L'Analyse Factorielle Discriminante (AFD)

I. Données

L'analyse factorielle discriminante (AFD) est une méthode descriptive et explicative, apparentée à l'analyse en composantes principales (ACP), s'appliquant à des données quantitatives sur lesquelles est déjà définie une typologie ou partition.

Par exemple des indicateurs associés à des clients d'une banque classés comme "bons payeurs", "mauvais payeurs" ou "ayant fait faillite".

II. Méthode

Le but de la méthode, comme en ACP, est de réduire le nombre de dimensions des données, en recherchant celles suivant lesquelles les classes se séparent le mieux. Les directions factorielles discriminantes successives sont déterminées, tandis que des graphiques factoriels plans permettent ici encore de visualiser les individus ou les variables. Divers indicateurs et tests sont également calculés, qui permettent de juger de l'intérêt et de la pertinence des résultats obtenus.

III. Mise en œuvre en SAS

La procédure SAS qui opère l'analyse factorielle discriminante est la procédure CANDISC. Sa syntaxe de base est la suivante :

```
PROC CANDISC / options;
    CLASS groupe;
    VAR variables;
```

Dans l'instruction CLASS, la variable *groupe* est la variable à items définissant la partition, tandis que les *variables* de l'instructions VAR sont les variables quantitatives utilisées.

IV. Détails mathématiques

Il existe plusieurs présentations mathématiques équivalentes de l'AFD, on donne ici l'une d'entre elles.

Ayant N individus, répartis en k groupes, sur lesquels ont été relevées p variables quantitatives, on note :

- y_{ih} , élément de R^p, est le h-ième individu du groupe i
- n_i est l'effectif du groupe i
- $y_{i.}$, élément de R^p, est le centre de gravité u groupe i
- $y_{..}$, élément de R^p, est le centre de gravité de l'ensemble

Les différents vecteurs sont reliés par des relations barycentriques non rappelées ici.

En convenant de noter les vecteurs précédents de R^p en ligne, on définit d'autre part les différentes matrices de variances-covariances :

$$T = \sum_{i,h} (y_{ih} - y_{..})'.(y_{ih} - y_{..}) / N \qquad \text{totale}$$

$$W_i = \sum_{h} (y_{ih} - y_{i.})'.(y_{ih} - y_{i.}) / n_{i.} \qquad \text{du groupe i}$$

$$W = \sum_{i} n_{i.}.W_i / N \qquad \textit{intraclasse}$$

$$B = \sum_{i} n_{i.}.(y_{i.} - y_{..})'.(y_{i.} - y_{..}) / N \qquad \textit{interclasse}$$

et on a l'égalité matricielle :

$$T = W + B$$

relation déjà rencontrée en classification automatique, avec des notations légèrement différentes toutefois.

On munit enfin l'espace R^p des observations de la métrique T^{-1}, dite *métrique de Mahalanobis*, pour laquelle le nuage total est *isotrope* (ou de même inertie dans toutes les directions).

L'inertie totale dans une direction de vecteur u, T^{-1}-unitaire, de R^p (noté en colonne) est en effet :

$$I_u = u'.T^{-1}.T.T^{-1}.u = 1$$

et se décompose en :

$$I_u = u'.T^{-1}.W.T^{-1}.u + u'.T^{-1}.B.T^{-1}.u$$

Ou, en notant a le vecteur T-unitaire $T^{-1}.u$:

$$I_u = a'.W.a + a'.B.a$$

Le pouvoir discriminant de la direction u sera d'autant meilleur que l'inertie interclasse a'.B.a est grande, ou ce qui revient au même que l'inertie intraclasse a'.W.a est faible.

On est donc conduit au problème classique d'optimisation sous contrainte :

Max a'.B.a avec a'.T.a = 1

qui a pour solution les vecteurs propres associés aux valeurs propres décroissantes successives λ de la matrice $T^{-1}.B$. Les a_λ sont T-orthogonaux et les u_λ correspondants T^{-1}-orthogonaux.

Pouvant être considérée comme une ACP du nuage des k centres de gravité pour une métrique particulière, l'AFD ne peut obtenir plus de k-1 directions discriminantes, et donne par ailleurs lieu aux diverses représentations et indices communs en ACP.

----===ooOoo===----

(03.07.2009)