



Les enseignements

- Bases de données et compléments en informatique [25-25]
- Modèles paramétriques de régression [20-20]
- Traitement des séries chronologiques [20-20]
- Recueil planifié des données: plans d'expérience et d'échantillonnage [20-20]
- Processus aléatoires [15-15]
- Introduction à la statistique bayésienne [15-15]
- Modèles et méthodes de statistique non paramétriques [15-15]
- Méthodes statistiques de Datamining [25-25]
- Analyse des données de survie [20-20]
- Epidémiologie et santé publique [20-20]
- Méthodes Statistique en assurance dommage [20-20]

Contenu du programme

Bases de données et compléments en informatique

Objectifs

- Acquérir les concepts et techniques sur lesquels reposent les systèmes de gestion de bases de données.
- Comprendre les fonctions principales d'un système de gestion de bases de données.
- Pouvoir recueillir des données, les structurer les manipuler avec le langage SQL et de présenter les résultats sous forme de rapport.
- Etre en mesure de concevoir la mise en place d'une base de données.
- Etre en mesure d'administrer une base de données: programmer des accès et des modifications à une base de données, etc.

Cahier de charges

- Bases du modèle relationnel: structures de données et algèbre.
- Langages relationnels de définition et de manipulation basés sur la logique du premier ordre.
- Langage SQL.
- Programmation d'applications de bases de données.
- Fonctions et architecture des logiciels de gestion de bases de données.
- Gestion des accès simultanés aux bases de données et techniques associées de récupération en cas de panne
- Préparation de tableaux résumés des données, production de rapports sous différents formats (texte, html...).

Modèles paramétriques de régression

Objectifs

Donner un aperçu critique des extensions des modèles linéaires et des modèles linéaires généralisé

Cahier des charges

1. Modèle linéaire paramétriques à effets aléatoires et mixtes: données longitudinales, mesures répétées, données manquantes
2. Modèles linéaires généralisés
3. Modèles linéaires généralisés mixtes et à effets aléatoires
4. Equation d'estimation généralisée,

Traitement des séries chronologiques

Objectifs

Acquérir une maîtrise opérationnelle des modèles stochastiques utilisables pour la prévision quantitative des séries chronologiques.

Cahier des charges

1. Modélisation des séries chronologiques : introduction
2. Modèles et méthodes de décomposition
3. Modèle linéaires paramétriques simples: Modèle ARMA et ARIMA, et méthodologie de Box et Jenkins
4. Modèles nonlinéaires et modèles hétéroscédastiques: modèles (G)ARCH,
5. Applications

Recueil planifié des données: plans d'expérience et d'échantillonnage (20h+20h)

Objectifs

Deux étapes sont cruciales dans toute étude statistique : le recueil des données et l'analyse de ces données. Souvent l'accent est mis sur l'ensemble des méthodes d'analyse statistique des données, mais la question des méthodes de recueil est souvent négligée, bien que essentielle.

On considère deux modes de recueil de l'information: l'expérimentation et l'échantillonnage. L'objet de ce module est d'apprendre à maîtriser les techniques de contrôle de l'hétérogénéité ou du choix des objets comparés en expérimentation, et les façons d'obtenir la représentativité en échantillonnage.

Cahier des charges

1. Notions de randomisation, de bloc (complet ou incomplet) et d'interaction pour les expériences factorielles. Confounding et de répétitions fractionnaires,
2. Plans d'expériences à plusieurs échelles et de plans optimaux.
3. Echantillonnage avec probabilités inégales: les estimateurs de HorvitzThompson et de HansenHurwitz.
4. Echantillonnage à partir d'unités complexes: échantillonnage stratifié, échantillonnage en grappes, échantillonnage en deux étapes. Plans de sondage à plusieurs degrés,
5. Amélioration d'estimateurs au moyen d'information auxiliaire (estimateur de quotient et de régression)
6. Echantillonnage de populations biologiques: problème de la base d'échantillonnage, estimation de la taille de la population.

Processus aléatoires

Objectifs

Présenter les principaux processus stochastiques avec une introduction à leur étude statistique

Cahiers des charges

- processus stochastiques et processus de Markov.
- Martingales a temps discret (sous-martingales, sur-martingales)
- Théorèmes du temps d'arrêt et de convergence

- Processus stationnaires.

Introduction à la statistique bayésienne (15h+15h)

Objectifs

- Acquérir les principes et les techniques de base de la statistique bayésienne,
- Etre capable de les utiliser et de mettre en évidence leurs avantages et leurs limites

Cahier des charges

- Le modèle bayésien: principes généraux.
- La fonction de vraisemblance et spécification a priori.
- Modèles à un paramètre: choix de la distribution a priori, calcul de la distribution a posteriori, résumer la distribution a posteriori.
- Modèles multi-paramètres: choix des distributions a priori et calcul des distributions a posteriori, paramètres de nuisance. Cas des modèles multinomial et gaussien multivarié.
- Inférence en grand échantillon et relation avec l'inférence fréquentiste.
- Méthodes de calcul en analyse bayésienne.

Modèles et méthodes de statistique non paramétriques (15h+15h)

Objectifs

1. Décrire les méthodes d'analyse statistique ne reposant pas sur un modèle paramétrique prédéfini.
2. Introduire à la théorie de l'estimation non paramétrique dans un espace fonctionnel.

Cahier des charges

1. Estimation non paramétrique d'une fonction de répartition
2. Estimation non paramétrique d'une fonction de densité : la méthode à noyau
3. Estimation non paramétrique d'une fonction de régression :
4. Estimation à noyau
5. Estimation par polynômes locaux et par projection
6. Estimation par splines

Méthodes statistiques de Datamining (25h+25h)

Objectifs

1. Acquérir des connaissances de base sur la méthodologie et les outils du data mining pour le traitement de grandes bases de données d'entreprises.
2. Il aura conscience des spécificités et de l'apport du data mining par rapport à la statistique classique
3. Etre en capacité de mettre les connaissances en pratique pour traiter des études de cas avec un logiciel dédié.

Cahier des charges

1. Echantillonnage et partitionnement de la base de données.
2. Pré-traitement et validation des données.
3. Exploration, réduction et transformation des variables.
4. Méthodes factorielles d'analyse de données
5. Outils de modélisation et de classification du data mining: Discrimination et classement;
6. Arbres de décision; Réseaux de neurones.
7. Outils de validation de modèle.

Analyse des données de survie (20h+20h)

Objectifs

1. L'étude du temps de survie de patients est d'une grande importance pour motiver l'introduction d'un nouveau traitement. Les données de survie se caractérisent par des temps de suivie avant l'évènement d'intérêt (décès, rechute, ...) ou le temps de censure (perte de vue).
2. L'étude de ce type de données nécessite des outils statistiques spécifiques qui seront abordés dans ce cours.

Cahier des charges

1. Concepts de base: mécanismes de censure et troncature, fonctions de survie paramétriques, courantes en analyse de survie
2. Grandeurs spécifiques des durées de vie et données incomplètes
3. Modèles paramétriques et méthodes d'estimation
4. Tests d'homogénéité de populations
5. Modèles de régression en durées de vie (vie accélérée, risques proportionnels, etc.)
6. Cas non paramétriques: estimateurs de NelsonAalen et KaplanMeier
7. Régression semiparamétrique: l'estimateur de la vraisemblance partielle de Cox
8. Modèles à risques compétitifs.

Epidémiologie et santé publique (20h+20h)

Objectifs

Donner à l'étudiant les outils de base pour être en capacité de:

1. contribuer à des études épidémiologiques,
2. de traiter les données des études cas-témoin, de cohortes, de survie en épidémiologie ainsi que les essais thérapeutiques;
3. assurer la surveillance de phénomènes de santé ;
4. utiliser l'épidémiologie comme aide à la décision individuelle et collective ;
5. comprendre et d'analyser de façon critique des études ayant trait à la santé, par exemple des essais cliniques en épidémiologie des populations ou en santé communautaire.

Cahier des charges

1. Introduction à l'épidémiologie, exemples cliniques, intérêt, limites,
2. Les différents types d'étude : de prévalence et de pronostic, cas-témoins, cohortes, les études d'intervention (essais thérapeutiques, dépistages, prévention primaire).
 - i. Etudes de cas témoins: questions, protocole, méthodes statistiques, limites, modèles multivariés.
 - ii. Etudes de cohortes: questions, protocole, méthodes statistiques, limites : modèles multivariés.
3. Système des essais cliniques: les phases pré-cliniques et les phases cliniques; phase III pour faire la preuve; critères d'évaluation et les indices cliniques; puissance et taille d'un essai de phase III; pronostic et courbe de survie
4. Les causalités en médecine
5. Gestion du bénéfice risque

Méthodes Statistique en assurance-dommage (20h+20h)

Objectifs

Développer les modèles stochastiques et les méthodes statistiques dans le cadre de leurs applications à des problèmes d'assurance.

Cahier des charges

1. Lois de probabilité du nombre de sinistres et du coût de sinistres: estimation des paramètres de ces lois. Tests d'adéquation de ces lois
2. Différentes méthodes de tarification et comparaison
3. Analyse du système bonus/malus et comparaison des différents systèmes proposés.
4. Techniques de règlements des sinistres.
5. Estimation de réserves, théorie des réassurances.