

Module : Méthodes statistiques et Etude de données				Code
				ING-4-SDIA-S8-P3
Période	S8-P1	Volume horaire	42	ECTS 4

Responsable	Ahmed Dhouibi	email	Ahmed.dhouibi@yahoo.fr
Equipe pédagogique	Ahmed Dhouibi.		

1. Objectifs de Module (*Savoirs, aptitudes et compétences*)

Ce module porte sur les méthodes d'analyse statistique dans les cas uni varié et bi varié.

Acquis d'apprentissage :

A la fin de cet enseignement, l'élève sera capable de :

- Maîtriser les notions de base des méthodes statistiques multi variées. (**C1.2**)
- Caractériser les spécificités de chaque méthode. (**C1.2**)
- Simuler et tester des hypothèses émises sur certains modèles de prédiction. (**C1.3**)
- Concevoir la méthode adéquate d'analyser un data frame. (**C1.1**)
- Communiquer les principaux résultats et les valider (**C3.3**)
- Modéliser une série de données (**C3.2**)
- Prédire la variable dépendante (**C2.2**)
- Faire un diagnostic du modèle (**C2.3**)

2. Pré-requis (*autres UE et compétences indispensables pour suivre l'UE concernée*)

- Mathématiques de l'ingénieur
- Théorie de la probabilité
- Analyse statistique
- Inférence statistique

3. Répartition d'Horaire de Module

Intitulé de l'élément d'enseignement	Total	Cours	TD	Atelier	PR
Module : Méthodes statistiques et étude de données	42	24	12	6	

4. Méthodes pédagogiques et moyens spécifiques au Module

(*pédagogie d'enseignement, ouvrages de références, outils matériels et logiciels*)

- Supports de Cours
- Projecteur et Tableau

- Travaux dirigés
- Logiciels de simulation : R, Python

Bibliographie

Titre	Auteur(s)	Edition
Introductory Econometrics A Modern Approach	Jeffrey M. Wooldridge Wooldridge	2017
Data mining for business analysis	Shumelly Bruce Patel	Wiley 2020
Statistics controversies and concepts	David Moore and Co	2019
"An Introduction to Multivariate Statistical Analysis" par T.W. Anderson.	R.A. Johnson	2022
"Applied Multivariate Statistical Analysis" .	D.W. Wichern	2021

5. Contenu (Descriptifs et plans des cours / Déroulement / Détail de l'évaluation de l'activité pratique)

Durée allouée

Module 1 : Méthodes statistiques et Etude de données

Séance 1	Cours	3H
<ul style="list-style-type: none"> • Régression simple et multiple <ul style="list-style-type: none"> - Modèle de régression simple - Modèle de régression multiple - Significativité d'un modèle de données - Prédiction 		
Séance 2	TD+ Atelier	3H
<ul style="list-style-type: none"> • TD activités et TP 		
Séance 3	Cours	3H
<ul style="list-style-type: none"> • Régression non linéaire <ul style="list-style-type: none"> - Polynomiale (quadratique, cubique) - Détection du meilleur modèle 		
Séance 4	TD+ Atelier	3H
<ul style="list-style-type: none"> • TD Activités et TP 		
Séance 5	Cours	3H
<ul style="list-style-type: none"> • Régression logistique <ul style="list-style-type: none"> - Cas où y est binaire - Cas où y multinomiale 		
Séance 6	TD + Atelier	3H
<ul style="list-style-type: none"> • Activités (TD3) et Atelier sur R 		
Séance 7	Cours-TD	3H
<ul style="list-style-type: none"> • Traitement des cas particuliers <ul style="list-style-type: none"> - Autocorrélation des erreurs 		

- Hémoscédasticité - Multicolinéarité des variables		
Séance 8 <ul style="list-style-type: none">● Analyse en composantes principales ACP (PCA)<ul style="list-style-type: none">- Utilité et intérêt- Démarche de l'ACP- Activités pratiques	Cours-TD	3H
Séance 9 <ul style="list-style-type: none">● Analyse factorielle des correspondances AFC (CA)<ul style="list-style-type: none">- Intérêt- Cadre d'étude- Nature des variables- Activités et Atelier	Cours-TD	3H
Séance 10 <ul style="list-style-type: none">● Analyse factorielle des correspondances multiples (ACM)<ul style="list-style-type: none">- Cas d'étude- Intérêt- Activités	Cours-TD	3H
Séance 11 <ul style="list-style-type: none">● Analyse discriminante<ul style="list-style-type: none">- Analyse discriminante linéaire- Analyse discriminante non linéaire- Validation des modèles : Matrices de confusion, taux d'erreur, courbes ROC- Activités	Cours-TD	3H
Séance 12 <ul style="list-style-type: none">● Analyse canonique<ul style="list-style-type: none">- Cadre d'étude- Activités	Cours-TD	3H
Séance 13 <ul style="list-style-type: none">● Clustering et Classification<ul style="list-style-type: none">- Clustering hiérarchique : Algorithmes agglomératifs et divisifs, dendrogrammes.- K-means et K-medoids : Algorithmes, choix du nombre de clusters, interprétation.- Techniques avancées : Clustering basé sur la densité (DBSCAN), méthodes de classification supervisée (k-NN, SVM).	Cours-TD	3H
Séance 14 <ul style="list-style-type: none">● TD 6 activités sur R et Présentation des projets	TD	3H

6. Mode d'évaluation de Module(nombre, types et pondération des contrôles)

Eléments d'enseignement	Coeff	DS	EX	TP	PR
Module – Méthodes statistiques et étude de données	2	30%	50%	20%	

Pour valider le module, les étudiants passeront un examen dont le coefficient est de 50%, un DS dont le coefficient est de 30% et un TP dont le coefficient est de 20%.

La durée de tous les examens (Examen, DS...) est de 1h30.

Le DS est planifié 7 semaines après le début du module et portera sur les thématiques des sept premières semaines.

Quant à l'examen, il est planifié après l'écoulement des 14 semaines et portera sur toutes les thématiques enseignées tout au long des 42 heures.

Concernant le TP, il est planifié une semaine avant l'examen et testera les connaissances acquises tout au long du module.

Le module est validé si l'étudiant obtient une moyenne supérieure ou égal à 10 sur 20.

Dans l'ensemble, l'examen vise à évaluer les étudiants à la fois sur leurs capacités à maîtriser les notions de base des méthodes statistiques multi variées, caractériser les spécificités de chaque méthode, simuler et tester des hypothèses émises sur certains modèles de prédiction, concevoir la méthode adéquate d'analyser un data frame et communiquer les principaux résultats et les valider.