

MASTER 1
« Économie du Développement » – « Intelligence Économique »
UNIVERSITÉ DE BORDEAUX
2017/2018

GESTION ET ANALYSE DES DONNÉES
SÉANCES 3 ET 4

Il est vivement conseillé de réaliser un do-file et d'enregistrer les commandes utilisées lors des séances de TD. Vous pouvez également enregistrer les sorties STATA® sur un fichier texte.

INTRODUCTION À STATA® : MANIPULER DES DONNÉES, GESTION DES VARIABLES ET SÉLECTIONS COMPLEXES

1. Ouvrir **STATA®**, importer la base de données **Macro2010.dta** et la sauvegarder sous **Macro2010_seance3.dta**.
2. Décrire les quatre éléments de l'interface. Afficher la fenêtre data Browser et celle du do-file (script). Selon vous, quel est l'intérêt de ces deux fenêtres pour l'utilisateur ?
3. Obtenir un tableau permettant de décrire et d'identifier les dix premières variables de la base de données.
4. Présenter un tableau des statistiques descriptives élémentaires sur les variables sélectionnées à la question précédente. Caractériser les différents types de variables.
5. Construire un indicateur composite de « *politiques environnementales* » (*Env_policies*) qui synthétisera les différents dispositifs législatifs en matière de lutte contre le réchauffement climatique (diminution des émissions des gaz à effet de serre (Co2, méthane), diminution de l'intensité énergétique, réduction de la consommation en énergies fossiles, développement transports collectifs à énergie renouvelables, augmentation de la part des énergies renouvelables dans la consommation énergétique des ménages). Que pouvez-vous dire de cet indice ? Quelle information nous donne-t-il ?
6. Caractériser la variable *HDI*. Construire une seule catégorie de développement humain élevé (rassemblant les modalités high et very high) pour ne conserver que trois modalités (high, medium et low). En quoi cette procédure est-elle importante dans l'analyse des données ?
7. Renommer les variables *VAR00345* et *VAR00346* avec des noms clairs et précis correspondant aux labels.
8. Construire et caractériser les variables suivantes :
 - a. *GDPcap* : PIB par habitant en 2010
 - b. *Density* : nombre d'habitants par km² en 2010
9. Éditer les labels des variables *GDPcap* et *Density*.
10. Créer et labéliser une variable *CATWB* classant les pays selon la typologie de la Banque Mondiale :
 - a. "Low income" lorsque le PIB par habitant est strictement inférieur à 1000 USD
 - b. "Lower middle income" lorsque le PIB par habitant est compris entre 1000 et 3 999 USD
 - c. "Upper middle income" lorsque le PIB par habitant est compris entre 4000 et 12 399 USD
 - d. "High income" lorsque le PIB par habitant est supérieur à 12 400 USD
11. Exclure les variables *Physicians*, *Forest*, *Security*.
12. Conserver uniquement dans la base de données les pays ayant au moins 450 000 habitants.
13. Sauvegarder le do-file et la base de données (**Macro2010_seance3.dta**).

ANALYSE DES DISTRIBUTIONS QUALITATIVES, QUANTITATIVES ET GRAPHIQUES DÉDIÉS

1. Ouvrir **STATA**[®], importer la base de données **Macro2010_seance3.dta** et la sauvegarder sous **Macro2010_seance4.dta**.
2. Quelle est la proportion de pays au sein de chaque région présentant un taux croissance du PIB élevé (supérieur à 5%) en 2010 ? Faire apparaître ces proportions dans un diagramme en secteurs circulaires pour chaque région.
3. Combien de pays développés appartiennent à la catégorie « croissance du PIB élevée » ? Quels sont ces pays ?
4. Parmi les différentes régions, quelle est la proportion de pays ayant un IDH élevé, moyen et faible ? Parmi les pays ayant un IDH élevé, moyen et faible quelle proportion représente chaque région ? Présenter vos résultats sous la forme d'un tableau.
5. A l'aide d'un graphique en tuyaux d'orgue, comparer la proportion de pays des différentes classes d'IDH en fonction des catégories de PIB/habitant établies par la Banque Mondiale. Modifier manuellement le graphique pour le rendre clair et lisible.
6. A l'aide du même type de graphique, comparer désormais la proportion de pays au sein des différentes catégories de PIB/habitant établies par la Banque Mondiale en fonction des classes d'IDH.
7. Reprendre les questions 6 et 7 et utiliser des diagrammes en tuyaux d'orgue « stacked ». Commenter.
8. Pour les pays latino-américains et d'Asie du Sud et de l'Est, présenter les statistiques descriptives permettant de comparer les deux « profils » régionaux en matière d'innovation (Patent et Research_exp). Vous ferrez apparaître en particulier les statistiques suivantes et commenterez celles qui sont en gras : la moyenne, la médiane, **le premier décile**, **le troisième quartile**, le **kurtosis**, le **skewness**, la variance, **l'écart-type**, le minimum et le maximum, le nombre de pays.
9. Dans quelle région les émissions de Co2 par habitant sont-elles les plus hétérogènes ? Quel outil faut-il mobiliser pour répondre à cette question ?
10. A l'aide d'un histogramme, analyser la distribution des émissions de Co2 par habitant au sein des pays de l'OCDE.
11. Déterminer graphiquement quelle est la proportion de pays de l'échantillon qui émettent au maximum 10 tonnes de Co2 par habitant en 2010 ? Quelle est la valeur maximale des émissions pour les 60% de pays qui émettent le moins ?
12. Parmi les pays de l'OCDE, donner les trois économies qui émettent le plus de Co2 en tonnes/habitant. Combien de tonnes de Co2 ont été émises en 2010 aux Etats-Unis ?
13. Sauvegarder le do-file et la base de données (**Macro2010_seance4.dta**).

[002]

TEST 2, durée : 1h**L'ensemble des résultats devront impérativement être interprétés.**
