

Sigle : INF5103 Gr. 01**Titre : Concepts statistiques pour la science des données****Session : Automne 2026, Horaire et local****Professeur : Allaoui, Mohamed Lamine****1. Description du cours paraissant à l'annuaire :****Objectifs**

Maîtriser les concepts statistiques avancés utilisés dans la science des données. Développer les connaissances pour développer des algorithmes d'analyse et de prédiction à partir des données en utilisant ces concepts statistiques.

Contenu

Rappels sur les concepts de probabilités et statistiques. Statistique et science des données. Modèles paramétriques vs. non paramétriques. Techniques d'échantillonnage des données et estimation de paramètres. Modèles statistiques pour la classification et la régression. Tests d'hypothèses. Propriétés d'un paramètre statistique: biais, consistance, efficacité. Maximum de vraisemblance. Statistique bayésienne. Analyse factorielle et analyse de variance. Réduction de dimensions. Modèles graphiques probabilistes. Méthodes de Monte-Carlo. Réduction du biais statistique dans l'analyse de données. Études d'applications (ex. régression, classification, ordonnancement, etc.).

Descriptif – Annuaire**2. Objectifs spécifiques du cours :**

Au terme de cette activité, l'étudiant(e) aura acquis des connaissances suivantes:

- Les principes de la science des données;
- Les concepts statistiques importants pour la science des données
- Les statistiques descriptives;
- Les statistiques inférentielles;
- L'apprentissage automatique à partir de données;

3. Stratégies pédagogiques :

Séances de cours en présentiel, de 3h/semaine comprenant une ou plusieurs stratégies :

- Cours magistral

Évaluations :

- Devoirs
- Examen intra (en présentiel)
- Examen final (en présentiel)

4. Heures de disponibilité ou modalités pour rendez-vous :

Sur demande.

L'étudiant(e) peut envoyer un courriel pour fixer un rendez-vous.

Courriel : mohamedlamine.allaoui@uqo.ca

5. Plan détaillé du cours sur 15 semaines :

Semaine	Thèmes	Dates
---------	--------	-------

1	Généralités • Phénomène des mégadonnées (big data). • Données structurées et non-structurées. • Principes de la science des données. • Statistique pour la science des données.	08 sept. 2026
2	Éléments de calcul des probabilités • Expérience aléatoire. • Variables aléatoires et propriétés. • Loi de Bayes et loi marginale. • Lois de probabilités discrètes & continues.	15 sept. 2026
3	Statistique descriptive • Prétraitement des données. • Population versus échantillon. • Techniques d'échantillonnage. • Mesures de tendance des données. • Visualisation des données. Distribution de l'énoncé devoir	22 sept. 2026
4	Statistique inférentielle • Modélisation à partir de données. • Modèles paramétriques et non-paramétriques. • Propriétés des paramètres : biais, consistance, efficacité. • Estimation de paramètres	29 sept. 2026
5	Modèles statistiques supervisés I (régression) • Principes de la régression. • Régression linéaire et non-linéaire. • Concept de sur-apprentissage. • Concept de validation croisée. Distribution de l'énoncé du projet Remise du devoir (Date limite : 09 octobre 2026 à 23 h 59)	06 oct. 2026
6	Semaine d'études	13 oct. 2026
7	Examen intra	20 oct. 2026
8	Modèles statistiques supervisés II (classification) • Méthode du Bayes naïf pour la classification. • Méthode de la régression logistique.	27 oct. 2026
9	Modèles non-paramétriques supervisés • Distributions non-paramétriques de données. • Distributions de données temporelles (chaines de Markov).	03 nov. 2026

	• Modèles supervisés basés sur les réseaux de neurones.	
10	Modèles statistiques non-supervisés I (regroupement) • Méthode des K-moyennes. • Méthode des mélanges Gaussiens. • Validation de groupes.	10 nov. 2026
11	Présentation des projets	17 nov. 2026
12	Modèles statistiques non-supervisés II (Reduction de dimensions) • Analyse à composantes principales • Analyse factorielle • Autoencoders	24 nov. 2026
13	Dérive de distribution • Origines de la dérive et effets sur les modèles d'apprentissage automatique • Types de dérive • Détection et surveillance • Méthodes d'atténuation	01 déc. 2026
14	Étude de quelques applications • Classification de documents multimédias • Détection d'intrusions en cybersécurité • Analyse de risques • Détection d'anomalies	08 déc.2026
15	Examen final	15 déc. 2026
6. Évaluation du cours :		

L'étudiant(e) dans ce cours sera évalué(e) par les examens de mi-session et final, ainsi que par des projets de session. La pondération de la note finale se fera comme suit :

- **Devoir : 15 %**
- **Projet : 25 %**
- **Examen de mi-session : 30 %**
- **Examen final : 30 %**

Pour les projets, l'évaluation sera répartie comme suit :

- **Devoir : 15 points** : Il s'agit d'implanter quelques méthodes statistiques pour la science des données en utilisant le langage Python.
- **Projet : 25 points** : Il s'agit de créer une application ou un outil de mesure statistique pour résoudre un problème du monde réel.

Tout retard dans la remise d'un travail entraîne une pénalité de **15 %/jour** sur la note attribuée à ce travail, **jusqu'au maximum d'une semaine**.

7. Politiques départementales et institutionnelles :

- Politique du département d'informatique et d'ingénierie relative à la tenue des examens
- Note sur le plagiat et sur la fraude
- Politique relative à la qualité de l'expression française écrite chez les étudiants et les étudiantes de premier cycle à l'UQO
- Absence aux examens : cadre de gestion, demande de reprise d'examen (formulaire)

Tolérance **ZÉRO** en matière de violence à caractère sexuel.

Le Bureau d'intervention et de prévention en matière de harcèlement (BIPH) a pour mission d'accueillir, soutenir et guider toute personne vivant une situation de harcèlement, de discrimination ou de violence à caractère sexuel. Le BIPH oriente ses actions afin de prévenir les violences à caractère sexuel pour que nous puissions étudier, travailler et s'épanouir dans un milieu sain et sécuritaire.

Vous vivez ou êtes une personne témoin d'une situation de violence à caractère sexuel ? Vous êtes une personne membre de la communauté étudiante ou une personne membre du personnel, autant à Gatineau qu'à Ripon et St-Jérôme, l'équipe du BIPH est là pour vous, sans jugement et en toute confidentialité.

Ensemble, participons à une culture de respect.

Pour de plus amples renseignements consultez UQO.ca/biph ou écrivez-nous au Biph@uqo.ca

8. Principales références :

1. Gareth James, Daniella Witten, Trevor Hastie, Robert Tibchirani, Jonathan Taylor. An Introduction to Statistical learning with Applications in Python. July 2023.
2. K-P. Murphy. Probabilistic machine Learning. MIT press, 2022.
3. Stanley H. Chan. Introduction to Probability for Data Science. Michigan Publishing Services, 2021. (<https://probability4datascience.com/index.html>)
4. Peter Bruce, Andrew Bruce, and Peter Gedeck. Practical Statistics for Data Scientists. O'Reilly books, 2020.
5. P-N. Tan, M. Steinbach, A. Karpatne, V. Kumar. Introduction to Data Mining (2nd Edition), Pearson 2018.
6. Witten et al. Data mining : Practical machine learning Tools and Techniques. Morgan Kaufmann, 2017

9. Page Web du cours :

<https://moodle.uqo.ca>