



Programme des enseignements de 2^e année

ANNÉE SCOLAIRE 2019 / 2020



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

PROGRAMME DES ENSEIGNEMENTS DE 2^e ANNÉE

ANNÉE SCOLAIRE 2019/2020

Table des matières

Présentation générale des enseignements.....	5
1 Les grands domaines d'enseignement.....	9
Tableau récapitulatif des enseignements 2A	17
Enseignements du premier semestre	19
Régression linéaire	21
Modèle linéaire généralisé et régression sur variables catégorielles	22
Séries temporelles 1	23
Apprentissage supervisé	24
Théorie des sondages.....	26
Droit et éthique des données / Sécurité des données.....	28
Complément d'informatique.....	29
Projet informatique.....	30
Remise à niveau en informatique	31
Modélisation microéconomique	32
Econométrie appliquée	33
Stage opérateur attachés.....	35
Stage opérateur ingénieurs.....	37
Anglais	38
Enseignements du deuxième semestre : Tronc commun	41
Projet statistique	43
Introduction aux modèles de durée.....	44
Statistique non paramétrique	45
Sondages 2 : Traitement de la non-réponse dans les enquêtes	46
Chaînes de Markov.....	48
Calculs Bayésiens.....	49
Microéconométrie appliquée.....	50
Macroéconométrie appliquée.....	52
Programmation orientée objet avec Java	53
C + + pour la statistique computationnelle	54
Outils pour le Big Data.....	56
Enseignements de deuxième semestre : cours électifs (pré-spécialisation UE 2-09).....	59
Modèles avancés de régression	60
Martingales et processus de Lévy	62
Méthodes de Rééchantillonnage	63

Séries temporelles 2	65
Statistique mathématique	66
Sondages 3 : Echantillonnage à plusieurs degrés	67
Statistique spatiale.....	69
Economie du risque	71
Théorie de la croissance économique	72
Economie Industrielle Empirique.....	74
Economie Industrielle	75
Démographie	76
Economie financière	78
Programmation avancée en R.....	80
Technologies mobiles	81
Traitement du signal	82
Conception de logiciel.....	84
Data Challenge	86
DataVisualisation	87
Projets professionnels, enseignements optionnels.....	89
Projet professionnel ingénieurs	90
Projet professionnel attachés	91
Participation aux activités associatives.....	92
Sport.....	94
Cours d'ouverture	95
Langues optionnelles	96
Forum des entreprises	97
Rendez-vous Carrières	98
R-Shiny	99
Visual Basic Application	100
Compléments de SAS	102
Libre Office Basic.....	103
Enseignements spécifiques aux élèves attachés.....	104
Module spécifique des attachés	105

Présentation générale des enseignements

Créée il y a 20 ans, l'Ensaï est l'une des deux grandes écoles d'ingénieurs avec l'Ensaë 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'Ensaï 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 élèves ingénieurs et en deux ans pour les élèves fonctionnaires (attachés statisticiens de l'INSEE) avec, pour ces derniers, la possibilité d'obtenir un master mention mathématiques appliquées – statistique, parcours « évaluation et décision publiques », coaccrédité avec les Universités de Rennes 1 et Rennes 2, l'Insa de Rennes, l'ENS de Rennes, Agrocampus Ouest..

Durant les deux premières années de scolarité à l'Ensaï, les élèves ingénieurs et les élèves fonctionnaires suivent en commun la majorité des enseignements. Au second semestre de 2^e année, les parcours se différencient. Les élèves ingénieurs ont des enseignements qui les renforcent dans les compétences d'ingénieur statisticien et les préparent aux filières de 3^e année. Les élèves fonctionnaires reçoivent une formation plus orientée vers les connaissances utiles au statisticien public.

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 mathématique oblige à de premiers enseignements fondamentaux distingués au premier semestre selon le cursus antérieur. Ainsi les élèves venant de la voie mathématique (concours communs polytechniques, L3 math...) ou des IUT Stid ont un enseignement renforcé en économie au 1^{er} semestre pour rattraper leur retard par rapport aux élèves venant de la voie « économie ». De façon symétrique, les élèves 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 mathématique 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 de un à deux mois conclut cette 1^{re} année : stage de découverte de la statistique publique pour les attachés stagiaires et stage opérateur pour les élèves 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^e année (élèves ingénieurs) ou leur entrée dans le monde professionnel (élèves fonctionnaires). Pour les élèves 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^e année, tous les élèves ingénieurs choisissent une filière de spécialisation dans laquelle interviennent de nombreux professionnels et ponctuée par un stage de 6 mois. Pour les élèves fonctionnaires, c'est le choix d'une option dans le parcours « évaluation et décision publiques » du master mention mathématiques appliquées – statistique, selon une approche métier : études statistiques, méthodologie statistique ou traitements informatiques. Il peut être suivi, sous conditions, directement à la suite de la 2^e année ou en formation continue.

LES SIX FILIERES DE SPECIALISATION DE 3^E ANNEE

Gestion des risques et ingénierie financière

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.

Marketing quantitatif et revenue management

Cette filière forme à la fois des data scientists ayant une très forte compréhension des enjeux métiers du marketing et des revenue managers aux capacités analytiques et quantitatives très élevées. L'Ensaï est d'ailleurs la seule école d'ingénieurs française à proposer un cursus complet dédié au revenue/yield management.

Statistique pour les sciences de la vie

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.

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 élèves 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...

Statistique et 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.

Ingénierie statistique des territoires et de la 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 élèves qui souhaitent faire de la recherche théorique ou appliquée après l'Ensaï 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'Ensaï, 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 – le Crest –, de bénéficier d'un encadrement personnalisé par un « tuteur » spécialiste du domaine dans lequel l'élève 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 élèves titularisés comme attachés statisticiens de l'Insee peuvent obtenir un master en « évaluation et décisions 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'Ensaï,
- Décalée, c'est-à-dire de façon discontinue au cours de leurs premières années de fonction.

Le master mention mathématiques appliquées-statistique est co-accrédité avec les Universités de Rennes 1 et Rennes 2, l'Insa de Rennes, l'ENS de Rennes, Agrocampus Ouest. Le parcours « évaluation et décisions publiques » comporte trois colorations au choix : statistiques et traitement des données, méthodologie de la statistique publique ou études statistiques.

1 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 de mathématiques nécessaires à une mise à niveau des étudiants n'ayant pas fréquenté 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 élèves réalisent également plusieurs projet 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'élève et les choix, par une initiation aux processus stochastiques, des cours de méthodes de régression non paramétrique, ré-échantillonnage, 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 élèves de mettre en œuvre sur des données réelles un large éventail des techniques étudiées au cours des deux premières années. Les élèves 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 1^{re} année. L'algorithmique est introduit dans un premier temps avec les notions algorithmiques de base. Les

étudiants issus des voies mathématique et IUT Stid bénéficient en outre d'un cours portant sur l'algorithmique et la complexité des algorithmes étudiés. Dans un deuxième temps, une introduction à la programmation orientée objet est abordée pour tous avec le langage Python. Le lien avec les enseignements de statistique est notamment effectuée 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 élèves réalisent dès le début de la deuxième année un projet dont l'objectif est de mettre en application les enseignements reçus en 1^{re} 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 Basic, 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 élèves 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'élèves ayant de bonnes connaissances en sciences économiques et sociales et un public d'élèves débutants ou n'ayant eu qu'une première initiation dans cette discipline. Pour les premiers, l'École propose un projet 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. L'enseignement de première année distingue un cours selon le statut d'ingénieur (principes de gestion des organisations) ou fonctionnaire (comptabilité nationale).

En seconde année, le cours d'économétrie du tronc commun vise à donner aux élèves les méthodes de validation empirique des modèles théoriques. Ce cours peut être complété au 2^e 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.

HUMANITES

Communication et gestion de projet

L'apprentissage des techniques de communication est progressif, basé sur la pratique et le conseil personnalisé. Il permet aux élèves de l'École de communiquer efficacement dans les situations les plus diverses de la vie sociale et professionnelle. Les élèves sont sensibilisés aux techniques de communication écrites et orales dans le cadre du projet de statistique descriptive en 1^{re} année, puis développent ces compétences en travaillant sur les projets qui jalonnent leur scolarité. En 1^{re} année, ils suivent ainsi un cours de techniques rédactionnelles et de gestion de projets, en parallèle du projet de statistique descriptive. Des coachs en communication interviennent directement auprès de chaque groupe d'élèves en 2^e année. En 3^e année, des simulations d'entretien d'embauche préparent les élèves à leur entrée sur le marché du travail.

Anglais

L'étude de l'anglais est obligatoire tout au long de la scolarité. Les élèves sont répartis en groupes de niveau. Un niveau B2 du CECRL¹ est obligatoire pour l'obtention du diplôme d'ingénieur. Des modules de préparation au TOEIC sont organisés toutes les années.

Cours libres optionnels

Les élèves en première et deuxième année doivent choisir 2 enseignements par année parmi les enseignements suivants :

Langues optionnelles : Les langues proposées sont l'allemand, le chinois, l'espagnol, l'italien, le japonais et le russe. Elles peuvent être étudiées du niveau débutant jusqu'au niveau perfectionnement. Pour ces langues optionnelles, **l'inscription aux deux semestres est obligatoire**.

Cours d'ouverture : D'autres options sont organisées chaque année. Elles visent à développer une culture générale dans un domaine scientifique ou des sciences sociales non couvert à l'Ensay (*philosophie, histoire, psychologie, sociologie, physique...*), à effectuer des travaux d'initiation dans un domaine artistique (*dessin, peinture, musique...*) ou à valoriser l'engagement citoyen des étudiants (participation à des activités associatives, à des projets avec des intervenants non statisticiens...). Au second semestre, le cours de physique doit obligatoirement être suivi (en première ou deuxième année) par les élèves ingénieurs des filières économiques et Stid.

¹ Ce niveau correspond au score minimal de 785 au TOEIC .

Enseignements de 2^e année Volumes horaires enseignés

	Total	Cours	TD	ECTS
UE 2-01 : MODELISATION STATISTIQUE				
Régression linéaire	33	15	18	3
Modèle linéaire généralisé et régression sur variables catégorielles	33	21	12	3
Séries temporelles 1	33	18	15	3
<i>Total UE 2-01</i>	99	54	45	9
UE 2-02 : COLLECTE ET APPRENTISSAGE				
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
UE 2-03 : INFORMATIQUE POUR LA DATA SCIENCE				
Compléments d'informatique	18	6	12	2
Projet Informatique	36	-	36	4
Remise à niveau en informatique (AST seulement)	21	21	-	1
<i>Total UE 2-03</i>	54	6	48	6
UE 2-04 : MODELISATION ECONOMIQUE				
Modélisation microéconomique	21	15	6	2
Econométrie appliquée	33	21	12	3
<i>Total UE 2-04</i>	54	36	18	5
UE 2-05 : HUMANITES ET STAGE OPERATEUR				
Stage opérateur (hors AST)				1
Cours libre optionnel n°1	21	-	21	1
Anglais 1 ^{er} semestre	30	-	30	2
<i>Total UE 2-05</i>	51	-	51	4
TOTAL 1^{ER} SEMESTRE	324	141	183	30
	Total	Cours	TD	ECTS
UE 2-06 : PROJET STATISTIQUE				
Projet statistique	33	-	33	5
<i>Total UE 2-06</i>	33	-	33	5
UE 2-07 : COMPLEMENTES DE MODELISATION STATISTIQUE				
Introduction aux modèles de durée	21	12	9	2
Statistique non paramétrique (ing.)	21	12	9	2
Traitement données manquantes enquêtes (att.)	21	12	9	2
Chaînes de Markov	21	12	9	2
<i>Total UE 2-07</i>	63	36	27	6
UE 2-08 : METHODES COMPUTATIONNELLES				
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	18	-	2
Outils pour le Big Data	24	24	-	2
<i>Total UE 2-08</i>	81	63	18	8
UE 2-09 : PRE-SPECIALISATION				
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
Cours électif n°5	21	21	-	2
<i>Total UE 2-09</i>	105	105	-	10
UE 2-10 : COURS LIBRE OPTIONNEL, COMMUNICATION				
Cours libre optionnel n°2	21	-	21	1
Projet professionnel et gestion de projets	20	-	20	
<i>Total UE2-10</i>	41	-	41	1
TOTAL 2ND SEMESTRE	323	204	140	30
TOTAL ANNEE	647	345	323	60

Enseignements de 2^e année
2nd semestre – UE 2-09
Les cours électifs

Les élèves doivent choisir 5 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.

COURS ELECTIFS DE 2^e ANNEE

Enseignements		Volume horaire				
		Cours	TD	TP	Projet	Total
ECONOMIE	Economie du risque	21				21
	Théorie de la croissance	21				21
	Economie industrielle appliquée	21				21
	Economie industrielle	15	6			21
	Démographie	21				21
	Economie financière	21				21
INFORMATIQUE	Programmation avancée en R	21				21
	Technologie mobile	6		15		21
	Traitement du signal	8	13			21
	Conception de logiciel	9		12		21
	Data Challenge	21				21
	DataVisualisation	21				21
STATISTIQUE	Modèles avancés de régression	21				21
	Martingales et processus de Lévy	12	9			21
	Méthodes de rééchantillonnage (bootstrap)	21				21
	Séries temporelles 2	21				21
	Statistique mathématique	21				21
	Statistique spatiale	3		18		21

Conseils concernant le choix des cours électifs et des options

Des cours du Tronc Commun sont à choisir:

- Un langage objet entre C++ et Java. Les enseignements de C++ et Java ne peuvent pas être suivis en même temps. Java est très fortement recommandé pour les élèves attachés (et obligatoire pour ceux désirant s'orienter vers une carrière informatique).
- R avancé a comme prérequis C++, conception de logiciel et technologies mobiles ont comme prérequis Java.

Un cours d'approfondissement en statistique/économétrie : calculs bayésiens ou micro/macro-économétrie. Micro/macro-économétrie est très fortement recommandée pour les élèves attachés.

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

Le choix de 5 cours électifs (un par bloc) vise à approfondir vos connaissances et à révéler vos préférences avant votre spécialisation de 3^{ème} année (filières ou parcours 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. 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 3^{ème} année. Les cours de « Data Visualisation » et « Data Challenge » sont adaptés à l'ensemble des spécialisations, les compétences mobilisées étant transversales. Le cours de statistique mathématique est très conseillé pour les élèves désirant s'orienter vers une formation par la recherche.

Enseignements		Filières					Master EDP			
		GDR	MQRM	SID	SSV	GS	ISTS	MSP	ES	STD
TC	Java		+	++	+		+	++	++	++
	C++	++	+		+	++	+			
	Calculs bayésiens	+	+	++	++	++	+			
	Micro et macro-économétrie	+	+				+	++	++	++
ECONOMIE	Economie du risque	++	+		+	+	++	++	++	+
	Economie industrielle appliquée		+		+		++		+	
	Economie industrielle		+		+		++	+	++	+
	Démographie							++	++	++
	Economie financière	++	+			+			+	
INFORMATIQUE	R avancé	+	+			+	+			
	Technologies mobiles		+	++						++
	Traitement du signal			+	+	+		+		+
	Conception de logiciel			++				+	+	++
	Data Challenge	+	+	+	+	+	+	+	+	+
	Data Visualisation	+	+	+	+	+	+	+	+	+
STATISTIQUE	Modèles avancés de régression	+	+	+	++	++	++	++		
	Martingales et processus de Lévy	++			+	+				
	Méthodes de rééchantillonnage (Bootstrap)	++	++	++	++	++	+	++	+	+
	Séries temporelles 2	++	++		+	+	+	+	+	
	Statistique mathématique	+			+	++		+		
	Echantillonnage à plusieurs degrés (sondages 3)				+		+	++	++	++
	Cartographie et statistique spatiale		+				++	+	++	+

Bloc 1 : statistique mathématique, économie du risque, data challenge

Bloc 2 : échantillonnage à plusieurs degrés (sondages 3), séries temporelles 2, économie industrielle, traitement du signal

Bloc 3 : martingales, modèles avancés de régression, DataViz

Bloc 4 : Méthodes de rééchantillonnage (bootstrap), technologies mobiles, démographie, économie financière

Bloc 5 : r avancé, conception de logiciel, économie industrielle appliquée, statistique spatiale

Tableau récapitulatif des enseignements 2A

Code	Matière	Enseignant	Correspondant
2AECO01	Modélisation microéconomique	Yutec SUN	Yutec SUN
2AECO02	Econométrie appliquée	Ronan LE SAOUT	Ronan LE SAOUT
2AECO03	Microéconométrie appliquée	Eric DELATTRE	Vincenzo CAPONI
2AECO04	Economie du risque	Olivier RENAULT	Vincenzo CAPONI
2AECO05	Théorie de la croissance économique	Vincenzo CAPONI	Vincenzo CAPONI
2AECO06	Economie industrielle empirique	Yutec SUN	Yutec SUN
2AECO07	Economie industrielle	Frédéric LOSS	Vincenzo CAPONI
2AECO08	Démographie	Laurent DI CARLO	Vincenzo CAPONI
2AECO09	Economie financière	Samuel DANTHINE	Samuel DANTHINE
2AECO10	Macroéconométrie appliquée	Stéphane AURAY	Stéphane AURAY
2AHUM01	Anglais	Divers intervenants	Todd DONAHUE / Emily BURMEISTER
2AHUM02	Droit et éthique des données / Sécurité des données	Stéphane ASTIER / Kamel GADOUCHE	Ronan LE SAOUT
2AHUM03	Projet professionnel ingénieur	Divers intervenants	Patrick GANDUBERT
2AHUM04	Projet professionnel attaché	Divers intervenants	Brigitte GELEIN
2AHUM05	Sport	Divers intervenants	Julien LEPAGE
2AHUM06	Stage	Divers intervenants	Patrick GANDUBERT/Ronan LE SAOUT
2AINF01	Compléments d'informatique	Rémi PÉPIN	Romarc GAUDEL
2AINF02	Projet informatique	Divers intervenants	Romarc GAUDEL & Rémi PÉPIN
2AINF03	Remise à niveau en informatique	Rémi PÉPIN	Rémi PÉPIN
2AINF04	POO avec Java	Olivier LEVITT	Romarc GAUDEL
2AINF05	C++ pour la statistique computationnelle	Thomas GUYET	Hong-Phuong DANG
2AINF06	Outils pour le Big Data	Arthur KATOSSKY/Rémi PÉPIN	Arthur KATOSSKY/Rémi PÉPIN
2AINF07	Programmation avancée en R	Matthieu MARBAC	Matthieu MARBAC
2AINF08	Technologies mobiles	Olivier LEVITT	Hong-Phuong DANG
2AINF09	Traitement du signal	Hong-Phuong DANG	Hong-Phuong DANG
2AINF10	Conception de logiciel	Mohamed GRAIET	Mohamed GRAIET
2AINF11	Data Challenge	Salima EL KOLEI	Salima EL KOLEI
2AINF12	Datavisualisation	Arthur KATOSSKY	Ronan LE SAOUT
2ASTA01	Régression linéaire	Valentin PATILEA	Valentin PATILEA
2ASTA02	Modèle linéaire généralisé et régression sur variables catégorielles	François COQUET	François COQUET
2ASTA03	Séries temporelles 1	Marie DU ROY DE CHAUMARAY	Marie DU ROY DE CHAUMARAY
2ASTA04	Apprentissage supervisé	Brigitte GELEIN	Brigitte GELEIN
2ASTA05	Théorie des sondages	LESAGE Eric	Guillaume CHAUVET
2ASTA06	Projet statistique	Divers intervenants	Khaled LARBI
2ASTA07	Modèles de durée	Valentin PATILEA	Valentin PATILEA
2ASTA08	Statistique non paramétrique	Marian HRISTACHE	Marian HRISTACHE
2ASTA09	Traitement de la non réponse dans les enquêtes	CHAUVET Guillaume	CHAUVET Guillaume
2ASTA10	Chaînes de Markov	Adrien SAUMARD	Adrien SAUMARD
2ASTA11	Calculs bayesiens	Marie DU ROY DE CHAUMARAY	Marie DU ROY DE CHAUMARAY
2ASTA12	Modèles avancés de régression	Gilles STUPFLER	Gilles STUPFLER
2ASTA13	Martingales et processus de Lévy	Benoit CADRE	Lionel TRUQUET
2ASTA14	Méthodes de rééchantillonnage	VIMOND Myriam	VIMOND Myriam
2ASTA15	Séries temporelles 2	Purevdorj TUVAANDORJ	Purevdorj TUVAANDORJ
2ASTA16	Statistique mathématique	Lionel TRUQUET	Lionel TRUQUET
2ASTA17	Echantillonnage à plusieurs degrés	CHAUVET Guillaume	CHAUVET Guillaume
2ASTA18	Statistique spatiale	Arthur KATOSSKY	Arthur KATOSSKY

2AOUT01	R-Shiny	Matthieu MARBAC	Matthieu MARBAC
2AOUT02	Visual Basic Application	Steven GOUICHOUX	Mohamed GRAIET
2AOUT03	Compléments de SAS	A déterminer	A déterminer
2AOUT04	Libre Office Basic	Samuel TOUBON	Mohamed GRAIET

La durée et les modalités d'examen sont données à titre indicatif. La direction des études se réserve le droit de les modifier. Les élèves seront prévenus par les enseignants en début de cours.

Enseignements du premier semestre

UE 2-01 – Matière 2ASTA01 – Semestre 1

Régression linéaire

Linear regression

Cours : 15h • TD : 12h • TP : 6h

Enseignant : Valentin PATILEA (Ensayi)

Correspondant : Valentin PATILEA

Objectif de la matière

Comprendre la construction d'un modèle de régression linéaire. Être capable de mener une étude complète de régression linéaire, d'avoir une lecture critique des sorties logicielles, d'effectuer une analyse constructive des résidus, d'identifier d'éventuels écarts au modèle, de les corriger, de faire une sélection de variables pertinente. Utilisation des packages du logiciel R.

Contenu de la matière

1. Introduction : rappels d'algèbre et de régression linéaire simple
2. Le modèle de régression linéaire multiple
3. Le modèle de régression linéaire multiple sous hypothèse gaussienne
4. Régions de confiance et tests d'hypothèses
5. Détection (et correction) des écarts au modèle - Analyse des résidus, effet de levier et mesure d'influence
6. Compléments de régression linéaire (en fonction du temps disponible) : bootstrap, pénalisation, tests d'adéquation, etc.

Documents pédagogiques

Polycopié de cours

Pré-requis

Lois de probabilité usuelles ; vecteurs gaussiens ; régions de confiance ; tests statistiques ; algèbre matricielle ; projection orthogonale

Contrôle des connaissances

Un examen final

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 (2^e éd .), 2012, Dunod*
- CORNILLON P.-A., MATZNER-LOBER E., *Régression avec R, 2010, Springer.*
- DODGE R., ROUSSON V., *Analyse de régression appliquée (2e éd.), 2004, Dunod.*
- KLEINBAUM D. G. et al., *Applied regression analysis and multivariate methods (4th ed.), 2008, Cengage Learning.*
- TOMASSONE R; et al., *La régression (2e éd.), 1992, Masson*

Langue d'enseignement

Français

UE 2-01 – Matière 2ASTA02 – Semestre 1

Modèle linéaire généralisé et régression sur variables catégorielles

Generalized Linear Models and Regression on Categorical Data

Cours: 21h • TD: 12h

Enseignant : François COQUET (Ensayi)

Correspondant : François COQUET

Objectif de la matière

A l'issue de cet enseignement, les élèves devront en fonction du problème posé et de la nature des données disponibles, savoir choisir un type de modèle, savoir le construire, le mettre en œuvre sous SAS et en interpréter les résultats.

Contenu de la matière

Dans de nombreux domaines d'études la variable expliquée est catégorielle, c'est-à-dire qu'elle est représentée par un nombre fini ou dénombrable de modalités. Ce cours fournira donc une méthode de régression sur variables catégorielles. Le modèle linéaire gaussien est clairement inadapté à l'étude statistique de ces variables, et nous introduirons le modèle linéaire généralisé, en l'illustrant de plusieurs exemples. Pour la partie technique, nous étudierons principalement le modèle logistique (variables binaires), ainsi que les modèles polytomiques ordonné et multinomial, ainsi que la régression de Poisson pour les variables à plus de deux catégories. Enfin, nous ferons une incursion dans le domaine des données dépendantes en introduisant la notion de modèles mixtes..

Pré-requis

Cet enseignement demande que tous les élèves maîtrisent les cours de modèle de régression et statistique 3 (statistique inférentielle).

Contrôle des connaissances

Un examen écrit d'une durée de deux heures (+ éventuellement rapport de TP).

Références bibliographiques

- AGRESTI A., *Categorical Data Analysis (3rd ed.)*, WILEY & Sons, 2012
- ALLISON, P., *Logistic regression using SAS : theory and application(2nd ed.)*, Cary, NC : SAS Institute Inc, 2012
- HOSMER D.W., LEMESHOW S., *Applied Logistic Regression (3rd ed.)*, WILEY & Sons, 2013
- KLEINBAUM, D., KLEIN, M., *Logistic regression : a self-learning text (3rd ed.)*, Springer, 2010
- STOKES MAURA E., DAVIS, CHARLES S., KOCH, GARY G., *Categorical Data Analysis Using the SAS System (3rd ed.)*, Cary, NC : SAS Institute Inc, 2012
- Tufféry, S., *Modélisation prédictive et apprentissage statistique avec R (2^e éd.)*, Technip, 2017

Langue d'enseignement

Français.

UE 2-01 – Matière 2ASTA03 – Semestre 1

Séries temporelles 1

Times Series 1

Cours: 18h • TD : 12h • TP : 6h

Enseignant : Marie DU ROY DE CHAUMARAY (Ensay)

Correspondant : Marie DU ROY DE CHAUMARAY

Objectif de la matière

L'objectif de ce cours d'introduction aux séries temporelles est de présenter les méthodes classiques d'analyse et de prévision de séries d'observations corrélées dans le temps. Tout d'abord, on présentera les outils et notions de base pour étudier les séries temporelles univariées (tendance, saisonnalité, notion de stationnarité) en introduisant notamment les processus ARMA et la méthodologie de Box-Jenkins. Dans un second temps, on s'intéressera au problème de la non-stationnarité, avant de présenter différentes méthodes de prédiction. Les concepts de ce cours seront illustrés à l'aide de données réelles ou simulées. A la fin de cet enseignement, les étudiants devront être capables d'analyser des séries temporelles classiques à l'aide de logiciels spécialisés.

Contenu de la matière

1. Tendence, 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.

Prérequis

Notions de base en Probabilités et Statistique inférentielle.

Contrôle des connaissances

Partiel et Examen final sur papier

Références bibliographiques

- Box G., Jenkins G., Reinsel G., *Time Series Analysis. Forecasting and Control*, Prentice Hall, 1994, 3^e Edition
- Hamilton J, *Time Series*, Princeton University Press, 1995.
- Shumway R. and Stoffer D. , *Time series analysis and its applications*, Springer 2006.

Langue d'enseignement

Français

UE 2-02 – Matière 2ASTA04 – Semestre 1

Apprentissage supervisé

Supervised learning

Cours : 18h • TP : 9h

Enseignant : Brigitte GELEIN (Ensay)

Correspondant : Brigitte GELEIN

Objectif de la matière

A l'issue de cet enseignement, les élèves devront être capables de mettre en œuvre et d'analyser les résultats des méthodes de discrimination et de segmentation par arbre.

Les applications de l'analyse discriminante et de la segmentation sont très nombreuses : scoring bancaire, marketing, aide au diagnostic ou au pronostic médical, reconnaissance de formes, prévision d'événements, etc.

Ce module présentera les fondements théoriques de ces méthodes et permettra leur mise en œuvre concrète lors de séances de TD et TP sur R. Elles consisteront en des études de cas réalisées sur des jeux de données réels pour lesquelles les élèves devront définir les méthodes à utiliser, les mettre en œuvre et en interpréter les résultats.

Contenu de la matière

- 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).

Documents pédagogiques

Polycopiés.

Pré-requis

Cet enseignement demande la maîtrise du cours de Statistique Exploratoire Multivariée 1A, de la notion de projection et d'optimisation (lagrangien).

Contrôle des connaissances

Un examen écrit.

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.

- CONFAIS 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.

Langue d'enseignement

Français

UE 2-02 – Matière 2ASTA05 – Semestre 1

Théorie des sondages

Sampling Theory

Cours : 18h • TD : 9h • TP : 3h

Enseignant : Eric LESAGE (Insee)

Correspondant : Guillaume CHAUVET

Objectif de la matière

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.

Contenu de la matière

Partie 1 : Echantillonnage en population finie

- Notations
- Estimation de Horvitz-Thompson
- Calcul de précision : estimateur de variance, intervalle de confiance.

Partie 2 : Méthodes d'échantillonnage

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

Partie 3 : Méthodes de redressement

- Estimateur par calage
- Applications : estimateur par la régression, estimateur post-stratifié
- Estimation d'un paramètre complexe

Documents pédagogiques

Support de cours.

Pré-requis

Cours de Probabilités, Introduction à la Statistique, Régression linéaire

Contrôle des connaissances

L'examen est un écrit. L'enseignement donne lieu à contrôle continu.

Compétences acquises

Choisir une stratégie d'échantillonnage et la mettre en œuvre dans des situations simples.

Produire les estimateurs et les indicateurs de précision associés.

Mobiliser de l'information auxiliaire pour redresser un estimateur.

Travail personnel

30h

Références bibliographiques

- ARDILLY P., *Les Techniques de Sondage (nouv. éd.)*, Paris, Technip, 2006
- COCHRAN W.G., *Sampling Techniques*, 3^e édition, New York, Wiley, 1977
- SÄRNDAL, C-E., SWENSSON, B, WRETMAN, J., *Model Assisted Survey Sampling*, Springer-Verlag, 2003
- TILLÉ Y., *Sampling Algorithms*, Wiley, 2006

Langue d'enseignement

Français

UE 2-02 – Matière 2AHUM02 – Semestre 1

Droit et éthique des données / Sécurité des données

Ethics and Law in Data and Analytics / Data security

Cours : 9h •

Enseignant : Stéphane ASTIER / Kamel GADOUCHE

Correspondant : Ronan LE SAOUT

Objectif de la matière

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 ».

En s'appuyant sur l'actualité (géopolitique, technologique, économique), ce séminaire d'une journée 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 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. Ce cours 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,...

Contrôle des connaissances

QCM en fin de séminaire.

Langue d'enseignement

Français

UE 2-03 – Matière 2AINF01 – Semestre 1

Complément d'informatique

Computer Science Project

Cours : 6h – TP : 12h

Enseignant : Rémi Pépin (Ensay)

Correspondant : Romaric GAUDEL

Objectif de la matière

Ce cours vient compléter les connaissances en informatique afin d'armer les étudiants pour le projet informatique.

Contenu de la matière

Le coeur du cours porte sur l'accès et la manipulation de données depuis un programme. Seront abordés aussi bien l'accès à une base de donnée relationnelle, que la lecture et l'écriture de fichiers dans des formats non-tabulaires (JSON, XML et HTML par exemple), ou encore l'accès à des données par le biais d'API-web. Tous ces aspects sont fréquemment rencontrés dans des contextes web et/ou Big Data.

Sera aussi abordé le principe de la gestion de version pour des codes partagés.

Langages / logiciels utilisés

Python, Python-SQL, JSON, XML, HTML, API REST, git.

Pré-requis

Connaître les bases d'algorithmique, les concepts objet, le langage Python, et le langage SQL.

Contrôle des connaissances

Aucun

Langue d'enseignement

Français.

UE 2-03 – Matière 2AINF02 – Semestre 1

Projet informatique

Computer Science Project

Suivi : 21h – Immersion : 3jours, dont 6h de suivi – travail en autonomie : 30h

Enseignant : Divers intervenants

Correspondant : Romaric GAUDEL & Rémi Pépin

Objectif de la matière

Le projet permet de mettre en pratique, sur un cas réel, les connaissances acquises dans les enseignements d'informatique. Il nécessite la mise en applications notamment des compétences

en gestion de projet (découpage du projet en sous-objectifs, suivi de ces objectifs, communication dans le groupe, partage de code / de documents, ...)

en programmation (conception de l'application, implémentation en langage objet, ...)

en manipulation de données (base de donnée relationnelles, formats arborescents, API web, ...)

Réalisé par groupe de 4 ou 5 étudiants sur un sujet proposé par l'encadrant, le projet démarre début septembre et les soutenances ont lieu fin novembre. En milieu de semestre une période dite d' « immersion » de 3 jours est consacrée au projet.

Documents pédagogiques

Notice.

Pré-requis

Connaître les bases d'algorithmique, les concepts objet, le langage Python, et le langage SQL. Suivre le cours complémentaire sur la manipulation de données.

Contrôle des connaissances

La notation sera à la fois collective et individuelle et portera sur le suivi de l'encadrant, le rapport rendu à mi-projet, le rapport final et la soutenance prévue début décembre.

Références bibliographiques

- JACOBSON I. (2004) *Object Oriented Software Engineering, A Use Case Driven Approach*, Addison Wesley (traduction française chez le même éditeur).
- C. CHRISTMENT, O. TESTE, M. TUFFERY, K. PINEL-SAUVAGNAT (2008), Bases de données relationnelles, Hermès Science.

Langue d'enseignement

Français.

UE 2-03 – Matière 2AINF03 – Semestre 1

Remise à niveau en informatique

Computer Science Project

Cours : 21h

Enseignant : Rémi Pépin (Ensay)

Correspondant : Rémi Pépin

Objectif de la matière

Remettre à niveau les étudiants qui n'auraient pas effectué leur première année à l'Ensay sur les concepts et langages informatiques vus en première année : bases de données relationnelles (en SQL), algorithmique et programmation orientée objet (Python et UML).

Compétences

- Effectuer des requêtes sur une base de données (en SQL)
- Concevoir une base de données adaptées à une application donnée.
- Programmer en Python.
- Concevoir (à l'aide d'UML) et implémenter (en Python) un logiciel composé de plusieurs classes d'objets réparties en divers bibliothèques.

Documents pédagogiques

Envoyés par courriel avant la rentrée (pour bases de données).

Pré-requis

Aucun

Contrôle des connaissances

À déterminer

Références bibliographiques

- UML : <http://openclassrooms.com/courses/debutez-l-analyse-logicielle-avec-uml>
- Python : <https://openclassrooms.com/courses/apprenez-a-programmer-en-python>

Langue d'enseignement

Français

UE 2-04 – Matière 2AECO01 – Semestre 1

Modélisation microéconomique

Microeconomic Theory

Cours : 15h • TD : 6h

Enseignant : Yutec SUN (Ensay)

Correspondant : Yutec SUN

Objectif de la matière

In this course, we aim to understand how consumer's demand and firm's supply decisions emerge as equilibrium outcome under more complex market conditions and what they imply for social welfare. We will develop our learning on the basis of Microeconomics 1 (Modélisation microéconomique 1). We will learn essential tools and concepts for the analysis such as game theory, dynamic models, uncertainty, externalities, and time preferences.

Contenu de la matière

1. Analysis of monopoly market
2. Firm competition and strategic interaction
3. Decision under risk and uncertainty
4. Externalities and welfare
5. Dynamic decisions and time preference

Pré-requis

The course will assume the knowledge provided by microeconomics 1. It will also require basic understanding of mathematical concepts such as continuity, differentiation, and vector space.

Contrôle des connaissances

A closed-book written exam

Références bibliographiques

- Hal Varian (2010). Introduction à la microéconomie, 6e édition, De Boeck.
- Hal Varian (2008). Analyse microéconomique, 3e édition, De Boeck.

Langue d'enseignement

Anglais.

UE 2-04 – Matière 2AECO02 – Semestre 1

Econométrie appliquée

Applied econometrics

Cours : 21h • TD : 12h

Enseignant : Ronan LE SAOUT (Ensay)

Correspondant : Ronan LE SAOUT

Objectif de la matière

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.

Le cours est organisé en trois grandes parties. La première partie présente les étapes d'une démarche économétrique, en distinguant le caractère descriptif et causal. La deuxième partie présente les problèmes liés à l'endogénéité des variables explicatives: leur origine et leurs effets sur le choix des outils d'inférence statistique (variables instrumentales...). La troisième partie présente une introduction à des méthodes spécifiques : évaluation des politiques publiques, données d'enquête, données de panel, données temporelles...

Principaux acquis de la formation : à l'issue du cours, l'étudiant saura

- 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.

Contenu de la matière

1. La démarche économétrique
2. Rappels sur les méthodes d'estimation
3. Les sources de l'endogénéité
4. Les variables instrumentales
5. Introduction à l'évaluation des politiques publiques, l'économétrie des données de panels, d'enquête

Documents pédagogiques

Disponible sous Moodle

Pré-requis

On mobilisera ici trois cadres d'inférence statistique, *i.e.* Moindres Carrés, Méthode des Moments et Maximum de Vraisemblance, et des raisonnements micro-économiques simples.

Contrôle des connaissances

Partiel et examen écrit avec documents.

Références bibliographiques

- ANGRIST, J. D. and J.-S. PISCHKE, *Mostly Harmless Econometrics, an empiricist companion*, Princeton University Press, 2008.
- CAMERON A.C. and P.K. TRIVEDI, *Microeconometrics. Methods and Applications*, Cambridge University Press, 2005.
- CREPON, B. et N. Jacquement, *Econométrie: méthode et applications*. De Boeck, 2010
- GREENE, W.H., *Econometric Analysis*. Prentice-Hall (7th ed), 2011.
- WOOLDRIDGE, J. M., *Econometric analysis of cross-section and panel data*, The MIT press (2nd rev. ed), 2010.

Langue d'enseignement

Français (cours et TD), Anglais (TD)

UE 2-05 – Matière 2AHUM06 – Semestre 1

Stage opérateur attachés

Professional internship for Insee civil servants

Durée : 4 à 8 semaines entre juin et août

Enseignant : Divers intervenants des Services statistiques publics français, européen et canadien

Correspondant : Ronan LE SAOUT

Objectif de la matière

- Découvrir et appréhender le service statistique public ;
- Réaliser un travail en mettant en application les enseignements reçus en 1^{re} année (utilisation des logiciels R ou SAS, statistique descriptive univariée, bivariée, analyses factorielles et classification ...etc).

Contenu de la matière

Le stage doit être conçu pour que les objectifs poursuivis soient atteints.

Objectif n°1 : découverte et appréhension du système statistique public

A l'issue de cette période, chaque élève doit avoir compris le fonctionnement de la structure d'accueil. Le stage comporte, à ce titre, différentes étapes d'observation :

1°) *observation des modes d'organisation du travail (centre(s) de décision, organigramme, intensité des lignes hiérarchiques, liens fonctionnels entre les différents services, ouverture sur l'extérieur...);*

2°) *observation des relations sociales (rôle de la Direction des ressources humaines, rôle des syndicats...) et de la problématique hygiène sécurité et risques psychosociaux au sein de la structure d'accueil (acteurs et dispositifs).*

Objectif n°2 : réalisation d'un travail nécessitant la mise en application des enseignements reçus en 1^{re} année

Ce travail aura été défini en amont par la structure d'accueil. Il constituera une participation concrète à la production de l'unité d'accueil.

Pré-requis

Aucun

Contrôle des connaissances

Le stage de découverte de la statistique publique est le seul moment de la scolarité au cours duquel la capacité d'intégration du stagiaire dans une équipe professionnelle est mesurée. Il a donc été décidé de donner plus de poids à l'évaluation par l'employeur, donc par le maître de stage.

Deux évaluations sont désormais prises en compte dans la note globale de stage :

A - Maître de stage : notation sur 20 valorisant l'intégration dans l'équipe, l'autonomie, l'investissement, les compétences techniques.... La note doit refléter le travail réel effectué par le stagiaire et sa capacité à travailler dans la statistique publique. Une note supérieure à

la moyenne signifie donc que le stagiaire est jugé apte à travailler dans la statistique publique.

B - Direction des études Ensay : notation sur 20 valorisant le rapport de stage sur les items suivants

- description de la structure d'accueil (activité...),
- étapes d'observation (organisation, relations sociales, hygiène et sécurité),
- travaux confiés et description des missions (partie rédigée de manière non technique),
- bilan professionnel et personnel,
- présentation et rédaction du rapport.

Le rapport de stage aura une longueur de 10 pages maximum. L'objectif est de juger de la capacité de synthèse de l'étudiant. Il n'est pas demandé une description technique des travaux conduits mais une description lisible par un lecteur non statisticien. La qualité technique des travaux menés est en effet évaluée dans la note du maître de stage.

La note finale du stage est la moyenne arithmétique de ces deux évaluations.

Pour les élèves internes, la note finale est la moyenne de la note de stage et de la note du rapport d'expérience professionnelle.

Références bibliographiques

Guide des stages des attachés 1A

Langue

Français, anglais ou autres langues de l'UE

UE 2-05 – Matière 2AHUM06 – Semestre 1

Stage opérateur ingénieurs

Professional internship

Durée : 4 semaines minimum entre début juin et fin août

Enseignant : Divers intervenants

Correspondant : Valérie CHRETIEN et Patrick GANDUBERT

Objectif

Le stage de Première année est un stage d'ouverture au monde professionnel, dit stage « opérateur ». Effectué en France ou à l'étranger, il a pour but :

- de découvrir le monde de l'entreprise dans sa dimension socio-organisationnelle ;
- d'identifier l'ensemble du processus de fabrication et de diffusion d'un produit ou d'un service de l'amont (études de besoins, conception...) à l'aval (mise sur le marché, études de satisfaction...).

S'il est effectué à l'étranger, ce stage a également pour but de découvrir une autre culture sociale et professionnelle et d'approfondir ses connaissances d'une langue étrangère. Faire son stage d'ouverture à l'étranger est fortement conseillé puisqu'il permettra de valider la **Période obligatoire à l'étranger** (POE).

Documents pédagogiques

Guide du stage 1A sur le Career center de l'école.

Contrôle des connaissances

Rapport de stage

UE 2-05 – Matière 2AHUM01 – Semestre 1

Anglais

English

Cours/TD : 30h au 1^{er} semestre (en cours libre optionnel au 2nd)

TOEIC blancs d'entraînement facultatif

Enseignant : Divers intervenants

Correspondant : Emily BURMEISTER

Objectif de la matière

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.

Compétences langagières visées

- Peut **comprendre** le contenu essentiel de sujets concrets ou abstraits dans un texte complexe, y compris une discussion technique dans sa spécialité.
- Peut **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.
- Peut **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.
- Possède **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.
- Montre un degré assez élevé de **contrôle grammatical**. Ne fait pas de fautes conduisant à des malentendus et peut le plus souvent les corriger lui/elle-même.
- Peut 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.
- Peut **prendre l'initiative de la parole** et son tour quand il convient et peut clore une conversation quand il le faut, encore qu'éventuellement sans élégance.
- Peut **faciliter la poursuite d'une discussion** sur un terrain familier en confirmant sa compréhension, en sollicitant les autres, etc.
- Peut 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.

Contenu de la matière

Pour les élèves ingénieurs qui n'ont pas passé ou qui n'ont pas réussi le TOEIC, les cours se concentrent en partie sur la préparation à cet examen. Les élèves attachés qui ne souhaitent pas passer le TOEIC suivent des cours dont l'objectif est d'améliorer l'expression et la compréhension orale. Tous les élèves dont le niveau le permet (score d'au moins 785 au TOEIC) choisissent deux thèmes. Voici quelques exemples des thèmes proposés par le passé : *Deba-*

ting, Travel, Current & Topical Affairs, English through Popular Music, British Humour et Minorities in America.

Les cours au 2nd semestre sont facultatifs pour les ingénieurs/attachés avec un niveau B2, le contenu reflète les souhaits des élèves. Les ingénieurs qui n'ont pas encore le niveau B2 CECR suivront obligatoirement un cours dédié à la préparation du TOEIC.

Pré-requis

Avoir passé le test de niveau pour les AST. Pour les cours thématiques, avoir certifié le niveau B2 CECR (obtention d'au moins 785 au TOEIC).

Contrôle des connaissances

La note finale prend en compte un TOEIC blanc ou le score TOEIC obtenu lors de la première année, la note moyenne des contrôles continus, la participation et l'assiduité. La note de contrôle continu est composée des interrogations et travaux divers définis par l'enseignant.

Références bibliographiques

- Azar, B., *Understanding and Using English Grammar*, New York: Longman, 1999.
- Council of Europe, Common European Framework of Reference for Languages <https://www.coe.int/en/web/common-european-framework-reference-languages/home>
- Lecomte, Stéphane, et. al, *La Grammaire au TOEIC et au TOEFL : Mode d'emploi*, Paris: Ophrys, 2008.
- Loughed, Lin, *Longman Preparation Series for the TOEIC Test*, New York: Longman, 2007.
- Loughed, Lin, *Préparation au Nouveau TOEIC*, Paris: Pearson Education France, 2007.
- *Oxford Preparation Course for the TOEIC*, Oxford: Oxford University Press.
- Rogers, Bruce, *Complete Guide to the TOEIC Test* (3rd ed.), Boston: Thomson, 2006.

Langue d'enseignement

Anglais

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

Enseignements du deuxième semestre : Tronc commun

UE 2-06 – Matière 2ASTA06 – Semestre 2

Projet statistique

Project in Statistics

Projet : 33h, dont 7h de suivi avec le tuteur et 3h de coaching

Enseignant : Divers intervenants

Correspondant : Khaled LARBI

Objectif de la matière

Les projets de statistique sont encadrés par des praticiens de la statistique et ont pour objectif l'application et l'approfondissement des connaissances statistiques acquises dans le cadre des enseignements. Les élèves, répartis par groupes de quatre, 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.

Contenu de la matière

Les sujets traités au cours des projets sont issus de domaines divers, mais présentent une forte composante statistique, de nature méthodologique et technique.

Exemples de sujets traités :

- Typologie de clientèle
- Étude de la dynamique de la structure par terme des taux
- Analyse de l'activité des hôpitaux locaux
- Typologie de consommation de soins
- Mesure de la précision du nouveau recensement
- Évolution des concentrations de nitrate dans les cours d'eau bretons

Documents pédagogiques

Un guide pratique.

Pré-requis

Cet enseignement demande la maîtrise de l'ensemble des enseignements de statistique de 1^{re} et 2^e année qui seront mobilisés de façon inégale, selon le sujet traité.

Contrôle des connaissances

Un rapport écrit, un poster et une soutenance orale.

Références bibliographiques

Selon sujet traité.

Langue d'enseignement

Français

UE 2-07– Matière 2ASTA07 – Semestre 2

Introduction aux modèles de durée

Introduction to Survival Analysis

Cours : 12h • TD : 6h • TP : 3h

Enseignant : Valentin PATILEA (Ensay)

Correspondant : Valentin PATILEA

Objectif de la matière

L'analyse de survie étudie les données qui représentent des durées. Les modèles d'analyse des données sont couramment utilisés en biostatistique, assurance, fiabilité, gestion de risque et finance, économie du travail, astrophysique, etc. Les objectifs de ce cours sont la présentation des aspects spécifiques de l'analyse statistique de durées, et des méthodes et modèles standard d'analyse, y compris pour le cas des variables explicatives sont disponibles. Un autre objectif est l'initiation aux logiciels statistiques spécialisés.

Contenu de la matière

1. Motivation et exemples
2. Notions statistiques pour les modèles de durée
3. Mécanismes de censure et de troncation
4. Modèles paramétriques (Weibull,...) et non paramétriques (Kaplan-Meier,...)
5. Comparaison des survies de deux échantillons
6. Modèles de régression (Cox Risques Proportionnels, Accelerated Failure Time,...)

Documents pédagogiques

Slides

Pré-requis

Bonne connaissance des notions de probabilités et de statistique inférentielle. Compétences de programmation en R sont recommandées.

Contrôle des connaissances

Un examen écrit de 2h.

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 Analysis*, CRC Press, 2014.
- D.G. KLEINBAUM, J.P. KLEIN: *Survival Analysis. A Self-Learning Text*, Springer, 2005.

Langue d'enseignement

Anglais

UE 2-07 – Matière 2ASTA08 – Semestre 2

Statistique non paramétrique

Nonparametric Econometrics

Cours : 21h

Enseignant : Marian HRISTACHE (Ensaï)

Correspondant : Marian HRISTACHE

Ce cours est destiné aux élèves ingénieurs

Objectif de la matière

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 P^k). Identifier la loi revient alors à estimer cette fonction, approche qui a connu un vigoureux développement au cours des vingt dernières années, et constituera l'objet de ce cours.

Contenu de la matière

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.

Documents pédagogiques

Support de cours.

Pré-requis

Bonne connaissance des cours de statistique 1-4, modèles de régression et économétrie.

Contrôle des connaissances

Examen écrit + devoirs.

Références bibliographiques

- BOSQ, D., LECOUTRE, J. P. (1987), *Théorie de l'estimation fonctionnelle*, Economica.
- 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.

Langue d'enseignement

Français

UE 2-07 – Matière 2ASTA09 – Semestre 2

Sondages 2 : Traitement de la non-réponse dans les enquêtes

Missing Survey Data

Cours : 12h • TD-TP : 9h

Enseignant : Guillaume CHAUVET (Ensayi)

Correspondant : Guillaume CHAUVET

Ce cours est destiné aux élèves attachés

Objectif de la matière

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.

Contenu de la matière

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

Documents pédagogiques

Support de cours

Pré-requis

Cours de Probabilités, Théorie des Sondages 1, Régression linéaire multivariée, Modèle linéaire généralisé et régression sur variables catégorielles.

Contrôle des connaissances

Examen écrit.

Compétences acquises

Identifier les facteurs d'incertitude liés à la non-réponse.

Choisir une stratégie de traitement de la non-réponse totale et la mettre en œuvre dans des situations simples (échantillonnage à un degré).

Produire des estimateurs corrigés de la non-réponse totale et les indicateurs de précision associés.

Choisir une stratégie de traitement de la non-réponse partielle et la mettre en œuvre dans des situations simples.

Travail personnel

20h

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.

Langue d'enseignement

Français

UE 2-07 – Matière 2ASTA10 – Semestre 2

Chaînes de Markov

Markov Chains

Cours : 12h • TD : 9h

Enseignant : Adrien SAUMARD (Ensay)

Correspondant : Adrien SAUMARD

Objectif de la matière

Ce cours revient des chaînes de Markov, famille importante de processus stochastiques, déjà rencontrée en ISAO, et dont le champ d'application est très large : file d'attente, gestion de stocks, fiabilité, traitement d'images, économétrie, génétique, physique, etc.

Le cours nécessite la maîtrise du calcul des probabilités. Il introduira les notions nécessaires à une bonne compréhension de la modélisation des phénomènes markoviens.

De nombreux exemples d'application seront étudiés en cours et sous forme d'exercice.

Contenu de la matière

Rappel des définitions.

Équation de Chapman-Kolmogorov et formules de conditionnement.

Classification des états, périodicité, temps d'atteinte, récurrence et transience.

Loi stationnaire et théorèmes limites, statistiques des chaînes de Markov.

Pré-requis

Aucun

Contrôle des connaissances

Examen écrit de 2h00.

Références bibliographiques

- BALDI P., MAZLIAK L. & PRIOURET P., *Martingales et chaînes de Markov : théorie élémentaire et exercices corrigés*, Hermann, 2001.
- BENAÏM M., EL KAROUI N., *Promenade aléatoire*, éditions de l'École Polytechnique, 2004.
- BOULEAU N., *Processus stochastiques et applications (nouv. éd.)*, Hermann, 2000.
- FOATA D. & FUCHS A., *Processus Stochastiques (2^e éd.)*, Dunod, 2004.
- GRAHAM C., *Chaînes de Markov*, Dunod, 2008.
- GRIMMETT G.R. & STIRZAKER D.R., *Probability and Random Processes*, Oxford Sciences Publications, 2001 (3rd ed.).
- NORRIS J.R., *Markov Chains*, Cambridge Series in Statistical and Probabilistic Mathematics, 1997.
- PARDOUX E., *Processus de Markov et applications: Algorithmes, réseaux, génome et finance*. Dunod, 2007.

Langue d'enseignement

Français.

UE 2-08– Matière 2ASTA11 – Semestre 2

Calculs Bayésiens

Bayesian computation

Cours : 21h • TP : 18h

Enseignant : Marie DU ROY DE CHAUMARAY (Ensay)

Correspondant : Marie DU ROY DE CHAUMARAY

Objectif de la matière

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.

Contenu de la matière

- 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)

Pré-requis

Notions de base en Probabilités, en Statistique inférentielle et en programmation avec R.

Contrôle des connaissances

Partiel et examen final sur papier.

Références bibliographiques

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

Langue d’enseignement

Français

UE 2-08– Matière 2ECO03 – Semestre 2

Microéconométrie appliquée

Applied Microeconometrics

Cours : 21h

Enseignant : Eric DELATTRE

Correspondant : Vincenzo CAPONI

Objectif de la matière :

Ce cours a pour objet de présenter les fondements économétriques de différentes techniques à disposition de l'économètre appliqué et de les illustrer par des exemples d'application (articles de recherche). Il sera également mis à disposition des élèves certaines commandes standards utilisées dans le logiciel STATA. L'élève devra être capable, à l'issue de ce cours de passer alternativement de la théorie aux aspects empiriques associés aux différentes méthodes présentées.

Contenu de la matière :

1. Problèmes de sélection sur inobservables (Heckman et switching endogène).
2. Modèles de comptage.
3. Modèles de durée.
4. Régressions quantiles.
5. Modèles polyatomiques et principes de la vraisemblance simulée.
6. Problèmes d'évaluation non marchande : évaluation contingente, prix hédoniques, couts de transports.
7. Introduction à l'économétrie bayésienne.

Documents pédagogiques

Les documents seront disponibles en ligne.

Pré-requis

Méthodes d'estimation linéaires (MCO, VI) et par maximum de vraisemblance.
Fondements microéconomiques.

Contrôle des connaissances :

Examen terminal avec documents.

Références bibliographiques :

- Limited-dependent and qualitative variable in econometrics. G.S. Maddala. CUP, 1983. [C1]
- Econometric analysis of cross section and panel data. J.M. Wooldridge. MIT Press, 2002. [C2, C3]
- Quantile regression. R. Koenker and K.F. Hallock. Journal of Economic Perspectives. Vol 5, Num 4, pp 1443-156. 2001. [C4]
- La régression quantile en pratique. P. Givord et X. D'Hautefeuille. DT INSEE 2013/01, 2013. [C4]
- Discrete choice methods with simulation. K.E. Train. CUP, 2003. [C5]
- Economic evaluation with stated preferences techniques. I.J Bateman et alii. Edward Elgar. 2002. [C6]
- Introduction to Bayesian econometrics. E. Greenberg. CUP, 2008. [C7]

- Contemporary Bayesian econometrics and statistics. J.K. Geweke. CUP, 2008. [C7]

Langue d'enseignement :

Français

UE 2-08– Matière 2ECO09 – Semestre 2

Macroéconométrie appliquée

Applied Microeconometrics

Cours : 21h

Enseignant : Stéphane AURAY (Ensaï)

Correspondant : Stéphane AURAY

Objectif de la matière

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.

Contenu de la matière

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)

Pré-requis

Les cours de macroéconomie de première année et de séries temporelles de deuxième année sont supposés assimilés.

Contrôle des connaissances

Projets

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

Langue d'enseignement

Anglais et français

UE 2-08– Matière 2INFO06 – Semestre 2

Programmation orientée objet avec Java

Object-Oriented Programming with Java

Cours : 6h . TP : 12h . travail personnel : 10h

Enseignant : Olivier LEVITT (Insee)

Correspondant : Romaric GAUDEL

Objectif de la matière

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".

Compétences

- Programmer en Java
- Traiter des flux/collections de données
- Fournir un livrable

Contenu de la matière

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

Pré-requis

Algorithmique. Introduction à la programmation orientée objet en Python. Compléments d'informatique.

Contrôle des connaissances

Contrôle continu : un examen écrit + éventuellement un TP noté

Références bibliographiques

- C. Delannoy, Programmer en java, Eyrolles. 2012 (8e éd.)
- B. Eckel, Thinking in Java, Prentice Hall, 2006 (4th ed.), www.BruceEckel.com, voir aussi la version française <http://penserenjava.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.

Langue d'enseignement

Français.

UE 2-08– Matière 2INFO05 – Semestre 2

C + + pour la statistique computationnelle

Programming in C + +

Cours : 6h – TP : 12h

Enseignant : Thomas GUYET (Agrocampus Ouest/IRISA)

Correspondant : Hong PHUONG DANG

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 2014. 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 les 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.

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.

Objectif de la matière

Comprendre le fonctionnement de la gestion de la mémoire des ordinateurs et donner aux élèves les principaux concepts du langage C++ pour écrire des programmes impliquant une gestion maîtrisée de la mémoire.

Contenu de la matière

Les notions de base :

- Structure des programmes
- Expressions, instructions et types de base
- La manipulation des pointeurs, modes références et gestion de la mémoire
- Représentation mémoire des tableaux et matrices

Structures de données avancées

- Introduction aux classes C++
- Structure algorithmiques : listes, arbres

Outils pour le C/C++

- Environnement de programmation (notions de compilation)
- Debugage (gdb, valgrind)
- Librairie standard : structures de données linéaires, itérateurs

Documents pédagogiques

Support de cours et de TP. Documents complémentaires en ligne (correction des TP, exercices corrigés en ligne, annales d'examen).

Pré-requis

Le contenu du cours d'algorithmique de première année, les concepts de programmation orientée objet (programmation Python et conception d'application UML) ainsi qu'une connaissance de l'environnement de développement Eclipse.

Contrôle des connaissances

Évaluation sur table.

Références bibliographiques

À compléter

Langue d'enseignement

Français

UE 2-08– Matière 2AINF06 – Semestre 2

Outils pour le Big Data

Big Data IT Tools

Cours :10,5h - TP : 13,5h – travail personnel : 12h

Enseignant	:	Arthur KATOSSKY/Rémi PÉPIN (Ensay)
------------	---	------------------------------------

Correspondant	:	Arthur KATOSSKY/Rémi PÉPIN
---------------	---	----------------------------

Objectif de la matière

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.

Contenu de la matière

- Les différents « v »
- Principes, avantages et inconvénients d'un système réparti
- connaître les stratégies de tolérance aux fautes (duplication des données, exécution avec erreurs)

Compétences

- Identifier l'architecture adaptée à une tâche (exécution séquentielle et/ou parallèle, exécution en mémoire et/ou en flux, exécution locale et/ou distante).
- Lancer des calculs sur une architecture Big Data (notamment, appliquer les paradigme Map-Reduce).
- exécuter des calculs volumineux – et en particulier des calculs statistiques – sur des prestataires de calcul (IaaS ou PaaS comme Amazon Web Service, Google Cloud Platform ou autre)

Pré-requis

Algorithmique.

Contrôle des connaissances

À déterminer.

Références bibliographiques

- Analyses des Big Data : quels usages, quels défis ? Note d'analyse du Commissariat général à la stratégie et la prospective
- Pirmin Lemberger, Marc Batty, Médéric Morel, Jean-Luc Raffaëlli. Big Data et machine learning - Manuel du data scientist, Dunod, 2015.
- Rudi Bruchez. Les bases de données NoSQL et le BigData : Comprendre et mettre en œuvre, Eyrolles (2015)

Langue d'enseignement

Français.

**Enseignements de deuxième semestre :
cours électifs (pré-spécialisation UE 2-09)**

UE 2-09 – Matière 2ASTA12 – Semestre 2

Modèles avancés de régression

Advanced Regression Models

Cours : 21h

Enseignant : Gilles STUPFLER (Ensay)

Correspondant : Gilles STUPFLER

Objectif de la matière :

Les procédures statistiques classiques reposent sur la connaissance des distributions théoriques des statistiques. La motivation du rééchantillonnage est d'approcher par simulation de Monte Carlo la distribution d'une statistique conditionnellement aux données observées lorsque l'on ne connaît pas sa loi. Le principe de cette technique est de substituer à la distribution 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. 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).

Une liste d'exercices non corrigés est fournie pour permettre aux étudiants de s'exercer en supplément des cours. Un volume horaire de 21h à 30h de travail personnel est estimé pour ce cours.

Contenu de la matière :

1. Régression non paramétrique : estimateurs par polynômes locaux
2. Régression partiellement linéaire
3. Réduction de dimension : modèles à direction révélatrice

Documents pédagogiques :

Support de cours.

Pré-requis :

Bonne connaissance des cours de modèles statistiques et estimation et modèles de régression. Avoir suivi le cours de modèles linéaires généralisés aidera à une meilleure compréhension.

Contrôle des connaissances :

Examen écrit.

Références bibliographiques :

- Fan, J. et Gijbels, I. (1992). Variable bandwidth and local linear regression smoothers, *The Annals of Statistics*, 20, 2008–2036.
- Fan, J. et Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*, Chapman & Hall.

- Friedman, J.H. et Stuetzle, W. (1981). Projection pursuit regression, *Journal of the American Statistical Association*, 76, 817–823.
- Härdle, W., Müller, M., Sperlich, S. et Werwatz, A. (2004). *Nonparametric and Semiparametric Models*, Springer.
- Härdle, W. et 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. et 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. et Spady, R.H. (1993). An efficient semiparametric estimator for binary response models, *Econometrica*, 61, 387–421.
- Powell, J.L., Stock, J.H. et 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.

Langue d’enseignement :

Français

UE 2-09 – Matière 2ASTA13 – Semestre 2

Martingales et processus de Lévy

Martingales and Levy Process

Cours : 12h • TD : 9h

Enseignant : Benoit CADRE (ENS)

Correspondant : Lionel TRUQUET

Objectif de la matière

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.

Avoir une intuition de la notion de martingale en temps continu et de la propriété de Lévy.

Être prêt à appliquer ces techniques dans un cadre applicatif.

Contenu de la matière

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
5. Processus de Lévy, processus de Poisson, mouvement brownien
6. Exemples de martingales à temps continu

Pré-requis

Une bonne maîtrise des techniques probabilistes telles qu'elles sont présentées dans le cours de Probabilités « maths ».

Contrôle des connaissances

Un examen final (2 h). Documents et calculatrices autorisés.

Références bibliographiques

- FOATA D., FUCHS A., *Processus stochastiques (2^e éd.)* (Dunod, 2004)
- WILLIAMS D., *Probability with martingales* (Cambridge University Press, 1991)

Langue d'enseignement

Français

UE 2-09– Matière 2ASTA14 – Semestre 2

Méthodes de Rééchantillonnage

Resampling Methods

Cours : 21h

Enseignant : Myriam VIMOND (Ensay)

Correspondant : Myriam VIMOND

Objectif de la matière :

Les procédures statistiques classiques reposent sur la connaissance des distributions théoriques des statistiques. La motivation du rééchantillonnage est d'approcher par simulation de Monte Carlo la distribution d'une statistique conditionnellement aux données observées lorsque l'on ne connaît pas sa loi. Le principe de cette technique est de substituer à la distribution 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. 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).

Une liste d'exercices non corrigés est fournie pour permettre aux étudiants de s'exercer en supplément des cours. Un volume horaire de 21h à 30h de travail personnel est estimé pour ce cours.

Contenu de la matière :

1. La mesure empirique
2. Tests par permutations
3. La validation croisée et le jackknife
4. Le Bootstrap d'Efron
5. Autres méthodes de rééchantillonnage
6. Les nombres pseudo-aléatoires

Compétences acquises

- connaissance et compréhension des principes des principales méthodes de rééchantillonnage ;
- savoir présenter et décrire de manière rigoureuse une procédure de rééchantillonnage ;
- savoir adapter les principes de rééchantillonnage selon l'objectif visé et la nature des données.

Pré-requis

Nombres Pseudo-Aléatoires, Méthodes de Monte Carlo, Données Manquantes (Cours de Calcul Bayésien), Statistique Inférentielle (Cours de première année), Probabilités, Régression, Séries temporelles, R et/ou Python

Contrôle des connaissances :

Examen écrit sans document

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.

Langue d'enseignement :

Français

UE 2-09 – Matière 2ASTA15 – Semestre 2

Séries temporelles 2

Times Series 2

Cours: 21h

Enseignant : Purevdorj TUVAANDORJ (Ensay)

Correspondant : Purevdorj TUVAANDORJ

Objectif de la matière

L'objectif de ce cours d'introduction aux séries temporelles multivariées est de présenter les différents modèles linéaires qui permettent de prendre en compte la corrélation entre les différentes séries et donc d'améliorer les prévisions de ces séries. Des outils spécifiques à l'analyse des séries temporelles multivariées tels que les problèmes de causalité ou de cointégration seront également introduits. Enfin, l'inclusion de régresseurs exogènes ou la prise en compte de la saisonnalité pour ces modèles sera discutée. Les séances de TP se feront avec le logiciel R.

Contenu de la matière

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

Pré-requis

Notions de base en Probabilités et Statistique inférentielle. Calcul matriciel de base. Cours de séries temporelles univariées.

Contrôle des connaissances

Examen final sur papier

Références bibliographiques

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

Langue d'enseignement

Français

UE 2-09– Matière 2ASTA16 – Semestre 2

Statistique mathématique

Mathematical Statistics

Cours : 21h

Enseignant : Lionel TRUQUET (Ensaï)

Correspondant : Lionel TRUQUET

Objectif de la matière

L'objectif de ce cours est de compléter le cours de d'introduction à la statistique vu en première année en présentant le formalisme mathématique qui permet de justifier certaines propriétés des estimateurs ou des procédures de test. En particulier, on justifiera les lois limites de certains estimateurs tels que le maximum de vraisemblance et on donnera un sens précis à la notion d'efficacité asymptotique pour un estimateur ou un test.

Contenu de la matière

- Compléments sur la convergence en loi. Delta méthode
- Estimateurs par minimum de contraste. Propriétés asymptotiques
- Normalité asymptotique locale. Efficacité des estimateurs
- Statistiques de rang, de signe et de permutations
- Efficacité asymptotique d'un test

Pré-requis

Le cours de statistique et de probabilités générales de 1A.

Contrôle des connaissances

Examen écrit.

Références bibliographiques

- van der Vaart, A.W., Asymptotic statistics, Cambridge University Press, 2000.

Langue d'enseignement

Français

UE 2-09 – Matière 2ASTA17 – Semestre 2

Sondages 3 : Echantillonnage à plusieurs degrés

Multistage Sampling

Cours : 12h • TD : 9h

Enseignant : Guillaume CHAUVET (Ensay)

Correspondant : Guillaume CHAUVET

Ce cours est principalement destiné aux élèves attachés.

Ce cours est indispensable pour les élèves du master Evaluation et décision publiques.

Ce cours est conseillé pour les filières ISTS et Sciences de la Vie.

Objectif de la matière

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.

Contenu de la matière

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)

Documents pédagogiques

Support de cours

Pré-requis

Cours de Probabilités, Théorie des Sondages 1, Régression linéaire multivariée.

Contrôle des connaissances

Examen écrit.

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.

Travail personnel

30h

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

Langue d'enseignement

Français

UE 2-09 – Matière 2ASTA18 – Semestre 2

Statistique spatiale

Spatial analysis

Cours : 3h TP : 18h

Enseignants : Arthur KATOSSKY (Ensay)

Correspondant : Arthur KATOSSKY

Objectif de la matière

Prendre en main un outil de cartographie et apprendre à réaliser des cartes sémiologique-ment correctes avec QGis.

Comprendre l'intérêt d'une analyse spatiale pour mieux appréhender des données, déceler les pièges du zonage, savoir s'en affranchir et acquérir quelques rudiments d'analyse spatiale.

Contenu de la matière

- Initiation aux projections et aux référentiels cartographiques
- Présentation des bonnes pratiques
- Création de nouvelles couches géographiques / Importation et exportation des= données
- Cartographie élémentaire et sémiologie
- La cartographie à l'Insee et les zonages d'études.
- Géocodage/géolocalisation et utilisation de données carroyées pour contourner les inconvénients des zonages
- Enjeux de la cartographie statistique, aperçu des méthodes d'analyse spatiale, impact de l'autocorrélation spatiale sur la modélisation statistique

Documents pédagogiques

Support disponible dans Moodle

Pré-requis

Aucun

Contrôle des connaissances

1 contrôle continu pour 1/3 de la note (quizz, TP noté, ...), et un atlas cartographique d'une dizaine de cartes à réaliser en binôme, pour 2/3 de la note.

Références bibliographiques

- Quantum Gis – Manuel de l'utilisateur
- P. Rouet (1991), Les données dans les systèmes d'information géographique, Ed. Hermes
- S-A Souiah & E. Minvielle (2003) - L'analyse démographique et spatiale. Statistiques, cartographie, télédétection, SIG. Editions du temps
- JM FLOCH (2012), Détection des disparités socio-économiques - L'apport de la statistique spatiale (Document de travail Insee - disponible en ligne)

Langue d'enseignement

Français, possibilité de faire un groupe de TP en anglais si des élèves le souhaitent.

UE 2-09 – Matière 2AECO04 – Semestre 2

Economie du risque

Economics of Risk

Cours : 21h

Enseignant : Olivier RENAULT

Correspondant : Vincenzo CAPONI

Objectif de la matière

Le cours se divise en 2 parties. Dans une première partie, nous étudierons le processus de décision en situation de risque et d'incertitude ainsi qu'une introduction à la théorie de l'assurance. La seconde partie sera consacrée à des applications (mesures d'inégalités, assurance, portefeuille) encore à déterminer.

Contenu de la matière

1. Eléments de théorie du risque et de l'incertitude
 - 1.1 Introduction : risque et incertitude
 - 1.2 Quelques critères de décision
 - 1.2.1 en situation d'incertitude
 - 1.2.2 en situation risquée
 - 1.3 Attitude vis-à-vis du risque
 - 1.3.1 Le critère de l'utilité espérée
 - 1.3.2 La dominance stochastique
 - 1.3.3 Prime de risque et Equivalent-certain
 - 1.3.4 Aversion au risque
 - 1.3.5 Mesure du risque
 - 1.4 Le risque et la théorie de l'assurance
 - 1.4.1 Aléa moral et anti-sélection
 - 1.4.2 Assurance optimale

2. Applications

Pré-requis :

Modélisation microéconomique du comportement du consommateur et du producteur

Contrôle des connaissances :

A déterminer

Références bibliographiques :

- The Economics of Risk and Time, Christian Gollier, MIT Press, 2001.
- Economic and financial decisions under uncertainty, Louis ECKHOUDT, Christian GOLLIER, Harris SCHLESINGER, Princeton University Press, 2005
- *Measuring inequality* (3rd ed.), Frank Cowell, Oxford University Press, 2011
<http://darp.lse.ac.uk/MI3>

Langue d'enseignement :

Français ou anglais

UE 2-09 – Matière 2AECO05 – Semestre 2

Théorie de la croissance économique

Economic Growth Theory

Cours : 21h

Enseignant : Vincenzo CAPONI (Ensay)

Correspondant : Vincenzo CAPONI

Objectif de la matière

The main objective of the course is to study the economic development from the beginning of economic activity (or from the point we start having some data) to most recent times and focus on human capital and education as one of the main sources of growth. The course will introduce a variety of modelling tools that will be used to interpret actual data and that, gradually, will give us a modelling strategy to describe the economic development of the world economy from stagnation, to moderate growth to the more recent sustained growth. The questions we will be able to answer by the end of the course are several: why are countries so much richer than others? Why our economies transit from economic stagnation with high fertility rates to economic growth with low ones? What is more likely to drive economic growth and the improvement of living conditions? What are the factors that generate inequality? Is there a trade-off between inequality and growth? What policies can foster growth and/or more equality?

Compétences acquises

Except for the specific tools that students will become familiar with, skills that are expected to be developed are as follows.

- The ability to abstract and simplify in order to identify and model the essence of a problem. We will look at the world economic history and capture a few economic fundamental facts and trends in order to develop our interpretative economic tools.
- Critical thinking. We will carefully look, for every economic modelling tools and assumptions, at their weaknesses and check of ways to improve our understanding.
- Draw economic policy inferences and to recognize the potential constraints in their implementation.

Contenu de la matière

Growth

Intro: What is Growth? Why it is Important?

A Model of non-Growth: Malthus

A Simple of Model of Growth: Solow

Causes of Growth

What is Human Capital? Its Importance for Growth.

Human Capital and Returns to Education

How to Finance Education

Inequality and Growth

Pré-requis

Microeconomics and Macroeconomics. In particular students are expected to be familiar with concepts of utility and profit maximization; general equilibrium; economic dynamics. Terms such as GDP, GDP per capita, wages, interest rate, consumption, steady state, etc... should also be familiar to the students.

Contrôle des connaissances

There will be a final exam worth no less than 60% of the final grade. The remaining (up to 40%) will be assessed either by writing a midterm exam, or with some alternative work to be discussed at the beginning of the course (small project, presentations, data based assignments).

Références bibliographiques

- ACEMOGLU D., (2008), *Introduction to Modern Economic Growth*, Princeton University Press
- JONES C. (2013), *Introduction to Economic Growth (Third Edition)*, W.W. Northon.
- CHECCHI D., (2008) *The Economics of Education: Human Capital, Family Background and Inequality*, Cambridge University Press
- BREWER D. J. and mcEwan P. J., (2010), *Economics of Education*, Elsevier

Langue d'enseignement

English

UE 2-09 – Matière 2AECO07 – Semestre 2

Economie Industrielle Empirique

Empirical Industrial Organization

Cours : 21h

Enseignant : Yutec SUN (Ensaï)

Correspondant : Vincenzo CAPONI

Objectifs de la matière

In this course, we will learn practical methods for conducting empirical analysis of consumer demand, firm strategy, and social welfare under various market environments. We will focus on the econometric analysis guided by the theoretical frameworks developed in the course Industrial Organization (Economie Industrielle).

Contenu de la matière

1. Estimation of consumer demand
2. Estimation of firm's market power
3. Estimation of entry costs and market structure
4. Estimation of bargaining model
5. Estimation of rational expectations model
6. Estimation of dynamic decision process model

Contrôle des connaissances

1. Examen écrit de 2h, sans aucun document.
2. Problem sets for take-home exercise

Pré-requis

The course will assume basic knowledge of econometrics and applied statistics such as maximum likelihood estimation, instrumental variables regression. The students are encouraged to take a course in Industrial Organization (Economie Industrielle) before this course.

Références bibliographiques

- TRAIN, K. *Discrete Choice Methods with Simulation*, Cambridge University Press, 2003.

Langue d'enseignement

Anglais.

UE 2-09 – Matière 2AECO07 – Semestre 2

Economie Industrielle

Industrial Organization

Cours : 21h

Enseignant : Frédéric LOSS (Université de Dauphine)

Correspondant : Vincenzo CAPONI

Objectifs de la matière

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.

Contrôle des connaissances

Examen écrit de 2h, sans aucun document.

Compétences acquises

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 ...)

Travail personnel à fournir

Le travail personnel à fournir dépendra des connaissances de chaque étudiant en microéconomie. Un étudiant ayant suivi un premier cours de microéconomie n'aura pas de difficulté à suivre ce cours. Il lui faudra alors compter environ une heure de travail personnel pour une heure de cours magistral. En revanche, un étudiant n'ayant suivi aucun cours de microéconomie devra faire un effort supplémentaire. Il lui faudra alors compter environ deux heures de travail personnel pour une heure de cours magistral.

Références bibliographiques

- TIROLE, J. *Théorie de l'Organisation Industrielle*, Economica, 1998.
- MOTTA, M. *Competition Policy. Theory and Practice*. Cambridge University Press, 2004.
- AMSTRONG, M. et PORTER, R. *Handbook of Industrial Organization, Vol3*, Elsevier, 2007

Langue d'enseignement

Français

UE 2-09 – Matière 2AECO08 – Semestre 2

Démographie

Demographics

Cours : 21h

Enseignant : Laurent DI CARLO (Insee)

Correspondant : Ronan LE SAOUT

Objectif de la matière

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.

Contenu de la matière

- | | |
|--|--|
| 1. Introduction à la démographe | 4.3. Le taux de fécondité global |
| 1.1. Définition | 4.4. Le taux de fécondité par âge |
| 1.2. Histoire de la démographie | 4.5. L'indicateur conjoncturel de fécondité |
| 1.3. Des bases de données démographiques | 4.6. L'âge moyen à la maternité |
| | 4.7. La descendance finale |
| 2. La croissance de la population | 4.8. Le taux brut de reproduction |
| 2.1. Croissance absolue et croissance relative | 4.9. Le taux net de reproduction |
| 2.2. Les facteurs d'évolution de la population : l'équation fondamentale | 5. L'analyse de la mortalité |
| 2.3. Évolution de la population en France sur longue période | 5.1. Taux brut de mortalité |
| 2.4. La démographie des régions françaises | 5.2. Les taux de mortalité par âge |
| | 5.3. La standardisation des taux de mortalité |
| 3. Outils de la démographie | 5.4. La mortalité infantile |
| 3.1. La pyramide des âges | 5.6. La table de mortalité |
| 3.2. L'âge | 5.7. L'espérance de vie |
| 3.3. Le diagramme de Lexis | 5.8. Les causes de la mortalité |
| 4. L'analyse de la fécondité | 6. Les indicateurs de développement humain (ONU) |
| 4.1. Les facteurs influençant le niveau des naissances | 7. La transition démographique |
| 4.2. Le taux brut de natalité | |

Pré-requis

Aucun.

Contrôle des connaissances

Un examen écrit d'une durée de 2h sans document.

Références bibliographiques

- PNUD, rapport sur le développement humain, 2014
- Catherine Rollet, Introduction à la démographie, 3^e 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

Langue d'enseignement

Français.

UE 2-09 – Matière 2AECO09 – Semestre 2

Economie financière

Financial Economics

Cours : 21h

Enseignant : Samuel DANTHINE (Ensayi)

Correspondant : Samuel DANTHINE

Objectif de la matière

Ce cours a pour objectif d'introduire les concepts de bases utilisés dans l'analyse du monde bancaire et de la finance, ainsi que des éléments d'économie financière. Il permettra aux élèves d'avoir une vision de base mais globale de ce secteur.

Contenu de la matière

1. Survol du système financier
2. La monnaie
3. Les taux d'intérêt
 - 3.1. Mesure
 - 3.2. Comportement
 - 3.3. Structure à risque et structure à terme
4. Marché boursier, théorie des anticipations rationnelles et hypothèse d'efficacité des marchés
5. Les produits financiers :
 - 5.1. Les actions
 - 5.2. Les obligations
 - 5.3. Les produits monétaires
 - 5.4. Les produits dérivés
 - 5.5. Les actifs alternatifs
6. Analyse économique de la structure financière
7. Banques et gestion des institutions financières
8. Analyse économique de la régulation financière
9. La banque aux US, en Europe, et en France
10. Introduction à l'économie financière

Pré-requis

Cours d'économie de 1^{re} année et du premier semestre de 2^e année.

Contrôle des connaissances

Examen écrit

Références bibliographiques

- Mishkin, F. The Economics of Money, Banking, and Financial Markets, 11th ed., Pearson.
- Mishkin, F., Bordes, C., Lacoue-Labarthe, D., Leboisne, N., Poutineau, J-C. Monnaie, banque et marchés financiers, 10^e édition, Pearson.
- Hull, J. Options, Futures, and other Derivatives, 9th edition, Pearson. (version française également)

- Hull, J. Risk Management and Financial Institutions, 4th edition, Wiley. (version française également)
- Danthine, J.P., Donaldson, J. Intermediate Financial Theory, 3rd edition, Academic Press.

Langue d'enseignement

Français et anglais

UE 2-09– Matière 2AINF07 – Semestre 2

Programmation avancée en R

Advanced Programming in R

Cours/TP : 21h

Enseignant : Matthieu MARBAC (Ensay)

Correspondant : Matthieu MARBAC

Objectif de la matière

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

Travail personnel

Un volume horaire de 3 à 4 h de travail personnel est estimé par séance de TP (sous réserve de maîtriser les pré-requis du cours). Un volume horaire de 20 à 30h de travail personnel, par étudiant, est estimé pour le projet.

Contenu de la matière

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

Pré-requis

Programmation en R (1A)

Cours de Programmation objet en C++ (2A)

Contrôle des connaissances

Un examen (TP noté).

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.

Langue d'enseignement

Français

UE 2-09 – Matière 2AINF08 – Semestre 2

Technologies mobiles

Mobile Technologies

Cours : 6h • TP : 15h

Enseignant : Olivier LEVITT (Insee)

Correspondant : Hong-Phuong DANG

Objectif de la matière

L'arrivée des appareils mobiles entraîne des changements dans le développement des applications pour tenir compte des capacités de stockage mais surtout de la spécificité des écrans de visualisation, notamment leur taille. L'objectif de ce cours est de sensibiliser les étudiants aux contraintes d'utilisation de ces supports et de leur montrer, à travers de cas pratiques, comment en tenir compte lors de la réalisation d'une application. Le développement sera effectué pour des appareils fonctionnant sous Android : smartphone ou tablette tactile.

Contenu de la matière

1. Android, les présentations
 - a. Contexte
 - b. Architecture
2. Android et les applications
 - a. Cycle de vie
 - b. Activités
3. Android et les interfaces
 - a. Layouts et Views
 - b. Menus et évènements
4. Android et les données
 - a. Accès aux ressources
 - b. Content Providers

Documents pédagogiques

Support de cours

Pré-requis

Les étudiants choisissant cette option devront être très à l'aise avec les concepts objets et la programmation abordés en première année.

Contrôle des connaissances

Un travail pratique à faire en binôme.

Références bibliographiques

Langue d'enseignement

Français

UE 2-09 – Matière 2AINF09 – Semestre 2

Traitement du signal

Signal Processing

Cours : 8 h • TD-TP : 13 h

Enseignant : Hong-Phuong DANG (Ensayi)

Correspondant : Hong-Phuong DANG

Objectif de la matière

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^e 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^e 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.

Contenu de la matière

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)

Documents pédagogiques

Support de cours

Pré-requis

- Un peu de maths : suite et série de fonction, intégration, nombre complexe, base de la théorie de la mesure
- Connaissance de langage Python pour les TPs

Langue d'enseignement

Français

UE 2-09 – Matière 2AINF10 – Semestre 2

Conception de logiciel

Software design

Cours : 9 h • TP : 12 h

Enseignant : Mohamed GRAIET (Ensayi)

Correspondant : Mohamed GRAIET

Objectif de la matière

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.

Contenu de la matière

- 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

Compétences acquises

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

Pré-requis

Programmation orientée objet en Python et Java

Documents pédagogiques

Support de cours

Contrôle des connaissances

Examen écrit

Références bibliographiques

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

Langue d'enseignement

Français

UE 2-09 – Matière 2AINF11 – Semestre 2

Data Challenge

Cours : 21h

Enseignant : Divers enseignants

Correspondant : Salima EL KOLEI

Objectif du cours électif

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 laquelle 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.

Contenu de la matière

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>

Compétences acquises

- 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...)

Prérequis

- Compétences en statistiques et informatiques de 1A et 2A.
- Compétences transversales mobilisées dans les projets 1A et 2A.

Modalités pédagogiques et d'évaluation

- Participation à un challenge et aux ateliers participatifs.
- Rapport écrit et une soutenance permettant d'évaluer leur compétences de data scientists sous l'angle à la fois de la statistique, de l'informatique et des softskills.

UE 2-09 – Matière 2AINF12 – Semestre 2

DataVisualisation

Cours : 21h

Enseignant : Arthur KATOSSKY (Ensayi)

Correspondant : Arthur KATOSSKY

Objectif de la matière

La visualisation de données (*data visualisation* ou *dataviz*) est l'utilisation des conventions graphiques pour la faciliter la compréhension de données trop volumineuses ou trop complexes pour être communiquées sous forme de listes ou de tableaux. L'objectif de ce cours est double:

- d'une part, fournir les éléments de bases de la sémiologie graphique (pour les conventions de représentations) et de la psychologie cognitive (pour la perception visuelle) pour permettre la critique constructive d'une infographie donnée
- d'autre part, **découvrir un nouveau paradigme de programmation** (la programmation événementielle) et un nouveau langage (Javascript) pour permettre la réalisation d'infographies interactives

Contenu de la matière

- Perception visuelle, conventions graphiques, échelles et couleurs
- *Trifecta Checkup* et critique d'infographies statiques
- Bibliothèques R et Python pour la visualisation de données
- Introduction à Javascript et à la bibliothèque d3.js
- Conception d'une infographie interactive
- Panorama des programmes et langages d'aide à la visualisation

Pré-requis:

Aucun

Contrôle des connaissances

Le contrôle des connaissances se fera en deux temps:

- à mi-cours, par groupe de 2, un commentaire critique d'une infographie, accompagné d'une représentation alternative plus pertinente des mêmes données
- en fin de cours, par groupe de 3, une infographie interactive en Javascript sur un sujet au choix

Références bibliographiques:

- Edward Tufte, *The Visual Display of Quantitative Information*, Cheshire, CT, Graphics Press, 2001, 2e éd. (1re éd. 1983)
- <https://beta.observablehq.com/@mbostock>
- <https://junkcharts.typepad.com>

Langue d'enseignement:

Français

Projets professionnels, enseignements optionnels

UE 2-10 Ing – Matière 2AHUM03 – Semestre 1

Projet professionnel ingénieurs

Professional Preparation

TD : 20h

Enseignant : Divers intervenants

Correspondant : Patrick GANDUBERT

Objectif de la matière

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

Contenu de la matière

Séance 1 :

- rédiger des CV et lettres de motivation efficaces ;
- se créer un profil professionnel pour les médias sociaux ;
- valoriser les stages et plus généralement les expériences passées ;
- communiquer efficacement sur son parcours.

Séance 2 :

- préparer des tests psychotechniques, fréquemment utilisés pour les recrutements.

Séance 3 :

- candidater à l'international avec un CV et une lettre de motivation adaptés aux usages des pays anglophones.

Lors de cette séance, des anciens élèves exposeront aussi leurs parcours et expériences de l'international.

Séance 4 : simulation d'entretien, avec des responsables de recrutement et opérationnels.

Deux sessions de 4 heures seront consacrées au questionnaire MBTI (MYERS-BRIGGS TYPE INDICATOR), dont l'objectif est 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) 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).

Pré-requis

Aucun

Contrôle des connaissances

L'enseignement est validé par la validation des CV et lettres de motivation en français et anglais, ainsi que la réponse au questionnaire MBTI.

Références bibliographiques

Seront données en cours.

Langue d'enseignement

Français et anglais

UE 2-10 Att – Matière 2AHUM07 – Semestre 1

Projet professionnel attachés

Professional Preparation

TD : 18h

Enseignant : Divers intervenants

Correspondant : Brigitte GELEIN

Objectif de la matière

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).

Contenu de la matière

Les séances seront consacrées à :

- 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.

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.

Pré-requis

Aucun

Contrôle des connaissances

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

Références bibliographiques

Notices Insee

Langue d'enseignement

Français

UE 2-10 et Bonus

Participation aux activités associatives

Participation in ENSAI associations

Correspondant : Ronan LE SAOUT

Objectif de la matière

La participation à des activités associatives favorise l'ouverture d'esprit, le développement des relations personnelles et l'engagement collectif. De ce point de vue, l'école souhaite l'encourager, dans la limite d'un investissement compatible avec la réussite académique de l'élève qui reste bien évidemment la priorité.

1. Les activités associatives liées à l'école

La participation comme dirigeant ou membre du bureau d'une association de l'école peut donner droit à des points bonus qui s'ajoutent à la moyenne du semestre concerné.

La liste précise des bénéficiaires et le niveau du bonus sont fixés par la direction des études, après concertation avec les associations concernées :

- en début d'année académique pour le semestre 1.
- en début d'année civile pour le semestre 2.

Pour bénéficier de ce bonus, les élèves doivent avoir une activité avérée au sein de l'association pendant le semestre concerné. La liste définitive des bénéficiaires est validée par le directeur des études avant chaque jury.

En plus de ce dispositif, des membres des associations de l'école, dont la liste est précisée en début d'année académique, peuvent bénéficier d'une prise en compte plus importante de leur engagement, afin de valoriser davantage l'ouverture procurée par les activités concernées (comptabilité, organisation, prospection, communication, animation...). Cette prise en compte supplémentaire consiste en la dispense d'un cours d'ouverture par semestre. Pour en bénéficier, les élèves concernés doivent faire remonter par le Président de l'association concernée une demande, puis remettre en fin de semestre un rapport d'activité détaillant les actions réalisées - à titre personnel - pour l'association. La note de ce rapport est prise en compte dans l'UE 10.

Hors cas de force majeure, si un élève ne remplit pas ses engagements et si son activité ne justifie pas la dispense, celle-ci est rétroactivement annulée. L'élève sera donc soumis au régime prévu dans le règlement de scolarité, c'est-à-dire qu'il devra donc composer sur deux cours d'ouverture (ou un cours de langue optionnelle) pour l'unité d'enseignement concernée.

2. Les autres activités associatives

La participation à des activités associatives dans un cadre hors scolaire peut aussi donner droit à des points bonus et à une dispense de cours d'ouverture. C'est un droit depuis la loi « Egalité et Citoyenneté » du 27 janvier 2017 et le décret n°2017-962 du 10 mai 2017 relatif à la reconnaissance des étudiants dans la vie associative, sociale ou professionnelle. C'est notamment le cas lorsque l'activité :

- traduit une compétence forte dans un domaine d'ouverture ;
- traduit un engagement dans un engagement dans une activité bénévole, dans une mission de service civique ou dans la réserve opérationnelle de la défense ;
- valorise l'école ;
- demande un investissement important.

Par exemple, la participation au 4L Trophy peut entrer dans ce cadre, lorsque les dates sont compatibles avec la période d'examens ou de rattrapages de l'école. Les élèves souhaitant en bénéficier doivent déposer une demande argumentée à la direction des études.

Bonus

Sport

Sports

TP : 30h

Enseignant : Divers intervenants

Correspondant : Julien LEPAGE

Enseignement destiné aux élèves-ingénieurs et aux attachés

Objectif de la matière

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

Contenu de la matière

9 activités sportives sont proposées par l'école :

- Badminton
- Basket
- Football
- Hand-ball
- Tennis de table
- Tennis débutant
- 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.

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 Ensay. Pour y prétendre, les élèves concernés doivent remplir les 3 conditions suivantes:

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

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

UE 2-10 – Semestres 1 et 2

Cours d'ouverture

General Culture Courses

Cours : 21h par semestre

Enseignant : Divers intervenants

Correspondant : Todd DONAHUE

Objectifs de la matière

L'objectif global est d'encourager les élèves à développer les *soft skills* dans les domaines de travail en équipe, connaissance de soi, innovation et curiosité intellectuelle et conscience civique et sociétale.

Pour les cours dans un domaine artistique (Cinéma, Dessin, Musique, Peinture et Théâtre)

Les élèves apprennent à analyser, à apprécier et à interpréter des œuvres artistiques (picturales, musicales ou théâtrales) leur permettant d'observer d'un œil critique. Ils participent activement à la création artistique dans le cadre d'un projet commun (représentations musicales et théâtrales, expositions de travaux artistiques).

Pour les cours visant à développer une culture générale dans un domaine scientifique ou des sciences sociales non couvert à l'Ensay (Développement durable, Architecture, Philosophie, Sociologie, Géopolitiques, Psychologie, Physique, Média, Histoire de l'Art)

Les élèves analysent et interprètent des disciplines majeures (de la pensée, de la culture, de l'expression et/ou de la science). Ils évaluent comment les idées influencent le caractère de la pensée humaine, la perception de la réalité et les normes qui gouvernent le comportement. Ils construisent un point de vue qui intègre les facteurs qui influencent l'activité humaine (artistiques, historiques et/ou scientifiques). Ils apprennent à s'exprimer de façon critique en examinant des sources variées qui mettent au défi les idées reçues.

Contenu de la matière

Différents cours d'ouverture sont proposés par l'école. Les cours ouverts en 2019/2020, leur contenu et leur mode de contrôle des connaissances sont décrits dans le fascicule « *Programme des enseignements : Cours libres optionnels.* »

Au second semestre, le cours de physique doit obligatoirement être suivi (en première ou deuxième année) par les élèves ingénieurs des filières économiques et Stid.

UE 2-10 – Matière TALV2 – Semestres 1 et 2

Langues optionnelles

Optional Languages

Cours : 30h par semestre

Enseignant : Divers intervenants

Correspondant : Todd DONAHUE

Objectif de la matière

Les élèves progressent dans les compétences communicatives langagières (expression écrite et orale, compréhension écrite et orale). Les élèves apprennent à décrire et analyser les contextes culturels spécifiques aux peuples qui parlent la langue étudiée. Ils prennent du recul sur leurs propres cultures en s'ouvrant vers celles étudiées en cours pour développer les *soft skills* notamment dans les domaines de l'ouverture au monde, à la vie sociale et civique et la connaissance de soi.

Contenu de la matière

6 langues optionnelles sont proposées par l'école :

- Allemand
- Chinois
- Espagnol
- Italien
- Japonais
- Russe

Le contenu et les modes de contrôle des connaissances sont décrits dans le fascicule « *Programme des enseignements : Langues.* »

Forum des entreprises

Ensaï Business Forum

Correspondant : Patrick GANDUBERT

Objectif de la matière

Le Forum des entreprises est organisé chaque année en novembre par l'association Forum Ensaï.

Environ 70 entreprises représentant la bancassurance, le conseil, l'industrie, la santé, les services numériques... participent à ce rendez-vous majeur de l'année.

Le Forum permet notamment :

- de se renseigner sur les nombreux débouchés qu'offre la formation d'ingénieur statisticien/data scientist
- de rencontrer des alumni
- de participer à des simulations d'entretiens avec Ensaï Alumni
- de trouver un stage

Rendez-vous Carrières

Career Rendez-Vous

Enseignant : Divers intervenants
Correspondant : Patrick GANDUBERT

Objectif de la matière

En complément du Forum des entreprises, les Rendez-Vous Carrière permettent aux entreprises partenaires de l'Ensay de venir présenter leurs activités les métiers de la data aux étudiants de 2^e année.

Les Rendez-Vous Carrière ont lieu entre janvier et mars.

Les étudiants choisissent au moins 5 Rendez-Vous Carrière parmi la vingtaine proposée.

Ils ont pour vocation :

- de mieux connaître le métier d'ingénieur statisticien/data scientist dans différents secteurs d'activité
- d'éclairer le choix de filière de spécialisation
- de commencer à se constituer un réseau professionnel
- de trouver un stage

Le format :

10 minutes : présentation corporate de l'entreprise/organisation

30 minutes : la statistique dans mon entreprise/organisation

- un exemple précis et concret d'application du métier d'ingénieur statisticien (chargé d'étude, data miner, data analyst, data scientist...) dans l'entreprise/organisation
- les qualités recherchées (hard & soft skills)
- l'œil du pro : un éclairage professionnel de l'intervenant lié à son propre parcours

20 minutes : échange, discussion

Matière 2AOUT01

R-Shiny

Cours : 4h30

Enseignant : Matthieu MARBAC (Ensay)

Correspondant : Matthieu MARBAC

Objectif de la matière

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

Contenu de la matière

Initiation au développement d'application R-Shiny.

Pré-requis

Programmation en R (1A)

Contrôle des connaissances

Aucun

Langue d'enseignement

Français

Matière 2AOUT02

Visual Basic Application

Visual Basic Programming Using Excel

Cours : 6h

Enseignant : Steven GOUICHOUX (Insee)

Correspondant : Mohamed GRAIET

Objectif de la matière

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.

Contenu de la matière

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 demie-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 demie-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.

Documents pédagogiques

Support de cours

Pré-requis

Les concepts algorithmiques abordés dans le cours algorithmique et programmation de 1^{ère} année doivent être maîtrisés.

Contrôle des connaissances

Aucun

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/>

Langue d'enseignement

Français

Matière 2AOUT03

Compléments de SAS

Advanced SAS

Cours : 1h30 h • TP : 9 h

Enseignant : A déterminer (Ensaï)

Correspondant : A déterminer

Objectif de la matière

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.

Contenu de la matière

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

Documents pédagogiques

Support du cours.

Pré-requis

SAS et langage SQL (1A)

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

Langue d'enseignement

Français

Matière 2AOUT04

Libre Office Basic

Libre Office Basic

Cours/TP : 6h

Enseignant : Samuel TOUBON Insee

Correspondant : Mohamed GRAIET

Objectif de la matière

Découvrir et comprendre l'environnement et la logique de développement en Basic pour LibreOffice Calc. Être capable de développer des applications dans ce langage pour automatiser des traitements. Savoir rechercher et utiliser l'information nécessaire dans ce domaine.

Contenu de la matière

1. Le langage de programmation Basic :
 - a. variables, structures conditionnelles et boucles, procédures et fonctions
 - b. les structures de données usuelles : tableaux et matrices
 - c. les formats, la gestion des dates
 - d. le traitement des erreurs
2. L'éditeur LibreOffice :
 - a. les boîtes de dialogue et les formulaires
 - b. exécution et débogage du code Basic
3. L'API LibreOffice :
 - a. manipulation des classeurs, feuilles, cellules
 - b. l'utilisation de formules de Calc dans Basic et de fonctions Basic dans Calc
4. L'accès aux bases de données

Contrôle des connaissances

Cet enseignement ne fait pas l'objet d'un contrôle des connaissances.

Références bibliographiques

Guide de programmation de OpenOffice.org Basic :

https://wiki.openoffice.org/wiki/Documentation/BASIC_Guide

Mémento sur la manipulation des feuilles de calcul :

https://wiki.openoffice.org/wiki/Spreadsheet_common

Langue d'enseignement

Français

Enseignements spécifiques aux élèves attachés

Matière

Module spécifique des attachés

Enseignant : Divers intervenants

Correspondant : Ronan LE SAOUT

Objectif de la matière

L'objectif de ce module est de donner des compléments de formation aux élèves statisticiens publics, utiles pour leur prise de poste et le suivi du master "évaluation et décision publiques".

Cette période de 2 semaines (juin) sera dédiée à 3 points:

- Une formation obligatoire au premier secours (PSC1, 1 journée);
- Des séminaires d'1h30 à 3h présentant des méthodes innovantes en statistique publique et les meilleurs mémoires de master "évaluation et décisions publiques";
- La réalisation d'un projet professionnel en autonomie (par groupe de 5 ou 6 élèves) sur des thèmes en production statistique, méthodologie, études économiques ou informatique. Pour ces projets, il est demandé de répondre à une commande, après un échange avec la personne proposant le sujet. Le travail est ensuite effectué en autonomie, sans encadrement spécifique. Une soutenance est effectuée le dernier jour devant l'ensemble des élèves et personnels impliqués.

Il n'y a pas de note associée à ces travaux mais des ECTS sont attribués. La présence et l'assiduité sont donc obligatoires. La validation de ce module est une condition de titularisation.

Liste de sujets proposés en 2019

Aspect géographique de l'analyse des données à la française

Revue de littérature sur le multimode

Travail méthodologique autour du multimode

Le temps de travail dans la fonction publique territoriale (revue de littérature et analyse méthodologique)

Appariement déterministe d'une enquête et d'un fichier fiscal

Serious Game (gestion du temps et de projet)

Grammaire appliquée de la datavisualisation

Maquette technique du nouveau intranet de l'Insee

Langue d'enseignement

Français