

Enseignant :

— Joseph Salmon : joseph.salmon@umontpellier.fr,

Horaire de consultation (*office hours*) :

— Jeudi 18h-19h, sur rendez-vous par mail uniquement.

Page web du cours :

Supports : <http://josephsalmon.eu/HMMA307.html>

Prérequis

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

Modalité d'évaluation : Scribe diaporamas + Projet final

Attention : les modalités d'évaluation sont susceptibles d'évolution en fonction de la situation sanitaire.

I) Scribe : 30% de la note finale ;

Par groupes de deux ou trois vous serez amenés à "latexiser" un cours d'une heure et demie accompagné d'un code Python (fichier .py uniquement). Le rendu sera en **anglais intégralement** et composé d'un dépôt `git` avec :

- un fichier `.tex` (beamer créant les diaporamas en `.pdf`)
- un fichier `.py` produisant les illustrations insérées dans les diaporamas.
- si besoin, des fichiers (svg, et le pdf correspondant) pour des croquis illustratifs.

La notation sera sur la qualité des fichiers rendus, des illustrations proposées et de la mise en page. Le modèle **obligatoire** du fichier `beamer` à utiliser est disponible ici : <https://github.com/josephsalmon/OrganizationFiles/tree/master/tex/beamer.tex> pour les diaporamas (si besoin cloner tout le dépôt pour assurer une compilation normale du fichier `.tex`).

Critères d'évaluation :

- Qualité de rédaction, d'orthographe, des graphiques (légendes, couleurs) si pertinent,
- Qualité d'écriture du code latex lui-même (label des sections, théorèmes, séparations visuelles avec des pourcentage, etc.),
- Proposition de code en Python permettant d'illustrer les méthodes (pour ce point des précisions seront apportées au cas par cas pour chaque cours),
- Absence de bug.

Le travail sera à envoyer par email au plus tard deux semaines après le cours (**2 pt de pénalité par jour de retard**), avec comme entête **[HMMA307]** (**1 pt de pénalité sinon**), suivant le planning ci-dessous :

Séance CM1 : SAYD YASSINE, SANTINELLI EMMA, DIEVAL MÉGANE

Séance CM2 : WANG RUOYU, DELAGE CINDY, BACAVE HANNA

Séance CM3 : LEFORT TANGUY, COIFFIER OPHELIE, GAIZI IBRAHIM

Séance CM4 : KHALIFI OUMAYMA, KANDOUCI WALID, Abdestar (?)

Séance CM5 : ISKOUNEN SELINA, AMAHJOUR WALID, Rudolf Romisch

Séance CM6 : LEPERCQUE CASSANDRE, HERMAN FANCHON, AMGHAR MOHAMED

Séance CM7 : SAUTON LOIC, LAKEHAL RYMA, YANI BOUAFFAD (?)

II) Projet final : 70% note finale ;

Critères d'évaluation :

- fournir un dépôt `git` contenant tout le travail.
- rapport court, <10 pages, (pdf)
- des diaporamas (`beamer`) pour une présentation de 10mn.

Les sujets et des détails seront fournis début octobre. Le code sera fait en Python. Les projets individuels seront à présenter le 5/11/2020, langue au choix entre français et anglais. Si besoin on pourra utiliser : https://www.statsmodels.org/stable/datasets/statsmodels.datasets.get_rdataset.html pour charger en Python des datasets R. Vous pourrez aussi utiliser toute autre base de donnée que vous jugerez utile pour améliorer votre travail.

Listes des sujets : il s'agit de décrire et de reproduire l'expérience associée à votre nom.

1. SAYD YASSINE

<http://www.biostathandbook.com/ancova.html>

2. SANTINELLI EMMA

<https://jakevdp.github.io/blog/2015/07/23/learning-seattles-work-habits-from-bicycle-counts/> en proposant une Anova en supplément.

3. DIEVAL MÉGANE

<https://towardsdatascience.com/maximum-likelihood-ml-vs-reml-78cf79bef2cf>

4. WANG RUOYU

<https://jbhender.github.io/Stats506/F18/GP/Group16.html> et adapter sur les données de pollution en Occitanie (e.g. https://data-atmo-occitanie.opendata.arcgis.com/datasets/4a648b54876f485e92f22e2ad5a5da32_0)

5. DELAGE CINDY

https://rh8liuqy.github.io/Example_Linear_mixed_model.html avec les outils décrits dans https://rh8liuqy.github.io/Linear_mixed_model_equations.html

6. BACAVE HANNA

<https://arxiv.org/pdf/1308.1334.pdf> en particulier (4.14) et le comparer avec des estimateurs naturels (Lasso, Ridge, ...)

7. LEFORT TANGUY

<https://www.tau.ac.il/~becka/bidual.pdf>

8. COIFFIER OPHELIE

<https://towardsdatascience.com/maximum-likelihood-ml-vs-reml-78cf79bef2cf>

9. GAIZI IBRAHIM <https://besjournals.onlinelibrary.wiley.com/doi/epdf/10.1111/2041-210X.13434>

10. AMGHAR MOHAMED : <http://environmentalcomputing.net/mixed-models-1/> et <http://environmentalcomputing.net/mixed-models-2/>
11. KANDOUCI WALID
<https://towardsdatascience.com/how-linear-mixed-model-works-350950a82911>
12. ABDESSTAR SAHBANE
<https://link.springer.com/article/10.1186/s12859-016-1441-7> data : <https://github.com/abbyyan3/BhGLM>
13. ISKOUNEN SELENA
<https://boostedml.com/2019/09/linear-mixed-models-making-predictions-and-evaluating-accuracy.html> avec
<https://boostedml.com/2018/12/linear-mixed-models-for-longitudinal-data.html>
14. RUDOLF RÖMISCH :
<https://arxiv.org/pdf/1904.06288.pdf>
15. LEPERCQUE CASSANDRE :
https://wiki.qcbs.ca/r_atelier6
16. HERMAN FANCHON :
https://vasishth.github.io/Freq_CogSci/linear-mixed-models.html#model-type-1-varying-intercepts
17. KHALIFI OUMAYMA : analyse faite sur des données simulées à la manière ici <https://stats.idre.ucla.edu/other/mult-pkg/introduction-to-generalized-linear-mixed-models/>
18. LAKEHAL RYMA : https://www.middleprofessor.com/files/applied-biostatistics_bookdown/_book/generalized-linear-models-i-count-data.html avec les données du problème en bas de page <https://datadryad.org/resource/doi:10.5061/dryad.bj001>
19. AMAHJOUR WALID : https://wiki.qcbs.ca/r_workshop7
20. YANI BOUAFFAD : <http://environmentalcomputing.net/mixed-models-3/>
21. SAUTON LOIC : Analyser et synthétiser les données Pollution Occitanie (cf. sujet 2019)

Grille de notation (NOTE sur 20 points) :

- Structuration (rapport) : 1
- Mise en page (rapport) : 1
- Style littéraire (rapport) : 1
- Orthographe (rapport) : 1
- Intérêt scientifique : (rapport) 4
- Présentation visuelle (oral) : 2
- Expression (oral) : 2
- Compréhension général (oral) : 3
- Absence de bogue (code) : 1
- Commentaires claires (code) : 1
- Ergonomie (code) : 3

Thèmes des séances (si le temps le permet)

- Introduction et rappels : moindres carrés, moindres carrés sous contraintes linéaires, KKT

- ANOVA à 1 et 2 facteurs
- Modèles mixtes
- Régression quantile

Bonus

1 pt supplémentaire sur la note finale peut être obtenu pour toute contribution à l'amélioration des cours (notebooks, polycopié, fichiers du git : <https://github.com/josephsalmon/OrganizationFiles> etc.) sous les contraintes suivantes :

Contraintes :

- on obtient 1 point par amélioration
- seule la première amélioration reçue est "rémunérée", les autres ne gagnent plus rien
- il faut déposer un fichier `.txt` (taille <10 ko) en créant une fiche dans la partie du Moodle intitulée "Bonus - Proposition d'amélioration"
- détailler précisément (ligne de code, page des présentations, 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** avec les bonus

Livres et ressources en ligne complémentaires

Cours d'Anova de Lukas Meier (en R) :

<https://stat.ethz.ch/~meier/teaching/anova/index.html>

Linear Mixed Models <https://ourcodingclub.github.io/tutorials/mixed-models/>

Tests usuels : <https://lindeloev.github.io/tests-as-linear/>

Cours d'optimisation (en ligne) de Laurent Lessard :

<https://laurentlessard.com/teaching/524-intro-to-optimization/>

Conseils généraux pour améliorer la visualisation de données, les couleurs et les présentations : <https://blog.datawrapperr.de/beautifulcolors/>

S. Boyd and L. Vandenberghe, "Convex optimization", 2004 ([pdf](#))

S. R. Searle and G. Casella and C. E. McCulloch, "Variance components", 2009

B. R. Clarke "Linear models : the theory and application of analysis of variance", 2008

H. Madsen, Thyregod "Introduction to general and generalized linear models", 2010

J. Pinheiro and D. Bates "Mixed-effects models in S and S-PLUS", 2006

A. Zuur, E.N. Ieno, N. Walker, A.A. Saveliev, G.M. Smith "Mixed Effects Models and Extensions in Ecology with R", 2009

X. A. Harrison and L. Donaldson and M. E. Correa-Cano and J. Evans and D. N. Fisher and C. E. D. Goodwin and B. S. Robinson and D. J. Hodgson and R. Inger "A brief introduction to mixed effects modelling and multi-model inference in ecology", 2018