

Note de lecture rédigée par Robin GENUER¹

ANALYSE DE DONNÉES AVEC R

François HUSSON, Sébastien LÊ et Jérôme PAGÈS

Livre (224 pages)

Édition : Presses Universitaires de Rennes – Collection Pratique de la statistique – 2009

L'*Analyse de données avec R* est un livre autonome concernant les principales méthodes d'analyse de données. « Autonome » car, en plus de décrire précisément les méthodes abordées, le livre contient de nombreux exemples d'applications de ces méthodes ainsi que toutes les commandes R nécessaires. Il est destiné aux étudiants, enseignants et aux utilisateurs souhaitant comprendre et mettre en œuvre les méthodes classiques d'analyse de données. Les auteurs s'appuient sur leur expérience d'enseignement à l'Agrocampus de Rennes. Il en résulte un exposé très pédagogique du sujet. De plus, ce livre peut être un très bon support de cours (particulièrement adapté pour des filières où les étudiants n'ont pas forcément de bases mathématiques très poussées).

Le livre est composé de quatre chapitres. Chacun des trois premiers chapitres est dédié à la description et l'utilisation pratique d'une méthode classique d'analyse factorielle (ACP, AFC, ACM). Leur exposé est très clair, et on comprend bien les points communs et les différences entre ces méthodes. Le dernier chapitre décrit deux méthodes classiques de classification (aussi appelée classification non-supervisée) : la classification hiérarchique ascendante (CAH) et l'agrégation autour des centres mobiles (k-means). L'application de ces méthodes est également illustrée sur des jeux de données réelles, avant de terminer par une « double-analyse » des données précédemment utilisées (à la fois par une analyse factorielle et une méthode de classification). Cet ouvrage finit par une annexe présentant en quelques pages le logiciel R, où l'essentiel des commandes de base pour prendre en main le logiciel sont présentées succinctement. Ceci est très utile pour un lecteur qui ne connaîtrait pas ou très peu R, et rajoute au caractère autonome du livre.

Un des points forts du livre est de réussir à donner toutes les informations nécessaires à la compréhension intuitive des méthodes tout en cachant les détails trop techniques (notamment pour les méthodes factorielles). Toutes les formules mathématiques utiles sont néanmoins présentes et commentées (notons toutefois que le lecteur désirant avoir les démonstrations en détail devra chercher l'information ailleurs). La lecture du livre est ainsi très agréable, surtout que beaucoup d'idées intuitives sont illustrées par des graphiques ou des commentaires très bien adaptés. Notons également que pour chaque méthode, un exemple est traité en même temps que la description de la méthodologie.

La mise en application des méthodes (qui peut être le but principal recherché par un lecteur de ce livre) est bien détaillée. Toutes les commandes sont clairement données et l'interprétation des résultats est très fournie. En plus de l'exemple qui sert de fil conducteur au

¹ Maître de Conférences, ISPED, Université Bordeaux Segalen, INSERM U897, Robin.Genuer@isped.u-bordeaux2.fr

Note de lecture : « Analyse de données avec R » (Fr. Husson et al., 2009)

chapitre abordé, plusieurs autres exemples sont ensuite traités de façon très détaillée. Les jeux de données proviennent de domaines d'application divers et originaux. De plus, tous les jeux de données ainsi que les lignes de commande utilisés dans le livre sont disponibles sur le site internet du livre. Les analyses sont effectuées à l'aide du paquet R, FactoMineR, développé par les auteurs en collaboration avec Julie Josse. Ce paquet permet d'obtenir directement les représentations graphiques utiles aux analyses, ce qui rend son utilisation très accessible et très agréable.

Enfin, un dernier avantage du livre est son prix qui est très raisonnable. Cela peut permettre à tout étudiant (ou utilisateur) de l'acquérir, voire même d'envisager un achat groupé pour toute une promotion (par exemple en master où les effectifs sont souvent de l'ordre de la vingtaine) par les départements d'enseignement.

En conclusion, l'*Analyse de données avec R* est un très bon livre clé-en-main (à la fois autonome et accessible) de description, d'explication et d'utilisation des principales méthodes d'analyse de données. A titre personnel, je l'ai utilisé pour l'élaboration d'un cours dispensé aux étudiants de master 2 de Biostatistique de l'ISPED (public non-mathématicien) et il m'a été extrêmement utile.