



# Programme des enseignements de 2<sup>ème</sup> année

ANNÉE SCOLAIRE 2023 / 2024



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

# PROGRAMME DES ENSEIGNEMENTS DE 2<sup>ème</sup> ANNÉE

## ANNÉE SCOLAIRE 2023-2024

### Table des matières

## Table des matières

<b>Présentation générale des enseignements .....</b>	<b>4</b>
<b>Les grands domaines d'enseignement .....</b>	<b>7</b>
<b>Enseignements de 2e année .....</b>	<b>10</b>
<b>Volumes horaires enseignés.....</b>	<b>10</b>
<b>Les cours électifs .....</b>	<b>12</b>
<b>Enseignements du premier semestre .....</b>	<b>17</b>
<b>UE 2-01 : Modélisation statistique .....</b>	<b>18</b>
<b>REGRESSION LINEAIRE .....</b>	<b>19</b>
<b>MODELES LINEAIRES GENERALISES .....</b>	<b>20</b>
<b>SERIES TEMPORELLES 1 .....</b>	<b>21</b>
<b>UE 2-02 : COLLECTE ET APPRENTISSAGE .....</b>	<b>22</b>
<b>APPRENTISSAGE SUPERVISE .....</b>	<b>23</b>
<b>THÉORIE DES SONDAGES.....</b>	<b>24</b>
<b>DROIT ET ETHIQUE DES DONNEES / SECURITE DES DONNEES .....</b>	<b>26</b>
<b>UE 2-03 : INFORMATIQUE POUR LA DATA SCIENCE .....</b>	<b>27</b>
<b>COMPLEMENT D'INFORMATIQUE .....</b>	<b>28</b>
<b>PROJET D'INFORMATIQUE .....</b>	<b>30</b>
<b>REMISE A NIVEAU EN INFORMATIQUE .....</b>	<b>32</b>
<b>UE 2-04 : MODELISATION ECONOMIQUE.....</b>	<b>34</b>
<b>MODELISATION MICROECONOMIQUE .....</b>	<b>35</b>
<b>ECONOMETRIE APPLIQUEE .....</b>	<b>36</b>
<b>UE 2-05 : HUMANITES ET STAGE OPERATEUR .....</b>	<b>37</b>
<b>ANGLAIS .....</b>	<b>39</b>
<b>Enseignements du second semestre : .....</b>	<b>41</b>
<b>Tronc commun .....</b>	<b>41</b>
<b>UE 2-06 : PROJET STATISTIQUE.....</b>	<b>42</b>
<b>UE 2-07 : COMPLEMENTS DE MODELISATION STATISTIQUE .....</b>	<b>43</b>
<b>INTRODUCTION AUX MODÈLES DE DURÉE.....</b>	<b>44</b>
<b>STATISTIQUE NON PARAMETRIQUE .....</b>	<b>45</b>
<b>TRAITEMENT DE LA NON-REPONSE DANS LES ENQUETES .....</b>	<b>46</b>
<b>CHAINES DE MARKOV.....</b>	<b>48</b>
<b>SÉRIES TEMPORELLES 2 .....</b>	<b>49</b>
<b>UE 2-08 : METHODES COMPUTATIONNELLES .....</b>	<b>50</b>
<b>CALCULS BAYESIENS.....</b>	<b>51</b>
<b>MICROECONOMETRIE APPLIQUEE .....</b>	<b>52</b>
<b>MACROÉCONOMÉTRIE APPLIQUÉE.....</b>	<b>53</b>

<b>PROGRAMMATION ORIENTEE OBJET AVEC JAVA .....</b>	<b>54</b>
<b>C++ POUR LA STATISTIQUE COMPUTATIONNELLE .....</b>	<b>55</b>
<b>OUTILS POUR LE BIG DATA.....</b>	<b>57</b>
<b>Enseignements du second semestre :.....</b>	<b>59</b>
<b>cours électifs (pré-spécialisation UE 2-09).....</b>	<b>59</b>
<b>UE 2-09 : PRE-SPECIALISATION.....</b>	<b>60</b>
<b>MODELES AVANCÉS DE RÉGRESSION .....</b>	<b>61</b>
<b>MARTINGALES ET PROCESSUS DE LÉVY .....</b>	<b>63</b>
<b>MÉTHODES DE RÉÉCHANTILLONNAGE.....</b>	<b>64</b>
<b>STATISTIQUE MATHÉMATIQUE .....</b>	<b>66</b>
<b>ÉCHANTILLONNAGE À PLUSIEURS DEGRÉS .....</b>	<b>67</b>
<b>STATISTIQUE SPATIALE .....</b>	<b>69</b>
<b>ECONOMIE DU RISQUE.....</b>	<b>70</b>
<b>ECONOMIE NUMERIQUE .....</b>	<b>71</b>
<b>DEMOGRAPHIE .....</b>	<b>74</b>
<b>PROGRAMMATION AVANCÉE EN R.....</b>	<b>76</b>
<b>CLOUD COMPUTING .....</b>	<b>77</b>
<b>TRAITEMENT DU SIGNAL .....</b>	<b>79</b>
<b>CONCEPTION DE LOGICIEL.....</b>	<b>81</b>
<b>DATA CHALLENGE.....</b>	<b>82</b>
<b>DATAVISUALISATION .....</b>	<b>83</b>
<b>UE 2-10 : PROJET PROFESSIONNEL.....</b>	<b>85</b>
<b>TEST DE PERSONNALITE DANS LE MILIEU PROFESSIONNEL .....</b>	<b>86</b>
<b>CONSTRUIRE SON PROJET PROFESSIONNEL.....</b>	<b>87</b>
<b>PROJET PROFESSIONNEL DES ATTACHÉS.....</b>	<b>88</b>
<b>STAGE D'APPLICATION .....</b>	<b>89</b>
<b>PRINCIPES DE GESTION DURABLE DES ORGANISATIONS .....</b>	<b>90</b>
<b>COURS LIBRE OPTIONNEL.....</b>	<b>91</b>
<b>DÉVELOPPEMENT DURABLE/RSE ET L'ÉTHIQUE .....</b>	<b>93</b>
<b>SPORTS .....</b>	<b>94</b>
<b>R-SHINY .....</b>	<b>95</b>
<b>VISUAL BASIC APPLICATION.....</b>	<b>96</b>
<b>COMPLEMENTS DE SAS .....</b>	<b>97</b>
<b>HYGIENE NUMERIQUE .....</b>	<b>98</b>

## Présentation générale des enseignements

---

Créée il y a 20 ans, l'ENSAI est l'une des deux grandes écoles d'ingénieurs avec l'ENSAE à être spécialisée dans le traitement de l'information et la statistique. Le secteur de la statistique est en pleine croissance, la donnée est devenue un actif stratégique des entreprises, et le métier de statisticien a évolué vers celui de Data Scientist. Par ses six filières de spécialisation, l'école offre ainsi des compétences reconnues dans des secteurs d'activités diversifiés, de l'industrie à la banque, en passant par les services aux entreprises ou la santé, en France ou à l'étranger.

L'ENSAI forme à la modélisation statistique, avec des compétences associées en économie et en informatique. Les compétences sont à la fois théoriques et opérationnelles, avec une grande part de la formation dédiée aux applications. La scolarité se déroule en trois ans pour les étudiants ingénieurs et en deux ans pour les étudiants fonctionnaires (attachés statisticiens de l'INSEE) avec, pour ces derniers, la possibilité d'obtenir un Master Science des Données pour la Décision Publique, parcours « évaluation et décisions publiques », co-accrédité avec les universités de Rennes 1 et Rennes 2, l'Insa de Rennes, Agrocampus Ouest.

Durant les deux premières années de scolarité à l'ENSAI, les étudiants ingénieurs et les étudiants fonctionnaires suivent en commun la majorité des enseignements.

### PREMIERE ANNEE : HOMOGENEISER LES COMPETENCES ET ACQUERIR LES CONNAISSANCES SCIENTIFIQUES DE BASE

La diversité des étudiants recrutés avec des connaissances préalables plus fortes en économie ou mathématiques oblige à de premiers enseignements fondamentaux distingués au premier semestre selon le cursus antérieur. Ainsi les étudiants venant de la voie « mathématiques » (concours communs polytechniques, L3 math...) ou des IUT Stid ont un enseignement renforcé en économie au 1<sup>er</sup> semestre pour rattraper leur retard par rapport aux étudiants venant de la voie « économie ». De façon symétrique, les étudiants venant de la voie « économie » (prépa BL, Cachan D2, L3 économie...) suivent des cours complémentaires de mathématiques (algèbre, TD d'analyse) pour acquérir les bases utiles dans l'apprentissage ultérieur des statistiques.

La pédagogie des enseignements des probabilités est aussi adaptée à ces différences de cursus d'origine, pour faciliter l'assimilation des notions nouvelles. En informatique, les étudiants issus de la voie « économie » et de la voie « IUT Stid » suivent un cours d'algorithmique, complexité et de calculabilité tandis que les autres étudiants bénéficient d'un cours d'introduction à l'algorithmique et à la programmation.

A l'issue de cette première année, tous les étudiants auront les connaissances scientifiques de base en statistique, économie et informatique. Ils sauront mener une étude descriptive à partir d'une base de données réelle, mettre en œuvre les premiers modèles statistiques, juger des qualités de différents algorithmes, et relier des problèmes économiques contemporains à la théorie économique. Les étudiants sont formés à différents langages informatiques qui les rendront agiles et opérationnels dans leur vie professionnelle future.

Un stage d'un à deux mois conclut cette 1<sup>re</sup> année : stage de découverte de la statistique publique pour les attachés stagiaires et stage opérateur pour les étudiants ingénieurs.

### DEUXIEME ANNEE : APPROFONDIR ET COMMENCER A SE SPECIALISER

C'est l'année où les concepts statistiques avancés sont abordés, renforcés par des applications concrètes. Des cours spécialisés aux types de données rencontrées sont ainsi proposés : données temporelles, durée, données économiques, Big Data... L'année est également marquée par deux projets majeurs : l'un en informatique, l'autre en statistique. Les étudiants gagnent en autonomie et des cours électifs sont proposés pour préparer leur spécialisation de 3<sup>e</sup> année (étudiants ingénieurs) ou leur entrée dans le monde professionnel (étudiants fonctionnaires). La présence des cours en anglais est renforcée lors du second semestre. Suite aux retours des entreprises partenaires et l'Insee sur l'importance de l'Anglais dans la formation des étudiants, l'école a décidé de mettre en œuvre un second semestre en anglais, avec un nombre de cours en anglais librement choisi par les étudiants. L'objectif est de donner la possibilité de suivre un cursus en anglais à tous les étudiants pendant un semestre "international", qu'ils soient partis en échange ou non. La mise en place de ce cursus en anglais vise également à permettre à nos partenaires européens de nous envoyer plus d'étudiants et ainsi de pérenniser nos accords Erasmus. Pour les étudiants ingénieurs, l'année se termine par un stage d'application en statistique de 2 à 3 mois.

### TROISIEME ANNEE : SE SPECIALISER

Ingénierie statistique appliquée à l'industrie, aux sciences de la vie, à l'analyse des territoires et de la santé, au traitement informatique de grands volumes de données, au marketing ou à la gestion des risques et à l'ingénierie financière... En fin de 2<sup>e</sup> année, tous les étudiants ingénieurs choisissent une filière de spécialisation dans laquelle interviennent de nombreux professionnels et ponctuée par un stage de 6 mois. Pour les étudiants fonctionnaires, c'est le choix d'une option dans le parcours « évaluation et décision publiques » du Master Science des Données pour la Décision Publique, selon une approche métier : études statistiques, méthodologie statistique ou traitements informatiques. Il peut être suivi, sous conditions, directement à la suite de la 2<sup>e</sup> année ou en formation continue.

### LES SIX FILIERES DE SPECIALISATION DE 3<sup>e</sup> ANNEE

La formation d'ingénieur de l'ENSAI inclut 6 filières de spécialisation. Toutes ces filières forment aux métiers de la Data Science, avec une maîtrise des outils permettant l'extraction, l'analyse et la fouille de données et une capacité à choisir les modalités de traitements des données massives (Big Data) et des techniques d'apprentissage automatique (Machine Learning). Selon les spécialisations, ces compétences sont spécifiques à un domaine ou transversales. L'ensemble des filières continue à former aux compétences transversales (Soft Skills) et à la valorisation des travaux menés dans un contexte professionnel et international. La séquence de tronc commun mêlant enseignements scientifiques, projets et anglais conclut la formation à l'autonomie et la capacité à mettre en œuvre des analyses de données en situation complexe.

#### **Data Science & Génie des risques**

Cette filière forme des ingénieurs spécialistes de la finance quantitative, capables d'innover et de proposer de nouvelles méthodes d'analyse. Elle s'articule autour de trois grands domaines de compétences : la réglementation et la gestion des risques bancaires, l'allocation et les stratégies d'investissement, l'innovation en ingénierie financière.

#### **Data Science & Marketing**

Cette filière forme à la fois des Data Scientists ayant une très forte compréhension des enjeux métiers du marketing et des revenus managers aux capacités analytiques et quantitatives très élevées.

#### **Data Science en Santé & Biostatistique**

Cette filière forme au métier de biostatisticien. Elle s'appuie sur des compléments en statistique, et fournit 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.

#### **Data Science & Génie statistique**

Cette filière renforce les connaissances en modélisation statistique, en abordant les thèmes de la qualité et la fiabilité, du traitement de l'image et du signal, ainsi que la prévision et ses applications, notamment dans le domaine de l'environnement. Les étudiants sont ainsi capables de s'adapter à des problématiques provenant de différents secteurs d'activité comme l'industrie, le secteur bancaire, l'environnement, les services...

#### **Data Science & ingénierie des données**

Cette filière vise à renforcer les connaissances en informatique pour l'analyse et la gestion de données, notamment dans le traitement des grandes bases de données. La formation permet une approche de la culture informatique ainsi qu'une présentation approfondie des technologies les plus récentes directement liées à l'analyse des données.

#### **Data Science, Modélisation économique & Santé**

Cette filière donne un bagage en ingénierie statistique et économétrie appliqué à la connaissance des dynamiques territoriales et à la santé, et permettant l'évaluation des politiques publiques. Cette filière ouvre sur de très nombreux domaines de la décision économique, principalement dans le secteur privé (cabinets d'étude, laboratoires pharmaceutiques, bureaux de conseils...) mais également dans le secteur public (ministères, santé, sécurité sociale...).

### **L'OPTION DE FORMATION PAR LA RECHERCHE**

Les étudiants qui souhaitent faire de la recherche théorique ou appliquée après l'ENSAI peuvent bénéficier de facilités offertes au cours de leur scolarité : possibilité, dans le cadre de conventions passées avec des universités, de suivre des cours de Master 2 pendant leur troisième année pour obtenir ce diplôme en même temps que celui de l'ENSAI, d'avoir des contacts privilégiés avec les laboratoires de recherche universitaires et ceux d'autres grandes écoles, avec le Centre de Recherche en Économie et Statistique du Groupe des Écoles Nationales d'Économie et Statistique (CREST), de bénéficier d'un encadrement personnalisé par un « tuteur » spécialiste du domaine dans lequel l'étudiant souhaite poursuivre ses recherches, possibilité d'effectuer le stage de troisième année dans un laboratoire de recherche, etc.

### **LE MASTER MENTION MATHÉMATIQUES APPLIQUÉES-STATISTIQUE, PARCOURS « ÉVALUATION ET DÉCISIONS PUBLIQUES »**

Les étudiants titularisés comme attachés statisticiens de l'INSEE peuvent obtenir un Master en « évaluation et décision publiques » dans le cadre de la formation continue de l'INSEE :

- Intégrée, c'est-à-dire dans le prolongement de leur deuxième année de scolarité à l'ENSAI,
- Décalée, c'est-à-dire de façon discontinue au cours de leurs premières années de fonction (5 ans maximum).

Le Master Science des Données pour la Décision Publique est co-accrédité avec les universités de Rennes 1 et Rennes 2, l'INSA de Rennes, Agrocampus Ouest. Le parcours « évaluation et décision publiques » comporte trois spécialisations au choix : statistiques et traitement des données, méthodologie de la statistique publique ou études statistiques.

## Les grands domaines d'enseignement

---

En dehors de quelques enseignements très spécialisés de troisième année, les cours peuvent être regroupés en quatre grands domaines :

1. Mathématiques, probabilités, statistique
2. Informatique
3. Économie et sciences sociales
4. Humanités

### **MATHEMATIQUES, PROBABILITES, STATISTIQUE**

La statistique fait partie intégrante des mathématiques appliquées. Elle se base sur le calcul des probabilités.

En première année, après des compléments en mathématiques nécessaires à une mise à niveau des étudiants n'ayant pas fréquentés les classes préparatoires scientifiques, quatre cours fondamentaux pour la compréhension scientifique des techniques statistiques enseignées par la suite sont abordés : intégration, probabilités, introduction à la statistique et statistique exploratoire multivariée. Les étudiants réalisent également plusieurs projets statistiques, en groupe, mettant en œuvre des méthodes de statistique descriptive ou des méthodes plus avancées. Les logiciels SAS et R, dédiés à la statistique, font également l'objet d'enseignement.

La seconde année est centrée sur l'apprentissage des techniques utiles au statisticien de profession : la modélisation, paramétrique ou non, d'une régression, l'étude des séries chronologiques modélisables par la méthode de Box-Jenkins, la théorie des sondages, l'analyse des modèles à choix discrets, l'apprentissage supervisé, les chaînes de Markov, les calculs bayésiens, les modèles de durée. Ces bases sont complétées, selon le statut de l'étudiant et ses choix, par une initiation aux processus stochastiques, des cours de méthodes de régression non paramétrique, rééchantillonnage, de compléments de séries temporelles, de statistique mathématique, de cartographie ou d'échantillonnage avancée.

Un projet statistique, encadré par des professionnels et fonctionnant en petits groupes, permet aux étudiants de mettre en œuvre sur des données réelles un large éventail des techniques étudiées au cours des deux premières années. Les étudiants peuvent également participer à un Data Challenge.

Les cours de troisième année s'inscrivent dans des voies de spécialisation. Ils présentent les développements spécifiques des probabilités et de la statistique utiles au domaine étudié, tout en apportant les connaissances indispensables sur l'environnement dans lequel sera amené à travailler le statisticien.

### **INFORMATIQUE**

L'enseignement informatique de première année est adossé à trois concepts principaux : l'algorithmique, la conception d'applications, et le stockage de données. Des liens sont effectués avec les enseignements de statistique. Python est le langage utilisé en 1ère année. L'algorithmique est introduit dans un premier temps avec les notions algorithmiques de base. Les étudiants issus des voies « économie » et « IUT Stid » bénéficient en outre d'un cours portant sur l'algorithmique et la programmation des algorithmes étudiés. Dans un deuxième temps, une introduction à la programmation orientée objet est abordée par tous avec le langage Python. Le lien avec les enseignements de statistique est notamment effectué à travers de TP sous Python dans le cours d'optimisation et méthodes numériques. La conception d'applications est abordée à travers des cours sur la documentation du code (y compris le métalangage de modélisation UML) et d'un projet de traitement de données. Enfin, les fichiers, les bases de données relationnelles et les tables statistiques SAS/R sont les trois principaux modes de stockage des données mis en pratique. Les accès aux fichiers sont abordés lors des cours de programmation objet et du projet. Le langage SQL est l'outil standard de mise en œuvre et d'interrogation de bases de données relationnelles.

Par la suite, tous les étudiants réalisent dès le début de la deuxième année un projet dont l'objectif est de mettre en application les enseignements reçus en première année et des compléments informatiques sur la gestion des bases de données dans des contextes Web et/ou Big Data. Au cours du second semestre, les ingénieurs suivent un cours de programmation orientée objet en C++ ou Java, ainsi qu'une introduction aux outils pour le Big Data. Plusieurs cours optionnels sont proposés en informatique, R avancé, technologies mobiles, conception de logiciel, traitement du signal, et data visualisation. Des compléments informatiques sur des outils (R Shiny, VBA, Libre Office Basics, Compléments de SAS) sont proposés sans ECTS.

La troisième année apporte les compléments nécessaires à la mise en œuvre informatique des méthodes statistiques dans les domaines de spécialisation proposés. Elle offre également une voie d'approfondissement dans le domaine spécifique du traitement de l'information (compléments sur les bases de données, génie logiciel, conception et programmation orientées objet, administration de projets informatiques, intelligence artificielle, Big Data, Datamining, réseaux, technologies Web, etc.).

### **ÉCONOMIE ET SCIENCES SOCIALES**

Les enseignements d'économie, de gestion et de sciences sociales ont pour objectif d'offrir à tous les étudiants une réelle capacité d'analyse et de compréhension des aspects essentiels du monde contemporain, à travers la mobilisation de la modélisation économique et de données fruits d'un comportement humain.

En première année, on distingue un public d'étudiants ayant de bonnes connaissances en sciences économiques et sociales et un public d'étudiants débutants ou n'ayant eu qu'une première initiation dans cette discipline. Pour les premiers, l'école propose un programme de macroéconomie appliquée abordant les développements récents en économie formalisée, et pour les seconds, des cours plus progressifs, avec en particulier une introduction à la modélisation macroéconomique (questions macroéconomiques contemporaines) et une introduction aux sciences sociales.

En deuxième année, le cours d'économétrie du tronc commun vise à donner aux étudiants les méthodes de validation empirique des modèles théoriques. Ce cours peut être complété au second semestre par des cours de micro et macro-économétrie appliquée. Des cours électifs complètent cette formation en ouvrant sur l'économie du risque, l'économie industrielle, l'économie des contrats ou l'économie financière notamment.

### **LES HUMANITES ET LES SOFTS-SKILLS**

Pour nos ingénieurs Data Scientists ou statisticiens publics, la notion de compétence renvoie aux notions d'autonomie et de savoir agir dans le champ de la production et de l'analyse de données et de la décision en situation complexe (tâches non répétitives).

Les connaissances en mathématiques, informatique, statistique, économie (l'ADN de l'ENSAI) forment le socle indispensable à la formation d'un ingénieur Data Scientist compétent.

Les compétences transversales ou Soft Skills sont également indispensables pour valoriser dans un contexte professionnel tout le potentiel acquis durant les 3 années de l'ENSAI.

L'école a produit un référentiel de ces domaines de compétences adapté aux futurs métiers des étudiants qu'elle forme. Il est attendu d'un Data Scientist qu'il présente des qualités importantes dans les domaines suivants : la rigueur intellectuelle, le raisonnement analytique et la conceptualisation, la créativité et le sens de l'innovation, la communication et la pédagogie, les relations humaines, le travail en équipe en mode projet et la gestion du temps, l'ouverture au monde et la responsabilité sociale et civique.

Ces qualités sont mises en œuvre dans les différents processus d'apprentissages : les cours, les TP et TD, le travail personnel demandé.

L'école a fait le choix de renforcer spécifiquement la montée en compétence sur « le savoir agir » de ses étudiants à partir principalement de situations s'approchant le plus de situations professionnelles : les projets, les Data Challenges, les stages, les pratiques associatives.

L'apprentissage des techniques de communication et de la gestion de projet est progressif, basé sur la pratique et le conseil personnalisé. Les étudiants sont sensibilisés à ces compétences dans le cadre du projet de statistique descriptive en première année, puis les développent en travaillant sur les projets qui jalonnent leur scolarité, en particulier le projet statistique de deuxième année, le projet informatique de deuxième année et le projet statistique de troisième année.

Durant le stage d'observation de première année, les étudiants élargiront leurs connaissances du monde professionnel et développeront leur capacité d'analyse à travers la rédaction du rapport.

Le « Projet professionnel » en deuxième et troisième année permet aux étudiants de mieux se connaître à travers, en particulier, l'utilisation d'un test de personnalité très utilisé pour les recrutements, de valoriser leurs compétences dans leur recherche de stage ou d'emploi en préparant les documents supports de leurs candidatures et de mettre en pratique les compétences scientifiques et comportementales durant leur stage d'application en statistique et leur stage de fin d'étude.

L'étude de l'anglais est obligatoire tout au long de la scolarité. Les étudiants sont répartis en groupes de niveau pour assurer un contenu et une pédagogie adaptés. Des modules de préparation au TOEIC sont intégrés aux cours mais l'acquisition et le développement des compétences linguistiques et communicationnelles restent les objectifs principaux des cours. Le niveau B2 du CECRL est obligatoire pour l'obtention du diplôme d'ingénieur mais les cours visent l'acquisition du niveau C1. Au-delà des compétences linguistiques, les cours d'anglais cherchent à équiper les futurs ingénieurs avec les compétences transversales qui leur permettront d'évoluer dans des contextes internationaux variés, d'effectuer la veille scientifique et de comprendre les normes culturelles dans les pays étrangers. En outre, les étudiants ayant un niveau A1 et A2 au CECRL en anglais auront un cours obligatoire d'anglais renforcé.

Les cours des Humanités visent à approfondir les connaissances des étudiants dans plusieurs disciplines des arts et des sciences sociales et humaines. Du fait de leur structure et contenu, ils permettent aux étudiants de développer des compétences transversales, notamment celles liées à l'ouverture au monde et la responsabilité sociale et civique. Les étudiants doivent choisir parmi deux types de cours :

Cours d'ouverture : Dans le cadre de la 2ème année, le cours d'ouverture est remplacé au S1 par deux cours obligatoires: développement durable/RSE (12 heures de cours) et 12h de cours sur l'éthique). Pour le S2, plusieurs options sont organisées. Certains cours visent l'ouverture via la culture générale (géopolitique, histoire de l'art, philosophie...), d'autres stimulent la créativité et la connaissance de soi via la pratique artistique (dessin, musique, peinture, théâtre...) Enfin, la possibilité de participer aux activités associatives ou aux projets spéciaux encourage l'engagement social et citoyen.

## **Enseignements de 2e année**

-

## **Volumes horaires enseignés**

	Total	Cours	TD	ECTS
<b>UE 2-01 : MODELISATION STATISTIQUE</b>				
Régression linéaire	33	15	18	3
Modèles linéaires généralisés	33	21	12	3
Séries temporelles 1	33	18	15	3
<i>Total UE 2-01</i>	99	54	45	9
<b>UE 2-02 : COLLECTE ET APPRENTISSAGE</b>				
Apprentissage supervisé	27	18	9	2,5
Théorie des sondages	30	18	12	3
Droit et éthique des données	9	9	-	0,5
<i>Total UE 2-02</i>	66	45	21	6
<b>UE 2-03 : INFORMATIQUE POUR LA DATA SCIENCE</b>				
Compléments d'informatique	18	6	12	2
Projet Informatique	33	-	33	4
Remise à niveau en informatique (AST seulement)	21	21	-	1
<i>Total UE 2-03</i>	51	6	45	6
<b>UE 2-04 : MODELISATION ECONOMIQUE</b>				
Modélisation microéconomique	21	21	-	2
Econométrie appliquée	33	21	12	3
<i>Total UE 2-04</i>	54	42	12	5
<b>UE 2-05 : HUMANITES ET STAGE OPERATEUR</b>				
Stage opérateur (hors AST)	-	-	-	1
Développement durable/RSE et l'éthique	24	-	24	1
Anglais 1 <sup>er</sup> semestre	30	-	30	2
<i>Total UE 2-05</i>	54	-	54	4
<b>TOTAL 1<sup>ER</sup> SEMESTRE</b>	<b>324</b>	<b>147</b>	<b>177</b>	<b>30</b>
	Total	Cours	TD	ECTS
<b>UE 2-06 : PROJET STATISTIQUE</b>				
Projet statistique	33	-	33	5
<i>Total UE 2-06</i>	33	-	33	5
<b>UE 2-07 : COMPLEMENTS DE MODELISATION STATISTIQUE</b>				
Introduction aux modèles de durée	21	12	9	2
Statistique non paramétrique (ing.) ou Traitement données manquantes enquêtes (att.)	21	21	-	2
Au choix : Chaînes de Markov ou Séries temporelles 2	21	12	9	2
<i>Total UE 2-07</i>	63	45	18	6
<b>UE 2-08 : METHODES COMPUTATIONNELLES</b>				
Au choix: Calculs bayésiens ou, micro et macroéconométrie appliquée	39	21	18	4
Au choix: POO avec Java ou, C++ pour la stat. Computationnelle	18	6	12	2
Outils pour le Big Data	24	10,5	13,5	2
<i>Total UE 2-08</i>	81	37,5	43,5	8
<b>UE 2-09 : PRE-SPECIALISATION</b>				
Cours électif n°1	21	21	-	2
Cours électif n°2	21	21	-	2
Cours électif n°3	21	21	-	2
Cours électif n°4	21	21	-	2
<i>Total UE 2-09</i>	84	84	-	8
<b>UE 2-10 : PROJET PROFESSIONNEL</b>				
Cours libre optionnel n°2	24	-	24	1
Principes de gestion durable des organisations	12	12	-	1
Projet professionnel et gestion de projets	20	-	20	1
Cours outils optionnels	30	-	30	-
<i>Total UE2-10</i>	86	12	74	3
<b>TOTAL 2ND SEMESTRE</b>	<b>347</b>	<b>178,5</b>	<b>168,5</b>	<b>30</b>
<b>TOTAL ANNEE</b>	<b>671</b>	<b>325,5</b>	<b>345,5</b>	<b>60</b>

## Enseignements de 2<sup>e</sup> année

### 2<sup>nd</sup> semestre – UE 2-09

### Les cours électifs

Les élèves doivent choisir 4 cours électifs (parmi la liste des cours figurant dans le tableau ci-dessous). Les choix sont subordonnés aux contraintes d'emploi du temps puisque certains cours sont programmés en même temps.

De plus, **les élèves ingénieurs devront choisir au moins une option statistique**. Le cours de Traitement du signal est considéré comme étant un cours de statistique dans le cadre de cette contrainte.

#### COURS ELECTIFS DE 2<sup>E</sup> ANNEE – UE 2-09

Enseignements		Langue	Volume horaire				
			Cours	TD	TP	Projet	Total
ECONOMIE	Economie du risque	FR	21				21
	Economie numérique	EN	21				21
	Economie industrielle	EN	21				21
	Démographie	FR	21				21
INFORMATIQUE	Programmation avancée en R	FR	21				21
	Cloud computing	FR	21				21
	Traitement du signal	FR	9		12		21
	Conception de logiciel	FR	21				21
	Data Challenge	FR	21				21
	DataVisualisation	EN	21				21
STATISTIQUE	Modèles avancés de régression	EN	21				21
	Martingales et processus de Lévy	FR	12	9			21
	Méthodes de rééchantillonnage	FR	21				21
	Echantillonnage à plusieurs degrés	FR	12	9			21
	Statistique mathématique	EN	21				21
	Statistique spatiale	FR			21		21

## Conseils concernant le choix des cours électifs et des options

Les élèves ont plusieurs choix de cours à suivre :

- Un cours d'approfondissement en statistique dans l'UE 2-07 : Chaînes de Markov ou Séries temporelles 2
- Un cours complémentaire au précédent dans l'UE 2-08 : Calculs bayésiens ou Micro/macro-économétrie. Les étudiants ayant choisi Chaînes de Markov dans l'UE 2-07 devront suivre Calculs bayésiens dans l'UE 2-08. Les étudiants ayant choisi Séries temporelles 2 dans l'UE 2-07 devront suivre Micro/macro-économétrie dans l'UE 2-08.
- Un langage objet entre C++ et Java dans l'UE 2-08. Java est obligatoire pour les élèves attachés.
- Quatre cours (un par bloc) doivent être choisis dans l'UE 2-09. La constitution de chaque bloc est disponible à la page suivante. Les élèves ingénieurs doivent choisir au moins une option statistique parmi ces 4 cours (voir détails page précédente).

Le cours de séries temporelles 2 est présent à la fois dans l'UE 2-07 et l'UE 2-09. Cela permet aux élèves qui souhaitent suivre à la fois le cours de Chaînes de Markov et celui de Séries temporelles 2 de choisir Chaînes de Markov dans l'UE 2-07 puis Séries temporelles 2 dans le bloc 4 de l'UE 2-09. Les élèves dans cette situation peuvent choisir librement entre Calculs bayésiens ou Micro/macro-économétrie pour le choix complémentaire de l'UE 2-08.

Certains enseignements sont requis afin de pouvoir en suivre d'autres. Les prérequis de cours électifs sont les suivants :

- Chaînes de Markov (UE 2-07) est un prérequis pour suivre Calculs bayésiens (UE 2-08)
- Séries temporelles 2 (UE 2-07) est un prérequis pour suivre Micro/macro-économétrie (UE 2-08)
- C++ (UE 2-08) est un prérequis pour suivre R avancé (UE 2-09)
- Calculs bayésiens (UE 2-08) est un prérequis pour suivre Méthodes de rééchantillonnage (UE 2-09)

Une dérogation peut être accordée (à la suite d'une demande spécifique argumentée) à un élève souhaitant choisir un cours sans avoir suivi l'enseignement prérequis. L'élève doit déjà disposer des connaissances et compétences associées au cours prérequis en question. Cela peut également s'appliquer à un élève attaché souhaitant suivre le cours de C++ au lieu de celui de Java normalement obligatoire.

Le choix de 4 cours électifs de l'UE 2-09 (un par bloc) vise à approfondir vos connaissances et à révéler vos préférences avant votre spécialisation de 3ème année (filières ingénieur ou parcours attaché du master évaluation et décision publiques). Il n'y a aucune prise en compte du choix des cours électifs suivis pour l'attribution des filières ou de parcours de master. Certains cours restent néanmoins plus ou moins cohérents avec les spécialisations de dernière année. Le tableau suivant précise les cours adaptés (+) ou très adaptés (++) aux spécialisations de dernière année. Certaines compétences et connaissances associées aux cours définis comme « très adapté » seront considérées comme acquises en début de 3ème année. Ces cours sont grisés dans le tableau de la page suivante.

Ce tableau peut servir à guider certains choix, mais les préférences personnelles de chaque étudiant doivent rester un facteur majeur de ces décisions. Ce tableau ne peut pas synthétiser de manière exhaustive tous les profils étudiants et certains choix de cours dépendent fortement de ce que l'élève souhaite faire à la sortie de l'école. De manière générale, les étudiants sont invités à se renseigner auprès des responsables des filières qu'ils envisagent.

Les cours de « Data Visualisation » et « Data Challenge » sont adaptés à l'ensemble des spécialisations, les compétences mobilisées étant transversales.

Le cours de Data Challenge sera limité en nombre d'étudiants (40) et peut revêtir différentes formes, challenges extérieurs ou organisé à l'Ensaï. Un tirage au sort déterminera si besoin les étudiants.

		Enseignement	Filières (Ing.)					Master SDDP (Att.)			
			GDR	MAR	SB	GS	ID	MES	MSP	ES	STD
UE 07	Choix A	Chaines de Markov	++	++	++	++	++				++
		Séries temporelles 2	++	++	+	+		+	++	+	+
UE 08	Choix B	Calculs bayésiens*		+	++	++	+	+			
		Micro et macro-économétrie*	+	+				+	+	++	
	Choix C	Java		+	+		++	+	++	++	++
		C++	++	+	+	+		+			
UE 09	Bloc 1	Statistique mathématique	+		+	++			+		
		Démographie							++	++	++
		Economie du risque	++	+	+	+		++	+	+	
		Traitement du signal				+	+				+
	Bloc 2	Modèles avancés de régression	+	+	++	++	+	++	++	++	
		Economie industrielle		+	+			++	+	++	
		Conception de logiciel					++		+		++
		Data Challenge	+	+	+	+	+	+	+	+	+
	Bloc 3	Martingales et processus de Lévy	++		+	+					
		Statistique spatiale		+	+			++	++	++	
		Data Visualisation	+	+	++	+	++	+	+	+	+
		R avancé*	+	+	++	++		+			+
	Bloc 4	Méthodes de rééchantillonnage*		+	++	++			++	+	+
		Echantillonnage à plusieurs degrés							++	++	+
		Economie numérique		+	+			++		+	
		Cloud computing		+			++				++
Séries temporelles 2		++	++		+		+	++	+		

**++** Case grisée : Les compétences et connaissances associées à cet enseignement doivent être acquises (pendant le cours ou de manière autonome) pour suivre la filière en troisième année

**\*** Les cours marqués d'une étoile ont un autre enseignement en prérequis (voir page précédente)

GDR : Gestion des Risques

MSP : Méthodologies de la statistique publique

MAR : Marketing

ES : Études statistiques

SB : Santé & biostatistique

STD : Statistique et traitement des données

GS : Génie statistique

ID : Ingénierie des données

MES : Modélisation économique & santé

Pour les étudiants qui suivent le parcours OFPR et qui souhaitent s'orienter vers une formation par la recherche en mathématique, les cours suivants sont fortement conseillés :

Bloc 1: Statistique Mathématique

Bloc 2: Modèles avancés de régression

Bloc 3: Martingales et processus de Levy

Bloc 4: Rééchantillonnage

Les étudiants OFPR qui souhaitent s'orienter vers une recherche autre que mathématique sont invités à se renseigner auprès du responsable de la filière correspondante.

Cours en plus, non soumis à évaluation :

Forum des entreprises (S1) : 1 journée

RDV Carrières (Début S2) : 5 sessions parmi une quinzaine

Cours logiciels : R Shiny (6h), Visual Basic (9h), Libre Office Basic (9h), Compléments de SAS (9h), R markdown ...



## Enseignements du premier semestre

UE 2-01 – Semestre 1

## UE 2-01 : Modélisation statistique

<i>Correspondant de l'UE</i>	: Lionel TRUQUET
<i>Nombre d'ECTS</i>	: 9
<i>Nombre d'heures d'enseignement</i>	: 99h
<p><b>Finalité de l'UE :</b>          Introduire les modèles linéaires ou linéaires généralisés de base qui permettent d'expliquer ou de prédire une variable d'intérêt ou réponse à partir d'un certain nombre de covariables associées. Mettre en œuvre des techniques d'inférence statistique adéquates pour ajuster ces modèles à partir d'un échantillon de données. Cette UE permet de faire le lien entre les notions de statistique descriptive (UE 2 de la première année) et les principes de base de la statistique inférentielle (UE 6 de la première année) à partir du formalisme probabiliste introduit dans l'UE 1 de la première année</p> <p><b>Structuration de l'UE :</b>          La matière Régression linéaire permet d'aborder le cas des données indépendantes avec réponse continue, la matière Modèle linéaire généralisé et régression sur variables catégorielles celui des données indépendantes avec réponse discrète et la matière Série temporelle 1 le cas où la variable d'intérêt est continue et présente une corrélation temporelle.</p> <p><b>Compétences ou acquis d'apprentissage à l'issue de l'UE :</b>          Connaître le cadre d'application et les hypothèses des différents modèles linéaires de base. Être capable d'ajuster et de valider ces modèles à partir d'un jeu de données, de produire et d'interpréter les résultats. Comprendre les fondements mathématiques de ces modèles et les méthodes d'inférence ou de test statistique utilisées.</p> <p><b>Les prérequis de l'UE :</b>          Les notions de probabilités, de statistique descriptive et de statistique inférentielle abordées en première année.</p>	

UE 2-01 – Matière 2ASTA01 – Semestre 1

## REGRESSION LINEAIRE

### *Linear regression*

<i>Enseignant</i>	: Basile DE LOYNES (Ensaï)
<i>Nombre d'ECTS</i>	: 3
<i>Répartition des enseignements</i>	: 33h dont 15h de cours, 12h de TD et 6h de TP
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: R et Python
<i>Documents pédagogiques</i>	: Polycopiés de cours
<i>Prérequis</i>	: Lois de probabilités usuelles, vecteurs gaussiens, régions de confiance, tests statistiques, algèbre matricielle, projection orthogonale

### Modalités d'évaluation :

Examen sur table

### Acquis d'apprentissage (objectifs) :

- Comprendre la construction d'un modèle de régression linéaire
- Mener une étude complète de régression linéaire
- Effectuer une analyse constructive des résidus
- Sélectionner des variables pertinentes
- Utiliser des packages du logiciel R et la librairie statsmodel sous Python

### Principales notions abordées :

- Rappels : projections orthogonales, vecteurs gaussiens et espérance conditionnelle
- Le modèle de régression linéaire multiple
- Le modèle de régression linéaire multiple sous hypothèse gaussienne
- Validation du modèle, sélection de variables, mesure de diagnostique

### Références bibliographiques :

- AZAIS J.M et BARDET J.M, Le modèle linéaire par l'exemple : régression, analyse de la variance et plans d'expérience illustrés avec R et SAS (2e éd.), Dunod, 2012
- CORNILLON P.-A., MATZNER-LOBER E., Régression avec R, Springer, 2010
- DODGE R., ROUSSON V., Analyse de régression appliquée (2e éd.), Dunod, 2004
- KLEINBAUM D. G. et al., Applied regression analysis and multivariate methods (4th ed.), Cengage Learning, 2008
- TOMASSONE R; et al., La régression (2e éd.), Masson, 1992
- JØRGENSEN B., The Theory of Linear Models, Chapman & Hall, 1993
- GROB J., Linear Regression, Lecture Notes in Statistics, Springer, 2003

UE 2-01 – Matière 2ASTA02 – Semestre 1

## MODELES LINEAIRES GENERALISES

### *Generalized Linear Models*

<i>Enseignant</i>	: Jad BEYHUM
<i>Nombre d'ECTS</i>	: 3
<i>Répartition des enseignements</i>	: 33h dont 21h de cours et 12h de TP
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: R
<i>Documents pédagogiques</i>	: Slides et Compte rendu de cours
<i>Prérequis</i>	: Cet enseignement demande que tous les élèves maîtrisent les cours relatifs aux modèles de régression linéaire et à la statistique inférentielle.

### Modalités d'évaluation :

Examen sur table

### Acquis d'apprentissage (objectifs) :

- Savoir choisir un modèle approprié pour expliquer une variable binaire ou catégorielle selon la nature du problème posé et des données disponibles
- Savoir construire ce modèle, comprendre son fonctionnement et le mettre en œuvre à l'aide d'un logiciel adapté
- Savoir interpréter finement les résultats obtenus

### Principales notions abordées :

- Modèle linéaire généralisé, modèle logistique, modèle probit
- Effets marginaux moyens
- Modèles logistiques et probit multinomiaux

### Références bibliographiques :

- Agresti A., Categorical Data Analysis (3rd ed.), WILEY & Sons, 2012
- Agresti, A., Foundations of Linear and Generalized Linear Models, Wiley & Sons, 2015
- Dobson, A.J., Barnett, A.G. An Introduction to Generalized Linear Models (3rd ed.), CRC Press, 2008
- Hansen, B., Econometrics, Princeton University Press, 2022

UE 2-01 – Matière 2ASTA03 – Semestre 1

## SERIES TEMPORELLES 1

### *Times Series 1*

<i>Enseignant</i>	: Marie DU ROY DE CHAUMARAY (Ensaï)
<i>Nombre d'ECTS</i>	: 3
<i>Répartition des enseignements</i>	: 33h dont 18h de cours, 6h de TD et 9h de TP
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: R
<i>Documents pédagogiques</i>	: Notes de cours, résumés de cours, transparents, anaes
<i>Prérequis</i>	: Probabilités et statistique inférentielle

### Modalités d'évaluation :

Examen final sur papier

### Acquis d'apprentissage (objectifs) :

- Connaître les principaux modèles linéaires utilisés et leurs caractéristiques.
- Savoir estimer les paramètres des modèles et tester leur validité.
- Reconnaître les caractéristiques principales d'une série temporelle en utilisant les outils graphiques usuels.
- Savoir conduire une démarche statistique complète: trouver des modèles adéquats, vérifier leur validité pour les données et faire des prévisions de valeurs futures.

### Principales notions abordées :

1. Tendances, saisonnalité et filtrage linéaire.
2. Processus stationnaires, autocovariance, autocorrélation.
3. Processus ARMA, causalité, inversibilité, innovation, estimation.
4. Méthode de Box-Jenkins, processus non-stationnaires (S)ARIMA, test de racine unitaire.
5. Prévision : meilleur prédicteur linéaire, lissage exponentiel.
6. Contributions exogènes, processus ARMAX et corrélation croisée. Hétéroscédasticité et processus (G)ARCH.

### Références bibliographiques :

- Hamilton J, Time Series, Princeton University Press, 1995.
- Box G., Jenkins G., Reinsel G., Time Series Analysis. Forecasting and Control, Prentice Hall, 1994, 3ème Edition
- Shumway R. and Stoffer D., Time series analysis and its applications, Springer 2006.
- Aragon Y., Séries temporelles avec R, Edp Sciences, 2016

UE 2-02 – Semestre 1

## UE 2-02 : COLLECTE ET APPRENTISSAGE

Correspondant de l'UE	: A déterminer
Nombre d'ECTS	: 6
Nombre d'heures d'enseignement	: 66h

### Finalité de l'UE :

Les données constituent la matière première qu'auront à manipuler les ingénieurs et élèves attachés de l'Insee à la sortie de l'Ensaï, tant dans les entreprises que les administrations. Il est crucial que cette matière première soit collectée dans le respect de règles juridiques et éthiques. Cette collecte doit se faire de façon optimale sur le plan scientifique, compte tenu de contraintes budgétaires. Ces données une fois collectées, les ingénieurs et attachés seront en mesure d'en extraire une information dans l'optique de l'aide à la décision dans les secteurs publics et privés. L'apprentissage supervisé recouvre une famille de techniques statistiques permettant d'offrir cette aide à la décision sous l'angle de la prédiction et/ou l'explication d'une grandeur d'intérêt (variable cible).

### Structuration de l'UE :

L'UE Collecte et apprentissage est structurée en matières. La matière **Droit et éthique des données** donne les éléments de base du respect de règles juridiques et éthiques de la collecte des données. La matière **Théorie des sondages** présente l'aspect scientifique de la collecte des données par des méthodes d'échantillonnage et de redressement. Enfin, la matière **Apprentissage supervisé** fournit aux étudiants un ensemble d'algorithmes de prédiction d'une grandeur d'intérêt ainsi que les outils d'évaluation de la qualité de cette prédiction. Compétences ou acquis d'apprentissage à l'issue de l'UE (de façon synthétique et en se référant aux compétences de l'ingénieur et compétences spécifiques de la formation ingénieur) :

### Compétences ou acquis d'apprentissage à l'issue de l'UE :

- Capacité à identifier les responsabilités éthiques et professionnelles, à prendre en compte les enjeux des relations au travail, de sécurité et de santé au travail et de la diversité
- Capacité à trouver l'information pertinente, à l'évaluer et à l'exploiter : «compétence informationnelle»
- Aptitude à étudier et résoudre des problèmes complexes souvent situés en univers incertain et risqué en mobilisant des connaissances scientifiques à dominante mathématique, statistique, économique et informatique.
- Concevoir et mettre en œuvre des projets de collecte et d'analyse d'informations en utilisant les outils informatiques associés, les techniques de production (bases de données, enquêtes, données non structurées issues d'internet) et les méthodes d'analyse quantitative de l'information, pour proposer des solutions et aider à la décision.
- Concevoir et mettre en œuvre des algorithmes prédictifs de machine learning, s'intégrant dans une architecture informatique de données volumineuses (big data) et en veillant au respect du cadre éthique et juridique d'utilisation des données.

### Les prérequis de l'UE :

Probabilités, Statistique avec R, Régression linéaire, Statistique Exploratoire Multivariée 1A, projection et d'optimisation (lagrangien).

UE 2-02 – Matière 2ASTA04 – Semestre 1

## APPRENTISSAGE SUPERVISE

### *Supervised learning*

<i>Enseignant</i>	: Sébastien DA VEIGA
<i>Nombre d'ECTS</i>	: 2.5
<i>Répartition des enseignements</i>	: 27h dont 18h de cours et 9h de TP
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: R
<i>Documents pédagogiques</i>	: Polycopié contenant le diaporama du cours, sous Moodle les sujets de TP
<i>Prérequis</i>	: Statistique Exploratoire Multivariée 1A, projection et d'optimisation (lagrangien)

### Modalités d'évaluation :

Examen sur table

### Acquis d'apprentissage (objectifs) :

- Distinguer l'apprentissage supervisé de l'apprentissage non-supervisé
- Choisir et mettre en œuvre différentes méthodes d'apprentissage supervisé
- Comparer les performances des méthodes d'apprentissage concurrentes

### Principales notions abordées :

- Apprentissage supervisé – définition et concepts généraux,
- k plus proches voisins (KNN),
- Classifieur bayésien naïf (Naive bayes),
- Analyse discriminante (factorielle et bayésienne)
- Segmentation par arbres (CART, CHAID), Probability Estimation Trees, arbres à cibles multiples (Ctree).
- Comparaison de méthodes (LIFT, ROC, avantages et inconvénients des différentes méthodes présentées).

### Références bibliographiques :

- BREIMAN L., FRIEDMAN J. H., OLSHEN R. A., STONE C. J. , Classification and regression trees. Wadsworth & Brooks/Cole Advanced Books & Software, 1984.
- HASTIE, T., TIBSHIRANI, R., and FRIEDMAN, J. . The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, 2009.
- HUBERTY C.J., Applied discriminant analysis, Wiley, 1994.
- CONFAYS J., NAKACHE J.-P., Statistique explicative appliquée, Technip, 2003.
- McLachlan G., Discriminant Analysis and Statistical Pattern Recognition, 2004.
- PROVOST F., DOMINGO P., Tree Induction for Probability-Based Ranking, Machine Learning, 52, 199–215, 2003, Kluwer Academic Publishers
- TUFFERY Stéphane, Modélisation prédictive et apprentissage statistique avec R, Technip, 2015

UE 2-02 – Matière 2ASTA05 – Semestre 1

## THÉORIE DES SONDAGES

### *Sampling Theory*

<i>Enseignant</i>	: Guillaume CHAUVET
<i>Nombre d'ECTS</i>	: 3
<i>Répartition des enseignements</i>	: 30h dont 18h de cours et 12h de TD
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: SAS/R
<i>Documents pédagogiques</i>	: Diaporama du cours
<i>Prérequis</i>	: Probabilités, Compléments de probabilités, Statistique descriptive et SAS, Statistique avec R

### Modalités d'évaluation :

Examen sur table

### Acquis d'apprentissage (objectifs) :

A l'issue de ce cours, les étudiants doivent connaître et maîtriser les principales méthodes d'échantillonnage utilisées dans le cas d'une population finie, ainsi que les propriétés des estimateurs associés. Une partie du cours est également consacrée à la présentation de méthodes de redressement, où une information externe est utilisée pour modifier les estimateurs afin de diminuer leur variance.

### Principales notions abordées :

Partie 1 : Echantillonnage en population finie

- Notations
- Plan de sondage, estimation de Horvitz-Thompson
- Calcul de précision : estimateur de variance, intervalle de confiance.

Partie 2 : Méthodes d'échantillonnage

- Sondage aléatoire simple
- Sondage aléatoire simple stratifié
- Sondage à probabilités inégales
- Sondage par grappes

Partie 3 : Méthodes d'estimation

- Approche assistée par un modèle
- Estimateur par calage
- Applications : estimateur par la régression, estimateur par le ration, estimateur post-stratifié

Partie 4 : Compléments sur les méthodes d'échantillonnage

- Sondage équilibré
- Coordination d'échantillons

Partie 5 : Compléments sur les méthodes d'estimation

- Estimation d'une fonction de totaux
- Traitement de la non-réponse

## Références bibliographiques :

- ARDILLY P., *Les Techniques de Sondage (nouv. éd.)*, Paris, Technip, 2006
- COCHRAN W.G., *Sampling Techniques*, 3<sup>e</sup> édition, New York, Wiley, 1977
- SÄRNDAL, C-E., SWENSSON, B, WRETMAN, J., *Model Assisted Survey Sampling*, Springer-Verlag, 2003
- TILLÉ Y., *Echantillonnage et estimation en populations finies*, Dunod, 2019

UE 2-02 – Matière 2AHUM02 – Semestre 1

## DROIT ET ETHIQUE DES DONNEES / SECURITE DES DONNEES

*Ethics and Law in Data Analytics / Data Security*

*Enseignant* : Stéphane ASTIER / Kamel GADOUCHE

*Nombre d'ECTS* : 0.5

*Répartition des enseignements* : 9h de cours

*Langue d'enseignement* : Français

### Modalités d'évaluation :

QCM en fin de séminaire

### Acquis d'apprentissage (objectifs) :

La première partie a pour objectif d'apporter aux étudiants une « culture générale » juridique leur permettant, lorsqu'ils seront en poste, d'échanger plus facilement avec leurs services juridiques sur les aspects spécifiques de la réglementation « informatique et libertés ».

La seconde partie a pour objectif d'apporter aux étudiants une « culture sur le management de la sécurité » incluant les enjeux de la sécurité des données, les responsabilités et les modalités d'application avec des exemples concrets.

### Principales notions abordées :

En s'appuyant sur l'actualité (géopolitique, technologique, économique), la première partie du séminaire mettra en perspective une série d'enjeux juridiques et éthiques liés à la digitalisation de l'ensemble de la société tant dans l'environnement professionnel que personnel.

Des exemples concrets de procédures en cours devant les instances françaises et internationales ainsi que des faits d'actualité viendront illustrer cette présentation et alimenter une réflexion globale sur l'impact de l'évolution technologique sur notre civilisation.

La seconde partie s'appuiera sur les standards et référentiels en cours (RGPD, ISO). Il s'agira de présenter les principes de la méthode d'établissement d'une politique de sécurité pour en comprendre les enjeux : objectifs, gouvernance, risques, mesures, plan d'action, indicateurs, évaluation et amélioration continue. Cette partie présentera aussi des recommandations, des exemples d'incident de sécurité et les principales mesures de sécurité généralement mises en œuvre : anonymisation, chiffrement, sauvegardes, permissions d'accès,...

UE 2-03 – Semestre 1

## UE 2-03 : INFORMATIQUE POUR LA DATA SCIENCE

<i>Correspondant de l'UE</i>	: À définir
<i>Nombre d'ECTS</i>	: 6 ( 7 pour les AST)
<i>Nombre d'heures d'enseignement</i>	: 51h (72h pour les AST)

### Finalité de l'UE :

À la fin de cette UE, les élèves seront capables de concevoir et de réaliser en groupe une application informatique complexe pour répondre aux besoins d'une maîtrise d'ouvrage. Pour ce faire ils devront être capables de répondre à un besoin précis, d'adapter leur communication au niveau attendu par leur interlocuteur, de valoriser leur travail et d'apprendre de nouveaux outils en autonomie et mobiliser des compétences de génie logiciel et de programmation en Python.

### Structuration de l'UE :

L'UE d'informatique pour la data science est constituée de deux modules, Compléments d'informatique, Remise à niveau et Projet informatique. Les élèves directement admis en seconde année bénéficient dans le cadre de l'UE d'une remise à niveau en informatique leur permettant de suivre les deux modules de l'UE.

Le premier module doit apporter les compétences supplémentaires nécessaires pour mener à bien le module Projet informatique, mais également donner les compétences en informatique que tout data scientist doit avoir.

Le module de Projet informatique place les élèves face à une problématique courante dans le monde professionnel : réaliser un projet dans un temps limité en groupe imposé. Ce projet sera l'occasion pour eux de mobiliser de nouvelles compétences en génie logiciel et sécurité informatique acquises dans le module de Compléments d'informatique. Il leur permettra également de développer leurs compétences communications et de travail en groupe (animation, organisation).

Le module de Remise à niveau informatique concerne les élèves admis à l'Ensaï sans avoir fait la première année dans l'établissement. Il doit leur fournir un socle minimal de compétences informatiques pour suivre les enseignements de l'UE. Ils acquerront également des compétences dans la sphère informatique pour mener à bien leur scolarité.

### Compétences ou acquis d'apprentissage à l'issue de l'UE :

- Réaliser une modélisation d'une application complexe indépendante du langage de programmation utilisé
- Développer une application Python complexe utilisant plusieurs services externes (service web, base de données)
- Comprendre des principes de base d'internet et du web et de sécurité informatique
- Synthétiser et mettre en avant le travail produit au travers de livrables écrits ou de présentations orales

### Les prérequis de l'UE :

Programmation orienté objets niveau initiation, python et SQL niveau intermédiaire

UE 2-03 – Matière 2AINF01 – Semestre 1

## COMPLEMENT D'INFORMATIQUE

*Computer Science – Complementary Course*

<i>Enseignant</i>	: À définir
<i>Nombre d'ECTS</i>	: 2
<i>Répartition des enseignements</i>	: 18h dont 6h de cours, 12h de TD
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: python3, Pycharm, PuTTY, git, phpadmin
<i>Documents pédagogiques</i>	: sous Moodle, pas de photocopié
<i>Prérequis</i>	: Programmation orienté objets niveau initiation, python et SQL niveau intermédiaire

### Modalités d'évaluation :

Des QCM seront disponibles sur moodle pour permettre aux élèves d'autoévaluer leurs acquis des séquences en autonomie. Ces QCM ne compteront pas dans la note finale

Note du code produit lors du projet informatique, examen écrit final

### Acquis d'apprentissage (objectifs) :

Ce cours vient compléter les connaissances en informatique des étudiants afin de leur en donner une vision plus globale. Les différents thèmes abordés permettront aux étudiants de mieux comprendre l'informatique d'aujourd'hui pour leur vie professionnelle tout en les armant pour leur projet informatique qui sera mené en parallèle.

- Appliquer des concepts de base d'ingénierie logicielle pour réaliser une modélisation fonctionnelle et indépendante du langage de programmation d'une application
- Expliquer le fonctionnement d'une application client server
- Expliquer le principe de l'injection de langage
- Déterminer le meilleur format d'échange de données pour un cas donné
- Mettre en place les outils pour mener à bien un projet informatique

### Principales notions abordées :

Ce cours a pour but d'apporter des connaissances supplémentaires en informatique aux élèves pour leur permettre d'évoluer dans le monde de la science de la donnée plus sereinement. Ainsi des notions de programmation orientée objet seront revues et approfondies et des nouvelles notions seront abordées.

La communication client-serveur, incontournable aujourd'hui, sera donc vu dans le cadre du cours. Cela sera l'occasion de démystifier certains termes informatiques, ainsi que de présenter le protocole http, et des notions de sécurité informatique comme les injections de langage avec l'exemple de l'injection SQL.

Une analyse des différents formats d'échange de données sera faite dans ce cours avec l'opposition donnée tabulaire (csv, base de données relationnelles) et non tabulaire (json, XML, HTML) ainsi que les moyens de traiter chacun de ces formats de données.

Des éléments de culture informatique seront également apportés aux élèves pour développer leurs connaissances générales dans le domaine.

## Références bibliographiques :

- Debrauwer L., Heyde F., & Van der Heyde, F. (2020). UML 2.5. Editions ENI.
- Alhir S. S., Oram, A., & O'Reilly & Associates. (1998). UML in a Nutshell. O'Reilly.
- Roques P. (2015). UML 2.5. Eyrolles.
- Gamma E. (1999). Design patterns. Vuibert.
- Larman C. (2005). Applying UML and Patterns. Prentice Hall.
- Larman C. (2005). UML 2 et les design patterns. Télé-université.
- Lutz M. (2014). Python Pocket Reference. O'Reilly.
- Lutz M., & Maniez, D. (2019). Python précis et concis. Dunod.
- Martelli A., Ravenscroft A., & Ascher D. (2005). Python Cookbook (Second éd.). O'Reilly.
- Slatkin B. (2019). Effective Python. Addison Wesley.
- Richard E. Silvermam (2013). Git Pocket Guide. O'Reilly
- Soutou C. (2007). UML 2 pour les bases de données. Eyrolles.
- Refactoring and Design Patterns. (s. d.). Refactor guru. Consulté 23 juin 2020, à l'adresse <https://refactoring.guru/design-patterns>

UE 2-03 – Matière 2AINFO02 – Semestre 1

## PROJET D'INFORMATIQUE

### *Computer Science project*

<i>Enseignant</i>	: À définir
<i>Nombre d'ECTS</i>	: 4
<i>Répartition des enseignements</i>	: 33h de projet
<i>Langue d'enseignement</i>	: Français (cours) et Anglais (ressources complémentaires fournies)
<i>Logiciels</i>	: python3, Pycharm, PuTTY, git, phpadmin, python3, SQL, json, XML, HTML, HTTP, SSH, webservice REST, UML
<i>Documents pédagogiques</i>	: sous Moodle, pas de polycopié
<i>Pré-requis</i>	: Programmation orienté objets niveau initiation, python et SQL niveau intermédiaire

### Modalités d'évaluation :

Livrable du projet et soutenance

### Acquis d'apprentissage (objectifs) :

- Concevoir une application multi-tiers via le formalisme UML
- Développer en python une application multi-tiers
- Communiquer avec une base de données via python
- Communiquer avec un serveur distant via python
- Planifier et mettre en place les outils pour un projet informatique

### Principales notions abordées :

Ce projet demande aux élèves par groupe de 4-5 mettre en pratique sur un cas réel les connaissances acquises en informatique. Ce projet a pour but de concevoir et développer une application multi-tiers en python. Ainsi les élèves vont devoir :

- Réaliser l'étude fonctionnelle de leur application pour en déterminer acteurs et les fonctionnalités principales ;
- Prioriser ces fonctionnalités ;
- Réaliser un rapport avec les conclusions de leur analyse ;
- Produire le code python de leur application tout en s'assurant d'une qualité minimale en test et documentation ;
- Réaliser un rapport final de leur projet et défendre ce projet lors d'une soutenance.

Les sujets seront proposés et encadrés par des professionnels de l'informatique qui pourront venir du secteur privé ou public. Leur rôle sera d'orienter les élèves sur les bonnes pratiques organisationnelles et de développement applicatif.

Pour simuler une période de « crunch » (période intense de travail dans le milieu informatique), les élèves vivront 3 jours d'immersion pour travailler et finaliser leur projet. Cette période intense mettra à l'épreuve leurs compétences de gestion du temps et de travail en groupe dans contexte proche de la vie professionnel quand un projet arrive à terme.

### Références bibliographiques :

- Debrauwer L., Heyde F., & Van der Heyde, F. (2020). UML 2.5. Editions ENI.
- Alhir S. S., Oram, A., & O'Reilly & Associates. (1998). UML in a Nutshell. O'Reilly.

- Roques P. (2015). UML 2.5. Eyrolles.
- Gamma E. (1999). Design patterns. Vuibert.
- Larman C. (2005). Applying UML and Patterns. Prentice Hall.
- Larman C. (2005). UML 2 et les design patterns. Télé-université.
- Lutz M. (2014). Python Pocket Reference. O'Reilly.
- Lutz M., & Maniez, D. (2019). Python précis et concis. Dunod.
- Martelli A., Ravenscroft A., & Ascher D. (2005). Python Cookbook (Second éd.). O'Reilly.
- Slatkin B. (2019). Effective Python. Addison Wesley.
- Richard E. Silvermam (2013). Git Pocket Guide. O'Reilly
- Soutou C. (2007). UML 2 pour les bases de données. Eyrolles.
- Chrisment C., Pinel-Sauvagnat K., Teste, O. & K. P. S. (2008). Bases de données relationnelles. Hermès Science.
- Hainaut J. L. (2018). Bases de données. Dunod.
- Jacobson I. (1992). Object Oriented Software Engineering : A Use Case Driven Approach (1re éd.). Addison-Wesley Professional.
- Jacobson I., Booch G., & Rumbaugh J. (2000). Le processus unifié de développement logiciel. Eyrolles.
- Refactoring and Design Patterns. (s. d.). Refactor guru. Consulté 23 juin 2020, à l'adresse <https://refactoring.guru/design-patterns>

UE 2-03 – Matière 2AINFO03 – Semestre 1

## REMISE A NIVEAU EN INFORMATIQUE

### *Computing Science refresh Course*

<i>Enseignant</i>	: À définir
<i>Nombre d'ECTS</i>	: 1
<i>Répartition des enseignements</i>	: 21h de cours
<i>Langue d'enseignement</i>	: Français (cours) et Anglais (ressources complémentaires fournies)
<i>Logiciels</i>	: python3, spyder, PyCharm, SQL, Phpadmin
<i>Documents pédagogiques</i>	: sous Moodle, pas de polycopié
<i>Prérequis</i>	: Aucun

### Modalités d'évaluation :

Examen sur table

### Acquis d'apprentissage (objectifs) :

- Produire des requêtes à une base de données en SQL
- Concevoir une base de données relationnelle en appliquant les 3 premières formes normales
- Réaliser un programme simple en python
- Concevoir en UML une application simple

### Principales notions abordées :

Ce cours va présenter les rudiments en informatique nécessaires à la scolarité des élèves qui n'auraient pas effectué la première année à l'Ensaï. Ainsi il traitera des bases de données relationnelle (conception et requêtes), de la programmation avec python (impérative et objet) et de la conception d'application avec UML. Le cours mêlera présentation type cours magistral et travaux pratique. Un accompagnement spécifique des élèves sera mis en place. Ainsi en fonction de leur niveau, les élèves pourront se voir dispenser de certaines parties du cours ou travailler en autonomie sur des notions plus poussées pendant que le cours suit son rythme normal.

## Références bibliographiques :

- Cormen T. H. (2002). Introduction à l'algorithmique. Dunod.
- Cormen T. H., Leiserson, C. E., Rivest, R. L., Stein, C., & Ronald L. Rivest. (2010). Algorithmique. Dunod.
- Baynat B, Chretienne P, Hanen C, Kedad-Sihoum S, Munier-Kordon A, Picouveau C. (2012). Exercices et problèmes d'algorithmique. Dunod
- Lutz M. (2014). Python Pocket Reference. O'Reilly.
- Lutz M., & Maniez, D. (2019). Python précis et concis. Dunod.
- Martelli A., Ravenscroft A., & Ascher D. (2005). Python Cookbook (Second éd.). O'Reilly.
- Slatkin B. (2019). Effective Python. Addison Wesley.
- Hainaut J-L. (2015), Bases de données : concepts, utilisation et développement (3e édition), Dunod
- Larman C. (2005). UML 2 et les design patterns. Télé-université.
- Soutou C. (2007). UML 2 pour les bases de données. Eyrolles.
- Chrisment C., Pinel-Sauvagnat K., Teste, O. & K. P. S. (2008). Bases de données relationnelles. Hermès Science.
- Hainaut J. L. (2018). Bases de données. Dunod.
- Goff, V. L. (2020, 23 juin). Apprenez à programmer en Python. OpenClassrooms.  
<https://openclassrooms.com/fr/courses/235344-apprenez-a-programmer-en-python>

UE 2-04 – Semestre 1

## UE 2-04 : MODELISATION ECONOMIQUE

<i>Correspondant de l'UE</i>	: Samuel DANTHINE
<i>Nombre d'ECTS</i>	: 5
<i>Nombre d'heures d'enseignement</i>	: 54h

### Finalité de l'UE :

Cette UE vise à renforcer les compétences en modélisation microéconomique et à mettre en pratique la modélisation statistique sur des données économiques. Les données économiques sont le fruit de comportements humains, non aléatoires. L'utilisation de ces données et la modélisation afférente doit tenir compte de cette caractéristique, la question étant de comprendre les effets de ce qui ne peut pas être mesuré (le terme d'erreurs des modèles) i.e. l'endogénéité ou la corrélation entre le terme d'erreurs et les variables explicatives. Des exemples de nombreux domaines sont présentés : économie de la santé, de l'éducation, de la concurrence, fonctionnement du marché de l'assurance... Les méthodes sont directement mobilisables pour des métiers de chargés d'études statistiques ou économiques. Elles permettent dans tous les cas d'avoir du recul sur l'usage des méthodes statistiques, pour l'ensemble des métiers de la statistique.

### Structuration de l'UE :

Le cours de microéconomie est un cours de théorie économique, pour lequel chaque concept est illustré par des applications empiriques. Il vise à présenter et utiliser des modèles économiques utiles pour discuter l'utilisation de données économiques. Le cours d'économétrie traite du lien entre modélisation statistique et théorie économique, et détaille l'utilisation de méthodes statistiques adaptées à l'utilisation de données économiques.

### Compétences ou acquis d'apprentissage à l'issue de l'UE :

Les principales compétences sont les suivantes :

- L'aptitude à mobiliser les ressources d'un (ou de plusieurs) champ scientifique et technique spécifique, l'UE étant à l'interaction de plusieurs champs scientifiques ;
- La capacité à prendre en compte les enjeux et les besoins de la société, l'UE traitant de problèmes sociétaux contemporains ;
- Aptitude à étudier et résoudre des problèmes complexes souvent situés en univers incertain et risqué en mobilisant des connaissances scientifiques à dominante mathématique, statistique, économique et informatique.
- Aptitude à mener des études économiques en valorisant les ressources informationnelles et en menant une veille scientifique sur les méthodes scientifiques.

### Les prérequis de l'UE :

Le cours de microéconomie s'appuie sur les cours de modalisation économique de 1<sup>ère</sup> année, le cours d'économétrie s'appuie sur l'UE de modélisation statistique.

UE 2-04 – Matière 2AECO01 – Semestre 1

## MODELISATION MICROECONOMIQUE

### *Microeconomic Theory*

<i>Enseignant</i>	: Samuel Danthine
<i>Nombre d'ECTS</i>	: 2
<i>Répartition des enseignements</i>	: 21h de cours
<i>Langue d'enseignement</i>	: English or French (to be determined)
<i>Documents pédagogiques</i>	: sous Moodle, pas de photocopié
<i>Prérequis</i>	: modélisation microéconomique 1

### Modalités d'évaluation :

A few quizzes and a final exam

### Acquis d'apprentissage (objectifs) :

- Understand and articulate main characteristics in microeconomic decisions
- Become familiar with canonical economic modeling approaches
- Formulate and derive models for analysis of specific empirical applications

### Principales notions abordées :

In this course, we will learn how to formulate models for microeconomic analysis with motivation for empirical applications. We will focus on models that allow us to understand how economic decisions are shaped by tradeoffs of underlying economic incentives, and how such incentives can be estimated by those models. Over the course, we will learn how we can extend the theoretical models learned in Modélisation microéconomique 1 to analyze more realistic decision environment with empirical data.

Our plan is to examine the following topics (depending on time):

- 1) Risk, uncertainty & learning
- 2) Strategic players
- 3) Asymmetric information
  - a) Adverse selection
  - b) Moral hazard
  - c) Model of insurance market
  - d) Auctions

### Références bibliographiques :

- Hal Varian (2010). Introduction à la microéconomie, 6e édition, De Boeck.
- To be determined

UE 2-04 – Matière 2AECO02 – Semestre 1

## ECONOMETRIE APPLIQUEE

*Applied econometrics*

<i>Enseignant</i>	: Marion Goussé
<i>Nombre d'ECTS</i>	: 3
<i>Répartition des enseignements</i>	: 33h dont 21h de cours, 9h de TD et 3h de TP
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: Stata
<i>Documents pédagogiques</i>	: sous Moodle, pas de polycopié
<i>Prérequis</i>	: Régression linéaire, micro-économie

### Modalités d'évaluation :

Les connaissances associées au cours seront évaluées à travers des QCM réguliers et compteront pour 1/3 de la note finale. La mobilisation des connaissances seront évaluées à travers la reproduction d'un article d'économétrie et sa présentation par groupe d'étudiants lors du dernier TP.

### Acquis d'apprentissage (objectifs) :

- Mobiliser les concepts de la théorie micro-économique et les techniques statistiques appropriées pour la construction d'indicateurs ou la conduite d'une étude économique ;
- Interpréter les résultats d'une étude économique, en distinguant approche descriptive et causale ;
- Analyser les sources et les effets de l'endogénéité des variables explicatives, et mettre en œuvre une analyse par variable instrumentale ;
- Identifier les biais et gains potentiels liés à l'usage de variables particulières: évaluation des politiques publiques, données d'enquête, données de panel, données temporelles...;
- Lire et discuter un texte économétrique avec un regard critique sur la qualité des données, le choix des méthodes et l'interprétation des résultats.

### Principales notions abordées :

Ce cours a pour but de présenter d'une part la démarche économétrique (spécification, identification, interprétation, lien avec la théorie économique), et d'autre part les modèles de base et les techniques d'inférence utilisés par les économètres. Les exemples seront principalement issus de la micro-économie.

### Références bibliographiques :

- ANGRIST, J. D. and J.-S. PISCHKE, Mostly Harmless Econometrics, an empiricist companion, Princeton University Press, 2008.
- CREPON, B. et N. Jacquement, Econométrie: méthode et applications. De Boeck, 2010
- WOOLDRIDGE, J. M., Introduction à l'économétrie – Une approche moderne, De Boeck (2nd rev. ed), 2018.

UE 2-05 – Semestre 1

## UE 2-05 : HUMANITES ET STAGE OPERATEUR

Correspondant de l'UE	: BURMEISTER Emily
Nombre d'ECTS	: 4
Nombre d'heures d'enseignement	: 54h

### Finalité de l'UE :

À la fin de cette UE, notamment grâce aux enseignements des langues, les élèves seront capables de mettre en œuvre les compétences linguistiques et culturelles qui facilitent la suivie des cours scientifiques dispensés en anglais ou d'autres langues, le travail dans un environnement professionnel international et la compréhension des normes culturelles dans les pays étrangers. À travers les cours d'ouverture, les élèves acquerront les connaissances dans des disciplines autres que la statistique, l'économie et l'informatique. Cette UE vise également le développement des compétences transversales (*soft skills*) qui aideront les élèves à réussir les projets académiques de leur formation, à intégrer le marché du travail et à devenir des citoyens éclairés. Grâce au stage de 1<sup>ère</sup> année, les élèves élargiront leurs connaissances du monde professionnel et développeront leur capacité d'analyse à travers la rédaction du rapport.

### Structuration de l'UE :

L'UE Humanités en 2<sup>e</sup> année se compose de 2 matières obligatoires (l'anglais et les cours libres optionnels).

#### Anglais et langues optionnelles

L'étude de l'anglais est obligatoire au premier semestre en 2<sup>ème</sup> année. Une deuxième langue est optionnelle mais se suit obligatoirement pendant 2 semestres. Les élèves sont répartis en groupes de niveau en anglais et en langue optionnelles. En anglais, les élèves travaillent toutes les compétences linguistiques pour atteindre le niveau B2 du CECR<sup>1</sup> et progresser vers un niveau C1. Pour les langues optionnelles (l'allemand, le chinois, l'espagnol, l'italien, le japonais et le russe), les élèves travaillent toutes les compétences linguistiques afin de progresser selon leur niveau de départ. Lors des cours de langues, les élèves développent également les connaissances interculturelles et les compétences transversales (*soft skills*), notamment dans les domaines de l'ouverture au monde, à la vie sociale et civique.

Pour tout complément d'information, chaque élève peut consulter le *Programme des enseignements : Langues*, disponible sur le site de l'école.

#### Les cours d'ouverture

Cette option vise à développer chez les élèves une culture générale dans un domaine scientifique ou des sciences sociales et humaines non couvert dans les autres l'UE à l'ENSAI (communication interculturelle, développement durable, géopolitiques, histoire de l'art, l'initiation à l'entrepreneuriat, philosophie, psychologie, physique...), à effectuer des travaux d'initiation dans un domaine artistique (cinéma, dessin, musique, peinture, théâtre...) ou à valoriser l'engagement citoyen des étudiants (participation à des activités associatives, à des projets avec des intervenants non statisticiens...). Les élèves développent ainsi les connaissances et des compétences transversales dans la matière suivie à travers les dispositifs pédagogiques proposés (projet individuel ou en groupe, réalisation artistique...).

Les cours d'ouverture proposés en 2022/2023, leur contenu et leurs modalités de contrôle des connaissances sont décrits dans le fascicule « Programme des enseignements : Cours libres optionnels »

<sup>1</sup> Le niveau B2 au CECR correspond à un score minimal de 785 points au TOEIC. Un niveau B2 du CECRL est obligatoire pour l'obtention du diplôme d'ingénieur.

### Stage

Le stage de première année correspond à un stage d'observation de 1 à 3 mois. Les élèves vivent une expérience professionnelle en autonomie, un stage de découverte de la statistique publique pour les attachés stagiaires et un stage opérateur pour les élèves ingénieurs. Le stage est validé par un rapport écrit de quelques pages qui détaille l'expérience vécue.

Une structuration du rapport sous forme d'une proposition de plan et de questions est proposée aux élèves ainsi que la liste des critères d'évaluation pour aider à développer chez les élèves une capacité à exprimer les apports de l'expérience.

### **Compétences ou acquis d'apprentissage à l'issue de l'UE :**

- Maîtriser une ou plusieurs langues étrangères
- Contextualiser et prendre en compte les enjeux et les besoins de la société
- Se connaître, s'auto-évaluer, gérer ses compétences, opérer ses choix professionnels
- Approfondir les connaissances du monde professionnel
- S'intégrer et évoluer dans un contexte professionnel y compris à l'international
- Savoir identifier les informations pertinentes, à les évaluer et à les exploiter

### **Les prérequis de l'UE :**

Pour les langues : test de positionnement

UE 2-05 – Matière 2AHUM01 – Semestre 1

## ANGLAIS

### English

<i>Enseignant</i>	: Emily BURMEISTER (ENSAI)
<i>Nombre d'ECTS</i>	: 2
<i>Répartition des enseignements</i>	: 30h de cours TD au semestre 1
<i>Langue d'enseignement</i>	: Anglais
<i>Logiciels</i>	: VLC, Audacity, Moodle, Teams
<i>Documents pédagogiques</i>	: Fournis sur Moodle par les intervenants
<i>Pré-requis</i>	: Avoir passé le test de niveau pour les AST. Pour suivre les cours thématiques, avoir le niveau B2 CECR

### Modalités d'évaluation :

La note finale prend en compte un TOEIC blanc à la fin du semestre ou le score TOEIC/Linguaskill obtenu lors d'un TOEIC/Linguaskill définitif (30%), la note moyenne des contrôles continus (40%) et la participation (30%). La note de contrôle continu est composé d'au moins 4 contrôles continus (interrogations et travaux divers) définis par les enseignants.

### Acquis d'apprentissage (objectifs) :

Les élèves travailleront toutes les compétences linguistiques pour atteindre le niveau B2 du CECR et progresser vers un niveau C1. Les élèves développeront également les *soft skills*, notamment dans les domaines de l'ouverture au monde, à la vie sociale et civique.

#### À l'issue du cours, l'étudiant aura développé les compétences langagières afin de :

- **objectif 1 : Comprendre** le contenu essentiel de sujets concrets ou abstraits dans un texte ou dialogue complexe ;
- **objectif 2 : Communiquer** avec un degré de spontanéité et d'aisance tel qu'une conversation avec un locuteur natif ne comportant de tension ni pour l'un ni pour l'autre ;
- **objectif 3 : Posséder une gamme assez étendue de langue** pour pouvoir faire des descriptions claires, exprimer son point de vue et développer une argumentation sans chercher ses mots de manière évidente ;
- **objectif 4 :** Montrer un degré assez élevé de **contrôle grammatical**. Ne pas faire de fautes conduisant à des malentendus et pouvoir le plus souvent les corriger lui/elle-même ;
- **objectif 5 : S'exprimer** de façon claire et détaillée sur une grande gamme de sujets, émettre un avis sur un sujet d'actualité et exposer les avantages et les inconvénients de différentes possibilités ;
- **objectif 6 :** Parler relativement longtemps avec un **débit assez régulier** ; bien qu'il /elle puisse hésiter en cherchant structures ou expressions, l'on remarque peu de longues pauses ;
- **objectif 7 : Prendre l'initiative de la parole** quand il convient et pouvoir clore une conversation quand il le faut, encore qu'éventuellement sans élégance ;
- **objectif 8 : Faciliter la poursuite d'une discussion** sur un terrain familier, en confirmant sa compréhension, en sollicitant les autres, etc. ;
- **objectif 9 :** Utiliser un nombre limité d'articulateurs pour **lier ses phrases** en un discours clair et cohérent bien qu'il puisse y avoir quelques "sauts" dans une longue intervention.

### Principales notions abordées :

Pour les élèves ingénieurs qui n'ont pas le niveau B2, les cours se concentrent en partie sur la préparation au TOEIC et au développement des compétences pour atteindre le niveau B2. Tous les élèves du niveau B2 ou plus suivront deux cours thématiques. Les thèmes varient tous les ans, voici quelques exemples des thèmes proposés par le passé : *Debating, British Humour, American Politics, English through Popular Music, Travel, Scientific English*.

### Références bibliographiques :

- Azar, B. S., & Hagen, S. A. (2009). *Understanding and using English grammar*. White Plains, NY: Pearson Longman.
- Council of Europe, *Common European Framework of Reference for Languages* <https://www.coe.int/en/web/common-european-framework-reference-languages/home>
- Merriam-Webster's Online Dictionary Springfield, Mass. : Merriam-Webster, Inc., 2020. <https://www.merriam-webster.com/>
- Oxford Advanced Learners Dictionary, 10<sup>th</sup> Edition, Oxford: Oxford University Press, 2020. <https://www.oxfordlearnersdictionaries.com/>
- Oxford Preparation Course for the TOEIC, Oxford: Oxford University Press.
- Rogers, Bruce, *Complete Guide to the TOEIC Test* (3<sup>rd</sup> ed.), Boston: Thomson, 2006.
- Divers ressources fournis selon les sujets traités par les enseignants qui fournissent un syllabus détaillé avec les contenus, objectifs et ressources spécifiques à leurs cours.

Pour tout complément d'information, chaque élève peut consulter le *Programme des enseignements : Langues*, disponible sur le site de l'école.

# Enseignements du second semestre : Tronc commun

UE 2-06 – Semestre 2

## UE 2-06 : PROJET STATISTIQUE

<i>Correspondant de l'UE</i>	: David AUDENAERT
<i>Nombre d'ECTS</i>	: 5
<i>Nombre d'heures d'enseignement</i>	: 8h dont 4h de suivi avec tuteur et 4h de coaching

### Finalité de l'UE :

Les projets de statistique sont encadrés par des praticiens de la statistique « extérieurs à l'école » et ont pour objectifs l'application et l'approfondissement des connaissances statistiques acquises dans le cadre des enseignements, le travail en mode projet et en autonomie, la communication professionnelle et les échanges scientifiques.

Les élèves, répartis par groupes de trois, traitent le sujet qu'ils ont choisi sous la conduite scientifique et technique du praticien qui l'a proposé. Ils doivent circonscrire le problème soumis et le traiter en utilisant les concepts, méthodes et outils appropriés.

### Structuration de l'UE :

Le projet a pour objectif de répondre à une problématique concrète en exploitant des données réelles fournies par le professionnel qui sera leur tuteur. Les élèves doivent

Utiliser les connaissances acquises durant leurs deux premières années et utiliser au moins une méthode de modélisation. Les groupes travaillent pour répondre à la problématique proposée et produisent un rapport d'étude accompagné d'un poster de présentation et d'une présentation orale.

Pour les compétences relevant de la communication et du coaching de projet, les élèves bénéficient de quatre séances avec des coachs en communication.

### Compétences ou acquis d'apprentissage à l'issue de l'UE :

Le projet statistique de deuxième année constitue une étape importante vers la professionnalisation :

- Travailler dans une relation de service avec un « client » extérieur pour réaliser une étude statistique complète,
- communiquer les résultats sous différentes formes : rapport, poster, soutenance,
- consolider les compétences en termes de gestion de projet et de travail en équipe.

### Les prérequis de l'UE :

Sans objet

UE 2-07 – Semestre 2

## UE 2-07 : COMPLEMENTS DE MODELISATION STATISTIQUE

<i>Correspondant de l'UE</i>	: Lionel TRUQUET
<i>Nombre d'ECTS</i>	: 6
<i>Nombre d'heures d'enseignement</i>	: 63h

### Finalité de l'UE :

Introduire des méthodes ou modèles plus avancés pour la statistique et qui sont utiles pour traiter des problèmes plus spécifiques ou complémentaires à ceux traités dans l'UE 1 ou l'UE 2 de la deuxième année.

Restituée dans une dimension professionnelle, restituée dans la formation (en lien avec d'autres UE)

### Structuration de l'UE :

Les quatre matières de cette UE sont indépendantes et traitent chacune d'un problème spécifique de modélisation et/ou d'inférence statistique.

### Compétences ou acquis d'apprentissage à l'issue de l'UE :

Maîtriser les techniques de modélisation statistique et les propriétés mathématiques de base des modèles aléatoires avancés. Savoir collecter et interpréter des données. Savoir utiliser les outils informatiques.

### Les prérequis de l'UE :

les notions de probabilité et de statistique inférentielle abordées en première année ainsi que celles abordées dans l'UE1 et l'UE2 de la deuxième année.

UE 2-07 – Matière 2ASTA07 – Semestre 2

## INTRODUCTION AUX MODÈLES DE DURÉE

*Introduction to Survival Analysis*

<i>Enseignant</i>	: Valentin PATILEA (Ensaï) pour le cours ; Sunny WANG (Ensaï) pour les TD/TP
<i>Nombre d'ECTS</i>	: 2
<i>Répartition des enseignements</i>	: 21h dont 12h de cours, 6h de TD et 3h de TP
<i>Langue d'enseignement</i>	: anglais
<i>Logiciels</i>	: R
<i>Documents pédagogiques</i>	: Support de cours, feuilles exercices, codes
<i>Prérequis</i>	: Probabilités, inférence statistique, régression ; programmation en R

### Modalités d'évaluation :

Devoir maison et examen écrit

### Acquis d'apprentissage (objectifs) :

- Connaître les aspects spécifiques de la modélisation statistique des durées
- Connaître les modèles de lois et les procédures d'inférences de base
- Comprendre les hypothèses sous-jacentes aux différentes approches, notamment par régression
- Ajuster des modèles appropriés aux données à l'aide des logiciels spécialisés

### Principales notions abordées :

Mécanismes de censure et de troncation

Modèles paramétriques (Weibull,...) et non paramétriques (Kaplan-Meier,...) et l'inférence associée

Comparaison des survies de plusieurs échantillons

Modèles de régression (Cox Risques Proportionnels, Accelerated Failure Time,...)

### Références bibliographiques :

- J.P. KLEIN, M.L. MOESCHBERGER, *Survival Analysis : Techniques for Censored and Truncated Data (2nd edition), Statistics for Biology and Health*, Springer, 2003.
- J.P. KLEIN, H.C. van HOUWELINGEN, J.G. IBRAHIM, T.H. SCHEIKE: *Handbook of Survival Anlysis*, CRC Press, 2014.
- D.G. KLEINBAUM, J.P. KLEIN: *Survival Anlysis. A Self-Learning Text*, Springer, 2005.

UE 2-07 – Matière 2ASTA08 – Semestre 2

## STATISTIQUE NON PARAMETRIQUE

### *Nonparametric statistics*

<i>Enseignant</i>	: Marian HRISTACHE (Ensay)
<i>Nombre d'ECTS</i>	: 2
<i>Répartition des enseignements</i>	: 21h de cours
<i>Langue d'enseignement</i>	: Anglais
<i>Logiciels</i>	:
<i>Documents pédagogiques</i>	:
<i>Prérequis</i>	: théorie des probabilités, statistique inférentielle, régression linéaire

### Modalités d'évaluation :

Examen écrit et devoir maison

### Acquis d'apprentissage (objectifs) :

La donnée d'un modèle statistique paramétrique fait intervenir une famille de lois caractérisée par un petit nombre de paramètres réels inconnus. Un tel cadre peut être parfaitement approprié lorsque la famille des lois de probabilité retenue apparaît imposée par le phénomène aléatoire que l'on veut décrire.

En pratique cependant, ce choix d'un modèle paramétrique n'est souvent qu'un procédé simplificateur commode, amenant des erreurs d'identification. L'approche alternative consiste à définir un modèle plus large, « non paramétrique », où une loi possible est caractérisée par une fonction (et non plus un élément de  $\mathcal{S}^k$ ). Identifier la loi revient alors à estimer cette fonction, approche qui a connu un vigoureux développement au cours des dernières années, et constituera l'objet de ce cours.

### Principales notions abordées :

1. Modèles non et semi-paramétriques ; principes de base de l'estimation fonctionnelle.
2. Estimation d'une densité par la méthode du noyau.
3. Estimation d'une régression par la méthode du noyau.
4. Estimation non paramétrique du score de propension.
5. Méthode des moments généralisés et instruments optimaux.

### Références bibliographiques :

- BOSQ, D., LECOUTRE, J. P. (1987), *Théorie de l'estimation fonctionnelle*, Economica.
- COMTE, F (2017), *Estimation non-paramétrique*, Spartacus
- HOROWITZ, J.L. (2009), *Semiparametric and Nonparametric Methods in Econometrics*, Springer
- LI, Q., RACINE, J. S. (2007), *Nonparametric Econometrics. Theory and Practice*, Princeton University Press.
- PAGAN, A., ULLAH, A. (1999), *Nonparametric Econometrics*, Cambridge University Press.
- TSYBAKOV, A. B. (2004), *Introduction à l'estimation non-paramétrique*, Springer.
- WASSERMAN, L. (2006), *All of Nonparametric Statistics*, Springer.

UE 2-07 – Matière 2ASTA09 – Semestre 2

## TRAITEMENT DE LA NON-REPONSE DANS LES ENQUETES

### *Missing Survey Data*

<i>Enseignant</i>	: Brigitte GELEIN (Ensaï)
<i>Nombre d'ECTS</i>	: 2
<i>Répartition des enseignements</i>	: 12 de cours, 9h de TP
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: Sans objet
<i>Documents pédagogiques</i>	: Support de cours tapuscrit
<i>Prérequis</i>	: Théorie des sondages, régression linéaire, modèle linéaire généralisé

### Modalités d'évaluation :

Quiz et examen

### Acquis d'apprentissage (objectifs) :

On rencontre des problèmes de données manquantes dans les enquêtes quand certaines des unités refusent de répondre, ou quand il est impossible de les contacter. On parle de non-réponse partielle lorsqu'un individu échantillonné renseigne une partie des questions de l'enquête, et de non-réponse totale lorsqu'aucune réponse n'est observée pour un individu.

La non-réponse a des conséquences en termes de variance des estimateurs (la taille de l'échantillon effectivement observé diminue) et surtout en termes de biais : les estimateurs non ajustés pour la non-réponse peuvent être fortement biaisés si les répondants diffèrent des non-répondants au regard des variables étudiées.

L'objectif de ce cours est de présenter les différents types de non-réponse, les facteurs qui peuvent permettre de limiter ce problème, et des méthodes classiques de traitement de la non-réponse dans les enquêtes.

### Principales notions abordées :

Partie 1 : Introduction

- Rappels sur l'échantillonnage en population finie
- Rappels sur les méthodes de calage
- Les types de non-réponse : non-réponse totale, non-réponse partielle

Partie 2 : Traitement de la non-réponse totale

- Echantillonnage à deux phases
- Redressement par repondération
- Groupes homogènes de réponse
- Applications

Partie 3 : Traitement de la non-réponse partielle

- Le modèle d'imputation
- Méthodes d'imputation simple

Applications

## Références bibliographiques :

- ARDILLY P., *Les Techniques de Sondage (nouv. éd.)*, Paris, Technip, 2006
- HAZIZA D. (2009). Imputation and inference in the presence of missing data, Handbook of Statistics, vol. 29, chap. 10.

UE 2-07 – Matière 2ASTA10 – Semestre 2

## CHAINES DE MARKOV

### Markov Chains

<i>Enseignant</i>	: François COQUET (ENSAI)
<i>Nombre d'ECTS</i>	: 2
<i>Répartition des enseignements</i>	: 21h dont 12h de cours, 9h de TD
<i>Langue d'enseignement</i>	: Français (support et un groupe de TD en Anglais)
<i>Logiciels</i>	: Sans objet
<i>Documents pédagogiques</i>	: notes de cours + fiche TD – cours au tableau ou tablette
<i>Prérequis</i>	: Cours de probabilités de base

### Modalités d'évaluation :

Examen sur table

### Acquis d'apprentissage (objectifs) :

- objectif1 : Reconnaître et prouver qu'un processus stochastique est une chaîne de Markov.
- objectif2 : Analyser la structure statique d'une chaîne de Markov (c'est-à-dire établir le graphe de transition associé, identifier la structure en classes de communication, montrer la récurrence ou la transience des états de la chaîne, calculer les périodicités des classes).
- objectif3 : Décrire le comportement limite d'une chaîne ergodique (par le calcul de la loi stationnaire, éventuellement réversible ; citer et appliquer les théorèmes limites du cours).

### Principales notions abordées :

Définition d'une chaîne de Markov homogène à temps discret sur un espace d'états discret. Illustration par des modèles fondamentaux introduits dans divers contextes de modélisation. Équation de Chapman-Kolmogorov et formules de conditionnement. Classification des états, périodicité, temps d'atteinte, récurrence et transience. Loi stationnaire et théorèmes limites, réversibilité d'une chaîne de Markov.

### Références bibliographiques :

- NORRIS J.R., Markov Chains, Cambridge Series in Statistical and Probabilistic Mathematics, 1997.
- GRIMMETT G.R. & STIRZAKER D.R., Probability and Random Processes, Oxford Sciences Publications, 2001 (3rd ed.).
- FOATA D. & FUCHS A., Processus Stochastiques (2e éd.), Dunod, 2004.
- BALDI P., MAZLIAK L. & PRIOURET P., Martingales et chaînes de Markov : théorie élémentaire et exercices corrigés, Hermann, 2001.

UE 2-07 – Matière 2ASTA15 – Semestre 2

## SÉRIES TEMPORELLES 2

### *Times Series 2*

<i>Enseignant</i>	: Valentin Patilea (ENSAI) pour le cours, Sunny Wang (ENSAI) pour les TD/TP
<i>Nombre d'ECTS</i>	: 2
<i>Répartition des enseignements</i>	: 21h dont 12h de cours, 9h TD/TP
<i>Langue d'enseignement</i>	: Anglais
<i>Logiciels</i>	: R
<i>Documents pédagogiques</i>	: Support de cours, fiches TD avec solutions, codes
<i>Prérequis</i>	: notions de base en probabilités et statistique inférentielle, calcul matriciel de base, notions de séries temporelles univariées, programmation en R.

### Modalités d'évaluation :

Examen sur table, devoir maison

### Acquis d'apprentissage (objectifs) :

- Connaître les propriétés mathématiques des processus linéaires de base qui permettent de modéliser la dynamique jointe de plusieurs séries temporelles.
- Savoir sélectionner et valider un modèle multivarié en fonction de la nature des séries temporelles rencontrées. Identifier les méthodes d'inférence statistique à mettre en oeuvre pour les modèles choisis.
- Etre en mesure d'identifier certaines propriétés propres aux séries temporelles multivariées (phénomènes de causalité ou de cointégration).

### Principales notions abordées :

1. Séries temporelles linéaires multivariées. Modèles VAR.
2. Inférence et tests dans les modèles VAR. Causalité au sens de Granger.
3. Racine unité. Séries temporelles cointégrées. Modèles à correction d'erreurs et tests de cointégration.
4. Régresseurs exogènes, modèles VARX.

### Références bibliographiques :

- LUTKEPOHL H. *New Introduction to Multiple Time Series Analysis*, Springer Verlag, 2006.
- TSAY, R.S. *Multivariate time series analysis: with R and financial applications*. Wiley. 2014.
- GOURIEROUX C., MONFORT A. *Séries temporelles et modèles dynamiques*, Economica (2e édition). 1990.
- HAMILTON J. *Time Series*, Princeton University Press, 1995.
- JOHANSEN S. *Likelihood Based Inference in Cointegrated Vector Auto-Regression Models*, Oxford University Press, 1995.

UE 2-08 – Semestre 2

## UE 2-08 : METHODES COMPUTATIONNELLES

<i>Correspondant de l'UE</i>	: Lionel TRUQUET
<i>Nombre d'ECTS</i>	: 8
<i>Nombre d'heures d'enseignement</i>	: 81h

### Finalité de l'UE :

Aujourd'hui, le développement logiciel s'appuie fortement sur la notion d'objet que ce soit pour les codes développés par les data-scientistes ou pour les bibliothèques qu'ils utilisent. Dans ce contexte, une part de cette UE vise à revoir et approfondir certains aspects de la programmation orientée objet et du développement modulaire. Cette part de l'UE est aussi l'occasion de découvrir un langage typé et compilé et les contraintes et avantages afférents. Enfin, cette part de l'UE présente la syntaxe d'un langage de calcul très utilisé pour le calcul intensif ou l'industrialisation de calcul. Au choix : C++ ou Java.

Au cours de la dernière décennie, nous avons assisté à l'émergence d'applications numériques nécessitant de faire face à de gigantesques quantités de données, générées de plus en plus rapidement. Ces applications (surveillance de réseaux, biologie et médecine, applications financières, réseaux sociaux, etc.) nécessitent un besoin grandissant de techniques capables d'analyser et de traiter ces grandes masses d'information, avec précision et efficacité. La statistique rejoint ici les sciences du numérique, et plus précisément l'informatique répartie, pour proposer de nouvelles approches, relatives au Big Data. Les techniques et les modèles doivent prendre en compte le volume pléthorique de ces données, mais également leur génération rapide en continu (vélocité) ainsi que la diversité de leur format (variété) et la qualité de l'information (véracité), appelés communément les 4V du Big Data.

### Structuration de l'UE :

cf. ci-dessus

### Compétences ou acquis d'apprentissage à l'issue de l'UE :

Concevoir et mettre en œuvre des projets de collecte et d'analyse d'informations en utilisant les outils informatiques associés, les techniques de production (bases de données, enquêtes, données non structurées issues d'internet) et les méthodes d'analyse quantitative de l'information, pour proposer des solutions et aider à la décision.

Dans son champ qui évolue très rapidement, capacité de réaliser une veille scientifique et d'apprendre les nouveaux outils et méthodes en autonomie (il possède les bases théoriques permettant la compréhension des nouveaux outils).

### Les prérequis de l'UE :

cf. fiches matière

UE 2-08 – Matière 2ASTA11 – Semestre 2

## CALCULS BAYESIENS

### *Bayesian Computation*

<i>Enseignant</i>	: Myriam Vimond
<i>Nombre d'ECTS</i>	: 4
<i>Répartition des enseignements</i>	: 39h dont 21h de cours, 18h de TD/TP
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: R
<i>Documents pédagogiques</i>	: notes de cours, slides
<i>Prérequis</i>	: Probabilités, SEM, Chaines de Markov, Statistique inférentielle

### Modalités d'évaluation :

Partiel et examen final sur papier

### Acquis d'apprentissage (objectifs) :

Ce cours est une introduction à la statistique Bayésienne et aux méthodes numériques qui sont classiquement utilisées dans ce cadre. On présentera tout d'abord les fondements de cette approche non fréquentiste ainsi que les techniques de calcul qui en découlent. Dans un second temps, on étudiera la méthode de Monte-Carlo par chaîne de Markov. Il s'agit d'une classe d'algorithmes pour l'échantillonnage de lois de probabilité basés sur la construction d'une chaîne de Markov ergodique. Ces algorithmes sont communément utilisés en statistique bayésienne mais aussi pour l'évaluation d'intégrales multiples. A l'issue de ce cours, les étudiants devront être capables de résoudre des problèmes de statistique inférentielle par une approche bayésienne et de mettre en œuvre des algorithmes classiques de Monte-Carlo par chaîne de Markov.

### Principales notions abordées :

- Distributions a priori et a posteriori. Estimateur de Bayes. Choix de la loi a priori.
- Générateur de nombres aléatoires (Nombre pseudo-aléatoires, inversion de la fonction de répartition, acceptation-rejet).
- Méthode de Monte-Carlo et MCMC (Metropolis-Hastings, Gibbs)

### Références bibliographiques :

- 2MARIN J.-M. et ROBERT C., *Bayesian Core : A Practical Approach to Computational Bayesian Statistics*, Springer, 2007
- 3RUBINSTEIN R., *Simulation and the Monte Carlo Method* (2<sup>nd</sup> ed.), Wiley, 2008
- 4ROBERT C., *Méthodes de Monte Carlo par chaînes de Markov*, Economica, 1996
- 5ROBERT C., *Le choix bayésien : principes et pratique*, Springer, 2006

UE 2-08 – Matière 2AECO03 – Semestre 2

## MICROECONOMETRIE APPLIQUEE

*Applied microeconometrics*

<i>Enseignant</i>	: Yutec Sun (Ensay)
<i>Nombre d'ECTS</i>	: 2
<i>Répartition des enseignements</i>	: 21h de cours
<i>Langue d'enseignement</i>	: Anglais
<i>Logiciels</i>	: Stata, R, Matlab, Julia
<i>Prérequis</i>	: Économétrie, Méthodes d'estimation linéaires, Fondements microéconomiques

### Modalités d'évaluation :

Examen

### Acquis d'apprentissage (objectifs) :

- Understand common source of bias in estimating causal effects in various empirical observations;
- Determine most effective econometric strategy to deal with particular identification challenge;
- Experience with estimating causal effects using econometric strategies adapted to panel data, discrete variables, observed or counterfactual program evaluation.

### Principales notions abordées :

The econometric strategies covered in this course can be categorized as follows:

1. Treatment effects
  - Propensity score matching
  - Difference in differences
  - Heckman selection
  - Regression discontinuity
2. Endogeneity
  - Instrumental variables
  - Control functions
  - Limited information maximum likelihood
3. Unobserved heterogeneity
  - Fixed effects
  - Random effects
4. Discrete choice models
  - Logit
  - Probit
  - Mixed logit
5. Equilibrium model (optional)
  - Rational expectations
  - Generalized method of moments

### Références bibliographiques :

- Angrist and Pischke, Mostly harmless econometrics, 2008
- Wooldridge, Econometric analysis of cross section and panel data, 2002
- Wooldridge, Discrete choice methods with simulation, CUP, 2003

UE 2-08 – Matière 2AECO08 – Semestre 2

## MACROÉCONOMÉTRIE APPLIQUÉE

*Applied Macroeconometrics*

<i>Enseignant</i>	: Stephane AURAY ( Ensai)
<i>Nombre d'ECTS</i>	: 2
<i>Répartition des enseignements</i>	: 21h de cours
<i>Langue d'enseignement</i>	: Français
<i>Prérequis</i>	: Les cours de macroéconomie de première année et de séries temporelles de deuxième année sont supposés assimilés.

### Modalités d'évaluation :

Projets

### Acquis d'apprentissage (objectifs) :

Cet enseignement a pour but d'apporter des compléments d'économétrie théorique utiles aux macro-économistes et de mettre en pratique, dans le cadre de la macroéconomie, certains outils vus dans le cours d'économétrie.

### Principales notions abordées :

Plusieurs thèmes seront abordés à la fois d'un point de vue théorique et d'un point de vue empirique, portant l'économie réelle. Ces thèmes permettent de comprendre les faits stylisés caractérisant les économies modernes (utilisation de différentes méthodes pour mettre en exergue ces faits) et de comprendre comment les modéliser au mieux.

Plan du cours :

- Introduction Part 1 : Evolution de la théorie macroéconomie au cours du temps (notes manuscrites)
- Introduction Part 2 : Les faits et la théorie
- Part 1 : Business Cycle Facts (slides)
- Part 2 : Inflation et chômage (notes manuscrites)
- Part 3 : Critique de Lucas (notes manuscrites)
- Part 4 : Comportement de consommation (notes manuscrites)
- Part 5 : Basic Model and extension with Labor Supply (slides)
- Part 6 : Fiscal Policy issues (slides)
- Part 7 : Search and Matching frictions (slides)

### Références bibliographiques :

- ACEMOGLU, D. (2009), *Introduction to Modern Economic Growth*, Princeton University Press
- LJUNGVIST, L. and SARGENT, T. (2015), *Recursive Macroeconomic Theory*, MIT University Press
- MCCANDLESS, G., (2017), *The ABCs of RBCs*, Harvard University Press
- PISSARIDES, C., (2000), *Equilibrium Unemployment Theory*, MIT University Press

UE 2-08 – Matière 2AINFO04 – Semestre 2

## PROGRAMMATION ORIENTEE OBJET AVEC JAVA

### *Object Oriented Programming with Java*

<i>Enseignant</i>	: IKKO YAMANE
<i>Nombre d'ECTS</i>	: 2
<i>Répartition des enseignements</i>	: 18h dont 6h de cours et 12h de TP
<i>Langue d'enseignement</i>	: Anglais
<i>Logiciels</i>	: Eclipse ou VSCode
<i>Documents pédagogiques</i>	: Transparents de cours
<i>Prérequis</i>	: Programmation orientée objet

### Modalités d'évaluation :

Contrôle continu : un examen écrit

### Acquis d'apprentissage (objectifs) :

Ce cours utilise Java comme langage d'application.

Vous découvrirez un langage typé et compilé. Ce contexte renforce l'importance des concepts de classe abstraite, d'interface et d'héritage inhérents à la programmation orientée objet. Un exemple concret d'interface est la notion d'itérateur, qui permet d'unifier le parcours de collections variées : vecteurs, listes, flux, arbres, ...

Vous découvrirez aussi les problématiques découlant de la livraison d'un code "clef en main".

### Principales notions abordées :

- Langage Java
- Typage, compilation
- Interfaces, classes abstraites, extensions
- Polymorphisme
- Collections et Itérateurs
- Livrable, build, gestion des dépendances

### Références bibliographiques :

- C. Delannoy, Programmer en java, Eyrolles. 2012 (8e éd.)
- B. Eckel, Thinking in Java, Prentice Hall, 2006 (4th ed.), [www.BruceEckel.com](http://www.BruceEckel.com), voir aussi la version française <http://penserensjava.free.fr/>
- R. Chevalier, Java 7, Pearson (Synthex Informatique), 2011-2013.
- Collectif, The Java Tutorial : A Short Course on the Basics - 6th Edition, Addison Wesley, 2014.
- David Flanagan, Java in a nutshell, O'Reilly, 2014 (6th ed.)
- Joshua Bloch, Effective Java, Pearson, 2009.

UE 2-08 – Matière 2AINFO05 – Semestre 2

## C++ POUR LA STATISTIQUE COMPUTATIONNELLE

### *Programming in C++*

<i>Enseignant</i>	: Laurent GEORGET
<i>Nombre d'ECTS</i>	: 2
<i>Répartition des enseignements</i>	: 18h dont 6h de cours et 12h de TP
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: Eclipse pour C/C++
<i>Documents pédagogiques</i>	:
<i>Prérequis</i>	: Bases de programmation dans un langage impératif (Python, Java, etc.)

### Modalités d'évaluation :

La note finale du cours est établie pour moitié avec 2 TP notés et pour moitié avec un examen écrit.

### Acquis d'apprentissage (objectifs) :

Comprendre le fonctionnement de la gestion de la mémoire des ordinateurs et donner aux élèves les principaux concepts du langage C++ moderne pour écrire des programmes plutôt orientés calculs et informatique scientifique.

### Principales notions abordées :

Le langage C++ est apparu en 1983. Il s'agit d'une extension du langage C permettant la programmation orientée objet. Ce langage reste très utilisé et a su évoluer dans le temps puisque sa dernière révision date de 2020. Les points forts du C++ sont :

- Les grandes performances de calcul en temps et en utilisation mémoire ;
- L'utilisation de concepts de développement orienté objet ;
- L'existence d'une grande communauté d'utilisateurs et d'un outillage de développement important (compilateurs, bibliothèques, outils de débogage, etc.).

Ce langage est utilisé dans la recherche ou l'industrie surtout lorsque le temps d'exécution ou la gestion de la mémoire deviennent des contraintes importantes : calcul scientifique, calculs financiers, traitement de grandes quantités de données, imagerie 3D, traitement vidéo, traitement du signal.

Sa capacité à traiter de l'information à bas niveau en fait un langage de prédilection pour la conception d'outils modulaires. Il est ainsi possible d'intégrer des bibliothèques de fonctions développées en C++ au sein de logiciels tels que Matlab, R, SAS ou encore Python pour dépasser les contraintes natives de ces outils et langages.

Ce langage pourra donc être utile pour les élèves devant faire face à des problèmes nécessitant des exécutions efficaces ou dont les solutions doivent s'intégrer dans les outils modulaires mentionnés plus haut. En particulier, les élèves de la filière gestion des risques utiliseront ce langage dans leur cours de méthodes numériques et le développement de modules en cours de R avancé pourra se faire en C++.

Les notions de base :

- Structure et syntaxe des programmes
- Expressions, instructions et types de base
- Manipulation des pointeurs, références et gestion de la mémoire
- Structures de données avancées et algorithmique
  
- Introduction aux classes C++ et concepts orientés objet : encapsulation, héritage, polymorphisme
- Outils pour le C/C++
- Environnement de programmation (notions de compilation)
- Debuggage (gdb, valgrind)
- Bibliothèque standard : structures de données, itérateurs, opérateurs

### Références bibliographiques :

Bjarne STROUSTRUP, *The C++ Programming Language*, Pearson Education, 2013

Herb SUTTER and The Standard C++ Foundation, *Chaîne YouTube de la CppCon*, [en ligne :

<https://www.youtube.com/user/CppCon>]

UE 2-08 – Matière 2AINFO06 – Semestre 2

## OUTILS POUR LE BIG DATA

### Big Data IT Tools

<i>Enseignant</i>	: Arthur Katosky (Ensaie)
<i>Nombre d'ECTS</i>	: 2
<i>Répartition des enseignements</i>	: 24h dont 10.5 de cours, 13.5h de TP
<i>Langue d'enseignement</i>	: Anglais
<i>Logiciels</i>	: Python et Spark
<i>Documents pédagogiques</i>	: Support de présentation, Fiches de TP, Tutoriels en autonomie
<i>Prérequis</i>	: Connaissances élémentaires en Python

### Modalités d'évaluation :

Quizz, examens et compte rendu de TP

### Acquis d'apprentissage (objectifs) :

- S'orienter parmi les technologies étiquetées « big data » les plus courantes
- Identifier les goulots d'étranglements dans l'exécution d'un traitement de données et adapter le traitement pour y remédier
- Choisir et mettre en œuvre une architecture adaptée à un traitement donné, et en particulier choix CPU vs. GPU, local vs. cloud, batch vs. flux, haut niveau vs. bas niveau etc.
- Produire des analyses statistiques simples avec Spark
- Provisionner une infrastructure simple sur AWS

### Principales notions abordées :

Le terme de *big data* est de plus en plus utilisé aussi bien en entreprise que dans les médias généralistes. Malheureusement, souvent, il est utilisé en tant que terme fourre-tout. Ce cours débute avec une déconstruction de la notion de *big data* en présentant les V du *big data* et l'introduction de la notion de traitement de données à haute performance.

Il présente ensuite un panorama des technologie étiquetées *big data* et les architectures informatiques associés en les mettant en parallèle à des solutions classiques :

- Architecture générale du calcul en local (processeur, mémoire vive, stockage) et en distribué (centralisé vs. pee-to-peer ; avantages et inconvénients des systèmes répartis)
- Architectures de stockages (systèmes de fichier vs. base de données, local vs. Distribué)
- Zoom sur le stockage distribué avec HDFS
- Zoom sur le calcul distribué avec Spark et MapReduce
- Présentation du *cloud computing* avec manipulation de *Amazon Web Service*

## Références bibliographiques :

- Carpenter, Jeff, and Eben Hewitt. 2020. *Cassandra: The Definitive Guide; Distributed Data at Web Scale*. 3rd ed. O'Reilly Media, Inc, USA.
- Chambers, Bill, and Matei Zaharia. 2018. *Spark: The Definitive Guide: Big Data Processing Made Simple*. O'Reilly Media, Inc, USA.
- Steen, Maarten van, and Andrew S. Tanenbaum. 2017. *Distributed Systems*. 3.01 ed. CreateSpace Independent Publishing Platform.
- White, Tom. 2015. *Hadoop - The Definitive Guide 4e-*. 4th ed. O'Reilly.
- Stephenson, D. (2018). *Big Data Demystified : How to use big data, data science and AI to make better business decisions and gain competitive advantage* (1re éd.). FT Press.
- Erl, T., Puttini, R., & Mahmood, Z. (2013). *Cloud Computing : Concepts, Technology & Architecture* (The Pearson Service Technology Series from Thomas Erl) (1re éd.). Pearson.

## **Enseignements du second semestre : cours électifs (pré-spécialisation UE 2-09)**

UE 2-09 – Semestre 2

## UE 2-09 : PRE-SPECIALISATION

<i>Correspondant de l'UE</i>	: A déterminer
<i>Nombre d'ECTS</i>	: 8
<i>Nombre d'heures d'enseignement</i>	: 84
<p><b>Finalité de l'UE :</b>          Cette UE vise à proposer une diversité de compétences aux étudiants, dans la perspective de leur choix de spécialisation. Les choix proposés de cours sont donc reliés aux spécialisations de 3<sup>ème</sup> année.</p> <p><b>Structuration de l'UE :</b>          L'UE est structurée en 5 blocs (i.e. un ensemble de cours proposés aux mêmes créneaux dans l'emploi du temps). Il est nécessaire de choisir un cours dans chaque bloc, ce qui permet de construire un parcours recherche, statistique appliquée, informatique, économie, ou un parcours propre à chaque élève.</p> <p><b>Compétences ou acquis d'apprentissage à l'issue de l'UE :</b>          Diverses selon les enseignements choisis. L'objectif général est de mobiliser la diversité des compétences d'un ingénieur statisticien.</p> <p><b>Les prérequis de l'UE :</b>          Connaissances scientifiques de base de 1<sup>ère</sup> et 2<sup>ème</sup> année.</p>	

UE 2-09 – Matière 2ASTA12 – Semestre 2

## MODELES AVANCÉS DE RÉGRESSION

### *Advanced Regression Models*

<i>Enseignant</i>	: Eftychia SOLEA (Ensay)
<i>Nombre d'ECTS</i>	: 2
<i>Répartition des enseignements</i>	: 21h de cours
<i>Langue d'enseignement</i>	: Anglais
<i>Logiciels</i>	: R
<i>Documents pédagogiques</i>	: Notes de cours, exercices intégrés, correction des exercices, code R
<i>Prérequis</i>	: Bonne compréhension des cours de régression et modèle linéaire généralisé

### Modalités d'évaluation :

Examen écrit

### Description and learning objectives :

In statistical modelling, the usual framework of a parametric model is most often a convenient simplification of the “true model” which describes the random phenomenon of interest. The parametric framework is typically sufficient for the description of simple situations, but quickly loses relevance when the focus is on modelling more complex situations such as regression problems. The opposite approach consists in defining a much more general model, called “nonparametric model”, where a distribution is characterised by a function and not by a vector of real-valued parameters. Identifying the underlying distribution is then equivalent to estimating this function.

In this course we discuss popular methods for estimating the regression function beyond linearity. The first topic discusses local linear regression and local polynomial regression that achieve flexibility in estimating the regression function. Next, we study how to estimate the regression function using basis expansions and regularisation such as smoothing splines. Finally, we will work on additive models and on semiparametric models. A semiparametric model contains a finite-dimensional parameter and a functional parameter. This kind of model allows to keep both a potential for interpretability of estimates and prediction of future events, due to the presence of the parametric component, and good flexibility brought by its functional part. The module focuses on two such models: partially linear regression models and single-index models.

## Main topic that might be covered :

- Smoothing in nonparametric regression: general concepts (such as bias-variance tradeoff, kernels).
- Basis expansion, penalized regression and regularization. Definition, statistical properties (bias, variance, consistency), choice of tuning parameters. Implementation in R.
- Kernel Smoothing Methods such as Local Linear Regression and Local polynomial regression: definition, statistical properties (bias, variance, consistency), choice of tuning parameters. Implementation in R. Limits of nonparametric approaches.
- Additive models, semi-parametric models (partial linear models and single-index models) Implementation in R

## Références bibliographiques :

- Fan, J. and Gijbels, I. (1992). Variable bandwidth and local linear regression smoothers, *Annals of Statistics*, 20, 2008-2036.
- Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*, Chapman & Hall.
- Friedman, J.H. and Stuetzle, W. (1981). Projection pursuit regression, *Journal of the American Statistical Association*, 76, 817-823.
- Härdle, W., Müller, M., Sperlich, S. and Werwatz, A. (2004). *Nonparametric and Semiparametric Models*, Springer.
  
- Härdle, W. and Stoker, T.M. (1989). Investigating smooth multiple regression by the method of average derivatives, *Journal of the American Statistical Association*, 84, 986-995.
- Hastie, T.J. and Tibshirani, R.J. (1990). *Generalized Additive Models*, Chapman & Hall.
- Horowitz, J.L. (2009). *Semiparametric and Nonparametric Methods in Econometrics*, Springer.
- Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models, *Journal of Econometrics*, 58, 71-120.
- Klein, R.W. and Spady, R.H. (1993). An efficient semiparametric estimator for binary response models, *Econometrica*, 61, 387-421.
- Powell, J.L., Stock, J.H. and Stoker, T.M. (1989). Semiparametric estimation of index coefficients, *Econometrica*, 51, 1403-1430.
- Robinson, P.M. (1988). Root-N-consistent semiparametric regression, *Econometrica*, 56, 931-954.
- Wasserman, L. (1999). *All of Nonparametric Statistics*, Springer.

UE 2-09 – Matière 2ASTA13 – Semestre 2

## MARTINGALES ET PROCESSUS DE LÉVY

### *Martingales and Levy Process*

<i>Enseignant</i>	: Basile DELOYNES (Ensay)
<i>Nombre d'ECTS</i>	: 2.
<i>Répartition des enseignements</i>	: 21h dont 12h de cours, 9h de TD
<i>Langue d'enseignement</i>	: Française
<i>Logiciels</i>	: N/A
<i>Documents pédagogiques</i>	: Notes de cours, planches de TD
<i>Prérequis</i>	: Théorie de la mesure, Probabilité générale

### Modalités d'évaluation :

Examen écrit.

### Acquis d'apprentissage (objectifs) :

Connaître et savoir appliquer les techniques de bases faisant appel aux outils de type martingale. Maîtriser notamment les notions de temps d'arrêt et de crochet de martingales.

Quelques notions de processus à temps continu : mouvement brownien, processus de Poisson, processus de Poisson composé.

### Principales notions abordées :

1. Processus stochastiques, filtrations, temps d'arrêt
2. Martingales sous-martingales et sur-martingales en temps discret
3. Théorèmes d'arrêt et de convergence
4. Processus en temps continu, processus de Poisson, mouvement brownien

### Références bibliographiques :

- DURETT R., *Probability : theory and examples* (Cambridge University Press, 2019)
- WILLIAMS D., *Probability with martingales* (Cambridge University Press, 1991)

UE 2-09 – Matière 2ASTA14 – Semestre 2

# MÉTHODES DE RÉÉCHANTILLONNAGE

## *Resampling Methods*

<i>Enseignant</i>	: Jad Beyhum
<i>Nombre d'ECTS</i>	: 2
<i>Répartition des enseignements</i>	: 21h de cours
<i>Langue d'enseignement</i>	: Français (cours) et Anglais (lecture de texte)
<i>Logiciels</i>	: R (packages coda, boot)
<i>Documents pédagogiques</i>	: notes de cours sous format papier
<i>Prérequis</i>	: probabilité, statistique inférentielle, régression, apprentissage supervisé, séries temporelles, calcul bayésien

### Modalités d'évaluation :

Examen écrit

### Acquis d'apprentissage (objectifs) :

- Mobiliser ses connaissances de probabilité et de statistiques inférentielles pour la définition d'une méthode de rééchantillonnage adaptée à l'inférence considérée
- Mettre en œuvre une méthode de rééchantillonnage
- Lire et discuter un texte scientifique utilisant des méthodes de rééchantillonnage à des fins statistiques.

### Principales notions abordées :

L'inférence statistique repose généralement sur la connaissance des distributions de probabilités des statistiques. Si la loi d'une statistique est inconnue, une méthode de rééchantillonnage cherche à approcher par simulation de Monte Carlo cette distribution conditionnellement aux données observées. Le principe de cette technique est de substituer à la loi de probabilité inconnue une distribution empirique construite à partir de l'échantillon d'apprentissage.

L'objectif de ce cours est de présenter les principales méthodes de rééchantillonnage : les tests par permutations, la validation croisée, le jackknife, le bootstrap. Le cours alternera description des procédures, preuves mathématiques, exercices et études de cas pratiques. Le cours portera essentiellement sur des modèles statistiques indépendant et identiquement distribués. En fonction du temps, nous verrons des extensions du Bootstrap appliquées par exemple aux modèles hétéroscédastiques (Wild Bootstrap), aux données dépendantes (Block Bootstrap, Subsampling) ou massives (Bag of Little Bootstrap).

#### 1. Tests par permutations

#### 2. La validation croisée et le jackknife

#### 3. Le Bootstrap d'Efron

#### 4. Cas pathologiques et dépendance temporelle

Une liste d'exercices non corrigés est fournie pour permettre aux étudiants de s'exercer en supplément des cours.

## Références bibliographiques :

- EFRON B, Tibshirani R (1993) *An introduction to the bootstrap*. New York: Chapman and Hall/CRC.
- GENTLE, J. E. (2009). *Computational statistics*. Springer Science & Business Media
- RIZZO, M.L. (2008) *Statistical Computing with R*. New York: Chapman and Hall/CRC.
- PESARIN, F, SALMASO, L. (2010) *Permutation Tests for Complex Data*. Wiley
- BHATTACHARYA, R. N., LIN, L., & PATRANGENARU, V. (2016). *A course in mathematical statistics and large sample theory*. Springer.

UE 2-09 – Matière 2ASTA16 – Semestre 2

## STATISTIQUE MATHÉMATIQUE

*Mathematical Statistics*

*Enseignant* : Adrien SAUMARD

*Nombre d'ECTS* : 2

*Répartition des enseignements* : 21h de cours

*Langue d'enseignement* : Anglais

*Logiciels* : Aucun

*Documents pédagogiques* : cours au tableau

*Prérequis* : statistique et probabilités générales de 1A

### Modalités d'évaluation :

Devoir maison et Examen écrit.

### Acquis d'apprentissage (objectifs) :

- Connaissance de la théorie rigoureuse, non-asymptotique, de la minimisation empirique du risque, en lien avec l'apprentissage statistique et les statistiques non-paramétriques. Théorie dite PAC, pour « Probably Approximately Correct ».
- Savoir discuter le compromis biais-variance et le phénomène de sur-apprentissage.
- Connaître les propriétés principales des variables aléatoires sous-Gaussiennes et sous-exponentielles, ainsi que les inégalités de concentration de base pour les sommes de variables indépendantes.
- Savoir établir une borne supérieure d'excès de risque, via l'utilisation de l'entropie métrique.
- Estimation optimale de la moyenne pour des variables à queue lourde.

### Principales notions abordées :

- Minimisation empirique de risque : perte, risque, excès de risque.
- Variables sous-Gaussiennes et sous-exponentielles.
- Méthode de Chernoff pour les inégalités de concentration.
- Inégalités de Hoeffding et de Bernstein pour les sommes de variables aléatoires indépendantes.
- Nombre de recouvrement par boules, entropie métrique.
- Notion de borne agnostique d'apprentissage et de borne d'apprenabilité.
- Lien entre dimension métrique et vitesse de convergence.
- Estimateur « Médiane-des-Moyennes » (MoM) de la moyenne.
- Lois à queues lourdes.
- Notion d'estimateur sous-Gaussien.

### Références bibliographiques :

- Shalev-Shwartz and Ben-David, Understanding Machine Learning, from theory to algorithms, Cambridge University Press
- Vershynin, High-dimensional Probability, an introduction with applications to data science, Cambridge University Press
- Wainwright, High-dimensional Statistics, A non-asymptotic viewpoint, Cambridge University Press

UE 2-09 – Matière 2ASTA17 – Semestre 2

## ÉCHANTILLONNAGE À PLUSIEURS DEGRÉS

### *Multistage Sampling*

<i>Enseignant</i>	: Guillaume CHAUVET (Ensaï)
<i>Nombre d'ECTS</i>	: 2
<i>Répartition des enseignements</i>	: 21h dont 12h de cours et 9h de TD/TP
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: SAS
<i>Documents pédagogiques</i>	: Diaporama du cours
<i>Prérequis</i>	: Théorie des sondages, régression linéaire, modèle linéaire généralisé

### Modalités d'évaluation :

Examen sur table

### Acquis d'apprentissage (objectifs) :

Lorsque l'on souhaite réaliser une enquête, il arrive fréquemment que l'on ne dispose pas d'une base de sondage, i.e. d'une liste des unités de la population. Dans ce cas, on utilise souvent des plans de sondage à plusieurs degrés où des échantillons sont sélectionnés de façon emboîtée. Les enquêtes auprès des ménages sont par exemple souvent réalisées en sélectionnant des communes, puis des pâtés de maison, puis finalement des ménages.

Dans ce type d'enquête, l'échantillonnage est une étape complexe car plusieurs tirages sont nécessaires. Cela complique à la fois le calcul des estimateurs, et l'estimation de leur variance. Par nature, cette procédure de sélection introduit des dépendances dans les observations, ce qui complique également la correction de la non-réponse et l'analyse des données d'enquête.

L'objectif de ce cours est de présenter les méthodes d'échantillonnage à plusieurs degrés, et de comprendre ses particularités. Des outils simples d'estimation de variance utilisant par exemple les méthodes de rééchantillonnage seront présentées. On discutera également des spécificités du traitement de la non-réponse pour ce type d'enquêtes.

### Compétences acquises :

- Choisir une stratégie d'échantillonnage et la mettre en œuvre pour des données en grappe.
- Produire les estimateurs et les indicateurs de précision associés.
- Choisir une stratégie de traitement de la non-réponse totale et la mettre en œuvre.

## Principales notions abordées :

### Partie 1 : Introduction

- Rappels sur l'échantillonnage en population finie
- Tirage à deux degrés : échantillonnage et estimation
- Tirage par grappes : échantillonnage et estimation

### Partie 2 : Précision d'un sondage à deux degrés

- Estimateur direct de variance
- Estimateurs simplifiés
- Estimation par bootstrap

### Partie 3 : Estimation pour un sondage à deux degrés

- Traitement de la non-réponse totale
- Calage au niveau ménage et individu
- (Analyse des données d'enquête)

## Références bibliographiques :

- ARDILLY P., *Les Techniques de Sondage (nouv. éd.)*, Paris, Technip, 2006
- SÄRNDAL, C-E., SWENSSON, B, WRETMAN, J., *Model Assisted Survey Sampling*, Springer-Verlag, 2003

UE 2-09 – Matière 2ASTA18 – Semestre 2

## STATISTIQUE SPATIALE

### *Spatial analysis*

<i>Enseignant</i>	: David AUDENAERT
<i>Nombre d'ECTS</i>	: 2
<i>Répartition des enseignements</i>	: 21h de TP
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: R, Qgis
<i>Documents pédagogiques</i>	: Supports de cours, liens bibliographiques
<i>Prérequis</i>	: aucun

### Modalités d'évaluation :

Devoir maison en groupe

### Acquis d'apprentissage (objectifs) :

Introduction aux différents types de données spatiales et à leurs méthodes d'analyse. Initiation à la cartographie avec Qgis et R. Sensibilisation au problème d'agrégation spatiale à celui de la méthode de discrétisation des données. L'accent est mis sur les applications avec des données réelles de type surfacique. Leur analyse se fera avec le logiciel R.

### Principales notions abordées :

- Les différents types de données spatiales (surfaciques, ponctuelles et continues) et leur méthode d'analyse
- Préparer ses données statistiques à la représentation cartographique
- Les différents types de représentation cartographique (notions de sémiologie graphique)
- L'importance de la méthode de discrétisation dans le cas de cartes choroplèthes.
- Le problème d'agrégation spatiale (MAUP) et le carroyage.
- Identifier un phénomène d'autocorrélation spatiale (données continues).

### Références bibliographiques :

- Manuel d'analyse spatiale : Théorie et mise en œuvre pratique avec R <https://www.insee.fr/fr/information/3635442>
- A. Baddeley (2010), "Analysing spatial point patterns in R", CSIRO.
- R. Bivand, E. Pebesma et V. Gómez-Rubio (2013), "Applied spatial data analysis with R", Springer-Verlag, 2<sup>e</sup> édition.

UE 2-09 – Matière 2AECO04 – Semestre 2

## ECONOMIE DU RISQUE

### *Economics of Risk*

<i>Enseignant</i>	: Olivier RENAULT
<i>Nombre d'ECTS</i>	: 2
<i>Répartition des enseignements</i>	: 21h de cours
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: N/A
<i>Documents pédagogiques</i>	: Cours et exercices en ligne
<i>Prérequis</i>	: Microéconomie standard

### Modalités d'évaluation :

Examens : exercices et questions de réflexion

### Acquis d'apprentissage (objectifs) :

Le cours se divise en 2 parties. Dans une première partie, nous étudierons les choix en contexte de risque et appliquerons les résultats théoriques à plusieurs champs d'application : assurance, finance, décision publique. Dans la seconde partie, nous étendrons notre analyse aux choix en contexte incertain en proposant une introduction à l'incertitude stratégique.

### Principales notions abordées :

1. Introduction : les préférences rationnelles, modélisation d'un choix, le risque et l'incertitude
2. Le choix en univers risqué
  - 2.1. Notions centrales : dominance stochastique, aversion pour le risque, prime de risque
  - 2.2. Deux premiers critères : espérance et espérance-variance
  - 2.3. Espérance d'utilité, mesure de l'attitude face au risque et paradoxes comportementaux
  - 2.4. Applications : choix d'assurance, choix d'épargne, choix de portefeuille
3. Le choix en univers incertain
  - 3.1. Deux premiers critères : Maximin et critère de Hurwitz
  - 3.2. Espérance subjective d'utilité
  - 3.3. Introduction à l'incertitude stratégique
  - 3.4. Prolongements : paradoxes comportementaux et aversion à l'ambiguïté

### Références bibliographiques :

- Eeckhoudt Louis, Christian Gollier et Harris Schlesinger (2005). Economic and financial decisions under uncertainty, Louis, Princeton University Press
- Gayant Jean-Pascal. Risque et décision, Vuibert
- Gollier Christian (2001). The Economics of Risk and Time, MIT Press
- Jokung Nguena Octove. Microéconomie de l'incertain, Dunod

UE 2-09 – Matière 2AECO05 – Semestre 2

## ECONOMIE NUMERIQUE

*Economics of Digital Technology, Innovation and Media*

<i>Enseignants</i>	: Yutec SUN (ENSAI)
<i>Nombre d'ECTS</i>	: 2
<i>Répartition des enseignements</i>	: 21 heures de cours
<i>Langue d'enseignement</i>	: Anglais
<i>Logiciels</i>	: Sans objet
<i>Documents pédagogiques</i>	: Diapositives de cours et articles de recherche
<i>Prérequis</i>	: Formation économétrique et microéconomique basique

### Modalités d'évaluation :

Students are expected to present an article of their choice and to participate in discussion. The performance of presentation and discussion will be the basis for evaluation.

### Acquis d'apprentissage (objectifs) :

- Savoir expliquer les mécanismes qui régissent l'économie numérique à partir d'exemples concrets.
- Savoir utiliser la boîte à outils de l'économiste pour comprendre et prédire l'évolution des marchés transformés par le numérique.
- Connaître les défis réglementaires auxquels font face les pouvoirs publics.

### Principales notions abordées :

Le cours abordera les notions introductives (offre des biens numériques, demande et rôle des évaluations par les consommateurs, concurrence et régulation) et plus avancées avec des exemples issus de l'économie industrielle appliquée et une introduction à la programmation dynamique. Mots clés : Transformation numérique ; marchés bifaces et plateformes; innovation; coûts de changement ; big data et intelligence artificielle; structure de marché et concurrence; collusion par les algorithmes.

### Références bibliographiques :

Brousseau, E., & Curien, N. (Eds.). (2007). Internet and digital economics: principles, methods and applications. Cambridge University Press.

Goldfarb, A., & Tucker, C. (2019). Digital economics. *Journal of Economic Literature*, 57(1), 3-43.

Parker, G. G., Van Alstyne, M. W., & Choudary, S. P. (2016). Platform revolution: How networked markets are transforming the economy and how to make them work for you. WW Norton & Company.

## Course outline

### Part 1: An overview of the Digital Economy (12 hours)

---

#### 1.1. Introduction (1 hour)

#### 1.2. Supply

1.2.1. Costs, prices and quality in the Digital Era

1.2.2. Platform markets

1.2.3. Data as a key input

#### 1.3. Demand

1.3.1. Ratings, reviews and other user-generated content

1.3.2. Substitution between consumption channels

1.3.3. Switching costs

#### 1.4. Platform Competition

1.4.1. Entry and exit of firms, barriers to entry and emergence of big players

1.4.2. Insights from the platform-market literature

#### 1.5. Regulation

1.5.1. Abuse of dominant position: Price parity clauses, BigTech mergers and Killer Acquisitions

1.5.2. Is Big Data a threat to competition?

#### 1.6. Innovation

1.6.1. Monopoly may not be so bad: evidence from the CPU war

1.6.2. Innovator's dilemma: why market leaders fail to innovate

1.6.3. Killing innovative competitors: evidence from pharmaceutical drugs

#### 1.7. Artificial intelligence

1.7.1. Can AI facilitate firm's collusion?

1.7.2. Empirical evidence from the gasoline prices

1.7.3. Advanced AI algorithms: promises and threats

UE 2-09 – Matière 2AECO06 – Semestre 2

## ECONOMIE INDUSTRIELLE

### *Industrial Organization*

<i>Enseignant</i>	: Yutec SUN
<i>Nombre d'ECTS</i>	: 2
<i>Répartition des enseignements</i>	: 21h de cours
<i>Langue d'enseignement</i>	: Anglais
<i>Logiciels</i>	: N/A
<i>Documents pédagogiques</i>	: supports de cours
<i>Prérequis</i>	: Microéconomie

### Modalités d'évaluation :

Examen

### Acquis d'apprentissage (objectifs) :

Les compétences acquises sont de trois ordres. En premier lieu, ce cours permet d'acquérir et d'appliquer des concepts de théories des jeux non coopératifs, indispensables à l'analyse et la compréhension de la concurrence imparfaite. En deuxième lieu, ce cours permet de comprendre le positionnement stratégique des firmes (tarification pratiquée, qualité des services, choix de produits...). Cela permet de comprendre une micro-économie plus proche des décideurs. La compréhension des mécanismes qui sont en jeu, de son langage sont désormais indispensables dans les cabinets de conseil, et parmi les économistes d'entreprise. Enfin, ce cours permet d'analyser les fusions d'entreprise du point de vue du bien-être social, et d'étudier la politique de la concurrence. Il permet ainsi d'acquérir tous les outils indispensables aux professionnels responsables de la concurrence travaillant en particulier dans des agences gouvernementales (autorité de la concurrence, DGCCRF, Commission européenne ...)

### Principales notions abordées :

Le cours d'économie industrielle vise à étudier de façon micro-économique des situations de concurrence imparfaite (monopole, oligopole en prix, en quantité...). Il vise aussi à analyser une situation de monopole discriminant du point de vue du bien-être social. Il a aussi pour objectif de montrer comment les entreprises en concurrence en prix peuvent sortir du paradoxe de Bertrand en différenciant leurs produits. Enfin, il cherche à analyser les relations verticales entre entreprises, les études des concentrations, et la politique de la concurrence.

### Références bibliographiques :

- TIROLE, J. *Théorie de l'Organisation Industrielle*, Economica, 1998.
- MOTTA, M. *Competition Policy. Theory and Practice*. Cambridge University Press, 2004.
- AMSTRONG, M. and PORTER, R. *Handbook of Industrial Organization, Vol3*, Elsevier, 2007
- Pepall, L., Richards D. and Norman G. *Industrial Organization: Contemporary Theory and Empirical Applications*, 2008.

UE 2-09 – Matière 2AECO07 – Semestre 2

## DEMOGRAPHIE

### Demographics

<i>Enseignant</i>	: A déterminer
<i>Nombre d'ECTS</i>	: 2
<i>Répartition des enseignements</i>	: 21h de cours
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: Pack office
<i>Documents pédagogiques</i>	: Supports commentés sur moodle (audio et diapo) + exercices
<i>Prérequis</i>	: Aucun

### Modalités d'évaluation :

Quiz, examens, devoirs maison, mini projets...

### Acquis d'apprentissage (objectifs) :

L'objectif est de comprendre les enjeux de l'analyse démographique, d'en connaître et maîtriser les outils et indicateurs, de savoir appréhender les caractéristiques actuelles des démographies française et mondiales, ainsi que de comprendre la construction des modèles principaux qui permettent de décrire et d'analyser la dynamique d'une population.

### Principales notions abordées :

- |   |  |
|---|--|
| 1. Introduction à la démographie  | 5. L'analyse de la fécondité                           |
| 1.1. Définition   | 5.1. Le taux brut de natalité                          |
| 1.2. Les grands instituts historiques   | 5.2. Le taux de fécondité par âge                      |
| 1.2. Les principales bases de données démographiques                            | 5.3. L'indicateur conjoncturel de fécondité            |
|   | 5.4. Les causes du niveau élevé de fécondité en France |
| 2. La croissance de la population   | 5.5. L'âge moyen à la maternité                        |
| 2.1. Mesure des facteurs d'évolution de la population : l'équation fondamentale | 5.6. L'âge moyen de fécondité                          |
| 2.2. Évolution de la population en France sur longue période                    | 5.7. Le taux brut de reproduction                      |
| 2.3. Les composantes du solde migratoire  | 5.8. Le taux net de reproduction                       |
| 2.4. Les immigrés   | 5.9. La descendance finale                             |
| 2.5. Les Français de l'étranger   | 6. L'analyse de la mortalité                           |
|   | 6.1. Taux brut de mortalité                            |
| 3. La localisation de la croissance démographique                               | 6.2. Les taux de mortalité par âge                     |
| 3.1. La démographie des régions françaises                                      | 6.3. La standardisation des taux de mortalité          |
| 3.2. L'étalement urbain   | 6.4. La mortalité infantile                            |
|   | 6.5. La table de mortalité                             |
|   | 6.6. L'espérance de vie                                |
| 4. Outils de la démographie   | 7. La transition démographique                         |
| 4.1. La pyramide des âges   |  |
| 4.2. Promenade au travers des pyramides   | 8. Les indicateurs de développement humain (ONU)       |
| - démographie de l'Inde   |  |
| - démographie de l'Allemagne  |  |

- démographie du Niger
  - démographie du Japon
- 4.3. Le diagramme de lexis

## Références bibliographiques

- PNUD, rapport sur le développement humain, 2019
- Rapport d'activité de l'OFII 2019
- Catherine Rollet, Introduction à la démographie, 3<sup>e</sup> édition, Armand Colin, 2011
- Jean-Claude Chesnais, la démographie, Que sais-je, PUF, 2010
- Gabriel Poulalion et Georges Pupion, Analyse démographique, L'Harmattan, 2009
- Jacques Vallin, La démographie, Etude (poche), collection repères, La Découverte, 2002

UE 2-09 – Matière 2AINF07 – Semestre 2

## PROGRAMMATION AVANCÉE EN R

*Advanced programming in R*

<i>Enseignant</i>	: Sébastien Da Veiga
<i>Nombre d'ECTS</i>	: 2
<i>Répartition des enseignements</i>	: 21h de cours /TP
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: R
<i>Documents pédagogiques</i>	: Supports de cours disponibles sous Moodle
<i>Prérequis</i>	: Statistiques avec R (1A)

### Modalités d'évaluation :

Compte-rendu et scripts du dernier TP.

### Acquis d'apprentissage (objectifs) :

Le langage R permet au statisticien de développer ses propres programmes. Cependant, le temps d'exécution de ces programmes peut être long. Lorsque ce temps d'exécution devient problématique il faut alors chercher à optimiser le programme.

Le but de cet enseignement est de donner aux étudiants :

- des outils permettant de rechercher les parties du programme qui posent problème
- des pratiques de programmation en R permettant d'obtenir des programmes plus rapides
- un aperçu de l'utilisation de bibliothèques C ou C++ en R

### Principales notions abordées :

1. Syntaxe de R – écrire des programmes efficaces en R
2. Les outils de profilage
3. Intégrer du code C/C++ dans R

### Références bibliographiques :

P.A. Cornillon, A. Guyader, F. Husson, N. Jégou, J. Josse, M. Kloareg, E. Matzner-Løber, L. Rouvière (2013). Statistiques avec R, 3e édition augmentée et en couleurs, Presses Universitaires de Rennes, France.

UE 2-09 Semestre 2

# CLOUD COMPUTING

## Cloud Computing

Enseignant	: À définir
Nombre d'ECTS	: 2
Répartition des enseignements	: 21h d'atelier (cours + TP)
Langue d'enseignement	: Français
Logiciels	: Un navigateur web, un IDE python, PC personnel recommandé
Documents pédagogiques	: Sur moodle, pas de document papier
Prérequis	: Python niveau initiation, BdD niveau initiation

### Modalités d'évaluation :

Quizz, mini projet

### Acquis d'apprentissage (objectifs) :

- Différencier le *cloud computing* du *on-premises computing*
- Sélectionner le meilleur type de service entre IaaS, PaaS, SaaS pour un besoin particulier
- Provisionner des services divers dans le cloud
- Manipuler différents services de base dans le cloud (stockage, calcul, base de données)
- Déployer une architecture hautement disponible dans le cloud

### Principales notions abordées :

Depuis l'arrivée en 2006 de l'offre commerciale d'*Amazon Web Service* (AWS) comportant un service de stockage et de calcul sur le marché, le monde de l'informatique s'est drastiquement métamorphosé. Plus besoin pour une entreprise d'investir dans un *datacenter*, il lui suffit en quelques clics de louer des machines pour répondre à ses besoins. Quinze ans après, AWS propose désormais plus de 200 services gérés, et n'est plus seul sur le marché (citons *Google Cloud Platform* (GCP), *Microsoft Azure*, OVH).

S'il est plus nécessaire d'investir dans des machines physiques, il faut néanmoins comprendre comment fonctionnent les services proposés par un *Cloud Provider* et comment les faire communiquer entre eux. Le cours va ainsi aborder les notions suivantes :

- Le stockage objet
- Les instances de calcul
- Les bases de données
- Les règles basiques de sécurité dans le cloud
- La création de réseau
- Le *serverless*
- L'*Infrastructure as Code*

À la fin du cours, les élèves seront capables de déployer une architecture hautement disponible dans le cloud.

## Références bibliographiques :

AWS Cookbook. (2022). O'Reilly Media.

Google Cloud Cookbook : Practical Solutions for Building and Deploying Cloud Services. (2022). O'Reilly Media.

Hurwitz, J. S., & Kirsch, D. (2020). Cloud Computing For Dummies, 2nd Edition (2e éd.). For Dummies.

Morris, K. (2021). Infrastructure as Code : Dynamic Systems for the Cloud Age (2e éd.). O'Reilly Media.

Piper, B., & Clinton, D. (2020). AWS Certified Solutions Architect Study Guide : Associate SAA-CO2 Exam (Aws Certified Solutions Architect Official : Associate Exam) (3e éd.). Sybex.

UE 2-09 – Matière 2AINF09 – Semestre 2

## TRAITEMENT DU SIGNAL

### *Signal Processing*

<i>Enseignant</i>	: Hong-Phuong DANG
<i>Nombre d'ECTS</i>	: 2
<i>Répartition des enseignements</i>	: 21h dont 9h de cours, et 12h de TP
<i>Langue d'enseignement</i>	: anglais
<i>Logiciels</i>	: Python - Jupyter notebook
<i>Documents pédagogiques</i>	: Support de cours
<i>Prérequis</i>	: Un peu de maths : suite et série de fonction, intégration, nombre complexe, base de la théorie de la mesure et ☐ Connaissance de langage Python pour les TPs

### Modalités d'évaluation :

Quiz, examens, devoirs maison, mini projets...

### Acquis d'apprentissage (objectifs) :

Acquérir les notions de base pour :

- Manipuler des signaux analogiques et numériques
- Modéliser et représenter le signal
- Effectuer des opérations simples de traitement

Le Traitement du Signal (TdS) apporte un autre point de vue sur les notions déjà vues dans les cours de statistiques et d'économie, par exemple corrélation, convolution, stationnarité, causalité...

Un des fondamentaux du TdS est l'analyse temps-fréquence (spectrogramme) avec la transformation fourrier. Nous introduisons les motivations vers la base d'ondelettes.

Nous abordons également dans cette matière la notion de représentation parcimonieuse à travers l'échantillonnage des signaux. La représentation parcimonieuse est à l'origine des modèles de régression pénalisée (LASSO) ou de la régularisation. Ces notions apparaîtront dans toutes les filières ingénieur de 3<sup>e</sup> année, en Master Smart Data et Master de Statistique Publique.

De plus, le signal intervient sous plusieurs formes dans la plupart des domaines des différentes technologies. On a, par exemple, le traitement de l'image, la télécommunication, l'acoustique, l'astronomie, l'économie, la biologie, l'optique, la mécanique, l'électricité, l'électronique. On retrouve de plus en plus de demandes d'entreprises sur les compétences dans cette discipline. Pour finir, cette matière permet, non seulement, aux élèves d'avoir un bagage complémentaire pour leurs cours en 3<sup>e</sup> année, mais aussi d'avoir une connaissance sur les notions de base de traitement du signal afin de pouvoir mieux comprendre les besoins des entreprises pour un stage ou un emploi.

### Principales notions abordées :

1. Définition d'un signal, énergie, puissance, périodicité, corrélation, convolution, bruit (signal perturbateur, rapport signal sur bruit,...). Notion de différents signaux (par dimension, phénomène, énergie, morphologie)
2. Représentation fréquentielle : notion de fréquence (fondamentale + harmonique), distribution de Dirac (élément neutre de l'opérateur de convolution), transformation Fourier (propriétés, exemples, spectre), théorème de Gabor (Principe d'incertitude de Heisenberg : mécanique quantique), égalité de Parseval Plancherel (conservation de l'énergie).
3. Théorie du filtrage : linéarité, principe de superposition, fonction de transfert, réponse impulsionnelle, passe-bas/bande/haut
4. La discrétion en temps et fréquence (lié au numérique), échantillonnage des signaux, Shannon-Nyquist, analyse temps-fréquence (spectrogramme)

### Références bibliographiques :

Martin Vetterli, Jelena Kovacevic, and Vivek K Goyal, *Foundations of Signal Processing*, 2014  
Kido Ken'iti, *Digital Fourier Analysis: Fundamentals*, Springer, 2015

UE 2-09 – Matière 2AINF10 – Semestre 2

## CONCEPTION DE LOGICIEL

### Software design

<i>Enseignant</i>	: Antoine Brunetti
<i>Nombre d'ECTS</i>	: À remplir
<i>Répartition des enseignements</i>	: 21h
<i>Langue d'enseignement</i>	: À remplir
<i>Logiciels</i>	: À remplir
<i>Documents pédagogiques</i>	: À remplir
<i>Prérequis</i>	: À remplir

### Modalités d'évaluation :

Quiz, examens, devoirs maison, mini projets...

### Acquis d'apprentissage (objectifs) :

En dehors des preuves de concept, la conception et la réalisation d'un logiciel nécessite de découper ce dernier en différentes fonctionnalités, en différents sous programmes ou composants logiciel,... On retrouve là le concept habituel en informatique : diviser pour régner ou diviser pour résoudre.

Ce principe s'applique aussi bien à la phase d'implémentation qu'à la phase de conception. Il s'appuie aujourd'hui sur des stratégies standards et le plus souvent sur le langage de modélisation UML.

A la fin de ce cours, les élèves connaîtront un ensemble de ces stratégies, et seront en mesure d'appliquer ces stratégies pour la conception et l'implémentation de petites applications. S'il y a lieu, les applications modélisées seront implémentées en Java.

- Concevoir une application
- Découper en fonctionnalités / composants / bibliothèques / ...
- Modéliser l'application
- Modéliser les données manipulées

### Principales notions abordées :

- Cycle de vie d'un logiciel (analyse, conception, implémentation, test, utilisation et maintenance corrective ou évolutive)
- Data UML (et approfondissement d'UML)
- Programmation par Design pattern (patrons E.GAMMA : les modèles de créations, les modèles de structures et les modèles de comportements)
- Programmation par composants
- Programmation orientée objet : généricité / traits / modèle de composant UML2.2
- Les méthodologies de conception OO : processus unifié / méthode AGILE ?
- Tests fonctionnels et structurels

### Références bibliographiques :

Design patterns : catalogue de modèle de conception réutilisable (1995 version anglaise, version française (traduction) 1990)

UE 2-09 – Matière 2AINF11 – Semestre 2

## DATA CHALLENGE

### *Datachallenge*

<i>Enseignant</i>	: Salima EL KOLEI
<i>Nombre d'ECTS</i>	: 2
<i>Répartition des enseignements</i>	: 21h
<i>Langue d'enseignement</i>	: A remplir
<i>Logiciels</i>	: Aucun
<i>Documents pédagogiques</i>	: Aucun
<i>Prérequis</i>	: Aucun

### **Modalités d'évaluation :**

Quizz, examens, devoirs maison, mini projets...

### **Acquis d'apprentissage (objectifs) :**

Le data challenge permet de rassembler sur une période très courte différentes équipes de profils variés afin de collaborer sur un projet. Cette expérience se rapproche des conditions réelles dans lesquelles évoluent les datascientists au sein des entreprises.

Il permet, à partir des mécanismes du jeu, de dynamiser et d'articuler la pédagogie autour d'un besoin concret d'entreprise et d'un événement qui s'achève par une évaluation objective. De nombreux challenges sont proposés autour de la Data ou présentant des problématiques Data importantes.

L'objectif de ce cours électif est de valoriser et d'évaluer les compétences transversales acquises dans ce contexte opérationnel.

- Comprendre les problèmes à résoudre.
- Travailler en mode projet avec des contraintes.
- S'intégrer et s'adapter dans un contexte pluridisciplinaire. Selon les challenges, les compétences seront mobilisées à géométrie variable.
- S'adapter à la réalité de la Data d'entreprise (données non structurées, manquantes, volumétrie...)
- Communication orale des résultats (pitch...)

### **Principales notions abordées :**

Les élèves inscrits au cours électif devront participer à au moins un challenge. Ils devront témoigner leur expérience dans un rapport écrit et une soutenance.

Ils devront participer aux deux séquences d'ateliers sur les thématiques liées au travail en équipe, en synergie et dans l'urgence.

Un apport de ressources permettant aux élèves de se préparer sera proposé en particulier sur les plateformes: <https://challengedata.ens.fr/en/home> et <https://www.datascience.net/fr/challenge>

UE 2-09 – Matière 2AINF12 – Semestre 2

## DATAVISUALISATION

### *Datavizualisation*

<i>Enseignant</i>	: Arthur KATOSSKY (Ensay)
<i>Nombre d'ECTS</i>	: 2
<i>Répartition des enseignements</i>	: 21h de cours
<i>Langue d'enseignement</i>	: Anglais
<i>Logiciels</i>	: R et JavaScript
<i>Documents pédagogiques</i>	: support de cours, vidéos et fiches de TP
<i>Prérequis</i>	: connaissance élémentaires de R ; une connaissance, même superficielle, des langages web (HTML, CSS, SVG, JavaScript) est un avantage, mais n'est pas nécessaire

### Modalités d'évaluation :

2 mini-projets

### Acquis d'apprentissage (objectifs) :

- Formuler une critique constructive d'une infographie donnée à partir d'une connaissance élémentaire de la sémiologie graphique (pour les conventions de représentation) et de la psychologie cognitive (pour la perception visuelle)
- Reproduire fidèlement une infographie statistique donnée avec le package ggplot2 de R
- Comprendre la logique d'une application codée en d3.js (une autonomie complète n'est pas attendue, même pour des applications simples)

### Principales notions abordées :

- Perception visuelle, conventions graphiques, échelles et couleurs
- *Trifecta Checkup* et critique d'infographies statiques
- La grammaire des graphiques
- Bibliothèque ggplot2 de R pour la visualisation de données
- Introduction à la programmation événementielle
- Initiation à la bibliothèque d3.js **sur Observable**
- Conception d'une infographie interactive

## Références bibliographiques :

- Munzner, T. 2014. *Visualization Analysis and Design*.
- Cairo, Alberto. 2019. *How Charts Lie: Getting Smarter about Visual Information*. 1 edition. W. W. Norton & Company.
- Tufte, Edward R. 2001. *The Visual Display of Quantitative Information*. 2nd edition. Graphics Pr.
- <https://junkcharts.typepad.com>
- Bertin, Jacques. 2013. *Sémiologie graphique : Les diagrammes, les réseaux, les cartes*. Editions de l'Ecole des Hautes Etudes en Sciences Sociales.
- Murray, Scott. 2017. *Interactive Data Visualization for the Web : An Introduction to Designing with D3*. 2nd ed. O'Reilly Media, Inc, USA.
- Wickham, Hadley. 2009. *ggplot2: Elegant Graphics for Data Analysis (Use R!) (English Edition)*. 1st ed. 2009. Corr. 3rd printing 2010. Springer. (<https://ggplot2-book.org>)

UE 2-10 – Semestre 2

## UE 2-10 : PROJET PROFESSIONNEL

<i>Correspondant de l'UE</i>	: Laurent TARDIF
<i>Nombre d'ECTS</i>	: 1
<i>Nombre d'heures d'enseignement</i>	: Sans objet

### Finalité de l'UE :

L'UE « projet professionnel » participe pleinement au développement des compétences comportementales et soft-skills permettant de valoriser le potentiel de data-scientist des étudiants.

L'objectif principal est de préparer les élèves aux candidatures pour les stages et entretiens de recrutement ainsi que leurs interactions dans le cadre professionnel.

### Structuration de l'UE :

Le projet professionnel permet aux étudiants de mieux se connaître à travers en particulier l'utilisation d'un test de personnalité très utilisé pour les recrutements, de valoriser leurs compétences dans leur recherche de stage ou d'emploi en préparant les documents supports de leurs candidatures et de mettre en pratique les compétences scientifiques et comportementales dans une expérience professionnelle durant leur stage d'application en statistique. Par ailleurs, des Rendez-Vous Carrière sont organisés en 2<sup>e</sup> année, ils permettent aux élèves ingénieurs d'appréhender leur futur métier en rencontrant des praticiens (souvent des alumni) sur un format interactif.

Un cours optionnel ou une langue optionnelle complète l'UE et participe via les compétences transversales travaillées au projet professionnel de l'étudiant.

### Compétences ou acquis d'apprentissage à l'issue de l'UE :

- Aptitude à étudier et résoudre des problèmes complexes souvent situés en univers incertain et risqué en mobilisant des connaissances scientifiques à dominante mathématique, statistique, économique et informatique.
- Capacité à s'insérer dans la vie professionnelle, à s'intégrer dans une organisation, à l'animer et à la faire évoluer : exercice de la responsabilité, esprit d'équipe, engagement et leadership, management de projets, maîtrise d'ouvrage, communication avec des spécialistes comme avec des non-spécialistes.
- Capacité à se connaître, à s'autoévaluer, à gérer ses compétences (notamment dans une perspective de formation tout au long de la vie), à opérer ses choix professionnels.
- Capacité à valoriser les ressources informationnelles pour contribuer à atteindre les objectifs de l'organisation qui l'emploie i.e. être en mesure de mettre en forme les résultats, les interpréter et leur donner du sens, les communiquer à l'oral comme à l'écrit auprès d'acteurs spécialistes ou non.

### Les prérequis de l'UE :

Aucun

UE 10 – Matière 2AHUM09 – Semestre 2

## TEST DE PERSONNALITE DANS LE MILIEU PROFESSIONNEL

### *Personality test*

<i>Enseignant</i>	: Matthieu FLOIRAT
<i>Nombre d'ECTS</i>	: 0,5
<i>Répartition des enseignements</i>	: 6h dont 2h de cours et 4h de TP dont 1h en autonomie
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: Sans objet
<i>Documents pédagogiques</i>	: Sans objet
<i>Prérequis</i>	: Sans objet

### **Modalités d'évaluation :**

Sans objet

### **Acquis d'apprentissage (objectifs) :**

Les sessions sont consacrées au questionnaire MBTI (MYERS-BRIGGS TYPE INDICATOR) et à la préparation à l'arrivée dans l'univers professionnel en resserrant et en articulant le concept des préférences psychologiques au monde de l'entreprise.

Les deux objectifs principaux de cette matière sont :

- la compréhension de soi, au travers de la découverte positive de son Type psychologique MBTI,
- la compréhension des autres au travers de la compréhension positive du fonctionnement des autres types

### **Principales notions abordées :**

L'utilisation accompagnée de l'outil psychométrique MBTI permet aux étudiants de mieux se connaître en termes de comportements, de motivations, et d'utiliser cette connaissance de soi pour leur orientation professionnelle en comprenant les situations et les contextes professionnels les plus adaptés à leur personnalité.

Par ailleurs, la prise de conscience de l'existence de fonctionnements différents selon les types de personnalité associée à une meilleure connaissance de soi, permet aussi de mieux situer son rôle dans une équipe et de mieux valoriser les atouts de chacun dans le cadre d'un projet. Ces points sont essentiels travailler en mode projet.

### **Références bibliographiques :**

Sans objet

UE 10 – Matière 2AHUM10 – Semestre 2

## CONSTRUIRE SON PROJET PROFESSIONNEL

### *Build your professional profile*

<i>Enseignant</i>	: Divers intervenants
<i>Nombre d'ECTS</i>	: 0,5
<i>Répartition des enseignements</i>	: 6h dont 3h de cours et 3h de TP
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: Word
<i>Documents pédagogiques</i>	: Fournis en cours
<i>Prérequis</i>	: Aucun

### **Modalités d'évaluation :**

L'évaluation est réalisée par des professionnels du recrutement à partir d'un CV et d'une lettre de motivation que l'élève rédige pour répondre à une offre d'emploi ou de stage qu'il a lui-même choisi.

### **Acquis d'apprentissage (objectifs) :**

- Réaliser un CV
  - Rédiger une lettre de motivation
- Créer et valoriser son profil sur les réseaux professionnels

### **Principales notions abordées :**

La matière « Construire son profil professionnel » comprend la rédaction d'un CV correspondant aux standards de recrutement actuels, la rédaction d'une lettre de motivation, et la création de son profil sur les réseaux sociaux professionnels tels que LinkedIn. L'objectif est de communiquer efficacement sur son parcours, en particulier valoriser ses expériences : parcours académiques, projets, stages, et expériences plus personnelles.

### **Références bibliographiques :**

Fournies en cours

UE 10 – Matière 2AHUM04 – Semestre 2

## PROJET PROFESSIONNEL DES ATTACHÉS

*Build your professional profile*

<i>Enseignant</i>	: Stéphane Legleye
<i>Nombre d'ECTS</i>	: 1
<i>Répartition des enseignements</i>	: 18h de TD
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: Word
<i>Documents pédagogiques</i>	: Fournis en cours
<i>Prérequis</i>	: Aucun

### Modalités d'évaluation :

Le module sera validé en fonction de l'implication de l'élève tout au long des TD.

### Acquis d'apprentissage, objectifs de la matière

- Comprendre l'activité et appréhender les domaines de compétences de l'INSEE.
- Connaître les familles de métiers à l'INSEE et dans la statistique publique.
- Comprendre les enjeux et modalités de la campagne de mobilité.
- Dégager ses motivations professionnelles et construire des parcours cohérents.

### Principales notions abordées

L'objectif est de sensibiliser les attachés stagiaires aux emplois et activités dans la statistique publique, afin de mieux cerner leur projet professionnel à court et moyen terme, et à les préparer à la campagne de mobilité (CV, lettre de motivation, choix de postes).

Les séances seront consacrées à :

Les 18 heures seront réparties sur 3 journées organisées comme suit :

- Journée 1 : Intervention sur la matinée, en séance plénière, du représentant de la Division mobilité et carrières (DMC) du Secrétaire général de l'Insee. Après-midi entretiens individuels avec la DMC et autres intervenants pour ceux qui le désirent,
- Journées 2 et 3 : alternance entre séances plénières et ateliers sur la campagne de mutation, les métiers de l'Insee et la rédaction de CV et lettre de motivation.

Un Forum des métiers de la statistique publique est organisé en Janvier, dans le cadre de la campagne de mutation.

### Prérequis

Aucun

### Références bibliographiques

Notices Insee

UE 2-10 – Matière 2AHUM06 – Semestre 2

## STAGE D'APPLICATION

*Application internship*

<i>Enseignant</i>	: Divers intervenants
<i>Nombre d'ECTS</i>	: l'évaluation du stage d'application intervient en 3e année avec une pondération de 5ECTS.

### Modalités d'évaluation :

Le stage de 2<sup>e</sup> année est évalué :

- **Par le Maître de stage** : résultats par rapport aux objectifs du stage, capacité d'initiative et d'intégration.
- **Par la Direction des études de l'ENSAI** : cette évaluation traduira la qualité méthodologique et prospective (propositions et préconisations) du rapport de stage, sa qualité rédactionnelle ainsi que l'esprit critique et la curiosité scientifique de l'élève.

**Attention** : le rapport de stage 2A étant rendu et évalué en début de 3<sup>e</sup> année, l'évaluation sera prise en compte dans l'Unité d'Enseignement « projet professionnel » de 3<sup>e</sup> année qui comprend également le stage de fin d'études.

### Acquis d'apprentissage (objectifs) :

- insertion dans un milieu professionnel
- application des enseignements 1 et 2A,
- formalisation de la démarche scientifique et communication écrite

### Principales notions abordées :

Le stage d'application (statistique, machine learning, économie appliquée) valide la compétence de statisticien généraliste normalement acquise par tous les élèves à l'issue de la 2<sup>e</sup> année. Il doit nécessairement mobiliser des enseignements de 2<sup>e</sup> année et notamment l'utilisation **d'au moins un outil de modélisation**. Le stage peut être effectué en France ou à l'étranger, dans une entreprise, un centre de recherche public ou privé, une administration ou une collectivité territoriale. S'il est effectué à l'étranger, ce stage peut permettre de valider la **Période obligatoire à l'étranger** (POE).

### Références bibliographiques :

Guide du stage 2<sup>e</sup> année, accessible sur Moodle.

UE 2-10 – 2AHUM07 – Semestre 2

## PRINCIPES DE GESTION DURABLE DES ORGANISATIONS

### Traduction

<i>Enseignant</i>	: à déterminer
<i>Nombre d'ECTS</i>	: 1
<i>Répartition des enseignements</i>	: 12h
<i>Langue d'enseignement</i>	: à remplir
<i>Logiciels</i>	: à remplir
<i>Documents pédagogiques</i>	: à remplir
<i>Prérequis</i>	: à remplir

### Modalités d'évaluation :

A remplir

### Acquis d'apprentissage (objectifs) :

A remplir

### Principales notions abordées :

A remplir

### Références bibliographiques :

À remplir

UE 2-5 – Semestre 1 / UE 2-10 – Semestre 2

## COURS LIBRE OPTIONNEL

<i>Enseignant</i>	: Divers intervenants
<i>Nombre d'ECTS</i>	: 1
<i>Répartition des enseignements</i>	: 24h
<i>Langue d'enseignement</i>	: Sans objet
<i>Logiciels</i>	: Sans objet
<i>Documents pédagogiques</i>	: Sans objet
<i>Prérequis</i>	: Sans objet

### Modalités d'évaluation :

Spécifique à chaque option

**Acquis d'apprentissage (objectifs) :**

- Développement d'une culture générale dans un domaine scientifique ou des sciences sociales et humaines non couvert dans les autres UE à l'ENSAI (communication interculturelle, développement durable, géopolitique, histoire de l'art, l'initiation à l'entrepreneuriat, philosophie, psychologie, physique...),
- Ou s'initier ou se perfectionner dans un domaine artistique (cinéma, dessin, musique, peinture, théâtre...).

**Principales notions abordées :**

Spécifique à chaque option

**Références bibliographiques :**

Sans objet

UE 2-5 – Semestre 1

## DÉVELOPPEMENT DURABLE/RSE ET L'ÉTHIQUE

<i>Enseignant</i>	: Divers intervenants
<i>Nombre d'ECTS</i>	: 1
<i>Répartition des enseignements</i>	: 24 h de cours
<i>Langue d'enseignement</i>	: Sans objet
<i>Logiciels</i>	: Sans objet
<i>Documents pédagogiques</i>	: Sans objet
<i>Prérequis</i>	: Sans objet

### Modalités d'évaluation :

À définir

### Acquis d'apprentissage (objectifs) :

À définir

### Principales notions abordées :

À définir

### Contenu de la matière

À définir

### Références bibliographiques :

À définir

Bonus

**SPORTS***Sports**Enseignant*

: Divers intervenants

**Prise en compte dans la scolarité :**

La participation à une activité sportive peut donner lieu à l'attribution d'un bonus (non cumulable) 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 Ensai. Pour y prétendre, les élèves concernés doivent remplir les 3 conditions suivantes:

- pratiquer régulièrement une activité sportive et participer aux compétitions liées dans le but de représenter l'Ensai ;
  - 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.

**Acquis d'apprentissage (objectifs) :**

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

**Principales notions abordées :**

10 activités sportives sont proposées par l'école :

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

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

Matière 2AOUT01

## R-SHINY

*Enseignant* : Mathieu MARBAC (Ensaï)

*Répartition des enseignements* : 3h de cours

*Langue d'enseignement* : Français

*Logiciels* : Sans objet

*Documents pédagogiques* : Sans objet

*Prérequis* : Programmation en R (1A)

### Modalités d'évaluation :

Sans objet

### Acquis d'apprentissage (objectifs) :

Savoir mettre en place une interface Shiny pour présenter des résultats ou effectuer des calculs/simulations simples

### Principales notions abordées :

Initiation au développement d'application R-Shiny

### Références bibliographiques :

Sans objet

Matière 2AOUT02

## VISUAL BASIC APPLICATION

<i>Enseignant</i>	: Steven GOUICHOUX (Insee)
<i>Répartition des enseignements</i>	: 6h de cours
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: Sans objet
<i>Documents pédagogiques</i>	: Sans objet
<i>Prérequis</i>	: Les concepts algorithmiques abordés dans le cours algorithmique et programmation de 1ère année doivent être maîtrisés.

### Modalités d'évaluation :

Sans objet

### Acquis d'apprentissage (objectifs) :

Découvrir et comprendre l'environnement et la logique de développement en VBA pour Excel. Être capable de développer des applications en VBA s'appuyant sur Excel. Savoir mobiliser l'information nécessaire au développement en VBA pour Excel.

### Principales notions abordées :

1. L'éditeur VBA : les fenêtres et leurs éléments
2. Les objets Visual Basic Application : classeur, feuille, cellule, graphique
3. Le langage de programmation : variables et constantes, structures conditionnelles et boucles, procédures et fonctions
4. Traitement des chaînes de caractères
5. Le code : compilation, exécution et débogage
6. L'utilisation des principales formules d'Excel dans VBA pour le traitement de données
7. Les boîtes de messages
8. Les formulaires et leur interaction avec Excel
9. Manipulation du système de fichiers Windows
10. Manipulation de fichiers textes : création, lecture, écriture

Une première demi-journée, à la frontière entre cours magistral et TP, permettra aux élèves de se familiariser avec l'interface, les bonnes pratiques, les concepts de base, autour d'exemples concrets. Des exercices d'application seront alors proposés aux élèves.

Une deuxième demi-journée, quelques semaines plus tard, sera consacrée à la correction des exercices, en insistant sur toutes les bonnes pratiques du langage, sur sa puissance et sa simplicité.

### Références bibliographiques :

- Débuter en VBA Excel : <https://openclassrooms.com/courses/analysez-des-donnees-avec-excel/premiers-pas-en-vba>
- FAQ VBA : <http://excel.developpez.com/faq/?page=VBA>
- Support de cours : <http://bidou.developpez.com/article/VBA/>

Matière 2AOUT03

## COMPLEMENTS DE SAS

<i>Enseignant</i>	: David AUDENAERT
<i>Répartition des enseignements</i>	: 9h de TP
<i>Langue d'enseignement</i>	: Français
<i>Logiciels</i>	: SAS
<i>Documents pédagogiques</i>	: Sans objet
<i>Prérequis</i>	: SAS et langage SQL

### Modalités d'évaluation :

Sans objet

### Acquis d'apprentissage (objectifs) :

L'objectif du cours est de savoir automatiser et paramétrer des traitements à partir du macro langage Sas : créer et manipuler des macro variables, macro programmes et macro fonctions.

### Principales notions abordées :

1. Définition du macro langage
2. Éléments du macro langage : macro variables, macro fonctions et macro programmes.
3. Étape data et macro langage
4. Stockage des macro programmes
5. Macro langage et procédure SQL

### Références bibliographiques :

- Decourt O, SAS l'essentiel, Dunod
- Ringuedé S, SAS, Pearson
- Forums et sites : support.sas.com, developpez.net, ...
- Chauvet-Peyrard Axelle, Polycopié ENSAE en version interactive : [http://acp.coursinfostat.free.fr/DOC\\_SAS/SAS4\\_Macro.htm](http://acp.coursinfostat.free.fr/DOC_SAS/SAS4_Macro.htm)

Matière 2AOUT02

## HYGIENE NUMERIQUE

*Enseignant* : Mathieu Goessens

*Répartition des enseignements* : 12h de cours

*Langue d'enseignement* : à remplir

*Logiciels* : à remplir

*Documents pédagogiques* : à remplir

*Prérequis* : à remplir.

### Modalités d'évaluation :

A remplir

### Acquis d'apprentissage (objectifs) :

A remplir

### Principales notions abordées :

A remplir

### Références bibliographiques :

A remplir