**TER : Scoring**

**II- Analyse discriminante**

1. Théorie de la méthode probabiliste

*Introduction de remplacement :*

Le but premier de cette méthode est de prédire au mieux les valeurs d’une variable Y qualitative à K modalités, à partir de p variables explicatives X = (X1 , . . . , Xp ) quantitatives.

Dans cette section, nous allons définir des règles de décision bayésiennes qui vont permettre d’affecter un nouvel individu à la classe “la plus probable” et non pas au groupe « le plus proche » comme c’est le cas pour l’analyse discriminante géométrique. Pour cela, il est nécessaire de faire des hypothèses probabilistes sur les données, d’où le nom de la méthode.

5.4.2 Sélection des variables

Le Lambda de Wilks est souvent utilisé dans les logiciels comme critère pour ne garder que les variables apportant de l’information sur l’appartenance ou non d’un individu à un groupe.

Le Lambda de Wilks est une approche paramétrique permettant de tester si plusieurs variables continues distinctes  sont liées à une variable qualitative Y à K ≥ 2 groupes, lorsqu’elles sont considérées avec leurs différentes interactions multivariées.

Les hypothèses d’utilisation de ce test sont:  suivent une loi normale et leur matrice de covariance respective sont égales (homoscédasticité).

La statistique du test du Lambda de Wilks se définie de la manière suivante :

Où W est la matrice de variance-covariance intragroupe et B la matrice de variance-covariance intergroupe.

Cette statistique de test suit une loi de Wilks à  degrés de liberté et l’hypothèse  est : « Indépendance entre  et  ».

Une variable a un bon pouvoir discriminant si la dispersion intra-groupe est faible et si la dispersion intergroupe est forte. Donc plus le Lambda de Wilks sera faible, plus la variable considérée est discriminante.

C’est ce critère qu’utilise la commande « greedy.wilks » de Rstudio que nous avons utilisée pour trouver les variables les plus discriminantes dans notre jeu de données et ainsi se focaliser sur un nombre de modèles plus réduit.

1. ACP préliminaire (présentation du jeu de données)
2. Exemple farms

Code sur git. Peut-être organisé la rédaction en suivant la méthode de construction d’un score de Mme Chavent ? (Pages 1 et 2)