



Licence 3

MATHEMATIQUES

Parcours : STATISTIQUE

Livret Pédagogique

Objectif

Le parcours STATISTIQUE a pour objectif l'acquisition de compétences qui couvrent les besoins des entreprises dans le domaine du traitement statistique et informatique de l'information.

Ce parcours offre une initiation aux méthodes et techniques utilisées dans les études statistiques, dans les domaines de l'aide à la décision, du suivi de protocoles expérimentaux, des systèmes d'information et des bases de données.

Un enseignement pratique axé sur des études de cas et des projets complète l'enseignement méthodologique en lui donnant une dimension concrète et opérationnelle.

Des professionnels interviennent dans l'enseignement aux côtés d'enseignants-chercheurs de différents laboratoires de l'UBS.

Contenu

La troisième année de la licence (L3 STAT) est composée d'UE de Statistique, d'Informatique et de Culture générale.

L'accent est mis sur l'utilisation de logiciels statistiques tels que SAS, Python, R.

Un grand nombre de postes de travail de configuration récente (sous Linux et Windows), accessibles en dehors des heures de formation, est mis à la disposition des étudiants. Ce qui permet à ces derniers, de développer et d'appliquer, en toute autonomie, leurs connaissances.

Poursuite d'études et débouchés

L'obtention de la licence MATHÉMATIQUES, Parcours STATISTIQUE, permet, de droit et sans conditions, une poursuite d'études dans le master DSMS (Data Science et Modélisation Statistique) proposé à l'UBS.

D'autres possibilités sont offertes dans le cycle préparatoire de l'école d'ingénieur de l'UBS (ENSIBS) ou dans d'autres universités pour un Master Recherche.

Les débouchés de la formation sont très variés et concernent les nombreux domaines d'activité où la statistique joue un rôle important :

- Industrie (fiabilité, sûreté de fonctionnement, contrôle qualité),
- Secteur tertiaire (banque, assurance, marketing),
- Santé (bio-statistique, épidémiologie, industrie pharmaceutique).

Contact

UFR Sciences et Sciences de l'Ingénieur
Département Mathématiques, Informatique, Statistiques
Campus de Tohannic - Rue Yves Mainguy
56000 Vannes

Scolarité

Elisabeth Riou (elisabeth.riou@univ-ubs.fr)

Directeur des Etudes

Professeur Gilles Durrieu (gilles.durrieu@univ-ubs.fr)

Site web

www.univ-ubs.fr

SEMESTRE 5

Unités d'Enseignement de Parcours

- **STA 1517** : Statistique Mathématique
- **STA 1520** : Séries Temporelles et Processus Stochastiques
- **STA 1516** : Programmation et Logiciels Statistiques
- **STA 1518** : Analyse Discriminante et Régression Logistique
- **STA 1513** : Enquête et Sondages

Unités d'Enseignement de Culture Générale

- **ANG 1502** : Anglais
- **COM 1502** : Communication
- **FE1 1501** : Français pour étranger (1)

SEMESTRE 6

Unités d'Enseignement de Parcours

- **STA 1627** : Inférence dans les Modèles à Espaces d'Etats
- **STA 1625** : Python et Machine Learning
- **STA 1616** : Introduction au Data Mining
- **INF 1611** : Gestion des Bases de Données
- **STA 1604** : Etude de Cas

Unités d'Enseignement de Culture Générale

- **ANG 1602** : Anglais
- **ECN 1602** : Économie
- **FE1 1601** : Français pour étranger (1)

UNITÉS D'ENSEIGNEMENT
DU SEMESTRE 5

STA1517 : Statistique mathématique

Modalités pédagogiques :

Cours : 20 h

TD : 22 h

ECTS : 5

Objectif : Acquérir les outils nécessaires à la compréhension et à la mise en oeuvre des techniques statistiques.

Prérequis : Analyse mathématique. Intégration. Algèbre.

Programme :

1. Estimation

- Notion de modèle statistique. Identifiabilité. Famille exponentielle.
- Notion de statistique. Exhaustivité. Critère de factorisation de Neyman.
- Estimateurs : Estimateur ponctuel. Estimateur de la fonction de répartition.
- Méthodes de construction des estimateurs : la méthode de moments, Z-estimateurs; M-estimateurs. Exemples. La méthode delta.
- Estimation basée sur la vraisemblance. L'information de Kullback-Leibler. Consistance. Information de Fisher. Normalité asymptotique : le cas univarié et le cas multivarié.
- Notion sur l'estimation Bayésienne
- Optimalité des estimateurs. La borne de Cramér-Rao. Efficacité asymptotique.
- Estimation sans biais de variance minimale. Théorème de Rao-Blackwell. Théorème de Lehmann-Schéffe. Exemples d'application.
- (Eléments de bootstrap. Exemples d'applications sur ordinateur TD)

2. Intervalles de confiance et la vérification d'hypothèses

- Intervalles de confiance. La méthode de pivot (exact et asymptotique). Exemples.
- Le test optimal de Neyman-Pearson. Le test du rapport de vraisemblance. Exemples.
- Loi asymptotique du rapport de vraisemblance : cas univarié et multivarié. Exemples d'application. Le test de Wald. Le test de score. Exemples.

Bibliographie :

- Saporta, G. Probabilités, analyse des données et statistique. TECHNIP, 1990.
- K. Knight. Mathematical statistics. Chapman and Hall 2000.
- Ph. Tassi. Methodes statistiques. Economica 1992.
- A.A. Borovkov. Mathematical Statistics. Gordon and Breach Press 1998.

STA 1520 : Séries Temporelles et Processus Stochastiques

Modalité pédagogiques :

Cours : 20 h

TD : 22 h

ECTS : 5

Objectif : L'analyse des séries temporelles et des processus stochastiques occupe, dans le domaine de la statistique, une place importante. Il est peu de disciplines qui ne soient confrontées à l'étude de phénomènes évoluant dans le temps et qu'on désire décrire, expliquer, contrôler ou prévoir. L'économie, la gestion des entreprises, la météorologie, la démographie, la santé, les sciences expérimentales, constituent autant d'exemples de domaines intéressés par ce type de problèmes.

Prérequis : Modèles linéaires ; Analyse de variance ; Tests ; Intervalles de confiance ; Calcul des probabilités ; Calcul Matriciel ; Equations différentielles.

Programme :

I – Séries Temporelles

1. Présentation des séries temporelles - Logiciels SAS et S-plus et langage de programmation.
2. Tendance des séries - Tests de la tendance linéaire de Student et Fisher - Prévision et fourchettes de prévision - Tendance linéaire multiple avec variables binaires sur les observations aberrantes, Prévisions - Régression glissante, rupture de tendance.
3. Tendances exponentielle et logarithmique - Taux d'accroissement des séries – Tendances polynomiales - Tendances logistique et Gompertz ; Prévisions
4. Saisonnalité; Description; coefficients saisonniers - Tests de saisonnalité de W-R et de Schuster - Périodogramme - Ajustements périodiques.
5. Décompositions saisonnières par la méthode des moyennes mobiles ; par régression – Séries CVS
6. Contrôle continu
7. Le programme Census-X11
8. Le LES ; Fourchettes de prévision ; Constante de lissage obtenue par prévision arrière
9. Modèles autorégressifs et modèles ARX autorégressifs avec variables exogènes
10. Autocorrélogrammes ; Estimation des paramètres du modèle autorégressif
11. Applications

II – Processus Stochastiques

1. Généralités sur les processus stochastiques
2. Espaces des états d'une chaîne de Markov et Matrice de transition
3. Classes d'équivalence de l'espace des états
4. Périodicité – Réurrence – Transience
5. Mesure invariante – Mesure réversible
6. Comportement asymptotique – Lois stationnaires
7. Calcul de lois limites

Bibliographie :

- Tenenhaus, M. (2007), Statistique ; Méthodes pour décrire, expliquer et prévoir, Dunod, Paris.
- Vaté, M. (1993), Statistique Chronologique et Prévision, Economica, Paris.
- Taylor, H.M., Karlin, S. (1984), An introduction to Stochastic Modeling, Wiley, 3ème Edition.

STA 1516 : Programmation et Logiciels Statistiques

Modalités pédagogiques :

Cours : 20 h

TD : 22 h

ECTS : 5

Objectif : Ce cours doit permettre aux étudiants d'acquérir une certaine aisance dans la manipulation de deux logiciels de statistique : SAS et R.

Prérequis : Bases de programmation et d'algorithmique.

Programme :

- Logiciel SAS
 1. Introduction à SAS
 2. Entrées/sorties, organisation des données sous SAS
 3. Mise en forme des données
 4. Fonctionnalités des procédures SAS BASE
 5. Fonctionnalités des procédures SAS GRAPH
 6. Macro-programmation sous SAS
 7. Introduction aux fonctionnalités avancées : SAS IML (Langage matriciel)
- Logiciel R
 1. Introduction à R
 2. Entrées/sorties, organisation des données sous R, notion d'objet
 3. Rappels de programmation, programmation vectorielle
 4. Fonction R de base
 5. Graphiques

Bibliographie :

- Decourt O., H. Kontchou-Kouomegni SAS - Maitriser SAS Base et SAS Macro, Dunod 2007.
- Matloff N., The Art of R Programming, No Starch Press 2011.

STA 1518 : Analyse discriminante et régression logistique

Modalités pédagogiques :

Cours : 20 h

TD : 22 h

ECTS : 5

Objectif : Acquérir la maîtrise de diverses méthodes d'Analyse Discriminante.

Prérequis : Algèbre et analyse matricielle. Analyse. Probabilités. Modèle linéaire.

Programme :

- Rappel des outils de probabilités. Analyse discriminante. Classement optimal des deux catégories. Règle de Bayes. Exemples.
- Le classement optimal des plusieurs catégories.
- Fonctions discriminantes pour des observations normales.
- Méthodes non paramétriques. Estimateur à noyau de la densité. La balance entre le biais et la variance. Le cas des données multidimensionnelles. Régression non paramétrique. Applications pour la discrimination.
- Régression logistique simple. Régression logistique multiple. Régression logistique avec des classes multiples.
- Analyse factorielle discriminante. Variances intra-classe et inter-classe. Choix des axes discriminantes et variables discriminantes.
- Techniques de réduction de la dimension et applications pour le classement des deux catégories. Discrimination des plusieurs catégories. Exemples. Validation croisée.

Bibliographie :

- R. Duda, P. Hart, D. Stork. Pattern Classification. Wiley, 2001.
- A. Webb. Statistical pattern recognition. Wiley 2002.
- L. Wasserman. All of Statistics. Springer 2004.
- T. Hastie, R. Tibshirani, J. Friedman. The elements of statistical learning, Springer 2009.

STA 1513 : Enquêtes et Sondages

Modalités pédagogiques :

Cours : 20 h

TD : 22 h

ECTS : 5

Objectif : Les méthodes de sondages ont pour objectif de tirer dans une population concrète, des échantillons destinés à estimer avec la meilleure précision possible, des paramètres d'intérêt. Les méthodes aléatoires et non aléatoires sont abordées,

Prérequis : Développement en série de Taylor ; Optimisation convexe ; Analyse de variance ; Tests ; Intervalles de confiance.

Programme :

- Procédure sous SAS. Présentation des Enquêtes
- Estimateurs de Hurwitz-Thomson. Plan de Sondage Aléatoire Simple
- Estimateurs non biaisés de la moyenne et de la variance pour des échantillons avec et sans remise. Estimation de l'effectif d'une population par capture-recapture
- Stratification a priori
- Analyse de la variance des échantillons stratifiés. Optimalité des allocations, Tests de Student pour deux strates
- Répartition de l'échantillon sous contrainte des coûts d'enquête
- Plan de sondage par grappes à probabilités égales
- Estimation et intervalle de confiance pour un Ratio
- Plan de sondage par grappes à probabilités inégales. Tirage systématique des grappes, Plan de sondage à 2 degrés
- Méthode des quotas. Redressement par le quotient, par la différence et par régression
- Redressement sur critères qualitatifs (post-stratification)

Bibliographie

- Ardilly, P. & Tillé, Y. (2003), Exercices Corrigés de Méthodes de Sondage, Ellipses, Paris.
- Tillé, Y. (2001), Théorie des Sondages, Dunod, Paris.

UNITÉS D'ENSEIGNEMENT DU SEMESTRE 6

STA 1627 : Inférence dans les Modèles à Espaces d'Etats

Modalités pédagogiques :

Cours : 20 h

TD : 22 h

ECTS : 5

Objectif :

Dans de nombreuses applications telles que les télécommunications, la télédétection, la géolocalisation, le traitement d'images ou de séquences vidéo, l'information que l'on souhaite obtenir n'est pas directement accessible mais est noyée dans un signal observable. La problématique qui se pose alors, au vu des observations et au regard d'hypothèses sur le système, est de savoir comment estimer au mieux le signal utile et/ou les paramètres du modèle. Le développement de méthodes permettant de reconstituer cette information « cachée » est en conséquence un enjeu important pour de très nombreuses applications. L'objectif de ce cours est de permettre aux étudiants d'acquérir les notions de lissage, filtrage et prédiction dans ces modèles et de maîtriser les techniques d'inférence Bayésienne associées.

Prérequis : Probabilités. Algèbre et analyse matricielle.

Programme :

- Rappel en probabilités. Lois multivariées. Propriétés de la loi normale multivariée.
- Modèles de Markov cachés. Propriétés principales. Notion de filtrage, lissage et prédiction. Formules forward-backward.
- Cas discret – algorithme de Baum-Welch.
- Cas linéaire et Gaussien – algorithme de Kalman
- Applications

Bibliographie :

- Särkkä, S., 2013. *Bayesian Filtering and Smoothing*, Cambridge University Press.
- Cappé, O., Moulines, E. & Ryden, T., 2005. *Inference in Hidden Markov Models*, Springer.
- Anderson, B.D.O. & Moore, J.B., 2005. *Optimal Filtering*, Dover Publications.
- Brown, R.G. & Hwang, P.Y.C., 2012. *Introduction to Random Signals and Applied Kalman Filtering with Matlab Exercises*, John Wiley & Sons.
- Grewal, M.S. & Andrews, A.P., 2008. *Kalman Filtering*, Wiley-IEEE Press.

STA 1625 : Python et Machine Learning

Modalités pédagogiques :

Cours : 20 h

TD : 22 h

ECTS : 5

Objectif : S'initier aux techniques de Machine Learning (Apprentissage Statistique) à l'aide du langage de programmation Python.

Prérequis : Bases de programmation et algorithmique.

Programme :

1. Introduction à Python
2. Introduction aux réseaux de neurones et à la discrimination linéaire
3. Perceptrons multicouches
4. Fonctions radiales de base et splines
5. Réduction de dimension
6. Apprentissage probabiliste
7. Introduction aux SVM
8. Optimisation et recherche
9. Apprentissage par algorithmes génétiques
10. Apprentissage par renforcement
11. Arbres de décision
12. Apprentissage non supervisé
13. MCMC avec Python
14. Processus gaussiens

Bibliographie :

- Marsland S., Machine Learning : An Algorithmic Perspective, 2^{ème} édition, CRC Press 2015.

STA 1616 : Introduction au Data Mining

Modalités pédagogiques :

Cours : 20 h

TD : 22 h

ECTS : 5

Objectif : Etudier les techniques permettant à un ordinateur d'identifier des structures ou des configurations dans les données.

Prérequis : Algèbre et Analyse Matricielle.

Programme :

- Introduction : qu'est-ce que le Data Mining ? Data Mining et Statistique. Les problèmes types du data Mining.
- Prétraitements : données manquantes, sélection de variables, ...
- Analyse du panier de la ménagère et algorithmes a priori
- K Plus proches voisins
- Méthodes de classification
- Régression multiple et application aux prévisions

Bibliographie :

- Tuffery S., Data Mining et Statistique décisionnelle, TECHNIP 2007.
- Franco J.-M., Le Data Warehouse, le Data Mining, Eyrolles 1998.

INF 1611 : Gestion des bases de données

Modalités pédagogiques :

Cours : 20 h

TD : 22 h

ECTS : 5

Objectif : L'objectif de ce cours est d'amener à une bonne compréhension et maîtrise des problèmes liés à l'administration et au tuning d'un SGBD. Le cours présente les composants d'un SGBD, son organisation et sa mise en oeuvre dans un environnement centralisé ou réparti. Les aspects liés à l'optimisation des performances sont ensuite étudiés.

Prérequis : Connaissance du modèle relationnel.

Programme :

- Architecture d'une base de données
Structure physique, Structure logique, Création, démarrage et arrêt d'une base
- Gestion des accès concurrents
Mécanisme transactionnel, Sérialisation, verrouillage, Mécanisme de lecture consistante
- Répartition
Bases de données réparties, Fragmentation, Réplication,
Transaction répartie
- Optimisation des performances
Optimisation des applications (Optimisation des requêtes, Différents types d'indexation, Stockage des objets, Partitionnement), Optimisation de la mémoire, Gestion des contentions

Bibliographie :

- Oracle 10g sous Windows, G. Briard, Eyrolles, 2006
- Oracle 10g, Guide du DBA, Kevin Loney, Bob Bryla, Oracle Press, 2005
- Gestion et administration des bases de données, Application à Sybase et Oracle, N. Boudjlida, Dunod, 2003
- G. et O. Gardarin, Le client-serveur, Eyrolles 1996
- E. Whalen, Oracle performance tuning and optimization, SAMS Publishing 1996

STA 1604 : Etudes de Cas

Modalités pédagogiques :

Cours : 20 h

TD : 22 h

ECTS : 5

Objectif : permettre aux étudiants de consolider ou de découvrir des techniques statistiques en les appliquant sur un cas concret.

Prérequis :

- Modèle linéaire.
- Régression.
- Techniques d'estimation et de tests d'hypothèses.
- Analyse Multidimensionnelle.

Programme :

Dans le cadre de cette UE, il est proposé aux étudiants un jeu de données qu'ils devront traiter, analyser, interpréter en appliquant les méthodes statistiques adaptées.

Les travaux sont menés en groupe et supervisés par les intervenants.

