



Méthodes quantitatives des sciences sociales

5. Les sciences sociales face à la multi-dimensionnalité

Sciences Po Saint-Germain-en-Laye, 1ère année

2016-2017



Introduction

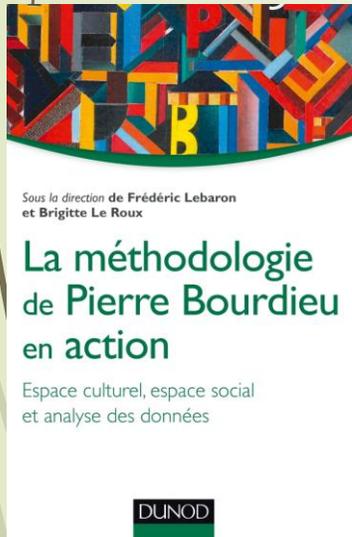
- Dans la réalité sociale, quel que soit le phénomène étudié, un très grand nombre de variables sont à l'œuvre simultanément. Notion d'**interdépendance**.
- C'est d'autant plus le cas que les phénomènes étudiés sont « multidimensionnels » (cf. la notion d'indicateur selon Lazarsfeld).
- Exemple de la « qualité de vie »: de nombreux facteurs sont susceptibles d'affecter cet objet. Idée de **décomposition** d'un phénomène en grandeurs différentes.
- On dispose rarement de la possibilité d'**isoler un facteur** de façon expérimentale, comme dans les sciences de la nature => recours à des modélisations statistiques.



Introduction

- Idée principale de cette cinquième séance: comment les méthodes statistiques utilisées dans les sciences sociales permettent-elles de traiter de phénomènes multidimensionnels et multivariés, sans avoir recours à l'expérimentation ?
- **Première stratégie:** le raisonnement « toutes choses égales par ailleurs » (modélisation, régression) = chercher l'effet « pur » d'une variable (dite indépendante) sur une autre (dépendante).
- **Deuxième stratégie:** l'analyse géométrique des données = réduire la dimensionnalité d'un phénomène, en dégagant ses facteurs (analyse factorielle), ses composantes ou dimensions.

Bibliographie



- Brigitte Le Roux, *Analyse géométrique des données multidimensionnelles*, Paris, Dunod, 2014.
- Frédéric Lebaron, Brigitte Le Roux (ed.), *La méthodologie de Pierre Bourdieu en action. Espace culturel, espace social et analyse des données*, Paris, Dunod, 2015.
- François Pétry, François Gélinau, *Guide pratique d'introduction à la régression en sciences sociales. Deuxième édition revue et augmentée*, Laval, Presses de l'université de Laval, 2009.



Webographie

- Site de Brigitte Le Roux: <http://www.math-info.univ-paris5.fr/~lerb/>
- DARES, La mesure des écarts de salaires entre les femmes et les hommes: http://travail-emploi.gouv.fr/IMG/pdf/La_mesure_des_ecarts_de_salaires_entre_femmes_et_hommes.pdf



1. Le raisonnement « toutes choses égales par ailleurs »

- Faute d'expérimentation, on essaie néanmoins de faire comme si l'on pouvait isoler l'effet de chacune des variables qui détermine un phénomène
- ⇒ C'est la recherche de l' « effet pur », une fois « contrôlées » les autres variables.
- Une distinction fondamentale: variable dépendante (expliquée) / variables indépendantes (explicatives). Exemple: niveau de bonheur subjectif des individus enquêtés. Il dépend de nombreuses variables : revenu, profession, âge, sexe, statut d'emploi, etc. **Modèle**.
- Les effets de chaque variable sont assimilés aux coefficients standardisés de l'équation de régression.
- En pratique, on regarde surtout le signe des effets et leur significativité statistique (inférence sur les coefficients).



1. Le raisonnement « toutes choses égales par ailleurs »

- Quel **type** de modèle de régression choisir ?
- Variables numériques => Régression linéaire multiple (généralisation de la régression simple à n variables)
- Variables catégorisées => Soit on transforme les variables catégorisées en variables numériques, soit on utilise un autre type de modèle (par exemple, le modèle logistique ou modèle Logit).
- Démarche: l'ajustement conduit à une estimation des **coefficients partiels** de la régression.
- Régression linéaire: (le plus souvent) ajustement par la méthode des **moindres carrés ordinaires** (OLS)
- Régression logistique (Logit): ajustement par la méthode du **maximum de vraisemblance**.

1. Le raisonnement « toutes choses égales par ailleurs »

Exemple 1 : les déterminants du salaire et l'effet du sexe.

- Les femmes ont globalement des salaires inférieurs à ceux des hommes (effet brut ou global du sexe sur les salaires)... Mais...
- ⇒ Elles ne travaillent pas dans les mêmes secteurs d'activité ;
- ⇒ Elles n'occupent pas les mêmes types de profession ;
- ⇒ Elles n'occupent pas les mêmes positions hiérarchiques dans l'entreprise ;
- ⇒ Elles travaillent plus souvent à temps partiel...
- ⇒ Différences dans le métier pas toujours observables...

On peut expliquer une partie de la différence globale de rémunération par ces **effets de structure**.

Ce qui « reste » est un effet net ou conditionnel (ou « TCEPA »).
Discrimination ? En partie.



1. Le raisonnement « toutes choses égales par ailleurs »

Exemple 2: déterminants de la « réussite scolaire » (exemple: accès sans année de retard au baccalauréat) en fonction de l'origine sociale et géographique. Louis-André Vallet, Jean-Paul Caille, « La scolarité des enfants d'immigrés », in VAN ZANTEN A. (dir.), **L'école : l'état des savoirs**, Paris, La Découverte, pp. 293-301.

- Données longitudinales : Panel national 1989 d'élèves du second degré, ministère de l'Éducation nationale, DPD.



« L'analyse longitudinale des parcours que connaissent les enfants d'immigrés dans le système scolaire français conduit à trois conclusions principales parmi lesquelles seule la première apparaît largement connue.

D'une part, les enfants d'immigrés comptent parmi les élèves qui encourent les plus grands risques de difficultés ou d'échec scolaires, d'orientation vers les filières peu prestigieuses du système éducatif comme de sortie précoce de celui-ci.

D'autre part, lorsque l'analyse statistique est utilisée pour démêler l'écheveau des influences et séparer ce qui tient en propre à l'origine nationale de ce qui relève des caractéristiques objectives du milieu familial et social, cette première conclusion doit être **renversée** : au sein des populations défavorisées, les enfants d'immigrés sont en moyenne inscrits dans une trajectoire scolaire plus positive que les autres élèves. Enfin, l'explication principale de ce dernier résultat est à rechercher dans les aspirations éducatives plus fortes et les demandes d'orientation plus ambitieuses qu'expriment les familles immigrées, comparativement aux autres familles dotées des mêmes ressources matérielles et culturelles » (Louis-André Vallet, Jean-Paul Caille, « La scolarité des enfants d'immigrés », in VAN ZANTEN A. (dir.), **L'école : l'état des savoirs**, Paris, La Découverte, p. 293).



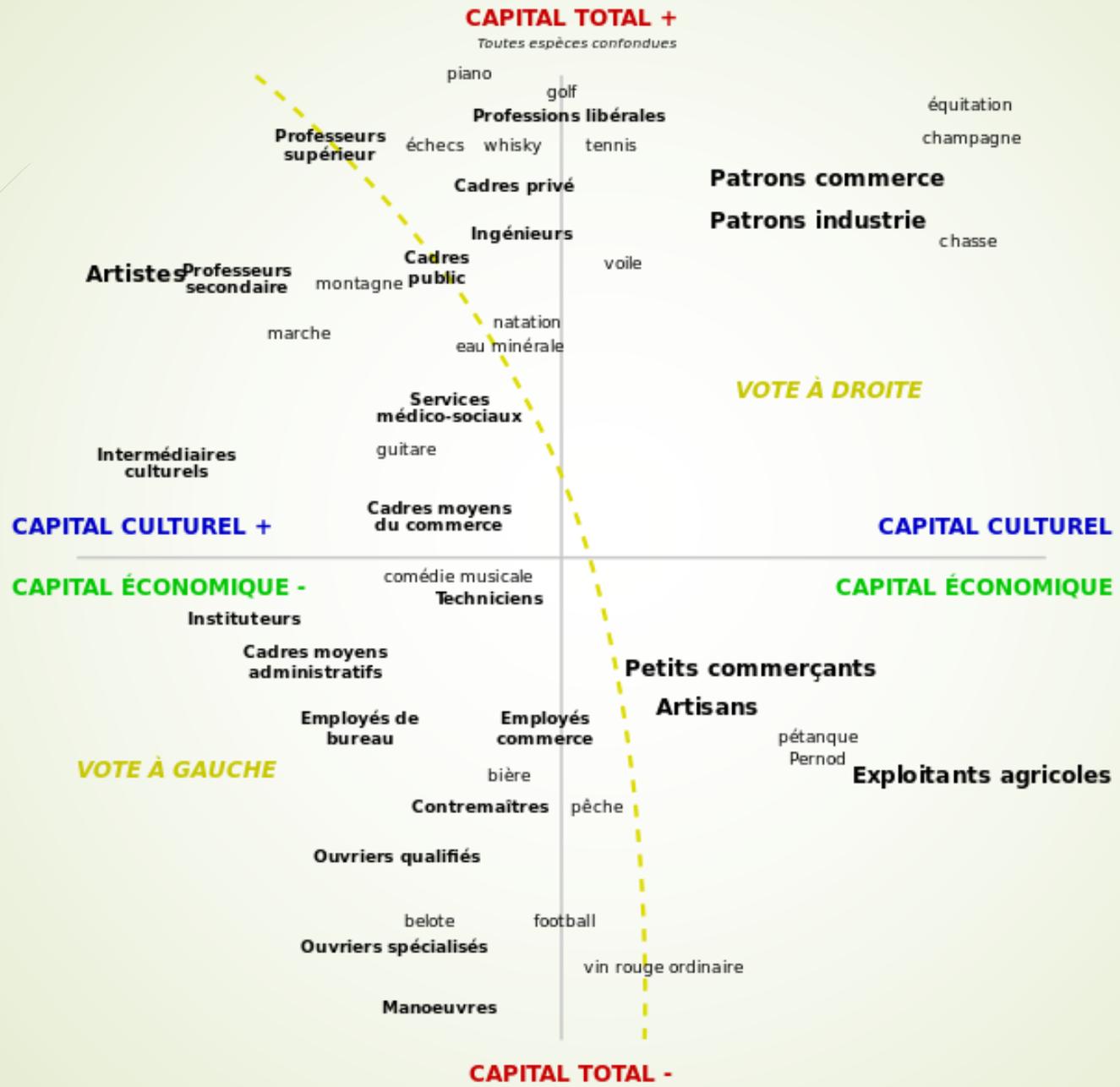
Comparer les effets bruts et les effets conditionnels (régression simple et régression multiple)

Plusieurs situations peuvent théoriquement se présenter lorsque l'on compare les coefficients standardisés pour la même variable indépendante, dans le cas simple (seule variable indépendante) et multiple:

- Stabilité (cas très fréquent)
- Disparition
- Apparition (cas rarissime)
- Atténuation (exemple des salaires)
- Accentuation
- Renversement (exemple de la réussite des enfants d'immigrés)

2. L'analyse géométrique des données

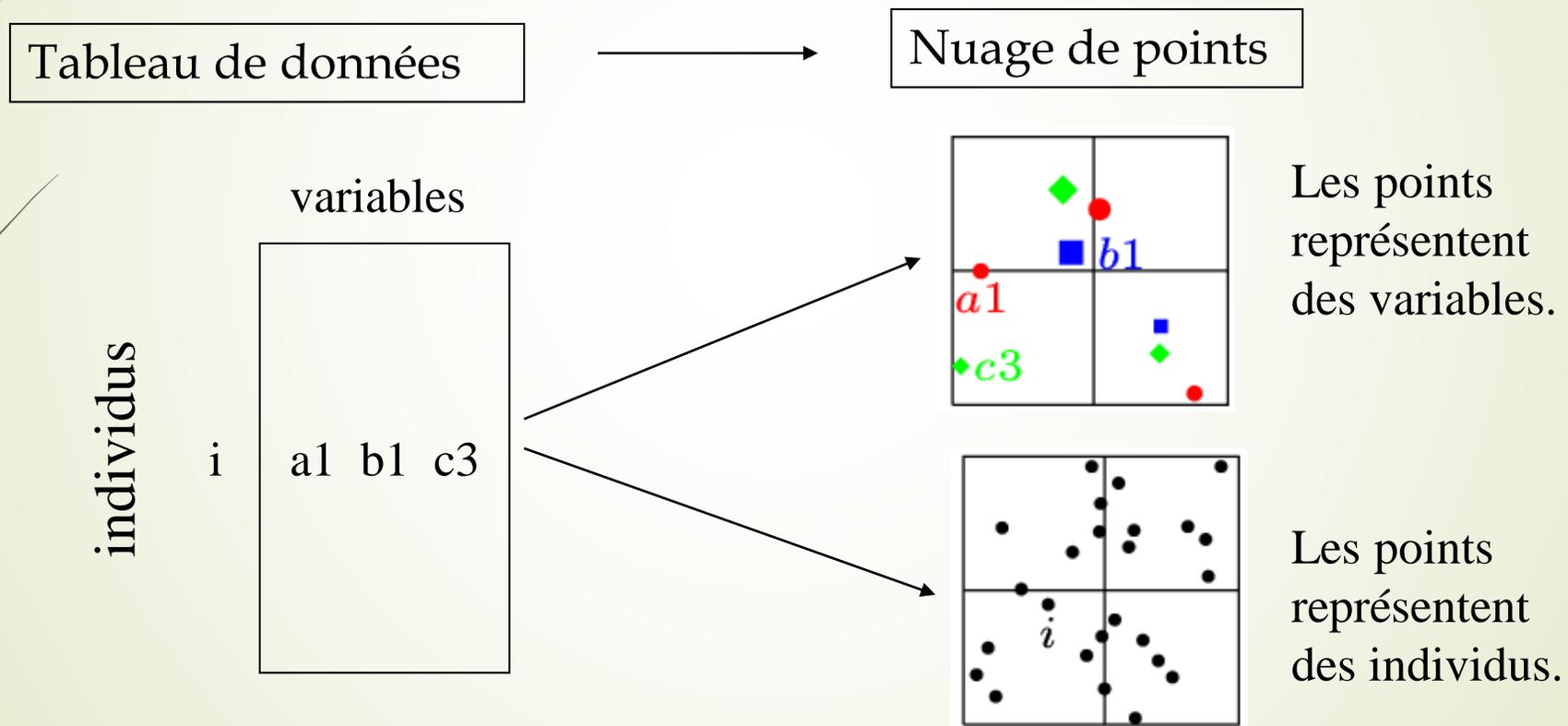
- Années 1960: augmentation des possibilités informatiques, nouveaux développements mathématiques dans les sciences humaines et sociales.
- Développement de méthodes permettant de traiter de tableaux de grande taille (beaucoup de lignes, beaucoup de colonnes) => en fournir un bon résumé visuel sous la forme d'un **nuage de points**. Cela suppose une définition de la **distance** entre les lignes (resp. les colonnes).
- « Bon » résumé: ajustement d'un nuage de points multidimensionnel par un nuage de plus petite dimension. **Moindre carrés orthogonaux**. Maximisation de la variance du nuage projeté.
- Utilisation dans les sciences sociales: de plus en plus importante jusqu'à aujourd'hui. Exemple de P.Bourdieu et des structures de l'espace social.



Qu'est-ce que l'AGD ?

14

Les données sont représentées sous forme de nuages de points.
L'interprétation statistique est fondée sur les nuages.



AGD

04/05/2017

Trois idées clés

15

Statistique classique	AGD
Quantitative (numérique)	Géométrique (multidimensionnelle)
Approche matricielle	Approche formelle (algèbre linéaire)
Orientée échantillonnage	Démarche inductive

- modélisation géométrique
- « finalement, faire une analyse des données, en bonne mathématique, c'est simplement chercher des vecteurs propres; toute la science (ou tout l'art) consiste à diagonaliser la bonne matrice » (Benzécri).
- la description d'abord ! Le modèle doit suivre les données et non l'inverse.

Trois paradigmes de l'AGD

16

Les méthodes de l'AGD furent introduites par Benzécri, avec l'analyse des correspondances (AC).

Elles sont bien connues en France sous l'appellation «Analyse des données».

tableau de contingence → analyse des correspondances (AC)

Tableau individus x variables :

variables numériques → analyse en composantes principales (ACP)

variables catégorisées → analyse des correspondances multiples (ACM)



Principes méthodologiques

17

L'AGD comme modèle cadre

- Homogénéité : toutes les grandeurs recensées dans le tableau sont des quantités de même nature.
- Exhaustivité : les marges du tableau représentent un inventaire complet d'un dossier réel dont le cadre n'est guère discutable.
- Construction de l'espace.
- Accent mis sur les individus.
- Schèmes explicatifs; analyse géométrique de questionnaires.



Analyse des donnée structurées

18

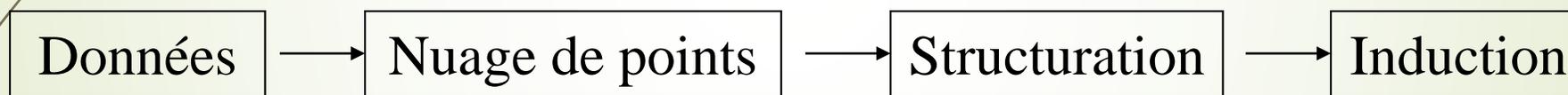
- Facteurs structurants
- Données structurées = données avec facteurs structurants



Inférence statistique en AGD

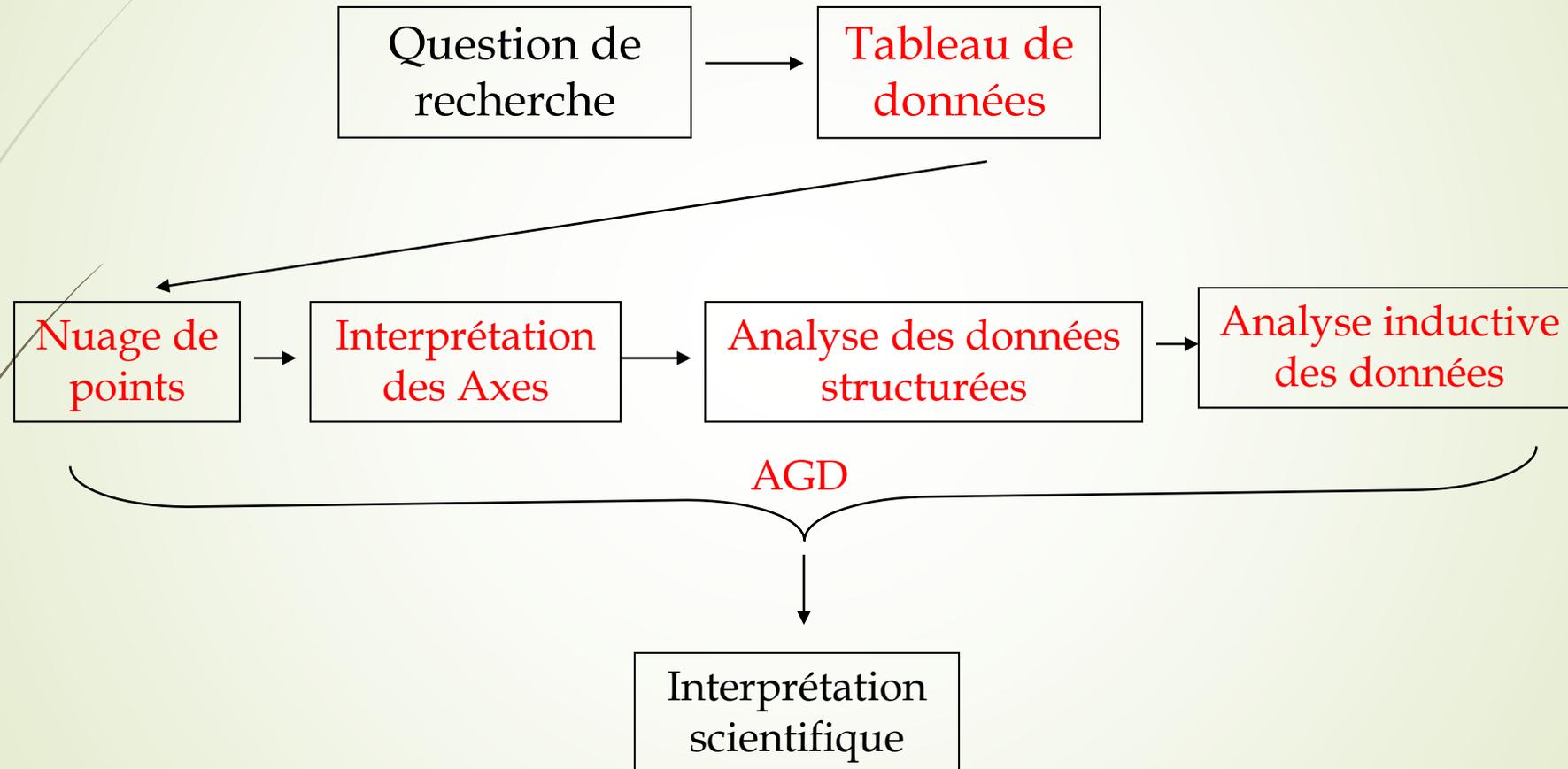
19

- L'inférence statistique doit être effectuée comme une analyse inductive des données, c'est-à-dire dans l'esprit de l'analyse des données, c'est-à-dire comme prolongement naturel de l'analyse descriptive.



Le modèle doit suivre les données, non l'inverse !

En résumé: qu'est-ce que l'AGD ?



Un exemple commenté

21

Jeu du portrait chinois

Hommes politiques et *Couleurs*

	<i>Blanc</i>	<i>Noir</i>	<i>Bleu</i>	<i>Orange</i>	<i>Jaune</i>	<i>Vert</i>
Giscard d'Estaing	35	10	29	6	9	12
Poniatowski	16	22	14	16	18	13
Chirac	16	9	25	12	18	18
Servan-Schreiber	14	9	12	23	23	19
Mitterrand	13	10	13	23	18	24
Marchais	6	40	7	20	14	14

◇ *Source : Sondages (1975), 3,4, 31-47. Référence : P. BOURDIEU (1979) : La distinction, p. 625-640. On trouvera dans ce texte les tableaux correspondant à dix autres questions (chapeaux, métiers, personnages de bandes dessinées, etc.).*

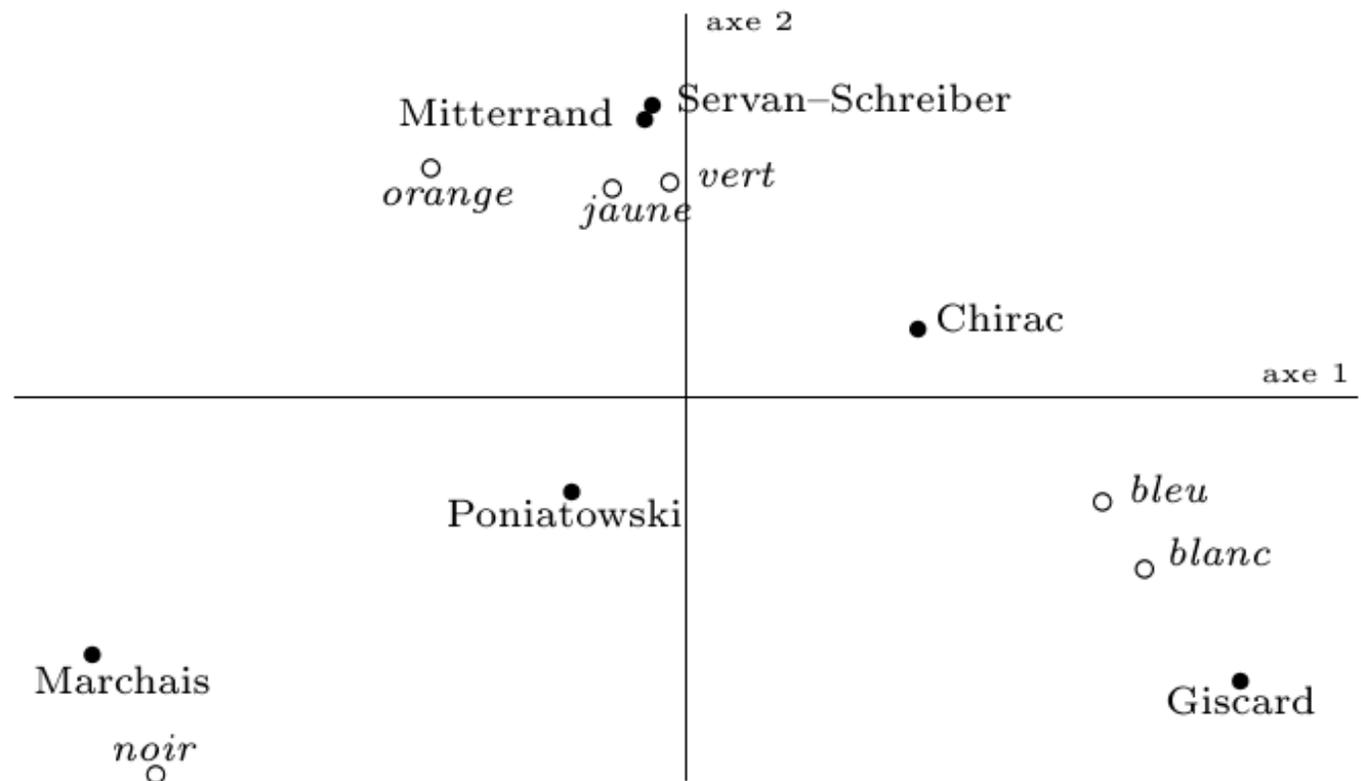
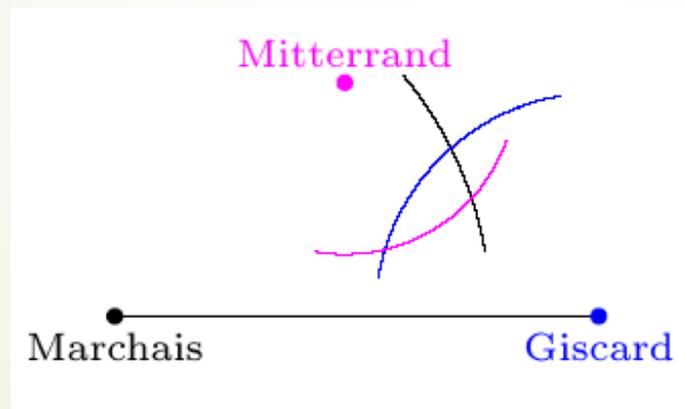
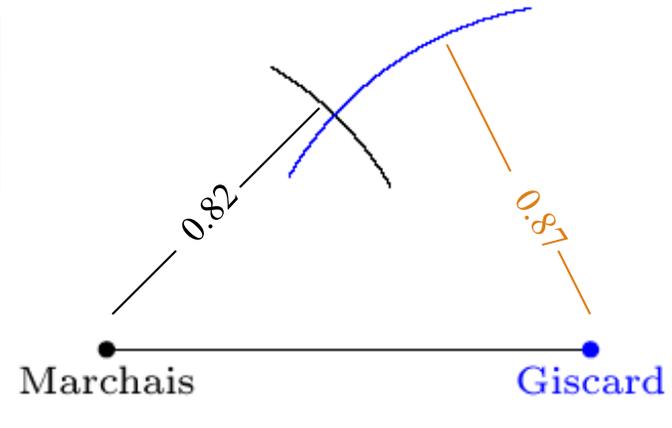
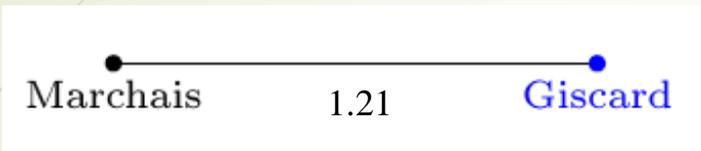


Diagramme plan simultané

	<i>Blanc</i>	<i>Noir</i>	<i>Bleu</i>	<i>Orange</i>	<i>Jaune</i>	<i>Vert</i>
Giscard d'Estaing	.35	.10	.29	.06	.09	.12
Marchais	.06	.40	.07	.20	.14	.14
Différence	.29	-.30	.22	-.14	-.05	-.02

$$d^2(\text{Giscard, Marchais}) = \frac{(.35 - .06)^2}{1/6} + \dots + \frac{(.12 - .14)^2}{1/6} = 1.4700$$

	Giscard	Mitterrand	Chirac	Poniatowski	Servan-Sch
Marchais	1.21	0.82	0.94	0.55	0.84
Giscard		0.87	0.56	0.74	0.87
Mitterrand			0.43	0.44	0.18



- Les valeurs propres (variances des axes)

Trace de la matrice: 0.20358

Nu- méro	valeur propre	Pourcentage	Pourcentage cumulé
1	0.1332	65.41	65.41
2	0.0592	29.10	94.51
3	0.0071	3.47	97.98
4	0.0041	2.00	99.98
5	0.0000	0.02	100.00

- Les coordonnées principales

	<i>Axe1</i>	<i>Axe2</i>
Giscard	+0.578	-0.297
Poniatowski	-0.119	-0.099
Chirac	+0.242	+0.072
Servan-Schreiber	-0.035	+0.306
Mitterrand	-0.043	+0.291
Marchais	-0.619	-0.269

	<i>Axe1</i>	<i>Axe2</i>
Blanc	+0.478	-0.180
Noir	-0.553	-0.394
Bleu	+0.434	+0.109
Orange	-0.266	+0.240
Jaune	-0.077	+0.219
Vert	-0.017	-0.224

- 1) Métrique du khi-deux
- 2) Représentation géométrique
- 3) Contraction des distances
- 4) Equivalence distributionnelle
- 5) Individus empiriques, individus épistémiques
- 6) Petits et grands tableaux
- 7) Aides à l'interprétation

<i>Ctr</i>	<i>Axe1</i>	<i>Axe2</i>	<i>Ctr</i>	<i>Axe1</i>	<i>Axe2</i>
Giscard	.42	.25	Blanc	.29	.09
Ponia	.02	.03	Noir	.38	.44
Chirac	.07	.01	Bleu	.24	.03
ServanSch.	.00	.26	Orange	.09	.16
Mitterrand	.00	.24	Jaune	.01	.14
Marchais	.48	.21	Vert	.00	.14

<i>Qlt</i>	<i>Axe1</i>	<i>Axe2</i>	<i>Qlt</i>	<i>Axe1</i>	<i>Axe2</i>
Giscard	.78	.21	Blanc	.82	.12
Ponia	.45	.31	Noir	.66	.34
Chirac	.61	.05	Bleu	.87	.05
ServanSch.	.01	.90	Orange	.53	.43
Mitterrand	.02	.86	Jaune	.09	.69
Marchais	.84	.16	Vert	.00	.82

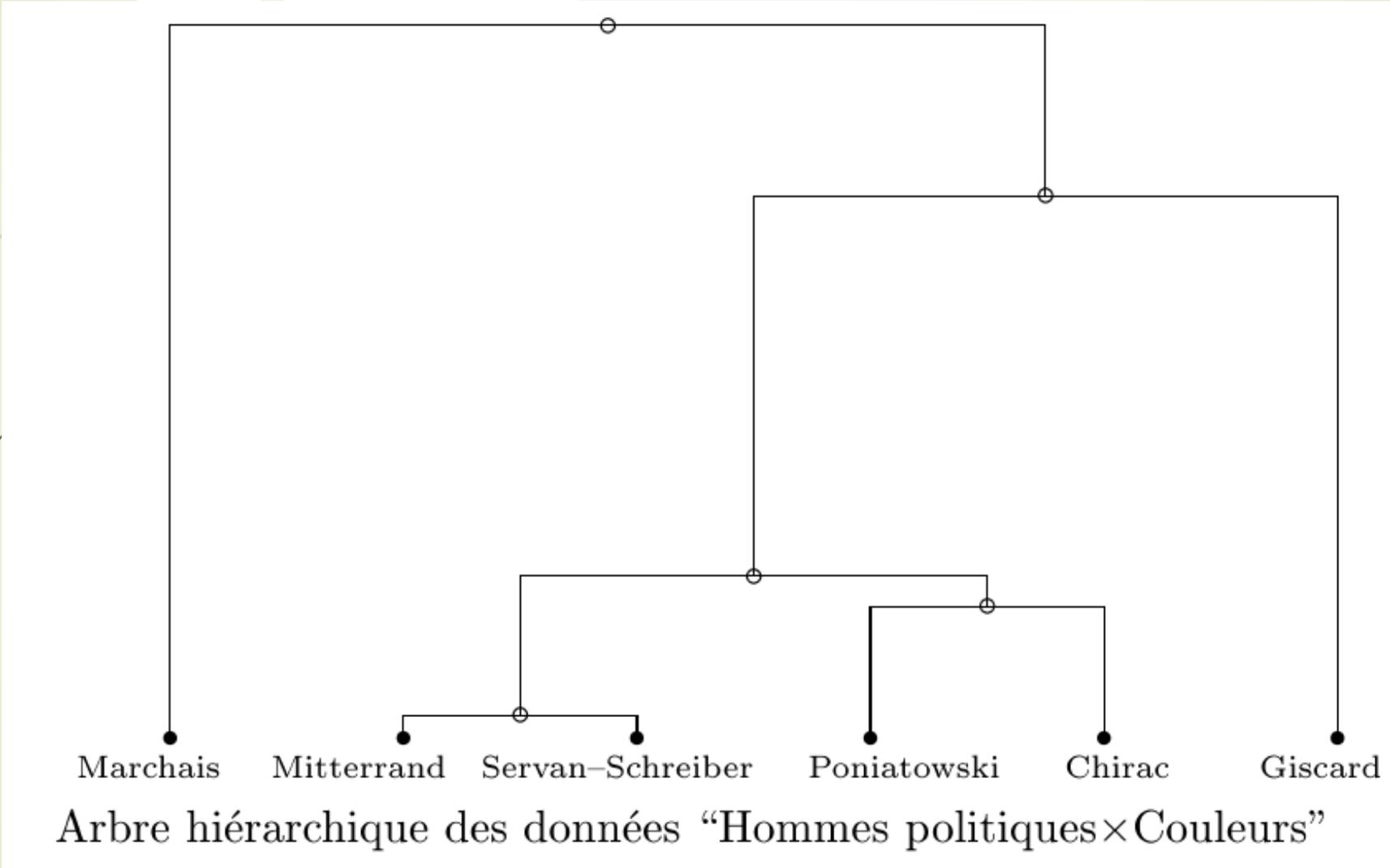
- 8) Points supplémentaires



AGD et classification

- On complète les méthodes d'analyse des données par des méthodes de classification automatique. Il en existe de nombreuses, des associations de classification existent partout dans le monde, et développent de nombreuses méthodes.
- Exemple proche de l'AGD: la classification ascendante hiérarchique ou CAH. Il s'agit de regrouper en partant de l'ensemble des individus statistiques les individus les plus « ressemblants » (selon un critère bien défini) dans des classes les plus « distinctes » possibles. On procède par **regroupements** successifs de points, jusqu'à ne plus avoir plus qu'un seul point: on obtient un arbre hiérarchique.
- On détermine ensuite le nombre de classes le plus pertinent à partir de l'arbre.
- On représente les classes dans le nuage issu de l'AGD.

Classification



AC & CAH

29

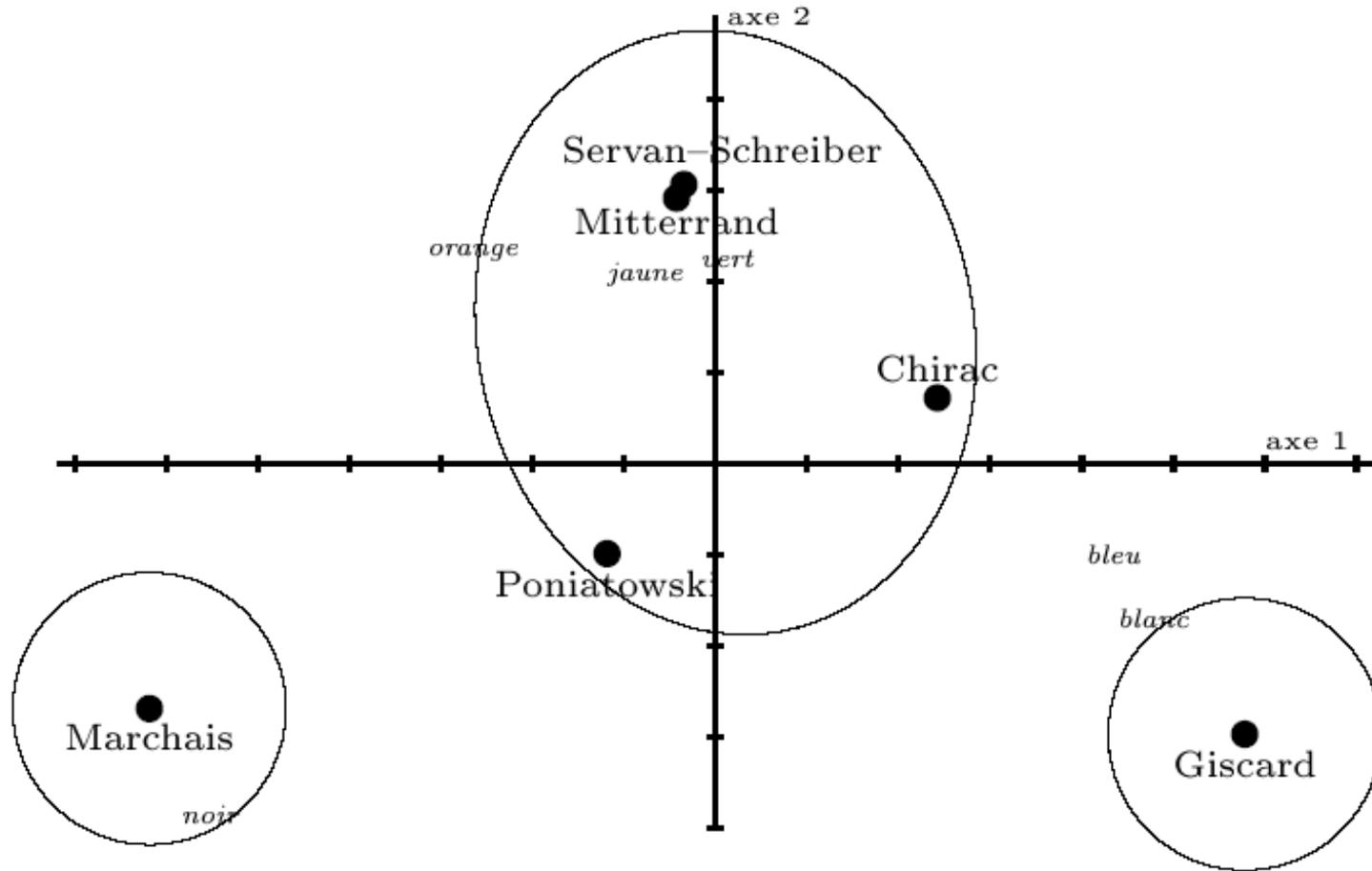
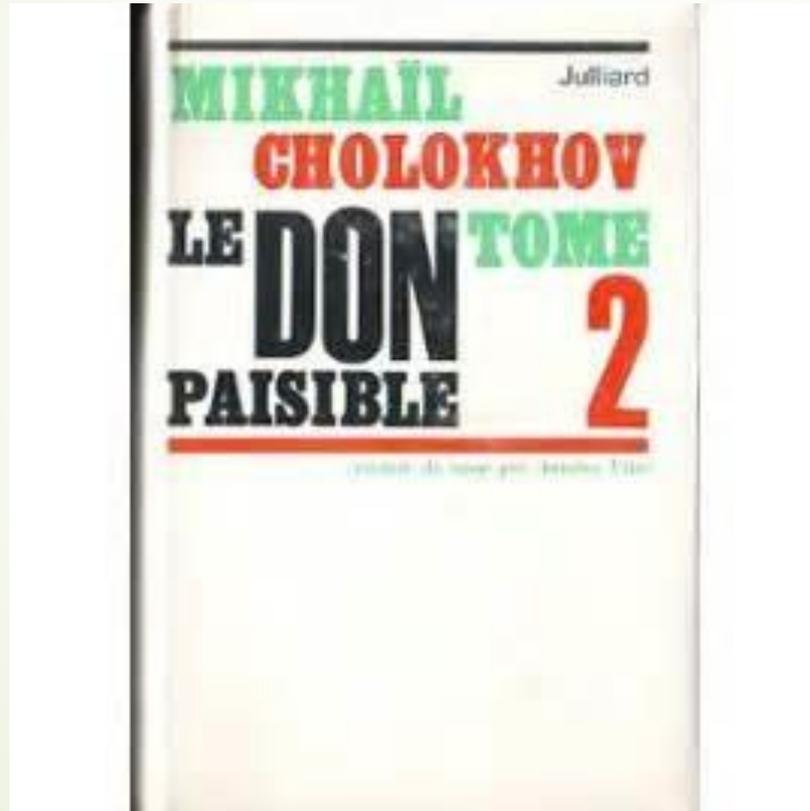


Diagramme simultané de l'analyse des correspondances et les 3 classes de la classification

AC et lexicométrie: l'exemple de l'affaire du Don paisible



L'ouvrage

En 1928, Mikhaïl Cholokhov publie le premier tome de ce qui sera son œuvre majeure, [Le Don paisible](#). Il s'agit d'un roman sous forme d'épopée qui dépeint la vie dans la région du [Don](#) au cours de la période de la [Première Guerre mondiale](#) et de la Guerre civile. Le roman, qui comprend quatre tomes, est centré sur le personnage de Grigori Melekhov, jeune officier cosaque. L'action commence en 1912, alors que la région est encore paisible et prospère, se poursuit avec la mobilisation de Grigori et ses proches au déclenchement de la Première Guerre mondiale à l'été 1914. Il décrit les premières escarmouches contre l'armée allemande, et la détérioration progressive de la situation à mesure que le conflit s'enlise. Avec la [Révolution d'octobre](#) 1917 et le retrait des troupes russes du conflit, Melekhov est démobilisé et cherche à reprendre la vie agricole, mais il est rapidement rattrapé par la Guerre civile. Il changera de camp à quelques reprises, assistera et participera à de nombreux combats et massacres, avant de finir dans le camp des perdants en 1922. (Source: Wikipedia).

La polémique

- ▶ Encensé par la critique officielle, le roman est traduit en de nombreuses langues et diffusé massivement avec l'aide des presses de l'État. Il est présenté comme la première grande œuvre de la nouvelle littérature soviétique. Ceci causera une réaction à partir des années 1960, et en particulier suite à l'attribution du [prix Nobel de littérature](#) à Choukhov en 1965, mais la question du [plagiat](#) s'est posée dès 1929.
- ▶ Plusieurs critiques, dont [Alexandre Soljenitsyne](#) et [Roy Medvedev](#), expriment leurs doutes quant à l'identité de l'auteur véritable du roman. Leurs arguments sont fondés sur le fait que Choukhov est trop jeune pour avoir été témoin des événements décrits ; qu'il n'a pas le niveau d'éducation requis pour avoir produit un tel chef-d'œuvre ; et que les positions officielles de l'écrivain et la qualité du reste de sa production littéraire ne s'accordent pas avec le traitement sympathique et objectif réservé aux Cosaques dans le roman. Ils avancent la thèse que le véritable auteur en serait [Fiodor Krioukov](#), écrivain cosaque et anti-bolchévique décédé de fièvre typhoïde en 1920. Cette controverse a longtemps fait rage, mais semble avoir trouvé aujourd'hui sa conclusion définitive.

L'ensemble des textes

Cholokov

- ▶ Le Don paisible (D1,D2,...D8)
- ▶ Terres défrichées (TS1,TS2)
- ▶ Ils se sont battus pour la patrie (R)
- ▶ Le destin d'un homme (S)

Krioukov

- ▶ Cinq courtes histoires (K1,...K5)

L'ensemble des mots

Les « mots-outils » permettent d'étudier le style:

- Prépositions
- Adverbes, adjectifs utilisés comme adverbes
- Conjonctions
- Pronoms démonstratifs
- Adjectifs possessifs

Tableau de données

35

	Mot1	Mot2	Mot3	...	Mot224
<i>D1</i>					
<i>D2</i>					
<i>D3</i>					
<i>D4</i>					
<i>D5</i>					
<i>D6</i>					
<i>D7</i>					
<i>D8</i>					
<i>TS1</i>					
<i>TS2</i>					
<i>R</i>					
<i>S</i>					
<i>K1</i>					
<i>K2</i>					
<i>K3</i>					
<i>K4</i>					
<i>K5</i>					

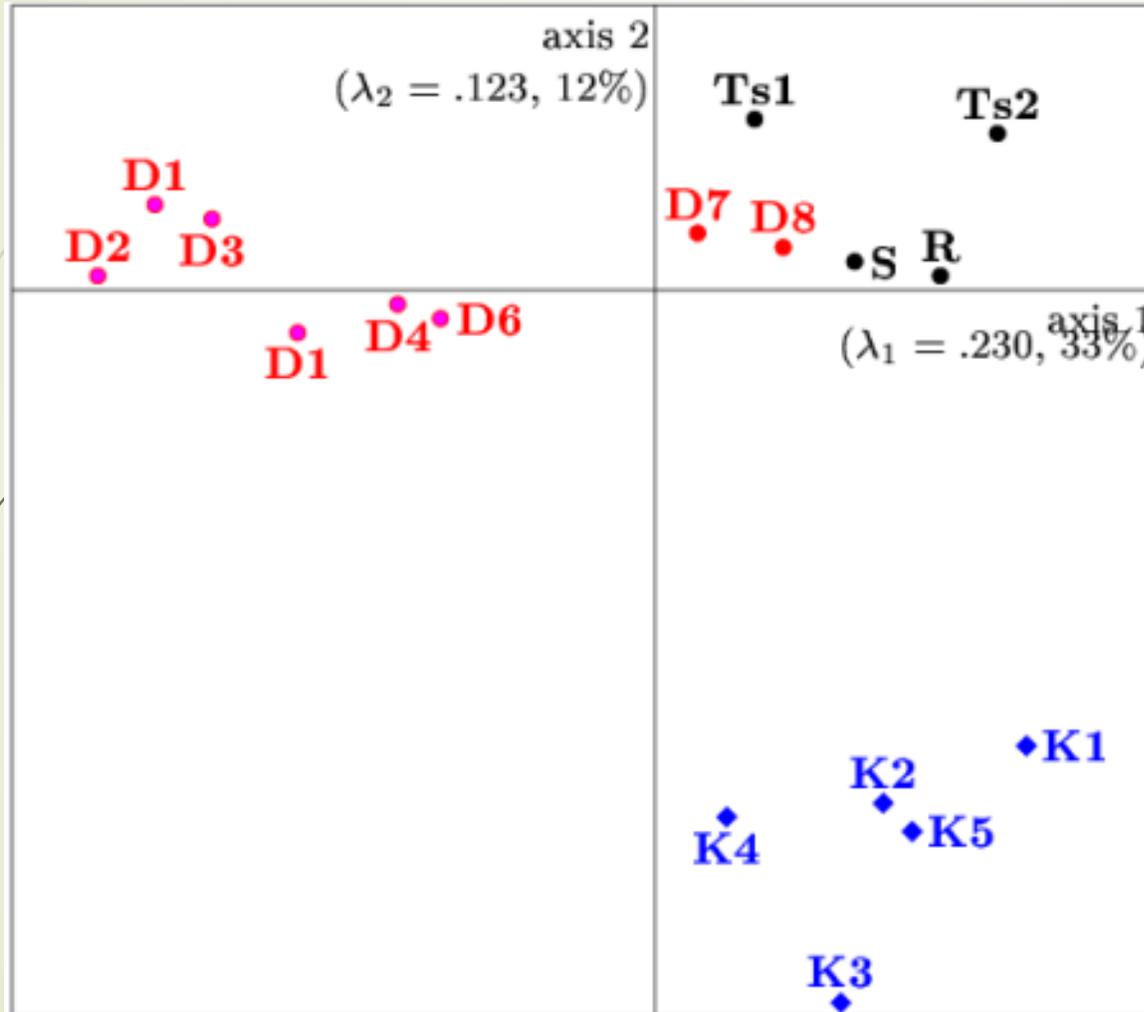
Nombre
d'occurrences
du mot 3 dans
le texte

TS2



Résultats de l'AC: Nuage des 17 textes

36



Cholokov:

Le Don paisible

(**D1-D8**);

Terres défrichées

(**Ts1-Ts2**);

Ils se sont battus pour la patrie

(**R**);

Le destin d'un homme (**S**).

K.D. Kryukov: 5 courtes histoires (**K1-K5**)

Plagiat ou pas ?

Référence: Volochine G. & Volochine A. Étude comparée de textes russes : le Don Tranquille et d'autres œuvres de M. A. Cholokhov ; les nouvelles de F. D. Kriukov. *Les cahiers de l'analyse des données*, 20 no. 1 (1995), p. 7-26 .

http://www.numdam.org/item?id=CAD_1995__20_1_7_0

Conclusion de l'histoire

En 1991, le journaliste russe [Lev Kolodny](#) (Лев Ефимович Колодный) a retrouvé le manuscrit original du *Don paisible* ainsi que d'autres papiers de travail, qui ont ensuite été formellement authentifiés par un groupe d'experts graphologues comme étant bien de la main de Cholokhov.



3. Introduction aux logiciels de statistique.

- Excel / OpenOffice Calc: stockage des données, statistiques descriptives élémentaires, régression simple, tests de signification basiques (exemple: test du Chi-deux), graphiques.
- Les logiciels de statistiques permettent d'effectuer des transformations sur les données (codages, création de nouvelles variables...);
- Ils permettent aussi d'exécuter des procédures statistiques plus sophistiquées et fournissent beaucoup plus de résultats et de graphiques.

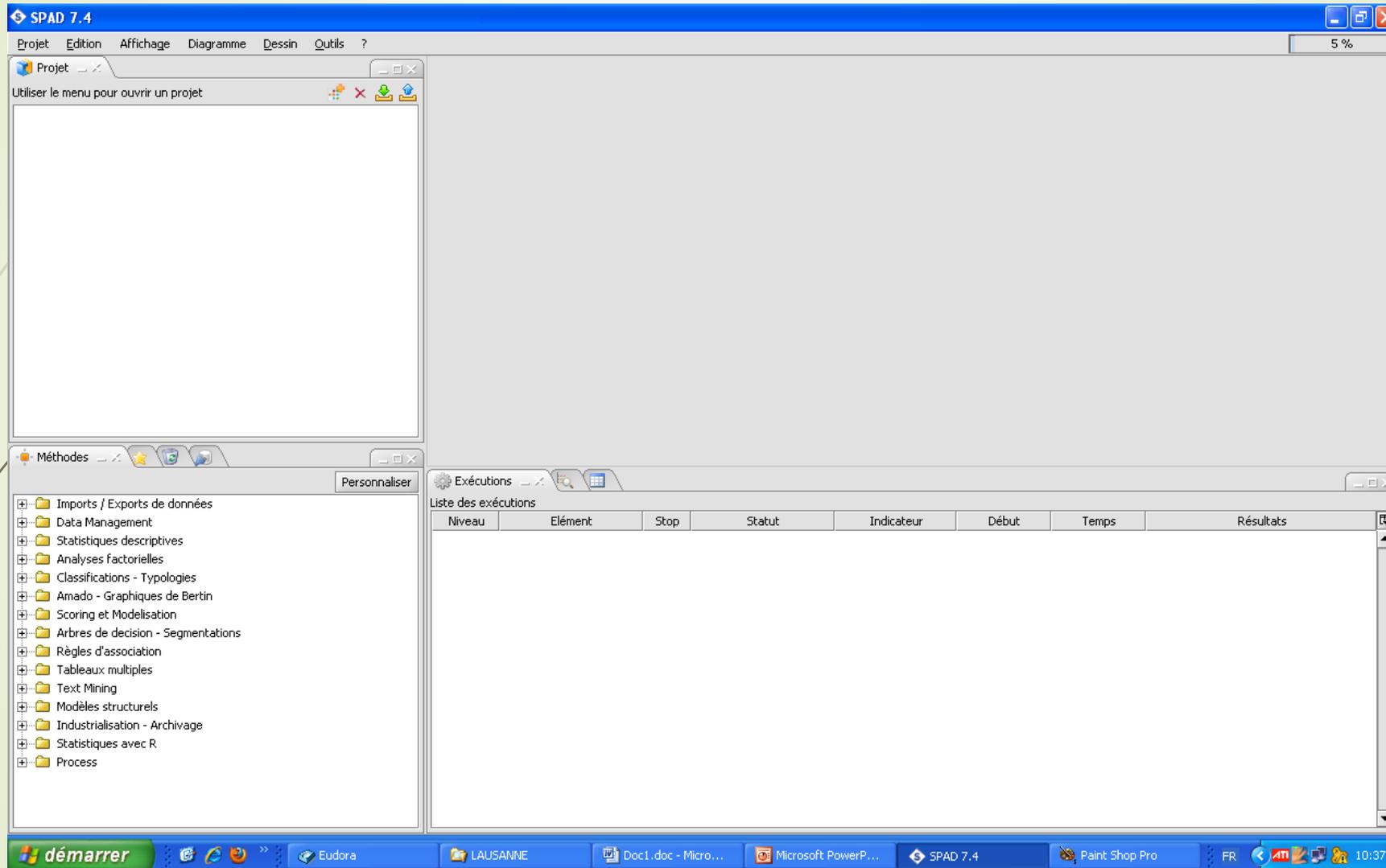


3. Introduction aux logiciels de statistique

- Des logiciels gratuits: R, ADDAD, etc.
- Des logiciels commerciaux: ExcelStat, SPSS (PASW), SAS, STATA, SPAD, etc.
- Dans la tradition de l'analyse des données: ADDAD et SPAD.
- Il existe des logiciels spécifiques pour : la cartographie (ArcGis...), l'analyse de réseaux (Ucinet), l'analyse textuelle (Alceste, Lexico, Tropes...)
- Bref tutoriel / illustration de SPAD.

Les 4 fenêtres de SPAD

41



Les 4 fenêtres de SPAD

42

The screenshot displays the SPAD 7.4 software interface with the following components:

- Project Window (top-left):** Shows a tree view under 'TASTE' with sub-items: 'Diagrammes de traitements', 'TASTEjan2011', 'TASTEjuin2011', and 'MONTREUX'.
- Diagram Window (top-right):** A large empty workspace for diagram creation.
- Methods Window (bottom-left):** A list of statistical methods including: 'Imports / Exports de données', 'Data Management', 'Statistiques descriptives', 'Analyses factorielles', 'Classifications - Typologies', 'Amado - Graphiques de Bertin', 'Scoring et Modelisation', 'Arbres de decision - Segmentations', 'Règles d'association', 'Tableaux multiples', 'Text Mining', 'Modèles structurels', 'Industrialisation - Archivage', 'Statistiques avec R', and 'Process'.
- Executions Window (bottom-right):** A table titled 'Liste des exécutions' with columns: 'Niveau', 'Élément', 'Stop', 'Statut', 'Indicateur', 'Début', 'Temps', and 'Résultats'.

Les 4 fenêtres de SPAD

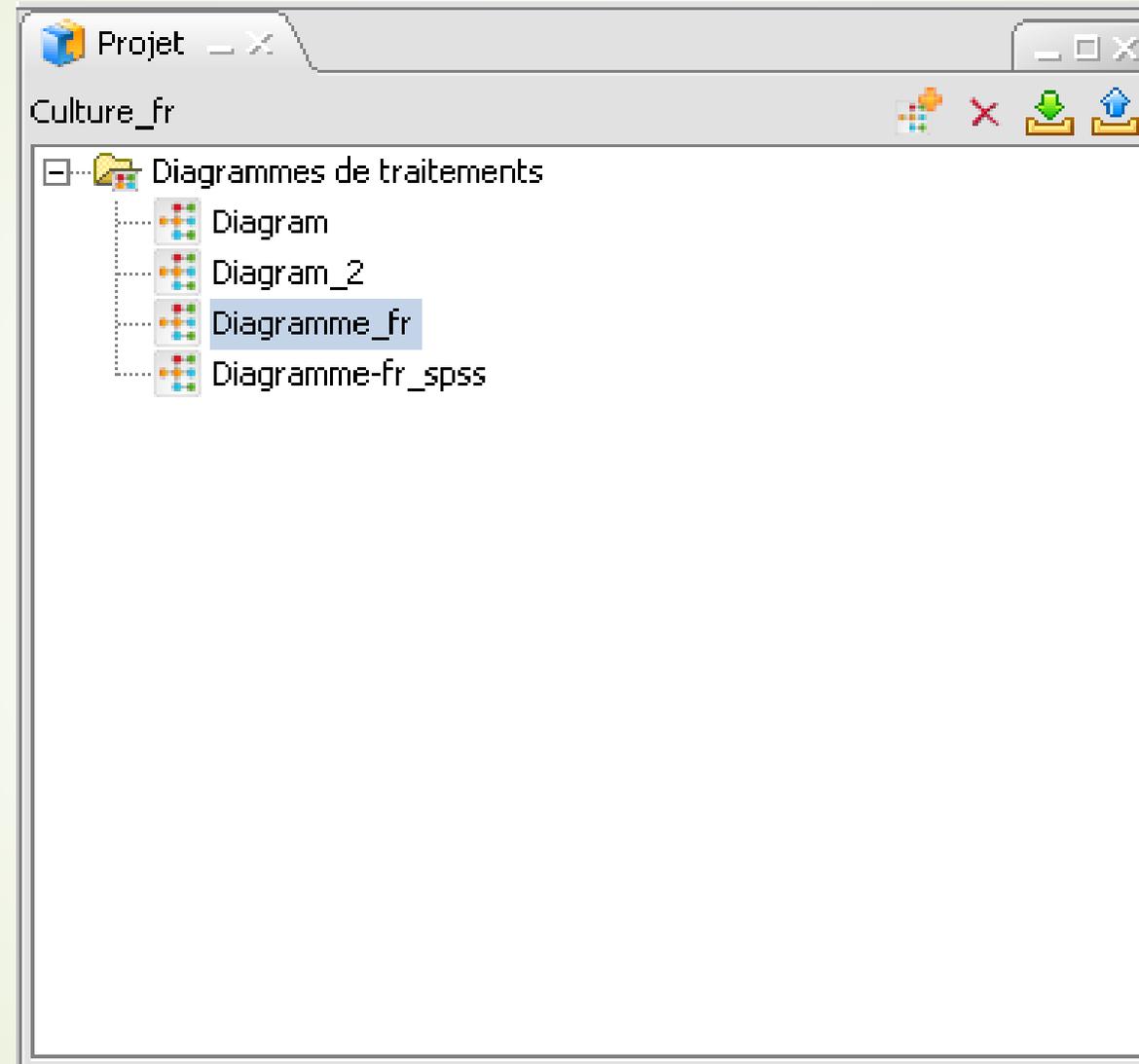
43

The screenshot displays the SPAD 7.4 interface for a project named 'Culture_fr'. The main window shows a flow diagram with the following steps: 'CULTURE.CULTURE' (input) leads to 'Edition de libellés', which leads to 'Remplacement données manquantes', then 'Générateur de nouvelles variables', followed by 'Mise en classes...', 'Générateur de nouvelles variables_1', and finally 'Statistiques de base_1'. A dashed arrow points from 'Générateur de nouvelles variables_1' to 'ACM'. The 'Exécutions' window at the bottom provides a detailed log of the process.

Niveau	Élément	Stop	Statut	Indicateur	Début	Temps	Résultats
0	CULTURE.CULTURE		✓ Terminé (ok)	3002	26/11/2012	00:01,235	
1	Edition de libellés		✓ Terminé (ok)		30/11/2012	00:00,31	
2	Remplacement données ...		✓ Terminé (ok)	3002	30/11/2012	00:00,109	
3	Générateur de nouvelle...		✓ Terminé (ok)	3002	30/11/2012	00:00,156	
1	Statistiques de base		✓ Terminé (ok)		30/11/2012	00:00,875	
4	Mise en classes...		✓ Terminé (ok)	3002	30/11/2012	00:00,46	
5	Générateur de nouvelle...		✓ Terminé (ok)	3002	30/11/2012	00:00,78	
6	Statistiques de base_1		✓ Terminé (ok)		30/11/2012	00:00,62	
6	ACM		Non démarré				

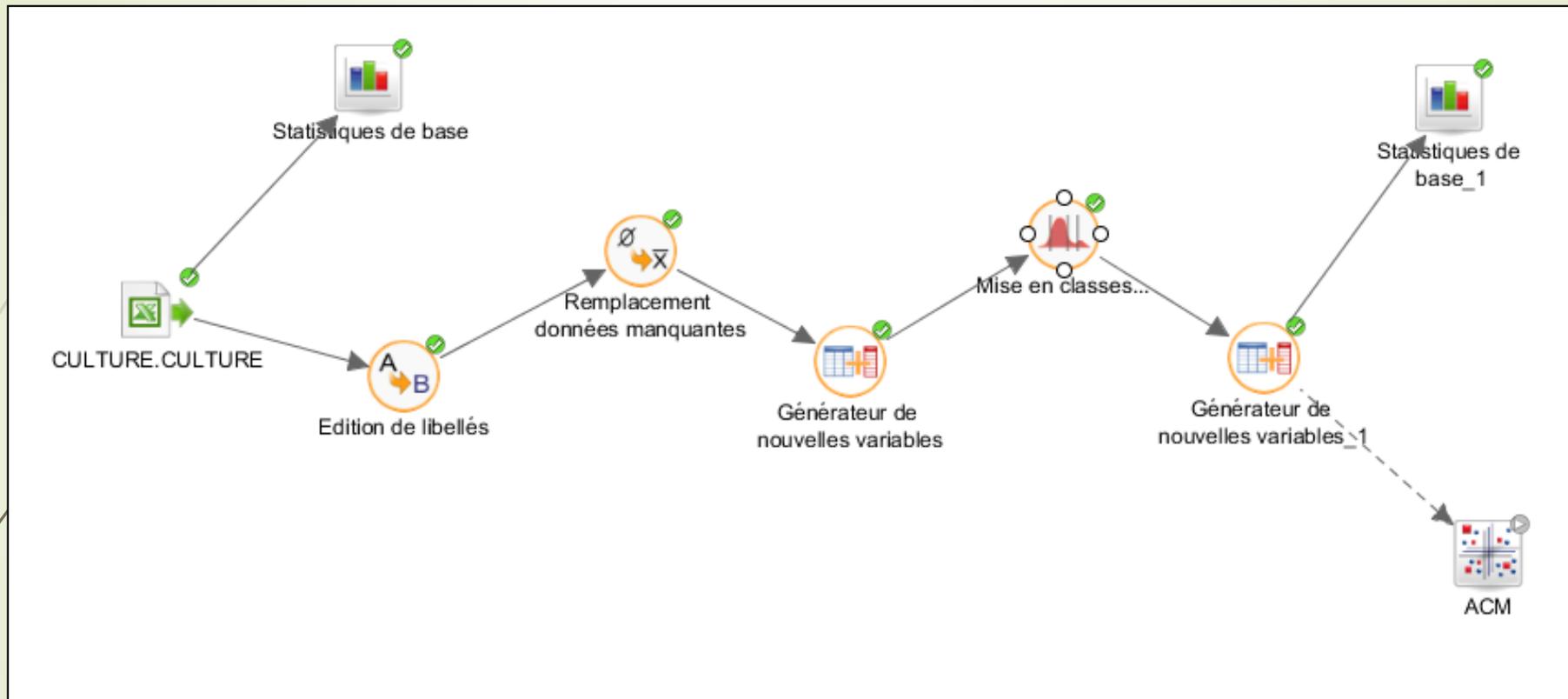
La fenêtre Projet

44



La fenêtre Diagramme

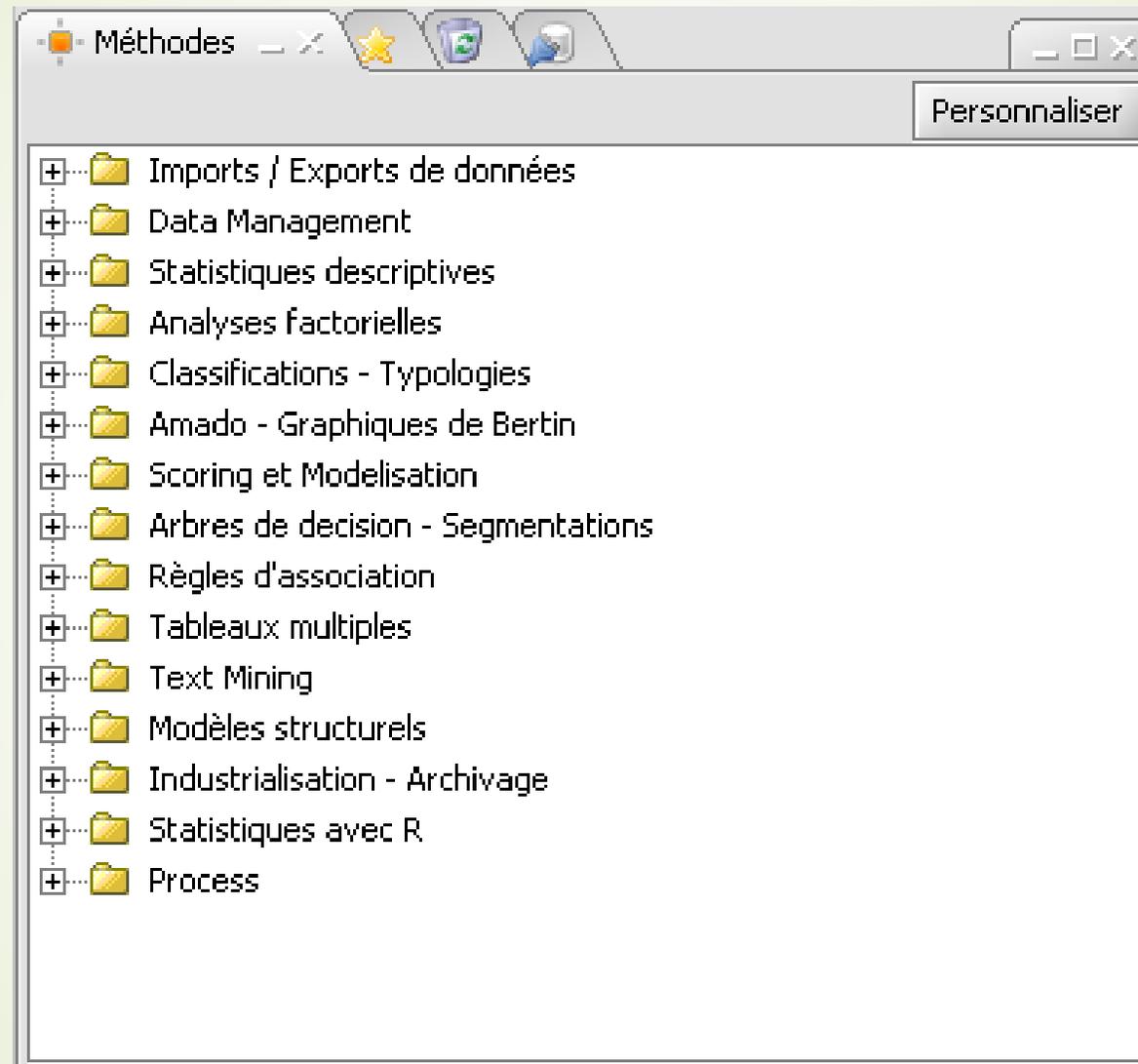
45



L'enchaînement des méthodes n'est pas obligatoire, sauf lorsqu'on veut que le tableau sur lequel on va faire l'analyse cumule les différents recodages.

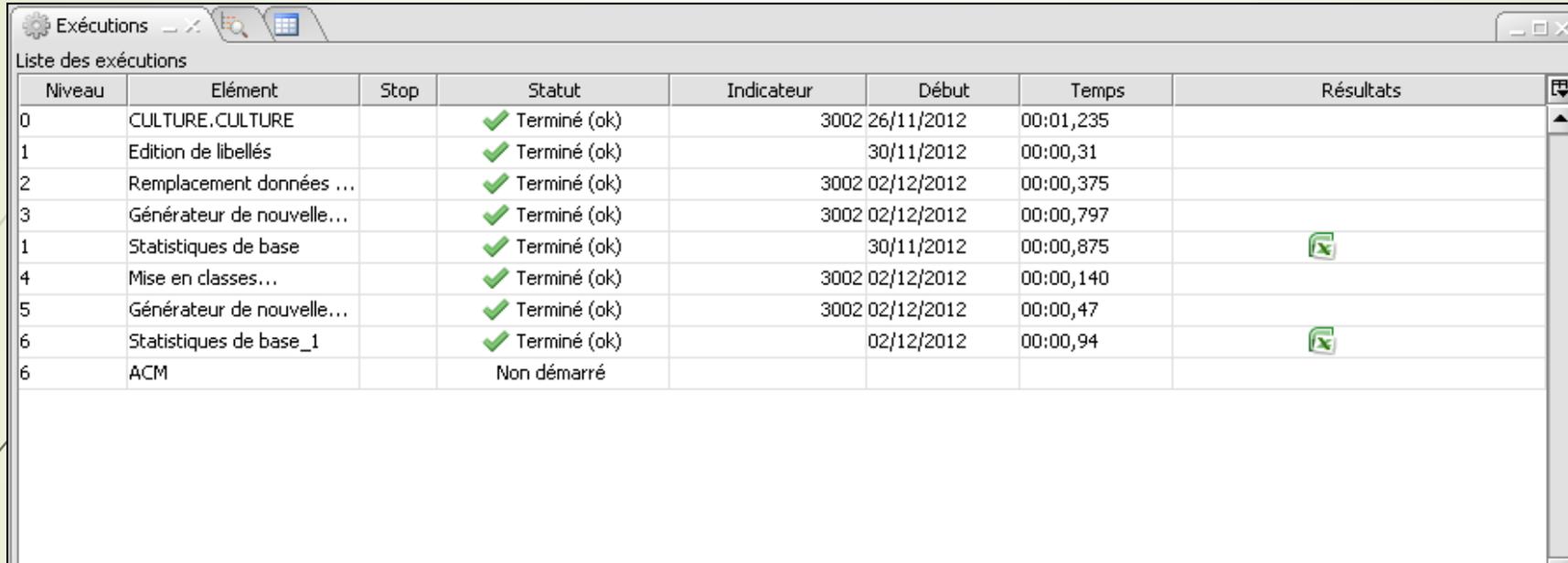
La fenêtre Méthodes

46



La fenêtre Résultats : onglet Exécutions

47



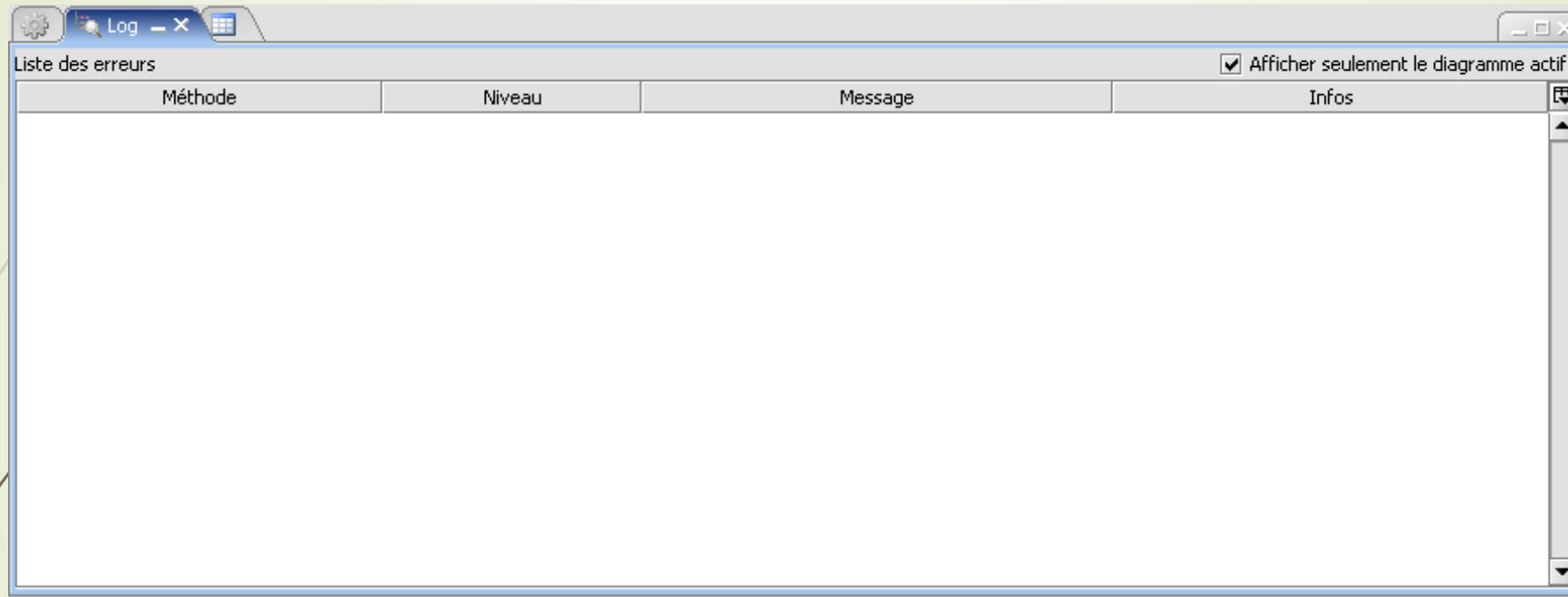
Exécutions

Liste des exécutions

Niveau	Élément	Stop	Statut	Indicateur	Début	Temps	Résultats
0	CULTURE.CULTURE		✓ Terminé (ok)	3002	26/11/2012	00:01,235	
1	Edition de libellés		✓ Terminé (ok)		30/11/2012	00:00,31	
2	Remplacement données ...		✓ Terminé (ok)	3002	02/12/2012	00:00,375	
3	Générateur de nouvelle...		✓ Terminé (ok)	3002	02/12/2012	00:00,797	
1	Statistiques de base		✓ Terminé (ok)		30/11/2012	00:00,875	
4	Mise en classes...		✓ Terminé (ok)	3002	02/12/2012	00:00,140	
5	Générateur de nouvelle...		✓ Terminé (ok)	3002	02/12/2012	00:00,47	
6	Statistiques de base_1		✓ Terminé (ok)		02/12/2012	00:00,94	
6	ACM		Non démarré				

La fenêtre Résultats : onglet Log

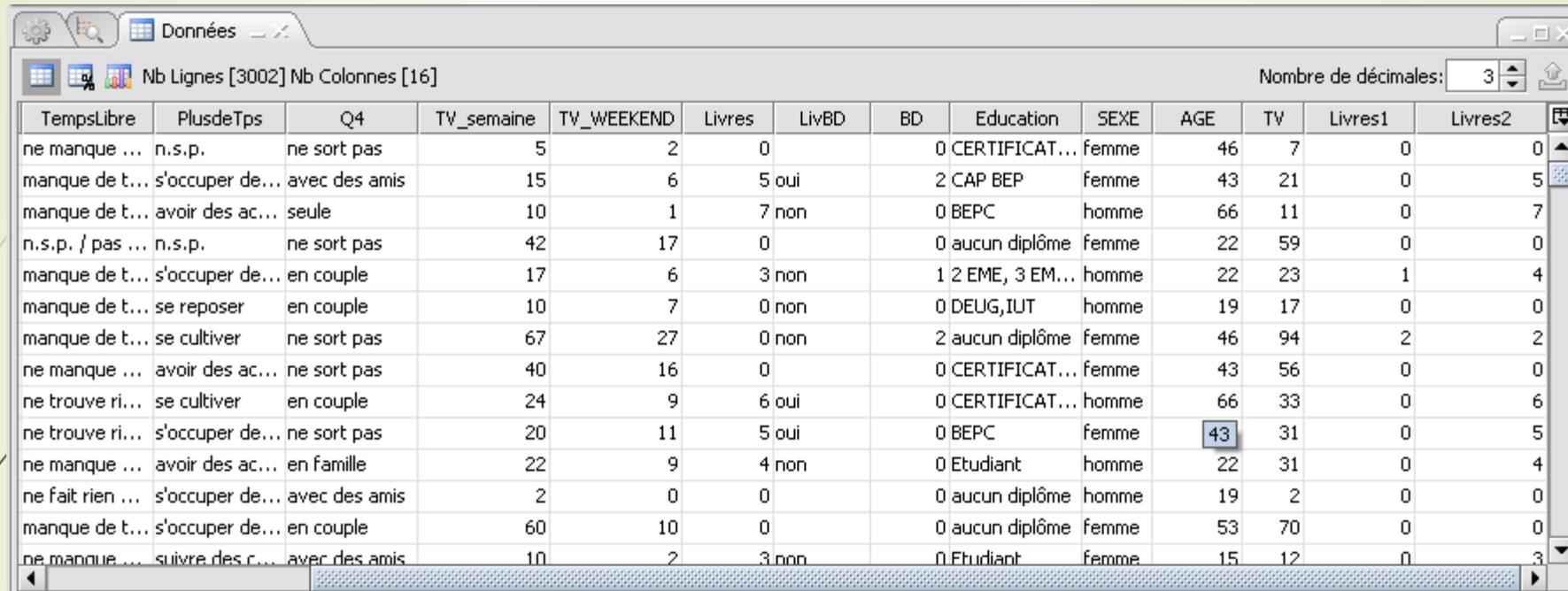
48



On trouve dans cette fenêtre les diagnostics lorsqu'une méthode n'a pas fonctionné correctement.

La fenêtre Résultats : onglet Données

49



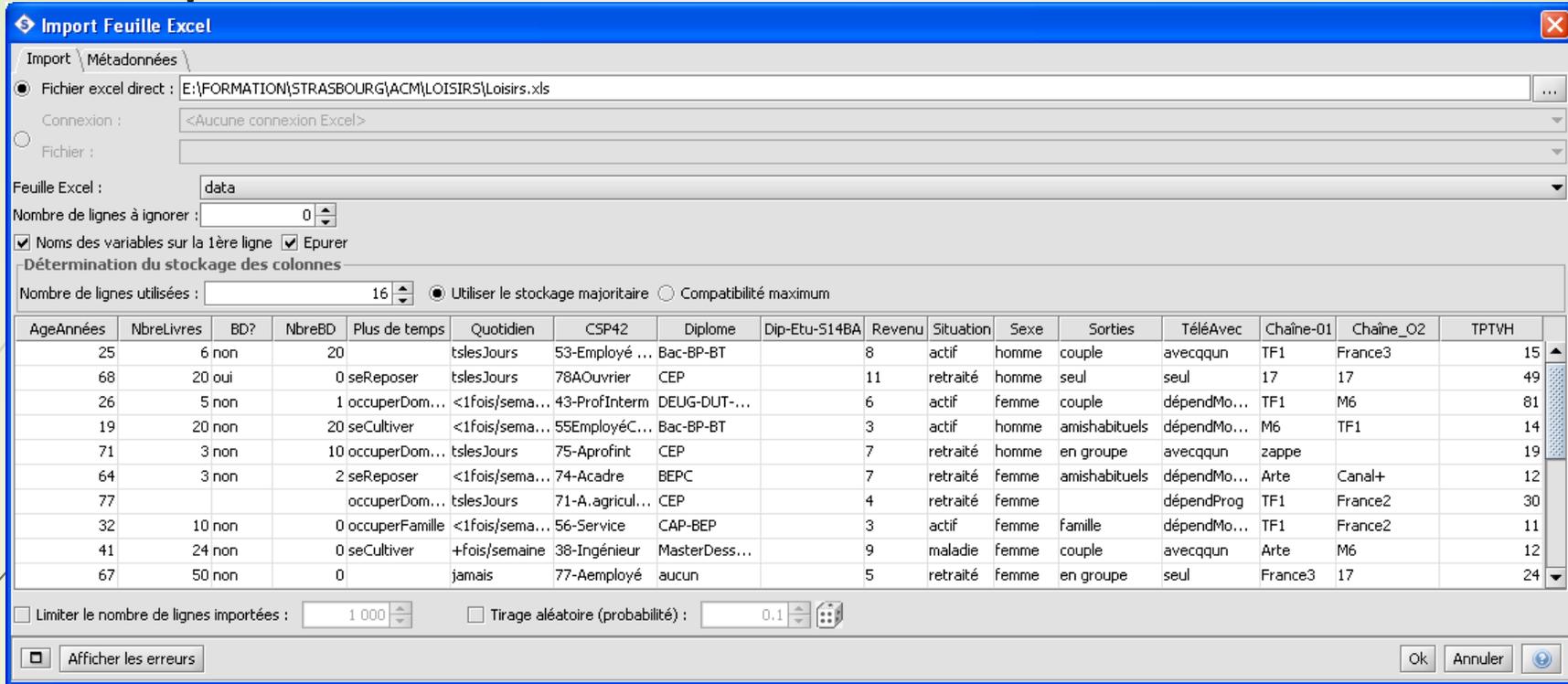
Nb Lignes [3002] Nb Colonnes [16] Nombre de décimales: 3

TempsLibre	PlusdeTps	Q4	TV_semaine	TV_WEEKEND	Livres	LivBD	BD	Education	SEXE	AGE	TV	Livres1	Livres2
ne manque ...	n.s.p.	ne sort pas	5	2	0		0	CERTIFICAT...	femme	46	7	0	0
manque de t...	s'occuper de...	avec des amis	15	6	5	oui	2	CAP BEP	femme	43	21	0	5
manque de t...	avoir des ac...	seule	10	1	7	non	0	BEPC	homme	66	11	0	7
n.s.p. / pas ...	n.s.p.	ne sort pas	42	17	0		0	aucun diplôme	femme	22	59	0	0
manque de t...	s'occuper de...	en couple	17	6	3	non	1	2 EME, 3 EM...	homme	22	23	1	4
manque de t...	se reposer	en couple	10	7	0	non	0	DEUG,IUT	homme	19	17	0	0
manque de t...	se cultiver	ne sort pas	67	27	0	non	2	aucun diplôme	femme	46	94	2	2
ne manque ...	avoir des ac...	ne sort pas	40	16	0		0	CERTIFICAT...	femme	43	56	0	0
ne trouve ri...	se cultiver	en couple	24	9	6	oui	0	CERTIFICAT...	homme	66	33	0	6
ne trouve ri...	s'occuper de...	ne sort pas	20	11	5	oui	0	BEPC	femme	43	31	0	5
ne manque ...	avoir des ac...	en famille	22	9	4	non	0	Etudiant	homme	22	31	0	4
ne fait rien ...	s'occuper de...	avec des amis	2	0	0		0	aucun diplôme	homme	19	2	0	0
manque de t...	s'occuper de...	en couple	60	10	0		0	aucun diplôme	femme	53	70	0	0
ne manque ...	s'occuper de...	avec des amis	10	2	3	non	0	Etudiant	femme	15	12	0	3

Cette fenêtre permet d'éventuels contrôles sur les transformations effectuées sur les données.

Exemple Culture : les données

50



Import \ Métadonnées

Fichier excel direct : E:\FORMATION\STRASBOURG\ACM\LOISIRS\Loisirs.xls

Connexion : <Aucune connexion Excel>

Fichier :

Feuille Excel : data

Nombre de lignes à ignorer : 0

Noms des variables sur la 1ère ligne Epurer

Détermination du stockage des colonnes

Nombre de lignes utilisées : 16 Utiliser le stockage majoritaire Compatibilité maximum

AgeAnnées	NbreLivres	BD?	NbreBD	Plus de temps	Quotidien	CSP42	Diplome	Dip-Etu-S14BA	Revenu	Situation	Sexe	Sorties	TéléAvec	Chaîne-01	Chaîne_02	TPTVH
25	6	non	20		tslesJours	53-Employé ...	Bac-BP-BT		8	actif	homme	couple	avecqqun	TF1	France3	15
68	20	oui	0	seReposer	tslesJours	78AOuvrier	CEP		11	retraité	homme	seul	seul	17	17	49
26	5	non	1	occuperDom...	<1fois/sema...	43-ProfInterm	DEUG-DUT-...		6	actif	femme	couple	dépendMo...	TF1	M6	81
19	20	non	20	seCultiver	<1fois/sema...	55EmployéC...	Bac-BP-BT		3	actif	homme	amishabituels	dépendMo...	M6	TF1	14
71	3	non	10	occuperDom...	tslesJours	75-Aprofint	CEP		7	retraité	homme	en groupe	avecqqun	zappe		19
64	3	non	2	seReposer	<1fois/sema...	74-Acadre	BEPC		7	retraité	femme	amishabituels	dépendMo...	Arte	Canal+	12
77				occuperDom...	tslesJours	71-A.agricul...	CEP		4	retraité	femme		dépendProg	TF1	France2	30
32	10	non	0	occuperFamille	<1fois/sema...	56-Service	CAP-BEP		3	actif	femme	famille	dépendMo...	TF1	France2	11
41	24	non	0	seCultiver	+fois/semaine	38-Ingénieur	MasterDess...		9	maladie	femme	couple	avecqqun	Arte	M6	12
67	50	non	0		jamais	77-Aemployé	aucun		5	retraité	femme	en groupe	seul	France3	17	24

Limiter le nombre de lignes importées : 1 000 Tirage aléatoire (probabilité) : 0.1

Afficher les erreurs

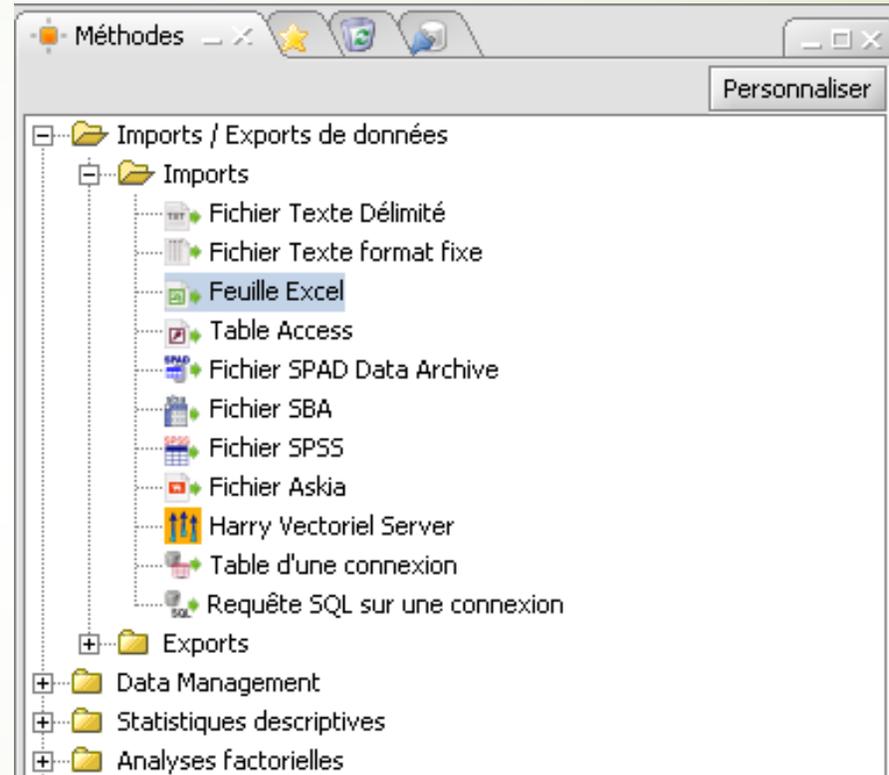
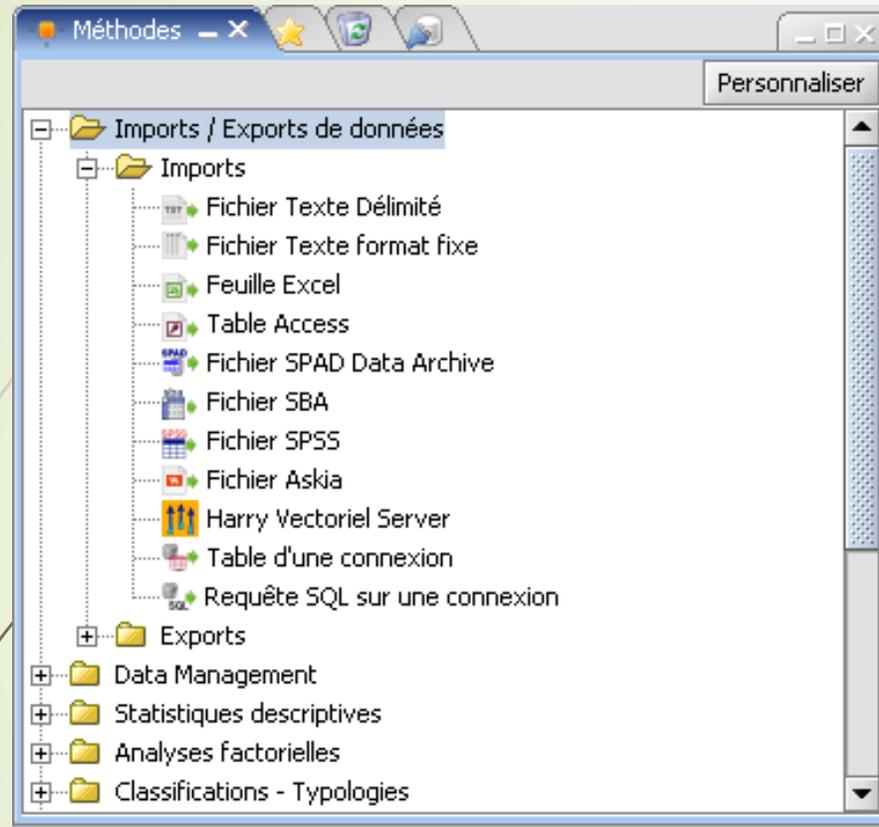
Ok Annuler

Cette enquête, effectuée par le département des études, de la prospective et des statistiques du Ministère de la culture et de la communication (Donnat, 2009), comporte 5004 individus âgés de 15 ans et plus.

Nous avons choisi sept questions ayant trait aux loisirs (une question), aux sorties le soir et à la sociabilité (deux questions), aux pratiques culturelles à savoir la lecture de quotidien, de livres et BD (deux questions) et la télévision (deux questions). On construit ainsi un espace des habitudes —ou pratiques— culturelles et de sociabilité.

Importer des données

51



Importer des données (suite)



52

Import Feuille Excel

Import \ Métadonnées

Fichier excel direct : E:\FORMATION\STRASBOURG\ACM\LOISIRS\Loisirs.xls

Connexion : <Aucune connexion Excel>

Fichier :

Feuille Excel : data

Nombre de lignes à ignorer : 0

Noms des variables sur la 1ère ligne Epurer

Détermination du stockage des colonnes

Nombre de lignes utilisées : 16 Utiliser le stockage majoritaire Compatibilité maximum

AgeAnnées	NbreLivres	BD?	NbreBD	Plus de temps	Quotidien	CSP42	Diplome	Dip-Etu-514BA	Revenu	Situation	Sexe	Sorties	TéléAvec	Chaîne-01	Chaîne_02	TPTVH
25	6 non		20		tslesJours	53-Employé ...	Bac-BP-BT		8	actif	homme	couple	avecqqun	TF1	France3	15
68	20 oui		0	seReposer	tslesJours	78AOuvrier	CEP		11	retraité	homme	seul	seul	17	17	49
26	5 non		1	occuperDom...	<1fois/sema...	43-ProfInterm	DEUG-DUT-...		6	actif	femme	couple	dépendMo...	TF1	M6	81
19	20 non		20	seCultiver	<1fois/sema...	55EmployéC...	Bac-BP-BT		3	actif	homme	amishabituels	dépendMo...	M6	TF1	14
71	3 non		10	occuperDom...	tslesJours	75-Aprofint	CEP		7	retraité	homme	en groupe	avecqqun	zappe		19
64	3 non		2	seReposer	<1fois/sema...	74-Acadre	BEPC		7	retraité	femme	amishabituels	dépendMo...	Arte	Canal+	12
77				occuperDom...	tslesJours	71-A.agricul...	CEP		4	retraité	femme		dépendProg	TF1	France2	30
32	10 non		0	occuperFamille	<1fois/sema...	56-Service	CAP-BEP		3	actif	femme	famille	dépendMo...	TF1	France2	11
41	24 non		0	seCultiver	+fois/semaine	38-Ingénieur	MasterDess...		9	maladie	femme	couple	avecqqun	Arte	M6	12
67	50 non		0		jamais	77-Aemployé	aucun		5	retraité	femme	en groupe	seul	France3	17	24

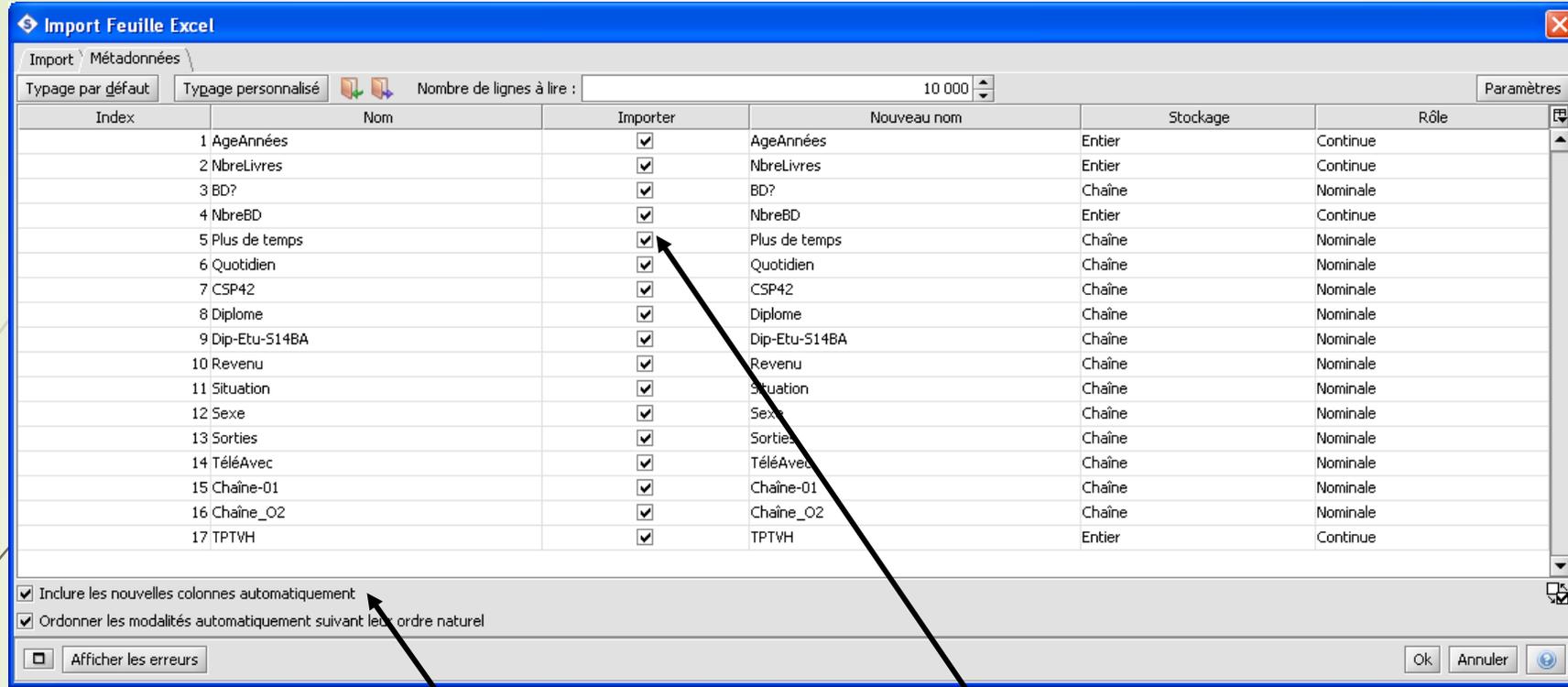
Limiter le nombre de lignes importées : 1 000 Tirage aléatoire (probabilité) : 0.1

Afficher les erreurs

Ok Annuler

Importer des données (suite)

53



Cocher l'option « Inclure les nouvelles colonnes automatiquement ».

Penser qu'il est possible de ne pas importer toutes les variables.

Penser à définir un identifiant

Créer un index

54



Aller chercher la fonction INDEX dans le dossier Statistiques. Mais l'index va commencer à 0. D'où la nécessité ensuite d'ajouter 1 à chaque valeur de l'index, d'où la calculatrice.

Générateur de nouvelles variables

Fonctions

- Fonctions
 - Conversion
 - Date & Heure
 - Inspection
 - Logique
 - Math et Trigo
 - Mémoire
 - Statistiques
 - Texte
- Outils
 - Assistant logique
 - Assistant logique (case)
 - Calculatrice**
 - Script Python

Variables initiales et créées

Nom	Stockage	Rôle	Conser...	Valeur
AgeAnnées	Entier	Continue	<input checked="" type="checkbox"/>	
NbreLivres	Entier	Continue	<input checked="" type="checkbox"/>	
BD?	Chaîne	Nominale	<input checked="" type="checkbox"/>	
NbreBD	Entier	Continue	<input checked="" type="checkbox"/>	
Plus de temps	Chaîne	Nominale	<input checked="" type="checkbox"/>	
Quotidien	Chaîne	Nominale	<input checked="" type="checkbox"/>	
CSP42	Chaîne	Nominale	<input checked="" type="checkbox"/>	
Diplome	Chaîne	Nominale	<input checked="" type="checkbox"/>	
Dip-Etu-514BA	Chaîne	Nominale	<input checked="" type="checkbox"/>	
Revenu	Chaîne	Nominale	<input checked="" type="checkbox"/>	
Situation	Chaîne	Nominale	<input checked="" type="checkbox"/>	
Sexe	Chaîne	Nominale	<input checked="" type="checkbox"/>	
Sorties	Chaîne	Nominale	<input checked="" type="checkbox"/>	
TéléAvec	Chaîne	Nominale	<input checked="" type="checkbox"/>	
Chaîne_01	Chaîne	Nominale	<input checked="" type="checkbox"/>	
Chaîne_02	Chaîne	Nominale	<input checked="" type="checkbox"/>	
TPTVH	Entier	Continue	<input checked="" type="checkbox"/>	
Index	Entier	Autre	<input checked="" type="checkbox"/>	
ID	Réel	Continue	<input checked="" type="checkbox"/>	

Calculatrice(Variables) Ligne : 0 Tester

Calculatrice
La calculatrice vous permet de définir une formule de calcul pour calculer les valeurs d'une variable à partir des variables numériques existantes. Les fonctions mathématiques usuelles sont disponibles.

Transformations successives

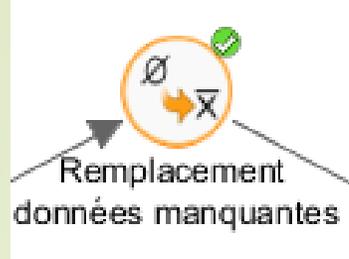
Fonctions	Paramètres	Valeurs	Résultats
INDEX			Int : Index
Calculatrice	Variables	[Index]	Num : ID

Ordonner les modalités automatiquement suivant leur ordre naturel Résultats : variable en préfixe

Afficher les erreurs Ok Annuler

Recoder les non-réponses

55



Remplacement des données manquantes

Variables disponibles :

Index	Nom	Stockage	Rôle	Valeurs manquantes	Méthode	Nb
1	AgeAnnées	Entier	Continue			
2	NbreLivres	Entier	Continue		Constante : 0	
3	BD?	Chaîne	Nominale		Constante : 0	
4	NbreBD	Entier	Continue		Constante : 0	
5	Plus de temps	Chaîne	Nominale		Constante : NeManquePastemps	
6	Quotidien	Chaîne	Nominale			
7	CSP42	Chaîne	Nominale		Constante : NR	
8	Diplome	Chaîne	Nominale		Constante : Etudiant	
9	Dip-Etu-514BA	Chaîne	Nominale		Constante : nonétudiant	
10	Revenu	Chaîne	Nominale			
11	Situation	Chaîne	Nominale			
12	Sexe	Chaîne	Nominale			
13	Sorties	Chaîne	Nominale		Constante : neSortPas	
14	TéléAvec	Chaîne	Nominale		Constante : PasTélé	
15	Chaîne-01	Chaîne	Nominale		Constante : pasTélé	
16	Chaîne_O2	Chaîne	Nominale		Constante : manquant	
17	TPTVH	Entier	Continue		Constante : 0	
18	Index	Entier	Autre			
19	ID	Entier	Identifiant			

Afficher les erreurs

Ok Annuler

Ici le recodage des non réponses (données manquantes) ne concerne que les variables continues. On décide, dans le cas présent, de donner la valeur 0 pour les variables continues, et une valeur particulière pour des variables nominales.

Recodage de données

56



Générateur de nouvelles variables

Fonctions

- Fonctions
 - Conversion
 - Date & Heure
 - Inspection
 - Logique
 - Math et Trigo
 - Mémoire
 - Statistiques
 - Texte
- Outils
 - Assistant logique
 - Assistant logique (case)
 - Calculatrice
 - Script Python

Variables initiales et créées

Nom	Stockage	Rôle	Conser...	Valeur
AgeAnnées	Entier	Continue	<input checked="" type="checkbox"/>	
NbreLivres	Entier	Continue	<input checked="" type="checkbox"/>	
BD?	Chaîne	Nominale	<input checked="" type="checkbox"/>	
NbreBD	Entier	Continue	<input checked="" type="checkbox"/>	
Plus de temps	Chaîne	Nominale	<input checked="" type="checkbox"/>	
Quotidien	Chaîne	Nominale	<input checked="" type="checkbox"/>	
CSP42	Chaîne	Nominale	<input checked="" type="checkbox"/>	
Diplome	Chaîne	Nominale	<input checked="" type="checkbox"/>	
Dip-Etu-514BA	Chaîne	Nominale	<input checked="" type="checkbox"/>	
Revenu	Chaîne	Nominale	<input checked="" type="checkbox"/>	
Situation	Chaîne	Nominale	<input checked="" type="checkbox"/>	
Sexe	Chaîne	Nominale	<input checked="" type="checkbox"/>	
Sorties	Chaîne	Nominale	<input checked="" type="checkbox"/>	
TéléAvec	Chaîne	Nominale	<input checked="" type="checkbox"/>	
Chaîne-01	Chaîne	Nominale	<input checked="" type="checkbox"/>	
Chaîne_02	Chaîne	Nominale	<input checked="" type="checkbox"/>	

Ligne : 0

Tester

Transformations successives

Fonctions	Paramètres	Valeurs	Résultats
Assistant logique			
	Filtre	[? Chaîne-01 : Chaîne_02]	Rés : CH1_CH2
Assistant logique			
	Filtre	[? NbreBD : 0.0]	Rés : BDlues
Calculatrice			
	Variables	[NbreLivres, BDlues]	Num : Lecture

Ordonner les modalités automatiquement suivant leur ordre naturel Résultats : variable en préfixe

Afficher les erreurs

Ok Annuler



ge de données

Assistant logique

Index	Nom	Stockage	Rôle
1	AgeAnnées	Entier	Continue
2	NbreLivres	Entier	Continue
3	BD?	Chaîne	Nominale
4	NbreBD	Entier	Continue
5	Plus de temps	Chaîne	Nominale
6	Quotidien	Chaîne	Nominale
7	CSP42	Chaîne	Nominale
8	Diplome	Chaîne	Nominale
9	Dip-Etu-S14BA	Chaîne	Nominale
10	Revenu	Chaîne	Nominale
11	Situation	Chaîne	Nominale
12	Sexe	Chaîne	Nominale
13	Sorties	Chaîne	Nominale
14	TéléAvec	Chaîne	Nominale
15	Chaîne-01	Chaîne	Nominale
16	Chaîne_02	Chaîne	Nominale
17	TPTVH	Entier	Continue
18	Index	Entier	Autre
19	ID	Entier	Identifiant

Valeurs	Variable	Index	Valeur	Effectif
=		1	11	78
<>		2	12	57
<		3	13	45
<=		4	14	40
>		5	15	25
>=		6	16	43
= ∅		7	17	68
<> ∅		8	19	18
		9	Arte	259
		10	Canal+	260
		11	France2	620
		12	France3	426
		13	France5	201
		14	M6	617
		15	TF1	1 479
		16	zappe	647
		17	pasTélé	121

Nombre de _Missing_ : 0

ET OU

(...)

Lien	Condition
	Chaîne-01 = TF1 ou France2 ou France3 ou Canal+ ou France5 ou Arte ou M6 ou zappe

Résultats

Type :

Si vrai : Constante : Variable : _Missing_

Si faux : Constante : Variable : _Missing_

Recodage de données



58

En clair : si pour la variable BD (avez-vous compté les bandes dessinées à la question « combien de livres avez-vous lu les 12 derniers mois ? ») il y a une valeur manquante ou la réponse « oui » alors pour la variable BD mettre 0, sinon laisser la valeur actuelle.

Assistant logique

Index	Nom	Stockage	Rôle
1	AgeAnnées	Entier	Continue
2	NbreLivres	Entier	Continue
3	BD?	Chaîne	Nominale
4	NbreBD	Entier	Continue
5	Plus de temps	Chaîne	Nominale
6	Quotidien	Chaîne	Nominale
7	CSP42	Chaîne	Nominale
8	Diplome	Chaîne	Nominale
9	Dip-Etu-514BA	Chaîne	Nominale
10	Revenu	Chaîne	Nominale
11	Situation	Chaîne	Nominale
12	Sexe	Chaîne	Nominale
13	Sorties	Chaîne	Nominale
14	TéléAvec	Chaîne	Nominale
15	Chaîne-01	Chaîne	Nominale
16	Chaîne_O2	Chaîne	Nominale
17	TPTVH	Entier	Continue
18	Index	Entier	Autre
19	ID	Entier	Identifiant
39	Chaîne1+2	Chaîne	Nominale

Index	Valeur	Effectif
1	non	2 819
2	oui	1 109
3	0	1 076

Nombre de _Missing_ : 0

ET OU Valider

(...)

Lien Condition

BD? = non

Résultats

Type : Réel

Si vrai : Constante : 1 Variable : NbreBD _Missing_

Si faux : Constante : 0 Variable : AgeAnnées _Missing_

OK Annuler

Recodage de données

59



Calculatrice

Variable	Stockage	Utilisée ?	Valeur
AgeAnnées	Entier	<input type="checkbox"/>	74.457
NbreLivres	Entier	<input checked="" type="checkbox"/>	93.036
NbreBD	Entier	<input type="checkbox"/>	68.206
TPTVH	Entier	<input type="checkbox"/>	35.169
Index	Entier	<input type="checkbox"/>	58.086
ID	Entier	<input type="checkbox"/>	9.378
NbrBDlues	Réel	<input checked="" type="checkbox"/>	89.475

Fonctions

- _IF_
- abs
- acos
- asin
- atan
- atan2
- ceil
- cos
- cosh
- e

Saisie de la formule

```
FieldValue("NbreLivres")+FieldValue("NbrBDlues")
```

Test

Attention, si des variables de l'expression contiennent des valeurs manquantes, le résultat sera valeur manquante

OK Annuler

Mise en classes, regroupement de modalités

60



Mise en classes, Regroupement de modalités

Variables disponibles : 22

Index	Nom	Stockage	Rôle	Nb modalités	Conserver
1	AgeAnnées	Entier	Continue		<input checked="" type="checkbox"/>
2	NbreLivres		Continue		<input type="checkbox"/>
3	BD?		Nominale	3	<input type="checkbox"/>
4	NbreBD		Continue		<input type="checkbox"/>
5	Plus de temps		Nominale	9	<input checked="" type="checkbox"/>
6	Quotidien		Nominale	5	<input checked="" type="checkbox"/>
7	CSP42		Nominale	36	<input checked="" type="checkbox"/>
8	Diplome		Nominale	10	<input type="checkbox"/>
9	Dip-Etu-S14BA		Nominale	9	<input type="checkbox"/>
10	Revenu		Nominale	12	<input type="checkbox"/>
11	Situation		Nominale	10	<input checked="" type="checkbox"/>
12	Sexe	Chaîne	Nominale	2	<input checked="" type="checkbox"/>
13	Sorties	Chaîne	Nominale	8	<input type="checkbox"/>
14	TéléAvec	Chaîne	Nominale	5	<input checked="" type="checkbox"/>
15	Chaîne-01	Chaîne	Nominale	17	<input type="checkbox"/>

Mise en classes

- Mise en classes automatique
- Mise en classes dynamique
- Centrer-réduire
- Transformation en rangs

Regroupement de modalités

- Nominale vers continue
- Croisement de variables
- Croisement de variables automatique

Variables recodées : 12

Index	Nom	Stockage	Rôle	Méthode	Variables d'origine
1	Rec_AgeAnnées	Chaîne	Nominale	Mise en classe	[AgeAnnées]
2	HeuresTélé	Chaîne	Nominale	Mise en classe	[TPTVH]
3	Rec_Livres+BD	Chaîne	Nominale	Mise en classe	[Livres+BD]
4	Rec_Chaine1+2	Chaîne	Nominale	Regroupement de mod...	[Chaine1+2]
5	Rec_Sorties	Chaîne	Nominale	Regroupement de mod...	[Sorties]
6	Rec_Revenu	Chaîne	Nominale	Regroupement de mod...	[Revenu]
7	Rec_Diplome	Chaîne	Nominale	Regroupement de mod...	[Diplome]
8	Rec_Situation	Chaîne	Nominale	Regroupement de mod...	[Situation]
9	CSP24	Chaîne	Nominale	Regroupement de mod...	[CSP42]
10	Rec_Dip-Etu-S14BA	Chaîne	Nominale	Regroupement de mod...	[Dip-Etu-S14BA]

Afficher les erreurs

Ok Annuler



Mise en classes (7 classes) de la variable âge pour devenir Rec_age

Mise en classe

Utiliser le dialogue de mise en forme

Utiliser les X premiers caractères du libellé des variables comme titre générique des modalités. X :

Libellés des classes

Autoriser les écarts entre les bornes des classes

Utiliser une variable de pondération :

Statistiques descriptives

Min[15]	Nb[5004]	Moyenne[46.331]
Max[95]	Nb_Missing_[0]	Ecart-type[19.019]

Colonne Effectif

Index	Classe	>=	<	Effectif / %
1	< 18	-∞	18	
2	[18;25[18	25	
3	[25;35[25	35	
4	[35;45[35	45	
5	[45;55[45	55	
6	[55;65[55	65	
7	>= 65	65	∞	

Paramètres spécifiques

Valeur spéciale mise dans une classe supplémentaire Valeur : Classe :

Données manquantes mises dans une classe supplémentaire Classe :



Mise en classes (5 classes) de la variable HeuresTélé

Mise en classe

Utiliser le dialogue de mise en forme

Utiliser les X premiers caractères du libellé des variables comme titre générique des modalités. X :

Libellés des classes

Autoriser les écarts entre les bornes des classes

Utiliser une variable de pondération :

Statistiques descriptives

Min[0] Nb[5004] Moyenne[20,186]
 Max[168] Nb_Missing_[0] Ecart-type[15,122]

Colonne Effectif

Index	Classe	>=	<	Effectif / %
1	< 5h	-∞	5	
2	5-9h	5	10	
3	10-19h	10	20	
4	20-34h	20	35	
5	>= 35h	35	∞	

Paramètres spécifiques

Valeur spéciale mise dans une classe supplémentaire Valeur : Classe :

Données manquantes mises dans une classe supplémentaire Classe :

63



Mise en classes (6 classes) de la variable Reclivres+BD

Mise en classe

Utiliser le dialogue de mise en forme

Utiliser les X premiers caractères du libellé des variables comme titre générique des modalités. X :

Libellés des classes

Autoriser les écarts entre les bornes des classes

Utiliser une variable de pondération :

Statistiques descriptives

Min[0]	Nb[5004]	Moyenne[25.065]
Max[1 998]	Nb _Missing_[0]	Ecart-type[107.417]

Colonne Effectif

Index	Classe	>=	<	Effectif / %
1	aucun	-∞		1
2	L1-4	1		5
3	L5-12	5		13
4	L13-40	13		41
5	L>40	41		900
6	n.s.p.	900	∞	

Paramètres spécifiques

Valeur spéciale mise dans une classe supplémentaire Valeur : Classe :

Données manquantes mises dans une classe supplémentaire Classe :

Mise en classes, regroupement de modalités

64

Mise en classes...



Regroupement de modalités

Variable(s) créée(s)
Nom : Stockage :

Valeur	Nouvelle valeur	Effectif ...
11	Autres	0.54 %
12	Autres	0.22 %
13	Autres	0.4 %
14	Autres	0.56 %
15	Autres	0.34 %
16	Autres	0.52 %
17	Autres	0.659 %
19	Autres	0.06 %
Arte	Arte	5.396 %
Canal+	Canal+	5.675 %
France2	France2	12.73 %
France3	France3	8.773 %
France5	France5	4.117 %
M6	M6	13.03 %
manquant	Autres	2.778 %
TF1	TF1	30.695 %

Nb_Missing_ : 0

Traitement des données manquantes

Sans changement
 Recodées en :

Traitement des autres valeurs

Mises en données manquantes
 Chaque valeur inconnue devient une nouvelle modalité
 Regroupées dans :

Ordre	Nouvelle valeur	Valeurs	Effectif / %
1	TF1	TF1	30.695 %
2	France2	France2	12.73 %
3	France3	France3	8.773 %
4	Canal+	Canal+	5.675 %
5	France5	France5	4.117 %
6	Arte	Arte	5.396 %
7	M6	M6	13.03 %
8	zappe	zappe	13.509 %
9	Autres	19 17 15 16 13 14 11 12 manquant	6.075 %

OK Annuler

Mise en classes (5 classes) de la variable TV pour devenir Télévision

Mise en classe

Utiliser le dialogue de mise en forme

Utiliser les X premiers caractères du libellé des variables comme titre générique des modalités. X :

Libellés des classes

Autoriser les écarts entre les bornes des classes

Utiliser une variable de pondération :

Statistiques descriptives

Min[0]	Nb[3002]	Moyenne[19.323]
Max[119]	Nb_Missing_[0]	Ecart-type[14.919]

Colonne Effectif

Index	Classe	>=	<	Effectif / %
1	< 1	-∞	1	
2	[1;10[1	10	
3	[10;19[10	19	
4	[19;31[19	31	
5	>= 31	31	∞	

Paramètres spécifiques

Valeur spéciale mise dans une classe supplémentaire Valeur : Classe :

Données manquantes mises dans une classe supplémentaire Classe :

Mise en classes (5 classes) de la variable Livres2 pour devenir Lecture

Mise en classes, Regroupement de modalités

Variables disponibles : 16

Index	Nom	Stockage	Rôle	Nb modalités	Conserver
2	Loisirs	Chaîne	Nominale	5	<input checked="" type="checkbox"/>
3	TempsLibre	Chaîne	Nominale	5	<input checked="" type="checkbox"/>
4	PlusdeTps	Chaîne	Nominale	8	<input checked="" type="checkbox"/>
5	Q4	Chaîne	Nominale	6	<input checked="" type="checkbox"/>
6	TV_semaine	Réel	Continue		<input checked="" type="checkbox"/>
7	TV_WEEKEND	Réel	Continue		<input checked="" type="checkbox"/>
8	Livres	Réel	Continue		<input checked="" type="checkbox"/>
9	LivBD	Chaîne	Nominale	2	<input checked="" type="checkbox"/>
10	BD	Réel	Continue		<input checked="" type="checkbox"/>
11	Education	Chaîne	Nominale	8	<input checked="" type="checkbox"/>
12	SEXE	Chaîne	Nominale	2	<input checked="" type="checkbox"/>
13	AGE	Réel	Continue		<input checked="" type="checkbox"/>
14	TV	Réel	Continue		<input checked="" type="checkbox"/>
15	Livres1	Réel	Nominale	37	<input checked="" type="checkbox"/>
16	Livres2	Réel	Continue		<input checked="" type="checkbox"/>

Variables recodées : 2

Index	Nom	Stockage	Rôle
1	Sorties	Chaîne	Nominale
2	Rec_TV	Chaîne	Nominale

Mise en classes

- Mise en classes automatique
- Mise en classes dynamique
- Centrer-réduire
- Transformation en rangs

- Regroupement de modalités
- Nominale vers continue
- Croisement de variables
- Croisement de variables automatique

Afficher les erreurs

Ok Annuler

Mise en classes (5 classes) de la variable Livres2 pour devenir Lecture

Mise en classes de variables continues

Méthode de mise en classes

Amplitudes égales Bornes manuelles : 1/5/13/40
Liste des bornes supérieures exclues séparées par / (exemple 15/18/25/50)

Fréquences égales Pourcentages manuels :
Liste des pourcentages pour chaque classe séparés par / (exemple 10/30/40/20)

Paramètres

Nombre de classes : 4

Amplitude : 100

Borne minimale (incluse) Automatique Imposée : 0

Borne maximale (exclue) Automatique Imposée : 100

Utiliser une variable de pondération : <Aucune>

Paramètres spécifiques

Valeur spéciale mise dans une classe supplémentaire Valeur : 0 Classe :

Données manquantes mises dans une classe supplémentaire Classe :

Extension des libellés des variables créées Préfixe : Rec_ Suffixe :

Utiliser les X premiers caractères du libellé des variables comme titre générique des modalités. X : 5

Libellés des classes créés : Générés par les bornes Numérotés

Générer les bornes des classes lors de la 1ère exécution

Préparation des données et détermination des bornes **terminé**

Ok Annuler

Mise en classes (5 classes) de la variable Livres2 pour devenir Lecture

Mise en classe

Utiliser le dialogue de mise en forme

Utiliser les X premiers caractères du libellé des variables comme titre générique des modalités. X :

Libellés des classes

Autoriser les écarts entre les bornes des classes

Utiliser une variable de pondération :

Statistiques descriptives

Min[0] Nb[3002] Moyenne[28.577]
 Max[1 100] Nb_Missing_[0] Ecart-type[101.922]

Colonne Effectif

Index	Classe	>=	<	Effectif / %
1	< 1	-∞	1	
2	[1;5[1	5	
3	[5;13[5	13	
4	[13;40[13	40	
5	>= 40	40	∞	

Paramètres spécifiques

Valeur spéciale mise dans une classe supplémentaire Valeur : Classe :

Données manquantes mises dans une classe supplémentaire Classe :

Fenêtre générale des Paramètres pour la mise en classes et /ou le regroupement de modalités

Mise en classes, Regroupement de modalités

Variables disponibles : 16

Index	Nom	Stockage	Rôle	Nb modalités	Conserver
1	ID	Entier	Identifiant		<input checked="" type="checkbox"/>
2	Loisirs	Chaîne	Nominale	5	<input checked="" type="checkbox"/>
3	TempsLibre	Chaîne	Nominale	5	<input checked="" type="checkbox"/>
4	PlusdeTps	Chaîne	Nominale	8	<input checked="" type="checkbox"/>
5	Q4	Chaîne	Nominale	6	<input checked="" type="checkbox"/>
6	TV_semaine	Entier	Continue		<input checked="" type="checkbox"/>
7	TV_WEEKEND	Entier	Continue		<input checked="" type="checkbox"/>
8	Livres	Entier	Continue		<input checked="" type="checkbox"/>
9	LivBD	Chaîne	Nominale	2	<input checked="" type="checkbox"/>
10	BD	Entier	Continue		<input checked="" type="checkbox"/>
11	Education	Chaîne	Nominale	8	<input checked="" type="checkbox"/>
12	SEXE	Chaîne	Nominale	2	<input checked="" type="checkbox"/>
13	AGE	Entier	Continue		<input checked="" type="checkbox"/>
14	TV	Réel	Continue		<input checked="" type="checkbox"/>
15	Livres1	Entier	Continue		<input checked="" type="checkbox"/>
16	Livres2	Réel	Continue		<input checked="" type="checkbox"/>

Variables recodées : 3

Index	Nom	Stockage	Rôle	Méthode	Variables d'origine
1	Sorties	Chaîne	Nominale	Regroupement de modali...	[Q4]
2	Télévision	Chaîne	Nominale	Mise en classe	[TV]
3	Lecture	Chaîne	Nominale	Mise en classe	[Livres2]

Afficher les erreurs

Ok Annuler



Fabrication d'une variable à deux modalités : actif, suppl.

70

Générateur de nouvelles variables

Fonctions

- Conversion
- Date & Heure
- Inspection
- Logique
- Math et Trigo
- Mémoire
- Statistiques
- Texte
- Outils
 - Assistant logique**
 - Assistant logique (case)
 - Calculatrice
 - Script Python

Variables initiales et créées

Nom	Stockage	Rôle	Conserver ?	Valeur
ID	Réel	Identifiant	<input checked="" type="checkbox"/>	
Loisirs	Chaîne	Nominale	<input checked="" type="checkbox"/>	
TempsLibre	Chaîne	Nominale	<input checked="" type="checkbox"/>	
PlusdeTps	Chaîne	Nominale	<input checked="" type="checkbox"/>	
Q4	Chaîne	Nominale	<input checked="" type="checkbox"/>	
TV_semaine	Réel	Continue	<input checked="" type="checkbox"/>	
TV_WEEKEND	Réel	Continue	<input checked="" type="checkbox"/>	
Livres	Réel	Continue	<input checked="" type="checkbox"/>	
LivBD	Chaîne	Nominale	<input checked="" type="checkbox"/>	
BD	Réel	Continue	<input checked="" type="checkbox"/>	
Education	Chaîne	Nominale	<input checked="" type="checkbox"/>	
SEXE	Chaîne	Nominale	<input checked="" type="checkbox"/>	
AGE	Réel	Continue	<input checked="" type="checkbox"/>	
TV	Réel	Continue	<input checked="" type="checkbox"/>	
Livres1	Réel	Nominale	<input checked="" type="checkbox"/>	

Assistant logique(Filtre) Ligne : 0

Assistant logique
Cet assistant est l'équivalent interfacé de la fonction `_IF_(Test logique, Valeur Si Vrai, Valeur Si Faux)`.
Il permet de spécifier un ensemble de conditions à tester renvoyant comme résultat soit Vrai soit Faux. Selon le résultat du test logique, l'utilisateur spécifie la valeur à renvoyer (Valeur définie,

Transformations successives

Fonctions	Paramètres	Valeurs	Résultats
-----------	------------	---------	-----------

Ordonner les modalités automatiquement suivant leur ordre naturel Résultats : variable en préfixe

Afficher les erreurs



Fabrication d'une variable à deux modalités : actif, suppl.

Générateur de nouvelles variables

Fonctions

- Fonctions
 - Conversion
 - Date & Heure
 - Inspection
 - Logique
 - Math et Trigo
 - Mémoire
 - Statistiques
 - Texte
 - Outils
 - Assistant logique
 - Assistant logique (case)
 - Calculatrice
 - Script Python

Variables initiales et créées

Nom	Stockage	Rôle	Conserver ?	Valeur
ID	Réel	Identifiant	<input checked="" type="checkbox"/>	
Loisirs	Chaîne	Nominale	<input checked="" type="checkbox"/>	
TempsLibre	Chaîne	Nominale	<input checked="" type="checkbox"/>	
PlusdeTps	Chaîne	Nominale	<input checked="" type="checkbox"/>	
Q4	Chaîne	Nominale	<input checked="" type="checkbox"/>	
TV_semaine	Réel	Continue	<input checked="" type="checkbox"/>	
TV_WEEKEND	Réel	Continue	<input checked="" type="checkbox"/>	
Livres	Réel	Continue	<input checked="" type="checkbox"/>	
LivBD	Chaîne	Nominale	<input checked="" type="checkbox"/>	
BD	Réel	Continue	<input checked="" type="checkbox"/>	
Education	Chaîne	Nominale	<input checked="" type="checkbox"/>	
SEXE	Chaîne	Nominale	<input checked="" type="checkbox"/>	
AGE	Réel	Continue	<input checked="" type="checkbox"/>	
TV	Réel	Continue	<input checked="" type="checkbox"/>	
LivresAct	Réel	Nominale	<input checked="" type="checkbox"/>	

Assistant logique(Filtre) Ligne : 0 Tester

Assistant logique
Cet assistant est l'équivalent interfacé de la fonction `_IF_(Test logique, Valeur Si Vrai, Valeur Si Faux)`.
Il permet de spécifier un ensemble de conditions à tester renvoyant comme résultat soit Vrai soit Faux. Selon le résultat du test logique, l'utilisateur spécifie la valeur à renvoyer (Valeur définie,

Transformations successives

Fonctions	Paramètres	Valeurs	Résultats
Assistant logique	Filtre	Double cliquer	Rés : If

Ordonner les modalités automatiquement suivant leur ordre naturel Résultats : variable en préfixe

Afficher les erreurs Ok Annuler



Fabrication d'une variable à deux modalités : actif, suppl.

72

Assistant logique

Index	Nom	Stockage	Rôle
1	ID	Réel	Identifiant
2	Loisirs	Chaîne	Nominale
3	TempsLibre	Chaîne	Nominale
4	PlusdeTps	Chaîne	Nominale
5	Q4	Chaîne	Nominale
6	TV_semaine	Réel	Continue
7	TV_WEEKEND	Réel	Continue
8	Livres	Réel	Continue
9	LivBD	Chaîne	Nominale
10	BD	Réel	Continue
11	Education	Chaîne	Nominale
12	SEXE	Chaîne	Nominale
13	AGE	Réel	Continue
14	TV	Réel	Continue

Valeurs

Variable

Index	Valeur	Effectif
-------	--------	----------

Nombre de _Missing_ : 540

ET OU

(...)

Lien	Condition
------	-----------

Résultats

Type :

Si vrai : Constante : Variable : _Missing_

Si faux : Constante : Variable : _Missing_



Fabrication d'une variable à deux modalités : actif, suppl.

73

Assistant logique

Index	Nom	Stockage	Rôle
1	ID	Réel	Identifiant
2	Loisirs	Chaîne	Nominale
3	TempsLibre	Chaîne	Nominale
4	PlusdeTps	Chaîne	Nominale
5	Q4	Chaîne	Nominale
6	TV_semaine	Réel	Continue
7	TV_WEEKEND	Réel	Continue
8	Livres	Réel	Continue
9	LivBD	Chaîne	Nominale
10	BD	Réel	Continue
11	Education	Chaîne	Nominale
12	SEXE	Chaîne	Nominale
13	AGE	Réel	Continue
14	TV	Réel	Continue

Valeurs	Variable	
Index	Valeur	Effectif
=	1 seul	476
<>	2 en couple	525
<	3 avec des amis	1 212
<=	4 en famille	749
>	5 n.s.p.	40

Nombre de _Missing_ : 0

ET OU

Lien: Condition:

Résultats

Type :

Si vrai : Constante : Variable : _Missing_

Si faux : Constante : Variable : _Missing_

74



Fabrication d'une variable à deux modalités : actif, suppl.

Assistant logique

Index	Nom	Stockage	Rôle
1	ID	Réel	Identifiant
2	Loisirs	Chaîne	Nominale
3	TempsLibre	Chaîne	Nominale
4	PlusdeTps	Chaîne	Nominale
5	Q4	Chaîne	Nominale
6	TV_semaine	Réel	Continue
7	TV_WEEKEND	Réel	Continue
8	Livres	Réel	Continue
9	LivBD	Chaîne	Nominale
10	BD	Réel	Continue
11	Education	Chaîne	Nominale
12	SEXE	Chaîne	Nominale
13	AGE	Réel	Continue
14	TV	Réel	Continue

Valeurs

Index	Valeur	Effectif
1	manque de temps pour faire tout ce ...	1 190
2	ne manque pas de temps, mais toujo...	1 251
3	ne trouve rien de précis ☐ faire	283
4	ne fait rien de précis	243
5	n.s.p. / pas ou peu de loisirs	35

Nombre de _Missing_ : 0

ET OU Valider

(...)

Lien	Condition
ou	Loisirs = n.s.p. TempsLibre = n.s.p. / pas ou peu de loisirs

Résultats

Type : Chaîne

Si vrai : Constante : Variable : _Missing_

Si faux : Constante : Variable : _Missing_

OK Annuler



Fabrication d'une variable à deux modalités : actif, suppl.

75

Assistant logique

Index	Nom	Stockage	Rôle
1	ID	Réel	Identifiant
2	Loisirs	Chaîne	Nominale
3	TempsLibre	Chaîne	Nominale
4	PlusdeTps	Chaîne	Nominale
5	Q4	Chaîne	Nominale
6	TV_semaine	Réel	Continue
7	TV_WEEKEND	Réel	Continue
8	Livres	Réel	Continue
9	LivBD	Chaîne	Nominale
10	BD	Réel	Continue
11	Education	Chaîne	Nominale
12	SEXE	Chaîne	Nominale
13	AGE	Réel	Continue
14	TV	Réel	Continue

Nombre de _Missing_ : 0

ET OU

Valeurs	Variable
=	1 se reposer 326
<>	2 suivre des cours 275
<	3 avoir des activités physiques 658
<=	4 avoir des activités artistiques 421
>	5 se cultiver 461
>=	6 s'occuper des enfants 322
= ∅	7 s'occuper de son domicile 443
<> ∅	8 n.s.p. 96

(...)

Lien	Condition
ou	Loisirs = n.s.p.
ou	TempsLibre = n.s.p. / pas ou peu de loisirs
ou	PlusdeTps = n.s.p.

Résultats

Type :

Si vrai : Constante : Variable : _Missing_

Si faux : Constante : Variable : _Missing_

76



Fabrication d'une variable à deux modalités : actif, suppl.

Assistant logique

Index	Nom	Stockage	Rôle
1	ID	Réel	Identifiant
2	Loisirs	Chaîne	Nominale
3	TempsLibre	Chaîne	Nominale
4	PlusdeTps	Chaîne	Nominale
5	Q4	Chaîne	Nominale
6	TV_semaine	Réel	Continue
7	TV_WEEKEND	Réel	Continue
8	Livres	Réel	Continue
9	LivBD	Chaîne	Nominale
10	BD	Réel	Continue
11	Education	Chaîne	Nominale
12	SEXE	Chaîne	Nominale
13	AGE	Réel	Continue
14	TV	Réel	Continue

Valeurs Variable

=

<>

<

<=

>

>=

= ∅

<> ∅

Moyenne : 44.123
Minimum : 15
Maximum : 96

Valeur :

Nombre de _Missing_ : 0

ET OU Valider

(...)

Lien	Condition
	Loisirs = n.s.p.
ou	TempsLibre = n.s.p. / pas ou peu de loisirs
ou	PlusdeTps = n.s.p.
ou	AGE < 18.0

Résultats

Type :

Si vrai : Constante : Variable : _Missing_

Si faux : Constante : Variable : _Missing_

OK Annuler



Fabrication d'une variable à deux modalités : actif, suppl.

Générateur de nouvelles variables

Fonctions

- Fonctions
 - Conversion
 - Date & Heure
 - Inspection
 - Logique
 - Math et Trigo
 - Mémoire
 - Statistiques
 - Texte
- Outils
 - Assistant logique
 - Assistant logique (case)
 - Calculatrice
 - Script Python

Variables initiales et créées

Nom	Stockage	Rôle	Conser...	Valeur
ID	Entier	Identifiant	<input checked="" type="checkbox"/>	
Loisirs	Chaîne	Nominale	<input checked="" type="checkbox"/>	
TempsLibre	Chaîne	Nominale	<input checked="" type="checkbox"/>	
PlusdeTps	Chaîne	Nominale	<input checked="" type="checkbox"/>	
Q4	Chaîne	Nominale	<input checked="" type="checkbox"/>	
TV_semaine	Entier	Continue	<input checked="" type="checkbox"/>	
TV_WEEKEND	Entier	Continue	<input checked="" type="checkbox"/>	
Livres	Entier	Continue	<input checked="" type="checkbox"/>	
LivBD	Chaîne	Nominale	<input checked="" type="checkbox"/>	
BD	Entier	Continue	<input checked="" type="checkbox"/>	
Education	Chaîne	Nominale	<input checked="" type="checkbox"/>	
SEXE	Chaîne	Nominale	<input checked="" type="checkbox"/>	
AGE	Entier	Continue	<input checked="" type="checkbox"/>	
TV	Réel	Continue	<input checked="" type="checkbox"/>	
Livres1	Entier	Continue	<input checked="" type="checkbox"/>	
Livres2	Réel	Continue	<input checked="" type="checkbox"/>	
Sorties	Chaîne	Nominale	<input checked="" type="checkbox"/>	
Télévision	Chaîne	Nominale	<input checked="" type="checkbox"/>	
Lecture	Chaîne	Nominale	<input checked="" type="checkbox"/>	
ISUP	Chaîne	Nominale	<input checked="" type="checkbox"/>	

Ligne : 0

Tester

Transformations successives

Fonctions	Paramètres	Valeurs	Résultats
Assistant logique			
	Filtre	[? supp : actif]	Rés : ISUP

Ordonner les modalités automatiquement suivant leur ordre naturel Résultats : variable en préfixe

Afficher les erreurs

Ok Annuler

Les fenêtres de SPAD

78

The screenshot displays the SPAD 7.4 interface for a project named 'Culture_fr'. The main window shows a workflow diagram with the following steps:

- CULTURE.CULTURE
- Statistiques de base
- Remplacement données manquantes
- Mise en classes...
- Statistiques de base_1
- ACM

Intermediate steps shown in the diagram include 'Edition de libellés', 'Générateur de nouvelles variables', and 'Générateur de nouvelles variables_1'. The 'Méthodes' (Methods) pane on the left lists various statistical and data management tools. The 'Exécutions' (Executions) pane at the bottom right shows a log of the workflow's execution.

Niveau	Élément	Stop	Statut	Indicateur	Début	Temps	Résultats
0	CULTURE.CULTURE		✓ Terminé (ok)	3002	26/11/2012	00:01,235	
1	Edition de libellés		✓ Terminé (ok)		30/11/2012	00:00,31	
2	Remplacement données ...		✓ Terminé (ok)	3002	02/12/2012	00:00,375	
3	Générateur de nouvelle...		✓ Terminé (ok)	3002	02/12/2012	00:00,797	
1	Statistiques de base		✓ Terminé (ok)		30/11/2012	00:00,875	
4	Mise en classes...		✓ Terminé (ok)	3002	02/12/2012	00:00,140	
5	Générateur de nouvelle...		✓ Terminé (ok)	3002	02/12/2012	00:00,47	
6	Statistiques de base_1		✓ Terminé (ok)		02/12/2012	00:00,94	
6	ACM		Non démarré				



Conclusion

Pour tirer pleinement profit de ce semestre, poursuivre l'apprentissage simultané :

- des méthodes statistiques (théorie mathématique et méthodologie générale) ;
- de la pratique d'un logiciel (au minimum Excel) ;
- de la mise en contexte d'une discipline de recherche et de problématiques propres: sciences économiques, sociologie, science politique, géographie, etc.