STA101 - Analyse des données : méthodes descriptives

Mis à jour le 25-09-2026

Code : STA101

Unité d'enseignement de type cours

6 crédits

Volume horaire de référence (+/- 10%) : **50 heures**

Responsabilité nationale :
EPN06 - Mathématique et statistique / Vincent AUDIGIER

Contact national :

EPN06 Mathématiques et statistiques

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

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

Présentation

Prérequis

Cette formation s'adresse à toute personne souhaitant :
- analyser des données statistiques multidimensionnelles (chargé d'études, ingénieur, chercheur, technicien, etc)
- se spécialiser par la suite dans le traitement de données massives.
Elle nécessite des connaissances de base en statistique descriptive (univariée et bivariée) et des notions de calcul matriciel.

Objectifs pédagogiques

Explorer, décrire et interpréter des données dans leur aspect multidimensionnel. Le cours s'appuiera sur la pratique du logiciel R.

Compétences

Etre en mesure à l'issue de l'enseignement de produire des études statistiques (rédaction d'un rapport) mettant en jeu les techniques d'analyse factorielle et de classification.

Programme

Contenu

Traitements préalables à une analyse factorielle
Analyses bivariées (analyses univariées supposées maîtrisées)

Outils mathématiques de l'analyse des données
Rappels d'algèbre linéaire
Notion de métrique, projecteur

L'analyse en composantes principales
Principe de l'ACP
Calcul des facteurs principaux et des composantes principales
Mesure de qualité des résultats, techniques d'interprétation
Utilisation de variables illustratives
Gestion des données manquantes

L'analyse des correspondances simples
Principe et métrique associée
Représentation des profils-lignes et des profils-colonnes, représentation simultanée
Règle d'interprétation des résultats

L'analyse des correspondances multiples
Principes de mise en oeuvre et interprétation
Application au dépouillement d'enquêtes

L'Analyse factorielle des données mixtes
Equilibre entre variables de natures différentes
Principe de mise en oeuvre et interprétation

Les méthodes de classification automatique
Méthodes non hiérarchiques : centres mobiles, nuées dynamiques
Méthodes hiérarchiques : méthode de Ward, construction et lecture du dendrogramme
Aspects pratiques de la classification : méthodes mixtes
Interprétation d'une partition à l'aide des variables initiales, en liaison avec une analyse factorielle

Modalités de validation

- Examen final

Description des modalités de validation

Le contrôle des connaissances est basé uniquement sur un devoir sur table

Bibliographie

Titre	Auteur(s)
Probabilités, analyse des données et statistique. 3 ème édition (Technip, 2011)	SAPORTA G.
Statistique. Méthodes pour décrire, expliquer et prévoir (Dunod, 2010)	TENENHAUS M.
Statistique exploratoire multidimensionnelle (Dunod, 4ème édition, 2006)	LEBART L., MORINEAU A., PIRON M.
Analyses factorielles simples (Economica, 1995)	BRY X.
L'analyse des données (Que-Sais-Je ? PUF, 9ème édition, 2006)	BOUROCHE J.M., G. SAPORTA
Approche pragmatique de la classification. Technip 2005	NAKACHE J.-P., CONFAIS J.
Statistiques avec R. Presses universitaires de Rennes 2012	Pierre-André Cornillon, Arnaud Guyader, Julie Josse ....
Analyse de données avec R. Presses universitaires de Rennes 2009	François Husson, Sébastien Lê et Jérôme Pagès