

INFÉRENCE, ÉVALUATION ET SÉLECTION DE MODÈLES

Cours MA, SA 2024

Prof. Dr Laurent Donzé

Groupe ASAM
Département d'informatique
Bd de Pérolles 90
CH – 1700 Fribourg (Suisse)



@asamunifr

Coordonnées

Prof. Dr Laurent Donzé

Bureau C320, Laurent.Donze@UniFr.ch, +41 26 300 82 75

Heures de réception

Sur rendez-vous ou via [MOD-Moodle](#) ou par [MS-Teams](#) .

Description et objectifs

Le cours fait partie de l'offre de cours Master en statistique appliquée. Celle-ci est constituée de quatre cours semestriels de 4.5 ECTS (1 par semestre), sur deux ans. Les cours forment une unité. Mais ils peuvent être choisis indépendamment les uns des autres. D'intérêt général, l'ensemble des quatre cours donne une bonne vue d'ensemble des problèmes et méthodes actuelles de la statistique appliquée. Les cours sont complétés par un atelier et appuyés par un serveur de calepins Jupyterhub.

Il n'y a pas de public cible privilégié. Bien que d'abord destinés aux étudiant-e-s de la Faculté des SES, les cours sont ouverts à d'autres Facultés ou Universités.

Le cours est une introduction à l'inférence, l'évaluation et la sélection de modèles. Il traitera d'inférence classique pour modèles linéaires et non linéaires, d'évaluation et sélection de modèles, de la technique du bootstrap, de l'algorithme EM. Le cours est à la fois théorique et pratique. Des applications avec des données réelles sont proposées aux étudiant-e-s. Le logiciel utilisé est R.

Informations générales

1. Cours : mardi, 8h.15 – 11h.00, salle PER21 G414 ;
2. L'étudiant-e a la possibilité de télécharger les notes de cours, soit sous forme de diapositives ou soit sous forme de script tout au long du semestre. L'essentiel du cours figure dans ces notes, mais elles n'en fournissent pas le contenu intégral ;
3. Un serveur de calepins Jupyterhub accompagne le cours ;
4. Se référer à [UniFr-timetable](#) pour d'autres renseignements ;
5. Travail de MA : En principe, le thème est imposé. S'inscrire auprès de L. Donzé. Attention aux prescriptions en vigueur : 1) les travaux écrits devront être composés en \LaTeX et adopter le format proposé à cet effet et 2) le logiciel utilisé est R ;
6. Bibliographie : les références seront données au fur et à mesure de l'avancement du cours.

Plates-formes Moodle et MS Teams

1. Les plates-formes du cours [MOD-Moodle](#) et [MS-Teams](#) sont en partie accessibles en accès invité ;
2. Les personnes inscrites au cours sur MyUniFr pourront accéder au contenu intégral des plates-formes, recevoir toutes les informations et participer au cours et à l'examen ;
3. Les étudiant-e-s qui désirent suivre le cours sans s'inscrire sur MyUniFr feront une demande motivée auprès du responsable du cours ;
4. Des formes particulières d'activité (devoirs, tests, lecture d'articles, etc.) seront proposées.

Exemples et exercices

1. Des exemples fournis dans des calepins Jupyter compléteront dans la mesure du possible la matière enseignée ;
2. La réalisation du projet devra permettre de traiter des cas concrets d'analyse de données à l'aide du logiciel R.

Projet

1. Les étudiant-e-s par groupe de deux ou trois effectuent un projet durant le semestre ;

2. Les étudiant-e-s devront constituer, selon un format prévu à cet effet, un calepin Jupyter ;
3. Les thèmes du projet et les modalités de réalisation sont définis en début du semestre ;
4. La réussite du projet est nécessaire pour s'inscrire à l'examen.

Prérequis, difficultés et exigences du cours

1. Les cours de base de statistique et de mathématique (Bachelor 1^{re} et 2^e année d'étude) ;
2. Une connaissance de base du logiciel R est recommandée.
3. Il s'agit d'un cours de **statistique**. En général, on tâchera dans la mesure du possible de démontrer tous les résultats exposés ;
4. Dans l'ensemble, on ne doit pas s'attendre à un cours facile. Mais il sera de toute façon adapté aux connaissances des étudiant-e-s.

Examens

1. Durée : 90 minutes ;
2. L'examen est de type QCM (Questions à choix multiples) ;
3. Mélange de questions théoriques et empiriques ;
4. L'examen se fait à « livres ouverts ». Mais l'usage de téléphones, de montres ou lunettes connectées ou autres tablettes est strictement interdit ;
5. La note finale résulte de la moyenne pondérée de la note d'examen (70%) et de celle du projet (30%).

Plan provisoire du cours

1. Inférence classique pour modèles linéaires et non linéaires
 - Tests de Wald
 - Tests basés sur la vraisemblance
2. Évaluation et sélection de modèles
 - Biais, variance et complexité
 - Critères de sélection
 - Validation croisée

3. Les méthodes de rééchantillonnage
 - Introduction au bootstrap
 - Tests et intervalles de confiance
 - Le modèle de régression linéaire
4. Algorithme EM

Références

- BURNHAM, Kenneth P. et David R. ANDERSON (2002). *Model Selection and Multimodel Inference. A Practical Information-Theoretic Approach*. Springer New York, NY. DOI : <https://doi.org/10.1007/b97636>.
- CAMERON, A. C. et P. K. TRIVEDI (2005). *Microeconometrics. Methods and Applications*. Cambridge University Press, p. 351. ISBN : 0-521-84805-9.
- DAVIDSON, Russell et James G. MACKINNON (1993). *Estimation and Inference in Econometrics*. Oxford University Press. 874 p. ISBN : 0-19-506011-3. DOI : [10.2307/2554780](https://doi.org/10.2307/2554780).
- DAVISON, A. C. (2003). *Statistical Models*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press. ISBN : 0-521-77339-3.
- DAVISON, A. C. et D. V. HINKLEY (1997). *Bootstrap Methods and their Application*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press. ISBN : 0-521-57471-4.
- EFRON, Bradley et Robert J. TIBSHIRANI (1993). *An Introduction to the Bootstrap*. 57 t. Monographs on Statistics and Applied Probability. Chapman et Hall/CRC. ISBN : 9780412042317.
- FLORENS, Jean-Pierre, Vêlayoudom MARIMOUTOU et Anne PÉGUIN-FEISSOLLE (2007). *Econometric Modeling and Inference*. Trad. par Josef PERKTOLD et Marine CARRASCO. Avec une préf. de James J. HECKMAN. Themes in modern econometrics. Cambridge University Press. ISBN : 0-521-70006-X.
- FREEDMAN, David A. (2005). *Statistical Models: Theory and Practice*. Cambridge University Press. ISBN : 978-0-521-85483-2.
- HASTIE, Trevor, Robert TIBSHIRANI et Jerome FRIEDMAN (2009). *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. Second Edition. Springer Series in Statistics. Springer. ISBN : 978-0-387-84857-0. DOI : [10.1007/b94608](https://doi.org/10.1007/b94608).
- OWEN, Art B. (2001). *Empirical likelihood*. DOI : [10.1002/0471667196.ess0629.pub2](https://doi.org/10.1002/0471667196.ess0629.pub2).