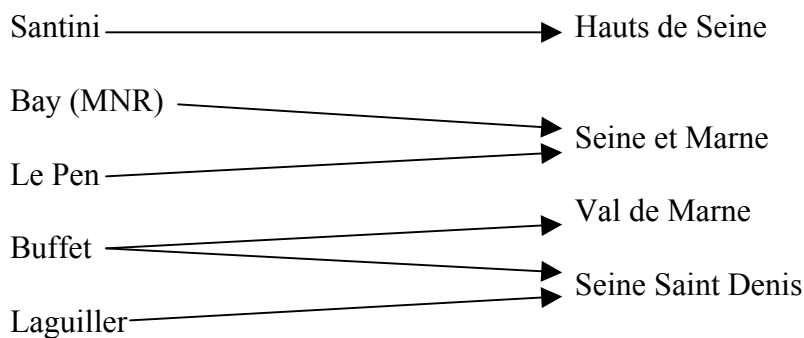


Examen de Statistiques (Corrigé)

A – Descriptions préliminaires.

- 1) Les trois catégories de vote les plus importantes, par ordre décroissant, sont :
Abstentions : 2 487 459 (40,44%) ;
Huchon : 1 170 448 (19,03%) ;
Copé : 908 264 (14,77%).
- 2) Le département où il y a eu le plus d'abstentions est celui de la Seine Saint Denis, avec 45,39% d'abstentions (tableau des pourcentages lignes).
- 3) La liste UDF (Santini) obtient son meilleur score dans les Hauts de Seine : 20,76% (tableau des pourcentages colonnes).
- 4) La liste FN (Le Pen) obtient son meilleur score en Seine et Marne : 16,01% (tableau des pourcentages colonnes).
- 5) Graphe des attractions pour les taux égaux ou supérieurs à 0,300.



B – Première analyse multidimensionnelle.

- 1) Cette première analyse multidimensionnelle est une analyse des correspondances, puisque le tableau des données est un tableau de contingence (tableau d'effectifs), croisant deux variables catégorisées, département (en lignes), candidats (en colonnes).
- 2) On retiendra les axes pour lesquels la valeur propre est supérieure à la valeur du Φ^2 divisé par le nombre d'axes possibles (ici 8), soit $0,02379/8=0.00297$. Ce qui amène à retenir les trois premiers axes, pour 97,5% de la variance totale. Certes on a déjà 83% de variance avec deux axes, mais en regardant les résultats du 3^o axe, on s'aperçoit qu'il y a des éléments qui y contribuent beaucoup. Ainsi Paris, Yvelines et Hauts de Seine pour les départements ; Huchon, Santini et Le Pen pour les candidats. Si on ne retient pas le 3^o axe, Huchon disparaît ainsi que les Yvelines. Enfin, la différence de pourcentage entre les 2^o et 3^o axe n'est pas très grande.

- 3) On va considérer les départements qui ont une forte contribution au premier axe (en l'occurrence supérieure à $1000/8=125$) et on y ajoutera ceux qui ont une bonne qualité de représentation. Le premier axe oppose donc la Seine Saint Denis (à gauche sur l'axe 1) aux trois départements de Paris, des Hauts de Seine et des Yvelines (à droite sur l'axe 1).
Le deuxième axe, sur la base du même critère, oppose la Seine Saint Denis et les Hauts de Seine à la Seine et Marne.
- 4) Dans l'espace des candidats, on retiendra ceux ayant une contribution supérieure à $1000/9=111$.
Le premier axe oppose Buffet et Le Pen à Santini et Copé. Il faudrait ajouter à Buffet et Le Pen, Laguiller et les Abstentions, qui, malgré leur très faible contribution, ont une excellente qualité de représentation sur cet axe.
Le deuxième axe, oppose Buffet et Santini à Le Pen.
- 5) Dans chacun des deux nuages, en considérant les éléments à forte contribution, on trouve un nuage de forme triangulaire, avec trois pôles.
A gauche dans le plan, le département de la Seine Saint Denis et Buffet. Cela correspond à son meilleur score et on retrouve l'attraction Buffet—Seine Saint Denis, constatée à la question A-5.
A droite dans le plan, le département des Hauts de Seine et Santini, à quoi il faut ajouter Paris. Là aussi, c'est dans les Hauts de Seine que Santini fait son meilleur score et c'est une attraction déjà vue à la question A-5.
Enfin, au sommet du triangle, en haut du plan 1-2, on a la Seine et Marne et Le Pen : meilleur score pour Le Pen et attraction déjà vue.

C – Autres analyses multidimensionnelles.

- 1) Une partition en deux classes offre l'inconvénient d'une part de variance inter trop faible (49,8%). En s'aidant de l'arbre représentant la CAH, on optera pour une partition en quatre classes (81,2% de variance inter). Il y a deux classes à un élément.
- 2) Les quatre classes retenues sont : Paris ; Hauts de Seine et Yvelines ; Val de Marne, Essonne, Val d'Oise et Seine et Marne ; Seine Saint Denis.
- 3) Comme pour la classification précédente, une partition en deux classes est insatisfaisante. On retiendra donc une partition en quatre classes (79,9% de variance inter). Là encore, on a deux classes à un élément.
- 4) Les quatre classes sont : Copé et Huchon ; Santini ; Buffet ; Bay (MNR), Le Pen, Laguiller, Pelegrin et les Abstentions.
- 5) Il est possible de faire un lien entre les résultats des classifications et ceux de l'analyse des correspondances. Dans l'espace des candidats on retrouve assez bien les quatre classes dans le plan des deux premiers axes. Dans l'espace des départements, les choses apparaissent moins évidentes, si on se limite au plan des deux premiers axes. En effet, Paris est situé entre les Yvelines et les Hauts de Seine (à droite sur le graphique). On aurait pu se limiter à une partition en trois classes, et dans ce cas Paris n'est plus séparé des deux autres départements. Mais si on regarde le troisième axe, on s'aperçoit que Paris a une très forte contribution à cet axe, ce qui justifierait d'en faire une classe à part. Une interprétation complète prendrait en compte le troisième axe.