

Note de lecture rédigée par Nathalie Villa-Vialaneix<sup>1</sup>

## RÉGRESSION AVEC R

Pierre-André CORNILLON et Eric MATZNER-LØBER

Livre (242 pages)

Édition : Springer-Verlag France, Collection Pratique R – 2011



Le livre « Régression avec **R** », de Pierre-André Cornillon et Eric Matzner-Løber, a pour objectif d'aborder les aspects théoriques et pratiques de la régression. Ce livre fait partie de la collection « Pratique **R** », éditée par Springer, qui se propose d'intégrer les aspects théoriques et pratiques de diverses méthodes statistiques par l'utilisation du logiciel libre multi-plateformes, **R**<sup>2</sup>. Cette collection compte actuellement quatre ouvrages (« Séries temporelles avec **R** » de Yves Aragon<sup>3</sup>, « Méthodes de Monte-Carlo avec **R** » de Christian Robert et Georges Casella<sup>4</sup> et « Psychologie statistique avec **R** » de Yvonnick Noël ; la collection sœur « Use **R** ! », également éditée par Springer, comprend, quant à elle, 45 ouvrages). Ce livre est une évolution du livre « Régression : théorie et applications »<sup>5</sup>, des mêmes auteurs et chez le même éditeur, dont il suit assez fidèlement la structure. Enfin, le livre est accompagné d'un fichier de correction des exercices (au format PDF, en français) ainsi que d'un fichier contenant les données et commandes utilisées dans l'ouvrage. Ces deux fichiers sont disponibles au téléchargement sur le site web de Springer<sup>6</sup> ; le fichier de données et

<sup>1</sup> Maîtresse de Conférences, Université de Perpignan Via Domitia – Dpt. STID, Domaine Universitaire d'Auriac, 11000 Carcassonne ; Equipe SAMM, Université Paris 1, 90 rue de Tolbiac, 75013 Paris ; [nathalie.villa@univ-paris1.fr](mailto:nathalie.villa@univ-paris1.fr)

<sup>2</sup> R Core Team. 2013. R: A Language and Environment for Statistical Computing. Vienne, Autriche. <http://www.r-project.org>

<sup>3</sup> Voir la note de lecture parue dans *Statistique et Enseignement*, 2(1) : <http://www.statistique-et-enseignement.fr/article/view/77/66>

<sup>4</sup> Voir la note de lecture parue dans *Statistique et Enseignement*, 3(1) : <http://www.statistique-et-enseignement.fr/article/view/117/106>

<sup>5</sup> Cornillon, P.-A. et E. Matzner-Løber (2006), *Régression : Théorie et Applications*, Springer, Paris.

<sup>6</sup> L'URL des fichiers complémentaires n'est pas donné dans l'ouvrage ; ils sont disponibles, respectivement, à

Note de lecture : « Régression avec R » (P.-A. Cornillon et E. Matzner-Løber, 2011)

commandes est fourni au format texte simple, contient près de 10 000 lignes, suivant strictement l'avancement du livre, et il est la concaténation de fichier de données au format texte et de fichier de commandes **R**.

L'ouvrage aborde diverses facettes de la régression, principalement dans un cadre linéaire : modèle linéaire, avec inférence et validation, régression sur variables qualitatives, sélection de variables et de modèles dans le cadre linéaire, moindres carrés généralisés, régularisation Ridge et Lasso, PCR, PLS. Une petite introduction à la régression non paramétrique est proposée dans le dernier chapitre avec la régression spline et la régression à noyau. Les chapitres sont organisés de manière similaire : ils débutent par un exposé plutôt théorique des notions abordées, puis les dernières sections de chaque chapitre sont consacrées à l'application de ces notions, avec le logiciel **R**, sur divers jeux de données réels. Chaque chapitre se termine par une série d'exercices offerts au lecteur, les exercices étant pour la plupart des exercices écrits et non des applications sur des jeux de données qui demanderaient l'utilisation du logiciel **R**.

L'ouvrage a pour ambition de traiter à la fois les aspects théoriques des notions abordées mais aussi les aspects pratiques et méthodologiques. Ce double objectif est atteint avec un exposé rigoureux des concepts, où les preuves mathématiques des théorèmes sont souvent fournies (pour les plus importantes) ou bien proposées en exercice. L'aspect pratique est abordé au travers d'exemples sur des données réelles : les commandes **R**, les sorties du logiciel (sorties numériques ou graphiques) et leurs commentaires sont intégrés dans le livre. Les commandes utilisées sont disponibles soit directement dans **R**, soit dans des packages standards comme *scatterplot3d*, *leaps*, *pls*, *splines*... L'application des méthodes sur des jeux de données réels est menée de manière progressive : les mêmes jeux de données sont utilisés dans divers chapitres, avec des conclusions qui ouvrent sur les limites des résultats produits dans un chapitre, limites qui sont éventuellement levées dans un chapitre plus avancé. Par ailleurs, de nombreux conseils pratiques généraux de mise en application des méthodes sont donnés, qui fournissent au praticien des bases rigoureuses pour l'utilisation des méthodes exposées et expliquent les principaux écueils d'interprétation ou d'application à éviter.

On peut regretter, toutefois, que les exercices n'incorporent pas plus de questions nécessitant l'utilisation pratique du logiciel **R** et des commandes introduites dans l'ouvrage. Même si certains exercices sont à visée applicative, ils consistent essentiellement en la compréhension de sorties logicielles déjà fournies. Une exception est faite dans le dernier chapitre, sur la régression spline et la régression à noyau, qui est une nouveauté par rapport à l'ouvrage de « Régression : théorie et applications ».

Contrairement à ce que son titre pourrait laisser paraître, l'ouvrage n'est donc pas destiné à apprendre l'utilisation de **R** au travers de son utilisation pour la régression mais, de manière beaucoup plus générale, il constitue un véritable exposé rigoureux sur la régression qui est illustrée avec le logiciel **R** : il s'adresse à un public ayant des bases sur le logiciel **R** et qui souhaite se former aux divers aspects de la régression et connaître les commandes spécifiques permettant de mettre celle-ci en œuvre sous **R**. Il nécessite aussi de posséder des connaissances de base en algèbre linéaire et en calcul matriciel (même si quelques concepts

---

[http://www.springer.com/cda/content/document/cda\\_downloadaddocument/correction\\_exercices.pdf?SGWID=0-0-45-1039238-p174033903](http://www.springer.com/cda/content/document/cda_downloadaddocument/correction_exercices.pdf?SGWID=0-0-45-1039238-p174033903)

et

[http://www.springer.com/cda/content/document/cda\\_downloadaddocument/Commandes\\_r%C3%A9gression\\_avec\\_R.txt?SGWID=0-0-45-1039237-p174033903](http://www.springer.com/cda/content/document/cda_downloadaddocument/Commandes_r%C3%A9gression_avec_R.txt?SGWID=0-0-45-1039237-p174033903)

*N. Villa-Vialaneix*

élémentaires sont rappelés en fin d'ouvrage). L'exposé des diverses notions y est très progressif et didactique, et les exemples, faisant le tour des commandes principales permettant d'utiliser les diverses méthodes présentées, illustrent très bien le propos. Le livre est d'un niveau très abordable pour un étudiant scientifique de premier ou deuxième cycle de statistique (DUT STID, Licence et Master de mathématiques ou de statistique, École d'ingénieurs) ; il constitue un très bon ouvrage de référence pour un enseignement sur cette thématique et peut également être utilisé par des chercheurs ou des professionnels de la statistique appliquée.