

# Table des matières

|   |           |
|---|-----------|
| <b>Avant-propos. Les statistiques : des outils indispensables à considérer avec rigueur . . . . .</b> | <b>1</b>  |
| <b>Introduction . . . . .</b>   | <b>3</b>  |
| <b>Chapitre 1. Travailler avec le logiciel R . . . . .</b>  | <b>19</b> |
| 1.1. Travailler avec le logiciel R . . . . .  | 19        |
| 1.1.1. Pourquoi et comment travailler avec le logiciel R ? . . . . .                                  | 19        |
| 1.1.2. Méthode de travail recommandée . . . . .   | 20        |
| 1.1.2.1. Créer un répertoire de travail . . . . .   | 20        |
| 1.1.2.2. Utiliser et classier son travail sous la forme de scripts . . . . .                          | 21        |
| 1.1.3. Trouver de l'aide sous R . . . . .   | 21        |
| 1.1.3.1. Pour débiter et se perfectionner sur des domaines particuliers . . . . .                     | 21        |
| 1.1.3.2. Pour chercher comment faire un test particulier ou un graphique . . . . .                    | 22        |
| 1.1.3.3. Usage . . . . .  | 24        |
| 1.1.3.4. Autres possibilités . . . . .  | 24        |
| 1.2. Opérations de base pour les statistiques sous R . . . . .  | 24        |
| 1.2.1. Manipuler le jeu de données sous R . . . . .   | 24        |
| 1.2.1.1. Opérations pour une suite de nombres, soit une variable . . . . .                            | 24        |
| 1.2.1.2. Observation d'un jeu de données . . . . .  | 25        |
| 1.2.1.3. Opérations sur les variables . . . . .   | 27        |
| 1.2.1.4. Sélection et fusion de jeux de données . . . . .   | 28        |
| 1.2.1.5. Importation/Exportation d'un jeu de données . . . . .  | 29        |
| 1.2.2. Formater le jeu de données . . . . .   | 30        |
| 1.2.3. Ordonner le jeu de données . . . . .   | 31        |
| 1.2.4. Tableaux croisés dynamiques . . . . .  | 31        |

|  |           |
|--|-----------|
| 1.3. Quelques graphiques pour résumer le jeu de données . . . . .  | 34        |
| 1.3.1. Affichage des graphiques . . . . .  | 34        |
| 1.3.2. Graphiques simples ( <i>plots</i> ) . . . . .   | 34        |
| 1.3.3. Histogrammes . . . . .  | 36        |
| 1.3.4. Boîtes à moustaches ( <i>boxplots</i> ) . . . . .   | 37        |
| 1.3.5. Diagrammes en barres ( <i>barplots</i> ) . . . . .  | 40        |
| 1.3.6. Représentations par paires . . . . .  | 42        |
| 1.3.7. Représentation graphique de diagrammes de contingence . . . . .                                   | 44        |
| 1.3.8. Étude de la dispersion et de l'interaction des données<br>sur modèle équilibré et croisé. . . . . | 46        |
| <b>Chapitre 2. Notions fondamentales en statistique . . . . .</b>  | <b>49</b> |
| 2.1. Le vocabulaire de base en statistique . . . . .   | 49        |
| 2.1.1. Élément, population, échantillon . . . . .  | 49        |
| 2.1.2. Variables . . . . .   | 51        |
| 2.2. Résumer un échantillon . . . . .  | 53        |
| 2.2.1. Paramètres de tendance centrale . . . . .   | 53        |
| 2.2.2. Paramètres de dispersion. . . . .   | 56        |
| 2.3. Les lois de probabilité. . . . .  | 57        |
| 2.3.1. Distribution normale ou loi de Laplace Gauss . . . . .  | 60        |
| 2.3.1.1. Les propriétés de la loi normale : notion d'intervalles<br>de confiance. . . . .                | 60        |
| 2.3.1.2. Inclusion des données entre deux intervalles : estimation<br>de la probabilité. . . . .         | 62        |
| 2.3.2. Les autres lois de probabilités utiles . . . . .  | 66        |
| <b>Chapitre 3. Concevoir un plan d'échantillonnage<br/>ou un plan d'expérience . . . . .</b>             | <b>69</b> |
| 3.1. Les plans d'échantillonnage . . . . .   | 69        |
| 3.1.1. Principe et critères à respecter . . . . .  | 69        |
| 3.1.2. Généralisation et fiabilité du plan . . . . .   | 73        |
| 3.1.3. Les variantes d'un échantillonnage aléatoire . . . . .  | 73        |
| 3.1.3.1. Inconvénient d'un échantillonnage aléatoire simple . . . . .                                    | 74        |
| 3.1.3.2. Échantillonnage aléatoire stratifié . . . . .   | 74        |
| 3.1.3.3. Échantillonnage aléatoire en grappes . . . . .  | 75        |
| 3.1.3.4. Échantillonnage aléatoire systématique. . . . .   | 77        |
| 3.2. Les plans expérimentaux . . . . .   | 78        |
| 3.2.1. Principe et critères à respecter . . . . .  | 78        |
| 3.2.2. Généralisation et sensibilité du plan . . . . .   | 80        |
| 3.2.3. Les différents plans expérimentaux . . . . .  | 81        |

|   |            |
|---|------------|
| 3.2.3.1. Les plans expérimentaux à un facteur contrôlé . . . . .  | 81         |
| 3.2.3.2. Les plans expérimentaux à plusieurs facteurs contrôlés . . . . .   | 82         |
| 3.2.3.3. Les plans expérimentaux à effet de bloc . . . . .  | 83         |
| 3.2.4. Plans expérimentaux et réalité terrain : des choix cruciaux<br>à considérer pour les analyses statistiques . . . . .               | 84         |
| 3.2.4.1. Plans équilibrés et plans déséquilibrés . . . . .  | 86         |
| 3.2.4.2. Facteurs fixes et facteurs aléatoires . . . . .  | 86         |
| 3.2.4.3. Plans croisés et plans hiérarchisés . . . . .  | 86         |
| 3.2.4.4. Plans additifs et plans avec interaction . . . . .   | 87         |
| <b>Chapitre 4. Principe d'un test statistique . . . . .</b>   | <b>89</b>  |
| 4.1. L'utilité des statistiques. . . . .  | 89         |
| 4.2. La théorie de la décision . . . . .  | 91         |
| 4.3. La démarche statistique . . . . .  | 93         |
| 4.3.1. Notions de risques d'erreur, de puissance et de robustesse. . . . .  | 94         |
| 4.3.2. Échantillons indépendants et appariés . . . . .  | 99         |
| 4.4. Exemple de l'application d'un test statistique . . . . .   | 99         |
| 4.4.1. Cas d'un test t de Student de deux échantillons indépendants<br>de type bilatéral . . . . .  | 102        |
| 4.4.2. Cas d'un test t de Student de deux échantillons indépendants<br>de type unilatéral . . . . .                                       | 104        |
| 4.4.3. Cas d'un test t de Student pour la comparaison d'un échantillon<br>à une norme . . . . .   | 105        |
| 4.4.4. Cas d'un test t de Student de deux échantillons appariés. . . . .  | 107        |
| 4.4.5. Puissance d'un test et recherche du nombre d'éléments<br>minimum à considérer par échantillon pour une puissance imposée . . . . . | 111        |
| <b>Chapitre 5. Clés de choix des tests statistiques . . . . .</b>   | <b>113</b> |
| 5.1. Comment utiliser les clés de choix ? . . . . .   | 113        |
| 5.2. Tests de vérification de conditions d'application. . . . .   | 121        |
| 5.2.1. Règle de Cochran . . . . .   | 121        |
| 5.2.2. Test de Kolmogorov-Smirnov . . . . .   | 122        |
| 5.2.3. Test de Shapiro-Wilks . . . . .  | 123        |
| 5.2.4. Test de Fisher-Snedecor . . . . .  | 124        |
| 5.2.5. Test de Bartlett . . . . .   | 126        |
| 5.2.6. Test de Levene . . . . .   | 128        |
| 5.2.7. Test de Goldfeld-Quandt. . . . .   | 129        |
| 5.2.8. Test de Durbin-Watson . . . . .  | 130        |
| 5.2.9. Test d'homogénéité des variances betadisp. . . . .   | 131        |

|  |                |
|--|----------------|
| <b>Chapitre 6. Tests de comparaison de paramètres en uni- ou bilatéral.</b>          | <b>133</b>     |
| 6.1. Comparaisons d'effectifs et de proportions                                      | 133            |
| 6.1.1. Test exact de Fisher.   | 133            |
| 6.1.2. Test du $\chi^2$ de conformité  | 135            |
| 6.1.3. Test du $\chi^2$ d'homogénéité.   | 138            |
| 6.1.4. Test de Mac Nemar.  | 140            |
| 6.1.5. Test binomial   | 142            |
| 6.1.6. Test de Cochran-Mantel-Hantzel.   | 142            |
| 6.2. Comparaisons de moyennes  | 145            |
| 6.2.1. Test de Student   | 145            |
| 6.2.1.1. Pour échantillons indépendants  | 146            |
| 6.2.1.2. Test de Welch.  | 148            |
| 6.2.1.3. Pour la comparaison d'un échantillon à une norme.                           | 150            |
| 6.2.1.4. Pour la comparaison d'échantillons appariés                                 | 152            |
| 6.2.1.5. Puissance du test de Student  | 154            |
| 6.2.2. Test U de Wilcoxon-Mann-Withney.  | 155            |
| 6.2.2.1. Test U de Wilcoxon-Mann-Withney pour échantillons indépendants              | 156            |
| 6.2.2.2. Test de conformité de Wilcoxon-Mann-Withney                                 | 158            |
| 6.2.2.3. Test U de Wilcoxon pour échantillons appariés                               | 160            |
| 6.2.2.4. Puissance du test de Wilcoxon-Mann-Withney                                  | 161            |
| 6.3. Test de corrélation de deux variables quantitatives                             | 163            |
| 6.3.1. Corrélation linéaire : le coefficient de Bravais-Pearson                      | 163            |
| 6.3.2. Corrélation monotone : les coefficients de Spearman ou de Kendall             | 166            |
| <br><b>Chapitre 7. Modèles linéaires classiques et généralisés</b>                   | <br><b>171</b> |
| 7.1. Principe des modèles linéaires.   | 173            |
| 7.1.1. Les modèles linéaires classiques : le principe des moindres carrés.           | 173            |
| 7.1.1.1. Principe général sur régression simple et ANOVA à un facteur                | 173            |
| 7.1.1.2. La régression multiple.   | 177            |
| 7.1.1.3. L'ANOVA multifactorielle.   | 180            |
| 7.1.1.4. Les analyses de covariance (ANCOVA).  | 180            |
| 7.1.1.5. Clé de choix des plans d'ANOVA à mettre en place.                           | 182            |
| 7.1.2. Les modèles linéaires généralisés : le principe des maximums de vraisemblance | 182            |
| 7.2. Les conditions d'application du modèle  | 186            |
| 7.2.1. Les modèles linéaires classiques  | 186            |
| 7.2.2. Les modèles linéaires généralisés   | 189            |

|  |     |
|--|-----|
| 7.3. Autres analyses utiles . . . . .  | 190 |
| 7.3.1. Transformations, linéarisation . . . . .  | 190 |
| 7.3.2. Les tests <i>post hoc</i> pour ANOVA ou GLM équivalent . . . . .                                  | 193 |
| 7.3.3. La partition de variance . . . . .  | 195 |
| 7.3.4. Critères de sélection de modèles linéaires : critère AIC<br>ou BIC . . . . .                      | 197 |
| 7.3.5. Du modèle à la prédiction . . . . .   | 198 |
| 7.4. Exemple d'application des différents modèles linéaires. . . . .                                     | 198 |
| 7.4.1. Régression linéaire simple. . . . .   | 198 |
| 7.4.1.1. Avec vérification des conditions d'application détaillée. . . . .                               | 198 |
| 7.4.1.2. Avec un exemple de prédiction à partir de l'équation<br>du modèle . . . . .                     | 205 |
| 7.4.2. Régression linéaire multiple. . . . .   | 207 |
| 7.4.2.1. Avec un exemple de partition de variance . . . . .  | 207 |
| 7.4.2.2. Avec un exemple d'optimisation de modèle<br>pour une régression multiple . . . . .              | 212 |
| 7.4.3. ANOVA à un facteur. . . . .   | 215 |
| 7.4.4. ANOVA multifactorielle. . . . .   | 221 |
| 7.4.4.1. ANOVA sur plan croisé équilibré . . . . .   | 221 |
| 7.4.4.2. ANOVA sur plan croisé déséquilibré . . . . .  | 227 |
| 7.4.4.3. ANOVA multifactorielle sur plan hiérarchisé . . . . .   | 233 |
| 7.4.4.4. ANOVA à mesures répétées : plan mixte,<br>plan avec appariement, plan à effet de bloc . . . . . | 238 |
| 7.4.5. Analyse de covariance (ANCOVA) . . . . .  | 241 |
| 7.5. Exemples d'application de modèles linéaires généralisés (GLM). . . . .                              | 244 |
| 7.5.1. GLM pour des données de proportions ou binaires<br>sous la loi binomiale . . . . .                | 244 |
| 7.5.2. GLM pour des données discrètes sous la loi de Poisson . . . . .                                   | 250 |
| 7.5.3. GLM pour des données réelles positives sous la loi gamma . . . . .                                | 257 |
| 7.5.4. GLM pour des données réelles sous la loi gaussienne . . . . .                                     | 263 |

## **Chapitre 8. Alternatives non paramétriques aux modèles linéaires . . . . .**

267

|   |     |
|---|-----|
| 8.1. Principe des tests non paramétriques . . . . .   | 267 |
| 8.2. Les alternatives d'ANOVA . . . . .   | 268 |
| 8.2.1. Les ANOVA basées sur les rangs . . . . .   | 268 |
| 8.2.2. Alternative non paramétrique à une ANOVA à un facteur :<br>le test de Kruskal-Wallis et <i>post hoc</i> associés . . . . . | 268 |
| 8.2.3. Alternative non paramétrique à une ANOVA à deux facteurs :<br>test de Scheirer-Ray-Hare . . . . .                          | 271 |
| 8.2.4. Alternative non paramétrique à une ANOVA avec appariement :<br>le test de Friedman . . . . .                               | 272 |

|  |            |
|--|------------|
| 8.3. Les ANOVA non paramétriques basées sur les tests<br>par permutation (PERMANOVA) . . . . . | 275        |
| 8.3.1. Principe . . . . .  | 275        |
| 8.3.2. Exemple d'application sous R . . . . .  | 277        |
| 8.4. Les modèles non linéaires . . . . .   | 283        |
| <b>Conclusion</b> . . . . .  | <b>285</b> |
| <b>Bibliographie</b> . . . . .   | <b>287</b> |
| <b>Index</b> . . . . .   | <b>289</b> |