



Projet Obe Maghreb

Ecole thématique gestion et analyse de données
20 au 29 avril 2010

Gestion et analyse de données d'enquêtes épidémiologiques

Introduction



Agnès Gartner, Pierre Traissac
UMR 204 « Prévention des malnutritions et pathologies associées »
IRD, Montpellier, France



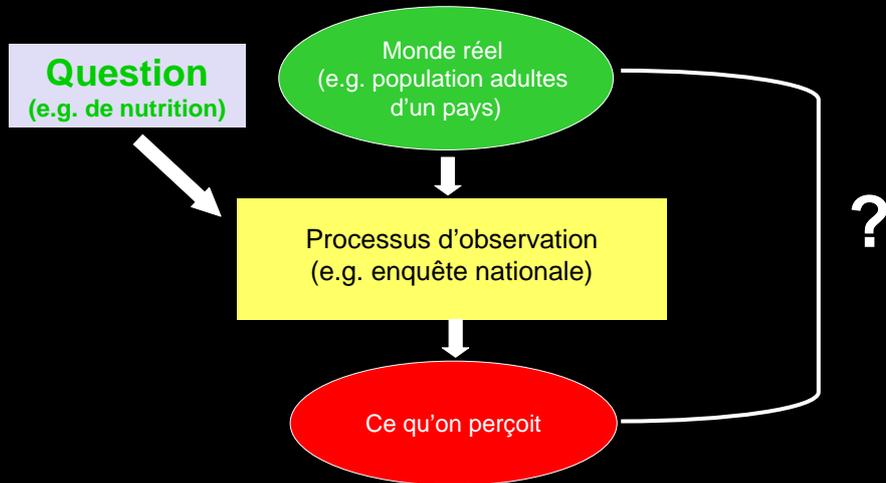
1

Introduction

- Des données à l'information

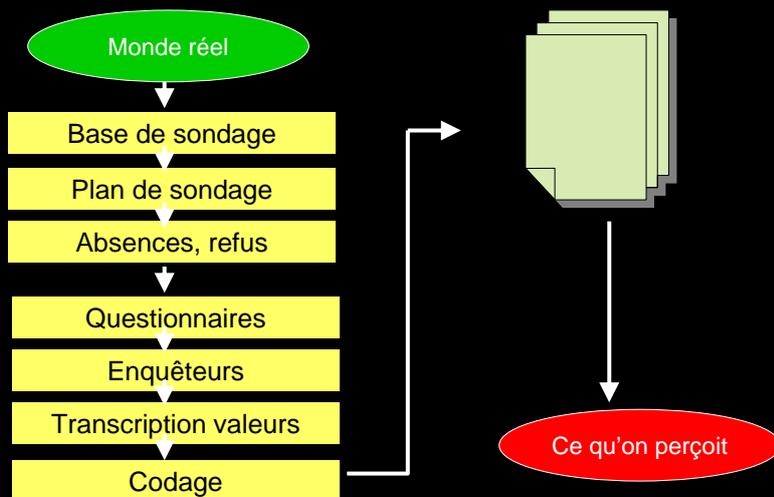
2

Des données à l'information



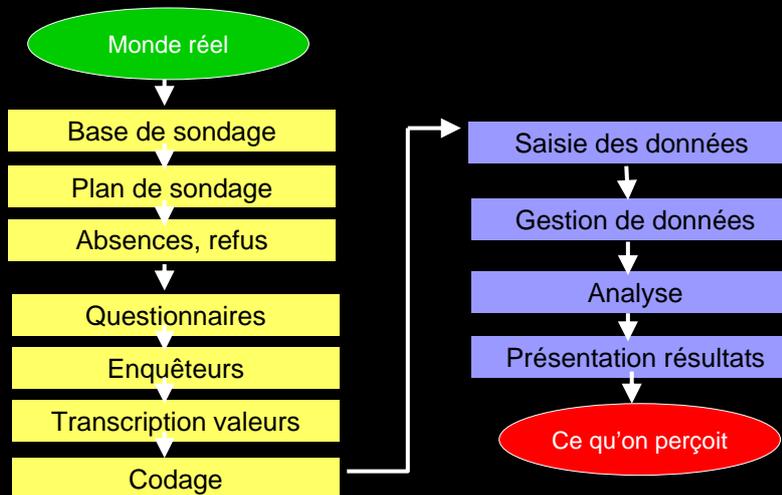
3

Des données à l'information



4

Des données à l'information



5

Des données à l'information

1- Accès aux données →

2- Gestion des données

3- Analyse des données

4- Présentation des données

- **Données informatisées**
 - importation
 - vérification (e.g. appareil de labo)
- **Données « papier »**
 - saisie (lecture, frappe)
 - contrôles
 - validation (double saisie totale / partielle)
 - apurement/vérifications logiques
 - documentation (dictionnaire de données)
- **Outil logiciel spécifique saisie**
 - création utilitaires de saisie
 - contrôle, validation, apurement
 - EpiData, ENTER dans EpiInfo
 - formulaire dans MS/ACCESS,
- **Data manager ↔ spécialiste discipline (corrections)**

6

Des données à l'information

1- Accès aux données

2- Gestion des données

3- Analyse des données

4- Présentation des données

- **Conservation**
 - sécurité
 - sauvegardes
- **Préparation avant analyse**
 - sélections (individus, variables)
 - mise en relation, fusion de fichiers
 - calcul de nouvelles variables
 - recodages
 - documentation (versions fichiers, dictionnaires)
- **Outil logiciel**
 - SGBD (e.g. MS- Access, Oracle, ...)
 - gestion de données dans logiciels statistiques (SAS, Stata, SPSS)
 - ou généralistes (EpiInfo, Epidata Analysis)
- **Data manager ↔ spécialiste discipline**

7

Des données à l'information

1- Accès aux données

2- Gestion des données

3- Analyse des données

4- Présentation des données

- **Synthèse des données**
 - question posée
 - données disponibles
 - caractéristiques de l'enquête
- **Analyse statistique**
 - plan d'analyse +++
 - résumés graphiques, numériques
 - type de variable (quantitative, qualitative)
 - analyses univariées, bivariées, multivariées
 - statistique descriptive, statistique inférentielle
 - documentation (processus, données)
- **Outils logiciels**
 - logiciels spécifiques (SAS, Stata, SPSS, R)
 - fonctions dans logiciels généralistes (Epidata Analysis)
- **Statisticien ↔ spécialiste discipline**

8

Des données à l'information



9

Des données à l'information

1- Accès aux données

2- Gestion des données

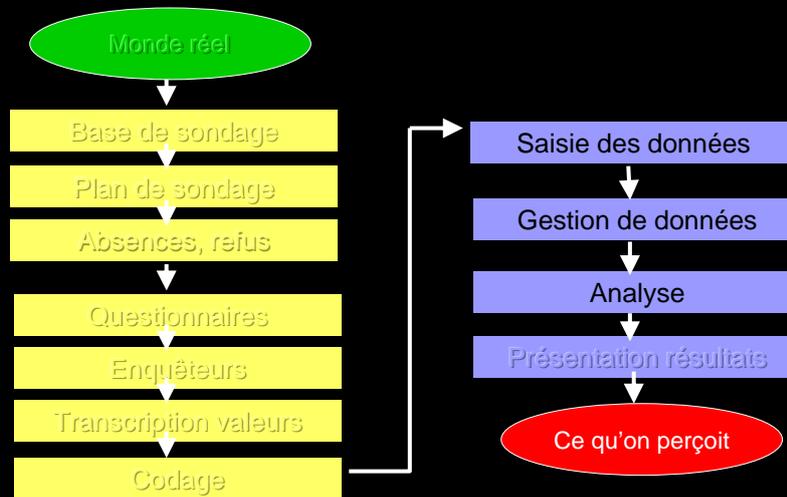
3- Analyse des données

4- Présentation des données →

- **Présenter l'information**
- **But, public ?**
- **Mode de présentation**
 - oral (durée, forme, contexte)
 - écrit (article, rapport, ouvrage)
 - hypermédia (CD, Web)
- **Tableaux, graphiques, cartes**
- **Outils logiciels divers**
 - fonctions graphiques SAS, Stata, Excel
 - spécifiques (PowerPoint, Dreamweaver)
- **Graphiste, concepteur site Web, journaliste ...**
↔ **spécialiste discipline**

10

Des données à l'information



11

Formation Epidata : introduction

- **Pourquoi Epidata, Epidata Analysis ?**

12

Pourquoi EpiData

1- Saisie des données

(Masques de saisie, contrôles, validation)

1- EpiData

2- Gestion des données

(Recodages, calculs de nouvelles variables, fusion de fichiers, sauvegardes)

2- EpiData Analysis (+ Anthro)

3- Analyse des données

(statistiques descriptive, inférentielle, tableaux, graphiques)

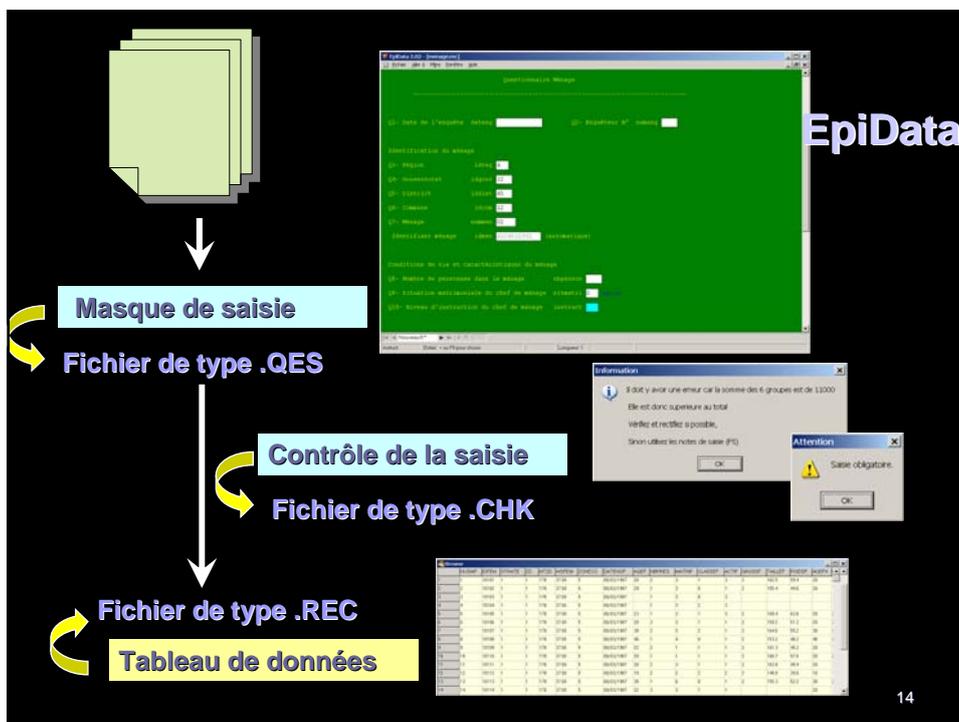
3- EpiData Analysis

4- Présentation des données

(tableaux, graphiques, document écrit, présentation orale)

4- MS Office (Word, Excel, Power Point)
(ou OPEN Office)

13



Pourquoi EpiData

1- Saisie des données

(Masques de saisie, contrôles, validation)

1- EpiData

2- Gestion des données

(Recodages, calculs de nouvelles variables, fusion de fichiers, sauvegardes)

2- EpiData Analysis (+ Anthro)

3- Analyse des données

(statistiques descriptive, inférentielle, tableaux, graphiques)

3- EpiData Analysis

4- Présentation des données

(tableaux, graphiques, document écrit, présentation orale)

4- MS Office (Word, Excel, Power Point) (ou OPEN Office)

15

Epidata Analysis

The screenshot displays the EpiData Analysis software interface. On the left, a 'TABLES ZONECO' summary shows the following data:

ZONECO	No.	%
1	306	12.60
2	518	21.33
3	782	32.26
4	353	14.24
5	459	18.50
Total	2429	100.00

Below the table is a histogram titled 'HISTOGRAM AGEF' showing the distribution of age groups. The x-axis represents age groups (13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49) and the y-axis represents frequency.

The main window displays a data table with the following columns: HUSAE, ICENF, OFEM, PALMENT, STRATE, DO, NTZO, WISENF, ZONECO, DATENQ, DATNAI. The data rows show individual records with values for these variables.

On the right, a command window shows the following commands:

```

1) Lecture table enfants initiale
2) CLOSE
3) READ TENFANTS.REC
4)
5) * calcul age en mois
6) DEFINE AGEEMOIS ###.##
7) AGEEMOIS= ( DATENQ - DATNAI ) / 30.4375
8) **codage age en années révolues
9) DEFINE AGEAN#
10) RECODE AGEEMOIS TO AGEAN 10-11.999=0 12.000-23.999=1 24.000-35.999=2
11) 36.000-47.999=3 48.000-59.999=4
    
```

At the bottom right, a small window displays the text 'l'efface (ATTENTION)'. The page number '16' is visible in the bottom right corner.

Pourquoi Epidata ?

- **Héritier de Epi Info 6**

- Epidata Association, Odense, Danemark (www.epidata.dk)
- logiciel du monde « libre » (non commercial)
- saisie de données, fonctionnalités ++ (Epidata)
- gestion / analyse de données simple (Epidata Analysis)
- +/- compatibilité fichiers / programmes Epi Info 6
- quelques spécificités épidémiologie

- **Langage de programmation (vs. menus)**

- bonnes pratiques, qualité : [traçabilité](#), [documentation](#)
- combine les deux (fenêtre « History », fichiers log)

- **Limites**

- fonctionnalités analyse de données (Epidata Analysis v 2.2.1)
- mises à jours fréquentes (Epidata Analysis v 2.2.1)

17

Formation Epidata : introduction

- **Objectifs de la formation**

18

Objectifs

- **Prise en main des logiciels**
 - Principales fonctionnalités
 - Epidata Analysis +++
- **Saisie de données (Epidata)**
 - Création masque de saisie simple (information)
 - Mise en œuvre de contrôles à la saisie (information)
- **Gestion des données (Epidata Analysis) +++**
 - Calcul variables, indices, recodages
 - Opérations sur tableaux de données : fusions, mise en forme
 - Indices nutritionnels enfants : logiciel ANTHRO OMS (information)
- **Analyse des données (Epidata Analysis) +++**
 - Analyses descriptives simples
 - En particulier analyse données épidémiologiques (RR, OR, confusion)

19

Objectifs

- **Mise en œuvre de bonnes pratiques + +**
 - Principes de qualité
 - Traçabilité
 - Documentation (données, processus gestion / analyse)
- **Logiciel = outil.**
Importance des notions théoriques de base
 - Gestion de données
 - Bio-statistique
 - Epidémiologie

20

Introduction

- **Modalités pratiques**

21

Modalités pratiques (1)

- **Intervenants IRD, Montpellier**
UMR 204 NUTRIPASS : IRD – Univ. Montp. 1 & 2
« Prévention des malnutritions et pathologies associées »

- **Agnès Gartner**

Chargée de recherche, Nutritionniste

- **Pierre Traissac**

Ingénieur de Recherche, Biostatistique-Epidémiologie

22

Modalités pratiques (2)

- **Dix participants**

- Yousra Al Karim	- Asmaa El Hamdouchi
- Otmene Ayyat	- Houda El Hsaini
- Hakim Belghiti	- Khalid El Kari
- Nissrine Choua	- Siham Goumi
- Nadia Derbali	- Ali Jafri

23

Modalités pratiques (3)

- **Logistique**

- IRD (Francis Delpeuch) et UIT (Abdellatif Bour)

- **Pédagogie**

- Exposés, cours magistraux
- Exercices sur ordinateur (en fonction des sujets)
- Participation active de l'auditoire souhaitée

- **Propositions de programme et d'horaires (c.f. document)**

24