

MODULE D'INITIATION A LA STATISTIQUE

MODULE 3 : ELABORATION ET TRAITEMENT DES STATISTIQUES

Section 1 : Sources de données statistiques

1. Recensements

Définition : Le recensement est une opération qui consiste à dénombrer tous les éléments, ou plus communément, tous les individus d'une population donnée.

Avantages : le principal avantage d'un recensement est la couverture complète de toute la population.

Inconvénients :

- Délai long pour la préparation et le traitement ;
- Coût élevé ;
- Impossibilité d'approfondir les sujets.

Section 1 : Sources de données statistiques

2. Sources administratives

Grand nombre de données générées par l'administration :

- Registre de l'Etat civil ;
- Déclarations d'impôt pour les particuliers et les entreprises ;
- Déclarations des effectifs employés par les entreprises pour les pensions et autres avantages sociaux ;
- Déclarations transfrontalières de marchandises et des personnes, etc.

Section 1 : Sources de données statistiques

3. Enquêtes par sondage

L'étude se fait sur une partie de la population appelée échantillon.

Avantages :

- **Rapidité** : l'enquête, si elle est bien préparée et bien ciblée, peut se réaliser assez rapidement sur le terrain et permettre ainsi de disposer de l'information à temps réel pour être utilisée.
- **Coût total** : moins élevé que celui d'un recensement. En effet, l'organisation d'une enquête est moins lourde que celle d'un recensement. L'argument du coût est souvent déterminant pour décider entre le recensement et l'enquête par sondage.

Section 1 : Sources de données statistiques

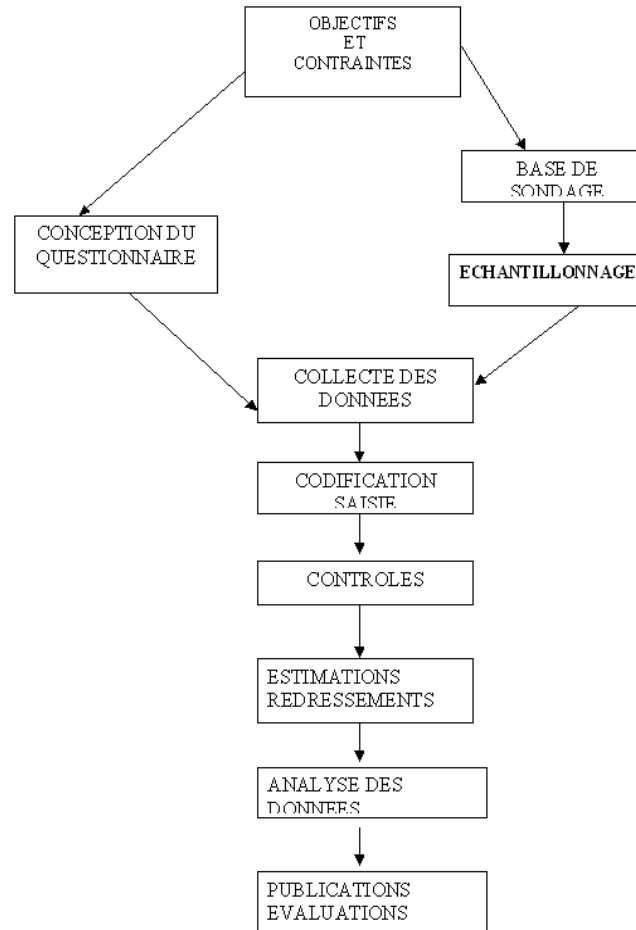
Enquêtes par sondage : Avantages (suite) :

Approfondissements : avec une enquête, les questions peuvent être plus détaillées et permettre ainsi d'approfondir les sujets abordés. De même, une enquête peut aborder plusieurs thèmes en même temps.

Inconvénients : Le principal inconvénient d'une enquête par rapport à un recensement est la non couverture totale de l'ensemble des unités de la population et la précision des résultats. La question de la précision dépend fortement de l'échantillon utilisé pour collecter les données. Si l'échantillon n'est pas suffisamment représentatif de la population, la fiabilité et la précision des données en pâtiront.

Section 2 : Etapes préliminaires à une enquête statistique

SCHEMA GENERAL DE PLANIFICATION D'UNE ENQUETE STATISTIQUE



Section 2 : Etapes préliminaires à une enquête statistique

1. Définition des objectifs

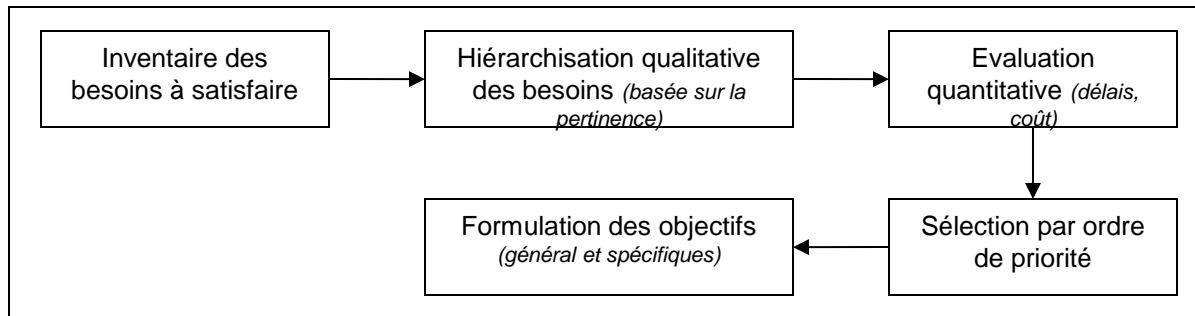
Étapes à suivre pour la définition des objectifs d'une enquête

Étape 1 : *inventaire des besoins à satisfaire*

Dans cette étape, on demande au statisticien de dialoguer en allant vers les commanditaires et d'autres utilisateurs potentiels des informations qui seront issues de l'enquête, afin de prendre en compte leurs préoccupations.

Les consultations sont utiles car les enquêtes sont des opérations coûteuses et il faut profiter de l'occasion pour analyser un problème dans tous ses aspects et obtenir tous les résultats utiles.

PROCEDURE DE DEFINITION DES OBJECTIFS D'UNE ENQUÊTE



Section 2 : Etapes préliminaires à une enquête statistique

1. Définition des objectifs (suite)

Étape 2 : hiérarchisation des besoins

La hiérarchisation des besoins se fait sur une base qualitative. Il est question ici de la pertinence des différents besoins répertoriés au regard des objectifs de l'enquête. **La décision doit être prise par les commanditaires de l'étude.**

Étape 3 : Évaluation quantitative

Dans cette étape apparaissent des contraintes qui sont susceptibles de peser sur la décision de faire ou de ne pas faire. Deux types de contraintes existent :

Les délais : ils doivent être raisonnables afin que les données puissent être utiles. La fixation des délais conditionne toute l'opération et est l'élément fondamental pour le calcul du budget de l'enquête. **Les délais sont fixés souvent par les commanditaires.**

Le coût de l'opération : il est intimement lié aux délais. Mais il est aussi fonction d'autres paramètres tels que la taille de l'échantillon, la méthode de collecte, la méthodologie.

Section 2 : Etapes préliminaires à une enquête statistique

1. Définition des objectifs (suite)

Étape 4 : Sélection par ordre de priorité

Comme dans toute planification, les contraintes ci-dessus évoquées imposent de faire une sélection des besoins en tenant compte des priorités. Ainsi, certains besoins peuvent être purement et simplement abandonnés compte tenu des moyens dont on dispose pour réaliser l'enquête.

Étape 5 : formulation des objectifs

Les objectifs sont de deux catégories :

L'objectif général : le centre d'intérêt recherché, la finalité de l'étude engagée.

Les objectifs spécifiques : ceux qu'on peut mesurer avec les données issues de l'enquête et donc bien ciblés.

Section 2 : Etapes préliminaires à une enquête statistique

1. Définition des objectifs (suite)

Étape 6 : la prise de décision.

Cette étape est fondamentale, puisqu'elle peut conduire, le cas échéant, à renoncer à mener l'opération, économisant ainsi des ressources. Ceci pourrait par exemple se produire si les besoins inventoriés peuvent être satisfaits par une enquête similaire réalisée par une autre structure, ou encore si une opération statistique ne pourra pas apporter une réponse satisfaisante au problème posé.

Section 2 : Etapes préliminaires à une enquête statistique

2. Élaboration de la méthodologie

La méthodologie comporte les étapes suivantes :

- **Champ de l'enquête**
- **Base de sondage**
- **Échantillonnage**
- **Taille de l'échantillon**

Section 2 : Etapes préliminaires à une enquête statistique

2. Élaboration de la méthodologie (suite)

1) Champ de l'enquête

But : délimitation de l'ensemble sur lequel va porter l'observation. Il comporte trois étapes :

Étape 1 : définition précise des unités à enquêter en faisant la distinction entre unités de sondage et unité statistique

Exemple: l'entreprise est l'unité de sondage et le salarié est l'unité statistique à enquêter.

Section 2 : Etapes préliminaires à une enquête statistique

2. Élaboration de la méthodologie (suite)

1) Champ de l'enquête (suite)

Étape 2 : dimension géographique ou spatiale de l'enquête

 Tout le territoire ou une partie

Étape 3 : dimension temporelle

 Période d'enquête et période d'observation

 Exemple: l'enquête se réalise en 2012 (période d'enquête) et porte sur les données de 2011 (période d'observation).

Section 2 : Etapes préliminaires à une enquête statistique

2. Élaboration de la méthodologie (suite)

2) Base de sondage

Définition : liste de toutes les unités faisant partie du champ de l'enquête

Exigences principales d'une base de sondage :

- Exacte et à jour
- Exhaustive
- Délimitations physiques claires et repérage facile sur le terrain

Exemples de base de sondage:

- Fichier : liste exhaustive des ménages dans une région du pays
- Fichier contenant la liste complète des entreprises

Section 2 : Etapes préliminaires à une enquête statistique

2. Élaboration de la méthodologie (suite)

2) Base de sondage(suite)

Qualités souhaitables d'une base de sondage :

- Effectifs des unités de sondage dans les zones à dénombrer assez homogènes
- Taille des unités de sondage connue approximativement avant le tirage
- Connaissance de certaines caractéristiques (informations préalables des unités aux fins d'une éventuelle stratification)

Si base de sondage inexistante, en créer une.

Si exigences non respectées, actualiser la base de sondage.

Section 2 : Etapes préliminaires à une enquête statistique

2. **Élaboration de la méthodologie (suite)**

3) **Echantillonnage**

- Chercher à voir s'il existe une liste des unités d'échantillonnage
- Chercher la cartographie (ou localisation) de ces unités
- Organiser l'actualisation des listes : cette étape peut nécessiter une descente sur le terrain
- Fixer la taille de l'échantillon
- Décider de la méthode d'échantillonnage qui convient pour l'étude envisagée et procéder au tirage de l'échantillon

Section 2 : Etapes préliminaires à une enquête statistique

2. Élaboration de la méthodologie (suite)

3) Echantillonnage

Démarche générale pour la construction d'un échantillon : Chercher à voir s'il existe une liste des unités d'échantillonnage

- Chercher la cartographie (ou localisation) de ces unités
- Organiser l'actualisation des listes : cette étape peut nécessiter une descente sur le terrain
- Fixer la taille de l'échantillon
- Décider de la méthode d'échantillonnage qui convient pour l'étude envisagée et procéder au tirage de l'échantillon

Section 2 : Etapes préliminaires à une enquête statistique

2. Élaboration de la méthodologie (suite)

4) Taille de l'échantillon

Relève d'un compromis entre précision souhaitée et budget alloué à l'enquête.

La taille de l'échantillon doit être suffisante pour permettre l'extrapolation des résultats au niveau de la population entière.

Prévoir un échantillon plus large afin d'anticiper les cas de refus, d'interruption, d'absence.

Section 2 : Etapes préliminaires à une enquête statistique

3. Méthode d'observation

Comparaison des méthodes d'observation

Méthode	Coût	Qualité des données
Enquête par voie postale	Réduit	Moindre (taux élevé de non-réponses)
Interview	Élevé	Plus fiable
Enquête par téléphone	Intermédiaire	Intermédiaire
Enquête par Internet	réduit	Moindre (taux élevé de non-réponses)

Section 2 : Etapes préliminaires à une enquête statistique

3. Instruments pour la collecte des données

1) Questionnaire

Document destiné à relever les caractères sur les unités à enquêter.

Questionnaire individuel : une fiche pour chacune des unités à enquêter.

Questionnaire collectif : destiné généralement à une institution

Section 2 : Etapes préliminaires à une enquête statistique

3. Instruments pour la collecte des données

1) Questionnaire (suite)

Qualités générales d'un bon questionnaire statistique

- Commodité pour les personnes interrogées
- Les questions doivent faciliter la réponse des personnes interrogées
- Précision et clarté des questions posées
- Le questionnaire doit être aussi précis si l'on veut que les réponses le soient aussi.
- Laisser peu d'initiative aux enquêtés
- Eviter des questions qui laissent trop d'initiative aux enquêtés. Au besoin, il faut proposer plusieurs alternatives de réponse (voir question à choix multiples).

Section 2 : Etapes préliminaires à une enquête statistique

3. Instruments pour la collecte des données

1) Questionnaire (suite)

Qualités générales d'un bon questionnaire statistique (suite)

- Agréable à la vue (ergonomie)
- Avoir un nombre raisonnable de questions (pour éviter les coûts superflus)
- Adapté au niveau intellectuel des répondants
- Adapté à la méthode d'observation
- Ne rien omettre d'essentiel
- Le questionnaire doit être adapté aux besoins et aux caractères que présentent les unités enquêtées.

Section 2 : Etapes préliminaires à une enquête statistique

3. Instruments pour la collecte des données

1) Questionnaire (suite)

Types de questions

✓ *Questions fermées*

Les questions fermées sont des questions pour lesquelles on impose au répondant une forme précise de réponse et un nombre limité de choix de réponse.

✓ *Questions ouvertes*

Une question ouverte laisse la réponse libre dans sa forme et dans sa longueur.

✓ *Questions semi-ouvertes :*

Les questions semi-ouvertes proposent une série de réponses à laquelle on ajoute une réponse « autre » (à préciser).

Section 2 : Etapes préliminaires à une enquête statistique

3. Instruments pour la collecte des données

1) Questionnaire (suite)

Validation du questionnaire

But : s'accorder sur les termes-clés utilisés dans le cadre de l'étude

La validation du questionnaire peut consister en :

- L'étude de la documentation existante sur le sujet ;
- Une réunion avec des personnes extérieures mais connaissant le domaine ;
- Un entretien de groupe permettant d'interroger des personnes compétentes ;
- Un test d'interview.

Section 2 : Etapes préliminaires à une enquête statistique

3. Instruments pour la collecte des données

2) Manuels d'instruction

Manuel d'instructions de l'agent enquêteur

Manuel d'instructions du Contrôleur

Manuel de l'agent de saisie

Guide du Superviseur

Section 2 : Etapes préliminaires à une enquête statistique

4. Chronogramme et calendrier

Représentation graphique de l'ordonnancement des tâches ou opérations

But : s'accorder

Exemple de chronogramme d'une étude

	1re	2e	3e	4e	5e	6e	7e	8e	9e	10e	11e	12e	semaine
tâche n° 1	■	■	■										
tâche n° 2			■	■									
tâche n° 3			■	■	■								
tâche n° 4					■	■							
tâche n° 5					■	■	■	■					
tâche n° 6							■	■					

Section 2 : Etapes préliminaires à une enquête statistique

5. Budget

Principales rubriques :

- dépenses de personnel,
- dépenses d'équipement,
- dépenses de fonctionnement,
- dépenses de publicité et de sensibilisation.

Section 2 : Etapes préliminaires à une enquête statistique

6. Recrutement et formation

- La formation est le moyen d'assurer l'uniformité des procédures et leur application correcte sur le terrain par tout le personnel engagé.
- La formation peut être distinguée en tronc commun pour tout le personnel et des modules spécialisés pour les agents de saisie, de collecte et les superviseurs.
- Insister sur les concepts et définitions.
- Utilisation de certains instruments de mesure par des simulations.

Section 2 : Etapes préliminaires à une enquête statistique

7. Pré-test et enquête pilote

Pré-test a pour objet :

- évaluer la viabilité du questionnaire
- Préparer la version du questionnaire de l'enquête pilote.

Enquête pilote a pour but :

- revoir l'ensemble des processus de collecte et de traitement des données ;
- Tester les manuels d'instructions aux **agents enquêteurs, aux Contrôleurs, aux agents de vérification et agents de saisie**
- Tester les procédures de dénombrement et de sélection des unités de l'enquête.

Section 3 : Traitement des données d'une enquête statistique

1. Traitement physique des questionnaires

Activités à mener :

- Contrôle de l'état physique des questionnaires
- Contrôle de la qualité du remplissage
- Synthèse par niveau de collecte pour faire le point par zone d'enquête
- Classement des questionnaires par zone d'enquête
- Calcul du taux de couverture

Section 3 : Traitement des données d'une enquête statistique

2. Codification des questionnaires

Transcription des réponses obtenues en codes numériques simples pour faciliter la saisie informatique

Activité particulièrement importante pour les questions ouvertes. Elle incombe aux agents de codification sous le contrôle de l'équipe de conception et d'analyse des données.



Il est conseillé de faire au maximum la pré codification des variables

Section 3 : Traitement des données d'une enquête statistique

3. Saisie des données

C'est le processus par lequel on transfère les données du support de collecte (le questionnaire) vers un fichier informatique.

- a) Nécessite l'utilisation de logiciels appropriés pour créer des masques de saisie contrôlée
Masque de saisie : programme informatique présentant une interface destinée à recevoir des codes représentant les réponses aux questions en vue de constituer une base de données
(Logiciels EPI INFO, CSPRO)
- b. Saisie directe : logiciels Excel, SPSS (peu conseillée)

Section 3 : Traitement des données d'une enquête statistique

3. Saisie des données (suite)

Informations requises pour la saisie des données :

- le nom de la variable,
- la longueur (nombre de positions occupées),
- le type (numérique, alphanumérique, date, ...),
- le nombre de décimales (si le type est numérique),
- le libellé de la variable,

Section 3 : Traitement des données d'une enquête statistique

3. Saisie des données (suite)

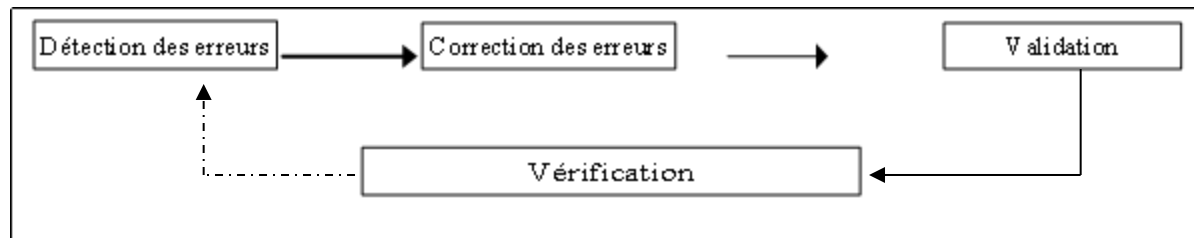
Informations requises pour la saisie des données (suite) :

- les valeurs des modalités (cas d'une variable qualitative),
- le traitement des valeurs manquantes (ne sait pas, n'est pas concerné, blanc, etc.),
- l'alignement de la variable (droit, gauche)
- Les sauts autorisés
- La logique et l'enchaînement des questions

Section 3 : Traitement des données d'une enquête statistique

4. Apurement des fichiers

L'apurement d'un fichier brut au terme de la saisie a pour objectif d'éliminer toutes les erreurs susceptibles de se retrouver dans un fichier de données



Section 3 : Traitement des données d'une enquête statistique

4. Apurement des fichiers (suite)

1) *Détection des erreurs : principales sources d'erreurs*

- **erreurs d'échantillonnage ou de sondage** sont dues à des défauts dans la base de sondage
- **erreurs d'observation ou de collecte** (concept, formulation des questions, qualité des manuels et de la formation, mauvaise foi de l'enquêté, méthodes d'observation inadéquates, mauvais dialogue entre l'enquêteur et les enquêtés, etc.) ;
- **erreurs de codification** (lors de la transcription des codes ou code mal défini) ;
- **erreurs de saisie** (lecture des chiffres, confusion des touches lors de la frappe) ;
- **erreurs algorithmiques** (imperfection des algorithmes de calcul et de redressement des données).

Section 3 : Traitement des données d'une enquête statistique

4. Apurement des fichiers (suite)

2) *Détection des erreurs : procédure de contrôle*

Des tests de cohérence : sont élaborés pour détecter de façon systématique les données erronées

- *tests sur les données* : permettent de vérifier la présence de toutes les données ; ils permettent de repérer les données manquantes.
- *tests sur les variables* : étudient la vraisemblance des valeurs prises par les variables ; ils examinent les valeurs extrêmes et les codes erronés.
- *Tests sur les observations* : permettent de vérifier l'exhaustivité et l'ordre de remplissage du questionnaire (les séquences).
- *Tests de contrôle interne* : permettent de faire le recoupement de plusieurs variables pour déceler des incohérences.
- *Tests graphiques* : les graphiques sont un moyen très utile pour mettre en exergue des incohérences concernant les données.

Section 3 : Traitement des données d'une enquête statistique

4. Apurement des fichiers (suite)

3) *Correction des erreurs*

Deux solutions : (dire quand on doit mettre en œuvre chacune des solutions)

- Imputation : à partir de valeurs exogènes ou de caractéristiques de tendance centrale
- Modification des pondérations

Section 3 : Traitement des données d'une enquête statistique

5. Traitement de données manquantes ou non-réponses

Deux types :

- *Non-réponse totale*
- *Non-réponse partielle*

Problèmes posés par les non-réponses

- *Sous-estimation ou surestimation*
- *Précision moindre des estimations et donc présence de biais.*

Section 3 : Traitement des données d'une enquête statistique

5. Vérification et validation des données

La validation des données est l'opération qui consiste à vérifier que chaque information comporte une donnée acceptable et que celle-ci n'est pas contradictoire par rapport à une autre connue auparavant. Elle vise donc à assurer la cohérence d'ensemble des informations saisies avec les données connues a priori.

Pour ce faire, on élabore des règles de vérification pour repérer et corriger les incohérences entre les réponses à l'intérieur du questionnaire

Section 3 : Traitement des données d'une enquête statistique

6. Vérification et validation des données (suite)

Méthode : tris à plat et tris croisés : permettent de :

- Contrôler la codification pour les variables qualitatives à partir des nomenclatures et codification élaborées ;
- Contrôler les plages pour une variable quantitative;
- Contrôler la compatibilité ;
- Contrôler la vraisemblance par exemple (si Y et X liées) ;
- Contrôler la cohérence. Pour des variables quantitatives on a des contrôles comptables ou algébriques ($x+y=z$), des contrôles logiques pour les variables qualitatives.

Section 3 : Traitement des données d'une enquête statistique

7. Extrapolation

C'est l'opération qui consiste à passer des indicateurs calculés sur les données de l'échantillon à l'estimation au niveau de la population toute entière.

Elle nécessite de calculer des coefficients de pondération qui sont en fait l'inverse des probabilités d'inclusion ou d'appartenance à l'échantillon.

Les coefficients sont calculés en utilisant les données du plan de sondage, corrigées avec les données de la structure réelle enquêtée

Des commandes dans le logiciel SPSS permettent d'appliquer des coefficients de pondération.

Section 3 : Traitement des données d'une enquête statistique

8. Tabulation

C'est l'opération qui consiste à extraire des tableaux (simples et croisés) de la base de données à des fins d'analyse. C'est la raison pour laquelle elle intervient après le nettoyage du fichier de données.

La tabulation doit suivre le remplissage des maquettes de tableaux conçus suivant le plan de tabulation

Section 4 : Rôle de l'informatique

1. **Intervention dans les différentes phases d'une enquête**
 - **Conception du questionnaire** et des autres documents de la collecte des données
 - **Traitement des données** : cette phase nécessite la connaissance et la manipulation d'au moins un logiciel spécialisé pour le traitement des données.
 - **Analyse des données** : pour l'édition des tableaux, le calcul des caractéristiques, les graphiques, etc.
 - **Publication** : l'utilisation de logiciels de publication peut s'avérer indispensable à ce stade.
 - **Archivage** : les logiciels spécifiques seront utilisés, notamment Data management toolkit du réseau IHSN pour les enquêtes.

Section 4 : Rôle de l'informatique

2. Bases et Banques de données

Elles permettent :

- Rassembler toutes les informations nécessaires dans un même ensemble ;
- Réaliser une conception intégrée et homogène de toutes les informations et de leurs relations en de fichiers inter reliés.

Une base de données est une collection de représentation de la réalité sous forme de données inter reliées :

- Aussi cohérentes possibles ;
- Mémorisées avec une redondance calculée ;
- Structurées de manière à faciliter leur exploitation.

Section 4 : Rôle de l'informatique

3. Archivage et diffusion des données

Les données sont archivées pour ne plus être modifiées, mais conservées comme telles. Elles pourront cependant être extraites et exploitées.

Les données seront archivées selon plusieurs objectifs:

- *Constituer des sauvegardes* : conserver une copie des données pendant que l'on travaille sur l'originale ;
- *Transmettre des données* : transmettre des données en un bloc plus facilement et rapidement à une ou plusieurs personnes ;
- *Clôturer une session de travail* : conserver les résultats d'un travail achevé pour une exploitation future ;
- *Diffuser les résultats d'un travail* : diffusion en fichiers ou en support CD/DVD Rom ou par Internet
- Réaliser une archive nationale des données d'enquête en accord avec la politique de diffusion du pays.

Je vous remercie