

Séquence 1 : Données numériques sous Excel et Rstudio

Les activités de cette séquence passent par le téléchargement d'une base de données sous excel et par l'utilisation d'un tableur puis d'un logiciel statistique (R avec l'interface de programmation R Studio) pour traiter des variables numériques, ainsi que d'un traitement de texte pour présenter les résultats et commentaires. On en profitera pour présenter les logiciels disponibles et mettre en place une méthode de travail en matière de sauvegarde et de rendus.

1. Téléchargez et ouvrez les données de la séquence 1 (indicateurs de développement humain par pays). Rappelez rapidement les individus statistiques et la population étudiées et caractérisez les principales variables disponibles.
2. Calculez la moyenne et la variance rapportée au nombre total de pays ainsi que l'écart-type correspondant pour les différents indicateurs.
3. Calculez directement la médiane, les quartiles, le premier et dernier décile de ces distributions.
4. Tracez une boîte à moustache pour la distribution des espérance de vie pour l'ensemble des pays.
5. Regroupez les données d'espérance de vie dans un nouvel onglet, tracez le graphique cumulé croissant de l'espérance de vie par pays et indiquez graphiquement l'emplacement de la médiane, des quartiles, du premier et dernier décile.
6. En vous appuyant sur la question précédente, constituez des classes appropriées puis tracez l'histogramme des pays.
7. Calculez l'ensemble des covariances et des corrélations entre variables. À nouveau, vous utiliserez le nombre de pays au dénominateur.
8. Tracez le nuage de points et la droite de régression des espérances de vie sur les PIB par tête.
9. Effectuez une régression linéaire de l'espérance de vie sur l'ensemble des variables.
10. Serait-il possible de faire les traitements précédents pour la population mondiale plutôt que par pays ? Quelles sont les données qui nous font défaut pour les indicateurs de dispersion ? Calculez l'espérance de vie des habitants de la planète et sa variance inter-pays.
11. Faites votre propre régression en choisissant vos variables et si vous le souhaitez une base de données alternatives. Vous pourrez compléter les données en utilisant la base intégrale ou encore vous rendre sur le site <http://hdr.undp.org/en/content/download-data> et télécharger l'intégralité des données du Rapport sur le développement humain (HDR) du Programme des Nations Unies pour le développement 2020.
12. Comparez les différentes étapes avec Excel et R Studio
13. Présentez et commentez l'intégralité de vos résultats.