

Programmation Statistique sous R
EXAMEN
Documents autorisés
Durée : 1H30

Antoine ROLLAND et Anthony SARDELLITTI

21 Février 2023

Table des matières

1	Exercice 1 : Importer les données	2
2	Exercice 2 : Statistiques descriptives	2
3	Exercice 3 : Tris et filtres	2
4	Exercice 4 : Modifier le dataframe	3

CONSIGNES

Vous avez 1h30 pour réaliser l'ensemble du devoir. Vous devez vous connecter sur les sessions EXAM. Il est conseillé de copier-coller le fichier .csv sur votre Bureau. Le cours en ligne a été téléchargé et est disponible sous format HTML. Si lors de la navigation quelques liens ne fonctionnent pas, il suffit d'ouvrir directement le fichier HTML souhaité.

Voici les recommandations pour que votre travail soit pris en compte et pour ne pas avoir de pénalités :

- Le script R est à rendre sur la session EXAM
- Votre fichier doit se nommer **NOM_PRENOM** et doit être un script R
- Le rendu est individuel
- Pour chaque question n'oubliez pas :
 - d'écrire le code R permettant de répondre à la question
 - de préciser le numéro de la question et de l'exercice en commentaire.

Exemple :

```
# Ex 1-b  
Data<-read.csv()
```

Présentation

Pour cet examen, on utilise le fichier `fao.csv` qui présente la production, l'importation et l'exportation des ressources alimentaires des pays du monde en 2020. les données sont issues du site de la FAO.

Voici une description des données :

- **Nom** : Nom du pays
- **Dispo_alim** : Disponibilité alimentaire totale (en Kcal/personne/jour) en 2020
- **Prod_v viande** : Production de viande en milliers de tonnes en 2020
- **Import_v viande** : Importations de viande en milliers de tonnes en 2020
- **Export_v viande** : Exportations de viande en milliers de tonnes en 2020
- **Population** : Nombre estimé d'habitant en 2018

Conseils

- Pensez à copier le dataset sur votre bureau
- N'ouvrez pas le fichier avec excel mais privilégiez le bloc note ou Notepad+
- Pensez à vous autocorriger avant de vous lancer dans la question suivante
- Pensez à sauvegarder votre script régulièrement

1 Exercice 1 : Importer les données

- Importez le jeu de données `fao.csv` avec la fonction `read.csv()` **uniquement** dans un objet appelé `df`.
- Combien de pays sont présents dans ce dataset ?
- Affichez un résumé des données avec la fonction adaptée. Il semble que ce jeu de données présente quelques valeurs manquantes.

2 Exercice 2 : Statistiques descriptives

- Quelle est la disponibilité alimentaire moyenne mondiale en Kcal/personne/jour ?
- Quelle est le nombre d'habitant dans le monde ?
- Quelle est l'écart-type du volume des exportations et importations de viande.
- Quelle est la médiane du volume de production de viande ?
- Calculez les **quartiles** du nombre de Kcal de disponibilité alimentaire.
- Calculez les **centiles** du volume d'importation de viande.

3 Exercice 3 : Tris et filtres

Pour chaque question, il est recommandé de sauvegarder le résultat de la requête dans un objet puis de le visualiser dans une vue pour vérifier.

- Construire une requête pour extraire les lignes du dataset avec les 5 pays les moins peuplés

- b. Construire une requête pour extraire les lignes du dataset avec les 5 pays les plus peuplés
- c. En moyenne, le besoin énergétique moyen d'une adulte est de 2300 kcal par jour. Construire une requête pour extraire les lignes du dataset avec les pays qui ont une disponibilité alimentaire supérieure ou égale à 2300 kcal. Combien de pays sont concernés ?
- d. Construire une requête pour extraire les lignes du dataset avec les pays qui ont une disponibilité alimentaire strictement supérieure à 3500 kcal et qui importe un volume de viande supérieure ou égale à 1 000 000 tonnes par an. Combien de pays sont concernés ?
- e. Construire une requête pour extraire les lignes du dataset avec la France et la Belgique.

4 Exercice 4 : Modifier le dataframe

- a. Ajouter une colonne nommée *part_export* qui correspond à la part des exportations de viande par rapport à la production de viande
- b. La colonne *Dispo_alim* présente la disponibilité alimentaire par personne. Ajouter une colonne nommée *dispo_alim_pays* qui correspond à la disponibilité total du pays en Kcal/jour.
- c. Calculer la somme de la disponibilité alimentaire mondiale.
- d. Sachant qu'en moyenne, le besoin énergétique moyen d'une adulte est de 2300 kcal par jour. Combien d'adulte pourrait-on nourrir avec la disponibilité alimentaire mondiale ?