Epidémiol. et santé anim., 1991, N°20, 133-138

COMPARAISON DES RESULTATS OBTENUS AVEC DIFFERENTES MISES EN CLASSES SUR UN EXEMPLE D'ANALYSE FACTORIELLE DES CORRESPONDANCES MULTIPLES*

Christine FOURICHON [1-2], *F. MADEC* [2]

RESUME : *Une analyse de sensibilité a été réalisée sur un exemple d'analyse factorielle des correspondances multiples de données épidémiologiques. Les variables quantitatives ont été mises en classes selon six méthodes différentes. Les résultats des analyses sur les variables qualitatives obtenues après transformation sont comparés. Avec des variables dont la distribution s'approche d'une distribution normale, l'analyse factorielle des correspondances multiples paraît être une méthode robuste.*

SUMMARY : *A sensitivity analysis of factorial analysis of correspondence was performed on a set of epidemiological data. Quantitative data were transformed into categorical data with six different methods. The results of the subsequent analysis were compared. With data which distribution is close to normal, factorial analysis of correspondence appears to be robust.*

*
* *

L'analyse factorielle des correspondances multiples est largement utilisée en écopathologie. En effet, elle permet de rechercher l'existence de relations entre variables dans des fichiers complexes. En particulier, elle autorise l'étude simultanée de groupes de variables, et la construction de modèles qui prennent en compte les interrelations, même avec un nombre élevé de variables. Enfin, elle n'impose aucune hypothèse sur la distribution des variables. Néanmoins, cette méthode ne peut être appliquée que sur des variables qualitatives (ou catégoriques) ; aussi, pour inclure des données quantitatives dans une analyse, il est nécessaire au préalable de les transformer par une mise en classes [Benzécri, Fénelon].

---

(1) *Ecole Nationale Vétérinaire, Département des Productions Animales, Laboratoire de Gestion de la Santé Animale, C.P. 3013, 44087 Nantes cedex 03, France.*

(2) *CNEVA, Laboratoire Central de Recherches Avicoles et Porcines, U.R., Station de Pathologie Porcine, B.P. 53, 22440 Ploufragan, France.*

* *Texte de l'exposé présenté lors de la réunion du 27 mars 1991.*

Les procédures de mise en classes habituellement admises reposent sur [Escofier et Pagès] :

1. Le respect de seuils connus antérieurement (ex. : norme d'une analyse biologique),

2. Le respect de structures évidentes dans la population mises en évidence lors de l'étude de la distribution de chaque variable (ex. : distribution bimodale),

3. ou, à défaut, le choix de classes en nombre et effectif équilibrés.

De nombreuses options peuvent donc être retenues pour définir les variables qualitatives analysées. Se pose alors la question de l'effet du mode de mise en classes sur les résultats de l'analyse.

L'objectif de notre étude était de décrire l'influence des méthodes de mise en classes sur les résultats d'une analyse factorielle des correspondances multiples, sur un exemple de données épidémiologiques.

MATERIEL ET METHODES

Dans un suivi de cohorte sur la pathologie respiratoire des porcs charcutiers, une analyse factorielle des correspondances multiples (A.F.C.M.) a pour objectif d'étudier les relations entre la croissance des animaux et la pathologie respiratoire, à l'échelle de l'individu [Fourichon et coll.]. Une analyse de sensibilité de cette A.F.C.M. est réalisée.

I. DONNEES

Une bande de porcs fait l'objet d'un suivi individuel des animaux, de la naissance à l'abattage. Chaque animal est pesé sept fois (naissance, sevrage, 8 semaines, 12 semaines, 16 semaines, 23 semaines, abattage) ; le gain moyen quotidien est calculé pour chaque intervalle successif. Les lésions respiratoires sont notées à l'abattoir selon une grille de notation semi-quantitative. Après élimination des individus présentant des données manquantes, le fichier comporte 163 porcs.

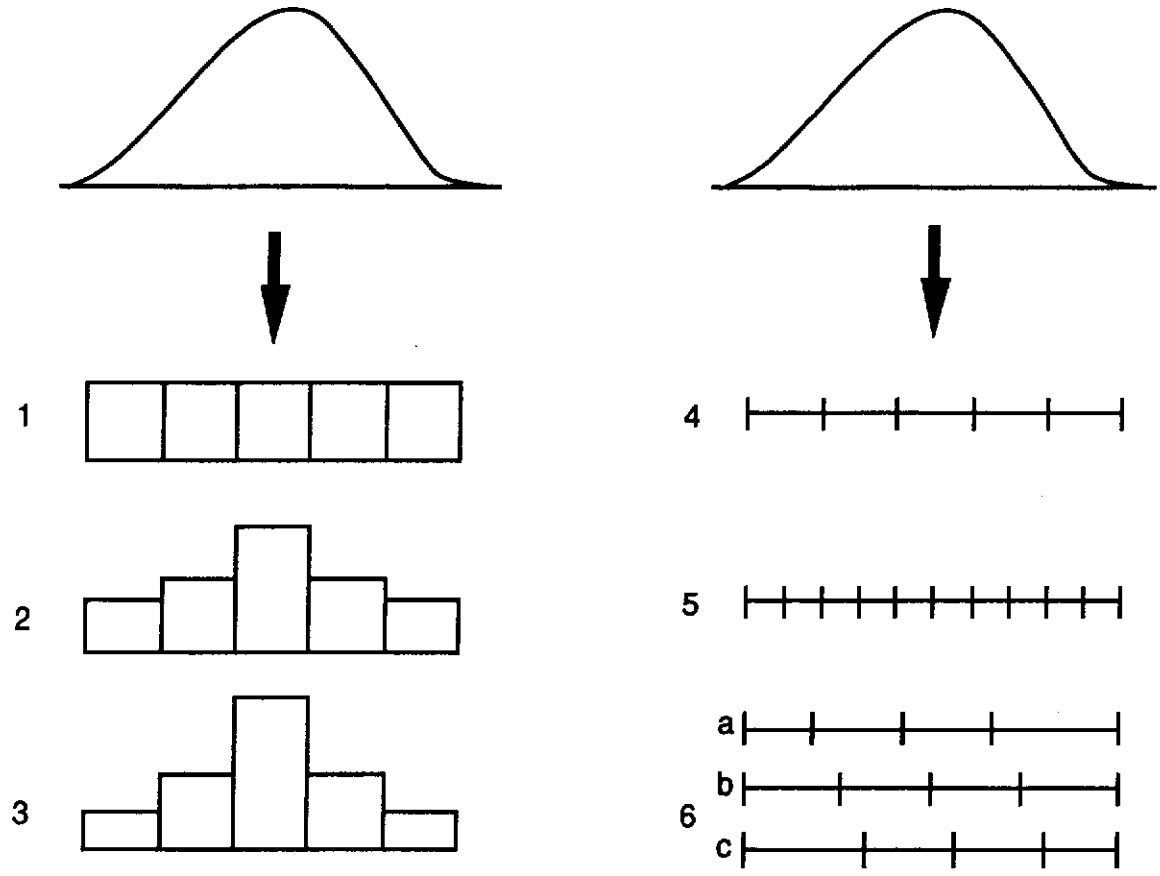
II. MISES EN CLASSES

Pour chaque gain moyen quotidien, six méthodes de mise en classes sont appliquées et définissent 6 jeux de données qualitatives (figure 1).

1. 5 classes d'effectifs égaux (les limites de classes sont les quintiles),
2. 5 classes comportant respectivement 15, 20, 30, 20, 15 % des animaux
3. 5 classes comportant respectivement 10, 20, 40, 20, 10 % des animaux,
4. 5 classes d'amplitude de variation égale,
5. 10 à 13 classes d'amplitude de variation égale et faible,
6. 3 séries de 4 classes se chevauchant, selon le principe des fenêtres glissantes.

   Dans ce jeu de données, chaque individu est donc représenté 3 fois et peut appartenir à des classes différentes s'il est situé à proximité d'une limite entre 2 classes.

Figure 1 : Six méthodes de mise en classes.



Pour les lésions respiratoires, une méthode unique de mise en 4 classes est appliquée au vu de la distribution des variables.

III. ANALYSE DE SENSIBILITE

Une A.F.C.M. est réalisée sur chaque jeu de variables, en incluant les gains moyens quotidiens comme variables actives, et les lésions respiratoires en variables supplémentaires. Pour comparer les résultats de chaque analyse, sont examinés :

- Les pourcentages d'inertie projetée sur chaque axe, rapportés au pourcentage d'inertie projetée attribuable au hasard,

- Les relations identifiées sur les plans factoriels définis par les 4 premiers axes factoriels.

RESULTATS

Les pourcentages d'inertie projetée sur les 4 premiers axes factoriels sont comparables pour les jeux de données 1 à 4 (2 fois l'inertie projetée due au hasard sur l'axe 1), légèrement supérieurs pour la série 5, et 3 fois supérieurs pour la série 6.

Les structures et relations mises en évidence avec les séries 1 à 4 sont globalement comparables. Ainsi, l'axe 1 est un axe de croissance précoce des animaux (de la naissance à 16 semaines). La croissance précoce n'apparaît pas liée aux lésions respiratoires (figure 2). Sur l'axe 2, les modalités moyennes des variables sont opposées aux modalités extrêmes ; l'examen de la répartition des individus sur le plan 1-2 révèle un effet Guttman (figure 3). L'axe 3 représente la croissance tardive des animaux (de 16 semaines à l'abattoir). L'importance des lésions respiratoires apparaît liée à la croissance tardive (figure 4).

La série 5 (petites classes d'égale amplitude) donne des résultats différents. L'axe 1 oppose les modalités très défavorables de croissance précoce aux autres (figure 5). En fait, quelques individus à profil de croissance particulier interviennent de façon prépondérante dans la constitution de cet axe (figure 6). Les autres modalités de croissance précoce sont disposées graduellement le long de l'axe 2, mais il existe des instabilités locales dans la répartition des classes (par exemple SE8 et SE9). Aucun axe ne peut être associé à la croissance tardive des animaux. Il n'est pas possible d'identifier de relation avec l'importance des lésions respiratoires.

L'analyse conduite sur la série 6 donne des résultats identiques aux séries 1 à 4 (distinction de 2 phases de croissance, association de la pathologie respiratoire à la croissance tardive mais pas à la croissance précoce). De plus, il est intéressant de constater que les modalités des trois sous-séries de données sont graduellement disposées le long des axes factoriels (figure 7). Ainsi, le décalage progressif des limites de classes de chaque variable est retrouvé sur les axes factoriels.

DISCUSSION - CONCLUSION

Les différentes méthodes de mise en classes utilisées sur cet exemple conduisent à des résultats globalement comparables. Seul le jeu de données 5 donne lieu à une interprétation différente des structures mises en évidence. La constitution de nombreuses classes d'effectif réduit permet de révéler la présence d'individus à profil particulier, en revanche l'information apportée par ces individus peut masquer les relations existant dans la population prise dans son ensemble.

Avec le type de données analysées ici, le choix d'une des méthodes de mise en classes testées, conformes aux recommandations admises, semble donc peu influencer les résultats. Il faut remarquer que toutes les variables sur lesquelles a été réalisée l'analyse de sensibilité ont des distributions de type normal, plus ou moins symétriques. L'influence du choix des classes avec des variables présentant d'autres types de distributions mériterait d'être étudiée.

Mis à part le cas des classes de petits effectif, l'analyse factorielle des correspondances multiples apparaît être une technique robuste.

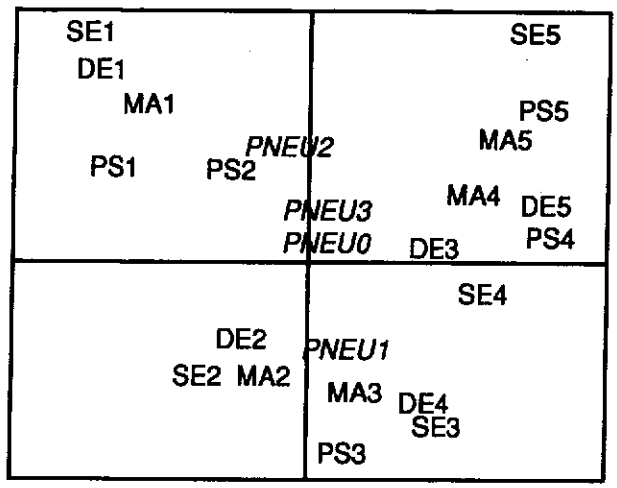


figure 2 : série 2, plan 1-2

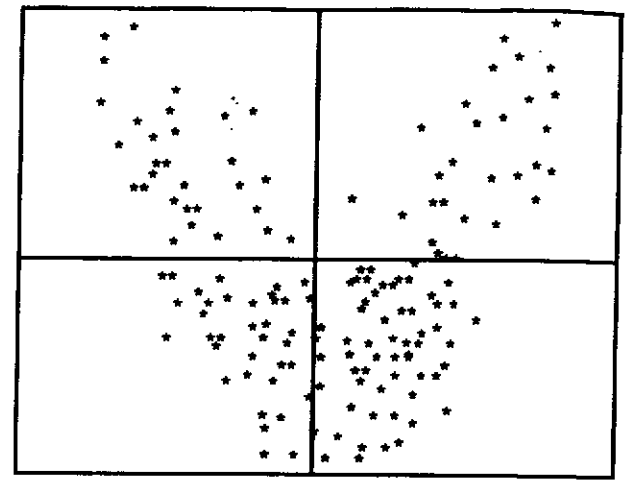


figure 3 : série 2, plan 1-2
position des individus

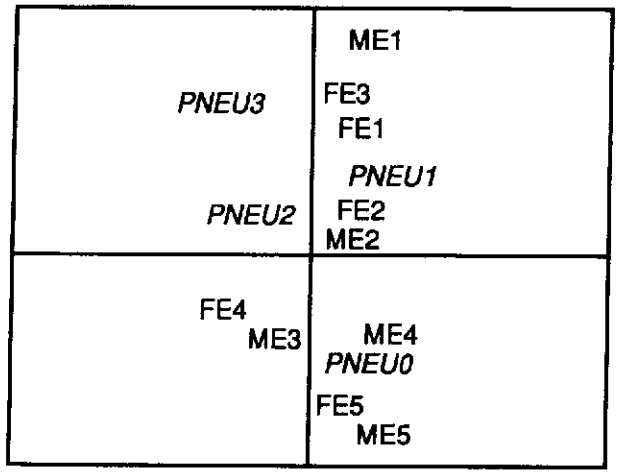


figure 4 : série 2, plan 1-3

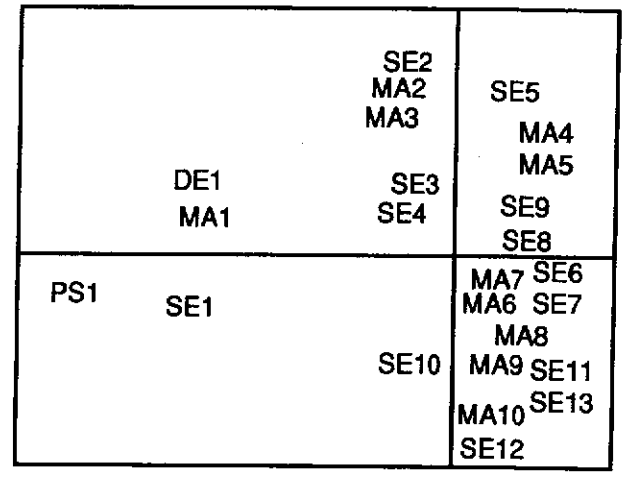


figure 5 : série 5, plan 1-2

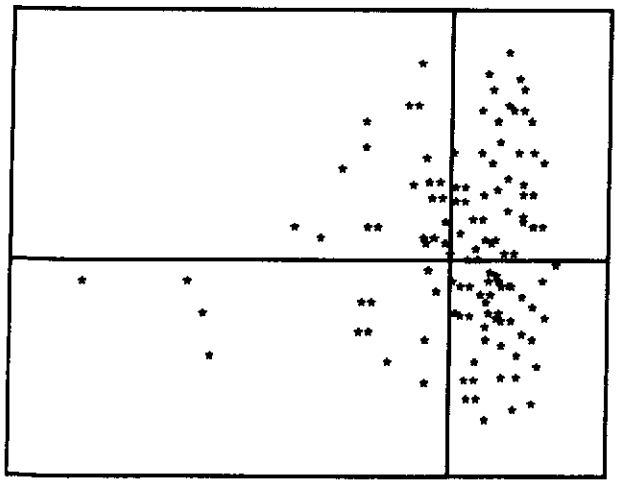


figure 6 : série 5, plan 1-2
position des individus

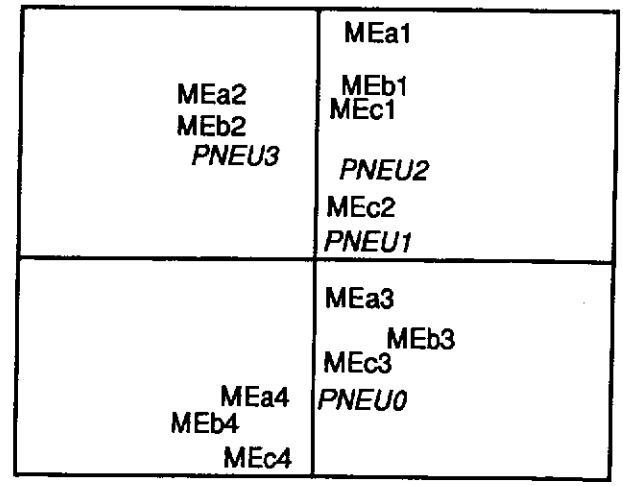


figure 7 : série 6, plan 1-2
(chaque variable est représentée 3 fois a,b,c)

légende figures 2 à 7 :

variables actives : croissances

MA : de la naissance au sevrage
SE : du sevrage à 8 semaines
PS : de 8 à 12 semaines
DE : de 12 à 16 semaines
ME : de 16 à 23 semaines
FE : de 23 semaines à l'abattage

1 = faible
à
5 = élevée

variable supplémentaire : pneumonie

PNEU0 : absence
PNEU1 : discrète
PNEU2 : modérée
PNEU3 : sévère

*** : individu**

REFERENCES BIBLIOGRAPHIQUES

BENZECRI J.P.- L'analyse des données : L'analyse des correspondances. Dunod, Paris, 1979, 616 p.

ESCOFIER B. et PAGES J.- Analyses factorielles simples et multiples, objectifs, méthodes et interprétation. Dunod, Paris, 1988, 241 p.

FENELON J.P.- Qu'est-ce que l'analyse des données ?, Lefonen, Paris, 1981, 311 p.

FOURICHON C., MADEC F., PANSART J.F.and PABOEUF F.- The application of some multivariate statistical methods in epidemiology - an example : a cohort study of respiratory diseases in pigs. In : Thrusfield (Ed.). Proceedings of the Soc. Vet. Epidemiol. Prev. Med., avril 1990, Belfast, 153-166.