Statistical analysis and document mining

Niveau d'étude
Bac +4

Composante
UFR IM2AG (informatique, mathématiques et mathématiques appliquées)

Période de l'année
Printemps (janv. à avril/mai)

> **Langue(s) d'enseignement:** Anglais

> **Méthodes d'enseignement:** En présence

> **Forme d'enseignement :** Cours magistral

> **Ouvert aux étudiants en échange:** Non

> **Code d'export Apogée:** GBX8AM14

Présentation

Description

The aim of this course is to present the statistical approaches for analysing multivariate data. The information age has resulted in masses of multivariate data in many different field: finance, marketing, economy, biology, environmental sciences,…The theoretical and practical aspects of multivariate data analysis are given equal importance. This balance is achieved through practicals involving actual data analysis using the R software.

Content :

- Multiple linear regression. Least squares, Gaussian linear model, test of linear hypotheses, one-way analysis of variance.
- Principal Components Analysis (PCA).
- Classification, linear discriminant analysis, perceptron, Naive Bayes
- Text mining, numeric representation of texts, connexion with graph clustering.

Heures d'enseignement

CM		CM	15h
TD		TD	15h
TP		TP	24h

Pré-requis recommandés

Elementary notions in probability theory (probability distribution, joint probability density function for random vectors, conditional distribution, expectation, variance, covariance, Gaussian distribution)
Elementary notions in mathematical statistics (estimator, confidence interval, statistical tests).
As a bonus: simple linear regression, linear algebra (matrix reductions), elementary notions in Rstudio and the R software.

Période : Semestre 8

Infos pratiques

Contacts

Responsable pédagogique
Pedro Luiz COELHO RODRIGUES
✉ pedro-luiz.coelho-rodrigues@univ-grenoble-alpes.fr

Lieu(x) ville

> Grenoble

Campus

> Grenoble - Domaine universitaire