

Table des matières

Avant-propos	3
Science des données ou statistique?	3
R, un outil pour la science des données	5
Plan de l'ouvrage	6
Avant de commencer	9
I Préparer les données	13
1 Structures de données dans R	15
1.1 Vecteurs	16
1.2 Matrices et tableaux	20
1.3 Listes	22
1.4 Data frames	23
1.5 Tibbles	28
1.6 Gestion des données manquantes	39
2 Importer les données	43
2.1 Données tabulaires (texte, csv)	43
2.2 Fichiers Excel	45
2.3 Extraction à partir d'une page HTML	45
2.4 Fichier JSON	46
II Transformer les données	47
3 Notions de base de programmation	49
3.1 Opérateurs logiques	49
3.2 Instructions conditionnelles	50
3.3 Boucle <code>for</code>	51
3.4 Boucle <code>while</code>	51
3.5 Boucle <code>repeat</code>	52
3.6 Actions dans une boucle	53

4 Les fonctions avec R	55
4.1 Construire des fonctions	55
4.2 Somme, minimum, etc.	56
4.3 Tri, rang et permutation	58
4.4 Comment appliquer une fonction à tout objet ?	60
4.5 Arrondis et troncatures	62
4.6 Opérations et transformations mathématiques	63
5 Programmation avancée	65
5.1 Calcul parallèle	65
5.2 Intégrer des fonctions C dans un programme R	69
5.3 Utiliser des fonctions Fortran dans un programme R	72
III Visualiser les données	77
6 Visualiser un tableau de données	79
6.1 Représenter un nuage de points	79
6.2 Visualiser les corrélations	79
6.3 Représenter la variabilité	82
7 Représenter une courbe	85
7.1 Tracer une ligne polygonale	85
7.2 Représenter une série temporelles	87
8 Réaliser des cartes	91
8.1 Les packages maps et maptools	91
8.2 Le package rCarto	97
9 Visualiser un graphe et des réseaux	101
9.1 Introduction aux graphes et réseaux	101
9.2 Quelques packages dédiés à la représentation de réseaux	103
10 Exporter des graphiques	113
IV Analyser les données	115
11 Analyse de données multivariées	117
11.1 Analyse en composantes principales	118
11.2 Analyse factorielle des correspondances	121
11.3 Package FactoInvestigate	125
12 Lissage des séries temporelles	129
12.1 Deux classes spécifiques pour les séries temporelles	129
12.2 Moyennes mobiles ou filtres linéaires	131
12.3 Méthodes de lissage exponentielle	133

13	Modèle linéaire	139
13.1	Régression linéaire	140
13.2	Analyse de la variance	171
13.3	Analyse de la covariance	185
13.4	Transformation de Box-Cox	186
13.5	Modèle linéaire généralisée	189
14	Modèles linéaires avec pénalisation	193
14.1	Régression ridge	193
14.2	Régression LASSO	198
14.3	Régression elastic net	202
15	Méthodes supervisées de classification	207
15.1	Régression logistique	207
15.2	Arbre de décision	210
15.3	Forêt aléatoire	216
15.4	Réseaux de neurones	219
15.5	Séparateur à vaste marge	225
16	Méthodes non-supervisées de classification	229
16.1	Méthode des k centres mobiles	229
16.2	Classification ascendante hiérarchique	236
16.3	Cartes de Kohonen	239
V	Diffuser les résultats	245
17	Rédiger et présenter une étude	247
17.1	Rédiger un rapport	248
17.2	Présenter une étude	256
18	Interface web avec Shiny	263
18.1	L'exemple de base	264
18.2	Un cas plus complet	265
	Index	270