

Devoir n°1 : R les bases

M. Hady DIALLO

2023-04-25

Révision Générale

Exercice 1

1. Créer un script R nommé `devoir_nom_exo_rg`
2. créer un data frame qui contient n lignes et m colonnes ($n > 100$, $m > 10$)
 - 5 variables quantitatives ;
 - 5 variables qualitatives de type factor avec une variable binaire. > Indication : vous pouvez utiliser les fonctions “rep”, “seq”, “rnorm”, “dnorm”, etc...
3. Enregistrer le dataframe dans un objet nommé `df.votreNom`
4. Exporter au format “csv” dont `sep=“;”`
5. générer un rapport au format “html”

Exercice 2

1. Créer un document markdown et le nommé `devoir_nom_exo_rg`
2. Inclure un “chunk” et le nommé “Setup”;
3. Définir d’une manière globale que l’ensemble des chunk ne vont pas afficher les erreurs, et qu’ils doivent afficher les résultats
4. Importer dans ce même chunk le data frame que vous avez créé dans l’exo 1 : `__df.votreNom`.
5. Statistique univarié
 - Créer une fonction qui décrit une variable quantitative : `d.var.quant`. La fonction devra afficher les tendances centrales, les graphiques (hist,boxplot, etc...), intervalle de confiance
 - Faire de même pour une variable qualitative : `d.var.quali`
 - Appliquer ces fonctions sur les 10 variables.
6. Statistique bivariée
 - Décrire la liaison entre
 1. Deux variables qualitatives
 2. Deux variables quantitatives
 3. Une variable qualitative et une var quantitative Cette description doit comprendre les tests d’indépendances, les liaisons linéaires, et les représentations graphiques.
7. Générer le rapport en pdf.

Révision sur l'apurement

Exercice 1

1. Importer la base "céréales"
2. Renommer l'ensemble des variables suivant le dictionnaire des variables, en des noms plus parlant et maximum 10 caractères;
3. Décrire les variables d'intérêt
4. identifier les valeurs manquantes et les imputer;
5. Importer la base table de conversion
6. Décrire la variable poids
7. Créer une clé d'identification dans les deux bases : une concaténation des variables produitID-UniteID-TailleID.
8. Fusionner les deux bases
9. Convertir les quantités consommées en unités standards. Ind: Créer une nouvelle variable qui calcule le poids consommé durant les 7 derniers jours ensuite la ramené en année.
10. Générer une variable "taille_men" comprise entre 1 et 30. La moyenne nationale est de 9. La taille est un entier. Cette variable sera unique pour chaque ménage qui est identifié par interview_key. Chaque interview_key est un ménage et donc chaque interview_key devra avoir une valeur unique.
11. Importer la base calories et fusionner celle avec la base. Calculer la consommation calorie par tête. Nettoyer d'abord les valeurs manquantes et les valeurs aberrantes.

Exercice 2 :

Imaginer qu'on doit répéter ce qu'on a fait à l'exercice précédent pour les groupes de produit suivants :

1. Légumes
2. Fruits
3. Boissons etc.

Il est alors demandé de construire des fonctions pour aller plus vite.

1. Fonction d'importation dont l'argument sera le chemin et nom du fichier
2. Fonction qui renomme les variables
3. Fonction qui fusionne deux bases
4. Fonction qui détecte les valeurs manquantes
5. Fonction qui impute les valeurs manquantes
6. Fonction qui détecte les valeurs aberrantes 7. Fonction qui les corrige.

NB : Devoir à rendre dimanche (**30 avril 2023**) avant 23h59. Le devoir est noté sur 5. Le cours prochain nous ferons un test écrit d'environ 30min et c'est sera noté sur 3. Chacun pourra présenter son devoir ce jour.

A
ren-
dre
au
plu-
tard
di-
manche
(**30**
avril
2023)
avant
23h59.
Vam
udatchi
!
