

## Plan de cours

Faculté des arts et des sciences  
Département de sciences biologiques

<b>Sigle du cours et section</b>	<b>BIO2042</b>	<b>Hiver 2019</b>
<b>Titre du cours</b>	<b>Biostatistique II</b>	
<b>Crédits</b>	<b>3</b>	
<b>Horaire</b>	<b>Théorie</b> : Lundi, 13h00 - 15h00 <b>Pratique</b> : Lundi, 15h00 - 18h00	
	<b>Salle</b> : B-4245, Pav. 3200 Jean-Brillant (théorique) et B-4250, Pav. 3200 Jean-Brillant (pratique)	
<b>Professeurs</b>	Guillaume Guénard et Patrick James	
<b>Local</b>	Patrick James : F-062 Mardi 11h00-16h00.	
<b>Courriel</b>	<a href="mailto:guillaume.guenard@umontreal.ca">guillaume.guenard@umontreal.ca</a> ; <a href="mailto:patrick.ma.james@gmail.com">patrick.ma.james@gmail.com</a>	
<b>Téléphone</b>	514-343-6111 ext. 1233	
<b>Démos</b>	Julian Wittische <a href="mailto:julian.wittische@umontreal.ca">julian.wittische@umontreal.ca</a> (Vendredi 13h00 à 15h00, Local F-062), Frédéric Dwyer-Samuel, Élodie Basque	

ÉVALUATION			
Examens	Pondération	Date d'attribution	Date de remise
Intra	30 %	25 février 2018	4 mars 2018; 23h59
Rapport d'analyse (Équipe de 2)	10 %	8 avril 2018	10 avril 2018 23h59
Examen Sommatif	38 %	15 avril 2018	26 avril 2018; 23h59
Travaux pratique	22 %	-	À la fin de chaque cours

Les examens prendront la forme de deux examens à la maison. Ces examens comporteront des questions sur l'analyse et la manipulation des données, de la programmation en R, et des critiques de lectures. Les étudiants auront une semaine pour les compléter. Les rapports seront remis à la date indiquée ci-dessus avant 23h59 sur Studium. Les remises de travaux en retard seront pénalisées à un taux de 1% par heure.

Il est attendu que les étudiants apportent leur propre ordinateur avec R (dernière version stable) et Rstudio Desktop installés. Veuillez informer le professeur si vous n'avez pas accès à un ordinateur portable pour ces séances d'informatique. R et Rstudio peuvent être téléchargés **librement et sans frais** à ces adresses: <https://cran.r-project.org/> et <https://www.rstudio.com/products/rstudio/download/>

### BUT DU COURS

Le but du cours est de présenter aux étudiants un ensemble de méthodes et techniques d'analyse statistique parmi les plus couramment utilisées dans les domaines de l'écologie et de la biologie. Le cours mettra l'emphase sur l'évaluation préliminaire des données, laquelle est préalable à l'ensemble des analyses statistique, ainsi que sur les modèles linéaires **généraux** (i.e., la régression linéaire et l'ANOVA) et **généralisés** (i.e., la régression logistique et de Poisson). Les étudiants acquerront également de l'expérience dans l'utilisation du langage et environnement de programmation R qui sera utilisé pour la manipulations des données et leur analyse.

Chaque cours débutera avec environ deux heures de théorie qui seront suivies de 2 à 3 heures de travaux pratiques. Durant ces travaux pratiques, les étudiants feront sur place (et sur leur propre ordinateur) des exercices formatifs en langage R portant sur le sujet du cours. Les étudiants soumettront ces exercices à la fin de chaque séance de travaux pratiques ; note leur sera attribuée à raison de 2% de la note finale par semaine.

## OBJECTIFS GÉNÉRAUX D'APPRENTISSAGE

À la fin du cours, l'étudiant devra répondre aux objectifs suivants :

### Exploratoire :

- Effectuer l'évaluation préliminaire de jeux de données biologiques à des fins d'analyse statistique afin de déceler les problèmes potentiels d'applicabilité des méthodes et d'identifier des solutions de rechange
- Choisir la ou les méthodes d'analyses appropriées à des situations pratiques
- Lire et critiquer des articles scientifiques employant de méthodes d'analyse statistique.

### Technique :

- Comprendre les fondements théoriques des modèles linéaires généraux et généralisés (LM et GLM) et savoir appliquer ces analyses dans des contextes appropriés
- Comprendre et savoir appliquer les outils de diagnostics ainsi que de validation des modèles linéaires généraux et généralisés
- Savoir interpréter les résultats des modèles linéaires généraux et généralisés
- Connaître les bases de la programmation en langage R pour la manipulations des données
- S'initier aux structures de contrôles avec le langage R


## SYLLABUS GÉNÉRAL

Semaine	Date	Sujet	Lecture
1	7 jan.	Introduction au cours et à la modélisation statistique Rappel sur la corrélation et la régression linéaire Rappel sur la programmation en R	Motulsky Ch. 32 – 33
2	14 jan.	Suppositions de la régression linéaire; diagnostics	Motulsky Ch. 34; Dalgaard Ch. 6
3	21 jan.	La régression multiple et polynomiale	Motulsky Ch. 37, pp 296 - 305; Dalgaard Ch. 11
4	28 jan.	ANOVA avec deux critères de classification, ANOVA hiérarchique	Motulsky Ch. 39; Dalgaard Ch. 12.1 - 12.6
5	4 fev.	ANCOVA; conception expérimentale	Whitlock & Schluter Ch. 18; Dalgaard Ch. 12.7 - 12.8
6	11 fev.	Comment identifier un bon modèle? La multicolinéarité	Motulsky Ch. 38
7	18 fev	Comparaison des modèles avec AIC, validation croisée	Johnson and Omland, 2004.
-	25 fev	Synthèse/Questions <b>15h00 : Attribution DE L'EXAMEN INTRA</b>	Whitman <i>et al.</i> 2004
-	4 mars	<b>SEMAINE DE LECTURE</b>	
8	11 mars	Modèles linéaires <b>généralisés</b> (GLMs); la famille exponentielle, estimation des paramètres par le maximum de vraisemblance (« <i>maximum likelihood</i> »)	Whitlock & Schluter Ch. 20
9	18 mars	La régression logistique I	Motulsky Ch. 37, pp 305 – 308 Dalgaard Ch. 13
10	25 mars	La régression logistique II, validation avec les « matrices de confusion »; ROC	Motulsky Ch. 42
11	1 avril	La régression Poisson pour les données de compte	Motulsky Ch. 6; Dalgaard Ch. 15
12	8 avril	Pour aller plus loin : Modèle non-linéaire, modèles mixtes, modèle additif généralisé ( <i>si les temps le permettent</i> ).	Motulsky Ch. 44; Dalgaard Ch. 16

		<b>TP : Attribution de la question/données pour le rapport d'analyse:</b>	
-	15 avril	Synthèse/Questions <b>15h00 : Attribution de L'EXAMEN FINAL</b>	

## RÉFÉRENCES ET DOCUMENTATION

- Motulsky. **Biostatistique – Une Approche Intuitive**, 2me ed. De Boeck. *En réserve à la bibliothèque EPC copies des chapitres numériques sont disponible sur StudiUM*
- Dalgaard. **Introductory Statistics with R**. *Copie numerique disponible sur StudiUM*
- Whitlock et Schluter. 2015. **The Analysis of Biological Data**, 2<sup>nd</sup> Ed. Roberts and Company. *En réserve à la bibliothèque EPC ; copies des chapitres numériques sont disponible sur StudiUM*

 <span style="float: right;">Bibliothèque ÉPC-Biologie</span>
<a href="http://guides.bib.umontreal.ca/disciplines/6--Sciences-biologiques">Guide en Sciences biologiques</a> (point de départ, ressources utiles, astuces) <a href="http://guides.bib.umontreal.ca/disciplines/6--Sciences-biologiques">http://guides.bib.umontreal.ca/disciplines/6--Sciences-biologiques</a>
Recherche dans le <a href="http://atrium.umontreal.ca/">catalogue Atrium</a> (livres, thèses UdeM, audiovisuel, titres de revues) <a href="http://atrium.umontreal.ca/">http://atrium.umontreal.ca/</a>
Recherche dans les <a href="http://www.bib.umontreal.ca/Maestro">bases de données</a> (articles scientifiques, statistiques, thèses, etc.) <a href="http://www.bib.umontreal.ca/Maestro">http://www.bib.umontreal.ca/Maestro</a>

## PLAGIAT

Nous vous invitons à consulter le règlement disciplinaire de l'Université sur le site suivant :  
<http://www.fas.umontreal.ca/plagiat/>

**NB : ce plan de cours provisoire est sujet à des changements jusqu'au moment de son acceptation officielle lors de la première séance de cours**