STA201 - Analyse multivariée approfondie

Mis à jour le 17-02-2025

Code : STA201

Unité d'enseignement de type cours

9 crédits

Volume horaire de référence (+/- 10%) : **70 heures**

Responsabilité nationale :
EPN06 - Mathématique et statistique / N'deye NIANG

Contact national :

EPN06 Mathématiques et statistiques

2 rue conté
Accès 35 3 ème étage porte 19
75003 Paris

Anne - Solenne Marroulle
anne-solenne.marroulle@lecnam.net

Présentation

Prérequis

Etre inscrit en M2 du master Statistique MR123 (ancien MR085) ou du master Actuariat MR126000A (ancien MR088) ou être agréé (niveau requis **STA101 et STA102**) .
Pour obtenir l'agrément, les auditeurs adresseront par courrier électronique à l'enseignant responsable, Mme Niang (ndeye.niang_keita@cnam.fr) un CV détaillé et une lettre de motivation indiquant les raisons de la demande et le projet pédagogique dans lequel elle s'inscrit.

Objectifs pédagogiques

Approfondir les méthodes statistiques d'analyse exploratoire, de régression et de classification

Programme

Contenu

Méthodes d'analyse exploratoire et confirmatoire :

- Analyse factorielle exploratoire et confirmatoire
- Codage optimal
- Positionnement Multidimensionnel
- Classification non supervisée et modèles de mélanges

Méthodes prédictives :

Méthodes de régression

- Estimation des paramètres par ré-échantillonnage (Bootstrap, Jackknife)
- Multicolinéarité et stabilité des estimations
- Complexité du modèle, trade-off biais-variance et précision des prédictions
- Méthodes de sélection des variables (Best subset regression, méthodes pas à pas)
- Méthodes de régularisation par composantes : Régression sur Composantes Principales, Régression PLS
- Méthodes de régularisation par contraintes : Régression Ridge, LASSO, Elastic Net
- Sélection de modèle par procédures d'apprentissage statistique
- Régression robuste
- Régression non paramétrique

Méthodes de classification supervisée

- Régression logistique binaire
- Régression Logistique multinomiale et ordinale
- Analyse factorielle discriminante
- Discrimination sur variables qualitatives
- Analyse discriminante probabiliste Linéaire et Quadratique
- Approche non paramétrique : Méthode des noyaux, Méthode des k plus proches voisins

Traitement des données manquantes

Modalités de validation

- Projet(s)

Description des modalités de validation

Le travail consistera à rédiger un rapport qui sera présenté et discuté lors d'une soutenance orale. L'étudiant pourra choisir entre deux types de projets : une étude de cas ou une synthèse d'articles.

Bibliographie

Titre	Auteur(s)
Analyse des données (Hermes,2003)	G.GOVAERT
Analyse discriminante (Dunod,2001)	M.BARDOS
Statistique explicative appliquée (Technip, 2003)	J.P.NAKACHE, J.CONFAIS
Probabilités, analyse des données, statistique 3ème édition (Technip,2010)	G.SAPORTA
Data mining et statistique décisionnelle, 3ème édition (Technip, 2010)	S.TUFFERY
The Elements of Statistical Learning (Springer 2009). Téléchargeable ici : http://www-stat.stanford.edu/~tibs/ElemStatLearn/download.html	T.HASTIE, J.FRIEDMAN, F.TIBSHIRANI
Multivariate Analysis, Academic Press, 1979	K.V. Mardia, J.T. Kent and J.M. Bibby
An Introduction to Statistical Learning with application in R, 2ème édition (Springer, 2021))	G. James, D. Witten, T. Hastie and R. Tibshirani
An Introduction to Statistical Learning with application in Python (Springer ,2023)	G. James, D. Witten, T. Hastie, R. Tibshirani and J.Taylor