

**Sigle : INF5103 Gr. 01**

**Titre : Concepts statistiques pour la science des données**

**Session : Hiver 2025, de 12 h 30 à 15 h 30 Horaire et local**

**Professeur : Hamadene, Assia**

**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 :**

Assia.hamadene@yahoo.fr

**5. Plan détaillé du cours sur 15 semaines :**

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

1	<b>Généralités</b> <ul style="list-style-type: none"> <li>• Phénomène des mégadonnées (big data).</li> <li>• Données structurées et non-structurées.</li> <li>• Principes de la science des données.</li> <li>• Statistique pour la science des données.</li> </ul>	15 Janv. 2025
2	<b>Éléments de cucul des probabilités</b> <ul style="list-style-type: none"> <li>• Expérience aléatoire.</li> <li>• Variables aléatoires et propriétés.</li> <li>• Loi de Bayes et loi marginale.</li> <li>• Loïs de probabilités discrètes &amp; continues.</li> </ul>	22 Janv. 2025
3	<b>Statistique descriptive</b> <ul style="list-style-type: none"> <li>• Prétraitement des données.</li> <li>• Population versus échantillon.</li> <li>• Techniques d'échantillonnage.</li> <li>• Mesures de tendance des données.</li> <li>• Visualisation des données.</li> </ul> <b>Travail pratique I</b>	29 Janv. 2025
4	<b>Statistique inférentielle</b> <ul style="list-style-type: none"> <li>• Modélisation à partir de données.</li> <li>• Modèles paramétriques et non-paramétriques.</li> <li>• Propriétés des paramètres : biais, consistance, efficacité.</li> </ul>	05 Fév. 2025
5	<b>Estimation de paramètres</b> <ul style="list-style-type: none"> <li>• Estimation par intervalle de confiance.</li> <li>• Tests d'hypothèses statistiques.</li> <li>• Méthode du maximum de vraisemblance.</li> <li>• Estimation Bayésienne.</li> </ul> <b>Travail pratique II</b>	12 Fév. 2025
6	<b>Modèles statistiques supervisés I (régression)</b> <ul style="list-style-type: none"> <li>• Principes de la régression.</li> <li>• Régression linéaire et non-linéaire.</li> <li>• Concept de sur-apprentissage.</li> <li>• Concept de validation croisée.</li> </ul>	19 Fév. 2025
7	<b>Modèles statistiques supervisés II (classification)</b> <ul style="list-style-type: none"> <li>• Méthode du Bayes naïf pour la classification.</li> <li>• Méthode de la régression logistique.</li> </ul> <b>Travail pratique III</b>	26 Fév. 2025
8	<b>Semaine d'études</b>	03 au 07 mars 2025
9	<b>Examen intra</b>	12 mars 2025

10	<b>Modèles non-paramétriques supervisés</b> <ul style="list-style-type: none"> <li>• Distributions non-paramétriques de données.</li> <li>• Distributions de données temporelles (chaines de Markov).</li> <li>• Modèles supervisés basés sur les réseaux de neurones.</li> </ul>	19 mars 2025
11	<b>Modèles statistiques non-supervisés I (regroupement)</b> <ul style="list-style-type: none"> <li>• Méthode des K-moyennes.</li> <li>• Méthode des mélanges Gaussiens.</li> <li>• Validation de groupes.</li> </ul> <b>Travail pratique IV</b>	26 mars 2025
12	<b>Modèles statistiques non-supervisés I (Reduction de dimensions)</b> <ul style="list-style-type: none"> <li>• Analyse à composantes principales</li> <li>• Analyse factorielle</li> <li>• Auto-encoders</li> </ul> <b>Étude de quelques applications</b> <ul style="list-style-type: none"> <li>• Détection d'anomalies</li> <li>• Classification de documents multimédias</li> <li>• Analyse de risques</li> </ul>	02 avril 2025
13	<b>Modèles Génératifs</b> <ul style="list-style-type: none"> <li>• GAN: Réseaux Génératifs Antagonistes.</li> <li>• VAE: Auto-encodeurs Variationnels.</li> </ul>	09 avril 2025
14	<b>Présentation des projets</b>	16 avril 2025
15	<b>Examen final</b>	23 avril 2025

## 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 :

- **Devoirs : 15 %**
- **Projets : 25 %**
- **Examen de mi-session : 30 %**
- **Examen final : 30 %**

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

- **Devoirs : 15 points** : Il s'agit d'implanter quelques méthodes statistiques pour la science des données en utilisant le langage Python.
- **Projets : 25 points** : Il s'agit de faire des projets sur les techniques statistiques pour l'analyse de données. Ces derniers peuvent être dans la description, l'analyse ou la modélisation statistique des données. Ils peuvent aussi être reliés à l'évaluation statistique des techniques d'apprentissage automatique. Quelques projets seront fournis par le professeur. Néanmoins, les étudiants auront la liberté de choisir des projets dans des domaines d'applications qu'ils souhaitent. L'évaluation se fera sur deux volets :

- Un rapport de 15 pages. Les normes de présentation de travaux (p. ex. page de garde, marge d'un pouce, interligne à 1,5, taille des caractères de 12 points) doivent être **absolument** respectées.
- Une présentation de 15 à 20 minutes sera faite en classe. Les projets se feront individuellement ou en équipes de deux si le nombre d'étudiants inscrits dépasse 10.

Une moyenne générale inférieure à **64 %** est éliminatoire et conduit automatiquement à l'échec de l'étudiant(e).

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**. La qualité du français sera considérée lors de la correction des travaux.

## 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](https://uqo.ca/biph) ou écrivez-nous au [Biph@uqo.ca](mailto: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>