

## **COURS DE THEORIES ET METHODES DE SONDAGE**

PROF DR. SISI KAYAN

La pratique des sondages s'est aujourd'hui étendue à tous les aspects de la vie. L'essentiel de nos connaissances, de la recherche scientifique, des politiques et de la planification, voire des décisions, actions et attitudes individuelles dans la vie de tous les jours, est dans une large mesure façonné et orienté par des sondages. Les décisions d'achat, le diagnostic et le traitement de maladies, les opinions sur les lieux, les peuples ou les institutions reposent sur des contacts et expériences limités, voire uniques, qui sont néanmoins considérés comme représentatifs de l'ensemble du phénomène considéré. Il n'est pas rare que des auteurs écrivent des livres ou des articles sur des lieux ou même des pays aussi grands que l'Inde, après n'y avoir séjourné que quelques jours. Ces écrits se fondent sur des observations et expériences limitées, qui font alors fonction d'échantillon.

Le phénomène des sondages s'est généralisé dans le monde moderne. L'acheteur d'un lot de céréales n'examine qu'une poignée de grains pour se faire une idée de la qualité de l'ensemble; le cuisinier ou la ménagère ne goûtent que quelques bouchées (ou cuillerées) de leurs préparations pour juger du goût de tout le plat, et il suffit de quelques gouttes de sang pour analyser tout le sang que contient le corps d'un individu. Cette omniprésence du recours aux sondages laisse à penser que ceux-ci pourraient avoir la même efficacité dans le domaine des statistiques de l'éducation que dans d'autres activités humaines ou dans les travaux universitaires.

La validité des conclusions tirées d'un sondage est néanmoins fonction de la représentativité de l'échantillon. Les déductions citées en exemple plus haut partent implicitement du principe que tous les éléments de la population considérée sont de même nature, du moins pour ce qui est de leurs grandes caractéristiques quantitatives ou qualitatives, ou bien que ces éléments sont suffisamment bien répartis dans la population pour qu'on puisse tirer de l'analyse d'un petit nombre d'éléments, voire d'un seul, autant d'informations que si l'on analysait l'ensemble de la population. Or, cette dernière peut n'être ni uniforme ni homogène, ce qui amène à se poser des questions sur la représentativité de l'échantillon et, partant, sur le bien-fondé de l'application à l'ensemble de la population des conclusions tirées de l'analyse de l'échantillon. Pour leur part, la représentativité de l'échantillon et la validité des conclusions tirées du sondage sont dans une large mesure fonction de la méthode de sondage, de la taille de l'échantillon et des techniques d'enquête et d'analyse.

Si l'on n'exclut de l'échantillon qu'un petit nombre d'éléments de la population, les estimations établies à partir de l'échantillon seront pratiquement identiques aux statistiques de la population tout entière, l'échantillon étant dans ce cas quasiment de même taille que la population. Lorsque, par contre, l'échantillon ne compte qu'un ou deux éléments de la population, il risque de ne pas être représentatif, sauf si la population est parfaitement homogène.

Dans l'optique de la représentativité, la méthode d'échantillonnage est aussi importante que la taille de l'échantillon. Il est communément admis que l'échantillonnage

aléatoire simple permet de sélectionner un échantillon véritablement représentatif, alors qu'un échantillon tiré à l'aveuglette peut s'avérer atypique, quelle qu'en soit la taille. Par ailleurs, les conclusions tirées de l'échantillon ne sont pertinentes et correctes que dans la mesure où les techniques d'analyse employées sont celles qui conviennent.

La taille de l'échantillon, la méthode d'échantillonnage et les techniques d'inférence des conclusions à partir de l'échantillon jouent donc un rôle capital tant dans la théorie que dans le pratique des sondages. Aussi a-t-on prêté une attention particulière à ces trois aspects dans la réalisation de l'enquête pilote.

Avantage de l'enquête par sondage : Substituer le sondage à la méthode actuelle de production des données (par recensement) ne peut se justifier que si, d'un point de vue tant analytique que pratique, le sondage présente des avantages certains par rapport au recensement. Tel est effectivement le cas, et ces avantages sont considérables lorsque la population est nombreuse. On peut considérer comme particulièrement importants, en théorie comme en pratique, les avantages ci-après que présentent les sondages par rapport aux recensements.

### **Objectif**

L'objectif de ce cours est d'introduire les principaux concepts de la théorie des sondages en soulignant la nature particulière de l'aléa et l'importance des calculs de précision, et de présenter les plans de sondage les plus classiques. Le rôle de l'information auxiliaire sera mis en exergue, que ce soit lors de la phase d'échantillonnage ou lors de l'estimation. Ce cours sera illustré par des exemples d'enquêtes, et les procédures de sondages de SAS seront brièvement présentées.

### **Principaux acquis de la formation :**

A l'issue du cours, l'étudiant saura

- Définir un plan de sondage (méthode et algorithme) adapté au problème posé ;
- Calculer la précision des estimateurs (moyenne, total, proportion) issus d'une enquête par sondage;
- Utiliser des variables auxiliaires pour améliorer la qualité des estimateurs;
- Repondérer les données pour corriger de la non-réponse totale et mettre en place des stratégies d'imputation pour corriger de la non-réponse partielle.

## **I. GENERALITES : Enquête et sondage**

### **I.1. L'enquête**

#### **I.1.1. Définition**

L'enquête sociologique est très fréquemment utilisée dans les sciences humaines et souvent par les gestionnaires. Le sondage électoral et les études de marché sont les formes d'enquête les plus répandues.

Mais comment définir une enquête ? Et plus particulièrement une enquête commerciale ?

Dans le langage courant le mot enquête ne ramène pas nécessairement à une démarche méthodologique de recherche. Il signifie le plus souvent : quête d'informations, collecte de témoignages, recherche pour savoir quelque chose. Mais il convient de distinguer l'enquête sociologique des autres types d'enquête (enquête administrative, juridique, de police ou encore journalistique).

L'enquête sociologique est une collecte d'informations sur une population donnée. Sa principale caractéristique est de recourir pour cette collecte de données à l'interrogation systématique de sujets de cette population pour décrire, comparer ou expliquer. Aussi, elle relève d'une démarche scientifique.

C'est pour cette raison qu'une enquête en sciences sociales ou commerciales doit être conduite avec la plus grande rigueur méthodologique pour réduire le plus possible la subjectivité de celui qui la réalise.

La mise en œuvre d'une enquête suppose avant tout un questionnaire bien pensé; strictement préparé autour de la construction d'une problématique. Par ailleurs, dans la collecte des informations une attention particulière doit être accordée au mode de recueil de l'information et au choix de l'échantillon. Au niveau de l'analyse des résultats, on va utiliser la quantification, et des lois statistiques.

L'utilisation des renseignements fournis par les sujets sur eux-mêmes, peut introduire un biais dans les conclusions que l'on pourrait tirer des résultats de l'enquête. En effet, le sujet peut répondre en exagérant ou en minimisant sa situation ou encore en répondant au hasard. Aussi, il faudra veiller à obtenir des réponses aussi fiables possibles.

Quand va-t-on recourir à l'enquête sociologique ?

En sciences humaines, il existe des méthodes concurrentes de la technique d'enquête pour obtenir des données sur une population.

Mais si nous avons pour objectif de comprendre la façon dont les individus se représentent un phénomène, de connaître leur perception des choses, leurs valeurs, et leurs normes, l'enquête sociologique (et par conséquent le sondage) est la plus adaptée. Autrement le chercheur a à sa disposition des outils comme l'étude documentaire, l'observation directe, l'entretien non-directif ou l'expérimentation.

Imaginons que l'on vous demande de mener une enquête, quelle serait selon vous la première étape dans la réalisation de cette enquête ?

La première chose à faire est de définir votre objectif général. Que voulez savoir (quel est le sujet que vous souhaitez étudier) et sur quelle population doit porter cette étude ? A quoi va servir cette enquête ?

Cette formalisation des objectifs de l'étude doit s'accompagner d'une formalisation des contraintes de l'étude. Il s'agit de mettre en évidence sa faisabilité pratique (aspect financier, calendrier prévisionnel, collaborations possibles, matériel disponible, possibilités d'accès aux personnes à interroger).

La rigueur garante de l'objectivité d'une enquête demande une préparation et une organisation très précise de celle-ci. Le respect de la démarche scientifique requiert le suivi de la chronologie des huit grandes étapes du déroulement d'une enquête.

### **I.I.2. Etapes d'une enquête**

- Étape 1 : La définition de l'objectif général
- Étape 2 : La définition des objectifs spécifiques
- Étape 3 : L'élaboration du plan d'observation
- Étape 4 : La préparation de l'instrument d'observation
- Étape 5 : Le recueil de l'information
- Étape 6 : La préparation des données
- Étape 7 : L'analyse des résultats
- Étape 8 : La rédaction du rapport final

Attardons nous quelques instants sur chacune de ces étapes :

#### ***Étape 1 : La définition de l'objectif général***

Nous avons vu tout à l'heure que l'objectif général devait être vu comme l'énoncé du problème qui nécessite le recours à l'enquête. On peut par exemple se demander s'il y a unité ou diversité des comportements de lecture des jeunes?

#### ***Étape 2 : La définition des objectifs spécifiques***

Une fois la question de départ posée, il s'agit de procéder à sa décomposition en objectifs spécifiques ou questions plus limitées. Cette formulation en objectifs spécifiques revient à se poser une série de questions de la forme : tel facteur a-t-il de l'effet sur tel phénomène ? Ils peuvent se formuler sous forme d'hypothèses que l'on va chercher à vérifier.

#### ***Étape 3 : L'élaboration du plan d'observation***

Ce plan d'observation est la façon d'organiser le déroulement de l'enquête. Les objectifs de l'enquête et les contraintes pratiques vont être déterminantes pour l'élaboration de ce plan. À cette étape un certain nombre de choix sont opérés : quelle population interroger ? Selon quels critères est-elle définie ? Qui sera soumis à cette enquête ? (La population entière ou un échantillon ?) Quel type de consultation faire ? (Des entretiens individuels ou un

questionnaire standardisé ?). S'agit-il d'une description d'une population à un moment donné ou faudra-t-il suivre son évolution dans le temps ?

#### ***Étape 4 : La préparation de l'instrument d'observation***

A ce stade du déroulement de l'enquête, l'échantillonnage et l'élaboration du questionnaire se font. Les questions à poser seront fonction des objectifs spécifiques de l'enquête. Plus ces derniers seront nombreux, et plus le questionnaire sera long. Ce questionnaire devra être mis en test réel pour s'assurer de son bon fonctionnement. (Bonne compréhension des questions, bon fonctionnement des filtres, liste des modalités de réponses pertinentes, ordre des questions logiques ...)

#### ***Étape 5 : Le recueil de l'information***

Les sujets sont interrogés à cette étape sur le terrain et selon la modalité choisie pour l'enquête

#### ***Étape 6 : La préparation des données***

Les données collectées ne peuvent être utilisées telles quelles, il sera donc nécessaire de les présenter sous une forme qui permette l'analyse prévue. Les données seront saisies en informatique suivant un codage précis en prévision d'analyses statistiques.

#### ***Étape 7 : L'analyse des résultats***

Les informations sont traitées en fonction de la nature des données recueillies et des objectifs de l'étude (description, comparaison ou vérification d'hypothèses). À ce stade il faut se préoccuper de la qualité des données obtenues. Trois niveaux d'analyse de plus en plus complexes sont possibles : les tris à plat (examen d'une seule variable à la fois), les tris croisés (deux variables prises en compte simultanément), l'analyse multivariée (plusieurs variables prises en compte simultanément.)

#### ***Étape 8 : La rédaction du rapport final***

Ce rapport final décrit à la fois les objectifs, la méthodologie, les résultats et leurs interprétations.

Comme vous pouvez le voir, ces étapes s'inscrivent dans une suite chronologique. L'enquête ne débute donc pas par la construction du questionnaire (étape n°4). Par ailleurs, l'étape du recueil de l'information est un point de non retour car, une fois la collecte terminée, vous ne pourrez plus repartir sur le terrain si telle ou telle information venait à vous manquer.

Autrement, les étapes 2, 3 et 4 ne peuvent être traitées de façon indépendante. Les objectifs spécifiques déterminent le plan d'observation mais ce dernier peut nous amener à revoir les objectifs spécifiques.

### **I.I.3. Le questionnaire d'une enquête**

Le questionnaire est l'instrument de base de l'enquête sociologique. Il est construit à partir des buts de recherche que l'on s'est fixés, il doit comporter l'ensemble des variables qui permettront au chercheur d'obtenir les informations qui lui seront nécessaires pour bâtir son analyse. Les principales hypothèses définies dans l'étape n°2 (objectifs spécifiques)

seront vérifiées par ces questions. La stratégie de construction d'un questionnaire est relativement simple.

***Il faut cependant s'assurer de quatre choses :***

- que les enquêtés (sondés) comprennent les questions
- qu'ils sont capables d'y répondre
- qu'ils acceptent d'y répondre
- et que la réponse est formulée de façon authentique et non-influencée.

Pour ce faire, le chercheur doit se mettre dans la peau de ceux à qui le questionnaire sera administré et prévoir leurs réactions aux questions. C'est le meilleur moyen de concevoir des questions favorisant l'obtention de l'information souhaitée.

Il est préférable, autant que faire ce peut, de revoir les questions déjà produites lors d'enquêtes antérieures sur le même thème de recherche ou sur des thèmes proches du sien.

***Il convient de réaliser un pré-test de ce questionnaire.***

Avant d'aller sur le terrain avec votre sondage, le pré-test est une opération qui consiste à soumettre le questionnaire à un nombre restreint de personnes susceptibles d'être interrogées dans le cadre de votre grande enquête. Son principal intérêt est de vous permettre de d'identifier les points faibles du questionnaire : mauvaise formulation des questions, insuffisance des informations recueillies. La détection de ces problèmes en amont de la phase de terrain permet de limiter les problèmes d'interprétation des résultats.

Globalement, la formulation des questions, le choix des questions (ouvertes ou fermées), ainsi que la forme et le choix des réponses proposées seront élaborés en fonction de trois critères : la clarté (compréhension), la neutralité (authenticité des réponses) et l'adéquation qui nous renvoie à la capacité des interviewés à répondre au questionnaire.

Je vous propose de préciser les choses sur les deux premiers critères :

**La clarté :** les questions doivent être formulées dans un vocabulaire simple compréhensible par le plus grand nombre. Les concepts techniques doivent être déclinés le plus simplement possible. La clarté s'obtient également par la formulation de questions courtes

**La neutralité :** Cette notion est importante dans une enquête ou un sondage. L'information à recueillir doit être demandée (via le questionnaire) de façon objective. Il s'agit ici de s'assurer de l'authenticité des réponses. Un questionnaire est considéré comme étant neutre quand il préjugera le moins possible des réponses éventuelles. Un choix assez large de possibilités de réponses devra être proposé aux sondés.

Exemple : une présélection d'items et une réponse "autre" dans laquelle on laisse le sondé préciser sa réponse.

La satisfaction de ces trois critères limite le nombre de refus de réponse et par conséquent les problèmes de pertinence des résultats obtenus.

La préparation matérielle du questionnaire se concrétise aussi par des efforts au niveau de la charpente de celui-ci. Trois éléments sont importants à ce niveau : la longueur du questionnaire, l'ordre et l'orientation des questions.

### ***La longueur du questionnaire.***

Globalement, il faut veiller à ce que les questionnaires auto-administrés (c'est-à-dire remplis par l'interviewé sans la présence d'un enquêteur) soient les plus courts possible, il en est de même pour certains questionnaires administrés par téléphone.

### ***L'ordre des questions.***

Un questionnaire comporte un certain nombre de sections qui correspondent chacune à une variable ou un groupe de variables. Il est donc nécessaire, pour s'assurer d'une bonne participation du sondé, de mettre des liaisons entre les différentes sections.

En règle générale, l'articulation retenue par les concepteurs de questionnaires est la suivante :

1) Une ou plusieurs questions pour s'assurer que l'interviewé fait partie de la population sur laquelle on fait notre recherche.

2) La première section doit comporter des questions faciles à répondre pour le mettre en confiance (il convient de soigner la formulation de ces questions). À l'intérieur de cette section, les questions devront être simples et attrayantes tandis que les questions les plus délicates y seront placées à la fin.

3) Les passages d'une section à l'autre devront être marqués par une courte phrase de transition informant l'interviewé de l'orientation du questionnaire.

4) En fin de sondage, on retrouvera les caractéristiques socio-économiques en fin de sondage.

Globalement, les questions générales précèdent les questions spécifiques.

### ***L'orientation du questionnaire.***

Il faut savoir qu'en matière de questionnaire il existe ce qu'on appelle un "biais de positivité" c'est à dire que toutes choses étant égales par ailleurs, les individus ont tendance à répondre plus facilement oui que non et plus facilement d'accord que pas d'accord.

Aussi il faut veiller à limiter l'effet de ce biais.

Juste une précision, par question fermée il faut entendre une question qui laisse le choix à l'interviewé entre deux modalités de réponses (oui/non) ou plus (choix multiples : un peu/ beaucoup/ pas du tout).

Une question ouverte laisse au sondé le choix d'utiliser son propre vocabulaire pour répondre à la question (Quel est votre âge ?).

Une question filtre est une question qui permet selon la réponse de choisir la question qui sera abordée ensuite.

### ***En résumé un bon questionnaire doit :***

- Être simple avec une présentation claire et attrayante (faire attention à la disposition et à la forme).

- Avoir des questions simples c'est à dire :

- courtes (phrases interrogatives n'excédant pas 20 mots)

- neutres (pas de questions tendancieuses ou biaisées)

- sans ambiguïté : utiliser un vocabulaire simple et adapté à la forme de recueil de l'information (écrit ou oral), éviter les mots à plusieurs sens et les formes grammaticales peu claires.

- Directes ou indirectes

- Il faut soigner les réponses proposées : OUI / NON / Ne sait pas

***Par ailleurs il convient de faire attention :***

- à l'ordre des questions

- aux réponses fourre-tout ou induites (questions dans laquelle il y a déjà la réponse)

- aux questions ouvertes.

### **Test du questionnaire.**

Comme je vous l'ai indiqué précédemment, il faut tester ce questionnaire avant de le lancer, et penser à son exploitation au moment de sa rédaction. En phase de test du questionnaire, une série de questions peut vous permettre d'évaluer votre questionnaire en vue de son amélioration.

Les questions sont-elles comprises ? Les enquêtés ont-ils du mal à répondre? Donnent-ils plusieurs réponses à une même question ? Écrivent-ils des commentaires dans la marge ?

Quelles questions passent mal et provoquent la gêne des enquêtés ? Ces questions sont-elles par exemple trop personnelles et entraînent-elles des refus de répondre, des rires ou d'autres manifestations de malaise ?

L'information demandée est-elle appropriée aux répondants ? N'a-t-on pas oublié d'introduire des filtres ? Les renvois prévus conviennent-ils ?

Le vocabulaire est-il adapté ? Y a-t-il des mots ou expressions qui nécessitent une explication ? L'enquêteur suit-il exactement le texte prévu pour la question ou est-il obligé de l'aménager ?

Les listes de modalités de réponses sont-elles pertinentes et exhaustives ? Les enquêtés en ont-ils ajouté ? La modalité "autre" a-t-elle reçu beaucoup de réponses ?

Les questions produisent-elles des variations de réponses suffisantes ? N'y a-t-il pas de questions qui font l'unanimité sur une seule réponse (et qui seraient donc inutiles.)

L'ordre des questions paraît-il logique ? N'y a-t-il pas des effets liés à l'ordre des questions ?



Y a-t-il des transitions, enchaînements, liens entre les questions ? Les textes de présentation ou de transition conviennent-ils ?

Des questions difficiles ne doivent-elles pas être positionnées plus loin dans le questionnaire ?

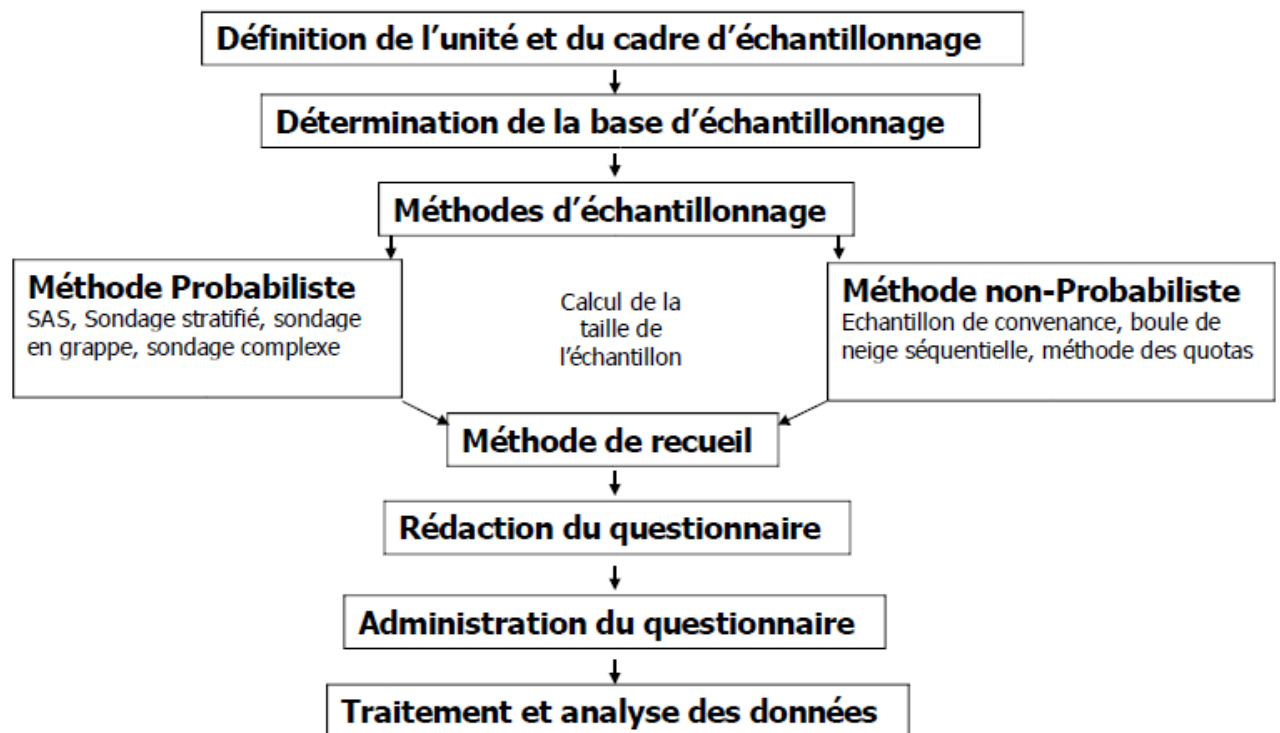
A-t-on indiqué aux enquêtés de quelle façon répondre (cocher, entourer, barrer, un choix, plusieurs réponses possibles, etc.) Les enquêtés comprennent-ils comment répondre aux questions ? Se conforment-ils aux consignes ?

Les enquêteurs (dans une enquête de face à face) comprennent-ils comment poser les questions et noter les réponses ? Les consignes pour l'enquêteur sont-elles claires et suffisantes.

En outre on peut, pour améliorer le questionnaire :

- Essayer plusieurs formes de questions pour choisir celle qui paraît la meilleure
- Fermer les questions ouvertes
- Évaluer la durée de l'entretien
- Relever les erreurs typographiques ou de mise en page.

## I.2. Le sondage



### I.2.1. Définition et notions

Au sens propre, une enquête est une recherche d'informations : Ex. enquêtes policières, journalistiques, parlementaires... Il existe aussi des enquêtes statistiques dont le but est le recueil de renseignement chiffrés et chiffrables. Certaines enquêtes sont exhaustives, ce sont les recensements (par exemple le recensement de la population).

Le sondage est une technique statistique qui permet de réaliser certaines enquêtes dans des conditions contrôlées, sur des sous-ensembles de la population appelés échantillons, et désignés soit par hasard, soit en fonction de caractéristiques particulières.

**C'est une recherche sur un échantillon de sujets qui se déroule dans la vie quotidienne avec des moyens standardisés de questions dont le but est de dégager une mesure quantitative. Les questions sont les mêmes pour tous les sujets. Ceci signifie en fait forcer la réalité, parce que tous ne comprennent les questions de la même façon. C'est ici que se trouve la faiblesse et le risque des méthodes standardisées du sondage.**

Le champ de l'enquête est **la population** (ou l'univers). Par exemple l'ensemble de ménages vivant à Kinshasa ou dans un quartier de Kinshasa ; l'ensemble des individus de 20 à 45 ans, l'ensemble des électeurs ...

Les éléments d'ensemble sont des **unités d'observation**, ou **individus**.

La **base de sondage** est la liste des unités, si possible exhaustive et sans double comptes. Une bonne base de sondage couvre au mieux l'univers de référence, elle comporte

en outre des informations permettant de contrôler le choix de l'échantillon et la production des résultats.

**L'échantillonnage** est la sélection des unités.

**Les variables d'intérêts** constituent l'objet de l'enquête. Elles peuvent être quantitatives (ex. les dépenses) ou qualitatives (ex. une opinion).

Le sondage ne s'intéresse pas aux individus en tant que tels, mais à des caractéristiques d'ensemble, **les paramètres** à estimer, par exemple une moyenne de revenus, une proportion d'électeurs votant pour tel candidat...

Un échantillon est **représentatif** s'il permet d'estimer les paramètres étudiés avec une précision acceptable, étant donnés les objectifs de l'enquête.

Il faut savoir que le sondage est une dimension un peu plus restreinte de l'enquête sociologique. La définition courante du sondage est la suivante : **le sondage est une enquête menée auprès d'un échantillon de personnes qui sont considérées comme représentatives d'un ensemble social sur lequel on souhaite recueillir des informations ou connaître l'opinion.**

On peut également dire que le sondage en sciences humaines est un instrument de recherche, de mesure, de structuration et de présentation de l'information fondée sur l'observation de réponses à un ensemble de questions posées à un échantillon d'une population.

De ces définitions on retient deux choses :

- le **choix de l'échantillon** est donc fondamental dans le cas d'un sondage d'opinion
- le principal objet du sondage est donc **d'obtenir des informations permettant de répondre à des questions et de vérifier les hypothèses** d'une recherche.

Par ailleurs, il faut savoir que **les résultats d'un sondage se traduisent par des chiffres** notamment par une série de pourcentages, contrairement à l'enquête pour laquelle l'analyse est plus approfondie.

Tout comme l'enquête, le sondage utilise des techniques quantitatives. **Les résultats** obtenus auprès de l'échantillon représentatif choisi seront **extrapolés** à l'ensemble de cette population qu'elle représente.

Il existe différentes catégories de sondages : les sondages ponctuels, de tendance et le panel.

• **Le sondage ponctuel** (ou à coupe transversale) est réalisé une fois. Il sert à décrire certaines caractéristiques d'une population.

• **Le sondage de tendance** est réalisé à plusieurs reprises. Il permet d'observer dans le temps l'évolution globale de certaines caractéristiques de la population. Il est basé sur la fidélité aux mêmes notions et au même questionnaire.

• **Le sondage panel** est administré plusieurs fois au même échantillon d'individus. Il permet d'étudier le changement au niveau individuel, à l'inverse du sondage de tendance qui étudie ce changement global.

Les principaux **avantages** du sondage sont la **flexibilité**, la **polyvalence**, et la **rapidité** d'utilisation.

Ses principaux **dangers** sont : le risque de **généralisation abusive** de ses résultats et le **mauvais échantillonnage**.

Pour s'assurer la bonne qualité d'un sondage, deux précautions sont à prendre : la méthode de **choix de l'échantillon et la taille de celui-ci**.

## I.2.2. Histoire des sondages

### *Bref profil historique:*

Le premier sondage important semble être "L'enquête ouvrière" de Karl Marx (1880), qui est resté non conclu.

En 1891, Max Weber fait une étude sur les conditions de travail dans le monde rural.

En réalité, le début systématique des sondages remonte à la doit disant « sociologie scientifique » de Quetelet et Le Play et aux intérêts du réformisme social » en Angleterre à la fin du XIX<sup>ème</sup> siècle (Booth).

Mais les antécédents les plus proches se trouvent aux Etats Unis dans les années 30 et 40, avec le développement de la recherche du marché, appliquée tout de suite à l'étude des attitudes électoraux.

Ainsi, Gallup, Crosley, Roper et Lazarsfeld réussissent à prévoir (prédire) le choix du président Roosevelt (1936) à partir d'un échantillon de 4.000/5.000 sujets.

Par contre la revue "Literary Digest", avec la méthode "Vote de paille", sur un échantillon de 2.400.000 personnes, se trompa.

Ce succès du premier sondage statistique attira l'attention de la recherche académique universitaire. Et Lazarsfeld fonda dans l'Université de Columbia (USA) le "Bureau of Applied Social Research" à la fin des années 30.

L'Institut de sociologie appliquée fut la première institution de recherche sociale empirique sur la communication et le modèle de toute une génération d'études des médias ».

Le statut des sondages comme discipline scientifique est de reconnaissance récente. Leur émergence dans l'histoire de la statistique a été décrite et analysée dans les travaux de Droysbeke-Fichet-Tassi (1987) et Droysbeke-Tassi (1990). Le besoin d'informations chiffrées remonte à plusieurs millénaires. On trouve les traces de recensements dans la civilisation sumérienne (5000 à 2000 acn), en Mésopotamie, Egypte, Chine, chez les Incas. C'est une manière pour les chefs d'Etat d'évaluer leur puissance : population, potentiel militaire, revenus... En France, des dispositifs généralisés de collecte de statistiques administratives se mettent en place au 13<sup>ème</sup> siècle.

C'est au 18<sup>ème</sup> siècle que l'école d'arithmétique politique, fondée par les Anglais John Graunt et William Petty suscite des enquêtes partielles pour évaluer les multiplicateurs. Le multiplicateur de naissance par exemple permet d'estimer la taille d'une population d'après la ratio, supposé constant, entre le nombre de naissances et la population d'une année donnée. En 1895, le norvégien Kiaer présenta à l'Institut international de statistique les fondements méthodologiques des « dénombremments représentatifs ». Les idées de Kiaer seront appuyées par les savants tels que Bowley, Von Borkiewicz, Tchuprow et reçoivent leur consécration en 1925 dans un rapport à l'Institut International des Statistiques (IIS) présenté par le Danois Jensen. Jerzy Neyman (1934 : sondages à deux degrés), Horvitz et Thompson (1952 : sondages à probabilités inégales), Dalenius (1957 : non-réponses). La pratique des sondages d'opinion naît au 19<sup>ème</sup> siècle aux Etats Unis avec les « votes de

paille » organisés par les journaux Harrisburg Pennsylvanian et Raleigh Star auprès de leurs lecteurs pour prédire les résultats des élections. La méthode des quotas connaît son jour de gloire le 3 novembre 1936, jour de l'élection de F.D.Roosevelt, prévue par 3 sondages par quotas réalisés indépendamment par Crossley, Roger et Gallup, alors que le vote de paille du Literary Digest auprès de 10 millions de personnes prédisait la victoire de Landon.

Jean Stoetzel importa la méthode en France et fonda l'IFOP en 1938, premier institut d'étude de l'opinion publique.

Dans le domaine de l'information économique et sociale, l'Institut national de la statistique et des études économiques (INSEE) est créé en 1946, faisant suite à la Statistique générale de la France (où s'illustra Alfred Sauvy) et le Service national de la statistique (créé en 1942 par René Carmille, mort en déportation). La tâche de l'INSEE était immense : bâtir le système d'information économique et sociale de la France en reconstruction : population, consommation, appareil productif. L'ampleur de la mission nécessitait le recours aux enquêtes par sondage, et le statisticien Pierre Thionet mit en place le dispositif dont les principes essentiels sont encore en vigueur, au même moment qu'il contribua à la formation des cadres statisticiens. Il écrivit le premier manuel français de sondage en 1946, dès la création de l'INSEE, pour son école d'application, devenue plus tard l'ENSAE.

### **I.2.3. Populations et échantillons**

Nous avons vu qu'une population était un ensemble d'individus, ou unités statistiques.

Un échantillon d'individus est une partie ou un sous ensemble d'une population plus vaste. Les données dont on dispose pour faire une étude proviennent parfois de la population dans son ensemble, mais le plus souvent ces données proviennent d'un échantillon. En règle générale, il est intéressant d'avoir des données les plus exhaustives possibles sur une population. C'est un moyen d'avoir une représentation, un portrait le plus fiable possible de cette population que l'on étudie. C'est le cas pour le dénombrement de la population d'un pays ou d'une région donnée, de ses naissances et de ses décès, de sa répartition suivant l'âge ou le sexe.

Mais il n'est pas toujours nécessaire de procéder au recueil d'information le plus large possible sur une population. C'est le cas du taux d'équipement des ménages, ou de leurs postes de dépenses de consommation. L'interrogation d'un échantillon est largement suffisante.

Quand faut-il interroger toute une population ou plutôt un échantillon ?

C'est l'importance attribuée à l'obtention d'un portrait exhaustif de la population, sa nécessité et sa faisabilité qui vont déterminer si un échantillon est suffisant ou pas ou si l'image complète de la population est nécessaire. (Exemple : construction d'hôpitaux, enquête de satisfaction/image.)

Chacun de ces choix a des avantages et des inconvénients :

- **La population complète** : son principal intérêt est son caractère exhaustif et donc la précision de l'information. Son principal défaut c'est son coût et sa lourdeur dans sa gestion administrative, les délais très longs.

• **L'échantillon** a pour principal avantage d'être moins coûteux et de donner une image assez fiable grâce aux techniques statistiques de l'ensemble de la population. Son inconvénient majeur c'est que l'erreur d'échantillonnage remet en question la fiabilité de l'image donnée par l'échantillon. Or il arrive que, malgré une grande rigueur dans la constitution de l'échantillon, celui-ci ne reflète pas la population qu'il est censé représenter. Dans ce cas, on parle de distorsion ou de biais entre l'échantillon et sa population d'origine. L'exemple le plus connu de l'enquête pour obtenir portrait exhaustif d'une population donnée est le recensement de la population.

L'exemple le plus connu d'enquête sur un échantillon est le sondage d'opinion.

#### **I.2.4. Comment va-t-on procéder à un échantillonnage ?**

Échantillonner, c'est choisir une partie d'une population pour représenter l'ensemble de la population.

Si vous travaillez sur un échantillon, c'est pour chercher une information sur l'ensemble de sa population mère. Les résultats observés sur cet échantillon n'auront de sens que s'ils sont rapportés à toute la population.

On donc a recours à des techniques bien précises pour construire un échantillon et s'assurer qu'il est représentatif de sa population mère.

Il existe deux grandes catégories de méthodes :

- **les méthodes d'échantillonnage non probabilistes (ou empiriques)**
- **les méthodes d'échantillonnage probabilistes (ou aléatoires)**

Si la population de référence est accessible à partir d'une base de sondage, on utilise des méthodes dites aléatoires (ou probabilistes). Des outils classiques du calcul des probabilités permettent de produire des indications chiffrées sur la précision des résultats. On distingue : le sondage aléatoire simple, le sondage aléatoire stratifié, le sondage par grappes, le sondage à deux degrés.

L'échantillonnage est d'autant plus efficace qu'on sait mobiliser les informations auxiliaires contenues dans la base de sondage pour guider la sélection. S'il n'y a pas de base de sondage, on applique des méthodes dites empiriques ou à choix raisonné. En réalité chaque type d'échantillonnage détermine un type de sondage. Si l'on dispose de données de cadrage de la population, la méthode la plus employée est la méthode des quotas. La qualité d'un échantillon dépend essentiellement de la fiabilité et de la pertinence des données de cadrage.

En sus des limites et des problèmes évoqués précédemment, chaque méthode d'échantillonnage comporte ses propres avantages et inconvénients, et n'est applicable que si certaines conditions précises, variant selon les méthodes, sont réunies. Il convient de peser soigneusement tous ces facteurs et de veiller à ce que les critères d'applicabilité soient bien respectés, faute de quoi les résultats risquent de n'être ni pertinents ni crédibles.

### **Les méthodes d'échantillonnage empiriques.**

• Elles regroupent les techniques où les éléments de l'échantillon sont choisis sur le terrain en fonction de jugements sur les caractéristiques de la population ou bien en reconstituant le hasard. Elles ne nécessitent pas de base de sondage, elles sont rapides, et pas très chères. Les échantillons obtenus à l'aide de ces méthodes sont les suivants :

- les échantillons par quotas ou proportionnel
- les échantillons " accidentels "
- les échantillons de volontaires

• Nous reviendrons un peu plus longuement dans quelques instants sur **la technique d'échantillonnage par quotas** qui est la plus utilisée dans cette catégorie.

### **Les méthodes d'échantillonnage probabilistes.**

Elles regroupent les techniques qui font intervenir le hasard pour désigner les éléments de l'échantillon. Elles nécessitent une base de sondage (liste de tous les individus de la population mère), sont plus coûteuses et longues que les méthodes empiriques. Les échantillons obtenus par ces techniques sont :

- les échantillons aléatoires simples
- les échantillons systématiques
- les échantillons par grappe
- les échantillons stratifiés (proportionnels ou non proportionnels)

Nous reviendrons dans un moment sur **la technique d'échantillonnage aléatoire simple** qui est la technique de base , dans cette catégorie.

Je vous propose d'aller à l'essentiel et de voir les échantillons les plus répandus pour chacune de ces grandes catégories. Les échantillons par quotas pour la première et l'échantillon aléatoire simple pour le second.

### **L'échantillonnage par quotas ou proportionnel**

C'est l'une des techniques les plus simples et les plus répandues. Elle repose sur un principe simple : reproduire le plus fidèlement possible les caractéristiques de la population étudiée grâce à l'application de règles très précises.

Par ce procédé, on va chercher à construire un échantillon qui ressemble à la population mère à partir d'informations statistiques (données de cadrage disponibles) sur la structure de cette population. On réduit ainsi le risque d'avoir un échantillon aberrant. Par ailleurs, cette méthode permet d'obtenir des précisions d'autant plus fines que les critères de quotas sont corrélés à l'objet du sondage (modèle de population.)

Ces données statistiques, sur les principaux caractères de la population visée par l'étude, sont appelées variables contrôlées (sexe, âge, taille de la commune de résidence, la profession du chef de ménage, la profession du répondant, le niveau d'étude etc.)

Le principe sur lequel repose cette méthode est qu'un échantillon identique à la population mère quant à la répartition des variables contrôlées sera peu différent de cette population quant aux autres variables non contrôlées.



Généralement se sont les statistiques produites par Statistique Canada qui sont utilisées pour établir ces quotas. Une fois sur le terrain, l'enquêteur est libre d'interroger qui il veut à condition de respecter ces quotas.

Exemple : s'il y a dans la population 48% d'hommes et 52% de femmes, 25% de jeunes de 25 à 30 ans, etc. L'enquêteur chargé d'enquêter auprès de 100 personnes devra interroger 48 hommes, 52 femmes et 25 personnes de 25 à 30 ans, etc.

Les quotas retenus peuvent être "simples" c'est à dire avec plusieurs séries de distribution qui seront respectées de façon indépendante ou encore croisée en combinant plusieurs caractéristiques.

Revenons à notre exemple si l'on sait que 19% des femmes de la population mère ont entre 25 à 30 ans et 31% des hommes ont de 25 à 30 ans. L'enquêteur devra respecter ces quotas et interroger 10 femmes et 15 hommes ayant de 25 à 30 ans.

Pour assurer une certaine objectivité aux résultats, il faut fixer quelques règles aux enquêteurs comme la gestion rigoureuse des feuilles de quotas, le contrôle strict de la démarche "aléatoire" de l'enquêteur (interdiction d'interroger des personnes qui se connaissent entre elles ou qui connaissent l'enquêteur; l'obligation d'une certaine dispersion géographique avec imposition d'un itinéraire précis à respecter ; ou encore la réalisation des enquêtes à des moments différents.

### **L'échantillonnage aléatoire simple.**

Comme je vous l'ai indiqué précédemment, cette méthode suppose l'établissement d'une base de sondage à partir de laquelle les éléments du sondage seront tirés au hasard.

Le principe de base est le suivant : tous les individus ou éléments constituant la population de base ont la même chance d'être choisis (par exemple : à partir de la liste complète des clients de l'entreprise.)

Il faudra donc, dans un premier temps établir la base de sondage qui est une liste exhaustive des individus de la population. Pour cela, on aura recours à des fichiers existants ou à des registres. Dans tous les cas de figure, il est très important que ces fichiers soient complets, à jour et adaptés à la population sur laquelle on veut enquêter.

A ce sujet, il est parfois tentant de vouloir utiliser des répertoires comme l'annuaire du téléphone ou les listes d'inscription électorale. Mais ce sont souvent de mauvaises bases de sondage. En effet, on évalue en effet à 6% le nombre de personnes qui ne figurent pas dans l'annuaire. Pour ce qui est des listes électorales, elles sont souvent inexactes en raison du nombre des non-inscrits ou des personnes inscrites mais résidant ailleurs ou l'inverse etc.

Une fois la base de sondage établie il existe trois techniques :

- le tirage de jetons numérotés ou de papiers avec le nom de chaque individu dans une urne
- la désignation de façon aléatoire de l'échantillon à l'aide d'un ordinateur
- la sélection de nombre identifiant les individus à l'aide d'une table de nombres aléatoires

Pour une population restreinte, la première technique suffira largement; pour une population de grande dimension, il est préférable d'attribuer un numéro distinct à chaque individu ou élément de la base de sondage. Ceux qui composeront l'échantillon seront sélectionnés à l'aide d'un programme informatique ou d'une table de nombres aléatoires.

L'intervention du hasard n'est pas synonyme d'anarchie. Il obéit tout de même à des règles strictes qu'il faut respecter dans la constitution d'un échantillon. Si on prend le cas de la population québécoise, un échantillonnage probabiliste représentatif nécessite le choix d'environ 1000 sujets.

Aussi, un échantillon construit suivant cette méthode, ne peut être modifié par les enquêteurs en cours d'enquête. En effet, l'enquêteur n'interrogera pas un profil, comme dans l'échantillon par quotas, mais une personne bien précise. Des moyens doivent être prévus pour procéder aux différentes relances nécessaires pour interroger les individus sélectionnés.

En règle générale, ces échantillons portent sur un grand nombre d'individus, environ 1000 ou plus.

### **Comparaison entre les méthodes par quotas et aléatoire.**

Le premier type de méthode d'échantillonnage par quotas est privilégié par les instituts de sondage et de marketing tandis que la seconde (aléatoire) est utilisée par Statistique Canada et les statisticiens professionnels.

Chacune de ces méthodes a ses avantages et ses inconvénients mais les études comparatives entre sondages probabilistes et sondages par quota ont montré l'équivalence des performances des deux méthodes, notamment dans le cadre des sondages électoraux.

Il faut toute fois préciser que l'utilisation de la méthode des quotas (empiriques) pour un échantillon de 3 000 individus et plus donnera des estimations biaisées. Tandis que l'utilisation de la méthode aléatoire donnera des résultats d'une grande imprécision pour des échantillons de moins de 80 individus.

Un autre aspect de la complémentarité de ces deux méthodes c'est la combinaison de ces deux méthodes pour la construction d'échantillons complexes.

En conclusion, sur ces techniques d'échantillonnage, il faut retenir que l'utilisation de ces techniques va vous permettre de reproduire fidèlement la population totale que vous étudiez et, de ce fait, vous garantir la représentativité de votre échantillon. Mais il faut savoir qu'en matière de sondage il n'existe pas de certitude, même si l'échantillon a été obtenu par utilisation d'une méthode rigoureuse. Les pourcentages obtenus, résultats du sondage, ont donc toujours une marge d'erreur liée au fait qu'on interroge une fraction de la population.

### **I.2.5. Les étapes d'une enquête par sondage**

a. Conception générale de l'enquête : Cette étape comprend la définition des objectifs de l'enquête, de la population étudiée, des informations que l'on souhaite obtenir, avec quelle périodicité, à quel niveau de finesse... Ici il faut s'assurer que les informations recherchées n'existent pas ailleurs (données dites secondaires). En fonction de ce qui précède, on définit une technique de recueil des informations (par téléphone, par voie postale, par internet, par enquêteur à domicile) et simultanément une méthode d'échantillonnage. On détermine aussi la taille de l'échantillon nécessaire, compte tenu du budget disponible et éventuellement de la précision souhaitée.

b. La rédaction du questionnaire : - Première rédaction du questionnaire (si possible en utilisant les résultats d'études exploratoires ou qualitatives préalables) ; - Pré test du questionnaire ; - rédaction définitive incluant le pré codage du questionnaire pour faciliter la saisie informatique.

c. Administration du questionnaire : - Le terrain : Commence avec la phase de collecte des informations auprès des interviewés, qui commence par une réunion d'information des enquêteurs (appelé aussi briefing) afin de leur transmettre les consignes d'échantillonnage et de passation du questionnaire ; Il s'en suit l'administration des questionnaires sur le terrain et, enfin par le contrôle de la qualité du travail des enquêteurs.

d. Traitement et analyse des données : Vérification, codage et saisie informatique des questionnaires, puis traitement informatique, généralement par des logiciels spécialisés de traitement d'enquêtes. Les résultats de ces traitements sont alors synthétisés dans un rapport final.

L'importance et la complexité de ces différentes phases varient beaucoup d'une enquête à l'autre.

### **I.3. Déroulement du sondage**

#### **I.3.1. Les phases**

Opération éminemment complexe et délicate, le sondage exige une préparation particulièrement minutieuse, encore que le degré de complexité puisse varier d'un sondage à l'autre. Le sondage peut être très simple à réaliser lorsque l'ensemble étudié est constitué de 100.000 fiches-auteur, par exemple, mais extrêmement difficile lorsque l'ensemble étudié est constitué de 100.00 ménages, appartenant à différentes couches socio-économiques et répartis sur toutes les régions d'un pays. Des principes généraux ont néanmoins été formulés qui permettent de surmonter les difficultés pratiques et de guider la réalisation effective des enquêtes par sondage, que l'échantillon à constituer soit simple ou complexe et délicat, petit ou grand.

#### **Enoncé des objectifs**

Le sondage ne doit être effectué ni in abstracto ni à l'aveuglette, hors de tout but ou objectif précis déterminant et orientant la mise en application du plan d'échantillonnage élaboré pour la circonstance. Les grands objectifs du sondage doivent au contraire être énoncés expressément et expliqués clairement dès la phase initiale, car ils constituent la référence première et constante de la réalisation effective du sondage. Toutes les étapes et phases ultérieures doivent être systématiquement axées sur ces objectifs, pour faire en sorte qu'il y ait une cohérence entre les objectifs, le plan d'échantillonnage élaboré a priori et l'organisation et la conduite effective du sondage. A chacune des étapes ultérieures du sondage, il est utile de déterminer si, jusqu'à la phase précédente, le déroulement du sondage est en accord avec le ou les objectifs premiers.

Ainsi, dans le cas de la présente enquête pilote, des objectifs bien définis et précis étaient visés. Comme il est indiqué au chapitre 1, ces objectifs sous-tendent le cadre, le schéma, le plan, la méthode et la conduite effective du sondage.

Dans quelle mesure ces objectifs ont été atteints, telle est l'aune à laquelle doit donc se mesurer le succès ou l'échec du sondage

#### **Population**

L'étape suivante du sondage consiste à cerner avec précision la population d'où l'échantillon va être tiré. On appelle population (ou univers) un ensemble regroupant tous les éléments qui possèdent certaines caractéristiques précises. Ainsi définie, la population doit regrouper tous les éléments possibles qui ont ces caractéristiques. Certains types de population - ensembles d'ampoules électriques ou d\*. bougies par exemple - sont faciles à cerner et définir, aussi bien en théorie qu'en pratique, alors que dans d'autres cas, la définition de la population peut poser des problèmes conceptuels, méthodologiques et opérationnels d'une difficulté extrême, Il faut donc préciser les règles ou critères permettant de définir et d'identifier sans ambiguïté, à des fins opérationnelles, la population d'où l'échantillon va être tiré.

Dans le cas de la présente enquête, qui est axée sur les statistiques de l'éducation, les éléments fondamentaux de la population seront tout naturellement les élèves, les

enseignants, les administrateurs de l'enseignement, les établissements scolaires et les lieux où ils se trouvent

Les ménages fournissent des élèves aux établissements. Elèves et ménages possèdent un nombre infini de caractéristiques et d'attributs, d'où la difficulté que poserait l'élaboration d'un échantillon exhaustif pouvant représenter l'infinité d'attributs de l'ensemble. Cette exhaustivité n'est d'ailleurs pas nécessairement requise dans la pratique. On n'a donc retenu pour le présent sondage qu'un certain nombre d'attributs particuliers de la population, laquelle a été définie en fonction de ces attributs.

### **Données à rassembler**

Il faut impérativement que les données à rassembler au moyen du sondage soient clairement définies et concrètement identifiées. Les besoins en données doivent être analysés par rapport aux objectifs du sondage, et la pertinence des données par rapport aux buts de l'enquête doit être analysée et vérifiée minutieusement avant la mise en œuvre du plan d'échantillonnage. Il faut veiller à ce qu'aucune donnée essentielle ne soit exclue de l'échantillon et éviter à tout prix de rassembler des données qui ne sont pas essentielles.

Les catégories de données identifiées aux fins de la présente enquête pilote visaient donc essentiellement à combler les lacunes du système d'information actuel et à vérifier la validité et l'exactitude des données émanant de sources secondaires. Le sondage a donc été axé sur la production de données relatives à la répartition par âge, année d'études, sexe et caste des effectifs de l'enseignement du premier degré, aux taux d'abandon scolaire et de redoublement, aux enseignants (répartis par année d'études et qualifications), aux infrastructures et à leur renforcement dans le cadre de ("Opération tableaux noirs" et aux postes de dépenses sur lesquels on ne disposait pas de données jusqu'à présent. Toutes les catégories de données retenues étaient donc directement reliées aux objectifs fondamentaux de l'enquête.

### **Exergue :**

Degré de précision recherché

Les résultats d'un sondage sont toujours marqués par un degré d'incertitude plus ou moins grand, qui s'explique par (i) la sélection d'un nombre limité d'unités, qui fait que l'échantillon est de taille beaucoup plus petite que la population; (ii) la sélection aléatoire des éléments de l'échantillon, pouvant donner lieu à des erreurs aléatoires ou à des fluctuations d'échantillonnage; (iii) les éventuelles erreurs d'observation et de mesure; (iv) les distorsions introduites par l'enquêteur ou par le répondant; (v) les déficiences éventuelles du plan et de la méthode d'échantillonnage; (vi) la mauvaise application d'un plan d'échantillonnage bien conçu par ailleurs.

Constituer des échantillons de plus grande taille, se doter d'outils d'observation plus perfectionnés et de méthodes modernes de mesure, adopter là méthode d'échantillonnage qui convient et l'appliquer minutieusement, disposer d'un nombre suffisant d'agents sur le terrain, convenablement formés : toutes ces conditions peuvent contribuer à réduire très largement les incertitudes, encore qu'il soit impossible d'éliminer toutes les causes possibles d'incertitude et d'erreurs d'échantillonnage.

Aussi est-il courant de déterminer a priori le degré de précision recherché dans les résultats du Sondage, et cette tâche incombe à l'utilisateur final desdits résultats.

La taille de l'échantillon joue un rôle important dans les sondages. Plus l'échantillon est grand, plus les résultats seront fiables. Dans sa grandeur maximale, l'échantillon coïncide avec la population étudiée, auquel cas il n'y a plus lieu d'appliquer les techniques du sondage. Le problème est donc de déterminer la taille optimale de l'échantillon.

### Mesures

La fiabilité de l'échantillon et sa pertinence par rapport à la population étudiée sont fonction du degré de précision atteint dans la mesure des variables et/ou l'observation des attributs. Or, il existe un large éventail d'outils de mesure et de méthodes d'échantillonnage.

### Unités d'échantillonnage

Lorsque la population est trop importante pour être étudiée dans son intégralité, ce qui est souvent le cas, il faut la diviser en petites unités, les unités d'échantillonnage, dont il faut déterminer la taille. L'agrégation de ces unités doit correspondre à l'intégralité de la population. Chaque membre de la population doit appartenir à une unité et une seule, ce qui signifie qu'il ne doit pas y avoir chevauchement de deux unités. Le territoire d'un grand pays comme l'Inde, par exemple, peut être divisé en Etats, les Etats en districts, les districts en cantons, et les cantons en villages, villes et agglomérations. Chacune de ces subdivisions peut être prise comme unité d'échantillonnage pour tel ou tel degré du sondage. Le plus souvent, il est possible de choisir entre plusieurs unités d'échantillonnage possibles. C'est ainsi que dans la présente enquête pilote, différentes unités ont été utilisées à différents degrés de l'échantillonnage multiple en grappes. Etats, districts, cantons, villages/villes et enfin, établissements scolaires : tous ces sous-ensembles ont constitué des unités d'échantillonnage pour des grappes stratifiées de la population étudiée.

### Base de sondage

La liste des unités d'échantillonnage est appelée base de sondage. Il est préférable que cette liste soit détaillée et exhaustive par rapport à la population étudiée, afin que la probabilité de figurer dans l'échantillon soit la même pour toutes les unités. "L'égalité des chances" de toutes les unités à chaque tirage n'est toutefois envisageable que dans le cas d'un échantillonnage aléatoire simple. Dans la pratique, les listes d'unités d'échantillonnage s'avèrent souvent partielles ou incomplètes, illisibles, voire, parfois, constituées d'éléments inconnus ou impossibles à identifier.

Plus la population est nombreuse, moins on en connaît les caractéristiques et plus il est difficile d'établir une base de sondage judicieuse. Les populations nombreuses posent aussi plus que d'autres le problème du double comptage. De même, plus l'ensemble étudié est spécialisé, plus il sera difficile de le connaître avec précision et, partant, d'établir une bonne base de sondage.

Il aurait été très difficile, voire impossible dans le cadre de la présente enquête, de dresser la liste complète de tous les élève\* inscrits dans les établissements d'enseignement en Inde.

En revanche, établir la liste exhaustive des Etats et Territoires constituant l'Union indienne, qui étaient les unités d'échantillonnage pour le premier degré de l'enquête pilote, n'a posé aucune difficulté, pas plus que les listes exhaustives des districts composant chacun des quatre Etats retenus pour le sondage. Une fois déterminés les districts, autrement dit les unités d'échantillonnage du second degré, le dénombrement de tous les cantons composant les districts retenus n'a guère posé de problèmes. Les listes des cantons, unités d'échantillonnage du troisième degré, pouvaient donc aussi être exhaustives. Une fois identifiés les cantons à retenir dans l'échantillon, les directions de l'éducation, à l'échelon de l'état, et les inspections à l'échelon du district, ont pu fournir les listes complètes des établissements dans chacun des cantons. Ces établissements représentaient les unités d'échantillonnage au dernier degré du sondage, et la constitution de l'échantillon a été là aussi matériellement possible. Cela dit, les établissements privés non subventionnés et non agréés ne figuraient pas sur les listes fournies par les inspections, parce que ni le ministère de l'éducation ni aucun autre organisme public ne disposent de renseignements ou d'une base de données sur les établissements de ce type. Il a donc fallu obtenir des renseignements préliminaires à ce sujet auprès des chefs des établissements figurant dans l'échantillon ou auprès des ménages pendant la mise en œuvre du plan de sondage.

### **Taille de l'échantillon**

Un échantillon trop grand est plus coûteux à constituer sans être nécessairement plus fiable et/ou plus précis, d'où un gaspillage de ressources; inversement, un échantillon trop petit peut donner des résultats qui ne sont ni fiables ni représentatifs, donc qui sont sans grand intérêt pratique. Il importe dès lors de savoir comment déterminer la taille optimale de l'échantillon. Une approximation de la taille optimale de l'échantillon correspondant à la méthode d'échantillonnage retenue, pour Je degré de précision voulu, peut être obtenue objectivement au moyen de la formule ci-après :

$$n = PQ/V,$$

où n représente la taille de l'échantillon, P la proportion d'unités élémentaires possédant la caractéristique étudiée dans la population totale, Q la proportion d'unités élémentaires n'ayant pas la caractéristique étudiée dans la population totale et V le niveau de variance de la moyenne recherché. Mais il arrive que des considérations de coût, donc de viabilité économique, de commodité administrative, de faisabilité technique et de temps interviennent autant dans la détermination de la taille de l'échantillon que la méthode d'échantillonnage retenue ou le degré de précision recherché dans les résultats. Dans l'enquête pilote, l'hétérogénéité de chaque strate de population, du point de vue de la répartition ville-campagne ou filles-garçons, du caractère public ou privé de la propriété et de la gestion des établissements ainsi que de leur caractère subventionné ou non, a joué un grand rôle dans la définition de la taille de l'échantillon. La méthode, d'échantillonnage qui a été jugée préférable combine donc les méthodes stratifiée, en grappes, aléatoire et systématique. En conséquence, la taille effective de l'échantillon ne peut pas être la même, et est parfois plus grande, que celle qui aurait été déterminée par l'ç choix d'une seule des méthodes mentionnées ci-dessus et par le degré de précision recherché:

### **Essai préalable du questionnaire**

Le questionnaire, outil principal de l'enquête, joue un rôle capital dans les sondages, dont la réussite (où l'échec) dépend très largement du caractère plus ou moins adéquat et adapté du questionnaire.

La marque d'un bon questionnaire est que les questions y figurent : (i) par leur nombre et leur nature, correspondent exactement aux besoins de l'enquête et ne sont ni trop ni trop peu nombreuses; (ii) sont conçues et agencées de manière à ce que la personne à interroger en perçoive clairement l'intention et le but, sans le moindre doute ou ambiguïté et sans qu'il y ait besoin d'explications supplémentaires; (iii) sont formulées de manière à ne susciter ni antipathie ni hostilité chez l'enquêté, n'exigent de lui aucun renseignement trop personnel ou confidentiel qu'il pourrait être réticent à divulguer à l'enquêteur et ne sont assorties d'aucun jugement de valeur pouvant l'indisposer; (iv) sont exactement celles qui permettent d'obtenir de l'enquêté les renseignements pertinents et précis requis aux fins du sondage.

Il est difficile d'élaborer du premier coup un questionnaire qui donne entièrement satisfaction. L'essai préalable du questionnaire sur le terrain constitue donc une étape essentielle du sondage. L'essai préalable consiste à remplir une version préliminaire du questionnaire au moyen de réponses obtenues sur le terrain afin de (i) déceler les lacunes structurelles du questionnaire; (ii) déterminer les modifications à apporter aux questions défectueuses afin d'éliminer les ambiguïtés, inexactitudes et incohérences au niveau du contenu, de la formulation et de l'expression; (iii) ajouter ou supprimer des questions pour remédier aux déficiences du questionnaire de départ. L'essai préalable permet généralement d'améliorer considérablement la première version du questionnaire. Il peut également faire apparaître des lacunes et des points qui risquent de poser des problèmes lorsque le sondage sera effectué en grandeur nature.

### **Organisation du travail sur le terrain**

Par organisation du travail sur le terrain, autre aspect essentiel du sondage, on désigne l'affectation judicieuse des enquêteurs dans les différentes zones couvertes par l'enquête. Il est préférable que les enquêteurs connaissent la ou les langues locales et aussi résident dans la région à laquelle ils sont affectés. Ce critère est particulièrement important pour la réussite du travail sur le terrain lorsqu'il existe des différences sociales et ethniques très marquées et que l'enquêteur venu de l'extérieur risque d'être suspect ou de susciter l'antipathie ou l'hostilité. Par ailleurs, dans un pays aussi étendu que l'Inde, recruter les agents de terrain parmi la population locale permet de réaliser des économies de temps et de frais de voyage, vu la distance qui sépare les différentes zones de l'enquête. C'est pour ces trois raisons que le travail sur le terrain a été confié à des agents locaux dans chaque Etat.

### **Administration de l'enquête**

Les grandes enquêtes sur le terrain posent de nombreux problèmes analogues à ceux que l'on rencontre dans l'administration des entreprises. Ces problèmes doivent être réglés par les agents de terrain, qui doivent, pour ce faire, avoir reçu une formation préalable. Celle-ci doit essentiellement consister à: (i) les initier aux principaux objectifs de l'enquête; (ii) leur apprendre les méthodes et instruments de mesure à utiliser sur le terrain; (iii) leur apprendre à faire en sorte que l'enquêté réponde convenablement aux questions et à déceler



les éventuelles contradictions, incohérences et inexactitudes dans les réponses; (iv) les initier dans le détail aux différentes étapes de la méthode de sondage retenue pour l'enquête; (v) leur donner les moyens de régler les cas de "non-réponse" ou de "réponse partielle". Le travail sur le terrain nécessite un système approprié d'orientation et de supervision, qui doit donc être mis en place. Tout aussi utile est la mise au point préalable d'un système de vérification détaillée de la qualité des questionnaires remplis sur place par les enquêteurs. Pour que ce travail sur le terrain donne des résultats fiables, il est essentiel qu'il soit soumis à une supervision efficace, assurée par des personnes compétentes.

Pour l'enquête pilote, un coordinateur a été désigné dans chaque inspection/direction des différentes zones de l'échantillon choisies à cette fin. Des assistants de projet et des membres de haut rang de l'équipe se sont également rendus sur le terrain durant le déroulement de l'enquête.

L'organigramme doit donc être établi en fonction de tous ces éléments pour faciliter le bon déroulement de, l'enquête. Dans le plan adopté pour l'enquête pilote, une importance à peu près égale a été accordée à tous ces aspects de l'administration du travail sur le terrain.

L'on se rend compte cependant, a posteriori, que la formation des enquêteurs aurait pu être prolongée d'une semaine environ, ce qui aurait permis au personnel de terrain d'effectuer des applications pratiques du plan d'enquête. Il aurait également fallu demander aux enquêteurs de tester les questionnaires auprès des écoles situées à Delhi, sous la supervision directe des membres de haut rang de l'équipe, et les familiariser concrètement avec tous les aspects mentionnés plus haut. L'on se serait alors épargné les mesures correctives qu'il a fallu prendre pendant la mise en œuvre effective du plan, lorsque des difficultés sont apparues sur le terrain. Mais le temps imparti a contraint à accélérer quelque peu la formation du personnel de terrain.

### **I.3.2. Méthode d'échantillonnage**

Il existe plusieurs méthodes d'échantillonnage, chacune ayant ses avantages et ses inconvénients. Ces méthodes diffèrent également par les conditions préalables devant être réunies dans chaque cas et par les ressources nécessaires à leur mise en œuvre. L'adéquation de la méthode d'échantillonnage est fonction : (i) de la taille de l'échantillon compte tenu du degré de précision recherché; (ii) des ressources nécessaires pour appliquer la méthode; (iii) du calendrier de l'enquête et des délais requis pour la mise en œuvre de la méthode; et (iv) de la taille, de la nature et de l'hétérogénéité de la population étudiée.

En théorie, la méthode d'échantillonnage la plus satisfaisante est l'échantillonnage aléatoire simple (EAS), qui permet d'obtenir un échantillon représentatif, quelle que soit la population étudiée. Elle consiste à tirer un échantillon de  $n$  unités sur une population de  $N$  unités en veillant à ce que la même probabilité de choix soit assignée à tous les échantillons  $NC_n$  qu'il est possible de constituer à partir de la population en question. Les unités constituant l'échantillon sont tirées une à une en veillant à ce que, à chaque tirage, toutes les unités qui n'ont pas déjà été tirées aient la même probabilité de l'être. Ces deux définitions de l'échantillonnage aléatoire simple sont en fait équivalentes.

Cependant, il est souvent préférable d'appliquer d'autres méthodes d'échantillonnage, pour des raisons pratiques. Le coût des ressources à mobiliser, la commodité, la recherche d'un plus grand degré de précision dans les résultats, l'adéquation à la réalité, les contraintes et difficultés pratiques sont autant de facteurs qui peuvent justifier que l'on renonce à l'EAS. La stratification et la constitution de grappes sont souvent des solutions de rechange pratiques, surtout lorsque la population à étudier est composée d'unités très diversifiées pouvant être regroupées en strates et/ou grappes clairement identifiables. Les principaux facteurs qui militent en faveur de la stratification sont : (i) la commodité administrative compte tenu de la population étudiée; (ii) la recherche d'un degré bien déterminé de précision des données, qui peut nécessiter que l'on traite chaque couche/strate comme une population à part entière; (iii) les différences entre les caractéristiques essentielles des unités élémentaires d'une strate de la population à l'autre, qui font que les problèmes d'échantillonnage varient très sensiblement selon les strates; ainsi, lors de l'enquête pilote, on a tout naturellement adopté pour les écoles les distinctions rurales/urbaines, filles/garçons, privées/publiques et subventionnées/non subventionnées, de manière à ce que les données d'échantillon soient représentatives des principaux aspects de la diversité de la population mère dans chaque strate; iv) le fait que la stratification permet souvent de mesurer avec une plus grande précision les caractéristiques de la population globale. La population, hétérogène, a donc été découpée en strates plus homogènes: sociolinguistiques, économiques-éducatives, administratives, etc.

#### **Gain de précision**

Le gain de précision réalisé dans les estimations grâce à la stratification de la population étudiée peut être assimilé à la différence entre l'estimation de la variance pondérée de la moyenne pondérée obtenue par échantillonnage stratifié et l'estimation de la variance de la moyenne qui aurait été obtenue par échantillonnage aléatoire simple. L'on peut, par exemple, calculer le carré moyen des écarts par rapport à la moyenne de

l'échantillon composite abstraction faite de la stratification et, partant, de la pondération des valeurs de strate:

$$s^2 = \frac{\sum (x(h) - \bar{y})^2}{n-1}$$

tandis que la variance pondérée de l'échantillon stratifié est donnée par la formule :

$$V(\bar{y}_{st}) = \frac{\sum W^2(h) s^2(h)}{n(h)} - \frac{W(h) s^2(h)}{N}$$

En règle générale,  $\bar{y}_{st}$ , moyenne pondérée de l'échantillon stratifié, est différent de  $\bar{y}$ , moyenne non pondérée de l'échantillon composite, en ce sens que :

$$\bar{y}_{st} = \sum_{h=1}^L N(h) \bar{y}(h) / N$$

Et

$$\bar{y} = \frac{\sum n(h) \bar{y}(h)}{n}$$

où la population a été divisée en  $h = 1, 2, L$  strates et où  $y(h_i)$  est la  $i$ -ième valeur du sous-échantillon tiré de la  $h$ -ième strate,  $N(h)$  l'effectif de la  $h$ -ième strate,  $n(h)$  la taille de l'échantillon tiré de la  $h$ -ième strate,  $s^2$  l'estimation de la variance par rapport à la moyenne de l'échantillon composite et  $W(h)$  les coefficients de pondération. Si les coefficients  $W(h) = N(h)/N = n(h)/n$  ou  $n(h)/N(h) = n/N$ , les deux moyennes coïncident.

### Taille des échantillons de strate

La constitution des échantillons tirés de chaque strate, pour ce qui est de leur taille en particulier, obéit à un certain nombre de règles de base : (i) une strate plus importante doit donner lieu à un échantillon plus important, sans pour autant qu'il y ait nécessairement proportionnalité. C'est en vertu de ce principe que dans l'Etat du Gujarat, l'échantillon tiré du district de Panchmahals avait un effectif supérieur à celui du district de Oandhinagar, les établissements scolaires étant beaucoup plus nombreux dans le premier district que dans le second. La taille des sous-échantillons correspondant aux districts de Barmer et Jaipur (Etat du Rajasthan) était également fonction de l'importance relative des effectifs des deux strates, sans que le rapport soit\* encore une fois, strictement proportionnel; (ii) plus la composition de la strate est diversifiée (hétérogénéité interne), plus l'échantillon doit être important. Ainsi, comme l'on supposait que des différences assez nettes existaient entre les établissements scolaires de différents cantons du district de Panchmahals, il a été décidé que le sous-échantillon correspondant à ce district serait de plus grande taille que celui du district de Oandhinagar; (iii) des échantillons plus importants peuvent être constitués lorsque le coût est moindre. Si les coûts étaient la considération dominante, des échantillons

de plus grande taille auraient été pris dans le district de Qandhinager, qui est moins étendu et moins hétérogène que celui de Panchmahals.

### **Avantages de la stratification**

Le gain de précision que l'on peut escompter tirer de la stratification est fonction d'un certain nombre de facteurs. Il est d'autant plus grand que (i) la population est composée d'établissements de taille très variable. L'école constitue l'unité d'échantillonnage au dernier degré du sondage et la taille des écoles varie beaucoup selon qu'il s'agit de zones rurales ou urbaines, d'établissements publics ou d'établissements privés de langue anglaise, d'établissements pour filles ou pour garçons, etc. L'on peut dès lors s'attendre à ce que la stratification augmente considérablement la précision des résultats. Telle est, en tout cas, l'une des raisons qui ont amené à opter pour la stratification. Faute de pondérations appropriées, les variations importantes entre des strates quasi homogènes comme le district de Gandhinagar et d'autres strates très diversifiées et numériquement importantes comme le district de Panchmahals risquent d'avoir un effet disproportionné sur les moyennes et les variances. En pratique, on a utilisé des coefficients de pondération proportionnels à l'effectif des sous-strates pour estimer les paramètres de l'échantillon stratifié; (ii) les principaux attributs/variables qui font l'objet de l'enquête sont étroitement liés à la taille de l'établissement. En règle générale, la population des villages est moins nombreuse que celle des villes et agglomérations urbaines. Les effectifs scolaires sont fonction, en dernière analyse, du nombre total d'enfants d'âge scolaire. Les effectifs des établissements sont donc généralement moins importants dans les zones rurales que dans les zones urbaines. De même, les effectifs des écoles de filles sont moins importants que ceux des écoles de garçons, ce qui s'explique essentiellement par des préjugés culturels contre la scolarisation des jeunes filles. Les établissements de ce type (non mixtes) ont de toute évidence des effets disproportionnés sur la précision des résultats du sondage. Par ailleurs, la densité de peuplement de zones tribales telles que le Panchmahals est généralement plus faible que dans le reste du pays. Toutes ces données de base ont rendu la stratification nécessaire; (iii) Ton dispose d'un bon instrument de mesure de la taille des établissements. Dans le cas d'espèce, la taille peut être clairement mesurée en termes d'effectifs scolaires. L'enquête pilote devrait donc gagner considérablement en précision, pratiquement à tous les points de vue, parce qu'à la base, toutes les conditions énumérées ci-dessus sont réunies.

Dans le cas de la présente enquête pilote, il s'agit, aux trois premiers degrés, d'un échantillonnage en grappes très stratifié, à savoir que les Etats constituent de grandes grappes de districts, les districts de grandes grappes de cantons et les cantons de grandes grappes d'écoles. L'école, unité d'échantillonnage au dernier degré, est assimilée à une grande grappe d'élèves, de moyens d'enseignement, d'infrastructures et d'enseignants. Les renseignements recueillis à partir des établissements retenus dans l'échantillon ont trait à des attributs tels que l'âge, l'année d'études, le sexe, la caste et le parcours scolaire (redoublements, abandons scolaires, etc.). S'agissant des enseignants, les caractéristiques à étudier étaient notamment les titres universitaires, l'expérience, l'ancienneté et le sexe. En outre, des données ont été recueillies sur les moyens d'enseignement et les infrastructures des établissements. La présente enquête pilote se distingue des opérations habituelles de ce type par le fait que la véritable unité de sondage, à savoir les élèves, est demeurée cachée derrière l'écran des établissements. Ce parti pris est pleinement justifié par des

considérations pratiques : faisabilité technique, viabilité économique, coût et commodité, Il aurait été en effet pratiquement impossible d'établir une liste complète de tous les élèves inscrits dans les établissements d'enseignement primaire d'un pays aussi vaste et peuplé que l'Inde. A supposer qu'une telle liste pût être établie, il aurait fallu, pour dénombrer tous les élèves, identifier et séparer les élèves des non-élèves au sein des familles, ce qui n'était guère possible étant donné les effectifs considérés et la présence d'élèves et de non-élèves dans une même famille d'une part, et l'existence de groupes importants de ménages n'ayant pas accès à l'éducation, de l'autre. Le dénombrement des ménages s'est effectivement avéré très difficile. Etant donné que les élèves résident, au sens restreint du terme, dans les écoles, les écoles retenues dans l'échantillon ont été assimilées, en définitive à l'ensemble des élèves qui y sont inscrits, ce qui a permis de réaliser des économies de temps et d'argent et d'échapper, au nombreuses difficultés pratiques que posait l'enquête directe auprès des élèves eux-mêmes. Toutefois, le fait de considérer l'école comme le répondant a entraîné un certain nombre d'autres contraintes en ce qui concernait notamment l'accès aux établissements» qui était limité aux heures de travail et aux jours ouvrables» Les grandes vacances et les jours de congés d'une part, l'emploi du temps des directeurs d'établissement, de l'autre, ont souvent gêné le travail sur le terrain.

Les notions de mise en grappes et de stratification semblent néanmoins converger dans l'enquête pilote, cette convergence étant justifiée par la nature des répondants et par les buts de l'enquête.

### **Sélection systématique d'échantillons aléatoires**

Même au niveau de grappes ou des strates, il n'y a pas eu sélection simple ou aléatoire des unités. La sélection systématique d'échantillons aléatoires a été jugée préférable, en raison de ses avantages propres. L'échantillonnage systématique (ES) semble à première vue tout à fait différent de l'échantillonnage aléatoire simple (EAS). Les unités d'une population sont d'abord numérotées de 1 à N, puis classées selon un ordre précis, l'ordre alphabétique par exemple. Ce nombre total d'unités de la population ou de la strate est divisé par la taille de l'échantillon ou du sous-échantillon, soit  $N(h)/n(h) = k$ . Pour constituer un échantillon de  $n(h)$  unités à partir de la  $h$ -ième strate, la première unité, que nous appellerons  $i$ , est tirée au hasard, puis l'on inclut dans l'échantillon chaque  $k$ -ième unité à partir de la première. La deuxième unité portera donc le numéro  $i+k$ , la suivante le numéro  $i+2k$ , etc. L'ensemble de l'échantillon tiré de cette strate/population est donc déterminé à partir du tirage aléatoire de la première unité.

L'échantillonnage systématique (ES) comporte des avantages évidents par rapport à l'EAS. Il est plus simple, de sorte que le plan de sondage est plus facile à élaborer et plus facile encore à mettre en œuvre. Cet avantage est particulièrement appréciable lorsque les essais en vue de la constitution de l'échantillon doivent être effectués in situ par le personnel de terrain. Même lorsque les unités à inclure dans l'échantillon sont déterminées au sein des services statistiques par des personnes très compétentes, les économies de temps et de ressources peuvent être non négligeables lorsque l'on opte pour TES.

Dans le cadre de la présente enquête, il a été demandé aux enquêteurs de procéder eux-mêmes au remplacement des non-répondants, en comptant de nouveau  $k$  unités à partir de l'unité qui Suit immédiatement celle qui n'a pas répondu, procédé qui ne pouvait

fonctionner dans le cadre d'un sondage aléatoire simple. L'ES semble a priori plus précis que l'EAS. De fait, la sélection systématique produit une stratification de la population. L'échantillonnage systématique peut donc être au moins aussi précis qu'un échantillonnage stratifié dans lequel serait tirée Une unité par strate, chaque strate éteint composée de k unités. La différence entré les deux méthodes de sondage est que dans l'échantillonnage systématique, les unités retenues dans l'échantillon occupent la même position relative dans la strate alors que dans l'échantillonnage stratifié, cette position est déterminée de manière aléatoire. L'échantillonnage systématique est donc même plus précis que l'échantillonnage stratifié, parce que les unités retenues dans l'échantillon se répartissent plus régulièrement au sein de population.

Cet avantage est particulièrement marqué lorsqu'à l'intérieur des strates, il existe de fortes disparités à l'égard de facteurs importants autres que ceux qui ont servi à déterminer les strates. C'est précisément pour cette raison que l'échantillonnage systématique a été retenu pour la présente enquête. A l'intérieur des strates, où les écoles ont été classées en fonction des distinctions écoles rurales/urbaines, privées/publiques et écoles de filles/écoles de garçons, l'on constate des différences très marquées entré établissements sur des aspects importants : écoles à un ou plusieurs maîtres, à une ou plusieurs classes, langue d'enseignement, moyens d'enseignement multilingue, etc. Il aurait été impossible dans la pratique de subdiviser délibérément les strates en fonction de l'importance des multiples facteurs qu'il y aurait fallu neutraliser ou contrôler lors de la mise en œuvre pratique du plan d'échantillonnage, d'où la préférence accordée à l'échantillonnage systématique à l'intérieur des strates, par rapport à l'EAS. La présente enquête combinait donc les méthodes de l'échantillonnage stratifié et systématique et de l'échantillonnage par choix raisonné. L'on peut donc dire qu'au dernier degré de l'enquête, la méthode appliquée était celle de l'échantillonnage systématique stratifié, qui convient bien lorsque les fractions fondées sont inégales. La Systématisation du choix des unités de sondage doit donc permettre un niveau raisonnable de différenciation à l'intérieur de chaque strate, afin que les différents caractères présents dans l'effectif de chaque strate puissent être représentés. L'échantillonnage systématique est donc étroitement lié à l'échantillonnage aléatoire et à l'échantillonnage stratifié. La distance k qui sépare chacune des unités prélevées de façon systématiques concrètement pour effet de stratifier la population en K grandes unités de sondage, dont chacune contient les k unités de départ. La sélection systématique de l'échantillon aléatoire revient donc en fait à choisir au hasard l'une de ces grandes unités de sondage. L'on peut donc considérer que l'échantillonnage systématique consiste à sélectionner une seule unité de sondage complexe qui constituera tout l'échantillon. L'échantillonnage systématique s'avère donc considérablement plus simple à manier et la mise en œuvre du plan de sondage plus économique.,

La variance des moyennes dans l'échantillonnage systématique peut être estimée par la formule ci-après :

$$V(y(sy)) = \{(N-1)/N\} S^2 - [k(n-1)/N] S^2 (wsy)$$

$$\text{où } S^2(wsy) = [1/k(n-1)] (y(ij) - y(i))^2,$$

qui désigne la variance pour les unités situées dans le même échantillon systématique.

Mais, dans la pratique, l'avantage comparatif de l'échantillonnage systématique par rapport à l'échantillonnage aléatoire simple ou à l'échantillonnage stratifié est dans une large mesure fonction de la nature de la population étudiée. Et c'est à la condition de bien connaître les données de base et la structure de cette population que l'on pourra tirer tout le parti possible de l'échantillonnage systématique. Dans le cas de l'enquête pilote, on avait effectivement une idée générale des caractéristiques principales de la population étudiée; même si l'ampleur de celles-ci n'était pas connue avec exactitude.

### **I.3.3. Classement, mise en tableaux et analyse des données**

Une fois les données rassemblées, après achèvement de l'enquête sur le terrain, l'étape suivante a consisté à dépouiller, classer et mettre en tableaux les données en vue de l'analyse finale. Il s'agit en l'occurrence d'une séquence d'étapes bien définies : (i) vérification et "conditionnement" des données; (ii) classement et mise en tableaux; (iii) analyse et interprétation.

#### **Vérification et "conditionnement" des données**

La première phase de l'étape finale du sondage consiste à vérifier minutieusement les questionnaires remplis, afin de déceler et, si possible, corriger les éventuelles contradictions, remédier aux erreurs d'enregistrement et enfin, supprimer les réponses erronées qui ne peuvent être ni corrigées ni modifiées objectivement.

Les données primaires ou brutes doivent alors parfois être "conditionnées" compte tenu des buts de l'enquête et/ou des conditions objectives dans lesquelles elle s'est déroulée. A titre d'exemple, il faut souvent convertir les valeurs monétaires en prix constants pour neutraliser les effets de l'inflation, comme il faut parfois modifier des unités de mesure pour, par exemple, passer de chiffres absolus à des chiffres par habitant, rapporter les effectifs scolaires aux effectifs des tranches d'âge ou la taille moyenne des établissements aux effectifs scolaires, etc.

La vérification des questionnaires remplis s'est avérée extrêmement utile dans le cas de l'enquête pilote. Elle a notamment permis de découvrir que deux ou plusieurs questionnaires comportaient des réponses identiques, ce qui a obligé - selon le nombre de questionnaires mal remplis - soit à les éliminer du dénombrement final des réponses, soit à retourner auprès du répondant pour obtenir de nouvelles réponses. Dans le cas de certaines valeurs répétitives (l'effectif total par exemple) obtenues à l'aide de plusieurs séries de questions ou tableaux, les erreurs d'enregistrement ou d'observation ont pu être corrigées. Il a en outre fallu demander de nouvelles réponses pour 35 questionnaires, parce que la répartition par âge des élèves n'était pas indiquée ou que les chiffres indiqués pour les effectifs étaient inexacts. Enfin, comme on l'a indiqué plus haut, la vérification des questionnaires a permis de découvrir des ambiguïtés dans la formulation de certaines questions concernant une sous-série de têtes de colonnes. L'on a également rejeté en bloc un très grand nombre de questionnaires parvenus de l'Haryana durant la phase finale des travaux, faute de disposer du temps nécessaire pour les faire remplir à nouveau par les enquêtés.

Le conditionnement des données a essentiellement consisté à convertir des valeurs absolues en valeurs relatives.

## **Classement et mise en tableaux**

Les données obtenues par échantillonnage doivent être présentées sous une forme épurée. Dans leur version originale, les données sont souvent trop brutes, trop nombreuses et trop détaillées pour être compréhensibles ou commodes à manier. Leur classement en catégories judicieusement définies contribue donc à réduire la masse des données, dans la mesure où cette agrégation permet d'éliminer des détails qui ne sont pas essentiels. Il faut néanmoins que ce classement soit conforme aux objectifs fondamentaux du sondage. Dans bien des cas, un bon classement des données aide souvent à trouver une présentation tabulaire adéquate, ce qui facilite l'analyse et l'interprétation des résultats.

Dans le cas de la présente enquête, la plupart des éléments du questionnaire étaient déjà présentés sous forme de tableaux, si bien que très rares étaient les séries de réponses qui ont dû être classées et mise en tableaux ultérieurement, une fois l'enquête sur le terrain achevée.

## **Analyse et interprétation**

Des données convenablement classées et mises en tableaux se prêtent à une analyse faisant appel à diverses techniques, statistiques et autres, qui peuvent être descriptives ou récapitulatives (moyennes et variances par exemple). C'est au cours de cette phase que sont produites les estimations qui doivent répondre aux objectifs ultimes du sondage.

Dans le cas de la présente enquête pilote, on a appliqué la théorie de l'échantillonnage d'acceptation pour déterminer la concordance ou la discordance entre les résultats obtenus par sondage et les données provenant de sources secondaires.

L'une des difficultés, les plus couramment rencontrées a trait à l'insuffisance de l'analyse et à l'incapacité à déceler le sens profond des relations et conclusions tirées de l'analyse des données, que celles-ci proviennent de sondages ou de recensements.

Les données obtenues par échantillonnage sont analysées et interprétées ci-après, afin de faire ressortir les caractéristiques et conséquences les plus importantes de la méthode de sondage utilisée pour la présente enquête.

### **I.3.4. Erreurs d'échantillonnage**

Dans la terminologie statistique, l'erreur désigne la différence entre les valeurs observées et les valeurs véritables. L'erreur peut également désigner la différence entre la valeur d'un paramètre dans la population étudiée et son estimation au moyen de l'échantillon. La valeur effective du paramètre s'obtient par observation de la population et son estimation par observation d'un échantillon. Dans ce dernier cas, il existe forcément des erreurs, dont les causes sont multiples. Des indicateurs statistiques tels que l'erreur absolue moyenne, l'erreur-type moyenne, l'écart-type ou l'erreur-type permettent d'évaluer l'ampleur de ces erreurs au moyen de tests de signification relatifs aux différences entre valeurs effectives et valeurs observées des paramètres. Ces tests, qui se fondent sur la loi normale, permettent de déterminer si l'on peut escompter que la valeur d'un paramètre donné estimée à partir de l'échantillon sera représentative de la valeur du même paramètre dans la population.

### **Certaines causes des erreurs d'échantillonnage**



Les erreurs d'échantillonnage peuvent s'expliquer par divers facteurs : mesures approximatives, erreurs d'approximation des chiffres calculés ou arrondis, distorsions au stade de la collecte et de l'analyse des données, distorsions propres aux enquêteurs et aux enquêtes, voire ignorance ou analphabétisme de ces derniers, etc. A titre d'exemple, les habitants des campagnes sont parfois incapables de dire leur âge ou celui de leurs enfants.

### **Types d'erreurs**

On distingue de manière générale deux grands groupes d'erreurs, selon qu'elles sont dues ou non à la technique des sondages. Seules les erreurs d'échantillonnage seront examinées succinctement ici, de même que les principaux problèmes liés aux méthodes d'estimation.

IL ne s'agit donc ici que des erreurs propres aux enquêtes par sondage et dont les recensements ne sont généralement pas entachés. Dans toute enquête par sondage, quelle que soit la taille de l'échantillon, le type de sondage et la méthode employée, une petite partie de la population seulement est examinée. On ne peut savoir avec certitude si telle ou telle unité particulière sera ou non retenue dans l'échantillon. L'on risque aussi de retenir dans l'échantillon des unités atypiques et d'exclure des unités plus représentatives de la population. Les résultats de l'échantillonnage sont donc rien moins que sûrs et sont même nécessairement différents de ceux qu'aurait donnés un recensement. D'un point de vue statistique, les erreurs de ce type sont expliquées par les fluctuations de l'échantillonnage, d'où leur appellation d'erreurs d'échantillonnage. Une enquête pilote contient inévitablement des erreurs de ce type.

### **Méthode d'échantillonnage et erreurs**

L'une des grandes causes des erreurs d'échantillonnage réside dans la mauvaise sélection de l'échantillon, Dans le cas de la présente enquête, la méthode adoptée était celle de l'échantillonnage systématique stratifié/en grappes à plusieurs degrés. La sélection systématique d'un échantillon aléatoire a été adoptée pour sélectionner les écoles au troisième degré de l'échantillonnage. Au premier degré, les Etats ont fait l'objet d'une sélection raisonnée; à cette fin, une stratification a été opérée en fonction du niveau de développement de l'éducation, mesuré par l'indice d'alphabétisation de chaque unité par rapport à la moyenne des taux d'alphabétisation pour l'ensemble de l'Inde. On a procédé ensuite à la sélection des districts et des cantons, stratifiés selon des critères analogues. Si des unités avaient été tirées dans la totalité des Etats et dans la quasi-totalité des districts de chaque Etat, les deux premiers degrés de l'échantillonnage n'auraient pas eu de raison d'être. Le choix de la procédure de sélection tient en fait à plusieurs raisons : commodité administrative, contraintes budgétaires, calendrier des travaux et rationalité relative de la procédure en égard à la situation à la base. S'il est difficile d'exclure que l'échantillon soit entaché de distorsion involontaires, on peut penser qu'il a permis d'obtenir des résultats raisonnablement fiables.

On aurait pu éviter et l'échantillonnage par choix raisonné et la stratification en s'en tenant à l'échantillonnage aléatoire pour les trois premiers degrés. Les objectifs concrets de la présente étude et les limites et contraintes imposées par divers facteurs ont empêché qu'il en fût ainsi. Quel que soit son bien-fondé sur le plan théorique, l'échantillonnage aléatoire simple n'est pas toujours applicable dans la pratique. Ainsi, si l'on avait choisi au hasard

quatre ou cinq Etats ou Territoires au sein de l'Union indienne, l'échantillon aurait pu comprendre les Etats et Territoires suivants : Orissa (est), Arunachal Pradesh (nord-est), Goa (ouest), Rajasthan (nord-ouest) et Jammu-et-Cachemire (nord). Les déplacements auraient été plus longs et plus coûteux, les difficultés administratives plus nombreuses et le calendrier d'exécution de l'enquête beaucoup plus difficile à respecter, sans pour autant que l'échantillon soit plus représentatif ou les résultats plus précis. La sélection des cantons au hasard n'était pas davantage possible, vu le très petit nombre de cantons dans chaque district, si l'on voulait qu'à chaque tirage la probabilité de sélection fût la même pour tous les cantons. L'échantillonnage aléatoire simple est en tout état de cause exclu lorsque la population étudiée est très différenciée, comme c'est le cas ici, vu les très grandes disparités qui existent entre les unités d'échantillonnage aux trois premiers degrés du sondage. Les conditions locales empêchent également que le nombre des écoles des zones rurales soit très important à chaque tirage. Le nombre limité d'unités spatiales à chaque degré du sondage fait que la probabilité de sélection varie d'un tirage à l'autre.

Au cours de l'enquête sur le terrain, un problème d'accès à certaines unités d'échantillonnage s'est inévitablement posé, en ce sens que des établissements sélectionnés étaient fermés, pour une raison ou une autre. En pareille situation, il faut remplacer l'unité d'échantillonnage sélectionnée par une autre unité située dans le voisinage. Ces unités de substitution faisant également partie de la même population mère, il aurait pu en résulter des distorsions imputables aux différences de caractéristiques entre l'unité initialement sélectionnée et l'unité de substitution. Toutefois, le remplacement des unités défaillantes a été effectué avec beaucoup de soin; en règle générale, on s'est efforcé de choisir des établissements ayant des caractéristiques similaires et situés dans la même circonscription scolaire.

### **Erreurs portant sur des populations nombreuses et hétérogènes**

Il peut également arriver que les erreurs d'échantillonnage soient dues au degré d'hétérogénéité de la population étudiée. Dans la présente étude, on s'est efforcé de limiter les risques d'erreurs de ce type en procédant à une stratification en fonction de plusieurs critères, à tous les degrés ou presque, et à une sélection systématique au dernier degré de l'échantillonnage.

### **I.3.5. Tests de signification**

Une autre série d'erreurs peut être imputable à la méthode d'échantillonnage ou à la méthode d'estimation des paramètres de la population lorsque les procédures adoptées ne remplissent pas les conditions requises. Dans le cas présent, des tests de signification ont été appliqués séparément aux quatre Etats retenus dans l'échantillon afin de déterminer si les estimations tirées de l'échantillon comportaient une différence significative par rapport aux valeurs correspondantes tirées des sources secondaires pour ce qui est de la proportion respective des jeunes filles et des élèves provenant des castes et tribus "énumérées" (CE/TE) dans les effectifs globaux de l'enseignement primaire. Les données provenant de sources secondaires à deux moments distincts sont considérées comme étant les paramètres/valeurs des caractéristiques considérées dans la population. Les projections de ces valeurs pour 1990-1991, année de l'enquête, ont été établies à partir des taux de croissance observés, pour permettre la comparaison des valeurs de l'échantillon et de la population. Ces projections, assimilées à des estimations de la population, ont été comparées aux valeurs estimées par sondage pour chacun des districts, ensemble et séparément. Les valeurs estimées pour la population (ou valeurs-population) étant elles-mêmes des projections, elles peuvent s'avérer en définitive majorées ou minorées par rapport aux valeurs aux valeurs effectives. Les résultats des tests de signification sont donc à prendre comme une indication du sens des erreurs et non de leur ampleur précise.

### **Déductions erronées**

Certaines erreurs sont liées aux déductions statistiques concernant la vérification des hypothèses, qui a pour objet d'obtenir, à partir des informations données par l'échantillon, des connaissances sur les aspects inconnus de la population étudiée. Il s'agit en l'occurrence de déterminer si telle hypothèse sur laquelle on s'est fondé est vraie. Dans le cas présent, on est parti de l'hypothèse que la proportion de jeunes filles dans l'effectif total des classes primaires des établissements retenus dans l'échantillon se retrouverait dans l'ensemble de la population. La même hypothèse a été adoptée en ce qui concerne les effectifs des élèves CE/TE, etc. Par ces hypothèses, que l'on appelle hypothèses nulles, on part du principe qu'il existe des différences entre les valeurs réelles et les valeurs estimées. A l'opposé de l'hypothèse nulle, il y a l'hypothèse alternative. Pour vérifier ces hypothèses jumelées, il faut définir le critère du test. On peut alors séparer l'ensemble de la population en régions où les hypothèses sont acceptées et régions où elles sont rejetées. Le critère du test est à caractère aléatoire, d'où le risque, aussi faible soit-il, d'en tirer des conclusions erronées. L'erreur commise ce faisant peut-être de première espèce, lorsque le test conduit à rejeter une hypothèse nulle qui est en réalité vraie (la condamnation d'un innocent en quelque sorte), ou de deuxième espèce, lorsque l'hypothèse nulle est acceptée alors qu'elle est fautive (le coupable est laissé en liberté).

### **I.3.6. Caractéristiques d'une bonne estimation**

Les estimateurs de la population doivent avoir les caractéristiques suivantes : (i) absence de distorsions; (ii) variance minimale; (iii) erreur quadratique moyenne minimale; (iv) efficacité; (v) meilleure estimation linéaire sans distorsion; et (vi) exhaustivité. Les tests statistiques permettent de déterminer la validité des statistiques d'échantillonnage en fonction des caractéristiques susmentionnées qui définissent un bon estimateur.

L'analyse qui précède a permis d'indiquer certains des problèmes et limites de la théorie et des méthodes des sondages appliqués à des situations concrètes, ainsi que les grandes lignes, la nature et le déroulement de l'enquête pilote. On examinera dans l'analyse ci-après, les résultats généraux du sondage obtenus par application empirique du plan et des méthodes retenus.

## II.3. Méthodes d'échantillonnage

### Certains concepts de probabilité et statistiques

#### Variables:

Le concept de variable concerne la dimension ou l'aspect qu'on veut mesurer et qui varie selon le sujet et le cas. Dans l'étude sur les caractéristiques sociologiques d'un groupe quelconque, par exemple d'une cité, l'appartenance au groupe, à la cité dans ce cas, serait la constante ; le sexe, le niveau d'instruction, le revenu, c'est-à-dire la dimension à mesurer serait la variable.

On retient les suivants types de variables :

*Variable quantitative*: La valeur de la variable est toujours numérique. Ex. combien de livres lus par an ?

*Variable qualitative*: Par exemple la couleur de la peau, l'état civil, l'affinité politique...

Il est important de savoir qu'il y a deux différents types de statistique, celle qui concerne les variables quantitatives et celle concernant les variables qualitatives.

On peut toujours convertir une variable quantitative en une variable qualitative, mais le contraire n'est pas possible. Néanmoins, en sociologie souvent on fait la conversion au moyen des variables d'intervalle, de manière à pouvoir utiliser les techniques de la statistique quantitative.

Exemple: Variable d'âge: 1 - 60 ans: 1-10: 1  
11-20: 2  
21-30: 3, ecc.

*Variable continue*: Il y a des valeurs intermédiaires. Par exemple : la hauteur, entre 1,71 cm et 1,72 on trouve des valeurs telles que 1,711, 1,712... La variable est continue ou non en fonction de l'unité de mesure choisie. Le niveau de revenu calculé en dollars sera une variable continue quand on se réfère aux centimes. Mais elle sera discontinue si l'unité est le dollar.

*Variable discrète ou discontinue*: Il n'y a pas de valeurs intermédiaires. Par exemple le nombre d'enfants : 1,2,3,4... On ne saurait avoir 1,2 fils. La mesure discrète ne prévoit pas de fraction entre les valeurs. Il s'agit en effet plus d'un comptage que d'une mesure. Quand on demande par exemple combien de fois on est allé à un théâtre par an.

La distinction entre variables quantitative et qualitatives s'utilisent en Statistique et dans les sciences expérimentales. Dans le cas des sciences sociales, on utilise plus un autre critère de classification: Variables nominales, ordinales, d'intervalle et de rapport. Ce sont en réalité de niveaux de mesure (mesuration).

*Variable nominale (categorical data)*: N'importe quelle mesure qualitative est nominale. En réalité c'est un système de classification qui consiste à mettre les cas dans des catégories.

Ex: Pour quel parti politique voterez-vous ? -Libéraux, -Conservateurs, - Radicaux, -Verts, -Je ne voterai pas, -Je ne sais pas.

Les catégories doivent être distinctes, exclusives et exhaustives. Elles ne doivent pas se superposer.

*Variable ordinale*: Quand les catégories sont ordonnées hiérarchiquement et ne sont pas au même niveau. Ex. : Ordonner les équipes de football selon le plus grand nombre de matches gagnés.

La majorité d'échelle d'attitude produit des données ordinales.

Ex: Selon vous, les médias sont:

1	2	3	4	5	6	7
Très impartiaux				Très partiels		

Ces échelles, très utilisés, ont certaines caractéristiques. Par exemple, l'échelle droite-gauche ou gauche droite.

1	2	3	4	5	6	7
---	---	---	---	---	---	---
Gauche				droite		

*Variable d'intervalle*: Nous avons une unité de mesure qui est la même pour tous les points de l'échelle. La distance entre les valeurs paires est égale. Par exemple la hauteur. La différence entre deux personnes qui ont une hauteur de 1,75 cm et 1,70 cm respectivement est la même que celle entre deux personnes de 1,60 et 1,55 cm.

Les mesures d'intervalle nous permettent de faire des opérations mathématiques de sommation ou de soustraction...

En sciences sociales, les variables ordinales sont considérées comme si elles étaient des variables d'intervalle. Ceci fait qu'on applique beaucoup de techniques statistiques valides pour les variables d'intervalle que autrement on ne pourrait utiliser. La discussion reste ouverte. Il ne faut pas oublier tout de même que les différences réelles persistent : la différence psychologique entre gagner 720.000\$ et 700.000\$ n'a certainement pas la même portée que gagner 70.000 et 50.000\$.

*Variable de rapport* ("ratio level"): Le poids, l'âge ... sont des variables de rapport. Le niveau de mesure de rapports permet d'affirmer qu'une valeur est le double, le triple... par rapport à un autre, c'est-à-dire permettant de multiplier ou de diviser, et ceci implique qu'il y ait dans l'échelle une valeur zéro absolue, non arbitraire.

Ex.; Les échelles de température Fahrenheit et le centigrade sont d'intervalle, parce qu'on ne peut pas affirmer que 30° C est 3 fois plus chaud que 10° C. L'échelle Kelvin est de rapport puisqu'elle mesure le mouvement des molécules: 0°K, le zéro absolu représente la température dans laquelle il n'y a pas de mouvement moléculaire...; 32°F et 0°K di rapporto poiché misura il movimento delle molecole: 0 ° K, lo zero assoluto rappresenta la temperatura nella quale non c'è nessun movimento molecolare...; 32° F e 0° C constituent

la température à laquelle l'eau devient gelée, il s'agit donc d'une convention par rapport à un phénomène physique.

Comme on peut déduire de l'exposition synthétique sur les sortes de mesure, chaque niveau supérieur conserve les caractéristiques du niveau antérieur et ajoute quelque chose en plus. Le niveau d'ordinal, outre le fait de classer les cas, les ordonne du plus petit au plus grand ou vice versa. Le niveau d'intervalle classe et ordonne les cas qui peuvent être comparés entre eux grâce au fait que l'unité de mesure de l'échelle est identique.

*Probabilité d'un événement: C'est le quotient qui dérive de la division du nombre de cas favorables par le numéro de cas possibles.*

Par exemple  $1/6$  est la probabilité de toucher la cible avec le tir à l'arc.

*Loi des grands nombres: Si un événement se répète plusieurs fois, la fréquence de l'événement se rapproche de sa probabilité jusqu'à s'identifier avec elle. Dans l'exemple du tir à l'arc,  $p$  de n'importe quelle valeur est de  $1/6$ , par exemple le "deux". Si nous répétons plusieurs fois le tir, il en dérivera une sixième part de ses faces.*

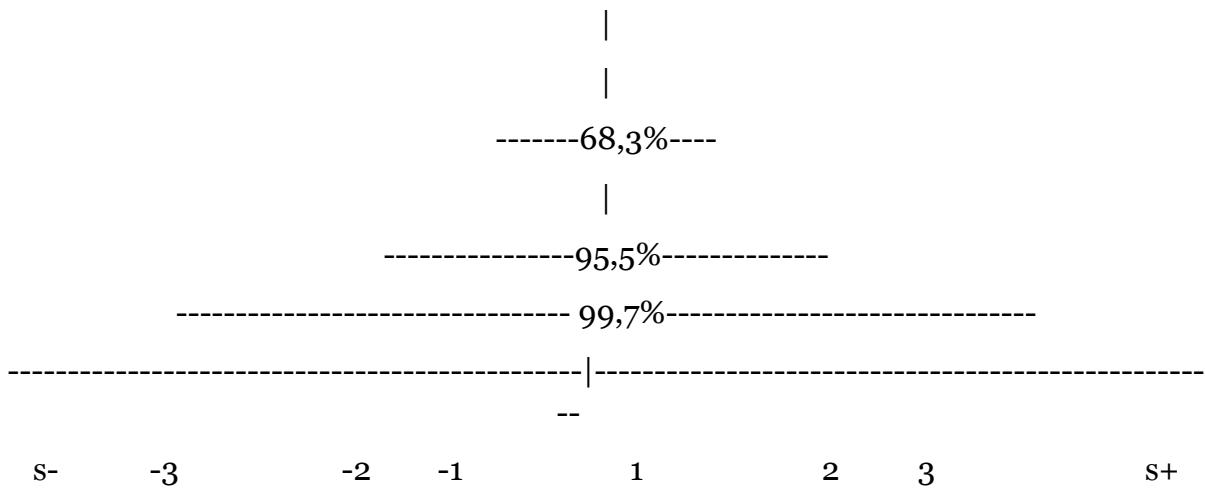
En outre, il y a un aspect de la loi des grands nombres qui concerne le nombre de fois qu'il faut répéter l'expérience. On en arrive à un nombre d'unité de l'échantillon, qui ne dépend pas de la grandeur de l'univers de la population, pour lequel (nombre), même si le nombre d'unité de l'échantillon augmente, le pourcentage d'erreur ne diminue pas de manière considérable. C'est à un certain niveau, cela ne sert pas à augmenter l'échantillon parce que les différences entre fréquence et probabilité ne diminuent pas en proportion aux augmentations de l'échantillon.

Dans l'exemple de tir, la probabilité de tirer juste sur 6:

Tir	fréquence (f)	probabilité (p)
6	0 (0)	$1/6$ (0,166)
12	4 (0,33)	$2/12$ (0,166)
24	2 (0,08)	$4/24$ (0,166)
60	8 (0,13)	$10/60$ (0,166)
120	18 (0,15)	$20/120$ (0,166)
600	104 (0,17)	$100/600$ (0,166)
6000	1002 (0,165)	$1000/6000$ (0,166)







Que signifie cette courbe? Comment la lire? On sait en mathématique que l’aire, la superficie incluse sous la courbe de la fonction entre les limites 1 et -1 est justement 68%, 1 étant justement la valeur de la déviation standard ; on sait aussi que l’aire entre le double de la déviation standard est justement 95,5% du total ; et on sait enfin que la superficie entre le triple de la déviation standard est justement le 99,7% du total. Par exemple, dans un échantillon de 100 malades avec une moyenne de 80 pulsations par minute et un écart quadratique de 10 p/m, nous savons que 68 auront une pulsation de  $80 \pm 10$ ; 95 se trouveront entre 60 et 100 pulsations par minute et 99 se trouveront entre 50 et 110.

D’autre part, sigma est justement le point d’inflexion, où la courbe change de concave en convexe.

Donc une fois la déviation standard se trouve le 68% des cas; deux fois la valeur sigma, le 95% des cas; et 3 fois le 99% des cas. Ce sont aussi les valeurs du coefficient de fiabilité de l’intervalle d’estime appelé aussi niveau de probabilité significativité. (LA de 68% = 1 sigma; LA de 95% = 2 sigma; LA de 99% = 3 sigma).

## **Théories des échantillons: l'échantillonnage.**

Parler de théorie des échantillons veut dire se référer à la notion complexe de techniques qu'il faut suivre afin qu'un échantillon représente bien son univers. La théorie des échantillon affirme qu'un grand nombre de cas choisi au hasard d'un univers, présente presque certainement les mêmes caractères dudit univers.

### **Quelques notions :**

- **Univers ou population** : C'est l'ensemble des sujets, des cas ou des éléments dont on veut mesurer les caractéristiques.

- **L'échantillon** : L'ensemble des éléments de la population choisis pour obtenir la mesure estimée. On appelle échantillonnage, le processus mise en place pour fixer un échantillon qui ne soit pas déformé. L'échantillon est représentatif quand il est bien choisi, et ne s'éloigne donc pas de l'univers réel.

- **Paramètres de l'échantillon** : Ce sont des observations ou données qui définissent les caractères d'un échantillon. Ils nous fournissent des mesures estimées de la population. Par exemple, les résultats électoraux sont un paramètre de la population. Par contre l'intention de vote qu'un sondage nous offre est un paramètre estimé – qui a une marge d'erreurs – de l'intention de vote de la population.

- **Unité statistique de l'échantillon** : Il s'agit de chacun des éléments qui interrogent l'échantillon. Parfois ce ne sont pas les personnes, mais les entités comme les familles, les sociétés, les clubs, les villages, les cités.... On ne devra donc pas confondre l'unité d'analyse ou d'échantillonnage avec l'élément de la population. Ex. : Dans une enquête sur la fécondité des femmes d'un village donné, l'unité d'analyse est la femme (mariée) à l'âge de procréer.

- **Base d'échantillon** : D'habitude, les éléments ou les sujets de l'échantillon ne sont pas tirés des listes des composants de l'univers : Ceci n'est pas toujours attribuable parce qu'on n'a pas toujours des listes et ce serait aussi très coûteux. Pour un sondage électoral par exemple, on ne prend pas l'échantillon de la liste électorale, mais on choisit proportionnellement les sujets des couches qui composent la population et qui sont connues : sexe, âge, niveau d'étude, démographie, ... La base de l'échantillonnage est donc le support matériel utilisé par le chercheur pour former un échantillon : listes, catalogues, cartes ... Attention ! Toutes les bases ne garantissent pas la même fiabilité. Si la base n'est pas bien choisie, la représentativité diminue. Ceci arrive quand par exemple on utilise la liste téléphonique pour composer l'échantillon. Tous les électeurs (votants) n'ont pas de téléphone. Elaborer la liste d'échantillonnage est assez complexe. Même dans les cas où les données de la population sont connues, il est difficile d'élaborer une liste complète. Ceci exige que les listes des recensements soient continuellement mises à jour (décès, nouvelles incorporations, transfert de domiciles ... Malgré le travail des bureaux de statistiques (dans les pays développés surtout), de temps en temps les plaintes des partis politiques sur les erreurs aléatoires.

## Techniques d'échantillonnage:

Il convient de distinguer deux grandes modalités d'échantillonnage: probabiliste et non probabilistes.

### 1) Echantillonnage probabilistes:

On l'appelle aussi aléatoire ou casuel. Il se base sur le principe suivant: tous les sujets de la population ont la même possibilité de former une partie de l'échantillon. Par conséquent, le choix est complètement aléatoire. Ce n'est sûrement pas la seule manière valide, mais c'est une modalité qui nous permet de bien contrôler les erreurs à l'intérieur de certaines marges et à l'intérieur d'un niveau de fiabilité.

On distingue:

#### a. L'échantillonnage aléatoire simple (Sondage Aléatoire Simple, SAS):

Chacune des unités de l'échantillon est une unité de la population. C'est un tirage au sort. Tous les éléments de la population sont tirés au sort. Il n'y a pas de réinsertion, on ne réintroduit pas dans la liste des sujets qui en sont extraits. Si la dimension de la population est relativement grande, le fait de ne pas remplacer les éléments ne pose aucun problème. Dans une liste de 1000 personnes, une fois on extrait 200, la probabilité du reste n'est pas de 1/1000 mais de 1/800. Dans l'échantillonnage aléatoire simple, il faut une liste complète de l'univers. On ne l'utilise pas souvent, parce que ce serait très coûteux et incommode, surtout dans le cas des sondages électoraux.

Cette situation ne reflète que rarement la réalité : dans la pratique, on essaie de restreindre le nombre de combinaisons en évitant celles qui seraient a priori non souhaitables.

Ce sondage fournit un cadre de référence indispensable pour deux raisons :

- Il sert d'étalon, car par ses propriétés on peut juger les autres modèles d'échantillonnage ;
- Il constitue en général « la brique » élémentaire des plans usuels ; par exemple, les sondages stratifiés et les sondages à deux degrés sont des assemblages de sondages simples.

La taille de la population de référence et le taux de sondage n'ont, contrairement aux idées reçues que peu d'influence sur la précision

b. **Echantillonnage aléatoire systématique:** C'est simplement une variation du précédent en ce qui concerne la manière de choisir l'unité de l'échantillon. On obtient avant tout le "coefficient de pas ou de saut" ( $K=N/n$ ) et on choisit casuellement un numéro  $x$  inférieur ou égal à  $K$  pour commencer et puis  $x+K$ ,  $x+2k$ ,  $x+3k$  et ainsi de suite jusqu'à atteindre le nombre d'unité de l'échantillon.

c. **Echantillonnage aléatoire stratifié:** On stratifie l'échantillon de manière proportionnelle à la stratification de l'univers selon les facteurs connus de la population (sexe, niveau de revenu, d'instruction, etc). Le critère de stratification le plus habituel est la dimension de l'habitat. Les groupes qui composent les couches ne doivent pas se superposer. La population est regroupée par couches et puis, dans chaque couche, on extrait un échantillon aléatoire. On peut faire la stratification avec deux ou plusieurs variables au

même moment, par exemple l'âge, le niveau d'instruction ... Il est assez fréquent trouver dans les fiches techniques que publient les journaux (quand ils le font), l'indication de « stratifié à double stade » ou « échantillonnage aléatoire stratifié à plusieurs stades ». Ceci signifie qu'une fois attribué à chacune des couches le nombre d'interviews qui correspondent relativement au facteur de stratification, parmi les grappes qui forment la couche, on en choisit un et on n'en tire l'un des éléments qui le constituent ; dans ce cas, on aura un échantillon stratifié à double stade. Si on fait le choix par hasard plusieurs fois, cela devient échantillon stratifié à plusieurs stades. Par exemple, une cité, parmi tant d'autres, une rue parmi toutes les rues, une maison parmi les autres maisons de la rue... ainsi de suite jusqu'à arriver au sujet (personne) à interviewer. La fixation de l'échantillon : Il s'agit de la distribution du nombre d'interviews par couche selon la population. Cette assignation peut être simple (même nombre d'unité dans chacune des couches, et donc pas aléatoire), proportionnelle (selon la population de la couche) ou optimale (quand elle se fait en tenant compte de la déviation standard ou l'écart quadratique moyen). Souvent on utilise la proportionnelle.

d. **Echantillonnage aléatoire à grappes:** L'idée de grappe est différente de celle de couches. La couche est un agrégat homogène d'éléments. La grappe est par contre un agrégat hétérogène. Une grappe est une boîte de vote qui dans un certain sens représente l'hétérogénéité de l'univers, de la population, tandis que la couche regroupe les sujets avec les mêmes critères. L'échantillonnage à grappe permet d'épargner temps et argent. Mais il est possible de se tromper dans le choix des grappes et dans l'évaluation des poids à donner au moment de l'analyse des données. Pour améliorer la représentativité, le chercheur profite de la pondération de l'échantillon. Il s'agit de réduire le poids d'un groupe quand ledit groupe est surreprésenté dans l'échantillon. Parfois la pondération sert pour masquer les défauts d'un échantillon.

## 2) Echantillonnage non probabiliste:

a. **Echantillonnage accidentel:** se réalise en interviewant la première personne que le l'enquêteur trouve. Par exemple: les téléphones ouverts dans les stations de radio ou de TV ; interviews dans la rue. Cet échantillonnage est commode et économique, mais il souffre d'un manque de représentativité.

b. **Echantillonnage par quotas:** Très fréquent dans le cas des études du marché. Il se base sur une connaissance de la population, que l'on divise en groupes ou sous-groupes en accord avec les principales variables démographiques de la population, facilement contrôlable par l'enquêteur : âge, sexe, zone de résidence, profession ... L'enquêteur emmène avec soi en général un profil des personnes à interviewer établi selon ces facteurs connus et selon les instructions reçues du conducteur de la recherche. En clair, il doit éviter les personnes qui s'éloignent de la norme (les personnes qui n'inspirent pas confiance, les résidences avec chien ... De cette manière on renforce la représentativité. Malgré cela, cet échantillonnage a le désavantage d'être laissé à l'arbitraire de l'enquêteur. Dans ce genre d'échantillonnage, outre la connaissance des groupes qui forment la population globale, il faut donner (attribuer) un poids à chacun des groupes.

c. **Echantillonnage par objectif:** On l'appelle aussi échantillonnage à "choix raisonné". Il est utilisé par les chercheurs dans des enquêtes sociologiques et pédagogiques

de petits groupes. Le choix se fonde sur la connaissance que le chercheur a de la population en question.

d. **Echantillonnage à cluster:** Quand l'objet de l'enquête est un phénomène particulièrement limité en extension dont il n'est pas possible de disposer d'une banque de données. Un cluster est un groupe d'unités semblables ou proches entre elles du point de vue de la position et de la composition. Ex. 200.000 conducteurs handicapés sur une population de 20.000.000.

e. **Echantillonnage par "opinion leaders":** Quand on veut étudier un phénomène dans certaines catégories particulières de personnes ou de professions: les directeurs des Isp, les directeurs des journaux ... par exemple.

### 3) **Autres méthodes empiriques de sondage:**

- **Le volontariat :** Il s'agit d'enquêtes réalisées auprès de lecteurs de tel ou tel journal, d'adhérents d'association acceptant de répondre à un questionnaire, des téléspectateurs acceptant de répondre par téléphone à une ou plusieurs questions. Le résultat d'un tel sondage doivent être traités avec beaucoup de précaution et ne peuvent être extrapolés à toute la population.

- **La méthode des itinéraires (ou *random route*)** dans laquelle on impose à l'enquêteur un point de départ dans une commune et un itinéraire à suivre, avec tirage systématique des logements dans lesquels il effectuera les interviews. Cette méthode a pour objectif de reproduire un certain tirage aléatoire des enquêtés mais sans donner explicitement des noms et adresses à l'enquêteur.

- **L'échantillonnage sur place**, c'est-à-dire le lieu d'achat ou d'activité lorsque la population étudiée est définie par son activité : clients d'un centre commercial, d'une chaîne de restaurants... Dans ce cas, on doit déterminer les endroits où enquêter, les périodes d'enquête, les pondérations a posteriori pour tenir compte de la probabilité de présence des répondants. Cet échantillonnage est une des techniques utilisées dans l'échantillonnage des populations mobiles ou rares.

- **L'utilisation des panels de consommateurs :** par exemple, la taille (20.000 ménages) du panel Mégascope de la Sofres en France permet parfois d'identifier un nombre suffisant de membres d'une sous-population étudiée.

- **L'utilisation du téléphone** permet aussi de filtrer des populations particulières. Cependant, comme dans le cas de panels postaux, si l'enquête doit être réalisée en face-à-face par enquêteur, cette méthode peut conduire à des adresses très dispersées et donc à des coûts d'enquête important.

## Erreurs dans la réalisation des sondages

Il existe deux types d'erreurs: l'erreur aléatoire ou naturelle et les erreurs systématiques.

### *Erreur aléatoire*

C'est une erreur qu'on peut calculer mathématiquement (on l'appelle aussi erreur d'échantillonnage). Chaque échantillon, puisque choisi aléatoirement, présente une différence entre ses caractéristiques et celles de la population. Le contraire n'est qu'une pure coïncidence.

La précision d'un statisticien est l'exactitude avec laquelle les données statistiques produisent leur paramètre. Quand ceci n'est pas connu, on parle de fiabilité qui est la constance d'une donnée statistique obtenue sur différents échantillons. Normalement on calcule sur un seul échantillon parce que ce n'est pas rentable le faire sur plusieurs échantillons.

Soit **P** le paramètre estimé d'une population : ceci équivaut à la somme/soustraction de l'erreur de l'échantillon(**e**) et de la donnée statistique (**p**) :

$$P = p \pm e$$

L'erreur de l'échantillon sur des populations d'un nombre infini (supérieur à 100.000) est donné par la formule suivante:

$$e = s \times \sqrt{\frac{p \times q}{n}}$$

- "s" (sigma) è la mesure ou le niveau de fiabilité ou coefficient de confiance de l'intervalle d'estime ou un niveau de probabilité qu'un échantillon représente dans sa population. Ses valeurs sont: s = 1 là où il y a un niveau de fiabilité de 68,3%; s = 2 là où il y a un niveau de fiabilité de 95,5%, et s = 3 où le niveau de fiabilité est de 99,7%. Ceci signifie que nous pouvons extrapoler ou projeter des données respectivement au 68,3%, 95,5% e 97,7% de la population.

- "n" est le nombre d'éléments qui font partie de l'échantillon.

- "p" e "q" sont des valeurs de pourcentage déjà données: normalement ils sont chacun de 50%. Ceci s'explique avec le fait qu'il faut assumer comme paramètre le pire des cas possible pour calculer l'erreur typique du pourcentage (toutes les questions des enquêtes divisent l'échantillon de l'univers de la même manière, en deux groupes parce que la réponse a au moins deux alternatives – oui ou non – et à plus forte raison dans le cas des sondages électoraux, parce qu'il y a au moins deux partis en lice). Le produit majeur qu'on obtient des deux nombre dont la somme est 100, est 50 x 50, pour cela, celle-ci est la pire des situations possibles.

L'erreur de l'échantillon sur des populations finies avec moins de 100.000 sujets se calcule avec la formule suivante :

$$s_p = \sqrt{\frac{p \times q}{n} \times \frac{N - n}{N - 1}}$$

### ***Erreurs systématiques***

Ces erreurs ne sont pas contrôlables et sont celles qui expliquent les fautes dans les prévisions des sondages électorales quand celles-ci vont au-delà d'une marge naturelle d'erreur. Toutefois, il faut considérer que ces erreurs de prévision dans les sondages électoraux peuvent aussi être dues aux changements d'opinion parmi les électeurs suite à certains événements advenus entre le moment de réalisation des sondages et les élections, lesquels événements peuvent avoir une certaine influence.

En considérant les phases du sondage, les erreurs systématiques auxquelles on peut faire face peuvent être synthétisées en ces termes :

- Des erreurs dans la construction de l'enquête (mauvais choix de l'échantillon, base de l'échantillon inadéquate, remplacement des interviews finales sur base des critères arbitraires, dispersion de l'échantillon qui ne garantit pas une représentation adéquate...)
- Des erreurs dans la phase de l'élaboration du questionnaire
- Les erreurs de celui qui fa les interviews (conditionnement, sournois ou évident)
- Erreur dans la mise en tableau, pendant la codification, pendant la procédure et le rapport.

## II. Sondage d'opinion publique en contexte électoral

### II.1. Notion et définition

Qu'est-ce que l'opinion publique ?

La chercheuse Noëlle-Neumann a mis les phares sur une notion appelée « climat d'opinion ». Chacun de nous a un sens presque statistique pour comprendre quelles sont les opinions qui circulent et qui gagnent de l'importance dans le domaine publique.

Elle perçut, en Allemagne, qu'une chose était demander qui on voterait, et l'autre est le résultat ou la réponse à la question « qui est-ce qu'on croyait qu'il aurait vaincu aux élections ? » ( climat d'opinion ). Ce climat affirme un mécanisme pas seulement rationnel, mais aussi psychologique.

Cette théorie l'a conduite au « spirale du silence », théorie qui fait référence aux opinions minoritaires qui, tout en donnant l'impression de disparaître dans l'opinion publique, existent en réalité. En faisant des recherches sur une telle attitude dans le passé, elle trouve un psychologue Asch, auteur de la théorie de conformisme. Selon Asch, si 9 sur 10 personnes sont d'accord sur quelque chose de faux, la 10<sup>ème</sup> personne finit par donner un point de vue erroné tout simplement pour se conformer à la majorité.

La spirale de silence est une réaction aux changements du climat d'opinion, qui répond à la peur d'isolement.

Il existe deux phénomènes très liés à l'opinion publique qui n'ont pas de connotation directement rationnelle : la mode et la censure.

La mode est une forme de comportement qui doit être exécutée pour ne pas tomber dans l'isolement.

L'opinion publique est bien différente de l'opinion publiée. La soit disant théorie de la cultivation, la fameuse théorie de « knowledge gap », la théorie de « l'agenda setting » et la théorie de la « spirale du silence » ... font partie de ce phénomène.

Les médias confluent dans la formation de l'opinion publique. Par rapport à celle-ci et de ses changements, les médias interviennent de plusieurs manières :

1. La minorité, soutenue par les médias, ont le courage de maintenir leur position
2. La présence intense et l'explicitation des médias ont de l'influence dans le domaine public
3. Ceux qui soutiennent une position majoritaire, avec le temps deviennent faibles s'ils ne trouvent de contrastes publiquement, perdant ainsi l'opportunité de discuter leur point de vue de manière ouverte.

Les médias créent-ils l'opinion publique ?

1. Ils fournissent une pression environnementale à laquelle les personnes répondent ou qu'elles accueillent avec le silence.
2. Ils constituent une des sources d'observation de l'opinion publique

Une institution doit faire attention, non pas à l'opinion publiée, mais bien à l'opinion publique. Les médias sont ambivalents par rapport à l'opinion publique. Ils sont facteurs de consensus social. Ils s'appuient sur l'ordre social courant et le reproduisent. Ils s'opposent



par contre à la vraie religion, aux vraies valeurs. Il est important d'enlever les obstacles. Et ceci est le rôle de la communication institutionnelle. Il ne faut pas sous-estimer le pouvoir des médias.

Le journalisme doit couvrir une fonction de surveillance et de corrélation entre l'élite et le public.

## **II.2. Les systèmes électoraux**

*Démocratie directe*: les décisions sont prises par les citoyens eux-mêmes, sans aucune intermédiation. Ceci est pour Jean Jacques Rousseau la vraie démocratie, parce que chaque représentation implique une certaine aliénation de la volonté populaire. En réalité ce système est presque une utopie qui peut s'appliquer seulement à des petites populations.

*Démocratie représentative*: Il y a des intermédiaires. C'est le système plus habituel.

*Démocratie sémi-directe*: Elle dérive de l'application des mécanismes de la démocratie directe au régime représentatif. Ici on note 3 mécanismes: (1) *Initiative populaire*, ou la capacité d'introduire dans le débat parlementaire des propositions de la part des citoyens, soit en demandant un referendum sur des sujets précis, soit pour faire des propositions de loi. Dans ce dernier cas, la capacité de légiférer reste à l'assemblée. En effet, la proposition devient loi dans l'a même assemblée où réside la souveraineté. (2) *Referendum*: Les électeurs choisissent, en répondant affirmativement ou négativement à des questions de loi. (3) *Plébiscite*: On demande aux électeurs le consensus sur un programme politique ou un personnage politique (Ex. Le plébiscite de la Vème République française sur De Gaulle qui, à la suite des résultats, devait démissionner comme président de la République).

### **Le suffrage**

Etant donné que le régime démocratique actuelle se fonde sur l'élection des représentants, on peut dire que l'histoire de la politique du régime démocratique s'identifie avec l'instauration et la diffusion du suffrage, du premier suffrage restreint ( censitaire et capacitaire) jusqu'au suffrage universel.

*Stricto sensu*, il n'existe pas de suffrage universel absolu, parce que il y a des limitations dont certaines sont injustes et on en parle en termes de restriction et d'autres sont logique comme par exemple le limite d'âge.

Techniquement on parle de restriction, inégalité et déformation.

- *Restriction* : A cause de l'instruction (suffrage capacitaire), à cause du niveau du rendement (suffrage censitaire), les électeurs sont ceux qui paient une taxe directe, recensement. Parfois les restrictions sont liées au sexe. En effet, jusqu'à la seconde guerre mondiale, les femmes n'étaient pas admises au vote. Les limites d'âge sont compréhensibles. L'âge minimum a évolué de 30 à 18 ans dans presque tous les pays du monde. Il existe aussi une limitation par « indignité pénale » qui dérive de la violation d'une norme pénale ou d'une action délictueuse. On comprendrait une telle restriction au candidat (celui qui doit être élu) et non aux électeurs.

- *Inégalité* : C'est la situation qui limite le principe d'égalité de vote entre les citoyens. Aujourd'hui on ne parle plus de vote multiple, vote pluriel ou vote familial. Cependant, il reste encore dans certaines parties de la terre des inégalités qui sont liées à la circonscription

ou dues aux différents systèmes électoraux (uninominal à majorité simple ou à majorité absolue, le système D'HONDT...).

- *Déformation*: Elles concernent les contraintes et les pressions sur les électeurs, les manipulations dans le processus électoral, surtout dans les pays où le régime démocratique n'est pas consolidé.

### **Principes et éléments du système électoral**

#### *1. Le vote doit être universel, égal, libre, secret et direct*

Ceci ne veut pas exclure l'exigence de certaines conditions requises pour exercer le vote, comme par exemple l'inscription au recensement, la pleine capacité juridique, avoir la nationalité, avoir l'âge minimal requis, avoir une résidence reconnue, etc.

Si l'élection est indirecte (comme aux USA par la désignation des candidats par les partis), les électeurs choisissent les **altérants**. Il peut arriver qu'ils ne respectent pas le mandat. Ni même ils s'opposent au principe de l'élection à liste fermée et bloquée, à l'exception des échanges ou remplacement au terme des votes.

Certains pays retiennent que le vote est plus un devoir qu'un droit pour les citoyens. Pour cela son exercice devient une obligation. Ainsi on pousse à la participation au vote.

#### *2. Eléments (composants) du vote*

a) *Corps électoral*: c'est l'ensemble de citoyens ayant le droit de voter et qui sont sur les listes électorales.

b) *Circonscription électorale*: La partie du territoire qui constitue la base pour la désignation des sièges. En général, la circonscription est une division territoriale (à l'exception des émigrés avec droit de vote qui forment une circonscription sans territoire). Parfois, tout le territoire national correspond à une seule circonscription, par exemple le cas des élections du parlement européen où chaque pays est une circonscription électorale. La circonscription peut être *uninominal* (si on élit un seul candidat) ou *plurinominale* (si on élit plusieurs candidats). Il faut bien prendre en considération le problème de la différence de densité de populations dans les différentes circonscriptions électorales qui peut conduire à une inégalité évidente ou inévitable sur le poids du vote.

#### *c) Candidature:*

1. Liste fermée et bloquée: L'électeur peut choisir seulement une liste parmi tant d'autres, telle que présentée, sans la moindre possibilité de modifier quoi que ce soit.

2. Liste fermée et non bloquée: L'électeur choisit l'ordre de préférence parmi les candidats appartenant à une liste.

3. Liste ouverte: L'électeur choisit les candidats les candidats qu'il préfère indépendamment à l'appartenance aux diverses listes.

#### *d) Processus d'élection:*

1. Vote unique.

2. Vote multiple: l'électeur a autant de votes que les sièges attribués à la circonscription.

3. Vote préférentiel: c'est le cas de l'élection avec listes ouvertes

4. Vote cumulable: C'est une variante du vote multiple, parce qu'on peut voter plusieurs fois le même candidat.

5. Vote alternatif: variante du vote préférentiel avec vote unique, où il est permis d'indiquer un second candidat, pourvu que le premier n'ait pas été élu ou ait eu plus des voix en plus.

e) *Règles et systèmes pour l'affectation des sièges*: C'est le nœud de chaque système électoral. Si le système est majoritaire, soit à majorité absolue qu'à majorité simple, il n'y a aucun problème, car il en sort le candidat le plus voté ou la liste la plus votée. Le premier cas c'est à majorité absolue porte souvent à deux tours de vote et cela facilite les coalitions électorales. Ici il n'y a pas de reste. Le problème de restes arrive dans les systèmes proportionnels, parce qu'il y a des votes excédents dans chaque parti et il manque de sièges non assignés dans la première répartition. Il faut donc distribuer les restes. Voyons en détails.

### II.3. Méthode d'affectation de sièges

La première chose à faire est de calculer le quotient électoral. Le quotient électoral est le "coût" (pour ainsi dire) de chaque siège parlementaire. Il s'obtient en divisant le nombre total des votes valables par le nombre de sièges disponibles dans la circonscription. Puis on procède à l'affectation des autres (le reste) selon des méthodes diverses:

*Méthode automatique*: Le pays est considéré comme une unique circonscription. On trouve le quotient électoral. Chaque parti obtient autant de sièges chaque fois qu'il atteint le quotient. On procède donc à une autre division entre leur vote et le quotient électoral, puis on partage les sièges qui restent entre les partis avec grand nombre de restes.

*Méthode de Hare*: C'est comme la méthode précédente, seulement qu'ici il s'agit des pays avec plusieurs circonscriptions électorales. Elle a trois variantes: restes majeurs, restes moyens et restes mineurs.

#### Metodo di Hare

Parti	Votes	Quotient électoral	Première répartition de siège	Moyenne Majeure	Restes
A	1000	400	2	500	200
B	650	400	1	650	250
C	350	400	-	-	350
Votes valides: 2.000    Nombre de sièges: 5					

*Méthode de Hagenbah-Bischof*: On l'appelle aussi méthode de quotient rectifié. On ajoute un ou deux au diviseur pour réduire le quotient, et dans cette méthode les sièges sont affectés tous à la première répartition.

Votes valides dans la circonscription

$$\text{Quotient} = \frac{\text{Siège dans la circonscription} + 1}{\text{-----}}$$

*Methode de D'Hondt:* On divise le nombre des votes valides de chaque parti par la série de nombres naturels – 1,2,3,4... - ou la série de nombre impaires ou paires. Puis, on attribue les sièges relativement, du quotient majeur au quotient mineur, jusqu'à compléter le nombre des sièges.

Ce système de D'Hondt récompense les partis qui ont reçu le plus grand nombre de votes.

Pour une certaine stabilité et gouvernabilité, certains pays ajoutent d'autres restrictions dans l'affectation des sièges : les partis qui n'atteignent pas un pourcentage minimum de votes (quorum), n'accède pas à la répartition des sièges. Ainsi par exemple en Allemagne le quorum est de 5%, en Espagne de 3%.

Exemple:

PARTIS	DIVIDEURS				
	1	2	3	4	5
A	168.000 (1)	84.000 (3)	56.000 (6)	42.000 (8)	33.600
B	104.000 (2)	52.000 (7)	34.666	26.000	20.800
C	72.000 (4)	36.000	24.000	18.000	14.400
D	64.000 (5)	32.000	21.333	16.000	12.800
E	40.000	20.000	13.333	10.000	8.000
F	32.000	16.000	10.666	8.000	6.400

480.000 Votes totaux — 8 sièges — 60.000 quotient

La répartition des sièges dans ce cas est:

Parti A: 4 sièges: 1,3,6 et 8

Parti B: 2 sièges: 2 et 7

Parti C: 1 siège: 4

Parti D: 1 siège: 5

## Sondage d'opinion publique

### **II.3. Le sondage d'opinion**

Par rapport aux autres instruments de mesure sociale, comme par exemple le panel ou l'expérimentation de laboratoire, le sondage offre certains avantages (Noelle-Neumann, 1993, 39,42) :

1. On observe un échantillon représentatif de la population, tandis que l'étude de laboratoire s'effectue d'habitude avec des sujets atypiques – étudiants, militaires, malades...  
7,
2. Le sondage se fait dans les conditions naturelles et ceci permet de rassembler des aspects décisifs de la vie réelle qui influencent la conduite.

Le panel ou interview aux mêmes sujets d'un échantillon pour mesurer les changements, est une modalité de sondage. Le panel permet un monitoring et intervalles constants des variables.

Définition: Le sondage est une recherche sur un échantillon des sujets qui représentent une grande collectivité, et qui se déroule dans le contexte de la vie quotidienne, avec des moyens standardisés de question, avec le but de relever des mesures quantitatives d'une variété de données objectives et subjectives d'une population. (García Ferrando, 1986, 123). Les éléments en sont donc: l'échantillon, l'univers, la probabilité et l'erreur statistique.

Quelques caractéristiques de sondages :

1. Ils mesurent les opinions et non l'opinion publique. Ils nous donnent les photographies statiques à l'intérieur d'un processus mutable qui ont référence à certaines questions, celles qui sont posées.
2. Ils nous fournissent une connaissance probable, parfois tout simplement indicative.
3. Comme ils sont des moyens standardisés – égaux pour tous les sujets interviewés – les sondages comportent des limites. On ne peut pas être sûr que toutes les personnes interrogées comprennent correctement les questions, qu'elles lisent dans le même ordre qu'elles leur sont présentées, et qu'elles n'imaginent pas les sens non présents dans les questions ...
4. Il convient de leur donner leur juste valeur, ni plus, ni moins. Malgré ceci, il faut dire que les sondages électoraux sont assez précis et très utilisés dans tous les pays. Aux USA par exemple, par deux fois seulement que les sondages se sont trompés : en 1948 Gallup avait pronostiqué la victoire de Dewey, alors que c'est Truman qui a gagné. La faute était due à un mauvais échantillonnage. En 1980, la victoire de Reagan n'était pas prévu par les sondages jusqu'une semaine avant les élections....(Bailey, 1995, ).

Il y a encore aujourd'hui un grand débat sur l'utilisation des sondages. Certains auteurs retiennent que l'abus de cette technique de recherche a amené des désavantages dans le développement de la sociologie.

#### **Répercussions politique et sociale des sondages d'opinion**

- **Perte de l'expérience du réel et désorientation de valeurs.**
- **Désinformation à cause d'une overdose d'informations.**

La perception réelle de la société est confondue avec la perception qui nous provient du travail des médias, engendrant de ce fait une désinformation. Sous l'influence de plusieurs types de médias, les données réelles diffèrent de celles que les personnes croient être certaines. Par exemple, si le nombre de filles qui avortent est de 8% à Kinshasa, les personnes influencées par un programme télévisé croient que ce pourcentage équivaut à 50%. Il y a une expérience immédiate et médiatique à travers les médias.

En ce qui concerne la désinformation par l'overdose d'information, disons qu'il s'agit de l'abondance ou l'excès de l'information qui nous fait croire que nous sommes informés, alors que nous sommes remplis de n'importe quoi, souvent accidentel, qui n'a rien à avoir avec la vraie information.

Mais le thème des effets des médias n'est pas souvent analysé, au contraire :

- La recherche est dérisoire (pauvre).
- L'insuffisance des programmes de recherche.
- Il y a des difficultés objectives:

Les limites naturelles de l'influence des médias: Les médias ont une forte influence sur ce dont nous ne pouvons avoir une expérience directe. Par contre sur nos expériences, ils n'ont pas d'influence significative.

## **II.4. Effets sociopolitiques des sondages**

### *II.4.1. Les effets positifs et négatifs*

#### *Effets positifs*

Il y a pas de doute. Les sondages permettent aux électeurs d'avoir des informations qui dans le passé étaient réservées seulement à l'élite. Ceci augmente leur capacité de décision.

- Les sondages permettent de connaître la position des autres citoyens, pour aider l'électeur à situer sa position ou son opinion dans le contexte social.

- Une accumulation d'informations pour les sociologues et les historiens sur le comportement électoral de la communauté. Ceci facilite les études scientifiques, établit les corrélations entre variables, permet la connaissance de changements dans la sociologie politique...

- Les sondages sont une source alternative d'information et empêchent la manipulation en tant en cas de contraste, permettant ainsi la confrontation d'informations qu'on a.

#### *Effets négatifs*

1. Ils stimulent le pessimisme anti démocratique (Schudson): ils fomentent le déterminisme, fatal pour un régime démocratique, dans le sens que la connaissance à l'avance de ce qui va arriver favorise le désengagement et l'indifférence. Cela ne vaut pas la peine de discuter de rien parce que les choses ne changeront pas. On observe par exemple que dans les campagnes électorales les candidats se concentrent sur les circonscriptions

électorales qui lui donnent l'avantage. Et les moyens de communication ne font plus attention aux candidats qui, dans le *ranking* de prévisions, sont au bas point.

2. Les sondages fomentent le populisme. En effet, l'attention excessive aux résultats des sondages favorise une politique basée sur l'image. Le cours de la politique est inconstant et change à tout moment, en suivant le baromètre de la popularité. Par contre, on peut déduire que les actions du *leadership* peuvent changer la grande partie de l'opinion publique. Ceci démontre combien l'opinion publique est éphémère.

3. La perte de la légitimité du système démocratique : C'est une conséquence de ce qui précède. Les élections ne sont pas le seul moyen de légitimation ou de contrôle démocratique, les sondages encore moins. En effet, un système démocratique se base aussi sur le respect des minorités, sur l'action et la fonction des autres organes par la diversité des thèmes et des décisions. Par exemple. En France pendant le mandat de Pompidou, on appliqua la peine de mort à deux condamnés après avoir su que 67% de la population était favorable à la condamnation capitale. Dans ce cas, on peut dire avec Moles que le fait plus grave est qu'on remplace le dialogue public avec le contrôle des sondages. La démocratie directe n'est toujours la plus profonde et la plus étendue ou ample.

4. Danger d'ingénierie sociale à travers les sondages : Il s'agit du risque de manipuler l'opinion publique en canalisant son attention au moyen d'une sélection de thème (Clemens e David Paletz). On se rappellera la fonction-effet de l'agenda-setting. Les sondages ne nous obligent pas à penser d'une certaine manière, mais sur une manière concrète plus que sur d'autres. Les sondages peuvent devenir un instrument de contrôle pour le pouvoir.

5. Renforcement de position intermédiaire ou de l'effet « spirale du silence ». C'est ce qu'on a affirmé avant sur la prédisposition à adhérer aux opinions majoritaires de ceux qui ont une opinion déjà formée ou à se concentrer sur la position de consensus. Il faut donner un fondement à ces affirmations comme l'a si bien fait Noelle-Neuman, auteur de cette théorie. (Cfr. Phénomène Rush Limbaugh).

6. Enfin, il y a lieu de signaler comme facteur nocif, la banalisation des campagnes électorales. On prête plus attention aux *hitparades* et non aux thèmes en discussion ou aux programmes.

#### *II.4.2. Les limites des sondages entendu comme reflet de la volonté populaire (Galtung)*

- Le sondage est individuel, c'est-à-dire les individus sont isolés au moment de l'interview du contexte réel.

- Il est trop démocratique, c'est-à-dire ne tient compte de plusieurs aspects et différences effectives.

- Il est statique. Il faut bien savoir que les opinions sont dynamiques et qu'elles changent souvent.

- Ils privilégient la position des classes moyennes. Les opinions minoritaires, radicales ou périphériques se perdent dans l'ensemble de l'enquête. Ceci se dénote déjà dans le choix des personnes à interroger.

- Les sondages s'adaptent seulement aux situations normales, harmonieuses. Quand il y a une situation de conflit social, crainte ou violence, le sondage est vu comme un espionnage.

### *II.4.3. Répercussions électorales des sondages politiques*

Voici certaines conclusions des études sur les possibles altérations de l'intention originariaire de vote :

1. “Effet *bandwagon*”: C'est le phénomène d'adhésion au probable vainqueur, qui permet la consolidation de la majorité et qui fait que beaucoup d'indécis choisissent le vainqueur prévu. C'est l'hypothèse du vote utile.
2. L'effet *underdog* (perdant): Si l'effet *bandwagon* mobilise pour le gagnant prévu, l'effet *underdog* mobilise pour le perdant prévu.
3. Effet *David* ou le renversement des prévisions : Parfois il arrive le contraire de l'effet *bandwagon*. Soit par le fait que le vainqueur prévu ne fait plus rien, ou alors les états-majors opposés s'arment de plus bel pour affronter le défi.
4. Compensation de l'effet de l'évolution du vote indépendamment des pronostics. Pour Wert et Lopez Pintor l'intention de vote évolue comme conséquence de la campagne électorale, indépendamment des pronostics.



**LIVRES CONSULTES**

L. Fabris, *L'indagine campionaria. Metodi, disegni e tecniche di campionamento*, Roma, NIS, 1989.

Andrea Cimenti, *Come si fa un sondaggio*, Cuen, Napoli, 1996.

Y. Harvatapoulos, Y.F. Livan, Ph. Sarnin, *L'art de l'enquête. Guide pratique*, Ed. Eyrolles, Paris, 1992 (si trova una traduzione spagnola: *El arte de la encuesta*, Ed. Deusto, Bilbao, 1992).

A. Fink, J. Kosecoff, *How to conduct surveys*, Sage, London, 1985.

Manuel García Ferrando, *Socioestadística. Introducción a la estadística en sociología*. Alianza Universidad, Madrid, 1985, 1995 (caps. 2-5).

Sharon Weiberg e K. Goldberg, *Statistics for the Behavioral Sciences*, Cambridge University Press, 1990 (caps. 1 a 4; 7 a 10).

Kenneth Bailey, *Metodi della ricerca sociale*, Il Mulino, Bologna, 1985, 1995 (caps. 4 e 5).

Vincenzo Cesareo e altri, *La religiosità in Italia*, Mondadori, Milano, 1995, in particolare la nota metodologica, pp. 293-312.

Elisabeth Noelle-Neuman. *The Spiral of Silence. Our Social Skin*. Chicago University Press, 1993.

W. S. Aquilino, *Interview Mode Effects in Surveys of Drug and Alcohol Use. A Field Experiment*, in "Public Opinion Quarterly", 1994, 58, 2, pp. 210-240.

A. Muñoz Alonso e altri, *Opinión pública y comunicación política*, Eudema, Madrid, 1990, pp.400-414.

M. Duverger, *Instituciones políticas y derecho constitucional*, Ariel, Barcelona, 1980 (classico, con molte riedizioni)

D. Nohlen, *Sistemas electorales del mundo*, Centro de Estudios Constitucionales, Madrid, 1981.

*Formes de participation politique*