

Avant-propos

3^e édition

Durant de trop longues années, nous avons ruminé le fameux adage selon lequel les auteurs de livres se divisent en deux catégories : ceux qui ont le droit à la couleur et ceux qui creusent dans la grisaille des jours¹. Autant le dire tout net, sur ce coup-là, notre mégalomanie a été mise à rude épreuve. Mais fi de ces sombres souvenirs, car nous pouvons désormais le clamer haut et fort, la littérature française en statistique est enfin dotée d'un ouvrage tant lumineux qu'enluminé!²

Entendons-nous bien, il ne s'agit pas ici d'un simple bariolage de titres de sections et de quelques figures jetées de-ci de-là : notre amour du travail bien fait et notre inclination à toujours repousser les frontières du possible nous ont en effet conduits à rédiger un chapitre entièrement nouveau sur l'utilisation de fonds de cartes géographiques, avec différents types de fonds, illustré d'une farandole d'exercices idoines, et bien sûr le tout en couleurs (précision pour ceux qui n'auraient pas suivi le fil rouge de cet avant-propos). Au terme de ce travail, nous espérons donc simplement, avec toute la fausse modestie de rigueur en ces circonstances, que vous aurez autant de plaisir à consulter ce livre que nous en avons eu à rédiger cette nouvelle version, tel un arc-en-ciel après des heures de pluie ininterrompue.

2^e édition

Après une première version publiée en 2008, cette deuxième édition a permis de prendre en compte les remarques bienveillantes de nombreux lecteurs. Certaines fiches décrivant une méthode statistique ont été complétées, d'autres utilisent de nouveaux exemples. Enfin des fiches ont été ajoutées ainsi que quelques exercices.

1^{ère} édition

Ce livre s'adresse aux praticiens de la statistique au sens large, c'est-à-dire aux personnes ayant à traiter des jeux de données : étudiants en sciences, biologistes, économistes, etc. Toute étude statistique impliquant de grandes quantités d'informations, l'appel à l'outil informatique est bien entendu indispensable. Il existe

1. "You see, in this world there's two kinds of people, my friend: Those with loaded guns and those who dig. You dig." Clint Eastwood dans *The Good, the Bad and the Ugly* (1966), de Sergio Leone.

2. Un lecteur sourcilieux noterait d'emblée que ceci est très exagéré, le classique Wonnacott & Wonnacott existant en couleurs depuis longtemps. Las, cet argument n'est pas recevable en l'état, puisqu'il est aujourd'hui aussi épuisé qu'un moujik qui aurait traversé la toundra à pieds.

ainsi actuellement de nombreux logiciels répondant à ces besoins. Nous avons opté ici pour le logiciel R, qui présente le triple intérêt d'être gratuit, très complet et en essor permanent. Néanmoins, aucune connaissance sur celui-ci n'est prérequis. Cet ouvrage se veut donc accessible et utile aussi bien au débutant qu'à toute personne familière du traitement informatique des données.

Ce livre s'articule en deux grandes parties : la première est centrée sur le logiciel R et son fonctionnement, la seconde sur la mise en œuvre des méthodes statistiques classiques avec R. Afin de les rendre aussi indépendantes que possible, un bref chapitre présentant une aide au démarrage (chapitre 5, StartR) tient lieu de transition : elle permettra au lecteur plus intéressé par les statistiques que par la maîtrise du logiciel d'être rapidement opérationnel.

Nous présentons dans le premier chapitre les concepts de base du logiciel. Le deuxième chapitre traite de la manipulation des données (importation, concaténation des niveaux d'un facteur, etc.), c'est-à-dire des opérations courantes dans la pratique statistique. Le bilan d'une étude statistique passant par une visualisation claire et synthétique des résultats, nous décrivons alors, dans le chapitre 3, les nombreuses possibilités offertes par R dans ce domaine. Nous y présentons tout d'abord la construction de graphiques simples avec les différentes options disponibles puis détaillons l'utilisation de graphiques plus avancés, notamment pour construire des cartes. Quelques bases de la programmation sont ensuite présentées dans le chapitre 4 : nous expliquons comment construire ses propres fonctions mais présentons aussi quelques-unes des nombreuses fonctions prédéfinies utiles pour automatiser des analyses répétitives. Centrée sur le logiciel R lui-même, cette première grande partie permet de comprendre les commandes utilisées dans les études statistiques ultérieures.

La seconde grande partie du livre propose de balayer un large spectre des techniques classiques en traitement statistique des données : intervalle de confiance, test du χ^2 , régression linéaire, analyse de la variance, méthodes d'analyse de données, régression logistique, etc. Chaque technique est motivée par un exemple concret et traitée de façon autonome dans une fiche qui lui est dédiée. Après une brève présentation de la méthode, les lignes de commandes de R sont détaillées explicitement et les résultats sont commentés. Ainsi, le lecteur qui le souhaite pourra télécharger les jeux de données à partir de l'adresse <http://www.agrocampus-ouest.fr/math/livreR> et retrouver facilement tous les résultats décrits. A l'origine, le logiciel R fonctionne par lignes de commandes, mais une interface graphique conviviale en menus déroulants est aussi disponible : nous présentons donc à la fin de chaque fiche « Le coin R Commander », qui permet de traiter l'exemple à partir de cette interface. Celle-ci a aussi un but pédagogique puisqu'elle permet de se familiariser avec les commandes R : en effet, à chaque manipulation du menu déroulant, les lignes de code correspondant à l'action effectuée sont générées.

Pour conclure cet avant-propos, remercions J. Pagès et D. Dehay qui nous ont permis d'écrire la première édition dans des conditions optimales.