

Sommaire

Avant-propos	ix
Table des figures	xxvii
Liste des tableaux	xxxi
Notations mathématiques	xxxiii
A Présentation du logiciel R	1
A.1 Présentation du logiciel	1
A.1.1 Origines	1
A.1.2 Pourquoi utiliser R ?	1
A.2 R et les statistiques	2
A.3 R et les graphiques	3
A.4 L'interface graphique de R	4
B Quelques jeux de données et problématiques	7
B.1 Indice de masse corporelle (IMC) chez des enfants	7
B.2 Poids de naissance	8
B.3 Épaisseur de l'intima-média	9
B.4 Alimentation chez des personnes âgées	10
B.5 Étude cas témoins sur l'infarctus du myocarde	11
B.6 Tableau résumant l'utilisation des jeux de données	12
Partie I Les bases du logiciel R	13
1 Les concepts de base, l'organisation des données	15
1.1 Votre première session	15
1.1.1 R est une calculatrice	16
1.1.2 Stratégie de travail	17
1.1.3 Affichage des résultats et redirection dans des variables	19
1.1.4 Utilisation de fonctions	21

1.2	Les données dans R	24
1.2.1	Nature (ou type, ou mode) des données	25
1.2.1.1	Type numérique (<code>numeric</code>)	25
1.2.1.2	† Type complexe (<code>complex</code>)	25
1.2.1.3	Type booléen ou logique (<code>logical</code>)	27
1.2.1.4	Données manquantes (<code>NA</code>)	27
1.2.1.5	Type chaînes de caractères (<code>character</code>)	28
1.2.1.6	† Données brutes (<code>raw</code>)	29
	Récapitulatif	29
1.2.2	Structure des données	29
1.2.2.1	Les vecteurs (<code>vector</code>)	30
1.2.2.2	Les matrices (<code>matrix</code>) et les tableaux (<code>arrays</code>)	31
1.2.2.3	Les listes (<code>list</code>)	32
1.2.2.4	Le tableau individus×variables (<code>data.frame</code>)	33
1.2.2.5	Les facteurs (<code>factor</code>) et les variables ordinaires (<code>ordered</code>)	34
	Récapitulatif	35
	Termes à retenir	36
	Exercices	36
	Fiche de TP	37
2	Importation-exportation et production de données	39
2.1	Importer des données	39
2.1.1	Importer des données depuis un fichier texte ASCII	39
2.1.1.1	Lecture de données avec <code>read.table()</code>	40
2.1.1.2	Lecture de données avec <code>read.ftable()</code>	43
2.1.1.3	Lecture de données avec la fonction <code>scan()</code>	44
2.1.2	Importer des données depuis Excel ou le tableur d'OpenOffice	45
2.1.2.1	Utiliser le copier-coller	45
2.1.2.2	Passer par un fichier ASCII intermédiaire	46
2.1.2.3	Utiliser des <i>packages</i> spécialisés	46
2.1.3	Importer des données depuis SPSS, Minitab, SAS ou Matlab	47
2.1.4	Les gros fichiers de données	47
2.2	Exporter des données	48
2.2.1	Exporter des données vers un fichier texte ASCII	48
2.2.2	Exporter des données vers Excel ou OpenOffice Calc	49
2.3	Création de données	49
2.3.1	Entrer des données jouets	49
2.3.2	Générer des données pseudo-aléatoires	50
2.3.3	Entrer des données issues d'un support papier	51
	Termes à retenir	53
	Exercices	53

Fiche de TP	54
3 Manipulation de données	59
3.1 Opérations sur les vecteurs, matrices et listes	59
3.1.1 Arithmétique vectorielle	59
3.1.2 Le recyclage	60
3.1.3 Fonctions basiques	61
3.1.4 Opérations sur les matrices ou les <i>data.frames</i>	62
3.1.4.1 Informations sur l'architecture	62
3.1.4.2 Fusion de tables	63
3.1.4.3 La fonction <code>apply()</code>	67
3.1.4.4 La fonction <code>sweep()</code>	68
3.1.4.5 La fonction <code>stack()</code>	68
3.1.4.6 La fonction <code>aggregate()</code>	69
3.1.4.7 La fonction <code>transform()</code>	70
3.1.5 Opérations sur les listes	70
3.2 Opérations logiques et relationnelles	71
3.3 Opérations ensemblistes	72
3.4 Structures de contrôle	73
3.4.1 Instructions de condition	73
3.4.2 Instructions de boucles	76
3.5 Extraction et insertion d'éléments	78
3.5.1 Extraction/Insertion dans les vecteurs	79
3.5.2 Extraction/Insertion dans les matrices	81
3.5.3 Extraction/Insertion dans les <i>arrays</i>	84
3.5.4 Extraction/Insertion dans les listes	85
3.6 Manipulation de chaînes de caractères	87
3.7 Manipulation de dates et d'unités de temps	90
3.7.1 Affichage de la date courante	90
3.7.2 Extraction de dates	91
3.7.3 Opérations sur des dates	93
Termes à retenir	95
Exercices	95
Fiche de TP	97
4 R et sa documentation	103
4.1 Aide intégrée au logiciel R	103
4.1.1 La commande <code>help()</code>	103
4.1.2 Quelques commandes complémentaires	105
4.2 † Aide accessible sur l'Internet	107
4.2.1 Moteurs de recherche	108
4.2.2 Forums de discussion	108
4.2.3 Listes de diffusion (<i>mailing lists</i>)	108

4.2.4	Discussion relayée par l'Internet (IRC)	109
4.2.5	<i>Wiki</i>	109
4.3	† Littérature sur R	109
4.3.1	Sur le web	109
4.3.2	En format papier	110
	Termes à retenir	111
	Exercices	111
	Fiche de TP	111
5	Techniques pour tracer des courbes et des graphiques	113
5.1	Les fenêtres graphiques	113
5.1.1	Fenêtre graphique de base, manipulation, sauvegarde	113
5.1.2	Découpage de la fenêtre graphique : <code>layout()</code> . . .	116
5.2	Les fonctions de tracé de bas niveau	118
5.2.1	Les fonctions <code>plot()</code> et <code>points()</code>	118
5.2.2	Les fonctions <code>segments()</code> , <code>lines()</code> et <code>abline()</code> . .	120
5.2.3	La fonction <code>arrows()</code>	122
5.2.4	La fonction <code>polygon()</code>	122
5.2.5	La fonction <code>curve()</code>	123
5.2.6	La fonction <code>box()</code>	124
5.3	La gestion des couleurs	124
5.3.1	La fonction <code>colors()</code>	124
5.3.2	Le codage hexadécimal des couleurs	125
5.3.3	La fonction <code>image()</code>	128
5.4	L'ajout de texte	130
5.4.1	La fonction <code>text()</code>	130
5.4.2	La fonction <code>mtext()</code>	131
5.5	Titres, axes et légendes	132
5.5.1	La fonction <code>title()</code>	132
5.5.2	La fonction <code>axis()</code>	134
5.5.3	La fonction <code>legend()</code>	135
5.6	L'interaction avec le graphique	136
5.6.1	La fonction <code>locator()</code>	136
5.6.2	La fonction <code>identify()</code>	137
5.7	† La gestion fine des paramètres graphiques : <code>par()</code>	137
5.8	† Graphiques avancés : <code>rgl</code> , <code>lattice</code> et <code>ggplot2</code>	149
	Termes à retenir	150
	Exercices	150
	Fiche de TP	152

6	Initiation à la programmation en R	157
6.1	Préambule	157
6.2	Développer des fonctions	158
6.2.1	Mise en route rapide : déclaration, création et appel de fonctions	158
6.2.2	Concepts de base sur les fonctions	159
6.2.2.1	Corps de fonction	159
6.2.2.2	Liste de paramètres formels et effectifs . .	159
6.2.2.3	Objet retourné par une fonction	163
6.2.2.4	Portée des variables dans le corps de la fonction	165
6.2.3	Application à la problématique	167
6.2.4	Opérateurs	168
6.2.5	Le R vu comme un langage fonctionnel	169
6.3	† Programmation orientée objets	170
6.3.1	Comment fonctionne le mécanisme orienté objet du R	170
6.3.1.1	Classe d'un objet et déclaration d'un objet	170
6.3.1.2	Déclaration et utilisation d'une méthode d'un objet	172
6.3.2	Retour à la problématique	175
6.3.3	Information sur les méthodes	177
6.3.4	Héritage de classe	179
6.4	† Aller plus loin en programmation R	183
6.4.1	Attributs R	183
6.4.1.1	Attribut <code>class</code>	185
6.4.1.2	Attribut <code>dim</code>	185
6.4.1.3	Attributs <code>names</code> et <code>dimnames</code>	189
6.4.2	Autres objets R	191
6.4.2.1	Expression R	192
6.4.2.2	Formule R	194
6.4.2.3	Environnement R	197
6.5	Gestion de son activité de développement	199
6.5.1	Débogage de fonctions	199
	Termes à retenir	201
	Exercices	201
	Fiche de TP	203
7	Maintenance des sessions	209
7.1	Les commandes R, les objets et leur stockage	209
7.2	Environnement de travail : les fichiers d'extension .RData .	210
7.3	Historique des commandes : les fichiers d'extension .Rhistory	213
7.4	Sauvegarder des graphiques	215
7.5	La gestion des <i>packages</i>	216

7.6	La gestion des chemins d'accès aux objets R	217
7.7	† Autres commandes utiles	219
7.8	† La gestion de la mémoire	220
7.9	† Utiliser R en mode BATCH	220
7.10	† Crédit d'un package R simplifié	221
	Termes à retenir	223
	Exercices	223
	Fiche de TP	224

Partie II Mathématiques et statistiques élémentaires 227

8	Mathématiques de base : calcul matriciel, intégration, optimisation	229
8.1	Les fonctions mathématiques de base	230
8.2	Calcul matriciel	231
8.2.1	Opérations de base	232
8.2.2	Produit extérieur	234
8.2.3	Produit de Kronecker	235
8.2.4	Matrices triangulaires	235
8.2.5	Opérateurs vec et demi-vec	236
8.2.6	Déterminant, trace, nombre de conditionnement	236
8.2.7	Données centrées, données réduites	237
8.2.8	Calcul des valeurs propres et vecteurs propres	238
8.2.9	Racine carrée d'une matrice hermitienne définie positive	238
8.2.10	Décomposition en valeurs singulières	239
8.2.11	Décomposition de Cholesky	240
8.2.12	Décomposition QR	241
8.3	Intégration numérique	241
8.4	Dérivation	242
8.4.1	Dérivation symbolique	242
8.4.2	Dérivation numérique	243
8.5	Optimisation	244
8.5.1	Fonctions d'optimisation	244
8.5.2	Racines d'une fonction	247
	Termes à retenir	249
	Exercices	249
	Fiche de TP	250

9 Statistique descriptive	255
9.1 Introduction	255
9.2 Structuration des variables suivant leur type	256
9.2.1 Structurer les variables qualitatives	257
9.2.2 Structurer les variables ordinaires	259
9.2.3 Structurer les variables quantitatives discrètes	259
9.2.4 Structurer les variables quantitatives continues	259
9.3 Tableaux de données	260
9.3.1 Tableaux des données individuelles	260
9.3.2 Tableaux des effectifs ou des fréquences d'une variable	260
9.3.3 Tableaux de données regroupées en classes	261
9.3.4 Tableaux croisant deux variables	261
9.3.4.1 Tableaux de contingence	261
9.3.4.2 Distribution conjointe	262
9.3.4.3 Distributions marginales	263
9.3.4.4 Distributions conditionnelles	263
9.4 Résumés numériques	264
9.4.1 Résumés de position d'une distribution	264
9.4.1.1 Le ou les modes	264
9.4.1.2 La médiane	265
9.4.1.3 La moyenne	267
9.4.1.4 Les fractiles	267
9.4.2 Résumés de dispersion d'une distribution	268
9.4.3 Résumés de forme d'une distribution	268
9.5 Mesures d'association	269
9.5.1 Mesures de liaison entre deux variables qualitatives	269
9.5.1.1 La statistique du χ^2 de Pearson	269
9.5.1.2 Φ^2 , V^2 de Cramér et coefficient de contin- gence de Pearson	270
9.5.2 Mesures de liaison entre des variables ordinaires (ou des rangs)	271
9.5.2.1 Le τ et le τ_b de Kendall	271
9.5.2.2 Coefficient de corrélation des rangs de Spear- man	272
9.5.3 Mesures de liaison entre deux variables quantitatives	272
9.5.3.1 Covariance et coefficient de corrélation de Pearson	272
9.5.4 Mesures de liaison entre une variable quantitative et une variable qualitative	273
9.5.4.1 Le rapport de corrélation $\eta^2_{Y X}$	273
9.6 Représentations graphiques	274
9.6.1 Graphiques pour les variables qualitatives	274
9.6.1.1 Diagramme en croix	274
9.6.1.2 Diagramme en tuyaux d'orgue	276

9.6.1.3	Diagramme de Pareto	276
9.6.1.4	Diagramme empilé	277
9.6.1.5	Diagramme circulaire	278
9.6.2	Graphiques pour les variables ordinaires	280
9.6.2.1	Diagramme en tuyaux d'orgue avec courbe des fréquences cumulées	280
9.6.3	Graphiques pour les variables quantitatives discrètes	280
9.6.3.1	Diagramme en croix	280
9.6.3.2	Diagramme en bâtons	281
9.6.3.3	Graph de la fonction de répartition empirique	281
9.6.3.4	Diagramme en tiges et feuilles	282
9.6.3.5	Boîte à moustaches (<i>boxplot</i>)	282
9.6.4	Graphiques pour les variables quantitatives continues	284
9.6.4.1	Graphe de la fonction de répartition empirique	284
9.6.4.2	Diagramme en tiges et feuilles	285
9.6.4.3	Boîte à moustaches	286
9.6.4.4	Histogramme en densité à amplitudes de classes égales ou inégales	286
9.6.4.5	Polygone des fréquences	287
9.6.4.6	Polygone des fréquences cumulées	288
9.6.5	Représentations graphiques dans un cadre bivarié	288
9.6.5.1	Croisement de deux variables qualitatives	288
9.6.5.2	Croisement de deux variables quantitatives	291
9.6.5.3	Croisement d'une variable qualitative et d'une variable quantitative	292
	Termes à retenir	293
	Exercices	293
	Fiche de TP	294
10	Variables aléatoires, lois et simulations	299
10.1	Notions sur la génération de nombres au hasard	299
10.2	La notion de variable aléatoire	301
10.2.1	Réalisations d'une variable aléatoire et loi de fonctionnement	301
10.2.2	Variables aléatoires <i>i.i.d.</i>	303
10.2.3	Caractériser la loi d'une variable aléatoire	304
10.2.3.1	Densité, fonction de répartition, fonction quantile	305
10.2.4	Paramètres de la loi d'une variable aléatoire	308
10.3	Loi des grands nombres et théorème de la limite centrale	311
10.3.1	Loi des grands nombres	311
10.3.2	Théorème de la limite centrale	312

10.4	La statistique inférentielle	313
10.4.1	Estimation (ponctuelle) de paramètres	313
10.4.2	La fonction de répartition empirique	315
10.4.3	Estimation par la méthode du maximum de vraisemblance	316
10.4.4	Fluctuation d'échantillonnage et qualités d'un estimateur	318
10.5	Quelques techniques de simulation d'une loi	320
10.5.1	Simuler à partir d'une autre loi	320
10.5.2	Méthode de la transformation inverse	321
10.5.3	Méthode du rejet	322
10.5.4	Simulation de variables aléatoires discrètes	322
10.6	La méthode du <i>bootstrap</i>	323
10.7	Lois usuelles et moins usuelles	324
10.7.1	Lois usuelles	324
10.7.2	† Lois moins usuelles	326
10.8	Modélisation d'un phénomène	327
	Termes à retenir	330
	Exercices	330
	Fiche de TP	330
11	Intervalles de confiance et tests d'hypothèses	335
11.1	Notations	335
11.2	Intervalles de confiance	336
11.2.1	Intervalles de confiance pour une moyenne	337
11.2.2	Intervalles de confiance pour une proportion p	338
11.2.3	Intervalles de confiance pour une variance	339
11.2.4	Intervalles de confiance pour une médiane	341
11.2.5	Intervalle de confiance pour un coefficient de corrélation	342
11.2.6	Tableau récapitulatif des intervalles de confiance	342
11.3	Tests d'hypothèses usuels	342
11.3.1	Tests paramétriques	345
11.3.1.1	Tests de moyenne	345
11.3.1.2	Tests de variance	348
11.3.1.3	Tests de proportion	350
11.3.1.4	Tests de coefficient de corrélation	352
11.3.2	Tests d'indépendance	354
11.3.2.1	Test du χ^2 d'indépendance	354
11.3.2.2	Test du χ^2 de Yates	356
11.3.2.3	Test de Fisher exact	357
11.3.3	Tests non paramétriques	358
11.3.3.1	Tests d'adéquation	358

11.3.3.2	Tests de position	361
11.3.4	Tableau récapitulatif des tests usuels	366
11.4	Autres tests d'hypothèses	367
	Termes à retenir	369
	Exercices	369
	Fiche de TP	370
12	Régression linéaire simple et multiple	375
12.1	Introduction	375
12.2	La régression linéaire simple	377
12.2.1	Objectif et modèle	377
12.2.2	Ajustement sur des données	377
12.2.3	Intervalle de confiance et de prédiction pour une nouvelle valeur	382
12.2.4	Analyse des résidus	385
12.2.5	Récapitulatif	388
12.3	La régression linéaire multiple	388
12.3.1	Objectif et modèle	388
12.3.2	Ajustement sur des données	389
12.3.3	Intervalle de confiance et de prédiction pour une nouvelle valeur	393
12.3.4	Test d'une sous-hypothèse linéaire : test de Fisher partiel	393
12.3.5	Cas des variables qualitatives à plus de deux modalités	394
12.3.6	Interaction entre les variables	396
12.3.7	Problème de la colinéarité	400
12.3.8	Sélection de variables	401
12.3.9	Analyse des résidus	409
12.3.10	Cas de la régression polynomiale	416
12.3.11	Récapitulatif	417
	Termes à retenir	418
	Exercices	418
	Fiche de TP	419
13	Analyse de variance élémentaire	423
13.1	Analyse de la variance à un facteur	423
13.1.1	Les objectifs, les données et le modèle	423
13.1.2	Exemple et inspection graphique	424
13.1.3	Table d'ANOVA et estimations des paramètres	426
13.1.4	Validation des hypothèses	429
13.1.5	Comparaisons multiples et contrastes	430
13.1.6	Récapitulatif	433
13.2	Analyse de la variance à deux facteurs	434

13.2.1	Objectifs, données et modèle	434
13.2.2	Exemple et inspection graphique	435
13.2.3	Table d'ANOVA, tests et estimation des paramètres	437
13.2.4	Validation des hypothèses	440
13.2.5	Contrastes	441
13.2.6	Récapitulatif	442
13.3	Analyses de variance à mesures répétées	443
13.3.1	Modèle à un facteur à mesures répétées	444
13.3.2	Modèle à deux facteurs à mesures répétées sur les deux facteurs	445
13.3.3	Modèle à deux facteurs à mesures répétées sur un seul facteur	447
	Termes à retenir	449
	Exercices	449
	Fiche de TP	449
Annexes : Installation du logiciel R et des packages R		453
C.1	Installation de R sous Microsoft Windows XP	453
C.2	Installation de <i>packages</i> supplémentaires	454
C.2.1	Installation à partir d'un fichier situé sur le disque .	454
C.2.2	Installation directement depuis l'Internet	455
C.2.3	Installation depuis la ligne de commande	456
C.2.4	Installation de <i>packages</i> sous Linux	457
C.3	Chargement des <i>packages</i> installés	458
Références		461
Index général		465
Index des commandes et des symboles R		473
Index des auteurs		485
Liste des packages R mentionnés dans le livre		487