

Introduction à l'ACM spécifique

❑ ***Quelques rappels en guise de préambule ?***

Eléments préliminaires et ACM standard

❑ ***La question des non-réponses et autres modalités rares en ACM : les limites de l'ACM standard***

- Le problème auquel nous sommes confrontés et les solutions habituellement adoptées
- Exemple à partir de l'enquête *ESS* : « Tolérances de la fraude et relations de confiance en Europe » (Lascoumes, Le Hay, 2010)
- Exemple à partir de l'enquête *Médiapolis (Cevipof-Carism)* : une écologie des pratiques médiatiques et informationnelles (Le Hay, Vedel, Chanvril, 2011)

❑ ***Une solution : l'ACM spécifique***

- La théorie de l'ACM spécifique : éléments mathématiques
- Application à partir de l'enquête *Médiapolis*

❑ ***Mise en œuvre à partir de l'enquête Médiapolis***

Éléments préliminaires

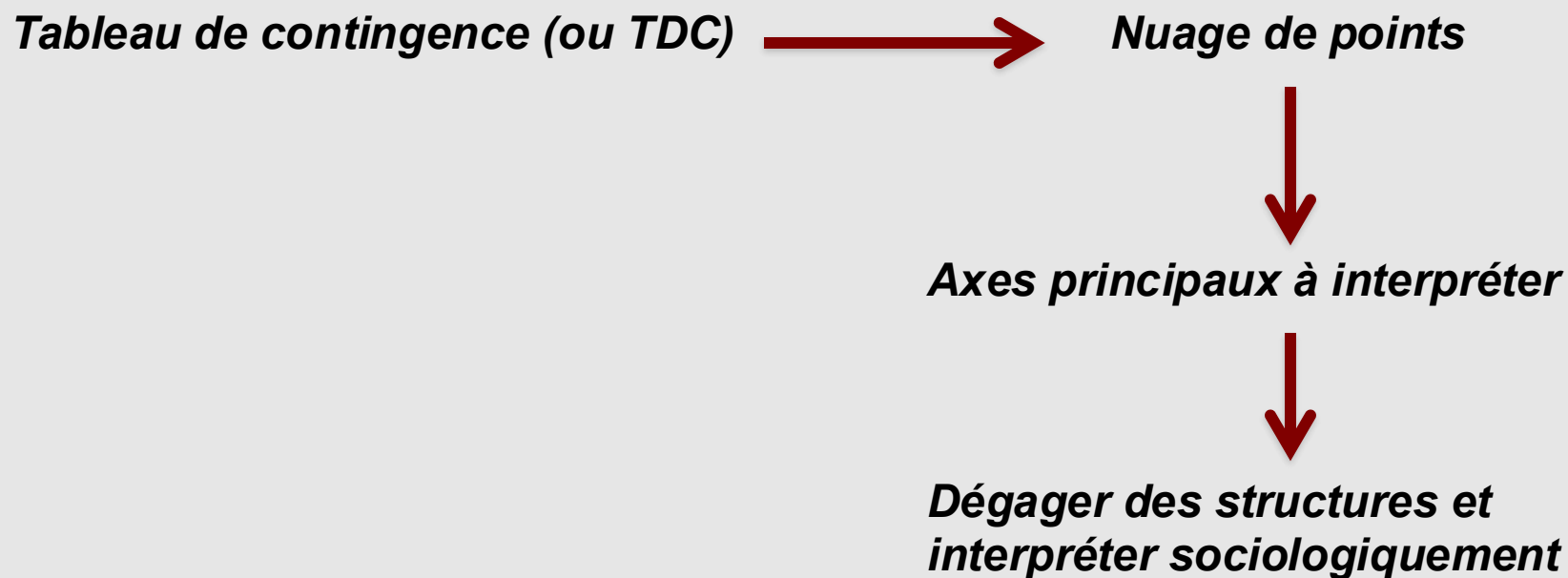
*Les grands principes
de l'analyse géométrique
et de l'ACM standard*

Tableau de contingence (tableau croisé en effectifs) : Analyse factorielle des correspondances (AFC)

Variables quantitatives : Analyse en composantes principales (ACP)

Variables qualitatives : Analyse des correspondances multiples (ACM)

Principes analytiques



Références Théoriques

- Benzécri J.P., Benzécri F., 1980, *Pratique de l'analyse des données: Analyse des correspondances, exposé élémentaire*, Paris, Dunod
- Chiche J., Le Roux B., 2006, « *Développements récents en analyse des correspondances multiples* », *Modulad*, n° XX
- Cibois P., *Les méthodes d'analyse d'enquêtes*, Paris, Puf, 2007.
- Lebart L., Morineau A., Piron M., 2000, *Statistique exploratoire multidimensionnelle*, Paris, Dunod
- Lemerrier C., Zalc C., 2010, *Méthodes quantitatives pour l'historien*, Paris, La Découverte
- Le Roux B., Rouanet H., 2010, *Multiple correspondance analysis*, Sage, Series: Quantitative applications in the social sciences, vol. 163, CA: Thousand Oaks, 115 p.
- Le Roux B., Rouanet H., 2004, *Geometric Data analysis: From correspondance analysis to structured data analysis*, Kluwer academic publishers
- Rouanet H., Le Roux B., 1993, *Analyse des données multidimensionnelles*, Paris, Dunod
- Saporta G., 1990, *Probabilités, analyse des données et statistiques*, technip

En quoi consiste une analyse géométrique ?

→ *C'est un changement de repère orthonormé qui s'appuie sur la direction du plus grand allongement du nuage de points initial*

Les droites obtenues (Axe 1, Axe 2, etc...), orthogonales entre elles, deviennent le nouveau repère.

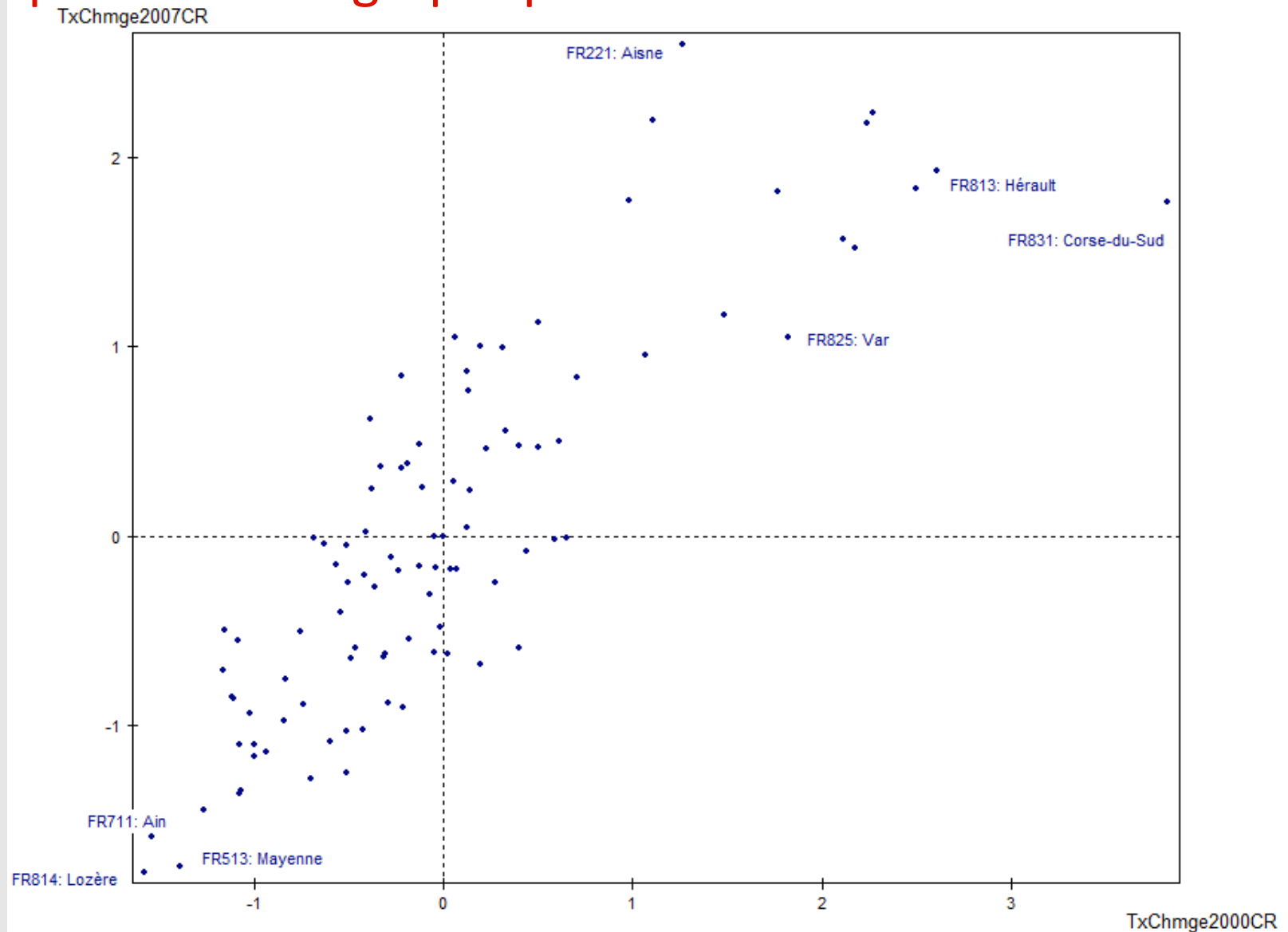
*Ces droites sont appelées **droites principales**.*

Propriétés de ce nouveau repère dans n dimensions :

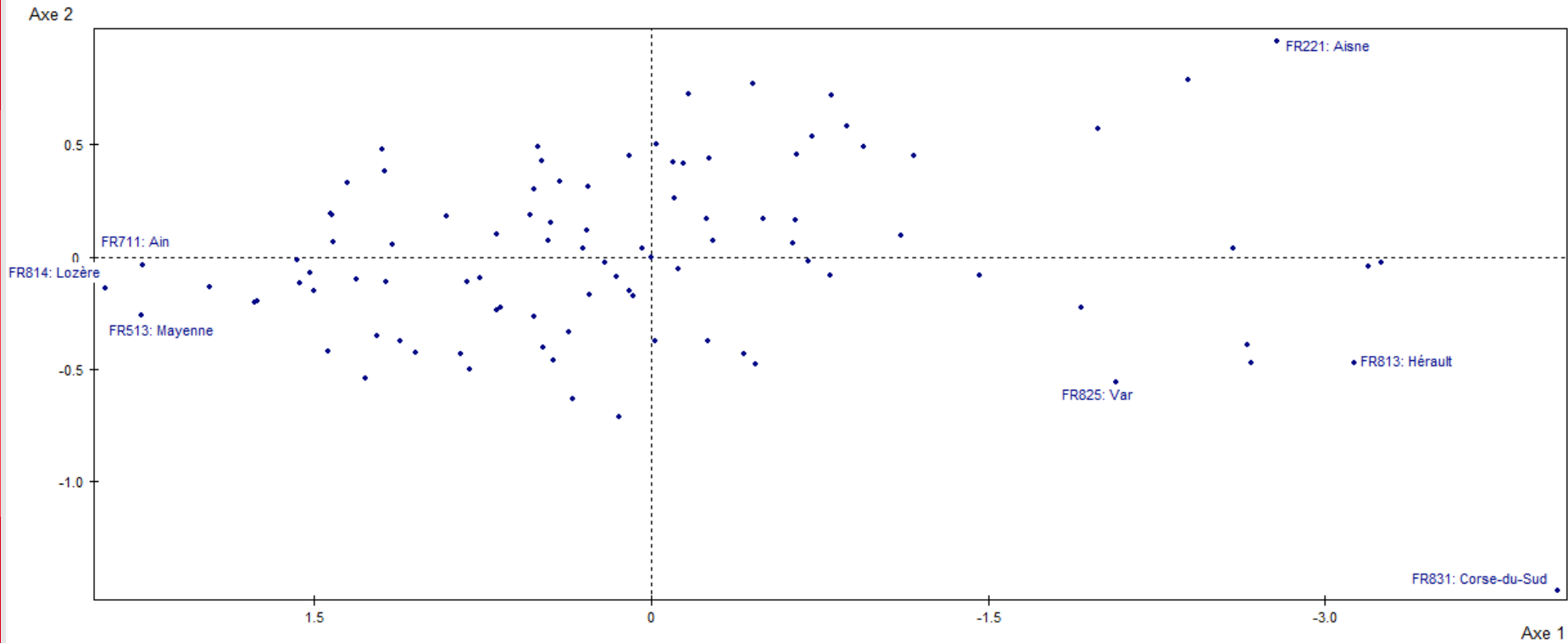
Dans n dimensions : il s'agit du nuage le moins déformant, le plus fidèle au nuage initial

Dans n dimensions : la meilleure synthèse, puisqu'elle limite la perte d'information

Représentation graphique des données centrées réduites



ACP des variables



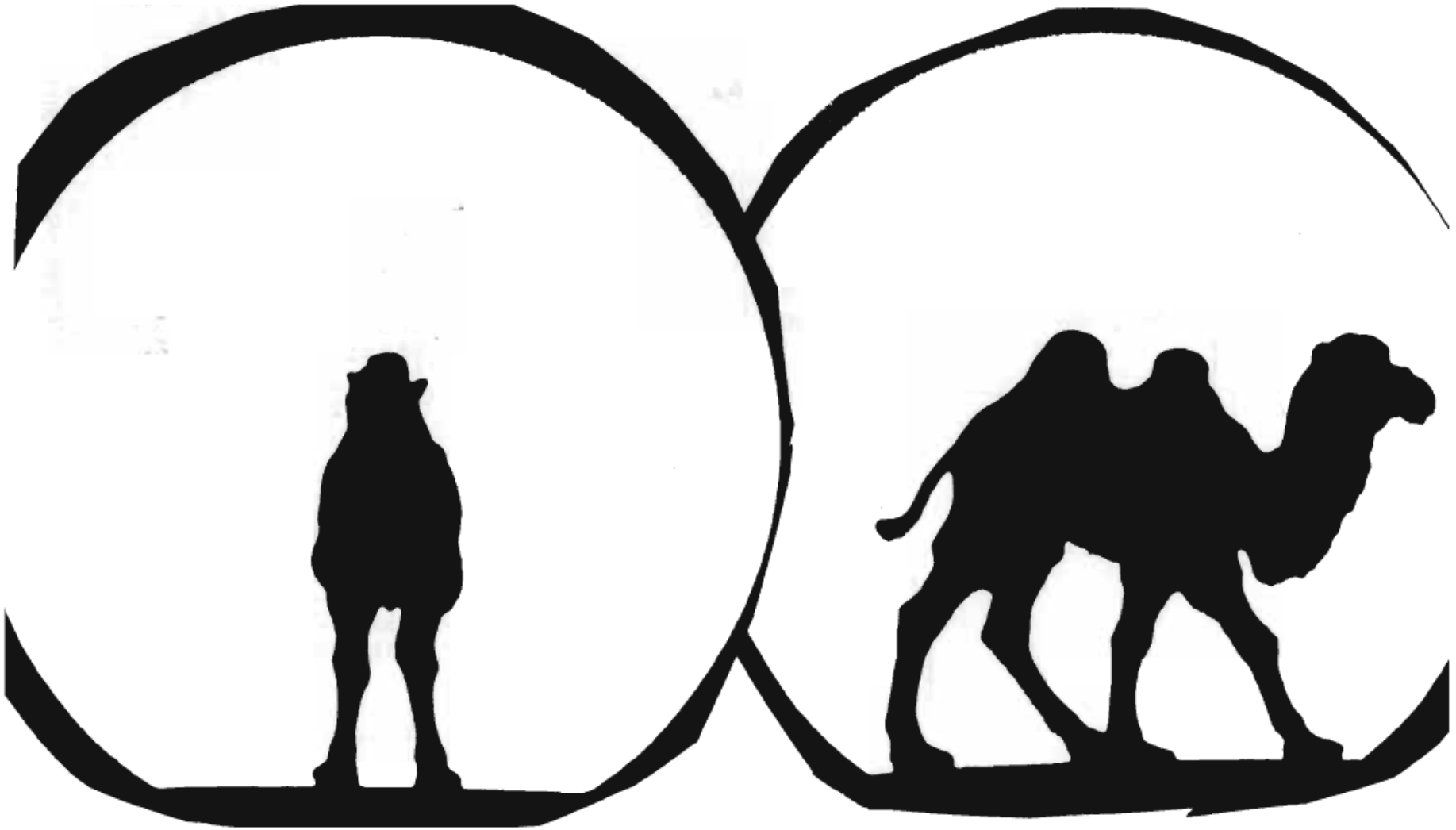
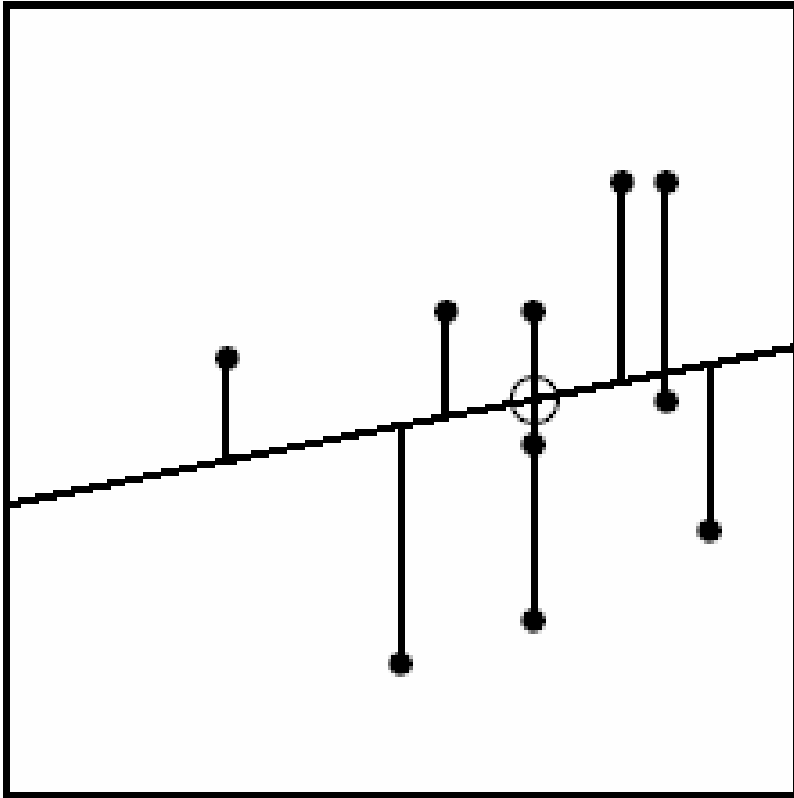


FIGURE: Quel animal ? (*illustration JP Fénélon*)

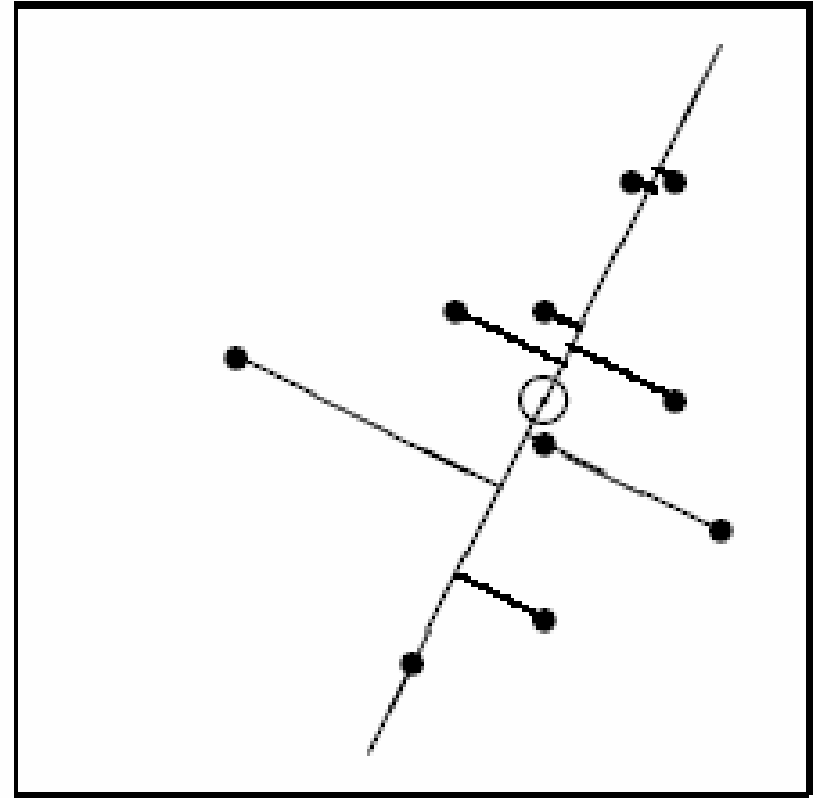


Droite de régression

passé par le point moyen G

Moindres carrés ordinaires

**(on conserve le repère
orthonormé de départ)**



Droite principale

passé par le point moyen G

Moindres carrés orthogonaux

**(on change de repère mais on
ajuste au plus près le nuage
initial)**

Principe géométrique de l'ACM :

- ❑ *Construire un nuage euclidien de points représentant les individus*
- ❑ *Interpréter des axes : déterminer des axes principaux et des variables principales*

Étape fondamentale :

- ❑ *Définir les distances entre individus sur la base de leurs réponses aux questions actives*

Le calcul de la distance entre individus dans l'ACM standard

Soit :

- ❑ $d_q(i, i')$ la distance entre les individus i et i' à la question q
- ❑ n_{κ} le nombre d'individus ayant choisi le modalité κ à la question q
- ❑ $n_{\kappa'}$ le nombre d'individus ayant choisi la modalité κ' à la question q

Propriété fondamentale de l'ACM :

la distance entre deux individus i et i' ne dépend que des questions de désaccord.

➔ *Par conséquent :*

si q est une question d'accord entre i et i' : $d_q(i, i') = 0$

Variance du nuage	$V = \frac{K}{Q} - 1$
Nombre de valeurs propres (nb de dim°) (non trivialement nulles)	$L = K - Q$
Moyenne des valeurs propres	$\bar{\lambda} = \frac{V}{K - Q} = \frac{1}{Q}$ <p><i>Variance totale = somme des valeurs propres</i></p>

Propriétés des
coordonnées de
l'individu l

$$\text{mean } y_l^I = \sum_{i \in I} \frac{y_l^i}{n} = 0$$
$$\text{Var } y_l^I = \sum_{i \in I} \frac{(y_l^i)^2}{n} = \lambda_l$$

Propriétés des
coordonnées de la
modalité k (pour toute
question q)

$$\sum_{k \in K} \frac{f_k}{Q} y_l^k = 0$$

Coordonnée du point
moyen modalité k

$$\sum_{k \in K} \frac{f_k}{Q} (y_l^k)^2 = \lambda_l$$
$$\bar{y}_l^k = \sum_{i \in I_k} \frac{y_l^i}{n_k} = \sqrt{\lambda_l} \times y_l^k$$

*Les limites de l'ACM standard :
la question des non-réponses
et autres modalités rares (peu choisies)*

Problème de l'ACM standard : les modalités à faibles effectifs (<5%)

- Modalités très éloignées du centre du nuage
- Contribution très forte à la variance des questions
- ➔ Très grande influence dans la détermination des axes factoriels

Les solutions généralement adoptées

- Regroupement de modalités (pas toujours satisfaisant, que faire des NR ?)
- Non prise en compte (suppression) des individus concernés par une NR
- ➔ affecte la représentativité de l'échantillon

Fin des années 1990, une solution : Ignorer les modalités rares dans le calcul des distances entre individus → l'ACM spécifique

- Réflexion sur le moyen de conserver ces individus et de reconstruire l'inertie sans toucher à la structure d'ensemble des nuages euclidiens.

Brigitte Le Roux, Jean Chiche, Henry Rouanet

- 1998 : Bases mathématiques (Le Roux, Chiche, « Analyse spécifique d'un questionnaire: cas particulier des non-réponses », XXXèmes journées de statistique de la SFDS, Rennes, mai 1998) puis programmation

- milieu des années 2000 : Spad intègre le programme de l'ACM spécifique à son logiciel

(<http://www.coheris.fr/fr/page/produits/SPAD-data-mining.html>)

- 2014 : un package pour R est élaboré par Nicolas Robette (GDATools)

Un premier exemple

* LASCOUMES (Pierre) et LE HAY (Viviane), « Tolérances de la fraude et relations de confiance », Daniel Boy, Bruno Cautrès et Nicolas Sauger (dir.), *Les Français : des Européens comme les autres ?*, Paris, Presses de Sciences Po, Académique, 2010, pp. 73-108. [ISBN 978-2-7246-1161-8]

Analyse à partir de la 2^e vague d'ESS (European Social Survey), 15 pays
Module rotatif :
«Probité économique : société de marché et citoyenneté »

Contribution à ce questionnement par une analyse empirique à partir de l'enquête ESS

- ❑ 2^e vague de l'ESS : module spécifique « probité économique, société de marché et citoyenneté »
- ❑ 15 pays
- ❑ Variables qui ne révèlent pas une trop grande polarisation des opinions et comportements (variables *clivantes*)

Objectif empirique

- ❑ Tester l'existence de relations entre les attitudes à l'égard des normes et des déviances et les rapports de confiance
- ❑ Éprouver une éventuelle diversité de ces rapports en Europe

Hypothèses sous-jacentes

SH1 - Diversité des rapports à la loi

(Becker, 1963 ; D. Mazda, 1969 ; Ogien 1995)

➔ des profils nationaux existent derrière une possible position de principe favorable au légalisme

SH2 – La confiance n'est pas unidimensionnelle

Confiance
privée

Confiance publique
(administrative et
politique)

2 profils non dissonants conceptuellement :

1/ Contexte social normé : importance attachée au respect des règles & confiance générale plus élevée

2/ Contexte social mal ou peu normé : importance des règles minimisée, tolérance vis-à-vis des déviations & défiance générale élevée.

Questions : dans quelle mesure le rigorisme s'accompagne-t-il de confiance ? Jusqu'à quel point la perception d'une relativité des normes sociales s'accompagne-t-elle de défiance ?

Les 12 questions prises en compte dans l'analyse

7 variables relatives aux dimensions de la confiance

Confiance privée :

- ❑ Artisans
- ❑ Institutions financières

Confiance publique :

Administration :

- ❑ Police
- ❑ Justice
- ❑ Fonctionnaires

Politique :

- ❑ Hommes politiques
- ❑ Partis politiques

5 variables relatives aux règles et aux lois (rigorisme/laxisme)

Jugements généraux :

- ❑ « Si on veut vraiment gagner de l'argent, on ne peut pas toujours se conduire honnêtement » (*échelle d'accord*)
- ❑ « On devrait toujours respecter strictement la loi, même si cela conduit à rater de bonnes occasions » (*échelle d'accord*)
- ❑ « De temps à autre, on peut se permettre d'ignorer la loi et faire ce que l'on veut » (*échelle d'accord*)

Jugements spécifiques :

- ❑ « Les citoyens ne devraient pas frauder le fisc » (*échelle d'accord*)
- ❑ « Payer quelqu'un en liquide sans facture pour éviter de payer la TVA ou d'autres taxes » (*condamnable ou non*)

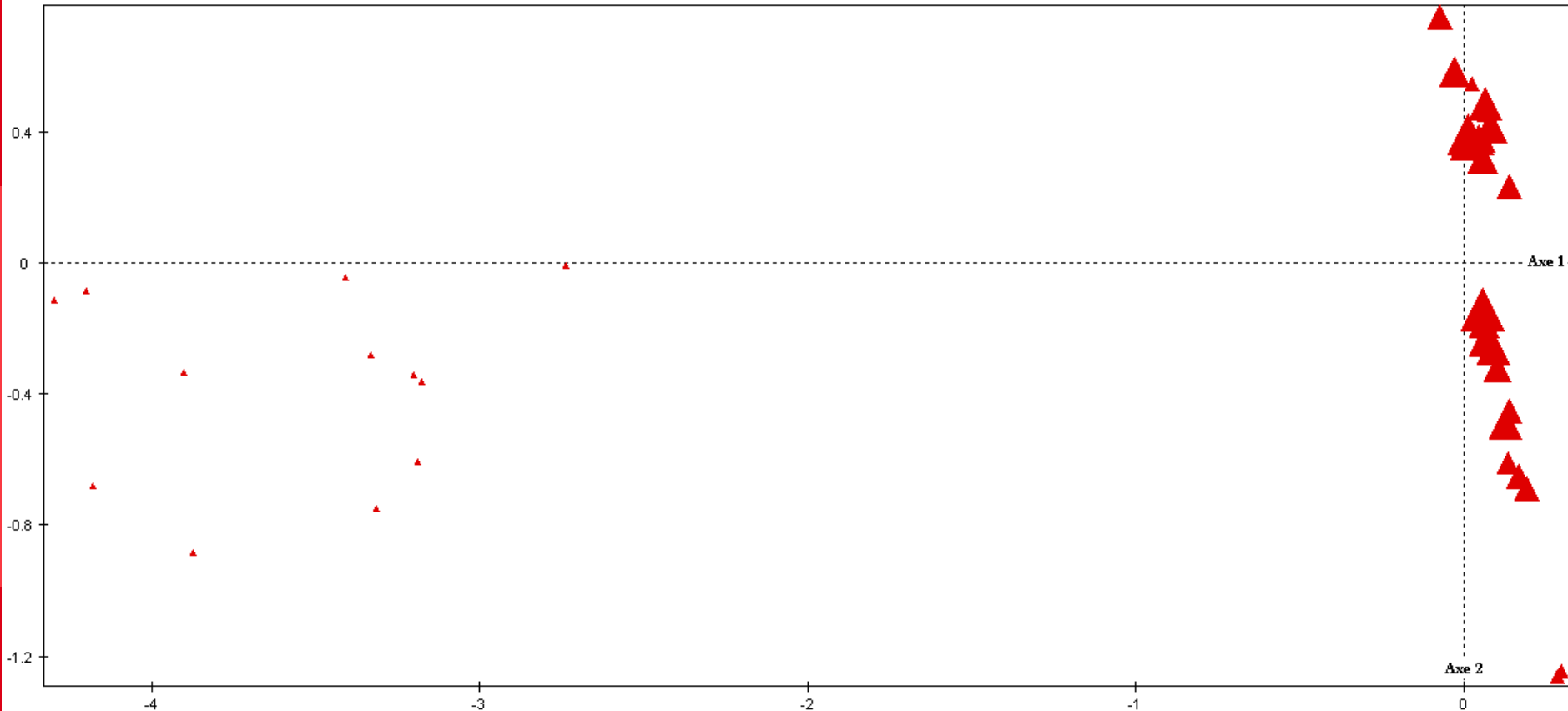
Pourcentages bruts de non-réponses aux 12 questions

Confiance dans la justice	2,2
Confiance dans la police	1,1
Confiance dans les fonctionnaires	3,2
Confiance dans les hommes politiques	1,9
Confiance dans les partis politiques	2,4
Confiance dans les plombiers etc.	2,6
Confiance dans les institutions financières	3,0
Les citoyens ne devraient pas frauder le fisc	1,5
Si on veut gagner de l'argent on ne peut pas toujours se comporter honnêtement	2,0
On devrait toujours respecter strictement la loi, même si cela conduit à rater de bonnes occasions	1,6
De temps à autre on peut se permettre d'ignorer la loi	1,8
Condamnable : payer quelqu'un en liquide pour éviter de payer la TVA	3,4

Les résultats issus d'une ACM standard

ACM standard qui conserve les modalités de non-réponses

Nuage des modalités dans le plan 1-2



Problème : les modalités rares sont éloignées du centre du nuage.
Trop forte contribution à la variance de leur question et trop forte influence dans la détermination des axes

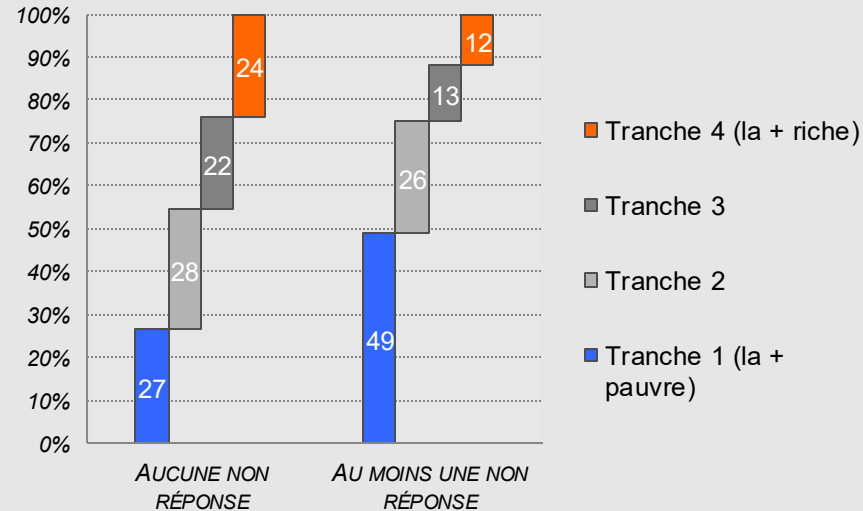
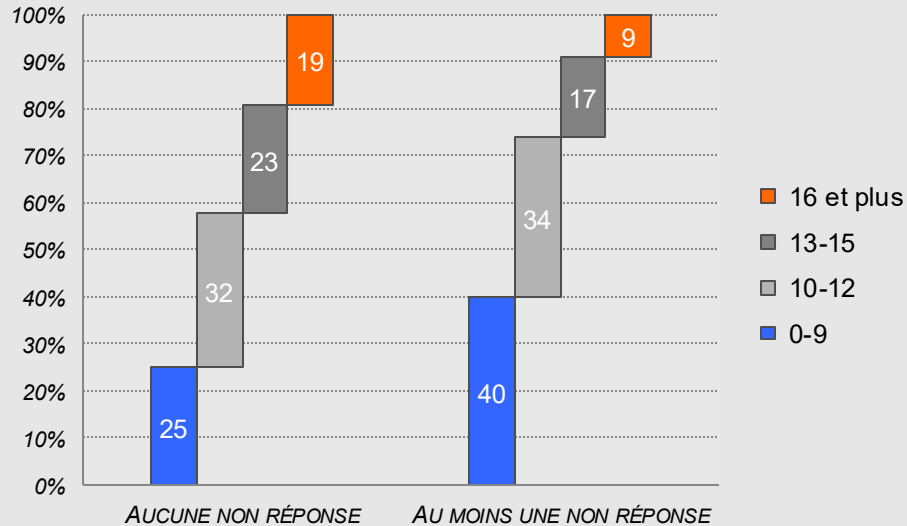
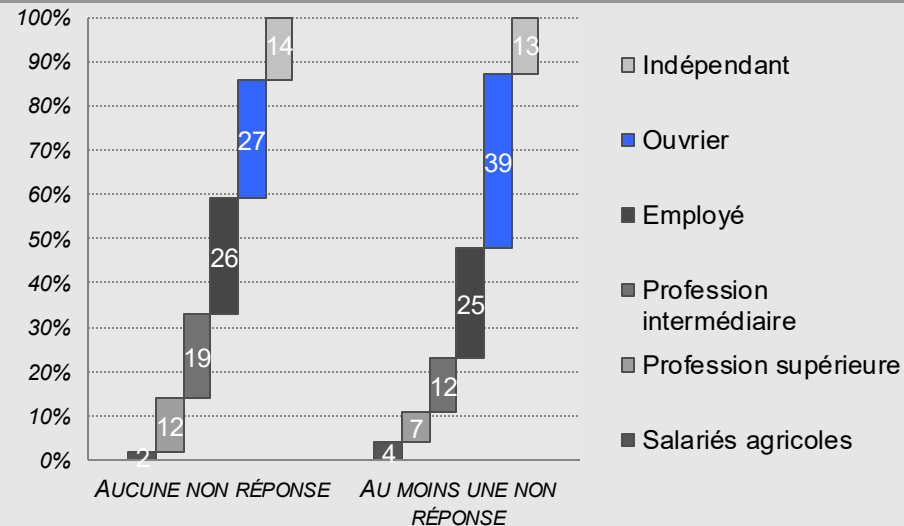
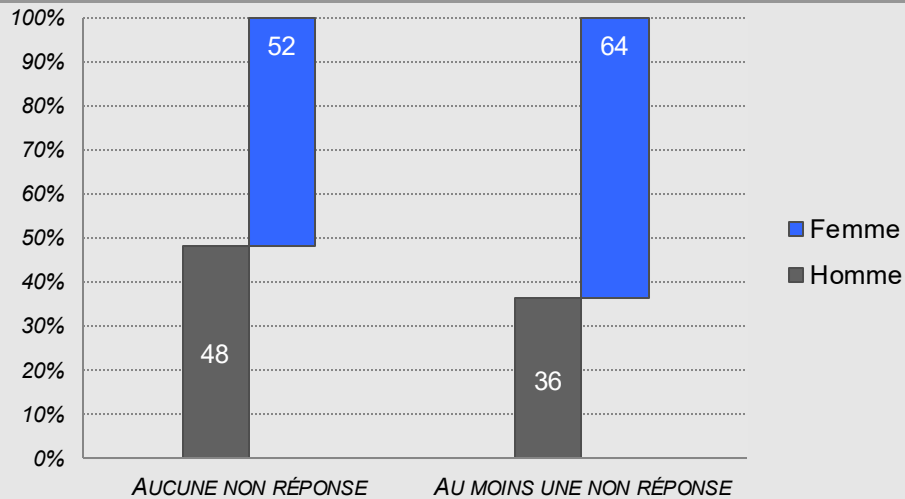
Occurrences des non-réponses		
	Effectif	Pourcentage
Aucune	25 509	87,3
Une	2 000	6,8
Deux	669	2,3
Trois	462	1,6
Quatre à six	452	1,5
Sept à douze	123	0,4
Total	29 215	100

La réalisation d'une ACM « classique » contraint à ne pas tenir compte de 13% de l'échantillon initial

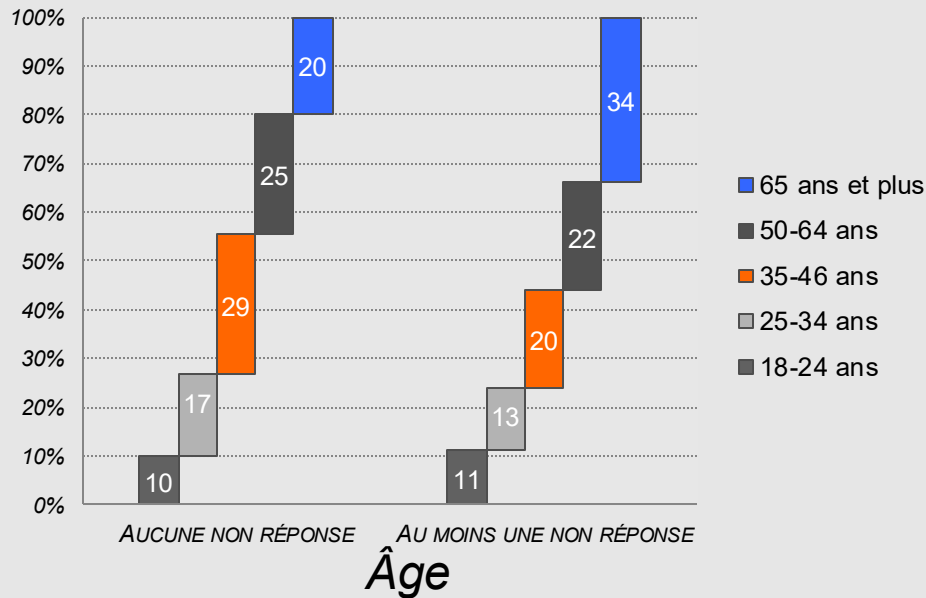
→ Quid de la représentativité d'un tel échantillon ?

En outre, une partie non négligeable de ce sous-échantillon (13%) a répondu à un grand nombre des questions

Le problème des non-réponses : exemple ESS



Le problème des non-réponses : exemple ESS



➔ Conclusion : le fait de supprimer les individus concernés par une non réponse biaise considérablement l'échantillon, qui n'est plus représentatif

Un second exemple : L'enquête Médiapolis (2009)

* LE HAY (Viviane), Vedel (Thierry), Chanvrlil (Flora), « usage des médias et politique : une écologie des pratiques informationnelles », *Réseaux*, oct.-nov. 2011, n° 170, p. 45-73

Projet ANR mené conjointement par le Cevipof (Centre de recherches politiques de Sciences Po) et le Carism (Centre d'analyse et de recherche interdisciplinaire sur les médias)

Les grandes questions du projet Médiapolis

- ❑ Comprendre les comportements d'information politique dans un univers changeant
- ❑ Identifier les nouvelles formes numériques de la citoyenneté, comprendre le rôle d'internet dans les nouvelles pratiques politiques, établir dans quelle mesure l'internet est source d'inégalités informationnelles et politiques
- ❑ Comprendre l'articulation entre pratiques politiques et pratiques informationnelles :
Les fractures civique, informationnelle et numérique se cumulent-elles et s'organisent-elles selon des logiques communes?
- ➔ Combinaison d'une démarche qualitative (entretiens, focus group, observation ethnographique) et d'une enquête quantitative

Présentation de l'enquête par questionnaire

- ❑ Enquête quantitative par sondage téléphonique sur un échantillon de 1754 personnes représentatif de la population française de 15 ans et plus (23 novembre au 12 décembre 2009)
- ❑ Méthode des quotas (genre, âge, profession ou ancienne profession du chef de ménage, niveau de diplôme) après stratification par régions administratives et catégories d'agglomération.

L'objet de l'ACM

Quels sont les grands profils de pratiques informationnelles en France aujourd'hui ?

La nature de ces profils s'articule-t-elle autour de logiques dans la structuration de leurs réseaux sociaux ?

Les variables de l'analyse : 7 variables actives (39 modalités actives et 5 modalités passives)

1- Le média utilisé le plus souvent en premier :

« Pour vous informer en matière politique, quel média utilisez-vous le plus souvent **en premier** ? »

		(<i>Etiquette graphique</i>)
La télévision	49.8%	(<i>Télévision1</i>)
La radio	20.8%	(<i>Radio1</i>)
Internet	12.4%	(<i>Internet1</i>)
La presse écrite nationale	8%	(<i>PresseNationale1</i>)
La presse écrite régionale	7.9%	(<i>PresseRégionale1</i>)
*Autre réponse	1.1%	(<i>*Media1_Autre</i>)

(L'item « La presse gratuite » a été remplacé par le second choix correspondant)

* signifie que la modalité est passive dans l'analyse

Les variables actives de l'analyse

2- Le média utilisé le plus souvent en second :

« Pour vous informer en matière politique, quel média utilisez-vous le plus souvent **en second** ? »

<i>(Etiquette graphique)</i>		
La télévision	29.8%	<i>(Télévision2)</i>
La radio	20.8%	<i>(Radio2)</i>
Internet	17.2%	<i>(Internet2)</i>
La presse écrite nationale	11%	<i>(PresseNationale2)</i>
La presse écrite régionale	16.6%	<i>(PresseRégionale2)</i>
*Autre réponse	4.6%	<i>(*Media2_Autre)</i>

(L'item « La presse gratuite » a été remplacé par le premier choix correspondant)

* signifie que la modalité est passive dans l'analyse

Les variables actives de l'analyse

3- Le temps passé à regarder les informations politiques à la télévision :

« Combien de temps par jour, en moyenne, passez-vous à regarder à la télévision les informations ou les émissions sur la politique ou l'actualité ? » (question ouverte)

(Etiquette graphique)

Moins de 30 minutes	24.6%	<i>(TVInfo_-30mn)</i>
30 minutes	27.1%	<i>(TVInfo_30mn)</i>
De 31 minutes à 1 heure	31%	<i>(TVInfo_31mn-1h)</i>
Plus d' 1 heure	13.7%	<i>(TVInfo_+1h)</i>
*Ne regarde jamais la TV	3.7%	<i>(*TV_Jamais)</i>

* signifie que la modalité est passive dans l'analyse

Les variables actives de l'analyse

4- Le journal télévisé le plus régulièrement regardé :

« Quel journal télévisé regardez-vous le plus régulièrement ? »
(question ouverte)

(Etiquette graphique)

Le 20h de TF1	29.3%	(<i>20h_TF1</i>)
Le 20h de France 2	19.3%	(<i>20h_France2</i>)
Le 13h de TF1	12.5%	(<i>13h_TF1</i>)
Le 19/20 de France 3	7.6%	(<i>19-20_France3</i>)
Le 19h45 de M6	5.3%	(<i>19h45_M6</i>)
Le 13h de France 2	5.1%	(<i>13h_France2</i>)
Les chaînes d'info continue	5.6%	(<i>InfoContinue</i>)
Ne regarde aucun JT	6.1%	(<i>AucunJT</i>)
Autre JT	8.8%	(<i>JT_Autre</i>)
*NSP	0.3%	(<i>*JT_NSP</i>)

* signifie que la modalité est passive dans l'analyse

Les variables actives de l'analyse

5- La radio la plus souvent écoutée pour s'informer en matière politique :

« Pour vous informer en matière politique, quelle radio écoutez-vous le plus souvent ? » (question ouverte)

(Etiquette graphique)

RTL	12.1%	<i>(Info_RTL)</i>
France Inter	10.3%	<i>(Info_FranceInter)</i>
France Info	8.2%	<i>(Info_FranceInfo)</i>
Europe 1	7.4%	<i>(Info_Europe1)</i>
RMC	5.9%	<i>(Info_RMC)</i>
N'écoute pas la radio	19.7%	<i>(PasEcouleRadio)</i>
N'écoute pas les infos à la radio	13.3%	<i>(PasEcouleInfoRadio)</i>
Autre Radio	20.8%	<i>(Info_AutreRadio)</i>
*NSP	2.4%	<i>(*InfoRadio_NSP)</i>

* signifie que la modalité est passive dans l'analyse

Les variables actives de l'analyse

6- La fréquence de lecture de la presse quotidienne :

« Lisez-vous la presse quotidienne ? »

Si oui : « Pouvez-vous me dire si vous lisez tous les jours ou presque, deux ou trois fois par semaine, moins souvent ou jamais : »

- Un quotidien national
- Un quotidien régional
- Un quotidien gratuit

(*Etiquette graphique*)

Ne lit pas la presse quotidienne 55.7% (*LecturePresse--*)

Ne lit aucun, ou seulement un des trois quotidiens,
au moins 2-3 fois par semaine 28.3% (*LecturePresse+-*)

Lit 2 à 3 de ces quotidiens,
au moins 2-3 fois par semaine 16.0% (*LecturePresse++*)

Les variables actives de l'analyse

7- Sujet d'actualité :

« Avez-vous entendu parler...

- Du referendum sur les minarets en Suisse ? (du 30/11/09 au 5/12/09)
- De la réforme de l'histoire-géographie en Terminale S ? (du 7/12/09 au 10/12/09)
- De l'instauration d'une taxe sur les bonus des banquiers ? (11 et 12/12/09) »

Si oui : « Comment avez-vous entendu parler de cet évènement pour la première fois ? »

N'en a pas entendu parler	33.6%
Par la télévision	34.7%
À la radio	18.5%
En lisant la presse	6.2%
Autre mode (dont internet)	7%

(Etiquette graphique)

(SujetActuNon)

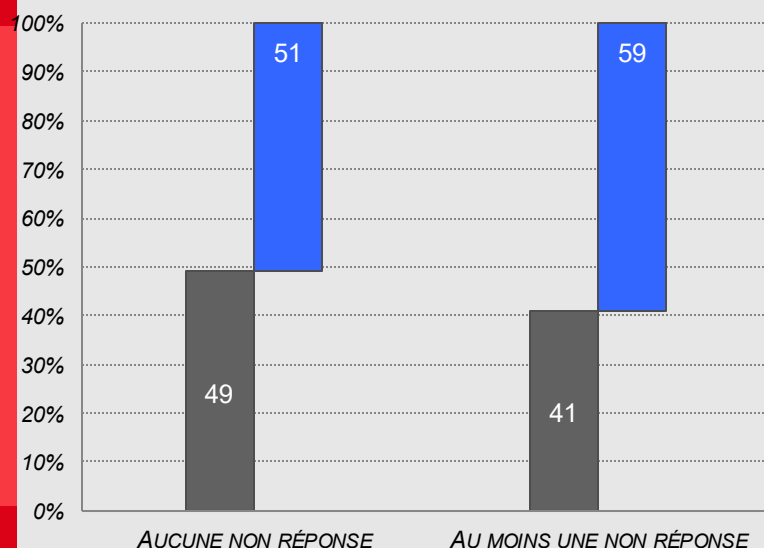
(SujetActuTV)

(SujetActuRadio)

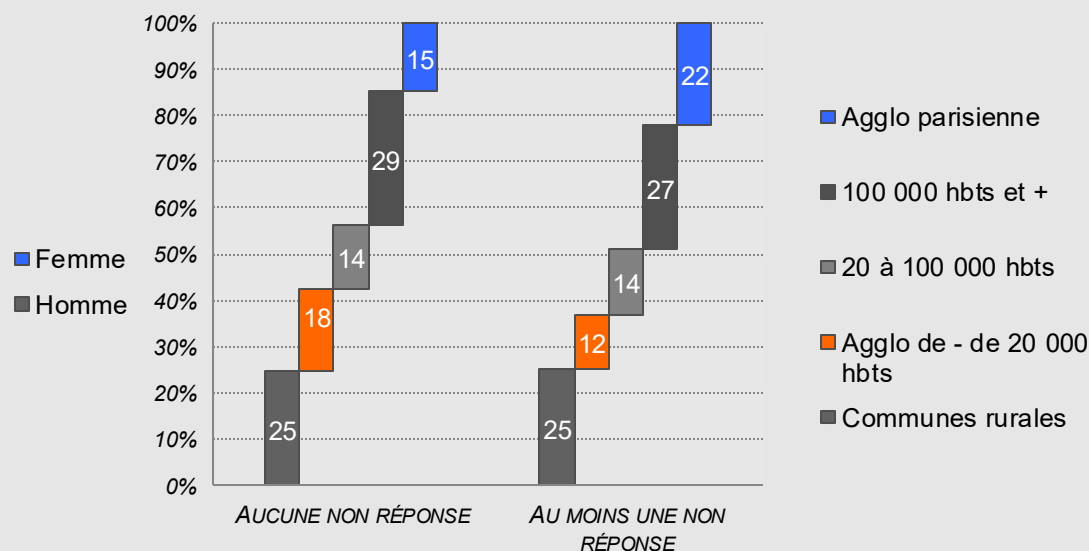
(SujetActuPresse)

(SujetActuAutre)

Qui sont les répondants concernés par les modalités rares & qui perturbent l'ACM standard ? → 11% de l'échantillon

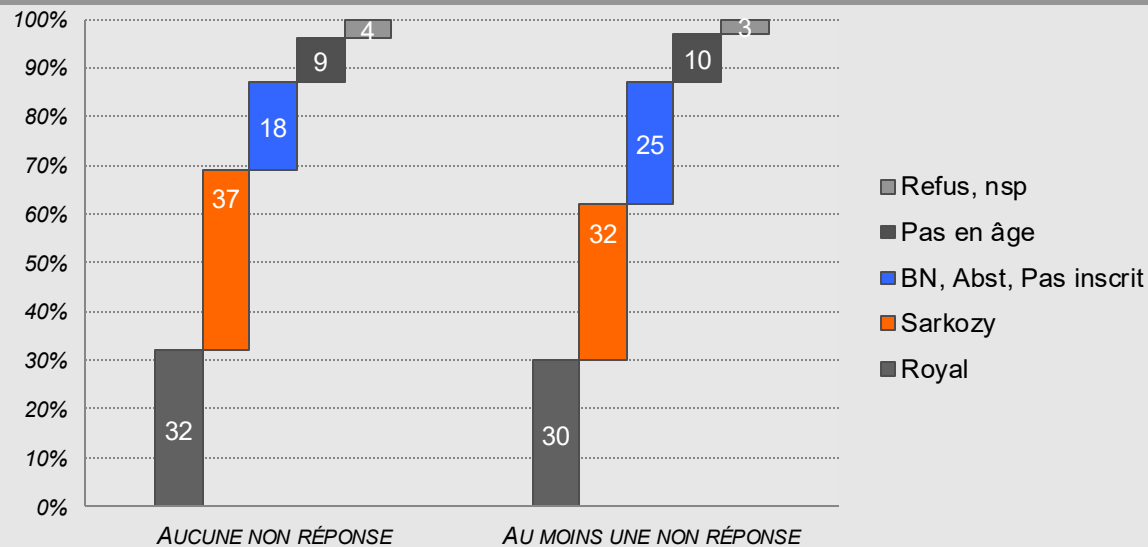


Sexe

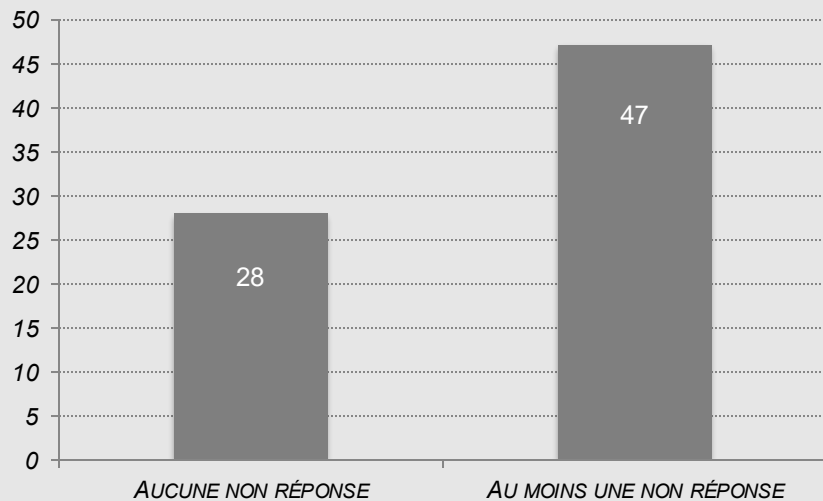


Catégorie
d'agglo

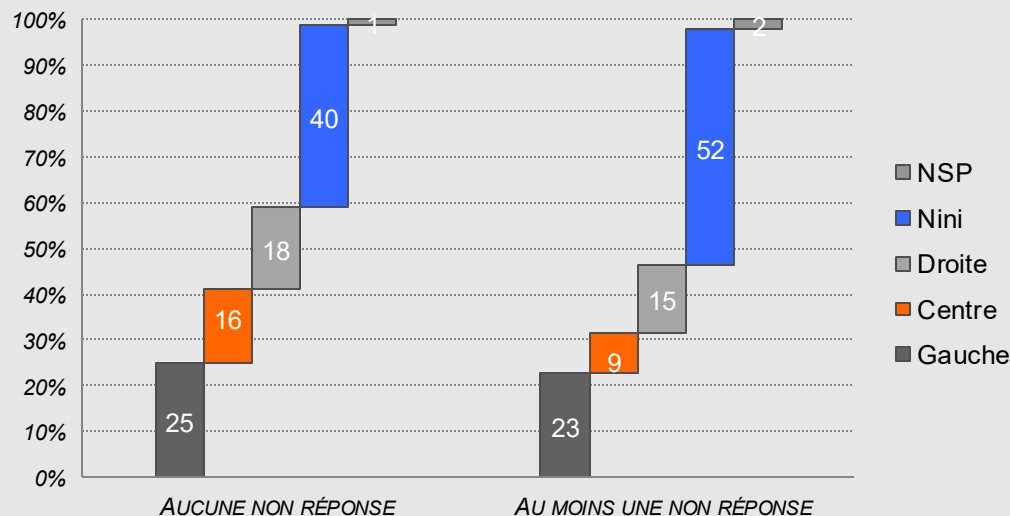
Le problème des non-réponses : exemple Médiapolis



Vote T2 2007



N'ont pas internet



Positionnement gauche / droite

Une solution : l'ACMspé

un traitement particulier des modalités rares
(fréquences $< 5\%$) et de non intérêt,
appelées « modalités passives ».

Variante de l'ACM standard, son avantage réside dans le fait
qu'elle permet de **conserver les propriétés constitutives**
(structurelles) de l'ACM standard.

Dans le calcul des distances entre individus, les modalités
passives ne sont pas supprimées, mais ignorées.
Le nuage des individus, ses directions et variables principales
sont gardés comme référence.

Principe géométrique de l'ACM :

- ❑ Construire un nuage euclidien de points représentant les individus
- ❑ Interpréter des axes : déterminer des axes principaux et des variables principales

Étape fondamentale :

- ❑ Définir les distances entre individus sur la base de leurs réponses aux questions actives

Le calcul de la distance entre individus dans l'ACM standard

Soit :

- $d_q(i, i')$ la distance entre les individus i et i' à la question q
- n_{κ} le nombre d'individus ayant choisi le modalité κ à la question q
- $n_{\kappa'}$ le nombre d'individus ayant choisi la modalité κ' à la question q

Propriété fondamentale de l'ACM :

la distance entre deux individus i et i' ne dépend que des *questions de désaccord*.

→ Par conséquent :

si q est une *question d'accord* entre i et i' : $d_q(i, i') = 0$

Le calcul de la distance entre individus dans l'ACM standard

Si q est une question de désaccord entre i et i' , l'un ayant choisi la modalité k et l'autre k' ($k \neq k'$) alors

$$d_q^2(i, i') = \frac{1}{f_k} + \frac{1}{f_{k'}}$$

Avec $f_k = n_k/n$, et $f_{k'} = n_{k'}/n$

La distance globale $d(i, i')$ entre i et i' est définie par :

$$d^2(i, i') = \frac{1}{Q} \sum_{q \in Q} d_q^2(i, i')$$

(moyenne quadratique des distances des questions)

À partir de ces distances, on définit le nuage euclidien des individus.

Le nombre de dimensions sera au plus égal à $(K-Q)$ (nombre de modalités actives moins nombre de questions actives) dont on détermine les directions principales.

(on définit les axes principaux en ajustant le nuage par la méthode des moindres carrés orthogonaux)

Propriété :

ACM (sur le tableau $I*Q$)

= AC du tableau disjonctif complet ($I*K$)

Un traitement particulier des modalités rares en ACMspé

Dans l'ACM spécifique, les modalités rares ou de non-intérêt sont appelées **modalités passives**.

- Pour deux individus i et i' qui n'ont pas choisi de modalités passives à la question q , la **distance spécifique** (notée $_{sp}d$) entre i et i' est inchangée :

$$_{sp}d_q(i, i') = d_q(i, i')$$

- Pour une question q d'accord, la distance est inchangée :

$$_{sp}d_q(i, i') = d_q(i, i') = 0$$

- Pour une question q de désaccord, l'un des individus ayant choisi une modalité active (notée k) et l'autre une modalité passive (notée k^{pass}):

$$_{sp}d_q^2(i, i') = \frac{1}{f_k}$$

Formule générale de la *distance spécifique* entre i et i' :

$${}_{sp}d^2(i, i') = \frac{1}{Q} \sum_{q \in Q} {}_{sp}d_q^2(i, i') = \frac{1}{Q} \sum_{q \in Q} \left(\sum_{k \in K' < q >} \frac{(\delta_{ik} - \delta_{i'k})^2}{f_k} \right)$$

Avec :

Q l'ensemble des questions actives

K' le sous-ensemble des modalités actives (modalités d'intérêt)

$(K'' = K - K'$ le sous-ensemble des modalités passives)

$$\delta_{ik} = \begin{cases} 1 & \text{si } i \text{ a choisi } k \\ 0 & \text{sinon} \end{cases}$$

Exemple (Le Roux, Rouanet, 2010, p. 36)

		Q			
		A	B	C	
I	i	a1	b2	c2	
	i'	a2	b2	c3	
		n	n	n	$3n=Qn$

		K							
		a1	a2	b1	b2	c1	c2	c3	
	i	1	0	0	1	0	1	0	$3=Q$
	i'	0	1	0	1	0	0	1	$3=Q$
		n_{a1}	n_{a2}	n_{b1}	n_{b2}	n_{c1}	n_{c2}	n_{c3}	$3n=Qn$

Accord sur la question B entre i et i'
 $c3$ est une modalité passive

Exemple (Le Roux, Rouanet, 2010, p. 36)

$\underline{\underline{K}}$

		a1	a2	b1	b2	c1	c2	c3	
i i'		1	0	0	1	0	1	0	$3=Q$
		0	1	0	1	0	0	1	$3=Q$
		n_{a1}	n_{a2}	n_{b1}	n_{b2}	n_{c1}	n_{c2}	n_{c3}	$3n=Qn$

$$_{sp}d_q^2(i, i') = \frac{1}{3} \times$$

$$\left[\frac{(1-0)^2}{f_{a1}} + \frac{(0-1)^2}{f_{a2}} \right] + \frac{(0-0)^2}{f_{b1}} + \frac{(1-1)^2}{f_{b2}} + \frac{(0-0)^2}{f_{c1}} + \frac{(1-0)^2}{f_{c2}} \right]$$

$$_{sp}d_q^2(i, i') = \frac{1}{3} \times \left[\left(\frac{1}{f_{a1}} + \frac{1}{f_{a2}} \right) + (0) + \left(\frac{1}{f_{c2}} \right) \right]$$



L'ACM spécifique ne revient pas à effectuer une analyse des correspondances du tableau disjonctif amputé des colonnes correspondant aux modalités passives

Une telle démarche engendre des propriétés indésirables notamment concernant les *questions d'accord* (distance entre individus non nulle)

AC du tableau disjonctif amputé

K

	a1	a2	b1	b2	c1	c2	c3	
i	1	0	0	1	0	1	0	$3=Q$
i'	0	1	0	1	0	0	1	$3=Q$
	n_{a1}	n_{a2}	n_{b1}	n_{b2}	n_{c1}	n_{c2}	n_{c3}	$3n=Qn$

K

	a1	a2	b1	b2	c1	c2	
i	1	0	0	1	0	1	$3=Q$
i'	0	1	0	1	0	0	2
	n_{a1}	n_{a2}	n_{b1}	n_{b2}	n_{c1}	n_{c2}	$3n-n_{c3}$

AC du tableau disjonctif amputé

	K					
	a1	a2	b1	b2	c1	c2
<i>i</i>	1	0	0	1	0	1
<i>i'</i>	0	1	0	1	0	0
	n_{a1}	n_{a2}	n_{b1}	n_{b2}	n_{c1}	n_{c2}

$3=Q$
 2
 $3n-n_{c3}$

Distance du Φ^2 pour la question B (question d'accord) :

$$d_{\mathcal{B}}^2(i, i') =$$

$$\sum_k \frac{n_{..}}{n_{.k}} \left(\frac{n_{ik}}{n_{i.}} - \frac{n_{i'k}}{n_{i'.}} \right)^2 =$$

$$d_{\mathcal{B}}^2(i, i') = \frac{3n - n_{c3}}{n_{b1}} \left(\frac{0}{3} - \frac{0}{2} \right)^2 + \frac{3n - n_{c3}}{n_{b2}} \left(\frac{1}{3} - \frac{1}{2} \right)^2 \neq 0!!$$

Le passage aux nuages principaux dans l'ACM spécifique

- ❑ Le *nuage spécifique des individus* \mathbb{R}^I est la projection orthogonale du nuage des individus (\mathbb{M}^I) de l'ACM standard sur le sous-espace engendré par les points idéaux associés aux seules modalités actives (modalités d'intérêt).

$$_{sp}d(i,i') \leq d(i,i')$$

- ❑ Le *nuage spécifique des modalités* est le sous-nuage des modalités actives

Les poids ($p_k = f_k / Q$) et les distances sont celles de l'ACM standard
($k \in K'$)

Les propriétés constitutives de l'ACM sont conservées :

❑ Les formules de transition (passage du nuage des individus au nuage des modalités) sont conservées et s'appliquent aux individus et modalités actives, passives et supplémentaires.

❑ Le calcul des contributions relatives et absolues est inchangé

❑ La diagonalisation de la matrice symétrique \mathbb{T} s'effectue sur les seules modalités actives

$${}_{sp}t_{kk'} = \frac{1}{Q} \left(\frac{n_{kk'}}{\sqrt{n_k n_{k'}}} - \frac{\sqrt{n_k n_{k'}}}{n} \right)$$

$$k \in K'$$

$$k' \in K'$$

	ACM standard	ACMspé	Propriétés
Variance du nuage	$V = \frac{K}{Q} - 1$	${}_{sp}V = \frac{K'}{Q} - \sum_{k \in K'} \frac{f_k}{Q}$	${}_{sp}V \leq V$
Nombre de valeurs propres (nb de \dim°)(no n trivialement nulles)	$L = K - Q$	${}_{sp}L = K' - Q'$ (Q' : nombre de questions actives sans modalité passive)	Si 1 modalité passive par question, alors ${}_{sp}L = L = K - Q$
Moyenne des valeurs propres	$\bar{\lambda} = \frac{V}{K - Q} = \frac{1}{Q}$	${}_{sp}\bar{\lambda} = \frac{\left(\frac{K'}{Q} - \sum_{k \in K'} \frac{f_k}{Q} \right)}{K' - Q'} = \frac{{}_{sp}V}{{}_{sp}L}$	
Variance totale = somme des valeurs propres			

Quelques éléments mathématiques de l'ACM spé

Propriétés conservées	ACM standard	ACMspé
Propriétés des coordonnées de l'individu l	$\text{mean } y_l^I = \sum_{i \in I} \frac{y_l^i}{n} = 0$ $\text{Var } y_l^I = \sum_{i \in I} \frac{(y_l^i)^2}{n} = \lambda_l$	$\text{mean}_{sp} y_l^I = \sum_{i \in I} \frac{_{sp} y_l^i}{n} = 0$ $\text{Var}_{sp} y_l^I = \sum_{i \in I} \frac{(_{sp} y_l^i)^2}{n} = _{sp} \lambda_l$
Propriétés des coordonnées de la modalité k (pour toute question q)	$\sum_{k \in K} \frac{f_k}{Q} y_l^k = 0$ $\sum_{k \in K} \frac{f_k}{Q} (y_l^k)^2 = \lambda_l$	$\forall q, \text{mean}_{sp} y_l^K = \sum_{k \in K_q} \frac{f_k}{Q} \times _{sp} y_l^k = 0$ $\sum_{k \in K'} \frac{f_k}{Q} \times (_{sp} y_l^k)^2 = _{sp} \lambda_l$
Coordonnée du point moyen modalité k	$\bar{y}_l^k = \sum_{i \in I_k} \frac{y^i}{n_k} = \sqrt{\lambda_l} \times y_l^k$	$_{sp} \bar{y}_l^k = \sqrt{_{sp} \lambda_l} \times _{sp} y_l^k$

Les étapes :

PHASE PRÉPARATOIRE

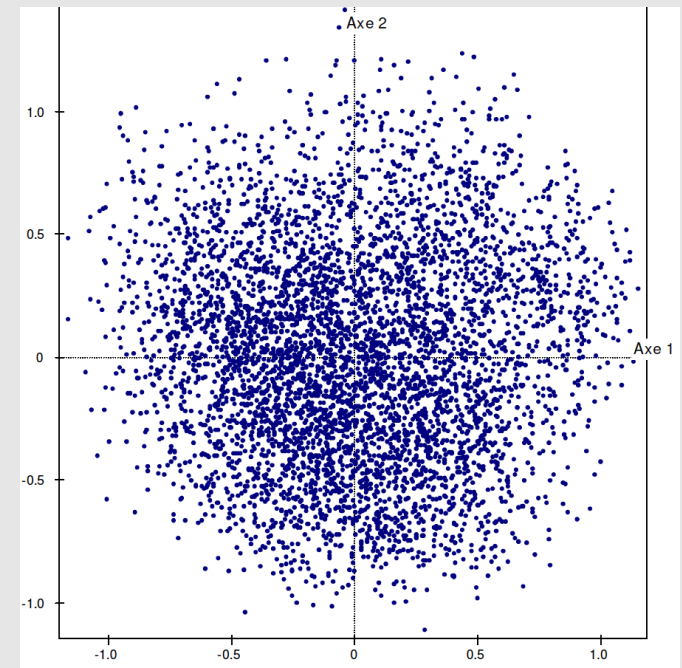
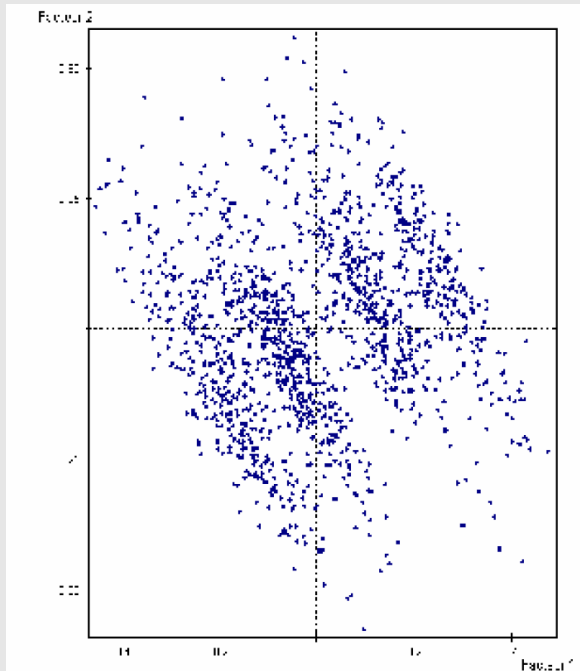
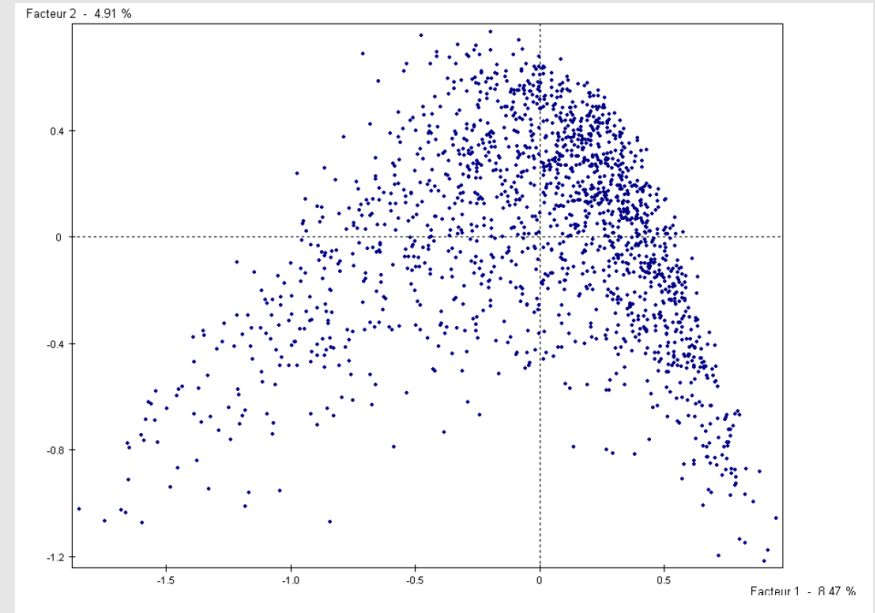
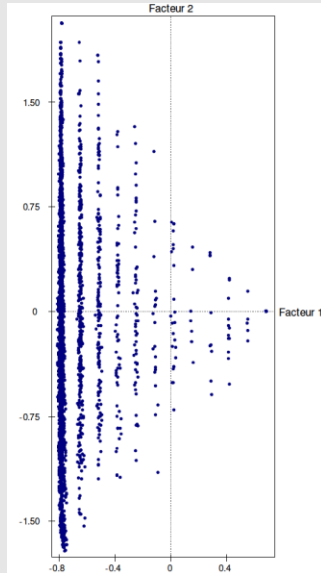
- ❑ Je détermine :
 - les variables actives et supplémentaires
(les individus actifs et supplémentaires)
→ ACM standard ou ACM spécifique (→ présence de modalités rares) ?
- ❑ Je prépare les variables (recodages éventuels) en m'assurant que les différentes dimensions de mon analyse sont également prises en compte (nb de variables équivalent) et qu'elles ont un nb de modalités comparable.
- ❑ Je réalise l'ACM spécifique

Les étapes :

PHASE DE TRAVAIL ET D'INTERPRÉTATION DES RÉSULTATS

- ❑ Je m'assure avec le listing de résultats (ou le fichier excel) que les tris à plat sont corrects et que les modalités que je souhaite passives ont bien ce statut
- ❑ Je contrôle que les valeurs propres (et les taux modifiés – que je calcule) sont satisfaisants → Nb d'axes à interpréter ?
- ❑ Je contrôle le nuage des individus et je m'assure qu'il possède une allure satisfaisante
- ❑ J'interprète les axes
 - ❑ En utilisant les contributions (fichier excel) > à la ctr moyenne
 - ❑ Je représente graphiquement les modalités qui contribuent
- ❑ J'archive les coordonnées factorielles individuelles dans le fichier de données initial (SPSS) afin de pouvoir mener des analyses complémentaires ensuite.

Allure générale du nuage des individus



Application de l'ACM spécifique Enquête Médiapolis

Les variables de l'analyse : 7 variables actives (39 modalités actives et 5 modalités passives)

- 1-Le média utilisé le plus souvent en premier (*Media1*)**
- 2-Le média utilisé le plus souvent en second (*Media2*)**
- 3-Le temps passé à regarder les informations politiques à la télévision (*TVInfo_*)**
- 4-Le journal télévisé le plus régulièrement regardé**
- 5-La radio la plus souvent écoutée pour s'informer en matière politique (*Info_*)**
- 6-La fréquence de lecture de la presse quotidienne (*LecturePresse*)**
- 7-Sujet d'actualité (*SujetActu*)**

Les variables supplémentaires de l'analyse :

Variables socio-démographiques

Positionnement politique

Les facteurs structurants de l'analyse :

La taille du réseau de discussions politiques (« Au quotidien, avec combien de personnes avez-vous l'habitude de parler de questions politiques ? »)

La fréquence des discussions politiques (« Vous arrive-t-il, souvent, rarement, parfois ou jamais de parler de politique avec des personnes de votre entourage ? »)

Valeurs propres, taux d'inertie et taux modifiés pour les 5 premiers axes

Variance totale du nuage : 4,6

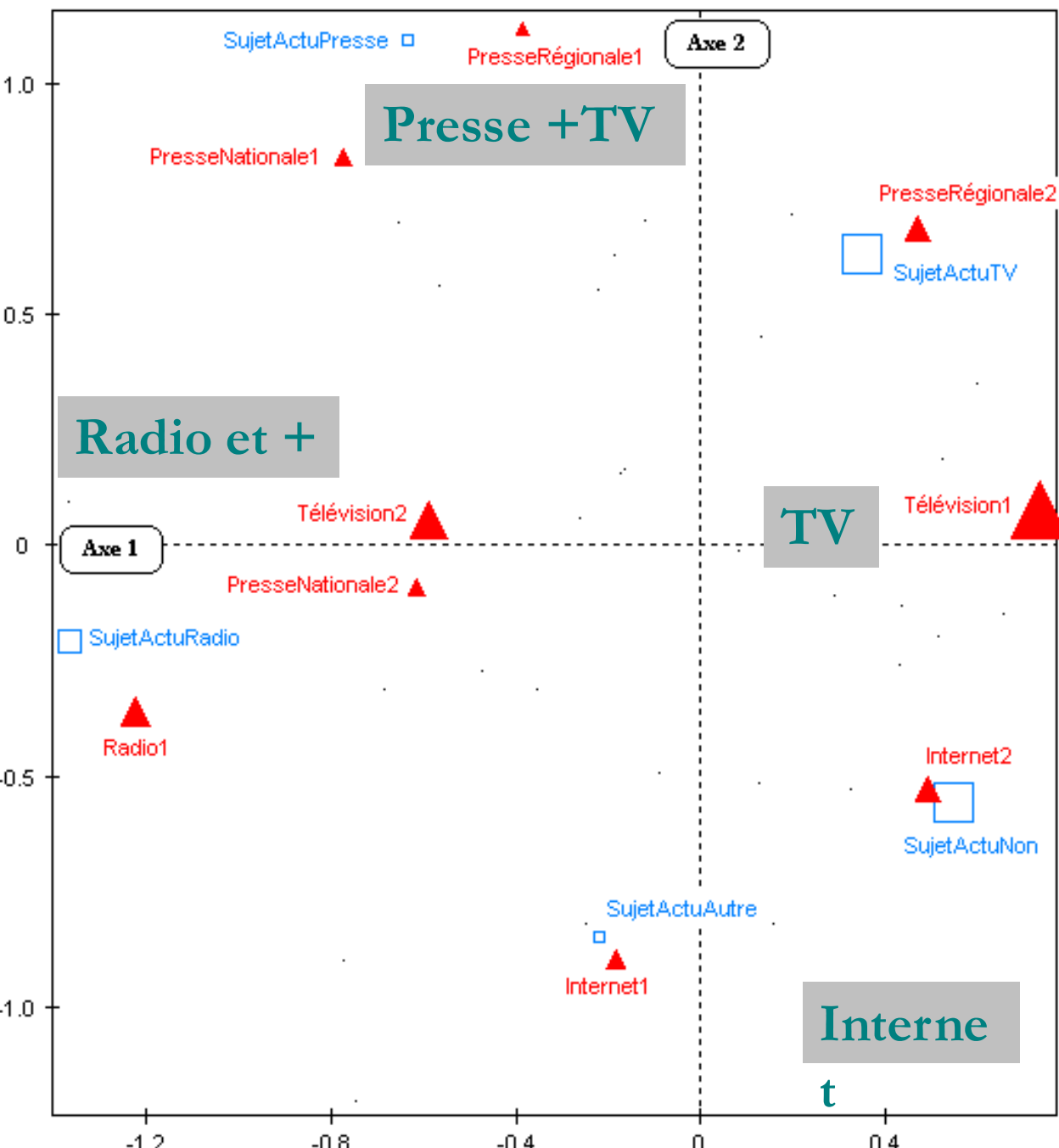
Axe	1	2	3	4	5
Valeur propre	0.303	0.249	0.224	0.195	0.183
Taux d'inertie	6.6%	5.4%	4.9%	4.2%	4%
Taux modifié	40.7%	19.7%	12.6%	6.3%	4.5%

→ Cumul:
83.9%

Le premier plan factoriel de l'ACM spécifique :

Les modes d'informations politiques et les pratiques informationnelles

Rmq : ont été conservées les modalités qui contribuent à l'axe 1 et/ou à l'axe 2 de telles sorte que leur cumul représente environ 85% de la contribution à l'axe

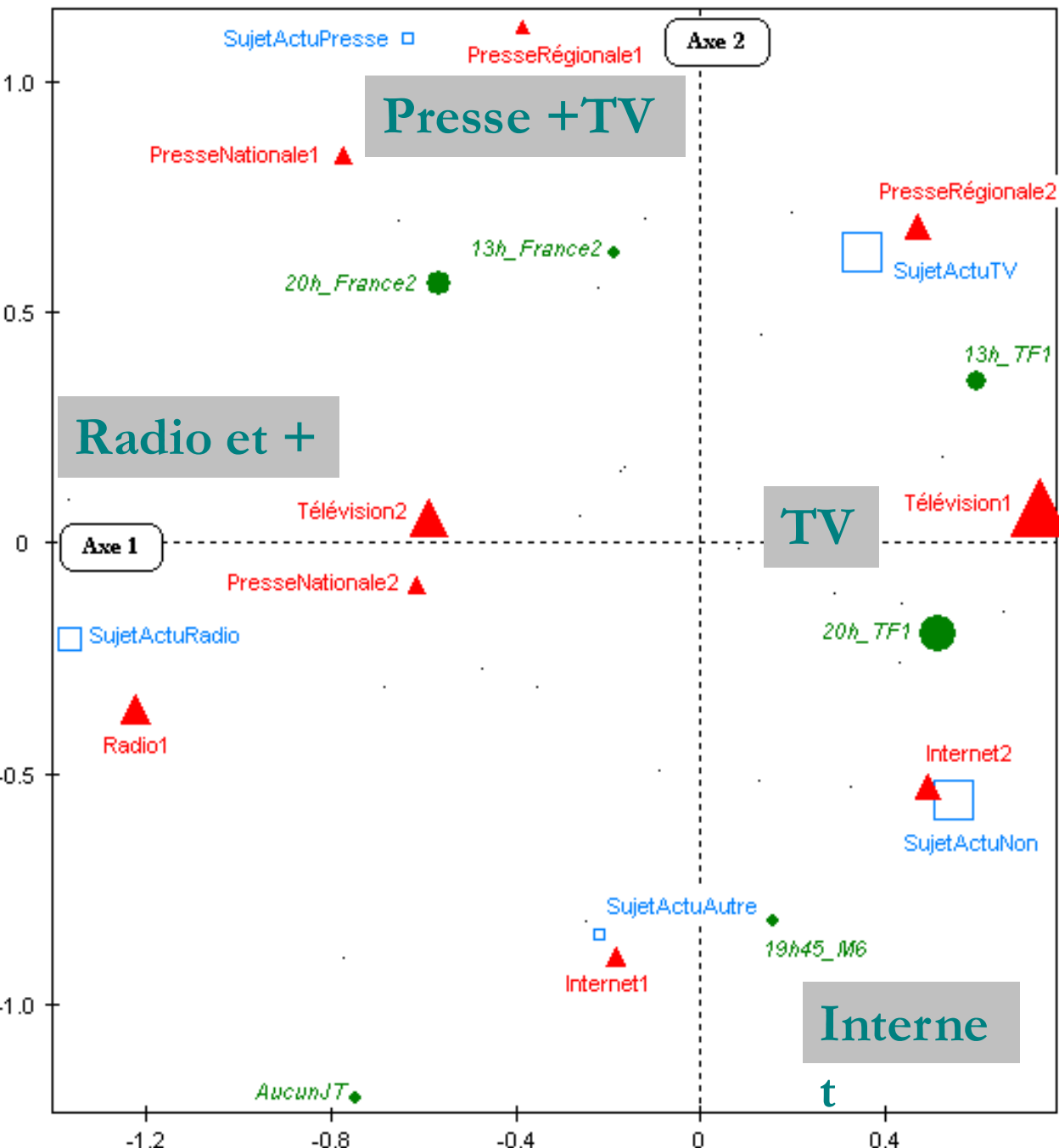


Le premier
medium utilisé

Le second
medium utilisé

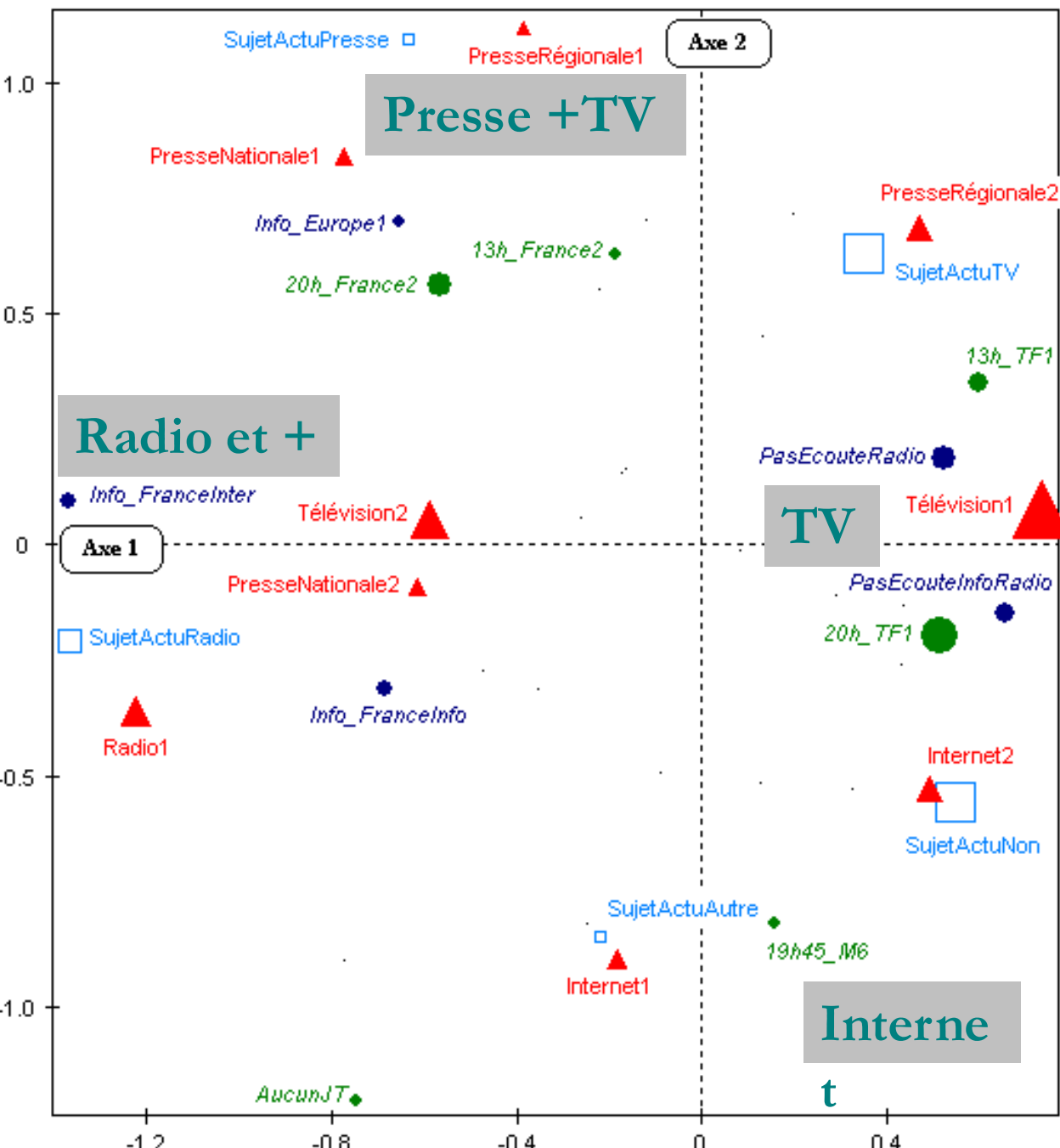
Comment le sujet
d'actualité a été
connu pour la
première fois

dans le plan 1-2



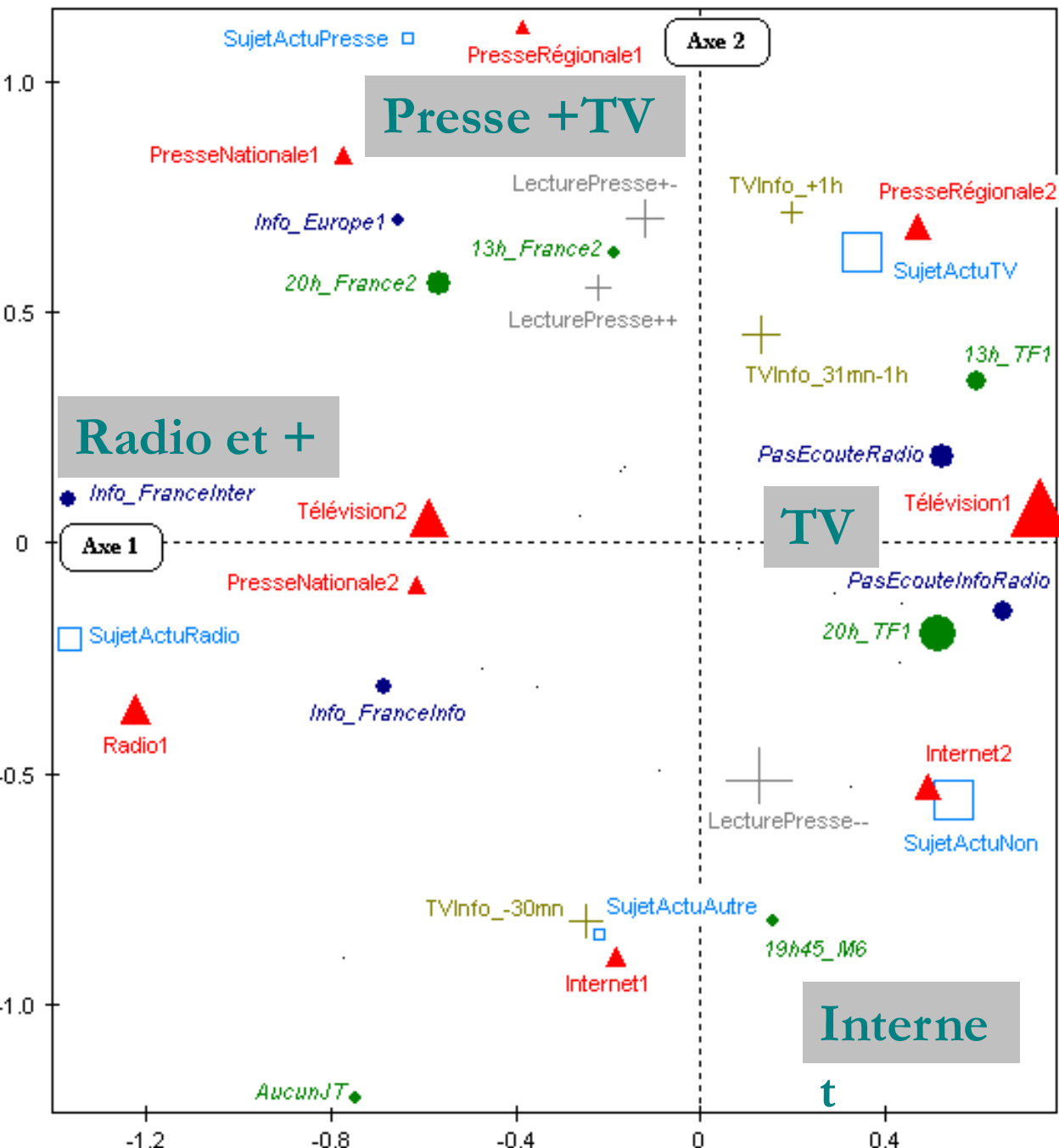
Le JT regardé le plus souvent

dans le plan 1-2



La radio la plus
souvent écoutée
en matière
politique

dans le plan 1-2



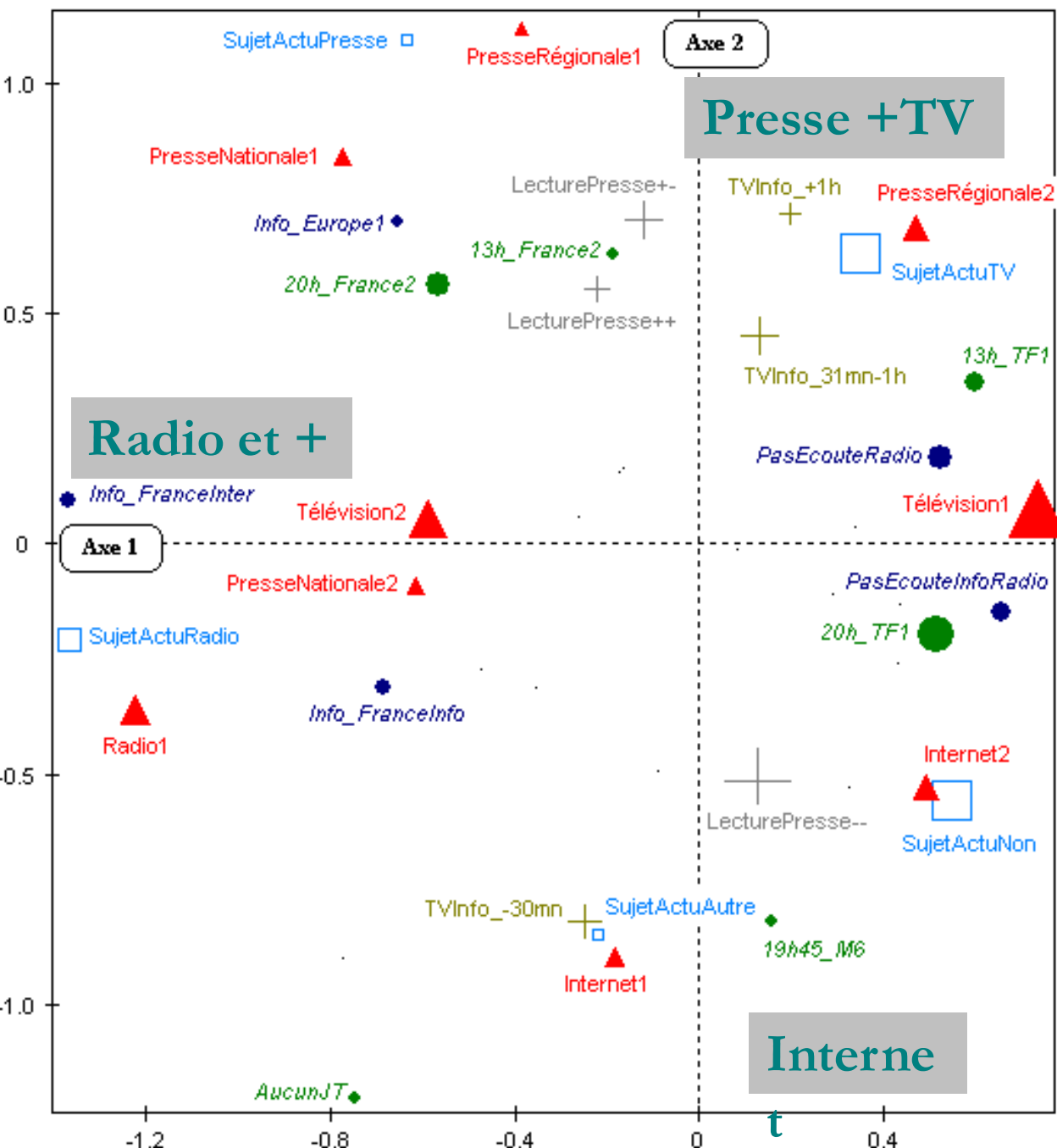
L'intensité de la lecture de la presse

Le temps passé à regarder les informations à la télévision

dans le plan 1-2

4 grands profils de pratiques informationnelles

Application de l'ACMspé à l'enquête *Médiapolis*



TV

Interne

Les caractéristiques sociales projetées dans l'espace des pratiques informationnelles :

Une forte structuration en fonction de
caractéristiques socio-démographiques et
politiques

