

# Analyse des données Licence Pro 2025-2026

## TD n°1- Les données en statistique

Florian Bayer

Pour l'ensemble de ce cours, les TD porteront sur un même jeu de données détaillant le niveau de vie dans les pays du Monde

Vous apprendrez à manipuler :

- un logiciel d'analyse de données Open Source : Orange3
- un logiciel de cartographie: Magrit

Les TD reprendrons le contenu des cours :

1. Les données statistiques
2. L'analyse univariée
3. La discrétisation
4. La corrélation

Les objectifs de ce TD sont de mettre en application les acquis du cours 1

- caractérisation du type de données
- et des tableaux

Nous verrons également quelques fonctionnalités Excel.

Organisation : Vous travaillerez en binômes tout au long de ce TD.

Ouvrez et observez le contenu du fichier Excel hdi-edu.xlsx: il contient un onglet Data et un onglet Meta.

En informatique, on évite de donner des noms trop longs aux variables. Les métadonnées permettent de mieux comprendre les informations relatives à chaque variable.

Parcourez la liste des données dans l'onglet META afin de vous approprier les thématiques présentes.

1. Dans l'onglet DATA, de quel type de tableau s'agit-il ?
2. De quel type sont les variables suivantes (détaillez le plus précisément possible en justifiant vos choix) :

- code
- transition\_demo
- esperance\_vie

## Étape 1 - Tri et observation (10 min) :

- Triez les pays par espérance de vie (ordre décroissant)
- Identifiez les 5 pays avec la plus longue espérance de vie
- Identifiez les 5 pays avec la plus courte espérance de vie
- Filtrez les données pour ne voir que les pays d'Afrique

## Étape 2 - Analyse comparative (10 min) :

- Choisissez 2 pays que vous connaissez et comparez leurs indicateurs principaux
- Notez 3 différences marquantes entre ces deux pays

Objectif : Visualiser la distribution de l'espérance de vie mondiale

Consignes :

- Sélectionnez la colonne `esperance_vie` (tous les pays)
- Créez un histogramme pour observer la distribution des valeurs
- Analysez la forme de la distribution : est-elle symétrique ? Y a-t-il des pics ?
- Interprétez : que révèle cette distribution sur les inégalités mondiales de santé ?

Aide : Insertion > Graphiques > Histogramme

Objectif : Calculer et comparer les statistiques par sous-régions du monde

Consignes :

- Calculez la moyenne, le minimum et le maximum d'espérance de vie par sous-région
- Créez un diagramme en bâtons (graphique en barres) des moyennes par sous-région
- Identifiez quelle sous-région a les plus grandes disparités internes

Aide pour les calculs :

- Étape 1 : Créez un tableau croisé dynamique (Insertion > Tableau croisé dynamique)
  - Lignes : sous\_region
  - Valeurs : esperance\_vie (configurez pour afficher moyenne, min, max)
- Étape 2 : Copiez les résultats du tableau croisé dynamique
  - Sélectionnez les moyennes par sous-région et copiez les valeurs dans une nouvelle feuille de votre classeur
- Étape 3 : Créez un diagramme en bâtons à partir de ces valeurs copiées
  - Sélectionnez vos données copiées > Insertion > Graphiques > Histogramme/Barres

Restitution rapide (5 min) :

Chaque binôme partage :

- Un constat marquant issu de vos analyses
- Une hypothèse pour expliquer une différence observée entre pays ou régions
- Une limite identifiée dans ce jeu de données

Préparation aux TD suivants :

- Ces premières observations vous serviront de base pour l'analyse statistique approfondie (TD2)
- La variable espérance de vie sera réutilisée pour les exercices de cartographie (TD3)