

**EVALUATION ORGANISATIONNELLE DES STRUCTURES DE
GESTION DES DEPISTAGES ORGANISES DES CANCERS EN
FRANCE METROPOLITAINE ET DANS LES DOM :**

**PHASE 3, CONSTITUTION DE GROUPES HOMOGENES DE
STRUCTURES DE GESTION**

Rapport de fin d'étude

Préparé pour l'institut National du Cancer

14 juin 2010

PRESENTATION DE LA SOCIETE

REES France, est un bureau d'études spécialisé dans l'évaluation des technologies et des actions de santé. Il regroupe des évaluateurs professionnels : économistes, bio-statisticiens, informaticiens, économètres. Son objectif est d'intensifier les liaisons entre la recherche universitaire et les professionnels de santé en mettant à la disposition des différents acteurs, des techniques d'analyse et de suivi innovantes.

Son domaine d'activité privilégié : la recherche de l'efficacité comparative clinique, humaine et économique des modes de prises en charge et des traitements en pratique médicale quotidienne.

- Enquêtes observationnelles et études quasi-expérimentales
- Logiciel interactif d'impact budgétaire Excel
- Comparabilité des groupes et analyses multivariées
- Appariement selon le score de propension
- Comparaisons indirectes – Méta-analyses multitraitements
- Modèle de décision et analyse de sensibilité probabiliste
- Bootstrap sur les données individuelles
- Cahier d'observations électronique – Evaluation des réseaux
- Qualité de vie et enquête de satisfaction
- Estimation de la disposition à payer par la méthode des choix discrets

REES a publié plus de 250 articles dans des revues à Comité de Lecture et réalisé plus de cent études évaluatives en appliquant ces techniques.

RECENTES PUBLICATIONS

- Launois R, « Synthèse des données probantes et évaluation des technologies médicales », Santé Décision Management, volume 1-2 : 27-47, 2009
- Riou França L., Payet S., Le Lay K., Launois R « Drotrecogin Alfa's impact on intensive care workload in real life practice: A propensity score approach ». Value in Health November/December 2008 Volume 11 N° 7: 1051-1060
- Launois R. « Les arcanes décryptées de l'analyse médico économique à l'usage du décideur ». Journal d'Économie Médicale décembre 2008 26 (6-7) : 331-349.
- Launois R, « Synthèse des données probantes et évaluation médico-économique : les deux doigts d'une même main », Le Nouveau Cancérologue, Oct-Nov-Déc 2008.
- Launois R, Garrigues B, « Télémedecine et urgencies cérébrovasculaires », Santé Décision Management, 2008, 11 (3-4) : 91-118.
- Launois R, Payet S, Francesconi C, Riou França, Boissier MC « Modèle d'impact budgétaire lors de l'utilisation du rituximab (MabThera®) après échec d'un ou plusieurs anti-TNF α dans le traitement de la polyarthrite rhumatoïde en France » Revue du Rhumatisme 2008 Vol.75 : 1229-1236.
- Payet S, Denis P, Berdeaux G, Launois R, « Assessment of the Cost Effectiveness of Travoprost versus Latanoprost as Single Agents for Treatment of Glaucoma in France », Clin Drug Invest, 2008, 28 (3) : 183-198
- Launois R., Payet S., Saidenberg-Kermanac'h N, Francesconi C, Riou-França L, Boissier MC « Budget impact of rituximab after failure of one or more TNF α inhibitor therapies in the treatment of rheumatoid arthritis in France ». Joint Bone Spine 2008, Vol 75 : 688-695.
- Launois R, Garrigues B. « Les Unités Cérébrovasculaires : s'organiser pour prendre immédiatement en charge les AVC où qu'ils soient », Information et Santé 2008 Fédération Hospitalière de France 2008, Lettre n° 16 : 11-23.
- Brown B, Diamantopoulos A, Bernier J, Schöffski P, Hieke K, Mantovani L, Launois R, Griebsh I, Robinson P. An economic evaluation of cetuximab combined with radiotherapy for patients with locally advanced head and neck cancer in Belgium, France, Italy, Switzerland and the United Kingdom. Value in Health 2008, Vol. 11, n° 5 : 791-799
- Launois R., Le Lay K., Chemali N., Vergnenègre A. « Cost-minimisation analysis of first line polychemotherapy regimens in the treatment of advanced non small cell lung cancer » Pan Arab Journal of Oncology 2008; 1(2):28-43.
- Étude du Réseau d'Évaluation en Économie de la Santé (REES France) in *Rapports Observatoire Français des Drogues et des Toxicomanie (OFDT)*. Collection Évaluation des Politiques Publiques. 2008. « Démarche qualité et accompagnement des soins aux usagers de drogues Étude DEQUASUD » 263 p. Lionel Riou França., Livia Velpry., Daniel Benamouzig., Robert Launois.
- Launois R « Plan cancer et évaluation de l'innovation » ; Le Nouveau Cancérologue, 2008; Vol. 1 – N° 1 : 9-10
- Payet S., Denis P., Berdeaux G., Launois R « Assessment of the cost-effectiveness of travopost versus latanoprost, as single agents for glaucoma treatment in France ». Clinical Drug Investigation Journal 2008, 28 (3): 183-98
- Dhainaut JF, Payet S, Vallet B, Riou França L, Annane D, Bollaert Pe, Le Tulzo Y, Runge I, Malledant Y, Guidet B, Le Lay K, Launois R. « Cost-effectiveness of activated protein C in real life clinical practice ». Critical Care 2007; 11 :R99 (7 septembre 2007)
- Etude du Réseau d'Évaluation en Économie de la Santé (REES France). Coordinateur Robert Launois in *Rapports Office Parlementaire d'Évaluation des Politiques de Santé (OPEPS)* septembre 2007. « La prise en charge précoce des personnes victimes d'un Accident Vasculaire Cérébral » Rapporteur Jean Bardet, Député. Assemblée Nationale n° 236 – Sénat n° 475 – 27 septembre 2007 : pp 71-315., en collaboration avec Yannick Bejot (CHU DIJON), Docteur François Bourdillon (Président de la Société Française de Santé Publique, CHU PITIE- SALPETRIERE), Laurent Eckert (REES-France), Docteur Bruno Fantino (AGORA-ADIM), Stéphane Finkel (Fédération Hospitalière de France), Camille Francesconi (REES-France), Docteur Paul Garassus (Vice-président du conseil scientifique du Bureau de l'assurance qualité et de l'information médico-économique de l'hospitalisation privée: BAQIMEHP), Docteur Bernard Garrigues (Président du Collège National d'Information Médicale, Fédération Hospitalière de France et CH DU PAYS D'AIX), Yves Gaubert (Fédération Hospitalière de France), Francis Megerlin (UNIV. PARIS V) et du Professeur Yves Samson (Chef de Service des Urgences cérébrovasculaires CHU PITIE- SALPETRIERE)
- Payet S, Launois R, Saidenberg-Kermanac'h Nathalie, Francesconi C, Letellier M, Dumas A, Boissier MC « Management and total direct medical cost of patients treated with a second-line TNF inhibitor therapy for rheumatoid arthritis (RA): The TC2 study ». Annals of the Rheumatic Diseases. Annual European Congress of Rheumatology (EULAR 2007); THU0190:183.

SOMMAIRE

1. METHODES	6
1.1. La gestion des données manquantes.....	6
1.1.1. Les méthodes simples de gestion des données manquantes	6
1.1.2. Typologie et pattern des données manquantes	7
1.1.3. Les imputations multiples.....	8
1.2. La Classification Ascendante Hiérarchique	10
1.2.4. La mise en œuvre de la classification ascendante hiérarchique.....	10
1.2.5. Le choix des variables représentant les individus :	11
1.2.6. Le choix d'un critère de ressemblance :	11
1.2.7. Notion d'inertie.....	12
1.2.8. Le choix d'un critère d'agrégation :	13
2. LES DONNEES DISPONIBLES.....	14
2.1. Structures de gestion et variables	14
2.1.1. Liste des structures de gestion	14
2.1.2. Listes des variables	15
2.2. Valeurs manquantes	17
2.2.1. Classement en ordre croissant des valeurs manquantes par variables.....	17
2.2.2. Classement en ordre croissant des valeurs manquantes par structures de gestion.....	19
2.3. Structures de gestion et variables non retenues pour la classification.....	21
2.3.1. Structures de gestion non incluses pour l'analyse	21
2.3.2. Variables non incluses pour l'analyse	22
3. CLASSIFICATION ASCENDANTE HIERARCHIQUE.....	22
3.1. Dendrogramme de la CAH.....	22
3.2. Tableau des nœuds de la hiérarchie et histogramme de la perte d'inertie inter-classe.....	25
3.3. Cubic Clustering Criterion	27
4. PARTITION EN 11 CLASSES.....	28
4.1. Consolidation des classes	29
4.2. Composition des classes	29
4.3. Décomposition de l'inertie totale	31
4.4. Caractérisation des classes par les variables : Moyennes, écarts-types et test de significativité.....	32
4.5. Caractéristiques des classes.....	40
5. PARTITION EN 7 CLASSES.....	42
5.1. Consolidation des classes	42
5.2. Composition des classes	42
5.3. Décomposition de l'inertie totale	44
5.4. Caractérisation des classes par les variables : Moyennes, écarts-types et test de significativité.....	45
5.5. Descriptif des classes	49
6. CONCLUSION	51
7. REFERENCES BIBLIOGRAPHIQUES.....	52

LISTE DES TABLEAUX

Tableau 1 : Liste des 90 structures de gestion de la base	14
Tableau 2 : Listes des 44 variables	15
Tableau 3 : Classement en ordre croissant des variables en fonction du nombre de valeurs manquantes	17
Tableau 4 : Classement des structures de gestion en fonction de leur nombre de valeurs manquantes ...	19
Tableau 5 : Liste des structures de gestion non incluses pour la classification	21
Tableau 6 : Liste des variables non incluses dans la classification	22
Tableau 7 : Historique des regroupements de classe effectués au cours de la CAH et détail de la perte d'inertie inter-classe associée à ces regroupements.....	25
Tableau 8 : Impact de la consolidation des classes sur l'inertie inter-classe	29
Tableau 9 : Détail de la composition des 11 classes	29
Tableau 10 : Décomposition par classe de l'inertie inter-classe et intra-classe.....	31
Tableau 11 : Moyennes, écarts-types et test de significativité par variables pour chacune des 11 classes et pour l'ensemble des structures de gestion.....	33
Tableau 12 : Impact sur l'inertie inter-classe de la consolidation des classes	42
Tableau 13 : Détails de la composition des 7 classes de la partition	42
Tableau 14 : Décomposition par classe de l'inertie inter-classe et de l'inertie intra-classe	44
Tableau 15 : Moyennes, écarts-types et test de significativité par variables pour chacune des 11 classes et pour l'ensemble des structures de gestion.....	45

LISTE DES FIGURES

Figure 1 : Structures des données manquantes	8
Figure 2 : Principe de l'imputation multiple.....	9
Figure 3 : Processus d'agrégations successives au cours d'une classification ascendante hiérarchique..	10
Figure 4 : Représentation graphique en dendrogramme de la classification ascendante hiérarchique.....	24
Figure 5 : Histogramme de la perte d'inertie inter-classe en fonction du nombre de classe	27
Figure 6 : Courbe du Cubic Clustering Criterion en fonction du nombre de classe.....	28
Figure 7 : Représentation géographique de la classification en 11 classes.....	31
Figure 8 : Représentation géographique de la classification en 7 classes.....	44

1. Méthodes

1.1. *La gestion des données manquantes*

Dans le cadre d'analyse de données, le statisticien peut être amené à travailler sur des bases de données comportant des observations n'ayant pas été complètement renseignées, on parlera alors de données manquantes. La présence de valeurs manquantes va le plus souvent entraîner la perte d'informations pertinentes ou informative et, en fonction du nombre de ces données non renseignées, résulter par l'apparition d'un biais dans les estimations et analyses qui vont suivre.

Le statisticien dispose de plusieurs possibilités qui vont lui permettre de composer avec une base de données non complète.

1.1.1. **Les méthodes simples de gestion des données manquantes**

Plusieurs méthodes, dites simples, de gestion des données manquantes peuvent être mises en application. Ces méthodes se distinguent car elles ne permettent de créer qu'un seul jeu de données complètes, qui va par la suite être analysé en considérant que la base ne contient que des valeurs observées. Cinq méthodes simples peuvent être utilisées : l'analyse des répondants, l'imputation de la moyenne simple, l'imputation par régression, la création d'une modalité supplémentaire (dans le cas des variables catégorielles) et la reprise de la dernière observation (dans le cadre d'analyse de séries temporelles).

L'analyse des répondants consiste à éliminer toutes les observations (les individus) pour lequel certaines réponses sont manquantes. Cette méthode pose notamment des problèmes de représentativité des échantillons qui peuvent potentiellement devenir de très petite taille (dans le cas de l'analyse des structures de gestion, seules 9 structures auraient été retenues !).

L'imputation de la moyenne simple remplace les valeurs manquantes d'une variable par la moyenne arithmétique calculée à partir des données observées de cette variable. Cette méthode n'est pas applicable dans le cas de variables catégorielles (une moyenne de 1,5 pour une variable catégorisée en 0, 1, 2 et 3 n'est pas interprétable). Dans le cas de variables continues, cette méthode conduit à de mauvaises estimations des mesures d'association et des coefficients de régression car elle tend à lisser les associations. Enfin, l'imputation par la moyenne cause d'important biais d'estimation de la variance (sous-estimation) puisque les écarts entre la moyenne et la valeur imputée sont nuls.

L'imputation par régression effectue dans un premier temps une régression linéaire de la variable avec

données manquantes sur les variables sans données manquantes, puis, une fois les coefficients de corrélations obtenus, on estime la valeur de la donnée non renseignée. Cette méthode présente l'avantage de ne pas générer d'estimation biaisée de la moyenne et des coefficients de régression tout en proposant un plus grand éventail de configuration (le choix de la méthode de régression). Cependant de par la faible variabilité des imputations, la précision des coefficients de régression peut être mal estimée et conduire à des estimations incorrectes.

La création d'une modalité supplémentaire s'applique aux variables catégorielles. Elle consiste, pour la variable avec données manquantes, en la création d'une modalité « valeur manquante » (par exemple pour une variable codée 0 pour « non » et 1 pour « oui », on ajoute une modalité 2 pour « non répondu »). La mise en pratique de cette méthode peut entraîner un biais sévère lors de la réalisation d'une classification car elle peut conduire à la dissolution de plusieurs classes très différentes au sein d'un même groupe.

La reprise de la dernière observation est une méthode d'imputation qui s'applique au cas particulier des séries temporelles (données longitudinales). Cette méthode propose de remplacer une donnée manquante pour une date T+1, T+2... par la valeur observée en date T. Cette méthode entraîne de très importants biais pour le calcul de la moyenne et des covariances.

1.1.2. Typologie et pattern des données manquantes

L'un des éléments clés d'une analyse avec des données manquantes est le « mécanisme d'apparition d'une valeur manquante ». Ce terme désigne la probabilité qu'un ensemble de valeurs soit manquant sachant les valeurs prises par les données observées et non observées :

$$P(R / y_o, y_m)$$

Où R est l'indicateur de valeur observée (R=1 si Y est observée, 0 sinon), y_o est la valeur observée et y_m est la valeur manquante.

Ce mécanisme met en avant les circonstances d'apparition de valeur manquantes.

Bien que très important dans l'analyse de données manquantes, le mécanisme d'apparition n'est que très rarement connu.

Les données Missing Completely At Random (MCAR) : ce sont les données pour lesquelles la probabilité d'être manquantes est indépendante à la fois des données observées et des valeurs que prend la variable.

$$P(R / y_o, y_m) = P(R)$$

Des données manquantes en raison de la chute d'échantillons de laboratoire est un exemple de données MCAR.

Les données Missing At Random (MAR) : l'apparition d'une valeur manquante dépend des données observées mais pas des données non observées.

$$P(R / y_o, y_m) = P(R / y_o)$$

Dans un questionnaire, une question faisant référence à la précédente (question 1 : oui/non ; question 2 : si oui...) est un exemple de données MAR.

Les données Missing Not At Random (MNAR) : la probabilité que la donnée soit manquante ne dépend pas des données observées mais des données non observées.

$$P(R / y_o, y_m) = P(R / y_m)$$

Un revenu non indiqué car trop important est un exemple de données MNAR.

En complément de la typologie des données manquantes, il est important d'analyser le pattern de ces données, c'est-à-dire leur implémentation. On distingue deux cas : les valeurs manquantes monotones et non monotones.

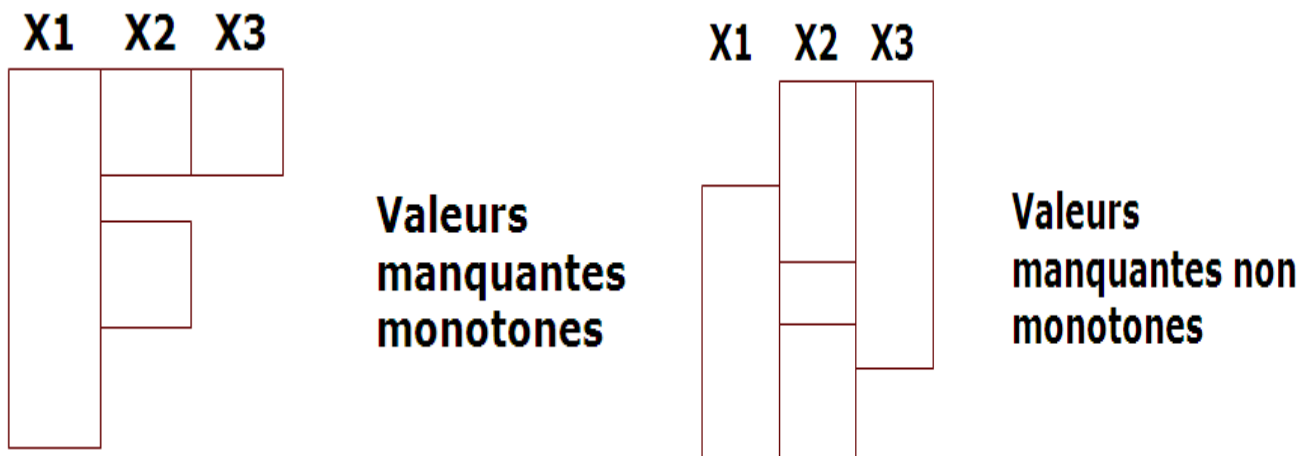


Figure 1 : Structures des données manquantes

1.1.3. Les imputations multiples

La méthode des imputations multiples est une méthode qui va permettre de compléter les données non

renseignées d'une base de données. Elle a été développée par Rubin en 1987.

Elle consiste en la création de plusieurs jeu de valeurs possibles pour remplacer une valeur manquante. L'objectif de cette méthode n'est pas de prédire les données manquantes avec la plus grande précision possible, ni de les décrire. Il s'agit de proposer une série de valeurs plausibles obtenues en cherchant à refléter correctement l'incertitude sur les valeurs non renseignées, à préserver les distributions et les relations entre les variables.

L'imputation multiple se déroule en trois étape : le repérage des données incomplètes, la constitution de $N > 1$ jeux de données pour lesquels les NA ont été remplacées, le pooling des résultats pour conduire à l'estimation de la valeur manquante.

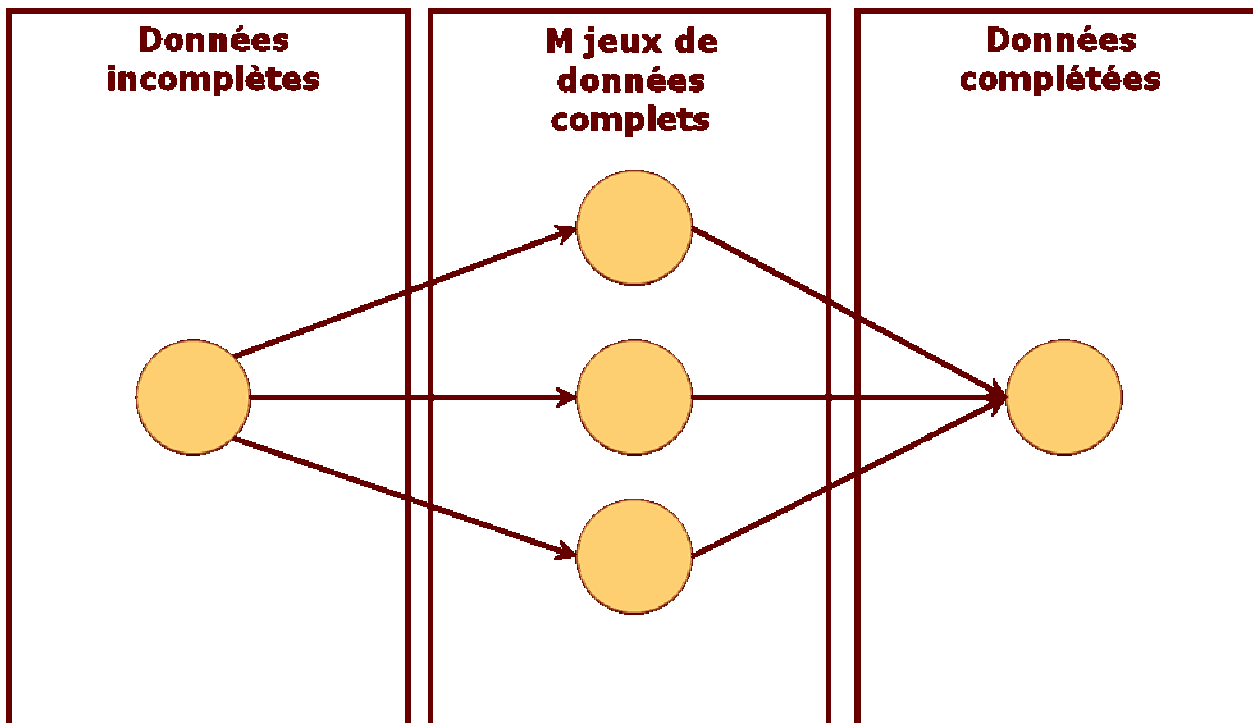


Figure 2 : Principe de l'imputation multiple

L'imputation multiple est une méthode qui apporte l'avantage de proposer des résultats qui incorporent l'incertitude de l'absence de réponse.

Deux méthodes sont principalement utilisées pour la mises en œuvre d'imputations multiples : l'algorithme EM (Expectation-Maximisation) et la méthode MCMC (Monte-Carlo Markov Chain).

L'algorithme EM est un processus en 2 étapes : on identifie les distributions des données manquantes en fonction des données observées et des variables explicatives, puis on remplace les données manquantes par les valeurs attendues. Ces étapes sont répétées jusqu'à voir une stabilisation des valeurs attendues.

La méthode MCMC consiste en la création d'une chaîne de Markov qui va converger vers la distribution

prédictive a posteriori et permettre l'estimation des données manquantes.

1.2. La Classification Ascendante Hiérarchique

Une classification a pour objectif de regrouper entre elles des observations dont la similarité a été établie en se basant sur un ou plusieurs critères. Les diverses techniques de classification existantes cherchent toutes à répartir les n observations, aussi appelées individus, caractérisées par p variables X_1, X_2, \dots, X_p en un certain nombre m de sous-groupes aussi homogènes que possible.

La classification hiérarchique est l'une des deux grandes familles de techniques de classification. En fonction du niveau de précision appliqué, la classification hiérarchique permet de rassembler deux individus dans un groupe, aussi appelé classe. Plus la précision est élevée, plus les individus seront distingués les uns des autres ; a contrario, plus la précision est faible, plus le nombre de classe sera faible. De la classification hiérarchique résulte une représentation graphique intuitive sous forme d'arbre, le dendrogramme, des regroupements successifs entre les individus et les groupes d'individus.

La classification hiérarchique devient ascendante lorsque son fonctionnement part du cas particulier (1 individu seul par classe) pour remonter vers le cas général (1 seule classe regroupant tous les individus).

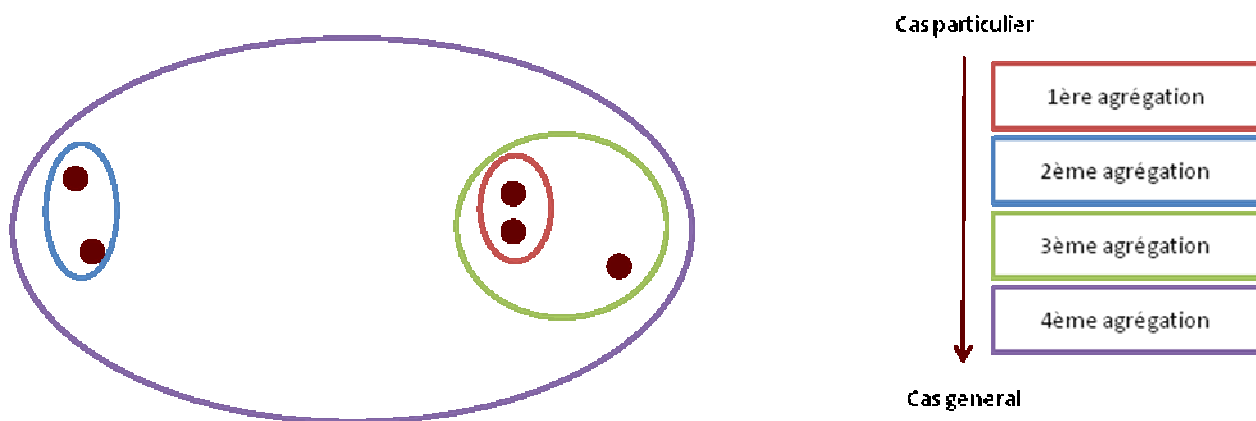


Figure 3 : Processus d'agréations successives au cours d'une classification ascendante hiérarchique

1.2.4. La mise en œuvre de la classification ascendante hiérarchique

La mise en œuvre de la classification ascendante hiérarchique peut se décomposer en 4 étapes :

- Le choix des variables représentant les individus
- Le choix d'un critère de ressemblance
- Le choix d'un critère d'agrégation
- Le choix du nombre de classe

1.2.5. Le choix des variables représentant les individus :

Les variables sont les critères sur lesquels les individus vont se différencier ou se ressembler. Lorsque le nombre de variables est trop important, il peut être pertinent d'en éliminer certaines pour ne garder que celles qui comportent le plus d'information (les plus discriminantes) en pratiquant une analyse en composantes principales (ACP) ou une analyse en composantes multiples (ACM).

Le calcul d'une distance entre les individus à partir des valeurs des variables peut causer un biais en ce qu'il repose sur des variables ne présentant pas les mêmes unités : la contribution de certaines variables aux distances entre les individus peut ainsi s'en trouver minimisée ou maximisée. Une solution pour contrer ce problème consiste à centrer puis réduire les variables.

1.2.6. Le choix d'un critère de ressemblance :

Afin de regrouper les individus les plus proches et dissocier les individus les plus éloignés, il est préalablement nécessaire de définir un critère de ressemblance, parfois appelé critère de dissimilarité. Cette notion de ressemblance est appréhendée par le calcul de la distance qui sépare deux individus.

Il existe plusieurs méthodes possibles pour calculer la distance entre les individus. Nous en présentons ici quelques unes :

- La distance euclidienne : il s'agit de la distance géométrique classique. C'est la distance la plus couramment utilisée et c'est celle que nous avons retenue pour effectuer la classification.

$$d_2(I_i, I_j) = \sqrt{\sum_k (x_{ik} - x_{jk})^2}$$

- La distance de Manhattan :

$$d_1(I_i, I_j) = \sum_k |x_{ik} - x_{jk}|$$

- La distance de Tchebychev :

$$d_\infty(I_i, I_j) = \text{Max}_k |x_{ik} - x_{jk}|$$

Une fois l'ensemble des distances entre les individus, il convient de les regrouper au sein d'une matrice des distances. Cette matrice est carrée (n lignes et n colonnes), symétrique ($d_{ij}=d_{ji}$), positive et présente des coefficients nuls sur la diagonale ($d_{ii}=0$).

1.2.7. Notion d'inertie

L'inertie d'un nuage de point est un indicateur de son homogénéité, c'est-à-dire qu'elle renseigne sur sa dispersion : une inertie faible traduit une forte homogénéité des points du nuage.

L'inertie totale correspond à l'inertie calculée sur l'ensemble des points du nuage. Elle est calculée à partir d'une distance (par exemple la distance euclidienne) et selon la formule :

$$I_T = \sum_{i=1}^n p_i d(M_i, G)^2$$

Où p_i est le poids associé à l'individu i et G est le centre de gravité du nuage.

L'inertie se calcule donc comme la somme des carrés des écarts au centre de gravité, et peut donc à ce titre être interprété comme une variance.

De la même manière, l'inertie peut être calculée pour une classe, en ne considérant que les points qui composent la classe et en utilisant le centre de gravité de la classe.

On peut alors définir l'inertie intra-classe comme la mesure de l'homogénéité de l'ensemble des classes qui composent une partition Γ : il s'agit de la somme de l'inertie de chacune des classes de cette partition.

$$I_{\text{intra}}(\Gamma) = \sum_{j=1}^k \gamma_j$$

L'inertie totale d'un nuage n'est pas égale à la somme des inerties des classes qui le compose, c'est-à-dire l'inertie intra-classe. Il faut tenir de l'inertie entre les classes, l'inertie inter-classe.

L'inertie interclasse est une mesure de l'hétérogénéité entre les classes, un indicateur du niveau de différenciation entre les classes. On la calcule de la façon suivante :

$$I_{\text{inter}}(\Gamma) = \sum_{j=1}^k \mu_j d(G_j, G)^2$$

Où G_j est le centre de gravité de la classe j , G est le centre de gravité du nuage et μ_j le poids de la classe j . Il s'agit donc de la somme des carrés des écarts entre les centres de gravité des classes et le centre de gravité du nuage, et peut donc être considérée comme une variance. Plus l'inertie inter-classe est importante, plus les classes sont différentes.

1.2.8. Le choix d'un critère d'agrégation :

La réalisation d'une classification ascendante hiérarchique implique le choix d'un critère de ressemblance entre les individus, pour pouvoir regrouper les individus les plus proches, mais aussi le choix d'un critère d'agrégation afin de fusionner les classes qui se ressemblent le plus. Plusieurs critères d'agrégation différents peuvent être mis en œuvre : single linkage, complete linkage, average, centroïd, méthode de Ward...

Pour cette étude, nous avons retenu la méthode de Ward.

La méthode de Ward repose sur le principe d'inertie inter-classe. Elle cherche à minimiser la somme des carrés des écarts de tous les couples de classes qui peuvent être formés à chaque étape, c'est-à-dire choisir le regroupement qui fera le moins baisser l'inertie inter-classe. La distance de Ward entre deux classes A et B, de centres de gravité respectifs G_A et G_B , et d'effectif n_A et n_B est définie par :

$$D(A, B) = \frac{d_2(G_A, G_B)}{\frac{1}{n_A} + \frac{1}{n_B}}$$

Où d_2 est la distance euclidienne.

2. Les données disponibles

2.1. Structures de gestion et variables

2.1.1. Liste des structures de gestion

Tableau 1 : Liste des 90 structures de gestion de la base

Structures de gestion			
1	Ain	46	Indre et Loire
2	Aisne	47	Isère
3	Allier (ABIDEC)	48	Landes (Centre de Gestion du Dépistage des Cancers 40)
4	Alpes de Haute-Provence et Alpes Maritimes	49	Loir et Cher
5	Ardèche et Drome	50	Loire
6	Ardennes	51	Loire-Atlantique
7	Ariège	52	Loiret
8	Aube	53	Lot
9	Aude (ADOC11)	54	Lot et Garonne (Structure de gestion du dépistage du Lot et Garonne)
10	Aveyron	55	Maine et Loire
11	Bas-Rhin & Haut-Rhin ADECA CCR (67+68)	56	Manche (IRIS MANCHE)
12	Bas-Rhin & Haut-Rhin ADEMAS CS (67+68)	57	Marne
13	Bouches-du-Rhône	58	Martinique
14	Calvados (Association MATHILDE)	59	Mayenne
15	Cantal, Haute-Loire, Puy de Dôme (ABIDOC)	60	Meurthe et Moselle
16	Charente	61	Meuse
17	Charente-Maritime	62	Morbihan (ADECAM)
18	Cher (IRSA ADOC18)	63	Moselle
19	Corrèze	64	Nord
20	Corse	65	Oise
21	Cote d'Or + Nièvre (ADECA 21)	66	Orne (Régie départementale des cancers)
22	Cotes d'Armor (ADEC)	67	Paris
23	Creuse	68	Pas-de-Calais
24	Deux-Sèvres	69	Pyrénées-Atlantiques (PYRADEC PAU)
25	Dordogne (vitalis)	70	Pyrénées-Orientales
26	Doubs, Jura, Haute Saône et Territoire de Belfort	71	Réunion
27	Essonne	72	Rhône
28	Eure	73	Saône et Loire (ADEMAS 71)
29	Eure et Loir (ADOC 28)	74	Sarthe

Structures de gestion			
30	Finistère (ADEC 29)	75	Savoie
31	Gard et Lozère	76	Seine et Marne
32	Gers	77	Seine-Maritime
33	Gironde (AGIDECA 33)	78	Seine-Saint-Denis
34	Guadeloupe	79	Somme
35	Guyane	80	Tarn
36	Haute-Garonne	81	Tarn et Garonne
37	Haute-Marne	82	Val de Marne
38	Hauts-Alpes	83	Val d'Oise
39	Haute-Savoie	84	Var
40	Hauts-Pyrénées	85	Vaucluse
41	Haute-Vienne	86	Vendée
42	Hauts-de-Seine	87	Vienne
43	Hérault	88	Vosges
44	Ille et Vilaine (ADEC 135)	89	Yonne (AIDEC)
45	Indre	90	Yvelines

La base d'origine comprend des données pour les 90 structures de gestion qui pratiquent le dépistage du cancer colorectal et/ou du cancer du sein. Par la suite, en raison de la nécessité de gérer les données manquantes de la base, certaines de ces structures ont été écartées de l'analyse.

2.1.2. Listes des variables

Après le choix des variables les plus pertinentes par les experts de l'Institut National du Cancer, 44 variables sont documentées dans la base de données.

Tableau 2 : Listes des 44 variables

Variables	
v2	Procédure complémentaire d'amélioration des fichiers par le fichier de la poste
v3	Délai de transmission : 0 : non transmis ou incomplets 1 : > 60 jours 2 : 45 < 60 jours 3 : 30 < 45 jours 4 : 15 < 30 jours 5 : < 15 jrs 6 : respect du délai ou en avance
v4	Densité médecins généralistes par rapport à la population totale 2008 (nombre de médecins pour 1 000 habitants)
v5	Densité radiologues par rapport à la population totale 2008 (nombre de médecins pour 1 000 habitants)
v6	Densité gynécologues par rapport à la population totale 2008 (nombre de médecins pour 1 000 habitants)

Variables	
v7	Densité HGE par rapport à la population totale 2008 (nombre de médecins pour 1 000 habitants)
v8	Statut juridique de la structure
v9	Autre dépistage pratiqué
v12	Retours d'information, au moins une fois par an, aux médecins généralistes sur leur activité
v13	Niveau de vie du département : 1 : entre 13740€ et 14820€ 2 : entre 14820€ et 15766€ 3 : entre 15766€ et 17350€ 4 : entre 17350€ et 20360€
v14	Nombre d'etp dans la structure 2008
v15	Personnel mis à disposition 2008 (oui/non)
v16	Une personne est dédiée aux tâches administratives et financières (oui à partir de 0,80ETP/non) 2008
v17	Budget (<i>global</i>) prévisionnel 2008 (ou réalisé 2008 /nombre d'etp dans la SG
v18	Coût par personne dépistée CS (coût CS/nombre personnes dépistées) 2008
v19	Pourcentage de participation INVITEES (CS) 2008
v20	Population invitée CS / nombre ETP 2008
v21	Date d'entrée de la SG dans le programme de dépistage du cancer du sein
v22	Mise à disposition des locaux 2008
v23	Budget réalisé/budget prévisionnel 2008
v24	Population INSEE cible CS 2008
v25	Date d'entrée de la SG dans le programme de dépistage du cancer colorectal
v26	Coût par personne dépistée CCR (coût CCR/nombre personnes dépistées) 2008
v27	Population INSEE cible CCR 2008
v28	Pourcentage de participation INVITEES (CCR) 2008
v29	Population invitée CCR / nombre ETP 2008
v30	Part du financement du Conseil Général dans le financement total, en pourcentage
v31	Mise en place de partenariats
v32	Externalisation de tâches par la SG
v33	Partage de moyens avec une autre SG
v34	Retours d'information, au moins une fois par an, aux radiologues sur leur activité
v35	Participation du médecin coordonnateur aux séminaires nationaux du programme de dépistage (journées INVS, DGS, inca)
v36	Part des coûts structurels (-loyer) dans le coût total 2008
v37	Mise en place d'outils favorisant la communication (site internet ou numéro vert par exemple)
v38	Formations organisées destinées au personnel de la SG
v39	Existence d'une convention collective au sein de la SG
v40	Mise en place de réunions avec l'ensemble de l'équipe de la SG
v41	Mise en place de procédures internes pour le suivi de l'assurance qualité
v42	Délai moyen, en jours, entre la mammographie et le rendu des résultats de 2ème lecture au patient
v43	Organisation d'évènements, au moins une fois par an, pour informer les professionnels de santé au dépistage organisé

Variables	
v44	Mise en place d'initiatives locales spécifiques (par exemple, initiatives pour la communication, offres de transport, actions envers des professionnels de santé particuliers, etc.)
v45	Ratio pour le cancer du sein : nombre de perdus de vue / population dépistée positive en 2008
v46	Ratio pour le cancer colorectal : nombre de perdus de vue (aucune nouvelle 1 an après le test hémocult) / population dépistée positive en 2008
v47	Existence d'un registre des tumeurs ou autre base de données au niveau départemental

En raison d'un taux de réponse trop faible (inférieur à 70%), deux de ces variables n'ont pas été retenues pour participer à l'analyse. Les deux variables présentant les dates d'entrée des SG dans les programmes de dépistages ont par la suite été recodée en variable d'ancienneté dans le dépistage.

2.2. Valeurs manquantes

2.2.1. Classement en ordre croissant des valeurs manquantes par variables

Le tableau 3 présente le nombre de réponses manquantes pour chacune des variables de la base. Deux variables atteignent et dépassent 30 données manquantes, ce qui correspond à un taux de réponse inférieur à 70%.

Tableau 3 : Classement en ordre croissant des variables en fonction du nombre de valeurs manquantes

Variables		# de valeurs manquantes	% sur 90 structures
v2	Procédure complémentaire d'amélioration des fichiers par le fichier de la poste	0	0%
v3	Délai de transmission : 0 : non transmis ou incomplets 1 : > 60 jours 2 : 45 < 60 jours 3 : 30 < 45 jours 4 : 15 < 30 jours 5 : < 15 jrs 6 : respect du délai ou en avance	0	0%
v4	densité médecins généralistes par rapport à la population totale 2008 (nombre de médecins pour 1 000 habitants)	0	0%
v5	densité radiologues par rapport à la population totale 2008 (nombre de médecins pour 1 000 habitants)	0	0%
v6	densité gynécologues par rapport à la population totale 2008 (nombre de médecins pour 1 000 habitants)	0	0%
v7	densité HGE par rapport à la population totale 2008 (nombre de médecins pour 1 000 habitants)	0	0%
v8	statut juridique de la structure	0	0%
v9	autre dépistage pratiqué	0	0%
v34	Retours d'information, au moins une fois par an, aux radiologues sur leur activité	1	1%

Variables		# de valeurs manquantes	% sur 90 structures
v13	Niveau de vie du département : 1 : entre 13740€ et 14820€ 2 : entre 14820€ et 15766€ 3 : entre 15766€ et 17350€ 4 : entre 17350€ et 20360€	4	4%
v14	nombre d'ETP dans la structure 2008	4	4%
v15	personnel mis à disposition 2008 (oui/non)	4	4%
v16	une personne est dédiée aux tâches administratives et financières (oui à partir de 0,80ETP/non) 2008	4	4%
v17	budget (<i>global</i>) prévisionnel 2008 (ou réalisé 2008 /nombre d'ETP dans la SG	4	4%
v18	coût par personne dépistée CS (coût CS/nombre personnes dépistées) 2008	4	4%
v20	Population invitée CS / nombre ETP 2008	5	6%
v21	date d'entrée de la SG dans le programme de dépistage du cancer du sein	5	6%
v22	mise à disposition des locaux 2008	5	6%
v23	budget réalisé/budget prévisionnel 2008	5	6%
v28	Pourcentage de participation INVITEES (CS) 2008	5	6%
v24	population INSEE cible CS 2008	6	7%
v25	date d'entrée de la SG dans le programme de dépistage du cancer colorectal	6	7%
v26	coût par personne dépistée CCR (coût CCR/nombre personnes dépistées) 2008	6	7%
v19	Pourcentage de participation INVITEES (CCR) 2008	8	9%
v27	population INSEE cible CCR 2008	8	9%
v29	Population invitée CCR / nombre ETP 2008	8	9%
v30	Part du financement du Conseil Général dans le financement total, en pourcentage	8	9%
v31	Mise en place de partenariats	8	9%
v32	Externalisation de tâches par la SG	8	9%
v33	Partage de moyens avec une autre SG	8	9%
v12	Retours d'information, au moins une fois par an, aux médecins généralistes sur leur activité	9	10%
v35	Participation du médecin coordonnateur aux séminaires nationaux du programme de dépistage (journées InVS, DGS, INCa)	9	10%
v36	part des coûts structurels (-loyer) dans le coût total 2008	10	11%
v37	Mise en place d'outils favorisant la communication (site internet ou numéro vert par exemple)	10	11%
v38	Formations organisées destinées au personnel de la SG	10	11%
v39	Existence d'une convention collective au sein de la SG	10	11%
v40	Mise en place de réunions avec l'ensemble de l'équipe de la SG	10	11%
v41	Mise en place de procédures internes pour le suivi de l'assurance qualité	11	12%
v42	Délai moyen, en jours, entre la mammographie et le rendu des résultats de 2ème lecture au patient	12	13%

Variables		# de valeurs manquantes	% sur 90 structures
v43	Organisation d'évènements, au moins une fois par an, pour informer les professionnels de santé au dépistage organisé	14	16%
v44	Mise en place d'initiatives locales spécifiques (par exemple, initiatives pour la communication, offres de transport, actions envers des professionnels de santé particuliers, etc.)	14	16%
v45	Ratio pour le cancer du sein : nombre de perdus de vue / population dépistée positive en 2008	17	19%
v46	Ratio pour le cancer colorectal : nombre de perdus de vue (aucune nouvelle 1 an après le test hémocult) / population dépistée positive en 2008	33	37%
v47	Existence d'un registre des tumeurs ou autre base de données au niveau départemental	67	74%
TOTAL		360	

2.2.2. Classement en ordre croissant des valeurs manquantes par structures de gestion

Tableau 4 : Classement des structures de gestion en fonction de leur nombre de valeurs manquantes

Structure	# de valeurs manquantes	% sur 44 variables
Calvados (Association MATHILDE)	0	0%
Cantal, Haute-Loire, Puy de Dôme (ABIDOC)	0	0%
Doubs, Jura, Haute Saône et Territoire de Belfort	0	0%
Gironde (AGIDECA 33)	0	0%
Indre et Loire	0	0%
Isère	0	0%
Marne	0	0%
Rhône	0	0%
Somme	0	0%
Ain	1	2%
Aisne	1	2%
Allier (ABIDEC)	1	2%
Aube	1	2%
Aveyron	1	2%
Charente-Maritime	1	2%
Corrèze	1	2%
Cote d'Or + Nièvre (ADECA 21)	1	2%
Deux-Sèvres	1	2%
Eure et Loir (ADOC 28)	1	2%
Haute-Garonne	1	2%
Loire-Atlantique	1	2%
Manche (IRIS MANCHE)	1	2%
Mayenne	1	2%
Meuse	1	2%
Moselle	1	2%
Paris	1	2%
Pyrénées-Atlantiques (PYRADEC PAU)	1	2%

Structure	# de valeurs manquantes	% sur 44 variables
Saône et Loire (ADEMAS 71)	1	2%
Seine et Marne	1	2%
Val d'Oise	1	2%
Vaucluse	1	2%
Yonne (AIDEC)	1	2%
Yvelines	1	2%
Alpes de Haute-Provence et Alpes Maritime	2	5%
Ardennes	2	5%
Aude (ADOC11)	2	5%
Charente	2	5%
Dordogne (vitalis)	2	5%
Essonne	2	5%
Finistère (ADEC 29)	2	5%
Haute-Marne	2	5%
Haute-Savoie	2	5%
Haute-Vienne	2	5%
Hérault	2	5%
Ille et Vilaine (ADEC 135)	2	5%
Loir et Cher	2	5%
Lot	2	5%
Lot et Garonne (Structure de gestion du dépistage du Lot et Garonne)	2	5%
Maine et Loire	2	5%
Morbihan (ADECAM)	2	5%
Oise	2	5%
Pas-de-Calais	2	5%
Sarthe	2	5%
Savoie	2	5%
Tarn	2	5%
Val de Marne	2	5%
Var	2	5%
Vosges	2	5%
Ariège	3	7%
Cher (IRSA ADOC18)	3	7%
Creuse	3	7%
Meurthe et Moselle	3	7%
Pyrénées-Orientales	3	7%
Réunion	3	7%
Seine-Saint-Denis	3	7%
Ardèche et Drome	4	9%
Cotes d'Armor (ADEC)	4	9%
Gard et Lozère	4	9%
Hautes-Alpes	4	9%
Loire	4	9%
Orne (Régie départementale des cancers)	4	9%
Tarn et Garonne	4	9%
Hautes-Pyrénées	5	11%
Hauts-de-Seine	5	11%
Indre	5	11%
Gers	6	14%

Structure	# de valeurs manquantes	% sur 44 variables
Vendée	6	14%
Guadeloupe	7	16%
Bas-Rhin & Haut-Rhin ADEMAS CS (67+68)	11	25%
Bas-Rhin & Haut-Rhin ADECA CCR (67+68)	12	27%
Bouches-du-Rhône	14	32%
Corse	16	36%
Eure	16	36%
Landes (Centre de Gestion du Dépistage des Cancers 40)	16	36%
Martinique	16	36%
Vienne	16	36%
Nord	17	39%
Seine-Maritime	17	39%
Guyane	22	50%
Loiret	30	68%
TOTAL	360	

Parmi les 90 structures de gestion considérées, 10 atteignent un taux de valeurs manquantes supérieur à 30%.

2.3. Structures de gestion et variables non retenues pour la classification

2.3.1. Structures de gestion non incluses pour l'analyse

Dix structures de gestion ont été écartées en raison de leur nombre trop important de valeurs manquantes. Celles-ci présentaient en effet un taux de réponse inférieur à 70%, seuil retenu pour garantir la cohérence de l'imputation des données manquantes.

Tableau 5 : Liste des structures de gestion non incluses pour la classification

Structure	# de valeurs manquantes	Pourcentage
Bouches-du-Rhône	14/44	32%
Corse	16/44	36%
Eure	16/44	36%
Landes (Centre de Gestion du Dépistage des Cancers 40)	16/44	36%
Martinique	16/44	36%
Vienne	16/44	36%
Nord	17/44	39%
Seine-Maritime	17/44	39%
Guyane	22/44	50%
Loiret	30/44	68%

De plus, trois autres structures ont été exclues de l'analyse : Guadeloupe, Bas-Rhin & Haut-Rhin

ADEMAS CS (67+68), et Bas-Rhin & Haut-Rhin ADECA CCR (67+68). Il s'agit de structure ne pratiquant que l'un des deux dépistages : cancer du sein (Guadeloupe, Bas-Rhin & Haut-Rhin ADEMAS CS (67+68)) ou cancer colorectal (Bas-Rhin & Haut-Rhin ADECA CCR (67+68)). Dans ces trois cas, le remplacement des données manquantes n'est pas justifiable, or la classification ascendante hiérarchique ne peut mathématiquement pas être mise en œuvre sur des données manquantes.

2.3.2. Variables non incluses pour l'analyse

Deux des quarante-quatre variables de la base de données ont été écartées de l'analyse, en raison du nombre de valeurs manquantes trop important (>30%) :

Tableau 6 : Liste des variables non incluses dans la classification

Variables	Nombre de valeurs manquantes	Pourcentage
Ratio pour le cancer colorectal : nombre de perdus de vue (aucune nouvelle 1 an après le test hémocult) / population dépistée positive en 2008	33/90	37%
Existence d'un registre des tumeurs ou autre base de données au niveau départemental	67/90	74%

3. Classification Ascendante hiérarchique

Une analyse de regroupement (cluster analysis) hiérarchique a été réalisée afin d'identifier des groupes (clusters) de structures de gestion qui ont en commun des caractéristiques similaires. Cette méthode statistique vise à créer des groupes d'objets de telle manière que les objets appartenant à un groupe soient similaires, et que les objets appartenant à des groupes distincts soient différents. Il s'agit de regrouper itérativement les structures de gestion, en commençant par les cas particuliers (une seule structure par classe) et en construisant progressivement un arbre, ou dendrogramme, regroupant finalement toutes les structures en une seule classe, à la racine à droite. Ceci suppose de savoir calculer, à chaque étape ou regroupement, la distance entre une structure et une classe ainsi que celle entre deux classes. Ceci nécessite donc de définir une distance. Différents choix, appelés saut en français et linkage en anglais, sont possibles. Le nombre de classes est déterminé a posteriori, principalement à la vue du dendrogramme ou d'un graphique représentant la décroissance de la hauteur de chaque saut, ou écart de distance, opéré à chaque regroupement

3.1. Dendrogramme de la CAH

Le dendrogramme représente graphiquement le résultat de l'analyse de regroupement réalisée en utilisant le critère de Ward comme distance de mesure et un mode de regroupement qui minimise la perte d'inertie interclasse. Les liens entre les objets sont représentés par des lignes en U « couchés horizontalement », dont la longueur indique la distance entre les objets. Onze niveaux de distance, sur l'axe des x du dendrogramme, ont été considérés successivement, de façon à identifier des clusters de plus en plus petits et homogènes.

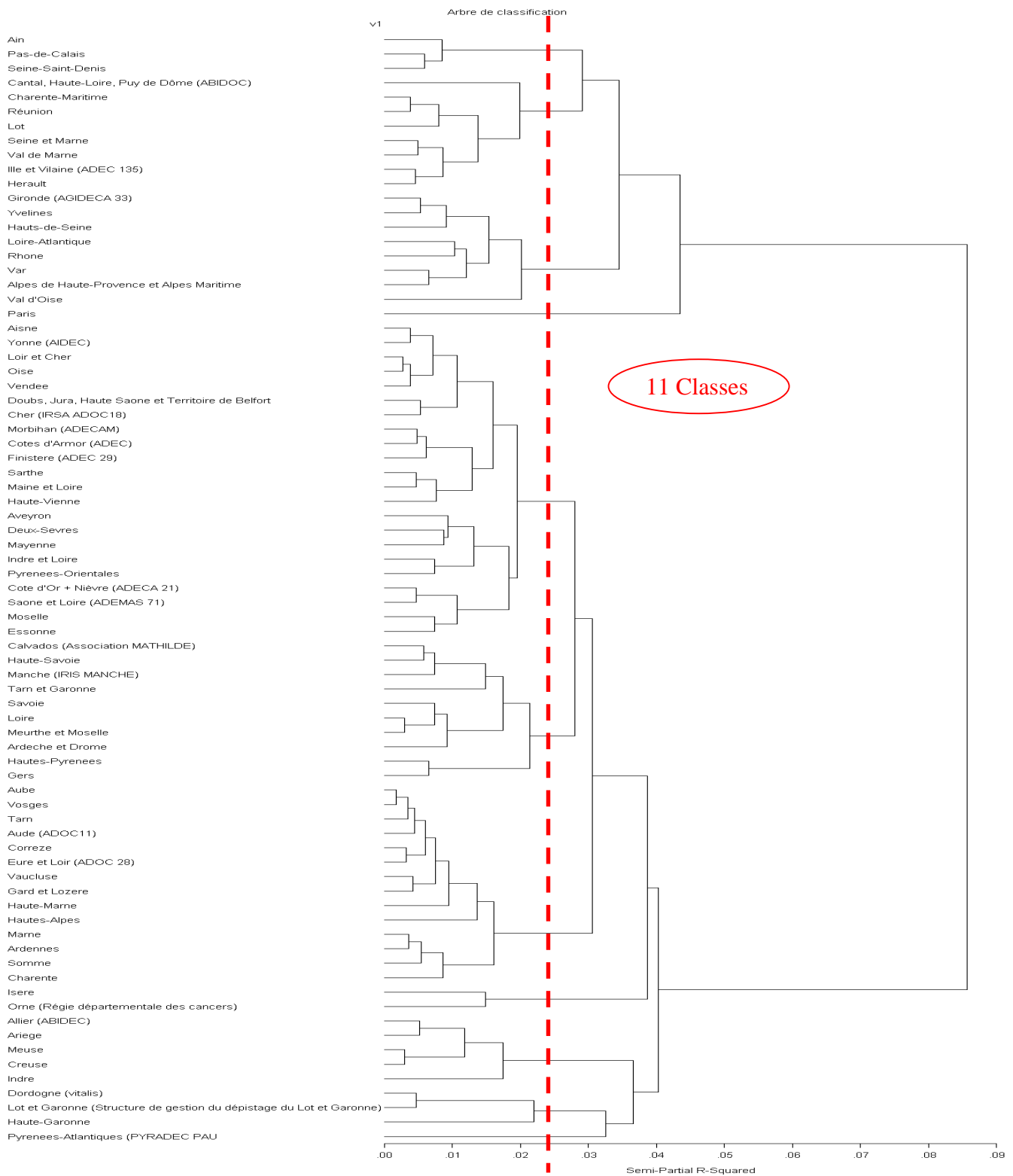


Figure 4 : Représentation graphique en dendrogramme de la classification ascendante hiérarchique

3.2. **Tableau des nœuds de la hiérarchie et histogramme de la perte d'inertie inter-classe**

Ce tableau retrace l'historique des regroupements de structures et de classes. Il apporte de plus le détail de la perte d'inertie inter-classe (la perte d'hétérogénéité entre les classes) à mesure que sont réalisés les regroupements.

Tableau 7 : Historique des regroupements de classe effectués au cours de la CAH et détail de la perte d'inertie inter-classe associée à ces regroupements

Nœud	Effectif pondéré	Classes jointes	Perte d'inertie inter-classe		
			0/00	Cumulée	Différentielle
CL1	77.0	CL2-CL3	86	86	.
CL2	20.0	CL6-Paris	43	129	42
CL3	57.0	CL4-CL5	40	169	3
CL4	48.0	CL8-CL23	39	208	2
CL5	9.0	CL17-CL7	37	245	2
CL6	19.0	CL9-CL13	34	279	2
CL7	4.0	CL11-Pyren	32	312	2
CL8	46.0	CL10-CL19	31	342	2
CL9	11.0	CL40-CL14	29	371	2
CL10	32.0	CL15-CL12	28	399	1
CL11	3.0	CL62-Haute	22	421	6
CL12	10.0	CL18-CL49	21	443	1
CL13	8.0	CL21-Val d	20	463	1
CL14	8.0	Canta-CL24	20	483	0
CL15	22.0	CL20-CL16	20	502	0
CL16	9.0	CL26-CL30	18	520	1
CL17	5.0	CL29-Indre	17	538	1
CL18	8.0	CL22-CL35	17	555	0
CL19	14.0	CL25-CL39	16	571	1
CL20	13.0	CL31-CL27	16	587	0
CL21	7.0	CL36-CL28	15	603	1
CL22	4.0	CL45-Tarn	15	617	0
CL23	2.0	Isère-Orne	15	632	0
CL24	7.0	CL41-CL38	14	646	1
CL25	10.0	CL33-Haute	14	660	0
CL26	5.0	CL34-CL44	13	673	0
CL27	6.0	CL51-CL42	13	686	0
CL28	4.0	CL32-CL50	12	698	1
CL29	4.0	CL58-CL74	12	710	0
CL30	4.0	CL63-CL47	11	720	1
CL31	7.0	CL48-CL57	11	731	0
CL32	2.0	Loire-Rhône	10	741	0

Nœud	Effectif pondéré	Classes jointes	Perte d'inertie inter-classe		
			0/00	Cumulée	Différentielle
CL33	9.0	CL43-Haute	10	751	1
CL34	3.0	Aveyr-CL37	9	760	0
CL35	4.0	CL46-Ardec	9	769	0
CL36	3.0	CL56-Hauts	9	779	0
CL37	2.0	Deux--Mayen	9	787	0
CL38	4.0	CL59-CL64	9	796	0
CL39	4.0	CL55-Chare	9	805	0
CL40	3.0	Ain-CL53	8	813	0
CL41	3.0	CL67-Lot	8	821	0
CL42	3.0	CL61-Haute	8	829	0
CL43	8.0	CL52-CL66	8	836	0
CL44	2.0	Indre-Pyren	7	844	0
CL45	3.0	CL54-Manch	7	851	0
CL46	3.0	Savoi-CL73	7	858	0
CL47	2.0	Mosel-Esson	7	866	0
CL48	5.0	CL69-CL68	7	873	0
CL49	2.0	Haute-Gers	6	879	1
CL50	2.0	Var-Alpes	6	886	0
CL51	3.0	CL60-Finis	6	892	0
CL52	6.0	CL65-CL72	6	898	0
CL53	2.0	Pas-d-Seine	6	904	0
CL54	2.0	Calva-Haute	6	910	0
CL55	3.0	CL70-Somme	5	915	0
CL56	2.0	Giron-Yveli	5	920	0
CL57	2.0	Doubs-Cher	5	926	0
CL58	2.0	Allie-Arieg	5	931	0
CL59	2.0	Seine-Val d	5	936	0
CL60	2.0	Morbi-Cotes	5	941	0
CL61	2.0	Sarth-Maine	5	945	0
CL62	2.0	Dordo-Lot e	5	950	0
CL63	2.0	Cote-Saône	5	955	0
CL64	2.0	Ille-Herau	5	959	0
CL65	4.0	CL71-Aude	4	964	0
CL66	2.0	Vaucl-Gard	4	968	0
CL67	2.0	Chare-Réuni	4	972	0
CL68	3.0	CL75-Vende	4	976	0
CL69	2.0	Aisne-Yonne	4	979	0
CL70	2.0	Marne-Arden	4	983	0
CL71	3.0	CL76-Tarn	4	986	0
CL72	2.0	Corre-Eure	3	990	0
CL73	2.0	Loire-Meurt	3	993	0

La représentation de la perte d'inertie inter-classe en histogramme permet de disposer rapidement

d'éléments permettant le choix d'un nombre de classes optimal. On remarquera ici la perte d'inertie plus importante entre 10 et 11 classes qui indique une bonne partition.

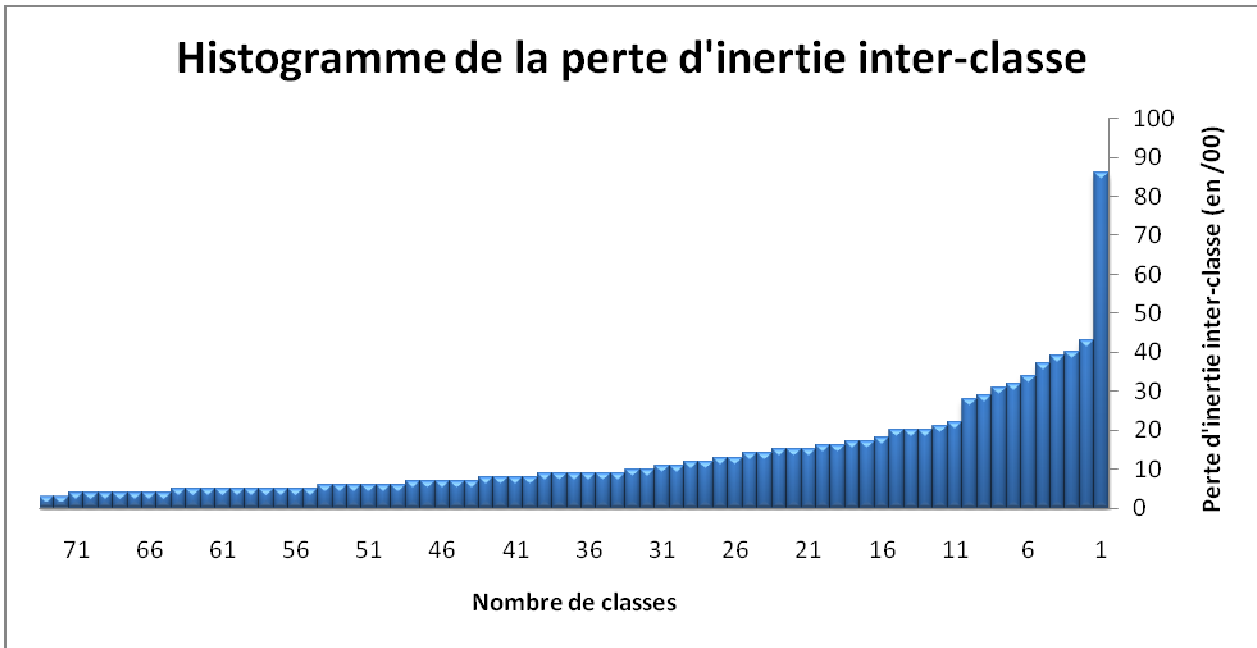


Figure 5 : Histogramme de la perte d'inertie inter-classe en fonction du nombre de classe

3.3. Cubic Clustering Criterion

Le Cubic Clustering Criterion (CCC) est un indicateur d'aide au choix du nombre de classe optimal pour une classification ascendante hiérarchique. Un CCC >2 indique une bonne classification, un CCC compris entre 0 et 2 indique une classification a priori correcte, un CCC <0 indique la présence d'outliers gênants (surtout si CCC <-30). Dans la pratique, on jugera qu'une classification en k classes est bonne si le graphe du CCC montre une pente importante entre k-1 et k, puis une pente plus modérée entre k et k+1. Ici, l'écart entre 10 et 11 classes (en rouge) est important, puis l'écart entre 11 et 12 classes est plus faible.

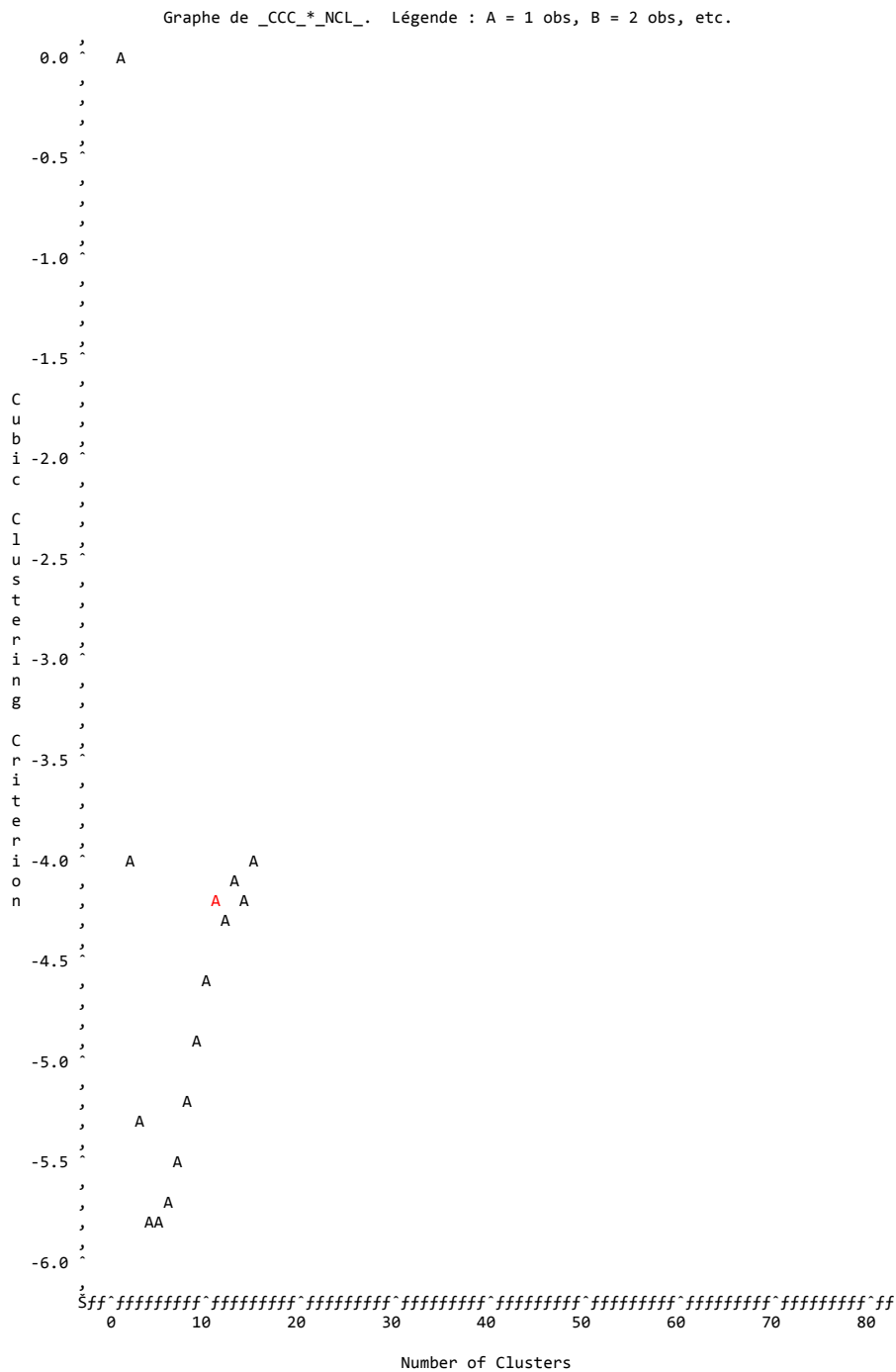


Figure 6 : Courbe du Cubic Clustering Criterion en fonction du nombre de classe

4. Partition en 11 classes

La classification en 11 classes est celle qui mathématiquement la plus justifiée.

4.1. Consolidation des classes

La consolidation des classes est une étape qui permet d'améliorer la qualité et la stabilité des classes constituées en augmentant l'inertie inter-classe (hétérogénéité entre les classes) et en diminuant l'inertie intra-classe (l'hétérogénéité au sein des classes). Ce processus consiste en un recentrage des classes : il s'agit de réaffecter de façon itérative les individus au centre de classe le plus proche.

Tableau 8 : Impact de la consolidation des classes sur l'inertie inter-classe

Résultats de la consolidation de la partition en 11 classes	
Inertie interclasse	avant la consolidation : 16.765
	après la consolidation : 17.028
	soit un gain de : 1.6 %

4.2. Composition des classes

Les classes 1 et 2 ont les effectifs les plus importants, avec respectivement 15 et 22 structures de gestion.

Tableau 9 : Détail de la composition des 11 classes

Classe numéro 1
Ardennes, Aube, Aude (ADOC11), Charente, Corrèze, Eure et Loir (ADOC 28), Gard et Lozère, Haute-Marne, Hautes-Alpes, Lot, Marne, Somme, Tarn, Vaucluse, Vosges
Classe numéro 2
Aisne, Aveyron, Cher (IRSA ADOC18), Cote d'Or + Nièvre (ADECA 21), Cotes d'Armor (ADEC), Deux-Sèvres, Doubs, Jura, Haute Saône et Territoire de Belfort, Essonne, Finistère (ADEC 29), Haute-Vienne, Indre et Loire, Loir et Cher, Maine et Loire, Mayenne, Morbihan (ADECAM), Moselle, Oise, Pyrénées-Orientales, Saône et Loire (ADEMAS 71), Sarthe, Vendée, Yonne (AIDEC)
Classe numéro, 3
Allier (ABIDEC), Ariège, Creuse, Indre, Meuse
Classe numéro 4
Ardèche et Drome, Calvados (Association MATHILDE), Gers, Haute-Savoie, Hautes-Pyrénées, Manche (IRIS MANCHE), Savoie, Tarn et Garonne
Classe numéro 5
Cantal, Haute-Loire, Puy de Dôme (ABIDOC), Charente-Maritime, Hérault, Ille et Vilaine (ADEC 135), Loire, Meurthe et Moselle, Réunion, Seine et Marne, Val de Marne
Classe numéro 6
Dordogne (vitalis), Haute-Garonne, Lot et Garonne (Structure de gestion du dépistage du Lot et Garonne)
Classe numéro 7

Alpes de Haute-Provence et Alpes Maritime, Gironde (AGIDECA 33), Hauts-de-Seine, Loire-Atlantique, Rhône, Val d'Oise, Var, Yvelines
Classe numéro 8
Ain, Pas-de-Calais, Seine-Saint-Denis
Classe numéro 9
Isère, Orne (Régie départementale des cancers)
Classe numéro 10
Pyrénées-Atlantiques (PYRADEC PAU)
Classe numéro 11
Paris

La représentation des classes par département. En blanc apparaissent les structures de gestion n'ayant pas participé à la CAH.

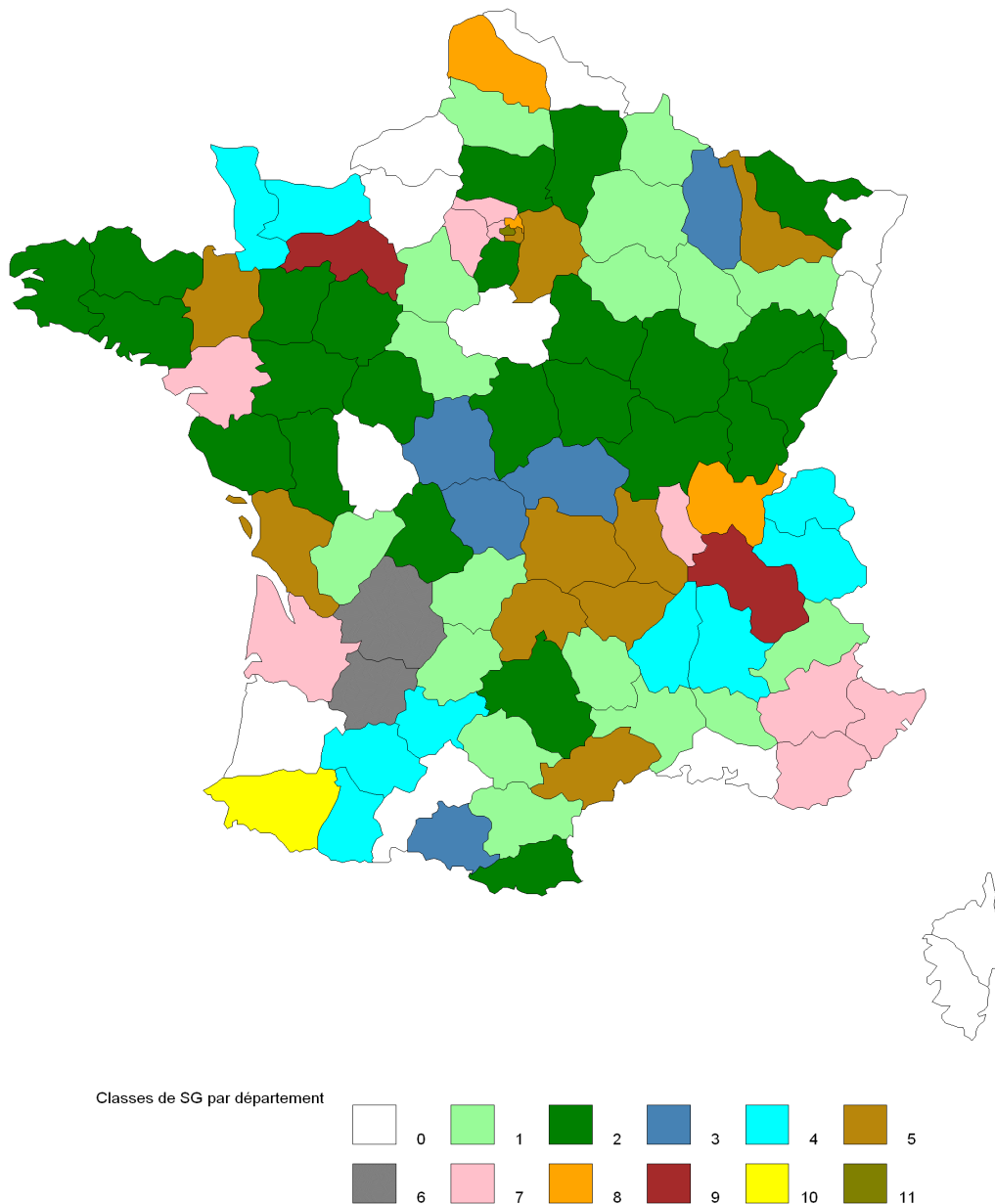


Figure 7 : Représentation géographique de la classification en 11 classes

4.3. Décomposition de l'inertie totale

L'inertie inter-classe correspond à l'hétérogénéité entre les classes. Plus elle est élevée, plus les classes diffèrent.

A contrario, l'inertie intra-classe est un indicateur de l'homogénéité au sein des classes. Plus elle est faible, plus les individus qui composent la classe sont semblables.

Selon la formule de Huygens : Inertie totale = Inertie inter + Inertie intra. L'inertie totale étant constante, la recherche d'une inertie inter-classe élevée implique une inertie intra-classe faible, et inversement.

Tableau 10 : Décomposition par classe de l'inertie inter-classe et intra-classe

_	Classes	Inertie	En 0/00 de l'inertie totale	Effectif absolu	Effectif pondéré (en 0/00)
Inertie totale		42.000	1000	77	1000
Inertie inter		17.028	405	.	.
	1	1.3527	32	15	195
	2	1.0525	25	22	286
	3	1.8239	43	5	65
	4	1.1483	27	8	104
	5	1.0103	24	9	117
	6	1.4162	34	3	39
	7	2.2346	53	8	104
	8	1.4020	33	3	39
	9	1.6006	38	2	26
	10	1.4720	35	1	13
	11	2.5150	60	1	13
Inertie intra		24.972	595	.	.
	1	4.0759	97	15	195
	2	7.7634	185	22	286
	3	1.5671	37	5	65
	4	3.1946	76	8	104
	5	2.7158	65	9	117
	6	1.1226	27	3	39
	7	3.3066	79	8	104
	8	0.6025	14	3	39
	9	0.6235	15	2	26
	10	0.0000	0	1	13
	11	0.0000	0	1	13

4.4. *Caractérisation des classes par les variables : Moyennes, écarts-types et test de significativité.*

Le tableau suivant permet, pour chaque classe, d'identifier les variables les plus significatives, c'est-à-dire les variables les plus importantes dans la constitution de la classe.

Les variables significatives pour une classe sont celles dont la valeur du paramètre « V-Test » est supérieure à 2 et dont la valeur du paramètre « Proba » est inférieure à 0,05.

Un bon exemple est la variable v9 (Autre dépistage pratiqué) qui est significative pour la classe 9 (dernières lignes, valeurs en rouge) : en effet, en parcourant la base de données, on constate que les deux structures de gestion qui composent la classe 9 (Orne et Isère) pratiquent un autre dépistage.

Une bonne caractérisation d'une classe est obtenue en comparant, pour une variable donnée, la moyenne de la classe et la moyenne sur l'ensemble des structures de gestion.

Tableau 11 : Moyennes, écarts-types et test de significativité par variables pour chacune des 11 classes et pour l'ensemble des structures de gestion

Variable	Statistique	Ensemble	Classe 1	Classe 2	Classe 3	Classe 4	Classe 5	Classe 6	Classe 7	Classe 8	Classe 9	Classe 10	Classe 11
	Eff.pond	77	15	22	5	8	9	3	8	3	2	1	1
v2 <i>Procédure amélioration fichier</i>	Moyenne	0.96	1	1	0.8	1	1	0.67	1	1	1	0	1
	Ec.type	0.19	0	0	0.4	0	0	0.47	0	0	0	0	0
	V.test	6.37	0.86	1.11	-1.91	0.6	0.64	-2.67	0.6	0.35	0.29	-4.97	0.2
	Proba	0	0.39	0.27	0.06	0.55	0.52	0.01	0.55	0.72	0.77	0	0.84
v3 <i>Délai de transmission</i>	Moyenne	2.7	1.93	3.23	4.4	1.88	1.56	3.33	3	3.33	2.5	4	4
	Ec.type	2.12	1.95	2	2.24	1.76	1.64	2.36	2	2.05	2.5	0	0
	V.test	1.2	-1.55	1.37	1.84	-1.16	-1.71	0.52	0.42	0.52	-0.14	0.61	0.61
	Proba	0.31	0.12	0.17	0.07	0.25	0.09	0.6	0.68	0.6	0.89	0.54	0.54
v4 <i>Densité médecins généralistes</i>	Moyenne	1.27	1.3	1.25	1.14	1.05	1.51	1.26	1.29	1.13	1.08	1.58	2.22
	Ec.type	0.34	0.28	0.2	0.15	0.29	0.61	0.08	0.25	0.21	0.21	0	0
	V.test	2.03	0.3	-0.37	-0.88	-1.96	2.16	-0.09	0.14	-0.73	-0.8	0.89	2.75
	Proba	0.04	0.76	0.71	0.38	0.05	0.03	0.93	0.89	0.46	0.42	0.37	0.01
v5 <i>Densité radiologues</i>	Moyenne	0.1	0.1	0.09	0.06	0.09	0.13	0.11	0.13	0.08	0.1	0.12	0.32
	Ec.type	0.04	0.02	0.02	0.03	0.02	0.04	0.03	0.02	0.02	0.02	0	0
	V.test	11.94	-0.44	-1.82	-2.35	-1.23	2.52	0.32	2.03	-1	-0.1	0.43	5.45
	Proba	0	0.66	0.07	0.02	0.22	0.01	0.75	0.04	0.32	0.92	0.66	0
v6 <i>Densité gynécologues</i>	Moyenne	0.06	0.05	0.05	0.03	0.05	0.06	0.06	0.09	0.04	0.05	0.08	0.23
	Ec.type	0.03	0.02	0.01	0.01	0.01	0.02	0.02	0.03	0.02	0.02	0	0
	V.test	12.35	-1.01	-1.56	-1.83	-0.44	0.57	0.06	3.04	-0.73	-0.27	0.82	5.85
	Proba	0	0.31	0.12	0.07	0.66	0.57	0.96	0	0.47	0.78	0.41	0
v7 <i>Densité HGE</i>	Moyenne	0.05	0.05	0.04	0.03	0.04	0.06	0.04	0.06	0.04	0.05	0.06	0.12
	Ec.type	0.02	0.01	0.01	0.01	0.01	0.02	0.02	0.01	0.01	0.01	0	0
	V.test	7.15	0.47	-1.73	-2.12	-1.73	2.14	-0.75	2.18	-0.75	-0.17	0.81	4.53
	Proba	0	0.64	0.08	0.03	0.08	0.03	0.45	0.03	0.45	0.87	0.42	0
v8 <i>Statut juridique</i>	Moyenne	1.13	1	1	1.2	1.38	1	1	1	1.33	2	4	1
	Ec.type	0.47	0	0	0.4	0.48	0	0	0	0.47	1	0	0

Variable	Statistique	Ensemble	Classe 1	Classe 2	Classe 3	Classe 4	Classe 5	Classe 6	Classe 7	Classe 8	Classe 9	Classe 10	Classe 11
	V.test	14.04	-1.2	-1.54	0.35	1.56	-0.88	-0.49	-0.83	0.77	2.66	6.16	-0.28
	Proba	0	0.23	0.12	0.73	0.12	0.38	0.62	0.41	0.44	0.01	0	0.78
v9	Moyenne	0.03	0	0	0	0	0	0	0	0	1	0	0
Autre dépistage pratique	Ec.type	0.16	0	0	0	0	0	0	0	0	0	0	0
	V.test	.	-0.7	-0.9	-0.38	-0.48	-0.52	-0.29	-0.48	-0.29	8.72	-0.16	-0.16
	Proba	.	0.48	0.37	0.71	0.63	0.6	0.77	0.63	0.77	0	0.87	0.87
v10	Moyenne	9.63	9.38	8.82	9.06	7.24	10.71	6.32	14.54	8.92	15.74	6.83	6.75
Ancienneté CS	Ec.type	4.76	5.34	3.83	3.58	3.12	4.09	0.25	5.67	3.17	3.43	0	0
	V.test	1.91	-0.22	-0.94	-0.28	-1.49	0.72	-1.22	3.06	-0.26	1.83	-0.59	-0.6
	Proba	0.06	0.82	0.35	0.78	0.14	0.47	0.22	0	0.79	0.07	0.56	0.55
v11	Moyenne	2.97	2.43	3.69	2.35	2.7	3.44	1.39	2.09	2.77	6.58	2	2.33
Ancienneté CCR	Ec.type	1.86	1.64	1.95	1.41	1.58	2.08	0.16	0.9	0.95	1	0	0
	V.test	1.98	-1.24	2.13	-0.77	-0.43	0.81	-1.49	-1.4	-0.19	2.77	-0.52	-0.34
	Proba	0.05	0.21	0.03	0.44	0.67	0.42	0.14	0.16	0.85	0.01	0.6	0.73
v12	Moyenne	0.96	1	1	1	1	1	1	1	0	1	1	1
Retours radiologues	Ec.type	0.19	0	0	0	0	0	0	0	0	0	0	0
	V.test	.	0.86	1.11	0.46	0.6	0.64	0.35	0.6	-8.72	0.29	0.2	0.2
	Proba	.	0.39	0.27	0.64	0.55	0.52	0.72	0.55	0	0.77	0.84	0.84
v13	Moyenne	2.03	1.6	2.02	1	1.94	2.44	1.67	3.19	1.67	2	2	4
Niveau de vie département	Ec.type	0.97	0.71	0.68	0	1.01	1.07	0.94	0.7	0.94	1	0	0
	V.test	3.48	-1.91	-0.06	-2.44	-0.29	1.34	-0.66	3.53	-0.66	-0.05	-0.03	2.03
	Proba	0	0.06	0.96	0.01	0.77	0.18	0.51	0	0.51	0.96	0.97	0.04
v14	Moyenne	7.6	5.58	7.95	3.91	6.44	9.21	5.83	10.77	11.11	10.75	5.75	8.71
Nb ETP	Ec.type	2.87	1.37	2.23	1	2.32	1.53	2.5	2.54	3.66	0.55	0	0
	V.test	6.58	-3.02	0.67	-2.95	-1.2	1.78	-1.09	3.27	2.14	1.56	-0.65	0.38
	Proba	0	0	0.5	0	0.23	0.08	0.28	0	0.03	0.12	0.52	0.7
v15	Moyenne	0.17	0.13	0.09	0	0.5	0	0	0.38	0.33	0.5	0	0

Variable	Statistique	Ensemble	Classe 1	Classe 2	Classe 3	Classe 4	Classe 5	Classe 6	Classe 7	Classe 8	Classe 9	Classe 10	Classe 11
<i>Personnel à disposition</i>	Ec.type	0.37	0.34	0.29	0	0.5	0	0	0.48	0.47	0.5	0	0
	V.test	1.7	-0.41	-1.15	-1.04	2.62	-1.43	-0.79	1.63	0.77	1.26	-0.45	-0.45
	Proba	0.1	0.68	0.25	0.3	0.01	0.15	0.43	0.1	0.44	0.21	0.65	0.65
v16	Moyenne	0.57	0.13	0.64	0.6	1	0.67	0.33	0.75	0.67	0.5	0	1
<i>Personne dédiée aux tâches administratives</i>	Ec.type	0.49	0.34	0.48	0.49	0	0.47	0.47	0.43	0.47	0.5	0	0
	V.test	2.72	-3.8	0.72	0.13	2.57	0.61	-0.84	1.07	0.34	-0.21	-1.15	0.87
	Proba	0.01	0	0.47	0.89	0.01	0.54	0.4	0.28	0.74	0.84	0.25	0.39
v17	Moyenne	163332.18	168682.58	149756.45	126699.86	139095.73	179506.78	193856.97	206874.9	170086.3	126893.46	179944.87	209310.56
<i>Budget global/Nb ETPD</i>	Ec.type	42744.43	31828.27	40411.87	14539.05	25197.9	19919.93	46801.84	59087.14	8632.05	43219.59	0	0
	V.test	2.75	0.54	-1.75	-1.97	-1.68	1.2	1.25	3.02	0.28	-1.21	0.39	1.08
	Proba	0.01	0.59	0.08	0.05	0.09	0.23	0.21	0	0.78	0.22	0.7	0.28
v18	Moyenne	18.21	20.61	17.3	21.27	17.83	18.09	16.89	16.17	17.89	17	14.62	18.12
<i>Coût par personnes dépistées CS</i>	Ec.type	3.69	4.19	3.14	2.58	1.96	2.59	5.65	3.37	2.92	3.28	0	0
	V.test	1.63	2.8	-1.35	1.91	-0.31	-0.1	-0.63	-1.64	-0.15	-0.47	-0.97	-0.02
	Proba	0.12	0.01	0.18	0.06	0.76	0.92	0.53	0.1	0.88	0.64	0.33	0.98
v19	Moyenne	0.54	0.6	0.59	0.54	0.55	0.45	0.57	0.46	0.47	0.53	0.67	0.24
<i>Pourcentage de participation CS</i>	Ec.type	0.13	0.08	0.14	0.04	0.03	0.05	0.19	0.15	0.07	0.03	0	0
	V.test	2.71	1.81	2.07	-0.04	0.15	-2.35	0.38	-2.05	-1	-0.2	1.01	-2.38
	Proba	0.01	0.07	0.04	0.97	0.88	0.02	0.7	0.04	0.32	0.85	0.31	0.02
v20	Moyenne	6328.62	4817.91	5369.66	4280.55	5414.25	7980.63	7948.62	9958.31	6877.17	4460.61	8681.57	18613.32
<i>Population invitée CS/Nb ETP</i>	Ec.type	2799.78	1321.31	1616.11	801.7	1424.93	1618.96	1375.81	2984.36	942.88	2549.98	0	0
	V.test	11.57	-2.31	-1.89	-1.68	-0.97	1.87	1.02	3.85	0.34	-0.95	0.84	4.39
	Proba	0	0.02	0.06	0.09	0.33	0.06	0.31	0	0.73	0.34	0.4	0
v22	Moyenne	0.09	0.07	0.05	0	0.13	0	0.33	0.38	0	0	0	0
<i>Mise à disposition des locaux</i>	Ec.type	0.29	0.25	0.21	0	0.33	0	0.47	0.48	0	0	0	0
	V.test	1.32	-0.36	-0.87	-0.73	0.35	-1	1.48	2.93	-0.56	-0.45	-0.32	-0.32
	Proba	0.24	0.72	0.38	0.47	0.72	0.32	0.14	0	0.58	0.65	0.75	0.75

Variable	Statistique	Ensemble	Classe 1	Classe 2	Classe 3	Classe 4	Classe 5	Classe 6	Classe 7	Classe 8	Classe 9	Classe 10	Classe 11
v23	Moyenne	0.94	0.97	0.92	1.07	0.95	0.98	0.94	0.89	0.88	0.83	0.9	0.93
<i>Budget réalisé/Budget prévisionnel</i>	Ec.type	0.12	0.17	0.1	0.13	0.11	0.07	0.06	0.11	0.03	0.14	0	0
	V.test	1.1	0.86	-0.87	2.34	0.14	0.84	-0.09	-1.3	-0.83	-1.29	-0.37	-0.14
	Proba	0.38	0.39	0.39	0.02	0.89	0.4	0.93	0.19	0.4	0.2	0.71	0.89
v24	Moyenne	91789.08	53763.33	81587.09	32280.2	64789.88	117663.12	90437	181445.01	143925	95972.5	98251	282863
<i>Population cible CS</i>	Ec.type	54712.03	22532.86	31878.63	13105.88	29686.28	24784.16	44243.54	24753.94	51235.29	54027.5	0	0
	V.test	15.8	-2.98	-1.03	-2.5	-1.46	1.5	-0.04	4.86	1.67	0.11	0.12	3.49
	Proba	0	0	0.3	0.01	0.14	0.13	0.97	0	0.09	0.91	0.91	0
v26	Moyenne	156.97	117.45	23.29	65.42	77.88	33.74	2581.07	80.13	25.3	17.83	35.82	27.45
<i>Coût par personne dépistée CCR</i>	Ec.type	512.57	180.3	24.53	51.92	110.42	19.63	598.83	109.19	6.25	2.35	0	0
	V.test	68.13	-0.33	-1.44	-0.41	-0.46	-0.76	8.3	-0.45	-0.45	-0.39	-0.24	-0.25
	Proba	0	0.74	0.15	0.68	0.65	0.45	0	0.66	0.65	0.7	0.81	0.8
v27	Moyenne	178240.22	104686.53	158033.68	62403.6	125246.88	237842.36	174748.67	347716.63	279471.33	185651.5	187474	519714
<i>Population cible CCR</i>	Ec.type	105354.71	44043.85	61847.99	24614.22	57152.28	62724.88	84448.02	46182.53	95905.3	104348.5	0	0
	V.test	14.63	-2.99	-1.06	-2.53	-1.49	1.79	-0.06	4.78	1.69	0.1	0.09	3.24
	Proba	0	0	0.29	0.01	0.14	0.07	0.95	0	0.09	0.92	0.93	0
v28	Moyenne	0.24	0.17	0.36	0.25	0.23	0.21	0.04	0.12	0.34	0.43	0.18	0.08
<i>Pourcentage de participation CCR</i>	Ec.type	0.16	0.13	0.14	0.13	0.1	0.12	0.05	0.08	0.14	0.03	0	0
	V.test	4.66	-1.92	4.22	0.02	-0.29	-0.62	-2.27	-2.42	1.11	1.65	-0.41	-1.06
	Proba	0	0.05	0	0.98	0.78	0.54	0.02	0.02	0.26	0.1	0.68	0.29
v29	Moyenne	9719.45	7485.18	10190.72	5197.16	7232.87	13142.02	1379.51	14138.13	9480.11	8583.1	6121.57	40825.26
<i>Population invitée CCR/Nb ETP</i>	Ec.type	6411.23	3445.63	4115.87	2235.7	4246.92	5544.87	935.03	7311.33	2010.34	4516.31	0	0
	V.test	7.51	-1.49	0.41	-1.62	-1.15	1.69	-2.28	2.05	-0.07	-0.25	-0.56	4.85
	Proba	0	0.14	0.69	0.11	0.25	0.09	0.02	0.04	0.95	0.8	0.57	0
v30	Moyenne	10.68	2.77	15.06	6.4	13.16	4.07	2.57	20.47	5	17.47	14	40.32
<i>Part financement Conseil Général</i>	Ec.type	12.73	5.98	13.56	12.8	12.57	7.75	2.04	12.07	7.07	0.47	0	0
	V.test	3.05	-2.67	1.9	-0.77	0.58	-1.65	-1.12	2.28	-0.78	0.76	0.26	2.33

Variable	Statistique	Ensemble	Classe 1	Classe 2	Classe 3	Classe 4	Classe 5	Classe 6	Classe 7	Classe 8	Classe 9	Classe 10	Classe 11
	Proba	0	0.01	0.06	0.44	0.56	0.1	0.26	0.02	0.43	0.45	0.79	0.02
v31	Moyenne	0.88	1	0.86	0	1	1	1	1	1	1	0	1
Mise en place partenariat	Ec.type	0.32	0	0.34	0	0	0	0	0	0	0	0	0
	V.test	13.65	1.56	-0.33	-6.31	1.08	1.15	0.64	1.08	0.64	0.52	-2.75	0.36
	Proba	0	0.12	0.74	0	0.28	0.25	0.52	0.28	0.52	0.6	0.01	0.72
v32	Moyenne	0.88	0.93	0.73	0.6	1	1	1	1	1	1	1	1
Externalisation des tâches	Ec.type	0.32	0.25	0.45	0.49	0	0	0	0	0	0	0	0
	V.test	1.47	0.67	-2.67	-2.02	1.08	1.15	0.64	1.08	0.64	0.52	0.36	0.36
	Proba	0.17	0.5	0.01	0.04	0.28	0.25	0.52	0.28	0.52	0.6	0.72	0.72
v33	Moyenne	0.66	0.67	0.59	1	0.75	0.78	1	0.38	0.33	0.5	1	1
Partage de moyen avec une autre SG	Ec.type	0.47	0.47	0.49	0	0.43	0.42	0	0.48	0.47	0.5	0	0
	V.test	1.11	0.04	-0.83	1.64	0.55	0.77	1.25	-1.8	-1.22	-0.49	0.71	0.71
	Proba	0.37	0.97	0.4	0.1	0.58	0.44	0.21	0.07	0.22	0.63	0.48	0.48
v34	Moyenne	0.74	0.53	0.82	0.6	0.38	1	0.67	0.88	1	1	1	1
Retour généralistes	Ec.type	0.44	0.5	0.39	0.49	0.48	0	0.47	0.33	0	0	0	0
	V.test	1.81	-2.02	0.98	-0.73	-2.47	1.88	-0.29	0.91	1.04	0.84	0.59	0.59
	Proba	0.08	0.04	0.33	0.46	0.01	0.06	0.77	0.36	0.3	0.4	0.55	0.55
v35	Moyenne	0.95	1	1	1	0.63	1	1	1	1	1	0	1
Participation médecins coordinateurs séminaires	Ec.type	0.22	0	0	0	0.48	0	0	0	0	0	0	0
	V.test	6.75	1	1.29	0.54	-4.32	0.74	0.41	0.69	0.41	0.33	-4.27	0.23
	Proba	0	0.32	0.2	0.59	0	0.46	0.68	0.49	0.68	0.74	0	0.81
v36	Moyenne	0.09	0.11	0.07	0.13	0.08	0.1	0.05	0.08	0.07	0.06	0.1	0.06
Part des coûts structurels dans les coûts	Ec.type	0.04	0.04	0.03	0.05	0.02	0.02	0.04	0.01	0.01	0.01	0	0
	V.test	3.36	2.44	-2.04	3.2	-0.87	1.13	-1.94	-0.62	-0.77	-1.23	0.42	-0.61
	Proba	0	0.01	0.04	0	0.38	0.26	0.05	0.54	0.44	0.22	0.67	0.54
v37	Moyenne	1.75	2	1.73	1.2	1.75	1.22	2	2	2	2	2	2
Mise en place outils	Ec.type	0.65	0	0.69	0.98	0.66	0.92	0	0	0	0	0	0

Variable	Statistique	Ensemble	Classe 1	Classe 2	Classe 3	Classe 4	Classe 5	Classe 6	Classe 7	Classe 8	Classe 9	Classe 10	Classe 11
	V.test	1.53	1.63	-0.22	-1.96	-0.01	-2.6	0.67	1.13	0.67	0.54	0.38	0.38
	Proba	0.15	0.1	0.83	0.05	0.99	0.01	0.5	0.26	0.5	0.59	0.7	0.7
v38	Moyenne	0.84	0.87	0.77	1	0.88	0.89	1	0.88	1	0.5	1	0
Formation personnel SG	Ec.type	0.36	0.34	0.42	0	0.33	0.31	0	0.33	0	0.5	0	0
	V.test	1.05	0.27	-1.09	0.99	0.25	0.39	0.75	0.25	0.75	-1.35	0.43	-2.33
	Proba	0.41	0.79	0.28	0.32	0.8	0.7	0.45	0.8	0.45	0.18	0.67	0.02
v39	Moyenne	0.58	0.93	0.86	0.2	0	0.11	0.67	0.88	0	0	1	0
Existence convention collective	Ec.type	0.49	0.25	0.34	0.4	0	0.31	0.47	0.33	0	0	0	0
	V.test	11.67	3.04	3.12	-1.79	-3.52	-3.05	0.29	1.75	-2.08	-1.69	0.84	-1.19
	Proba	0	0	0	0.07	0	0	0.77	0.08	0.04	0.09	0.4	0.24
v40	Moyenne	0.81	1	0.95	0.2	0.75	0.78	0.67	1	0	0.5	1	0
Réunions avec l'équipe de la SG	Ec.type	0.6	0.37	0.37	0.4	0.97	0.79	0.47	0.5	0	0.5	0	0
	V.test	1.83	1.38	1.36	-2.3	-0.27	-0.14	-0.4	0.96	-2.34	-0.72	0.32	-1.33
	Proba	0.07	0.17	0.17	0.02	0.79	0.89	0.69	0.34	0.02	0.47	0.75	0.18
v41	Moyenne	4.59	5.71	3.85	4.29	4.48	5.43	5.15	4.29	4.3	4.29	0	6
Mise en place procédures internes suivi de l'assurance qualité	Ec.type	1.84	0.75	2.19	1.53	1.36	1.14	0.69	2.14	0.7	1.71	0	0
	V.test	2.1	2.6	-2.23	-0.38	-0.19	1.44	0.53	-0.49	-0.28	-0.24	-2.5	0.76
	Proba	0.04	0.01	0.03	0.71	0.85	0.15	0.59	0.62	0.78	0.81	0.01	0.44
v42	Moyenne	14.99	12.86	15.61	10.49	22.94	11.75	14.81	16.38	14.7	12.4	17	14.6
Délai moyen rendu des résultats	Ec.type	6.19	2.85	5.53	0.98	11.13	3.59	1.06	4.59	2.81	4.4	0	0
	V.test	2.5	-1.47	0.55	-1.67	3.81	-1.66	-0.05	0.67	-0.08	-0.59	0.33	-0.06
	Proba	0.01	0.14	0.58	0.1	0	0.1	0.96	0.51	0.93	0.55	0.74	0.95
v43	Moyenne	2.27	1.73	2.91	2.2	1.75	1.33	3.33	2.13	2	5.5	1	3
Organisation information professionnels dépistage	Ec.type	1.65	1.06	2.07	0.98	1.09	1.05	0.47	1.45	1.41	0.5	0	0
	V.test	2.08	-1.4	2.13	-0.1	-0.94	-1.81	1.13	-0.27	-0.29	2.79	-0.77	0.44
	Proba	0.04	0.16	0.03	0.92	0.35	0.07	0.26	0.79	0.77	0.01	0.44	0.66
v44	Moyenne	2.14	2.73	1.36	1	2.13	2.44	1.33	3.25	2.33	2	6	3

Variable	Statistique	Ensemble	Classe 1	Classe 2	Classe 3	Classe 4	Classe 5	Classe 6	Classe 7	Classe 8	Classe 9	Classe 10	Classe 11
<i>Mise en place initiatives locales</i>	Ec.type	1.5	1.44	0.93	0.63	1.62	1.57	1.25	1.56	0.47	0	0	0
	V.test	2.94	1.69	-2.86	-1.75	-0.04	0.64	-0.95	2.19	0.22	-0.14	2.57	0.57
	Proba	0	0.09	0	0.08	0.97	0.52	0.34	0.03	0.82	0.89	0.01	0.57
v45	Moyenne	0.07	0.05	0.05	0.04	0.09	0.08	0.03	0.17	0.05	0.14	0.06	0.28
<i>Nb perdus de vue CS/Population dépistée pos.</i>	Ec.type	0.09	0.06	0.06	0.03	0.06	0.05	0.03	0.19	0.04	0.07	0	0
	V.test	2.26	-1.1	-1.49	-0.89	0.41	0.34	-0.91	3.17	-0.46	1	-0.16	2.21
	Proba	0.02	0.27	0.14	0.38	0.68	0.74	0.37	0	0.64	0.32	0.88	0.03

4.5. **Caractéristiques des classes**

La **classe 1** est composée de 15 structures de gestion. Elle se caractérise principalement par l'absence d'une personne dédiée aux tâches administratives (v16 : une personne dédiée est présente dans seulement 13% des structures de la classe 1 contre 57% sur l'ensemble des structures). La classe 1 se démarque également par des populations cibles pour le dépistage des cancers du sein et colorectaux inférieures à la moyenne (v24 et v27 : respectivement 53763 vs 91789 et 104686 vs 178240), par un nombre d'ETP plus faible que sur l'ensemble des structures (v14 : 5,58 vs 7,6) et par des coûts de dépistage du cancer du sein plus importants que la moyenne (v18 : 20,61€ vs 18,21€). La quasi-totalité des structures de gestion de la classe 1 possèdent une convention collective (v39 : 93% vs 58% sur l'ensemble des SG).

La **classe 2** comprend 22 structures de gestion. Elle se démarque surtout par un pourcentage de participation au dépistage du cancer colorectal plus important que la moyenne (v28 : 36% vs 24%) et par une part de structures dotées d'une convention collective plus importante que sur l'ensemble des structures (v39 : 86% vs 58%). Elle se distingue également par un recours moins fréquent à l'externalisation des tâches (v32 : 73% vs 88% sur l'ensemble des structures) et par une quantité moins importante d'initiatives locales spécifiques (v44 : 1,3 vs 2,1 en moyenne).

La **classe 3** est caractérisée par l'absence totale de partenariat au sein des structures de gestion qui la compose (v31 : 0% vs 88% sur l'ensemble des SG). On notera aussi que la part des coûts structurels (hors loyer) dans les coûts est plus élevée dans cette classe en comparaison avec la moyenne des structures (v36 : 13% vs 9%). La classe 3 compte 5 structures de gestion.

La **classe 4** est composée de 8 structures de gestion. Elle se distingue de l'ensemble des structures principalement par une absence totale de convention collective (v39 : 0% vs 58%), par des délais moyens de rendus des résultats plus importants (v42 : 23 jours vs 15 jours) et par une participation moins importante des médecins aux séminaires organisés (v35 : 63% vs 95% en moyenne). Le retour aux médecins généralistes pratiqué par les structures de la classe 4 est inférieur à la moyenne (v34 : 38% des SG le pratiquent contre 74% sur l'ensemble des SG). La classe 4 se démarque également du point de vue du personnel : la part des SG qui ont profité de personnel mis à disposition en 2008 est plus importante que sur l'ensemble des SG (v15 : 50% vs 17%), de même que pour la présence d'une personne dédiée aux tâches administratives (v16 : 100% vs 57%).

La **classe 5** se caractérise principalement par un nombre de structures avec convention collective moins important que dans l'ensemble (v39 : 11% vs 58%). Elle se caractérise également par un taux de participation au dépistage du cancer du sein inférieur de presque 10 points par rapport à la moyenne (v19 : 45% vs 54%). La classe 5 est composée de 9 structures de gestion.

La **classe 6** se démarque par un coût par personne dépistée pour le cancer colorectal très supérieur à celui que l'on retrouve sur l'ensemble des structures de gestion (v26 : 2581€ vs 157€). Les structures de la classe 6 ont également moins recours au fichier de la poste pour améliorer leur fichier (v2 : 67% vs 96% en moyenne).

La **classe 7** comprend 8 structures de gestion. Elle se caractérise par des populations cibles pour le cancer du sein et le cancer colorectal très supérieur à la moyenne (v24 et v27 : respectivement 181445 vs 91789 pour le cancer du sein et 347716 vs 178240 pour le cancer colorectal). Le rapport population invitée / Nombre d'ETP observé dans la classe 7 pour le cancer du sein est également plus important que pour l'ensemble des structures (v20 : 9958/ETP vs 6328/ETP). Le niveau de vie des départements dans lesquels sont situées les structures de la classe 7 est plus élevé que celui de l'ensemble des SG (v13 : 15766€ - 17350€ pour la classe 7 vs 14820€ - 15766€ pour l'ensemble).

La **classe 8** est composée de 3 structures et se distingue par l'absence de retour aux radiologues (v12).

La **classe 9** est caractérisée par la pratique d'un autre dépistage (v9 : 100% vs 3% en moyenne). Elle se distingue également par une ancienneté dans le dépistage du cancer colorectal plus importante (v11 : 6 ans vs 3 ans en moyenne) et par l'organisation de nombreux événements d'information sur le dépistage organisé à l'intention des professionnels de santé (v43). Deux structures de gestion forment la classe 9.

La **classe 10** n'est composée que d'une seule structure (Pyrénées-Atlantiques). Elle se démarque par l'absence de recours au fichier de la poste pour améliorer son propre fichier (v2 : 0% vs 96% sur l'ensemble des structures), par son absence de partenariat avec les autres structures (v31 : 0% vs 88%) et sur la non participation des médecins coordinateurs aux séminaires organisés (v35 : 0% vs 95%). Il s'agit également de la seule classe où le statut juridique est « œuvre de la CPAM » (contre « Association loi 1901 » dans la plupart des SG, v8).

A l'instar de la classe 10, la **classe 11** ne comprend qu'une seule structure de gestion (Paris). Cette classe se distingue par une densité de médecins supérieur à celle observée sur l'ensemble des SG (v5, v6

et v7 : 0,32 vs 0,1 pour les radiologues, 0,23 vs 0,06 pour les gynécologues et 0,12 vs 0,05 pour les HGE). Elle se caractérise également par un rapport population cible invitée / Nombre d'ETP supérieur à la moyenne que ce soit pour le cancer du sein ou le cancer colorectal (v20 et v29 : 18613/ETP vs 6328/ETP pour le cancer du sein et 40825/ETP vs 9719/ETP pour le cancer colorectal). La classe 11 possède une population cible pour le cancer du sein deux fois supérieure à celle observée sur l'ensemble des structures (v24 : 282863 vs 91789).

5. Partition en 7 classes

La classification en 7 ensembles est moins rigoureuse sur le plan mathématique, mais elle pourrait éventuellement être jugée plus opérationnelle pour la conduite des entretiens à venir c'est pourquoi cette variante a été étudiée dans le cadre d'une analyse de sensibilité.

5.1. Consolidation des classes

Tableau 12 : Impact sur l'inertie inter-classe de la consolidation des classes

Résultats de la consolidation de la partition en 7 classes	
Inertie interclasse	avant la consolidation : 11.722
	après la consolidation : 11.722
	soit un gain de : -0.0 %

La consolidation des classes n'apporte ici aucun bénéfice au niveau de l'inertie inter-classe.

5.2. Composition des classes

Tableau 13 : Détails de la composition des 7 classes de la partition

Classe numéro 1
Aisne, Ardèche et Drome, Ardennes, Aube, Aude (ADOC11), Aveyron, Calvados (Association MATHILDE), Charente, Cher (IRSA ADOC18), Corrèze, Cote d'Or + Nièvre (ADECA 21), Cotes d'Armor (ADEC), Deux-Sèvres, Doubs, Jura, Haute Saône et Territoire de Belfort, Essonne, Eure et Loir (ADOC 28), Finistère (ADEC 29), Gard et Lozère, Gers, Haute-Marne, Haute-Savoie, Haute-Vienne, Hautes-Alpes, Hautes-Pyrénées, Indre et Loire, Loir et Cher, Loire, Maine et Loire, Manche (IRIS MANCHE), Marne, Mayenne, Meurthe et Moselle, Morbihan (ADECAM), Moselle, Oise, Pyrénées-Orientales, Saône et Loire (ADEMAS 71), Sarthe, Savoie, Somme, Tarn, Tarn et Garonne, Vaucluse, Vendee, Vosges, Yonne (AIDEC)
Classe numéro 2
Allier (ABIDEC), Ariège, Creuse, Indre, Meuse
Classe numéro 3

Ain, Cantal, Haute-Loire, Puy de Dôme (ABIDOC), Charente-Maritime, Hérault, Ille et Vilaine (ADEC 135), Lot, Pas-de-Calais, Réunion, Seine et Marne, Seine-Saint-Denis, Val de Marne

Classe numéro 4

Dordogne (vitalis), Haute-Garonne, Lot et Garonne (Structure de gestion du dépistage du Lot et Garonne), Pyrénées-Atlantiques (PYRADEC PAU)

Classe numéro 5

Alpes de Haute-Provence et Alpes Maritime, Gironde (AGIDECA 33), Hauts-de-Seine, Loire-Atlantique, Rhône, Val d'Oise, Var, Yvelines

Classe numéro 6

Isère, Orne (Régie départementale des cancers)

Classe numéro 7

Paris

La représentation graphique traduit bien la surreprésentation de la classe 1 par rapport aux autres classes, en sous effectif. Une classification en un nombre plus restreint d'ensembles entraîne inévitablement une réduction des différences entre les classes et une augmentation des différences entre les structures qui composent la classe.

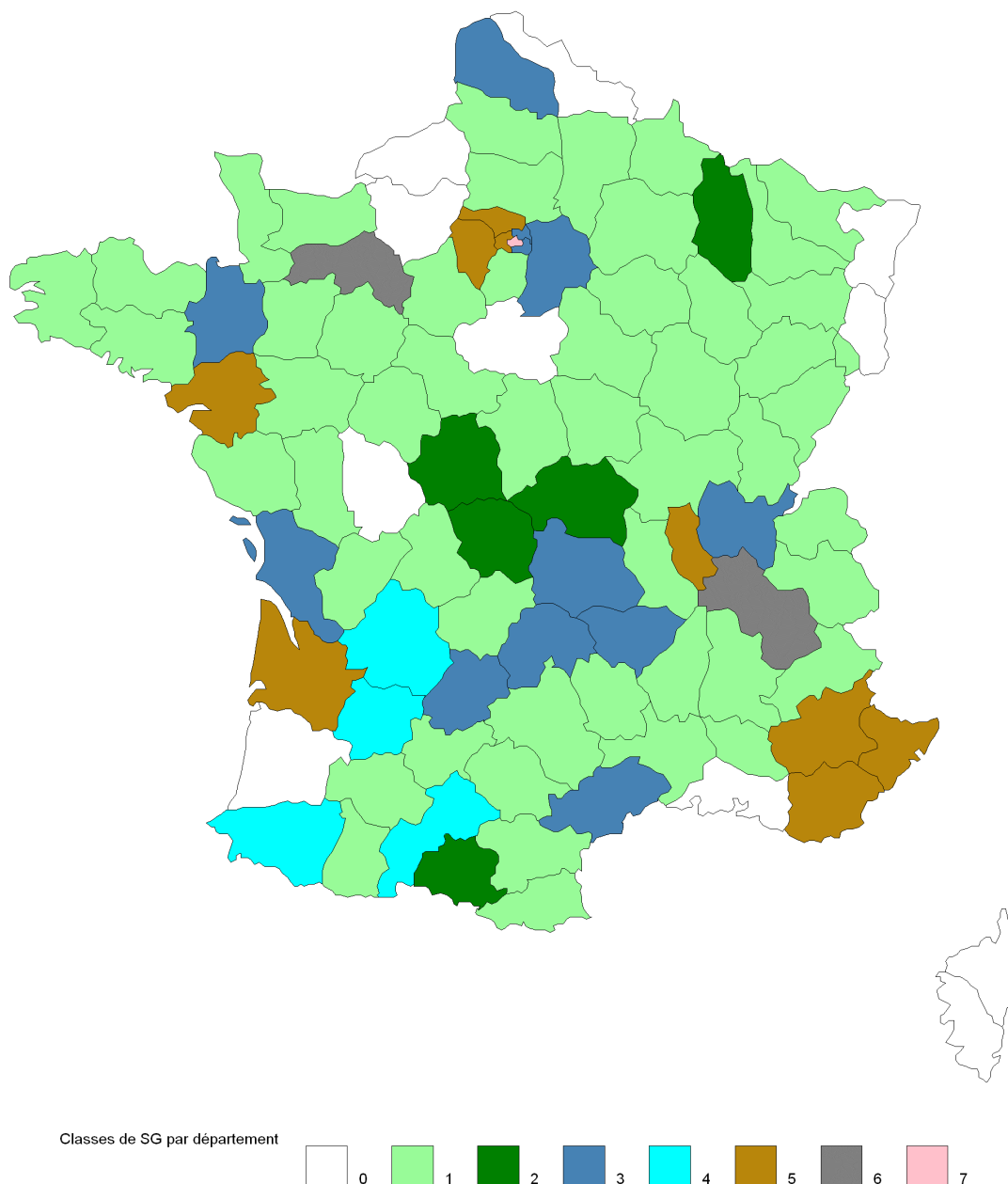


Figure 8 : Représentation géographique de la classification en 7 classes

5.3. Décomposition de l'inertie totale

On remarquera que l'inertie inter-classe est plus faible que pour une partition en 11 classes (Baisse de 7 points) et que l'inertie intra-classe est plus élevée (Hausse de 7 points), ce qui signifie que les classes sont moins différenciées entre elles et regroupent des structures de gestion plus hétérogènes.

Tableau 14 : Décomposition par classe de l'inertie inter-classe et de l'inertie intra-classe

	Classes	Inertie	En 0/00 de l'inertie totale	Effectif absolu	Effectif pondéré (en 0/00)
Inertie totale		42.000	1000	77	1000
Inertie inter		11.722	279	.	.
	1	0.8358	20	46	597
	2	1.8239	43	5	65
	3	1.1883	28	11	143
	4	1.5234	36	4	52
	5	2.2346	53	8	104
	6	1.6006	38	2	26
	7	2.5150	60	1	13
Inertie intra		30.278	721	.	.
	1	17.798	424	46	597
	2	1.5671	37	5	65
	3	4.4956	107	11	143
	4	2.4873	59	4	52
	5	3.3066	79	8	104
	6	0.6235	15	2	26
	7	0.0000	0	1	13

5.4. Caractérisation des classes par les variables : Moyennes, écarts-types et test de significativité.

Tableau 15 : Moyennes, écarts-types et test de significativité par variables pour chacune des 11 classes et pour l'ensemble des structures de gestion

Variable	Statistique	Ensemble	Classe 1	Classe 2	Classe 3	Classe 4	Classe 5	Classe 6	Classe 7
	Eff.pond	77.00	46.00	5.00	11.00	4.00	8.00	2.00	1.00
v2 Procédure amélioration fichier	Moyenne	0.96	1.00	0.80	1.00	0.50	1.00	1.00	1.00
	Ec.type	0.19	0.00	0.40	0.00	0.50	0.00	0.00	0.00
	V.test	7.02	2.14	-1.91	0.72	-4.86	0.60	0.29	0.20
	Proba	0.00	0.03	0.06	0.47	0.00	0.55	0.77	0.84
v3 Délai de transmission	Moyenne	2.70	2.54	4.40	2.00	3.50	3.00	2.50	4.00
	Ec.type	2.12	2.02	2.24	2.04	2.06	2.00	2.50	0.00
	V.test	0.95	-0.79	1.84	-1.18	0.77	0.42	-0.14	0.61
	Proba	0.47	0.43	0.07	0.24	0.44	0.68	0.89	0.54
v4	Moyenne	1.27	1.23	1.14	1.44	1.34	1.29	1.08	2.22

Variable	Statistique	Ensemble	Classe 1	Classe 2	Classe 3	Classe 4	Classe 5	Classe 6	Classe 7
<i>Densité médecins généralistes</i>	Ec.type	0.34	0.26	0.15	0.58	0.16	0.25	0.21	0.00
	V.test	2.28	-1.44	-0.88	1.75	0.38	0.14	-0.80	2.75
	Proba	0.05	0.15	0.38	0.08	0.71	0.89	0.42	0.01
v5 <i>Densité radiologues</i>	Moyenne	0.10	0.09	0.06	0.12	0.11	0.13	0.10	0.32
	Ec.type	0.04	0.02	0.03	0.05	0.03	0.02	0.02	0.00
	V.test	14.22	-2.35	-2.35	1.14	0.50	2.03	-0.10	5.45
	Proba	0.00	0.02	0.02	0.26	0.62	0.04	0.92	0.00
v6 <i>Densité gynécologues</i>	Moyenne	0.06	0.05	0.03	0.05	0.06	0.09	0.05	0.23
	Ec.type	0.03	0.01	0.01	0.02	0.02	0.03	0.02	0.00
	V.test	19.60	-2.27	-1.83	-0.25	0.46	3.04	-0.27	5.85
	Proba	0.00	0.02	0.07	0.80	0.64	0.00	0.78	0.00
v7 <i>Densité HGE</i>	Moyenne	0.05	0.04	0.03	0.05	0.05	0.06	0.05	0.12
	Ec.type	0.02	0.01	0.01	0.02	0.02	0.01	0.01	0.00
	V.test	9.22	-2.38	-2.12	1.69	-0.24	2.18	-0.17	4.53
	Proba	0.00	0.02	0.03	0.09	0.81	0.03	0.87	0.00
v8 <i>Statut juridique</i>	Moyenne	1.13	1.07	1.20	1.09	1.75	1.00	2.00	1.00
	Ec.type	0.47	0.25	0.40	0.29	1.30	0.00	1.00	0.00
	V.test	3.02	-1.47	0.35	-0.30	2.72	-0.83	2.66	-0.28
	Proba	0.01	0.14	0.73	0.77	0.01	0.41	0.01	0.78
v9 <i>Autre dépistage pratiqué</i>	Moyenne	0.03	0.00	0.00	0.00	0.00	0.00	1.00	0.00
	Ec.type	0.16	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	V.test	.	-1.73	-0.38	-0.58	-0.33	-0.48	8.72	-0.16
	Proba	.	0.08	0.71	0.56	0.74	0.63	0.00	0.87
v10 <i>Ancienneté CS</i>	Moyenne	9.63	9.07	9.06	8.95	6.45	14.54	15.74	6.75
	Ec.type	4.76	4.50	3.58	3.48	0.31	5.67	3.43	0.00
	V.test	2.79	-1.24	-0.28	-0.51	-1.36	3.06	1.83	-0.60
	Proba	0.02	0.21	0.78	0.61	0.17	0.00	0.07	0.55
v11 <i>Ancienneté CCR</i>	Moyenne	2.97	3.08	2.35	3.35	1.54	2.09	6.58	2.33
	Ec.type	1.86	1.87	1.41	1.96	0.30	0.90	1.00	0.00
	V.test	2.37	0.63	-0.77	0.73	-1.57	-1.40	2.77	-0.34
	Proba	0.04	0.53	0.44	0.46	0.12	0.16	0.01	0.73
v12 <i>Retours radiologues</i>	Moyenne	0.96	1.00	1.00	0.73	1.00	1.00	1.00	1.00
	Ec.type	0.19	0.00	0.00	0.45	0.00	0.00	0.00	0.00
	V.test	3.75	2.14	0.46	-4.30	0.41	0.60	0.29	0.20
	Proba	0.00	0.03	0.64	0.00	0.68	0.55	0.77	0.84
v13 <i>Niveau de vie département</i>	Moyenne	2.03	1.89	1.00	2.18	1.75	3.19	2.00	4.00
	Ec.type	0.97	0.77	0.00	1.19	0.83	0.70	1.00	0.00
	V.test	4.86	-1.54	-2.44	0.55	-0.59	3.53	-0.05	2.03
	Proba	0.00	0.12	0.01	0.58	0.55	0.00	0.96	0.04
v14 <i>Nb ETP</i>	Moyenne	7.60	7.04	3.91	9.33	5.81	10.77	10.75	8.71
	Ec.type	2.87	2.22	1.00	3.07	2.16	2.54	0.55	0.00
	V.test	6.59	-2.09	-2.95	2.14	-1.28	3.27	1.56	0.38
	Proba	0.00	0.04	0.00	0.03	0.20	0.00	0.12	0.70
v15 <i>Personnel à disposition</i>	Moyenne	0.17	0.17	0.00	0.09	0.00	0.38	0.50	0.00
	Ec.type	0.37	0.38	0.00	0.29	0.00	0.48	0.50	0.00
	V.test	1.08	0.14	-1.04	-0.74	-0.92	1.63	1.26	-0.45

Variable	Statistique	Ensemble	Classe 1	Classe 2	Classe 3	Classe 4	Classe 5	Classe 6	Classe 7
	Proba	0.39	0.89	0.30	0.46	0.36	0.10	0.21	0.65
v16 <i>Personne dédiée aux tâches administratives</i>	Moyenne	0.57	0.57	0.60	0.55	0.25	0.75	0.50	1.00
	Ec.type	0.49	0.50	0.49	0.50	0.43	0.43	0.50	0.00
	V.test	0.57	-0.13	0.13	-0.19	-1.33	1.07	-0.21	0.87
	Proba	0.75	0.89	0.89	0.85	0.19	0.28	0.84	0.39
v17 <i>Budget global/Nb ETPD</i>	Moyenne	163332.18	154509.37	126699.86	177821.41	190378.94	206874.90	126893.46	209310.56
	Ec.type	42744.43	36814.75	14539.05	17679.16	40976.81	59087.14	43219.59	0.00
	V.test	3.93	-2.19	-1.97	1.21	1.29	3.02	-1.21	1.08
	Proba	0.00	0.03	0.05	0.23	0.20	0.00	0.22	0.28
v18 <i>Coût par personnes dépistées CS</i>	Moyenne	18.21	18.37	21.27	18.53	16.33	16.17	17.00	18.12
	Ec.type	3.69	3.63	2.58	2.93	4.99	3.37	3.28	0.00
	V.test	1.23	0.47	1.91	0.31	-1.04	-1.64	-0.47	-0.02
	Proba	0.30	0.64	0.06	0.76	0.30	0.10	0.64	0.98
v19 <i>Pourcentage de participation CS</i>	Moyenne	0.54	0.58	0.54	0.45	0.60	0.46	0.53	0.24
	Ec.type	0.13	0.11	0.04	0.05	0.17	0.15	0.03	0.00
	V.test	4.50	3.46	-0.04	-2.71	0.85	-2.05	-0.20	-2.38
	Proba	0.00	0.00	0.97	0.01	0.40	0.04	0.85	0.02
v20 <i>Population invitée CS/Nb ETP</i>	Moyenne	6328.62	5243.24	4280.55	7725.77	8131.86	9958.31	4460.61	18613.32
	Ec.type	2799.78	1500.98	801.70	1744.31	1233.03	2984.36	2549.98	0.00
	V.test	18.86	-4.12	-1.68	1.78	1.31	3.85	-0.95	4.39
	Proba	0.00	0.00	0.09	0.08	0.19	0.00	0.34	0.00
v22 <i>Mise à disposition des locaux</i>	Moyenne	0.09	0.04	0.00	0.09	0.25	0.38	0.00	0.00
	Ec.type	0.29	0.20	0.00	0.29	0.43	0.48	0.00	0.00
	V.test	1.96	-1.75	-0.73	0.00	1.13	2.93	-0.45	-0.32
	Proba	0.08	0.08	0.47	1.00	0.26	0.00	0.65	0.75
v23 <i>Budget réalisé/Budget prévisionnel</i>	Moyenne	0.94	0.95	1.07	0.94	0.93	0.89	0.83	0.93
	Ec.type	0.12	0.13	0.13	0.09	0.06	0.11	0.14	0.00
	V.test	1.45	0.34	2.34	-0.19	-0.27	-1.30	-1.29	-0.14
	Proba	0.21	0.73	0.02	0.85	0.79	0.19	0.20	0.89
v24 <i>Population cible CS</i>	Moyenne	91789.08	71646.98	32280.20	119515.19	92390.50	181445.01	95972.50	282863.00
	Ec.type	54712.03	31036.32	13105.88	45715.84	38465.14	24753.94	54027.50	0.00
	V.test	20.04	-3.91	-2.50	1.80	0.02	4.86	0.11	3.49
	Proba	0.00	0.00	0.01	0.07	0.98	0.00	0.91	0.00
v26 <i>Coût par personne dépistée CCR</i>	Moyenne	156.97	62.84	65.42	35.09	1944.76	80.13	17.83	27.45
	Ec.type	512.57	121.76	51.92	19.80	1218.04	109.19	2.35	0.00
	V.test	23.40	-1.95	-0.41	-0.85	7.12	-0.45	-0.39	-0.25
	Proba	0.00	0.05	0.68	0.40	0.00	0.66	0.70	0.80
v27 <i>Population cible CCR</i>	Moyenne	178240.22	138773.17	62403.60	240403.93	177930.00	347716.63	185651.50	519714.00
	Ec.type	105354.71	60005.29	24614.22	94346.35	73341.42	46182.53	104348.50	0.00
	V.test	18.83	-3.98	-2.53	2.10	-0.01	4.78	0.10	3.24
	Proba	0.00	0.00	0.01	0.04	1.00	0.00	0.92	0.00
v28 <i>Pourcentage de participation CCR</i>	Moyenne	0.24	0.28	0.25	0.24	0.08	0.12	0.43	0.08
	Ec.type	0.16	0.15	0.13	0.15	0.08	0.08	0.03	0.00
	V.test	2.99	2.20	0.02	-0.00	-2.19	-2.42	1.65	-1.06
	Proba	0.01	0.03	0.98	1.00	0.03	0.02	0.10	0.29
v29	Moyenne	9719.45	8967.47	5197.16	11686.51	2565.02	14138.13	8583.10	40825.26

Variable	Statistique	Ensemble	Classe 1	Classe 2	Classe 3	Classe 4	Classe 5	Classe 6	Classe 7
<i>Population invitée CCR/Nb ETP</i>	Ec.type	6411.23	4205.46	2235.70	5531.75	2207.27	7311.33	4516.31	0.00
	V.test	10.54	-1.25	-1.62	1.09	-2.28	2.05	-0.25	4.85
	Proba	0.00	0.21	0.11	0.27	0.02	0.04	0.80	0.00
<i>v30 Part financement Conseil Général</i>	Moyenne	10.68	10.73	6.40	3.30	5.43	20.47	17.47	40.32
	Ec.type	12.73	12.58	12.80	7.14	5.25	12.07	0.47	0.00
	V.test	2.98	0.04	-0.77	-2.06	-0.84	2.28	0.76	2.33
<i>v31 Mise en place partenariat</i>	Proba	0.01	0.97	0.44	0.04	0.40	0.02	0.45	0.02
	Moyenne	0.88	0.93	0.00	1.00	0.75	1.00	1.00	1.00
	Ec.type	0.32	0.25	0.00	0.00	0.43	0.00	0.00	0.00
<i>v32 Externalisation des tâches</i>	V.test	14.42	1.71	-6.31	1.29	-0.85	1.08	0.52	0.36
	Proba	0.00	0.09	0.00	0.20	0.40	0.28	0.60	0.72
	Moyenne	0.88	0.85	0.60	1.00	1.00	1.00	1.00	1.00
<i>v33 Partage de moyen avec une autre SG</i>	Ec.type	0.32	0.36	0.49	0.00	0.00	0.00	0.00	0.00
	V.test	1.33	-1.17	-2.02	1.29	0.74	1.08	0.52	0.36
	Proba	0.26	0.24	0.04	0.20	0.46	0.28	0.60	0.72
<i>v34 Retour généralistes</i>	Moyenne	0.66	0.63	1.00	0.73	1.00	0.38	0.50	1.00
	Ec.type	0.47	0.48	0.00	0.45	0.00	0.48	0.50	0.00
	V.test	1.49	-0.72	1.64	0.49	1.46	-1.80	-0.49	0.71
<i>v35 Participation médecins coordinateurs séminaires</i>	Proba	0.20	0.47	0.10	0.63	0.15	0.07	0.63	0.48
	Moyenne	0.74	0.65	0.60	1.00	0.75	0.88	1.00	1.00
	Ec.type	0.44	0.48	0.49	0.00	0.43	0.33	0.00	0.00
<i>v36 Part des coûts structurels dans les coûts</i>	V.test	1.36	-2.13	-0.73	2.11	0.05	0.91	0.84	0.59
	Proba	0.24	0.03	0.46	0.04	0.96	0.36	0.40	0.55
	Moyenne	0.95	0.93	1.00	1.00	0.75	1.00	1.00	1.00
<i>v37 Mise en place outils communication</i>	Ec.type	0.22	0.25	0.00	0.00	0.43	0.00	0.00	0.00
	V.test	0.78	-0.63	0.54	0.83	-1.82	0.69	0.33	0.23
	Proba	0.59	0.53	0.59	0.40	0.07	0.49	0.74	0.81
<i>v38 Formation personnel SG</i>	Moyenne	0.09	0.08	0.13	0.09	0.06	0.08	0.06	0.06
	Ec.type	0.04	0.03	0.05	0.02	0.04	0.01	0.01	0.00
	V.test	2.61	-0.22	3.20	0.29	-1.47	-0.62	-1.23	-0.61
<i>v39 Existence convention collective</i>	Proba	0.02	0.82	0.00	0.77	0.14	0.54	0.22	0.54
	Moyenne	1.75	1.83	1.20	1.36	2.00	2.00	2.00	2.00
	Ec.type	0.65	0.56	0.98	0.88	0.00	0.00	0.00	0.00
<i>v40 Réunions avec l'équipe de la SG</i>	V.test	1.82	1.19	-1.96	-2.14	0.78	1.13	0.54	0.38
	Proba	0.11	0.23	0.05	0.03	0.44	0.26	0.59	0.70
	Moyenne	0.84	0.83	1.00	0.91	1.00	0.88	0.50	0.00
<i>v39 Existence convention collective</i>	Ec.type	0.36	0.38	0.00	0.29	0.00	0.33	0.50	0.00
	V.test	1.62	-0.53	0.99	0.64	0.88	0.25	-1.35	-2.33
	Proba	0.15	0.60	0.32	0.52	0.38	0.80	0.18	0.02
<i>v39 Existence convention collective</i>	Moyenne	0.58	0.72	0.20	0.09	0.75	0.88	0.00	0.00
	Ec.type	0.49	0.45	0.40	0.29	0.43	0.33	0.00	0.00
	V.test	5.57	2.87	-1.79	-3.56	0.69	1.75	-1.69	-1.19
<i>v40 Réunions avec l'équipe de la SG</i>	Proba	0.00	0.00	0.07	0.00	0.49	0.08	0.09	0.24
	Moyenne	0.81	1.00	0.20	0.27	0.75	1.00	0.50	0.00
	Ec.type	0.60	0.55	0.40	0.45	0.43	0.50	0.50	0.00
<i>v40 Réunions avec l'équipe de la SG</i>	V.test	4.52	3.43	-2.30	-3.14	-0.19	0.96	-0.72	-1.33

Variable	Statistique	Ensemble	Classe 1	Classe 2	Classe 3	Classe 4	Classe 5	Classe 6	Classe 7
SG	Proba	0.00	0.00	0.02	0.00	0.85	0.34	0.47	0.18
v41	Moyenne	4.59	4.61	4.29	5.07	3.87	4.29	4.29	6.00
<i>Procédures internes suivi assurance qualité</i>	Ec.type	1.84	1.87	1.53	1.18	2.31	2.14	1.71	0.00
	V.test	0.37	0.11	-0.38	0.92	-0.81	-0.49	-0.24	0.76
	Proba	0.90	0.92	0.71	0.36	0.42	0.62	0.81	0.44
v42	Moyenne	14.99	15.92	10.49	12.50	15.36	16.38	12.40	14.60
<i>Délai moyen rendu des résultats</i>	Ec.type	6.19	7.14	0.98	3.70	1.32	4.59	4.40	0.00
	V.test	1.02	1.59	-1.67	-1.43	0.12	0.67	-0.59	-0.06
	Proba	0.42	0.11	0.10	0.15	0.90	0.51	0.55	0.95
v43	Moyenne	2.27	2.37	2.20	1.18	2.75	2.13	5.50	3.00
<i>Organisation information professionnels dépistage</i>	Ec.type	1.65	1.71	0.98	1.03	1.09	1.45	0.50	0.00
	V.test	2.42	0.62	-0.10	-2.36	0.59	-0.27	2.79	0.44
	Proba	0.03	0.53	0.92	0.02	0.55	0.79	0.01	0.66
v44	Moyenne	2.14	2.07	1.00	2.00	2.50	3.25	2.00	3.00
<i>Mise en place initiatives locales</i>	Ec.type	1.50	1.48	0.63	1.04	2.29	1.56	0.00	0.00
	V.test	1.36	-0.55	-1.75	-0.34	0.49	2.19	-0.14	0.57
	Proba	0.24	0.58	0.08	0.73	0.63	0.03	0.89	0.57
v45	Moyenne	0.07	0.06	0.04	0.06	0.04	0.17	0.14	0.28
<i>Nb perdus de vue CS / Population dépistée pos.</i>	Ec.type	0.09	0.06	0.03	0.05	0.03	0.19	0.07	0.00
	V.test	3.49	-1.61	-0.89	-0.50	-0.87	3.17	1.00	2.21
	Proba	0.00	0.11	0.38	0.62	0.38	0.00	0.32	0.03

5.5. Descriptif des classes

La **classe 1** est la classe qui comporte le plus grand nombre de structures de gestion (46 SG), elle tend donc à se rapprocher des observations moyennes faites sur l'ensemble des structures de gestion. Elle est principalement caractérisée par des populations cibles pour le dépistage du cancer du sein et le dépistage du cancer colorectal inférieures à la moyenne (respectivement 71647 vs 91789 pour le cancer du sein, et 138773 vs 178240 pour le cancer colorectal) et par un rapport population invitée / nombre d'ETP dans le cadre du cancer du sein moins important que pour l'ensemble des structures (5243/ETP vs 6328/ETP). On notera également que dans la classe 1, le taux de participation au dépistage du cancer du sein est légèrement supérieur à celui de l'ensemble des SG (58% vs 54%).

La **classe 2** compte 5 structures de gestion. Elle se caractérise par une absence totale de partenariat avec d'autres structures. Elle se démarque également avec une part plus importante des coûts structurels dans les coûts (13% dans la classe 2 vs 9% sur l'ensemble des structures).

La **classe 3** est composée de 11 structures de gestion. Elle se distingue par l'absence d'organisation de réunion avec l'ensemble des membres de l'équipe (seulement 20% des structures de la classe 3

organisent des réunions, contre 88% si l'on considère l'ensemble des structures), ainsi que par un faible nombre de structures à posséder une convention collective (9% vs 58%). Les structures de cette classe présentent un taux de participation au dépistage du cancer du sein inférieur de 10 points à celui observé sur l'ensemble des SG. Le retour d'information aux radiologues est plus faible au sein de la classe 3 que sur la totalité des structures.

La **classe 4** se caractérise principalement par un coût par personne dépistée pour le cancer colorectal qui est plus de 10 fois supérieur à celui observé sur l'ensemble des structures (1945€ vs 157€). Au sein de la classe 4 seulement 1 structures de gestion sur 2 à recours au fichier de La Poste pour améliorer son fichier. La classe 4 est composée de 4 structures de gestion.

La **classe 5** compte 8 structures de gestion. Elle se démarque d'abord par des populations cibles pour les dépistages du cancer du sein et le cancer colorectal très supérieures à la moyenne (respectivement 181445 vs 91789 pour le cancer du sein, et 347716 vs 178240 pour le cancer colorectal). Le rapport population cible invitée / nombre d'ETP pour le cancer du sein est plus élevé dans cette classe que sur l'ensemble des structures (9958/ETP vs 6329/ETP). Le niveau de vie des départements où sont situées les structures de cette classe est plus élevé que la moyenne (15766€ - 17350€ vs 14820€ - 15766€).

La **classe 6** est caractérisée par la pratique d'un autre dépistage (100% vs 13% en moyenne). Elle se distingue également par une ancienneté dans le dépistage du cancer colorectal plus importante (6 ans vs 3 ans en moyenne) et par l'organisation de nombreux événements d'information sur le dépistage organisé à l'intention des professionnels de santé. Cette classe est composée de 2 structures de gestion.

La **classe 7** ne comprend qu'une seule structure de gestion (Paris). Cette classe se distingue par une densité de médecins supérieur à celle observée sur l'ensemble des SG (0,32 vs 0,1 pour les radiologues, 0,23 vs 0,06 pour les gynécologues et 0,12 vs 0,05 pour les HGE). Elle se caractérise également par un rapport population cible invitée / Nombre d'ETP supérieur à la moyenne que ce soit pour le cancer du sein ou le cancer colorectal (18613/ETP vs 6328/ETP pour le cancer du sein et 40825/ETP vs 9719/ETP pour le cancer colorectal). La classe 7 possède une population cible pour le cancer du sein deux fois supérieure à celle observée sur l'ensemble des structures (282863 vs 91789).

6. Conclusion

La classification en 7 ensembles est celle qui s'avère la plus fonctionnelle. Cependant, en raison la relation qui unie l'inertie inter-classe et l'inertie intra-classe, la réduction du nombre de groupes construits au cours de la classification tend à gommer les différences qui apparaissent entre les groupes de structures de gestion et, dans le même temps, accentue les différences à l'intérieur de chacun de ces groupes.

La classification en 11 classes remplit les critères mathématiques d'une bonne classification (perte d'inertie inter-classe et Cubic Clustering Criterion), de ce fait nous recommandons de séparer les structures de gestion en 11 groupes homogènes.

7. Références bibliographiques

Andridge R, A review of Hot-deck imputation for survey Non-response, *International Statistical Review*, 78, 1, 40-64

Donders ART, Geert JMG, Review : a gentle introduction to imputation of missing values, *Journal of Clinical Epidemiology*, 2006, 59, 1087-1091

Donzé L, L'imputation des données manquantes, la technique de l'imputation multiple, les conséquences sur l'analyse des données : l'enquête 1999 KOF/ETHZ sur l'innovation, 2001, Ecole Polytechnique de Zurich

Durrant G, Imputation methods for handling item-nonresponse in the social sciences : A methodological review, *National Centre for Research Methods Working Paper Series*, 2005

Horton N, Lipsitz S, Multiple imputation in practice : comparison of software packages for regression models with missing variables, *The American Statistician*, 2001, 55(3), 244-254

Lerman IC, Bachar K, Comparaison de deux critères en classification ascendante hiérarchique sous contrainte de contiguïté. Application en imagerie numérique, *Journal de la Société Française de Statistique*, 2008, Tome 149, 2 : 45-74

Rubin D, *Multiple imputation for nonresponse in surveys*, 1987, Wiley Series

Van Buuren S, Multiple imputation of discrete and continuous data by fully conditional specification, *Statistical Methods in Medical Research*, 2007, 16 : 219-242

Vanderschelden M, Homogamie et choix du conjoint Traitement de la non-réponse Imputation de variables qualitatives corrélées, *INSEE*, 2005