



Programme des enseignements

3^e année

Filière Statistique pour les sciences de la vie
ANNEE SCOLAIRE 2018 / 2019



École nationale
de la statistique
et de l'analyse
de l'information

FILIÈRE STATISTIQUE POUR LES SCIENCES DE LA VIE

ANNÉE SCOLAIRE 2018/2019

BIOSTATISTICS SPECIALIZATION

2018/2019 ACADEMIC YEAR

Table des matières

Présentation de la filière.....	5
Thématiques abordées.....	5
Option Formation Par la Recherche	5
Les entreprises partenaires	6
Descriptifs des enseignements du tronc commun	9
UE – Tronc commun	11
Droit du Travail.....	11
Anglais	13
Sport.....	15
Descriptifs des enseignements de la filière	17
UE - Méthodologie statistique 1.....	19
Modélisation non linéaire	19
Compléments de Statistique Bayésienne.....	20
Statistique des processus	21
Régularisation.....	22
UE - Méthodologie statistique 2.....	23
Plans d'expériences.....	23
Modèles mixtes	24
Compléments de modèles de durée	26
UE - Méthodologie statistique 3.....	28
Mesures de qualité de vie	28
Modélisation compartimentale.....	30
Traitement des données manquantes dans les essais cliniques	31
Méta-Analyse	33
UE - Essais cliniques.....	35
Essais cliniques : méthodologie et analyses statistiques	35
Pharmacométrie.....	36
Projet d'essais cliniques	37
UE - Epidémiologie	38
Epidémiologie quantitative	38
Atelier d'épidémiologie.....	39
Projet de santé et environnement	40
UE - Statistique pour données Omics.....	41
Introduction à l'analyse de données « Omics »	41
Analyse des données « Omics »	42
Projet de génétique.....	43
UE - Séminaires professionnels	44
Analyse sensorielle.....	44
Introduction au développement clinique.....	45

Statistical methods for Phase I dose-finding studies in oncology	46
Multiplicité dans les essais cliniques	48
L'analyse de survie en épidémiologie et en pharmaco-épidémiologie	50
Analyse de combinaisons de drogues	52
Evaluation médico-économique	53
BIG DATA, data Mining et machine Learning	54
Deep Learning	56

Présentation de la filière

Thématiques abordées

Pour répondre aux exigences de la filière, en plus du tronc commun à tous les élèves de troisième année, l'enseignement est divisé en six unités d'enseignement (UE).

Après des compléments en statistique, notamment en données de survie, modèles mixtes, régularisation et plans d'expériences, les enseignements apportent les outils nécessaires pour une spécialisation dans le domaine de l'expérimentation. Les cours d'épidémiologie, d'essais cliniques et l'analyse des données Omics permettent en particulier aux étudiants de recevoir une solide formation pour des applications dans le secteur de la santé.

Tout au long de l'année, les étudiants auront à gérer différents projets. Les trois principaux correspondent aux cours d'épidémiologie, d'essais cliniques et l'analyse des données Omics. Ils permettent de compléter et de mettre en pratique les connaissances théoriques vues en cours. Ces projets sont aussi l'occasion de développer l'esprit d'équipe par des travaux en groupes de plusieurs étudiants.

Transversalement à ces unités d'enseignement, les applications en informatique (R, SAS, WINBUGS...etc) sont omniprésentes. Des séminaires professionnels présentent la richesse des métiers offerts en statistique pour les sciences de la vie. Ils sont l'occasion d'une présentation, par les praticiens, des outils ou modèles statistiques utilisés dans les entreprises et instituts de recherche.

La langue anglaise n'est pas négligée puisque des enseignements donnent lieu à l'écriture de mémoires en anglais et deux enseignements sont dispensés dans cette langue.

Option Formation Par la Recherche

L'Ensay offre la possibilité, aux élèves de 3^e année qui le souhaitent, de se préparer en vue d'une carrière de chercheur en entreprise au sein des services Recherche et Développement ou dans le secteur académique. Dans le cadre de l'option formation par la recherche (Ofpr), ces élèves bénéficient d'un aménagement de leur scolarité leur permettant de suivre au choix un des trois masters recherche suivants :

- Master Recherche, mention Mathématiques et applications, spécialité Statistique Mathématique
- Master Recherche, mention Santé Publique, spécialité Modélisation en Pharmacologie Clinique et Epidémiologie (MPCE)
- Master Recherche, mention Bio-informatique, spécialité Bio-informatique en santé

A l'issue de ce parcours, ils pourront poursuivre cette formation par une thèse académique ou de type Cifre (Convention Industrielle de Formation par la Recherche).

Les thèses académiques sont en général encadrées dans des laboratoires de recherche tels que ceux du CNRS ou de l'Inserm. En ce qui concerne les entreprises signataires de thèses Cifre ou organismes de recherche, on peut citer par exemple l'I.R.I.S. –Laboratoires Servier associé à l'INSERM ou encore l'IRSN associé à l'Université de Paris 11.

Les entreprises partenaires

La filière bénéficie de partenariats avec des acteurs économiques de premier plan. Ces partenariats permettent de développer des échanges privilégiés notamment via des cours, des séminaires professionnels et des stages.



Enseignements de 3^{ème} année
Filière Statistique pour les sciences de la vie

	Volume horaire				Crédits
	Cours	Atelier	Projet	Total	Coefficients
Tronc commun - UE-3A-00					
Droit du Travail	3	6		9	1,5
Anglais	30			30	2,5
Sport		30		30	
Total	33	36		69	4
Méthodologie statistique 1 UE-3ASV-01					
Modélisation non linéaire ¹	9	3		12	0,5
Compléments de Statistique bayésienne	9	6		15	1
Statistique des processus ¹	12	3		15	1
Régularisation	3	9		12	0,5
Total	33	21		54	3
Méthodologie statistique 2 UE-3ASV-02					
Plans d'expériences ¹	18			18	1
Modèles mixtes	15	9		24	2
Compléments de modèles de durée	18	6		24	2
Total	51	15		66	5
Méthodologie statistique 3 UE-3ASV-03					
Mesures de qualité de vie	12	6		18	1,5
Modélisation compartimentale	6	6		12	0,5
Traitement des données manquantes	6	6		12	0,5
Méta-analyse ²	12	6		18	1,5
Total	36	24		60	4
Essais cliniques UE-3ASV-04					
Essais cliniques ²	18			18	1,5
Pharmacométrie	6	12		18	1
Projet d'essais cliniques ²		6	18	24	2,5
Total	24	18	18	60	5
Épidémiologie UE-3ASV-05					
Épidémiologie quantitative	18			18	1,5
Atelier d'épidémiologie		18		18	1
Projet de santé et environnement			24	24	2,5
Total	18	18	24	60	5
Statistique pour données Omics UE-3ASV-06					
Introduction à l'analyse des données Omics	6	6		12	1
Analyse des données Omics	6	12		18	1
Projet de génétique			18	18	2
Total	12	18	18	48	4
Séminaires professionnels		63		63	
Total		63		63	
Total	207	213	60	480	30

¹ Enseignement commun aux filières Statistique pour les Sciences de la Vie et Génie statistique

² Enseignement commun aux filières Statistique pour les Sciences de la Vie et ISTS

Partie stage

UE Stages	Crédits
	30

Descriptifs des enseignements du tronc commun

UE – Tronc commun

Droit du Travail

Work Law

Cours : 3h • Atelier : 6h

Enseignant : Charlotte GRUNDMAN, Avocat au Barreau de Paris.

Correspondant : Ronan Le Saout

Objectif pédagogique :

La matière étant extrêmement vaste et complexe, il est ici proposé aux étudiants une approche didactique et vivante du sujet, l'objectif de l'enseignement étant de permettre aux étudiants qui travailleront dans un futur proche en entreprise d'avoir compris certaines notions pratiques essentielles en droit du travail.

Contenu de la matière

A cette fin, et hormis le cours d'amphi, il sera systématiquement proposé aux étudiants, après l'étude d'une notion, un exercice visant à mettre en pratique la notion abordée.

Afin de satisfaire le plus possible à cet objectif, il est ainsi proposé l'organisation suivante des cours :

Cours commun (3 heures) :

Chapitre 1 : Comprendre d'où l'on vient pour savoir où on va :

- Introduction historique au droit du travail
- Les sources du droit du travail
 - sources imposées,
 - sources négociées
- Ordre public absolu et ordre public social

Chapitre 2 : les instances de contrôle du droit du travail

- L'inspecteur du travail
- Les multiples juges du droit du travail
- Point sur la procédure prud'homale

Chapitre 3 : Formation et exécution du contrat de travail

- la qualification du contrat de travail : « faux artisans, faux auto-entrepreneurs et vrai salarié ».
- le contrat à durée indéterminée, norme juridique et sociale
- la période d'essai après la loi du 25 juin 2008 : définition, durée et rupture
- les principales clauses du contrat de travail :
 - la clause de mobilité
 - la clause de non-concurrence

Chapitre 4 : la rupture du contrat à durée indéterminée

- le licenciement pour motif personnel
- le licenciement pour motif économique
- la démission du salarié
- les autres modes de rupture

Les TD :

La première heure de cours sera consacrée à l'étude d'un chapitre. Cet exposé sera suivi d'une mise en situation pratique, où les étudiants devront par groupe répondre à un cas pratique. Un rapporteur sera désigné par groupe, et la notation se fera à cette occasion.

Chapitre 1 : La modification du contrat de travail

Modification du contrat de travail et changement des conditions de travail

- la durée du travail (focus sur le forfait-jour)
- le lieu de travail
- la rémunération

Chapitre 2 : Le recrutement

Chapitre 3 : les droits fondamentaux du salarié

- Le fait religieux en entreprise
- Vie personnelle et technologies de l'information et de la communication (TIC)
- La mise en place de moyens de contrôle via les TIC en entreprise
- Harcèlements
- Maladie et maternité du salarié

Langue d'enseignement

Français

UE – Tronc commun

Anglais

English

Cours : 30h (dont 15h d'aide au projet)

Enseignant : Divers intervenants

Correspondant : Todd DONAHUE

Objectif pédagogique

Les élèves qui n'ont pas passé ou qui n'ont pas réussi le TOEIC l'année dernière auront progressé dans les compétences requises – c'est à dire, la compréhension orale, la reconnaissance des erreurs, les pièges grammaticaux, et la compréhension écrite. Les autres auront acquis les compétences nécessaires pour affronter le monde professionnel. Ils auront vu les tournures qui aident à diriger et à participer à des réunions, à prendre des décisions, et à négocier. Ils se seront entraînés à faire des présentations. Ils auront rédigé un projet en anglais et préparé la soutenance de ce projet.

Contenu de la matière

Pour les élèves qui n'ont pas eu un score d'au moins 785 au TOEIC : pendant les 5 premières séances, la plupart des cours seront basés sur la préparation à cet examen. Les ressources informatiques de l'Ecole doivent aussi être mises à profit (pages Moodle, TOEIC Mastery), ainsi que les méthodes disponibles à la bibliothèque.

Pour les autres élèves, les cours seront organisés par groupe de niveau et conçus afin de les préparer à affronter le monde professionnel sur le plan international. Les thèmes suivants seront traités : « Leading meetings », « Interviews », « Presentations », « Taking decisions », et « Negotiating deals », et « Cultural and Political Current Events ».

Ensuite, les 5 dernières séances seront consacrées au travail de rédaction/correction des rapports faits en anglais dans chaque filière ainsi qu'à la préparation des soutenances orales. Chaque responsable de filière indiquera aux élèves, en début d'année, le projet concerné et les modalités de notation. Les élèves recevront des consignes détaillées avant de démarrer ces cinq séances, afin d'arriver à la première séance avec une première version ou extrait de leur rapport en anglais prêt pour correction et relecture.

Pré-requis

Aucun

Contrôle des connaissances

L'examen final prend la forme d'une simulation d'entretien d'embauche. Cet examen oral durera environ 25 minutes, sera noté, et permettra d'évaluer le niveau d'expression orale sur l'échelle CECRL*. Le CV et la lettre faite pour cet exercice seront évalués et feront partie de la note finale. Le niveau acquis apparaîtra sur le Supplément au diplôme. L'objectif de la CTI[†] pour tous les élèves ingénieurs est d'atteindre le niveau B2.

* le Cadre européen commun de référence pour les langues.

[†] la Commission des Titres d'Ingénieur.

Références bibliographiques

- Arbogast, B., *30 Days to the TOEIC Test*, Canada: Peterson's, 2002.
- Schramper-Azar, B., *Understanding and Using English Grammar*, New York: Longman, 1999.
- Buckwalter, Elvis, et al, *Boostez votre score au TOEIC-spécial étudiants*, Paris: Eyrolles, 2009.
- Gear, Jolene, *Cambridge Grammar and Vocabulary for the TOEIC Test*, Cambridge: Cambridge University Press, 2010.
- Lecomte, Stéphane, et al, *La Grammaire au TOEIC et au TOEFL : Mode d'emploi*, Paris: Ophrys, 2008.

- Loughed, Lin, *Tests complets pour le nouveau TOEIC* (4^{ème} ed.), Paris: Pearson Education France, 2008.
- MBA Center, *New TOEIC Study Book*, Paris: MBA Center Publications, 2007.

Langue d'enseignement

Anglais

Pour tout complément d'information, chaque élève peut consulter le Programme des enseignements : Langues étrangères, distribué au début de l'année académique.

UE – Tronc commun

Sport

Sport

TP : 30h

Enseignant : Divers intervenants

Correspondant : Julien LEPAGE

Cours facultatif

Objectif de la matière

L'objectif est d'amener les élèves à maintenir un esprit sportif, sortir du strict cadre académique et développer leurs capacités physiques.

Contenu de la matière

9 activités sportives sont proposées par l'école :

- Badminton
- Basket
- Cross-Training
- Football
- Hand-ball
- Tennis de table
- Tennis débutant
- Volley-ball
- Course à pied/préparation physique/coaching sportif

Outre les entraînements, les élèves inscrits peuvent être amenés à participer à des compétitions.

Prise en compte dans la scolarité

La participation à une activité sportive peut donner lieu à l'attribution d'un bonus ajouté sur la moyenne du semestre concerné. Le niveau de ce bonus est précisé dans une circulaire d'application en début d'année académique. Il varie selon l'assiduité aux séances, l'engagement et la participation aux compétitions tout au long de l'année.

Pour être définitive, la liste des élèves bénéficiant de ces bonus doit être validée par le directeur des études.

Un bonus peut être exceptionnellement attribué en dehors des activités sportives réalisées dans le cadre Ensay. Pour y prétendre, les élèves concernés doivent remplir les 3 conditions suivantes :

- pratiquer régulièrement une activité sportive et participer aux compétitions liées ;
- posséder un niveau national (voir très bon niveau régional suivant le sport en question) ;
- déposer une demande argumentée auprès de la direction des études et du service sport en début d'année scolaire, afin de faire valider le programme d'entraînement, des compétitions et les modalités de diffusion des performances.

Pour certains ayant des contraintes sportives, des aménagements horaires pourront d'ailleurs être ainsi envisagés si besoin.

Descriptifs des enseignements de la filière

UE - Méthodologie statistique 1

Modélisation non linéaire

Nonlinear Modeling

Cours : 9h • Atelier : 3h

Enseignant : Fabien NAVARRO (Ensay)

Correspondant : Salima EL KOLEI

Enseignement destiné aux élèves des filières « Statistique pour les sciences de la vie » et « Génie statistique »

Objectif pédagogique

Ce cours a pour but d'introduire les modèles de régression paramétrique non-linéaires et une approche non paramétriques alternative à la méthode des polynômes locaux, pour l'estimation d'une régression.

Contenu de la matière

1. Extension du modèle linéaire
 - 1.1. base de fonctions
 - 1.2. les splines
 - 1.3. les ondelettes
2. Modèle paramétrique non linéaire
 - 2.1. M-estimation
 - 2.2. intervalles de confiance et tests
3. Splines pénalisées et modèles mixtes

Pré-requis

Les cours de régression de deuxième année

Contrôle des connaissances

A déterminer

Références bibliographiques

- ANTONIADIS, J. BERRUYER, R. CARMONA. *Régression non linéaire et applications*. Economica, 1992
- R.L. EUBANK. *Nonparametric Regression and Spline Smoothing* : 2nd Edition. Dekker 1999
- D. RUPPERT, M. P. WAND, R. J. CARROLL. *Semiparametric Regression*. Cambridge University Press, 2003
- J. O. RAWLINGS, S. G. PANTULA, D. A. DICKEY. *Applied Regression Analysis. A Research Tool (2nd ed.)*. Springer, 1998.
- T. P. RYAN. *Modern Regression Methods (2nd ed.)*. Wiley, 2008

Langue d'enseignement

Français

UE - Méthodologie statistique 1

Compléments de Statistique Bayésienne

Advanced Bayesian Statistics

Cours : 9h • Atelier : 6h

Enseignant : Sophie ANCELET (IRSN)

Correspondant : Lionel TRUQUET

Enseignement destiné aux élèves de la filière "Génie Statistique", « Ingénierie statistique des territoires et de la santé » et « Statistique pour les sciences de la vie »

Objectif pédagogique

A l'issue de cet enseignement, les élèves devront maîtriser les connaissances de base pour l'analyse de données par approche statistique bayésienne. Les problèmes traités seront empreints aux sciences de la vie où l'emploi des méthodes bayésiennes progresse considérablement. Cependant, le caractère général de ces méthodes, utiles dans bien d'autres domaines d'application, sera clairement établi. À l'issue de ce module, l'étudiant doit être capable de mettre en œuvre les méthodes enseignées pour mener des inférences bayésiennes de données, notamment à l'aide des logiciels WINBUGS, OPENBUGS et JAGS.

Contenu de la matière

Un rappel de cours est fait concernant les principes de la modélisation statistique bayésienne. L'accent sera mis sur l'analyse bayésienne par les méthodes de Monte Carlo par Chaînes de Markov (MCMC). Aux travers d'exemples, seront abordés les notions de graphe d'indépendance conditionnelle, réseau bayésien, convergence des Chaînes de Markov, inférence, prédiction, validation et comparaison de modèles dans un cadre bayésien. Les exemples seront traités sous le logiciel WINBUGS ou JAGS en salle informatique.

Pré-requis

Cours de statistique bayésienne en deuxième année

Contrôle des connaissances

Projet court

Références bibliographiques

- Collectif BIOBAYES: Albert I., Ancelet S., David O., Denis J.B., Makowski D., Parent E., Soubeyrand S. (2015) Méthodes statistiques bayésiennes. Bases théoriques et applications en alimentation, environnement et génétique. *ELLIPSES*, ISBN : 978234000501
- Carlin, B. P. and Louis, T.A. (2009). Bayesian Methods for Data Analysis. Chapman & HALL/CRC, third edition, (535 pp.)
- Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B (2004). Bayesian data analysis. Texts in Statistical Science. Chapman & HALL/CRC, second edition, (668 pp.)
- Robert, C. P. (2001). The Bayesian choice. Springer, (second edition) (604 pp.)
- Lunn, D.J., Thomas, A., Best, N. and Spiegelhalter, D. (2000). WinBUGS -- a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10: 325-337.
- Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (1996). Markov chain Monte Carlo in practice. Chapman and Hall, (486 pp.)

Langue d'enseignement

Français

UE - Méthodologie statistique 1

Statistique des processus

Statistics of Stochastic Processes

Cours : 12h Atelier : 3h

Enseignant : Myriam VIMOND (Ensay)

Correspondants : Salima EL KOLEI

Enseignement avec une partie commune aux élèves des filières « Statistique pour les sciences de la vie » et « Génie statistique »

Objectif pédagogique

Il s'agit de présenter des modélisations des principaux phénomènes aléatoires dépendants du temps rencontrés dans l'industrie et en sciences de la vie (hors-séries temporelles et traitement du signal) à partir de processus markoviens. Dans chaque cas la présentation portera autant sur les outils probabilistes que sur l'inférence statistique dans ces modèles.

Contenu de la matière

1. Compléments sur les chaînes de Markov. Théorèmes ergodiques. Estimation de la matrice de transition.
2. Introduction aux chaînes de Markov à temps continu.
3. Exemples d'utilisation des chaînes de Markov en oncologie et en épidémiologie.
4. Introduction aux équations différentielles stochastiques et applications en pharmacocinétique.

Pré-requis

Les cours de Probabilités et Statistique de première année et le cours de chaînes de Markov

Contrôle des connaissances

Examen écrit.

Références bibliographiques

- ASMUSSEN, S. Applied probability and queues. Second edition. Springer 2003.
- BOSQ, D. Statistique mathématique et statistique des processus. Lavoisier, 2012.
- DACUNHA-CASTELLE, D., DUFLO, M. Probabilités et statistiques II, Problèmes à temps mobile. Masson, 1993.
- DELMAS, J-F., JOURDAIN, B. Modèles aléatoires. Applications aux sciences de l'ingénieur et du vivant. Springer 2006.
- FOATA, D., FUCHS, A. Processus Stochastiques (2^e éd.). Dunod 2004.
- FUCHS, C., Inference for Diffusion Processes, With Application in Life Sciences. Springer 2013.
- KIMMEL, M., AXELROD, D. Branching processes in biology. Springer 2002.
- PARDOUX, E. Processus de Markov et applications. Algorithmes, génome et finance. Dunod 2007.

Langue d'enseignement

Français

UE - Méthodologie statistique 1

Régularisation

Regularisation

Cours : 3h • Atelier : 9h

Enseignant Julien CHIQUET (Université de Rennes 1)
Correspondant Lionel TRUQUET

Enseignement destiné aux élèves de la filière « Statistique pour les sciences de la vie »

Objectif pédagogique

L'objet de ce module est de présenter une introduction aux méthodes régularisées qui sont devenues un outil indispensable à l'analyse de données, en particulier issues des sciences du vivant. Il vise également à fournir aux étudiants des compétences pratiques quant à l'utilisation de ces méthodes à l'aide du logiciel R.

Contenu de la matière

Les deux journées commenceront par 3 heures de cours

- Jour 1: les limites du modèles linéaires; la régression stepwise; introduction à la régularisation
- Jour 2: la régularisation ridge, le lasso et leurs variantes pour le modèle linéaire

Les après-midi sont consacrées à des séances machines. Les méthodes régularisées classiques seront évaluées sur des données simulées dont le protocole sera mis en place par les étudiants. Des données issues de la génomique seront ensuite analysées.

Pré-requis

Régression linéaire

Contrôle des connaissances

L'évaluation se fera sous la forme d'un rapport associé aux séances machines.

Langue d'enseignement

Français

UE - Méthodologie statistique 2

Plans d'expériences

Experiment Design

Cours : 18h

Enseignant Walter TINSSON (Université de PAU)
Correspondant Lionel TRUQUET

Enseignement destiné aux élèves des filières « Statistique pour les sciences de la vie » et « Génie statistique », et « Ingénierie statistique des territoires et de la santé »

Objectif pédagogique

Comprendre les principes fondateurs des stratégies d'expérimentation
Apprendre à choisir, construire un dispositif expérimental
Acquérir les outils d'analyse des plans d'expériences (utilisation du logiciel R)

Contenu de la matière

Principes fondateurs et présentation des grandes familles de plans
Les Outils d'Analyse : modèle linéaire, analyse de la variance
Plans factoriels complets et fractionnaires, Optimalité
Expériences Accélérées

Pré-requis

Ce qui a été vu en classes préparatoires et lors des deux premières années de l'Ecole est largement suffisant pour suivre le cours (Calcul matriciel, Optimisation, Bases de la régression et d'analyse de variance...)

Contrôle des connaissances

Examen pratique en salle informatique

Références bibliographiques

- AZAIS, J.-M., BARDET, J.-M. Le modèle linéaire par l'exemple (2^e éd.). Dunod, 2012.
- DROESBEKE, J.-J., FINE, J., SAPORTA, G. (Eds Scientifiques). Plans d'expériences : Applications à l'entreprise. Technip, 1997.

Langue d'enseignement

Français

UE - Méthodologie statistique 2

Modèles mixtes

Mixed Models

Cours : 15h • Atelier : 9h

Enseignant : Marie-Cécile FOURNIER (Université de Nantes)

Correspondant : Lionel TRUQUET

Enseignement destiné aux élèves de la filière « Statistique pour les sciences de la vie »

Objectif pédagogique

A l'issue de cet enseignement, les élèves devront connaître les fondements de la théorie statistique du modèle mixte afin d'en assurer une bonne compréhension, maîtrise et interprétation.

Contenu de la matière

Dans le modèle linéaire, on peut prendre en compte diverses structures de corrélation (intra-classe, temporelle, spatiale) grâce à une version dite "mixte" du modèle qui fait intervenir à la fois des effets fixes et des effets aléatoires.

Il est à noter que nombre de problèmes peuvent être abordés sous une formulation de modèle mixte, cf. le filtre de Kalman et les splines cubiques par exemple. Ce type de modèles peut faire l'objet d'une grande variété de traitements et d'interprétation statistique (algorithme EM, méthodes MCMC, approche bayésienne). Le concept permet également de nombreuses extensions (modèle linéaire généralisé, modèle non linéaire).

Le modèle mixte connaît actuellement un développement important de ses applications dans maints secteurs de l'industrie, des sciences économiques et sociales, de la biologie et de la médecine. Il est servi par de bons logiciels tels que les Proc GLM, Mixed, Nlmixed, Glimmix de SAS, Asreml, nlme de R et S plus, Monolix, Genstat et Winbugs.

1. Introduction: écriture générale, hypothèses, exemples
2. Prédiction des effets aléatoires: meilleure prédiction, meilleure prédiction linéaire et meilleure prédiction linéaire sans biais (BLUP)
3. Equations du modèle mixte d'Henderson
4. Inférence par la méthode du maximum de vraisemblance : estimation et tests d'hypothèse des effets fixes et des composantes de variance
5. Concept de vraisemblance résiduelle (REML) : présentation classique et présentation bayésienne
6. Théorie de l'algorithme EM
7. Application aux composantes de variance
8. Aperçu sur le modèle linéaire généralisé mixte

Pré-requis

Lois de probabilité, algèbre linéaire et théorie du modèle linéaire (régression, anova)

Contrôle des connaissances

Examen écrit.

Références bibliographiques

- B.P. CARLIN, T.A. LOUIS, *Bayesian Methods for Data Analysis*, Chapman & Hall/CRC Press, 3rd édition, 2009
- A.P. DEMPSTER, N.M. LAIRD, D.B. RUBIN, *Maximum likelihood from incomplete data via the EM algorithm*, J. R. Statist. Soc. B 39 1-38, 1977
- P.J. DIGGLE, P. HEAGERTY, K-Y LIANG, S. ZEGER, *Analysis of Longitudinal Data*, Oxford Statistical Science Series, 2nd édition, 2002

- L. FAHRMEIR, G. TUTZ, *Multivariate statistical modelling based on generalized linear models* (2nd ed.), Springer Verlag, Berlin, 2001
- H.O. HARTLEY, J.N.K. RAO, Maximum likelihood estimation for the mixed analysis of variance model, *Biometrika* 54 93-108, 1967
- M. FITZMAURICE, N LAIRD, J.H. WARE, *Applied Longitudinal Analysis* (2nd ed.), Wiley series in probability and statistics, 2011
- D.A. HARVILLE, Maximum likelihood approaches to variance component estimation and related problems, *J. Am. Stat. Assoc.* 72, 320-338, 1977
- C.R. HENDERSON, *Applications of linear models in animal breeding*, University of Guelph, Guelph, 1984 (en ligne)
- N.M. LAIRD, J.H. WARE, *Random effects models for longitudinal data*, *Biometrics* 38 963-974, 1982
- K.Y. LIANG, S.L. ZEGER, Longitudinal data analysis using generalized linear models, *Biometrika* 73 13-22, 1986
- P. Mc CULLAGH, J.A. NELDER, *Generalized linear models*, 2nd edition, Chapman & Hall, London, 1989
- G. MOLENBERGHS, G. VERBEKE, *Models for Discrete Longitudinal Data*, Springer Verlag, New York, 2005
- H.D. PATTERSON, R. THOMPSON, Recovery of interblock information when block sizes are unequal, *Biometrika* 58 545-554, 1971
- J.C. PINHEIRO, D.M. BATES, *Mixed effects models in S and S-plus*, Springer, Berlin, 2000
- C.R. RAO, J. KLEFFE, *Estimation of variance components and applications*, North Holland Series in Statistics and probability, Elsevier, 1988
- C. ROBERT, *Méthodes de Monte Carlo par chaînes de Markov*, Economica, Paris, 1996
- S.R. SEARLE, G. CASELLA, C.E. Mc CULLOCH, *Variance components*, J Wiley & Sons, New York, 1992
- G. VERBEKE, G. MOLENBERGHS, *Linear mixed models in practice*, Springer Verlag, New York, 1997

Langue d'enseignement

Français

UE - Méthodologie statistique 2

Compléments de modèles de durée

Survival Analysis Applied to Biostatistics

Cours : 18h • Atelier : 6h

Enseignant Yann DE RYCKE (INSERM Hôpital Bichat)

Correspondant Lionel TRUQUET

Enseignement destiné aux élèves de la filière « Statistique pour les sciences de la vie »

Objectif pédagogique

A l'issue de cet enseignement, les élèves devront maîtriser les connaissances de base pour analyser des données censurées. Ces problèmes concernent tous ceux qui sont impliqués dans les sciences de la vie où ce type de données est très fréquemment rencontré. A l'issue de ce module, l'étudiant doit être capable de mettre en œuvre les méthodes enseignées pour résoudre les problèmes classiques de comparaison de plusieurs échantillons, de modélisations (paramétriques ou non) des distributions de survie.

Contenu de la matière

Les quatre premières séances seront consacrées à l'enseignement théorique des différentes méthodes. Une dernière séance consistera en un apprentissage des différentes procédures disponibles dans les logiciels (SAS en particulier). Le cours sera illustré de nombreux exemples et des exercices seront, à chaque étape, proposés aux étudiants pour vérifier qu'ils ont bien compris et acquis l'essentiel des méthodes. Ces exemples et exercices seront principalement issus de travaux effectués en recherche clinique et en épidémiologie. Cependant, le caractère général de ces méthodes, utiles dans bien d'autres domaines d'application, sera clairement établi.

- 1- Généralités et particularités des données de survie.
- 2-Définition des différentes fonctions de survie.
- 3-Recueil des données de survie et préparation de la base de données à analyser.
- 4-Estimations non paramétriques des fonctions de survie.
- 5-Estimations et comparaisons paramétriques des distributions de survie - Cas du modèle exponentiel. Modèle exponentiel généralisé. Etude graphique de l'adéquation de la modélisation.
- 6-Comparaisons non paramétriques de plusieurs distributions de survie : logrank pondérés, liens avec les tests de rangs.
- 7-Calcul du nombre de sujets nécessaires : cas du modèle exponentiel et du logrank.
- 8-Modélisation semi-paramétrique des fonctions de survie par le modèle de Cox : définition de la vraisemblance conditionnelle, problème de codages, modèles avec interactions, étude de l'adéquation de l'hypothèse des taux proportionnels.
- 9-Introduction à différents problèmes : variables dépendantes du temps, risques concurrents, recherche d'interactions qualitatives, analyses intermédiaires ...
- 10-Etude d'un cas en vraie grandeur : Recherche de facteurs pronostiques chez des enfants atteints de tumeurs cérébrales.
- 11-Mise en application (Y. De Rycke).

Pré-requis

Aucun

Contrôle des connaissances

Examen écrit (2h et documents autorisés).

Références bibliographiques

- C. HILL, C. COM-NOUGUE, A. KRAMAR, T. MOREAU, J. O'QUIGLEY, R. SENOUSI, C. CHASTANG, *Analyse statistique des données de survie (2^e éd.)*, Collection statistique en biologie et médecine, Flammarion, 1996.

Langue d'enseignement

Français

UE - Méthodologie statistique 3

Mesures de qualité de vie

Measuring Quality of Life

Cours : 12h • Atelier : 6h

Enseignants : Jean Benoît HARDOUIN (Université de Nantes)

Correspondant : Lionel TRUQUET

Enseignement destiné aux élèves de la filière « Statistique pour les sciences de la vie »

Objectif pédagogique

Les mesures de qualité de vie associée à la santé (Health-Related Quality of Life Measures) ont connu au cours des 20 dernières années un développement important, notamment dans le domaine de l'évaluation des stratégies thérapeutiques du cancer, du SIDA ou encore de maladies chroniques. Les mesures de qualité de vie font partie des « mesures subjectives en santé », par opposition aux mesures d'efficacité cliniques objectives, traditionnellement utilisées dans les évaluations. Elles font appel à des méthodes et concepts développés en psychométrie, mais aussi en économie de la santé (approche par les préférences individuelles).

Contenu de la matière

1. Introduction
 - 1.1. Contexte et approches : Les mesures de qualité de vie en santé
 - 1.2. Définition d'une échelle de qualité de vie – Typologie des échelles
2. L'approche psychométrique classique
 - 2.1. Construction d'un questionnaire (génération et réduction d'items, choix des dispositifs de réponse, méthodes de scoring des échelles multi-items)
 - 2.2. Les propriétés de fiabilité, validité, sensibilité au changement
 - 2.2.1. Fiabilité : mesures de reproductibilité et de cohérence interne (alpha de Cronbach)
 - 2.2.2. Validité structurelle et clinique des échelles
 - 2.2.2.1. Validation de la structure d'une échelle par analyses factorielles exploratoires (EFA) et confirmatoires (CFA – Modèles d'équations structurelles)
 - 2.2.2.2. Validation clinique d'une échelle : discrimination de groupes cliniques a priori
 - 2.2.3. Sensibilité au changement : indicateurs internes et externes (effect-sizes, courbes ROC ...)
 - 2.3. Application : atelier de validation psychométrique d'une échelle de qualité de vie en cancérologie
3. L'approche économique des échelles de qualité de vie
 - 3.1. Rappel des méthodes d'évaluation économique en santé
 - 3.2. Les index de santé pondérés par les préférences des individus
4. Les nouvelles approches psychométriques : la Théorie de Réponse aux Items (Item Response Theory - IRT)
 - 4.1. L'IRT non paramétrique : Modèle de Guttman et modèles de Mokken
 - 4.2. Les modèles de la famille de Rasch
 - 4.2.1. Le modèle de Rasch
 - 4.2.2. Les modèles polytomiques de la famille de Rasch : Partial Credit Model et Rating Scale Model
 - 4.3. Les autres modèles paramétriques de la Théorie de Réponse aux Items
 - 4.4. Extensions : analyse longitudinale, modèles multidimensionnels, le fonctionnement différentiel des items, l'introduction de covariables
 - 4.5. Applications
 - 4.5.1. Exemple de validation d'échelle par un modèle IRT : le Knee injury and Osteoarthritis Outcome Score (KOOS)

4.5.2. Exemple d'analyse de données par un modèle IRT : analyse spatio-temporelle de la qualité de vie en France

4.5.3. Atelier machine sous SAS

4.5.4.

Pré-requis

Contrôle des connaissances

Projet

Références bibliographiques

- NUNNALLY, BERNSTEIN, *Psychometric Theory*, 3rd edition, 1994 Mc GRAW HILL
- P.M. FAYERS, D. MACHIN, *Quality of Life : assessment, analysis and interpretation* (2nd ed.), Wiley & Sons, 2007
- B. FALISSARD, *Mesurer la subjectivité en santé* (2^e éd.) Collection évaluation et statistique, Masson, 2008
- K. Sijtsma, I. W. Molenaar. *Introduction to Nonparametric Item Response Theory*, Collection Measurement Methods for the Social Science. Sage Publications, Inc, 2002
- H. Fischer, I. W. Molenaar. *Rasch Models: Foundations, Recent Developments, and Applications*. Springer, 1995.
- Mesbah M, Cole BF, Ting Lee ML (editors) *Statistical methods for Quality of Life studies : Designs, measurements and analysis*. Luwer, 2002
- Streiner DL, Norman GR. *Health Measurement scales* (4th ed.). Oxford University Press, 2008

Langue d'enseignement

Français

UE - Méthodologie statistique 3

Modélisation compartimentale

Compartmental Modeling

Cours : 6h • Atelier : 6h

Enseignant Audrey LAVENU (Université de Rennes 1)
Correspondant Lionel TRUQUET

Enseignement destiné aux élèves de la filière « Statistique pour les sciences de la vie »

Objectif pédagogique

A l'issue de cet enseignement, les élèves devront être capables de simuler des épidémies par des modèles Susceptible-Infectious-Removed (SIR), de comprendre l'interprétation des paramètres et de construire des modèles dérivés du modèle standard.

Contenu de la matière

Introduction : Contexte de la modélisation compartimentale. Définition d'une épidémie.

1. Modèle SIR déterministe
 - 1.1. Construction du modèle (système d'équations différentielles)
 - 1.2. Calcul du taux de reproduction de base (R_0)
 - 1.3. Simulation d'épidémies
 - 1.4. Exemples de données de surveillance épidémiologique
 - le réseau Sentinelles
 - application sur logiciel
 - 1.5. Exemples de modèles dérivés du modèle SIR.
2. Modèle SIR stochastique

Pré-requis

Aucun

Contrôle des connaissances

Rapport basé sur exercices commencés en séance avec commentaires étoffés du cours

Références bibliographiques

- Anderson, R. A. and May R. M. 1982. Directly transmitted infectious diseases: Control by vaccination, *Science* 215, 1053-1060.
- Anderson R. A. and May R. M., 1992: *Infectious Diseases of Humans: Dynamics and Control* (2nd ed.). Oxford University Press, Oxford.
- Bailey N. T. J., 1975: *The mathematical theory of infectious diseases and its application*. Griffin, London, 2nd edition (épuisé).
- Bartlett M. S. 1960. *Stochastic Population Models in Ecology and Epidemiology*, Methuen, London.

Langue d'enseignement

Français

UE - Méthodologie statistique 3

Traitement des données manquantes dans les essais cliniques

Handling of missing data in clinical trials

Cours : 6h • Atelier : 6h

Enseignant : Mélanie PRAGUE (INRIA)

Correspondant : Lionel TRUQUET

Enseignement destiné aux élèves de la filière « Statistique pour les sciences de la vie »

Contenu de la matière

PART I

Introduction to missing data – issues with missing data

Assumptions about missing data and missing data mechanisms – A formal taxonomy: MCAR, MAR and MNAR

Exercise 1

Regulatory considerations – The FDA-commissioned NRC/NAS report – Notion of estimand

Single imputation methods: LOCF, BOCF

Commonly used analytic methods under MAR:

- Complete-Case-Analysis (CCA)
- A modification of CCA under MAR: Inverse Probability Weighting (IPW)

PART II

Commonly used analytic methods under MAR (continued):

- Single imputation
- Multiple imputation
 - The multivariate normal regression framework
 - Monotone missing data patterns
 - Non-monotone missing data patterns: MCMC algorithms
 - Case of binary outcomes

PART III

Parametric analyses under MAR:

- Likelihood-based methods (Mixed Model with Repeated Measures = MMRM)
- Case study – Computer practical using SAS (version 9.4 or higher) – Questions 1-5
- Marginal versus conditional models
- Exercise 2

Introducing MNAR assumptions via the conditional model:

- Shift parameter
- Power considerations

Pattern-Mixture-Models (PMM)

PART IV

Principles and methods of sensitivity analyses under MNAR via PMM:

- Control-based imputation: Copy Reference (CR) and Jump To Reference (J2R) inference
- Case study – Computer practical using SAS – Question 6

- Delta adjustment and tipping point analysis
- Case study – Computer practical using SAS – Question 7
- Power considerations for MMRM and delta-adjusted PMM analyses
- Case study – Computer practical using SAS – Question 8

Key messages to take home

Contrôle des connaissances

A déterminer

Langue d'enseignement

Français mais supports écrits en anglais

UE - Méthodologie statistique 3

Méta-Analyse

Meta Analysis

Cours : 12h • Atelier : 6h

Enseignant : Vanessa TAIEB (Creativ Ceutical)

Correspondant : Samuel DANTHINE

Enseignement destiné aux élèves des filières « Ingénierie Statistique des Territoires et de la Santé » et « Statistique pour les sciences de la vie »

Objectif pédagogique

"La méta-analyse est une démarche, plus qu'une simple technique, qui a pour but de combiner les résultats de plusieurs essais thérapeutiques, pour en faire une synthèse reproductible et quantifiée. Cette synthèse produit un gain de puissance statistique dans la recherche de l'effet d'un traitement, une précision optimale dans l'estimation de la taille de l'effet et permet en cas de résultats apparemment discordants d'obtenir une vue globale de la situation".

Trois types de méta-analyses sont distingués, en fonction des données utilisées:

1. La méta-analyse des données résumées de la littérature, donc uniquement des essais publiés (ce qui expose au biais de publication)
2. La méta-analyse exhaustive sur données résumées se basant sur les études publiées et sur les travaux non publiés
3. La méta-analyse sur données individuelles se basant sur les données de tous les patients inclus dans les essais pris en considération dans la méta-analyse.

Dans la démarche de la méta-analyse, la variabilité (l'hétérogénéité) est considérée comme un paramètre de nuisance; elle contredit l'hypothèse de l'existence d'un effet traitement commun à tous les essais. La méta-analyse est très utilisée, notamment dans les analyses médico-économiques qui utilisent dans leur modélisation des indicateurs de résultats issus de publications diverses.

Contenu de la matière

1. Introduction
 - 1.1. What is it? / Why do we do it?
 - 1.2. The use of meta-analysis in clinical trial / Health economic evaluation
 - 1.3. Meta-analysis vs. randomised clinical trials
2. Protocol development
 - 2.1. Objectives
 - 2.2. Outcome measure and baseline information
 - 2.3. Data sources / Study selection
 - 2.4. Data extraction
 - 2.5. Analyses / Sensitivity analyses
 - 2.6. Presentation of results
3. Estimating treatment difference
 - 3.1. Binary data
 - 3.1.1. Log-odds ratio
 - 3.1.2. Log-relative risk
 - 3.2. Normally distributed data
 - 3.2.1. Absolute mean difference
 - 3.2.2. Standardised mean difference
 - 3.3. Ordinal data
 - 3.3.1. Log-odds ratio (proportional odds model)
 - 3.4. Survival data
 - 3.4.1. Log hazard ratio
4. Combining estimates of treatment difference
 - 4.1. Fixed-effects parametric approach (FE)

- 4.1.1. Definition/assumption
- 4.1.2. Model
- 4.1.3. Estimation of the treatment difference and hypothesis test
- 4.1.4. Testing for heterogeneity
- 4.2. Random-effect parametric approach (RE)
 - 4.2.1. Definition/assumption
 - 4.2.2. Model
 - 4.2.3. Estimation of the treatment difference and hypothesis test
 - 4.2.4. Testing for between studies heterogeneity
5. Dealing with heterogeneity
 - 5.1. Limited power of heterogeneity tests
 - 5.2. Choice between FE and RE models
 - 5.3. Can we always present an overall estimate of treatment difference?
 - 5.4. Choice of appropriate measure of treatment difference
 - 5.5. Meta regression
6. Presentation of results
7. Selection / publication bias
8. Direct comparison vs. Indirect comparison
 - 8.1. Eg. Drug A vs placebo & Drug B vs. placebo => Drug A vs. Drug B
9. An introduction to Bayesian approach
10. Conclusion
 - 10.1. The use of meta-analysis
 - 10.2. Contrast between useful and useless meta-analysis

Pré-requis

Basic Winbugs knowledge

Contrôle des connaissances

A déterminer

Références bibliographiques

Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions. Chichester (UK): John Wiley & Sons, 2008.

Dias, S., Welton, N.J., Sutton, A.J. & Ades, A.E. NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network Meta-Analysis of Randomised Controlled Trials. 2011; last updated September 2016; available from <http://www.nicedsu.org.uk>

Langue d'enseignement

Anglais

UE - Essais cliniques

Essais cliniques : méthodologie et analyses statistiques

Clinical Trials

Cours : 18h

Enseignants : Yann DE RYCKE (INSERM Hôpital Bichat) Responsable du cours
David HAJAGE (INSERM APHP)

Correspondant : Lionel TRUQUET

Enseignement destiné aux élèves des filières « Statistique pour les sciences de la vie » et « Ingénierie des territoires et de la santé »

Objectif pédagogique

La nature et la structure des données recueillies dans le cadre d'essais cliniques (qui incluent les études sur les médicaments, les interventions médicales novatrices et les nouveaux matériels) nécessitent de recourir à des méthodes statistiques adaptées.

Cet enseignement permettra aux élèves de se familiariser avec les différents types d'études, les enjeux, les acteurs et plus particulièrement les méthodes statistiques utilisées dans le domaine des essais cliniques.

Contenu de la matière

Après une présentation générale des essais cliniques, le cours comportera 2 parties. La première aura comme objectif de permettre aux élèves de se familiariser avec la méthodologie des essais cliniques et de découvrir le déroulement d'une étude du point de vue du biostatisticien. La seconde s'attachera à détailler certaines méthodes utilisées dans l'analyse des études cliniques.

Présentation générale des essais cliniques :

Les différentes étapes d'une étude, les intervenants, le rôle du biostatisticien

Aspects réglementaires et éthiques

Déroulement d'une étude pour le biostatisticien

L'analyse statistique : du plan d'analyse aux résultats

Choix d'une méthode adaptée aux objectifs et aux données

Puissance et nombre de sujets nécessaires

Rédaction d'un rapport statistique

Divers éléments à prendre en considération : Biais, Indépendance, Normalité, Bilatéral / Unilatéral, Populations ITT et PP, ...

Approfondissement de quelques méthodes d'analyse

Mesures répétées

Essais de différence, d'équivalence, de supériorité, de non infériorité

Courbes ROC

Concordance, fiabilité, reproductibilité

Contrôle des connaissances

Examen écrit

Références bibliographiques

Sera distribuée en séance

Langue d'enseignement

Français

UE - Essais cliniques

Pharmacométrie

Pharmacometrics

Cours : 6h • Atelier : 12h

Enseignants : Jérémie GUEDJ, Thu Thuy NGUYEN (INSERM, Institut Claude Bernard University Paris Diderot)

Correspondant : Lionel TRUQUET

Enseignement destiné aux élèves de la filière « Statistique pour les sciences de la vie »

Objectif pédagogique

A l'issue de cet enseignement, les élèves devront maîtriser les principales méthodes statistiques utilisées par le biostatisticien lors de l'analyse ou la conception des essais cliniques.

Contenu de la matière

Introduction à la pharmacométrie, la pharmacocinétique et la pharmacodynamie (principes, rôle dans le développement des médicaments, exemples de modèles)

Modèles non-linéaires à effets mixtes (introduction, historique, modèles statistiques, méthodes d'estimation, logiciels)

Construction et évaluation de modèles

TP en Monolix (présentation du logiciel, construction d'un modèle pharmacocinétique)

Optimisation de protocoles dans les modèles non linéaires simples ou mixtes, théorie et applications, simulation d'essais cliniques

TP en Monolix et PFIM (graphes diagnostiques, simulation, optimiation)

Séminaire

Pré-requis

Modèles Mixtes

Contrôle des connaissances

Examen

Références bibliographiques

Langue d'enseignement

Français

UE - Essais cliniques

Projet d'essais cliniques

Project in Clinical Trials

Atelier : 6h • Projet : 18h

Enseignant : Yann DE RYCKE (INSERM Hôpital Bichat) et David HAJAGE (INSERM APHP)

Correspondant : Lionel TRUQUET

Enseignement destiné aux élèves de la filière « Statistique pour les sciences de la vie »

Objectif pédagogique

Le but de ces projets est de mettre en application quelques-unes des méthodes vues pendant le cours sur les essais cliniques. Dans cet objectif, une base de données correspondant à un essai clinique réel est fournie aux élèves, ainsi que les documents nécessaires (cahier d'observation, protocole...) afin qu'ils réalisent suivant des directives précises l'analyse demandée, en utilisant le logiciel R.

Contenu de la matière

Ce projet, réalisé en général par groupe de 2 ou 3 élèves, demande un important travail de programmation, de réflexion statistique, un effort de recherche sur la pathologie étudiée ainsi qu'une bonne compréhension des recommandations internationales sur l'analyse statistique et la rédaction d'un rapport dans le cadre d'un essai clinique.

A la fin du projet, les élèves doivent remettre un rapport statistique d'une trentaine de pages donnant lieu à une soutenance réalisée devant des professionnels des essais cliniques. Le rapport ainsi que la soutenance seront réalisés en anglais, langue de travail utilisée dans l'industrie pharmaceutique. Les élèves auront à ce titre un soutien assuré par un enseignant d'anglais.

Pré-requis

Introduction aux essais cliniques, compléments d'essais cliniques. Bonne maîtrise des méthodes statistiques générales, en particulier: statistiques descriptives, régression linéaire, analyse de variance, régression logistique, analyse des données de survie, modèles mixtes.

Contrôle des connaissances

L'évaluation des projets repose à la fois sur le contenu des rapports et le comportement à la soutenance, devant un jury composé de spécialistes.

Références bibliographiques

Sera distribuée en séance

Langue d'enseignement

Français. Le rapport sera rédigé en anglais, et donnera lieu à une soutenance en anglais.

UE - Epidémiologie

Epidémiologie quantitative

Quantitative Epidemiology

Cours : 18h

Enseignant : Olivier GRIMAUD (EHESP)

Correspondant : Lionel TRUQUET

Enseignement destiné aux élèves de la filière « Statistique pour les sciences de la vie »

Objectif pédagogique

L'objet de l'épidémiologie est l'étude la distribution dans le temps et dans l'espace des états de santé des populations humaines et l'analyse leurs déterminants.

A l'issue du cours, l'étudiant sera capable de :

- Procéder à une description simple de la situation épidémiologique à partir des indicateurs épidémiologiques usuels.
- Formuler les hypothèses de liaison, discuter de la possibilité d'une relation de cause à effets entre les facteurs de risque et l'état de santé.
- Choisir les types d'études appropriés à la mise en évidence des relations entre maladies et facteurs de risque.
- Interpréter les résultats d'études épidémiologiques.

Contenu de la matière

Intitulé de la séance de cours
Introduction de l'épidémiologie – Causalité – Enquêtes de cohorte
Epidémiologie descriptive: prévalence, incidence, standardisation
Les mesures d'association
Enquêtes cas-témoins. Biais
Facteurs de confusion
Analyses multivariées

Pré-requis

Les méthodes statistiques de base sont supposées connues (fluctuations d'échantillonnage, intervalle de confiance, méthodes d'estimation et tests statistiques usuels).

Contrôle des connaissances

L'évaluation consistera en un examen écrit de 1 heure et demi, tous documents et calculatrice sont autorisés.

Références bibliographiques

- J. BOUYER, D. HEMON, S. CORDIER, F. DERRIENNIC, I. STUCKER, B. STENGEL, J. CLAVEL, *Epidémiologie. Principes et méthodes quantitatives*, Paris, Les éditions INSERM, 1993
- K.J. ROTHMAN, S. GREENLAND, *Modern epidemiology*, Little, Brown and Company, Boston, 1998
- D.G. KLEI NBAUM, L.L. KUPPER, H. MORGENSTERN, *Epidemiologic Research. Principles and quantitative methods*, New York, Van Nostrand Reinhold, 1982

Langue d'enseignement

Français

UE - Epidémiologie

Atelier d'épidémiologie

Epidemiology Workshop

Atelier : 18h

Enseignant : Mélanie PRAGUE (Université de Bordeaux)

Correspondant : Lionel TRUQUET

Enseignement destiné aux élèves de la filière « Statistique pour les sciences de la vie »

Objectif pédagogique

L'objectif de cet atelier est de se confronter à des exemples d'études épidémiologiques concrètes. Les étudiants utiliseront l'ensemble des connaissances en méthodes épidémiologiques et en biostatistiques acquises au cours de leur formation pour analyser différents articles rapportant les résultats d'études épidémiologiques. Ils devront être capables d'interpréter les résultats et de mettre en évidence les points forts et les limites du protocole, de la procédure de collecte des données et des analyses statistiques afin d'identifier les biais potentiels et d'évaluer la portée des conclusions de l'étude. Si nécessaire, les étudiants proposeront des protocoles ou plans d'analyses statistiques alternatifs. L'objectif final est d'acquérir du discernement dans l'utilisation des méthodes épidémiologiques et biostatistiques.

Contenu de la matière

1. Introduction à la lecture critique d'articles scientifiques
2. Rappels sur les biais dans les études épidémiologiques
3. Analyses de plusieurs articles d'épidémiologie analytique
4. Discussions à propos des articles présentés par les étudiants

Pré-requis

Ces ateliers s'appuient sur les notions abordées précédemment dans les cours d'épidémiologie quantitative et biostatistiques.

Contrôle des connaissances

Présentation orale de l'analyse critique d'un article.

Références bibliographiques

- Bouyer J, Hémon D, Cordier S, Derriennic F., Stücker I, Stengel B, Clavel J. *Epidémiologie : principe et méthodes quantitatives*, Les Editions InsermArnette, Paris, 1995

Langue d'enseignement

Français

UE - Epidémiologie

Projet de santé et environnement

Health and Environmental Topics Project

Projet : 24h

Enseignant : Sophie ANCELET (IRSN)

Correspondant : Lionel TRUQUET

Enseignement destiné aux élèves de la filière « Statistique pour les sciences de la vie »

Objectif pédagogique

Toute démarche d'analyse statistique s'intègre dans un processus où, dans le meilleur des cas, les attentes sont explicitées en amont et les résultats sont utilisés en aval pour la décision ou l'action. Les « projets d'épidémiologie » consistent à produire un rapport et une présentation orale basés sur l'étude statistique des données. L'objectif de ce projet est l'acquisition d'un savoir-faire dans la valorisation de données épidémiologiques en vue d'apporter des éléments de réponse à des questions en s'adressant à une ou plusieurs catégories précises de destinataires.

Contenu de la matière

Pour ce projet, les élèves travailleront par groupe de 3. Des contacts destinés au cadrage initial et au suivi du projet sont planifiés sur 10 heures d'enseignement au total pour les 4 groupes.

1. Présentation et choix des sujets :
 - 1.1. Contexte, enjeux, attentes
 - 1.2. Aspects techniques : matériel statistique, logiciels
2. Travail par groupe :
 - 2.1. Choix de techniques d'analyse statistique
 - 2.2. Analyse des données
 - 2.3. Interprétation des résultats
3. Atelier d'évaluation sur le travail accompli et de discussion sur les propositions relatives au rapport et à la communication orale.
4. Travail par groupe :
 - 4.1. Production du rapport
 - 4.2. Préparation de la communication orale

Pré-requis

Contrôle des connaissances

L'évaluation des projets repose à la fois sur le contenu des rapports et le comportement à la soutenance, devant un jury composé de spécialistes.

Références bibliographiques

En fonction des projets.

Langue d'enseignement

Français

UE - Statistique pour données Omics

Introduction à l'analyse de données « Omics »

Introduction to Omics Data Analysis

Cours : 6h • Atelier : 6h

Enseignant : Guillaume DESACHY (IRIS Servier)

Correspondant : Lionel TRUQUET

Enseignement destiné aux élèves de la filière « Statistique pour les sciences de la vie »

Objectif pédagogique

Les séminaires sur le génome permettent aux étudiants de rencontrer différents chercheurs travaillant sur des problèmes liés aux génomes.

Contenu de la matière

L'étude de la génomique a de nombreuses applications dans le domaine des sciences de la vie. Les problèmes liés au génome font appel à de nombreuses méthodes statistiques. Différentes approches pour l'analyse des séquences génomiques seront présentées.

Le cours consiste en une vaste initiation à la Génomique et Post-Génomique (Transcriptomique, Protéomique) et de façon plus générale à la Bioinformatique. Les objectifs de ce cours sont l'acquisition des connaissances de bases (du Génome à l'Organisme en passant par l'ADN, l'ARN, la protéine, la cellule et les bases de données associées) permettant la compréhension des principales problématiques liées à l'application de la Statistique à la Biologie. Les principales méthodes statistiques évoquées sont les chaînes-de-markov, chaînes-de-markov cachées, test d'hypothèse, test-multiple, la classification et l'alignement de séquences biologiques. Le cours est accompagné de travaux pratiques réalisés sous R.

Pré-requis

Pas de pré-requis

Contrôle des connaissances

Mini-projets en anglais, compte-rendus de TP.

Langue d'enseignement

Français

UE - Statistique pour données Omics

Analyse des données « Omics »

Omics Data Analysis

Cours : 6h • Atelier : 12h

Enseignants : Isabelle BRITO (Institut Curie) et Pierre NEUVIAL (IMT)
Correspondant : Lionel TRUQUET

Enseignement destiné aux élèves de la filière « Statistique pour les sciences de la vie »

Objectif pédagogique

Ces séminaires permettent aux étudiants de rencontrer différents chercheurs travaillant sur l'analyse d'une grande variété de données « omics », y compris les données génomiques, transcriptomiques, métabolomiques, protéomiques, épigénétiques, et métagénomiques.

Contenu de la matière

Les technologies telles que le séquençage à haut débit, les puces à ADN, et la spectrométrie de masse ont modifié l'échelle de données biologiques disponibles et permettent de générer des quantités de données importantes à plusieurs niveaux biologiques. L'analyse de ces données fait appel à de nombreuses méthodes statistiques qui seront présentées.

Plusieurs problèmes statistiques se posent dans l'analyse des données transcriptomiques : normalisation des données, modélisation des variances pour la construction des statistiques de tests afin de détecter les gènes différentiellement exprimés, corrections pour les tests multiples, classification et modèles de mélange afin d'obtenir des groupes de gènes à fonction biologique similaire. Des approches de la biologie des systèmes seront également abordées, tels que l'inférence de réseaux de gènes à partir de données transcriptomiques. Les questions posées par l'analyse de données épigénétiques et métagénomiques seront brièvement abordées. Le cours est accompagné de travaux pratiques (sur R/Bioconductor) centrés sur une analyse différentielle de données issues des puces à ADN et du séquençage à haut débit (RNA-seq). Des problématiques statistiques similaires seront abordées pour l'analyse de données métabolomiques.

Pré-requis

Introduction à l'analyse de données « Omics », théorie des tests, modèle linéaire (ANOVA et modèle mixte), classification.

Contrôle des connaissances

Un examen écrit qui comprend les notions abordées en cours et des extraits de scripts R.

Références bibliographiques

- Mary-Huard T., Picard F., Robin S. Introduction to Statistical Methods for Microarray Data Analysis, in *Mathematical and Computational Methods in Biology*, Hermann : Paris, 2007.
- Gentleman R.C., Carey V.J., Dudoit S., Irizarry R., Huber W., *Bioinformatics and Computational Biology Solutions using R and Bioconductor*, New York: Springer, 397-420, 2005.

Langue d'enseignement

Français

UE - Statistique pour données Omics

Projet de génétique

Genetics Project

Projet : 18h

Enseignant : Hadrien LORENZO (Université de Bordeaux)

Correspondant : Lionel TRUQUET

Enseignement destiné aux élèves de la filière « Statistique pour les sciences de la vie »

Objectif pédagogique

Ce projet fait suite aux séminaires sur l'analyse de données « Omics ».

Contenu de la matière

Etude d'un jeu de données.

Pré-requis

Contrôle des connaissances

L'évaluation des projets repose à la fois sur le contenu des rapports et le comportement à la soutenance, devant un jury composé de spécialistes.

Références bibliographiques

Langue d'enseignement

Français. Le rapport sera rédigé en anglais.

UE - Séminaires professionnels

Analyse sensorielle

Sensory Analysis

Atelier : 6h

Enseignant : Pascal SCHLICH (INRA Dijon)

Correspondant : Lionel TRUQUET

Enseignement destiné aux élèves de la filière «Statistique pour les sciences de la vie»

Objectif pédagogique

Présenter aux étudiants la nature et les enjeux de l'analyse sensorielle, quelques-unes de ses méthodes et surtout l'utilisation d'outils statistiques pour l'exploitation de ses données.

Contenu de la matière

Le cours est présenté sous la forme de séminaires animés par un chercheur professionnel de l'analyse sensorielle.

Pascal Schlich

1. Objectifs, principes, moyens et acteurs de l'analyse sensorielle
2. L'analyse des données de profil sensoriel par ANOVA, ACP, MANOVA et CVA
3. Le contrôle des performances des dégustateurs
4. Les méthodes multidimensionnelles non descriptives : du tri libre au Positionnement Sensoriel Polarisé (PSP)
5. La méthode de la Dominance Temporelle des Sensations (DTS). Application à une dégustation de chocolats et analyse statistique des données acquises en ligne
6. La cartographie des préférences. Cas réels

Pré-requis

Contrôle des connaissances

Pas d'évaluation.

Références bibliographiques

Sera distribuée en séance

Langue d'enseignement

Français

UE - Séminaires professionnels

Introduction au développement clinique

Introduction to clinical development

Atelier : 6h

Enseignants : A déterminer (I.R.I.S. – Institut de Recherches Internationales des laboratoires Servier)
Correspondant : Lionel TRUQUET

Enseignement destiné aux élèves de la filière «Statistique pour les sciences de la vie»

Objectif pédagogique

Le développement d'un nouveau médicament, depuis la découverte d'une molécule, jusqu'à la mise sur le marché et les études post-AMM, est un processus complexe, long et coûteux, dans lequel le biostatisticien tient un rôle important.

Le but de cet enseignement est de familiariser les élèves au processus de développement d'un nouveau médicament et d'introduire les principales notions qui seront abordées lors de l'UE Essais cliniques : méthodologie et analyses statistiques.

Contenu de la matière

- Développement d'un médicament
- Introduction aux essais cliniques
- Rôle du biostatisticien dans les essais cliniques
- Présentation d'un essai clinique

Pré-requis

Aucun.

Contrôle des connaissances

Pas d'évaluation.

Références bibliographiques

- Essais cliniques théorie, pratique et critique, 4^{ème} édition, G. BOUVENOT, M. VRAY
- Guidelines ICH

Langue d'enseignement

Français

UE - Séminaires professionnels

Statistical methods for Phase I dose-finding studies in oncology

Atelier : 6h

Enseignants : Marie-Karelle RIVIERE (SANOFI)

Correspondant : Lionel TRUQUET

Enseignement destiné aux élèves de la filière «Statistique pour les sciences de la vie»

Objectif pédagogique

Les essais cliniques de phase I ont pour objectif de déterminer chez l'Homme la dose maximale tolérée (DMT) d'un nouveau traitement ou d'une nouvelle association, en vue de l'évaluation secondaire de son efficacité dans les essais de phase II, notamment par la détermination de la dose minimale efficace (DME). En général ces essais sont conduits chez des volontaires sains mais en cancérologie, la particularité de ces essais de recherche de dose est d'être conduits chez le sujet malade, du fait de la toxicité potentielle sévère des traitements liée à leur mode d'action. Ces essais comprennent plusieurs doses fixes et sont conduits sur un petit nombre de sujets. Cependant, il a été démontré que plus 50% des traitements n'arrivent pas devant les agences de régulation car l'estimation de la dose et de l'efficacité lors des essais cliniques de phase précoce par manque d'évaluation suffisante de la relation dose-réponse. Ce qui prouve qu'il y a un manque de connaissances et de bonnes pratiques statistiques concernant cette partie de la recherche clinique.

Dans ce cours nous proposons, une journée introductive aux méthodes statistiques pour les essais de recherche de dose.

Contenu de la matière

- Introduction to dose-escalation
- Algorithm based dose-finding methods
- Model-based dose finding methods - Continual Reassessment Method (CRM)
- Using R to implement the methodology for a cancer study
- Reducing the duration of phase I oncology trials using incomplete observations
- TITE-CRM with R
- Conclusions and futures trend in dose-finding

Pré-requis

Une connaissance des essais cliniques ainsi qu'une connaissance du logiciel R est fortement souhaitée.

Contrôle des connaissances

Pas d'évaluation.

Références bibliographiques

- Ivanova, A. Escalation, group and A + B designs for dose-finding trials. *Stat Med* 25, 21 (Nov 2006), 3668–3678.
- Chevret, S. *Statistical Methods for dose-Finding Experiments*. Statistics in Practice. John Wiley and Sons Ltd., Chichester, 2006.
- O'Quigley, J., Pepe, M., and Fisher, L. Continual reassessment method: a practical design for phase 1 clinical trials in cancer. *Biometrics* 46 (1990), 33–48.
- Garrett-Mayer, E. The continual reassessment method for dose-finding studies: a tutorial. *Clin Trials* 3, 1 (2006), 57–71.
- Cheung, Y. K., and Chappell, R. Sequential designs for phase I clinical trials with late-onset toxicities. *Biometrics* 56, 4 (Dec 2000), 1177–1182.

Langue d'enseignement

Les diapositives seront en anglais, mais le cours sera disposé en français.

UE - Séminaires professionnels

Multiplicité dans les essais cliniques

Multiplicity issues in Clinical Trials

Atelier : 6h

Enseignants : Gautier PAUX (I.R.I.S. – Institut de Recherches Internationales des laboratoires Servier)
Correspondant : Lionel TRUQUET

Enseignement destiné aux élèves de la filière «Statistique pour les sciences de la vie»

Objectif pédagogique

Il est devenu de plus en plus courant dans un essai clinique d'évaluer l'efficacité d'un traitement sur plusieurs critères, comparer deux ou plusieurs doses à un contrôle, ou caractériser l'effet du traitement dans plusieurs sous-groupes de patients. Cependant, si plusieurs tests statistiques sont réalisés et que le succès de l'étude est déclaré si au moins un des tests est statistiquement significatif, la probabilité de revendiquer que le traitement est efficace alors qu'en réalité il ne l'est pas, i.e. l'erreur de Type I, s'accroît. Dans un contexte réglementaire confirmatoire, un contrôle fort de l'erreur de Type I est indispensable et des procédures de tests multiples doivent être utilisées.

Le but de cet enseignement est de familiariser les élèves aux problématiques associées à la multiplicité des tests dans les essais cliniques et d'introduire les méthodes statistiques traditionnelles permettant de contrôler l'erreur de Type I.

Contenu de la matière

- Introduction and context
- Multiple Testing Procedures (MTPs)
 - o Classification of MTPs
 - o Single-family multiplicity problems
 - o Multi-family multiplicity problems
- Power calculation in clinical trials with multiplicity issues
- Software implementation
-

Pré-requis

Une connaissance de la théorie des tests statistiques et des essais cliniques est fortement souhaitée.

Contrôle des connaissances

Pas d'évaluation.

Références bibliographiques

Benda, N., Branson, M., Maurer, W., Friede, T. (2010). Aspects of modernizing drug development using clinical scenario planning and evaluation. *Drug Information Journal*. 44, 299–315.

Bretz, F., Maurer, W., Brannath, W., Posch, M. (2009). A graphical approach to sequentially rejective multiple test procedures. *Statistics in Medicine*. 28, 586–604.

Dmitrienko, A., Bretz, F., Westfall, P.H., Troendle, J., Wiens, B.L., Tamhane, A.C., Hsu, J.C. (2009). Multiple testing methodology. *Multiple Testing Problems in Pharmaceutical Statistics*. Dmitrienko, A., Tamhane, A.C., Bretz, F. (editors). Chapman and Hall/CRC Press, New York.

Dmitrienko, A., Millen, B., Brechenmacher, T., Paux, G. (2011a). Development of gatekeeping strategies in confirmatory clinical trials. *Biometrical Journal*. 53, 875–893.

Dmitrienko, A., Paux, G., Brechenmacher, T. Power calculations in clinical trials with complex clinical objectives. *Journal of Japanese Society of Computational Statistics*. To appear.

Dmitrienko, A., Tamhane, A.C. (2009). Gatekeeping procedures in clinical trials. *Multiple Testing Problems in Pharmaceutical Statistics*. Dmitrienko, A., Tamhane, A.C., Bretz, F. (editors). Chapman and Hall/CRC Press, New York.

Dmitrienko, A., Tamhane, A.C. (2011). Mixtures of multiple testing procedures for gatekeeping applications in clinical trials. *Statistics in Medicine*. 30, 1473–1488.

Dmitrienko, A., D'Agostino, R.B. (2013). Tutorial in biostatistics: Traditional multiplicity adjustment methods in clinical trials. *Statistics in Medicine*. 32, 5172–5218.

Dmitrienko, A., D'Agostino, R.B., Huque, M.F. (2013). Key multiplicity issues in clinical drug development. *Statistics in Medicine*. 32, 1079–1111.

Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple significance testing. *Biometrika*. 75, 800–802.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*. 6, 65–70.

Hommel, G. (1988) A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika*. 75, 383–386.

Marcus, R. Peritz, E., Gabriel, K.R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*. 63, 655-660.

Langue d'enseignement

Français

UE - Séminaires professionnels

L'analyse de survie en épidémiologie et en pharmaco-épidémiologie

Using survival analysis in epidemiology and pharmaco-epidemiology

Atelier : 3h

Enseignant : Anne THIEBAUT (Inserm)

Correspondant : Lionel TRUQUET

Enseignement destiné aux élèves de la filière «Statistique pour les sciences de la vie»

Objectif pédagogique

Ce séminaire a pour but de présenter des questions de recherche que soulève l'analyse par des modèles de durée de données épidémiologiques en général, pharmaco-épidémiologiques en particulier. L'utilisation de l'analyse de survie en épidémiologie peut nécessiter quelques aménagements par rapport au cadre initial des essais randomisés. Il en est de même dans le domaine de la pharmaco-épidémiologie qui vise à évaluer l'usage des produits de santé et l'impact de l'exposition médicamenteuse sur l'état de santé.

Contenu de la matière

1. Rappels sur l'analyse de survie « classique »
2. Utilisation en épidémiologie : choix de l'origine des délais avant événement
3. Utilisation en pharmaco-épidémiologie : variables dépendantes du temps, fonctions de risque complexes
- 4.

Pré-requis

Modèles de durée

Notions d'épidémiologie

Contrôle des connaissances

Pas d'évaluation.

Références bibliographiques

- Korn EL, Graubard BI, Midthune D. Time-to-event analysis of longitudinal follow-up of a survey: choice of the time-scale. *American Journal of Epidemiology* 1997 Jan 1; 145(1): 72-80.
- Pencina MJ, Larson MG, D'Agostino RB. Choice of time scale and its effect on significance of predictors in longitudinal studies. *Statistics in Medicine* 2007 Mar 15; 26(6): 1343-53.
- Cologne J, Hsu WL, Abbott RD, Ohishi W, Grant EJ, Fujiwara S, Cullings HM. Proportional hazards regression in epidemiologic follow-up studies: an intuitive consideration of primary time scale. *Epidemiology* 2012 Jul; 23(4): 565-73. doi: 10.1097/EDE.0b013e318253e418.
- van Staa TP, Abenhaim L, Leufkens H (1994). A study of the effects of exposure misclassification due to the time-window design in pharmacoepidemiologic studies. *Journal of Clinical Epidemiology* 1994 Feb; 47(2): 183-9.
- Schneeweiss S, Stürmer T, Maclure M. Case-crossover and case-time-control designs as alternatives in pharmacoepidemiologic research. *Pharmacoepidemiology and Drug Safety* 1997 Oct; 6(Suppl.3): S51-S59.
- Guess HA. Exposure-time-varying hazard function ratios in case-control studies of drug effects. *Pharmacoepidemiology and Drug Safety* 2006 Feb; 15(2): 81-92.

Langue d'enseignement

Français

UE - Séminaires professionnels

Analyse de combinaisons de drogues

Analysis of drug combinations

Atelier : 6h

Enseignant : Julie FOUCQUIER (Pharnext)

Correspondant : Lionel TRUQUET

Enseignement destiné aux élèves de la filière «Statistique pour les sciences de la vie»

Descriptif en cours d'élaboration

Objectif pédagogique

Contenu de la matière

Pré-requis

Contrôle des connaissances

Pas d'évaluation.

Références bibliographiques

Langue d'enseignement

Français

UE - Séminaires professionnels

Evaluation médico-économique

Economic Evaluation in Health Care

Atelier : 12h

Enseignant : Hélène CAWSTON (Amaris)

Correspondant : Samuel DANTHINE

Enseignement destiné aux élèves des filières « Statistique pour les sciences de la vie » et « Ingénierie des territoires et de la santé »

Objectif pédagogique

L'objectif de ce cours est de présenter les méthodes de l'évaluation économique dans le domaine de la santé. Ce cours abordera d'abord les principes fondamentaux de l'évaluation sur le terrain, puis les différentes méthodes seront présentées et discutées, notamment du point de vue de résultats utilisés : l'évaluation coût-efficacité, l'évaluation coût-utilité, l'évaluation coût-bénéfice. Les aspects théoriques et techniques de l'évaluation médico-économique seront abordés, en particulier les techniques de modélisation statistique avancées permettant d'intégrer l'incertitude dans le calcul économique.

Contenu de la matière

1. Introduction à l'évaluation économique en santé
 - 1.1. Concepts
 - 1.2. Méthodes d'évaluation
2. Modélisation économique
 - 2.1. Arbres de décision
 - 2.2. Modèles de MARKOV
3. Analyse de l'incertitude
 - 3.1. Analyse classique
 - 3.2. Analyse probabiliste
4. Analyse critique d'un article

Pré-requis

Contrôle des connaissances

Examen écrit

Références bibliographiques

- M.F. DRUMMOND, B.J. O'BRIEN, G.L. STODDART, G.W. TORRANCE, *Méthodes d'évaluation économique des programmes de santé*, Economica, 2^{ème} édition, 1998
- M.F. DRUMMOND, McGUIRE A, *Economic evaluation in health care*, 2001.
-

Langue d'enseignement

Français

UE - Séminaires professionnels

BIG DATA, data Mining et machine Learning

BIG DATA, data Mining and machine Learning

Atelier : 12h

Enseignant : Xavier VAN AUSLOOS (SOGETI)

Correspondant : Lionel TRUQUET

Enseignement destiné aux élèves de la filière «Statistique pour les sciences de la vie»

Objectif pédagogique

Introduction aux problématiques BIG DATA, data Mining et machine Learning en santé

Contenu de la matière

- Métier et les enjeux en santé :
 - projets d'innovations avec Institut Paoli Calmettes de Marseille autour de la médecine de précision/traitement
 - projet d'innovation mené par une collègue de SOGETI HT : <http://open-intelligence.fr/genomique-et-big-data-lassociation-integragen-igr-inserm-et-sogeti-hightech/>
 - projet SIRIC : <http://www.e-cancer.fr/Professionnels-de-la-recherche/Recherche-translationnelle/Les-SIRIC>: mise en place d'une infra type Big Data pour la recherche translationnelle en cancérologie
 - freins actuels : sécurité, peu de connaissances techniques, confidentialité des données
 - projet HANDILIGHT <http://www.lemondeinformatique.fr/defih/zone-presse.html>
- Intro au Big Data :
 - pourquoi le big data vs info traditionnelle
 - Ecosystèmes
 - Hadoop / HDFS
 - Hive/Spark
 - noSQL
 - Data Processing with YARN/MapReduce2
- TD etTP
 - Lab 1 Hadoop Hands-On Lab
 - Hadoop Languages
 - Lab 2 Hadoop Languages
 - Hive
 - Lab 3 Hive
 - Spark Fundamentals
 - Lab 4 Spark Fundamentals
 - Lab 4 Spark Fundamentals (continued)
 - Using SQL with Hadoop (IBM BigSQL or equivalent)
 - Using R with Hadoop (IBM BigR and equivalent)
 - HBase

- Lab 5 HBase
- Ambari
- Lab 6 Ambari
- IBM BigSheets, MS PowerBI,
- Lab 7 BigSheets or MS PowerBI
- Text Analytics Fundamentals
- Lab 8 Text Analytics Fundamentals lab
- Final Considerations

Pré-requis**Contrôle des connaissances**

Pas d'évaluation.

Références bibliographiques**Langue d'enseignement**

Français

UE - Séminaires professionnels

Deep Learning

Deep Learning

Atelier : 6h

Enseignant : Romaric GAUDEL (Ensay)

Correspondant : Lionel TRUQUET

Enseignement destiné aux élèves de la filière «Statistique pour les sciences de la vie»

Objectif pédagogique

Introduction aux réseaux de neurones et aux réseaux de neurones convolutifs.

Mise en pratique sous R ou Python.

Pré-requis

Une connaissance des enjeux de l'apprentissage automatique.

Contrôle des connaissances

Pas d'évaluation.

Références bibliographiques

- Deep Learning. Ian Goodfellow, Yoshua Bengio and Aaron Courville. MIT Press. Version gratuite en ligne : <http://www.deeplearningbook.org>

Langue d'enseignement

Français