

## **L’analyse statistique implicative au service d’une étude sur l’anticipation des départs à la retraite**

**Jean-Claude Oriol\***, **Anicée Chancel\*\***,

\*Université Lyon 2, IUT Lumière Cerral, Campus Porte des Alpes,  
Bron

\*\*Chargée d’études statistiques, IUT Lumière, Campus Porte des  
Alpes, Bron

Jean-Claude.Oriol@univ-lyon2.fr, anicee.chancel@hotmail.fr,

**Résumé :** La réforme des retraites en 2008 dans un certain nombre d’entreprises a remplacé un système rigide par un système flexible. Des décisions individuelles permettent aux agents de prolonger ou non leur activité. La gestion des ressources humaines doit donc anticiper des départs par des embauches. Elle a donc commandé une étude statistique afin de déterminer les variables en jeu dans ces décisions. Notre propos est d’exposer les méthodes utilisées et de montrer qu’en ce cas une ACM n’a pas donné de résultat et qu’en revanche l’ASI a permis de montrer un certain nombre de résultats.

**Abstract.** The reform of the pensions of 2008 to state French companies replaced a stiff system by a flexible system. Individual decisions allow the agents to prolong or not their activity. The human resources management thus has to anticipate departures by hiring. She thus commanded a statistical study to determine variables in set in these decisions. Our comment is to expose the used methods and to show that in this case an ACM did not give result and that on the other hand the ASI allowed to show certain number of results.

### **1 1. Introduction**

Dès les premier mois qui ont suivi la réforme de la retraite en juillet 2008, un certain nombre d’entreprises ont constaté que celle-ci entraînait des modifications importantes du comportement des agents ou des employés selon le cas vis-à-vis des départs – avec pour conséquences immédiate une augmentation de plusieurs dizaines de millions d’euros par an sur ses dépenses de personnels (dans l’exemple traité ici sur une masse totale d’environ 8,6 milliards d’euros par an). Une mesure phare de la réforme est en effet la suppression de la mise à la retraite d’office. Elle implique que les agents peuvent dorénavant choisir leur date de départ : notre première étude concerne la SNCF qui ne peut plus anticiper les cessations comme elle le faisait auparavant ; l’objectif de notre travail est de déterminer les différents

paramètres qui influencent la prise de décision de la prolongation d’activité, de repérer leur importance et d’en établir un classement.

## 2 Population, individus, variables

Pour mener notre étude nous avons un certain nombre de contraintes matérielles et nous allons décrire la population enquêtée. Nous avons choisi de présenter ici la première population étudiée dans le cadre d’une étude plus générale qui concerne le personnel de la SNCF.

### 2.1 Qui prendre en compte dans l’étude ?

L’étude concerne les agents affiliés au régime de retraite de la SNCF. Les contractuels (CDD et CDI) ne sont donc pas pris en compte. Les cadres supérieurs n’ont pas non plus été retenus dans l’étude : leur comportement vis-à-vis de la retraite était déjà spécifique avant la réforme. De plus, l’étude ne prend en compte que les agents pouvant partir à la retraite entre le 1er janvier 2008 et le 31 décembre 2009 (voir tableau 1).

L’âge d’ouverture des droits n’étant pas disponible dans l’infocentre RH de la SNCF, nous ne pouvons pas tenir compte des changements de catégorie en cours de carrière<sup>1</sup>.

**Table 1.** Age d’ouverture des droits en fonction de l’année de naissance.

Année de naissance	Année d’ouverture des droits pour les agents roulants (1)	Année d’ouverture des droits pour les agents sédentaires (2)
1953	2003	2008
1954	2004	2009
1955	2005	2010
1956	2006	2011
1957	2007	2012
1958	2008	2013
1959	2009	2014

Selon la nomenclature (1) désignent les Agents ayant travaillé pendant au moins 15 ans dans un emploi de catégorie roulant et (2) Agents ayant travaillé entre 0 et 15 ans dans un emploi de catégorie roulant

<sup>1</sup> Par exemple, un agent né en 1959, actuellement dans un emploi de sédentaire et ayant eu un emploi de roulant pendant vingt ans peut partir à la retraite en 2009. Cependant, il ne peut pas être pris en compte dans l’étude.

## 2.2 Population, individus

Parmi les 12 058 agents retenus, 7 061 (58,6%) ont prolongé leur activité et 4 997 (41,4%) ne l'ont pas prolongée<sup>2</sup>. Ce sera donc notre population, les individus sont les agents retenus.

## 2.3 Variables

Les variables étudiées ont été choisies selon deux critères : Dans un premier temps, elles ont été choisies en fonction de leur disponibilité et de leur valeur qualitative réputée. Dans un second temps, après discussion avec les professionnels du métier, les variables les plus pertinentes ont été retenues.

- Situation personnelle
  - Genre
  - Situation familiale
  - Interruption de Longue Durée (ILD)
  - Région
- Éléments de rémunération
  - Qualification
  - Échelon
  - Ancienneté à la SNCF
  - Pourcentage d'utilisation
- Activité
  - Catégorie (roulant ou sédentaire)
  - Branche
  - Pénibilité

## 3 Etude individuelle des variables : synthèse

Après une étude individuelle des variables nous avons entrevu quelques facteurs expliquant la prolongation d'activité ; ils semblent être multiples et chaque facteur ne peut expliquer seul la volonté de prolongation d'activité. D'après cette étude, seules les variables concernant la situation familiale et la catégorie ne sont pas des facteurs explicatifs de la décision de prolonger son activité ou non. Concernant la situation personnelle, nous avons constaté que les femmes, contrairement aux hommes, prolongent leur activité. De plus, les agents étant en situation d'interruption de longue durée prolongent leur activité. Les éléments de rémunération sont des facteurs explicatifs de la prolongation d'activité. Nous avons remarqué dans un premier temps que les cadres ont tendance à prolonger leur

---

<sup>2</sup> La prolongation d'activité dépend de l'âge auquel l'agent part à la retraite. Un agent ne prolonge pas son activité lorsqu'il part au plus tard dans le mois suivant la date d'ouverture de ses droits.

activité contrairement aux agents d'exécution (hors agent de conduite). Ensuite, plus un agent est ancien à la SNCF, moins il prolonge son activité. De plus, les agents accédant à l'échelon maximal (8 pour les agents de conduite et 10 pour les autres) prolongent leur activité. Enfin, les agents travaillant à temps partiel prolongent leur activité contrairement aux agents travaillant à temps plein. Les conclusions concernant l'ancienneté et l'échelon ne sont pas contradictoires. En effet, ces variables ne sont pas équivalentes et elles ont été étudiées de façon individuelle. Concernant l'activité, les agents des branches direction de la circulation ferroviaire et fonctions transverses ont tendance à prolonger leur activité alors que les agents des branches traction, infrastructure et matériel ont tendance à ne pas prolonger leur activité. Les agents travaillant en Ile de France, Alsace, Lorraine ou Provence-Alpes-Côte d'azur prolongent leur activité contrairement à ceux travaillant en Aquitaine, Auvergne, Centre, Bourgogne ou Champagne-Ardenne. Les agents occupant un emploi considéré comme pénible ne prolongent pas leur activité.

#### **4 Analyse statistique implicative**

Après avoir analysé dans la précédente partie chaque variable séparément, nous allons maintenant nous attacher à les étudier ensemble. Ainsi, nous pourrions connaître leur importance respective et surtout leurs inter-relations.

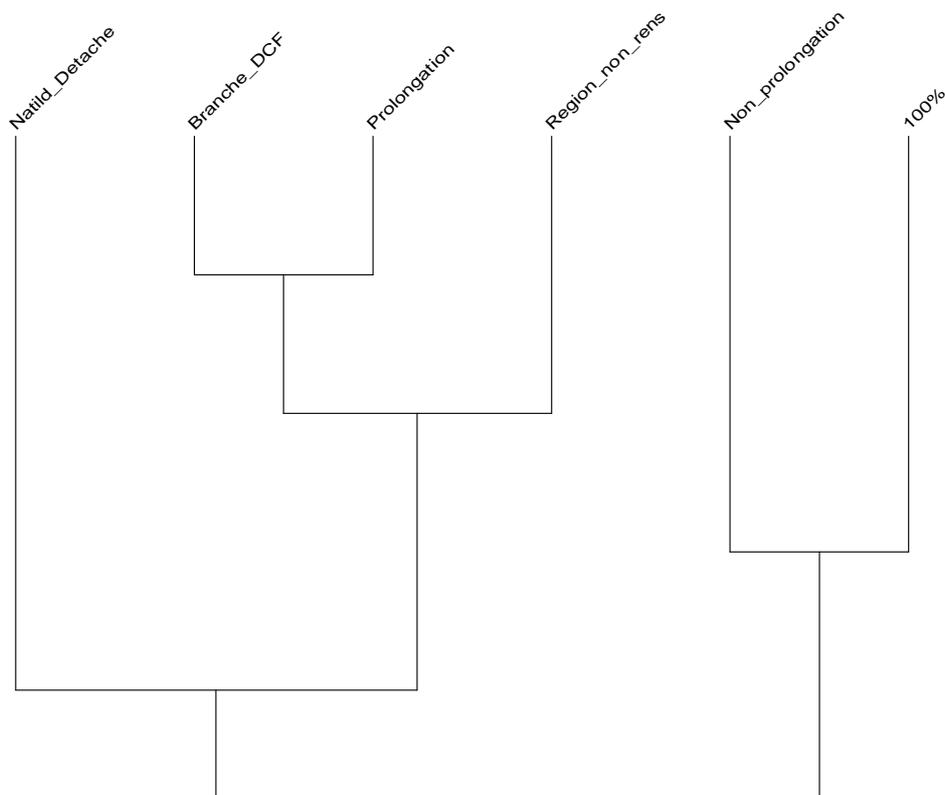
##### *4.1 . Détermination des implications entre variables*

Dans les situations réelles, «l'implication stricte de la variable a sur la variable b, situation où tout individu qui vérifierait a vérifierait également b», n'est observée qu'exceptionnellement. En effet, il est fréquent de rencontrer des contre-exemples, situations dans lesquelles des individus vérifient a mais ne vérifient pas b. L'analyse statistique implicative permet d'extraire d'un ensemble de données des règles ou quasi-règles d'association<sup>3</sup> entre les variables et de fournir un indice de qualité de ces associations. Elle va ainsi modéliser des relations du type « si a alors presque b ». « Autrement dit, les individus ont tendance à être b s'ils sont a ».

- *Arbre des similarités*

Pour déterminer les similarités, on compte le nombre de personnes qui vérifient simultanément les variables a et b. On calcule aussi la valeur théorique, où seul le

hasard interviendrait. Deux variables se ressemblent d'autant plus que la valeur calculée est supérieure à la valeur théorique. L'arbre est construit de la manière suivante : à la première étape, on a n groupes composés d'une seule variable (autant de groupes que de variables). On regroupe alors les deux groupes qui se ressemblent le plus selon le critère de similarité. On obtient donc n-1 groupes composés d'une ou plusieurs variables. A la deuxième étape, on regroupe les deux groupes qui se ressemblent le plus. Et ainsi de suite jusqu'à ce que les groupes obtenus soient très différents les uns des autres. Remarque : la hiérarchie ascendante est représentée par un graphique descendant.



**Figure 1 : Arbre des similarités. Branche DCF <sup>4</sup>**

Nous constatons que les variables « non prolongation » et « 100% » sont similaires. (IDS<sup>5</sup> : 0,955217) C'est-à-dire qu'il y a beaucoup d'agents à temps plein et qui ne prolongent pas leur activité eu égard du nombre d'agents qu'il y aurait eu dans le

3 Nous rappelons qu'une quasi-règle d'association est une règle du type : si une personne présente la caractéristique a, alors elle a une probabilité relativement forte de présenter aussi la caractéristique b.

4 Branche\_DCF : Branche direction de la circulation ferroviaire / region\_non\_rens : région non renseignée

5 IDS : Indice de similarité. Les valeurs sont des arrondis de probabilité. La valeur 1 n'existe qu'en théorie.

cas où seul le hasard interviendrait. De même, les variables « branche Direction des Circulations Ferroviaires » et « prolongation » sont similaires (IDS: 1). Au niveau suivant (IDS : 0,99997), la variable « région non renseignée » leur ressemble. Au niveau suivant (IDS: 0,925657), la variable « détaché » est similaire à ces trois variables.

- *Arbre cohésitif*

L'arbre cohésitif se construit de la même manière que l'arbre des similarités. A la première étape, on a n groupes composés d'une seule variable (autant de groupes que de variables). On regroupe alors les deux groupes qui se ressemblent le plus selon le critère de cohésion. On obtient donc n-1 groupes composés d'une ou plusieurs variables. A la deuxième étape, on regroupe les deux groupes qui se ressemblent le plus. Et ainsi de suite jusqu'à ce que les groupes obtenus soient très différents les uns des autres.

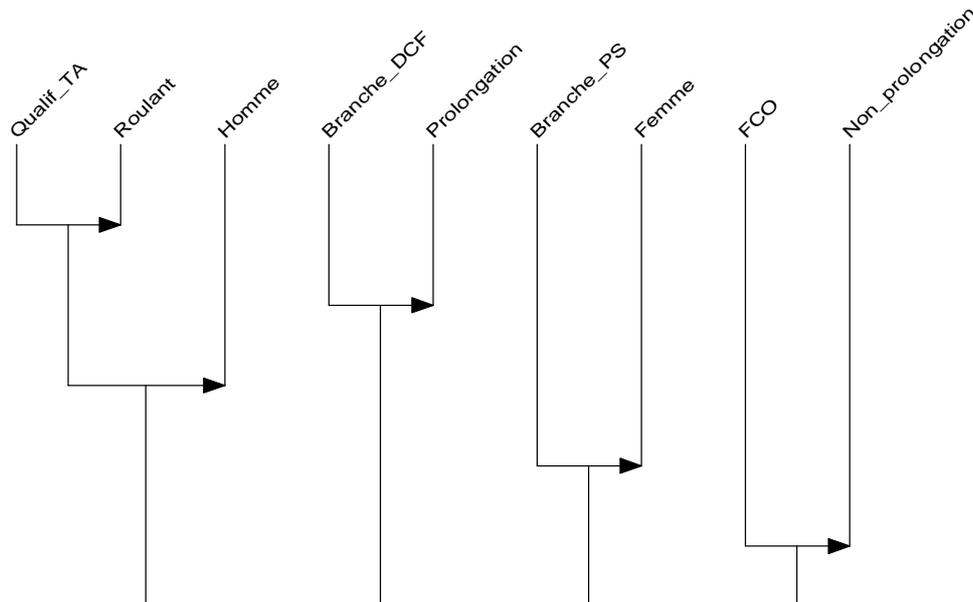


Figure 2 : **Arbre cohésitif. DCF**<sup>6</sup>

Nous remarquons que l'emploi dans la branche Direction des Circulations Ferroviaires implique la prolongation d'activité avec un critère de cohésion de 1. De plus, nous constatons que le fait de travailler dans la région Franche-Comté implique la non prolongation d'activité mais avec un critère de cohésion de seulement 0,532. Ensuite, l'emploi dans la branche Protection Sociale implique le fait d'être une femme (critère de cohésion : 0,99). Enfin, un emploi de qualification

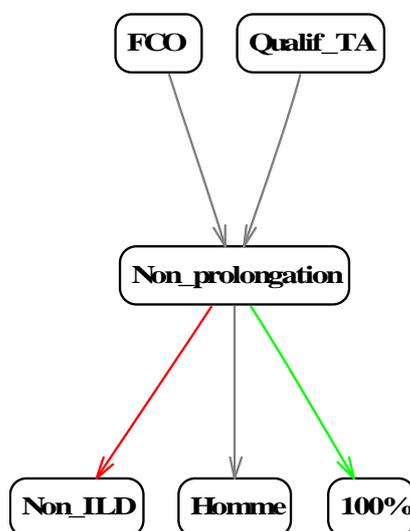
TA implique un emploi de roulant (critère de cohésion : 1). Au niveau suivant (critère de cohésion : 1), ces variables impliquent le fait d’être un homme.

- *Graphes implicatifs*

Pour déterminer les implications, on compte le nombre de contre-exemples observés dans le fichier et on calcule le nombre de contre exemple théorique que l’on obtiendrait s’il y avait indépendance entre les variables a et b. Plus le nombre de contre-exemples observés est petit par rapport au nombre de contre-exemples théorique et plus la règle « si a alors presque b » est admissible.

Le graphe dessine toutes les implications dont le critère d’implication est supérieur ou égal au seuil donné.

- Non prolongation d’activité



**Figure 3 : Graphe implicatif obtenu aux seuils 0.95 (en rouge), 0.90 (en bleu), 0.80 (en vert) et 0.70 (en gris) en ne sélectionnant que les chemins dans lesquels la variable non prolongation est présente.**<sup>7</sup>

Les agents travaillant dans la région Franche-Comté et les agents de qualification TA ne prolongent pas leur activité. De plus, les agents qui ne prolongent pas leur activité sont des hommes, sont à temps plein et ne sont pas en situation d’interruption de longue durée. Les agents travaillant dans la région Franche-Comté et les agents de qualification TA ne prolongent pas leur activité. Les agents

---

6 Branche\_DCF : Branche direction de la circulation ferroviaire / Branche\_PS : branche protection sociale / FCO : Franche-Comté  
7 FC : Région Franche-Comté / Non\_ILD : agent n’étant pas en situation d’interruption de longue durée

qui ne prolongent pas leur activité sont des hommes, sont à temps plein et ne sont pas en situation d'interruption de longue durée.

- Prolongation d’activité

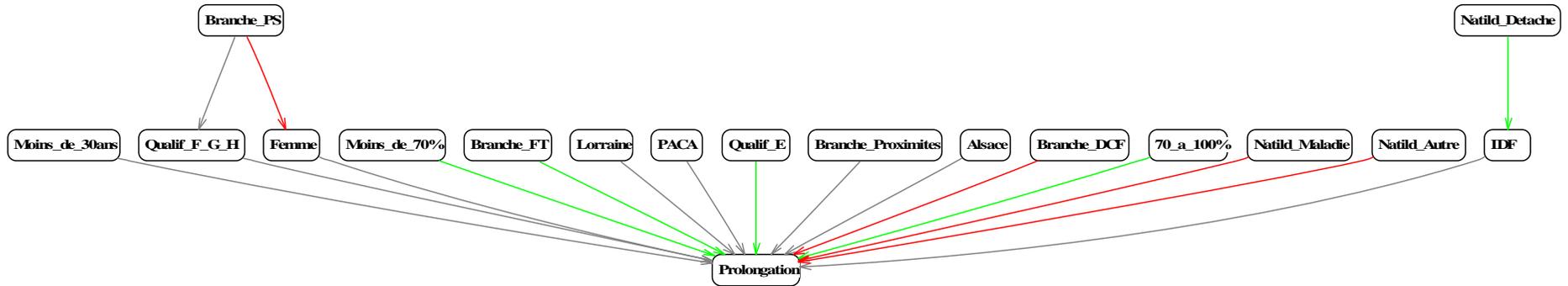


Figure 4 : Graphe implicatif obtenu aux seuils 0.95 (en rouge), 0.90 (en bleu), 0.80 (en vert) et 0.70 (en gris) en ne sélectionnant que les chemins dans lesquels la variable prolongation est présente.<sup>8</sup>

---

8 Moins\_de\_30ans : ancienneté inférieure à 30 ans à la SNCF  
 Branche\_PS : Branche protection sociale  
 Moins\_de\_70% : Temps partiel inférieur à 70%  
 Branche\_FT : Branche fonctions transverses  
 PACA : Région Provence-Alpes-Côte d’azur  
 Branche\_DCF : Branche direction de la circulation ferroviaire  
 70\_a\_100% : Temps partiel supérieur ou égal à 70%  
 Natild\_Autre : Nature d’interruption de longue durée autre que détaché ou maladie

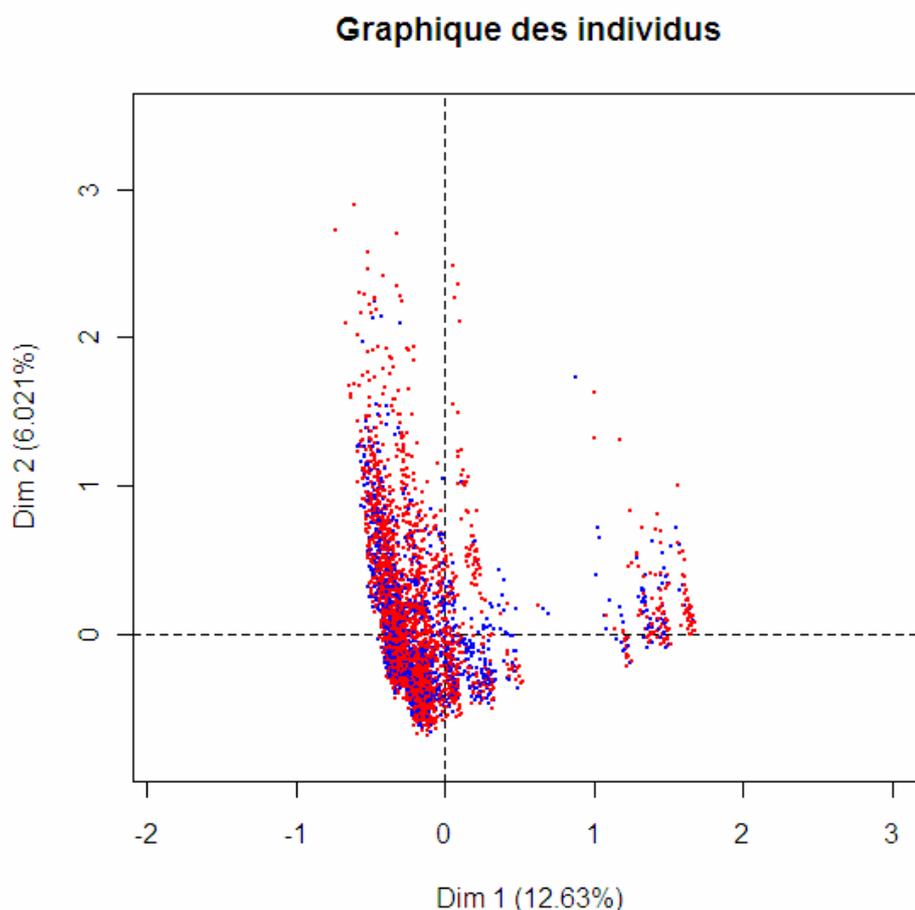
On constate que les agents des branches protection sociale, fonctions transverses, proximités et direction de la circulation ferroviaire prolongent leur activité. Les agents en situation d'interruption de longue durée, quelle qu'en soit la raison (détaché, maladie ou autre) prolongent eux aussi leur activité. On remarque la même chose concernant les personnes travaillant dans les régions Ile de France, Provence-Alpes-Côte d'Azur, Alsace et Lorraine et les agents de qualification E à H. Les femmes, les agents dont l'ancienneté à la SNCF est inférieure à 30 ans et les agents employés à temps partiel prolongent eux aussi leur activité. On constate aussi que la branche protection sociale est principalement composée de femmes. On remarque donc que les cadres (qualification F à H) prolongent leur activité. Cela peut s'expliquer par leur représentation de leur travail et de la retraite. De plus, dans la majorité des cas, les cadres occupent des emplois qui ne sont pas considérés comme pénible. Enfin, l'aspect financier joue un rôle important. En effet, ces agents ont un niveau de vie élevé et souhaitent le conserver. De plus, plus le salaire est important et plus la prolongation d'activité est avantageuse. Les femmes sont plus souvent dans une situation d'emploi à temps partiel et/ou d'ancienneté inférieure à 30 ans que les hommes (35% des femmes contre 26% des hommes). C'est l'une des raisons qui incitent les femmes à prolonger leur activité. Dans cette situation, c'est l'aspect financier qui est en jeu. De plus, les femmes ont plus souvent des absences liées aux besoins de leurs enfants que les hommes. Une fois encore c'est le facteur financier qui influence la décision de prolonger son activité. Les résultats concernant les régions sont assez surprenants. En effet, on pouvait s'attendre à une opposition nord-sud.

## **5 Analyse des correspondances multiples<sup>9</sup>**

Les techniques d'analyse factorielle permettent des visualisations graphiques et l'établissement de profils-types. La majorité des variables de mon étude sont de type qualitatif. L'Analyse des Correspondances Multiples (ACM) est la méthode adaptée à ce type de variables. De plus, les variables quantitatives peuvent aussi être prises en compte. Il suffit pour cela de les transformer en classes.

---

<sup>9</sup> Le détail des variables utilisées est disponible en annexe



**Figure 5 : graphique des individus. En rouge on retrouve les agents ayant prolongé leur activité et en bleu ceux ne l’ayant pas prolongé**

Ce graphique représente les individus sur les deux premiers axes factoriels. Le premier axe représente 12.6% de la dispersion et le deuxième en représente 6%. Ces deux axes ont été choisis car, ensemble, ils représentent le mieux la réalité, soit 18.6% de dispersion. Ce taux peut paraître très faible de premier abord mais il est relativement élevé comparativement à une situation de distribution uniforme. En effet, si nous étions dans une situation de distribution uniforme, les deux axes représenteraient 6% de la réalité.

Ce graphique nous montre que les agents prolongeant leur activité et ceux ne le prolongeant pas n’ont pas été séparés par l’ACM. On peut donc en déduire que les motivations de départ en retraite sont complexes et dépendent des individus ou que les facteurs explicatifs n’ont pas tous été pris en compte dans l’étude. Malgré cela, l’ACM permet de définir des tendances.

Grâce au trois premiers axes, qui représentent 23,4% de la réalité nous pouvons dire que les agents prolongeant leur activité sont des femmes sédentaires, travaillant à la SNCF depuis au moins 35 ans, notamment dans la branche fonctions transverses des agents et qui sont à l'échelon maximum à la date d'ouverture de leurs droits.

A contrario, les agents ne prolongeant pas leur activité sont des roulants de qualification TA ou TB travaillant dans la branche traction et qui, à la date d'ouverture des droits, sont au maximum à l'échelon 09 (07 pour les roulants).

## **6 Conclusion**

Cet article a été complété par d'autres travaux analogues en particulier une étude concernant le même problème avec des agents de la fonction publique avec des résultats assez semblables. Ici on constate aisément que face à un problème d'une certaine taille (plus de 12000 individus, 67 variables) et d'une question complexe (raisons de continuer son activité professionnelle ou non), dans laquelle les raisons s'entrecroisent, l'ACM ne nous a pas permis de dégager des éléments forts tout au plus quelques tendances. Sans doute aurions nous pu diminuer le nombre de variables à prendre en compte mais nous aurions peut-être ignoré des facteurs importants. En revanche l'analyse statistique implicite par le biais de la sélection des graphes implicatifs d'un certain niveau d'implication contenant soit la variable « non prolongation » soit la variable « prolongation » permet une lecture immédiate et de dégager des niveaux d'implication intéressants entre certaines variables.

## **Références**

- Couturier, R. (2005). Un système de recommandation basé sur l'a.s.i. In *Troisième rencontre internationale de l'Analyse Statistique Implicative (ASIS)*, pp. 157–162.
- Oriol, J-C. (2007). *Formation à la statistique par la pratique d'enquêtes par questionnaires et la simulation : étude didactique d'une expérience d'enseignement dans un département d'IUT (sous la direction de Régnier J-C)*, Université Lumière - Lyon 2 (17/11/2007), Retrieved from <http://tel.archives-ouvertes.fr/tel-00191166/fr/> Accessed 17 dec. 2010
- Gras, R. (1996). *L'implication statistique*, La Pensée Sauvage Editions, Grenoble.
- Gras, R. (2005). Panorama du développement de l'A.S.I. à travers des situations fondatrices. In *Actes de la 3ème Rencontre Internationale A.S.I.*, pp. 9–33. Université de Palerme.

Gras, R., Régnier J-C., Guillet, F. (Eds) (2009). *Analyse Statistique Implicative. Une méthode d'analyse des données pour la recherche de causalités*. RNTI E-16, Toulouse : Cépaduès éditions

### Annexe : Variables concernant l'étude des agents SNCF

L'étude porte sur les agents SNCF dont la date d'ouverture des droits à la retraite est comprise entre le 1<sup>er</sup> janvier 2008 et le 31 décembre 2009. Le fichier étudié comporte 12 476 individus décrits par 67 variables.(binaires)

Variable	Remarque	Variable	Remarque
Femme		Branche_Proximites	
Homme		Branche_Traction	
Celibat		Branche_TL	Transports et logistique
Divorce		Branche_VFE	Voyageurs France Europe
Mariage		Alsace	
Séparation		Aquitaine	
Veuvage		Auvergne	
Natild_Autre		BN	Basse Normandie
Natild_Detache		Bourgogne	
Natild_Maladie		Bretagne	
Non_ILD		Centre	
Qualif_A_B_C		CA	Champagne Ardenne
Qualif_D_E		FC	Franche Comté
Qualif_F_G_H		HN	Haute Normandie
Qualif_TA_TB		IDF	Ile de France
Qualif_A		LR	Languedoc Roussillon
Qualif_B		Limousin	
Qualif_C		Lorraine	
Qualif_D		MP	Midi Pyrénées
Qualif_E		NPC	Nord Pas de calais
Qualif_F		PL	Pays de la Loire
Qualif_G		Picardie	
Qualif_H		PC	Poitou Charente
Qualif_TA		PACA	Provence Alpes Cote d'Azur
Qualif_TB		Region_non_rens	Région non renseignée
Roulant		RA	Rhône Alpes

<b>Variable</b>	<b>Remarque</b>	<b>Variable</b>	<b>Remarque</b>
Sédentaire		Metier_non_penible	
Echelon_00_a_08		Metier_penible	
Echelon_09		Penibilite_non_rens	Pénibilité non renseignée
Echelon_10		Non_prolongation	
Branche_non_rens	Branche non renseignée	Prolongation	
Branche_DCF	Direction Circulation Ferroviaire	Anciennete1	Moins de 30 ans
Branche_FT	Fonctions transverses	Anciennete2	Entre 30 et 34 ans
Branche_Gares		Anciennete3	35 ans et plus
Branche_Infra	Infrastructure	Moins_de_70%	
Branche_Materiel		70_a_100%	
Branche_PS	Protection sociale	100%	