

Table des matières

Introduction	9
Chapitre 1. Éléments du langage	13
1.1. Représentation des données sous Stata	13
1.1.1. Le langage Stata	13
1.1.2. Créer et manipuler des variables	15
1.1.3. Sélection indexée ou critériée d'observations	17
1.1.4. Traitement des valeurs manquantes	18
1.2. Gestion de données	19
1.2.1. Importation de données externes	19
1.2.2. Gestion de variables	21
1.2.3. Convertir une variable numérique en variable catégorielle	23
1.3. Statistiques descriptives univariées, estimation ponctuelle et par intervalles	24
1.3.1. Résumer une variable numérique	24
1.3.2. Résumer une variable catégorielle	26
1.4. Statistiques descriptives bivariées	27
1.4.1. Décrire une variable numérique par une variable qualitative	27
1.4.2. Décrire deux variables qualitatives	30
1.5. Ce qu'il faut retenir	31
1.6. Pour aller plus loin	31
1.7. Applications	31
Chapitre 2. Mesures d'association, comparaisons de moyennes ou de proportions pour deux échantillons ou plus	39
2.1. Comparaisons de deux moyennes de groupe	39
2.1.1. Échantillons indépendants	39
2.1.2. Échantillons non indépendants	43

2.1.3. Approche non paramétrique	43
2.2. Comparaisons de deux proportions	44
2.2.1. Echantillons indépendants	44
2.2.2. Echantillons non indépendants	47
2.3. Mesures de risque et <i>odds-ratio</i>	48
2.4. Analyse de variance	51
2.4.1. ANOVA à un facteur	51
2.4.2. Comparaisons de paires de moyennes	53
2.4.3. Test de tendance linéaire	54
2.4.4. Utilisation de contrastes	56
2.4.5. Approche non paramétrique	58
2.4.6. ANOVA à deux facteurs	59
2.5. Ce qu'il faut retenir	61
2.6. Pour aller plus loin	62
2.7. Applications	62
Chapitre 3. Régression linéaire	75
3.1. Mesures d'association entre deux variables numériques	75
3.1.1. Statistiques descriptives bivariées	75
3.1.2. Corrélation de Pearson	77
3.1.3. Corrélation non paramétrique	78
3.2. Régression linéaire	78
3.2.1. Estimation des paramètres du modèle	78
3.2.2. Prédiction ponctuelle et par intervalle	80
3.2.3. Diagnostic du modèle	81
3.2.4. Régression linéaire multiple	83
3.3. Ce qu'il faut retenir	86
3.4. Pour aller plus loin	86
3.5. Applications	86
Chapitre 4. Régression logistique et analyses épidémiologiques	97
4.1. Mesures d'association en épidémiologie	97
4.1.1. Etudes pronostiques et mesures de risque	97
4.1.2. Etudes diagnostiques	102
4.2. Régression logistique	104
4.2.1. Estimation des paramètres du modèle	104
4.2.2. Régression logistique et études diagnostiques	106
4.2.3. Prédiction ponctuelle et par intervalle	107
4.2.4. Cas des données groupées	109
4.3. Ce qu'il faut retenir	111
4.4. Pour aller plus loin	111
4.5. Applications	111

Chapitre 5. Analyse de données de survie	121
5.1. Représentation des données et statistiques descriptives	121
5.1.1. Format de représentation des données de survie	121
5.1.2. Statistiques descriptives	122
5.2. Fonction de survie et courbe de Kaplan-Meier	123
5.2.1. Table de mortalité	123
5.2.2. Courbe de Kaplan-Meier	125
5.2.3. Fonction de risque cumulé	127
5.2.4. Test d'égalité de fonctions de survie	127
5.3. Régression de Cox	129
5.4. Ce qu'il faut retenir	131
5.5. Pour aller plus loin	131
5.6. Applications	131
Bibliographie	139
Index	141