

Introduction à l'analyse statistique des données d'enquêtes ... simple et amusante

Pascal Marchand

***LERASS / IUT Information & Communication
(UPS Toulouse 3)***

**pascal.marchand@iut-tlse3.fr
<http://www.infocom.iut-tlse3.fr/~marchand/>**

Quelles précautions ?

- ☞ La statistique n'est pas la seule approche possible ...
- ☞ *Aux chiffres, on leur fait dire n'importe quoi !*
- ☞ La question est de savoir si, pour connaître le temps qu'il fera demain, on préfère se fier aux dictons populaires, à l'intuitive sagesse des anciens ou aux calculateurs de Météo France. Tout est possible et est question de choix épistémologique, dans le meilleur des cas ...
- ☞ La statistique ne peut pas tout faire: il faut prévoir ce qu'on lui demandera : HYPOTHESES
- ☞ Ce n'est pas la statistique qui garantit la qualité d'une recherche, mais le protocole (Cf. TD).

Que faut-il savoir ?

- ☞ Ce que sont une addition, une soustraction, une multiplication, une division et comment on les écrit (points, étoiles, fractions...).
- ☞ (éventuellement) manipuler une machine à calculer ou un ordinateur
- ☞ on note (parfois) des nombres par des lettres:
 - ☞ n pour des effectifs (n_i = nombre d'individus),
 - ☞ x ou y pour des observations (x_i =observation de chaque individu),
 - ☞ m pour la moyenne ...
 - ☞ Ces lettres remplacent donc des observations réelles.
- ☞ on note (parfois) des « tests » (qui correspondent à des calculs) par des lettres grecques. Elles remplacent donc des calculs théoriques.

Quelques signes cabalistiques:

- ☞ Σ est la *somme*, donc renvoie à une addition
- ☞ $\sum x_i$ est la *somme* des observations des individus sur la variable x
- ☞ $\sum_{i=1}^n x_i$ est la *somme* des observations sur la variable x de l'individu 1 jusqu'à l'individu n . C'est-à-dire : $x_1 + x_2 + x_3 + \dots + x_n$
- ☞ $|x|$ est la *valeur absolue de x* , c'est-à-dire sa valeur positive
- ☞ $|x - y|$ est la *valeur absolue de $x - y$* , c'est-à-dire sa valeur positive
- ☞ x^2 est x « au carré », donc $(x \times x)$ Cf. Pythagore
- ☞ \sqrt{x} est la « racine carrée » de x
 - ☞ Donc: $\sqrt{x} \times \sqrt{x} = x$

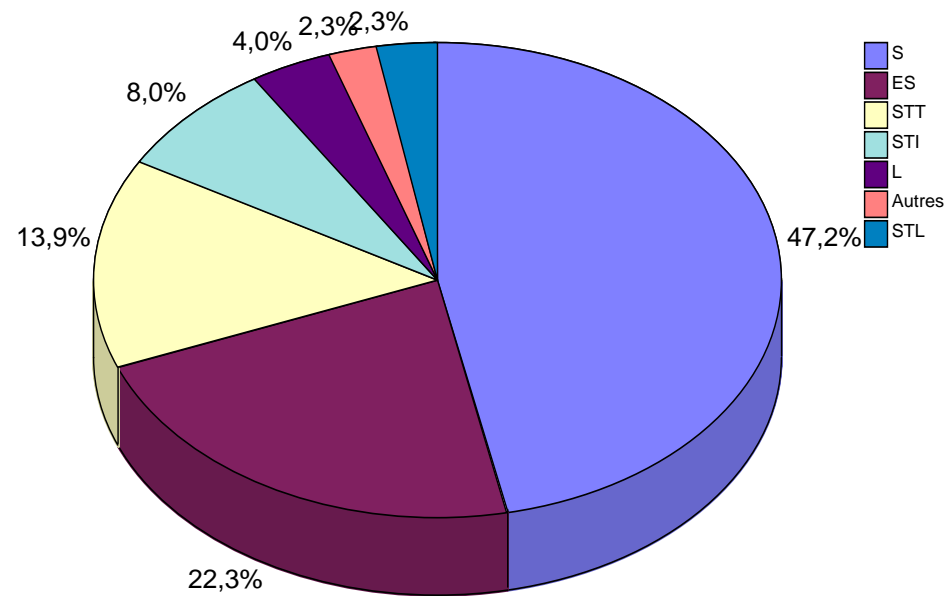
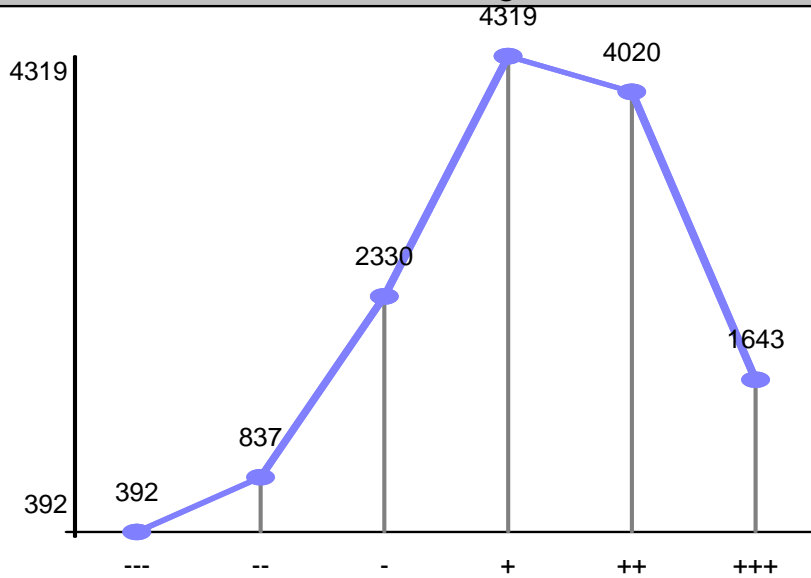
$$VD = f(VI)$$

Comportements, attitudes,
images, représentations ...

Groupes sociaux
(naturels / expérimentaux)

METIER/bonne image

Bac d'origine



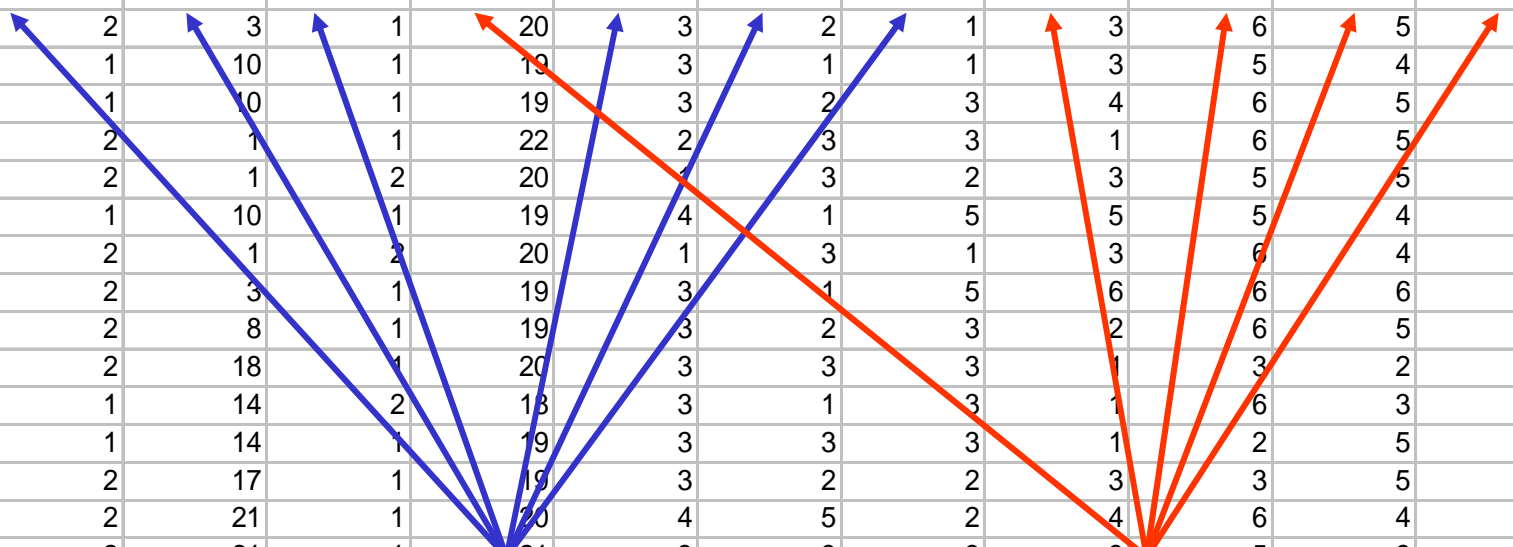
Tableaux de données

ANNEE	DEPARTEMENT	SEXE	AGE	BAC	CSP PERE	CSP MERE	Poursuivre	Court	Encadrement	Débouchés	FORMATION
2	4	2	19	3	1	1	4	5	6	6	adaptabilité, polyvalence, fo
2	3	1	20	3	2	1	3	6	5	4	courte, spécifique, complèt
1	10	1	19	3	1	1	3	5	4	4	apprentissage, emploi du te
1	10	1	19	3	2	3	4	6	5	5	compétence, professionnalis
2	1	1	22	2	3	3	1	6	5	4	spécialisées, variées, interm
2	1	2	20	3	3	2	3	5	5	4	assiduité, efficace, enseign
1	10	1	19	4	1	5	5	5	4	3	être motivé, régularité, prog
2	1	2	20	1	3	1	3	6	4	4	assidu, régulier, instructif, b
2	3	1	19	3	1	5	6	6	6	5	complète, adaptée, sérieux
2	8	1	19	3	2	3	2	6	5	5	technique, motivant, concre
2	18	1	20	3	3	3	1	3	2	2	amusante, instructive, pous
1	14	2	18	3	1	3	1	6	3	3	pratique, technique, pas for
1	14	1	19	3	3	3	1	2	5	4	sérieux, complet, encadré,
2	17	1	19	3	2	2	3	3	5	4	polyvalente, professionnelle
2	21	1	20	4	5	2	4	6	4	3	intéressante, qualificative, s
2	21	1	21	2	3	2	2	5	2	3	sérieux, bonne ambiance, c
2		1								3	théorique, conformiste, rapi
2		3								4	enseignement technique, e
2		1								3	professionnelle, multiple, co
2		3								4	polyvalence, professionnalis
1	17	2	18	2	2		3	5	6	6	5 variée, enrichissante, très in
1	17	2	18	1	1		1	1	1	6	5 intéressant, enrichissant, o
2	9	1	20	4	3		3	3	4	5	3 polyvalent, technologique, n
2	8	1	22	4	2		3	2	3	3	3 bien,po mol,peu mieux faire
2	11	1	20	4	2		3	3	5	6	2 technique suivant matiere, r
2	25	2	19	5	2		5	2	6	4	5 généraliste, intéressant, for
2	25	2	20	2	1		3	1	5	4	3 pratique , general, enrichiss
2							3	3	6	5	2 générale, technique selon n
2							1	5	6	6	5 complète, dynamique, profé
2							1	2	1	5	5 rapide, trop scolaire, pas de
2	11	1	19	4	5		1	4	5	5	5 technique, communication,
2	18	1	19	1	3		3	3	3	4	4 rapide, assez complète, un

Variables nominales
ou qualitatives

Variables continues
ou quantitatives

Variable textuelle











Libellé :

Variable :

Contrôles...

Bibliothèque...

Type

-  Fermée (unique)
-  Fermée (multiple)
-  Fermée (échelle)
-  Numérique
-  Texte
-  Code
-  Date / heure
-  Image

Modalités :

Barème...

Indiquer les modalités de réponse en les séparant par des points-virgules (;).



Nouveau titre...

Nouvelle question ...

OK

Annuler

Libellé :

Diplôme

Variable :

DIPLOME

Contrôles...

Bibliothèque...

Type

 Fermée (unique) Fermée (multiple) Fermée (échelle) Numérique Texte Code Date / heure Image

Modalités :

Aucun ; CEP BEPC ; CAP BEP ; Bac ; Bac+2 BTS ;
Bac+4 DESS ; Autres ;

7 modalités

Indiquer les modalités de réponse en les séparant par des points-virgules (;).

<

>

Nouveau titre...

Nouvelle question ...

OK

Annuler

Libellé :

Pouvez-vous indiquer votre revenu mensuel personnel ?









Variable :

REVENU

Contrôles...

Bibliothèque...

Type

-  Fermée (unique)
-  Fermée (multiple)
-  Fermée (échelle)
-  Numérique
-  Iexte
-  Code
-  Date / heure
-  Image

Utiliser le bouton 'Contrôles...' pour donner éventuellement des limites de valeurs possibles.

<

>

Nouveau titre...

Nouvelle question ...

OK

Annuler

$$VD = f(VI)$$

Hypothèse générale: les groupes comparés (VI) ne font pas la même réponse (VD). On cherche donc à rejeter *l'hypothèse nulle* (« hasard »): *le mètre, le four et le lave-linge*.

Hypothèses opérationnelles: La VI étant **nominale** (2 groupes ou plus)

☞ Si la VD_? est aussi **nominale** (2 réponses ou plus) : les groupes (G1 et G2...) ne font pas la même réponse (R1 et R2...).

☞ Si la VD est **continue** (1 valeur quantitative): les groupes (G1 et G2...) n'ont pas la même moyenne.

☞ Ho : $m_{G1} = m_{G2}$ ou $m_{G1} - m_{G2} = 0$

☞ Ha : $m_{G1} \neq m_{G2}$ ou $m_{G1} < m_{G2}$ ou $m_{G1} > m_{G2}$

☞ ou : $m_{G1} - m_{G2} \neq 0$ ou $m_{G1} - m_{G2} < 0$ ou $m_{G1} - m_{G2} > 0$

☞ Donc: l'hypothèse détermine les statistiques à appliquer

$$VD = f(VI)$$

- ☞ Problème : Quelle confiance puis-je avoir dans la différence observée entre G1 et G2?
- ☞ Notion de *probabilité* : nombre de cas favorables / nombre de cas possibles (dés, cartes, loto ...).
- ☞ On compare donc un événement observé à tous les événements possibles.
- ☞ La probabilité associe à un événement un nombre (réel) compris entre 0 (= événement impossible) et 1 (ou 100% = événement certain).
- ☞ Les tests statistiques cherchent à comparer une observation particulière (les réponses de l'enquête) au hasard (Cf. définition de la science).

1. $\chi^2 = f(VI)$

Variables **nominales**
(qualitatives)

Variables **nominales**
(qualitatives)

?

Sig.

ns

	G1	G2
R1	0%	50%
R2	50%	0%

	G1	G2
R1	25%	25%
R2	25%	25%

Effectifs observés :

Effectifs théoriques (si < 5, Yates) :

	G1	G2	
R1	n1	n2	M1
R2	n3	n4	M2
	M3	M4	N

	G1	G2	
R1	n'1=M1xM3/N	n'2=M1xM4/N	M1
R2	n'3=M2xM3/N	n'4=M2xM4/N	M2
	M3	M4	N

$$\chi^2 = \sum \frac{(n_i - n'_i)^2}{n'_i}$$

Cette valeur-test, associée au DDL = $(n_i - 1)(n_j - 1)$, permet de repérer les seuils de probabilité (p) de rejet de H_0 (généralement .05)

La représentation des études

Tableau croisé (effectifs observés (n) et pourcentages) :

Département :S/T	Tertiaires	Secondaires	TOTAL
Sexe			
h	19,1% (2586)	37,3% (5050)	56,4% (7636)
f	30,8% (4172)	12,8% (1733)	43,6% (5905)
TOTAL	49,9% (6758)	50,1% (6783)	

Effectifs théoriques (n') et observés (n) :

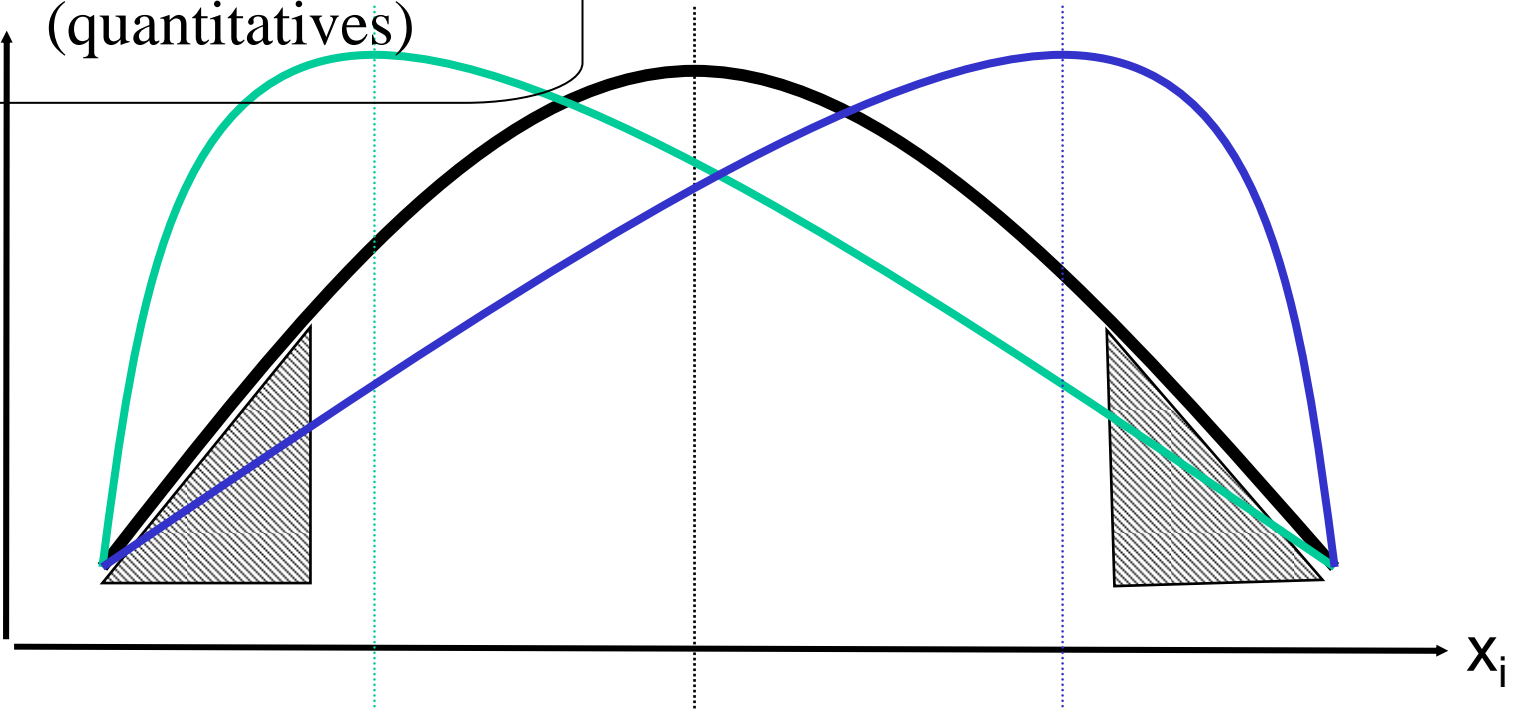
Département :S/T	Tertiaires	Secondaires	TOTAL
Sexe			
h	3810,95 (2586)	3825,05 (5050)	7636 (7636)
f	2947,05 (4172)	2957,95 (1733)	5905 (5905)
TOTAL	6758 (6758)	6783 (6783)	13541 (13541)

$$\chi^2 = \sum \frac{(n_i - n'_i)^2}{n'_i} = 1802,45 \quad p < .0001$$

2. $\mathcal{VD} = f(\mathcal{VI})$

- Tendence centrale (mode, moyenne, médiane)

Variables **continues**
 n_i (quantitatives)

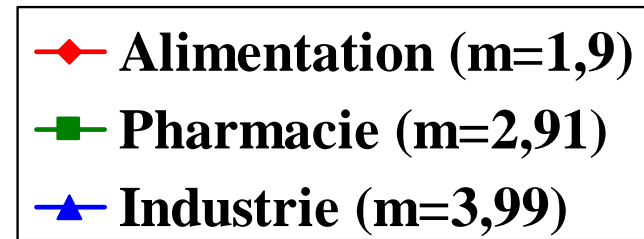
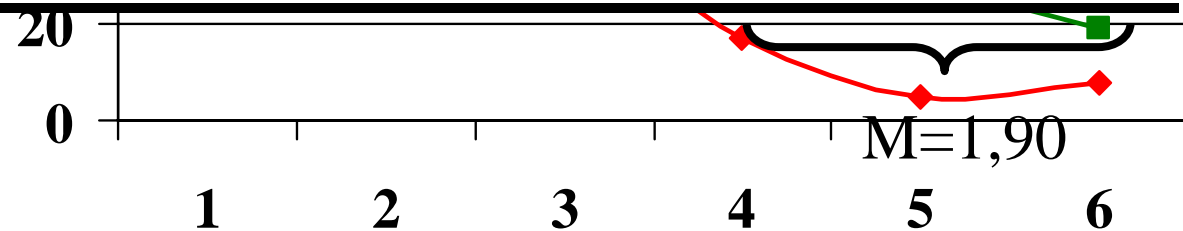


$$m = \frac{\sum (x_i \cdot n_i)}{N}$$

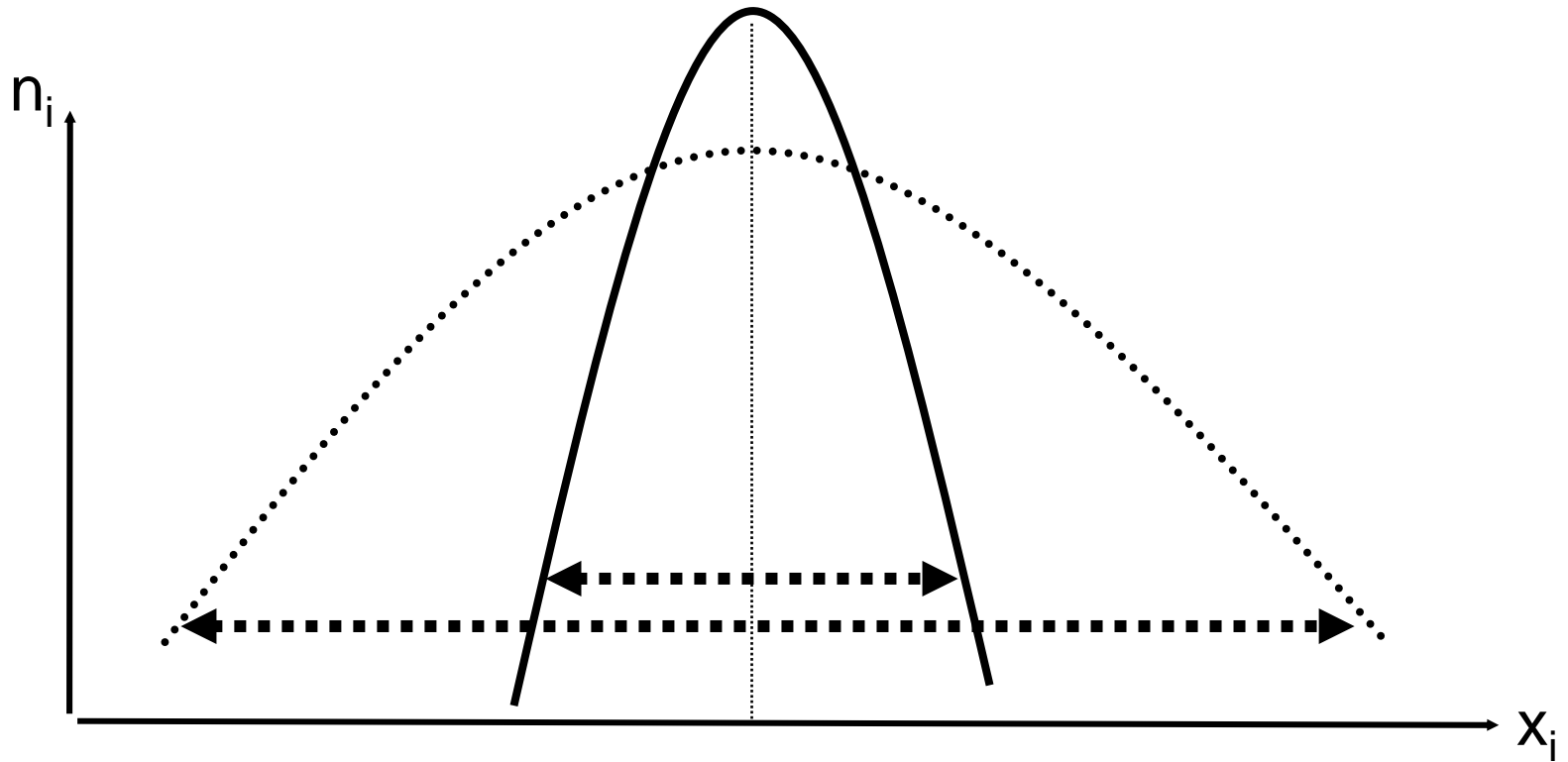
A quel point êtes-vous favorable aux OGM ?

Item d'attitude de type Likert (6 points)

xi	ni	xi.ni
1	156	156
2	80	160
3	45	135
4	17	68
5	5	25
6	8	48
Somme	311	592



- Dispersion (variance, écart-type)

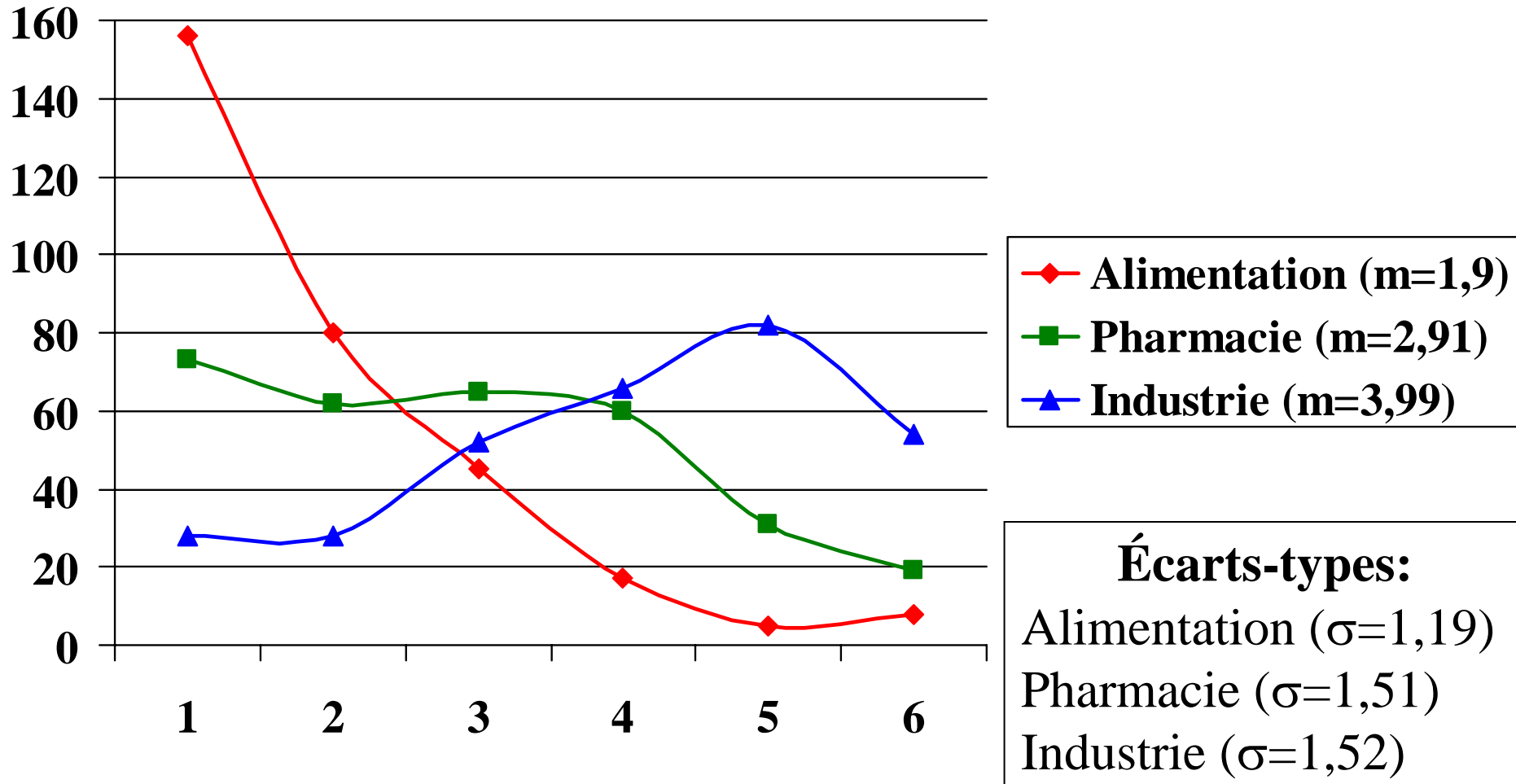


$$\text{Variance} = \frac{\sum (x_i - m)}{N}$$

$$\sigma = \sqrt{\frac{\sum (x_i - m)}{N}}$$

A quel point êtes-vous favorable aux OGM ?

Item d'attitude de type Likert (6 points)



$$VD = f(VI)$$

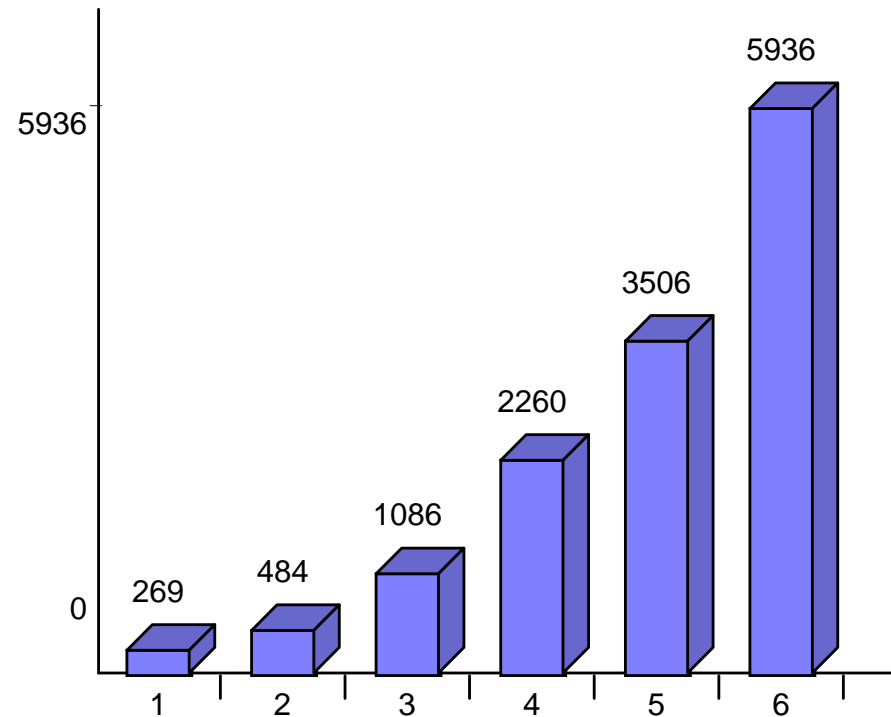
Variables **continues**
(quantitatives)

Variables **nominales**
(qualitatives)

Croisement VI = *sexe* avec l'item d'attitude « les études demandent un certain travail » (6 points) :

Sexe	ETUDES/demandent un certain travail
h	7636 (4,67)
f	5905 (5,25)
TOTAL	13541 (4,92)

ETUDES/demandent un certain travail



Comparaison de moyennes

- L'analyse de variance (pour N sujets et k groupes) :
 - F = variance **inter** / variance **intra**

... entre les moyennes

$$\sum n_g \cdot (m_g - m)^2 / (k - 1)$$

... à l'intérieur des groupes

$$\sum \sum (x - m_g)^2$$

- Cette valeur-test, associée aux DDL (intra et inter), permet de repérer les seuils de rejet de H_0 ($p < .05$).
- Degrés De Liberté :
 - Inter = $k - 1$
 - Intra = $N - k$
 - Total = $N - 1$

N.B. $(N - 1) = (k - 1) + (N - k)$

Tableau	Graphique	Valeurs	Comparer	Ordonner	Regrouper	Renommer	Légende	Carrousel	Suivant
ACP	Barème	Tests	Caract.	Permuter	Supprimer	Revenir	Dimensions	Transposer	Autre

Sexe	ETUDES/demandent un certain travail
h	7636 (4,6740)
f	5905 (5,2481)
TOTAL	13541 (4,9244)

Les valeurs du tableau sont, pour chaque critère et chaque modalité, le nombre d'observations hormis les non-réponses.

Les noms des critères discriminants sont encadrés.

Les nombres encadrés correspondent à des moyennes par catégorie significativement différentes (test t) de l'ensemble de l'échantillon (au risque de 95%).

Résultats du test de Fisher :

ETUDES/demandent un certain travail : $V_{inter} = 1097,3263$, $V_{intra} = 1,4791$, $F = 741,8989$, $1-p = 99,9990\%$

Les paramètres sont établis sur une notation de 1 (1) à 6 (6).

Exemple : La représentation des études

Croisement sexe/secteur ($k = 4$) avec des items d'attitude (en 6 points) sur les études (N=13541) :

LES ETUDES ...	h/T	h/S	f/T	f/S	V_inter	V_intra	F
préparation à l'avenir	4,52	4,59	4,73	4,7	29,92	1,15	25,95
acquérir certaines connaissances	4,91	5,06	5,21	5,3	72,48	0,86	83,86
acquérir certaine culture	4,33	4,23	4,6	4,53	118,27	1,63	72,41
comportent certaines difficultés	4,43	4,61	4,69	4,96	103,65	1,48	70,25
permettent certains diplômes	5,13	5,17	5,35	5,36	43,47	0,89	48,96
enrichissement intellectuel	4,6	4,67	4,94	4,98	108,31	1,29	83,9
accéder à certaines professions	4,7	4,91	4,92	4,98	37,93	1,3	29,13
acquérir certaine qualification	4,75	4,94	5	5,09	49,14	1,03	47,6
capacités de réflexion	4,46	4,67	4,69	4,81	49,21	1,27	38,63
demandent un certain travail	4,64	4,69	5,2	5,37	378,83	1,48	256,59
valorisation sociale	4,16	4,17	4,34	4,17	30,1	1,69	17,85
demandent une certaine volonté.	4,8	4,89	5,2	5,31	164,07	1,08	151,71

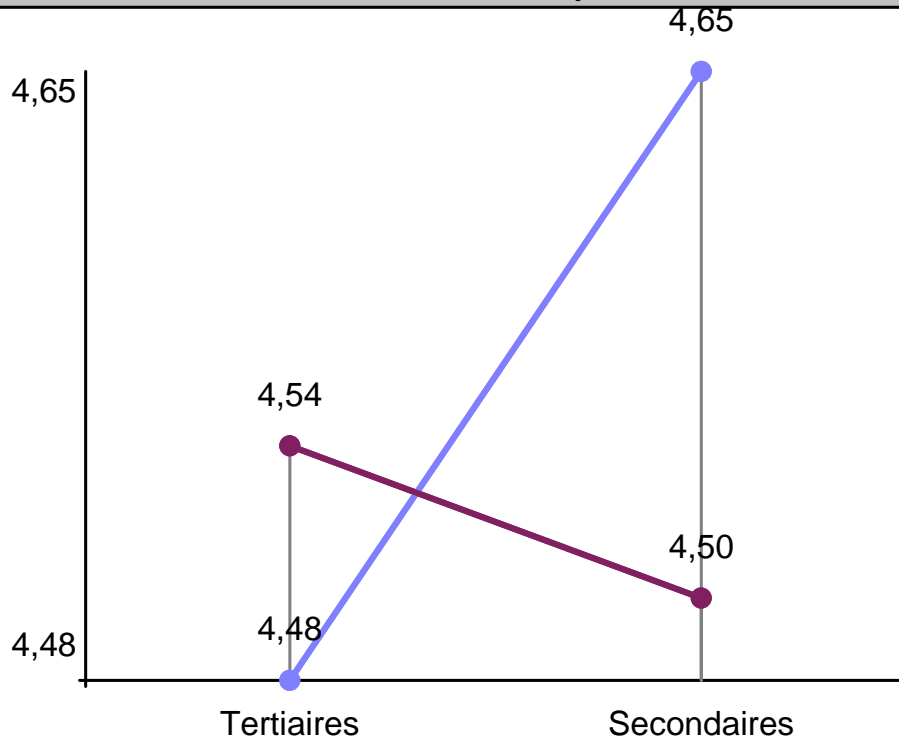
DDL-inter= 3 ; DDL-intra=13537 ; DDL-total= 13540

3. $VD = f(VI_1 * VI_2)$

Variables **continues**
(quantitatives)

Variables **nominales**
(qualitatives)

'Sexe' x 'S/T' x 'Débouchés professionnels'



Analyse de la variance à deux facteurs (2586 répétitions) :

- L'effet principal de 'Sexe' est très significatif ($F = 6,71, p < .01$)
- L'effet principal de 'S/T' est très significatif ($F = 14,18, p < .01$)
- L'interaction de 'Sexe' et 'S/T' est très significative ($F = 15,13, p < .01$)

—●— h —●— f

4. $\mathcal{VD} = f(\mathcal{VI})$

Variables **continues**
(quantitatives)

Variables **continues**
(quantitatives)

La **covariance** mesure le degré de dépendance de deux variables X et Y:

$$\text{covar}(X, Y) = \frac{\sum_{i=1}^n (y_i - m_y)(x_i - m_x)}{n-1} \quad (= 0 \text{ si indépendance})$$

La **régression** indique l'intensité du lien entre deux variables quand l'une est contrôlée. C'est une fonction $y_i = f(x_i)$.

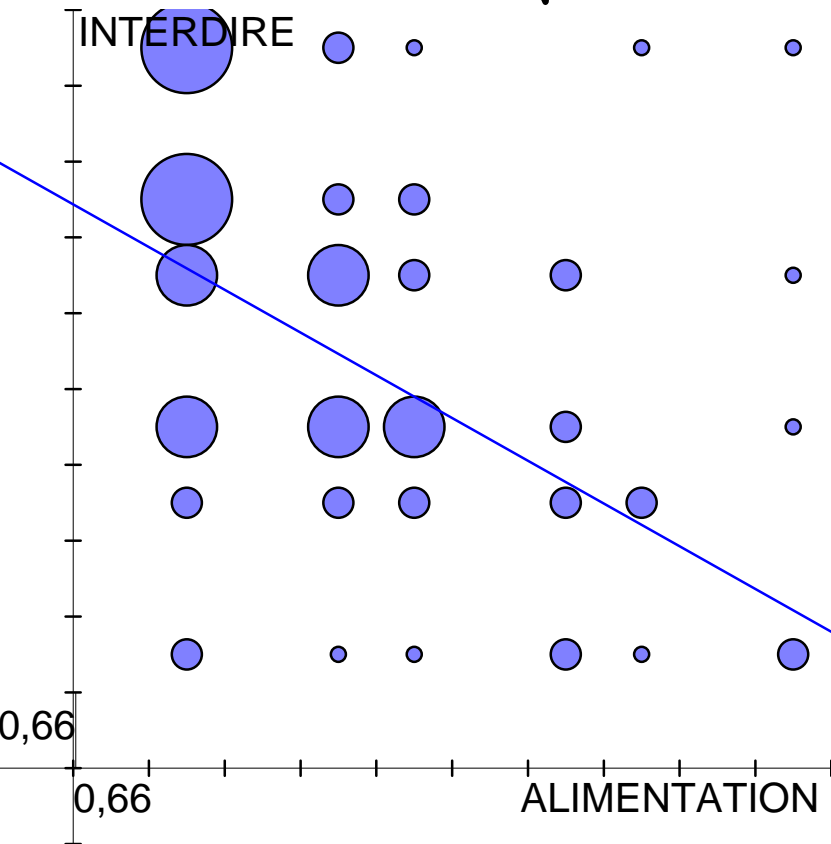
La **corrélation** (paramétrique) indique l'intensité du lien entre deux variables aléatoires.

Dans les deux cas, il s'agit d'ajuster une droite sur un ensemble de points d'observation (x_i, y_i) et de l'utiliser pour tester l'intensité de la liaison entre les deux variables.

La corrélation (paramétrique)

Le « r » de Bravais-Pearson :

$$r = \frac{\text{cov}(x, y)}{\sqrt{\text{var}(x) \times \text{var}(y)}} = \frac{\sum_{i=1}^n \frac{(y_i - m_y)(x_i - m_x)}{n}}{\sqrt{\sum_{i=1}^n \frac{(y_i - m_y)^2}{n} \sum_{i=1}^n \frac{(x_i - m_x)^2}{n}}}$$



Interdire les OGM * Favorable aux OGM dans l'alimentation

Équation de la droite de régression :

• $\text{INTERDIRE} = -0,56 * \text{ALIM.} + 4,91$

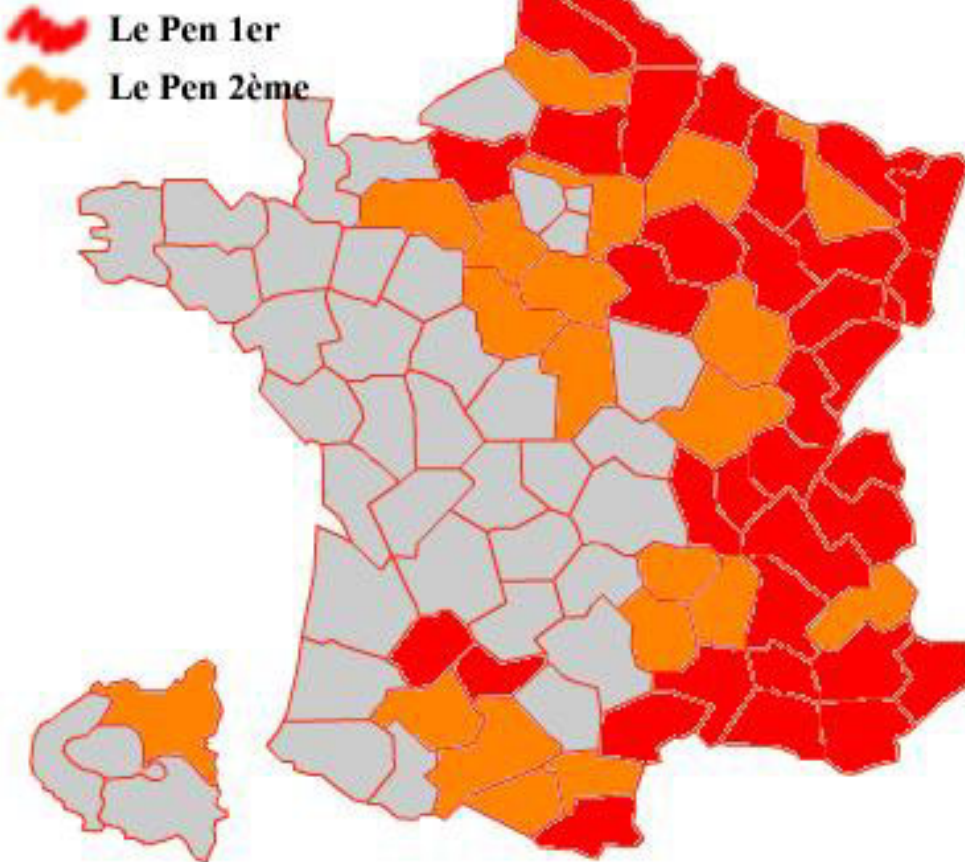
• Coefficient de corrélation : -0,46 (ns)

(ALIMENTATION explique 21% de la variance de INTERDIRE)

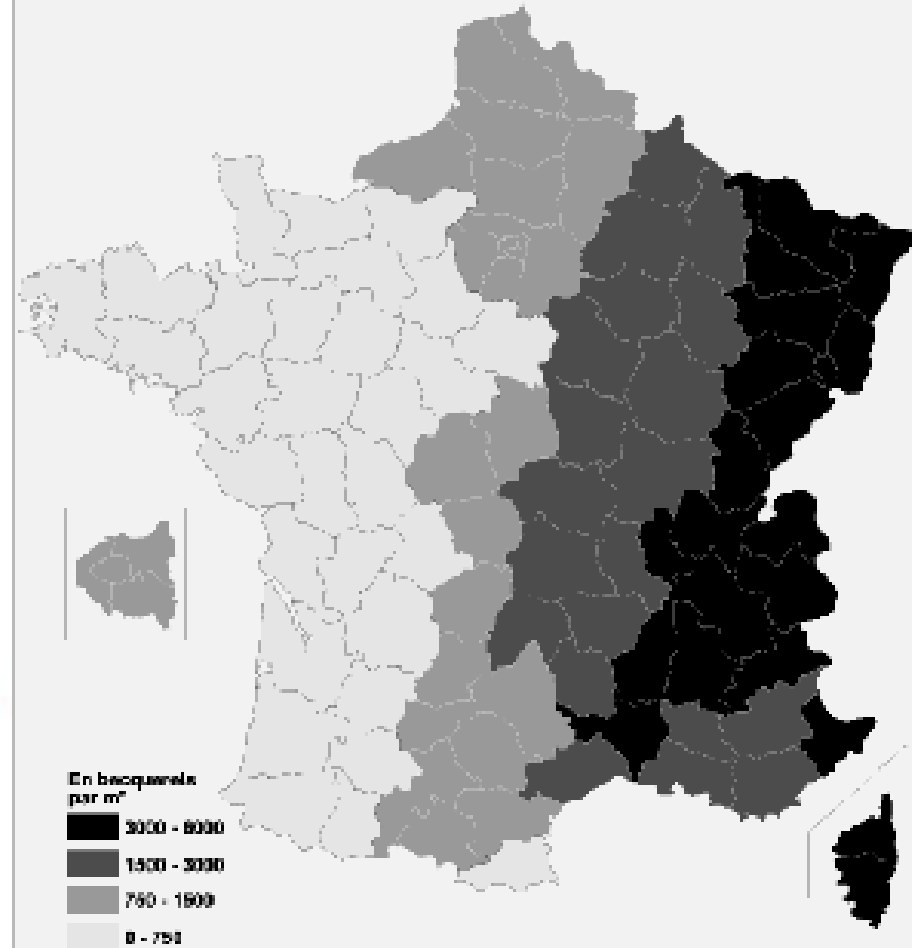
Écart-type du coefficient de régression :
0,062

Interpréter des corrélations ...

La carte qui fait peur



Les retombées radioactives de Tchernobyl
Activité surfacique de CESIUM 137 (en mai 1986)



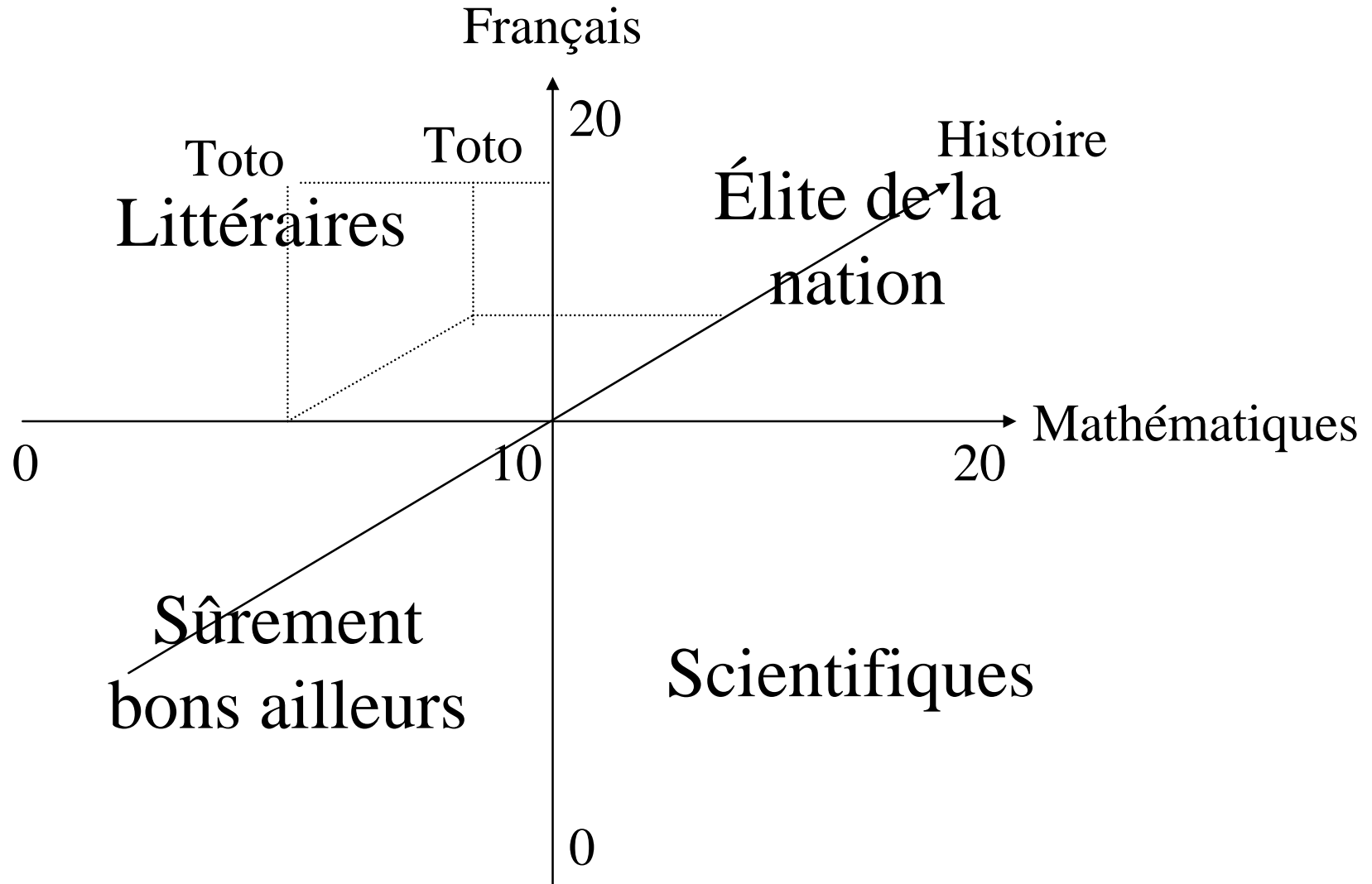
Source : MapInfo Institut de protection et de sûreté nucléaire

Copyright : KTEHCARTO 2006

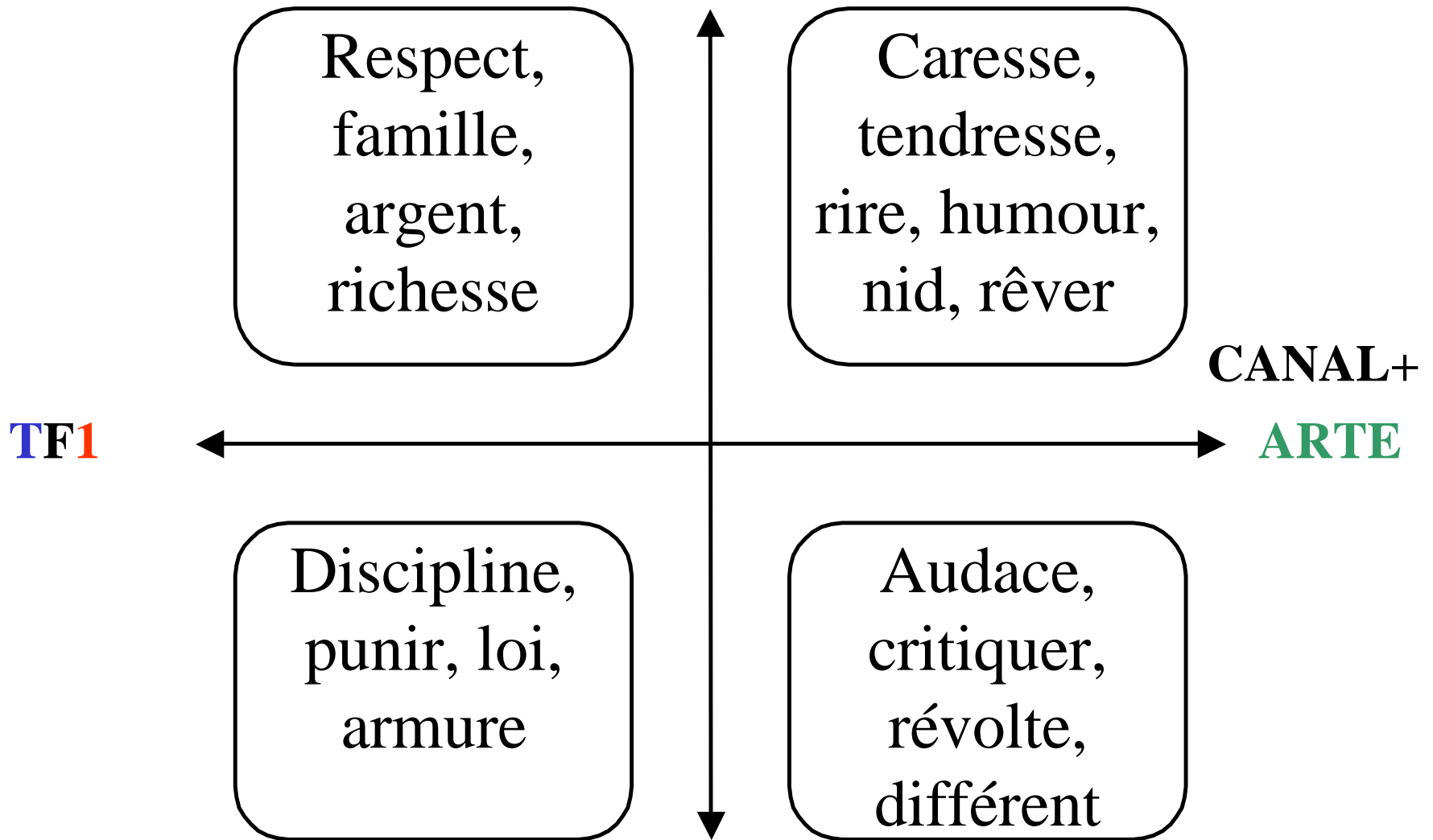
5. L'analyse des données

- extraire des structures pouvant rendre compte de liens entre différentes réponses d'un tableau multidimensionnel
- opérer des typologies d'individus, c'est-à-dire des classifications prenant en compte simultanément plusieurs réponses ou plusieurs caractéristiques de base.
- Faire une comparaison des typologies d'individus avec les structures de réponses.
- Plus précisément, on s'attend à ce que :
 - il y ait moins de structures que de variables
 - ces structures rendent compte à la fois des réponses des individus et des liens entre ces réponses
 - ces structures aient un sens psychologique ou sociologique (c'est-à-dire qu'on puisse les nommer).

- **Notion de nuage de points:** Interprétation des positions d'élèves en fonction du croisement de leurs notes



- Le Diberder, A. (1993). L'image des chaînes : les mots pour le dire. *Les écrits de l'image*, 1, 112-131.



Tableaux de données (terrorisme)

GROUPE	POLITIQUE	Ils_ont_souffert_y_a_trop_d	Notre_société	Ils_refusent_c	La_société_n	Ils_ont_des_c	Ils_souffrent_	Ils_sont_indif	Ils_se_laisser	
1	6	3	4	3	3	4	3	4	2	4
1	3	1	2	1	7	1	1	6	7	5
1	4	1	2	3	7	5	1	5	4	6
1	5	2	6	1	3	4	1	3	7	4
1	2	1	7	4	7	6	2	2	4	3
1	3	3	7	3	7	3	2	3	6	6
1	1	1	7	5	7	2	1	1	5	1
2	4	1	4	1	1	1	1	1	2	2
2	5	1	5	4	2	2	2	3	6	5
2	7	2	1	3	6	6	1	3	7	2
2	3	1	3	4	5	2	1	1	3	3
2	4	2	5	4	6	6	2	2	5	4
2	2	1	7	2	1	7	1	1	5	4
1	9	1	5	4	7	7	1	1	6	6
3	1	1	5	2	4	1	1	2	6	7
3				1				4	1	6
3				2				3	5	6
3				4				5	7	6
3	3	1	7	1	4	4	1	1	1	1
3	4	4	4	2	2	1	1	4	5	6
3	4	5	4	4	7	5	6	7	7	6
3	8	4	3	6	7	7	1	5	3	4
4	1	3	1	1	2	4	1	1	1	1
4	4	1	2	2	6	1	1	3	5	5
4	7	1	1	1	1	1	1	1	4	6
4	4	2	6	4	6	5	1	5	4	4
4	4	1	4	4	3	4	1	1	6	4
4	5	2	4	4	5	4	5	6	6	3
4	3	2	5	2	5	6	2	3	4	2
4	2	2	5	4	5	2	1	2	6	6
4	6	3	7	4	6	6	2	1	4	7

Variables illustratives

Variables actives

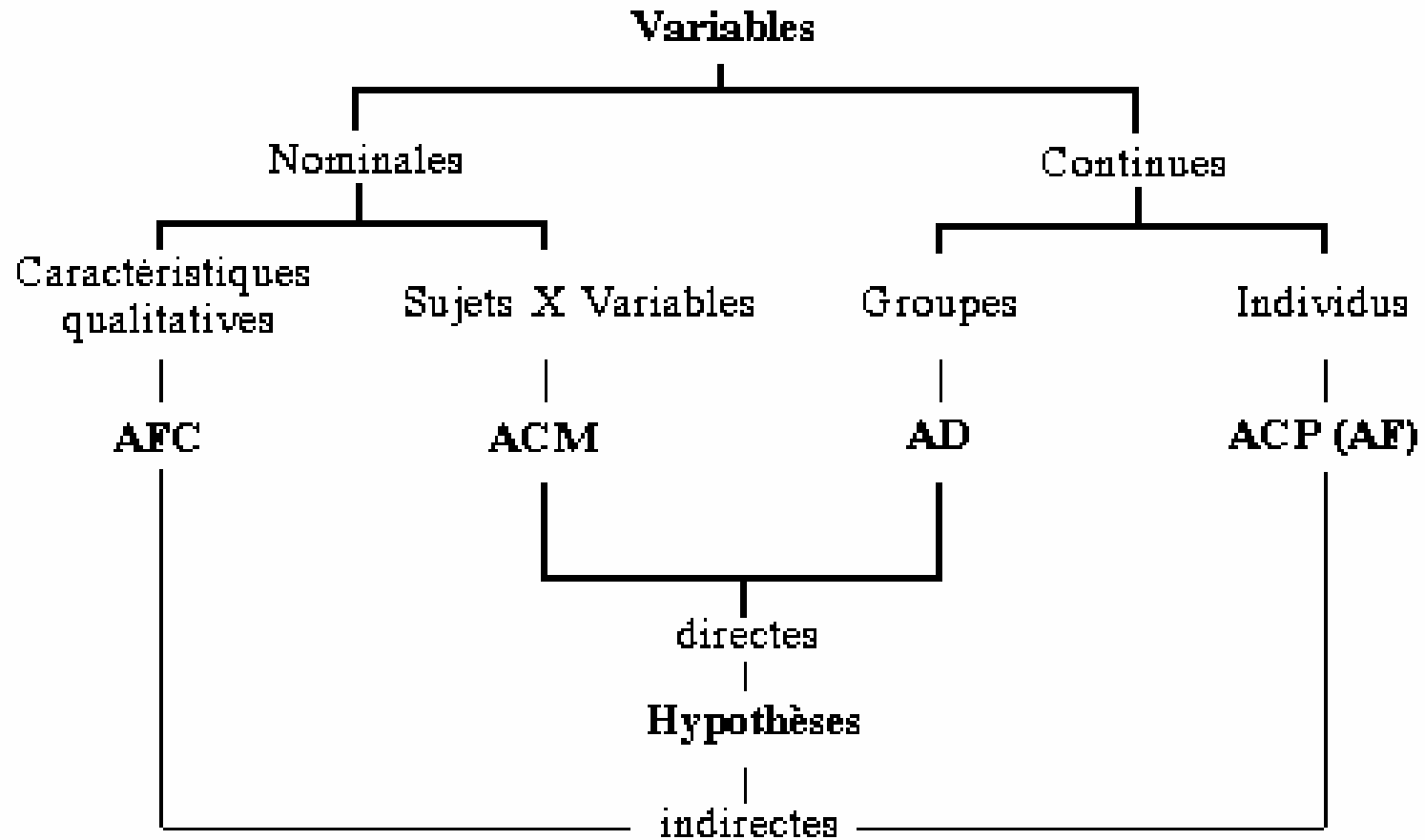
L'analyse des données

En analyse multidimensionnelle, on cherche à représenter les n lignes et les p colonnes du tableau de données par des points dont les coordonnées sont précisément les éléments de ce tableau. Deux nuages de points sont construits:

- le nuage des n individus (nuage des points-lignes) situé dans l'espace à p dimensions des variables (colonnes): chacune des n lignes est représentée par un point à p coordonnées
- le nuage des p variables (nuage des points-colonnes) situé dans l'espace à n dimensions des individus (lignes): chacune des p colonnes est représentée par un point à n coordonnées

Le tableau de données est donc une matrice dans laquelle chaque *vecteur* (ligne ou colonne) représente un point soit dans l'espace des individus, soit dans l'espace des variables. Chacune des deux dimensions du tableau permet de définir les distances (ou proximités) entre les éléments définissant l'autre dimension: les colonnes permettent de définir les distances entre les lignes et inversement.

L'analyse des données



Scores factoriels à expliquer par :

variables nominales { a priori : ANOVA
a posteriori : SEGMENTATION

variables continues : REGRESSION

L'analyse en composantes principales

L'ACP s'effectue sur des variables continues (quantitatives).

Pour pouvoir analyser des variables d'étendues diverses, le tableau de données est centré et réduit.

- Centrer revient à soustraire à chaque observation la moyenne. La somme des observations (colonne) est égale à 0.
- Réduire consiste (grosso modo) à diviser chaque observation par l'écart-type.

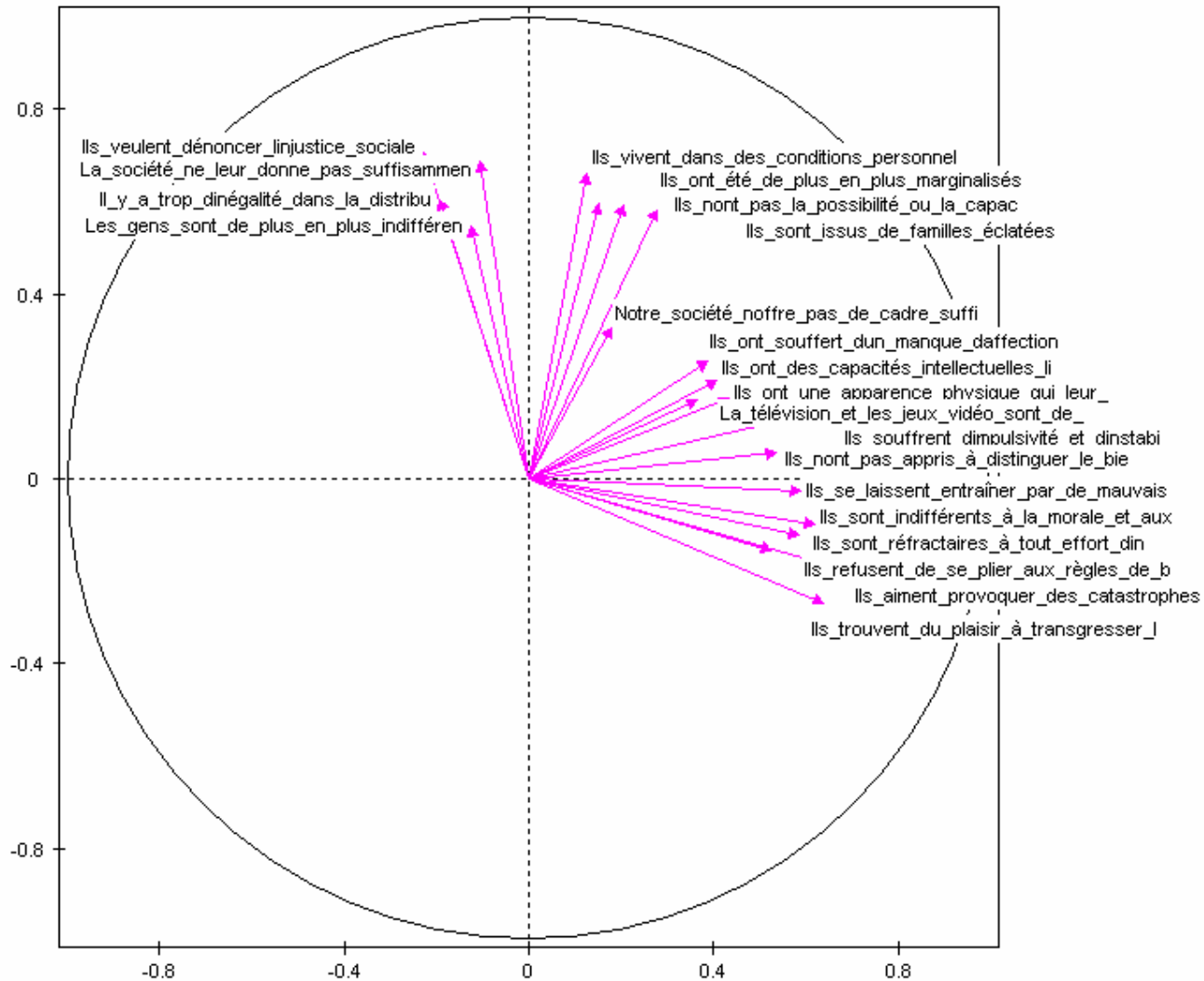
L'analyse en composantes principales

Matrice des corrélations										
Ils_ont_souffert_dun_manque_daffection	1,00									
Il_y_a_trop_dinégalité_dans_la_distribu	-0,05	1,00								
Notre_société_noffre_pas_de_cadre_suffi	0,17	0,14	1,00							
Ils_refusent_de_se_plier_aux_règles_de_b	0,10	-0,06	0,16	1,00						
La_société_ne_leur_donne_pas_suffisammen	0,19	0,36	0,21	-0,06	1,00					
Ils_ont_des_capacités_intellectuelles_li	0,19	-0,04	0,14	0,02	0,08	1,00				
Ils_souffrent_dimpulsivité_et_dinstabi	0,27	-0,10	-0,02	0,28	0,08	0,36	1,00			
Ils_sont_indifférents_à_la_morale_et_aux	0,07	-0,10	0,24	0,37	-0,15	0,25	0,27	1,00		
Ils_se_laissent_entraîner_par_de_mauvais	0,18	-0,10	-0,04	0,22	-0,08	0,23	0,39	0,41	1,00	
Ils_ont_été_de_plus_en_plus_marginalisés	0,22	0,32	0,26	-0,05	0,26	0,03	0,21	0,12	0,17	1,00
Ils_nont_pas_appris_à_distinguer_le_bie	0,15	-0,05	0,09	0,08	-0,06	0,17	0,32	0,34	0,44	0,10
Ils_aiment_provoquer_des_catastrophes	0,17	-0,19	0,03	0,45	-0,12	0,18	0,34	0,30	0,28	-0,11
Les_gens_sont_de_plus_en_plus_indifféren	-0,04	0,47	0,20	-0,17	0,39	-0,01	-0,02	-0,11	-0,13	0,22
Ils_trouvent_du_plaisir_à_transgresser_l	0,14	-0,15	0,04	0,41	-0,19	0,03	0,36	0,23	0,21	-0,07
Ils_ont_une_apparence_physique_qui_leur_	0,22	-0,03	0,20	0,02	-0,05	0,24	0,20	0,17	0,06	0,09
Ils_veulent_dénoncer_linjustice_sociale	0,04	0,47	0,14	-0,08	0,44	-0,01	0,00	-0,17	-0,11	0,38
Ils_sont_issus_de_familles_éclatées	0,32	0,22	0,13	-0,07	0,31	0,25	0,30	-0,05	0,09	0,29
Ils_sont_réfractaires_à_tout_effort_din	0,03	-0,07	-0,07	0,29	-0,15	0,08	0,29	0,45	0,29	0,06
Ils_nont_pas_la_possibilité_ou_la_capac	0,07	0,17	0,09	0,01	0,38	0,26	0,14	0,09	0,04	0,22
Ils_vivent_dans_des_conditions_personnel	0,11	0,30	0,06	0,00	0,33	0,13	0,12	-0,02	0,09	0,36
La_télévision_et_les_jeux_vidéo_sont_de_	0,26	-0,03	0,19	0,23	-0,02	0,08	0,19	0,08	0,06	0,13

Alpha de Cronbach = 0,76

L'analyse en composantes principales

Facteur 2 - 17.04 %



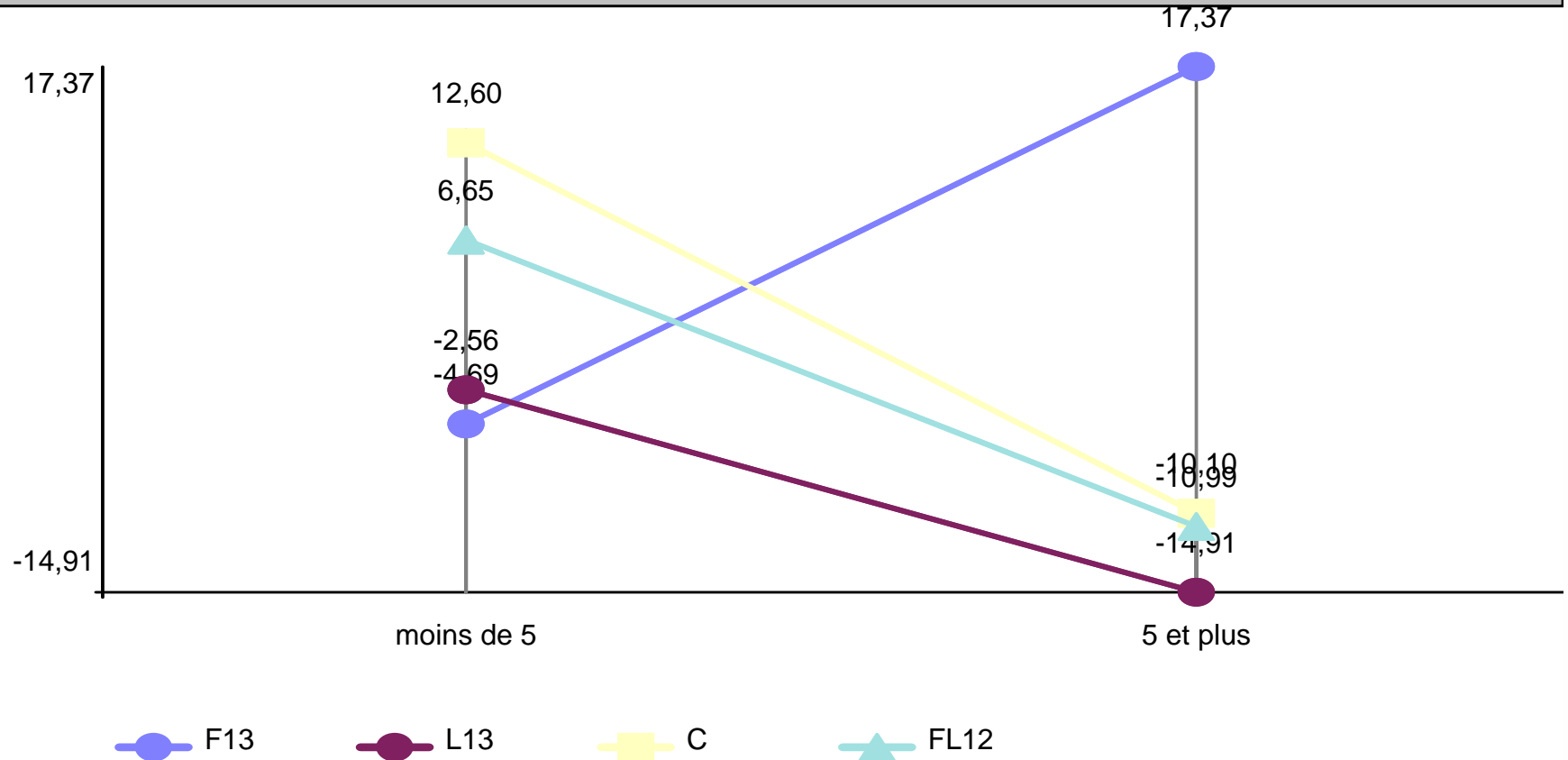
Facteur 1 - 18.66 %

L'analyse en composantes principales

Coordonnées des variables actives					
Libellé de la variable	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5
Ils veulent dénoncer linjustice sociale	-0,23	0,71	0,10	0,10	-0,01
La société ne leur donne pas suffisamment	-0,11	0,69	-0,05	0,06	0,05
Ils vivent dans des conditions personnelles	0,12	0,66	0,13	-0,01	0,31
Il y a trop dinégalité dans la distribution	-0,19	0,60	0,11	0,38	0,03
Ils ont été de plus en plus marginalisés	0,15	0,60	0,17	0,09	-0,27
Ils nont pas la possibilité ou la capacité	0,20	0,59	0,06	-0,01	0,35
Ils sont issus de familles éclatées	0,28	0,58	-0,20	-0,39	0,12
Les gens sont de plus en plus indifférents	-0,13	0,55	-0,02	0,35	0,09
Notre société noffre pas de cadre suffisant	0,18	0,33	-0,25	0,29	-0,69
Ils ont souffert dun manque daffectif	0,39	0,25	-0,32	-0,32	-0,26
Ils ont des capacités intellectuelles limitées	0,41	0,21	0,04	-0,45	-0,10
Ils ont une apparence physique qui leur déplaît	0,43	0,18	-0,28	-0,24	-0,07
La télévision et les jeux vidéo sont devenus des distractions	0,36	0,17	-0,54	0,16	0,03
Ils souffrent dimpulsivité et dinstabilité	0,66	0,15	0,07	-0,19	0,12
Ils nont pas appris à distinguer le bien du mal	0,54	0,06	0,45	-0,10	-0,12
Ils se laissent entraîner par de mauvais exemples	0,59	-0,03	0,44	-0,13	-0,10
Ils sont indifférents à la morale et aux valeurs	0,62	-0,10	0,34	0,23	-0,34
Ils sont réfractaires à tout effort disciplinaire	0,58	-0,12	0,37	0,25	0,24
Ils refusent de se plier aux règles de bon sens	0,52	-0,15	-0,14	0,49	0,02
Ils aiment provoquer des catastrophes	0,71	-0,21	-0,29	0,13	0,23
Ils trouvent du plaisir à transgresser les règles	0,64	-0,27	-0,29	0,25	0,23

L'analyse en composantes principales

'GROUPE' x 'POLITIQUE' x 'COORDONNEES F2'



Analyse de la variance à deux facteurs (17 répétitions) :

- L'effet principal de '*Groupe*' n'est pas significatif ($F = 0,46$, $p = ns$)
- L'effet principal de '*Politique*' n'est pas significatif ($F = 0,53$, $p = ns$)
- L'interaction est significative ($F = 2,96$, $p < .04$)

L'analyse factorielle des correspondances

L'AFC s'effectue sur des variables nominales (qualitatives).

La corrélation entre les points-lignes et entre les points-colonnes est calculée par la **distance du chi²**.

Distance entre points-lignes

$$d^2(i,i') = \sum_{j=1}^p \left(\frac{1}{f_{.j}} \right) \left(\frac{f_{ij}}{f_i} - \frac{f_{i'j}}{f_{i'}} \right)^2$$

Distances entre points-colonnes

$$d^2(j,j') = \sum_{i=1}^n \left(\frac{1}{f_{i.}} \right) \left(\frac{f_{ij}}{f_j} - \frac{f_{ij'}}{f_{j'}} \right)^2$$

Le principe repose sur la décomposition du tableau de données: On montre qu'un tel tableau peut être décomposé en d'autres tableaux autant de fois que le plus petit des nombres de ses lignes et de ses colonnes.

Il y a donc autant de facteurs que le plus petit des nombres de ses lignes et de ses colonnes

L'analyse factorielle des correspondances

La première décomposition est celle du tableau des effectifs théoriques.

Effectifs observés (T) :

n1	n2	M1
n3	n4	M2
M3	M4	N

Effectifs théoriques (T0)

$n'1=M1 \times M3/N$	$n'2=M1 \times M4/N$	M1
$n'3=M2 \times M3/N$	$n'4=M2 \times M4/N$	M2
M3	M4	N

« En analyse des correspondances, le premier facteur correspond au tableau d'indépendance. Comme en général on s'intéresse (...) aux écarts à l'indépendance, on ne prend pas en compte ce facteur et on l'appelle facteur zéro, réservant le numéro un pour la première approximation des écarts à l'indépendance » (Cibois, 1994, p.59).

Il y a donc autant de facteurs à retenir que le plus petit des nombres de ses lignes et de ses colonnes -1.

L'analyse factorielle des correspondances

Le deuxième tableau est celui des différences entre effectifs observés et effectifs théoriques (tableau des « restes »). Il y a donc des valeurs « positives » (« attractions » entre lignes et colonnes) et des valeurs « négatives » (« répulsions » entre lignes et colonnes).

R1 = Eff. théoriques - Eff. observés

n1-n'1	n2-n'2	M'1
n3-n'3	n4-n'4	M'2
M'3	M'4	

Ce tableau peut être décomposé en « vecteurs marginaux », c'est-à-dire en suites de valeurs représentant ses marges : on aura donc un vecteur horizontal composé d'une suite de valeurs représentant les colonnes (M'3, M'4...), et un vecteur vertical composé d'une suite de valeurs représentant les lignes (M'1, M'2...).

Ce tableau peut, à son tour, être décomposé (T1, R2), et ainsi de suite (T2, T3, etc.) jusqu'au plus petit des nombres de ses lignes et de ses colonnes -1 (principe de convergence mathématique).

L'analyse factorielle des correspondances

Interprétation:

- Inertie ou variance expliquée: L'information totale (100%) est représentée par la valeur du Chi^2 du tableau initial (T). La variance de chaque facteur est calculée à partir du Chi^2 de chaque tableau reconstitué (T1, T2, etc.). La somme des Chi^2 des sous-tableau est égale au Chi^2 du tableau T. L'inertie de chaque facteur est exprimée en pourcentage du Chi^2 du tableau T.

- Contributions absolues ou corrélations *variables * facteurs* décrivent la part prise par un élément (ligne ou colonne) dans la construction d'un axe factoriel. On les détermine en divisant la marge correspondante par le Chi^2 du tableau. Pour chaque facteur, la somme des contributions est égale à 1, et on retiendra, pour l'interprétation d'un facteur, les variables dont les contributions sont les plus fortes.

NB. la contribution d'une variable illustrative est donc nulle.

L'analyse factorielle des correspondances

- Cosinus carrés ou contribution relative est la part prise par chaque facteur pour expliquer une variable, ou encore la qualité de la représentation de chaque forme par les axes. Pour chaque variable, la somme des \cos^2 (en ligne) est égale à 1.
- Coordonnées factorielles et graphiques: La représentation graphique s'établit sur la base des vecteurs propres de T1, T2, etc.
- L'AFC peut être appliquée à un tableau croisant des variables (colonnes) et, soit des individus, soit du lexique (lignes).