



Intégration des perceptions et des préférences utilisateur en conception de produit

L'analyse conjointe

Jean-François PETIOT

Jean-Francois.Petiot@ec-nantes.fr

Resp. Option Science et Musique

Resp. Mastere [Marketing Design et Création \(Coll. Audencia\)](#)

Ecole Centrale Nantes
LS2N, UMR CNRS 6004



Objectifs du cours

- ◆ Intégrer les **perceptions** utilisateur en conception de produit
- ◆ Identifier les caractéristiques du produits sur lesquels repose la **satisfaction client**
- ◆ Modéliser les relations entre la **préférence** utilisateur et les caractéristiques produit (**attributs de design**)
 - L'analyse conjointe
 - Les cartographies de préférences
- ◆ Comprendre la **sémantique produit** en fonction des choix de conception (attributs de design)
- ◆ Optimiser/segmenter l'offre produit

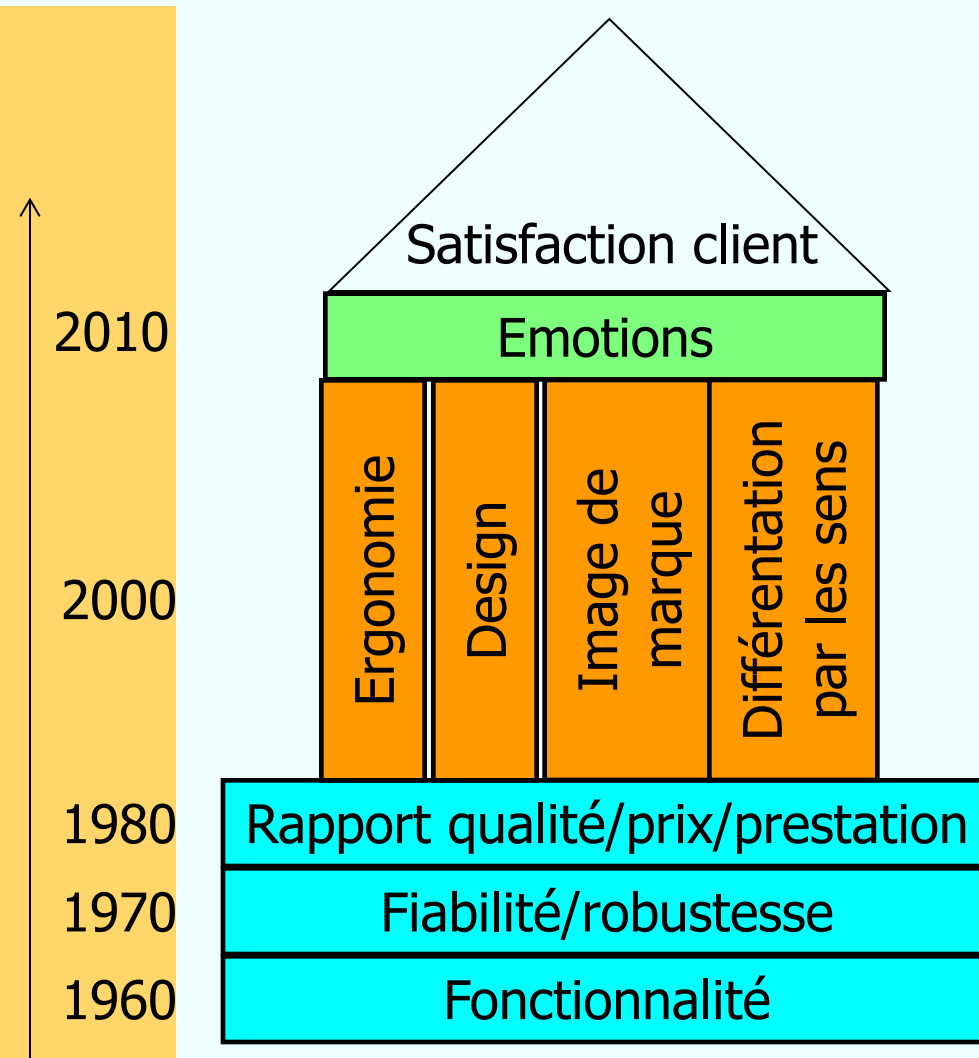


Plan (4h cours+4h TD)

- ◆ La sémantique produit
 - Design et sémiotique – exemples
 - Le différentiel sémantique
 - Perceptions utilisateur - Qualité perçue – exemples
- ◆ L'analyse conjointe (AC)
 - Petit exemple
 - Méthodologie de l'AC
 - Rappel : le modèle linéaire et l'ANOVA
 - Estimation du modèle – interprétation des résultats
- ◆ Classification
 - La Classification ascendante hiérarchique (CAH)
 - Exemples
- ◆ Les cartographies de préférence



Diversité des attentes clients



- Les dimensions sur lesquelles repose la satisfaction client sont de + en + diverses
- Les aspects sensoriels, symboliques, émotionnels, sont prépondérants

⇒ Motivent la décision d'achat.



Exemples



Ambiance lumineuse

**Toucher des surfaces
de planches de bord**

Confort acoustique



Agrément de conduite

Confort thermique

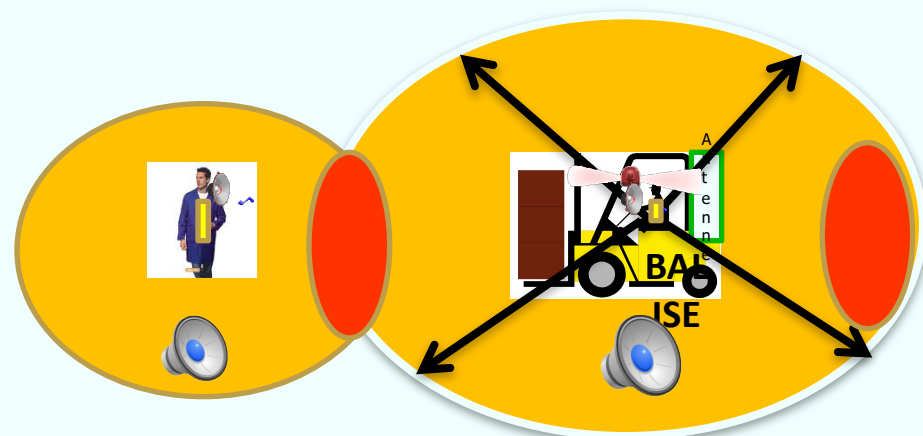




Exemples (2)



Revêtement de surface



Design sonore



Bruit de portières [PSA]



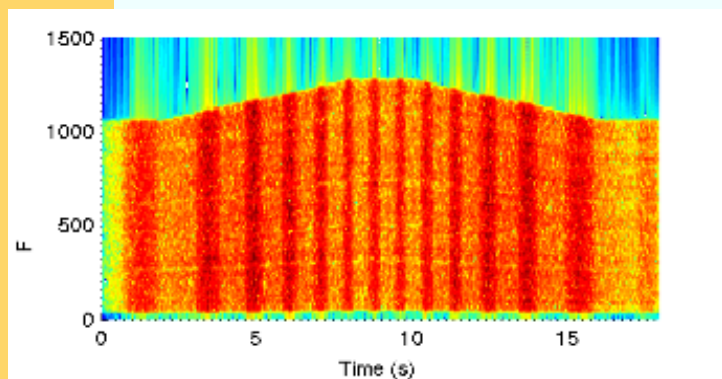
Exemples (3)

- Sécurité des piétons
- Aide à la conduite
- Image du véhicule

Sonification d'un véhicule électrique



- Quel(s) son(s) proposer ?
 - Analyse des contraintes
 - Propositions
 - Optimisation



Experts: Sound S4

monotonous
dubstep-like
high-frequency
sharp
poor
train-like
high-pitched
noisy
modulated
annoying
spooky
clumsy
quality
tremolo
broken
band-pass
tonal
low-frequency
rough
fast
house-like
filtered
synthetic
deep
futuristic
80's
keyboard-like
nerve-wracking
haunted
lousy



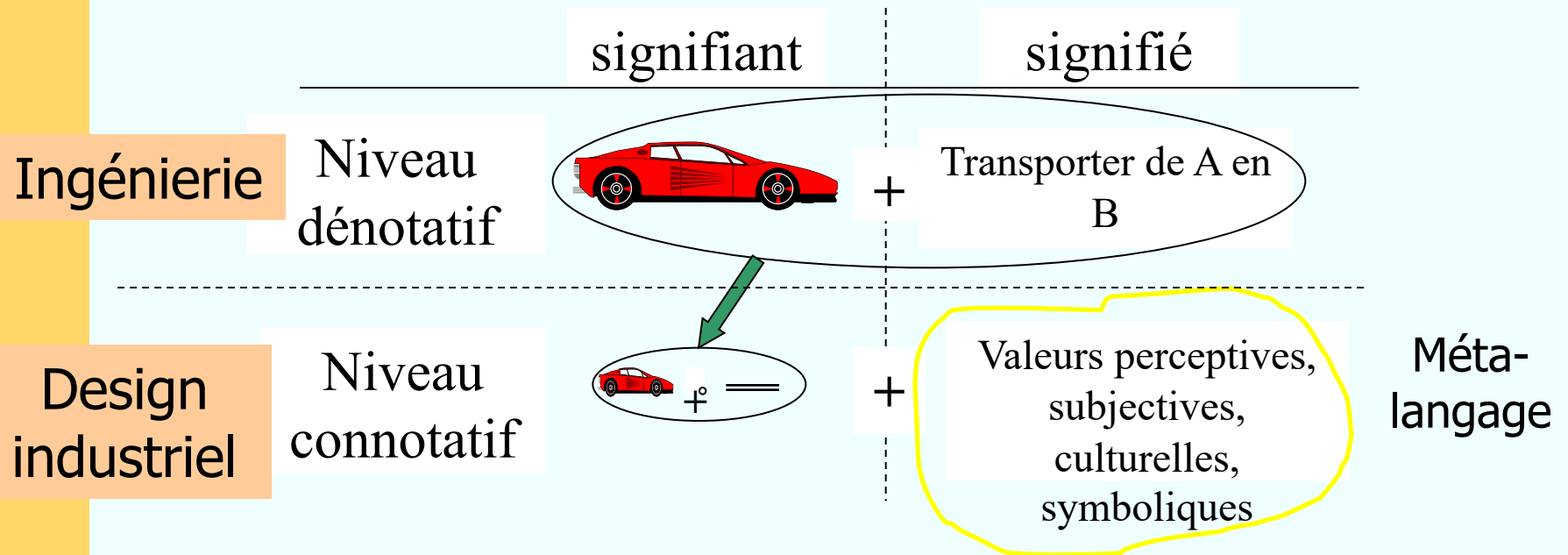


Design et sémiotique (Saussure, Peirce)

étude des signes

Signe = signifiant + signifié

produit = moyen de communication



Design $\neq$ art : l'intention à communiquer est explicite en design industriel

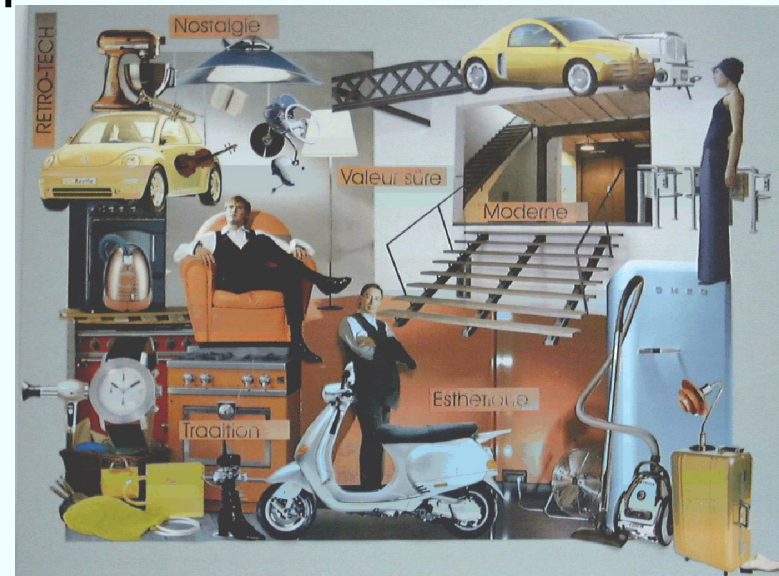


◆ Définitions

- 1) « activité créatrice qui consiste à déterminer les propriétés formelles des objets que l'on veut produire industriellement »
 - propriétés formelles : caractéristiques extérieures, mais aussi relations structurelles qui font de l'objet un système cohérent
- 2) « activité qui consiste à créer, selon des paramètres économiques, techniques et esthétiques, des produits, objets ou systèmes qui seront ensuite fabriqués et commercialisés »

◆ 3 types de valeurs du produit

- Valeurs techniques
 - Cahier des charges technique
- Valeurs d'usage
 - Liés à l'utilisation (ergonomie)
- Valeurs de communication
 - Fonctions d'estime





Valeurs de communication ?



Rétro, classique,

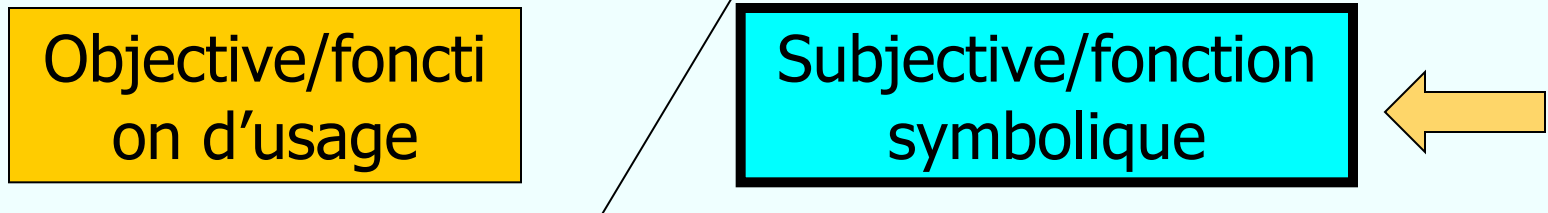
High tech,
épuré, ...



Sémantique produit

“The study of the symbolic qualities of man-made forms in the context of their use, and application of this knowledge to industrial design” [Krippendorff]

- ◆ Une dimension particulière du besoin



- ◆ Nécessité d'une méthodologie pour déterminer les fonctions symboliques (esthétique, style, communications)
- ◆ Évaluer la satisfaction de ces fonctions durant le processus de conception
 - Développer les produits de manière rationnelle



Le design brief

- ◆ « cahier des charges », destiné au dialogue designer/entreprise, décrivant le projet
- ◆ Éléments à inclure
 - Objectifs et résultats attendus
 - Public cible
 - Budget et livrables
 - Contraintes (normes, brevet, marque, ...)
 - Description de la demande
- ◆ Description de la demande
 - décrire la sémantique produit, par des termes ou des images,
 - décrire le produit et son univers d'un point de vue connotatif
 - Utiliser des planches de tendance (mood board)



Exemple de design brief

- ◆ Design sonore automobile
L'univers DS (PSA, 2013)



Avant-garde
Charismatique Non-conventionnel
À l'opposé du mainstream

Technologique
Digital Connecté
En phase avec le monde numérique

Raffinement
Matières inédites & uniques
ciselé

Hyper confort
Agréable Intuitif Profondeur

Les mots-clés DS



◆ Les planches de tendances. Exemple : Retro-Tech







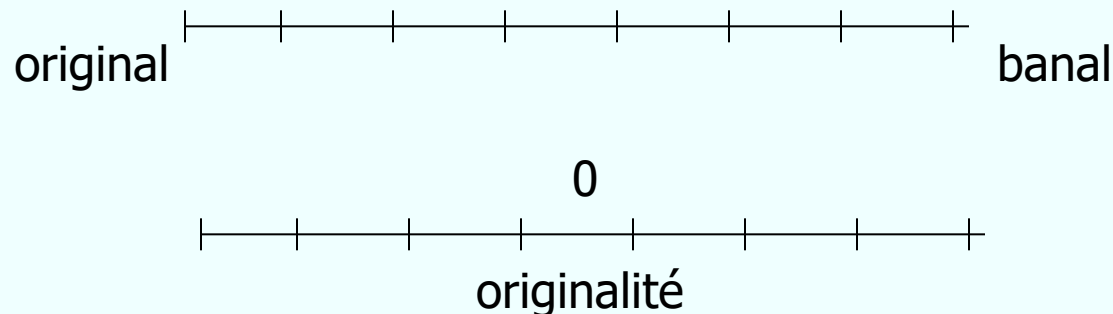
Le différentiel sémantique [Osgood]

◆ Objectif

- outil psychométrique pour la mesure d'une attitude
- évaluation des valeurs connotatives des objets ou des images

◆ Méthode

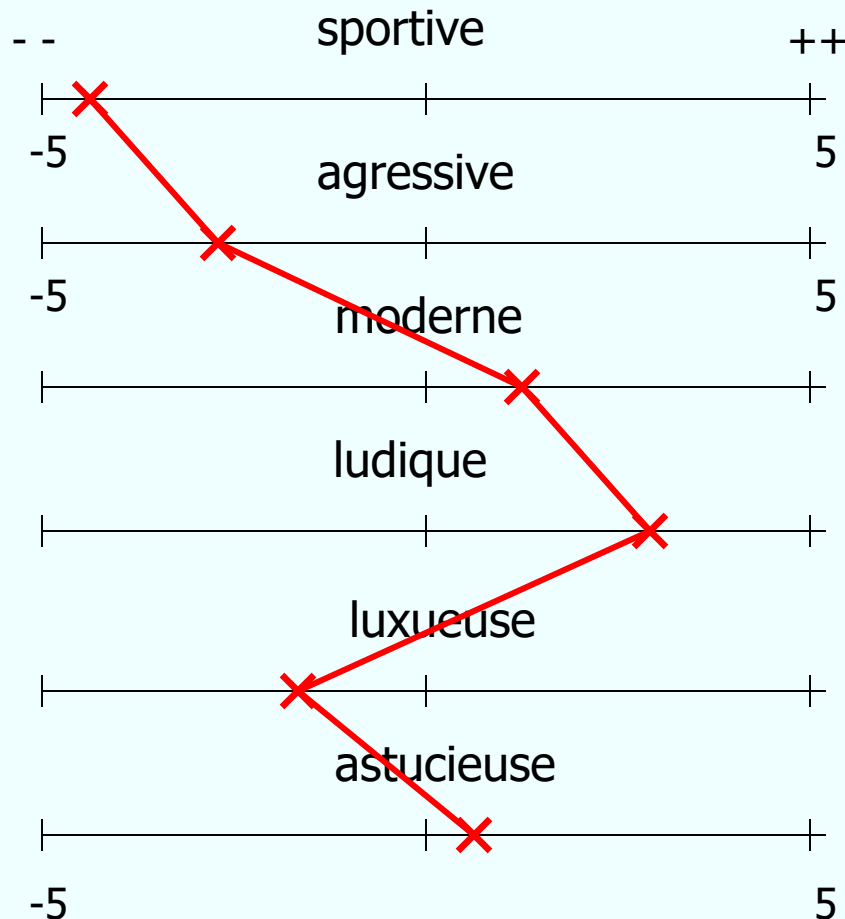
- établir une liste de qualificatifs (items ou paires d'adjectifs)
- constituer une grille de 5 ou 7 cases pour permettre au sujet de « positionner » son jugement





Différentiel sémantique

- Le sujet « évalue » les produits sur une liste d'attributs sémantiques prédéfinis



profil sémantique

Le point central est
considéré comme un
point neutre



La perception

Perception = ensemble des processus de traitement de l'information sensorielle

- Permettent , par l'interprétation des données reçues, , de construire une représentation interne du monde

Interaction Homme/environnement est fonction :

- De l'information captée par le sujet
- De la rétroaction exercée par lui sur ces données (sélection – organisation, interprétation)

La perception mobilise 2 formes de traitement

- Traitements **ascendants** (bottom-up) dirigés par les données
- Traitements **descendants** (top-down) dirigés par les concepts ou représentations

Réduire la perception à une association de sensations ne permet pas de rendre compte des phénomènes en jeu lors de notre perception de l'environnement (théorie Gestalt)

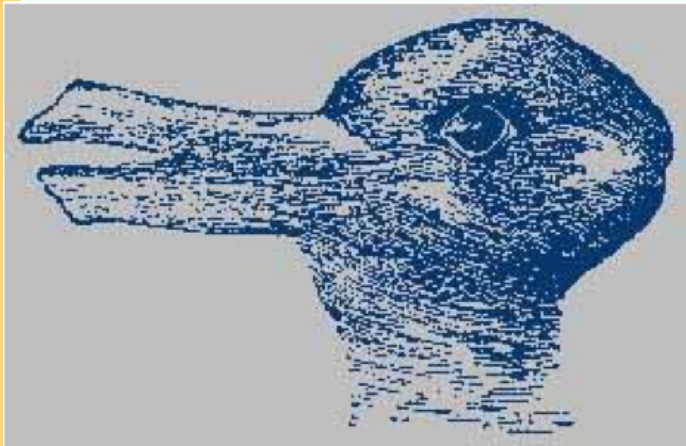


La perception

La perception est active

- Nous **sélectionnons**, choisissons ce que nous percevons
- Nous **organisons**, structurons, **interprétons** et donnons un sens aux signes et signaux qui nous parviennent
- Nous créons ce que nous voyons ou entendons

La perception résulte d'un **processus physique et mental** qui utilise nos cinq sens



Bistabilité visuelle : Canard ou lapin ?



vase ou visages ?



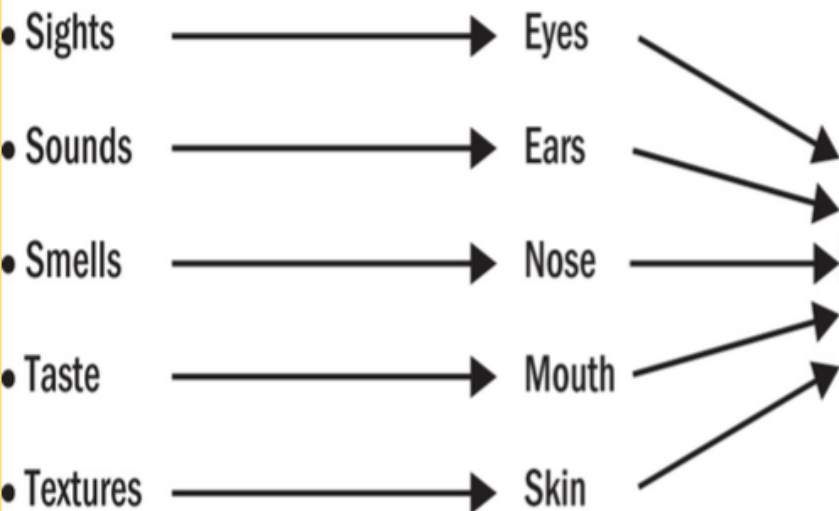
Le processus de perception

On reçoit des
stimuli
externes

Interprétation
par des
processus
cognitifs
internes

SENSORY STIMULI

SENSORY RECEPTORS

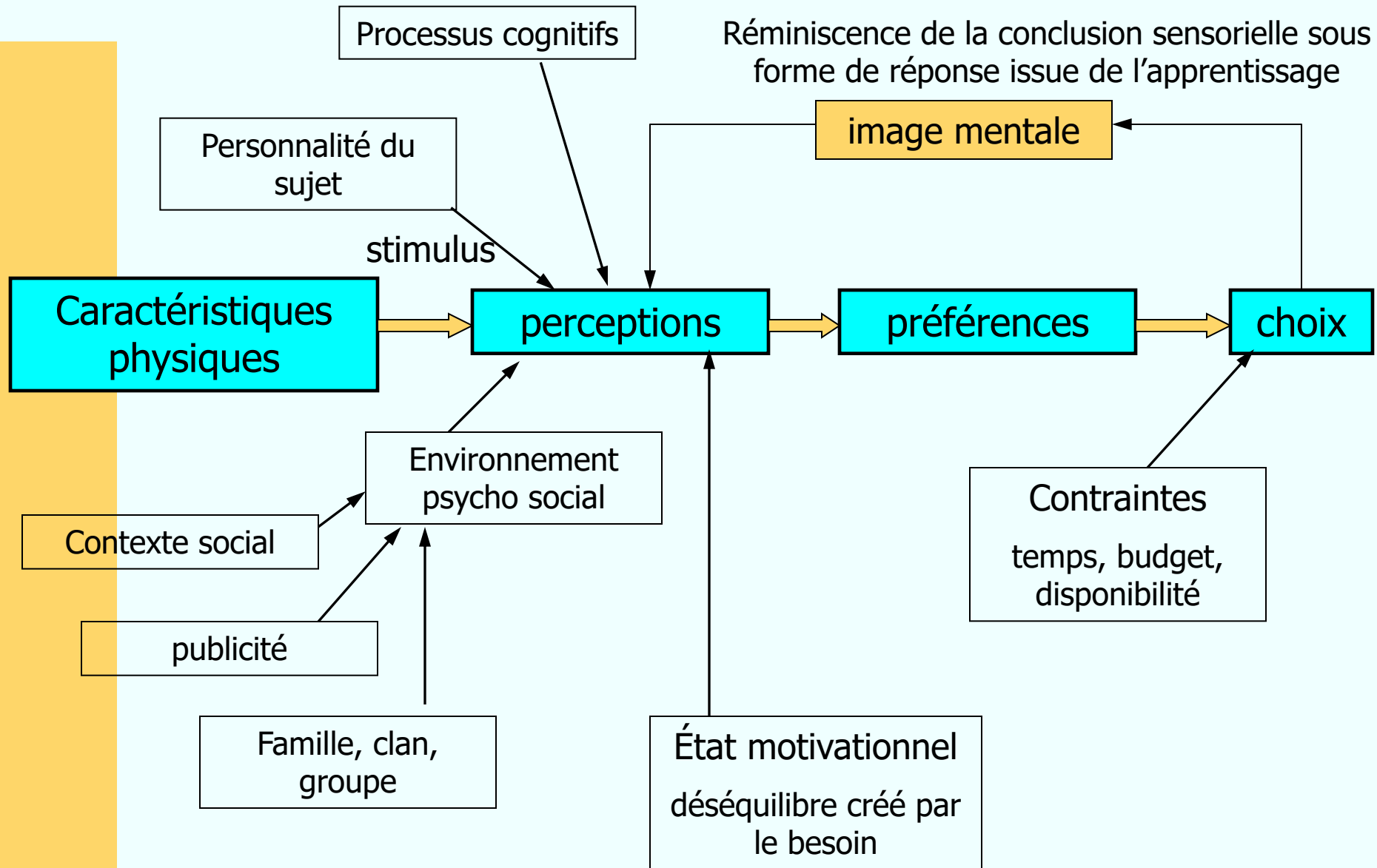


La perception est ACTIVE

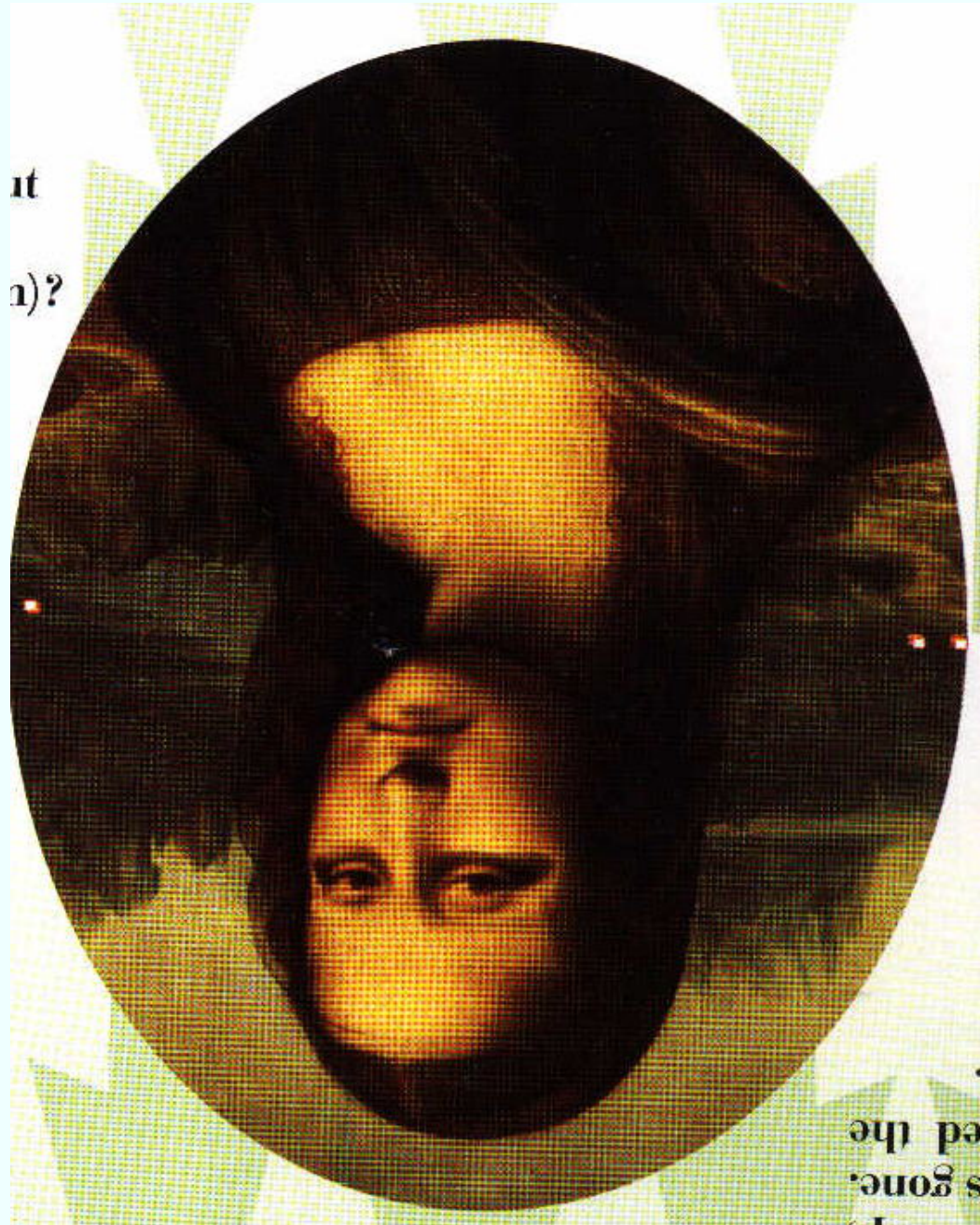
L'Humain se construit une représentation cohérente du
monde



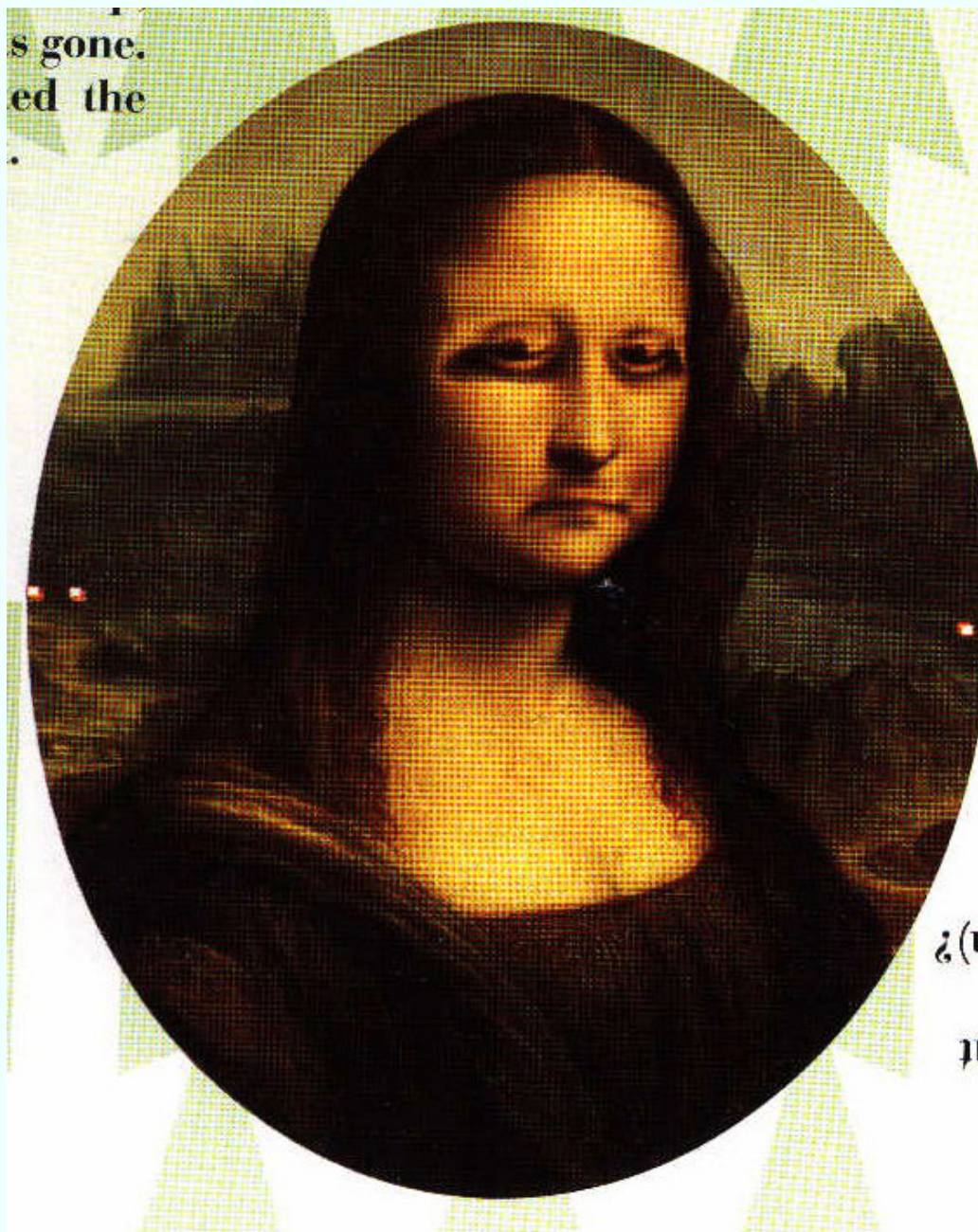
Perceptiologie [Brunswik]

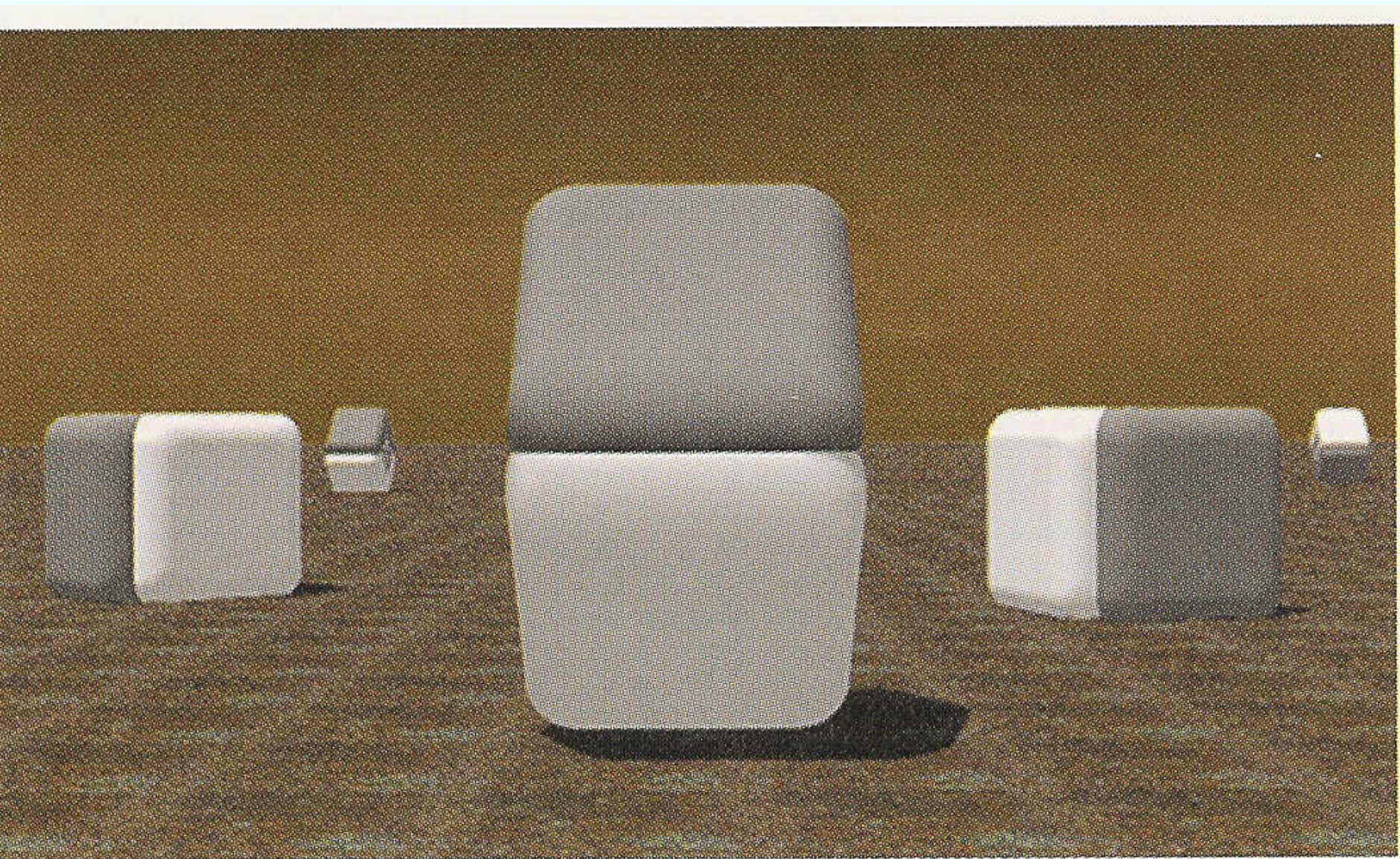


it
n)?

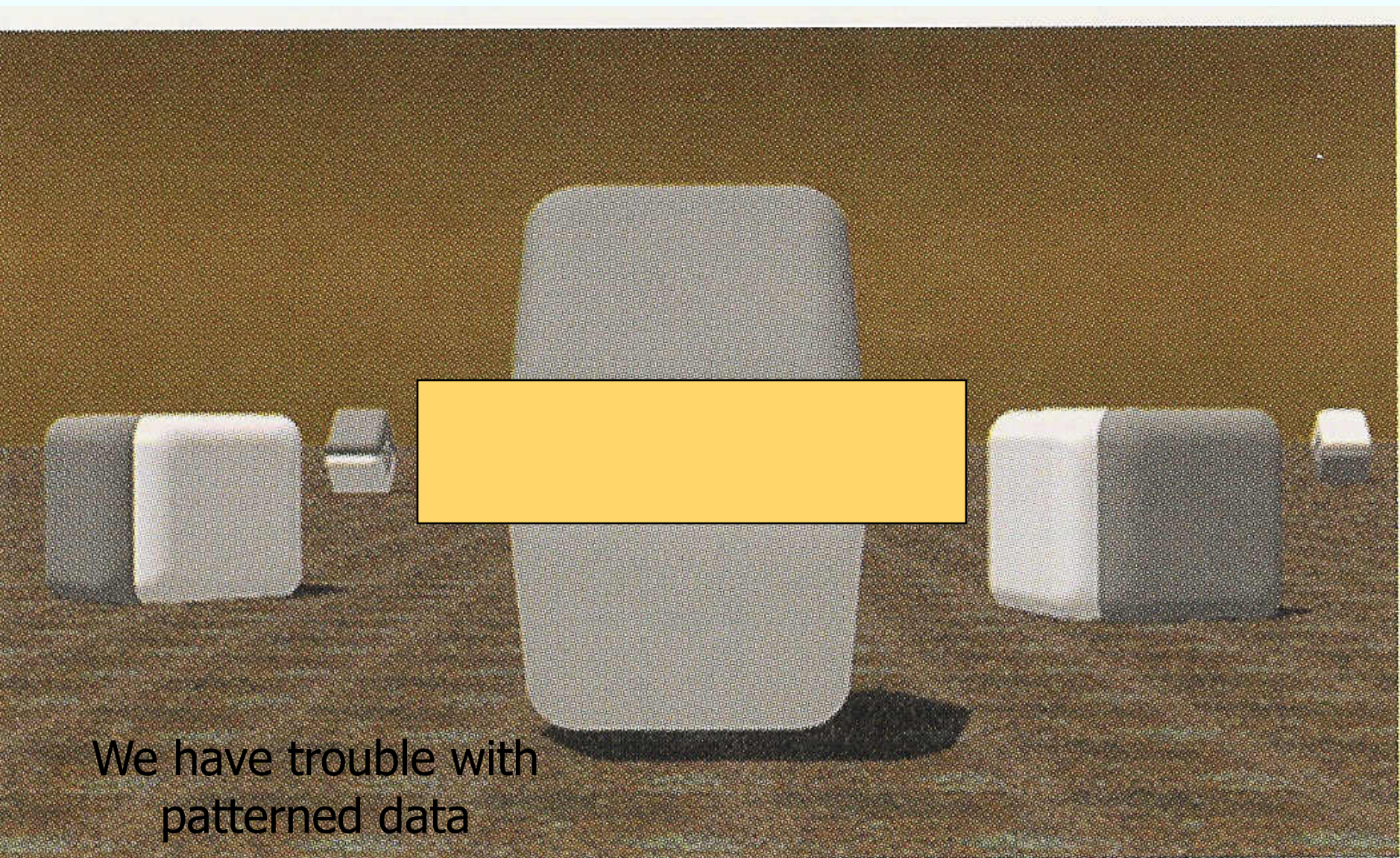


s gone.
ed the



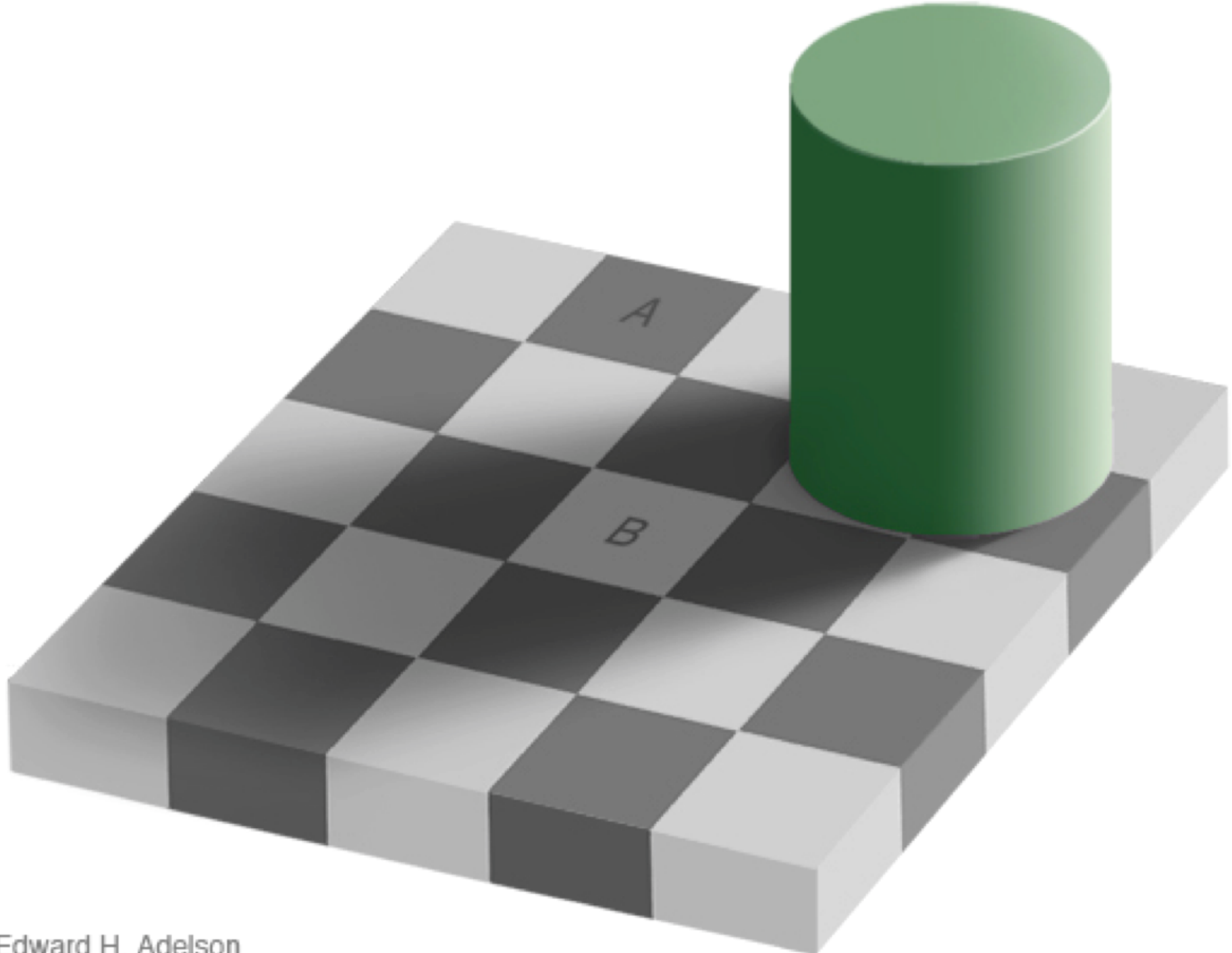


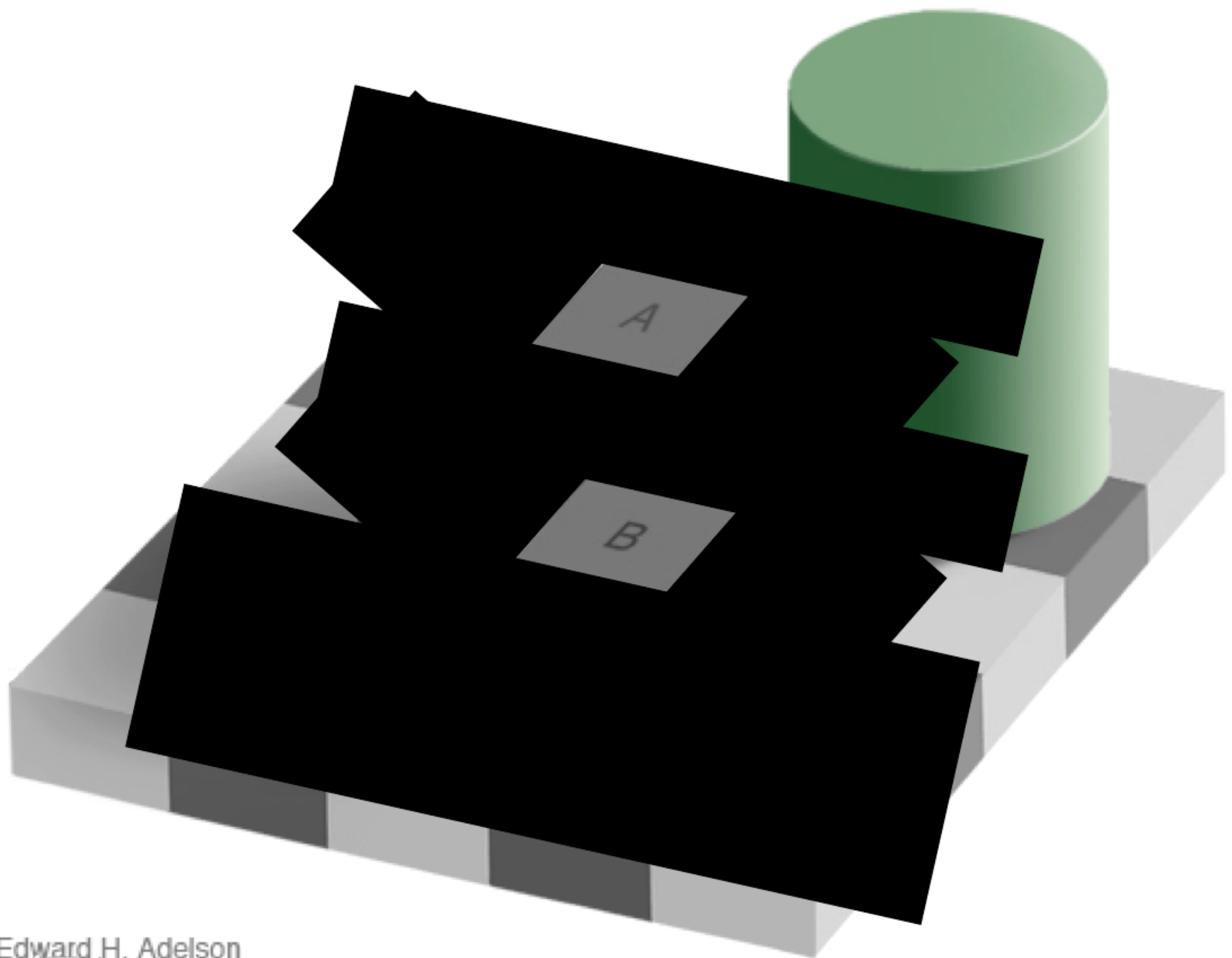
CORNSWEET ILLUSION © DALE PURVES AND



We have trouble with
patterned data

CORNSWEET ILLUSION © DALE PURVES AND

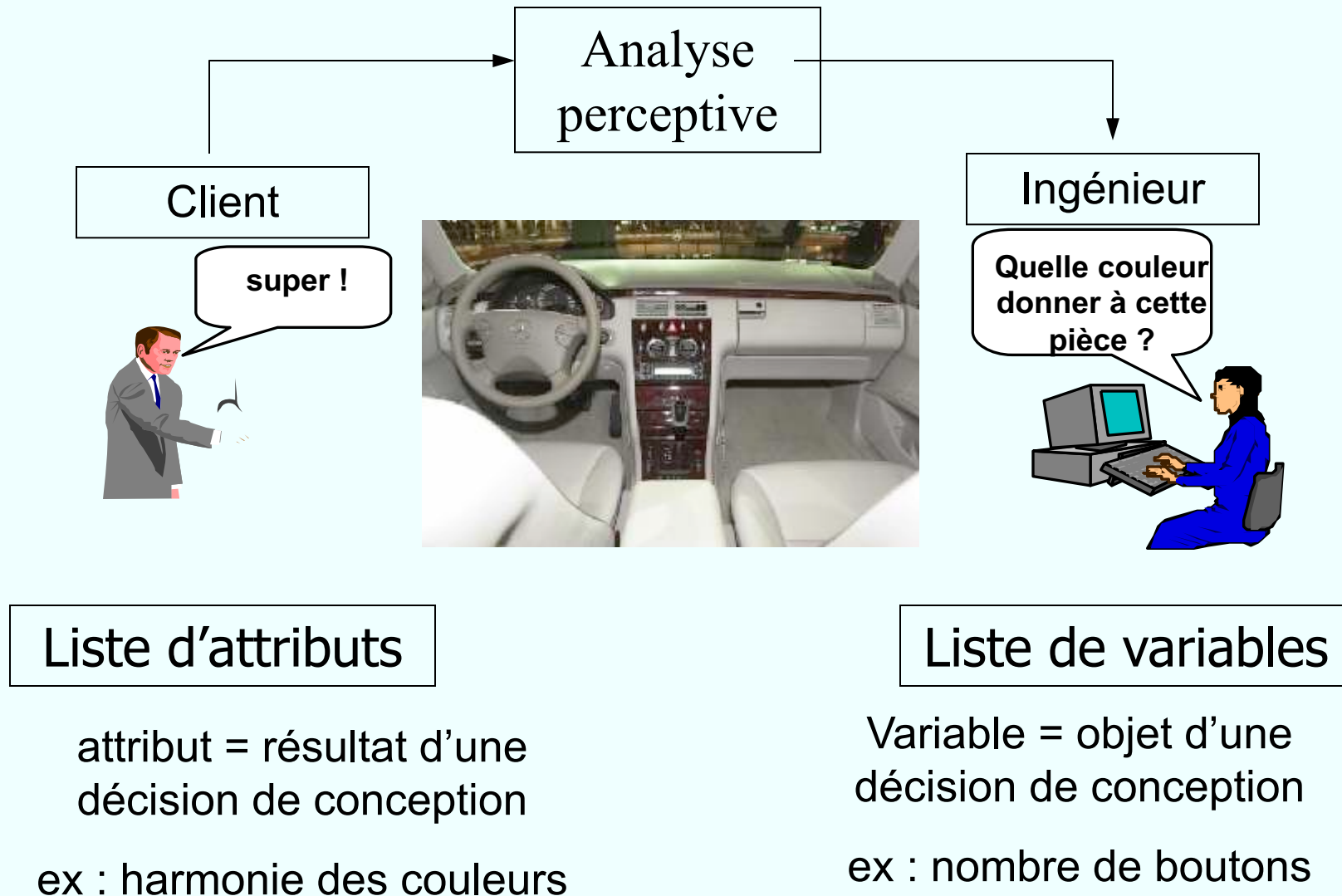






Qualité perçue

- ♦ Passer de la qualité “dure” à la qualité “perçue” ?



Exemple : « tapotement » de planches de bord (« toqué ») [Bezatz, 2007]

- ◆ Une entreprise a des retours clients négatifs concernant le bruit de « toqué » de ses planches de bord (enquêtes)
- ◆ Quelles actions mettre en œuvre pour traiter ce problème ?



s18



s12



s16

Quantifier l'attente client

Vérifier que la proposition produit est conforme aux attentes

Assurer la conformité tout au long du processus de développement du produit

Expert ACOUSTIQUE = Le son produit dépend de différentes variables de conception

-Choix matériaux

-Formes

-Qualité des assemblages

Exemple « toqué » (2)

♦ Étude sensorielle et analyse signal

- 31 sons enregistrés

Descripteurs

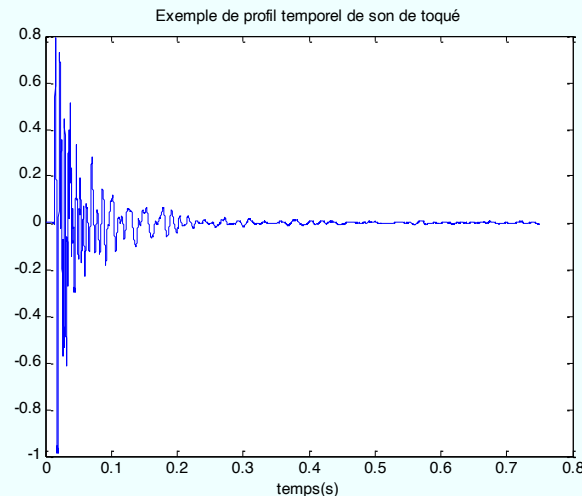
hauteur

détonant

intensité

attaquant

éloignement



s18



s12



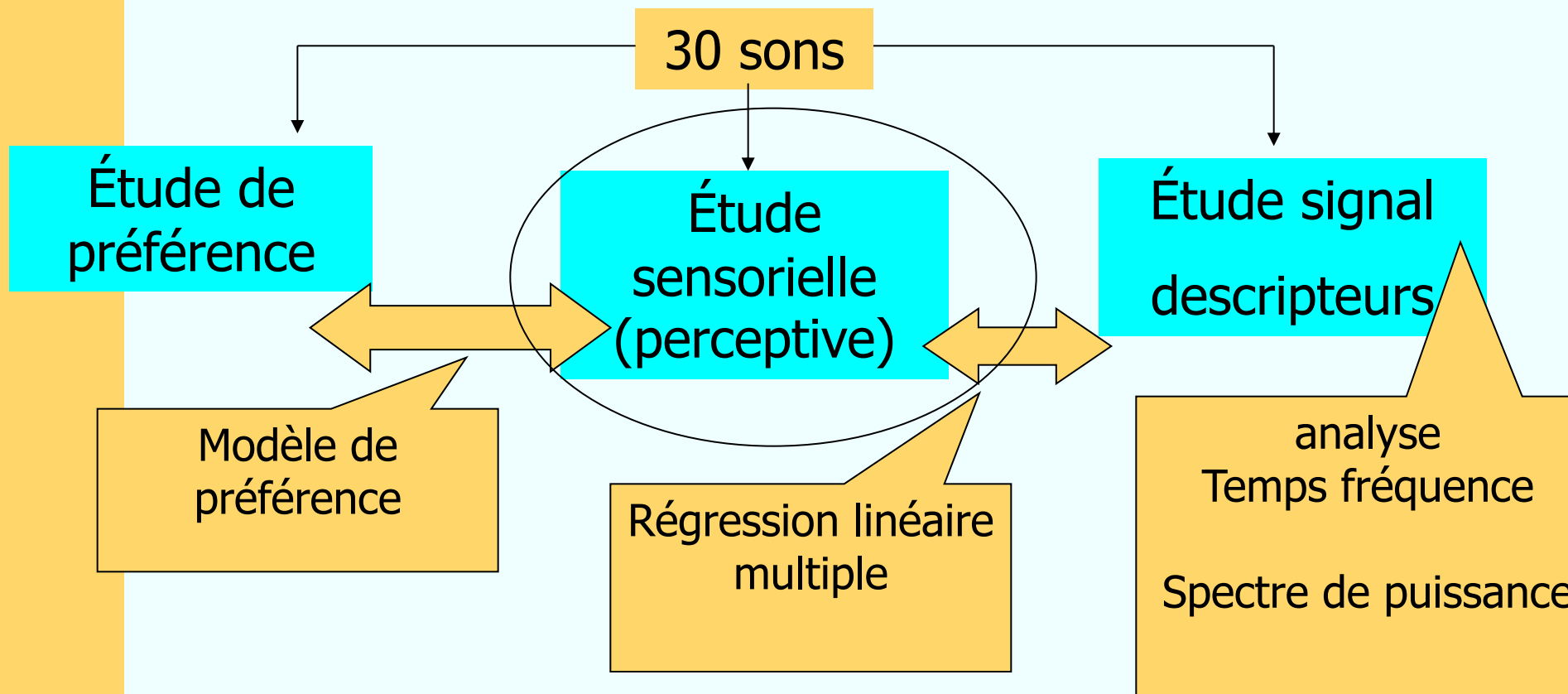
s16

=> Interprétation des descripteurs sensoriels par des critères liés au signal acoustique



Exemple « toqué (3) » : Démarche d'étude

- ♦ 3 types d'étude





Etude sensorielle et perceptive

- ◆ Comment comprendre ce qui structure la perception ?
 - ◆ Comment définir les dimensions perceptives principales qui sous-tendent la perception ?
- | Méthodes verbales | Méthodes non verbales |
|--|--|
| <ul style="list-style-type: none">• Profil sensoriel (panel d'experts)• Différentiel sémantique | <ul style="list-style-type: none">• Tri libre• Épreuve de dissimilarité |



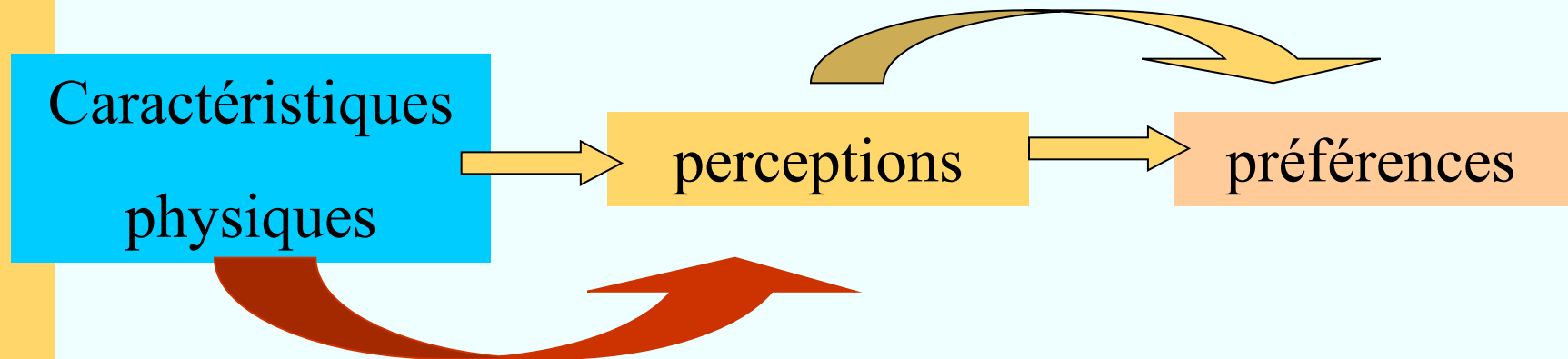
Intégrer les perceptions en design

- ◆ Méthodes qualitatives
 - Enquêtes
 - Focus groups

- ◆ Méthodes quantitative
 - Méthode du différentiel sémantique
 - Analyse conjointe
 - Psychophysique
 - Analyse sensorielle et sémantique

Cadre de travail

- Étude des liens :



- Secteurs d'application

- Industrie alimentaire
- Cosmétique
- Packaging
- Design sonore
- Transport - automobile
- ...

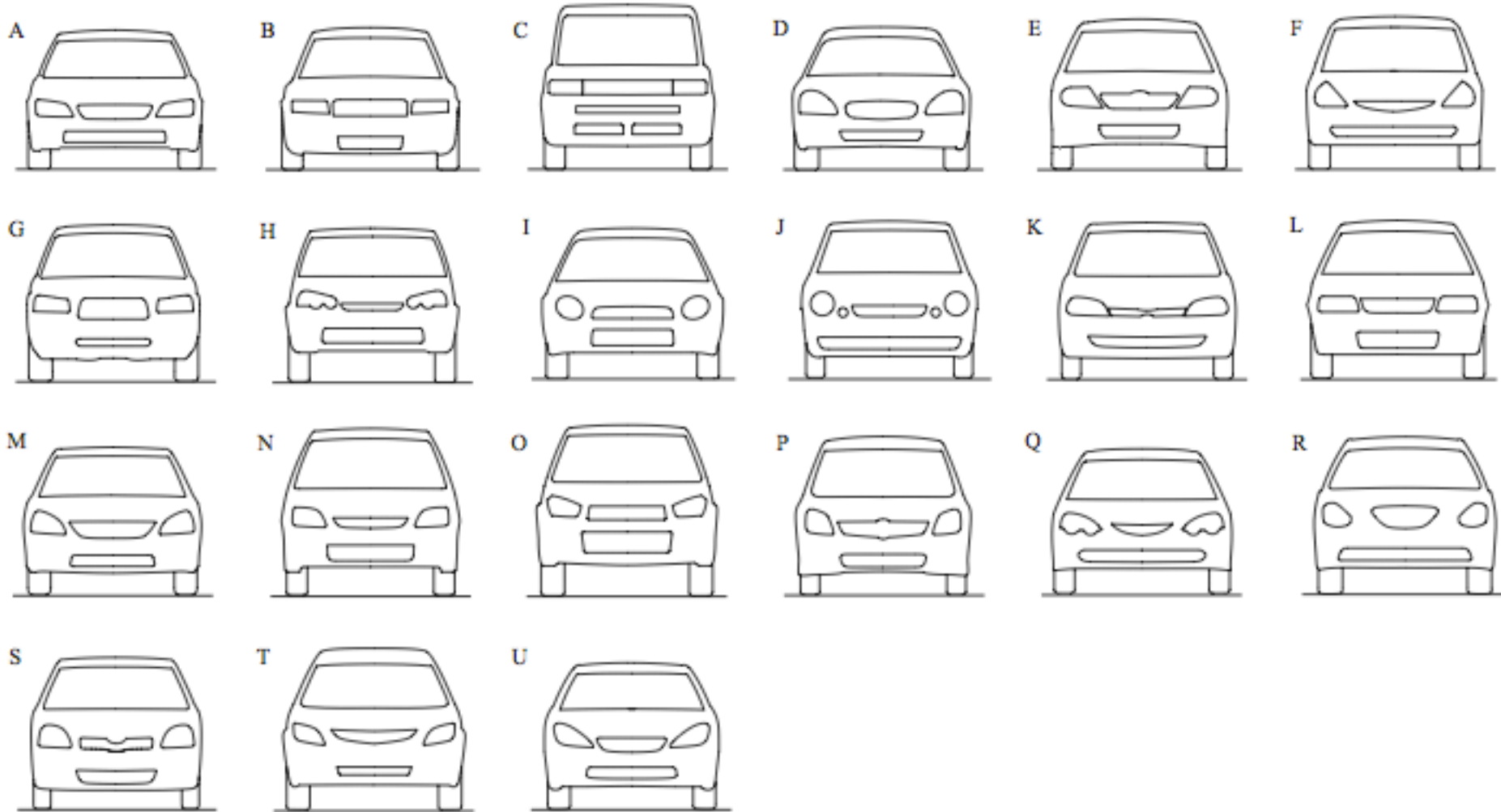
- Méthodes

- Analyse sensorielle
- Enquêtes quali/quantitatives
- Psychophysique
- Analyse conjointe
- ...



Exemple : face avant de véhicule

◆ Espace produit





Exemple : face avant

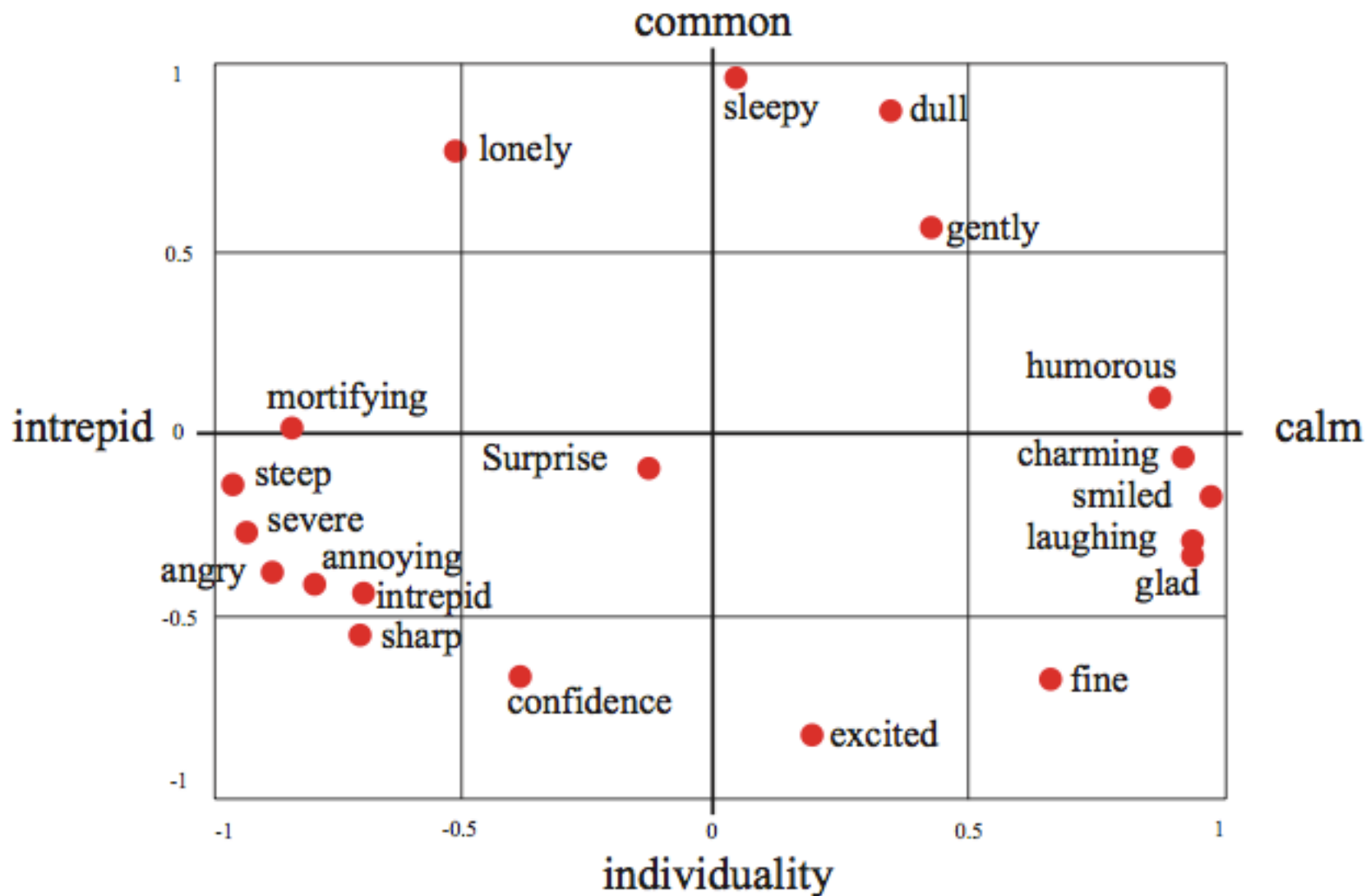
- ◆ Evaluation sémantique
 - 20 termes

	-3	-2	-1	0	1	2	3
laughing					×		
sharp			×				
surprise							×
fine				×			
intrepid						×	
angry						×	
lonely		×					
excited				×			
steep				×			
smiled						×	
confident					×		
sleepy			×				
charming		×					
severe				×			
gently					×		
dull					×		
mortifying					×		
annoying					×		
glad			×				
humorous			×				



Exemple : face avant

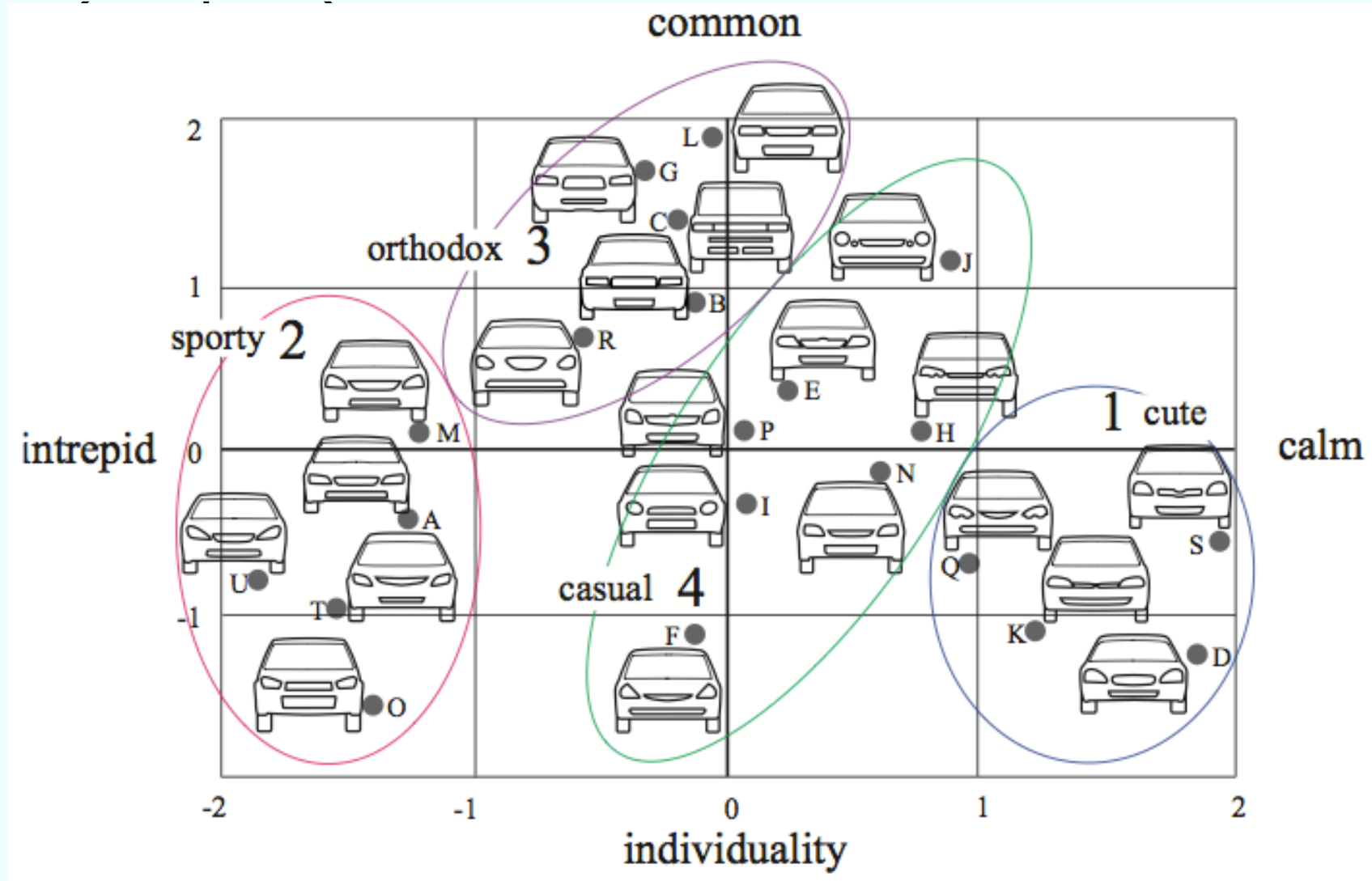
- ◆ Analyse en composantes principales (variables





Exemple : face avant

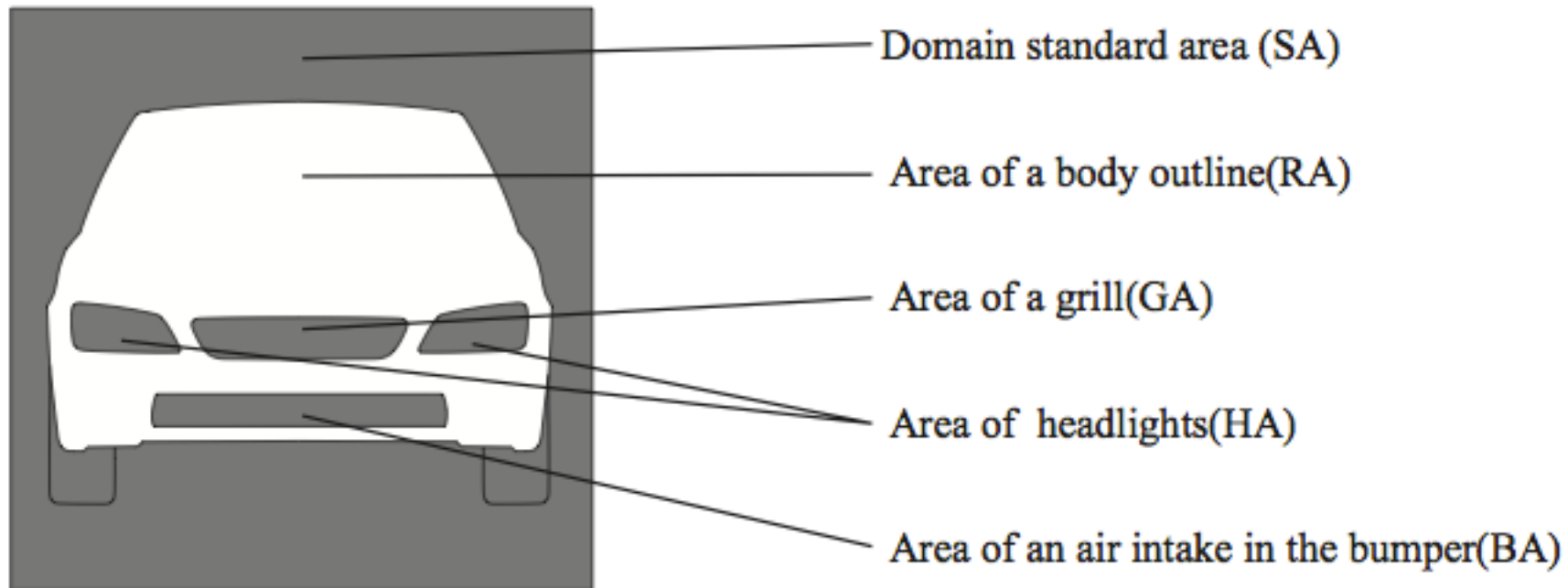
- ◆ Analyse en composantes principales (produits)



Exemple : face avant

- ◆ Relation entre perceptions et design attributs

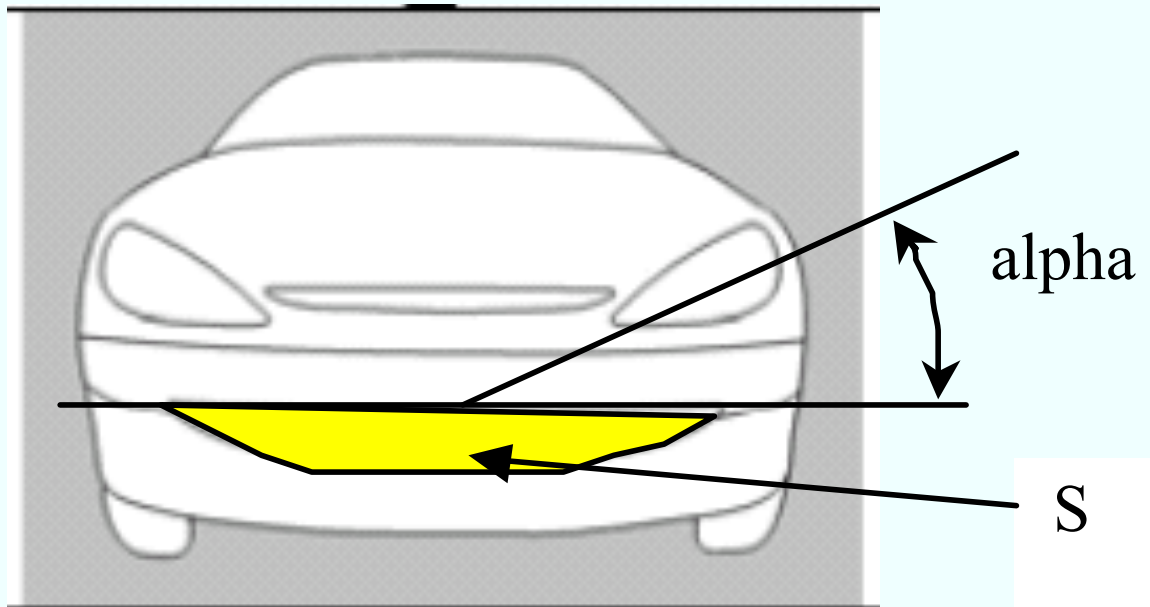
Exemple d'éléments formels



Exemple : face avant

- ◆ Relation entre perceptions et design attributs

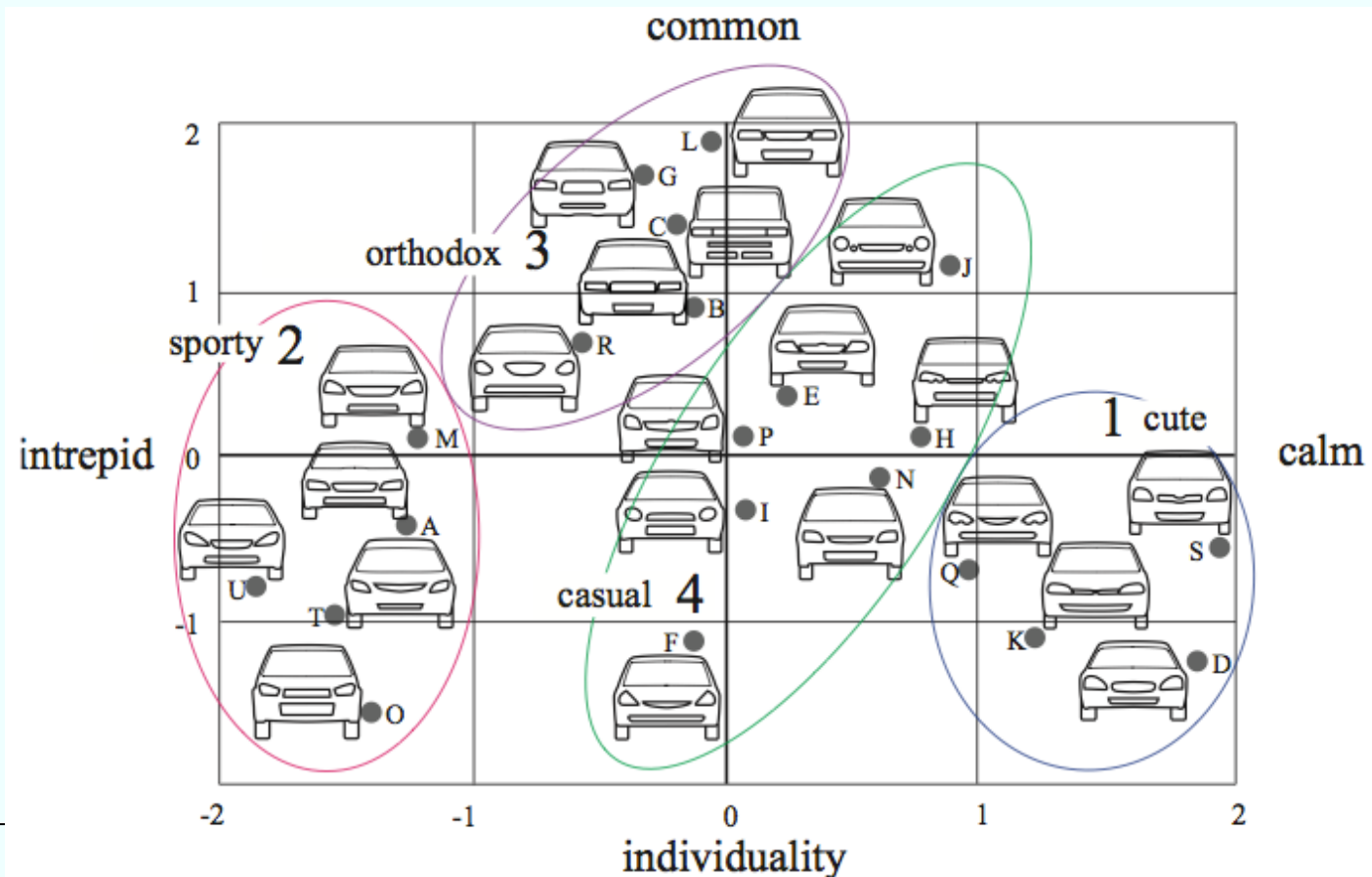
Exemple d'éléments formels

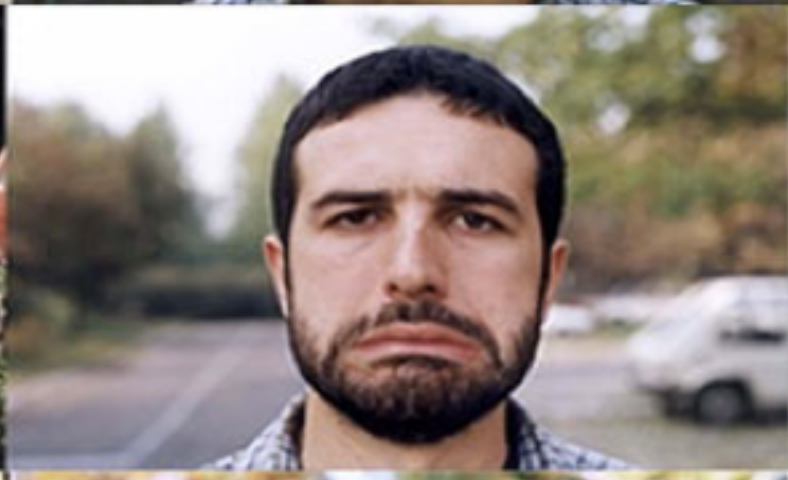
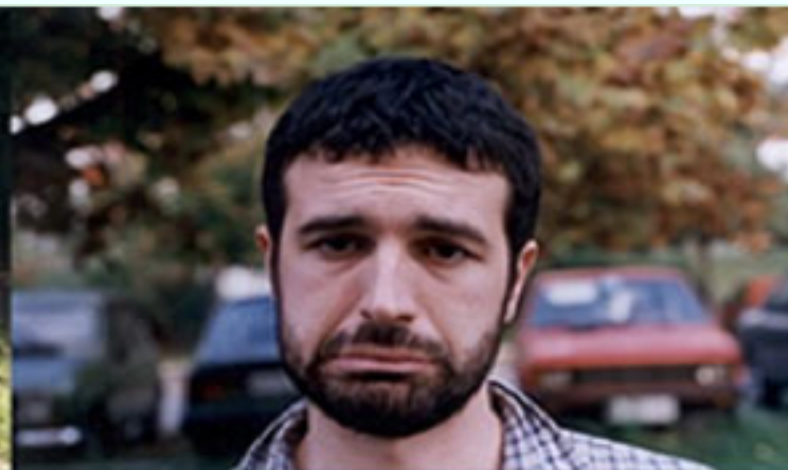




Exemple : face avant

- ◆ Typologie des groupes
 - Quels caractères sont représentatifs d'un groupe ?
 - => analyse de données
 - => raisonnement par analogie











L'analyse conjointe

quelles “ options ” donner à un nouveau produit ?
quels prix leurs attribuer ?

...ou de design

Quels sont les liens entre des **attributs formels** du produit et les **dimensions sémantiques** du produit, ou les **préférences**, de l'utilisateur ?



Petit exemple

- ◆ mise sur le marché d 'une nouvelle balle de golf :
 - 3 caractéristiques primordiales pour le joueur :
 - La distance moyenne de “ drive ”
 - La durée de vie moyenne
 - Le prix
 - ces caractéristiques peuvent prendre plusieurs valeurs (modalités)

<i>distance</i>
250 m
200 m
180 m

<i>durée de vie</i>
54 trous
36
18

<i>prix</i>
1€
1.5€
2€

Plan d'expériences complet : $3*3*3 = 27$ designs



Scénarii possibles

♦ hypothèses :

- Le coût de production d'une balle est d'autant moins grand que la distance de drive et la durée de vie sont faibles
- Le client recherche une balle peu chère et aux performances maximales

Produit #1 optimal du point de vue
« demande »

Distance	250 m
Durée de vie	54 trous
prix	1€

On va en vendre beaucoup, mais on
risque de ne pas faire de bénéfices

Produit #2 optimal du point de vue
« offre »

Distance	180 m
Durée de vie	18 trous
prix	2 €

On a bétonné sur le prix, mais elle
risque de ne pas se vendre



Recherche d'un compromis

- ◆ Problème de marketing :
 - *On risque de “ perdre notre chemise ” à vendre le produit #1, et on va perdre notre énergie à tenter de vendre le produit#2 dont personne ne voudra.*

Rappel : $\text{Prix} = \text{cout} + \text{marge}$

- ◆ Le produit le plus viable est donc un produit intermédiaire entre #1 et #2, mais comment en trouver les caractéristiques ?
 - Il nous manque en fait ici une donnée capitale : les préférences des clients.
 - L'analyse conjointe va permettre de répondre à cette question.



Recherche d'un compromis

- ◆ Etape 1 : classer les différentes modalités des caractéristiques du produit :

<i>rang</i>	<i>distance</i>	<i>rang</i>	<i>durée de vie</i>
1	250	1	54
2	200	2	36
3	180	3	18

- ◆ Etape 2 : considérer “ conjointement ” les 2 caractéristiques “ durée de vie ” et “ distance ” :
 - établir (par enquêtes) le classement de préférence des 9 balles possibles pour différents clients, le prix étant le même pour toutes les combinaisons



Recherche d'un compromis

◆ classement des préférences selon les modalités

Client 1

Durée de vie

distance

	54 trous	36 trous	18 trous
250 m	1	2	4
200 m	3	5	7
180 m	6	8	9

Client 2

Durée de vie

distance

	54 trous	36 trous	18 trous
250 m	1	3	6
200 m	2	5	8
180 m	4	7	9

Les deux clients sont d'accord sur les choix extrêmes, mais le client 2 met plus l'accent sur la durée de vie, le client 1 sur la distance. Leurs “ arbitrages ” (trade-off) sont différents

L'information que l'on retire de ces deux tableaux est fondamentale et est l'essence même de l'analyse conjointe



Définition des utilités partielles

- Étape 3 : Considérons maintenant pour chaque modalité une valeur arbitraire telle que la somme de ces valeurs pour chaque balle conduit à un classement conforme au classement du client

Client 1

Durée de vie

distance

	54 trous	36 trous	18 trous
	70	25	0
250 m	1	2	4
100	170	125	100
200 m	3	5	7
50	120	75	50
180 m	6	8	9
0	70	25	0

Cette valeur s'appelle **l'utilité partielle** du niveau du facteur

- l'utilité partielle du niveau "200 m" du facteur distance est égale à +50



Définition des utilités partielles

◆ NB

- Il n'y a pas de solution unique
- Seules les différences entre les valeurs des utilités partielles sont importantes
- Nous avons choisi arbitrairement la valeur 0 pour « 18 trous » et « 180m ». On aurait pu imposer une somme nulle pour les « utilités » des niveaux d'un même facteur (paramétrage centré)

Client 1

Durée de vie

distance

	54 trous 38,3	36 trous -6,6	18 trous -31,7
250 m 50	1 88,3	2 43,4	4 18,3
200 m 0	3 38,3	5 -6,6	7 -31,7
180 m -50	6 -11,7	8 -56,6	9 -81,7



Définition des utilités partielles

- ◆ Étape 4 : ”, considérons maintenant les arbitrages entre “ prix ” et “ durée de vie ” par ex. pour le client 1
 - - > on détermine les utilités partielles des niveaux du facteur prix

Client 1

Durée de vie

prix

	54 trous	36 trous	18 trous
1€	1	4	7
1,5€	2	5	8
2€	3	6	9

Client 1

Durée de vie

prix

	54 trous	36 trous	18 trous
	70	25	0
1€	1	4	7
20	90	45	20
1.5€	2	5	8
5	75	30	5
2€	3	6	9
0	70	25	0



Utilités partielles et utilité totale

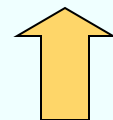
- ◆ Etape 5 : Nous avons maintenant pour chaque modalité un ensemble complet de valeurs (appelées **utilités partielles** ou “part-worths”) qui ont capté tous les arbitrages du client n°1

	Distance			Durée de vie			Prix		
	250m	200m	180m	54 trous	36 trous	18 trous	1 euro	1,5 euro	2 euros
Utilité partielle	100	50	0	70	25	0	20	5	0

- ◆ Etape 6 : utilisation : sélection entre 2 balles

	Balle A		Balle B	
Distance	250m	100	200 m	50
Durée	18 trous	0	54 trous	70
Prix	1,5 euros	5	2 euros	0
utilité		105		120

Utilité balle B > utilité balle A





Utilité totale de tous les designs du plan factoriel ($3 \times 3 = 27$)

Avec ce modèle d'utilités partielles, il est possible de prédire l'utilité totale de l'ensemble des designs du plan complet

design	distance	durée vie	prix	utilité totale
n1	d1	dv1	p1	190
n2	d1	dv1	p2	175
n3	d1	dv1	p3	170
n4	d1	dv2	p1	145
n5	d1	dv2	p2	130
n6	d1	dv2	p3	125
n7	d1	dv3	p1	120
n8	d1	dv3	p2	105
n9	d1	dv3	p3	100
n10	d2	dv1	p1	140
n11	d2	dv1	p2	125
n12	d2	dv1	p3	120
n13	d2	dv2	p1	95
n14	d2	dv2	p2	80
n15	d2	dv2	p3	75
n16	d2	dv3	p1	70
n17	d2	dv3	p2	55
n18	d2	dv3	p3	50
n19	d3	dv1	p1	90
n20	d3	dv1	p2	75
n21	d3	dv1	p3	70
n22	d3	dv2	p1	45
n23	d3	dv2	p2	30
n24	d3	dv2	p3	25
n25	d3	dv3	p1	20
n26	d3	dv3	p2	5
n27	d3	dv3	p3	0



Recherche d'un compromis

- ◆ Étape 7 : modèle de calcul du coût de production

Coût

Centimes d'euros

Durée de vie

distance

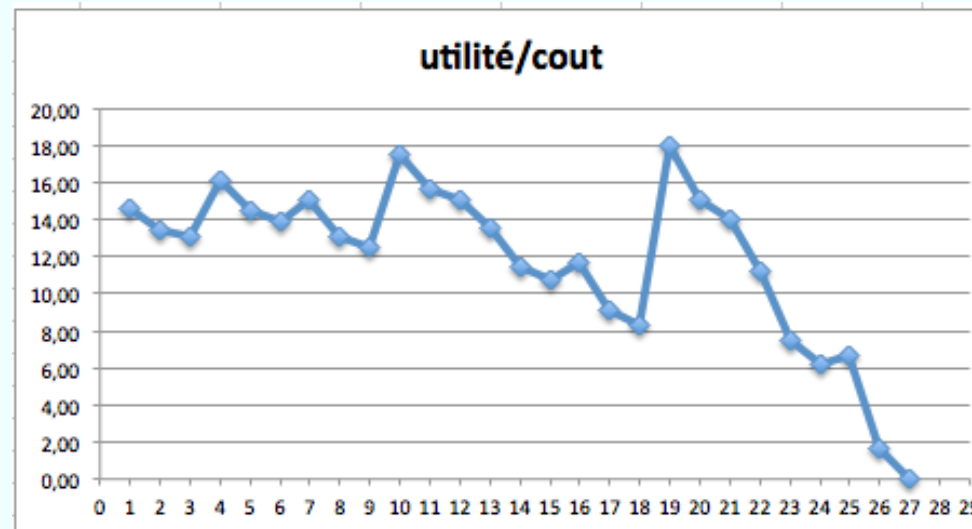
	54 trous (dv1)	36 trous (dv2)	18 trous (dv3)
250 m (d1)	13	9	8
200 m (d2)	8	7	6
180 m (d3)	5	4	3

- ◆ Etape 8 : optimisation du rapport rapport utilité/coût

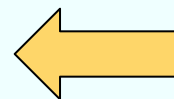


Produit optimal (utilité/cout)

design	distance	durée vie	prix	utilité totale	cout	utilité/cout
n1	d1	dv1	p1	190	13	14,62
n2	d1	dv1	p2	175	13	13,46
n3	d1	dv1	p3	170	13	13,08
n4	d1	dv2	p1	145	9	16,11
n5	d1	dv2	p2	130	9	14,44
n6	d1	dv2	p3	125	9	13,89
n7	d1	dv3	p1	120	8	15,00
n8	d1	dv3	p2	105	8	13,13
n9	d1	dv3	p3	100	8	12,50
n10	d2	dv1	p1	140	8	17,50
n11	d2	dv1	p2	125	8	15,63
n12	d2	dv1	p3	120	8	15,00
n13	d2	dv2	p1	95	7	13,57
n14	d2	dv2	p2	80	7	11,43
n15	d2	dv2	p3	75	7	10,71
n16	d2	dv3	p1	70	6	11,67
n17	d2	dv3	p2	55	6	9,17
n18	d2	dv3	p3	50	6	8,33
n19	d3	dv1	p1	90	5	18,00
n20	d3	dv1	p2	75	5	15,00
n21	d3	dv1	p3	70	5	14,00
n22	d3	dv2	p1	45	4	11,25
n23	d3	dv2	p2	30	4	7,50
n24	d3	dv2	p3	25	4	6,25
n25	d3	dv3	p1	20	3	6,67
n26	d3	dv3	p2	5	3	1,67
n27	d3	dv3	p3	0	3	0,00



◆ Rapport utilité sur cout maximal : produit n19





Cas de la cotation de préférence (de 0 à 10)

◆ Notation des préférences selon les modalités

Client 1

Durée de vie

distance

	54 trous	36 trous	18 trous
250 m	8,5	6,25	5
200 m	6	3,75	2,5
180 m	3,5	1,25	0

◆ Recherche des utilités partielles

Client 1

Durée de vie

distance

	54 trous	36 trous	18 trous
250 m	8,5	6,25	5
200 m	6	3,75	2,5
180 m	3,5	1,25	0



Cas de la cotation de préférence (de 0 à 10)

Cotation de préférence entre “ prix ” et “ durée de vie ”

Client 1

Durée de vie

prix

	54 trous	36 trous	18 trous
1€	4,5	2,25	1
1,5€	4	1,75	0,5
2€	3,5	1,25	0

◆ Recherche des utilités partielles

Client 1

Durée de vie

prix

	54 trous 3,5	36 trous 1,25	18 trous 0
1€ 1	4,5	2,25	1
1.5€ 0,5	4	1,75	0,5
2€ 0	3,5	1,25	0



Cas de la cotation de préférence (de 0 à 10)

Méthode pour trouver les utilités partielles a_{jk} ?

- ♦ On choisit un modèle de préférence du type :

$$\hat{y}_i = \mu + \sum_{j=1}^p \sum_{k=1}^{k_j} a_{jk} \cdot \delta_i(jk)$$

Modèle linéaire
compensatoire

utilité globale du design i

k_j = nombre de niveaux du
facteur j

p = nombre de facteurs

a_{jk} : utilité partielle du niveau k pour le facteur j

$$\delta_i(jk) \begin{cases} = 1 \text{ si niveau du facteur } j \text{ pour le design } i \text{ est } k \\ = 0 \text{ sinon} \end{cases}$$

a_{jk} est le coefficient d'une régression linéaire multiple, déterminé par la minimisation de :

$$\text{MIN} \left(\sum_{i=1}^n (y_i - \hat{y}_i)^2 \right)$$



1^{ier} bilan sur l'analyse conjointe

- ◆ À partir de la préférence client, on construit la part (**utilité partielle**) de chaque modalité du produit dans la préférence
- ◆ C'est un modèle de **décomposition** : l'attractivité de chaque modalité est estimée à partir des préférences déclarées du consommateur
- ◆ On suppose un modèle linéaire additif de la contribution de chaque modalité des facteurs
- ◆ applications
 - Marketing
 - Finance
 - Production industrielle
 - Design industriel [Kansei engineering]
 - Quantification theory Type I



Méthodologie de l'analyse conjointe



L'analyse conjointe

♦ objectif

Évaluer les critères de choix d'un consommateur confronté à des alternatives

Expliquer et comprendre les choix des individus

Définir la proposition optimale dans un contexte concurrentiel

Mesure l'influence d'une modification d'un ou plusieurs éléments

Pratiquer des stratégies de segmentation en connaissant les priorités des différentes cibles



Hypothèses

Les choix du consommateur portant sur un produit sont :

- ◆ **Multi-attribut** : le consommateur perçoit un produit comme un ensemble d'attributs
- ◆ **Cognitifs** : le consommateur tient compte de ce qu'il sait, ou croit savoir de l'objet, pour se construire une attitude
- ◆ Basés sur la **théorie de l'utilité**

Modèle de Lancaster : « *Le consommateur opte pour la catégorie de bien qui maximise son utilité, sous la contrainte budgétaire* »

Tout choix implique des **compromis** qui dépendront de l'importance accordée par le consommateur à chacun des attributs

Analyse conjointe = méthode des TRADE-OFF



Méthodologie : 3 étapes

◆ 1 : Recueil des données

- Créer une situation qui amène l'interviewé à révéler ses attentes/produit par un processus de **choix** ou de **classement**

⇒ Scénarii (plan d'expériences)

- Deux grandes familles de méthodes :
 - L'analyse traditionnelle (évaluation - **rating based** - (ex. de la balle de golf))
 - L'analyse conjointe discrète (choix - **choice based conjoint** sur profils complets)

◆ 2. Estimation du modèle

- Déterminer les **utilités partielles** et les **importances** des attributs, à partir des évaluations des consommateurs

◆ 3. Utilisation du modèle

- Définition du produit « idéal »
- Prédiction de parts de marché - Segmentation de consommateurs



Etape 1 : recueil des données

- ◆ Choisir les attributs et leurs modalités
 - Déterminants dans le choix du consommateur
 - Indépendants
 - Exhaustifs (décrivent complètement le produit)
 - Manipulables par l'entreprise

 - ◆ Choisir les designs proposés aux consommateurs
 - Plan factoriel complet ou fractionnaire
 - Plan optimal (D-optimaux)
- Théorie des plans d'expériences
- ◆ Choisir le mode de présentation des designs
 - Texte, dessins, photos, maquettes CAO, produits réels, ...

 - Définir l'échelle de mesure de la préférence
 - Classement, cotation, comparaisons par paires



Etape 1 : exemple jus d'orange

◆ Les attributs (facteurs)

- Recette (8 modalités)
- Marque (4 modalités)
- Type (2 modalités)

⇒ R1, R2, R3, ..., R8

⇒ joker/Solevita/Jafaden/Carrefour

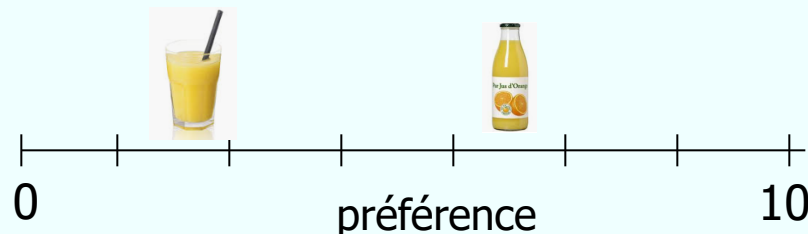
⇒ Brick/bouteille verre

Nbre de designs du **plan complet** ?

$$8 * 4 * 2 = 64$$

Choix d'un plan **fractionnaire**
à 16 designs

Cotation sur une échelle de 0 à 10



Scenario	Recette	Marque	Type
1	1	1	1
2	2	1	2
3	3	2	1
4	4	2	2
5	5	3	1
6	6	3	2
7	7	4	1
8	8	4	2
9	1	2	2
10	2	3	1
11	3	3	2
12	4	4	1
13	5	4	2
14	6	1	1
15	7	1	2
16	8	2	1



N.B: orthogonalité d'un PLEX

Propriété requise pour minimiser les incertitudes sur les coefficients et estimer les effets des facteurs indépendamment les uns des autres

Orthogonalité : pour tout couple de facteurs, chaque niveau de l'un est associé à chaque niveau de l'autre un même nombre de fois

Si ce nombre est le même pour tous les couples, le plan est équilibré

Plan équilibré : C'est un plan orthogonal pour lequel le nombre de fois qu'interviennent les niveaux de deux facteurs est le même, quelque soit les couples de facteurs considérés

On peut vérifier l'équilibre en formant, pour chaque couple de facteur, la **table de contingence** comptant les effectifs qui croisent les modalités
À l'équilibre, ces effectifs sont les mêmes pour tout couple de facteurs



Etape 1 : exemple carte bancaire

◆ Les facteurs

- Brand (3 levels)
- Annual fee (3 levels)
- Interest rate (3 levels)
- Credit limit (3 levels)

⇒ visa, Mastercard, Discover

⇒ 40\$, 20\$, 0\$

⇒ 10%, 14%; 18%

⇒ 1000\$, 2000\$, 5000\$

Nbre de designs du **plan complet** ?

$$3^4 = 81$$

Epreuve de choix de 1 parmi 4 options, répété n fois

If you were shopping for a credit card, and these were your only options, which would you choose?			
VISA \$40 annual fee 10% interest rate \$2,000 credit limit	Mastercard \$20 annual fee 18% interest rate \$5,000 credit limit	Discover No annual fee 14% interest rate \$1,000 credit limit	NONE: I would defer my purchase



Etape 2 : estimation du modèle

- L'attractivité d'un produit « i » peut s'écrire :

$$U_i = V_i + \varepsilon_i$$

Avec :

U_i : Attractivité latente, non mesurable

V_i : Attractivité observée sur le PLEX

ε_i : erreur inexpliquée, associée au produit "i"

On suppose que V_i est une fonction des attributs du produit « i », codés en X_i

Avec :

$$V_i = \sum_{k=1}^m a_{ik} \cdot x_{ik}$$

a_{ik} : utilité de l'attribut k du produit "i"
 x_{ik} : niveau de l'attribut k du produit "i"

m : nombre d'attributs



Etape 2 : estimation du modèle

- Cas du modèle « rating based » (ANOVA)

- L'utilité estimée est donnée par :
$$\hat{U}_i = V_i = \sum_{k=1}^m a_{ik} \cdot x_{ik}$$

Les utilités partielles a_{ik} sont données par minimisation de l'erreur quadratique

$$a_{ik} = \arg \min \left(\sum_{i=1}^n (U_i - \hat{U}_i)^2 \right)$$

Avec :

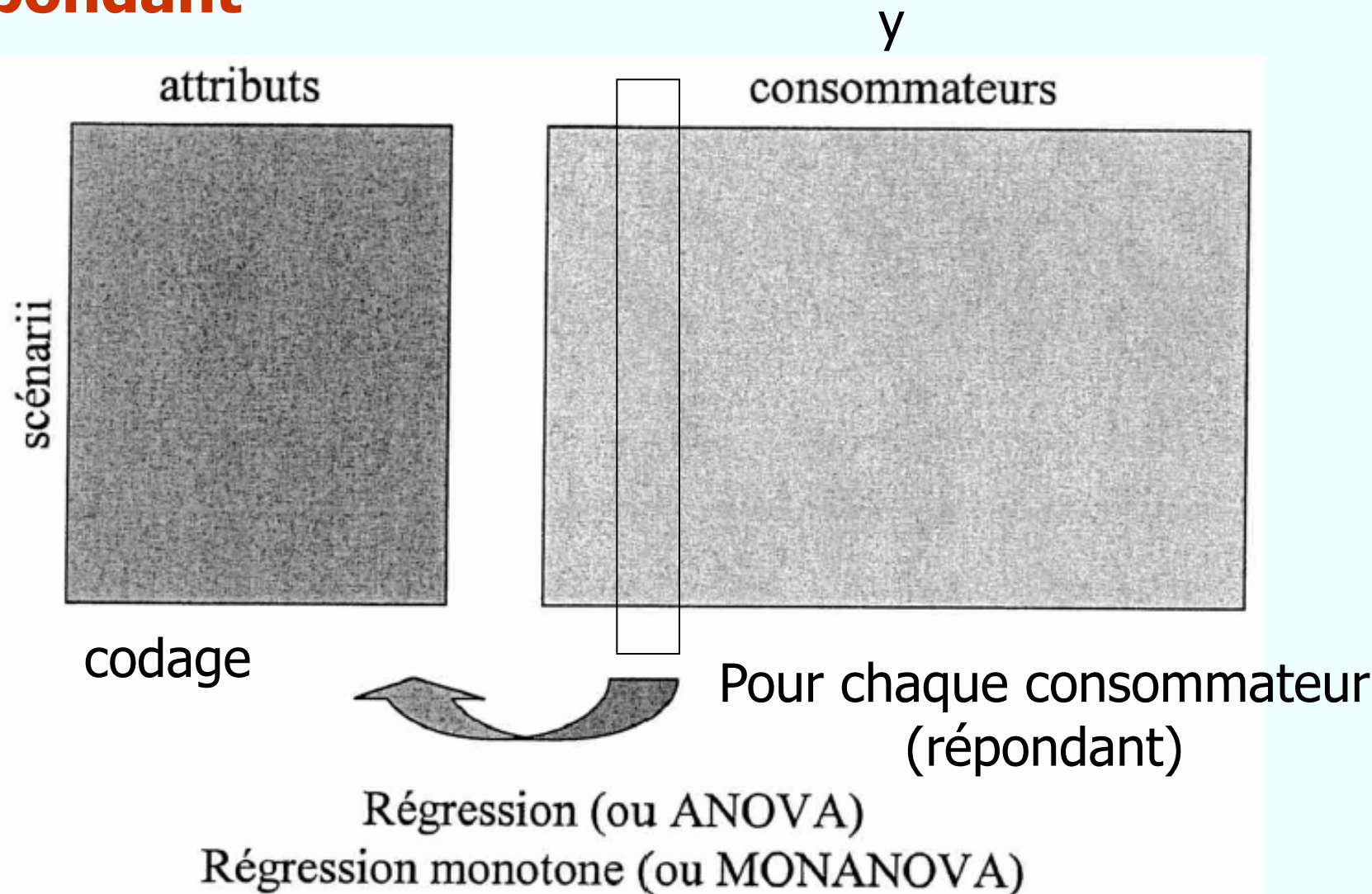
U_i : cotation de préférence du produit i , donnée par le sujet

- Lorsque U_i est une variable quantitative, le modèle correspond à l'Analyse de la variance (ANOVA)
- Lorsque U_i est une variable de rang, une transformation monotone est nécessaire (MONANOVA – monotonic ANOVA)

Dans le cas d'un modèle « rating based », on peut estimer un modèle pour chaque répondant (individuel)



Estimation des modèles pour chaque répondant





Etape 2 : estimation du modèle « choice »

- Cas du modèle « choice based » (multinomial Logit)
- La probabilité de choisir le produit « i » dans l'ensemble J est donnée par :

$$P_i = \frac{e^{V_i}}{\sum_{k \in J} e^{V_k}}$$

Le log de la vraisemblance est donné par : $LL = \sum_i n_i \cdot \ln(P_i)$

Avec : n_i : nombre de répondants qui ont choisi le produit i

Les utilités partielles a_{ik} sont données par maximisation de la log.vraisemblance

$$a_{ik} = \arg \max(LL)$$

Dans le cas d'un modèle « choice based », le modèle est estimé globalement, pour tous les répondants



Etape 3 : utilisation du modèle (rating based)

- pour chaque répondant, on peut estimer:
 - La **valeur** qu'il attribue à chaque modalité d'un attribut
 - Le **produit idéal** (optimal du point de vue de la préférence) en déterminant les modalités qui ont la plus grande utilité partielle
 - Sa **sensibilité** aux différents attributs (prix, marque, couleur, ...)

⇒ Utilité partielle a_{ik}

⇒ Importance I_j de l'attribut « j »

$$I_j = \frac{\max_k(a_{jk}) - \min_k(a_{jk})}{\sum_{i=1}^p (\max_k(a_{ik}) - \min_k(a_{ik}))}$$



Rappel Régression linéaire multiple



Méthodes explicatives

		<i>Variables à expliquer y</i>	
		quantitatives	qualitatives
<i>Variables explicatives x</i>	quantitatives	Régression linéaire Analyse canonique	Analyse discriminante
	qualitatives	Analyse de la variance segmentation Analyse conjointe	Analyse conjointe



Regression linéaire multiple

- Écriture matricielle

$$y = X.\alpha + \varepsilon$$

$$y_i = \begin{bmatrix} 1 \\ 1 \\ 1 & x_{i1} & \dots & x_{ip} \\ 1 \\ 1 \end{bmatrix} \times \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \vdots \\ \alpha_p \end{bmatrix} + \varepsilon_i$$

- Les coefficients $(a_0, a_1, \dots, a_p)$, estimateurs des $(\alpha_0, \alpha_1, \dots, \alpha_p)$, qui minimisent les résidus, sont donnés par :

$$a = (X^t.X)^{-1}.X^t.y$$



Regression linéaire multiple

- ◆ $X(n, p+1)$: tableau des variables indépendantes VI
- ◆ $Y(n)$: vecteur de la variable dépendante VD
- ◆ La matrice $(X^t . X)$ doit être inversible (de rang plein)
- ◆ La solution est instable si les VI sont très corrélées
 - grande incertitude sur les coefficients de la régression
 - Pouvoir prédictif faible
- ◆ Solution : décorréler les VI
 - Régression régularisée (régression sur les composantes principales)



Régression linéaire multiple

◆ On note

$SCT = \sum (y_i - \bar{y})^2$ Somme des carrés totale (variance totale)

$SCR = \sum (\hat{y}_i - \bar{y})^2$ Somme des carrés expliquée par le modèle (variance expliquée)

$SC_{RES} = \sum (y_i - \hat{y}_i)^2$ Variance résiduelle

Décomposition de la variance :

$$SCT = SCR + SC_{RES}$$



Régression linéaire multiple

- ◆ Qualité d'ajustement du modèle : le coefficient de détermination R^2

R^2 représente le **pourcentage de variance capté par le modèle**

$$R^2 = \frac{\text{variance expliquée}}{\text{variance totale}} = 1 - \frac{SC_{RES}}{SCT} = \frac{SCR}{SCT} = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2}$$

- Il est compris entre 0 (le modèle n'explique pas mieux que $y = \bar{y}$) et 1 (toute la variance est expliquée par le modèle)
- R^2 indicateur de qualité uniquement si le nombre d'essais est supérieur au nombre de paramètres à estimer.

Le R^2 n'est pas un « bon » indicateur

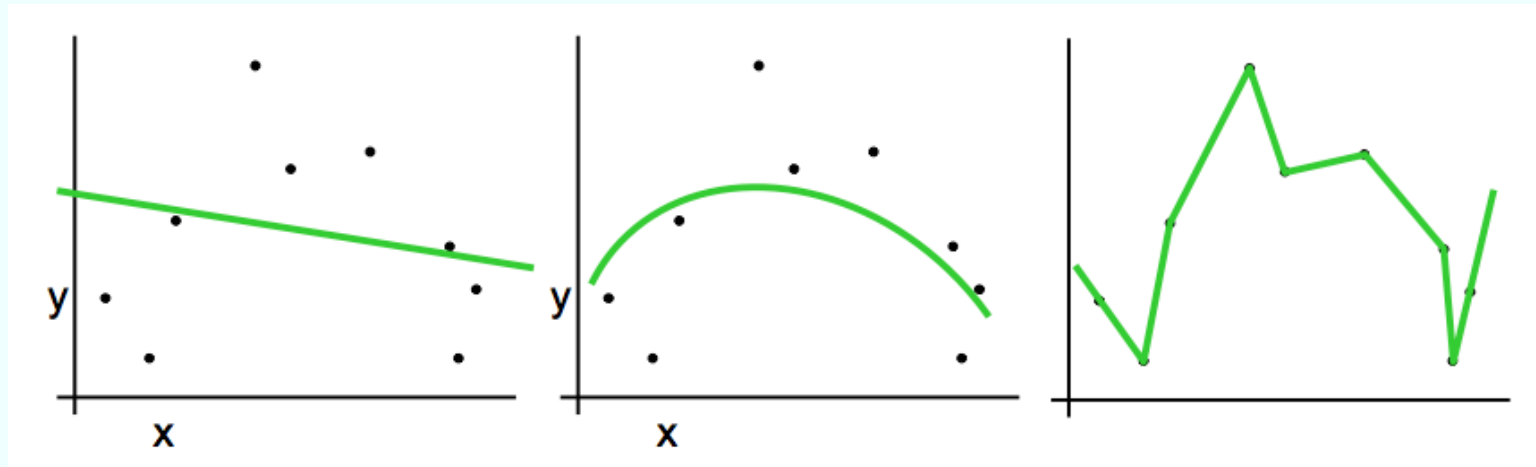
-Quelques soient les VI ajoutées au modèle, le R^2 ne peut que croître

Un « meilleur » indicateur : le R^2 ajusté

-On pénalise les modèles qui ont beaucoup de paramètres (p)

$$R_a^2 = 1 - \frac{SC_{RES}/(n-p-1)}{SCT/(n-1)}$$

Validation d'un modèle



$$\hat{y} = a_0 + a_1x$$

$$\hat{y} = a_0 + a_1x + a_2x^2$$

Relier les points

Quelle méthode est la meilleure pour modéliser le système ?



Validation d'un modèle

- Pour valider un modèle, on teste sa capacité à la **généralisation**, c'est à dire à prédire des valeurs pour des observations qui n'ont pas servi à le construire (données externes)
- Plusieurs méthodes de validation
 - Validation croisée (2-fold ou k-fold)
 - On découpe les observation en k groupes : on apprend sur un groupe et on calcule l'erreur de prédiction sur l'autre.
 - « leave one out cross validation » (LOOCV équivalent à n-fold)
 - On écarte une observation pour construire le modèle et on calcule l'erreur de prédiction pour cette observation
 - On répète cela n fois

La recherche d'un modèle explicatif fiable est un compromis entre **bon ajustement sur les données** et **bonne capacité à généraliser**



Validation croisée

- Résidus :

$$e_{(i)i} = y_i - \hat{y}_{(i)i}$$

$$\hat{y}_{(i)i}$$

Prévision de y_i calculée
avec un modèle estimé
sans la $i^{\text{ème}}$ observation

- PRESS (Predicted residual sum of squares) :

$$PRESS = \frac{1}{n} \sum_{i=1}^n e_{(i)i}^2$$

C'est une estimation de la qualité
de prédiction d'un modèle

$$PRESS_{RMCE} = \sqrt{PRESS}$$

Une comparaison du $PRESS_{RMCE}$ et du
RMCE renseigne sur la sensibilité du
modèle aux observations

$$RMCE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Un RMCE faible et un $PRESS_{RMCE}$ grand
indique un surajustement du modèle sur les
données



Validation d'un modèle

Qualité de prédiction du modèle : le coefficient de détermination Q^2 en cross validation

Il représente le R^2 en cross validation :

$$Q^2 = R_{CV}^2 = 1 - \frac{PRESS}{SCT}$$

$$PRESS = \sum_i (y_i - \hat{y}_{i/i})^2$$

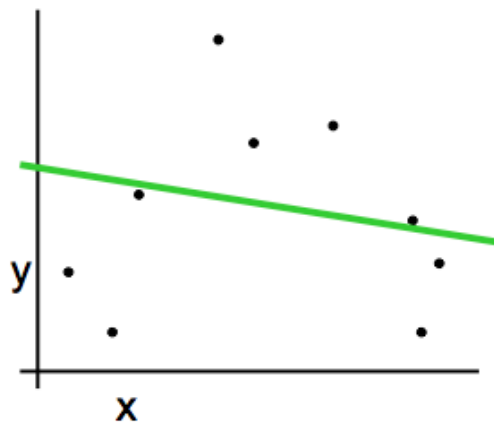
$\hat{y}_{i/i}$: Représente la valeur de y_i prédite par le modèle lorsque l'observation i est retirée pour l'estimation du modèle

- Q^2 est borné par 1
- Il peut être négatif
- Il représente la capacité du modèle à généraliser (prédire correctement lorsqu'on lui présente de nouvelles données)

Une différence importante entre le R^2 et le Q^2 peut être le signe d'un **surajustement** du modèle



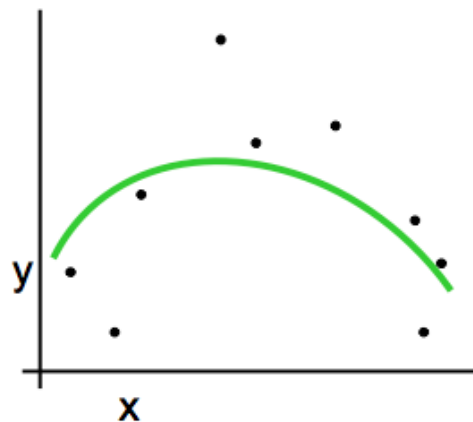
Validation d'un modèle



$$\hat{y} = a_0 + a_1x$$

$$R^2 = 0,67$$

$$Q^2 = 0,43$$

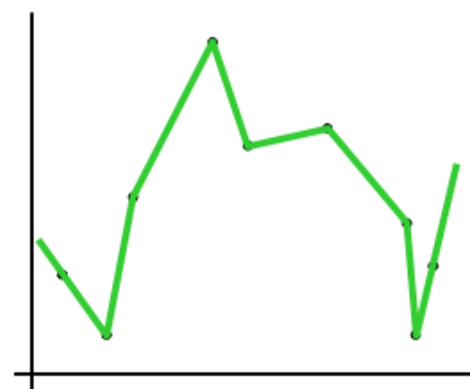


$$\hat{y} = a_0 + a_1x + a_2x^2$$

S'ajuste pas trop mal
Généralise bien

$$R^2 = 0,87$$

$$Q^2 = 0,64$$



Relier les points

S'ajuste parfaitement
Généralise très mal

$$R^2 = 1$$

$$Q^2 = 0,36$$

Le modèle quadratique est
le plus apte à généraliser



Significativité globale du modèle

F : statistique de Fisher

$$F = \frac{SCR_m/p}{SCR_{es}/(n-p-1)} \approx F_{\alpha}(p; n-p-1)$$

SCRes : somme des carrés des résidus

$$SCRes = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

SCRm : somme des carrés régression sans la moyenne

$$SCRm = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

◆ Test de Fisher

- Est-ce que la régression est significative dans son ensemble ?

H0 : les α sont tous nuls (sauf α_0)

- ◆ Si $F > F_{\alpha}(p; n-p-1)$: rejet H0 : régression significative
- ◆ Si $F \leq F_{\alpha}(p; n-p-1)$: accepter H0 : régression non significative au seuil α



Significativité des coefficients

- ◆ Test de Fisher généralisé
 - Modèle **complet** (avec p VI) .vs. Modèle **réduit** (avec r VI)

H_0 : les coefficient α pour passer du modèle réduit au modèle complet sont tous nuls

$$F_{\text{partiel}} = \frac{(SCR_p - SCR_r)/(p - r)}{SCR_{\text{res}}/(n - p - 1)} \approx F_{\alpha}(p - r; n - p - 1)$$

- ◆ Test
 - Est-ce que les $(p-r)$ VI apportent une contribution significative ?

- ◆ Si $F > F_{\alpha}(p-r; n-p-1)$: rejet H_0 (les $(p-r)$ VI apportent une contribution significative)
- ◆ Si $F \leq F_{\alpha}(p-r; n-p-1)$: accepter H_0



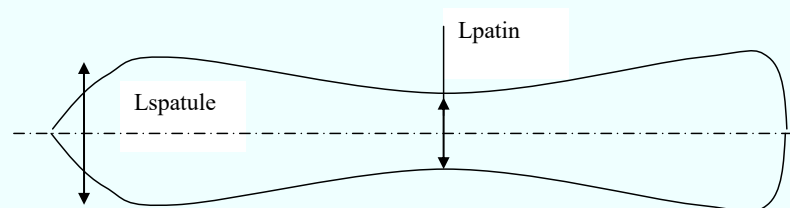
Sélection optimale de variables

- Construction du modèle « optimal »
 - Optimisation d'un critère (choisir le MCE ou le R^2 ajusté, jamais le R^2)
 - Imposer le nombre min et max de variables : le logiciel teste toutes les combinaisons possibles $\Rightarrow$ très long
 - Méthode descendante (à conseiller)
 - » On commence avec toutes les variables, on calcule le F correspondant au retrait de chaque variable, on retire celle qui a le F le plus faible
 - Méthode ascendante
 - » On part d'un petit modèle et on introduit à chaque pas la variable qui a le F d'introduction le plus élevé
 - Algorithme “pas à pas”
 - on applique en alternance une itération de sélection avant et une itération d'élimination arrière. On arrête lorsqu'on ne peut ajouter une variable ni en éliminer une.

Exemple : les skis

- On s'intéresse à deux propriétés subjectives des skis :
 - l'accroche (capacité à accrocher dans la neige)
 - La maniabilité
- On considère 2 caractéristiques géométriques du ski :
 - Lspatule
 - Taille de guêpe

$$taille_de_guêpe = (Lspatule - Lpatin) / 2$$



Coefficient de corrélation critère objectif/descripteur

	Largeur patin	Taille de guepe
Accroche	-0,077	0,882
Maniable	0,926	0,030

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x}) * (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$



Exemple : les skis

- ♦ variable à expliquer Y : **accroche**
- ♦ Variables explicatives X : x1 : Lspatule ; x2 : Taille de guêpe
- ♦ Plan d'expériences :

	Largeur patin	Taille de guepe	Accroche
ski1	72	115	7,40
ski2	75	102	4,63
ski3	65	93	2,92
ski4	76	105	6,59
ski5	62	114	9,41
ski6	75	104	7,10
ski7	68	105	5,02



Exemple : les skis

- ♦ variable à expliquer Y : accroche
- ♦ Variables explicatives X : x1 : Lspatule ; x2 : Taille de guêpe
- ♦ Modèle :

$$Y_i = a_0 + a_1 x1_i + a_2 x2_i + \varepsilon_i$$

: R2 = 77,8%

Analyse de la
variance

Source	DDL	Somme des carrés	Moyenne des carrés	F	Pr > F
Modèle	2	21,255	10,628	7,022	0,049
Erreur	4	6,054	1,513		
Total corrigé	6	27,309			

Calculé contre le modèle Y=Moyenne(Y)

- ♦ P-value = 4,9% < 5% : rejet de H0 : régression significative dans son ensemble



Exemple : les skis

Analyse Type I Sum of Squares :

Source	DDL	Somme des carrés	Moyenne des carrés	F	Pr > F
Largeur patin	1	0,161	0,161	0,107	0,760
Taille de guepe	1	21,094	21,094	13,938	0,020

Test de Fisher généralisé

- ◆ Pour **Largeur patin** : $F \ll F_{5\%}$ (p-value = 76%) : accepter H_0 : largeur patin n'apporte pas une contribution significative
- ◆ Pour **taille de guepe** : $F \gg F_{5\%}$: rejet H_0 : taille de guepe apporte une contribution significative



L'ANOVA (Analysis of variance)

Cas particulier des variables explicatives X qualitatives

Variable qualitative : ne peut prendre qu'un nombre discret de valeurs (modalités ou niveaux). *Exemple la couleur*

Cas d'un facteur : le modèle est donné par :

$$y_{ik} = \mu + \alpha_i + \varepsilon_{ik} \quad \varepsilon_{ik} \sim \mathcal{N}(0, \sigma)$$

y_{ij} : réponse associée à l'observation k pour le niveau i du facteur

α_i : effet moyen du niveau i du facteur

μ : effet moyen général

On choisit d'avoir $\sum_i \alpha_i = 0$: paramétrage centré

- Écriture matricielle $Y = X.a + e$

- Estimation des coefficients $\hat{a} = (X^t.X)^{-1}.X^t.Y$



L'ANOVA

L'Analyse de la variance (ANOVA) : cas particulier de la régression

Techniquement, on réalise la régression avec la matrice X correspondant au **tableau disjonctif** recodé des niveaux des facteurs

Tableau disjonctif : codage d'une variable qualitative avec des 1 et 0 (1 si présent, 0 si absent)

- Exemple : cas de deux facteurs

cas	F1 : bracelet	F2 : type
1	métal	aiguille
2	cuir	aiguille
3	plastique	aiguille
4	métal	digitale
5	cuir	digitale
6	plastique	digitale

Tableau disjonctif complet

	F2 : type		F1 : Bracelet		
cas	aiguille	digital	métal	cuir	plastique
1	1	0	1	0	0
2	1	0	0	1	0
3	1	0	0	0	1
4	0	1	1	0	0
5	0	1	0	1	0
6	0	1	0	0	1

L'ANOVA

Tableau disjonctif recodé

On intègre la contrainte $\sum \alpha_i = 0$ pour chaque facteur en retranchant arbitrairement une modalité à toutes les autres

- Exemple

- Facteur F2 : codage de (a-d)
- Facteur F1 : codage de (m-p) et (c-p)

N.B.: ce choix arbitraire n'a aucune incidence sur le résultat

tableau disjonctif					
cas	F2 : type		F1 : bracelet		
	aiguille (a)	digital (d)	metal (m)	cuir c	plastique (p)
1	1	0	1	0	0
2	1	0	0	1	0
3	1	0	0	0	1
4	0	1	1	0	0
5	0	1	0	1	0
6	0	1	0	0	1



tableau recodé			
cas	F2 : type	F1 : bracelet	
	a-d	m-p	c-p
1	1	1	0
2	1	0	1
3	1	-1	-1
4	-1	1	0
5	-1	0	1
6	-1	-1	-1

On réalise ensuite une régression de la réponse sur le tableau disjonctif recodé



L'ANOVA

Écriture matricielle de la régression sur variables qualitatives

$$y = X.a + e$$

$$y_i = \begin{array}{c|ccc} & 1 & F1 & F2 \\ \hline 1 & 1 & 1 & 0 & 1 \\ 2 & 1 & 0 & 1 & -1 \\ 3 & 1 & -1 & -1 & 1 \\ 4 & 1 & -1 & -1 & -1 \\ 5 & 1 & 1 & 0 & -1 \\ 6 & 1 & 0 & 1 & 1 \end{array} \times \begin{array}{c} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \end{array} + e_i$$

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



L'ANOVA : exemple

Données

exercice1

Réponse	Facteurs	
preference y	Bracelet	cadran
1	métal	aiguille
7	cuir	digitale
1	plastique	aiguille
9	plastique	digitale
6	métal	digitale
2	cuir	aiguille
2	métal	aiguille
6	cuir	digitale
2	plastique	aiguille
8	plastique	digitale
8	métal	digitale
2	cuir	aiguille

Coefficients d'ajustement :

Observations	12,000
Somme des poids	12,000
DDL	8,000
R²	0,932
R ² ajusté	0,906
MCE	0,896
RMCE	0,946
MAPE	26,498
DW	2,647
Cp	4,000
AIC	1,814
SBC	3,754
PC	0,137
<u>Press RMCE</u>	1,420

Source	Valeur
Constante	4,500
Bracelet-m	-0,250
Bracelet-c	-0,250
Bracelet-p	0,500
cadran-a	-2,833
cadran-d	2,833

Equation du modèle :

$$\text{preference y} = 4,5 - 0,25 \cdot \text{Bracelet-m} - 0,25 \cdot \text{Bracelet-c} + 0,5 \cdot \text{Bracelet-p} - 2,83 \cdot \text{cadran-a} + 2,83 \cdot \text{cadran-d}$$



L'ANOVA : exemple

Données

ANALYSE DE LA VARIANCE

Source	DDL	Somme des carrés	Moyenne des carrés	F	Pr > F
Modèle	3	97,833	32,611	36,403	< 0,0001
Erreur	8	7,167	0,896		
Total corrigé	11	105,000			

Analyse Type III Sum of Squares :

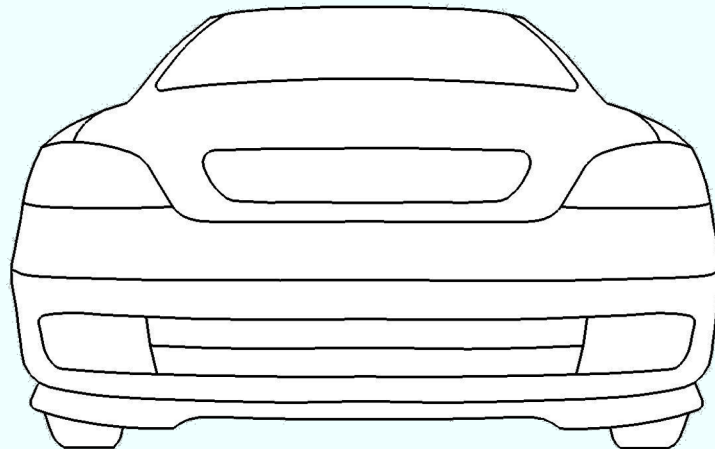
Source	DDL	Somme des carrés	Moyenne des carrés	F	Pr > F
Bracelet	2	1,500	0,750	0,837	0,468
cadran	1	96,333	96,333	107,535	< 0,0001

← Régression significative dans son ensemble

← Seul le critère « cadran » est significatif



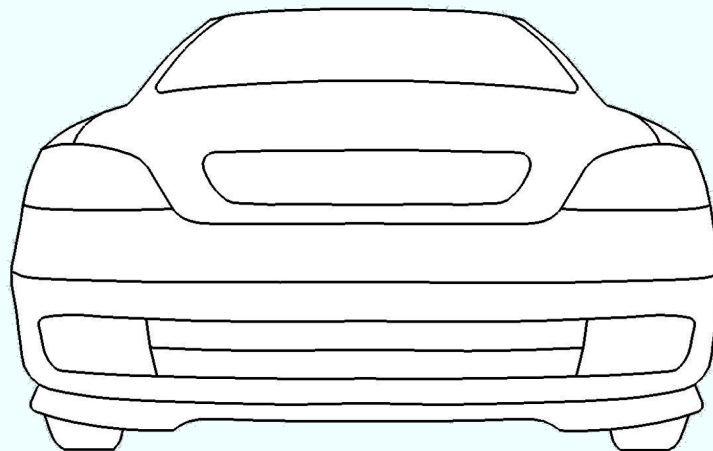
Analyse conjointe - Exemple



Exemple : face avant de véhicule

◆ Motivation

- Faire évoluer le design (restylage) ou créer de nouveaux designs
- Définir un modèle quantitatif entre les perceptions utilisateur (ou la préférence) et les attributs de design



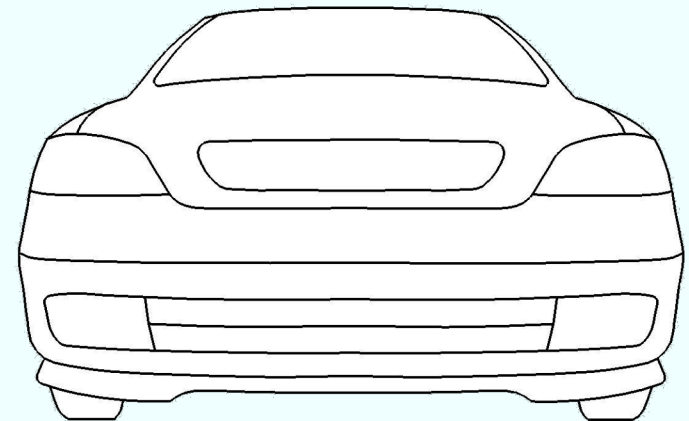
La méthode
ne remplace
pas le
designer

Analyse conjointe

Évaluer les produits

- Les produits sont vus comme un “panier” d’attributs
- Les clients attribuent une valeur à ces attributs
- La valeur totale du produit est fonction de la valeur de ses attributs

On détermine les attributs pertinents après une recherche qualitative

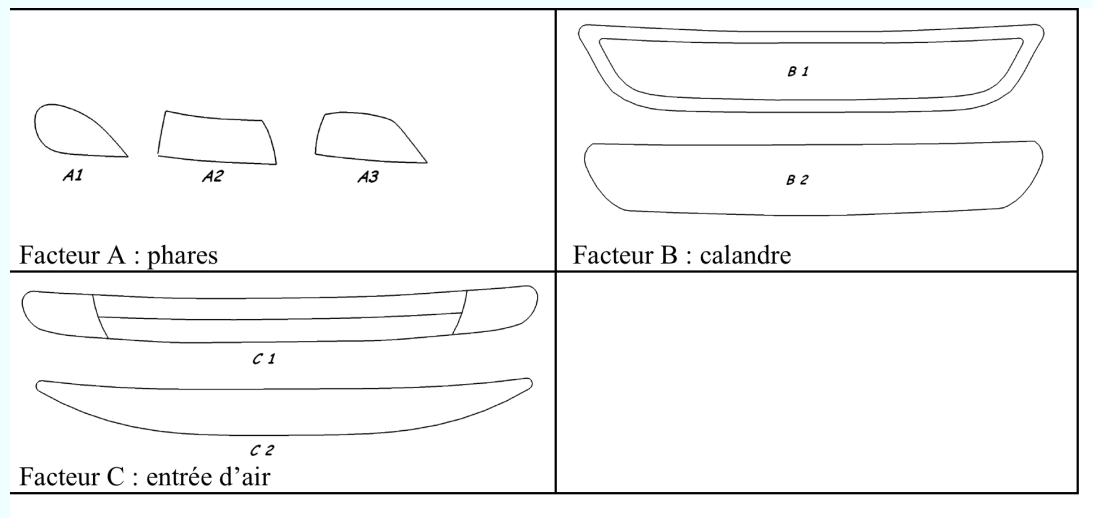




Exemple 1 : face avant

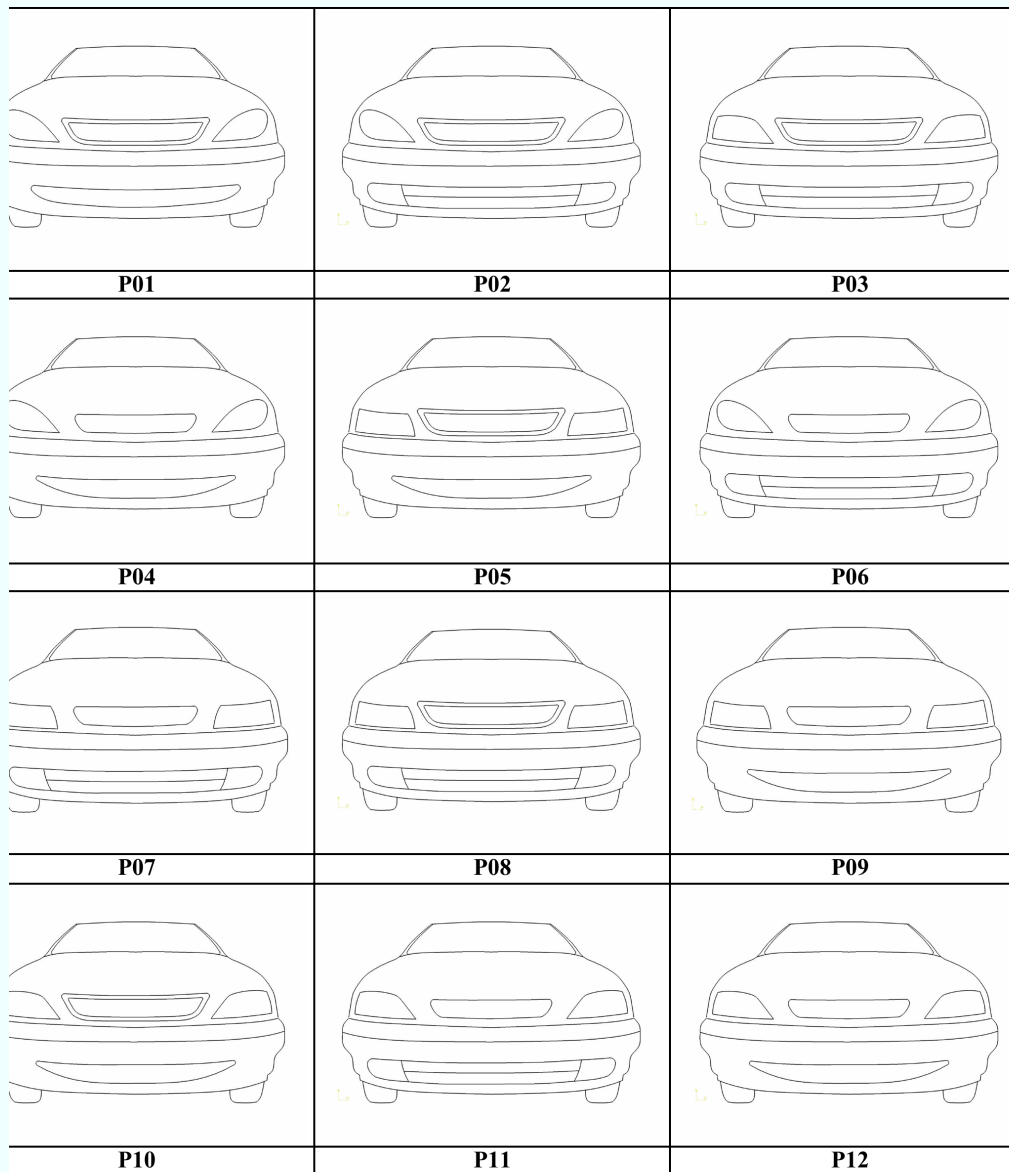
- ◆ Attribut A : phare
 - Niveaux : A1, A2, A3
- ◆ Attribut B : radiateur
 - Niveaux : B1, B2
- ◆ Attribut C : entrée d'air
 - Niveaux : C1, C2

$3 \times 2 \times 2 = 12$ designs





Example 1 : face avant



$3*2*2=12$ design

Plan complet



Définition du plan d'expérience

- Plan complet

#	A Headlights	B Radiator grill	C Air intake	Preference Rating : $u(X)$
1	A1	B1	C1	3
2	A1	B2	C1	4
3	A1	B1	C2	5
4	A1	B2	C2	4
5	A2	B1	C1	3
6	A2	B2	C1	2
7	A2	B1	C2	8
8	A2	B2	C2	7
9	A3	B1	C1	5
10	A3	B2	C1	3
11	A3	B1	C2	7
12	A3	B2	C2	8

On demande au **répondant** une cotation de la **préférence** de 1 "très peu probable que j'achète" à 10 "très probable que j'achète".



Le modèle en analyse conjointe

- ◆ On choisit un modèle de préférence du type :

$$\hat{y}_i = \mu + \sum_{j=1}^p \sum_{k=1}^{k_j} a_{jk} \cdot \delta_i(jk)$$

Modèle linéaire
compensatoire

k_j = nombre de niveaux du
facteur j

p = nombre de facteurs

$\hat{y}_i$ utilité globale du design i

a_{jk} : utilité partielle du niveau k pour le facteur j

$$\delta_i(jk) \begin{cases} = 1 \text{ si niveau du facteur } j \text{ pour le design } i \text{ est } k \\ = 0 \text{ sinon} \end{cases}$$

a_{jk} est le coefficient d'une régression linéaire multiple, déterminé par la minimisation de :

$$\text{MIN} \left(\sum_{i=1}^n (y_i - \hat{y}_i)^2 \right)$$

Exemple : face avant

Codage : Tableau disjonctif complet

	A			B		C	
	A1	A2	A3	B1	B2	C1	C2
y1	1	0	0	1	0	1	0
y2	1	0	0	1	0	0	1
y3	1	0	0	0	1	1	0
y4	1	0	0	0	1	0	1
y5	0	1	0	1	0	1	0
y6	0	1	0	1	0	0	1
y7	0	1	0	0	1	1	0
y8	0	1	0	0	1	0	1
y9	0	0	1	1	0	1	0
y10	0	0	1	1	0	0	1
y11	0	0	1	0	1	1	0
y12	0	0	1	0	1	0	1



P4: A1, B2, C2

◆ Design P4 :

$$\delta_3(11) =$$

$$\delta_3(13) =$$

$$\delta_3(12) =$$



Détermination des utilités partielles a_{jk}

- ◆ On utilise l'ANOVA (analysis of variance)
- ◆ Seules les différences entre coefficients pour un même facteur sont estimables

$$\hat{y}_i = \mu + \sum_{j=1}^p \sum_{k=1}^{k_j} a_{jk} \cdot \delta_i(jk)$$

- ◆ Il faut se fixer une référence pour le calcul des utilités partielles

Niveau 1 du
facteur i
nulle

- ◆ Solution 1 : on choisit arbitrairement un produit de référence d'utilité nulle

$$a_{i1} = 0$$

- ◆ Solution 2 : on impose d'avoir la somme des utilités partielles nulles, pour chaque facteur

$$\sum_j a_{ij} = 0$$

Paramétrage centré

- ◆ => cf balle de golf

Utiliser cette
solution



Estimation des utilités partielles a_{jk} : résultat

1. Utilités partielles

1. Ce sont directement les coefficients du modèle ANOVA

$$\sum_k a_{jk} = 0$$

$$R^2 = 77\%$$

Design factor	niveau	Utilité partielle	Amplitude A_j	Importance I_j
Phare	A1	$a_{11}=-0,91$	1,75	32 %
	A2	$a_{12}=0.08$		
	A3	$a_{13}=0,83$		
Radiateur	B1	$a_{21}=0,25$	0.5	9 %
	B2	$a_{22}=-0,25$		
Entrée air	C1	$a_{31}=-1.58$	3,16	59 %
	C2	$a_{32}=1.58$		

On choisit (solution 2)

$$\sum_k a_{jk} = 0$$

Pour tous les facteurs



Calcul des amplitudes A_j

1. Tableau de synthèse

Design factor	niveau	Utilité partielle	Amplitude A_j	Importance I_j
Phare	A1	$a_{11}=-0,91$	1,75	32 %
	A2	$a_{12}=0.08$		
	A3	$a_{13}=0,83$		
Radiateur	B1	$a_{21}=0,25$	0.5	9 %
	B2	$a_{22}=-0,25$		
Entrée air	C1	$a_{31}=-1.58$	3,16	59 %
	C2	$a_{32}=1.58$		

◆ Amplitude

- Représente la contribution du facteur à la préférence

$$A_j = \underset{i}{MAX}(a_{ji}) - \underset{i}{MIN}(a_{ji})$$



Calcul des importances I_j

1. Tableau de synthèse

Design factor	niveau	Utilité partielle	Amplitude A_j	Importance I_j
Phare	A1	$a_{11}=-0,91$	1,75	32 %
	A2	$a_{12}=0,08$		
	A3	$a_{13}=0,83$		
Radiateur	B1	$a_{21}=0,25$	0.5	9 %
	B2	$a_{22}=-0,25$		
Entrée air	C1	$a_{31}=-1.58$	3,16	59 %
	C2	$a_{32}=1.58$		

◆ Importance, I_j

- Représente la sensibilité du client au facteur j

$$I_j = \frac{A_j}{\sum_{j=1}^p A_j}$$

$$59\% = \frac{3,16}{1,75 + 0,5 + 3,16}$$



Ajustement du modèle sur les données

$$R^2 = \frac{\text{variation expliquée}}{\text{variation totale}} = \frac{SCR}{SCT} = \frac{\sum (\hat{U}_i - \bar{U})^2}{\sum (U_i - \bar{U})^2}$$

◆ R^2 : coefficient de détermination

- Représente la qualité de l'ajustement du modèle (%)

- $R^2 > 0.9$: très bon ajustement
- $R^2 > 0.8$: bon ajustement
- $R^2 < 0.6$: mauvais ajustement

Cf. test stat.
sur R^2

◆ Si le R^2 est faible

- Il est inutile d'interpréter les coefficients du modèle – le modèle n'explique pas les préférences
- R^2 faible
 - Le sujet utilise d'autres attributs que ceux proposés pour fonder sa préférence
 - Le modèle du sujet est plus complexe qu'un modèle linéaire simple avec effets simples
 - Le sujet a répondu n'importe quoi aux tests...



Interprétation des résultats

Coefficient de détermination $R^2 = 77\%$

Design factor	niveau	Utilité partielle	Amplitude Aj	Importance Ij
Phare	A1	$a_{11}=-0,91$	1,75	32 %
	A2	$a_{12}=0,08$		
	A3	$a_{13}=0,83$		
Radiateur	B1	$a_{21}=0,25$	0.5	9 %
	B2	$a_{22}=-0,25$		
Entrée air	C1	$a_{31}=-1.58$	3,16	59 %
	C2	$a_{32}=1.58$		

- ◆ 77% de la variance de la préférence est expliquée par les 3 facteurs => on considère que c'est suffisant pour interpréter le modèle
- ◆ La forme de l'entrée d'air à une importance primordiale pour expliquer la préférence (59%), puis la forme du phare (32%), puis le radiateur (9%)
- ◆ Le design « idéal » est constitué des modalités A3 (+0,83), B1 (+0,25) et C2 (+1,58)
- ◆ Le design « anti-idéal » est constitué des modalités A1 (-0,91), B2 (-0,25) et C1 (-1,58)



Pouvoir prédictif du modèle

- ◆ En AC, évalué par des designs en « holdout »
 - Designs qui sont évalués par le répondant, mais qui ne servent pas à créer le modèle (2-fold cross validation)
- ◆ Calcul de différents indices de performance
 - **Hit rate** : % de cas pour lesquels le meilleur design (évalué par le répondant) correspond au meilleur design (donné par le modèle)
 - Calcul du **PRESS** (ou du Q^2)
 - Coefficient de corrélation de Pearson ou coefficient de Kendall entre les le modèle et les évaluations
 - Mean Absolute share error (**MASE**)

$$MASE = \frac{\sum_{k=1}^K |S_{j.Pred} - S_{j.Obs}|}{K}$$

$S_{k.Pred}$: part de marché prédite

$S_{k.Obs}$: part de marché observée

K : nombre de produits dans le holdout

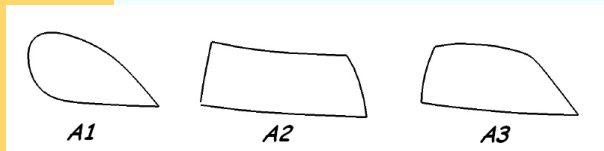
Exemple plus complet

Génération de l'espace produit

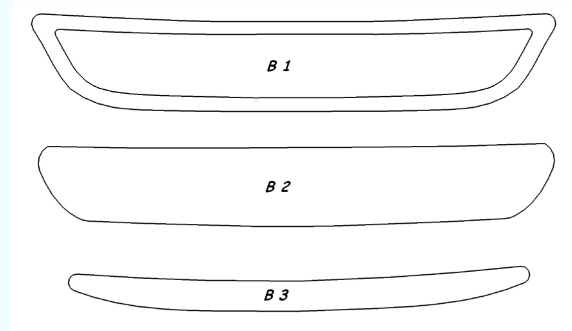
- Définition de 4 principaux facteurs et de leurs niveaux :
 - Facteur A: phares– 3 niveaux
 - Facteur B: Radiateur – 3 niveaux
 - Facteur C: entrée air– 3 niveaux
 - Facteur D: parechoc– 2 niveaux



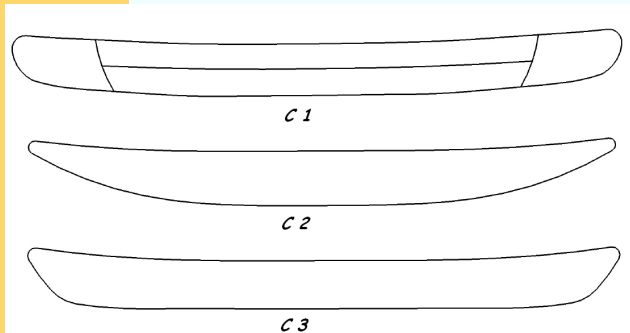
**54 produits
différents**



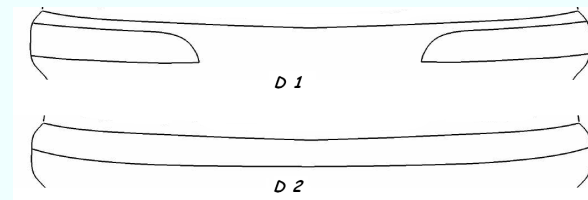
*Factor A:
Headlights*



*Factor B:
Radiator Grill*



*Factor C:
Air Intake*



*Factor D:
Front Bumper*



Exemple plus complet

Latin square

Plan fractionnaire

#	1 Headlights	2 Radiator grill	3 Air intake	bumper	Rating : $u(X)$
1	A1	B1	C1	D1	8
2	A1	B2	C2	D1	7
3	A1	B3	C3	D1	9
4	A2	B1	C2	D1	7
5	A2	B2	C3	D1	2
6	A2	B3	C1	D1	1
7	A3	B1	C3	D1	3
8	A3	B2	C1	D1	2
9	A3	B3	C2	D1	8
10	A1	B1	C1	D1	6
11	A1	B2	C2	D2	9
12	A1	B3	C3	D2	8
13	A2	B1	C2	D2	6
14	A2	B2	C3	D2	2
15	A2	B3	C1	D2	1
16	A3	B1	C3	D2	2
17	A3	B2	C1	D2	5
18	A3	B3	C2	D2	5



Modélisation de la préférence

Modèle analyse conjointe

- L'analyse de la variance est utilisée pour déterminer les utilités partielles des niveaux des facteurs, ainsi que l'importance des facteurs

Source	Valeur	Erreur standard	t	Pr > t	Borne inférieure (95%)	Borne supérieure (95%)
Constante	5,047	0,417	12,110	<0,0001	4,119	5,976
1 Headlights-A1	2,761	0,593	4,654	0,001	1,439	4,083
1 Headlights-A2	-1,881	0,587	-3,201	0,009	-3,190	-0,572
1 Headlights-A3	-0,881	0,587	-1,499	0,165	-2,190	0,428
2 Radiator grill-B1	0,261	0,593	0,440	0,669	-1,061	1,583
2 Radiator grill-B2	-0,547	0,587	-0,932	0,373	-1,856	0,762
2 Radiator grill-B3	0,286	0,587	0,487	0,637	-1,023	1,595
3 Air intake-C1	-1,239	0,593	-2,088	0,063	-2,561	0,083
3 Air intake-C2	1,953	0,587	3,324	0,008	0,644	3,262
3 Air intake-C3	-0,714	0,587	-1,215	0,252	-2,023	0,595
bumper-D1	0,074	0,433	0,172	0,867	-0,891	1,039
bumper-D2	-0,074	0,433	-0,172	0,867	-1,039	0,891

- ◆ coefficient de détermination du modèle

$$R^2 = 78\%$$

R^2 large : l'ajustement du modèle est correct



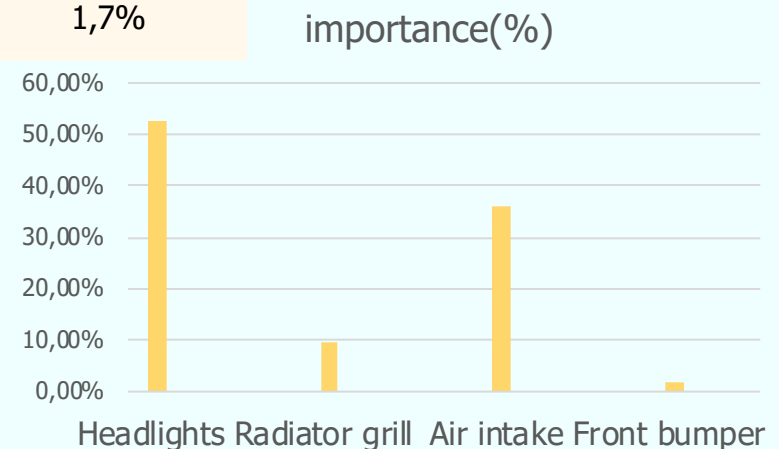
Modélisation de la préférence

Modèle analyse conjointe

- Calcul de l'importance des facteurs

Design factor	Factor level	Utility (part-worth)	A: amplitude	I: Importance
Headlights	A1	2,761	4,642	52,7%
	A2	-1,881		
	A3	-0,881		
Radiator grill	B1	0,261	0,833	9,5%
	B2	-0,547		
	B3	0,286		
Air intake	C1	-1,239	3,191	36,2%
	C2	1,953		
	C3	-0,714		
Front bumper	D1	0,074	0,149	1,7%
	D2	-0,074		

“headlights” and “air intake ” are the 2 most important factors





Utilisation du modèle de préférence

- Calcul des utilités totale sur tout l'espace experimental

					total utility	partworth			
#	1 Headlights	2 Radiator grill	3 Air intake	bumper	u(X)	1 Headlights	2 Radiator grill	3 Air intake	bumper
1	A1	B1	C1	D1	6,905	2,761	0,261	-1,239	0,074
2	A1	B1	C2	D1	10,097	2,761	0,261	1,953	0,074
3	A1	B1	C3	D1	7,430	2,761	0,261	-0,714	0,074
4	A1	B2	C1	D1	6,097	2,761	-0,547	-1,239	0,074
5	A1	B2	C2	D1	9,288	2,761	-0,547	1,953	0,074
6	A1	B2	C3	D1	6,621	2,761	-0,547	-0,714	0,074
7	A1	B3	C1	D1	6,930	2,761	0,286	-1,239	0,074
8	A1	B3	C2	D1	10,121	2,761	0,286	1,953	0,074
9	A1	B3	C3	D1	7,455	2,761	0,286	-0,714	0,074
10	A2	B1	C1	D1	2,263	-1,881	0,261	-1,239	0,074
11	A2	B1	C2	D1	5,455	-1,881	0,261	1,953	0,074
12	A2	B1	C3	D1	2,788	-1,881	0,261	-0,714	0,074
13	A2	B2	C1	D1	1,455	-1,881	-0,547	-1,239	0,074
14	A2	B2	C2	D1	4,646	-1,881	-0,547	1,953	0,074
15	A2	B2	C3	D1	1,979	-1,881	-0,547	-0,714	0,074
16	A2	B3	C1	D1	2,288	-1,881	0,286	-1,239	0,074
17	A2	B3	C2	D1	5,479	-1,881	0,286	1,953	0,074
18	A2	B3	C3	D1	2,813	-1,881	0,286	-0,714	0,074
19	A3	B1	C1	D1	3,263	-0,881	0,261	-1,239	0,074
20	A3	B1	C2	D1	6,455	-0,881	0,261	1,953	0,074
21	A3	B1	C3	D1	3,788	-0,881	0,261	-0,714	0,074
22	A3	B2	C1	D1	2,455	-0,881	-0,547	-1,239	0,074
23	A3	B2	C2	D1	5,646	-0,881	-0,547	1,953	0,074
24	A3	B2	C3	D1	2,979	-0,881	-0,547	-0,714	0,074
25	A3	B3	C1	D1	3,288	-0,881	0,286	-1,239	0,074
26	A3	B3	C2	D1	6,479	-0,881	0,286	1,953	0,074
27	A3	B3	C3	D1	3,813	-0,881	0,286	-0,714	0,074



Utilisation du modèle de préférence

- Calcul des utilités totale sur tout l'espace experimental

					total utility	partworth			
#	1 Headlights	2 Radiator grill	3 Air intake	bumper	u(X)	1 Headlights	2 Radiator grill	3 Air intake	bumper
28	A1	B1	C1	D2	6,756	2,761	0,261	-1,239	-0,074
29	A1	B1	C2	D2	9,948	2,761	0,261	1,953	-0,074
30	A1	B1	C3	D2	7,281	2,761	0,261	-0,714	-0,074
31	A1	B2	C1	D2	5,948	2,761	-0,547	-1,239	-0,074
32	A1	B2	C2	D2	9,139	2,761	-0,547	1,953	-0,074
33	A1	B2	C3	D2	6,473	2,761	-0,547	-0,714	-0,074
34	A1	B3	C1	D2	6,781	2,761	0,286	-1,239	-0,074
35	A1	B3	C2	D2	9,973	2,761	0,286	1,953	-0,074
36	A1	B3	C3	D2	7,306	2,761	0,286	-0,714	-0,074
37	A2	B1	C1	D2	2,115	-1,881	0,261	-1,239	-0,074
38	A2	B1	C2	D2	5,306	-1,881	0,261	1,953	-0,074
39	A2	B1	C3	D2	2,639	-1,881	0,261	-0,714	-0,074
40	A2	B2	C1	D2	1,306	-1,881	-0,547	-1,239	-0,074
41	A2	B2	C2	D2	4,497	-1,881	-0,547	1,953	-0,074
42	A2	B2	C3	D2	1,831	-1,881	-0,547	-0,714	-0,074
43	A2	B3	C1	D2	2,139	-1,881	0,286	-1,239	-0,074
44	A2	B3	C2	D2	5,331	-1,881	0,286	1,953	-0,074
45	A2	B3	C3	D2	2,664	-1,881	0,286	-0,714	-0,074
46	A3	B1	C1	D2	3,115	-0,881	0,261	-1,239	-0,074
47	A3	B1	C2	D2	6,306	-0,881	0,261	1,953	-0,074
48	A3	B1	C3	D2	3,639	-0,881	0,261	-0,714	-0,074
49	A3	B2	C1	D2	2,306	-0,881	-0,547	-1,239	-0,074
50	A3	B2	C2	D2	5,497	-0,881	-0,547	1,953	-0,074
51	A3	B2	C3	D2	2,831	-0,881	-0,547	-0,714	-0,074
52	A3	B3	C1	D2	3,139	-0,881	0,286	-1,239	-0,074
53	A3	B3	C2	D2	6,331	-0,881	0,286	1,953	-0,074
54	A3	B3	C3	D2	3,664	-0,881	0,286	-0,714	-0,074



Utilisation du modèle de préférence

Détermination des meilleurs designs (classement par utilité décroissante)

					total utility	partworth			
#	1 Headlights	2 Radiator grill	3 Air intake	bumper	u(X)	1 Headlights	2 Radiator grill	3 Air intake	bumper
8	A1	B3	C2	D1	10,121	2,761	0,286	1,953	0,074
2	A1	B1	C2	D1	10,097	2,761	0,261	1,953	0,074
35	A1	B3	C2	D2	9,973	2,761	0,286	1,953	-0,074
29	A1	B1	C2	D2	9,948	2,761	0,261	1,953	-0,074
5	A1	B2	C2	D1	9,288	2,761	-0,547	1,953	0,074
32	A1	B2	C2	D2	9,139	2,761	-0,547	1,953	-0,074
9	A1	B3	C3	D1	7,455	2,761	0,286	-0,714	0,074
3	A1	B1	C3	D1	7,430	2,761	0,261	-0,714	0,074
36	A1	B3	C3	D2	7,306	2,761	0,286	-0,714	-0,074
30	A1	B1	C3	D2	7,281	2,761	0,261	-0,714	-0,074

- ◆ The “best” design is “8”, followed by “2” and “35”

« 8 » = (A1, B3, C2, D1)

- ◆ Total utility u(8)

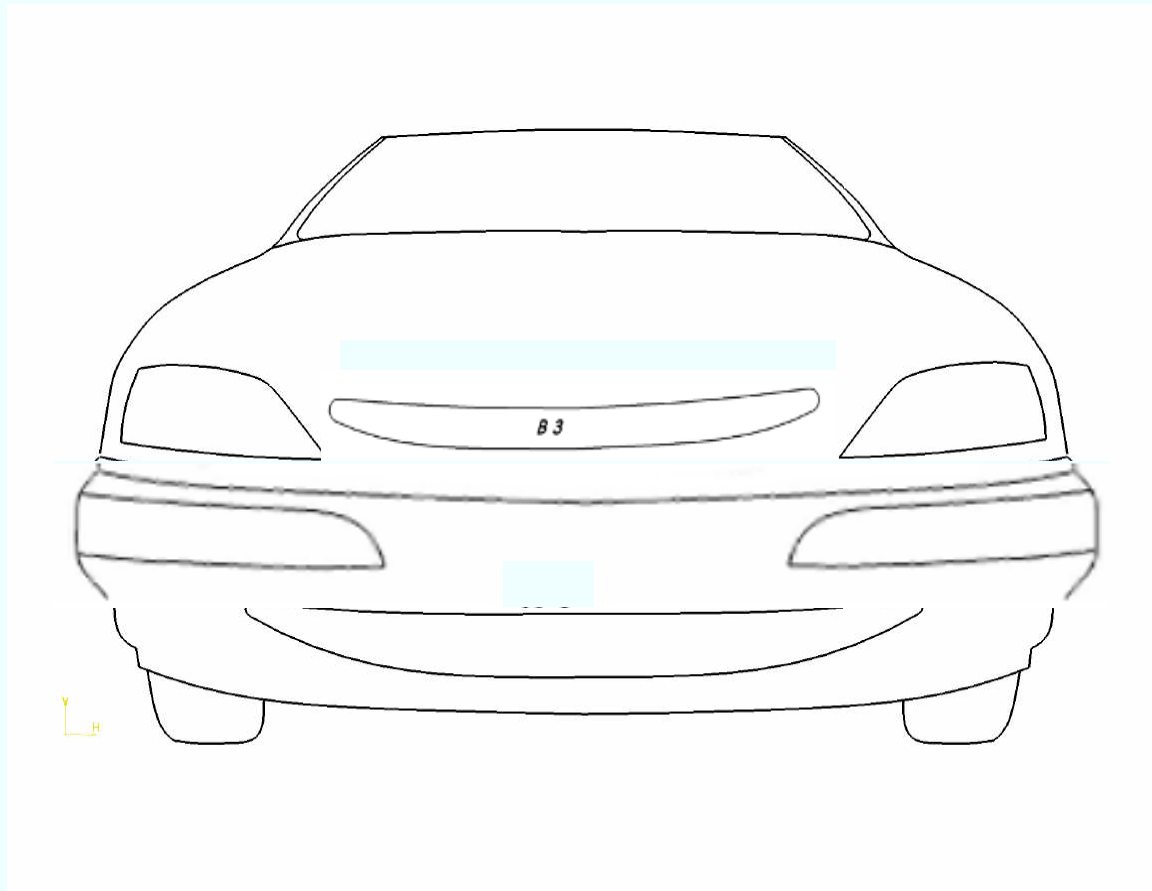
$$u(8) = 5,047 + 2,761(A1) + 0,286(B3) + 1,953(C2) + 0,074(D1) = 10,121$$



Utilisation du modèle de préférence

Profil du "meilleur design" ("8")

« 8 » = (A1, B3, C2, D1)





Utilisation du modèle de préférence

Détermination des " designs (classement par utilité décroissante)

					total utility	partworth			
#	1 Headlights	2 Radiator grill	3 Air intake	bumper	u(X)	1 Headlights	2 Radiator grill	3 Air intake	bumper
8	A1	B3	C2	D1	10,121	2,761	0,286	1,953	0,074
2	A1	B1	C2	D1	10,097	2,761	0,261	1,953	0,074
35	A1	B3	C2	D2	9,973	2,761	0,286	1,953	-0,074
29	A1	B1	C2	D2	9,948	2,761	0,261	1,953	-0,074
5	A1	B2	C2	D1	9,288	2,761	-0,547	1,953	0,074
32	A1	B2	C2	D2	9,139	2,761	-0,547	1,953	-0,074
9	A1	B3	C3	D1	7,455	2,761	0,286	-0,714	0,074
3	A1	B1	C3	D1	7,430	2,761	0,261	-0,714	0,074
36	A1	B3	C3	D2	7,306	2,761	0,286	-0,714	-0,074
30	A1	B1	C3	D2	7,281	2,761	0,261	-0,714	-0,074

- ◆ The "best" design is "8", followed by "2" and "35"

$$\ll 8 \gg = (A1, B3, C2, D1)$$

- ◆ Total utility $u(8)$

$$u(8) = 5,047 + 2,761(A1) + 0,286(B3) + 1,953(C2) + 0,074(D1) = 10,121$$



Autre exemple en design

Exemple : montre

- ◆ F_1 : Bracelet
 - Métal
 - Cuir
 - Plastique
- ◆ F_2 : Type
 - À aiguille
 - Digitale



cas	bracelet	type
1	métal	aiguille
2	cuir	aiguille
3	plastique	aiguille
4	métal	digitale
5	cuir	digitale
6	plastique	digitale



Exemple : les montres

- ◆ Cotation de la préférence sur le plan complet

cas	bracelet	type	préférence
1	métal	aiguille	1
2	cuir	aiguille	7
3	plastique	aiguille	1
4	métal	digitale	6
5	cuir	digitale	2
6	plastique	digitale	9

Exemple : les montres

◆ Résultat (ajustement du modèle)

- $R^2 = 0,62\%$: (62% de variance expliquée) ajustement du modèle moyen : les conclusions sont considérées comme fiables

• Utilités partielles

Scores				
bracelet			type	
cuir	métal	plastic	aiguille	digitale
+0,16	-0,83	+0,66	-3	+3

$$Y = 4,33 - 0,833 \text{ meta} + 0,167 \text{ cuir} + 0,667 \text{ plastic} + 3 \text{ digital} - 3 \text{ aiguille}$$

◆ Interprétation des utilité partielles

- Le design préféré est « plastic+digitale »
- Le design rejeté est « metal + aiguille »



Amplitude et importance des facteurs

attribute	level	utility	amplitude A_j	importance
F1 strap	Leather	0,167	1,5	20%
	metal	-0,833		
	plastic	0,667		
F2 Type	hand	-3	6	80%
	digital	3		

◆ Amplitude A_j

- Rend compte de la **contribution du facteur** à la préférence

$$A_j = MAX(a_1; a_2; a_3) - MIN(a_1; a_2; a_3)$$

◆ Importance I_j d'un facteur

- ◆ Le sujet est principalement sensible au type de bracelet (80%)

$$I_j = \frac{A_j}{\sum_j A_j}$$



Nota Bene

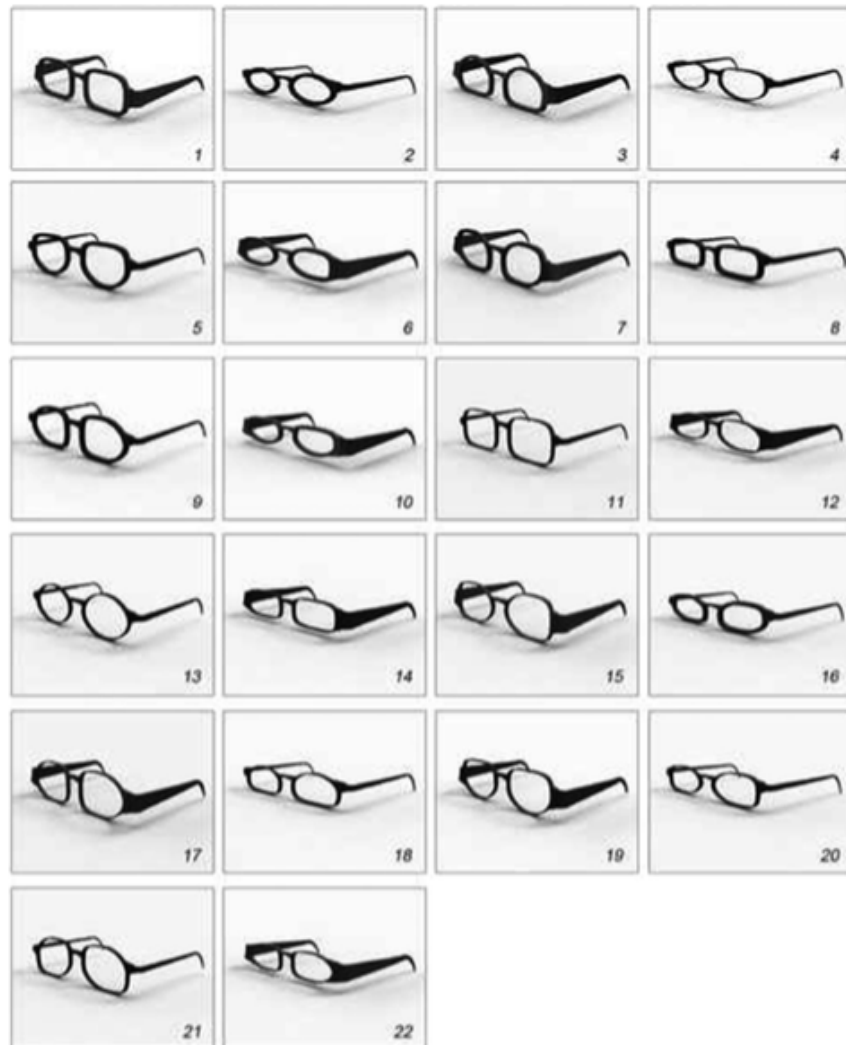
- ◆ Il est généralement irréaliste de présenter le plan complet
 - Ex. Jus d'orange (64 produits)
 - On se limite aux effets simples **sans interaction**
- ◆ Possibilité de prendre en compte l'effet d'interaction entre facteurs
 - Mais attention au nombre de variables à estimer ! (avoir suffisamment d'observations (scénarios))

$$\hat{y} = a_0 + a_1.(m - p) + a_2.(c - p) + a_3(a - d) + a_4(m - p)(a - d) + a_5(c - p)(a - d)$$

=> Au moins 6 scénarios nécessaires pour déterminer les 6 variables a_i ($i=1$ à 5)

Exemple : évaluation sémantique

Évaluation de montures de lunettes



Factors

Level Settings

A: frame leg width

A1: wide  A2: narrow 

B1:   B2:  

B: rim profile

B3:   B4:  

B5:   B6:  

B7:   B8:  

C: rim aspect ratio

C1: 0.75 (high) 

C2: 0.5 (wide) 

D: rim thickness

D1: thick  D2: thin 



Exemple

Évaluation de montures de lunettes



Table 2

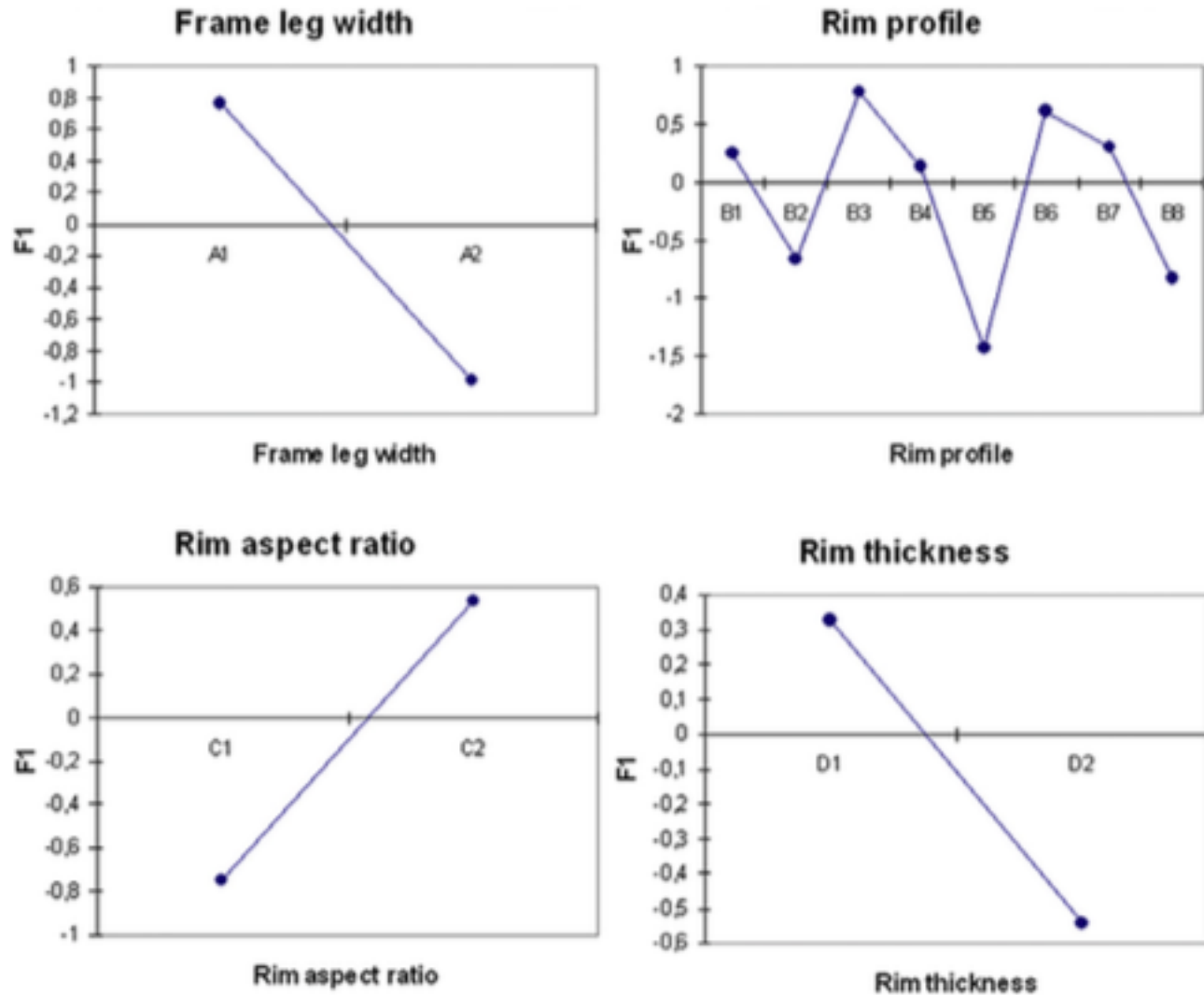
List of the semantic attributes.

No.	Adjectives
1	Feminine – masculine
2	Common – special
3	Original – dull
4	Delicate – rough
5	Funny – serious
6	Obedient – rebellious
7	Modern – retro
8	Smart – ordinary



Exemple

- ◆ Résultat pour la dimension modern-retro



Exemple

- ◆ Résultat pour la dimension **modern-retro**



(A2, B6, C1, D2)



(A1, B3, C2, D1)



Conclusions sur l'AC

◆ Avantages

- Méthode basée sur un « bon sens marketing solide »
- Permet de gérer des phénomènes assez complexes, avec un grand nombre de facteurs (exemple : packaging : on peut introduire des dizaines d'options)

◆ Limites

- Se prête mieux à la mesure de sensibilité pour des facteurs « objectifs » que « subjectifs (cas du luxe, de la parfumerie,...)
- Concernant l'utilisation prédictive, limite des plans expérimentaux fractionnaires (absence d'effets d'interaction)
- Limites liées au modèle compensatoire (pas d'effet de seuil)



Bibliographie

- ◆ Fiche de synthèse sur l'analyse conjointe. J-F PETIOT. Les techniques de l'ingénieur, 2013.

L'analyse conjointe, la statistique et le produit idéal – Méthodes et applications. Cisia – Ceresta éditeur, 1998.

Jean Claude Riquet et Christophe Benavent L'analyse conjointe et ses applications en marketing., IAE de Lille. <http://christophe.benavent.free.fr>

Crochemore S. Application originale de l'analyse conjointe pour un design *robuste* de volants ou comment concevoir utile. Actes de Confere 2005.

Y. Evrard, B. Pras, E. Roux. MARKET : Études et Recherches en Marketing. DUNOD 2000.

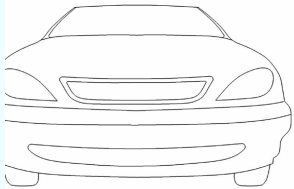
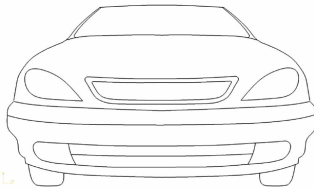
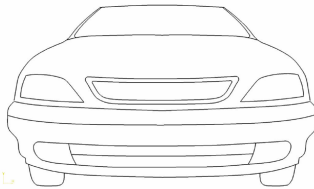
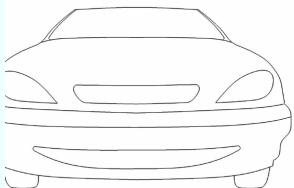
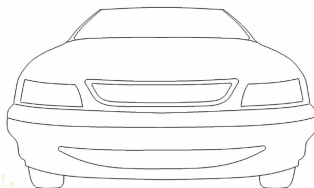
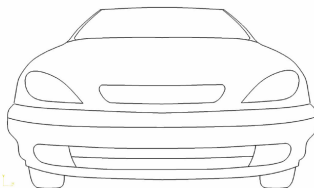
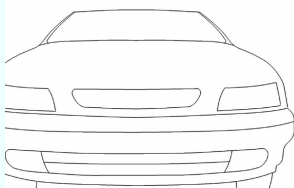
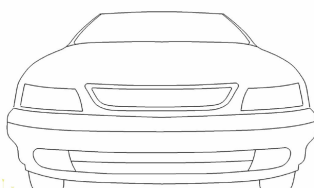
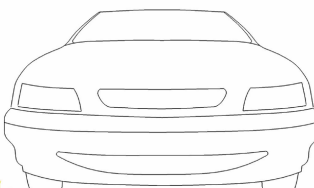
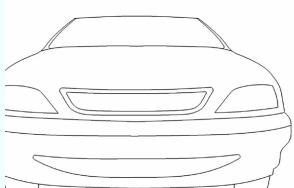
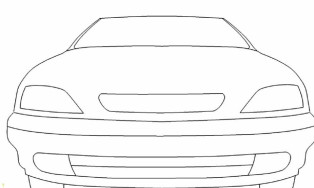
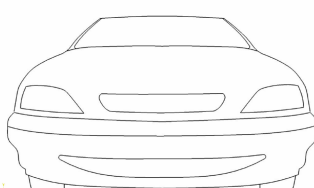
Lebart. Morineau. Piron. Statistique exploratoire multidimensionnelle. DUNOD, Sciences Sup.

 - ◆ <http://www.sawtoothsoftware.com/conjoint-analysis-software>
- ◆ Green, P. Carroll, J. and Goldberg, S. A general approach to product design optimization via conjoint analysis, *Journal of Marketing*, vol 43, summer 1981, pp 17-35.
- ◆ Helm R., Scholl A., Manthey L., Steiner M. Measuring customer preferences in new product development: comparing compositional and decompositional methods. *International Journal of Product Development* 2004 - Vol. 1, No.1 pp. 12 – 29.



Exercice analyse conjointe

Donner une note de préférence pour chaque design

		
P01	P02	P03
		
P04	P05	P06
		
P07	P08	P09
		
P10	P11	P12



La Classification ascendante hiérarchique CAH

La classification automatique

- ◆ Étant donnés N objets, définis par des mesures sur p variables, constituer des groupes d'objets tels que
 - les objets dans un même groupe sont « aussi semblables » que possible
 - les objets dans des groupes différents sont « aussi dissemblables » que possible
- ◆ Applications
 - segmentation du marché (consommateurs ou produits)
 - compréhension des différences perceptives
 - recherche de classes sociales (sociologie)
 - classification des espèces (paléontologie)

Il n'y a pas unicité à un problème de classification





Exemple : étude de préférences

- ◆ Recherche de groupes de sujets
- ◆ Tableau des notes de préférence (produit*sujet)

Produits	sujet 1	sujet 2	sujet 3	sujet 4	...	sujet 123	sujet 124	sujet 125
P1	4	0,	2	4		0	2	2.3
P2	3,9	7	3,91	3,91		9	3	3
P3	1,00	3,00	1,00	1,00		4,00	1,20	1
P4	1,3	3	1	1		4	1	1
P5	7	2	2,76	7		2	2,6	2
P6	5	10	1,70	4		10	1	2
P7	6	5	2	6		4	2	2
P8	1	2,59	4	1		2.4	5	4

- À partir de données métriques, déterminer les groupes de consommateurs qui « répondent de la même façon » et qui constituent un groupe « homogène »



définitions

E ensemble des objets à classer

Distance : application de $E \times E$ dans $\mathbb{R}$

$$\left. \begin{array}{l} d(i, j) = d(j, i) \\ d(i, j) \geq 0 \\ d(i, j) = 0 \Leftrightarrow i = j \\ d(i, j) \leq d(i, k) + d(k, j) \end{array} \right\} \text{Dissimilarité}$$

définitions

$$d^2(e_i; e_j) = (e_i - e_j)' M (e_i - e_j)$$

➤ $M = I$ $d^2(e_i; e_j) = \sum_{k=1}^p (x_{ik} - x_{jk})^2$

➤ $M = D_{1/s^2}$ $d^2(e_i; e_j) = \sum_{k=1}^p \frac{(x_{ik} - x_{jk})^2}{s_k^2}$

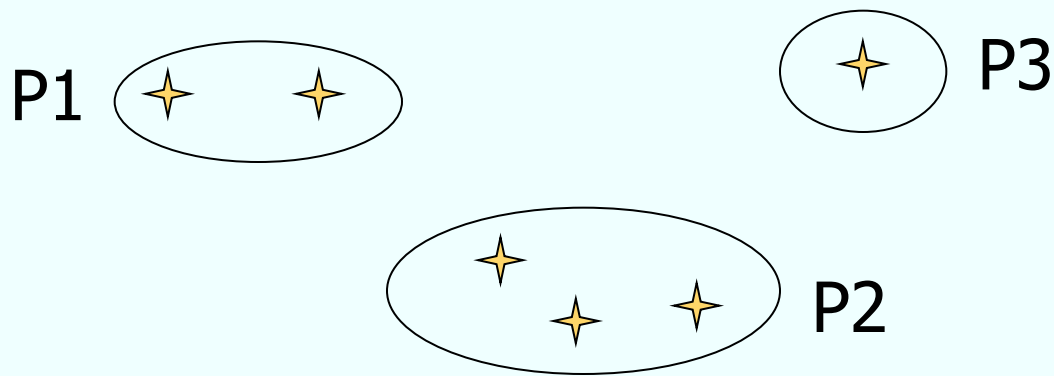
➤ Mahalanobis $M = V^{-1}$

➤ L_q : $d(e_i; e_j) = \left(\sum_{k=1}^p |x_{ik} - x_{jk}|^q \right)^{1/q}$

➤ L_1 : $d(e_i; e_j) = \sum_{k=1}^p |x_{ik} - x_{jk}|$

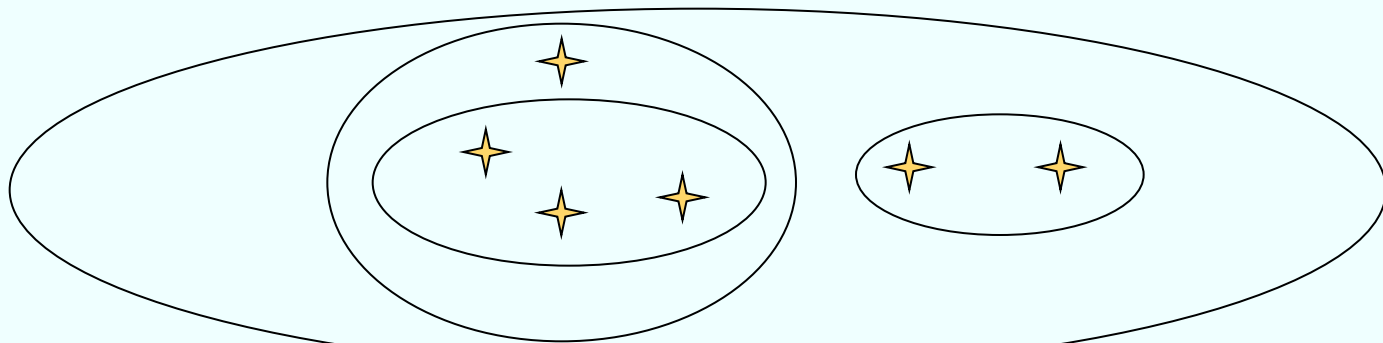
◆ Partition

- Ensemble de parties, disjointes, non vides, qui réunies forment l'ensemble complet



◆ Hiérarchie

- Partitions emboîtées





Différents algorithmes de classification

- ◆ Méthode hiérarchique : construction progressive d'un arbre de classification
 - ➡
 - Classification ascendante (CAH) : regroupement progressif
 - Méthode descendante : division du groupe total en sous groupe
- ◆ Méthode non hiérarchique (nodales) : construire k groupes par allocation des individus dans les groupes
 - Adaptées aux populations de taille importante
 - Nuées dynamiques
 - Centres mobiles

Classification supervisée :
inférer, à partir d'un échantillon
de données classées, une
procédure de **classification**

Classification non supervisée
(automatique):

méthodes destinées à répartir
des objets dans une structure
organisée (hiérarchie)

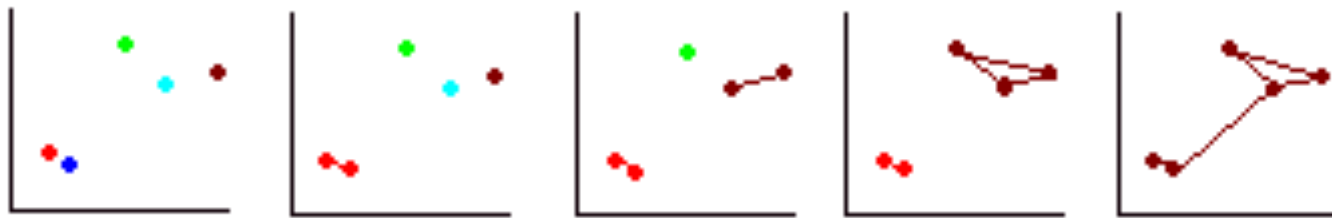


Principe de la CAH

- ◆ Au départ : on regroupe les individus les + ressemblants
- ◆ De manière itérative, on construit des classes de + en + importantes :
 - Soit en regroupant 2 individus
 - Soit en agrégeant 1 individu et 1 classe
 - Soit en agrégeant 2 classes

CAH : méthode

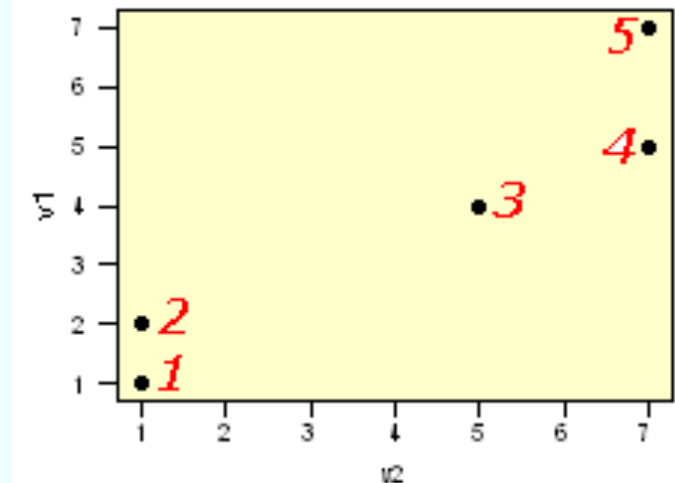
- ◆ 1. Choisir les données brutes
- ◆ 2. standardiser les données brutes
- ◆ 3. calculer la distances entre chaque groupe (matrice des distances)
- ◆ 4. Recherche les 2 groupes les plus similaires
- ◆ 5. fusionner les deux groupes pour former un **nouveau groupe**.
- ◆ 6. Calculer les distances entre le **nouveau groupe** et les autres groupes
- ◆ 7. Répéter l'étape 3 jusqu'à ce que tous les objets soient dans un seul groupe.



Exemple : CAH

- ◆ Construction d'un arbre de classification passant de n individus au groupe total : dendrogramme
 - définir la mesure de proximité entre individu
 - définir la règle de constitution des groupes (mesure de la proximité entre groupes)
- ◆ Exemple

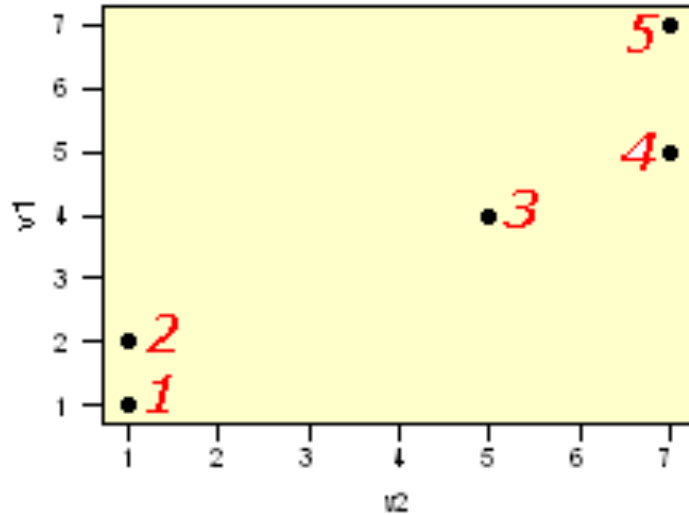
case	X1	X2
1	1	1
2	2	1
3	4	5
4	5	7
5	7	7





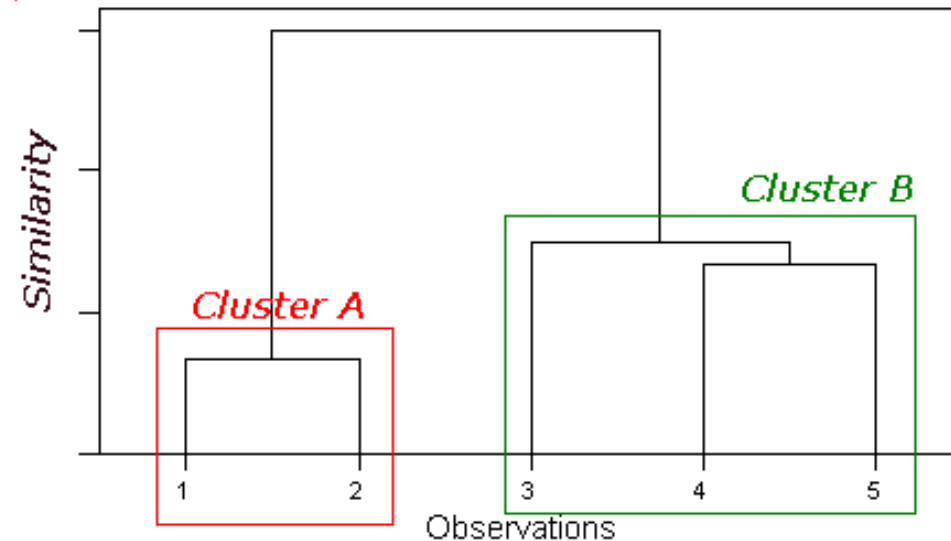
Exemple : CAH

◆ Exemple

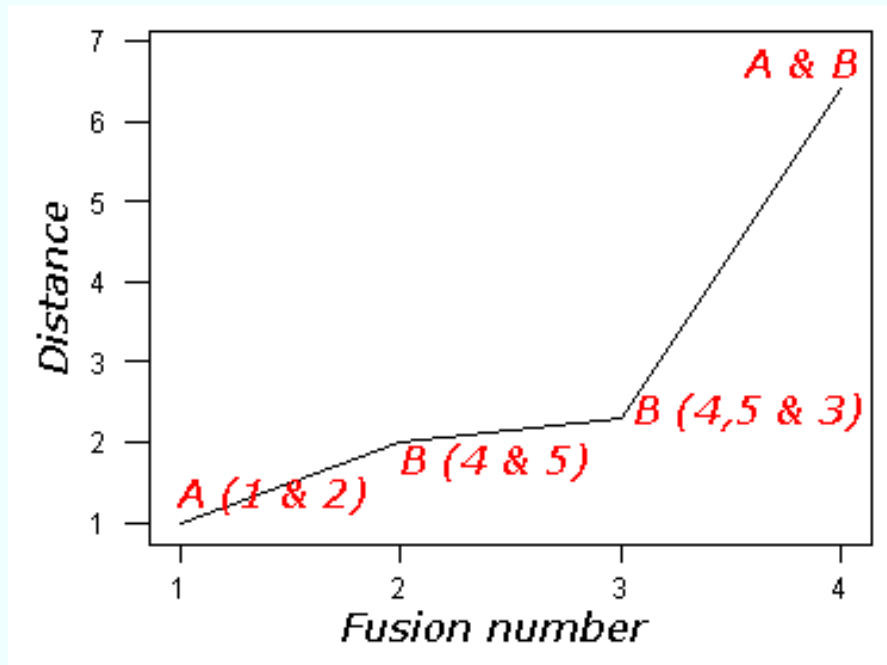


	1	2	3	4	5
1	0				
2	1	0			
3	5	4.5	0		
4	7.2	5.8	2.2	0	
5	8.5	6.7	3.6	2	0

	A	3	4	5
A	0			
3	4.7	0		
4	6.9	2.2	0	
5	8.1	3.6	2	0

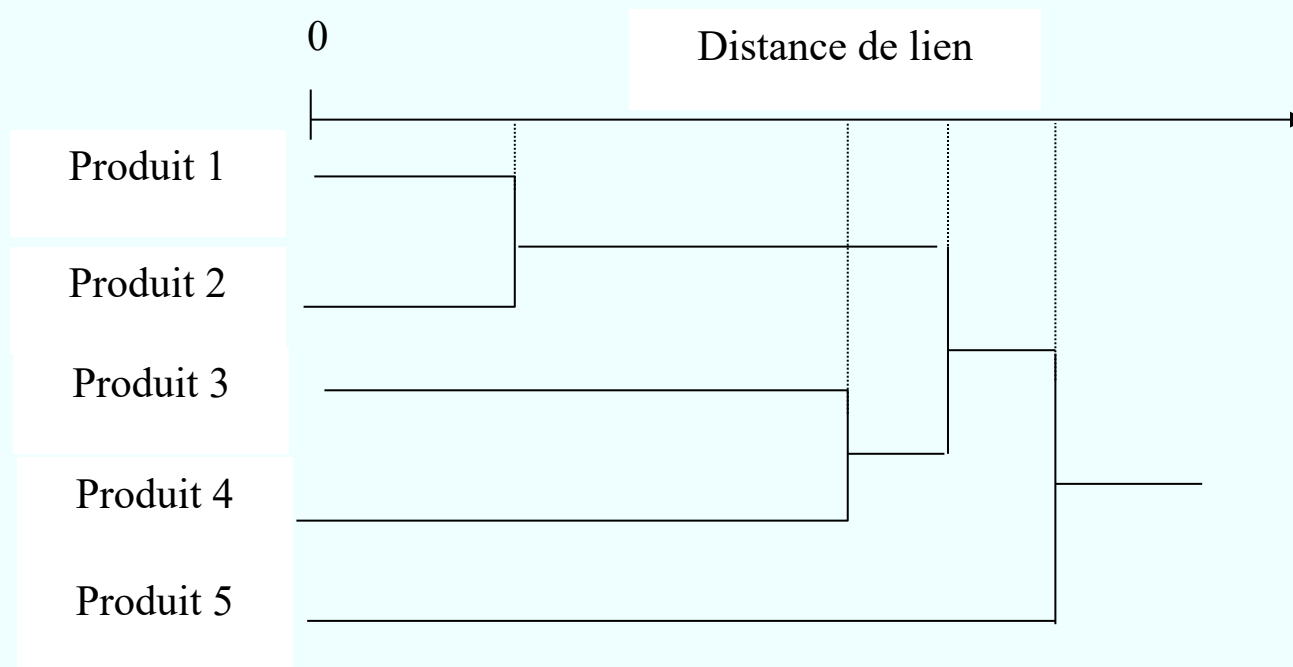


◆ Distance de lien/n° de fusion



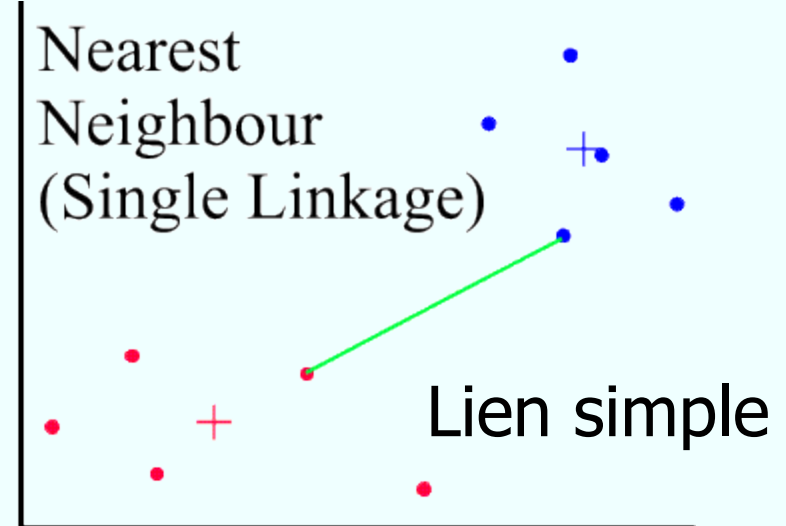
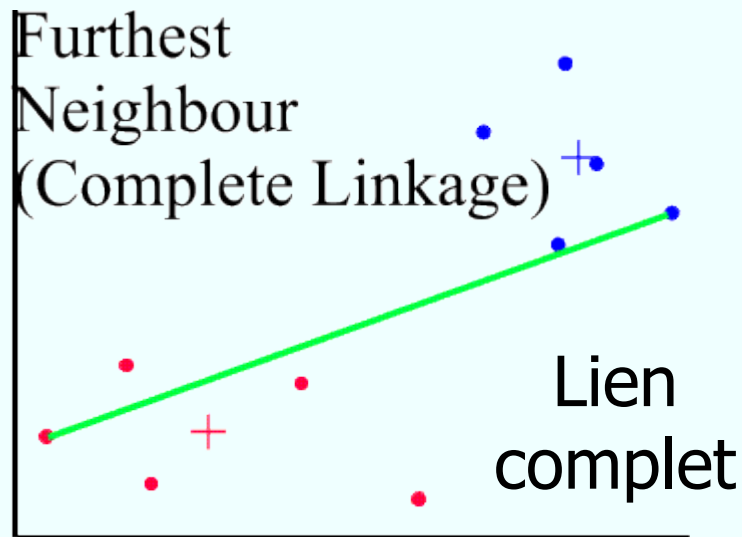
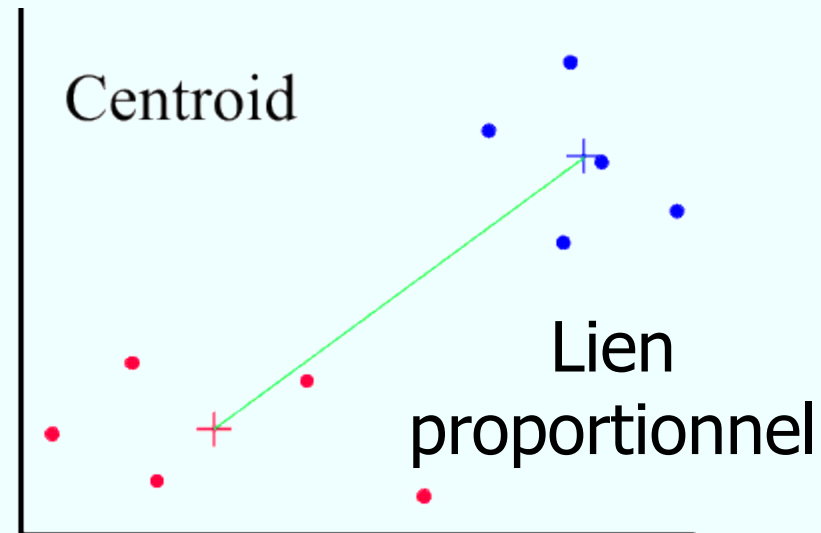
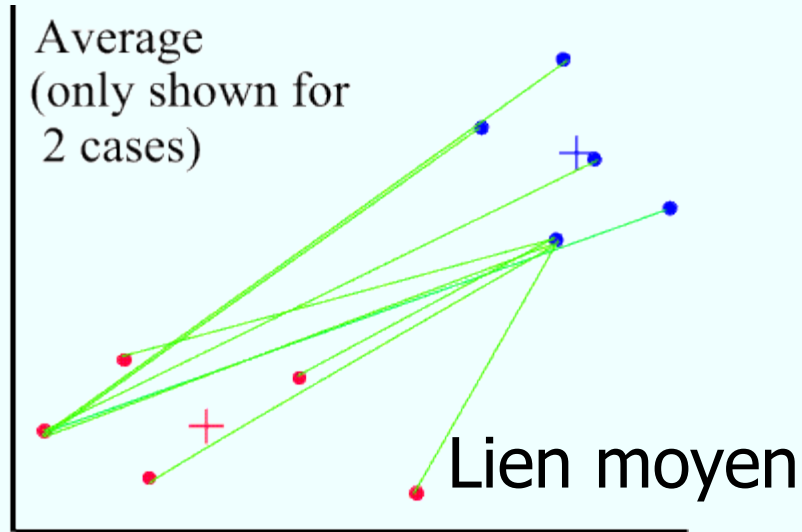
- ◆ nombre de groupe à prendre en compte :
 - avant un « saut » important de la distance de lien
 - ici deux groupes A et B

◆ dendrogramme



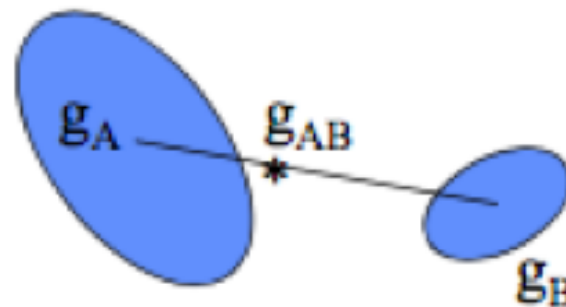


Différentes règles de regroupement



Méthode de Ward

On agrège les individus qui font le moins varier
l'inertie intra-classe



$$g_{AB} = \frac{p_A g_A + p_B g_B}{p_A + p_B}$$

Variation d'inertie due à l'agrégation de A et B :

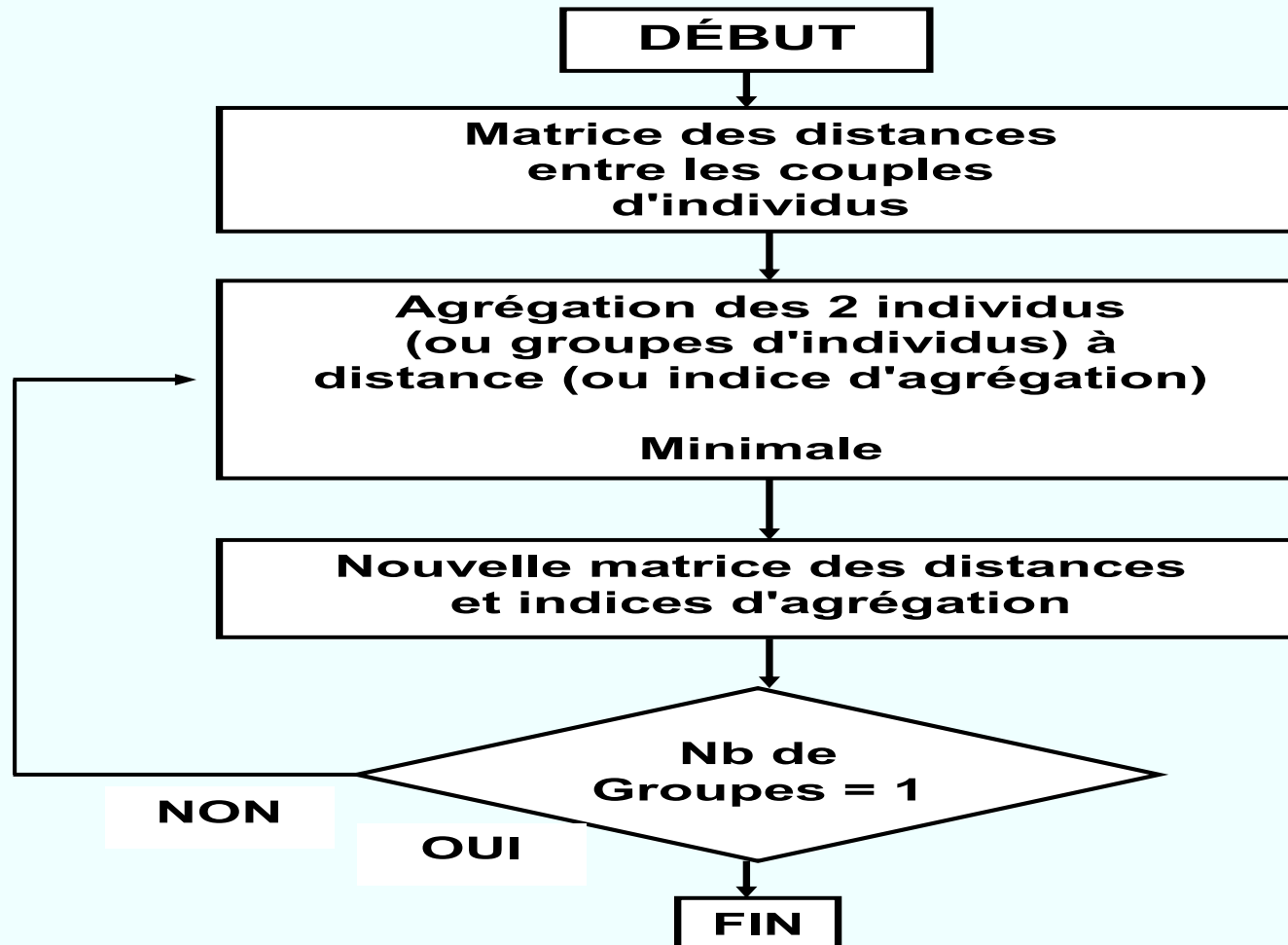
$$p_A d^2(g_A; g) + p_B d^2(g_B; g) - (p_A + p_B) d^2(g_{AB}; g)$$

$$= \frac{p_A p_B}{p_A + p_B} d^2(g_A; g_B)$$



méthode hiérarchique (CAH)

ALGORITHME DE BASE





Interprétation des résultats

- ◆ Quelle est la validité de la classification obtenue ?
 - analyse de la variance $\frac{\text{variance_intergroupe}}{\text{variance_totale}}$
 - distance entre le cdg et les éléments du groupe
 - Pour chaque variable, Test stat. de signif. / à la valeur moyenne (Student, s...)

- ◆ Comment décrire les groupes constitués ?
 - décrire l'individu « centre de gravité »
 - ajouter des variables explicatives supplémentaires



La typologie : applications en marketing

◆ Sur des individus

- Enquêtes d'images de marques
- Analyse de besoins (produits nouveaux)
 - Profil clientèle

◆ Sur des produits

- A partir des caractéristiques produits+profil des consommateurs -> segmentation marché



EXEMPLE : MARQUES DE NETTOYANTS POUR SOL

	A	B	C	D	E
Odeur agréable	60	38	40	39	62
Puissant	24	82	92	80	30
Sans rinçage	35	68	65	72	29
Écologique	68	52	30	28	75

points
d'image
attribués
à 5 marques
de
nettoyants
pour sol



EXEMPLE : MARQUES DE NETTOYANTS POUR SOL

MATRICE DES DISTANCES (*) ENTRE MARQUES

	A	B	C	D	E
A	0	129	156	154	21
B	129	0	37	31	138
C	156	37	0	22	165
D	154	31	22	0	163
E	21	138	165	163	0

(*) CITY BLOCK (Manhattan)



EXEMPLE : suite

Distance minimale : 21 entre A et E

- Première agrégation : on regroupe A et E
- Poursuite de l'algorithme avec crit. d'agrégation lien moyen :
 $D^* (B, (A,E)) = 1/2 (D (B,A) + D (B,E)) = 1/2 (129 + 138) = 133,5$
→ Nouvelle matrice de distances et d'indices d'agrégation :

	(A,E)	B	C	D
(A,E)	0	133,5	160,5	158,5
B	133,5	0	37	31
C	160,5	37	0	22
D	158,5	31	22	0



EXEMPLE : suite

Distance minimale : 22 entre C et D

→ Nouvelle matrice de distances et d'indices d'agrégation :

	(A,E)	B	(C,D)
(A,E)	0	133,5	159,5
B	133,5	0	29
(C,D)	159,5	29	0

Puis, distance minimale : 29 entre B et (C,D)

→ On regroupe B et (C,D)



EXEMPLE : suite

Distance minimale : 22 entre C et D

→ Nouvelle matrice de distances et d'indices d'agrégation :

	(A,E)	B	(C,D)
(A,E)	0	133,5	159,5
B	133,5	0	29
(C,D)	159,5	29	0



EXEMPLE suite

Puis, distance minimale : 29 entre B et (C,D)

→ On regroupe B et (C,D)

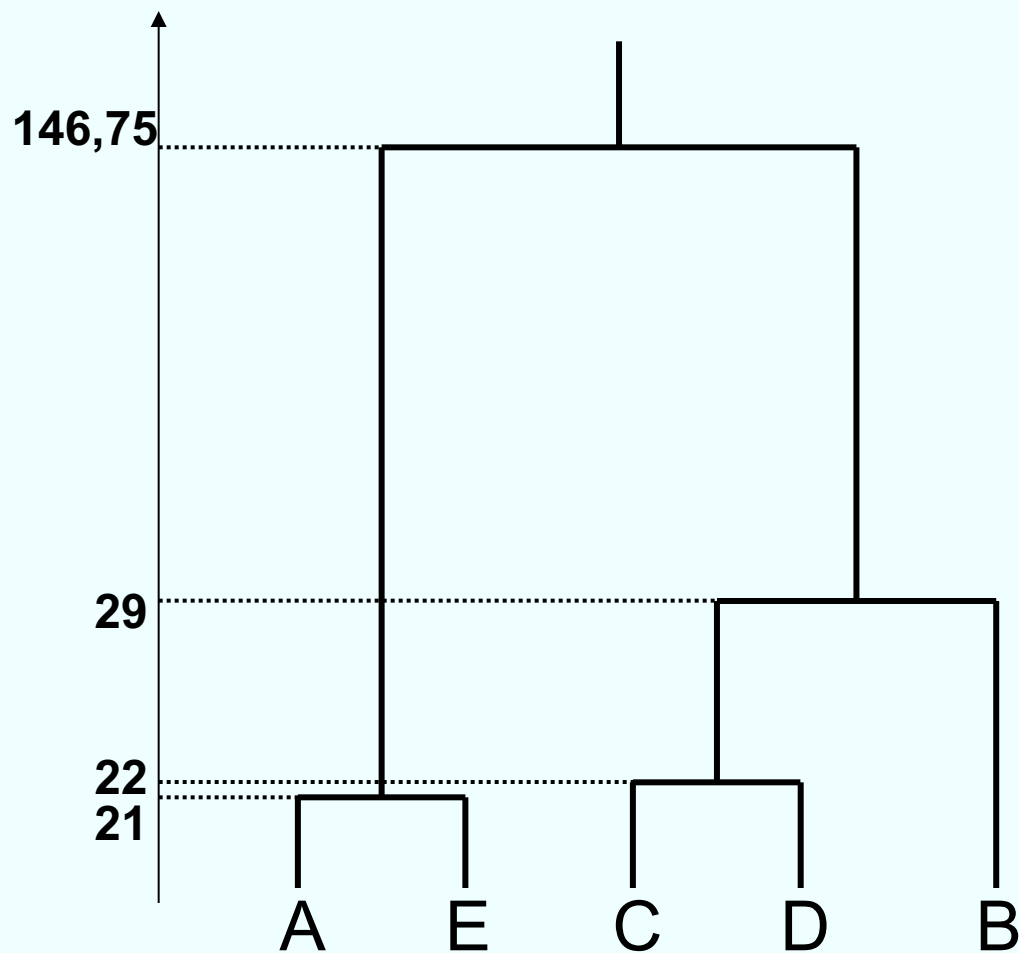
→ Nouvelle matrice de distances et d'indices d'agrégation :

	(A,E)	(C,D,B)
(A,E)	0	146,75
(C,D,B)	146,75	0



EXEMPLE suite

Arbre de classification:





EXEMPLE : MARQUES DE NETTOYANTS POUR SOL

Typologie à 3 groupes

Groupes :		(A,E)	B	(C,D)
Effectif :		2	1	2
CDG	Odeur agréable	61	38	39.5
	Puissant	27	82	86
	Sans rinçage	32	68	68.5
	Écologique	71.5	52	29



EXEMPLE : MARQUES DE NETTOYANTS POUR SOL

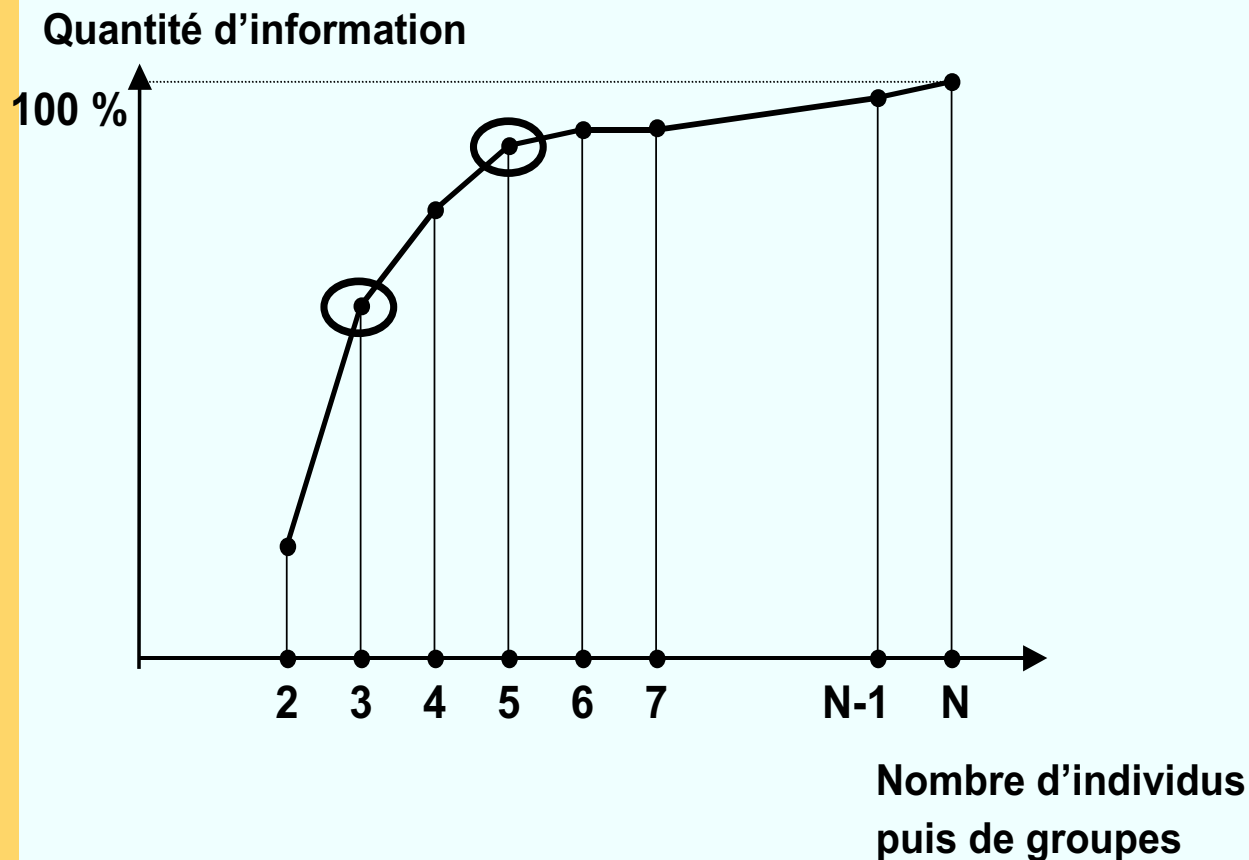
Typologie à 2 groupes

Groupes :	(A,E)	(B,C,D)
Effectifs :	2	3
Odeur agréable	61	38,8
Puissant	27	84,7
Sans rinçage	32	68.3
Écologique	71.5	40,5



CHOIX DU NOMBRE DE GROUPES

COURBE DE « RENDEMENT » D'UNE ANALYSE → Règle du « COUDE »





Exemple : les étudiants en vacances

- ◆ Objectif : segmenter une population d'étudiants/attentes concernant les voyages
- ◆ Questionnaire

item	Importance		
	Pas du tout 0	5	très important 10
Se dépenser physiquement			
Faire la fête			
Prendre le temps de vivre			
Sortir des normes de la vie courante			
Etre émerveillé par la nouveauté			
S'ouvrir l'esprit par une autre culture			
Se faire une opinion pays/culture			
S'enrichir d'images et de souvenirs			
Vivre des situations risquées			
Cherche à se valoriser/entourage			

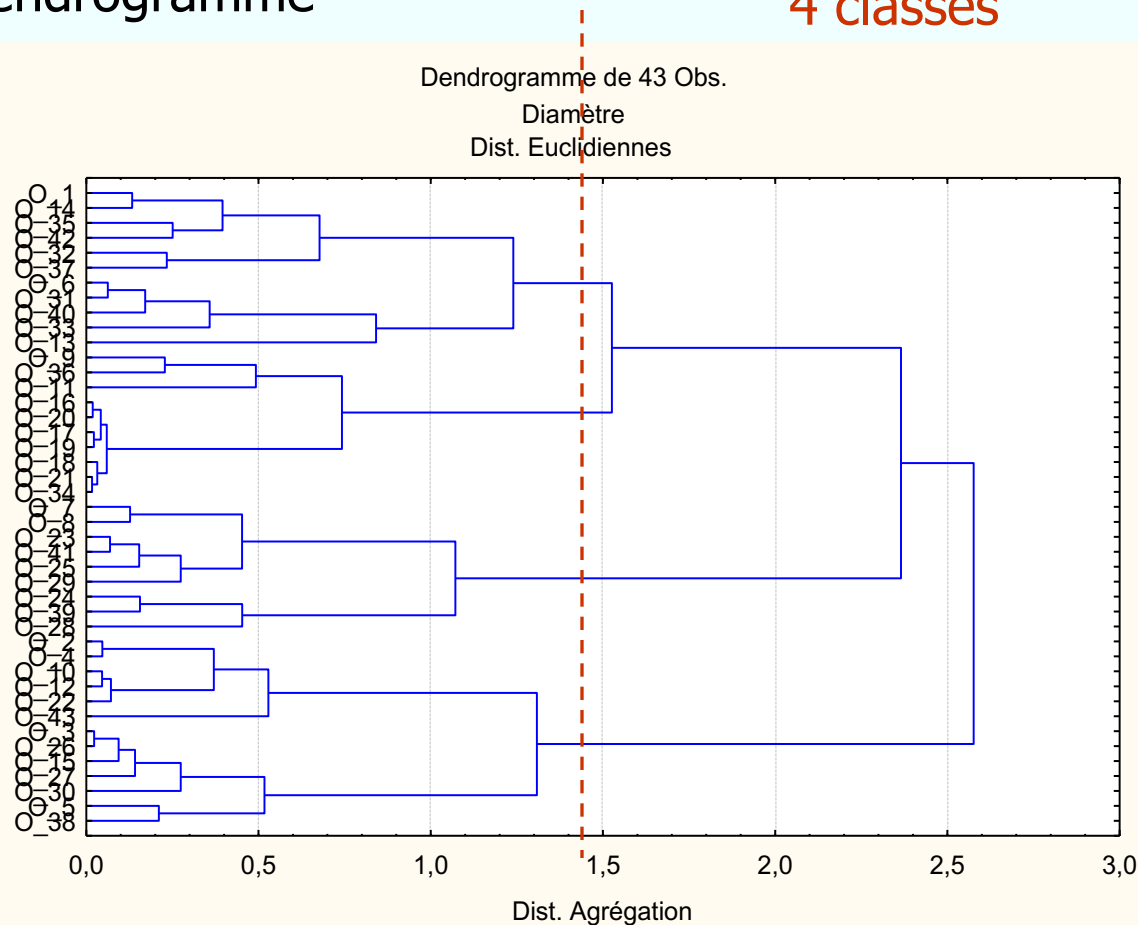


Exemple : les étudiants en vacances

◆ Analyse ascendante hiérarchique

- données métrique : distance euclidienne
- distance de lien : diamètre (lien complet)
- dendrogramme

4 classes





Exemple : les étudiants en vacances

item	Moyenne des variables				
	Glob.	C 1	C 2	C3	C4
Se dépenser physiquement	3.8	4.06	4.91	2.3	3.67
Faire la fête	5.75	6.11	5.91	5.9	4.92
Prendre le temps de vivre	6.84	6.83	7.55	3.7	8.83
Sortir des normes de la vie courante	7.65	8.78	9.27	6.2	5.67
Etre émerveillé par la nouveauté	7.88	8.78	8.55	7.6	8.08
S'ouvrir l'esprit par une autre culture	8.18	8.5	8.73	7.6	6.17
Se faire une opinion pays/culture	6.8	7.72	5.27	4.8	8.5
S'enrichir d'images et de souvenirs	7.61	8.72	5.64	7.6	7.75
Vivre des situations risquées	5.96	7.33	7.36	3.5	4.67
Cherche à se valoriser/entourage	5.35	6.39	3.18	5.6	5.58

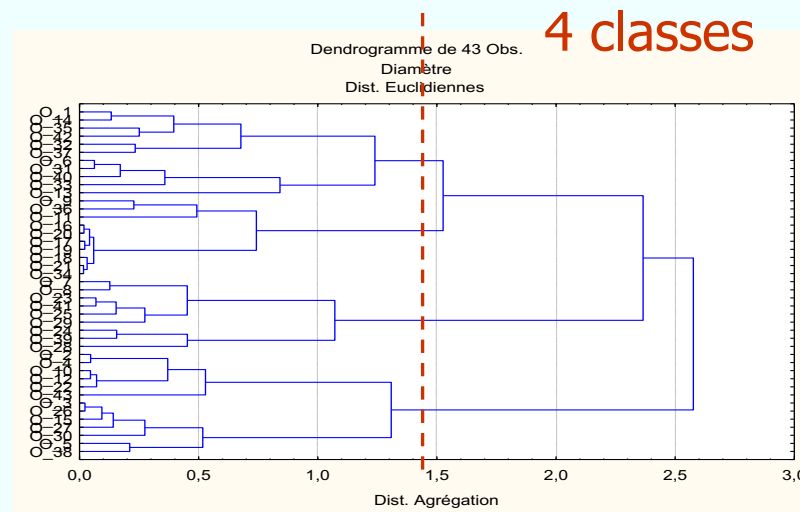
Gras : significativement supérieur à la moyenne des notes

Italique : sign. inférieur

Exemple : les étudiants en vacances

◆ Description des classes

- Étudier les valeurs des *cdg* des classes sur les variables initiales
- L'interprétation est parfois complexe car les groupes ne représentent pas forcément des types « purs »
 - Ajouter des variables supplémentaires (qui n'ont pas servi à faire la classification) peut aider à l'interprétation

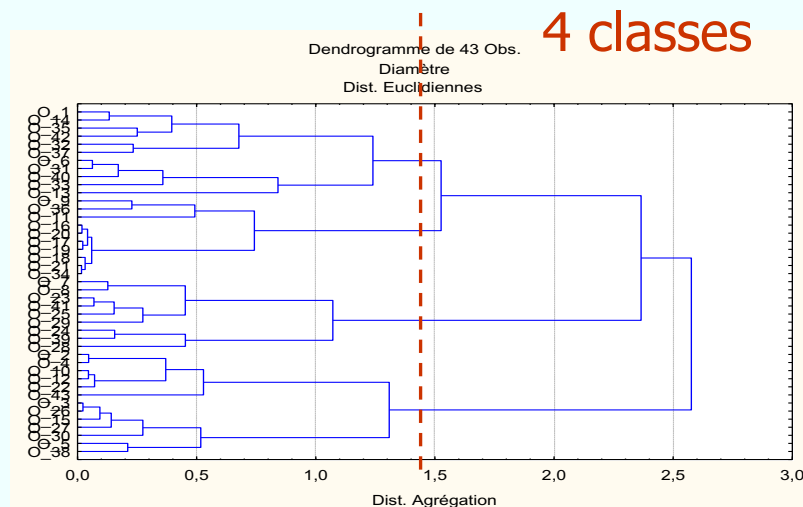




Exemple : les étudiants en vacances

◆ Description des classes

classe	caractérisation
C1	Les aventuriers modérés
C2	Les aventuriers fanatiques
C3	Les fêtards
C4	Les intellos





Exemple : les étudiants en vacances

◆ Caractérisation des classes

- caractériser la population de chaque classe avec des questions socio-économiques (afin d'orienter des actions de communications et de publicité)
 - tri selon l'appartenance à une classe,
 - analyse discriminante (idem que la régression multiple mais avec une variable à expliquer nominale)

classe	caractérisation						
		%gars	%filles	%sciences	%littéraires	%paris	%province
C1	Les aventuriers modérés	55	45	55	45	50	50
C2	Les aventuriers fanatiques	70	30	55	45	45	55
C3	Les fêtards	60	40	65	35	55	45
C4	Les intellos	15	85	25	75	60	40

Conclusion

- cette étude permettrait à un tour-opérateur de segmenter les clients en fonction des bénéfices recherchés et de proposer les produits adaptés.



Exemple 3 : parfums

Regroupement a partir du profil sensoriel moyen (panel)

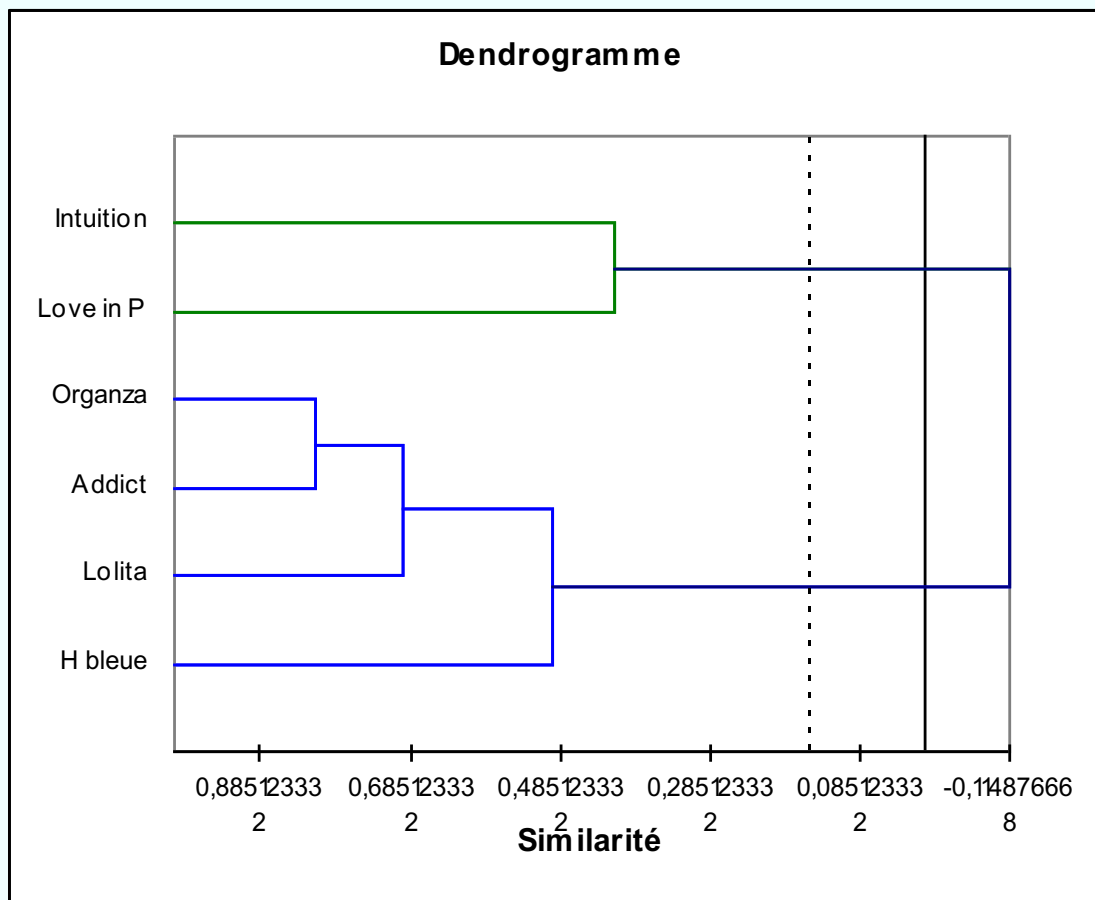
Notes moyennes du panel

	Addict	Love in P	H bleue	Intuition	Organza	Lolita
vanille	4,25	3,00	1,63	1,38	2,81	4,75
citron	1,63	4,00	2,81	5,75	2,81	2,69
eau de cologn	2,75	3,00	5,56	4,63	3,31	2,00
épicé	5,00	2,81	5,56	3,69	4,94	4,25
poivré	4,31	2,81	5,50	3,69	4,75	3,69
floral	5,19	6,25	3,31	4,63	3,94	4,06
aquatique	1,81	3,25	1,25	2,56	1,81	2,38
sucré	5,44	5,75	3,31	3,50	5,06	7,56
frais	2,63	6,25	2,19	5,31	3,81	4,06
lourd	6,38	3,19	7,38	4,06	4,75	5,13



Exemple 3 : parfums

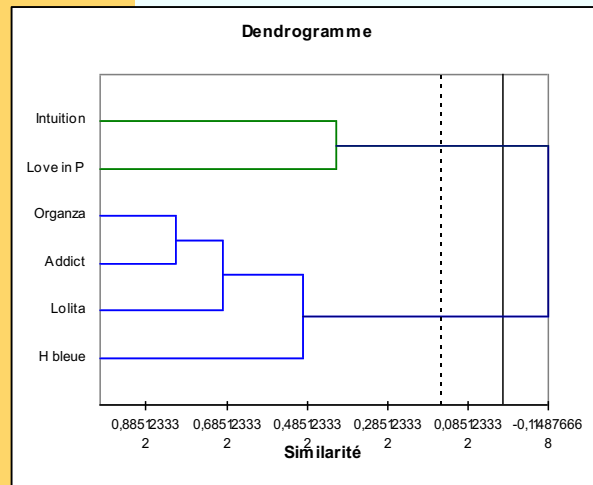
Regroupement a partir du profil sensoriel moyen (panel)





Exemple 3 : parfums

Regroupement a partir du profil sensoriel moyen (panel)



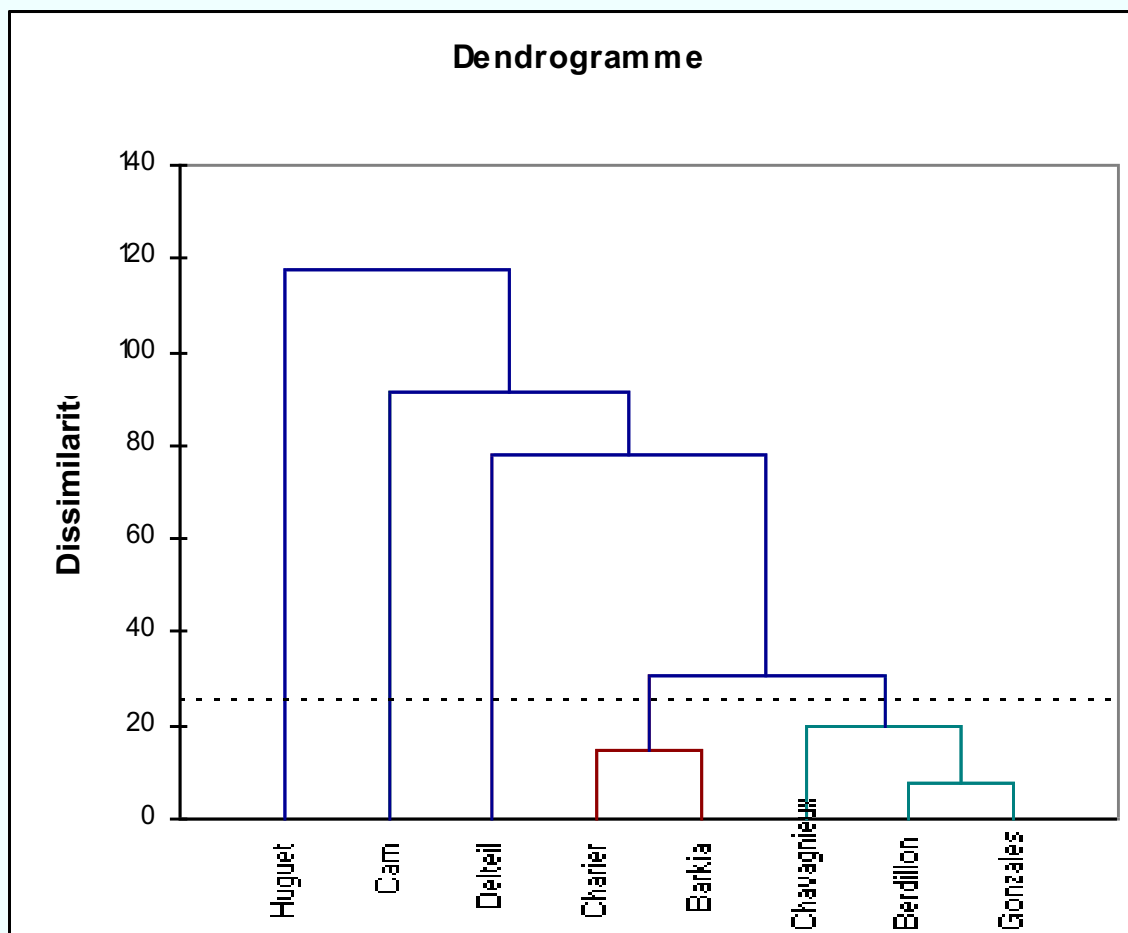
Groupe 1 : floral–citron-frais

Groupe 2 : lourd – pas citron

	vanille	citron	eau de cologne	épicé	poivré	floral	aquatiqu e	sucré	frais	lourd
groupe 1	2,19	4,88	3,81	3,25	3,25	5,44	2,91	4,63	5,78	3,63
groupe 2	3,36	2,48	3,41	4,94	4,56	4,13	1,81	5,34	3,17	5,91
moyenne	2,97	3,28	3,54	4,38	4,13	4,56	2,18	5,10	4,04	5,15

Exemple 4 : parfums

Regroupement des sujets



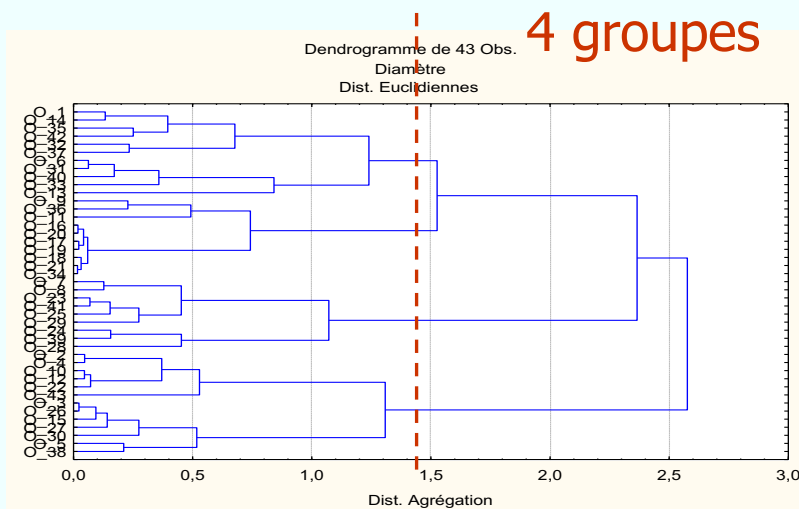


Exemple 5 : étude de préférences

Classement de préférence de p produits par n conso.

	Produit 1	Produit 2	...		Produit p
1					
2					
3					
4					
5					
...					
n					

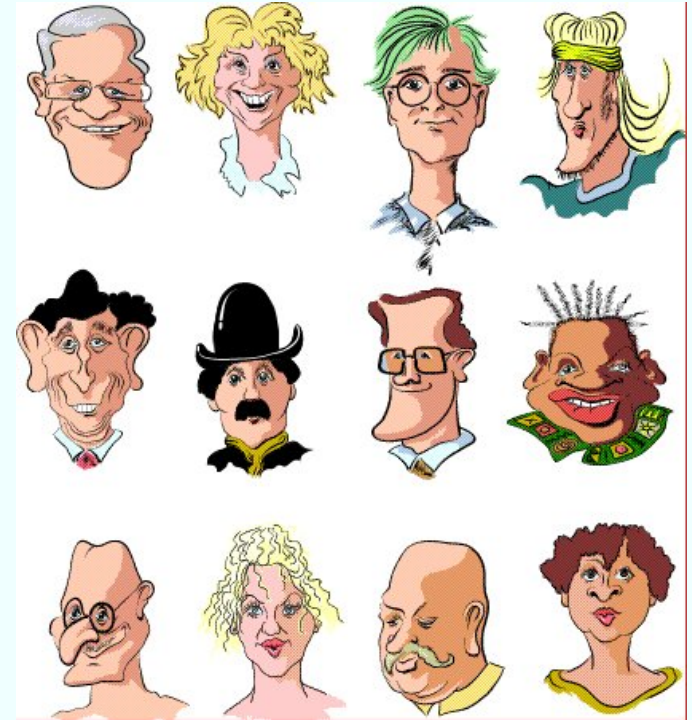
segmentation



Cas des variables nominales

◆ Classification de « figures »

case	genre	lunettes	moustache	sourire	chapeau
1	m	y	n	Y	n
2	f	n	n	Y	n
3	m	y	n	N	n
4	m	n	n	N	n
5	m	n	n	y	n
6	m	n	y	N	y
7	m	y	n	Y	n
8	m	n	n	Y	n
9	m	y	y	Y	n
10	f	n	n	N	n
11	m	n	y	N	n
12	f	n	n	N	n



- Pb du paléontologiste
- définition d 'une distance ou d 'une similarité sur des données nominales (binaires)



Cas des variables nominales

N.B. Variables nominales : on code chaque variable avec des 0 et des 1 (selon le nbre de modalités)

forme disjonctive complète

case	genre	lunettes	moustache	sourire	chapeau
1	m	y	n	Y	n
2	f	n	n	Y	n
3	m	y	n	N	n
4	m	n	n	N	n
5	m	n	n	y	n
6	m	n	y	N	y
7	m	y	n	Y	n
8	m	n	n	Y	n
9	m	y	y	Y	n
10	f	n	n	N	n
11	m	n	y	N	n
12	f	n	n	N	n

cas e	mas culin	fémi nin	lunet tes	Pas lunet te	mou stach e	Pas mou stach e	souri re	Pas sourir e	chap eau	Pas chap eau
1	1	0	1	0	0	1	1	0	0	1
2	0	1	0	1	0	1	1	0	0	1
3	1	0	1	0	0	1	0	1	0	1
4	1	0	0	1	0	1	0	1	0	1
5	1	0	0	1	0	1	1	0	0	1
6	1	0	0	1	1	0	0	1	1	0
7	1	0	1	0	0	1	1	0	0	1
8	1	0	0	1	0	1	1	0	0	1
9	1	0	1	0	1	0	1	0	0	1
10	0	1	0	1	0	1	0	1	0	1
11	1	0	0	1	1	0	0	1	0	1
12	0	1	0	1	0	1	0	1	0	1

- On se ramène alors à des variables quantitatives binaires (forme disjonctive complète)



Cas des variables nominales

Définition d'une dissimilarité D entre les individus

Table de contingence

	1	0	Tot
1	<i>a</i>	<i>b</i>	<i>a+b</i>
0	<i>c</i>	<i>d</i>	<i>c+d</i>
Tot	<i>a+c</i>	<i>b+d</i>	<i>n</i>

Exemple :

Individu 1 : 01100001010010

Individu 2 : 01010001100010

Indice de Sokal et Michener

$$S_{12} = \frac{a + d}{n}$$

dissimilarité

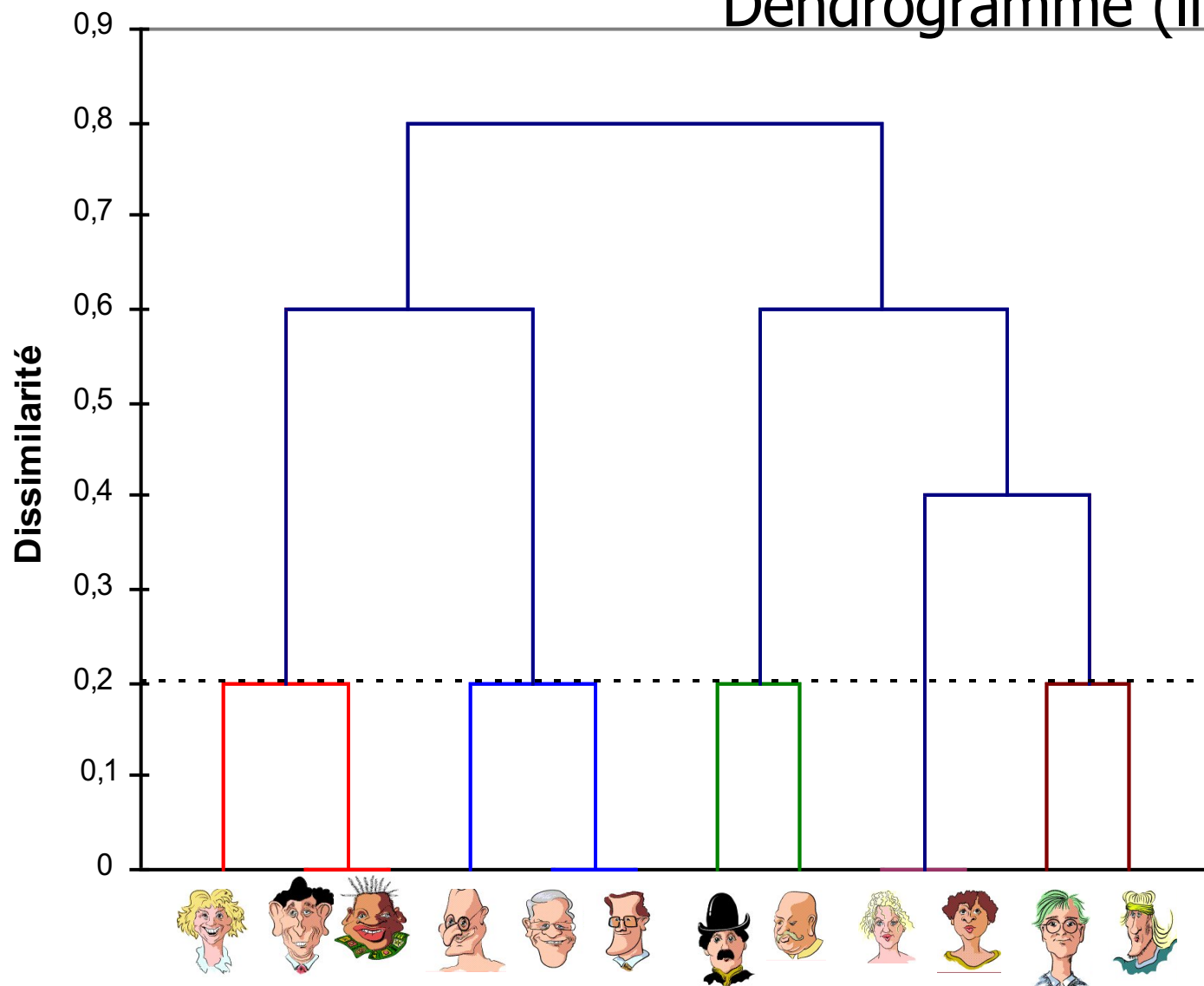
$$D_{12} = \sqrt{1 - S_{12}}$$

$$S_{12} = \frac{3 + 7}{14}$$



Cas des variables nominales

Dendrogramme (lien complet)





La classification : points importants

- ◆ Quelles données utiliser pour le regroupement ?
- ◆ Comment mesurer les similarités ?
- ◆ Comment former les « clusters » ?
- ◆ Comment identifier le nombre optimum de clusters ?

Il n'y a jamais
unicité à un
problème de
classification

Il y a
seulement des
classifications
plus ou moins
appropriées au
but fixé

Ex : livres dans une bibliothèque



Exemple 2 : les jus d'orange

◆ Attributs (facteurs)

- Recette (8 modalités)

⇒ R1, R2, R3, ..., R8

- Marque (4 modalités)

⇒ joker/Solevita/Jafaden/Carrefour

- Type (2 modalités)

⇒ Brick/bouteille verre

Plan
fractionnaire

Scenario	Recette	Marque	Type
1	1	1	1
2	2	1	2
3	3	2	1
4	4	2	2
5	5	3	1
6	6	3	2
7	7	4	1
8	8	4	2
9	1	2	2
10	2	3	1
11	3	3	2
12	4	4	1
13	5	4	2
14	6	1	1
15	7	1	2
16	8	2	1

Traitement des données

Consommateur h fournit la réponse $y_h = [y_{1h} \dots y_{nh}]$

↓ consommateur

recettes								marques types				y_h		
1	0	0	0	0	0	0	0	1	0	0	0	1	0	8
0	1	0	0	0	0	0	0	1	0	0	0	0	1	5
0	0	1	0	0	0	0	0	0	1	0	0	1	0	6
0	0	0	1	0	0	0	0	0	1	0	0	0	1	7
0	0	0	0	1	0	0	0	0	0	1	0	1	0	6
0	0	0	0	0	1	0	0	0	0	1	0	0	1	4
0	0	0	0	0	0	1	0	0	0	0	1	1	0	7
0	0	0	0	0	0	0	1	0	0	0	0	1	1	6
1	0	0	0	0	0	0	0	0	1	0	0	0	1	7
0	1	0	0	0	0	0	0	0	0	1	0	1	0	5
0	0	1	0	0	0	0	0	0	0	1	0	0	1	8
0	0	0	1	0	0	0	0	0	0	0	1	1	0	5
0	0	0	0	1	0	0	0	0	0	0	1	0	1	6
0	0	0	0	0	1	0	0	1	0	0	0	1	0	4
0	0	0	0	0	0	1	0	1	0	0	0	0	1	7
0	0	0	0	0	0	0	1	0	1	0	0	1	0	7

Codage : tableau disjonctif complet

Utilités individuelles

Pour chaque consommateur

Utilités:

≈ Coefficients de régression

≈ Effet en ANOVA

recette

Moyenne	5.98
R1	1.15
R2	-1.36
R3	0.81
R4	0.69
R5	-0.21
R6	-2.61
R7	0.93
R8	0.60

marque

M1	0.39
M2	-0.14
M3	0.46
M4	-0.71

type

T1	-0.13
T2	0.13



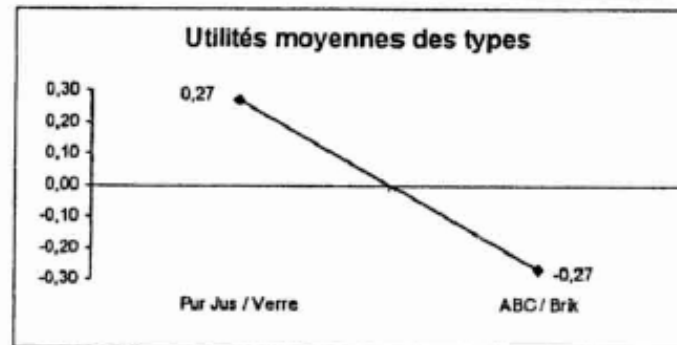
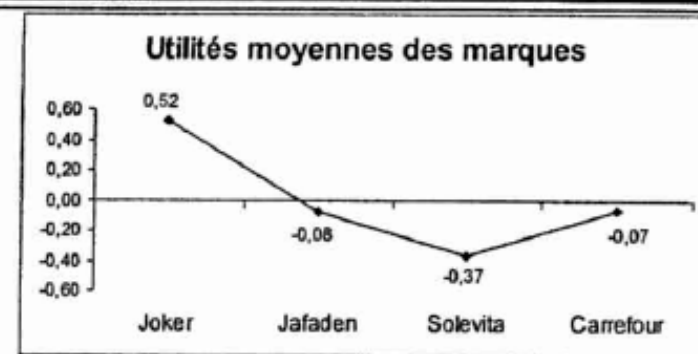
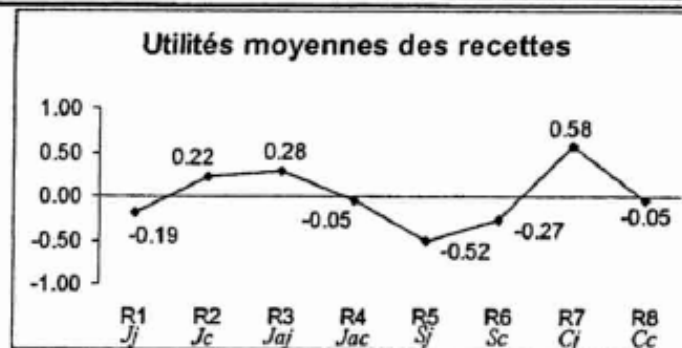
importances individuelles

Moyenne	5.98	Amplitudes	Importances
R1	1.15	$[1.15 - (-2.61)]$ 3.76	72 %
R2	-1.36		
R3	0.81		
R4	0.69		
R5	-0.21		
R6	-2.61		
R7	0.93		
R8	0.60		
M1	0.39	1.17	22 %
M2	-0.14		
M3	0.46		
M4	-0.71		
T1	-0.13	0.26	6 %
T2	0.13		

Importance d'un attribut = $\frac{\text{Amplitude des utilités de cet attribut}}{\text{Somme des amplitudes des utilités}}$



Utilités moyennes (pour tous les conso.)



Hypothèse d'additivité
(modèle compensatoire)



Produit idéal :

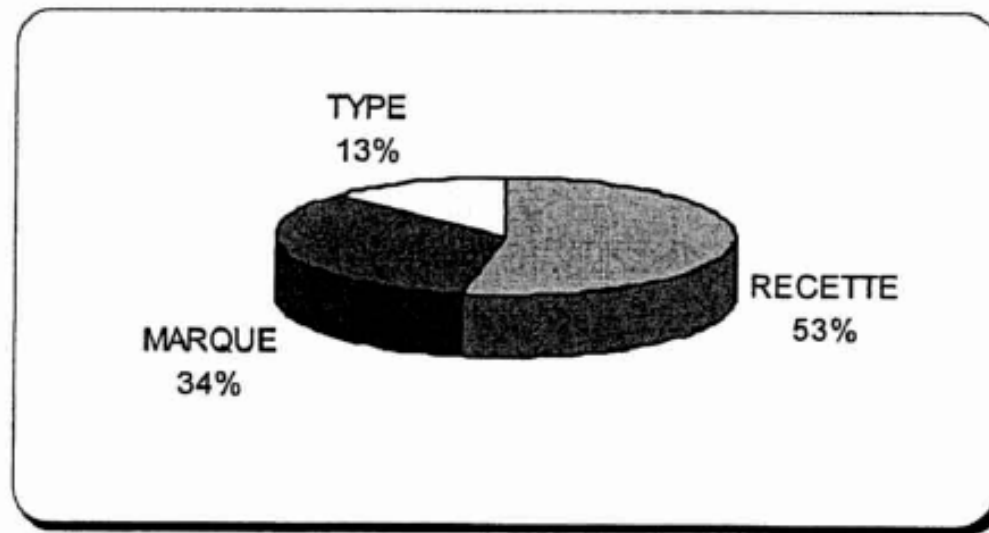
Recette : Pur Jus Carrefour

Marque : Joker

Type : Bouteille verre

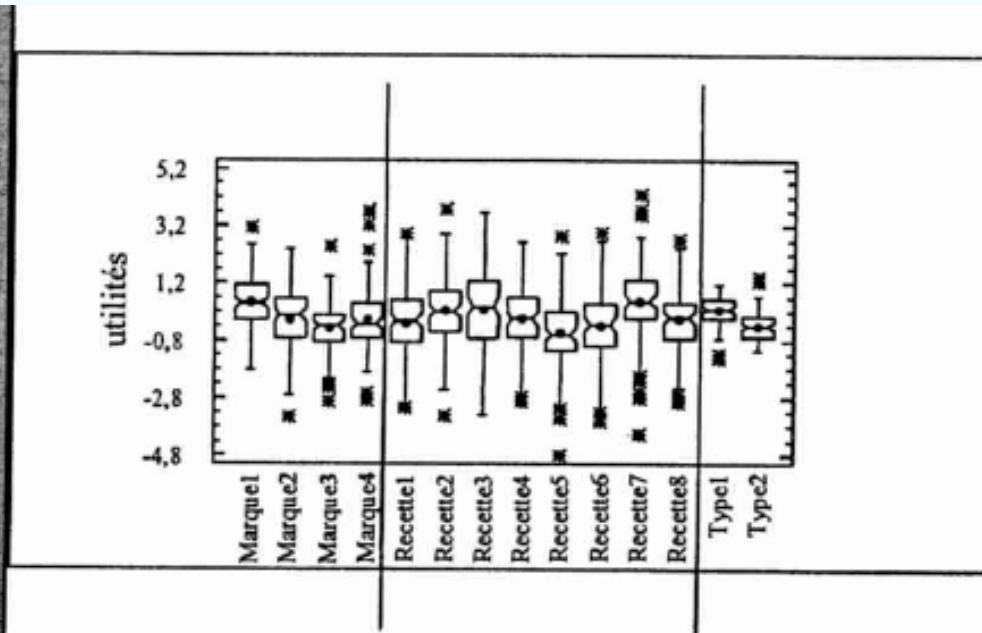


Importances moyennes

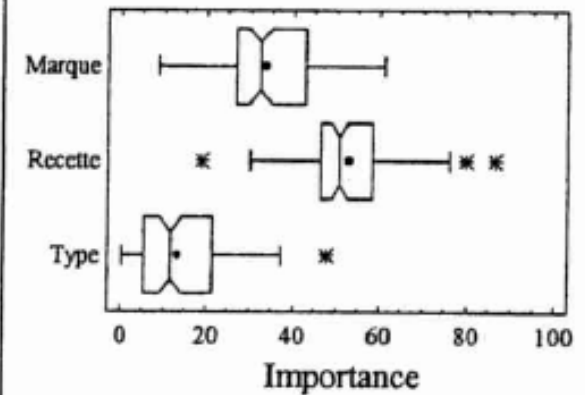


Importances moyennes

Dispersion des jugements



Utilités individuelles



Importances individuelles

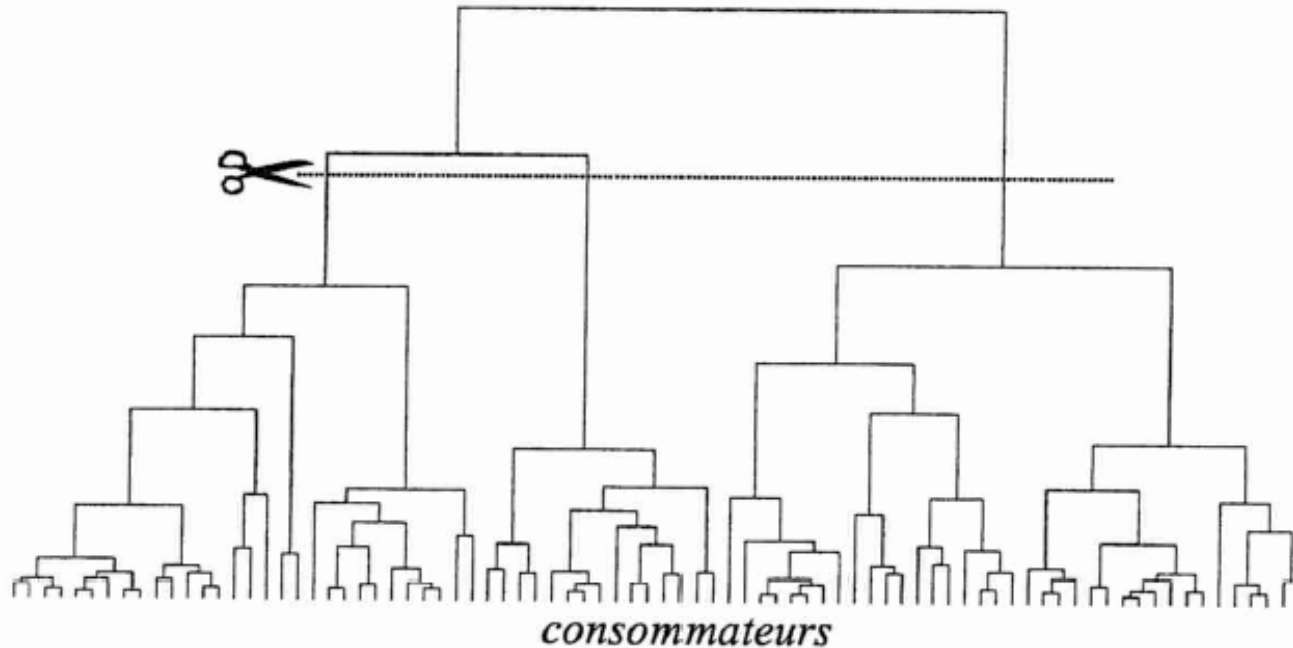


Segmentation du panel

- ◆ Possibilités de faire une classification
 - Sur les notations directes
 - Sur les utilités partielles
 - Sur les importances



Classification sur les importances

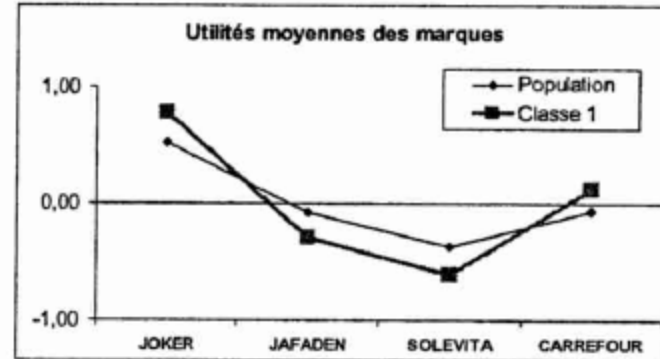
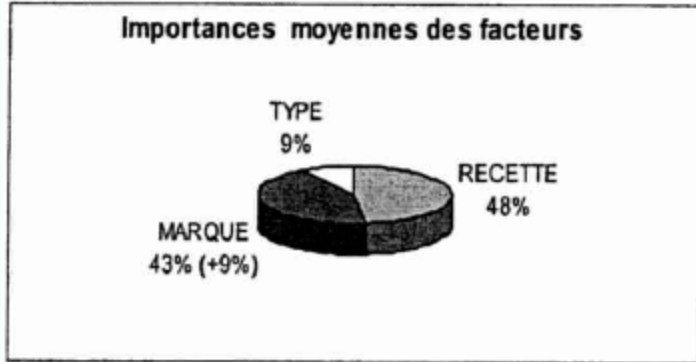


Segmentation du panel

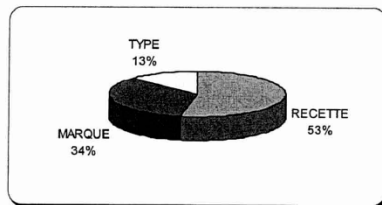


Définitions de cibles

classe 1 du panel



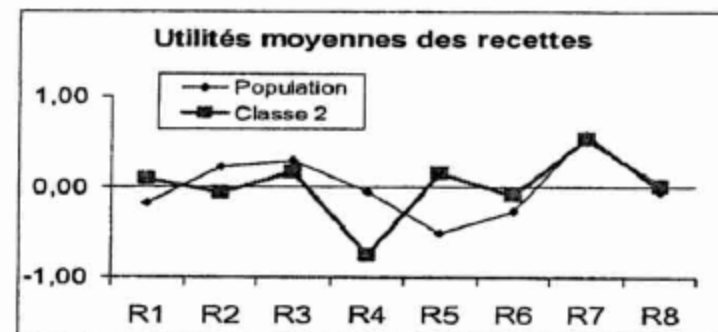
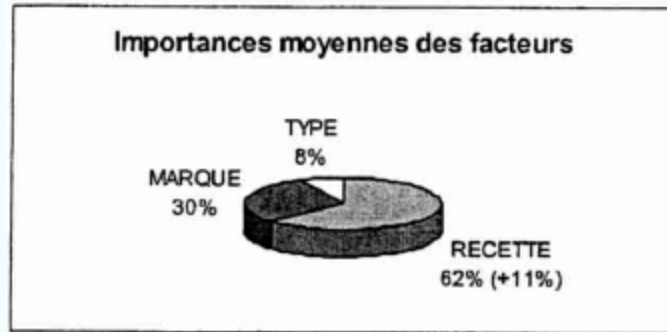
**Les individus de la 1ère classe
donnent une importance plus élevée
à la marque que le panel global.**



Importances moyennes

44.7%

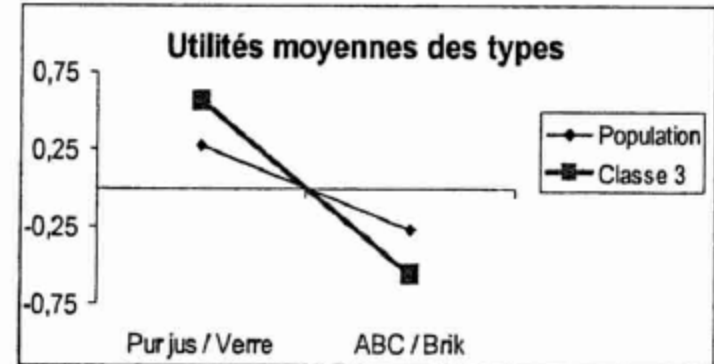
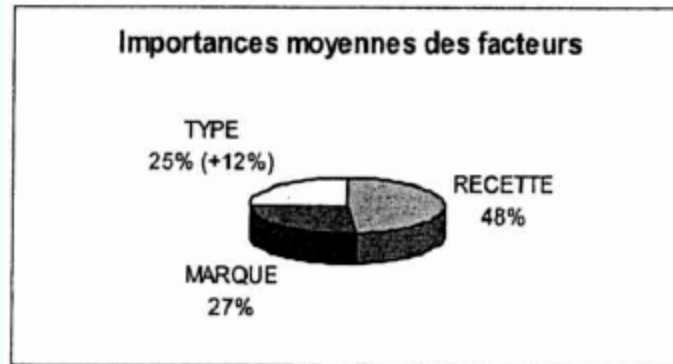
classe 2 du panel



**Les individus de la 2è classe
donnent une importance plus
élevée à la recette.**

24%

classe 3 du panel



**Les individus de la 3è classe
donnent une importance
plus élevée au type.**

31.3%



Cartographie des préférences



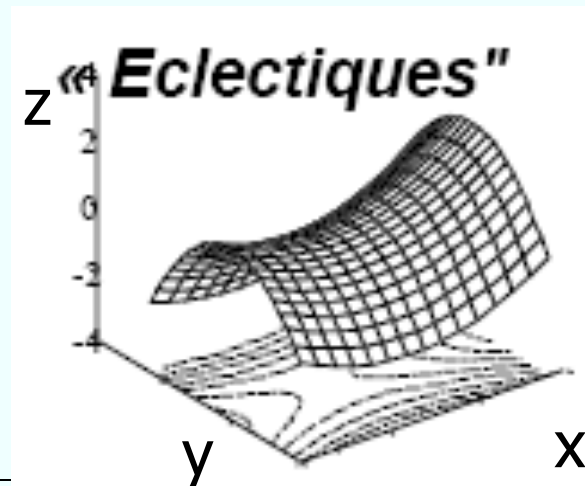
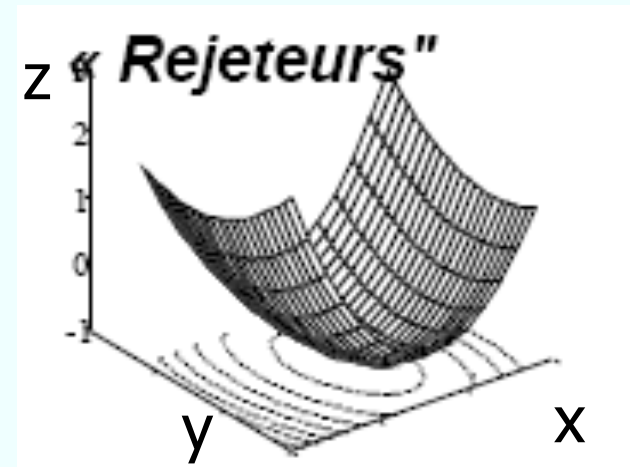
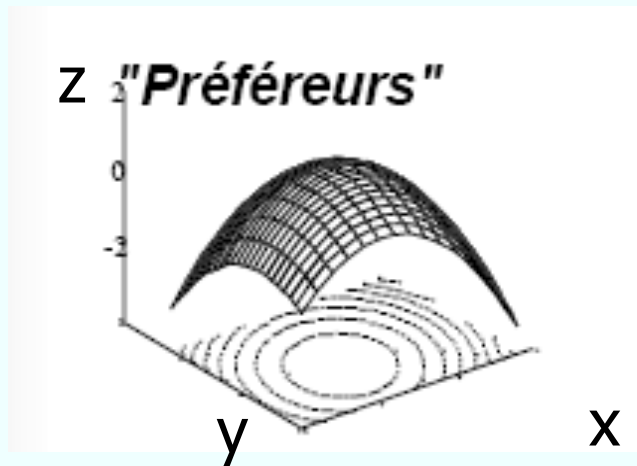
Cartographie des préférences

Objectifs

- Pour chaque sujet, construire un modèle expliquant la préférence par des grandeurs descriptives des produits
 - Grandeurs sensorielles
 - Grandeurs instrumentales (mesurées instrumentalement)
- Interpréter graphiquement les préférences
- Orienter la conception vers des formules préférables
- Segmenter les clients en fonction des préférences

Différentes formes de la préférence

Profils typiques de preference (axe z) en fonction de 2 variables sensorielles (x,y)





Exemple introductif

- ◆ Test d'acceptabilité d'un **dessert à la vanille**



- Objectif
 - Définir la formulation préférable au niveau sensoriel

- Données sensorielles

Jury d'experts

flaveur

texture

	Saveur de vanille : d1	Onctuosité : d2
C 1	2	1
C 2	8	1
C 3	9	2
C 4	3	3
C 5	8	6
C 6	1	3
C 7	2	8

7 produits
2 descripteurs



Exemple introductif

Consommateur
courant

- Données « hédoniques »
 - Notes de préférence P du sujet

7 produits : C1 à C7

3 sujets : A, B, C

P	A	B	C
C 1	2	6	5
C 2	7	6	4
C 3	8	6,5	7
C 4	4,5	7	8
C 5	9	7,5	2
C 6	1,5	7	1
C 7	5	8	1



Exemple introductif

- Méthode
 - tenter d'expliquer les préférences des sujets par les données sensorielles définies par les experts
- Outil
 - Régression linéaire multiple

Variable explicative : les descripteurs d1, d2

Variable à expliquer : note de préférence

observations
(P_i , $d1_i$, $d2_i$)



Estimateurs des
coefficients a , b , c
du modèle

$$\hat{P}_i = a.d1_i + b.d2_i + c$$

Exemple introductif

- Valeur prédite par la régression

$$\hat{P}_i = a.d1_i + b.d2_i + c$$

Tableau de synthèse des modèles de préférence

	a	b	c	R ²	F	Significatif (p=0.05)
A	0,82	0,42	-0,02	0,98	97,84	Oui
B	-0,02	0,27	6,02	0,94	31,17	Oui
C	0,15	-0,59	5,33	0,37	1,17	Non

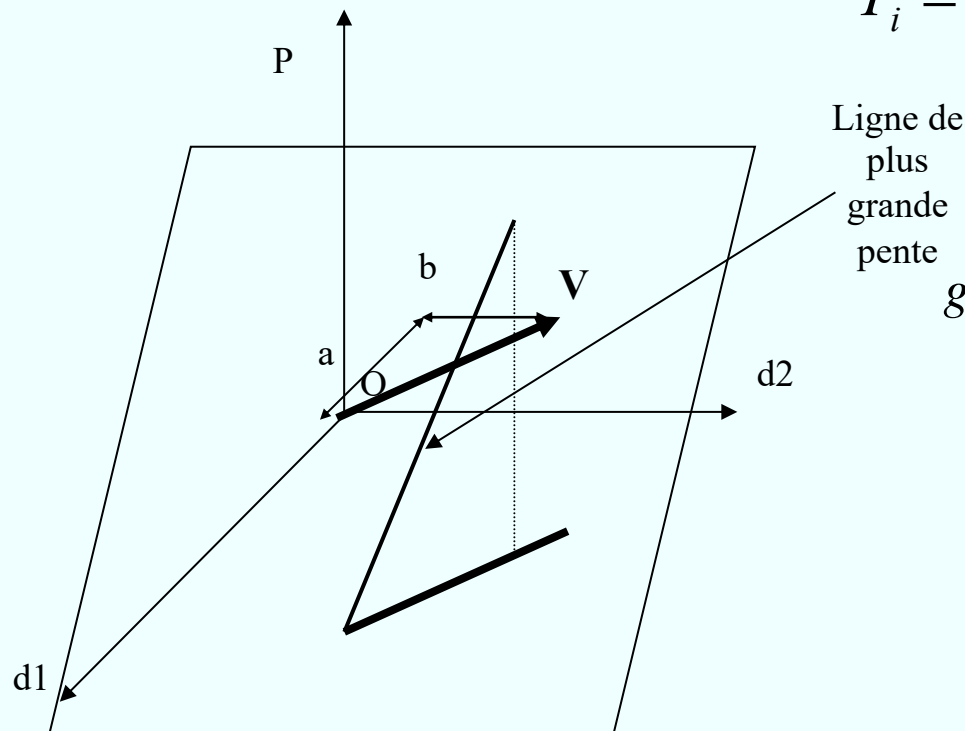
- Interprétation graphique de la régression
- Pour chaque sujet, tracé du **modèle vecteur (vecteur gradient)**

$$grad(\hat{P}) = \begin{pmatrix} \frac{\partial \hat{P}}{\partial d_1} \\ \frac{\partial \hat{P}}{\partial d_2} \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix}$$



Le modèle vecteur

- Situé arbitrairement à l'origine
- Colinéaire à la projection sur (d1, d2) de la ligne de plus grande pente du plan de régression (vecteur gradient)
 - Orienté vers l'augmentation de P
 - de norme prop. à R^2

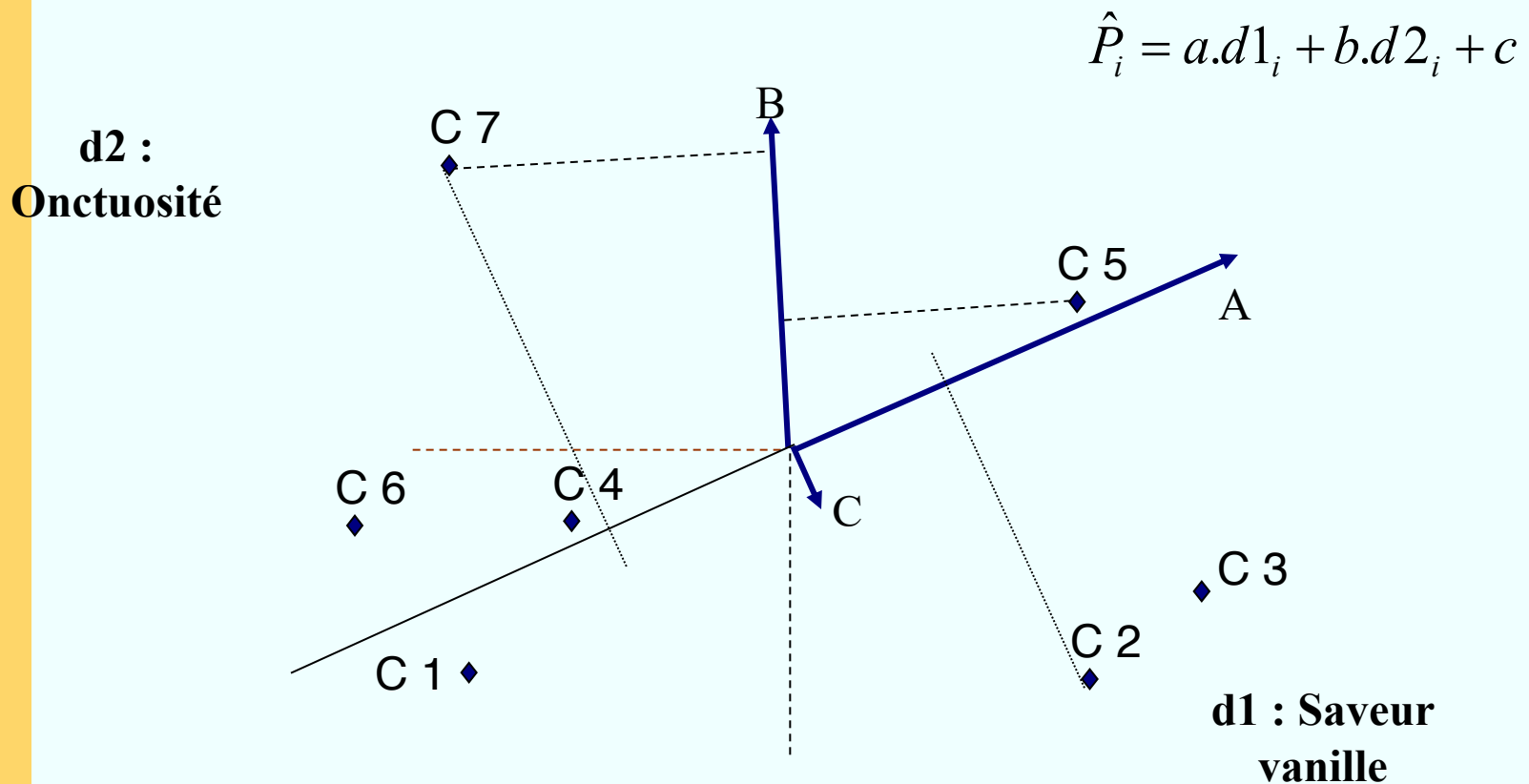


$$\hat{P}_i = a.d1_i + b.d2_i + c$$
$$\text{grad}(\hat{P}) = \begin{pmatrix} \frac{\partial \hat{P}}{\partial d_1} \\ \frac{\partial \hat{P}}{\partial d_2} \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix}$$

$$\mathbf{V} \sim \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix}$$

Exemple introductif

- **Modèle vecteur**
 - orienté vers l'augmentation de la préférence
 - les perpendiculaires au vecteur sont les iso-préférences





Exemple introductif

- Conclusions
 - sujet A :
 - Très bon ajustement du modèle ($R^2 = 98\%$)
 - aime bien à la fois une saveur vanille marquée et les crèmes lisses ($a = 0,82$; $b = 0,42$)
 - sujet B :
 - Très bon ajustement du modèle ($R^2 = 94\%$)
 - apprécie une crème très lisse ; indifférent au goût de vanille ($a = -0,02$; $b = 0,27$)
 - sujet C :
 - mauvais ajustement du modèle ($R^2 = 37\%$)
 - Pas d'interprétation des préférences à cause du modèle inadapté. La notation de préférence n'est pas expliquée par les descripteurs



Exemple introductif

- Cas du sujet C
 - Trois explications possibles :
 - C n'exprime pas de préférence,
 - » il en est incapable,
 - » il a répondu au hasard aux tests,
 - C est influencé par d'autres descripteurs que ceux proposés,
 - C a une structure de préférence plus complexe que les précédents, modèle vectoriel n'est pas adapté pour la décrire

Modèle vectoriel

$$\Rightarrow \hat{P}_i = a.d1_i + b.d2_i + c$$

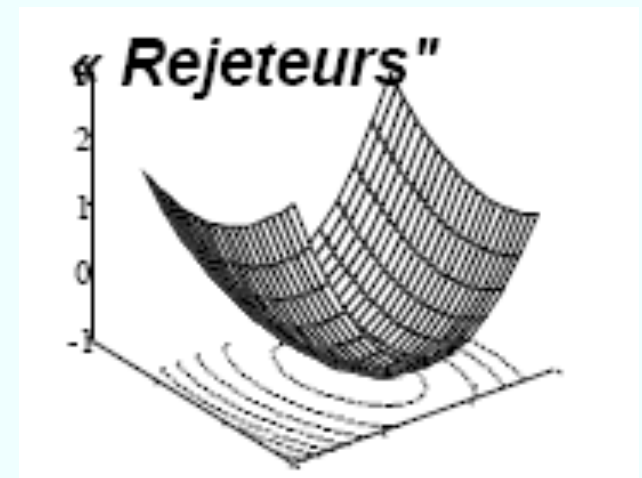
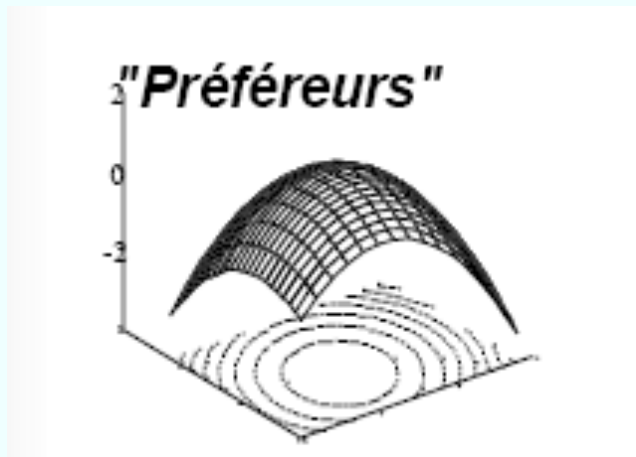
Modèle circulaire

$$\hat{P}_i = a_1.d1_i + a_2.d2_i + a_3.(d1_i^2 + d2_i^2) + c$$

Modèle circulaire

$$\hat{P}_i = a_1.d1_i + a_2.d2_i + a_3.(d1_i^2 + d2_i^2) + c$$

- Point idéal
 - $a_3 < 0$
 - Optimum de préférence
- Point anti idéal
 - $a_3 > 0$
 - Point de rejet





Modèle circulaire : sujet C

$$\hat{P}_i = a_1.d1_i + a_2.d2_i + a_3.(d1_i^2 + d2_i^2) + c$$

$$R^2=0.52$$

Coefficients du modèle :

	Constante	a1	a2	a3
C	0,443	1,996	0,979	-0,184

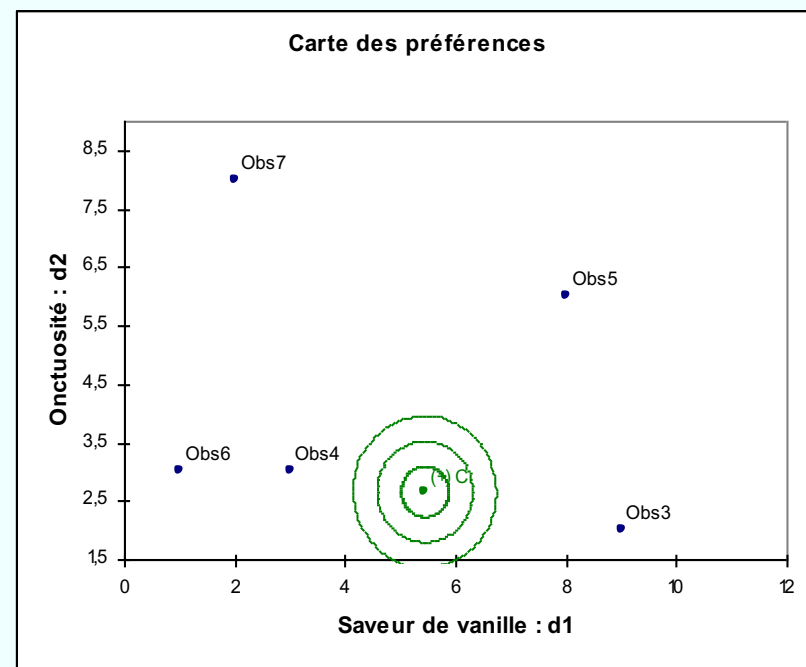
- Point idéal

- $a3 < 0$

- Optimum de préférence

Position de l'optimum

$$\left\{ \begin{array}{l} -\frac{a_1}{2a_3} \\ -\frac{a_2}{2a_3} \end{array} \right.$$





Cartographie des préférences

- ♦ interpréter graphiquement les préférences
- ♦ orienter la conception vers des formules préférables
- ♦ Construction d'un modèle individuel
 - on régresse le jugement hédonique sur les données perceptuelles, sensorielles ou instrumentales
- ♦ On utilise généralement les 2 premiers facteurs F_1 et F_2 de l'ACP pour régresser la préférence

Modèle vectoriel
$$\hat{P}_i = a.F1_i + b.F2_i + c$$

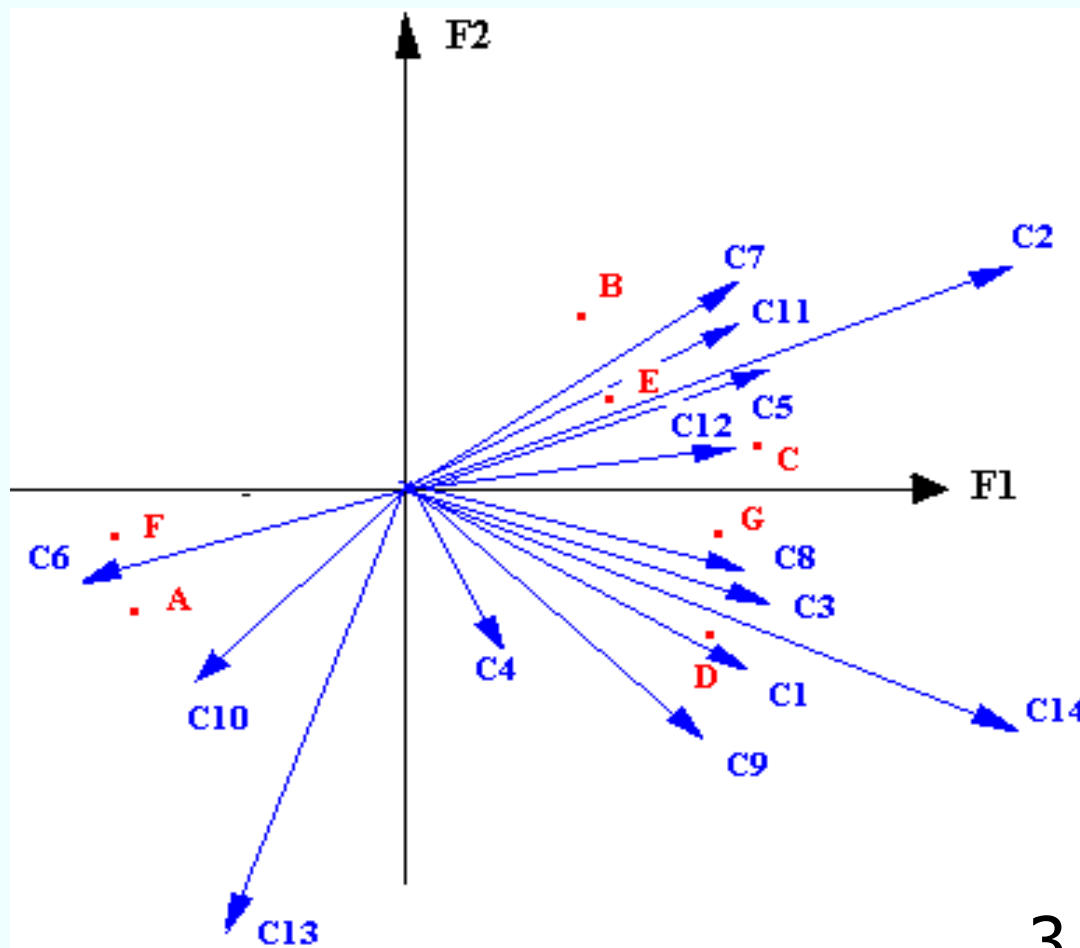
Modèle circulaire
$$\hat{P}_i = a_1.F1_i + a_2.F2_i + a_3.(F1_i^2 + F2_i^2) + c$$

Modèle elliptique
$$\hat{P}_i = a_1.F1_i + a_2.F2_i + a_3.(F1_i^2) + a_4.F2_i^2 + c$$

Modèle complet
$$\hat{P}_i = a_1.F1_i + a_2.F2_i + a_3.(F1_i^2) + a_4.F2_i^2 + a_5.F1.F2 + c$$



Panel de consommateurs : Modèle vectoriel



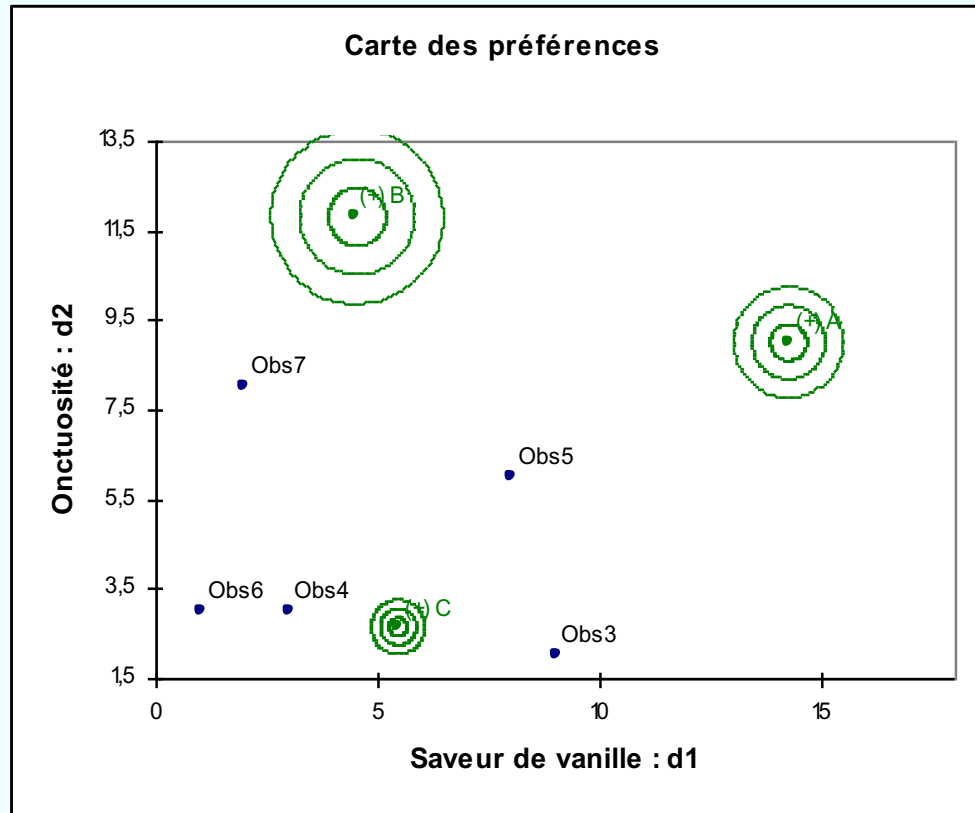
7 produits A, B, C, D, E, F, G

14 consommateurs C1 à C14

3 groupes de
consommateur
S



Modèle circulaire

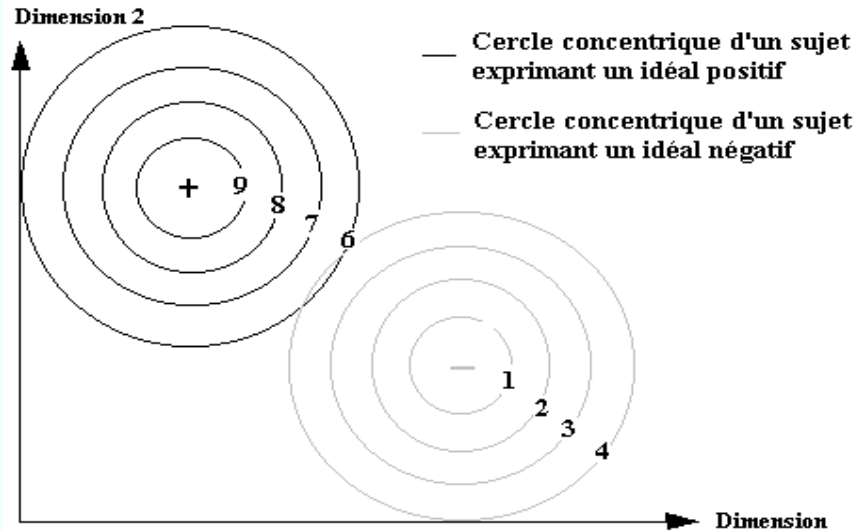


points idéaux
des 3 sujets A,
B, C

Modèle circulaire (point extreme)

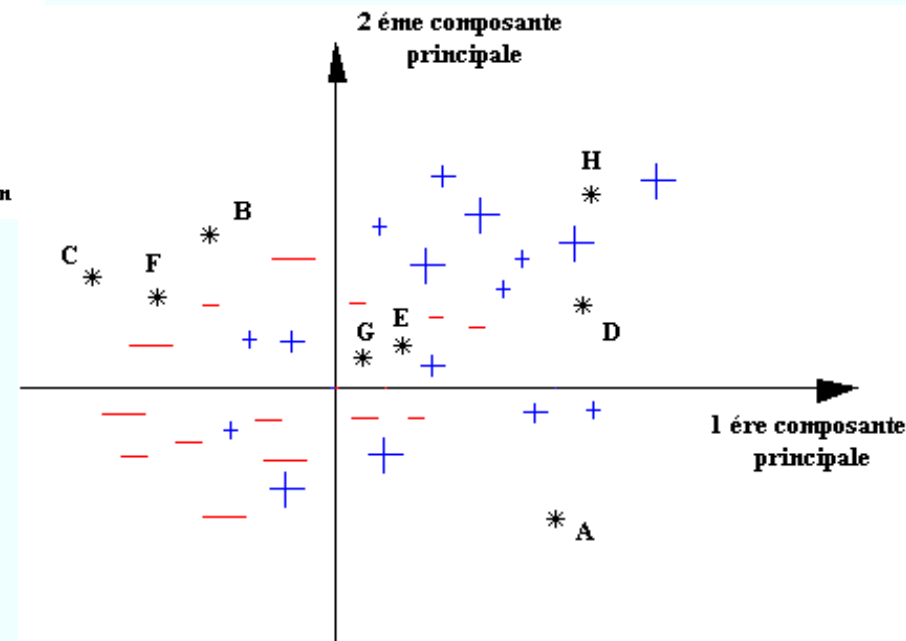
$$\hat{P}_i = a_1.d1_i + a_2.d2_i + a_3.(d1_i^2 + d2_i^2) + c$$

pt extreme $\begin{pmatrix} -a1/2a3 \\ -a2/2a3 \end{pmatrix}$



Produit segmentant

Idéal si $a3 < 0$
Anti-idéal si $a3 > 0$



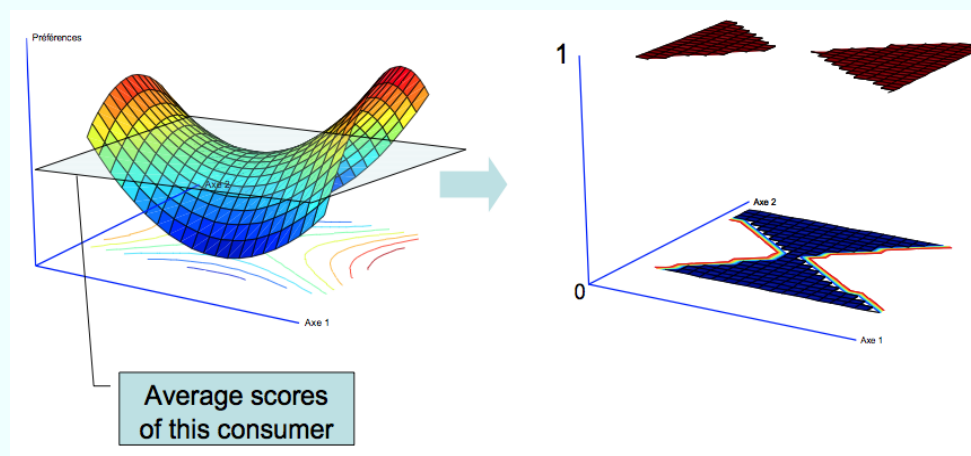
Agrégation des données individuelles

Seuil pour chaque sujet i

$$C_i(x,y) = 1 \text{ si } P_i(x,y) \geq \overline{P_i(x,y)}$$

$$C_i(x,y) = 0 \text{ si } P_i(x,y) < \overline{P_i(x,y)}$$

=1 si la préférence est supérieure à la valeur moyenne, 0 sinon



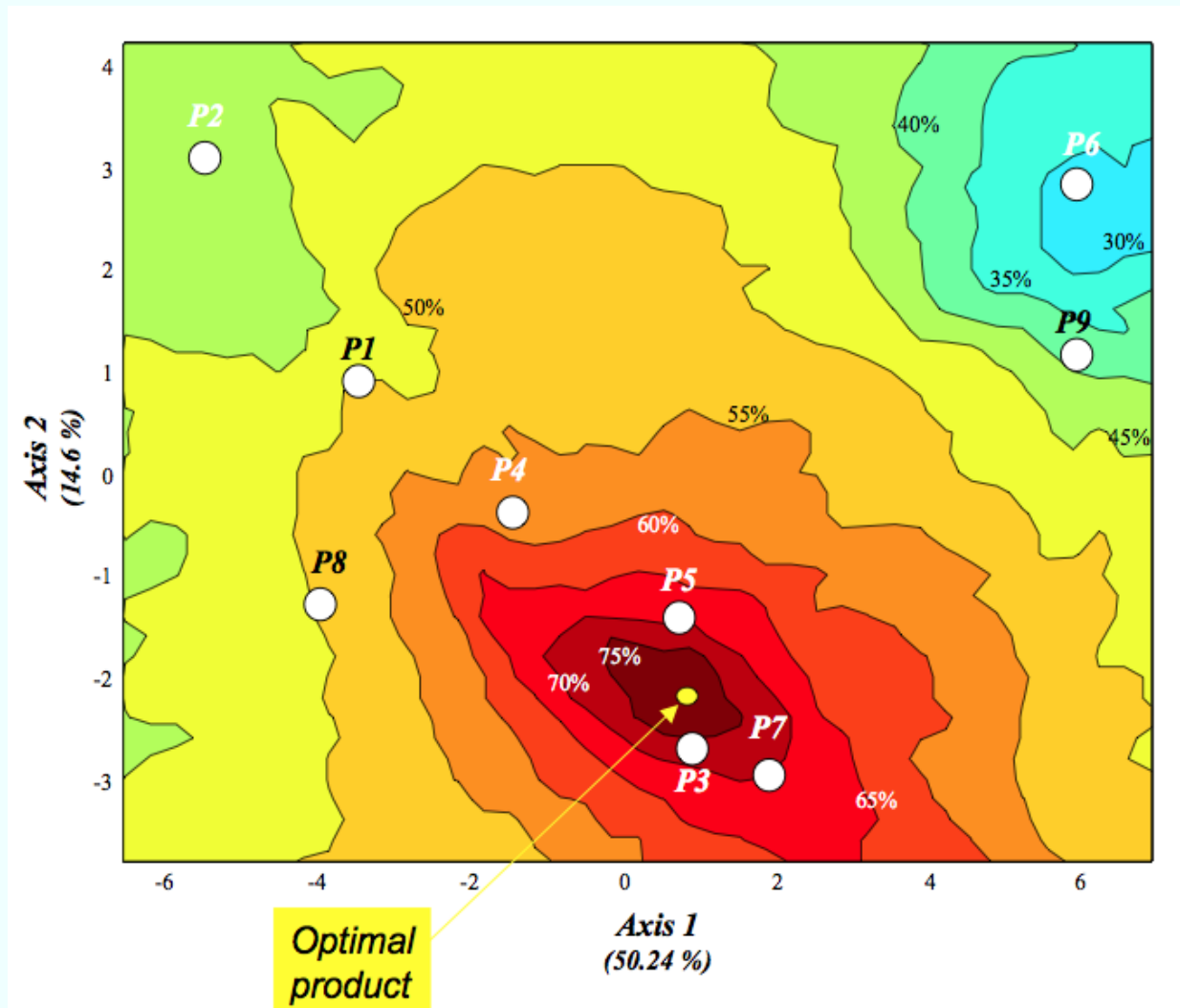
Agrégation de la préférence pour tous les sujets

$$P(x,y) = 100 \cdot \frac{\sum_{i=1}^n C_i(x,y)}{n}$$

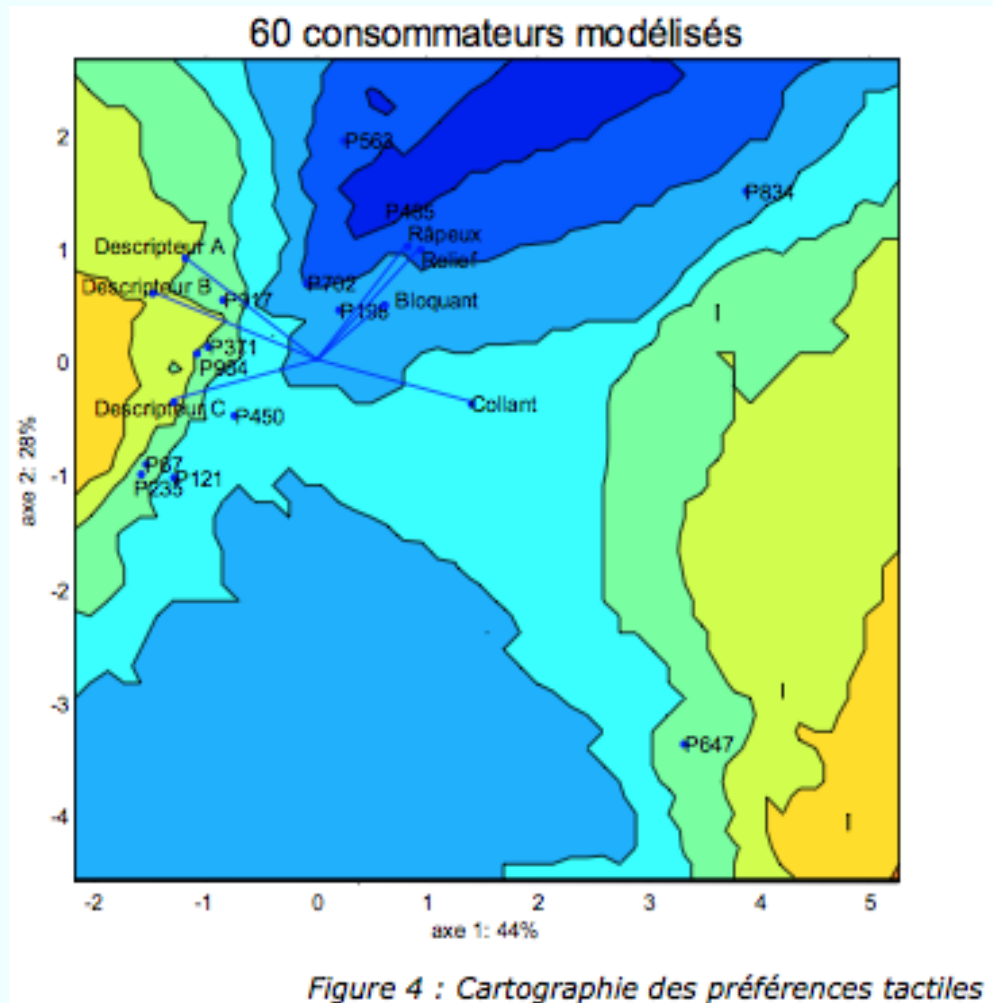
Surface de préférence

Surface de préférence

Courbes de niveau



Exemple – matériau volant



- ◆ « Rapeux » et « relief » inhibe la préférence
- ◆ Statut segmentant de « collant » et « descripteur A et B »



RESUME : cartographie des préférences

- ◆ Outil du marketing pour le positionnement concurrentiel
- ◆ Outil de synthèse de solution lorsque les attributs « perçus » du produit sont déterminants
 - alimentaire
 - cosmétique
 - vision
 - acoustique
- ◆ limites :
 - ne propose qu'un modèle très simplifié des préférences
 - effets de seuils, d'interaction, de saturation, non considérés
 - difficulté pour trouver les caractéristiques produit pertinentes
 - innovation de rupture
 - Méthode peu adaptée pour l'innovation...