

## Analyse des données

Intitulé du cours	Analyse des données
Niveau /semestre	L3 / S1
Composante	École d'Économie de Toulouse
Enseignant responsable	Sandrine CASANOVA
Autre enseignant	Eve LECONTE
Volume Horaire CM	25,5
Volume horaire TD	
Volume horaire TP	15
Langue du cours	Français
Langue des TD et/ou TP	Français

### Coordonnées de l'équipe pédagogique :

Sandrine Casanova : [sandrine.casanova@tse-fr.eu](mailto:sandrine.casanova@tse-fr.eu), bureau T 215

Eve Leconte : [eve.leconte@tse-fr.eu](mailto:eve.leconte@tse-fr.eu), bureau T 218

Les étudiants peuvent poser des questions à la fin du cours, par mail ou prendre rendez-vous par mail pour rencontrer les enseignantes.

### Objectifs du cours :

Ce cours d'option de statistique appliquée comporte deux parties :

#### **- Analyse des données (statistique exploratoire) avec Sandrine Casanova**

A l'issue de cette partie, l'étudiant connaîtra trois méthodes d'analyse des données qui permettent de résumer l'information contenue dans de grands tableaux de données et saura appliquer ces méthodes sur des données réelles à l'aide du logiciel R. Plus largement, l'étudiant aura acquis un savoir-faire dans l'exploration de données à l'aide de méthodes statistiques multidimensionnelles.

#### **- Tests d'hypothèses (statistique inférentielle) avec Eve Leconte**

Cette partie du cours reprend et approfondit le chapitre sur les tests d'hypothèses du cours de Statistique Inférentielle de la L2 Economie-Gestion et initie les étudiants à la pratique des tests avec le logiciel R. A l'issue de cette partie, l'étudiant devra pouvoir formuler des hypothèses statistiques pour répondre à un problème donné et savoir choisir le test statistique adéquat pour trancher entre deux hypothèses en fonction de la nature des variables et du problème posé. Seront notamment étudiés les tests permettant d'étudier la liaison entre deux variables. L'étudiant saura mettre en œuvre ces tests sur des jeux de données réelles à l'aide du logiciel R.

### Prérequis :

Statistique descriptive, variables aléatoires discrètes et continues usuelles (loi, espérance, variance), convergence en loi, estimateurs empiriques de l'espérance et de la variance.

### **Modalités pratiques de gestion du cours :**

Les ordinateurs et tablettes sont acceptés en cours et en TP.

Nous attendons des étudiants une participation active en cours et en TP.

De plus, la plupart des exercices de TD seront donnés sur Moodle et seront à faire en autonomie en dehors des cours. Des corrigés seront ensuite mis à disposition sur Moodle.

L'acceptation des arrivées tardives est laissée à l'appréciation des enseignantes.

### **Modalités d'évaluation :**

Le contrôle continu compte pour 50 % et l'examen final pour 50 %.

Le contrôle continu se compose pour 80 % de trois projets :

- pour la partie Analyse des données : deux projets avec le logiciel R au cours du semestre (le premier donné début novembre et le deuxième donné début décembre).
- pour la partie Tests d'hypothèses : un projet avec le logiciel R donné début décembre à rendre après les vacances de Noël.

Les 20 % restants du contrôle continu correspondent à des évaluations en ligne sous forme de QCM.

L'examen terminal est un examen écrit sans documents sur l'ensemble du cours et des TD.

### **Bibliographie/références :**

#### **Pour les prérequis :**

« Statistique pour économistes et gestionnaires ». Brigitte Tribout, Pearson Education

#### **Pour l'apprentissage du logiciel R :**

R peut être téléchargé gratuitement sur le site du CRAN : <https://cran.r-project.org/>

Plusieurs manuels en anglais y sont accessibles.

Un manuel en ligne en français :

[https://cran.r-project.org/doc/contrib/Paradis-rdebuts\\_fr.pdf](https://cran.r-project.org/doc/contrib/Paradis-rdebuts_fr.pdf)

#### **Pour le cours (partie Analyse des données) :**

« Statistique : méthodes pour décrire, expliquer et prévoir », Michel Tenenhaus. Dunod

« Analyse de données avec R », François Husson, Sébastien Lê et Jérôme Pagès, PUR.

« Statistiques avec R », Pierre-André Cornillon et autres, PUR.

#### **Pour le cours (partie Tests d'hypothèses) :**

« Probabilités, analyse des données et statistique ». Gilbert Saporta, Technip.

### **Planification des séances :**

Chaque semaine, il y aura un cours sur la partie Analyse des Données (le mardi matin) et un cours sur la partie Tests d'hypothèses (le vendredi matin).

Les TP avec R auront lieu le vendredi matin après le cours du vendredi en alternant les deux parties.

### **Plan du cours :**

#### **Partie Analyse des données**

- Introduction à l'analyse des données
- Analyse en composantes principales (ACP)
- Analyse factorielle des correspondances (AFC)
- Classification non supervisée

#### **Partie Tests d'hypothèses**

- Rappels sur la théorie des tests d'hypothèses
- Tests d'ajustement du khi-deux de Pearson et de Kolmogorov-Smirnov
- Test du khi-deux d'indépendance
- Tests pour un échantillon (paramétriques et non paramétriques)
- Tests pour deux échantillons (comparaisons de variances et de moyennes) paramétriques et non paramétriques
- Test de comparaison de plusieurs moyennes (analyse de variance) et de plusieurs médianes (test de Kruskal-Wallis)

### **Enseignement à distance :**

Si la situation sanitaire l'exige, les enseignements pourront avoir lieu à distance sous forme de classe en ligne interactive, QCM et exercices en ligne, TP à distance, forum sous Moodle.