**Statistiques et analyse de données 1 (semestre 2)**

**Référent enseignement : Antoine Géré**

**Enseignant : Antoine Géré**

* Cours : 24 heures (16 séances)
* TD : 10.5 heures (7 séances)
* TP : 1.5 heures (1 séance)

**Objectifs pédagogiques :**

* Exploiter, résumer et représenter des données statistiques brutes
* Maitriser les connaissances de base en statistiques descriptives à une seule variable
* Maitriser les connaissances de base en statistiques descriptives à deux dimensions, permettant de réaliser des prévisions et d’établir des relations en coupe transversale et en série temporelle

**Programme de l’enseignement :**

1. Introduction
   1. Notions fondamentales
   2. Les tableaux et les représentations graphiques
2. Les valeurs centrales
   1. Les moyennes
   2. Les autres paramètres de position
3. Dispersion, concentration et forme
   1. Les paramètres de dispersion et de forme
   2. La concentration
4. Distribution à deux dimensions
   1. Les tableaux de contingence
   2. Les caractéristiques des lois marginales
   3. La covariance
   4. Les courbes de régression
5. Régression
   1. L'ajustement affine
   2. Le coefficient de corrélation
   3. Les équations des droites de régression
   4. L'analyse de la variance
6. Analyse des séries temporelles
   1. Les taux de croissance
   2. Les représentations graphiques des séries chronologiques
   3. La décomposition d'une série chronologique

**Modalités d’évaluations :**

* 1 Examen de 1h30 (80% de la note finale)
* 1 TD noté (20% de la note finale)

**Statistiques et analyse de données 2 (semestre 3)**

**Référent enseignement : Antoine Géré**

**Enseignant : Antoine Géré**

* Cours : 15 heures (10 séances)
* TD : 6 heures (4 séances)
* TP : 7.5 heures (5 séances)

**Objectifs pédagogiques :**

* Maitriser les lois usuelles des probabilités
* Travailler sur les variables aléatoires
* Maitriser les premières notions d’échantillonnages
* Appréhender les incertitudes associées aux données
* Estimer les paramètres d’une population
* Réaliser des tests statistiques sur les analyses comparatives

**Programme de l’enseignement :**

1. Rappels sur la statistique descriptive
2. Rappels sur les notions de probabilité et d’analyse combinatoire
   1. Permutations, arrangement, combinaison. Notions d’analyse combinatoire.
   2. Opérations générales sur les probabilités.
3. Variables aléatoires et lois des phénomènes aléatoires
   1. Variables aléatoires
   2. Variables aléatoires discrètes
   3. Variables aléatoires continues
   4. Lois usuelles des phénomènes aléatoires
   5. Lois discrètes et continues usuelle
4. Echantillonnage : deux catégories de méthodes d’échantillonnages
   1. Méthodes empiriques ou non-probabilistes ou raisonnées
      1. Méthode des quotas
      2. Echantillonnage à l’aveuglette
      3. Echantillonnage volontaire
      4. Echantillonnage au jugé
   2. Méthodes aléatoires – randomisation
   3. Echantillonnage Aléatoire Simple (EAS)
   4. Echantillonnage aléatoire stratifié
   5. Echantillonnage par grappe
   6. Echantillonnage à plusieurs degrés
5. Estimateurs et estimations
   1. Base de l’estimation
   2. Estimation d’une proportion
   3. Estimation d’une moyenne et d’un écart-type
6. 6- Tests statistiques paramétriques
   1. Notions générales sur les tests d’hypothèses
   2. Test du Chi-2
   3. Test de comparaisons de 2 moyennes
   4. Tests sur les fréquences et les variances.

**Modalités d’évaluations :**

* 1 Examen de 1h30 (80% de la note finale)
* 1 TD noté (20% de la note finale)

**Statistiques et analyse de données 3 (semestre 4)**

**Référent enseignement : Antoine Géré**

**Enseignant : Antoine Géré**

* Cours : 12 heures (8 séances)
* TD : 7.5 heures (5 séances)
* TP : 0 heure (0 séance)

**Objectifs pédagogiques :**

* Analyser et interpréter les traitements pour répondre à une problématique posée par :
  + Traiter des données quantitatives, qualitatives, mixtes dans le cadre des analyses de corrélations, des analyses de liaisons et des analyses causales.
  + Réaliser une ANOVA à un ou à deux facteurs, faire des Régressions Multiples.
* Acquérir des compétences dans le management, le traitement et l’analyse de données puisées dans un lac de données (data Lake).

**Programme de l’enseignement :**

1. Tests sur les fréquences
2. Tests sur les variances
3. Analyse de la variance (ANOVA à un et deux Facteurs)
   1. Conditions d’application
   2. Modèle d’analyse de la variance.
   3. Pratique (data, R-traitement, interprétation) de l’analyse de la variance.
4. Régression linéaire multiple (avec tests des coefficients)
5. Tests non- paramétriques essentiels
   1. Test de Dixon (valeurs aberrantes)
   2. Test de Grubbs
   3. Test de Shapiro-Wilk (Normalité)
   4. Test de Wilcoxon (symétrie)
   5. Test de Mann- Whitney
   6. Test de Kruskal-Wallis
   7. Test des échantillons appariés.
6. Première Expérience en analyse de données (PEAD)

**Première Expérience en analyse de données (PEAD)**

**Nous avons engagé en 2021-2022 un exercice PEAD (Première expérience en Analyse de Données) pour les étudiants 2A, dans le programme Statistiques et analyse de données 3. C’est un apprentissage qui se veut pluridisciplinaire axé sur l’organisation, le traitement et l’interprétation de** données.

**Objectif :**

Réaliser une première expérience en Data Mining.

* Pister des données pour répondre à une problématique formuler à partir de la méthodologie scientifique
* Mobiliser une méthodologie statistique formulée à partir de la méthodologie scientifique pour exploiter des informations.
* Prendre conscience des marges de vigilance nécessaire aux égards des données.

**Périmètre :** Pays « des Suds »

**Thématique :** en lien avec les objectifs de formation à l’ISTOM

**Dimension des données :** inf (dim) = 5 ( # variables), min ( Echantillon) = 100

**Matériels et outils :**

* Logiciel R + différents packages
* R-data iris, R-data crabs ( package MASS)
* <http://www.fao.org/faostat/en/#data>
* <https>[://comtrade.un.org/data/](https://comtrade.un.org/data/)
* <https://unctadstat.unctad.org/wds/ReportFolders/reportFolders.aspx?sCS_ChosenLang=en>
* <https://www.imf.org/en/Data>
* <https://data.worldbank.org/>
* <https://ec.europa.eu/eurostat/data/database>

NB : d’autres bases de données sont possibles après proposition au corp enseignant.

**Etapes et procédures :**

* 1 semaine : : construction des groupes (3 étudiants par groupe)
* Entre 1 et 2 mois : recherche de bases de données
* Au bout des 2 mois : dépôt d’un document pdf (4 diapos) d’investigation sur Moodle (thématique choisie et questionnement initial, justification et pertinence, lien thématique et base de données, interrogation de la base de données, démarche statistique proposée)
* 1 semaine plus tard : validation (données-thématiques)
* 2 jours plus tard : travaux d’analyses - encadrements
* 3 jours plus tard : Fin des travaux, le document à remettre

**Design document final :**

Document pdf structuré de 04 pages.

* Page 1 : Titre de l’étude, noms et prénoms des auteurs, année universitaire, problématisation, questionner la base de données. Proposer une méthode d’analyse statistique (Statistique descriptive, analyse de corrélation, analyse explicative, tendance et prédiction, analyse comparative et Test statistique). Résumer la démarche en exhibant les principaux socles de l’analyse visée.   
    
  **NB**: cette première page doit être suffisamment attractive pour inciter le lecteur à poursuivre son investigation.
* Pages 2 et 3 : présentation données, sources, origine, population cible, variables et des livrables (tableau de test, graphe, etc…) et ce, conformément aux étapes mentionnées à la page 01. Discussion sur les limites de l’analyse
* Page 4 : conclusion, références, annexes.

**NB :** il n’est pas nécessaire de lister tous les traitements statistiques entrepris ou réalisables. Seules les analyses qui répondent aux attentes de l’objectif principal seront présentées. La pertinence recherchée est l’obtention d’une intégration harmonieuse des objets statistiques tels que (courbes, graphes, nuage, tableaux de test, etc…) ainsi que les commentaires associés. Chacun de ces objets statistiques est numéroté, accompagné d’une légende en italique.

**Modalités d’évaluations :**

* 1 Examen de 1h30 (50% de la note finale)
* 1 Examen intermédiaire (25% de la note finale)
* PEAD (25% de la note finale)

**Statistiques et analyse de données 4 (semestre 5)**

**Référent enseignement : Antoine Géré**

**Enseignant : ??**

* Cours : 12 heures (8 séances)
* TD : 12 heures (8 séances)
* TP : 0 heure (0 séance)

**Objectifs pédagogiques :**

* Maîtriser les outils statistiques d’analyse de données en ayant développé sa capacité de raisonnement
* Maîtriser quelques méthodes d’analyse à plusieurs variables couramment utilisées lors du traitement des données
* Utiliser le logiciel R et certains de ses packages : Rcmdr, FactoMiner, factoextra, ade4, tseries, etc

**Programme de l’enseignement :**

1. Introduction
   1. Domaines d’application
   2. Place de l’analyse des données dans la démarche scientifique
   3. Choix des méthodes de l’analyse des données
2. Analyse Factorielle des Correspondances
   1. Tableau de contingence
   2. Notions de base théorique
   3. Test de Khi 2
   4. Démarche d’interprétation
3. Analyse des Composantes Principales
   1. Tableau de mesure
   2. Notions de base théorique
   3. Matrice des corrélations
   4. Démarche d’interprétation
4. Analyse des Correspondances Multiples
   1. Tableau disjonctif complet
   2. Hyper tableau de contingence
   3. Transformation des variables
   4. Démarche d’interprétation
5. Classification
   1. Méthode des nuées dynamiques (K-Means)
   2. Méthode CAH
6. Modélisation
   1. Modélisation par la Régression Linéaire Multiple (RLM)
   2. Modèle Logit (Régression logistique)

**Modalités d’évaluations :**

* 1 Examen de 1h30 (80% de la note finale)
* 1 Examen intermédiaire (20% de la note finale)

**Statistiques et analyse de données 5 (semestre 8)**

**EMOA** **1 : TECHNIQUES D'ENQUETES - CONSTRUCTION ET ANALYSE STATISTIQUE**

Contact Adrien Guétté (voir avec Brice)

**Référent enseignement : Antoine Géré**

**Enseignant : ??**

* Cours : 15 heures (10 séances)
* TD : 18 heures (12 séances)
* TP : 0 heure (0 séance)

**Objectifs pédagogiques :**

L’art de l’entretien : Introduction

* Comprendre l’intérêt potentiel et les conditions de pertinence de différentes techniques d’enquête (questionnaire, semi-directif, non directif), de leur conception à l’interprétation de leurs résultats
* Être capable de choisir une technique en fonction des objectifs visés et des contraintes du contexte
* Commencer à saisir l’importance de :
  + la posture, à l’attitude de l’enquêteur, notamment dans la qualité de l’écoute
  + la formulation des questions et la conception d’un guide d’enquête
  + l’art de la relation
* Acquérir quelques réflexions ou réflexes pratiques et tactiques variés : comment tisser la confiance, confidentialité, comment obtenir un rendez-vous, comment faire face aux difficultés de rencontrer les enquêtés, etc.

DATA-MINING - Traitement des données d’enquête

L’objectif principal de ce cours est de savoir utiliser des outils, mobiliser des méthodes de la statistique pour extraire, valoriser les données de natures diverses (qualitatives, quantitatives, mixtes, textuelles) provenant des travaux d’enquêtes, et ce, dans le but de classer, regrouper, comparer les individus ou groupe d’individus  d’une population statistique ciblée ou de faire des prédictions.

Signalons que les données peuvent provenir de R-data, de sites internet, d’expérimentations, des réseaux sociaux, etc.

Outils mobilisés : R et packages (ade4, adegraphics, car, Rcmdr- FactoMineR, factoextra, Rtemis, MASS, Tserie).

**Programme de l’enseignement :**

* Demi-journée 1 : 2h conduire un entretien / repro (guide d’observation)
* Demi-journée 2 : 1h formuler des questions / 30min pratique de questionnaire en ligne
* Demi-journée 3 : 3h concevoir un dispositif d’enquête
* Demi-journée 4 : ?

**Modalités d’évaluations :**

* 1 TD noté (100% de la note finale)

**Statistiques et analyse de données 6 (semestre 8)**

**EMOA 2 : Analyse de données et économétrie appliquée au développement**

Contact Adrien Guétté (voir avec Brice)

**Référent enseignement : Antoine Géré**

**Enseignant : ??**

* Cours : - heures (- séances)
* TD : - heures (- séances)
* TP : - heure (- séance)

**Objectifs pédagogiques :**

??

**Programme de l’enseignement :**

??

**Ressources Année universitaire 2024 - 2025**

**Statistiques et analyse de données 7 (semestre 9 ou 10)**

Cours dispensés en 5A dans les différentes options/DA.