

Le but de ces TP est que vous soyez capable de mener une étude statistique sur des données expérimentales, qu'elles soient qualitatives, quantitatives discrètes ou continues. Il est important que vous compreniez les commandes et que vous sachiez les reproduire.

Partie 1 :

Etude sur l'indice de masse corporelle

Un échantillon de dossiers d'enfants a été saisi. Ce sont des enfants vus lors d'une visite en 1^{er} section de maternelle en 1996-1997 dans des écoles de Bordeaux (Gironde, France).

L'échantillon présent ici est constitué de dix enfants âgés de 3 ou 4 ans.

Les données disponibles pour chaque enfant sont :

- Le sexe : F pour fille et G pour garçons ;
- Le fait que leur école soit située en ZEP (zone d'éducation prioritaire, c'est-à-dire réunissant plusieurs indices de précarité sociale) : O pour oui et N pour non;
- L'âge en années et en mois à la date de la visite (deux variables, une pour le nombre d'années et une pour le nombre de mois) ;
- Le poids en kg arrondi à 100 g près;
- La taille en cm arrondie à 0,5 cm près.

Prénom	Erika	Celia	Eric	Eve	Paul	Jean	Aden	Louis	Jules	Léo
Sexe	F	F	G	F	G	G	G	G	G	G
ZEP	O	O	O	O	N	O	N	O	O	O
Poids	16	14	13.5	15.5	16.5	16	17	14.8	17	16.7
An	3	3	3	4	3	4	3	3	4	3
Mois	5	10	5	0	8	0	11	9	1	3
Taille	100.0	97.0	95.5	101.0	100.0	98.5	103.0	98.0	101.5	100.0

En statistique, il est très important de connaître le type des variables étudiées : qualitatives, ordinales ou quantitatives. R permet de spécifier explicitement ce type au moyen des fonctions de structure que nous avons vues dans ce chapitre.

Voilà quelques manipulations à effectuer avec R. Pensez à bien utiliser la stratégie de travail introduite en début de chapitre.

1. Choisissez la fonction R appropriée pour enregistrer les données de chacune des variables précédentes dans des vecteurs que vous nommerez Individus, Poids, Taille et Sexe.
 2. Calculez la moyenne des variables pour lesquelles cela est possible.
 3. Calculez l'IMC des individus et regroupez les valeurs obtenues dans un vecteur nommé IMC (faites attention aux unités).
 4. Regroupez ces variables dans la structure R qui vous paraît la plus adaptée.
 5. Utilisez l'aide en ligne de R afin d'obtenir des informations sur la fonction plot ().
 6. Tracez le nuage de points du Poids en fonction de la Taille.
- Pensez à fournir un titre à votre graphique et à annoter vos axes

Partie 2 :

Fichier pn.xls

1. Importer le fichier pn.xls
2. Ordonner le fichier selon les poids de naissance (BWT) croissants
3. Sachant que la variable RACE désigne la race de la mère (1=blanche,2=noire,3=autre) et la variable SMOKE décrit le tabagisme durant la grossesse (oui=1,non=0),extraire les individus ayant des mères noires ou blanches et qui fument.

Partie 3 :

(A faire chez vous)

La validation d'une matière dépend des résultats au contrôle continu (40 % de la note finale) et de la note à l'examen (60 % de la note finale).

Voici les notes de 8 étudiants :

- Contrôle continu : 11, 13, 15.5, 12, 8, 9, 13, 16
- Examen : 8.5, 7, 16, 9, 7, 13, 14, 17

1. Rentrer ces valeurs dans les variables **ccont** et **exam**.
2. Créer un tableau « **Note** » à 8 lignes et 3 colonnes telles que : la première colonne contienne la variable **ccont**, la deuxième colonne contienne la variable **exam**, la troisième colonne soit la moyenne. Nommer les colonnes de ce tableau de manière adéquate.
3. Ordonner le tableau selon les moyennes croissantes.

Une matière est validée si la moyenne de l'étudiant est supérieure à 10.

4. Déterminer le nombre d'étudiants ayant validés la matière.
5. Extraire la moyenne la plus élevée et la moyenne la plus basse.
6. Déterminer le nombre d'étudiant dont la moyenne est supérieure à 15.
7. Exporter le tableau « **Note** » sur votre ordinateur sous le nom « **Validation** ».