

Multivariate and high-dimensional statistics

Titulaire

Thomas VERDEBOUT (Coordonnateur)

Mnémonique du cours

STAT-F600

Crédits ECTS

5 crédits

Langue(s) d'enseignement

Anglais

Période du cours

Premier quadrimestre

Campus

Plaine

Contenu du cours

Lectures are given in English. The description of the course is therefore given in English too.

In the first part of the course we review some basic statistical concepts in point estimation and hypothesis testing. Then we properly define the multivariate Gaussian distribution and present some classical Multivariate Methods of inference for this model including the Hotelling test and the sphericity test. Then we discuss the general linear model and present some classical multivariate methods such as Principal Component Analysis and Discriminant analysis. The last part of the course is dedicated to the extension of these various methods to the high-dimensional situations where the number of observed variables may be as big or bigger than the number of observations.

Objectifs (et/ou acquis d'apprentissages spécifiques)

The main objective of this course is to provide the more recent methodologies in statistics that offer the possibility to deal with "big data". More precisely, we show how inference can be performed when both the number of observations and the number of observed variables are big. We start the course by reviewing some classical Multivariate Methods and show how they can be adapted to deal with such massive data sets. At the end of this UE, the student will be able to determine a correct methodology to deal with multidimensional data and to use R to implement it. She/he will be able to use methods that are adapted to large databases.

Méthodes d'enseignement et activités d'apprentissages

Mainly based on oral presentations with slides. We also propose a group project related to the last part of the course.

Contribution au profil d'enseignement

1) Constituer, développer et entretenir des connaissances dans différents domaines de statistique et d'informatique
2) Résoudre des problèmes en acteur scientifique
3) Concevoir et mettre en oeuvre de manière autonome des projets de recherche scientifique
4) Communiquer dans un langage adapté au contexte et au public
5) Se développer, dans un souci du respect des questions éthiques liées à son domaine d'expertise

Références, bibliographie et lectures recommandées

1) Theory of Multivariate Statistics, Bilodeau and Brenner, 1999, Springer
2) Statistics for High-dimensional data, Buhlmann and van de Geer, 2011, Springer

Support(s) de cours

Podcast

Autres renseignements

Lieu(x) d'enseignement

Plaine

Contact(s)

Thomas Verdebout (thomas.verdebout@ulb.be)

Méthode(s) d'évaluation

Autre, Examen écrit et Projet

Méthode(s) d'évaluation (complément)

A research work provided by groups of students on recent developments in high-dimensional statistics and a written exam in January.

Construction de la note (en ce compris, la pondération des notes partielles)

30% Group works 70% Written exam

Langue(s) d'évaluation principale(s)

Anglais

Programmes

Programmes proposant ce cours à la faculté des Sciences

MA-STAT | Master en statistique, orientation générale | finalité Approfondie, bloc 1 et MS-BGDA | Master de spécialisation en science des données, Big data | bloc U

Programmes proposant ce cours à la Solvay Brussels School of Economics and Management

MS-BGDA | Master de spécialisation en science des données, Big data | bloc U

Programmes proposant ce cours à l'école polytechnique de Bruxelles

MS-BGDA | Master de spécialisation en science des données, Big data | bloc U

