

Analyse factorielle de données de catégorisation

Application aux données sensorielles

Soutenance de thèse de
Marine Cadoret

Préparée au
Laboratoire de mathématiques appliquées
Agrocampus Ouest

30 août 2010



Sommaire

- 1 Introduction
- 2 Analyse factorielle des données de catégorisation
- 3 Recueil et traitement des données incomplètes
- 4 Éléments de validité
- 5 Deux extensions de la catégorisation
- 6 Conclusions et perspectives

Sommaire

- 1 Introduction
- 2 Analyse factorielle des données de catégorisation
- 3 Recueil et traitement des données incomplètes
- 4 Éléments de validité
- 5 Deux extensions de la catégorisation
- 6 Conclusions et perspectives

La catégorisation

- Méthode de recueil de données simple
- Issue du domaine de la psychologie
- Apparition en analyse sensorielle à la fin des années 80 (Lawless, 1989)
- Utilisée sur tous types de produits, notamment :
 - Fromages : Lawless *et al.*, 1995
 - Eaux : Falahee et MacRae, 1995
 - Bières : Chollet et Valentin, 2001
 - Pièces automobiles : Faye *et al.*, 2006
 - Vins : Parr *et al.*, 2007
 - Sauces barbecues : Blancher *et al.*, 2008

La tâche de catégorisation

- Approche holistique (globale)
- Il est demandé de faire des groupes d'objets en fonction de leurs ressemblances :
 - 2 objets sont dans un même groupe s'ils sont perçus comme semblables
 - 2 objets sont dans 2 groupes différents s'ils sont perçus comme différents
- Aucun critère imposé aux sujets
- Evaluation simultanée
- Le nombre de groupe doit être compris entre 2 et $I - 1$ (avec I le nombre d'objets)
- Une tâche de verbalisation peut aussi être demandée pour décrire les groupes fournis (catégorisation "qualifiée")

Données

98 sujets - 12 parfums féminins :



Angel



Lolita
Lempicka



L'Instant



Cinéma



J'adore
(ET)



J'adore
(EP)



Shalimar



Aromatics
Elixir



Coco
Mademoiselle



Chanel n°5



Pure
Poison



Pleasures

Produits présentés anonymement

Données

98 sujets - 12 parfums féminins :



482



193



240



548



493



935



968



357



173



674



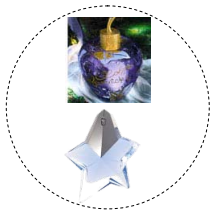
475



117

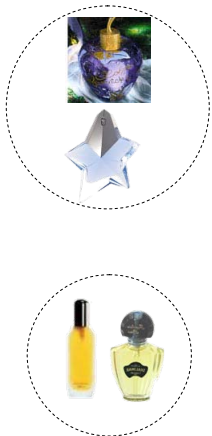
Produits présentés anonymement

Exemple de catégorisation : sujet 1



Exemple de catégorisation : sujet 1

sucré, vanille

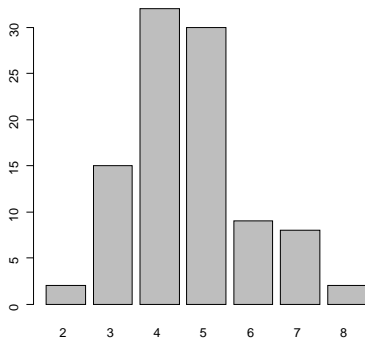


entêtant

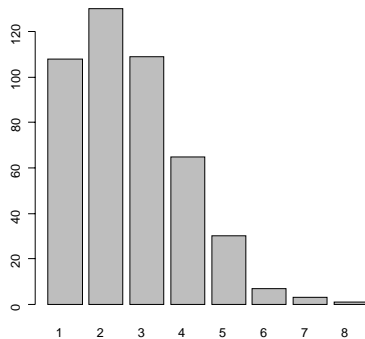


fleuri, vert

Première description : les groupes

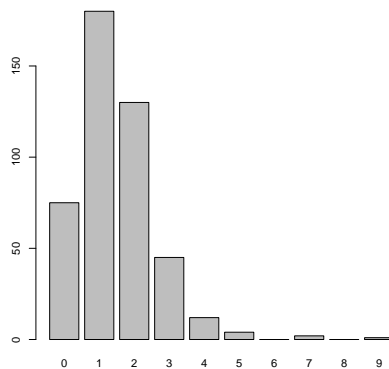


Nombre de groupes formés



Nombre de parfums par groupe

Première description : les mots



Nombre de mots utilisés par groupe

Première description : matrice de cooccurrences (réordonnée)

	J'adore (ET)	J'adore (EP)	Pleasures	Coco Melle	Cinéma	Pure Poison	L'instant	Lolita Lemp.	Angel	Chanel n°5	Arom. Elixir	Shalimar
J'adore (ET)	98	56	48	38	24	28	22	18	12	14	7	7
J'adore (EP)	56	98	38	28	23	29	28	18	11	12	12	6
Pleasures	48	38	98	28	22	30	23	18	11	14	11	6
Coco Melle	38	28	28	98	30	33	20	21	10	11	12	9
Cinéma	24	23	22	30	98	28	26	42	18	9	8	10
Pure Poison	28	29	30	33	28	98	25	18	10	21	12	11
L'instant	22	28	23	20	26	25	98	22	14	10	13	13
Lolita Lemp.	18	18	18	21	42	18	22	98	36	8	6	9
Angel	12	11	11	10	18	10	14	36	98	15	27	21
Chanel n°5	14	12	14	11	9	21	10	8	15	98	51	30
Arom. Elixir	7	12	11	12	8	12	13	6	27	51	98	42
Shalimar	7	6	6	9	10	11	13	9	21	30	42	98

Première description : matrice de cooccurrences (réordonnée)

	J'adore (ET)	J'adore (EP)	Pleasures	Coco Melle	Cinéma	Pure Poison	L'instant	Lolita Lemp.	Angel	Chanel n°5	Arom. Elixir	Shalimar
J'adore (ET)	98	56	48	38	24	28	22	18	12	14	7	7
J'adore (EP)	56	98	38	28	23	29	28	18	11	12	12	6
Pleasures	48	38	98	28	22	30	23	18	11	14	11	6
Coco Melle	38	28	28	98	30	33	20	21	10	11	12	9
Cinéma	24	23	22	30	98	28	26	42	18	9	8	10
Pure Poison	28	29	30	33	28	98	25	18	10	21	12	11
L'instant	22	28	23	20	26	25	98	22	14	10	13	13
Lolita Lemp.	18	18	18	21	42	18	22	98	36	8	6	9
Angel	12	11	11	10	18	10	14	36	98	15	27	21
Chanel n°5	14	12	14	11	9	21	10	8	15	98	51	30
Arom. Elixir	7	12	11	12	8	12	13	6	27	51	98	42
Shalimar	7	6	6	9	10	11	13	9	21	30	42	98
<i>Seul</i>	2	2	8	8	5	7	9	7	13	17	6	24

Sommaire

- 1 Introduction
- 2 Analyse factorielle des données de catégorisation
- 3 Recueil et traitement des données incomplètes
- 4 Éléments de validité
- 5 Deux extensions de la catégorisation
- 6 Conclusions et perspectives

Bref historique : représentation des objets

- Matrice de cooccurrences

	1	i'	l
1		.	
i	.	$c(i, i')$	
l			

Table: $c(i, i')$ correspond au nombre de sujets ayant placés les objets i et i' dans un même groupe

- Matrice de cooccurrences transformée en matrice de distances : $d(i, i') = J - c(i, i')$
- Représentation des objets obtenue par MDS non-métrique (Lawless, 1989)

Bref historique : représentation des mots

- Matrice de fréquences

	1	..	m	..	M
1		
.		
i		..	$f(i, m)$..	
.		
l		

Table: $f(i, m)$ correspond au nombre de sujets ayant utilisé le mot m pour décrire un groupe contenant l'objet i

- Coefficient de corrélation entre chaque dimension et chaque vecteur de fréquences (Faye *et al.*, 2004, 2006)

Bref historique : représentation des sujets

- DISTATIS (Abdi *et al.*, 2007) basée sur MDS et STATIS
- Matrice de cooccurrences individuelle du sujet j

	1	i'	l
1		.	
i	.	$c_j(i, i')$	
l			

Table: $c_j(i, i') = 1$ si les objets i et i' dans le même groupe par le sujet j et 0 sinon

- Matrice de cooccurrences individuelle transformée en matrice de distances : $d_j(i, i') = 1 - c_j(i, i')$

Autre codage des données

		Sujet 1	...	Sujet j	...	Sujet J
1	Angel	G1		G1		G2
	Arom. El.	G3		G5		G1
.		.		.		.
.		.		.		.
i	Cinéma	G1		G2		G3
	Loli. Lemp.	G1		G2		G3
	L'instant	G2		G2		G3
.		.		.		.
.		.		.		.
l	Pleasures	G1		G2		G4
	Shalimar	G2		G3		G1

- Un sujet = Une variable qualitative

Autre codage des données

		Sujet 1	...	Sujet j	...	Sujet J
1	Angel	vert		fleuri, vert		floral
	Arom. El.	amer		amer		fort, acide
.		.		.		.
.		.		.		.
i	Cinéma	vert		doux		G3
	Loli. Lemp.	vert		doux		G3
	L'instant	sucré		doux		G3
.		.		.		.
.		.		.		.
l	Pleasures	vert		doux		frais
	Shalimar	sucré		vanille		fort, acide

- Un sujet = Une variable qualitative
- Un label = La séquence de mots utilisée pour décrire le groupe

Autre codage des données

		Sujet 1	...	Sujet j	...	Sujet J
1	Angel	vert		fleuri, vert		floral
	Arom. El.	amer		amer		fort, acide
.
.
i	Cinéma	vert		doux		G3
	Loli. Lemp.	vert		doux		G3
	L'instant	sucré		doux		G3
.
.
l	Pleasures	vert		doux		frais
	Shalimar	sucré		vanille		fort, acide

- Un sujet = Une variable qualitative
- Un label = La séquence de mots utilisée pour décrire le groupe

⇒ Traitement par Analyse des Correspondances Multiples (ACM; Lebart *et al.*, 1977)

Tableau disjonctif complet

	Sujet 1		Sujet j		Sujet J
	1		k		K
1					
i	1 0 0		x_{ik}		
l					
	l_1		l_k		l_K

- $x_{ik} = 1$ si l'objet i appartient au groupe k (du sujet j) et 0 sinon
- l_k : nombre d'objets appartenant au groupe k (du sujet j)

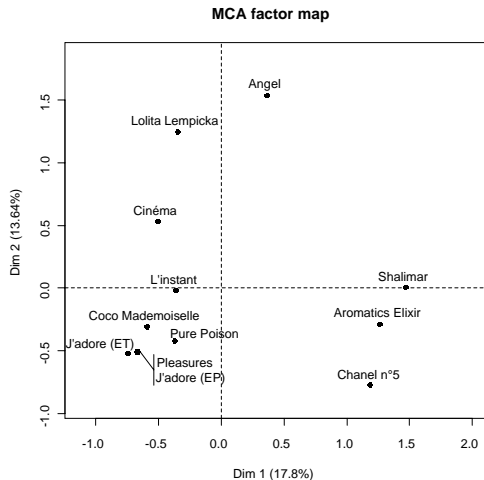
Représentation des objets en ACM

Distance entre 2 objets i et l :

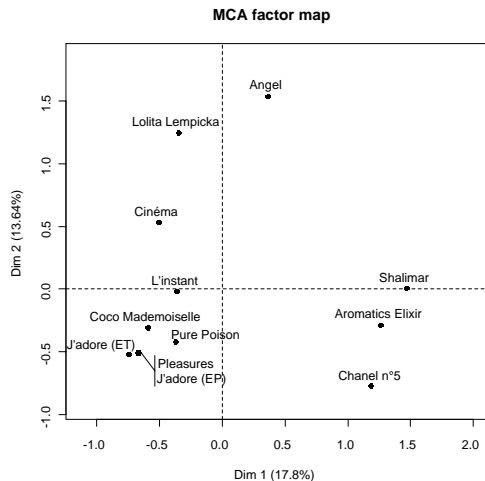
$$d^2(i, l) = \frac{1}{J} \sum_k \frac{l}{l_k} (x_{ik} - x_{lk})^2$$

- 2 objets sont d'autant plus proches qu'ils ont été placés dans un même groupe par de nombreux sujets
- 2 objets sont d'autant plus éloignés qu'ils ont été placés dans des groupes différents par de nombreux sujets

Représentation des parfums



Représentation des parfums

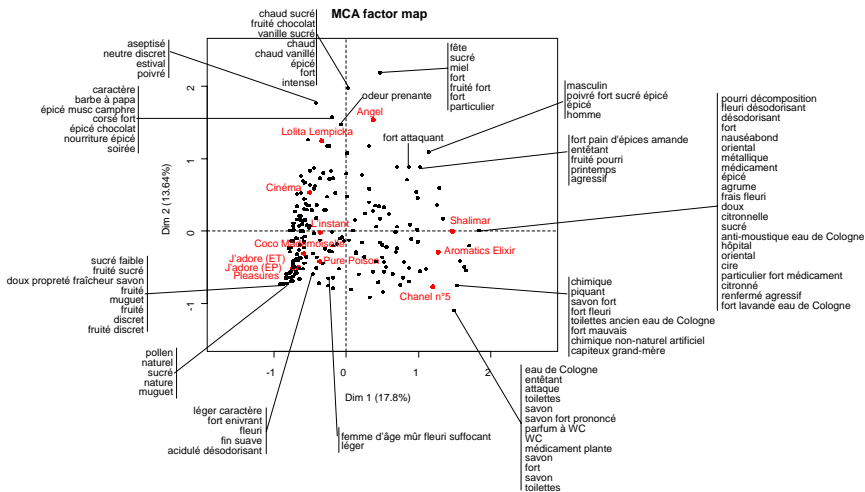


	Arom. El.	Shalimar	J'adore (EP)	J'adore (ET)
Arom. El.	98	42	12	7
Shalimar	42	98	6	7
J'adore (EP)	12	6	98	56
J'adore (ET)	7	7	56	98
<i>Seul</i>	6	24	2	2

Représentation des mots en ACM

- Représentation individuelle des mots
- Distance entre deux mots k et h :
 - $d^2(k, h) = \sum_i l \left(\frac{x_{ik}}{l_k} - \frac{x_{ih}}{l_h} \right)^2$
 - Deux mots seront d'autant plus proches qu'ils sont associés aux mêmes objets
- Coordonnée du mot k sur l'axe de rang s :
 - $G_s(k) = \frac{1}{\sqrt{\lambda_s}} \sum_{i=1}^l \frac{x_{ik}}{l_k} F_s(i)$
 - Double propriété barycentrique : un mot sera du côté des objets auxquels il a été associé et un objet sera du côté des mots auxquels il a été associé

Représentation des mots et des parfums

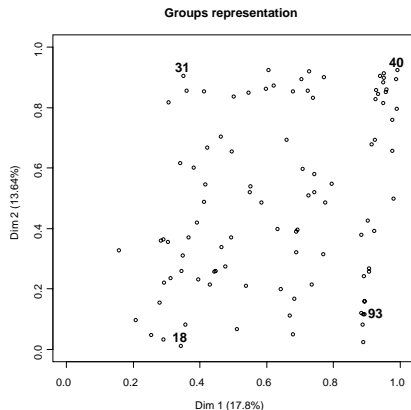


Représentation des sujets en ACM

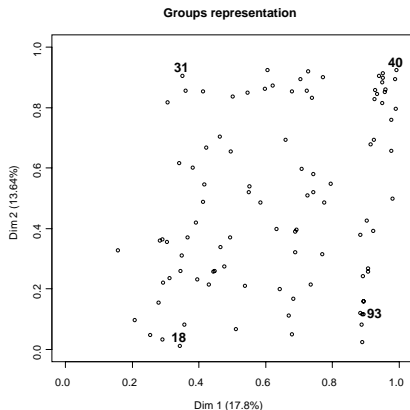
- Pas de représentation usuelle des variables (sujets) elles-mêmes en ACM
- Solutions possibles pour la coordonnée du sujet j sur l'axe de rang s :
 - $\frac{1}{\sqrt{K_j-1}}\eta^2(F_s, j)$, avec K_j le nombre de groupes fournis par le sujet j et F_s le vecteur de coordonnées des objets sur l'axe de rang s (Escofier, 1979)
 - $\eta(F_s, j)$ (Cazes, 1982)
- ACM \Leftrightarrow Recherche des z_s tels que : $\max \sum_j \eta^2(z_s, j)$

\Rightarrow Choix du carré du rapport de corrélation : $\eta^2(F_s, j)$

Représentation des sujets

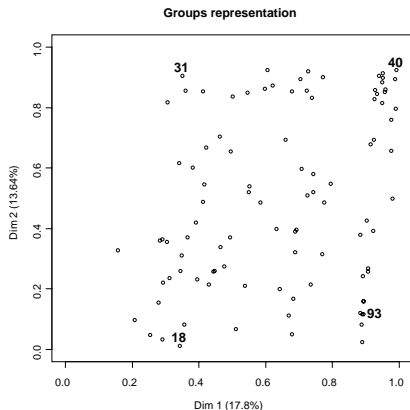


Représentation des sujets



	18	31	40	93
Aromatics Elixir	2	2	5	3
Chanel n°5	3	4	5	3
Shalimar	2	1	4	4
Coco Mademoiselle	3	1	2	2
J'adore (EP)	1	1	1	1
J'adore (ET)	3	1	1	2
L'instant	2	1	2	1
Pleasures	3	1	1	1
Pure Poison	1	2	2	2
Cinéma	3	3	3	2
Lolita Lempicka	1	3	3	2
Angel	3	5	6	1

Représentation des sujets



	18	31	40	93
Aromatics Elixir	2	2	5	3
Chanel n°5	3	4	5	3
Shalimar	2	1	4	4
Coco Mademoiselle	3	1	2	2
J'adore (EP)	1	1	1	1
J'adore (ET)	3	1	1	2
L'instant	2	1	2	1
Pleasures	3	1	1	1
Pure Poison	1	2	2	2
Cinéma	3	3	3	2
Lolita Lempicka	1	3	3	2
Angel	3	5	6	1

Sommaire

- 1 Introduction
- 2 Analyse factorielle des données de catégorisation
- 3 Recueil et traitement des données incomplètes**
- 4 Éléments de validité
- 5 Deux extensions de la catégorisation
- 6 Conclusions et perspectives

Problèmes

Lorsque il y a trop d'objets à évaluer : choix d'un sous-ensemble d'objets pour chaque sujet

- Comment traiter les données incomplètes?
- Comment choisir les objets évalués par chaque sujet?

Traitement des données incomplètes

	Sujet 1	Sujet 2	Sujet 3	Sujet 4	Sujet 5
Angel	S1.miss	4	1	5	S5.miss
Aromatics Elixir	3	3	S3.miss	2	S5.miss
Chanel n° 5	4	3	4	1	3
Cinéma	2	S2.miss	6	4	2
Coco Mademoiselle	S1.miss	5	2	4	3
J'adore (EP)	1	6	2	3	3
J'adore (ET)	2	6	2	S4.miss	3
L'Instant	1	S2.miss	6	2	4
Lolita Lempicka	1	5	1	5	2
Pleasures	3	4	6	S4.miss	4
Pure Poison	1	1	S3.miss	4	4
Shalimar	2	2	3	2	1

Pour chaque sujet : objets non-présentés = groupe additionnel

Choix d'un sous-ensemble d'objets (1)

Chaque sujet évalue p objets parmi I :

- aléatoirement :
 - chaque objet i est évalué un nombre différent de fois : r_i
 - chaque paire d'objets (i, l) est évaluée un nombre différent de fois : λ_{il}

Choix d'un sous-ensemble d'objets (1)

Chaque sujet évalue p objets parmi I :

- aléatoirement :
 - chaque objet i est évalué un nombre différent de fois : r_i
 - chaque paire d'objets (i, l) est évaluée un nombre différent de fois : λ_{il}
- selon un plan \rightarrow plan en blocs incomplets équilibrés (BIE) :
 - chaque objet est évalué un même nombre de fois : r
 - chaque paire d'objets est évaluée un même nombre de fois : λ

Choix d'un sous-ensemble d'objets (2)

Distance entre les 2 objets i et l décomposée selon :

- part due aux groupes réellement constitués
- part due aux groupes d'objets manquants

Choix d'un sous-ensemble d'objets (2)

Distance entre les 2 objets i et l décomposée selon :

- part due aux groupes réellement constitués
- part due aux groupes d'objets manquants

Plan aléatoire : $d^2(i, l) = d_{k'}^2(i, l) + \frac{1}{j} \frac{l}{I-p} ((r_i - \lambda_{il}) + (r_l - \lambda_{il}))$

Plan en BIE : $d^2(i, l) = d_{k'}^2(i, l) + \frac{1}{j} \frac{l}{I-p} 2(r - \lambda)$

Choix d'un sous-ensemble d'objets (2)

Distance entre les 2 objets i et l décomposée selon :

- part due aux groupes réellement constitués
- part due aux groupes d'objets manquants

Plan aléatoire : $d^2(i, l) = d_{k'}^2(i, l) + \frac{1}{j} \frac{l}{I-p} ((r_i - \lambda_{il}) + (r_l - \lambda_{il}))$

Plan en BIE : $d^2(i, l) = d_{k'}^2(i, l) + \frac{1}{j} \frac{l}{I-p} 2(r - \lambda)$

Dans quelle mesure un plan en BIE sera “meilleur” qu'un plan aléatoire?

Choix d'un sous-ensemble d'objets (2)

Distance entre les 2 objets i et l décomposée selon :

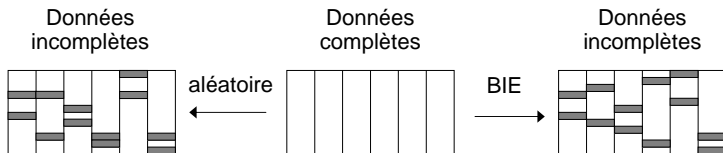
- part due aux groupes réellement constitués
- part due aux groupes d'objets manquants

Plan aléatoire : $d^2(i, l) = d_{k'}^2(i, l) + \frac{1}{j} \frac{l}{I-p} ((r_i - \lambda_{il}) + (r_l - \lambda_{il}))$

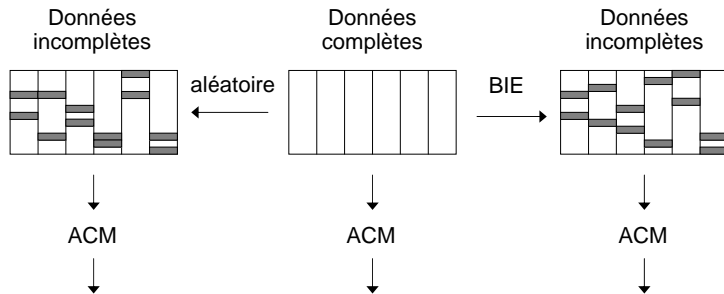
Plan en BIE : $d^2(i, l) = d_{k'}^2(i, l) + \frac{1}{j} \frac{l}{I-p} 2(r - \lambda)$

Dans quelle mesure un plan en BIE sera "meilleur" qu'un plan aléatoire? \Rightarrow Simulation de données incomplètes

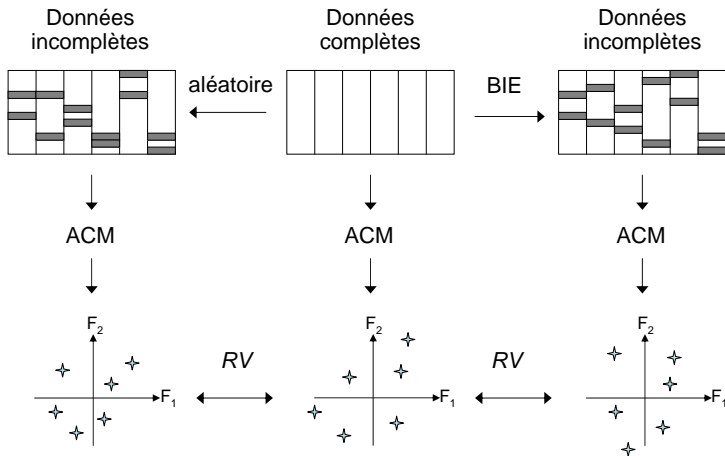
BIE vs. aléatoire (1)



BIE vs. aléatoire (1)



BIE vs. aléatoire (1)



BIE vs. aléatoire (2)

Simulations de différentes situations à partir du jeu de données “parfums” :

	10 parfums (17% mqt)		8 parfums (33% mqt)		6 parfums (50% mqt)	
	aléatoire	BIE	aléatoire	BIE	aléatoire	BIE
30 sujets						
60 sujets						
98 sujets						

Table: Valeurs moyennes des coefficients $RV(\text{complet}, \text{incomplet})$ pour 100 simulations.

BIE vs. aléatoire (2)

Simulations de différentes situations à partir du jeu de données “parfums” :

	10 parfums (17% mqt)		8 parfums (33% mqt)		6 parfums (50% mqt)	
	aléatoire	BIE	aléatoire	BIE	aléatoire	BIE
30 sujets						
60 sujets						
98 sujets	0.97		0.88		0.721	

Table: Valeurs moyennes des coefficients $RV(\text{complet}, \text{incomplet})$ pour 100 simulations.

BIE vs. aléatoire (2)

Simulations de différentes situations à partir du jeu de données “parfums” :

	10 parfums (17% mqt)		8 parfums (33% mqt)		6 parfums (50% mqt)	
	aléatoire	BIE	aléatoire	BIE	aléatoire	BIE
30 sujets	0.76		0.652		0.471	
60 sujets	0.92		0.77		0.621	
98 sujets	0.97		0.88		0.721	

Table: Valeurs moyennes des coefficients $RV(\text{complet}, \text{incomplet})$ pour 100 simulations.

BIE vs. aléatoire (2)

Simulations de différentes situations à partir du jeu de données “parfums” :

	10 parfums (17% mqt)		8 parfums (33% mqt)		6 parfums (50% mqt)	
	aléatoire	BIE	aléatoire	BIE	aléatoire	BIE
30 sujets	0.76	0.781	0.652	0.746	0.471	0.579
60 sujets	0.92	0.942	0.77	0.869	0.621	0.758
98 sujets	0.97	0.98	0.88	0.94	0.721	0.855

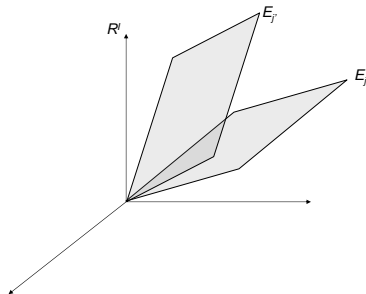
Table: Valeurs moyennes des coefficients $RV(\text{complet}, \text{incomplet})$ pour 100 simulations.

Sommaire

- 1 Introduction
- 2 Analyse factorielle des données de catégorisation
- 3 Recueil et traitement des données incomplètes
- 4 Éléments de validité**
- 5 Deux extensions de la catégorisation
- 6 Conclusions et perspectives

Données de catégorisation

- Nombre de colonnes \gg nombre de lignes
- ACM : une variable = un sous-espace \rightarrow production automatique de fortes dimensions communes



- Validité de la première dimension (consensus)?

Consensus entre les sujets

- Première valeur propre $\Rightarrow \lambda_1 = \frac{1}{J} \sum_j \eta^2(F_{1,j})$

Consensus entre les sujets

- Première valeur propre $\Rightarrow \lambda_1 = \frac{1}{J} \sum_j \eta^2(F_{1,j})$
- Hypothèses :
 - Absence de consensus (catégorisations indépendantes)
 - Existence d'au moins une dimension "consensuelle"

Consensus entre les sujets

- Première valeur propre $\Rightarrow \lambda_1 = \frac{1}{J} \sum_j \eta^2(F_{1,j})$
- Hypothèses :
 - Absence de consensus (catégorisations indépendantes)
 - Existence d'au moins une dimension "consensuelle"
- Evolution de l'indicateur pour des données indépendantes simulées

Consensus entre les sujets

- Première valeur propre $\Rightarrow \lambda_1 = \frac{1}{j} \sum_j \eta^2(F_{1,j})$
- Hypothèses :
 - Absence de consensus (catégorisations indépendantes)
 - Existence d'au moins une dimension "consensuelle"
- Evolution de l'indicateur pour des données indépendantes simulées

		Nombre de sujets			
		10	20	50	100
Nombre d'objets	10				
	20				
	50				
	100				

Table: Valeurs moyennes des λ_1 pour 100 simulations.

Consensus entre les sujets

- Première valeur propre $\Rightarrow \lambda_1 = \frac{1}{j} \sum_j \eta^2(F_{1,j})$
- Hypothèses :
 - Absence de consensus (catégorisations indépendantes)
 - Existence d'au moins une dimension "consensuelle"
- Evolution de l'indicateur pour des données indépendantes simulées

		Nombre de sujets			
		10	20	50	100
Nombre d'objets	10	0.6957	0.6246	0.5578	0.5242
	20	0.471	0.3892	0.3196	0.2863
	50	0.3062	0.2288	0.1678	0.14
	100	0.2364	0.164	0.1099	0.0864

Table: Valeurs moyennes des λ_1 pour 100 simulations.

Consensus entre les sujets

- Première valeur propre $\Rightarrow \lambda_1 = \frac{1}{j} \sum_j \eta^2(F_{1,j})$
- Hypothèses :
 - Absence de consensus (catégorisations indépendantes)
 - Existence d'au moins une dimension "consensuelle"
- Evolution de l'indicateur pour des données indépendantes simulées

		Nombre de sujets			
		10	20	50	100
Nombre d'objets	10	0.6957	0.6246	0.5578	0.5242
	20	0.471	0.3892	0.3196	0.2863
	50	0.3062	0.2288	0.1678	0.14
	100	0.2364	0.164	0.1099	0.0864

Table: Valeurs moyennes des λ_1 pour 100 simulations.

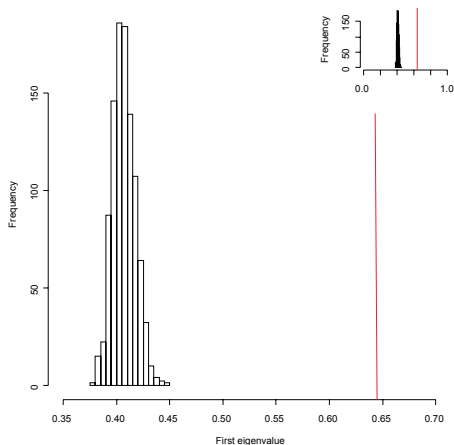
- Nécessité d'un test sur la première valeur propre
 - H_0 : absence de consensus (catégorisations indépendantes)
 - H_1 : existence d'au moins une dimension "consensuelle"

Significativité de l'indicateur pour un tableau donné

Associer une probabilité critique à la première valeur propre de l'ACM :

- Faire un grand nombre de fois :
 - 1 Permutations indépendantes des lignes pour chaque colonne
 - 2 Calcul de la première valeur propre associée au tableau permuté
- Distribution de la première valeur propre sous H_0
- Calcul de la probabilité critique

Significativité de λ_1 pour les données “parfums”

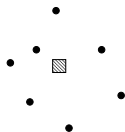


Ellipses de confiance autour des produits

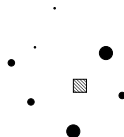
- Ellipses de confiance autour des produits
- Principe :
 - Ré-échantillonnage des sujets (colonnes) : création de jurys virtuels
 - Utilisation de la propriété barycentrique de l'ACM :

$$F_s(i) = \frac{1}{\sqrt{\lambda_s}} \sum_{k \in K} \frac{x_{ik}}{J} G_s(k)$$

Véritable produit



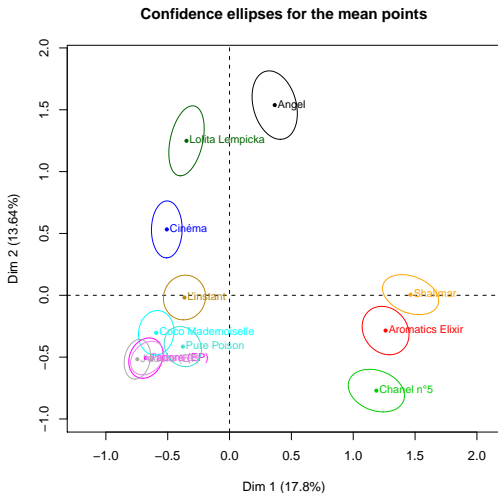
Produit ré-échantillonné



Ellipse



Ellipses de confiance autour des produits

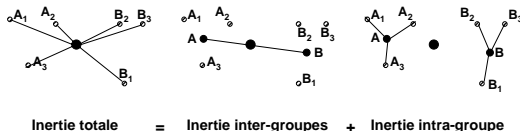


Recouvrement des ellipses

- Le non-chevauchement des ellipses est-il de même importance que celui susceptible d'être obtenu par hasard?

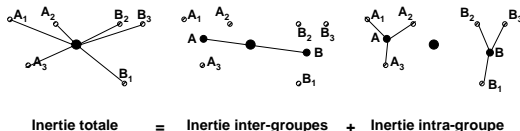
Recouvrement des ellipses

- Le non-chevauchement des ellipses est-il de même importance que celui susceptible d'être obtenu par hasard?
- Recherche d'un indicateur de recouvrement :
 - Ellipses représentent les produits ré-échantillonnés
 - Recouvrement des "zones" représentées par les produits ré-échantillonnés
 - Théorème de Huygens sur le nuage des produits ré-échantillonnés :



Recouvrement des ellipses

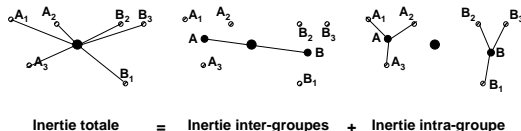
- Le non-chevauchement des ellipses est-il de même importance que celui susceptible d'être obtenu par hasard?
- Recherche d'un indicateur de recouvrement :
 - Ellipses représentent les produits ré-échantillonnés
 - Recouvrement des "zones" représentées par les produits ré-échantillonnés
 - Théorème de Huygens sur le nuage des produits ré-échantillonnés :



- Indicateur de recouvrement : $0 \leq \frac{\text{Inertie Inter-groupes}}{\text{Inertie totale}} \leq 1$

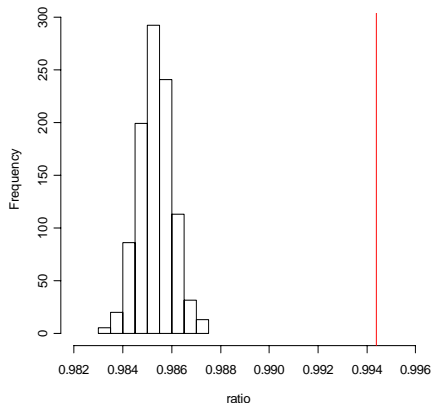
Recouvrement des ellipses

- Le non-chevauchement des ellipses est-il de même importance que celui susceptible d'être obtenu par hasard?
- Recherche d'un indicateur de recouvrement :
 - Ellipses représentent les produits ré-échantillonnés
 - Recouvrement des "zones" représentées par les produits ré-échantillonnés
 - Théorème de Huygens sur le nuage des produits ré-échantillonnés :



- Indicateur de recouvrement : $0 \leq \frac{\text{Inertie Inter-groupes}}{\text{Inertie totale}} \leq 1$
- Même procédure de permutation que pour la première valeur propre

Significativité de l'indicateur de recouvrement pour les données "parfums"

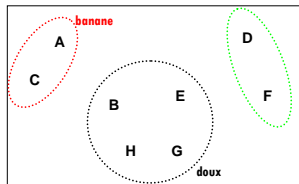


Sommaire

- 1 Introduction
- 2 Analyse factorielle des données de catégorisation
- 3 Recueil et traitement des données incomplètes
- 4 Éléments de validité
- 5 Deux extensions de la catégorisation**
- 6 Conclusions et perspectives

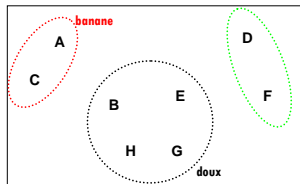
Deux extensions de la catégorisation

- Napping[®] catégorisé (Pagès *et al.*, 2010)

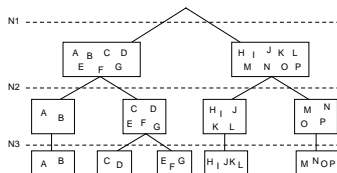


Deux extensions de la catégorisation

- Napping[®] catégorisé (Pagès *et al.*, 2010)



- Tri hiérarchique (Egoroff, 2005)



Sommaire

- 1 Introduction
- 2 Analyse factorielle des données de catégorisation
- 3 Recueil et traitement des données incomplètes
- 4 Éléments de validité
- 5 Deux extensions de la catégorisation
- 6 Conclusions et perspectives

Conclusions

- Mise au point de méthodologies statistiques répondant à des problématiques sensorielles
- Méthodologies qui contiennent :
 - Indicateurs unidimensionnels : description élémentaire des données
 - Approches factorielles : représentation des objets, des mots, des sujets, etc. liées entre elles
 - Éléments de validité
- Développement pour les utilisateurs : mise à disposition de fonctions, données et aides dans le package SensoMineR

Perspectives

- Construction des ellipses de confiance :
 - Ellipses optimistes
 - Nouvelle stratégie à envisager dans leur construction : bootstrap total?

Perspectives

- Construction des ellipses de confiance :
 - Ellipses optimistes
 - Nouvelle stratégie à envisager dans leur construction : bootstrap total?
- Envisager les données incomplètes pour :
 - Tri hiérarchique : un groupe additionnel quel que soit le niveau?
 - Mapping[®] catégorisé : obtenir des données qualitatives et quantitatives qui soient en accord

Perspectives

- Construction des ellipses de confiance :
 - Ellipses optimistes
 - Nouvelle stratégie à envisager dans leur construction : bootstrap total?
- Envisager les données incomplètes pour :
 - Tri hiérarchique : un groupe additionnel quel que soit le niveau?
 - Napping[®] catégorisé : obtenir des données qualitatives et quantitatives qui soient en accord
- Visualisation :
 - Graphique dynamique avec R : dynGraph (Lê et Durand, 2010)
 - Adaptation de logiciels dédiés aux données textuelles : Qnomis (Kerbaol *et al.*, 1997)

Evolution de la première valeur propre pour un nombre d'objets fixé

