

Analyse des correspondances multiples

Realisé par:
IDABELLA ZINEB
MARROUSS IMANE

Plan

Données -Problématique et objectifs

Transformation du tableau de données

Étude des individus

Étude des modalités

Tableau de Burt

Données- Problématique & Objectifs

L'Analyse Factorielle des Correspondances (AFC) est conçue pour les situations où les individus sont décrits par deux variables qualitatives.

Cette situation est courante lorsque les données sont issues d'une enquête structurée autour de questions fermées.

Souvent, on se retrouve avec des individus caractérisés par plusieurs (deux ou plus) variables nominales ou ordinales.

Exemple :

Participants questionnés sur leurs **genres de films** favoris ("Action" ou "Comédie"...), leurs **salles de cinéma** préférées ("Cinéma A" ou "Cinéma B"), et **la fréquence de leurs visites** ("Jamais", "Rarement", ou "Souvent").

La nécessité de traiter des données provenant de multiples variables accentue l'importance d'approches analytiques adaptées, comme l'Analyse des Correspondances Multiples (ACM).

L'ACM est souvent appliquée dans les enquêtes à choix fermés, où les participants sélectionnent parmi des options prédéfinies.

Nous reprenons l'exemple précédent de l'enquête sur les préférences cinématographiques, sous forme du tableau suivant:

	Genre de Film	Salle de Cinema	Frequence de visite
indiv1	Action	Cinema A	Rarement
indiv2	Comedie	Cinema B	Souvent
indiv3	Action	Cinema B	Jamais
indiv4	Action	Cinema A	Souvent
indiv5	Comedie	Cinema B	Rarement

Variables :

Genre de Film

Salle de Cinéma

Fréquence de Visites

Modalités des Variables :

Genre de Film : Action, Comédie.

Salle de Cinéma : Cinéma A, Cinéma B.

Fréquence de Visites : Jamais, Rarement, Souvent.

- **Objectif général:**

L'ACM partage le même objectif que l'ACP ou l'AFC : Visualiser synthétiser l'information dans des volumes étendus de données.

- **Données:**

I individus décrits par J variables nominales ou ordinales.

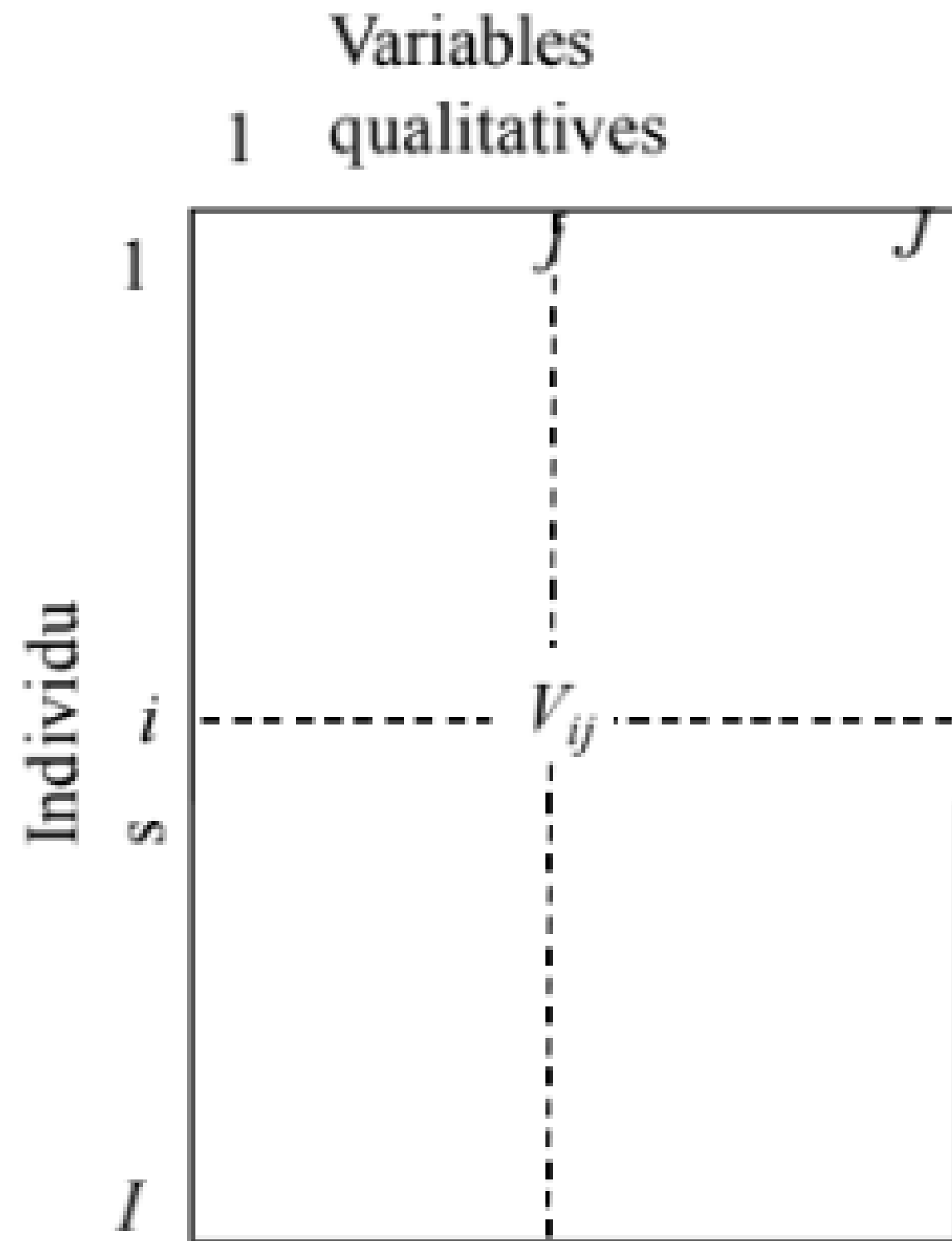
Un individu = une ligne du tableau = ensemble de modalités

L'ACM vise à mettre en lumière

- Les relations entre les modalités des différentes variables.
- Les liens entre individus et modalités.
- Les relations entre les variables, basées sur les relations entre les modalités.

Transformation de tableau de données

Tableau de Données



I individus
 J variables qualitatives

V_{ij} : modalité de la variable j
possédée par l'individu i

Exemple :

enquête où I personnes sont interrogées sur J
questions à choix multiples.

Tableau Disjonctif Complet

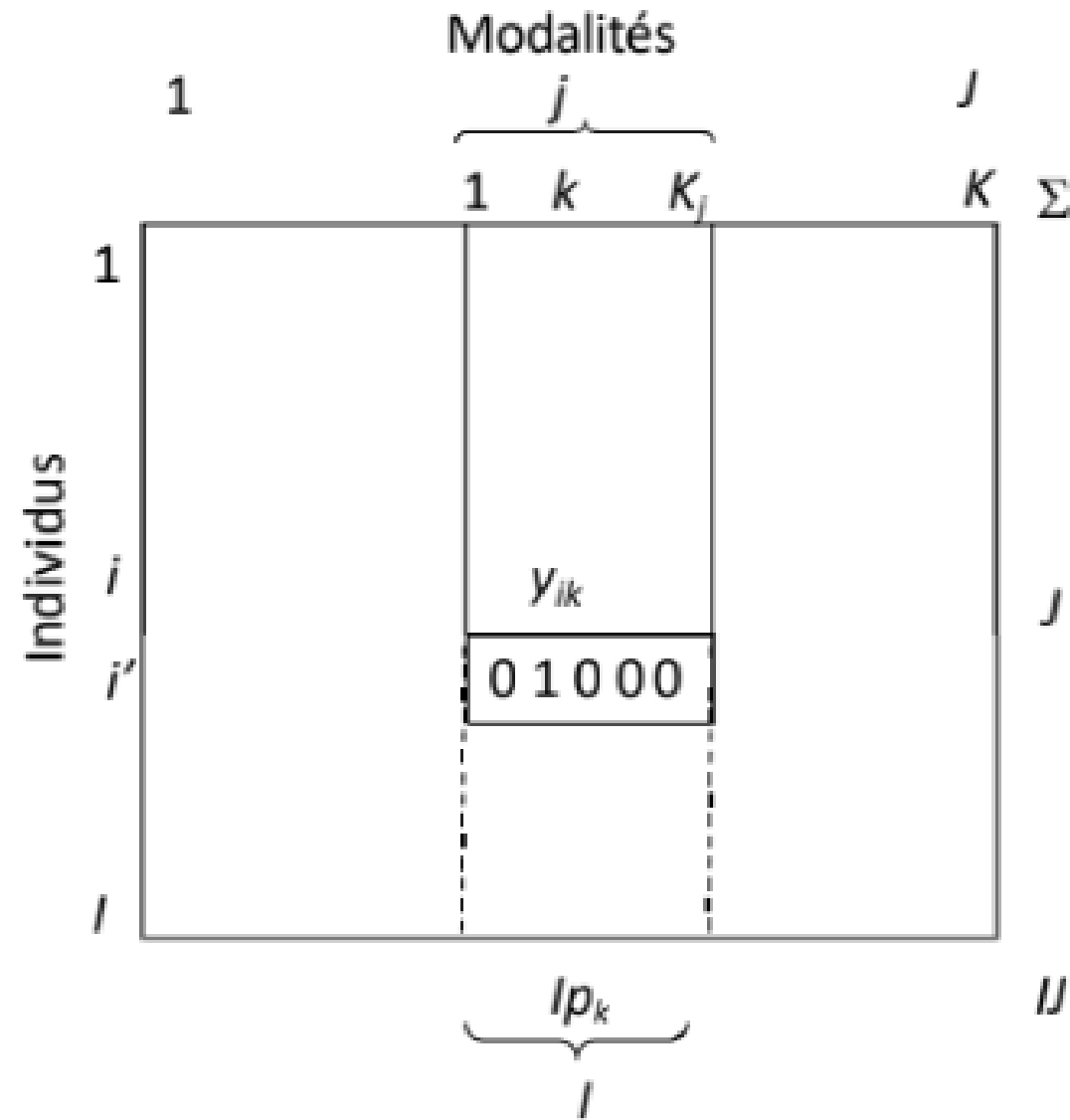
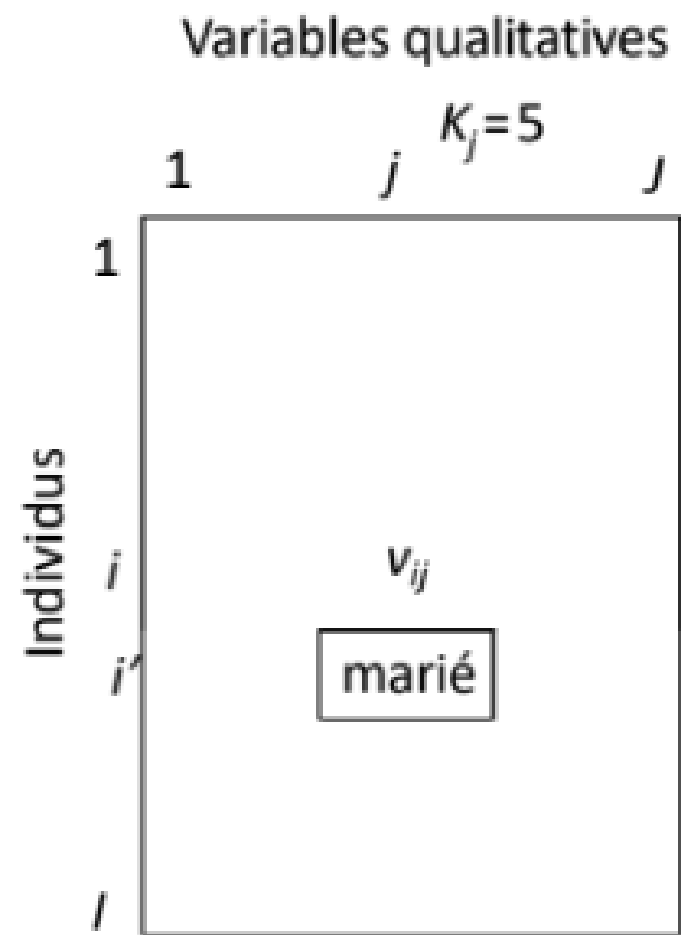


Tableau disjonctif complet (TDC)

Dans le TDC :

les lignes sont les individus et les colonnes sont les modalités des variables qualitatives .

I individus

J variables

K modalités

K_j : modalité de la variable j

y_{ik} : vaut 1 si l'individu i possède la modalité k de la variable j et 0 sinon.

$I * P_k$: nombre d'individus fois la proportion d'individus possédant la modalité k .

On résume que la disjonction complète est **une codification en données binaires**

Transformation du tableau disjonctif complet

Le poids d'un individu est $1/I$

$Y_{ik} = 1$ si i possède la modalité k de la variable j
 $= 0$ sinon

p_k : proportion d'individus possédant la modalité k .

X_{ik} : est la variable continue transformée de Y_{ik} , calculée par : **$X_{ik} = Y_{ik} / p_k$**

Pour centrer la variable continue, on soustrait 1 du résultat : **$X_{ik} = (Y_{ik} / p_k) - 1$**

$$X_{ik} = Y_{ik} / p_k$$

$$\frac{\sum_{i=1}^I X_{ik}}{I} = \frac{1}{I} \frac{\sum_{i=1}^I Y_{ik}}{p_k} = \frac{1}{I} \frac{I \times p_k}{p_k} = 1$$

$$\text{Centrage : } X_{ik} = Y_{ik} / p_k - 1$$

Exemple:

	genre- movie	lecture	sport
1	action	roman	football
2	comédie	BD	natation
3	drame	magazine	basket
4	action	BD	football
5	comédie	roman	natation

tableau de données

3 questions et 9 modalités

	movie- action	movie- comédie	movie- drame	lecture- roman	lecture- BD	lecture- magazin	sport- football	sport- natation	sport- basket
1	1	0	0	1	0	0	1	0	0
2	0	1	0	0	1	0	0	1	0
3	0	0	1	0	0	1	0	0	1
4	1	0	0	0	1	0	1	0	0
5	0	1	0	1	0	0	0	1	0

tableau disjonctif complet

centrage de données:

Toutes les variables ont trois modalités, donc $pk=3$.

Le résultat centré serait obtenu en divisant chaque élément de toutes les colonnes dans le TDC par pk et puis on soustrait 1 $\Rightarrow 1/3 - 1 = -0.66$ ou : $0/3 - 1 = -1$.

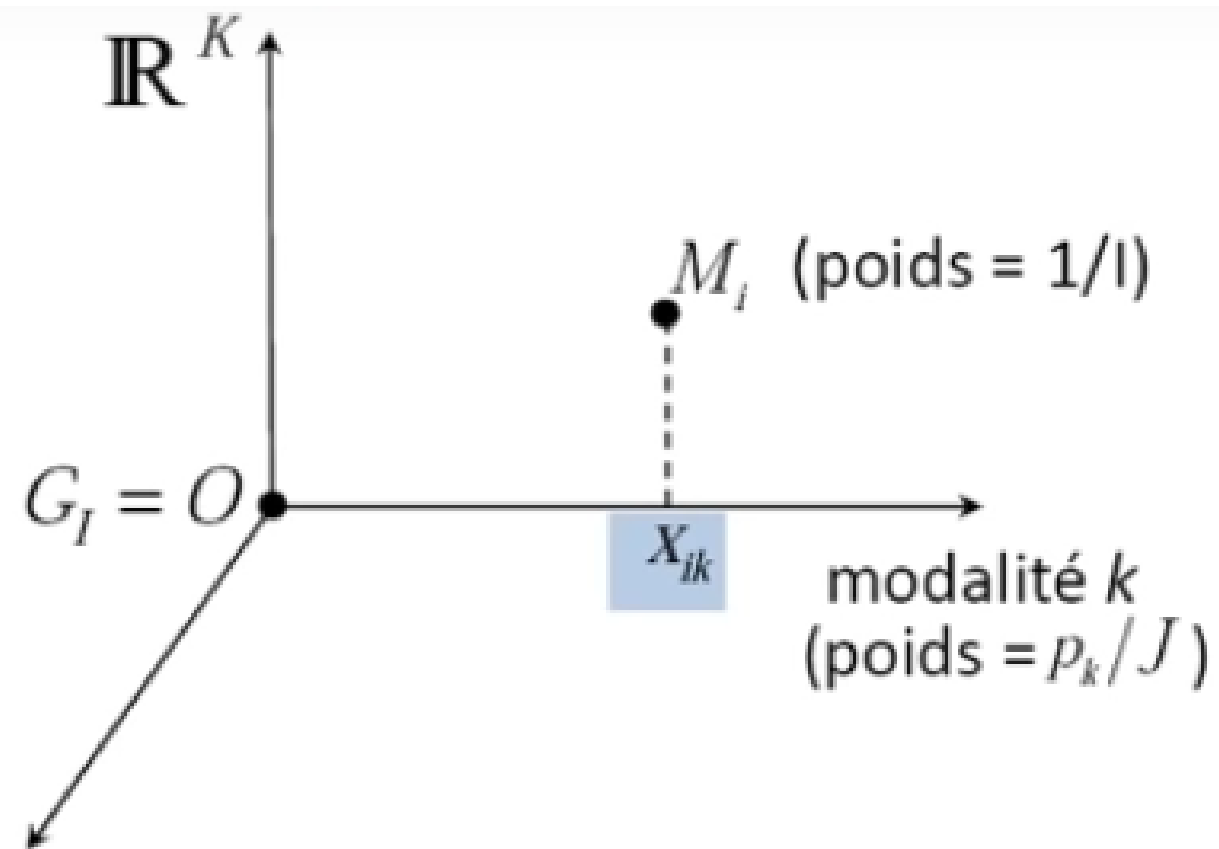
	movie-action	movie-comedie	movie-drame	lecture-roman	lecture-BD	lecture-magazin	sport-football	sport-natation	sport-basket
1	-0.66	-1	-1	-0.66	-1	-1	-0.66	-1	-1
2	-1	-0.66	-1	-1	-0.66	-1	-1	-0.66	-1
3	-1	-1	-0.66	-1	-1	-0.66	-1	-1	-0.66
4	-0.66	-1	-1	-1	-0.66	-1	-0.66	-1	-1
5	-1	-0.66	-1	-0.66	-1	-1	-1	-0.66	-1

Etude des individus

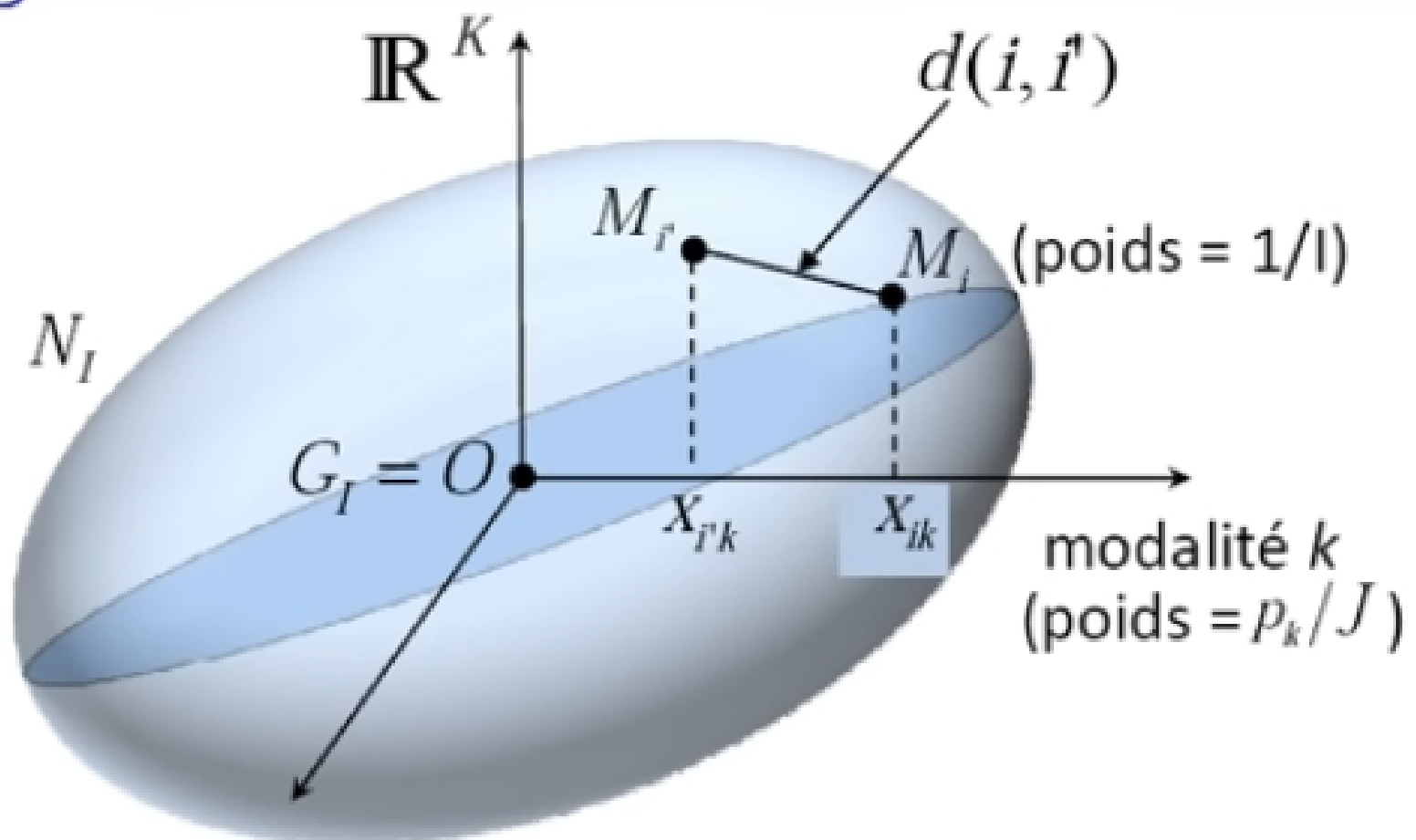
Nuage des individus

Tableau Disjonctif Complet

	Modalités		
	1	k	K
Individus	1		
	i	x_{ik}	
	I		



*un individu $i \Rightarrow$ une ligne de TDC
 \Rightarrow un ensemble de valeurs $k \Rightarrow$
 un point dans un espace de K
 dimension*



*N_I : nuage de tous les individus
 I centré en O*

Individus - Distance

distance entre individus:

$$d_{i,i'}^2 = \sum_{k=1}^K \frac{p_k}{J} (x_{ik} - x_{i'k})^2 = \sum_{k=1}^K \frac{p_k}{J} \left(\frac{y_{ik}}{p_k} - \frac{y_{i'k}}{p_k} \right)^2 = \frac{1}{J} \sum_{k=1}^K \frac{1}{p_k} (y_{ik} - y_{i'k})^2$$

- 2 individus prennent les mêmes modalités : distance = 0
- 2 individus ont en commun beaucoup de modalités : distance petite
- 2 individus dont l'un des 2 possède une modalité rare: distance grande pour prendre en compte la spécificité d'un des 2.
- 2 individus ont en commun une modalité rare: distance petite pour prendre en compte leur spécificité commune.

distance des individus a l'origine:

$$d(i, G_I)^2 = \sum_{k=1}^K \frac{p_k}{J} (x_{ik})^2 = \sum_{k=1}^K \frac{p_k}{J} \left(\frac{y_{ik}}{p_k} - 1 \right)^2 = \frac{1}{J} \sum_{k=1}^K \frac{y_{ik}}{p_k} - 1$$

- Un individu est d'autant plus loin de l'origine qu'il comporte des modalités rares

Inertie

Inertie totale du nuage

$$\text{Inertie}(N_I) = \sum_{i=1}^I \underbrace{\frac{1}{I} d^2(i, O)}_{\text{inertie de } i} = \sum_{i=1}^I \left(\frac{1}{IJ} \sum_{k=1}^K \frac{y_{ik}}{p_k} - \frac{1}{I} \right) = \frac{K}{J} - 1$$

- Différemment en AFC dont l'inertie total est égale au Khi-deux associé au tableau de contingence divisé par le nombre d'individus.
En ACM l'inertie est égale au nombre moyen de modalités diminué de 1 (c-a-d l'inertie dépend de la forme de tableau et pas de son contenu)

Ajustement du nuage des individus

- Comme pour toute analyse factorielle, nous allons maintenant projeter le nuage des individus pour le visualiser dans un espace réduit de dimension plus petite.
- souvent on se contentera de représenter le nuage sur un plan (2 dimensions).
- Pour ce faire, on projette ce nuage sur une suite d'axes orthogonaux d'inertie maximum.

BUT:

recherche d'un axe qui maximise l'inertie et qui est orthogonal aux axes précédemment trouvés.

Exemple

Soit un tableau disjonctif complet sur des réponses à un questionnaire (26 enquêtés, 4 questions posées sur un produit) :

	Q1-1	Q1-2	Q1-3	Q2-1	Q2-2	Q3-1	Q3-2	Q4-1	Q4-2	Q4-3
A	1	0	0	1	0	1	0	1	0	0
B	1	0	0	0	1	0	1	1	0	0
C	1	0	0	1	0	1	0	0	1	0
D	1	0	0	1	0	0	1	0	1	0
E	1	0	0	1	0	1	0	1	0	0
F	1	0	0	0	1	0	1	0	0	1
G	1	0	0	0	1	0	1	1	0	0
H	1	0	0	0	1	0	1	1	0	0
I	0	1	0	1	0	1	0	1	0	0
J	0	1	0	1	0	1	0	1	0	0
K	0	1	0	1	0	0	1	1	0	0
L	0	1	0	1	0	1	0	0	1	0
M	0	1	0	1	0	1	0	0	1	0
N	0	1	0	1	0	1	0	0	1	0
O	0	1	0	0	1	1	0	0	1	0
P	0	1	0	0	1	1	0	0	1	0
Q	0	0	1	1	0	1	0	1	0	0
R	0	0	1	1	0	1	0	0	0	1
S	0	0	1	1	0	0	1	0	1	0
T	0	0	1	0	1	0	1	0	1	0
U	0	0	1	0	1	0	1	0	1	0
V	0	0	1	0	1	0	1	0	1	0
W	0	0	1	0	1	0	1	0	0	1
X	0	0	1	0	1	0	1	0	0	1
Y	0	0	1	0	1	0	1	0	0	1
Z	0	0	1	1	0	0	1	0	0	1

4 variables: Q1 , Q2 ,Q3 ,Q4

26 individus :A,B ...Z

10 modalités

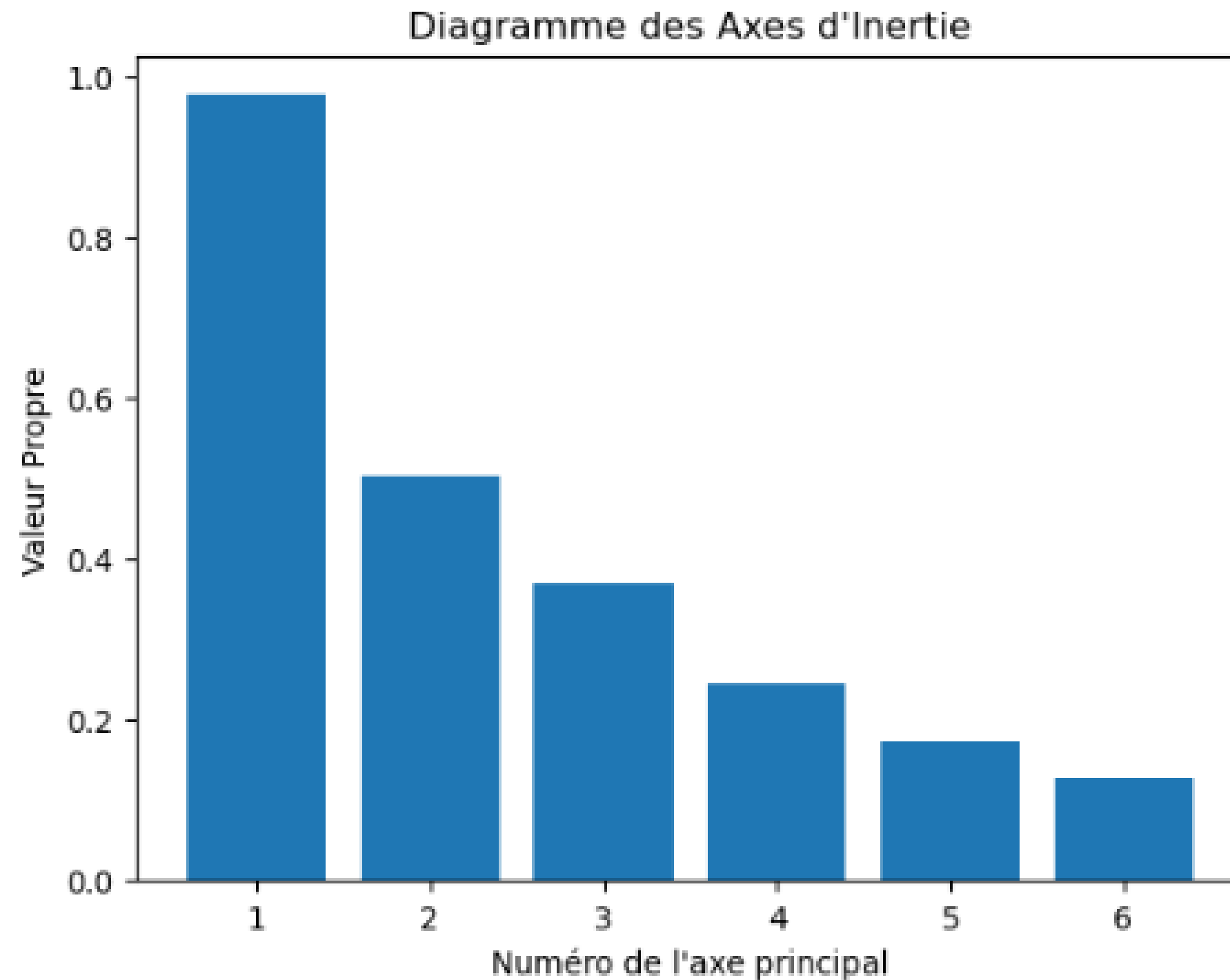
1. Ajustement de nuage

L'inertie totale est égale au nombre moyen de modalités moins une ($K/J - 1 = 10/4 - 1$), soit 1,5. Le nombre de valeurs propres non triviales à se « partager » ce montant est de dix modalités moins quatre variables, soit **six**

Axes	F1	F2	F3	F4	F5	F6
Valeurs Propres	0.977279	0.502124	0.370890	0.243821	0.174087	0.128722
Inerties	0.407722	0.209487	0.154736	0.101722	0.072629	0.053703
Inerties Cumulées	0.407722	0.617209	0.771945	0.873668	0.946297	1.000000

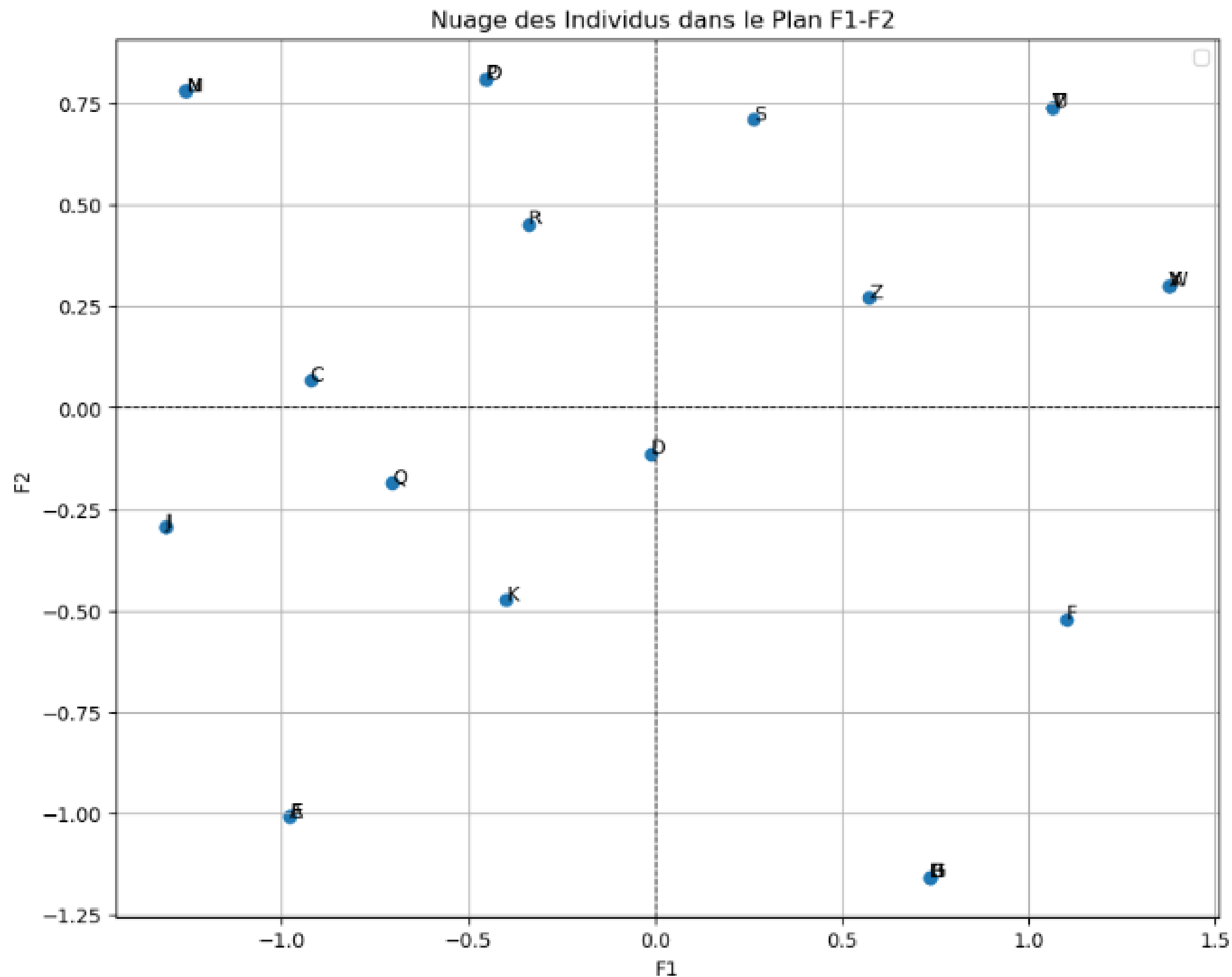
- Les 2 premiers axes (F1 et F2) expliquent une proportion significative de la variance, avec une inertie cumulée de 61.72% jusqu'à l'axe F2.
- les trois premiers axes (F1, F2, et F3) expliquent ensemble 77.19% de l'inertie totale.
- L'inertie cumulée atteint 100% à l'axe F6, ce qui signifie que la totalité de la variance des données est expliquée par l'ensemble des six axes.

D'ou le diagramme des inerties:



On choisit donc de représenter le nuage des individus dans le plan factoriel réduit formé par les deux premiers axes qui capturent une proportion significative de la variance des données

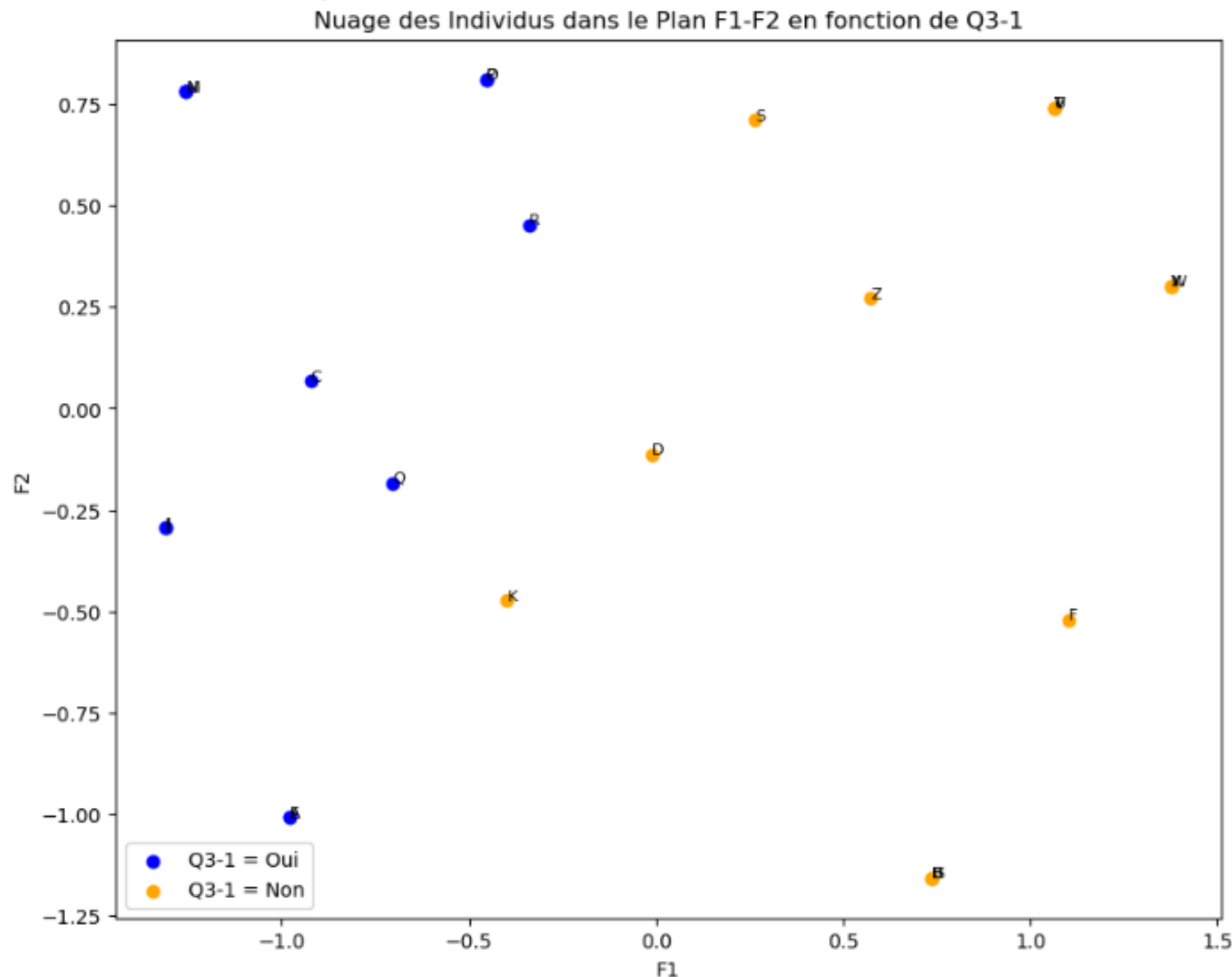
2. Nuage des individus



- les individus W et Y (ainsi D et P) sont des individus très similaires en termes de relations entre les variables originales
- les individus S et Z (Q et K) sont peu similaires
- les autres individus sont clairement indépendants
- l'individu B est clairement isolé de tous les autres points ce qui explique que B a des caractéristiques différentes de la majorité des autres individus.

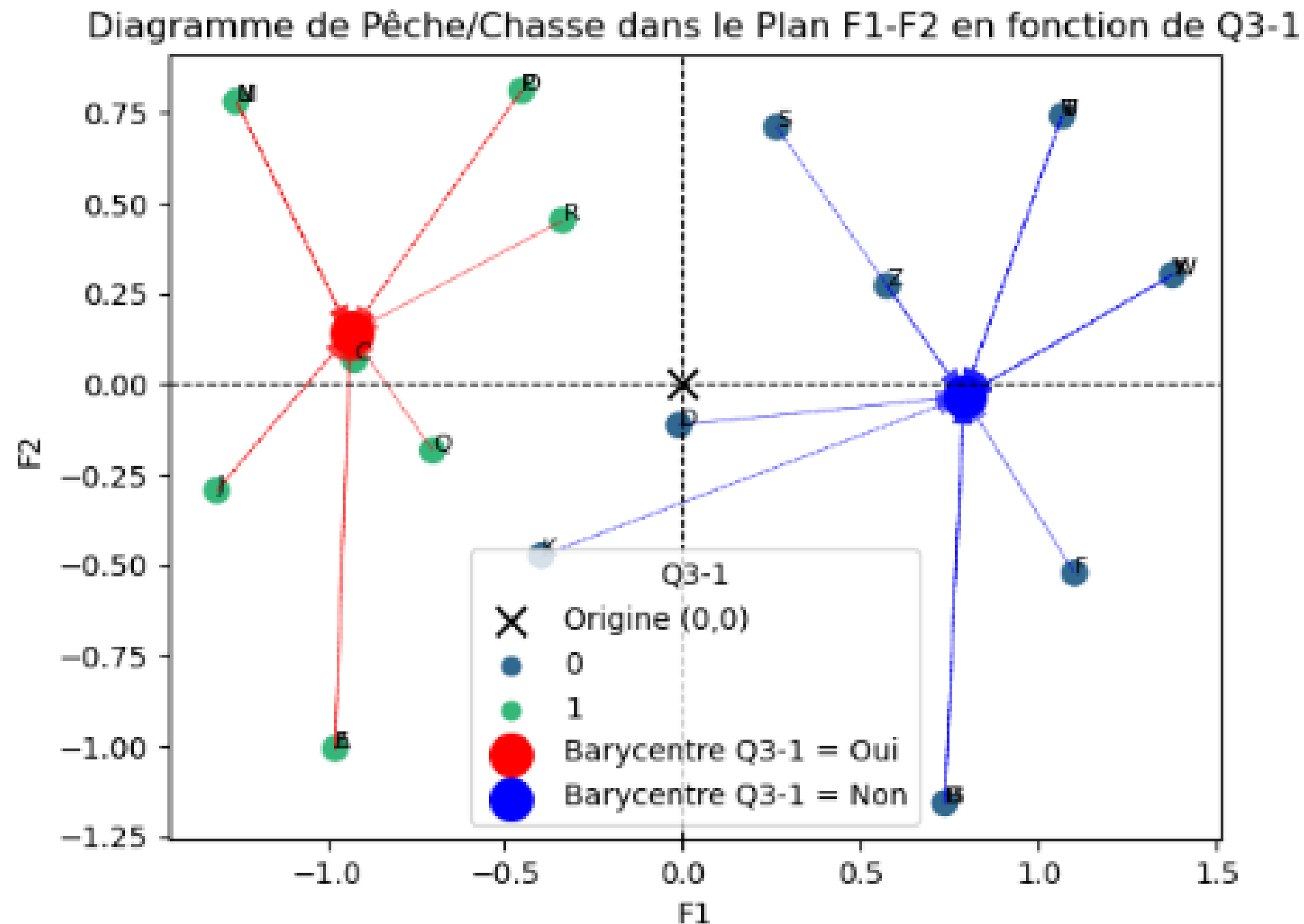
3.Représentation des individus en fonction d'une modalité

pour offrir une visualisation plus éclairante sur les relations entre les variables et faciliter la caractérisation des individus. Examinons le graphe suivant qui représente le nuage des individus selon la modalité **Q3-1** (*disant que : Q3(est le prix de produit) et Q3-1 est la modalité qui dit que le prix est très élevé*)



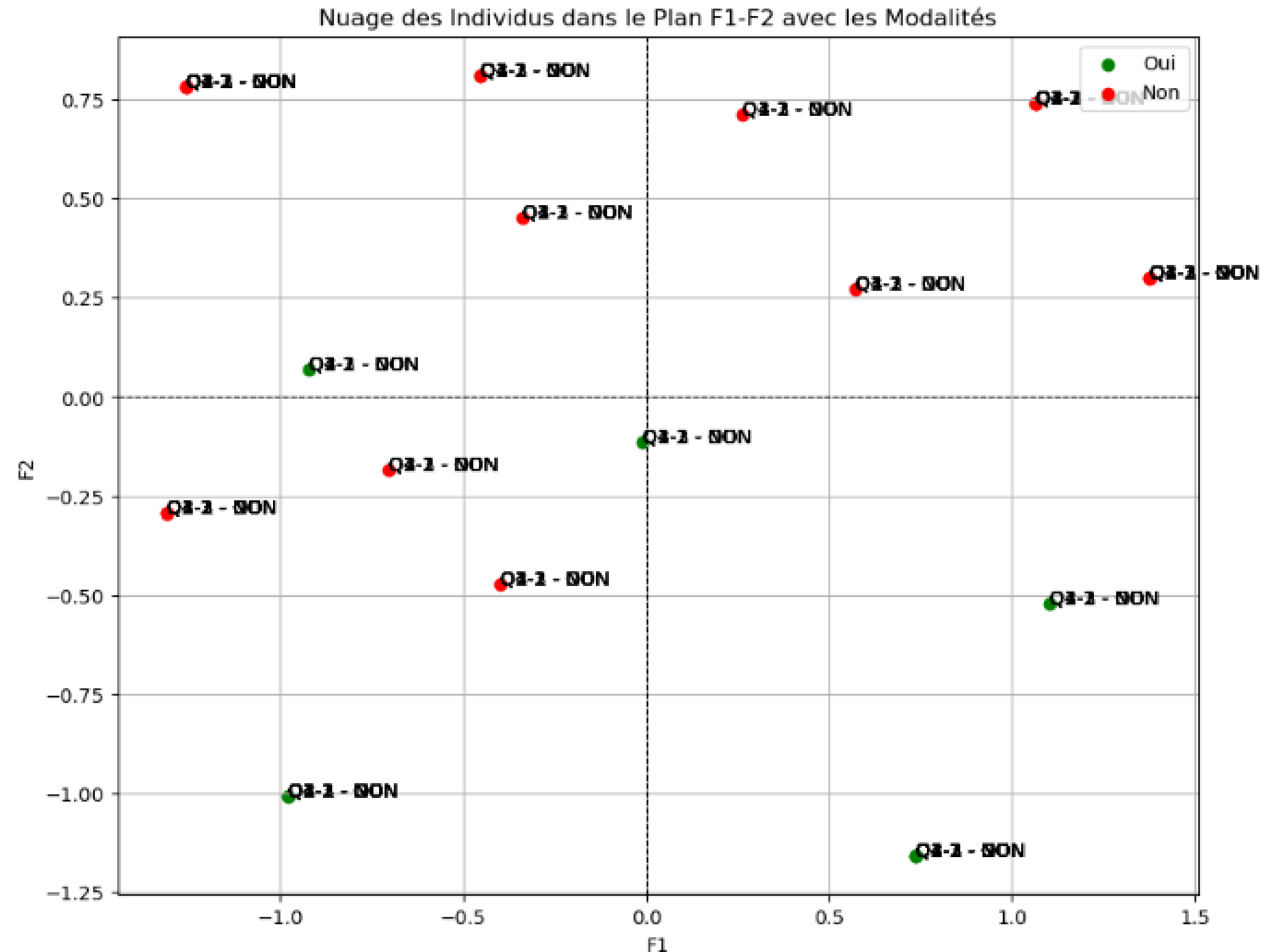
- le nuage des individus est divisé en deux groupes par rapport a la modalité "prix élevé", **groupe jaune** (individus répondants que le prix n'est pas très élevé) et **groupe bleu** des individus disant que le prix de produit est très élevé
- l'axe F2 semble séparer les individus selon qu'ils trouvent le produit est cher ou non
- On peut voir donc qu'**une modalité est un groupe d'individus**

Une autre représentation des individus en fonction de la modalité Q3-1: **Pêche -Chasse**



- **nuage bleu:** individus disant que le prix n'est pas élevé
- **nuage rouge:** individus disant que le produit est très cher
- Pour la modalité Prix élevé de produit, on a donc 2 barycentres :
 - "Prix élevé oui" en rouge
 - "Prix élevé non" en bleu
- Ici, le barycentre prix élevé "non" est plus proche de l'origine que le barycentre prix élevé "oui". Cela signifie qu'il y a peu de personnes qui trouvent le produit est cher que de personnes qui n'en font pas
- Ainsi les modalités de faibles effectifs seront plus éloignées de l'origine
- **On dit que la modalité Q3-1"non" est plus fréquente, et que Q3-1"oui" est rare**

4. Représentation individus/modalités



- **Chaque modalité est au barycentre des individus qui la prennent**

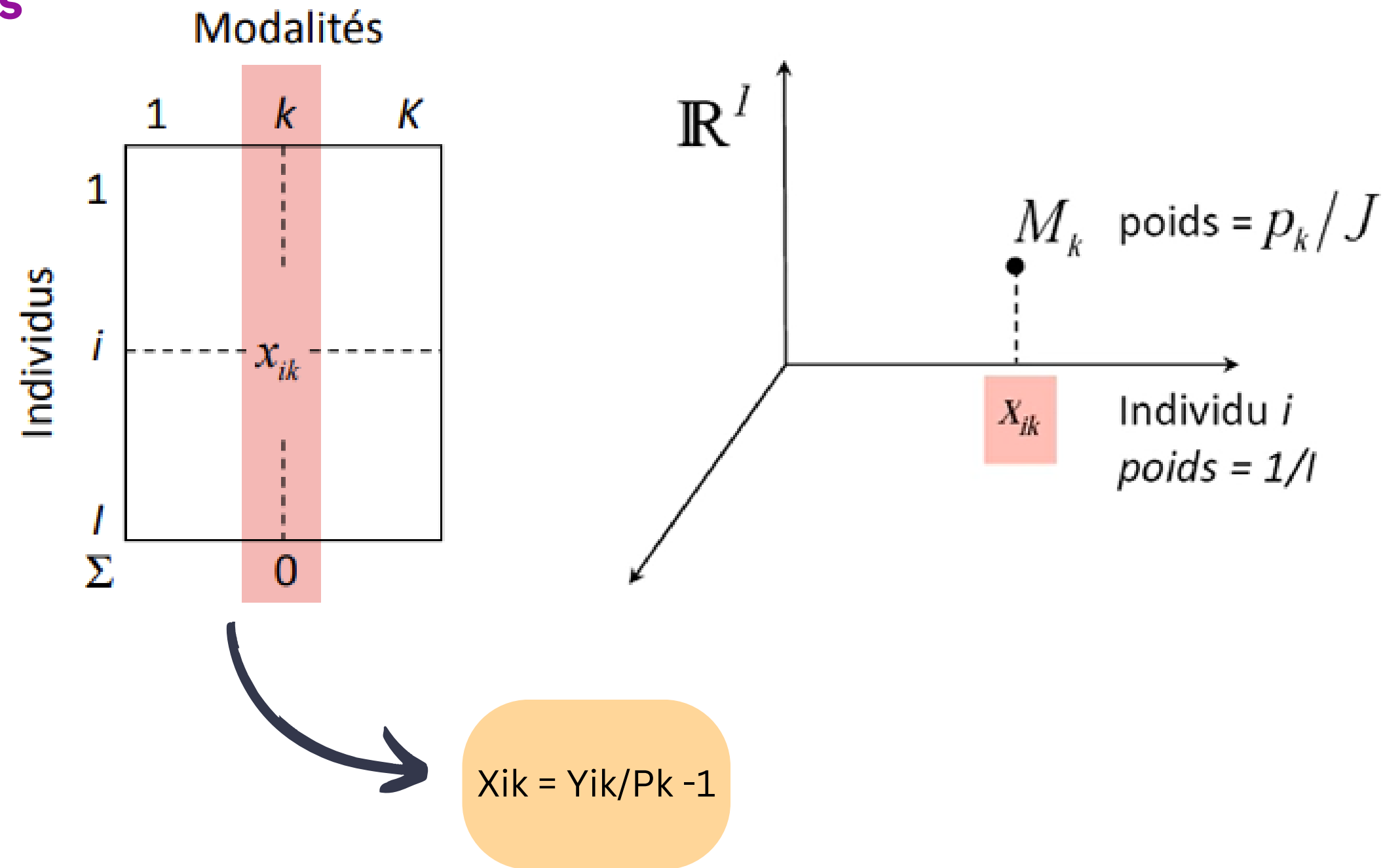
La moyenne des coordonnées pour les individus qui ont une modalité particulière sera égale aux coordonnées de la modalité dans cet espace

- **Chaque individu au barycentre des modalités qu'il possède**

pour un individu donné, le barycentre des modalités qu'il possède est situé au niveau de cet individu dans l'espace de l'analyse des correspondances multiples.

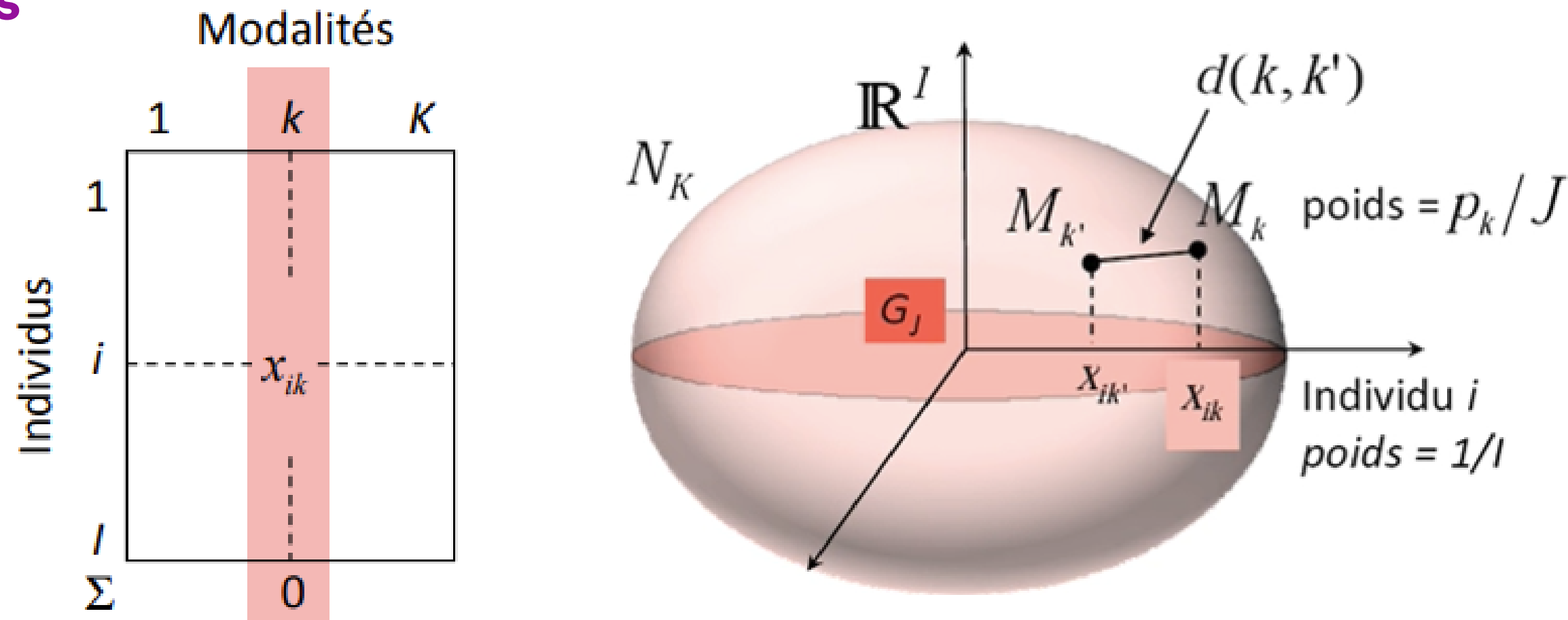
Etude des modalités

Nuage des modalités



- 1 colonne \Rightarrow I valeurs \Rightarrow 1 point dans un espace à I dimensions.

Nuage des modalités



$$Var(k) = d^2(k, O) = \sum_{i=1}^I \frac{1}{I} x_{ik}^2 = \sum_{i=1}^I \left(\frac{y_{ik}}{p_k} - 1 \right)^2 = \frac{1}{p_k} - 1$$

	p_k	1/2	1/5	1/10	1/101
	$d(k, O)$	1	2	3	10
(si $J = 10$)	$Inertie(k)$	0.05	0.08	0.09	0.099

$$Inertie(k) = \frac{p_k}{J} d^2(k, O) = \frac{1 - p_k}{J}$$

$$d^2(k, k') = \sum_{i=1}^I \left(\frac{y_{ik}}{p_k} - \frac{y_{ik'}}{p_{k'}} \right)^2 = \frac{p_k + p_{k'} - 2p_{kk'}}{p_k p_{k'}}$$

Inertie-Definitions

=>L'inertie totale du nuage des modalités est déterminée uniquement par le nombre total de modalités K et le nombre de variables J

Inertie totale du nuage des modalités :

$$I = \sum_{j=1}^J Inertie(j) = \frac{K - J}{J}$$

- Inertie d'une variable : $Inertie(j) = \frac{K_j - 1}{J}$

- Inertie d'une modalité : $Inertie(k) = \frac{1 - p_k}{J}$

Variable	NB de modalites	Inertie	NB de dimensions du ss-espace
sexe	2	1/J	1
region	21	20/J	20
departement	96	95/J	95

k_j : est le nombre de modalités de la variable j

p_k : la proportion d'individus possédant la modalité k de la variable j

Proximité entre les modalités

- Si deux modalités d'une même variable sont proches, cela signifie que les individus qui possèdent l'une des modalités et ceux qui possèdent l'autre sont globalement similaires du point de vue des autres variables
- Si deux modalités de deux variables différentes sont proches, cela peut signifier que ce sont globalement les mêmes individus qui possèdent l'une et l'autre

Tableau de Burt

Tableau de Burt

- Ensemble des liaisons entre variables prises 2 à 2 (tableau analogue à la matrice des corrélations entre variables quantitatives)
 - Le tableau est symétrique
- ⇒ L'ACM ne dépend que des liaisons entre les variables prises 2 à 2
(comme l'ACP ne dépend que de la matrice des corrélations)

		variable l		variable j	
		1	k	q	K
variable j	1				
	k		0 I_k	I_{qk}	
	q		I_{qk}	0 I_q	
	K				
Σ			$J I_k$	$J I_q$	

Exemple de Questionnaire

- Il s'agit d'une partie des données issues de l'enquête "Les étudiants et la ville" effectuée en 2001 par des étudiants de sociologie sous la direction de S. Denèfle à l'Université François Rabelais de Tours.
- L'analyse porte sur cinq questions en rapport avec le logement étudiant. L'ensemble des individus statistiques est ici un échantillon de 383 étudiants. Les questions sont les suivantes :

Question	N°	Réponses possibles	Poids (%)	Abréviation
Habitez-vous (variable "mode d'occupation")	1	seul	48,30%	Seul
	2	colocataires	13,84%	Coloc
	3	en couple	13,05%	Couple
	4	avec les parents	23,50%	Parents
	5	non réponse	1,31%	NR1
Quel type d'habitation occupez-vous ? (variable "type d'habitation")	6	cité universitaire	10,70%	Cité
	7	studio	28,20%	Studio
	8	appartement	30,29%	Appart
	9	chambre chez un particulier	5,22%	Chambre
	10	autre	19,84%	Autre
	11	non réponse	5,74%	NR2
Si vous vivez en dehors du foyer familial, depuis combien de temps ? (variable "ancienneté")	12	moins de 1 an	20,89%	< 1 an
	13	1 à 3 ans	24,80%	1-3 ans
	14	plus de 3 ans	28,72%	> 3 ans
	15	non applicable	24,80%	NA
	16	non réponse	0,78%	NR3
A quelle distance approximative de la Fac vivez-vous ? (variable "éloignement")	17	moins de 1 km	26,89%	< 1 km
	18	1 à 5 km	49,87%	1 à 5 km
	19	plus de 5 km	20,89%	> 5 km
	20	non réponse	2,35%	NR4
Quelle est la superficie de votre logement ? (variable "superficie")	21	moins de 10 m ²	9,14%	< 10 m ²
	22	10 à 20 m ²	17,75%	10 à 20 m ²
	23	20 à 30 m ²	24,80%	20 à 30 m ²
	24	plus de 30 m ²	39,16%	> 30 m ²
	25	non réponse	9,14%	NR5

Tableau de Burt

	Seu l	Col oc	Co upl e	Par ents & NR	Cit é	Stu dio	Ap part	Cha mbr e	Aut re	NR 2	<= 1 an	1-3 ans	> 3 ans	NA & NR	- de 1k m	1 à 5 km	+5 km & NR	- de 10 ² m	10 à 20 ² m	20 à 30 ² m	+ de 30 ² m	NR 5
Seul	185	0	0	0	34	90	40	13	3	5	61	60	59	5	70	101	14	32	61	71	21	0
Colo	0	53	0	0	5	6	32	2	3	5	13	18	21	1	13	33	7	1	4	8	40	0
Coup	0	0	50	0	2	10	34	0	3	1	5	14	28	3	15	23	12	2	2	14	32	0
Par / NR	0	0	0	95	0	2	10	5	67	11	1	3	2	89	5	34	56	0	1	2	57	35
Cité	34	5	2	0	41	0	0	0	0	0	17	13	9	2	15	23	3	27	9	1	4	0
Stud	90	6	10	2	0	108	0	0	0	0	29	33	45	1	41	61	6	1	33	57	17	0
App	40	32	34	10	0	0	116	0	0	0	23	35	47	11	37	62	17	1	10	29	74	2
Cha	13	2	0	5	0	0	0	20	0	0	6	6	3	5	6	10	4	4	7	5	4	0
Autr	3	3	3	67	0	0	0	0	76	0	2	4	4	66	2	29	45	0	1	1	50	24
NR2	5	5	1	11	0	0	0	0	0	22	3	4	2	13	2	6	14	2	8	2	1	9
- de1	61	13	5	1	17	29	23	6	2	3	80	0	0	0	30	44	6	14	26	24	16	0
1-3	60	18	14	3	13	33	35	6	4	4	0	95	0	0	25	60	10	11	22	28	32	2
+de3	59	21	28	2	9	45	47	3	4	2	0	0	110	0	43	53	14	10	14	41	45	0
NA / NR	5	1	3	89	2	1	11	5	66	13	0	0	0	98	5	34	59	0	6	2	57	33
- 1k	70	13	15	5	15	41	37	6	2	2	30	25	43	5	103	0	0	12	26	38	26	1
1 à 5	101	33	23	34	23	61	62	10	29	6	44	60	53	34	0	191	0	20	35	47	82	7
+ 5k/NR	14	7	12	56	3	6	17	4	45	14	6	10	14	59	0	0	89	3	7	10	42	27
- 10	32	1	2	0	27	1	1	4	0	2	14	11	10	0	12	20	3	35	0	0	0	0
10-20	61	4	2	1	9	33	10	7	1	8	26	22	14	6	26	35	7	0	68	0	0	0
20-30	71	8	14	2	1	57	29	5	1	2	24	28	41	2	38	47	10	0	0	95	0	0
30+	21	40	32	57	4	17	74	4	50	1	16	32	45	57	26	82	42	0	0	0	150	0
NR5	0	0	0	35	0	0	2	0	24	9	0	2	0	33	1	7	27	0	0	0	0	35

- Le tableau de Burt comporte une ligne et une colonne pour chaque modalité des variables étudiées.
- Chaque cellule du tableau indique le nombre d'individus statistiques qui possèdent en même temps la modalité ligne et la modalité colonne correspondantes.

Calcul des Valeurs propres

- L'analyse du tableau de Burt produit au plus K-Q valeurs propres non nulles

- La somme des valeurs propres est égale à l'inertie totale, c'est-à-dire $\frac{K-Q}{Q}$

- la moyenne des valeurs propres

$$\lambda_m = \frac{1}{Q} = \frac{1}{Nb \text{ de questions}}$$

- La moyenne des valeurs propres (λ_m) est de 0,2. Les 6 premières valeurs propres, celles qui sont supérieures à cette moyenne, sont conservées.

- K est le nombre total de modalités
- Q est le nombre de questions

Nombre de Dims.	Valeurs Propres et Inertie de toutes les Dimensions (Etudiants-ville-2006.sta)			
	Inertie Totale = 3,40			
	ValSing.	ValProp.	%age Inertie	%age Cumulé
1	0,8494	0,7215	21,22	21,22
2	0,6351	0,4033	11,86	33,08
3	0,5734	0,3288	9,67	42,75
4	0,5019	0,2519	7,41	50,16
5	0,4859	0,2361	6,94	57,11
6	0,4588	0,2105	6,19	63,30
7	0,4455	0,1985	5,84	69,14
8	0,4318	0,1865	5,48	74,62
9	0,4133	0,1708	5,02	79,65
10	0,4084	0,1668	4,90	84,55
11	0,3521	0,1240	3,65	88,20
12	0,3323	0,1104	3,25	91,44
13	0,3229	0,1042	3,07	94,51
14	0,2733	0,0747	2,20	96,71
15	0,2352	0,0553	1,63	98,33
16	0,1961	0,0385	1,13	99,47
17	0,1348	0,0182	0,53	100,00

Calcul des Valeurs propres

- On ne conserve que les valeurs propres λ supérieures à λ_m et on calcule pour chacune d'entre elles :

$$\lambda' = (\lambda - \lambda_m)^2$$
- La méthode de Jean-Paul Benzécri vise à évaluer l'importance des valeurs propres en calculant des "taux modifiés" d'inertie expliquée.
- les taux modifiés représentent l'écart du nuage de points par rapport au nuage parfaitement sphérique qui serait obtenu si aucun lien n'existait entre les modalités.

$$\frac{\lambda'}{\sum \lambda'}$$

- Le taux d'inertie modifié moyen est de $100\%/6 = 16,7\%$. Seule la première valeur propre dépasse ce taux, mais une étude limitée seulement au premier axe principal présenterait peu d'intérêt. Nous étudierons donc les deux premiers axes (voire, éventuellement, le 3ème).

Nb de dim.	Val Prop.	$\lambda' = (\lambda - \lambda_m)^2$	Taux d'inertie modifié
1	0,7215	0,2720	81,43%
2	0,4033	0,0413	12,37%
3	0,3288	0,0166	4,97%
4	0,2519	0,0027	0,81%
5	0,2361	0,0013	0,39%
6	0,2105	0,0001	0,03%

Résultats relatifs aux modalités

- le tableau fournit des informations sur les coordonnées, la contribution (inertie relative), la qualité de représentation (cos2) de chaque modalité, ainsi que le poids de chaque modalité et sa contribution à la formation de l'inertie totale.
- On retrouve dans ce tableau l'inertie relative de chaque variable, comme somme des inerties relatives des modalités qui la compose. Par exemple, pour la première variable :

$$\frac{I(X_1)}{I} = \frac{0,6}{3,4} = 0,0304 + 0,0507 + 0,0511 + 0,0442 = 0,1765$$

- L'interprétation devrait se concentrer sur les modalités ayant les meilleures qualités de représentation selon chaque axe factoriel, en utilisant les colonnes "Cosinus2" ou "Qualité".

	Ligne	Coord. dim 1	Coord. dim 2	Coord. dim 3	Masse	Qualité	Inertie	Inertie dim 1	Cos ² dim 1	Inertie dim 2	Cos ² dim 2	Inertie dim 3	Cos ² dim 3
MODE:Seul	1	-0,6921	0,5251	-0,2405	0,0966	0,7592	0,0304	0,0641	0,4475	0,0661	0,2576	0,0170	0,0540
MODE:Coloc	2	-0,2275	-1,0201	0,7869	0,0277	0,2749	0,0507	0,0020	0,0083	0,0714	0,1671	0,0521	0,0995
MODE:Couple	3	-0,2219	-1,3155	0,2448	0,0261	0,2762	0,0511	0,0018	0,0074	0,1120	0,2598	0,0048	0,0090
MODE:Parents et NR	4	1,5914	0,2388	-0,0995	0,0496	0,8575	0,0442	0,1741	0,8354	0,0070	0,0188	0,0015	0,0033
TYPE:Cité	5	-0,7505	1,4395	1,9098	0,0214	0,7532	0,0525	0,0167	0,0675	0,1100	0,2484	0,2375	0,4373
TYPE:Studio	6	-0,7176	0,1556	-1,0249	0,0564	0,6243	0,0422	0,0403	0,2022	0,0034	0,0095	0,1802	0,4125
TYPE:Appart	7	-0,2390	-1,0433	0,3581	0,0606	0,5534	0,0410	0,0048	0,0248	0,1635	0,4729	0,0236	0,0557
TYPE:Chambre	8	-0,2358	0,7976	0,0280	0,0104	0,0382	0,0558	0,0008	0,0031	0,0165	0,0351	0,0000	0,0000
TYPE:Autre	9	1,5850	0,1322	-0,0137	0,0397	0,6263	0,0472	0,1382	0,6219	0,0017	0,0043	0,0000	0,0000
TYPE:NR2	10	0,9207	0,8725	-0,3942	0,0115	0,1075	0,0554	0,0135	0,0517	0,0217	0,0464	0,0054	0,0095
ANC:< 1 an	11	-0,6570	0,5958	0,1375	0,0418	0,2127	0,0465	0,0250	0,1140	0,0368	0,0937	0,0024	0,0050
ANC:1-3 ans	12	-0,4743	-0,0604	0,1251	0,0496	0,0806	0,0442	0,0155	0,0742	0,0004	0,0012	0,0024	0,0052
ANC:> 3ans	13	-0,4839	-0,6110	-0,1348	0,0574	0,2521	0,0419	0,0186	0,0944	0,0532	0,1504	0,0032	0,0073
ANC:NA et NR	14	1,5393	0,2579	-0,0823	0,0512	0,8400	0,0438	0,1681	0,8148	0,0084	0,0229	0,0011	0,0023
ELOIGN:< 1km	15	-0,6523	0,0477	-0,2336	0,0538	0,1774	0,0430	0,0317	0,1565	0,0003	0,0008	0,0089	0,0201
ELOIGN:1 à 5 km	16	-0,2129	-0,0935	0,1793	0,0997	0,0858	0,0295	0,0063	0,0451	0,0022	0,0087	0,0097	0,0320
ELOIGN:>5 km - NR	17	1,2118	0,1454	-0,1143	0,0465	0,4549	0,0452	0,0946	0,4445	0,0024	0,0064	0,0018	0,0040
SUP:< 10 m²	18	-0,7762	1,5959	2,0488	0,0183	0,7389	0,0534	0,0153	0,0606	0,1154	0,2561	0,2333	0,4222
SUP:10 à 20 m²	19	-0,6118	0,7532	-0,5435	0,0355	0,2670	0,0484	0,0184	0,0808	0,0499	0,1225	0,0319	0,0638
SUP:20 à 30 m²	20	-0,6689	-0,1651	-0,8958	0,0496	0,4213	0,0442	0,0308	0,1476	0,0034	0,0090	0,1210	0,2647
SUP:> 30 m²	21	0,4169	-0,7995	0,4513	0,0783	0,6545	0,0358	0,0189	0,1119	0,1241	0,4115	0,0485	0,1311
SUP:NR5	22	1,9938	0,8154	-0,4956	0,0183	0,4914	0,0534	0,1007	0,3998	0,0301	0,0669	0,0136	0,0247

Etude des variables

==> Contributions des variables à l'inertie des axes

• Nous savons que les contributions des variables à la formation de l'inertie du nuage dépendent essentiellement du nombre de leurs modalités. On peut cependant comparer leur contribution à l'inertie d'un axe à leur contribution à l'inertie du nuage, ce qui donne une idée de l'importance prise par chacune d'elles dans la formation des axes. Par exemple, nous obtenons ici :

Variable	Contribution à l'inertie du nuage	Contribution à l'inertie de l'axe 1	Contribution à l'inertie de l'axe 2	Contribution à l'inertie de l'axe 3
Mode d'occupation	0,1765	0,2420	0,2565	0,0754
Type de logement	0,2941	0,2142	0,3168	0,4467
Ancienneté	0,1765	0,2272	0,0988	0,0090
Eloignement	0,1176	0,1326	0,0049	0,0205
Superficie	0,2353	0,1840	0,3230	0,4484

- On voit sur ce tableau que la part des variables "Mode d'occupation" et "Ancienneté" dans la formation du premier axe est supérieure à leur part dans l'inertie totale du nuage.

Etude des variables

Constitution du tableau précédent :

- Dans le tableau des résultats relatifs aux modalités, on additionne les inerties des différentes modalités d'une même variable, pour le nuage entier d'une part, et pour chacun des axes retenus d'autre part. Par exemple, pour la première variable :

	Inertie	Inertie dim 1	Inertie dim 2	Inertie dim 3
MODE:Seul	0,0304	0,0641	0,0661	0,0170
MODE:Coloc	0,0507	0,0020	0,0714	0,0521
MODE:Couple	0,0511	0,0018	0,1120	0,0048
MODE:Parents et NR	0,0442	0,1741	0,0070	0,0015
Total	0,1764	0,2420	0,2565	0,0754

- Cette démarche permet de quantifier l'inertie globale associée à chaque variable dans le nuage et dans chaque dimension, ce qui est essentiel pour comprendre l'impact de chaque variable sur la formation des axes

Etude des axes

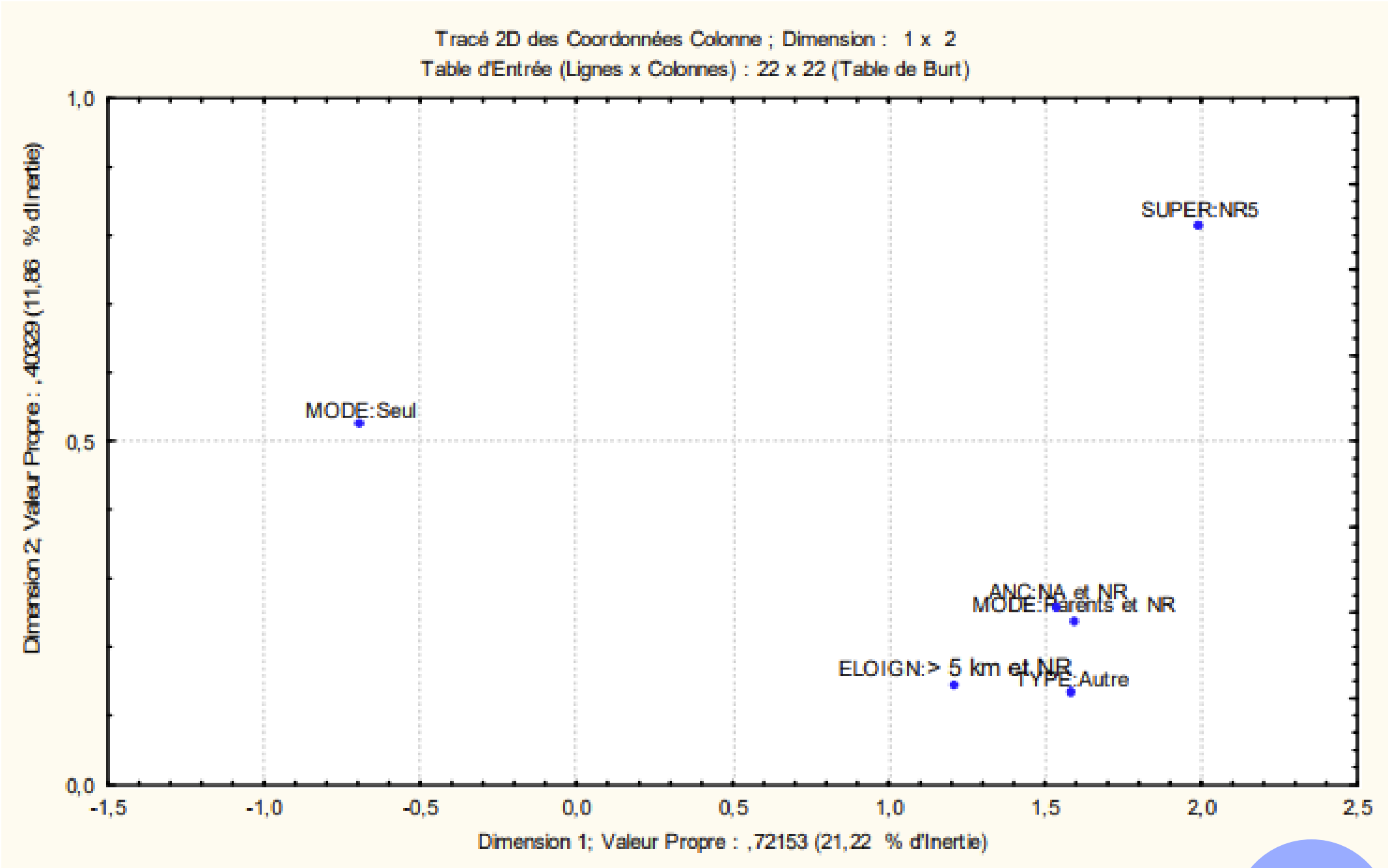
-	+
MODE: Seul (6,41%)	MODE: Parents / NR (17,4%) ANC: NA et NR (16,8%) TYPE: Autre (13,8%) SUPER: NR5 (10,1%) ELOIGN:>5km (9,5%)

- Tableau :Contributions des Modalités au Premier Axe

Premier axe

Interprétations :

- Les modalités "NA", "NR", "Autre" et "Parents" de la variable "MODE" ont des contributions élevées, suggérant une distinction entre les étudiants vivant au foyer familial et ceux avec un logement indépendant.



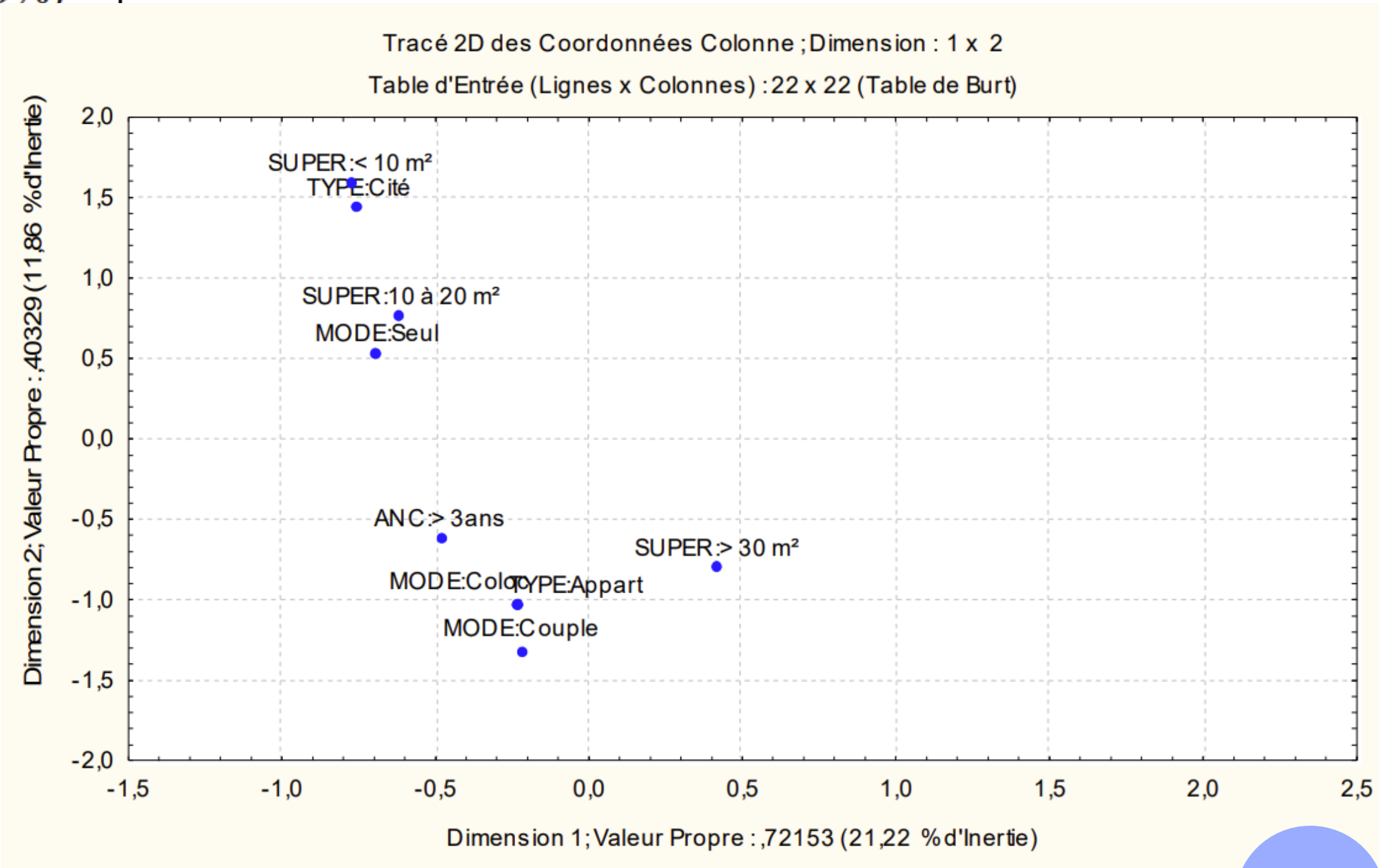
Etude des axes

-	+
TYPE: Appart. (16,4%)	SUPER: < 10m ² (11,5%)
SUPER: > 30 m ² (12,4%)	TYPE: Cité (11%)
MODE: Couple (11,2%)	MODE: Seul (6,6%)
MODE: Coloc. (7,1%)	SUPER: 10 à 20 m ² (4,99%)
ANC: >3ans (5,3%)	

Interprétations :

- L'axe se focalise sur les variables TYPE, SUPERFICIE, et MODE, principalement liées aux étudiants vivant hors du foyer familial.
- Il oppose les étudiants en cité universitaire, vivant seuls dans des logements de petite taille, à ceux en couple ou en colocation, habitant des appartements plus spacieux.

Second axe



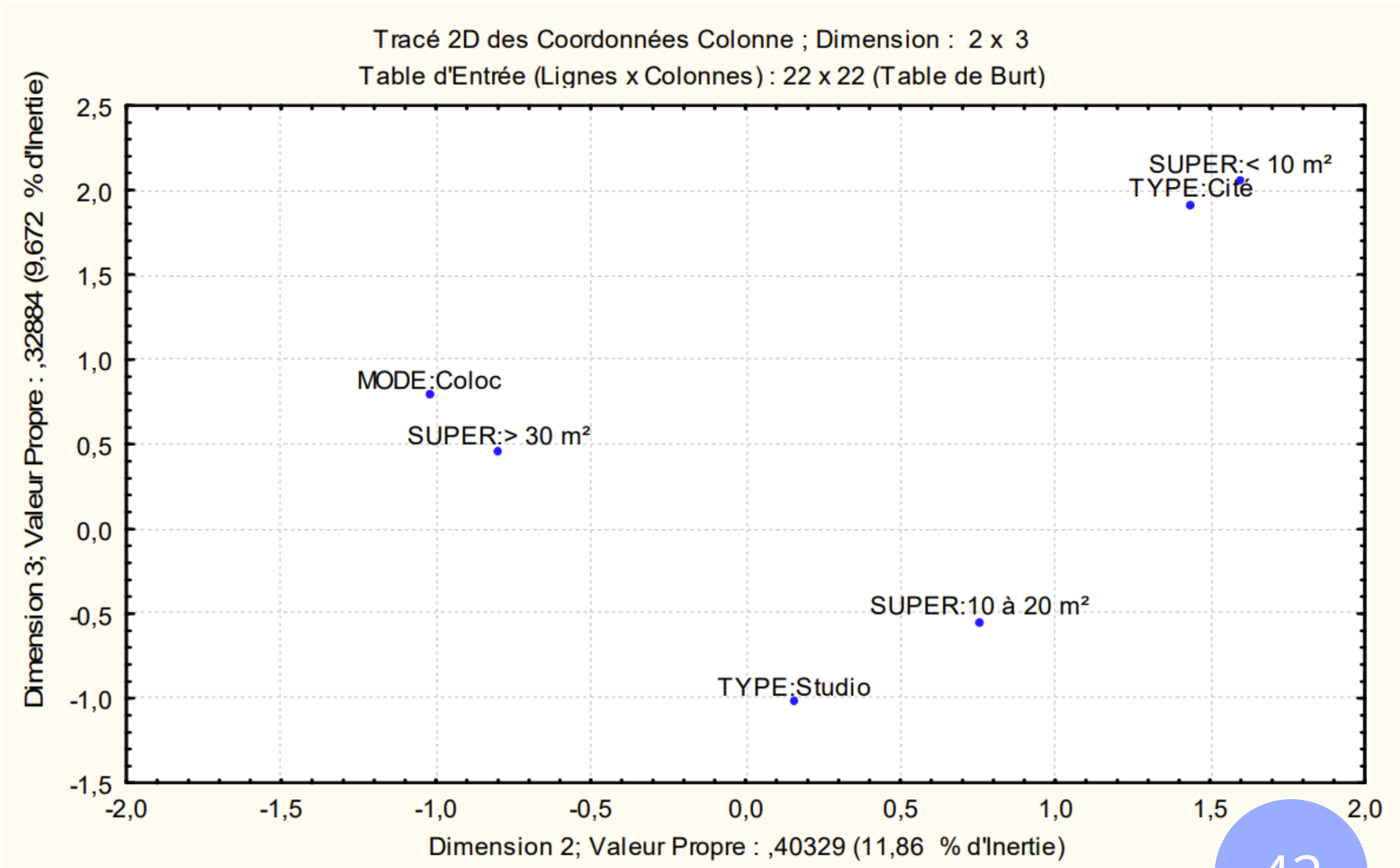
Etude des axes

-	+
TYPE: Studio (18,0%) SUPER: 20 à 30m ² (12,1%)	TYPE: Cité (23,8%) SUPER: <10m ² (23,3%) MODE: Coloc (5,2%) SUPER: >30m ² (4,85%)

Troisième axe

Interprétations :

- C'est essentiellement la superficie, et le type de logement correspondant, qui interviennent ici : logement en cité universitaire, de moins de 10 m2, studio de 20 à 30 m2 et colocation (plutôt en appartement) de plus de 30 m2. Encore une fois, les étudiants habitant au domicile familial interviennent peu dans la formation de cet axe.



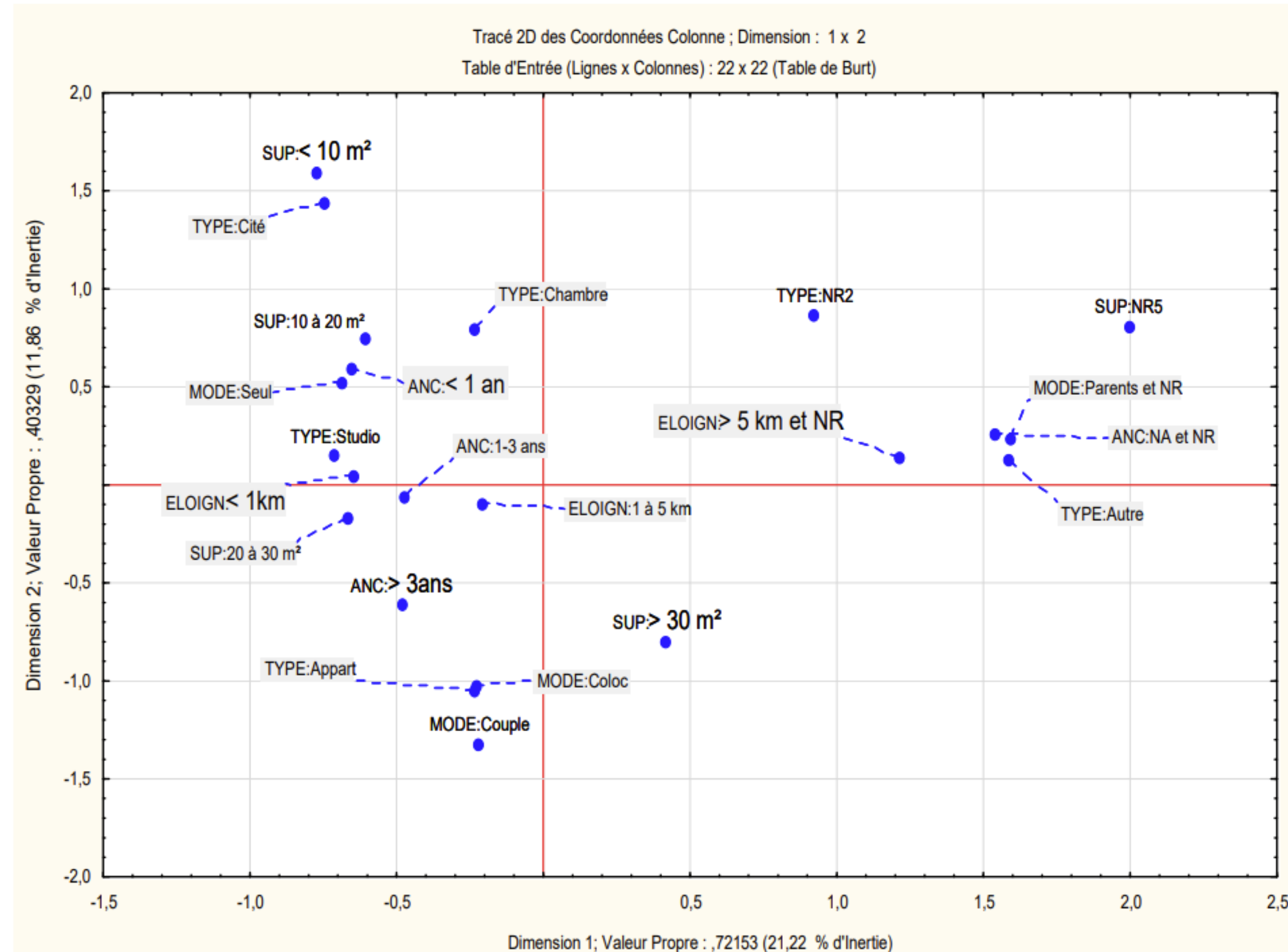
Graphiques selon les deux premiers axes principaux

Interprétations :

==>Trois groupes distincts dans l'étude :
logement chez les parents (ou non
réponse), étudiants solitaires en
cité/chambre, et étudiants en
couple/colocation.

==>Les étudiants en studio forment un
groupe intermédiaire.

==>Les variables "Eloignement" et
"Ancienneté" ont un impact modeste,
avec des tendances spécifiques liant
l'ancienneté et l'éloignement aux types de
logements.



MERCI!