

# STA101 - Analyse des données : méthodes descriptives

## 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 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 SPAD.

### 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

Étude des variables : corrélation linéaire, corrélation des rangs

Représentation des individus : diagramme de dispersion, diagramme sous forme d'icônes:

étoiles, rayons de soleil, profils

#### 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

#### 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

#### L'analyse des correspondances simples

Tableau de contingence : test de l'indépendance de deux variables qualitatives, khi-deux

L'analyse des correspondances simples

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

Classification des individus à l'issue d'une ACM.

*Cet enseignement comprend une formation au logiciel statistique SPAD.*

*Logiciels recommandés outre SPAD : STATGRAPHICS, UNIWIN, XLSTAT,R*

## Description des modalités de validation

🌟 Valide le 23-02-2019

**Code : STA101**

6 crédits

**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

Sabine Glodkowski

[formation.maths@cnam.fr](mailto:formation.maths@cnam.fr)

Le contrôle des connaissances est basé sur une étude de cas et la rédaction d'un projet réalisés avec le logiciel SPAD.

## Bibliographie

<b>Titre</b>	<b>Auteur(s)</b>
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