Gestion et analyse de données sur STATA, Univ.Bdx (2016-18)

Pierre Levasseur

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

GESTION ET ANALYSE DES DONNÉES
SÉANCES 5 ET 6

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.

FUSION, AGRÉGATION, TESTS DE DÉPENDANCE ET DE NORMALITÉ

1. Ouvrir STATA®, importer la base de données Macro_merge_seance4.dta et la sauvegarder sous Macro2010_seance5.dta.
2. Créer une nouvelle variable (env_policies_cont) standardisant l'indicateur composite de politiques environnementales (*cf.* séance 1, variable COMPO_INDEX). Quel est l'intérêt de la standardisation ?
3. Créer une nouvelle variable Poor (« discrétisée »), prenant la valeur 0 si le pays fait partie des 25% (exclure) ayant le moins de pauvres dans la population totale, 1 entre 25% (inclure) et 75% (exclure) et enfin 2 au-delà. Renommer ces modalités et indiquer le label de cette variable.
4. Ajouter les individus issus de la base de données OECD2010.dta. Ouvrir la fenêtre data Browser. Que constatez-vous pour les nouvelles observations et la variable Poor ?
5. Effacer la variable Poor (sans l'inscrire dans le do-file) et relancer la commande réalisée à la question 3. Que remarquez-vous désormais pour les individus ajoutés ? Que faut-il modifier ?
6. Ajouter les variables présentes dans la base de données GDP2010.dta. Sauvegarder la nouvelle base de données sous Macro2010_seance5.dta.
7. Existe-t-il une relation de dépendance entre la région d'appartenance et le niveau de développement humain pour les pays n'appartenant pas à l'OCDE ? Est-elle significative (au seuil de 5%) ? Expliquer et commenter.
8. Existe-t-il une relation de dépendance entre le niveau de développement humain et le fait d'appartenir à la catégorie « croissance du PIB élevée »pour les pays d'Amérique latine et d'Asie du Sud et de l'Est qui n'appartiennent pas à l'OCDE ? Est-elle significative (au seuil de 5%) ? Expliquer et commenter.
9. A l'aide de boites à moustaches (boxplot), analyser et comparer graphiquement la distribution des émissions de Co2 par habitant entre les pays latino-américains, africains et d'Asie du Sud et de l'Est.
10. Graphiquement, la distribution des émissions de Co2 par habitant dans les pays de l'OCDE vous semble-t-elle normale ? Même question en retirant les 10% de pays qui émettent le plus. Réaliser les tests de Shapiro-Wilk et de Shapiro-Francia puis commenter la normalité de la distribution dans les deux cas.
11. Programmer une boucle permettant en même temps de présenter les statistiques élémentaires et de tester la normalité des variables suivante : Export, Import, Inf, Cell, Pop_growth.
12. Sauvegarder le do-file et la base de données (Macro2010_seance5.dta).

CROISEMENT DE VARIABLES QUANTITATIVES, TESTS DE COMPARAISON DE MOYENNES ET DE VARIANCES

1. Ouvrir **STATA®**, importer la base de données **Macro2010_seance5.dta** et la sauvegarder sous **Macro2010_seance6.dta**.
2. Représenter et analyser graphiquement la distribution de la variable *Cell*, puis *GDPcap*.
3. Effectuer les mêmes représentations graphiques en retirant les 25% de valeurs extrêmes dans chacune des deux variables (en queue de distribution).
4. Quels types de transformation pouvons-nous effectuer sur la variable *GDPcap* afin de faciliter l'analyse de sa distribution ? Comparer.
5. Représenter graphiquement le nuage de points croisant le logarithme du PIB/hab (Y) et le nombre d'abonnements téléphoniques pour 100 habitants (X). Tracer la droite d'ajustement linéaire. Commenter.
6. Pour chacune des régions, représenter graphiquement le croisement entre la croissance du PIB (Y) et le nombre d'abonnements téléphoniques pour 100 habitants (X). Commenter par rapport au résultat de la question précédente.
7. Interpréter le sens, l'intensité et la significativité de la corrélation entre les variables suivantes : *Cell*, *Elecrural*, *GDPcap*, *Poverty* (au seuil de 1%).
8. Pouvons-nous dire qu'il existe une relation causale entre le nombre d'abonnements téléphoniques et le PIB par habitant ? Expliquer.
9. A l'aide de boîtes à moustaches, comparer la distribution du nombre d'abonnements téléphoniques pour les pays appartenant ou non à l'OCDE.
10. Le nombre moyen d'abonnements téléphoniques est-il significativement différent entre les deux groupes de pays (au seuil de 5%) ?
11. A l'aide de boîtes à moustaches, comparer la distribution du nombre d'abonnements téléphoniques par niveau de développement humain.
12. Les dépenses de santé en proportion du PIB sont-elles significativement différentes d'un groupe à l'autre d'IDH (au seuil de 5%) ? Les variances sont-elles significativement différentes (au même seuil) ?
13. Sauvegarder le do-file et la base de données (**Macro2010_seance6.dta**).

[004]

TEST 3, durée : 1h

L'ensemble des résultats devront impérativement être interprétés.