



Organisation des Nations Unies  
pour l'alimentation et l'agriculture

# STRATEGIE MONDIALE POUR L'AMELIORATION DES STATISTIQUES AGRICOLES ET RURALES

## FORMATION EN STATISTIQUES AGRICOLES

### Module0: Quelques rappels statistiques



# Quelques rappels statistiques

# 0.1 Quelques définitions (1/14)

- La population
- La population cible
- L'unité statistique
- La population échantillonnée
- L'unité d'échantillonnage ou de sondage
- La donnée
- La statistique / Les statistiques
- La période d'enquête
- Période de référence

# 0.1 Quelques définitions (2/14)

- **La population** est composée de l'ensemble des individus ayant en commun des critères d'identification ou de définition.
- Exemple:
  - Population des exploitations agricoles
  - Population des exploitants agricoles
  - Main d'œuvre agricole masculine
  - Actifs agricoles, équipements et machines
  - Tous les actifs agricoles, équipements et machines en bon état pour la campagne agricole actuelle
- ✓ *Cette population pourrait, dans certaines situations, coïncider avec la population cible\* ou la population échantillonnée\**

# 0.1 Quelques définitions (3/14)

- **La population cible** ou champ d'une enquête (target population): est la population sur laquelle porte les informations désirées. Il convient de la définir soigneusement pour chaque étude.
- **L'unité statistique** (ou individu) est un élément quelconque, dont l'ensemble compose la population ou l'univers. En d'autres termes, c'est un élément de la population (cible). Elle peut être de natures différentes (villages, ménages, hameaux, parcelles cultivées, entreprises, etc.), par exemple :

Population	Unité statistique
Main d'œuvre agricole	Un employé agricole quelconque ou un manœuvre agricole quelconque
Actifs agricoles, équipements et machines	Un tracteur

# 0.1 Définition des concepts (4/14)

- **La population échantillonnée** (sampled population) : population effectivement concernée par le champ de l'étude. Elle ne coïncide pas toujours avec la population cible. Cette situation pourrait se produire dans les cas suivants:
  - Problème méthodologique : difficulté d'identifier directement les unités statistiques désirées et les sélectionner.
  - Problème de couverture : exclusion de certaines unités. Difficulté d'accès aux unités répondantes.

*Dans le cas où les deux populations sont différentes, la population échantillonnée devrait être raisonnablement cohérente en termes de couverture et de correspondance avec la population cible pour que les résultats de l'enquête soient pertinents.*

# 0.1 Définition des concepts (5/14): Population échantillonnée

- Problème méthodologique : difficulté d'identifier directement les unités statistiques désirées et les sélectionner.
  - Dans la pratique, il est quelquefois difficile d'identifier et sélectionner directement les unités statistiques et y avoir accès.
  - Exemple: Inventaire des équipements agricoles disponibles et leur état (fonctionnel) au cours d'une campagne agricole donnée
  - Dans ce cas, l'unité statistique est l'équipement agricole. Par ailleurs, il s'avère presque impossible d'avoir accès directement aux équipements agricoles pour les inventorier et apprécier leur état fonctionnel.

# 0.1 Définition des concepts (6/14): Population échantillonnée

- Problème méthodologique : difficulté d'identifier directement les unités statistiques désirées et les sélectionner.
  - Toutefois, il semble aisé d'avoir accès aux ménages agricoles (exploitants / exploitations agricoles). Dans ce cas de figure, le ménage agricole (exploitant/ exploitation agricole) constituera l'unité d'échantillonnage.
  - Les équipements ainsi disponibles au sein des ménages (exploitants / exploitations agricoles) sont répertoriés et les informations d'intérêt sont recueillies.



# 0.1 Définition des concepts (7/14): Population échantillonnée

- Problème de couverture: exclusion de certaines unités. Difficulté d'accéder aux unités répondantes.
- Quelques exemples:
  - Exclusion des zones isolées (du faite des coûts de déplacement relativement importants)
  - Exclusion des activités agricoles effectuées par certaines institutions (ex: prisons...)
  - Base de sondage liste non exhaustive, résultant de l'absence d'information sur certaines unités

# 0.1 Définition des concepts (8/14)

- **L'unité d'échantillonnage ou de sondage**
  - est constitué de chaque « membre » de la base d'échantillonnage. Cette base est la liste exhaustive des unités statistiques d'une population donnée
  - est l'unité directement soumise à une opération de sélection.
  - peut être :
    - une exploitation agricole
    - une parcelle agricole
    - un ménage
    - un enfant
    - un logement
    - une école
    - une formation sanitaire
- **L'unité d'analyse:** niveau sur lequel porte l'analyse. L'exploitation agricole peut être par exemple l'unité d'échantillonnage (on sélectionne donc des exploitations agricoles), mais l'analyse peut porter sur des parcelles qui sont alors les unités d'analyse.

# 0.1 Définition des concepts (9/14)

- **L'unité répondante** (ou déclarante) est l'intermédiaire qui fournit l'information sur chaque unité statistique (ex. : un chef d'exploitation agricole interviewé sur ses parcelles; une mère interrogée sur ses enfants, un directeur interrogé sur son école).
- **L'unité d'observation ou de référence** (observation unit, unit of interest)
  - est celle sur laquelle on demande des informations. Par exemple :
    - les exploitations agricoles, les parcelles pour lesquelles le chef de l'exploitation agricole a donné des informations ;
    - les enfants pour lesquels la mère a donné des informations ;
    - l'école primaire publique pour laquelle le directeur a donné des informations.
  - C'est donc l'objet sur lequel on fait une mesure
  - C'est l'unité de base observée
  - Pour des populations humaines, c'est un individu

# 0.1 Définition des concepts (10/14)

- **La donnée** est l'élément de base d'un système d'information plus large. Lorsque les statisticiens produisent des données, ils essaient de mesurer ou de compter les phénomènes (individus ou activités) qui font partie du monde réel.

Exemples de données:

- Nombre de vaches dans une ferme
- Superficie d'un champ
- Nombre de personnes dans un ménage
- Nombre d'enfants dans une famille

*Les données ne sont pas très utiles par elles-mêmes. Elles doivent être organisées dans des statistiques pour les rendre compréhensibles et utilisables.*

# 0.1 Définition des concepts (11/14)

- **Les statistiques** sont les compilations de faits numériques et de chiffres. Ces faits et chiffres sont créés à partir de données et sont organisés pour leur utilisation. Elles apparaissent dans les tableaux, diagrammes, graphiques ou cartes.

Elles peuvent provenir :

- D'enquêtes (recensement ou par sondage);
- De sondages d'opinion;
- De données administratives (par exemple, les importations et exportations).

*Il faut distinguer les statistiques officielles (productions des structures étatiques reconnues par le système statistique national) et non officielles (production des privés).*

# 0.1 Définition des concepts (12/14)

- **La statistique** est une science mathématique qui se concentre sur la collecte, l'analyse, l'interprétation ou l'explication, et la présentation des données.
- **L'information** est de la donnée traitée et communiquée (c'est à dire mise, sous une forme ou une autre, à la disposition du public):
  - Ainsi, une donnée non diffusée n'est pas une information
  - Elle ne peut être à usage général, mais plutôt, l'information porte sur une question particulière.
  - L'information dans ce contexte peut être utilisée pour soutenir les décisions dans une variété de situations.

# 0.1 Définition des concepts (13/14)

- **La période d'enquête** correspond à celle pendant laquelle la collecte des données sera effectuée sur le terrain. Elle doit garantir un bon contrôle des procédés d'identification de l'échantillon (neutralisation des effets saisonniers en particulier).

## Exemple :

Les enquêtes agricoles se déroulent en général pendant la période couvrant le cycle cultural. Pour organiser le passage des enquêteurs, il faut tenir compte du calendrier des cultures, de leur cycle de croissance et des différences entre les céréales, les tubercules et racines, les légumes, les cultures de rente, les cultures fruitières et autres cultures.

# 0.1 Définition des concepts (14/14)

- **Période de référence** : C'est la période à laquelle se rapportent les données. Elle dépend des objectifs de l'enquête. Selon les cas c'est un intervalle de temps (semaine, mois, année, campagne agricole, etc.) ou d'une date précise. Il faut noter que dans une même enquête des variables peuvent avoir des périodes de référence différentes.

## Exemple :

La période de référence d'une enquête de production végétale est la campagne agricole. La période de référence pour les naissances, les acquisitions et les morts naturelles des animaux dépend du type d'animal. En général, elle est d'un an pour les bovidés, de six mois pour les petits ruminants et porcs et d'un mois pour la volaille.



# 0.2 Les étapes d'une enquête statistique (1/2)

Les principales étapes du déroulement d'une enquête sont :

- L'identification des besoins en information / Définition des objectifs de l'enquête et des moyens
- Détermination de la période d'enquête et de la période de référence
- Plan de travail, budget
- Choix de la base et des unités de sondage
- Choix du plan de sondage
- Méthode de collecte de données

## 0.2 Les étapes d'une enquête statistique (2/2)

- Conception des documents techniques (questionnaires, manuels d'instruction...)
- Recrutement et formation du personnel
- Tests et/ou enquêtes pilotes
- Organisation et suivi des activités sur le terrain
- Traitement des données (saisie, tabulation, traitement et analyse)
- Préparation du rapport et la diffusion des résultats

# Figure 1 : Différentes étapes d'une enquête

Source : GSBPM v5

Gestion de la qualité/Gestion des métadonnées							
Définition des besoins	Conception	Élaboration	Collecte	Traitement	Analyse	Diffusion	Évaluation
1.1 Identification des besoins	2.1 Conception des produits	3.1 Élaboration de l'instrument de collecte	4.1 Création d'un cadre et sélection d'un échantillon	5.1 Intégration des données	6.1 Élaboration du projet de produits	7.1 Actualisation des systèmes de produits	8.1 Recueil des produits d'évaluation
1.2 Consultations et validation des besoins	2.2 Conception des descriptions de variables	3.2 Élaboration ou renforcement des composantes du processus	4.2 Organisation de la collecte	5.2 Classification et codage	6.2 Validation des produits	7.2 Élaboration des produits de diffusion	8.2 Conduite de l'évaluation
1.3 Établissement des objectifs en matière de produits	2.3 Conception de la collecte	3.3 Élaboration ou renforcement des composantes de diffusion	4.3 Réalisation de la collecte	5.3 Examen et validation	6.3 Interprétation et explication des produits	7.3 Gestion de la publication des produits de diffusion	8.3 Adoption d'un plan d'action
1.4 Identification des concepts	2.4 Conception du cadre et de l'échantillon	3.4 Configuration du déroulement des travaux	4.4 Finalisation de la collecte	5.4 Édition et imputation des données	6.4 Mise en place du contrôle de la divulgation	7.4 Promotion des produits de diffusion	
1.5 Vérification de la disponibilité des données	2.5 Conception du traitement et de l'analyse	3.5 Mise à l'essai du système de production		5.5 Calcul de nouvelles variables et unités	6.5 Finalisation des produits	7.5 Gestion de l'assistance aux utilisateurs	
1.6 Préparation de l'étude de cas	2.6 Conception de systèmes de production et du déroulement des travaux	3.6 Mise à l'essai du processus de production statistique		5.6 Calcul des coefficients de pondération			
		3.7 Finalisation du système de production		5.7 Calcul des agrégats			
				5.8 Finalisation des fichiers de données			

## 0.3. Méthode d'échantillonnage (1/10)

- Le choix d'un plan de sondage optimal est une phase capitale. En effet, il aide à déterminer:
  - la taille de l'échantillon
  - les procédures de tirage
  - les estimateurs calculés, ainsi que leur précision théorique
- Il existe principalement trois types de plans de sondage :
  - Le plan de sondage par liste
  - Le plan de sondage aréolaire
  - Le plan de sondage à bases multiples
- *Pour choisir un plan de sondage, il faut toujours prendre en compte la précision des données souhaitée et les ressources disponibles.*

## 0.3. Méthode d'échantillonnage (2/10)

- La méthode d'échantillonnage est établie à partir d'une base de sondage. Cependant, dans le cas où la base n'est pas fiable, on adopte généralement un plan à au moins 2 degrés:
  - Au premier degré on tire un échantillon d'unités primaires (UP) (village, section d'énumération, zone de dénombrement...) préférablement selon des strates liées aux variables d'intérêt à mesurer (par ex, zone agroécologique, zone à forte, moyenne ou faible densité agricole...) et contenant les unités d'enquêtes. On établit ensuite à l'intérieur de chaque UP échantillonnée la liste des unités d'enquête;
  - Au second degré, on tire l'échantillon d'unités secondaires (US) auprès duquel se fera l'enquête dans les listes d'UP établies au premier degré.

## 0.3. Méthode d'échantillonnage (3/10)

- Le principe du plan de sondage retenu est d'aboutir à un échantillon où toute unité d'enquête a une chance (probabilité) non nulle d'être retenue dans l'échantillon (échantillon probabiliste).
- Cependant, pour éviter de biaiser les estimations en privilégiant certaines unités d'enquête par rapport à d'autres quant aux chances d'appartenir à l'échantillon,
  - il est parfois préférable de considérer aussi **l'égalité des probabilités** : toutes **les exploitations** de type **semblable** (même strate) **ont la même** (ou presque) **chance d'appartenir à l'échantillon** sera le principe à adopter dans le tirage à 2 degrés retenu.

## 0.3. Méthode d'échantillonnage (4/10)

Faut-il, une fois un échantillon d'exploitations tiré, le conserver indéfiniment ou le renouveler périodiquement pour les enquêtes agricoles ?

- L'idéal est d'enquêter régulièrement les mêmes exploitations. En effet, la précision de l'estimation de l'évolution des grandeurs suivies dans ce cas est beaucoup plus grande.
- Mais en réalité cela présente entre autres des inconvénients :
  - la lassitude des unités enquêtées qui entrainera à terme le refus de collaborer
  - la construction de données d'une campagne à partir des données des campagnes antérieures par l'enquêteur
  - Le changement de la structure des unités enquêtées (ex. ménages) après un certain temps

## 0.3. Méthode d'échantillonnage (5/10)

### La technique de tirage à 2 degrés

- Le tirage des UP est presque toujours fait avec probabilités inégales. Cette probabilité pour chaque UP est :

$$p_i = \frac{P_i}{P} = \text{taille relative de l'unité } i$$

Avec:

$P$  = taille totale des UP

$P_i$  = taille de l'UP  $i$  ( $i = 1, 2, \dots, N$ )

$N$  = nombre total d'unités primaires (UP)

- Ainsi, la probabilité de chaque UP est liée à une variable « taille » qui peut être le nombre d'habitants, le nombre de ménages, le nombre d'exploitations, etc.



## 0.3. Méthode d'échantillonnage (6/10)

### La technique de tirage à 2 degrés

- Le tirage des UP est effectué avec remise ou sans remise;
- Tirage avec remise: il permet à une unité d'échantillonnage d'être sélectionnée plus d'une fois. La probabilité de sélection d'une unité à un tirage quelconque est constante;
- Tirage sans remise: Contraire du tirage précédent. La probabilité de sélection d'une unité à chaque tirage varie et dépend du tirage précédent. Cette procédure est complexe et difficilement applicables dans des enquêtes;
- On peut approcher le sondage avec remise par le sondage sans remise. En effet, ces deux types de tirages diffèrent très peu lorsque la fraction de sondage est faible.

## 0.3. Méthode d'échantillonnage (7/10)

### La technique de tirage à 2 degrés

- Le tirage des exploitations dans les UP échantillonnées se fait suite à un dénombrement exhaustif; Ce dénombrement est déterminant dans l'extrapolation des données et le bon tirage de l'échantillon des exploitations à enquêter;
- Le choix des US peut se faire en constituant des strates (Grandes, Moyennes et Petites exploitations par exemple) pour éviter les biais. À défaut de données suffisantes pour faire cette stratification, il est possible d'utiliser des connaissances empiriques.

## 0.3. Méthode d'échantillonnage (8/10)

### Taille des échantillons

- La taille de l'échantillon dans un sondage à deux degrés se compose de la taille de l'échantillon des UP (ex. : village, zone de dénombrement...) et du nombre d'US (ex. : exploitations) à tirer par UP échantillonnée;
- Le nombre d'US échantillonnées par UP dépend:
  - Du degré de dispersion des US dans les UP eu égard à la variable d'étude
  - De la contribution du second degré dans la précision des estimations

## 0.3. Méthode d'échantillonnage (9/10)

### Taille des échantillons

- Le nombre d'UP échantillonnées est déduit après avoir déterminé la taille totale  $n$  d'unités à enquêter avec:

$$n = \frac{CV(y)^2}{CV(\bar{y})^2} = \frac{CV(y)^2}{d^2}$$

$CV(y)$  = Coefficient de variation de  $y$

$\bar{y}$  = Estimateur de la moyenne

$d$  = Précision relative souhaitée pour la moyenne

- Cette taille  $n$  doit être ajustée par l'effet de sondage ( $Deff$ ) et le taux de non-réponses ( $r$ ) tel que:

$$n' = n \times Deff \times (1 + r)$$

# 0.3. Méthode d'échantillonnage (10/10)

## Taille des échantillons

- Lorsque l'échantillonnage prend en compte une stratification, la taille des strates est déterminée selon l'objectif recherché;
  - Objectif de précision locale:  
On détermine séparément la taille nécessaire de chaque strate en fonction de la précision fixée. La taille totale est alors la somme des différentes tailles.
  - Objectif de précision globale:  
On détermine d'abord la taille  $n$  de l'échantillon. Cette taille est ensuite répartie entre les strates suivant les méthodes ci-dessous:
    - Répartition égale
    - Répartition proportionnelle
    - Répartition optimale de Neyman

## 0.4. Collecte des données (1/2)

- La collecte de données est la phase opérationnelle encore appelée phase de terrain;
- Bien avant cette phase, il est important de procéder à une sensibilisation. Les personnes à sensibiliser sont:
  - Les autorités administratives, locales et coutumières, les leaders d'opinion et de groupes sociaux organisés pour solliciter leur participation à relayer l'information relative aux objectifs de l'étude auprès de leur populations.
  - La population cible pour solliciter sa coopération et sa disponibilité à fournir l'information recherchée et l'informer de la période de mise en œuvre.

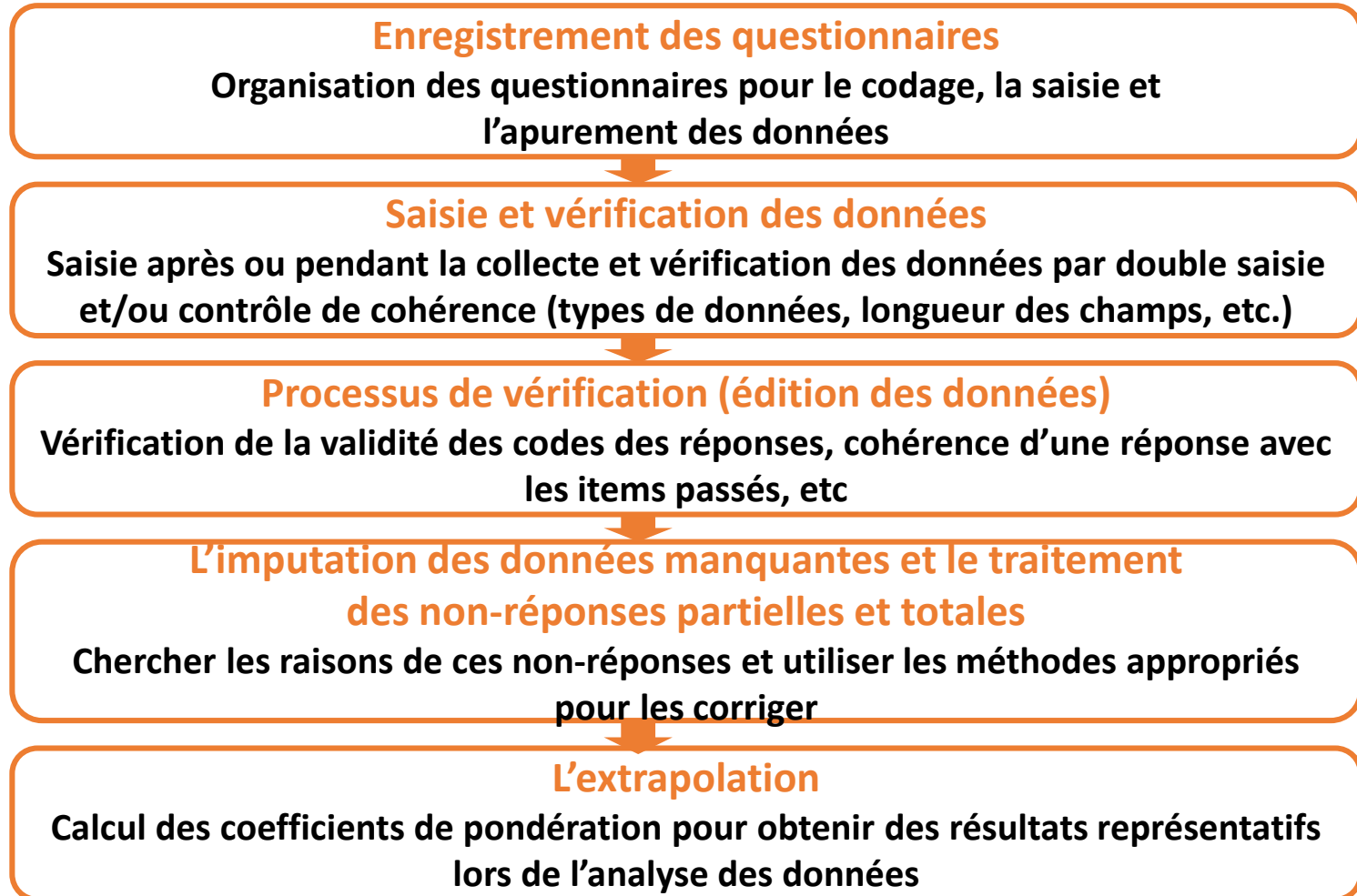
## 0.4. Collecte des données (2/2)

### Points importants pour la réussite de l'opération de collecte

- La chronologie des différentes phases qui doivent obligatoirement être effectuées successivement
- L'identification des unités de sondage
- L'identification des unités d'enquête
- La pédagogie du personnel de collecte (enquêteurs, contrôleurs)
- La garantie de la qualité statistique

# 0.5. Traitement des données (1/7)

- Les principales étapes:





## 0.5. Traitement des données (2/7)

### Les différents types de données manquantes

- **Non-réponse complètement aléatoire (Missing Completely at Random: MCAR):** aucune relation entre l'information manquante et l'ensemble du jeu de données. Elle survient de manière entièrement aléatoire et ne peut s'expliquer.
- **Non-réponse aléatoire (Missing at Random : MAR):** Existence d'une relation entre l'information manquante et une quelconque variable du jeu de données. Ce cas est aussi « non-réponse conditionnellement aléatoire »;  
Ex: Les femmes sont moins enclines à déclarer leur âge et poids comparativement aux hommes.

## 0.5. Traitement des données (3/7)

### Les différents types de données manquantes

- Non-réponse non aléatoire (Missing Not at Random: **MNAR**): Existence d'une relation entre l'absence de l'information et sa valeur. Ce cas est aussi « non-réponse non-ignorable ».  
Ex: Les personnes qui ont un revenu élevé ont une faible propension à les déclarer.

## 0.5. Traitement des données (4/7)

### Méthodes de traitement des données manquantes

- **L'analyse de tous les cas complets** (valable pour les MCAR, MAR): Elle consiste à exclure toutes les unités dont des données ou des résultats sont manquants.
- **L'analyse des cas disponibles** (valable pour les MCAR, MAR): Elle consiste à exclure une variable ou une série de variables à cause de leur taux de données manquantes
- **La pondération en fonction du taux de non-réponse** (valable pour les MCAR, MAR): Elle consiste à trouver le moyen de repondérer l'échantillon pour en rétablir la représentativité

# 0.5. Traitement des données (5/7)

## Méthodes de traitement des données manquantes

- **L'imputation des valeurs manquantes:** Cette méthode consiste à remplacer les données manquantes par des valeurs plausibles. Ci-dessous quelques méthodes d'imputation:
  - **L'imputation par la moyenne.** Dans ce cas, il conviendra de s'assurer que la distribution ne contient pas de valeurs extrêmes (susceptible d'affecter la moyennes)
  - **Le report des dernières valeurs connues** (imputation historique)
  - **L'imputation par ratio-régression**
  - **L'utilisation d'informations déduites d'observations connexes** (imputation déterministe ou déductive). Ceci est valable pour les MNAR;
  - **L'imputation selon le voisin le plus proche.** C'est l'imputation à partir d'un donneur qui possède des caractéristiques similaires Ceci est valable pour les MAR et les MNAR;

## 0.5. Traitement des données (6/7)

### Méthodes de traitement des données manquantes

- **L'imputation « hot-deck »**: Elle consiste à remplacer la valeur manquante par la valeur de l'unité précédente dans le fichier, ou celle de la dernière unité rencontrée de préférence dans la même entité géographique. Cette unité de remplacement doit être suffisamment semblable à l'unité pour laquelle l'information manque. Cette méthode peut être facilement automatisée.
- **L'imputation « cold-deck »**: Elle consiste à utiliser une information extérieure à l'enquête relative à l'unité pour laquelle l'information manque. Par exemple, cette approche est surtout applicable dans les enquêtes par panel.

## 0.5. Traitement des données (7/7)

### Extrapolation: Exemple d'un sondage à deux degrés

- $n$  : Nombre d'unités dans l'échantillon
- $p_i$  : probabilité d'inclusion de l'unité  $i$
- $W_i$  : poids de sondage de l'unité  $i$
- $W$  : poids de sondage total du plan de sondage
- On a :

$$W_i = \frac{1}{p_i} \quad \text{et} \quad W = \prod_{i=1}^n W_i = \prod_{i=1}^n \frac{1}{p_i}$$

## 0.6. Analyse des données (1/5)

- Elle peut se définir comme le processus consistant à transformer les données brutes en informations pertinentes grâce à un raisonnement analytique et logique;
- Cette phase débute par la définition des indicateurs et ainsi que leurs formules de calcul. Mais, l'on constate des différences à ce niveau dans les statistiques agricoles produites;
- Ces différences entraînent des problèmes de comparabilité des données entre les pays et parfois à l'intérieur des régions d'un même pays.

## 0.6. Analyse des données (2/5)

### Raisons de ces différences dans les statistiques agricoles produites

- Différence de conception et de définition en fonction du pays, de ses traditions, de sa culture, de ses pratiques statistiques;
- Différence dans les sources de données pour calculer le même indicateur;
- Différences dans les dénominateurs et les estimations démographiques: utilisation de différentes méthodes de projection;
- Coordination institutionnelle insuffisante entre les acteurs nationaux, régionaux et internationaux d'une part et entre les acteurs nationaux d'autre part;
- Métadonnées (informations sur le contenu des données) incomplètes.



## 0.6. Analyse des données (3/5)

### Tabulation et logiciels statistiques

- La tabulation est processus qui consiste à dériver des nombres d'individus ou de cas correspondant aux combinaisons spécifiées de caractéristiques à partir des enregistrements constituant un ensemble de données.
- Les spécifications des tabulations doivent être compréhensibles à la fois pour les spécialistes et le personnel chargé du traitement des données et être suffisamment détaillées pour que le personnel chargé du traitement des données ne prenne pas de décisions quant au contenu des tabulations.

# 0.6. Analyse des données (4/5)

## Tabulation et logiciels statistiques

- Il existe un grand nombre de logiciels statistiques pour procéder à l'analyse des données dont quelques uns sont listés ci-après:
  - Les logiciels libres: R (une application libre du langage S) et DAP (la version libre du programme SAS);
  - Les programmes dans le domaine public: CPro (Census and Survey Processing System, essentiellement utilisé pour saisir, tabuler, cartographier et diffuser les données d'enquêtes et de recensements), Survey Solution (SuSo) et Epi Info (spécialisé dans l'épidémiologie);
  - Les logiciels gratuits: GeoDa (un logiciel gratuit d'analyse de données spatiales, de géovisualisation, d'autocorrélation spatiale et de modélisation spatiale), QGis (un système d'information géographique à libre accès) et WinBUGS [logiciel d'analyse bayésienne avec la méthode de Monte-Carlo par chaînes de Markov (MCMC)];

## 0.6. Analyse des données (5/5)

### Tabulation et logiciels statistiques

- Les logiciels commerciaux:
  - **EViews** (logiciel d'analyse économétrique),
  - **Stata** (logiciel statistique global),
  - **SAS** (logiciel statistique global),
  - **S-PLUS** (logiciel statistique global) et
  - **SPSS** (logiciel statistique global).

## 0.7. Diffusion des données

- Cette phase renvoie à tous les moyens par lesquels les données sont rendues publiques tels que:
  - La publication de documents, notamment les communiqués de presse, les périodiques et des documents ponctuels hors-série
  - La diffusion électronique de statistiques – par exemple sur cédérom (CD-ROM), sur clé USB ou via Internet
  - L'envoi de statistiques en version imprimée ou électronique en réponse à des demandes directes
  - La mise en place de systèmes automatisés pour donner accès aux statistiques sur demande par téléphone ou par Internet
- Pour s'assurer que les données soient utilisées à leur plein potentiel, il est important de consulter les utilisateurs afin de déterminer le mode de diffusion le plus adapté à leurs besoins.

## 0.8. La gestion de la qualité des données (1/6)

- La qualité doit être recherchée dans le système de production (aspects institutionnels, ressources humaine, matérielle et financière) et dans les produits (outils, méthodologie, opérations);
- Elle doit être une partie intégrante des activités statistiques à travers des contrôles internes et externes à toutes les étapes du processus de programmation et de production statistique;
- Cette recherche de la qualité est une exigence pour la comparabilité internationale. Mais elle doit être bien planifiée dans les activités courantes de gestion et de production des statistiques officielles afin de réduire les coûts liées à celle-ci.

# 0.8. La gestion de la qualité des données (2/6)

## Les conditions préalables de la qualité des données :

- Cadre juridique et institutionnel favorable
- Ressources humaines, financières et matérielles adaptées aux programmes de productions des statistiques agricoles
- Admettre que :
  - Les statistiques contiennent de l'information pertinente pour le domaine;
  - La qualité est une condition qui gouverne tout travail de production de statistiques.

## 0.8. La gestion de la qualité des données (3/6)

- Plusieurs références internationales existent en matière de gestion de la qualité des données. Elles ont été formulées par les organismes tels que:
  - Le Fonds monétaire international (FMI) avec Le Système général de diffusion des données (SGDD), la Norme spéciale de diffusion des données (NSDD) et le Cadre d'évaluation de la qualité des données (CEQD);
  - La Division statistique des Nations Unies avec le Cadre national d'assurance de la qualité (CNAQ)
  - Paris21 avec le questionnaire sur les Indicateurs de renforcement des capacités statistiques (IRCS)

# 0.8. La gestion de la qualité des données (4/6)

- ..... Elles ont été formulés par les organismes tels que:
  - Les instituts de statistique : Statistique Canada, Eurostat, Insee, etc.
  - La FAO avec la mise en place d'un système d'enquêtes intégré, la base de données intégrée et l'émergence des nouvelles technologies (PDA, GPS, télédétection)
- La plupart des cadres internationaux sur la qualité des données s'inspirent du cadre développé par le FMI: le *cadre d'évaluation de la qualité des données* ou CEQD (DQAF en anglais, *Data quality assessment framework*).



## 0.8. La gestion de la qualité des données (5/6)

Les dimensions retenues par les organismes internationaux pour définir la qualité des statistiques officielles:

- **La pertinence** de l'information exprimant comment l'information répond aux besoins réels des utilisateurs
- **L'exactitude et la fiabilité** de l'information statistique exprimant la mesure dans laquelle l'information décrit correctement le phénomène qu'elle devrait évaluer
- **L'actualité et la ponctualité** de l'information statistique qui tient compte de sa date de publication par rapport à sa date de référence

## 0.8. La gestion de la qualité des données (6/6)

Les dimensions retenues par les organismes internationaux pour définir la qualité des statistiques officielles:

- **L'accessibilité et la clarté** qui renvoie à la facilité avec laquelle on peut obtenir l'information auprès du producteur
- **L'interprétabilité ou les métadonnées** se caractérisant par la disponibilité d'informations supplémentaires nécessaires à son interprétation (métadonnées)
- **La cohérence et la comparabilité** de l'information statistique qui est assurée lorsque l'on peut réussir à recouper cette information avec d'autres renseignements statistiques dans un cadre analytique général

*Figure 2 : Cadre d'évaluation de la qualité des données*

FAO	UNSD	UNECE	OECD	EUROSTAT	FMI
La pertinence	La pertinence	La pertinence	La pertinence	La pertinence	Condition préalable à la qualité
					Méthodologie rigoureuse
L'exactitude et la fiabilité	L'exactitude et la fiabilité	L'exactitude	L'exactitude	L'exactitude	L'exactitude et la fiabilité
L'actualité et la ponctualité	L'actualité et la ponctualité	L'actualité et la ponctualité	L'actualité	L'actualité et la ponctualité	Utile : Disponible à temps pour la prise de décision
L'accessibilité et la clarté	L'accessibilité et la clarté	L'accessibilité et la clarté	L'accessibilité et l'interprétabilité	L'accessibilité et la clarté	L'accessibilité
	Les métadonnées				Assurance de l'intégrité
La cohérence et la comparabilité	La cohérence et la comparabilité	La comparabilité	La cohérence	La cohérence et la comparabilité	Utile : Disponible à temps pour la prise de décision
		La pertinence au niveau institutionnel	La crédibilité		Condition préalable à la qualité
					Prérequis de l'intégrité

# Exercices

- **Exercice 1 : Définition de la population**
- **Exercice 2 : Population échantillonnée vs population cible**

MERCI