

# BIostatistique « DES STATISTIQUES FREQUENTISTES AUX STATISTIQUES BAYESIENNES »

(Diplôme d'Université des)

- **Objectifs** : Les objectifs de cette formation sont d'apporter à l'étudiant des compétences statistiques à jour, c'est-à-dire que l'étudiant devra être capable de lire, comprendre et discuter un article scientifique récent utilisant inférence fréquentiste ou bayésienne, de montrer son autonomie dans l'interprétation de données et dans la rédaction des sections « méthode » et « résultats » d'un article ainsi que dans la rédaction d'un rapport d'analyse.

**L'objectif principal** est décliné en objectifs spécifiques :

1. Elicitation des lois *a priori*
  - a. Connaître les règles générales de l'élicitation
  - b. Connaître les lois de distribution essentielles utilisables comme prior (densité, paramètres, propriétés)
  - c. Savoir choisir les hyperparamètres d'une loi *a priori*
  - d. Connaître les sources d'information utilisation pour l'élicitation (littérature, experts,...)
  - e. Juger de l'informativité d'un prior
  - f. Mener une analyse de sensibilité et en interpréter les résultats
2. Modèles de vraisemblance
  - a. Savoir choisir un modèle de vraisemblance en fonction des données
3. Conduite à tenir *a posteriori*
  - a. Maitriser les fondements théoriques du passage des lois *a priori* aux lois *a posteriori*
  - b. Connaître les méthodes analytiques (situation conjuguée) d'extraction des lois *a posteriori* et les méthodes par simulation et/ou approximation (McMC)
  - c. Maitriser la méthode de fonctionnement des McMC (nombre d'itération, burn-in, thinning,...)
  - d. Savoir poser un diagnostic de convergence probable des algorithmes
  - e. Connaître la conduite à tenir en cas de non convergence des algorithmes
4. Présentation des résultats
  - a. Savoir choisir des modèles graphiques et descriptifs des données
  - b. Savoir présenter et discuter des résultats d'inférence (tendance centrale, étendue de distribution,...)
  - c. Savoir présenter et discuter des résultats de convergence des algorithmes
5. Autres considérations méthodologiques
  - a. Savoir mettre en évidence les données manquantes, en discuter les mécanismes de survenue et choisir une méthode éventuelle de traitement
  - b. Connaître les schémas des études épidémiologiques et de recherche clinique et adapter les méthodes d'analyse en conséquence
  - c. Maitriser un calcul de nombre nécessaire de sujets

- **Durée** : 1 an                      **Nombre de places** : 20 (minimum :10)

- **Droits à payer** :

➤ **Si inscription unique à ce seul diplôme ou comme 1<sup>er</sup> DU (si plusieurs DU)**

**En 1<sup>ère</sup> année** : droits administratifs (261,10 €) + droits pédagogiques (500 € ou 1000 € ou 200 €\*)

➤ **Si inscription seconde : comme 2<sup>ème</sup> ou énième DU (car plusieurs DU)**

**En 1<sup>ère</sup> année** : droits administratifs réduits (168 €) + droits pédagogiques (500 € ou 1000 € ou 200 €\*)

Droits administratifs – droits pédagogiques :

- Formation initiale : 500€
- Formation continue : 1000€
- Réinscription : 200€

- **Enseignant responsable** :

**Pr. Erik-André SAULEAU**

Laboratoire de Biostatistique – Faculté de Médecine – 4 rue Kirschleger – F-67085 STRASBOURG

☎ 03 68 85 39 51            Mèl : [ea.sauleau@unistra.fr](mailto:ea.sauleau@unistra.fr)

- **Public concerné** :

Médecin ou non médecin, **ayant déjà suivi** une formation initiale de base à la statistique et/ou biostatistique (application de la statistique aux données issues des sciences de la vie) :

- Internes en médecine
- Etudiants en 2<sup>ème</sup> Cycle des études médicales
- Médecins cliniciens, Spécialistes de recherche clinique (ARC,...), Statisticiens de l'industrie pharmaceutique,...

Les inscriptions se feront soit au titre de la Formation Initiale, soit au titre de la Formation Continue.

Sans qu'il y ait d'examen probatoire, l'étudiant sera invité à vérifier qu'il a atteint un niveau minimal requis au moyen d'un auto-questionnaire à remplir et valider sur internet. Des conseils d'ouvrages statistiques pourront l'aider le cas échéant à se hisser à ce niveau. **Un entretien avec l'enseignant responsable du DU finalisera l'acceptation ou non du candidat.**

• **Cours et programme :**

Le DU se déroule sur une année universitaire, sans stage. Un volume total de 100 heures sera dispensé sous la forme de 60 heures de cours magistraux et de 40 heures de TD, à raison de d'une session de deux jours consécutifs par mois.

Les cours auront lieu dans **la Salle informatique 107 du Forum de la Faculté de Médecine**

- 23/11/2016 (13h30-18h00), 24/11/2016 (9h00-18h00), 25/11/2016 (9h00-18h00)
- 04/01/2017 (13h30-18h00), 05/01/2017 (9h00-18h00), 06/01/2017 (9h00-18h00)
- 25/01/2017 (13h30-18h00), 26/01/2017 (9h00-18h00), 27/01/2017 (9h00-18h00)
- 08/03/2017 (13h30-18h00), 09/03/2017 (9h00-18h00), 10/03/2017 (9h00-18h00)
- 29/03/2017 (13h30-18h00), 30/03/2017 (9h00-18h00), 31/03/2017 (9h00-13h00)
- 03/05/2017 (13h30-18h00), 04/05/2017 (9h00-18h00), 05/05/2017 (9h00-13h00)
- 31/05/2017 (13h30-18h00), 01/06/2017 (9h00-18h00), 02/06/2017 (9h00-18h00)
- 21/06/2017 (13h30-18h00), 22/06/2017 (9h00-18h00), 23/06/2017 (9h00-18h00)

En plus des 100 heures de présence, l'étudiant aura à assurer des travaux personnels (éventuellement en groupe) pour chaque TD ainsi que la rédaction d'un rapport pour la validation finale du DU.

Les enseignements se répartissent en 60 heures de cours magistraux et 40 heures de TD dont les intitulés sont les suivants :

| Chapitres des CM   | (total de 60 heures) |
|--|----------------------|
| 1 Principal général et probabilités                            | 12h                  |
| 2 Lois de distribution   | 8h                   |
| 3 Statistiques descriptives et inférence (posterior explicite) | 4h                   |
| 4 Techniques   | 8h                   |
| 5 Modélisation   | 12h                  |
| 6 Adéquation des modèles                                       | 4h                   |
| 7 Données manquantes, données aberrantes                       | 8h                   |
| 8 Epistémologie bayésienne et conclusion                       | 4h                   |

| TD  | (total de 40 heures) |
|---|----------------------|
| TD 0: rédaction scientifique et lecture critique d'articles | 6h                   |
| TD A: introduction au logiciel R                            | 6h                   |
| TD B: introduction à WinBUGS et interface avec R            | 8h                   |
| TD C: Statistiques descriptives et inférence (McMC)         | 6h                   |
| TD D: équivalent de modèles mixtes                          | 8h                   |
| TD E: Essai thérapeutique                                   | 6h                   |

Détails du programme :

- 1. Principal général et probabilités**
  - 1.1. Brève histoire des probabilités et de la statistique (EAS) 4h
    - 1.1.1. Bayes-Laplace
    - 1.1.2. Galton, Pearson K, Fisher, Neyman-Pearson
    - 1.1.3. Jeffreys, Savage, Lindley, de Finetti, les modernes
  - 1.2. Probabilités (NM) 4h
    - 1.2.1. Axiomatique
    - 1.2.2. Théorème de Bayes-Laplace
  - 1.3. Discussion philosophique sur les types de probabilités (EAS) 4h
- TD 0 : rédaction scientifique et lecture critique d'articles (NM+autre intervenant) 6h**
- 2. Lois de distribution**
  - 2.1. Des événements aux lois de probabilité (élicitation) (NM) 4h
  - 2.2. Lois discrètes (NM) : Bernoulli, binomiale, multinomiale, Poisson 2h
  - 2.3. Lois continues (EAS) : Gauss, beta, Dirichlet, gamma, Wishart, uniforme 2h
- TD A : introduction au logiciel R (EAS+autre intervenant) 6h**
- 3. Statistiques descriptives et inférence: Deux situations courantes (posterior explicite)**
  - 3.1. Une proportion=beta, descriptif et inférence (NM) 2h
  - 3.2. Une moyenne=normale, descriptif et inférence (EAS) (variance connue et variance inconnue) 2h
- 4. Techniques (EAS)**
  - 4.1. Algorithmes 2h
    - 4.1.1. MCMC
    - 4.1.2. Gibbs
    - 4.1.3. Autres

|   |    |
|---|----|
| 4.2. Diagnostics de convergence, stationnarité, analyses de sensibilité         | 6h |
| <b>TD B : introduction à WinBUGS et interface avec R (NM+autre intervenant)</b> | 8h |
| <b>TD C : Statistiques descriptives et inférence: retour</b>                    |    |
| C1 Deux situations courantes : version MCMC                                     |    |
| C1.1 Une moyenne (EAS)  | 1h |
| C1.1 Une proportion (NM)  | 1h |
| C2 Autres cas   |    |
| C2.1 Poisson, deux comptes (EAS)  | 1h |
| C2.2 Gamma (NM)   | 1h |
| C2.3 Deux moyennes (EAS)  | 1h |
| C2.4 Deux proportions (NM)  | 1h |
| <b>5. Modélisation</b>  |    |
| 5.1. ANOVA 1F et 2F(NM)   | 1h |
| 5.2. Régression linéaire simple et multiple (EAS)                               | 2h |
| 5.3. ANCOVA (EAS)   | 1h |
| 5.4. GLM : régression logistique (NM)   | 1h |
| 5.5. GLM : régression de Poisson (EAS)  | 1h |
| 5.6. Table de contingence (NM)  | 2h |
| 5.7. Modèles de survie (EAS)  | 2h |
| 5.8. Comparaisons multiples (NM)  | 1h |
| 5.9. Autres (NM) : concordance (kappa,ICC), ...                                 | 1h |
| <b>6. Adéquation des modèles (EAS)</b>  | 4h |
| <b>7. Données manquantes, données aberrantes (NM)</b>                           | 8h |
| <b>TD D : équivalent de modèles mixtes (EAS+autre intervenant)</b>              | 8h |
| D1 Données répétées   |    |
| D2 Modèles hiérarchiques (avec modèle conditionnel)                             |    |
| D3 Effets aléatoires  |    |
| D4 Meta-analyses  |    |
| <b>TD E : Essai thérapeutique (NM+autre intervenant)</b>                        | 6h |
| E1 Comparaison de deux groupes  |    |
| E2 Essais séquentiels   |    |
| E3 Nombre nécessaire de sujets  |    |
| E4 Nombre de sujets à traiter   |    |
| <b>8. Epistémologie bayésienne et conclusion</b>                                |    |
| 8.1. Principe de vraisemblance (NM)   | 1h |
| 8.2. Popper a tort (contre la réfutation) (NM)                                  | 1h |
| 8.3. Synthèse argumentée Bayes-Laplace vs fréquentistes (EAS)                   | 2h |

Sur le **site internet** institutionnel du Laboratoire de Biostatistique et Informatique Médicale les étudiants disposeront des documents pédagogiques à l'issue de chaque cours et TD (diapositives de cours). En parallèle, les sujets traités en TD feront l'objet d'un dépôt sur le site Buxiv (<http://buxiv.unistra.fr>) sous la forme problème-résolution-code et un niveau de difficulté minimal. Ainsi chaque problème sera résumé dans un fichier, la méthode de résolution exposée dans un autre et les données ainsi que le code (R et/ou WinBUGS) joints.

- **Examens :**

L'obtention du DU est conditionné à la **présence** à toutes les heures de cours et de TD ainsi qu'à l'obtention ;

- de la **moyenne** à une épreuve de rédaction **d'un rapport d'étude finale**  
et
- de la **moyenne** à la rédaction **des rapports d'étude** portant sur chaque TD.

L'étudiant n'ayant pas la moyenne au rapport final pourrait le représenter l'année suivante sans devoir assister aux cours ni TD.

▪ **Par rapport à l'objectif principal, les compétences professionnelles attendues**, en termes de « savoir faire » sont les suivantes :

- Maitriser des deux logiciels pour l'ensemble des analyses (logiciel bayésien : WinBUGS, logiciel plus généraliste : R Core Team)
- Savoir manipuler un jeu de données (nettoyage, recherche de données manquantes, descriptif,...)
- Savoir mener une analyse inférentielle simple (descriptif et univarié)
- Savoir décider d'un modèle adéquat compte tenu de la question initiale et du contexte, savoir l'implémenter et l'utiliser
- Maitriser la rédaction d'un rapport détaillé d'analyse
- Savoir rédiger les sections « méthode » et « résultats » d'un article scientifique

- **Evaluation du DU**

Un questionnaire anonyme sera remis à chaque étudiant à la fin de chaque journée d'enseignement ainsi qu'un questionnaire final, concernant l'ensemble du DU. Pour chaque cours, l'étudiant se prononcera (« correct » ou « à améliorer ») sur différents

aspects : réponse aux objectifs du DU, réponse aux attentes de l'étudiant (en précisant quelles étaient ces attentes), qualités pédagogiques de l'intervenant (sur chacun des cours).