

# Sur une analyse des différences entre les scores PISA obtenus en mathématiques par les filles et les garçons.

Valérie Girardin

Laboratoire de Mathématiques N. Oresme, UMR6139, Campus II,  
Université de Caen Normandie, BP5186, 14032 Caen, France

Justine Lequesne

Centre Henri Becquerel, Unité de Recherche Clinique,  
rue d'Amiens, CS 11517, 76038 Rouen Cedex 1, France

Nous utilisons les tests associés au modèle log-linéaire avec la paramétrisation de Zighera pour analyser les différences de score obtenus par les garçons et les filles en fonction des différentes années d'enquête, du statut socio-professionnel du père et du statut migratoire des parents. Ce modèle nous permet de considérer l'ensemble des relations entre ces variables pouvant expliquer les différences observées. La distinction se fait d'une enquête à l'autre entre ce qui relève de la composition des échantillons ou de changements dans la composition de la population et des changements effectifs dans les relations entre les caractéristiques étudiées et les résultats des élèves.

De nombreuses analyses de résultats d'enquêtes PISA ont été faites, notamment pour étudier l'évolution des scores des élèves sur plusieurs années ; voir par exemple Ammermüller (2008), Barrera *et al.* (2011), Meunier (2014), Goussé et Le Donné (2007). Pour cela, des méthodes de décomposition de différentiels de scores sont utilisées, comme celles d'Oaxaca-Blinder (1973) ou de Juhn, Murphy et Pierce (1993). Ces analyses s'inscrivent dans la lignée de recherches sur le lien entre les performances scolaires et plusieurs facteurs, tels que le genre, l'origine migratoire, ou l'efficacité des établissements. Cependant, ces techniques de décomposition ne permettent pas d'évaluer d'éventuelles relations entre le score des élèves et plusieurs facteurs simultanément.

L'utilisation de modèles log-linéaires est fréquente pour décomposer les liaisons existant entre des variables socio-démographiques. Elle est néanmoins contrainte par le nombre d'interactions, qui croît de façon multiplicative avec le nombre de variables considérées, ce qui complexifie fortement le traitement. Une nouvelle paramétrisation des modèles log-linéaires, dite de Zighera, initiée par Zighera (1985) et développée dans Lequesne (2015) et Girardin, Lequesne et Ricordeau (2017), ne requiert qu'une procédure d'estimation et de tests unique et rapide pour identifier un modèle parcimonieux, ne contenant que les variables

les plus significatives, et ceci indépendamment du nombre de variables étudiées. Les paramètres estimés permettent en outre de décrire de manière précise les associations existant entre les variables, selon des relations qui peuvent être causales ou non, en termes d'information et d'odds-ratios. Plusieurs usages de ce modèle peuvent être envisagés, pour analyser des évolutions de population, identifier des biais d'enquêtes ou réaliser des comparaisons internationales.

Nous appliquons ici cette paramétrisation pour analyser les résultats des acquis scolaires en compréhension des mathématiques des garçons et des filles, tels que mesurés par les enquêtes PISA de l'OCDE entre 2000 et 2012. Cette analyse fait suite à Girardin et Lequesne (2017), centrée sur les différences d'observation entre les différentes années d'enquêtes, d'une part en France et d'autre part en Allemagne. Les résultats de compréhension de l'écrit sont également analysés dans Lequesne (2015).

### **Bibliographie :**

Ammermüller, A. (2008) PISA : What Makes the Difference? *Empirical Economics*, V.33(2), pp.263-287.

Barrera-Osorio, F., Garcia-Moreno, V. Patrinos, H.A. & Porta, E. (2011) Using the Oaxaca-Blinder decomposition technique to analyze learning outcomes changes over time : an application to Indonesia's results in PISA Mathematics. *World Bank Policy Research Working Paper Series*, 5584.

Blinder, A. (1973) Wage discrimination : reduced form and structural estimates, *The Journal of Human Resources*, V.8(4), pp. 436-455.

Girardin, V., Lequesne, J. (2017) How variation of scores of the Program for International Student Assessment can be explained through analysis of information. *Stochastic and Data Analysis Methods and Applications in Statistics and Demography* Ed. : J. Bozeman, T. Oliveira, S. Silvestrov, C. Skiadas à paraître.

Girardin, V., Lequesne, J. and Ricordeau (2017) An information-based parameterization of the log-linear model for categorical data analysis *en révision* pour *Methodology and Computing in Applied Probability*.

Goussé, M. & Le Donne N. (2014) Why do inequalities in 15-year-old cognitive skills increase in France between 2000 and 2009? *AMCIS Conference on Educational Systems : Inequalities, Labour Markets and Civic engagement*.

Juhn, C., Murphy, K & Pierce, B. Wage (1993) Inequality and the rise in returns to skill, *The Journal of Political Economy*, V.101(3), pp. 410-442.

Lequesne, J. (2015) Tests statistiques basés sur la théorie de l'information, applications en biologie et en démographie. *Thèse de doctorat*, Université de Normandie, Caen, France.

Meunier, M. (2007) Origine migratoire et performance scolaire en lecture : analyse des scores PISA 2000 *Rapport de recherche Université de Genève*.

Oaxaca, R. (1973) Male-female wage differentials in urban labor markets, *International Economic Review*, V.14(3), pp. 693-709.

Zigheera, J. A. (1985) Partitioning information in a multidimensional contingency table and centring of loglinear parameters *Applied Stochastic Models and Data Analysis*, V.1, p 93-108.