

## Chapitre B

# Quelques jeux de données et problématiques

### Objectif

Ce chapitre présente quelques jeux de données provenant d'études épidémiologiques traitées par différentes équipes de l'Institut de santé publique d'épidémiologie et de développement (ISPED) de Bordeaux. Chacun de ces jeux de données s'accompagne d'une courte problématique qui permettra de mieux comprendre le contexte de l'étude. Ils serviront de fil conducteur tout au long de l'ouvrage. Nous montrerons comment il est possible d'utiliser les différentes fonctionnalités du logiciel R afin d'importer, de manipuler et d'effectuer les analyses statistiques adéquates sur ces jeux de données. Pour chaque jeu de données présenté, un tableau comprenant descriptif, variables et codage est fourni. Le lecteur devra se référer à ce chapitre lorsque ces études seront évoquées. Un tableau qui indique les chapitres dans lesquels chaque jeu de données est utilisé est fourni en fin de chapitre. Notez enfin que ces données sont accessibles sur le site internet associé au livre : <http://www.biostatisticien.eu/springeR>

SECTION B.1

### Indice de masse corporelle (IMC) chez des enfants

#### Présentation :

Un échantillon de dossiers d'enfants a été saisi. Ce sont des enfants vus lors d'une visite en 1<sup>re</sup> section de maternelle en 1996-1997 dans des écoles de Bordeaux (Gironde, France). L'échantillon est constitué de 152 enfants âgés de 3 ou 4 ans.

**Variables et codage :**

Description	Unité ou Codage	Variable
Sexe	F pour fille ; G pour garçon	SEX
École située en zone d'éducation prioritaire	O pour oui ; N pour non	zep
Poids	Kg (arrondi à 100 g près)	poids
Âge à la date de la visite	Années	an
Âge à la date de la visite	Mois	mois
Taille	Cm (arrondi à 0,5 cm près)	taille

**Jeu de données : IMC-ENFANT****Fichier : imcenfant.xls**

SECTION B.2

**Poids de naissance****Présentation :**

Il s'agit d'une enquête concernant les facteurs de risque associés au faible poids de naissance de nourrissons (données collectées au centre médical de Baystate dans le Massachusetts pendant l'année 1986). Le faible poids de naissance est un événement qui intéresse les médecins depuis plusieurs années en raison du taux de mortalité infantile et du taux d'anomalies infantiles très élevés chez les nourrissons de faible poids. Le comportement d'une femme pendant la grossesse (régime alimentaire, habitudes tabagiques...) peut altérer de façon importante les chances de mener la grossesse à terme, et, par conséquent, de donner naissance à un enfant de poids normal. Le fichier de données contient les informations sur 189 femmes (numéro d'identification : ID) venant consulter dans le centre médical. On considère qu'un enfant a un faible poids de naissance si celui-ci est inférieur à 2 500 g.

**Variables et codage :**

Description	Unité ou Codage	Variable
Âge de la mère	Années	AGE
Poids de la mère lors du dernier cycle menstruel	Livres	LWT
Race de la mère	1=Blanche ; 2=Noire ; 3=Autre	RACE
Tabagisme durant la grossesse	Oui=1 ; Non=0	SMOKE
Nombre d'antécédents de prématurité	0=Non ; 1=Un ; 2=Deux ; etc.	PTL
Antécédents d'hypertension	Oui=1 ; Non=0	HT
Présence d'irritabilité utérine	Oui=1 ; Non=0	UI
Nombre de visites à un médecin durant le premier trimestre de la grossesse	0=Aucune ; 1=Une ; etc.	FVT
Poids de naissance	Grammes	BWT
Poids de naissance inférieur ou égal à 2 500 g	Oui=1 ; Non=0	LOW

**Jeu de données : POIDS-NAISSANCE****Fichier : Poids\_naissance.xls**

## SECTION B.3

**Épaisseur de l'intima-média****Présentation :**

L'athérosclérose est la principale cause de décès chez l'homme après 35 ans et chez la femme après 45 ans dans la plupart des pays développés. C'est un épaississement et une perte d'élasticité des parois internes des artères, dont une des conséquences est l'infarctus du myocarde. La paroi artérielle est constituée de trois couches qui sont respectivement à partir de la lumière artérielle : l'intima, la média et l'aventice. L'épaisseur de l'intima-média est un marqueur reconnu d'athérosclérose. Elle a été mesurée par échographie sur un échantillon de 110 sujets en 1999 dans les CHU de Bordeaux. Des informations sur les principaux facteurs de risque ont aussi été recueillies.

**Variables et codage :**

Description	Unité ou Codage	Variable
Sexe	1=Homme ; 2=Femme	SEX
Âge le jour de la visite	Années	AGE
Taille	Cm	taille
Poids	Kg	poids
Statut tabagique	0=Ne fume pas 1=A arrêté de fumer 2=Fume	tabac
Estimation de consommation pour les fumeurs et ex-fumeurs	Nombre de paquets/année	paqan
Activité physique	0=Non ; 1=Oui	SPORT
Mesure de l'intima-média	Mm	mesure
Consommation d'alcool	0=Ne boit pas 1=Boit occasionnellement 2=Boit régulièrement	alcool

**Jeu de données : INTIMA-MÉDIA****Fichier : Intima\_Media.xls**

## SECTION B.4

**Alimentation chez des personnes âgées****Présentation :**

Un échantillon de personnes âgées résidant à Bordeaux (Gironde, France) a été interrogé en 2000 dans le cadre d'une enquête nutritionnelle. L'échantillon est constitué de 226 sujets.

**Variables et codage :**

Description	Unité ou Codage	Variable
Sexe	2=Femme ; 1=Homme	sex
Situation familiale	1=Vit seul 2=Vit en couple 3=Vit dans sa famille 4=Autre type de cohabitation	situation
Consommation journalière de thé	Nombre de tasses	the
Consommation journalière de café	Nombre de tasses	cafe
Taille	Cm	taille
Poids	Kg	poids
Âge le jour de l'entretien	Années	age
Consommation de viande	0=Jamais 1=Moins d'une fois par semaine 2=Une fois par semaine 3=2/3 fois par semaine 4=4/6 fois par semaine 5=Tous les jours	viande
Consommation de poisson	Idem	poisson
Consommation de fruits crus	Idem	fruit_crus
Consommation de fruits et légumes cuits	Idem	fruit_legume_cuits
Consommation de chocolat	Idem	chocol
Matière grasse préférentiellement utilisée pour la cuisson	1=Beurre 2=Margarine 3=Huile d'arachide 4=Huile de tournesol 5=Huile d'olive 6=Mélange d'huile (type Isio4) 7=Huile de colza 8=Graisse de canard ou d'oie	matgras

**Jeu de données : NUTRIAGE****Fichier : nutriage.xls**

## SECTION B.5

## Étude cas témoins sur l'infarctus du myocarde

**Présentation :**

Les données suivantes sont issues d'une enquête cas témoins dont le but était d'évaluer l'existence d'un risque plus élevé de survenue d'un infarctus du myocarde chez les femmes qui utilisent ou ont utilisé des contraceptifs oraux. L'étude a été menée auprès de 149 femmes ayant eu un infarctus du myocarde (cas) et 300 femmes n'en n'ayant pas eu (témoins). Le facteur d'exposition principal est la prise de contraceptifs oraux, les autres facteurs recueillis sont : l'âge, le poids, la taille, la consommation de tabac, l'hypertension artérielle, les antécédents familiaux de maladies cardio-vasculaires.

**Variables et codage :**

Description	Unité ou Codage	Variable
Infarctus du myocarde	0=Témoins ; 1=Cas	infarct
Prise de contraceptifs oraux	0=Jamais ; 1=Oui	co
Consommation de tabac	0=Non 1=Fumeuse actuelle 2=Ancienne fumeuse	tabac
Âge	Années	age
Poids	Kg	poids
Taille	Cm	taille
Antécédents familiaux de maladie cardio-vasculaire	0=Non ; 1=Oui	atcd
Hypertension artérielle	0=Non ; 1=Oui	hta

**Jeu de données : INFARCTUS****Fichier : Infarct.xls**

## SECTION B.6

**Tableau résumant l'utilisation des jeux de données**

	Méthodes					
	Import-Export	Manipulation	Statistique descriptive	Tests	ANOVA	Régression
Jeux de données	IMC-ENFANT	×	×		×	
	POIDS-NAISSANCE	×	×			×
	INTIMA-MÉDIA	×	×		×	×
	NUTRIAGE	×	×	×	×	
	INFARCTUS	×		×		