

Heure et salle

- Heure : Mercredi, 9h45—11h15
- Salle : (visio : lien Zoom/password disponible sur moodle)

Page du cours

Supports : <http://josephsalmon.eu/HLMA408.html>

Dépôts (Moodle) : <https://moodle.umontpellier.fr/course/view.php?id=440>

Enseignants

- Joseph Salmon, joseph.salmon@umontpellier.fr (Univ. Montpellier)
- Pierre-Louis Montagard, pierre-louis.montagard@umontpellier.fr (Univ. Montpellier)
- Thierry Mignon, thierry.mignon@univ-montp2.fr (Univ. Montpellier)
- Tristan Roget, tristan.roget@gmail.com (Univ. Montpellier)
- Raphaël Paegelow, raphael.paegelow@ens-lyon.fr (Univ. Montpellier)
- Colin Thomas, colin.thomas@cirad.fr (CIRAD)
- Tiffany Cherchi, tiffany.cherchi@umontpellier.fr (Univ. Montpellier)

Horaires de consultation (*office hours*)

- Mercredi 17h-19h, uniquement sur rendez-vous préalable par mail.

Prérequis et ressources pour se mettre à niveau

Les étudiants sont supposés connaître les bases de la théorie des probabilités et de l'algèbre linéaire.

Probabilités : Pour un démarrage en probabilité, les ouvrages suivants en français, permettent une première assimilation des concepts [CGDM01], [Ouv08], [Ouv07], [GS01].

Algèbre Linéaire : Il faut être suffisamment à l'aise avec le calcul matriciel pour bien commencer avec les modèles linéaires. Les livres [Sch05] et [Gv13] peuvent être un bon début. Les classiques du genre sont notamment [Gv13, HJ94]. Enfin, [TB97] est un bon point d'entrée. D'autre part, le contenu de calcul matriciel de HLMA314 est supposé suffisant pour cette partie.

Algorithmique Les étudiants doivent aussi être capables d'implémenter des méthodes numériques basiques en utilisant un langage de haut niveau : Python sera la choix par défaut.

Pour démarrer en Python, des documents d'introduction sont disponibles à cette adresse : <http://josephsalmon.eu/HLMA310.html>, cours HLMA310 "logiciels scientifiques". Pour aller plus loin, une référence contenant la plupart des algorithmes usuels en informatique est le livre [CLRS01].

Description du cours

Après un bref aperçu de statistiques descriptives, ce cours introduira les méthodes élémentaires de statistiques et les concepts fondamentaux que sont les tests, les intervalles de confiance et l'estimation ponctuelle. Nous aborderons avec des exemples applicatifs les problématiques des modèles linéaires ainsi que de l'analyse de variance (ANOVA).

Notation

Ce cours est 100% CC et comporte deux quizzes et un TP noté, qui porteront tant sur les aspects statistiques de sur leur implémentation. La répartition des notes se fait comme suit :

- Quiz : 25% (**17/03/2021**)
- TP Noté : 35% (donné le 12/04/2021, et à rendre pour le **16/04/2021, 23h59**)
- Quiz final : 40% (**05/05/2021**)

Bonus

2 pts supplémentaires sur **la note finale** peuvent être obtenus pour toute contribution à l'amélioration des cours (présentations, codes, etc.) sous les contraintes suivantes.

Contraintes :

- on obtient 1 point par amélioration
- seule la première amélioration reçue est "rémunérée", les autres ne rapportent plus rien
- déposer un fichier au format **.md**¹ (de taille <10 ko) en créant une fiche sur Moodle, dans la section "**Bonus - Proposition d'amélioration**"
- détailler précisément (ligne de code, page des présentations, page du poly, etc.) l'amélioration proposée, ce qu'elle corrige et/ou améliore
- pour les fautes d'orthographe : proposer au minimum 5 corrections par contribution
- chaque élève ne peut gagner que 2 points maximum de cette manière

Thème de chaque séance

Semaine 1 : Consignes et introduction

Semaine 2 : Statistiques descriptives

Semaine 3 : Échantillonnage aléatoire simple

Semaine 4 : Modèle statistique

Semaine 5 : Estimation de paramètre

Semaine 6 : Intervalle de confiance / tests cas gaussiens

Semaine 7 : Tests

Semaine 8 : Régression linéaire (I)

Semaine 9 : Régression linéaire (II)

Semaine 10 : Analyse de la variance

1. <https://fr.wikipedia.org/wiki/Markdown>

Semaine 11 : Quiz

Livres et ressources pour l'introduction aux statistiques

- **Modern Statistics for Modern Biology** ; S. Holmes, W. Huber, 2018, [HH18]
<http://web.stanford.edu/class/bios221/book/> et le cours en ligne (en anglais)
https://www.huber.embl.de/msmb/course_spring_2020/lectures.html
- **Statistique générale pour utilisateurs. Méthodologie**, J. Pagès, 2010, [Pag10]
- **Statistical Rethinking : A Bayesian Course with Examples in R and Stan**
R. McElreath, 2015 [McE15]

Références

- [CGDM01] M. Cottrell, V. Genon-Catalot, C. Duhamel, and T. Meyre. *Exercices de probabilités, Licence - Master - Écoles d'ingénieur*. Cassini, 3^e édition, 2001.
- [CLRS01] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to algorithms*. MIT press, 2001.
- [GS01] G. R. Grimmett and D. R. Stirzaker. *Probability and random processes*. Oxford University Press, New York, third edition, 2001.
- [Gv13] G. H. Golub and C. F. van Loan. *Matrix computations*. Johns Hopkins University Press, Baltimore, MD, fourth edition, 2013.
- [HH18] S. Holmes and W. Huber. *Modern statistics for modern biology*. Cambridge University Press, 2018.
- [HJ94] R. A. Horn and C. R. Johnson. *Topics in matrix analysis*. Cambridge University Press, Cambridge, 1994. Corrected reprint of the 1991 original.
- [McE15] R. McElreath. *Statistical Rethinking : A Bayesian Course with Examples in R and Stan*. Chapman & Hall/CRC Texts in Statistical Science. Chapman and Hall/CRC, 2015.
- [Ouv07] J-Y. Ouvrard. *Probabilités : Tome 2, Licence - CAPES*. Enseignement des mathématiques. Cassini, 2 édition, 2007.
- [Ouv08] J-Y. Ouvrard. *Probabilités : Tome 1, Licence - CAPES*. Enseignement des mathématiques. Cassini, 2 édition, 2008.
- [Pag10] J. Pagès. *Statistique générale pour utilisateurs. Méthodologie*. Presses universitaires de Rennes, 2010.
- [Sch05] J. R. Schott. *Matrix analysis for statistics*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], second edition, 2005.
- [TB97] L. N. Trefethen and D. III Bau. *Numerical linear algebra*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997.