

# Présentation du parcours Biostatistique de la spécialité Recherche en santé publique du Master 2 Santé Publique de l'université Paris-Sud, cohabilité Paris Descartes

A. Chambaz, L. Meyer, P. Tubert-Bitter

22 juin 2010

## Table des matières

<b>1</b>	<b>Positionnement du parcours Biostatistique dans la spécialité Recherche en santé publique du M2 Recherche</b>	<b>1</b>
1.1	Hiérarchie. . . . .	1
1.2	Objectif. . . . .	1
1.3	Parti-pris. . . . .	2
<b>2</b>	<b>Programme du parcours Biostatistique</b>	<b>2</b>
2.1	Programme proposé. . . . .	2
2.2	Contenu des cours. . . . .	3

## 1 Positionnement du parcours Biostatistique dans la spécialité Recherche en santé publique du M2 Recherche

### 1.1 Hiérarchie.

Master 2 Santé Publique

→ Spécialité Recherche en Santé Publique (biostatistique, épidémiologie, recherche clinique, économie de la santé)

→ Parcours Biostatistique

### 1.2 Objectif.

Offrir une formation de haut niveau en biostatistique orientée “santé” à des à des étudiants issus de filières mathématiques appliquées, à des élèves en dernière année d'école d'ingénieurs, ainsi qu'à des étudiants de formation “santé” (médecins, pharmaciens, odontologistes, vétérinaires, biologistes).

### 1.3 Parti-pris.

Former des promotions dont les étudiants viennent d'horizons divers ; assurer un socle commun de connaissance en statistique mathématique et de maîtrise initiale du langage de programmation R grâce à deux UE couplées concentrées sur deux semaines au début du premier semestre.

## 2 Programme du parcours Biostatistique

### 2.1 Programme proposé.

Le programme affiche une coloration *originale, différenciée et de pointe*. Chaque cours compte une ou deux séances de consacrées aux applications sous R. Convention : (m) mise-à-niveau, (x) obligatoire, (o) optionnel, (c) mutualisé avec les autres parcours

1. Cours du mois de septembre (mise-à-niveau) et des mois d'octobre et novembre :
  - (o) *Matrices, pas d'examen*
  - (o) *Mise à niveau en épidémiologie, pas d'examen*
  - (o) *Mise à niveau en essais thérapeutiques, pas d'examen*
  - (x,m) *Socle commun Statistique mathématique, 18h, 2 ECTS* (cours de mise-à-niveau)
  - (x,m) *Socle commun Biostatistique avec R, 18h, 2 ECTS* (cours de mise-à-niveau)
  - (x) *Analyse de survie, données censurées, 20h, 3 ECTS*
  - (x,c) *Analyse causale, 16h, 3 ECTS*
  - (x) *Données longitudinales et répétées, pharmaco-cinétique/métrie, 18h, 3 ECTS*
  - (x) *Recherche en... , 15h, 2 ECTS*  
un moment privilégié au cours duquel tous les étudiants de la spécialité se retrouvent
  - (o) *Modèle linéaire généralisé, 18h, 3 ECTS*
  - (o) *Epidémiologie quantitative, 21h, 3 ECTS*
  - (o) *Mesure en recherche clinique et épidémiologique, 13.5h, 3 ECTS*
  - (o) *Epidémiologie descriptive et sondages, 15h, 3 ECTS*
2. Cours des mois de décembre et janvier :
  - (x,c) *Modélisation bayésienne pour la santé, 15h, 3 ECTS*
  - (x) *Réseaux bayésiens pour l'analyse statistiques de données post-génomiques, 15h, 3 ECTS*
  - (o,c) *Analyses cliniques à schémas adaptatifs, 15h, 3 ECTS*
  - (o,c) *Tests multiples en santé, 15h, 3 ECTS*
  - (o,c) *Statistique pour la pharmacoépidémiologie et la pharmacovigilance, 15h, 3 ECTS*
  - (o) *Estimation fonctionnelle pour la survie, 15h, 3 ECTS*
  - (o) *Algorithmes stochastiques pour la santé, 15h, 3 ECTS*
  - (o) *Modélisation markovienne pour la santé, 15h, 3 ECTS*
  - (o) *Séries chronologiques, 12h, 2 ECTS*

Les descriptions des modules sont reportées en fin de document.

Les modules obligatoires rapportent **21 ECTS**. Nécessité pour les étudiants de choisir au moins trois modules parmi les dix modules proposés en option, afin d'atteindre le compte de 30 ECTS.

**Les étudiants qui suivent *Analyses cliniques à schémas adaptatifs* doivent impérativement suivre trois séances du cours *Essais clinique* (dates 18/11, 25/11, 2/12).**

Pour la lecture de l'emploi-du-temps, Option1a et Option1b sont l'une de *Analyses cliniques à schémas adaptatifs*, *Tests multiples pour la santé* et *Statistique pour la pharmacoépidémiologie et la pharmacovigilance* (options ouvertes aux autres parcours). Option2 est l'une de *Algorithmes stochastiques pour la santé*, *Estimation fonctionnelle pour la santé* et *Modélisation markovienne pour la santé* (options réservées aux étudiants du parcours Biostatistique).

## 2.2 Contenu des cours.

Chaque cours compte une ou deux séances de consacrées aux applications sous R. Convention : (m) mise-à-niveau, (x) obligatoire, (o) optionnel, (c) mutualisé avec les autres parcours

1. Cours du mois de septembre (mise-à-niveau) et des mois d'octobre et novembre :

(o) **Matrices, pas d'examen**

M. Chavance (INSERM)

La mise à niveau de calcul matriciel doit permettre aux étudiants de comprendre et d'utiliser les notations matricielles. Il est construit sur la présentation matricielle de la régression linéaire et de l'estimateur des moindres carrés.

(o) **Mise à niveau en épidémiologie, pas d'examen**

L. Meyer (INSERM, Paris 11)

Cet enseignement est destiné aux étudiants de toutes les options dont la formation n'a pas inclus d'équivalent d'un module d'épidémiologie de niveau M1. Cette mise à niveau couvre les différentes notions suivantes : (i) grands types d'enquêtes épidémiologiques, (ii) mesures de fréquence, (iii) mesures d'association, (iv) biais de sélection et biais de mesure, (v) confusion et interaction.

(o) **Mise à niveau en essais thérapeutiques, pas d'examen**

M-C. Le Deley (Gustave Roussy) et Y. de Rycke (Institut Curie)

Cet enseignement est destiné aux étudiants de toutes les options qui n'ont pas suivi l'UE de Recherche clinique ni le STARC. Il couvre les notions de (i) démarche expérimentale dans l'évaluation d'un nouveau traitement, (ii) définition des essais de phase I, II, III, (iii) essai randomisé, de sa planification à son analyse, (iv) analyse des risques d'erreur, du nombre de sujets nécessaires, (v) essais de supériorité, d'équivalence, de non infériorité.

(x,m) **Socle commun Statistique mathématique, 18h, 2 ECTS** (cours de mise-à-niveau)

P. Tubert-Bitter (INSERM, Paris 11)

Cours de mise-à-niveau *sur deux ou trois semaines en début d'année* pour s'assurer d'un socle commun de connaissance, couplé avec *socle commun Biostatistique avec R*. Contenu : (i) notions de convergence, théorème limite centrale ; (ii) vraisemblance, score, matrice d'information ; (iii) exhaustivité ; (iv) famille exponentielle ; (v) estimation, inégalité de Cramer-Rao ; (vi) maximum de vraisemblance ; (vii) théorie générale des tests, lemme de Neyman-Pearson ; (viii) tests exacts, tests de permutation ; (ix) test du rapport de vraisemblance, du score, de Wald, tests UPP et UMPU ; (x) tests de rangs ; (xi) tests en présence de paramètre nuisible ; (xii) comparaisons de tests.

(x,m) **Socle commun Biostatistique avec R, 18h, 2 ECTS** (cours de mise-à-niveau)

Y. Rozenholc (UFR MI, Paris 5)

Cours de mise-à-niveau pour s'assurer d'un socle commun de connaissance, couplé avec *socle commun Statistique mathématique*. Cette UE comprendra une initiation à la maximisation de fonctions de plusieurs variables.

(x) **Analyse de survie, données censurées, 20h, 3 ECTS**

C. Guihenneuc (UFR Biomédicale, Paris 5) et ML. Taupin (Université d'Evry)

C. Guihenneuc et M-L. Taupin assurent ce cours dans le cadre du Master mention Mathématiques et Informatique, spécialité Ingénierie Mathématique pour les Sciences

du Vivant de l'UFR Math-Info de Paris 5. Propose de permettre aux étudiants du parcours Biostatistique de suivre le cours de ML dans ce cadre.

Les objectifs de ce cours sont d'analyser des données issues d'une étude de survie, de savoir interpréter les résultats et de savoir discuter les hypothèses des différents modèles. Dans un premier temps, la modélisation des durées de survie via des approches non paramétriques ou paramétriques sera étudiée. Dans un second temps, l'ajustement sur des covariables sera considéré via des modèles de régression (paramétriques ou semi-paramétriques) comme le modèle de Weibull et le modèle de Cox.

(x,c) *Analyse causale*, **16h, 3 ECTS**

A. Chambaz (UFR MI, Paris 5)

Diagrammes acycliques de causalité, systèmes d'équations structurelles et intervention, identifiabilité de probabilité causale, confusion ; estimation robuste de paramètres causaux (via modèles de travail de type linéaires généralisés).

(x) *Données longitudinales et répétées, pharmaco-cinétique/métrie*, **18h, 3 ECTS**

F. Mentré (INSERM, Paris 7)

Analyse statistique des données longitudinales, modèles à effets mixtes, HMM.

(x) *Recherche en...*, **15h, 2 ECTS**

Série de séminaires en commun avec autres parcours pour faire le tour d'un sujet (exemples récents : *Epidémiologie prénatale* ou *Données de survie*).

(o) *Modèle linéaire généralisé*, **18h, 3 ECTS**

J-P. Jaïs (Necker, Paris 5)

Le point sur les modèles linéaires et linéaires généralisés.

(o) *Epidémiologie quantitative*, **21h, 3 ECTS**

J. Bouyer (INSERM, Paris 11)

Notions de facteur de risque et de mesure d'association ; méthodes d'analyse multivariée en épidémiologie pour l'étude de l'association entre une maladie et un facteur de risque, exemple du modèle logistique : théorie et application.

(o) *Mesure en recherche clinique et épidémiologique*, **13.5h, 3 ECTS**

Joël Coste (Cochin, Paris 5)

L'objectif est d'exposer les principes de la mesure et les conséquences des erreurs de mesure en recherche clinique et en épidémiologie (erreurs non différentielles) ; d'initier la planification et à l'interprétation d'une étude métrologique pour des mesures nominales, ordinales, quantitatives et pour des échelles composites ; d'initier aux principales utilisations et à l'interprétation des échelles composites dans le domaine de la psychiatrie et de la qualité de la vie.

(o) *Epidémiologie descriptive et sondages*, **15h, 2 ECTS**

J. Warszawski (INSERM) et J. Bouyer (INSERM)

Le but de ce cours est de fournir des méthodes pour estimer la fréquence de morbidité ou la distribution de facteurs de risque dans une population et étudier leurs variations dans l'espace et dans le temps. Ce cours comportera deux volets : (a) l'utilisation des statistiques sanitaires ; (b) la mise en œuvre des enquêtes ad hoc permettant de répondre à des objectifs éventuellement multiples et comportant au moins un objectif descriptif.

2. Cours des mois de décembre et janvier :

(x,c) *Modélisation bayésienne pour la santé*, **15h, 3 ECTS**

C. Guihenneux (UFR Biomédicale, Paris 5)

Distributions *a priori* et *a posteriori*, inférence bayésienne (fonction de coût, risque et risque intégré associé, estimateur bayésien) ; calculs et mise en pratique ; méthode MCMC par chaînes de Markov ; applications via différentes modélisations.

(x) *Réseaux bayésiens pour l'analyse statistiques de données post-génomiques*, **15h, 3 ECTS**

G. Nuel (CNRS, Paris 5)

Propagation dans les réseaux bayésiens, modélisation des séquences biologiques, chaînes

de Markov cachées, reconstruction phylogénétique, données familiales en génétique, réseaux de gènes.

(o,c) *Analyses cliniques à schémas adaptatifs*, **15h, 3 ECTS**

A. Chambaz (UFR MI, Paris 5)

Analyses cliniques à schémas fixes, analyses cliniques séquentielles, analyses cliniques séquentielles adaptatives.

(o,c) *Tests multiples en santé*, **15h, 3 ECTS**

Y. Rozenholc (UFR MI, Paris 5)

Le but de ce cours est d'introduire la notion de test multiple et son usage en statistique appliquée. Motivée par le besoin de comparer des signaux obtenus par tomographie pour l'étude de l'angiogenèse tumorale ou bien encore des niveaux d'expressions sur la base de données micro-array, le problème élémentaire de la comparaison à zéro de la moyenne d'un vecteur gaussien en grande dimension est introduit. Le cours se développe alors autour des concepts fondamentaux de (i) contrôle de Bonferroni, (ii) False Discovery Rate, (iii) sélection de modèles et multi-tests, (iv) ondelettes et tests. Quelques résultats importants de statistique mathématique seront présentés durant le cours, en se focalisant non pas sur leurs preuves mais sur le commentaire des hypothèses et de la nature des résultats.

(o,c) *Statistique pour la pharmacoépidémiologie et la pharmacovigilance*, **15h, 3 ECTS**

P. Tubert-Bitter (INSERM, Paris 11)

Ce cours est consacré à l'après-commercialisation des médicaments. Y sont abordées (i) les notions de vigilance, de détection de signaux (data-mining), de surveillance (par méthodes séquentielles); (ii) les notions de biais spécifique, de score de propension, d'études cas-croisé, d'études de séries de cas dans le cadre de la pharmacoépidémiologie; (iii) l'étude de sources de données de grande taille (comme les bases de données médico-administratives), avec notamment la mise au point de notifications spontanées d'effets indésirables.

(o) *Estimation fonctionnelle pour la survie*, **15h, 3 ECTS**

F. Comte (UFR MI, Paris 5)

Estimateurs à noyau, et sélection de fenêtre; estimateur par projection et pénalisation. Application à l'estimation de densité en présence de censure, du taux de risque instantané, à la durée de vie résiduelle, à la répartition conditionnelle avec censure par intervalle.

(o) *Algorithmes stochastiques pour la santé*, **15h, 3 ECTS**

J-C. Fort (UFR MI, Paris 5)

Méthode de Gibbs, algorithme SAEM, algorithme de Métropolis-Hastings.

(o) *Modélisation markovienne pour la santé*, **15h, 3 ECTS**

A. Samson (IUT Paris Descartes, Paris 5) ou F. Comte (UFR MI, Paris 5)

Les objectifs de ce cours sont de présenter des modèles markoviens utilisés en santé. Dans un premier temps, des modèles de propagation d'épidémie du type SIR, SIS entre autres seront étudiés. Dans un second temps, nous considérerons des modèles stochastiques construits par perturbation aléatoire de systèmes différentiels utilisés en pharmacocinétique, oncologie, diabétologie, neurologie, etc.

(o) *Séries chronologiques*, **12h, 2 ECTS**

L. Watier (INSERM, Paris 11)

Une introduction à l'analyse des séries temporelles.