

**ENSEIGNEMENTS  
DU DISPOSITIF DE LA**

**FORMATION CONTINUE  
DIPLOMANTE DES ATTACHÉS**

**PROGRAMME  
2011-2012**



---

# Sommaire

---

	<i>pages</i>
<b>Modélisation économique</b>	<b>5</b>
↳ Economie du travail et de l'emploi	6
↳ Economie industrielle	7
↳ Modélisation macroéconomique	8
↳ Analyse de la conjoncture	9
↳ Économétrie sur données de panel	10
↳ Economie de la santé	11
<b>Economie et sciences sociales</b>	<b>12</b>
↳ La famille contemporaine	13
↳ L'éducation	14
↳ Les entreprises	15
↳ L'Union européenne, l'Etat et les collectivités locales	16
↳ Techniques et pratique de la démographie	17
<b>Méthodes d'enquête</b>	<b>19</b>
↳ Nomenclatures et codage	20
↳ Traitement de la non-réponse	21
↳ Enquêtes répétées dans le temps	22
↳ Collecte d'informations économiques et sociales	23
↳ Redressement d'échantillons et méthode de calage	24
↳ Techniques d'échantillonnage	25
↳ Estimation sur petits domaines	26
<b>Analyse statistique</b>	<b>27</b>
↳ Analyse exploratoire des données	28
↳ Économétrie des durées	29
↳ Analyse spatiale	30
↳ Analyse des données approfondie	31
↳ Analyse et traitement des séries temporelles	32
↳ Compléments de séries temporelles	33
↳ Compléments de modèle linéaire	34
↳ Régression sur données catégorielles et données de comptage	35
↳ Théorie des indices de prix	36
↳ Approche non standard en régression	37
↳ Statistique bayésienne	38
<b>Informatique</b>	<b>39</b>
↳ Compléments de bases de données	40
↳ Compléments de SAS	41
↳ Datawarehouse - datamining	42
↳ Réseaux et sécurité	43
↳ XML	44

<b>Environnement professionnel</b>	<b>45</b>
↳ <b>Anglais</b>	<b>46</b>
↳ <b>Analyse des organisations</b>	<b>47</b>
↳ <b>Management et administration de projets</b>	<b>48</b>
↳ <b>Ecrit &amp; oral : les techniques de l'exposé</b>	<b>49</b>

# MODELISATION ECONOMIQUE

Les enseignements du domaine «*Modélisation économique*» portent sur différents domaines d'application de la modélisation économique, que celle-ci s'effectue à partir de données individuelles ou agrégées. Les thèmes abordés sont le travail, l'emploi et les stratégies d'entreprises. Deux modules sont consacrés aux aspects méthodologiques de la modélisation : élaboration d'informations conjoncturelles, estimation de modèles synthétiques.

↳ Economie du travail et de l'emploi	6
↳ Economie industrielle	7
↳ Modélisation macroéconomique	8
↳ Analyse de la conjoncture	9
↳ Econométrie sur données de panel	10
↳ Economie de la santé	11

## Economie du travail et de l'emploi

### Objectif :

L'objectif de ce module est d'analyser les causes et les solutions apportées au problème du chômage, à partir de la présentation des principaux apports des théories récentes du marché du travail et de l'étude de cas ou d'articles récents relatifs à cette littérature.

Une introduction présentera les définitions et les mesures du chômage. Le développement de l'analyse du chômage se déroulera autour de cinq thèmes, articulés autour des débats actuels (voir ci-dessous).

### Contenu :

#### **Evolutions récentes et disparités internationales en matière d'emploi :**

Quelles sont les performances et principales différences entre les pays de l'OCDE en matière d'emploi ?

On présentera les disparités en matière de taux d'emploi, de travail à temps partiel, de productivité, de chômage et de protection sociale liée aux différents statuts au regard de l'emploi. Ces disparités renvoient-elles à différents régimes d'emploi ?

#### **Demande de travail et nature du chômage :**

Cette partie présentera les principaux débats théoriques relatifs aux causes du chômage. Seront alors abordés d'une part les approches en termes de « rigidités » du marché du travail, d'autre part les problèmes de coordination des politiques économiques, notamment au sein de l'Union Européenne.

#### **Les politiques d'emploi : modèles nationaux et coordination européenne :**

Cette partie présentera dans un premier temps une typologie des mesures de lutte contre le chômage, telles qu'elles existent dans les principaux pays de l'OCDE. On différenciera dans le même temps des modèles nationaux de politique de l'emploi, en fonction des spécificités nationales. Dans un second temps, on proposera une synthèse critique des principales méthodologies utilisées pour évaluer l'efficacité de ces interventions, ainsi qu'un bilan des résultats existants. On discutera également des perspectives de coordination des politiques de l'emploi dans le cadre de la Stratégie Européenne pour l'Emploi

#### **L'analyse de l'emploi et du chômage en termes d'offre de travail :**

Les différences de comportements d'offre de travail éclairent-elles les disparités en matière d'emploi ? Comment analyser les différences entre hommes et femmes en matière d'offre de travail ? Quels sont les effets de la protection sociale et des politiques visant à 'rendre le travail payant' sur l'offre de travail et la recherche d'emploi ?

Cette partie présentera les développements récents des modèles d'offre de travail pour étudier les effets des politiques économiques (PPE, politiques ciblées sur les seniors...) sur la participation à l'emploi, le volume de travail tout au long de la vie et le choix d'occupation.

Il envisagera également des questions d'ordre méthodologique, en particulier l'utilisation de méthodes de microsimulation pour les analyses ex ante de l'impact des politiques publiques.

## Economie industrielle

### Objectif :

Ce module a pour objet de fournir aux participants des clés d'interprétation des motifs économiques justifiant les choix stratégiques et organisationnels des entreprises.

Différents rappels concernant la théorie de l'organisation industrielle (notions de concurrence pure et parfaite, de pouvoir de marché...) seront faits. On examinera la question de la détermination des prix, du choix des produits, de la localisation industrielle, des barrières à l'entrée. En exemple, les questions actuelles relatives à l'économie des industries à rendements croissants seront traitées et l'économie des biens d'information sera présentée.

### Pré-requis :

Ce module requiert une bonne connaissance du cours de microéconomie enseigné en 1ère année à l'Ensaï.

### Contenu :

#### Rappels de la théorie de l'organisation industrielle

- Rappels sur la théorie de la firme
- Le monopole
- Le pouvoir de marché
- Introduction à la théorie des jeux et étude de l'oligopole

#### Concentration horizontale et politique de la concurrence

- Notion de marché pertinent
- Analyse économique des concentrations horizontales
- La politique de la concurrence européenne

#### Analyse économique des relations verticales

- Intégration et désintégration verticale
- Distribution et restriction verticale

#### La collusion explicite, la collusion tacite et la prédation

- Parallélisme de comportement, collusion tacite, collusion explicite
- Facteurs favorisant la collusion
- Comportement prédateurs et autres comportements anti-concurrentiels

#### La concurrence dans les industries de réseau

- Caractéristiques économiques d'un réseau (externalités de réseau, verrouillage, etc.)
- Équilibres et dynamiques d'un réseau
- Guerre de standards
- Stratégies dans les industries en réseau

#### Economie des biens d'information

- Spécificités des biens d'information
- Tarification, versioning, bundling, marchés bi-faces
- Numérique et biens d'information

## Modélisation macroéconomique

### Objectif :

L'objet de ce module est d'initier les participants à la méthodologie de la modélisation macroéconomique et à ses développements actuels. Il rappellera les acquis de la modélisation macroéconomique quantitative et ses liens avec les techniques des séries temporelles multidimensionnelles (modèles VAR, VECM ...). Il abordera les problèmes de la dynamique macroéconomique et les fondements des cycles conjoncturels tels les cycles endogènes, la dynamique des prix et les cycles d'accumulation des stocks et du capital. Ce module consacrera une part importante à la construction et à la comparaison de modèles opérationnels permettant la maîtrise de variantes dans un but d'analyse de la politique macroéconomique ou de prévision.

### Pré-requis :

Ce module requiert la compréhension du cours de Macroéconomie enseigné en 1ère année à l'Ensaï.

### Contenu :

#### Introduction à la macro-économie

- 1) **Rappels de macro-économie** : les controverses en macro-économie ; le modèle IS-LM ; le modèle de la synthèse ; modélisation macro-économétrique "à l'ancienne".
- 2) **Le « renouveau » de la macro-économie** : la critique de Lucas ; les anticipations rationnelles ; fondements micro-économiques de la macro-économie ; les problèmes de l'agrégation et la notion d'agents représentatifs ; choix inter-temporels et programmes des agents.

#### Concepts de séries temporelles :

- 1) Rappels de séries temporelles ; présentation de la méthodologie VAR et VECM : critique de Sims ; estimation ; prévision ; causalité au sens de Granger ; fonctions impulsion-réponse (IRF) ; VAR structurel..
- 2) Applications : estimation d'équations de comportement sous la forme VECM (consommation des ménages, échanges extérieurs)

#### Présentation de modèles opérationnels de la macroéconomie

- 1) **Les modèles macro-économétriques : présentation générale**
  - Histoire et théorie
  - Les propriétés des modèles macro-économétriques
- 2) **Présentation détaillée des modèles Opale et Mésange :**
  - Les hypothèses structurelles
  - Les équations du modèle.
- 3) **Un point de vue opérationnel :**
  - Estimation
  - Prévision
  - Utilisation en variante : chocs exogènes, politique budgétaire et fiscal, réformes structurelles

#### Les modèles de cycles réels et monétaires

- 1) **Le modèle canonique des cycles réels** : la quantification du phénomène cyclique ; le modèle de référence ; résolution du système dynamique à anticipations rationnelles.
- 2) **Vers une nouvelle synthèse néoclassique** : les fluctuations de l'emploi ; autres voies d'approfondissement.

## Analyse de la conjoncture

### Objectif :

Les techniques de l'analyse conjoncturelle sont destinées à la prévision de l'évolution économique dans les six à huit prochains mois à partir d'informations actuelles ou issues d'un passé proche. Ce module attirera l'attention sur les erreurs à éviter dans l'observation du passé récent. Il dressera un état des données disponibles en France et à l'étranger et des instruments (enquêtes conjoncturelles). Ce module fera une large place à l'application des méthodes exposées au cas de la conjoncture française et internationale du moment.

### Pré-requis :

La révision des cours de Comptabilité Nationale et de Macroéconomie est recommandée.

### Contenu :

#### Les sources de l'information conjoncturelle :

- **Les indicateurs conjoncturels quantitatifs**

Les indicateurs conjoncturels quantitatifs sont des statistiques économiques, mensuelles ou trimestrielles, issues d'enquêtes ou résultant de l'activité administrative.

Ces indicateurs permettent, après traitements appropriés (CVS,...) une appréciation quantitative de la situation économique. Ils sont souvent utilisés sous forme de graphiques ou de « tableaux de bord ».

- **Les comptes trimestriels**

Présentation et utilisation par le conjoncturiste.

- **Les enquêtes de conjoncture**

- Methodologie

- Avantages

- Interprétation

- Utilisation pour la prévision à court terme

#### Principes de l'analyse conjoncturelle

- **La synthèse de l'information conjoncturelle**

On présentera l'analyse et la prévision conjoncturelle par grands domaines, avec l'application à une analyse en vraie grandeur pour l'économie française et les principaux secteurs : revenus, épargne, consommation, échanges extérieurs, production de l'industrie et du BTP, prix, emploi et chômage ; taux d'intérêt et taux de change.

- **La conjoncture internationale**

Les données et leur interprétation

La détermination de la demande mondiale vers la France

- **La conjoncture monétaire et financière**

Les données et leur interprétation

Analyse des conséquences des fluctuations financières sur les principales économies industrialisées.

## Econométrie sur données de panel

### Objectif :

L'objectif de cette formation est de fournir les outils utiles à l'estimation de modèles économétriques sur données de panel. Les méthodes et leur mise en application pratique seront décrites. Des exemples numériques ainsi que des travaux de recherche empiriques seront présentés tout au long de la formation pour illustrer l'utilisation de ces méthodes.

### Pré-requis :

La révision des méthodes économétriques usuelles sur données en coupe transversale (méthode des moindres carrés ordinaires, méthode des variables instrumentales, méthode des moindres carrés généralisés) est recommandée.

### Contenu :

#### 1 Introduction

- 1.1 Illustration
- 1.2 Définitions
- 1.3 Notations
- 1.4 Ecritures du modèle de panel
- 1.5 Les opérateurs

#### 2 Modèle linéaire statique

- 2.1 Le modèle à effets fixes
- 2.2 Le modèle à effets aléatoires

#### 3 Le modèle linéaire dynamique

- 3.1 La Méthode des Moments Généralisés : rappel
- 3.2 Le modèle dynamique de panel : spécification et hypothèses
- 3.3 L'estimation du modèle dynamique de panel
- 3.4 Exemples

#### 4 Extensions

- 4.1 Double effet, individuel et temporel
- 4.2 Panel non cylindré
- 4.3 Modèles de panel non linéaires
- 4.4 Systèmes de régression avec erreurs composées

## **Economie de la santé**

### **Objectif :**

A partir d'un rappel du fonctionnement du système de santé français, du financement à la description des acteurs, nous nous focaliserons sur l'analyse du comportement de certains acteurs.

Après avoir rappelé les enjeux théoriques, l'accent sera mis sur l'estimation économétrique des comportements individuels.

### **Contenu :**

#### **1 Rappels institutionnels : l'organisation du système de santé français**

#### **2 Les grandes sources d'information statistique en économie de la santé**

#### **3 Equité et discrimination en santé**

#### **4 La demande de santé :**

- Demande de santé, demande de soins ;
- La mesure de l'état de santé ;
- Estimation de la demande de soins ;
- Estimation de la demande de santé.

#### **5 L'assurance santé**

##### **Concepts en assurance-santé**

- Tarification actuarielle et anti-sélection ;
- Sélection des risques ;
- Sélection adverse ;
- Risque moral :
  - a. Le risque moral ex ante ;
  - b. Le risque moral ex post ;
  - c. Le risque moral ex post combiné à la demande induite.

##### **Estimation des phénomènes d'aléa moral et de sélection adverse**

#### **6 L'offre de soins ambulatoire**

##### **La rémunération des médecins libéraux : aspects incitatifs**

- Le paiement à l'acte ;
- La capitation ;
- Le case-mix.

##### **Estimation des comportements des médecins libéraux français**

#### **7 L'offre de soins hospitalière**

- Production et fonction de production des établissements de soins (introduction) ;
- Règles de paiement et stratégies des offreurs de soins ;
- Exploitation des données PMSI.

# ECONOMIE ET SCIENCES SOCIALES

L'objet des enseignements du domaine « *Economie et sciences sociales* » est d'initier les participants à l'étude d'un certain nombre d'institutions comme l'école, la famille, les entreprises ou l'Etat en faisant appel aux outils et aux méthodes développés par les sciences sociales. Il s'agira de repérer les zones d'accord ou de désaccord entre disciplines, principalement l'économie et la sociologie, tout en s'interrogeant sur la place de la régulation juridique dans la construction et le fonctionnement de ces institutions.

↳ <b>La famille contemporaine</b>	<b>13</b>
↳ <b>L'éducation</b>	<b>14</b>
↳ <b>Les entreprises</b>	<b>15</b>
↳ <b>L'Union européenne, l'Etat et les collectivités locales</b>	<b>16</b>
↳ <b>Techniques et pratique de la démographie</b>	<b>17</b>

## **La famille contemporaine**

### **Objectif :**

Partant de l'évolution de la définition de la catégorie de « ménage » en usage dans les enquêtes statistiques et d'un rapide inventaire de la modification des dispositions qui dans les différentes branches du système juridique se préoccupent de la famille, ce module examinera les transformations qui marquent la vie familiale dans les sociétés contemporaines : élargissement du champ d'intervention de l'Etat et de la société, émergence de modèles familiaux qui s'éloignent du modèle traditionnel fondé sur la stabilité du lien conjugal, la division hiérarchique des rôles et la dépendance entre les générations. Ce module présentera les résultats de la sociologie et de l'économie appliquées à la famille qui éclairent le mieux ces transformations.

### **Pré-requis :**

Une bonne connaissance de la théorie microéconomique du consommateur et de l'histoire de la pensée sociologique est recommandée.

### **Contenu :**

#### **La famille : histoire et démographie**

- Présentation des doctrines démographiques
- La démographie historique et ses sources (registres paroissiaux / recensements / Enquêtes de démographie historique)
- Régimes démographiques et formes familiales :
  - \* La famille avant la transition démographique
  - \* Conséquences de la transition démographique sur la famille

#### **Changements familiaux et des politiques familiales**

- Comment décrire une politique familiale (approche historique, outils, effets...)
- Quelques éléments de comparaison européenne

#### **La famille contemporaine : un cadrage socio-démographique**

- Approche statistique de la famille à travers principalement le recensement, l'état civil, les « enquêtes familles »
- La famille contemporaine : les étapes de la vie (accents sur certains thèmes, tels que formation des unions et arrivée des enfants, grand parentalité, etc)
- Changements et continuités

#### **Introduction au droit de la famille**

- Les différentes formes de famille
- Mariage, concubinage
- Divorce et séparation
- La détermination des filiations
- Les droits de l'enfant

#### **Ménages et famille dans les pays du Sud (accent sur l'Afrique subsaharienne)**

- Quelle adéquation des notions de famille et de ménage ?
- Les traits généraux des systèmes familiaux en Afrique
- Etudier la famille dans les pays du Sud : sources et données...sur les ménages
- Apports de la démographie du ménage et de la famille
- L'évolution de la famille en Afrique : entre permanences et changements

## L'éducation

### Objectif :

Le sommet de Lisbonne a promu le lifelong learning comme objectif commun aux pays de l'union pour faire de l'Europe la région la plus compétitive du monde. L'école est à la fois une institution, une organisation, un espace d'actions. Elle a des objectifs fixés par la loi, des ressources fournies par la collectivité, et avec les secondes elle offre des services pour atteindre les premiers. Quels sont-ils ? Comment les atteint-elle ? Comment marie-t-elle éducation et instruction, socialisation et sélection des jeunes, transmission d'une culture et préparation à la vie professionnelle ? Comment évoluent les inégalités personnelles, géographiques et sociales devant l'école, et quelle est la contribution de cette dernière à la réduction des inégalités sociales ? En présentant les méthodes utilisées pour décrire les trajectoires scolaires, ce module montrera la portée sociologique des enquêtes de démographie scolaire et soumettra aux participants quelques-unes des questions posées par la sociologie et l'économie de l'éducation. On s'efforcera dans l'analyse des carrières et des résultats scolaires de départager le rôle respectif du milieu d'origine, des enseignants et de l'institution. Des pistes de recherche seront suggérées aux participants : étude de l'élaboration des programmes scolaires, observation des stratégies des élèves, des parents et des enseignants, analyse des critères sur lesquels repose l'évaluation des élèves ou des étudiants, estimation de la rentabilité, interne et externe, de l'investissement scolaire selon une optique économique ou encore examen des liens entre l'Etat et l'institution scolaire en adoptant une démarche historique.

### Contenu :

#### Les objectifs de l'école en France et les débats actuels

Introduction : La situation française : une loi d'orientation, des questions et un débat en cours.

1. Quatre objectifs généraux : transmettre des connaissances ; préparer à la vie professionnelle ; préparer à vivre ensemble dans une démocratie ; être équitable.
2. Eléments de diagnostic quant à l'atteinte de ces objectifs, des points en débat
3. Pérennité et universalité des objectifs : éléments de comparaison internationale et temporelle. Objectifs de l'école et demande sociale.
4. La mise en œuvre concrète des objectifs : orientations nationales et déconcentration / décentralisation ; éléments de comparaisons internationales à ce sujet ; les forces et les faiblesses de l'évaluation et du pilotage de notre système éducatif ; les personnels et les cadres de l'éducation ; spécificités françaises.

Conclusion : rendre compte du fonctionnement et des résultats du système éducatif et le piloter.

#### Problématique des inégalités

1. L'avancée internationale de la recherche sur les inégalités devant l'enseignement - 2. Distinguer entre massification et démocratisation de l'école : l'intérêt des odds ratios - 3. Modèles statistiques pour étudier l'évolution temporelle des inégalités sociales devant l'école - 4. L'évolution temporelle des inégalités sociales devant l'école en France depuis le début du XXe siècle - 5. Un regard plus large sur les inégalités de parcours scolaires des élèves en fonction de leur environnement familial : les enseignements des panels d'élèves - 6. Théories explicatives des inégalités sociales devant l'école.

**Les chiffres clés de l'Education Nationale :** 1. Des indicateurs pour un état des lieux du système éducatif : niveau national : « Etat de l'école » ; niveau régional : « Géographie de l'école » ; niveau établissement : « IPES » et « IVAL » (+ INPEC et INFOSUP). 2. Trois domaines : moyens et personnels, activité et scolarisation, résultats. 3. L'Education nationale : quelques grands chiffres

#### Le compte de l'Education

Le compte satellite de l'Education : la structure générale du compte, les sources utilisées, la Dépense Intérieure d'Education (DIE) et les dépenses moyennes par élève et par niveau scolaire, les comparaisons internationales.

#### La formation continue

1. Position de la France en Europe en ce qui concerne la formation continue des adultes : les définitions européenne et leur discussion ; les chiffres du module ad hoc de la LFS.
2. Première enquête française sur la formation continue post-scolaire : AES 2006, les différentes manières d'apprendre et la problématique des inégalités ; la réforme : l'accord interprofessionnel de 2003 et la loi de 2004 relative à la formation tout au long de la vie, changements et permanence ; l'ANI 2009 et la loi qui lui fait suite ; présentation des grandes enquêtes européennes en cours CVTS4, AES et du croisement de leur version française (DIFES 2).

## Les entreprises

### Objectif :

Ce module s'attachera à décrire la complexité de l'environnement dans lequel évoluent les entreprises et leurs transformations dans les vingt dernières années (introduction de nouvelles techniques, développement des pôles de recherche et d'ingénierie et des fonctions financières, changements rapides des produits et des services offerts, amélioration des moyens d'information économique, internationalisation de la production, des réseaux de distribution et des capitaux). Cette approche descriptive sera complétée par une analyse de la structure juridique qui permet à une entreprise d'entretenir des rapports avec la puissance publique (fiscalité, normes comptables, politique industrielle, droit de la concurrence), de nouer des liens avec les membres de son personnel, ses fournisseurs, ses clients, et de réunir la masse des capitaux nécessaires à la production.

### Contenu :

#### Stratégie de la firme et management

- Le concept de stratégie
- Le diagnostic stratégique
- Les outils d'aide au choix
- Les options stratégiques

#### Droit des entreprises

- Les différentes structures de société
- Leur régime fiscal
- Le statut des dirigeants
- La constitution des sociétés
- L'entreprise en difficulté

#### Analyse des coûts et rentabilité

- Les coûts : consommations de la valeur ajoutée
- Les méthodes de calcul des coûts et le problème majeur de la répartition des charges indirectes
- L'analyse de la rentabilité :
  - 1 - les coûts variables
  - 2 - le seuil de rentabilité

#### Financement de l'entreprise

Introduction : le circuit financier

- La politique d'investissement : choix des investissements
  - 1 - détermination des flux nets de trésorerie
  - 2 - critères de rentabilité
- La politique de financement
  - 1 - présentation des différents modes de financement
  - 2 - sélection
- Les prévisions financières
  - 1 - différentes prévisions fonction de l'horizon
  - 2 - présentation du plan de financement

## L'Union européenne, l'Etat et les collectivités locales

### Objectif :

Les objectifs de cette formation sont de :

1. Comprendre, à travers l'étude de la construction communautaire, les transferts de compétences consentis par les Etats membres au profit de la Communauté européenne
2. Mesurer les enjeux de l'élargissement de l'Union européenne aux pays d'Europe centrale, orientale et baltique en 2004
3. Faciliter la compréhension du fonctionnement de l'Union européenne par la présentation synthétique du système institutionnel communautaire et par le suivi du processus décisionnel communautaire
4. Comprendre les enjeux actuels de la décentralisation
5. Connaître les évolutions récentes ou prochaines
6. Percevoir les impacts de ces évolutions sur les réalités territoriales et professionnelles
7. Présenter l'organisation des systèmes publics d'informations statistiques au sein de l'Union Européenne
8. Comprendre l'ordre juridique communautaire (les principales sources du droit communautaire et les principes régissant les rapports entre le droit communautaire et le droit national) et évaluer l'implication des autorités nationales dans la mise en œuvre du droit communautaire.

### Contenu :

#### **Les étapes décisives de la construction communautaire et les enjeux de l'élargissement à l'Est :**

La construction communautaire de la CECA (1951) au traité de Maastricht (1992) ; la réforme de l'UE dans la perspective de l'élargissement à l'Est : les traités d'Amsterdam (1997), de Nice (2000) et de Lisbonne (2007).

**Les principaux acteurs du droit communautaire – Le processus décisionnel :** Commission européenne ; Conseil de l'UE ; Parlement européen ; Cour de Justice et Tribunal de Première Instance ; Processus décisionnel ; Réforme des institutions.

**L'ordre juridique communautaire :** les principales sources du droit communautaire (traités, règlements, directives, décisions et autres actes) ; les grands principes du droit communautaire : primauté et effet direct ; l'implication des autorités nationales dans la mise en œuvre du droit communautaire.

**L'acquis statistique communautaire :** historique du développement des statistiques communautaires, les débats d'actualité dans le domaine statistique.

**La gouvernance du système statistique européen :** rôle des institutions européennes dans la gouvernance du système statistique européen ; relations de négociation avec les Etats membres.

**Le système administratif français :** de la centralisation à la déconcentration/décentralisation.

**La décentralisation - nouvelle donne :** Les grandes étapes de la décentralisation : de l'acte 1 (les lois de 1982-1983) à l'acte 2 (la révision constitutionnelle de 2003 et la loi du 13 août 2004).

**La répartition des compétences et leur évolution :** Compétences transférées par la loi du 13 août 2004 aux collectivités territoriales et aux EPCI.

**La nouvelle organisation territoriale :** Le maillage du territoire français par de l'intercommunalité et le développement des territoires de projet (pays, agglomérations, parcs naturels, ...).

## Techniques et pratiques de la démographie

### Objectif :

Les objectifs de cette formation sont de :

- 1- Maîtriser les concepts et outils de la démographie. Il s'agira notamment de revoir les indicateurs de base de la démographie (fécondité, espérance de vie, quotients migratoires...), leurs limites et intérêts, et de les calculer en s'appuyant sur des bases de données disponibles à l'Insee. L'accent sera aussi mis sur la connaissance des tendances démographiques afin de mieux comprendre le sens à donner à un résultat et de mieux situer un territoire.
- 2- Etre sensibilisé à la place de la démographie dans le débat public (impact du vieillissement, impact sur la ressource en main d'œuvre, le développement économique...). La prospective démographique, outil essentiel dans la réponse aux préoccupations des acteurs publics, sera présentée.
- 3- Construire une analyse démographique en s'appuyant sur des cas pratiques : bilan démographique, analyse de cohortes (utilisation de l'échantillon démographique permanent). Cette étape inclura à la prise en main de fichiers, le calcul d'indicateurs, puis l'analyse et la présentation des résultats.

### Pré-requis :

Une bonne connaissance des notions démographiques est recommandée.

### Contenu :

#### 1. Les grandes tendances démographiques

Phases d'évolution - la transition démographique  
France, Union européenne, Monde

#### 2. Les migrations internationales

Grandes phases.  
Conséquences économiques et sociales : pays d'accueil et d'origine.  
Cas français

#### 3. Les migrations internes

Mesures  
Grandes phases  
Conséquences économiques et sociales

#### 4. Les projections

Présentation des principes des modèles.  
Principales problématiques des acteurs publics.

#### 5. Démographie et société

Mesure et impact :  
· du vieillissement de la population (équilibre des retraites, services à la personne, ...).  
· des comportements d'activité (niveau de la ressource en main d'œuvre, adéquation offre-demande, politique de l'emploi...).  
· de la réduction de la taille des ménages (politique du logement...).

#### 6. Cas pratiques

· Construction d'un bilan démographique et d'une analyse longitudinale (exploitations issues de l'échantillon démographique permanent)  
· Détermination des causes de migrations tout au long d'un cycle de vie



# METHODES D'ENQUETE

Centrés sur la production statistique, les modules du domaine « *Méthodes d'enquête* » permettent de comprendre et de maîtriser les différentes phases de l'élaboration d'une enquête : construction de nomenclatures, conception de questionnaires, choix de la technique d'échantillonnage, traitement et prévention des non-réponses. Ce domaine d'enseignement ouvre aussi sur les méthodes de collecte fondées sur les sources administratives ou réglementaires.

↳ Nomenclatures et codage	20
↳ Traitement de la non-réponse	21
↳ Enquêtes répétées dans le temps	22
↳ Collecte d'informations économiques et sociales	23
↳ Redressement d'échantillons et méthode de calage	24
↳ Techniques d'échantillonnage	25
↳ Estimation sur petits domaines	26

## Nomenclatures et codage

### Objectif :

La construction d'une nomenclature, sa mise en oeuvre à travers des opérations de classement et de codage puis son usage pour interpréter des tableaux statistiques sont des moments essentiels du travail du statisticien. On étudiera les nomenclatures économiques et plus spécialement les nomenclatures d'activités et de produits, la nomenclature socioprofessionnelle et d'autres exemples, en analysant quelques problèmes soulevés aux divers moments de la construction et de l'usage de ces outils statistiques. Des éléments de comparaisons entre la France et d'autres pays européens seront présentés, pour le cas des nomenclatures de professions et de groupes sociaux, souvent différentes.

### Contenu :

#### La nomenclature socioprofessionnelle française (et les CITP de l'ONU et d'Eurostat)

- Historique
- Catégories retenues
- Principes de classement adoptés lors du passage des CSP aux PCS (début des années 80)
- Principales évolutions des PCS lors des vingt dernières années :
  - procédure de chiffrage automatique (QUID, SICORE)
  - position de l'INSEE face à la remise en cause de la nomenclature dans les années 90)
  - harmonisation européenne des nomenclatures socio-économiques

#### Théorie des nomenclatures

- Présentation des principales règles et définitions.
- Critères discriminants et structurants

#### Les nomenclatures d'activités et de produits

- Construction d'un système standard international et européen intégrant les échanges (douaniers), le système productif (branches, secteurs d'activités et unités) et les produits (enquêtes de production, marchés) : le réseau des nomenclatures économiques et les nomenclatures fonctionnelles
- La gestion d'un réseau de nomenclatures
- Unités administratives et statistiques : un autre réseau
- Classement d'une unité selon l'activité principale exercée (APE) (indépendamment de la nomenclature d'activités)
- Les nomenclatures fonctionnelles : COICOP pour la consommation des ménages et l'indice des prix ; COFOG pour les dépenses des administrations ; COPNI ; COPP...
- Caractéristiques des nomenclatures actuelles d'activités et de produits
- Les révisions des nomenclatures de 2007 : analyse des nouvelles structures et pointage des principaux changements de méthode, de structure et de frontières

#### Sémantique des nomenclatures

- Les méthodes textuelles d'accès aux nomenclatures
- La sémantique dans les nomenclatures
- Les traitements linguistiques (principes, procédures, recherche, codages automatique et assisté)
- Les outils

## Traitement de la non-réponse

### Objectif :

L'objectif est d'exposer aux participants les questions liées à l'analyse et au traitement des non-réponses. Outre une présentation de la formalisation classique des principales méthodes de traitement des non-réponses (modélisation, imputation, hot-deck, calage ...), l'enseignement intégrera divers aspects méthodologiques comme la prévention des non-réponses. Des études de cas sur des applications en taille réelle compléteront la formation théorique. Le cours de deuxième année intitulé *Théorie des sondages* est une base indispensable pour aborder ce module. A la fin de la session, les participants devront être capables de résoudre des problèmes théoriques, méthodologiques et pratiques, relatifs aux problèmes posés par la non-réponse.

### Pré-requis :

Révision du cours de Théories des Sondages de 2<sup>ème</sup> année.

### Contenu :

#### Introduction

- Contexte ; définitions ; causes de la non-réponse ; exemples de non-réponse ; prévention

#### Théorie : 100% réponse

- Estimateur de Horvitz-Thompson ; estimateur par le régression généralisée ; l'échantillonnage à deux-phases

#### Mécanisme de réponse

- Mécanisme uniforme ; mécanisme ignorable vs non-ignorable

#### Vérification

- Types de règles ; localisation des erreurs

#### Options de traitement

- Utilisation des répondants complets seulement ; repondération ; imputation ; substitution

#### Repondération

- Repondération par l'inverse de la probabilité de réponse ; repondération par calage

#### Formation des classes de repondération

- Justification théorique ; méthode par croisement ; méthode des scores

#### Méthodes d'imputation

- Méthodes déterministes ; méthodes aléatoires

#### Estimation ponctuelle en présence de données imputées

- L'estimateur imputé ; biais de l'estimateur imputé ; biais quand les hypothèses ne sont pas valides ; distorsion des distributions

#### Formation des classes d'imputation

- Justification théorique ; construction des classes

#### Estimation de la variance : imputation

- Pourquoi estimer la variance due à la non-réponse et à l'imputation ; méthodes d'estimation des composantes de la variance ; autres méthodes

#### Estimation de la variance : repondération

- Variance des estimateurs ; méthodes

#### Autres sujets

- Imputation pondérée vs non pondérée ; moyenne d'un domaine ; coefficient de corrélation

## Enquêtes répétées dans le temps

### Objectif :

L'objectif visé est de fournir aux participants une présentation générale et opérationnelle des problèmes d'échantillonnage répétée dans le temps. L'accent sera mis sur la problématique des panels dans le domaine de la statistique publique et inclura, outre la formalisation classique de la théorie des panels, un examen approfondi des questions méthodologiques et des études de cas sur la base d'applications en taille réelle. A la fin de la session, les participants devront être capables de résoudre les problèmes théoriques, méthodologiques et pratiques, relatifs aux problèmes posés par les enquêtes répétées dans le temps.

### Pré-requis :

Révision du cours de Théories des Sondages de 2<sup>ème</sup> année.

### Contenu :

#### Introduction

- Concepts de base et définitions
- Avantages et désavantages des enquêtes répétées

#### Plans de sondage, échantillonnage, collecte des données

- Plans de sondage typiques
- Paramètres reliés aux plans de sondage
- Définition des unités longitudinales
- Utilisation de répertoires et autres bases de sondage
- Sélection de l'échantillon
- Conception du questionnaire
- Collecte des données : modes de collecte ; erreurs de mesure ; interviews dépendantes ; règles concernant les répondants ; dépistage ; incitatifs
- Types et causes de non-réponse

#### Pondération

- Pondération longitudinale et transversale
- Calcul du poids de base ; méthode du partage des poids ; traitement de la non-réponse ; calage aux marges ; poids intégré

#### Imputation

- Méthodes ; considérations ; impacts

#### Estimation & statistiques descriptives

- Estimation ; définitions des estimateurs ; tableaux de transition
- Estimation de la variance ; Jackknife & Bootstrap

#### Considérations analytiques

#### Contrôle de la divulgation et accès aux données

## Collecte d'informations économiques et sociales

### Objectif :

On étudiera les différents aspects que prend la collecte d'informations économiques et sociales au sein du système statistique public. En particulier, on analysera l'organisation choisie, du côté entreprises et du côté ménages. Si la collecte d'information se trouve relativement "cloisonnée", auprès des entreprises d'une part, auprès des ménages d'autre part, en partie en raison de l'organisation même du système statistique, cette séparation n'est pas toujours pertinente. Nous chercherons ainsi, au cours de cette formation, à mettre en évidence les points communs et les différences des contraintes et méthodes de la collecte d'information dans ces deux domaines.

La formation débutera par les aspects juridiques et déontologiques de la pratique du statisticien en quête d'information, en traitant notamment des dimensions législatives de l'exploitation de fichiers administratifs et de l'acquisition des données au moyen d'enquêtes, que l'unité concernée soit un ménage ou une entreprise.

Elle se poursuivra par la présentation des conditions de collecte par enquête statistique, et du travail d'élaboration d'un questionnaire, tout d'abord concernant les enquêtes auprès des entreprises, puis celles auprès des ménages. Des concepteurs d'enquête présenteront leurs travaux. Un exposé est également dévolu à la présentation du Recensement de la population.

### Contenu :

#### Déontologie et législation des enquêtes

- Le secret statistique et les textes fondateurs : Code ESOMAR, loi de 1951
- La nouvelle donne : la loi Informatique et libertés (1978) et sa modification en 2004
- La programmation des statistiques publiques : le Conseil national de l'information statistique (CNIS)
- Aspects déontologiques non totalement couverts par la loi

#### Conception de questionnaires « entreprises »

- Les sources d'information sur les entreprises
- Exemples qui serviront de fil conducteur
- Contexte institutionnel
- Champ de l'enquête et base de sondage
- Echantillonnage
- Mise au point du questionnaire et tests
- Collecte et contrôle de premier niveau

#### Conception de questionnaires « ménages »

- Les étapes préliminaires à la conception d'un questionnaire : objectif, population visée...
- Les règles de base de la conception d'un questionnaire : présentation, contenu des questions, questions ouvertes et questions fermées, ordre des questions, test du questionnaire, les erreurs et les pièges à éviter
- L'administration du questionnaire
- Les spécificités liées au mode de collecte : en face à face, par téléphone, par voie postale, avec assistance informatique
- Les spécificités liées à la nature de l'enquête : panels...
- Les méthodes d'évaluation des questionnaires
- Les interactions enquêteur-questionnaire-enquêté

#### Comparaison questionnaires ménages/entreprises : y-a-t-il vraiment des différences ?

#### Présentation de l'enquête « Statistiques sur les Ressources et les Conditions de Vie (SRCV) »

#### Présentation de l'enquête sur la production de déchets non dangereux dans l'industrie

#### Présentation du recensement de la population

#### Présentation des enquêtes de victimation

## Redressement d'échantillons et méthode de calage

### Objectif :

L'objectif de ce module est de présenter l'ensemble des méthodes de redressement d'échantillons reposant sur l'utilisation d'information auxiliaire : comment améliorer la précision d'estimateurs calculés à partir d'un échantillon, grâce à la connaissance de totaux ou d'effectifs connus sur la population. On présentera différentes méthodes d'estimation. **L'accent sera mis sur les méthodes de calage.** Les techniques de calage simultané sur différentes unités d'une même enquête (ex : ménages et individus) seront présentées. On abordera succinctement les liens entre calage (simple ou généralisé) et correction de la non-réponse.

Les participants utiliseront la macro SAS Calmar2 ("calage sur marges") sur des données réelles. A la fin de la session, les participants seront capables de choisir la méthode de redressement adaptée aux données dont ils disposent, et de la mettre en oeuvre informatiquement.

### Pré-requis :

La relecture, avant la formation, des chapitres suivants d'un cours de sondage est vivement recommandée : sondage aléatoire simple, stratification, sondage à probabilités inégales.

### Contenu :

#### Généralités, rappels en théorie des sondages

##### Estimation par le ratio

- Principe de la méthode.
- Définition de l'estimateur par le ratio.
- Propriétés : biais, erreur quadratique moyenne, estimation de l'EQM.
- Ratio par la taille.

##### Post-stratification

- Principe de la méthode.
- Définition (cas d'un sondage aléatoire simple).
- Propriétés : sans biais, variance, variance estimée, comparaison avec l'estimateur Horvitz-Thompson.

##### Estimation par régression

- Estimation par différence : définition, propriétés, exemple.
- Estimation par régression : définition, différentes expressions de l'estimateur par régression.
- Propriétés : espérance, variance, variance estimée.
- Exemples : le cas d'une seule variable auxiliaire.

##### Estimation par calage

- Le problème du calage.
- Aspects théoriques du calage.
- Fonctions de distance.
- Propriétés des estimateurs par calage.

##### La macro Calmar2

- Mise en œuvre de la macro..
- Un petit exemple de calage sur marges.

##### Calage et correction de la non-réponse

- Les méthodes de repondération.
- Calage direct sur l'échantillon des répondants.

##### Un exemple historique de calage

- Choix des variables de calage.
- Choix de la fonction de distance, comparaison des distributions de poids.
- Précision des estimateurs calés.

##### Calages simultanés pour différentes unités d'une même enquête

- Calage au niveau individu ; exemple de l'enquête emploi.
- Calage aux niveaux ménage et individu ; exemple de l'enquête sur la consommation alimentaire.
- Calage aux niveaux ménage, individu et individu-Kish.

## Techniques d'échantillonnage

### Objectif :

Passage obligé d'une opération d'enquête, l'échantillonnage est loin d'être une action anodine. Elle demande d'abord une certaine technicité algorithmique même dans les cas les plus simples. C'est aussi un moment où, on risque de perdre plus ou moins définitivement une information pré-existante. Contraindre un échantillon c'est le structurer par une information et de ce fait augmenter son efficacité.

On étudiera les contraintes suivantes :

- taille fixe et probabilités égales (sondage aléatoire simple)
- probabilités variables
- probabilités variables et taille fixe
- stratification
- contraintes multiples (échantillonnage équilibré)

### Pré-Requis :

Révision du cours de Théories des Sondages de 2<sup>ème</sup> année.

### Contenu :

#### Introduction, utilisation de l'échantillonnage.

Généralités  
Conception d'un plan de sondage  
Estimation

#### Tirages à probabilités égales.

Tirage bernoullien  
Sondage aléatoire simple  
Tirage systématique

#### Tirages à probabilités inégales.

Principe  
Tirage stratifié  
Tirage poissonien, tirage réjectif  
Estimation de variance

#### Echantillonnage équilibré.

Principe  
La méthode du cube  
Exemples d'utilisation : le Nouveau Recensement, l'Echantillon-Maître  
Application à l'imputation d'une variable

#### Travaux pratiques :

Conception d'une stratégie d'échantillonnage : échantillonnage à probabilités égales ou inégales, échantillonnage équilibré.

## Estimation sur petits domaines – Echantillonnage indirect

### Objectif :

**Estimation sur petits domaines :** Lorsqu'on souhaite estimer des paramètres à partir de données d'enquête par sondage (taux de chômage, proportion de « pauvres », etc...), on se trouve souvent confronté au problème de la médiocre qualité des estimations relatives aux sous populations (appelées « domaines ») de petite taille. C'est une conséquence mécanique de la faible taille de l'échantillon qui recoupe ces domaines, qui sont en général de « petites » aires, comme par exemple des départements ou des agglomérations. Il est alors nécessaire d'utiliser de l'information auxiliaire venant de sources exhaustives ou issues d'enquêtes par sondage de très grande taille (par exemple les enquêtes de recensement) et de modéliser les indicateurs en fonction des variables ayant un bon pouvoir explicatif. Sur ce point, la théorie offre d'assez nombreuses pistes, mais l'expérience française en la matière est encore fort limitée.

Dans ce contexte, le module présentera les principales méthodes d'estimation sur petits domaines en s'appuyant à la fois sur la théorie des sondages et sur la modélisation classique (modèles linéaires mixtes, modèles linéaires généralisés, et modèles linéaires mixtes généralisés). Des exemples d'application seront commentés. Des exercices d'application des techniques enseignées sont également proposés sous forme de travaux pratiques.

### Pré-requis :

Une bonne connaissance de la théorie des sondages est recommandée.

### Contenu :

#### 1 Estimation sur petits domaines

##### 1.1 Introduction et rappels de théorie des sondages

##### 1.2 Estimation directe

1.2.1. L'estimateur de Horvitz-Thompson : estimation du total ; estimation de la moyenne

1.2.2. L'estimateur par la régression : formulation générale ; formulations spécifiques

1.2.3. L'estimateur par calage généralisé

##### 1.3 Estimation indirecte avec modélisation implicite

1.3.1. Principe de base de cette approche

1.3.2. L'estimation synthétique

1.3.3. L'estimation composite

##### 1.4 Estimation indirecte avec modélisation explicite

1.4.1. Principe de base de cette approche

1.4.2. La classe des estimateurs sans biais optimums et linéaires (SBOL et ESBOL)

1.4.3. La classe des prédicteurs de comptage (modèles de Poisson, logistique,...)

1.4.4. La classe des prédicteurs optimums (dits « Bayésiens empiriques »)

1.4.5. Applications, aspects informatiques et mesures de qualité

##### 1.5 Conclusion

#### 2 Travaux pratiques

##### 2.1 Estimation directe

2.1.1. L'estimateur de Horvitz-Thompson

2.1.2. L'estimateur par la régression

2.1.3. L'estimateur par calage généralisé

##### 2.2 Estimation indirecte avec modélisation implicite:

2.2.1. Les estimateurs synthétiques (sans information auxiliaire et avec information auxiliaire)

2.2.2. Les estimateurs composites

# ANALYSE STATISTIQUE

Les enseignements du domaine «*Analyse statistique*» sont axés sur les méthodes d'analyse des données statistiques les plus utiles à l'étude des phénomènes économiques ou sociaux. Ainsi, les participants pourront approfondir leur connaissance des modèles applicables aux données qualitatives, y compris textuelles, aux données temporelles ou spatiales, agrégées ou individuelles.

↳ Analyse exploratoire des données	28
↳ Econométrie des durées	29
↳ Analyse spatiale	30
↳ Analyse des données approfondie	31
↳ Analyse et traitement des séries temporelles	32
↳ Compléments de séries temporelles	33
↳ Compléments de modèle linéaire	34
↳ Régression sur données catégorielles et données de comptage	35
↳ Théorie des indices de prix	36
↳ Approche non standard en régression	37
↳ Statistique bayésienne	38

## Analyse exploratoire des données

### Objectif :

Les objectifs de cette formation sont de :

- Présenter les principaux outils de la statistique exploratoire, que ce soit dans le domaine de la statistique descriptive, de l'analyse des données, de la régression, des séries temporelles.
- Profiter des connaissances des étudiants en statistique classique pour revenir sur les outils qu'ils ont étudiés afin, en les comparant à ceux de la statistique exploratoire, d'en mieux connaître les avantages et les inconvénients.
- Proposer une démarche pratique du statisticien face à des données.
- Sensibiliser à l'apport d'une démarche graphique et visuelle en statistique.

### Contenu :

#### Statistique unidimensionnelle

- Quantiles, box-plots, stem-&-leaf et histogramme
- La transformation des variables : la famille de Tukey, lien avec les transformations de Box-Cox
- La recherche de la loi d'une variable, Qqplot et autres outils de diagnostic

#### Statistique bidimensionnelle

- Outils graphiques
- Régression simple : la « résistant line », l'analyse des résidus
- Autres méthodes de régression : régression « bi-weight », LOWESS
- Analyse d'un tableau de contingence : le « median polish »

#### Etude des séries chronologiques

- Outils graphiques, lissage d'une série par médianes mobiles

#### Statistique multidimensionnelle : régression et analyse des données

- Visualisation multidimensionnelle
- Analyse par graphiques interactifs : scatterplot matrix, brush, spin
- Projections : Directions révélatrices (« projection pursuit ») etc..
- Exploration et analyse des données

#### Démarches et concepts

- Philosophie de l'exploration et de la visualisation statistique
- Perception et design graphique
- Hypothèses implicites et robustesse

## Econométrie des durées

### Objectif :

Les variables de durée sont fréquemment soumises à des phénomènes de censure, de troncature ou de biais de sélection. La formation fournit les outils méthodologiques pour les analyser. Trois grandes familles de modèles seront présentées, se rapportant à des analyses paramétriques, semi-paramétriques ou non paramétriques, ainsi que des travaux portant sur des thèmes variés (durées de chômage par exemple).

### Pré-requis :

Une bonne connaissance des modèles de régression linéaire est nécessaire.

### Contenu :

#### Modèles de durée

- Exemples introductifs
- Durée de vie et fonctions associées
- Censure et troncature
- Lois usuelles
- Quelques notions de vieillissement ou rajeunissement

#### Estimation paramétrique

- Maximum de vraisemblance en présence de données censurées

#### Estimation non paramétrique

- Cas des données complètes
- Cas des données censurées à droite : estimateurs de Nelson-Aalen et de Kaplan-Meier
- Comparaison d'échantillons
- Tests graphiques d'adéquation à une loi paramétrique

#### Modèles de régression

- Modèles de régression paramétriques
- Modèles à risques proportionnels
- Modèle semi-paramétrique de Cox, vraisemblance partielle

#### L'économétrie des durées avec SAS

- Estimation non paramétrique et paramétrique : PROC LIFETEST et LIFEREG.

#### L'analyse des durées de chômage : une synthèse des résultats (USA, GB, France)

- L'effet de l'indemnisation du chômage sur les taux de sortie du chômage : comment tenir compte d'un régresseur variable au cours du temps
- L'application des modèles à risques concurrents : le taux de sortie vers les emplois précaires varie-t-il avec le temps passé en chômage ?
- L'impact des stages sur les durées de chômage : les leçons d'un modèle de transition

## Analyse spatiale

### Objectif :

On abordera dans ce module les deux composantes de l'analyse spatiale : les méthodes de représentation cartographique et de mesure de la configuration spatiale (qualitatives ou quantitatives) et l'interprétation géographique de données économiques ou sociales. Le module définira les concepts-clés de l'analyse spatiale. Il présentera un panorama des modèles statistiques ou mathématiques utilisés en géographie. On traitera également de l'application de l'analyse des données à la géographie, des mesures de concentration, d'homogénéité spatiale et de polarisation et enfin des différentes méthodes de zonage. Les participants mettront en pratique les principes de base de la sémiologie graphique en apprenant à utiliser des logiciels de cartographie.

### Contenu :

#### **Le rôle de la statistique publique dans l'analyse spatiale et le diagnostic de territoire**

Qui sont les acteurs locaux destinataires de ces analyses, quels sont leurs moyens d'intervention, leurs problématiques ? - Les principaux indicateurs de la connaissance statistique actuelle des territoires - Panorama des outils d'analyse (développés dans les sessions suivantes).

#### **Outils et méthodes de la géostatistique**

Notions de variable régionalisée, de support, de sélectivité - Analyse de la variabilité spatiale : nuée variographique, variogramme - Choix d'un modèle - Prédiction par krigeage - Aperçu d'autres méthodes géostatistiques : méthodes multivariées et analyse krigeante, modèles non stationnaires, simulations conditionnelles.

#### **Outils pour l'analyse territoriale en infra communal**

- 1) Infrastructures géographiques et données pour l'infra-communal. Zonages, référentiels géographiques. Géolocalisation de sources administratives. Questions de diffusion.
- 2) Décrire la diversité spatiale. L'approche classique par les zonages : outils et limites. Approche non zonée : estimation de densité et indices locaux.
- 3) Modéliser avec des données spatialisées. Régressions avec auto-corrélation spatiale, régressions géographiques.

#### **Modèles et méthodes d'analyse des données bilocalisées**

#### **Le langage cartographique :**

La cartographie à l'INSEE, principe de sémiologie - Présentation et mise en œuvre des outils cartographiques élaborés à l'INSEE, essentiellement les outils d'analyse spatiale ; lissages, résultantes vectorielles...

## Analyse des données approfondie

### Objectif :

Ce module portera sur les développements récents de l'analyse des données.

Les méthodes de segmentation constituent une approche non-paramétrique de l'analyse discriminante. Ces méthodes permettent de prendre en compte un grand nombre de variables, qualitatives et quantitatives (arbre de régression) et d'en sélectionner les plus pertinentes. Leur principe est de segmenter un ensemble d'individus en sous-populations, de manière séquentielle. La méthode CART sera présentée. Elle consiste à construire un arbre binaire utilisant le critère d'homogénéité des sous-populations obtenues à chaque nœud (étape), puis à élaguer cet arbre afin de déterminer une règle de décision finale.

L'AFM est dédiée aux tableaux individus x variables dans lesquels les variables sont structurées en groupes (exemple : mêmes questions prises à différentes dates). Les variables peuvent être quantitatives et/ou qualitatives. A partir d'un tel tableau, l'AFM produit les résultats usuels de l'analyse factorielle (représentation des individus, des variables quantitatives (comme en ACP) et des modalités des variables qualitatives (comme en ACM) et des résultats spécifiques de la structure des variables en groupes ; représentation des individus du point de vue de chaque groupe, représentation des groupes etc.

L'AFM Duale est dédiée aux tableaux individus x variables dans lesquels les individus sont structurés en groupes. Cette analyse fournit une représentation superposée des G nuages de variables associés aux G groupes d'individus ainsi qu'une représentation des matrices de corrélations associées chacune à un groupe d'individus. On peut par exemple analyser l'évolution à différentes dates des corrélations entre variables.

La complémentarité entre AFM et classification d'individus sera abordée notamment avec la classification de trajectoires temporelles.

Si les méthodes de classification les plus connues s'appliquent à la réalisation de typologies d'individus, la classification de variables s'avère également utile. En effet, l'émergence de bases de données toujours plus volumineuses confère à la réduction des dimensions une place cruciale. Or les méthodes factorielles (ACP, AFC et ACM) ne répondent pas à tous les besoins. Par la création de groupes de variables corrélées, la classification de variables offre la possibilité de construire des représentants de classes (variables synthétiques) - ou encore de choisir parmi les variables initiales celles qui sont les plus représentatives de leur classe (parangons). Ce nouvel ensemble plus restreint de variables sera plus facilement gérable et interprétable.

### Pré-requis :

La révision du cours d'Analyse des Données de 2<sup>ème</sup> année constitue un préalable indispensable.

### Contenu :

#### Segmentation

#### Méthodologie des méthodes classiques d'analyse factorielle

- Analyse en composantes principales
- Analyse factorielle des correspondances
- Analyse des correspondances multiples

#### Analyse factorielle multiple : AFM et AFM duale

#### Complémentarité entre AFM et classification d'individus

#### Classification de variables

## Analyse et traitement des séries temporelles

### Objectif :

Il vise à rappeler les concepts liés au traitement des séries chronologiques. Dans un premier temps, les séries univariées et stationnaires seront étudiées notamment à l'aide de la méthode de Box et Jenkins. L'étude de la non-stationnarité permettra, dans un deuxième temps, la généralisation aux séries multidimensionnelles et non-stationnaires au travers des modèles VAR et VECM. Enfin, les deux dernières demi-journées seront consacrées à la désaisonnalisation des séries.

L'objectif est qu'à l'issue du module le stagiaire sache modéliser de manière pratique les séries temporelles qu'il a à sa disposition. Une large place sera donc laissée à la mise en pratique de ces techniques sur données réelles.

L'aspect non stationnaire, très riche en développements, sera étudié plus en détails dans le module « Compléments de séries temporelles ».

### Pré-requis :

La révision du cours de Séries Temporelles de 2<sup>ème</sup> année constitue un préalable indispensable pour suivre cette formation.

### Contenu :

#### **Le traitement de la saisonnalité et les effets de jours ouvrables**

- Introduction
- La désaisonnalisation avec X11-ARIMA et avec TRAMO-SEATS
- Introduction à la correction des effets de jours ouvrables
- Travaux pratiques

#### **Rappels sur la modélisation des séries temporelles univariées**

- Introduction
- Quelques outils
- Les modèles AR, MA et ARMA : définition, estimation, tests et prévision
- Travaux pratiques

#### **Etalonnages et introduction à la non-stationnarité**

- Introduction
- Définition, estimation et tests
- Les modèles ARIMA
- Travaux pratiques

#### **Introduction aux séries multidimensionnelles : les modèles VAR et VECM**

## Compléments de séries temporelles

### Objectif :

L'objectif est de proposer aux participants de compléter leurs connaissances dans le domaine des séries temporelles en étudiant le cas multivarié, et le cas non stationnaire.

Après avoir examiné le caractère non standard des séries non stationnaires, et observé le caractère inadapté des tests classiques dans les régressions par moindres carrés ordinaires, on s'intéresse aux tests permettant de détecter de telles séries.

Dans le cadre multivarié, on définira la notion de cointégration, et les représentations associées. Des estimations et tests complèteront le tout.

### Pré-requis :

La révision du cours de Séries Temporelles de 2ème année constitue un préalable indispensable. En outre, il est conseillé d'avoir suivi le module de formation « Analyse et traitement des séries temporelles » avant « Compléments de séries temporelles ».

### Contenu :

#### Cadre univarié :

- Séries stationnaires et la méthodologie de Box et Jenkins ;
- Séries non stationnaires ;
- Séries avec racines unité : test de Dickey-Fuller et de Phillips-Perron.

#### Cadre multivarié :

- Les modèles VAR, et ses propriétés ;
- Quelques résultats asymptotiques ;
- Modèles VAR avec racine unité ;
- Cointégration (estimation et tests variés)

**Cas pratiques :** quelques exemples seront étudiés pour illustrer les points précédents.

## Compléments de modèle linéaire

### Objectif :

Cette formation a pour objectif de répondre à certains problèmes liés à la modélisation de variables continues. Par exemple, le calcul de l'indice des prix nécessite une mesure explicite de la qualité d'un produit afin de tenir compte des progrès technologiques. Cette mesure de qualité est essentiellement modélisée à partir de caractéristiques techniques qui peuvent être des variables qualitatives ou quantitatives. De même, dans les études sur le revenu, un grand nombre de variables à la fois qualitatives et quantitatives entre en jeu. Un des problèmes souvent rencontré est la forte corrélation entre les différentes variables explicatives, entraînant de sérieux problèmes de stabilité à la fois dans la calibration du modèle et dans les prévisions qui en découlent. Un autre problème souvent rencontré est la nature aléatoire de certains facteurs. Par exemple, le plan d'échantillonnage utilisé pour étudier les entreprises françaises va rendre aléatoire les variables endogènes associées. La non-prise en compte du caractère aléatoire de ces variables peut modifier de manière sensible l'interprétation des résultats obtenus.

Dans cette formation, nous aborderons certaines techniques permettant de résoudre en partie les problèmes évoqués ci-dessus. Dans un premier temps, nous abordons l'analyse de la variance, partiellement vue en 2<sup>ème</sup> année d'ENSAI ; l'approche est entièrement revue dans une optique qui permet sa généralisation à des cas plus complexes. Le deuxième point concerne les modèles à effets aléatoires ou mixtes. La troisième partie a pour but de présenter les modèles d'analyse de la covariance qui mélange à la fois les facteurs explicatifs quantitatifs et qualitatifs. Enfin, la quatrième partie aborde le problème délicat de régression sur données corrélées.

Toutes les techniques présentées feront l'objet d'une mise en œuvre sous SAS lors des travaux pratiques.

### Pré-requis :

Il est recommandé de maîtriser le cours de modèle linéaire de 2<sup>ème</sup> année à l'ENSAI.

### Contenu :

#### Analyse de la variance

- Identifiabilité, fonctions estimables, théorème de Gauss-Markov général
- Facteurs croisés, hiérarchisés et d'interactions, étude des facteurs par le diagramme de Hasse
- Tests d'un sous-modèle, degrés de liberté, tests de type I, II et III, comparaisons multiples
- Présentation de la PROC GLM de SAS pour l'analyse de la variance (lsmeans, estimate, contrasts,...)

#### Introduction aux modèles à effets aléatoires et modèles mixtes

- Effets fixes, effets aléatoires, modèles à structures de covariance linéaire
- Estimation des effets fixes et des composantes de la variance
- Présentation de la PROC MIXED de SAS

#### Analyse de la covariance

- Facteurs qualitatifs, facteurs quantitatifs, écriture du modèle d'analyse de la covariance
- Test d'égalité des pentes, test d'égalité des constantes
- Mise en œuvre pratique (TP) avec la PROC GLM de SAS

#### Régressions sur données corrélées, prévision

- Instabilité de la prévision en présence de forte colinéarité des régresseurs
- Revues des techniques de sélection de variables
- Comparaisons des principales méthodes de régression biaisée
- Calibration de modèles par validation croisée
- Utilisation des PROC REG et PLS de SAS

#### Travaux Pratiques

## Régression sur données catégorielles et données de comptage

### Objectif :

La formation se divise en quatre parties :

Dans une première partie, nous considérons des réponses catégorielles, c'est-à-dire des variables représentées par un nombre fini de modalités. La réponse est dichotomique lorsqu'elle n'a que deux modalités. C'est le cas lorsque l'étude porte sur l'occurrence d'un événement tel que le chômage, l'apparition d'une maladie, l'acceptation d'un prêt, le décès d'un patient... Plus généralement, on dit que la réponse est polytomique lorsqu'elle admet plusieurs modalités. Par exemple lorsque l'on étudie des catégories socioprofessionnelles ou bien les différents stades d'une maladie. Dans ce cas la variable peut être ordonnée ou non. L'étude statistique de ces variables nécessite alors un autre outil que le modèle linéaire. Nous présenterons les méthodes de régression sur variables catégorielles, en particulier la régression logistique. Nous étudierons les modèles dichotomiques, polytomique ordonné et multinomial.

Dans une deuxième partie, la réponse est une variable de comptage. Il s'agit donc d'une variable pouvant prendre des valeurs entières 0,1,2... Ces variables jouent un rôle important dans de nombreux domaines (science sociale, génétique, épidémiologie, assurance). Ce sont par exemple le nombre d'accidents d'un individu, le nombre de pannes d'une machine, le nombre d'enfants d'un foyer... Encore une fois, l'étude de telles variables sort du cadre du modèle linéaire classique. Le cours traitera donc des méthodes de régression sur variables de comptage. Nous étudierons plus particulièrement le modèle de Poisson et le modèle binomial négatif.

Dans une troisième partie, nous verrons que les régressions sur variables catégorielles et sur variables de comptage sont des cas particuliers d'un modèle plus vaste : le modèle linéaire généralisé. Nous définirons ce modèle et nous en présenterons les différentes propriétés.

Dans une dernière partie, nous considérerons des données longitudinales, c'est-à-dire des variables observées plusieurs fois sur des mêmes individus. Il peut s'agir d'études réalisées à différentes périodes sur un même échantillon. Nous étendrons alors les études précédentes du modèle de Poisson et du modèle logistique dans ce cadre des données répétées.

Les différentes méthodes introduites durant ce cours seront accompagnées d'exemples. Nous présenterons également les procédures SAS correspondant aux modèles étudiés.

### Pré-requis :

Révision du cours de Modèles de Régression de 2<sup>ème</sup> année.

### Contenu :

- **Données catégorielles** : exemple d'introduction ; la régression logistique simple ; la régression logistique multinomiale ; la régression logistique ordonnée ; la régression logistique pour les enquêtes cas-témoins ; les tables de contingence
- **Données de comptage** : contexte ; la régression de Poisson ; la surdispersion ; la régression binomiale négative
- **Une généralisation : le modèle linéaire généralisé** : définitions ; les propriétés du modèle linéaire généralisé ; illustrations
- **Modélisations longitudinales** : la méthode GEE ; données catégorielles longitudinales ; données de comptage longitudinales

## **Théorie des indices de prix**

### **Objectif :**

L'objectif de ce module est de relier la pratique et la théorie des indices.

Il s'orientera principalement vers les indices temporels de prix, mesure de l'inflation, sans exclure des exemples provenant d'autres domaines.

Ce module bénéficiera de l'expérience de l'harmonisation européenne des indices de prix à la consommation acquise à partir de la fin des années 90 et des débats autour des effets de substitution et de l'effet qualité du milieu des années 2000.

Une journée de cette formation sera consacrée à la présentation des deux indices principaux de l'INSEE, à savoir l'indice des prix à la consommation et l'indice des prix à la production.

### **Pré-requis :**

Connaissance de la théorie microéconomique du consommateur et des modèles de régression.

### **Contenu :**

#### **Une apparence simple, mais beaucoup de questions en suspens**

- Un panier fixe dont on suit l'évolution de la valeur selon l'évolution des prix des ses composantes. L'indice de Laspeyres. Une vision orientée par les possibilités de collecte.
- Mais ce qui paraît simple ne l'est pas : l'effet de structure illustré par le GVT des salaires ; un problème d'agrégation, quel indice choisir : Laspeyres, Paasche, Jevons....
- Une réponse, le chaînage : le biais résolu par le chaînage : mise à jour des pondérations ; le biais introduit par le chaînage : le Bouncing . L'intérêt de la moyenne géométrique. D'autres pondérations pour éviter le biais de chaînage.
- Chaînage et soldes .

#### **Refondation théorique**

- L'approche mathématique : la décomposition de la valeur en volume x prix, l'intégrale sur le contour de la courbe de consommation et l'indice de Divisia.
- L'approche économique : l'indice à utilité constante.
- L'unification des approches mathématiques et économiques
- Les hypothèses qui sous-tendent les indices historiques à la lumière de l'approche théorique.
- Introduction du plan de sondage, conséquences sur la pondération.
- Liste des indices connus Laspeyres, Paasche, Fisher, moyenne géométrique, le Törnqvist, le Balk-Walch...
- Introduction éventuelle selon le temps disponible à l'approche axiomatique.
- Les possibilités futures offertes par les nouveaux modes de collecte.

#### **Principales difficultés d'un indice de prix à la consommation**

- L'effet qualité : un mal théorique nécessaire ; quelles méthodes : remplacement en équivalent, dissemblable, recouvrement - méthodes hédoniques (indicatrice temporelle ou prix de base) ; les enjeux : le rapport Boskin et le contre-rapport de l'Académie des sciences ; une mesure générale: les IQI proposés par Eurostat ; les produits saisonniers ; large champ : principalement les produits frais et l'habillement ; fortes variations des indices ; un problème théorique : «la saison moyenne n'existe pas» ; estimation d'un Paasche par la méthode de Rothwell, ou imputation des prix hors saison.

#### **Cadre légal et l'expérience internationale**

En France : instruments contractuels mais laissant place à la négociation - le cas particulier du tabac ; règlements européens, les principaux textes de l'IPCH - normes FMI ; d'autres organisations et d'autres indices ; un peu de vocabulaire anglophone ; des sites utiles.

#### **Les indices français de prix à la consommation et à la production**

- Indice des Prix à la Consommation : concepts, principes méthodologiques, couverture, pondérations, organisation de la collecte, échantillon, méthode de calcul, remplacement de produits, analyse (IPC, CVS, ISJ, IPC des ménages modestes, IPCH), diffusion.
- Indice des Prix à la Production : des concepts liés à la nature de l'indicateur (prix, transactions, produits) ; les différents usages de l'indice et l'introduction subséquente de variantes (comptabilité nationale et prix de base vs. Indexation de contrats et prix de marché) ; l'organisation particulière de l'enquête en deux phases (témoins et pondérations puis suivi des prix) ; le traitement de l'effet qualité en pratique dans OPISE ; les indices de prix de production à l'étranger.

## Approche non standard en régression

### Objectif :

L'analyse de la liaison entre variables observées, qu'elles soient continues ou catégorielles, est une des bases de l'analyse statistique. Elle présuppose la définition par le statisticien d'un modèle paramétrique qui fixe les liaisons existant entre ces variables, et requiert ensuite un travail classique d'ajustement aux données, via les techniques usuelles d'estimation et de tests. Le célèbre « modèle linéaire gaussien » est l'exemple typique de cette démarche.

Depuis une vingtaine d'années ont été développées des techniques non paramétriques qui n'imposent aucune condition sous-jacente quant à la forme de la loi des observations. L'objectif de la formation est ainsi de faire un tour d'horizon de ces techniques et de leurs apports ou limites quant aux techniques paramétriques appartenant à la boîte à outils du statisticien. Une attention particulière sera donnée lors de la formation à la mise en œuvre pratique de ces techniques. Le développement de bibliothèques de fonctions, ou « packages », dans le cadre du logiciel libre R permet, en effet, d'envisager l'implémentation des techniques d'estimation non paramétrique et leur comparaison aux techniques paramétriques usuelles.

### Pré-requis :

Une relecture du cours de Statistique non paramétrique (ENSAI 2) est conseillée. Il existe divers ouvrages présentant les techniques d'estimation paramétriques. Parmi ceux-ci, notons les ouvrages récents de Li et Racine (2007), *Nonparametric Econometrics*, Princeton University Press, et Ahamada et Flachaire (2008), *Econométrie Non Paramétrique*, Economica.

### Contenu :

La formation sera organisée chaque jour en deux temps. Les matinées seront consacrées à un cours introduisant les techniques non et semi paramétriques et leurs propriétés. La mise en œuvre de ces techniques sera illustrée, lors des après-midi, à partir d'exemples traités sur ordinateur en utilisant le logiciel libre R où différents « packages » ont été écrits permettant l'estimation de modèles non ou semi paramétriques.

D'où le plan de cours :

- 1ère après-midi : introduction et rappels sur l'estimation non paramétrique d'une densité de probabilité (histogramme et estimateur de Parzen et Rosenblatt).
- 2ème jour : (1) le matin, estimation non paramétrique d'un modèle de régression (estimateur de Nadaraya et Watson et polynômes locaux) et (2) l'après-midi, initiation au logiciel R et estimation de la densité à partir de données simulées et de données de salaire.
- 3ème jour : (1) le matin, estimation de la densité conditionnelle et de la fonction de répartition et (2) l'après-midi, estimation non paramétrique de modèles de régression à partir de données simulées et application aux données de salaire lors de l'estimation de fonction de gain.
- 4ème jour : (1) le matin, tour d'horizon sur les approches semi paramétriques (modèles partiellement linéaires, modèles additifs et modèles à indices simples) et (2) l'après-midi, applications dans le cadre de l'estimation d'une fonction de prix hédoniste.
- 5ème jour : le matin, projet par les participants à la formation mettant en œuvre des techniques vues pendant la semaine.

Lors de chacune de ces études réalisées, seront mises en lumière les convergences et divergences de résultats obtenues selon les approches paramétriques classiques (régression linéaire ou modélisation logit) ou celles qui auront été introduites lors de la formation, pour modéliser les données.

## Statistique bayésienne

### Objectif :

L'approche bayésienne de la statistique connaît à l'heure actuelle un essor considérable notamment grâce aux progrès de l'informatique et des méthodes numériques de type MCMC. Lorsque l'on réalise une étude, on a souvent des informations a priori provenant soit d'études antérieures soit d'avis d'expert. La statistique bayésienne permet d'utiliser ces connaissances a priori et de les combiner avec l'information apportée par les données pour obtenir une information a posteriori. La statistique bayésienne est également très utilisée dans les meta-analyses, c'est à dire les analyses qui mettent ensemble plusieurs études réalisées dans des conditions parfois différentes pour en extraire de l'information avec une meilleure précision.

Au cours de la formation nous nous efforcerons de comparer les avantages et les inconvénients de l'approche bayésienne par rapport à l'approche classique (ou fréquentiste).

Toutes les notions abordées seront illustrée à l'aides des logiciels WINBUGS ou SAS.

### Contenu :

#### **Le paradigme bayésien (1 journée) :**

- Exemple introductif ;
- La formule de Bayes ;
- Lois a priori, lois a posteriori ;
- Choix des lois a priori, lois informatives, lois non informatives, lois conjuguées.

#### **Lois a posteriori (cours 1/2 journée, TP 1/2 journée)**

- Nécessité de recourir aux méthodes « computationnelles » pour calculer la loi a posteriori ;
- Initiation aux méthodes MCMC (chaînes de Markov par Monte-Carlo) ;
- Mise en œuvre avec le logiciel Winbugs.

#### **Méthodes d'estimation bayésiennes (cours 1/2 journée, TP 1/2 journée)**

- Rappels de théorie de la décision ; notions de prédicteurs ;
- Comparaison des estimateurs bayésiens et fréquentistes ;
- Intervalles de crédibilité ;
- Mise en œuvre avec Winbugs.

#### **Modèles de régressions bayésiens (TP, 1/2 journée)**

- Analyse bayésienne des modèles de régressions les plus courants (régression linéaire, logistique, poisson) ;
- Applications à l'aide du logiciel Winbugs et SAS (procédures expérimentales).

#### **Tests bayésiens (TP, 1/2 journée)**

- Notions de tests bayésiens, facteur de Bayes ;
- P-value versus Q-value.

# INFORMATIQUE

Les modules du domaine « *Informatique* » contribuent à un objectif commun : la maîtrise, par les participants, du traitement automatique de l'information dans ses différentes phases : extraction, analyse, stockage, restitution. Ce qui nécessite, de leur part, une bonne compréhension des méthodes de conception de systèmes d'informations statistiques.

↳ Compléments de bases de données	40
↳ Compléments de SAS	41
↳ Datawarehouse - datamining	42
↳ Réseaux et sécurité	43
↳ XML	44

### Objectif :

Ce module s'inscrit dans le prolongement du cours de *Bases de données* de deuxième année. L'objectif de cet enseignement est d'élargir les connaissances en interrogation des données et de présenter les évolutions des modèles de données utilisés.

Les limites de l'interrogation des Bases de Données relationnelles avec SQL sont rapidement atteintes, pour étendre les possibilités d'interrogation les requêtes SQL sont intégrées au sein d'un programme écrit en C, C++ ou JAVA.

Ceci nécessite d'utiliser des outils particuliers, ils seront présentés dans ce module en utilisant le langage de programmation d'ORACLE PL/SQL.

D'autre part, dans ce module seront présentés les architectures 3-tiers, le client-serveur et les SGBD distribués. Les constructeurs de SGBD relationnels étendent leurs systèmes pour permettre la gestion d'objet (SGBD objet-relationnel). Les dernières extensions d'Oracle dans ce domaine seront également étudiées.

### Contenu :

#### 1. Rappels

- 1.1. **Le modèle relationnel** : normalisation ; algèbre relationnelle.
- 1.2. **Le langage SQL** : création de tables ; interrogation ; mises à jour.

#### 2. Bases de données sécurisées

##### 2.1. Introduction à la protection des SGBD

Mécanismes de protection d'un SGBD ; types de politiques de contrôle d'accès

##### 2.2. Protection des SGBD : mise en oeuvre d'une politique de sécurité

Introduction ; architectures de sécurité ; granularité de l'information ; choix des classifications ; polyinstantiation ; décomposition des relations ; composants de sécurité Oracle

#### 3. Tuning (Optimisation des bases de données)

L'objectif de ce cours sur le "Tuning" des bases de données est de permettre de maîtriser les notions fondamentales gouvernant l'obtention de performances. Pour cela, nous détaillerons quelques briques internes des SGBD relationnels, à savoir les briques assurant la gestion du contrôle de concurrence et de la reprise sur pannes, l'indexation et l'optimisation de requêtes. Ce cours comportera aussi des travaux dirigés où nous résolverons des cas d'écoles, exemplaires de ce que vous pouvez rencontrer dans le monde réel. Il est important de noter que ce cours n'est pas une mise en pratique sur tel ou tel SGBD du commerce ; nous nous focalisons sur les principes et non sur les manières dont le Tuning se fait ici ou là.

#### 4. Tatouage de bases de données

Aujourd'hui la copie des données est très facile, ceci en facilite le vol. Le tatouage des données permet de les protéger, son principe consiste à cacher des informations "invivables" dans les données de façon à les identifier lorsqu'elles ont été volées.

Plan de cette partie :

- 4.1. **Problématique**
- 4.2. **Modélisation**
- 4.3. **Caractéristiques du tatouage**
- 4.4. **Attaques**
- 4.5. **Interprétation géométrique**

## Compléments de SAS

### Pré-requis :

Un bon niveau en SAS de base (étape DATA) et en méthode de programmation est indispensable.

### Contenu :

#### I) - Macro-langage

Le macro-langage est un langage de programmation qui permet d'écrire des blocs d'instructions mis en mémoire pour être rappelé à tout instant. Quelques plus du macro-langage : amélioration des possibilités du langage de base, rapidité et simplicité des applications, utilisation de programmes paramétrés, réalisation de programmes SAS à insérer dans des traitements répétitifs.

##### A) Introduction

Eléments du macro-langage ; fonctionnement du processeur macro

##### B) Macro-variables

Création ; appel et résolution

##### C) Macro-fonctions

Catégories de macro-fonctions

##### D) Macro

Définition ; paramétrage ; environnement de référence

##### E) Macro-instructions

Macro-instructions générales ; macro-instructions dans une macro

##### F) Interfaces

Lien avec l'étape DATA

##### G) Compléments

Macro-quoting ; débogage ; gestion des macros

#### II) - SAS IML

SAS/IML (Interactive Matrix Language) est un langage de programmation matriciel interactif. Ce langage est très flexible et ses capacités de traitement sont importantes. L'intérêt d'utiliser ce module de SAS est la programmation directe en éliminant la plupart des spécifications itératives (boucles), appels à des sous-routines mathématiques, déclarations de variables, etc.

#### Introduction : principe, fonctionnement

##### Le langage SAS/IML

- Définition d'une matrice
- Création de matrice
- Principaux opérateurs
- Différentes instructions : instructions de contrôle (par exemple IF-THEN-ELSE), fonctions et commandes
- Gestion des valeurs manquantes

##### Travailler avec des matrices

- Saisie et génération de matrices
- Opérateurs binaires, sous-matrices, réduction de matrices
- Tutoriel : régression linéaire

##### Graphiques sous SAS/IML

##### Applications statistiques

- Corrélation
- Résolution de systèmes non-linéaires par la méthode de Newton
- Programmation linéaire et quadratique

## Datawarehouse - Datamining

### Objectif :

**Datawarehouse :** Les différentes composantes d'un entrepôt de données sont présentées en utilisant un cas concret de mise en place d'un entrepôt. Les spécificités de chacune des composantes sont illustrées avec certains produits du marché. Un lien est fait entre les connaissances théoriques acquises à l'Ensaï et leurs utilisations dans chacune des composantes : alimentation, stockage et restitution.

Une place importante est accordée à la présentation des bases de données multidimensionnelles. Une première partie est consacrée à la modélisation ensuite les différentes opérations sont présentées. Cette présentation se termine par une étude de cas qui permet de mettre en pratique ce type de système, l'outil utilisé est celui d'Oracle.

**Datamining :** Le DataMining se présente comme une discipline aux confluent de la statistique "classique" (modélisation, imputation, classification, ...) et des méthodes "informatiques" de traitement des données (arbres de décision, réseaux de neurones, raisonnements basés sur la mémoire, cartes auto-organisées, ...). Elle emprunte à ces deux sphères pour construire des modèles explicatifs, mais surtout prédictifs (des scores).

Les techniques nouvelles du Datamining : un bref aperçu théorique, des exemples de résultats.

Mise en oeuvre du Datamining : les outils existants dans SAS/BASE et SAS/STAT ; un outil dédié : SAS ENTERPRISE MINER (SAS/EM)

Après une rapide présentation des qualités et des défauts de SAS/EM, et de son principe de fonctionnement, l'essentiel de l'apprentissage sera articulé autour d'un TD sur les données utilisées dans le cours Datawarehouse et consacré à l'élaboration d'un modèle.

### Contenu :

#### Datawarehouse

##### 1) Exemple de mise en œuvre d'un entrepôt de données

- 1.1 Présentation générale du Datawarehouse : définition d'un entrepôt de données ; concepts de base ; vision globale de l'architecture
- 1.2 Exemple de construction d'un entrepôt de données au Crédit Du Nord : les différentes étapes du projet ; les problèmes rencontrés

##### 2) Bases de Données multidimensionnelles

- 2.1. Concepts de base : contexte d'un datawarehouse ; définition ; architecture technique d'un système décisionnel ; méthode de modélisation
- 2.2. Les architectures OLAP
- 2.3. Les principaux éditeurs
- 2.4. Etude pratique

##### 3) Exemple d'utilisation des bases de données multidimensionnelles à l'Insee :

**3.1. Le projet Pélican :** généralités ; concepts et référentiel de Pélican ; les données : consultation, utilisation ; quelques cubes de Pélican ; architecture informatique de l'application ; bilan sur l'utilisation de Pélican.

##### **3.2. Le projet EDL (Entrepôt de Données Locales)**

##### 4) Mise en pratique des bases de données multidimensionnelles : Utilisation de Palo et Excel

#### Datamining

5) **Présentation générale du Datamining :** le concept de DataMining : historique ; buts ; méthodologie ; intérêt dans une étude statistique "normale"

##### 6) Mise en œuvre du datamining (1 journée)

- 6.1. SAS et le DataMining : les procédures statistiques au service du DataMining ; l'intérêt d'une interface ; un outil spécialisé : SAS Enterprise Miner ; les autres outils de DataMining
- 6.2. Découverte de SAS Enterprise Miner sous forme de TD sur des données d'assurance automobile

## Réseaux et sécurité

### Contenu :

De plus en plus d'applications migrent ou sont conçues dans un environnement ouvert, utilisant les technologies IP (Internet Protocol) et les différents services applicatifs existants dans cet environnement. Cette ouverture induit des risques nouveaux vis-à-vis de la sécurité des données et des applications, en terme de confidentialité, disponibilité et intégrité.

Ce module prévoit une sensibilisation aux risques informatiques (réseaux, systèmes ou organisationnels) et la formation aux différentes techniques permettant de limiter ces risques.

En particulier, ce module prévoit d'acquérir des notions de base sur le fonctionnement des réseaux (en particulier, IP Internet Protocol), des services (DNS, telnet, FTP, SNMP, SMTP, HTTP, HTTPS, IPSEC, etc) et les bases de sécurité nécessaires (firewall, cryptographie, etc) pour aborder la mise en place d'applications dans un domaine ouvert (Intranet, Extranet, Internet).

Ce module prévoit, de plus, une présentation de vulnérabilités « classiques » des systèmes d'exploitation ainsi que les principes d'une protection active d'application en production (veille technologique et détection d'intrusion). Ces principes permettent d'assurer le maintien de la sécurité des applications au quotidien.

### Contenu :

#### 1) Introduction à la sécurité des systèmes d'information.

- 1.1. Pourquoi une politique SSI : a) Les risques encourus ; b) l'environnement juridique dans lequel travaille l'INSEE
- 1.2. Classification des données
- 1.3. Administration de la SSI

#### 2) Réseaux IP et sécurité

- 2.1. Présentation des réseaux IP : a) Rappels sur l'Internet ; b) Rappels sur les protocoles TCP/IP : le modèle OSI et TCP/IP ; caractéristiques IP, TCP, UDP, ICMP ; principaux protocoles applicatifs ; les vulnérabilités des protocoles ; typologie des attaques réseaux ; les attaques de niveau applicatif (DNS, Mail, WWW) ; introduction à IP V6.
- 2.2. La sécurité des réseaux IP : solutions : a) Introduction aux principes de sécurisation des architectures réseaux : cloisonnement ; translation d'adresses ; concepts des Firewalls ; b) Quelle sécurité dans quel équipement ? Hubs ; switchs ; routeurs ; firewall & Proxy applicatif.

#### 3) Cryptographie

- 3.1. Mécanismes de base de cryptographie : a) Chiffrement symétrique : principes et applications standards (RC4, DES, AES) ; b) Chiffrement à clés publiques : principes et applications (RSA, Diffie-Hellmann, ECC) ; c) Signature : principes et exemples de mise en œuvre (MD4, SHA, DSA)
- 1.2. Systèmes cryptographiques : sécurisation de protocoles réseaux : SSL IPSEC ; sécurisation de la messagerie : PGP ; infrastructure de gestion de clés ; protocoles d'authentification : Kerberos, SecurId, etc.

#### 4) Vulnérabilités applicatives :

- 4.1. Vulnérabilités des logiciels
- 4.2. Programmes malveillants (virus, chevaux de Troie, ...)
- 4.3. Sécurisation des systèmes d'exploitation
- 4.4. Sécurisation des applications

#### 5) Détection d'intrusion

- 5.1. Introduction à la détection d'Intrusion
- 5.2. Méthodes de détection d'intrusion
- 5.3. Les systèmes de détection d'intrusion (IDS)
- 5.4. Faiblesses des systèmes de détection d'intrusion

## XML

### Objectif :

Les objectifs de ce cours sont de donner aux étudiants les connaissances techniques nécessaires pour appréhender le langage XML et ses applications : créer et valider des documents XML, modéliser des données métiers avec une DTD, avoir une vision globale de l'environnement XML, manipuler des documents XML avec DOM, utiliser XSLT pour transformer et afficher des documents XML, utiliser XML Schema pour structurer les données, avoir une vision globale de l'utilisation d'XML dans différents domaines. De nombreux exercices illustrent l'apprentissage technique de XML.

### Pré-requis :

Un léger rappel du langage html sera fait en début de formation mais il est préférable de connaître ce langage pour suivre ce cours (CSS serait un plus).

### Contenu :

#### L'environnement XML

- Le concept
- Le langage XML
- La DTD
- SAX & DOM
- XPath
- Le langage de style XSL
- Les autres standards
- XML Schéma

#### Les applications d'XML

- La configuration d'applications
- La publication de données
- L'échange de données

## **ENVIRONNEMENT PROFESSIONNEL**

Les enseignements du domaine «*Environnement professionnel*» ont pour but de favoriser l'acquisition des compétences nécessaires à l'exercice des fonctions de responsabilité dans les domaines suivants : diffusion de données statistiques, encadrement, réalisation d'études, et administration de projets. Un module est plus spécifiquement destiné à aider les stagiaires dans la rédaction de leur mémoire professionnel.

↳ <b>Anglais</b>	<b>46</b>
↳ <b>Analyse des organisations</b>	<b>47</b>
↳ <b>Management et administration de projets</b>	<b>48</b>
↳ <b>Ecrit &amp; oral : techniques de l'exposé</b>	<b>49</b>

## Anglais

### Objectif :

Ce module est conçu pour faire progresser les niveaux d'expression et de compréhension orales des attachés afin de leur permettre une communication aisée.

Il a également pour but d'approfondir les connaissances culturelles du monde professionnel et quotidien anglo-saxon

### Contenu :

#### I) General English

##### Oral expression

1. Role plays
2. Expressing opinions
3. Présentations
4. Public speaking
5. Predicting
6. Negotiating/arguing

##### Oral comprehension

7. Radio programmes
8. Television news and documentaries
9. Film excerpts

##### Written comprehension

10. Film Reviews
11. Newspaper articles
12. Excerpt(s) from literature

#### II) Themes

Cinema : documentaries and films

Meetings and Negotiating/Decision Making

Currents events : TV news and newspapers

#### III) English for Statistics :Analysis of Complex Survey Data

- 1 Overview of complex surveys (sources of error, randomization theory, complex designs, complex estimation problems, nonresponse, calibration)
- 2 Model-based estimation and inference (issues in modeling with survey data, checking for sampling effects, accounting for sampling design, computing)
- 3 Design-based estimation and inference (linearization methods, fitting regressions to survey data, random groups, jackknife, balanced repeated replication, software)
- 4 Comparison of methods

## Analyse des organisations

### Objectif :

Comprendre les règles de fonctionnement et/ou de dysfonctionnement de groupements organisés comme les entreprises ou les administrations ou toute unité socio-organisationnelle de travail, tel est l'objectif de cet enseignement. Le module rappellera les grandes orientations de la sociologie des organisations et retracera les débats théoriques suscités par la rationalisation des grandes organisations : comment identifier un dysfonctionnement et accroître les performances d'une organisation ? Quelles sont les stratégies des différents acteurs ? Comment sont articulées la division du travail et la coordination des activités ? Comment circulent les innovations technologiques ? En quoi consistent les frontières d'une organisation ? Quels sont les rapports entre une administration (ou une entreprise) et son environnement ? Le module fera appel à la comparaison d'organisations ainsi qu'à l'analyse de situations concrètes de travail. Il sera organisé autour de la présentation de cas concrets comme la réforme d'une grande organisation, le perfectionnement d'une technologie, le lancement d'un nouveau produit ou la création d'un établissement, la fusion d'unités, mais aussi autour de questions relatives à la réglementation du travail ou toutes autres réformes conduisant l'organisation à repenser l'articulation et le fonctionnement de ses composantes. Ce module sera également l'occasion de présenter l'organisation en mode projet et la pratique de la démarche « maîtrise d'œuvre locale en organisation du travail » (Maiol) menée à l'Insee dans le cadre des projets.

### Contenu :

#### L'approche empirique des organisations

Une organisation = une « action organisée » qui se caractérise toujours par une division et une coordination du travail, un découpage de la structure et une mise en liaison, un système de pouvoir et de contrôle, une stabilité relative et des éléments évolutifs : les buts et les frontières. L'organisation est en partie déterminée par des facteurs contingents : l'âge et la taille qui influent sur la complexification de l'organisation, la technologie, la nature de l'environnement, le contexte culturel et la stratégie et les relations externes

#### Les différents repères et théories sur l'organisation

1. L'école classique des organisations – L'approche rationaliste : l'organisation scientifique du travail ; la rationalisation de l'administration ; le modèle bureaucratique – 2. L'école des relations humaines - L'approche psychosociologique : les expériences dans les ateliers de Hawthorne ; la théorie des besoins et des motivations – 3. L'école néo-classique des organisations - L'analyse concurrentielle : les matrices de positionnement ; l'excellence – 4. Les théories de la contingence – L'analyse de l'environnement : influence de la technique ; l'instabilité de l'environnement – 5. La sociologie des organisations – L'acteur et le système : l'analyse stratégique – Pouvoir et autorité ; théorie des conventions – Mondes et accords ; théorie de la traduction – Coopération et traduction ; identité au travail et culture d'entreprise – 6. La synthèse de Henry Mintzberg : perspectives rationnelle, contingente et politique ; théorie des configurations

#### Les évolutions des organisations

1. La remise en cause du modèle taylorien et l'émergence de nouvelles formes d'organisation du travail : la critique du taylorisme dans les sociétés industrielles avancées ; une évolution progressive des organisations du travail productif ; une évolution de la conception et du fonctionnement de l'organisation industrielle ou commerciale - 2. L'évolution des structures d'entreprises et d'organisation : les structures classiques ; les structures plus récentes par projet ; les structures matricielles ; les structures en réseaux - 3. Les modèles de gestion et de prise de décision : le modèle classique ; la distinction entre différents types de décisions ; vers un modèle plus sociologique mettant en évidence le jeu d'interactions existant entre les différents acteurs intervenant dans le processus décisionnel - 4. Les mutations du travail : le modèle du poste de travail ; le modèle compétence ; le passage de la logique du poste de travail à la logique de compétence

#### Les pratiques d'analyse et d'action visant à introduire et conduire des processus de changements au sein des organisations

1. Les grands types de méthodes : l'analyse du travail, des circuits et processus, système et stratégique - 2. Les étapes d'une démarche de conduite de projet de changement organisationnel : la formalisation de la problématique organisationnelle (la clarification des objectifs ; la recherche des points d'appui et la décision ; le choix de la communication) ; l'analyse de la situation et le diagnostic organisationnel ; la recherche et le choix de solutions ; la mise en œuvre du projet de changement organisationnel ; le suivi et l'ajustement du projet - 3. L'accompagnement des changements organisationnels : les conditions de réussite

#### La démarche Maiol appliquée dans le cadre des projets à l'INSEE

1. L'organisation en mode projet – 2. La prise en compte de l'organisation du travail à l'Insee – 3. La démarche Maiol : contexte, objectifs, acteurs, étapes, livrables, bénéfices...

## **Management et administration de projets**

### **Objectif :**

Ce module doit permettre aux participants d'améliorer leurs compétences dans la conduite de projets. Le module mettra d'abord en valeur les éléments nécessaires à la conception d'un projet : analyse de la mission, mise en place du dispositif, planification, pilotage, indicateurs d'évaluation et maîtrise des coûts puis il présentera les apports du management dans le domaine de la délégation, de la communication et de la gestion des compétences. Le module se termine par la présentation de la conduite de projet à L'Insee. Les participants auront l'occasion de mettre en pratique certains apports théoriques grâce à des études de cas ou des exercices filmés..

### **Contenu :**

#### **ADMINISTRATION DE PROJETS :**

##### **Phase introductive du cours**

- Projet : définition et enjeux
- La gestion de projet
- Acteurs et responsabilités

##### **La méthodologie de conduite de projet**

- Fondements de la méthodologie
- Découpage et cycle de vie d'un projet
- Organisation du projet
- Gestion et pilotage
- Assurance et maîtrise de la qualité
- Communication et conduite du changement

##### **Détails des étapes et outils d'aide à la décision associés**

- Phase 1 : Etude de faisabilité ou étude préalable
- Phase 2 : La sélection du scénario
- Phase 3 : La conception
- Phase 4 : La réalisation
- Phase 5 : La fin de projet

#### **MANAGEMENT :**

On s'attachera dans un premier temps à :

- définir le management par projet,
- l'intégrer dans la représentation que chacun a du management,
- mesurer la part du management par projet dans l'organisation actuelle,
- d'échanger sur les atouts, les limites d'un tel type de management.

#### **LA CONDUITE DE PROJET A L'INSEE**

## Ecrit & oral : techniques de l'exposé

### Objectif :

Ecrire pour être lu... et entendu

### Pré-requis :

Cette formation commence par un rappel des « Techniques rédactionnelles ». Avoir suivi cet enseignement en formation initiale ou continue est un plus mais n'est pas indispensable.

### Contenu :

Cette formation commence par un rappel des techniques rédactionnelles, en particulier sur la structure des textes.

Elle enchaîne sur les techniques de l'exposé oral. Les premières portent sur les vecteurs de la communication orale : les attitudes, les gestes et la voix.

Sont ensuite abordés les principes qui doivent sous-tendre un exposé : une pensée organisée, une structure visible, un discours illustré, répété, adapté à son public.

Le bon usage des diaporamas fait l'objet de la dernière séquence.

Au cours de la formation, les participants sont mis en situation d'exposé afin d'appliquer les règles qui président à une prestation orale efficace, à la fois en termes de forme et sur le fond.

### À préparer avant la formation :

Les participants devront préparer un exposé d'une durée de 3 à 5 minutes sur un sujet de leur choix, personnel (vacances, hobby ou livre préféré...).

Pendant la formation, ils bâtiront un exposé professionnel d'une durée de 10 à 12 minutes. Pour ce faire, les participants devront **prévoir la documentation nécessaire**.

Remarque : pour chaque exposé, les participants pourront utiliser un diaporama.

