

## Projet de Statistique

F. Saïd

---

### 1. Objectif

Le but de ce projet est de réaliser une étude statistique approfondie sur un jeu de données réel à l'aide de Python, en appliquant les techniques apprises dans le cours d'Analyse Statistique des Données. Ce projet vise à renforcer vos compétences en analyse de données et en utilisation d'outils scientifiques pour résoudre des problématiques spécifiques.

### 2. Instructions

- Sélectionnez une base de données parmi celles proposées sur le [UCI Machine Learning Repository](#). Justifiez votre choix en expliquant la pertinence des données pour des questions de recherche significatives.
- Formulez des hypothèses claires et définissez des questions de recherche spécifiques à aborder avec un seuil de signification de 5%.

### 3. Procédures analytiques

Votre étude doit inclure :

1. Analyse descriptive de variables catégorielles et quantitatives, création de graphiques et calcul d'indicateurs statistiques. Conversion de variables quantitatives en catégorielles si nécessaire.
2. Tests de normalité, d'égalité de 2 variances, d'égalité de 2 moyennes, ANOVA, tests d'association de deux variables catégorielles, régressions linéaires simples et multiples, et analyse en composantes principales.
3. Optionnel : classification non supervisée des données.

### 4. Exigences de restitution

- Les problématiques doivent être clairement formulées. Justifiez les méthodes statistiques utilisées et interprétez les résultats dans leur contexte, en considérant les limites et implications de votre étude.
- Présentez vos résultats sous forme d'un diaporama sonore (maximum 10 pages) et d'un notebook Jupyter, à soumettre via Moodle avant la date limite. Les noms des membres du binôme doivent être indiqués sur tous les documents et fichiers.

## 5. Grille d'évaluation

Critère	Points Possibles	Détails
Sélection et justification de la base de données	10	Pertinence et applicabilité des données.
Clarté des hypothèses et questions de recherche	10	Précision et pertinence.
Complétude et rigueur des analyses statistiques	30	Couverture et exactitude des analyses, vérification des conditions de validité des procédures utilisées.
Interprétation des résultats	30	Justesse, profondeur d'analyse, discussion des implications.
Présentation et restitution des résultats	10	Qualité du diaporama et du notebook, clarté de communication.
Respect des délais et des consignes de format	10	Ponctualité, conformité aux exigences de format et de présentation.