

**Gestion et analyse de données**

**Contrôle continu 2 – durée : 1h**

---

*Ouvrir le logiciel STATA®, puis importer le fichier urban.dta. L'ensemble des tests statistiques sont à interpréter au seuil de 5%.*

1. Le salaire horaire (*hourlywage*) est-il normalement distribué parmi les femmes ? Répondre en vous aidant du test statistique adapté et d'un graphique.
2. Commenter la dispersion de la variable *hourlywage* en excluant les 5% des valeurs extrêmes en fin de distribution.
3. Est-il possible de transformer la variable *hourlywage* pour que sa distribution se rapproche de celle de la loi normale ? Utiliser l'outil le plus adapté pour effectuer cette transformation.
4. Représenter sur le même graphique la fonction de densité de kernel du salaire horaire normalisé des femmes et celle du salaire horaire normalisé des hommes.
5. Existe-t-il une relation statistique entre la consommation d'alcool et le fait de savoir parler anglais ? Interpréter cette relation.
6. Rajouter les individus issus de la base de données rural.dta. Renommer la base de données urban\_rural.dta. Quel est le nombre total d'observations ?
7. Rajouter les variables issues de la base de données anthropométrie.dta. Renommer la base urban\_rural\_complete.dta.
8. Analyser et commenter la relation entre le salaire horaire et le poids des individus ? Représenter graphiquement cette relation.
9. Le poids moyen des hommes est-il significativement différent de celui des femmes ?