

## THÈMES CHOISIS DE STATISTIQUE MULTIVARIÉE

Cours MA, SA 2023

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](mailto:Laurent.Donze@UniFr.ch), +41 26 300 82 75

**Heures de réception**

Sur rendez-vous ou via [MULT-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 étudiants de la Faculté des SES, les cours sont ouverts aux étudiant-e-s d'autres Facultés ou Universités.

Le cours se concentrera sur certains sujets classiques de statistique multivariée. En particulier, le cours traitera de méthodes spécifiques de régression, d'analyse en composantes principales, d'analyse factorielle, de modèles d'équations structurelles. Le cours sera à la fois théorique et pratique. Des applications avec des données réelles seront proposées aux étudiant-e-s. Le logiciel utilisé sera R.

## Informations générales

1. Cours : mardi, 8h.15 – 11h.00, salle PER21 A403 ;
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  $\text{\LaTeX}$  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 [MULT-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<sup>re</sup> et 2<sup>e</sup> année d'étude) ;
2. Les cours d'économétrie sont vivement recommandés ;
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. Méthodes spécifiques de régression
  - Types de variables, terminologie, notation
  - Le modèle linéaire de régression et les moindres carrés
  - Les méthodes de rétrécissement
  - La régression en composantes principales et la régression par moindres carrés partiels
2. Analyse en composantes principales
  - Décomposition d'une matrice de données en facteurs

- Analyse en composantes principales
- Facteurs principaux et composantes principales

### 3. Analyse factorielle

- Introduction
- Le modèle orthogonal de l'analyse factorielle
- L'estimation du modèle de l'analyse factorielle
- Variance totale et ajustement du modèle
- Détermination du nombre de facteurs
- La rotation des facteurs
- Estimation des scores des facteurs

### 4. Modélisation d'équations structurelles

- Introduction
- Modélisation PLS-SEM
- Modélisation CB-SEM

## Références

- HAIR, Joseph F. et al. (2014). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*.
- HÄRDLE, W. K. et L. SIMAR (2015). *Applied multivariate statistical analysis*. Springer. DOI : [10.1007/978-3-662-05802-2](https://doi.org/10.1007/978-3-662-05802-2).
- HASTIE, Trevor, Robert TIBSHIRANI et Jerome FRIEDMAN (2009). *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. Springer.
- IZENMAN, Alan Julian (2008). *Modern multivariate statistical techniques. Regression, Classification, and Manifold Learning*. Springer.
- JOHNSON, Richard A. et Dean W. WICHERN (2007). *Applied Multivariate Statistical Analysis*. Pearson Education, Inc. DOI : [10.2307/2288368](https://doi.org/10.2307/2288368).