Utilisation des modèles à équations structurelles en analyse sensorielle

Michel Tenenhaus
HEC School of Management (GRECHEC),
1 rue de la Libération, Jouy-en-Josas, France
tenenhaus@hec.fr

Résumé

Deux écoles concurrentes se sont imposées dans le domaine de la modélisation des équations structurelles. La première approche, appelée « Covariance-based SEM » (*SEM = Structural Equation Modeling*), s'est développée autour de Karl Jöreskog. Cette approche a pour objectif de modéliser la matrice de covariance des variables observées. Elle peut être considérée comme une généralisation de l'analyse factorielle en facteurs communs et spécifiques au cas de plusieurs tableaux de données reliés entre eux par des liens de causalité. La seconde approche s'est développée autour de Herman Wold sous le nom de « PLS » (*Partial Least Squares*), puis de « Component-based SEM ». Elle a pour objectif la construction de scores résumant au mieux chaque bloc de variables tout en tenant compte du réseau de causalité. C'est une généralisation de l'analyse en composantes principales au cas de plusieurs tableaux de données reliés entre eux par des liens de causalité. Plus récemment Hwang et Takane (2004) ont proposé une méthode « Component-based SEM » : la méthode GSCA (*Generalized Structural Component Analysis*). Elle permet une recherche de scores optimisant un critère global. Nous allons discuter de l'utilisation de ces trois approches en donnant des exemples d'utilisation en analyse sensorielle.

Mots-clés : PLS, ULS, SEM, GSCA.

1. Utilisation de la méthode d'estimation ULS en modélisation d'équations structurelles

Pour une présentation détaillée des modèles à équations structurelles, on peut consulter Bollen (1989). Un modèle à équations structurelles est formé de deux modèles : le modèle structurel et le modèle de mesure.

Le modèle structurel

On considère un ensemble de variables latentes centrées (variables inobservables) reliées entre elles par des liens de causalité. Notons $\boldsymbol{\eta}$ un vecteur colonne formé de *m* variables latentes endogènes (c'est à dire expliquées par d'autres variables latentes) et $\boldsymbol{\xi}$ un vecteur colonne formé de *k* variables latentes exogènes (c'est à dire purement explicatives). Le modèle structurel reliant le vecteur $\boldsymbol{\eta}$ aux vecteurs $\boldsymbol{\eta}$ et $\boldsymbol{\xi}$ s'écrit

(1) $$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}$$

où $\mathbf{B}$ est une matrice $m \times m$ de coefficients de régression à diagonale nulle, $\boldsymbol{\Gamma}$ une matrice $m \times k$ de coefficients de régression et $\boldsymbol{\zeta}$ un vecteur aléatoire de dimension *m*. On suppose que la matrice **I-B** est inversible.

Le modèle de mesure

Chaque variable latente (inobservable) est décrite par un ensemble de variables manifestes (observables) supposées centrées. Le vecteur $\mathbf{y}_j = (y_{j1},...,y_{jp_j})'$ des variables manifestes reliées à la variable latente endogène η_j peut s'écrire en fonction de η_j : $\mathbf{y}_j = \boldsymbol{\lambda}_j^y \eta_j + \boldsymbol{\varepsilon}_j$. Le vecteur $\mathbf{y} = (\mathbf{y}_1',...,\mathbf{y}_m')'$ s'écrit

(2) $$\mathbf{y} = \boldsymbol{\Lambda}_\mathbf{y}\boldsymbol{\eta} + \boldsymbol{\varepsilon}$$