

Table des matières

Introduction	9
Chapitre 1. Eléments du langage	13
1.1. Avant de démarrer	13
1.1.1. Installer R	13
1.1.2. RStudio	13
1.1.3. Liste de packages utiles	14
1.1.4. Rechercher de l'aide	14
1.1.5. Scripts R	14
1.2. Représentation des données sous R	15
1.2.1. Gestion de variables numériques	15
1.2.2. Opérations sur une variable numérique	16
1.2.3. Gestion de variables catégorielles	18
1.2.4. Manipulation de variables catégorielles	19
1.3. Sélection d'observations	21
1.3.1. Sélection indexée d'observations	21
1.3.2. Sélection critériée d'observations	21
1.4. Représentation et traitement des valeurs manquantes	22
1.5. Importation et sauvegarde de données	24
1.5.1. Données univariées	24
1.5.2. Données multivariées	24
1.5.3. Sauvegarde de données dans un fichier externe	26
1.6. Gestion de données multidimensionnelles	26
1.6.1. Construction d'un tableau structuré de données	26
1.6.2. Les données <i>birthwt</i>	28
1.7. Ce qu'il faut retenir	30
1.8. Pour aller plus loin	30
1.9. Applications	31

Chapitre 2. Statistiques descriptives et estimation	35
2.1. Résumer une variable numérique	35
2.1.1. Tendance centrale et forme de la distribution	35
2.1.2. Indicateurs de dispersion	36
2.2. Résumer une variable catégorielle	37
2.3. Représenter graphiquement la distribution d'une variable	39
2.3.1. Cas des variables numériques	40
2.3.2. Cas des variables catégorielles	43
2.4. Estimation par intervalles pour une moyenne ou une proportion	44
2.4.1. Intervalles de confiance pour une moyenne	44
2.4.2. Intervalles de confiance pour une proportion	47
2.5. Ce qu'il faut retenir	48
2.6. Applications	48
Chapitre 3. Mesures et tests d'association entre deux variables	53
3.1. Statistiques descriptives bivariées	53
3.1.1. Décrire une variable numérique selon les modalités d'une variable qualitative	53
3.1.2. Décrire deux variables qualitatives	56
3.2. Comparaisons de deux moyennes de groupe	58
3.2.1. Échantillons indépendants	59
3.2.2. Échantillons non indépendants	61
3.3. Comparaisons de proportions	64
3.3.1. Cas de deux proportions	64
3.3.2. Test du chi-deux	66
3.3.3. Cas de deux échantillons non indépendants	67
3.4. Mesures de risque et odds-ratio	67
3.5. Approches non paramétriques et tests exacts	69
3.6. Ce qu'il faut retenir	70
3.7. Pour aller plus loin	71
3.8. Applications	71
Chapitre 4. Analyse de variance et plans d'expérience	77
4.1. Représentation des données et statistiques descriptives	77
4.1.1. Format de représentation des données	77
4.1.2. Statistiques descriptives et structuration des données	78
4.2. ANOVA à un facteur	80
4.2.1. Le modèle d'ANOVA à un facteur	80
4.2.2. Comparaisons par paires de traitement	83
4.2.3. Test de tendance linéaire	84

4.3. Approche non paramétrique de l'ANOVA à un facteur	88
4.4. ANOVA à deux facteurs	88
4.4.1. Construction du tableau d'ANOVA	89
4.4.2. Diagnostic du modèle	91
4.5. Ce qu'il faut retenir	92
4.6. Applications	92

Chapitre 5. Corrélation et régression linéaire 101

5.1. Corrélation	101
5.1.1. Statistiques descriptives	101
5.1.2. Diagramme de dispersion et courbe loess	103
5.1.3. Mesures d'association paramétrique et non paramétrique	103
5.1.4. Estimation par intervalles et test d'inférence	104
5.2. Régression linéaire simple	105
5.2.1. Droite de régression	105
5.2.2. Estimation par intervalles et tableau d'analyse de variance	108
5.2.3. Prédications à partir du modèle de régression	109
5.2.4. Diagnostic du modèle et analyse des résidus	110
5.2.5. Lien avec l'ANOVA	112
5.3. Régression linéaire multiple	113
5.4. Ce qu'il faut retenir	114
5.5. Pour aller plus loin	114
5.6. Applications	114

Chapitre 6. Mesures d'association en épidémiologie et régression logistique 121

6.1. Tableaux de contingence et mesures d'association	121
6.1.1. Tableaux de contingence	121
6.1.2. Mesures et tests d'association	122
6.1.3. Odds-ratio et stratification	122
6.2. Etudes diagnostiques	125
6.2.1. Sensibilité et spécificité d'un test diagnostique	125
6.2.2. Valeurs prédictives positive et négative	126
6.2.3. Tableau de synthèse des propriétés diagnostiques d'un test	127
6.3. Régression logistique	128
6.3.1. Estimation des paramètres du modèle	128
6.3.2. Prédiction par intervalles	131
6.3.3. Cas des données groupées	133
6.4. Courbe ROC	134
6.5. Ce qu'il faut retenir	135
6.6. Applications	136

Chapitre 7. Analyse des données de survie	147
7.1. Représentation des données et statistiques descriptives	147
7.2. Fonction de survie et courbe de Kaplan-Meier	148
7.2.1. Table de mortalité/survie	148
7.2.2. Courbe de Kaplan-Meier	150
7.2.3. Fonction de risque cumulé	152
7.2.4. Test d'égalité des fonctions de survie	153
7.3. Régression de Cox	154
7.4. Ce qu'il faut retenir	156
7.5. Pour aller plus loin	157
7.6. Applications	158
Annexe A. Introduction à RStudio	165
Annexe B. Les graphiques avec le package lattice	169
Annexe C. Les packages Hmisc et rms	181
Bibliographie	195
Index	199