

# Analyse des leviers: effet de colinéarité et hiérarchisation des impacts, recherche des structures causales dans les études de marché et sociales

## Mots clés :

- **Directeur de thèse** : Gilbert Saporta
- **Co-encadrant(s)** :
- **Unité de recherche** : Centre d'Étude et de Recherche en Informatique et Communications
- **Ecole doctorale** : École Doctorale Informatique, Télécommunications, Électronique de Paris
- **Domaine scientifique principal**: Divers

## Résumé du projet de recherche (Langue 1)

Les études de marché et les études sociales font appel à des modèles statistiques afin de déterminer quelles actions doivent être mises en œuvre en priorité pour obtenir un résultat donné, comme par exemple augmenter la fidélité ou la satisfaction de clients et aussi pour simuler un impact attendu en agissant sur ces leviers. Ainsi des praticiens d'études de marché vont par exemple utiliser la régression linéaire pour tenter d'attribuer une importance relative à des leviers d'action. Cependant les corrélations entre prédicteurs créent des difficultés : instabilité dans les coefficients mesurés entre des enquêtes comparables successives, ou encore apparition de coefficients négatifs ce qui peut être contre-intuitif. Cette situation (appelée multicollinéarité) est fréquemment rencontrée dans les études par enquêtes quand des questions voisines (par exemple sur l'attrait d'un produit ou sur une pratique sociale) sont formulées dans les questionnaires présentés aux répondants. Les variables utilisées pour la modélisation construites à partir des réponses obtenues sont alors corrélées. La recherche proposée s'intéressera successivement au modèle direct sans approche structurelle puis aux approches avec modèles structurels. Dans le cas d'un modèle direct, cet effet induit par la multi-collinéarité a fait l'objet de nombreux travaux et publications. Ce sujet a évolué depuis les années 1980 en raison de l'accroissement des capacités de calcul informatique, et sont apparues différentes méthodes fondées sur une décomposition de la variance expliquée (décomposition du  $R^2$  entre les prédicteurs : cf en référence les méthodes dérivées de la théorie des jeux comme la valeur de Shapley, ou la décomposition proposée par Johnson), en vue de générer des importances relatives toutes positives et plus stables en cas de renouvellement de l'enquête ou de la mesure. La première partie de la recherche portera sur la mise en œuvre de la décomposition de la variance dans ce domaine et sur une analyse critique de la validité de ces approches pour l'analyse des leviers. Il s'agira de vérifier si la décomposition de la variance est une réponse adaptée à la situation créée par la corrélation entre prédicteurs de proposer éventuellement des recommandations sur la manière de traiter la corrélation entre prédicteurs. On étudiera dans quelle mesure les techniques de régularisation (ridge, PLS, lasso etc) peuvent constituer un prétraitement utile de la multicollinéarité pour l'étude des leviers. Dans une deuxième partie, la recherche portera sur l'analyse structurelle, s'intéressant d'une part à l'utilisation et aux conditions de validation de modèles conceptuels prédéfinis incluant des hypothèses cause-effet, et aux moyens de choisir un modèle particulier, et d'autre part envisagera l'apport potentiel de l'analyse structurelle causale (cf en référence Pearl) pour tenter de déterminer directement partir des données des structures causales. La recherche visera à proposer des moyens de sélectionner un modèle parmi les nombreuses structures causales possibles, de s'assurer de la stabilité du choix et de définir des méthodes de calcul des impacts des leviers en tenant compte de la structure identifiée. L'approche envisagée consistera à combiner bootstrap et algorithmes de recherche causale. En effet le thème même de l'analyse des leviers est essentiellement lié à une perspective de causalité. Il paraît donc pertinent de rechercher si les travaux des 15 dernières années dans la recherche sur la causalité peuvent être appliqués au cas des études de marché et dépasser l'analyse des simples associations et corrélations.