ANALYSE DE LA COVARIANCE (J3, L)
(31 / 01 / 2020, © Monfort, Dicostat2005, 2005-2020)

L'**analyse de la covariance** a pour objet l'étude d'une **variable quantitative** en fonction de variables dont certaines sont des **variables qualitatives** (eg des **variables indicatrices**) et d'autres des variables quantitatives. Elle généralise donc, en un sens, l'**analyse de la variance** classique et elle étudie notamment la **matrice des covariances** reliant des **va**.

Elle constitue un exemple de **relation fonctionnelle** déduite d'une **loi multivariée**.

Outil de base de ce type d'analyse, le **modèle d'analyse de la covariance** est un **modèle mixte** qui combine, dans la liste des **variables exogènes**, des variables non numériques, mais codées (variables de classification, facteurs expérimentaux, etc) (cf **codage**), et des variables numériques.

(i) En général, on appelle **modèle d'analyse de la covariance** un **modèle linéaire** qui se présente sous la forme suivante dans l'**espace des observations** :

(1) $y = X b + Z c + u,$ avec $E u = 0,\ V u = \sigma^2 . I_N$,

forme dans laquelle X est une (N,K)-**matrice** dont les éléments $x_{nk} \in \{0,1\}$ représentent des variables indicatrices (va de classification d'**unités statistiques** selon des catégories données, ou selon les niveaux des **traitements** qui leurs sont appliqués, etc) et Z est une (N,L)-matrice dont les éléments $z_{nl} \in \mathbf{R}$ représentent des observations relatives à d'autres variables, numériques, supposées influer sur y.

(ii) Lorsque c est supposé connu, b peut être estimé par la **méthode des mco**, à l'aide d'un estimateur linéaire (ou affine) tq $b^\wedge (c) = B (y - Zc)$, où B est une (K,N)-matrice dépendant de X (lorsque rg X = K, alors $B = (X' X)^{-1} X'$) (cf **singularité**). Par suite, l'**estimateur des mco** de (b, c) est l'estimateur des mco du nouveau modèle linéaire (ou modèle transformé) :

(2) $(I_N - X B) y = (I_N - X B) Z c + u,$ avec $E u = 0, V u = \sigma^2 . I_N$.

Ainsi, lorsque rg $(I_N - X B) y = L$, $c^\wedge = (Z_\#' Z_\#)^{-1} Z_\#' y_\#$ est l'estimateur des mco du modèle (2), avec :

$$V c^\wedge = \sigma^2 (Z_\#' Z_\#)^{-1} ,$$

(3) $$Z_\# = (I_N - X B) Z,$$

$$y_\# = (I_N - X B) y.$$

On définit ainsi un procédé itératif de calcul d'un estimateur de (b, c).

(iii) L'analyse de la covariance suit une démarche semblable à celle de l'**analyse de la variance**.

En particulier, une **équation d'analyse de la covariance** peut s'en déduire, permettant l'application de certaines **procédures statistiques** (**tests d'hypothèses**, notamment) : ce type d'équations décompose la **variabilité** de la variable endogène en fonction de celles des variables exogènes (aussi bien numériques que qualitatives).

Une difficulté courante vient de la **singularité** de la **matrice des observations** des variables exogènes : ceci entraîne un défaut d'**estimabilité** de certains paramètres et nécessite l'imposition de contraintes sur ces derniers.

(iv) Il existe des versions multidimensionnelles de ce type de modèle, versions dont la **variable endogène** est vectorielle (cf **vecteur aléatoire**).