

# RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE

## MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE

### **Master professionnel de Biostatistiques**

#### **Programme détaillé par matière**

Le Master professionnel de Biostatistiques est composé de 4 grandes parties :

- Probabilités et Bio-statistiques ;
- Epidémiologie quantitative, Méthodologie et Recherche clinique ;
- Utilisation du logiciel SPSS ;
- Utilisation du langage R

Les aspects théoriques et pratiques des différentes parties du programme sont présentés dans le programme suivant et sont assurés au cours du Master 1 (les 2 premiers semestres) et au cours du 1<sup>er</sup> semestre du Master 2.

La 2<sup>ème</sup> partie du Master 2 (Février - Juin de la 2<sup>ème</sup> année) est consacrée au mémoire de fin d'études.

Des thèmes sont proposés aux étudiants qui travaillent par binôme. Les mémoires sont encadrés par les enseignants du Master et par les experts des thèmes proposés.

## **Semestre 1**

### **1- Intitulé de la matière : Probabilités**

**Objectifs de l'enseignement :** Les participants doivent :

- Avoir assimilé les notions de probabilités à partir de l'analyse combinatoire ;
- Savoir utiliser ces notions dans la construction des lois de probabilités usuelles ;
- Savoir utiliser les tables statistiques.

**Connaissances préalables recommandées :**

Connaissances mathématiques de base acquises dans le cadre d'une licence en biologie ou de tout autre diplôme en graduation (sciences médicales, humaines ou vétérinaires).

**Contenu de la matière (Tout le contenu de la matière se fait sous forme de Cours + TD) :**

- Rappels sur l'analyse combinatoire : 4 heures ;
- Généralités sur les probabilités / Probabilités conditionnelles / Théorème de Bayes : 4 heures ;
- Généralités sur les distributions de probabilité / Lois discrètes : 4 heures ;
- Lois de probabilité continues : Loi uniforme / Loi normale : 4 heures ;
- Lois dérivées de la loi normale : 4 heures.

### **2- Intitulé de la matière : Bio-statistique 1**

**Objectifs de l'enseignement :** Les participants doivent être capables :

- D'utiliser les méthodes statistiques de base pour répondre à des questions pratiques (Etude de l'efficacité d'un traitement, recherche de facteurs de risque, ...) ;
- D'interpréter correctement les résultats des tests statistiques ;
- De savoir lire et interpréter un listing obtenu lors d'une analyse faite par le logiciel.

**Connaissances préalables requises :**

Avoir suivi le cours sur les Probabilités

**Contenu de la matière (Tout le contenu de la matière se fait sous forme de Cours + TD) :**

- Eléments de statistique descriptive : 4 heures ;
- Estimation par intervalle de confiance d'un paramètre dans une population : 4 heures ;
- Généralités sur les tests / Comparaison de 2 ou de plusieurs pourcentages / Comparaison d'un % observé à un % théorique : 4 heures ;
- Comparaison de 2 moyennes / Comparaison d'une moyenne observée à une moyenne théorique : 4 heures ;
- Tests non paramétriques dans la comparaison de 2 moyennes : 4 heures ;
- Puissance d'un test : 4 heures.

### 3- Intitulé de la matière : Epidémiologie quantitative

#### Objectifs de l'enseignement :

Les participants doivent pouvoir interpréter les résultats des enquêtes épidémiologiques, qu'elles soient descriptives, étiologiques ou évaluatives.

#### Connaissances préalables recommandées :

- Estimation par intervalle de confiance ;
- Tests statistiques usuels.

#### Contenu de la matière (Tout le contenu de la matière se fait sous forme de Cours + TD) :

- Les différents types d'études en épidémiologie clinique : 4 heures ;
- Mesures de fréquences : 4 heures ;
- Mesures d'association / Risque attribuable : 4 heures ;
- Critères de validité interne et externe d'un test diagnostique : 4 heures ;
- Applications pratiques : 4 heures.

### 4- Intitulé de la matière : Initiation au logiciel R

#### Objectifs de l'enseignement :

Le langage R se présente sous la forme d'un logiciel **libre** (Open Access).

Ce logiciel permet de gérer des tableaux de données et de faire des **analyses** statistiques, des représentations graphiques, de l'analyse d'images et du **calcul** numérique.

R fonctionne avec le principe des **Packages** ; un Package est une librairie qui contient une collection de fonctions utilisable sous R.

R est en **évolution** continue ; plusieurs chercheurs proposent leurs méthodes sous forme de package R.

Ce cours est destiné pour les Masters Bio-statistiques (SNV Université Alger 1 et Institut Pasteur d'Algérie). Il aborde l'installation du logiciel R avec son interface graphique R-studio.

#### Connaissances préalables recommandées :

Connaître : Variables quantitatives et qualitatives, paramètres de distribution : moyenne, variance, écart type, principes des tests statistiques, comparaisons de deux moyennes (Test t).

#### Contenu de la matière :

- Installation du logiciel R (Rcmdr et r-Studio) ;
- Présentation du logiciel R ;
- Présentation du logiciel r-studio ;
- Présentation des objets et du fonctionnement de R ;
- Les packages R.

## **5- Intitulé de la matière : Initiation au logiciel SPSS**

### **Objectifs de l'enseignement :**

Les participants doivent être capables d'interpréter les résultats des enquêtes épidémiologiques, qu'elles soient descriptives, étiologiques ou évaluatives.

### **Connaissances préalables recommandées :**

Utilisation du Microsoft Office : Word, Excel, ...

### **Contenu de la matière :**

- Introduction (Installation, Configuration, Apprendre à connaître et à familiariser avec les différentes commandes SPSS) ;
- Créer un masque de saisie sur le logiciel SPSS ;
- Saisir les données d'une étude, d'un projet sur le logiciel SPSS ;
- Importer un fichier de données (SPSS, dbase files, ...) ;
- Procéder à la sélection, création, transformation et recodage des variables sur SPSS ;
- Analyser et exploiter des données sur SPSS ;
- Les tableaux de fréquences pour une variable qualitative ;
- Détermination des paramètres de tendance centrale (moyenne, mode, médiane, quartiles..) et les paramètres de dispersion (écart-type, variance, erreur-type, ...) ;
- Procédure Explorer : La commande Explorer donne accès à des statistiques descriptives additionnelles ainsi qu'à des graphiques qui aident à mieux comprendre les données ;
- Intervalle de confiance d'un pourcentage et d'une moyenne ;
- Loi binomiale.
- Appliquer les tests d'hypothèses usuels sur SPSS :
  - Test de Khi-deux pour comparer deux pourcentages de deux variables qualitatives ;
  - Tester la normalité d'une distribution (Test Kolmogorov Smirnov, test de Shapiro-wilk) ;
  - Test T pour échantillons indépendants (comparaison entre deux moyennes).

## **Semestre 2**

### **1- Intitulé de la matière : Bio-statistique 2**

**Objectifs de l'enseignement** : Les participants doivent être capables :

- D'utiliser les méthodes statistiques permettant d'étudier le lien entre 2 variables, quantitatives ou qualitatives ;
- D'évaluer le taux d'incidence d'un évènement et son évolution au cours du temps ;
- D'évaluer la concordance entre 2 tests diagnostiques, qu'ils soient basés sur un critère quantitatif ou qualitatif.

**Connaissances préalables requises** :

Avoir suivi les cours de Biostatistiques 1 et d'épidémiologie quantitative.

**Contenu de la matière (Tout le contenu de la matière se fait sous forme de Cours + TD) :**

- Régression et corrélation linéaire : 4 heures ;
- Test de tendance entre plusieurs pourcentages et entre plusieurs moyennes : 4 heures ;
- Comparaison de plusieurs moyennes : 4 heures ;
- Test de linéarité : 4 heures ;
- Concordance entre deux tests diagnostiques : 4 heures ;
- Courbes de survie : méthode de Kaplan Meier et Méthode actuarielle : 4 heures ;
- Comparaison de 2 ou de plusieurs courbes de survie : 4 heures ;
- Courbe ROC et domaines d'utilisation : 4 heures.

### **2- Intitulé de la matière : Méthodologie**

**Objectifs de l'enseignement** : Les participants doivent être capables :

D'élaborer un protocole d'étude avec toutes ses composantes : bien définir la question de recherche, définir les objectifs de l'étude et leur critères d'évaluation, définir la population d'étude et les modalités d'échantillonnage, élaborer le plan d'analyse statistique.

**Connaissances préalables requises** :

Avoir suivi les cours de Bio-statistiques et d'épidémiologie quantitative.

**Contenu de la matière (Tout le contenu de la matière se fait sous forme de Cours + TD) :**

- Elaboration d'un protocole : 4 heures ;
- Méthodes d'échantillonnage : 4 heures ;
- Elaboration d'un questionnaire : 4 heures ;
- Etudes cas-témoins : 4 heures ;
- Etudes exposés - non exposés : 4 heures ;
- Calcul de la taille minimale d'un échantillon selon le type d'étude : 4 heures ;
- Applications pratiques : Travaux personnels sur l'élaboration d'un protocole : 4 heures.

### **3- Intitulé de la matière 1 : Analyse statistique 1 avec logiciel R**

#### **Objectifs de l'enseignement :**

Maîtriser les bases du logiciel R pour le traitement des données et pour l'analyse statistique.

#### **Connaissances préalables recommandées :**

- Connaître les principes de l'informatique bureautique (Traitements de données avec un outil de type tableau, ex : Excel...)
- Avoir des connaissances en statistiques.

#### **Contenu de la matière (Tout le contenu de la matière se fait sous forme de Cours + TD +TP) :**

- Calculs élémentaires dans R ;
- Import et export des données dans R ;
- Le traitement des données ;
- Création d'objets avec R ;
- Création de vecteurs avec R ;
- Création d'arguments avec R ;
- Organisation des données tidyverse ;
- But de l'étude statistique ;
- Statistique descriptive simple à une variable ;
- Représentations graphiques ;
- Statistique Double et courbe ROC ;
- Data visualisation.

### **4- Intitulé de la matière 1 : Analyse statistique 1 avec le logiciel SPSS**

#### **Objectifs de l'enseignement :**

Les participants doivent être capables d'interpréter les résultats des enquêtes épidémiologiques qu'elles soient descriptives ou analytiques, d'évaluer la qualité d'un test diagnostique et d'évaluer la survie dans un essai clinique ou une étude pronostique.

#### **Connaissances préalables recommandées :**

Avoir suivi et validé le module « Introduction au logiciel SPSS ».

#### **Contenu de la matière :**

- Corrélation bi-variée et l'équation de la régression sur SPSS ;
- Analyse de la variance (ANOVA à un seul facteur) ;
- Les mesures d'association épidémiologiques (OR, RR, Intervalle de confiance) ;
- Maîtrise des différentes procédures d'analyses des données (uni-variées, bi-variées) ;
- Présentation graphique sur SPSS ;
- Analyse à trois variables sur SPSS (Khi-deux de MH, OR ajusté, interaction, confusion, ...)
- Analyse de survie sur SPSS ;
- Test de Mc-Nemar pour des échantillons appariés ;
- Test t pour échantillons appariés - Analyse de fiabilité (Kappa, coefficient de corrélation interclasse) ;
- La courbe de ROC sur SPSS.

## **Semestre 3**

### **1- Intitulé de la matière : Bio-statistique 3**

**Objectifs de l'enseignement** : Les participants doivent être capables :

- De choisir et utiliser les modèles mathématiques adaptés à la problématique posée ;
- D'interpréter les résultats des analyses multi-variées à partir des listings obtenus lors de l'analyse par SPSS ou par le langage R.

**Connaissances préalables recommandées** :

Avoir suivi les cours de Bio-statistiques et d'épidémiologie quantitative.

**Contenu de la matière (Tout le contenu de la matière se fait sous forme de Cours + TD) :**

- Généralités sur le calcul matriciel : opérations sur les matrices, diagonalisation : 8 heures ;
- Généralités sur l'analyse multi variée : 2 heures ;
- Analyse multi-variée descriptive : Analyse en composantes principales : 4 heures ;
- Analyse multi-variée descriptive : Analyse des correspondances multiples : 4 heures ;
- Analyse multi variée explicative : Régression logistique binaire, Régression logistique multinomiale : 4 heures ;
- Analyse multi-variée explicative : Régression linéaire multiple : 4 heures ;
- Modèle de Cox dans la survie : 4 heures ;
- Séries chronologiques : 4 heures ;
- Applications pratiques / Interprétation de listings : 8 heures.

### **2- Intitulé de la matière : Recherche clinique**

**Objectifs de l'enseignement** : Les participants doivent être capables :

- De lire et de vérifier la validité méthodologique et statistique d'un article médical ;
- D'interpréter une méta-analyse ;
- De rédiger un article scientifique en vue d'une publication.

**Connaissances préalables requises** :

Avoir suivi tous les cours du Master 1 et avoir validé l'année.

**Contenu de la matière (Tout le contenu de la matière se fait sous forme de Cours + TD) :**

- Essais cliniques de supériorité / Essais clinique de non infériorité : 6 heures ;
- Analyse des données d'un essai clinique : 4 heures ;
- Lecture critique d'un article médical : 4 heures ;
- Méta-analyse : 4 heures ;
- Rédaction d'un article médical : 4 heures.

### 3- Intitulé de la matière : Analyse statistique 2 avec le langage R

#### Objectifs de l'enseignement :

Maîtriser les bases du logiciel R pour le traitement des données et pour l'analyse statistique.

#### Connaissances préalables recommandées :

- Connaître les principes de l'informatique bureautique (traitements de données avec un outil de type tableau, ex : Excel...);
- Avoir des connaissances en statistiques.

#### Contenu de la matière (Tout le contenu de la matière se fait sous forme de Cours + TP) :

- Tests Statistiques avec Rcmdr ;
- Analyse de variance à 1 facteur (ANOVA) ;
- Régression linéaire généralisée (GLM) ;
- Analyse Multi-variées ;
- L'analyse en composantes principales (ACP) ;
- L'analyse factorielle des correspondances (AFC) ;
- L'analyse des correspondances multiples (ACM) ;
- Facto Mine R sur Rcmdr ;
- Le package 'Shiny' pour le développement d'application Web interactive sous R.

### 4- Intitulé de la matière : Analyse statistique 2 avec le logiciel SPSS

#### Objectifs de l'enseignement :

Les participants doivent être capables d'interpréter les résultats des analyses multi-variées, descriptives ou explicatives.

#### Connaissances préalables recommandées :

Avoir suivi et validé les modules SPSS assurés au Master 1.

#### Contenu de la matière :

##### - Analyse multi variée descriptive :

- . Analyse en composantes principales ;
- . Analyse factorielle des correspondances ;
- . Analyse des correspondances multiples.

##### - Analyse multi variée explicative :

- . Régression logistique binaire et multinomiale ;
- . Modèle de Cox dans la survie ;
- . Régression linéaire multiple ;
- . MANOVA (modèles linéaires généralisés).

##### - Séries chronologiques.