

## L'Analyse Discriminante (AD)

L'analyse discriminante, à ne pas confondre avec la classification automatique, traite des données déjà classées en différents groupes, et vise à produire une *fonction discriminante* permettant d'affecter une nouvelle observation à l'un de ces groupes.

C'est une méthode *décisionnelle*, et non seulement descriptive, telles l'ACP ou l'AFC.

### I. Données

Les observations de base sont des variables quantitatives relevées sur des objets ou individus répartis en groupes déjà connus.

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

### II. Méthode

Supposant éventuellement des *probabilités a priori* données sur les différents groupes, l'analyse discriminante calcule les probabilités *a posteriori* qu'une nouvelle observation appartienne à chacun de ces groupes (fonction discriminante) et attribue cette observation au groupe le plus probable.

Les variantes de la méthodes tiennent à la manière dont est définie et calculée cette probabilité.

#### Méthodes paramétriques

Les méthodes *paramétriques* supposent que les observations des différents groupes suivent des distributions normales multidimensionnelles.

Il y a deux cas, selon que la dispersion (mesurée par les variances-covariances) est supposée la même dans les différents groupes ou pas.

#### Méthodes non paramétriques

Les méthodes *non-paramétriques* estiment la densité associée à chaque groupe par des techniques utilisant la notion de *noyau* (densité élémentaire de forme

particulière dans un domaine borné construite autour de chaque point du groupe).

### **Autres méthodes**

D'autres méthodes, mathématiquement moins complexes, s'appuient sur des comptages de *plus proches voisins*...

### **Test**

On éprouve la qualité de la fonction discriminante obtenue en la testant sur un échantillon indépendant des observations de base, mais pour lequel on connaît le groupe d'appartenance des différentes observations. On mesure ainsi *le taux de bien classés*.

Il est clair qu'opérer ce test sur l'échantillon de base lui-même produirait un biais en faveur de la méthode...

## **IV. Mise en œuvre en SAS**

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

```
PROC DISCRIM / options;
  CLASS groupe;
  PRIORS [EQUAL/PROP/probabilités];
  VAR variables;
```

Parmi les options de la procédure : METHOD = NORMAL, choix par défaut, ou NONPAR. Dans le cas paramétrique (NORMAL), les options POOL = NO, TEST ou YES sont claires. On lira dans la documentation les options du cas non paramétrique.

Dans l'instruction CLASS, la variable *groupe* est la variable à items à discriminer.

L'instruction PRIORS EQUAL donne des probabilités a priori égales, c'est le choix par défaut, et PROP proportionnelles aux effectifs, sinon les probabilités des différents items sont indiquées d'une manière comparable à celle-ci :

```
PRIORS A = 0.5 B = 0.2 C = 0.3;
```

L'instruction VAR enfin est suivie de la liste des variables quantitatives discriminantes à utiliser.

Exemple

```
PROC DISCRIM data = banque;  
  CLASS payeur;  
  PRIORS B = 0.7 M = 0.25 F = 0.05;  
  VAR x y z w;
```

-----ooOoo-----

(22.05.2008)