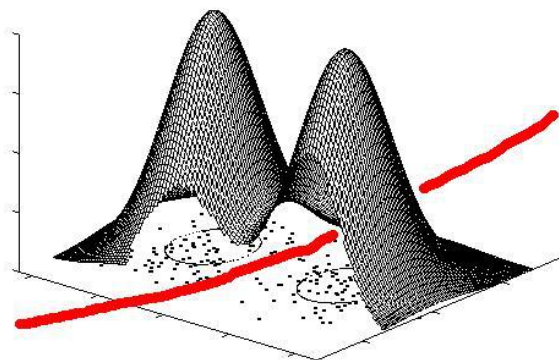


MASTER 2 Sciences, Technologies, Santé
Mention Mathématiques et Applications
Parcours Ingénierie Statistique Numérique – Data sciences

Année universitaire 2020 – 2021



<https://sciences-technologies.univ-lille.fr/les-departements-de-formation/mathematiques/>

PRESENTATION

OBJECTIFS

Le master 2 d'Ingénierie Statistique et Numérique (ISN) a été créé en 1990. Son contenu est régulièrement adapté à l'évolution des moyens, des méthodes et des besoins des entreprises, en particulier en observant les cursus professionnels des étudiants diplômés (plus de 300). Le master ISN fournit un savoir-faire :

- En statistique (analyse des données, théorie de l'apprentissage, apprentissage par renforcement, séries temporelles, statistiques appliquées au marketing et à la biologie)
- En calcul scientifique (recherche opérationnelle, méthodes numériques)
- En informatique (programmation, construction et gestion de bases de données, virtualisation, cloud computing)
- Ainsi qu'une approche de la culture générale de l'entreprise (marketing, gestion, anglais)

Les étudiants ont accès à tous les outils informatiques de statistique et de calcul en accompagnement des cours. Tout se déroule (sauf l'anglais) dans un espace réservé ISN où ils travaillent toute la semaine, avec un équipement informatique adapté. Il est possible de suivre également le M2 ISN en alternance avec une présence régulière en entreprise entre septembre et février, puis à temps plein à compter de mars.

EQUIPE ENSEIGNANTE

Elle se compose de personnels du laboratoire de Mathématiques Paul Painlevé de Lille ainsi que d'autres enseignants universitaires. L'implication de nombreux intervenants issus du milieu professionnel assure aux étudiants une ouverture d'esprit et des contacts utiles (projets, stages, emplois) pour appréhender au mieux la palette des métiers possibles à l'issue de leur formation.

DEBOUCHES

Les diplômés du master ISN sont typiquement recrutés avec un statut de cadre au sein des entreprises de toutes tailles. Ils sont appréciés pour leur esprit d'initiative et pour leur profil vite opérationnel. Les fonctions exercées sont celles de :

- Responsable (qualité, production, organisation, prospection, prévision, ciblage, études et scores, plan commercial)
- Chargé d'études (intelligence artificielle, biostatistique, data-mining, datascience, marketing, crédit)
- Ingénieur (consultant développement, informatique décisionnelle, administrateur bases de données, big data, jeux vidéo)

Le master ISN est en phase avec les débouchés offerts par les grandes entreprises régionales et nationales :

- Dans le secteur de la vente à distance (La Redoute, les Trois Suisses, Damart...)
- Dans les administrations publiques ou territoriales (INSEE, INRETS, ASSEDIC, DRASS, SNCF...)
- Dans le secteur banque/assurance (Crédit Agricole, Banque Directe, CACF, La Poste, Cofidis, Banque Accord, Swiss Life...)
- Dans le domaine des services (Altao, Softcomputing...)
- Dans de nombreux autres secteurs d'activités (Décathlon, Leroy Merlin, France Télécom, ArcelorMittal, Ikea, EDF...)

Ces dernières années, il n'est pas rare de voir les étudiants contactés par les entreprises pour un CDI avant même la soutenance de leur diplôme. Une possibilité de poursuite d'études en doctorat existe aussi occasionnellement (en moyenne, un étudiant par an).



TEMOIGNAGES

Ronald OMORES (Promotion 2018) - Consultant Data science chez Soft Computing Paris :

J'ai intégré le master 2 ISN à la suite d'un premier master en Statistique-Econométrie à la Sorbonne et d'une première expérience en entreprise. J'ai émis le souhait de faire un second master 2 afin de renforcer mes compétences sur les sujets d'analyse de données (Machine Learning) et aussi du numérique. De tous les masters de statistique, le master ISN cochant pour moi toutes les cases en termes de thématiques abordées, mais aussi de masse horaire consacrée. A l'issue de cette année supplémentaire, j'ai rejoint les équipes Data science de Soft Computing d'abord pour mon stage de fin d'année, puis ensuite comme consultant. Au sein de Soft Computing, je travaille sur des sujets divers comme entre-autre la connaissance client pour le marketing, l'analyse des réseaux sociaux, le Deep Learning, mais aussi la construction de modèles de lutte contre le blanchiment d'argent.

Noamane CHALLIOUI (Promotion 2017) – Data Scientist chez Mobivia :

À l'issue de la première année MAS, j'ai effectué un stage chez CEREMA. Les connaissances acquises durant cette première année étaient indispensables pour réussir mes missions autour de la modélisation statistique. Par la suite j'ai intégré le M2 ISN en alternance chez Humanis, où j'ai pu développer mon expérience professionnelle tout en consolidant un bagage scientifique de qualité. J'occupe actuellement un poste de Data Scientist au sein du groupe Mobivia (Norauto, Midas, Carter-Cash, ATU...) où je participe à la mise en place d'une architecture Big Data dans un environnement de Cloud, et j'évolue dans un milieu avec des projets impliquant la maintenance prédictive, la connaissance client, et la supply chain.

Aurélien MAILLET (Promotion 2012) – Responsable études actuarielles chez Swisslife :

Après avoir passé mon Master 1 à Reims, j'étais à la recherche d'un Master 2 dans mon domaine avec une bonne réputation et un bon réseau. A la sortie du Master, j'ai trouvé un emploi rapidement (moins d'un mois) et j'ai eu le choix entre 3 entreprises : Aviva à Paris, Lincoln (SSII) et Swisslife.

Dorothee THUILLIER (Promotion 2011) – Ingénieure d'études en biostatistiques au CNRS :

A l'entrée en ISN, l'envie de partir dans la recherche médicale était claire. Etre à l'écoute, se faire comprendre des différents acteurs du laboratoire (médecins, chercheurs, biologistes, bioinformaticiens, biostatisticiens), les nombreux projets réalisés en master permettant de développer ces qualités humaines. La diversification des enseignements, l'ouverture aux méthodes numériques, à l'informatique, proposées en M2 ISN me permettent d'accompagner les évolutions techniques du laboratoire et de m'ouvrir notamment à la bioinformatique. Sans connaissance en génétique à mon arrivée, j'ai su gagner la confiance des différents acteurs pour mener à bien des projets de recherche. Cela par l'envie d'apprendre, l'autonomie et savoir se remettre en question. Le master m'a appris à cultiver cette différence au sein du laboratoire aujourd'hui.

Pierre THOREL (Promotion 2010) – Chargé d’activités marketing chez Crédit Agricole Nord de France : Doté d'un goût prononcé pour les activités analytiques et attiré par le monde de l'entreprise, c'est tout naturellement que j'ai opté pour le master ISN à l'issue d'une licence de mathématiques. Le diplôme qu'il délivre est reconnu par les entreprises de la région, et offre une large gamme de compétences en statistiques et en informatique. Il permet d'exercer toute la palette des métiers analytiques, allant de l'informatique décisionnelle au marketing analytique en passant par la gestion du risque et la recherche opérationnelle... A titre personnel, j'ai pu, en 4 ans, exercer les métiers de chargé d'études statistiques, de gestionnaire de projets informatiques, puis de projets bancaires. Aujourd'hui rattaché au marketing analytique, j'ai en charge les activités datamining du service.

Viviane TAHON (Promotions 2008) - Contrôleur de gestion chez Banque Accord : Passionnée par les mathématiques (bac S spé Maths, classes préparatoires MPSI), j'ai suivi un module d'initiation à la statistique durant mon année de licence de maths. Ayant adoré cette option, je me suis orientée vers le Master ISN qui offre des débouchés vers des applications concrètes des statistiques et de l'informatique. Après un stage de fin d'études de 6 mois chez Cofidis, j'ai rejoint Banque Accord en tant qu'analyste risque à l'international pendant un an, le poste étant basé en France et comprenant des déplacements à l'étranger. Puis j'ai occupé la même fonction sur le périmètre français, mais dans un département plus axé pilotage, prévisions et budget, en lien direct avec la direction financière que j'ai intégrée 3 ans et demi après en tant que contrôleur de gestion.

Aurélien DEPLANCKE - Data Analyst chez Crédit Agricole Consumer Finance : Data Analyst au sein de la Direction Crédit de Crédit Agricole Consumer Finance j'ai pour mission principale le pilotage des scores d'octroi de crédit (accorder ou non un prêt à un client particulier ou professionnel) de sa conception au suivi des indicateurs et éventuellement jusqu'à sa refonte. Ce métier exige de solides connaissances en analyses de données (Discrétisation, ACM, régression logistique, ...) ainsi qu'une bonne gestion du temps. En effet la conception d'un score d'octroi s'étale sur plusieurs mois, il faut donc savoir gérer son emploi du temps avec les autres missions confiées. Ayant une forte appétence pour le domaine scientifique je me suis finalement orienté vers les mathématiques. D'abord vers la théorie mathématique (Licence puis M1 Recherche) et enfin vers les mathématiques appliquées. C'est la réputation de l'Université de Lille I et du master ISN qui ont orienté mon choix. Sans attente particulière ce n'est que vers la fin du cursus que je me suis rendu compte des opportunités offertes par le Master. En effet celui-ci étant reconnu dans la métropole lilloise voire en Ile de France il permet de trouver un emploi sans difficulté dès l'obtention du diplôme. La qualité de l'enseignement au sein du Master et l'exigence demandée aux étudiants pour leurs travaux permet de rejoindre de grands groupes dans le domaine de la Grande Distribution, de la Banque ou du Crédit à la Consommation ainsi que des entreprises de services du numérique ou encore des PME dans des domaines variés. Suite à un stage de fin d'étude en tant que chargé d'études statistiques au Crédit Agricole Consumer Finance j'ai choisi de rejoindre une PME dans le domaine de la Stratégie Hospitalière. Mes connaissances en Bases de Données et mon expérience avec le logiciel SAS m'ont rapidement amené à occuper une place de référent dans l'équipe Etudes. Après 6 ans j'ai choisi de retourner au Crédit Agricole Finance du côté du service Marketing. Ce poste

m'a permis de développer mes compétences en communication orale et écrite et d'échanger avec des Responsables de Produits et/ou de Marchés. 4 années plus tard j'ai souhaité évoluer en interne en rejoignant une nouvelle Direction. Récemment arrivé au poste que j'occupe je souhaite y rester à court terme afin de développer plusieurs scores d'octroi (Auto, Partenaires Grande Distribution, Partenaires bancaires, ...). A moyen terme j'aimerais m'orienter vers le management d'équipes d'études statistiques. Le Master ISN est un bon tremplin pour acquérir cette souplesse et cette diversité professionnelles.

Mathieu LEFEBVRE (Promotion 2001) - Directeur Risque, Développement et Operations chez Oney Banque Russie : J'ai choisi de faire le master ISN pour sa formation concrète et appliquée. Je voulais comprendre les usages des mathématiques et les mettre en œuvre. J'ai eu l'opportunité dans ma vie professionnelle de couvrir beaucoup de fonctions. J'ai commencé analyste risque chez Finaref puis grâce au réseau du DESS ISN j'ai eu l'opportunité de changer et de rejoindre Oney en tant qu'analyste risque, puis contrôleur de gestion groupe, puis de directeur financier de la filiale hongroise et finalement de rejoindre la Russie. Les mathématiques sont toujours très présentes dans mon quotidien (Modèles de régression, extrapolation, segmentation, modélisation de comportement client, big data). Au-delà de ces aspects les mathématiques apportent une vision logique, un esprit d'analyse qui sert à tous les niveaux. J'aime relever des challenges et j'espère que le futur m'en apportera de nouveaux !

ORGANISATION DES ETUDES 2020 - 2021

SEMESTRE 3

Outils statistiques et numériques pour les Datasciences, niveau expert	Machine learning	<ul style="list-style-type: none"> - Théorie de l'apprentissage - Méthodes d'apprentissage - Apprentissage autonome 	12 ECTS
	Modélisation : méthodes et outils	<ul style="list-style-type: none"> - Statistiques spatiales - Séries temporelles - Recherche opérationnelle - Méthodes pour l'ingénieur 	12 ECTS
	Outils informatiques pour les big data	<ul style="list-style-type: none"> - Programmation avancée - Base de données et web - Virtualisation et conteneurisation 	6 ECTS

SEMESTRE 4

Au choix : Orientation entreprise	Language et culture d'entreprise	<ul style="list-style-type: none"> - Anglais - Intervenants conférences métiers - Statistiques pour l'entreprise 	9 ECTS
	Projets et stage	<ul style="list-style-type: none"> - Projets de statistiques - Stage 	21 ECTS
Au choix : Orientation recherche	Spécialité recherche 1		9 ECTS
	Spécialité recherche 2		9 ECTS
	Mémoire ou stage		12 ECTS

CONTACTS

Responsable :

NICOLAS WICKER

Laboratoire Paul Painlevé UMR CNRS 8254
Université de Lille
Faculté des sciences et technologies
Département de Mathématiques
Cité scientifique - Bâtiment M2
59655 VILLENEUVE D'ASCQ CEDEX (France)
Mail : nicolas.wicker@univ-lille.fr

Secrétariat pédagogique :

Aurore SMETS

Université de Lille
Faculté des sciences et technologies
Département de Mathématiques
Cité scientifique - Bâtiment M2 – Bureau 010
59655 VILLENEUVE D'ASCQ CEDEX (France)
Mail : math-masters2@univ-lille.fr
Tel : +33 (0) 3.20.43.42.33

Association des Anciens :

Association "A.I.S.N." : <http://aisn.free.fr/>

ADMISSION

PHASE 1 : CANDIDATURE (PRÉ-REQUIS ET PUBLIC CONCERNÉ)

S'agissant des étudiants ayant validé le master 1 de la mention Mathématiques Appliquées, Statistique de l'Université de Lille, l'accès au master 2 Mathématiques et applications, parcours Ingénierie Statistique et Numérique - Data sciences (ISN) est de droit ; il vous appartiendra donc de procéder directement à votre inscription administrative (cf phase 2).

S'agissant des étudiants ayant validé un autre master 1 d'un parcours national de master dans une mention compatible (Ingénierie Mathématique, Econométrie, Mathématiques Appliquées aux Sciences Sociales, Mathématiques Appliquées aux Sciences Humaines) ou un diplôme à l'étranger (hors procédure Campusfrance), l'accès en master 2 ISN se fait par sélection sur dossier, suivi le cas échéant d'un entretien de motivation. **Toute candidature doit passer par la plateforme ecandidat :**

<https://ecandidat.univ-lille.fr/>

Par ailleurs, le master peut accueillir en formation continue des étudiants issus d'une entreprise ou des demandeurs d'emploi.

Les auditeurs souhaitant suivre la formation dans le cadre de la formation continue pourront se renseigner auprès du Service de la Formation Continue (SFC) :

<http://formation-continue.univ-lille.fr/>

Toutefois, la procédure de candidature est la même que citée précédemment (<https://ecandidat.univ-lille.fr/>).

En outre, il est possible de suivre la formation en alternance dans le cadre d'un contrat d'apprentissage. L'apprentissage impliquant une charge de travail plus importante, il est réservé à des étudiants ayant eu de bons résultats en première année et reste soumis à l'approbation du responsable de la formation.

PHASE 2 : INSCRIPTION

Une fois que la décision d'acceptation dans la formation a été notifiée, l'**inscription administrative** se fait au début de l'année universitaire (le calendrier est en ligne sur le site de l'université de Lille : <https://www.univ-lille.fr/etudes/candidater-sinscrire/>) auprès des services administratifs de l'université.

Elle sera complétée, à la rentrée, par l'**inscription pédagogique** qui permet d'établir le contrat pédagogique de chaque étudiant et de l'inscrire aux examens correspondants (se rapprocher du secrétariat pédagogique).

PROGRAMMES DES COURS

2020 - 2021

- **S3 : Théorie de l'apprentissage – 20h CM et 20h TD**

Objectif

Savoir calculer des bornes sur les risques théoriques des méthodes d'apprentissage.

Programme

- Rappel sur les inégalités de concentration
- Risque théorique et empirique en apprentissage
- Nombres de croissance et dimension de Vapnik-Chervonenkis
- Complexité de Rademacher
- Nombres de couverture et dimension de pulvérisation

Compétences acquises (directes/indirectes) :

Compréhension des capacités de prédiction des méthodes d'apprentissage.

Modalités d'évaluation

Examen final (3h)

- **S3 : Méthodes d'apprentissage – 20h CM et 20h TD**

Objectif

Mettre en œuvre des techniques standards par des logiciels spécialisés.

Mettre en concurrence ces méthodes pour sélectionner la plus adaptée.

Définir un plan d'expériences en amont de l'acquisition des données.

Programme

- Classification supervisée :
 - Analyse factorielle discriminante
 - Méthodes probabilistes : modèle de mélange Gaussien, régression logistique binaire et polytomique, méthode non paramétrique des k plus proches voisins
 - Méthodes de segmentation : arbre de classification CART (construction, élagage de l'arbre), forêts aléatoires
 - Méthodes d'ensemble : bagging, boosting
- Classification non supervisée :
 - Méthode des k-means
 - Classification Ascendante Hiérarchique :
 - Méthodes de partitionnement spectral
 - Méthodes probabilistes

Compétences acquises (directes/indirectes) :

Identifier une problématique comme relevant de l'analyse des données

Interpréter/présenter les résultats en vue d'un échange avec des non-statisticiens

Modalités d'évaluation

2/3 * Examen final (2h) + 1/3 * mémoire

- **S3 : Apprentissage autonome – 20h CM et 20h TD**

Objectif

Mettre en œuvre des techniques standards par des logiciels spécialisés.

Mettre en concurrence ces méthodes pour sélectionner la plus adaptée.

Programme

- Apprentissage par renforcement
- Processus markovien de décision
- Choix d'une politique d'apprentissage : Q-learning, SARSA
- Utilisation des réseaux de neurones
- Recherche arborescente par Monte-Carlo

Compétences acquises (directes/indirectes) :

Modélisation d'un problème en vue d'un apprentissage autonome.

Interpréter/présenter les résultats en vue d'un échange avec des non-statisticiens.

Modalités d'évaluation

2/3 * Examen final (2h) + 1/3 * projet

- **S3 : Statistiques spatiales – 15h CM et 15h TD**

Objectif

Ce cours de statistique spatiale vise à donner aux étudiants les bases théoriques et pratiques de la statistique spatiale, pour des échantillons de taille finie dans un premier temps et puis dans le cadre asymptotique.

Les statistiques spatiale et spatio-temporelle comprennent toutes les techniques (statistiques) qui étudient les phénomènes observés sur les ensembles spatiaux ou spatio-temporels. De tels phénomènes apparaissent dans une variété de domaines : épidémiologie, sciences de l'environnement, physique, économétrie, traitement d'images et bien d'autres. La modélisation des données spatiales ou spatio-temporelles est motivée par le nombre croissant de situations provenant de différents domaines où les données sont observées dans une zone géographique (régions, communes, stations, pays, pixels d'une image,). C'est le cas, par exemple, en épidémiologie, économie, environnement, où les données sont souvent spatiales ou spatio-temporelles, de sorte que l'emplacement spatial peut servir de substitut aux facteurs de risque.

Programme

Les différents types de données spatiales y sont abordés ainsi que la modélisation spatiale dans un cadre géostatistique, latticiel ou de processus ponctuels spatiaux :

- Différents types de données spatiales
- Dépendance spatiale
- Géostatistique : Préviation spatiale (krigeage, ...)

- Processus ponctuels spatiaux
- Économétrie spatiale (processus autorégressifs, modèle linéaire spatiale, ...)

Compétences acquises (directes/indirectes) :

Connaissance des propriétés statistiques des outils d'inférence (estimation et tests) dans le contexte de données spatiales, spatio-temporelles et leurs applications

Modalités d'évaluation

CC, CT

- **S3 : Séries temporelles – 15h CM et 15h TD**

Objectif

Objectifs (en termes de savoir-faire) : Savoir modéliser une série temporelle, savoir réaliser des prévisions de séries temporelles, savoir établir des intervalles de confiance pour ces prévisions.

Programme

- Séries temporelles univariées à temps discret : définitions et exemples.
- Etude des composantes tendancielle et saisonnière d'une série :
 - Désaisonnalisation par différentes méthodes
 - Lissages par moyenne mobile et lissages exponentiels
- Généralités sur les processus :
 - Processus stationnaires du second ordre
 - Autocorrélations et autocorrélations partielles
 - Estimation de ces caractéristiques
- Prévision par des méthodes paramétriques :
 - Modèles ARMA : identification, prévision, intervalles de prévision
 - Modèles ARIMA et SARIMA
- Extensions :
 - Modèles non linéaires et modèles multivariés
- Applications : Travaux pratiques permettant d'implanter et de tester les méthodes sur des données réelles ou simulées

Compétences acquises (directes/indirectes) :

Maîtrise des techniques usuelles pour la modélisation et la prévision des séries temporelles.

Modalités d'évaluation

Un examen écrit de 2h + un projet (pondérations : 2/3 – 1/3)

- **S3 : Recherche opérationnelle – 15h CM et 15h TD**

Objectif

Savoir identifier et modéliser un problème d'optimisation combinatoire.

Savoir résoudre un problème d'optimisation combinatoire par le logiciel spécialisé AMPL.

Savoir interpréter l'impact économique d'un problème d'optimisation.

Savoir identifier le lien avec des classes et techniques classiques d'optimisation et de la théorie des graphes.

Programme

- Modélisation et interprétation économique de l'optimisation linéaire
- Traitement numérique par le logiciel AMPL
- Techniques de modélisation par nombres entiers et variables bivalentes
- Problèmes d'optimisation et algorithmique sur les graphes

Compétences acquises (directes/indirectes) :

Savoir se servir des techniques modernes d'optimisation dans un processus d'aide à la décision en entreprise

Modalités d'évaluation

Note finale = $(DM1+DM2+2*CT) / 4$: DM1 et DM2 devoirs maison, CT examen écrit

- **S3 : Méthodes pour l'ingénieur – 15h CM et 15h TD**

Objectif

Dans de nombreux domaines (image, séries temporelles, etc.), la manipulation, la compression ainsi que l'extraction d'informations d'un signal sont vitales pour comprendre ses caractéristiques. Ce cours présentera les bases des principales méthodes numériques pour cela, ainsi que des applications concrètes.

Programme

- Méthodes spectrales
- Décomposition en valeurs singulières de grandes matrices
- Applications

Compétences acquises (directes/indirectes) :

Traitement de signal en grande dimension/large quantité ; détection de formes ; compression ; etc.

Modalités d'évaluation

1/3 CC + 1/4 Projet + 7/12 Examen terminal

- **S3 : Programmation avancée – 15h CM et 15h TD**

Objectif

Maîtriser des outils de développement

Programme

- Concepts de programmation objets POO
- Modélisation UML
- Phases de développement informatique (CDC, analyse, tests, versions)
- Langage de programmation (techniques de programmation)

- Développement de solution distribuée

Compétences acquises (directes/indirectes) :

Compréhension des infrastructures informatiques et savoir-faire dans le développement de logiciels

Modalités d'évaluation

Examen final (3h)

- **S3 : Base de données et web – 15h CM et 15h TD**

Objectif

Apprécier la qualité d'une base de données

Maîtriser les bases d'un langage d'interrogation SQL

Développer des sites web interagissant avec des bases de données

Programme

- Schéma conceptuel d'une base de données
- Schéma relationnel et opérateurs algébriques
- Langage de requête
- Langage SQL
- Langage de programmation web
- Interaction avec une base de données

Compétences acquises (directes/indirectes) :

Savoir élaborer et mettre en place un site web interagissant avec des données

Modalités d'évaluation

Examen final (4h)

- **S3 : Virtualisation et conteneurisation – 9h CM et 9h TD**

Objectif

Comprendre les nouvelles infrastructures informatiques.

Programme

- Concept des systèmes d'exploitation
- Concept de la virtualisation
- Concept de la conteneurisation et des orchestrateurs SWARM et Kubernetes

Compétences acquises (directes/indirectes) :

Maîtrise des environnements de développement et de production orientation devops

Modalités d'évaluation

Examen final (2h)

- **S4 : Anglais – 20hCTD**

Objectif

L'anglais de communication scientifique sera abordé par le biais de l'anglais de spécialité (Compréhension de textes de spécialité, de vulgarisation scientifique, rapide synthèse et présentation sur Power-Point etc...).

Programme

On abordera la pratique de l'anglais de communication professionnelle en entreprise au travers de mises en situation, de jeux de rôle etc... (rédaction de CV, lettre de motivation, entretien etc...). Ce travail s'accompagnera d'un entraînement aux certifications. (Ici, certification TOEIC)

Compétences acquises (directes/indirectes) :

Dans les 4 compétences de compréhension et de production, on vise le **niveau B2** du Cadre Européen Commun de Référence en langues [CECR].

Sensibilisation aux différentes certifications et examens de langues afin de mieux préparer leur projet professionnel.

Modalités d'évaluation

L'évaluation se compose d'une note de présentation orale (PE (powerpoint) +PO et IO) et d'un 'test blanc' de certification type TOEIC qui donnera lieu à une note/20 :

1^{ère} session : Note/20 Présentation orale + Note eq. TOEIC Blanc/20 = NOTE GLOBALE/20

2^{ème} session : Note eq. TOEIC blanc/20. (+ Bonus éventuel)

- **S4 : Statistiques pour l'entreprise – 15h CM et 15h TD**

Objectif

Au premier semestre, les étudiants ont vu les estimateurs et les tests dans le cas où la loi des observations était connue à un paramètre près. Quand la loi n'est pas connue, on peut toutefois estimer l'espérance des données, et en donner un intervalle de confiance.

Grâce au théorème centrale limite, on peut construire un intervalle de confiance et estimer l'espérance de variables aléatoires. Cependant, si on dispose de peu de données ou si les queues de distribution sont trop lourdes, la moyenne est trop instable. Dans ce cas, on peut utiliser la médiane, qui est plus robuste, pour construire des intervalles de confiance et des tests.

Programme

- Estimation de la fonction de répartition et la densité
- Tests sur la loi (Kolmogorov-Smirnov)
- Tests sur la médiane
- Tests de comparaison d'échantillons
- Mesure d'association (données bivariées)

Compétences acquises (directes/indirectes) :

Comprendre le principe des tests non paramétriques : la loi de la statistique de test ne doit pas dépendre de la loi des données.

Pouvoir traiter des échantillons de petite taille.

Estimation non paramétrique d'une fonction (compromis biais/variance, choix de la fenêtre).

Modalités d'évaluation

Examen final