

## Magistère d'Économiste Statisticien 1ère année

---

### Analyse de données

*Projet 1 à réaliser en binôme et à déposer sous Moodle au plus tard le mardi 4 novembre 2025 à 23h55*

**Recommandations :** Votre projet doit être sous la forme d'un rapport rédigé (**20 pages maximum, annexes comprises**). **Tous les tableaux et graphiques présentés doivent être commentés (même succinctement)**. Vous devez mettre le code R à la fin du rapport (seule annexe qui ne compte pas dans le nombre de pages). Votre fichier doit être exclusivement au format pdf et vous devez indiquer dans le nom du fichier les noms des deux auteurs. Un seul des deux auteurs doit déposer le rapport sur Moodle.

#### Sujet du projet

Vous disposez d'un fichier contenant des caractéristiques de 50 pays. Les variables quantitatives considérées sont les suivantes :

- **pays** : nom du pays,
- **mortinfant** : taux de mortalité des enfants de moins de 5 ans (pour 1000 naissances),
- **exportation** : exportations de biens et services (en % du PIB total),
- **sante** : dépenses en santé (en % du PIB total),
- **importation** : importations de biens et services (en % du PIB total),
- **revenu** : revenu disponible moyen annuel des ménages (en standard de pouvoir d'achat),
- **inflation** : taux d'inflation (en %),
- **espvie** : espérance de vie (en années),
- **fertilite** : nombre moyen d'enfants par femme,
- **pihb** : PIB par habitant (en standard de pouvoir d'achat).

#### Plan du rapport

1. Introduction : présentation des données, des objectifs et du plan du rapport.
2. Statistiques univariées et bivariées : présenter succinctement chaque variable (statistiques descriptives classiques, histogramme et boîtes à moustaches pour chaque variable) et **commenter vos sorties en « français »**. Donner la matrice des corrélations et la commenter.
3. ACP
  - (a) Justifier l'emploi de l'ACP sur ces données,
  - (b) Choisir le nombre de composantes principales en arbitrant entre les différents critères vus en cours,
  - (c) Interpréter les composantes principales retenues,
  - (d) Donner, en justifiant, les pays bien représentés sur les axes retenus,

- (e) Interpréter les graphiques des pays sur les composantes principales retenues. Pour simplifier, ne commenter que deux pays bien représentés "opposés" par axe retenu.
  - (f) Quels sont les pays ayant fortement contribué à la première composante principale ?
4. Conclusion : reprendre les principaux résultats et donner un avis sur la capacité de l'ACP à bien résumer l'information contenue dans le fichier de données.
  5. Mettre en annexe le code R.