

### 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	... lib. 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



## Avant l'ACM : analyse univariée !

- **Étape préliminaire** : analyse variable par variable
- **Objectif** : « supprimer » les modalités rarement choisies

### Tris à plat

#### Marche

Mobilier / Ameublement : 17  
 Moto : 8  
 Renovation : 18  
 Scooter : 5  
 Side-car : 1  
 Voiture : 17

#### Apport

Apport : 35  
 pas\_Apport : 31

#### Impaye

Imp\_0 : 42  
 Imp\_1\_ou\_2 : 8  
 Imp\_3 et + : 16

#### Assurance

AID : 31  
 AID + Chomage : 13  
 Sans Assurance : 12  
 Senior : 10

#### Endettement

End\_1 : 18  
 End\_2 : 15  
 End\_3 : 19  
 End\_4 : 14

#### Famille

Celibataire : 17  
 Divorce : 5  
 Marie : 25  
 Union libre : 13  
 Veuf : 6

#### Enfants

Enf\_0 : 39  
 Enf\_1 : 8  
 Enf\_2 : 11  
 Enf\_3 : 6  
 Enf\_4 : 2

#### Logement

Accedant a la propriete : 6  
 Locataire : 23  
 Loge par employeur : 3  
 Loge par la famille : 6  
 Proprietaire : 28

#### Profession

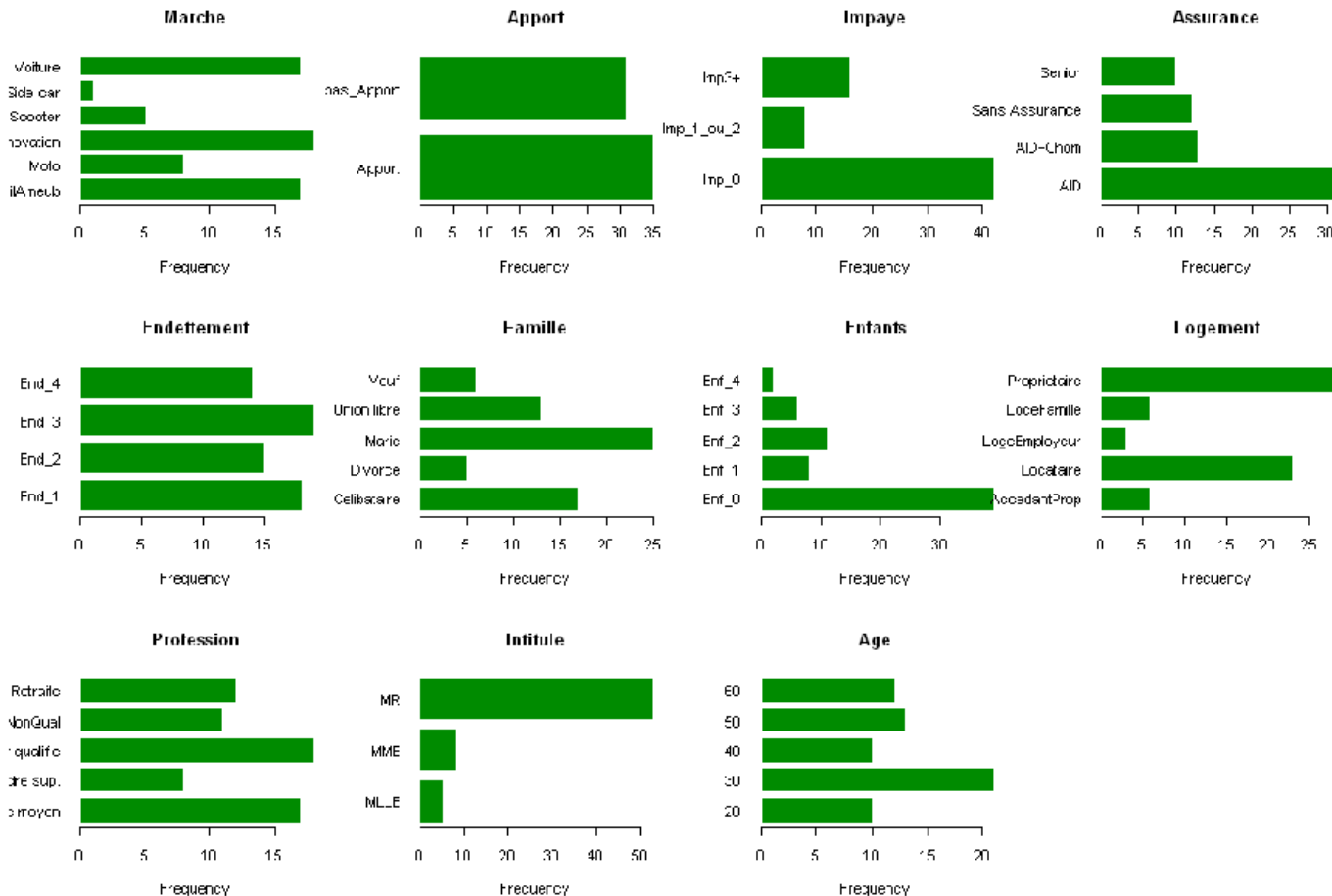
Cadre moyen : 17  
 Cadre sup. : 8  
 Ouvrier non qualifie : 11  
 Ouvrier qualifie : 18  
 Retraite : 12

#### Intitule

MLLE : 5  
 MME : 8  
 MR : 53

#### Age

20 : 10  
 30 : 21  
 40 : 10  
 50 : 13  
 60 : 12



```

par(mfrow=c(3,4))
for (i in 1:11){
  barplot(table(credit[,i]), xlab="Frequency", main=names(credit)[i],
    horiz=T, las=1, col="green4", border="white")
}

```

## 3.2 – Le tableau de données étudié

### Nature des variables

- Un tableau *individus x variables* **qualitatives**
- Des variables qualitatives ou quantitatives peuvent jouer un rôle illustratif (= supplémentaire)

#### variables actives

#### variables supplémentaires

Ind	Marche	Apport	Impaye	Assurance	Endettement	Famille	Enfants	Logement	Profession	Intitule	Age
1	<b><u>Qualitatives</u></b>					<b>Qualitatives et/ou Quantitatives</b>					
2											
3											
4											
$i$											
⋮	⋮					⋮					

## Le tableau disjonctif complet (TDC)

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

### Tableau des données initiales Codage « condensé »

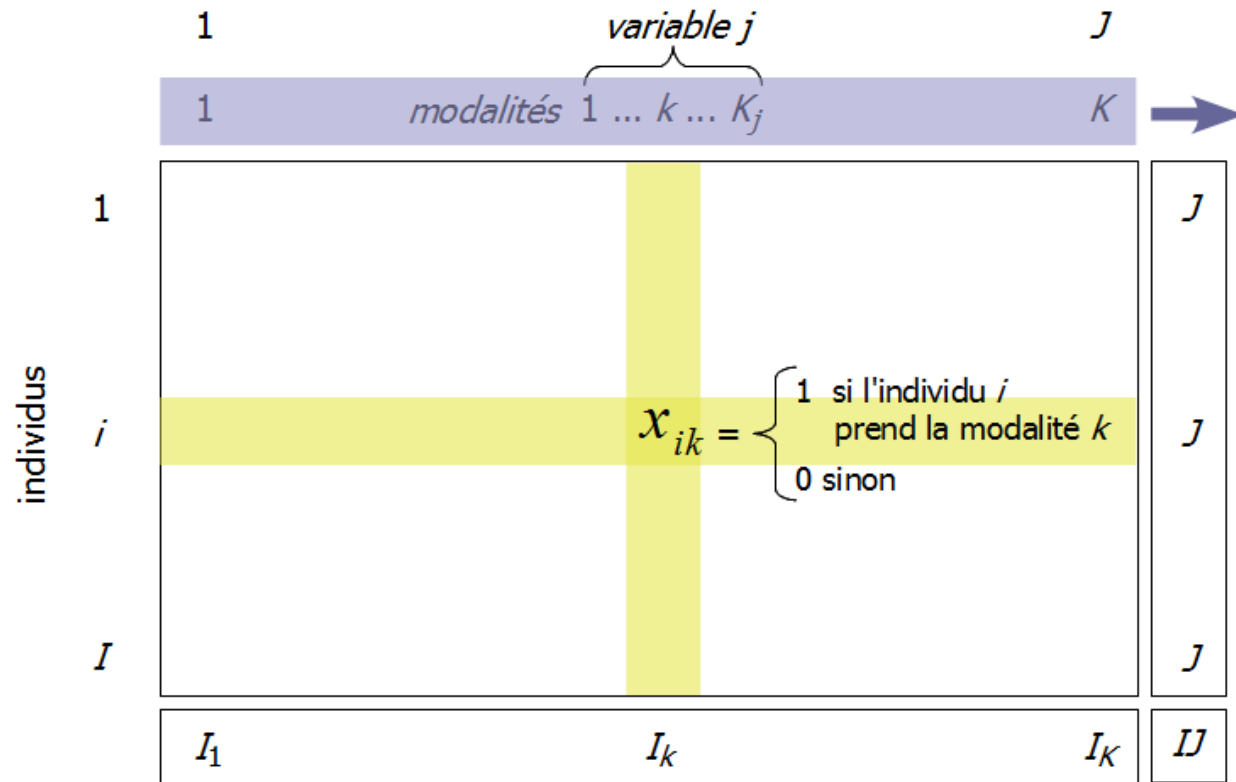
Apport	Endettement	Intitulé
Apport	End_3	MR
pas_Apport	End_2	MR
Apport	End_4	MR
pas_Apport	End_3	MR
Apport	End_2	MME

*extrait*

### Tableau des données transformées Codage « binaire » (TDC)

Apport	pas Apport	End_1	End_2	End_3	End_4	MELLE	MME	MR
1	0	0	0	1	0	0	0	1
0	1	0	1	0	0	0	0	1
1	0	0	0	0	1	0	0	1
0	1	0	0	1	0	0	0	1
1	0	0	1	0	0	0	1	0

## Notations

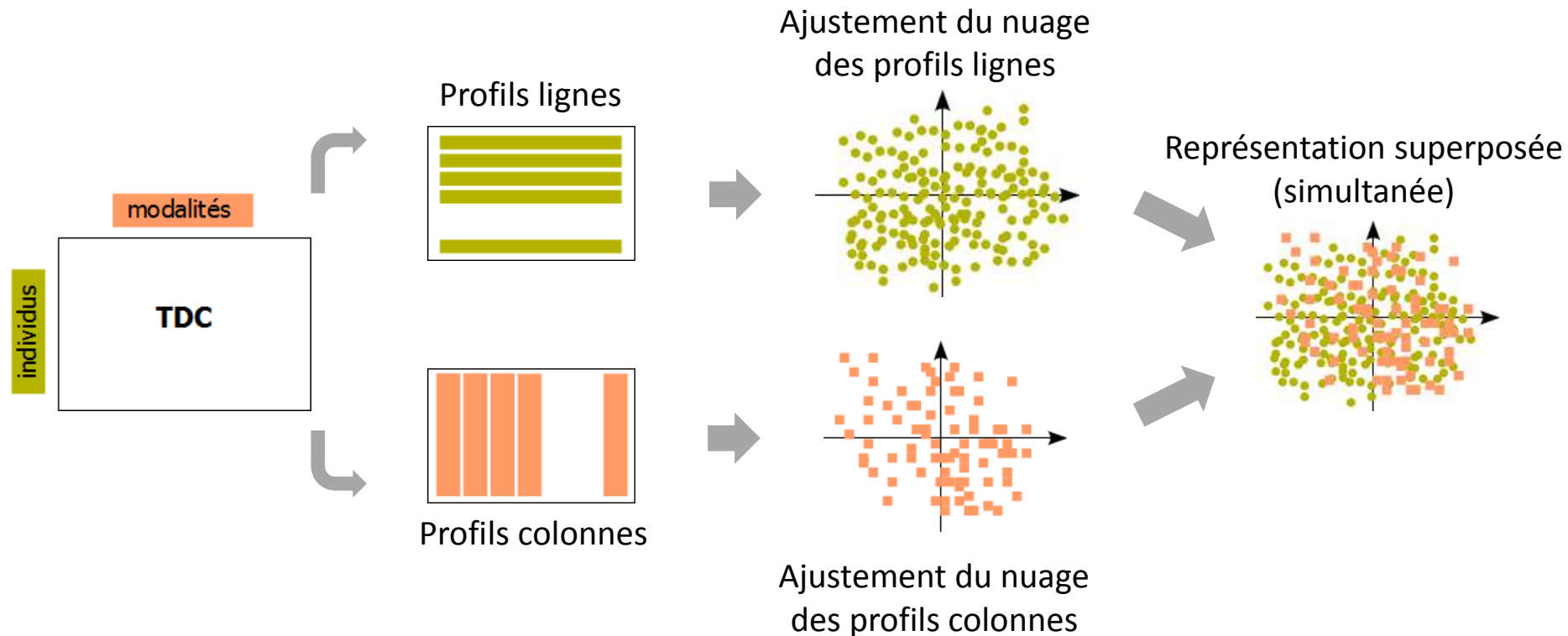


Rôle important  
joué par les  
**modalités**  
en ACM

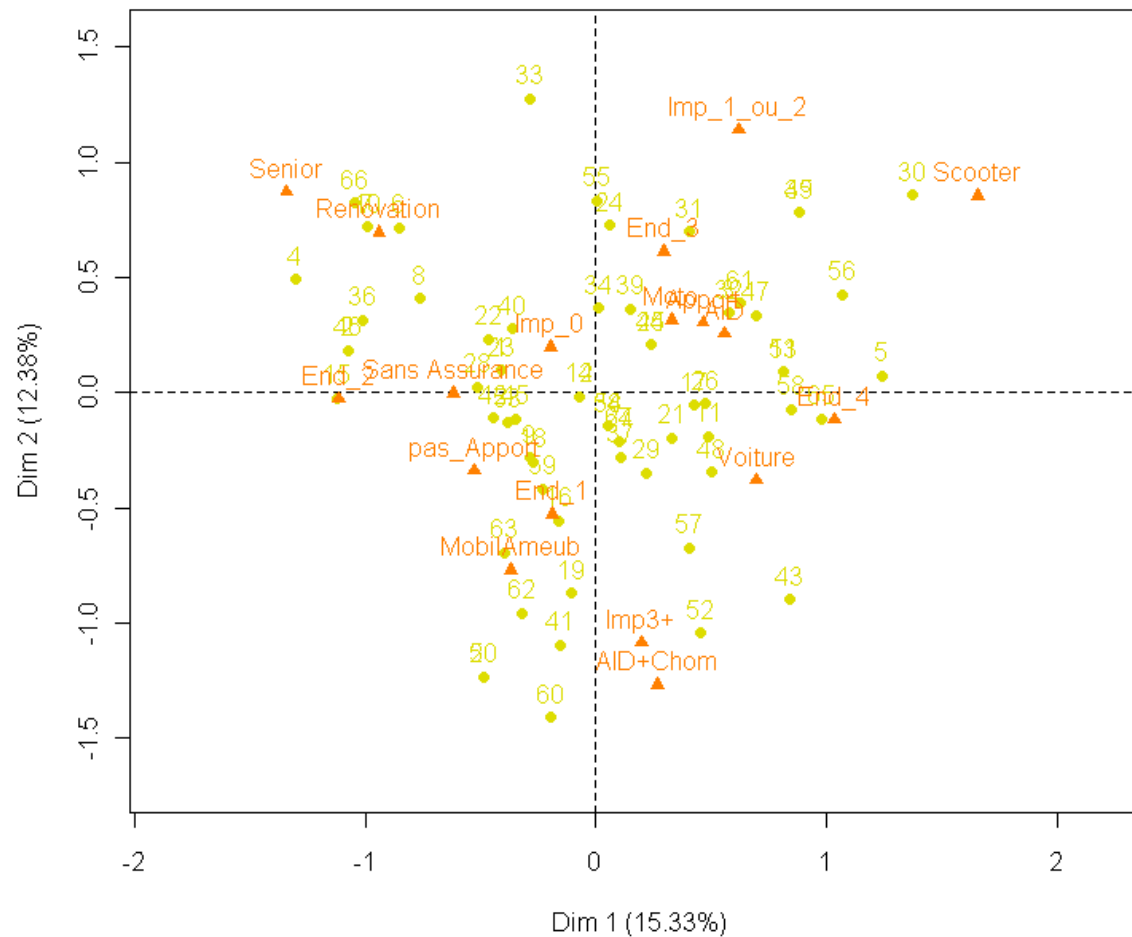
- **Lignes** : ressemblances entre individus
- **Colonnes** : étudier les liaisons (correspondances multiples) entre les modalités

### 3.3 – Le principe d'une ACM

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



## Exemple « Crédit » - Représentation superposée, axes 1 et 2



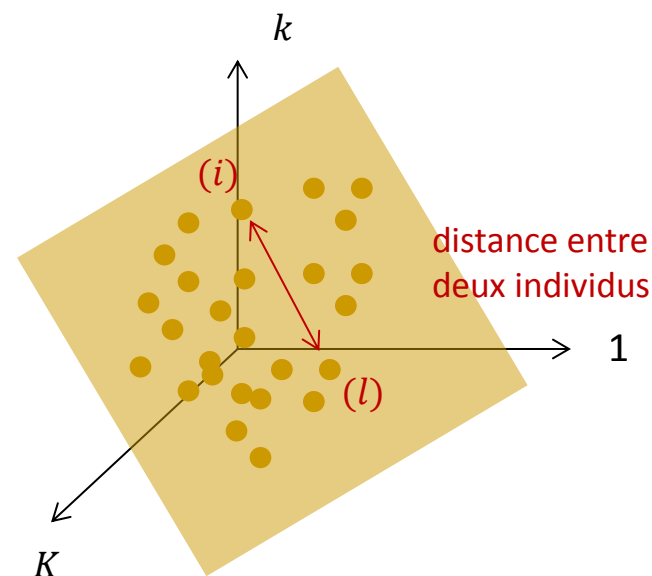
## 3.4 – L'ajustement des deux nuages de profils

- 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	...	$k$	...	$K$	$\Sigma$
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 ?

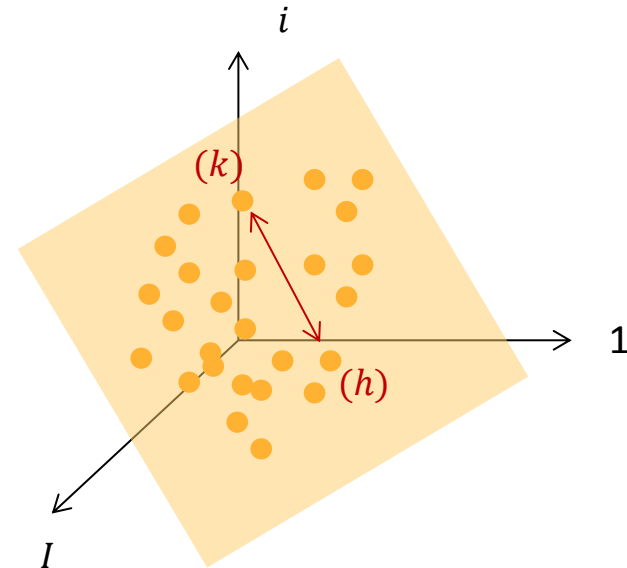
Mobil	Renov	Moto	Scooter	Side car	auto	Apport	pas Apport	Imp0	Imp1_2	Imp3+
1	0	0	0	0	0	1	0	1	0	0
0	0	0	0	0	1	0	1	0	1	0
1	0	0	0	0	0	0	1	0	0	1
0	0	0	0	0	1	0	1	0	1	0
1	0	0	0	0	0	0	1	0	1	0

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

- 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			$\frac{x_{1k}}{I_k}$		
$\vdots$					
$i$			$\frac{x_{ik}}{I_k}$		
$\vdots$					
$I$			$\frac{x_{Ik}}{I_k}$		
$\Sigma$	1		1		1



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

## Distance entre deux modalités

Mobil	Renov	Moto	Scooter	Side car	auto	Apport	pas Apport	Imp0	Imp1_2	Imp3+
1	0	0	0	0	0	1	0	1	0	0
0	0	0	0	0	1	0	1	0	1	0
1	0	0	0	0	0	0	1	0	0	1
0	0	0	0	0	1	0	1	0	1	0
1	0	0	0	0	0	0	1	0	1	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

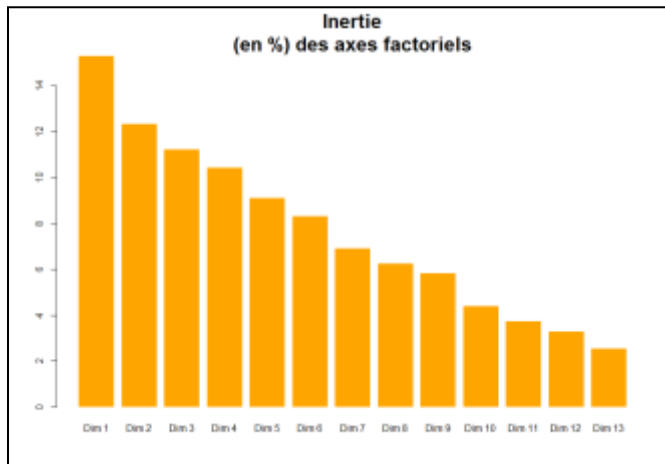
$$I(N_K) = I(N_I) = \frac{K}{J} - 1$$

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

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

```
> res$eig
```

	eigenvalue	percentage of variance	cumulative percentage of variance
dim 1	0.39858654	15.330251	15.33025
dim 2	0.32188055	12.380021	27.71027
dim 3	0.29241671	11.246797	38.95707
dim 4	0.27223718	10.470661	49.42773
dim 5	0.23766789	9.141073	58.56880
dim 6	0.21655909	8.329196	66.89800
dim 7	0.18031383	6.935147	73.83315
dim 8	0.16322576	6.277914	80.11106
dim 9	0.15178184	5.837763	85.94882
dim 10	0.11489708	4.419118	90.36794
dim 11	0.09739129	3.745819	94.11376
dim 12	0.08622601	3.316385	97.43015
dim 13	0.06681622	2.569855	100.00000



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

## Eloignement ou proximité avec l'origine des axes



## 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
	$\frac{I_1}{IJ}$	$\dots$	$\frac{I_k}{IJ}$	$\dots$	$\frac{I_K}{IJ}$
				1	

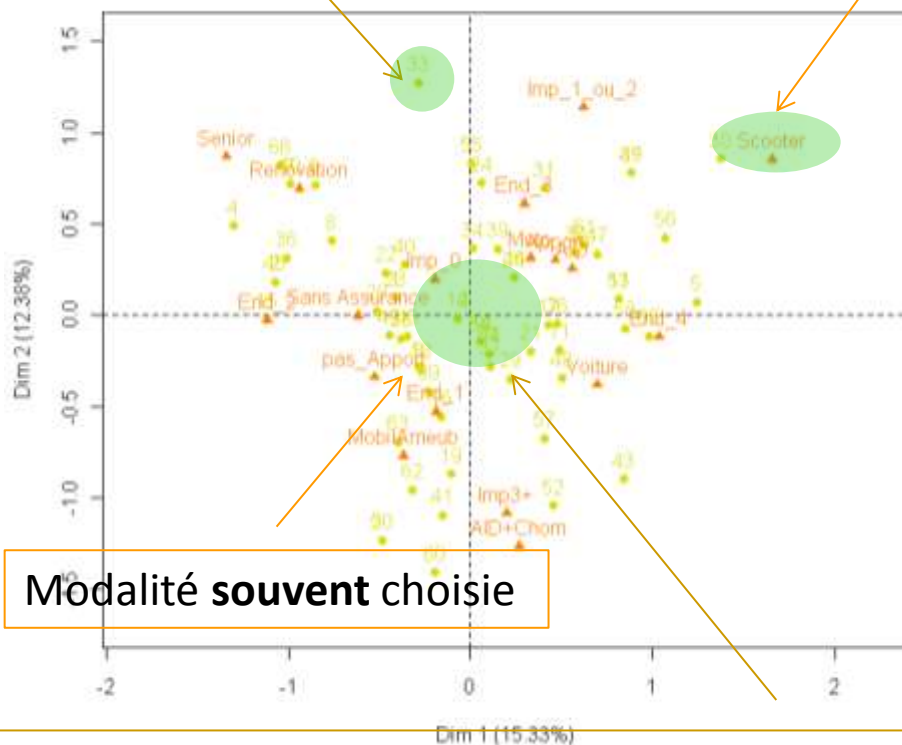
Centre de gravité des **modalités**  $G_K$

Centre de gravité des **individus**  $G_I$

## Eloignement par rapport au centre de gravité

Individu prenant des modalités **rares**

Modalité **peu** choisie



Modalité **souvent** choisie

Individu prenant des modalités **souvent** 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*

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

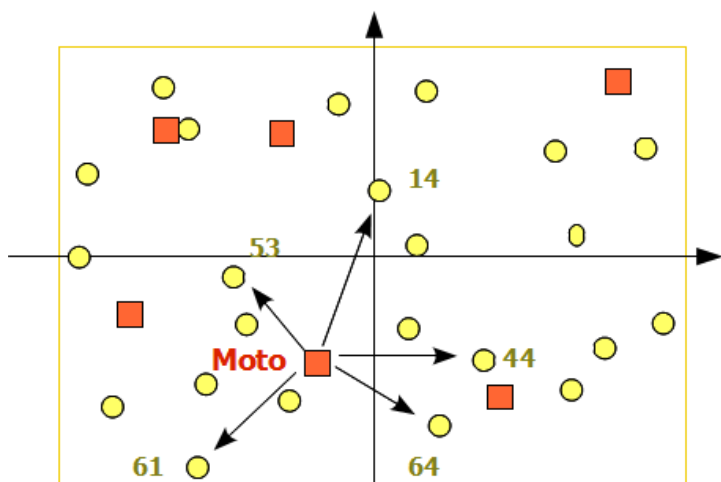
*Un individu est proche de l'origine s'il choisit des modalités fréquentes*



# Relations quasi barycentriques ou de transition

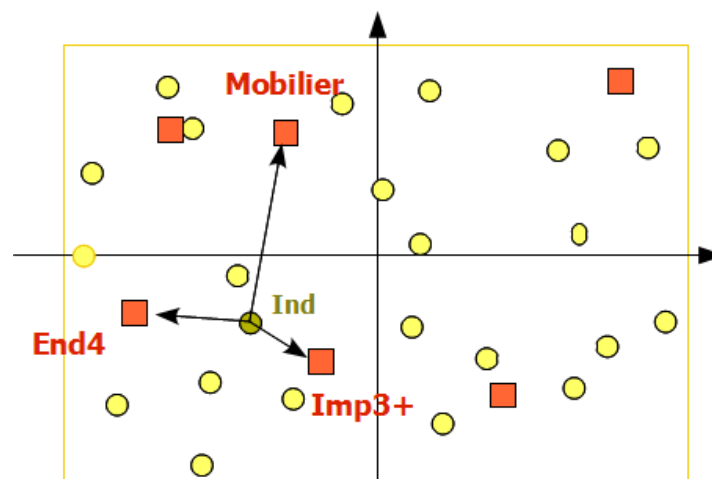
## Proximité entre lignes et colonnes

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



*Représentation  
superposée*

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



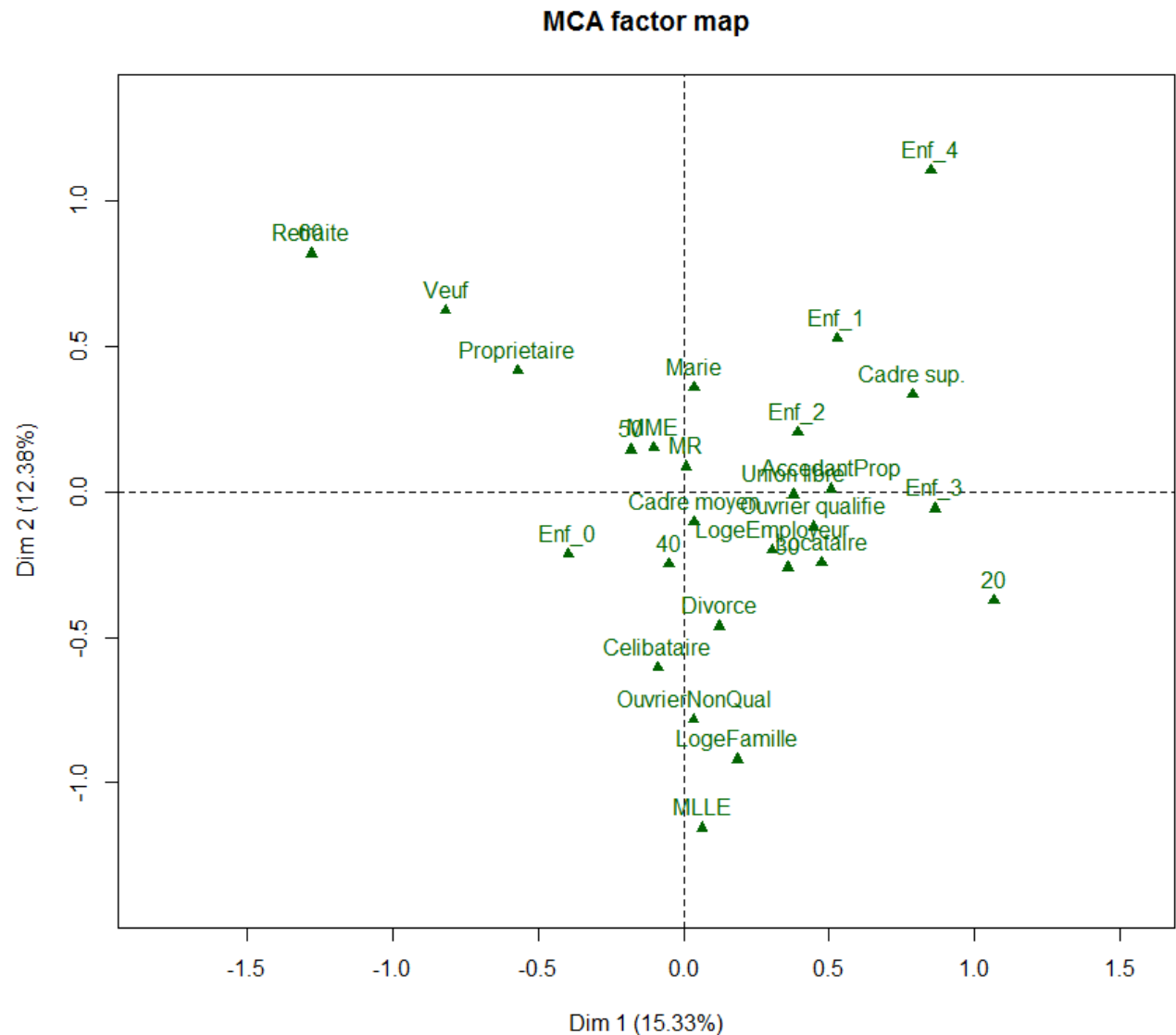
Une modalité est au (quasi) centre de gravité des individus qui la possèdent

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

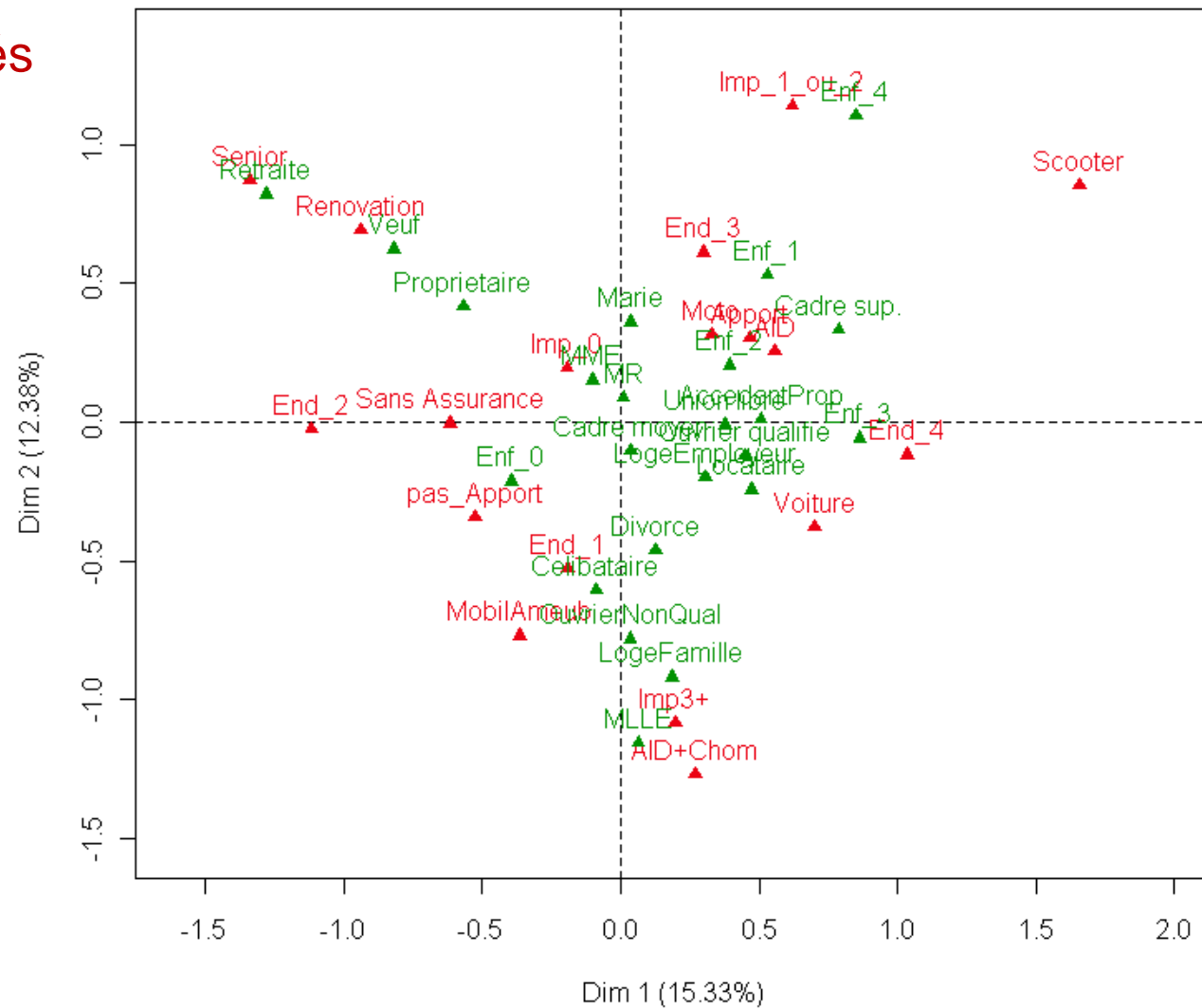
## Plan (1,2) des individus et des modalités actives



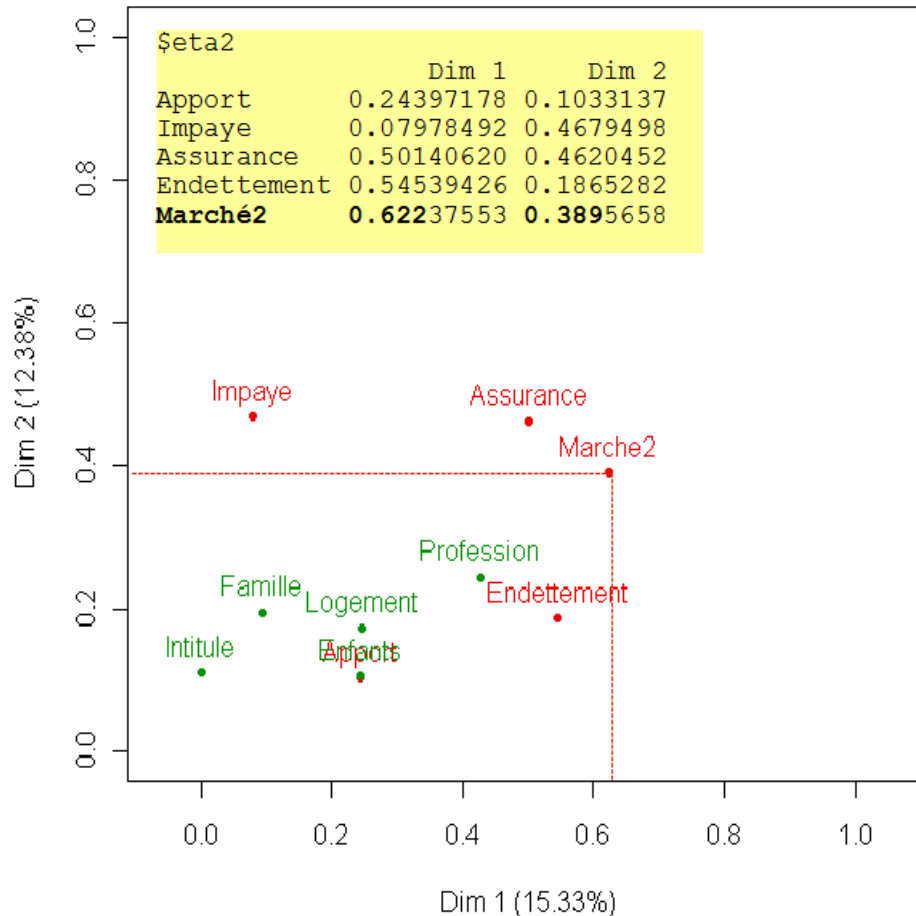
## Plan (1,2) des modalités illustratives



## Plan (1,2) des modalités actives et illustratives



## La représentation graphique des variables

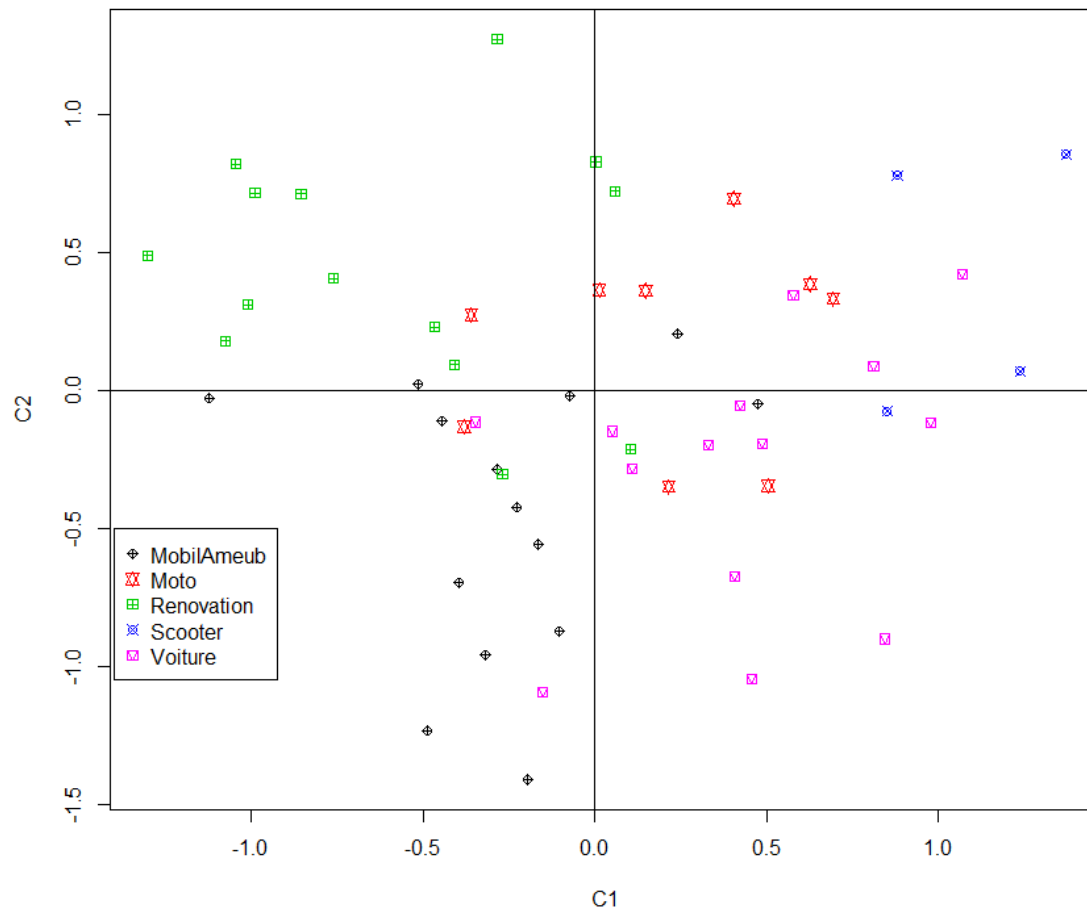


La coordonnée d'une variable ( $j$ ) sur l'axe ( $s$ ) traduit **l'intensité de la liaison** entre la variable et l'axe

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

La liaison est calculée par le **rapport de corrélation** entre la variable et la composante principale associée à l'axe

## Les sous-populations sur le plan (1,2)

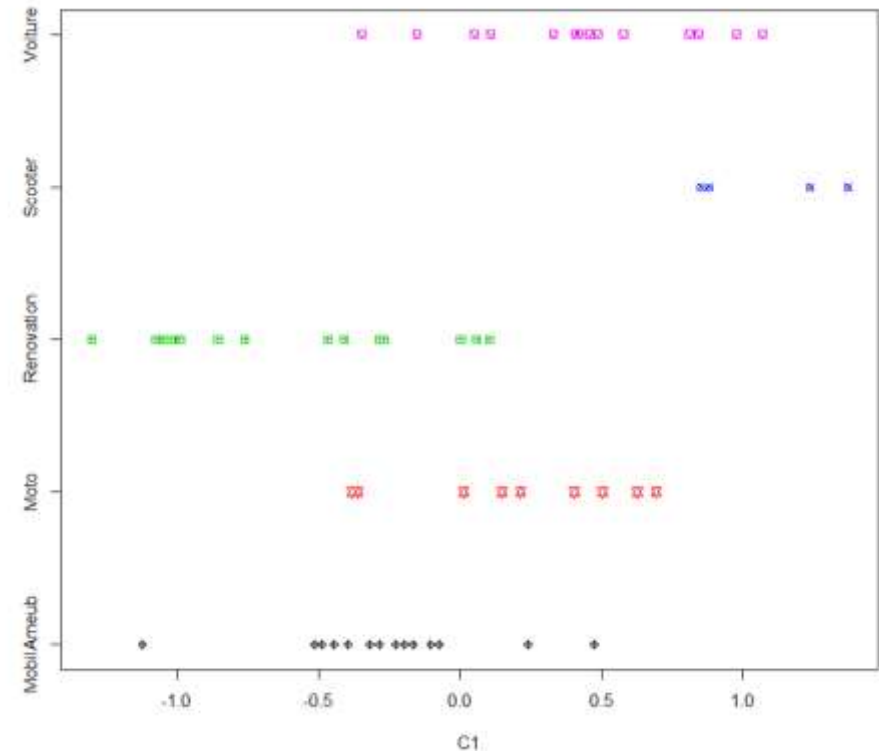


1 <sup>e</sup> comp. princ.	Marché
-0.407912134	Renovation
-1.073705596	Voiture
0.578066865	Renovation
-1.302326720	Scooter
1.238843153	Renovation
.	.
.	.
.	.
-1.044908996	Renovation

```
> summary(lm(C1~credit$Marche2))
...
Multiple R-squared: 0.6224
```

**Rapport de corrélation élevé**  
 les modalités de la variable forment  
 autant de sous-populations  
**homogènes** et **bien séparées** le  
 long de l'axe

## Les sous-populations le long de l'axe 1



```
stripchart(C1~credit$Marche2,
+ pch=(10:14), col=c(1:4,6))
```

## 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

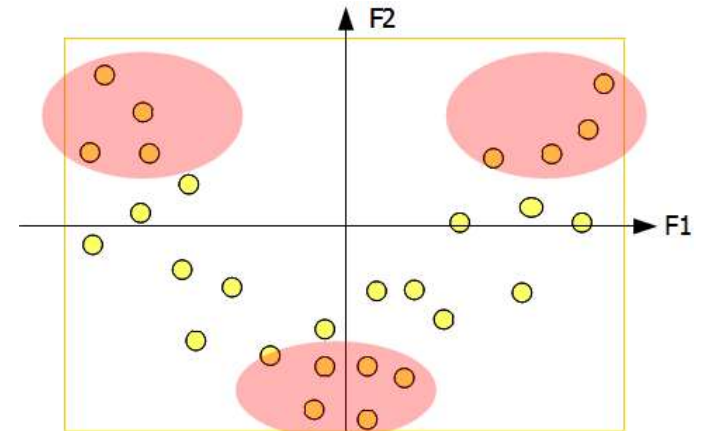
$$I_s = \frac{1}{J} \sum_{j=1}^J \eta^2(j, C_s)$$

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

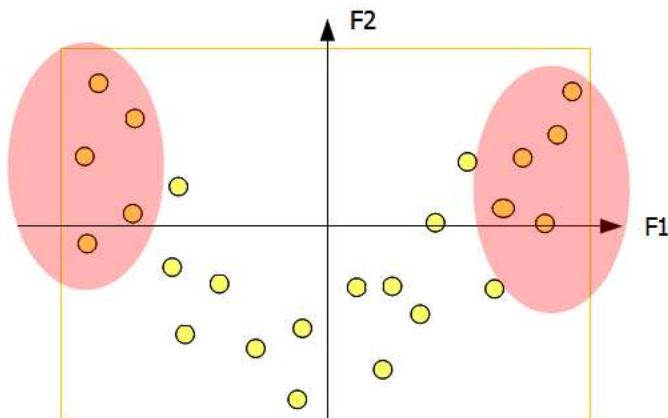


## L'effet Gutman

**Facteur 2** : opposition extrêmes / moyens

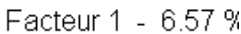


**Facteur 1** : facteur d'échelle (« effet taille »)



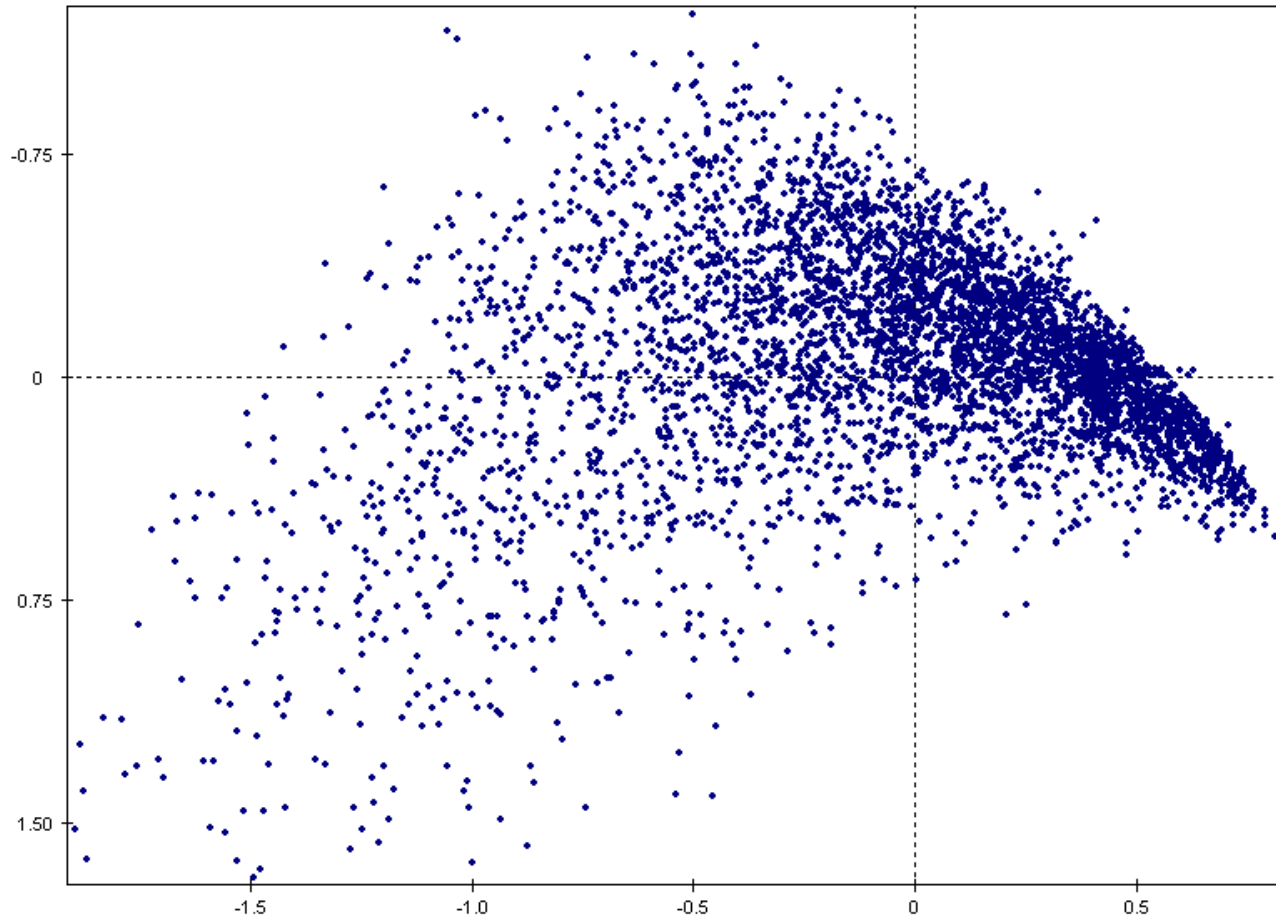
Apparaît très souvent lorsque les modalités de la variable possèdent un ordre naturel ou lorsqu'elles proviennent d'une variable quantitative mise en classes

# Plan des modalités



## Plan des individus

Facteur 4 - 3.41 %



Facteur 1 - 6.57 %