

Cet ouvrage expose de manière détaillée et exemples à l'appui, différentes façons de répondre à un des problèmes statistiques les plus courants : la régression.

Cette nouvelle édition se décompose en 5 parties. La première donne les grands principes des régressions simple et multiple par moindres carrés. Les fondamentaux de la méthode tant au niveau des choix opérés que des hypothèses et leur utilité sont expliqués. La deuxième partie est consacrée à l'inférence et présente les outils permettant de vérifier les hypothèses mises en œuvre. Les techniques d'analyse de la variance et de la covariance sont également présentées dans cette partie. Le cas de la grande dimension est ensuite abordé dans la troisième partie. Différentes méthodes de réduction de la dimension telles que la sélection de variables, les régressions sous contraintes (lasso, elasticnet ou ridge) et sur composantes (PLS ou PCR) sont notamment proposées. Un dernier chapitre propose des algorithmes, basés sur des méthodes de ré-échantillonnage comme l'apprentissage/validation ou la validation croisée, qui permettent d'établir une comparaison entre toutes ces méthodes.

La quatrième partie se concentre sur les modèles linéaires généralisés et plus particulièrement sur les régressions logistique et de Poisson avec ou sans technique de régularisation. Une section particulière est consacrée aux comparaisons de méthodes en classification supervisée. Elle introduit notamment des critères de performance pour scorer des individus comme les courbes ROC et lift et propose des stratégies de choix seuil (Youden, macro F1...) pour les classer. Ces notions sont ensuite mises en œuvre sur des données réelles afin de sélectionner une méthode de prévision parmi plusieurs algorithmes basés sur des modèles logistiques (régularisés ou non). Une dernière section aborde le problème des données déséquilibrées qui est souvent rencontré en régression binaire.

Enfin la dernière partie présente l'approche non paramétrique à travers les splines, les estimateurs à noyau et des plus proches voisins.

La présentation témoigne d'un réel souci pédagogique des auteurs qui bénéficient d'une expérience d'enseignement auprès de publics très variés. Les résultats exposés sont replacés dans la perspective de leur utilité pratique grâce à l'analyse d'exemples concrets. Les commandes permettant le traitement des exemples sous R figurent dans le corps du texte. Enfin chaque chapitre est complété par une suite d'exercices corrigés. Les codes, les données et les corrections des exercices se trouvent sur le site <https://github.com/regressionavecR.io>

Cet ouvrage s'adresse principalement à des étudiants de Master et d'écoles d'ingénieurs ainsi qu'aux chercheurs travaillant dans les divers domaines des sciences appliquées.