

# UNE NOUVELLE APPROCHE DANS LE CHOIX DES RÉGRESSEURS DE LA RÉGRESSION MULTIPLE EN PRESENCE D'INTERACTIONS ET DE COLINEARITES

Michel Lesty

*Coryent Conseil, 28, rue Sainte Adélaïde 78000 Versailles France, e-mail coryent@freenet.fr*

## Résumé

A partir d'exemples simples, on présente une méthode originale de choix des variables et interactions d'ordre 2 dans un modèle de régression multiple, en présence de multicollinéarités ou d'interactions. La méthode "CORICO" (Iconographie des Corrélations) est fondée sur les corrélations totales et partielles. La détection des points aberrants est obtenue au moyen de variables indicatrices des observations. Nous montrons comment l'introduction de fonctions logiques non-linéaires d'ordre 2 améliore l'interprétation du plan d'expérience.

## Abstract

On the basis of simple examples, we present an original method of choice of variables and interactions in a multiple regression model, in the presence of multicollinearities or interactions. CORICO method is based on total and partial correlations. We display how this practice improves the interpretation of design experiment.

*Mots-clés : Régression multiple, interactions, corrélations partielles, méthode CORICO, plan d'expérience, prédiction.*

## 1 - Introduction

L'instabilité des coefficients de régressions issus des données d'une expérience non parfaitement planifiée est à l'origine de la majorité des difficultés d'interprétation. Le problème du choix des régresseurs se pose avec une acuité d'autant plus grande que les régresseurs ne sont pas orthogonaux (corrélation nulle).

Une première approche, qui est celle de l'Analyse en Composantes Principales consiste, par une transformation de variables, à retrouver des propriétés d'orthogonalité, absentes à l'origine. Une fois la régression orthogonalisée, on retourne aux régresseurs initiaux tout en maîtrisant mieux les sources d'instabilité. Cette technique ainsi que celle de la régression pseudo-orthogonalisée a fait l'objet d'une abondante littérature (cf. Tomassone, Lesquoy et Millier [1983]).

Une seconde approche, celle de la régression PLS (Partial Least Square), permet de relier un ensemble de variables dépendantes  $Y$  à un ensemble de variables indépendantes  $X$ . On effectue les analyses en composantes principales des ensembles  $X$  et  $Y$  sous la contrainte que les composantes principales des  $X_j$  soient aussi corrélées que possible à celles des  $Y_k$ . Les corrélations entre variables prédites et les variables initiales sont plus faibles qu'avec le modèle de régression linéaire classique, mais l'on peut obtenir, si les  $Y_k$  sont bien choisis, un modèle plus robuste sur un nouvel échantillon (voir Tenenhaus M., Gauchi J.P. et Ménardo C. [1995]).