

**Positionnement Multidimensionnel : Concepts, théorie et applications**

par

Marie-Claude PARÉ

Mémoire présenté au Département de mathématiques  
en vue de l'obtention du grade de maître ès sciences (M.Sc.)

FACULTÉ DES SCIENCES  
UNIVERSITÉ DE SHERBROOKE

Sherbrooke, Québec, Canada, novembre 2007

111-1861



Library and  
Archives Canada

Published Heritage  
Branch

395 Wellington Street  
Ottawa ON K1A 0N4  
Canada

Bibliothèque et  
Archives Canada

Direction du  
Patrimoine de l'édition

395, rue Wellington  
Ottawa ON K1A 0N4  
Canada

*Your file    Votre référence*  
*ISBN: 978-0-494-49556-8*  
*Our file    Notre référence*  
*ISBN: 978-0-494-49556-8*

**NOTICE:**

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

**AVIS:**

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

---

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

■ ■ ■  
**Canada**

Le 23 novembre 2007

*le jury a accepté le mémoire de Mme Marie-Claude Paré dans sa version finale.*

*Membres du jury*

M. Ernest Monga  
Directeur  
Département de mathématiques

M. Jean Cadieux  
Membre  
S.I.M.Q.G.

M. André Mayers  
Président-rapporteur  
Département d'informatique

# SOMMAIRE

Le positionnement multidimensionnel permet d'estimer les coordonnées d'un ensemble d'objets dans un espace de dimension spécifiée. Ce travail se fait à l'aide de distances, de similarités ou de préférences basées sur les jugements des répondants. Ce mémoire décrit diverses méthodes n'utilisant que des données de distances ou de similarités (dissimilarités). Ces méthodes sont détaillées à l'aide d'exemples mettant en pratique la théorie. De plus, une section est réservée aux programmes informatiques spécialisés dans ce domaine de l'analyse de données. Une attention particulière est faite sur le programme ALSCAL provenant du logiciel SPSS.

# REMERCIEMENTS

Je remercie mon directeur de recherche, M. Ernest Monga, pour son appui autant au niveau mathématiques que financier qui fut bénéfique à mon apprentissage. Je remercie par ailleurs M. Bernard Colin du temps qu'il a alloué à l'écoute de mes questionnements et au partage de ses vastes connaissances. Je tiens aussi à remercier M. Mario Lambert, coordonnateur du département de mathématiques, pour sa grande disponibilité et sa confiance. Son support m'a grandement aidé dans le cheminement de mes années de maîtrise. Je souhaite aussi souligner l'aide financière offerte par le département de mathématiques. J'offre un merci tout spécial à mon copain, Keven Bosa, qui m'a appuyé durant toutes mes études en m'offrant un soutien autant moral que mathématique. Merci pour cette grande générosité. Pour terminer, je tiens à remercier mes parents pour leur confiance, leur support et leur amour.

Marie-Claude Paré  
Sherbrooke, Juin 2007

# TABLE DES MATIÈRES

<b>SOMMAIRE</b>	<b>iii</b>
<b>REMERCIEMENTS</b>	<b>iv</b>
<b>TABLE DES MATIÈRES</b>	<b>v</b>
<b>LISTE DES TABLEAUX</b>	<b>ix</b>
<b>INTRODUCTION</b>	<b>1</b>
0.1 Exemple introductif . . . . .	1
<b>CHAPITRE 1 — Initiation au positionnement multidimensionnel</b>	<b>4</b>
1.1 Conception de la carte des perceptions . . . . .	4
1.1.1 Quatre étapes de la conception . . . . .	4
1.1.2 Sélection de la dimensionnalité . . . . .	9
1.1.3 Identification des dimensions . . . . .	12
1.1.4 Sélection des données similaires . . . . .	14

1.2	Distances, similarités et dissimilarités . . . . .	18
1.2.1	Distances . . . . .	19
1.2.2	Similarités et dissimilarités . . . . .	20
1.3	Théorie du positionnement multidimensionnel . . . . .	23
1.3.1	Solution classique de positionnement multidimensionnel . . . . .	23
1.3.2	Dualité entre l'analyse en composantes principales et le positionnement multidimensionnel . . . . .	30
1.3.3	Propriétés optimales de la solution et indice de concordance . . .	32
1.3.4	Méthode non-métrique de positionnement multidimensionnel . . .	33
1.3.5	Indices de concordances et standardisations rotationnelles . . . . .	46
1.3.6	Analyse de l'espace conjoint . . . . .	54
1.3.7	Positionnement multidimensionnel pondéré . . . . .	57
1.3.8	Ajustement des attributs influents . . . . .	66

**CHAPITRE 2 — Le positionnement mutidimensionnel : Exemples introductifs détaillés** **72**

2.1	Représentation graphique des villes d'Angleterre . . . . .	72
2.2	L'effet du fer à cheval . . . . .	78
2.3	Positionnement de produits . . . . .	80
2.3.1	Positionnement des nouveaux concepts diététiques . . . . .	82
2.3.2	Positionnement de journaux médicaux . . . . .	86

2.3.3	Positionnement d'un service financier . . . . .	88
2.3.4	Positionnement d'un détaillant . . . . .	90
2.4	Étude des breuvages colas à l'aide de l'ajustement des attributs influents	93
<b>CHAPITRE 3 — Comparaison de programmes traitant de positionnement multidimensionnel</b>		<b>103</b>
3.1	ALSCAL . . . . .	103
3.2	Programmes spécialisés dans tous les types de problèmes de positionnement multidimensionnel . . . . .	117
3.3	Une application à l'aide de ALSCAL . . . . .	120
3.3.1	Introduction des jeux de données utilisés . . . . .	120
3.3.2	Application des jeux de données dans ALSCAL . . . . .	122
<b>CONCLUSION</b>		<b>134</b>
<b>Annexe</b>		<b>136</b>

# LISTE DES TABLEAUX

1.1	Approche de composition dans l'étude des espadrilles . . . . .	6
1.2	Données des préférences des répondants . . . . .	56
1.3	Coordonnées des répondants et des journaux . . . . .	56
1.4	Longueurs des vecteurs $\mathbf{w}_i$ . . . . .	62
1.5	Coordonnées des 4 configurations . . . . .	64
2.1	Données de chaque répondant pour chaque sorte de cola pour l'attribut chimique . . . . .	96
2.2	Coefficients de corrélation entre les attributs . . . . .	101
3.1	Tableau des distances utilisées dans les différents types de mesures . . . . .	106
3.2	Tableau 1 des caractéristiques des 6 programmes . . . . .	118
3.3	Tableau 2 des caractéristiques des 6 programmes . . . . .	120
3.4	Valeurs de SSTRESS . . . . .	123
3.5	Stress et corrélation carrée . . . . .	124
3.6	Coordonnées des stimuli . . . . .	124

3.7	Pondération des dimensions . . . . .	129
3.8	Étendues des poids des sujets . . . . .	129
3.9	Tableau des nombres d'attributs . . . . .	139

# LISTE DES FIGURES

1	Carte des perceptions pour les chiffres en code morse . . . . .	3
1.1	Graphique d'importance/performance . . . . .	8
1.2	Graphique de la mesure de stress en fonction de la dimension . . . . .	11
1.3	Carte des perceptions des voitures . . . . .	13
1.4	Carte des perceptions des voitures ainsi que leurs vecteurs attributs . . .	14
1.5	Diagramme de l'algorithme de Shepard-Kruskal . . . . .	35
1.6	Diagramme de Shepard pour l'algorithme de la violation du groupe adjacent	39
1.7	Diagramme de Shepard pour l'approche de classement d'image de Guttman	40
1.8	Diagramme d'images pour l'algorithme de la violation du groupe adjacent	41
1.9	Diagramme d'images pour l'approche de classement d'image de Guttman	42
1.10	Espace conjoint des journaux et des répondants . . . . .	57
1.11	Espace dérivé des journaux . . . . .	61
1.12	Poids des dimensions des journaux . . . . .	62
1.13	Espace multidimensionnel du premier répondant . . . . .	64

1.14	Espace multidimensionnel du deuxième répondant . . . . .	65
1.15	Espace multidimensionnel du troisième répondant . . . . .	65
1.16	Espace multidimensionnel du quatrième répondant . . . . .	66
1.17	Carte des perceptions des animaux . . . . .	67
1.18	Exemple de vecteur d'attribut provenant de l'ajustement des attributs influents . . . . .	68
2.1	Points des coordonnées calculés des 12 villes d'Angleterre . . . . .	75
2.2	Carte des 12 villes d'Angleterre . . . . .	78
2.3	Représentation graphique en deux dimensions de la matrice de Kendall .	79
2.4	Carte des perceptions des aliments et des produits diététiques . . . . .	84
2.5	Carte des perceptions avec vecteurs attributs objectifs et subjectifs . . .	85
2.6	Espace conjoint des journaux et des scénarios . . . . .	87
2.7	Espace conjoint des services financiers et des emplois . . . . .	89
2.8	Carte des perceptions des magasins . . . . .	91
2.9	Espace dérivé des breuvages colas . . . . .	94
2.10	Espace dérivé des breuvages colas avec le vecteur attribut chimique . . .	100
3.1	Espace dérivé des breuvages colas pour l'évaluation individuelle des 10 sujets dans un modèle non-pondéré . . . . .	125
3.2	Espace dérivé des breuvages colas pour l'évaluation moyenne des 10 sujets dans un modèle non-pondéré . . . . .	126

3.3	Espace dérivé des breuvages colas pour l'évaluation individuelle des 10 sujets dans un modèle pondéré . . . . .	128
3.4	Configuration de la pondération des dimensions selon les 10 sujets . . . . .	130
3.5	Graphique des étendues des poids des sujets . . . . .	131
3.6	Espace conjoint des breuvages colas et des attributs . . . . .	132
3.7	Espace conjoint des breuvages colas et des attributs . . . . .	133

# INTRODUCTION

Le positionnement multidimensionnel est une méthode d'analyse de données qui a pour but de transformer des similarités ou des préférences en distances représentées dans un espace multidimensionnel en conservant le plus fidèlement possible l'ordre des proximités d'objets. Elle permet au chercheur d'expliquer les similarités ou les différences obtenues entre ces objets à l'étude. Dans plusieurs cas, les dimensions de cette configuration peuvent être interprétées et utilisées pour analyser les données. Cette méthode de représentation graphique contribue aussi à la construction d'échelles communes à un ensemble d'attributs subjectifs. Il est à noter que les données de préférences seront exclues de ce mémoire puisque l'on se spécialisera dans les distances et les similarités. Pour bien introduire le positionnement multidimensionnel, l'exemple suivant, provenant du livre de Mardia, Kent et Bibby [1], illustre une partie de ce que l'on peut accomplir avec cette technique.

## 0.1 Exemple introductif

Dans l'exemple, 598 répondants ont comparé les dix chiffres en code morse en donnant une cote entre 0 et 100 sur leur ressemblance. Plus les chiffres se ressemblent, plus la valeur de la cote de ressemblance est élevée. Deux chiffres parfaitement identiques ont

une cote de 100. Voici la liste des dix chiffres en code morse :

0 : - - - - -	5 : . . . . .
1 : . - - - -	6 : - . . . .
2 : . . - - -	7 : - - . . .
3 : . . . . -	8 : - - - . .
4 : . . . . -	9 : - - - . .

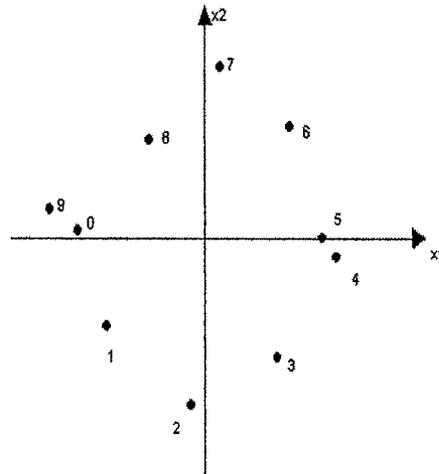
Pour l'ensemble des répondants, on obtient une matrice dite de similarités qui est donnée ci-dessous :

	1	2	3	4	5	6	7	8	9	0
1	84	62	16	6	12	12	20	37	57	52
2	62	89	59	23	8	14	25	25	28	18
3	16	59	86	38	27	33	17	46	9	9
4	6	23	38	89	56	34	24	13	7	7
5	12	8	27	56	90	30	18	10	5	5
6	12	14	33	34	30	86	65	22	8	18
7	20	25	17	24	18	65	85	65	31	15
8	37	25	46	13	10	22	65	88	58	39
9	57	28	9	7	5	8	31	58	91	79
0	52	18	9	7	5	18	15	39	79	94

Les cotes dans la matrice sont les cotes moyennes obtenues sur l'ensemble des répondants. Par exemple, la cote de similarité moyenne entre les symboles morses pour les chiffres 1 et 4 est 6. On remarque que lorsque les répondants sont amenés à comparer les

chiffres identiques, la cote de similarité est très élevée. La figure 1 représente la carte des perceptions pour le code morse créée à l'aide du positionnement multidimensionnel.

Figure 1 – Carte des perceptions pour les chiffres en code morse



Plus les points sont éloignés les uns des autres, plus les objets sont différents. On remarque que les chiffres 4 et 5 sont très similaires ainsi que 9 et 0. Par contre, 1 et 6 ne se ressemblent pas du tout. Ce résultat semble bien normal car si on regarde de près les dix chiffres en code morse, on voit une certaine continuité de 0 à 1, de 1 à 2 et ainsi de suite. Notons que sur la carte des perceptions, les chiffres se suivent dans le sens inverse des aiguilles d'une montre.

# CHAPITRE 1

## Initiation au positionnement multidimensionnel

### 1.1 Conception de la carte des perceptions

#### 1.1.1 Quatre étapes de la conception

La carte des perceptions est l'espace multidimensionnel qui montre le positionnement relatif de tous les objets selon les attributs à l'étude. On donne ci-après les quatre étapes pour construire et interpréter cette carte.

Notons que ces étapes s'appliquent à toutes les méthodes d'analyse de données qui ont pour but de concevoir une carte de perceptions et non spécifiquement pour la positionnement multidimensionnel.

*1<sup>ère</sup>* étape : Identification des objets à évaluer.

Le chercheur identifie tous les objets (aussi nommés stimuli) à évaluer en faisant bien attention à n'oublier aucun objet important ou à ne pas en inclure un qui est inapproprié. L'inclusion d'un objet inapproprié pourrait modifier grandement la carte des perceptions résultante, ce qui peut avoir pour conséquence de fausser l'analyse subséquente.

**Exemple 1.** *Dans une étude sur les grands magasins de Sherbrooke sont sélectionnés les objets suivants : LaBaie, Sears, Zellers, Stokes. Un magasin inapproprié, Stokes, a été inclus et un magasin important, Wal-Mart, a été omis.*

2<sup>ème</sup> étape : Sélection de l'approche.

Il existe deux sortes d'approches dans le but de collecter les données d'analyse :

1) L'approche de composition :

L'évaluation est faite directement par rapport aux attributs. Ainsi, chaque répondant évalue chaque objet, attribut par attribut.

2) L'approche de décomposition :

Dans cette approche, le répondant donne son avis de façon générale sur les objets en question selon les attributs sélectionnés. Il n'évalue pas chaque objet selon chaque attribut comme dans l'approche de composition, mais plutôt chaque objet en se basant sur l'ensemble des attributs.

L'exemple suivant expose les différences entre ces deux approches.

**Exemple 2.** *On souhaite comparer des marques d'espadrilles.*

*Objets : 3 marques d'espadrilles*

*- Nike*

- Adidas
- New Balance

*Attributs :*

- Confort
- Durabilité
- Esthétique

*1- L'approche de composition :*

*Chaque répondant dit s'il croit que l'espadrille possède l'attribut ou non (binaire 0-1). Ensuite, les résultats sont compilés pour tous les répondants (aussi appelés sujets). Supposons que, dans cet exemple, il y ait en tout 18 répondants. Les résultats sont présentés dans le tableau 1.1.*

Tableau 1.1 – Approche de composition dans l'étude des espadrilles

Attributs/Marques	Nike	Adidas	New Balance
Confort	16	15	9
Durabilité	14	13	11
Esthétique	13	18	10

*Si on analyse ce tableau, on observe que 16 personnes croient que Nike possède l'attribut confort et 11 répondants sur 18 trouvent que New Balance dure longtemps.*

*2- L'approche de décomposition :*

*Le répondant fait les comparaisons de similarités de façon générale en se basant sur les trois attributs. Pour illustrer nos résultats, on utilise une matrice à tableau croisé.*

	<i>Nike</i>	<i>Adidas</i>	<i>NewBalance</i>
<i>Nike</i>	6	3	5
<i>Adidas</i>	3	6	2
<i>NewBalance</i>	5	2	6

- 6 : *Similaires*

- 1 : *Non similaires*

*La matrice précédente n'est donnée que par un seul répondant. Pour évaluer les données de tous les sujets, on n'a qu'à faire les moyennes de tous les tableaux. On remarque que pour ce répondant en particulier, Nike et New Balance sont vraiment similaires tandis qu'Adidas et New Balance ne le sont pas du tout.*

3<sup>ème</sup> étape : Sélection des techniques de base appropriées dans l'approche sélectionnée.

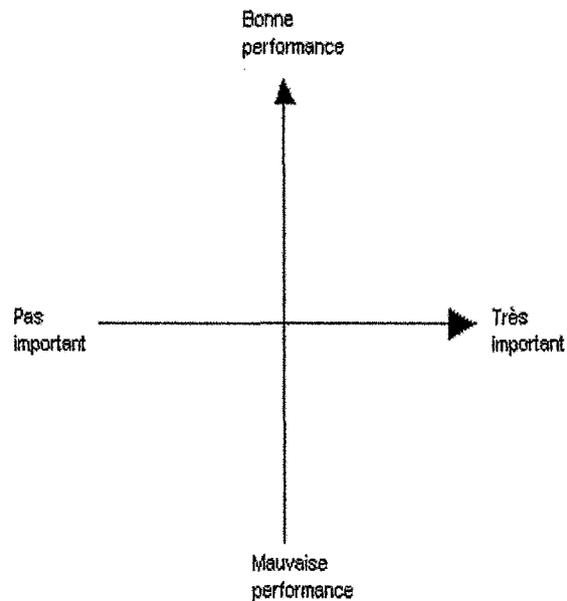
La sélection des techniques de base se fait en fonction de l'approche sélectionnée à l'étape précédente. Ces techniques sont introduites ci-après.

1) Techniques utilisées dans l'approche de composition :

Les trois techniques de composition les plus utilisées sont :

1. L'approche graphique ;
2. Les techniques statistiques multivariées conventionnelles ;
3. Les méthodes spécialisées pour la conception de la carte des perceptions.

Figure 1.1 – Graphique d’importance/performance



1. L’approche graphique :

Ce sont des analyses de graphiques comme par exemple celui d’importance/performance représenté à la figure 1.1. Ces analyses reposent sur le jugement du chercheur et sur les représentations unidimensionnelles et bidimensionnelles des objets.

2. Techniques statistiques multivariées conventionnelles :

Ces techniques sont utilisées lorsque l’étude comporte beaucoup d’attributs et que l’on souhaite représenter les objets dans ces dimensions. On peut citer par exemple les analyses discriminantes et factorielles.

3. Méthodes spécialisées pour la conception de la carte des perceptions :

Ces techniques servent à développer une carte des perceptions avec des données qualita-

tives ou provenant d'échelles nominales comme dans l'analyse des correspondances.

## 2) Techniques utilisées dans l'approche de décomposition :

Le choix spécifique de la méthode se fait selon la nature des données (similaires ou de préférences) et selon le fait que l'on veut une carte des perceptions individuelle ou conjointe. Les cartes de perceptions conjointes seront détaillées dans la section portant sur l'analyse de l'espace conjoint (section 1.3.6).

### 4<sup>ème</sup> étape : Interprétation des résultats :

Une fois qu'on a la carte des perceptions, les interprétations des deux approches se font de façons distinctes.

Pour l'approche de composition, puisque les positions des objets sont entièrement définies par les attributs spécifiés par le chercheur, la carte des perceptions ne tient pas compte des autres mesures de perception.

Pour l'approche de décomposition, la description des dimensions de perceptions et leurs correspondances aux attributs objectifs sont les éléments clés de l'interprétation des résultats.

## 1.1.2 Sélection de la dimensionnalité

Cette sous-section expose plusieurs techniques utilisées pour sélectionner le nombre idéal de dimensions. Les trois techniques les plus utilisées sont

1. l'indice de concordance ;
2. la mesure de stress ;

### 3. l'évaluation subjective.

#### 1. L'indice de concordance :

L'indice de concordance est un indice basé sur le carré d'un coefficient de corrélation qui peut être interprété comme un indicateur de proportion des variances des disparités. C'est des disparités les différences entre les distances qui représentent les similarités et les distances qui sont fournies par les répondants qui peuvent être justifiées par le positionnement multidimensionnel. En terme plus simple, c'est une mesure qui indique à quel degré coïncident les données avec le modèle de positionnement multidimensionnel. Ainsi, on peut déterminer la meilleure dimensionnalité. Les mesures considérées comme acceptables sont 0.60 et plus.

#### 2. La mesure de stress :

La mesure de stress est une mesure de proportion de variances des disparités qui ne sont pas expliquées par des modèles de positionnement multidimensionnel. La mesure de stress la plus utilisée est celle de Kruskal.

$$\sqrt{\frac{\sum_{i=1}^n \sum_{j=1}^n (d_{ij} - \hat{d}_{ij})^2}{\sum_{i=1}^n \sum_{j=1}^n (d_{ij} - \bar{d})^2}} \quad (1.1)$$

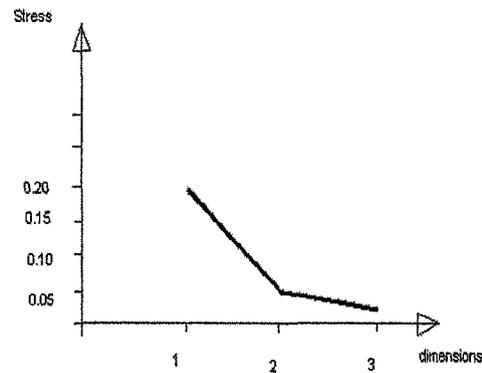
$d_{ij}$  : distances entre les stimuli i et j pour le nombre de dimensions en question

$\hat{d}_{ij}$  : distances entre les jugements de similarité donnés par les répondants

$$\bar{d} : \frac{\sum_{i=1}^n \sum_{j=1}^n d_{ij}}{n^2}$$

On sait que plus  $d_{ij}$  est près de  $\hat{d}_{ij}$ , meilleur est la représentation dans la dimension en question.

Figure 1.2 – Graphique de la mesure de stress en fonction de la dimension



Donc, plus la mesure de stress est basse, plus le nombre de dimensions est approprié. Pour faire une bonne analyse de la mesure de stress, on utilise un graphique comme celui de la figure 1.2. Habituellement, on dit que le nombre idéal de dimension se situe au niveau du coude dans le graphique que l'on peut remarquer à la dimension 2 de la figure en exemple. D'une dimension à l'autre, une droite est créée et sa pente diffère à chaque section. Le coude cherché est créé par un changement radical de pente, ce qui signifie que la diminution de stress d'une dimension à l'autre est beaucoup moins grande que précédemment. En prenant le graphique de la figure 1.2, on remarque que l'on ne gagnera pas assez de précision dans la représentation en trois dimensions pour que ça vaille la peine d'augmenter le degré de difficulté de l'interprétation du graphique. En somme, l'ajout de la troisième dimension possède un coût plus élevé que la valeur de l'information que l'analyste reçoit en retour. On donnera plus de détail sur la mesure de

stress dans la section du positionnement multidimensionnel non-métrique (section 1.3.4).

### 3. L'évaluation subjective :

On procède à une évaluation subjective pour déterminer le nombre idéal de dimensions en établissant un équilibre entre la difficulté à interpréter les résultats si la dimension est trop grande et la justesse de la représentation. De plus, si la problématique a déjà fait l'objet d'une étude antérieure, il est possible, par souci de comparaison ou de validité numérique, de s'inspirer des autres études pour déterminer le nombre de dimensions.

### 1.1.3 Identification des dimensions

Afin d'identifier les dimensions, il n'existe pas de procédure universelle avec le positionnement multidimensionnel. Les méthodes d'identification se répartissent en deux : les méthodes basées sur les procédures subjectives et celles basées sur les procédures objectives.

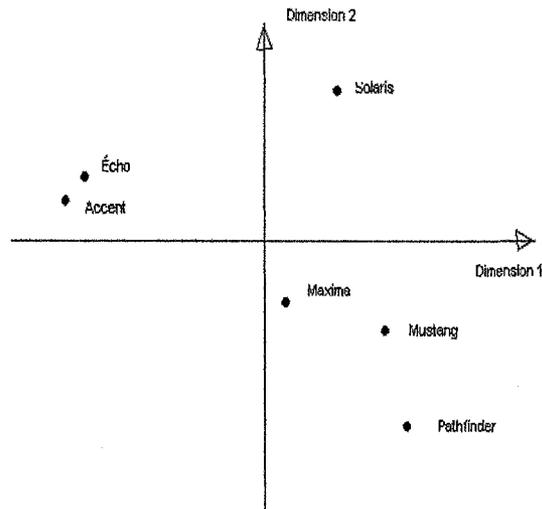
⇒ Procédures subjectives ;

Puisque la mathématique ne voit pas la réalité qu'elle représente, l'interprétation peut dépendre du jugement humain. Le chercheur peut, par lui-même, définir les dimensions, mais parfois, il doit se référer à un spécialiste s'il ne possède pas de bonnes connaissances dans le domaine de l'étude.

**Exemple 3.** *La figure 1.3 représente une carte des perceptions où sont placées différentes voitures. En évaluant cette carte des perceptions, on peut déterminer que la dimension 1 est le prix d'achat et la dimension 2 est écologique.*

⇒ Procédures objectives ;

Figure 1.3 – Carte des perceptions des voitures

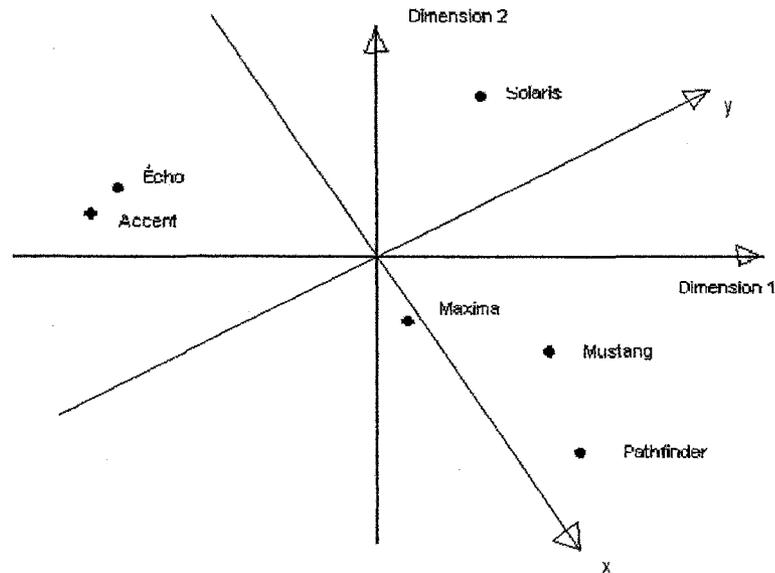


Il existe plusieurs méthodes formalisées. La plus utilisée est l'ajustement des attributs influents (property fitting). Elle collecte les évaluations des attributs de chaque objet et trouve la meilleure correspondance de chaque attribut. Le but est d'identifier les attributs déterminants dans les jugements de similarités fait par les répondants. Les indices de concordance sont ensuite déterminées pour chaque attribut. L'analyste peut ensuite identifier quel attribut décrit le mieux les positions de perceptions et lesquels sont illustratifs des dimensions. On reviendra sur la notion d'ajustement des attributs influents dans la section 1.3.8.

Une des caractéristiques du positionnement multidimensionnel est le fait que l'on a une invariance par rotation, par réflexion et par translation. Il peut donc s'avérer très utile d'utiliser un de ces types de transformation pour permettre à un attribut particulier de

se retrouver sur un axe selon ce que le chercheur veut privilégier. Cette façon de faire est une des méthodes qui permet d'identifier les dimensions.

Figure 1.4 – Carte des perceptions des voitures ainsi que leurs vecteurs attributs



**Exemple 4.** *L'ajustement des attributs influents donne donc le sens des vecteurs représentant les attributs confort ( $y$ ) et coût d'essence ( $x$ ). Ces vecteurs sont présentés dans la figure 1.4. En appliquant une rotation des axes de  $40^\circ$  dans le sens horaire de la figure 1.3, l'ordonnée devient le confort et l'abscisse le coût d'essence.*

#### 1.1.4 Sélection des données similaires

Pour obtenir la perception des similarités des répondants, les trois procédures habituellement utilisées sont

1. la comparaison par paires d'objets ;
2. la combinaison croisée ;
3. la mesure dérivée.

### 1. Comparaison par paires d'objets :

Les répondants doivent donner une cote de similarité à chaque paire d'objets. L'analyse de positionnement multidimensionnel peut se faire à partir des données de chaque individu ou à partir de la moyenne des cotes de similarités affectée à chaque paire de stimuli. L'exemple suivant applique la sélection de données similaires par comparaisons de paires d'objets.

**Exemple 5.** *On a 4 objets : A, B, C, D.*

*Ainsi, on a 6 paires d'objets : AB, AC, AD, BC, BD, CD.*

*Données non-métriques :*

*1 : Non similaires*

*6 : Similaires*

$$\begin{array}{l}
 AB=5 \quad AD=6 \quad BD=4 \\
 AC=3 \quad BC=2 \quad CD=1
 \end{array}$$

*Données métriques :*

*Échelle de 1 à 10 où 10 représente la similarité.*

$$\begin{array}{l}
 AB=8 \quad AD=10 \quad BD=7 \\
 AC=5 \quad BC=3 \quad CD=2
 \end{array}$$

Dans une échelle non-métrique, il faut ranger de 1 à n (n : le nombre de paires d'objets) toutes les paires d'objets. Comme dans l'exemple ci-haut, on a 6 paires d'objets. Ainsi, 1

signifie que les deux objets en question sont peu similaires contrairement à 6 qui montre une grande ressemblance. Dans une échelle métrique, il faut associer un degré de similarité à chaque paire de stimuli. Cette partie est affichée dans la deuxième section de l'exemple où l'on retrouve une échelle de 1 à 10 où 1 signifie encore dissimilaire mais la similarité est représentée par 10 et non par 6.

## **2. Combinaison croisée :**

Si le répondant juge qu'il y a similarité entre les objets, il les combine ensemble. Cette méthode ne permet qu'un agrégat de similarités. Il s'agit donc d'une technique binaire. Les objets sont similaires ou pas. Les réponses de tous les répondants sont combinées pour créer une matrice des similarités. L'exemple ci-après explique les combinaisons croisées dans une étude constituée de 14 répondants et 6 stimuli.

**Exemple 6.** *Les 14 répondants possèdent chacun 6 cartons représentant les groupes suivants :*

- *A. Nirvana*
- *B. Our Lady Peace*
- *C. New Kids on the Block*
- *D. Gun's and Roses*
- *E. Beatles*
- *F. Bad Religion*

*Ils combinent deux par deux tous les groupes qu'ils trouvent similaires. Voici la matrice des données résultante :*

	A	B	C	D	E	F
A	-	12	5	8	2	9
B	12	-	4	10	2	7
C	5	4	-	1	4	5
D	8	10	1	-	2	4
E	2	2	4	2	-	5
F	9	7	5	4	5	-

*On remarque qu'il y a donc 8 répondants qui croient que Nirvana et Gun's and Roses se ressemblent et seulement 2 pour Our Lady Peace et les Beatles.*

### **3. Mesure dérivée :**

Les similarités sont typiquement basées sur la cote donnée au stimulus par le répondant. Cette technique de collecte de données sera surtout utilisée pour l'ajustement des attributs influents et l'analyse de l'espace conjoint que l'on développera dans les sous-sections 1.3.8 et 1.3.6 respectivement. L'exemple 7 expose cette technique de collection de données.

**Exemple 7.** *Les 4 objets à l'étude sont :*

*A. Salade de poulet;*

*B. Filet de saumon;*

*C. Big Mac;*

*D. Fondue parmesan.*

*Les 3 échelonnements pour l'évaluation sont :*

1	10
<i>Moins calorique</i>	<i>Plus calorique</i>
<i>Moins protéiné</i>	<i>Plus protéiné</i>
<i>Moins consistant</i>	<i>Plus consistant</i>

On obtient ainsi pour chaque répondant une matrice  $3 \times 4$  comme par exemple :

$$\begin{array}{l} \text{Calories} \\ \text{Protéines} \\ \text{Consistance} \end{array} \begin{pmatrix} A & B & C & D \\ 2 & 2 & 10 & 7 \\ 8 & 8 & 4 & 2 \\ 7 & 8 & 5 & 3 \end{pmatrix}$$

On remarque ainsi que le plus calorique est le Big Mac mais qu'il contient peu de protéines. Les meilleurs repas santé, selon le tableau de ce répondant, sont donc le filet de saumon et la salade de poulet.

Toutes les matrices des répondants sont ensuite évaluées pour créer les mesures des similarités.

Les trois hypothèses les plus importantes pour les mesures dérivées sont :

1. Le chercheur sélectionne les dimensions appropriées pour l'échelonnement sur lequel on évalue les objets.
2. L'échelle peut être pondérée (de façon égale ou inégale) pour réaliser les données similaires pour un sujet ou un groupe de sujets.
3. Même si le poids de l'échelle peut être déterminé, tous les individus ont le même poids.

## 1.2 Distances, similarités et dissimilarités

Cette section est consacrée à la définition et la démonstration de certains résultats liés aux distances, aux similarités et aux dissimilarités. On se limitera aux définitions dans

$\mathbb{R}^p$ .

### 1.2.1 Distances

**Définition 1.** Soient  $P$  et  $Q$  deux points dans  $\mathbb{R}^p$ . On dira que  $d(P, Q)$  est une distance si et seulement si les trois conditions suivantes sont satisfaites :

- (a) symétrie :  $d(P, Q) = d(Q, P)$  ;
- (b) non-négativité :  $d(P, Q) \geq 0$  ;
- (c) marque d'identification :  $d(P, P) = 0$ .

**Définition 2.** Pour tout point  $P, Q$  et  $R$  de  $\mathbb{R}^p$ , une distance est métrique si :

- $d(P, Q) = 0 \Leftrightarrow P = Q$  ;
- $d(P, Q) \geq d(P, R) + d(R, Q)$ .

**Exemple 8.** - Distance euclidienne :

Soient  $\mathbf{x}_i$  et  $\mathbf{x}_j$  éléments de  $\mathbb{R}^p$  tels que  $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})'$  et  $\mathbf{x}_j = (x_{j1}, x_{j2}, \dots, x_{jp})'$ .

La distance euclidienne entre  $\mathbf{x}_i$  et  $\mathbf{x}_j$  est

$$E_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|_2 = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2}.$$

- Distance de Karl-Pearson :

Dans les mêmes conditions que pour la distance euclidienne, la distance de Karl-Pearson est basée sur un ensemble de  $n$  points  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$  et est donnée par :

$$K_{ij} = \sqrt{\sum_{k=1}^p \frac{(x_{ik} - x_{jk})^2}{s_k^2}}$$

où  $s_k^2$  est la variance de la  $k^{\text{ième}}$  variable et est donnée par  $s_k^2 = \frac{1}{n-1} \sum_{i=1}^n (x_{ik} - \bar{x}_k)^2$  et où  $\bar{x}_k = \frac{1}{n} \sum_{i=1}^n x_{ik}$  est la moyenne sur les  $n$  points de la  $k^{\text{ième}}$  composante.

- Distance de Mahalanobis :

Soient  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$   $n$  points de  $\mathbb{R}^p$ . Désignons par  $\mathbf{S} = \frac{1}{n} \sum_{r=1}^n (\mathbf{x}_r - \bar{\mathbf{x}})(\mathbf{x}_r - \bar{\mathbf{x}})'$  la matrice des covariances empiriques des  $n$  points où  $\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$  est le vecteur moyenne. Ainsi, l'élément  $s_{ij}$  de  $\mathbf{S}$  indique la covariance entre les points  $i$  et  $j$ . La distance de Mahalanobis entre les points  $\mathbf{x}_i$  et  $\mathbf{x}_j$  est donnée par :

$$M_{ij} = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)' \mathbf{S}^{-1} (\mathbf{x}_i - \mathbf{x}_j)}.$$

## 1.2.2 Similarités et dissimilarités

**Définition 3.** Soient  $A$  et  $B$  deux points dans  $\mathbb{R}^p$ . On dit que  $s(A, B)$  est une similarité si et seulement si les trois conditions suivantes sont satisfaites :

- (a) symétrie :  $s(A, B) = s(B, A)$  ;
- (b) positivité :  $s(A, B) \geq 0$  ;
- (c)  $s(A, B)$  augmente quand la similarité entre  $A$  et  $B$  augmente.

**Définition 4.** Supposons qu'on ait  $p$  attributs pour  $n$  objets. La matrice  $\mathbf{X}$  de taille  $n \times p$  représente les données d'absences/présences des  $p$  attributs sur ces  $n$  objets. Cette matrice  $\mathbf{X}$  est appelé la matrice d'incidence ( $x_{ij} = 0$  ou  $1$ ).

**Exemple 9.** Désignons par  $\mathbf{C}$  la matrice des similarités et par  $\mathbf{X}$  la matrice d'incidence. On peut déduire des matrices de similarités à partir de matrices d'incidence à l'aide des opérations suivantes :

-  $\mathbf{C}_1 = \frac{\mathbf{X}\mathbf{X}'}{p}$

$$- C_2 = \frac{XX' + (F - X)(F - X)'}{p}$$

où  $F$  est une matrice qui possède la même taille que  $X$ , c'est-à-dire  $n \times p$  et qui est uniquement composée de 1.

**Exemple 10.** Soit  $X$  ( $3 \times 5$ ) une matrice d'incidence définie telle que :

$$X = \begin{matrix} & U & V & W & X & Y \\ \begin{matrix} A \\ B \\ C \end{matrix} & \begin{pmatrix} 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \end{matrix}$$

où  $A, B, C$  sont les 3 individus et  $U, V, W, X$  et  $Y$  sont les 5 attributs. À partir de cette matrice d'incidence, créons les matrices de similarités  $C_1$  et  $C_2$ .

$$\begin{aligned} C_1 &= \frac{XX'}{p} \\ &= \frac{XX'}{5} \\ &= \frac{1}{5} \begin{pmatrix} 3 & 3 & 1 \\ 3 & 4 & 1 \\ 1 & 1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 3/5 & 3/5 & 1/5 \\ 3/5 & 4/5 & 1/5 \\ 1/5 & 1/5 & 1/5 \end{pmatrix} \end{aligned}$$

Les éléments de la matrice  $C_1$  représentent la proportion d'attributs que les individus possèdent tous les deux. En examinant  $C_1$ , on peut conclure que ce sont les individus  $A$  et  $B$  qui se ressemblent le plus.

Notons que  $C_3 = XX'$  est aussi une matrice de similarité où ces éléments indiquent une parfaite similitude lorsqu'ils atteignent 5 et une différence absolue lorsqu'ils atteignent 0.

Pour trouver  $C_2$ , calculons d'abord  $F-X$ .

$$\begin{aligned} F - X &= \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} - \begin{pmatrix} 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 \end{pmatrix} \end{aligned}$$

$$\begin{aligned} C_2 &= \frac{XX' + (F - X)(F - X)'}{p} \\ &= \begin{pmatrix} 3/5 & 3/5 & 1/5 \\ 3/5 & 4/5 & 1/5 \\ 1/5 & 1/5 & 1/5 \end{pmatrix} + \begin{pmatrix} 2/5 & 1/5 & 2/5 \\ 1/5 & 1/5 & 1/5 \\ 2/5 & 1/5 & 4/5 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 4/5 & 3/5 \\ 4/5 & 1 & 2/5 \\ 3/5 & 2/5 & 1 \end{pmatrix} \end{aligned}$$

Dans le cas de  $C_2$ , en plus de calculer les proportions d'attributs que les deux individus ont en commun, on calcule les proportions d'attributs que ni un ni l'autre ne possèdent. Ainsi, on est assuré que la diagonale est uniquement composée de 1, ce qui indique une similitude parfaite dans le cas présent.

Notons que  $C_4 = XX' + (F - X)(F - X)'$  est aussi une mesure de similarités où ces éléments indiquent une parfaite similitude lorsqu'ils atteignent 5 et une différence absolue lorsqu'ils atteignent 0.

**Propriété 1.** Pour toute matrice de similarités  $C$  on a

$$c_{rs} = c_{sr}$$

et

$$c_{rs} \leq c_{rr}$$

$\forall r, s.$

Dans la majorité des cas, les algorithmes de positionnement multidimensionnel utilisent des matrices de distances ou de dissimilarités comme données d'entrées. Ainsi, il faut convertir les similarités en distances ou en dissimilarités. Les deux propriétés suivantes indiquent comment procéder.

**Propriété 2.** *À partir d'une matrice de similarités  $C$ , on obtient une matrice de distances  $D$  d'éléments  $d_{ij}$  en posant :*

$$d_{ij} = \sqrt{c_{ii} + c_{jj} - 2c_{ij}} \geq 0$$

**Propriété 3.** *Soient  $x_i$  et  $x_j$  deux points dans  $\mathbb{R}^p$ . On calcule la dissimilarité  $\delta_{ij}$  entre les points  $x_i$  et  $x_j$  comme suit :*

$$\delta_{ij}^2 = c - c_{ij}$$

*où  $c_{ij}$  est la similarité entre les points  $x_i$  et  $x_j$  et  $c$  est la valeur maximale que peut prendre la mesure de similarité.*

## 1.3 Théorie du positionnement multidimensionnel

Considérons une matrice de distances  $D$  de dimension  $n \times n$  dont les éléments sont les  $d_{ij}$ . En partant de cette matrice de distances  $D$ , le but du positionnement multidimensionnel est de trouver les points  $P_1, P_2, \dots, P_n$  dans  $k$  dimensions tels que si  $\hat{d}_{rs}$  représente la distance entre  $P_r$  et  $P_s$ , alors  $\hat{D}$  est similaire à  $D$  dans un certain sens. En pratique, on se limite à un nombre de dimensions  $k = 1, 2$  ou  $3$  pour faciliter la représentation.

### 1.3.1 Solution classique de positionnement multidimensionnel

**Définition 5.** *Une matrice des distances  $D$  est dite euclidienne s'il existe une configuration de points dans un espace euclidien quelconque où les distances d'interpoints sont*

données par  $D$ . Ainsi, pour un certain  $p$ , il existe des points  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n \in \mathbb{R}^p$  tels que

$$d_{rs}^2 = (\mathbf{x}_r - \mathbf{x}_s)'(\mathbf{x}_r - \mathbf{x}_s)$$

Introduisons quelques notations utiles à la théorie de cette section. Pour toute matrice de distances  $D$  composée des éléments  $d_{rs}$ , soit

$$\mathbf{A} = (a_{rs}) = \left(-\frac{1}{2}d_{rs}^2\right)$$

et soit

$$\mathbf{B} = \mathbf{H}\mathbf{A}\mathbf{H}$$

où  $\mathbf{H} = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'$  est la matrice centrée  $n \times n$  et où  $\mathbf{1}$  est le vecteur unitaire  $n \times 1$ .

**Théorème 1.** *Soit  $D$  une matrice de distances et  $\mathbf{B} = \mathbf{H}\mathbf{A}\mathbf{H}$ , alors  $D$  est euclidienne si et seulement si  $\mathbf{B}$  est semi-définie positive.*

Ainsi, on a les résultats suivants :

1. Si  $D$  est une matrice euclidienne des distances d'interpoints pour une configuration  $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n)'$  alors  $b_{rs} = (\mathbf{z}_r - \bar{\mathbf{z}})'(\mathbf{z}_s - \bar{\mathbf{z}})$  et  $\mathbf{B} = \mathbf{H}\mathbf{A}\mathbf{H}$  devient  $\mathbf{B} = (\mathbf{H}\mathbf{Z})(\mathbf{H}\mathbf{Z})'$ . Ainsi,  $\mathbf{B} \geq 0$

2. Inversement, si  $\mathbf{B} \geq 0$  de rang  $p$ , alors, la configuration correspondante à  $\mathbf{B}$  peut être construite comme suit :

Soient  $\lambda_1 > \lambda_2 > \dots > \lambda_p > 0$  les valeurs propres strictement positive de  $\mathbf{B}$  avec les vecteurs propres correspondants  $\mathbf{X} = (\mathbf{x}_{(1)}, \mathbf{x}_{(2)}, \dots, \mathbf{x}_{(p)})$  normalisés tels que  $\mathbf{x}'_{(i)}\mathbf{x}_{(i)} = \lambda_i$  ( $i = 1, 2, \dots, p$ ). Ainsi, le point  $P_r$  dans  $\mathbb{R}^p$  avec les coordonnées  $\mathbf{x}_r = (x_{r1}, x_{r2}, \dots, x_{rp})'$  ( $\mathbf{x}_r$  est la  $r^{\text{ième}}$  ligne de  $\mathbf{X}$ ) a des distances entre les points données par  $D$  pour une configuration qui a comme centre de gravité  $\bar{\mathbf{x}} = 0$  et  $\mathbf{B}$  représente la matrice des produits scalaires pour cette configuration ( $b_{rs} = (\mathbf{x}_r - \bar{\mathbf{x}})'(\mathbf{x}_s - \bar{\mathbf{x}})$ ).

*Démonstration.* ( $\implies$ )

On sait que  $\mathbf{D}$  est une matrice des distances euclidiennes.

Supposons que

$$\begin{aligned} d_{rs}^2 &= -2a_{rs} \\ &= (\mathbf{z}_r - \mathbf{z}_s)'(\mathbf{z}_r - \mathbf{z}_s) \\ &= \sum_{i=1}^n (z_{ri} - z_{si})^2 \end{aligned}$$

On peut écrire

$$\begin{aligned} \mathbf{B} &= \mathbf{H}\mathbf{A}\mathbf{H} \\ &= \mathbf{A} - \frac{1}{n}\mathbf{A}\mathbf{J} - \frac{1}{n}\mathbf{J}\mathbf{A} + \frac{1}{n^2}\mathbf{J}\mathbf{A}\mathbf{J} \end{aligned}$$

où  $\mathbf{J}=\mathbf{1}\mathbf{1}'$ . Ainsi,

$$\frac{1}{n}\mathbf{A}\mathbf{J} = \begin{pmatrix} \bar{a}_{1.} & \cdots & \bar{a}_{1.} \\ \vdots & \ddots & \vdots \\ \bar{a}_{n.} & \cdots & \bar{a}_{n.} \end{pmatrix}$$

$$\frac{1}{n}\mathbf{J}\mathbf{A} = \begin{pmatrix} \bar{a}_{.1} & \cdots & \bar{a}_{.n} \\ \vdots & \ddots & \vdots \\ \bar{a}_{.1} & \cdots & \bar{a}_{.n} \end{pmatrix}$$

$$\frac{1}{n^2}\mathbf{J}\mathbf{A}\mathbf{J} = \begin{pmatrix} \bar{a}_{..} & \cdots & \bar{a}_{..} \\ \vdots & \ddots & \vdots \\ \bar{a}_{..} & \cdots & \bar{a}_{..} \end{pmatrix}$$

où  $\bar{a}_{r.} = \frac{1}{n} \sum_{s=1}^n a_{rs}$ ,  $\bar{a}_{.s} = \frac{1}{n} \sum_{r=1}^n a_{rs}$  et  $\bar{a}_{..} = \frac{1}{n^2} \sum_{r=1}^n \sum_{s=1}^n a_{rs}$

Ainsi,

$$b_{rs} = a_{rs} - \bar{a}_{r.} - \bar{a}_{.s} + \bar{a}_{..} \tag{1.2}$$

Puisque  $a_{rs} = \frac{-1}{2}(\mathbf{z}_r - \mathbf{z}_s)'(\mathbf{z}_r - \mathbf{z}_s)$ , alors,

$$\begin{aligned}\bar{a}_{r.} &= \frac{-1}{2n} \sum_{i=1}^n (\mathbf{z}_r - \mathbf{z}_i)'(\mathbf{z}_r - \mathbf{z}_i) \\ \bar{a}_{.s} &= \frac{-1}{2n} \sum_{j=1}^n (\mathbf{z}_j - \mathbf{z}_s)'(\mathbf{z}_j - \mathbf{z}_s) \\ \bar{a}_{..} &= \frac{-1}{2n^2} \sum_{i=1}^n \sum_{j=1}^n (\mathbf{z}_i - \mathbf{z}_j)'(\mathbf{z}_i - \mathbf{z}_j).\end{aligned}$$

Ainsi,

$$b_{rs} = (\mathbf{z}_r - \bar{\mathbf{z}})'(\mathbf{z}_s - \bar{\mathbf{z}}).$$

D'où,

$$\mathbf{B} = (\mathbf{HZ})'(\mathbf{HZ}).$$

Donc,

$$\mathbf{B} \geq 0.$$

( $\Leftarrow$ )

On sait que  $\mathbf{B} \geq 0$ .

Soient  $\lambda_1 > \lambda_2 > \dots > \lambda_p$  les valeurs propres positives de  $\mathbf{B}$  et  $\mathbf{X} = (\mathbf{x}_{(1)}, \mathbf{x}_{(2)}, \dots, \mathbf{x}_{(p)})$  les vecteurs propres de  $\mathbf{B}$  associés aux  $\lambda_i$  standardisés par  $\mathbf{x}_{(i)}' \mathbf{x}_{(i)} = \lambda_i$ .

Soient

$$\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$$

et

$$\mathbf{\Gamma} = \mathbf{X}\mathbf{\Lambda}^{-1/2}$$

où

$$\mathbf{\Lambda}^{-1/2} = \text{diag}(\lambda_1^{-1/2}, \lambda_2^{-1/2}, \dots, \lambda_p^{-1/2}).$$

Ainsi, les colonnes de  $\mathbf{\Gamma}$  sont

$$\gamma_{(i)} = \lambda_i^{-1/2} \mathbf{x}_{(i)}$$

Les  $\gamma_{(i)}$ ,  $i=1,2,\dots,p$ , sont les vecteurs propres normalisés de  $\mathbf{B}$  car

$$\begin{aligned} \gamma'_{(i)} \gamma_{(i)} &= (\lambda_i^{-1/2} \mathbf{x}_{(i)})' (\lambda_i^{-1/2} \mathbf{x}_{(i)}) \\ &= \lambda_i^{-1} (\mathbf{x}'_{(i)} \mathbf{x}_{(i)}) \\ &= \lambda_i^{-1} \lambda_i \\ &= 1. \end{aligned}$$

Selon le théorème de la décomposition spectrale (voir Annexe),

$$\begin{aligned} \mathbf{B} &= \mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}' \\ &= \mathbf{X}\mathbf{\Gamma}^{-1/2}\mathbf{\Gamma}\mathbf{\Gamma}^{-1/2}\mathbf{X}' \\ &= \mathbf{X}\mathbf{X}'. \end{aligned}$$

Alors,  $b_{rs} = \mathbf{x}'_r \mathbf{x}_s$  et  $\mathbf{B}$  représente la matrice des produits scalaires de cette configuration. Montrons maintenant que  $\mathbf{D}$  est la matrice des distances entre les points pour cette configuration. Utilisons 1.2 pour écrire  $\mathbf{B}$  en terme de  $\mathbf{A}$ .

Ainsi,

$$\begin{aligned}
(\mathbf{x}_r - \mathbf{x}_s)'(\mathbf{x}_r - \mathbf{x}_s) &= \mathbf{x}'_r \mathbf{x}_r - 2\mathbf{x}'_r \mathbf{x}_s + \mathbf{x}'_s \mathbf{x}_s \\
&= b_{rr} - 2b_{rs} + b_{ss} \\
&= (a_{rr} - \bar{a}_{r.} - \bar{a}_{.r} + \bar{a}_{..}) \\
&\quad - 2(a_{rs} - \bar{a}_{r.} - \bar{a}_{.s} + \bar{a}_{..}) \\
&\quad + (a_{ss} - \bar{a}_{s.} - \bar{a}_{.s} + \bar{a}_{..}) \\
&= a_{rr} - 2a_{rs} + a_{ss} \\
&= -2a_{rs} \\
&= d^2_{rs}
\end{aligned}$$

car  $a_{rr} = \frac{-1}{2}d^2_{rr} = 0$ .

De plus, on peut montrer que cette configuration a un centre de gravité  $\bar{\mathbf{x}} = \mathbf{0}$ . Sachant que  $\mathbf{B}$  est la matrice  $\mathbf{A}$  doublement centrée, on peut dire que

$$\mathbf{B}\mathbf{1} = \mathbf{H}\mathbf{A}\mathbf{H}\mathbf{1} = \mathbf{0}.$$

Ainsi,  $\mathbf{1}$  est un vecteur propre de  $\mathbf{B}$  correspondant à la valeur propre 0. De plus,  $\mathbf{1}$  est orthogonale aux colonnes de  $\mathbf{X}$ , car  $\mathbf{0} = \mathbf{B}\mathbf{1} = (\mathbf{X}\mathbf{X}')\mathbf{1} = \mathbf{X}(\mathbf{X}'\mathbf{1})$  et comme  $\mathbf{X} \neq \mathbf{0}$ , alors  $\mathbf{X}'\mathbf{1} = \mathbf{0}$ .

Ainsi,

$$\begin{aligned}
n\bar{\mathbf{x}} &= \sum_{i=1}^n \mathbf{x}_i \\
&= \mathbf{X}'\mathbf{1} \\
&= \mathbf{0}
\end{aligned}$$

D'où,

$$\bar{\mathbf{x}} = \mathbf{0}$$

Donc, le centre de gravité de cette configuration est à l'origine.  $\square$

Pour obtenir une configuration en  $k$  dimensions à l'aide de la solution classique du positionnement multidimensionnel, on utilise l'algorithme suivant.

**Algorithme 1.** - À partir de  $D$ , construire la matrice  $A = (-\frac{1}{2}d_{rs}^2)$ .

- Obtenir la matrice  $B$  avec  $b_{rs} = a_{rs} - \bar{a}_{r.} - \bar{a}_{.s} + \bar{a}_{..}$  ou avec  $B = HAH$ .
- Obtenir les  $k$  plus grandes valeurs propres  $\lambda_1 > \lambda_2 > \dots > \lambda_k > 0$  de  $B$  ( $k$  choisi d'avance) avec les vecteurs propres correspondants  $X = (x_{(1)}, x_{(2)}, \dots, x_{(k)})$  qui sont standardisés par  $x'_{(i)}x_{(i)} = \lambda_i$  ( $i = 1, 2, \dots, k$ ).
- Les coordonnées requises des points  $P_r$  sont  $x_r = (x_{r1}, x_{r2}, \dots, x_{rp})'$ ,  $r = 1, 2, \dots, k$

Dans plusieurs cas de positionnement multidimensionnel, on récolte des données de similarités plutôt que de distances. Le théorème suivant permet de travailler directement avec ces similarités sans avoir à les transformer préalablement en distances.

**Théorème 2.** Si  $C$  est semi-définie positive, alors la matrice des distances  $D$  est euclidienne avec comme matrice centrée des produits scalaires  $B = HCH$ .

*Démonstration.* Si  $C \geq 0$ , alors on a que

$$\begin{aligned} d_{rs}^2 &= c_{rr} - 2c_{rs} + c_{ss} \\ &= \mathbf{x}'\mathbf{C}\mathbf{x} \geq 0 \end{aligned}$$

où  $\mathbf{x}$  est un vecteur possédant  $+1$  à la  $r^{\text{ième}}$  ligne et  $-1$  à la  $s^{\text{ième}}$  ligne pour tout  $r \neq s$ . Ainsi, la transformation standard est bien définie et  $D$  est une matrice des distances car on a les trois conditions suivantes :

- symétrie :

$$d_{rs} = \sqrt{c_{rr} - 2c_{rs} + c_{ss}} = \sqrt{c_{ss} - 2c_{sr} + c_{rr}} = d_{sr}$$

– non-négativité :

$$d_{rs} = \sqrt{c_{rr} - 2c_{rs} + c_{ss}} \geq 0$$

– marque d'identification :

$$d_{rr}^2 = c_{rr} - 2c_{rr} + c_{rr} = 0$$

Maintenant, soient  $\mathbf{A} = \frac{-1}{2}d_{rs}^2$  et  $\mathbf{B} = \mathbf{H}\mathbf{A}\mathbf{H}$  et puisque  $\mathbf{H}\mathbf{C}\mathbf{H}$  est aussi semi-définie positive, on montrera que  $\mathbf{B} = \mathbf{H}\mathbf{C}\mathbf{H}$  pour en conclure que  $\mathbf{D}$  est euclidienne avec comme matrice des produits scalaires  $\mathbf{H}\mathbf{C}\mathbf{H}$ .

En remplaçant  $a_{rs}$  par  $\frac{-1}{2}d_{rs}^2$  dans la formule 1.2, on obtient

$$b_{rs} = c_{rs} - \bar{c}_{r.} - \bar{c}_{.s} + \bar{c}_{..}$$

D'où,

$$\mathbf{B} = \mathbf{H}\mathbf{C}\mathbf{H}.$$

□

### 1.3.2 Dualité entre l'analyse en composantes principales et le positionnement multidimensionnel

Dans les autres techniques d'analyse de données, il est plus fréquent de partir avec une matrice des données et de calculer la matrice des distances à partir de celle-ci. La fonction de distance la plus utilisée et la plus simple est sans aucun doute la distance euclidienne. Dans ce cas, la solution classique du positionnement multidimensionnel se rapproche beaucoup de l'analyse en composantes principales.

Soient  $\mathbf{X}(n \times p)$  une matrice de données et  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p (\neq 0)$  les valeurs propres

de  $n\mathbf{S} = \mathbf{X}'\mathbf{H}\mathbf{X}$ . Alors,  $\lambda_1, \lambda_2, \dots, \lambda_p$  sont aussi les valeurs propres de  $\mathbf{B} = (\mathbf{X}\mathbf{H})(\mathbf{X}\mathbf{H})' = \mathbf{H}\mathbf{X}\mathbf{X}'\mathbf{H} (\Rightarrow b_{rs} = (\mathbf{x}_r - \bar{\mathbf{x}})(\mathbf{x}_s - \bar{\mathbf{x}})')$

**Définition 6.** Soit  $\mathbf{v}_{(i)}$  le  $i^{\text{ième}}$  vecteur propre de  $\mathbf{B}$  ( $\mathbf{B}\mathbf{v}_{(i)} = \lambda_i\mathbf{v}_{(i)}$ ) standardisé par  $\mathbf{v}_{(i)}'\mathbf{v}_{(i)} = 1, i = 1, 2, \dots, p$ .

Pour un  $k$  fixé, les lignes de  $\mathbf{V}_k = (\mathbf{v}_{(1)}, \mathbf{v}_{(2)}, \dots, \mathbf{v}_{(k)})$  sont appelées les coordonnées principales de  $\mathbf{X}$  en  $k$  dimensions.

**Théorème 3.** Les principales coordonnées de  $\mathbf{X}$  en  $k$  dimensions sont données par les résultats centrés des  $n$  objets des  $k$  premières composantes principales.

*Démonstration.* Soit  $\gamma_{(i)}$  le  $i^{\text{ième}}$  vecteur de composante principale standardisé par  $\gamma_{(i)}'\gamma_{(i)} = 1$ . Par le théorème de la décomposition spectrale (voir Annexe),

$$n\mathbf{S} = \mathbf{X}'\mathbf{H}\mathbf{X} = \mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}'$$

où  $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$  et  $\mathbf{\Gamma}$  est la matrice orthogonale où les colonnes sont les vecteurs propres standardisés ( $\mathbf{\Gamma} = (\gamma_{(1)}, \gamma_{(2)}, \dots, \gamma_{(p)})$ ).

Par le théorème de la décomposition de la valeur singulière (voir Annexe), on peut choisir les signes de  $\gamma_i$  et de  $\mathbf{v}_i$ . Ainsi,  $\mathbf{H}\mathbf{X}$  peut être écrit en termes de ces vecteurs propres tel que

$$\mathbf{H}\mathbf{X} = \mathbf{V}\mathbf{\Gamma}'$$

où  $\mathbf{V} = (\mathbf{v}_{(1)}, \mathbf{v}_{(2)}, \dots, \mathbf{v}_{(p)})$ .

D'où,

$$\sum_{i=1}^n v_{ir} \gamma_{is} = x_{rs} - \frac{1}{n} \sum_{j=1}^n x_{js}$$

Les valeurs des  $n$  lignes de  $\mathbf{HX}$  sur la  $i^{\text{ème}}$  composante principale sont données par les  $n$  éléments de  $\mathbf{HX}\gamma_{(i)}$ . Ainsi, les valeurs des  $k$  premières composantes principales sont données par

$$\mathbf{HX}\Gamma_k = \mathbf{V}\Gamma_k' = \mathbf{V}(\mathbf{I}_k, \mathbf{0})' = \mathbf{V}_k$$

où  $\Gamma_k = (\gamma_1, \gamma_2, \dots, \gamma_k)$ . □

### 1.3.3 Propriétés optimales de la solution et indice de concordance

Soient  $\mathbf{X}$  une configuration dans  $\mathbb{R}^p$ ,  $\mathbf{L} = (\mathbf{L}_1, \mathbf{L}_2)$  une matrice orthogonale telle que  $\mathbf{L}_1$  est  $p \times k$  et  $\mathbf{L}_2$  est  $p \times (n - k)$ ,  $\mathbf{D}$  une matrice des distances euclidiennes correspondant à la configuration  $\mathbf{X}$ , un  $k$  fixé ( $1 \leq k \leq p$ ) et

$$\Phi = \sum_{r,s=1}^n (d_{rs} - \hat{d}_{rs})^2 = \sum_{r,s=1}^n \sum_{i=k+1}^p (x_{ri} - x_{si})^2$$

Ainsi, pour toutes les projections  $\mathbf{XL}_1$  de  $\mathbf{X}$  sur un sous-espace de  $k$  dimensions de  $\mathbb{R}^p$ ,  $\Phi$  est minimisé quand  $\mathbf{X}$  est projeté sur ses coordonnées principales en  $k$  dimensions. Le théorème suivant est introduit dans le but de conduire à des indices de concordance utilisés pour savoir si la configuration finale explique bien les données de départ. On retrouve ses indices de concordances à la définition 7.

**Théorème 4.** *Si  $\mathbf{D}$  est une matrice (pas nécessairement euclidienne) ainsi, pour un  $k$  fixé,  $\Psi = \text{tr}(\mathbf{B} - \hat{\mathbf{B}})^2$  est minimisé pour toutes les configurations  $\hat{\mathbf{X}}$  dans  $k$  dimensions quand  $\hat{\mathbf{X}}$  est la solution classique du problème de positionnement multidimensionnel.*

*Démonstration*

Voir la référence [1] à la page 408.

**Définition 7.** *Les indices de concordance pour les proportions d'une matrice des distances  $D$  expliquées par une solution classique de positionnement multidimensionnel en  $k$  dimensions sont, pour  $\lambda_k > 0$ ,*

$$\alpha_{1,k} = \left( \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^n |\lambda_i|} \right) \times 100\%$$

et

$$\alpha_{2,k} = \left( \frac{\sum_{i=1}^k \lambda_i^2}{\sum_{i=1}^n \lambda_i^2} \right) \times 100\%.$$

### 1.3.4 Méthode non-métrique de positionnement multidimensionnel

Dans le cas où on travaille avec des similarités, on peut transformer ces données en distances pour ensuite utiliser la solution classique de positionnement multidimensionnel pour créer l'espace dérivé. Cette technique est une méthode métrique. On peut aussi ordonner chaque paire de données de la plus similaire à la moins similaire. On utilisera ces données pour appliquer la méthode non-métrique. Cette méthode a pour but d'obtenir un aperçu de la nature des similarités à l'étude.

Dans la sous-section 1.1.4, où l'on parle de la sélection des données, on a énoncé trois procédures pour collecter les données de similarités. Dans le positionnement multidimensionnel non-métrique, on pourrait utiliser la comparaison par paires ou la combinaison croisée. Il existe aussi d'autres techniques pour arriver à nos fins comme par exemple

- les partitions où l'on divise l'ensemble des objets en plusieurs petites catégories ;
- la comparaison triadique où l'on ordonne les degrés de similarités des trois paires possibles provenant de trois objets et ce, pour tous les triplets possibles ;

- la tétrade où l'on compare toutes les paires d'objets possibles et indiquer la plus similaire et/ou la moins similaire.

Il est à noter que la sélection des données se fait au début de l'enquête et que ce n'est qu'à titre informatif que les procédures sont énoncées ci-haut.

### **Échelonnement ordinal**

Dans le positionnement multidimensionnel non-métrique, les dissimilarités données  $\delta_{rs}$  sont utilisées pour générer un ensemble de distances dérivées  $\hat{d}_{rs}$  qui sont approximativement reliées aux dissimilarités données  $\delta_{rs}$  de façon monotone, c'est-à-dire

$$\delta_{rs} < \delta_{uv} \Leftrightarrow \hat{d}_{rs} < \hat{d}_{uv}.$$

Si les  $\hat{d}_{rs}$  sont reliés de façon monotone aux  $\delta_{rs}$ , on le notera  $\delta_{rs} \sim^{mon} \hat{d}_{rs}$ . Puisque seul l'ordre de classement est important, l'échelonnement est ordinal et est communément appelé l'échelonnement non-métrique.

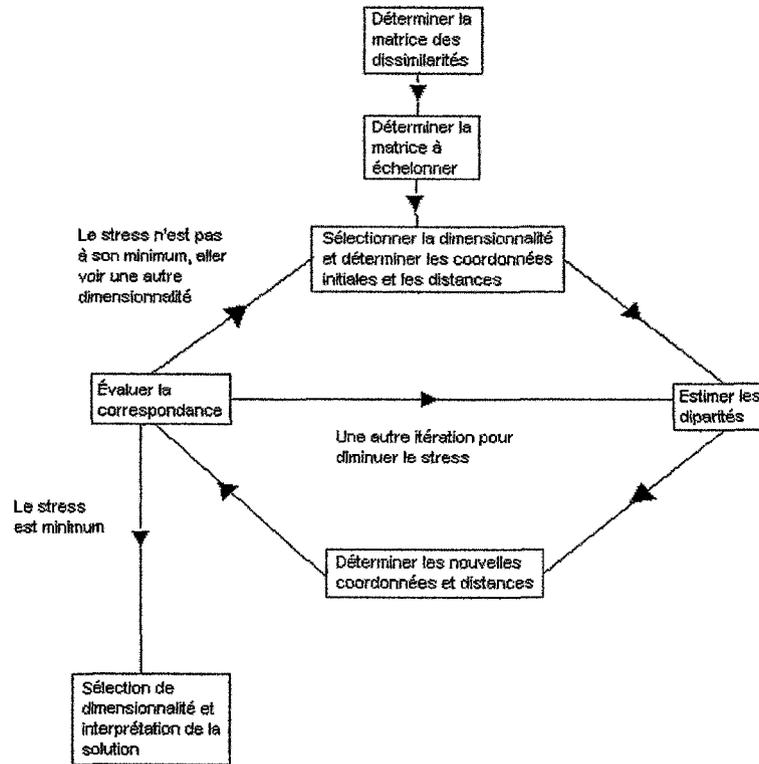
### **L'algorithme de Shepard-Kruskal**

L'approche utilisée pour créer l'espace multidimensionnel dans le cas présent est un processus itératif. Il est communément appelé algorithme de Shepard-Kruskal. Le but de ce processus est d'obtenir une représentation spatiale dans une dimensionnalité donnée telle que les distances euclidiennes sur les objets sont reliées de façon monotone aux dissimilarités originales.

La figure 1.5 représente l'algorithme de Shepard-Kruskal qui sera détaillé dans les paragraphes suivants.

Comme première étape, on crée de façon instinctive (sans procédure précise), une confi-

Figure 1.5 – Diagramme de l’algorithme de Shepard-Kruskal



guration  $X^{(0)}$  en  $k$  dimensions. On peut aussi la construire en sélectionnant des points aléatoires provenant d’une distribution normale multivariée. On y calcule les distances euclidiennes  $d_{rs}^{(0)}$  entre chaque point de la configuration. Ensuite, on compare ces distances avec les dissimilarités originales. Si les  $d_{rs}^{(0)}$  ont le même ordre de classement que les  $\delta_{rs}$ , on a une configuration satisfaisante. Par contre, cela n’arrive que très rarement. Ainsi, on doit trouver une configuration qui fera en sorte que l’ordre des distances soit aussi près que possible de l’ordre des dissimilarités.

Comme deuxième étape, on utilise donc la régression monotone pour relier les  $d_{rs}^{(0)}$  et les  $\delta_{rs}$ . La régression estimée produit un nouvel ensemble de pseudo-dissimilarités  $\hat{d}_{rs}^{(0)}$  appelées "disparités" qui sont reliées de façon monotone aux  $\delta_{rs}$ .







**Exemple 12.** Les données de l'exemple 11 sont traitées ici avec l'approche de classement d'image de Guttman.

$\delta_{rs}$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$d_{rs}^{(0)}$	0.6	3.2	1.6	2.2	2.8	2.6	2.5	2.1	3.3	3.6	3.7	3.5	3.3	3.6	4.0
$\hat{d}_{rs}^{(0)}$	0.6	1.6	2.1	2.2	2.5	2.6	2.8	3.2	3.3	3.3	3.5	3.6	3.6	3.7	4.0

Les figures 1.6 et 1.7 représentent les diagrammes de Shepard qui montrent les liens entre les dissimilarités, les distances et les disparités. On remarque que les courbes représentant les disparités (tracées en pointillés) sont croissantes. Celle provenant de l'algorithme de la violation du groupe adjacent possède plusieurs plateaux. Ce phénomène provient du fait que cet algorithme satisfait à la monotonie faible et non à la forte. Les figures 1.8 et 1.9 représentent les diagrammes d'images qui montrent le lien entre les distances et les disparités. On remarque qu'il y a en effet une tendance à la croissance.

Figure 1.6 – Diagramme de Shepard pour l'algorithme de la violation du groupe adjacent

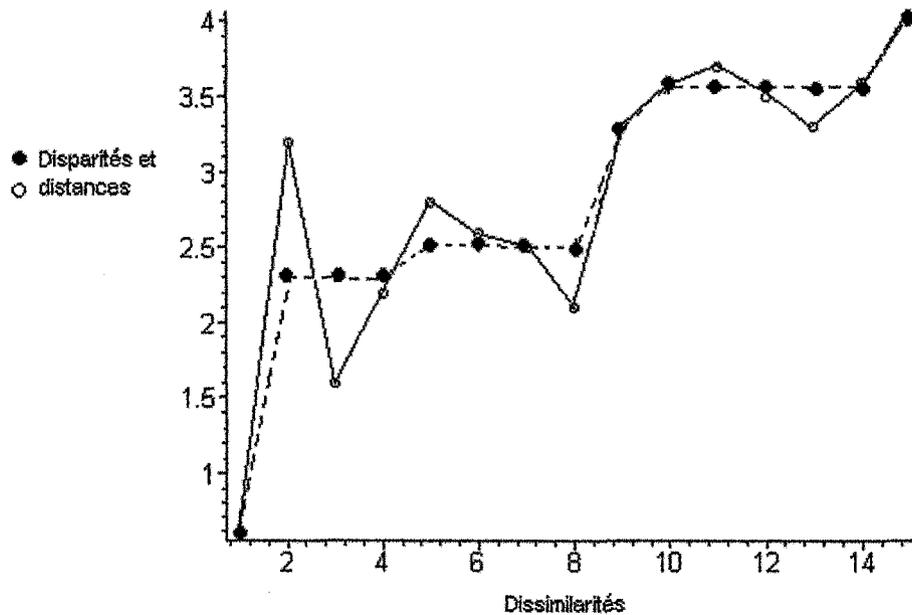
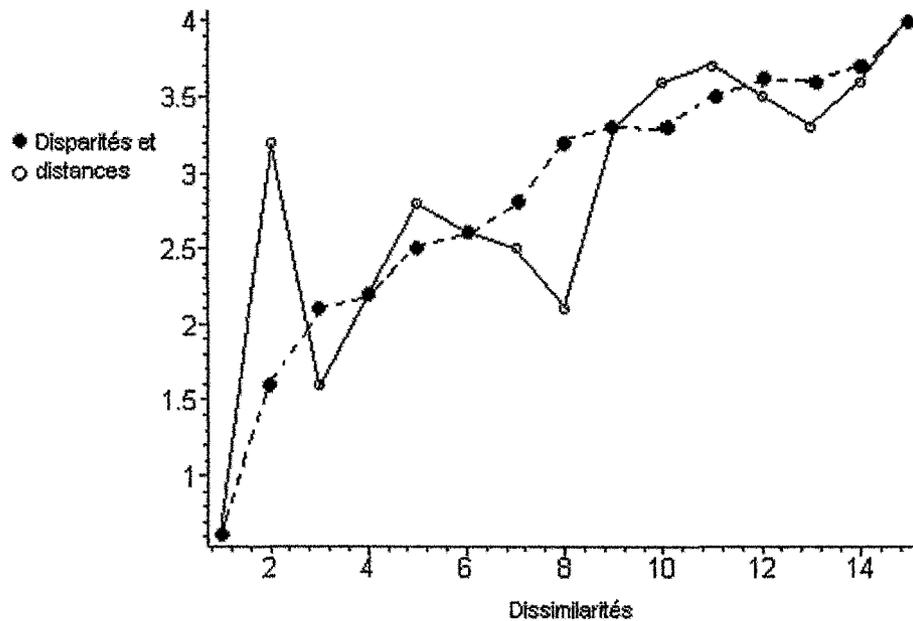


Figure 1.7 – Diagramme de Shepard pour l’approche de classement d’image de Guttman



### Liens entre les dissimilarités originales

S’il y a un lien entre les données de dissimilarités originales tels que  $\delta_{rs} = \delta_{uv}$ , alors il n’y a aucune garantie que  $d_{rs}^{(0)} = d_{uv}^{(0)}$ . Il existe deux approches pour estimer les  $d_{rs}^{(0)}$  quand deux ou plusieurs  $\delta_{rs}$  sont liés.

– 1<sup>ère</sup> approche :

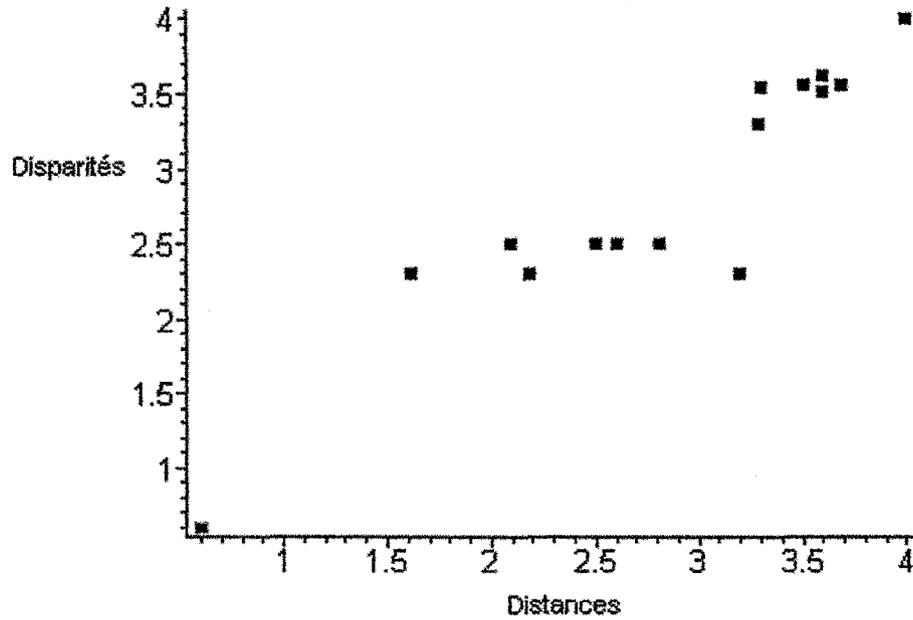
Permet aux  $d_{rs}^{(0)}$  d’être inégales même si les  $\delta_{rs}$  sont liés.

– 2<sup>ème</sup> approche :

Requiert que  $d_{rs}^{(0)} = d_{uv}^{(0)}$  si  $\delta_{rs} = \delta_{uv}$ .

En général, on utilise la première approche. La seconde demande un niveau de précision qui est souvent non nécessaire.

Figure 1.8 – Diagramme d’images pour l’algorithme de la violation du groupe adjacent



Comme troisième phase, on détermine le degré de concordance à l’aide de la mesure de stress. Elle consiste à calculer si les mesures dérivées des dissimilarités, c’est-à-dire les distances et les disparités, sont près les unes des autres. Si c’est le cas, l’échelonnement est considéré comme étant bon. La formule de stress est donnée par

$$STRESS = \frac{\sum_{r=1}^n \sum_{s=r+1}^n [\hat{d}_{rs}^{(0)} - d_{rs}^{(0)}]^2}{\sum_{r=1}^n \sum_{s=r+1}^n d_{rs}^{(0)^2}}$$

qui est l’écart de la somme des carrés normalisée par  $\sum_{r=1}^n \sum_{s=r+1}^n d_{rs}^{(0)^2}$ .

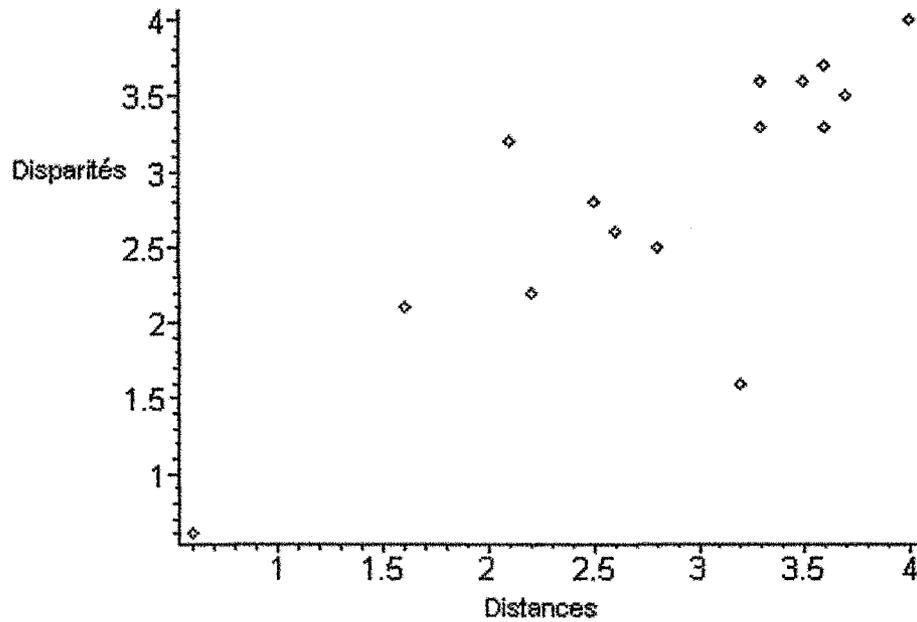
Il existe d’autres dénominateurs utilisés pour normaliser le stress comme

$$\sum_{r=1}^n \sum_{s=r+1}^n \hat{d}_{rs}^{(0)^2}$$

ou

$$\sum_{r=1}^n \sum_{s=r+1}^n [d_{rs}^{(0)} - \bar{d}_{..}]^2.$$

Figure 1.9 – Diagramme d’images pour l’approche de classement d’image de Guttman



La mesure de stress utilisée par le programme ALSCAL (discuté dans le chapitre 3) est la suivante

$$SSTRESS = \sqrt{\frac{\sum_{r=1}^n \sum_{s=r+1}^n [\hat{d}_{rs}^{(0)^2} - d_{rs}^{(0)^2}]^4}{\sum_{r=1}^n \sum_{s=r+1}^n \hat{d}_{rs}^{(0)^4}}}. \quad (1.3)$$

Les mesures de concordance STRESS et SSTRESS mettent plus d’emphase sur les écarts correspondants aux plus grandes dissimilarités. Plus les écarts sont élevées, plus ces indicateurs croient de manière importante. Cette affirmation s’applique encore plus sur SSTRESS puisque cette mesure utilise le carré des distances. Ainsi, les positions relatives des objets qui sont près les uns des autres ne devrait pas être utilisée pour tirer des conclusions. Ces objets représentés dans l’espace dérivé peuvent être vus simplement comme une grappe d’objets similaires.

Une mesure du degré de concordance basée sur les différences relatives est donnée par

$$STRESSR = \sum_{r=1}^n \sum_{s=r+1}^n \left[ 1 - \frac{\hat{d}_{rs}^{(0)}}{d_{rs}^{(0)}} \right]^2$$

Dans cette phase, on calcule une nouvelle configuration  $\mathbf{X}^{(1)}$  pour remplacer l'ancienne. On la construira de sorte que ces nouvelles mesures dérivées  $d_{rs}^{(1)}$  minimisent la formule de stress. Nos mesures de disparités ne changent pas à cette étape, contrairement à nos distances. Cette procédure nécessite donc une procédure numérique approximative comme les deux méthodes suivantes

- la descente par étape ;
- la méthode de Newton-Raphson.

Ces procédures donnent une expression de  $\mathbf{X}^{(1)}$  en fonction de  $\mathbf{X}^{(0)}$ ,  $\hat{d}_{rs}^{(0)}$  et  $d_{rs}^{(0)}$ . Elles déterminent les directions dans lesquelles la meilleure amélioration de niveau de concordance peut être atteinte. Ensuite, elles déplacent les points de la configuration  $\mathbf{X}^{(0)}$  dans ces directions d'une petite incrémentation pour créer la configuration  $\mathbf{X}^{(1)}$ .

La quatrième phase est la phase évaluative. On compare toutes les valeurs des degrés de concordance calculées dans la méthode itérative pour comparer les dimensionnalités. La procédure se termine habituellement quand le résultat de la dernière itération est assez petit. Par exemple, dans le programme ALSCAL, la procédure s'arrête lorsque la valeur entre le stress d'une itération à la suivante est de 0.001. On obtient ainsi une dimension qui donne la concordance optimale.

La cinquième phase est la phase de sélection et d'interprétation. Une fois que l'on a obtenu la solution minimum du stress pour une certaine dimensionnalité, on construit le graphique des valeurs minimums de stress. Cette fonction des dimensionnalités donne habituellement une courbe ayant une allure exponentielle de pente vers le bas. À chaque

fois que l'on arrive sur une dimension précise dans la courbe, on y trouve un coude. À partir d'une certaine dimension, la fonction devient plane. La dimension correspondante au point de changement de pente est ensuite sélectionnée comme étant la dimension appropriée. L'allure de la courbe dépend du fait que lorsque la dimension augmente, la valeur de stress diminue.

L'ampleur de la valeur de stress dépend de plusieurs facteurs comme de la formule utilisée et du critère choisi pour faire la régression monotone. À moins d'indication contraire, on utilisera le critère de monotonie faible avec la première approche dans le cas des liens. Il existe d'autres facteurs d'influence comme le nombre de points à l'étude, les erreurs et le bruit. Pour plus de détails sur leur influence, il est possible de consulter l'étude de Monte Carlo sur la fonction de stress dans le livre de J.D. Jobson [5].

Une règle de pouce existe pour utiliser la valeur de stress comme indicateur. Ainsi, la valeur idéale de la solution devrait être au maximum 0.10 et il est préférable qu'elle soit aussi basse que 0.05 [6]. Il faut tout de même être prudent dans l'analyse de la valeur de stress. Il faut tenir compte du nombre de points à l'étude. Plus on a de points à comparer, plus le nombre de termes à additionner dans la formule du stress sera élevé. En d'autres termes, plus on a un grand nombre de points à mettre en concordance, plus grand est le nombre des différences des carrés dans la formule de stress ce qui fait augmenter sa valeur. Ceci impliquera que la valeur sera plus grande, mais la représentation n'en sera peut-être pas plus mauvaise.

Pour garder un équilibre entre le nombre de points et le nombre de dimensions, il existe ici aussi une règle du pouce qui indique le nombre de points idéal tout dépendant du nombre de dimensions. Cette règle va comme suit : le nombre de points devrait excéder  $4k-1$  où  $k$  est le nombre de dimensions. De plus, comme l'erreur de la variance augmente,

la vraie configuration est moins distincte et plus difficile à concorder. Une portion de la différence des carrés entre les distances et les disparités est due à l'erreur dans les dissimilarités initiales.

Le paragraphe suivant énonce entièrement l'algorithme de Shepard-Kruskal.

**Algorithme 2.** *L'algorithme de Kruskal-Shepard [6] sert à construire une configuration basée sur l'information de l'ordre du rang.*

1. *Soit  $\mathbf{D}$  une matrice de dissimilarité. Ordonner les éléments qui ne se trouvent pas sur la diagonale. Ainsi,  $\delta_{r_1, s_1} \leq \dots \leq \delta_{r_m, s_m}$  où  $m = \frac{1}{2}n(n-1)$  ( $m$  représentant le nombre d'éléments en dessus de la diagonale) et où  $(r_1, s_1), (r_2, s_2), \dots, (r_m, s_m)$  dénote toutes les paires d'indices inégaux où  $r_i < s_i$ . Sélectionner le nombre de dimensions  $k, k = 1, 2, \dots, p$ . Déterminer la configuration  $\mathbf{X}^{(0)}$  ainsi que les distances  $d_{rs}^{(0)}$  associées à  $\mathbf{X}^{(0)}$ .*
2. *Utiliser la régression monotone pour relier  $d_{rs}^{(0)}$  et  $\delta_{rs}$  pour créer les disparités  $\hat{d}_{rs}^{(0)}$  qui sont reliées de façon monotone aux  $\delta_{rs}$ .*
3. *Définir le stress (carré) de  $\mathbf{X}^{(0)}$  par*

$$S^2(\mathbf{X}^{(0)}) = \min \frac{\sum_{r=1}^n \sum_{s=r+1}^n (\hat{d}_{rs}^{(0)} - d_{rs}^{(0)})^2}{\sum_{r=1}^n \sum_{s=r+1}^n d_{rs}^{(0)2}}$$

*où le minimum est pris sur  $\hat{\mathbf{d}}^{(0)}$  quand  $\delta_{rs} \sim^{mon} \hat{d}_{rs}^{(0)}$ . Le  $\hat{d}_{rs}^{(0)}$  qui minimise  $S^2(\mathbf{X}^{(0)})$  représente la régression monotone des moindres carrées de  $d_{rs}^{(0)}$  sur  $\delta_{rs}$ . De plus,  $S^2(\mathbf{X}^{(0)})$  représente la superficie sur laquelle l'ordre du rang de  $d_{rs}^{(0)}$  ne concorde pas avec l'ordre du rang de  $\delta_{rs}$ . Si les ordres du rang concordent exactement (ce qui arrive rarement en pratique) alors  $S^2(\mathbf{X}^{(0)}) = 0$ .*

*La présence du dénominateur dans l'expression de  $S^2(\mathbf{X}^{(0)})$  normalise le stress et le rend invariant à d'autres transformations de la forme  $\mathbf{y}_r = c\mathbf{x}_r, r = 1, 2, \dots, n, c \neq 0$ . Le stress est aussi invariant aux transformations de la forme  $\mathbf{y}_r = \mathbf{A}\mathbf{x}_r + \mathbf{b}$  quand  $\mathbf{A}$  est orthogonale.*

Déterminer si le degré de concordance est bon entre  $\hat{d}_{rs}^{(0)}$  et  $\delta_{rs}$ . Si c'est bon, passer à l'étape suivante, sinon, modifier la configuration spatiale pour obtenir  $\mathbf{X}^{(1)}$  ainsi que les distances  $d_{rs}^{(1)}$  de façon à ce qu'elles soient plus reliées monotonement aux disparités  $\hat{d}_{rs}^{(0)}$  générées à l'étape 2 et retourner à l'étape 2 avec comme nouvelles distances  $d_{rs}^{(1)}$ .

4. Pour chaque dimension  $k$ , la configuration qui a la plus petite valeur de stress est appelée la configuration ayant la meilleure concordance en  $k$  dimensions, soit  $S_k = \min_k S(\mathbf{X}^{(k)})$  le stress minimal.

5. Pour choisir la bonne dimension, calculer  $S_1, S_2, \dots$  jusqu'à ce que les valeurs deviennent basses. Disons par exemple que  $S_k$  est bas et que  $k = k_0$ . Puisque  $S_k$  est une fonction décroissante en  $k$ ,  $k = k_0$  est la bonne dimension. Kruskal [6] a fournit une règle pour juger le degré de tolérance de  $S_k$ ;

- $S_k \geq 20\%$  :vraiment pas bon
- $S_k = 10\%$  :moyen
- $S_k \leq 5\%$  :bien
- $S_k = 0$  :parfait

### 1.3.5 Indices de concordances et standardisations rotationnelles

On va maintenant définir un indice d'ajustement utilisé pour comparer deux configurations. Soit  $\mathbf{X}$  une matrice ( $n \times p$ ) des coordonnées de  $n$  points (en  $p$  dimensions) obtenue de  $\mathbf{D}$  par une certaine technique. Supposons que  $\mathbf{Y}$  est la matrice des coordonnées ( $n \times q$ ) d'un autre ensemble de points obtenus par une autre technique ou utilisant une autre mesure de distance. Soit  $q \leq p$ . En ajoutant des colonnes de 0 à  $\mathbf{Y}$ , on pourra ainsi prétendre sans perte de généralité que  $\mathbf{Y}$  est une matrice ( $n \times p$ ). La mesure de l'indice d'ajustement adopté est obtenue en déplaçant les points  $\mathbf{y}_r$ , relatifs aux points  $\mathbf{x}_r$ , quand

la somme résiduelle des carrées est minimale.

Somme résiduelle des carrées :

$$\sum_{r=1}^n (\mathbf{x}_r - \mathbf{y}_r)' (\mathbf{x}_r - \mathbf{y}_r)$$

On peut déplacer  $\mathbf{y}_r$  relativement à  $\mathbf{x}_r$  par des rotations, des réflexions et des translations, c'est-à-dire par :

$$\mathbf{A}'\mathbf{y}_r + \mathbf{b}$$

où  $r = 1, 2, \dots, n$  et  $\mathbf{A}'$  est une matrice orthogonale  $p \times p$ . Ainsi, on souhaite résoudre

$$R^2 = \min_{\mathbf{A}, \mathbf{b}} \sum_{r=1}^n (\mathbf{x}_r - \mathbf{A}'\mathbf{y}_r - \mathbf{b})' (\mathbf{x}_r - \mathbf{A}'\mathbf{y}_r - \mathbf{b})$$

pour  $\mathbf{A}$  et  $\mathbf{b}$ . Notons que  $\mathbf{A}$  et  $\mathbf{b}$  sont trouvés par les moindres carrés. Leurs valeurs sont données dans le théorème qui suit.

**Théorème 5.** Soit  $\mathbf{X}$  ( $n \times p$ ) et  $\mathbf{Y}$  ( $n \times p$ ) deux configurations de  $n$  points, prenons-les centrés à l'origine sans perte de généralité, ainsi  $\bar{\mathbf{x}} = \bar{\mathbf{y}} = 0$ . Soit  $\mathbf{Z} = \mathbf{Y}'\mathbf{X}$  et utilisons le théorème de décomposition de la valeur singulière (voir Annexe). Ainsi,

$$\mathbf{Z} = \mathbf{V}\mathbf{\Gamma}\mathbf{U}'$$

où  $\mathbf{V}$  et  $\mathbf{U}$  sont des matrices orthogonales ( $p \times p$ ) et  $\mathbf{\Gamma}$  est une matrice diagonale d'éléments non-négatifs. Ainsi, les valeurs minimisantes de  $\mathbf{A}$  et  $\mathbf{b}$  dans la formule de  $R^2$  sont

$$\hat{\mathbf{b}} = 0 \text{ et } \hat{\mathbf{A}} = \mathbf{V}\mathbf{U}'$$

et, de plus,

$$R^2 = \text{tr}\mathbf{X}\mathbf{X}' + \text{tr}\mathbf{Y}\mathbf{Y}' - 2\text{tr}\mathbf{\Gamma}$$

Si on effectue un changement d'échelle de la forme  $c\mathbf{A}\mathbf{y}_r + \mathbf{b}$ ,  $r = 1, 2, \dots, n$ ,  $c > 0$ , on aura que la valeur minimisante de  $c$  est

$$\hat{c} = \frac{\text{tr}\Gamma}{\text{tr}\mathbf{Y}\mathbf{Y}'}$$

et

$$\begin{aligned} R^2 &= \text{tr}\mathbf{X}\mathbf{X}' + \hat{c}^2 \text{tr}\mathbf{Y}\mathbf{Y}' - 2\hat{c} \text{tr}\Gamma \\ &= \text{tr}\mathbf{X}\mathbf{X}' + \hat{c}^2 \text{tr}\mathbf{Y}\mathbf{Y}' - 2\hat{c} \text{tr}\mathbf{X}'\mathbf{Y}\mathbf{Y}'\mathbf{X}^{1/2} \end{aligned}$$

où  $\Gamma = \text{diag}(\lambda_1^{1/2}, \lambda_2^{1/2}, \dots, \lambda_p^{1/2})$  et  $\mathbf{V}$  est composé des vecteurs propres de  $\mathbf{Z}\mathbf{Z}'$  et  $\mathbf{U}$  est composé des vecteurs propres de  $\mathbf{Z}'\mathbf{Z}$ .

*Démonstration.* Trouvons d'abord la valeur minimisante de  $\mathbf{b}$ .

Posons

$$\begin{aligned} k &= \sum_{r=1}^n (\mathbf{x}_r - \mathbf{A}'\mathbf{y} - \mathbf{b})' (\mathbf{x}_r - \mathbf{A}'\mathbf{y} - \mathbf{b}) \\ &= \sum_{r=1}^n \sum_{j=1}^n (x_{rj} - b_j - \sum_{i=1}^p y_{ri} a_{ij})^2 \end{aligned}$$

Pour trouver la valeur de  $\mathbf{b}$  lorsque  $k$  est minimisé, dérivons  $k$  par rapport à  $\mathbf{b}$ .

Posons  $m \in \{1, 2, \dots, n\}$ .

$$\begin{aligned} \frac{\partial k}{\partial b_m} &= -2 \sum_{r=1}^n (x_{rm} - b_m - \sum_{i=1}^p y_{ri} a_{im}) \\ &= 2nb_m - 2 \sum_{r=1}^n x_{rm} - 2 \sum_{r=1}^n \sum_{i=1}^p y_{ri} a_{im} \end{aligned}$$

$$\begin{aligned}
\frac{\partial k}{\partial b_m} = 0 &\Leftrightarrow 2nb_m = 2 \sum_{r=1}^n x_{rm} + 2 \sum_{i=1}^p a_{im} \left( \sum_{r=1}^n y_{ri} \right) \\
&\Leftrightarrow b_m = \frac{1}{n} \sum_{r=1}^n x_{rm} + \sum_{i=1}^p a_{im} \left( \frac{1}{n} \sum_{r=1}^n y_{ri} \right) \\
&\Leftrightarrow \mathbf{b} = \bar{\mathbf{x}} - \mathbf{A}'\bar{\mathbf{y}}
\end{aligned}$$

Donc, pour minimiser la formule, on doit poser

$$\hat{\mathbf{b}} = \bar{\mathbf{x}} - \mathbf{A}'\bar{\mathbf{y}}.$$

Si les deux configurations sont centrées,  $\hat{\mathbf{b}} = \mathbf{0}$ . On montrera donc que l'on peut écrire  $R^2$  comme suit :

$$R^2 = \min_A [\text{tr}(\mathbf{X} - \mathbf{YA})(\mathbf{X} - \mathbf{YA})'] \quad (1.4)$$

$$= \text{tr}\mathbf{XX}' + \text{tr}\mathbf{YY}' - 2\max_A [\text{tr}\mathbf{X}'\mathbf{YA}]. \quad (1.5)$$

La formule 1.4 est facile à montrer en calculant que

$$\text{tr}(\mathbf{X} - \mathbf{YA})(\mathbf{X} - \mathbf{YA})' = \sum_{r=1}^n (\mathbf{x}_r - \mathbf{A}'\mathbf{y})'(\mathbf{x}_r - \mathbf{A}'\mathbf{y}).$$

Montrons maintenant l'égalité 1.5.

$$\begin{aligned}
\min_A [\text{tr}(\mathbf{X} - \mathbf{YA})(\mathbf{X} - \mathbf{YA})'] &= \min_A [\text{tr}(\mathbf{X} - \mathbf{YA})(\mathbf{X}' - \mathbf{A}'\mathbf{Y}')] \\
&= \min_A [\text{tr}\mathbf{XX}' - 2\text{tr}\mathbf{YAX}' + \text{tr}\mathbf{YAA}'\mathbf{Y}'] \\
&= \min_A [\text{tr}\mathbf{XX}' - 2\text{tr}\mathbf{X}'\mathbf{YA} + \text{tr}\mathbf{YY}'] \\
&= \text{tr}\mathbf{XX}' + \text{tr}\mathbf{YY}' - 2\max_A [\text{tr}\mathbf{X}'\mathbf{YA}]
\end{aligned}$$

Les contraintes sur  $\mathbf{A}$  sont  $\mathbf{AA}' = \mathbf{I}$ , c'est-à-dire

$$\mathbf{a}'_i \mathbf{a}_j = \begin{cases} 1 & \text{si } i=j \\ 0 & \text{si } i \neq j \end{cases}$$

où  $\mathbf{a}'_i$  est la  $i^{\text{ème}}$  ligne de  $\mathbf{A}$ . Ainsi, il y a  $p(p+1)/2$  contraintes.

Soit  $\frac{1}{2}\Lambda$  la matrice symétrique de dimensions  $p \times p$  des multiplicateurs de Lagrange (voir Annexe) pour ces contraintes. Le but est de maximiser

$$\text{tr}\{\mathbf{Z}'\mathbf{A} - \frac{1}{2}\Lambda(\mathbf{A}\mathbf{A}' - \mathbf{I})\}$$

où  $\mathbf{Z} = \mathbf{X}'\mathbf{Y}$ . Par différentiation, on obtient que

$$\frac{\partial}{\partial \mathbf{A}} \text{tr}\mathbf{Z}'\mathbf{A} = \mathbf{Z}$$

et

$$\frac{\partial}{\partial \mathbf{A}} \text{tr}\Lambda\mathbf{A}\mathbf{A}' = 2\Lambda.$$

Ainsi, en dérivant  $\text{tr}\{\mathbf{Z}'\mathbf{A} - \frac{1}{2}\Lambda(\mathbf{A}\mathbf{A}' - \mathbf{I})\}$  et en égalant cette dérivée à zéro, on trouve que  $\mathbf{A}$  doit satisfaire à

$$\mathbf{Z} = \Lambda\mathbf{A}.$$

En effet,

$$\begin{aligned} \frac{\partial}{\partial \mathbf{A}} \text{tr}\{\mathbf{Z}'\mathbf{A} - \frac{1}{2}\Lambda(\mathbf{A}\mathbf{A}' - \mathbf{I})\} &= \frac{\partial}{\partial \mathbf{A}} \text{tr}\mathbf{Z}'\mathbf{A} - \frac{1}{2} \frac{\partial}{\partial \mathbf{A}} \text{tr}\Lambda\mathbf{A}\mathbf{A}' - \frac{1}{2} \text{tr}\Lambda \\ &= \mathbf{Z} - \frac{1}{2}(2\Lambda\mathbf{A}) \\ &= \mathbf{Z} - \Lambda\mathbf{A}. \end{aligned}$$

Ainsi,

$$\begin{aligned} \frac{\partial}{\partial \mathbf{A}} \text{tr}\{\mathbf{Z}'\mathbf{A} - \frac{1}{2}\Lambda(\mathbf{A}\mathbf{A}' - \mathbf{I})\} = 0 &\Leftrightarrow \mathbf{Z} - \Lambda\mathbf{A} = 0 \\ &\Leftrightarrow \mathbf{Z} = \Lambda\mathbf{A}. \end{aligned}$$

Utilisons l'égalité du théorème de la décomposition de la valeur singulière, c'est-à-dire  $\mathbf{Z} = \mathbf{V}\mathbf{\Gamma}\mathbf{U}'$  (voir Annexe). Notons que  $\mathbf{\Lambda}$  est symétrique et que  $\mathbf{A}$  est orthogonale. On obtient du fait que  $\mathbf{Z} = \mathbf{\Lambda}\mathbf{A}$  et du fait que  $\mathbf{U}$  et  $\mathbf{V}$  sont orthogonales, que

$$\begin{aligned}
 \mathbf{\Lambda}^2 &= \mathbf{Z}\mathbf{A}'\mathbf{A}\mathbf{Z}' = \mathbf{Z}\mathbf{Z}' = (\mathbf{V}\mathbf{\Gamma}\mathbf{U}')(\mathbf{U}\mathbf{\Gamma}\mathbf{V}) \\
 &= \mathbf{V}\mathbf{\Gamma}\mathbf{U}'\mathbf{U}\mathbf{\Gamma}\mathbf{V} \\
 &= \mathbf{V}\mathbf{\Gamma}^2\mathbf{V}' \\
 &= \mathbf{V}\mathbf{\Gamma}\mathbf{V}'\mathbf{V}\mathbf{\Gamma}\mathbf{V} \\
 &= (\mathbf{V}\mathbf{\Gamma}\mathbf{V}')(\mathbf{V}\mathbf{\Gamma}\mathbf{V}).
 \end{aligned}$$

D'où,

$$\mathbf{\Lambda} = \mathbf{V}\mathbf{\Gamma}\mathbf{V}'.$$

En substituant cette valeur de  $\mathbf{\Lambda}$  dans  $\mathbf{Z} = \mathbf{\Lambda}\mathbf{A}$ , on obtient

$$\mathbf{V}\mathbf{\Gamma}\mathbf{U}' = \mathbf{Z} = \mathbf{\Lambda}\mathbf{A} = (\mathbf{V}\mathbf{\Gamma}\mathbf{V}')\mathbf{A}.$$

D'où,

$$\begin{aligned}
 \mathbf{V}\mathbf{\Gamma}\mathbf{U}' = \mathbf{V}\mathbf{\Gamma}\mathbf{V}'\mathbf{A} &\Leftrightarrow \mathbf{V}'\mathbf{V}\mathbf{\Gamma}\mathbf{U}' = \mathbf{V}'\mathbf{V}\mathbf{\Gamma}\mathbf{V}'\mathbf{A} \\
 &\Leftrightarrow \mathbf{\Gamma}\mathbf{U}' = \mathbf{\Gamma}\mathbf{V}'\mathbf{A} \\
 &\Leftrightarrow \mathbf{\Gamma}^{-1}\mathbf{\Gamma}\mathbf{U}' = \mathbf{\Gamma}^{-1}\mathbf{\Gamma}\mathbf{V}'\mathbf{A} \\
 &\Leftrightarrow \mathbf{U}' = \mathbf{V}'\mathbf{A} \\
 &\Leftrightarrow \mathbf{V}\mathbf{U}' = \mathbf{V}\mathbf{V}'\mathbf{A} \\
 &\Leftrightarrow \mathbf{V}\mathbf{U}' = \mathbf{A}.
 \end{aligned}$$

Donc, pour minimiser  $R^2$ ,

$$\hat{\mathbf{A}} = \mathbf{V}\mathbf{U}'.$$

On remarque ainsi que  $\hat{\mathbf{A}} = \mathbf{V}\mathbf{U}'$  est une solution de  $\mathbf{Z} = \mathbf{\Lambda}\mathbf{A}$ . Notons que  $\hat{\mathbf{A}}$  est orthogonale.

Finallement, pour vérifier si  $\hat{\mathbf{A}}$  maximise bien  $tr\{\mathbf{Z}'\mathbf{A} - \frac{1}{2}\mathbf{\Lambda}(\mathbf{A}\mathbf{A}' - \mathbf{I})\}$  (et que ce n'est pas seulement un point stationnaire), on doit différencier  $tr\{\mathbf{Z}'\mathbf{A} - \frac{1}{2}\mathbf{\Lambda}(\mathbf{A}\mathbf{A}' - \mathbf{I})\}$  une seconde fois. Dans ce but, il est préférable d'écrire  $\mathbf{A}$  comme un vecteur  $\mathbf{a} = (\mathbf{a}'_{(1)}, \mathbf{a}'_{(2)}, \dots, \mathbf{a}'_{(p)})'$ . Ensuite,  $tr\{\mathbf{Z}'\mathbf{A} - \frac{1}{2}\mathbf{\Lambda}(\mathbf{A}\mathbf{A}' - \mathbf{I})\}$ , est une fonction quadratique des éléments de  $\mathbf{a}$  et la seconde dérivation de  $tr\{\mathbf{Z}'\mathbf{A} - \frac{1}{2}\mathbf{\Lambda}(\mathbf{A}\mathbf{A}' - \mathbf{I})\}$  avec respect pour  $\mathbf{a}$  peut être exprimée comme la matrice  $-\mathbf{I}_p \otimes \mathbf{\Lambda}$ . Puisque  $\mathbf{\Lambda} = \mathbf{V}\mathbf{\Gamma}\mathbf{V}'$  et les éléments de la diagonale de  $\mathbf{\Gamma}$  sont non-négatifs, on voit que la seconde matrice dérivative est semi-définie négative. Ainsi,  $\hat{\mathbf{A}}$  maximise  $tr\{\mathbf{Z}'\mathbf{A} - \frac{1}{2}\mathbf{\Lambda}(\mathbf{A}\mathbf{A}' - \mathbf{I})\}$ .

Si on utilise cette valeur de  $\hat{\mathbf{A}}$  dans  $R^2 = tr\mathbf{X}\mathbf{X}' + tr\mathbf{Y}\mathbf{Y}' - 2max_A[tr\mathbf{X}'\mathbf{Y}\mathbf{A}]$ , on obtient

$$R^2 = tr\mathbf{X}\mathbf{X}' + tr\mathbf{Y}\mathbf{Y}' - 2tr\mathbf{\Gamma}.$$

En effet,

$$\begin{aligned} \mathbf{X}'\mathbf{Y}\hat{\mathbf{A}} &= \mathbf{X}'\mathbf{Y}\mathbf{V}\mathbf{U}' \\ &= \mathbf{Z}'\mathbf{V}\mathbf{U}' \\ &= (\mathbf{V}\mathbf{\Gamma}\mathbf{U}')'\mathbf{V}\mathbf{U}' \\ &= \mathbf{U}\mathbf{\Gamma}\mathbf{V}'\mathbf{V}\mathbf{U}' \\ &= \mathbf{U}\mathbf{\Gamma}\mathbf{U}' \end{aligned}$$

et

$$tr\mathbf{U}\mathbf{\Gamma}\mathbf{U}' = tr\mathbf{\Gamma}\mathbf{U}\mathbf{U}' = tr\mathbf{\Gamma}.$$

De plus,

$$\mathbf{X}'\mathbf{Y}\mathbf{Y}'\mathbf{X} = \mathbf{Z}'\mathbf{Z} = \mathbf{U}\mathbf{\Gamma}\mathbf{V}'\mathbf{V}\mathbf{\Gamma}\mathbf{U}' = \mathbf{U}\mathbf{\Gamma}^2\mathbf{U}'$$

implique que

$$\text{tr}\mathbf{\Gamma} = \text{tr}(\mathbf{X}'\mathbf{Y}\mathbf{Y}'\mathbf{X})^{1/2}.$$

D'où

$$R^2 = \text{tr}\mathbf{X}\mathbf{X}' + \text{tr}\mathbf{Y}\mathbf{Y}' - 2\text{tr}(\mathbf{X}'\mathbf{Y}\mathbf{Y}'\mathbf{X})^{1/2}.$$

Maintenant, si on a un changement d'échelles sur la transformation  $\mathbf{A}'\mathbf{y}_r + \mathbf{b}$ ,  $r = 1, 2, \dots, n$ , il sera de la forme

$$c\mathbf{A}'\mathbf{y}_r + \mathbf{b} \tag{1.6}$$

où  $c > 0$ . Ainsi,

$$\begin{aligned} R^2 &= \min_A [\text{tr}(\mathbf{X} - c\mathbf{Y}\mathbf{A})(\mathbf{X} - c\mathbf{Y}\mathbf{A})'] \\ &= \min_A [\text{tr}(\mathbf{X} - c\mathbf{Y}\mathbf{A})(\mathbf{X}' - c\mathbf{A}'\mathbf{Y}')] \\ &= \min_A [\text{tr}\mathbf{X}\mathbf{X}' - c\mathbf{Y}\mathbf{A}\mathbf{X}' - c\mathbf{X}\mathbf{A}'\mathbf{Y}' + c^2\mathbf{Y}\mathbf{A}\mathbf{A}'\mathbf{Y}'] \\ &= \min_A [\text{tr}\mathbf{X}\mathbf{X}' + c^2\mathbf{Y}\mathbf{Y}' - 2c\mathbf{X}'\mathbf{Y}\mathbf{A}] \\ &= \text{tr}\mathbf{X}\mathbf{X}' + c^2\text{tr}\mathbf{Y}\mathbf{Y}' - 2c \times \max_A [\text{tr}\mathbf{X}'\mathbf{Y}\mathbf{A}] \\ &= \text{tr}\mathbf{X}\mathbf{X}' + c^2\text{tr}\mathbf{Y}\mathbf{Y}' - 2c \times \text{tr}\mathbf{\Gamma}. \end{aligned}$$

La valeur de  $c$  qui minimise  $R^2$  se trouve comme suit :

$$\frac{\partial R^2}{\partial c} = 2c\text{tr}\mathbf{Y}\mathbf{Y}' - 2\text{tr}\mathbf{\Gamma}$$

$$\begin{aligned}
\frac{\partial R^2}{\partial c} = 0 &\Leftrightarrow 2c \times \text{tr} \mathbf{Y} \mathbf{Y}' - 2 \text{tr} \mathbf{\Gamma} = 0 \\
&\Leftrightarrow c \times \text{tr} \mathbf{Y} \mathbf{Y}' = \text{tr} \mathbf{\Gamma} \\
&\Leftrightarrow c = \frac{\text{tr} \mathbf{\Gamma}}{\text{tr} \mathbf{Y} \mathbf{Y}'}
\end{aligned}$$

D'où on obtient le fait que

$$\hat{c} = \frac{\text{tr} \mathbf{\Gamma}}{\text{tr} \mathbf{Y} \mathbf{Y}'}$$

et que

$$\begin{aligned}
R^2 &= \text{tr} \mathbf{X} \mathbf{X}' + \hat{c}^2 \text{tr} \mathbf{Y} \mathbf{Y}' - 2 \hat{c} \times \text{tr} \mathbf{\Gamma} \\
&= \text{tr} \mathbf{X} \mathbf{X}' + \hat{c}^2 \text{tr} \mathbf{Y} \mathbf{Y}' - 2 \hat{c} \times \text{tr} \mathbf{X}' \mathbf{Y} \mathbf{Y}' \mathbf{X}^{1/2}
\end{aligned}$$

On remarque que  $R^2 = 0$  si et seulement si  $\mathbf{y}_r$  peut être rotationné exactement à  $\mathbf{x}_r$ .

□

Pour un approfondissement de ce sujet, voir la référence [4].

### 1.3.6 Analyse de l'espace conjoint

Dans l'analyse du positionnement multidimensionnel, on a toujours travaillé en comparant les objets à l'étude entre eux seulement. Cette technique se faisait à l'aide d'une matrice carrée des données. On peut aussi placer conjointement dans la carte des perceptions les objets et les répondants ou les objets et leurs attributs. Pour ce faire, on utilise une matrice rectangulaire.

On introduira seulement le cas où on a les objets et les répondants par le biais de l'analyse des différences individuelles. En utilisant cette procédure, on représente conjointement par des points chaque répondant et chaque stimuli dans l'espace multidimensionnel. Lorsqu'un sujet est situé près d'un certain objet, c'est qu'il lui apporte une grande affinité. Au contraire, plus ils sont distancés, plus le sujet le considère comme inintéressant. Si deux répondants sont représentés près l'un de l'autre, cela signifie qu'ils ont des goûts similaires. De plus, comme dans le positionnement multidimensionnel classique, si deux objets sont près dans l'espace dérivé, ils se ressemblent et, à l'inverse, s'ils sont loin, ils sont dissimilaires.

L'exemple suivant tiré du livre de Dillon et Goldstein [6] montre globalement ce que donne l'analyse des différences individuelles.

**Exemple 13.** *Les objets à l'étude sont les 5 journaux suivants.*

- $X_1$ . *Boston Herald*;
- $X_2$ . *New York Daily News*;
- $X_3$ . *New York Post*;
- $X_4$ . *New York Times*;
- $X_5$ . *Boston Globe*.

*Le tableau 1.2 donne le classement que chaque sujet  $Y_i$ ,  $i = 1, 2, 3, 4$ , fait des 5 journaux selon leurs préférences.*

*Dans le tableau 1.3, on a les coordonnées des points des journaux et des répondants dans l'espace dérivé.*

*À la figure 1.10, on a l'espace conjoint entre les journaux et les répondants. On y remarque que le sujet  $Y_3$  préfère en effet le Boston Herald et que le sujet  $Y_2$  aime mieux le New York*

Tableau 1.2 – Données des préférences des répondants

	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$
$Y_1$	1	3	2	5	4
$Y_2$	2	1	3	4	5
$Y_3$	1	2	3	5	4
$Y_4$	4	5	3	1	2

Tableau 1.3 – Coordonnées des répondants et des journaux

	1	2
$X_1$	-0.77	-0.18
$X_2$	-0.91	0.60
$X_3$	-0.29	-0.31
$X_4$	1.32	0.59
$X_5$	0.65	-0.70
$Y_1$	-0.71	-0.55
$Y_2$	-0.66	0.82
$Y_3$	-0.87	0.04
$Y_4$	1.11	-0.11

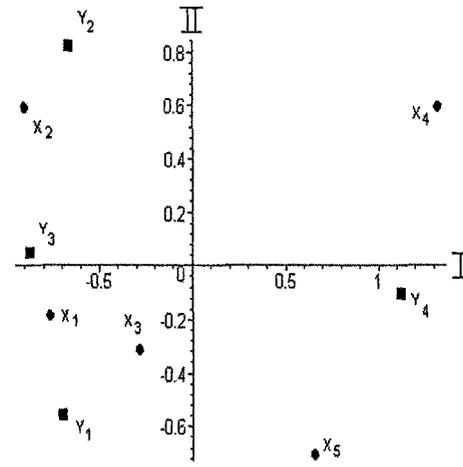
*Daily News. De plus, l'espace dérivé reflète le fait que les trois premiers répondants ont des goûts similaires contrairement au quatrième. On peut aussi voir que ces répondants trouvent que le Boston Herald et le New York Post sont très similaires.*

*La matrice  $D$  donne les distances entre les positions des journaux et des répondants. Ces distances sont les distances euclidiennes des points de l'espace conjoint calculés en deux dimensions par*

$$d_{ij} = \left[ \sum_{k=1}^2 (Y_{ik} - X_{jk})^2 \right]^{1/2}$$

*où les  $Y_{ik}$  sont les coordonnées du sujet  $i$  pour la dimension  $k$  et les  $X_{jk}$  sont les coordonnées du journal  $j$  pour la dimension  $k$ .*

Figure 1.10 – Espace conjoint des journaux et des répondants



$$D = \begin{matrix} & \begin{matrix} X_1 & X_2 & X_3 & X_4 & X_5 \end{matrix} \\ \begin{matrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{matrix} & \begin{pmatrix} 0.37 & 1.16 & 0.49 & 2.33 & 1.37 \\ 1.00 & 0.33 & 1.19 & 2.00 & 2.00 \\ 0.24 & 0.56 & 0.68 & 2.26 & 1.69 \\ 1.88 & 2.14 & 1.41 & 0.72 & 0.75 \end{pmatrix} \end{matrix}$$

### 1.3.7 Positionnement multidimensionnel pondéré

Jusqu'à présent, on a toujours travaillé avec des techniques que l'on appelle positionnement multidimensionnel non-pondéré. Ces techniques utilisent des matrices carrées ou rectangulaires des données. Dans ce premier cas, on peut représenter les objets qui sont comparés au niveau des similarités ou des préférences dans l'espace dérivé. Dans le deuxième cas, on crée une carte des perceptions contenant les stimuli à laquelle on ajoute les points représentant soit les répondants, soit les attributs à l'étude. On crée ainsi un espace conjoint.

Le positionnement multidimensionnel pondéré peut travailler avec ces deux genres de

matrices. Dans le cas présent, on exposera la situation où on a des matrices de données carrées. Cette technique est communément appelée différences individuelles d'échelonnement. L'espace pondéré représente la variation entre les matrices des proximités des sujets puisque l'on affiche l'importance que chaque répondant donne à chaque dimension de l'espace dérivé. Ainsi, le modèle pondéré peut être conceptualisé comme incluant des espaces additionnels qui reflètent la façon dont pense chacun des répondants.

Les différences individuelles d'échelonnement modifient le modèle de distances euclidiennes normales en incluant des sujets différentiels pondérés dans la formule de distances. Posons  $x_{jt}$  les coordonnées de la  $t^{\text{ième}}$  dimension pour le  $j^{\text{ième}}$  stimulus où  $t = 1, 2, \dots, r$ . Supposons que l'on a  $m$  répondants et  $n$  stimuli. La formule de la distance euclidienne utilisée dans le modèle pondéré de positionnement multidimensionnel pour le  $i^{\text{ième}}$  répondant est

$$d_{ijk} = \left[ \sum_{t=1}^r w_{it} (x_{jt} - x_{kt})^2 \right]^{1/2}$$

où  $w_{it}$  représente le poids que le répondant  $i$  donne à la dimension  $t$ . En posant  $x_{ijt} = \sqrt{w_{it}} x_{jt}$ , on obtient la forme standard de la formule de distance euclidienne

$$d_{ijk} = \left[ \sum_{t=1}^r (x_{ijt} - x_{ikt})^2 \right]^{1/2} .$$

En appliquant cette transformation à l'espace dérivé, on l'étend ou le contracte en modifiant l'importance des dimensions pour le sujet  $i$ .

Les poids que chaque répondant donne à chaque axe seront regroupés dans la matrice  $\mathbf{W}$  où les lignes représentent les répondants et les colonnes, les dimensions. Pour approximer ces poids, on se base sur le fait que les distances  $d_{ijk}$  sont similaires aux disparités  $\hat{d}_{ijk}$  où  $\hat{d}_{ijk}$  est la disparité entre le  $j^{\text{ième}}$  et le  $k^{\text{ième}}$  stimuli pour le  $i^{\text{ième}}$  répondant.

Sachant que

$$d_{ijk}^2 \cong \hat{d}_{ijk}^2,$$

on a

$$\hat{d}_{ijk}^2 \cong d_{ijk}^2 = \sum_{t=1}^r w_{it} (x_{jt} - x_{kt})^2.$$

Posons

$$p_{jkt} = (x_{jt} - x_{kt})^2.$$

Ainsi,

$$\hat{d}_{ijk}^2 \cong d_{ijk}^2 = \sum_{t=1}^r w_{it} p_{jkt}.$$

Puisqu'on cherche à isoler les  $w_{it}$ , il sera plus facile de travailler en notation matricielle.

On obtient donc

$$\hat{\mathbf{D}} = \mathbf{W}\mathbf{P}'$$

où  $\hat{\mathbf{D}}$  est une matrice  $m \times \frac{1}{2}n(n-1)$  où les lignes représentent les sujets et les colonnes, chaque paire de stimuli,  $\mathbf{W}$  est une matrice  $m \times r$  où les lignes sont encore les répondants et les colonnes, les dimensions et où  $\mathbf{P}'$  est une matrice  $r \times \frac{1}{2}n(n-1)$  où les lignes représentent les dimensions et les colonnes, chaque paire de stimuli. L'objectif étant de calculer  $\mathbf{W}$ , on l'isolera de  $\hat{\mathbf{D}} = \mathbf{W}\mathbf{P}'$  comme suit :

$$\mathbf{W}\mathbf{P}' = \hat{\mathbf{D}}$$

$$\mathbf{W}\mathbf{P}'\mathbf{P} = \hat{\mathbf{D}}\mathbf{P}$$

$$\mathbf{W}\mathbf{P}'\mathbf{P}(\mathbf{P}'\mathbf{P})^{-1} = \hat{\mathbf{D}}\mathbf{P}(\mathbf{P}'\mathbf{P})^{-1}$$

$$\mathbf{W} = \hat{\mathbf{D}}\mathbf{P}(\mathbf{P}'\mathbf{P})^{-1}$$

Ainsi, on obtient l'estimé de  $\mathbf{W}$ .

Pour mieux comprendre le positionnement multidimensionnel pondéré, étudions l'exemple trouvé dans le livre de Dillon et Goldstein [6]. On a cinq journaux à comparer listés comme suit

- $X_1$ . Boston *Herald* ;
- $X_2$ . New York *Daily News* ;
- $X_3$ . New York *Post* ;
- $X_4$ . New York *Times* ;
- $X_5$ . Boston *Globe*.

En appliquant une technique de positionnement multidimensionnel à la moyenne des données des répondants, on obtient les coordonnées suivantes en deux dimensions

$$\begin{array}{c} \\ \\ \\ \\ \\ \end{array} \begin{array}{cc} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \end{array} & \begin{pmatrix} 1.21 & -.81 \\ -.85 & -.93 \\ -.8 & -.7 \\ -.8 & 1.3 \\ 1.24 & 1.14 \end{pmatrix} \end{array} .$$

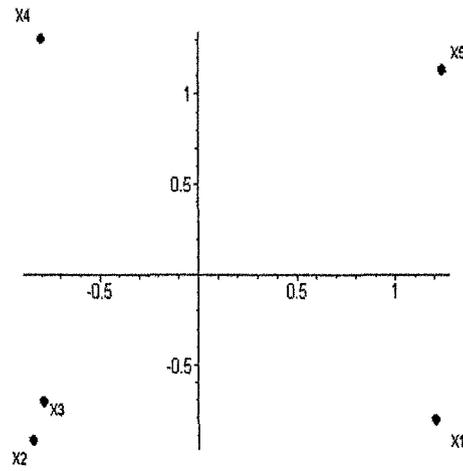
D'où on obtient l'espace dérivé représenté à la figure 1.11

À l'aide de la formule des distances euclidiennes entre les points  $j$  et  $k$  en  $r$  dimensions

$$d_{jk} = \left[ \sum_{t=1}^r (x_{jt} - x_{kt})^2 \right]^{1/2} ,$$

on calcule la matrice des distances  $\mathbf{D}$

Figure 1.11 – Espace dérivé des journaux



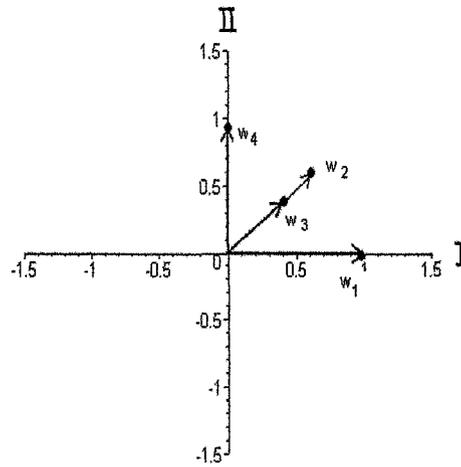
$$\mathbf{D} = \begin{matrix} & \begin{matrix} X_1 & X_2 & X_3 & X_4 & X_5 \end{matrix} \\ \begin{matrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \end{matrix} & \begin{pmatrix} 0 & 2.06 & 2.01 & 2.91 & 1.95 \\ 2.06 & 0 & 0.24 & 2.23 & 2.94 \\ 2.01 & 0.24 & 0 & 2.00 & 2.74 \\ 2.91 & 2.23 & 2.00 & 0 & 2.04 \\ 1.95 & 2.94 & 2.74 & 2.04 & 0 \end{pmatrix} \end{matrix} .$$

Une fois que toutes ces données sont recueillies, on calcule les valeurs des poids de chacune des dimensions pour chacun des sujets. Dans cet exemple, on obtient les poids trouvés dans la matrice  $\mathbf{W}$ . Chaque répondant est représenté par  $Y_i$ .

$$\mathbf{W} = \begin{matrix} & \begin{matrix} 1 & 2 \end{matrix} \\ \begin{matrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{matrix} & \begin{pmatrix} 0.98 & 0.00 \\ 0.60 & 0.60 \\ 0.40 & 0.40 \\ 0.00 & 0.92 \end{pmatrix} \end{matrix}$$

On peut représenter graphiquement ces poids comme à la figure 1.12.

Figure 1.12 – Poids des dimensions des journaux



La longueur de chacun de ces vecteurs  $w_i$  est représentée dans le tableau 1.4.

Tableau 1.4 – Longueurs des vecteurs  $w_i$

$w_1$	$w_2$	$w_3$	$w_4$
0.98	0.85	0.57	0.92

Ces vecteurs de poids expriment l'importance de chaque dimension de l'espace des stimuli pour chaque répondant. Dans la matrice  $\mathbf{W}$ , on a donc les coordonnées des sommets des vecteurs  $w_i$ . En faisant les projections des  $w_i$  sur les dimensions 1 et 2, on a l'importance de chaque dimension pour chaque répondant. De façon générale, plus la projection d'un vecteur sur un axe est loin de l'origine, plus cette dimension a de l'importance pour le répondant en question. À l'inverse, plus la projection est près de l'origine, moins la dimension a de l'importance. Prenons dans l'exemple le sujet 2 où  $w_2 = (0.60, 0.60)$ . La dimension I a la même importance que la II à un niveau de 0.60. Comparons-le au sujet 3 où  $w_3 = (0.40, 0.40)$ . Encore une fois, les deux dimensions ont la même importance de niveau 0.40. Par contre, ce dernier considère que les dimensions ont moins d'importance que le deuxième répondant. Le premier sujet, quant à lui, donne un poids de 0.98 à la

dimension 1 et une importance nulle pour la dimension 2. Pour faire l'analyse, on peut aussi étudier la longeurs des vecteurs que l'on retrouve dans le tableau 1.4. Ces longeurs sont calculées en sommant le carré de tous les poids du sujet et en extrayant la racine carrée.

À l'aide des coordonnées de  $\mathbf{X}$  et des poids  $\mathbf{W}$ , on peut produire un espace personnel pour chaque sujet que l'on dénote  $\mathbf{X}_i$ . Les coordonnées de ces espaces  $i$  sont

$$x_{ijt} = \sqrt{w_{it}}x_{jt}$$

où  $w_{it}$  est l'importance que le sujet  $i$  porte à la dimension  $t$  et  $x_{jt}$  sont les coordonnées de  $j^{\text{ième}}$  objet pour la  $t^{\text{ième}}$  dimension. Les distances euclidiennes pour la configuration  $\mathbf{X}_i$  sont

$$\begin{aligned} d_{ijk} &= \left[ \sum_{t=1}^r (x_{ijt} - x_{ikt})^2 \right]^{1/2} \\ &= \left[ \sum_{t=1}^r w_{it} (x_{jt} - x_{kt})^2 \right]^{1/2} . \end{aligned}$$

Dans l'exemple, une fois que l'on a calculé les coordonnées des configurations de chacun des sujets, on obtient les résultats présentés dans le tableau 1.5.

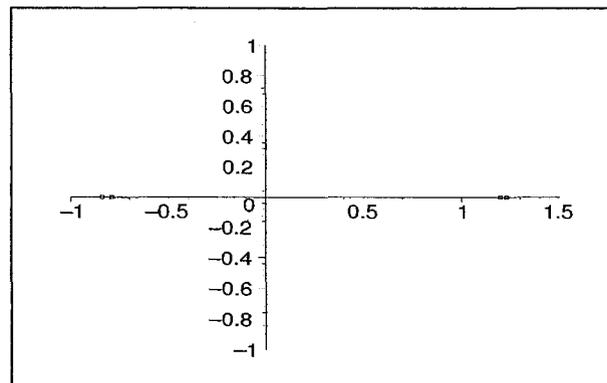
Ces configurations sont représentées graphiquement par les figures 1.13,1.14, 1.15 et 1.16.

On sait que plus la dimension est petite, moins cette dimension est importante pour le sujet. Ainsi, cette dimension sera réduite de beaucoup. Au contraire, si le poids du sujet est grand, la dimension aura de l'importance pour celui-ci et la dimension sera à peine réduite. Donc, elle ne rapetissera pas beaucoup. Les poids des sujets compris entre 0 et 1 réduisent l'espace dérivé tandis que ceux qui sont plus grand que 1 l'étirent. Il est à noter que dans le cas présent, il n'y aura pas d'étirement de dimension puisque tous les

Tableau 1.5 – Coordonnées des 4 configurations

	$X_1$		$X_2$		$X_3$		$X_4$	
	1	2	1	2	1	2	1	2
$X_1$	1.20	0.00	0.94	-0.63	0.77	-0.51	0.00	-0.78
$X_2$	-0.84	0.00	-0.66	-0.72	-0.54	-0.59	0.00	-0.89
$X_3$	-0.79	0.00	-0.62	-0.54	-0.51	-0.44	0.00	-0.67
$X_4$	-0.79	0.00	-0.62	1.01	-0.51	0.82	0.00	1.25
$X_5$	1.23	0.00	0.96	0.88	0.78	0.72	0.00	1.09

Figure 1.13 – Espace multidimensionnel du premier répondant



poids sont compris entre 0 et 1. Ainsi, pour le premier répondant, la majeure partie de la réduction se fait sur la deuxième dimension car son poids est nul. Pour le deuxième, on a des poids identiques pour les deux dimensions. L'espace dérivé est donc le même mais en plus petit. On retrouve le même cas pour le troisième répondant mais comme il apporte moins d'importance aux dimensions, sa configuration sera encore plus petite que la deuxième. Le quatrième répondant se situe dans la même position que le premier mais en inversant les dimensions.

On peut aussi comparer ces configurations en calculant les distances entre chaque objet pour chaque espace dérivé personnel. Les matrices  $D_{1,2}$  et  $D_{3,4}$  expriment ces distances où la matrice  $D_{i,j}$  représente les distances des points pour le répondant  $i$  sous la diagonale

Figure 1.14 – Espace multidimensionnel du deuxième répondant

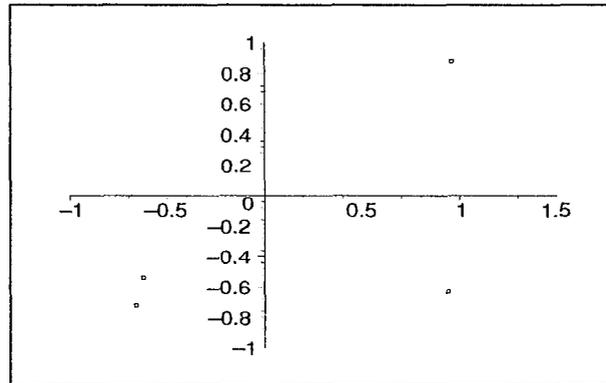
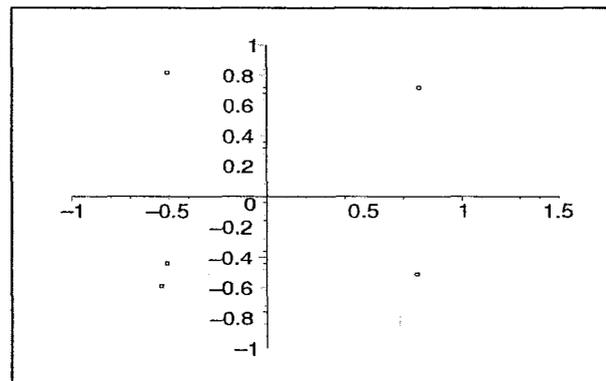


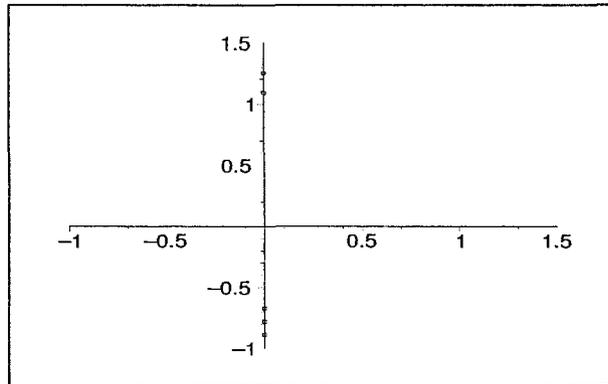
Figure 1.15 – Espace multidimensionnel du troisième répondant



et les distances des points pour le répondant j par-dessus la diagonale.

$$\mathbf{D}_{1,2} = \begin{matrix} & \begin{matrix} X_1 & X_2 & X_3 & X_4 & X_5 \end{matrix} \\ \begin{matrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \end{matrix} & \begin{pmatrix} 0 & 1.60 & 1.56 & 2.26 & 1.51 \\ 2.04 & 0 & 0.18 & 1.73 & 2.28 \\ 1.99 & 0.05 & 0 & 1.55 & 2.12 \\ 1.99 & 0.05 & 0 & 0 & 1.59 \\ 0.03 & 2.07 & 2.02 & 2.02 & 0 \end{pmatrix} \end{matrix}$$

Figure 1.16 – Espace multidimensionnel du quatrième répondant



$$\mathbf{D}_{3,4} = \begin{matrix} & \begin{matrix} X_1 & X_2 & X_3 & X_4 & X_5 \end{matrix} \\ \begin{matrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \end{matrix} & \begin{pmatrix} 0 & 0.11 & 0.11 & 2.03 & 1.87 \\ 1.31 & 0 & 0.22 & 2.14 & 1.98 \\ 1.28 & 0.15 & 0 & 1.92 & 1.76 \\ 1.85 & 1.41 & 1.26 & 0 & 0.16 \\ 1.23 & 1.86 & 1.73 & 1.29 & 0 \end{pmatrix} \end{matrix}$$

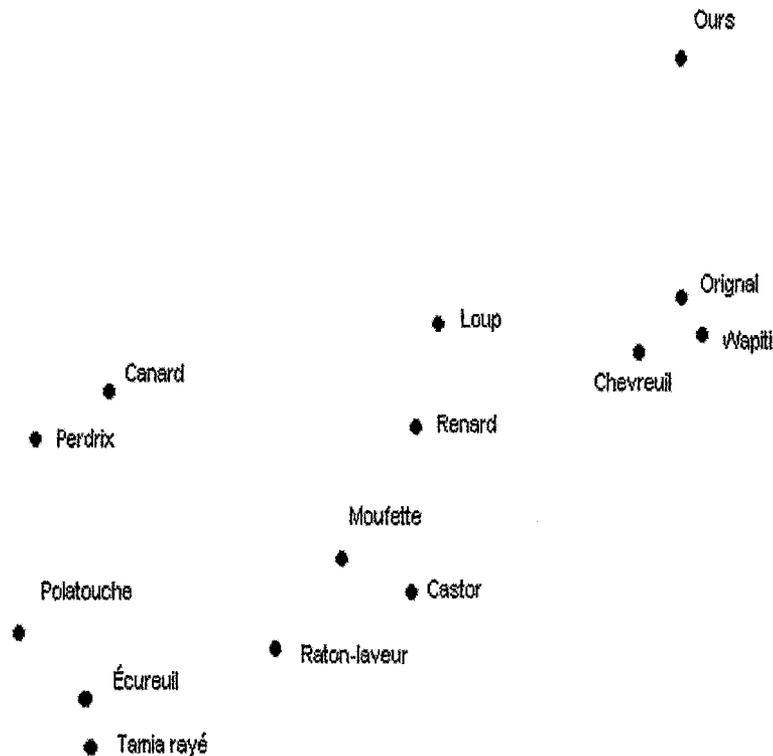
Le positionnement multidimensionnel pondéré permet donc de créer plusieurs espaces dérivés, c'est-à-dire un par répondant, qui représentent l'importance que chacun porte à chaque dimension. Les poids des sujets ont donc pour effet d'étirer ou de réduire les dimensions de l'espace dérivé obtenu par une méthode non-pondérée de positionnement multidimensionnel.

### 1.3.8 Ajustement des attributs influents

Une fois que l'espace multidimensionnel est créé, il existe une procédure objective, mentionnée plus tôt, pour inclure les vecteurs attributs dans cet espace. Cette procédure se nomme l'ajustement des attributs influents (property fitting). Elle permet d'interpréter l'espace dérivé et d'identifier les dimensions. Si on a une variable mesurant une caracté-

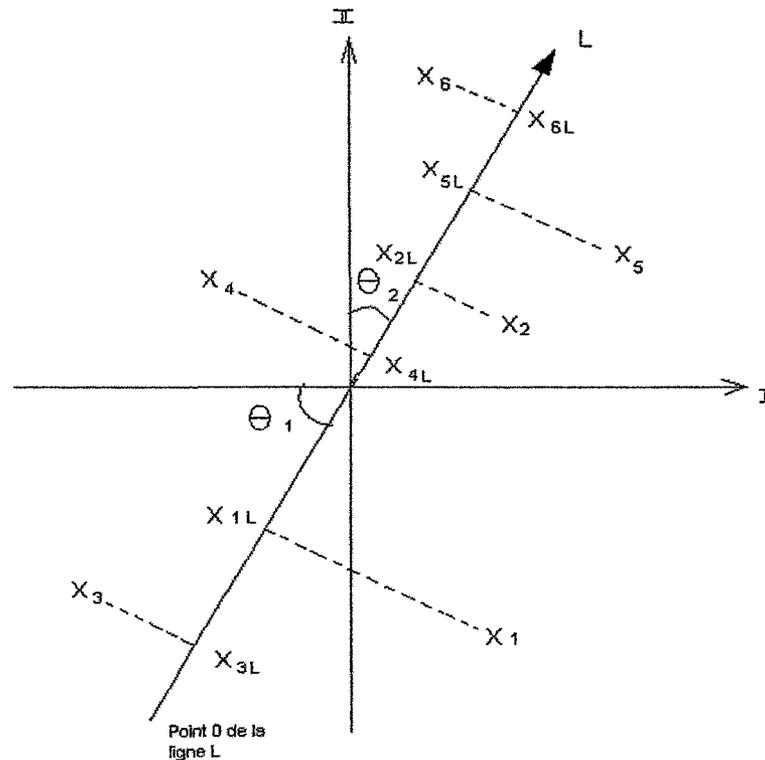
ristique des objets à l'étude et qu'on la suspecte d'avoir une meilleure relation avec la position des objets dans l'espace dérivé, on tentera d'utiliser cette variable pour expliquer la configuration.

Figure 1.17 – Carte des perceptions des animaux



Supposons que l'on ait une liste de 14 animaux et que l'on ait la carte des perceptions représentée à la figure 1.17. On peut donc remarquer qu'en partant d'en bas à gauche jusqu'à en haut à droite, la taille des animaux augmente selon la place qu'ils occupent dans l'espace. Il serait donc intéressant d'aller calculer exactement la direction de ce vecteur qui représenterait l'attribut de la taille d'un animal. L'évaluation de ces attributs à l'étude est faite par les répondants qui donnent leur avis sur le degré de concordance entre chacun des objets et chacun des attributs.

Figure 1.18 – Exemple de vecteur d'attribut provenant de l'ajustement des attributs influents



Prenons par exemple la figure 1.18. On a une représentation graphique de ce que l'on tente de faire. On a la carte des perceptions basées sur les similarités des éléments  $X_1, X_2, \dots, X_6$ . On cherche une direction à travers cet espace qui correspond à un accroissement de la signification de l'attribut. Dans la figure X se trouve une ligne L qui se nomme vecteur d'attribut. Il a comme propriété que la projection de chaque objet  $X_i$  sur lui-même correspond le mieux possible avec la signification qu'a cet objet sur cet attribut. Les points  $X_{iL}$  sont les projections des points  $X_i$  sur la droite L. Si l'attribut en question est fortement relié à l'espace de stimuli dérivé, alors les valeurs actuelles des attributs s'accorderont grandement avec la projection du stimulus et ils auront un haut niveau de corrélation. Si l'attribut en question n'est pas fortement relié à l'espace, la corrélation entre la valeur actuelle de l'attribut et de la projection de l'objet sera faible.

La valeur du coefficient de corrélation indique la force de la relation entre l'attribut en question et l'espace des stimuli. Si la valeur est basse, on peut conclure que les répondants n'utilisaient pas cet attribut latent quand ils ont fait leur jugement de similarités. Si la valeur est élevée, on est porté à croire que le sujet a utilisé cet attribut latent ou d'autres qui sont fortement liés.

Pour trouver la direction recherchée d'un vecteur, il s'agit d'utiliser la régression multiple. Cette procédure trouve la direction du vecteur attribut telle que la corrélation entre la projection du stimulus et la valeur de l'attribut est aussi grande que possible. On peut décrire ce processus de façon détaillée en 4 étapes.

Étape 1 :

Il faut obtenir la moyenne de chaque attribut caractéristique.

Étape 2 :

Il faut régresser le vecteur moyen de l'évaluation de l'attribut en question que l'on nomme  $y$  sur les coordonnées de l'espace dérivé et traiter chacune des données comme une variable indépendante. On peut supposer, sans perte de généralité, que  $y$  est centré à l'origine. On doit prendre  $y_i$  comme valeur spécifique du stimulus  $i$  sur l'attribut  $y$  où  $i = 1, 2, \dots, n$  et  $n$  représente la nombre d'objets. De plus, prendre  $X_{i1}, X_{i2}, \dots, X_{ir}$  comme coordonnées du stimulus  $i$  sur chacune des  $r$  dimensions. On peut supposer, sans perte de généralité, que les vecteurs  $\mathbf{X}_i$  sont centrés à l'origine.

L'équation de régression multiple est donnée par

$$y_i = b_0 + b_1 X_{i1} + b_2 X_{i2} + \dots + b_r X_{ir}.$$

Les valeurs inconnues de  $b_1, b_2, \dots, b_r$  sont appelés les paramètres de régression et  $b_0$  est l'ordonnée à l'origine.

La valeur de  $\hat{y}_i$  est la meilleure estimation de la projection du stimulus  $i$  sur le vecteur attribut  $\mathbf{y}$  provenant des coordonnées du stimulus  $X_{it}$ ,  $t = 1, 2, \dots, r$  et de la valeur de l'attribut  $y_i$ . Pour estimer le vecteur  $\mathbf{b} = (b_0, b_1, b_2, \dots, b_r)'$ , on doit calculer  $\beta = (\beta_0, \beta_1, \dots, \beta_r)'$  où

$$\beta = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

et où  $\mathbf{X} = (\mathbf{1}, \mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_r)$  et  $\mathbf{1}$  est le vecteur unitaire de dimension  $r \times 1$ .

Une fois que l'on a la valeur de  $\beta$ , on peut calculer  $\hat{\mathbf{y}}$  comme suit

$$\hat{\mathbf{y}} = \mathbf{X}\beta$$

Étape 3 :

On doit calculer le coefficient de corrélation multiple, noté  $R$ , à l'aide de la formule

$$R^2 = 1 - \frac{(\mathbf{y} - \hat{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}})}{(\mathbf{y} - \bar{\mathbf{y}})'(\mathbf{y} - \bar{\mathbf{y}})}$$

Puisqu'on travaille avec un vecteur  $\mathbf{y}$  centré à l'origine, alors  $\bar{\mathbf{y}} = 0$ . Ainsi,

$$R^2 = 1 - \frac{(\mathbf{y} - \hat{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}})}{\mathbf{y}'\mathbf{y}}$$

La racine de ce coefficient  $R^2$  donne la corrélation  $R$  entre la projection du stimulus et les valeurs de l'attribut. Il est à noter que la valeur de  $R^2$  est comprise entre 0 et 1. Plus la valeur de  $R$  est basse, plus on est porté à croire que les répondants n'ont pas porté attention à cet attribut lorsqu'ils ont fait leur évaluation des similarités. Au contraire, plus il est haut, plus ils ont évalué sous l'influence de cet attribut.

Étape 4 :

Pour représenter graphiquement le vecteur attribut, il faut trouver le point dans l'espace

dérivé tel que ses coordonnées sont  $(\beta_1, \beta_2, \dots, \beta_r)$ . On nommera ce point  $\beta^*$ .

Puisque la moyenne des coordonnées est nulle, il suffit de tracer une ligne de l'origine de l'espace dérivé des stimuli au point  $\beta^*$ . La tête de la flèche est positionnée au bout de la droite, c'est-à-dire où le point  $\beta^*$ .

On peut ensuite calculer les angles formés entre le vecteur attribut et chacune des dimensions comme suit

$$\cos\theta_i = \frac{\beta_i}{\|\beta\|}$$

où  $i = 1, 2, \dots, r$  et

$$\|\beta\| = \left[ \sum_{i=1}^r \beta_i^2 \right]^{1/2}.$$

On a dénoté les directions des cosinus dans la figure 1.18 par  $\theta_1$  et  $\theta_2$ .

Il est à noter que si, dans l'étude, on a à évaluer plusieurs attributs, il n'est pas nécessaire de tous les représenter. Pour choisir les attributs que l'on souhaite éliminer, on examine les coefficients de corrélation de Pearson entre les attributs en calculant

$$\begin{aligned} r_{ij} &= \frac{\text{Cov}(\mathbf{X}_i, \mathbf{X}_j)}{\sqrt{\text{Var}(\mathbf{X}_i)}\sqrt{\text{Var}(\mathbf{X}_j)}} \\ &= \frac{(\mathbf{X}_i - \bar{\mathbf{X}}_i)'(\mathbf{X}_j - \bar{\mathbf{X}}_j)}{((\mathbf{X}_i - \bar{\mathbf{X}}_i)'(\mathbf{X}_i - \bar{\mathbf{X}}_i))^{1/2}((\mathbf{X}_j - \bar{\mathbf{X}}_j)'(\mathbf{X}_j - \bar{\mathbf{X}}_j))^{1/2}} \end{aligned}$$

S'il y en a plusieurs qui montrent tous deux-à-deux un haut niveau de corrélation entre eux, on n'en choisira qu'un seul pour les représenter.

## CHAPITRE 2

# Le positionnement mutidimensionnel : Exemples introductifs détaillés

### 2.1 Représentation graphique des villes d'Angleterre

Le premier exemple introductif présenté est tiré de [1]. On a donc dans la matrice des distances  $D$ , toutes les distances entre les 12 villes d'Angleterre. Ces distances sont celles calculées par les routes. En partant de cette matrice et en utilisant le théorème 1, on va recréer la carte de cette ville. Ensuite, en utilisant les standardisations rotationnelles, on superposera cette carte avec la vraie représentation cartographique pour les comparer.

$$D = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \\ 11 \\ 12 \end{matrix} & \begin{pmatrix} 0 & 244 & 218 & 284 & 197 & 312 & 215 & 469 & 166 & 212 & 253 & 270 \\ 244 & 0 & 350 & 77 & 167 & 444 & 221 & 583 & 242 & 53 & 325 & 168 \\ 218 & 350 & 0 & 369 & 347 & 94 & 150 & 251 & 116 & 298 & 57 & 287 \\ 284 & 77 & 369 & 0 & 242 & 463 & 236 & 598 & 257 & 72 & 340 & 164 \\ 197 & 167 & 347 & 242 & 0 & 441 & 279 & 598 & 269 & 170 & 359 & 277 \\ 312 & 444 & 94 & 463 & 441 & 0 & 245 & 169 & 210 & 392 & 143 & 378 \\ 215 & 221 & 150 & 236 & 279 & 245 & 0 & 380 & 55 & 168 & 117 & 143 \\ 469 & 583 & 251 & 598 & 598 & 169 & 380 & 0 & 349 & 531 & 264 & 514 \\ 166 & 242 & 116 & 257 & 269 & 210 & 55 & 349 & 0 & 190 & 91 & 173 \\ 212 & 53 & 298 & 72 & 170 & 392 & 168 & 531 & 190 & 0 & 273 & 111 \\ 253 & 325 & 57 & 340 & 359 & 143 & 117 & 264 & 91 & 273 & 0 & 256 \\ 270 & 168 & 284 & 164 & 277 & 378 & 143 & 514 & 173 & 111 & 256 & 0 \end{pmatrix} \end{matrix}$$

Pour obtenir les coordonnées en deux dimensions, on doit calculer les valeurs propres de  $\mathbf{B} = \mathbf{H}\mathbf{A}\mathbf{H}$  où  $\mathbf{A} = (a_{rs})$  et  $a_{rs} = -\frac{1}{2}d_{rs}^2$ . Voici la liste des 12 valeurs propres de  $\mathbf{B}$  :

$$\begin{aligned} \lambda &= (\lambda_1, \lambda_2, \dots, \lambda_{12}) \\ &= (394472.7934, 63634.37442, 13544.07385, 10245.31428, 2464.677912, 1450.264499, \\ &\quad 500.8959775, -0.06384438, -19.87868556, -214.0725462, -1141.217055, -7063.413056) \end{aligned}$$

Les vecteurs propres associés aux deux plus grandes valeurs propres de  $\mathbf{B}$  sont

$$\mathbf{x}_1 = (-0.0985935492, -0.4418627588, 0.3015707457, -0.4619185961, -0.4122912057, 0.5102100535, 0.01694603830, 0.8322993545, 0.0691057130, -0.3332921459, 0.2617687852, -0.2439437336)$$

et

$\mathbf{x}_2 = (0.5591644698, -0.07047269668, 0.1234321028, -0.3047103160, 0.5601533026, 0.1241938226, -0.2010594489, -0.1027158556, -0.02189906405, -0.1073797054, -0.1346109535, -0.4240956601)$ .

Une fois que l'on a standardisé les vecteurs propres, les coordonnées de nos douze points sont donc

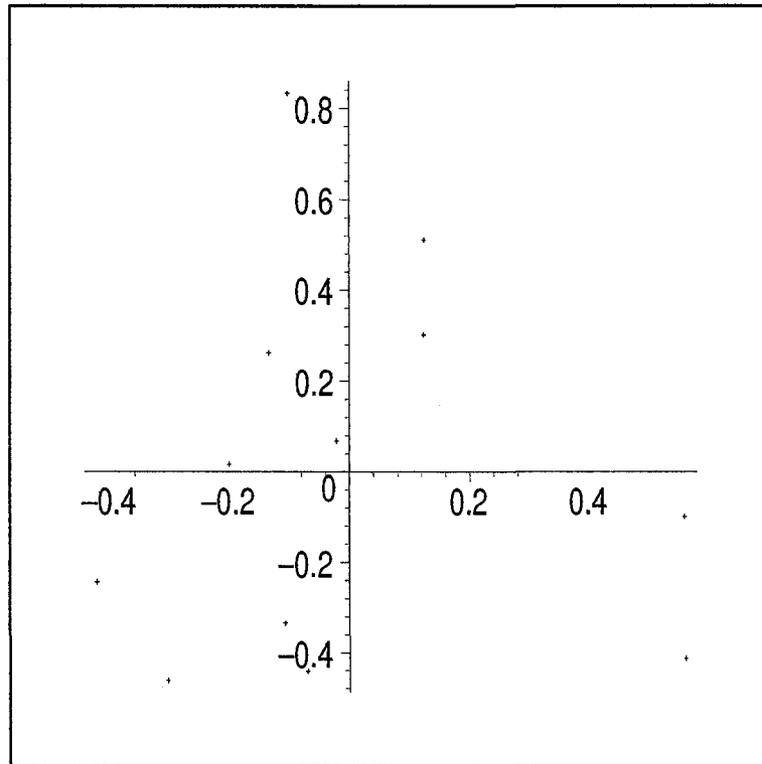
$$\begin{aligned}
 P_1 &= (-45.20477992, 139.9476497) \\
 P_2 &= (-202.5930672, -17.63790227) \\
 P_3 &= (138.2694992, 30.89257923) \\
 P_4 &= (-211.7886229, -76.26287948) \\
 P_5 &= (-189.0341191, 140.1951347) \\
 P_6 &= (233.9301460, 31.08322242) \\
 P_7 &= (7.769719915, -50.32114663) \\
 P_8 &= (381.6073560, -25.70771814) \\
 P_9 &= (31.68481181, -5.480896428) \\
 P_{10} &= (-152.8136888, -26.87498618) \\
 P_{11} &= (120.0203911, -33.69042131) \\
 P_{12} &= (-111.8476455, -106.1426362)
 \end{aligned}$$

et ils sont représentés à la figure 2.1

On va maintenant utiliser la définition 9 pour calculer les mesures d'acceptation pour les proportions de la matrice des distances D.

$$\begin{aligned}
 \alpha_{1,k} &= \left( \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^n |\lambda_i|} \right) \times 100\% \\
 &= \left( \frac{\lambda_1 + \lambda_2}{\sum_{i=1}^{12} |\lambda_i|} \right) \times 100\% \\
 &= .9259403463
 \end{aligned}$$

Figure 2.1 – Points des coordonnées calculés des 12 villes d'Angleterre



$$\begin{aligned}
 \alpha_{2,k} &= \left( \frac{\sum_{i=1}^k \lambda_i^2}{\sum_{i=1}^n \lambda_i^2} \right) \times 100\% \\
 &= \left( \frac{\lambda_1^2 + \lambda_2^2}{\sum_{i=1}^{12} \lambda_i^2} \right) \times 100\% \\
 &= .9978245999
 \end{aligned}$$

Ces mesures de concordance montrent que cette représentation en deux dimensions sera très près de la réalité. On pourra la comparer visuellement en utilisant les standardisations rotationnelles. On ajustera donc cette configuration pour l'amener dans le même espace que la carte originale.

Les véritables coordonnées de ces 12 villes d'Angleterre sont :

$$\mathbf{X} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \end{matrix} \\ \begin{matrix} Est \\ Nord \end{matrix} & \begin{pmatrix} 257 & 529 & 339 & 629 & 292 & 259 & 508 & 265 & 433 & 533 & 420 & 627 \\ 279 & 104 & 554 & 142 & 90 & 665 & 433 & 842 & 438 & 183 & 563 & 308 \end{pmatrix} \end{matrix}$$

Pour pouvoir utiliser les standardisations rotationnelles, on travaillera avec le  $\mathbf{X}$  centré.

Voici maintenant les coordonnées qu'on a calculé à l'aide du théorème 1.

$$\mathbf{Y} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \end{matrix} \\ \begin{matrix} Est \\ Nord \end{matrix} & \begin{pmatrix} 45 & 203 & -138 & 212 & 189 & -234 & -8 & -382 & -32 & 153 & -120 & 112 \\ 140 & -18 & 31 & -76 & 140 & 31 & -50 & -25 & -5 & -27 & -34 & -106 \end{pmatrix} \end{matrix}$$

Puisque  $\mathbf{Y}$  est déjà centrée, on travaille directement avec cette matrice. On va donc calculer les valeurs des matrices  $\mathbf{U}$ ,  $\mathbf{V}$ ,  $\mathbf{\Gamma}$ . Pour ce faire, commençons par les valeurs propres de  $\mathbf{Z}^t\mathbf{Z}$  (qui sont les mêmes que celles de  $\mathbf{Z}\mathbf{Z}^t$ ). On obtient que ces valeurs propres sont :

$$\lambda_1 = 0.8976323347 \times 10^{10}$$

$$\lambda_2 = 0.2787904118 \times 10^{10}.$$

Ainsi,

$$\mathbf{\Gamma} = \begin{pmatrix} 528006.0718 & 0 \\ 0 & 94743.46071 \end{pmatrix}.$$

Calculons les vecteurs propres de  $\mathbf{Z}^t\mathbf{Z}$  et de  $\mathbf{Z}\mathbf{Z}^t$  pour trouver respectivement  $\mathbf{U}$  et  $\mathbf{V}$ .

On obtient

$$\mathbf{U} = \begin{pmatrix} -0.3478133903 & 0.9375647821 \\ 0.9375637821 & 0.3478133903 \end{pmatrix}$$

et

$$\mathbf{V} = \begin{pmatrix} -0.9998921152 & -0.01468870883 \\ 0.01468870883 & -0.9998921152 \end{pmatrix}.$$

On obtient donc que

$$\mathbf{A} = \begin{pmatrix} 0.3340042651 & -0.9425715628 \\ -0.9425715628 & -0.3340042651 \end{pmatrix}$$

$$\mathbf{b} = \begin{pmatrix} 424.250 \\ 383.417 \end{pmatrix}$$

$$c = 1.357912493.$$

On trouve donc la nouvelle configuration en appliquant les transformations suivantes :

$$c\mathbf{A}'\mathbf{y}_r + \mathbf{b}.$$

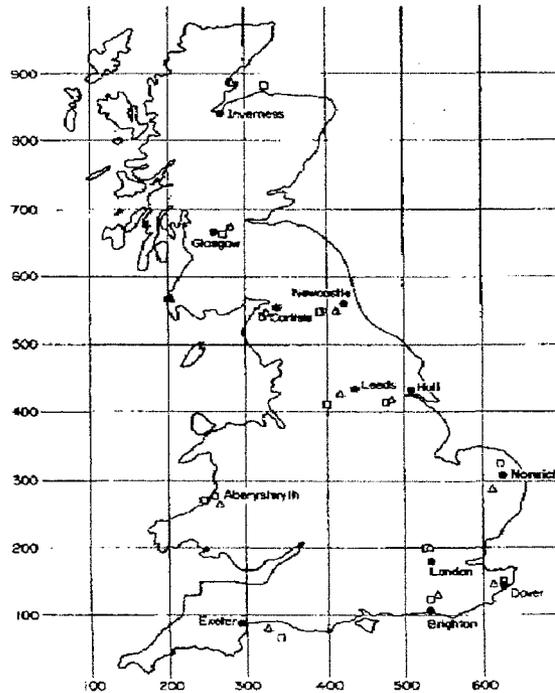
On peut remarquer sur la carte représentée à la figure 2.2 que la vraie configuration et celle calculée sont très proches puisque les points représentent les positions originales et les triangles, la solution classique.

De plus, la valeur minimale de la somme résiduelle des carrées est

$$R^2 = 8277.797$$

ce qui signifie que la configuration calculée par la solution classique de positionnement multidimensionnel est bonne. On arrive à cette conclusion en comparant ce  $R^2$  avec celui calculé pour la solution de Shepard-Kruskal donnée dans le livre de Mardia, Kent et Bibby [1] qui est 13 749.

Figure 2.2 – Carte des 12 villes d'Angleterre



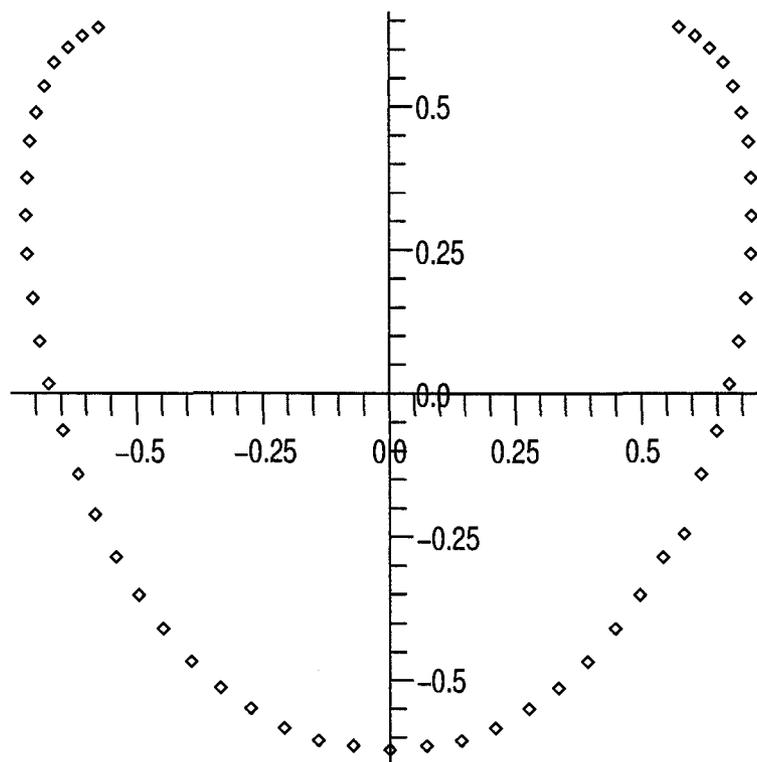
## 2.2 L'effet du fer à cheval

Il existe un cas particulier où il est aisé de calculer les distances des points qui sont près les uns des autres. Par contre, s'ils sont trop loin, il n'y a plus de lien entre les deux et pourtant ils sont représentés près les uns des autres. Cet effet a été étiqueté par David G.Kendall [1] en 1971 comme étant l'effet du fer à cheval. L'exemple suivant caractérise très bien cet effet.

**Exemple 14.** *Considérons la matrice des similarités  $C$  ( $51 \times 51$ ) définie par*

$$c_{rs} = \begin{cases} 9 & \text{si } r=s \\ 8 & \text{si } 1 \leq |r-s| \leq 3 \\ 7 & \text{si } 4 \leq |r-s| \leq 6 \\ \dots & \\ 1 & \text{si } 22 \leq |r-s| \leq 24 \\ 0 & \text{si } |r-s| \geq 25 \end{cases}$$

Figure 2.3 – Représentation graphique en deux dimensions de la matrice de Kendall



*Il faut calculer la matrice des distances à partir de nos similarités à l'aide de la formule donnée à la définition 2. En utilisant l'algorithme 1, on obtient les valeurs propres de  $\mathbf{B} = \mathbf{H}\mathbf{A}\mathbf{H}$ . Au total, huit valeurs propres sont négatives comprises entre -0.0945 et -2.0743 et quarante-trois positives 126.0857, 65.9353, 18.1710, 7.8208, 7.6103, ..., 0. En calculant les vecteurs propres des deux plus grandes valeurs propres, on trouve les coordonnées de nos cinquante et un points. Ils sont illustrés à la figure 2.3. On remarque que les points*

*les plus éloignés s'attirent. Cette configuration ressemble à la circonférence d'un cercle.*

## 2.3 Positionnement de produits

Le positionnement de produits est une application du positionnement multidimensionnel dans le domaine du marketing. Cette technique sert à déterminer la place qu'occupe un produit dans un marché. Le but est donc de concevoir le marché comme un espace multidimensionnel dans lequel chaque produit est positionné. À la base, le concept de positionnement a été créé en rapport au travail effectué par les économistes sur la structure du marché, la position de la compétitivité de la compagnie en question et sur les concepts de compétition entre les produits.

Le positionnement multidimensionnel appliqué au positionnement de produits donne une perspective mettant l'accent sur la perception des consommateurs concernant la place qu'occupe le produit sur le marché. Cette approche donne donc une évaluation beaucoup plus subjective puisque ce sont les acheteurs et les consommateurs qui donnent leur perceptions et leurs préférences. De plus, ces perceptions sont fondées sur des pensées structurées, mais intangibles pour le chercheur. La découverte et la dénomination des axes simplifieront grandement la compréhension des concepts latents, préalablement inconnu des chercheurs.

Créé par les experts en marketing, l'ensemble des stimuli doit tenir compte de tous les produits en compétition. Il est donc préférable d'inclure deux types de produits, une même classe de produits et d'autres qui peuvent servir de substituts.

**Exemple 15.** *Lors d'une étude de soupe, les éléments sélectionnés dans l'ensemble primaire sont plusieurs marques de soupes ayant des saveurs différentes. L'ensemble de substituts sera donc composé de repas pouvant remplacer la soupe mais qui concernent certains de ces attributs comme la rapidité de préparation et la consistance légère. On pourrait y inclure par exemple la salade, les sandwiches ou les crudités.*

Tel que vu dans le chapitre 1, une fois que les objets sont sélectionnés et que les données similaires, de préférences ou les deux sont recueillies, il faut choisir la technique utilisée pour faire la collection des données. Cette procédure dépend en majeure partie du nombre d'items ou des préférences du chercheur.

En résumé, le but est de concevoir le marché comme un espace multidimensionnel où les produits, les marques ou les compagnies sont positionnés en fonction soit de leurs similarités avec les compétiteurs, de leurs attributs ou du fait qu'ils sont les préférés ou non.

Pour illustrer le positionnement multidimensionnel dans le positionnement de produits, l'article de Wind et Robinson [3] propose plusieurs exemples dont :

- l'étude des produits diététiques ;
- l'étude des journaux médicaux ;
- l'étude des services financiers ;
- l'étude des détaillants.

### 2.3.1 Positionnement des nouveaux concepts diététiques

Dans cette étude, les stimuli sont les nouveaux produits diététiques. Le but est donc d'évaluer leur positionnement en tentant d'obtenir une meilleure compréhension du positionnement de produits avec une comparaison des évaluations objectives et subjectives.

La première tâche des répondants, qui sont des femmes à la diète dans le cas présent, sera de classer les aliments et treize concepts de produits diététiques. Les aliments à classer seront des produits diététiques comme Metrecal et d'autres types d'aliments comme des croustilles, des hot-dog, des pommes ou du yogourt. Le classement se fera donc par similarités.

La seconde tâche sera d'évaluer les produits et les concepts selon leurs préférences générales et sur certains types de repas servis ou certaines occasions pour manger comme par exemple :

- des aliments servis pour le lunch ;
- un régime draconien pour perdre du poids rapidement ;
- améliorer son apparence ;
- des aliments servis pour le dîner ;
- des aliments servis quand on est seul ;
- etc.

La troisième tâche sera d'évaluer les concepts et les produits sur un ensemble de douze attributs :

1. calories ;
2. nutrition ;

3. goût ;
4. commodité de livraison ;
5. vitamines ;
6. capacité d'être un bon substitut ;
7. cholestérol ;
8. protéines ;
9. gras ;
10. sucre ;
11. prix ;
12. hydrate de carbone.

En résumé, les trois ensembles de données subjectives sont :

- les données de produits similaires ;
- le classement des préférences des produits ;
- le classement des produits sur plusieurs attributs.

Les données objectives seront fournies par un groupe de techniciens alimentaires qui évalueront les multiples produits selon les attributs objectifs actuels.

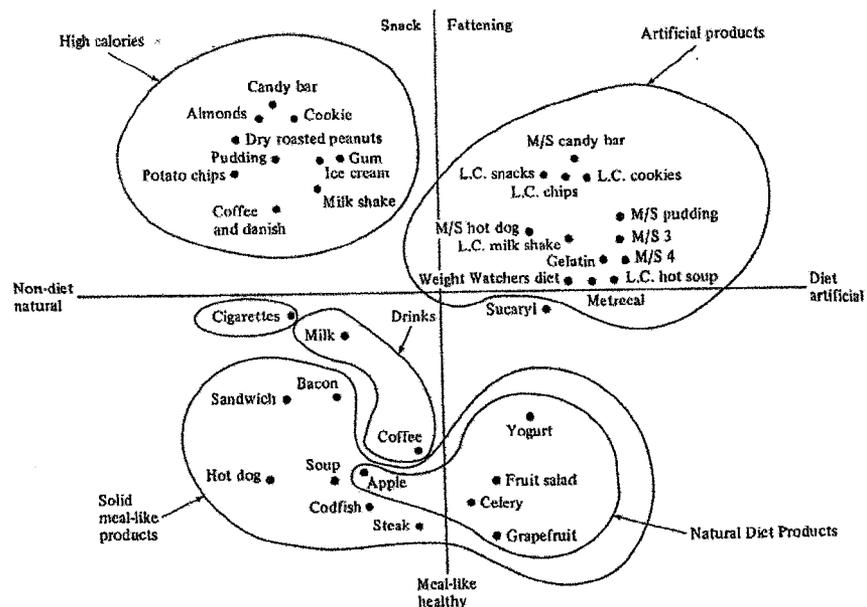
La première carte des perceptions conçue est celle des données similaires et est donnée à la figure 2.4. En examinant ce que chaque grappe contient, l'analyste peut ainsi leur suggérer un titre. Voici donc les cinq grappes suggérées :

1. des aliments à hautes teneurs en calories comme la crème glacé, les croustilles, les biscuits, etc. ;
2. des aliments diététiques non-naturels comme les biscuits L.C. et des concepts diététiques comme Metrecal ;

3. des aliments diététiques naturels comme du yogourt et du céleri ;
4. des repas solides comme du steak et de la soupe ;
5. des breuvages comme du café et du lait.

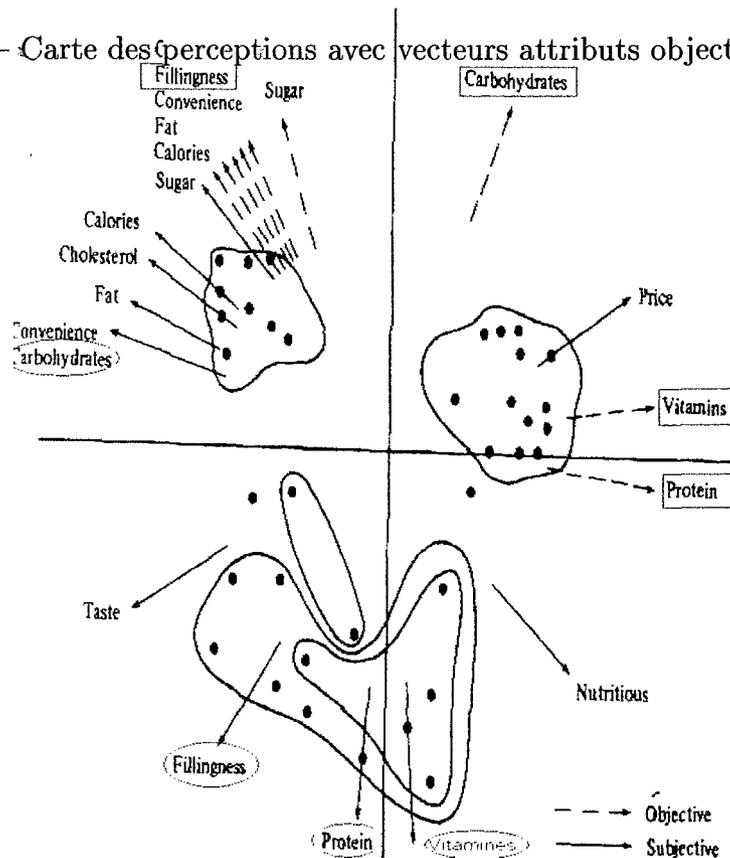
En examinant la disposition des produits, on peut ainsi déterminer les dimensions. La dimension 1 va de produits non-diététiques naturels à produits diététiques artificiels. La dimension 2 varie de repas sants à casse-croûtes engraisants.

Figure 2.4 – Carte des perceptions des aliments et des produits diététiques



Tous les outils sont maintenant présents pour faire l'analyse de la carte des perceptions. Ainsi, il découle que les nouveaux concepts diététiques compétitionneront avec les autres produits diététiques comme Metrecal et Weight Watcher. Par contre, un plus grand aperçu de la position de chaque concept et produit a été obtenu par l'examen de la configuration de l'espace conjoint des produits et concepts et leurs attributs aperçus. La figure 2.5 représente cette carte des perceptions.

Figure 2.5 – Carte des perceptions avec vecteurs attributs objectifs et subjectifs



Les vecteurs ayant les lignes pleines représentent les attributs jugés de façon subjective par les femmes qui sont à la diète. Les vecteurs ayant des pointillés représentent les attributs observés de façon objective par les techniciens alimentaires. On peut ainsi remarquer que les personnes à la diète ne perçoivent pas nécessairement les différents aliments selon leurs attributs objectifs. Cinquante pourcent ou plus des divergences entre les vecteurs attributs objectifs et subjectifs se situent au niveau des hydrates de carbone, de la capacité d'être un bon substitut, des protéines et des vitamines. Une grande congruence, par contre, se situe au niveau des calories, du sucre, du gras, du cholestérol et de la facilité de préparation. Ceci suggère que les personnes à la diète sont plus consciencieuses et intéressées par cet ensemble d'attributs.

Les chercheurs s'attardent maintenant à l'analyse subjective de cette dernière carte des

perceptions. Les produits diététiques comme le yogourt et les pommes sont vus comme étant nutritifs et sains. Les produits comme les croustilles et le chocolat sont perçus comme étant riches en calories, en gras et très mauvais pour le cholestérol. Les aliments des repas naturels sont riches en protéines et en vitamines. Les nouveaux concepts diététiques sont perçus comme étant dispendieux, moins nutritifs que les produits diététiques naturels, mauvais pour la santé, pauvre en protéines et n'ayant pas une bonne capacité de substituts. De plus, ils n'ont pas bon goût.

### **2.3.2 Positionnement de journaux médicaux**

L'éditeur d'une revue médicale souhaite connaître le positionnement de son journal face à ses concurrents. Une étude a donc été entreprise sur un échantillon de scientifiques qui devront classer dix journaux selon les six catégories suivantes :

- A. Préférences globales ;
- B. Préférences au niveau de la facilité de lectures ;
- C. Informations utiles à la pratique quotidienne ;
- D. Offre le plus de développement ;
- E. Le meilleur dans l'urgence ;
- F. La plus grande accessibilité.

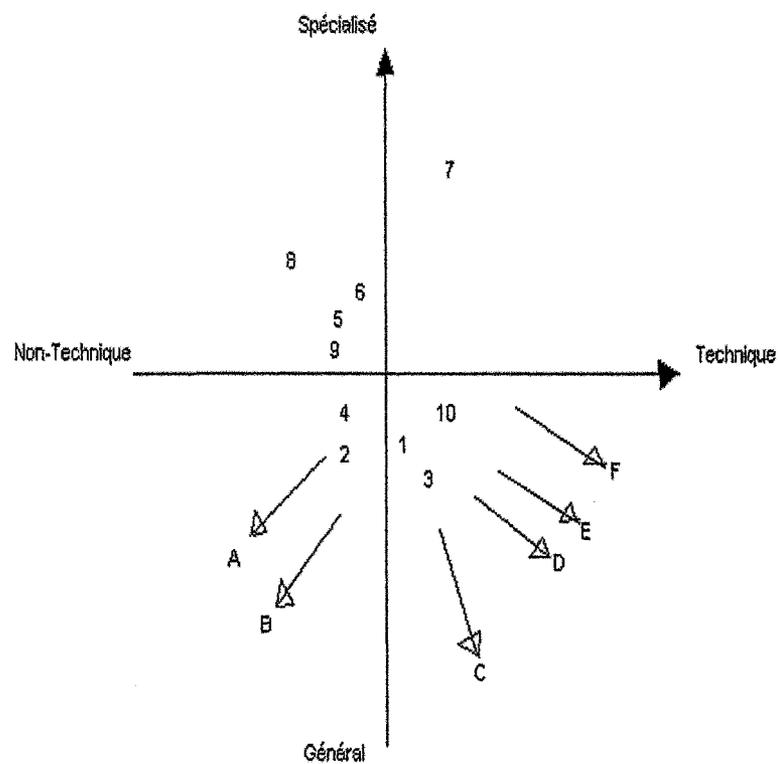
Les dix journaux sont les suivants :

1. Modern Medicine ;
2. Med. World News ;
3. JAMA ;
4. Medical Tribune ;

5. Medical economics ;
6. M.D. ;
7. Geriatrics ;
8. Human sexuality ;
9. Emergency medicine ;
10. New England Journal of Medicine.

Ainsi, pour chaque scénario, une matrice  $10 \times 10$  des similarités des données d'entrées est utilisée pour créer la carte des perceptions.

Figure 2.6 – Espace conjoint des journaux et des scénarios



On retrouve l'espace conjoint des journaux et des scénarios à la figure 2.6. En observant cet espace conjoint, les chercheurs peuvent déterminer ce que signifient les dimensions. Bien sûr, on ne connaît pas ces journaux. Il serait donc impossible de les déterminer,

mais les scientifiques de l'échantillon, eux, y sont arrivés sans problème. La dimension 1 est donc de non-technique à technique et la dimension 2, de général à spécialisé. Ainsi, le journal " New England Journal of Medicine " serait le plus technique et " Geriatrics ", le plus spécialisé.

Les analystes peuvent aussi arriver à plusieurs conclusions à l'aide des six vecteurs d'attributs. Par exemple, les journaux " Medical Tribune " et " Med World News " semblent être les meilleurs lorsque c'est urgent et pour avoir la plus grande largeur d'appel. De façon générale, les journaux préférés sont " JAMA ", " Modern Medicine " et " New England Journal of Medicine ".

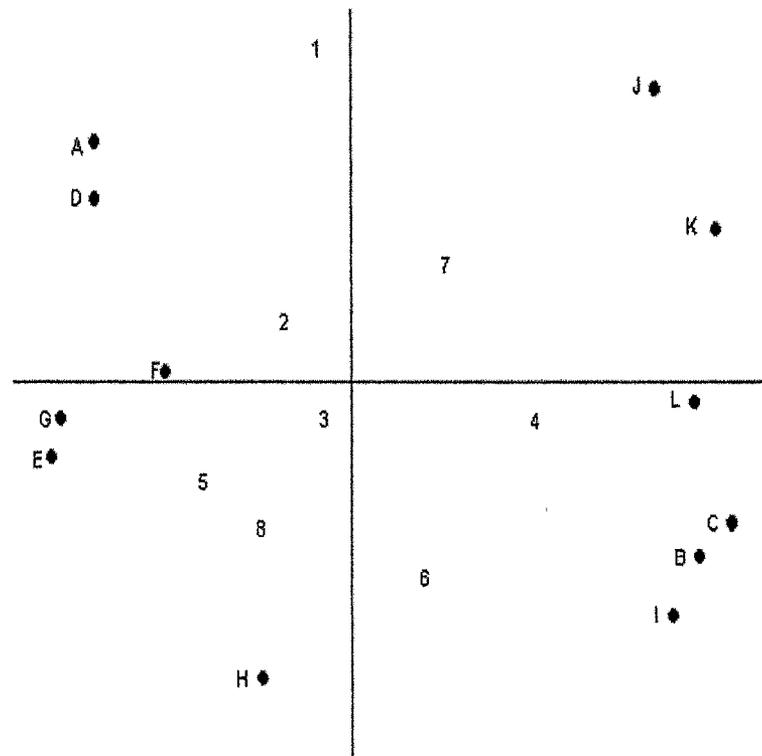
Le but de cette étude était d'avoir un espace conjoint des journaux et des scénarios pour que le journal en question puisse le présenter à ses futurs clients en leur montrant ses forces et ses positions par rapport à ses compétiteurs.

### **2.3.3 Positionnement d'un service financier**

Cette étude emploie une méthode plus indirecte. L'objectif est de positionner de nouveaux services financiers. Pour ce faire, il est demandé aux répondants, qui sont des hommes qui possèdent une maison, de faire l'évaluation habituelle des similarités et, sur un ensemble de douze occupations, choisir les cinq à qui les détenteurs sont les plus susceptibles à utiliser ces services financiers. Ils devront de plus classer les cinq occupations du plus au moins probable à utiliser ce service.

La figure 2.7 représente l'espace conjoint des services financiers et des emplois qui sont définis ci-bas.

Figure 2.7 – Espace conjoint des services financiers et des emplois



Services financiers :

1. Préparation d'impôt;
2. Évaluations des investissements;
3. Service de rapport des revenus et des dépenses;
4. Programmes financiers;
5. Fonds d'investissement;
6. Bulletin fiscal;
7. Assistant monétaire;
8. Avis spéciaux par téléphone.

Emplois :

- A. Physicien
- B. Employé dans un magasin
- C. Électricien
- D. Membre de la direction d'une grande corporation
- E. Vendeur de porte-à-porte pour une firme
- F. Musicien dans un orchestre symphonique
- G. Chimiste
- H. Professeur dans une école publique
- I. Secrétaire
- J. Serveur dans un restaurant
- K. Opérateur de machine dans une usine
- L. Policier

Il y a une séparation entre les services prestigieux et ceux non prestigieux. Par exemple, il y a comme services prestigieux les fonds d'investissements et les avis spéciaux de téléphone. À l'opposé se trouve les programmes financiers et l'assistant monétaire.

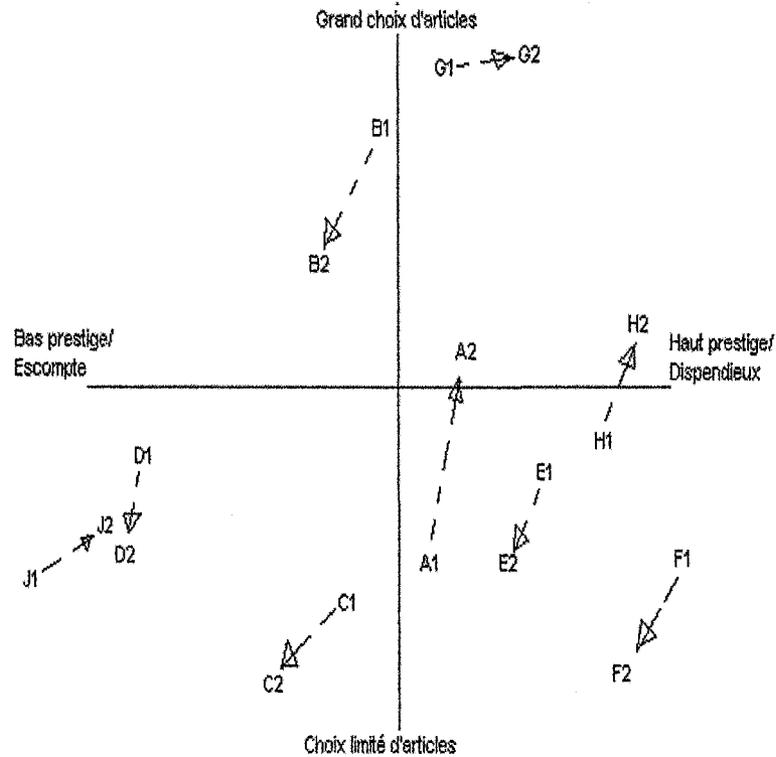
Le fait d'utiliser un positionnement indirect en présentant les occupations où les personnes sont les plus susceptibles d'utiliser certains services fournit plusieurs indications utiles pour une approche stratégique de promotion. En résumé, cette étude donne une nouvelle approche dans l'étude du positionnement.

#### **2.3.4 Positionnement d'un détaillant**

La différence entre cette étude et les précédentes est l'introduction du temps. Les chercheurs souhaitent donc regarder les changements de positionnement des détaillants selon

une période de deux ans. L'échantillon sera formé de femmes aux foyers. Les détaillants seront choisis dans une ville métropolitaine.

Figure 2.8 – Carte des perceptions des magasins



La figure 2.8 représente la configuration des neuf magasins ainsi que les neuf vecteurs pointillés qui représentent les déplacements des configuration sur deux ans.

Les vrais noms des détaillants sont cachés dans cette étude mais ceux qui ont créé cette configuration en deux dimensions les connaissaient et ont ainsi pu interpréter les dimensions. Ils en sont venus à dire que la dimension 1 est de bas prestige/escompte à haut prestige/dispendieux et que la dimension 2 va de choix limité d'articles à grand choix d'articles.

Il est maintenant possible de faire une bonne analyse à l'aide des dimensions ainsi que

de la direction et de la longueur des vecteurs. Ce qui donne donc une bonne image du positionnement du marché à travers le temps. Bien sûr, il aurait été possible d'évaluer sur une plus longue période de temps pour couvrir certains changements. De plus, le positionnement et le traçage qui bouge peuvent suggérer un nombre d'implications de stratégies.

On a donc présenté quatre exemples qui utilisent le positionnement multidimensionnel pour créer la carte des perceptions ainsi que des techniques de regroupement par grappe. De ces programmes résultent les similarités ou les différences des stimuli.

On peut donc appliquer ces techniques pour le positionnement de produits. Ce concept de positionnement s'applique à plusieurs produits comme les calculatrices, les aliments diététiques et les journaux médicaux, à plusieurs services comme les services financiers et à des entreprises comme les détaillants.

La plus grande charge de travail dans ce genre d'analyse va au chercheur. La tâche du répondant est plus simple avec ces techniques non-métriques de positionnement multidimensionnel et regroupement par grappe puisqu'elle ne consiste qu'en un classement ordonné des objets à l'étude avec des similarités ou des préférences. Pour qu'une étude de positionnement de produits soit vraiment efficace, on devrait y ajouter une étude de segmentation de marché. Cette combinaison peut fournir des instructions pour le design et la coordination des stratégies de marché des firmes.

Finalement, cette technique peut s'avérer très efficace et peut amener une analyse sous un autre angle dans les études de marché.

## 2.4 Étude des breuvages colas à l'aide de l'ajustement des attributs influents

Cet exemple sur le positionnement multidimensionnel est tiré du livre de Schiffman, Reynolds et Young [7] et du livre de Dillon et Goldstein [6]. Les stimuli à l'étude sont les 10 breuvages colas énumérés ci-bas.

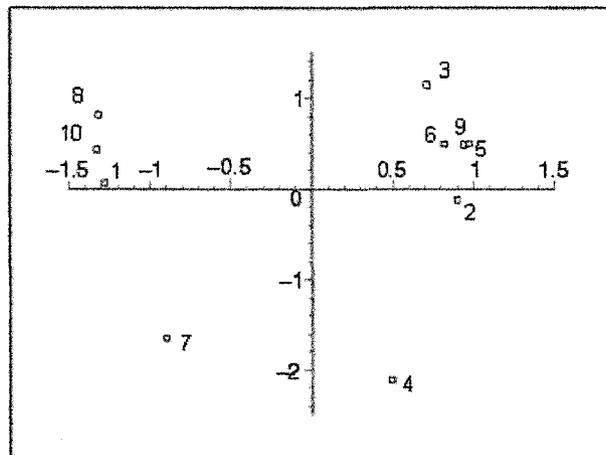
1. Pepsi diète ;
2. RC Cola ;
3. Yukon ;
4. Dr. Pepper ;
5. Shasta ;
6. Coca-Cola ;
7. Dr. Pepper diète ;
8. Tab ;
9. Pepsi-cola ;
10. Rite diète ;

L'ensemble des répondants est constitué de 5 filles et 5 garçons âgés entre 18 et 21 ans et qui sont étudiants à l'université. L'ensemble complet de ces données se retrouve en annexe. La solution en 2 dimensions provenant d'un modèle de positionnement multidimensionnel non-pondéré donne les coordonnées que l'on retrouve dans la matrice  $\mathbf{X}$ .

$$\mathbf{X} = \begin{bmatrix} -1.28 & 0.06 \\ 0.9 & -0.13 \\ 0.71 & 1.14 \\ 0.49 & -2.11 \\ 0.94 & 0.48 \\ 0.82 & 0.49 \\ -0.90 & -1.65 \\ -1.32 & 0.81 \\ 0.97 & 0.49 \\ -1.33 & 0.43 \end{bmatrix}$$

On obtient ainsi l'espace dérivé de la figure 2.9.

Figure 2.9 – Espace dérivé des breuvages colas



En analysant cet espace, on peut déduire quel attribut chaque dimension signifie.

Puisque le Pepsi diète, le Dr.Pepper diète et le Rite diète se situent tous à la gauche de ce graphique, on sera donc portés à nommer cette dimension diète versus non-diète. Donc, plus un produit est situé à droite dans l'espace dérivé, moins il est diète. Pour la deuxième dimension, on utilisera le fait que Dr.Pepper et Dr.Pepper diète sont tout en bas du graphique pour la déterminer. Donc, plus un point se retrouve dans le bas du

graphique, plus la boisson qu'il représente aura une saveur de cerise. Plus il sera vers le haut, plus la boisson aura une saveur régulière.

Chacun des 10 répondants a évalué les dix sortes de boissons colas sur 13 attributs qui sont énoncés ci-bas.

1. Bon versus mauvais ;
2. fort versus délicat ;
3. sucré versus non-sucré ;
4. amer versus non-amer ;
5. aigre versus non-aigre ;
6. fruité versus non-fruité ;
7. épicé versus non-épicé ;
8. enrobe la bouche versus n'enrobe pas la bouche ;
9. âpre versus non-âpre ;
10. fait plisser les lèvres versus ne fait pas plisser les lèvres ;
11. frais versus vieux ;
12. chimique versus non-chimique ;
13. complexe versus non-complexe.

L'évaluation fut faite sur un échelle de 0 à 100 où 0 représente l'adjectif de gauche et 100, celui de droite. Les données de chaque répondant pour chaque produit et chaque attribut sont données dans le livre de Schiffman, Young et Reynolds [7]. À l'aide de ces valeurs, on illustrera l'ajustement des attributs influents. Dans cet exemple précis, on travaillera avec l'attribut chimique. Ainsi, dans le tableau 2.1, on retrouve les données de chaque répondant pour chaque sorte de cola pour l'attribut chimique.

Tableau 2.1 – Données de chaque répondant pour chaque sorte de cola pour l'attribut chimique

	Breuvages colas									
	1	2	3	4	5	6	7	8	9	10
Répondant 1	8	93	84	38	39	1	42	90	41	7
Répondant 2	82	17	14	83	0	84	92	21	74	37
Répondant 3	49	19	4	94	12	39	41	25	65	30
Répondant 4	22	76	74	99	19	88	76	22	91	21
Répondant 5	8	45	17	91	13	87	7	1	79	0
Répondant 6	39	52	41	88	75	56	36	44	51	21
Répondant 7	85	99	94	99	89	18	14	75	99	27
Répondant 8	38	62	57	43	20	6	18	20	56	17
Répondant 9	42	81	19	35	11	11	24	13	47	17
Répondant 10	74	35	10	18	28	12	25	7	6	35

On trouvera son vecteur attribut en suivant les 4 étapes développées au chapitre 1 dans la section 1.3.8.

Étape 1 :

Trouvons la moyenne de chacun des 10 breuvages pour l'attribut chimique.

Deux exemples de calculs sont présentés ci-bas.

**Exemple 16.** *Exemples de calculs pour Pepsi diète et RC Cola.*

*Pepsi diète :*

$$\frac{1}{10}(8 + 82 + 49 + 22 + 8 + 39 + 85 + 38 + 42 + 78) = 44.7 \approx 45$$

*RC Cola :*

$$\frac{1}{10}(93 + 17 + 19 + 76 + 45 + 52 + 99 + 62 + 81 + 35) = 57.9 \approx 58$$

On obtient ainsi toutes les valeurs moyennes qui sont représentées dans la matrice  $y$  suivante.

$$y' = \begin{matrix} & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ \begin{pmatrix} 45 & 58 & 41 & 69 & 31 & 40 & 35 & 32 & 61 & 21 \end{pmatrix} \end{matrix}$$

Étape 2 :

On réutilisera les coordonnées données dans le vecteur  $\mathbf{1}$  augmenté de la matrice  $\mathbf{X}$  qui donnent la solution en 2 dimensions provenant d'un modèle de positionnement multidimensionnel. Rappelons que cette matrice, notée  $\mathbf{X}_a$  est donnée par

$$\mathbf{X}_a = \begin{bmatrix} 1 & -1.28 & 0.06 \\ 1 & 0.9 & -0.13 \\ 1 & 0.71 & 1.14 \\ 1 & 0.49 & -2.11 \\ 1 & 0.94 & 0.48 \\ 1 & 0.82 & 0.49 \\ 1 & -0.9 & -1.65 \\ 1 & -1.32 & 0.81 \\ 1 & 0.97 & 0.49 \\ 1 & -1.33 & 0.43 \end{bmatrix}$$

où l'espace dérivé est représenté à la figure 2.9. On note les colonnes 2 et 3 de la matrice  $\mathbf{X}_a$ ,  $\mathbf{X}_{a1}$  et  $\mathbf{X}_{a2}$ . On doit maintenant centrer ces deux colonnes. Puisque  $\mathbf{X}_{a1}$  l'est déjà, on n'aura qu'à centrer  $\mathbf{X}_{a2}$ . Une fois que c'est fait, on obtient la nouvelle matrice  $\mathbf{X}_c$  telle que

$$\mathbf{X}_c = \begin{bmatrix} 1 & -1.28 & 0.059 \\ 1 & 0.9 & -0.131 \\ 1 & 0.71 & 1.139 \\ 1 & 0.49 & -2.111 \\ 1 & 0.94 & 0.479 \\ 1 & 0.82 & 0.489 \\ 1 & -0.9 & -1.651 \\ 1 & -1.32 & 0.809 \\ 1 & 0.97 & 0.489 \\ 1 & -1.33 & 0.429 \end{bmatrix}$$

On doit aussi centrer  $y$  que l'on notera  $y_c$ .

$$y'_c = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 1.7 & 14.7 & -2.3 & 25.7 & -12.3 & -3.3 & -8.3 & -11.3 & 17.7 & -22.3 \end{pmatrix}$$

Calculons maintenant le vecteur  $\beta$  pour trouver nos coordonnées avec

$$\beta = (\mathbf{X}_c' \mathbf{X}_c)^{-1} \mathbf{X}_c' y_c.$$

Ainsi,

$$\beta = \begin{pmatrix} 0 \\ 8.20 \\ -6.84 \end{pmatrix}.$$

On remarque que  $\beta_0 = 0$  puisque  $\mathbf{X}_c$  et  $y_c$  sont centrés à l'origine.

Étape 3 :

Calculons maintenant la coefficient de corrélation multiple. Pour ce faire, commençons par calculer  $\hat{y}_c$ .

$$\hat{\mathbf{y}}'_c = (\mathbf{X}_c \beta)' =$$

$$\left( -10.90 \quad 8.28 \quad -1.97 \quad 18.46 \quad 4.43 \quad 3.38 \quad 3.91 \quad -16.36 \quad 4.61 \quad -13.85 \right)$$

Ainsi,

$$R^2 = 1 - \frac{(\mathbf{y}_c - \hat{\mathbf{y}}_c)'(\mathbf{y}_c - \hat{\mathbf{y}}_c)}{\mathbf{y}'_c \mathbf{y}_c}$$

$$= 0.5157$$

et  $R=0.7181$ . Cette valeur de  $R$  n'est que modérément élevée. Le fait que  $R^2 = 0.5157$  indique qu'environ 52% des variations dans l'évaluation des attributs est expliqué par la position des breuvages colas dans l'espace dérivé.

Étape 4 :

Pour représenter graphiquement le vecteur attribut, on peut soit simplement tracer une droite entre l'origine et le point  $\beta^* = (\beta_1, \beta_2)$  avec la tête de la flèche au point  $\beta^*$  ou utiliser les angles  $\theta_1$  et  $\theta_2$ . Pour appliquer la deuxième méthode, calculons  $\cos \theta_1$  et  $\cos \theta_2$ .

Tout d'abord, calculons la norme de  $\beta$ .

$$\|\beta\| = \left[ \sum_{i=1}^r \beta_i^2 \right]^{1/2} = \left[ \sum_{i=1}^2 \beta_i^2 \right]^{1/2} = 10.68$$

Ensuite, calculons  $\cos \theta_1$  et  $\cos \theta_2$ .

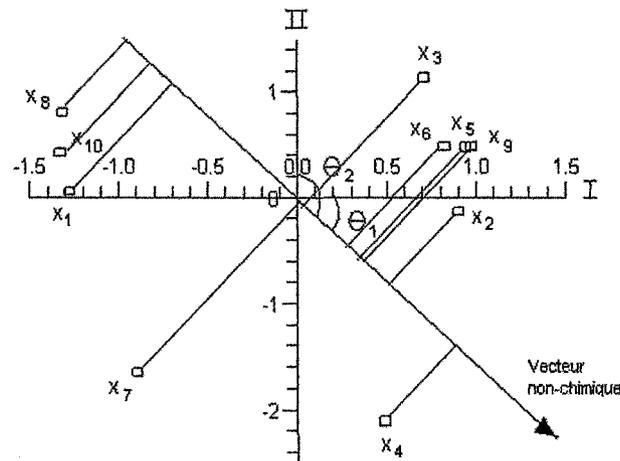
$$\cos \theta_1 = \frac{\beta_1}{\|\beta\|} = \frac{8.20}{10.68} = 0.7679$$

$$\cos \theta_2 = \frac{\beta_2}{\|\beta\|} = \frac{-6.84}{10.68} = -0.6405$$

Ainsi,  $\theta_1 = 39.8^\circ$  et  $\theta_2 = 129.8^\circ$ .

La figure 2.10 donne l'espace dérivé des stimuli avec le vecteur attribut chimique. On y retrouve aussi les projections des breuvages sur ce vecteur ainsi que les représentations de  $\theta_1$  et  $\theta_2$ . Dans cet étude, plus les données sont élevées, moins le produit est chimique. Ainsi, plus on avance dans le sens du vecteur, moins les boissons sont chimiques. En étudiant la figure 2.10, on remarque que les répondants considèrent les boissons diètes chimiques.

Figure 2.10 – Espace dérivé des breuvages colas avec le vecteur attribut chimique



On n'est par contre pas obligé de représenter les 13 attributs. Supposons qu'ici, on souhaite n'en choisir que 6. Pour les choisir, on doit tout d'abord calculer tous les coefficients de corrélation entre chacun des attributs à l'aide de la formule

$$r = \frac{Cov(\mathbf{X}_i, \mathbf{X}_j)}{\sqrt{Var(\mathbf{X}_i)}\sqrt{Var(\mathbf{X}_j)}} = \frac{(\mathbf{X}_i - \bar{\mathbf{X}}_i)'(\mathbf{X}_j - \bar{\mathbf{X}}_j)}{((\mathbf{X}_i - \bar{\mathbf{X}}_i)'(\mathbf{X}_i - \bar{\mathbf{X}}_i))^{1/2}((\mathbf{X}_j - \bar{\mathbf{X}}_j)'(\mathbf{X}_j - \bar{\mathbf{X}}_j))^{1/2}}$$

Lorsque l'on a un haut niveau de corrélation entre chaque membre d'un petit groupe, on peut choisir d'en afficher un seul qui représentera les autres. Le tableau 2.2 contient tous les coefficients de corrélation de cet exemple où 1 à 13 représentent les 13 attributs

énoncés au début de cette section.

Tableau 2.2 – Coefficients de corrélation entre les attributs

	1	2	3	4	5	6	7	8	9	10	11	12	13
1													
2	-0.38												
3	0.94	-0.18											
4	-0.97	0.33	-0.96										
5	-0.93	0.51	-0.84	0.93									
6	0.56	0.32	0.67	-0.56	-0.30								
7	-0.22	0.80	-0.09	0.23	0.37	0.52							
8	-0.85	0.63	-0.77	0.87	0.85	-0.30	0.59						
9	-0.88	0.63	-0.83	0.87	0.84	-0.29	0.58	0.94					
10	-0.89	0.54	-0.81	0.84	0.88	-0.20	0.55	0.85	0.93				
11	0.84	-0.21	0.80	-0.84	-0.72	0.75	0.12	-0.69	-0.65	-0.56			
12	-0.94	0.41	-0.91	0.92	0.82	-0.63	0.19	0.86	0.90	0.82	-0.90		
13	-0.45	0.48	-0.50	0.54	0.52	0.10	0.62	0.66	0.74	0.67	-0.31	0.55	

Bon, sucré, amer et âpre sont éliminés car ils ont un haut niveau de corrélation avec les attributs aigre, frais et "fait plisser les lèvres". L'attribut "enrobe la bouche" est éliminé puisqu'il est hautement corrélé avec les attributs "fait plisser les lèvres" et chimique. Complexe et fort, quand à eux, sont éliminés car les deux montrent de petites déviations standards indiquant qu'ils sont de mauvais discriminateurs. Même si épice montre aussi une petite déviation standard, il est retenu comme étant un descripteur de saveur possible.

Les 6 attributs retenus sont donc

- aigre;
- fruité;
- épice;
- fait plisser les lèvres;
- frais;

-chimique.

# CHAPITRE 3

## Comparaison de programmes traitant de positionnement multidimensionnel

L'importance de l'utilisation de programmes informatiques en analyse de données est très grande puisqu'ils nous permettent d'analyser un plus grand nombre de données plus rapidement et plus aisément. Parmi les plus courants, on compte ALSCAL, INDSCAL, KYST, MINISSA, MULTISCALE et POLYCON. Dans la section 3.1, on a choisi de détailler l'utilisation du programme ALSCAL. C'est celui qui est le plus complet, le plus polyvalent et aussi le plus accessible. La section 3.2 est consacrée à la comparaison des six programmes retenus sur la base d'un certain nombre de critères.

### 3.1 ALSCAL

L'algorithme ALSCAL est extrêmement flexible et peut fournir des échelonnements métriques et non-métriques avec ou sans des modèles de différences individuelles (pondérés). Le mot ALSCAL provient de "Alternating Least-Squares SCAling" et ses fondements ont

paru dans [8]. Il est disponible sur SAS, SPSS et IMSL. Dans le cadre de ce travail, on a testé cette procédure sur le logiciel SPSS.

ALSCAL dans SPSS permet de construire des structures géométriques multidimensionnelles, le plus souvent en 2 ou 3 dimensions, s'ajustant au mieux aux similarités ou dissimilarités observées ou calculées. Il supporte jusqu'à 100 objets et 6 dimensions. Afin de fonctionner, le nombre d'objets doit être plus grand que le nombre de dimensions sinon on se retrouve dans un espace avec une infinité de solution. Il y a une contrainte supplémentaire sous ALSCAL, le nombre d'objets doit être quatre fois plus élevé que le nombre de dimensions plus un.

Pour ce qui est des données d'entrées, l'algorithme ALSCAL est mieux adapté avec des matrices de dissimilarités et de distances qu'avec des matrices de similarités. Pour cette raison, SPSS requiert des matrices de distances et de dissimilarités et non de similarités. Si l'on dispose de matrices de similarités métriques, il existe une conversion dans SPSS pour les transformer en matrices de distances. Par contre, s'il le désire, le chercheur devra les convertir lui-même avant d'entreprendre l'analyse de positionnement multidimensionnel avec SPSS. Pour ce faire, il devra utiliser les formules de la section 1.2.

Pour faire une étude appropriée, il est très important de sélectionner les bonnes conditions parmi celles que le programme offre. Dans ce but, l'étude complète des choix spécifiques qu'on peut faire en utilisant l'algorithme ALSCAL dans SPSS se fera dans cette section. Notons que pour ce travail, la version utilisée est 15.0 SPSS pour Windows. Pour commencer, dans SPSS, il faut sélectionner "Analyze", "Scale" et ensuite "Multidimensional Scaling(ALSCAL)". Il s'ouvrira une boîte de dialogue dans laquelle on fait entrer les objets dans l'espace réservé aux variables. Il est aussi permis d'entrer un groupe de variables dans l'espace "Individual matrices for". Ainsi, les données de sorties sont produites pour

chaque valeur discrète du groupe de variables.

Dans la fenêtre de positionnement multidimensionnel se trouve un encadré nommé "Distances". On peut soit choisir le cas où les données sont des distances ou soit le cas où l'on crée les distances selon nos données.

Si les données sont déjà des distances, on choisit le premier cas. En cliquant sur l'option "Shape", on sélectionne le genre de matrice qu'on a. Il existe 3 cas, la matrice carrée symétrique, carrée asymétrique ou rectangulaire.

La matrice carrée symétrique est celle choisie par défaut dans SPSS. Ces lignes et ces colonnes représentent les mêmes objets. Il n'est pas nécessaire d'entrer les données de la partie triangulaire supérieure. Il faut par contre entrer les valeurs 0 sur la diagonale.

Pour la matrice carrée asymétrique, les lignes et les colonnes représentent toujours les mêmes objets. Par contre, les valeurs des parties triangulaires supérieure et inférieure ne sont pas nécessairement les mêmes. Par exemple, la distance entre X et Y n'est pas la même que celle entre Y et X. Dans ce cas, on doit entrer toutes les valeurs de la matrice.

La forme rectangulaire est choisie lorsque les lignes et les colonnes ne représentent pas les mêmes entités. Les objets sont représentés par les colonnes, mais les lignes peuvent être, par exemple, les sujets ou des attributs. Par défaut, il est supposé que l'on a une seule matrice. Ainsi, on peut laisser la case "Number of rows" vide. Par contre, si on a plusieurs matrices représentant chacun des sujets, il faut indiquer le nombre de lignes que possèdent une matrice. En d'autres termes, si une étude comporte 8 répondants, 12 objets et 5 attributs, il faut entrer 8 matrices (une par répondant) de taille  $5 \times 12$  une à la suite de l'autre. On entrerait ainsi dans la case "Number of rows" le nombre 5 qui

signifie que chaque matrice possède 5 lignes. Pour le programme ALSCAL, le nombre minimal de lignes (d'attributs) qui peut être traité est 4.

À la différence des matrices carrées, la carte des perceptions montrera de façon conjointe les points des colonnes et des lignes.

Si l'on souhaite créer nos propres mesures de distances, il faut aller dans la section "Create distances from data". En effet, puisque le positionnement multidimensionnel utilise des données de dissimilarités pour créer les solutions, il faut modifier celles qui ne sont pas des dissimilarités ou des distances mais plutôt des valeurs des variables mesurées.

Une fois que cette option a été sélectionnée, il faut aller choisir la mesure de distances ou de dissimilarités désirée dans la sous-section "Measure". Il y a 3 types de mesures : intervalle, de comptage ou binaire. Le tableau 3.1 indique quelle distance chaque type de mesure peut utiliser.

Tableau 3.1 – Tableau des distances utilisées dans les différents types de mesures

	Intervalle	De comptage	Binaire
Distance euclidienne	X		X
Distance euclidienne carrée	X		X
Chebychev	X		
Block	X		
Minkowski	X		
Customized	X		
Mesure du Chi-carrée		X	
Mesure du Phi-carrée		X	
"Size difference"			X
"Pattern difference"			X
Variance			X
Lance et William			X

Il est à noter que pour la mesure de distance de Minkowski, nous pouvons choisir une puissance allant de 1 à 4 et 2 est le choix par défaut. Pour la mesure de distance "Customized", on peut faire la même chose que Minkowski et choisir la racine allant aussi de 1 à 4 et 2 est le choix par défaut. Dans le cas des données binaires, nous pouvons choisir lesquels signifient "absent" et "présent". Une version détaillée de ces mesures se trouvent en annexe.

On peut choisir dans la sous-section "Create distance matrix", l'unité de l'analyse. Les alternatives sont entre variables (colonnes) ou entre cas (lignes). Dans la sous-section "Transform Values", il est possible de standardiser nos valeurs lorsqu'elles ne sont pas toutes mesurées sous la même échelle. Si on ne souhaite pas les standardiser, il faut sélectionner "None" dans la liste de standardisation. Sinon, il faut y sélectionner une méthode dont les choix sont énumérés ci-bas.

- Z score :

Les valeurs sont standardisées au Z score avec une moyenne de 0 et un écart-type de 1. Par exemple, si l'observation  $x$  provient d'un échantillon de moyenne  $\bar{x}$  et d'écart-type de  $s$ , alors  $z = \frac{x - \bar{x}}{s}$ .

- Range -1 to 1 :

(Ranger de -1 à 1)

Chaque valeur à être standardisée est divisée par l'étendue des données.

Exemple : -2 à 5, on divise par 7.

- Range 0 to 1 :

(Ranger de 0 à 1)

La procédure soustrait la valeur minimale de chaque item à être standardisé et divise par

l'étendue des données.

Exemple : 5 à 12 devient de 0 à 7 qu'on divise par 7.

- Maximum magnitude of 1 :

(Grandeur maximale de 1)

Cette procédure divise chaque valeur à être standardisée par celle qui est maximale.

- Mean of 1 :

(Moyenne de 1)

La procédure divise chaque valeur à être standardisée par la moyenne de ces valeurs.

- Standard deviation of 1 :

(Écart-type de 1)

La procédure divise chaque variable ou cellule à être standardisée par l'écart-type des valeurs.

De plus, lorsque l'option "Create distances from data" est sélectionnée, une matrice carrée symétrique est créée. Si l'option "By variable" est choisie, les objets de cette matrice seront donc les variables des données initiales. Si le choix s'arrête sur l'option "By case", les objets seront maintenant les éléments qui définissent les lignes de la matrice.

La section "Model" permet de choisir le type de mesures de notre matrice ainsi que les conditions qui s'y appliquent. On sélectionne aussi si on est dans le cas métrique ou non-métrique, pondéré ou non-pondéré et classique ou répété (respectivement une ou plusieurs matrices en entrée).

Dans la sous-section "Scaling model", il faut sélectionner si on travaille avec un cas de positionnement multidimensionnel pondéré ou non. Si c'est le cas, il faut choisir "Individual differences Euclidean distance". Dans le cas contraire, on prend "Euclidean distance". Il est évident que le premier cas requiert au moins deux matrices. Si l'on se trouve dans le cas de positionnement multidimensionnel non-pondéré et qu'on a deux matrices ou plus, nous rentrons automatiquement dans un cas de positionnement multidimensionnel répliqué. Nous n'avons donc pas à le sélectionner nulle part. Il est à noter que, par défaut, "Euclidean distance" est sélectionné et que, dans le cas de "Individual differences Euclidean distance", il est permis d'avoir des poids négatifs. Il suffit de cocher la case "Allow negative subject weight".

Dans la sous-section "Level of measurement", on identifie le type de mesure des données que l'on retrouve à l'intérieur de la matrice. Il y est offert 3 types de mesures : ordinales, intervalles et ratios.

Le choix par défaut dans cette sous-section est la mesure ordinale. Dans ce cas, les nombres assignés à chaque paire d'objets représentent un ordre de rang. Ainsi, si on possède  $n$  objets, on aura une matrice  $n \times n$ . Les paires seront donc classées de 1 à  $\frac{1}{2}n(n-1)$  puisqu'on utilise seulement les données sous la diagonale dans une matrice symétrique. Dans le cas des mesures ordinales, l'analyse est non-métrique. De plus, le traitement des données se fait à l'aide de l'algorithme de violation du groupe adjacent aussi nommé transformation monotone des moindres carrées de Kruskal. Si la case "Untie tied observations" est cochée, ALSCAL traite les données comme étant continue. Si, au contraire, elle n'est pas cochée, il traite les données comme des dissimilarités. Par défaut, elle n'est pas cochée.

Pour les mesures intervalles, les nombres assignés à chaque paire de stimuli représentent

aussi un ordre mais cette fois-ci, les différences entre les mesures qui sont égales représentent des intervalles équivalents. Ainsi, les différences entre les paires d'objets peuvent être comparées de façon significative. Un exemple de mesure d'intervalle fréquemment utilisée en positionnement multidimensionnel est le cas où l'on mesure la température en degré Celsius. Lorsque l'on utilise les mesures intervalles, l'analyse est métrique.

Pour les mesures ratios, les nombres assignés à chaque paire d'objets possèdent les mêmes caractéristiques que les mesures intervalles mais, en plus, la division et la multiplication des données sont significatives dans le cas des ratios. Prenons par exemple une longueur en mètres d'une maison en supposant que 0 signifie le coin gauche. Si la porte est située à 4 mètres de ce coin et que la maison mesure 8 mètres au total, on peut dire que la porte se situe au milieu ( $1/2$ ) de la maison. Dans le cas présent, l'analyse est aussi métrique.

Lorsque, dans la sous-section "Scaling Model" de la fenêtre "Model", on choisit "Matrix", c'est que les données d'une matrice sont comparées uniquement avec les données de cette matrice en particulier. Ceci s'applique autant dans le cas où l'on n'a qu'une seule matrice que dans le cas où on en a plusieurs (une par répondant). Cette condition est celle sélectionnée par défaut.

Il faut sélectionner "Row" lorsque les éléments de la matrice de données ne peuvent pas être comparés librement avec tous les autres éléments de la matrice, mais seulement avec les éléments de la même ligne. Ce cas utilise donc seulement les matrices carrées asymétriques et rectangulaires. Par exemple, si on a  $n$  objets, chaque sujet commence l'évaluation en se basant sur le premier. Il le compare avec tous les autres en notant 1 celui qui est le plus similaire à ce premier objet, 2 au deuxième jusqu'à  $n-1$  pour celui qui est le plus différent. Il recommence le même processus jusqu'au  $n^{\text{ième}}$  objet. Ainsi, on crée une matrice carrée asymétrique. L'exemple suivant le démontre très bien.



*les lignes ensembles.*

Si on travaille toujours avec nos  $n$  objets mais que l'on possède aussi  $m$  attributs, on a bien évidemment une matrice rectangulaire. Puisque chaque attribut représente une ligne, il est donc impossible de comparer les éléments de différentes lignes puisqu'ils n'ont pas la même signification. L'exemple 18 expose ce cas de matrice rectangulaire.

**Exemple 18.** *Reprenons les 5 objets de l'exemple 17. Dans le cas des matrices rectangulaires, on étudie les 4 critères suivants sur un échelle de 1 à 10 où 1 représente l'élément de gauche.*

- A. Histoire réelle versus histoire fantastique ;*
- B. Pas original versus original ;*
- C. Décors simples versus décors complexes ;*
- D. Palette d'acteurs peu connus versus palette d'acteurs connus ;*

*Selon l'évaluation d'un sujet, on obtient la matrice suivante :*

$$\begin{array}{c} \phantom{A} \\ \phantom{B} \\ \phantom{C} \\ \phantom{D} \end{array} \begin{array}{ccccc} F1 & F2 & F3 & F4 & F5 \\ \left( \begin{array}{ccccc} 8 & 7 & 10 & 2 & 8 \\ 8 & 8 & 9 & 6 & 10 \\ 10 & 7 & 9 & 7 & 9 \\ 10 & 8 & 9 & 6 & 7 \end{array} \right) \end{array}$$

*Puisque chaque ligne représente un attribut différent, leurs éléments ne peuvent pas être comparés ensemble.*

La condition "Unconditional" est aussi utilisée uniquement pour les matrices carrées asymétriques et rectangulaires. On peut comparer tous les éléments d'une matrice ensemble

et même entre d'autres matrices.

Il faut sélectionner dans la sous-section "Dimensions" de la section "Model" le nombre maximal et minimal de dimensions que l'on souhaite avoir comme solution. Le programme créera une solution à chaque dimension. Si on souhaite qu'une seule dimension soit affichée, il faut choisir le même nombre de dimensions maximale et minimale. Par défaut, le positionnement multidimensionnel dans SPSS sélectionne une solution en 2 dimensions. On peut par contre choisir des dimensions de 1 à 6 puisque, parfois, deux dimensions ne suffisent pas pour bien représenter les objets. Il est ainsi possible d'explorer les mesures de stress pour chaque dimension.

Il faut être prudent puisqu'en augmentant la dimensionnalité, on augmente aussi la difficulté d'interprétation. La majorité des analyses de positionnement multidimensionnel se font en 2 ou 3 dimensions. Il est à noter que, dans le cas des modèles pondérés, la dimension minimale devrait être 2 et que le minimum de 1 dimension est permis uniquement lorsque la distance euclidienne est sélectionnée dans la sous-section "Scaling Model".

Pour ce qui est des données de sorties, il faut aller dans la section "Options". On y retrouve la sous-section "Group plots" qui génère la carte des perceptions des objets pour le positionnement multidimensionnel et un nuage de points de la concordance linéaire entre les données et le modèle que l'on nomme aussi diagramme de Shepard. De plus, lorsque c'est possible, les nuage de points des concordances non-linéaires et des données de transformations sont générés. Pour les modèles pondérés, les matrices de poids sont aussi affichées.

L'option "Group plots" génère tout d'abord la carte du positionnement multidimensionnel sur les dimensions de la solution. La carte des perceptions étant créée, il est bon de

savoir que la dimension 1 est représentée par l'axe des x et la dimension 2, par l'axe des y dans une solution à 2 dimensions. Pour identifier la signification des dimensions, ALSCAL ne possède pas de procédure objective. Il faut donc que le chercheur ou le spécialiste analyse la position des objets pour déduire la signification des dimensions. Il est aussi propice d'étudier les grappes d'objets qui indiquent un groupe d'éléments similaires. Pour l'étudier de façon plus approfondie, il est aussi possible de refaire un analyse de positionnement multidimensionnel mais uniquement avec les éléments de cette grappe en particulier.

Lorsque l'on sélectionne l'option "Group plots", ALSCAL affiche aussi un nuage de points des concordances linéaires. Ce graphique porte aussi le nom de diagramme de Shepard. Tel que mentionné à la section 1.3.4, ce diagramme affiche les distances sur l'ordonnée et les disparités sur l'abscisse. Rappelons que les disparités sont des distances dont l'ordre croissant suit celui du classement des dissimilarités. Dans un modèle parfait, les distances et les disparités partagent le même point. Ainsi, plus le nuage de points tend à former une ligne de  $45^\circ$  par rapport aux axes, meilleure est la concordance du modèle de positionnement multidimensionnel avec les données.

Quand le type de mesure choisi est ordinal, le programme génère le graphique des concordances non-linéaires. On y retrouve les observations sur l'abscisse et les distances sur l'ordonnée. Les observations sont rangées en ordre croissant. Un modèle possédant une bonne concordance est un modèle qui suppose l'homoscédasticité avec des points très près de la ligne pour les petites valeurs autant que pour celles élevées. Le programme peut aussi générer le graphique des transformations lorsque l'on est toujours dans le cas où le type de mesure est ordinal. Il montre sur l'axe des x, les observations et sur l'axe des y, les disparités qui sont en d'autres termes, les distances après transformations monotones. Par définition de monotonie, plus on passe au crible les observations (de gauche à droite),

plus les valeurs des disparités croissent. La ligne non-linéaire formée par le graphique de la transformation est une ligne de régression non-linéaire pour les données ordinales. Plus cette ligne de régression est régulière plutôt qu'en escaliers, plus les données ordinales ont une ressemblance métrique et moins il y a de différence dans les données de sorties du positionnement multidimensionnel entre les spécifications de type ordinal et intervalle.

Il est utile de choisir "Individual subject plots" seulement si le type de mesure sélectionné est ordinal et si la condition sur le modèle est "Matrix". Ainsi, cette option génère des graphiques différents pour les données de chaque sujet.

Les données des matrices d'entrées et les données échelonnées seront affichées si la case "Data Matrix" est cochée.

Pour l'affichage des données du modèle choisi, des données de sorties et des options algorithmiques sélectionnées, il faut cocher "Model and options summary".

Dans ALSCAL, deux types de mesure de stress sont utilisées. Le premier est SSTRESS que l'on retrouve à la formule 1.3 de la sous-section 1.3.4. Il est utilisé comme critère d'arrêt de l'algorithme non-métrique. Le deuxième type de mesure est le stress de Kruskal qui se trouve à la formule 1.1 à la sous-section 1.1.2. Celui-ci est calculé avec les valeurs de la configuration finale.

Dans la sous-section "Criteria", il faut indiquer le critère de convergence du SSTRESS, sa valeur minimum ainsi que le nombre maximal d'itérations d'après lequel l'algorithme s'arrête. Par défaut, le critère de convergence est à 0.001, la valeur minimum du SSTRESS est de 0.005 et le nombre d'itération par défaut est de 30. Il est à noter que si l'on indique "0" comme nombre d'itération, automatiquement, SPSS en fera 30. On peut aussi fixer

le nombre d'itérations aussi grand qu'il en sera nécessaire.

Dans le bas de la fenêtre "Options", il est possible d'identifier sous quelle valeur le programme doit traiter les distances comme étant manquante. Par défaut, cette valeur est 0.

Dans les données de sorties affichées se trouve toujours la corrélation carrée des distances entrées avec les distances calculées à partir des données provenant de l'analyse du positionnement multidimensionnel. Cette corrélation carrée est notée RSQ. Elle reflète la proportion des variances des données de distances entrées expliquée par les données positionnées et vice-versa. La valeur RSQ est affichée pour chaque sujet dans les modèles pondérés.

Dans le cas du positionnement multidimensionnel pondéré, on obtient en sortie l'indice d'influence pour chaque sujet et l'étendue des poids des sujets. L'indice d'influence est un indice utilisé pour montrer qui sont les sujets influents qui affectent l'analyse. En effet, les individus peuvent attacher une importance différente à certaines dimensions . Un indice de 0 signifie que le sujet pondère chaque dimension comme la moyenne et un indice de 1, que le sujet a donné toute son attention à une seule dimension et aucune aux autres. Dans le logiciel SPSS, ALSCAL produit en sortie les étendues des poids des sujets ainsi que le graphique qui les représente qui possède une seule dimension. Les étendues des poids des sujets sont représentés par les angles que forment les droites entre l'origine et les points des répondants dans le graphique représentant les poids des sujets. Ainsi, plus l'angle est près de  $45^\circ$ , plus l'étendue des poids des sujets est près de 0. Plus l'angle est près de  $0^\circ$ , plus l'étendue des poids des sujets s'éloigne de 0 dans le sens positif et plus l'angle est près de  $90^\circ$ , plus il s'éloigne dans le sens négatif. En d'autres termes, dans le graphique de l'étendue des poids des sujets, les sujets ayant donné une petite

importance aux dimensions apparaîtront au milieu du graphique, c'est-à-dire près de l'origine. Les répondants qui donnent une plus haute importance à la première dimension se retrouveront dans le côté positif du graphique et ceux qui donnent plus d'importance à la dimension 2 seront représentés dans la partie négative. Plus la différence d'importance entre les dimensions est grande, plus le sujet se retrouve à l'extrémité du graphique.

### **3.2 Programmes spécialisés dans tous les types de problèmes de positionnement multidimensionnel**

Dans cette section, 6 programmes informatiques spécialisés dans les problèmes de positionnement multidimensionnel sont comparés. On les retrouve dans SPSS, dans SAS et dans d'autres logiciels. Ces 6 programmes sont :

- A : ALSCAL ;
- I : INDSCAL ;
- K : KYST ;
- MI : MINISSA ;
- MU : MULTISCALE ;
- P : POLYCON.

Ils ont été choisis selon les critères suivants :

- une grande accessibilité d'utilisation ;
- une grande couverture de techniques d'analyse de positionnement multidimensionnel comme une analyse non-métrique d'une seule matrice ou une analyse d'un modèle pondéré possédant plusieurs matrices de données ;

- une utilisation de différentes façons de calculer les configurations recherchées.

Tableau 3.2 – Tableau 1 des caractéristiques des 6 programmes

	A	I	K	MI	MU	P
⇒ Types de matrices :						
carrées symétriques	X	X	X	X	X	X
carrées asymétriques	X		X		X	X
rectangulaires	X		X			X
⇒ Nombre de matrices d'entrées :						
une seule matrice	X		X	X	X	X
plus d'une matrice	X	X	X		X	X
⇒ Types de modèles						
métriques	X	X	X		X	X
non-métriques	X		X	X		X
pondérés	X	X			X	
non-pondérés	X		X	X	X	X
⇒ Types de mesures						
ordinales	X		X	X		X
intervalles	X	X	X		X	X
ratios	X	X	X		X	X
⇒ Conditions sur les matrices :						
matrice	X	X	X	X	X	X
ligne	X		X			X
pas de condition	X		X		X	X
⇒ Types de données						
similarités	X	X	X	X	X	X
dissimilarités	X	X	X	X	X	X

Il est à noter que dans la sous-section "Conditions sur les matrices" du tableau 3.2, ce sont les mêmes conditions que celles définies à la section 3.1. Lorsqu'en entrée on a une matrice pour chaque répondant et que l'on est dans la condition "matrice", les éléments d'une matrice en particulier ne peuvent être comparés qu'avec ceux de cette même matrice. Dans le cas où le type de matrice est carrée asymétrique ou rectangulaire, les conditions

sur les matrices peuvent être "ligne" ou "pas de condition". Pour la condition "ligne", les éléments d'une ligne sont comparés uniquement avec les éléments de cette même ligne. Dans la situation où il n'y a pas de condition, les éléments de toutes les matrices peuvent être comparés ensemble.

On remarque en effet qu'ALSCAL est le plus polyvalent des programmes présentés ici puisqu'il satisfait à toutes les conditions présentées dans le tableau 3.2. Il peut traiter une multitude de cas comme on l'a vu dans la section précédente.

Les deux autres programmes les plus polyvalents sont KYST et POLYCON qui se ressemblent d'ailleurs beaucoup. Le seul modèle qu'ils ne traitent pas est le modèle pondéré.

Contrairement à KYST et POLYCON, INDSCAL est spécialisé uniquement dans les modèles pondérés. Par le fait même, il traite les cas où il y a plusieurs matrices d'entrées (au moins 2) et où le modèle est métrique. De plus, les données de chaque matrice sont comparées uniquement entre elles et non avec les données des autres matrices.

MULTISCALE est un peu limité aux niveaux de ses caractéristiques comparativement à ALSCAL, KYST et POLYCON puisqu'il ne traite que les matrices carrées. Il n'utilise d'ailleurs que les modèles métriques puisqu'il ne traite que des types de mesures intervalles ou ratios.

Pour sa part, MINISSA est un programme particulier puisqu'il ne traite qu'une seule matrice en entrée avec des modèles non-métriques et non-pondérés où les types de mesures sont uniquement ordinales. Il est donc beaucoup plus limité que les autres avec lesquels il est comparé. Il faut tout de même préciser que ces restrictions ne font pas de lui un mauvais programme. Il est d'ailleurs considéré comme étant remarquablement robuste

[7].

Le tableau 3.3 présente le nombre maximal de stimuli, de matrices en entrée et de dimensions pour chacun des 6 programmes. Le symbole  $\infty$  signifie qu'il n'y a pas de limite au nombre maximal de matrices en entrée.

Tableau 3.3 – Tableau 2 des caractéristiques des 6 programmes

	A	I	K	MI	MU	P
Nombre maximal de stimuli	100	100	100	100	50	60
Nombre maximal de matrices en entrée	$\infty$	99	$\infty$	1	100	$\infty$
Nombre maximal de dimensions	6	10	6	10	10	10

POLYCON et MULTISCALE sont désavantagés pour la taille de la matrice d'entrée puisqu'ils ne prennent qu'au maximum respectivement 60 et 50 stimuli, c'est-à-dire 60 et 50 objets ou colonnes, contrairement à ALSCAL, INDSCAL, KYST et MINISSA qui acceptent tous 100 stimuli. Au niveau du nombre de matrice en entrée, MINISSA se démarque des autres puisqu'il ne prend qu'une seule matrice tandis que ALSCAL, KYST et POLYCON n'ont pas de limite. Le seul point un peu plus faible de ALSCAL par rapport aux autres est le nombre de dimensions maximal qui est le même que celui de KYST, mais est plus petit que tous les autres.

### 3.3 Une application à l'aide de ALSCAL

#### 3.3.1 Introduction des jeux de données utilisés

On utilise ici les jeux de données d'une étude faite sur les breuvages colas qui a été introduit à la section 2.4. Ces jeux de données proviennent du livre de Schiffman, Reynolds et Young [7]. Rappelons que cette étude comporte 10 répondants et que les stimuli sont

les 10 sortes de breuvages suivants :

1. Pepsi diète ;
2. RC Cola ;
3. Yukon ;
4. Dr. Pepper ;
5. Shasta ;
6. Coca-Cola ;
7. Dr. Pepper diète ;
8. Tab ;
9. Pepsi-cola ;
10. Rite diète ;

Dans un premier temps, chaque répondant a évalué chaque paire de breuvages en notant de 0 à 100 où 0 signifie "pareil" et 100, "différent". De cela découle les jeux de données trouvés en annexe à la section 3 où on a 10 matrices représentant chacun des 10 répondants. On y retrouve aussi une matrice indiquant les moyennes de chaque comparaison et ce, pour les 45 paires de breuvages.

Dans un deuxième temps, les répondants ont évalué individuellement ces 10 breuvages sur les 13 attributs suivants :

1. Bon versus mauvais ;
2. fort versus délicat ;
3. sucré versus non-sucré ;
4. amer versus non-amer ;
5. aigre versus non-aigre ;
6. fruité versus non-fruité ;

7. épiché versus non-épiché ;
8. enrobe la bouche versus n'enrobe pas la bouche ;
9. âpre versus non-âpre ;
10. fait plisser les lèvres versus ne fait pas plisser les lèvres ;
11. frais versus non-frais ;
12. chimique versus non-chimique ;
13. complexe versus non-complexe.

Ici encore, les attributs furent évalués de 0 à 100 où 0 représente le terme de gauche et 100, celui de droite. Les deux autres jeux de données de la troisième section de l'annexe représentent ces évaluations individuelles ainsi que ces évaluations moyennes.

### **3.3.2 Application des jeux de données dans ALSCAL**

Dans la section 3.1, on a défini de façon détaillée toutes les options disponibles dans le programme ALSCAL. Dans celle-ci, on applique les jeux de données des breuvages colas et on interprète les résultats. On étudiera les cas de positionnement multidimensionnel non-pondérés, pondéré ainsi que le cas où l'on produit un espace conjoint avec les attributs.

#### **Positionnement multidimensionnel non-pondéré**

Dans cette partie, on utilisera deux jeux de données, celui possédant l'évaluation par paires de chaque sujet et celui possédant l'évaluation par paires pour la moyenne des sujets.

Dans ces deux évaluations, on est dans le cas de matrices carrées symétriques. De plus, on garde comme option de type de mesure "ordinal" mais on coche la case "Untie tied

observation" pour que les données soient traitées comme continues. On conserve aussi comme condition "Matrix" qui fait que les données sont comparées entre elles mais pas d'une matrice à l'autre s'il y en a plusieurs en entrée comme c'est le cas pour les dix sujets. Les nombres maximal et minimal de dimensions restent 2 et, puisque l'on est dans un modèle non-pondéré, on conserve le modèle de distance euclidienne. Pour ce qui est des critères de convergence, on laisse ceux par défaut.

Pour l'évaluation faite avec le jeu de données contenant les opinions des 10 sujets individuellement, on obtient tout d'abord en sortie chaque valeur du SSTRESS pour chaque itération de l'algorithme que l'on retrouve au tableau 3.4.

Tableau 3.4 – Valeurs de SSTRESS

Itération	SSTRESS	Amélioration
1	0.44185	
2	0.41811	0.02375
3	0.41444	0.00367
4	0.41306	0.00138
5	0.41247	0.00059

L'algorithme s'est arrêté puisque l'amélioration était plus petite que 0,001.

On obtient ensuite la valeur de stress de Kruskal ainsi que la corrélation carrée RSQ pour chacune des 10 matrices. Rappelons que la valeur de la corrélation carrée est la proportion de variance expliquée des données positionnées qui sont expliquées par leur distance correspondante. Ces valeurs sont présentées dans le tableau 3.5.

On obtient aussi la valeur moyenne de stress qui est de 0.32224 et de RSQ, 0.40285. Puisque cette valeur de stress est élevé et que celle de RSQ ne l'est pas, on est porté

Tableau 3.5 – Stress et corrélation carrée

Matrice	Stress	RSQ
1	0.303	0.482
2	0.352	0.288
3	0.332	0.366
4	0.328	0.372
5	0.314	0.430
6	0.302	0.474
7	0.341	0.322
8	0.316	0.441
9	0.302	0.473
10	0.328	0.381

à croire que la représentation graphique en deux dimensions n'expliquera pas toutes les dissimilarités initiales.

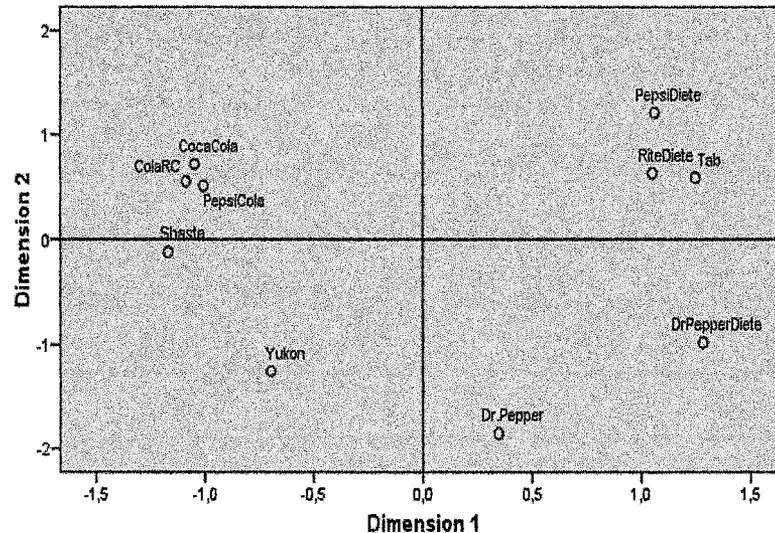
Les coordonnées des stimuli sont données dans le tableau 3.6.

Tableau 3.6 – Coordonnées des stimuli

Stimuli	Dimension 1	Dimension 2
1. Pepsi diète	1.0639	1.2090
2. RC Cola	-1.0861	0.5529
3. Yukon	-0.6946	-1.2585
4. Dr. Pepper	0.3474	-1.8568
5. Shasta	-1.1668	-0.1212
6. Coca-Cola	-1.0452	0.7192
7. Dr. Pepper diète	1.2840	-0.9783
8. Tab	1.2497	0.5903
9. Pepsi-cola	-1.0063	0.5120
10. Rite diète	1.0543	0.6313

Ce qui amène à l'espace dérivé des breuvages colas qui est représenté à la figure 3.1.

Figure 3.1 – Espace dérivé des breuvages colas pour l'évaluation individuelle des 10 sujets dans un modèle non-pondéré



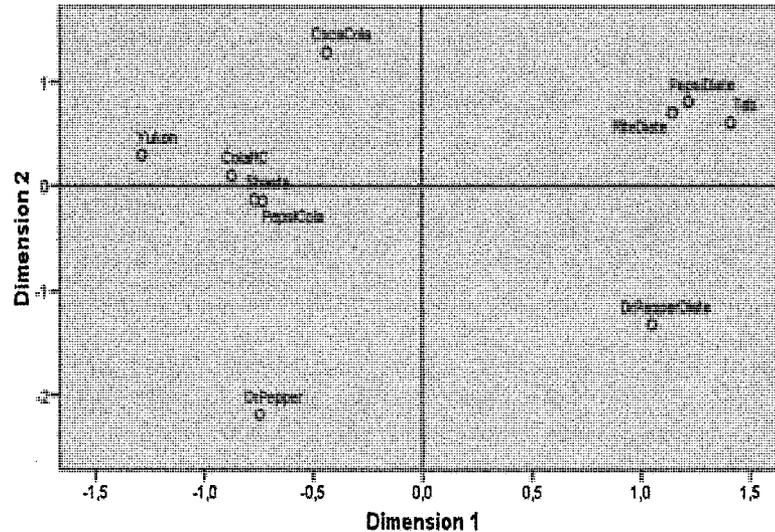
En examinant la configuration résultante, on peut subjectivement déterminer les dimensions. On note immédiatement que la dimension 1 représente de non-diète à diète puisque Pepsi diète, Rite diète et Dr.Pepper diète se retrouvent tous à l'extrême gauche du graphique. Pour ce qui est de la dimension 2, on remarque que Dr.Pepper et Dr.Pepper diète se situent dans le bas de la configuration. Ceci laisse croire que cette dimension représente les boissons cerisées versus non-cerisées. Il est à noter qu'un spécialiste des boissons gazeuses pourrait faire un analyse plus approfondie de cette configuration.

Pour ce qui est du cas où on a le jeu de données contenant la moyenne des évaluations des sujets, on obtient aussi la valeur de SSTRESS pour chaque itération. L'arrêt de l'algorithme est encore dû au fait que l'amélioration était plus petite que 0,001. À la première itération, la valeur était de 0.27437 et après les 10 itérations, elle était de 0.22027. On obtient pour cette évaluation, une valeur du stress de Kruskal de 0.19869. De plus, la valeur de la corrélation carrée est de 0.79987 qui est beaucoup plus élevé que la précédente. On peut ainsi supposer que cette configuration représente bien les données

initiales.

L'espace dérivé se retrouve à la figure 3.2.

Figure 3.2 – Espace dérivé des breuvages colas pour l'évaluation moyenne des 10 sujets dans un modèle non-pondéré



On tire les mêmes conclusions que précédemment au sujet des dimensions puisque Rite diète, Pepsi diète et Dr.Pepper diète se retrouvent tous encore à la droite du graphique et Dr.Pepper et Dr.Pepper diète se situent dans le bas. On remarque par contre que Yukon est beaucoup plus haut dans notre configuration, ce qui a plus de sens si la dimension 2 représente les boissons cerisées à non-cerisées. De plus, Coca-Cola se distingue plus dans cette configuration du groupe ColaRC, Shasta et Pepsi Cola.

### Positionnement multidimensionnel pondéré

Il est intéressant d'utiliser le positionnement multidimensionnel pondéré pour connaître l'importance que chaque répondant donne à chaque dimension. On peut ensuite comparer ces sujets selon ces importances.

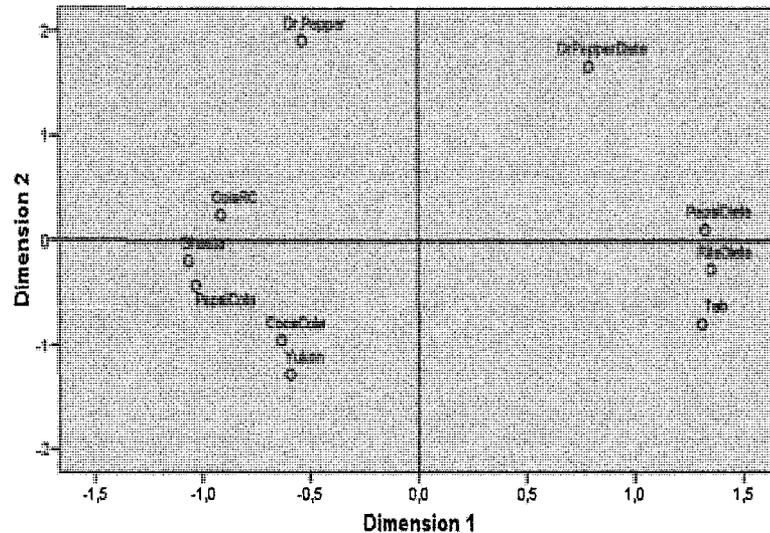
Pour travailler avec le positionnement multidimensionnel pondéré dans ALSCAL, il faut obligatoirement posséder plusieurs matrices d'évaluation, c'est-à-dire une par répondant. Dans le cas présent, on utilise le jeu de données ayant les 45 évaluations pour chacun des 10 sujets. Il faut aussi sélectionner "Individual differences Euclidean distances" comme modèle.

Les autres choix d'options se font comme dans le cas du positionnement multidimensionnel non-pondéré. Dans le cas évalué ici, on a choisi un type de mesure ordinal et on a coché à nouveau "Untie tied observation". On possède des matrices carrées symétriques et on conserve la condition "Matrix". Les nombres minimal et maximal de dimensions restent 2 et on ne permet pas de poids négatifs. Pour ce qui est des critères d'arrêt de l'algorithme, on maintient les mêmes que ceux par défaut.

On obtient comme précédemment toutes les valeurs du SSTRESS pour chaque itération de l'algorithme. La valeur initiale du SSTRESS est de 0.45326 et la valeur finale qui se trouve à l'itération 7 est de 0.40284. L'algorithme arrête encore à cause de l'amélioration qui est plus petite que 0,001. La valeur finale du SSTRESS étant élevé laisse croire que la configuration finale ne représentera pas très bien les données initiales. Le stress de Kruskal moyen des 10 matrices est de 0.30140 et le RSQ moyen est de 0.48200. Ayant un stress élevé et un RSQ plutôt bas, on est porté à croire, en effet, que la configuration finale n'est pas très bonne. Cette configuration se trouve à la figure 3.3.

Pepsi diète, Rite diète et Dr.Pepper diète se retrouvent encore à l'extrême droite du graphique, ce qui laisse supposer à nouveau que la dimension 1 va de non-diète à diète. Par contre, la dimension 2 est inversée par rapport à ce qu'on a retrouvé précédemment puisque Dr.Pepper et Dr.Pepper diète se retrouvent dans le haut du graphique. Ainsi, on peut dire que la dimension 2 signifie de plus à moins cerisé.

Figure 3.3 – Espace dérivé des breuvages colas pour l'évaluation individuelle des 10 sujets dans un modèle pondéré



Puisqu'on est dans un cas de positionnement multidimensionnel pondéré, on souhaite obtenir les poids des sujets qui mesurent l'importance de chaque dimension pour chaque sujet. Le tableau 3.7 donne ces poids ainsi que l'indice d'influence. Rappelons qu'un répondant ayant des poids proportionnels aux poids moyens auront un indice d'influence de 0 tandis que les sujets ayant un gros poids pour une dimension et un petit pour l'autre aura un indice d'influence près de 1.

L'importance moyenne pour la dimension 1 est 0.3016 et pour la dimension 2, 0.1804.

La figure 3.4 représente l'emplacement des sujets selon l'importance qu'ils donnent à chaque dimension.

On y remarque que le sujet 5, par exemple, donne une grande importance à la dimension 1 mais une minuscule à la dimension 2. C'est tout le contraire pour le sujet 10. Pour ce qui est du 2<sup>ième</sup> sujet, il donne une petite importance aux deux dimensions. Ainsi, le 5<sup>ième</sup> sujet aura donc un étendue des poids des sujets élevé. Pour le 10<sup>ième</sup> sujet, il sera

Tableau 3.7 – Pondération des dimensions

Sujet	Indice d'influence	Dimension 1	Dimension 2
1	0.0683	0.5934	0.4172
2	0.0996	0.3716	0.3404
3	0.3093	0.3601	0.4677
4	0.1541	0.6738	0.4131
5	0.4647	0.7612	0.2664
6	0.3768	0.7624	0.3180
7	0.1665	0.3773	0.3849
8	0.3359	0.4430	0.6035
9	0.1272	0.5488	0.3514
10	0.3845	0.3771	0.5625

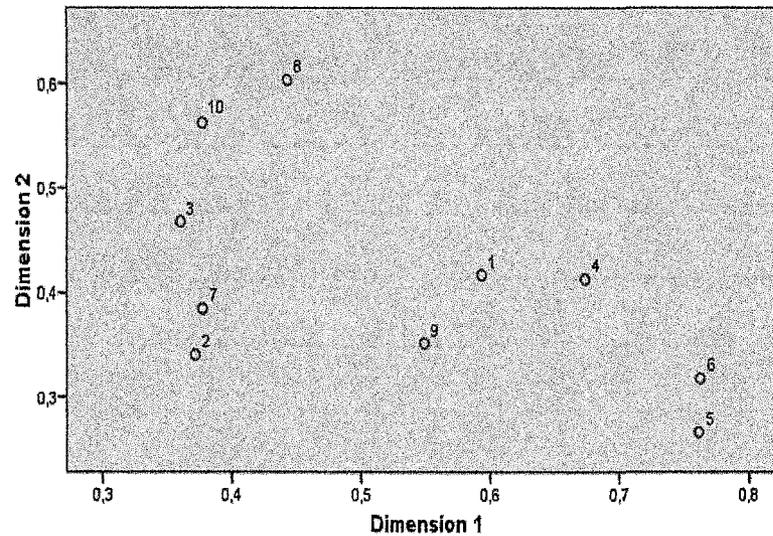
bas et pour le 2<sup>ième</sup> sujet, il sera près de 0. Le tableau 3.8 donne toutes les valeurs de l'étendue des poids des sujets pour les 10 sujets et la figure 3.5 affiche ces données dans un graphique en une dimension.

Tableau 3.8 – Étendues des poids des sujets

Sujet	Étendue des poids des sujets"
1	0.2963
2	-0.2863
3	-1.0622
4	0.5887
5	1.6673
6	1.3544
7	-0.5262
8	-1.1667
9	0.4971
10	-1.3623

On peut ainsi facilement voir que les sujets 1,4,5,6 et 9 accordent plus d'importance à la dimension 1 et les sujets 2,3,7,8 et 10, pour leur part, donnent plus d'importance à la

Figure 3.4 – Configuration de la pondération des dimensions selon les 10 sujets



dimension 2.

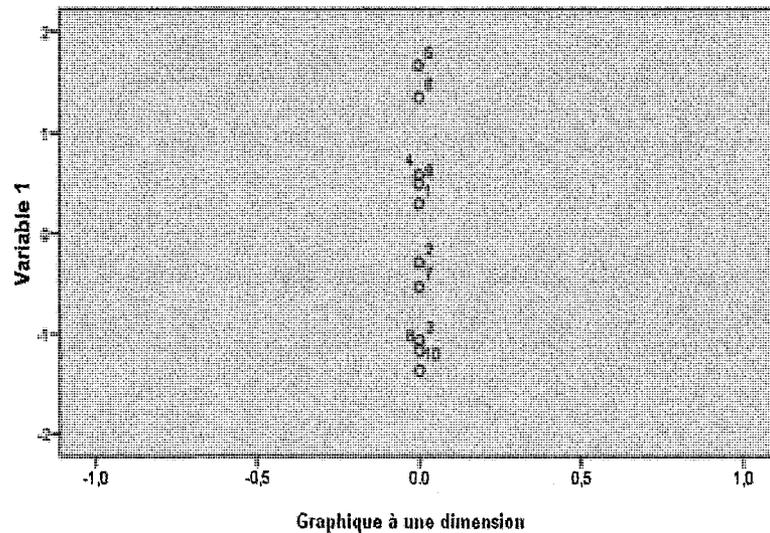
### Analyse de l'espace conjoint

Le positionnement multidimensionnel traite des matrices carrées symétriques et asymétriques. On sait aussi qu'il peut utiliser des matrices rectangulaires pour faire son analyse où les lignes et les colonnes ne représentent pas la même chose. Ce type de positionnement multidimensionnel produit un espace conjoint où les éléments qui représentent les lignes sont affichés dans la même configuration que les objets qui représentent les colonnes.

Dans le cas des breuvages colas, on utilise deux des jeux de données de la troisième section de l'annexe. Le premier est celui où les 10 sujets évaluent les 10 objets sur les 13 attributs et le deuxième, celui où on a l'évaluation moyenne des 10 sujets pour les 10 objets et les 13 attributs.

Dans l'utilisation de ces deux jeux de données, on travaillera encore avec des données

Figure 3.5 – Graphique des étendues des poids des sujets



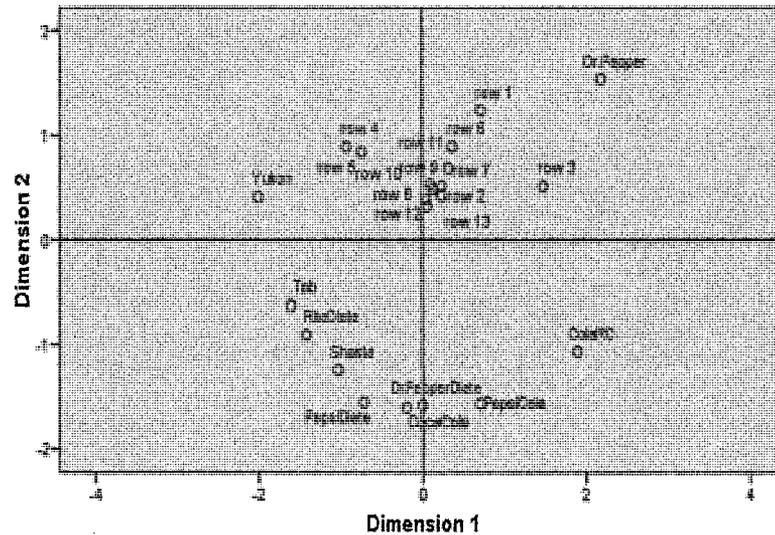
ordinales continues, avec un modèle de distances euclidiennes et avec un nombre minimal et un nombre maximal de dimensions de 2. De plus, on sélectionne l'option "Rectangular" et on indique le nombre de lignes qui, dans notre cas, est 13. La condition sur l'évaluation des données dans la ou les matrices ne sera plus "Matrix" ici mais bien "Row" puisque chaque ligne représente un attribut différent. Ainsi, on ne peut pas comparer deux valeurs provenant de deux lignes différentes ensembles puisque ces valeurs n'évaluent pas le même attribut. On garde les mêmes critères d'arrêt de l'algorithme que ceux par défaut.

Dans un premier temps, on traite le jeu de données contenant l'évaluation individuelle des 10 sujets. Contrairement aux applications faites précédemment, l'algorithme s'arrête suite au nombre maximum d'itérations, c'est-à-dire 30. À la première itération, on avait une valeur de SSTRESS de 0.48228 et à la trentième, 0.17543.

On obtient aussi en sortie toutes les valeurs de stress de Kruskal ainsi que les valeurs des corrélations carrées pour chaque attribut de chaque matrice (de chaque sujet). Le stress de Kruskal moyen est de 0.223 et le RSQ moyen est de 0.954, ce qui est énorme et qui montre

que la configuration finale représente bien les données initiales. Cette configuration est représentée dans la figure 3.6.

Figure 3.6 – Espace conjoint des breuvages colas et des attributs

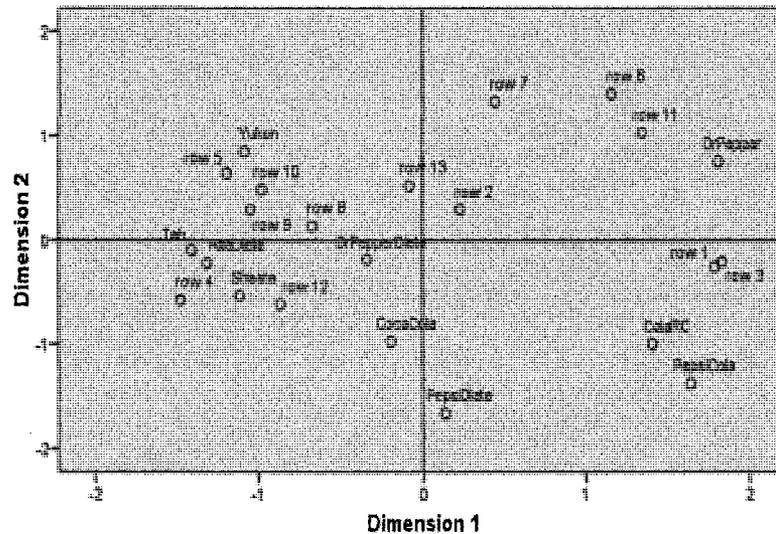


Dans l'espace conjoint, plus deux objets sont près l'un de l'autre, plus ils se ressemblent. C'est la même chose pour deux attributs. Plus un objet est près d'un attribut, plus cet objet possède cet attribut. Dans notre cas, cela signifie que l'objet se rapproche de l'adjectif de gauche dans la liste des attributs énoncée au début de cette section. Par exemple, les répondants trouvent que la boisson Dr.Pepper est bonne contrairement au Pepsi diète qui se situe très loin du point de l'attribut "bon". De plus, Yukon possède les caractéristiques d'être amer et aigre. On remarque dans ce graphique que tous les attributs sont amassés au centre tandis que les boissons les encerclent mais sont surtout situées dans le bas.

En second lieu, on traite l'évaluation moyenne des sujets pour les attributs sur chaque objet. Le critère d'arrêt pour l'algorithme est encore le nombre d'itérations maximal. La première valeur du SSTRESS est de 0.21501 et la trentième est de 0.12661, ce qui est très petit. Le stress de Kruskal moyen est de 0.143 et la valeur moyenne de la corrélation

carrée pour les 13 attributs est 0.981. On peut ainsi conclure que la configuration finale créée par le positionnement multidimensionnel explique très bien les données moyennes initiales. Cet espace dérivé est représenté à la figure 3.7.

Figure 3.7 – Espace conjoint des breuvages colas et des attributs



On remarque, encore ici, que le Dr.Pepper est situé près de l'attribut "bon". Par contre, il n'est pas le seul. ColaRC et Pepsi Cola le sont aussi. Tab, Shasta et Rite diète possèdent tous l'attribut amer selon le graphique et Dr.Pepper possède la caractéristique d'enrober la bouche. De plus, on peut voir dans cette configuration qu'aucun breuvage ne possède l'attribut épicé puisque tout le monde se situe loin de son point. Ce résultat est logique puisque les breuvages colas ne sont normalement pas épicés.

# CONCLUSION

Ce mémoire expose plusieurs techniques de positionnement multidimensionnel qui affichent l'espace dérivé des objets à l'étude. Ces techniques utilisent des modèles non-pondérés ou pondérés, classiques ou répliqués et des modèles utilisant des matrices carrées symétriques, asymétriques ou rectangulaires. Rappelons que l'utilité des matrices rectangulaires est de produire un espace conjoint qui affiche dans la configuration finale, autant les variables qui représentent les lignes que les variables qui représentent les colonnes. Ce genre de matrice peut aussi produire des vecteurs d'attributs lorsqu'on travaille avec la méthode d'ajustement des attributs influents. Il aurait été intéressant d'utiliser le programme PROFIT (PROperty FITting) pour approfondir l'étude des jeux de données des breuvages colas. Dans un même ordre d'idée, puisqu'il existe une multitude de programmes informatiques avec lesquels on peut tester tous les genres de modèles, il serait intéressant de les tester sur les jeux de données utilisés ici et ensuite les comparer.

Ce travail a aussi affiché ce que peut faire le programme ALSCAL, mais n'en n'a exploré qu'une infime partie. On a en effet traité les cas de positionnement multidimensionnel pondéré et non-pondéré et le cas d'espace conjoint. On aurait, par contre, pu explorer les cas où l'on traite les données comme étant des données ordinales discrètes ou intervalles et où il n'y a pas de condition sur l'évaluation des données, c'est-à-dire quand on coche "Unconditionnal" dans "Conditionnality". On aurait aussi pu vérifier ce que l'on obtient

en 3 dimensions ou plus et ce que l'on peut changer en modifiant les critères d'arrêt de l'algorithme. On remarque ainsi qu'ALSCAL est un programme qui nous donne une grande variété d'options pour obtenir la meilleure configuration possible. Il est à noter, par contre, qu'il faut être prudent car un mauvais choix d'option pourrait fausser les résultats.

Le positionnement multidimensionnel peut être appliquée à plusieurs genres de données. Dans ce mémoire, on s'est uniquement concentré aux distances, aux dissimilarités et aux similarités. Il aurait été très intéressant de travailler aussi avec des préférences. Ce genre de données apporterait une perspective complémentaire aux études effectuées. Par exemple, en plus de comparer plusieurs produits sur le marché, l'étude afficherait également ceux qui sont les préférés des consommateurs. Ceci donnerait une étude beaucoup plus complète.

# Annexe

## 1. Résultats importants

**Théorème 6.** *Théorème de la décomposition spectrale :*

*Toute matrice symétrique  $\mathbf{A}$  ( $p \times p$ ) peut être écrite comme*

$$\mathbf{A} = \mathbf{\Gamma} \mathbf{\Lambda} \mathbf{\Gamma}' = \sum_{i=1}^p \lambda_i \gamma_{(i)}' \gamma_{(i)}$$

*où  $\mathbf{\Lambda}$  est la matrice diagonale des valeurs propres de  $\mathbf{A}$  et  $\mathbf{\Gamma}$  est la matrice orthogonale où les colonnes sont les vecteurs propres normalisés de  $\mathbf{A}$ .*

Démonstration :

Voir la référence [1].

**Théorème 7.** *Théorème de décomposition de la valeur singulière :*

*Si  $\mathbf{A}$  est une matrice  $n \times p$  de rang  $r$ , alors  $\mathbf{A}$  peut être écrit comme*

$$\mathbf{A} = \mathbf{U} \mathbf{L} \mathbf{V}'$$

*où  $\mathbf{U}$  ( $n \times r$ ) et  $\mathbf{V}$  ( $p \times r$ ) sont des matrices orthogonales et  $\mathbf{L}$  est une matrice diagonale d'éléments positifs.*

Démonstration :

Voir la référence [1].

## Méthode des multiplicateurs de Lagrange

Soit le problème d'optimisation

$$\text{Min}F(X)$$

où  $X = (x_1, x_2, \dots, x_n)$  soumis à  $g_k(x) = 0$ ,  $k = 1, 2, \dots, m$  ( $m$  contraintes).

La méthode des multiplicateurs de Lagrange stipule qu'on atteint la valeur optimale de  $F(X)$  pour les valeurs  $x_i$  de  $X$  qui satisfont le système :

$$\begin{aligned} \frac{\partial F(X)}{\partial x_1} - \sum_{k=1}^m \lambda_k \frac{\partial g_k(x)}{\partial x_1} &= 0 \\ \frac{\partial F(X)}{\partial x_2} - \sum_{k=1}^m \lambda_k \frac{\partial g_k(x)}{\partial x_2} &= 0 \\ &\vdots \\ \frac{\partial F(X)}{\partial x_n} - \sum_{k=1}^m \lambda_k \frac{\partial g_k(x)}{\partial x_n} &= 0 \\ g_1(x) &= 0 \\ g_2(x) &= 0 \\ &\vdots \\ g_m(x) &= 0 \end{aligned}$$

Les  $\lambda_k$ ,  $k = 1, 2, \dots, m$ , sont les multiplicateurs de Lagrange.

## 2. Mesures de distances

L'exemple suivant énonce plusieurs mesures de distances utilisées dans ce mémoire.

**Exemple 19.** Soient  $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^p$ ,  $x_{ik}$  la  $k^{\text{ième}}$  coordonnée de  $\mathbf{x}_i$ ,  $x_{jk}$  la  $k^{\text{ième}}$  coordonnée de  $\mathbf{x}_j$  et  $r \in \mathbb{N}$ .

1- La distance euclidienne est définie comme suit :

$$\sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2}$$

2- La distance euclidienne au carrée est définie comme suit :

$$\sum_{k=1}^p (x_{ik} - x_{jk})^2$$

3- La distance de Karl-Pearson est définie comme suit :

$$\sqrt{\sum_{k=1}^p \frac{(x_{ik} - x_{jk})^2}{s_k^2}}$$

où  $s_k$  est la variance de la  $k^{\text{ième}}$  variable. ( $s_k^2 = \frac{1}{n-1} \sum_{i=1}^n (x_{ik} - \bar{x}_k)$ )

4- La distance de Mahalanobis est définie comme suit :

$$\sqrt{(\mathbf{x}_i - \mathbf{x}_j)' \mathbf{S}' (\mathbf{x}_i - \mathbf{x}_j)}$$

où  $\mathbf{S}$  est la matrice des covariances. ( $\mathbf{S} = \frac{1}{n} \sum_{r=1}^n (\mathbf{x}_r - \bar{\mathbf{x}})(\mathbf{x}_r - \bar{\mathbf{x}})'$ )

5- La distance de Minkowski est définie comme suit :

$$\left( \sum_{k=1}^p |x_{ik} - x_{jk}|^r \right)^{1/r}$$

6- La distance de Chebychev est définie comme suit :

$$\text{Max}_{k=1, \dots, p} |x_{ik} - x_{jk}|$$

7- La distance des blocs ou de Manhattan est définie comme suit :

$$\sum_{k=1}^p |x_{ik} - x_{jk}|$$

**Exemple 20.** Mesures de distances pour des données binaires :

Supposons que l'on a  $n$  objets et  $p$  attributs. Si l'objet  $k$  possède l'attribut  $l$ , on note 1. Au contraire, s'il ne possède pas l'attribut, on note 0. Ces données sont regroupées dans une matrice notée  $X$  où  $x_i$  est le vecteur qui représente le  $i^{\text{ème}}$  objet.

Dans le tableau 3.9,  $a, b, c$  et  $d$  possèdent les significations suivantes :

- $a$  : nombre d'attributs que les points  $x_i$  et  $x_j$  possèdent en commun ;
- $b$  : nombre d'attributs que le point  $x_i$  possède mais que le point  $x_j$  ne possède pas ;
- $c$  : nombre d'attributs que le point  $x_i$  ne possède pas mais que le point  $x_j$  possède ;
- $d$  : nombre d'attributs que ni le point  $x_i$  et ni le point  $x_j$  ne possèdent.

Tableau 3.9 – Tableau des nombres d'attributs

	1	0
1	$a$	$b$
0	$c$	$d$

Il est à noter que  $n = a + b + c + d$

La liste des mesures de distance de données binaires est la suivante :

1- Distance euclidienne :

$$\sqrt{b + c}$$

2- Distance euclidienne carrée :

$$b + c$$

3- "Pattern distance" :

$$\frac{(b * c)}{n^2}$$

4- Variance :

$$\frac{(b + c)}{4n}$$

5- Lance et William (aussi connu sous le nom de "coefficient non-métrique de Bray-Curtis") :

$$\frac{(b + c)}{2a + b + c}$$

### 3. Jeux de données des breuvages colas

Les jeux de données des breuvages colas proviennent du livre de Schiffman, Young et Reynolds [7].

Le jeu de données suivant représente les évaluations de dissimilarités de 10 répondants sur les 10 breuvages colas. Chaque ensemble de 10 lignes représente un répondant. Par exemple, 2-01 à 2-10 représentent l'évaluation faites par le deuxième répondant où 2-01 signifie que cette ligne est la comparaison entre le premier breuvage cola faite par le deuxième répondant et ainsi de suite.

	1	2	3	4	5	6	7	8	9	10
1 - 01	,00									
1 - 02	16,00	,00								
1 - 03	81,00	47,00	,00							
1 - 04	56,00	32,00	71,00	,00						
1 - 05	87,00	68,00	44,00	71,00	,00					
1 - 06	60,00	35,00	21,00	98,00	34,00	,00				
1 - 07	84,00	94,00	98,00	57,00	99,00	99,00	,00			
1 - 08	50,00	87,00	79,00	73,00	19,00	92,00	45,00	,00		
1 - 09	99,00	25,00	53,00	98,00	52,00	17,00	99,00	84,00	,00	
1 - 10	16,00	92,00	90,00	83,00	79,00	44,00	24,00	18,00	98,00	,00
2 - 01	,00									
2 - 02	9,00	,00								
2 - 03	90,00	70,00	,00							
2 - 04	87,00	65,00	6,00	,00						
2 - 05	87,00	77,00	83,00	83,00	,00					
2 - 06	33,00	79,00	25,00	89,00	39,00	,00				
2 - 07	86,00	86,00	99,00	22,00	90,00	40,00	,00			
2 - 08	81,00	30,00	57,00	88,00	69,00	39,00	97,00	,00		
2 - 09	74,00	20,00	94,00	78,00	5,00	81,00	92,00	88,00	,00	
2 - 10	23,00	26,00	72,00	94,00	2,00	76,00	81,00	20,00	5,00	,00

	1	2	3	4	5	6	7	8	9	10
3-01	,00									
3-02	49,00	,00								
3-03	96,00	96,00	,00							
3-04	97,00	92,00	94,00	,00						
3-05	68,00	12,00	90,00	93,00	,00					
3-06	77,00	44,00	88,00	90,00	26,00	,00				
3-07	97,00	93,00	94,00	25,00	93,00	49,00	,00			
3-08	54,00	76,00	92,00	94,00	20,00	24,00	93,00	,00		
3-09	47,00	48,00	92,00	94,00	35,00	18,00	94,00	23,00	,00	
3-40	21,00	47,00	90,00	92,00	68,00	67,00	87,00	55,00	15,00	,00
4-01	,00									
4-02	23,00	,00								
4-03	99,00	51,00	,00							
4-04	99,00	23,00	78,00	,00						
4-05	90,00	16,00	22,00	49,00	,00					
4-06	74,00	55,00	50,00	99,00	13,00	,00				
4-07	14,00	88,00	77,00	75,00	50,00	70,00	,00			
4-08	25,00	95,00	48,00	99,00	99,00	79,00	99,00	,00		
4-09	60,00	36,00	69,00	24,00	21,00	53,00	99,00	99,00	,00	
4-10	,00	89,00	72,00	81,00	77,00	71,00	74,00	51,00	71,00	,00

	1	2	3	4	5	6	7	8	9	10
5 - 01	,00									
5 - 02	62,00	,00								
5 - 03	77,00	16,00	,00							
5 - 04	98,00	14,00	55,00	,00						
5 - 05	76,00	22,00	40,00	47,00	,00					
5 - 06	84,00	16,00	16,00	81,00	7,00	,00				
5 - 07	17,00	80,00	36,00	93,00	60,00	90,00	,00			
5 - 08	76,00	93,00	86,00	80,00	94,00	36,00	19,00	,00		
5 - 09	74,00	20,00	16,00	38,00	5,00	18,00	6,00	71,00	,00	
5 - 10	10,00	72,00	78,00	92,00	92,00	86,00	16,00	2,00	99,00	,00
6 - 01	,00									
6 - 02	85,00	,00								
6 - 03	82,00	15,00	,00							
6 - 04	97,00	28,00	56,00	,00						
6 - 05	51,00	31,00	36,00	43,00	,00					
6 - 06	79,00	27,00	7,00	82,00	7,00	,00				
6 - 07	13,00	84,00	38,00	87,00	76,00	82,00	,00			
6 - 08	82,00	99,00	73,00	68,00	80,00	40,00	20,00	,00		
6 - 09	69,00	24,00	30,00	27,00	16,00	12,00	28,00	80,00	,00	
6 - 10	15,00	80,00	78,00	90,00	72,00	66,00	17,00	5,00	95,00	,00

	1	2	3	4	5	6	7	8	9	10
7-01	,00									
7-02	10,00	,00								
7-03	53,00	75,00	,00							
7-04	99,00	99,00	99,00	,00						
7-05	87,00	27,00	65,00	99,00	,00					
7-06	60,00	66,00	72,00	99,00	99,00	,00				
7-07	96,00	99,00	90,00	10,00	90,00	75,00	,00			
7-08	98,00	99,00	91,00	98,00	88,00	34,00	99,00	,00		
7-09	73,00	15,00	90,00	99,00	9,00	56,00	95,00	75,00	,00	
7-10	54,00	62,00	84,00	99,00	95,00	53,00	85,00	91,00	49,00	,00
8-01	,00									
8-02	14,00	,00								
8-03	61,00	47,00	,00							
8-04	79,00	96,00	77,00	,00						
8-05	72,00	21,00	12,00	73,00	,00					
8-06	66,00	12,00	28,00	81,00	13,00	,00				
8-07	66,00	64,00	75,00	41,00	71,00	82,00	,00			
8-08	51,00	67,00	32,00	93,00	49,00	66,00	86,00	,00		
8-09	7,00	20,00	67,00	71,00	15,00	56,00	76,00	69,00	,00	
8-10	19,00	51,00	6,00	88,00	25,00	81,00	50,00	8,00	83,00	,00

	1	2	3	4	5	6	7	8	9	10
9-01	,00									
9-02	11,00	,00								
9-03	90,00	69,00	,00							
9-04	72,00	26,00	90,00	,00						
9-05	93,00	17,00	69,00	24,00	,00					
9-06	39,00	34,00	36,00	98,00	80,00	,00				
9-07	26,00	82,00	77,00	85,00	53,00	99,00	,00			
9-08	80,00	74,00	75,00	99,00	93,00	87,00	13,00	,00		
9-09	73,00	8,00	91,00	35,00	17,00	17,00	99,00	91,00	,00	
9-10	24,00	62,00	90,00	76,00	85,00	64,00	77,00	24,00	65,00	,00
10-01	,00									
10-02	69,00	,00								
10-03	63,00	58,00	,00							
10-04	76,00	85,00	79,00	,00						
10-05	52,00	14,00	51,00	81,00	,00					
10-06	61,00	39,00	35,00	83,00	36,00	,00				
10-07	80,00	90,00	93,00	6,00	78,00	85,00	,00			
10-08	28,00	87,00	83,00	94,00	64,00	44,00	90,00	,00		
10-09	80,00	20,00	92,00	98,00	51,00	23,00	80,00	33,00	,00	
10-10	78,00	28,00	40,00	99,00	36,00	71,00	82,00	62,00	13,00	,00

La matrice suivante représente l'évaluation moyenne que les 10 répondants font des 10 breuvages. Ainsi, les lignes et les colonnes représentent les boissons.

	1	2	3	4	5	6	7	8	9	10
1	,00									
2	34,00	,00								
3	79,00	54,00	,00							
4	86,00	56,00	70,00	,00						
5	76,00	30,00	51,00	66,00	,00					
6	63,00	40,00	37,00	90,00	35,00	,00				
7	57,00	86,00	77,00	50,00	76,00	77,00	,00			
8	62,00	80,00	71,00	88,00	67,00	54,00	66,00	,00		
9	65,00	23,00	69,00	66,00	22,00	35,00	76,00	71,00	,00	
10	26,00	60,00	70,00	89,00	63,00	67,00	59,00	33,00	59,00	,00

Les 10 matrices suivantes représentent les évaluations des 10 répondants. Ainsi, chaque matrice possède 10 colonnes qui sont les breuvages colas et 13 colonnes qui sont les 13 attributs.

	1	2	3	4	5	6	7	8	9	10
1 - 01	98,00	3,00	2,00	47,00	90,00	99,00	10,00	39,00	37,00	76,00
1 - 02	98,00	91,00	74,00	31,00	5,00	,00	33,00	39,00	18,00	33,00
1 - 03	91,00	2,00	7,00	7,00	43,00	96,00	3,00	34,00	,00	99,00
1 - 04	4,00	99,00	99,00	78,00	8,00	1,00	87,00	99,00	80,00	15,00
1 - 05	99,00	99,00	76,00	99,00	81,00	17,00	89,00	18,00	10,00	85,00
1 - 06	99,00	91,00	27,00	62,00	98,00	99,00	40,00	17,00	10,00	99,00
1 - 07	98,00	96,00	50,00	31,00	96,00	97,00	78,00	99,00	32,00	99,00
1 - 08	48,00	96,00	99,00	99,00	99,00	98,00	60,00	99,00	95,00	6,00
1 - 09	27,00	99,00	99,00	49,00	2,00	,00	99,00	79,00	99,00	7,00
1 - 10	10,00	99,00	22,00	99,00	99,00	3,00	31,00	31,00	81,00	70,00
1 - 11	47,00	6,00	6,00	48,00	99,00	99,00	47,00	35,00	49,00	98,00
1 - 12	8,00	93,00	84,00	38,00	39,00	1,00	42,00	90,00	41,00	7,00
1 - 13	3,00	93,00	75,00	38,00	41,00	1,00	47,00	91,00	34,00	9,00

	1	2	3	4	5	6	7	8	9	10
2-01	35,00	84,00	97,00	13,00	94,00	22,00	29,00	88,00	40,00	84,00
2-02	89,00	82,00	18,00	70,00	28,00	73,00	77,00	37,00	33,00	26,00
2-03	11,00	82,00	96,00	3,00	87,00	13,00	16,00	85,00	34,00	82,00
2-04	92,00	83,00	5,00	90,00	16,00	80,00	87,00	15,00	64,00	21,00
2-05	91,00	85,00	5,00	84,00	3,00	84,00	85,00	15,00	68,00	14,00
2-06	79,00	77,00	36,00	2,00	47,00	20,00	17,00	89,00	17,00	82,00
2-07	83,00	71,00	16,00	33,00	11,00	26,00	46,00	7,00	48,00	46,00
2-08	78,00	73,00	16,00	88,00	29,00	63,00	83,00	12,00	69,00	13,00
2-09	85,00	74,00	10,00	89,00	6,00	76,00	80,00	15,00	70,00	14,00
2-10	82,00	81,00	3,00	91,00	3,00	83,00	89,00	19,00	72,00	16,00
2-11	87,00	87,00	47,00	30,00	64,00	43,00	77,00	60,00	69,00	50,00
2-12	82,00	17,00	14,00	83,00	,00	84,00	92,00	21,00	74,00	37,00
2-13	82,00	34,00	21,00	75,00	40,00	78,00	84,00	38,00	51,00	42,00

	1	2	3	4	5	6	7	8	9	10
3-01	40,00	89,00	99,00	25,00	47,00	69,00	70,00	67,00	46,00	84,00
3-02	83,00	45,00	75,00	19,00	35,00	42,00	22,00	45,00	41,00	63,00
3-03	44,00	76,00	78,00	13,00	19,00	50,00	23,00	37,00	15,00	72,00
3-04	87,00	87,00	49,00	95,00	96,00	96,00	92,00	72,00	96,00	92,00
3-05	95,00	88,00	47,00	95,00	94,00	95,00	76,00	81,00	96,00	94,00
3-06	93,00	94,00	95,00	10,00	94,00	95,00	11,00	97,00	96,00	93,00
3-07	93,00	92,00	94,00	97,00	94,00	93,00	96,00	95,00	96,00	96,00
3-08	75,00	78,00	94,00	50,00	49,00	49,00	49,00	53,00	55,00	49,00
3-09	94,00	89,00	49,00	95,00	29,00	37,00	39,00	73,00	95,00	93,00
3-10	83,00	21,00	5,00	93,00	50,00	95,00	97,00	86,00	96,00	84,00
3-11	26,00	95,00	97,00	8,00	70,00	49,00	28,00	79,00	71,00	92,00
3-12	49,00	19,00	4,00	94,00	12,00	39,00	41,00	25,00	65,00	30,00
3-13	46,00	50,00	50,00	50,00	21,00	38,00	37,00	40,00	57,00	49,00

	1	2	3	4	5	6	7	8	9	10
4-01	78,00	73,00	74,00	8,00	85,00	28,00	70,00	99,00	32,00	93,00
4-02	35,00	64,00	46,00	51,00	29,00	79,00	50,00	,00	53,00	53,00
4-03	89,00	21,00	73,00	,00	99,00	39,00	70,00	90,00	53,00	99,00
4-04	14,00	80,00	20,00	99,00	,00	78,00	25,00	9,00	82,00	18,00
4-05	51,00	99,00	52,00	99,00	14,00	99,00	52,00	19,00	84,00	18,00
4-06	99,00	80,00	26,00	20,00	99,00	90,00	35,00	99,00	21,00	87,00
4-07	99,00	99,00	80,00	58,00	99,00	73,00	80,00	99,00	86,00	99,00
4-08	51,00	79,00	24,00	87,00	7,00	26,00	80,00	82,00	85,00	51,00
4-09	18,00	99,00	51,00	99,00	24,00	89,00	16,00	17,00	99,00	31,00
4-10	72,00	89,00	25,00	88,00	,00	88,00	17,00	,00	99,00	10,00
4-11	79,00	68,00	51,00	15,00	82,00	23,00	65,00	61,00	50,00	77,00
4-12	22,00	76,00	74,00	99,00	19,00	88,00	76,00	22,00	91,00	21,00
4-13	50,00	77,00	25,00	26,00	82,00	76,00	29,00	48,00	90,00	36,00

	1	2	3	4	5	6	7	8	9	10
5 - 01	88,00	51,00	80,00	9,00	92,00	15,00	99,00	85,00	23,00	10,00
5 - 02	79,00	28,00	16,00	75,00	12,00	39,00	11,00	16,00	53,00	9,00
5 - 03	83,00	31,00	78,00	13,00	86,00	16,00	94,00	98,00	10,00	96,00
5 - 04	18,00	64,00	19,00	91,00	12,00	94,00	8,00	5,00	70,00	2,00
5 - 05	74,00	74,00	27,00	92,00	19,00	99,00	9,00	15,00	90,00	30,00
5 - 06	58,00	49,00	92,00	24,00	85,00	48,00	98,00	98,00	98,00	10,00
5 - 07	44,00	67,00	21,00	53,00	53,00	89,00	71,00	99,00	99,00	86,00
5 - 08	11,00	57,00	25,00	91,00	14,00	66,00	10,00	10,00	62,00	8,00
5 - 09	21,00	71,00	12,00	83,00	12,00	77,00	8,00	7,00	78,00	5,00
5 - 10	17,00	84,00	19,00	87,00	25,00	94,00	5,00	6,00	83,00	4,00
5 - 11	95,00	26,00	18,00	25,00	82,00	20,00	98,00	99,00	26,00	99,00
5 - 12	8,00	45,00	17,00	91,00	13,00	87,00	7,00	1,00	79,00	,00
5 - 13	9,00	57,00	12,00	83,00	29,00	72,00	14,00	37,00	85,00	25,00

	1	2	3	4	5	6	7	8	9	10
6-01	88,00	61,00	81,00	96,00	97,00	98,00	91,00	99,00	93,00	97,00
6-02	54,00	43,00	38,00	29,00	40,00	22,00	59,00	60,00	35,00	17,00
6-03	82,00	44,00	73,00	29,00	78,00	40,00	34,00	85,00	37,00	75,00
6-04	35,00	68,00	33,00	85,00	30,00	79,00	79,00	14,00	84,00	32,00
6-05	77,00	74,00	26,00	82,00	30,00	81,00	78,00	13,00	85,00	32,00
6-06	80,00	70,00	43,00	48,00	71,00	79,00	80,00	58,00	35,00	73,00
6-07	96,00	86,00	79,00	95,00	93,00	94,00	94,00	99,00	96,00	86,00
6-08	41,00	74,00	22,00	50,00	10,00	70,00	10,00	40,00	34,00	20,00
6-09	50,00	72,00	17,00	45,00	75,00	76,00	66,00	60,00	69,00	44,00
6-10	41,00	75,00	31,00	49,00	55,00	65,00	15,00	52,00	67,00	20,00
6-11	98,00	55,00	96,00	96,00	95,00	99,00	96,00	95,00	93,00	94,00
6-12	39,00	52,00	41,00	88,00	75,00	56,00	36,00	44,00	51,00	21,00
6-13	65,00	43,00	50,00	80,00	64,00	66,00	48,00	38,00	47,00	21,00

	1	2	3	4	5	6	7	8	9	10
7-01	31,00	5,00	67,00	99,00	52,00	36,00	99,00	94,00	30,00	98,00
7-02	90,00	51,00	57,00	12,00	92,00	21,00	11,00	72,00	68,00	59,00
7-03	19,00	3,00	84,00	24,00	37,00	19,00	16,00	89,00	20,00	96,00
7-04	85,00	98,00	19,00	90,00	94,00	81,00	90,00	13,00	82,00	7,00
7-05	27,00	98,00	30,00	91,00	80,00	94,00	35,00	31,00	67,00	10,00
7-06	99,00	99,00	10,00	,00	99,00	61,00	1,00	45,00	49,00	99,00
7-07	99,00	99,00	92,00	98,00	99,00	93,00	9,00	40,00	95,00	99,00
7-08	89,00	97,00	6,00	89,00	49,00	2,00	85,00	84,00	99,00	96,00
7-09	94,00	97,00	13,00	98,00	95,00	9,00	11,00	44,00	95,00	4,00
7-10	99,00	96,00	22,00	99,00	99,00	5,00	9,00	99,00	98,00	4,00
7-11	17,00	19,00	56,00	49,00	31,00	82,00	42,00	86,00	47,00	95,00
7-12	85,00	99,00	94,00	99,00	89,00	18,00	14,00	75,00	99,00	27,00
7-13	3,00	99,00	35,00	9,00	33,00	5,00	5,00	91,00	22,00	9,00

	1	2	3	4	5	6	7	8	9	10
8-01	79,00	57,00	71,00	19,00	51,00	67,00	61,00	47,00	66,00	78,00
8-02	87,00	92,00	59,00	26,00	16,00	25,00	25,00	37,00	46,00	21,00
8-03	37,00	28,00	68,00	31,00	56,00	70,00	73,00	74,00	4,00	77,00
8-04	33,00	84,00	45,00	98,00	14,00	14,00	13,00	20,00	84,00	24,00
8-05	84,00	99,00	63,00	76,00	28,00	30,00	42,00	30,00	89,00	34,00
8-06	88,00	90,00	74,00	22,00	72,00	86,00	52,00	84,00	46,00	90,00
8-07	82,00	83,00	63,00	12,00	49,00	35,00	31,00	41,00	87,00	22,00
8-08	59,00	73,00	28,00	51,00	10,00	17,00	34,00	34,00	36,00	45,00
8-09	87,00	77,00	40,00	33,00	20,00	22,00	37,00	52,00	89,00	29,00
8-10	99,00	83,00	45,00	42,00	28,00	25,00	24,00	43,00	47,00	55,00
8-11	98,00	85,00	72,00	22,00	75,00	85,00	76,00	92,00	90,00	81,00
8-12	38,00	62,00	57,00	43,00	20,00	6,00	18,00	20,00	56,00	17,00
8-13	55,00	71,00	64,00	10,00	70,00	47,00	46,00	57,00	40,00	30,00

	1	2	3	4	5	6	7	8	9	10
9-01	22,00	10,00	88,00	15,00	28,00	93,00	91,00	99,00	29,00	69,00
9-02	32,00	30,00	19,00	36,00	30,00	11,00	15,00	5,00	77,00	17,00
9-03	81,00	21,00	80,00	21,00	13,00	72,00	91,00	97,00	14,00	84,00
9-04	82,00	87,00	25,00	91,00	82,00	7,00	8,00	2,00	80,00	14,00
9-05	65,00	75,00	17,00	81,00	85,00	20,00	21,00	5,00	73,00	17,00
9-06	97,00	34,00	87,00	11,00	8,00	82,00	86,00	93,00	10,00	75,00
9-07	71,00	78,00	35,00	80,00	17,00	15,00	79,00	83,00	64,00	20,00
9-08	84,00	84,00	29,00	39,00	32,00	13,00	19,00	5,00	56,00	10,00
9-09	56,00	75,00	18,00	52,00	72,00	20,00	26,00	11,00	71,00	12,00
9-10	85,00	90,00	37,00	58,00	85,00	27,00	15,00	21,00	60,00	26,00
9-11	67,00	55,00	74,00	24,00	73,00	73,00	78,00	75,00	20,00	73,00
9-12	42,00	81,00	19,00	35,00	11,00	11,00	24,00	13,00	47,00	17,00
9-13	27,00	89,00	15,00	30,00	17,00	23,00	33,00	8,00	29,00	11,00

	1	2	3	4	5	6	7	8	9	10
10 - 01	25,00	63,00	79,00	45,00	92,00	93,00	33,00	78,00	86,00	84,00
10 - 02	44,00	52,00	39,00	8,00	28,00	30,00	12,00	35,00	39,00	25,00
10 - 03	20,00	13,00	53,00	16,00	60,00	57,00	10,00	16,00	21,00	25,00
10 - 04	79,00	81,00	20,00	70,00	7,00	7,00	87,00	13,00	10,00	20,00
10 - 05	72,00	76,00	17,00	60,00	32,00	24,00	62,00	26,00	22,00	24,00
10 - 06	50,00	31,00	43,00	2,00	44,00	37,00	3,00	30,00	42,00	57,00
10 - 07	79,00	76,00	64,00	55,00	25,00	17,00	45,00	65,00	25,00	23,00
10 - 08	88,00	41,00	15,00	38,00	10,00	14,00	54,00	28,00	7,00	18,00
10 - 09	84,00	66,00	37,00	32,00	7,00	19,00	17,00	27,00	20,00	19,00
10 - 10	95,00	70,00	16,00	52,00	25,00	16,00	40,00	5,00	89,00	36,00
10 - 11	51,00	40,00	61,00	25,00	75,00	89,00	30,00	86,00	88,00	61,00
10 - 12	74,00	35,00	10,00	18,00	28,00	12,00	25,00	7,00	6,00	35,00
10 - 13	78,00	41,00	20,00	20,00	40,00	26,00	8,00	25,00	87,00	56,00

Cette dernière matrice représente la moyenne des évaluations des 10 répondants. Encore une fois, les colonnes sont les boissons colas et les lignes, les attributs.

	1	2	3	4	5	6	7	8	9	10
1	58,00	48,00	74,00	38,00	73,00	62,00	65,00	80,00	50,00	77,00
2	69,00	46,00	44,00	36,00	32,00	34,00	32,00	35,00	58,00	32,00
3	56,00	21,00	69,00	16,00	58,00	47,00	43,00	71,00	32,00	81,00
4	53,00	73,00	33,00	89,00	36,00	54,00	58,00	26,00	83,00	25,00
5	74,00	68,00	36,00	86,00	47,00	64,00	55,00	25,00	87,00	36,00
6	84,00	42,00	53,00	20,00	72,00	70,00	42,00	71,00	72,00	77,00
7	84,00	73,00	59,00	61,00	64,00	63,00	63,00	73,00	85,00	68,00
8	62,00	60,00	36,00	68,00	31,00	42,00	48,00	45,00	75,00	32,00
9	62,00	79,00	35,00	68,00	34,00	43,00	40,00	39,00	82,00	26,00
10	68,00	79,00	23,00	76,00	47,00	50,00	34,00	36,00	79,00	33,00
11	67,00	60,00	58,00	34,00	75,00	66,00	64,00	77,00	54,00	82,00
12	45,00	61,00	41,00	69,00	31,00	40,00	35,00	32,00	58,00	21,00
13	42,00	49,00	37,00	42,00	44,00	43,00	35,00	47,00	65,00	29,00

# Bibliographie

- [1] K.V.Mardia, J.T.Kent et J.M.Bibby, *Multivariate Analysis*, 1979.
- [2] J.F.Hair, R.Anderson, R.L.Tatham et W.C.Black, *Multivariate data analysis with readings*, third edition, 1992.
- [3] Y.Wind et P.J. Robinson, *Product Positioning, An Application of Multidimensional Scaling*.
- [4] Sibson R., *Studies in the Robustness of Multidimensional Scaling : Procrustes Statistics*, 1978
- [5] J.D. Jobson, *Applied Multivariate Data Analysis, Volume II : Categorical and Multivariate Methods*, 1992
- [6] W.R.Dillon et M.Goldstein, *Multivariate Analysis : Methods and applications*, 1984
- [7] S.S.Schiffman, L.Reynolds et F.W. Young, *Introduction to multidimensionnal scaling : Theory Methods and Applications*, 1981
- [8] Y.Takane, F.W.Young, J.DeLeeuw, *Nonmetric individual differences multidimensional scaling : an alternating least squares method with optimal scaling features. Psychometrika* 42(1977) :7-67.(ALSCAL)