Cours de Bio-Statistiques

**(Jean-François Martin, Ingénieur Biostatisticien, INRA de Toulouse)**

Les analyses statistiques dans le domaine de la biologie, de l’Agronomie et des analyses « omiques ». Description, modélisation et fouilles de données.

**A - Présentation générale**

1. Introduction :
2. Analyses descriptives
3. Tests statistiques et modèle linéaire (statistiques univariées)
   1. Principes, hypothèses et risques
   2. Le choix d’un test univarié
   3. Les tests paramétriques
      1. Test de Student et dérivés
      2. Anova et modèle linéaire
      3. Modèle mixte, analyse temporelle
   4. Les tests non paramétriques (Man & Whitney, Wilcoxon, Kruskal & Wallis)
4. L’analyse des données (analyses multivariées)
   1. Corrélation et covariance
   2. Régressions linéaires.
   3. Analyses descriptives non supervisées
      1. L’analyse en composante principale (ACP)
      2. Les méthodes de classification
         1. Classification ascendante hiérarchique (CAH)
         2. Classification par centres mobiles (Kmeans)
   4. Méthodes supervisées
      1. La partial least square régression (PLSR)
      2. La partial least square discriminant analysis (PLSDA)
5. La planification expérimentale ou comment optimiser les résultats et conclusions par le choix d’un design approprié. Utilisation de plans factoriels et plan incomplet de criblage (Placket Burman)

**B - Applications dans quelques domaines de la Biologie**

1. Application aux données « omiques »
   1. Correction de p-value, false discovery rate (FDR)
2. Intégration de données multi-tableaux transcriptomique-protéomique-métabolomique
   1. Modélisation PLS (Partial least square), sparse PLS
   2. Analyse canonique
3. Validation de biomarqueurs
   1. Les courbes ROC

**C – Les logiciels**

1. Le langage R : package FactoMineR et package mixOmics
2. Les outils sur le web sous Galaxy

**Déroulement des cours et travaux dirigés**

|  |  |  |
| --- | --- | --- |
|  | MATIN (cours) | APRES-MIDI (TD) |
| Jour1 | * Introduction * Statistiques descriptives * Tests univariés (Test de student et ANOVA) * Test non paramétriques (Wilcoxon, Kruskal Wallis) * Correction des p-values * Les courbes ROC * Planification expérimentale | * Présentation de R et Rstudio * Installation des packages * Les différents types d’objets R et les fonctions de bases. * Fonctions graphiques (outils R standards et ggplot2) * Différentes fonctions de tests univariées (t.test, aov, lm) |
| Jour2 | * Corrélation et covariance * Régressions linéaires * Fouilles des données par analyses non supervisées (Analyse en composantes principales et clustering) | * Calcul de corrélations et représentations graphiques (package corrgram) * Fonctions graphiques scatter plots. * ACP avec le package FactoMineR * Clustering Hiérarchique et heatmap. |
| Jour3 | * Analyses multivariées supervisées (Partial Least Square) * Analyses discriminantes et canoniques * Intégration de données * Présentation du package mixOmics | * Analyses PLS et intégration de données avec le package mixOmics. |
|  |  |  |