

TP 4 et 5: Rstudio.
Listes et dataframe (tableaux de données statistiques)

1. Listes

Créer et afficher une liste contenant des numériques, des caractères et des booléens.

Créer la liste "Stid", composée du vecteur NbEtud=c(39,26) et du facteur Bac contenant 'S', 'ES', 'ST'.

Quels sont les attributs de la liste ? Obtenir le nom des composantes de la liste.

Quelle est la classe de Stid ? Quelle est la longueur de la liste Stid ?

Obtenir la 3^{ème} valeur de la composante Bac de la liste Stid, avec le nom ou avec l'indice.

Obtenir les longueurs des composantes de la liste Stid.

Créer la liste "notes" des notes par matières d'un élève sous la forme d'une liste de 3 vecteurs de numériques, nommés par le nom de la matière.

Obtenir le vecteur des moyennes dans chaque matière.
en sélectionnant successivement chaque composante
en utilisant la fonction lapply

Extraire la moyenne en math de la liste des moyennes.

Calculer la moyenne équipondérée des 3 moyennes par matière.

Calculer la moyenne de toutes les notes:

avec un calcul de moyenne pondérée

en construisant le vecteur de toutes les notes (fonction unlist).

anglais	math	gym
12	14	17
9	15	15
		13
		17

2. Listes en résultat de fonctions

Appliquer la fonction lm (fitting linear model) pour effectuer une régression linéaire entre deux variables numériques; par exple x : 10 valeurs aléatoires selon une loi normale de moy 5 et d'écart type 2 et y de 10 valeurs aléatoires selon une loi uniforme continue sur [-5,0].

Quels sont les noms des composantes de la liste résultat de la fonction ?

Obtenir les coefficients (a coefficient directeur et b ordonnée à l'origine) de la droite de régression de y en x ?

3. Création data frame

Créer cinq vecteurs, de numériques ou de caractères, nommés v1,v2,v3, v4 et v5 contenant

- ✓ le prénom 'abel', 'ben', 'cati', 'didi', 'eli', 'fani', 'gari'
- ✓ le sexe ,
- ✓ l'âge (4,4,5,5,6,6,6 ans) ,
- ✓ le poids
- ✓ la taille pour 7 individus.

Pour les poids et tailles utiliser le générateur de nombres aléatoires de loi normale (fonction rnorm) et la fonction round pour arrondir.

Créer le data frame, nommé eleves, contenant ces variables, avec comme noms de variables : prenom, sexe, age, poids, taille.

Quelle est la classe de l'objet « eleves » ? sa longueur ? Quels sont ses autres attributs ?

Vérifier le nom des variables en utilisant la fonction names.

Ajouter un nom aux lignes du dataframe, avec la fonction row.names : elev1, ...elev7

str donne un premier aperçu du dataframe et *summary* fournit un premier résumé statistique.

4. Accès aux éléments du data frame

Accès aux colonnes.

Appliquer la fonction `is.factor` à chaque colonne du dataframe : Utiliser le nom ou l'indice des variables et `[...]`, `[[...]]` ou `$`.

Quel est le type des éléments de la variable « sexe » ?

Quelle est la classe de la variable « sexe »

Obtenir les modalités (levels) de cette variable « sexe »

Accès à un élément d'une colonne.

Obtenir l'âge du 2^{ème} élève, de plusieurs manières.

Quelle est la taille de ben ?

5. Sélection d'individus et/ou de variables

- Avec l'opérateur de sélection `[...]`, `[,...,...]`, `$`

Obtenir le prénom de tous les élèves. De ceux qui ont plus de 5ans.

Obtenir le tableau de toutes les variables des 3 premiers élèves.

Obtenir le tableau des tailles et poids.

Sélectionner les enfants dont la taille dépasse la taille moyenne.

Sélectionner les enfants qui ont plus de 5 ans ou qui mesurent plus de la moyenne.

Sélectionner les enfants qui n'ont pas 4 ans.

Supprimer l'élève prénommée fani.

- Avec la fonction `subset` (voir aide)

Refaire les sélections avec cette fonction et les arguments `subset` ou `select`

Créer un data frame contenant la taille et le poids des élèves de plus de 5ans, avec les deux méthodes précédentes. Vérifier l'égalité des deux dataframes construits.

6. Ordonner les individus

Ordonner les élèves selon l'ordre croissant de leur taille. (revoir TP3 et fonction `order`)

7. Création d'un data frame par importation d'un fichier texte

Le fichier `donnees.txt` est sur le WebCampus.

Observer le contenu du fichier avec le bloc note ou notepad++

Dans RStudio : choisir un répertoire de travail. Copier le fichier de données dans ce répertoire de travail.

Utiliser le bouton import dataset pour importer les données dans un dataframe de R.

Quelle est la commande d'importation de ces données ?

Consulter l'aide pour la fonction `read.table`.

Importer les données sous R en utilisant la fonction `read.table`.

Observer le dataframe obtenu, le type des variables... Corriger si besoin...

8. Importation de données d'un tableur

Récupérer le fichier Excel des données sur:

<http://www.biostatisticien.eu/springeR/jeuxDonnees1.html>.

Sous Excel, enregistrer les données au format `.csv`, puis importer sous R, comme indiqué en 7.

Importer les données du fichier pris en exemple au TP1.