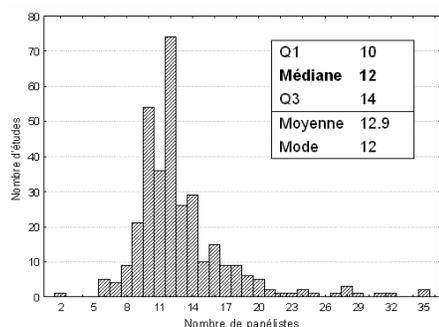


Figure 5-1.a : distribution du nombre de panélistes par étude

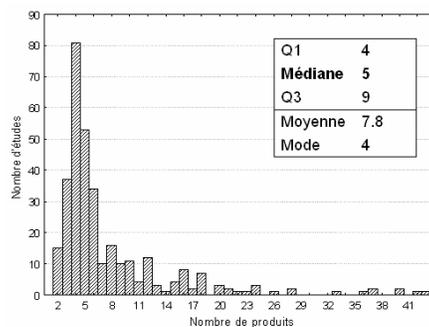


Figure 5-1.b: distribution du nombre de produits par étude

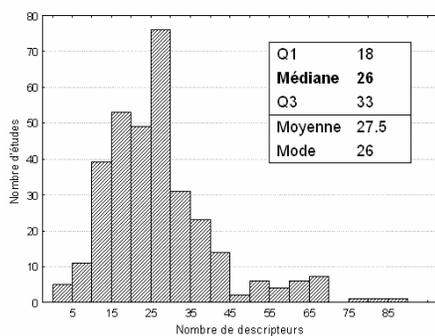


Figure 5-1.c: distribution du nombre de descripteurs par étude

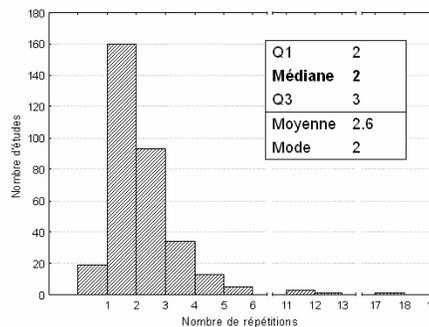


Figure 5-1.d: distribution du nombre moyen de répétitions (par produit) par étude

2.2 Variables intra-étude

2.2.1 Description des variables

L'âge, le sexe, le niveau d'expérience en analyse sensorielle et le niveau d'éducation sont quatre variables qui caractérisent le panéliste et dont nous souhaitons mesurer l'impact sur les niveaux de performances. Les informations sur le niveau d'expérience en analyse sensorielle et le niveau d'éducation ont été renseignées par moins d'un tiers des fournisseurs et représentent seulement 15% des études (tableau 5-1). Bien que l'âge et le sexe soient des variables que les fournisseurs doivent obligatoirement renseigner, quelques études ne comportant pas ces informations ont exceptionnellement été acceptées dans la SensoBase. La variable "type de descripteur" a également été incluse dans les analyses, bien qu'elle ne soit pas panéliste dépendante.

Tableau 5-1 : Nombre de jeux de données et de fournisseurs selon la variable analysée

Variables	Nombre d'études	Nombre de fournisseurs
Age	324	35
Sexe	328	35
Niveau d'expérience	51	13
Niveau d'éducation	49	11
Type de descripteur	329	35

2.2.1.1 Unité expérimentale

L'objectif de ce travail est d'apporter de la connaissance sur les variables qui pourraient avoir un impact sur le niveau de performance des panélistes. Par conséquent, l'unité expérimentale devrait être le panéliste. Cependant, certains sujets ont participé à plusieurs études sensorielles, donc leur âge et leur niveau d'expérience peut évoluer entre les études. L'unité expérimentale correspond alors au panéliste à un temps de mesure (i.e. une étude sensorielle). Cela signifie qu'une personne contribuera à un nombre de degrés de liberté égal à son nombre de participations à une étude sensorielle. Cette définition de l'unité expérimentale a également été retenue parce que les conditions d'entraînement et la variabilité des produits changent d'une étude à l'autre. Il est donc préférable de ne pas moyenner des indices de performances issus de différentes études pour un panéliste donné. Par conséquent, cette unité expérimentale sera également considérée lors de l'analyse de l'effet du sexe et du niveau d'éducation, bien que leur niveau ne varie pas entre les études pour chaque sujet. La variable "type de descripteur" occupe un statut particulier puisque son niveau (apparence, texture, saveur...) n'est pas le même pour tous les descripteurs d'un profil sensoriel. Dans ces conditions, l'unité expérimentale est le panéliste dans une étude pour chaque type de descripteur. Par rapport aux autres variables analysées dans ce chapitre, ce changement revient à multiplier le nombre de degrés de liberté par le nombre moyen de catégories de descripteurs.

2.2.1.2 Age

L'âge des panélistes varie entre 16 et 81 ans. La distribution des niveaux de cette variable, présentée dans la figure 5-2 montre que la majorité des personnes ont un âge compris entre 35 et 55 ans. Les personnes âgées de 20 à 25 ans sont cependant très représentées. Ce phénomène s'explique par le fait que certaines études ont été réalisées avec une majorité d'étudiants. L'analyse de l'effet de l'âge est réalisée en constituant trois groupes de panélistes : les sujets de moins de 30 ans (30-), les sujets entre 30 et 45 ans (30-45) et les sujets de plus de 45 ans (45+). Les effectifs de chaque classe sont respectivement 586, 1202 et 1886.

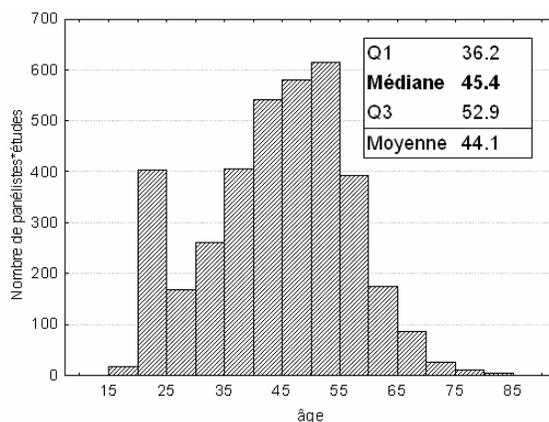


Figure 5-2 : Distribution des panélistes en fonction de l'âge

2.2.1.3 Sexe

Les hommes et les femmes ne sont pas représentés de manière équilibrée dans les données. Seulement 23,3% des panélistes sont des hommes. La distribution du pourcentage d'hommes dans chaque étude est présentée dans la figure 5-3. Elle met en évidence un nombre important d'études (plus de 100) dans lesquelles les panélistes sont uniquement des femmes (première barre de la distribution = 0% d'homme). Lors de l'analyse de l'effet du sexe, ces études ne seront pas prises en compte parce qu'elles n'apportent aucune information sur la comparaison entre les hommes et les femmes. Les sujets féminins seront codés F et les sujets masculins M dans les analyses.

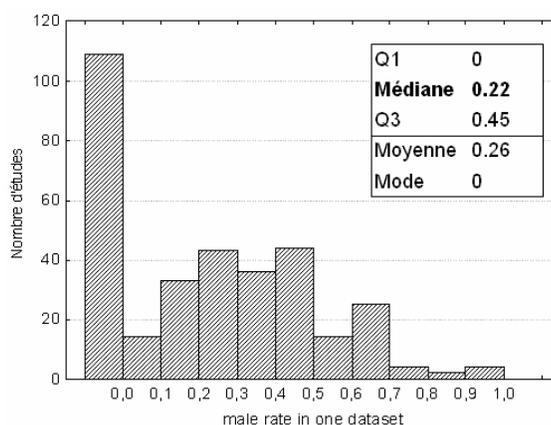


Figure 5-3 : Distribution des proportions d'homme par étude

2.2.1.4 Niveau d'expérience

Le niveau d'expérience est codé en trois catégories: "EXP 0" pour les panélistes qui participent à la réalisation d'un profil sensoriel pour la première fois, "EXP 1-3" pour les panélistes ayant entre 1 et 3 ans d'expérience et "EXP 3+" pour les personnes plus expérimentées. Ces catégories, regroupant respectivement 278, 180 et 228 sujets, ont été déterminées ainsi afin d'obtenir des effectifs suffisamment importants et relativement équilibrés selon les classes.

2.2.1.5 Niveau d'éducation

Le niveau d'éducation est directement codé par les fournisseurs de données selon trois catégories: éducation primaire (avant le collège), éducation secondaire (collège et lycée) et éducation supérieur (après le baccalauréat). Puisque la première catégorie ne concerne que 9 panélistes de la SensoBase, les données correspondantes sont exclues de l'analyse. La seconde catégorie, appelée "Bac-", représente 123 individus et la troisième catégorie, appelée "Bac+", rassemble 284 panélistes.

2.2.1.6 Type de descripteur

Les descripteurs sont classés en cinq catégories : apparence (couleur, forme...), texture (croquant, granuleux...), saveur (sucré, salé, acide, amer), odeur (fraise, caramel...) et arôme (lorsque le produit est mis en bouche). La figure 5-4 montre que les descripteurs visuels sont moins représentés que les autres, donc moins utilisés dans les données. Il est important de noter que chaque catégorie n'est pas représentée dans chaque étude. Certains profils peuvent avoir été réalisés avec deux catégories de descripteurs seulement. Cet aspect peut avoir un impact lors de la comparaison des types de descripteurs.

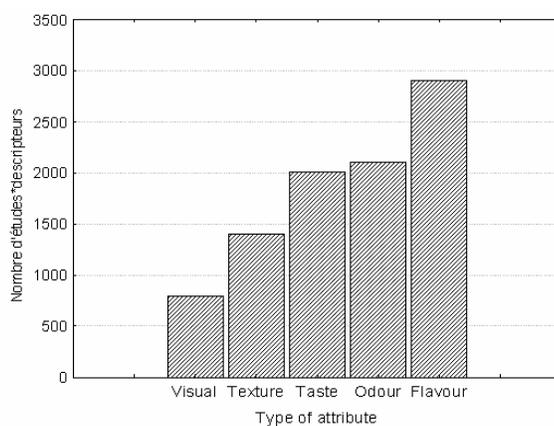


Figure 5-4 : Distribution des types de descripteurs

2.2.2 Indices de performance

En analyse sensorielle, la notion de performance est généralement scindée en trois aspects : la répétabilité, la discrimination et l'accord (ISO, 1993). Un indice de performance, calculé pour

chaque sujet et chaque descripteur, a donc été choisi pour chaque aspect de la performance. Deux de ces indices sont basés sur un modèle d'ANOVA au niveau individuel, réalisé pour chaque descripteur :

$$Score = produit + erreur \quad (13)$$

- L'**indice de répétabilité**, noté ETres, est la racine carrée de la variance de l'erreur issue du modèle (13). Cela représente l'écart-type moyen entre les répétitions pour chaque produit. Un indice faible indique que les notes des répétitions d'un produit étaient très proches, donc que le niveau de répétabilité est élevé. Afin d'obtenir des valeurs comparables d'une étude sensorielle à l'autre, toutes les notes sont transposées sur des échelles bornées entre 0 et 10.

- L'**indice de discrimination**, r_{CMprod} , correspond au rapport du carré moyen de l'effet produit sur la somme des carrés moyens de l'erreur et de l'effet produit (Fisher, 1946):

$$r_{CMprod} = \frac{CM_{prod}}{CM_{prod} + CM_{erreur}} \quad (14)$$

Comparé à la statistique de Fisher de l'effet produit, F_{prod} , égale au rapport CM_{prod}/CM_{erreur} , l'indice r_{CMprod} n'est pas sensible aux très faibles valeurs du carré moyen de l'erreur. En effet, la valeur de F_{prod} peut être très élevée dès lors que CM_{erreur} est très faible. Dans le cas particulier où $CM_{erreur}=0$, la statistique de Fisher de l'effet produit ne peut pas être calculée alors que l'indice proposé atteint la valeur maximale de 1.

- L'**indice d'accord** est le coefficient de corrélation de Pearson, entre les notes des produits d'un sujet et la moyenne des notes des produits des autres panélistes. Une valeur élevée de cet indice indique que le panéliste est en accord avec le groupe.

La performance de chaque sujet est donc caractérisée par un indice de répétabilité, un indice de discrimination et un indice d'accord pour chaque descripteur de chaque profil. Etant donné que l'unité expérimentale est le panéliste au sein d'un profil, les données analysées correspondent à la moyenne des indices de performance sur l'ensemble des descripteurs. Lorsque l'analyse porte sur le type de descripteur, les indices sont moyennés par type de descripteur.

2.2.3 Modèle d'analyse

Avant d'analyser l'effet de chaque variable, les niveaux moyens des trois indices de performance sont calculés afin d'obtenir des niveaux généraux de référence. Le même type d'analyse est ensuite exécuté pour chacun des trois indices et pour chaque variable (âge, sexe, niveau d'expérience, niveau d'éducation et type de descripteur). Le modèle d'ANOVA utilisé est:

$$\text{Indice de performance} = \text{profil} + \text{variable} + \text{profil*variable} \quad (15)$$

Où "profil" (effet aléatoire) représente le niveau moyen propre à chaque profil sensoriel, "variable" le niveau moyen de chaque catégorie de la variable d'intérêt, et "profil*variable" l'interaction entre les deux effets principaux.

Il est important de noter que le nombre de panélistes n'est pas obligatoirement équilibré entre les catégories d'une variable dans chaque profil sensoriel. Par exemple, un panel peut être constitué de 11 femmes et d'un seul homme. Dans ces conditions, l'estimation de l'effet de la variable (sexe, dans le cas présent) peut être biaisée. En effet, lorsque la catégorie faiblement représentée dans une étude atteint un niveau inhabituel (élevé ou faible) par rapport aux autres études incluses dans les calculs, l'écart entre les catégories est anormalement élevé pour l'ensemble des panélistes du groupe (i.e. les 12 sujets). Afin d'éviter ce problème, la répartition des effectifs selon les catégories est prise en compte en affectant la pondération suivante à chaque unité expérimentale :

$$w_i = \frac{\text{effectif minimum observé dans une catégorie}}{\text{effectif moyen par catégorie}} \quad (16)$$

Un exemple de calcul des poids, w_i , est présenté dans la figure 5-5:

Catégories d'âge	-30	30-45	+45	poids, w_i (par sujet)
Etude n°1	6	6	6	6/6 = 1
Etude n°2	8	4	6	4/6 = .67
Etude n°3	14	3	1	1/6 = .17
Etude n°4	11	7	0	7/9 = .78

Les données du tableau sont des nombres de panélistes. Chaque étude comprend 18 sujets répartis dans trois catégories.
 Nombre moyen de sujet par catégorie : $N_{theo} = 18/3 = 6$
 Si seulement deux catégories sont représentées, $N_{theo} = 18/2 = 9$

Figure 5-5 : Exemple de calcul des poids affectés aux panélistes d'une étude

Le poids affecté à chaque panéliste d'un profil est égal à 1 lorsque le nombre de sujets de cette étude est parfaitement équilibré entre les catégories (profil 1, figure 5-5). Le poids total de ce profil est alors égal au nombre de panélistes de l'étude. Au contraire, lorsqu'une catégorie est faiblement représentée, le poids affecté à chaque sujet est diminué (profils 2 et 3). Le poids total du profil est alors égal au nombre de panélistes présents dans la catégorie peu représentée multipliée par le nombre de catégories. Si une catégorie n'est représentée par aucun sujet, les calculs sont effectués en considérant que le nombre de catégories est moins élevé pour cette étude (profil 4 de la figure 5-5). Lorsqu'une seule catégorie est représentée, le profil est exclu de l'analyse.

L'analyse de chaque variable se termine par un test de comparaison multiple entre les moyennes des catégories afin de mettre en évidence les différences entre les niveaux.

Les effets des variables âge, sexe, niveau d'éducation et niveau d'entraînement sont également testés en fonction du type de descripteur en utilisant le modèle suivant :

$$\text{Indice de performance} = \text{Profil} + \text{Variable} + \text{TypeDescr} + \text{Profil*Variable} + \text{Profil *TypeDescr} + \text{Variable*TypeDescr} \quad (17)$$

L'unité expérimentale est alors le panéliste pour chaque type de descripteur lors d'un profil sensoriel. "Variable" représente l'effet d'une des quatre variables individuelles (âge, sexe...), "Profil", l'effet lié à chaque étude et "TypeDescr", l'effet du type de descripteur. Lorsque l'interaction "Variable*TypeDescr" est significative, cela signifie que l'effet de la variable est différent d'un type de descripteur à l'autre. Dans ce cas, l'analyse de cette variable sera également effectuée par type de descripteur.

Les analyses ont donné lieu à la création d'un programme SAS[®] permettant d'automatiser les calculs pour chaque variable. Les calculs reposent essentiellement sur l'utilisation de la procédure GLM avec l'ordre WEIGHT.

2.3 Variables inter-études

2.3.1 Description des variables

Les variables qualifiées d'inter-études conservent le même niveau au sein d'une étude. En revanche, les niveaux peuvent différer d'un profil à l'autre. Les variables étudiées sont le type d'échelle (continue ou discrète), le pays d'origine, le type de laboratoire (privé ou public) et la famille de produits. Les valeurs de ces variables sont renseignées pour chacun des 329 profils sélectionnés. En marge de ces quatre variables, le facteur "fournisseur" est également considéré. Celui-ci constitue le premier niveau d'organisation des études sensorielles rassemblées dans la SensoBase. Ce facteur sera donc étudié en priorité.

2.3.1.1 Unité expérimentale

Contrairement aux variables intra-études, les variables inter-études permettent de caractériser le niveau de performance du panel plutôt que des panélistes. L'unité expérimentale considérée devrait donc être le panel. Cependant, lorsqu'un jury réalise plusieurs études sensorielles, le type d'échelle et la famille de produits sont susceptibles de changer. La constitution du panel peut également être légèrement modifiée (certains panélistes se retirent et sont remplacés par de nouveaux sujets). De plus, les conditions expérimentales, telles que la nature de l'entraînement ou l'animateur de panel, peuvent varier. L'unité expérimentale retenue est donc le profil sensoriel réalisé à un temps donné par un panel, c'est-à-dire une étude sensorielle, quelle que soit la variable analysée.

2.3.1.2 Le facteur "fournisseur"

L'analyse du facteur "fournisseur" en tant que tel ne permet pas de faire progresser la connaissance en analyse sensorielle, mais cet effet, s'il existe, doit être pris en considération afin de ne pas parasiter les résultats des analyses des autres variables. La figure 5-6 montre que, parmi les 35 fournisseurs qui participent au programme SensoBase, 13 établissements ont envoyé un seul profil sensoriel. Le nombre d'études est ensuite très variable selon les fournisseurs. Rappelons que le nombre de profils retenus par fournisseur est limité à 30 dans les analyses réalisées dans ce chapitre.

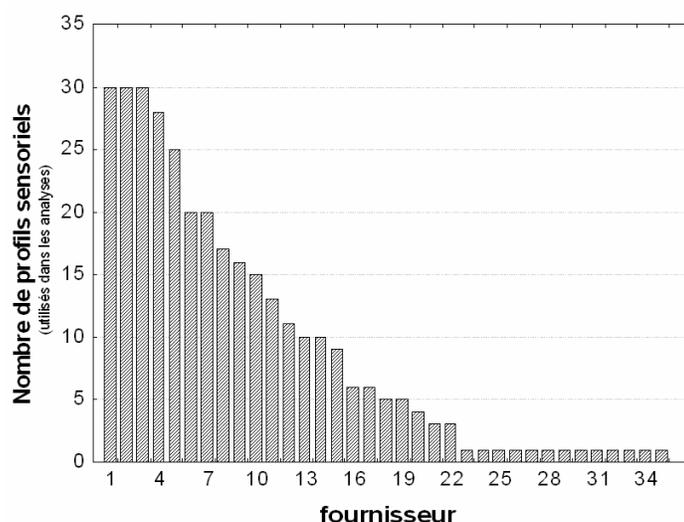


Figure 5-6 : représentation du nombre de profils sensoriels selon les fournisseurs

2.3.1.3 Le type d'échelle

La variable "type d'échelle" est composée de deux catégories uniquement : les échelles discrètes et les échelles continues. En revanche, chaque catégorie regroupe indifféremment des échelles aux caractéristiques différentes en terme de forme (rectangulaires ou linéaires), de longueur ou de libellés (vocabulaire précisant la définition de chaque borne et/ou de plusieurs intervalles). Le nombre de cases des échelles discrètes peut également varier (entre 5 et 15). Parmi les 329 profils, 84 études utilisent des échelles discrètes et 245 des échelles continues. Seuls quatre fournisseurs utilisent les deux types d'échelle (pour des études différentes).

2.3.1.4 Le pays d'origine

La plupart des fournisseurs de données de la SensoBase sont français (26 sur 35). L'étude de l'effet pays sera donc limitée à la comparaison entre les profils réalisés en France et les profils réalisés à l'étranger. Chaque catégorie représente respectivement 254 et 75 études. Nous cherchons ici à évaluer si les différences culturelles générales ou des différences liées à la pratique de l'analyse sensorielle entre les pays peuvent conduire à des différences de niveaux de performance.

2.3.1.5 Laboratoire privé ou public

Cette classification entre laboratoires privés et publics distingue les équipes d'analyse sensorielle qui évoluent au sein d'une entreprise, de celles qui évoluent dans le milieu académique. La justification de la comparaison de ces deux catégories est que les habitudes de travail sont souvent différentes. Les laboratoires publics mènent généralement des études relativement longues, qui impliquent un nombre important de produits et/ou de panélistes. Au contraire, les laboratoires privés réalisent davantage des études courtes, focalisées sur une question particulière. Cet aspect est corroboré par les différences de nombres moyens de produits, sujets, descripteurs et d'observations selon les types d'études (tableau 5-2).

Tableau 5-2 : Effectifs moyens selon le type de laboratoire

	Laboratoires privés	Laboratoires publics
Produit	7,0	13,8
Sujet	12,5	15,6
Descripteur	27,6	35,3
Observation	6 124	21 426

Les 329 études analysées se scindent en 290 profils issus de laboratoires privés (30 fournisseurs) et 39 profils issus de laboratoires publics (5 fournisseurs).

2.3.1.6 Type de produit

Lors de la saisie des données d'un profil dans le fichier Excel[®] SensoBase (chapitre 3, paragraphe 2), le fournisseur peut choisir le type de produits de l'étude parmi 53 possibilités (cf. annexe 5). Afin d'obtenir des effectifs assez élevés par catégorie, les différentes réponses possibles sont regroupées en sept classes : fruits et légumes (38 profils), poisson et viande (38), produits de panification (15), produits laitiers (108), plats préparés (28), boissons (62) et sauces (33). Sept études ne correspondent à aucune catégorie. Elles sont donc écartées de cette analyse.

2.3.2 Indice de performance

L'analyse des variables inter-études considère la performance au niveau du panel. Chaque indice de performance n'est donc pas calculé au niveau du sujet mais au niveau du groupe. Cependant, la répétabilité et l'accord d'un groupe peuvent respectivement se concevoir comme la moyenne des répétabilités et des accords individuels. Les indices proposés dans le paragraphe précédent sont donc conservés. L'indice de répétabilité du panel est la moyenne des répétabilités individuelles, c'est-à-dire la moyenne des écart-types de l'erreur issus des modèles d'ANOVA individuels par descripteur (équation 13). De manière similaire, l'indice d'accord est la moyenne des coefficients de corrélation de Pearson calculés entre chaque individu et le reste du groupe, pour chaque descripteur. En revanche, la moyenne des discriminations individuelles n'est pas

équivalente à la discrimination du groupe. Les panélistes peuvent être discriminants individuellement sans être en accord, ne créant ainsi aucun consensus quant aux différences entre produits. Deux indices de discrimination seront donc confrontés. Le premier correspond à la moyenne des discriminations individuelles (moyenne des r_{CMprod}). Cet indice permet de présenter des résultats comparables aux analyses des variables intra-étude. Le second indice est calculé pour chaque descripteur à partir du modèle d'ANOVA à deux facteurs avec interaction :

$$Score = produit + sujet + produit * sujet + erreur \quad (18)$$

Dans lequel le facteur sujet est aléatoire. La statistique de Fisher de l'effet produit traduit le niveau de discrimination du groupe. Cependant, cet indice, constitué du rapport de deux carrés moyens (paragraphe 3.1.1, chapitre 1) peut conduire à une valeur extrêmement forte pour un descripteur, lorsque le carré moyen du dénominateur est très proche de zéro. Une seule valeur très élevée, obtenue pour un descripteur d'un profil, influence alors fortement la discrimination moyenne du panel, calculée en tenant compte de tous les descripteurs. Nous souhaitons cependant conserver un indice qui reflète ce rapport entre le carré moyen de l'effet produit et celui de l'interaction. Par conséquent, nous utiliserons la probabilité associée à l'effet produit. Ce rapport nous semble en effet très pertinent pour décrire le pouvoir discriminant d'un groupe d'individu. L'indice de discrimination est donc la moyenne des probabilités obtenues pour tous les descripteurs d'une étude. A l'inverse de la statistique de Fisher, cet indice a l'inconvénient de surpondérer les descripteurs peu discriminants dans le calcul de la moyenne, mais ce biais est moins marqué que lors du calcul de la statistique de Fisher moyenne.

2.3.3 Modèle d'analyse

Le premier facteur de variabilité entre les profils sensoriels est naturellement le fournisseur. L'étude de cette variable permet d'évaluer l'intensité des écarts de performances entre les participants au programme SensoBase. Pour chaque indice de performance, un modèle d'ANOVA à un facteur (Score=fournisseur) est utilisé pour analyser l'effet de cette variable.

Etant donné qu'un fournisseur ne peut appartenir qu'à un pays et un type de laboratoire (privé ou public), le facteur "fournisseur" est hiérarchisé dans les facteurs "pays" et "laboratoire privé ou public". Pour ces variables, le modèle d'ANOVA utilisé est donc :

$$Score = variable + fournisseur(variable) \quad (19)$$

Où "variable" représente soit "pays", soit "laboratoire privé ou public". Le test de l'effet "variable" permet alors de déterminer si les différences entre les niveaux de "variable" sont significatives par rapport à la variabilité des profils sensoriels entre les fournisseurs au sein de chaque niveau de "variable" (i.e. au sein de chaque pays ou type de laboratoire).

La situation est différente pour les variables "type d'échelle" et "famille de produit" puisqu'un fournisseur peut intervenir dans plusieurs catégories d'une même variable. L'effet "fournisseur" n'est donc plus hiérarchisé dans chaque variable et il existe une interaction entre les deux

facteurs. Cependant, tous les fournisseurs ne participent pas à tous les niveaux de chaque variable. En effet, seuls 4 fournisseurs présentent des profils qui utilisent soit des échelles discrètes, soit des échelles continues et 11 fournisseurs ont envoyé des études portant sur des familles de produits différentes. De plus, chaque fournisseur intervient au maximum dans 4 des 7 catégories de produits. L'interaction "variable*fournisseur" est donc incomplète pour ces variables. Dans cette situation, nous avons préféré ne pas utiliser le modèle avec interaction. Par ailleurs, les situations dans lesquels un fournisseur intervient dans plusieurs catégories d'une variable sont relativement rares. Dans l'approche exploratoire que constitue ce chapitre, nous avons donc choisi de simplifier la structure des données. Lorsque les profils sensoriels d'un même fournisseur appartiennent à différentes catégories, les études sont attribuées à autant de pseudo-fournisseurs que de catégories représentées. Chaque pseudo-fournisseur appartient alors à une seule catégorie. Cette approximation de la nature des données de la SensoBase permet d'utiliser le modèle de l'équation (19) pour les variables "type de produit" et "type d'échelle".

3 Résultats - Discussion

Les niveaux de performance globaux sont donnés dans le tableau 5-3 et les distributions correspondantes sont présentées dans la figure 5-7. Les calculs sont réalisés sans système de pondération. Les niveaux obtenus seront utilisés comme référence lors de la présentation des résultats de chaque variable au cours de ce chapitre.

Tableau 5-3 : Niveau moyen de chaque indice de performance

	REPETABILITE (ETres)	DISCRIMINATION (pCMprod)	ACCORD (CORR)
Moyenne	1.200	0.620	0.395
Médiane	1.142	0.615	0.394
N	3814	3814	4222
Std	0.540	0.113	0.217

Niveaux moyens calculés sur 329 jeux de données sélectionnés dans la SensoBase.
Unité expérimentale = le panéliste dans un jeu de données

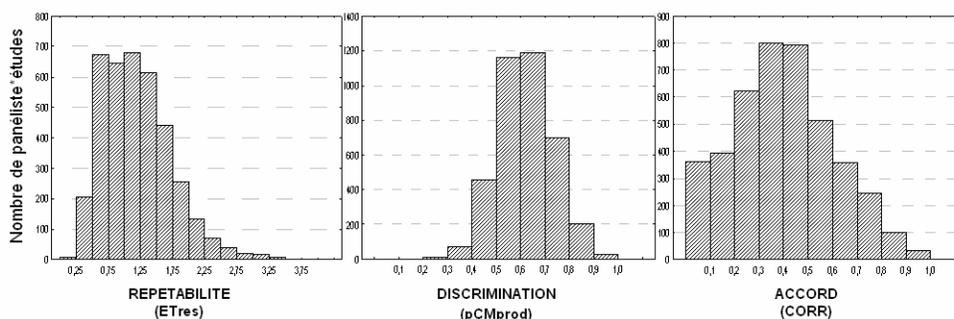


Figure 5-7 : Distributions des indices de performance (unité expérimentale = le panéliste dans une étude)

Le niveau moyen de répétabilité est de 1,2 sur une échelle de notation de 0 à 10. Cela signifie que, en moyenne sur tous les profils sensoriels de la SensoBase, une répétition s'écarte de la note moyenne du produit de 12% de la longueur de l'échelle. Ce niveau a été calculé à partir de 3814 observations, soit autant que de panélistes repartis dans 329 profils sensoriels. L'écart-type des indices de répétabilité est de 0,54. Cela représente 45% de la valeur de la moyenne ($0,54/1,2=45\%$), ce qui souligne la variabilité des situations rassemblées dans la SensoBase. Ce résultat est mis en évidence graphiquement par la forme de la distribution (figure 5-7) pour laquelle les valeurs comprises entre 0,5 et 1,5 sont toutes aussi fréquentes.

Le niveau moyen de discrimination est de 0,62. Cela signifie que le carré moyen de l'effet produit représente 62% du total des carrés moyens dans le modèle d'ANOVA individuel (équation 13). En utilisant l'équation (14), la valeur de la statistique de Fisher de l'effet produit correspondante serait $0,62 / (1 - 0,62) = 1,63$. En supposant que cette valeur ait été obtenue dans un profil ayant les caractéristiques médianes (5 produits et 2 répétitions, figures 5-1.b et 5-1.c), une estimation de la probabilité associée est calculée. Les degrés de liberté sont $5-1=4$ pour le numérateur et $5*2-4-1=5$ pour le dénominateur, donc la probabilité est $p(1.63,4,5)=0.30$. La probabilité moyenne, calculée d'après les valeurs de probabilité obtenues pour chaque modèle d'ANOVA individuel (équation 13) donne un résultat similaire : $p=0,31$ (calcul basé sur $N=126110$ observations = nombre de profil*panéliste*descripteur). La distribution des probabilités associées à l'effet produit est présentée dans la figure 5-8. Elle indique que la probabilité est respectivement inférieure à 0,05 et 0,10 pour 28,8% et 36,2% des descripteurs. Le niveau de discrimination individuel n'atteint pas le seuil de significativité pour la majorité des descripteurs.

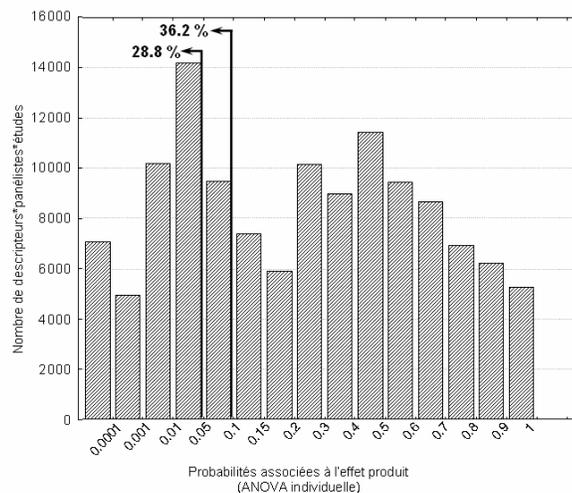


Figure 5-8 : Distribution des probabilités associées au test de l'effet produit dans les modèles d'ANOVA individuels

La figure 5-9 représente les mêmes données, considérées cette fois au niveau du panéliste. Les observations deviennent, pour chaque sujet dans chaque étude, le pourcentage de descripteurs dont l'effet produit est significatif au risque de 0,05 (distribution de gauche) et de 0,10 (distribution de droite). Les graphiques montrent que le panéliste médian est discriminant pour 24% des

descripteurs lorsque le risque est fixé à 0,05 (respectivement 33% au risque de 0,10) dans une étude sensorielle. La forme de la distribution indique la présence de certains résultats extrêmes. Le nombre de panélistes non discriminants (au risque 0,05) pour tous les descripteurs d'une étude représente en effet 9,5% des observations. A l'opposé, certains panélistes sont discriminants pour la quasi-totalité des descripteurs. Ce cas reste cependant plus rare.

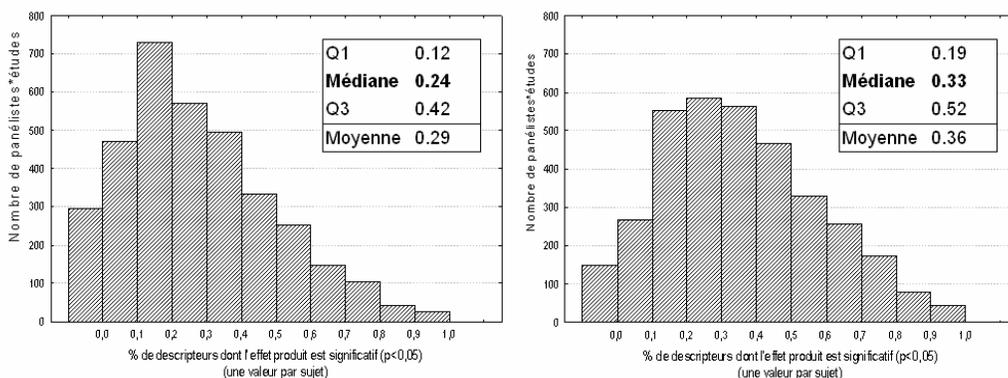


Figure 5-9 : Distributions des pourcentages de descripteurs pour lesquels un panéliste est discriminant dans une étude

Le niveau moyen d'accord (coefficient de corrélation de Pearson) est de 0,395. Cette valeur est positive, donc les panélistes semblent, en moyenne, en accord entre eux dans chaque étude. Le nombre d'observations (N=4222) est plus élevé dans ce calcul parce que les études sans répétitions peuvent être prises en compte. Puisque la médiane de la distribution du nombre de produits (figure 5-1.b), est 5, la significativité du niveau d'accord moyen est testée dans cette situation. Les probabilités obtenues sont respectivement 0,25. L'accord entre panélistes n'est donc pas significatif en moyenne. La figure 5-10 présente la distribution des probabilités associées au test de significativité du coefficient de corrélation de Pearson (pour chaque sujet) calculé dans la SensoBase. Ces probabilités sont inférieures à 0,05 pour 35,0% des descripteurs et inférieures à 0,10 pour 44,8% des descripteurs.

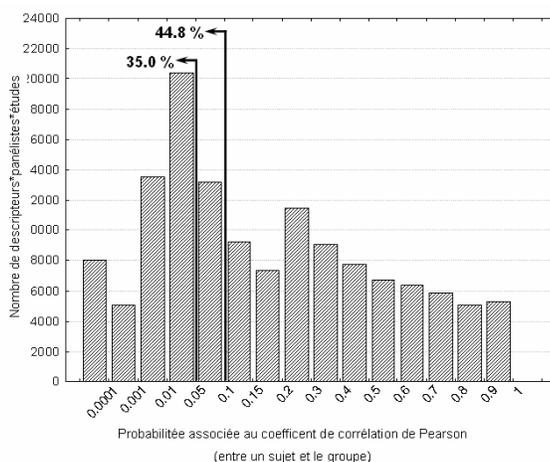


Figure 5-10 : Distribution des probabilités associées coefficients de corrélation de Pearson individuels

La figure 5-11 représente les mêmes résultats en se plaçant au niveau du panéliste. Les observations deviennent, pour chaque sujet de chaque étude, le pourcentage de descripteurs pour lesquels le sujet est en accord avec le groupe (risque fixé à 0,05 pour la distribution de droite et à 0,10 pour la distribution de gauche). Ces graphiques montrent que, en moyenne, les panélistes sont en accord ($p < 0,05$) pour seulement 33% de descripteurs d'une étude.

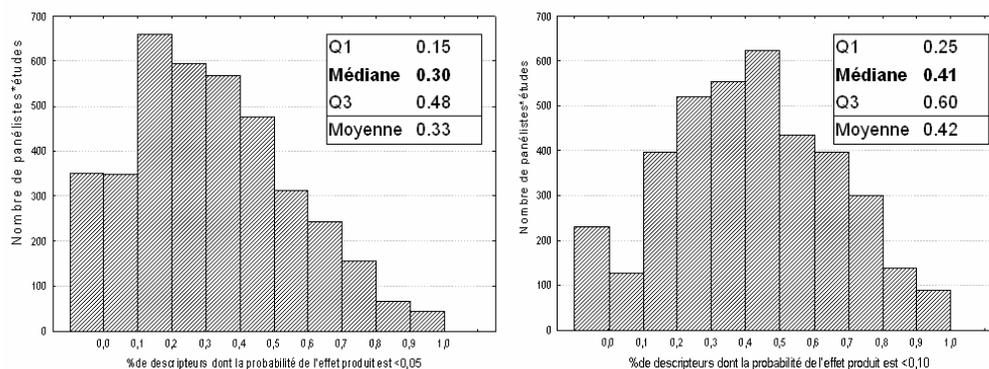


Figure 5-11: Distributions des pourcentages de descripteurs pour lesquels un panéliste est en accord avec le groupe dans une étude

Ces premiers résultats permettent d'appréhender les niveaux moyens de chaque indice. Ils mettent également en évidence l'hétérogénéité des situations rassemblées dans la SensoBase. L'analyse de chaque variable, présentée dans les paragraphes suivants, vise à expliquer cette variabilité.

3.1 Variables intra-étude

3.1.1 Effet de l'âge

Le tableau 5-4 synthétise les résultats liés à l'effet de l'âge. Chaque ligne du tableau correspond à un indice de performance. La statistique de Fisher (F) et la probabilité associée (prob) sont présentées pour chaque facteur du modèle d'ANOVA (équation 15). Les effets significatifs sont surlignés en gras. Des informations additionnelles (Moyenne, N et StdErr) sont également disponibles pour chaque indice de performance. La partie droite de la table contient les résultats des tests de comparaison de moyennes. Les effets moyens de chaque catégorie sont écrits en gras lorsqu'une différence significative entre les produits existe.

Tableau 5-4 : Effet de l'âge sur le niveau de performance des panélistes

AGE	AGE		Dataset		Interaction		Mean	N	StdErr	levels	Mean	N	Std
	F	prob	F	prob	F	prob							
AGREEMENT (Pearson)	2,35	0,097	14,05	<0,001	1,12	0,058	0,387	3572	0,088	30-	0,007 A	484	0,069
										30-45	0,006 A	1202	0,087
										45+	-0,006 A	1886	0,088
DISCRIMINATION (pCMprod)	9,51	<0,001	8,80	<0,001	1,09	0,130	0,618	3202	0,053	30-	-0,003 B	403	0,044
										30-45	0,009 A	1093	0,055
										45+	-0,006 B	1706	0,051
REPETABILITY (ETres)	2,32	0,099	13,25	<0,001	0,99	0,551	1,207	3202	0,242	30-	0,034 A	403	0,306
										30-45	-0,003 A	1093	0,214
										45+	-0,004 A	1706	0,22

L'effet de l'âge est très significatif sur le niveau de discrimination ($F=9,51$; $p<0.001$). Les tests de comparaison multiples indiquent que le niveau de discrimination est meilleur pour la classe d'âge médiane (30 à 45 ans). L'effet de l'âge sur le niveau d'accord ($F=2,35$; $p<0.097$) et sur le niveau de répétabilité ($F=2,32$; $p<0,099$) n'est pas clairement significatif mais une tendance existe. Bien que les tests de comparaison multiples ne soient pas significatifs, cette tendance indiquerait un niveau d'accord plus faible pour les personnes de plus de 45 ans et un niveau de répétabilité moindre pour les personnes plus jeunes.

Puisque la variable "âge" doit obligatoirement être renseignée pour chaque étude insérée dans la Sensobase, la taille de l'échantillon disponible est importante ($N=3572$, ou 3202 lorsque des répétitions des mesures sont nécessaires). Cependant, elle n'est pas égale à celle présentée dans le tableau 5-3 en raison des études qui ne contiennent qu'une classe d'âge. Ces dernières sont évincées de l'analyse car elles n'apportent aucune information sur les différences entre les catégories d'âge. Les scores moyens obtenus pour les indices d'accord (0,387), de discrimination (0,618) et de répétabilité (1,207) du tableau 5-4 sont cependant très proches des valeurs du tableau 5-3 (respectivement 0,395, 0,620 et 1,200). Ce résultat indique que le système de pondération n'affecte pas le niveau moyen de performance mesuré. L'effet "profil", très significatif pour chaque indice, souligne la forte hétérogénéité des situations rassemblées dans la SensoBase vis-à-vis des niveaux de performance. Les raisons de ces différences pourront en partie être expliquées lors des analyses portant sur les variables inter-études. L'interaction n'est significative pour aucun indice. Les écarts de performance entre les catégories d'âge ne sont donc statistiquement pas différents d'une étude sensorielle à l'autre. Ce résultat renforce l'importance de l'effet de l'âge observé. Cependant, l'effet de l'interaction est à la limite de la significativité dans l'analyse du niveau d'accord ($F=1.124$, $p=0.0583$). L'effet de l'âge est donc légèrement différent d'un profil sensoriel à l'autre pour cet indice de performance.

3.1.2 Effet du sexe

Les statistiques de Fisher de l'effet du sexe sur les niveaux de performances sont particulièrement faibles (accord : $F=0,24$; discrimination: $F=0,10$; répétabilité: $F=0,002$) et les probabilités associées ne sont pas significatives (tableau 5-5).

Tableau 5-5 : Effet du sexe sur le niveau de performance des panélistes

SEXE	SEXE		Dataset		Interaction		Mean	N	StdErr	levels	Mean	N	Std
	F	prob	F	prob	F	prob							
AGREEMENT (Pearson)	0,24	0,623	14,81	<0,001	1,16	0,060	0,385	2696	0,103	F	0,001	A 1791	0,096
										M	-0,001	A 905	0,107
DISCRIMINATION (pCMprod)	0,10	0,750	8,36	<0,001	1,22	0,023	0,616	2381	0,060	F	0,001	A 1593	0,058
										M	-0,002	A 788	0,057
REPETABILITY (ETres)	0,00	0,961	12,90	<0,001	0,84	0,938	1,185	2381	0,291	F	-0,002	A 1593	0,288
										M	0,003	A 788	0,255

Aucun effet du sexe n'est mis en évidence. Cette conclusion peut être surprenante par rapport aux résultats de la littérature qui concluent souvent à la supériorité des femmes concernant les

tâches sensorielles et mémorielles. Cette supériorité n'avait cependant pas été mise en évidence concernant les tâches de discrimination. De plus, il est important de noter que les panélistes qui participent à une étude sensorielle sont généralement sélectionnés en fonction de leurs aptitudes sensorielles (tests de mémoire, épreuves sensorielles et descriptives...). Par conséquent, seuls les meilleurs dégustateurs sont sélectionnés. Le processus de sélection pourrait ainsi expliquer en partie la non significativité de l'effet sexe. Cependant, le même raisonnement devrait être appliqué lors de l'étude de l'effet de l'âge et du niveau d'éducation. L'effet de ces variables sur les niveaux de performance ne devrait donc pas être significatif, alors que le contraire est observé.

3.1.3 Effet du niveau d'expérience

Tableau 5-6 : Effet du niveau d'expérience sur le niveau de performance des panélistes

EXPERIENCE	EXPERIENCE		Dataset		Interaction		Mean	N	StdErr	levels	Mean	N	Std
	F	prob	F	prob	F	prob							
AGREEMENT (Pearson)	3,13	0,048	13,64	<0,001	0,99	0,499	0,399	540	0,078	EXP 0	-0,027 B	173	0,083
										EXP13	0,003 A	174	0,08
										EXP3+	0,025 A	193	0,064
DISCRIMINATION (pCMprod)	4,13	0,018	13,76	<0,001	0,87	0,732	0,626	507	0,046	EXP 0	-0,010 B	172	0,048
										EXP13	-0,006 B	158	0,045
										EXP3+	0,019 A	177	0,04
REPETABILITY (ETres)	1,60	0,206	11,78	<0,001	0,97	0,537	1,361	507	0,226	EXP 0	0,005 A	172	0,254
										EXP13	0,032 A	158	0,215
										EXP3+	-0,044 A	177	0,179

L'effet du niveau d'expérience est significatif concernant l'accord et le pouvoir discriminant, montrant une amélioration des niveaux de performance en fonction du niveau d'entraînement. Le niveau moyen d'accord est en effet plus faible pour la catégorie "EXP 0" et les panélistes ayant plus de trois années d'expérience en profil sensoriel sont plus discriminants. Les résultats de la littérature sur l'effet du niveau d'entraînement sont donc confirmés. Cependant, aucun effet sur le niveau de répétabilité n'est observé. Notons que l'analyse de l'effet du niveau d'expérience est basée sur un petit nombre d'observations par rapport aux autres analyses (approximativement 500 au lieu de plusieurs milliers). Toutefois, 500 "panélistes" représentent un nombre important comparé à l'analyse des données d'un seul profil sensoriel. La valeur moyenne de chaque indice garde en outre un niveau proche de la moyenne calculée sur l'ensemble des données de la SensoBase (tableau 5-3). Cela indique que l'échantillon constitué des études pour lesquelles le niveau d'expérience était renseigné est représentatif de l'ensemble de données de la SensoBase.

3.1.4 Effet du niveau d'éducation

Tableau 5-7 : Effet du niveau d'éducation sur le niveau de performance des panélistes

EDUCATION	EDUCATION		Dataset		Interaction		Mean	N	StdErr	levels	Mean	N	Std
	F	prob	F	prob	F	prob							
AGREEMENT (Pearson)	1,58	0,212	4,84	<0,001	1,10	0,340	0,363	298	0,113	Bac-	-0,021 B	117	0,13
										Bac+	0,015 A	181	0,092
DISCRIMINATION (pCMprod)	4,02	0,048	2,76	<0,001	1,99	0,048	0,606	267	0,058	Bac-	-0,024 B	106	0,066
										Bac+	0,017 A	161	0,052
REPETABILITY (ETres)	0,05	0,817	6,63	<0,001	0,60	0,937	1,353	267	0,290	Bac-	0,023 A	106	0,284
										Bac+	-0,017 A	161	0,261

L'effet du niveau d'éducation a un impact uniquement sur le pouvoir discriminant. De meilleurs résultats sont obtenus pour les sujets dont le niveau d'étude est élevé (supérieur au Baccalauréat). Cet effet pourrait s'expliquer par les capacités que permettent de développer par le

travail scolaire, comme la compréhension d'un vocabulaire spécifique et l'utilisation d'un système de notation sur une échelle. Le niveau de discrimination pourrait être accru pour ces raisons. En revanche, le niveau d'éducation ne semble pas avoir d'effet sur les performances sensorielles au sens strict puisque le niveau de répétabilité n'est pas amélioré.

3.1.5 Effet du type de descripteur

Tableau 5-8 : Effet du type de descripteur sur le niveau de performance des panélistes

TYPE DE DESCRIPTEUR	TYPE DESCR.		Dataset		Interaction		Mean	N	StdErr	levels	Mean	N	Std
	F	prob	F	prob	F	prob							
AGREEMENT (Pearson)	53,20	<0,001	3,82	<0,001	5,45	<0,001	0,406	14101	0,254	Visuel	0,166 A	2129	0,256
										Texture	0,007 B	2798	0,29
										Arôme	-0,033 C	3382	0,277
										Saveur	-0,045 C	3641	0,282
										Odeur	-0,046 C	2151	0,309
DISCRIMINATION (pCMprod)	25,64	<0,001	5,41	<0,001	3,79	<0,001	0,621	12420	0,125	Visuel	0,055 A	1795	0,134
										Arôme	-0,002 B	2997	0,126
										Odeur	0,008 BC	1936	0,144
										Texture	-0,013 C	2453	0,137
										Saveur	-0,014 C	3239	0,132
REPETABILITY (ETres)	34,03	<0,001	10,71	<0,001	3,27	<0,001	1,174	12420	0,421	Saveur	-0,113 A	3239	0,413
										Arôme	-0,005 B	2997	0,443
										Odeur	-0,001 B	1936	0,48
										Visuel	0,038 C	1795	0,442
										Texture	0,128 D	2453	0,46

L'effet du type de descripteur est particulièrement fort pour chaque indice de performance (accord : $F=53.20$, $p<0.0001$; discrimination: $F=25.64$, $p<0.0001$; répétabilité: $F=34.03$, $p<0.0001$). Les niveaux de performance sont donc très différents selon ce critère. Dans le tableau 5-8, contrairement aux autres tableaux de résultats, les catégories (types de descripteurs) sont classées en fonction de leur niveau de performance par ordre décroissant. L'effet "profil" est toujours significatif, comme dans toutes les analyses. L'interaction est également significative, donc il existe des différences de classement des types de descripteurs entre les profils. Mais cet effet reste minime puisque les différences globales entre catégories de descripteurs sont très significatives.

Les niveaux de performance sont plus élevés pour les descripteurs visuels que pour les autres descripteurs, concernant les indices d'accord et de discrimination. En revanche, leur niveau de répétabilité est inférieur à celui des descripteurs d'odeur, d'arôme et de saveur. Le niveau de répétabilité relativement faible des descripteurs visuels est donc probablement lié à une dispersion plus étendue des notes sur les échelles de notation. En effet, l'étendue moyenne des notes pour les descripteurs visuels est de 4,73 (échelles de 0 à 10). Cette valeur est supérieure de plus de un point aux étendues des descripteurs d'odeurs (3.30), d'arômes (3.69) et de saveurs (3.61). Pour les descripteurs de texture le niveau de répétabilité est particulièrement faible et l'étendue moyenne des notes est inférieure à celle des descripteurs visuels (4,51). La médiocre répétabilité des descripteurs de texture n'est donc pas uniquement le reflet d'une forte dispersion des notes. Le niveau d'accord est plus élevé pour les descripteurs visuels, puis pour les descripteurs de texture.

3.1.6 Interaction entre le type de descripteur et les autres variables

Le tableau 5-9 présente les résultats des analyses réalisées d'après l'équation 17 afin de vérifier si l'effet des variables âge, sexe, niveau d'expérience et niveau d'études dépend du type de descripteur. L'interaction Variable*TypeDescr est significative seulement dans le modèle qui étudie l'effet de l'âge sur le niveau de répétabilité. Les ANOVA effectuées dans ce cas pour chaque type de descripteur, en utilisant le modèle de l'équation 15, mettent en évidence un effet de l'âge sur le niveau de répétabilité, uniquement pour les descripteurs visuels (F=9,16 ; p=0,003). Pour cette catégorie, le niveau de répétabilité des femmes est significativement plus élevé. Cependant, les niveaux de discrimination ne sont pas différents entre les sexes pour les descripteurs visuels (F=0,02, p=0,89). Le même résultat est observé lorsque tous les types de descripteurs sont pris en compte (les statistiques de Fisher de l'effet du sexe sur les niveaux de performances sont particulièrement faibles (accord : F=0,24; discrimination: F=0,10; répétabilité: F=0,002) et les probabilités associées ne sont pas significatives (tableau 5-5). La différence de niveau de répétabilité est donc la conséquence d'une utilisation différente de l'échelle selon le sexe : les hommes dispersent davantage leurs notes que les femmes pour les descripteurs visuels. Ceci est confirmé par le calcul des étendues pour ces descripteurs (étendue pour les femmes : 5,12 ; étendue pour les hommes : 6,15).

Tableau 5-9 : Résultats des ANOVA faisant intervenir l'effet du type de descripteur

INDICE PERF.	SOURCE	AGE			SEXE			EXPERIENCE			EDUCATION		
		ddl	F	Prob	ddl	F	Prob	ddl	F	Prob	ddl	F	Prob
ACCORD (Pearson)	Variable	1	1,84	0,175	1	0,01	0,939	2	0,68	0,509	1	0,71	0,401
	Etude	291	13,61	<,001	214	14,51	<,001	42	15,06	<,001	28	9,94	<,001
	TypeDescr	4	171,27	<,001	4	193,62	<,001	4	78,94	<,001	4	74,7	<,001
	Variable*Etude	291	0,91	0,849	212	1,23	0,014	55	1,9	<,001	28	1,91	0,003
	Variable*TypeDescr	4	1,15	0,329	4	0,32	0,868	8	1,22	0,284	4	1,26	0,286
	Etude*TypeDescr	713	5,12	<,001	494	5,2	<,001	117	5,03	<,001	84	3,44	<,001
DISCRIMINATION (pCMprod)	Variable	1	9,05	0,003	1	0,14	0,705	2	4,67	0,010	1	9,3	0,002
	Etude	273	13,8	<,001	200	13,4	<,001	39	13,9	<,001	25	8,35	<,001
	TypeDescr	4	57,05	<,001	4	61,24	<,001	4	31,82	<,001	4	38,86	<,001
	Variable*Etude	270	1,4	<,001	198	1,51	<,001	51	1,63	0,003	25	2,22	0,001
	Variable*TypeDescr	4	1,7	0,147	4	1,58	0,176	8	0,9	0,512	4	0,56	0,693
	Etude*TypeDescr	650	3,38	<,001	447	3,8	<,001	105	2,53	<,001	72	1,54	0,003
REPETABILITE (ETres)	Variable	1	0,75	0,387	1	0,4	0,528	2	1,785	0,168	1	1,02	0,312
	Etude	273	28,65	<,001	200	18,09	<,001	39	21,88	<,001	25	8,25	<,001
	TypeDescr	4	61,88	<,001	4	67,04	<,001	4	22,98	<,001	4	21,11	<,001
	Variable*Etude	270	1,64	<,001	198	1,96	<,001	51	2,39	<,001	25	1,46	0,070
	Variable*TypeDescr	4	2,17	0,070	4	4,22	0,002	8	1,18	0,306	4	1,1	0,357
	Etude*TypeDescr	650	3,19	<,001	447	2,38	<,001	105	2,84	<,001	72	2,2	<,001

La significativité de l'interaction Age*TypeDescr sur le niveau de répétabilité est presque atteinte également (tableau 5-9, prob=0,0702). Cependant, les analyses par type de descripteur ne mettent aucun effet de l'âge en évidence. Le type de descripteur ne semble pas modifier l'effet de l'âge. L'interaction Variable*TypeDescr est donc non significative, quels que soient la variable (age, sexe, niveau d'éducation, niveau d'expérience) et l'indice de performance (accord, discrimination ou répétabilité). Les résultats présentés dans les tableaux 5-4 à 5-7 sont donc vérifiés quel que soit le type de descripteur. Notons enfin que les résultats concernant les effets principaux "variable", présentés dans le Tableau 5-9, sont cohérents avec les résultats des tableaux 5-4 à 5-7, excepté pour l'effet du niveau d'expérience sur les indices d'accord et de répétabilité. L'effet sur le niveau d'accord était à la limite de la significativité dans le tableau 5-6 et

devient non significatif dans le modèle tenant compte du type de descripteur. La modification inverse se produit concernant l'effet du niveau d'expérience sur la répétabilité.

Les résultats présentés dans ce chapitre sont basés sur un vaste ensemble de profils sensoriels, au moins pour l'étude de l'effet de l'âge, du sexe et du type de descripteur. Cependant, ces premiers résultats issus de la SensoBase représentent "seulement" 35 fournisseurs de données (dont 26 laboratoires français). La représentativité des résultats doit donc encore être améliorée. Outre ce problème de biais de représentativité, il existe un biais inhérent au mode de recueil des données, comparable au "biais de publication" très fréquent en méta-analyse. Le biais de publication est dû au fait que les données publiées ne reflètent qu'une partie des résultats obtenus, correspondant le plus souvent aux expérimentations dont les résultats sont satisfaisants. Les études ne montrant aucun effet ne sont en général pas publiées. Les données analysées ne couvrent donc qu'une partie des résultats. Ce biais est transposable à notre situation puisque les fournisseurs choisissent les données qu'ils envoient dans la SensoBase. Ils peuvent, par exemple, sélectionner uniquement une partie du panel et/ou des descripteurs, ou envoyer seulement les mesures acquises dans les meilleures conditions. De plus, les informations saisies sur les sujets, les produits et les descripteurs ne peuvent être contrôlées que par les fournisseurs. Ces différentes sources potentielles de biais doivent être gardées à l'esprit lors de l'analyse des résultats. Toutefois, pour chaque fournisseur, l'intérêt principal du programme SensoBase est de recevoir une analyse des données de son panel. Il n'a donc aucune raison d'envoyer seulement une partie de ses données ou des informations erronées. Le principal risque réside dans les erreurs de saisie des données dans le fichier Excel[®], qui ne peut être totalement maîtrisé.

Signalons également que les valeurs moyennes, présentées dans le tableau 5-3, donnent des niveaux de performance de référence aux acteurs de l'analyse sensorielle. Cependant, du fait de la très forte variabilité des situations rencontrées dans le domaine, ces valeurs doivent uniquement être utilisées comme des indicateurs du niveau de performance généralement atteint. Ce ne sont en aucun cas des valeurs seuils que le panel doit atteindre afin de déclarer que celui-ci est performant. A notre avis, il est préférable de considérer la distribution de chaque indice afin d'obtenir une information sur la dispersion des valeurs rencontrées.

3.2 Variables inter-études

3.2.1 Le facteur "fournisseur"

L'étude du facteur "fournisseur" permet de mesurer l'hétérogénéité de l'échantillonnage des fournisseurs rassemblés dans la SensoBase. L'analyse de chaque indice de performance met en évidence un effet significatif de la provenance des profils sensoriels (tableau 5-10). Cet effet est beaucoup plus marqué concernant l'indice de répétabilité ($F=27,29$).

Tableau 5-10 : Effet du fournisseur sur les niveaux de performance des panels

	FOURNISSEUR		StdErr
	F	p	
ACCORD (Pearson)	4.00	<0.001	0.55
DISCRIMINATION (prob)	3.80	<0.001	0.37
DISCRIMINATION (pMCprod)	2.85	<0.001	0.28
REPETABILITE (ETres)	27.29	<0.001	0.79

Afin d'appréhender plus finement les raisons de la significativité de ces effets, la figure 5-12 présente, pour chaque indice de performance, les moyennes de chaque fournisseur et l'intervalle de confiance à 95% associé. Les fournisseurs, en abscisse, sont classés par ordre décroissant selon le nombre de profils sensoriels qui leur appartient dans cette analyse. Pour l'indice d'accord et les deux indices de discrimination, les niveaux de performance moyens des fournisseurs sont relativement homogènes, à l'exception de quelques individus dont le niveau s'écarte de la moyenne. L'hétérogénéité entre les fournisseurs semble, de plus, s'accroître pour les fournisseurs dont le nombre d'observations, c'est à dire le nombre de profils sensoriels, est moindre (fournisseurs de la partie droite des graphiques). Ce résultat appuie le fait qu'il est nécessaire d'avoir un échantillonnage de situations assez vaste par fournisseur pour obtenir une estimation fiable. Les effets significatifs de ces trois indices (tableau 5-10) sont donc principalement dus à quelques fournisseurs qui s'écartent de la moyenne générale. Notons de plus que les deux indices de discrimination ne semblent pas équivalents. Par exemple, la probabilité moyenne associée à la statistique de Fisher de l'effet produit (graphique en bas à gauche) du fournisseur 6 est relativement élevée (probF=0,294). Cela traduit un niveau de discrimination moyen du panel plutôt faible pour les profils sensoriels de ce fournisseur. Cependant, le niveau de discrimination individuel moyen (graphique en bas à droite) se situe exactement dans la moyenne pour le même fournisseur. Les panélistes issus des profils de ce fournisseur sont donc individuellement capables de faire autant de différences entre produits que les juges des autres fournisseurs, mais au niveau du panel, le consensus quant aux différences entre produits est moins bon, comme l'atteste la valeur plutôt basse de la statistique d'accord du panel 6. Les deux notions doivent donc être considérées de manière distincte.

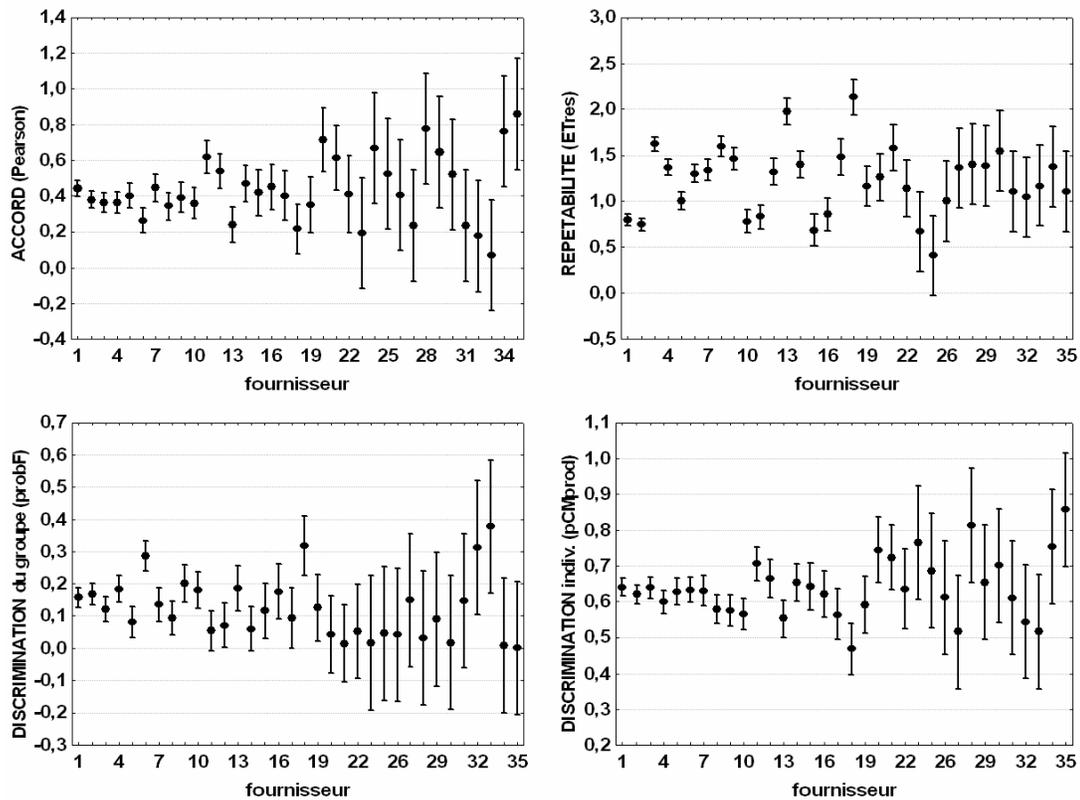


Figure 5-12 : Représentation des moyennes et des intervalles de confiance de chaque fournisseur pour chaque indice de performance

Le graphique de la répétabilité montre en revanche une forte hétérogénéité des niveaux entre les fournisseurs. Ceci indique que l'écart-type entre les répétitions est différent selon le laboratoire d'origine. Ce résultat pourrait traduire des différences réelles de niveaux de répétabilité des sujets, mais la relative homogénéité des niveaux de discrimination tend à infirmer cette hypothèse. Les différences de niveaux de répétabilité reflèteraient donc davantage des différences d'étendue des notes sur les échelles de notation. L'hétérogénéité observée indique donc des différences de comportement selon les fournisseurs. Cet effet pourrait éventuellement s'expliquer par des consignes d'utilisation des échelles différentes selon les fournisseurs.

3.2.2 Le type d'échelle

Tableau 5-11 : Effet du type d'échelle sur les niveaux de performance des panels

ECHELLE	ECHELLE		fourn(ECHELLE)		Moyenne	N	StdErr	niveaux	Moyenne	N	Std
	F	p	F	p							
ACCORD (Pearson)	0.22	0.639	4.07	<0.001	0.40	329	0.54	continue	0.434 A	245	0.176
								discrète	0.415 A	84	0.188
DISCRIMINATION (prob)	0.51	0.476	4.27	<0.001	0.85	317	0.35	continue	0.122 A	241	0.116
								discrète	0.141 A	76	0.130
DISCRIMINATION (pCImprod)	0.24	0.622	2.67	<0.001	0.62	317	0.28	continue	0.640 A	241	0.084
								discrète	0.630 A	76	0.104
REPETABILITE (ETres)	3.88	0.050	24.49	<0.001	1.18	317	0.79	continue	1.182 A	241	0.417
								discrète	1.30 B	76	0.448

L'effet du type d'échelle (continue ou discrète) n'est pas significatif pour l'indice d'accord et les deux indices de discrimination. La significativité est juste atteinte ($p=0,05$) pour l'indice de répétabilité et met en lumière de meilleurs résultats pour les échelles continues. Ce résultat va à

l'encontre de l'idée fréquemment évoquée selon laquelle les échelles discrètes forcent les panélistes à faire un choix parmi un nombre de possibilités restreint. En supposant que ce résultat n'est pas uniquement dû à la corrélation avec l'effet fournisseur, cela indiquerait que les échelles discrètes perturbent la notation des panélistes et ne permettent pas d'accroître la répétabilité. Cependant, l'effet "fournisseur", hiérarchisé dans le facteur "échelle", exhibe une variabilité beaucoup plus importante que l'effet "échelle" (effet fournisseur(échelle) : $F=24,49$; effet échelle : $F=3,88$). Cette conclusion est donc à relativiser puisque les différences entre fournisseurs au sein de chaque catégorie sont plus importantes que les différences de type d'échelle. Notons également que l'effet fournisseur(échelle) est nettement significatif ($p<0,001$) quel que soit l'indice de performance considéré. La variabilité qui existe entre les fournisseurs est donc le premier critère de différence entre les niveaux de performance des panels.

3.2.3 Le pays d'origine

Tableau 5-12 : Comparaison des niveaux de performance entre la France et l'étranger

PAYS	PAYS		fourn(PAYS)		Moyenne	N	StdErr	niveaux	Moyenne	N	Std
	F	prob	F	prob							
ACCORD (Pearson)	0.12	0.726	4.08	<0.001	0.40	329	0.55	France	0.438 A	254	0.175
								Etranger	0.422 A	75	0.192
DISCRIMINATION (prob)	0.13	0.720	3.88	<0.001	0.85	317	0.37	France	0.117 A	244	0.118
								Etranger	0.128 A	73	0.122
DISCRIMINATION (pMCprod)	0.53	0.467	2.94	<0.001	0.62	317	0.28	France	0.642 A	244	0.086
								Etranger	0.625 A	73	0.100
REPETABILITE (ETres)	59.47	<0.001	19.96	<0.001	1.18	317	0.79	France	1.351 B	244	0.410
								Etranger	0.853 A	73	0.179

Les résultats concernant le pays d'origine montrent également un effet significatif sur le niveau de répétabilité. La valeur de la statistique de Fisher est cette fois beaucoup plus élevée ($F=59.47$) et supérieure à celle de l'effet fournisseur(pays) ($F=19.96$). Ce résultat tend donc à montrer que l'une des principales causes de différences de niveaux de répétabilité serait lié à l'effet pays. Le test de comparaison de moyenne indique de meilleurs résultats pour les pays étrangers. Cet écart de performance ne se répercute pas sur la discrimination. L'effet observé serait donc principalement lié à des différences de dispersion des notes sur les échelles. Signalons que la majorité des pays autres que le France, impliqués dans la SensoBase, sont majoritairement d'origine anglo-saxonne ou de l'Europe du Nord. L'effet observé tendrait ainsi à montrer que le mode de travail en analyse sensorielle est sensiblement différent entre ces pays et la France, puisque le comportement des panélistes semble varier. Toutefois, cela n'a pas d'impact sur le niveau d'accord ou le niveau de discrimination.

3.2.4 Laboratoire privé ou public

Tableau 5-13 : Comparaison des niveaux de performance entre les laboratoires privés et publics

TYPE DE LABORATOIRE	LABO		fourn(LABO)		Moyenne	N	StdErr	niveaux	Moyenne	N	Std
	F	p	F	p							
ACCORD (Pearson)	2.11	0.147	3.99	<0.001	0.40	329	0.55	public	0.383 A	39	0.173
								privé	0.443 A	290	0.180
DISCRIMINATION (prob)	0.44	0.510	3.79	<0.001	0.85	317	0.37	public	0.136 A	38	0.102
								privé	0.117 A	279	0.120
DISCRIMINATION (pMCprod)	2.16	0.143	2.77	<0.001	0.62	317	0.28	public	0.6101 A	38	0.084
								privé	0.642 A	279	0.090
REPETABILITE (ETres)	24.55	<0.001	21.68	<0.001	1.18	317	0.79	public	1.472 B	38	0.389
								privé	1.176 A	279	0.401

L'analyse de l'effet du type de laboratoire (privé ou public) produit des résultats comparables à ceux des variables étudiées dans les paragraphes précédents. L'effet fournisseur(labo) est significatif pour tous les indices. Le type de laboratoire a un effet significatif sur le niveau de répétabilité uniquement. Cet effet montre un niveau de répétabilité moins satisfaisant des laboratoires publics. La cause de cet écart pourrait être lié à de multiples facteurs, comme la complexité des protocoles mis en place dans les laboratoires publics ou le nombre plus élevé d'échantillons à déguster (cf. tableau 5-2). Cette seconde hypothèse a déjà été envisagée lors de l'étude de l'évolution de la performance au fil des dégustations (chapitre IV). Les résultats montraient une baisse du niveau de répétabilité avec l'augmentation du nombre d'échantillons dégustés. Toutefois, ces conclusions doivent être relativisées puisque la plupart des laboratoires publics impliqués dans la SensoBase sont français (4 fournisseurs sur 5). Les résultats observés sont donc probablement corrélés à l'effet pays abordé précédemment.

3.2.5 Type de produit

Tableau 5-14 : Effet du type de produit sur les niveaux de performance des panels

TYPE DE PRODUIT	TYPE DE PRODUIT		fourn(TypeProd)		Moyenne	N	StdErr	niveaux	Moyenne	N	Std								
	F	p	F	p															
ACCORD (Pearson)	1.2340582	0.2889761	2.9794	<0.001	0.40	322	0.5289	Bakery	0.477 A	15	0.180								
								MeatFish	0.468 A	38	0.169								
								FruitVeg	0.459 A	38	0.168								
								Sauce	0.457 A	33	0.194								
								ReadyMM	0.451 A	28	0.147								
								DairyPdt	0.447 A	108	0.175								
								Beverage	0.363 A	62	0.163								
								Bakery	0.051 A	13	0.095								
								Sauce	0.092 A	33	0.137								
								DairyPdt	0.112 A	106	0.109								
DISCRIMINATION (prob)	0.7764222	0.5890681	2.85172	<0.001	0.85	310	0.3511	FruitVeg	0.115 A	38	0.091								
								ReadyMM	0.127 A	23	0.078								
								MeatFish	0.130 A	36	0.131								
								Beverage	0.152 A	61	0.130								
								Bakery	0.692 A	13	0.082								
								Sauce	0.653 A	33	0.107								
								ReadyMM	0.644 A	23	0.070								
								FruitVeg	0.635 A	38	0.094								
								DairyPdt	0.631 A	106	0.082								
								MeatFish	0.614 A	36	0.089								
DISCRIMINATION (pMCprod)	0.7941716	0.5751996	2.16595	<0.001	0.62	310	0.2783	Beverage	0.612 A	61	0.081								
								Sauce	1.094 A	33	0.238								
								ReadyMM	1.105 A	23	0.373								
								Beverage	1.153 A	61	0.429								
								DairyPdt	1.202 A	106	0.489								
								FruitVeg	1.35 A	38	0.331								
								Bakery	1.401 A	13	0.176								
								MeatFish	1.434 A	36	0.386								
								REPETABILITE (ETres)	5.933385	<0.001	21.4946	<0.001	1.17	310	0.7425				

Une fois encore, seul le niveau de répétabilité est influencé par la variable étudiée. Le type de produit ne modifierait donc pas le niveau d'accord ou de discrimination. Les écarts de répétabilité observés selon les catégories de produits doivent être considérés avec précaution car les effectifs sont très variables et certaines catégories reflètent les résultats de très peu de fournisseurs. Il nous semble donc plus raisonnable de laisser la question de l'effet du type de produit en suspens pour le moment. Elle pourra être envisagée à nouveau lorsque le nombre de fournisseurs représentés dans chaque catégorie de produit aura augmenté. Il serait également souhaitable d'affiner les catégories de produits car celles-ci regroupent des situations très différentes en l'état actuel.

4 Discussion - Conclusion

Le programme SensoBase a permis de rassembler des profils sensoriels d'origines très diverses. Dans ce chapitre, ces données ont été utilisées afin de mesurer les niveaux moyens de trois indices de performance puis l'effet de cinq variables intra-étude et cinq variables inter-études sur les niveaux de performance des panélistes et des panels.

Concernant les variables intra-étude, les résultats montrent que le niveau d'accord et de discrimination augmentent avec le niveau d'expérience en analyse sensorielle. Le niveau de discrimination s'élève également avec le niveau scolaire, mais ce dernier ne modifie pas les niveaux d'accord et de répétabilité. L'âge agit significativement sur l'indice de discrimination uniquement, mettant en avant des performances plus élevées pour les personnes entre 30 et 45 ans. L'effet du type de descripteur est significatif pour tous les indices de performance. La répétabilité observée est meilleure pour les descripteurs de saveurs, d'arômes et d'odeurs, mais ce résultat pourrait être lié à l'utilisation d'une portion plus restreinte de l'échelle de notation pour ces descripteurs. Les niveaux d'accord et de discrimination sont beaucoup plus élevés pour les descripteurs visuels. Le niveau d'accord est également meilleur pour les descripteurs de texture que pour les termes d'arôme, de saveur et d'odeur, mais la discrimination de la texture est relativement faible. L'effet sexe n'est jamais significatif parmi les panélistes sélectionnés pour participer à des profils sensoriels, mais les hommes auraient tendance à utiliser une portion de l'échelle plus importante que les femmes, lors de la notation de descripteurs visuels uniquement.

L'analyse des variables inter-études montre que la source majeure de variabilité entre les profils sensoriels réunis dans la SensoBase provient de l'effet "fournisseur", hautement significatif ($p < 0,001$) quel que soit l'indice de performance considéré. Cette variabilité de la performance est en outre plus importante au regard de l'indice de répétabilité. Pour les quatre autres variables étudiées (type d'échelle, laboratoire privé ou public, pays et type de produit), seul le niveau de répétabilité est significativement modifié selon les catégories de chaque facteur. L'effet est plus marqué concernant le facteur pays et montre un niveau de répétabilité plus faible des fournisseurs français. Puisque cet écart de performance n'est pas significatif concernant le niveau de discrimination, cela semble indiquer que l'utilisation de l'échelle est différente entre les fournisseurs français et étrangers. Une approche différente de l'entraînement des panélistes pourrait alors être en cause. Les différences observées sur le niveau de répétabilité pour les

variables "type d'échelle" et "laboratoire privé ou public" sont moins flagrantes. Elles pourraient en outre être liées à la corrélation entre ces facteurs et les variables "fournisseurs" et "pays". Les niveaux d'accord et de discrimination ne sont pas modifiés en fonction des variables inter-études.

Ce chapitre présente les premiers résultats exploratoires obtenus à l'aide de l'ensemble des profils de la SensoBase. Il serait maintenant intéressant de vérifier la validité de ces résultats par des analyses de corrélation entre les différents effets étudiés. En fonction de l'enclin des fournisseurs à envoyer des données toujours plus complètes, nous espérons à l'avenir pouvoir étudier l'effet d'autres variables, comme la fréquence de consommation du produit de l'étude, la motivation des panélistes (curiosité, occupation du temps libre ou rémunération), le fait qu'ils soient fumeurs ou non... Afin d'améliorer la qualité des résultats présentés dans ce chapitre, en particulier concernant les variables inter-études, nous devons à présent diversifier l'origine des données et créer des partenariats avec un grand nombre de fournisseurs différents. Il serait en effet intéressant de pouvoir considérer le fournisseur en tant qu'unité expérimentale de nos analyses.

Chapitre 6 : Etude de l'Assessor Model à l'aide des données de la SensoBase

1 Introduction

Dans le domaine de l'analyse sensorielle, la communication entre les sensométriciens et les praticiens a parfois la réputation de ne pas être très efficace. Les premiers développent en effet des méthodes statistiques toujours plus sophistiquées, mais ont assez peu accès à des études sensorielles réelles pour tester leurs modèles. Les seconds produisent énormément de données sensorielles mais utilisent une palette d'outils statistiques restreinte et parfois inadaptée. Les causes de ce manque de communication sont multiples. Tout d'abord, la complexité des méthodes proposées par les statisticiens peut être un frein à leur utilisation. Les acteurs de l'analyse sensorielle ont besoin de comprendre le fonctionnement des outils statistiques qu'ils utilisent pour interpréter de manière exacte leurs résultats. La vulgarisation des méthodes statistiques est donc une étape indispensable qui n'est pas systématiquement réalisée. De plus, les méthodes développées ne sont pas toujours facilement accessibles. La plupart sont programmées dans des langages comme SAS[®], MATLAB[®] ou R[®] et leur utilisation requiert un minimum de connaissances en informatique et en programmation. Les méthodes statistiques ne sont véritablement utilisées en analyse sensorielle que lorsqu'un logiciel qui les intègre existe. Enfin, le faible intérêt porté aux nouvelles méthodes statistiques est dû au manque d'information sur le gain réel que constituerait l'utilisation de ces méthodes. En effet, les publications mettent en général en avant l'intérêt théorique des méthodes par rapport aux techniques précédentes, puis illustrent leur propos à l'aide d'un ensemble de données restreint. Cependant, les méthodes ne sont jamais testées sur un grand nombre d'études sensorielles. Cela permettrait de vérifier que les améliorations théoriques proposées se traduisent par une amélioration de la finesse des interprétations, voire une modification des conclusions.

L'Assessor Model, développé par Brockhoff en 1994, fait, à notre avis, partie de ces méthodes statistiques sous-exploitées en analyse sensorielle. Celle-ci permet de prendre en compte l'hétérogénéité entre panélistes de la dispersion des notes, ce que très peu de méthodes réalisent à notre connaissance. Un programme SAS[®] automatisé est disponible en ligne sur le site de

l'auteur (<http://www.dina.kvl.dk/~per/>). Les publications sur ce sujet (Brockhoff et Skovgaard, 1994; Brockhoff, 1994; Brockhoff, 1998) détaillent l'intégralité des calculs et mettent en évidence le gain d'information apporté par ce modèle. Cette méthode reste cependant très peu utilisée par les praticiens de l'analyse sensorielle.

L'objectif du travail présenté dans ce chapitre est d'utiliser l'Assessor Model sur un grand nombre de profils sensoriels afin de vérifier que le modèle théorique défini dans cette méthode s'ajuste bien à la réalité des données sensorielles. Ce modèle d'analyse de la performance d'un jury est ensuite comparé à l'approche traditionnelle, faisant intervenir l'analyse de la variance. La réflexion menée avec Brockhoff autour de ces premiers résultats a permis de proposer un nouveau modèle, qui sera présenté dans un deuxième temps.

2 Etude de l'Assessor Model et comparaison avec l'approche classique

2.1 Matériel et Méthodes

2.1.1 Le modèle classique

Par modèle "classique", nous entendons le modèle d'analyse de la variance le plus couramment utilisé pour l'analyse des données de profil sensoriel. Ce modèle a déjà été présenté au chapitre 1 (paragraphe 3.1.2). Les notations ont cependant été modifiées afin de correspondre à celles utilisées avec l'Assessor Model. La signification des indices i (produit), j (sujet) et k (répétition) reste cohérente avec les notations de l'ensemble de ce manuscrit.

$$\begin{aligned}
 Y_{ijk} &= \mu + \alpha_j + v_i + \gamma_{ij} + \varepsilon_{ijk} \\
 \alpha_j &\sim N(0, \sigma_{\text{suj}}^2) \\
 \gamma_{ij} &\sim N(0, \sigma_{\text{suj}*\text{prod}}^2) \\
 \varepsilon_{ijk} &\sim N(0, \sigma_c^2) \quad \text{indépendants entre eux}
 \end{aligned}
 \tag{6-1}$$

Rappelons que ce modèle suppose l'homogénéité des variances des notes de chaque sujet (hypothèse d'homoscédasticité). La dispersion des notes est donc supposée identique à tous les panélistes. Rappelons également que le terme d'interaction est souvent considéré comme l'expression du désaccord entre les panélistes. Cette définition est exacte uniquement si les dispersions des notes sont effectivement homogènes entre les panélistes. Lorsque cette hypothèse est infirmée, le terme d'interaction reflète le désaccord entre les juges ainsi que la variabilité des dispersions. En pratique, cette distinction est très rarement prise en considération.

2.1.2 L'Assessor Model

Outre le fait que le niveau moyen de notation peut varier entre les panélistes (effet sujet), Brockhoff part du principe que la source majeure de variabilité entre les panélistes vient des différences de dispersion des notes et non d'un désaccord réel concernant les différences entre produit. L'Assessor Model se présente alors sous la forme suivante Brockhoff et Skovgaard, 1994 :

$$Y_{ijk} = \alpha_j + \beta_j \cdot v_i + \varepsilon_{ijk} \quad (6-2)$$
$$\varepsilon_{ijk} \sim N(0, \sigma_j^2) \quad \text{indépendants entre eux}$$

Le paramètre α_j représente l'effet du facteur sujet (comme dans le modèle classique). Le paramètre v_i représente l'effet moyen de chaque produit d'après l'ensemble du panel (équivalent à l'effet produit du modèle classique). Le paramètre β_j est un coefficient de dispersion des notes de chaque sujet par rapport aux autres panélistes. Ce coefficient multiplie l'effet produit. Un panéliste dont la valeur de β_j est supérieure à la moyenne du groupe reflète un sujet dont la dispersion des notes est élevée. D'un point de vue géométrique, ce coefficient peut être perçu comme la pente de la régression des moyennes des produits du juge j sur les moyennes des produits du panel^{*}. La figure 6-1 présente un exemple pour lequel le juge A disperse peu ses notes. Le paramètre β de ce panéliste est donc faible. Au contraire, la dispersion des notes du sujet B est importante, donc β est élevé.

^{*} En pratique, lors de la régression, les notes du panel (en abscisse sur la figure 6-1) sont centrées et réduites alors que les notes de chaque sujet (en ordonnées) sont uniquement centrées. La moyenne des β_j n'est donc pas égale à 1. Elle reflète la dispersion moyenne des notes des sujets. Sa valeur dépend également de l'unité de l'échelle de mesure.

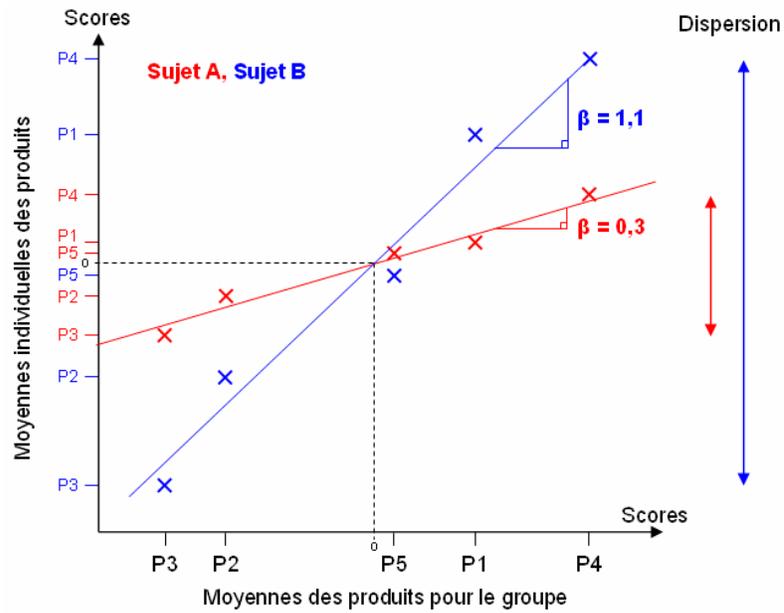


Figure 6-1 : Représentation géométrique du paramètre β de l'Assessor Model

Par rapport au modèle classique, l'Assessor Model spécifie en outre une variance résiduelle pour chaque sujet (σ_j^2). Cette particularité va permettre de tester l'homogénéité de la répétabilité individuelle des panélistes. En revanche, ce modèle ne fait pas apparaître de terme lié au désaccord entre les panélistes. Cette source de variabilité est donc incluse dans l'erreur ε_{ij} . Afin de comparer graphiquement les deux approches, le schéma de la figure 6-2 représente la décomposition de la variance totale en fonction de l'expression de chaque modèle.

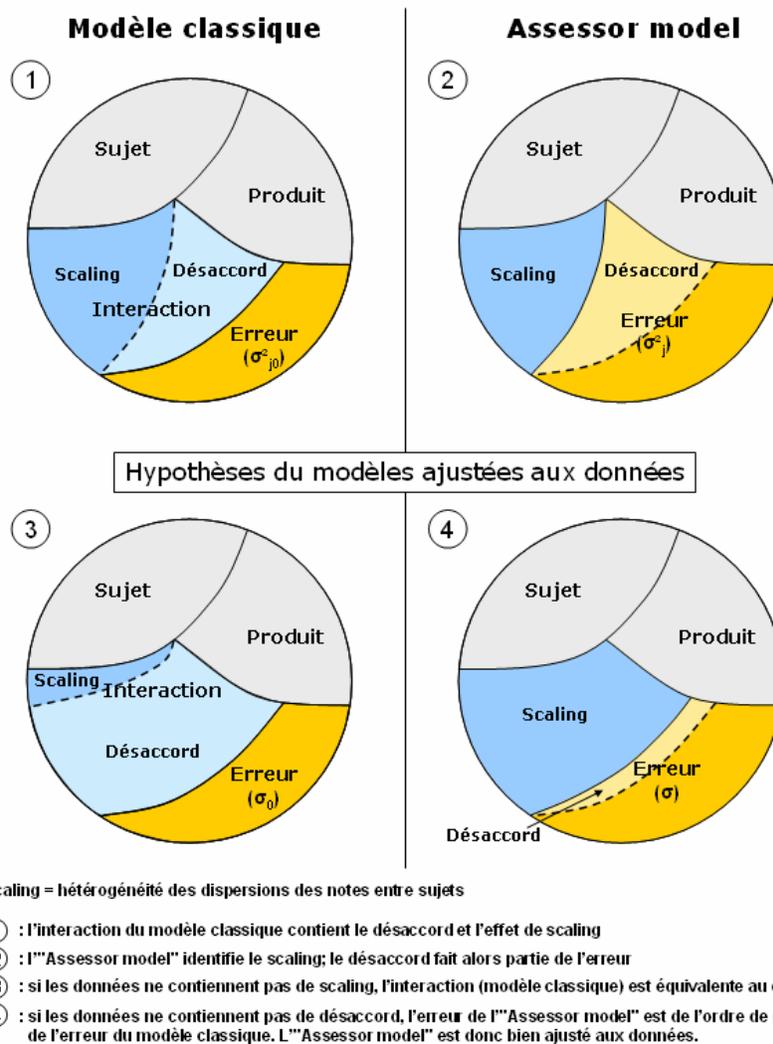


Figure 6-2 : Décomposition de la variance selon le modèle classique ou l'Assessor Model

La situation 1 met en évidence que le terme d'interaction du modèle classique renferme à la fois le désaccord et l'hétérogénéité de la dispersion des notes entre les panélistes (également appelée "scaling" dans ce chapitre). La partie haute de cette figure (situations 1 et 2) montre que l'Assessor Model explique une part de variabilité totale (zone bleue et grise) plus faible que le modèle classique. L'Assessor Model est en outre imbriqué dans le modèle classique. En effet, l'estimation des effets des facteurs sujet et produit sont identiques dans les deux modèles, mais l'estimation du scaling (paramètre β) n'utilise que J-1 degrés de libertés (un coefficient de dispersion β_j par juge) alors que l'interaction en utilise (I-1)(J-1). Les situations 1 et 2 mettent également en évidence que la différence entre l'erreur du modèle classique* (estimée à l'aide des variances résiduelles individuelles σ_{j0}^2) et l'erreur du modèle de Brockhoff estimée à l'aide des variances résiduelles individuelles σ_j^2) correspond au désaccord "pur".

* en spécifiant cependant une variance par sujet pour être homogène avec l'Assessor Model

Si le modèle classique est bien ajusté aux données (figure 6-2, situation 3), cela sous-entend que la part de variance expliquée par le scaling est faible et que l'approximation de l'interaction par le désaccord est justifiée. Si l'Assessor Model est bien ajusté aux données (figure 6-2, situation 4) la part de désaccord est faible et l'erreur de ce modèle est comparable à celle du modèle classique.

Afin de valider son modèle et de tester la performance des panélistes, Brockhoff propose quatre tests, présentés dans les paragraphes suivants.

2.1.2.1 Test d'homogénéité des variances résiduelles

Ce test n'utilise pas l'Assessor Model mais le modèle classique dans lequel une variance résiduelle, σ_{j0}^2 , est calculée pour chaque sujet. L'estimation de cette variance résiduelle peut être obtenue en réalisant un modèle d'ANOVA à un facteur (produit) pour chaque sujet. Le test d'hypothèse d'homogénéité des variances résiduelles individuelles est :

$$H_0 : \sigma_{10}^2 = \dots = \sigma_{j0}^2$$

La statistique de test utilisée est un khi-deux corrigé de Bartlett (Bartlett, 1937), dont le détail est présenté en annexe 8. Si le test est significatif, alors les variances résiduelles individuelles ne sont pas égales entre elles. Cela signifie que le niveau de répétabilité est différent d'un panéliste à l'autre. Dans ce cas, l'hypothèse d'homoscédasticité stipulée dans le modèle classique n'est pas respectée.

2.1.2.2 Test de la présence de désaccord

Ce test utilise le fait que l'Assessor Model est imbriqué dans le modèle classique (figure 6-2) et que la différence entre les variances résiduelles de ces deux modèles donne accès à la variance liée au désaccord "pur". Le test d'hypothèse est donc :

$$H_0 : \sqrt{\sigma_1^2 - \sigma_{10}^2} = \dots = \sqrt{\sigma_j^2 - \sigma_{j0}^2} = 0$$

Si l'hypothèse nulle est rejetée, l'écart entre les deux variances résiduelles est significatif. Il existe donc du désaccord "pur" entre les sujets. Ce résultat indique également que l'Assessor Model, emboîté dans le modèle classique, n'est pas aussi bien ajusté aux données que le modèle classique. Le modèle de Brockhoff ne devrait donc pas être utilisé. Au contraire, si l'hypothèse nulle est acceptée, l'Assessor Model est valide puisqu'il n'est pas significativement moins bien ajusté aux données que le modèle classique. Il pourrait donc être utilisé à la place du modèle classique. En outre, le désaccord entre les sujets est négligeable. Dans ces conditions, les paramètres estimés dans l'Assessor Model sont pertinents. Les deux derniers tests peuvent alors être conduits.

Le fait que les deux modèles soient emboîtés permet de tester leur égalité en calculant le logarithme du rapport de vraisemblance :

$$-2 \log \text{Vraisemblance} = IK \sum_{j=1}^J \log \left(\frac{\sigma_j^2}{\sigma_{j0}^2} \right) \quad (6-3)$$

Si l'hypothèse d'égalité des deux modèles est vraie, la distribution de cet indice suit asymptotiquement une distribution du khi-deux avec $IJ - (I+2J-2)$ degrés de libertés.

2.1.2.3 Test d'homogénéité de la dispersion des notes

Le coefficient de dispersion des notes de chaque panéliste est évalué par le paramètre β_j . Le test d'hypothèse d'homogénéité des dispersions est donc :

$$H_0 : \beta_1 = \dots = \beta_J$$

Lorsque l'hypothèse nulle est rejetée, les paramètres individuels de dispersion des notes ne sont pas égaux. Certains sujets utilisent donc une portion de l'échelle de notation plus importante que d'autres. La stratégie de test utilise également le logarithme du rapport de vraisemblance. L'Assessor Model est testé par rapport au modèle (emboîté) qui suppose l'égalité des coefficients β_j , soit :

$$Y_{ijk} = \alpha_j + v_i + \varepsilon_{ijk} \quad (6-4)$$

$$\varepsilon_{ijk} \sim N(0, \sigma_j^2)$$

Le calcul de la vraisemblance associée permet de déterminer si le modèle (6-4) est équivalent à l'Assessor Model (équation 6-2). Lorsque la statistique de test du khi-deux indique que les modèles sont équivalents, l'ajout du paramètre β n'est pas justifié et la dispersion des notes est relativement homogène entre les sujets. Dans le cas contraire, l'Assessor Model apporte significativement plus d'information grâce au paramètre β : la dispersion des notes change d'un panéliste à l'autre.

2.1.2.4 Test d'homogénéité de la sensibilité

Brockhoff définit la sensibilité de chaque panéliste par le rapport (Brockhoff et Skovgaard, 1994) :

$$\text{sensibilité}_j = \frac{\beta_j^2}{\sigma_j^2} \quad (6-5)$$

Pour le panéliste j , une sensibilité élevée indique que la dispersion des moyennes des produits (β_j) est forte par rapport à la dispersion moyenne des répétitions de chaque produit (σ_j). La

signification de cet indice est similaire à celle de la statistique de Fisher de l'effet produit. Les auteurs démontrent d'ailleurs la relation suivante (Brockhoff et Skovgaard, 1994) :

$$\frac{\beta_j^2}{\sigma_j^2} \approx a \cdot \left(\frac{CM_{produit}}{CM_{erreur}} - 1 \right) = a \cdot (F_{produit} - 1) \quad (6-6)$$

Où "a" désigne une constante.

Le test d'hypothèse d'homogénéité de la sensibilité est donc :

$$H_0 : \frac{\beta_1^2}{\sigma_1^2} = \dots = \frac{\beta_j^2}{\sigma_j^2}$$

Lorsque l'hypothèse nulle est rejetée, la sensibilité des sujets est hétérogène. En d'autres termes, le pouvoir discriminant varie entre les panélistes. L'évaluation de l'homogénéité de la sensibilité utilise également le logarithme du rapport de vraisemblance. L'Assessor Model est alors testé par rapport au modèle qui impose l'égalité des rapports β_j^2 / σ_j^2 . Brockhoff démontre que la définition d'un tel modèle revient à ajouter une contrainte lors du calcul du coefficient β (l'explication détaillée est disponible dans Brockhoff, 2003). Lorsque l'égalité des modèles est vérifiée, les sensibilités entre panélistes sont homogènes.

Les quatre tests proposés permettent de caractériser la performance des panélistes. En revanche, aucun test de l'effet produit n'est proposé. Ceci est lié au fait que, dans l'Assessor Model, l'effet produit v_i , est multiplié au coefficient de dispersion de chaque sujet β_j . Cette structure particulière ne permet pas de calculer directement la part de variabilité totale due à l'effet produit, donc l'effet produit ne peut pas être évalué avec ce modèle.

2.1.3 Comparaison des résultats des deux modèles

L'objectif de la comparaison des résultats entre le modèle classique et l'Assessor Model est d'évaluer si les différences théoriques entre ces deux approches ont une répercussion sur la détection du désaccord entre les panélistes et sur la significativité de l'effet produit. Afin d'obtenir une réponse fiable à cette question, 264 profils sensoriels de la SensoBase (parmi les 329 études utilisées dans le chapitre 5) ont été analysés en utilisant les deux modèles. Les études non sélectionnées sont celles sans répétitions, ou dans lesquelles le plan d'expérience était très déséquilibré.

La comparaison des modèles porte d'abord sur la conclusion quant à l'existence du désaccord entre les panélistes. L'interaction produit* sujet du modèle classique est souvent interprétée comme un terme de désaccord entre les panélistes. L'Assessor Model utilise le test du logarithme du rapport de vraisemblance entre les deux modèles (paragraphe 2.1.2.2) pour détecter la présence de désaccord. Pour chaque descripteur de chaque profil, la comparaison peut aboutir à l'un des quatre diagnostics présentés dans le tableau 6-1.

Tableau 6-1 : comparaison de la significativité du désaccord entre le modèle classique et l'Assessor Model. Diagnostics possibles.

		Interaction sujet*produit (modèle classique)	
		significative	non significative
Test du Désaccord (Assessor model)	significatif	✓	désaccord = erreur
	non significatif	interaction = scaling	✓

Lorsque les deux tests sont significatifs ou non significatifs simultanément (cases vertes), les conclusions sur la présence de désaccord sont identiques. Les résultats ne sont pas modifiés lors de l'utilisation du modèle de Brockhoff. Il est donc possible de conserver le modèle classique, plus simple à utiliser. En revanche, lorsque l'interaction du modèle classique est significative mais que le test du désaccord utilisé avec l'Assessor Model n'est pas significatif (case rouge), cela implique que l'interaction détecte en réalité des différences de dispersion des notes entre les panélistes (scaling). La conclusion du modèle classique sur la présence de désaccord est donc erronée. La dernière situation (case orange) correspond à la situation inverse : le test du désaccord est significatif, contrairement à l'interaction du modèle classique. L'explication d'un tel résultat est liée à la nature des deux tests. Ceux-ci ne mesurent pas la significativité du désaccord de la même manière, ce qui provoque parfois des différences de résultats. Ce type de conclusion devrait cependant rester marginale. Dans les deux situations problématiques (cases rouge et orange), nous chercherons à déterminer si le choix du modèle a un impact sur la significativité de l'effet produit. Cette question est primordiale pour l'animateur de panel, dont l'objectif final est de mesurer ces différences avec précision.

2.1.3.1 Détection du désaccord à tort par le modèle classique (case rouge)

Lorsque le modèle classique détecte du désaccord à tort (case rouge), le problème vient de l'hétérogénéité des dispersions des notes entre les panélistes. Celle-ci n'est pas prise en compte dans le modèle classique. Afin de palier ce problème, nous proposons de standardiser les données, de manière à homogénéiser la dispersion des notes entre les panélistes à l'aide de la transformation suivante :

$$Z_{ijk} = \frac{Y_{ijk} - \alpha_j}{s_j} \quad (6-7)$$

Les notes de chaque sujet sont donc centrées puis divisées par leur variance. Le modèle classique est alors réutilisé avec les données transformées. Pour chaque descripteur, la significativité du nouvel effet produit est ensuite comparée à sa valeur sans transformation. Si l'effet devient significatif, l'hétérogénéité des dispersions des notes entre les panélistes augmentait artificiellement le terme d'interaction et dissimulait l'existence des différences entre produits. Dans ce cas, le modèle classique n'est pas adapté aux données et entraîne des conclusions fausses au niveau des différences entre produits. Il est alors fortement recommandé de tenir compte de l'hétérogénéité de la dispersion des notes entre les panélistes. Si l'effet produit ne devient pas significatif, la prise en compte des différences de dispersion ne modifie pas les conclusions. Par conséquent, bien que la variabilité des dispersions inter-panélistes soit significative et que le modèle classique n'en tienne pas compte, cela n'a pas d'influence sur l'interprétation de l'effet produit. Afin de vérifier l'impact de la transformation des données sur les différences entre produits, le coefficient de corrélation de Pearson est également calculé sur les moyennes des produits (sur l'ensemble du panel) avec ou sans transformation. Une valeur faible indiquerait que la transformation altère l'ordonnement des produits.

2.1.3.2 Détection du désaccord à tort par l'Assessor Model (case orange)

Lorsque le modèle de Brockhoff détecte du désaccord alors que l'interaction n'est pas significative, cela implique que le terme d'interaction est de l'ordre de grandeur du terme d'erreur. Cette affirmation sera cependant vérifiée par le calcul dans chaque situation. De plus, si l'interaction n'est pas significative, elle représente une autre estimation de l'erreur résiduelle, mais elle possède moins de degrés de liberté que le terme d'erreur (lorsque tous les sujets ont effectué au moins deux répétitions de chaque produit). La statistique de Fisher de l'effet produit sera donc également calculée en utilisant le terme d'erreur (et non le terme d'interaction) au dénominateur, afin de détecter d'éventuelles modifications de la significativité. Notons que ce test n'est pas lié à la comparaison des deux modèles. Nous profitons simplement de l'occasion pour évaluer l'impact de la modification du terme d'erreur dans le calcul de l'effet produit.

2.2 Résultats

2.2.1 Tests de l'Assessor Model

Les résultats des quatre tests proposés par Brockhoff, réalisés sur 264 études sensorielles, soit 7024 descripteurs, sont les suivants :

- **la répétabilité des panélistes est hétérogène pour 73% des descripteurs.** Ce résultat indique que dans près de trois quarts des cas, l'hypothèse d'homoscédasticité stipulée dans le modèle classique n'est pas valide. Ce modèle ne devrait donc théoriquement pas être utilisé. Sans pour autant utiliser l'Assessor

Model, il serait donc conseillé de pondérer les analyses de la variance par l'inverse des écart-types individuels.

- **Le test du désaccord entre panélistes est significatif pour 25,2% des descripteurs.** Par conséquent, l'Assessor Model est significativement moins bien ajusté aux données que le modèle classique pour un quart des descripteurs. Dans ces situations, les estimations des dispersions individuelles des notes, et donc des sensibilités individuelles données par le modèle de Brockhoff, sont biaisées par le désaccord.
- Parmi les descripteurs pour lequel le désaccord est non significatif, **la dispersion des notes entre panélistes est hétérogène dans 76,4% des cas.** L'approche proposée par l'Assessor Model, qui fait intervenir explicitement un terme de dispersion dans son expression, est donc en accord avec la réalité des données dans une large majorité des cas. En outre, l'interaction du modèle classique n'est en général pas équivalente au désaccord entre les panélistes ; elle inclut également la variabilité due au scaling. La prise en compte de l'hétérogénéité de la dispersion, à l'aide du modèle de Brockhoff ou d'une transformation des données permettant de réduire ces différences, est donc généralement souhaitable.
- Parmi les descripteurs pour lequel le désaccord est non significatif, **la sensibilité des panélistes est hétérogène pour 70,5% des descripteurs.** Le pouvoir discriminant n'est donc pas homogène entre les sujets. Cette information n'était pas disponible avec le modèle classique.

L'Assessor Model propose donc une modélisation pertinente du mode de notation des panélistes dans la majorité des cas. Les indices proposés permettent en outre d'accéder à une information sur l'hétérogénéité de la répétabilité, de la dispersion des notes et de la sensibilité des sujets, ce que ne permet pas le modèle classique. Le test du désaccord de l'Assessor Model permet enfin d'approcher la significativité du désaccord "pur", contrairement au modèle classique dans lequel désaccord et scaling sont réunis dans le terme d'interaction. Le modèle classique est donc remis en question, en particulier parce qu'il ne tient pas compte de l'hétérogénéité de la dispersion des notes selon les panélistes.

2.2.2 Comparaison des deux modèles

La comparaison des résultats, basée sur la significativité du désaccord selon chaque approche, est synthétisée dans le tableau 6-2.

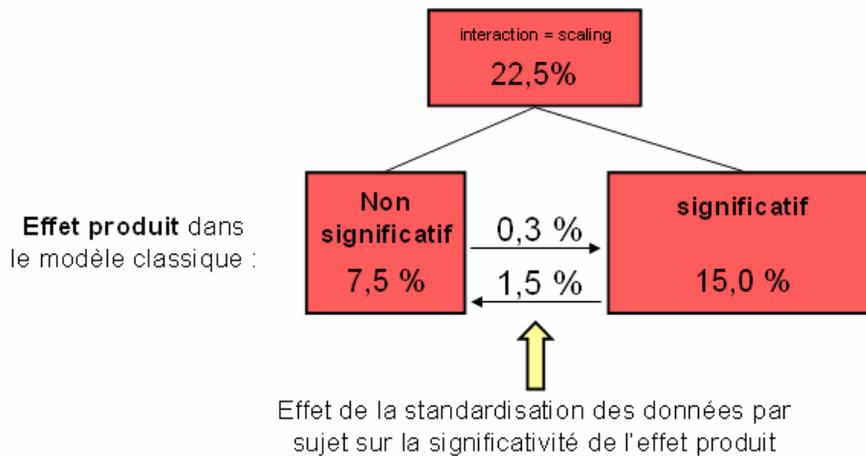
Tableau 6-2 : Résultats de la comparaison des deux modèles par rapport au désaccord

		Interaction sujet*produit (modèle classique)	
		significative	non significative
Test du Désaccord (Assessor model)	significatif	22,3%	désaccord = erreur 2,9%
	non significatif	interaction = scaling 22,5%	52,3%

Malgré les différences théoriques entre les deux modèles et les résultats des tests liés à l'Assessor Model, la détection du désaccord concorde dans 74,6% des cas (cases vertes). Le modèle classique reste donc satisfaisant dans ces situations. Cependant, lorsque l'interaction du modèle classique est significative (44,8% des descripteurs), elle reflète l'hétérogénéité des dispersions des notes entre les panélistes, plutôt que le désaccord, dans la moitié des situations (case rouge = 22,5%). Le désaccord n'est donc pas correctement appréhendé dans ce cas. Ce phénomène peut avoir des répercussions sur la significativité de l'effet produit. Enfin, lorsque l'interaction est non significative (55,2%), le test du désaccord l'est généralement aussi. En effet, seuls 2,9% des descripteurs exhibent un désaccord entre les panélistes d'après le test de Brockhoff alors que l'interaction n'est pas significative. Cette situation rare s'explique par la nature des deux tests (cf. paragraphe 2.1.3). Pour les deux situations où les modèles divergent, l'analyse se porte ensuite sur l'impact sur la significativité de l'effet produit.

2.2.2.1 Prise en compte de l'hétérogénéité de la dispersion entre panélistes

Lorsque l'interaction est significative, contrairement au test du désaccord (case rouge), les données sont standardisées afin d'homogénéiser la dispersion des notes de chaque panéliste. Le modèle classique est ensuite réutilisé sur ces données transformées pour observer d'éventuelles modifications de la significativité de l'effet produit. Les résultats, présentés dans la figure 6-3, montrent une légère modification des résultats. Lorsque l'effet produit était non significatif (7,5% des descripteurs), le test de l'effet produit sur les données transformées devient significatif dans 4,19% des cas (soit 0,3% de l'ensemble des descripteurs de l'étude). A l'inverse, la transformation provoque une perte de la significativité de l'effet produit pour 9,95% des descripteurs significatifs avant transformation (soit 1,5% du nombre total de descripteur). De manière générale, la standardisation des données n'améliore donc pas les résultats. Par conséquent, bien que le modèle classique ne tienne pas compte des différences de dispersion des notes entre les panélistes, l'incidence sur la significativité de l'effet produit est très faible avec la transformation proposée. Cette conclusion remet également en question la pertinence de la transformation utilisée pour réduire l'hétérogénéité des dispersions des notes.



→ % de descripteurs pour lequel l'effet produit est devenu significatif après standardisation des données par sujet
 ← % de descripteurs pour lequel l'effet produit est devenu NON significatif après standardisation des données par sujet
 NB : tous les pourcentages présentés dans cette figure sont exprimés en fonction du nombre total de descripteurs (soit 7024).

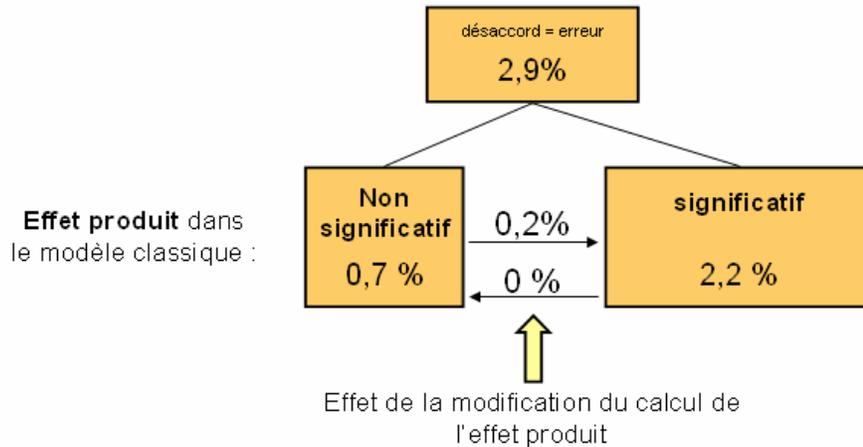
Figure 6-3 : Effet de la transformation des données sur la significativité de l'effet produit dans le modèle classique

Afin de mesurer l'impact de la transformation des notes sur les différences entre produits sous un autre angle, le coefficient de corrélation de Pearson est calculé sur les moyennes des produits entre les deux situations. Lorsque ce coefficient est significatif, cela indique que la standardisation des données n'a pas modifié les différences entre produits. Les résultats montrent que le coefficient de corrélation est positif dans 99,9% des cas et significatif dans 72% des cas (risque $\alpha = 0,05$). La transformation modifie donc la nature des différences entre produits dans 28% des situations. Cette valeur relativement élevée indique que la transformation n'est pas tout à fait neutre et qu'elle peut avoir un impact sur les différences observées entre les produits.

2.2.2.2 Substitution du terme d'interaction par le terme d'erreur dans le calcul de l'effet produit

Lorsque le test du désaccord est significatif alors que l'interaction du modèle classique ne l'est pas (case orange), les calculs montrent que la somme des carrés de l'interaction est relativement élevée par rapport à la somme des carrés totale. En effet, l'interaction représente en moyenne 20,8% de la variance totale dans ce cas, alors qu'elle ne représente que 16,8% lorsque les tests de l'interaction et du désaccord sont non significatifs. Le test de la différence entre ces deux proportions est hautement significatif ($p < 0,001$). Ceci explique la significativité du test du désaccord proposé par Brockhoff. Cependant, les sommes des carrés de l'interaction et de l'erreur sont du même ordre de grandeur. Le rapport $SC_{interaction}/SC_{erreur}$ est de 0,89 en moyenne et sa valeur ne dépasse jamais 1,2. Ceci explique que l'interaction ne soit pas significative. Dans ce cas, la variance de l'interaction équivaut à une estimation de la variance résiduelle (paragraphe 3.1.2, chapitre 1). Nous évaluons alors si la substitution du carré moyen de l'interaction par le carré moyen de l'erreur, dans le calcul de l'effet produit, modifie la significativité de ce dernier. Les résultats sont présentés dans la figure 6-4. La modification réalisée a un impact sur l'effet produit

dans 24,2% des cas pour lesquels les descripteurs non significatifs (0,7%) le deviennent (soit 0,2% du nombre total de descripteurs). Ce résultat est principalement lié au fait que le nombre de degrés de liberté de l'erreur est généralement supérieur à celui de l'interaction. Par conséquent, si le carré moyen de l'erreur est égal à celui de l'interaction, les statistiques de Fisher de l'effet produit selon les deux modes de calcul seront identiques, mais le seuil de significativité sera plus faible lorsque le terme d'erreur est utilisé.



→ % de descripteurs pour lequel l'effet produit est devenu significatif en utilisant CM_{erreur} au dénominateur du calcul de la statistique de Fisher de l'effet produit

← % de descripteurs pour lequel l'effet produit est devenu NON significatif en utilisant CM_{erreur} au dénominateur du calcul de la statistique de Fisher de l'effet produit

NB : tous les pourcentages présentés dans cette figure sont exprimés en fonction du nombre total de descripteurs (soit 7024).

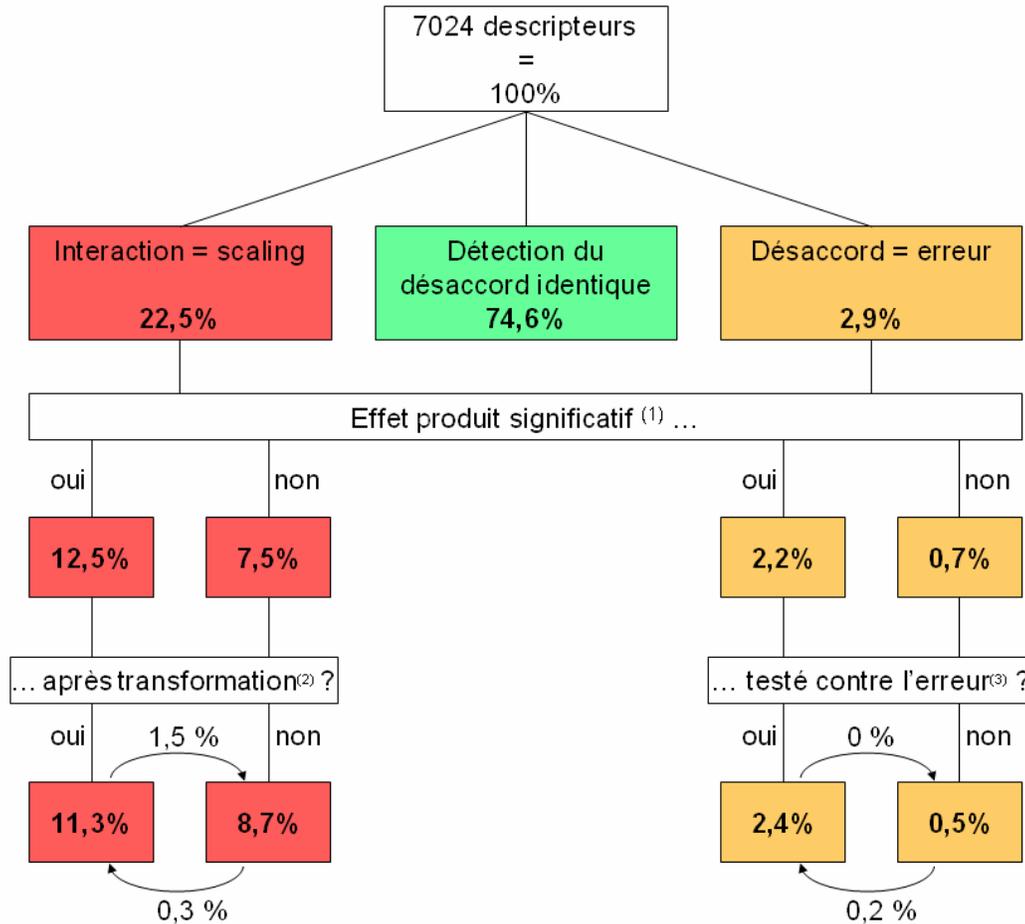
Figure 6-4 : Effet de la modification du dénominateur (CM_{erreur} au lieu de $CM_{interaction}$) dans le calcul de l'effet produit

2.3 Conclusion

Les résultats des quatre premiers tests proposés par Brockhoff démontrent clairement que la modélisation des données sensorielles est plus satisfaisante avec l'Assessor Model dans l'optique de la mesure de la performance. En effet, la répétabilité des sujets se révèle très souvent hétérogène dans les 267 études analysées. L'hypothèse d'homoscédasticité du modèle classique est donc assez rarement vérifiée. De plus la dispersion des notes est très souvent hétérogène et cette source de variabilité n'est pas explicitement prise en compte dans le modèle classique.

La figure 6-5 propose une synthèse des résultats de la comparaison des deux modèles. Elle met en évidence que, dans la pratique, le modèle classique et l'Assessor Model s'accordent dans la majorité des cas (74,6%). Néanmoins, l'interaction du modèle classique est significative alors que le test du désaccord accepte l'hypothèse nulle pour 22,5% des descripteurs. Dans ce cas, le terme d'interaction est surestimé par rapport à sa vraie valeur et l'hétérogénéité de la dispersion des notes n'est pas négligeable. Les considérations théoriques mises en œuvre dans le modèle proposé par Brockhoff modifient donc les conclusions sensorielles concernant le désaccord dans un nombre important de cas. Au niveau de la significativité de l'effet produit, les répercussions

semblent en revanche mineures. La standardisation des données améliore en effet très peu la significativité de l'effet produit (4,51% des descripteurs concernés, soit 0,3% du nombre total de descripteur). Cette transformation a même tendance à provoquer l'effet inverse : lorsque l'effet produit était significatif, il ne l'est plus dans 10% des cas (soit 1,5% du nombre total de descripteurs).



(1): lorsque les résultats du test du désaccord de Brockhoff et du test de l'interaction dans le modèle classique ne concordent pas, les descripteurs sont classés en fonction de la significativité de l'effet produit dans le modèle classique

(2): standardisation des données par sujet afin de supprimer l'hétérogénéité des variances

(3): statistique de Fisher de l'effet produit = $CM_{produit} / CM_{erreur}$

↔ % de descripteurs pour lesquels la significativité de l'effet produit a été modifiée

NB : tous les pourcentages présentés dans cette figure sont exprimés en fonction du nombre total de descripteurs (soit 7024).

Figure 6-5 : synthèse de la comparaison entre le modèle classique et l'Assessor Model

Toutefois, la transformation utilisée peut être remise en question pour deux raisons. D'une part, la somme des carrés de l'effet produit est modifiée par la standardisation des données, donc l'intensité des différences entre produits au niveau du groupe est perturbée. D'autre part, lorsque la valeur de l'écart-type d'un juge est très proche de zéro, la transformation conduit à une expansion très importante des notes du panéliste. Cela provoque une forte augmentation de l'interaction alors que la transformation vise à diminuer l'importance de ce terme. Techniquement, il serait préférable d'utiliser une transformation qui ne modifie pas la variance expliquée par l'effet

sujet et l'effet produit, mais qui diminue la variance de l'effet de l'interaction. Au cours de nos discussions avec Per Brockhoff, il nous a alors proposé une autre transformation, qui a finalement débouché sur un nouveau modèle aux caractéristiques intéressantes. Cette nouvelle approche est décrite dans la fin de ce chapitre.

3 Transformation proposée par Brockhoff et modification du modèle

3.1 Méthode

Plutôt que de standardiser les données par sujet, Brockhoff propose la transformation suivante (communication personnelle):

$$Z_{ijk} = Y_{ijk} - (\hat{\beta}_j - \bar{\beta})\hat{v}_i \quad (6-8)$$

Le terme $(\hat{\beta}_j - \bar{\beta})$ correspond à la différence entre le coefficient de dispersion du panéliste j et le coefficient moyen. Il est positif si le panéliste utilise une portion de l'échelle plus importante que la moyenne du groupe pour effectuer ses notations. Le terme \hat{v}_i est la moyenne du produit i estimée d'après l'Assessor Model (équation 6-2). La figure 6-6 représente graphiquement la transformation :

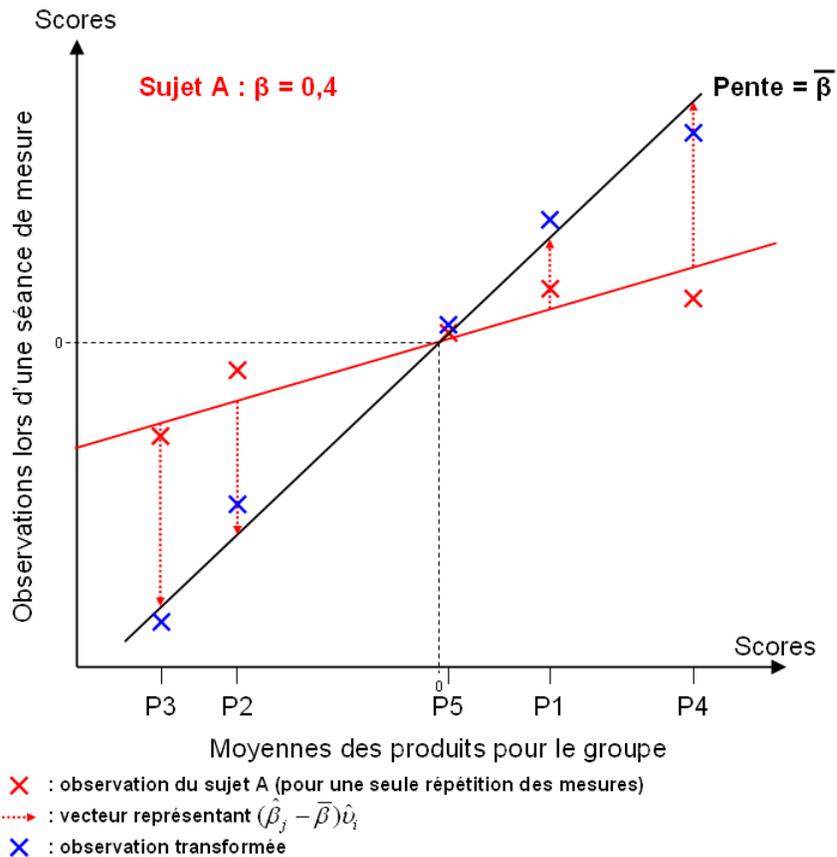


Figure 6-6 : Représentation graphique de la transformation proposée par Brockhoff

Pour le point d'abscisse \hat{v}_i , l'expression $(\hat{\beta}_j - \bar{\beta})\hat{v}_i$ correspond à l'écart entre les deux droites de pentes $\hat{\beta}_j$ et $\bar{\beta}$. Cela s'interprète comme l'écart entre la dispersion des notes du sujet j et la dispersion moyenne du panel. Cette valeur est retranchée à la note donnée par le sujet j au produit i. La transformation permet donc de faire coïncider la dispersion de chaque sujet avec celle du groupe. Au niveau de l'ensemble du panel, cette transformation revient à supprimer l'hétérogénéité des pentes entre les sujets.

Cette étape peut également être réalisée directement dans le modèle classique, en ajoutant le terme multiplicatif : $\beta'_j \cdot \hat{v}_i$. Cette modification revient à ajouter la covariable contenant les effets produits \hat{v}_i , estimés à partir de l'Assessor Model^{*}, et de prendre en compte la variabilité des pentes individuelles β'_j , calculées entre les notes moyennes du groupe et les notes de chaque panéliste. Par rapport à l'expression du modèle de Brockhoff, on a la relation :

$$\hat{\beta}'_j = \hat{\beta}_j - \bar{\beta} \quad (6-9)$$

Le modèle d'ANOVA tenant compte de l'hétérogénéité de la dispersion des notes entre sujet s'écrit alors :

$$\begin{aligned} Y_{ijk} &= \mu + \alpha_j + v_i + \beta'_j \cdot \hat{v}_i + \gamma_{ij} + \varepsilon_{ijk} \\ \alpha_j &\sim N(0, \sigma_{\text{suj}}^2) \\ \gamma_{ij} &\sim N(0, \sigma_{\text{suj} \cdot \text{prod}}^2) \\ \varepsilon_{ijk} &\sim N(0, \sigma_c^2) \quad \text{indépendants entre eux} \end{aligned} \quad (6-10)$$

Nous avons vérifié par le calcul que ce modèle était équivalent au modèle classique après avoir transformé les données selon l'équation (6-8). Les sommes de carrés et les carrés moyens des effets produit, sujet, interaction sujet*produit et de l'erreur sont strictement identiques.

Ce modèle, utilisé avec des sommes de carrés séquentielles (type I de SAS[®]), permet en fait d'extraire la part de variance attribuée à l'hétérogénéité de la dispersion des notes ($\beta'_j \cdot \hat{v}_i$), de l'interaction habituelle (γ_{ij}). Cette dernière, ne contient alors plus que la variance imputable au désaccord entre les panélistes. Il devient donc possible de tester, dans un seul modèle, la significativité du désaccord et de l'hétérogénéité de la dispersion, ainsi que les effets sujet et produit.

Ce modèle a l'avantage de pouvoir être calculé dans tous les logiciels de statistiques disposant d'un module d'analyse de la variance avec covariable. Son utilisation présente néanmoins une difficulté : le calcul du terme \hat{v}_i . En effet, l'obtention des valeurs de cette covariable implique

^{*} L'estimation des moyennes des produits au niveau du groupe est réalisée en pondérant les observations individuelles en fonction de la sensibilité de chaque panéliste. Les sujets plus discriminants influencent donc davantage le calcul de la moyenne.

d'utiliser l'Assessor Model au préalable, disponible uniquement sous SAS® actuellement. Notons que cette covariable contient les estimations des moyennes des produits du panel en tenant compte de l'hétérogénéité de la dispersion de leurs notes. Une première approximation triviale pourrait alors être la moyenne simple de chaque produit sur l'ensemble du panel (notée \bar{x}_i). L'objectif de la suite de ce travail est donc de comparer les résultats obtenus en utilisant les deux covariables possibles (\hat{v}_i ou \bar{x}_i) dans le modèle 6-10, aux résultats obtenus avec le modèle classique.

3.2 Résultats

Le tableau 6-3 rassemble, pour les trois modèles envisagés, les pourcentages de descripteurs dont la statistique de Fisher est significative pour les effets sujet*produit, produit et sujet*covariable (\hat{v}_i ou \bar{x}_i). Les résultats (tableau 6-3) montrent clairement que l'ajout du facteur faisant intervenir la covariable permet de réduire la taille de l'interaction sujet*produit, donc sa significativité. En effet, cette interaction est significative pour 44,8% des descripteurs avec le modèle classique, mais seulement dans 18,41% des cas avec le modèle de l'équation 6-10. Le remplacement des moyennes des produits (\hat{v}_i), calculés avec l'Assessor Model, par les moyennes simples (\bar{x}_i), constitue une approximation relativement acceptable puisque l'interaction devient significative pour seulement 8,8% de descripteurs supplémentaires, soit 27,21%. Le nouveau modèle proposé par Brockhoff entraîne également la détection plus fréquente d'un effet produit significatif (68,46% avec le nouveau modèle ; 59,36% avec le modèle classique). L'approximation par la moyenne \bar{x}_i donne un résultat intermédiaire. L'interaction sujet*covariable est significative pour 68,15% des descripteurs avec le modèle de l'équation 6-10. Cette valeur est proche du pourcentage obtenu lors du test de l'hétérogénéité des coefficients β_j (paragraphe 2.2.1). Ce pourcentage diminue lors de l'utilisation des moyennes simples (\bar{x}_i).

Tableau 6-3 : Pourcentage de descripteurs dont l'effet est significatif selon le modèle utilisé

	Modèle utilisant \hat{v}_i	Modèle utilisant \bar{x}_i	Modèle classique
Sujet*Produit	18,41 %	27,21 %	44,82 %
Produit	68,46 %	63,26 %	59,36 %
Sujet*Covariable	68,02 %	53,09 %	--

3.3 Conclusion

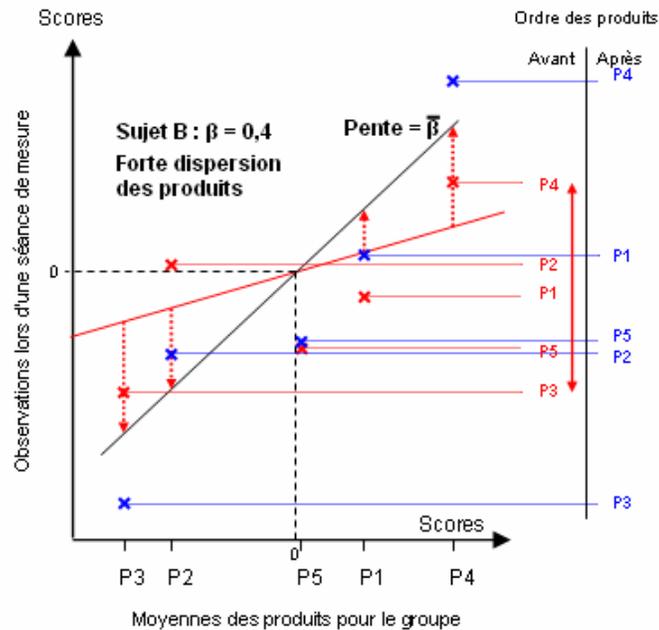
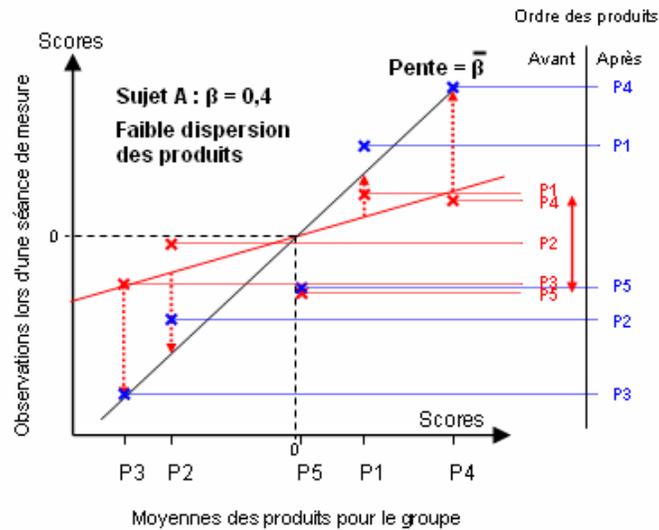
Conformément à nos attentes, le nouveau modèle proposé par Brockhoff permet de réduire le terme d'interaction et ainsi de détecter plus fréquemment un effet produit significatif qui pouvait être masqué. L'approximation des moyennes des produits, obtenues à l'aide de l'Assessor Model

(\hat{v}_i), par les moyennes simples (\bar{x}_i) conduit à des résultats intermédiaires entre l'approche classique et la nouvelle méthode. Cette dernière possibilité pourrait donc constituer une alternative intéressante au modèle classique et à l'Assessor Model puisqu'elle allie la prise en compte (non optimale, cependant) de l'hétérogénéité des dispersions entre panélistes et la simplicité d'utilisation (elle ne nécessite pas l'utilisation de l'Assessor Model). Cependant, à l'instar du modèle classique, ce modèle suppose également l'hypothèse d'homoscédasticité des variances. Celle-ci n'est pourtant généralement pas vérifiée d'après les résultats donnés par l'Assessor Model. Il serait donc préférable de spécifier des variances hétérogènes en pondérant les notes de chaque sujet par l'inverse de l'écart-type de ses notes moyennes (dans le modèle d'analyse de la variance).

4 Discussion-conclusion

L'utilisation de l'Assessor Model sur un grand nombre d'études sensorielles a permis de montrer l'existence d'une hétérogénéité des dispersions des notes entre les panélistes dans une proportion importante des situations (76,4% des descripteurs). Ce modèle a également permis de montrer que les variances résiduelles individuelles sont souvent hétérogènes. Théoriquement, le modèle classique ne devrait donc pas être utilisé dans ce cas. Enfin, les tests proposés permettent aussi de caractériser l'hétérogénéité de la sensibilité des panélistes (i.e. leur pouvoir discriminant) et de conclure sur la présence de désaccord "pur". L'Assessor Model apporte donc beaucoup d'informations nouvelles sur les niveaux de performance des panélistes par rapport au modèle classique. Son utilisation par les praticiens de l'analyse sensorielle leur permettrait d'obtenir, à notre avis, un gain d'information non négligeable.

Au niveau de la conclusion sur la significativité de l'effet produit, la standardisation des données ne génère pas de meilleurs résultats. En revanche, la transformation proposée par Brockhoff améliore sensiblement la significativité de l'effet produit. Celle-ci repose sur le calcul des moyennes des produits au niveau du groupe en tenant compte de l'hétérogénéité des dispersions des notes selon les panélistes. Elle consiste ensuite à supprimer l'hétérogénéité des pentes, issues de la régression des notes individuelles sur les notes moyennes, calculées au préalable. L'approximation de ces moyennes, tenant compte de la variabilité de la dispersion des notes, par la moyenne simple, constitue une alternative intéressante. Elle ne permet pas de capturer aussi bien l'hétérogénéité des dispersions mais permet néanmoins de tenir compte en partie de cette source de variabilité. En outre, cette seconde transformation n'utilise pas l'Assessor Model. La dernière solution proposée peut donc être mise en place avec n'importe quel logiciel de statistique permettant de spécifier des modèles d'analyse de la variance avec covariable.



- ✗ : observation du sujet A (pour une seule répétition des mesures)
- ⋯→ : vecteur représentant $(\hat{\beta}_j - \bar{\beta})\hat{O}_i$
- ✕ : observation transformée

Figure 6-7 : Relation entre le coefficient β et la dispersion des notes

Cependant, l'interprétation du paramètre β de l'Assessor Model comme un coefficient qui représente la dispersion individuelle des notes est discutable. En effet, la figure 6-7 illustre le fait que deux panélistes avec le même coefficient β peuvent avoir des dispersions des notes très différentes selon leur accord avec le groupe. En effet, les observations d'un sujet parfaitement en accord avec le groupe seront alignées sur la droite de régression. Au contraire, les observations d'un sujet en désaccord, avec le même coefficient β , seront plus écartées de la droite. La

dispersion de ses notes sera donc obligatoirement plus élevée. Le coefficient β ne s'interprète donc pas seulement comme un coefficient individuel de dispersion des notes. Sa valeur dépend également de l'accord du panéliste avec le groupe. De plus, la transformation proposée entraîne une modification du classement des produits pour le descripteur en question. Dans l'exemple proposé, le sujet A avait donné le score le plus faible au produit P5. Cependant, après transformation des données, ce produit occupe le rang médian. Par conséquent, dans les situations où les panélistes ne sont pas parfaitement en accord avec le groupe, le traitement effectué a pour conséquence d'aligner les notes des produits sur le classement du groupe. Ceci peut être considéré comme une transformation un peu excessive par rapport à l'objectif de réduction de l'hétérogénéité de la dispersion des notes uniquement.

La discussion sur la pertinence de ce nouveau modèle est donc ouverte. Celui-ci semble légitime lorsque le paramètre $(\hat{\beta}_j - \bar{\beta})\hat{v}_i$ est considéré comme l'expression de l'hétérogénéité des dispersions des notes selon les panélistes. Cependant, ce modèle a pour conséquence d'aligner les notes de chaque sujet sur la moyenne du groupe, et non de réduire uniquement la dispersion des notes entre les panélistes.

Discussion générale - Conclusion

Le travail présenté dans ce manuscrit s'articule autour de la conception de la SensoBase et de son utilisation dans différentes applications. Ce système repose sur un échange entre des données de profils sensoriels et leur analyse. Le site de la Sensobase, fonctionnel depuis avril 2005, permet aux fournisseurs d'envoyer leurs études sensorielles dans une base de données, puis de recevoir des analyses, effectuées à l'aide d'un programme SAS[®]. L'outil de traitement statistique des données rassemble plusieurs méthodes sélectionnées en fonction des connaissances que nous avons à un instant donné. Cependant, chaque type de traitement pourrait être modifié ou amélioré, en fonction des attentes des fournisseurs et des nouveaux développements dans le domaine de la sensométrie. La critique de l'outil d'analyse des données et les évolutions envisageables seront donc abordées dans le premier point de cette discussion.

La création d'une base de données, qui rassemblera bientôt plus de 500 études de profil descriptif, a permis de traiter des questions d'analyse sensorielle sous un angle nouveau. Ainsi, en s'intéressant aux études successives réalisées avec un même panel, nous avons pu suivre l'évolution de la performance de plusieurs panels au cours du temps. La démarche statistique proposée pour l'analyse de ce type de données sera discutée dans un second point. La réflexion sera en particulier menée sur la possibilité d'intégrer de nouveaux temps de mesures à l'analyse. Le troisième point sera consacré à la critique des résultats obtenus grâce à la méta-analyse de l'ensemble de la base et sur les conséquences que pourraient avoir ces résultats sur les choix quotidiens des utilisateurs de l'analyse sensorielle. Enfin, le quatrième point récapitulera les connaissances apportées par l'étude d'une méthode statistique permettant de tenir compte de l'hétérogénéité de la dispersion des notes selon les panélistes (Assessor Model). La pertinence d'une approche "base de données" sur ce type de travail sera ensuite discutée.

1 Evolution de l'outil statistique de traitement des données de profil sensoriel

Le premier challenge de ce travail de thèse était de proposer un outil de traitement qui convienne à l'analyse de tout profil sensoriel contenant des répétitions des mesures. L'enjeu était de couvrir les différents aspects de la performance sans pour autant proposer des analyses redondantes.

L'outil utilisé actuellement correspond à une sélection de méthodes complémentaires, proposant des analyses unidimensionnelles et multidimensionnelles, abordant un ou plusieurs aspects de la performance (répétabilité, discrimination, accord). La difficulté majeure, liée à la réalisation d'un tel outil de traitement, résidait dans l'arbitrage qu'il a fallu effectuer entre la finesse des résultats proposés et la nécessité de délivrer une information synthétique. Dans notre cas, la sélection a été guidée par les habitudes d'analyses du laboratoire. Certaines méthodes additionnelles ont cependant été développées pour compléter certaines approches. La CAPMAP a été conçue pendant ce travail de thèse afin de proposer une cartographie des sujets et de visualiser les corrélations entre descripteurs en fonction du niveau de discrimination. La table de l'accord multidimensionnel fait également partie des outils développés récemment par le laboratoire. Elle permet d'appréhender l'accord entre les panélistes de manière plus complète. Notons toutefois que la sélection proposée ne reflète qu'une vision de la question du traitement des données de profil sensoriel. D'autres techniques auraient évidemment pu être utilisées pour remplacer ou compléter nos choix. Les méthodes statistiques utilisées dans l'outil actuel peuvent donc être appelées à évoluer en fonction de l'expérience de leur utilisation et des commentaires des utilisateurs. Le retour d'information de leur part a déjà permis de modifier certaines analyses. Par exemple, le tableau appelé "analyse multidimensionnelle" a été simplifié afin de ne garder que les indices principaux. L'information sur le niveau de performance multidimensionnel est devenue moins complète mais la compréhension des indices proposés semble s'être améliorée.

L'attention s'est également portée sur la présentation des résultats. Cet aspect ne relève pas d'une question de recherche, mais il est indispensable de prendre cet élément en considération afin de faciliter la lecture des résultats et ainsi inciter les fournisseurs à participer au projet SensoBase. D'après les échanges avec les utilisateurs, la présentation synthétique des résultats est un point particulièrement apprécié. De plus, la présence d'une notice explicative rédigée dans un langage non statistique (cf. annexe 2) a souvent été remarquée et s'est avérée très utile auprès d'un nombre important d'utilisateurs. La vulgarisation des techniques statistiques semble donc être un élément indispensable à la diffusion et à l'utilisation effective des méthodes statistiques.

Afin de proposer un outil toujours plus en adéquation avec les attentes des utilisateurs, tout en proposant des méthodes satisfaisantes d'un point de vue technique, il est indispensable d'entretenir la communication entre les praticiens de l'analyse sensorielle et les sensométriciens. Une première enquête informelle auprès des utilisateurs réguliers a permis de montrer qu'aucun fournisseur n'utilisait toutes les informations disponibles, mais que chacun se focalisait sur une ou deux analyses en particulier. Ce phénomène est probablement lié au fait que la compréhension et l'intégration de chaque méthode par les fournisseurs demande un temps relativement long. En outre, certains aspects ressortent régulièrement dans les commentaires. Le plan bidimensionnel proposé par la CAPMAP est très riche mais il est difficile à interpréter pour une large majorité des utilisateurs. En revanche, les diagnostics individuels, accessibles depuis cette analyse (en cliquant sur les descripteurs ou les sujets), semblent très appréciés, en particulier lors de la

restitution des résultats aux panélistes. Une évolution de l'outil pourrait consister à rendre ces tableaux accessibles directement depuis la table CAP, dont la lecture des résultats est plus aisée. Afin de poursuivre les recherches dans cette direction, la réalisation d'une enquête systématique auprès de tous les fournisseurs pourrait être un point de départ à l'évolution de l'outil proposé. Ce type de démarche est néanmoins relativement lourde et la modification des analyses peut générer des mécontentements de la part des personnes habituées à un type de résultat. De plus, l'outil ne doit pas tenter de satisfaire les désirs de chaque fournisseur mais proposer une structure générale qui permette à chacun d'évaluer la performance de son jury. La multiplication des méthodes disponibles dans l'outil ne serait donc pas nécessairement un bon choix car la variété des résultats risquerait de perturber les utilisateurs.

2 Evolution de la performance

L'étude de l'évolution de la performance a conduit à mettre en place une méthodologie basée sur l'utilisation du modèle mixte. Cette approche permet d'étudier à la fois l'évolution du panel dans son ensemble et l'évolution de chaque panéliste par rapport au groupe. L'analyse est complétée par une représentation graphique basée sur l'utilisation des cartes de contrôle afin de détecter les événements atypiques. Cette méthodologie nous semble bien adaptée au suivi de la performance au cours du temps, mais les caractéristiques de l'indice étudié peuvent avoir un impact sur la pertinence des résultats. En effet, la composition du panel et l'ensemble des produits étudiés peuvent varier au fil des évaluations. Ces modifications sont susceptibles d'influencer plus ou moins le niveau de performance du panel selon l'indice considéré. Par exemple, la discrimination d'un panel peut évoluer au cours du temps seulement à cause de la variabilité de l'espace produit à chaque temps de mesure, et non en fonction d'une réelle évolution de la performance de chaque panéliste. Il est donc important de choisir un indice peu sensible aux variations des conditions de l'expérimentation, ou de tenir compte de ces variations dans l'analyse. Dans ce dernier cas, le problème est généralement d'identifier et d'isoler les sources de variation parasites afin d'étudier uniquement l'évolution du niveau de performance du panel à proprement parler. Il est par exemple difficile de mesurer l'effet de la variabilité des produits à chaque temps de mesure sur l'évolution de la performance. En revanche, certaines covariables, comme le nombre d'échantillons présentés ou le niveau d'expérience de chaque panéliste à chaque temps de mesure peuvent être introduites dans le modèle.

L'indice sélectionné doit enfin satisfaire des contraintes de normalités des résidus. Cette condition n'était pas toujours remplie dans les analyses proposées dans ce manuscrit. Différentes transformations ont alors été envisagées afin de mieux satisfaire cette condition. Les conclusions au niveau du profil d'évolution de la performance étaient cependant peu modifiées. De plus, l'utilisation d'une transformation des données rend la lecture des résultats moins directe.

En marge de ces considérations techniques, les premiers résultats obtenus posent des interrogations relatives à la prise en compte de nouveaux points de mesures. En effet, lorsqu'un

nouveau profil sensoriel est réalisé par un panel, une possibilité consiste à ajuster un nouveau modèle qui tienne compte des données supplémentaires. L'analyse porte alors sur la modification éventuelle de la structure du modèle mixte. Une autre possibilité consiste à considérer le modèle établi à l'aide des premiers temps de mesure comme un outil de contrôle de l'évolution des futures études sensorielles. L'analyse se porte alors sur la position des nouveaux points de mesure par rapport à l'évolution prévue. De notre point de vue, les premières analyses obtenues sur quelques séries d'études avaient principalement pour objectif de caractériser et de décrire l'évolution d'un panel au cours du temps. L'adjonction de nouveaux temps de mesures seraient donc plutôt utile à la validation de la modélisation initiale. Dans ce cas, le modèle serait ajusté à nouveau en tenant compte des données supplémentaires, puis les nouveaux résultats seraient comparés aux anciens afin de détecter d'éventuelles modifications. En revanche, lorsque les résultats de plusieurs séries d'études auront été analysés et que les principaux mécanismes d'évolution de la performance auront été appréhendés, il pourrait être envisageable d'utiliser cette méthodologie dans une démarche de contrôle de la performance.

L'objectif initial était de créer un outil de décision permettant de détecter à quel moment un panel doit être ré-entraîné, ou à quel moment un panéliste doit être exclu du groupe, si nécessaire. Nous disposons aujourd'hui d'une méthodologie qui permet de décrire l'évolution de la performance et de détecter les événements atypiques. La décision finale, concernant l'éviction d'un panéliste ou son réentraînement, reste cependant de la responsabilité de l'animateur de panel.

Notons enfin qu'au commencement de ce travail, la caractérisation de l'évolution de la performance était la question essentielle posée par les centres ACTIA. Une méthodologie ad hoc a été développée parallèlement à la mise en place de la sensobase. Il serait maintenant envisageable d'intégrer cette méthodologie du modèle mixte longitudinal aux traitements de la SensoBase. L'ajout de chaque nouvelle étude pourrait entraîner une mise à jour automatique des résultats. Cette opération demanderait toutefois quelques modifications de la structure de la SensoBase afin de définir explicitement les liens chronologiques entre les études.

3 Méta-analyse des niveaux de performance

L'analyse simultanée d'un grand nombre d'études descriptives a permis de faire progresser la connaissance sur les niveaux de performance généralement atteints en analyse sensorielle et sur quelques facteurs d'explication de la variation de ces niveaux.

Les résultats montrent que la répétabilité est relativement homogène au sein de chaque étude. En revanche, le niveau de répétabilité varie de manière significative entre les études sensorielles. Les causes des différences observées pourraient être liées aux conditions de réalisation de chaque profil. Par exemple, les différences entre produits peuvent être plus évidentes dans certaines études. Les panélistes sont alors plus catégoriques sur leurs jugements, donc plus répétables. Concernant l'analyse du niveau de discrimination des panélistes, les résultats montrent qu'en moyenne, chaque juge ne discrimine les produits que pour un tiers des

descripteurs. Cette valeur relativement faible amène à se demander s'il est nécessaire d'utiliser autant de descripteurs (26 en moyenne) dans chaque étude alors que les panélistes ne sont capables de faire des différences entre les produits que pour un nombre réduit d'entre eux. Cette conclusion remet également en cause la capacité des panélistes à discriminer des produits sur un grand nombre d'attributs sensoriels différents. Le niveau de discrimination semble en revanche relativement homogène entre les études. Le constat est très similaire concernant le niveau d'accord entre les panélistes. En effet, bien que la tendance soit positive dans la majorité des cas, l'accord de chaque juge avec le groupe est significatif dans un tiers des situations seulement.

Les résultats de l'analyse des variables intra-étude montrent un fort effet du type de descripteur. Les performances sont globalement meilleures pour les descripteurs visuels et plus mitigées pour les descripteurs de texture. En outre, le niveau de discrimination est sensiblement plus élevé pour les personnes âgées de 30 à 45 ans, ayant un niveau scolaire supérieur au baccalauréat et/ou une expérience en analyse sensorielle relativement longue. De plus, le niveau d'accord est significativement meilleur pour les panélistes les plus expérimentés. Le sexe n'influence pas la performance. Notons cependant que les conclusions concernant l'effet de la durée de l'entraînement et du niveau d'études reposent sur un nombre de données relativement faible. Il est donc important d'encourager les utilisateurs de la SensoBase à donner un maximum d'information pour permettre d'étudier l'effet de différents facteurs avec un niveau de confiance satisfaisant. Il serait également intéressant de mesurer les interactions entre les différentes variables, mais la nature des données de la SensoBase ne le permet pas. En effet, nous disposons rarement des informations sur toutes les variables dans une étude. Les interactions sont donc très déséquilibrées, voire incomplètes. La qualité et l'exhaustivité des informations données par les fournisseurs est donc primordiale pour réaliser des analyses robustes et puissantes. Afin d'obtenir, de manière plus systématique, des données plus complètes de la part des fournisseurs, la liste des informations optionnelles a été réduite dans le fichier Excel[®] SensoBase. En contrepartie, les fournisseurs sont sensibilisés aux types d'études que nous souhaitons réaliser et encouragés à fournir les données correspondant à une courte liste d'items (dont les variables étudiées dans ce manuscrit).

Concernant les facteurs inter-études, la principale source de variation du niveau de la performance est le fournisseur. Cela indique que la mise en place et la réalisation d'un profil sensoriel peuvent être très différentes d'un établissement à l'autre. Ce résultat peut traduire le fait que les objectifs de chaque profil sont différents. Il est par exemple normal que les panélistes ne soient pas discriminants ou en accord sur certains descripteurs lors d'une démarche exploratoire, visant à conserver les descripteurs les plus discriminants. En revanche, une analyse menée sur des descripteurs éprouvés, avec un panel habitué à réaliser ce type d'analyse, devrait en général donner de meilleurs résultats. De plus, l'animateur de panel joue probablement un rôle déterminant sur la qualité de l'analyse, tant par la façon dont il organise les épreuves d'entraînement (nombre de séances, tâches effectuées...) que par la nature des consignes qu'il donne aux panélistes. Ce type d'information n'est cependant pas disponible dans les données recueillies par la SensoBase. Les analyses montrent que l'effet fournisseur peut en partie

s'expliquer par le pays d'origine. Le niveau de répétabilité des laboratoires français est significativement inférieur à celui des autres pays (majoritairement de l'Europe du nord). Cependant, étant donné que cet effet n'est pas vérifié avec l'indice de discrimination, la différence observée s'interprète davantage comme une différence d'utilisation de l'échelle. Les panélistes français ont donc tendance à utiliser une portion de l'échelle de notation plus importante. Les autres facteurs inter-études ne sont pas significatifs (échelle continue ou discrète, laboratoire privé ou public et type de produit). Ce résultat peut en partie être lié au fait que le nombre de fournisseurs participant au programme SensoBase est encore relativement faible (35 fournisseurs). Il est donc difficile de mettre à jour certains effets dans ces conditions. Par exemple, afin de donner des résultats robustes quant aux différences de niveaux de performances selon le type de produit, il serait préférable de disposer de profils provenant de 30 fournisseurs différents pour chaque catégorie. Afin de réaliser ce type de calculs, l'objectif pour les années à venir est donc en priorité d'augmenter le nombre de fournisseurs participant au programme SensoBase plutôt que le nombre global d'études sensorielles.

Par rapport aux ambitions initiales, qui étaient, entre autres choses, de définir des niveaux de performance de référence pour chaque catégorie de chaque variable, il s'est avéré que la variabilité entre les études sensorielles dépendait d'un nombre trop important de facteurs non maîtrisés, ce qui ne permet pas vraiment de donner une estimation fiable. En revanche, la comparaison des niveaux d'un facteur à l'intérieur d'une étude est envisageable puisque les conditions d'entraînement sont les mêmes pour tous les panélistes. La SensoBase nous semble donc aujourd'hui plus apte à détecter des effets liés à des variables intra-étude qu'inter-études.

Parmi les utilisateurs de la SensoBase, nombreux sont ceux qui désirent se positionner en terme de performance par rapport à l'ensemble des données recueillies. Le plus souvent, ils souhaiteraient également comparer leurs performances à celles obtenues dans les études précédentes, quel que soit le panel ou l'espace produit considéré. La structure de la SensoBase permet de répondre à ce type d'attente. Cela nécessiterait simplement de développer une procédure capable de rassembler les résultats précédemment obtenus et de renvoyer, avec l'analyse de chaque étude, le positionnement des performances par rapport à l'ensemble des données de la base, ou à l'ensemble des données du fournisseur en question. Ce type de résultat pourrait être fourni pour chaque indice de performance de la sélection de méthodes.

Enfin, notons que les analyses présentées dans le chapitre 5 ne portent que sur trois indices de performance unidimensionnels. Ces derniers avaient été sélectionnés afin d'explorer les résultats sur les trois facettes de la notion de performance. Il serait maintenant intéressant de poursuivre ce genre d'investigation avec des indices de performance multidimensionnels ou des indices unidimensionnels différents afin d'observer si les conclusions évoluent. Nous pourrions par exemple utiliser les tests proposés par l'Assessor Model de Brockhoff.

4 Assessor Model

L'Assessor Model propose une approche différente de l'étude de la performance en analyse sensorielle. L'utilisation de cette méthode sur un grand nombre de profils sensoriels a permis de vérifier que les hypothèses utilisées dans ce modèle correspondaient en général à la réalité des données recueillies. Les résultats montrent que l'hypothèse d'homoscédasticité, stipulée dans le modèle d'ANOVA classique (modèle à deux facteurs avec interaction), était rarement vérifiée et que la dispersion des notes étaient souvent différentes entre les panélistes. L'Assessor Model, dans lequel une variance résiduelle par sujet est calculée et qui fait intervenir explicitement un terme d'hétérogénéité des dispersions, serait donc mieux adapté à l'étude de la performance des panélistes dans la majorité des situations.

Ce modèle est accompagné de quatre tests sur les niveaux de performance des panélistes. La variance résiduelle de chaque sujet permet de tester l'hétérogénéité de la répétabilité des panélistes. Le coefficient de dispersion des notes permet quant à lui de tester l'hétérogénéité des dispersions entre les sujets. De plus, l'Assessor Model propose un test du désaccord qui, au contraire du modèle classique, ne prend en compte que le désaccord "pur" et non l'hétérogénéité des dispersions des notes selon les panélistes. Enfin, ce modèle propose un test de l'hétérogénéité des sensibilités (i.e. du pouvoir discriminant). L'Assessor Model permet donc de disposer de quatre tests pour caractériser l'hétérogénéité de la notation et de la performance des panélistes, qui ne sont pas disponibles dans l'approche classique. Dans les mois à venir, l'objectif est donc d'implémenter cette méthode dans l'outil de traitement automatisé proposé par la SensoBase pour compléter l'analyse de la performance proposée.

En marge des calculs des indices de performance, l'étude s'est portée sur la comparaison des résultats obtenus entre le test du désaccord de l'Assessor Model et le test de l'interaction produit*sujet, généralement considérée comme l'expression du désaccord entre les sujets dans le modèle classique. Les résultats montrent que le modèle classique détecte à tort un désaccord entre les panélistes dans 22,5% des cas. Nous avons ensuite cherché à déterminer si ce résultat influence les conclusions concernant l'effet produit. Lorsque les données sont simplement standardisées, afin de réduire l'hétérogénéité des dispersions des notes, les conclusions concernant l'effet produit dans le modèle classique étaient peu modifiées. Ce résultat tendrait plutôt à indiquer qu'il n'est donc pas nécessaire d'utiliser l'Assessor Model lors de l'étude des différences entre produits puisque la modification au niveau de l'effet produit est faible. Cependant, Brockhoff propose une transformation des données, en accord avec son modèle, qui permet de supprimer l'hétérogénéité des dispersions entre sujets. Les résultats montrent alors que la significativité de l'effet produit passe de 59% des descripteurs avec le modèle classique à 68% avec la transformation. La prise en compte de l'hétérogénéité des dispersions permet donc de mettre à jour des différences entre produits pour une proportion non négligeable de descripteurs.

Les investigations sur l'Assessor Model devraient cependant être poursuivies afin de s'assurer que la transformation proposée ne retire effectivement que la part de variabilité due à

l'hétérogénéité des dispersions. En effet, cette transformation, ainsi que le modèle de Brockhoff de manière générale, repose sur l'hypothèse que l'hétérogénéité des dispersions entre panélistes correspond à l'hétérogénéité des pentes, ajustées entre les notes de chaque sujet et les notes de la moyenne du groupe. Cette équivalence nous semble acceptable lorsque les panélistes sont en accord mais elle devient discutable lorsque les panélistes sont en désaccord (figure 6-7, chapitre 6). La transformation proposée ne devrait donc pas être utilisée systématiquement, mais uniquement dans les situations dans lesquelles le test du désaccord est non significatif.

Le type d'investigation réalisée sur l'Assessor Model à l'aide des données de la SensoBase pourrait être conduit avec d'autres méthodes. Cela permettrait d'évaluer si les différences théoriques entre plusieurs modèles entraînent souvent une modification des conclusions dans la pratique. L'analyse en composante principale pourrait par exemple être comparée à l'analyse en variable canonique par le biais du calcul du coefficient RV entre les deux configurations obtenues. La SensoBase offre ici une nouvelle possibilité pour comparer des méthodes.

5 Conclusion générale

Les travaux présentés dans ce manuscrit ont pour dénominateur commun l'évaluation de la performance à l'aide des données recueillies par la SensoBase. Cette approche "base de données", déjà développée dans de nombreux autres domaines, s'étend donc désormais à l'analyse sensorielle. L'intérêt majeur d'une telle démarche est le gain de puissance conféré par le grand nombre de données prises en compte. En contrepartie, la réunion de vastes ensembles de données nécessite la définition et l'organisation d'une structure commune à chaque élément (i.e. chaque étude sensorielle dans notre cas). Jusqu'à présent, la SensoBase a permis d'étudier l'évolution de la performance d'un panel au cours du temps, de mesurer l'effet de plusieurs variables (sexe, âge, type de produit, type de descripteur, ...) sur les niveaux de performances et de comparer deux méthodes statistiques (Assessor Model et approche classique) en se basant sur les résultats obtenus sur près de 300 études sensorielles. Le vaste ensemble de données recueillies pourrait également permettre de répondre à d'autres questions. Par exemple, le problème du nombre optimal de panélistes que doit contenir une étude afin d'obtenir des performances satisfaisantes pourrait être abordé grâce à la SensoBase. A l'aide d'une méthode de ré-échantillonnage (bootstrap, par exemple), appliquée aux sujets de chaque panel, les niveaux de performance pourraient être évalués pour différents effectifs. Les résultats obtenus seraient alors représentatifs d'un nombre important de situations réelles. L'étude du nombre de panélistes pourrait également se baser sur la variabilité naturelle des effectifs observés dans les profils de la SensoBase. Cette méthode éviterait alors l'utilisation d'une méthode de ré-échantillonnage et permettrait de rester proche des conditions d'acquisition des données. Ce type de démarche pourrait aussi s'appliquer à l'étude de l'effet du nombre de produits et du nombre de descripteurs sur les niveaux de performance. La SensoBase pourrait en outre être utilisée pour étudier des

questions moins liées à la notion de performance, comme l'importance des effets d'ordre et de report, l'influence de l'effet session ou la comparaison de méthodes multidimensionnelles.

Enfin, signalons que l'intérêt de la SensoBase n'est pas uniquement de disposer d'un nombre important d'études sensorielles. Cette structure pourrait également constituer une plateforme d'échange privilégiée entre les praticiens de l'analyse sensorielle et les sensométriciens. Cet aspect du projet doit à notre avis être encouragé afin d'améliorer l'adéquation entre les techniques statistiques et les besoins de l'analyse sensorielle.

Références bibliographiques

- A**FNOR, Ed. (1993). Guide général pour la sélection, l'entraînement et le contrôle des sujets. Partie 1: Sujets qualifiés. Analyse Sensorielle. Paris - La Défense, AFNOR.
- AFNOR, Ed.** (1999). Guide général pour l'évaluation sensorielle. Description, différenciation et mesure hédonique. Analyse Sensorielle. Paris - La Défense, AFNOR.
- Amerine, M. A., Pangborn, R. M. and Roessler, E. B., Eds.** (1965). Principles of sensory evaluation of food. Food Science and Technology. A Series of Monographs. New York, Academic Press.
- Aranceta, J., Perez-Rodrigo, C., Ribas, L. and Serra-Majem, L. (2003).** "Sociodemographic and lifestyle determinants of food patterns in Spanish children and adolescents: the enKid study." *European Journal of Clinical Nutrition* **57**(Suppl. 1): S40-S44.
- Arnold, G. M. and Williams, A. A. (1986).** The use of generalised procustes techniques in sensory analysis. Statistical procedures in food research. J. R. Piggott. Barking, Essex, Elsevier Applied Science: 233-253.
- ASTM (1981).** Guidelines for the selection and training of sensory panel members, ASTM.
- Aubry, V., Schlich, P., Issanchou, S. and Etiévant, P. (1999).** "Comparison of wine discrimination with orthonasal and retronasal profilings. Application to Burgundy Pinot Noir wines." *Food Quality and Preference* **10**(4/5): 253-259.
- B**äckström, A., Pirttilä-Backman, A.-M. and Tuorila, H. (2003). "Dimensions of novelty: a social representation approach to new foods." *Appetite* **40**(3): 299-307.
- Bartlett, M. S. (1937).** "Properties of sufficiency and statistical test." *Proceedings of the Royal Society London Series A*(160): 268-282.
- Bi, J. (2003).** "Agreement and reliability assessments for performance of sensory descriptive panel." *Journal of Sensory Studies* **18**(1): 61-76.
- Bitnes, J., Martens, H., Ueland, Ø. and Martens, M. (2005).** "Longitudinal study of taste identification of sensory panellists: Effect of Ageing, Experience and Exposure." *Food Quality and Preference In Press, Corrected Proof*.
- Brockhoff, P. (2003).** "Statistical testing of individual differences in sensory profiling." *Food Quality and Preference. Elsevier. Volume 14*: 425-434.
- Brockhoff, P. and Skovgaard, M. (1994).** "Modelling individual differences between assessors in sensory evaluations." *Food Quality and Preference Volume 5*: 215-224.
- Brockhoff, P. M. (1994).** Statistical analysis of sensory data, Royal veterinary and agricultural university Copenhagen.
- Brockhoff, P. M. (1998).** "Assessor modelling." *Food Quality and Preference* **9**(3): 87-9.
- C**arlucci, A. and Monteleone, E. (2001). "Statistical validation of sensory data: a study on wine." *Journal of the Science of Food and Agriculture* **81**(8): 751-758.
- Chabanet, C. and Pineau, N. (2005).** Using linear mixed models to handle variability of consumer's liking. 7th Pangborn Sensory Science Symposium. Harrogate, United-Kingdom.
- Chollet, S. and Valentin, D. (2001).** "Impact of training on beer flavor perception and description: are trained and untrained subjects really different?" *Journal of Sensory Studies* **16**(6): 601-618.
- Choudhury, E. S., Moberg, P. and Doty, R. L. (2003).** "Influences of age and sex on a microencapsulated odor memory test." *Chem. Senses* **28**(9): 799-805.
- Clapperton, J. F. and Piggott, J. R. (1979).** "Flavour characterization by trained and untrained assessors." *J. Inst. Brew.* **85**(September-October): 275-277.
- Collett, D. (1996).** Modeling binary data. Londres (G.B), T.J. Press.
- Couronne, T. (1997).** "A study of assessors' performance using graphical methods." *Food Quality and Preference* **8**(5/6): 359-365.

Couronne, T. (2001). Etude de la performance des sujets. Traité d'évaluation sensorielle. Aspect cognitifs et métrologiques des perceptions. I. Urdapilletta. Paris, Dunod: 393-415.

Dijksterhuis, G. (1995). "Assessing panel consonance." *Food Quality and Preference* **6**(1): 7-14.

Dijksterhuis, G. (1998). "European dimensions of coffee : rapid inspection of data set using Q-PCA." *Food Quality and Preference* **9**(3): 95-98.

Dixon, R. A., Wahlin, A., Maitland, S. B., Hultsch, D. F., Hertzog, C. and Bäckman, L. (2004). "Episodic memory change in late adulthood: Generalizability across samples and performance indices." *Memory & Cognition* **32**(5): 768-778.

Escoufier, B. and Pagès, J. (1984). L'étude factorielle multiple. Paris, Université Pierre et Marie Curie.

Escoufier, B. and Pagès, J. (1985). "Mise en oeuvre de l'analyse factorielle multiple pour des tableaux numériques qualitatifs ou mixtes." *Rapport ISIRA* **263**.

Escoufier, B. and Pagès, J., Eds. (1988). Analyses factorielles simples et multiples : objectifs, méthodes et interprétation. Paris, Dunod.

Escoufier, Y. (1973). "Le traitement des variables vectorielles." *Biometrics* **29**: 751-760.

ESN (1996). A European sensory and consumer study. A case study on coffee. Chipping Campden, Gloucestershire, GL55 6LD UK, European Sensory Network. Available from Campden & Chorleywood Food Research Association.

Everitt, B. S. and Dunn, G. (1991). Applied Multivariate Data Analysis. London.

Finn, R. H. (1970). "A note on estimating the reliability of categorical data." *Educ. Psychological Meas.* **30**: 71-76.

Fisher, R. A. (1946). Statistical methods for research workers. New-York.

Gabriel, K. R. (1971). "The biplot graphic display of matrices with application to principal component analysis." *Biometrika* **58**(3): 453-467.

Glass, G. V. (1976). "Primary, secondary and meta-analysis of research." *Educational researcher* **5**: 3-8.

Gou, P., Guerrero, L. and Romero, A. (1995). Cartographie externe des préférences à partir des cartes sensorielles générées par groupes avec différentes préparations en utilisant un petit nombre d'échantillons. 4ièmes Journées Agro-Industries et Méthodes Statistiques. A. G. Agro-Industrie. Dijon, 7-8 December 1995: 171-183.

Gower, J. C. (1975). "Generalized procrustes analysis." *Psychometrika* **40**(1): 33-50.

Hanafi (2003). "Analyse des composantes communes et poids spécifiques (ACCPs), Définition, propriétés de la méthode et principes de mise en oeuvre sous Matlab." *Version provisoire*.

Harris, J. T., Seppala, C. T. and Desborough, L. D. (1999). "A review of performance monitoring and assessment techniques for univariate and multivariate control systems." *Journal of process control* **9**: 1-17.

Hedges, L. V., Bushman, B. J. and Cooper, H. (1992). "Testing the null hypothesis in meta-analysis: a comparison of combined probability and confidence interval procedure." *Psychol. Bull.* **111**: 188-194.

Hummel, S. T., Futschik, T., Frasnelli, J. and Huttenbrink, K. B. (2003). "Effects of olfactory function, age, and gender on trigeminally mediated sensations: a study based on the lateralization of chemosensory stimuli." *Toxicology Letters* **140-141**: 273-280.

Hunter, E. A. and Muir, D. D. (1995). "A comparison of two multivariate methods for the analysis of sensory profile data." *Journal of Sensory Studies* **10**: 89-104.

Husson, F., Bocquet, V. and Pagès, J. (2004). "Use of confidence ellipses in a PCA applied to sensory analysis application to the comparison of monovarietal ciders." *Journal of Sensory Studies* **19**(6): 510-518.

Husson, F., Le, S. and Pagès, J. (2005). "Confidence ellipse for the sensory profiles obtained by principal component analysis." *Food Quality and Preference* **16**(3): 245-250.

ISO (1993). Sensory Analysis - General guidance for the selection, training and monitoring of assessors - Part 1 : Selected assessors. 8586-1. Paris- La Défense, AFNOR.

James, L. R., Demaree, R. G. and Wolf, G. (1984). "Estimating within-group interrater reliability with and without response bias." *Journal of applied Psychology* **69**(85-98).

Joliffe, I. T. (1986). Principal component analysis. New York, Springer-Verlag.

Karpinski, K. F. (1989). "Method performance." *J.Assoc.Off.Anal.Chem.* **72**(6): 931-935.

Kazi-Aoual, F. (1993). Approximations to permutation tests for data analysis, E.N.S.A. Montpellier, Unité de Biométrie.

King, M. C., Hall, J. and Cliff, M. A. (2001). "A comparison of methods for evaluating the performance of a trained sensory panel." *Journal of Sensory Studies* **16**(6): 567-581.

Kunert, J. and Qannari, E. M. (1999). "A simple alternative to generalized procrustes analysis : application to sensory profiling data." *Journal of Sensory Studies* **14**(2): 197-208.

L'Hermier des Plantes, H. (1976). Structuration des tableaux à trois indices de la statistique. Thèse de 3ème cycle. Montpellier, Université de Montpellier.

Labbe, D., Rytz, A. and Hugi, A. (2004). "Training is a critical step to obtain reliable product profiles in a real food industry context." *Food Quality and Preference* **15**(4): 341-348.

Larsson, M., Lovden, M. and Nilsson, L.-G. (2003). "Sex differences in recollective experience for olfactory and verbal information." *Acta Psychologica* **112**(1): 89-103.

Latreille, J., Mauger, E., Ambroisine, L., Tenenhaus, M., Vincent, M., Navarro, S. and Guinot, C. (2006). "Measurement of the reliability of sensory panel performances." *Food Quality and Preference* **17**(5): 369-375.

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

Lawless, H. and Heymann (1998). The trigeminal flavor senses.

Le Dien, S., Husson, F. and Pagès, J. (2004). Construction de région de confiance en AFMH: application à l'analyse sensorielle. 8èmes Journées Européennes Agro-Industrie et Méthodes Statistiques, Rennes.

Le Dien, S. and Pagès, J. (2003). "Hierarchical multiple factor analysis: application to the comparison of sensory profiles." *Food Quality and Preference* **14**(5-6): 397-403.

Le Dien, S., Pagès, J. and Husson, F. (2004). Confidence ellipse in HMFA applied to sensory profiles of chocolates. 7th Sensometrics symposium, Davis, California USA.

Le Quéré, J. M., Husson, F., Renard, C. and Primault, J. (2006). "French cider characterization by sensory, technological and chemical evaluations." *LWT* **39**(9): 1033-1044.

Lea, P., Rødbotten, M. and Næs, T. (1995). "Measuring validity in sensory analysis." *Food Quality and Preference* **6**(4): 321-326.

Ledauphin, S., Hanafi, M. and Qannari, E. M. (2006). "Assessment of the agreement among the subjects in fixed vocabulary profiling." *Food Quality and Preference* **17**(3-4): 277-280.

Lesschaeve, I. (1997). Etude des performances de sujets effectuant l'analyse descriptive quantitative de l'odeur ou de l'arôme de produits alimentaires. Recherche de liens entre épreuves de sélection et épreuves de profil, Université de Bourgogne.

Li, W. K., Monroe, K. B. and Chan, D. K. S. (1994). "The effects of country, of origin, brand, and price information: a cognitive-effective model of buying intentions." *Advances in Consumer Research* **21**: 449-457.

Lundahl, D. S. and Daniel, M. R. M. (1990). "Use of contrasts for the evaluation of panel inconsistency." *J. Sensory Studies* **5**: 265-277.

Lundahl, D. S. and McDaniel, M. R. (1988). "The panelist effect. Fixed or random?" *Journal of Sensory Studies* **3**: 113-121.

Magnusson, M. K. and Koivisto Hursti, U.-K. (2002). "Consumer attitudes towards genetically modified foods." *Appetite* **39**(1): 9-24.

Mandel, J. (1991). "The variation of measurement through inter laboratory studies." *Chemometrics and intelligent laboratory systems* **11**: 109-119.

- Marchant, L. F. and McGrew, W. C. (1991).** "Laterality of function in apes: a meta-analysis of methods." *Journal of Human Evolution* **21**(6): 425-438.
- Martin, N., Molimard, P., Spinnler, H. E. and Schlich, P. (2000).** "Comparison of odour sensory profiles performed by two independent trained panels following the same descriptive analysis procedures." *Food Quality and Preference* **11**(6): 487-495.
- McEwan, J. A., Hunter, E. A., van Gemert, L. J. and Lea, P. (2002).** "Proficiency testing for sensory profile panels: measuring panel performance." *Food Quality and Preference* **13**(3): 181-190.
- Monrozier, R. and Danzart, M. (2001).** "A quality measurement for sensory profile analysis. The contribution of extended cross-validation and resampling techniques." *Food Quality and Preference* **12**(5-7): 393-406.
- Montgomery, D. C. and Mastrangelo, C. M. (1991).** "Some statistical process control methods for autocorrelated data." *Journal of quality technology* **23**(3): 179-191.
- Morand, E. and Pagès, J. (2006).** "Procrustes multiple factor analysis to analyse the overall perception of food products." *Food Quality and Preference* **17**(1-2): 36-42.
- Mortell, R. R. and Runger, G. C. (1995).** "Statistical process control of multiple stream processes." *Journal of quality technology* **27**(1): 1-12.
- Moskowitz, H. (2000).** "On the reliability of expert panel data - differences due to attribute type." *Food Quality and Preference* **11**(1/2): 145-146.
- Moskowitz, H. R. (1996).** "Experts versus consumers: A comparison." *Journal of Sensory Studies* **11**(1): 19-37.
- Moskowitz, H. R. and Jacobs, B. E. (1988).** Simultaneous optimization of products and concepts for foods. Applied sensory analysis of foods. H. Moskowitz. Boca Raton, CRC press. II: 141-174.
- Moskowitz, H. R., Jacobs, B. E. and Lazar, N. (1985).** "Product response segmentation and the analysis of individual differences in liking." *Journal of Food Quality* **8**(2/3): 169-181.
- Mucci, A., Hough, G. and Ziliani, C. (2004).** "Factors that influence purchase intent and perceptions of genetically modified foods among Argentine consumers." *Food Quality and Preference* **15**(6): 559-567.
- Mullen, B. and Suls, J. (1982).** "The effectiveness of attention and rejection as coping styles: A meta-analysis of temporal differences." *Journal of Psychosomatic Research* **26**(1): 43-49.
- Muller, N., Heymann, H. and Sadie, A. (2003).** A sensory study of South African food products: The selection and training of panelists with different literacy levels. 5th Pangborn symposium, Boston, Massachusetts.
- Murphy, C. (1985).** "Cognitive and chemosensory influences on age-related changes in the ability to identify blended foods." *Journal of Gerontology* **40**(1): 47-52.

- N****æs, T. (1990).** "Handling individual differences between assessors in sensory profiling." *Food Quality and Preference* **2**(3): 187-199.
- Næs, T. (1998).** "Detecting individual differences among assessors and differences among replicates in sensory profiling." *Food Quality and Preference* **9**(3): 107-110.
- Næs, T., Hirst, D. and Baardseth, P. (1994).** "Using cumulative ranks to detect individual differences in sensory profiling." *Journal of Sensory Studies* **9**(1): 87-99.
- Næs, T. and Langsrud, V. (1998).** "Fixed or random assessors in sensory profiling?" *Food Quality and Preference* **9**(3): 145-152.
- Næs, T. and Solheim, R. (1991).** "Detection and interpretation of variation within and between assessors in sensory profiling." *Journal of Sensory Studies* **6**(3): 159-177.

- P****agès, J. (1996).** "Eléments de comparaison entre l'analyse factorielle multiple et la méthode STATIS." *Revue de Statistique Appliquée* **XLIV**(4): 81-95.
- Pagès, J., Asselin, C., Morlat, R. and Robinet, J. (1987).** "L'analyse factorielle multiple dans le traitement des données sensorielles. Application à des vins rouges de la Vallée de la Loire." *Sciences des Aliments* **7**(4): 549-571.
- Pagès, J. and Husson, F. (2001).** "Inter-laboratory comparison of sensory profiles: methodology and result." *Food Quality and Preference* **12**(5-7): 297-309.
- Pangborn, R. M. (1981).** Individuality in responses to sensory stimuli. Criteria of food acceptance. J. Solms and R. L. Hall. Zürich, Forster-Verlag AG: 177-219.
- Peterson, R. A. and Jolibert, A. J. P. (1995).** "A meta-analysis of country-of-origin effects." *Journal of International Business Studies* **26**: 883-900.

- Piggott, J. R. and Hunter, E. A. (1999).** "Evaluation of assessor performance in sensory analysis." *Italian Journal of Food Science* **11**(4): 289-304.
- Pineau, N., Chabanet, C. and Schlich, P. (2006).** "Evolution of the repeatability of a panel in sensory analysis: a mixed model and control chart approach." *Journal of Sensory studies* **Accepted, sous presse.**
- Pinheiro, J. C. and Bates, D. M. (2000).** Mixed-effects models in S and S-PLUS. New-York, Springer-Verlag.
- Polit, D. F. (1996).** Data Analysis and Statistics for Nursing Research. Stamford, Connecticut.
- Porcherot, C., Issanchou, S. and Schlich, P. (1998).** Influence of cooking modes on potato sensory properties. Fourth Sensometrics meeting. KVL. Copenhagen, 6-8 August 1998.
- Porcherot, C. and Schlich, P. (2000).** "FLASH table and canonical mapping of potato varieties." *Food Quality and Preference* **11**(1/2): 163-165.
- Pritchett-Mangan, P. A. (1992).** "Performance assessment of sensory panelists." *Journal of Sensory Studies* **7**: 229-252.
- Prutkin, J., Fisher, E.M., Etter, L., Fast, K., Gardner, E., Lucchina, L.A., Snyder, D.J., Tie, K., Weiffenbach, J., Bartoshuk, .LM. (2000).** "Genetic variation and inferences about perceived taste intensity in mice and men". *Physiol Behav.* Apr 1-15;69(1-2):161-73.

- Qannari, E. M., Courcoux, P., Lejeune, M. and Maystre, O. (1995).** Comparaison de trois stratégies de détermination d'un compromis en évaluation sensorielle. 4ièmes Journées Agro-Industries et Méthodes Statistiques. A. G. Agro-Industrie. Dijon, 7-8 December 1995: 59-66.
- Qannari, E. M., Courcoux, P., Lejeune, M. and Maystre, O. (1997).** "Comparaison de trois stratégies de détermination d'un compromis en évaluation sensorielle." *Revue de Statistiques Appliquées XLV*(1): 61-74.
- Qannari, E. M., MacFie, H. J. H. and Courcoux, P. (1999).** "Performance indices and isotropic scaling factors in sensory profiling." *Food Quality and Preference* **10**(1): 17-21.
- Qannari, E. M. and Meyners, M. (2001).** "Identifying assessor differences in weighting the underlying sensory dimensions." *Journal of Sensory Studies* **16**(5): 505-515.
- Qannari, E. M., Wakeling, I., Courcoux, P. and Macfie, H. J. H. (2000).** "Defining the underlying sensory dimensions." *Food Quality and Preference* **11**(1/2): 151-154.

- Risvik, E., Colwill, J. S., McEwan, J. A. and Lyon, D. H. (1992).** "Multivariate analysis of conventional profiling data: a comparison of a British and a Norwegian trained panel." *Journal of Sensory Studies* **7**(2): 97-118.
- Roberts, A. K. and Vickers, Z. M. (1994).** "A comparison of trained and untrained judges' evaluation of sensory attributes intensities and liking of Cheddar cheeses." *Journal of Sensory Studies* **9**(1): 1-20.
- Rossi, F. (2001).** "Assessing sensory panelist performance using repeatability and reproducibility measures." *Food Quality and Preference* **12**(5-7): 467-479.
- Rousset-Akrim, S., Martin, J. F., Bayle, M. C. and Touraille, C. (1996).** Intercountry study on consumers preferences on dry-cured hams. Cross-cultural studies on food preferences and consumer behaviour. A.-C. M. o. c. a. a. t. i. o. f. c. a. acceptability. Ås, NORWAY, AIR-CAT project, c/o MATFORSK: 40-44.

- S.A.S. (2003).** SAS/STAT user's guide. (Version 9.1.3), MIXED procedure. Cary, NC, SAS Institute.
- Saporta, G. (1990).** Probabilité, analyse des données statistiques.
- Sauvant, D., Schmidely, P. and Daudin, J. J. (2005).** "Les méta-analyses des données expérimentales : applications en nutrition animale." *INRA production alimentaire* **18**(1): 63-73.
- Sawyer, F. M., Cardello, A. V. and Prell, P. A. (1988).** "Consumer evaluation of the sensory properties of fish." *Journal of food science* **53**(1): 12-24.
- Schlich, P. (1993).** Contributions à la sensométrie, Paris XI Orsay.
- Schlich, P. (1994).** "Grapes: a method and a SAS program for graphical representations of assessor performances." *Journal of Sensory Studies* **9**(2): 157-169.

- Schlich, P. (1996).** Defining and validating assessor compromises about product distances and attribute correlations. *Multivariate Analysis of Data in Sensory Science*. T. Naes and E. Risvik. Amsterdam, Elsevier Science: 259-306.
- Schlich, P. (1997).** CAP : une méthode et un outil de contrôle rapide et synthétique des performances des sujets en évaluation sensorielle descriptive. 5èmes Journées Agro-industries et Méthodes Statistiques. A. G. Agro-Industrie. Versailles, 3-5 december 1997: 8.1-8.10.
- Schlich, P. (1998).** "What are the sensory differences among coffees? Multi-panel analysis of variance and FLASH analysis." *Food Quality and Preference* **9**(3): 103-106.
- Schlich, P. (2000).** CAP: Control of Assessor Performances. Fifth Sensometrics, University of Missouri, Columbia.
- Schlich, P. (2004).** L'analyse en variables canoniques des données de profils sensoriels. Applications pour le contrôle des performances des sujets et pour la comparaison des produits. ème journées européennes agro-industrie et méthodes statistiques, Rennes.
- Schlich, P. (2004).** On the multivariate analysis of sensory profiling data. 7th Sensometrics, Davis, California.
- Schlich, P. and Degoud, S. (1995).** Flash analysis of profiling data. 4^{èmes} Journées Agro-Industries et Méthodes Statistiques, Dijon, ASU Groupe Agro-Industrie.
- Schlich, P., Pineau, N., Brajon, D. and Qannari, E. M. (2004).** Multivariate control of assessor performances. 7th Sensometrics, UC Davis, California.
- Sinesio, F., Moneta, E. and Saba, A. (1991/2).** "Comparison of multivariate methods of analysis to evaluate panellists' performances." *Food Quality and Preference* **3**(4): 201-208.
- Sinesio, F., Risvik, E. and Rodbotten, M. (1990).** "Evaluation of panelist performance in descriptive profiling of rancid sausages : a multivariate study." *Journal of Sensory Studies* **5**(1): 33-52.

- T**homas-Danguin, T., Rouby, C., Sicard, G., Vigouroux, M., Barkat, S., Brun, V. et al. (2003). Sensory analysis and olfactory perception: some sources of variation. Handbook of flavor characterization. K. D. Deibler and J. Delwiche. New York, Marcel Dekker, Inc.: 65-91.
- Trondsen, T., Braaten, T., Lund, E. and Eggen, A. E. (2004).** "Consumption of seafood--the influence of overweight and health beliefs." *Food Quality and Preference* **15**(4): 361-374.
- Tucker, L. R. (1966).** "Some mathematical notes on three-mode factor analysis." *Psychometrika* **31**: 279-311.

- V**annier, A., Brun, O. X. and Feinberg, M. H. (1999). "Application of sensory analysis to champagne wine characterisation and discrimination." *Food Quality and Preference* **10**(2): 101-107.
- Verlegh, P. W. J. and Steekamp, J. B. E. M. (1999).** "A review and meta-analysis of country-of-origin research." *Journal of Economic Psychology* **20**: 521-546.

- W**olfinger, R. D. (1996). "Heterogeneous variance-covariance structure for repeated measures." *Journal of Agricultural, Biological, and Environmental statistics* **1**(2): 205-230.
- Wolters, C. J. and Allchurch, E. M. (1994).** "Effect of training procedure on the performance of descriptive panels." *Food Quality and Preference* **5**(3): 203-214.

- Y**ao, E., Lim, J., Tamaki, K., Ishii, R., Kim, K. O. and O'Mahony, M. (2003). "Structured and unstructured 9-point hedonic scales: A cross cultural study with American, Japanese and Korean consumers." *Journal of Sensory Studies* **18**(2): 115-139.

Index des tableaux

Tableau 1-1 : DDL et SC des facteurs de l'ANOVA du groupe	16
Tableau 1-2 : Espérance des carrés moyens selon la nature du facteur sujet	17
Tableau 1-3 : DDL et SC des facteurs de l'ANOVA du groupe (pas de répétition)	21
Tableau 1-4 : Définition des 6 indices de performances individuels de GRAPES	27
Tableau 1-5 : Définition des 6 indices de performances au niveau du groupe de GRAPES	29
Tableau 1-6 : Exemple de table CAP (d'après Schlich, 2000)	31
Tableau 2-1 : exemple de table CAP	52
Tableau 2-2 : couleurs associées aux risques des indices FPROD et FINTER	52
Tableau 2-3 : Signification des diagnostics de la table CAP	55
Tableau 2-4 : Accord multidimensionnel	62
Tableau 2-5 : Discrimination multidimensionnelle	64
Tableau 2-6 : table FLASH	68
Tableau 4-1 : Participation des panélistes selon les temps de mesure	88
Tableau 4-2 : Tests de différence des modèles emboîtés basés sur le logarithme du rapport de vraisemblance – étude n°1	97
Tableau 4-3 : indice AIC pour chaque descripteur en fonction de la structure de covariance	99
Tableau 4-4 : Estimation des paramètres de l'effet fixe temps	100
Tableau 4-5 : Estimation des paramètres individuels	101
Tableau 4-6 : Résumé des alertes de chaque panéliste en fonction des temps de mesure (performance significativement moins élevée que celle du groupe)	107
Tableau 4-7 : Synthèse du choix des structures de covariance selon les tests de modèles emboîtés et le critère AIC	110
Tableau 4-8 : Résumé de la significativité et des signes des paramètres de l'effet fixe temps pour l'ensemble des descripteurs selon le modèle mixte retenu	112
Tableau 4-9 : bilan de l'estimation des paramètres individuels	113
Tableau 4-10 : Résumé des alertes de chaque panéliste en fonction des temps de mesure (performance significativement moins élevée que celle du groupe)	116
Tableau 4-11 : Estimations des paramètres des modèles mixtes avec et sans la covariable "nbobs"	118
Tableau 5-1 : Nombre de jeux de données et de fournisseurs selon la variable analysée	128
Tableau 5-2 : Effectifs moyens selon le type de laboratoire	135
Tableau 5-3 : Niveau moyen de chaque indice de performance	137
Tableau 5-4 : Effet de l'âge sur le niveau de performance des panélistes	140
Tableau 5-5 : Effet du sexe sur le niveau de performance des panélistes	141
Tableau 5-6 : Effet du niveau d'expérience sur le niveau de performance des panélistes	142
Tableau 5-7 : Effet du niveau d'éducation sur le niveau de performance des panélistes	142
Tableau 5-8 : Effet du type de descripteur sur le niveau de performance des panélistes	143
Tableau 5-9 : Résultats des ANOVA faisant intervenir l'effet du type de descripteur	144
Tableau 5-10 : Effet du fournisseur sur les niveaux de performance des panels	146
Tableau 5-11 : Effet du type d'échelle sur les niveaux de performance des panels	147
Tableau 5-12 : Comparaison des niveaux de performance entre la France et l'étranger	148
Tableau 5-13 : Comparaison des niveaux de performance entre les laboratoires privés et publics	149
Tableau 5-14 : Effet du type de produit sur les niveaux de performance des panels	149
Tableau 6-1 : comparaison de la significativité du désaccord entre le modèle classique et l'Assessor Model. Diagnostics possibles	160
Tableau 6-2 : Résultats de la comparaison des deux modèles par rapport au désaccord	163
Tableau 6-3 : Pourcentage de descripteurs dont l'effet est significatif selon le modèle utilisé	169

Index des illustrations

Figure 1-1 : Différences d'utilisation de l'échelle	11
Figure 1-2 : Illustration des concepts de répétabilité, discrimination et accord.....	12
Figure 1-3 : Représentation graphique du terme d'interaction.....	19
Figure 1-4 : Construction des graphiques « Eggshell plot »	25
Figure 1-5 : Graphique des indices de GRAPES concernant l'utilisation de l'échelle	28
Figure 1-6 : Graphique des indices de GRAPES concernant les différences entre produit.....	28
Figure 1-7 : Graphique des indices de GRAPES concernant la dispersion des répétitions	29
Figure 1-8 : Définition de la boîte à moustache construite pour chaque indice de performance de GRAPES.....	30
Figure 1-9 : Détermination du diagnostic CAP	31
Figure 1-10 : Projection des variables sur le plan factoriel 1-2 (ACP)	33
Figure 1-11 : Projection des individus sur le plan factoriel 1-2 (ACP).....	34
Figure 1-12 : Représentation « biplot » sur le plan factoriel 1-2 (ACP).....	34
Figure 1-13 : ACP (pour un descripteur) des données centrées par juge	37
Figure 1-14 : ACP pour un descripteur avec les sujets en variables	37
Figure 1-15 : parallèle ente ANOVA et MANOVA (d'après Schlich, 2004)	39
Figure 1-16 : Plan factoriel (axes 1 et 2) de la CVA.....	40
Figure 1-17 : Représentation géométrique de la démarche de la méthode GPA.....	42
Figure 1-18 : Comparaison des méthodes STATIS, STATIS DUAL et GAMMA	47
Figure 2-1 : Calcul de l'indice RANKF	54
Figure 2-2 : Exemple de CAPMAP	57
Figure 2-3 : diagnostic individuel du juge 7 issu de la CAPMAP	58
Figure 2-4 : plan factoriel de l'analyse en variable canonique (AVC)	66
Figure 2-5 : Aperçu de la présentation des résultats issus du programme PERF	70
Figure 3-1 : Schéma de fonctionnement de la SensoBase	75
Figure 3-2 : Instructions générales pour renseigner le fichier SensoBase Excel2003® (feuille 1) .	76
Figure 3-3 : Extraits de la feuille de caractérisation des données du fichier SensoBase Excel2003 (feuille 2).....	77
Figure 3-4 : Récapitulatif des données ou liste des erreurs affichée à la fin de la saisie des caractéristiques des données.....	77
Figure 3-5 : Structure des données dans le fichier SensoBase excel2003 (feuille 3).....	78
Figure 3-6 : Tableau croisé du nombre de répétitions par sujet et par produit	78
Figure 3-7 : Schéma structurelle de la base de données SensoBase	80
Figure 3-8 : Page d'accueil du site de la SensoBase.....	81
Figure 3-9 : Espace personnel accessible après inscription au projet SensoBase	82
Figure 3-10 : page contenant les paramètres qui peuvent être modifiés lors d'une demande d'analyse.....	84
Figure 3-11 : Récapitulatif de la demande d'analyse	84
Figure 4-1 : Données utilisées pour étudier l'évolution de la performance	89
Figure 4-2 : Représentations graphiques des modèles mixtes en fonction de la partie fixe et de la partie aléatoire.....	91
Figure 4-3 : Arbre de décision permettant d'établir la sélection de la structure de covariance de chaque modèle	98
Figure 4-4 : carte de contrôle de l'évolution du panel pour le descripteur ar_4	103
Figure 4-5 : carte de contrôle de l'évolution des panélistes pour le descripteur ar_4.....	103
Figure 4-6 : carte de contrôle de l'évolution du panel pour le descripteur tx_3	104
Figure 4-7 : carte de contrôle de l'évolution des panélistes pour le descripteur tx_3	104
Figure 4-8 : carte de contrôle de l'évolution du panel pour le descripteur od_6	105
Figure 4-9 : carte de contrôle de l'évolution des panélistes pour le descripteur od_6.....	105
Figure 4-10 : carte de contrôle de l'évolution du panel pour le descripteur ap_4	106
Figure 4-11 : carte de contrôle de l'évolution des panélistes pour le descripteur ap_4	106
Figure 4-12 : représentations de la normalité des résidus pour le descripteur od_4 (modèle mixte avec effet fixe n.s. et structure de covariance RCL).....	108
Figure 4-13 : Evolution de l'AIC de chaque descripteur selon la structure de corrélation	111

Figure 4-14 : carte de contrôle de l'évolution de la répétabilité du panel pour le descripteur ap_red (structure de covariance CS, effet temps quadratique).....	114
Figure 4-15 : carte de contrôle de l'évolution des panélistes pour le descripteur ap_red.....	115
Figure 4-16 : représentations de la normalité des résidus pour le descripteur t_burnt (modèle mixte avec effet fixe quadratique et structure de covariance CS).....	119
Figure 5-1.a : distribution du nombre de panélistes par étude.....	127
Figure 5-2 : Distribution des panélistes en fonction de l'âge.....	129
Figure 5-3 : Distribution des proportions d'homme par étude.....	129
Figure 5-4 : Distribution des types de descripteurs.....	130
Figure 5-5 : Exemple de calcul des poids affectés aux panélistes d'une étude.....	132
Figure 5-6 : représentation du nombre de profils sensoriels selon les fournisseurs.....	134
Figure 5-7 : Distributions des indices de performance (unité expérimentale = le panéliste dans une étude).....	137
Figure 5-8 : Distribution des probabilités associées au test de l'effet produit dans les modèles d'ANOVA individuels.....	138
Figure 5-9 : Distributions des pourcentages de descripteurs pour lesquels un panéliste est discriminant dans une étude.....	139
Figure 5-10 : Distribution des probabilités associées coefficients de corrélation de Pearson individuels.....	139
Figure 5-11 : Distributions des pourcentages de descripteurs pour lesquels un panéliste est en accord avec le groupe dans une étude.....	140
Figure 5-12 : Représentation des moyennes et des intervalles de confiance de chaque fournisseur pour chaque indice de performance.....	147
Figure 6-1 : Représentation géométrique du paramètre β de l'Assessor Model.....	155
Figure 6-2 : Décomposition de la variance selon le modèle classique ou l'Assessor Model.....	156
Figure 6-3 : Effet de la transformation des données sur la significativité de l'effet produit dans le modèle classique.....	164
Figure 6-4 : Effet de la modification du dénominateur (CMerreur au lieu de CMinteraction) dans le calcul de l'effet produit.....	165
Figure 6-5 : synthèse de la comparaison entre le modèle classique et l'Assessor Model.....	166
Figure 6-6 : Représentation graphique de la transformation proposée par Brockhoff.....	167
Figure 6-7 : Relation entre le coefficient β et la dispersion des notes.....	171

Annexes

Annexe 1

Comparaison des méthodes multidimensionnelles
STATIS, STATIS DUAL et GAMMA

Méthode	STATIS	STATIS DUAL	GAMMA
Données	M juges évaluent les mêmes N produits sur P _i descripteurs 	M juges évaluent N _i produits sur les mêmes P descripteurs 	N produits M juges
Centrage	Les tableaux de notes X _i sont toujours centrés en colonnes		
Objet représentant le sujet i	Matrice N*N de produits scalaires entre produits $W_i = X_i X_i'$	Matrice P*P de covariances ou corrélations entre descripteurs (à 1/n près) $V_i = X_i' X_i$	Tableaux N*P de notes X _i
Standardisation des objets	$W_i = \frac{W_i}{\ W_i\ }$ où $\ W_i\ ^2 = \sum_{k,l} w_{kl}^2$	$V_i = \frac{V_i}{\ V_i\ }$ où $\ V_i\ ^2 = \sum_{k,l} v_{kl}^2$	$X_i = \frac{X_i}{\ X_i\ }$ où $\ X_i\ ^2 = \sum_{k,l} x_{kl}^2$
Similarité entre 2 sujets i et j	Produit scalaire $\langle W_i, W_j \rangle = r_{ij} = \text{trace}(W_i W_j)$ $r_{ij}=1$ si les configurations de produits des 2 sujets sont identiques	Produit scalaire $\langle V_i, V_j \rangle = r_{ij} = \text{trace}(V_i V_j)$ $r_{ij}=1$ si les 2 sujets ont les mêmes matrices de corrélations entre descripteurs	Transformation linéaire du produit scalaire $\langle X_i, X_j \rangle : t_{ij} = [1 + \text{trace}(X_i' X_j)] / 2$ $t_{ij}=1$ si les 2 sujets ont caractérisé les produits sur les mêmes descripteurs.
Matrice de similarité entre sujets	$RV = \frac{RV}{\ RV\ }$ où $\ RV\ ^2 = \sum_{k,l} r_{kl}^2$ $0 \leq r_{ij} \leq 1$	$RV_D = \frac{RV_D}{\ RV_D\ }$ où $\ RV_D\ ^2 = \sum_{k,l} r_{ij}^2 d_{ij}^2$ $0 \leq r_{ij} d_{ij} \leq 1$	$T = \frac{T}{\ T\ }$ où $\ T\ ^2 = \sum_{k,l} t_{ij}^2$ $0 \leq t_{ij} \leq 1$
Carte des Sujets	Coordonnées des sujets : 2 premiers vecteurs propres β_1 et β_2 de la PCO double centrée ⁴ de la matrice de similarités (RV, RVD ou T)		
Poids des sujets	$w_i = \beta_i / \sum \beta_i$ où β_i est le 1 ^{er} vecteur propre de norme 1 de la matrice de similarité entre sujets		
Compromis	$W = \frac{\sum_{i=1}^M w_i W_i}{\left\ \sum_{i=1}^M w_i W_i \right\ }$	$V = \frac{\sum_{i=1}^M w_i V_i}{\left\ \sum_{i=1}^M w_i V_i \right\ }$	$C = \frac{\sum_{i=1}^M w_i X_i}{\left\ \sum_{i=1}^M w_i X_i \right\ }$
Carte produits	Les 2 premiers vecteurs propres de la PCO double centrée de la matrice compromise permettent d'obtenir un plan factoriel de la carte des produits		

⁴ Double centrage d'une matrice $A \rightarrow A_{ij} = A_{ij} - A_{i.} - A_{.j} + A_{..}$.

Annexe 2

Notice d'interprétations des analyses statistiques
de la SensoBase

Notice d'interprétation des sorties de la batterie d'indices

NB : Cette notice est une version bêta qui peut encore contenir des erreurs de typographie et/ou de syntaxe. Si vous rencontrez des erreurs ou des problèmes de compréhension de cette notice, j'apprécierai beaucoup que vous m'en fassiez part (phineau@cesg.cmr.fr).

Table CAP

La table CAP permet de rendre compte des performances d'un panel d'analyse sensorielle ayant réalisé le profil descriptif d'un ensemble de produits sur un certain nombre de descripteurs. Pour que les calculs puissent être menés, les sujets doivent avoir effectué des répétitions de la dégustation de chaque produit. Cette table se présente de la manière suivante :

DIAGNOSTIC CAP Table of Data_potatoes (nsudj=10 - nprod=16 - nattr=12)

attribut	fprod	finfer	mean	rmse	_1	_8	_6	_5	_3	_7	_4	_10	_2	_9
SWEET	9.49	1.10	47.2	22.1	=	=	+	+	+	=	+	=	=	=
CHESNUT	6.02	1.10	46.6	24.3	-	-	+	+	=	+	=	=	=	+
PASTRY	4.42	1.51	19.9	20.7	-	-	=	+	=	=	+	+	=	=
HERBACE	3.23	1.55	22.8	22.4	+	=	=	=	=	=	=	+	=	=
BITTER	2.87	1.28	32.6	22.1	+	=	=	+	=	=	+	+	=	=
RAWPOTAT	2.81	1.09	7.68	13.9	-	=	+	+	=	=	=	=	=	=
CEREAL	2.32	1.05	27.0	21.0	+	=	+	=	=	=	=	=	=	=
CELERY	1.86	1.38	20.0	17.4	=	+	=	=	=	+	=	=	=	=
BUTTER	1.73	1.52	13.7	16.2	=	=	=	=	=	=	=	=	=	=
NUTS	1.63	1.65	17.2	17.4	=	-	-	=	=	=	=	=	=	=
CARDOAR	1.45	1.30	25.8	22.0	+	=	=	=	=	=	=	=	=	=
MUSHROOM	1.38	1.29	17.5	18.2	=	+	+	=	=	=	=	=	=	=
rankf				3.33	4.17	4.50	4.75	4.83	6.08	6.50	6.75	6.83	7.25	

ATTRIBUT
La première colonne contient le nom de chaque attribut ou descripteur sensoriel évalué dans l'étude. Chaque ligne présente les résultats relatifs à un attribut différent.

FPROD, FINFER
Il s'agit de la statistique du test de Fisher de l'effet produit et de l'interaction sujet*produit dans le modèle d'analyse de la variance :
sujet + produit + sujet*produit

ou l'effet sujet et l'interaction sujet*produit sont des effets aléatoires.
Pour chaque attribut, FPROD mesure, sur l'ensemble du groupe, l'importance des différences entre les moyennes des produits et est relatif à l'accord des sujets quant à ces différences. Si les moyennes des produits sont très dispersées pour cet attribut et que la plupart des sujets s'accordent sur l'ordonnement de ces moyennes, alors FPROD sera grand. Si au contraire les moyennes des produits sont peu dispersées et/ou si les sujets diffèrent sur l'ordonnement de ces moyennes, alors FPROD sera petit.

FINTER mesure, sur l'ensemble du groupe, l'importance des différences entre le classement des produits de chaque sujet. Si tous les sujets ont classé les produits dans le même ordre pour un descripteur donné, alors FINTER sera petit. Si au contraire l'ordre des produits pour un descripteur est très différent d'un sujet à l'autre, alors FINTER sera grand. Il faut néanmoins noter que lorsque les sujets notent les produits dans le même ordre mais qu'ils n'utilisent pas l'échelle de la même manière (certains n'utilisent qu'une partie de l'échelle, par exemple), cela contribue également à augmenter l'effet d'interaction. L'interaction peut donc également être significative du fait de la variabilité entre les utilisations d'échelles. FINTER ne reflète donc pas exclusivement le désaccord entre le classement des produits.

Ces valeurs de FPROD et FINTER sont testées à l'aide de la loi de Fisher afin de pouvoir conclure sur leur significativité. Cela consiste à calculer le risque, noté α (entre 0 et 1), de conclure qu'un effet est significatif à tort. Plus FPROD ou FINTER est grand, plus le risque de conclure qu'il existe un effet à tort sera faible, donc α diminue. Plus FPROD ou FINTER est petit, plus le risque α augmente. En analyse sensorielle, le risque maximum classiquement autorisé est de 5%. Dans le cas du tableau CAP, plusieurs seuils ont été utilisés pour moduler les conclusions. Cela donne lieu à l'étiquetage d'un code couleur comme suit :

Pour les valeurs de FPROD, si le risque α correspondant est :

< 5%, l'effet produit est significatif. Il existe des différences nettes entre les moyennes des notes des produits pour ce descripteur. La valeur du FPROD est donc sur fond **VERT**.

compris entre 5% et 15%, l'effet produit est de l'ordre de la tendance pour ce descripteur. Il existe une légère différence entre les produits. La valeur du FPROD est donc sur fond **ORANGE**.

> 15%, il n'y a pas d'effet produit. Les produits n'ont pas été différenciés pour ce descripteur. La valeur du FPROD est donc sur fond **ROUGE**.

Pour les valeurs de FINTER, si le risque α correspondant est :

> 10%, il n'y a pas d'effet de l'interaction. Les sujets ne sont pas en désaccord sur le classement des produits pour ce descripteur. La valeur du FINTER est donc sur fond **VERT**.

compris entre 1% et 10%, l'effet de l'interaction est de l'ordre de la tendance pour ce descripteur. Il existe un léger désaccord entre les sujets. La valeur du FINTER est donc sur fond **ORANGE**.

< 1%, l'effet de l'interaction est significatif. On observe un désaccord net entre les classements des produits par les sujets pour ce descripteur. La valeur du FINTER est donc sur fond **ROUGE**.

NB : il est tout à fait possible que les sujets soient en accord et qu'il n'y ait pas d'effet produit ou inversement. FPROD et FINTER ne sont pas directement liés.

MEAN

Cette colonne contient la moyenne des notes données à un descripteur sur l'ensemble du panel. Elle permet de voir le niveau moyen de notation sur l'échelle. Une valeur très faible pour un descripteur peut indiquer que la sensation en question est très peu perçue, au moins pour la majorité des produits.

RMSE

Le Root Mean Square of Error (écart-type résiduel) donne une indication sur la variabilité non expliquée par les effets du modèle d'anova (produit, sujet et interaction ici). Il est exprimé dans la même unité que l'échelle de notation. Il correspond à la valeur moyenne dont s'écarte une observation (la note d'un sujet pour un produit) de la moyenne générale, après avoir tenu compte de l'écart expliqué par le produit et le sujet en question. Dans le cas présent, il représente la valeur moyenne dont s'écartent les répétitions entre elles pour un sujet et un produit donné. Si le RMSE est faible, la variabilité moyenne entre les répétitions est faible donc le panel est répétable sur ce descripteur. Si le RMSE est élevé, la variabilité entre les répétitions est forte donc les sujets ne sont pas répétables.

RANKE

Les colonnes restantes correspondent chacune à un des sujets de l'étude. Ils sont classés selon leur RANKE, dont la valeur est donnée sur la dernière ligne. Cette valeur représente une moyenne de rangs. Pour chaque descripteur, les sujets sont classés selon leur pouvoir discriminant (capacité à faire des différences entre les produits). Pour cela, une anova à un facteur produit est calculée par sujet. Le pouvoir discriminant est la valeur du test de Fisher de l'effet produit. Le sujet ayant le meilleur pouvoir discriminant reçoit le rang 1, le dernier sujet reçoit le dernier rang. Chaque sujet reçoit au total

autant de rangs que de descripteurs. RANKF correspond à la moyenne de ces rangs pour un sujet. Si un sujet obtient un RANKF de 1, c'est qu'il a toujours obtenu le meilleur pouvoir discriminant pour tous les descripteurs. A l'opposé, le RANKF maximum correspond au nombre de descripteurs. Les sujets sont donc classés par RANKF croissants (du plus discriminant en moyenne au moins discriminant).

DIAGNOSTICS INDIVIDUELS +/-/=!!

En fonction de son pouvoir discriminant, de son accord avec le groupe et de sa répétabilité, chaque sujet se voit attribuer un diagnostic parmi +, -, +!, -!, =!, =. La détermination de ce diagnostic se fait de la manière suivante pour chaque descripteur :

+ , - ou =

Si le pouvoir discriminant (Fproduit de l'anova à un facteur produit) du sujet n'est pas significatif ($\alpha > 10\%$), le sujet aura le signe =. Le sujet ne fait pas de différence significative entre les produits pour le descripteur considéré.

Si le pouvoir discriminant est significatif et que le sujet est en accord avec le groupe, le sujet aura le signe +. S'il est discriminant et en désaccord, le sujet aura le signe -.

L'accord est calculé en utilisant le coefficient de Kendall. Celui-ci compte les paires de produits pour lesquelles les produits sont classés dans le même ordre entre le sujet considéré et la moyenne des autres sujets discriminants. Par conséquent, plus les produits sont classés dans le même ordre entre les sujets, plus le Kendall est grand. Lorsque ce coefficient est significatif au seuil de 20%, le sujet reçoit le signe + pour ce descripteur. Sinon il reçoit le signe - (sujet en désaccord avec le panel).

! ou non

Lorsqu'un sujet est significativement moins répétable que la moyenne du panel il reçoit le signe ! en plus du signe +, - ou =.

La répétabilité est basée sur le calcul des écart-types entre les répétitions d'un même produit pour un descripteur. Plus cette valeur est faible, plus le sujet est répétable. L'effet sujet de l'anova par descripteur selon le modèle produit + sujet (où les objets sont ces écart-types entre répétitions) permet de dire si tous les sujets sont aussi répétables sur les produits. Un test de contraste (au seuil de 1%) permet de détecter les sujets ayant des valeurs d'écart-types significativement plus grandes. Ces sujets sont les moins répétables du panel. Ils reçoivent le signe ! en plus du signe +, - ou =.

Le diagnostic final se trouve donc parmi les 6 cas suivants :

+	Discriminant, en accord, répétable*
+!	Discriminant, en accord, non répétable*
=	Non discriminant, répétable*
=!	Non discriminant, non répétable*
-	Discriminant, en désaccord, répétable*
-!	Discriminant, en désaccord, non répétable*

* : répétabilité par rapport au panel

? ou non

Il est possible que le symbole "?" apparaisse dans une table CAP. Cela signifie que, pour le descripteur en question, le sujet n'avait pas effectué de répétition (d'au moins un des produits). Les symboles "+", "?" et "-?" sont alors à interpréter avec précaution. Ils ne permettent pas de conclure que le sujet est discriminant (impossible de calculer une discrimination individuelle sans répétition) mais simplement de dire si le classement des produits de ce sujet est en accord avec le groupe. Dans de très rares cas, le symbole "=?" peut apparaître. Cela signifie alors que le sujet n'a pas effectué de répétition pour ce descripteur (?), et qu'il a donné la même note à tous les produits (=).

TABLE CAP p=1

Cette table CAP est identique à la première mais un des paramètres de décision du diagnostique (+/-/=) a été modifié. Cela a pour conséquence de prendre en compte tous les panélistes dans le calcul de l'accord de chaque panéliste avec le reste du groupe (au lieu de ne prendre en compte que les sujets qui étaient discriminants).

Techniquement, cela consiste à déclarer tous les sujets comme étant discriminants, quel que soit le niveau de leur Fproduit individuel (voir paragraphe "+, - ou =" des explications de la table CAP), donc à fixer le seuil de significativité de la probabilité du Fproduit à p=1. Le diagnostic "=" disparaît donc de la table pour ne laisser que des "+", "accord avec le groupe) ou des "-" (désaccord avec le groupe).

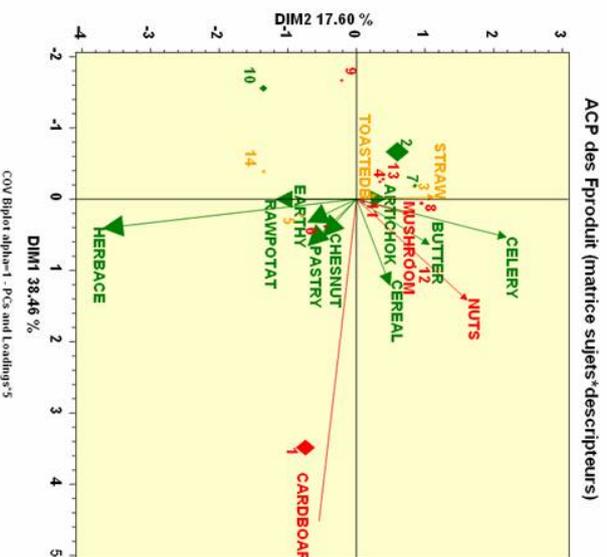
DIAGNOSTIC CAP Table of pdt (nsubj=10 - nprod=16 - nattr=12)

attribut	fprod	finter	mean	rmse	_1	_8	_6	_5	_3	_7	_4	_10	_2	_9
SWEET	9.49	1.10	42.4	21.9	+	+	+	+	+	+	+	+	+	+
CHESNUT	6.02	1.10	41.1	24.3	+	+	-	+	+	+	+	+	+	+
PASTRY	4.42	1.51	19.2	20.1	+	-	-	+	+	+	+	+	+	+
HERBACE	3.23	1.55	30.5	21.2	+	-	+	-	+	+	-	+	+	+
BITTER	2.87	1.28	26.9	20.2	+	+	+	+	+	+	-	+	-	-
RAWPOTAT	2.81	1.09	22.7	21.9	+	-	-	+	+	+	+	+	+	+
CEREAL	2.32	1.05	26.3	22.5	+	-	+	-	-	+	-	+	+	-
CELERY	1.86	1.38	16.9	15.9	+	+	+	-	+	+	+	+	+	-
BUTTER	1.73	1.52	16.8	17.3	-	+	+	-	-	+	-	+	+	+
NUITS	1.63	1.65	16.3	16.5	+	-	-	+	+	+	-	+	+	+
CARDBOAR	1.45	1.30	8.07	13.8	+	+	+	+	-	+	+	+	+	+
MUSHROOM	1.38	1.29	12.0	14.9	+	-	-	+	-	+	+	+	+	+
rankf	3.33	4.17	4.50	4.75	4.83	6.08	6.50	6.75	6.83	7.25

Par défaut, cette table n'est pas disponible dans les sorties proposées. Pour y avoir accès, lors d'une demande d'analyse depuis le site Internet de la Sensobase, il faut sélectionner "OU" comme réponse à la proposition : "Table CAP supplémentaire où tous les sujets participent au calcul de l'accord".

CAPMAP

La CAPMAP est une ACP de la matrice sujets*descripteurs contenant les Fproduit (correspondent au pouvoir discriminant) des sujets pour chaque descripteur. Elle a pour but de faire apparaître les axes (de l'espace des descripteurs) selon lesquels les différences de pouvoir discriminants sont maximales entre les sujets.



En cliquant sur un point représentant la position d'un sujet sur la carte ou sur la pointe de la flèche d'un descripteur, cette carte donne également accès à des informations plus détaillées sur les performances liées à ce sujet ou à ce descripteur. Ces informations complémentaires sont commentées à la suite de l'aide sur la CAPMAP (partie « Tableaux individuels de performances »).

NB : en raison d'un problème informatique qui reste à résoudre, il est possible que vous deviez cliquer non pas **sur** le point ou **sur** la flèche mais à quelques millimètres à **gauche** du point ou de la flèche. Sinon, il est aussi possible d'ouvrir directement le tableau que vous souhaitez par l'explorateur windows en allant dans le même répertoire que le fichier index.html. Le nom du fichier à ouvrir se nomme "obs + nom du sujet ou du descripteur + .html" (par exemple, vous devez ouvrir "obsCELEERY.html" pour obtenir les informations sur le descripteur celery ou "obs10.html" pour obtenir les informations sur le sujet 10). Cette astuce peut aussi être utile quand plusieurs point se chevauchent.

LES DESCRIPTEURS

Longueur de la flèche

Plus un descripteur présente une flèche longue, plus le pouvoir discriminant est différent entre les sujets. Certains sujets sont donc meilleurs que d'autres pour faire des différences entre les produits pour ce descripteur. Plus la flèche est courte, plus le niveau de discrimination est homogène au sein du panel. Cela ne traduit en rien le niveau global de discrimination ! Une flèche courte peut correspondre à un niveau de discrimination élevé ou faible du panel.

Taille de la pointe

La taille de la pointe indique si le panel est globalement discriminant pour ce descripteur ou non. Si la pointe est grosse, le panel est très discriminant. Si la pointe est petite, le panel n'est pas discriminant. Elle dépend directement de la valeur du FPROD.

Couleur de la flèche

La couleur de la flèche représente l'accord entre les sujets du panel. Elle dépend à la fois des valeurs de FPROD et de FINITER (respectivement les Fproduit et l'interaction de l'anova à deux facteurs sujet et produit + interaction : plus d'explication dans la notice de la table CAP). Si FPROD est significatif ($\alpha < 5\%$), cela signifie que les sujets sont suffisamment en accord pour qu'il existe des différences significatives entre les produits au niveau du panel (effet produit significativement plus grand que l'interaction). La flèche sera donc verte, que le FINITER (désaccord entre les sujets) soit significatif ou non. Si FPROD n'est pas significatif, le panel n'est pas discriminant. La couleur de la flèche est alors déterminée par le FINITER. S'il n'est pas significatif, le désaccord entre les sujets n'est pas significatif. Dans ce cas (pas de différences entre les produits mais pas de désaccord non plus entre les sujets), la flèche sera orange. Si FINITER est significatif, les sujets sont en désaccord entre eux sur le classement des produits pour le descripteur considéré. Une éventuelle discrimination entre les produits a donc pu être masquée par le désaccord entre les sujets. La flèche sera donc rouge.

LES SUJETS

Position d'un sujet

La place d'un sujet sur la carte dépend des FPROD qu'il a obtenus pour chaque descripteur. Plus son FPROD sera élevé (par rapport au groupe) pour un descripteur, plus ce sujet sera éloigné du centre de la carte dans la direction de la flèche du descripteur. Plus son FPROD sera faible, plus le sujet se situera dans la direction opposée. Si un sujet obtient un Fprod moyen, il sera proche de zéro.

Taille du point

La taille du point d'un sujet est fonction de son RANKF (présenté dans l'aide de la table CAP). Il représente le rang moyen d'un sujet en sein du panel en terme de pouvoir discriminant. Si RANKF d'un sujet est petit (proche de 1), celui-ci a mieux différencié les produits que les autres sujets sur l'ensemble des descripteurs. La taille du point correspondant sera donc grande. Si le RANKF d'un sujet est grand (proche du nombre de sujets de l'étude) si celui-ci a moins bien différencié les produits que les autres sujets sur l'ensemble des descripteurs. La taille du point correspondant sera donc petite.

Couleur du point

La couleur du point représente l'accord ou le désaccord entre le sujet et les autres panelistes. Elle dépend de la valeur du ALPHA (différent du α utilisé dans les probabilités) présenté dans le tableau « Accord multidimensionnel ». La couleur du point sur la carte correspond à celle de ce tableau. Si un sujet a généralement classé les produits dans le même ordre que les autres sujets sur la plupart des descripteurs, son ALPHA sera élevé (proche de 1). Plus le classement de ses produits sera différent des autres sujets et sur un grand nombre de descripteurs, plus ALPHA sera faible (proche de zéro). Le coefficient ALPHA peut ensuite être Normalisé, ce qui permet de lui associer une probabilité de déclarer que le sujet est en accord avec le groupe à tort.

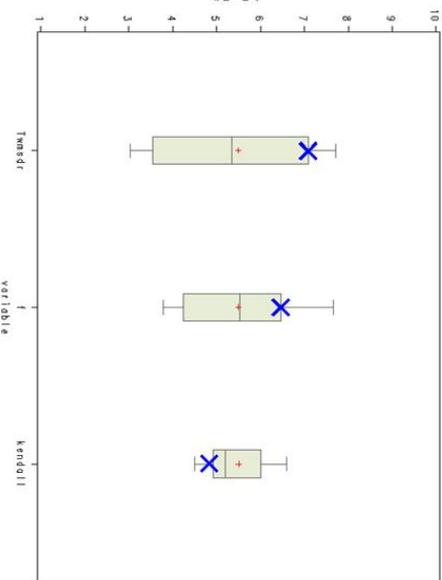
Lorsque la valeur normalisée est supérieure à 2 (correspondant approximativement à un risque de 5% pour la loi normale centrée réduite), le sujet est dit significativement en accord avec le groupe ; le point sera de couleur verte. Si cette valeur est comprise entre 1,5 et 2, le sujet a plutôt tendance à être en accord avec le groupe ; le point sera beige. Si la valeur est entre 1 et 1,5, le sujet est plutôt en léger désaccord avec le groupe ; le point sera orange. En dessous de 1, le sujet est en désaccord avec le panel ; le point sera rouge.

Tableaux individuels de performances

Ces tableaux sont accessibles depuis la CAPMAP en cliquant sur un point représentant un sujet sur la carte ou sur la pointe de la flèche d'un descripteur. Que l'on choisisse un sujet ou un descripteur, les outils statistiques utilisés sont les mêmes. Cette partie ne présente donc en détail que les résultats obtenus pour un sujet. Quelques précisions seront toutefois apportées dans le cas des descripteurs.

Lorsque vous cliquez sur un sujet (sujet "6" dans cet exemple) de la carte de la CAPMAP, une fenêtre s'ouvre et contient les résultats suivants :

Boxplot de Variable pour subject=6



Diagnostic individuel univarié, sujet 6

attribut	diagnostic +/-/!	discrimination Fprod	répétabilité Twmsdr	accord Kendall
MUSHROOM	-	4.94	-1.44	-0.103
HERBACE	-	4.74	-1.47	0.008
SWEET	+	4.23	-2.02	0.225
BITTER	+	3.37	-3.05	0.218
NUTS	-	2.92	0.21	-0.263
CARDBOAR	+	2.19	-1.38	0.233
PASTRY	=	1.86	-0.66	.
CHESNUT	=	1.70	-0.07	.
RAWPOTAT	=	0.83	-4.43	.
CELERY	=	0.79	-4.86	.
BUTTER	= 1	0.74	1.84	.
CEREAL	=	0.43	0.27	.

Bien que ces résultats présentent d'abord les boxplots puis le tableau de performances, les boxplots ne seront présentés que dans une deuxième temps.

TABLEAU DE PERFORMANCES

Ce tableau de 5 colonnes présente des résultats sur les performances du sujet sélectionné sur la CAPMAP (sujet 6 ici) pour chaque descripteur.

Attributs

Cette colonne contient le nom de chacun des descripteurs triés par ordre décroissant de Fprod (colonne 3). Les descripteurs pour lesquels le sujet est le plus discriminant se trouvent donc en haut du tableau.

Diagnostic : +/-/!

Cette colonne reprend les symboles présentés dans la Table CAP pour le sujet considéré. La signification des symboles se trouve à la fin de la partie « Table CAP ». Le même code couleur est utilisé.

Discrimination : Fprod

Cette colonne contient les valeurs du test de Fisher de l'anova par sujet à un facteur produit (cf "+", - ou "=" de la partie « Table CAP »). Plus cette valeur est élevée, plus le sujet a fait de différences entre les produits pour ce descripteur. Les descripteurs sont classés par ordre décroissant de Fprod. Si le risque α associé au Fprod est inférieur à 10%, la case est verte (correspond au symbole "+", ou "=", dans la colonne diagnostic). Sinon elle est rouge (symbole "-").

Répétabilité : Twmsdr

Cette colonne contient les valeurs "studentisées" des écart-types entre les répétitions (cf "i" ou non" de la partie « Table CAP »). Pour chaque descripteur, ces écart-types sont calculés par sujet et par produit. Plus ces écart-types sont petits, plus le sujet est répétable pour ce produit sur ce descripteur. Sur l'ensemble du panel, ces indices peuvent avoir un niveau moyen très différent d'un descripteur à l'autre. Par exemple, si un descripteur est beaucoup plus facile à noter qu'un autre, la moyenne des écart-types sur tout le panel sera forcément plus petite. Un test de Student est donc réalisé pour chaque sujet et sur chaque descripteur entre la moyenne des écart-types d'un sujet et la moyenne des écart-types du reste du panel, de manière à mesurer si son niveau de répétabilité est significativement inférieur à la moyenne du groupe pour ce descripteur. Cette valeur "studentisée" des écart-types d'un sujet est appelée Twmsdr. Si les écart-types d'un sujet sont plus petits que ceux du panel, Twmsdr est négatif et le sujet est plus répétable que la moyenne du groupe sur ce descripteur. Si les écart-types d'un sujet sont plus grands que ceux du panel pour ce descripteur, Twmsdr est positif et le sujet est moins répétable que la moyenne du groupe. Si Twmsdr est positif et significativement différent de 0 au seuil $\alpha = 1%$ (test unilatéral à droite), la case sera rouge (correspond à la présence du symbole "i" de la colonne diagnostic). Sinon, la case sera verte (pas de symbole "i").

Accord avec le groupe : Kendall

Cette colonne contient les valeurs des coefficients de Kendall (cf "+", - ou "=" de la partie « Table CAP »), calculés uniquement pour les descripteurs pour lesquels le sujet est discriminant. Plus les produits sont classés dans le même ordre entre le sujet considéré et les autres sujets discriminants sur ce descripteur, plus le Kendall est grand (proche de 1). Si ce coefficient est significatif au seuil de 20%, la case est verte (correspond au symbole "+", de la colonne diagnostic), pour ce descripteur. Sinon la case est rouge (symbole "-").

LES BOXPLOTS

Les trois indices utilisés pour créer les trois boxplots sont les mêmes que ceux du tableau décrit ci-dessus. Chaque boxplot représente la dispersion des sujets concernant une valeur statistique. Le premier boxplot se base sur l'indice de répétabilité des sujets appelé Twmsdr (nom assez barbare qui sera changé ultérieurement), le second sur un indice de discrimination (le Fprod) et le troisième sur un indice d'accord (le coefficient de Kendall).

Chaque boxplot est construit de la manière suivante :

Le rectangle central s'étend du quantile 25% au quantile 75% de l'étendue des valeurs de l'indice considéré pour tous les sujets. Il contient donc la moitié des sujets du panel ayant obtenu les valeurs les plus médianes pour cet indice.

Le trait horizontal à l'intérieur du rectangle représente la valeur médiane.

La petite croix rouge représente la valeur moyenne.

Les moustaches (traits qui sortent du rectangle) peuvent mesurer jusqu'à 1,5 fois la taille du rectangle au maximum de chaque côté. Mais leur taille ne dépasse pas non plus la plus grande valeur observée dans les données pour la moustache haute, et la plus petite valeur observée pour la moustache basse.

La croix bleue représente la valeur du sujet considéré dont le nom est rappelé dans le titre du graphique (sujet "6" dans cet exemple).

Utilisation des indices : T_{WMSDR} , $F_{Produit}$, Kendall

Les données utilisées pour construire chaque boxplot se sont pas directement les valeurs des trois indices (T_{WMSDR} , $F_{Produit}$ et Kendall). Pour chaque descripteur, chaque indice est d'abord transformé en rang. Le meilleur sujet reçoit le rang le plus élevé (qui correspond au nombre total de sujets) ; le moins bon sujet reçoit le rang 1. Ces rangs sont ensuite moyennés sur l'ensemble des descripteurs pour chaque sujet. Chaque sujet se caractérise alors par trois rangs moyens, un par indice statistique. Ce sont ces valeurs qui vont permettre de construire les trois boxplots.

Cette transformation en rangs moyens permet de comparer les boxplots entre eux puisqu'ils utilisent la même échelle : le nombre de sujets. Par construction, les trois boxplots ont la même moyenne qui correspond à la moyenne des rangs. En revanche, l'étendue du boxplot est propre à chaque indice. Plus un boxplot est étendu, plus les rangs moyens des sujets sont différents donc plus le niveau de répétabilité, de discrimination ou d'accord est différent entre les sujets du panel sur l'ensemble des descripteurs. Au contraire, plus un boxplot est petit, plus les rangs moyens sont proches entre les sujets donc plus les différences entre sujets sont faibles sur l'ensemble des descripteurs (ce qui révèle une certaine homogénéité globale du panel pour cet indice). Dans l'exemple présenté ci-dessus, le boxplot de T_{WMSDR} est plus grand que celui de Kendall. Le niveau de répétabilité est donc plus différent d'un sujet à l'autre que le niveau d'accord.

Les boxplots sont donc les mêmes pour tous les sujets ; seule la place de la croix bleue varie. Cela permet donc d'évaluer rapidement, tous descripteurs confondus, les performances d'un sujet par rapport au panel en terme de répétabilité (T_{WMSDR}), de discrimination (F) et d'accord (Kendall). Les boxplots sont construits de manière à ce que plus une croix bleue se trouve vers le haut du boxplot, plus le sujet est performant pour cet indice. D'après le graphique montré en exemple, on peut donc conclure que le sujet 6 est très répétable par rapport aux autres sujets du panel, très discriminant également, mais qu'il est moins en accord avec le reste du panel que les autres sujets. Cette dernière conclusion est à relativiser à cause de la petite taille du boxplot pour le Kendall.

CAS PARTICULIER POUR LES RANGS DU KENDALL

Le fait que le coefficient de Kendall ne soit calculé que lorsque le sujet est discriminant pour un descripteur entraîne une difficulté pour le calcul des rangs. En effet, si seulement 5 sujets sur 12 sont discriminants pour un descripteur, seulement 5 des 12 rangs possibles (de 1 à 12) seront attribués. Afin de conserver une moyenne des rangs identique aux autres boxplots, nous avons choisi d'affecter les rangs les plus proches de la moyenne. Dans notre cas, la moyenne des rangs de 1 à 12 est égale à 6,5, donc les cinq rangs affectés auront pour valeur : 4,5, 5,5, 6,5, 7,5 et 8,5.

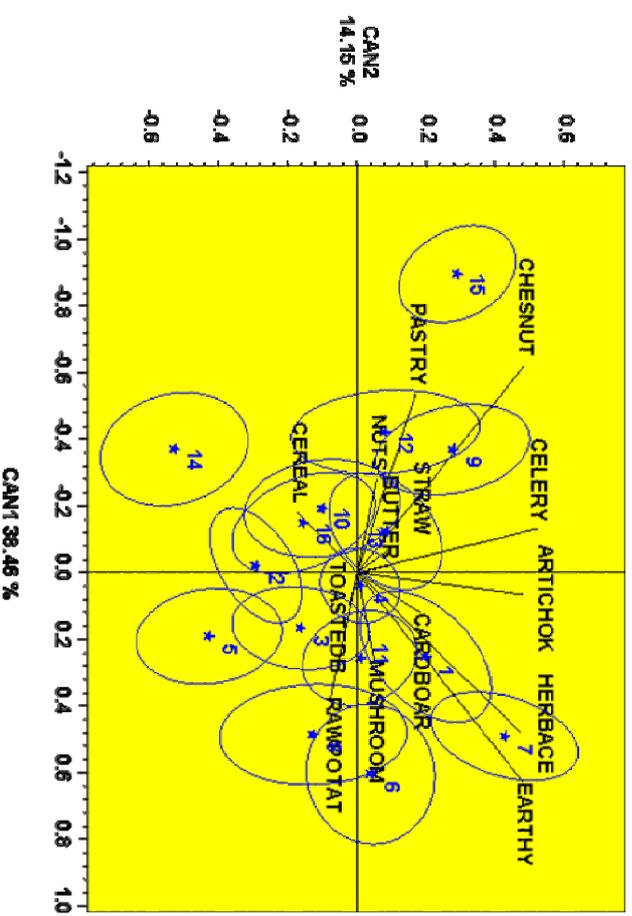
PERFORMANCES PAR DESCRIPTEUR

Lorsque vous cliquez sur la flèche d'un descripteur de la CAPMAP, le même type de boxplot et de tableau apparaît. Dans le tableau, chaque ligne ne correspond plus à un descripteur mais à un sujet et ceux-ci sont classés du plus discriminant au moins discriminant.

Pour la construction des boxplots, le procédé est identique à une exception près : les rangs ne sont pas affectés à chaque sujet mais à chaque descripteur. Par conséquent, pour calculer un rang moyen pour un descripteur, il faut affecter un rang à chaque descripteur pour chaque sujet (classement des descripteurs) puis faire la moyenne des rangs obtenus par chaque descripteur. De cette manière, un boxplot représente la dispersion des rangs moyens des descripteurs.

La CVA (canonical Variate Analysis)

Canonical Variate Analysis



L'analyse en variables canoniques (CVA) est en quelque sorte une ACP des moyennes des produits qui va tenir compte de la dispersion des notes des sujets autour de chaque moyenne. Elle a pour but de trouver les dimensions de l'espace des descripteurs qui maximisent la séparation entre les produits. Bien qu'il soit possible de réaliser une CVA qui prennent en compte les répétitions, ce n'est pas le cas présent. Dans l'analyse retournée par ce programme, seules les moyennes entre les répétitions sont utilisées, de manière à obtenir une seule valeur par sujet, produit descripteur. Chaque descripteur est représenté par une flèche. Plus cette flèche est longue, plus les sujets ont discriminé les produits de la même manière sur ce descripteur. Les produits sont représentés par des points. Plus un point est éloigné dans la direction d'un descripteur, plus ce produit a obtenu des intensités fortes pour ce descripteur par rapport aux autres produits.

De plus, cette méthode permet de tracer les **ellipses de confiance** autour de la moyenne de chaque point (dont le pourcentage de confiance est précisé en bas de page : 80% dans notre exemple). Plus les ellipses sont grandes, moins la zone de confiance autour de la moyenne est précise. Si deux ellipses ne se touchent pas, cela veut dire que les moyennes des produits sont significativement différentes (au seuil défini par l'ellipse, soit 100 - 80 = 20% ici).

F₁ et p associée

Cette valeur correspond au F de Fisher de la statistique de Hotelling-Lawley. Elle indique de manière globale si les produits ont été discriminés dans l'espace des descripteurs. La probabilité associée donne un renseignement sur la significativité de ce test. En général, ce test est toujours hautement

significatif car les sujets ont été capable de faire des différences entre les produits sur quelques descripteurs au moins. Il est donc souvent plus intéressant de regarder directement la valeur du $F_{i,j}$ (qui peut s'interpréter comme un rapport entre un effet produit et une résiduelle ou un désaccord (comme en anova)). Un F de 2 signifie par exemple qu'il existe en général un écart deux fois plus grand entre les moyennes de deux produits qu'entre les notes de deux sujets pour un même produit.

NDIMSIG

Cet entier représente le nombre de dimensions significatives de l'espace des descripteurs. Il donne un ordre d'idée de la dimensionnalité de l'espace produit. Néanmoins, il semble que cette valeur soit souvent surestimée. C'est donc un indice à utiliser en comparaison avec d'autres indices mais il est difficilement interprétable seul.

Nombre de descripteurs

Pour des raisons purement calculatoires, le nombre de descripteurs utilisés dans la CVA est limité à $p = (nb \text{ produits} - 1) * (nb \text{ sujets} - 1)$. Si le nombre de descripteurs dépasse ce seuil, la CVA ne peut pas aboutir. Par conséquent, avant de lancer cette analyse et si le nombre de descripteurs de l'étude est trop important, les descripteurs les moins discriminants sont supprimés afin de ne retenir que les p descripteurs ayant les meilleurs F produit. L'utilisateur peut également agir sur la sélection des descripteurs à l'aide du paramètre « seuil de la sélection de variables de la CVA » lors de la demande d'analyse via le site SENSODATABASE. Il régle ainsi le seuil de la probabilité en dessous de laquelle un descripteur ne sera pas considéré comme discriminant. Par défaut, cette valeur est fixée à 15%. Ce seuil peut être diminuer pour obtenir une sélection de variables plus sévère.

Accord multidimensionnel

CONF-SCOR-CORR

14 panelists - 16 products - 14 Flavor attributes

Panelist	RV	α	RVD	NRV	Na	NRVD
1	0.29	0.56	0.53	0.40	1.25	2.46
2	0.38	0.59	0.54	1.32	2.04	2.51
3	0.37	0.59	0.53	1.91	1.74	3.53
4	0.35	0.56	0.51	1.07	1.20	3.12
5	0.37	0.58	0.60	1.35	1.62	3.39
6	0.28	0.51	0.52	-0.02	0.18	2.98
7	0.37	0.61	0.52	1.20	2.46	3.12
8	0.29	0.53	0.41	0.21	0.80	2.45
9	0.33	0.57	0.54	1.17	1.39	2.95
10	0.40	0.61	0.54	2.15	2.27	3.14
11	0.27	0.57	0.52	0.61	1.39	2.56
12	0.32	0.55	0.45	0.75	1.29	2.67
13	0.27	0.54	0.50	0.40	0.72	3.22
14	0.36	0.57	0.44	0.72	1.66	2.14
Mean	0.33	0.57	0.51	0.94	1.43	2.88

Ce tableau contient, pour chaque sujet, trois indices d'accord avec le reste du panel (RV, α et RVD) et leur version normalisée (respectivement NRV, Na et NRVD). Chaque indice aborde la notion d'accord avec le groupe d'une manière différente. Les versions normalisées des indices permettent de conclure sur la significativité de l'accord. Dans les sorties dont vous disposez pour chaque jeu de données, ce tableau ne contient que trois colonnes. Il s'agit dans la mesure du possible des valeurs normalisées. Toutefois, si le jeu de données compte moins de 5 attributs, les valeurs normalisées Na

et NRVD sont peu fiables. Le tableau contiendra alors les valeurs de α et RVD. De même, si le jeu de données compte moins de 5 sujets, les valeurs normalisées NRV et Na sont peu fiables. Le tableau contiendra alors les valeurs de RV et α .

RV-NRV

Le RV, calculé dans la méthode STATIS, détermine si 2 sujets (ou un sujet et le reste du groupe) ont donné les mêmes "distances" entre les produits. En d'autres termes, si un sujet donne des notes similaires à deux produits sur l'ensemble des descripteurs, ces produits sont proches d'un point de vue sensoriel pour le paneliste. La proximité calculée entre ces produits (=corrélation entre les notes des deux produits) sera donc élevée. Cette proximité est calculée pour toutes les paires de produits du sujet. Afin de mesurer l'accord entre deux sujets, la méthode STATIS compare les valeurs de proximité de toutes les paires de produits entre les deux sujets. Plus les proximités entre les paires de produits sont similaires, plus le coefficient RV est proche de 1. Plus les proximités entre les paires de produits sont différentes (les sujets ne s'accordent pas sur les distances entre produits), plus le coefficient RV sera proche de 0.

Il faut ici noter que si l'un des sujets confond deux descripteurs (en inversant les notes entre ces descripteurs pour tous les produits par exemple), cela n'a pas d'effet sur les proximités calculées entre les paires de produits de ce sujet. Par conséquent, la comparaison avec les proximités d'un autre sujet ne changent pas non plus. Le RV sera donc strictement le même.

Pour un sujet donné, un RV est d'abord calculé entre ce sujet et chacun des autres. Ces (nombre de sujets-1) RV sont ensuite moyennés. La première colonne du tableau ci-dessus présente ces valeurs moyennes par sujet.

NRV est la version normalisée du RV. Elle permet, grâce à un test de permutations, de savoir si la valeur de RV obtenue est significativement supérieure au hasard. Si cette valeur est supérieure à 2 (quasiment significatif à 5%, pour une loi normale), le sujet est en accord avec le panel. La case sera alors verte. Si cette valeur est comprise entre 1,5 et 2, le sujet est en léger accord avec le panel. La case sera alors beige. Si cette valeur est comprise entre 1 et 1,5, le sujet est plutôt en désaccord avec le panel. La case sera alors orange. Si cette valeur est inférieure à 1, le sujet est en désaccord avec le panel. La case sera rouge.

α -Na

Le α calculé par la méthode GAMMA regarde cette fois si 2 sujets ont mis des notes proches aux mêmes descripteurs pour chaque produit. Plus les 2 sujets s'accordent à classer les produits de la même manière sur tous les descripteurs, plus le α sera grand. Par rapport au coefficient RV, le α tient compte dans ses calculs de la nature des descripteurs. Une inversion entre deux descripteurs se traduira cette fois par une baisse du coefficient α .

De même que pour l'indice précédent, la version normalisée Na permet de tester la significativité du coefficient α . Le même code couleur est adopté.

RVD-NRVD

Le RVD, calculé par la méthode STATIS DUAL, reprend les mêmes concepts que ceux du RV, mais appliqués aux descripteurs. Le but est ici de savoir si deux sujets font les mêmes associations entre les descripteurs. Si lorsqu'un sujet donne une note élevée pour un descripteur, il a tendance à donner également une note élevée à un autre ou plusieurs autres descripteurs quel que soit le produit, et que l'autre sujet fait la même chose, le coefficient RVD entre ces deux sujets sera proche de 1. En revanche, si un sujet ne fait pas les mêmes associations entre les descripteurs qu'un autre sujet, le coefficient RVD sera plus faible. Si les associations sont complètement opposées (lorsqu'un sujet donne une note élevée à un descripteur, il donne généralement une note élevée à un deuxième descripteur, alors que le deuxième sujet fait l'inverse), le coefficient RVD est proche de 0.

De même que pour les indices précédents, la version normalisée NRVD permet de tester la significativité du coefficient RVD. Le même code couleur est adopté.

CAP multivarié

subject	n	nprod	natt	nattsel	f	probf	ncan	nrvrep	nrvsubj	BETA
14	32	16	17	2	8.06	0.0001	1	-0.225	1.067	5.805
1	32	16	17	7	7.09	0.0001	4	2.243	0.690	4.926
6	32	16	17	7	6.28	0.0001	3	1.647	0.143	5.025
3	32	16	17	5	5.43	0.0001	2	1.866	2.012	3.239
13	32	16	17	1	3.98	0.0046	1	-0.227	0.940	3.233
8	32	16	17	3	3.84	0.0002	2	0.202	0.447	4.116
4	32	16	17	1	3.30	0.0116	1	-0.872	0.922	3.248
10	32	16	17	3	2.89	0.0025	3	1.769	2.278	3.623
11	32	16	17	5	2.72	0.0022	4	1.134	0.809	2.885
12	32	16	17	4	2.71	0.0026	2	2.187	1.050	4.261
5	32	16	17	3	2.63	0.0049	2	1.779	1.834	4.487
9	32	16	17	1	2.60	0.0335	1	0.169	1.550	4.088
7	32	16	17	6	2.13	0.0153	4	1.396	1.811	5.063
2	32	16	17	3	2.12	0.0214	1	-0.065	1.534	5.702

Le tableau CAP multivarié présente un ensemble d'indices multidimensionnels pour chaque sujet du panel. Ces différents indices visent pour la plupart à approcher la complexité de la perception sensorielle du sujet.

Subject

La première colonne contient le nom des sujets de l'étude. Ils sont classés par ordre décroissant de leur F de MANOVA (colonne "F").

n est le nombre de dégustations effectuées par le sujet lors de cette étude. Lorsque le plan est équilibré, cela correspond au nombre de produits multiplié par le nombre de répétitions.

nprod

nprod correspond au nombre de produits dégustés par le sujet. En calculant $n/nprod$, cela permet de retrouver le nombre de répétitions moyen de chaque produit (soit 32/16=2 répétitions de chaque produit dans l'exemple).

natt

natt est le nombre d'attributs évalué pour chaque produit dans le jeu de données.

nattsel

La colonne nattsel indique le nombre d'attributs pour lesquels le sujet était discriminant, sans que l'information ne soit redondante entre les descripteurs sélectionnés. Ainsi, si les différences perçues par le sujet entre les produits sont les mêmes (et significatives) pour deux descripteurs, seul un des deux descripteur sera retenu. Le choix des variables est obtenue à partir d'un algorithme qui sélectionne les variables discriminantes (au seuil de 15%) une à une, à condition qu'elle apporte une information supplémentaire sur les différences entre produits. A chaque sélection d'un nouveau descripteur, l'algorithme vérifie que les descripteurs précédemment sélectionnés apportent encore une information pertinente sur les différences entre produits. Dans le cas contraire, certains descripteurs peuvent être exclus malgré leur sélection dans un premier temps.

Dans le cas extrême où nattsel=1, cela signifie que pour tous les descripteurs pour lesquels le sujet a fait des différences significatives entre les produits, ceux-ci sont classés dans le même ordre. Un seul descripteur aurait donc pu suffire à ce sujet pour qu'il exprime toutes les différences perçues. Lorsque nattsel augmente, cela signifie qu'il faut retenir davantage de descripteurs pour ce sujet afin d'appréhender toutes les différences qu'il a pu faire.

F, probf

F est le F de MANOVA de l'effet produit. Comme le Fproduit en ANOVA, cette valeur caractérise l'importance des différences faites par le sujet entre les produits, mais au niveau multidimensionnel, c'est-à-dire dans l'espace des descripteurs. Le modèle utilisé peut s'écrire :

Descripteurs = sujet + produit + sujet*produit (avec sujet en effet aléatoire)

Dans le cas où le jeu de données ne contient pas de répétitions, l'interaction n'intervient plus dans le modèle. La probabilité associée (probf) indique le niveau de significativité de la valeur de F.

ncan

ncan correspond au nombre de variables canoniques significatives parmi les variables retenues après sélection (valeur de nattsel). Ncan ne peut donc pas être supérieur à nattsel. Les variables canoniques sélectionnées (au seuil de 10%) sont les dimensions de l'espace des descripteurs pour lesquelles le sujet maximise les différences entre les produits (en tenant compte des répétitions). Plus ncan est grand, plus le sujet discrimine les produits sur un grand nombre de dimensions sensorielles. Cette valeur peut donc s'interpréter comme étant un indice de la complexité de l'espace sensoriel pour le sujet. Néanmoins, cette valeur ne doit être utilisée qu'à titre comparatif entre les sujets et non considéré comme le nombre de dimensions sensorielles qu'est capable de percevoir le sujet.

NRVrep, NRVsubj

La notion de NRV est la même que celle du tableau CONF-SCOR-CORR. Il s'agit de la valeur du RV Normalisée (donc comparable à une loi normale). Le RV donne une proximité entre la configuration des objets dans un espace et celle des mêmes objets dans un autre espace. Pour le NRVrep, on s'intéresse à la proximité de la configuration des produits entre les différentes répétitions. Plus le NRVrep est grand, plus la configuration des produits est similaire entre les répétitions donc plus le sujet est répétable d'un point de vue multidimensionnel. A l'inverse plus le NRVrep est petit, plus les configurations sont différentes donc le sujet est peu répétable. Le NRVsubj est strictement identique au NRV de la table CONF-SCOR-CORR. Il s'agit d'un indice d'accord du sujet avec le reste du panel. Une grande valeur du NRVsubj indique que la configuration moyenne des produits de ce sujet est similaire à celle des autres sujets.

Beta

Calculer le coefficient Beta revient à faire une ACP des données du sujet (moyennées sur les répétitions) puis à regarder combien de dimensions apportent une information pertinente. Cela revient à examiner l'histogramme des valeurs propres pour savoir combien faut-il garder de dimensions. Au minimum, beta=1 ; cela correspond à la situation où toute l'information se résume par un seul axe. Au maximum, beta est la plus petite valeur entre le nombre de descripteurs et le nombre de produits - 1 ; cela correspond au cas où toutes les dimensions portent autant d'information. Par rapport au ncan, beta donne plus une suggestion sur le nombre d'axes à retenir sur les données moyennes du sujet (moyenne des répétitions) alors que le ncan donne la dimensionnalité de la discrimination qui tient compte des répétitions. Par conséquent, dans la plupart des cas, beta est plus grand que ncan car le fait d'oublier la variabilité autour de chaque produit à tendance à augmenter le nombre de dimensions sur lesquelles le pourcentage de variance expliquée n'est pas négligeable. Mais le cas contraire (beta-ncan) est malgré tout possible.

Table Flash

La table FLASH permet de constituer des **groupes de variables (descripteurs sensoriels) fortement corrélés**. A l'intérieur d'un groupe de variables, les produits ayant de faibles (respectivement fortes) notes sont donc le plus souvent les mêmes pour tous les descripteurs de ce groupe. Pour chaque descripteur (i.e. pour chaque ligne du tableau), les produits les plus éloignés de la moyenne globale pour ce descripteur sont surlignés :

- en **jaune** pour les produits dont la moyenne **est inférieure à la moyenne globale**
- en **vert** pour les produits dont la moyenne **est supérieure à la moyenne globale**

ATTRIBUTE	F	PROB	gmean	P7	P8	P1	P3	P5	P6	P11	P4	P10	P14	P16	P2	P13	P9	P12	P15
PASTRY	4.43	0.00	19.2	-4.25	-9.84	-0.1	-11.1	-11.7	-11.7	15.6	17.9	20.4	20.9	22.5	22.4	-27.1	27.9	-28.1	-43.8
SWEET	3.40	0.00	42.4	-22	-29	-27.6	30.1	-32.6	-31.5	42.9	37.8	-58.3	47.9	33.9	37	-63.6	55.5	-59.4	-69.4
NUTS	1.64	0.07	16.3	12.9	-8.75	12.8	14.1	14.2	-9.75	16.9	14.8	19.1	-27.6	17.5	19.3	11.5	22.5	15.4	-23.4
BUTTER	1.74	0.07	16.8	14.2	11.9	11.9	10.6	-22.3	12.7	11.4	20.5	17.5	4.54	22.5	18.4	19.3	22	16.8	-24.8
CHESNUT	6.07	0.00	41.1	37.1	-20.9	33.6	37.7	-32.9	35.6	-29.5	37.3	42.6	-6.8	43.3	-29	-42.2	-63.9	47	-67.2
CARBOAR	1.45	1.26	8.07	14.4	-14.3	6.54	10.8	4.88	11.2	7.88	7.43	7.21	11.8	-1.75	5.75	6.88	8.21	4.54	5.88
MUSHROOM	1.97	1.12	12	18.4	14.3	15.6	14.1	11.5	15.4	14.6	15.5	7.97	12.9	11.6	10.4	4.5	9.86	7.71	-5.43
CEREAL	2.33	0.04	26.3	23.4	30.8	-16.6	21.4	26.3	31.9	22.1	23.3	-18.9	-23.8	18.4	30.7	28.9	29.4	27.2	23.3
CELERY	1.88	0.29	18.9	16.5	14.5	17.9	-11	14.4	14.8	20.3	12.8	-23.2	11.7	18.5	13	15	23.8	21.5	15.6
BITTER	2.88	0.00	28.9	16.5	18.4	30.7	28.8	30.6	28.3	24.7	28.9	24.8	-19.8	22.9	18.2	-17.5	24.2	20	32.2
HERBACE	3.24	0.00	30.5	14.5	36.3	38.4	31.3	31.9	35.8	-48.9	33.9	32.9	-13.3	24.5	-19.9	28.2	23.3	28.1	-20.3
RAMPOUTAT	3.81	0.01	22.7	28.2	18.8	19.6	19.6	22.3	24.9	25.8	18.9	17.1	18.5	20.9	22	-15.3	14.5	25	-8.86

Attribut

Cette colonne contient les noms des descripteurs utilisés dans l'étude, classés par groupes. Ces groupes sont définis en utilisant la procédure VARCLUS (option centroid) de SAS®. Il s'agit d'un processus itératif qui forme des groupes de manière à maximiser la somme des variances expliquées par la première composante principale de chacun des groupes (cf. paragraphe sur VARCLUS pour plus d'explications, à la suite des explications sur la table Flash). Chaque groupe est donc formé des variables les plus corrélées entre elles. Le nombre de groupes est déterminé par le pourcentage de variance totale expliquée. Lorsque la variance expliquée par n groupes dépasse 50% (valeur fixée arbitrairement) de la variance totale, alors le nombre de groupes est fixé à n.

Ces groupes sont ensuite ordonnés. Le premier groupe est celui dont la variance expliquée est la plus forte. Les groupes suivants sont classés par ordre décroissant de corrélation avec le premier groupe. Ainsi, le deuxième groupe est celui qui présente le plus de similarités avec le premier. Cela signifie que l'ordonnement des produits du deuxième groupe est proche de celui du premier groupe. A l'opposé, le dernier groupe est celui qui a la corrélation la plus faible avec le premier. L'ordonnement de ses produits est donc à l'opposé de celui du premier groupe.

A l'intérieur de chaque groupe, l'attribut placé en première position est celui dont la corrélation à la moyenne du groupe est la plus élevée. De cette manière, le premier descripteur de chaque groupe est sensé être le plus représentatif du groupe. Les descripteurs suivants sont classés par ordre de corrélation décroissante au premier descripteur afin de regrouper au maximum les descripteurs dont la notation des produits est proche.

F Cette colonne donne la valeur du F de l'effet produit dans l'anova : attribut = produit + sujet. Plus cette valeur est importante, plus le groupe est discriminant pour ce descripteur.

Rq : au sein d'un groupe, le descripteur avec le F le plus élevé n'est pas forcément le descripteur le plus représentatif du groupe au sens de la corrélation à la moyenne du groupe. Par conséquent, les descripteurs avec les F les plus élevés ne sont pas forcément les premiers de leur groupe. Néanmoins, s'il ne fallait retenir qu'une variable par groupe, cela n'aurait pas de sens de garder une variable dont le F n'est pas significatif. Dans la plupart des cas, il serait donc préférable de garder la variable pour laquelle les panélistes sont le plus discriminant, plutôt que de garder la première variable de chaque groupe.

PROB

Prob est la probabilité associée à F, c'est-à-dire le risque de déclarer à tort qu'il existe des différences entre les produits pour un descripteur donné. Si prob<0.05, alors la valeur du F est sur fond vert. Si

0.05<prob<0.15, alors la valeur du F est sur fond orange. Si prob>0.15, alors la valeur du F est sur fond rouge.

gmean

La valeur de gmean correspond à la moyenne des notes données par le groupe, tous produits confondus, pour le descripteur considéré. Cela permet de détecter les descripteurs pour lesquels les notes données par le panel sont systématiquement faibles, par exemple. Cette valeur est aussi à la base de la comparaison des moyennes des produits (paragraphe suivant).

Moyennes des produits

Le reste du tableau est constitué d'autant de colonnes que de produits. Chaque cellule contient la moyenne des notes données par le panel pour le produit et le descripteur considéré. Les produits sont classés par ordre croissant de leur moyenne pour le premier descripteur du premier groupe. Lorsque la moyenne d'un produit est significativement supérieure à gmean (test de contraste dont le seuil est fixé à 10%), elle est précédée d'un "+" et est écrite sur fond vert. Lorsque la moyenne d'un produit est significativement inférieure à gmean, elle est précédée d'un "-" et est écrite sur fond jaune. Lorsque la moyenne d'un produit n'est pas significativement différente de gmean, elle est écrite sur fond gris.

Precisions sur l'algorithme VARCLUS

Varclus est une procédure de SAS® qui permet de réaliser une classification de variables. Pour cela, une première ACP de l'ensemble des variables est réalisée. Les deux premiers groupes sont constitués des variables corrélées positivement à la première composante principale (PC1) de l'ACP pour un groupe, et des variables corrélées négativement à PC1 pour l'autre groupe. Une ACP de chaque groupe de variables est alors réalisée. Le pourcentage de variance expliquée par la PC1 d'un groupe de variables caractérise donc la force ou l'importance des corrélations entre les variables de ce groupe. Cette variance expliquée (par la PC1 d'un groupe) est ensuite rapportée au nombre total de variables dans le jeu de données (et non pas au nombre de variables dans le groupe uniquement). Ce pourcentage de variance "totale" expliquée par (la PC1 de) chaque groupe est ainsi comparable aux autres groupes. C'est cette valeur que l'on cherche à maximiser, ou plutôt la somme de ces valeurs pour l'ensemble des groupes.

A ce stade, l'algorithme autorise certaines variables à changer de groupe si cela peut permettre d'augmenter le pourcentage de variance totale expliquée. Lorsque la composition des groupes est stable, pour un nombre de groupes donné, n, le passage à n+1 groupes va commencer en prenant le groupe dont la variance expliquée est la plus faible afin de le diviser en deux nouveaux groupes de variables. Le même processus d'optimisation des groupes est alors utilisé.

Particularité des données sans répétitions : signification et interprétation des résultats proposés

Table CAPSR

La table CAPSR permet de rendre compte des performances d'un panel d'analyse sensorielle ayant réalisé le profil descriptif d'un ensemble de produits sur un certain nombre de descripteurs sans répétition. Cette table se présente de la manière suivante :

attribut	fprod	mean	_4	_11	_13	_2	_1	_3	_6	_10	_9	_14	_12	_5	_7	_8
EARTHLY	4.25	29.3	+	+	+	-	-	-	-	+	+	+	+	+	+	+
CHESSNUT	3.78	41.7	-	+	-	+	-	+	-	+	+	+	+	+	+	+
PASTRY	3.47	19.2	+	+	+	+	-	-	-	+	+	+	+	+	+	-
HERBACE	3.20	30.8	-	+	+	+	+	+	+	+	-	+	-	+	+	-
RAWPOTAT	2.95	24.0	-	+	+	+	+	-	-	+	+	+	+	+	+	-
CEREAL	1.97	27.4	-	-	+	+	-	-	+	+	+	+	+	+	+	-
ARTICHOK	1.64	38.5	+	+	+	-	-	-	-	+	+	+	+	+	+	-
BUTTER	1.61	17.2	-	+	+	-	-	-	-	+	+	+	+	+	+	+
CELERY	1.44	16.9	-	-	-	-	-	+	-	+	+	+	+	+	+	+
NUTS	1.15	16.3	=	-	-	+	-	-	-	=	=	-	+	-	+	-
CARDBOAR	1.11	7.70	-	-	-	+	+	-	-	=	=	-	+	-	+	+
STRAW	1.03	29.4	+	+	+	-	+	-	-	-	-	-	-	-	-	-
TOASTEDB	0.77	8.82	=	-	-	-	-	-	-	=	-	+	+	-	-	-
MUSHROO	0.73	12.0	-	-	-	-	-	-	-	-	-	-	-	-	+	+
rankvar	-	-	3.36	4.36	4.57	4.79	4.86	6.88	7.43	7.50	8.50	8.57	9.38	10.5	12.0	12.2

ATTRIBUT

La première colonne contient le nom de chaque attribut ou descripteur sensoriel évalué dans l'étude. Chaque ligne présente les résultats relatifs à un attribut différent.

FPROD

Il s'agit de la statistique du test de Fisher de l'effet produit dans le modèle d'analyse de la variance : sujet + produit.

FPROD mesure, sur l'ensemble du groupe, l'importance des différences entre les moyennes des produits pour chaque descripteur et est relatif à l'accord des sujets quant à ces différences. Si les moyennes des produits sont très dispersées et que tous les sujets s'accordent sur l'ordonnement de ces moyennes, alors FPROD sera grand. Si au contraire les moyennes des produits sont peu dispersées et/ou que les sujets diffèrent sur l'ordonnement de ces moyennes, alors FPROD sera petit.

Cette valeur de FPROD est testée à l'aide de la loi de Fisher afin de pouvoir conclure sur sa significativité. Cela consiste à calculer le risque, noté α (entre 0 et 1), de conclure qu'un effet est significatif à tort. Plus FPROD est grand, plus le risque de conclure qu'il existe un effet à tort sera faible donc α diminue. Plus FPROD est petit, plus le risque α augmente. En analyse sensorielle, le risque maximum classiquement autorisé est de 5%. Dans le cas du tableau CAP, plusieurs seuils ont été utilisés pour moduler les conclusions. Cela donne lieu à l'élaboration d'un code couleur comme suit :

Si $\alpha < 5\%$, l'effet est significatif. Il existe des différences nettes entre les moyennes des notes des produits. Dans le tableau, lorsque le α associé à un FPROD est inférieur à 5%, la case est verte car on observe un effet produit significatif.

Si α est compris entre 5% et 10%, l'effet est de l'ordre de la tendance. FPROD sera beige (légère différence entre les produits).

Si α est compris entre 10% et 20%, il n'y a qu'une légère tendance. FPROD sera orange (faible différence entre les produits).

MEAN

Cette colonne contient la moyenne des notes données à un descripteur sur l'ensemble du panel. Elle permet de voir le niveau moyen de notation sur l'échelle. Une valeur très faible pour un descripteur peut indiquer que la sensation en question est très peu perçue, au moins pour la majorité des produits.

RANKVAR

Les colonnes suivantes correspondent chacune à un des sujets de l'étude. Ils sont classés selon leur RANKVAR, dont la valeur est donnée sur la dernière ligne. Cette valeur représente une moyenne de rangs. Pour chaque descripteur, les sujets sont classés selon leur dispersion des notes des produits sur l'échelle. Cette dispersion est représentée par la variance des notes des produits. Plus la variance est grande, plus le sujet a donné des notes différentes aux produits pour ce descripteur. Le sujet ayant la plus grande variance reçoit le rang 1, le dernier sujet reçoit le dernier rang. Chaque sujet reçoit au total autant de rangs que de descripteurs. RANKVAR correspond à la moyenne de ces rangs pour un sujet. Si un sujet obtient un RANKVAR de 1, c'est qu'il a toujours obtenu la plus grande variance pour tous les descripteurs. Le RANKVAR maximum correspond au nombre de descripteurs. Les sujets sont donc classés par RANKVAR croissants.

DIAGNOSTICS INDIVIDUELS +/-/=

En fonction de la dispersion de ses notes et de son accord avec le groupe, chaque sujet se voit attribuer un diagnostic parmi +, -, et =. Si l'écart-type (racine de la variance) du sujet est inférieur à 1% de l'échelle (0, 1 sur une échelle de notation de 0 à 10 par exemple), ou s'il est très inférieur à la moyenne des écart-types du panel (<1.5 x l'interquartile) le sujet aura le signe =. Il ne fait pas de différence significative entre les produits. Si l'écart-type est supérieur à ces deux seuils et que le sujet est en accord avec le groupe, le sujet aura le signe +. Si l'écart-type est supérieur à ces deux seuils mais en désaccord, le sujet aura le signe -.

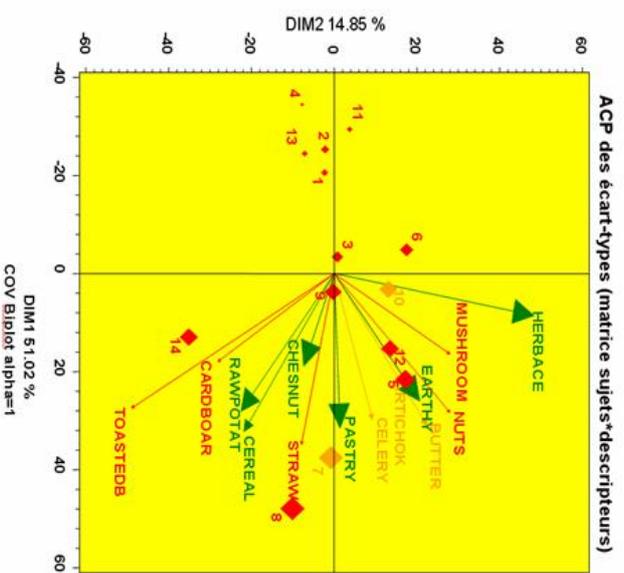
L'accord est calculé en utilisant le coefficient de Kendall. Celui-ci compte les paires de produits pour lesquelles les produits sont classés dans le même ordre entre deux sujets. Par conséquent, plus les produits sont classés dans le même ordre entre les sujets, plus le Kendall est grand. Lorsque ce coefficient est significatif au seuil de 20%, le sujet reçoit le signe + pour ce descripteur. Sinon il reçoit le signe - (sujet en désaccord avec le panel).

Le diagnostic final se trouve donc parmi les 3 cas suivants :

+	Discriminant, en accord
=	Non discriminant
-	Discriminant, en désaccord

CAPMAPSR

La CAPMAPSR est une ACP de la matrice sujets*descripteurs contenant les écart-types (mesure de la dispersion des notes). Elle a pour but de faire apparaître les axes (de l'espace des descripteurs) selon lesquels les différences de dispersion des notes sont maximales entre les sujets.



LES DESCRIPTEURS

Longueur de la flèche

Plus un descripteur présente une flèche longue, plus les écart-types sont différents entre les sujets. Certains sujets ont donc plus tendance à faire des différences entre les produits pour ce descripteur. Plus la flèche est courte, plus le niveau de discrimination est homogène au sein du panel. Cela ne traduit en rien le niveau global de dispersion des notes ! Une flèche courte peut correspondre à un niveau d'écart-type élevé ou faible du panel.

Taille et couleur de la flèche

Lorsque les données ne contiennent pas de répétition, le FPROD de l'anova par descripteur mesure à la fois le pouvoir discriminant du groupe et leur répétibilité. Plus le FPROD est élevé, plus les sujets sont discriminants et en accord. La flèche aura donc une pointe large (pour la discrimination) et verte (pour l'accord). Plus le FPROD est petit, moins le panel est discriminant et en accord. La flèche aura donc une pointe étroite (pour la non discrimination) et rouge (pour le désaccord).

LES SUJETS

Position d'un sujet

La place d'un sujet sur la carte dépend des écart-types qu'il a obtenus pour chaque descripteur. Plus son écart-type sera élevé (par rapport au groupe) pour un descripteur, plus ce sujet sera éloigné du centre de la carte dans la direction de la flèche du descripteur. Plus son écart-type sera faible, plus le sujet se situera dans la direction opposée. Si un sujet obtient un Fprod moyen, il sera proche de zéro.

Taille du point

La taille du point d'un sujet dépend des valeurs des écart-types (et donc des variances) obtenues. La valeur utilisée correspond presque à la moyenne des écart-types. Il s'agit en fait de la moyenne de la racine carrée de la somme des variances. Plus cette valeur est grande, plus le sujet a observé des différences entre les produits, tous descripteurs confondus. Le point sera alors gros.

Couleur du point

La couleur du point représente l'accord ou le désaccord entre le sujet et les autres panélistes. Elle dépend de la valeur du ALPHA (différent du α utilisé dans les probabilités) présenté dans le tableau « Performances multidimensionnel ». La couleur du point sur la carte correspond à celle de ce tableau. Si un sujet a généralement classé les produits dans le même ordre que les autres sujets sur la plupart des descripteurs, son ALPHA sera élevé (proche de 1). Plus le classement de ses produits sera différent des autres sujets et sur un grand nombre de descripteurs, plus ALPHA sera faible (proche de zéro). Le coefficient ALPHA peut ensuite être Normalisé, ce qui permet de lui associer une probabilité de déclarer que le sujet est en accord avec le groupe à tort.

Lorsque la valeur normalisée est supérieure à 2 (correspondant approximativement à un risque de 5% pour la loi normale centrée réduite), le sujet est dit significativement en accord avec le groupe ; le point sera de couleur verte. Si cette valeur est comprise entre 1,5 et 2, le sujet a plutôt tendance à être en accord avec le groupe ; le point sera beige. Si la valeur est entre 1 et 1,5, le sujet est plutôt en léger désaccord avec le groupe ; le point sera orange. En dessous de 1, le sujet est en désaccord avec le panel ; le point sera rouge.

Auteur : Nicolas Pineau, pineau@cesga.cnrs.fr, tel : 03 80 68 16 34

Annexe 3

Fonctionnement du programme SAS[®] %PERF
utilisé dans la SensoBase

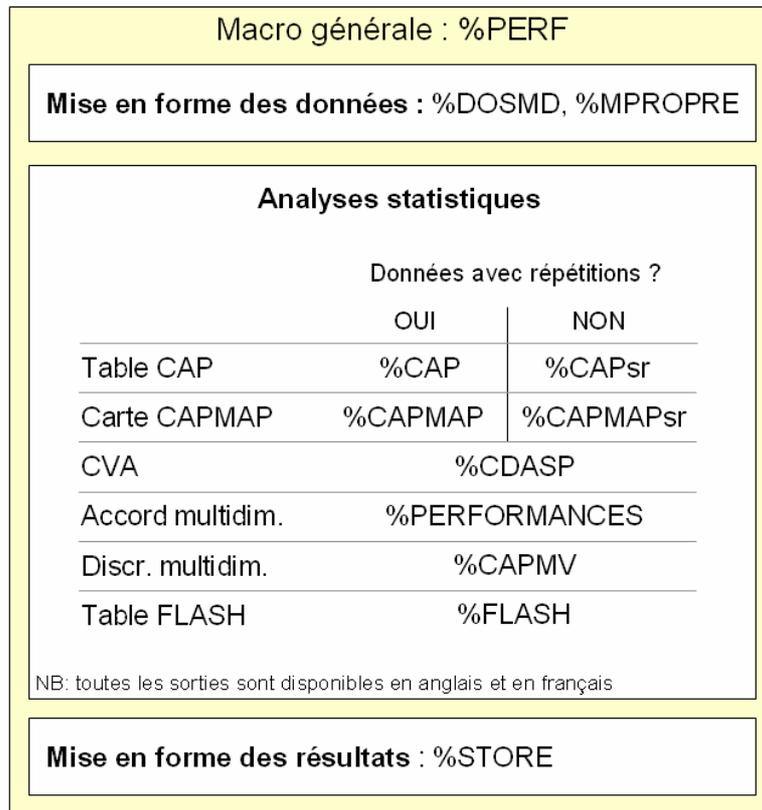
Données

Le tableau de données doit être structuré de la manière suivante :

- une colonne Sujet
- une colonne Produit
- éventuellement une colonne Répétition
- autant de colonnes que de Descripteurs

Structure du programme SAS®

Les principales macros utilisées par %PERF dépendent de la présence de répétitions ou non dans le fichier :



La macro %PERF compte un grand nombre de paramètres. Certains doivent être obligatoirement renseignés, d'autres sont optionnels.

Paramètres de %PERF

```
data=,                nom du dataset
sujet=sujet,          nom de la variable sujet
produit=produit,     nom de la variable produit
rep=,                 nom de la variable rep si elle existe
storelib=,            libname de la librairie de stockage
study=,               nom que l'on souhaite donner à l'étude
NB : les fichiers importants seront stockés dans &storelib\&study_nomdFichier
nomdesc=_numeric_,   liste des variables descripteurs
```

```
/*paramètres de %CAP et %CAPSR*/
pglob=0.05,           seuil pour effets sujet et prod des anova par descr
preliab=0.01,         seuil pour wmsdr>wmsdr moyen ( ! ou non)
pind=0.10,            seuil pour anova par sujet descripteur (+/- ou =)
pkendall=0.20,        seuil pour accord du sujet avec le groupe (+ ou -)
fmean=4.2,            format de la variable des F produit par descripteur
```

```
/*paramètres de %performances*/
centrage=Oui,         les données doivent-elles être centrées ?
nbperm=,              nb de perm pour Statis et Gamma
```

<pre> nbpermCOL=150, HTML=oui, statis=oui, gamma=oui, dual=oui, missvalue=non, allperm=non, allpermCOL=non, stdperf=CORR, outperf=mdperf, cheminHTML='c:\Temp', /*param de l'acp des F pprod=0.05, pinter=0.05, /*param du biplot*/ stdbiplot=COV, alpha=1, coefbiplot=0, weight=, varbiplot=&nomdesc, supvar=, supcoef=1, nonsup=Y, poly=, inc=, id=, arna=A, angle=, labelbiplot=No, size1=0.75, size2=0.75, size3=0.75, cbcoefvar=1, cbcoefind=1, style=swiss, stylev=swiss1, device=CGMMW6C, xmax=15, ymax=21, print=YES, plot=YES, display=NO, title=, footnote=, gsf=NO, star=YES, cgm=, make=NO, class=, partial=, dim=2, gout=work.perf, name=, color=black blue red g out=biplott, ods= /*param de la CVA*/ CVAprb=0.15, size4=2, ellconf=0.90 </pre>	<pre> nb de perm pour Dual sortie HTML ? run STATIS ? run GAMMA ? run STATIS DUAL ? le fichier contient des valeurs manquantes ? réaliser toutes les permutations possible (si n<7) Idem pour Statis Dual CORRélation ou COVariance nom du dataset de sortie chemin pour stockage du fichier Html prod*/ couleur des flêches : pp<&pprod => vert, pp>&pprod : (pi>&pinter => orange, pi<&pinter => rouge) CORRélation ou COVariance 1: distance respectées pour dilater les variables / individus variable des poids variables des descripteurs du jeu de données variables supplémentaires coefficient des variables supplémentaires angle des labels des descripteurs label des biplot ou nom des variables ? taille du titre taille de l'échelle taille des noms sur le graphique taille max des x des graphs (en cm) taille max des y des graphs (en cm) sorties dans l'ouput ? sortie des graphs Texte du titre Texte de la footnote étoiles pour les points ? créer un cgm ? nb de dimensions de l'ACP catalogue de sortie des graphs nom du graphique reen pink blue yellow, couleur de : text obs var ellipses back star frame dataset des infos pour construire le biplot dataset de sortie valeur de la proba pour la sélection de variables épaisseur des ellipses intervalle de confiance des ellipses </pre>
--	--

Sorties

Les sorties se présentent sous plusieurs formes. Tous les fichiers importants sont stockés sous SAS dans un répertoire. De plus, les tableaux de résultats sont également produits en html et les figures en GIF et en CGM de manière à pouvoir accéder à toute l'information via une page web. Ces fichiers s'organisent de la manière suivante.

Deux fichiers de mise en page, non produits par la macro et standards pour toutes les analyses :

- **index.html** : page contenant deux frames (le sommaire/index à gauche et les tableaux de résultats à droite. **Il s'agit de l'unique fichier à ouvrir pour accéder à l'intégralité des résultats**)
- **sommaire.html** : page contenant les titres des différentes analyses. En cliquant sur chaque titre, l'analyse correspondante s'ouvre dans la page de droite.

Les fichiers de résultats :

- **captable.html** : contient de tableau CAP édité (couleurs)
- **capmap.gif** : ACP des fprod (ou des variances quand il n'y a pas de répétitions)
- **cva12.gif** : CVA du jeu de données
- **indicesperf.html** : accord multidimensionnel (RV, RVD, ALPHA)
- **perf.html** : contient différents résultats sur les données
- **capstd.html** : diagnostique par descripteur (non inclus dans le fichier global et produit que lorsque le jeu de données contient des répétitions)

Annexe 4

Manuel d'utilisation du fichier SensoBase

MANUEL D'UTILISATION DU FICHIER SENSOBASE

Un jeu de données doit donner lieu à la création d'un fichier Excel ou d'un fichier CSV.

1 jeu de données = 1 étude sensorielle (qui peut avoir été réalisée en plusieurs jours). Si vos données contiennent plusieurs répétitions, elles doivent toutes faire partie du même jeu de données.

Exemple : Profil descriptif sur 4 yaourts de marques différentes sur 2 séances d'évaluation sensorielle (2 sessions).

* Etape 1 : Recommandations

Lisez ce document en entier avant de commencer votre saisie. Les informations de la page « Remarques générales » pourront vous être utiles.

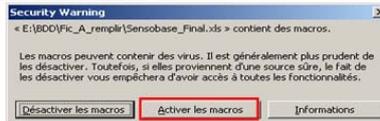
Commencez par **enregistrer ce fichier sous un autre nom** afin de garder une copie toujours vierge du fichier de départ.

En cas de perte du fichier, vous pouvez le télécharger sur le site liris.cesg.cnrs.fr sur votre espace personnel (s'identifier) ou le demander

par mail à SensoBase@cesg.cnrs.fr

Le niveau de sécurité d'EXCEL vis-à-vis des macros ne doit pas être élevé ou très élevé sinon les macros contenues dans ce fichier ne pourront pas être exécutées.

A l'ouverture du fichier EXCEL, vous pouvez peut être apparaître le message suivant :

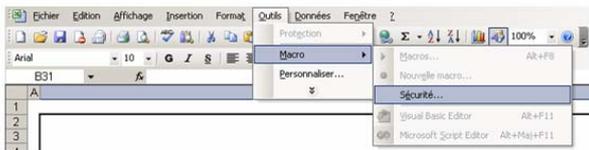


Cliquez sur **Activer les macros**

Si non, vous verrez apparaître le message suivant :



Pour diminuer manuellement le niveau de sécurité, cliquez sur le menu **Outil > Macros > Sécurité ...**



Sélectionner le niveau de sécurité **Moyen** puis cliquer sur **Ok**.



1

* Etape 2 : Informations sur le fournisseur de données, les sujets et leurs habitudes alimentaires, les produits, les descripteurs utilisés et le protocole du jeu de données

1. Saisie des informations

DESCRIPTION DES DONNEES		Page Précédente		Vérifier la saisie et passer à l'étape suivante	
FOURNISSEUR					
Nom de la Société / login		CESG		obligatoire facultatif	
SUJETS				Autant de colonnes que de sujets Ici : 6 sujets	
Code sujet	Est-ce une Clé ?	S1	S2	S3	S4
Date de naissance	Sexe	12/01/1960	21/10/1990	31/01/1920	31/01/1908
Pays de naissance	Autre pays de naissance	F	M	F	F
Nationalité	Autre nationalité				
Etat civil	Pays de domicile	1	1	1	1
Autre pays de domicile	Département				
Niveau d'études	Situation professionnelle	52	39	21	77
					28
					52

Commentaires (S'affichent en survolant la cellule)

Si vous ne connaissez que l'année de naissance du sujet, respectez tout de même le format JJ/MM/AAAA. (Ex ici : 1979, saisissez 01/01/1979)

2

2. Vérifications

	A	B	C	D	E	F	G	H	I	
1	DESCRIPTION DES DONNEES				Page Précédente	Vérifier la saisie et passer à l'étape suivante				
134		Prof : participation antérieure ?								
135		Prof : nombre d'études								
136		Passé								
137	Habitudes vis à vis du produit	Consommation du type de produit de l'étude								
138		Approvisionnement								
139										
140	PRODUITS									
141		Code produit		AF1	AF2	KG	LX	TgT		
142		Est-ce une Clé ?		0	0	0	0	0	0	
143		Famille d'aliments		121	121	121	121	121	121	
144		Produit du commerce ?								
145		Description détaillée								
146		Marque commerciale								
147		Spécificités liées à l'étude								
148		Taille échantillon		10	10	10	10	10	10	
149		Unité de taille		ml	ml	ml	ml	ml	ml	
150		Quantité à consommer		1	1	1	1	1	1	
151		Autre quantité								
152		Date de fabrication								
153		Date de récolte								
154		Date de mise en bouteille / conditionnement								
155		Date limite de consommation								
156		Date de congélation								
157										
158										
159	DESCRIPTEURS									
160		Code descripteur		Sweet	Bitter	AR2	Tx1	Tx2		
161		Est-ce une Clé ?		0	0	0	0	0	0	
162		Libellé du descripteur		sweet	Bitter	AR2	Tx1	Tx2		
163		Type de descripteur		3	3	3	4	4	4	
164		Valeur minimale de l'échelle		0	0	0	0	0	0	
165		Valeur maximale de l'échelle		10	10	10	10	10	10	
166		Type d'échelle		3	3	3	3	3	3	
167		Graduations								
168		Echelle discrète : précision								
169		Profil : définition du descripteur								
170		Profil : référence de la définition								
171										

Vérifier la cohérence des informations saisies et passer à l'étape suivante (saisie des données du profil)

Corrections / modifications

Erreur(s) de saisie

Journal des Erreurs

2 erreurs

Cellule :#E20 Le sexe doit être M ou F

Cellule :#H26 La valeur doit être 0 ou 1

OK

Tout va bien

Récapitulatif

5 produits : AF1 / AF2 / KG / LX / TgT

6 sujets : S1 / S2 / S3 / S4 / S5 / S6

6 descripteurs : Sweet / Bitter / Leonon / AR2 / Tx1 / Tx2

OK Annuler

Etape 3

Etape 3 : Saisie du tableau de données

	A	B	C	D	E	F	G	H	I	J	K
	Code sujet	Code produit	Session	Répétition	Ordre	Sweet	Bitter	AR2	Tx1	Tx2	
2	S1	AF1	1	1	1	7,9	2,9	1,2	12	1,2	
3	S1	AF2	1	1	2	1,5	9,8	2,9	8,8	2,9	
4	S1	KG	1	1	3	9,9	1,2	3,2	6,4	3,2	
5	S1	LX	1	1	4	7,7	2,9	0,3	6,8	0,3	
6	S1	TgT	1	1	5	1,9	3,2	6,4	5,1	0,3	
7	S1	AF1	1	2	6	5,1	0,3	6,8	0,3	8,3	
8	S2	AF1	1	1	6	0,3	8,3	3,4	2,9	5,4	
9	S2	AF2	1	1	3	2,9	5,4	9,8	6,4	5,3	
10	S2	KG	1	1	2	6,4	5,3	2,6	1,9	4,8	
11	S2	LX	1	1	4	1,9	4,8	1,7	5,9	4,9	
12	S2	TgT	1	1	1	5,9	4,9	4	0,9	5,1	
13	S2	AF2	1	2	5	0,9	5,1	0,8	2,4	2,4	
14	S3	AF1	1	1	3	2,4	2,4	3,2	0,1	5,6	
15	S3	AF2	1	1	1	0,1	5,6	4	5,8	7,5	
16	S3	KG	1	1	2	5,8	7,5	1,1	1,9	4,8	
17	S3	LX	1	1	5	1,5	3,6	6,9	5,9	4,9	
18	S3	TgT	1	1	4	1	8,4	2	0,9	5,1	
19	S3	KG	1	2	6	7,3	5,4	2,5	2,4	2,4	
20	S4	AF1	1	1	5	1	3,1	9,5	0,1	5,6	
21	S4	AF2	1	1	4	4,5	3,4	4,9	4	0,9	
22	S4	KG	1	1	1	4,1	0,6	5,1	0,8	2,4	
23	S4	LX	1	1	3	0,4	6,9	2,4	3,2	0,1	
24	S4	TgT	1	1	6	6,5	3,3	5,6	4	5,8	
25	S4	LX	1	2	2	4,9	4	7,5	1,1	1,9	
26	S5	AF1	1	1	2	5,1	0,8	9,6	6,9	5,9	
27	S5	AF2	1	1	6	2,4	3,2	8,4	2	0,9	
28	S5	KG	1	1	3	5,6	4	5,8	7,5	0,1	
29	S5	LX	1	1	5	7,5	1,1	1,9	4,8	5,8	
30	S5	TgT	1	1	1	9,6	6,9	5,9	4,9	1,9	
31	S5	TgT	1	2	4	8,4	2	0,9	5,1	5,9	
32											

Remarque : Les données sont impérativement à saisir dans la plage bleue sans ligne vide dans celle-ci (ni dans les données, ni entre la ligne d'entête et les données).

Identifiants des sujets

Identifiants des produits

Notes sur les différents descripteurs

Ordre : Les sujets évaluent des produits dans un certain ordre, l'ordre doit faire partie d'une session mais ne doit pas obligatoirement être indiqué.

Numéro de répétition : Un ou plusieurs sujet évalue(nt) plusieurs fois le même produit (si pas de répétition → 1 pour toute la colonne)

Numéro de session : Les sujets évaluent les produits au cours de plusieurs séances d'évaluation sensorielle (dates différentes) (si toutes les notations proviennent de la même séance → 1 pour toute la colonne session)

Etape 4

*** Etape 4 : Tableau croisé Produits * Sujets & enregistrement au format CSV**

Ce tableau est un moyen de vérifier les données. Dans cet exemple, le sujet « S1 » a évalué 2 fois le produit « AF1 ».

Etape 3 Génération du fichier CSV

	A	B	C	D	E	F	G	H	I
1									
2		Retour aux Données							
3									
4			Enregistrer votre fichier au format CSV (= fichier à envoyer)						
5									
6									
7	Count of Code sujet	Code produit							
8	Code sujet	AF1	AF2	KG	LX	TgT	(blank)	Grand Total	
9	S1	2	1	1	1	1	1	6	
10	S2	1	2	1	1	1	1	6	
11	S3	1	1	2	1	1	1	6	
12	S4	1	1	1	2	1	1	6	
13	S5	1	1	1	1	2	1	6	
14	(blank)								
15	Grand Total	6	6	6	6	6	6	30	
16									
17									

D'autre part, chaque produit a été évalué 6 fois sur l'ensemble du groupe

La saisie du fichier SENSOBASE est maintenant terminée. Vous devez sauvegarder votre travail au format CSV à l'aide du bouton d'enregistrement correspondant. C'est ce fichier que vous devrez ensuite envoyer via votre compte personnel sur le site internet (<https://liris.cesg.cnrs.fr>) pour recevoir ensuite les analyses statistiques.

Cependant, il est préférable de sauvegarder vos données au format XLS également. En effet, si vous désirez effectuer une modification dans les données avant de les envoyer, vous ne devez pas le faire dans le fichier CSV (il serait alors refusé par le site SensoBase). Vous devez ouvrir le fichier XLS, faire les modifications et générer à nouveau le fichier CSV.

Enregistrement au format CSV

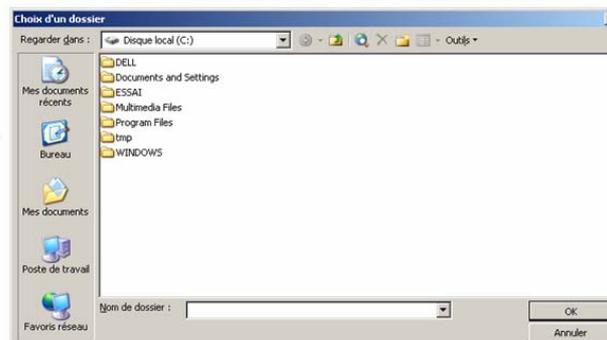
Lors du clic sur le bouton d'enregistrement, une **fenêtre de confirmation** s'affiche. Confirmez afin de continuer l'enregistrement.



Vous devez définir le **nom de votre fichier** à enregistrer. Choisissez un nom clair qui vous permettra facilement d'identifier votre jeu de données lorsqu'il sera inséré dans la SensoBase (ex : Vin_jdd12, etudeX_22dec2006, ...)



Ensuite il suffit de choisir l'**emplacement du fichier** à enregistrer. Vous devez choisir le répertoire destinataire et confirmer.



Remarques générales concernant le fichier EXCEL :

- * A tout moment, vous pouvez enregistrer votre travail afin de le reprendre plus tard comme un fichier Excel classique. Il est d'ailleurs conseillé d'enregistrer régulièrement le fichier (Fichier → Enregistrer)
- * Il est interdit de renommer ou ajouter une feuille (ou onglet) : Toutes les informations doivent être renseignées dans les seules feuilles proposées
- * Vous pouvez faire des copier/coller dans tout le fichier mais pas de couper/coller.

Étape 2 :

- * **Attention à la saisie des dates, respectez bien le format JJ/MM/AAAA ou MM/AAAA.**
- * **Il faut obligatoirement saisir les descripteurs dans l'ordre où ils seront placés dans le tableau de données.**
- * En règle générale, ne laissez pas de colonnes ou de lignes vides (dans l'étape 2 comme dans l'étape 3) qui risqueraient de compromettre la vérification automatique des données.
- * Certaines cellules sont protégées, vous ne pouvez pas saisir de données dans celles-ci. Vous verrez le message suivant apparaître :



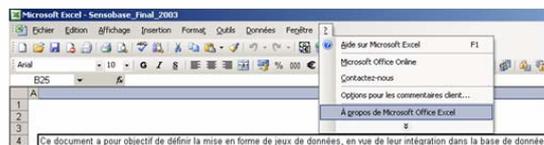
- * Si vous découvrez un BUG dans le programme, comme celui-ci, enregistrez votre travail et fermez EXCEL. Puis, essayez d'ouvrir votre fichier à nouveau. Si le problème persiste, n'hésitez pas à nous téléphoner au 03.80.68.16.05 (Vous pouvez également envoyer par mail le contenu du message d'erreur à SensoBase@cesg.cnrs.fr)



Si vous rencontrez un problème avec ce fichier ou si vous avez des suggestions pour l'améliorer, n'hésitez pas à nous contacter au 03.80.68.16.05 ou par mail : SensoBase@cesg.cnrs.fr.

Versions d'EXCEL antérieures à EXCEL 2003

Pour connaître votre version d'Excel, Cliquez sur ? → A propos de Microsoft Excel



• EXCEL 2002 :

La saisie des noms des descripteurs ne se fait pas automatiquement, vous devez saisir vous-même les noms des descripteurs. L'ordre des descripteurs n'a pas d'importance. En revanche, les libellés doivent être **exactement** les mêmes que ceux définis dans l'étape 2).

• EXCEL 2000 :

Mêmes remarques que pour EXCEL 2002

- * Les cases interdites ne sont pas protégées, veuillez à bien remplir les cellules comme indiqué dans ce manuel.

Dans l'étape 2 :

- Ne pas modifier les cellules des colonnes A, B et C
- Ne pas insérer ou supprimer de lignes ou de colonnes

Dans l'étape 3 :

- Ne pas saisir de données en dehors de la plage colorée en bleu clair

- * Dans l'étape 3, il n'y a pas de bouton « Tableau Sujets*Produits » pour vérifier le contenu des données.

A la fin de l'étape 3, vous pouvez seulement enregistrer et fermer le fichier.

Annexe 5

Liste des items du fichier Excel[®] SensoBase

SUJET

Code permettant d'identifier chaque sujet au sein du jeu de données.

(Un code par sujet)

Chaque sujet est constitué d'une colonne

Est-ce une Clé ?

Clé=code qui fera toujours référence au même sujet entre les différents jeux de données

OUI=1

NON=0

Si un déjudestateur reçoit toujours le même Code sujet (ligne précédente) entre les différents jeux de données que vous pourrez nous envoyer

et que ce code ne sera jamais utilisé pour identifier un autre déjudestateur, alors ce Code sujet est une clé.

Si d'un jeu de données à l'autre un Code sujet peut correspondre à deux personnes différentes, ce code sujet n'est pas une clé.

Format :J/JM/M/AAAA ou MM/AAAA

AAAA=année

MM=mois

J=jour

J/JM/M/AAAA=jour/mois/année

MM/AAAA=mois/année

AAAA=année

Nombre de personnes <18ans

Nombre de personnes de moins de 18 ans vivant sous le même toit que le sujet, y compris le sujet lui-même

Si le sujet est étudiant et partage son temps entre la résidence de ses parents et sa résidence d'étudiant, considérer le foyer parental.

1=1

2=2

3=3

4=4

5=5 ou plus

6=le sujet habite dans une résidence collective (maison de retraite, foyer de travailleurs...)

Niveau net moy(en) (avant impôts) de la totalité des ressources mensuelles du foyer du sujet, en tenant compte des allocations familiales.

Si le sujet est étudiant et partage son temps entre la résidence de ses parents et sa résidence d'étudiant, considérer le foyer parental.

1=moins de 750 euros

2=de 750 à 1500 euros

3=de 1500 à 2000 euros

4=de 2000 à 3000 euros

5=de 3000 à 4000 euros

6=de 4000 à 5000 euros

7=plus de 5000 euros

8=ne sait pas

9=Combien le sujet a d'enfants (en comptant aussi les enfants qui ont quitté le domicile)

0=0

1=1

2=2

3=3

4=4

5=5 ou plus

(en règle générale)

1=au domicile

2=sur le lieu de travail

3=à la cantine / restaurant d'entreprise

4=dans un fast-food

5=dans un restaurant

6=dans un café, une cafétéria

7=dans un autre lieu

(en règle générale)

1=au domicile

2=sur le lieu de travail

3=à la cantine / restaurant d'entreprise

4=dans un fast-food

5=dans un restaurant

6=dans un café, une cafétéria

7=dans un autre lieu

(en règle générale)

1=au domicile

2=sur le lieu de travail

3=à la cantine / restaurant d'entreprise

4=dans un fast-food

5=dans un restaurant

6=dans un café, une cafétéria

7=dans un autre lieu

(en règle générale)

1=au domicile

2=sur le lieu de travail

3=à la cantine / restaurant d'entreprise

4=dans un fast-food

5=dans un restaurant

6=dans un café, une cafétéria

7=dans un autre lieu

(en règle générale)

1=au domicile

2=sur le lieu de travail

3=à la cantine / restaurant d'entreprise

4=dans un fast-food

5=dans un restaurant

6=dans un café, une cafétéria

7=dans un autre lieu

(en règle générale)

1=au domicile

2=sur le lieu de travail

3=à la cantine / restaurant d'entreprise

4=dans un fast-food

5=dans un restaurant

6=dans un café, une cafétéria

7=dans un autre lieu

(en règle générale)

1=au domicile

2=sur le lieu de travail

3=à la cantine / restaurant d'entreprise

4=dans un fast-food

5=dans un restaurant

6=dans un café, une cafétéria

7=dans un autre lieu

(en règle générale)

1=au domicile

2=sur le lieu de travail

3=à la cantine / restaurant d'entreprise

4=dans un fast-food

5=dans un restaurant

6=dans un café, une cafétéria

7=dans un autre lieu

(en règle générale)

1=au domicile

2=sur le lieu de travail

Nombre de personnes >18ans

Nombre de personnes de plus de 18 ans vivant sous le même toit que le sujet, y compris le sujet lui-même

Si le sujet est étudiant et partage son temps entre la résidence de ses parents et sa résidence d'étudiant, considérer le foyer parental.

1=1

2=2

3=3

4=4

5=5 ou plus

6=le sujet habite dans une résidence collective (maison de retraite, foyer de travailleurs...)

Niveau net moy(en) (avant impôts) de la totalité des ressources mensuelles du foyer du sujet, en tenant compte des allocations familiales.

Si le sujet est étudiant et partage son temps entre la résidence de ses parents et sa résidence d'étudiant, considérer le foyer parental.

1=moins de 750 euros

2=de 750 à 1500 euros

3=de 1500 à 2000 euros

4=de 2000 à 3000 euros

5=de 3000 à 4000 euros

6=de 4000 à 5000 euros

7=plus de 5000 euros

8=ne sait pas

9=Combien le sujet a d'enfants (en comptant aussi les enfants qui ont quitté le domicile)

0=0

1=1

2=2

3=3

4=4

5=5 ou plus

(en règle générale)

1=au domicile

2=sur le lieu de travail

3=à la cantine / restaurant d'entreprise

4=dans un fast-food

5=dans un restaurant

6=dans un café, une cafétéria

7=dans un autre lieu

(en règle générale)

1=au domicile

2=sur le lieu de travail

3=à la cantine / restaurant d'entreprise

4=dans un fast-food

5=dans un restaurant

6=dans un café, une cafétéria

7=dans un autre lieu

(en règle générale)

1=au domicile

2=sur le lieu de travail

3=à la cantine / restaurant d'entreprise

4=dans un fast-food

5=dans un restaurant

6=dans un café, une cafétéria

7=dans un autre lieu

(en règle générale)

1=au domicile

2=sur le lieu de travail

3=à la cantine / restaurant d'entreprise

4=dans un fast-food

5=dans un restaurant

6=dans un café, une cafétéria

7=dans un autre lieu

(en règle générale)

1=au domicile

2=sur le lieu de travail

3=à la cantine / restaurant d'entreprise

4=dans un fast-food

5=dans un restaurant

6=dans un café, une cafétéria

7=dans un autre lieu

(en règle générale)

1=au domicile

2=sur le lieu de travail

3=à la cantine / restaurant d'entreprise

4=dans un fast-food

5=dans un restaurant

6=dans un café, une cafétéria

7=dans un autre lieu

(en règle générale)

1=au domicile

2=sur le lieu de travail

3=à la cantine / restaurant d'entreprise

4=dans un fast-food

5=dans un restaurant

6=dans un café, une cafétéria

7=dans un autre lieu

(en règle générale)

1=au domicile

2=sur le lieu de travail

3=à la cantine / restaurant d'entreprise

4=dans un fast-food

5=dans un restaurant

6=dans un café, une cafétéria

7=dans un autre lieu

(en règle générale)

1=au domicile

2=sur le lieu de travail

Activité

1=à l'école, étudiant ou recherche du premier emploi

2=apprenant/explaining

3=artisan, commerçant

4=chef d'entreprise

Préparation des repas	Concernant les repas pris au domicile, le sujet participe à la préparation... 1=de tous les repas pris au domicile 2=de plus de la moitié des repas pris au domicile 3=de la moitié des repas pris au domicile 4=de moins de la moitié des repas pris au domicile 5=le sujet ne participe jamais à la préparation des repas pris au domicile
Temps de préparation	Lorsque le sujet participe à la préparation d'un repas pris au domicile, combien de temps y passe-t-il en moyenne ? 1=moins de 5 minutes 2=de 5 à 10 minutes 3=de 11 à 15 minutes 4=de 16 à 30 minutes 5=de 31 minutes à 1 heure
Achats de denrées alimentaires	Le sujet ne participe jamais à la préparation 7=le sujet ne participe jamais à la préparation Fréquence à laquelle le sujet participe aux achats des denrées alimentaires pour son foyer 1=plusieurs fois par semaine 2=environ 1 fois par semaine 3=1 à 3 fois par mois 4=moins d'une fois par mois 5=jamais
Fréquence restaurants	Fréquentation de restaurants (non compris fast-food et cafétérias): 1=plusieurs fois par semaine 2=environ 1 fois par semaine 3=1 à 3 fois par mois 4=moins d'une fois par mois 5=jamais
Restaurant français : déjà fréquenté ?	0=non 1=oui
Restaurant italien : déjà fréquenté ?	0=non 1=oui
Restaurant mexicain : déjà fréquenté ?	0=non 1=oui
Restaurant chinois : déjà fréquenté ?	0=non 1=oui
Restaurant japonais : déjà fréquenté ?	0=non 1=oui
Restaurant réunionnais : déjà fréquenté ?	0=non 1=oui
Restaurant cubain : déjà fréquenté ?	0=non 1=oui
Restaurant grec : déjà fréquenté ?	0=non 1=oui
Restaurant libanais : déjà fréquenté ?	0=non 1=oui
Restaurant tunisien : déjà fréquenté ?	0=non 1=oui
Restaurant espagnol : déjà fréquenté ?	0=non 1=oui
Restaurant indien : déjà fréquenté ?	0=non 1=oui
Autre restaurant étranger : déjà fréquenté ?	0=non 1=oui
Consommation de plats français	Lors des repas pris au domicile, fréquence de consommation de plats français 1=au moins 1 fois par semaine 2=moins d'1 fois par semaine, mais 1 à plusieurs fois par mois 3=moins d'1 fois par mois, mais 1 à plusieurs fois par an 4=moins d'1 fois par an 5=jamais
Consommation de plats italiens	Lors des repas pris au domicile, fréquence de consommation de plats italiens 1=au moins 1 fois par semaine 2=moins d'1 fois par semaine, mais 1 à plusieurs fois par mois 3=moins d'1 fois par mois, mais 1 à plusieurs fois par an 4=moins d'1 fois par an 5=jamais
Consommation de plats mexicains	Lors des repas pris au domicile, fréquence de consommation de plats mexicains 1=au moins 1 fois par semaine 2=moins d'1 fois par semaine, mais 1 à plusieurs fois par mois 3=moins d'1 fois par mois, mais 1 à plusieurs fois par an 4=moins d'1 fois par an 5=jamais
Consommation de plats chinois	Lors des repas pris au domicile, fréquence de consommation de plats chinois 1=au moins 1 fois par semaine 2=moins d'1 fois par semaine, mais 1 à plusieurs fois par mois 3=moins d'1 fois par mois, mais 1 à plusieurs fois par an 4=moins d'1 fois par an 5=jamais

Consommation de plats japonais	Lors des repas pris au domicile, fréquence de consommation de plats japonais 1=au moins 1 fois par semaine 2=moins d'1 fois par semaine, mais 1 à plusieurs fois par mois 3=moins d'1 fois par mois, mais 1 à plusieurs fois par an 4=moins d'1 fois par an 5=jamais
Consommation de plats réunionnais	Lors des repas pris au domicile, fréquence de consommation de plats réunionnais 1=au moins 1 fois par semaine 2=moins d'1 fois par semaine, mais 1 à plusieurs fois par mois 3=moins d'1 fois par mois, mais 1 à plusieurs fois par an 4=moins d'1 fois par an 5=jamais
Consommation de plats cubains	Lors des repas pris au domicile, fréquence de consommation de plats cubains 1=au moins 1 fois par semaine 2=moins d'1 fois par semaine, mais 1 à plusieurs fois par mois 3=moins d'1 fois par mois, mais 1 à plusieurs fois par an 4=moins d'1 fois par an 5=jamais
Consommation de plats grecs	Lors des repas pris au domicile, fréquence de consommation de plats grecs 1=au moins 1 fois par semaine 2=moins d'1 fois par semaine, mais 1 à plusieurs fois par mois 3=moins d'1 fois par mois, mais 1 à plusieurs fois par an 4=moins d'1 fois par an 5=jamais
Consommation de plats libanais	Lors des repas pris au domicile, fréquence de consommation de plats libanais 1=au moins 1 fois par semaine 2=moins d'1 fois par semaine, mais 1 à plusieurs fois par mois 3=moins d'1 fois par mois, mais 1 à plusieurs fois par an 4=moins d'1 fois par an 5=jamais
Consommation de plats tunisiens	Lors des repas pris au domicile, fréquence de consommation de plats tunisiens 1=au moins 1 fois par semaine 2=moins d'1 fois par semaine, mais 1 à plusieurs fois par mois 3=moins d'1 fois par mois, mais 1 à plusieurs fois par an 4=moins d'1 fois par an 5=jamais
Consommation de plats espagnols	Lors des repas pris au domicile, fréquence de consommation de plats espagnols 1=au moins 1 fois par semaine 2=moins d'1 fois par semaine, mais 1 à plusieurs fois par mois 3=moins d'1 fois par mois, mais 1 à plusieurs fois par an 4=moins d'1 fois par an 5=jamais
Consommation de plats indiens	Lors des repas pris au domicile, fréquence de consommation de plats indiens 1=au moins 1 fois par semaine 2=moins d'1 fois par semaine, mais 1 à plusieurs fois par mois 3=moins d'1 fois par mois, mais 1 à plusieurs fois par an 4=moins d'1 fois par an 5=jamais
Voyages à l'étranger	Fréquence des voyages à l'étranger (seuls les voyages de plus de 2 jours sont pris en compte) 1=1 à plusieurs fois par mois 2=moins d'1 fois par mois mais au moins 1 fois par an 3=moins d'1 fois par an mais au moins 1 fois tous les 5 ans 4=moins d'1 fois tous les 5 ans 5=jamais
Consommation de cuisine locale	Lors des voyages à l'étranger, consommation de cuisine locale 1=à tous les repas 2=à plus de la moitié des repas 3=à moins de la moitié des repas 4=à aucun des repas 5=le sujet ne voyage jamais à l'étranger
Prothèses dentaires	Le sujet porte-t-il une ou plusieurs prothèse(s) dentaire(s) ? 0=non 1=oui
Femme enceinte	Le sujet est-il une femme enceinte ? 0=non 1=oui
Fumeur	Le sujet est-il fumeur ? 0=non (il a jamais fumé) 1=oui, fumeur actuel 2=ancien fumeur
Fumeur : nombre de cigarettes	*-nombre moyen de cigarettes fumées quotidiennement (si le sujet fume la pipe ou le cigare, faire une approximation d'équivalence) 0=le sujet est non fumeur ou ancien fumeur 1=moins d'une cigarette par jour 2=1 à 4 cigarettes par jour 3=5 à 19 cigarettes par jour 4=20 cigarettes ou plus par jour
Fumeur : durée	Fumures actuels *-durées combiées d'années ? Nombre à 1 ou 2 chiffres. Si non fumeur ou ancien fumeur, saisir 0.

Fumeur : dernière cigarette
Fumeurs : à quand remonte la dernière cigarette ?
0=non fumeur ou ancien fumeur
1=à moins d'1 heure
2=à plus d'1 heure mais à moins de 2 heures
3=à plus de 2 heures
Anciens fumeurs : nombre moyen de cigarettes fumées quotidiennement à l'époque (si le sujet fumait la pipe ou le cigare, faire une approximation d'équivalence)

Ancien fumeur : nombre de cigarettes
Date sujet est non fumeur ou fumeur actuel
1=moins d'une cigarette par jour
2=1 à 4 cigarettes par jour
3=5 à 19 cigarettes par jour
4=20 cigarettes ou plus par jour
Anciens fumeurs : pendant combien d'années ?
Nombre à 1 ou 2 chiffres.

Ancien fumeur : durée
Si non fumeur ou fumeur actuel, saisir 0.
Anciens fumeurs : depuis combien d'années le sujet a-t-il arrêté de fumer ?
Nombre à 1 ou 2 chiffres.

Taille
Si non fumeur ou fumeur actuel, saisir 0.
En centimètres.
Nombre à 3 chiffres.

Poids
Nombre à 2 ou 3 chiffres.
raison principale pour laquelle le sujet a accepté de participer à l'étude.

Motivation
1=curiosité, goût pour les produits nouveaux
2=intérêt pour la recherche
3=pour rencontrer d'autres personnes
4=pour la rémunération
5=pour combler une partie de son temps libre
6=pour une autre raison

Profil : participation antérieure ?
Uniquement pour les données de profil :
Homnis cette étude, le sujet a-t-il déjà participé à une étude d'analyse saisonnière en tant que sujet entrainé ?
0=non
1=oui

Profil : nombre d'années
Uniquement pour les données de profil :
Depuis combien d'années le sujet participe à des études en tant que sujet entrainé ?
Nombre à 1 ou 2 chiffres.

Profil : nombre annuel de séances
Si c'est la première fois que le sujet participe, saisir 0.
Uniquement pour les données de profil :
Depuis sa première expérience en tant que sujet entrainé, nombre moyen annuel de séances auxquelles le sujet a participé, en tant que sujet entrainé. (Y compris les séances d'entraînement)

Profil : nombre d'études
0=c'est la première fois que le sujet participe en tant que sujet entrainé
1=plus de 150 séances par an
2=de 100 à 150 séances par an
3=de 50 à 99 séances par an
4=de 30 à 49 séances par an
5=de 20 à 29 séances par an
6=de 10 à 19 séances par an
7=moins de 10 séances par an

Profil : autres produits
Uniquement pour les données de profil :
Nombre total d'études différentes auxquelles le sujet a participé, en tant que sujet entrainé, sans tenir compte de l'étude actuelle. (On entend par "étude" un ensemble de séances portant sur un même type de produit, et comportant une phase d'entraînement suivie d'une phase d'évaluation des produits).
Nombre à 1 ou 2 chiffres. Si c'est la première fois que le sujet participe à une étude en tant que sujet entrainé, saisir 0.

Profil : pains, biscuits ?
Uniquement pour les données de profil :
Le sujet a-t-il déjà évalué, en tant que sujet entrainé, des pains ou biscuits ?
0=non
1=oui

Profil : Biscuits secs salés
Uniquement pour les données de profil :
Le sujet a-t-il déjà évalué, en tant que sujet entrainé, des biscuits secs salés ?
0=non
1=oui

Profil : Biscuits sucrés ?
Uniquement pour les données de profil :
Le sujet a-t-il déjà évalué, en tant que sujet entrainé, des biscuits sucrés ?
0=non
1=oui

Profil : pâtisseries et viennoiseries ?
Uniquement pour les données de profil :
Le sujet a-t-il déjà évalué, en tant que sujet entrainé, des pâtisseries ou viennoiseries ?
0=non
1=oui

Profil : pâtes alimentaires et riz ?
Uniquement pour les données de profil :
Le sujet a-t-il déjà évalué, en tant que sujet entrainé, des pâtes alimentaires ou riz ?
0=non
1=oui

Profil : autres produits céréaliers ?
Uniquement pour les données de profil :
Le sujet a-t-il déjà évalué, en tant que sujet entrainé, des autres produits céréaliers ?
0=non
1=oui

Profil : yaourts, petits suisses, fromages blancs ?
Uniquement pour les données de profil :
Le sujet a-t-il déjà évalué, en tant que sujet entrainé, des yaourts, petits-suisses ou fromages blancs ?
0=non
1=oui

Profil : glaces ?
Uniquement pour les données de profil :
Le sujet a-t-il déjà évalué, en tant que sujet entrainé, des glaces ?
0=non
1=oui

Profil : desserts lactés ?
Uniquement pour les données de profil :
Le sujet a-t-il déjà évalué, en tant que sujet entrainé, des desserts lactés ?
0=non
1=oui

Profil : laits ?
Uniquement pour les données de profil :
Le sujet a-t-il déjà évalué, en tant que sujet entrainé, des laits ?
0=non
1=oui

Profil : fromages ?
Uniquement pour les données de profil :
Le sujet a-t-il déjà évalué, en tant que sujet entrainé, des fromages ?
0=non
1=oui

Profil : autres produits laitiers ?
Uniquement pour les données de profil :
Le sujet a-t-il déjà évalué, en tant que sujet entrainé, des autres produits laitiers ?
0=non
1=oui

Profil : œufs et dérivés ?
Uniquement pour les données de profil :
Le sujet a-t-il déjà évalué, en tant que sujet entrainé, des œufs et dérivés ?
0=non
1=oui

Profil : graisses et huiles ?
Uniquement pour les données de profil :
Le sujet a-t-il déjà évalué, en tant que sujet entrainé, des graisses et huiles ?
0=non
1=oui

Profil : viandes ?
Uniquement pour les données de profil :
Le sujet a-t-il déjà évalué, en tant que sujet entrainé, des viandes ?
0=non
1=oui

Profil : volailles et gibiers ?
Uniquement pour les données de profil :
Le sujet a-t-il déjà évalué, en tant que sujet entrainé, des volailles et gibiers ?
0=non
1=oui

Profil : abats ?
Uniquement pour les données de profil :
Le sujet a-t-il déjà évalué, en tant que sujet entrainé, des abats ?
0=non
1=oui

Profil : autres produits dérivés de la viande ?
Uniquement pour les données de profil :
Le sujet a-t-il déjà évalué, en tant que sujet entrainé, des autres produits dérivés de la viande ?
0=non
1=oui

Profil : saucisses, saucissons, boudins ?
Uniquement pour les données de profil :
Le sujet a-t-il déjà évalué, en tant que sujet entrainé, des saucisses, saucissons, boudins ?
0=non
1=oui

Profil : toie gras, pâtes, rillettes ?
Uniquement pour les données de profil :
Le sujet a-t-il déjà évalué, en tant que sujet entrainé, des foies gras, pâtes, rillettes ?
0=non
1=oui

Profil : jambons ?
Uniquement pour les données de profil :
Le sujet a-t-il déjà évalué, en tant que sujet entrainé, des jambons ?
0=non
1=oui

Profil : autres produits de charcuterie ?
Uniquement pour les données de profil :
Le sujet a-t-il déjà évalué, en tant que sujet entrainé, des autres produits de charcuterie ?
0=non
1=oui

Profil : coquilles, mollusques, crustacés ?
Uniquement pour les données de profil :
Le sujet a-t-il déjà évalué, en tant que sujet entrainé, des coquillages, mollusques, crustacés ?
0=non
1=oui

Profil : œufs de poisson, tarاما ?
Uniquement pour les données de profil :
Le sujet a-t-il déjà évalué, en tant que sujet entrainé, des œufs de poissons, tarاما ?
0=non
1=oui

Profil : poissons frais ?
Uniquement pour les données de profil :
Le sujet a-t-il déjà évalué, en tant que sujet entrainé, des poissons frais ?
0=non
1=oui

Profil : poissons panés ?
Uniquement pour les données de profil :
Le sujet a-t-il déjà évalué, en tant que sujet entrainé, des poissons panés ?
0=non
1=oui

Profil : poissons en conserve ?
Uniquement pour les données de profil :
Le sujet a-t-il déjà évalué, en tant que sujet entrainé, des poissons en conserve ?
0=non
1=oui

Profil : poissons fumés ?
Uniquement pour les données de profil :
Le sujet a-t-il déjà évalué, en tant que sujet entrainé, des poissons fumés ?
0=non
1=oui

Profil : autres produits dérivés de poissons ?
Uniquement pour les données de profil :
Le sujet a-t-il déjà évalué, en tant que sujet entrainé, des autres produits dérivés de poissons ?
0=non
1=oui

Profil : légumes cuits ?
Uniquement pour les données de profil :
Le sujet a-t-il déjà évalué, en tant que sujet entrainé, des légumes cuits ?
0=non
1=oui

Profil : crudités ?
Uniquement pour les données de profil :
Le sujet a-t-il déjà évalué, en tant que sujet entrainé, des crudités ?
0=non
1=oui

Profil : pommes de terre ?
Uniquement pour les données de profil :
Le sujet a-t-il déjà évalué, en tant que sujet entrainé, des pommes de terre et dérivés ?
0=non
1=oui

Profil : légumes secs et féculents ?
Uniquement pour les données de profil :
Le sujet a-t-il déjà évalué, en tant que sujet entrainé, des légumes secs et féculents ?
0=non
1=oui

Profil : fruits frais ?
Uniquement pour les données de profil :
Le sujet a-t-il déjà évalué, en tant que sujet entrainé, des fruits frais ?
0=non
1=oui

Profil : fruits secs, graines ?
Uniquement pour les données de profil :
Le sujet a-t-il déjà évalué, en tant que sujet entrainé, des fruits secs, amandes, noix, graines ?
0=non
1=oui

Profil : fruits au sirop, compotes ?
Uniquement pour les données de profil :
Le sujet a-t-il déjà évalué, en tant que sujet entrainé, des fruits au sirop, compotes ?
0=non
1=oui

Profil : confiseries de sucre ?
Uniquement pour les données de profil :
Le sujet a-t-il déjà évalué, en tant que sujet entrainé, des confiseries de sucre ?
0=non
1=oui

DESCRIPTEUR

Code descripteur	Code permettant d'identifier chaque descripteur au sein du jeu de données. (Un code par descripteur) Chaque descripteur est constitué d'une colonne
Est-ce une Clé ?	Où=1 Non=0 Si un descripteur reçoit toujours le même Code (ligne précédente) entre les différents jeux de données que vous pourrez nous envoyer et que ce code ne sera jamais utilisé pour identifier un autre descripteur, alors ce Code descripteur est une clé. Si d'un jeu de données à l'autre un Code peut correspondre à deux descripteurs différents, ce code descripteur n'est pas une clé.
Libellé du descripteur	Libellés, tel qu'il a été donné aux sujets. Si confidentiel, le coder.
Type de descripteur	1=apparentive 2=odeur 3=saveur 4=texture 5=arôme 6=arrière-goût 7=flétrissimal 8=notation hédonique
Valeur minimale de l'échelle	
Valeur maximale de l'échelle	
Type d'échelle	1=rectaire continue rectangulaire 2=échelle continue linéaire fermée 3=échelle continue linéaire ouverte 4=échelle discrète 5=autre échelle Uniquement si Type d'échelle = 1, 2 ou 3. 6=non L'échelle est-elle graduée ? 1=oui Uniquement si Type d'échelle = 4. Préciser le nombre de valeurs possibles de cette échelle.
Graduations	
Echelle discrète : précision	
Profil : définition du descripteur	Uniquement pour les données de profil. Définition du descripteur donnée aux sujets pour l'évaluation. Si une méthode particulière d'évaluation a été donnée, la détailler.
Profil : référence de la définition	Uniquement pour les données de profil. Si la définition est tirée d'une norme ou d'une publication, indiquer sa référence.
Profil : nombre de standards	Uniquement pour les données de profil. Combien de références (standards) ont été utilisés pour définir ce descripteur ?
Label borne inférieure	Label de la borne inférieure de l'échelle
Label borne supérieure	Label de la borne supérieure de l'échelle

PROTOCOLE

Date début étude	Date de début de l'étude = date de la première séance d'évaluation. Format MM/AA/AA
Date fin étude	Date de fin d'étude =date de la dernière séance d'évaluation. Format MM/AA/AA
Phase de sélection ?	Y a-t-il eu une phase de sélection des sujets ? 0=non 1=oui
Sélection : nombre de sujets recrutés	Uniquement s'il y a eu une phase de sélection. Nombre de sujets recrutés pour effectuer les tests de sélection.
Sélection : détection des incapacités	Uniquement s'il y a eu une phase de sélection : types de tests effectués Tests de détection des incapacités (agouésie, anosmie, vision des couleurs...)? 0=non effectué 1=effectué
Sélection : détermination de l'acuité sensorielle	Uniquement s'il y a eu une phase de sélection : types de tests effectués Tests de détermination de l'acuité sensorielle (tests d'appariement, de détection, de discrimination, de classement ? 0=non effectué 1=effectué
Sélection : évaluation du potentiel à décrire	Uniquement s'il y a eu une phase de sélection : types de tests effectués Tests d'évaluation du potentiel à décrire et communiquer des informations sur des perceptions sensorielles ? 0=non effectué 1=effectué
Sélection : autres types de tests	Uniquement s'il y a eu une phase de sélection : types de tests effectués Autres types de tests ? 0=non effectué 1=effectué
Sélection : précisions	Uniquement s'il y a eu une phase de sélection. La phase de sélection pourra être décrite plus précisément dans un document Word (description précise des tests utilisés, critères de sélection des juges...) Si les tests sont normalisés, on pourra indiquer la référence de la norme. Indiquer dans cette ligne le nom du fichier Word.
Entraînement : nombre de sujets	Uniquement pour les données de profil. Nombre de sujets ayant participé à la phase d'entraînement.
Entraînement : nombre de séances	Uniquement pour les données de profil. Nombre de séances d'entraînement (sans compter les éventuelles séances tests qui ont lieu après les séances d'entraînement et avant les séances d'évaluation)
Entraînement : durée d'une séance	Uniquement pour les données de profil.
Entraînement : tests de détection	Durée moyenne d'une séance d'entraînement (en minutes) Uniquement pour les données de profil : types de tests effectués pendant l'entraînement Tests de détection ? 0=non effectué 1=effectué
Entraînement : tests de reconnaissance	Uniquement pour les données de profil : types de tests effectués pendant l'entraînement Tests de reconnaissance ? 0=non effectué 1=effectué
Entraînement : tests d'appariement	Uniquement pour les données de profil : types de tests effectués pendant l'entraînement Tests d'appariement ? 0=non effectué 1=effectué
Entraînement : tests de discrimination	Uniquement pour les données de profil : types de tests effectués pendant l'entraînement Tests de discrimination ? 0=non effectué 1=effectué
Entraînement : tests de classement	Uniquement pour les données de profil : types de tests effectués pendant l'entraînement Tests de classement ? 0=non effectué 1=effectué
Entraînement : tests d'évaluation descriptive	Uniquement pour les données de profil : types de tests effectués pendant l'entraînement Tests d'évaluation descriptive ? 0=non effectué 1=effectué
Entraînement au développement de descripteurs	Uniquement pour les données de profil : types de tests effectués pendant l'entraînement Entraînement au développement de descripteurs ? 0=non effectué 1=effectué
Entraînement à l'utilisation de descripteurs	Uniquement pour les données de profil : types de tests effectués pendant l'entraînement Entraînement à l'utilisation de descripteurs ? 0=non effectué 1=effectué
Entraînement à l'utilisation de l'échelle	Uniquement pour les données de profil : types de tests effectués pendant l'entraînement Entraînement à l'utilisation de l'échelle ? 0=non effectué 1=effectué
Entraînement : stimuli utilisés	Uniquement pour les données de profil : stimuli utilisés pendant l'entraînement. 1=stimuli non alimentaires 2=stimuli alimentaires 3=stimuli mixtes

Entraînement : stimuli alimentaires

Uniquement pour les données de profil : si utilisation de stimuli alimentaires pour les tests d'entraînement
établi- ce des produits du même type que ceux à tester ?

1=oui, tous

2=non, aucun

3=certains oui, d'autres non

1=générés par le jury

2=fournis par l'expérimentateur

3=certains générés par le jury, d'autres fournis par l'expérimentateur

Entraînement : précisions

Uniquement pour les données de profil.

La phase d'entraînement pourra être décrite plus précisément dans un document Word (description des tests, des stimuli utilisés, dépenses de génération des descripteurs, utilisation de standards...). Si les tests sont normalisés, on pourra indiquer la référence de la norme.

Indiquer dans cette ligne le nom du fichier Word.

Uniquement pour les données de profil.

Entraînement : vérification des performances
Y'a-t-il eu une vérification des performances des sujets à l'issue de l'entraînement ?

0=non 1=oui

Entraînement : sélection finale
Uniquement pour les données de profil.
Y'a-t-il eu une sélection finale des sujets au vu des résultats de l'entraînement ?

0=non 1=oui

Critères de sélection finale : répétabilité

Uniquement pour les données de profil.

Si y a eu une sélection finale au vu des résultats de l'entraînement, était-ce sur un critère de répétabilité ?

0=non 1=oui

Critères de sélection finale : discrimination

Uniquement pour les données de profil.

Si y a eu une sélection finale au vu des résultats de l'entraînement, était-ce sur un critère de bonne discrimination des produits ?

0=non 1=oui

Critères de sélection finale : comportement

Uniquement pour les données de profil.

Si y a eu une sélection finale au vu des résultats de l'entraînement, était-ce sur un critère de comportement général du sujet ?

0=non 1=oui

Critères de sélection finale : disponibilité

Uniquement pour les données de profil.

Si y a eu une sélection finale au vu des résultats de l'entraînement, était-ce sur un critère de disponibilité du sujet ?

0=non 1=oui

Critères de sélection finale : autre critère

Uniquement pour les données de profil.

Si y a eu une sélection finale au vu des résultats de l'entraînement, était-ce sur un autre critère que ceux détaillés dans les lignes précédentes ?

0=non 1=oui

Durée d'une séance d'évaluation

Concernant l'évaluation (théorique ou de profil), quelle était la durée moyenne d'une séance (en minutes) ?

0=non 1=oui

Mode de collecte des données

Nombre de séances d'évaluation (y compris les éventuelles séances tests qui ont lieu après l'entraînement et avant l'évaluation)

1= système informatique

2= questionnaires papier

type de numéro unique pendant les séances.

1= lumière naturelle

2= lumière blanche

3= lumière rouge

4= lumière verte

5= autre type de lumière

Protocole de préparation des produits

Le protocole de préparation des produits pourra être décrit dans un document Word (conservation, préparation, maintien en température, photos éventuelles des produits ou des emballages...).

Indiquer le nom du fichier Word.

1= tous les sujets ont reçu les produits dans le même ordre

2= carré latin

3= bloc incomplet équilibré

4= autre type d'ordres de présentation

Y'avait-il un produit d'échauffement pour lequel les données ne sont pas fournies ?

0=non 1=oui

Même produit d'échauffement ?

Si y a eu un produit d'échauffement, était-ce le même pour tous les sujets ?

0=non 1=oui

Place du produit d'échauffement ?

Si y a eu un produit d'échauffement, celui-ci apparaissait...

1= uniquement lors de la première séance

2= à chaque séance

3= autre présentation

Instructions données aux sujets

Les instructions données aux sujets pour l'évaluation des produits pourront être décrites dans un document Word (préciser si expectation, rinçage... ; préciser l'intervalle de temps entre les échantillons...).

Indiquer dans cette ligne le nom du fichier Word.

Température de service des échantillons, en degrés Celsius.

1= normale 2= comparatif

3= autre température, voir remarques pour leur pertinence et leur

1= présent sous forme de saleté

2= indommision sous forme de bons d'achat dans un magasin

3= échantillons de produits alimentaires liés à l'étude

4= échantillons de produits alimentaires non liés à l'étude

5= autre cadeau

6= ils ne recevaient aucune indemnité

Uniquement si l'indemnisation=1 ou 2.

Somme reçue

Quelle somme chaque sujet recevait-il par heure, en euros ?

Description du protocole

Toute autre information concernant le protocole pourra être décrite dans un document Word

Préciser dans cette ligne le nom du fichier Word.

Annexe 6

Données fournies par Unilever et tableaux de résultats

Résultats concernant le choix de la partie aléatoire

Tests de différence des modèles emboîtés basés sur le logarithme du rapport de vraisemblance – étude n°2

DESCR	ID versus									CS versus						RCL versus			choix
	CS			RCL			RCQ			RCL			RCQ			RCQ			
	ddl	logV	prob	ddl	logV	prob	ddl	logV	prob	ddl	logV	prob	ddl	logV	prob	ddl	logV	prob	
ap_brown	1	73,3	<0,001	2	81,2	<0,001	3	81,2	<0,001	1	7,9	0,005	1	7,9	0,005	0	<0,01	0,999	RCL
ap_filmlayer	1	45,8	<0,001	2	51,8	<0,001	3	73,0	<0,001	1	6,0	0,014	2	27,2	<0,001	1	21,2	<0,001	RCQ
ap_red	1	51,4	<0,001	2	52,7	<0,001	3	52,7	<0,001	1	1,3	0,254	1	1,3	0,254	0	0,0	0,998	CS
ap_transparant	1	65,9	<0,001	2	68,6	<0,001	3	68,7	<0,001	1	2,7	0,100	2	2,8	0,247	0	0,1	0,784	CS
ap_yellow	1	169,5	<0,001	2	203,7	<0,001	3	211,2	<0,001	1	34,2	<0,001	2	41,7	<0,001	1	7,5	0,006	RCQ
m_astringent	1	120,4	<0,001	2	120,8	<0,001	3	123,5	<0,001	1	0,4	0,527	2	3,1	0,212	1	2,7	0,100	CS
m_prickling	1	113,0	<0,001	2	113,2	<0,001	3	114,2	<0,001	0	0,2	0,655	1	1,2	0,273	1	1,2	0,273	CS
m_salivastim	1	124,0	<0,001	2	124,1	<0,001	3	126,8	<0,001	0	0,1	0,752	1	2,8	0,094	1	2,8	0,094	CS
o_alcohol	1	121,5	<0,001	2	122,7	<0,001	3	123,7	<0,001	1	1,2	0,273	2	2,2	0,333	1	1,0	0,317	CS
o_animal	1	57,0	<0,001	2	57,2	<0,001	3	67,6	<0,001	1	0,2	0,655	2	10,6	0,005	1	10,4	0,001	RCQ
o_cheese	1	32,4	<0,001	2	38,1	<0,001	3	51,4	<0,001	1	5,7	0,017	2	19,0	0,000	1	13,3	<0,001	RCQ
o_fermen_fruit	1	93,7	<0,001	2	97,9	<0,001	3	99,8	<0,001	1	4,2	0,040	2	6,1	0,047	1	1,9	0,168	RCL
o_liquorice	1	81,4	<0,001	2	82,6	<0,001	3	84,2	<0,001	1	1,2	0,273	2	2,8	0,247	1	1,6	0,206	CS
o_medecin	1	83,2	<0,001	2	84,7	<0,001	3	84,8	<0,001	1	1,5	0,221	2	1,6	0,449	1	0,1	0,752	CS
o_mushroom	1	193,1	<0,001	2	199,5	<0,001	3	201,0	<0,001	1	6,4	0,011	2	7,9	0,019	1	1,5	0,221	RCL
o_roastpork	1	39,1	<0,001	2	41,2	<0,001	3	41,4	<0,001	1	2,1	0,147	2	2,3	0,317	1	0,2	0,655	CS
o_ryebread	1	103,8	<0,001	2	107,5	<0,001	3	108,4	<0,001	1	3,7	0,054	2	4,6	0,100	1	0,9	0,343	CS
o_treacle	1	79,0	<0,001	2	79,2	<0,001	3	91,7	<0,001	1	0,2	0,655	2	12,7	0,002	1	12,5	<0,001	RCQ
t_burnt	1	120,8	<0,001	2	130,2	<0,001	3	130,3	<0,001	1	9,4	0,002	2	9,5	0,009	1	0,1	0,752	RCL
t_salt	1	108,3	<0,001	2	111,2	<0,001	3	111,3	<0,001	1	2,9	0,089	2	3,0	0,223	1	0,1	0,752	CS
t_sour	1	125,0	<0,001	2	125,1	<0,001	3	125,6	<0,001	1	0,1	0,752	2	0,6	0,741	1	0,5	0,480	CS
t_sweet	1	46,6	<0,001	2	46,7	<0,001	2	57,2	<0,001	1	0,1	0,752	2	10,6	0,005	1	10,6	0,001	RCQ
ta_bitter	1	134,9	<0,001	2	136,1	<0,001	3	136,3	<0,001	1	1,2	0,273	2	1,4	0,497	1	0,2	0,655	CS
ta_lavas	1	113,1	<0,001	2	114,8	<0,001	3	115,6	<0,001	1	1,7	0,192	2	2,5	0,287	1	0,8	0,371	CS

DESCR: nom des descripteurs, classés par type; ddl: nombre de degrés de liberté du test du Khi-deux; logV: logarithme du rapport de vraisemblance entre les deux modèles testés; prob: probabilité associée au test, les valeurs inférieures à 0,05 sont surlignées en gras.

indice AIC pour chaque descripteur en fonction de la structure de covariance

descripteurs	ID*	CS*	RCL*	RCQ*
ap_brown	506,5	435,2	429,3	429,3
ap_filmlayer	705,9	662,1	658,1	638,9
ap_red	536,9	487,5	488,2	488,2
ap_transparant	831	767,1	766,4	766,4
ap_yellow	653,8	486,3	454,1	448,6
m_astringent	743,8	625,4	627	626,3
m_prickling	854,5	743,5	743,5	744,3
m_saliva_stimul	832,2	710,2	710,2	709,4
o_alcohol	801,4	681,9	682,7	683,7
o_animal	906,4	851,4	853,2	844,8
o_cheese	619,1	588,7	585	573,7
o_fermen_fruit	774,3	682,6	680,4	680,5
o_liquorice	567,2	487,8	488,6	489
o_medecin	814,8	733,6	734,1	736
o_mushroom	466,7	275,6	271,2	271,7
o_roastpork	442,3	405,2	405,1	406,9
o_ryebread	803,8	702	700,3	701,4
o_treacle	834,2	757,2	759	748,5
t_burnt	930,9	812,1	804,7	806,6
t_salt	634,9	528,6	527,7	529,7
t_sour	651,9	528,9	530,8	532,3
t_sweet	850,1	805,5	805,5	796,9
ta_bitter	860,6	727,7	728,5	730,3
ta_lavas	822,4	711,3	711,6	712,8
nb sel ¹	0	9	8	7

*: meilleure structure en gras pour chaque descripteur (l'AIC est d'autant plus faible que l'ajustement est adéquat)

¹: nombre de sélection de chaque structure de covariance

Résultats concernant la partie fixe

Estimation des paramètres liés aux effets fixes et tests de Fisher
(résultats pour le modèle sélectionné pour chaque descripteur uniquement)

Descripteur	fix	scorr	Effet	Estimation*	StdErr	DF	tValue	Probt**
ap_brown	t	CS	Intercept	-0.9069	0.06175	12	-14.69	<.0001
ap_brown	t	CS	tps0	0.005549	0.001616	376	3.43	0.0007
ap_filmlayer	t2	CS	Intercept	-0.8077	0.08098	12	-9.97	<.0001
ap_filmlayer	t2	CS	tps0	0.01267	0.002287	375	5.54	<.0001
ap_filmlayer	t2	CS	tps0*tps0	-0.00092	0.000194	375	-4.75	<.0001
ap_red	t2	CS	Intercept	-0.8478	0.06646	12	-12.76	<.0001
ap_red	t2	CS	tps0	0.01276	0.001737	375	7.35	<.0001
ap_red	t2	CS	tps0*tps0	-0.00045	0.000147	375	-3.08	0.0022
ap_transparent	t2	CS	Intercept	-1.0245	0.1021	12	-10.03	<.0001
ap_transparent	t2	CS	tps0	-0.00445	0.002900	375	-1.53	0.1261
ap_transparent	t2	CS	tps0*tps0	0.000666	0.000246	375	2.71	0.0070
ap_yellow	cst	CS	Intercept	-0.8025	0.1019	12	-7.88	<.0001
m_astringent	t2	CS	Intercept	-1.1099	0.1018	12	-10.90	<.0001
m_astringent	t2	CS	tps0	0.01223	0.002113	375	5.79	<.0001
m_astringent	t2	CS	tps0*tps0	-0.00047	0.000179	375	-2.63	0.0089
m_prickling	t2	CS	Intercept	-1.1963	0.1240	12	-9.64	<.0001
m_prickling	t2	CS	tps0	0.01397	0.002449	375	5.70	<.0001
m_prickling	t2	CS	tps0*tps0	-0.00070	0.000207	375	-3.36	0.0009
m_saliva_stimul	t2	CS	Intercept	-1.1919	0.1213	12	-9.83	<.0001
m_saliva_stimul	t2	CS	tps0	0.01239	0.002374	375	5.22	<.0001
m_saliva_stimul	t2	CS	tps0*tps0	-0.00046	0.000201	375	-2.27	0.0241
o_alcohol	t2	CS	Intercept	-0.4009	0.1157	12	-3.46	0.0047
o_alcohol	t2	CS	tps0	0.01449	0.003042	375	4.76	<.0001
o_alcohol	t2	CS	tps0*tps0	-0.00327	0.000258	375	-12.67	<.0001
o_animal	t2	CS	Intercept	-1.0417	0.1078	12	-9.66	<.0001
o_animal	t2	CS	tps0	0.006298	0.002941	375	2.14	0.0329
o_animal	t2	CS	tps0*tps0	-0.00107	0.000249	375	-4.30	<.0001
o_cheese	t2	CS	Intercept	-0.2633	0.06804	12	-3.87	0.0022
o_cheese	t2	CS	tps0	0.01439	0.002364	375	6.09	<.0001
o_cheese	t2	CS	tps0*tps0	-0.00160	0.000200	375	-7.97	<.0001
o_ferment_fruit	t2	CS	Intercept	-0.5167	0.1030	12	-5.02	0.0003
o_ferment_fruit	t2	CS	tps0	0.01414	0.002768	375	5.11	<.0001
o_ferment_fruit	t2	CS	tps0*tps0	-0.00309	0.000235	375	-13.17	<.0001
o_liquorice	t2	CS	Intercept	-0.9610	0.1363	12	-7.05	<.0001
o_liquorice	t2	CS	tps0	0.02175	0.004102	245	5.30	<.0001
o_liquorice	t2	CS	tps0*tps0	-0.00159	0.000276	245	-5.75	<.0001
o_medecin	t2	CS	Intercept	-0.4424	0.1167	12	-3.79	0.0026
o_medecin	t2	CS	tps0	0.004824	0.002787	391	1.73	0.0843
o_medecin	t2	CS	tps0*tps0	-0.00297	0.000243	391	-12.23	<.0001
o_mushroom	t2	CS	Intercept	-0.6085	0.1651	12	-3.69	0.0031
o_mushroom	t2	CS	tps0	0.01443	0.003335	205	4.33	<.0001
o_mushroom	t2	CS	tps0*tps0	-0.00089	0.000338	205	-2.64	0.0089
o_roastpork	cst	CS	Intercept	-1.3178	0.09916	12	-13.29	<.0001
o_ryebread	t2	CS	Intercept	-0.4976	0.1128	12	-4.41	0.0009
o_ryebread	t2	CS	tps0	0.02234	0.002822	375	7.92	<.0001
o_ryebread	t2	CS	tps0*tps0	-0.00295	0.000239	375	-12.33	<.0001
o_treacle	t2	CS	Intercept	-1.2733	0.09753	12	-13.06	<.0001
o_treacle	t2	CS	tps0	0.01146	0.002520	391	4.55	<.0001
o_treacle	t2	CS	tps0*tps0	-0.00097	0.000220	391	-4.43	<.0001
t_burnt	t2	CS	Intercept	-1.3205	0.1360	12	-9.71	<.0001
t_burnt	t2	CS	tps0	0.007301	0.002686	375	2.72	0.0069
t_burnt	t2	CS	tps0*tps0	-0.00046	0.000228	375	-2.03	0.0428
t_salt	t2	CS	Intercept	-1.0707	0.08904	12	-12.03	<.0001
t_salt	t2	CS	tps0	0.009944	0.001891	375	5.26	<.0001
t_salt	t2	CS	tps0*tps0	-0.00037	0.000160	375	-2.30	0.0218
t_sour	t	CS	Intercept	-1.2076	0.08974	12	-13.46	<.0001
t_sour	t	CS	tps0	0.006813	0.001838	376	3.71	0.0002
t_sweet	t2	CS	Intercept	-1.2363	0.08959	12	-13.80	<.0001
t_sweet	t2	CS	tps0	0.002042	0.002535	391	0.81	0.4209
t_sweet	t2	CS	tps0*tps0	-0.00046	0.000221	391	-2.09	0.0376
ta_bitter	t	CS	Intercept	-1.3480	0.1211	12	-11.13	<.0001
ta_bitter	t	CS	tps0	0.004903	0.002239	392	2.19	0.0292
ta_maggi_lavas	t2	CS	Intercept	-1.1942	0.1053	12	-11.34	<.0001
ta_maggi_lavas	t2	CS	tps0	0.003628	0.002258	391	1.61	0.1090
ta_maggi_lavas	t2	CS	tps0*tps0	-0.00075	0.000197	391	-3.80	0.0002

Les probabilités non significatives au seuil de 0,05 sont écrites sur fond rouge

Annexe 7

Utilisation de la procédure MIXED de SAS®

La procédure MIXED du logiciel SAS® permet d'estimer des modèles contenant des effets fixes et des effets aléatoires (comme la procédure GLM) en autorisant des corrélations entre les paramètres aléatoires et des variabilités non constantes dans les données. Le modèle le plus complexe utilisé dans le chapitre 4, pour lequel l'effet fixe 'temps' est quadratique et la structure de covariance de l'effet aléatoire est RCQ, est exécuté à l'aide du programme suivant :

```
ODS output  solutionF=SolF      covparms=CovP
             tests1=Tst1       solutionR=SolR;

proc MIXED data=exemple;
  CLASS panellist;
  MODEL REP= time0 time0*time0 / htype=1 solution;
  RANDOM int time0 time0*time0 / subject=panellist solution;
run;
```

La ligne de commande "ODS output" n'est pas obligatoire mais elle permet de créer les tableaux SAS®, nommés SolF, CovP, Tst1 et SolR dans cet exemple, qui contiennent les différents résultats de l'analyse.

La variable 'panellist' est déclarée dans l'ordre CLASS (variable de classification, ou facteur). La partie fixe du modèle est déterminée dans l'ordre MODEL. REP est la variable qui contient les valeurs de répétabilité pour chaque sujet à chaque temps de mesure. "time0 time0*time0" correspond à l'effet temps quadratique. La variable temps est appelé time0 pour rappeler que les valeurs des temps doivent être centrées afin d'obtenir des estimations correctes des paramètres aléatoires (procédure MIXED, user's guide S.A.S., 2003). Dans le premier exemple, les valeurs des temps s'échelonnent entre 1 et 4,5 (années) et la valeur moyenne est de 2,75. Les valeurs transformées s'étendent donc de -1,75 à 1,75. L'option htype=1 permet de spécifier l'utilisation des carrés moyens de type I pour le test des effets fixes (les résultats sont stockés dans le fichier Tst1). L'option "solution" permet d'obtenir les estimations des paramètres des effets fixes (stockés dans le fichier SolF).

La partie aléatoire du modèle est déterminée dans l'ordre RANDOM. La structure RCQ est spécifiée avec la commande "int time0 time0*time0 / subject=panellist". Les estimations des paramètres de la structure de covariance sont stockés dans le fichier CovP. L'option "solution" donne accès aux estimations des paramètres individuels (stockés dans le fichier SolR).

Les résultats affichés dans la fenêtre de sortie de SAS® pour le descripteur od_7 (première étude) sont présentés ci-dessous :

```
----- attribute=od_7 -----
```

Iteration History			
Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	214.24460768	
1	4	199.73678559	.
2	3	199.36443773	.
3	1	199.30991603	0.00005311
4	1	199.30839397	0.00000008
5	1	199.30839179	0.00000000

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
Intercept	panellist	0.007872
time0	panellist	0.01578
time0*time0	panellist	0.1017
Residual		0.4793

Fit Statistics

-2 Res Log Likelihood	199.3
AIC (smaller is better)	207.3
AICC (smaller is better)	207.9
BIC (smaller is better)	209.6

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	-1.2963	0.1184	12	-10.95	<.0001
time0	-0.01446	0.08300	12	-0.17	0.8646
time0*time0	-0.04270	0.1125	12	-0.38	0.7109

Solution for Random Effects

Effect	panellist	Estimate	Std Err	DF	Value	Pr > t
Intercept	101	0.002614	0.08662	42	0.03	0.9761
time0	101	-0.02971	0.1104	42	-0.27	0.7893
time0*time0	101	-0.06673	0.1566	42	-0.43	0.6721
Intercept	102	-0.00123	0.08721	42	-0.01	0.9888
time0	102	0.04344	0.1226	42	0.35	0.7249
time0*time0	102	-0.4755	0.1947	42	-2.44	0.0189
.
.
.
Intercept	115	-0.01541	0.08699	42	-0.18	0.8602
time0	115	-0.02569	0.1229	42	-0.21	0.8354
time0*time0	115	0.2726	0.1951	42	1.40	0.1696

Type 1 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
time0	1	12	0.02	0.8886
time0*time0	1	12	0.14	0.7109

La table **Iteration History** retrace les itérations nécessaires au calcul de la vraisemblance estimé par REML (-2 Res Log Like). Lors de la dernière itération, -2 Res Log Like=199.3. Cette valeur est reportée dans la table **Fit statistics** dans laquelle l'indice AIC ainsi que deux critères d'informations supplémentaires (AICC and BIC) sont calculés. En utilisant l'équation (6), il est possible de vérifier que $AIC = -2 \text{ Res log likelihood} + 2*k = 199.3 + 2*4 = 207.3$, k étant le nombre de paramètres aléatoires du modèle (variances associée à l'ordonnée à l'origine, time0, time0*time0 et la variance résiduelle). Le résultat obtenu pour l'AIC du descripteur od_7 est différent de celui présenté dans le tableau 4-3 (page 99) parce que la partie fixe utilisée dans cet exemple est un effet temps quadratique alors qu'il s'agit d'un effet temps déclaré en facteur catégoriel dans les résultats du tableau 4-3. La vraisemblance correspondante était -2 Res log likelihood=189.9, donc $AIC = 189.9 + 2*4 = 197.9$.

La table **Covariance Parameter Estimates** donne les estimations des paramètres de variance-covariance associés à l'effet aléatoire sélectionné. Puisque la structure RCQ a été choisie dans cet exemple, les variances des trois paramètres liés à l'effet panéliste (Intercept, time0 and time0*time0) et la variance résiduelle sont estimées.

La table **Solution for Fixed Effects** fournit les estimations des paramètres liés à l'effet fixe et les statistiques de Student associées. Dans cet exemple, ni le paramètre de pente (estimate=-

0.01446, $p=0.8646$) ni le paramètre de courbure (estimate=-0.04270, $p=0.7109$) ne sont significativement différent de zéro.

Les estimations des paramètres individuels sont recueillies dans la table **solution for Random Effects**. Le paramètre de courbure du panéliste 102 est significativement plus faible que celui des autres sujets (estimate= -0.4755, $p=0.0189$). Les paramètres estimés pour chaque panéliste sont également stockés dans la dernière colonne du tableau 4-5.

La dernière table présente les tests des effets fixes calculés avec les sommes de carrés de type I. Les effets temps linéaires et quadratiques ne sont pas significatifs.

Afin de réaliser cette analyse avec un effet temps linéaire ou sans effet du temps (partie fixe), l'ordre **model** du programme SAS® doit être respectivement remplacé par :

```
model REP= time0 / htype=1 solution; (effet temps linéaire)
model REP= / solution; (pas d'effet temps; le fichier Tst1 n'est pas disponible)
```

Les structures de covariances RCL et CS sont exécutées en remplaçant l'ordre **random** respectivement par :

```
random int time0 / subject=panellist solution; (structure RCL)
random int / subject=panellist solution; (structure CS)
```

Afin de spécifier la structure de covariance ID, l'ordre **random** doit simplement être supprimé (le fichier **SOLR** n'est alors plus disponible).

Il est important de noter que lorsque le modèle mixte est utilisé afin de sélectionner les meilleures structures de covariance (paragraphe 2.4, chapitre 4), l'effet temps doit alors être choisi en tant que facteur catégoriel (ou qualitatif). 'time0' est alors déclaré dans l'ordre **CLASS**. Le programme SAS® utilisant la procédure MIXED devient :

```
proc MIXED data=example;
  CLASS time0 panellist;
  MODEL REP= time0 / htype=1 solution;
  RANDOM int t0 t0*t0 / subject=panellist solution;
run;
```

Puisque la variable temps doit être définie en tant que facteur catégoriel pour être utilisé dans l'ordre **MODEL**, une autre variable, strictement identique (appelée t0), doit être créée dans le tableau de données 'example' et utilisée dans l'ordre **RANDOM**.

La comparaison entre deux structures de covariance (pour des modèles avec la même partie fixe) est réalisée en calculant le logarithme du rapport de vraisemblance correspondant. Sous l'hypothèse nulle d'égalité des modèles, ce rapport suit une loi du Khi-deux, dont le nombre de degrés de liberté correspond à la différence entre les nombres de paramètres de chaque modèle. Par exemple, pour comparer les structures de covariance RCL et RCQ pour le descripteur od_7, les deux modèles sont ajustés avec un effet fixe temps déclaré en tant que facteur catégoriel. -2 Res Log Likelihood = 189.9 pour la structure RCQ et -2 Res Log Likelihood = 195.6 pour la structure RCL. Le logarithme du rapport des vraisemblances, Rloglik, est égal à la différence entre les logarithmes des vraisemblances : Rloglik = 195,6 – 189,9 = 5,7. Le nombre de degrés de liberté est égal à 4 – 3 = 1. La probabilité correspondant à une statistique de Khi-deux égale à 5,7 avec un degré de liberté est $p=0,017$. Le modèle avec la structure RCQ est donc retenu. Ces valeurs sont reportées dans le tableau 4-2 (page 97), sur la ligne correspondant au descripteur od_7 (comparaison des structures RCL et RCQ).

Annexe 8

Statistique de test du Khi-deux corrigé de Bartlett

La statistique du Khi-deux corrigé de Bartlett (1937), s'exprime de la manière suivante (d'après Brockhoff et Skovgaard, 1994) :

$$\chi^2_{Bartlett} = c \left\{ JK(I-1) \log \left(\frac{1}{J} \sum_{j=1}^J CM_e^{(j)} \right) - K(I-1) \sum_{j=1}^J CM_e^{(j)} \right\}$$

$$avec \quad c = \left\{ 1 + \frac{1}{3(J-1)} \left(\frac{J}{K(I-1)} - \frac{1}{JK(P-1)} \right) \right\}^{-1}$$

Le coefficient c est le facteur de correction de Bartlett et $CM_e^{(j)}$ représente le carré moyen de l'erreur du modèle individuel : "note = produit" pour le sujet j.

Meta-analysis of panel performances from a database of hundreds of datasets

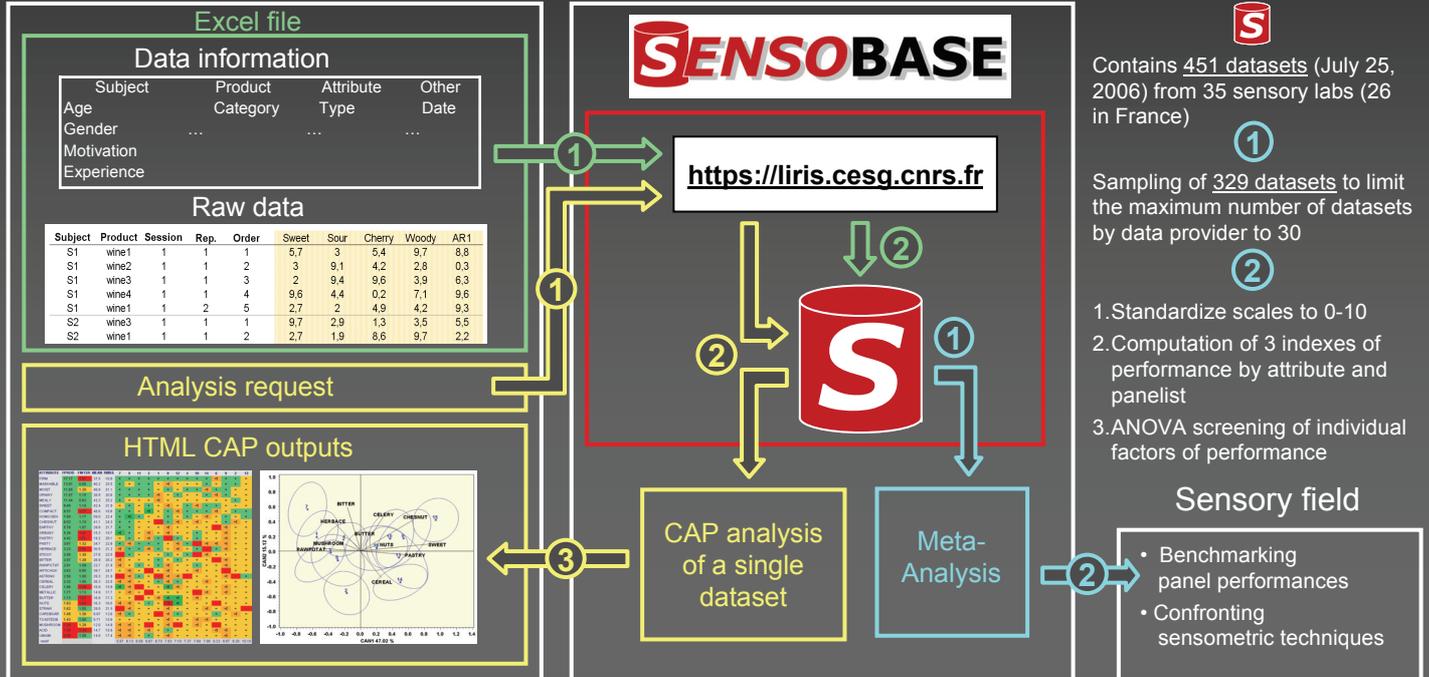
N. Pineau, P. Schlich, D. Brajon, S. Cordelle
Centre Européen des Sciences du Goût (CESG), Dijon, France (schlich@cesg.cnrs.fr)

The SensoBase

is a database containing the raw data of sensory studies (descriptive analysis with trained panelists). For each dataset brought to the SensoBase, the data provider receives a free statistical analysis (CAP) of the data. One of the aims of the SensoBase project was to gain knowledge on panelist performances. This poster reports the first results on that topic.

DATA PROVIDER

CESG - Dijon



S
Contains 451 datasets (July 25, 2006) from 35 sensory labs (26 in France)

①
Sampling of 329 datasets to limit the maximum number of datasets by data provider to 30

- ②
1. Standardize scales to 0-10
 2. Computation of 3 indexes of performance by attribute and panelist
 3. ANOVA screening of individual factors of performance

Sensory field

- Benchmarking panel performances
- Confronting sensometric techniques

Indexes of performance

- Agreement = Pearson correlation coefficient (panelist versus others)
- Discrimination = $MS_{\text{product}} / (MS_{\text{product}} + MS_{\text{residual}})$ (from individual one-way ANOVA)
- Repeatability = $\text{Root } MS_{\text{residual}}$ (from a 0-10 scale)

Weighted ANOVA of a performance index

- Index first averaged over attributes to get a single value per panelist
- Model: $\text{Index} = \text{Factor} + \text{Dataset} + \text{Factor} \times \text{Dataset}$ (for instance: Factor=AGE)
- Dataset is considered as a random effect
- Experimental unit: the panelist (n from 267 to 3,202 depending on the factor analyzed)
- Each dataset has a weight proportional to the balance of the factor level frequencies and to the total number of panelists in this dataset

Level of performances by age, gender, panelist education and sensory experience

AGE (n=3,202)	F-tests in ANOVA			AGE Level	Mean
	AGE	Dataset	AGE*Dataset		
Agreement	2.35	14.10	1.12	All	0.387
Discrimination	9.52	8.80	1.09	30-	0.615 b
				30-45	0.627 a
				45+	0.612 b
Repeatability	2.31	13.22	0.99	All	1.207

EDUCATION (n=267)	F-tests in ANOVA			EDU Level	Mean
	EDU	Dataset	EDU*Dataset		
Agreement	1.72	5.27	1.01	All	0.363
Discrimination	4.02	2.76	1.99	Secondary	0.582 b
				Higher	0.619 a
Repeatability	0.05	6.60	0.60	All	1.353

GENDER (n=2,381)	F-tests in ANOVA			GEN Level	Mean
	GEN	Dataset	GEN*Dataset		
Agreement	0.24	14.86	1.16	All	0.385
Discrimination	0.10	8.39	1.22	All	0.616
Repeatability	0.01	12.96	0.84	All	1.185

EXPERIENCE (n=486)	F-tests in ANOVA			EXP Level	Mean
	EXP	Dataset	EXP*Dataset		
Agreement	3.13	13.65	0.99	none	0.372 b
				1-3 years	0.402 a
				>3 years	0.424 a
Discrimination	4.11	13.70	0.87	none	0.616 b
				1-3 years	0.620 b
				>3 years	0.645 a
Repeatability	1.60	11.76	0.97	All	1.361

When significant (p=0.05), the F statistic is in yellow and the levels of the factor are compared. Otherwise, just the grand mean (All) is given.

Conclusion

- Although the levels of performances are very different among the datasets (dataset effect always significant), the findings below are valid across datasets, since interaction was never significant
- No gender effect on performances
- Best discrimination with the 30-45 years old group
- Education and experience in sensory profiling improve discrimination
- Experience in sensory profiling improves also agreement with the panel

Context

CAP-V3 is the last version of the CAP (Control of Assessor Performances) system that complements the classical CAP table by a CAP map and enhances earlier versions with a multivariate approach of panellist performances. It also features both univariate and multivariate product-oriented analyses, since, even while monitoring performances, the panel leader likely seeks after those analyses.

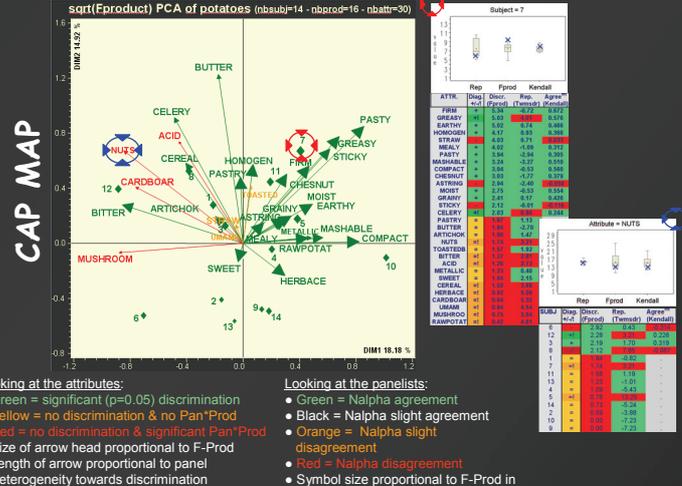
Dataset

- 16 potatoes
- 14 panellists
- 30 attributes
- 2 replications

Univariate panellist performances

CAP TABLE

attrib	fprod	finter	mean	rmse	7	5	11	3	1	9	12	4	10	14	6	9	2	13
FIRM	17.17	1.51	37.5	18.8														
MASHABLE	13.91	0.88	40.2	25.5														
MOIST	11.65	1.20	46.8	21.1														
GRAINY	11.57	1.10	26.9	20.6														
MEALY	11.44	0.83	42.3	25.2														
SWEET	8.49	1.10	42.4	21.9														
COMPACT	8.57	1.77	49.8	19.8														
HOMOGEN	7.40	1.17	59.0	22.4														
CHESNUT	6.02	1.10	41.1	24.3														
EARTHY	5.78	1.07	29.8	21.7														
GREASY	5.26	1.47	15.3	15.7														
PASTRY	4.42	1.51	19.2	20.1														
PASTY	3.67	1.32	34.7	22.8														
HERBACE	3.23	1.55	30.5	21.2														
STICKY	2.95	1.30	27.6	22.5														
BITTER	2.87	1.28	26.9	20.2														
RAWPOTA	2.81	1.09	22.7	21.9														
ARTICHOK	2.62	0.95	39.7	24.7														
ASTRING	2.59	1.90	29.3	21.9														
CEREA	2.32	1.05	28.3	22.5														
CELERY	1.86	1.38	16.9	15.9														
METALLIC	1.77	1.13	14.9	17.7														
BUTTER	1.73	1.52	16.8	17.3														
NUTS	1.53	1.65	16.3	16.5														
STRAW	1.62	1.05	29.0	21.5														
CARDBOA	1.45	1.30	8.07	13.8														
TOASTEDB	1.43	1.06	8.71	15.8														
MUSHROOM	1.38	1.29	12.0	14.9														
ACID	1.24	1.58	14.7	15.9														
UMAMI	0.99	1.08	14.8	17.4														
rankf					6.97	6.13	6.00	6.97	6.73	7.03	7.10	7.37	7.60	7.90	8.23	8.87	9.20	10.10



Multivariate panellist performances

Conf - Scor - Corr

Subject	NRV	Nalpa	NRVD
1	2.68	3.45	3.59
10	2.92	3.73	5.96
11	1.94	3.18	5.41
12	2.57	3.15	5.49
13	1.79	2.81	6.32
14	3.32	4.02	4.00
2	3.75	4.34	6.35
3	3.23	3.74	7.23
4	3.17	3.74	6.16
5	3.20	3.88	7.55
6	0.97	2.45	4.29
7	3.22	4.24	6.50
8	2.36	2.92	5.42
9	2.06	3.21	5.03
Mean	2.66	3.49	5.66

Green=Agreement
Orange=Slight Disagreement
Red=Disagreement

• NRV, Nalpa and NRVD are normalised indices of agreement in terms of product configuration, product scores and attribute correlations

• NRV and Nalpa are obtained by formulas whereas 150 permutations were drawn for NRVD

CAP - MV

subject	nattsel	f	probf	ncan	nrvrep	nrvsubj	BETA
12	15	924.61	.0011	2	2.12262	2.570	6.529
14	11	20.32	.0001	4	0.19387	3.321	6.797
2	10	11.51	.0001	4	-0.73818	3.750	6.725
3	11	10.84	.0001	3	2.41757	3.232	4.835
10	7	8.25	.0001	4	2.79518	2.923	5.195
7	11	8.14	.0005	4	1.27951	3.222	6.169
1	9	7.57	.0001	4	2.74002	2.677	5.509
6	7	6.28	.0001	3	0.62057	0.973	7.029
13	5	4.92	.0001	3	-1.28380	1.790	5.874
5	5	4.71	.0001	4	1.69260	3.200	5.812
9	3	4.69	.0001	2	1.68636	2.063	5.785
8	7	4.24	.0001	4	0.54402	2.356	5.166
4	6	3.50	.0003	3	2.33226	3.174	2.958
11	6	3.34	.0004	4	3.02012	1.941	4.582

• nattsel : nb of discriminative attributes

• F : F-prod MANOVA of the nattsel attributes

• probfI : p-value associated with F

• ncan : nb of canonical variables

• nattsel : nb of discriminative attributes

• nrvrep : NRV coefficient between replications

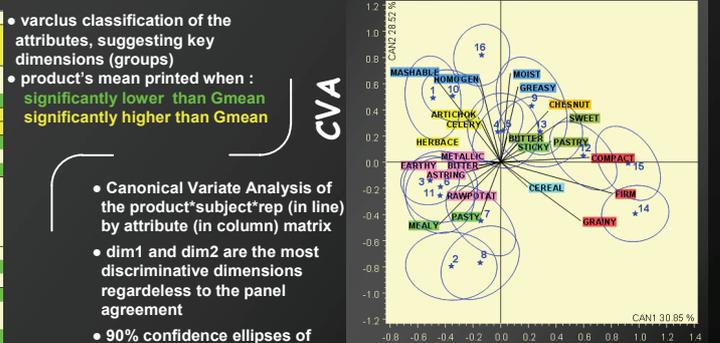
• nrvsubj : same as NRV in Conf-Scor-Corr table

• beta : ≈ nb of significative dimensions in a product*attribute PCA of the subject

Product differences

FLASH Table

ATTRIBUTE	F	PROB	GMEAN	16	3	10	1	6	2	8	5	11	4	13	9	12	15	14
FIRM	17.20	0.000	37	19-	21-	21-	22-	22-	24-					49+	63+	69+	77+	
COMPACT	8.58	0.000	49		29-	40-	27-	39-	36-					59+		60+	73+	73+
GRAINY	11.60	0.000	27	6-	14-	12-	15-	18-	52+							39+	51+	46+
PASTRY	4.43	0.000	19			10-		10-						4-	28+	44+		
SWEET	9.49	0.000	42	54+	30-	56+	27-	32-	29-	33-				54+	56+	22-	51+	69+
BUTTER	1.74	0.047	17						25+							25+	9-	
CHESNUT	6.02	0.000	41					29-	21-	21-				30-	53+	58+		63+
CEREA	2.33	0.004	26			17-												44+
ARTICHOK	2.63	0.001	40			52+			23-									20-
CELERY	1.86	0.029	17			29+												44+
HERBACE	3.23	0.000	31					20-										20-
EARTHY	5.78	0.000	30					40+	42+					43+				42+
PASTY	2.58	0.011	29											21-				47+
BITTER	2.88	0.000	27											17-				44+
RAWPOTAT	2.81	0.001	23											17-				46+
METALLIC	1.78	0.040	15															10-
PASTY	3.68	0.000	35					22-	47+	56+								23-
MEALY	11.40	0.000	42	29+	60+			54+	78+	55+	31-	54+						23-
STICKY	2.95	0.000	28	37+	16-									17-				24-
MASHABLE	13.90	0.000	40	69+		58+	66+		29-	25-	51+							26-
HOMOGEN	7.41	0.000	59	78+	72+	70+			42-	31-	69+							27-
GREASY	5.27	0.000	15	26+	7-				8-	2-	7-	26+						45-
MOIST	11.70	0.000	47	68+		63+	69+	38-	17-	29-	58+	35-						34-



All those analyses are available on <https://linis.cesg.cnrs.fr/>, the SENSORBASE website. Download a guide-line file to send your dataset in the right format and receive the analyses in a few day.