

3. ACM – Analyse des Correspondances Multiples

Plan

1. Exemples, problématique
2. Le tableau de données étudié
3. Principe d'une ACM
4. L'ajustement des deux nuages de profils
5. Axes factoriels et inerties
6. Interprétation des résultats d'une ACM

3.1 – Exemples, problématique

Exemple 1. Crédit à la consommation

- 66 clients d'un organisme de crédit
- 5 variables de situation bancaire
- 6 variables socio – démographiques

Extrait du tableau des données

Cient	Marche	Apport	Impaye	Assurance	Endettement	Famille	Enfants	Logement	Profession	Intitule	Age
1	Renovation	pas_Apport	Imp_0	AID	End_1	Union libre	Enf_0	Loge Famille	Ouvrier non qualif	MR	50
2	Renovation	pas_Apport	Imp_0	Sans Assur	End_2	Marie	Enf_0	Proprietaire	Ouvrier qualifie	MR	40
3	Voiture	Apport	Imp_0	AID	End_3	Marie	Enf_1	Acced Propri	Ouvrier qualifie	MR	30
4	Renovation	pas_Apport	Imp_0	Senior	End_2	Veuf	Enf_0	Proprietaire	Retraite	MR	60
5	Scooter	Apport	Imp_3 et +	AID	End_4	Union libre	Enf_0	Loge par employ	Cadre moyen	MR	20
6	Renovation	pas_Apport	Imp_0	Senior	End_3	Marie	Enf_0	Proprietaire	Retraite	MR	60
7	Renovation	Apport	Imp_0	Senior	End_2	Veuf	Enf_0	Proprietaire	Retraite	MME	60
8	Renovation	Apport	Imp_0	Sans Assur	End_2	Veuf	Enf_0	Proprietaire	Retraite	MME	60
9	Mobil Ameub	Apport	Imp_0	Sans Assur	End_1	Union libre	Enf_0	Loge par employ	Ouvrier non qualif	MR	40
10	Renovation	Apport	Imp_0	Senior	End_2	Marie	Enf_0	Proprietaire	Retraite	MR	60
⋮	⋮					⋮					⋮
64	Renovation	pas_Apport	Imp_3 et +	AID	End_4	Divorce	Enf_3	Proprietaire	Ouvrier non qualif	MME	30
65	Voiture	Apport	Imp_1_ou_2	AID+Chomage	End_4	Marie	Enf_1	Locataire	Ouvrier qualifie	MR	20
66	Renovation	pas_Apport	Imp_1_ou_2	Senior	End_2	Marie	Enf_0	Proprietaire	Retraite	MR	60

Objectifs (1) : niveau des **individus** [lignes]

- Connaître, étudier, caractériser la clientèle de l'organisme de crédit
- Typologie des clients selon leur profil « crédit / bancaire »
- Quels clients ont des profils similaires ?

Objectifs (2) : niveau des **variables** [colonnes]

- Étude des liaisons (correspondances) entre les variables « crédit / bancaire »
Ex : Le choix d'une assurance est-il lié à un niveau d'endettement ?
- Étude des liaisons entre les variables socio démographiques et les variables « crédit / bancaire »
*Ex : Comment le niveau d'endettement est-il relié à la classe d'âge ?
À la profession ?*

Correspondances entre variables actives



crédit - bancaire

socio - demo

Cient	Marche	Apport	Impaye	Assurance	Endettement	Famille	Enfants	Logement	Profession	Intitule	Age
1	Renovation	pas_Apport	Imp_0	AID	End_1	Union libre	Enf_0	Loge Famille	Ouvrier non qualif	MR	50
2	Renovation	pas_Apport	Imp_0	Sans Assur	End_2	Marie	Enf_0	Proprietaire	Ouvrier qualifie	MR	40
3	Voiture	Apport	Imp_0	AID	End_3	Marie	Enf_1	Acced Propri	Ouvrier qualifie	MR	30
4	Renovation	pas_Apport	Imp_0	Senior	End_2	Veuf	Enf_0	Proprietaire	Retraite	MR	60
5	Scooter	Apport	Imp_3 et +	AID	End_4	Union libre	Enf_0	Loge par employ	Cadre moyen	MR	20
6	Renovation	pas_Apport	Imp_0	Senior	End_3	Marie	Enf_0	Proprietaire	Retraite	MR	60
7	Renovation	Apport	Imp_0	Senior	End_2	Veuf	Enf_0	Proprietaire	Retraite	MME	60
8	Renovation	Apport	Imp_0	Sans Assur	End_2	Veuf	Enf_0	Proprietaire	Retraite	MME	60
9	Mobil Ameub	Apport	Imp_0	Sans Assur	End_1	Union libre	Enf_0	Loge par employ	Ouvrier non qualif	MR	40
10	Renovation	Apport	Imp_0	Senior	End_2	Marie	Enf_0	Proprietaire	Retraite	MR	60
:	:	:	:	:	:	:	:	:	:	:	:
64	Renovation	pas_Apport	Imp_3 et +	AID	End_4	Divorce	Enf_3	Proprietaire	Ouvrier non qualif	MME	30
65	Voiture	Apport	Imp_1_ou_2	AID+Chomage	End_4	Marie	Enf_1	Locataire	Ouvrier qualifie	MR	20
66	Renovation	pas_Apport	Imp_1_ou_2	Senior	End_2	Marie	Enf_0	Proprietaire	Retraite	MR	60



Correspondances entre variables actives
et illustratives (supplémentaires)

Exemple 2. Enquête sur la prison

- 1500 adultes français
- 50 questions sur leur connaissance de la prison
- 10 questions sur leur opinion de la prison
- 6 variables socio – démographique

Extrait du tableau des données

Enquêté	WC cloisonnés	frigo autorisé	télé autorisée	superficie cellule	... libé. anticipée normale	conditions de détention	...	sexe	...	diplôme	situation prof	orientation politique
1	non	oui	oui	10	non	a.bonnes		homme		DPS	temps complet	centre droit
2	non	non	oui	6	non	a.mauvaises		femme		DPS	temps complet	autre
3	non	non	oui	6	non	mauvaises		femme		sans diplôme	temps complet	centre
4	non	oui	oui	10	non	bonnes		homme		BAC	étudiant	droite
5	non	non	oui	6	oui	mauvaises		femme		>BAC+2	retaité	centre gauche
6	non	non	oui	5	oui	a.mauvaises		homme		DPS	temps complet	centre gauche
7	non	non	oui	10	oui	mauvaises		femme		BAC+2	étudiant	gauche
8		oui	oui	10	oui	bonnes		femme		BAC	temps complet	centre gauche
9	non	non	non	6	non	mauvaises		homme		>BAC+2	chômeur	autre
10	non	non	oui	8	non	a.bonnes		femme		sans diplôme	femme au foyer	centre droit
	non	non	oui	10	oui	a.mauvaises		homme		BAC	temps complet	centre droit
:												
1494	non		oui	7,5	oui	a.bonnes		homme		sans diplôme	temps complet	centre droit
1495	non	non	oui	6	non	mauvaises		femme		BEPC	temps complet	centre droit
1496	non	non	oui	4	non	a.bonnes		homme		BAC	temps complet	centre
1497	oui	oui	oui	20	non	a.bonnes		femme		BAC+2	temps complet	droite
1498	non	non	oui	4	non	bonnes		femme		BAC+2	étudiant	centre droit
1499	non		non	8	oui			femme		BAC	temps complet	centre droit
1500	non		oui		oui	a.mauvaises		homme		>BAC+2	temps complet	autre

Exemple 3. Enquête « cinéma »

- Enquête en milieu étudiant – 93 étudiants
- Sexe, âge, abonnement, fréquence
- Questions d'appréciation de 14 types de films :

documentaire
fantastique
arts

action
psychologique
historique

comédie
romantique
aventure

horreur
DA/Manga
thriller

policier
comédie musicale

Extrait du tableau des données

Etud	Sexe	Age	Fréquence	Abonnement	Festival	Policier	Thriller	Fantastique	Psycho	Action	Aventures	Horreur	Comédie	...
1	f	21	4	3	n	4	3	3	3	2	2	1	3	
2	f	21	3	1	n	1	1	1	3	1	1	1	4	
3	m	21	2	3	o	2	4	1	3	3	2	1	5	
4	f	18	2	2	n	3	3	1	4	2	3	1	4	
5	m	18	2	3	o	2	4	4	4	4	2	4	2	
6	m	23	3	4	o	3	3	2	3	1	1	1	2	
7	m	19	3	4	o	3	4	5	5	3	2	2	3	
8	m	19	3	4	o	3	2	2	3	4	4	1	3	
9	f	19	4	2	o	1	3	5	4	2	3	1	4	
10	f	19	3	1	o	1	2	3	4	1	1	1	3	
11	f	22	2	3	n	1	2	2	3	1	2	1	3	
12	f	26	2	4	n	1	4	2	4	2	1	1	5	
13	f	24	3	4	n	2	4	1	3	1	1	3	3	
:														
92	f	24	2	4	o	1	1	2	3	1	2	2	3	
93	m	18	3	3	o	2	3	4	4	2	2	1	1	

Exemple 4. Enquête « Ouest France »

- 340 personnes enquêtées
- 8 variables socio démographiques
- 26 rubriques du journal (lu = 1 ; non lu = 2)

Extrait du tableau des données

Ident	Habitat	Sexe	SituFamille	Age	C SP	NbEnfants	NivInstruct	Habitat	InfoLocales	InfoPolitique	Sports
1	0	1	2	1	8	1	5	4	2	1	1
2	0	2	2	3	3	1	4	4	1	1	2
3	4	2	0	0	5	1	5	4	1	1	2
4	0	1	1	1	0	1	5	0	2	1	1
5	4	2	1	0	5	1	5	4	1	1	2
6	3	2	0	1	5	1	3	3	2	2	2
7	0	1	2	4	4	2	5	1	1	1	1
8	1	2	1	1	0	1	5	0	1	2	2
9	2	0	2	3	5	3	1	1	1	1	2
10	2	1	2	4	9	1	0	1	1	1	2
11	0	1	2	3	5	3	5	1	1	1	2
12	1	1	1	2	9	1	4	0	2	1	2
13	0	2	0	2	1	3	3	2	1	1	2
14	1	2	2	2	1	3	1	0	2	2	2
15	1	2	2	0	7	4	1	1	1	2	2
16	3	1	2	2	2	2	0	3	1	2	1
17	3	0	1	2	5	1	3	4	2	1	2
18	3	2	2	0	4	3	5	2	1	2	1
19	3	0	2	2	6	4	1	1	1	1	2
20	2	1	2	2	4	3	5	0	1	1	2
21	4	1	1	2	4	1	0	3	2	1	1

Exemple 5. Test appareil photo numériques

- Test de 27 modèles d'appareil photo numérique (reflex et hybrides)
- Caractéristiques techniques, performances, notes d'appréciation labo
- Objectif : comparer les appareils du point de vue de leurs caractéristiques techniques

Extrait des données

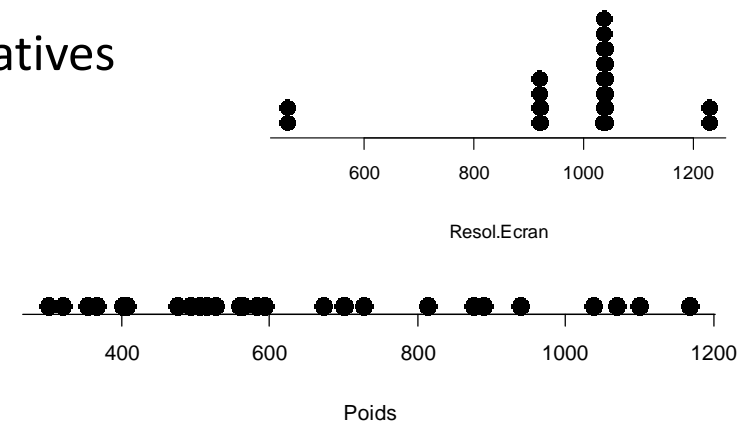
Marque	Modèle	Type	Prix	Caractéristiques techniques								Notes, performances				
				Capteur	FocaleMin	FocaleMax	Grossissement	Resolution	Poids	Viseur	Tactile	NoteGlobale	QualitéPhoto	QualitéVidéo	QualitéEcran	NoteDép
NIKON	D7200 18-55	Reflex	1140	24	27	82,5	3	1229	940	oui	non	15,1	4	5	4	5
NIKON	D7200 18-105	Reflex	1150	24	27	157,5	5,8	1229	1168	oui	non	14,5	4	5	4	5
NIKON	D3300	Reflex	430	24	27	82,5	3	921	700	oui	non	14,1	4	4	5	4
NIKON	D5500 18-55	Reflex	750	24	27	82,5	3	1037	674	oui	oui	14,1	4	4	5	5
CANON	EOS 1200D	Reflex	350	18	29	88	3	460	728	oui	non	13,9	4	4	4	3
NIKON	D5500 18-105	Reflex	950	24	27	157,5	5,8	1037	890	oui	oui	13,9	4	4	5	5
CANON	EOS 760D	Reflex	1100	24	29	216	7,5	1040	1101	oui	oui	13,8	4	4	4	4
CANON	EOS 750D 18-55	Reflex	800	24	29	88	3	1040	815	oui	oui	13,8	4	4	4	4
CANON	EOS 750D 18-135	Reflex	1050	24	29	216	7,5	1040	1070	oui	oui	13,5	3	4	4	4
PENTAX	K S2	Reflex	700	20	27,5	76,5	2,8	921	877	oui	non	13,3	4	3	4	4
PENTAX	K 3 II	Reflex	990	24	27	82,5	3	1037	1039	oui	non	13,1	4	3	4	3
PANASONIC	LUMIX DMC G7	Hybride	800	16	28	84	3	1040	561	oui	oui	15	4	4	5	4
PANASONIC	LUMIX DMC G85	Hybride	420	16	24	64	2,7	1040	355	non	oui	14,6	4	4	5	4

Choix des variables actives

- Celles qui déterminent les ressemblances entre modèles d'appareil photo
→ les **caractéristiques techniques**
- Mais : certaines caract. sont de nature quantitative (ex. : poids, résolution)
- Les caractéristiques quanti. doivent être transformées en quali. afin de pouvoir jouer un rôle actif

Recodage des variables

- **Découpage en classes** des variables quantitatives
- Choix du découpage
→ en se basant sur des graphiques
→ mise en classe automatique



Objectifs (1) : niveau des **individus [lignes]**

- Selon quels critères les modèles se différencient-ils le plus ?
Quels sont les principaux facteurs de variabilité des modèles d'appareils ?
- Peut-on établir une typologie des modèles selon leur profil « technique » ?
- Hybrides et Reflex sont-ils clairement discriminés, séparés ?

Objectifs (2) : niveau des **variables [colonnes]**

- Étude des liaisons (correspondances) entre les caractéristiques techniques
Ex : un modèle avec un bon capteur est-il toujours associé à un bon écran ?
- Étude des liaisons entre les caractéristiques techniques et les performances ou les notes obtenues par les modèles
- Étude des liaisons entre le type de modèle (Reflex, Hybride) et les caractéristiques techniques ou les performances

Avant l'ACM : analyse univariée : les tris à plat

Démarche

- Par variable, analyser les fréquences d'apparition des modalités
- En cas de modalité « rare » (ex. Logiciel SPAD : %min = 2%)
 1. Effectuer des regroupements entre modalités « voisines » (si possible...)
 2. Remplacer par la modalité rare par une modalité *au hasard*
 3. Lui substituer la modalité la plus fréquente

Intérêt en ACM

- Une modalité rare joue souvent un rôle prédominant en ACM
- Certains axes factoriels sont engendrés par des modalités trop rarement choisies (non souhaitable !)

Distributions de fréquences : variables actives

Viseur	Tactile
Vise.non:10	Tact.non:11
Vise.oui:17	Tact.oui:16

Focale.min.F	Focale.Max.F	Grossissement.F
Foc.m.<25: 7	Foc.M.<80 : 8	Gross<3:23
Foc.m.>27:20	Foc.M.>140 : 4	Gross>3: 4
	Foc.M.80-100:15	

Poids.F	Ecran.Reso.F	Capteur.Reso.F
Pds.513-711:9	Res.Ecr.bon :15	Capt.12-16: 9
Pds<513 :9	Res.Ecr.max : 2	Capt.16-22: 4
Pds>711 :9	Res.Ecr.moins: 2	Capt<12 : 1
	Res.Ecr.moy : 8	Capt>22 :13

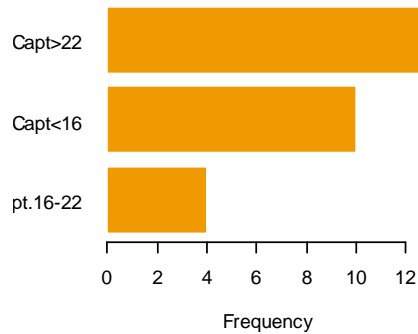
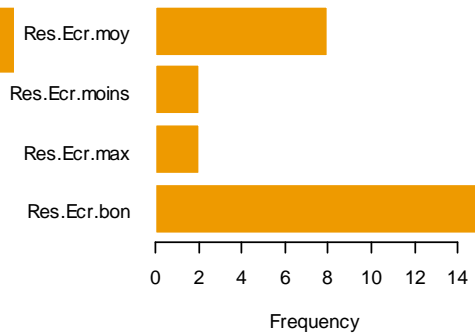
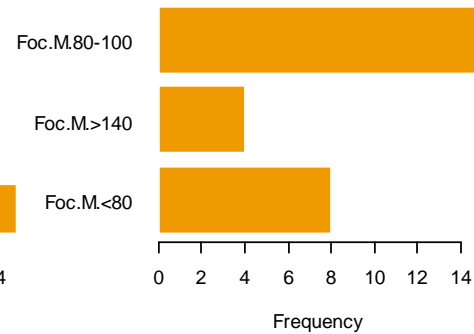
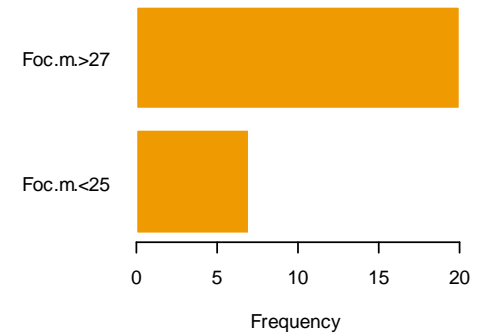
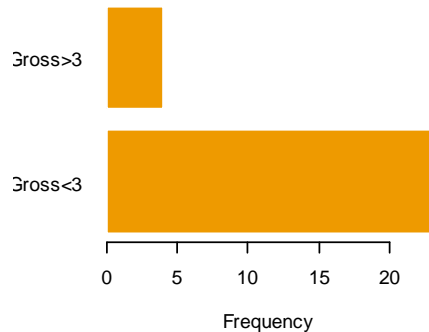
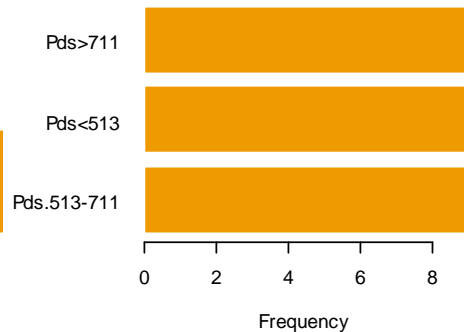
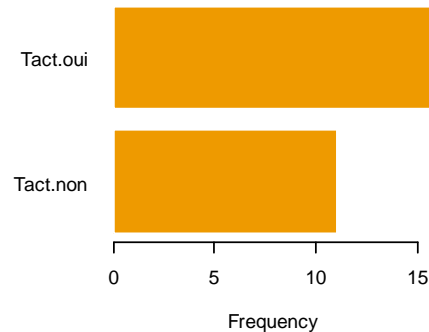
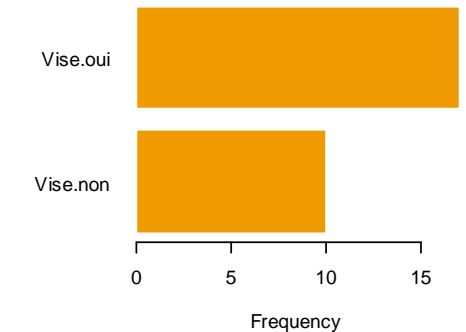
Distributions de fréquences : variables actives

Après recodage de la modalité « rare »

Viseur	Tactile
Vise.non:10	Tact.non:11
Vise.oui:17	Tact.oui:16

Focale.min.F	Focale.Max.F	Grossissement.F
Foc.m.<25: 7	Foc.M.<80 : 8	Gross<3:23
Foc.m.>27:20	Foc.M.>140 : 4	Gross>3: 4
	Foc.M.80-100:15	

Poids.F	Ecran.Reso.F	Capteur.Reso.F
Pds.513-711:9	Res.Ecr.bon :15	Capt.16-22: 4
Pds<513 :9	Res.Ecr.max : 2	Capt<16 :10
Pds>711 :9	Res.Ecr.moins: 2	Capt>22 :13
	Res.Ecr.moy : 8	

Capteur.Reso.F

Ecran.Reso.F

Focale.Max.F

Focale.min.F

Grossissement.F

Poids.F

Tactile

Viseur


3.2 – Le tableau de données étudié

Nature des variables

- Un tableau *individus x variables*
- Variables **actives** : **qualitatives**
- Variables **illustratives** (= supplémentaires) : **qualitatives** et / ou **quantitatives**

Variables **actives**

Variables **illustratives**

Modèle														
	Viseur	Tactile	Focale.min.F	Focale.Max.F	Grossissement.F		Ecran.Reso.F	Capteur.Reso.F	Type	Prix	NoteGlobale	Qualité photo	Qualité vidéo	
D7200 18-55	Vise.oui	Tact.non	Foc.m.>27	Foc.M.80-100	Gross<3	Pds>711	Res.Ecr.max	Capt>22	Reflex	1140	15.1	2	3	2
D7200 18-10	Vise.oui	Tact.non	Foc.m.>27	Foc.M.>140	Gross>3	Pds>711	Res.Ecr.max	Capt>22	Reflex	1150	14.5	2	3	2
D3300	Vise.oui	Tact.non	Foc.m.>27	Foc.M.80-100	Gross<3	Pds.513-711	Res.Ecr.moy	Capt>22	Reflex	430	14.1	2	2	3
D5500 18-55	Vise.oui	Tact.oui	Foc.m.>27	Foc.M.80-100	Gross<3	Pds.513-711	Res.Ecr.bon	Capt>22	Reflex	750	14.1	2	2	3
EOS 1200D	Vise.oui	Tact.non	Foc.m.>27	Foc.M.80-100	Gross<3	Pds>711	Res.Ecr.moin	Capt.16-22	Reflex	350	13.9	2	2	2
D5500 18-10	Vise.oui	Tact.oui	Foc.m.>27	Foc.M.>140	Gross>3	Pds>711	Res.Ecr.bon	Capt>22	Reflex	950	13.9	2	2	3
EOS 760D	Vise.oui	Tact.oui	Foc.m.>27	Foc.M.>140	Gross>3	Pds>711	Res.Ecr.bon	Capt>22	Reflex	1100	13.8	2	2	2
EOS 750D 18	Vise.oui	Tact.oui	Foc.m.>27	Foc.M.80-100	Gross<3	Pds>711	Res.Ecr.bon	Capt>22	Reflex	800	13.8	2	2	2

Le tableau disjonctif complet (TDC)

Le tableau de données initial est transformé en un *tableau disjonctif complet*

Tableau des données initiales
Codage « condensé »

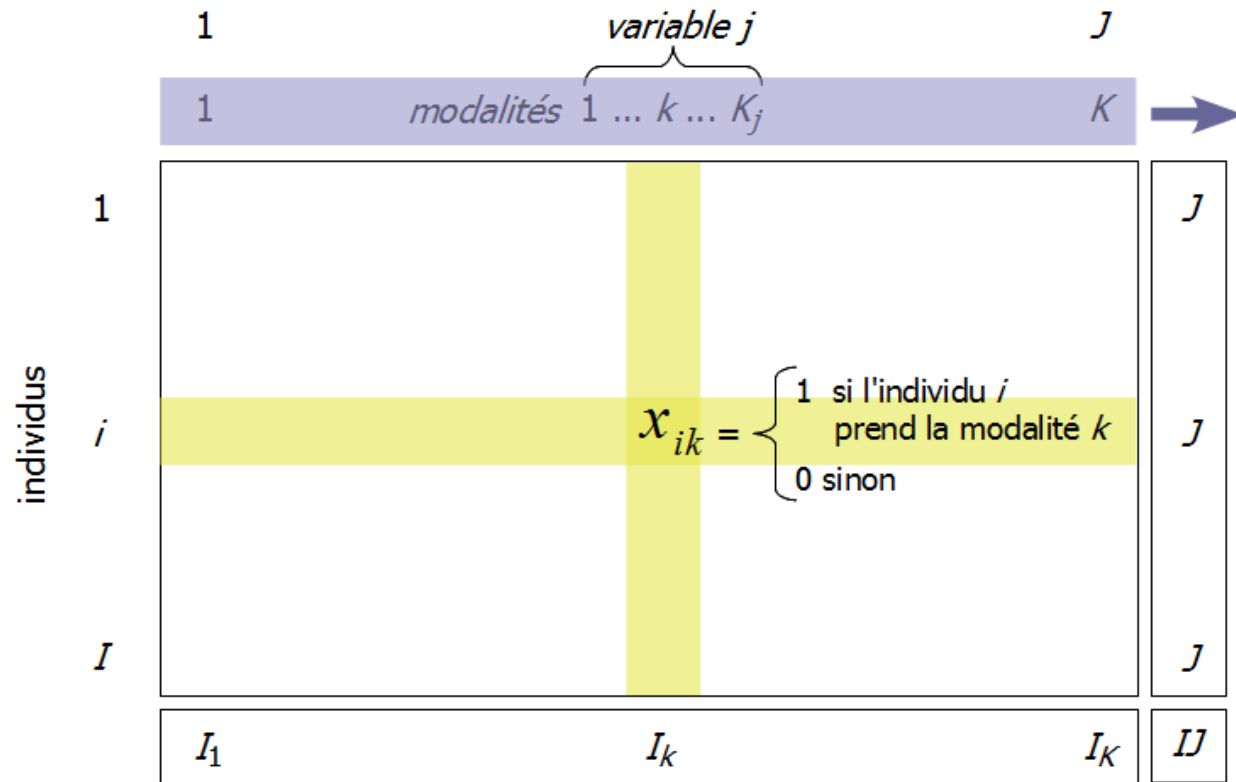
Tactile	Focale.min.F	Focale.Max.F
Tact.oui	Foc.m.>27	Foc.M.>140
Tact.non	Foc.m.>27	Foc.M.<80
Tact.non	Foc.m.>27	Foc.M.80-100
Tact.oui	Foc.m.>27	Foc.M.80-100
Tact.oui	Foc.m.<25	Foc.M.<80

extrait

Tableau des données transformées
Codage « binaire » (TDC)

Tactile		Focale min		Focale Max		
oui	non	< 25	> 27	< 80	80 - 100	> 140
1	0	0	1	0	0	1
0	1	0	1	1	0	0
0	1	0	1	0	1	0
1	0	0	1	0	1	0
1	0	1	0	1	0	0

Notations

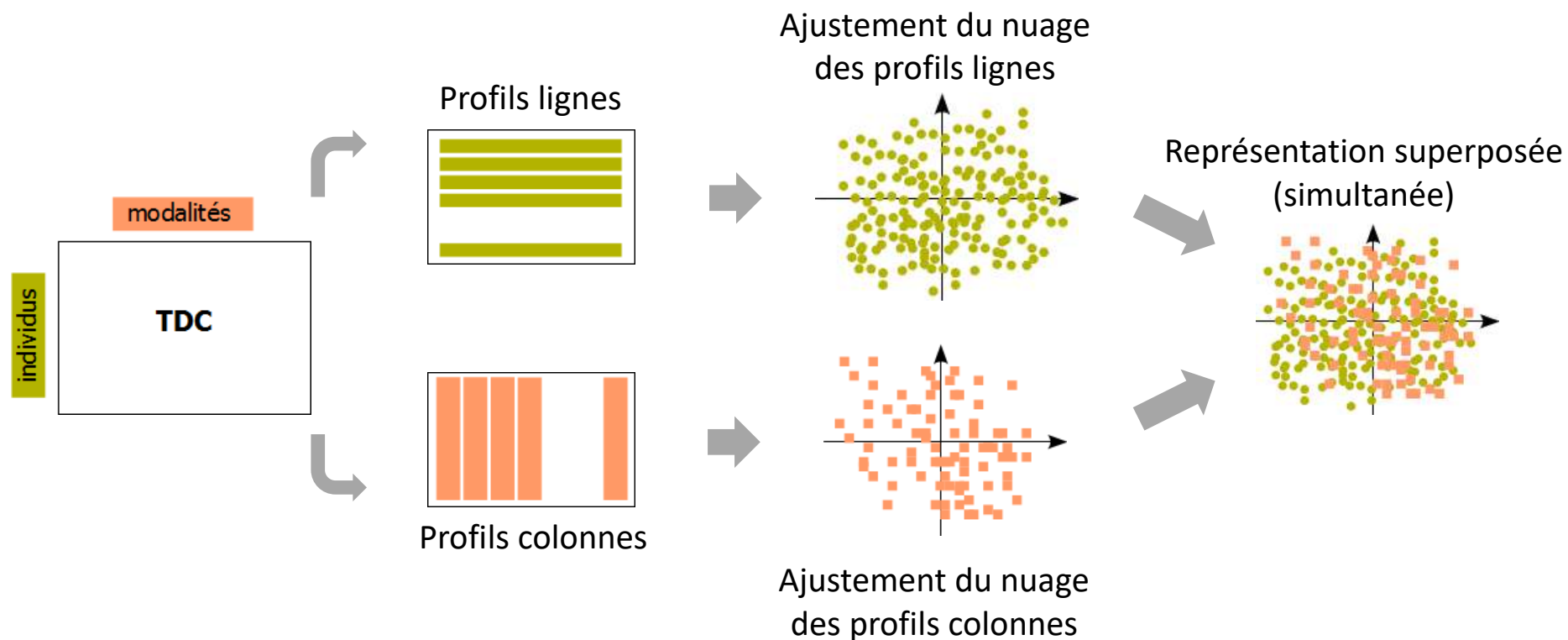


Rôle important
joué par les
modalités
en ACM

- **Lignes** : ressemblances entre individus
- **Colonnes** : étudier les liaisons (correspondances multiples) entre les modalités

3.3 – Le principe d'une ACM

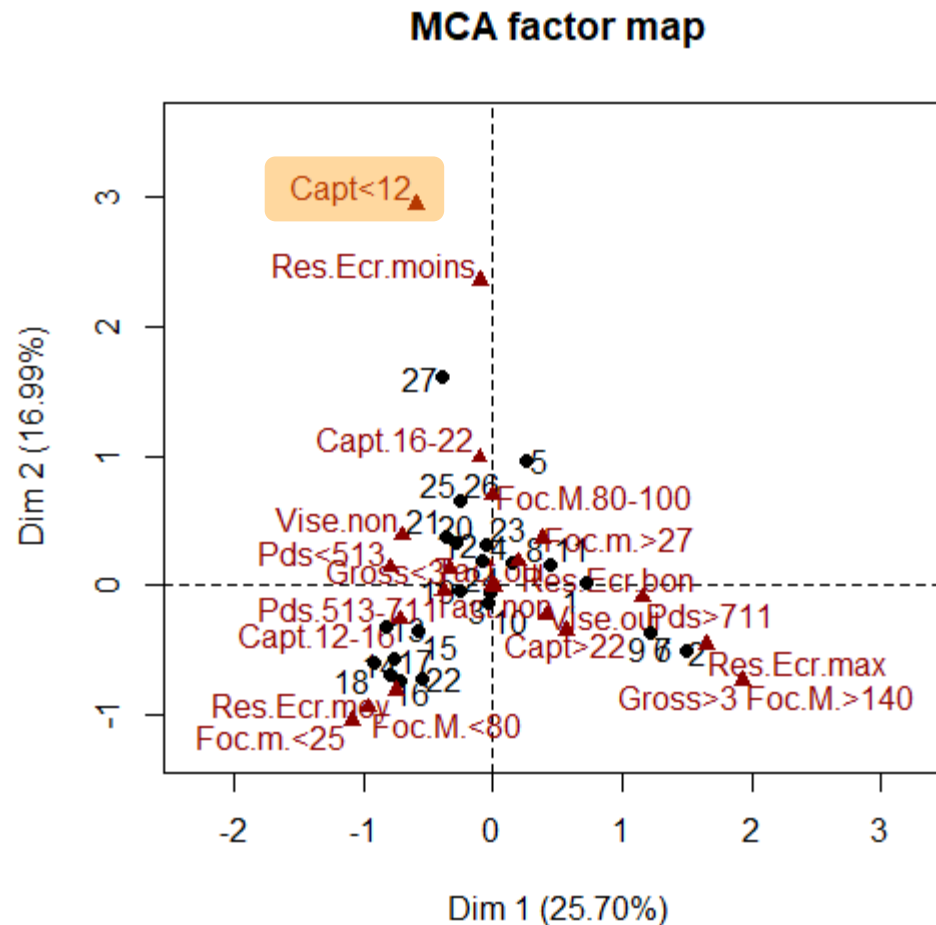
- L'ACM est une **AFC** !
- Une AFC appliquée au **tableau disjonctif complet**



MCA factor map



Représentation superposée, axes 1 et 2 – Avec la modalité rare



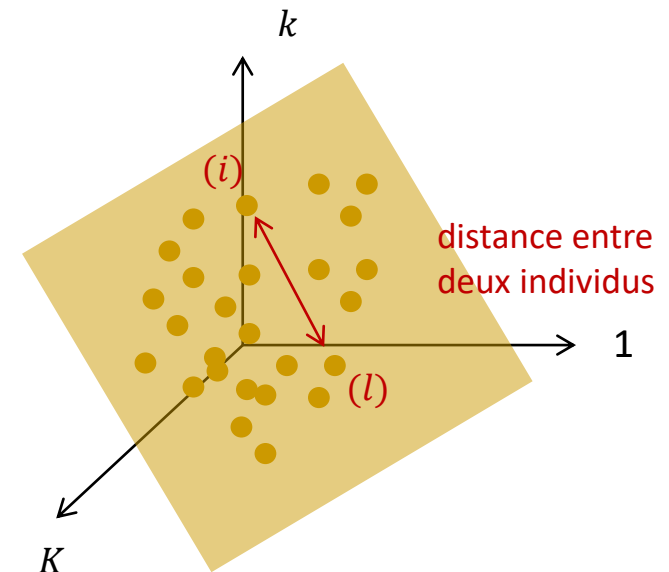
3.4 – L'ajustement des deux nuages de profils

- Comme pour toute analyse factorielle : maximisation de l'inertie projetée
- Mesure de distance utilisée : celle du **khi2**, comme en AFC

Ajustement du nuage des profils - individus

	1	...	k	...	K	Σ
1						1
\vdots						
i	$\frac{x_{i1}}{J}$...	$\frac{x_{ik}}{J}$...	$\frac{x_{iK}}{J}$	1
\vdots						
I						1

Poids d'un individu (i) : $m_i = \frac{1}{I}$



Distance entre deux individus

Que donne l'application de la distance du khi2 entre deux profils lignes ?

Tactile		Focale min		Focale Max		
oui	non	< 25	> 27	< 80	80 - 100	> 140
1	0	0	1	0	0	1
0	1	0	1	1	0	0
0	1	0	1	0	1	0

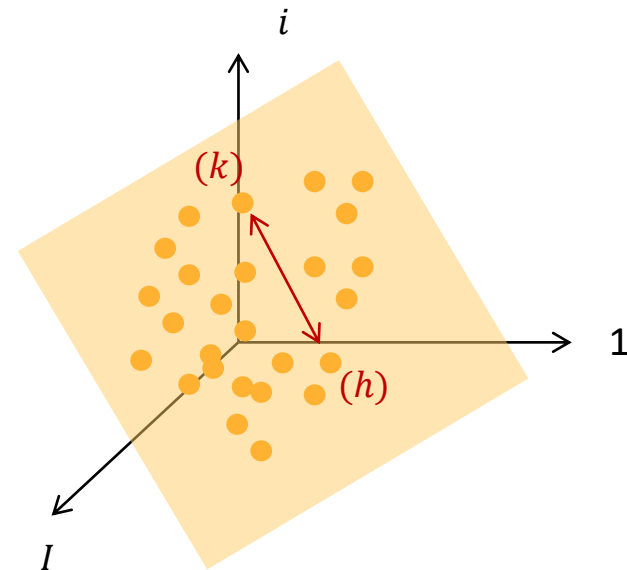
$$d^2_{\chi^2}(i, l) = \frac{1}{J} \sum_{k=1}^K \frac{I}{I_k} (x_{ik} - x_{lk})^2$$

- Deux individus sont d'autant plus proches qu'ils possèdent un grand nombre de **modalités en commun**
- Une modalité rare éloigne son possesseur de tous les autres

Ajustement du nuage des profils - modalités

	1	...	k	...	K
1			$\frac{x_{1k}}{I_k}$		
\vdots					
i			$\frac{x_{ik}}{I_k}$		
\vdots					
I			$\frac{x_{Ik}}{I_k}$		

Σ	1		1		1
----------	---	--	---	--	---



Poids d'une modalité (k) : $m_k = \frac{I_k}{IJ}$

Distance entre deux modalités

Tactile		Focale min		Focale Max		
oui	non	< 25	> 27	< 80	80 - 100	> 140
1	0	0	1	0	0	1
0	1	0	1	1	0	0
0	1	0	1	0	1	0
1	0	0	1	0	1	0
1	0	1	0	1	0	0

$$d^2_{\chi^2}(k, h) = \frac{I}{I_k I_h} (I_k + I_h - 2I_{kh})$$

- Deux modalités sont d'autant plus distantes qu'elles ont été **choisies simultanément** par un petit nombre d'individus
- La distance est d'autant plus grande que le nombre de personnes ayant choisi les modalités h et k est faible

3.5 – Axes factoriels et inerties

L'ACM permet d'extraire **un nombre d'axes** égal à

$$(K - J)$$

Inerties

- L'inertie **totale** en ACM est égale à
- L'inertie d'une **variable** (j) est définie par
- L'inertie d'une **modalité** (k) est calculée par

$$I(N_K) = I(N_I) = \frac{K}{J} - 1$$

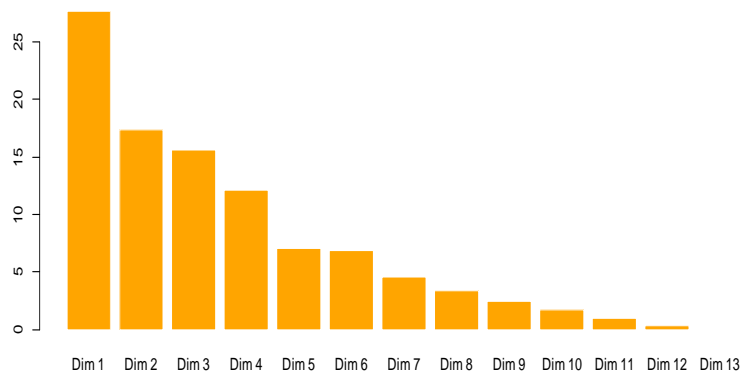
$$I(j) = \frac{1}{J} (K_j - 1)$$

$$I(k) = \frac{1}{J} \left(\frac{I - I_k}{I} \right)$$

```
> res$eig
```

	eigenvalue	percentage of variance	cumulative percentage of variance
dim 1	4.496801e-01	2.767262e+01	27.67262
dim 2	2.825727e-01	1.738909e+01	45.06171
dim 3	2.536131e-01	1.560696e+01	60.66867
dim 4	1.964699e-01	1.209045e+01	72.75912
dim 5	1.140490e-01	7.018398e+00	79.77752
dim 6	1.115975e-01	6.867536e+00	86.64506
dim 7	7.371341e-02	4.536210e+00	91.18127
dim 8	5.535890e-02	3.406701e+00	94.58797
dim 9	3.948641e-02	2.429933e+00	97.01790
dim 10	2.716035e-02	1.671406e+00	98.68931
dim 11	1.615603e-02	9.942171e-01	99.68352
dim 12	5.142742e-03	3.164764e-01	100.00000
dim 13	2.144711e-32	1.319822e-30	100.00000

Inertie (en %) des axes factoriels

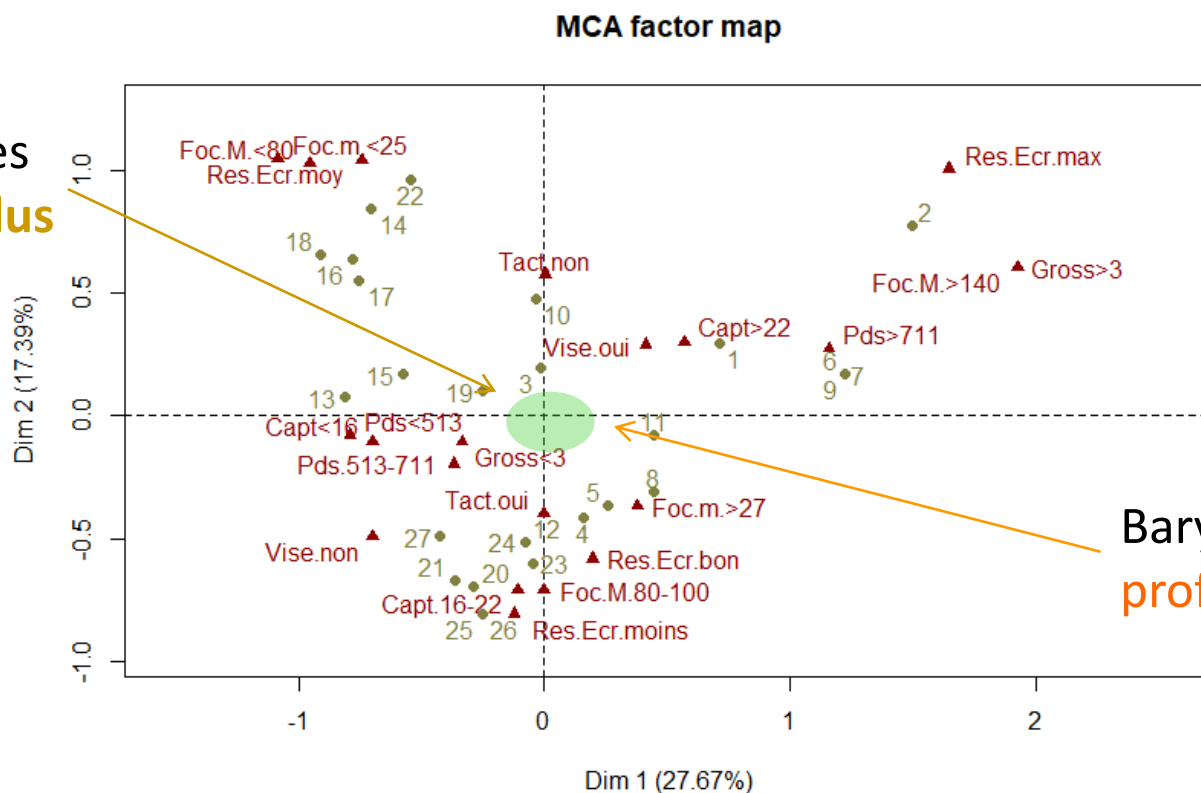


- Inertie moyenne : $\frac{1}{J}$
- Faibles taux d'inertie en général en ACM
- Possibilité de « corriger » les taux d'inertie (*correction de Benzecri – Greenacre*)

3.6 – Interprétation d'une ACM

Eloignement ou proximité avec l'origine des axes

Barycentre des
profils **individus**



Barycentre des
profils **modalités**

Définition et interprétation des deux barycentres

	1	k	K		
1				J	$\frac{1}{I}$
i		x_{ik}		J	\vdots
					$\frac{1}{I}$
I				J	\vdots
					$\frac{1}{I}$
	I_1	I_k	I_K	IJ	1
	$\frac{I_1}{IJ}$...	$\frac{I_k}{IJ}$...	$\frac{I_K}{IJ}$
				1	

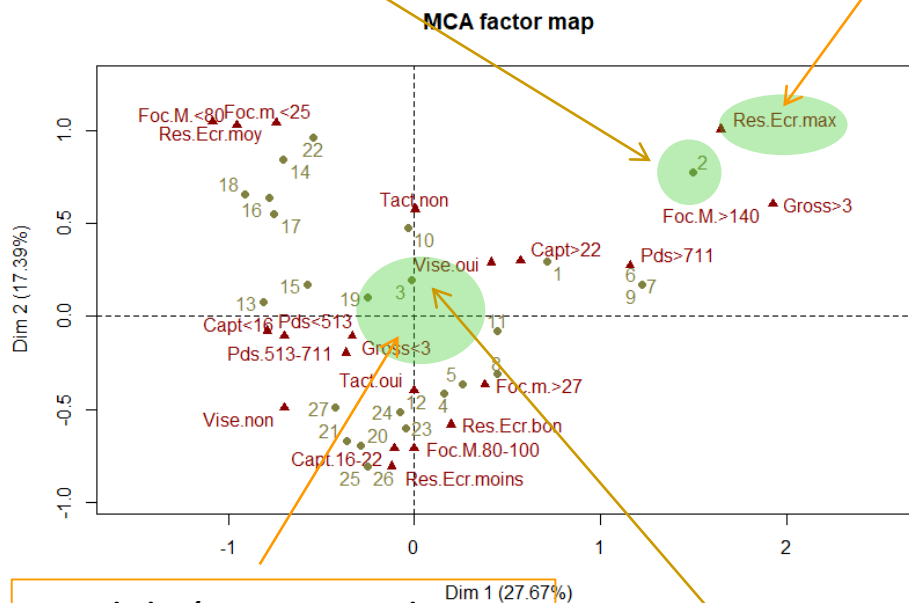
Centre de gravité des **modalités** G_K

Centre de gravité des **individus** G_I

Eloignement par rapport au centre de gravité

Individu prenant des modalités **rares**

Modalité **peu** choisie



$$d^2_{\chi^2}(k, G_K) = \frac{I}{I_k} - 1$$

Une modalité choisie par un nombre faible d'individus est éloignée du centre de gravité

$$d^2_{\chi^2}(i, G_I) = \frac{I}{J} \sum_k \left(\frac{x_{ik}}{I_k} \right) - 1$$

Un individu est proche de l'origine s'il choisit des modalités fréquentes

Individu prenant des modalités **souvent** choisie

Relations quasi barycentriques ou de transition

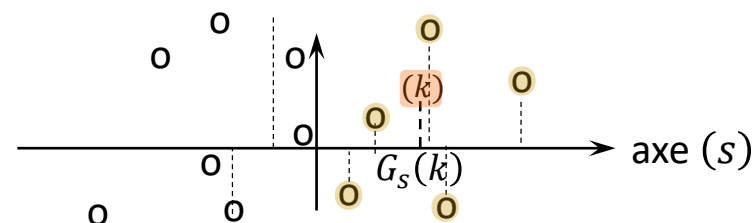
Proximité entre lignes (individus) et colonnes (modalités)

$F_s(i)$: coordonnée de l'individu (i) sur l'axe (s)

$G_s(k)$: coordonnée de la modalité (k) sur l'axe (s)

$$G_s(k) = \frac{1}{\sqrt{\lambda_s}} \sum_{i=1}^I \left(\frac{x_{ik}}{I_k} \right) F_s(i)$$

« Une modalité est au (quasi-) centre de gravité des individus qui la possèdent »



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

« Un individu est au (quasi-) centre de gravité des modalités qu'il possède »

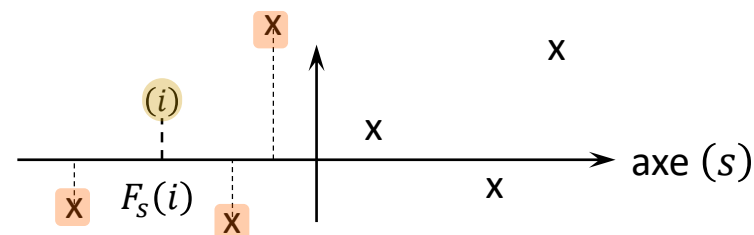
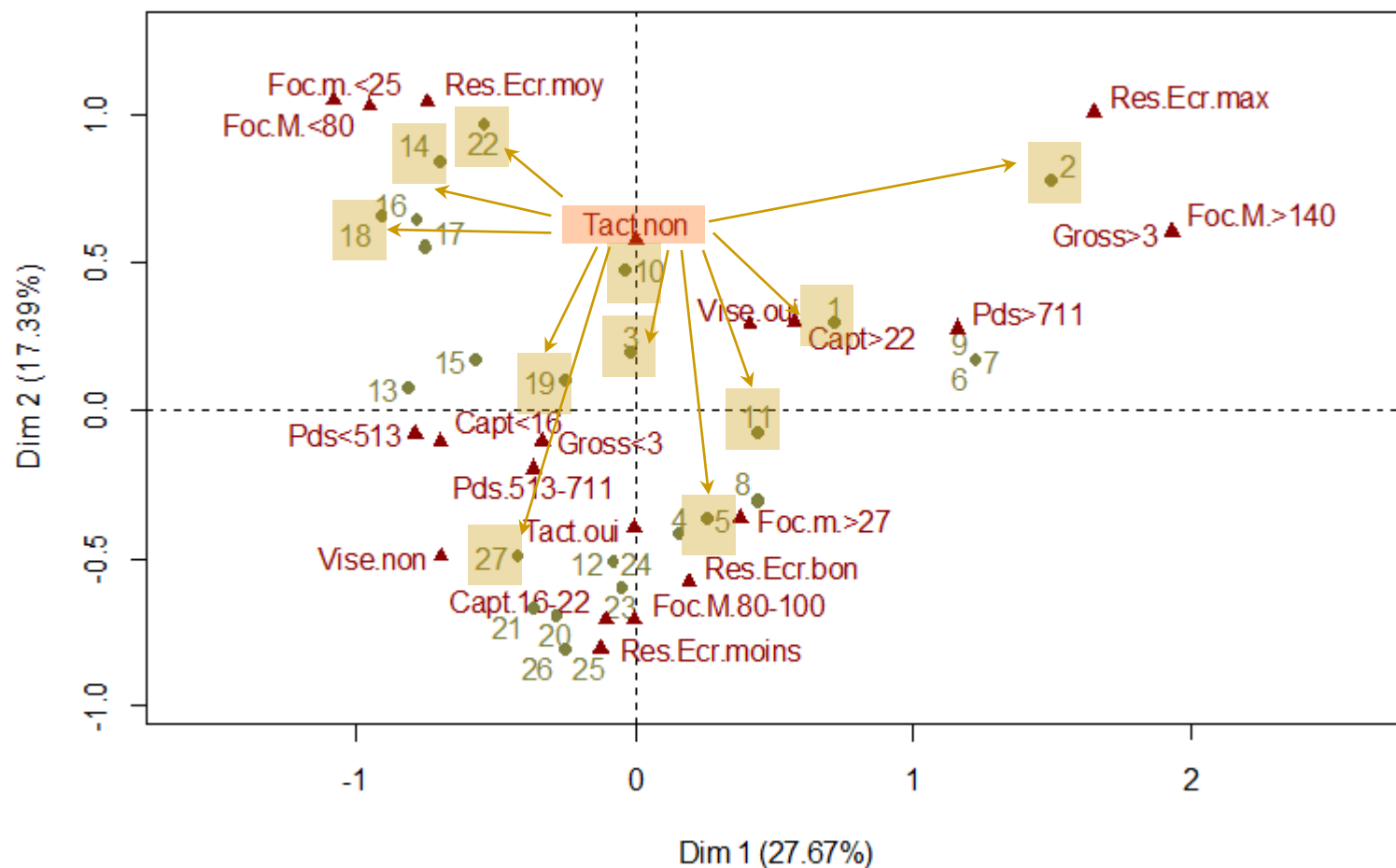
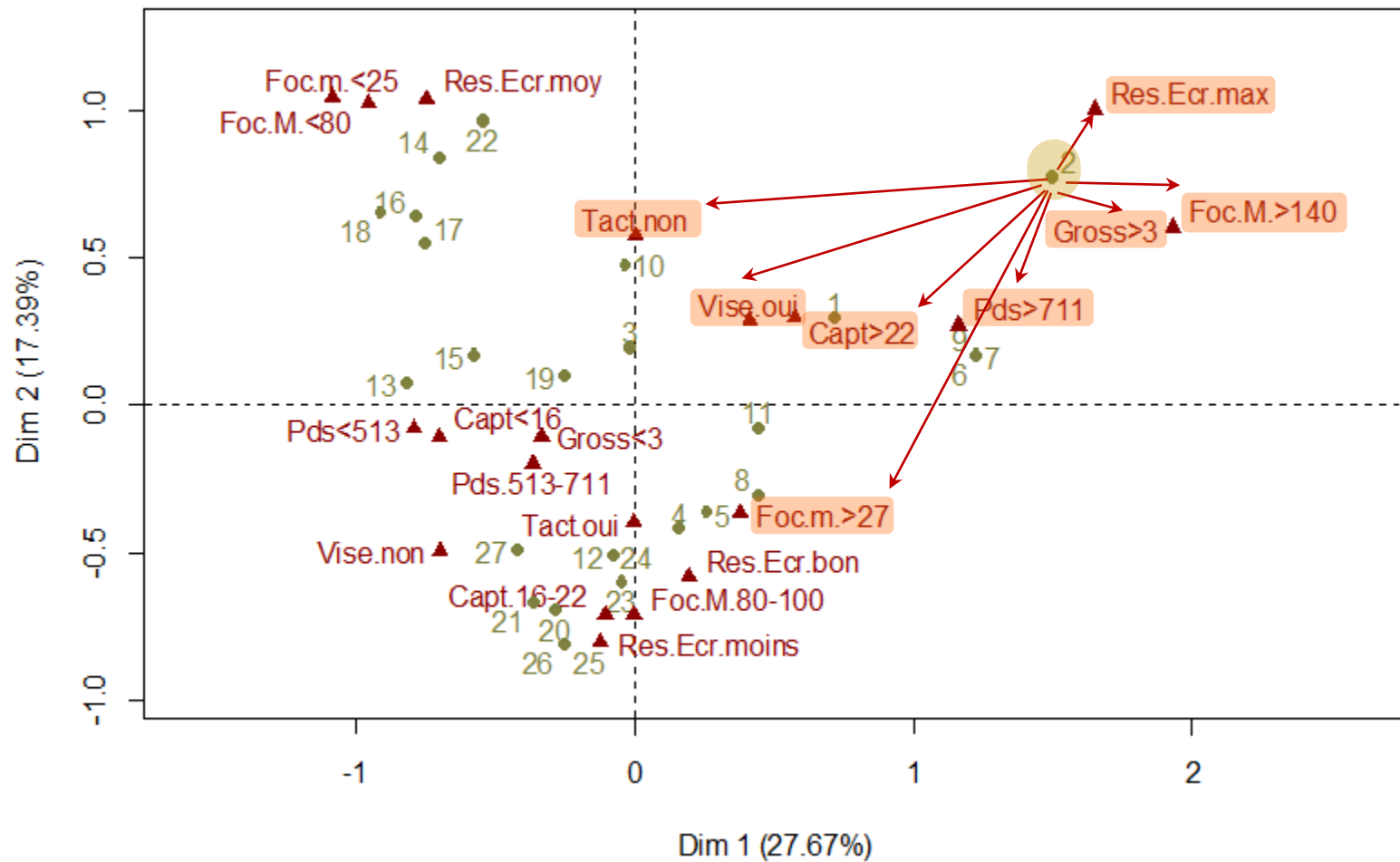


Illustration - Plan (1,2) des individus et des modalités actives

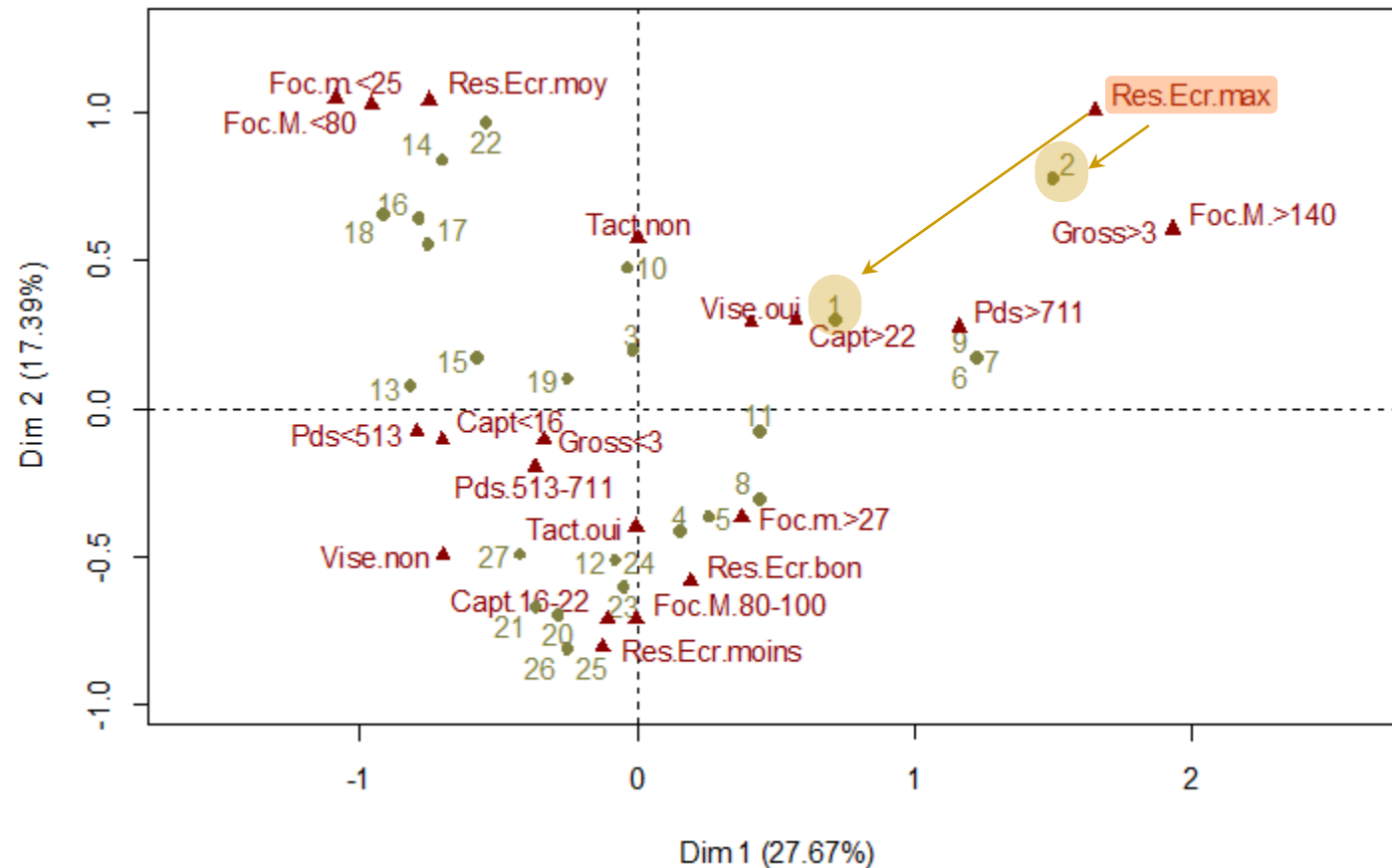




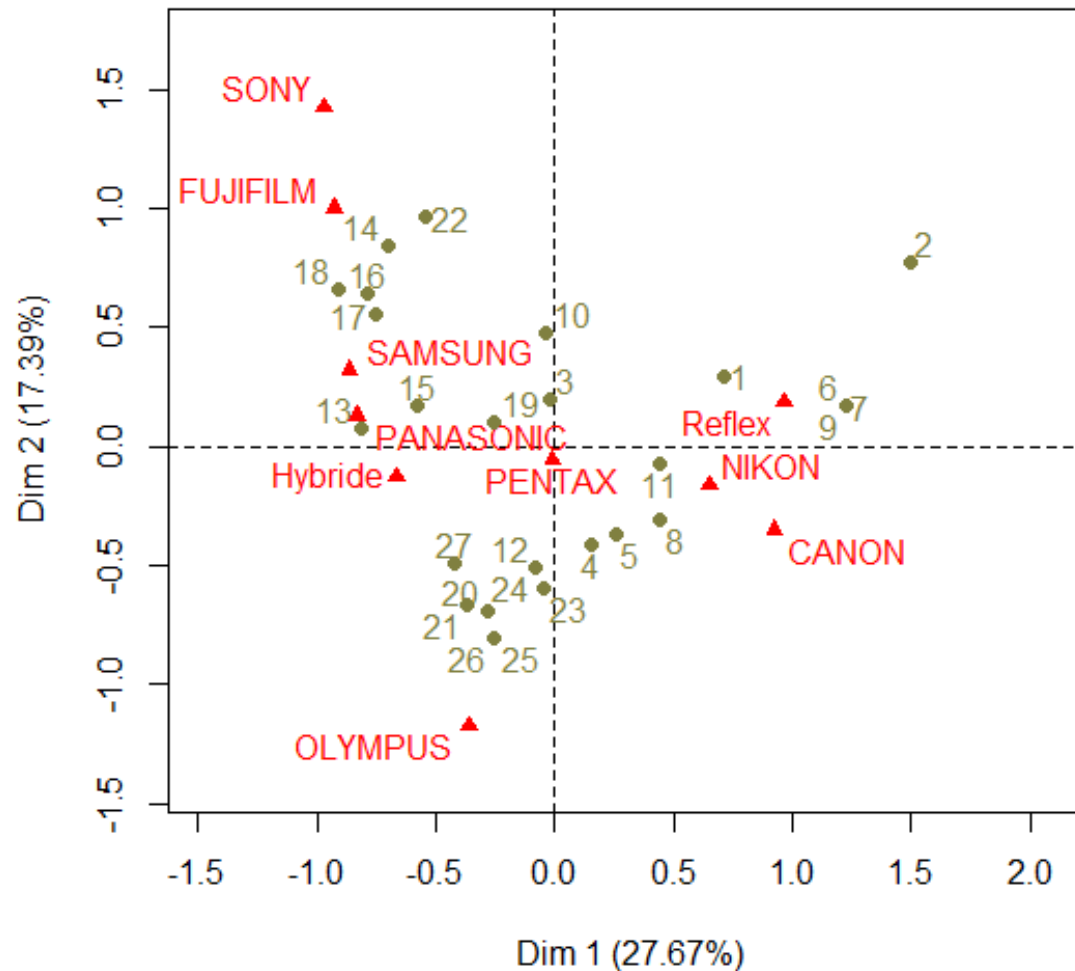
Un quasi barycentre !

$$G_s(k) = \frac{1}{\sqrt{\lambda_s}} \sum_{i=1}^I \left(\frac{x_{ik}}{I_k} \right) F_s(i)$$

Facteur de dilatation



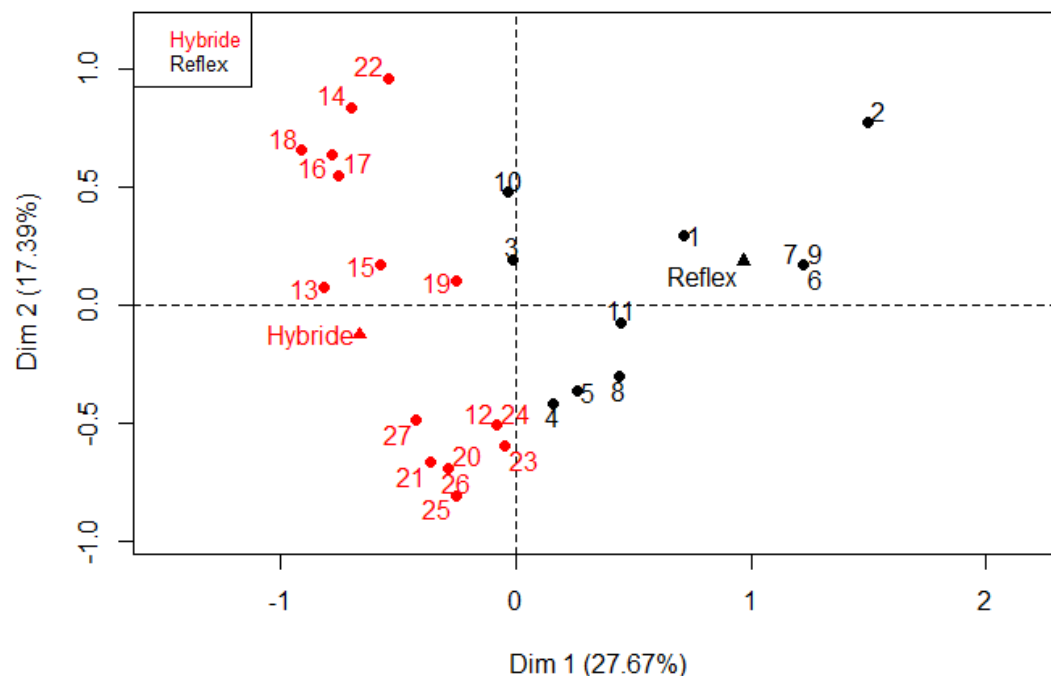
Plan (1,2) des individus et modalités illustratives



Plan (1,2) des modalités illustratives Avec « habillage » de sous-populations

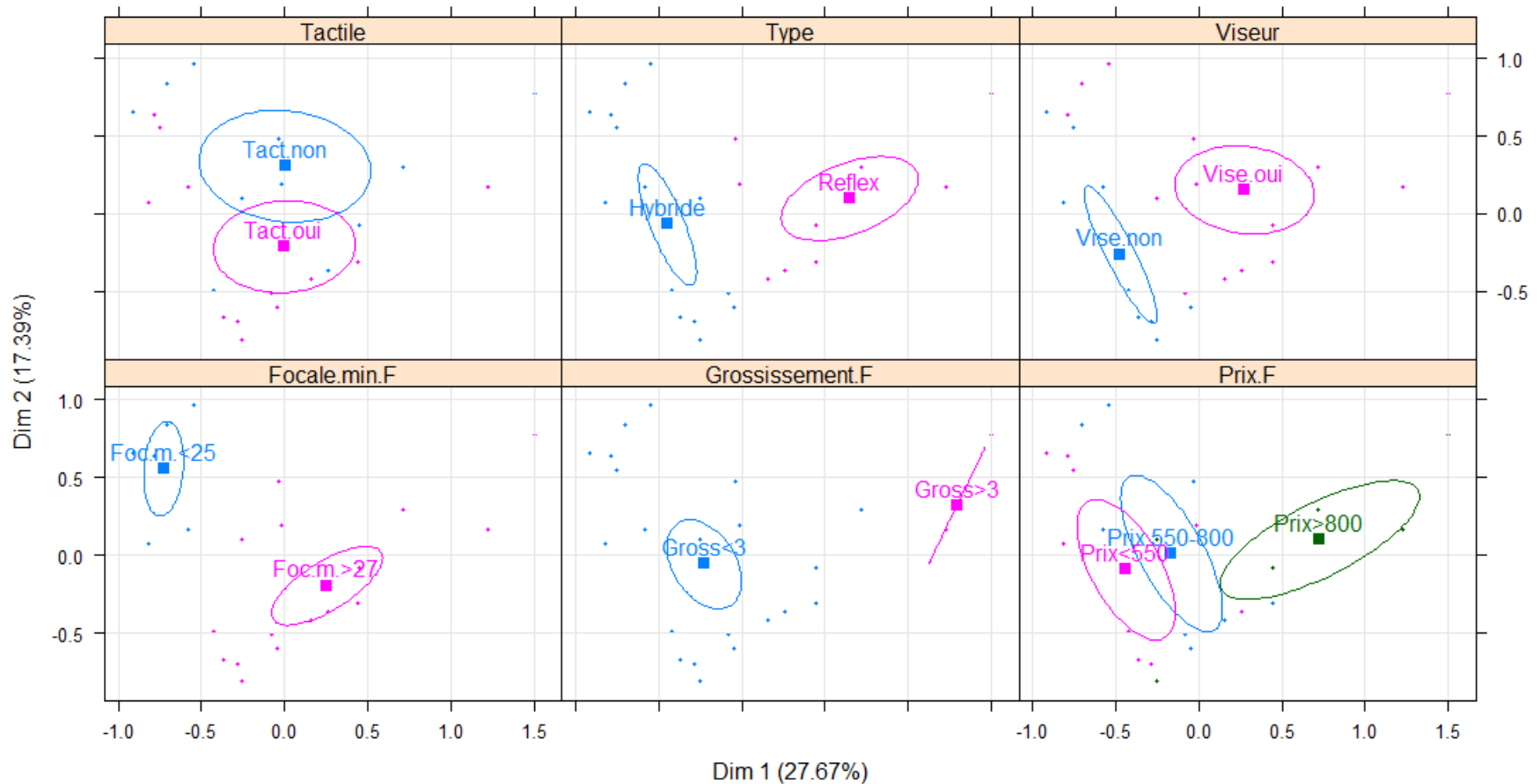
```
photo.MCA<-photo[, c("Viseur",
"Tacile",
"Focale.min.F", "Focale.Max.F",
"Grossissement.F", "Poids.F",
"Ecran.Reso.F", "Capteur.Reso.F"
, "Prix", "NoteGlobale",
"NoteFlash", "NoteDélai2prises",
"Marque", "Type")]
```

```
res<-MCA(photo.MCA,
ncp=5, ind.sup=NULL,
quanti.sup=9:12,
quali.sup=13:14, graph = FALSE)
```



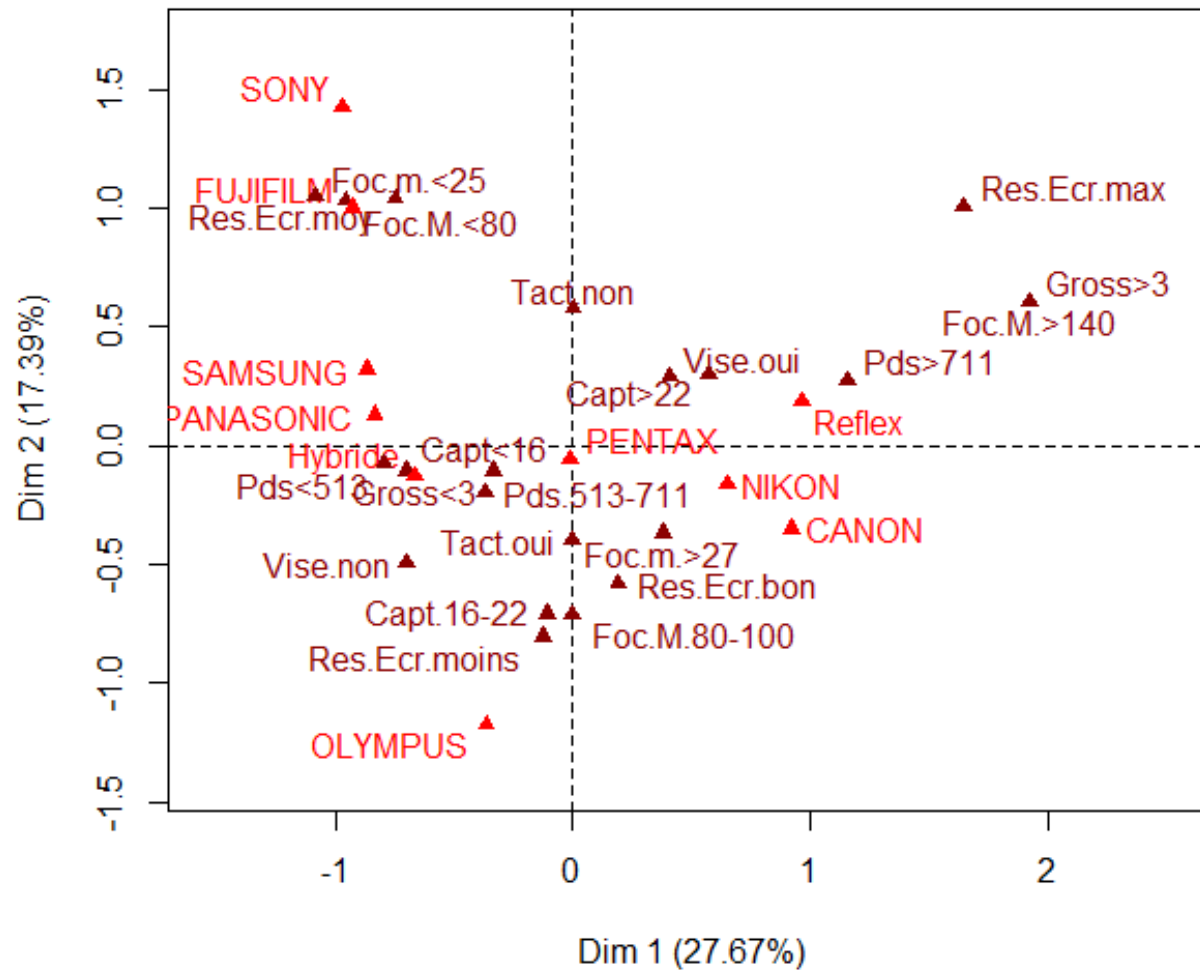
```
plot.MCA(res, axes=c(1, 2), new.plot=TRUE, col.ind="#808040", col.ind.sup="blue",
col.var="darkred", col.quali.sup="red", label=c("ind", "ind.sup", "quali.sup"),
invisible=c("var"), habillage=14, palette=palette(c("red", "black")))
```

Plan (1,2) des modalités – Avec ellipses de confiance



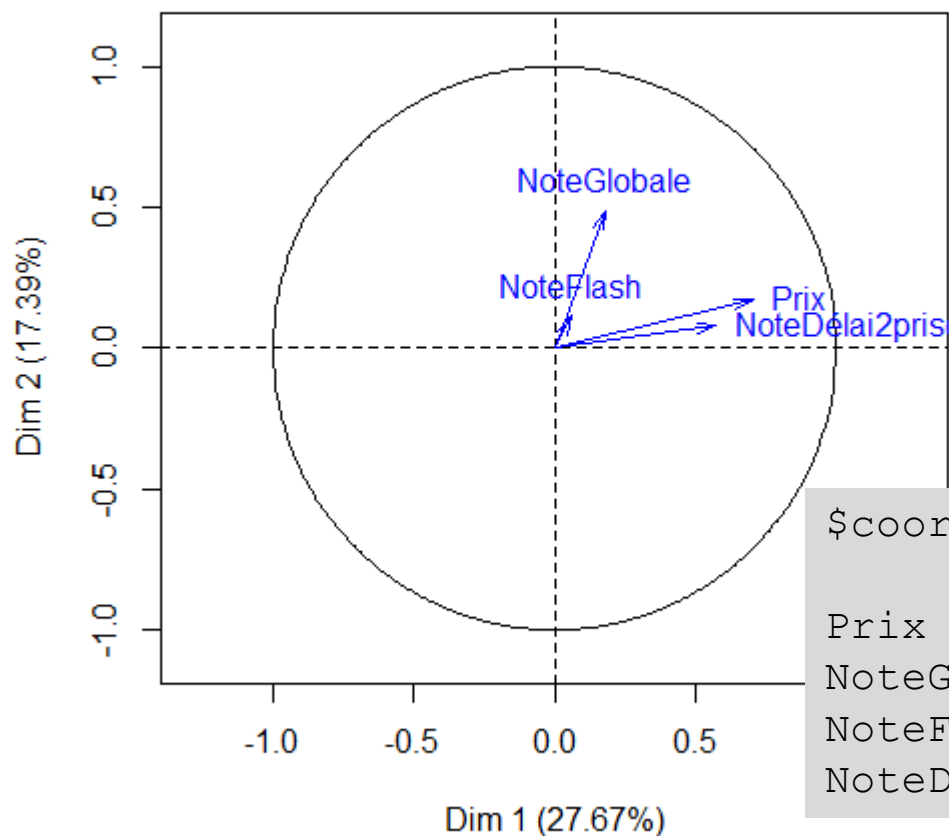
```
plotellipses(res, keepvar=c("Type", "Marque", "Prix.F"), level=0.95)
```

Plan (1,2) des modalités actives et illustratives



Plan (1,2) des variables illustratives quantitatives

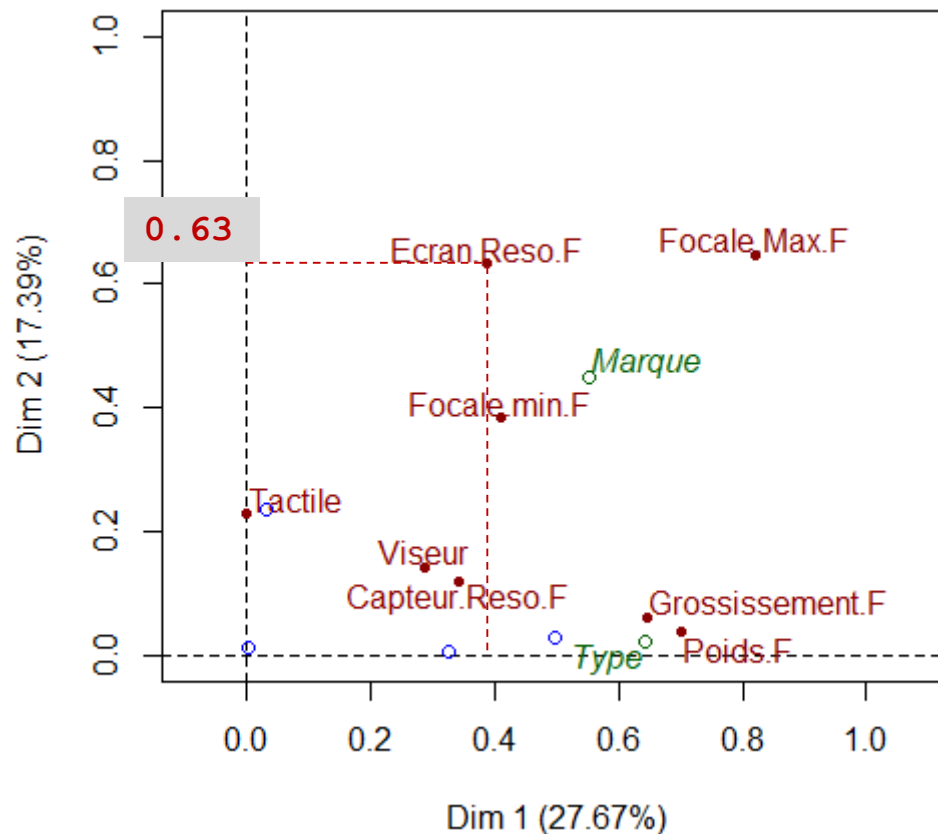
Supplementary variables on the MCA factor map



\$coord	Dim 1	Dim 2
Prix	0.70443725	0.17297875
NoteGlobale	0.17754562	0.48724429
NoteFlash	0.05805497	0.11322736
NoteDélai2prises	0.57135308	0.08132291

La représentation graphique des variables

Intérêt : identifier les variables les plus pertinentes pour interpréter une dimension



La coordonnée d'une variable (j) sur l'axe (s) traduit **l'intensité de la liaison** entre la variable et l'axe

\$eta2	Dim 1	Dim 2
Visiteur	2.878250e-01	0.14226609
Tactile	7.514901e-06	0.22874067
Focale.min.F	4.115739e-01	0.38443397
Focale.Max.F	8.209336e-01	0.64811495
Grossissement.F	6.461547e-01	0.06346555
Poids.F	7.012347e-01	0.03993716
Ecran.Reso.F	3.884393e-01	0.63 232453
Capteur.Reso.F	3.412720e-01	0.12129880

\$eta2	Dim 1	Dim 2
Marque	0.5539402	0.45065970
Type	0.6433143	0.02392529

Le rapport de corrélation comme mesure de liaison

L'intensité de la liaison est déterminée par le **rapport de corrélation** entre la variable (j) et la composante principale C_s associée à l'axe (s)

$$\eta^2(j, C_s)$$

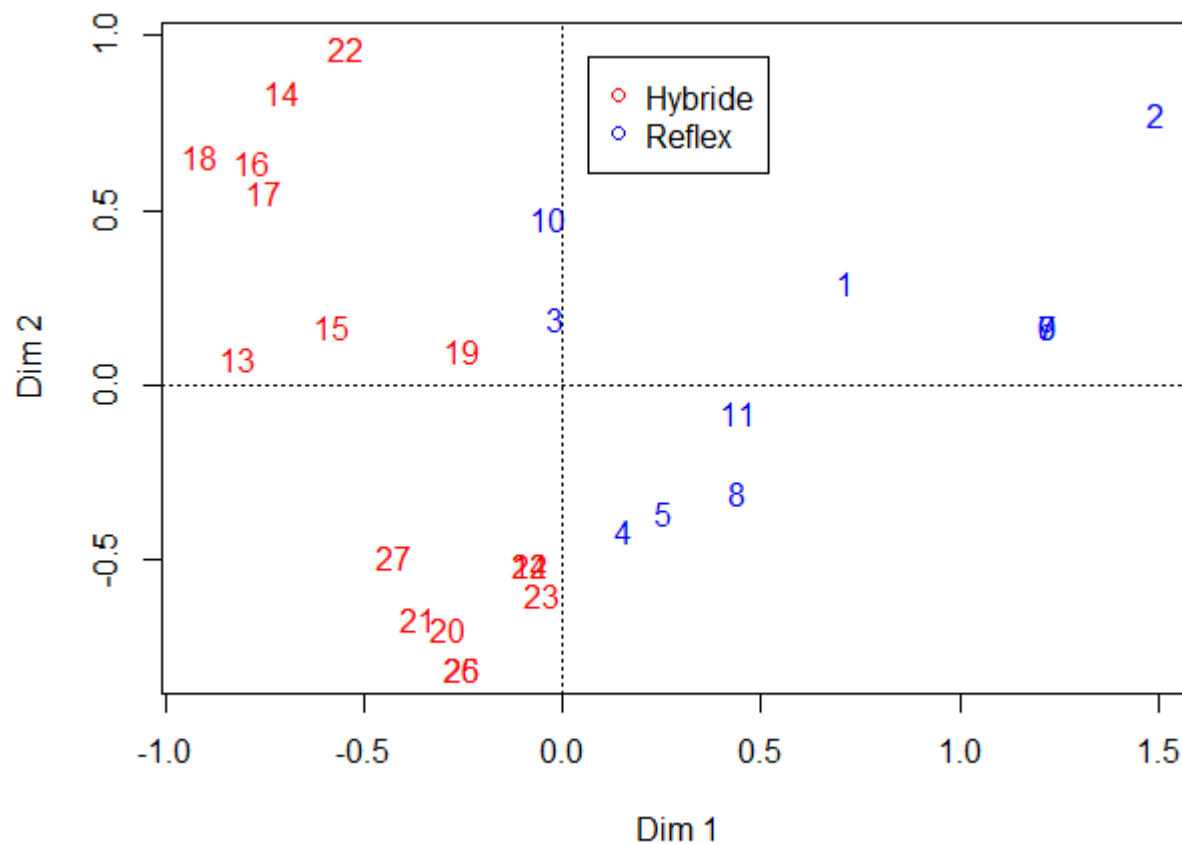
$$\eta^2 = R^2 \in [0; 1]$$

Rapport de corrélation

- Mesure de liaison entre une variable quantitative et une variable qualitative
- Il est calculé par exemple de façon systématique en analyse de la variance à un facteur

Illustration sur un exemple

Comment la variable *Type* est-elle liée à l'axe 1 ?



Coord.Axe1 Type

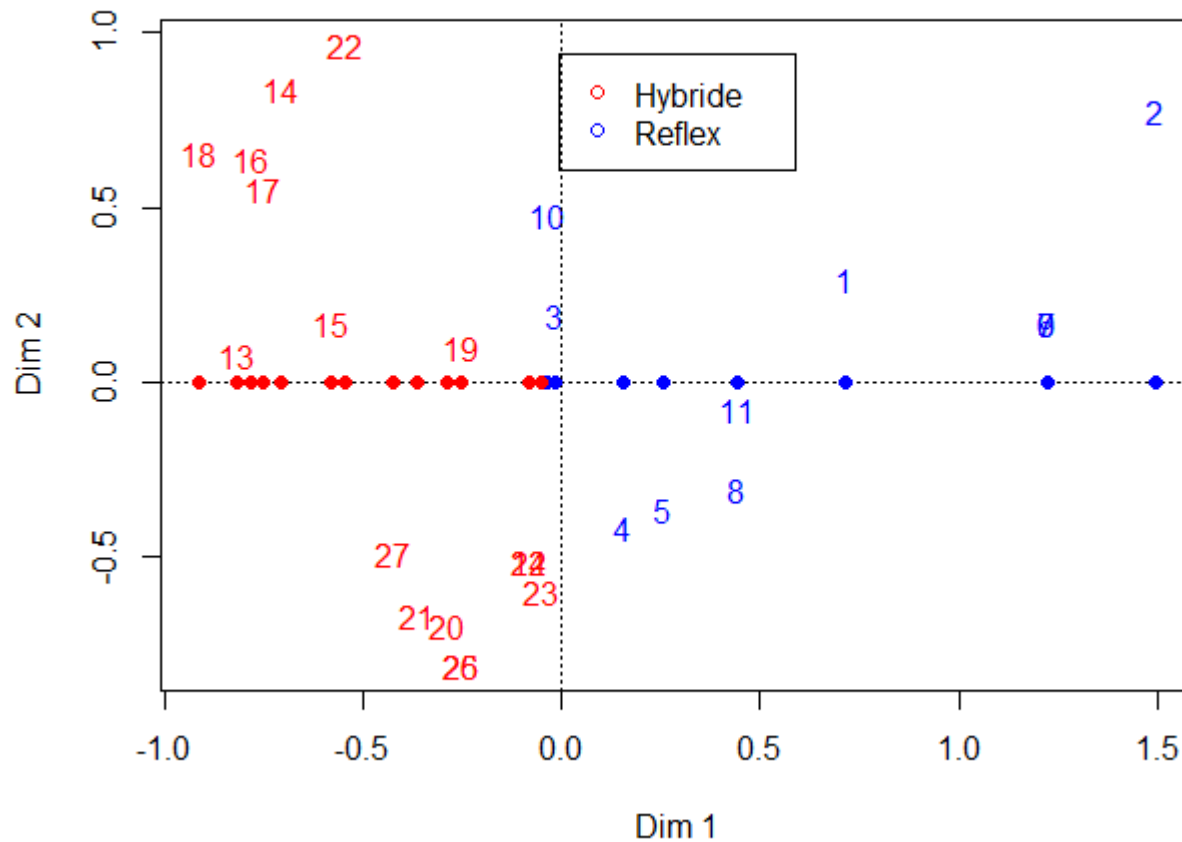
```

1  "0.7148"  "Reflex"
2  "1.4967"  "Reflex"
3  "-0.0162" "Reflex"
4  "0.158"   "Reflex"
5  "0.2578"  "Reflex"
.
.
.
22 "-0.5453" "Hybride"
23 "-0.0491" "Hybride"
24 "-0.0793" "Hybride"
25 "-0.2545" "Hybride"
26 "-0.2545" "Hybride"
27 "-0.4235" "Hybride"

```

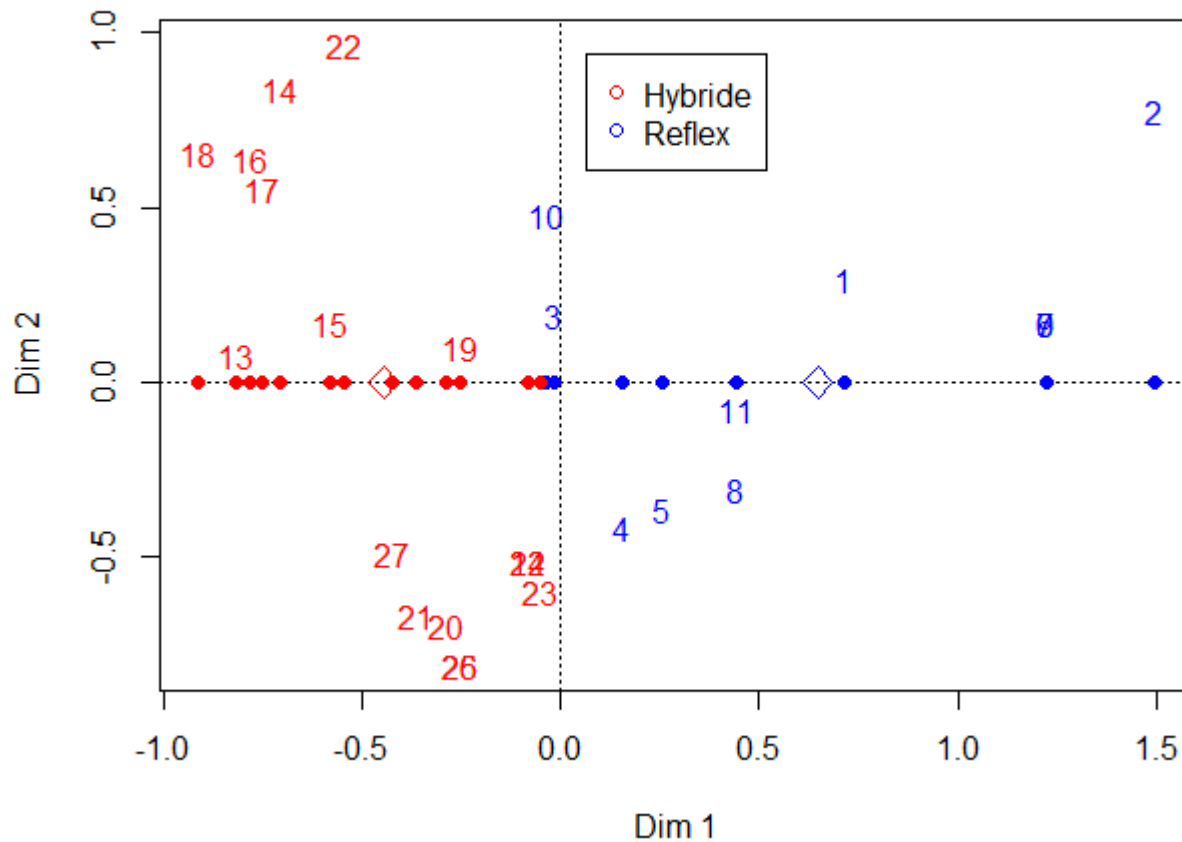
On identifie les deux sous-populations liées à la variable *Type* : *Reflex*, *Hybride*

Liaison entre *Type* et Dim 1



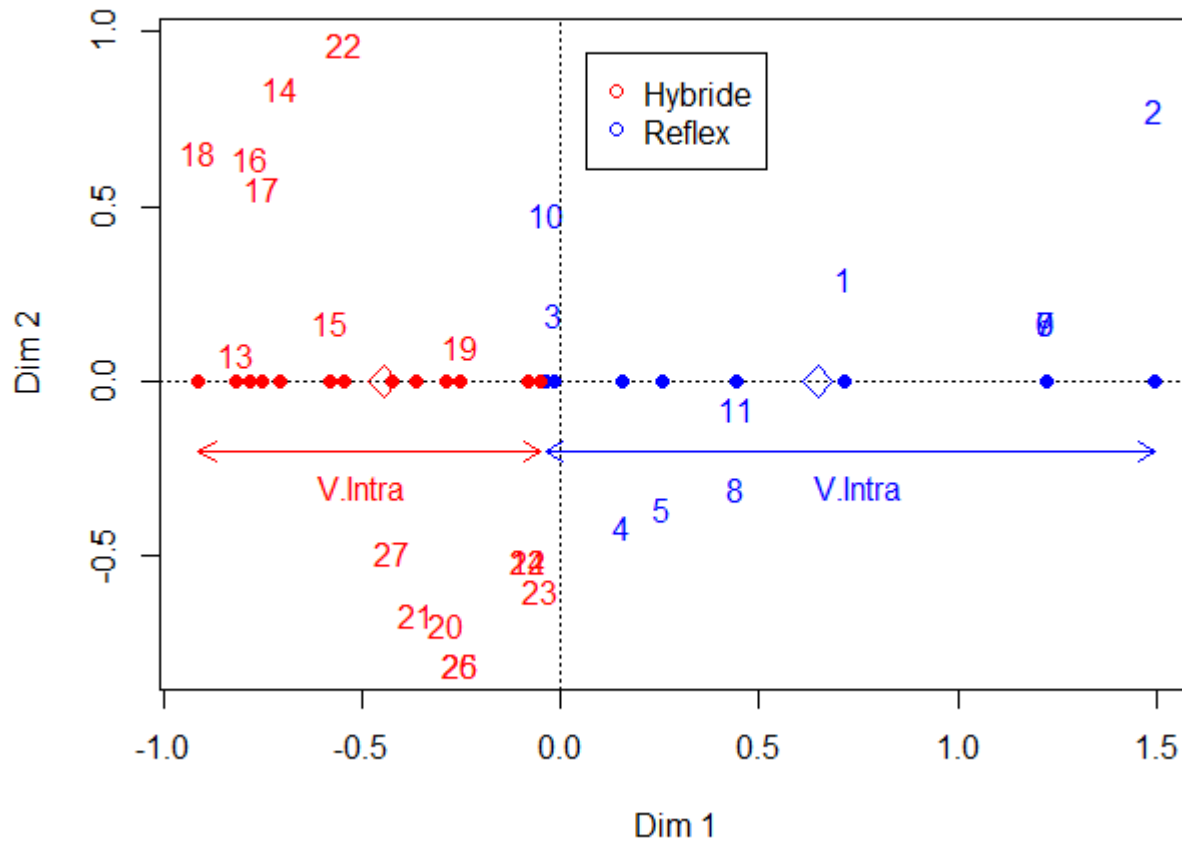
On détermine les **coordonnées** des individus le long de l'axe 1

Liaison entre *Type* et Dim 1



On détermine les **points moyens** des deux sous-populations le long de l'axe

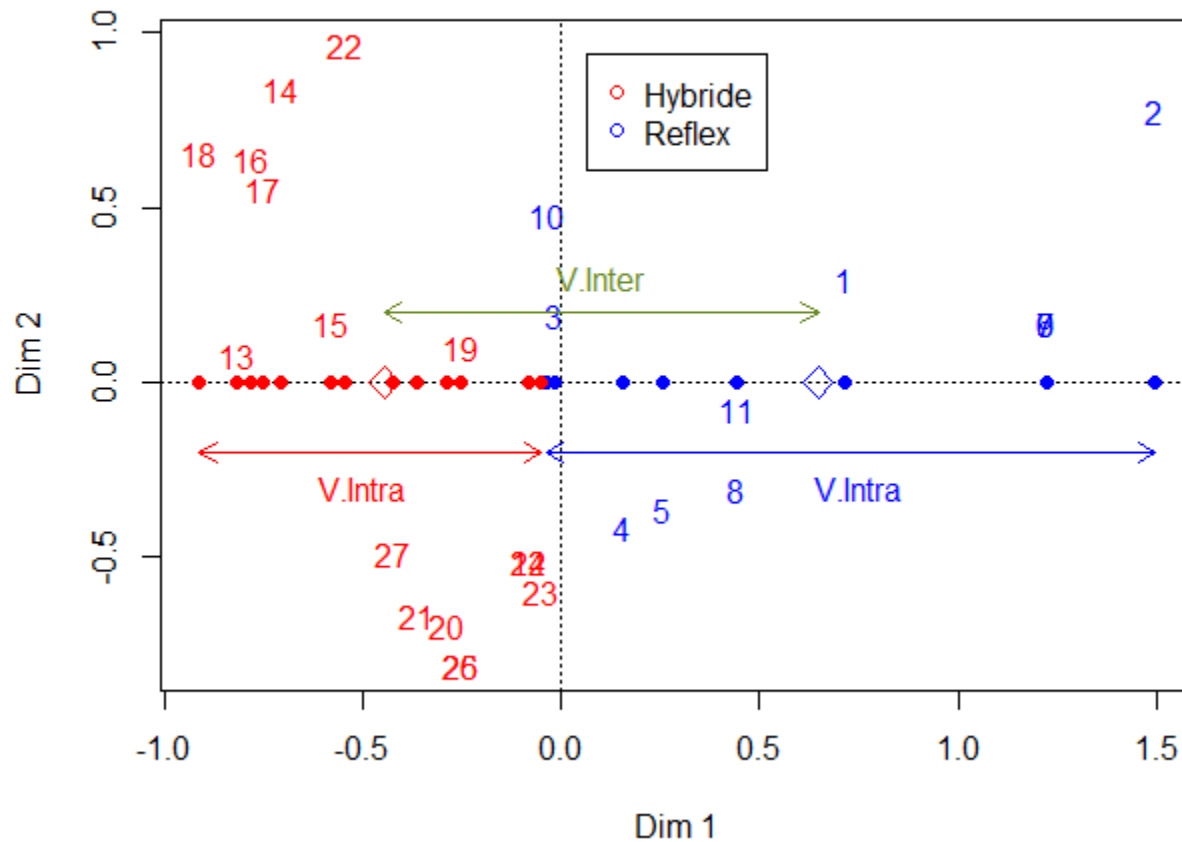
Liaison entre *Type* et Dim 1



On calcule la
variabilité INTRA
sous-population

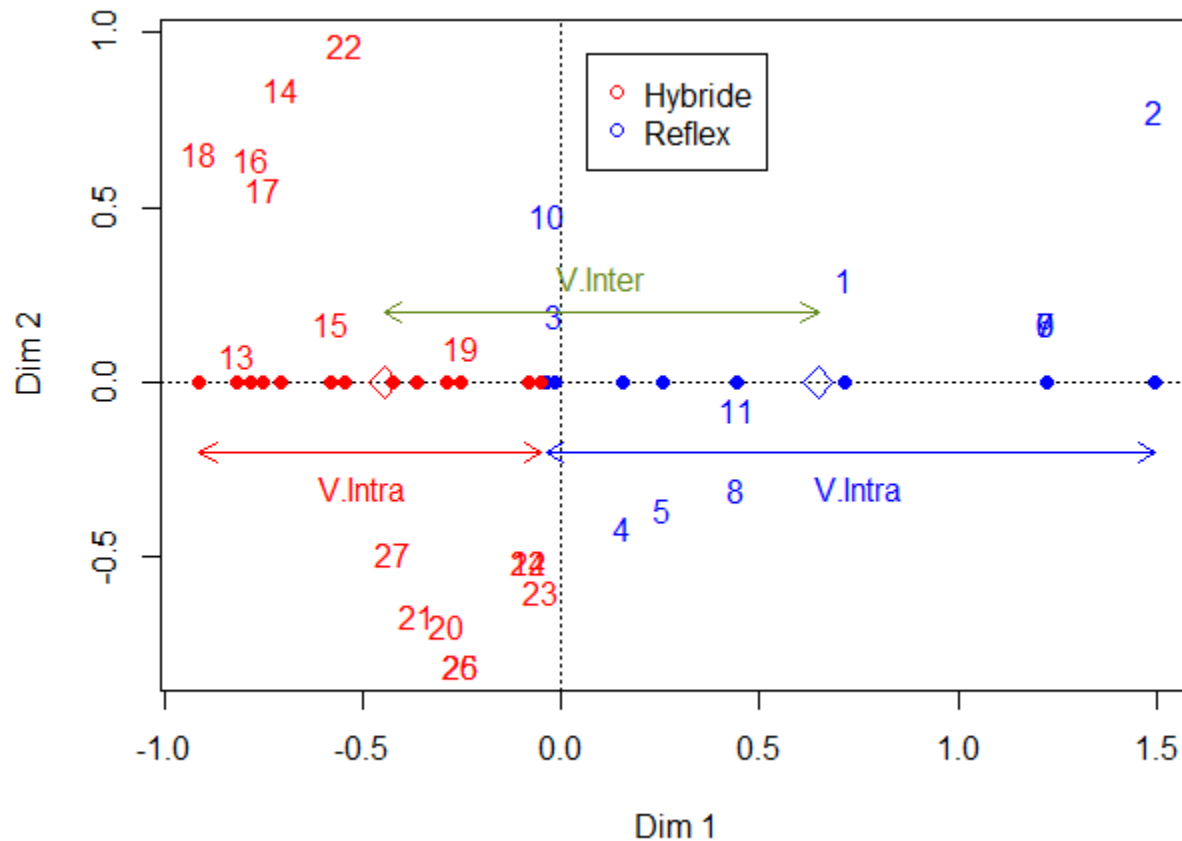
Mesure globale de
l'hétérogénéité au
sein des sous-
populations

Liaison entre *Type* et Dim 1



On calcule la **variabilité INTER** sous-populations

Mesure globale de séparation des sous-populations



La part de variabilité INTER dans la variabilité totale détermine le **rapport de corrélation**

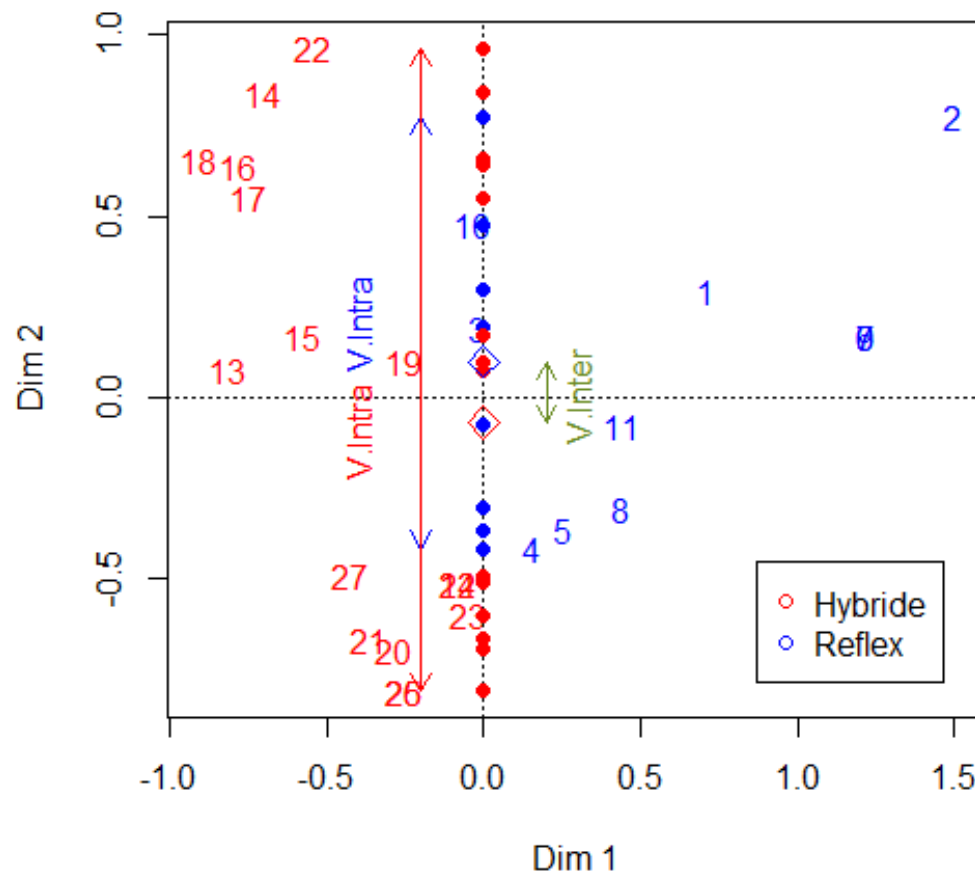
$$\eta^2 = \frac{V.Inter}{V.Totale}$$

$$\eta^2 = \frac{V.Inter}{V.Inter + V.Intra}$$

\$eta2		
	Dim 1	Dim 2
Marque	0.5539402	0.45065970
Type	0.6433143	0.02392529

La liaison est d'autant plus intense que la variable forme des sous-populations **homogènes** et **bien séparées** le long de l'axe

Liaison entre *Type* et Dim 2



```
$eta2
      Dim 1      Dim 2
Marque 0.5539402 0.45065970
Type   0.6433143 0.02392529
```

Les deux sous-nuages sont quasi superposés le long de l'axe deux (leurs points moyens sont très proches)

La variable *Type* est très peu liée à la dimension 2

Propriétés

$$\eta^2(j, C_s) = J \times \sum_{k=1}^{K_j} I_s(k)$$

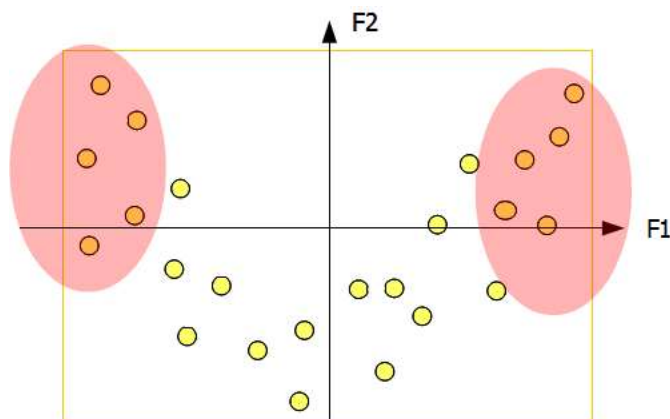
Une variable dont les modalités apportent beaucoup d'inertie à un axe est très liée à cet axe

$$I_s = \frac{1}{J} \sum_{j=1}^J \eta^2(j, C_s)$$

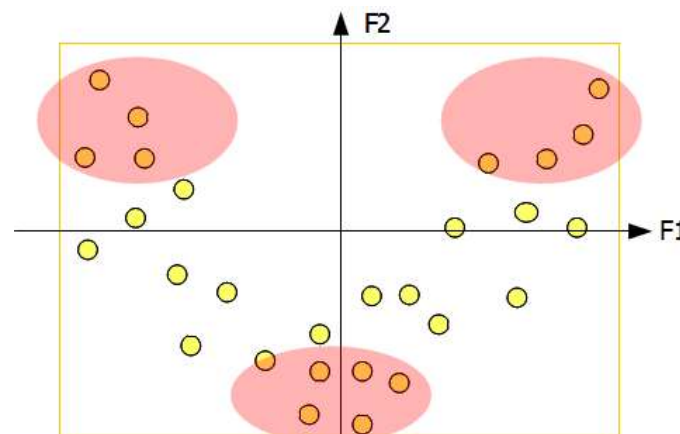
L'inertie d'un axe est égale à la moyenne des rapports de corrélations des variables avec l'axe

L'effet Gutman

Facteur 1 : facteur d'échelle (« effet taille »)

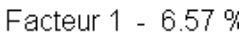


Facteur 2 : opposition extrêmes / moyens



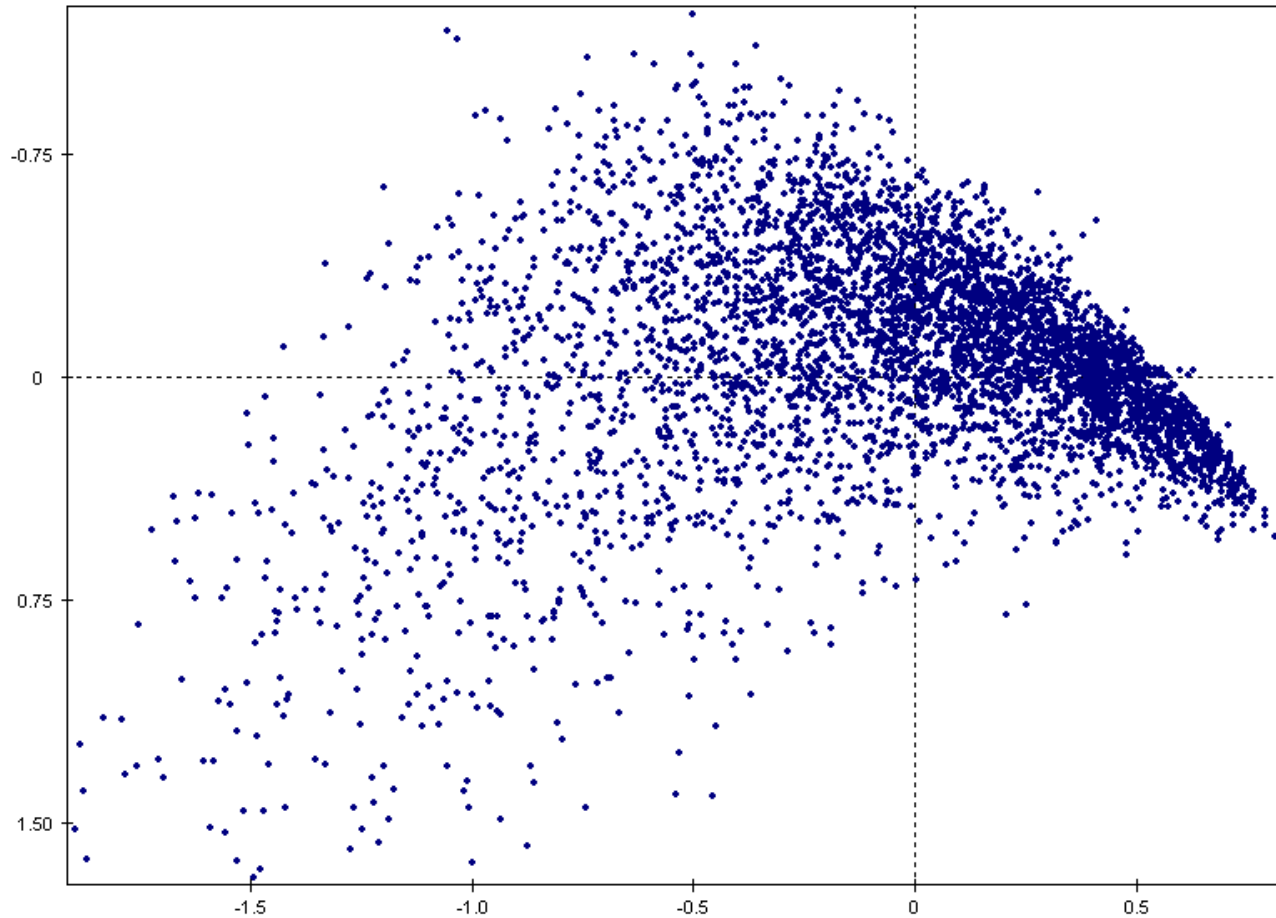
Apparaît très souvent lorsque les modalités de la variable possèdent un ordre naturel ou lorsqu'elles proviennent d'une variable quantitative mise en classes

Plan des modalités



Plan des individus

Facteur 4 - 3.41 %



Facteur 1 - 6.57 %

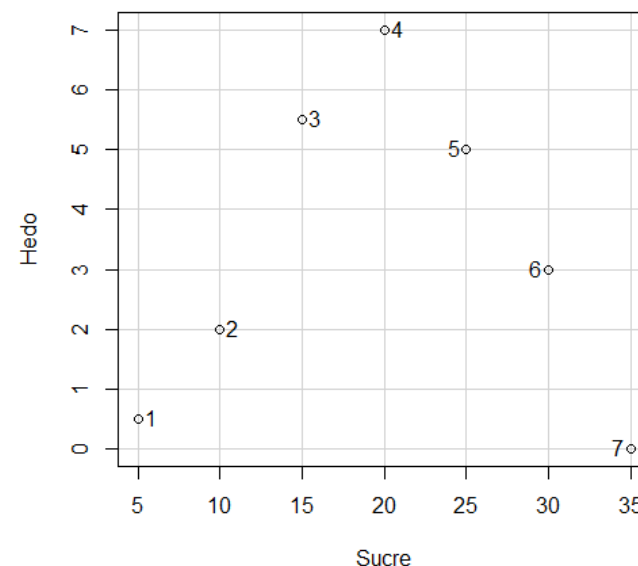
ACM et étude de liaisons non linéaires

Lorsque les données sont **quantitatives**, il peut s'avérer intéressant de les **découper en classes** afin de mettre en évidence d'éventuelles liaisons non linéaires entre variables

Un exemple (très...) simple !

Dégustation de 7 boissons plus ou moins sucrées par des consommateurs qui attribuent une note hédonique

	Sucre	Hedo
1	5	0.5
2	10	2.0
3	15	5.5
4	20	7.0
5	25	5.0
6	30	3.0
7	35	0.0



Une petite ACP...

La note d'appréciation est non corrélée
(linéairement) à la quantité de sucre...
Mais non indépendante !

Recodage en classes puis ACM...

R Sucre.Hedo					
	Sucre	Hedo	Hedo.F	Sucre.F	
1	5	0.5	beurk	pas sucré	
2	10	2.0	moy	peu sucré	
3	15	5.5	moy	peu sucré	
4	20	7.0	bon	moy sucré	
5	25	5.0	moy	bcp sucré	
6	30	3.0	moy	bcp sucré	
7	35	0.0	beurk	max sucré	

