

Fionn Murtagh, "Correspondence analysis and data coding with Java and R", Computer Science and Data Analysis Series, London, Chapman and Hall/CRC, 2005

Fionn Murtagh, "Correspondence analysis and data coding with Java and R", Computer Science and Data Analysis Series, London, Chapman and Hall/CRC, 2005

Henry Rouanet et Brigitte Le Roux

**Édition électronique**

URL : <http://journals.openedition.org/msh/9173>

ISSN : 1950-6821

Éditeur

Centre d'analyse et de mathématique sociales de l'EHESS

Édition imprimée

Date de publication : 20 mars 2008

Pagination : 113

ISSN : 0987-6936

Référence électronique

Henry Rouanet et Brigitte Le Roux, « Fionn Murtagh, "Correspondence analysis and data coding with Java and R", Computer Science and Data Analysis Series, London, Chapman and Hall/CRC, 2005 », *Mathématiques et sciences humaines* [En ligne], 181 | Printemps 2008, mis en ligne le 27 mars 2008, consulté le 01 mai 2019. URL : <http://journals.openedition.org/msh/9173>

ANALYSE BIBLIOGRAPHIQUE

Fionn Murtagh, “Correspondence analysis and data coding with Java and R”, *Computer Science and Data Analysis Series*, London, Chapman & Hall/CRC, 2005.

Voici un nouvel ouvrage en langue anglaise sur l’analyse des données, rédigé par un auteur qui, depuis sa thèse avec Jean-Paul Benzécri (soutenue en 1981) est resté proche de celui-ci.

Le chapitre 1, dans la ligne de l’ouvrage de Benzécri *Histoire et préhistoire de l’analyse des données* (Dunod, 1962), replace dans une perspective historique les développements qui ont conduit à l’analyse des correspondances et aux méthodes d’analyse géométrique des données.

Le chapitre 2, rappelle les formules de base de l’analyse des correspondances et de la classification hiérarchique.

Les chapitres 3 et 4 (études de cas) sont centrés sur le domaine d’excellence de Murtagh : le *codage des données*, étape indispensable et cruciale, qui (paraphrasant une formule de Murtagh) permet au statisticien d’être un partenaire à part entière de l’entreprise de recherche, aux antipodes d’un aveugle *data mining*. Changement d’échelle, dédoublement, codage flou, codage barycentrique, codage disjonctif (le format quasi-universel du questionnaire) sont passés en revue et présentés en détail. Le codage flou, solution efficace à l’épineux problème des valeurs extrêmes, fournit à l’auteur l’occasion d’une ouverture sur la théorie des possibilités, à propos d’une analyse de données astronomiques.

Le chapitre 5 est spécifiquement consacré à la *lexicologie*, domaine d’intérêt central de Benzécri, où le codage des données, en vue de l’analyse de croisements Textes \times Termes, est à l’évidence crucial. Dans ce chapitre, on trouvera d’abord une judicieuse sélection d’articles commentés des *Cahiers d’analyse des données* – cette fabuleuse mine, encore si peu exploitée, d’analyses effectuées tantôt à partir des mots-outils, qui identifient avant tout le genre littéraire, tantôt à partir des mots pleins, qui nous informent sur le contenu. On retrouvera notamment, dans ce chapitre, la fameuse controverse sur le *Don paisible*, dont la paternité avait été contestée à Cholokhov. La méthode est appliquée ensuite par l’auteur à des textes en anglais, visant à souligner l’actualité de l’analyse des données, qui aux yeux de l’auteur reste insurpassée.

Ce livre est bienvenu, de lecture agréable, accompagné de programmes (Java et R) rédigés par un spécialiste de l’algorithmique (noter le site www.correspondance.info, avec le «[A](#)» dans «[Correspondances](#)») il va assurément contribuer à accroître la diffusion des méthodes d’analyse géométrique des données dans le monde, y compris en France où elles virent le jour...

H. Rouanet et B. Le Roux